# COMPRENSIÓN DE CONCEPTOS ELEMENTALES DEL ANÁLISIS DE VARIANZA POR ESTUDIANTES UNIVERSITARIOS

**Osmar Darío Vera**

**TESIS DOCTORAL**

**Directoras: Dra. Carmen Batanero Bernabeau**

**Dra. Carmen Díaz Batanero**

**Granada, 2015**

# RESUMEN

En la actualidad hay un amplio consenso en la necesidad de una formación suficiente en inferencia estadística, que permita al graduado de múltiples especialidades comprender la literatura científica de su área de conocimiento y realizar estudios propios con las debidas garantías.

En este trabajo nos hemos interesado por la comprensión que alcanzan los estudiantes de Psicología sobre los conocimientos elementales de análisis de varianza y sus prerrequisitos, una vez finalizada su enseñanza. Dicho contenido se estudia en la mayoría de las Universidades, aunque ha habido muy pocas investigaciones sobre el grado de comprensión o la capacidad de aplicación de los estudiantes. Este estudio se justifica por el gran número de estudiantes que se enfrentan al aprendizaje de estos conceptos así como por los resultados de la investigación didáctica sobre algunos pre-requisitos para su comprensión que indican la existencia de sesgos y errores que creemos se pueden trasladar y ampliar en el análisis de varianza.

Nos basamos en el enfoque ontosemiótico de la cognición e instrucción matemática, el estudio matemático del tema, las directrices curriculares sobre la materia en la licenciatura de Psicología y las investigaciones previas. A partir de estos fundamentos se presentan tres estudios interrelacionados:

- El Estudio 1 se orienta a la construcción de un cuestionario con suficientes garantías de fiabilidad y validez para la evaluación de la comprensión sus contenidos, siguiendo un proceso metodológico riguroso. Tiene en cuenta la enseñanza recibida por los estudiantes a los que va dirigido, así como el análisis conceptual del objeto de estudio.
- En el Estudio 2 se realiza la evaluación en una muestra de 224 alumnos de la Universidad de Huelva, utilizando el cuestionario construido. Se analizan con detalle las respuestas en cada ítem y la puntuación global. También se realizan estudios de validación y de la fiabilidad y generalizabilidad del instrumento.
- En el Estudio 3, de tipo cualitativo, se realiza un análisis semiótico de las respuestas abiertas a dos problemas, con cinco apartados cada uno. Esto permite completar el estudio de la comprensión del objeto e identificar conflictos semióticos ligados al análisis de varianza como a contraste de hipótesis.

Se finaliza la Memoria con las principales conclusiones sobre los objetivos e hipótesis formuladas y con la descripción de las líneas futuras de investigación abiertas. Se analizan también las principales aportaciones, entre las que destacamos las siguientes; a) la caracterización detallada del significado institucional de referencia del análisis de varianza; b) la construcción de un cuestionario, el análisis detallado de las respuestas al mismo y a los problemas abiertos; c) la clasificación de los conflictos semióticos identificados. Resultados parciales de todas estas contribuciones se han dado a conocer a través de diversas publicaciones que se detallan a lo largo de la Memoria.

# ABSTRACT

Today there is broad consensus on the need for an adequate training in statistical inference, which allows graduate students to understand scientific literature in their area of expertise and to perform their own studies with due guarantee.

In this work we are interested in the understanding that psychology students obtain on elementary analysis of variance and its prerequisites, after having studied the topic. Such content is today included in most universities, although there has been very little research on the degree of understanding and competence for applying the topic by the students. The study is justified by the large number of students following courses of variance analysis, the results of educational research on some prerequisites that indicate the existence of biases and errors that could be translated or expanded to variance analysis.

We rely on the onto-semiotic approach to mathematics cognition and instruction, the mathematical study of the topic, the curricular guidelines for the subject in the Bachelor of Psychology and the previous research. From these foundations three interrelated studies are presented:

- Study 1 is aimed to build a valid and reliable instrument that can assess understanding of these contents. After following a rigorous methodological process we built a questionnaire that takes into account the education received by the students and the conceptual analysis of the study object.
- Study 2 performs an assessment, using the previously built questionnaire with a sample of 224 students of the University of Huelva. We analyse in detail the answers to each item and the overall score. Validation studies and the reliability and generalizability estimation of the instrument are also performed.
- In Study 3, qualitative, semiotic analysis of open responses to two problems, with five sections each is performed. This allows the study of complete understanding of the object and identifies semiotic conflicts linked to analysis of variance as hypothesis testing.

The report with the main conclusions related to the research objectives and hypotheses and the description of new research lines complete the thesis. Main contributions are also analysed, among which we highlight the following; a) a detailed characterization of the institutional reference meaning for the analysis of variance; b) the built questionnaire and the detailed analysis of the responses to it and to the open problems; c) the classification of semiotic conflicts. Partial results of all these contributions have been diffused through various publications which are detailed throughout the report.

# ÍNDICE


INTRODUCCIÓN……..……. 1

1. PROBLEMA DE INVESTIGACIÓN: CONTEXTUALIZACIÓN Y ESPECIFICACIÓN DEL CONTENIDO ……..…… 3

1.1. Introducción……..…… 3

1.2. El objeto de estudio……...…… 4

1.2.1. Relevancia del análisis de varianza en Psicología……..…….. 4

1.2.2. Retos didácticos que plantea…….…… 5

1.3. Contexto educativo…….……..… 6

1.3.1. La formación estadística en Psicología…….…… 6

1.3.2. Asignaturas de estadística en la licenciatura de Psicología en la Universidad de Huelva…….……..… 9

1.3.2.1. La asignatura Análisis de datos en Psicología I…..……..…. 9

1.3.2.2. La asignatura Análisis de datos en Psicología II…..……..… 11

1.4. Objetivos de la investigación…….……..… 12

1.5. Hipótesis iniciales…….……..…. 14

1.6. Etapas en la investigación y características metodológicas generales…..……. 14

1.7. Significado de referencia del análisis de varianza…..……..…. 16

1.7.1. Introducción……..……..…….. 16

1.7.2. Pre-requisitos…….……..… 17

1.7.3. Intervalo de confianza…….……... 22

1.7.4. Contraste de hipótesis…….…… 24

1.7.5. Tipos de variables en análisis de varianza…….…… 29

1.7.6. Análisis de varianza de un factor efectos fijos…….……. 30

1.7.7. Análisis de varianza de varios factores…….…… 34

1.7.8. Análisis de varianza de medidas repetidas…….…….. 36

1.7.9. Supuestos del análisis de varianza…….…….…… 38

1.8. Delimitaciones de los contenidos del estudio…….…… 41

2. FUNDAMENTOS TEÓRICOS……. 45

2.1. Introducción……..……. 45

2.2. Marco teórico…….……. 46

2.2.1. Introducción…….……. 46

2.2.2. Sistemas de prácticas operativas y discursivas ligadas a campos o tipos de problemas…….……..… 46

2.2.3. Objetos intervinientes y emergentes de los sistemas de prácticas…… 48

2.2.4. Relaciones entre objetos: función semiótica……..…… 50

2.2.5. Evaluación de la comprensión……. 51
2.2.6. Idoneidad de un cuestionario de evaluación…… 51
2.3. Antecedentes de nuestra investigación……. 52
2.3.1. Introducción…….. 52
2.3.2. Percepción de la variabilidad aleatoria……. 53
2.3.3. Comprensión de las distribuciones muestrales……. 55
2.3.3.1. Muestreo……. 55
2.3.3.2. Distribución muestral……. 56
2.3.3.3. Teorema central del límite…….. 57
2.3.4. Investigaciones sobre comprensión del intervalo de confianza…… 58
2.3.4.1. Comprensión del intervalo de confianza por los investigadores… 59
2.3.4.2. Evaluación del aprendizaje en cursos universitarios……. 59
2.3.4.3. Trabajos apoyados por el enfoque onto-semiótico……. 62
2.3.5. Investigaciones sobre comprensión del contraste de hipótesis…… 63
2.3.5.1. Controversias sobre el contraste de hipótesis……. 63
2.3.5.2. Errores sobre el nivel de significación y el valor p…… 65
2.3.5.3. Errores sobre el concepto de hipótesis……. 66
2.3.5.4. Otros errores en el contraste…… 67
2.3.6. Investigaciones relacionadas con el diseño experimental y el análisis de la varianza…… 68
2.3.6.1. Tipos de variables…… 68
2.3.6.2. Interacción…….. 69
2.3.6.3. Análisis de pruebas de valoración de la comprensión del diseño experimental…… 70
2.3.6.4. Dificultades causadas por el software…… 70
2.4. Conclusiones del estado de la cuestión……. 71

3. ESTUDIO 1. CONSTRUCCIÓN DE UN CUESTIONARIO DE EVALUACIÓN…… 73
3.1. Introducción……. 73
3.2. Objetivos del Estudio 1……. 74
3.3. Definición de la variable objeto de medición…….. 75
3.3.1. Fundamentos de la definición……. 75
3.3.2. Unidades de contenido semántico ……. 76
3.4. Construcción de un banco de ítems……. 78
3.4.1. Tipo y número de ítems……. 78
3.4.2. Proceso de construcción del banco de ítems……. 79

3.5. Análisis a priori de los ítems de evaluación……. 79
3.5.1.Parte I: Prerrequisitos……. 80
3.5.2.Parte II: Análisis de la Varianza…… 86
3.5.3.Resumen del contenido de los ítems……. 97
3.6. Pruebas piloto de ítems…… 100
3.6.1.Muestra y material…… 100
3.6.2.Método……... 101
3.6.3.Resultados y discusión……..….. 101
3.7. Evaluación de ítems mediante juicio de expertos…….…..….. 107
3.7.1. Expertos participantes……..…..……. 107
3.7.2. Material y método…….…… 108
3.7.3. Resultados y discusión…… 109
3.8. Selección final y revisión de ítems para el cuestionario……..… 111
3.9. Revisión del cuestionario en función del juicio de expertos…… 119
3.10. Conclusiones del Estudio 1…….……. 126

4. ESTUDIO 2. EVALUACIÓN DE CONOCIMIENTOS MEDIANTE CUESTIONARIO……. 129
4.1. Introducción……. 129
4.2. Objetivos del Estudio 2……..……. 130
4.3. Hipótesis del Estudio 2…….……. 130
4.4. Descripción de la muestra…….……. 131
4.5. Material y método……...……. 131
4.6. Resultados detallados por ítem……..…. 135
4.6.1.Intervalo de confianza……. 135
4.6.2.Contraste de hipótesis…….……. 136
4.6.3.Modelo de análisis de varianza. Selección y supuestos……. 141
4.6.4.Cálculo en análisis de varianza…… 144
4.6.5.Interpretación de resultados…….……. 146
4.7. Índice de dificultad de la prueba…….……. 150
4.8. Estudio de la puntuación global…….. 152
4.9. Estudios de validación…….……. 153
4.9.1.Validez de contenido…….…… 154
4.9.2.Fiabilidad de consistencia interna ……. 154
4.9.3.Validez discriminante de los ítems…… 155
4.9.4.Análisis factorial y validez de constructo……..…… 156
4.10. Generalizabilidad…….……..…… 157

4.11. Conclusiones sobre la evaluación del cuestionario…… 159

5. ESTUDIO 3. ANÁLISIS SEMIÓTICO DE RESPUESTAS A PROBLEMAS ABIERTOS…… 163

5.1. Introducción…… 163

5.2. Objetivos e hipótesis del estudio semiótico 164

5.3. Método…… 165

5.4. Análisis a priori de los problemas propuestos…… 167

5.4.1. Análisis a priori del problema 1. Contraste de hipótesis 167

5.4.2. Análisis a priori del problema 2. Análisis de varianzas de medidas repetidas…… 170

5.4.3. Síntesis de contenido evaluado en los problemas abiertos…… 174

5.5. Resultados en el problema 1: Contraste de hipótesis…… 175

5.5.1. Definición de las hipótesis estadísticas…… 175

5.5.2. Elección del contraste…… 182

5.5.3. Cálculo de las regiones y decisión sobre las hipótesis…… 188

5.5.4. Interpretación de resultados…… 193

5.5.5. Cálculo de la potencia del contraste…… 197

5.5.6. Síntesis de resultados en el problema 1…… 203

5.6. Resultados en el problema 2: Análisis de varianza de medidas repetidas…….. 204

5.6.1. Definición de las hipótesis estadísticas…… 204

5.6.2. Variable dependiente e independiente…… 210

5.6.3. Elección de un contraste…… 213

5.6.4. Cálculo de las regiones y decisión sobre las hipótesis…… 217

5.6.5. Interpretación de resultados…… 225

5.6.6. Síntesis de resultados en el problema 2…… 231

5.7. Comparación de resultados en los dos problemas…….. 232

5.8. Conclusiones sobre los problemas abiertos…… 233

6. CONCLUSIONES…… 235

6.1. Introducción…… 235

6.2. Conclusiones respecto a los objetivos de la investigación……. 235

6.3. Conclusiones respecto a las hipótesis iniciales …… 240

6.4. Idoneidad del cuestionario de evaluación…… 241

6.5. Aportaciones y limitaciones del estudio…… 242

6.6. Líneas de investigación abiertas……. 243

REFERENCIAS……... 245

ANEXO 01. Cuestionario de Evaluación……... 253

ANEXO 02. Pruebas Piloto……..……... 259
ANEXO 03. Cuestionario Expertos……... 277

# INTRODUCCIÓN

Las ciencias experimentales, como la Psicología, la Química, o la Medicina, se apoyan en la recogida, análisis e interpretación de datos, para poder obtener nuevo conocimiento, partiendo de teorías previas y de experimentos, que se deben controlar y contrastar. Esto hace que, en dichas ciencias, la actualidad y los métodos estadísticos sean la herramienta metodológica básica para la realización de nuevas investigaciones o para generalizar a diferentes contextos o situaciones resultados ya confirmados.

Todo ello lleva al debate acerca de cuáles son los contenidos estadísticos básicos para la formación de los profesionales de estas ciencias. Un acuerdo generalizado es la necesidad de una formación suficiente en inferencia estadística y en diseño experimental, que permita al graduado en el futuro leer la literatura científica de su área de conocimiento e igualmente recoger con las debidas garantías sus propios datos. Dicha formación le permitirá obtener elementos para discernir entre las falsas teorías y las mejor fundamentadas y le proporcionará bases metodológicas suficientes para utilizar y seleccionar los instrumentos requeridos en su profesión y aplicarlos significativamente.

Una de estas ciencias es la Psicología, donde la formación estadística a nivel de Grado, Máster y Doctorado incluye elementos básicos de inferencia estadística. En este trabajo nos hemos interesado en particular por el análisis de varianza elemental. Dicho contenido es enseñado desde hace tiempo en la mayoría de las Universidades que forman a los psicólogos. Sin embargo, ha habido muy pocas investigaciones sobre el grado de comprensión o la capacidad de aplicación de esta técnica por parte de los estudiantes que se preparan en esta especialidad. El interés de realizar un estudio al respecto se deduce no sólo del gran número de estudiantes que siguen cursos de análisis de varianza, sino de los resultados de la investigación didáctica sobre algunos pre-requisitos para su comprensión que indican la existencia de sesgos y errores (descritos, por ejemplo en Batanero, 2000 y Harradine, Batanero y Rossman, 2011).

Para abordar esta problemática, en esta memoria presentamos tres estudios relacionados orientados a la evaluación de la comprensión adquirida por los estudiantes del grado de Psicología una vez estudiado el tema. La Memoria se organiza en cinco capítulos, con el contenido descrito a continuación.

El Capítulo 1 presenta el problema de investigación, justificando su relevancia. Se describe también el contexto de enseñanza y el marco teórico utilizado que es el enfoque onto- semiótico del conocimiento y la instrucción matemática (Godino, 2002, 2012). Este marco teórico sirve para fijar los objetivos e hipótesis del trabajo, así como para determinar el significado de referencia del análisis de varianza, que será la base de la construcción de los instrumentos de evaluación.

En el Capítulo 2 se incluyen los antecedentes de investigación sobre análisis de varianza y sobre otros temas relacionados con el mismo: distribuciones muestrales y teorema central del límite, intervalo de confianza y contraste de hipótesis.

En el resto de la memora se presentan tres estudios empíricos relacionados: El primero de ellos (descrito en el Capítulo 3) se orienta a la construcción de un cuestionario válido y fiable para la evaluación de estos contenidos en estudiantes del grado de Psicología. La construcción de este cuestionario sigue un proceso metodológico riguroso y tiene en cuenta la enseñanza recibida por los estudiantes a los que va dirigido y el análisis conceptual del análisis de varianza (que se incluye en el Capítulo 1).

En el segundo estudio, descrito en el Capítulo 4, se realiza una evaluación, utilizando el cuestionario construido anteriormente con una muestra de 224 alumnos de la Universidad de Huelva, analizando con detalle las respuestas en cada ítem y estudiando la puntuación global y en ítems relacionados con conceptos importantes en el análisis de la varianza. También se realiza un estudio de fiabilidad y generalizabilidad del instrumento.

Para completar el anterior estudio, de carácter eminentemente cuantitativo, en el Capítulo 5 se incluye un tercer estudio de tipo cualitativo en el que se realiza un análisis semiótico de las respuestas abiertas a dos problemas, cada uno de ellos con cinco apartados. Este análisis permite completar el estudio de la comprensión de los estudiantes sobre el análisis de varianza y el contraste de hipótesis y también identificar conflictos semióticos ligados a objetos matemáticos que intervienen en estos dos procedimientos estadísticos.

Se finaliza la Memoria con las principales conclusiones sobre los objetivos planteados y las hipótesis formuladas y con la descripción de las líneas de investigación abiertas. Se analizan también las principales aportaciones, entre las que destacamos la caracterización detallada del significado institucional de referencia del análisis de varianza. El cuestionario construido, así como el análisis detallado de las respuestas al mismo y a los problemas abiertos; la clasificación de errores en los pasos en la realización de un análisis de varianza y la lista de conflictos semióticos identificados son otras aportaciones relevantes. Resultados parciales de todas estas contribuciones se han dado a conocer a través de diversas publicaciones que se detallan a lo largo de la Memoria.

# CAPÍTULO 1.
# PROBLEMA DE INVESTIGACIÓN: CONTEXTUALIZACIÓN Y ESPECIFICACIÓN DEL CONTENIDO


1.1. Introducción
1.2. El objeto de estudio
1.2.1. Relevancia del análisis de varianza en Psicología
1.2.2. Retos didácticos que plantea
1.3. Contexto educativo
1.3.1. La formación estadística en Psicología
1.3.2. Asignaturas de estadística en la licenciatura de Psicología en la Universidad de Huelva
1.3.2.1. La asignatura Análisis de datos en Psicología I
1.3.2.2. La asignatura Análisis de datos en Psicología II
1.4. Objetivos de la investigación
1.5. Hipótesis iniciales
1.6. Etapas en la investigación y características metodológicas generales
1.7. Significado de referencia del análisis de varianza
1.7.1. Introducción
1.7.2. Pre-requisitos
1.7.3. Intervalo de confianza
1.7.4. Contraste de hipótesis
1.7.5. Tipos de variables en análisis de varianza
1.7.6. Análisis de varianza de un factor, efectos fijos
1.7.7. Análisis de varianza de varios factores
1.7.8. Análisis de varianza de medidas repetidas
1.7.9. Supuestos del análisis de varianza
1.8. Delimitaciones de los contenidos del estudio


## 1.1. INTRODUCCIÓN

Las investigaciones sobre la didáctica de la inferencia estadística tienen hoy en día muy amplio impacto a nivel internacional, como se muestran en los trabajos presentados en las International Conferences on Teaching Statistics (ICOTS) y en Harradine, Batanero y Rossman (2011). En España, la investigación en esta temática fue iniciada por Vallecillos con su tesis doctoral (Vallecillos, 1994), que fundamentó las bases de una amplia investigación posterior, incluyendo algunos trabajos realizados en la Universidad de Granada (Alvarado, 2007; Díaz, 2007; Olivo, 2008; Vera, 2008; Cañadas, 2012). Gran parte de estos trabajos señalan la existencia de errores conceptuales en la interpretación de los conceptos y dificultades de aplicación de la inferencia entre estudiantes e investigadores.

Con ánimo de contribuir a la investigación en didáctica de la estadística en el nivel universitario, siguiendo la línea iniciada con la tesis de Máster (Vera, 2008) y con la intención de profundizar sobre la inferencia estadística, hemos elegido para esta tesis

doctoral el objeto estadístico análisis de varianza. La investigación empírica sobre su comprensión o aplicación es escasa, aunque encontramos algunos trabajos, que incluyen sugerencias didácticas o describen la experiencia de profesores en su enseñanza, así como sobre su uso en la investigación (Green, 2007; Pardo, Garrido, Ruiz y San Martín, 2007).

En este primer capítulo de la Memoria nos dedicamos a contextualizar el problema de estudio y proporcionar los fundamentos necesarios para abordarlo. Se describe el contexto educativo en que se tomaron los datos, se justifica la importancia del tema en la preparación de los estudiantes de Psicología y otras especialidades universitarias. También presentaremos los objetivos de nuestro trabajo, así como las hipótesis iniciales y la organización general de la investigación. Finalmente se analiza, desde el punto de vista matemático, el objeto de estudio para delimitar el significado de referencia en nuestra investigación y fijar los contenidos de los instrumentos de evaluación.

## 1.2. EL OBJETO DE ESTUDIO

El análisis de varianza es una herramienta importante en la investigación para gran número de disciplinas, debido a su amplia aplicación en el diseño de experimentos, el control de variables extrañas y el análisis de los datos, con la finalidad de identificar factores que contribuyen a la variabilidad de las variables de interés (Pardo et al., 2007). La importancia de esta técnica estadística en Psicología no es menor; siendo reflejo de dicha importancia la inclusión de estos contenidos en cualquier plan de estudios de esta titulación. La enseñanza del análisis de varianza y, en general, de la inferencia estadística a alumnos de Psicología plantea especiales problemas didácticos, pues los estudiantes no poseen una base matemática tan amplia como en carreras científicas o técnicas. Estas mismas dificultades se pueden hallar entre estudiantes de Educación, Sociología, Antropología, entre otros, por lo que pensamos que parte de los resultados que encontraremos se pueden extender a estudiantes de otras ramas de las Ciencias Sociales. En lo que sigue desarrollamos con más detalle la relevancia del tema dentro de la Psicología y los retos didácticos que plantea su enseñanza.

### 1.2.1. RELEVANCIA DEL ANÁLISIS DE LA VARIANZA EN PSICOLOGÍA

En los estudios de Grado de Psicología, generalmente se incluye el análisis de varianza como parte de un segundo curso de estadística, dirigido a alumnos que han finalizado un primer curso de introducción a los métodos estadísticos. También suele ser un tema que se imparte en cursos de postgrado o de formación profesional**,** no sólo en Psicología, sino a otros profesionales de las ciencias, ciencias humanas y sociales o ingenierías; así como para técnicos en ejercicio y para investigadores en una amplia gama de disciplinas.

El motivo de la enseñanza tan amplia de este tema, es que constituye una herramienta imprescindible en el trabajo de científicos y profesionales. Los conceptos y procedimientos incluidos en el análisis de varianza son fundamentales en las diferentes etapas del proceso de investigación:

a. En el diseño de la investigación, donde el investigador ha de decidir qué variables introducirá en su estudio y el papel que jugarán en el mismo (dependiente, independiente, extraña). Se resuelve, por tanto, el nivel de medición de la variable como la forma en que se clasificará;

b. En el análisis de los datos, donde se debe concretar el tipo de análisis de varianza, modelo, número, tipos de factores y sus niveles; comprobar que se cumplen los supuestos de aplicación y realizar los cálculos adecuadamente y;
c. En la toma de decisiones a partir de los resultados del análisis, el investigador debe interpretar correctamente los resultados y sus limitaciones.

La necesidad e importancia de este método ha ido aumentando durante los últimos años dentro de las Ciencias Sociales y, más concretamente, en la Psicología. Por ejemplo, Livacic-Rojas, Vallejo y Fernández (2006) indican la importancia de implementar un diseño ajustado a los objetivos de la investigación, así como a la complejidad de este tema de estudio. Apuntan que, tanto los diseños de medidas repetidas, como de medidas parcialmente repetidas han mostrado ser una solución en ámbitos experimentales para evaluar programas de intervención en Psicología, Educación y Ciencias de la Salud. Por otro lado, el avance de la investigación dentro de la Psicometría confirma día a día que las diferencias individuales existen y son medibles, por lo que el psicólogo que investiga en este campo debe conocer los modelos multivariados incluyendo el análisis de varianza (Koan, 2001).

También la Psicología clínica exige ya un dominio bastante profundo de estas técnicas, pues el psicólogo en ejercicio puede usarlas en el diseño y desarrollo de experiencias clínicas. Entre las ventajas que justifican el uso tan extendido de los modelos factoriales de análisis de varianza, qutitalizá la más destacable sea que también ofrecen la posibilidad de estudiar el efecto de la interacción entre factores (Green, 2007; Pardo et al., 2007). No es sin embargo suficiente que el psicólogo sepa aplicar mecánicamente unas fórmulas, sino que sería apropiado que conozca el fundamento y la deducción de las mismas, así como las condiciones que exigen las técnicas estadísticas en cada caso concreto. El objetivo de nuestro trabajo es precisamente evaluar si durante la formación de los Graduados en Psicología se alcanzan estas competencias.

### 1.2.2. RETOS DIDÁCTICOS QUE PLANTEA

A pesar de la importancia del tema, la enseñanza del análisis de varianza no está exenta de dificultades, como han señalado algunos autores. Por ejemplo, Pardo *et al*. (2007) analizan 150 artículos de investigación, observando un alto porcentaje de ellos donde se llega a conclusiones incorrectas cuando se usa el análisis de varianza con varios factores. El principal error encontrado fue no analizar la interacción.

En el mismo sentido, Castro Sotos, Vanhoof, Van den Noortgate y Onghena (2007) presentan una revisión sistemática de publicaciones que aportan evidencia empírica de interpretación incorrecta de técnicas inferenciales. Los autores señalan errores relacionados con: no tener en cuenta el paralelismo entre un contraste de hipótesis y los procesos de decisión; confusión entre la definición de hipótesis nula y alternativa; creencia que una hipótesis puede referirse tanto a la población como a la muestra o interpretación equívoca del *p*-valor, entre otros.

Kurji (2002) examina algunos de los problemas que plantea la enseñanza de la estadística a nivel universitario en un curso en el cual se incluyen conceptos de diseño y análisis de experimentos. Apunta que los estudiantes de su investigación tuvieron grandes dificultades para integrar conceptos estadísticos en sus proyectos finales de investigación. Propone un cambio en la enseñanza de la estadística en los dos niveles (grado y posgrado), enfatizando el proceso de planificación y ejecución de proyectos de

investigación. Sugiere enseñar los conceptos básicos poniendo énfasis en la comprensión y la forma de interpretar los resultados más que en el cálculo estadístico. También se alentaría a los estudiantes a discutir cuestiones entre ellos, hacer presentaciones tanto orales como escritas, y que sean evaluados por sus pares.

Como consecuencia, y puesto que algunas investigaciones citadas se han centrado con investigadores (que podrían haber olvidado lo que aprendieron en sus estudios), creemos necesario evaluar la comprensión alcanzada por los estudiantes al finalizar el estudio del tema. Nos centraremos en los métodos elementales y conceptos necesitados como prerrequisito, que estos estudiantes han estudiado con anterioridad. La finalidad es comprobar cuáles de los objetos y procedimientos ligados al análisis de varianza son bien comprendidos y cuáles plantean dificultades a los estudiantes. Esto sería un primer paso para revisar y mejorar la enseñanza del tema.

## 1.3. CONTEXTO EDUCATIVO

Nuestro estudio se ha llevado a cabo en un contexto educativo específico. Los datos de evaluación se tomaron en estudiantes de Psicología en la Universidad de Huelva. Para que se pueda valorar mejor la posible aplicación del cuestionario o de los resultados de evaluación a otros estudiantes, a continuación se analiza con detalle el contexto de la investigación.

### 1.3.1. LA FORMACIÓN ESTADÍSTICA EN PSICOLOGÍA

En lo que sigue analizamos, en primer lugar, la formación estadística de los estudiantes de Psicología que formaron parte de la muestra, y seguidamente la formación concreta que recibieron en el tema de investigación elegido. El grupo de estudiantes participantes cursaron la Licenciatura en Psicología (plan de estudio anterior al actual Grado de Psicología).

La formación estadística para estos estudiantes debe permitir al psicólogo analizar los datos que genere en su práctica profesional, y comprender los resultados en los documentos publicados en las distintas áreas de la Psicología. Carpintero (1990) indica que, el psicólogo necesita conocer y diferenciar las características numéricas y métricas de los instrumentos que maneja para utilizar la estadística de manera adecuada. Requiere distinguir el tipo de escala de medida y relaciones entre las variables además de conocer la lógica general de la investigación. También precisa comprender qué es y qué no es estadísticamente significativo y la manera en que esto puede variar de acuerdo con el tamaño de las muestras o con el uso de una prueba con mayor potencia.

Santisteban (1990) analizó distintos programas de las facultades de Psicología en el plan de estudio vigente en esa fecha en las universidades españolas, afirmando que, en los primeros dos años de curso, la mayoría de las universidades españolas ofrecían en aquél momento al menos dos cursos de Estadística. En cuanto a los temas que se trataban en la Licenciatura, en el primero de los cursos de Estadística se incluía la estadística descriptiva y el cálculo de probabilidades, aplicados al área de Psicología. Los contenidos que se incluían eran: tablas y gráficos estadísticos; estadísticos de posición, dispersión y forma; distribuciones de frecuencias; correlación lineal simple; regresión lineal simple y probabilidad.

En el segundo curso el temario, se orientaba a la estadística inferencial, centrada en la propuesta de Fisher y Neyman y Pearson. Los programas más completos incluían: introducción al muestreo; estimación de parámetros (puntual y mediante intervalos de

confianza); contrastes paramétricos de hipótesis (medias, varianzas proporciones), correlación y regresión; pruebas de independencia y de bondad de ajuste; introducción al análisis de la varianza y de la covarianza y técnicas no paramétricas.

Aunque la enseñanza de la inferencia en Psicología se ha realizado con enfoque clásico, se encuentra falta de perspectiva bayesiana. Díaz y Batanero (2006) sugieren la necesidad de que en los cursos de pregrado de Psicología se introduzcan algunas ideas de inferencia bayesiana, junto con los de inferencia clásica y proponen los siguientes contenidos:

- Conceptos básicos fundamentales: Población, parámetro, muestra estadístico, función de verosimilitud, distribución inicial y final.
- Estimación puntual: Métodos clásicos y bayesianos.
- Estimación por intervalos: Intervalos de confianza y de credibilidad.
- Contrastes de hipótesis: Pruebas clásicas y bayesianos, problemas de decisión múltiple.

Por otro lado, Santisteban (1990) especificó que un elemento común en las universidades era *"...Introducir a los psicólogos en el manejo y aplicación de paquetes de programas estandarizados con objeto de que analicen sus datos y sepan interpretarlos…*" (p. 469). Sobre este punto, Díaz y Batanero (2006) indican que debe tenerse cuidado de no caer en el supuesto que el ordenador resuelve la enseñanza de la Estadística.

Actualmente son muchas las publicaciones que analizan la enseñanza de estadística en Psicología. Un ejemplo es el texto de Dunn, Smith, y Beins, (2012) que recoge aspectos como la evaluación, elección de buenos ejemplos, colaboración entre alumnos y profesores, actividades y enfoques innovadores para este tema. Igualmente los libros, como el de Aron, Coups y Aron (2012) que tratan de aproximarse a la estadística con un enfoque práctico e intuitivo, incluyen entre los temas el de análisis de varianza en que centramos nuestra investigación.

**Los actuales estudios de Grado**

El interés por la formación estadística del psicólogo se mantiene en los estudios de Grado. Como se indica en el libro blanco del título de grado (ANECA, 2005) una de las competencias consideradas necesarias en la mayoría de las universidades europeas en estos estudios es el conocimiento de metodología de investigación y técnicas estadísticas. Asimismo, en dicho documento, se recogen, dentro de las competencias básicas a alcanzar, entre otras las siguientes, que están relacionadas con nuestro tema de investigación:

- Conocer distintos diseños de investigación, los procedimientos de formulación y contrastación de hipótesis y la interpretación de los resultados (p. 81).
- Ser capaz de medir y obtener datos relevantes para la evaluación de las intervenciones (p. 82).
- Saber analizar e interpretar los resultados de la evaluación (p. 82)
- Capacidad de gestión de la información (p. 83).

Se incluye específicamente un bloque de contenidos formativos de carácter

obligatorio sobre *Métodos, diseños y técnicas de investigación en Psicología* con una carga lectiva de 18 créditos, o 450 horas-trabajo del alumno). En sus contenidos formativos mínimos se incluyen los métodos cuantitativos y cualitativos. Diseños de investigación. Técnicas de Análisis de datos. Fundamentos científicos y distintos diseños de investigación.

**Análisis de varianza**

El análisis de varianza se ha incluido generalmente en la formación del psicólogo en la Licenciatura y se mantiene en los estudios de Grado. Estas asignaturas pueden variar un poco, dependiendo de la institución. Seguidamente, a modo de ejemplo, analizamos los conceptos incluidos del tema en algunas universidades españolas en asignaturas de carácter obligatorio; todas ellas precedidas de un primer curso de análisis estadístico (con diversas denominaciones) de 6 créditos, en que se estudia los fundamentos de estadística descriptiva, cálculo de probabilidades e iniciación a la inferencia.

En la Universidad Nacional a Distancia, dentro de la asignatura "Diseño de Investigación y Análisis de Datos" (segundo curso, 6 créditos), se incluyen los siguientes contenidos

- Introducción a la estadística inferencial y al contraste de hipótesis en los diseños de investigación.
- Análisis de datos en diseños de un grupo.
- Análisis de datos en diseños de dos grupos independientes.
- Análisis de datos en diseños de dos grupos relacionados.
- Análisis de datos en diseños de más de dos grupos independientes (un factor).
- Análisis de datos en diseños intrasujetos.
- Análisis de datos en diseños de más de dos grupos independientes (dos factores).

La Universidad de Salamanca, en la asignatura "Estadística aplicada a la Psicología II" (Segundo curso, 6 créditos), se incluyen, entre otros, los siguientes contenidos

- Principios generales de la inferencia estadística: conceptos básicos; Muestreo, Distribuciones teóricas, Distribución muestral;
- Estimación de parámetros; Contrastes de Hipótesis; Errores y Potencia;
- Análisis de varianza (ANOVA);
- Análisis de varianza bifactorial.

Contenidos sobre diseños experimentales con grupos diferentes y con los mismos grupos y diseños factoriales se incluyen en la asignatura "Diseño y análisis de datos en Psicología" (6 créditos, segundo curso) en la Universidad a Distancia de Madrid. En la Universidad Abierta de Cataluña, dentro de la asignatura "Modelos estadísticos y Psicométricos" (6 créditos) se estudian el análisis de varianza unifactorial, su lógica y modelos, contrastes a priori y posteriori y supuestos; contrastes factoriales, interacción, análisis de efectos simples y compuestos y diseños complejos.

En resumen, el análisis de varianza suele considerarse un tema necesario para la formación del psicólogo en los nuevos Grados, igual que se consideraba en las antiguas Licenciaturas. En lo que sigue describimos con detalle las asignaturas que han cursado los estudiantes de la muestra que cursaron sus estudios en la Universidad de Huelva. En el periodo en que se recogieron los datos, estaba vigente, como se ha indicado, los estudios de Licenciatura.

## 1.3.2. ASIGNATURAS DE ESTADÍSTICA EN LA LICENCIATURA DE PSICOLOGÍA EN LA UNIVERSIDAD DE HUELVA

### 1.3.2.1. LA ASIGNATURA ANÁLISIS DE DATOS EN PSICOLOGÍA I

La asignatura Análisis de Datos en Psicología I de la Licenciatura en Psicología en la Universidad de Huelva era una asignatura troncal de 6 créditos (3 créditos teóricos, impartidos en grupos de hasta 100 alumnos) y 3 prácticos, impartidos en grupos pequeños de 30 alumnos. Cuando se realizó el estudio se estaba participando en el Plan Piloto de Experimentación de ECTS, y contaba con aproximadamente 220 alumnos distribuidos en dos turnos: mañana y tarde. Este Plan Piloto ensayaba la forma en que se trabaja actualmente en los estudios de Grado; por tanto, constituía un periodo de transición hacia los nuevos grados.

Era una del conjunto de asignaturas de metodología científica de investigación psicológica que debía cursar el alumno, quienes durante el primer cuatrimestre cursaron la asignatura "Fundamentos metodológicos en Psicología". En ella se estudiaron los conceptos básicos sobre conocimiento científico así como las diferentes metodologías de investigación (observacional, de encuesta, experimental y cuasi-experimental). Por tanto, al iniciar "Análisis de datos en Psicología I" contaban con ciertas nociones del esquema general de una investigación.

#### Objetivos

El objetivo principal de la asignatura era introducir al alumno en el uso de algunas de las herramientas conceptuales y metodológicas necesarias para el análisis de los resultados de investigaciones psicológicas. Éste se concretaba en los siguientes objetivos específicos.

1. Conocer y comprender los conceptos básicos de análisis de datos.
2. Conocer algunas técnicas de análisis estadístico y su adecuación en función del tipo de datos recogidos y de la naturaleza del problema científico planteado.
3. Saber interpretar correctamente los resultados de diversas técnicas de análisis.

#### Competencias

Las competencias, genéricas y específicas que se esperaba del alumno eran las siguientes:

- Competencias genéricas o transversales: habilidades de investigación, motivación para el aprendizaje y manejo de herramientas informáticas.
- Competencias específicas, divididas en cognitivas, procedimentales y actitudinales. Entre las cognitivas encontramos: capacidad para delimitar el problema de

investigación e interpretar resultados y generalizarlos. Entre las procedimentales: contribuir al trabajo en equipo, manejo de informática al nivel de usuario, planificar conceptualmente un discurso y trasladarlo a un texto en un formato adecuado. Por último, entre las actitudinales, se destacan: estar motivado por el trabajo, interés por el aprendizaje y por la investigación en el área.

**Contenidos**

El programa de la asignatura constaba de los siguientes temas:

1. El análisis de datos en el proceso de investigación. Conceptos básicos: variables, muestreo, escalas de medida, población y muestra, diseños.
2. Distribuciones de frecuencias y representaciones gráficas (diagramas de barras, sectores, histogramas, tallo y hojas, diagrama de caja).
3. Medidas de tendencia central, posición, dispersión y forma.
4. Puntuaciones individuales y curva normal.
5. Estadística descriptiva bivariable: análisis de la relación entre dos variables categóricas. Estadísticos de asociación.
6. Estadística descriptiva bivariable I: análisis de la relación entre una variable categórica y una variable cuantitativa.
7. Estadística descriptiva bivariable II: análisis de la relación entre dos variables cuantitativas. Covarianza, correlación y regresión lineal simple.

**Metodología**

En las clases teóricas, el profesor realizaba una exposición del contenido y se resolvían problemas de aplicación para potenciar el aprendizaje. Las clases prácticas se realizaban en las aulas de informática. En cada práctica el profesor exponía inicialmente los objetivos de la práctica, así como las tareas a realizar en la misma. Los alumnos usaban el programa SPSS para resolver las tareas y algunos recursos disponibles en internet (tutoriales, programas java para demostraciones, etc.). El programa de prácticas previsto incluía los siguientes bloques:

1. Introducción al paquete estadístico SPSS.
2. Matrices de datos. Definir variables, introducir datos, editar datos, importar/exportar archivos, crear nuevas variables.
3. Análisis estadístico descriptivo univariable y bivariable. Estadísticos, gráficos y tablas univariables. Estadísticos, gráficos y tablas bivariables. Correlación, regresión lineal simple, tablas de contingencia, diagramas de dispersión y diagramas de barras múltiple.

**Evaluación**

La calificación final en la asignatura era la suma de las puntuaciones obtenidas en el examen final y en otras actividades realizadas en grupo. El examen final de la asignatura constaba de dos partes: (a) preguntas con alternativas de respuesta donde se evalúan los conceptos desarrollados durante el curso; (b) problemas abiertos a resolver.

### 1.3.2.2. LA ASIGNATURA ANÁLISIS DE DATOS EN PSICOLOGÍA II

La asignatura Análisis de Datos en Psicología II se ubicaba en el segundo curso y primer cuatrimestre y era continuación de "Análisis de Datos en Psicología I", siguiendo el mismo enfoque. Esta asignatura era también troncal de 6 créditos, 3 teóricos en grupo grande y 3 créditos prácticos en grupos pequeños. La asignatura contaba con aproximadamente 190 alumnos distribuidos en dos turnos, uno de mañana y otro de tarde.

#### Objetivos

El objetivo principal era afianzar los conocimientos sobre análisis de datos y metodología de investigación que el alumno obtuvo en el curso anterior, así como adquirir nuevas habilidades que le permitiesen el análisis detallado de los datos que se recogen en Psicología. Este objetivo se concretaba a través de los siguientes objetivos específicos:

1. Ampliar/completar los conocimientos que el alumno posee sobre procedimientos de inferencia en modelos estadísticos paramétricos.
2. Comprender los conceptos básicos y la aplicabilidad del modelo lineal general.
3. Saber identificar la prueba estadística más adecuada para los objetivos de la investigación y en correspondencia con la naturaleza de los datos.
4. Saber aplicar distintas pruebas y análisis; además analizar los resultados obtenidos.

#### Competencias

Las competencias definidas para la asignatura eran:

- Competencias genéricas o transversales: similares a las de la asignatura Análisis de Datos en Psicología I.
- Competencias específicas. Entre las cognitivas encontramos: conocer los conceptos de probabilidad e inferencia estadística más usuales en Psicología, conocer y comprender la lógica subyacente a la inferencia estadística, comprender las nociones básicas y la aplicabilidad del modelo lineal general. Entre las competencias procedimentales: (a) aprender a establecer e interpretar las relaciones entre variables y ejercitar el razonamiento matemático-estadístico en Psicología; (b) adquirir capacidad para realizar una investigación; (c) identificar la prueba estadística más adecuada a los objetivos de investigación y naturaleza de los datos, aplicarla e interpretarla; (e) valorar críticamente las pruebas y análisis utilizados en distintas publicaciones de carácter científico. Por último entre las actitudinales: capacidad de valorar y discutir el propio trabajo, y desarrollar habilidades para dirigir y coordinar trabajos en equipo.

#### Contenidos

El programa de la asignatura constaba de los siguientes temas:

1. Introducción: conceptos básicos de la inferencia estadística: población, muestra, parámetro y estadístico; distribución muestral; estimación de parámetros: estimación puntual y estimación por intervalos.

2. Contraste de hipótesis sobre medias y proporciones (muestras independientes y relacionadas).
3. Análisis de la varianza de un factor completamente aleatorizado y de medidas repetidas.
4. Análisis de la varianza de dos o más factores.
5. Correlación y Regresión lineal simple.

**Metodología y evaluación**

La metodología y sistema de evaluación son similares a la de la asignatura Análisis de Datos I. También contempla una parte de trabajo autónomo, donde se pide a los alumnos la ejecución de dos trabajos durante el curso, tales como:

1. Lectura reflexiva y comentario crítico de un artículo de investigación publicado en una revista científica del área.
2. Desarrollo de una pequeña investigación en grupos de cuatro o cinco alumnos. El profesor los guiará en cuestiones relativas a la formulación de hipótesis, construcción de la base de datos en SPSS, tratamiento y análisis de los mismos. Al final del curso el alumno elaboraba un informe del trabajo en forma de artículo según las normas APA.

## 1.4. OBJETIVOS DE LA INVESTIGACIÓN

Tal como se ha mencionado, en esta tesis continuamos nuestro trabajo de fin de Máster (Vera, 2008). En la misma línea del trabajo citado, esta investigación *se orienta a realizar un estudio de evaluación de la comprensión que muestran los estudiantes de Psicología sobre algunos objetos básicos del análisis de varianza y su aplicación en la resolución de problemas relacionados.*

Nuestro propósito es contribuir a la mejora de la enseñanza de este tema en Psicología y otras carreras relacionadas con las Ciencias Sociales y por tanto a la mejor aplicación de la estadística en estas áreas. La importancia del tema ha sido debidamente justificada.

Este objetivo general requiere previamente la construcción de un instrumento de evaluación válido y fiable, por lo que, a su vez se descompone en otros específicos que presentamos a continuación. Para ello, seguimos el método de trabajo en otras tesis doctorales presentadas en nuestro equipo de investigación, tales como las de Alvarado (2007) sobre el Teorema Central del Límite; Díaz (2007) en relación a la probabilidad condicional; y Olivo (2008) sobre los intervalos de confianza.

*Objetivo 1: Realizar un análisis de los objetos matemáticos básicos que intervienen en el análisis de varianza que nos fundamente la definición semántica de la variable objeto de medición, delimitando las unidades de contenido semántico de dicha variable.*

Como argumentamos posteriormente la "comprensión del análisis de varianza" es un constructo psicológico y su caracterización ha de hacerse a través de las respuestas de los estudiantes a tareas cuyas respuestas proporcionen indicadores empíricos de dicha comprensión. El primer paso esencial en la construcción de dicho instrumento es

la delimitación de las unidades de contenido que han de estar suficientemente fundamentadas para asegurar posteriormente la validez de contenido (Barbero, 2003; Morales, 2011). Este fundamento se obtendrá, por un lado, a partir de la revisión (ya realizada) de las asignaturas de análisis de datos cursadas por los participantes. Además, en la Sección 1.7, se realizará un análisis epistémico del tema, a partir de textos de Estadística dirigidos a estudiantes de Psicología y Ciencias Sociales. Con ello determinaremos el *significado de referencia* del estudio comparando posteriormente con el significado adquirido por los estudiantes.

*Objetivo 2: Construir un instrumento para evaluar la comprensión de los elementos básicos del análisis de la varianza, siguiendo unos criterios metodológicos rigurosos en la selección de los ítems que lo compongan.*

Una vez delimitado el contenido objeto de evaluación se comenzará un proceso de construcción de un instrumento (Estudio 1) compuesto de dos partes: (a) un cuestionario formado por ítems de opción múltiples que permitan evaluar en un tiempo limitado la mayor cantidad posible de elementos de significado del análisis de varianza elemental; y (b) algunos problemas abiertos que permitan profundizar mejor en los razonamientos y la comprensión de los estudiantes.

El proceso de construcción del cuestionario se describe en el Capítulo 3 y tiene también en cuenta el estudio de los antecedentes (Capítulo 2) que permite, por un lado, seleccionar algunos de los ítems del cuestionario y por otro, realizar un análisis a priori de las posibles dificultades esperadas en los estudiantes. Como parte del proceso será necesario construir un banco de ítems de acuerdo con normas psicométricas, realizar pruebas empíricas de dichos ítems en muestras de estudiantes de Psicología y evaluarlos mediante un panel de expertos. Con todo ello se finalizará con la selección de los ítems y elaboración del cuestionario.

*Objetivo 3: Llevar a cabo un estudio de evaluación con el citado cuestionario en una muestra de estudiantes de Psicología.*

Como se ha indicado, esta evaluación, tendrá dos partes diferenciadas. En primer lugar, mediante el cuestionario previamente elaborado, se lleva a cabo un estudio cuantitativo de la comprensión de una muestra de estudiantes sobre un amplio conjunto de ítems que cubren los diferentes contenidos incluidos (Estudio 2). Aunque de una forma más superficial, este estudio (descrito en el Capítulo 4) permite informar sobre algunos errores y dificultades de los estudiantes en relación al tema. Igualmente permite informar de las características psicométricas del cuestionario: dificultad, discriminación, fiabilidad y validez.

En el Capítulo 5 se lleva a cabo un análisis semiótico de las respuestas a problemas abiertos realizados por los mismos estudiantes (Estudio 3). Ello nos permite evaluar de una forma más profunda el razonamiento y la comprensión de los estudiantes sobre elementos básicos del análisis de varianza. Utilizaremos el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores para identificar los conflictos semióticos.

## 1.5. HIPÓTESIS INICIALES

En este apartado tratamos de formular, con carácter provisional, algunas hipótesis sobre lo que esperamos encontrar. Puesto que apenas hay investigaciones sobre las dificultades de los estudiantes en el análisis de varianza, estas hipótesis deben entenderse como conjeturas o expectativas ya que nuestro estudio es exploratorio (no confirmatorio).

> Hipótesis 1. *Esperamos encontrar dificultades en algunos pre-requisitos del tema, en particular, en la interpretación de intervalos de confianza, el nivel de significación, la asignación de hipótesis en un contraste y el cálculo de la potencia.*

Esta hipótesis se fundamenta en los estudios de los antecedentes, en los cuáles se han descripto estas dificultades en otros estudiantes. Por ejemplo en Olivo (2008) en relación a los intervalos de confianza y Vallecillos (1994), Díaz, Batanero y Wilhelmi (2008) o Harradine, Batanero y Rossman (2011) en relación con varios objetos que intervienen en el contraste de hipótesis.

> Hipótesis 2. *Se prevé que también existan dificultades en relación con el análisis de varianza, por ejemplo, en la selección del modelo, la interpretación de la tabla o salida de software y la comprensión de los supuestos.*

Aunque no tenemos antecedentes en este punto, pero puesto que el análisis de varianza es un tipo particular de contraste de hipótesis, pensamos se reproducirán algunos problemas descritos respecto al contraste de hipótesis, por ejemplo en las investigaciones de Vallecillos (1994) y Cañadas (2012).

> Hipótesis 3. *El análisis semiótico de problemas abiertos permitirá explicar parte de las dificultades anteriores en términos de conflictos semióticos.*

Esta hipótesis se apoya en la experiencia previa obtenida por otros miembros del equipo de investigación, como Alvarado (2007), Olivo (2008), y Cañadas, Batanero, Díaz y Roa (2012). Estos últimos realizan un análisis de las respuestas abiertas a un problema que les permite identificar conflictos semióticos relacionados con el contraste de hipótesis.

## 1.6. ETAPAS DE LA INVESTIGACIÓN Y CARACTERÍSTICAS METODOLÓGICAS GENERALES

Nuestro trabajo se compone de una primera parte, donde se exponen sus fundamentos, y una segunda que recoge los estudios empíricos realizados (ver Figura 1.6.1).

La investigación se compone de una parte cuantitativa (especialmente en los Estudios 1 y 2) y otra cualitativa. Según las clasificaciones de Bisquerra (1989) se trata de una investigación aplicada y descriptiva. Aunque es esencialmente exploratoria, se incluyen algunos elementos confirmatorios a través del contraste de algunas hipótesis.

Figura 1.6.1. Organización de la investigación

La parte de fundamentación a su vez se descompone en los siguientes apartados:

- *Análisis del contenido matemático*, que nos permite describir el significado institucional de referencia en nuestra investigación. Se expone en este mismo capítulo y consiste en el análisis detallado, desde el punto de vista elemental del objeto matemático análisis de varianza y sus prerrequisitos.
- *Análisis curricular*, presentado en este mismo capítulo, que ofrece una perspectiva de las asignaturas cursadas por los estudiantes participantes en los Estudios 2 y 3 y valorar mejor sus conocimientos.
- *Marco teórico*, que se presenta en el Capítulo 2. Utilizaremos para este trabajo algunos elementos del enfoque onto-semiótico desarrollados en por Godino y sus colaboradores (Godino, 1996; 2002; 2003; 2012).
- *Antecedentes de la investigación* que nos permite relacionar nuestro trabajo con otros anteriores. Tanto el marco teórico como los antecedentes se presentan en el Capítulo 2.

La parte empírica del trabajo se organiza en tres estudios independientes, descritos en la introducción. El primero de ellos consiste en la construcción de un cuestionario formado por ítems de opción múltiple. Se utiliza para ello la metodología usual recomendada por APA, AERA y NCME (1999) para reforzar su validez de contenido, su fiabilidad y generalizabilidad. Es un instrumento de medición que, mediante las preguntas planteadas, proporciona una estimación de conocimientos y capacidades de los sujetos, que no son accesibles por simple observación o encuesta (Barbero, 2003).

La definición semántica de la variable objeto de medición se apoya en

fundamentos adecuados para definir "el conocimiento sobre el análisis de varianza", que es un constructo inobservable (León y Montero, 2002). Se basó, en primer lugar, en la revisión de las investigaciones previas, que permitió conocer los errores relacionados con nuestro objeto de estudio que han sido investigados. Una definición más precisa de la variable se basa en el significado de referencia incluido en este capítulo. Para construir el cuestionario, se comienza por la elaboración de una lista de unidades de contenido de la variable objeto de medición. Se sigue con la recopilación de un banco de ítems, se realizan juicios de expertos y pruebas piloto de los ítems. Con todo ello se seleccionan los ítems del cuestionario definitivo y una primera determinación de las características psicométricas de los mismos.

En el Estudio 2 se lleva a cabo un estudio de evaluación con el cuestionario, en un total de 224 alumnos de segundo año de la Licenciatura en Psicología de la Universidad de Huelva. Aunque la muestra es intencional, su amplitud permite que los resultados puedan ser generalizables a otros estudiantes de similares características. Se realiza un estudio detallado por ítems, analizando los distractores, así como los índices de dificultad y discriminación. Asimismo se analizan las puntuaciones totales y se estudian las características psicométricas del instrumento. Para ello se hace una aproximación desde la teoría de la Generalizabilidad (Martínez Arias, 1995), calculando un coeficiente de generalizabilidad, tanto para ítems como para sujetos, atendiendo, de este modo, a estas dos posibles fuentes de error en el muestreo.

En el Estudio 3 se analizan dos problemas abiertos respondidos por la misma muestra de estudiantes. Para ello se emplea el análisis semiótico propuesto por Godino (2003,), que es un método de análisis cualitativo, empleado en trabajos previos, por ejemplo por Cañadas (2012). Nos apoyamos en el enfoque onto-semiótico pues se consideran como categorías primarias los tipos de objetos elementales considerados en este enfoque. Estas categorías se desarrollan posteriormente en forma inductiva utilizando la técnica de análisis de contenido.

## 1.7. SIGNIFICADO DE REFERENCIA DEL ANÁLISIS DE VARIANZA

### 1.7.1. INTRODUCCIÓN

Como se expondrá en el Capítulo 2, el marco teórico que fundamenta nuestra investigación es el onto-semiótico de la cognición matemática, que Godino y colaboradores vienen desarrollando desde 1994. Dentro de este enfoque es importante fijar un significado de referencia del objeto de estudio, ya que se admite una diversidad de significados institucionales y personales para los objetos matemáticos.

En esta sección realizamos un análisis del objeto matemático "análisis de varianza" que nos posibilitará definir el significado de referencia del mismo en nuestra investigación. Ello permitirá delimitar mejor los contenidos concretos en que se centra nuestro estudio, y será el fundamento: en la definición semántica de la variable objeto de medición y para la elaboración del cuestionario de evaluación que describiremos en el Capítulo 3. También lo usaremos en la discusión del análisis de las respuestas a dicho cuestionario (Capítulo 4) así como en el análisis semiótico realizado sobre los problemas abiertos presentados en el Capítulo 5.

El análisis de varianza es la base del diseño experimental o metodología estadística orientada a la planificación y análisis de un "experimento". La palabra "experimento" se refiere a la concepción y realización de ensayos que verifiquen la

validez de las hipótesis establecidas sobre las causas de un determinado problema, o que afectan a una cierta variable objeto de estudio (Manzano, 1997; Pagano, 2011). El diseño de experimentos está basado en tres principios básicos: (a) asignación aleatoria de los individuos a las diferentes condiciones en el experimento, de forma que se puedan determinar efectos que producen rechazando explicaciones alternativas; (b) replicación, o repetición del experimento en varios individuos para poder estimar el error experimental y (c) control de variables extrañas, que permite reducir la variabilidad del error (Ferré y Rius, 2008).

Como auxiliar de estos diseños, el objetivo del análisis de varianza en su forma más básica es comprobar si una variable con varios niveles, llamada independiente puede explicar las variaciones observadas en otra u otras llamadas dependientes (Manzano, 1997).

De manera general, de acuerdo a este autor, se puede describir el análisis de varianza como la técnica que permite comprobar si *k* muestras provienen o no de poblaciones con la misma media, cuando se dan ciertas condiciones establecidas. Los procedimientos que se aplican permiten descomponer la variabilidad total en variabilidades aportadas por los distintos factores puestos en juego, además de sus interacciones mutuas. Esta es una descripción muy simplificada, pues dentro del término análisis de varianza encontramos una pluralidad de métodos, aunque todos ellos tienen unas bases comunes, que es la diferenciación de diversos tipos de variables y la realización de unos contrastes específicos, así como unos supuestos comunes de aplicación.

### 1.7.2. PRE-REQUISITOS

Para comenzar, describiremos los requerimientos básicos que el estudiante ha de conocer para poder comprender el análisis de varianza, que incluye ciertas distribuciones de probabilidad, el intervalo de confianza, necesario para realizar estimaciones, y el contraste de hipótesis, ya que el análisis de varianza involucra la realización de contrastes.

#### Lectura de tablas de distribuciones teóricas y cálculo de probabilidades

Supondremos que los estudiantes que se inician en el análisis de varianza manejan el concepto de *variable aleatoria*, así como la diferencia entre *estadístico* (por ejemplo, la media o la varianza obtenida en una muestra) y *parámetro* (el mismo valor en la población). Deberían comprender que los datos que manejan son muestras de valores de variables aleatorias extraídas de una o varias poblaciones. Además tendrían que saber utilizar las tablas de distribuciones teóricas para el cálculo de probabilidades en distintas distribuciones.

Por ejemplo, si en los cálculos, suponemos que tenemos una variable aleatoria *X* con una distribución Normal con parámetros 0 y 1 y necesitara determinar $P(X < 1{,}2)$ debería leer directamente la tabla de la distribución normal estándar. Para calcular $x_0$ tal que $P(X < x_0) = 0{,}5$, tendría que hacer una lectura inversa, buscando el percentil de la distribución, $x_0$, que corresponde a dicha probabilidad.

**Muestras relacionadas e independientes**

Debido a la importancia de un correcto diseño en estadística, para obtener conclusiones confiables, se insiste en la necesidad de tomar muestras aleatorias, esto es, donde cada una de las muestras posibles tiene igual probabilidad de ser escogida (Montgomery, 2005). Un punto importante (para diferenciar el método de análisis de varianza a aplicar) es discriminar si los resultados obtenidos a través de dos muestras pueden haber sido tomadas de dos poblaciones distintas (muestras independientes), o tratarse de mediciones diferentes en una misma población (muestras relacionadas). En general podemos decir, que dos muestras son independientes cuando las observaciones de una de ellas no condicionan en ningún sentido a la otra. Hablamos de dos muestras relacionadas cuando se considera una sola población en que se han tomado dos medidas (por ejemplo medidas antes y después de un tratamiento) (Andrés y Luna, 2004).

Siguiendo igual criterio, serán independientes $k$ muestras, cuando las observaciones de una de ellas no condicionan en ningún sentido a alguna de las otras, siendo las dependientes o relacionadas el caso contrario. En el primer caso se usa el análisis de varianza de efectos fijos o aleatorios y en segundo análisis de varianza de medidas repetidas (Vallejo y Lozano, 2006). En este último método se registran diversas medidas de la variable dependiente a un mismo grupo de sujetos, siendo muy útil para la evaluación de los procesos de cambios en contextos evolutivos, sociales y educativos (Balluerka y Vergara, 2002).

**Distribución muestral del estadístico**

Los estudiantes deben comprender también el concepto de *distribución muestral* de un estadístico, y conocer algunas de estas distribuciones de uso frecuente para el análisis de varianza. En cualquier situación de muestreo, se entiende que podríamos generar muchas muestras a partir de una población, y calcular en cada una de ellas un estadístico (media, mediana, varianza, proporción de éxitos…). Por tanto el estadístico es una variable aleatoria y su distribución se denomina distribución muestral (Harradine, Batanero y Rossman, 2011).

Como ejemplos de estadísticos que usaremos para esta investigación se tienen: la media muestral $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ y la varianza muestral $S^2 = \frac{1}{n-1}\sum_{i=1}^{2}(Y_i - \overline{Y})^2$, siendo $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria tomada de la población en estudio. Revisaremos a continuación, siguiendo a Montgomery (2005), algunas de las distribuciones más importantes que involucra el análisis de varianza, a saber: distribución Normal, distribución $\aleph^2$ (Chi Cuadrado), distribución *t* (Student) y la distribución *F* (Fisher).

*Distribución Normal*

La distribución normal desempeña un papel fundamental para nuestro objeto de estudio, ya que una de las hipótesis básicas para la aplicación del modelo es la normalidad de los datos. Es también una de las distribuciones teóricas más utilizada en Psicología (por ejemplo el cociente intelectual sigue esta distribución). Por otro lado, el teorema central del límite asegura que la media muestral se distribuye, para tamaño grande de la muestra en forma aproximadamente normal y la distribución normal también interviene en algunos contrastes a posteriori.

Supongamos que *X* es una variable aleatoria que distribuye en forma Normal, entonces la distribución de probabilidades de *X* es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \; ; \; -\infty < x < \infty \quad (1)$$

donde en la expresión (1), los valores del parámetro $\mu$ (valor medio de la distribución) varían a lo largo de toda la recta real, mientras los de $\sigma$ (desviación estándar de la distribución) lo hacen en los reales positivos. En la Figura 1.7.1, se representan dos distribuciones Normales con diferentes parámetros.

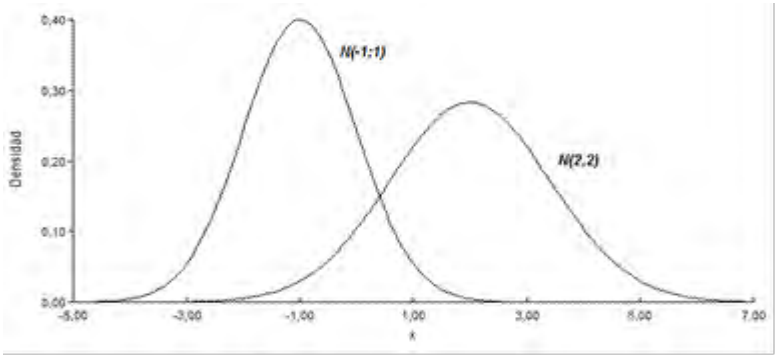



Figura 1.7.1. Distribución Normal con distintos parámetros

Si $X \sim N(\mu, \sigma^2)$ es una variable aleatoria, $X$ sigue una distribución Normal con valor medio $\mu$ y varianza $\sigma^2$ (parámetros que caracterizan la distribución), el cambio de variable $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ la transforma en otra variable aleatoria, tal que distribuye también normal, llamada normal estándar, es decir con parámetros 0 y 1.

*Distribución Chi Cuadrado*

Otra distribución relevante para el análisis de varianza es la Chi cuadrado. Supongamos una secuencia de variables aleatorias $N(0,1)$ independientes, digamos $Z_1, Z_2, \ldots Z_h$; entonces la variable aleatoria $X$, formada por la suma de sus cuadrados, $X = Z_1^2 + Z_2^2 + \cdots Z_h^2$ , sigue una distribución Chi cuadrado ($\aleph^2$) con $h$ grados de libertad, y su función de densidad es de la forma:

$$f(x) = \frac{1}{2^{\frac{h}{2}}\Gamma\left(\frac{h}{2}\right)} x^{\frac{h}{2}-1} e^{-\frac{x}{2}} \; , \; x > 0 \quad (2)$$

En la expresión (2), $\Gamma$ es la función matemática gamma y $h$ son los grados de libertad. En la Figura 1.7.2, se observa el cambio en la forma de la distribución, de acuerdo con los grados de libertad (la cantidad de variables aleatorias independientes involucradas).

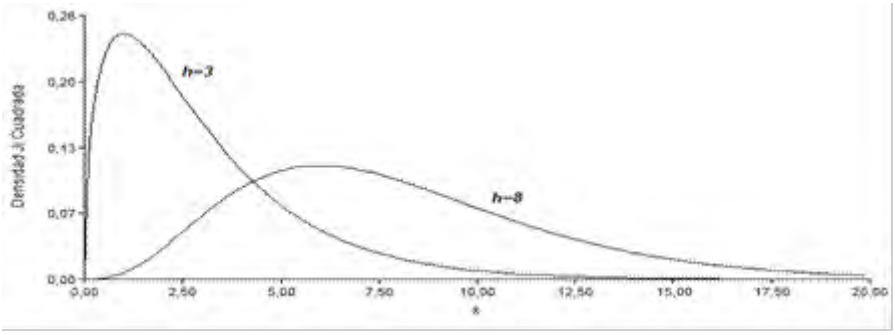



Figura 1.7.2. Distribución Chi cuadrado

Parte de las técnicas que se utilizan en el análisis de varianza se basan en el cálculo y comprensión de esta distribución. Por ejemplo, en cálculos del tipo $Z_i^2 = \left(\frac{X_i-\mu}{\sigma}\right)^2$ , puesto que si $X_i \sim N(\mu, \sigma^2)$ entonces $Z_i^2 \sim \aleph_1^2$. Esta distribución servirá, entre otras cosas, para justificar la distribución *F* del estadístico de contraste, ya que éste no es otra cosa que el cociente entre dos variables con distribución Chi cuadrado independientes, asociadas con su grado de libertad.

Por otro lado, si una secuencia $Y_1, Y_2, \ldots, Y_n$ forman una muestra aleatoria de una distribución $N(\mu, \sigma^2)$, entonces se tiene que

$$\frac{SS}{\sigma^2} = \frac{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{Y_i-\bar{Y}}{\sigma}\right)^2 \sim \aleph_{n-1}^2 \quad (3)$$

Análogamente y mediante un cálculo elemental podemos ver que si $S^2 = SS/(n-1)$, la distribución de $S^2$ es de la forma $\left[\frac{\sigma^2}{n-1}\right]\aleph_{n-1}^2$ , o sea una constante por una Chi- cuadrado.

*Distribución t de Student*

Esta es otra distribución relevante para nuestra investigación, ya que se recurre a ella para comparar las medias de dos poblaciones, bien sea en muestras independientes o relacionadas (Oliver, 2011). Además el análisis de varianza, en su versión más simple, puede ser visto como una extensión de la prueba *t* para muestras independientes cuando se deben comparar tres o más grupos independientes (Peers, 2006).

Supongamos dos variables aleatorias *Z* y $\aleph_h^2$ , respectivamente independientes con distribuciones normal estándar y Chi-cuadrado con *h* grados de libertad, entonces se tiene que la variable aleatoria $t_h = \frac{Z}{\sqrt{\frac{\aleph_h^2}{h}}}$ sigue una distribución *t* de Student con *h* grados de libertad, y la función de distribución de $t_h$ es de la forma:

$$f(x) = \frac{\Gamma\left[\frac{h+1}{2}\right]}{\sqrt{h\pi}\,\Gamma\left(\frac{h}{2}\right)} \frac{1}{\left[\frac{x^2}{h}+1\right]^{\frac{h+1}{2}}} \quad , \qquad -\infty < x < \infty \quad (4)$$

En la expresión (4), $\Gamma$ es la función matemática gamma y *h* son los grados de

libertad; además la media y la varianza de $t_h$ son respectivamente 0 y $h/(h-2)$ para valores de $h > 2$. En la Figura 1.7.3 ilustramos distribuciones *t* con distintos grados de libertad, también superpuestas. Observamos que, a medida que aumentan los grados de libertad, la distribución se asemeja más a la Normal estándar, para la práctica suele aproximarse a la distribución normal cuando el tamaño muestral es superior a 30. Esta semejanza la notamos especialmente observando las colas de las distribuciones.

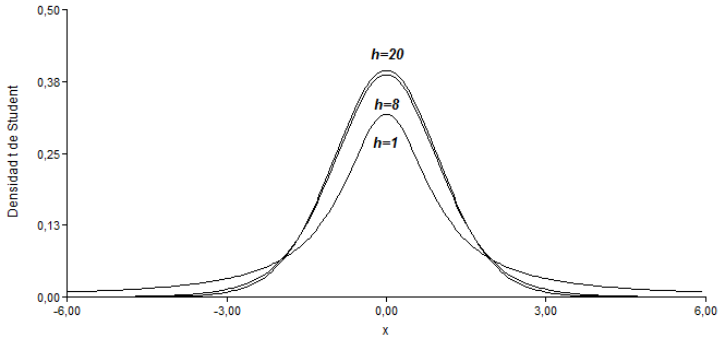



Figura 1.7.3. Distribución *t* de Student para distintos grados de libertad

Finalmente, es posible mostrar que la variable aleatoria $t = \frac{\overline{Y}-\mu}{S/\sqrt{n}}$ tiene una distribución *t* con *n-1* grados de libertad; siendo $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de una distribución $N(\mu, \sigma^2)$. Vemos así la relación estrecha entre las distribuciones *t* y la Normal.

*Distribución F de Fisher*

Presentaremos para finalizar la distribución *F* que está directamente relacionada con el análisis de varianza, pero requiere de las anteriores para existir, puesto que se forma partiendo del cociente de dos variables aleatorias que distribuyen según Chi cuadrado (5); y a su vez cada una de éstas está relacionada con distribuciones normales.

Supongamos dos variables aleatorias independientes cada una con distribución Chi Cuadrado con *u* y *v* grados de libertad respectivamente, las anotamos: $\aleph_u^2$ y $\aleph_v^2$. Formamos el cociente que denotamos por $F_{u,v}$ del siguiente modo:

$$F_{u,v} = \frac{\aleph_u^2/u}{\aleph_v^2/v} \qquad (5)$$

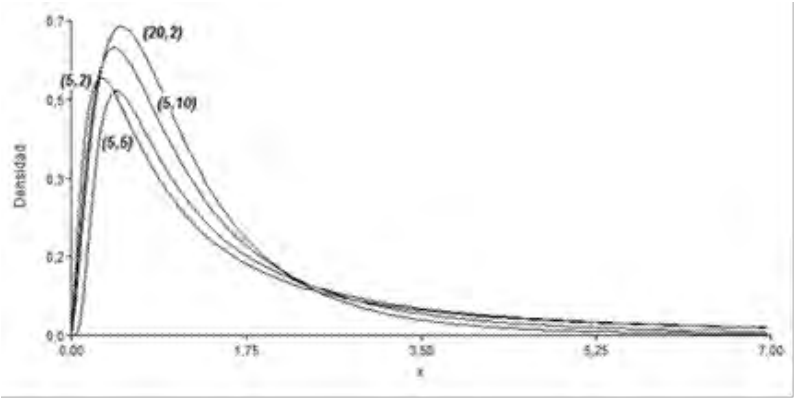



Figura 1.7.4. Distribuciones *F* de Fisher para algunas combinaciones de gl

La expresión (5), sigue una distribución $F$ con $u$ grados de libertad en el numerador y $v$ grados de libertad en el denominador. La función de distribución de $X \sim F(u, v)$ está dada por:

$$h(x) = \frac{\Gamma\left(\frac{u+v}{2}\right)\left(\frac{u}{v}\right)^{\frac{u}{2}} x^{\frac{u}{2}-1}}{\Gamma\left(\frac{u}{2}\right).\Gamma\left(\frac{v}{2}\right)\left[\frac{u}{2}.x+1\right]^{\frac{u+v}{2}}}, \qquad 0 < x < \infty \quad (6)$$

Presentamos los gráficos de varias distribuciones para diversos pares de grados de libertad, observando sus diferencias en la Figura 1.7.4. Un ejemplo importante de aplicación para esta distribución es la del cociente de varianzas muestrales. Supongamos dos poblaciones con distribución Normal independientes e igual varianza $\sigma^2$; se toman dos muestras aleatorias de tamaños $n_1$ y $n_2$ respectivamente de cada población. Entonces, si $S_1^2$ y $S_2^2$ son sus varianzas muestrales respectivamente, se tiene

$$\frac{S_1^2}{S_2^2} \sim F_{n_1-1,\, n_2-1} \quad (7)$$

En particular en el análisis de varianza se realizará un contraste de cociente de varianzas entre grupos y dentro de los grupos que, formalmente es similar al anterior (7).

### 1.7.3. INTERVALO DE CONFIANZA

Aunque los contrastes de hipótesis son un procedimiento muy útil, en múltiples oportunidades han de complementarse con la estimación de parámetros a través de los intervalos de confianza (Batanero, 2001).

El intervalo de confianza es necesario para nuestro objeto de investigación, especialmente cuando se determinan los efectos de cada uno de los factores. Por ejemplo, en el modelo de efectos fijos, cuando se rechaza la igualdad de medias, significa que al menos uno de los efectos no es nulo; pero no se sabe con exactitud cuáles de esas medias difieren, o cual de los efectos es distinto de cero. Para ello son muy útiles las comparaciones y los análisis adicionales entre grupos de las medias de los tratamientos. También se realizan intervalos de confianza para los contrastes, éstos no son otra cosa que una combinación lineal de los parámetros, donde las constantes de la combinación suman cero (Montgomery, 2005).

El intervalo de confianza según Olivo (2008) es uno de los procedimientos generales de la inferencia estadística. El autor analiza una doble definición.

- En el contexto de estimar un parámetro poblacional, lo define como un rango de valores (calculado a partir de los datos de una muestra) en el cual podría encontrarse el verdadero valor del parámetro, junto con un coeficiente de confianza que indica el porcentaje de muestras tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro.
- Como procedimiento general, describe una regla de construcción de dicho rango de valores a partir de un estadístico calculado en los datos de la muestra, para el parámetro correspondiente. La idea general de intervalo de confianza se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y según las condiciones (tipo de distribución, qué se conoce de la misma, etc.).

Dos niveles son importantes para el estudiante a la hora de comprender este objeto: el conocimiento del cálculo correcto del intervalo para su construcción y su interpretación.

**Construcción del intervalo de confianza**

Para encontrar un intervalo de confianza para un parámetro particular es útil determinar un procedimiento para obtenerlo. Siguiendo a López Sánchez, Pérez de Vargas, Zamora Romero, Murciano Cespedosa, Alonso Fernandez, Riviriego Eiros y Lahoz Beltrá (2004), mostraremos el cálculo del intervalo para la media poblacional. Consideramos una muestra aleatoria $X_1, X_2, \dots, X_n$ de tamaño *n* extraída de la población *X*.

El estadístico que emplearemos es: $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ , siendo $\overline{X}$ la media de la muestra y $S$ su desviación típica. Este estadístico sigue una distribución *t* de Student con *n-1* grados de libertad. El nivel de confianza $1-\alpha$, lo establece a priori el experimentador (los usuales son 0,90; 0,95; y 0,99). Dada la distribución del estadístico y el nivel de confianza, se tiene la siguiente igualdad $P\left(-t_{(n-1),\alpha/2} < \frac{\overline{X}-\mu}{S/\sqrt{n}} < t_{(n-1),\alpha/2}\right) = 1-\alpha$, siendo $t_{(n-1),\alpha/2}$ el valor crítico en la *t* de Student.

La expresión anterior es equivalente a: $P\left(\frac{\overline{X}-t_{(n-1),\alpha/2}}{S/\sqrt{n}} < \mu < \frac{\overline{X}+t_{(n-1),\alpha/2}}{S/\sqrt{n}}\right) = 1-\alpha$, el cual hace referencia a que con una probabilidad $1-\alpha$ el intervalo aleatorio $\left(\overline{X} - t_{(n-1),\frac{\alpha}{2}}.\frac{S}{\sqrt{n}}; \overline{X} + t_{(n-1),\frac{\alpha}{2}}.\frac{S}{\sqrt{n}}\right)$ contendrá el valor medio poblacional $\mu$ . En el caso que la muestra tenga un tamaño suficiente o la distribución de partida sea normal, el procedimiento que se deberá emplear aquí para la determinación del intervalo considerado es componer la fórmula $\overline{X} \pm z_{\frac{\alpha}{2}}.\frac{S_d}{\sqrt{n}}$ , usando los datos antes mencionados.

**Interpretación del intervalo de confianza**

Será importante, como hemos señalado la interpretación que se hace sobre el intervalo de confianza, pues tanto Behar (2001) como Olivo (2008) encuentran errores al interpretar este objeto estadístico; tanto en investigadores el primero, como alumnos de grado el segundo. En la investigación de Olivo (2008), los estudiantes no alcanzan a interpretar qué representa un intervalo de confianza, dando en general una interpretación bayesiana del mismo, es decir como la probabilidad de que el parámetro esté contenido en el intervalo. Sólo el 36,5% (*n* =252) de los estudiantes en su muestra dieron una respuesta correcta.

Nótese que los extremos del intervalo varían aleatoriamente de muestra a muestra, mientras que el valor del parámetro es fijo (constante) aunque desconocido. Dado que sus extremos se determinan a partir de los estimadores: media muestral y desviación típica muestral, que son variables aleatorias; el intervalo es aleatorio.

La probabilidad a que se refiere dicho intervalo puede interpretarse de manera informal pero quizás sea más clara del siguiente modo: si consideramos todas las muestras distintas de tamaño *n* que puedan ser extraídas de la población *X*, y con cada una construimos los correspondientes intervalos, el (1-$\alpha$)100% de estos intervalos contendrán el verdadero parámetro $\mu$. Como vemos, el coeficiente de confianza sólo nos

da la proporción de intervalos calculados de la misma población con igual tamaño de muestra que cubrirían el valor del parámetro, pero no sabemos si el intervalo calculado en alguna de las muestras lo cubre o no (Cumming, Williams y Fidler, 2004; Cumming y Fidler, 2005).

Otra observación es que el nivel de confianza establece en alguna medida la amplitud del correspondiente intervalo de confianza. Aumentando el nivel de confianza (mayor certeza), aumenta la amplitud (menor precisión).

### 1.7.4. CONTRASTE DE HIPÓTESIS

Siguiendo a Montgomery (2005), recordamos que una hipótesis estadística es un enunciado o afirmación ya sea de los parámetros de un modelo o de una distribución de probabilidad que refleja alguna conjetura acerca de la situación del problema.

Supongamos a modo de ejemplo, que se desea conocer si la autoestima en sujetos extrovertidos es mayor que en sujetos introvertidos. Para esto consideremos que $y_{11}, y_{12}, \dots, y_{1n_1}$ representan las $n_1$ observaciones correspondientes a una muestra de una población de sujetos extrovertidos; e $y_{21}, y_{22}, \dots, y_{2n_2}$ las $n_2$ observaciones de otra muestra de una población de sujetos introvertidos. Se quiere estudiar si difieren en su puntuación media de autoestima. Batanero (2000) describe el razonamiento típico de un contraste, en el que habría que seguir los pasos siguientes:

1. *Describir el efecto a contrastar en función de los parámetros de una o varias poblaciones.* En este primer paso se caracteriza la hipótesis alternativa ($H_1$). En el ejemplo, consideramos que la puntuación media de la autoestima en los sujetos extrovertidos, $\mu_E$ será mayor que en los sujetos introvertidos, $\mu_I$. Esperamos un efecto ($\mu_E > \mu_I$ ), y será aquel que describe la hipótesis alternativa: $H_1: \mu_E > \mu_I$.
2. *Establecer la hipótesis nula ($H_0$) indicando que el efecto no se presenta:* En el mismo ejemplo, $H_0: \mu_E = \mu_I$ (no hay diferencias entre las puntuaciones medias en los dos grupos). El contraste de hipótesis intenta evaluar la fuerza de la evidencia de los datos en contra de la hipótesis nula. Siguiendo a Batanero (2000) suponemos una hipótesis nula simple, para facilitar la discusión.
3. *Calcular un estadístico a partir de los datos*. Supuesta la hipótesis nula cierta, calculamos como estadístico la diferencia de las medias en dos muestras extraídas aleatoriamente. La distribución del estadístico quedará especificada (en el ejemplo sería una distribución *t* de Student). Supongamos en dicho ejemplo los siguientes resultados para las medias de las dos muestras de tamaño *n*=40: $\bar{x}_E = 115{,}1$; $\bar{x}_I = 101{,}78$ y que $s_E^2 = 197{,}66$; $s_I^2 = 215.19$ son los estimadores insesgados de las varianzas en las poblaciones. En este caso se obtiene: $\bar{x}_E - \bar{x}_I = 13{,}32$; la estimación de la varianza conjunta $s^2 = 202{,}48$; y *t*=4,16. En el contraste tratamos de evaluar si este valor es improbable, bajo la hipótesis nula cierta.
4. *Determinación del valor-p:* Se llama *valor-p* a la probabilidad de obtener un valor del estadístico de contraste tan extremo o más que el valor *t* observado cuando la hipótesis nula es cierta. En nuestro ejemplo, el valor-p correspondiente al valor *t* observado es extremadamente pequeño, menor que 0,001. Por tanto este valor es muy improbable, en caso de ser la hipótesis nula cierta y por ello rechazamos dicha hipótesis.
5. Esta decisión no implica que hayamos demostrado la falsedad de la hipótesis nula; pues sólo hemos analizado una parte de la población. La demostración deductiva de una hipótesis por métodos estadísticos no es posible; por lo que la aceptación o rechazo de

las hipótesis se considera provisional. Será la comunidad científica, según Batanero (2000) quien a la larga decida qué hipótesis se dan por establecidas.

La descripción anterior muestra la complejidad del contraste de hipótesis. Describimos a continuación los conceptos y procedimientos que los estudiantes deben manejar al plantear la resolución de un contraste de hipótesis, según los pasos anteriormente señalados.

**Asignación de hipótesis estadísticas. Hipótesis nula y alternativa**

Chow (1996) recuerda la diferencia entre hipótesis de investigación e hipótesis experimental y entre éstas y las hipótesis estadísticas. Lo primero que el alumno ha de diferenciar es la naturaleza de dichas hipótesis. En el ejemplo usado en la sección anterior, la hipótesis de investigación sería la creencia que los sujetos extrovertidos tienen mayor autoestima que los introvertidos.

En una investigación empírica la hipótesis de investigación suele ser difusa al comienzo; en el ejemplo, la autoestima se podría medir con diferentes instrumentos. Es necesario diseñar una variable dependiente específica (en el ejemplo la puntuación total derivada de un test específico) y ponerla a consideración a algunos sujetos (los sujetos extrovertidos e introvertidos). Con esta base podemos construir una *hipótesis experimental* (que la puntuación del test será mejor en la población extrovertida que en la introvertida).

Una consecuencia de la hipótesis experimental se usará para llevar a cabo el análisis estadístico (que la puntuación media en el test de autoestima será más alta en los sujetos extrovertidos que los introvertidos). Esta implicación es la *hipótesis estadística alternativa*, $(H_1: \mu_E > \mu_I)$, que no es coincidente con la hipótesis experimental, sino que es una consecuencia de ella. Finalmente, la negación de la hipótesis estadística alternativa es que la puntuación media en las dos poblaciones de estudiantes es la misma, $H_0: \mu_E = \mu_I$. El establecer la hipótesis nula sirve para especificar la distribución del estadístico del contraste en el muestreo y comenzar los cálculos que llevarán a rechazarla o no.

En resumen, a nivel estadístico el alumno ha de diferenciar entre las hipótesis que se plantean en cada problema (Hipótesis nula $H_0$ e hipótesis alternativa $H_1$), que han de ser complementarias entre ellas (además la igualdad siempre se debe verificar bajo $H_0$). La hipótesis alternativa es la que interesa probar al investigador; por el contrario el razonamiento se hace en base a la hipótesis nula.

Los alumnos confunden con frecuencia todas estas hipótesis, por lo que al rechazar la hipótesis nula podrían pensar que han probado la hipótesis experimental o incluso la de investigación (Chow, 1996). Vallecillos (1996) indica, además, que algunos estudiantes no comprenden que las hipótesis se refieren a los parámetros de la población y no a los estadísticos muéstrales. Asimismo, es importante que los alumnos entiendan que las hipótesis se deben formular con anterioridad a la toma de datos y no una vez observados.

**Región crítica y de aceptación**

Como se ha indicado, el contraste de hipótesis se lleva a cabo tomando como criterio de decisión el valor obtenido en un estadístico relacionado con el parámetro de interés (Batanero, 2000; Harradine, Batanero y Rossman, 2011). La distribución de este estadístico, por ejemplo, la distribución de la media muestral, depende del parámetro

(en el ejemplo: la media poblacional).

Teniendo en cuenta la distribución de ese estadístico cuando la hipótesis nula sea verdadera, podremos rechazar o mantener la hipótesis nula $H_0$. Llamaremos valor observado del estadístico de contraste al obtenido usando los datos de la muestra. Como parte de este procedimiento también se especifica el conjunto de valores del estadístico de prueba que llevan al rechazo de $H_0$. Es decir, se divide el conjunto de sus posibles valores en dos conjuntos mutuamente excluyentes (Díaz y Batanero, 2006), a saber:

- Uno de ellos contiene todos los valores del estadístico para los cuales no se rechaza la hipótesis nula y se llama región de aceptación.
- El otro, los valores para los cuales se rechaza la hipótesis nula, y se conoce como región crítica ó zona de rechazo.

Llamaremos *puntos críticos* a los valores que separan las dos regiones. Si obtenemos un punto crítico como valor del estadístico, al no poder tomar una decisión, deberemos aumentar el tamaño de la muestra y repetir el contraste. Es importante resaltar que las regiones crítica y de aceptación cubren todo el rango de variación del estadístico, es decir, constituyen una partición de dicho rango. La forma de la región de aceptación y rechazo (o región crítica) dependerá si el contraste es unilateral o bilateral, se verá en el siguiente apartado.

**Contraste unilateral y bilateral**

Otro concepto importante a considerar es si el contraste es unilateral o bilateral, porque ello afecta al tipo de hipótesis que se contrastan y a la construcción de la región crítica y de aceptación. El llevar a cabo un contraste unilateral o bilateral (también llamados de una o dos colas) viene determinado por el problema que se haya planteado, lo que a su vez determina las hipótesis (Montgomery, 2005; Peers, 2006). Cuando las hipótesis se plantean únicamente en términos de igualdad para la hipótesis nula y desigualdad sin dirección para la alternativa, se tendrá un contraste bilateral. En este caso la región crítica estará formada por las dos colas de la distribución del estadístico de contraste.

Si en la hipótesis alternativa se indica no sólo desigualdad sino en qué dirección se da dicha desigualdad (Batanero, 2000) estaremos en un caso unilateral (por ejemplo, si se quiere contrastar que la autoestima en los sujetos extrovertidos es mayor que en los introvertidos). En este caso la región crítica está formada por una cola de la distribución del estadístico.

Para clarificar, centraremos la atención en un contraste de hipótesis para una media, con una muestra de tamaño *n* y distribución normal con desviación estándar conocida. Manteniendo la hipótesis nula de la forma: $H_0: \mu = \mu_0$, las hipótesis alternativas posibles para una prueba unilateral podrían ser: $H_1: \mu > \mu_0$ (contraste unilateral a derecha) ó $H_1: \mu < \mu_0$ (contraste unilateral a izquierda).

Supongamos se tiene el caso $H_1: \mu > \mu_0$. Puesto que $Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1)$, cuando $H_0$ es verdadera $P(Z \leq z_\alpha) = 1 - \alpha$ (siendo $\alpha$ el nivel de significación de la prueba). Entonces rechazaremos la hipótesis nula si con los datos se tiene evidencia de la alternativa, es decir si $\bar{x} \gg \mu_0$ , así rechazamos la hipótesis nula (siendo ésta verdadera) a favor de la alternativa, si $\bar{x} \in [\mu_0 + z_\alpha . \frac{\sigma}{\sqrt{n}}; +\infty)$ ó $\frac{(\bar{x}-\mu)}{\sigma/\sqrt{n}} \in [z_\alpha; +\infty)$ y este intervalo es la región crítica para este contraste.

Si el contraste es bilateral con hipótesis $H_0: \mu = \mu_0$ y $H_1: \mu \neq \mu_0$ , rechazaremos la hipótesis nula en favor de la alternativa cuando $\bar{x} \ll -\mu_0$ ó $\bar{x} \gg \mu_0$ , es decir cada vez que $\bar{x}$ este en la siguiente unión de intervalos:

$\left(-\infty; \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \cup [\mu_0 + z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}; +\infty)$ y esta es la región crítica del contraste.

**Valor p; nivel de significación, probabilidad error tipo I**

En las pruebas de Neyman-Pearson, se consideran dos tipos de errores (Pagano, 2011). En este apartado analizamos el error tipo I, que consiste en rechazar la hipótesis nula $H_0$ siendo en realidad $H_0$ verdadera. Su probabilidad se conoce como *nivel de significación* de la prueba, anotado $\alpha$. Es decir: $\alpha = P(\text{error tipo I})$.

Aunque es posible reportar los resultados de una prueba de hipótesis sólo en base al nivel de significación $\alpha$ especificado a priori, esta costumbre, según Montgomery (2005) "*es con frecuencia inadecuada porque no le ofrece al responsable de la toma de decisiones idea alguna de si el valor calculado del estadístico de prueba apenas rebasó la región de rechazo o si se adentró bastante en la misma*" (p. 37). Por ello se ha adoptado muy extensamente en la práctica reportar el valor *p*, o probabilidad de que el estadístico de prueba sea al menos igual al valor que se observa del estadístico, siempre que la hipótesis nula sea verdadera. Se decide rechazar la hipótesis nula siempre que el valor *p* observado es menor o a lo sumo igual que el nivel de significación $\alpha$ establecido de antemano.

Como veremos en el Capítulo 2, los conceptos de nivel de significación y valor p suelen ser interpretados en forma incorrecta, siendo uno de los errores más frecuentes el intercambio en los términos de la probabilidad condicional que los definen.

**Significación de un resultado**

Para que un resultado se considere estadísticamente significativo, no existe una regla fija acerca de cuan pequeño debe ser el valor-*p*, o significación estadística del resultado. Aunque, convencionalmente en la práctica adoptamos un valor fijo del nivel de significación, que se elige a criterio del investigador, los valores más comunes de usar son 0,05 y 0,01; con alguno de éstos se compara el valor-*p* para decidir sobre la significación estadística del resultado de la prueba.

Según Díaz (2007), es importante diferenciar entre significación estadística y significación práctica. La significación práctica implica significación estadística más un efecto experimental suficientemente elevado (Cohen, 1994); en el ejemplo de comparaciones de nivel de autoestima, una diferencia suficientemente alta entre las medias, para que sea relevante en la investigación.

Lecoutre (1999) hace notar que podemos obtener una significación estadística, a pesar de que no haya significación práctica, porque el efecto sea pequeño, puesto que la significación estadística se alcanza siempre que tomemos una muestra suficientemente grande.

**Potencia, probabilidad de error tipo II**

Mientras que la probabilidad de cometer error tipo I es constante, la probabilidad de error tipo II o probabilidad de aceptar una hipótesis nula $H_0: \theta = \theta_0$, cuando es falsa,

resulta variable, y se suele conocer P(error tipo II) como $\beta$. La razón es que, en caso de ser verdadera la hipótesis alternativa $H_1$, hay un conjunto infinito de valores que hacen que la hipótesis nula no sea verdadera. Por ello $\beta$, depende del verdadero valor de $\theta$ (parámetro genérico a contrastar) por ello es una función de $\theta$. Cuando $\theta = \theta_0$ (la hipótesis nula es verdadera), $\beta(\theta_0) = 1 - \alpha$ que es la probabilidad de aceptar la hipótesis nula, siendo cierta (no se comete error) (Díaz y Batanero, 2008; Pagano, 2011).

Al complemento de $\beta(\theta)$ ó probabilidad de rechazar la hipótesis nula cuando no sea cierta se la denomina *potencia del test*, la cual también es variable pues depende del verdadero valor que se le asigna al parámetro. A la potencia del test se la suele simbolizar por $\pi(\theta)$, de esta forma resultará que: $1 - \beta(\theta) = \pi(\theta)$. Otra manera de definir la potencia del criterio es como la probabilidad de que el estadístico tome un valor de la región crítica, cuando es cierta la hipótesis alternativa. Esta probabilidad es la contraria de que el estadístico caiga en la región de aceptación, siendo cierta $H_1$.

Es claro que cuando menores sean los valores de $\alpha$ y $\beta(\theta)$ tanto mejor será el contraste. Sin embargo estos errores están ligados entre sí, en el sentido que, al disminuir uno aumenta el otro (Vallecillos, 1994). Para comprender esto basta considerar que si hacemos $\alpha$ =0, es decir si aceptamos siempre la hipótesis $H_0$, $\beta(\theta) = 1$ para cualquier valor posible de $\theta$ distinto del supuesto, ya que siempre rechazamos la hipótesis alternativa, aún en el caso de que sea cierta. La única forma de disminuir a la vez las dos probabilidades de errores es aumentar el tamaño de la muestra, situación que no siempre es posible en los procesos de experimentación.

Para un nivel de significación $\alpha$ fijo, es posible sin embargo, determinar la región crítica, de forma que se obtenga la potencia máxima, y por tanto disminuya lo más posible $\beta(\theta)$ para los valores de $\theta$ distintos del supuesto (Pagano, 2011).

**Regla de decisión**

Una vez elegido el estadístico y formadas las regiones de aceptación y rechazo se establece una regla de decisión, que permitirá rechazar o no la hipótesis nula. Se trata de comparar el valor del estadístico de contraste con el valor o valores críticos que separan las dos regiones. Dependiendo del tipo de contraste (si es unilateral o bilateral) y de la distribución del estadístico de prueba formaremos nuestra regla de decisión. Decidimos rechazar la hipótesis nula cuando el valor del estadístico de contraste está en la zona de rechazo o zona crítica; y por el contrario, mantendremos la hipótesis nula cuando el estadístico de contraste caiga en la zona de aceptación.

Según Batanero (2000), además de comprender los diferentes conceptos implicados, se ha de comprender el razonamiento subyacente al procedimiento del contraste de hipótesis. La autora indica que la estructura formal de los contrastes estadísticos es superficialmente parecida a la de la prueba por contradicción, pero hay diferencias fundamentales, no siempre bien comprendidas entre estos dos tipos de razonamientos. En una prueba por contradicción razonamos en la forma siguiente:

- Si A, entonces B no puede ocurrir.
- B ocurre; entonces deducimos que A es falso.

En un contraste estadístico tratamos de aplicar un razonamiento similar en la siguiente forma:

- Si A, entonces es muy poco verosímil que B ocurra;

- Ocurre B, y rechazamos A

Sin embargo no sería una conclusión válida deducir que es muy improbable que A sea cierto y aquí es donde encontramos la confusión de muchos investigadores (Batanero, 2000; Harradine, Batanero y Rossman, 2011).

### 1.7.5. TIPOS DE VARIABLES EN ANÁLISIS DE VARIANZA

Una vez analizados los conocimientos básicos necesarios para abordar el análisis de varianza, comenzamos con el estudio del tema propiamente dicho, en primer lugar con la clasificación de los tipos de variables que pueden considerarse. Para el análisis de varianza se consideran diferentes tipos de variables (Box, Hunter y Hunter, 2008):

- Las variables *dependientes* o *respuesta* son aquellas en las que está interesado un investigador, quien desea evaluar su variación en distintos grupos o poblaciones y son cuantitativas.

  En nuestro trabajo nos interesamos por el análisis de varianza con una sola variable dependiente, aunque pudiera haber más de una, si consideramos el Análisis multivariante de varianza o MANOVA. En el campo de la Psicología algunos ejemplos de variables dependientes serían puntuaciones en un test, tiempo necesario para realizar una tarea o resultados de mediciones de algún atributo psicológico.

- Las variables *independientes* o *factores,* que usualmente son cualitativas o bien tienen un conjunto finito de valores, que se denominan niveles del factor. Estas variables son las manipuladas por el investigador o bien controladas por el investigador en su estudio. En el análisis de varianza, estamos interesados en determinar en qué medida los *factores*, podrían producir cambios (que se denominan *efectos*) sobre la variable dependiente. Algunos ejemplos de variables independientes en Psicología podrían ser sexo, edad o grupo (control-experimental).

#### Factores y sus tipos. Muestras independientes y relacionadas

En cada estudio concreto se puede tener uno o varios factores, y además estos se pueden dividir en varios tipos:

- *Factor de efectos fijos:* Cuando incluimos en el estudio todos los niveles posibles del factor (Dunn y Clark, 1987). Por ejemplo, al estudiar la diferencia de rendimiento escolar dependiendo del tipo de establecimiento que provienen (como privado- publico-concertado), la variable dependiente es el rendimiento, el factor es el tipo de establecimiento del cual proviene y todos los niveles posibles se incluyen en el estudio. Siempre que los experimentos de varios factores de efectos fijos se realicen combinando todos los niveles posibles, se hablará de *diseños completos.*

- *Factor de efectos aleatorios:* En este caso el factor tiene un conjunto muy amplio de niveles que no se pueden incluir en su totalidad, de modo que se toma una muestra aleatoria en el estudio (Molinero, 2003). En el ejemplo anterior podríamos incluir un segundo factor "barrio" y tomar una muestra aleatoria de barrios en una ciudad.

Otra división posible de los factores es intersujetos y factores intrasujeto (Batanero y Díaz, 2008):

- *Factor intrasujeto* es aquél en que los varios niveles del factor se toman en la misma unidad estadística. Por ejemplo, podríamos tomar la tensión arterial (variable

respuesta) a un conjunto de personas antes y contemplando tres etapas de un tratamiento médico a lo largo de dos meses.

El factor intrasujeto sería el tiempo en que se toma la tensión con cuatro niveles a lo largo del tratamiento

- *Factor intersujeto*, cuando los diversos niveles se contemplan en diferentes unidades estadísticas; en el ejemplo anterior sería el caso de diferenciar el factor género (hombre/mujer).

Además, pueden aparecer variables que, no siendo de interés al investigador, afecten a la variable dependiente. Estas variables se denominan *extrañas* y el investigador ha de tratar de controlarlas para que no afecten al experimento. La finalidad del análisis de varianza es poder comprobar las hipótesis objeto de estudio, controlando el error aleatorio y las variables extrañas (Tejedor, 1999).

Las variables extrañas pueden controlarse manteniéndolas constantes o por medio de la asignación aleatoria de los sujetos experimentales a cada nivel de cada factor, pero esto suele ser difícil. Otra posibilidad es medirlas e incluirlas como parte del análisis estadístico; en este caso se suelen denominar *covariables* (León y Montero, 2002). Por ejemplo, si se quieren comparar varios medicamentos para un mismo tipo de síntoma y sospechamos que la edad puede afectar su efecto sobre los pacientes, podríamos incluir la edad como una covariable en el estudio cuando no se pueda asignar aleatoriamente los medicamentos a los pacientes teniendo en cuenta la edad.

### 1.7.6. ANÁLISIS DE VARIANZA DE UN FACTOR EFECTOS FIJOS

En un estudio estadístico, en general se tienen variables independientes con más de dos niveles diferentes, lo que lleva al problema de comparación de varias muestras. Este problema no puede solucionarse estudiando las muestras dos a dos para intentar detectar las diferencias significativas, pues se produce el problema de las "comparaciones múltiples" (Moses, 1992). Este autor indica que cuando efectuamos $k$ comparaciones teniendo en una de ellas un nivel de significación individual $\alpha$, el nivel de significación global toma el valor $k\alpha$, por lo que crece la probabilidad de encontrar diferencias significativas simplemente por azar.

El análisis de varianza intenta paliar este problema, a la vez que resulta un procedimiento más eficaz de análisis. Dependiendo del número y tipo de factores hay varios modelos que se pueden elegir. El modelo más simple es el conocido como análisis de varianza de un factor, con efectos fijos, que describiremos a continuación, siguiendo a Dunn y Clark (1987).

Supongamos que disponemos de $k$ grupos, y sólo $k$ grupos fijos (son los únicos con los que cuenta la población), de modo que en el grupo $i$-ésimo hay $n_i$ observaciones (caso no balanceado). Sea $y_{ij}$ la j-ésima observación en el grupo $i$, ésta se modela del siguiente modo:

$$y_{ij}=\mu+\alpha_i+\varepsilon_{ij} \quad \begin{cases} i=1,2,\ldots,k \\ j=1,2,\ldots,n_i \end{cases}$$

donde $\mu$ es una constante que representa la media global del conjunto de las $k$ muestras; $\alpha_i$ el efecto del $i$-ésimo tratamiento, constante dentro del grupo $i$. Si $\mu_i$ es la media de la población sometida al $i$-ésimo tratamiento, entonces $\alpha_i$ representa la diferencia de la

media de este grupo con la media global; es decir usando simbología matemática esto es $\alpha_i = \mu_i - \mu$; finalmente, $\varepsilon_{ij}$ es una cantidad aleatoria.

Suponemos para este modelo que $\varepsilon_{ij}$ tiene distribución normal con media 0 y varianza $\sigma^2$, y que la suma de todos los efectos es nula, es decir: $\sum_{i=1}^{k} \alpha_i = 0$. Todos los elementos del mismo grupo *i* se dice están "sujetos al tratamiento *i*" (Tejedor, 1999). Además, los valores del subíndice *i* varían desde el valor 1 hasta el valor *k* (número de grupos), el subíndice *j* varía según la cantidad de observaciones disponibles para el grupo *i*.

El análisis de varianza, consiste en la realización de un contraste estadístico para decidir entre las dos hipótesis siguientes (Batanero y Díaz, 2008):

$H_0$: $\alpha_1=\alpha_2=...=\alpha_k=0$
$H_1$:*al menos un* $\alpha_i$ *es distinto de 0*

Dicho de otro modo, bajo la hipótesis nula, todos los grupos provienen de poblaciones con igual media, y, en consecuencia, todos los tratamientos producen un efecto nulo. Bajo la hipótesis alternativa, al menos un tratamiento produce un efecto no nulo, por lo que algunas de las poblaciones tendrán diferentes medias. Las conclusiones que podamos obtener se referirán, por tanto a esos *k* valores de $\alpha_i$ y no a otros distintos, de ahí que el modelo se denomine de "efectos fijos".

**Modelo lineal correspondiente**

Según Batanero y Díaz (2008) para realizar el contraste, representaremos la desviación de la observación $y_{ij}$ respecto a la media total en la forma

$$y_{ij} - \bar{y} = \left(y_{ij} - \bar{y}_{i.}\right) + (\bar{y}_{i.} - \bar{y})$$

En este modelo lineal, el primer sumando de la derecha representa la desviación de la observación respecto a la media de la muestra *i*, y nos muestra la variabilidad dentro de la población *i.* El sumando $(\bar{y}_{i.} - \bar{y})$ es la desviación entre la media de la muestra *i* y la media global, y mide la variabilidad entre los grupos. Según las autoras, si la variabilidad entre grupos fuese grande respecto a la que hay dentro de cada grupo, pensaríamos que nos hallamos ante poblaciones con medias diferentes, y rechazaríamos la hipótesis nula de igualdad de efectos. En la Figura 1.4.6, se muestran dos ejemplos. En el ejemplo (a), la variabilidad entre grupos es mayor que la que hay en cada grupo. Por el contrario, en el (b) al haber una gran dispersión dentro de cada grupo, no permite apreciar si hay una diferencia real de medias en las poblaciones.

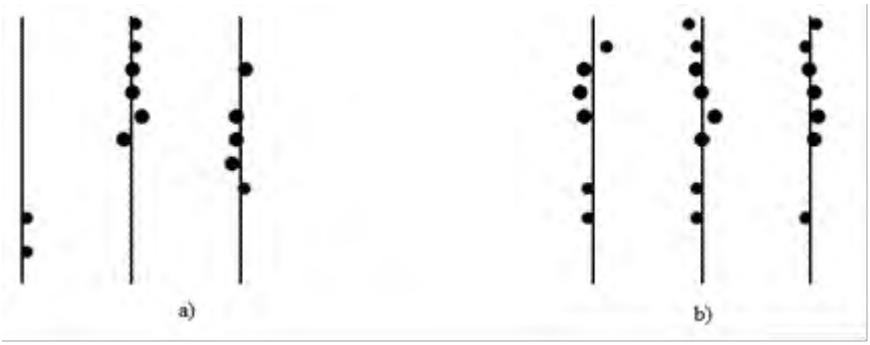



Figura 1.4.6. Ejemplos de variabilidad (adaptado de Batanero y Díaz, 2008)

**Cálculo del estadístico *F* e interpretación de resultados**

Para efectuar el análisis estadístico, se debe proceder a calcular los elementos de la tabla de análisis de varianza, dada en la Tabla 1.7.1 (Dunn y Clark, 1987). En la Tabla 1.7.1, se muestra el cálculo de los distintos componentes, donde $SC_T$ representa la suma de cuadrados total (variabilidad total), $SC_D$ la suma de cuadrados dentro de los grupos (variabilidad dentro de cada grupo) y $SC_E$ la suma de cuadrados entre grupos (variabilidad de un grupo a otro) también llamado suma de cuadrados del error, y se verifica que $SC_T = SC_D + SC_E$, con:

$$SC_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2,\ SC_D = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_{i.}\right)^2,\ SC_E = \sum_{i=1}^{k}(\bar{y}_{i.} - \bar{y})^2$$

Tabla 1.7.1. Análisis de varianza para un factor fijo

| Fuente de variación | Suma de cuadrados | Grados de libertad | Cuadrados medios | *F* |
|---|---|---|---|---|
| Entre grupos | $SC_E$ | $k-1$ | $CM_E = \frac{SC_E}{k-1}$ | $F_{obs} = \frac{CM_E}{CM_D}$ |
| Dentro de los grupos (residual) | $SC_D$ | $n-k$ | $CM_D = \frac{SC_D}{n-k}$ | |
| Total | $SC_T$ | $n-1$ | | |

Dunn y Clark (1987) indican que $SC_D$ es en realidad la varianza de los valores $y_{ij}$ respecto a la media global de todas las muestras, multiplicada por $n$. Por tanto, el cuadrado medio $CM_D$ estima la varianza de los valores $y_{ij}$ respecto a la media global, esto es, $\sigma^2$. Por su parte $CM_E$ estima la varianza de las medias de cada muestra respecto a la media global. Si la hipótesis nula fuese cierta, estas dos varianzas serían aproximadamente iguales, siendo las diferencias observadas pequeñas y debidas únicamente al error del muestreo.

El estadístico $F_{obs}$ sigue la distribución *F* con $k-1$ y $n-k$ grados de libertad. En el caso de ser mayor las diferencias entre grupos a las diferencias dentro de los grupos, el valor $F_{obs}$ será mayor que la unidad. Adoptaremos, en consecuencia, la siguiente regla de decisión: Si el valor $F_{obs} > F_{\alpha;k-1;n-k}$ se rechaza $H_0$ con nivel de significación $\alpha$, en caso contrario no se rechaza $H_0$ (Batanero y Díaz, 2008).

En consecuencia, para saber aplicar el análisis de varianza de un factor y efectos fijos, el estudiante tendrá que alcanzar las siguientes competencias y conocimientos, que trataremos de evaluar en nuestro estudio:

- Diferenciar las situaciones en que hay que aplicar el análisis de varianza de efectos fijos, respecto a otras situaciones.
- Comprender el modelo lineal que subyace en el análisis de varianza de efectos fijos.
- Diferenciar la variable dependiente y el factor, así como los niveles del factor.
- Comprender las hipótesis involucradas en el modelo.
- Llevar a cabo correctamente los cálculos de la tabla de análisis de varianza (Tabla 1.7.1) e interpretar los diferentes componentes de la misma.
- Interpretar correctamente los resultados para llegar a una decisión sobre las hipótesis.

**Estimación de los efectos del factor**

Una vez rechazada la hipótesis nula, pueden calcularse intervalos de confianza para los valores $\alpha_i$. Puesto que las $k$ poblaciones tienen una varianza común $\sigma^2$, y ésta viene estimada por $CM_D$, aplicando el procedimiento para el cálculo del intervalo de confianza de la media de una población, cuando no se conoce la desviación típica, obtenemos para $\alpha_i$ el intervalo de confianza, con coeficiente de confianza $1-\alpha$.

$$\bar{x}_i - t_\alpha \cdot \sqrt{\frac{CM_D}{n_i}} \leq \alpha_i \leq \bar{x}_i + t_\alpha \cdot \sqrt{\frac{CM_D}{n_i}}$$

En dicha expresión $t_\alpha$ es el percentil del $(1-\alpha)100\%$ de la distribución $t$ con $n$-$k$ grados de libertad (Dunn y Clark, 1987). También puede hallarse un intervalo de confianza para la diferencia de efectos entre dos tratamientos, que coincide con la diferencia de medias entre las dos subpoblaciones, mediante la expresión (Batanero y Díaz, 2008):

$$(\bar{x}_i - \bar{x}_j) - t_\alpha \cdot \sqrt{CM_D\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \leq \mu_i - \mu_j \leq (\bar{x}_i - \bar{x}_j) + t_\alpha \cdot \sqrt{CM_D\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

**Comparaciones específicas de medias de tratamientos**

Una vez rechazada la hipótesis de igualdad de las medias en los diferentes tratamientos, estaremos interesados, en general, en decidir si algunos tratamientos prefijados son o no significativamente diferentes. Aunque este tema no se contempla en la enseñanza a los alumnos de la muestra, haremos un breve resumen siguiendo a Díaz y Batanero y Wilhelmi (2008) para que el lector comprenda todos los pasos del análisis de varianza. Las hipótesis a contrastar son las siguientes:

$$H_0: \alpha_i = \alpha_j$$
$$H_1: \alpha_i \neq \alpha_j$$

Adoptaremos la siguiente regla de decisión: si la diferencia de medias entre los grupos *i, j* es menor que $t_{\frac{\alpha}{2}} \cdot \sqrt{CM_D\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ no rechazamos la hipótesis nula, y en caso contrario la rechazamos. Nótese que hemos realizado una prueba $t$ de diferencia de medias, estimando la varianza común mediante $CM_D$. Puesto que en el cálculo de $CM_D$ hay $n$-$k$ grados de libertad, estos serán los utilizados.

**Comparaciones múltiples**

Cuando se debe realizar un número grande de comparaciones, crece la probabilidad de detectar como diferencias significativas algunas que no lo son realmente. El contraste anterior puede aplicarse para probar todas las posibles diferencias entre medias, cuando el número de éstas es pequeño.

Sin embargo, por ejemplo, si tenemos 10 pares de medias, es de esperar $0{,}05 * \binom{10}{2} = 0{,}05 * 45 = 2$ diferencias significativas simplemente por azar (Moses, 1992).

La solución al problema está en utilizar un método que permita efectuar una serie

de comparaciones, manteniendo un nivel de significación global $\alpha$ prefijado. Entre los diversos métodos que existen, expondremos el de Scheffé (Tejedor, 1999) quien considera un contraste entre los parámetros $\alpha_1, \alpha_2, \dots \alpha_k$, a una función lineal $\Phi = \sum \alpha_i * c_i = 0 \; y \; \sum c_i = 0$ , siendo los valores $c_i$ los coeficientes del contraste. Un contraste como el definido, admite como estimador insesgado el siguiente: $\widehat{\Phi} = \sum \bar{x}_i * c_i$ , y su varianza se estima mediante: $\sigma_{\Phi}^2 = CM_D \cdot \sum \frac{c_i}{n_i}$

Aceptada esta definición, puede mostrarse (Dunn y Clark, 1987) que existe una probabilidad $1 - \alpha$ que todos los contrastes satisfagan a la vez las desigualdades: $\widehat{\Phi} - \sigma_{\Phi} \cdot \sqrt{(k-1)F} \leq \Phi \leq \widehat{\Phi} + \sigma_{\Phi} \cdot \sqrt{(k-1)F}$. Siendo *F* el percentil del $(1 - \alpha/2) \cdot 100\%$ de la distribución *F* con *k-1*, *n-k* grados de libertad. Esta expresión se utiliza para calcular intervalos de confianza o determinar las regiones de aceptación y crítica al realizar las comparaciones múltiples.

### 1.7.7. ANÁLISIS DE VARIANZA DE VARIOS FACTORES

En ocasiones, los investigadores tendrán más de una variable independiente y necesitan modelos más complicados para ver el efecto de cada una y la forma como interaccionan. Explicaremos el caso más sencillo, análisis de varianza de dos factores, con efectos fijos (Molinero, 2003).

**Modelo lineal correspondiente**

En algunos modelos, podría encontrarse que la diferencia en la respuesta entre los niveles de un factor, no es la misma para todos los niveles del otro factor. Si esto ocurriese, se dice existe una *interacción*. Llamamos *interacción* entre dos variables al efecto de una de ellas cuando éste depende del nivel de la otra.

Si consideraremos dos factores *A* y *B* con efectos fijos y balanceado, se puede modelar la *k*-ésima observación de la celda (*i,j*) según la siguiente igualdad:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1,2, \dots, a \\ j = 1,2, \dots, b \\ k = 1,2, \dots, n \end{cases}$$

donde $\mu$ es una constante que representa la media global, $\alpha_i$ representa el efecto del nivel *i* del factor *A,* $\beta_j$ representa el efecto de nivel *j* del factor *B,* $(\alpha\beta)_{ij}$ representa la interacción del nivel *i* del factor *A* con el *j* del factor *B*, $\varepsilon_{ijk}$ es el término que dota al modelo de aleatoriedad. Se la considera $\varepsilon_{ijk}$ una variable aleatoria con una distribución $N(0, \sigma^2)$. Suponemos, se cumple que: $\sum \alpha_i = \sum \beta_j = \sum (\alpha\beta)_{ij} = 0$.

En este modelo, también es posible descomponer como en el anterior, las desviaciones entre los valores observados y la media global de la muestra en la forma:

$$y_{ijk} - \bar{y}_{\dots} = (y_{ijk} - \bar{y}_{ij\cdot}) + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\dots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\dots}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\dots}) \,,$$

en esta descomposición, cada uno de los sumandos tiene una interpretación, veamos:

$(y_{ijk} - \bar{y}_{ij\cdot})$ representa la desviación de una observación respecto a la media de su grupo, y se conoce como término del error; $(\bar{y}_{i\cdot\cdot} - \bar{y}_{\dots})$ es la desviación de la media del grupo con el nivel de *i* en el factor A, respecto a la media global, representando el efecto de dicho nivel; $(\bar{y}_{\cdot j\cdot} - \bar{y}_{\dots})$ es el efecto del nivel *j* del factor B. El último sumando es el efecto de la interacción (Tejedor, 1999).

**Cálculo del estadístico *F* e interpretación de resultados**

Con el fin de realizar el análisis estadístico del modelo bifactorial de efectos fijos, es necesario descomponer de la variación total (a través de la suma de cuadrados total, $SS_T$) como sigue (Dunn y Clark, 1987):

$$SS_T = SS_A + SS_B + SS_{AB} + SS_E$$

Tabla 1.7.2. Análisis de varianza para dos factores fijos

| Fuente de variación | Suma de cuadrados | Grados de libertad | Cuadrados medios | *F* |
|---|---|---|---|---|
| Factor A | $SC_A$ | $a-1$ | $CM_A = \frac{SC_A}{a-1}$ | $F_{obsA} = \frac{CM_A}{CM_E}$ |
| Factor B | $SC_B$ | $b-1$ | $CM_B = \frac{SC_B}{b-1}$ | $F_{obsB} = \frac{CM_B}{CM_E}$ |
| Interacción | $SC_{AB}$ | $(a-1)(b-1)$ | $CM_{AB} = \frac{SC_{AB}}{(a-1)(b-1)}$ | $F_{obsAB} = \frac{CM_{AB}}{CM_E}$ |
| Dentro de los grupos (residual) | $SC_E$ | $ab(n-1)$ | $CM_E = \frac{SC_E}{ab(n-1)}$ | |
| Total | $SC_T$ | $abn-1$ | | |

Cada una de las sumas de cuadrados representa una parte de la variabilidad total, descompuesta como suma de las variabilidades para cada uno de los factores presentes en el modelo (factor *A*, factor *B*, la interacción *AB* y la variabilidad residual). Además a cada suma de cuadrados se le asocia un número entero positivo que representa el grado de libertad (gl), esto es, la cantidad de elementos independientes que hay en dicha suma de cuadrados.

Así, la suma cuadrados total $SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{...}\right)^2$ para un modelo con *a* niveles del factor A, *b* niveles del factor B y balanceado, tiene *abn -1* grados de libertad (gl). Análogamente, $SS_A = bn\sum_{i=1}^{a}(\bar{Y}_{i..} - \bar{Y}_{...})^2$ , suma de cuadrados correspondiente al factor A, tiene $a-1$ grados de libertad; para el factor B es $SS_B = an\sum_{j=1}^{b}\left(\bar{Y}_{.j.} - \bar{Y}_{...}\right)^2$ y tienen $b-1$ grados de libertad, mientras que los correspondientes a la interacción AB son $(a-1)(b-1)$ sus grados de libertad.

Finalmente la suma de cuadrados del error resulta (Montgomery, 2005),

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \bar{Y}_{ij.}\right)^2 \quad \text{con} \quad ab(n-1) \text{ gl}$$

**Contraste de hipótesis relacionado con el factor *A***

En el análisis de varianza de dos factores fijos estamos interesados en contrastar varios grupos de hipótesis. Para esto, procedemos en primer lugar al cálculo de los elementos de la Tabla 1.7.2 (Dunn y Clark, 1987). Para comprobar el efecto del factor *A* se definen las siguientes hipótesis estadísticas:

$$H_{0A}: \text{todas las } \alpha_i \text{ son iguales a } 0 \quad vs \quad H_{1A}: \text{algún } \alpha_i \text{ es distinto de } 0$$

Calculamos para este contraste $F_{obsA}$ y rechazamos $H_{0A}$ con nivel de significación $\alpha$, en el caso de que su valor sea mayor que el percentil $\alpha \cdot 100\%$ de la distribución *F* cuyos gl son $a-1$ y $ab(n-1)$. En caso contrario, se decide no rechazarla.

**Contraste de hipótesis relacionado con el factor B**

Para comprobar el efecto del factor *B* se definen las siguientes hipótesis estadísticas:

$$H_{0B}: \text{todas las } \beta_i \; son \; iguales \; a \; 0 \quad vs \quad H_{1B}: algún \; \beta_i \; es \; distinto \; de \; 0$$

Calculamos para este contraste $F_{obsB}$ y rechazamos $H_{0B}$ con nivel de significación $\alpha$, en el caso de que su valor sea mayor que el percentil $\alpha \cdot 100\%$ de la distribución *F* cuyos gl son $b-1$ y $ab(n-1)$. En caso contrario, se decide no rechazarla.

**Contraste de hipótesis relacionado con la interacción**

Para comprobar el efecto de la interacción se definen las siguientes hipótesis estadísticas :

$$H_{0AB}: todas \; las \; (\alpha\beta)_{ij} \; son \; iguales \; a \; 0$$

$$H_{1AB}: algún \; (\alpha\beta)_{ij} \; es \; distinto \; de \; 0$$

Calculamos para este contraste $F_{obsAB}$ y rechazamos $H_{0AB}$ con nivel de significación $\alpha$, en el caso de que su valor sea mayor que el percentil $\alpha \cdot 100\%$ de la distribución *F* cuyos gl son $(a-1)(b-1)$ y $ab(n-1)$. En caso contrario, se decide no rechazarla.

En resumen, para saber aplicar el análisis de varianza de dos factores y efectos fijos, el estudiante tendrá que alcanzar las siguientes competencias y conocimientos:

- Diferenciar las situaciones en que hay que aplicar el análisis de varianza de dos factores con efectos fijos, respecto a otras situaciones.
- Comprender el modelo lineal que subyace.
- Comprender las hipótesis involucradas en el modelo.
- Llevar a cabo correctamente los cálculos de la tabla de análisis de varianza (Tabla 1.7.2) e interpretar los diferentes componentes de la misma.
- Interpretar correctamente los resultados para llegar a una decisión sobre las hipótesis.

## 1.7.8. ANALISIS DE VARIANZA DE MEDIDAS REPETIDAS

El procedimiento de análisis de varianza con medidas repetidas analiza grupos de variables dependientes que representan diferentes mediciones del un mismo atributo sobre los mismos sujetos. En general, aparece un factor intra-sujetos que indica las variables que contienen las medidas repetidas. Su relevancia en Psicología es debido a que es común que se analicen reacciones de tratamientos en distintos tiempos a un mismo individuo (Bono, Roser y Vallejo, 2008) con objeto de reducir la variabilidad del error.

**Análisis estadístico**

El análisis estadístico en estas pruebas es similar al realizado para el modelo bifactorial con efectos fijos, donde un factor son los tratamientos y el otro los sujetos. Se tiene una observación por tratamiento en cada sujeto, y el orden en que se realizan los tratamientos se determina al azar (Montgomery, 2005). Una diferencia con el

modelo bifactorial es que no interesa estudiar el efecto interacción. El modelo se puede escribir del siguiente modo:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,b \end{cases},$$

donde se tienen $a$ observaciones para cada sujeto. Las hipótesis de interés son:

$$H_0: \text{todas las } \alpha_i \; son \; iguales \; a \; 0 \quad vs \quad H_1: algún \; \alpha_i \; es \; distinto \; de \; 0$$

con $\alpha_i$ efecto del i-ésimo tratamiento.

En forma análoga a lo desarrollado para los otros dos modelos anteriores de análisis de varianza, llamemos $y_{i.}$ e $y_{.j}$ respectivamente a la suma total de los valores en las observaciones realizadas bajo el tratamiento *i* y a la suma total de valores de las observaciones sobre el sujeto j, siendo $y_{..}$ la gran suma total, y $N = ab$ el número total de observaciones registradas. De igual forma, $\bar{y}_{i.}$ es el promedio de los valores de las observaciones que se hacen bajo el tratamiento *i*, $\bar{y}_{.j}$ el promedio de los valores en las observaciones del sujeto *j*, e $\bar{y}_{..}$ el valor del gran promedio de todas las observaciones (Montgomery, 2005).

Tabla 1.7.3. Análisis de varianza para medidas repetidas

| Fuente de variación | Suma de cuadrados | Grados de libertad | Cuadrados medios | *F* |
|---|---|---|---|---|
| Tratamientos (IntraSujeto) | $SC_{Tratamiento}$ | $a-1$ | $CM_{Tratam} = \frac{SC_{Tratam}}{a-1}$ | $F_0 = \frac{CM_{Tratam}}{CM_E}$ |
| Sujeto (Intersujeto) | $SC_{sujeto}$ | $b-1$ | $CM_{Sujeto} = \frac{SC_{Sujeto}}{b-1}$ | |
| Residual) | $SC_E$ | $(a-1)(b-1)$ | $CM_E = \frac{SC_E}{(a-1)(b-1)}$ | |
| Total | $SC_T$ | $N-1$ | | |

Es necesario también aquí, descomponer la variación total como sigue:

$SS_T = SS_{InterSujeto} + SS_{IntraSujeto} + SS_E$. También es común utilizar la nomenclatura *tratamiento* en lugar de intersujeto, y *bloques* en lugar de intrasujeto. Procedemos al cálculo de la tabla del análisis de varianza, como en los casos anteriores, que está representada en la Tabla 1.7.3.

Cada suma de cuadrado en la Tabla 1.7.3 representa una parte de la variabilidad total.

Así, la suma cuadrados total se la puede computar como:

$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \bar{Y}_{..}\right)^2$, como hay N observaciones, esta suma de cuadrados tiene *N -1* gl.

Análogamente, $SS_{Tratamiento} = b\sum_{i=1}^{a}(\bar{Y}_{i.} - \bar{Y}_{..})^2$ suma de cuadrados del factor Tratamientos, tiene $a-1$ elementos independientes, por lo que sus gl son $a-1$; para el factor sujeto es $SS_{Sujeto} = a\sum_{j=1}^{b}\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$ con $b-1$ gl.

Finalmente $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$ es la suma de cuadrados del error, con $(a-1)(b-1)$ gl.

Para estudiar la hipótesis nula usamos el estadístico de prueba $F_0$ que tiene distribución *F* con $a-1$ y $(a-1)(b-1)$ gl cada vez que $H_0$ es verdadera. La hipótesis nula se rechazaría con nivel de significación $\alpha$ siempre que el valor

computado $F_0$ sea mayor que el percentil $\alpha \cdot 100\%$ de la distribución $F$ con $(a-1)$ y $(a-1)(b-1)$ gl. En caso contrario, se decide no rechazarla.

En consecuencia, para saber aplicar el análisis de varianza de medidas repetidas, el estudiante tendrá que alcanzar las siguientes competencias y conocimientos:

- Diferenciar las situaciones en que hay que aplicar el análisis de varianza de medidas repetidas, respecto a otras situaciones.
- Comprender el modelo lineal que subyace.
- Comprender las hipótesis involucradas en el modelo.
- Llevar a cabo correctamente los cálculos de la Tabla 1.7.3. e interpretar los diferentes componentes de la misma.
- Determinar cuál es la única hipótesis de interés presente en el problema.
- Interpretar correctamente los resultados para llegar a una decisión sobre esa única hipótesis de interés.

### 1.7.9. SUPUESTOS DEL ANÁLISIS DE VARIANZA

Además de elegir el modelo, e interpretar los resultados; los estudiantes deben comprender los supuestos de aplicación exigidos para poder emplear el análisis de varianza y saber cómo comprobarlos. Sabemos que los modelos nunca se ajustarán exactamente a la realidad y la descomposición de la variabilidad presente en las observaciones, expresada para todos los modelos presentados, no es más que una relación algebraica. Pero la aplicación de los contrastes requiere del cumplimiento de ciertos supuestos para verificar si el modelo ajustado describe de forma adecuada las observaciones (Tejedor, 1999).

Los supuestos de aplicación se pueden comprobar examinando los residuos (diferencia entre el valor observado y el ajustado con el modelo). Se recomienda realizar el examen de los mismos de forma automática; puesto que si el modelo es adecuado, en los residuos no deberá observarse ninguna estructura.

Si anotamos con $Y$ a la variable aleatoria, $y_{ij}$ sus observaciones, entonces el residual que se corresponde a la j-ésima observación del i-ésimo tratamiento lo definimos como:

$$e_{ij} = y_{ij} - \hat{y}_{ij}$$

donde $\hat{y}_{ij}$ es el valor ajustado mediante el modelo elegido, en otras palabras es una "estimación" de la observación $y_{ij}$ .

Además de comprobar que los residuos no tienen ninguna estructura particular, se deben comprobar los supuestos generales, que son los siguientes (Dunn y Clark, 1987): independencia, linealidad (adecuación al modelo), homocedasticidad (igualdad de varianzas), normalidad y esfericidad (para análisis de varianza de medidas repetidas). A continuación describimos cada uno de los supuestos mencionados.

#### Independencia

Este supuesto asume que cada una de las observaciones es independiente de las demás. Una posibilidad para comprobarlo es el gráfico de los residuos, realizado en el orden temporal en que se recogieron los datos; este gráfico se puede utilizar para

detectar correlaciones entre las medidas residuales. De existir correlación entre los residuos puede ser un problema grave, además de difícil de solucionar. Se deben tomar medidas preventivas, tal como la aleatorización adecuada en la realización del experimento.

**Linealidad (adecuación al modelo)**

Cada uno de los valores observados $y_{ij}$ puede descomponerse en varios términos que están relacionados linealmente, según el modelo de análisis de varianza elegido. En este supuesto se pide que el modelo elegido describa de forma adecuada a las observaciones, y que los errores tengan una distribución con media 0 y varianza constante pero desconocida. Un gráfico muy valioso para evaluar el incumplimiento de este supuesto es el gráfico de dispersión de residuos versus valores ajustados. Si en este gráfico aparecieran tendencias curvilíneas, no sería correcto el modelo utilizado.

**Homocedasticidad (igualdad de varianzas)**

La igualdad de varianzas u homocedasticidad supone una varianza común, notada $\sigma^2$ para todos los grupos. La prueba de Levene es una prueba de hipótesis estadística, aunque no la única, que lo comprueba. Se contrastaría la hipótesis nula que las varianzas de los grupos poblacionales son iguales, versus la alternativa que existe al menos un par de ellas que sean distintas.

Es posible también utilizar la gráfica de los residuos contra los valores ajustados por el modelo para analizar este supuesto, y esta gráfica no deberá mostrar algún patrón obvio. A menudo ocurre que la varianza de las observaciones aumenta conforme aumenta la magnitud de las observaciones; en este caso los residuos serían más grandes a medida que crece $y_{ij}$ (en el modelo de un solo factor de efectos fijos completamente aleatorizado), y la gráfica propuesta para analizar tendría forma de embudo con la boca hacia la derecha.

Si este supuesto no se cumple, la prueba *F* sólo se ve afectada levemente cuando se está en presencia de un modelo balanceado con efectos fijos; pero para otras opciones el problema podría ser grave. Esta es una razón por la que se opta, siempre que sea posible, por adecuar modelos con tamaños de muestras iguales para todos los niveles del factor. Para solucionar el problema de varianza no constante, es posible aplicar una transformación con la finalidad de estabilizarla, y nuevamente realizar el análisis usando los datos transformados (Montgomery, 2005). Generalmente la transformación otorga la ventaja que la distribución del error se acerque a la normalidad.

**Normalidad**

Otro supuesto para confirmar la adecuación del modelo de análisis de varianza, es la verificación de la normalidad de los residuos. Durante este proceso, es posible observar también la presencia de puntos atípicos que pudiesen introducir distorsiones serias en la aplicación del modelo. Estas anomalías suelen ponerse de manifiesto con el trazado de la gráfica de probabilidad normal de los residuos (verificación descriptiva, no confirmatoria), a través de un residual que sea mucho más grande que los demás, en el sentido de estar muy alejado del centro de esas observaciones; también por su forma, ya que debería ser al menos una línea recta en los valores centrales de la gráfica.

Una prueba estadística utilizada para la verificación de este supuesto es la

denominada de Kolmogorov-Smirnov, cuyo estadístico se conforma tomando las distancias entre la función de distribución empírica de los datos y una distribución teórica especificada (en este caso la Normal). Su objetivo es señalar si los datos provienen de una población que tiene la distribución teórica especificada.

Los modelos de efectos fijos, que son los que presentamos en este trabajo no se ven muy afectados por el incumplimiento de la normalidad, por ello se dicen modelos robustos con respecto a este supuesto, ya que solo se afecta ligeramente el verdadero nivel de significación en la prueba *F* y a la potencia (Montgomery, 2005); sin embargo, es primordial estar atentos a la presencia de puntos atípicos en el gráfico de normalidad.

**Esfericidad**

En el análisis de varianza con medidas repetidas, comprobar la esfericidad significa comprobar el supuesto de homogeneidad de varianzas para este modelo. La prueba estadística de Mauchly, es la que confirma estadísticamente la esfericidad. Para la hipótesis nula en este contraste, tendremos que todos los pares de varianzas para todos los niveles del factor son iguales, mientras que para la alternativa habrá al menos un par que no lo es. Si el supuesto se cumple, los datos podrán analizarse mediante el modelo de medidas repetidas, sin correr el riesgo de obtener un valor muy sesgado de la prueba *F*.

Cuando decimos que todos los pares de varianzas para todos los niveles de factor son estadísticamente igual, ocurre que la matriz de covarianza del error de variables dependientes transformadas es proporcional a una matriz identidad. Esto implica que la matriz de covarianza define una esfera, de allí el nombre a esta prueba.

Cada vez que se rechace la hipótesis de esfericidad deben corregirse los grados de libertad, multiplicándolos por un factor, épsilon, que mide el alejamiento de la esfericidad. Si esta fuese perfecta, épsilon sería igual a 1, y se puede mantener el modelo univariado de análisis de varianza. En caso de incumplirse este supuesto, se recomienda recurrir al análisis de tendencias en los niveles de las variables independientes en lugar de utilizar la prueba general de análisis de varianza, ya que este tipo de comparaciones no requieren de la esfericidad.

En consecuencia, para poder comprobar los supuestos del modelo análisis de varianza, el estudiante tendrá que alcanzar las siguientes competencias y conocimientos:

- Interpretar que los modelos análisis de varianza no se adecuan exactamente a la realidad de cada problema.
- Distinguir cada uno de los supuestos a verificar y cuál es su objetivo.
- Calcular adecuadamente los residuos resultantes de la aplicación de algún modelo.
- Realizar gráficas para comprobar los supuestos del modelo.
- Obtener los datos para elaborar las gráficas.
- Aplicar los test de Levene, Kolmogorov y Mauclhy adecuadamente a los datos, los residuos y a las varianzas del factor respectivamente.
- Interpretar los gráficos que resultan de la aplicación de un modelo, ya sea utilizando residuales como observaciones y ajustes.

## 1.8. DELIMITACIONES DE LOS CONTENIDOS DEL ESTUDIO

Como hemos indicado, en este trabajo nos centraremos en la evaluación de la comprensión del análisis de varianza elemental, sus supuestos, modelos e interpretación, así como de los prerrequisitos para llevarlo a cabo. El proceso de construcción del cuestionario para la evaluación de dichos conocimientos se recoge en el Capítulo 3 y seguirá las normas psicométricas habituales (APA, AERA y NCME, 1999).

El cuestionario se orienta a medir el constructo psicológico "comprensión de conceptos básicos del análisis de varianza". Como cualquier constructo psicológico, este también es inobservable en forma directa, de modo que ha de deducirse de los comportamientos de los individuos ante tareas que sean indicadores de dicha comprensión. La dificultad de su evaluación se deriva de que sólo pueden ser medido indirectamente, y puede haber diferentes aproximaciones a su medida (Osterlind, 1989). Por otro lado, esta medida debía estar basada en muestras limitadas de tareas y conductas; por tanto siempre hay un error, ligado al muestreo (Martínez Arias, 1995).

Tabla 1.8.1. Unidades de contenido de la variable objeto de medición

| | |
|---|---|
| Prerre-quisitos | - Lecturas de las tablas de la distribución y cálculo de probabilidades<br>- Diferenciar muestras independientes y relacionadas<br>- Distribución muestral del estadístico (Normal, Chi-cuadrado, *t* de Student y *F* de Fisher) y variabilidad del estadístico. |
| Intervalo de confianza | - Construcción de un intervalo de confianza<br>- Interpretación de un intervalo de confianza |
| Contraste de hipótesis | - Hipótesis; Asignación de hipótesis: diferenciar hipótesis nula y alternativa<br>- Región crítica y de aceptación<br>- Contraste unilateral y bilateral<br>- Valor p; nivel de significación, probabilidad error I<br>- Significación de un resultado<br>- Potencia, probabilidad de error II<br>- Regla de decisión |
| Análisis de varianza | - Tipos de variables en análisis de varianza: factor y sus tipos<br>- Modelo de un factor y efectos fijos<br>Modelo lineal asociado<br>Cálculo del estadístico *F*<br>Interpretación de resultados<br>- Modelo de dos factores y efectos fijos. Interacción<br>Modelo lineal asociado<br>Cálculo del estadístico F<br>Interpretación de resultados<br>- Modelo de medidas repetidas<br>Modelo lineal asociado<br>Cálculo del estadístico *F*<br>Interpretación de resultados<br>- Supuestos del análisis de varianza. Comprobación de supuestos<br>Independencia, linealidad, homocedasticidad, normalidad, esfericidad (caso medidas repetidas) |

En términos de Millman y Greene (1989) el cuestionario se sitúa en un dominio curricular, pues evalúa habilidades y conocimientos que se desarrollan como resultado

de la instrucción deliberada sobre un cierto contenido curricular; siguiendo a este autor es un instrumento de *medida centrada en el sujeto* pues permite evaluar a los alumnos en diferentes puntos de un continuo teórico, en función de sus respuestas al cuestionario. Las respuestas correctas pueden determinarse de modo objetivo y se refieren a un *dominio específico* (se trata de un *test referido a criterio*). Tiene también una finalidad *diagnóstica,* pues se trata de identificar dificultades en el aprendizaje de un grupo de sujetos, con el objetivo de llevar a cabo una intervención.

Para asegurar el cumplimiento de estos fines se requiere una descripción cuidadosa de las áreas de contenido que serán evaluadas (Osterlind, 1989; Morales, 2011). Ello nos ha llevado a realizar el análisis sistemático presentado en los párrafos anteriores, así como el análisis de las asignaturas cursadas por el alumno.

Como consecuencia de dicho análisis, en la Tabla 1.8.1 se recogen las unidades de análisis que definirán nuestro constructo y la lista de unidades de contenido, que conforman la definición semántica de la variable objeto de medición.

Dichas unidades determinan la definición semántica de la variable y se clasifican en los siguientes grandes grupos:

- *Prerrequisitos:* Se ha justificado la necesidad de que los alumnos manejen las distribuciones de probabilidad que intervienen en el análisis de varianza, así como el concepto de distribución muestral. Deben conocer tanto la distribución Normal como la distribución t de Student cuando construyan intervalos de confianza o un estadístico de prueba para un contraste. En cuanto a la *F* de Fisher es esencial a la hora tanto de construir el estadístico del análisis de varianza y en la comprobación del supuesto de igualdad de varianzas. Finalmente, la distribución Chi-Cuadrado está implícita en la *F,* por todo lo que no se incluye.
- *Intervalos de confianza*; puesto que el rechazar la hipótesis de igualdad de medias en el análisis de varianza debe acompañarse con la construcción e interpretación de intervalos de confianza para los efectos.
- *Contraste de hipótesis*; ya que el análisis de varianza contiene al menos un contraste de hipótesis. En particular prestamos atención a la asignación de hipótesis, la diferencia entre un contraste unilateral y bilateral, los tipos de errores y sus probabilidades, valor *p*, nivel de significación y potencia, las regiones críticas y de aceptación y la regla de decisión.
- *Análisis de varianza:* como nuestro objetivo es evaluar la comprensión de este objeto a nivel elemental, incluimos sólo modelos de efectos fijos: unifactorial, bifactorial y de medidas repetidas; analizamos los tipos de variables, los factores y sus tipos-muestras independientes y relacionadas presentes en estos modelos. En cada caso se analiza el modelo lineal correspondiente, el cálculo del estadístico *F* e interpretación de resultados, así como el/los contrastes de hipótesis que se le asocian. Creemos fundamental dentro de nuestro nivel del análisis considerar los supuestos del análisis de varianza. Finalmente analizamos los conocimientos que el estudiante deberá alcanzar a fin de saber aplicar cada modelo, y que tratamos de evaluar a través de nuestro estudio.
- *Lecturas de las tablas de la distribución y cálculo de probabilidades*: Siempre que los alumnos construyan un intervalo de confianza o realicen un contraste han de leer las tablas de la distribución y llevar a cabo cálculo de probabilidades y valores críticos. Por tanto, se puede considerar que este contenido está implícito en los otros.

- *Diferenciar muestras independientes y relacionadas:* Cada vez que los estudiantes deban decidir si aplicar un modelo de análisis de varianza con medidas repetidas o completamente aleatorizado (ya sea de uno ó de dos factores fijos); así como un contraste para muestras pareadas ó independientes, también al analizar el tipo de variables para los modelos, deberán examinar la diferencia entre estos dos tipos de muestras.
- *Distribución muestral del estadístico* (Normal, Chi-cuadrado, *t* de Student y *F* de Fisher) y variabilidad del estadístico. La distribución muestral del estadístico también puede ser evaluada, desde los contrastes de hipótesis ya sean para las prueba de una y dos muestras, como en el análisis de varianza (la prueba *F*, también los contrastes para probar supuestos)
- *Región crítica y aceptación y regla de decisión,* ya que los estudiantes para tomar la decisión han de conocer y obtener las regiones críticas y aceptación. Además también son evaluados estos contenidos al momento de interpretar los resultados de una tabla de análisis de varianza y decidir sobre el efecto de algún factor sobre los cambios en la respuesta.
- *Valor p, nivel de significación, probabilidad de error tipo I, potencia, probabilidad de error tipo II*, ya que el conocimiento de estas probabilidades condicionales es fundamental para la evaluación de las probabilidades de error en un contraste de hipótesis. Esta unidad también se evalúa a la hora de tomar la decisión cuando se contrasta una igualdad de medias en un modelo de análisis de varianza, así como en la toma de decisión para la prueba *F*.

La lista resultante (Tabla 1.8.1) resulta todavía un poco amplia, pues se desea que el cuestionario pueda ser contestado cómodamente por los estudiantes en un tiempo aproximado de dos horas. Por otro lado, algunas unidades están contenidas o pueden evaluarse desde otras, tal como lo explicamos a continuación:

- *Modelos en análisis de varianza*, cualquiera sean los modelos (uno/dos factores fijos, de medidas repetidas), los fundimos en una única unidad de contenido: selección de un modelo en análisis de varianza, ya que de acuerdo con las características de una problema, debería el estudiante poder adaptar cual es el modelo que mejor se le adecua para resolverlo y tomar una decisión final.
- *Supuestos de aplicación y comprobación de supuestos*, los hemos dividido en dos unidades de contenido, ya que el estudiante, deberá discriminar tanto cuales son los supuestos a comprobar para cada modelo, como seleccionar las técnicas estadísticas adecuadas a aplicar en cada caso.
- *Cálculo de Estadístico F*, lo hemos solo considerado para el caso de análisis de varianza con dos factores y medidas repetidas, ya que si el alumno puede demostrar competencia con estos modelos, para el caso de un factor (menos engorroso) tendrá la competencia también.
- Análogamente, y por la misma razón explicada en el punto último anterior, la *interpretación de resultados* la resumimos a la de la tabla de análisis de varianza para dos factores fijos, pues mostrando esa competencia para este modelo, también lo hará para los otros dos cuya comprensión estudiamos en este trabajo, pues tienen análogos procedimientos y argumentos.
- *Interpretación de resultados* con ordenador, decimos tres niveles de contenidos

relacionados con la interpretación de los resultados y usando las salidas obtenidas a partir de software, pues los estudiantes de nuestra muestra trabajan con SPSS, y a través de estos contenidos será posible tanto identificar modelos, corroborar supuestos y evaluar resultados.

En consecuencia, se obtiene la siguiente Tabla 1.8.2 simplificada, y es la que se usará para la construcción de ítems del cuestionario. Retomaremos estos contenidos en el Capítulo 3, donde se describe la construcción del cuestionario.

Tabla 1.8.2. Unidades de contenido de la variable objeto de medición

| Prerrequisitos |
|---|
| 1. Construcción de un intervalo de confianza |
| 2. Interpretación del significado de un intervalo de confianza. |
| 3. Contraste bilateral y unilateral. Asignación de hipótesis |
| 4. Hipótesis nula y alternativa. Asignación de hipótesis |
| 5. Errores tipo I y II, nivel de significación y potencia |
| 6. Cálculo de valor crítico, dado el nivel de significación |
| 7. Regla de decisión en un contraste |
| **Análisis de varianza** |
| 8. Selección de un modelo de acuerdo con las características de un problema |
| 9. Supuestos de aplicación del análisis de varianza. |
| 10. Comprobación de supuestos en un modelo |
| 11. Modelo lineal asociado a un tipo de análisis de varianza |
| 12. Cálculo en el análisis de varianza, Medidas repetidas |
| 13. Cálculo en el análisis de varianza. Modelo de dos factores |
| 14. Interpretación de una tabla de análisis de varianza. Modelo de dos factores |
| 15. Interpretación de salida de software. Identificación del modelo |
| 16. Interpretación de salida de software. Resultados |
| 17. Interpretación de salida de software. Supuestos del modelo |

# CAPÍTULO 2.
# FUNDAMENTOS TEÓRICOS

2.1. Introducción
2.2. Marco teórico
2.2.1. Introducción
2.2.2. Sistemas de prácticas operativas y discursivas ligadas a campos o tipos de problemas
2.2.3. Objetos intervinientes y emergentes de los sistemas de prácticas
2.2.4. Relaciones entre objetos: función semiótica
2.2.5. Evaluación de la comprensión
2.2.6. Idoneidad de un cuestionario de evaluación
2.3. Antecedentes de nuestra investigación
2.3.1. Introducción
2.3.3. Percepción de la variabilidad aleatoria
2.3.3. Comprensión de las distribuciones muestrales
2.3.3.1. Muestreo
2.3.3.2. Distribución muestral
2.3.3.3. Teorema central del límite
2.3.4. Investigaciones sobre comprensión del intervalo de confianza
2.3.4.1. Evaluación del aprendizaje en cursos universitarios
2.3.4.2. Trabajos apoyados por el enfoque onto-semiótico
2.3.4.3. Comprensión del intervalo de confianza por los investigadores
2.3.5. Investigaciones sobre comprensión de los contrastes de hipótesis
2.3.5.1. Controversias sobre el contraste de hipótesis
2.3.5.2. Errores sobre el nivel de significación y el valor p
2.3.5.3. Errores sobre el concepto de hipótesis
2.3.5.4. Otros errores en el contraste
2.3.6. Investigaciones sobre comprensión de diseño de experimentos y análisis de la varianza
2.3.6.1. Tipos de variables
2.3.6.2. Interacción
2.3.6.3. Análisis de pruebas de valoración de la comprensión del diseño experimental
2.3.6.4. Dificultades causadas por el software
2.4. Conclusiones del estado de la cuestión

## 2.1. INTRODUCCIÓN

A lo largo de este capítulo desarrollaremos los fundamentos que subyacen en nuestro trabajo de investigación. Primeramente, presentamos los elementos que utilizamos del enfoque onto-semiótico de la cognición matemática, que Godino y colaboradores vienen desarrollando desde 1994 y que es nuestro marco teórico. En primer lugar, describimos la teoría de significados personales e institucionales de los objetos matemáticos (Godino y Batanero, 1994, 1998). Además nos interesará la teoría de funciones semióticas (Godino, 2002; 2003; 2012; 2014; Godino, Batanero y Roa,

2005; Godino, Batanero y Font, 2007) pues uno de nuestros objetivos es la identificación de conflictos semióticos de los estudiantes en relación con nuestro objeto de estudio. Finalmente utilizamos los criterios de idoneidad de un cuestionario de evaluación que Batanero y Díaz (2005) adaptan de los criterios de idoneidad didáctica (Godino, 2003; 2011; Godino, Batanero, Rivas y Arteaga, 2013).

En la segunda parte de este capítulo presentamos los antecedentes de nuestra investigación, examinando investigaciones previas, que se relacionan con la comprensión del análisis de varianza y los objetos en los que este se apoya, en particular los relacionados con la comprensión de las distribuciones muestrales, del Teorema Central del Límite, los intervalos de confianza y los contrastes de hipótesis.

## 2.2. MARCO TEÓRICO

### 2.2.1. INTRODUCCIÓN

De acuerdo a Godino, Batanero, y Font (2007), para especificar y afrontar los problemas de investigación en didáctica de las matemáticas, son necesarias las teorías. En nuestro caso, hemos tomado el "enfoque onto-semiótico" (EOS) de la cognición matemática, propuesto por Godino y colaboradores (Godino y Batanero, 1994; 1998; Godino, 2002; 2003; 2012; 2014) como el más adecuado para nuestra investigación. Dicho modelo, nos da un punto de vista pragmático-antropológico, partiendo del papel clave que tiene la actividad de resolución de problemas.

Los autores han ido precisando y desarrollando las nociones de "significado institucional y personal de un objeto matemático" (entendidos ambos en términos de sistemas de prácticas en las que el objeto es determinante para su realización) y su relación con la noción de comprensión (Godino y Batanero, 1994; 1998; Godino, 1996; Godino, 2002; 2012; 2014). Su distinción entre dimensión institucional y personal del conocimiento es fundamental para nuestra investigación, donde partimos del análisis del significado institucional del objeto de estudio, pero, al observar las respuestas de los alumnos al instrumento de evaluación, analizamos los significados personales (sistemas de prácticas adquiridos por un sujeto).

Godino, Batanero y Font (2007) han pretendido elaborar una ontología suficientemente rica para describir la actividad matemática y los procesos de comunicación de sus "producciones", completando también la idea de función semiótica y de la ontología matemática asociada. El concepto de idoneidad didáctica también se usará para valorar la idoneidad del cuestionario de evaluación.

### 2.2.2. SISTEMAS DE PRÁCTICAS OPERATIVAS Y DISCURSIVAS LIGADAS A CAMPOS O TIPOS DE PROBLEMAS

Godino (2003) diferencia los siguientes tipos, dentro del significado institucional:

- *Referencial:* sistema de prácticas que se usa como referencia para una investigación o en la enseñanza y que será una parte del significado holístico del objeto matemático. Nosotros hemos presentado este significado referencial en el análisis llevado a cabo del objeto análisis de varianza en el Capítulo 1.
- *Pretendido:* Sistema de prácticas incluidas en la planificación del proceso de

estudio. Se puede deducir del análisis de los programas oficiales o de las guías docentes de los profesores que imparten el curso. En nuestro estudio viene marcado por las guías docentes analizadas de la asignatura y se resume también en la Tabla 1.5.1 donde definimos el contenido de la variable objeto de medición.

- *Implementado:* En un proceso de estudio específico es el sistema de prácticas efectivamente implementadas por el docente; en nuestra investigación son las prácticas relativas a las nociones básicas del diseño experimental propuestas por el docente en el seno de la institución. No se han analizado con detalle estas prácticas en nuestro estudio; en principio asumimos que no hay diferencia entre el significado pretendido y el implementado.
- *Evaluado:* El subsistema de prácticas que utiliza el docente para evaluar los aprendizajes. En nuestra investigación, se trata de los objetos elementales del diseño experimental que hemos evaluado a través del cuestionario elaborado. En los Capítulos 3 y 4 se incluyen unas tablas que especifican con detalle este significado evaluado con nuestros instrumentos.

Respecto a los tipos de significados personales, se consideran los siguientes:

- *Global:* corresponde a la totalidad del sistema de prácticas personales que es capaz de manifestar potencialmente un sujeto relativo a un objeto matemático. Sería el que nos interesaría evaluar, pero no nos es accesible.
- *Declarado*: da cuenta de las prácticas efectivamente expresadas a propósito de las pruebas de evaluación propuestas, incluyendo tanto las correctas como las incorrectas desde el punto de vista institucional. Este es el significado que analizamos en nuestra investigación, a través del cuestionario de evaluación implementado.
- *Logrado*: corresponde a las prácticas manifestadas que son conformes con la pauta institucional establecida. Este significado en nuestra investigación lo analizamos de acuerdo con el éxito manifestado en la elección de las respuestas a los ítems. Somos conscientes que de este modo solo obtenemos una muestra de los significados personales logrados por nuestros alumnos, que constituye una primera aproximación al constructo que nos interesa.

Figura 2.2.1. Sistemas de prácticas operativas y discursivas ligadas a campos o tipos de problemas (Godino, Font y Wilhelmi, 2008, p. 5)

Esta clasificación la podemos apreciar gráficamente a través de la Figura 2.2.1, donde en la parte central los autores, indican las relaciones dialécticas entre enseñanza y aprendizaje, lo cual da como consecuencia el acoplamiento progresivo entre los significados personales e institucionales.

### 2.2.3. OBJETOS INTERVINIENTES Y EMERGENTES DE LOS SISTEMAS DE PRÁCTICAS

En los sistemas de prácticas matemáticas, y, como consecuencia de éstas, intervienen y emergen objetos matemáticos cuya definición o propiedades necesitamos recordar. El EOS propone la siguiente tipología:

- *Lenguaje* (términos, expresiones, notaciones, tablas, gráficos) en sus diversos registros (escrito, oral, gestual). Para nuestra investigación, algunos ejemplos son los términos *intervalo de confianza, contraste de hipótesis, análisis de varianza, modelos, supuestos*. En cuanto a expresiones nos podemos referir a: *"aleatoriedad de la toma de datos para el diseño", "homocedasticidad de la muestra", o "interpretación de una tabla de análisis de varianza"*, entre otros ejemplos. Podemos hallar muchos ejemplos de notaciones en el Apartado 1.7, donde hicimos el estudio del significado de referencia. Respecto al lenguaje tabular y gráfico, en nuestro cuestionario hemos incluido ítems donde se pide analizar tablas (Ítem 13, Ítem 16) y responder de acuerdo a la observación de gráficos (Ítem 17).
- *Situaciones-problemas*: son las aplicaciones extra-matemáticas o matemáticas de las que surge el objeto. En nuestro trabajo contextualizan el conocimiento de los estudiantes y se trata de problemas de comparación de muestras en su mayoría a situaciones provenientes de la Psicología.
- Conceptos- definición (son introducidos mediante definiciones o descripciones). Ejemplos para nuestra investigación serán: parámetro muestral, parámetro poblacional, definiciones de error Tipo I, error Tipo II en un contraste de hipótesis.
- *Proposiciones* (son los teoremas y propiedades que ligan los conceptos); por ejemplo el Teorema Central del Límite, o la descomposición de la varianza en partes.
- *Procedimientos* (algoritmos, operaciones, técnicas de cálculo). En nuestra investigación los encontramos en la determinación de los intervalos de confianza, la aplicación de un contraste de hipótesis, y la determinación de los elementos que conforman una tabla de análisis de varianza.
- *Argumentos* usados para justificar o explicar las proposiciones y procedimientos, o las soluciones a los problemas. En nuestra investigación, el alumno debe argumentar en los problemas abiertos; por ejemplo, al tomar una decisión sobre rechazar o no una hipótesis o al interpretar los datos en el contexto del problema.

Creemos muy importante esta clasificación en los seis tipos de entidades primarias postuladas, pues amplían la tradicional que solo distingue entre conocimiento conceptual y procedimental en la enseñanza, a la vez que muestran la complejidad de la labor del profesor. Godino, Font y Wilhelmi (2008) indican que estos objetos se relacionan entre sí mediante configuraciones epistémicas (en el caso de las instituciones) o cognitivas (en el caso de las personas) (Figura 2.2.2).

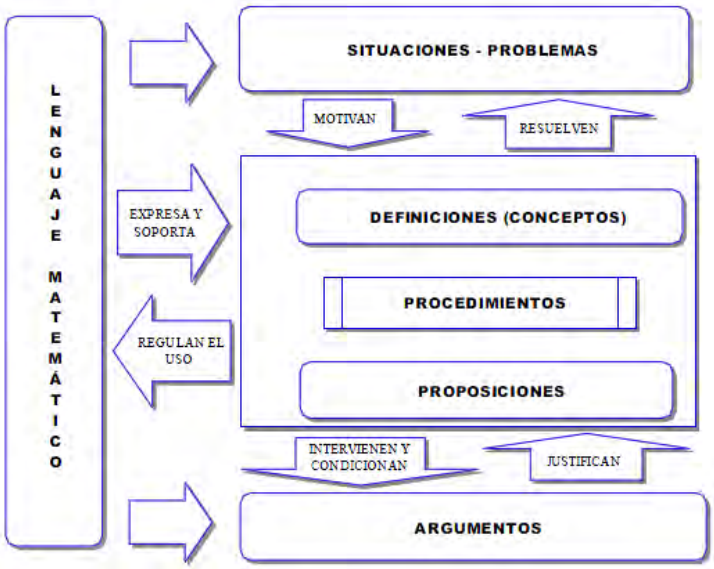



Figura *2*.2.2. Configuración de objetos primarios (Godino *et al*, 2008, p.7)

Además, todos estos objetos pueden ser considerados desde las siguientes facetas duales (Godino, 2002):

- *Personal—institucional.* Cuando se comparten en una institución, los objetos que emergen son objetos institucionales, mientras que cuando resultan de una práctica matemática específica de una persona emergen objetos personales.
- *Ostensivo – no ostensivo.* Cualquier objeto que es público y se puede mostrar a otro, se entiende que es ostensivo. Otros objetos poseen una naturaleza no ostensiva; pero se los utiliza en las prácticas públicas por medio de sus ostensivos asociados (una notación, un símbolo, una gráfica, una tabla, una figura, etc.).
- *Expresión – contenido:* Los objetos que intervienen en la actividad matemática, se relacionan a través de funciones semióticas. Se tratan de una relación entre un antecedente y un consecuente que un sujeto establece, en concordancia con algún criterio o alguna correspondencia. Cada tipo de objeto descripto puede ser el antecedente o el consecuente en una función semiótica.
- *Extensivo – intensivo*. Un objeto emergente de un sistema de práctica, puede ser considerado como un caso particular; por ejemplo un contraste de hipótesis para una media poblacional, ($H_0$: $\mu=0$ *vs* $H_1$: $\mu>0$) es un caso particular de otra clase mas general (en el ejemplo una familia de contrastes para la media poblacional). Esta dualidad es una característica fundamental y clave de la actividad matemática, es decir: el uso de elementos genéricos y su particularización a casos dados.
- *Unitario—sistémico*. Existen posibilidades en que los objetos matemáticos participan de la actividad matemática, como entidades simples (por ejemplo, el percentil de una distribución, la desviación estándar de una muestra). En otros casos se descomponen para su estudio. Por ejemplo, el estadístico de un contraste de hipótesis para la media poblacional puede definirse mediante una fórmula en el cual intervienen como componentes la media y desviación estándar de la muestra, así como su tamaño muestral.

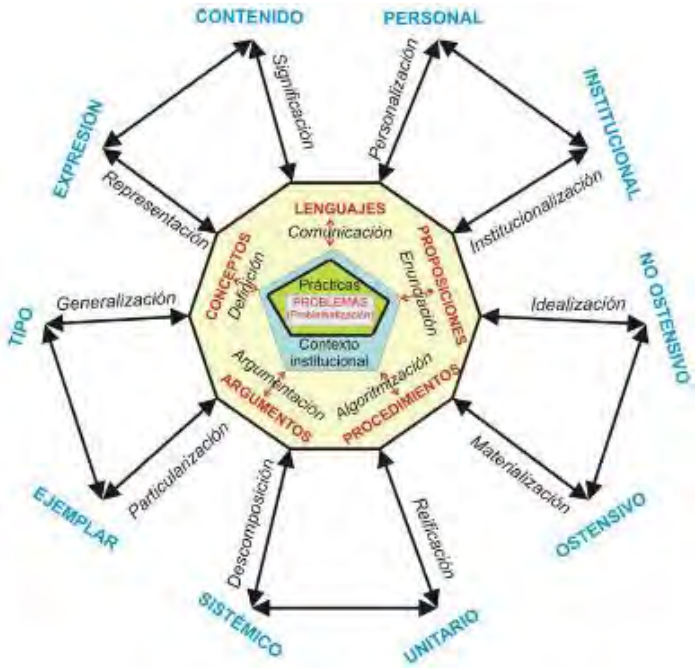



Figura 2.2.3. Configuración de objetos y procesos (Godino *et al.*, 2008, p. 8)

En la Figura 2.2.3 se representan estas nociones, formando la configuración de objetos primarios y secundarios, también analizadas desde la perspectiva proceso-producto. Tanto la intervención como la emergencia de los objetos de la configuración tienen lugar mediante los procesos matemáticos respectivos de comunicación, problematización, definición, enunciación, elaboración de procedimientos y argumentación. Además las dualidades dan lugar a una serie de procesos cognitivos /epistémicos de institucionalización/ personalización; generalización/ particularización; descomposición/ reificación; materialización/ idealización y representación/ significación (Godino et al., 2007).

### 2.2.4. RELACIONES ENTRE OBJETOS: FUNCIÓN SEMIÓTICA

Esta idea aparece en la faceta dual expresión-contenido. Los autores, la introducen a partir de trabajos de Hjelmslev (1971) y Eco (1979). Más precisamente, según Godino (2002) la función semiótica, pone en juego tres componentes: *un plano de expresión* (signo u objeto inicial); un *plano de contenido* (significado de tal signo, lo representado, objeto final); y un *criterio ó regla de correspondencia* que sirve para interpretar la relación entre ambos planos señalados. Los autores sugieren que cualquier tipo de objeto (situaciones-problemas, conceptos, proposiciones, procedimientos y argumentos), pueden participar en la función semiótica como expresión o contenido.

Por ejemplo, el contenido de la expresión "nivel de significación del 5%" es que en una prueba de hipótesis habrá aproximadamente 5 ocasiones en 100 en que se rechazaría la hipótesis nula cuando esta es cierta. El antecedente "nivel de significación" en este ejemplo es un elemento del lenguaje e informa los sujetos sobre cuáles son los objetos que se deben poner en correspondencia con el mismo (una propiedad que relaciona los conceptos de hipótesis nula y la distribución muestral del estadístico).

Las funciones semióticas, generalmente vienen dadas por su expresión, de manera que los otros componentes quedarían implícitos. Alguien debería hacer una posible interpretación de dicha expresión o signo, ya que este por sí mismo no explicita la correspondencia entre expresión y contenido. Cuando se produce un desacuerdo entre el significado que ha establecido el autor de la función semiótica y el que hace el interpretante de la misma se habla de *conflicto semiótico.* En particular, en la enseñanza o en la evaluación, cada vez que un estudiante hace una interpretación que no va de acuerdo con lo esperado desde la institución que imparte la enseñanza (de una expresión en un libro de texto, o una explicación del profesor, de un ítem de evaluación, etc.), se produce un *conflicto semiótico*. Esta interpretación permite explicar muchos de los errores y de las dificultades que se observan en el aprendizaje. En nuestra investigación trataremos de caracterizar algunos conflictos semióticos de los estudiantes en relación con el análisis de varianza, realizando un análisis semiótico con dos problemas abiertos (Capitulo 5).

### 2.2.5. EVALUACIÓN DE LA COMPRENSIÓN

Para analizar la comprensión que de un cierto objeto matemático poseen un grupo de sujetos, Godino (1996) propone un modelo que tiene en cuenta dos ejes: uno *descriptivo*, que indica los aspectos o componentes de los objetos de los cuales queremos evaluar su comprensión, y otro *procesual* que indica las fases o niveles necesarios en el logro de la comprensión. La comprensión personal de un determinado objeto por parte de un sujeto, no puede ser observada directamente, pero sí indirectamente a partir de las prácticas personales. Por lo tanto, en el enfoque onto-semiótico la evaluación de la comprensión sería el estudio de la correspondencia entre los significados personales e institucionales. Esta evaluación, por tanto, depende de la institución desde donde se lleva a cabo, ya que es la que determina si un sujeto "comprende" el significado de un objeto.

Por otro lado, al considerar a la funciones semióticas como esenciales en el proceso relacional entre entidades, permite entender la comprensión en términos de funciones semióticas (Godino, Batanero y Font, 2007). De este modo, la comprensión de un objeto por parte de un sujeto tiene en cuenta las funciones semióticas que es capaz de establecer. Así cada función semiótica constituye un conocimiento.

En esta investigación, y siguiendo los lineamientos del marco teórico que presentamos, inferiremos la comprensión personal de los alumnos que participan sobre el significado del análisis de varianza en estudiantes de Psicología mediante el análisis de la respuestas (entendidas como prácticas), que realiza la persona para resolver tanto ítems de evaluación propuestos como situaciones-problemas de respuesta abierta.

### 2.2.6. IDONEIDAD DE UN CUESTIONARIO DE EVALUACIÓN

En nuestro trabajo nos proponemos la construcción de un cuestionario. Es menester entonces, interesarnos hasta qué punto el cuestionario construido resulta eficaz para el resultado que se propone. Nuestro interés en este estudio es que los resultados no se limiten nada más a los alumnos de nuestra muestra, sino que estamos interesados en poder inferir nuestros resultados a otros estudiantes que realizaran la misma prueba en las mismas condiciones.

Si las tareas del cuestionario resultan lo suficientemente representativas, y nuestras unidades de contenido están bien definidas, podremos hacer tal inferencia sobre aquello que los alumnos sean capaces de hacer y decir en relación al objeto, en otras tareas (Batanero y Díaz, 2005). De este modo podríamos tener conocimiento generalizable aplicable a las dificultades y capacidades de otros alumnos con características similares a los de nuestra muestra. En este sentido, se aplica el concepto *idoneidad* y sus tipos (Godino, 2011; Godino, Batanero, Rivas y Arteaga, 2013) al caso de la evaluación (Batanero y Díaz, 2005):

- La *dificultad* de un ítem o tarea daría una medida de su *idoneidad cognitiva*; es decir del grado de representatividad de los significados evaluados respecto a los significados personales. En nuestro trabajo hemos obtenido ítems con un grado de dificultad que varía entre 0,93 (ítem 3) y 0,27 (ítem 2 e ítem 18), el 50% de nuestros ítems se encuentran entre una dificultad de 0,49 y 0,70 con lo cual la mayoría tiene una dificultad moderada; además podemos aceptar cómodamente una distribución normal (p-valor 0,731). Tenemos ítems aplicables a una gran variedad de conocimientos de los estudiantes desde los más fáciles a difíciles, abundando aquellos con dificultad intermedia.
- La *discriminación* de un ítem valoraría su *idoneidad evaluadora,* un ítem puede ser adecuado cognitivamente, pero no diferenciar (por ser demasiado fácil) los alumnos que tienen un mayor o menor conocimiento del concepto. Esta idoneidad podría ser un componente de la idoneidad *instruccional*, en cuanto uno de los objetivos de la instrucción es la función evaluadora.
- La *validez de contenido* de un cuestionario indicaría una *idoneidad epistémica*, o grado de representatividad del instrumento en cuanto al significado, objeto de evaluación.
- La *fiabilidad* o *generalizabilidad a otros ítems* daría una medida de la estabilidad de la respuesta, es decir sería otro componente de la *idoneidad evaluadora.*
- La *generalizabilidad a otros estudiantes,* sugeriría una *idoneidad generalizadora o externa* en cuanto a que los resultados se generalizarían a otros estudiantes.

## 2.3. ANTECEDENTES DE NUESTRA INVESTIGACIÓN

### 2.3.1. INTRODUCCIÓN

Las investigaciones que presentamos en este apartado están centradas sobre la enseñanza y el aprendizaje de los objetos estadísticos que se incluyen en el cuestionario de evaluación y que son básicos en la comprensión del análisis de la varianza, aunque no todos estos objetos han sido analizados en la investigación didáctica con la misma intensidad.

Para el estado de la cuestión que presentamos nos basamos en las síntesis previas incluidas en las tesis doctorales de Vallecillos (1994), Alvarado (2007), Díaz (2007), Olivo (2008), Tauber (2001) y Cañadas (2012). También utilizamos algunos trabajos de síntesis como los de Batanero (2000), Castro Soto *et al.*, (2007), Díaz, Batanero y Wilhelmi (2008) y Harradine, Batanero y Rossman (2011) que realizan una exploración muy exhaustiva de la literatura con la finalidad de reunir publicaciones que reporten estudios con evidencia empírica de conceptos erróneos en relación con la inferencia estadística en estudiantes universitarios. Finalmente analizamos trabajos recientes no

incluidos en los anteriores.

En lo que sigue comenzamos analizando las investigaciones previas sobre la idea de variabilidad, cuya comprensión subyace todo el razonamiento estadístico y en particular el análisis de varianza. Se sigue con el estudio de las distribuciones muestrales y el intervalo de confianza, que son prerrequisitos para la aplicación de las pruebas de hipótesis, elemento fundamental del análisis de varianza. Se continúa con una síntesis de la investigación sobre contraste de hipótesis, siendo éste el objeto inferencial que ha recibido más atención en esta investigación didáctica. Finalmente nos ocupamos de las pocas investigaciones que han tratado el análisis de varianza y el diseño experimental.

### 2.3.2. PERCEPCIÓN DE LA VARIABILIDAD ALEATORIA

Son muchos los autores que investigan sobre lo que denominan "variación" en estadística. De acuerdo con Reading y Shaughnessy (2004) la variabilidad es una propiedad de un objeto para cambiar o variar (variable), mientras que la variación sería la descripción o medida de cambio en la variable.

Nosotros utilizaremos acá el término variabilidad aleatoria; pues el término variación es más amplio, ya que incluye la variación determinista, que es sobre abundante, ya que la mayoría de los fenómenos naturales son manifestaciones del cambio (Estepa, 2013).

Estepa y Del Pino (2013) indican que la variabilidad aleatoria está ligada a los conceptos de incertidumbre y variable y analizan los diferentes significados del término variación en estadística. También estudian sus diferentes significados en relación a la variabilidad en la medida, los datos, el muestreo y las variables estadísticas. Además, dentro de la variabilidad aleatoria podríamos diferenciar la variabilidad en los experimentos aleatorios, en las variables aleatorias, en los datos de las muestras y las distribuciones univariantes o multivariantes. Generalmente, el concepto de dispersión, que estudia la variabilidad respecto a una medida de posición central, también se incluye dentro del de variación en las investigaciones citadas.

Garfield y Ben Zvi (2008) indican que la idea de variación puede ser aplicada a una innumerable cantidad de contextos y es un tema importante para la educación estadística. Además, esta idea es un componente clave para la comprensión del concepto de distribución, y es esencial para la toma de decisiones en estadística inferencial.

Recientemente se ha incrementado el interés en describir y medir la comprensión de la variabilidad aleatoria con sus diferentes componentes, a partir del *Third Internactional Research Forum on Statistical, Reasoning and Thinking* que se centró en el tema y en dos números especiales de la revista *Statistics Education Research Journal* (Garfield y Ben-Zvi, 2005).

Dicha variabilidad es también considerada como un objeto fundamental del pensamiento estadístico por parte de Wild y Pfannkuch (1999) quienes la incluyen en su modelo sobre este tipo de razonamiento. Estos autores proponen los siguientes componentes para su comprensión en la enseñanza:

1. Reconocimiento de la omnipresencia de la variabilidad aleatoria y de la necesidad de registrar esa variabilidad en los estudios estadísticos.

2. Medición y modelado de variabilidad en los datos para los fines de predicción, explicación, o control. Creación de resúmenes (numéricos o gráficos) para representar la variabilidad encontrada en los datos y uso de estos resúmenes para estudiar su impacto.
3. Explicar y controlar la variabilidad de los datos. Buscar los factores que la producen y considerar su impacto en el diseño y toma de muestras.
4. Utilizar estrategias de investigación en relación a la variabilidad (procedimientos formales para examinar sus propiedades). Uno de estos procedimientos sería precisamente el análisis de varianza, que permite diferenciar factores que contribuyen a la variabilidad en los datos y diferenciar la parte explicada (por estos factores) y no explicada (o aleatoria) de la variabilidad en los datos.

La desviación estándar es la medida más popular de la variación o dispersión, pero al mismo tiempo es una noción muy compleja y siendo un concepto bastante formal, su tratamiento se ha retrasado en los niveles secundarios y universitarios. Sin embargo, Lann y Falk (2003) indican que cuando los estudiantes de primer año de Universidad son interrogados acerca de la dispersión de los datos, una gran proporción de los mismos elige el rango antes que otra medida más formal para resumirla. Watson y Kelly (2003) exploraron la comprensión de la palabra variación entre 379 estudiantes entre 7º y 9º año con las siguientes preguntas: a) que significa variación, b) Use la palabra variación en una oración, y c) Escriba un ejemplo de algo que varía. Usando los datos y entrevistas a cuatro estudiantes (Torok y Watson, 2000) identificaron cuatro niveles de desarrollo del concepto de variabilidad de los estudiantes: débil apreciación de la variabilidad, apreciación aislada de aspectos de la variabilidad, apreciación inconsistente de la variabilidad y, buena apreciación de la variabilidad.

Reading y Reid (2005) desarrollan la anterior jerarquía en una nueva investigación con 33 estudiantes universitarios dentro de un curso de estadística. Los autores, presentan un informe centrado en el análisis de las respuestas a una pregunta que requiere que los estudiantes describan y comparen distribuciones. La mayoría de las respuestas (más del 95%), mostraron alguna evidencia de comprensión de la variabilidad. Los niveles considerados son los siguientes:

- *No consideración de la variabilidad*. Cuando el estudiante en su respuesta no relaciona la pregunta con algún tipo de variabilidad.
- *Débil consideración de la variabilidad.* Se reconoce la existencia de la variabilidad, pero la discusión se limita generalmente a una descripción básica, incorrecta o pobremente expresada. Las respuestas sugieren una falta de lenguaje y herramientas para describir o usar la variabilidad adecuadamente.
- *Consideración de la variabilidad en desarrollo.* En este nivel, las respuestas proveen una descripción más detallada y exacta de al menos una de las dos fuentes de variabilidad, a saber: entre grupos, y dentro de grupos.
- *Fuerte consideración de la variabilidad*. Las respuestas clasificadas en este nivel relacionan la variabilidad entre grupos y dentro de grupos, por lo cual preparan al estudiante hacia una comprensión intuitiva del análisis de la varianza.

En la Tabla 2.3.1 se presentan los niveles antes detallados con sus descriptores, estos permiten la codificación de respuestas a tareas personales, ó en relación a algún concepto estadístico particular, o bien aplicar la jerarquía a un rango más amplio de

tareas. Reading y Reid encontraron en su investigación pocas respuestas que pudiesen ser codificada como consideración *fuerte de la variabilidad,* y por lo tanto la información de los descriptores podría ser insuficiente, por lo que su jerarquía necesita ser ampliada, y ver su eficacia en tareas provenientes de clases de estadística más avanzada.

Tabla 2.3.1. Jerarquía de comprensión de la variabilidad. Reading y Reid (2005)

| |
|---|
| *No* consideración de la variabilidad<br>- Declaraciones de carácter general que no muestran ninguna consideración significativa de la variabilidad |
| *Débil* consideración de la variabilidad<br>- Identifica características de una sola fuente de variabilidad (entre grupos o dentro de los grupos);<br>- Reconoce la variabilidad en relación a otros conceptos<br>- Describe incorrectamente la variabilidad<br>- No se basa en los datos para describir la variabilidad<br>- Las expresiones usadas para describir la variabilidad pueden ser pobres<br>- Sugiere factores irrelevantes para explicar la variabilidad<br>- No usa variabilidad para apoyar la inferencia.<br>- No relaciona la variabilidad dentro de los grupos y la variabilidad entre grupos |
| Consideraciones de la variabilidad *en desarrollo*<br>- Claramente describe ambas variaciones, tanto entre como dentro de los grupos<br>- Reconoce la variabilidad en relación a otros conceptos<br>- Describe correctamente la variabilidad<br>- Se basa en los datos para describir la variabilidad.<br>- Prevé razonablemente la cantidad de variabilidad<br>- Se refiere correctamente a los factores que explican la variabilidad<br>- Usa la variabilidad para apoyar la inferencia<br>- No relaciona la variabilidad entre grupos y dentro de grupos |
| *Fuerte* consideración de la variabilidad<br>- Relaciona la variabilidad entre grupos y dentro de grupos para confirmar una inferencia |

Otras investigaciones relacionadas con la idea de variabilidad se describen en Sánchez, Silva y Coutinho (2011), quienes también describen las dificultades de los profesores con este concepto y su falta de discriminación respecto al de dispersión. Además, algunos profesores en su estudio no discriminan las situaciones en que es conveniente utilizar la varianza frente a otra medida de dispersión como el rango.

### 2.3.3. COMPRENSIÓN DE LAS DISTRIBUCIONES MUESTRALES

Un prerrequisito para comprender el análisis de la varianza, así como los intervalos de confianza y el contraste de hipótesis, es el muestreo y las distribuciones muestrales, tema sobre el que resumimos a continuación los estudios previos.

#### 2.3.3.1.MUESTREO

Moses (1992) afirma que la comprensión de que una muestra puede proveer información sobre la población de la que ha sido tomada es una idea clave del razonamiento inferencial. Comprender esta propiedad implica aceptar la

representatividad de las muestras y la variabilidad de las mismas, pero la investigación ha mostrado que la variabilidad es pobremente comprendida.

Un ejemplo de ello es la heurística de la representatividad (Kahneman, Slovic y Tversky, 1982), que consiste en enfatizar sólo la representatividad y prescindir de la variabilidad del muestreo, teniendo demasiada confianza en las pequeñas muestras. Estos autores describen la "ley de los pequeños números", o creencia en la que pequeñas muestras serían representativas en todas sus características estadísticas de las poblaciones de donde proceden. Este error puede llevar a sobreestimar la potencia de los métodos estadísticos, subestimar la amplitud de sus intervalos de confianza o esperar replicabilidad de experimentos realizados con pequeñas muestras. Otros errores relacionados con la idea de muestreo incluyen tomar una muestra de tamaño insuficiente o interpretar resultados de muestras no aleatorias como si el muestreo lo hubiese sido. (Díaz, Batanero y Wilhelmi, 2008).

### 2.3.3.2.DISTRIBUCIÓN MUESTRAL

Harradine, Batanero y Rossman (2011) afirman que la inferencia estadística consiste en tres elementos distintos (que a la vez interactúan): a) el razonamiento sobre el proceso, b) los conceptos y c) los cálculos, pero que las principales dificultades en la comprensión de la inferencia estadística caen en los dos primeros elementos.

En cuanto a los conceptos, indican que las distribuciones muestrales son más abstractas que la distribución de una población o de una muestra, por lo que suelen ser muy difíciles de comprender para los estudiantes. Además, razonar acerca de distribuciones muestrales requiere que los estudiantes integren varios conceptos estadísticos y ser capaces de prever los comportamientos de futuras muestras. Aconsejan como estrategia para ayudar a los estudiantes a comprender este concepto, hacerlos participar en actividades que involucren la toma repetida de muestras aleatorias de una población.

Schuyten (1991) indica la dificultad que supone el uso de diversos niveles para el mismo concepto, por ejemplo, la media en inferencia. En estadística descriptiva se usa la media $\overline{x}$ de una muestra, que es un valor numérico constante. En inferencia, se introduce la media teórica o esperanza matemática  de *la población concreta,* y la media $\bar{X}$ de la muestra particular con la que estamos trabajando, y es una variable aleatoria, un caso particular de todas las medias de las posibles muestras de tamaño dado que podríamos extraer de la población de interés. Al considerar la población de todas las medias se considera la distribución muestral que tiene una media o esperanza matemática: $E(\bar{X})$. Para el alumno diferenciar entre estos niveles de medias supone una gran dificultad conceptual (Batanero, 2000).

Son muchas las investigaciones que dan cuenta de esta dificultad. DelMas, Garfield y Chance (1999) diseñaron el software *Sampling Distribution* para trabajar las distribuciones muestrales. Con este software, guiaron a los alumnos con diversas actividades usando poblaciones con distribución normal u otras diversas formas, como U, asimétrica o uniforme. En su experimento, observaron que los estudiantes cambiaban la forma de la distribución teórica de la población sin inconvenientes, y podían simular la distribución muestral de diversos estadísticos, incluso con varios tamaños de muestras. El software incluía un programa que permitía poder ver la distribución y los estadísticos de cada muestra. Se pidió estudiar las siguientes cuestiones:

- Relación entre tamaño muestral y dispersión de la distribución muestral;
- Determinar a partir de qué tamaño muestral los estadísticos de la distribución comienzan a estabilizarse;
- Analizar si los estadísticos obtenidos son buenos estimadores de los parámetros de la población;
- Estudiar si el comportamiento de los estadísticos de la distribución muestral, concuerda con lo previsto por el teorema central del límite.

Los resultados del experimento mostraron el aprendizaje de los estudiantes, aunque un número significativo de alumnos parecía no comprender las implicaciones básicas del teorema central del límite.

En otro trabajo posterior Chance, delMas y Garfield (2004) puntualizan que los estudiantes aprenden mejor cuando las actividades están estructuradas de tal manera que los ayuden a evaluar la diferencia entre sus propias creencias y los resultados empíricos. La simulación, en general, mejoró la comprensión de los estudiantes; sin embargo, el ordenador añade nueva información en el aprendizaje y esto requiere un coste y esfuerzo añadido por parte de los estudiantes.

Lipson (1997), trabaja las distribuciones muestrales con dos grupos de alumnos usando los programas *Minitab* y *Sampling Laboratory*. En la primera sesión de su experiencia utilizan un ejercicio de muestreo con una caja con bolas blancas y negras, y en forma tradicional se introdujo las distribuciones en el muestreo, obteniendo finalmente el histograma adecuado. A partir de allí se dividió a los alumnos en forma aleatoria en dos grupos que usarían los programas señalados anteriormente, y responderían a algunas cuestiones después de dos sesiones más, durante las cuales realizaron simulaciones con el computador, analizando distribuciones muestrales con diferentes tamaños de muestra.

La autora analiza los mapas conceptuales que hicieron los alumnos antes y después de las simulaciones con ordenadores con la ayuda de expertos, permitiendo observar si había habido alguna evolución en el aprendizaje después del uso de ordenadores. La autora no encontró grandes diferencias entre los estudiantes que utilizaron distintos software, pero las actividades de simulación mostraron la mejora de la comprensión de muchos de ellos.

### 2.3.3.3.TEOREMA CENTRAL DEL LÍMITE

El Teorema Central del Límite justifica el uso de la distribución normal en inferencia. Castro Sotos el al. (2007) indican que pocos estudiantes comprenden que la variabilidad de la media muestral tiene a cero cuando el tamaño muestral tiende a infinito. Estos estudiantes no verán el tamaño muestral como un factor que afecte la validez de las conclusiones de la inferencia, o a la generalización de la muestra hacia la población.

Un resumen de las investigaciones sobre el teorema central del límite lo encontramos en Alvarado (2007). El autor, nos remite a Méndez (1991) quien recomienda contar con material manipulable para obtener representaciones de procesos muestrales y de la distribución muestral de la media. También sugiere el uso de ordenadores para realizar simulaciones, pudiendo propiciar así la comprensión intuitiva

antes de la formal. Méndez identifica cuatro propiedades elementales que deben comprenderse para llegar a entender en profundidad el teorema:

1. La media de la distribución muestral es igual a la media de la población, e igual a la media de una muestra cuando el tamaño de la muestra tiende al infinito;
2. La varianza de la distribución muestral es menor que la de la población (cuando $n>1$);
3. La forma de la distribución muestral tiende a ser acampanada a medida que se incrementa el tamaño muestral, y aproximadamente normal, independientemente de la forma de la distribución en la población;
4. La forma de la distribución muestral crece en altura y decrece en dispersión a medida que el tamaño muestral crece.

Alvarado (2007) realiza un análisis histórico de la evolución del teorema central del límite y un estudio de 16 textos dirigido a estudiantes de ingeniería. Describe a partir de ello nueve campos diferentes de problemas de donde surge el teorema. Indica que los libros de textos carecen de la mayoría de los campos históricos del teorema central del límite, teniendo prioridad los campos de problemas indirectos.

Basándose en el anterior estudio, Alvarado diseña un experimento de enseñanza seleccionando los elementos de significados más adecuados del teorema para la construcción de su propuesta didáctica, organizándolos en un proceso de estudio pretendido, que ponen en contexto en ejemplos vinculados con la ingeniería. Su experiencia de enseñanza incorpora recursos computacionales a la acción didáctica con el uso de la hoja Excel y el programa estadístico @risk para simular distribuciones de probabilidad en el estudio de algunas situaciones de riesgo en la ingeniería.

Como resultado de este experimento de enseñanza compara los elementos del significado institucional pretendidos con los significados implementados y adquiridos por los estudiantes. Estos mostraron una buena comprensión de la aproximación normal a la distribución binomial. Tampoco tuvo mayor dificultad el lenguaje simbólico, con algunas excepciones en los problemas algebraicos abiertos. El procedimiento algebraico y la tipificación fueron los mejor utilizados en general, así como la simulación con apoyo de Excel. Se destaca la apropiación de las propiedades de la media y varianza de la suma de variables aleatorias.

Algunas dificultades que se observaron fueron: la identificación de la distribución de probabilidad que modela la variable estadística; expresión verbal incorrecta en una gama de conceptos relacionadas con el teorema, errores de estandarización al confundir la suma de variables aleatorias con el promedio, no aplicar la corrección de continuidad y no ser capaces de dar una decisión en base al análisis del teorema central del límite. Asimismo, los estudiantes erróneamente lo extrapolan a situaciones inadecuadas.

### 2.3.4. INVESTIGACIONES SOBRE COMPRENSIÓN DEL INTERVALO DE CONFIANZA

Las investigaciones sobre el intervalo de confianza las podemos dividir en dos apartados: (a) las que se centran en la interpretación y uso de intervalos de confianza por parte de los investigadores y b) las que tratan de evaluar el aprendizaje de estudiantes, dentro de cursos universitarios.

### 2.3.4.1.COMPRENSIÓN DE LOS INTERVALOS DE CONFIANZA POR INVESTIGADORES

Ante las numerosas críticas al uso exclusivo de los contrastes de hipótesis en la investigación, varios investigadores sugieren que dichos contrastes han de estar acompañados por intervalos de confianza de los parámetros estudiados (Díaz, 2007). Se supone que los intervalos de confianza, entre otros beneficios, pueden ayudar a los investigadores a superar sus concepciones erróneas asociadas a la significación estadística.

Cumming, William y Fidler (2004) se interesan por analizar si la facilidad de comprensión de los intervalos es sólo aparente, y para ello estudian sistemáticamente los errores en la interpretación de intervalos de confianza de los investigadores que han publicado artículos en los que usan dichos intervalos. Uno de los puntos en los que se interesaron fue la predicción de cómo sería un intervalo de confianza si se replica el experimento, observando que muchos investigadores piensan (erróneamente) que hay una alta probabilidad de que el parámetro caiga de nuevo en el intervalo de confianza original, cuando esta probabilidad es menor de lo supuesto por estos investigadores. Los autores observan que este error conceptual es consistente con la intuición que se tiene de la *ley de los pequeños números*, donde se subestima la variabilidad proveniente del muestreo.

Continuando el trabajo anterior, Belia, Fidler y Cumming (2005) pidieron a algunos investigadores mediante una encuesta hecha por correo electrónico que juzgaran cuando eran significativamente diferentes las medias de dos grupos independientes, a partir de la interpretación de los intervalos de confianza presentados gráficamente, observando que sólo unos cuantos sujetos se acercaron a la respuesta correcta.

Un grupo de los que contestaron interpretaron los resultados como si se refiriesen a una sola población, aunque, ni siquiera para datos apareados la respuesta es correcta, porque el intervalo de confianza de la diferencia de medias no es igual a la diferencia de dos intervalos de confianza por separado. Otro grupo resolvió el problema como si se tratase de dos medias independientes. Además muchos investigadores creían que dos intervalos de confianza del 95% de confianza distinto para la misma media no se pueden solapar. También pensaban que cuando dos intervalos para medias independientes no se solapan, ello implica una diferencia significativa en el nivel de 0,05 entre las medias, y que el solapamiento de los dos intervalos implica que no existe una diferencia significativa.

### 2.3.4.2. EVALUACIÓN DEL APRENDIZAJE EN CURSOS UNIVERSITARIOS

Son varias las investigaciones que encontramos en este apartado. La primera de ella, desarrollada por Behar (2001), quien prepara un cuestionario para valorar la comprensión del intervalo de confianza, que aplica a un grupo de expertos en estadística ($n$=47), y otro a un grupo de estudiantes universitarios ($n$=297). Señala la falta de comprensión de la manera como se relacionan los distintos objetos asociados con un intervalo de confianza; en especial el ancho del intervalo y el nivel de confianza. Se cree que no es posible disminuir la confianza del intervalo, como si esta fuese inseparable del ancho del mismo. Otro error básico es pensar que los valores que constituyen un intervalo de confianza, se refieren a la variable aleatoria o al estadístico que se usa como

estimador y no al parámetro en estudio.

Una buena proporción de los participantes de ambos grupos, interpreta incorrectamente el coeficiente de confianza, pues suponen que da la probabilidad de que el parámetro se encuentre en el intervalo. En realidad la verdadera definición es el porcentaje de intervalos calculados a partir de muestras de igual tamaño en la población que contiene al parámetro. Olivo (2008) sugiere, no se asocia el nivel de confianza con la frecuencia relativa a largo plazo con la que los intervalos generados aleatoriamente contengan al verdadero parámetro de la población.

delMas (2001) realiza un estudio centrado en el intervalo de confianza con el software antes descrito. Para esto uso un pretest, con el cual se acercaba al conocimiento previo requerido antes de usar la herramienta, sirviéndole como una medida del diagnóstico. Utilizando el programa realiza una enseñanza del intervalo de confianza, con 82 preguntas que organizan en ocho grupos. Cada grupo de preguntas comienza con la población normal con media $\mu$ y $\sigma$ conocida. En la primera parte se simula la selección de una muestra de tamaño 16. Se toman luego otras muestras de la misma población, se registran los valores de la media muestral, la desviación estándar, el ancho de cada intervalo y el margen de error. Luego se repite el proceso con otra familia que no sigue una distribución normal, y con cada muestra el estudiante calcula los límites inferior y superior, y también hace lo propio el software. En cuanto al postest, contiene además de cuestiones análogas al pretest, otra que se relacionan con el software, y problemas de aplicación, observando el aprendizaje de los estudiantes.

Garfield, delMas, y Chance (2004), desarrollan un módulo sobre el intervalo de confianza. Su investigación está basada en un modelo de cambio conceptual, ya que los estudiantes prueban sus predicciones usando simulación con un software.

Indican que los estudiantes han de tener los siguientes conocimientos previos para comprender el intervalo de confianza: población, muestra, parámetro, estadístico, variabilidad del muestreo/ error del muestreo, error estándar, distribución muestral. Han de entender que las estimaciones provienen de muestras, y el verdadero valor puede únicamente ser obtenido conociendo toda la población. Esas estimaciones son cercanas al parámetro poblacional, pero no exactas. Además, deben comprender las siguientes ideas sobre los intervalos:

- Un intervalo de confianza para la media poblacional es la estimación de un parámetro poblacional desconocido (la media), basado en una muestra aleatoria de la población.
- Es un conjunto de valores plausibles del parámetro (μ) que pudiera haber generado el dato observado como un posible resultado.
- Un intervalo de confianza para la media poblacional se calcula mediante un estadístico muestral más o menos una medida del error del muestreo.
- El nivel de confianza nos da la probabilidad de que el método construya un intervalo que incluya el parámetro desconocido. La probabilidad se relaciona con el método (datos, intervalo), no con el parámetro.
- Un incremento en el tamaño de la muestra conduce a decrecer el ancho del intervalo: grandes muestras tienen intervalos más estrechos que muestras más pequeñas (si todos los otros elementos permanecen constantes).
- Altos niveles de confianza tienen intervalos más anchos que niveles de confianza más bajos (todos los otros elementos permaneciendo constantes).

- Intervalos estrechos y altos niveles de confianza son deseables, pero estas dos cosas se afectan una a la otra.
- Si se recogen muchas muestras aleatorias independientemente de una población y se construyen los intervalos de confianza del 95% para cada muestra, pudiéramos esperar que alrededor de un 5% de los intervalos no incluyan la media poblacional (el parámetro poblacional). Este 95% se refiere al proceso de tomar muestras repetidas y construir intervalos de confianza para cada muestra.
- Después de calcular un intervalo de confianza, el parámetro puede estar incluido o no, pero eso no se sabe.

También señalan las siguientes concepciones erróneas frecuentes:

- Hay un 95% de probabilidad de que el intervalo de confianza incluya la media muestral.
- Hay un 95% de probabilidad de que la media poblacional estará entre los dos valores (límite superior y límite inferior).
- 95% de los datos están incluidos en el intervalo.
- Un intervalo ancho significa menos confianza.
- Un intervalo de confianza más angosto es siempre mejor (a pesar del nivel de confianza)

En un estudio con 180 alumnos universitarios Fidler y Cumming (2005) tratan de ver si relacionan los intervalos con los contrastes de hipótesis y si saben interpretar su significado práctico en artículos publicados en revistas científicas. Para ello les preguntó a un grupo de estudiantes de ciencias (ultimo año de estudios y postgrado) si los intervalos y el valor *p* en los contrastes proporcionan o no evidencia empírica a favor de la hipótesis nula o de la alternativa. Observan que el número de interpretaciones incorrectas es mayor en los contrastes de hipótesis, aunque alrededor de un 20 % (n=55) de alumnos en su estudio también interpretan incorrectamente un intervalo de confianza. Por ejemplo, algunos alumnos suponen que el intervalo se refiere a los valores de la media de la muestra o a los valores individuales de la variable, ignorando su naturaleza inferencial.

Los autores, sugieren que el uso del intervalo de confianza ayudaría a disminuir errores conceptuales, tales como la creencia que un resultado estadísticamente no significativo es equivalente a que no haya efecto. En su investigación, pidió a un grupo de estudiantes que interpretaran resultados con el formato de intervalo de confianza, y la misma pregunta con el formato de contraste de hipótesis. Resultó mayor la cantidad de participantes que respondieron correctamente con el primer formato que con el segundo (24 frente a 10 de un total de 55 participantes).

Fidler y Cumming también advierten sobre ideas equivocadas de los estudiantes de cómo se relacionan entre si las distintas componentes de los intervalos de confianza. Frente a la pregunta elemental acerca de la relación entre el ancho del intervalo y el tamaño de la muestra, solo un 16% pudo contestarla correctamente.

### 2.3.4.3. TRABAJOS APOYADOS EN EL ENFOQUE ONTO-SEMIÓTICO

La contribución más importante que hemos encontrado es la investigación de Olivo (2008) que abarca diversos aspectos, utilizando también el enfoque onto-semiótico. Comienza con un estudio histórico, donde muestra que la primera aproximación a la idea de intervalo surge de Bayes y de su discípulo Laplace, quien estudia la estimación de la proporción y la media a partir del teorema de Bayes, llegando a una idea (no explícita) de lo que sería el intervalo de credibilidad y de este modo aparecen los fundamentos de la actual inferencia bayesiana. La alternativa de los intervalos fiduciarios introducidos por Fisher, no generó éxito, pues no brinda una distribución sobre los parámetros desconocidos, aunque, a través de ellos es posible determinar los valores de verosimilitud de los parámetros. También indica que la verosimilitud y la máxima verosimilitud, representan hoy las ideas principales sobre las que se apoya la teoría de estimación. Destaca que

> los intervalos de confianza tal como se los conoce en la actualidad, surgen luego de la resistencia de muchos investigadores a aceptar las probabilidades subjetivas, que son básicas para el procedimiento bayesiano (Olivo, 2008, p. 268).

Este estudio servirá para justificar uno de los conflictos encontrados en los estudiantes en su investigación, que consiste en dar una interpretación bayesiana al intervalo y al coeficiente de confianza; interpretación que se presentó en un 36,5% ($n$=252) de la muestra estudiada.

Un segundo punto en la investigación de Olivo (2008) es analizar y clasificar elementos de significado de los intervalos de confianza, en una muestra de libros de texto universitarios. Describe una gran diversidad de campos de problemas, acompañados con una gran variedad de lenguaje, argumentos y propiedades. El autor se ha basado en estudios de textos universitarios de estadística (siguiendo el método de Tauber, 2001 y Alvarado, 2007), y ha confirmado conflictos de significado en cuanto a la definición del objeto de estudio; además en un muy alto porcentaje de libros se da prioridad a las definiciones y propiedades y no se tratan las aplicaciones.

El tercer punto abordado en el trabajo de Olivo (2008) es la construcción de un instrumento válido para evaluar la comprensión del intervalo de confianza, luego de recibir instrucción. En la construcción del mismo utiliza el método sugerido por Batanero y Díaz (2005). Analizando las respuestas a los ítems abiertos, destaca un elenco de conflictos semióticos, observados en los alumnos de las carreras de ingeniería, que resumimos a continuación. Los clasificamos dentro de cada tipo de objeto matemático primario emergente de los sistemas de prácticas:

- *Definición, propiedades y conceptos:* No se comprende que el coeficiente de confianza da un porcentaje de intervalos tomados en las mismas condiciones que contienen al parámetro (confirmando los resultados Behar, 2001). Hay una interpretación Bayesiana del intervalo, resultado aportado por la investigación y también detecta que el alumno confunde estadístico con parámetro. Del mismo modo, confunde varianza poblacional y muestral de acuerdo con otras investigaciones (Vallecillos y Batanero, 1997; Behar, 2001). Tampoco asocian correctamente el ancho del intervalo con el nivel de confianza, conflicto también denunciado en las investigaciones de Fidler y Cumming (2005) y Behar (2001). Finalmente encuentra un conflicto, consistente en que los alumnos visualizan los intervalos de confianza como estadísticos descriptivos; resultado también confirmado por la investigaciones de Cumming y Fidler (2005).

- *Campos de problemas y procedimientos.* Encuentra un conflicto cuando el alumno confunde las condiciones de diferentes campos de problemas, por lo que no elige la distribución de muestreo apropiada. También confunde los parámetros en las distribuciones muestrales, por ejemplo el número de grados de libertad. Observa además que no determina correctamente un valor crítico a partir de la tabla de la distribución.
- *Lenguaje y argumentación.* Finalmente denuncia que los alumnos hacen una interpretación incorrecta del intervalo de confianza a partir de una salida de ordenador, confunden los símbolos de varianza poblacional con desviación típica y observa una gran cantidad de conflictos con la notación. Encuentra también conflictos semióticos en la interpretación de intervalos de confianza a partir de gráficos.

### 2.3.5. INVESTIGACIONES SOBRE COMPRENSIÓN DEL CONTRASTE DE HIPÓTESIS

La investigación sobre este tema es muy abundante, pues los contrastes de hipótesis son herramientas básicas en investigación empírica. Batanero (2000) indica que, para realizar esas pruebas de hipótesis se requiere que los estudiantes comprendan y sean capaces de relacionar muchos conceptos abstractos. Esto explica que el uso de las pruebas de hipótesis en las ciencias empíricas en general y en particular en la Psicología ha sido muy criticado (Harlow, Mulaik y Steiger, 1997; Batanero, 2000).

En lo que sigue se hace un breve resumen de estas investigaciones, partiendo de los trabajos de Vallecillos (1994), Chow (1996), Batanero (2000), Díaz y de la Fuente (2004), Díaz (2007), Castro-Soto et al. (2007) y Díaz, Batanero y Wilhelmi (2008), debidamente ampliados con investigaciones no recogidas en los mismos.

#### 2.3.5.1. CONTROVERSIA SOBRE EL CONTRASTE DE HIPÓTESIS

Batanero (2000) indica que hay dos concepciones sobre los contrastes estadísticos: (a) las pruebas de significación, que fueron introducidas por Fisher y (b) los contrastes como reglas de decisión entre dos hipótesis, que fue la concepción de Neyman y Pearson. También sugiere que muchos errores actuales en el contraste de hipótesis se deben a que se mezclan y confunden estas dos metodologías.

**Los tests de significación de Fisher**

Según Batanero, Fisher concibió las pruebas de significación para confrontar una hipótesis nula con las observaciones; para él un valor *p* indicaba la fuerza de la evidencia contra la hipótesis nula. Esta hipótesis $H_{o,}$ se refiere al valor supuesto de un parámetro y para ponerla a prueba se organiza un experimento aleatorio asociado a $H_{o.}$ El experimento termina con datos de una muestra y un estadístico relacionado con el parámetro (Diaz, Batanero y Wilhelmi, 2008).

De acuerdo a de la Fuente y Díaz (2003) el objetivo es poner a prueba la hipótesis nula; no se identifica por una hipótesis alternativa concreta y sólo se considera el error tipo I. Además no hay un criterio estándar para elegir el valor *p,* sino que este valor lo fija el investigador.

**El contraste de hipótesis de Neyman y Pearson**

Neyman y Pearson proponen el contraste de hipótesis como una regla que se puede aplicar para tomar una decisión entre dos hipótesis complementarias $H_0$ o nula y $H_1$ o alternativa (Batanero, 2000). Al decidir entre una de las hipótesis a partir de los resultados del contraste se consideran dos tipos de error:

- *Error tipo I:* Rechazar una hipótesis nula que de hecho sea verdadera. La probabilidad de cometer este tipo de error es un valor $\alpha$ fijo o *nivel de significación*.
- *Error tipo II:* aceptar la hipótesis nula que de hecho es falsa. Beta ($\beta$ ) es la probabilidad de cometer este tipo de error y el complemento de beta $\pi = (1 - \beta)$ sería la *potencia* del contraste. Mientras que $\alpha$ es un número preestablecido, tanto $\beta$ como $\pi$ son variables.

De la Fuente y Díaz (2003) indican que este contraste proporciona un criterio para decidir entre una de las dos hipótesis; añade que los errores tipo II y las probabilidades de error tienen una interpretación frecuencial. Se trata de hallar un procedimiento que a la larga controle el porcentaje de los dos tipos de error cuando se repite muchas veces en la misma población. La relación entre las probabilidades de error hace que, al aumentar una disminuya la otra, si se mantiene constante el tamaño de la muestra.

**Críticas al mal uso del contraste por parte de los investigadores**

Según Batanero (2000), muchos investigadores mezclan las dos metodologías anteriores, práctica que ha ocasionado las siguientes críticas, descritas por Díaz y de la Fuente (2004) y que son otra prueba de la dificultad de comprensión del tema, pues son infundadas:

1. Se indica que el método es inapropiado porque todo el razonamiento del contraste de hipótesis parte del supuesto de que la hipótesis nula es cierta, aunque generalmente es falsa. A esta crítica Díaz y de la Fuente indican que la lógica del contraste de hipótesis no depende de que la hipótesis nula sea falsa o verdadera, sino que se deduce matemáticamente de la distribución muestral del estadístico, que sigue siendo válida incluso cuando la hipótesis nula es falsa.
2. Se critica que el nivel de significación se establece en forma subjetiva; las autoras indican que esto no invalida el contrate y recomiendan para ser más objetivos seguir el método de Fisher, usando el *p*-valor, en lugar del nivel de significación.
3. Se indica que la significación estadística no informa de la probabilidad de que la hipótesis sea cierta. De la Fuente y Díaz revelan que esta probabilidad solo se puede obtener en inferencia bayesiana.

En un intento de mejora de la práctica de la investigación, la American Psychological Association y otras asociaciones científicas recomiendan a los investigadores que no se limiten a presentar los contrastes de hipótesis, incluso cuando sean estadísticamente significativos, sino se tomen en cuenta los intervalos de confianza y la estimación de los efectos.

En lo que sigue se analizan errores relacionados con conceptos específicos que intervienen en el contraste de hipótesis y que se han detectado en investigaciones con estudiantes o analizando la práctica estadística de los investigadores.

### 2.3.5.2. ERRORES SOBRE EL NIVEL DE SIGNIFICACIÓN Y EL VALOR P

Los conceptos peor entendidos asociados con los contrastes de hipótesis son el nivel de significación, α (probabilidad de rechazar la hipótesis nula siendo cierta), y el valor *p*.

#### Intercambio del condicional en la definición

La interpretación incorrecta más extendida del nivel de significación es cambiar los términos de la probabilidad condicional en su definición, interpretándolo como probabilidad de que la hipótesis nula sea cierta, habiendo tomado la decisión de rechazarla (Batanero, Díaz y Wilhelmi, 2008). Este error fue descrito por Birnbaum (1982), quien encontró estudiantes que daban la siguiente definición del concepto:

> Un nivel de significación del 5% "indica que, en promedio, 5 de cada 100 veces que rechacemos la hipótesis nula estaremos equivocados" aunque la definición correcta sería "un nivel de significación del 5% indica que, en promedio, 5 de cada 100 veces que la hipótesis nula sea cierta la rechazaremos" (p. 25)-.

El mismo intercambio de condicional se hace en la interpretación del valor *p* (Probabilidad de obtener un valor igual o más extremo al dado, si la hipótesis es cierta) que se interpreta como probabilidad de que la hipótesis sea cierta si se obtuvo el valor dado del estadístico.

El mismo error lo denuncia Falk (1986). Vallecillos (1994) con una muestra de 436 estudiantes universitarios de diferentes especialidades, lo encontró incluso en aquellos que eran capaces de discriminar entre una probabilidad condicional y su inversa (Vallecillos y Batanero, 1997). También se han descrito en profesores de metodología e investigadores en Psicología (Lecoutre, 1999; Lecoutre, Lecoutre y Poitevineau, 2001; Haller y Kraus, 2002). Vallecillos (1994) encuentra alumnos que realizan otras explicaciones erróneas sobre el significado de $\alpha$, llegando a suprimir la condición en la probabilidad ó confundiéndola con su probabilidad complementaria. Estos errores los encontramos también en las investigaciones de Cañadas *et al* (2012) donde los estudiantes de su muestra confunden nivel de significación con el concepto de *p* valor.

#### Interpretación de resultados significativos

La interpretación incorrecta del nivel de significación se une normalmente a la confusión entre significación estadística y significación práctica que no siempre van unidas, como muestra Lecoutre (1999) en un experimento con 20 investigadores en Psicología y 25 profesionales en métodos de investigación.

Una consecuencia de esta interpretación incorrecta del nivel de significación es, en opinión de Díaz, Batanero y Wilhelmi (2008), que los investigadores tienen una confianza excesiva en que una repetición del experimento llevará consigo una replicación de los resultados. Los autores indican que esta creencia está relacionada con lo que describieron como la "creencia en la ley de los pequeños números", consistente en pensar que las características de una población se han de reproducir incluso en una muestra pequeña (Kahneman, Slovic y Tversky, 1982). Esto puede explicar también la falta de costumbre de analizar la potencia del contraste. También en Cañadas *et al* (2012) hemos encontrado que: responden según el contexto del problema, pero rechazan la hipótesis nula

cuando no deberían hacerlo y tienen en cuenta el valor *p* pero no $\alpha$ explicitamente, entre otros.

Vallecillos (1994) encuentra alumnos que confunden el nivel de significación con la probabilidad de ocurrencia de un resultado significativo, así como un resultado significativo lo ven como el resultado que corrobora la hipótesis nula.

**Otros errores**

En su investigación sobre el contraste de hipótesis, Vallecillos (1994) encuentra otros errores en relación a la i*nterpretación de las probabilidades de error y sus relaciones.* Indica que los estudiantes no aprecian la posibilidad de ocurrencia de los errores de tipo I y II. Por otro lado, la autora indica que los alumnos tampoco comprenden que el *nivel de significación y potencia son riesgos del decisor*: se confunde el nivel de significación con la probabilidad de obtener un resultado correcto, lo que indica en general la confusión de resultado significativo con resultado correcto. También suelen pensar que ambos riesgos están predeterminados por el tamaño de la muestra.

Además no se es consciente de que cuando se realizan contrastes consecutivos en el mismo conjunto de datos, el valor del nivel de significación global sube; por lo cual pueden aparecer resultados significativos simplemente por haber realizado demasiadas pruebas sobre el mismo conjunto de datos (problema de las comparaciones múltiples, descrito por Moses (1992)). Lo que sería adecuado en estos casos es usar métodos estadísticos que permiten dar un nivel de significación global. Por ejemplo, el análisis de varianza sustituye con ventaja el realizar comparaciones entre todos los pares de medias de diferentes grupos.

### 2.3.5.3. ERRORES SOBRE EL CONCEPTO DE HIPÓTESIS

Vallecillos (1994; 1999) encontró un 13% aproximadamente de alumnos que confunden la hipótesis nula con la alternativa al plantear las hipótesis. Vallecillos (1999) asegura que el definir la hipótesis nula y alternativa presenta graves problemas a estudiantes, quienes no son capaces de identificarlas adecuadamente. Este error, según la autora, está indicando la falta de comprensión de la diferencia entre la demostración matemática de una hipótesis y el contraste estadístico de hipótesis.

En la investigación de Vallecillos se planteó también un problema abierto a los estudiantes de su muestra, encontrando que sólo el 26% de los participantes plantea correctamente las hipótesis y un 18,7% confunde el estadístico y el parámetro al plantearlas. La entrevista posterior de Vallecillos y Batanero (1997) a 7 estudiantes considerados brillantes por sus profesores indica que todos ellos conciben el concepto de hipótesis como una afirmación que se refiere a una población y está sujeta a confirmación. Comprenden que las hipótesis nula y alternativa son complementarias, pero tienen diferente papel en el contraste; sin embargo no parece comprenderse que el contraste es un mecanismo de decisión, sino más bien lo conciben como una "demostración débil" o probabilística de la hipótesis. Finalmente, aunque todos admiten en la entrevista que la hipótesis se debe referir a un parámetro, algunos de ellos plantean la hipótesis en función del estadístico muestral. En dicho trabajo se detectan entre otros, dos errores conceptuales a saber:

- Creencia que la hipótesis nula y la región de aceptación son la misma cosa, y
- Creencia que la hipótesis se puede referir tanto a la población como a la muestra.

En Cañadas *et al* (2012), donde se estudian las respuestas a un problema abierto sobre el contraste Chi cuadrado en una muestra de estudiantes de Psicología, encuentran que casi en un 46% de estudiantes no logran formular correctamente las hipótesis. Manifiestan errores tales como confundir: la hipótesis nula y la alternativa, parámetro y estadístico, el contraste de homogeneidad con el de independencia, valor y variable ó se formulan hipótesis que no cubren todo el espacio paramétrico.

Chow (1996) indica que también se confunden la hipótesis estadística alternativa y la hipótesis de investigación El autor explica que hay distintos niveles de hipótesis dentro de una investigación:

- *Hipótesis substantiva*: es la explicación teórica que nos planteamos acerca del fenómeno de estudio. Suele hacer referencia a un constructo teórico inobservable (por ejemplo la inteligencia, actitud, etc. depende de una cierta variable), por lo que su estudio directo no es posible.
- *Hipótesis de investigación*: es una deducción observable de la hipótesis substantiva (por ejemplo, puesto que el rendimiento de los estudiantes se puede deducir de su inteligencia, podemos estudiar si el rendimiento de los estudiantes depende de la variable en cuestión). Trataremos de encontrar apoyo para la hipótesis sustantiva a través de la hipótesis de investigación.
- *Hipótesis estadística alternativa*: De la hipótesis de investigación pasamos a la hipótesis estadística Hace referencia a una población de sujetos, descrita mediante un modelo matemático que se especifica por uno o varios parámetros (en el ejemplo, suponemos que la distribución de puntuaciones en una prueba de rendimiento en matemáticas de dos grupos de alumnos sigue una distribución normal; y además por ejemplo que la media de uno de los grupos será mayor que la del otro).
- *Hipótesis estadística nula*: el complemento lógico de la hipótesis alternativa. Si la hipótesis alternativa era que la puntuación media en el cuestionario era diferente en dos grupos de sujetos, la hipótesis nula especificaría que no hay diferencias de medias entre los grupos control y experimental.

La teoría estadística se ocupa del último nivel (hipótesis estadísticas nula y alternativa), pero alumnos e investigadores confunden todas estas hipótesis y cuando encuentran un resultado significativo, piensan que se refiere a la hipótesis de investigación o incluso a la substantiva (Chow, 1996). Es decir, si se rechaza que la media de los dos grupos de estudiantes en la prueba de rendimiento en matemáticas es igual, se deduce (injustificadamente) que el rendimiento (en general) o bien la inteligencia de uno de los grupos es mayor que la del otro.

### 2.3.5.4. OTROS ERRORES EN EL CONTRASTE

Finalmente, otros errores encontrados en la literatura son los siguientes:

1. *Parámetro y estadístico:* Vallecillos (1994) indica que los estudiantes confunden estos dos conceptos, por lo que a veces plantean sus hipótesis utilizando el estadístico muestral (aproximadamente un 20% de estudiantes en su trabajo). Los estudiantes no dan cuenta de la variabilidad de la distribución muestral del

estadístico, así como tampoco se reconoce la dependencia de dicha distribución muestral respecto al parámetro poblacional. Del mismo modo Cañadas *et al.* (2012), encuentran que: el alumno calcula correctamente el Chi-cuadrado pero confunde los grados de libertad y obtiene por tanto un valor *p* incorrecto, así obtienen probabilidades con valor mayor que la unidad ó no reconoce el valor de la probabilidad asociada al valor obtenido.

2. *Criterio de decisión.* Respecto a la relación entre el nivel *de significación y el criterio de decisión*, Vallecillos observa una confusión entre región de rechazo y de aceptación, así como de la forma cómo hay que construirlas al tener un test unilateral ó bilateral. En relación al nivel de significación y la distribución del estadístico, hay una falta de apreciación de que la hipótesis alternativa determina, junto con el nivel de significación, la región crítica, y que un mayor nivel de significación da una menor área determinada por la función de densidad de la distribución del estadístico en el muestreo y la región crítica bajo la hipótesis nula cierta.
3. *Control experimental y condiciones de aplicación de los métodos:* Una dificultad importante en las ciencias humanas es que no es posible controlar todas las variables relevantes: por ello la estadística sugiere la aleatorización, que garantiza la existencia de técnicas para medir la probabilidad de que un efecto haya sido producido aleatoriamente. En la práctica esta situación no se da y no es posible controlar simultáneamente todas las variables de interés. En otras ocasiones se violan otros supuestos, por ejemplo, la frecuencia es demasiado pequeña para aplicar un test de Chi-cuadrado (Díaz, Batanero y Wilhelmi, 2008).

### 2.3.6. INVESTIGACIONES RELACIONADAS CON EL ANÁLISIS DE VARIANZA Y EL DISEÑO EXPERIMENTAL

El último punto del estado de la cuestión es el específico de nuestro trabajo, donde hemos encontrado pocas investigaciones.

#### 2.3.6.1.TIPOS DE VARIABLES

Rubin y Rosebery (1990) planificaron y observaron un experimento de enseñanza dirigido a estudiar las dificultades interpretativas en algunas ideas básicas del diseño experimental. Usaron una actividad de lanzamiento a una canasta de baloncesto. Se varió la distancia de lanzamiento (de 1 a 9 metros) y el ángulo posicional del lanzador (0º, 45º y 90º). El objetivo de la lección era explorar los efectos separados de la distancia y el ángulo. El estudio mostró la confusión entre las ideas de variables independientes, dependientes y extrañas. Algunos estudiantes sugirieron como posibles variables independientes características individuales del lanzador, como su altura o su habilidad para encestar, que no eran manipulables. Incluso la altura de la canasta, que se conservó inalterable durante el experimento fue considerada como variable independiente por algunos estudiantes. Los autores concluyen la dificultad en distinguir entre las características de los sujetos que no tenían influencia sobre el resultado del experimento de otras variables que si podrían tenerla. El papel de la asignación aleatoria como medio de compensar estas diferencias individuales tampoco fue comprendido.

### 2.3.6.2.INTERACCIÓN

Otro concepto importante en el análisis de varianza es el de interacción, que aparece cuando la diferencia entre los niveles de un factor, no es la misma en los niveles del otro. La presencia de la interacción es lo que se destaca en los modelos de análisis de varianza de varios factores, pero, no hay acuerdo generalizado sobre la estrategia que debe seguirse para interpretar correctamente una interacción significativa.

Rosnow y Rosenthal (1991) indicaron que la interacción es el resultado peor interpretado en Psicología; también presentan explicaciones acerca de su cálculo y la interpretación. Zukerman, Hodgins, Zuckerman y Rosenthal (1993) realizan un estudio empírico de artículos publicados en revistas de prestigio de Psicología, encontrando que la tercera parte de 551 investigaciones analizadas fallan en identificar el método correcto para interpretar la interacción,

Pardo *et al*, (2007) indican que la interpretación de la interacción conduce en muchos casos a conclusiones incorrectas. Los autores revisan 699 artículos publicados en *Psicothema* (entre 2000 y 2004) y 149 publicados en n *International Journal of Clinical and Health Psicology* (entre 2001 y 2005), estudiando los que habían usado un análisis de varianza de dos o más factores. Así se obtuvieron 166 artículos (19,6% del total), 16 de los cuáles fueron descartados, por no tener efecto de interacción significativo, o por haber abordado el análisis estadístico con pruebas no paramétricas. Sólo en 13 se analiza e interpreta correctamente la interacción.

Un error frecuente (79,1% de los artículos revisados) es analizar e interpretar la interacción de cada efecto simple por separado, aunque para evaluar el efecto del tratamiento, se ha de comparar la diferencia observada después de analizar el efecto de la interacción. Los autores sugieren que este error puede ser debido al uso frecuente del paquete estadístico SPSS para el análisis, ya que este no permite realizar un análisis post hoc de la interacción en forma directa, ni efectuar los contrastes que aíslan el efecto de la interacción en los diseños factoriales.

Green (2007) enfatiza que los estudiantes tienen dificultades para comprender el concepto de interacción. La autora propone ejercicios para la interpretación de la interacción, mediante una superposición de tareas. Los estudiantes primero escucharon las explicaciones y ejemplos sobre que es la interacción y respondieron preguntas de elección múltiple orientadas a la comprensión y aplicación, Además, trabajaron en pequeños grupos para calcular los términos de la interacción a mano y usando software estadístico. Seguidamente escribieron resúmenes técnicos de los resultados. Los estudiantes recibieron diferentes conjuntos de datos y se pidió repetir el análisis, y una breve escritura de resultados como ejercicio para la casa. La comprensión final de la interpretación de la interacción fue evaluada con tareas escritas, que incluían tablas, gráficos y diagramas, encontrando que esta secuencia resulto efectiva.

La autora recomienda para la enseñanza de interacción usar lenguajes múltiples (vocabulario adecuado, tablas, gráficos), pedir a los estudiantes que generen una tabla de valores mostrando los efectos principales y los efectos de interacción y destacar que frente a la presencia de una interacción significativa, las conclusiones sobre los efectos principales deben ser revisadas. Finalmente hay que diferenciar los efectos simples de los efectos principales y las fuentes de variabilidad de la interacción.

### 2.3.6.3. ANÁLISIS DE PRUEBAS DE VALORACIÓN DE LA COMPRENSIÓN DEL DISEÑO EXPERIMENTAL

Trigo, López, Martínez y Moreno (2005) analizan las propiedades psicométricas de un conjunto de ítems de opción múltiple utilizadas en evaluaciones oficiales de una asignatura relacionada con el diseño de experimentos en la Licenciatura en Psicología de la Universidad de Sevilla. Su finalidad fue construir un banco de ítems depurados que se pueda muestrear para futuras pruebas. Para ello analizan 465 ítems utilizados en exámenes de diferentes convocatorias de la asignatura. Cada ítem tenía 6 posibles opciones, entre las cuales dos de ellas eran: (a) ninguna de las anteriores es correcta y (b) todas las anteriores son correctas. Los ítems se clasificaron de acuerdo con siete categorías de contenidos correspondientes a los distintos temas del programa docente:

1. Hipótesis (identificación del tipo de hipótesis y de los contrastes adecuados);
2. Diseño (tipo de diseño utilizado, técnicas de control, fuentes de variabilidad del diseño y supuestos relevantes);
3. Cálculo de estimaciones de efecto y estimadores de parámetros;
4. Cálculos necesarios para la comprobación de supuestos y el cuadro de análisis de varianza;
5. Conclusiones estadísticas (aceptación o rechazo de hipótesis nulas);
6. Interpretación de resultados (significado de las conclusiones estadísticas en términos de la comprobación de supuestos o de las hipótesis de investigación);
7. Mixtas (ítems con opciones referidas a más de una de las categorías anteriores y que por tanto fueron excluidos del análisis).

Los autores analizan la validez y fiabilidad, que osciló entre 0.70 y 0.86. Analizaron los índices de dificultad y discriminación, encontraron un valor medio de 0.54 para el índice de dificultad, y distribución aproximadamente Normal. El índice de discriminación promedio fue de 0.4. Se analizaron también diferencias en las propiedades psicométricas de las pruebas en su conjunto en función del número y tipos de alternativas de las mismas. Mientras las pruebas de seis opciones, mostraron un mayor grado de consistencia interna, los ítems de tres alternativas resultaron más discriminativos. Por último, se analizaron diferencias entre los indicies de dificultad en función del tipo de contenido al que hacían referencia las preguntas. Las preguntas sobre hipótesis fueron más fáciles que las de diseño; y las relacionadas con la interpretación más difíciles que el resto.

### 2.3.6.4. DIFICULTADES CAUSADAS POR EL SOFTWARE

Chiocca (2006) analiza la hoja de cálculo como una herramienta para enseñar estadística, e indica que aparecen algunos obstáculos debido al uso de este instrumento, específicamente en el aprendizaje del análisis de varianza. Entre la construcción del conocimiento estadístico y el uso de la hoja de cálculo puede aparecer un obstáculo (en el sentido de Brousseau (1997), incluso si los estudiantes tienen experiencia en la aplicación de técnicas estadísticas y en el uso de la hoja de cálculo.

En su experiencia se explica el algoritmo del análisis de varianza, a parir de una tabla que contiene 36 resultados distribuidos en las 4 variedades, pero no es balanceado. Para resolver el problema será necesario obtener promedios, tanto por variedad como el promedio general. La autora indica que el conocimiento de la media ordinaria se convierte en obstáculo en el cálculo de la media ponderada, pues un alto porcentaje de alumnos (el 68%) calculan la media de las medias, sin tener en cuenta que el tamaño de

las muestras, y las limitaciones a las ordenes en la hoja de cálculo. En este sentido sugiere la existencia de un *obstáculo instrumental* en el siguiente sentido: en la situación planteada el uso de ordenadores tiene un doble objetivo, el aprendizaje de estadística y el uso de la hoja de cálculo; al usar la herramienta utiliza un esquema particular para obtener el promedio y este esquema genera la emergencia de un obstáculo para algunos estudiantes.

Para remediar estos problemas encontramos propuestas, como la de Castle (2002), quien propone agregar la posibilidad de una enseñanza por aprendizaje visual, para desarrollar la comprensión de conceptos estadísticos. Propone el software Excel, pues tiene la facilidad de visualizar el impacto en un gráfico al variar el valor de un dato, viendo el efecto que tiene sobre estimaciones estadísticas. Destaca la importancia de mostrar el uso de datos reales a problemas reales, especialmente con estudiantes provenientes de otras disciplinas, para que puedan ver la relevancia de las técnicas estadísticas sobre sus propios temas. Enfatiza el fuerte efecto que tiene en los estudiantes el aprendizaje interactivo que supone el cambio de valores en los datos, tanto para el intervalo de confianza, como para la tabla de análisis de varianza.

### 2.3.7. CONCLUSIONES DEL ESTADO DE LA CUESTIÓN

El estudio de los antecedentes muestra una gran variedad de dificultades en relación a objetos estadísticos tales como muestreo, distribución muestral, intervalo de confianza y contraste de hipótesis y también alguna investigación aislada sobre otros objetos relacionados con el diseño experimental.

En el cuestionario que hemos construido se incluyen algunos ítems relacionados con la comprensión del intervalo de confianza y contraste de hipótesis pues son objetos básicos en el tema. Haremos énfasis en aspectos como la potencia en las pruebas de hipótesis que no han sido tan tratados en la investigación previa.

Por otro lado no hemos encontrado investigaciones didácticas relacionadas con la comprensión del análisis de varianza, sus diversos modelos, elección del modelo, supuestos y comprobación de los mismos o la interpretación y comprensión de cómo se obtiene una tabla de análisis de varianza. Por ello todos estos puntos serán incluidos en el cuestionario, dándoles un peso importante.

Tendremos también en cuenta las sugerencias de Trigo *et al.* (2005) respecto a que los ítems de evaluación deben tener un número de opciones entre tres y cuatro. Resaltamos el hecho de que las categorías de contenido consideradas por los autores están todas incluidas en mayor o menor medida en el análisis conceptual llevado a cabo en el Capítulo 1, que se utilizará para definir la variable objeto de medición. Dicho análisis, junto con los resultados del estudio del estado de la cuestión, resumido en este capítulo se tendrá en cuenta, tanto en la definición de la variable, como en la elaboración de los ítems.

Finalmente, los ítems de opción múltiple se complementarán con problemas abiertos, que han sido mucho menos utilizados en la investigación empírica y nunca en relación al análisis de varianza. Por ello, nuestro análisis semiótico de las respuestas de los estudiantes a los diversos apartados aportará la identificación de conflictos semióticos específicos en la comprensión del análisis de varianza y matizará otros relacionados con la comprensión del contraste de hipótesis.

# CAPÍTULO 3.
# ESTUDIO 1. CONSTRUCCIÓN DE UN CUESTIONARIO DE EVALUACIÓN


3.1. Introducción
3.2. Objetivos del Estudio 1
3.3. Definición de la variable objeto de medición
3.3.1. Fundamentos de la definición
3.3.2. Tabla de especificaciones del instrumento
3.4. Construcción de un banco de ítems
3.4.1. Tipo y número de ítems
3.4.2. Proceso de construcción del banco de ítems
3.5. Análisis a priori de los ítems de evaluación
3.5.1. Parte I. Prerrequisitos
3.5.2. Parte II. Análisis de varianza
3.5.3. Resumen de contenido de los ítems
3.6. Pruebas piloto de ítems
3.6.1. Muestra y material
3.6.2. Método
3.6.3. Resultados y discusión
3.7. Evaluación de ítems mediante juicio de expertos
3.7.1. Expertos participantes
3.7.2. Material y método
3.7.3. Resultados y discusión
3.8. Selección final y revisión de ítems para el cuestionario
3.9. Revisión del cuestionario mediante juicio a expertos
3.10. Conclusiones sobre el Estudio 1


## 3.1.INTRODUCCIÓN

En el Capítulo 1 se expuso como segundo objetivo de la investigación el siguiente: "C*onstruir un instrumento para evaluar la comprensión de los elementos básicos del análisis de la varianza, siguiendo unos criterios metodológicos rigurosos en la selección de los ítems que lo compongan".*

Para cumplir el citado objetivo en este capítulo describiremos el Estudio 1, orientado a la construcción del cuestionario que será utilizado para evaluar los conocimientos de los estudiantes de Psicología sobre el análisis de varianza. Algunos resultados e han publicado en Vera y Batanero (2008) y Vera, Olivo y Díaz (2009).

Comenzamos con la descripción de los objetivos de este estudio y a continuación se detalla la metodología utilizada en la construcción de nuestro instrumento, seguiremos la metodología ya probada en Díaz (2007) y Olivo (2008), que consta de los siguientes pasos:

1 Definición de la variable objeto de medición (que parte del análisis del objeto matemático realizado en el Capítulo 1 y que se describe brevemente en la Sección 3.3).

2 Construcción de un banco de ítems partiendo del estudio de los antecedentes y de ítems disponibles en pruebas de evaluación realizadas anteriormente a estudiantes similares a los que participan en el estudio (Sección 3.4), así como de la redacción de ítems nuevos. De entre todos los ítems recopilados se seleccionan los que a priori podrían contribuir a nuestros objetivos.

3 Análisis a priori de los ítems (Sección 3.5). Este análisis nos ayuda a comprobar que el significado que se pretende evaluar queda cubierto por los ítems.

4 Pruebas piloto de ítems (Sección 3.6). Se administran grupos de ítems a estudiantes similares a los que formarán parte del estudio definitivo, con el fin de comprobar su legibilidad, tiempo necesario para completarlos, dificultad y proporción de respuesta a cada distractor.

5 Valoración de ítems mediante el juicio de expertos (Sección 3.7). Se valora tanto la adecuación del ítem para un contenido específico, como las características del ítem.

6 Selección, en base a estas dos pruebas, de los ítems que conformarán el cuestionario piloto, y revisión de los mismos, teniendo en cuenta las indicaciones de los expertos para, mejorar su legibilidad.

De esta forma, se finaliza el Estudio 1 con una versión del cuestionario que se utilizará posteriormente en el Capítulo 4 para realizar un estudio de evaluación con estudiantes de Psicología (Estudio 2). En dicho Estudio 2 también se analizará la validez y fiabilidad del instrumento implementado. A continuación describimos el Estudio 1.

## 3.2.OBJETIVOS DEL ESTUDIO 1

El *objetivo principal* del Estudio 1 es construir un cuestionario que incluya ítems que globalmente permitan proporcionar indicadores de la comprensión del objeto estadístico "análisis de varianza" a nivel elemental. Más concretamente, se probará el cuestionario con estudiantes de Psicología después de la instrucción (ya descripta en el Capítulo 1), aunque nuestra intención es que su aplicación pueda ser más global para otro tipo de estudiantes con iguales características. Entenderemos la comprensión, de acuerdo a nuestro marco teórico, es decir, como correspondencia entre los significados personales e institucionales y dependiente de la institución de enseñanza (Godino, 1996). Las respuestas de los estudiantes a los ítems serían los indicadores empíricos que nos permiten realizar esta evaluación

Un *segundo objetivo* es que el instrumento permita identificar las principales dificultades de comprensión de los estudiantes en este tema, para utilizar esta información en la mejora de materiales curriculares, como los libros de texto y en la planificación y evaluación de la enseñanza. Cumpliendo este segundo objetivo contribuiríamos con la investigación sobre un tema que ha recibido muy poca atención hasta el momento.

El cuestionario a elaborar en este estudio será nuestro instrumento de medida y se aplicará posteriormente a grupos de sujetos, similares al grupo de ensayo o normativo (Losada y López, 2003). Pretendemos que el instrumento desarrollado pueda ser aplicado a la población general de estudiantes (no sólo en la Universidad donde se toma la muestra piloto), por la cual aspiramos a conseguir suficiente validez (Kline, 2013). Según León y Montero (2002) la comprensión (en este caso es la de elementos básicos del análisis de varianza), es un constructo que no se puede observar directamente; por lo

que al tratar de evaluarla, sus características serán inferidas a partir de las respuestas dadas por estudiantes.

En consecuencia, se clasifica el método de recogida de datos como *medida*, pues queremos estimar las capacidades y los conocimientos de los individuos, en forma indirecta, no por simple encuesta u observación, pues los individuos no tienen conciencia plena de estas capacidades (Fox, 1981; Sánchez, 2012). Se desea asimismo que el cuestionario cumpla los requisitos de idoneidad descriptos por Batanero y Díaz (2005).

El proceso de construcción del instrumento seguirá las normas psicométricas recomendadas (APA, AERA y NCME, 1999; Kline, 2013). Se tendrá en cuenta las características de los alumnos, y la naturaleza de las inferencias que se extraerán de las puntuaciones obtenidas a partir de su aplicación a una muestra de sujetos. Esto nos pondrá en condiciones de especificar la variable dentro de un marco global, y también el desarrollo, prueba y revisión de los ítems (Sax, 1989).

## 3.3. DEFINICIÓN DE LA VARIABLE OBJETO DE MEDICIÓN

### 3.3.1. FUNDAMENTOS DE LA DEFINICIÓN

Como cualquier constructo psicológico, la comprensión de los conceptos básicos del análisis de varianza, caracteriza los comportamientos de los individuos y permite explicar patrones de comportamiento ante ciertas tareas (León y Montero. 2002). El primer paso en la construcción de un instrumento es definir lo que deseamos medir, pues, como se ha indicado, un constructo no es directamente observable pero sí se infiere a través de la conducta mediante indicadores a partir de los cuales es posible evaluar su presencia-ausencia o grado en se encuentra (Kline, 2013). Dichos indicadores se deducirán de las respuestas a un instrumento de medición, y pueden traducirse a valores numéricos (por ejemplo, la puntuación en un cuestionario) o categorías cualitativas (como tipo de respuesta).

Según Martínez Arias (1995) el diseño de instrumentos para la medida de constructos educativos y psicológicos presenta algunos problemas, según se indican a continuación:

- La aproximación a la medida de un constructo no suele ser única.
- Las situaciones, el muestreo de tareas y las ocasiones fortuitas que están presentes al momento de medir, siempre traen aparejado un margen de error.
- Las unidades de medidas y el origen de los significados no están, en general bien definidos.
- No hay una única escala y su origen y unidades de medida pueden no ser bien definidas.
- Dificultad de identificar un constructo, así como sus definiciones operativas, con total independencia del investigador.

Esta autora describe la definición del constructo a dos niveles, a saber:

- Nivel semántico, según el cual el constructo queda definido mediante la especificación detallada de los comportamientos observables o reglas de correspondencia entre el constructo y la conducta, y

- Nivel sintáctico ó relacional, a través del cual ponemos en relación el constructo con otros constructos que suponemos relacionados con él dentro de una misma teoría.

Nosotros abordamos la definición semántica de la variable objeto de medición, a partir de la delimitación de contenidos propuesta en el Capítulo 1, que se resume brevemente a continuación.

### 3.3.2. TABLA DE ESPECIFICACIONES DEL INSTRUMENTO

La elaboración de la tabla de especificaciones que abarca la selección de los contenidos a evaluar, partió por un lado, de la revisión del contexto educativo de los estudiantes y de los contenidos desarrollados sobre el tema en las asignaturas Análisis de datos I y II realizadas por los participantes (Sección 1.3). Adicionalmente, se efectuó un análisis epistémico de los objetos matemáticos que constituyen el análisis de varianza elemental en la Sección 1.7, a partir de diversos libros de texto, la mayoría de ellos orientados a la Psicología o las Ciencias Sociales (Manzano, 1997; Balluerka y Vergara, 2002; Andrés y Luna, 2004; Montgomery, 2005; Peers, 2006; Vallejo y Lozano, 2006; Ferré y Rius, 2008; Pagano, 2011).

Tabla 3.1. Tabla de especificaciones del contenido del cuestionario

| PRERREQUISITOS |
|---|
| 1. Construcción de un intervalo de confianza |
| 2. Interpretación del significado de un intervalo de confianza. |
| 3. Contraste bilateral y unilateral. Asignación de hipótesis |
| 4. Hipótesis nula y alternativa. Asignación de hipótesis |
| 5. Errores tipo I y II, nivel de significación y potencia |
| 6. Cálculo de valor crítico, dado el nivel de significación |
| 7. Regla de decisión en un contraste |
| ANÁLISIS DE VARIANZA |
| 8. Selección de un modelo de Análisis de Varianza, de acuerdo con las características de un problema |
| 9. Supuestos de aplicación de un modelo de Análisis de Varianza. |
| 10. Comprobación de supuestos en un modelo de Análisis de Varianza |
| 11. Modelo lineal asociado a un tipo de Análisis de Varianza |
| 12. Cálculo en el Análisis de varianza, Medidas repetidas |
| 13. Cálculo en el Análisis de varianza. Modelo de dos factores |
| 14. Interpretación de una tabla de Análisis de Varianza. Modelo de dos factores |
| 15. Interpretación de salida de software. Identificación del modelo de Análisis de Varianza |
| 16. Interpretación de salida de software. Resultados |
| 17. Interpretación de salida de software. Supuestos del modelo de Análisis de Varianza |

Con estos dos tipos de análisis fijamos el significado de referencia del análisis de varianza elemental en nuestro estudio (más concretamente, significado pretendido). Se construyó una primera tabla de unidades de contenido, clasificando dichos objetos en una serie de unidades principales de contenido semántico, que abarcaba tanto prerrequisitos del tema como objetos propios del análisis de varianza. Posteriormente, y observando que la lista lograda era excesivamente amplia para nuestros objetivos, y que algunos contenidos estaban incluidos en otros, se simplificó, quedando la que reproducimos en la Tabla 3.1 para facilitar la lectura de este capítulo. La lista de especificaciones del contenido de la variable incluye como pre-requisitos, la interpretación y construcción de intervalos de confianza, el contraste de hipótesis, ya que nuestro objeto de estudio requiere la realización de al menos un contraste.

Respecto al análisis de varianza propiamente dicho, incluye los modelos de efectos fijo (uni y bifactorial) y medidas repetidas; su selección, supuestos, comprobación de los mismos, y el cálculo e interpretación de resultados.

Cada uno de los elementos de esta lista es una configuración de objetos matemáticos, que incluyen los diferentes tipos de objetos considerados en nuestro marco teórico (Godino, 2003; 2014):

- *Lenguaje* (términos, expresiones, notaciones, gráficos). En los ítems propuestos del cuestionario se tendrá en cuenta el lenguaje verbal, simbólico y gráfico que el estudiante tendrá que interpretar en los enunciados. Asimismo, en los problemas abiertos que se describen en el Capítulo 5, podremos evaluar el uso que hacen los estudiantes de estos diferentes tipos de lenguaje matemático.
- *Situaciones-problemas*. Se consideran las siguientes situaciones: estimación de parámetros mediante un intervalo de confianza (contenido 1); probar hipótesis sobre uno o dos parámetros (contenidos 3 a 7); comparar medias en muestras independientes (contenidos 8 a 11; 13 a 17); comparar medias en muestras relacionadas (contenidos 8 a 12; y 15 a 17)
- *Conceptos- definición* (introducidos mediante definiciones o descripciones). Se consideran en los diferentes contenidos población y muestra, variable aleatoria y distribución, estadístico muestral, parámetro poblacional, intervalo y coeficiente de confianza, regiones en un contraste, error Tipo I y Tipo II, nivel de significación y potencia; variable dependiente e independiente, factor, efecto, interacción, modelo lineal.
- *Proposiciones* (teoremas y propiedades). Se consideran las propiedades básicas de la variable aleatoria y distribuciones teóricas, de la distribución muestral (dependencia del valor supuesto, relación de la varianza con el tamaño muestral; parámetros de las distribuciones *t*, *F* y Chi-cuadrado), propiedades de la dependencia e independencia, Teorema central del límite, descomposición de la variabilidad total para los diversos modelos lineales considerados.
- *Procedimientos* (algoritmos, operaciones, técnicas de cálculo). En nuestro cuestionario se tendrá en cuenta la lectura de tablas (mayoría de contenidos), construcción e interpretación de intervalos de confianza (contenidos 1 y 2), búsqueda de valores críticos (contenido 6), elección de un modelo lineal (contenido 11), completar tablas de Análisis de Varianza (contenido 12), toma de decisión en un contraste (contenido 7), interpretación de salidas de ordenador (contenidos 15 a 17), comprobación de supuestos (contenido 10).
- *Argumentos* (enunciados usados para validar o explicar las proposiciones y procedimientos, deductivos o de otro tipo). Se tendrá en cuenta en los problemas abiertos descritos en el Capítulo 5, ya que para los ítems de opción múltiple no se requiere argumentación. Esperamos sobre todo argumentos de tipo deductivo, bien formales (con empleo de notación matemática) o informales (lenguaje verbal).

En lo que sigue se describen el resto de los pasos en la construcción del cuestionario, donde tratamos de recoger las unidades de contenido fijadas.

## 3.4.CONTRUCCIÓN DE UN BANCO DE ÍTEMS

Una vez definida la variable objeto de medición, se procedió a recopilar un banco de ítems de entre los cuáles se pudieran elegir los que formarían el cuestionario. De acuerdo a León y Montero (2002), dicho "banco de ítems" consiste en una colección de estímulos recopilado para una finalidad determinada, en general, producir una respuesta que permita inferir acerca de algún constructo psicológico. En la selección de los ítems seguimos los criterios de calidad habituales que sugieren Wang y Osterlind, (2013):

- Congruencia entre cada ítem y el objetivo del cuestionario, de vital importancia para lograr la validez del instrumento. Este criterio se garantiza mediante el análisis a priori de cada ítem (Sección 3.4.3) y el juicio de expertos que han valorado dicha congruencia (Sección 3.6).
- Reducir al máximo la contribución de cada ítem al error de la medida de puntuación en el test, tratando de no usar ítems de excesiva dificultad o facilidad. Lo aseguraremos mediante las pruebas piloto de ítems (Sección 3.5) y luego se comprobará el resultado global mediante el estudio de la fiabilidad del cuestionario final en el Capítulo 4.
- Consistencia entre los fines del cuestionario y el formato de los ítems (en el siguiente apartado se describen los tipos de ítems utilizados).

Además, siguiendo a Kline (2013), los ítems satisfacen criterios de homogeneidad e independencia local (la respuesta a una unidad no está relacionada con la respuesta a otra). Por otro lado, el número total de ítems permite cubrir el contenido pretendido y asegurar una fiabilidad suficiente, limitando la longitud total plausible del instrumento. También se puso atención en la amplitud del contenido, su estructura, forma de aplicación, y característica de los estudiantes.

### 3.4.1. TIPO Y NÚMERO DE ÍTEMS

Para tomar la decisión sobre el formato de ítems a usar, tuvimos en cuenta los componentes específicos del dominio de rendimiento o habilidad que se desean evaluar (Millman y Greene, 1989). Dichos componentes, que incluyen tanto el recuerdo, comprensión de símbolos y conceptos, como su aplicación y la capacidad de argumentación sugieren la necesidad de usar más de un tipo de ítems.

En consecuencia se decidió utilizar dos formatos: en primer lugar, para poder evaluar una gama amplia de conocimientos de conceptos y propiedades en un tiempo limitado, decidimos usar ítems de opción múltiple, donde el alumno ha de seleccionar la respuesta correcta partiendo de un conjunto de alternativas dadas (Viladrich, 2012). El número de respuestas presentadas en estos ítems en todos los casos ha sido tres, para aumentar la fiabilidad (pues se recomienda usar un número similar de opciones y no demasiadas). Sólo una de las opciones incluidas era la correcta, y las otras dos representaban distractores evitando además usar respuestas tales como “ninguna de las anteriores” o “todas las anteriores” (Trigo, López, Martínez y Moreno, 2005). Se tuvo también en cuenta que los estudiantes de la muestra estaban habituados a este tipo de formato, ya que el profesorado lo usa habitualmente en sus exámenes; y en segundo lugar se incluyeron también algunos ítems de respuesta abierta, que permitirán además de acceder a la solución final; al lenguaje, los procedimientos y argumentos empleados, valorar el uso del lenguaje estadístico, y hacer un análisis de conflictos semióticos.

En la decisión del número de ítems a elegir, tuvimos en cuenta los siguientes aspectos (Barbero, 2003):

- Administración del instrumento: Sería administrado por razones de eficiencia de forma grupal. Para que fuese eficaz la redacción del ítem debe ser lo suficientemente clara y las instrucciones para completarlo fácil de seguir, pues en esta administración el investigador no tiene tanta interacción con el encuestado.
- Limitaciones de tiempo en la administración: Hicimos un balance entre la cantidad de ítems requeridos para cubrir el contenido y el tiempo que se disponía para su administración (así como la posible fatiga de los estudiantes). Finalmente consideramos proporcionar un tiempo de una hora y media para, de este modo evitar que algún ítem no sea respondido sólo por falta de tiempo.

### 3.4.2. PROCESO DE CONSTRUCCIÓN DEL BANCO DE ÍTEMS

Para construir los ítems, se comenzó analizando la investigación previa, entre ellas tesis doctorales sobre didáctica de la inferencia estadística (Vallecillos, 1994; Behar, 2001; Tauber, 2001; Alvarado, 2007; Díaz, 2007; Olivo 2008 y Cañadas, 2012), así como los trabajos reseñados en el Capítulo 2. Esta lectura facilitó la elaboración de una lista de dificultades relacionadas con el tema, y nos ayudó a recopilar algunos ítems potencialmente útiles, sobre todo en relación con el intervalo de confianza y las pruebas de hipótesis.

Asimismo se examinaron las pruebas de evaluación utilizadas los años anteriores en la asignatura Análisis de Datos I y II en la Universidad de Huelva, que fueron facilitadas por los profesores de la asignatura. Dichas pruebas contenían ítems que cubrían gran parte del contenido pretendido, puesto que dichos contenidos eran objeto de enseñanza y, como se ha dicho, habitualmente se utilizan ítems de opción múltiple en las pruebas de evaluación.

Elaborado un primer banco de ítems se comenzó un proceso de análisis para determinar los contenidos evaluados por cada uno de los ítems. Se comparó cada ítem con la tabla de especificaciones de la variable, determinando si evaluaba alguno de dichos contenidos, descartándolo en caso contrario. Para la mayoría de ítems seleccionado fue necesario cambiar algunos de los distractores, para poder evaluar algunos de los sesgos descritos en la investigación previa y que pensábamos podrían presentarse en los estudiantes. En otros casos fue necesario cambiar los datos proporcionados para lograr que los cálculos fuesen sencillos y no aumentar innecesariamente la dificultad del ítem.

Todo se hizo llevando un proceso iterativo, completando progresivamente el conjunto de ítems, y revisando el formato (añadiendo o suprimiendo distractores o la redacción, en caso necesario). Finalmente, cuando no se encontraron reactivos por alguno de los métodos anteriores, se elaboraron algunos nuevos, siempre tratando de respetar el formato y número de distractores. Se llegó de este modo a un conjunto de 41 ítems, que cubrían la tabla de especificaciones (al menos dos ítems en cada unidad de contenido).

## 3.5. ANÁLISIS A PRIORI DE LOS ÍTEMS DE EVALUACIÓN

En lo que sigue realizamos un análisis a priori de los ítems del banco de ítems, agrupados por contenido y finalizando con una tabla de especificaciones del contenido

cubierto por el conjunto de ítems. En cada ítem se marca en negrita la respuesta correcta. Analizamos con detalle cada uno de los distractores del primer ítem en cada contenido, indicando cuál de los distractores del resto es equivalente, sin repetir el análisis.

### 3.5.1. PARTE I. PRERREQUISITOS

**Contenido 1: Construcción de un intervalo de confianza**

**Ítem 1a.** Un psicólogo escolar desea estimar la puntuación media de la población en un test de rendimiento de lectura. Para ello administra el test a una muestra de 36 estudiantes, obteniendo una media de 48 y desviación típica 10. Calcular, al nivel de significación $\alpha$=0,05, los límites del intervalo de confianza para la puntuación en el test. (Error típico de la media: $\frac{S_d}{\sqrt{n}}$).

a. [38 , 58]
b. [15,3 , 80,7]
**c. [44,73 , 51,27]**

**Ítem 1b**. Se toma una muestra de 48 estudiantes con la finalidad de conocer su rendimiento en lectura. Un psicólogo escolar administra un test, así obtiene una media de 54,7 y desviación típica 5,23. Calcular, al nivel de significación α=0,01, los límites del intervalo de confianza para la puntuación en el test. (Error típico de la media: $\frac{S_d}{\sqrt{n}}$) son:

a. [53,95 ; 55,45]
b. [41,23 ; 68,17]
**c. [52,75 ; 56,65]**

Se pretende evaluar a través de estos ítems la comprensión del procedimiento de construcción de un intervalo de confianza, particularizado para la media del "rendimiento en lectura" de una población. Por los tamaños de las muestras, podemos asumir normalidad. El procedimiento para la determinación del intervalo pedido es desarrollar la fórmula $\bar{x} \pm z_{\frac{\alpha}{2}}.\frac{S_d}{\sqrt{n}}$, utilizando los datos extraídos del enunciado. Si tomamos el ítem 1a, se tiene que los datos son: $\bar{x} = 48$, $\alpha = 0{,}05$, $S_d = 10$ y $n = 36$, de donde resulta $z_{\frac{\alpha}{2}} = z_{0{,}025} = 1{,}96$ (el estudiante deberá buscar valores en tablas de distribución). Un procedimiento equivalente se sigue para el ítem 1b, aunque aquí $\alpha = 0{,}01$.

- El distractor (c) nos proporciona el resultado correcto en los dos casos. Para el ítem 1a, resulta: $48 \pm 1{,}96 \cdot \frac{10}{\sqrt{36}} = 48 \pm 3{,}27$, y el intervalo de confianza siguiente: $[44{,}73;\ 51{,}27]$.
- En la opción (a) para el ítem1a se comete un error en la fórmula, que consiste en sumar y restar simplemente la desviación típica a la media muestral, no usando la tabla de la distribución. En el 1b tampoco se usa la tabla, pero se le suma y resta el error típico de la media. Para el ítem 1a se obtiene $48 \pm 10$, generando el intervalo $[38;\ 58]$.
- En la opción (b) del ítem 1a se evalúa si el alumno lee correctamente las tablas de la distribución, ya que se considera como valor del percentil el número 19,62 en lugar del 1,962 que es el correcto. Obtendría $48 \pm 19{,}62 \cdot \frac{10}{\sqrt{36}} = 48 \pm 32{,}7$, es decir: $[15{,}3;\ 80{,}7]$. Para el ítem 1b, en este distractor evalúa si el alumno utiliza correctamente el error típico de la media, ya que en este intervalo de confianza propuesto como respuesta, no se ha dividido la desviación típica por el facto $\sqrt{48}$ .

**Contenido 2: Interpretación del significado de un intervalo de confianza.**

**Ítem 2a**. En relación con la pregunta anterior, también podríamos afirmar que:

a. **Si extrajéramos 100 muestras y calculáramos en cada una el intervalo de confianza para la media, en 95 de ellos, se encontraría el verdadero valor del parámetro.**
b. La puntuación del test está en el intervalo estimado con una probabilidad del 95%
c. Si extrajéramos 100 muestras y calculáramos en cada una el intervalo de confianza, en 95 de los intervalos, se encontraría la puntuación de los sujetos de la muestra.

**Ítem 2b**. Un investigador ha calculado 1000 intervalos de confianza al 99% para μ. Suponiendo que los datos en los que se basan los intervalos se seleccionan de manera independiente entre sí. ¿Cuántos de estos 1000 intervalos se espera que contengan el valor verdadero de μ?

a. **Sólo 990 intervalos contendrán el valor de μ**
b. Los 1000 intervalos contendrán el valor de μ
c. No se puede saber a priori la cantidad de intervalos se espera que capturen el verdadero valor del parámetro μ.

Estos ítems, adaptados de Olivo (2008), evalúan la interpretación del intervalo de confianza, siendo la respuesta correcta la opción (a) para ambos ítems. Si hacemos la interpretación frecuentista de la probabilidad, usando un nivel de significación $\alpha = 0{,}05$ (0,01 en el ítem 2b); el valor 0,95 indica el nivel de confianza del intervalo (0,99 para el ítem 2b). Por tanto, sobre 100 muestras distintas extraídas de la misma población, y calculando con ellas 100 intervalos, 95 de ellos incluirían el verdadero valor de la media poblacional $\mu$ (análoga interpretación se generaría para el ítem 2b). Los errores que se evalúan en estos ítems son los siguientes:

- El distractor (b) en el ítem 2a da la interpretación bayesiana del intervalo de confianza, la cual consiste en considerar los extremos como fijos y el parámetro como aleatorio. Es decir, interpretar el coeficiente de confianza como coeficiente de credibilidad que da una probabilidad a posteriori para el parámetro, dado el conocimiento de los datos de la muestra. Para el ítem 2b, este mismo distractor plantearía un modelo determinista para la interpretación del intervalo de confianza, pues supone que nunca se produciría error alguno.
- Para el distractor (c) en el ítem 2a se confunde entre el valor de la variable y el parámetro a estimar, error descrito por Olivo (2008). Para el ítem 2b, este distractor supone no interpretar el coeficiente de confianza en forma frecuencial.

**Contenido 3: Contraste bilateral y unilateral. Asignación de hipótesis.**

**Ítem 3a.** Queremos conocer si los sujetos extrovertidos e introvertidos difieren en la puntuación media en autoestima y no disponemos de ninguna información previa. El tipo de hipótesis nula razonable que debo plantear es:

a. $\mu_I \leq \mu_E$
b. $\boldsymbol{\mu_I = \mu_E}$
c. $\mu_I \geq \mu_E$

**Ítem 3b**. Supongamos que no se dispone de alguna información previa y se quiere conocer si los sujetos hombres y mujeres difieren en la variabilidad de autoestima. El tipo de hipótesis nula razonable que debo plantear es:

a. $\sigma^2_H \leq \sigma^2_M$
b. **$\sigma^2_H = \sigma^2_M$**
c. $\sigma^2_H \geq \sigma^2_M$

Estos dos ítems evalúan la discriminación entre contraste unilateral y bilateral. El enunciado plantea una hipótesis de investigación (en términos verbales), y el alumno debe traducirlo a una expresión simbólica, diferenciando entre hipótesis nula y alternativa y entre un test unilateral y bilateral. Vallecillos (1994) plantea un ítem similar, al estimar una proporción y encuentra un 31,4% de errores en su muestra, indicando que los estudiantes confunden hipótesis estadística nula con hipótesis de investigación en ese ítem.

La respuesta correcta en los dos ítems es la opción b), ya que no se dispone de información previa sobre la hipótesis alternativa. Si designamos por $\mu_I$: la puntuación media poblacional del primer grupo de sujetos y con $\mu_E$: la puntación media poblacional del segundo grupo en el ítem 3a, resulta $H_0: \mu_I = \mu_E$. Análogamente, si designamos por $\sigma_M^2$ la varianza poblacional de la autoestima de las mujeres y con $\sigma_H^2$ la varianza poblacional de la autoestima de los hombres en el ítem 3b, resulta $H_0: \sigma_M^2 = \sigma_H^2$. Los errores que se evalúan en este ítem son los siguientes:

- El distractor (a), para ambos ítems, lo elegiría un alumno que confunde contraste bilateral y unilateral y asigna una hipótesis nula compuesta (que un grupo tiene mayor media (ítem 3a) o variabilidad (ítem 3b)); confusión que se ha encontrado en investigaciones previas, como la de Vallecillos (1994).
- El distractor (c), para ambos ítems, lo elegiría un alumno que confunde el contraste unilateral y bilateral (Vallecillos, 1994); la diferencia con el anterior es la dirección de la hipótesis.

## Contenido 4: Hipótesis nula y alternativa. Asignación de hipótesis

**Ítem 4a**. ¿Cuál de las siguientes hipótesis está bien formulada?:

a. Ho: $\mu = 3$; Hi: $\mu \neq 4$
b. Ho: $\mu = 3$; Hi: $\mu \geq 3$
c. **Ho: $\mu = 3$; Hi: $\mu \neq 3$**

**Ítem 4b**. De los siguientes pares de afirmaciones, indique cual NO cumple con las reglas para plantear hipótesis estadísticas:

**a. Ho: $\sigma = 15$; Ha: $\sigma \leq 15$**
b. Ho: $\mu \leq 100$; Hi: $\mu > 100$
c. Ho: $\mu 1 - \mu 2 = 0$; Hi: $\mu 1 - \mu 2 \neq 0$

Estos ítems evalúan la comprensión de la formulación de hipótesis estadísticas, un tema en el cuál también Vallecillos (1994) encontró dificultades. La respuesta correcta es la (c) para el 4a, y la opción (a) en 4b. Recordemos que las hipótesis estadísticas son complementarias. Así, para el ítem 4a, si la media poblacional es $\mu$ se tendrá que, en el caso (c) $\mu \in \{3\}$ bajo $H_0$ y $\mu \in R - \{3\}$ bajo $H_1$, que son complementarios.

En el ítem 4b, se pide indicar cuál de las hipótesis estadísticas no cumple las condiciones. Si la desviación estándar poblacional es $\sigma$, se tendrá en el caso (a) $\sigma \in \{15\}$ bajo $H_0$ y $\sigma \in (-\infty; 15]$ bajo $H_1$, que no sólo no son complementarios, sino que además no tienen intersección vacía. Los errores que se evalúan en estos ítems son los siguientes:

- Si en el ítem 4a se elige el distractor (b) se toman conjuntos no complementarios para el parámetro bajo ambas hipótesis (con el agravante que la constante 3 está como valor posible del parámetro, tanto en la hipótesis nula como en la alternativa). En el ítem 4b el distractor (b) cumple con las condiciones de ser hipótesis estadística.
- Al elegir el distractor (a) en 4a además de elegir hipótesis no complementarias, también tienen valores comunes. Mientras que si elige el distractor (c) en el ítem 4b, mostraría que el alumno desconoce la formulación para hipótesis estadísticas, pues las hipótesis para este distractor están bien formuladas.

## Contenido 5: Errores Tipo I y II, nivel de significación y potencia

**Ítem 5a$_1$.** Supongamos un contraste bilateral sobre la media, siendo la variable estudiada la inteligencia. Para $H_0$: $\mu = 100$, $H_1$: $\mu = 110$, $\alpha = 0{,}05$ y $\beta = 0{,}4406$. ¿Cuál es la probabilidad de rechazar la $H_0$ cuando no es „cierta"?

a. 0,05
b. 0,4406
**c. 0,5594**

**Ítem 5a$_2$**. Supongamos que conocemos la „verdad absoluta" sobre la eficacia de dos tratamientos (A y B), y sabemos que existen diferencias en la efectividad de ambos para curar la depresión. Un investigador que realice un estudio y parta de la hipótesis *„no existen diferencias en la efectividad de los tratamientos A y B para curar la depresión"* cometerá un *error tipo II* cuando:

a. Concluya que A y B no son efectivos para curar la depresión
**b. Concluya que A y B no difieren en su efectividad para curar la depresión**
c. Concluya que A y B difieren en su efectividad para curar la depresión

**Ítem 5b$_1$**. Se desea realizar un contraste bilateral sobre la media de una población, siendo la variable estudiada el nivel de adaptación al medio. Para ello se quiere decidir entre $H_0$: $\mu = 130$, $H_1$: $\mu = 131$, $\alpha = 0{,}01$ y $\beta = 0{,}0228$. ¿Cuál es la probabilidad de rechazar $H_0$ cuando sea „falsa"?.

a. 0,01
b. 0,0228
**c. 0,9772**

**Ítem 5b$_2$**. Si en una investigación lees que la potencia de un contraste vale 0,5594, entonces interpretas que:

a. La probabilidad de rechazar la hipótesis nula siendo falsa vale (1 -0.5594) = 0,4406
**b. La probabilidad de rechazar la hipótesis nula siendo falsa vale 0,5594**
c. La probabilidad de mantener la hipótesis nula siendo falsa vale 0,5594

Estos ítems de elaboración propia pretenden evaluar la comprensión de los errores que pueden cometerse en un contraste (errores de Tipo I y Tipo II), y la potencia; también la diferencia entre hipótesis estadísticas e hipótesis de investigación. En todos estos conceptos se producen errores (Vallecillos, 1994; Batanero, 2000; Contreras, 2011), al interpretar una probabilidad condicional.

En 5a$_1$, se está calculando la potencia para un valor particular dado del parámetro

bajo la alternativa, $\mu = 110$ (respectivamente $\mu = 131$ en $5b_1$). La respuesta correcta es la (c) para ambos ítems. Si denotamos por $\alpha\ y\ \beta$ las probabilidades de cometer Errores de Tipo I y de Tipo II, siendo $\beta = 0{,}4406$ para el ítem $5a_1$ ($\beta = 0{,}0228$ para el ítem $5b_1$); la probabilidad pedida para la potencia es $\pi = 1 - 0{,}4406 = 0{,}5594$ (ítem $5a_1$), y $\pi = 1 - 0{,}0228 = 0{,}9772$ (ítem $5b_1$).

Tanto en $5b_2$ como en $5a_2$, la respuesta correcta es la (b). Consideremos $\mu_A\ y\ \mu_B$ el verdadero valor medio de la puntuación de la efectividad al aplicar el tratamiento A y B para curar la depresión. Con esto, las hipótesis estadísticas a contrastar en este problema son: $H_0{:}\,\mu_A = \mu_B\ vs\ H_1{:}\,\mu_A \neq \mu_B$. Se cometerá un Error de Tipo II si decide que no existe diferencia entre los dos tratamientos para curar la depresión, cuando en realidad si existen. Los errores que se evalúan en este ítem son los siguientes:

- Al elegir la opción (a) para los ítems $5a_1$ y $5b_1$, estaría confundiendo la potencia del test con el nivel de significación, siendo esta la probabilidad de cometer Error de Tipo I, $P(rechazar\ H_0, siendo\ cierta)$. Si eligiese esta opción para $5b_2$ estaría confundiendo la probabilidad de Error Tipo II con la potencia del test. De elegirla en el ítem $5a_2$ $se\ olvida\ una\ hipótesis\ explícita\ en\ el\ problema$ , ya que se afirma que conocemos la „verdad absoluta" sobre la eficacia de los dos tratamientos A y B.
- Al elegir la opción (b) en el ítem $5a_1$ o $5b_1$ estaría confundiendo la potencia del test con la probabilidad de cometer Error de Tipo II ó $P(no\ rechazar\ H_0, siendo\ falsa)$, ($\beta = 0{,}4406$, ítem 5a1; $\beta = 0{,}0228$, $5b_1$).
- Si elige la opción (c) para $5b_2$ estaría confundiendo la probabilidad de cometer Error Tipo II con la potencia del test. Finalmente si elige la opción (c) en el $5a_2$ estaría confundiendo el Error de Tipo I con el de Tipo II, ya que el Error de Tipo I en este problema es afirmar que existe diferencia entre los dos tratamientos aplicados para curar la depresión, cuando en realidad no la hay.

**Contenido 6: Cálculo de valor crítico, dado el nivel de significación**

**Ítem 6a**. La puntuación típica correspondiente a una α= 0,01 en un contraste unilateral derecho es:

**a. 2,33**
b. -2,33
c. 3,10

**Ítem 6b.** La puntuación típica correspondiente a una α= 0,05 en un contraste unilateral izquierdo es:

a. 1,64
**b. -1,64**
c. 2,33

En estos ítems se evalúa la comprensión de la relación entre el nivel de significación y el valor crítico, la diferenciación entre test unilateral y bilateral, y el concepto de puntuación típica. Implícitamente se asume que se puede aplicar la distribución normal. Además evaluamos la comprensión alcanzada para el manejo de tablas estadísticas.

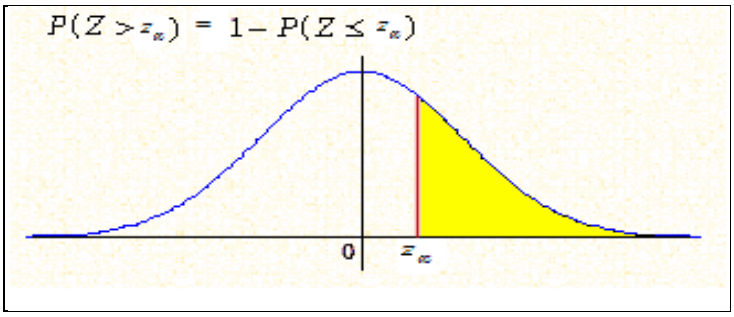



Figura 3.1. Cálculo de áreas en la Normal (0,1)

En el ítem 6a la respuesta correcta es la (a). La puntuación $z_\alpha$, es aquél valor que deja un área $\alpha$ a su derecha bajo la curva de densidad *N* (0,1) (parte sombreada de la Figura 3.1), teniendo en cuenta que se trata de un test unilateral derecho, o sea bajo las hipótesis $H_0: \mu = \mu_0 \; vs \; H_1: \mu > \mu_0$. Como el área bajo la curva es de una unidad, si $\alpha = 0{,}01 < 0{,}5\,, z_\alpha > 0$. Análogamente la respuesta correcta es la (b) para el ítem 6b, donde la hipótesis alternativa $H_1: \mu < \mu_0$, indica una región de rechazo de la forma $z_{\text{observado}} \leq$ c, donde c es un número negativo. En nuestro problema, resulta $z_\alpha = -z_{0{,}05} = -1{,}64$. Los errores que se evalúan en este ítem son los siguientes:

- Si la respuesta elegida fuese el distractor (b) para el ítem 6a, se toma el test unilateral izquierdo (respectivamente derecho si elige opción (a) en 6(b)) o lo que es lo mismo confunde hipótesis nula y alternativa o bien la región crítica y de aceptación en un contraste unilateral. Además, si la puntuación típica fuese negativa, el valor de $\alpha$ debería ser mayor que 0,5 en el ítem 6a.

  Si se elige el distractor (c) en 6a, hay un error en la lectura de la tabla, ya que $P\left(Z_{N(0,1)} > 3{,}10\right) \approx 0{,}00097 \approx 0$. Mientras que esta misma opción en 6b se estaría cometiendo un doble error, considerando un test unilateral derecho, y usando un valor $\alpha = 0{,}001$.

### Contenido 7: Regla de decisión en un contraste

**Ítem 7a**. Cuando realizamos un contraste, la regla de decisión nos lleva a rechazar la hipótesis nula siempre que:

a. El estadístico de contraste caiga en la región de rechazo
b. La probabilidad asociada al estadístico de contraste (el valor de significación) sea menor que el valor de alfa
**c. a) y b) son correctas**

**Ítem 7b.** Cuando realizamos un contraste de hipótesis, la regla de decisión nos lleva a no rechazar la hipótesis nula siempre que:

a. El estadístico de contraste no caiga en la región de rechazo
b. La probabilidad asociada al estadístico de contraste (el valor de significación) sea mayor que el valor de alfa
**c. a) y b) son correctas**

Evaluamos con estos dos ítems, adaptados de otros de Vallecillos (1996), el conocimiento de la regla de decisión de rechazar la hipótesis nula, no comprendida por algunos estudiantes. También se evalúa la comprensión de la elección de la región de aceptación y la de rechazo. Si el estadístico de prueba cae en la primera zona

(respectivamente la segunda), se decide que existe suficiente evidencia como para rechazar la hipótesis nula (respectivamente no rechazarla). Es decir, es correcta la respuesta (a) tanto en 7a como en 7b.

- Si la probabilidad asociada al estadístico de contraste es menor (respectivamente mayor) que el valor de $\alpha$ (nivel de significación) el valor observado del estadístico cae en la zona de rechazo (respectivamente aceptación), de donde la opción (b) también es correcta tanto en 7a como en 7b. Por tanto la respuesta correcta en ambas unidades de medida es la (c), ya que tanto las opciones (a) como la (b) son correctas.
- Si eligiera sólo la respuesta (a) ó la (b) estaría desconociendo la relación de equivalencia entre las dos respuestas.

### 3.5.2. PARTE II. ANÁLISIS DE VARIANZA

El resto de los ítems fueron construidos para este trabajo, pues no encontramos antecedentes en la investigación previa. Nos basamos en los libros de texto utilizados en el análisis del significado de referencia.

**Contenido 8: Selección de un modelo de Análisis de Varianza, de acuerdo con las características de un problema**

**Ítem 8a.** Una maestra cree que unas nuevas actividades de lectura ayudarán a mejorar la capacidad lectora de los niños de primaria. La maestra tiene una clase con 21 niños a los que les pasa la prueba de lectura DRP (Degree of Reading Power), para conocer el nivel del que parten. Después realiza estas actividades en clase durante 8 semanas. Al final del período vuelve a pasarles la prueba. De las siguientes técnicas, ¿cuál debería aplicar los investigadores para comprobar si las actividades modifican las capacidades lectoras?

a. Contraste de hipótesis sobre dos medias independientes
**b. Contraste de hipótesis sobre dos medias relacionadas**
c. Análisis de varianza de un factor completamente aleatorizado

**Ítem 8b$_1$.** Para mejorar la psicomotricidad de los niños de primaria, una maestra cree que ayudarán unas nuevas actividades físicas. La maestra divide aleatoriamente su grupo de trabajo en tres partes iguales. A cada grupo le aplica un tipo de ejercicio diferente, pues desea saber qué tipo de ejercicios le dará mejores resultados. De las técnicas estadísticas que siguen, ¿cuál debería aplicar un investigador para comprobar si los métodos que aplica son diferentes?

a. Contraste de hipótesis sobre medias independientes
b. Contraste de hipótesis sobre medias relacionadas
**c. Análisis de varianza de un factor completamente aleatorizado**

**Ítem 8b$_2$.** Un investigador utilizará un Análisis de Varianza de dos factores, efectos fijos y completamente aleatorizado cuando:

a. En el estudio haya dos variables dependientes.
b. En el estudio haya una variable independiente, con dos niveles seleccionados al azar.
c. **En el estudio haya dos variables independientes, cada una con dos o más niveles.**

A través de estos ítems se desea evaluar si el alumno es capaz de escoger el modelo que mejor se adecua al problema. La respuesta correcta es la (b) para el ítem 8a, ya que la maestra pasa a los mismos individuos la prueba antes y después de ser sometidos a un tratamiento. El procedimiento adecuado es el análisis de la varianza de un factor fijo (Montgomery, 2002).

La respuesta correcta es la (c) para el ítem 8b$_1$, ya que la maestra divide aleatoriamente su grupo de trabajo en tres partes iguales, antes de aplicar algún

tratamiento diferente, generando así los niveles del único factor que quiere analizarse. El interés aquí se centraría en contrastar la igualdad de las tres medias de los tratamientos, versus que existe algún par de medias que son distintas. El análisis de varianza de dos factores exige la presencia de dos variables independientes (los factores) con dos o más niveles de estudio. Los errores que se evalúan en este ítem son los siguientes:

- Al elegir la opción (a), en el ítem 8a se estaría confundiendo el concepto de independencia de muestras, ya que aquí se toma la misma población para pasar la prueba al cabo de un período, y claramente las muestras son dependientes. Si elige igual opción en el ítem $8b_1$, estaría aplicando un contraste de hipótesis válido a lo sumo para dos muestras de datos. Mientras que en $8b_2$, de elegir esta opción no tendría en cuenta que se trata de un análisis de varianza de dos factores, siendo un requisito para aplicarlo la independencia (Dunn y Clark, 1987).
- Al elegir como correcta la opción (c), en 8a, estaría desconociendo que para elegir este modelo, se requiere al menos comparar tres muestras independientes. Mientras que si elige la opción (b), en el $8b_1$; estaría pensando en muestras relacionadas, y aquí son independientes. Finalmente si elige la opción (b) en $8b_2$, estaría confundiendo el análisis de Varianza de un factor con el de dos, y además se confunde los factores fijos con los aleatorios, donde los niveles del factor se eligen al azar entre una población de niveles (Bono, Roser y Vallejo, 2008).

### Contenido 9: Supuestos de aplicación de un modelo Análisis de varianza

**Ítem 9a.** Los supuestos del Análisis de Varianza de dos factores, efectos fijos y completamente aleatorizado son:

a. Independencia de las observaciones, normalidad de las distribuciones, y aditividad
b. Independencia de las observaciones, igualdad de varianza y aditividad
c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas**

**Ítem 9b.** Los supuestos del Análisis de varianza de un factor, efectos fijo y completamente aleatorizado son:

a. Independencia de las observaciones, normalidad de las distribuciones, y aditividad
b. Independencia de las observaciones, igualdad de varianza y aditividad
c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas**

En estos ítems, se quiere evaluar la comprensión de los supuestos para aplicar el análisis de varianza, que son: independencia, aleatoriedad, homocedasticidad y normalidad (Dunn y Clark, 1987). La respuesta correcta es la (c). Los errores que se evalúan en este ítem son los siguientes:

- De escoger como correcto el distractor (a) para ambos ítems no estaría considerando el supuesto de homocedasticidad. En algunas ocasiones la varianza de las observaciones se incrementa con la magnitud de las observaciones; en este caso, los residuos se harían mayores conforme las observaciones se hicieran más grandes (Montgomery, 2002). El problema podría ser grave, si alguna de la varianzas se disparara con respecto a las demás.
- De escoger como correcto el distractor (b) en ambos ítems no estaría considerando el supuesto de normalidad que debe estar presente en el modelo pues se requiere para aplicar las pruebas *F*. Si no se cumple, una posibilidad es transformar las

variables para que se ajusten a la normalidad o bien usar métodos no paramétricos adecuados (Tejedor, 1999). En 1.4.9, se destaca como se ven afectados los modelos frente a la falta de normalidad.

### Contenido 10: Comprobación de supuestos en un modelo de análisis de varianza

**Ítem 10a**. La prueba que se utiliza para comprobar el supuesto de esfericidad en un estudio con una V.I. con medidas repetidas es:

a. La prueba de Levene
b. La prueba de Kolmogorov
**c. La prueba de Mauchly**

**Ítem $10b_1$**. La prueba que se utiliza para comprobar el supuesto de normalidad en un estudio con una V.I. es:

a. La prueba de Levene
**b. La prueba de Kolmogorov**
c. La prueba de Mauchly

*Ítem $10b_2$*. La hipótesis nula que somete a contraste la prueba de Levene es la siguiente:

a. **Ho: $\sigma^2_A = \sigma^2_B$**
b. Ho: $\sigma^2_A \neq \sigma^2_B$
c. Ho: $\mu_{SI} = \mu_{NO}$.

En estos ítems se quiere evaluar la comprensión de las diversas pruebas estadísticas que se utilizan para evaluar los supuestos en análisis de varianza. Uno de ellos, en el análisis de varianza de medidas repetidas es el de esfericidad (Dunn y Clark, 1987). La respuesta correcta es la opción (c) en el 10a. La respuesta correcta en el $10b_1$ es la opción (b), ya que este es el contraste usado para verificación de la normalidad. En el ítem $10b_2$, se busca que el alumno conozca cual es la hipótesis nula que se contrasta al aplicar una prueba de Levene, de donde la respuesta correcta es la opción a). Los errores son:

- De escoger el distractor (a) tanto en 10a como en $10b_1$, estaría confundiendo la prueba de Mauncly y la de Kolmogorov con la de Levene (utilizada para testear la igualdad de varianzas). Si en $10b_2$, eligiese el distractor (b) no se trataría además de una hipótesis nula adecuada.
- De escoger el distractor (b) para el 10a, confunde con la prueba de Kolmogorov para chequear el grado de acercamiento de los datos a alguna distribución prefijada (la normal en este particular). Mientras que si escoge la opción (c) en el $10b_1$ confunde con la prueba de esfericidad. Análogamente si elige en $10b_2$ la opción (c) estaría calculando una prueba de medias para dos muestras.

### Contenido 11: Modelo lineal asociado a un tipo de análisis de varianza

**Ítem 11a.** El modelo lineal utilizado para representar las fuentes de variabilidad presentes en un Análisis de varianza de un factor, efectos fijos completamente aleatorizado es:

**a. $Y_{ij}=\mu + \alpha_j + E_{ij}$**
b. $Y_{ij}=\mu + \alpha_j + \beta_i + E_{ij}$
c. $Y_{ijk}=\mu + \alpha_j + \beta_i + (\alpha\beta)_{ij} + E_{ijk}$

**Ítem 11b.** El Análisis de Varianza de dos factores, con efectos fijos, descompone la variabilidad total en los siguientes componentes:

a. Variabilidad total = V. entre grupos + V. error
b. Variabilidad total = V. entre grupos + V. entre sujetos + V. error
**c. Variabilidad total = V. factor A + V. factor B + V. interacción + V. error**

En estos ítems tratamos de evaluar la comprensión para asociar un modelo estadístico, de acuerdo con el estudio. El modelo lineal que mejor se adecua al problema 11a es: $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, de modo que la respuesta correcta es la opción (a). Para el ítem 11b, el formato para la descomposición de la variación total es: Variación Total = *V*. factor *A* + *V*. factor *B* + *V*. Interacción + *V*. error, por lo que la opción (c) es la correcta. Los errores que se evalúan en este ítem son los siguientes:

- En el caso de elegir la opción (b) en 11a estaría considerando un modelo aditivo, con dos factores cuyos efectos serían $\alpha_i \; y \; \beta_j$ . El alumno confunde los modelos para uno y dos factores. Si eligiese esta misma opción en 11b, no estaría considerando en la descomposición de la variabilidad la presencia e dos factores, y la interacción.
- En el caso de elegir la opción (c) en 11a estaría considerando un modelo completo para dos factores, con un efecto interacción entre los mismos, a través del término $(\alpha\beta)_{ij}$. Mientras que si eligiese la opción (a) en 11b, desconoce la presencia de dos factores en el modelo, sólo estaría considerando uno de ellos.

### Contenido 12: Cálculo en el análisis de varianza, medidas repetidas

**Ítem 12a**. Si en un Análisis de Varianza de un factor, y medidas repetidas encuentro que la F empírica u observada toma un valor de 8,16 esto quiere decir que:

a. CM inter grupos / CM error = 8,16
**b. CM intra sujetos / CM error = 8,16**
c. CM inter grupos / CM intrasujetos = 8,16

**Ítem 12b.** Si en un Análisis de Varianza de un factor, y medidas repetidas encuentro que la F empírica u observada toma un valor de 8,16 esto quiere decir que:

**a. CM entre grupos / CM error = 8,16**
b. CM entre sujetos / CM error = 8,16
c. CM entre grupos / CM intra sujetos = 8,16

A través de estos ítems evaluamos la comprensión sobre algunos de los cálculos realizados para completar una tabla de análisis de varianza para un factor con medidas repetidas. También se particulariza la construcción del valor *F* observado para el único factor presente, el factor intra-sujetos ó entre-grupos. La respuesta correcta es la (b) en 12a (así como la opción a) para 12b), pues en este diseño se analiza la variabilidad intra sujetos (entre grupos), ya que la aleatorización sólo se aplica a los sujetos dentro de cada grupo (Tejedor, 1999). Los errores que se evalúan en este ítem son los siguientes:

- De elegir el distractor (a) en ítem 12a, el sujeto estaría analizando la variabilidad introducida por la formación de los grupos, que no es relevante para este tipo de experimentos, sino para el diseño ordinario de un factor completamente aleatorizado. El alumno confunde medidas repetidas con factor de efectos fijos
- De elegir como correcto el distractor (c) consideraría el análisis de la interacción entre dos factores, que no está presente para este modelo.

**Contenido 13: Cálculo en análisis de varianza. Modelo de dos factores**

En estos ítems se quiere evaluar la comprensión de los procedimientos que se siguen para completar la tabla de análisis de varianza en el diseño factorial de dos factores con efectos fijos. En particular, la descomposición de la variación total, los grados de libertad (*GL*) y el cálculo del cuadrado medio (*CM*) para todas las fuentes de variación; (en este caso dos factores *A* y *B* y la interacción *AB*).

**Ítem 13a$_1$** y **13a$_2$.** Se quiere estudiar el efecto de ciertas variables motivacionales sobre el rendimiento en tareas de logro. Se manipularon dos variables: "tipo de entrenamiento motivación" (A1: instrumental; A2: atribucional y A3: control) y "clima de clase" (B1: cooperativo; B2: competitivo y B3: individual). Se seleccionaron a 45 sujetos y se dividieron en grupos para cada condición experimental.
A continuación presentamos la tabla del análisis de varianza incompleta. A partir de la información de la tabla contesta a las preguntas del Ítem 13a y del Ítem 13b.

Tabla 1

| Fuente de variación | SC | GL | MC | F |
|---|---|---|---|---|
| Factor A | 70 | | | |
| Factor B | | | 20 | |
| Interacción AB | | | | 3,91 |
| Error | 46 | | 1,278 | |
| Total | 176 | 44 | | |

**Ítem 13a$_1$.**
El valor del sumatorio cuadrado para el factor B (ver Tabla 1) es:
a. 15,65
b. 35
**c. 40**

**Ítem 13a$_2$**. El valor de la media cuadrática para el factor A (ver Tabla 1) es:
a. 5
b. **35**
c. 15,65

**Ítem 13b.** Se quiere estudiar el efecto de ciertas variables laborales sobre la "transferencia negativa del trabajo a la familia" en mujeres. Se recogieron datos de dos variables: "tipo de trabajo" (A1: cualificación alta; A2: cualificación media y A3: cualificación baja) y "estabilidad en el trabajo" (B1: si; B2: no). Se seleccionaron a 35 mujeres y se dividieron en grupos para cada condición experimental.

A continuación presentamos la tabla del análisis de varianza incompleta. A partir de la información de la tabla contesta a las preguntas del Ítem 13 y del Ítem 14.

Tabla 2

| Fuente de variación | SC | Gl | MC | F |
|---|---|---|---|---|
| Factor A (tipo de trabajo) | 70 | | | |
| Factor B (estabilidad) | 10 | | | |
| Interacción | 30 | | | |
| Error | 86 | | | |
| Total | 196 | | | |

**Ítem 13b.** El valor de la media cuadrática del error (ver Tabla 2) es:
a. 86
**b. 2,96**
**c.** 3,38

La respuesta correcta es la (c) para el ítem 13a$_1$, y opción b) para el ítem 13a$_2$, que proporciona los sumatorios cuadrados del factor *B* y *A* respectivamente; y opción (b) del

ítem 13b, representando la media cuadrática del error ($CM_{Error}$). En las Tabla 3.2 y 3.3, se dan las soluciones. Los valores críticos son los que han de compararse con el valor observado del estadístico para ver si caemos en la zona de rechazo ó de no rechazo del contraste (ver 1.4.7) y tomar la decisión.

Tabla 3.2. Tabla de análisis de varianza, solución del problema. Ítem $13a_1$ y $13a_2$

| | Fuente de variación | SC | GL | MC | F | Valor crítico | Decisión |
|---|---|---|---|---|---|---|---|
| F1 | A | 70 | 2 | 35 | 27,3 | 3,2 | Rechazar |
| F2 | B | **40** | 2 | 20 | 15,65 | 3,2 | Rechazar |
| F3 | INT. | 20 | 4 | 5 | 3,91 | 2,6 | Rechazar |
| | ERROR | 46 | 36 | 1,2 | | | |
| | TOTAL | 176 | 44 | | | | |

Los errores que se evalúan en este ítem son los siguientes:

- De escoger como correcto el distractor (a), confunde la suma de cuadrados pedida para el factor *B* con el valor *F* para el mismo factor (ítem $13a_1$); (valor observado del estadístico de prueba). Mientras que si lo elige para el ítem $13a_2$, confunde con la media de cuadrados de la interacción (Tabla 3.2), y si lo elige en 13b, confunde con el sumatorio cuadrado para el error (Tabla 3.3).
- De escoger como correcto el distractor (b) confunde con media de cuadrados del factor A ($CM_{Factor\ A}$) (ítem $13a_1$); y si elige como correcto el distractor (c) tanto en $13a_2$ como en 13b; confunde con el *F* observado para el factor *B* (ver Tablas 3.2 y 3.3 respectivamente).

Tabla 3.3. Tabla análisis de varianza, solución del problema. Ítem 13b

| | Fuente de v*ariación* | *SC* | GL | MC | F | Valor crítico | Decisión |
|---|---|---|---|---|---|---|---|
| F1 | A | 70 | 2 | 35 | 13,46 | 3,33 | Rechazar |
| F2 | B | 10 | 1 | 10 | 3,38 | 4,18 | No Rechazar |
| F3 | $I_{NT.}$ | 30 | 2 | 15 | 5,07 | 3,33 | Rechazar |
| | ERROR | 86 | 29 | **2,96** | | | |
| | TOTAL | 196 | 34 | | | | |

## Contenido 14: Interpretación de una tabla análisis de varianza. Modelo de dos factores

**Ítem 14a**. Una de las conclusiones del estudio sería (alfa = 0,05) (ver Tabla 1)

a. **Hay efecto del factor A ("entrenamiento") sobre el rendimiento en tareas de logro**
b. No hay efecto del factor B ("clima de clase") sobre el rendimiento en tareas de logro
c. No hay interacción de los factores

**Ítem 14b.** Una de las conclusiones del estudio sería (alfa = 0,05) (ver Tabla 2)

a. **No hay efecto del factor B ("estabilidad en el trabajo") sobre la transferencia negativa trabajo-familia**
b. No hay efecto del factor A ("tipo de trabajo") sobre la transferencia negativa trabajo-familia
c. No hay interacción de los factores A y B

Con estos ítems se quiere evaluar la interpretación de los valores obtenidos en la tabla análisis de varianza, para dar una respuesta estadísticamente válida a cada problema. El modelo de efectos ajustado es: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$, y en la Tabla 1.4.2, se dan los resultados de las pruebas de hipótesis necesarias.

La respuesta correcta es (a) para ambos ítems, ya que, para asegurar una influencia tanto de factores principales (*A* y/o *B*) como de la interacción *AB*, deberá comprobarse que el estadístico de prueba cae en la zona de rechazo del contraste. Esto último es equivalente a decir que el valor del estadístico observado $F_{Obs}$ es mayor al percentil correspondiente (respectivamente menor que el percentil correspondiente). Los errores que se evalúan en este ítem son los siguientes:

- De elegir como correcto el distractor (b) estaría confundiendo el criterio para rechazar la hipótesis nula con el criterio de no rechazarla para el factor *B* en 14a (confusión de análoga naturaleza pero a la inversa, para el factor *A* en el ítem 14b).
- De elegir como correcto el distractor (c) estaría confundiendo criterio para rechazar la hipótesis nula de igualdad de efectos para las interacciones en ambas unidades de medida, pues aquí, por todo lo explicado existe efecto de la interacción de ambos factores sobre la respuesta.

## Contenido 15: Interpretación de salida de software. Identificación del modelo de análisis de varianza

El estudio HBSC sobre Conductas de los Escolares relacionadas con la Salud en su edición de 2006 realizó encuestas a niños de 38 países en edad escolar. Hemos extraído los datos de un grupo de niños a los que se les pidió que evaluaran la "calidad en la relación con su mejor amigo" en una escala de 0 a 10. También se clasificó a los niños según su "sexo" y la "autoestima" medida con la escala de Rosenberg considerándose dos valores (baja / alta autoestima). Los resultados obtenidos tras analizar los datos fueron los siguientes:

Tabla 1. Estadísticos descriptivos

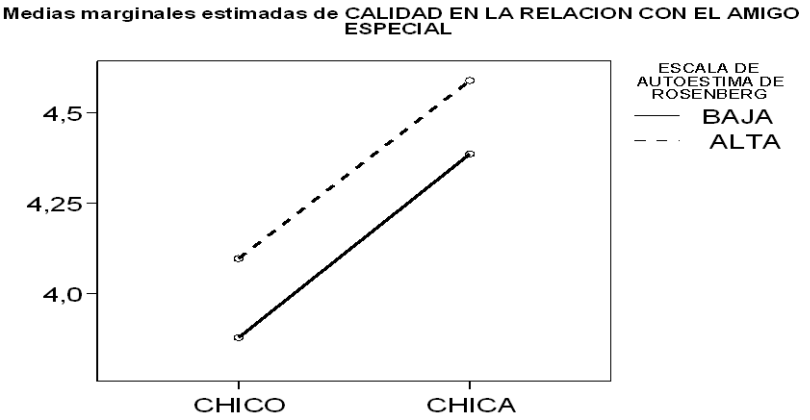
Variable dependiente: CALIDAD EN LA RELACION CON EL AMIGO ESPECIAL

| SEXO | AUTOESTIMA | Media | Desv. típ. | N |
|---|---|---|---|---|
| CHICO | BAJA | 3,877 | ,9808 | 13 |
| | ALTA | 4,096 | ,6636 | 26 |
| CHICA | BAJA | 4,386 | ,6637 | 29 |
| | ALTA | 4,590 | ,4932 | 10 |

Tabla 2. Pruebas de efectos inter-sujetos.

| Fuente | Suma de cuadrados tipo III | gl | Media cuadrática | F | Significación |
|---|---|---|---|---|---|
| Modelo corregido | 4,090(a) | 3 | 1,363 | 2,721 | ,050 |
| Intersección | 1149,723 | 1 | 1149,723 | 2294,721 | ,000 |
| Sexo | 4,027 | 1 | 4,027 | 8,038 | ,006 |
| Autoestima | ,716 | 1 | ,716 | 1,429 | ,236 |
| sexo * autoestima | ,001 | 1 | ,001 | ,002 | ,965 |
| Error | 37,076 | 74 | ,501 | | |
| Total | 1437,320 | 78 | | | |
| Total corregida | 41,166 | 77 | | | |

R cuadrado = ,099 (R cuadrado corregida = ,063)




De acuerdo con este estudio (Tabla 2) responder el ítem 15a.

**Ítem 15a**. El tipo de análisis aplicado en este estudio es:

a. Análisis de varianza de dos factores, efectos fijos y con medidas repetidas
**b. Análisis de varianza de dos factores, efectos fijos y completamente aleatorizado**
c. Análisis de varianza de un factor, efectos aleatorios y completamente aleatorizado

La incorporación de la mujer al mercado de trabajo es un fenómeno que se verifica en todos los países de nuestro entorno geográfico y cultural. La clave de este proceso ha sido el interés creciente de la mujer por adquirir mayor formación. Pero el proceso es lento, con dificultades, sobre todo cuando se trata de integrar la vida familiar y laboral. En este año hemos estudiado la influencia que tienen diferentes variables sobre la satisfacción laboral.

Se han decido estudiar si hay efecto del "tipo de trabajo" (A1: trabajos de cualificación baja; A2: trabajos de cualificación media o A3: trabajos de cualificación alta) influyen sobre la "satisfacción laboral". También se ha tenido en cuenta, al mismo tiempo, si la "flexibilidad en el horario" (B1: si; B2: no) influye en la "satisfacción laboral".

Tabla 3. Pruebas de los efectos inter-sujetos

| Fuente | Suma de cuadrados | *GL* | Media cuadrática | *F* | Significación |
|---|---|---|---|---|---|
| Modelo corregido | 465,614(a) | 5 | 93,123 | ,852 | ,516 |
| Intersección | 17192,763 | 1 | 17192,763 | 157,355 | ,000 |
| Tipo de trabajo | 49,120 | 2 | 24,560 | ,225 | ,799 |
| Flexibilidad horaria | 270,878 | 1 | 270,878 | 2,479 | ,008 |
| Tipo * Flexibilidad | 34,383 | 2 | 17,191 | ,157 | ,855 |
| Error | 12564,981 | 115 | 109,261 | | |
| Total | 33298,000 | 121 | | | |
| Total corregida | 13030,595 | 120 | | | |

De acuerdo con este estudio responder el item 15b

**Ítem 15b**. El tipo de análisis aplicado en este estudio es:

a. Contraste de dos medias independientes.
b. Análisis de varianza de un factor, con medidas repetidas.
**c. Análisis de varianza de dos factores, completamente aleatorizado.**

Con de estos ítems se quiere evaluar si los estudiantes saben interpretar las salidas que ofrece un software estadístico, en este caso, SPSS. La respuesta correcta es la (b) en 15a (y la (c) en 15b)), pues se analiza la influencia en una variable dependiente, a saber:

"calidad en la relación con el amigo especial", (respectivamente "satisfacción laboral" en 15b) de dos factores independientes: la autoestima y el sexo, (respectivamente el tipo de trabajo y la flexibilidad horaria en 15b). De donde se trata de un Análisis de varianza de dos factores con efectos fijos y completamente aleatorizado. Los errores que se evalúan en este ítem son:

- De escoger el distractor a) estaría confundiendo el modelo de medidas repetidas con el de dos factores completamente aleatorizado, para 15a. En el modelo de medidas repetidas entra en juego un factor de bloqueo, para evitar problemas que podrían surgir por la heterocedasticidad, ó para casos en los que no sea posible realizar todas las corridas en un mismo momento. En cambio si decidiese por esta opción en 15b, reduce el problema a una prueba *t* de dos muestras independientes para medias poblacionales.
- De escoger (c) estaría confundiendo el modelo unifactorial con el bifactorial en 15a; y con el unifactorial con medidas repetidas si optara por la opción b) en 15b.

### Contenido 16: Interpretación de salida de software. Resultados

**Ítem 16a$_1$**. La condición experimental en la que el valor para la variable "calidad en la relación con el amigo especial" es menor es

a. **Chicos con baja autoestima**
b. Chicas con baja autoestima
c. Chicos con alta autoestima

**Ítem 16a$_2$**. Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha$=0,05 (Tabla 2):

a. Existe interacción entre el "sexo" y la "autoestima".
b. La "autoestima" influye sobre la "calidad en la relación con los amigos".
c. **El "sexo" influye sobre la "calidad en la relación con los amigos".**

**Ítem 16b**. Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha$=0,05 (Tabla3):

a. No hay diferencia estadísticamente significativas en la "satisfacción laboral" en función del "tipo de trabajo".
b. Hay efecto del factor "flexibilidad horaria" sobre la "satisfacción laboral".
c. **Hay interacción entre la "flexibilidad horaria" y "tipo de trabajo".**

Este ítem también evalúa la interpretación de las salidas de un software estadístico, que proporciona los resultados obtenidos en la tabla de análisis de varianza, y también un gráfico estadístico. La respuesta correcta es la (a) en 16a$_1$, y se desprende del gráfico de *medias marginales estimadas de calidad en la relación con el amigo especial*. Cada uno de los cuatro puntos en el gráfico nos indica el promedio de puntuación de la calidad en la relación con el amigo especial con respecto a una baja/alta autoestima, siendo chico ó chica. En el ejemplo no hay interacción, ya que los chicos tanto con baja como con alta autoestima, mantienen igual diferencia de la calidad en relación con el amigo especial como las chicas.

En el ítem 16a$_2$ la respuesta correcta es la (c), ya que en la última columna de la Tabla 2 se observa un valor de significación igual a 0,006, mucho menor que el alfa considerado para el estudio, a saber: 0,05, por lo que se decide rechazar la hipótesis nula. Análogo razonamiento aparece para 16b, usando el factor "flexibilidad horaria" sobre la variable satisfacción laboral. Los errores que se evalúan en este ítem son:

- De escoger el distractor (b) en el ítem 16a$_1$ se estaría desconociendo la interpretación del gráfico de interacciones, ya que las chicas, tengan tanto alta como

baja autoestima siempre mantienen los niveles más altos en la calidad en la relación con el amigo especial. Si opta por este distractor en $16a_2$ se hace una interpretación incorrecta de la significación del factor autoestima, que es superior a la puntuación 0,05. Análogo razonamiento hacemos en el caso que elija la opción (b) en 16b

- De escoger el distractor (c) en $16a_1$ también desconocería la interpretación del gráfico de interacciones. Si eligiera como correcto el distractor (a) en $16a_2$, estaría respondiendo que la interacción sería significativa, aunque el valor p correspondiente es superior al nivel 0,05 establecido. Análogo razonamiento hacemos si eligiera el distractor (a) para 16b.

## Contenido 17. Interpretación de salida de software. Supuestos del modelo de análisis de varianza

Posteriormente se realizó una comparación en la misma variable ("calidad en la relación con el amigo especial") entre niños de familias monoparentales y biparentales. Los resultados son los siguientes:

Tabla 5. Prueba de muestras independientes

| | | Prueba de Levene | | Prueba T para la igualdad de medias | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | F | Sig. | T | Gl | Sig. (bilateral) | Diferencia de medias | Error típ. diferencia | 95% I. confianza diferencia | |
| | | | | | | | | | Inferior | Superior |
| Calidad relación con el amigo especial | Varianzas iguales | ,900 | ,346 | 4,437 | 87 | ,000 | ,7327 | ,1651 | ,4045 | 1,0609 |
| | Varianzas desiguales | | | 3,342 | 24,079 | ,003 | ,7327 | ,2192 | ,2803 | 1,1851 |

Usar esta información para responder ítem 17a.

**Ítem 17a**. De acuerdo con la prueba de Levene (Tabla 5):

a. No se cumple el supuesto de homocedasticidad
b. **Se cumple el supuesto de homocedasticidad**
c. No se puede concluir sobre la homocedasticidad

Una vez analizados los supuestos de aplicación, vamos a estudiar si la "estabilidad en el trabajo" influye sobre la "satisfacción laboral". Estos son los resultados.

Tabla 6. Estadísticos de grupo

| estabilidad del trabajo | N | Media | Desviación típ. |
|---|---|---|---|
| Si | 92 | 14,34 | 11,546 |
| No | 29 | 8,52 | 2,339 |

Tabla 7. Prueba de muestras

| | Prueba de Levene | | Prueba T para la igualdad de medias | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | F | Sig. | t | Gl | Sig. (bilateral) | Diferencia de medias | Error típ. diferen. | 95% IC para la diferencia | |
| Varianzas | | | | | | | | Inferior | Superior |
| iguales | 3,39 | ,16 | 2,69 | 119 | ,008 | 5,820 | 2,164 | 1,535 | 10,104 |
| Varianzas distintas | | | 4,54 | 110,17 | ,000 | 5,820 | 1,280 | 3,284 | 8,356 |

Usar esta información para responder ítem $17b_1$.

**Ítem $17b_1$**. De acuerdo con la prueba de Levene (Tabla 7):

a. Se cumple el supuesto de homocedasticidad
b. **No se cumple el supuesto de homocedasticidad**
c. No se puede concluir sobre la homocedasticidad

La incorporación de la mujer al mercado de trabajo es un fenómeno que se verifica en todos los países de nuestro entorno geográfico y cultural. La clave de este proceso ha sido el interés creciente de la mujer por adquirir mayor formación. Pero el proceso es lento, con dificultades, sobre todo cuando se trata de integrar la vida familiar y laboral. En este año hemos estudiado la interrelación entre variables laborales y familiares. Vamos a analizar la influencia que tienen diferentes variables sobre la satisfacción laboral.

Antes de comenzar a realizar análisis hemos realizado la prueba de Kolmogorov. Obteniendo los siguientes resultados

Tabla 8. Prueba de Kolmogorov-Smirnov para una muestra

| | | Satisfacción |
|---|---|---|
| Parámetros normales(a,b) | Media | 12,94 |
| | Desviación típica | 10,421 |
| Diferencias más extremas | Absoluta | ,372 |
| | Positiva | ,372 |
| | Negativa | -,223 |
| Z de Kolmogorov-Smirnov | | 4,094 |
| Sig. Asintót. (bilateral) | | ,384 |

Usar esta información para responder ítem $17b_2$

**Ítem $17b_2$**. De acuerdo con la prueba de Kolmogorov ($\alpha$=0,05) (Tabla 8):

a. No se cumple el supuesto de homocedasticidad
b. No se cumple el supuesto de normalidad
c. **Se cumple el supuesto de normalidad**

A través de estos tres ítems, se quiere evaluar la interpretación de las salidas de un software estadístico; se particulariza en la interpretación de los supuestos del modelo: la homocedasticidad y la normalidad.

Al analizar la Tabla 5, el valor *p* es 0,346 en la prueba de Levene para responder al Ítem 17a (el valor *p* en igual prueba, en Tabla 7 para el ítem $17b_1$ es 0,16). Con esta evidencia en 17a, asumir desigualdad de varianzas sería un riesgo muy alto; no así para $17b_1$, ya que un riesgo del 16%, es razonable. De allí que la respuesta correcta sea la (b) para ambos ítems. Para el Ítem $17b_2$ se analiza el supuesto de normalidad usando la prueba de Kolmogorov. Observando en la Tabla 8 que la significación asintótica toma el valor 0,384; no se tiene suficiente evidencia para rechazar la hipótesis nula (los datos provienen de una población con distribución normal), de donde la respuesta correcta es la opción (c). Los errores que se evalúan en estos ítems son:

- De elegir el distractor (a) en 17a no haría una interpretación correcta de la Tabla 3, pues de la misma se desprende que, no se cumple el supuesto de homocedasticidad; para este problema por cada 100 muestras tomadas en iguales condiciones, en casi 35 de ellas, las varianzas no resultarían estadísticamente iguales. Si se elige este distractor en $17b_1$, también se hace una interpretación incorrecta (asumiendo un gran riesgo de aceptar la normalidad en datos). Si elige esta opción en $17b_2$ estaría pensando en otra prueba estadística (la de Levene en este contexto) y no en la de Kolmogorov como se pide en el enunciado.
- Si eligiese el distractor (c) en 17a cometería error pues, aunque generalmente cuando el valor p es menor que el nivel de significación no tomamos conclusiones, en este caso, como no rechazamos la hipótesis que las varianzas son diferentes, la única opción es asumir que son estadísticamente iguales. Situación análoga ocurre

para $17b_1$. Si elige la opción (c) en $17b_2$ cometería un error del orden del 38%, lo estimamos como alto.

### 3.5.3. RESUMEN DEL CONTENIDO DE LOS ÍTEMS

Resumimos el análisis del banco de ítems en las Tablas 3.4 y 3.5., en las que hemos marcado con una X el contenido estadístico que es evaluado directamente por esa unidad, y con una S el que es evaluado en forma secundaria. El contenido se ha dividido en dos grandes grupos, el de los prerrequisitos y el bloque de análisis de la varianza propiamente dicho. Observamos que dicho contenido incluye los diferentes objetos matemáticos descritos en el enfoque ontosemiótico que intervienen o emergen de los sistemas de prácticas:

- *Lenguaje:* verbal y simbólico, tabular y gráfico, así como salidas del software.
- *Situaciones-problemas*: Se proponen problemas de estimación de parámetros (intervalos de confianza) y contraste de hipótesis, tanto referidos a la comparación de dos muestras (contraste t), como la comparación de varias muestras independientes (análisis de varianza de efectos fijos) o relacionadas (análisis de varianza con medidas repetidas)
- *Conceptos-definición:* diferencia entre muestras independientes y relacionadas, distribución muestral, funciones de distribución, distribución de probabilidad, aleatoriedad, intervalo de confianza, hipótesis estadísticas y de investigación, valor p, nivel de significación, error tipo I, tipo II, función de potencia, región crítica y de aceptación, tipos de factores, efecto, media global, grupal, error, interacción entre factores, normalidad, homocedasticidad, linealidad, esfericidad, grado de libertad.
- *Proposiciones*: relacionadas con la esperanza y varianza matemática, con la descomposición de la varianza, condiciones de independencia de variables aleatorias y de dependencia, estandarización de variables aleatorias. Proposiciones en relación con propiedades de la distribución normal, de Student, Chi cuadrado, de Fisher y la relación entre ellas, en relación con el teorema central del límite.
- *Procedimientos*: Lectura de tablas, cálculo de probabilidades, construcción del intervalo de confianza, asignar hipótesis estadísticas, para encontrar el valor p, hallar tanto la potencia de una prueba de hipótesis como el error tipo II, cálculo de efectos, de errores, realizar los distintos tipos de gráficos, comprobar supuestos, para calcular los estadísticos (tanto para contraste de dos medias, como de dos varianzas, de más de dos varianzas y del *F*).
- *Argumentos*: Los alumnos necesitarán comprobar ejemplos y contraejemplos de propiedades y aplicar razonamiento deductivo.

Tabla 3.4. Contenido de los ítems del banco. Ítems 1 al 9

| | Contenido evaluado | Ítems | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1a | 1b | 2a | 2b | 3a | 3b | 4a | 4b | $5a_1$ | $5b_1$ | $5b_2$ | $5a_2$ | 6a | 6b | 7a | 7b | 8a | $8b_1$ | $8b_2$ | 9a | 9b |
| PRERREQUISTOS | Lecturas de tablas de distribución y cálculo de probabilidades | x | x | s | s | | | | | x | x | x | s | x | x | s | s | | | | s | s |
| | Diferenciar muestras independientes y relacionadas | s | s | x | x | s | s | | s | | | | s | | | | | x | x | x | x | x |
| | Distribución muestral Variabilidad del estadístico en diferentes muestras | x | x | x | x | | | | | | | | | | | | | | | | s | s |
| | ***Intervalo de confianza*** | | | | | | | | | | | | | | | | | | | | | |
| | Construcción | x | x | s | s | | | | | | | | | | | | | | | | | |
| | Interpretación | | | x | x | | | | | | | | | | | | | | | | | |
| | ***Contraste de hipótesis*** | | | | | | | | | | | | | | | | | | | | | |
| | Hipótesis; Asignación de hipótesis | | | | | x | x | x | x | | | | s | s | s | | | s | s | s | | |
| | Hipótesis nula y alternativa | | | | | x | x | x | x | | | | s | s | s | x | x | s | s | s | | |
| | Valor p; nivel de significación, probabilidad error I | | | | | | | | | x | x | x | x | x | x | x | x | | | | | |
| | Potencia, probabilidad error II | | | | | | | | | x | x | x | x | | | | | | | | | |
| | Selección de un contraste | | | | | x | x | s | s | s | s | | | | | s | s | x | x | x | s | s |
| | Región crítica y de aceptación | | | | | | | | | | | | | | x | x | | | | | | |
| | Contraste unilateral y bilateral | s | s | | | x | x | s | s | s | s | | s | x | x | | | s | s | s | | |
| | Regla de decisión | s | s | | | | | | | | | | s | | | x | x | s | s | | | |
| ANOVA | Supuesto de independencia | | | | | | | | | | | | | | | | | x | x | x | x | x |
| | Normalidad y homocedasticidad | | | | | | | | | s | s | | | s | s | | | s | s | s | x | x |
| | Comprobación de supuestos | s | s | | | | | | | s | s | | s | | | | | s | s | s | x | x |
| | Modelo lineal | | | | | | | | | | | | | | | | | x | x | x | s | s |
| | Cálculo del estadístico *F* | | | | | | | | | | | | | | | | | | | | | |
| | Interpretación de la tabla y de gráficos de Análisis de varianza | | | | | | | | | | | | | | | | | | | | | |
| | Factor y sus tipos | | | | | | | | | | | | | | | | | x | x | x | x | x |
| | Interacción | | | | | | | | | | | | | | | | | | | | | |
| | Modelo de efectos fijos | | | | | | | | | | | | | | | | | x | x | x | x | x |
| | Modelo de efectos aleatorios | | | | | | | | | | | | | | | | | s | s | s | s | s |
| | Modelo de un factor | | | | | | | | | | | | | | | | | x | x | x | x | x |
| | Modelo de medidas repetidas | | | | | | | | | | | | | | | | | x | x | x | s | s |
| | Modelo de dos factores | | | | | | | | | | | | | | | | | | | x | x | x |

Tabla 3.5. Contenido de los ítems del banco. Ítems 10 al 17

| | contenido evaluado | ítems | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 a | 10 $b_1$ | 10 $b_2$ | 11 a | 11 b | 12 a | 12 b | 13 $a_1$ | 13 $a_2$ | 13 b | 14 a | 14 b | 15 a | 15 b | 16 $a_1$ | 16 $a_2$ | 16 b | 17 a | 17 $b_1$ | 17 $b_2$ |
| prerrequisitos | Lecturas de tablas de distribución y cálculo de probabilidades | | | | | | s | s | x | x | x | x | x | | | | s | s | s | s | s |
| | Diferenciar muestras independientes y relacionadas | x | x | x | | | s | s | s | s | s | x | x | x | x | x | x | x | s | s | s |
| | Distribución muestral variabilidad del estadístico en diferentes muestras | | | | | s | | | s | s | s | s | s | | | | s | s | s | s | s |
| | ***Intervalo de confianza*** | | | | | | | | | | | | | | | | | | | | |
| | Construcción | | | | | | | | | | | | | | | | | | | | |
| | Interpretación | | | | | | | | | | | | | | | | | | | | |
| | ***Contraste de hipótesis*** | | | | | | | | | | | | | | | | | | | | |
| | Hipótesis; asignación de hipótesis | x | x | x | | | | | | | | s | s | s | s | | | | s | s | s |
| | Hipótesis nula y alternativa | x | x | x | | | | | | | | | | | | | | | s | s | s |
| | Valor p; nivel de significación, probabilidad error I | | | | | | | | | | | x | x | | | | x | x | x | x | x |
| | Potencia, probabilidad error II | | | | | | | | | | | | | | | | | | | | |
| | Selección de un contraste | x | x | x | | | | | | | | s | s | | | | | | | | |
| | Región crítica y de aceptación | | | | | | | | | | | s | s | | | | s | s | s | s | s |
| | Contraste unilateral y bilateral | | | x | | | | | | | | s | s | | | | | | s | s | s |
| | Regla de decisión | | | | | | | | | | | x | x | | | x | x | x | x | x | x |
| | Supuesto de independencia | | | | | | | | | | | | | s | s | | s | s | | | |
| | Normalidad y homocedasticidad | x | x | x | | | | | | | | | | s | s | | | | x | x | x |
| | Comprobación de supuestos | x | x | x | | | | | | | | | | | | | | | x | x | x |
| | Modelo lineal correspondiente a un análisis de varianza | | | | x | x | | | | | | | | s | s | s | s | s | | | |
| análisis de varianza | Cálculo del estadístico *f* | | | | | | x | x | x | x | x | x | x | s | s | | | | s | s | |
| | Interpretación de la tabla y de gráficos de análisis de varianza | | | | | | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Factor y sus tipos | x | x | | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x | x |
| | Interacción | | | | x | x | s | s | s | s | s | x | x | x | x | x | x | x | | | |
| | Modelo de efectos fijos | x | x | | x | x | s | s | s | s | s | x | x | x | x | x | x | x | | | |
| | Modelo de efectos aleatorios | s | s | | x | x | s | s | s | s | s | | | s | s | s | s | s | | | |
| | Modelo de un factor | x | x | x | x | x | s | s | | | | s | s | s | s | s | s | s | | | |
| | Modelo de medidas repetidas | x | x | x | s | s | x | x | s | s | s | s | s | x | x | s | s | s | | | |
| | Modelo de dos factores | x | x | x | x | x | | | x | x | x | s | s | x | x | x | x | x | | | |

Al sintetizar las Tablas 3.4 y 3.5 y clasificar los ítems respecto a las especificaciones fijadas en la Sección 3.3 obtenemos la Tabla 3.6, donde observamos que nuestro banco contiene entre 2 y 4 ítems por contenido. En la siguiente fase usaremos el juicio de expertos y las pruebas empíricas de ítems para seleccionar de entre todos los ítems del banco, al menos uno por cada contenido para nuestro cuestionario definitivo.

Tabla 3.6. Especificaciones del contenido del cuestionario por ítems

| PRERREQUISITOS | Ítems |
|---|---|
| 1. Construcción de un intervalo de confianza | 1a, 1b |
| 2. Interpretación del significado de un intervalo de confianza. | 2a, 2b |
| 3. Contraste bilateral y unilateral. Asignación de hipótesis | 3a, 3b |
| 4. Hipótesis nula y alternativa. Asignación de hipótesis | 4a, 4b |
| 5. Errores tipo I y II, nivel de significación y potencia | $5a_1$,$5b_1$,$5b_2$,$5a_2$ |
| 6. Cálculo de valor crítico, dado el nivel de significación | 6a, 6b |
| 7. Regla de decisión en un contraste | 7a, 7b |
| ANÁLISIS DE VARIANZA | |
| 8. Selección de un modelo Análisis de varianza, de acuerdo con las características de un problema | 8a,$8b_1$,$8b_2$ |
| 9. Supuestos de aplicación de un modelo Análisis de varianza. | 9a, 9b |
| 10. Comprobación de supuestos en un modelo Análisis de varianza | 10a,$10b_1$,$10b_2$ |
| 11. Modelo lineal asociado a un tipo de Análisis de varianza | 11a,11b |
| 12. Cálculo en el Análisis de varianza, Medidas repetidas | 12a,12b |
| 13. Cálculo en el Análisis de varianza. Modelo de dos factores | $13a_1$,$13a_2$,13b |
| 14. Interpretación de una tabla Análisis de varianza. Modelo de dos factores | 14a,14b |
| 15. Interpretación de salida de software. Identificación del modelo de Análisis de varianza | 15a,15b |
| 16. Interpretación de salida de software. Resultados | $16a_1$,$16a_2$,16b |
| 17. Interpretación de salida de software. Supuestos del modelo Análisis de varianza | 17a,$17b_1$,$17b_2$ |

## 3.6. PRUEBAS PILOTO DE ÍTEMS

Elaborado el banco de ítems, realizamos unas pruebas piloto de los mismos para obtener información empírica sobre sus características psicométricas, en particular la dificultad y discriminación de los ítems, y ayudados por el juicio de expertos seleccionar los que formarían el cuestionario definitivo. La valoración empírica de los ítems, se llevó a cabo administrándolos a algunas muestras de estudiantes que pasamos a describir.

### 3.6.1. MUESTRA Y MATERIAL

La muestra de participantes en las pruebas piloto estuvo formada por alumnos de segundo año de la Licenciatura en Psicología de la Universidad de Huelva, que habían cursado la asignatura de Análisis de Datos II, descrita en el Capítulo 1 y cuya profesora colaboró con nosotros de dos cursos diferentes (2007-2008) y (2008-2009). Para ello se dividió el conjunto de ítems en dos cuestionarios, con objeto de no aumentar la fatiga de los participantes en los ensayos piloto. Se repartieron los ítems en forma que quedase al menos uno de cada contenido en cada uno de los cuestionarios.

Para evitar la copia de los alumnos se prepararon varias versiones con la única diferencia en el orden en que aparecen las preguntas en la prueba, y el orden en que se colocan los distractores en cada unidad de medida, manteniendo la redacción. También con ello se evita que las preguntas finales quedasen, por efecto del cansancio sin responder. Uno de los cuestionarios fue respondido por 93 alumnos y el otro por 130 alumnos.

### 3.6.2. MÉTODO

La administración de estos ítems se llevó a cabo en condiciones lo más semejantes posibles a las que se utilizarán en la forma del cuestionario final, con la intención de conseguir información objetiva y específica tanto de los ítems, como de las dificultades, legibilidad y ajuste al tiempo previsto.

Cada instrumento se completó como una actividad de evaluación ordinaria para los estudiantes, asegurando de este modo que las condiciones eran similares a las que se tendrían al momento de pasar nuestro instrumento definitivo, y que los participantes habían estudiado el contenido a evaluar. Se penalizó las respuestas incorrectas para evitar que respondieran al azar y se resolvieron posibles dudas en el enunciado, tomando nota para mejorar la redacción de los ítems.

### 3.6.3. RESULTADOS Y DISCUSIÓN

Los datos que obtuvimos al administrar los cuestionarios a las dos muestras, fueron codificados y analizados usando la misma metodología empleada por Díaz (2007) y Olivo (2008). En primer lugar calculamos el número de respuestas en cada categoría de cada ítem junto con los porcentajes, los índices de dificultad y discriminación. En la Tabla 3.7, presentamos los resultados en cada ítem y distractor, marcando en negrita las respuestas correctas. Se detectan ítems, donde algunos distractores apenas fueron elegidos (por ejemplo: ítems $5b_1$; $5b_2$, así como ítems 7a; 7b; $10b_2$; 17a y $17b_1$); pero, en la mayoría se presentó bastante equilibrio entre los distractores.

En la Tabla 3.8, se presenta el índice de dificultad o proporción de alumnos que contestaron correctamente el ítem: a mayor dificultad del ítem, menor será su índice (Backhoff, Larrazolo y Rosas, 2000; Losada y López Feal, 2003; Viladrich, 2012). Pudimos también calcular, intervalos de confianza para la proporción verdadera que contestaría correctamente el ítem en la población (Tabla 3.8) de la cual se obtuvo la muestra, utilizando la aproximación normal, puesto que nuestras muestras eran de más de 30 elementos. Los métodos clásicos de la inferencia estadística parten de los supuestos tales como: ausencia de datos atípicos, verificar alguna estructura en los datos y comprobar si se cumplen algunas hipótesis de interés como puede ser su ajuste a la distribución normal. Para comprobar estas condiciones siguiendo a Salvador Figueras y Gargallo (2003) se representa gráficamente mediante histogramas y gráfico de caja y bigotes los índices de dificultad obtenidos, con el fin de estudiar la forma de la distribución y determinar la posible existencia de varias modas en la misma que pongan de manifiesto la presencia de diversos grupos homogéneos en la muestra o de ítems con una dificultad extrema.

Estas gráficas (Figura 3.2) nos proporcionan la distribución de índices de dificultad en el conjunto de ítems, donde observamos la presencia de dos valores atípicos (ítem 2a y el 12a); se trata de una distribución unimodal (da idea de un grupo homogéneo), pero no coincide el centro de simetría con el valor central, vemos desde la figura suavizada del histograma existe una cierta asimetría a la izquierda (los valores atípicos están hacia este lado de la distribución, y el coeficiente de asimetría es negativo). Diremos que el ajuste a la distribución normal no es muy bueno. Los valores tienden a acumularse hacia la derecha de la distribución de los datos, lo cual correspondería a afirmar que los cuestionarios poseen una tendencia a ser resueltos con facilidad por la mayoría de los alumnos, reflejando el aprendizaje de la asignatura.

Tabla 3.7. Frecuencia (y porcentaje) de respuestas en cada distractor del ítem

| Ítem | **N** | Opción a | Opción b | Opción c | Blanco |
|---|---|---|---|---|---|
| 1a | 93 | 1 (1,1) | 2(2,2) | **72(77,4)** | 18 (19,4) |
| 1b | 130 | 15(11,5) | 0(0) | **91(70)** | 24(18,5) |
| 2a | 93 | **10(10,8)** | 35(37,6) | 4(4,3) | 44(47,3) |
| 2b | 130 | **29(22,3)** | 4(3,1) | 18(13,8) | 79(60,8) |
| 3a | 93 | 4(4,3) | **86(92,5)** | 2(2,2) | 1(1,1) |
| 3b | 130 | 0(0) | **123(94,6)** | 2(1,5) | 5(3,8) |
| 4a | 93 | 3(3,2) | 1(1,1) | **87(93,5)** | 2(2,2) |
| 4b | 130 | **121(93,1)** | 3(2,3) | 5(3,8) | 1(0,8) |
| $5a_1$ | 93 | 10(10,8) | 6(6,5) | **70(75,3)** | 7(7,5) |
| $5b_1$ | 130 | 10(7,7) | **50(38,5)** | 44(33,8) | 26(20) |
| $5b_2$ | 130 | 11(8,5) | **66(50,8)** | 41(31,5) | 12(9,2) |
| $5a_2$ | 93 | 1(1,1) | **60(64,5)** | 17(18,3) | 15(16,1) |
| 6a | 93 | **52(55,9)** | 23(24,7) | 5(5,4) | 13(14) |
| 6b | 130 | 5(3,8) | **118(90,8)** | 1(0,8) | 6(4,6) |
| 7a | 93 | 24(25,8) | 2(2,2) | **55(59,1)** | 12(12,9) |
| 7b | 130 | 26(20) | 4(3,1) | **89(68,5)** | 11(8,5) |
| 8a | 93 | 3(3,2) | **76(81,7)** | 3(3,2) | 11(11,8) |
| $8b_1$ | 130 | 22(16,9) | 10(7,7) | **94(72,3)** | 4(3,1) |
| $8b_2$ | 130 | 5(3,8) | 11(8,5) | **102(78,5)** | 12(9,2) |
| 9a | 93 | 12(12,9) | 5(5,4) | **48(51,6)** | 28(30,1) |
| 9b | 130 | 9(6,9) | 4(3,1) | **65(50)** | 52(40) |
| 10a | 93 | 5(5,4) | 5(5,4) | **72(77,4)** | 11(11,8) |
| $10b_1$ | 130 | 9(6,9) | **106(81,5)** | 5(3,8) | 10(7,7) |
| $10b_2$ | 130 | **75(57,7)** | 3(2,3) | 34(26,2) | 18(13,8) |
| 11a | 93 | **49(52,7)** | 11(11,8) | 3(3,2) | 30(32,3) |
| 11b | 130 | 5(3,8) | 9(6,9) | **108(83,1)** | 8(6,2) |
| 12a | 93 | 58(62,4) | **7(7,5)** | 15(16,1) | 13(14) |
| 12b | 130 | **100(76,9)** | 5(3,8) | 6(4,6) | 19(14,6) |
| $13a_1$ | 47 | 0(0) | 4(8,5) | **30(63,8)** | 13(27,7) |
| $13a_2$ | 46 | 0(0) | **34(73,9)** | 2(4,3) | 10(21,7) |
| 13b | 130 | 7(5,4) | **91(70)** | 3(2,3) | 29(22,3) |
| 14a | 93 | **24(25,8)** | 6(6,5) | 5(5,4) | 58(62,4) |
| 14b | 130 | **56(43,1)** | 10(7,7) | 10(7,7) | 54(41,5) |
| 15a | 93 | 7(7,5) | **65(69,9)** | 5(5,4) | 16(17,2) |
| 15b | 130 | 1(0,8) | 13(10) | **106(81,5)** | 10(7,7) |
| $16a_1$ | 93 | **79(84,9)** | 3(3,2) | 1(1,1) | 10(10,8) |
| $16a_2$ | 93 | 22(23,7) | 10(10,8) | **38(40,9)** | 23(24,7) |
| 16b | 130 | 28(21,5) | **40(30,8)** | 19(14,6) | 43(33,1) |
| 17a | 93 | 10(10,8) | **58(62,4)** | 2(2,2) | 23(24,7) |
| $17b_1$ | 130 | **80(61,5)** | 21(16,2) | 9(6,9) | 20(15,4) |
| $17b_2$ | 130 | **94(72,3)** | 4(3,1) | 11(8,5) | 21(16,2) |

Tabla 3.8. Estimación de los índices de dificultad

| Ítem | N | Índice dificultad | Intervalo confianza 95% Límite inferior | Límite superior | Ítem | N | Índice dificultad | Intervalo confianza 95% Límite inferior | Límite superior |
|---|---|---|---|---|---|---|---|---|---|
| 1a | 93 | 0,77 | 0,65 | 0,90 | 9b | 130 | 0,50 | 0,35 | 0,65 |
| 1b | 130 | 0,70 | 0,57 | 0,83 | 10a | 93 | 0,77 | 0,65 | 0,90 |
| 2a | 93 | 0,11 | 0,02 | 0,20 | $10b_1$ | 130 | 0,82 | 0,70 | 0,93 |
| 2b | 130 | 0,22 | 0,10 | 0,34 | $10b_2$ | 130 | 0,58 | 0,43 | 0,72 |
| 3a | 93 | 0,92 | 0,85 | 1,00 | 11a | 93 | 0,53 | 0,38 | 0,67 |
| 3b | 130 | 0,95 | 0,88 | 1,01 | 11b | 130 | 0,83 | 0,72 | 0,94 |
| 4a | 93 | 0,94 | 0,86 | 1,01 | 12a | 93 | 0,08 | 0,00 | 0,15 |
| 4b | 130 | 0,93 | 0,87 | 0,99 | 12b | 130 | 0,77 | 0,65 | 0,89 |
| $5a_1$ | 93 | 0,75 | 0,63 | 0,88 | $13a_1$ | 47 | 0,64 | 0,50 | 0,78 |
| $5b_1$ | 130 | 0,38 | 0,24 | 0,53 | $13a_2$ | 46 | 0,74 | 0,61 | 0,87 |
| $5b_2$ | 130 | 0,51 | 0,36 | 0,65 | 13b | 130 | 0,70 | 0,57 | 0,83 |
| $5a_2$ | 93 | 0,65 | 0,51 | 0,78 | 14a | 93 | 0,26 | 0,13 | 0,39 |
| 6a | 93 | 0,56 | 0,41 | 0,70 | 14b | 130 | 0,43 | 0,29 | 0,57 |
| 6b | 130 | 0,91 | 0,82 | 0,99 | 15a | 93 | 0,70 | 0,57 | 0,83 |
| 7a | 93 | 0,59 | 0,45 | 0,73 | 15b | 130 | 0,82 | 0,70 | 0,93 |
| 7b | 130 | 0,68 | 0,55 | 0,82 | $16a_1$ | 93 | 0,85 | 0,75 | 0,95 |
| 8a | 93 | 0,82 | 0,70 | 0,93 | $16a_2$ | 93 | 0,41 | 0,27 | 0,55 |
| $8b_1$ | 130 | 0,72 | 0,59 | 0,85 | 16b | 130 | 0,31 | 0,17 | 0,44 |
| $8b_2$ | 130 | 0,78 | 0,66 | 0,90 | 17a | 93 | 0,62 | 0,48 | 0,76 |
| 9a | 93 | 0,52 | 0,37 | 0,66 | $17b_1$ | 130 | 0,62 | 0,47 | 0,76 |
| | | | | | $17b_2$ | 130 | 0,72 | 0,59 | 0,85 |

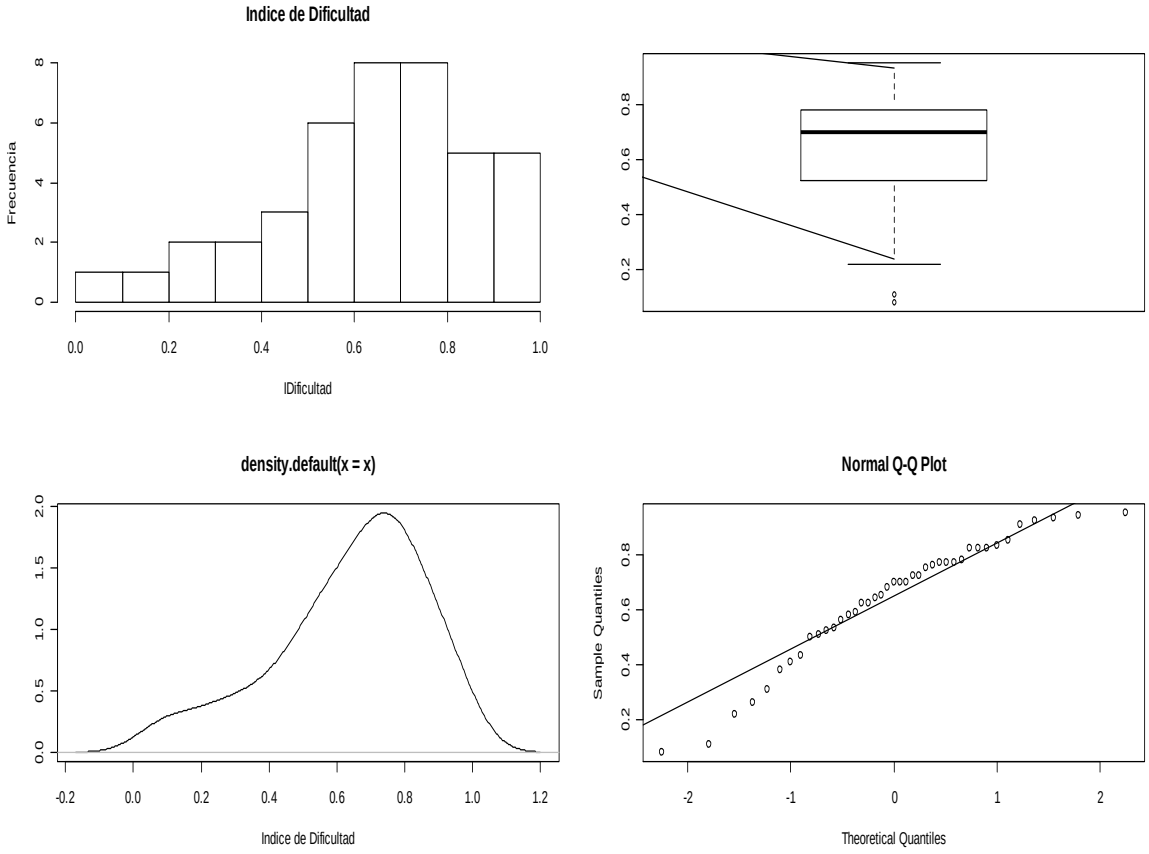



Figura 3.2. Ajuste a la normalidad de los Índices de dificultad (n=41)

En la Tabla 3.9, se presenta un resumen de estimaciones de estadísticos del índice de dificultad, donde se destaca que la mitad de los valores oscilan entre 0,50 y 0,78 (primer y tercer cuartil de la Tabla 3.9). Esto indica que la mitad de los reactivos se pueden resolver con un lógico nivel de instrucción; la mediana y la media de la distribución difieren poco.

Tabla 3.9. Resumen de estimación de estadísticos del Índice de Dificultad (n=41)

| Medidas Resumen (n=41) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mín. | 1er.C. | Mediana | Media | 3er.C. | Máx. | D. E. | Coef.Asim. | Curtosis | CV |
| 0,08 | 0,50 | 0,70 | 0,64 | 0,78 | 0,95 | 0,22 | -0,83 | 0,22 | 34,1 |

También hemos realizado pruebas de normalidad (Tabla 3.10), cuyos resultados no reflejan un valor aceptable de valor *p* en la prueba de bondad de ajuste de Shapiro-Wilks (más conveniente en muestras pequeñas), pero si en la de Kolmogorov-Smirnov con la corrección de Lilliefors (aconsejado para muestras de tamaño mayor a 30) (Salvador Figueras y Gargallo, 2003). Todo ello nos autorizará a realizar contrastes en condiciones de normalidad supuesta.

Tabla 3.10. Pruebas de normalidad para índice de dificultad (n= 41)

| Pruebas de normalidad | p-valor |
|---|---|
| Prueba de Shapiro-Wilks. W - Estadístico = 0,91 | 0,020 |
| Prueba de Kolmogorov. D-Estadístico = 0,123 | 0,565 |

El índice de dificultad correspondiente a los ítem 2a y el 12a (ambos tomados de la prueba piloto con muestra de tamaño 93) muy cercanos al 0, se comportan como datos atípicos, sin embargo con las unidades 2b y 12b (tomados en la prueba piloto de tamaño 130) de contenido análogo para sendos ítems, se ha logrado mejorar esta medida, y de manera muy notable en el caso del 12b. Veamos ahora los cambios y mejoría que se obtienen con las medidas, si quitásemos de nuestros datos los dos valores atípicos mencionados.

Observando la figura (Figura 3.3) como las tablas (Tablas 3.11 y 3.12), es posible comparar ambos estudios estadísticos en donde: se mejora la distancia entre media y mediana (disminuye), el coeficiente de asimetría continúa siendo negativo pero mayor que el anterior y más cercano al 0; la curtosis más cercana a 0,25 (mesocurtica), indicando que el grado de apuntalamiento de esta distribución es semejante al de la normal, su coefiente de variación (medida que sirve para comparar muestras) disminuye indicando una menor variabilidad.

Tabla 3.11. Resumen de estimación de estadísticos del Índice de Dificultad (n=39)

| Medidas Resumen (n=39) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Mín. | 1er.C. | Mediana | Media | 3er.C. | Máx. | D. E. | Coef.Asim. | Curtosis | CV |
| 0,22 | 0,55 | 0,70 | 0,67 | 0,8 | 0,95 | 0,19 | -0,59 | 0,27 | 28,5 |

Asimismo, ambas pruebas estadísticas de ajuste a la normalidad reflejan una mucho mejor adecuación a esta distribución, basta observar los valores p en la tabla 3.12.

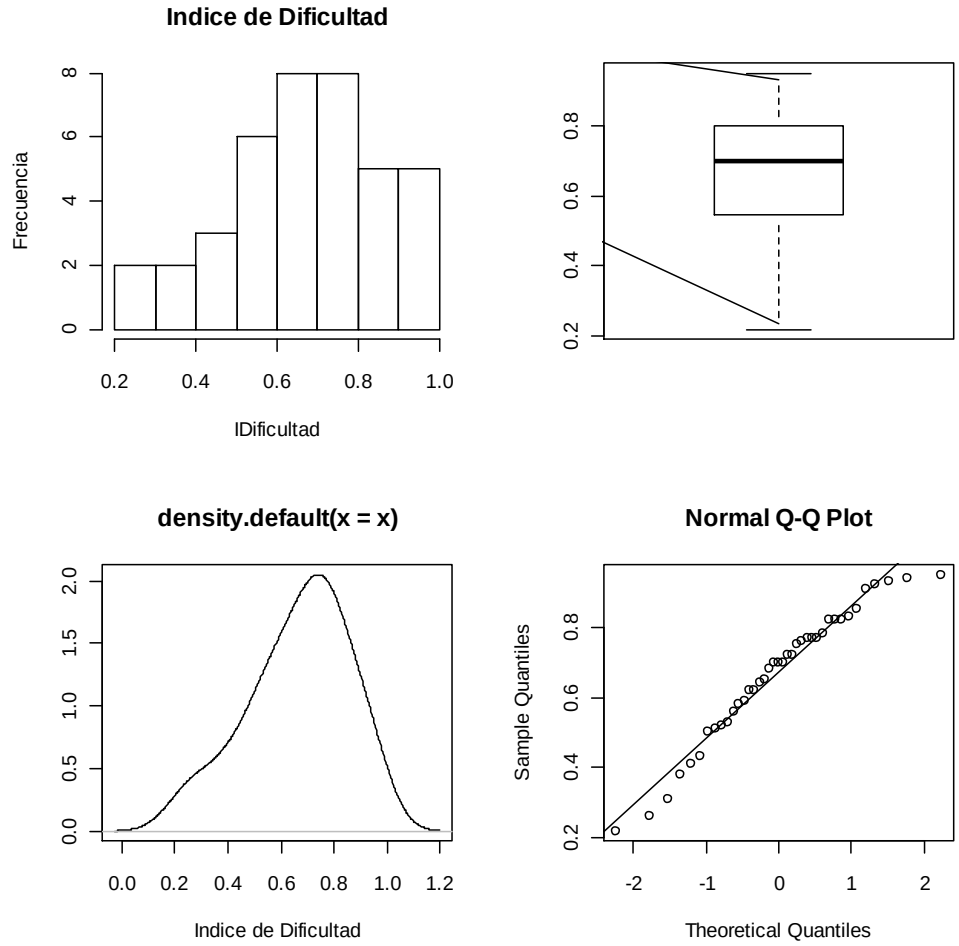



Figura 3.3. Ajuste a la normalidad de los Índices de dificultad (n=39)

Tabla 3.12. Pruebas de normalidad para índice de dificultad (n=39)

| Pruebas de normalidad | p-valor |
|---|---|
| Prueba de Shapiro-Wilks. W - Estadístico = 0,9574 | 0,1457 |
| Prueba de Kolmogorov. D-Estadístico = 0,1115 | 0,717 |

Con esta nueva muestra, quitando los dos valores atípicos, nos aseguramos de la normalidad de los índices de dificultad, y será posible realizar estudios de inferencia posteriores sin dudar sobre la distribución de los mismos.

En cuanto al índice de discriminación brinda información acerca de la capacidad para diferenciar entre alumnos con poco o mucho conocimiento (Barbero, 2003). Un ítem tiene un alto poder de discriminación si lo aciertan casi todos los que han obtenido muchos puntos en la prueba, y fallan la mayoría de los que han obtenido pocos puntos. Para calcular este índice, se seleccionan dos grupos, en función de sus puntuaciones totales de acuerdo con el detalle que a continuación se da:

- Grupo inferior: Todas las puntuaciones inferiores o iguales al percentil 27.
- Grupo superior: Todas las puntuaciones superiores o iguales al percentil 73.

Tabla 3.13. Prueba de muestras independientes para el índice de discriminación (n = 93)

| Ítem | *t* | *p*-valor | Dif. de medias | Error típ. de la dif. | 95% Interv. de conf. | |
|---|---|---|---|---|---|---|
| | | | | | Lím.Inf. | Lím. Sup. |
| Item1a | 4,05 | 0,00 | 0,39 | 0,10 | 0,20 | 0,58 |
| Item2a | 0,75 | 0,46 | 0,07 | 0,09 | -0,11 | 0,24 |
| Item3a | 2,40 | 0,02 | 0,16 | 0,07 | 0,03 | 0,3 |
| Item4a | 0,85 | 0,40 | 0,07 | 0,08 | -0,09 | 0,22 |
| Item5a$_1$ | 0,85 | 0,40 | 0,10 | 0,12 | -0,13 | 0,33 |
| Item5a$_2$ | 2,77 | 0,01 | 0,32 | 0,12 | 0,09 | 0,56 |
| Item6a | 2,95 | 0,01 | 0,36 | 0,12 | 0,11 | 0,6 |
| Item7a | 1,29 | 0,20 | 0,16 | 0,13 | -0,09 | 0,41 |
| Item8a | 2,24 | 0,03 | 0,23 | 0,10 | 0,02 | 0,43 |
| Item9a | 6,55 | 0,00 | 0,65 | 0,10 | 0,45 | 0,84 |
| Item10a | 3,24 | 0,00 | 0,32 | 0,10 | 0,12 | 0,52 |
| Item11a | 4,67 | 0,00 | 0,52 | 0,11 | 0,30 | 0,74 |
| Item12a | 3,73 | 0,00 | 0,42 | 0,11 | 0,20 | 0,64 |
| Item13a$_1$ | 1,77 | 0,09 | 0,31 | 0,17 | -0,05 | 0,66 |
| Item13a$_2$ | 4,86 | 0,00 | 0,55 | 0,11 | 0,32 | 0,78 |
| Item14a | 4,13 | 0,00 | 0,42 | 0,10 | 0,22 | 0,62 |
| Item15a | 3,65 | 0,00 | 0,39 | 0,11 | 0,18 | 0,6 |
| Item16a$_1$ | 1,82 | 0,07 | 0,16 | 0,09 | -0,02 | 0,34 |
| Item16a$_2$ | 6,20 | 0,00 | 0,61 | 0,10 | 0,42 | 0,81 |
| Item17a | 4,27 | 0,00 | 0,45 | 0,11 | 0,24 | 0,66 |

Tabla 3.14. Prueba de muestras independientes para el índice de discriminación (n=130)

| Ítem | *t* | *p*-valor | Dif. de medias | Error típ. de la dif. | 95% Interv. de conf.. | |
|---|---|---|---|---|---|---|
| | | | | | Lím.Inf. | Lím. Sup. |
| Item1b | 4,77 | 0,00 | 0,42 | 0,09 | 0,25 | 0,60 |
| Item2b | 3,36 | 0,00 | 0,29 | 0,09 | 0,12 | 0,47 |
| Item3b | 1,73 | 0,09 | 0,09 | 0,05 | -0,01 | 0,20 |
| Item4b | 2,64 | 0,01 | 0,14 | 0,05 | 0,03 | 0,25 |
| Item5b$_1$ | 2,14 | 0,04 | 0,22 | 0,10 | 0,02 | 0,43 |
| Item5b$_2$ | 3,52 | 0,00 | 0,36 | 0,10 | 0,16 | 0,56 |
| Item6b | 2,82 | 0,01 | 0,19 | 0,07 | 0,06 | 0,32 |
| Item7b | 3,79 | 0,00 | 0,35 | 0,09 | 0,17 | 0,54 |
| Item8b$_1$ | 2,29 | 0,02 | 0,21 | 0,09 | 0,03 | 0,40 |
| Item8b$_2$ | 5,35 | 0,00 | 0,44 | 0,08 | 0,28 | 0,61 |
| Item9b | 3,26 | 0,00 | 0,33 | 0,10 | 0,13 | 0,54 |
| Item10b$_1$ | 5,05 | 0,00 | 0,37 | 0,07 | 0,23 | 0,52 |
| Item10b$_2$ | 7,15 | 0,00 | 0,61 | 0,09 | 0,44 | 0,78 |
| Item11b | 4,52 | 0,00 | 0,35 | 0,08 | 0,20 | 0,50 |
| Item12b | 4,14 | 0,00 | 0,35 | 0,09 | 0,18 | 0,52 |
| Item13b | 6,24 | 0,00 | 0,54 | 0,09 | 0,37 | 0,71 |
| Item14b | 8,2 | 0,00 | 0,66 | 0,08 | 0,50 | 0,82 |
| Item15b | 2,13 | 0,04 | 0,19 | 0,09 | 0,01 | 0,37 |
| Item16b | 3,65 | 0,00 | 0,36 | 0,10 | 0,16 | 0,56 |
| Item17b$_1$ | 7,24 | 0,00 | 0,58 | 0,08 | 0,42 | 0,74 |
| Item17b$_2$ | 5,18 | 0,00 | 0,44 | 0,09 | 0,27 | 0,61 |

Se considera que el ítem "discrimina", si existe diferencia significativa entre la proporción de respuestas en ambos grupos (Tabla 3.13 y 3.14). Las pruebas *t* para diferencias entre estos valores medios para cada ítem se presentan en la Tabla 3.13 y 3.14. Se observa que la mayoría de los ítems discriminan, se tendrá en cuenta esta medida a la hora de la elección definitiva de las unidades para formar el cuestionario objeto de esta investigación.

## 3.7. EVALUACIÓN DE ÍTEMS MEDIANTE JUICIO DE EXPERTOS

Simultáneamente con la prueba de ítems se llevó a cabo un juicio de expertos (Losada y López, 2003). Dicho procedimiento trata de establecer el grado en que el conjunto de ítems del cuestionario representa adecuadamente nuestro dominio de conductas, para asegurar la validez de contenido del cuestionario (Martínez Arias, 1995). Las muestras de expertos se utilizan con asiduidad, proveyendo medios intelectuales para el diseño del instrumento (Hernández, Fernández y Baptista, 1998). En nuestro caso se ha tratado de valorar a comunicabilidad, el nivel adecuado de dificultad y la adecuación a las especificaciones de la prueba (Millman y Greene, 1989; Thorndike, 1989: Viladrich, 2012).

### 3.7.1. EXPERTOS PARTICIPANTES

Se trató que el grupo elegido de expertos representase una diversidad relevante de puntos de vista y capacidades, de acuerdo con Millman y Greene (1989) quienes señalan que el experto lo define el propósito del instrumento. Se exhiben en la Tabla 3.15 las características de dicha muestra, donde se destaca tanto su formación como su experiencia.

Tabla 3.15. Expertos participantes en la valoración del contenido

| | Formación y Experiencia |
|---|---|
| 1 | Matemático. Doctor en Estadística. Profesor de estadística con perfil en docencia e investigación. Investiga en Psicometría e imparte cursos de postgrado de Psicometría. España. |
| 2 | Ingeniero. Doctor en Didáctica de las Matemáticas. Profesor de Probabilidad y Estadística. Catedrático universitario durante 30 años. Realizó una tesis doctoral sobre inferencia para estudiantes universitarios. Investiga en Didáctica de la Estadística. México |
| 3 | Matemático y Estadístico. Doctor en Didáctica de la Matemática, profesor de didáctica de la estadística. Tesis doctoral sobre educación estadística. Argentina. |
| 4 | Matemático. Doctor en Matemática Educativa. Magister en Estadística, profesor de didáctica de la matemática. , Colombia. |
| 5 | Profesor de matemáticas. Doctor en Didáctica de las Matemáticas. Magíster en Estadística. Línea de investigación Didáctica de la Probabilidad y Estadística. Sus trabajos están orientados a la enseñanza y aprendizaje de objetos estadísticos en los estudios universitarios. Chile. |
| 6 | Matemático. Doctor en didáctica de la Matemática, profesor de estadística. Autor de libros para la enseñanza. Ha publicado innumerables artículos de investigación relacionados con enseñanza de la Estadística. España. |
| 7 | Matemático. Doctor en Matemáticas, tesis doctoral en educación estadística. Catedrático de Universidad del área de Conocimiento de Didáctica de la Matemática. Profesor de didáctica de la matemática y de estadística. España. |
| 8 | Estadístico. Doctor, Profesor de estadística en ingeniería y matemáticas. Tesis doctoral en Educación Estadística. Colombia. |
| 9 | Matemático y Estadístico. Licenciado en Pedagogía de las Matemáticas, con título profesional de Estadístico. Máster en estadística, y Doctor en Estadística con tesis sobre aprendizaje de la Estadística. Profesor en la carrera y en la maestría de estadística. España |
| 10 | Matemático, Doctor. Maestro en Ciencias con mención en Estadística. Consultor estadístico. México. |
| 11 | Psicólogo. Máster en Educación Matemática. Doctorado en Ciencias con especialidad en Matemática Educativa. Diplomado en Estadística Aplicada. Profesor y coordinador académico diplomado en Bioestadística y Metodología de la Investigación. México. |

Los expertos fueron localizados durante la búsqueda de antecedentes; por su participación en eventos científicos, en publicaciones internacionales, o tesis doctorales realizadas en los últimos diez años vinculados con la inferencia estadística. Se contactaron a través de sus correos electrónicos, por sus departamentos de trabajo, grupos de investigación, también en algún caso usamos listas de miembros de

Sociedades Internacionales. Son en total 11 los expertos que aceptaron colaborar; todos ellos cuentan con una amplia experiencia docente en estadística, específicamente en inferencia.

Algunos además cuentan con doctorados en didáctica de la matemática, en temas relacionados con el nuestro. Hemos recibido respuesta de investigadores de España, Colombia, México, Argentina y Chile; son matemáticos, estadísticos, profesores de matemáticas, todos ellos con doctorados o postgrados en Estadística. Uno de ellos es investigador en Psicometría, los demás en didáctica de la probabilidad y la estadística, incluyendo también a un Psicólogo. Todos cuentan con al menos 5 años de experiencia en investigación en el área, y edad comprendida entre los 31 y los 59 años.

### 3.7.2. MATERIAL Y MÉTODO

Una vez seleccionado el conjunto de ítems, siguiendo los lineamientos de Martínez Arias (1995), se preparó un cuestionario (Anexo 01) para ser completado por todo el grupo de expertos, y a través del cual se pudiese generar evidencias en la validación de contenido del cuestionario. Análogo a lo señalado por Díaz (2007), nuestro propósito ha sido doble, en sentido de obtener un consenso sobre la tabla de especificaciones para el instrumento, y también de sus opiniones sobre cómo cada unidad se ajusta en la evaluación de ese contenido determinado para el cual se lo ha elaborado.

Junto con el cuestionario (ver la Figura 3.4) se envió a los expertos una carta explicando los propósitos perseguidos en esta investigación, y pidiéndoles su colaboración. En el mismo cuestionario se incluía instrucciones sobre cómo completarlo. A continuación se describían los contenidos, seguido por los ítems, y qué se evaluaba con los distractores.

Finalmente para cada contenido e ítems, se agregó una tabla de datos que el juez debía rellenar, utilizando una escala de tipo Likert con cinco puntos. Los valores extremos toman los valores nada (1) y mucho (5). Se siguió el criterio de Wang y Osterlind, (2013) quienes sugieren usar una escala de valoración de 1 a 5, para pedir algo más que un acuerdo/desacuerdo a los jueces; además el mismo autor, agrega que usar una escala de puntuación mayor solo puede contribuir a complicar el proceso para llegar al acuerdo deseado.

Con el propósito de evitar contaminaciones, sesgos por persuasión, o por efectos de prestigio personal, nuestros ítems fueron valorados por los expertos con el esquema conocido como "panel ciego"; de esta forma los involucrados nunca tuvieron un encuentro cara a cara, tampoco conocen la identidad de los otros (Wang y Osterlind, 2013).

También con la finalidad de contar con otro criterio de validación de contenido, al finalizar su cuestionario se le proveyó de nuestra tabla de especificaciones de contenido, dejándoles la posibilidad que agregaran los que consideraban nos faltaban para completar el universo de contenido definido.

**CUESTIONARIO PARA RECOGIDA DE DATOS DE EXPERTOS**

A continuación presentamos una lista de contenidos que consideramos relevantes para evaluar la comprensión del tema análisis de varianza (pre-requisitos y aspectos básicos del análisis de varianza) en estudiantes de Psicología. Junto a cada contenido, presentamos una lista de ítems de evaluación en los cuáles presentamos con negrita la respuesta correcta.

Requerimos su colaboración para evaluar:

- El grado en que el contenido propuesto es relevante para la comprensión del Análisis de la Varianza (aspectos básicos y fundamentales).
- El grado en que cada ítem es adecuado para evaluar la comprensión del contenido específico que se propone.

Le agradecemos marque para cada uno de estos dos aspectos su opinión en la escala 1 a 5, donde 1 indica: nada relevante y 5 muy relevante. Agradecemos también cualquier sugerencia que ayude a mejorar la redacción del ítem

---

***PARTE I. PRERREQUISITOS***

**Contenido 1: Construcción de un intervalo de confianza**

Ítem 1a. Un psicólogo escolar desea estimar la puntuación media de la población en un test de rendimiento de lectura. Para ello administra el test a una muestra de 36 estudiantes, obteniendo una media de 48 y desviación típica 10. Calcule, al nivel de significación $\alpha$=0,05, los límites del intervalo de confianza para la puntuación en el test. (Error típico de la media: $S_d/\sqrt{n}$).

a. [38 , 58]
b. [15,3 , 80,7]
c. **[44,73 , 51,27]**

Ítem 1b. Se toma una muestra de 48 estudiantes con la finalidad de conocer su rendimiento en lectura. Un psicólogo escolar administra un test, así obtiene una media de 54,7 y desviación típica 5,23. Calcule, al nivel de significación α=0,01, los límites del intervalo de confianza para la puntuación en el test. (Error típico de la media: Sd/√n) son:

a. [53,95 ; 55,45]
b. [41,23 ; 68,17]
**c. [52,75 ; 56,65]**

**Nota:** Los distractores a y b evalúan confusión en uso de fórmulas y lectura de las tablas de la distribución.

| | 1: Nada | 2 | 3 | 4 | 5: Mucho |
|---|---|---|---|---|---|
| El contenido "**Construcción de un intervalo de confianza**" es relevante | | | | | |
| El ítem 1a es adecuado para este contenido | | | | | |
| El ítem 1b es adecuado para este contenido | | | | | |

Figura 3.4. Ejemplo del contenido del cuestionario de expertos (ver Anexo 03)

## 3.7.3. RESULTADOS Y DISCUSIÓN

En la Tabla 3.16, se presentan la media, mediana y desviación estándar de las puntuaciones que el grupo de expertos asignaron a cada uno de los contenidos de la tabla de especificaciones del cuestionario. Todos los contenidos que recibieron una alta valoración (por encima de 4 puntos). Algunos, como el ítem 1, 15 y 17, a pesar de tener un valor medio alto, han tenido una alta desviación estándar, ya que una pequeña proporción de jueces los han calificado con una puntuación baja, pero la mayoría tuvo una desviación típica adecuada. Por tanto, se decidió conservar todos los contenidos y

no agregar más, debido a que los jueces tampoco recomendaron otros contenidos diferentes a los presentados en el cuestionario.

Tabla 3.16. Frecuencia de asignación de acuerdo por los expertos (1-5) a cada contenido, media, mediana y desviación típica

| PRERREQUISITOS | 1 | 2 | 3 | 4 | 5 | $\bar{x}$ | Med | Sd |
|---|---|---|---|---|---|---|---|---|
| 1. Construcción de un intervalo de confianza | | 1 | 1 | 5 | 4 | 4,09 | 4 | 0,94 |
| 2. Interpretación del significado de un intervalo de confianza. | | | 3 | 4 | 4 | 4,09 | 4 | 0,83 |
| 3. Contraste bilateral y unilateral. Asignación de hipótesis | | | 1 | 3 | 7 | 4,56 | 5 | 0,68 |
| 4. Hipótesis nula y alternativa. Asignación de hipótesis | | | | 5 | 6 | 4,54 | 5 | 0,52 |
| 5. Errores tipo I y II, nivel de significación y potencia | | | 1 | 3 | 7 | 4,54 | 5 | 0,68 |
| 6. Cálculo de valor crítico, dado el nivel de significación | | 1 | 1 | 6 | 3 | 4 | 5 | 0,89 |
| 7. Regla de decisión en un contraste | | | | 2 | 9 | 4,81 | 5 | 0,40 |
| ANÁLISIS DE VARIANZA | | | | | | | | |
| 8. Selección de un modelo análisis de varianza, de acuerdo con las características de un problema | | | | 2 | 9 | 4,81 | 5 | 0,40 |
| 9. Supuestos de aplicación de un modelo análisis de varianza. | | | 1 | 1 | 9 | 4,72 | 5 | 0,64 |
| 10. Comprobación de supuestos en un modelo análisis de varianza | | | | 2 | 8 | 4,80 | 5 | 0,42 |
| 11. Modelo lineal asociado a un tipo de análisis de varianza | | | | 2 | 9 | 4,81 | 5 | 0,40 |
| 12. Cálculo en el análisis de varianza, Medidas repetidas | | | 1 | 3 | 7 | 4,54 | 5 | 0,68 |
| 13. Cálculo en el análisis de varianza. Modelo de dos factores | | | 2 | 2 | 7 | 4,45 | 5 | 8,82 |
| 14. Interpretación de una tabla análisis de varianza. Modelo de dos factores | | | | | 11 | 5 | 5 | 0 |
| 15. Interpretación de salida de software. Identificación del modelo de análisis de varianza | | 1 | 1 | 1 | 7 | 4,40 | 5 | 1,07 |
| 16. Interpretación de salida de software. Resultados | | | 1 | 1 | 9 | 4,72 | 5 | 0,64 |
| 17. Interpretación de salida de software. Supuestos del modelo análisis de varianza | | 1 | | 1 | 8 | 4,60 | 5 | 0,96 |

En la Tabla 3.17, se exhiben los resultados del grado de acuerdo que los expertos han valorado, entre cada uno de los ítems y su contenido principal. En estos resultados podemos apreciar, en general, ítems bien valorados, con puntuaciones medias de al menos 4 puntos. Los ítems que tienen menos valor o bien están ligeramente alrededor de ese promedio son: $5a_1$, $5b_1$, 6a, 6b, 7b, 10a, 15a, todos ellos presentan un valor alto de su desviación estándar, a pesar que la mayoría de las puntuaciones dadas por los jueces se concentran en valores altos. Para algunos de éstos ítems, como de destaca más adelante, varios expertos han propuesto cambios en parte de su redacción.

Tabla 3.17. Frecuencia de asignación de acuerdo (1-5) por los expertos a cada ítem.

| Ítem | 1 | 2 | 3 | 4 | 5 | $\bar{x}$ | Med | Sd |
|---|---|---|---|---|---|---|---|---|
| 1a | | 1 | 1 | 4 | 4 | 4,00 | 4 | 1 |
| 1b | | 1 | 1 | 4 | 5 | 4,18 | 4 | 0,98 |
| 2a | | | 3 | 4 | 4 | 4,09 | 4 | 0,83 |
| 2b | | | 2 | 4 | 4 | 4,20 | 4 | 0,78 |
| 3a | | 1 | | 2 | 8 | 4,54 | 5 | 0,93 |
| 3b | | 2 | 1 | 3 | 5 | 4,00 | 4 | 1,18 |
| 4a | | | 1 | 5 | 5 | 4,36 | 4 | 0,67 |
| 4b | | | | 4 | 6 | 4,60 | 5 | 0,51 |
| $5a_1$ | | 1 | 3 | 4 | 3 | 3,81 | 4 | 0,98 |
| $5b_1$ | 1 | 1 | 2 | 3 | 4 | 3,72 | 4 | 1,34 |
| $5b_2$ | | | 2 | 5 | 4 | 4,18 | 4 | 0,75 |
| $5a_2$ | | | 1 | 4 | 6 | 4,45 | 5 | 0,68 |
| 6a | | 2 | 2 | 5 | 2 | 3,63 | 4 | 1,02 |
| 6b | | 2 | 1 | 6 | 2 | 3,72 | 4 | 1,00 |
| 7a | | | 6 | 6 | 3 | 4,09 | 4 | 0,70 |
| 7b | 1 | | 2 | 4 | 4 | 3,90 | 4 | 1,22 |
| 8a | | | 1 | 2 | 8 | 4,63 | 5 | 0,67 |
| $8b_1$ | | 1 | | 4 | 6 | 4,36 | 5 | 0,92 |
| $8b_2$ | | 1 | 2 | 4 | 4 | 4,00 | 4 | 1 |
| 9a | | | 2 | 4 | 5 | 4,27 | 4 | 0,78 |
| 9b | | 1 | 2 | 3 | 5 | 4,09 | 4 | 1,04 |
| 10a | | 2 | 3 | 1 | 4 | 3,70 | 3,5 | 1,25 |
| $10b_1$ | | | 1 | 4 | 5 | 4,40 | 4,5 | 0,69 |
| $10b_2$ | | | 2 | 3 | 5 | 4,30 | 4,5 | 0,82 |
| 11a | | | 1 | 3 | 7 | 4,54 | 5 | 0,68 |
| 11b | | | 1 | 3 | 7 | 4,54 | 5 | 0,68 |
| 12a | | | 2 | 6 | 3 | 4,09 | 4 | 0,70 |
| 12b | | | 2 | 5 | 4 | 4,18 | 4 | 0,75 |
| $13a_1$ | | | 2 | 6 | 3 | 4,09 | 4 | 0,70 |
| $13a_2$ | | | 2 | 6 | 3 | 4,09 | 4 | 0,70 |
| 13b | | | 3 | 4 | 4 | 4,09 | 4 | 0,83 |
| 14a | | 1 | | 5 | 5 | 4,27 | 4 | 0,90 |
| 14b | | 1 | 1 | 4 | 5 | 4,18 | 4 | 0,98 |
| 15a | 1 | | 3 | 1 | 5 | 3,90 | 4,5 | 1,37 |
| 15b | | 1 | 2 | 3 | 4 | 4,00 | 4 | 1,05 |
| $16a_1$ | | | 2 | 3 | 6 | 4,36 | 5 | 0,80 |
| $16a_2$ | | | | 4 | 7 | 4,63 | 5 | 0,50 |
| 16b | 1 | | | 3 | 7 | 4,36 | 5 | 1,20 |
| 17a | | 1 | 1 | 3 | 5 | 4,20 | 4,5 | 1,03 |
| $17b_1$ | | 1 | 1 | 2 | 5 | 4,20 | 4,5 | 1,03 |
| $17b_2$ | | 1 | 1 | 2 | 5 | 4,22 | 5 | 1,09 |

## 3.8.SELECCIÓN FINAL Y REVISIÓN DE ÍTEMS PARA EL CUESTIONARIO

A continuación describiremos los resultados en la selección de los ítems que formarían parte del cuestionario. La posibilidad de contar con al menos dos ítems por contenido, nos permitió elegir el mejor valorado por los jueces y que también tuviera una buena ejecución en las pruebas piloto.

Empleando una metodología similar a Díaz (2007) y Olivo (2008) estudiamos, en cada contenido y para cada ítem los resúmenes obtenidos del análisis de los cuestionarios dados a los expertos y los índices de dificultad del ítem. El criterio fue seleccionar aquellos ítems que tuviesen la mayor media ($\bar{x}$) posible en la puntuación que dieron los expertos, y la menor desviación estándar ($s$). Hemos descartado aquellos cuya valoración haya alcanzado valores inferiores a 4 para la media, pues estos indican una

valoración poco positiva del ítem o del contenido (recordemos que la escala ha sido de 1-5). Además, una desviación típica pequeña indica que hay un grado de acuerdo alto, ya que hay poca dispersión en las puntuaciones.

Por otro lado, entre dos ítems valorados de igual modo por los expertos, en la mayoría de los casos se tomó el de dificultad más moderada, pues los valores del índice de dificultad más cercanos al 0,5 proporcionan mejor discriminación. Se trató asimismo de incluir ítems con una amplia gama de dificultad, incluyendo índices de alta y baja dificultad. De este modo podríamos dar una valoración más equitativa a todos los alumnos, en función de sus conocimientos y, al mismo tiempo conseguir una distribución aproximadamente normal de los índices de dificultad.

**Contenido 1: Construcción de un intervalo de confianza**

Tabla 3.18. Resultados en la valoración de expertos en el Contenido 1

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | $\bar{x}$ | Sd | | | |
| | | 1 | 1 | 5 | 4 | 4,09 | 0,94 | | | |
| Ítem 1a | | 1 | 1 | 4 | 4 | 4,00 | 1,00 | 0,77 | 93 | Equilibrio en distractores |
| Ítem 1b | | 1 | 1 | 4 | 5 | 4,18 | 0,98 | 0,70 | 130 | La opción b, no fue escogida por ningún alumno |

Este contenido, tal como se indica en la Tabla 3.18, ha sido bien valorado por los expertos, ya que nueve de los once expertos consultados lo han valorado con 4 y 5 puntos, logrando de este modo una media, $\bar{x} = 4{,}09$. Aunque dos expertos lo han valorado con 2 y 3 puntos respectivamente, la desviación típica es razonable $s = 0{,}94$. Los dos ítems tienen muy buena discriminación (Tabla 3.13 y 3.14). De entre los dos ítems que comprenden a este contenido (Tabla 3.18) tomamos el 1b, porque resulta un poco más alta la puntación de jueces y discrimina mejor (Tabla 3.14), a pesar que uno de los distractores no es elegido. Se decide tomar el ítem y revisar este distractor.

**Contenido 2: Interpretación del significado de un intervalo de confianza.**

Tabla 3.19. Resultados en la valoración de expertos al Contenido 2

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | DT. | | | |
| | | | 3 | 4 | 4 | 4,09 | 0,83 | | | |
| Ítem 2a | | | 3 | 4 | 4 | 4,09 | 0,83 | 0,11 | 93 | Casi el 50% no lo contesta |
| Ítem 2b | | | 2 | 4 | 4 | 4,20 | 0,78 | 0,22 | 130 | Desequilibrio en distractores |

Este contenido también ha sido bien valorado por el grupo de expertos, con media, $\bar{x} = 4{,}09$ y desviación estándar $s = 0{,}83$ (ver Tabla 3.16), aunque tres expertos lo evaluaron con 3 puntos, indicando que no es un tema estrictamente relacionado con el análisis de la varianza. Mantenemos este contenido, para comprobar que los alumnos no caen en errores descritos en la literatura para la interpretación de los intervalos, ya que

podría influir posteriormente en la comprensión de cálculos que se hacen en Análisis de Varianza, por ejemplo, estimar el tamaño de un efecto.

De los dos ítems presentados para este contenido (Tabla 3.19), hemos decidido tomar el 2b, que posee mejor puntuación de jueces y resulta algo más fácil que el 2a; también admite una buena discriminación (Tabla 3.14) y mejor equilibrio entre los distractores. Además el contenido del mismo se da en términos frecuenciales, que según Gigerenzer y Hoffrage (1995) facilitan la resolución de problemas estadísticos, en comparación con otros problemas en los que los datos se dan mediante probabilidades.

**Contenido 3: Contraste bilateral y unilateral. Asignación de hipótesis.**

Observando la Tabla 3.16, también nos encontramos que este contenido ha sido muy bien valorado por el elenco de jueces ($\bar{x} = 4{,}54, s = 0{,}68$), ya que un 64% le otorgaron la máxima calificación (5), lo cual da idea de una mayor concentración de valores alrededor de la media.

Tabla 3.20. Resultados en la valoración de expertos al contenido 3

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 1 | 3 | 7 | 4,54 | 0,68 | | | |
| Ítem 3a | | 1 | | 2 | 8 | 4,54 | 0,93 | 0,92 | 93 | Equilibrio de |
| Ítem 3b | | 2 | 1 | 3 | 5 | 4,00 | 1,18 | 0,95 | 130 | distractores |

De los dos ítem que se corresponden a este contenido, (Tabla 3.20) hemos decidido tomar el 3a por estar mejor puntuado por los expertos consultados, siendo el nivel de dificultad planteado prácticamente el mismo para ambos, coincidiendo con un equilibrio entre los distractores. Además discrimina correctamente (Tabla 3.13).

**Contenido 4: Hipótesis nula y alternativa. Asignación de hipótesis.**

Las medidas de posición y dispersión resultaron: $\bar{x} = 4{,}54;\ s = 0{,}52$ , para este contenido ya que el 100% de los jueces lo puntuaron con valores 4 y 5 (ver Tabla 3.16), lo que indica una muy buena valoración y alto grado de acuerdo.

Tabla 3.21. Resultados en la valoración de expertos al Contenido 4

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | 5 | 6 | 4,54 | 0,52 | | | |
| Ítem 4a | | | 1 | 5 | 5 | 4,36 | 0,67 | 0,94 | 93 | Equilibrio en |
| Ítem 4b | | | | 4 | 6 | 4,60 | 0,51 | 0,93 | 130 | distractores |

Escogimos para este contenido el ítem 4b por estar mejor puntuado por los jueces, tener buena discriminación (Tabla 3.14) y presentar menor variabilidad (mayor grado de acuerdo); además se descubre un equilibrio para la elección de los distractores y resulta algo más difícil (Tabla 3.21).

**Contenido 5: Errores Tipo I y II, nivel de significación y potencia**

También aquí el 64% de jueces puntúa este contenido con la mayor calificación (5), y un solo experto le asigna un puntaje medio (3). Resulta una media $\bar{x} = 4{,}54$ y una desviación estándar, $s = 0{,}68$ (ver Tabla 3.16), estando muy bien valorado por todo el grupo de expertos en general. De este contenido, puesto que tenemos 4 ítems vamos a tomar dos.

Tabla 3.22. Resultados en la valoración de expertos al Contenido 5

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 1 | 3 | 7 | 4,54 | 0,68 | | | |
| Ítem $5a_1$ | | 1 | 3 | 4 | 3 | 3,81 | 0,98 | 0,75 | 93 | Equilibrio en distractores |
| Ítem $5b_1$ | 1 | 1 | 2 | 3 | 4 | 3,78 | 1,34 | 0,38 | 130 | Desequilibrio en |
| Ítem $5b_2$ | | | 2 | 5 | 4 | 4,18 | 0,75 | 0,51 | 130 | distractores. |
| Ítem $5a_2$ | | | 1 | 4 | 6 | 4,45 | 0,68 | 0,65 | 93 | |

Hemos decidido escoger el $5b_2$ y $5a_2$, sus puntuaciones están mejor concentradas alrededor de la media, discriminan mejor (Tabla 3.13 y 3.14) y los otros tienen puntuación de los jueces más baja, aunque siempre cercana al 4 (Tabla 3.22). Además estos ítems resultan tener una dificultad moderada, aunque es necesario controlar los distractores en una revisión posterior.

**Contenido 6: Cálculo de valor crítico, dado el nivel de significación**

Este contenido presenta una puntuación media, $\bar{x} = 4$ y una desviación estándar, $s = 0{,}89$; también indica buena puntuación de los jueces aunque con una mayor variabilidad (ver Tabla 3.16). En cuanto a los dos ítems que corresponden a este contenido, hemos decidido tomar el 6a, puesto que discrimina mejor (Tabla 3.13), presenta una dificultad moderada; pero deberemos estudiar algo más los distractores ya que hay un desequilibrio (Tabla 3.23). Se revisará la redacción del ítem para aumentar la discriminación.

Tabla 3.22. Resultados en la valoración de expertos al Contenido 6

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | 1 | 1 | 6 | 3 | 4,00 | 0,89 | | | |
| Ítem 6a | | 2 | 2 | 5 | 2 | 3,63 | 1,02 | 0,56 | 93 | Desequilibrio en distractores. |
| Ítem 6b | | 2 | 1 | 6 | 2 | 3,72 | 1,00 | 0,91 | 130 | Equilibrio en distractores. |

**Contenido 7: Regla de decisión en un contraste**

Este contenido ha sido altamente valorado por los expertos, como se observa en la Tabla 3.16, pues presenta una media $\bar{x} = 4{,}81$ , muy cerca del máximo de la escala y una $s = 0{,}40$, y una muy baja variabilidad (solo dos expertos lo votaron con 4 puntos). Hemos escogido el Ítem 7a que está mejor puntuado por los expertos consultados, y además presenta una dificultad moderada (Tabla 3.24), aunque no discrimina tan bien

como el 7b. No obstante se deben observar los distractores pues presentan desequilibrio y por ello revisaremos la redacción.

Tabla 3.24. Resultados en la valoración de expertos al Contenido 7

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | 2 | 9 | 4,81 | 0,40 | | | |
| Ítem 7a | | | 2 | 6 | 3 | 4,09 | 0,70 | 0,59 | 130 | Desequilibrio en |
| Ítem 7b | 1 | | 2 | 4 | 4 | 3,90 | 1,22 | 0,68 | 130 | distractores. |

**Contenido 8: Selección de un modelo Análisis de Varianza, de acuerdo con las características de un problema**

Este contenido ha sido igualmente valorado que el último, $\bar{x} = 4{,}81$ para la media y $s = 0{,}40$ para la desviación estándar, teniendo igual distribución en las puntuaciones dadas por los jueces (Tabla 3.16). Los tres ítems que conforman este contenido han sido muy bien puntuados por los expertos consultados (Tabla 3.25), pero decidimos tomar el $8b_1$ y el $8b_2$ (dificultad moderada y discriminación aceptable en $b_1$ y muy buena para el $b_2$) (Tabla 3.14) y mayor tamaño de muestra.

Tabla 3.25. Resultados en la valoración de expertos al Contenido 8

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | 2 | 9 | 4,81 | 0,40 | | | |
| Ítem 8a | | | 1 | 2 | 8 | 4,63 | 0,67 | 0,82 | 93 | Equilibrio en distractores. |
| Ítem $8b_1$ | | 1 | | 4 | 6 | 4,36 | 0,78 | 0,72 | 130 | |
| Ítem $8b_2$ | | 1 | 2 | 4 | 4 | 4,00 | 1,00 | 0,78 | 130 | Equilibrio en distractores. |

**Contenido 9: Supuestos de aplicación de un modelo Análisis de Varianza.**

Casi el 82% del total de jueces le han dado el más alto puntaje de la escala a este contenido, lo cual proporciona una medida de tendencia central y dispersión respectivamente de $\bar{x} = 4{,}72\ y\ s = 0{,}64$ (Tabla 3.16). A pesar de presentar muy parejos su índice de dificultad, el 9a está mejor puntuado por el grupo de expertos, con mayor media y menor desviación estándar (Tabla 3.26), mejor discriminación (Tabla 3.13) y equilibrio en distractores, de donde nos hemos decidido por él.

Tabla 3.26. Resultados en la valoración de expertos al Contenido 9

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 1 | 1 | 9 | 4,72 | 0,64 | | | |
| Ítem 9a | | | 2 | 4 | 5 | 4,27 | 0,78 | 0,52 | 93 | |
| Ítem 9b | | 1 | 2 | 3 | 5 | 4,09 | 1,04 | 0,50 | 130 | Equilibrio en distractores. |

**Contenido 10: Supuestos de aplicación de un modelo Análisis de Varianza.**

A pesar que no ha sido puntuado por uno de los expertos, presenta medidas altas, $\bar{x} = 4{,}80$ y $s = 0{,}42$ (ver Tabla 3.16), esto lo transforma en un contenido altamente valorado. Para este grupo de ítems que conforman este contenido el $10b_2$ además de estar bien puntuado por los jueces, presenta una dificultad de media a moderada al compararse con los otros dos (Tabla 3.27) y discrimina bien. Esto nos decide por este ítem, pero debemos observar los distractores.

Tabla 3.27 Resultados en la valoración de expertos al Contenido 10

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | 2 | 8 | 4, 80 | 0,42 | | | |
| Ítem 10a | | 2 | 3 | 1 | 4 | 3,70 | 1,25 | 0,77 | 93 | Equilibrio en distractores. |
| Ítem $10b_1$ | | | 1 | 4 | 5 | 4,40 | 0,69 | 0,82 | 130 | Equilibrio en distractores. |
| Ítem $10b_2$ | | | 2 | 3 | 5 | 4,30 | 0,82 | 0,58 | 130 | |

**Contenido 11: Modelo lineal asociado a un tipo de Análisis de Varianza**

Contenido muy bien puntuado por el grupo de expertos, con media y desviación estándar respectivamente de $\bar{x} = 4{,}81$ y $s = 0{,}40$. Estos dos ítems para este contenido, han sido igual puntuados por el grupo de expertos (Tabla 3.28), pero decidimos tomar la unidad de medida 11b porque el enunciado del 11a es muy abstracto, además resulta más fácil; el 11b discrimina mejor, presenta equilibrio en distractores y es mayor el tamaño de la muestra.

Tabla 3.28. Resultados en la valoración de expertos al Contenido 11

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | 2 | 9 | 4,81 | 0,40 | | | |
| Ítem 11a | | | 1 | 3 | 7 | 4,54 | 0,68 | 0,53 | 93 | |
| Ítem 11b | | | 1 | 3 | 7 | 4,54 | 0,68 | 0,83 | 130 | Equilibrio entre distractores. |

**Contenido 12: Cálculo en el Análisis de Varianza, medidas repetidas**

La puntuación de este contenido es algo más moderada, pero sin bajar los cuatro puntos de media. Se observa algo más de variabilidad, ya que hubo un experto que lo ha puntuado con tres puntos. La media ha sido $\bar{x} = 4{,}54$ y la desviación estándar $s = 0{,}68$ (ver la Tabla 3.16). Escogemos el Ítem 12b, argumentando los mismos motivos que se tuvieron en cuenta para escoger entre los ítems del contenido anterior, es decir mejor puntuación dada por el grupo de expertos, discrimina mejor y más fácil. Finalmente, además presenta equilibrio entre los distractores y mayor tamaño de muestra (Tabla 3.29).

Tabla 3.29. Resultados en la valoración de expertos al Contenido 12

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 1 | 3 | 7 | 4,54 | 0,68 | | | |
| Ítem 12a | | | 2 | 6 | 3 | 4,09 | 0,70 | 0,08 | 93 | |
| Ítem 12b | | | 2 | 5 | 4 | 4,18 | 0,75 | 0,77 | 130 | Equilibrio entre los distractores. |

**Contenido 13: Cálculo de Análisis de Varianza. Modelo de dos factores**

Muy buena valoración de jueces, a través de su puntuación media, pero algo dispersos los valores, se puede observar en su puntaje medio $\bar{x} = 4{,}45$ ; y el de su desviación estándar, $s = 0{,}82$ en la Tabla 3.16.

Tabla 3.30. Resultados en la valoración de expertos al Contenido 13

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 2 | 2 | 7 | 4,45 | 0,82 | | | |
| Ítem $13a_1$ | | | 2 | 6 | 3 | 4,09 | 0,70 | 0,64 | 47 | Ningún alumno |
| Ítem $13a_2$ | | | 2 | 6 | 3 | 4,09 | 0,70 | 0,74 | 46 | elije la opción a. |
| Ítem 13b | | | 3 | 4 | 4 | 4,09 | 0,83 | 0,70 | 130 | Equilibrio entre los distractores. |

Decidimos tomar los ítems $13a_1$, y $13a_2$ dado que los dos tienen igual puntaje medio, desviación estándar e índice de dificultad. Ambas unidades discriminan correctamente. El 13b, a pesar de tener equilibrados sus distractores, presenta mayor dispersión y es algo más fácil (Tabla 3.30). Ambos ítems elegidos se corresponden con el mismo enunciado, evitando al alumno complicaciones con cuentas.

**Contenido 14: Interpretación de una tabla Análisis de Varianza. Modelo de dos factores**

Fue uno de los contenidos mejores puntuados, con 5 puntos de media y 0 desviación estándar (ver Tabla 3.16), lo cual significa que todos los expertos lo han puntuado con el máximo puntaje. Hemos escogido el ítem 14a (Tabla 3.31) porque se refiere a la misma tabla de datos que el $13a_2$ y se evita a los alumnos realizar dos veces un cálculo análogo, además para tener algún ítem con mayor dificultad, de este modo se abre el abanico posible de dificultades presentes en la prueba. Este ítem discrimina bien.

Tabla 3.31. Resultados en la valoración de expertos al Contenido 14

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | | | 11 | 5,00 | 0,00 | | | |
| Ítem 14a | | 1 | | 5 | 5 | 4,27 | 0,90 | 0,26 | 93 | Equilibrio entre |
| Ítem 14b | | 1 | 1 | 4 | 5 | 4,18 | 0,98 | 0,43 | 130 | distractores. |

**Contenido 15: Interpretación de salida de software. Identificación del modelo de Análisis de Varianza**

Tabla 3.32. Resultados en la valoración de expertos al Contenido15

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | 1 | 1 | 1 | 7 | 4,40 | 1,07 | | | |
| Ítem 15a | 1 | | 3 | 1 | 5 | 3,90 | 1,37 | 0,70 | 93 | Equilibrio distractores |
| Ítem 15b | | 1 | 2 | 3 | 4 | 4,00 | 1,05 | 0,82 | 130 | Desequilibrio entre distractores |

Está algo disperso el puntaje dado por el grupo de expertos a este contenido ($s = 1{,}07$) a pesar de estar muy bien puntuado, ya que su valor medio es $\bar{x} = 4{,}40$ (ver Tabla 3.16) y casi el 64% de los jueces le dan el máximo puntaje de la escala, 5. Aquí decidimos escoger el ítem 15b, ya que resulto mucho más fácil, y el otro ítem que disponemos está bajo puntuado por el grupo de expertos (Tabla 3.32). Tiene una discriminación aceptable y es mayor el tamaño de muestra, pero debemos revisar la redacción por el desequilibrio entre los distractores.

**Contenido 16: Interpretación de salida de software. Resultados**

Para este contenido han dado una alta puntuación los jueces, resultando un valor medio $\bar{x} = 4{,}72$ y una deviación estándar, $s = 0{,}64$ (ver la Tabla 3.16). De los tres ítems que componen este contenido, decidimos escoger la unidad 16b, que está muy bien puntuado por los jueces. Aunque es algo más difícil, quisimos incluir un ítem que discriminase bien los alumnos con alto conocimiento (Tabla 3.14).

Tabla 3.33. Resultados en la valoración de expertos al Contenido 16

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | | 1 | 1 | 9 | 4,72 | 0,64 | | | |
| Ítem $16a_1$ | | | 2 | 3 | 6 | 4,36 | 0,80 | 0,85 | 93 | Distractores equilibrados. |
| Ítem $16a_2$ | | | | 4 | 7 | 4,63 | 0,50 | 0,41 | 93 | 25% no lo contesta |
| Ítem 16b | 1 | | | 3 | 7 | 4,36 | 1,20 | 0,31 | 130 | 32% no lo contesta. |

**Contenido 17: Interpretación de salida de software. Supuestos del modelo Análisis de Varianza**

Si bien, el puntaje medio otorgado a este contenido resulta alto ($\bar{x} = 4{,}60$), la valoración de jueces es algo dispersa, generando una desviación estándar algo alta ($s = 0{,}96$) (ver Tabla 3.16). Aquí, para este grupo de ítems, observando la Tabla 3.34 decidimos tomar el $17b_1$, pues discrimina muy bien (Tabla 3.14) y los otros dos presentan desequilibrio en los distractores.

Tabla 3.34. Resultados en la valoración de expertos al Contenido 17

| | Valoración expertos | | | | | | | Índice Dificultad | N | Observaciones |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | Media | Desv.Est. | | | |
| | | 1 | | 1 | 8 | 4,60 | 0,96 | | | |
| Ítem 17a | | 1 | 1 | 3 | 5 | 4,20 | 1,03 | 0,62 | 93 | Desequilibrio distractores. |
| Ítem $17b_1$ | | 1 | 1 | 3 | 5 | 4,20 | 1,03 | 0,62 | 130 | |
| Ítem $17b_2$ | | 1 | 1 | 2 | 5 | 4,22 | 1,09 | 0,72 | 130 | Equilibrio distractores. |

## 3.9. REVISIÓN DEL CUESTIONARIO EN FUNCIÓN DEL JUICIO DE EXPERTOS

Una vez elegidos los ítems que formarían el cuestionario, siguiendo el procedimiento anteriormente descripto, se procedió a revisar los ítems elegidos en función de los comentarios y sugerencias de mejora realizadas por los expertos, quienes habían incluido comentarios muy detallados en algunos de los ítems. A continuación se describe el proceso, mostrando en cada caso, en primer lugar el ítem seleccionado inicialmente, a continuación los comentarios de los expertos sobre el mismo, y finalmente la redacción definitiva.

### Contenido 1: Construcción de un intervalo de confianza

El ítem finalmente elegido fue el 1b, respecto al cual algunos expertos señalaron que sería más correcto utilizar el término "nivel de confianza", en lugar de "nivel de significación", para evitar confusión con el contraste de hipótesis:

> *La mayoría de los libros al hallar intervalos de confianza hablan de intervalo de confianza con nivel de confianza del 95% y no con nivel de significación del 5%.* (Experto 1).

También dos expertos agregaban al respecto, que aunque hubiese relación entre ambos conceptos, el nivel de confianza está más asociado al contraste de hipótesis, y lo hizo con las siguientes palabras:

> *Aunque hay una relación estrecha entre intervalos de confianza y contraste de hipótesis, generalmente el término "nivel de significación" está más asociado con contraste de hipótesis y en su lugar el término "nivel de confianza (1-α)" para estimación por medio de intervalos de confianza* (Experto 9).
> *Usar la expresión nivel de confianza de 95% (99%). En la definición de intervalo de confianza, el objeto que aparece es nivel de confianza, no nivel de significación* (Experto 2).

En otro orden de cosas, advertimos que no especificábamos que se trataba de un intervalo para la puntuación media poblacional, dejándolo como obvio, utilizando en la pregunta simplemente que se calculase un intervalo de confianza para la puntuación del test administrado. Por ello atendimos a otra sugerencia de otro experto, quien indica que esta indefinición puede causar dificultad al estudiante:

> *Frecuentemente los alumnos no tienen idea acerca de la variable que se está midiendo y en consecuencia dudan acerca de que media se está obteniendo. Por otro lado conviene resaltar el parámetro que se está estimando* (Experto 2).

Finalizando ya con las modificaciones realizadas, atendimos a una propuesta para que la fórmula de la desviación estándar de la media (incluida en el enunciado)

apareciese en letra cursiva y al final del ítem (experto 11). Con todo ello, la redacción final del ítem queda modificada del siguiente modo:

**Ítem 1b.** Se toma una muestra de 48 estudiantes con la finalidad de conocer su rendimiento en lectura. Un psicólogo escolar administra un test, así obtiene una media de 54,7 y desviación típica 5,23. Los límites del intervalo de confianza para la puntuación media poblacional del test administrado, con un nivel de confianza α=0,99 son:

a. [53,95 ; 55,45]
b. [41,23 ; 66,17]
**c. [52,75 ; 56,65]**

(Error típico de la media: $S_d/\sqrt{n}$ )

**Contenido 2: Interpretación del significado de un intervalo de confianza.**

El ítem elegido ha sido el 2b, al cual también decidimos aplicar cambios en su redacción. Varios expertos nos aconsejaban refrendar sobre qué parámetro se calculaban los 1000 intervalos de confianza (experto 1, 2, 3, 5 y 9), agregamos como aposición la frase "media poblacional", para atender a ese pedido. En el distractor a, también hemos realizado cambios en la redacción, ya que dos de los expertos señalaron:

> *Moderar el determinismo cuando se dice: "Sólo 990 intervalos,…" con alguna expresión que dé cuenta que se trata de un valor promedio a largo plazo* (Experto 1).
> *La opción a., la cual es la correcta no me parece apropiada ya que la palabra sólo debería sustituirse por aproximadamente ya que de 1000 intervalos al nivel de confianza del 99%, es posible cualquier valor en torno a 990* (Experto 4).

Nos pareció adecuado además, que la letra griega $\mu$ apareciese siempre con igual formato. Finalmente, el ítem modificado nos quedo de la siguiente forma:

**Ítem 2b.** Un investigador ha calculado 1000 intervalos de confianza al 99% para la media poblacional $\mu$. Suponiendo que los datos en los que se basan los intervalos se seleccionan de manera independiente entre si. ¿Cuántos de estos 1000 intervalos se espera que contengan el valor verdadero de $\mu$?

a. **Aproximadamente 990 intervalos contendrán el valor de $\boldsymbol{\mu}$**
b. Los 1000 intervalos contendrán el valor de $\mu$
c. No se puede saber a priori la cantidad de intervalos se espera que capturen el verdadero valor del parámetro $\mu$.

**Contenido 3: Contraste bilateral y unilateral. Asignación de hipótesis.**

El ítem elegido ha sido el 3a, cuyo enunciado original reproducimos a continuación. Éste ítem les ha parecido a todos los Expertos consultados y a nosotros mismos apropiado, sin necesidad de cambio tampoco para su redacción. Por esta razón conserva su formato y redacción original.

**Ítem 3a**. Queremos conocer si los sujetos extrovertidos e introvertidos difieren en la puntuación media en autoestima y no disponemos de ninguna información previa. El tipo de hipótesis nula razonable que debo plantear es:

a. $H_0: \mu_I \leq \mu_E$
b. $\boldsymbol{H_0: \mu_I = \mu_E}$
c. $H_0: \mu_I \geq \mu_E$

**Contenido 4: Hipótesis nula y alternativa. Asignación de hipótesis.**

El ítem elegido ha sido el 4b, cuyo enunciado original reproducimos a continuación. Debido a razones análogas dadas para el último ítem anterior este tampoco ha sufrido modificaciones de naturaleza alguna.

**Ítem 4b**. De los siguientes pares de afirmaciones, indique cual NO cumple con las reglas para plantear hipótesis estadísticas:
a. $\boldsymbol{H_0: \sigma = 15;\ H_a: \sigma \leq 15}$
b. $H_0: \mu \leq 100;\ H_a: \mu > 100$
c. $H_0: \mu_1 - \mu_2 = 0;\ H_a: \mu_1 - \mu_2 \neq 0$

**Contenido 5: Errores Tipo I y II, nivel de significación y potencia.**

Para esta unidad de contenido se han elegido dos ítems, el $5b_2$ y el $5a_2$, cuyo enunciado original reproducimos a continuación. Estos últimos dos ítems, por consenso general de expertos, y también por el nuestro propio se mantienen sin modificaciones

**Ítem $5b_2$.** Si en una investigación lees que la potencia de un contraste vale 0,5594, entonces interpretas que:
a. La probabilidad de rechazar la hipótesis nula siendo falsa vale (1 -0.5594) = 0,4406
b. **La probabilidad de rechazar la hipótesis nula siendo falsa vale 0,5594**
c. La probabilidad de mantener la hipótesis nula siendo falsa vale 0,5594

**Ítem $5a_2$.** Supongamos que conocemos la „verdad absoluta" sobre la eficacia de dos tratamientos (A y B), y sabemos que existen diferencias en la efectividad de ambos para curar la depresión. Un investigador que realice un estudio y parta de la hipótesis *„no existen diferencias en la efectividad de los tratamientos A y B para curar la depresión"* cometerá un *error tipo II* cuando:
a. Concluya que A y B no son efectivos para curar la depresión
b. **Concluya que A y B no difieren en su efectividad para curar la depresión**
c. Concluya que A y B difieren en su efectividad para curar la depresión

**Contenido 6: Cálculo de valor crítico, dado el nivel de significación.**

Se ha elegido el ítem 6a en el que también hemos decidido introducir modificaciones, en atención a la opinión de los expertos consultados y al desequilibrio en sus distractores. Cambiamos la frase "puntuación típica", ya que al respecto se recogió la siguiente opinión:

> *El término puntuación típica está más relacionado con la tipificación de una variable. Creo que debiera usarse el término "valor crítico". De hecho, es el término que se usa para definir globalmente a estos dos ítems* (Experto1).

Además el Experto 2, nos aconseja especificar que el valor crítico se corresponde con una distribución normal, para que no se confunda con cualquier otra. Finalmente la redacción del ítem es:

**Ítem 6a.** El valor crítico correspondiente a una $\alpha = 0{,}01$ en un contraste unilateral derecho, suponiendo normalidad de los datos es:
a. **2,33**
b. -2,33
c. 3,10

**Contenido 7: Regla de decisión en un contraste.**

Se ha elegido el ítem 7a para este contenido. Atendiendo la opinión de dos expertos, que se expresaban del siguiente modo hemos decidido cambiarlo por el enunciado que sigue inmediatamente más abajo:

> *Me parece que "valor de significación" y "nivel de significación" se prestan a confusión. Creo que en la literatura científica en general, a la primera expresión se le conoce con el nombre poco creativo de "p-value" (p valor o valor P), inclusive esta letra "P" aparece en la mayoría de las salidas de Software estadístico* (Experto 9).
> *La frase: la probabilidad asociada al estadístico de contraste (el valor de significación) sea menor que el valor de alfa no me resulta familiar. Particularmente preferiría la frase: La probabilidad de observar un valor del estadístico tan extremo o más que el observado. Otra alternativa es decir el p-valor de la prueba. Una breve consulta en la Web parece confirmar que efectivamente la frase: "la probabilidad asociada al estadístico de contraste", es utilizada en psicología.* (Experto 1)

**Ítem 7a**. Cuando realizamos un contraste, la regla de decisión nos lleva a rechazar la hipótesis nula siempre que:

a. El estadístico de contraste caiga en la región de rechazo
b. La probabilidad asociada al estadístico de contraste (el nivel de significación) sea menor que el valor de alfa
c. **a) y b) son correctas**

**PARTE II. Análisis de varianza**

**Contenido 8: Selección de un modelo Análisis de varianza, de acuerdo con las características de un problema.**

Se han elegido el ítem $8b_1$ $8b_2$ para este contenido. El Experto 11, nos sugiere cambiar la pregunta, refiriéndola directamente a la técnica estadística que aplicaría la maestra frente al problema en el $8b_1$, y no como lo planteamos nosotros, haciendo referencia a un investigador. Asimismo otro experto, aconseja referirnos directamente a Análisis de varianza de un factor, en vez de escribir en la opción (c), Análisis de varianza de un factor completamente aleatorizado. Sin embargo, se decidió dejar el ítem con nuestra redacción, pues los alumnos han estudiado ese contenido usando este término:

**Ítem $8b_1$.** Para mejorar la psicomotricidad de los niños de primaria, una maestra cree que ayudarán unas nuevas actividades físicas. La maestra divide aleatoriamente su grupo de trabajo en tres partes iguales. A cada grupo le aplica un tipo de ejercicio diferente, pues desea saber qué tipo de ejercicios le dará mejores resultados. De las técnicas estadísticas que siguen, ¿cuál debería aplicar la maestra para comprobar si los métodos que aplica son diferentes?

a. Contraste de hipótesis t sobre medias independientes
b. Contraste de hipótesis t sobre medias relacionadas
c. **Análisis de varianza de un factor completamente aleatorizado**

El ítem $8b_2$ ha tenido ningún cambio respecto a su redacción original, no ha sido comentado por los expertos, de donde la siguiente es la redacción definitiva.

**Ítem $8b_2$**. Un investigador utilizará un Análisis de varianza de dos factores, efectos fijos y completamente aleatorizado
cuando:

a. En el estudio haya dos variables dependientes.
b. En el estudio haya una variable independiente, con dos niveles seleccionados al azar.
c. **En el estudio haya dos variables independientes, cada una con dos o más niveles**

**Contenido 9: Supuestos de aplicación de un modelo Análisis de varianza.**

En este contenido el ítem ha quedado sin modificación respeto de la original, pues no ha habido sugerencias por parte de los expertos.

**Ítem 9a.** Los supuestos del Análisis de varianza de dos factores, efectos fijos y completamente aleatorizado son:
a. Independencia de las observaciones, normalidad de las distribuciones, y aditividad
b. Independencia de las observaciones, igualdad de varianza y aditividad
c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas**

**Contenido 10: Comprobación de supuestos en un modelo Análisis de varianza**

En esta unidad el ítem ha quedado sin modificación respeto de la original, pues no ha habido sugerencias por parte de los expertos, además lo consideramos en su mismo redacción a pesar del desequilibrio en sus distractores.

**Ítem 10b$_2$.** La hipótesis nula que somete a contraste la prueba de Levene es la siguiente:
a. **Ho: $\sigma^2_A = \sigma^2_B$**
b. Ho: $\sigma^2_A \neq \sigma^2_B$
c. Ho: $\mu_{SI} = \mu_{NO}$.

**Contenido 11: Modelo lineal asociado a un tipo de Análisis de varianza**

Tampoco el ítem elegido tiene comentarios de jueces, lo consideramos con su redacción original que se presenta inmediatamente debajo.

**Ítem 11b**. El Análisis de varianza de dos factores, con efectos fijos, descompone la variabilidad total en los siguientes
componentes:
a. Variabilidad total = V. entre grupos + V. error
b. Variabilidad total = V. entre grupos + V. entre sujetos + V. error
c. **Variabilidad total = V. factor A + V. factor B + V. interacción + V. error**

**Contenido 12: Cálculo en el Análisis de varianza, medidas repetidas**

En el ítem que hemos atendido al siguiente comentario:

> *Debería uniformizarse la notación. Para referirse a la media cuadrática aquí se utiliza CM mientras que en las tablas posteriores aparece MC* (Experto 1).

**Ítem 12b.** Si en un Análisis de varianza de un factor, y medidas repetidas encuentro que la F empírica u observada toma
un valor de 8,16 esto quiere decir que:
a. **CM entre grupos / CM error = 8,16**
b. CM entre sujetos / CM error = 8,16
c. CM entre grupos / CM intra sujetos = 8,16

**Contenido 13: Cálculo de Análisis de varianza. Modelo de dos factores**

Ítem $13a_1$y $a_2$. Se quiere estudiar el efecto de ciertas variables motivacionales sobre el rendimiento en tareas de logro. Se manipularon dos variables: "tipo de entrenamiento motivación" (A1: instrumental; A2: atribucional y A3: control) y "clima de clase" (B1: cooperativo; B2: competitivo y B3: individual).
Se seleccionaron a 45 sujetos y se dividieron en grupos para cada condición experimental.
A continuación presentamos la tabla del Análisis de varianza incompleta.

Tabla 1

| Fuente de variación | SC | Gl | CM | F |
|---|---|---|---|---|
| Factor A | 70 | | | |
| Factor B | | | 20 | |
| Interacción AB | | | | 3,91 |
| Error | 46 | | 1,278 | |
| Total | 176 | 44 | | |

A partir de la información de la tabla contesta a las preguntas de los ítems $13a_1$ y $13a_2$.

Los ítems elegidos no han tenido modificaciones, no ha habido sugerencias por parte de los expertos consultados, hemos decidido dejar los dos ítems invariantes, ya que sólo el $13a_1$ presentaba desequilibrio en distractores y ambos tienen análogo formato de redacción.

**Ítem $13a_1$**. El valor del sumatorio cuadrado para el factor B (ver Tabla 1)es:
a. 15,65
b. 35
c. **40**

**Ítem $13a_2$**. El valor de la media cuadrática para el factor A (ver Tabla 1) es:
a. 5
b. **35**
c. 15,6

**Contenido 14: Interpretación de una tabla Análisis de varianza. Modelo de dos factores.**

Tampoco esta unidad ha tenido comentarios por parte de los expertos que mejore su redacción, de donde queda con su formato original.

**Ítem 14a.** Una de las conclusiones del estudio sería (alfa = 0,05) (ver tabla 1)
a. **Hay efecto del factor A ("entrenamiento") sobre el rendimiento en tareas de logro**
b. No hay efecto del factor B ("clima de clase") sobre el rendimiento en tareas de logro
c. No hay interacción de los factores

**Contenido 15: Interpretación de salida de software. Identificación del modelo de Análisis de varianza**

No ha sufrido cambios este ítem, se mantiene invariante respecto del original, no ha habido sugerencias realizadas por parte de los expertos consultados. Si bien presentaba desequilibrio en los distractores decidimos dejarlo con igual redacción, ya que la misma, coincidiendo con los expertos, nos pareció clara.

La incorporación de la mujer al mercado de trabajo es un fenómeno que se verifica en todos los países de nuestro entorno geográfico y cultural. La clave de este proceso ha sido el interés creciente de la mujer por adquirir mayor formación. Pero el proceso es lento, con dificultades, sobre todo cuando se trata de integrar la vida familiar y laboral. En este año hemos estudiado la influencia que tienen diferentes variables sobre la satisfacción laboral.

Se han decido estudiar si hay efecto del "tipo de trabajo" (A1: trabajos de cualificación baja; A2: trabajos de cualificación media o A3: trabajos de cualificación alta) influyen sobre la "satisfacción laboral". También se ha tenido en cuenta, al mismo tiempo, si la "flexibilidad en el horario" (B1: si; B2: no) influye en la "satisfacción laboral".

**Tabla 5. Pruebas de los efectos inter-sujetos**

| Fuente | Suma de cuadrados | GL | Media cuadrática | F | Sig |
|---|---|---|---|---|---|
| Modelo corregido | 465,614(a) | 5 | 93,123 | ,852 | ,516 |
| Intersección | 17192,763 | 1 | 17192,763 | 157,355 | ,000 |
| Tipo de trabajo | 49,120 | 2 | 24,560 | ,225 | ,799 |
| Flexibilidad horaria | 270,878 | 1 | 270,878 | 2,479 | ,008 |
| Tipo * Flexibilidad | 34,383 | 2 | 17,191 | ,157 | ,855 |
| Error | 12564,981 | 115 | 109,261 | | |
| Total | 33298,000 | 121 | | | |
| Total corregida | 13030,595 | 120 | | | |

**Ítem 15b**. El tipo de análisis aplicado en este estudio (Tabla 5) es:

a. Contraste de dos medias independientes.
b. Análisis de varianza de un factor, con medidas repetidas.
c. **Análisis de varianza de dos factores completamente aleatorizado.**

## Contenido 16: Interpretación de salida de software. Resultados

No ha sufrido cambios este ítem, se mantiene invariante respecto del original, no habiendo sugerencias realizadas por parte de los expertos consultados.

Ítem 16b. Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha$=0,05 (ver Tabla 5):

a. **No hay diferencia estadísticamente significativas en la "satisfacción laboral" en función del "tipo de trabajo".**
b. Hay efecto del factor "flexibilidad horaria" sobre la "satisfacción laboral".
c. Hay interacción entre la "flexibilidad horaria" y "tipo de trabajo".

## Contenido 17. Interpretación de salida de software. Supuestos del modelo Análisis de varianza

Una vez analizados los supuestos de aplicación del modelo Análisis de varianza, vamos a estudiar si la "estabilidad en el trabajo" influye sobre la "satisfacción laboral". Estos son los resultados.

Tabla 8. Prueba de muestras

| | *Prueba de Levene* | | *Prueba T para la igualdad de medias* | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Varianzas | F | Sig. | T | Gl | Sig. (bilateral) | Diferencia de medias | Error típ. diferencia | 95% IC para la diferencia | |
| | | | | | | | | Inferior | Superior |
| Iguales | 3,39 | ,168 | 2,69 | 119 | ,008 | 5,82 | 2,164 | 1,535 | 10,104 |
| distintas | | | 4,54 | 110,17 | ,000 | 5,82 | 1,280 | 3,284 | 8,356 |

Tampoco se le han practicado cambios a este ítem, pues no se han encontrado sugerencias por lo que, la redacción entonces es la que sigue:

**Ítem 17b$_1$.** De acuerdo con la prueba de Levene (ver tabla 8):
a. Se cumple el supuesto de homocedasticidad
b. **No se cumple el supuesto de homocedasticidad**
c. No se puede concluir sobre la homocedasticidad

## 3.10. CONCLUSIONES DEL ESTUDIO 1

En este Capítulo se ha descrito el Estudio 1, orientado a la construcción de un cuestionario para evaluar la comprensión de los estudiantes de Psicología sobre el análisis de la varianza. Para ello se delimitó el contenido del instrumento, se construyó un banco de ítems y se seleccionaron los que constituirán el cuestionario, mediante juicio de expertos y pruebas empíricas de ítems. A continuación se exponen las principales conclusiones respecto a los objetivos planteados.

El objetivo principal del Estudio 1 fue construir un cuestionario siguiendo reglas psicométricas adecuadas (en particular las recomendadas por APA, AERA y NCME, 1999). Se entendería la comprensión, de acuerdo a nuestro marco teórico, y en el sentido descrito en la Sección 2.2.5, es decir, como correspondencia entre los significados personales e institucionales y dependiente de la institución de enseñanza (Godino, 1996).

Para ello, se realizó en el Capítulo 1 un análisis detallado del objeto de estudio, que nos permitió dar una base sólida a la variable objeto de medición; en la Sección 3.3.2 se presenta la lista de unidades semántica, en cuya elección se tuvo en cuenta, abarcar el contenido analizado el Capítulo 1 y por otra parte no ser excesivamente amplios, para lograr un cuestionario de longitud conveniente. En la Sección 3.3.2 también se detallan los objetos matemáticos englobados en la definición semántica de la variable, clasificados de acuerdo a nuestro marco teórico. Con ello se finaliza la delimitación del significado institucional pretendido en el cuestionario.

Se construyó un banco de ítems partiendo de las investigaciones previas y de pruebas anteriores de evaluación en el curso seguido por los estudiantes de la muestra, así como de algunos libros de estadística en Psicología. Se tuvo en cuenta las recomendaciones sobre redacción de ítems de diversos textos como Kline (2013). El análisis a priori del conjunto de ítems mostró que se disponía de al menos dos ítems por cada una de las unidades de contenido. Puesto que para cada contenido, se incluye finalmente al menos un ítem, se proporciona una primera justificación de la validez de contenido del cuestionario.

El análisis de los ítems también permite cumplir el segundo objetivo, el cual es que el instrumento permita identificar las principales dificultades de comprensión de los alumnos en este tema, para planificar una intervención, en caso necesario.

Finalizada la construcción del banco de ítems, se sometieron a una doble prueba. El estudio de los ítems, mediante el juicio de expertos, por medio de un procedimiento estructurado, ha permitido emparejar los ítems con la definición semántica de la variable. Este emparejamiento ha determinado que el contenido de nuestro futuro instrumento de medición sea relevante y representativo de un universo de observaciones admisibles ligadas a la comprensión de nuestro objeto de estudio y por tanto, constituye otra evidencia de la validez de contenido.

Las pruebas empíricas de ítems (con muestras de 93 y 130 estudiantes) permitieron estimar el tiempo de respuesta, mejorar la redacción en algún caso y obtener estimaciones de los índices de dificultad y discriminación. Respecto a los primeros, se obtuvo una gama amplia de dificultad, y una distribución normal, que permite elegir finalmente ítems fáciles, intermedios y difíciles para poder evaluar diferentes grados de comprensión. Respecto a la discriminación, en general todos discriminan (diferencia media entre el grupo superior e inferior positiva) aunque en algunos casos la discriminación es pequeña o no estadísticamente significativa. También se encuentra un amplio conjunto de ítems de alta discriminación.

Finalizado este estudio de los ítems, con el doble procedimiento de juicio de expertos y de prueba empírica, se seleccionaron los que constituirán el cuestionario. Para cada contenido se tiene en cuenta el grado de acuerdo entre jueces, alta puntuación y valores adecuados de dificultad y discriminación. Con todo ello se finaliza el Estudio 1 con un cuestionario validado que consta de un total de 20 ítems que cubren las unidades de contenido fijadas en la definición semántica de la variable.

# CAPITULO 4.
# ESTUDIO 2: EVALUACIÓN DE CONOCIMIENTOS MEDIANTE CUESTIONARIO


4.1. Introducción
4.2. Objetivos del Estudio 2
4.3. Hipótesis del Estudio 2
4.4. Descripción de la muestra
4.5. Material y método
4.6. Resultados detallados por ítem
4.6.1. Intervalo de confianza
4.6.2. Contraste de hipótesis
4.6.3. Modelo de análisis de varianza. Selección y supuestos
4.6.4. Cálculo en análisis de varianza
4.6.5. Interpretación de resultados
4.7. Índice de dificultad de la prueba
4.8. Estudio de la puntuación global
4.9. Estudios de validación
4.9.1. Validez de contenido
4.9.2. Fiabilidad de consistencia interna
4.9.3. Validez discriminante de ítems
4.9.4. Análisis factorial y validez de constructo
4.10. Generalizabilidad
4.11. Conclusiones sobre la evaluación con cuestionario


## 4.1. INTRODUCCIÓN

En este capítulo presentaremos los resultados del estudio de evaluación realizado con el cuestionario, cuya construcción se ha descrito en el Capítulo 3, para cumplir el tercer objetivo de nuestro trabajo. A partir de las respuestas obtenidas en una muestra amplia de estudiantes de Psicología que han estudiado el tema, podremos informar sobre algunos errores y dificultades de los estudiantes, respecto al análisis de varianza y prerrequisitos para su comprensión. Asimismo el estudio nos permitirá analizar algunas hipótesis sobre la comprensión de los objetos estadísticos incluidos en el estudio del análisis de varianza. Con todo ello informaremos sobre el significado personal logrado por los estudiantes que participan en el estudio.

Comenzamos el capítulo describiendo los objetivos e hipótesis específicas de este nuevo estudio. Seguidamente se describe la muestra de estudiantes que respondió el cuestionario, el material y método. A continuación analizamos los resultados de la evaluación en cada uno de los ítems, agrupándolos de acuerdo a su contenido principal. También realizamos un análisis de la puntuación global en el cuestionario. El análisis se completa con un estudio de la validez de contenido, fiabilidad y generalizabilidad, finalizando con la exposición de las principales conclusiones del estudio. Resultados parciales de este estudio se han publicado en Vera y Díaz (2010, 2013 y 2014), Vera, Díaz y Batanero (2011) y Batanero, Vera y Díaz (2012).

## 4.2. OBJETIVOS DEL ESTUDIO 2

El *objetivo principal del Estudio 2 es evaluar la comprensión del objeto estadístico análisis de la varianza por parte de estudiantes de Psicología*. Dicha evaluación se realiza después de que los estudiantes han estudiado el tema, siguiendo la instrucción descrita en el Capítulo 1, y utilizando el cuestionario cuya construcción rigurosa se describió en el Capítulo 3.

Puesto que la muestra es amplia, nos permitirá encontrar tendencias y visualizar la variabilidad del significado personal de los estudiantes acerca del objeto de estudio; todo ello teniendo en cuenta el significado institucional de referencia, que sirve de pauta de comparación con los significados personales que se evalúan. Como indicadores de esta comprensión se cuantifican las respuestas correctas, elección de distractores y respuestas en blanco, comparando tanto con los resultados de las pruebas piloto como con los de investigaciones previas.

*Un segundo objetivo es obtener evidencias de fiabilidad y generalizabilidad de la prueba, que podría ser utilizada en investigaciones posteriores.* A través de los coeficientes de fiabilidad y generalizabilidad se valorará el grado posible de generalización de resultados a otros estudiantes con características similares y a otros ítems con semejanzas con los utilizados. Justificaremos, igualmente, la validez de contenido y validez discriminante del instrumento.

## 4.3. HIPÓTESIS DEL ESTUDIO

Como se ha indicado, casi no existen investigaciones didácticas que reflejen la comprensión del análisis de varianza, por lo que este estudio ha de interpretarse como exploratorio. Sin embargo, puesto que los ítems del cuestionario se probaron con muestras pilotos, es de esperar que se reproduzcan los resultados de aquellas, lo que nos permite formular algunas hipótesis. Con todo, las hipótesis se deben entender como conjeturas o expectativas acerca de lo que esperaríamos obtener de nuestros resultados, y para ayudar a orientar el desarrollo del trabajo y elaborar los informes.

> H1: Una primera hipótesis es que *se espera confirmar los errores y dificultades ya encontrados en la investigación previa realizada, en relación con algunos de los prerrequisitos al análisis de varianza*.

Es posible, que, como en las investigaciones de Alvarado (2007) y Olivo (2008) nuestros estudiantes no usen la información del tamaño muestral en forma adecuada en la distribución muestral. Respecto al intervalo de confianza esperamos encontrar estudiantes que no tengan en cuenta los factores que determinan el ancho del intervalo, al igual que en la investigación de Behar (2001) y otros que den una interpretación bayesiana al significado del intervalo, como ocurrió en la tesis de Olivo (2008). Los errores descritos en el Capítulo 2 sobre la comprensión del contraste de hipótesis, y encontrados, entre otras investigaciones en las de Vallecillos (1994; 2002) y Cañadas (2012) también podrían presentarse.

> H2: Una segunda hipótesis es que *se espera la existencia de dificultades a la hora de interpretar los resultados de los cálculos respecto a los contextos de los problemas planteados*.

Esta hipótesis se apoya en la investigación previa, puesto que algunos estudiantes, luego de calcular un intervalo de confianza interpretaron incorrectamente su significado en un contexto dado en la investigación de Olivo (2008). También en Arteaga (2011) como en Cañadas *et al.* (2012) se reporta un alto índice de estudiantes que no alcanza a interpretar los resultados en términos del problema. Los autores señalan como causa de ello a que este es el último paso en el proceso de modelización y no suele trabajarse mucho dentro de la clase de estadística.

H3. Una tercera hipótesis es que *observaremos dificultad por parte del estudiante a la hora de interpretar un resultado proveniente del cálculo que realiza un software estadístico*.

Se apoya esta hipótesis en otros trabajos que indican que el estudiante pueda no reconocer el tipo de análisis estadístico que se aplica o no interpretar el p-valor presentado en un contraste de hipótesis (Vallecillos, 1994) o que no sea capaz de interpretar pruebas de diagnóstico. El uso de software puede mejorar la comprensión; sin embargo, el ordenador añade nueva información en el aprendizaje y esto requiere un coste y esfuerzo añadido por parte de los estudiantes (Chance, delMas y Garfield, 2004).

## 4.4. DESCRIPCIÓN DE LA MUESTRA

La muestra de participantes ha estado formada por un total de 224 estudiantes de segundo año de la Licenciatura en Psicología de la Universidad de Huelva, que cursaban una asignatura de Análisis de Datos II, cuya profesora colaboró con nosotros en la recogida de datos de los estudiantes. Los contenidos que abarca esta asignatura, así como de otras cursadas por el estudiante, que se enmarcan dentro del área de Métodos de Investigación en las Ciencias del Comportamiento se describieron con detalle en el Capítulo 1. Esta muestra no fue aleatoria, sino de tipo intencional; sin embargo todos los estudiantes de los cursos donde se pasó el instrumento han participado, entonces se la puede considerar como representativa del grupo objeto de estudio en la investigación (Ghiglione y Matalón, 1991). Los estudiantes completaron los ítems como parte de la evaluación final de la asignatura, en la que habían estudiado entre otros temas, el análisis de varianza.

Se prepararon dos cuestionarios diferentes (versiones A y B) con las mismas preguntas pero en orden diferente. De los 224 estudiantes que componen la muestra, 113 contestaron el cuestionario A y 111 el cuestionario B, que fueron aplicados de manera simultánea en una sesión de una hora y media. La proporción de varones y mujeres fue la habitual en esta titulación y la edad entre 19 y 20 años, con pocos casos en que se superase. La versión A se presenta en el Anexo 1.

## 4.5. MATERIAL Y MÉTODO

El material que hemos utilizado es el cuestionario que se presenta en la Tabla 4.1, (versión A) resaltando en negrita la opción correcta consta de 20 ítems, cada uno con tres distractores. Está dividido en dos partes: una primera donde se evalúan contenidos teóricos y otra donde se evalúan contenidos aplicados al software.

Tabla 4.1. Cuestionario, versión A

**PARTE 1: EVALUACIÓN DE CONTENIDOS TEÓRICOS**

**Ítem 1.** Se toma una muestra de 48 estudiantes con la finalidad de conocer su rendimiento en lectura. Un psicólogo escolar administra un test, así obtiene una media de 54,7 y desviación típica 5,23. Los límites del intervalo de confianza para la puntuación media poblacional del test administrado, con un nivel de confianza α=0,99 son:

a. [53,95; 55,45].
b. [42,53; 66,87].
c. **[52,76; 56,64].**

(Error típico de la media: $S_d/\sqrt{n}$ ).

**Ítem 2.** Un investigador ha calculado 1000 intervalos de confianza al 99% para la media poblacional μ. Suponiendo que los datos en los que se basan los intervalos se seleccionan de manera independiente entre sí. ¿Cuántos de estos 1000 intervalos se espera que contengan el valor verdadero de μ?

a. **Aproximadamente 990 intervalos contendrán el valor de μ.**
b. Los 1000 intervalos contendrán el valor de μ.
c. No se puede saber a priori la cantidad de intervalos que se espera que capturen el verdadero valor del parámetro μ.

**Ítem 3**. Queremos conocer si los sujetos extrovertidos e introvertidos difieren en la puntuación media en autoestima y no disponemos de ninguna información previa. El tipo de hipótesis nula razonable que debo plantear es:

a. Ho: $\mu_I \leq \mu_E$.
b. Ho: $\mu_I \geq \mu_E$.
c. **Ho: $\boldsymbol{\mu_I = \mu_E}$.**

**Ítem 4.** De los siguientes pares de afirmaciones, indique cual NO cumple con las reglas para plantear hipótesis estadísticas:

a. Ho: $\mu \leq 100$; $H_1$: $\mu > 100$.
b. **Ho: $\sigma = 15$; $H_1$: $\sigma \leq 15$.**
c. Ho: $\mu_1 - \mu_2 = 0$; $H_1$: $\mu_1 - \mu_2 \neq 0$.

**Ítem 5.** Si en una investigación lees que la potencia de un contraste vale 0,5594, entonces interpretas que:

a. La probabilidad de rechazar la hipótesis nula siendo falsa vale (1 -0,5594) = 0,4406.
b. **La probabilidad de rechazar la hipótesis nula siendo falsa vale 0,5594.**
c. La probabilidad de mantener la hipótesis nula siendo falsa vale 0,5594.*

**Ítem 6**. Supongamos que conocemos la „verdad absoluta" sobre la eficacia de dos tratamientos (A y B), y sabemos que existen diferencias en la efectividad de ambos para curar la depresión. Un investigador que realice un estudio y parta de la hipótesis „no existen diferencias en la efectividad de los tratamientos A y B para curar la depresión" cometerá un error tipo II cuando:

a. Concluya que A y B no son efectivos para curar la depresión.
b. **Concluya que A y B no difieren en su efectividad para curar la depresión.**
c. Concluya que A y B difieren en su efectividad para curar la depresión.

**Ítem 7.** El valor crítico correspondiente a un α= 0,01 en un contraste unilateral derecho, suponiendo normalidad de los datos es:

a. -2,33.
b. **2,33.**
c. 3,10.

**Ítem 8.** Cuando realizamos un contraste, la regla de decisión nos lleva a rechazar la hipótesis nula siempre que:

a. El estadístico de contraste caiga en la región de rechazo.
b. La probabilidad asociada al estadístico de contraste (el nivel de significación) sea menor que el valor de alfa.
c. **a) y b) son correctas.**

**Ítem 9.** Para mejorar la psicomotricidad de los niños de primaria, una maestra cree que ayudarán unas nuevas actividades físicas. La maestra divide aleatoriamente su grupo de trabajo en tres partes iguales. A cada grupo le aplica un tipo de ejercicio diferente, pues desea saber qué tipo de ejercicios le dará mejores resultados. De las técnicas estadísticas que siguen, ¿cuál debería aplicar la maestra para comprobar si los métodos que aplica son diferentes?

a. Contraste de hipótesis *t* sobre medias independientes.
b. Contraste de hipótesis *t* sobre medias relacionadas.
c. **Análisis de varianza de un factor completamente aleatorizado.**

**Ítem 10.** Un investigador utilizará un análisis de varianza de dos factores, efectos fijos y completamente aleatorizado cuando:

a. En el estudio haya una variable independiente, con dos niveles seleccionados al azar.
b. **En el estudio haya dos variables independientes, cada una con dos o más niveles.**
c. En el estudio haya dos variables dependientes.

**Ítem 11.** Los supuestos de aplicación del análisis de varianza de dos factores, efectos fijos y completamente aleatorizado son:

a. Independencia de las observaciones, normalidad de las distribuciones, y aditividad.
b. Independencia de las observaciones, igualdad de varianzas y aditividad.
c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas.**

**Ítem 12**. El análisis de varianza de dos factores, con efectos fijos, descompone la variabilidad total en los siguientes componentes:

a. Variabilidad total = V. entre grupos + V. error.
b. Variabilidad total = V. entregrupos + V. entresujetos + V. error.
c. **Variabilidad total = V. factor A + V. factor B + V. interacción + V. error.**

**Ítem 13.** Si en un análisis de varianza de un factor, y medidas repetidas encuentro que la *F* empírica u observada toma un valor de 8,16 esto quiere decir que:

a. **CM entregrupos / CM error = 8,16.**
b. CM entresujetos / CM error = 8,16.
c. CM entregrupos / CM intrasujetos = 8,16.

Se quiere estudiar el efecto de ciertas variables motivacionales sobre el rendimiento en tareas de logro. Se manipularon dos variables: "tipo de entrenamiento motivacional" (A1: instrumental; A2: atribucional y A3: control) y "clima de clase" (B1: cooperativo; B2: competitivo y B3: individual). Se seleccionaron a 45 sujetos y se dividieron en grupos para cada condición experimental. A continuación presentamos la tabla del Análisis de varianza incompleta.

**Tabla 1**

| Fuente de variación | *SC* | *GL* | *MC* | *F* |
|---|---|---|---|---|
| Factor A | 70 | | | |
| Factor B | | | 20 | |
| Interacción AB | | | | 3,91 |
| Error | 46 | | 1,278 | |
| Total | 176 | 44 | | |

A partir de la información de la Tabla 1 contesta a las preguntas de los ítems 14, 15 y 16.

**Ítem 14.** El valor del sumatorio cuadrado para el factor B (ver Tabla 1) es:

a. 15,65.
b. 35.
c. **40.**

**Ítem 15**. El valor de la media cuadrática para el factor A (ver Tabla 1) es:

a. 5.
b. **35.**
c. 15,65.

**Ítem 16.** Una de las conclusiones del estudio sería (alfa = 0,05) (usar Tabla 1 completa)
a. **Hay efecto del factor A ("entrenamiento") sobre el rendimiento en tareas de logro.**
b. No hay efecto del factor B ("clima de clase") sobre el rendimiento en tareas de logro.
c. No hay interacción de los factores.

**PARTE 2: CONTENIDOS APLICADOS AL SPSS**

La incorporación de la mujer al mercado de trabajo es un fenómeno que se verifica en todos los países de nuestro entorno geográfico y cultural. La clave de este proceso ha sido el interés creciente de la mujer por adquirir mayor formación. Pero el proceso es lento, con dificultades, sobre todo cuando se trata de integrar la vida familiar y laboral. En este año hemos estudiado la influencia que tienen diferentes variables sobre la satisfacción laboral.

Se ha decidido estudiar si hay efecto del "tipo de trabajo" (A1: trabajos de cualificación baja; A2: trabajos de cualificación media ó A3: trabajos de cualificación alta) sobre la "satisfacción laboral". También se ha tenido en cuenta, al mismo tiempo, si la "flexibilidad en el horario" (B1: si; B2: no) influye en la "satisfacción laboral".

Tabla 2. Pruebas de los efectos inter-sujetos

| Fuente | Suma de cuadrados | GL | Media cuadrática | F | Significación |
|---|---|---|---|---|---|
| Modelo corregido | 465,614(a) | 5 | 93,123 | ,852 | ,516 |
| Intersección | 17192,763 | 1 | 17192,763 | 157,355 | ,000 |
| Tipo de trabajo | 49,120 | 2 | 24,560 | ,225 | ,799 |
| Flexibilidad horaria | 270,878 | 1 | 270,878 | 2,479 | ,008 |
| Tipo * Flexibilidad | 34,383 | 2 | 17,191 | ,157 | ,855 |
| Error | 12564,981 | 115 | 109,261 | | |
| Total | 33298,000 | 121 | | | |
| Total corregida | 13030,595 | 120 | | | |

**Ítem 17.** El tipo de análisis aplicado en este estudio (Tabla 2) es:
a. Contraste de dos medias independientes.
b. Análisis de varianza de un factor, con medidas repetidas.
c. **Análisis de varianza de dos factores completamente aleatorizado.**

**Ítem 18.** Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha$=0,05 (ver Tabla 2) :
a. Hay diferencias estadísticamente significativas en la "satisfacción laboral" en función del "tipo de trabajo".
b. **Hay efecto del factor "flexibilidad horaria" sobre la "satisfacción laboral".**
c. Hay interacción entre la "flexibilidad horaria" y "tipo de trabajo".

Una vez analizados los supuestos de aplicación del modelo análisis de varianza, vamos a estudiar si la "estabilidad en el trabajo" influye sobre la "satisfacción laboral". Estos son los resultados.

Tabla 3. Prueba de muestras

| | Prueba de Levene | | Prueba T para la igualdad de medias | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | 95% IC | |
| Varianzas | *F* | Sig. | *t* | gl | Sig. (bilateral) | Diferencia de medias | Error típ. | Inferior | Superior |
| iguales | 3,391 | ,168 | 2,690 | 119 | ,008 | 5,820 | 2,164 | 1,535 | 10,104 |
| distintas | | | 4,548 | 110,174 | ,000 | 5,820 | 1,280 | 3,284 | 8,356 |

**Ítem 19.** La hipótesis nula que se somete a contraste en la prueba de Levene es la siguiente:
a. Ho: $\mu_{SI} = \mu_{NO}$.
b. **Ho: $\sigma^2_A = \sigma^2_B$.**
c. Ho: $\sigma^2_A \neq \sigma^2_B$.

**Ítem 20**. De acuerdo con la prueba de Levene (ver Tabla 3):
a. **Se cumple el supuesto de homocedasticidad.**
b. No se cumple el supuesto de homocedasticidad.
c. No se puede concluir sobre la homocedasticidad.

## 4.6. RESULTADOS DETALLADOS POR ITEM

En primer lugar, se presentan los resultados para cada ítem, analizando las respuestas y comparadas con los resultados en las pruebas piloto, así como con los de otras investigaciones previas.

En las tablas que siguen, se marca en negrita la respuesta correcta para cada caso, agrupamos el análisis de resultados en varios apartados, que corresponden al intervalo de confianza, contraste de hipótesis, modelo de análisis de varianza, cálculos e interpretación de resultados.

### 4.6.1. INTERVALO DE CONFIANZA

**Ítem 1.** Se toma una muestra de 48 estudiantes con la finalidad de conocer su rendimiento en lectura. Un psicólogo escolar administra un test, así obtiene una media de 54,7 y desviación típica 5,23. Los límites del intervalo de confianza para la puntuación media poblacional del test administrado, con un nivel de confianza α=0,99 son:

a. [53,95 ; 55,45]
b. [42,53 ; 66,87]
**c. [52,76 ; 56,64]**

(Error típico de la media: $S_d/\sqrt{n}$ )

Para responder el ítem, los estudiantes han de recordar y aplicar correctamente la fórmula de cálculo del intervalo de confianza para la media. Aunque no se especifique la distribución, puesto que el tamaño de la muestra es mayor que 30, los estudiantes aplicarían la distribución normal, debiendo leer sus tablas correctamente y encontrar el valor crítico.

Tabla 4.6.1. Distribución de respuestas al ítem 1

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 40 | 17,9 |
| b | 9 | 4,0 |
| **c** | **137** | **61,2** |
| sin contestar | 38 | 17,0 |
| Total | 224 | 100,0 |

Los resultados se presentan en la Tabla 4.6.1 con un 61,2% de respuestas correctas. Aunque suponen una buena capacidad para la construcción de los intervalos en más de la mitad de la muestra (por tanto lo consideramos bueno), están algo alejados de los resultados de Olivo (2008), quien encuentra 79% de respuestas correctas en un ítem similar, incluso cuando el autor no da la fórmula del error típico, mientras que nosotros la damos. El porcentaje de respuestas correctas en las pruebas piloto para este ítem, fue también mayor: 78,3% (n = 93) y 70% (n = 130), a pesar que los tres grupos de estudiantes tuvieron la misma instrucción y profesora.

En cuanto a los distractores el distractor (a) elegido por un 17,9% indica error en el uso de la tabla de distribución (pues no se calcula correctamente el percentil), y el distractor (b) elegido por el 4,0% evalúa el conocimiento de la desviación estándar para la media (que no se divide por la raíz cuadrada del tamaño de la muestra). Olivo (2008) encuentra un 8,3% y 6,7% de fallos para estos errores, y otro 4,4% comete las dos faltas a la vez. En las pruebas piloto los porcentajes de faltas para distractores que evalúan los

mismos errores han sido 1,1% y 2,2% (n=93); 11,5% y 0% (n=130). Un porcentaje del 17% no responde al ítem, ocurriendo en las pruebas piloto, 19,6% (n=93) y 18,5% (n=130).

Ítem 2.Un investigador ha calculado 1000 intervalos de confianza al 99% para la media poblacional μ. Suponiendo que los datos en los que se basan los intervalos se seleccionan de manera independiente entre sí. ¿Cuántos de estos 1000 intervalos se espera que contengan el valor verdadero de μ?

a. **Aproximadamente 990 intervalos contendrán el valor de μ**
b. Los 1000 intervalos contendrán el valor de μ
c. No se puede saber a priori la cantidad de intervalos que se espera que capturen el verdadero valor del parámetro μ.

El segundo ítem evalúa la comprensión del significado de un intervalo de confianza. Aunque de acuerdo con la Tabla 4.6.1 no son muchos los estudiantes que presentan dificultades para construir un intervalo de confianza, son muy pocos los que realmente comprenden su definición (27,2%, Tabla 4.6.2). En las pruebas piloto encontramos un 10,9% (n=93) y 22,3% (n=130) de aciertos, por lo que los resultados, aunque malos son algo mejores en la muestra definitiva.

Tabla 4.6.2. Distribución de respuestas al ítem 2

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| **a** | **61** | **27,2** |
| b | 11 | 4,9 |
| c | 25 | 11,2 |
| sin contestar | 127 | 56,7 |
| Total | 224 | 100,0 |

Olivo (2008) obtiene mejores resultados, pues encuentra un 36,5% de interpretaciones correctas en un ítem similar, en estudiantes de ingeniería. Se debe contemplar que se trata de estudiantes con mayor base en conocimientos de matemáticas que los de nuestra muestra. Al mismo tiempo Behar (2001) en un ítem similar, observa en un grupo de expertos en estadística y otro de estudiantes universitarios con instrucción especial en estadística un 48,9% y 35,7% de aciertos respectivamente.

Un 56,7% en nuestra muestra deja el problema en blanco (47,8% (n=93) y 60,8% (n=130) en las pruebas piloto), lo que refuerza nuestra conclusión acerca de lo difícil que resulta a los estudiantes interpretar un intervalo de confianza. En cuanto a los distractores para este ítem los porcentajes de estudiantes que los eligen no son significativos, ya dijimos que a falta de seguridad, en su mayoría prefieren no contestarlo. Esa puede ser la causa de la gran diferencia que se tiene con los porcentajes obtenidos en la investigación de Olivo (2008) para estos distractores (34,1% y 29%).

### 4.6.2. CONTRASTE DE HIPÓTESIS

Un segundo punto analizado son los conocimientos sobre contraste de hipótesis. Sobre este contenido, se incluyen a continuación varios ítems, cada uno de los cuáles evalúa la comprensión de alguno de los componentes de este objeto estadístico.

**Ítem 3.** Queremos conocer si los sujetos extrovertidos e introvertidos difieren en la puntuación media en autoestima y no disponemos de ninguna información previa. El tipo de hipótesis nula razonable que debo plantear es:

a. Ho: $\mu_I \leq \mu_E$
b. Ho: $\mu_I \geq \mu_E$
**c. Ho: $\mu_I = \mu_E$**

Este ítem evalúa el conocimiento del planteamiento de las hipótesis nula y alternativa, partiendo de un contexto de aplicación. Da un 93,3% (Tabla 4.6.3) de respuestas correctas, mientras que en las pruebas piloto se encontraron un 93,5% (n=93) y 94,6% (n=130) de respuestas correctas. Por tanto ha resultado fácil para los estudiantes, el planteamiento de las hipótesis en un contraste estadístico y parecen comprender la diferencia entre hipótesis estadística nula y alternativa: la hipótesis nula es siempre la de no diferencia, aunque el investigador, de hecho esté interesado en la existencia de diferencias.

Nuestros resultados son mejores que los de Vallecillos (1994), quien, en un ítem similar obtuvo un 68,6% de respuestas correctas en la muestra global de su estudio, y un 78,6% en una submuestra de estudiantes de Psicología. Puesto que nuestros estudiantes pasaron el cuestionario como parte de una evaluación final de la asignatura, esto podría explicar los mejores resultados en este ítem.

Tabla 4.6.3. Distribución de respuestas al ítem 3

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 4 | 1,8 |
| b | 10 | 4,5 |
| **c** | **209** | **93,3** |
| sin contestar | 1 | 0,4 |
| Total | 224 | 100,0 |

Un 1,8% confunde las hipótesis nula y alternativa (distractor (a)), confusión que se ha encontrado en investigaciones previas, como la de Vallecillos (1994) en un 6,4%, para la muestra total y 5,7% para Psicología. Rubin y Rosebery (1990) sugieren que esta confusión también se debe a que los estudiantes buscan resultados confirmatorios (y no falsatorios) de las hipótesis estadísticas; por ello no comprenden que las hipótesis nulas están redactadas en términos de la no existencia de efectos. En las pruebas piloto esta confusión la tiene un 4,3% (n=93) y 0% (n=130).

Otro 4,5% elige el distractor (b), confundiendo un contraste bilateral con uno unilateral, error denunciado por Batanero (2000) y que aparece en Cañadas (2012). En cuanto a las pruebas piloto este error lo cometen el 2,2% (n=93) y 1,5% (n=130). Finalmente, solo un estudiante (0,4%) no lo contesta, en las piloto 1,1% (n=93); 3,8% (n=130).

**Ítem4**. De los siguientes pares de afirmaciones, indique cual **NO** cumple con las reglas para plantear hipótesis estadísticas:

a. Ho: $\mu \leq 100$; $H_1$: $\mu > 100$
b. **Ho: $\sigma = 15$; $H_1$: $\sigma \leq 15$**
c. Ho: $\mu_1 - \mu_2 = 0$; $H_1$: $\mu_1 - \mu_2 \neq 0$

Este ítem, al igual que el anterior, evalúa el conocimiento del planteamiento de las hipótesis nula y alternativa. Resultó bastante sencillo a los estudiantes (84,8% de respuestas correctas, ver Tabla 4.6.4). En las pruebas piloto, hubo un 94,6% (n=93) en un ítem similar y 93,1% (n=130).

Tabla 4.6.4. Distribución de respuestas al ítem 4

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 13 | 5,8 |
| **b** | **190** | **84,8** |
| c | 13 | 5,8 |
| sin contestar | 8 | 3,6 |
| Total | 224 | 100,0 |

Vallecillos (1994) encontró un 56% de respuestas correctas y 77,1% para la muestra de estudiantes en Psicología, en un ítem similar. Es de destacar que el mayor porcentaje de respuestas correctas, lo obtiene la especialidad Matemáticas, y que el segundo mayor porcentaje lo tienen los estudiantes de la carrera de Psicología. Nuestros resultados son mejores que los de Vallecillos, incluso en estudiantes de Psicología, a pesar de ser más difícil el ítem, teniendo en cuenta que en nuestro caso, los estudiantes han de valorar conjuntamente las dos hipótesis (en lugar de una sola).

Los distractores (a) (hipótesis que no cubren el espacio muestral) y (c) (intercambio entre hipótesis nula y alternativa) tan sólo son elegidos por un 5,8% en ambos casos (Vallecillos, 1994 no incluye estos distractores); para las pruebas piloto los distractores son elegidos en un 2,3% y 3,8% (n=130), y 3,2% con 1,1% (n=93). Finalmente es muy bajo el porcentaje que no responde al ítem (3,6%), mientras que Vallecillos (1994) presenta un 17% el índice de no respuesta, que en la especialidad de Psicología se reducen al 11,4%. En las piloto 2,2% (n=93) y 0,8% (n=130) no resuelve.

**Ítem 5**. Si en una investigación lees que la potencia de un contraste vale 0,5594, entonces interpretas que:

a. La probabilidad de rechazar la hipótesis nula siendo falsa vale (1 -0,5594) = 0,4406.
b. **La probabilidad de rechazar la hipótesis nula siendo falsa vale 0,5594.**
c. La probabilidad de mantener la hipótesis nula siendo falsa vale 0,5594.

Tabla 4.6.5. Distribución de respuestas al ítem 5

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 39 | 17,4 |
| **b** | **114** | **50,9** |
| c | 45 | 20,1 |
| sin contestar | 26 | 11,6 |
| Total | 224 | 100,0 |

En este ítem, que evalúa la comprensión de los conceptos de error tipo I, II y de potencia, obtenemos un 50,9% de aciertos (Tabla 4.6.5). Las pruebas piloto con ítems similares, arrojaron un 76,1% (n=93) y 38,5% (n=130) y 50,8% (n=130) de respuestas correctas. Vallecillos (1994) obtiene 22,9% de respuestas correctas en un ítem similar (20% en Psicología), aunque ella da los distractores mediante símbolos; por ello puede haber sido más difícil que en nuestro caso, en que están dados verbalmente.

Además la autora pregunta por la probabilidad de error tipo II (y no por la potencia) y su discriminación con la probabilidad de error tipo I. La autora no incluye ítems sobre el concepto de potencia, razón por lo que con este ítem aportamos un resultado original.

Un 17,4% elige el distractor (a), confundiendo la potencia con nivel de significación (probabilidad de rechazar la hipótesis nula siendo verdadera), mientras que en la pruebas pilotos fue de 6,5% (n=93), 33,8% (n=130) y 8,5 % (n=130). Un 20,1% de nuestros estudiantes elige el distractor (c), confundiendo potencia y error tipo II (probabilidad de no rechazar la hipótesis nula siendo falsa). Para las pruebas piloto en sus ítems análogos se ha encontrado que lo escoge un 10,9% (n=93) y 7,7% (n=130) y 31,5% (n=130).

También hay un alto porcentaje de estudiantes que no responden el ítem (11,6%). En resumen, la comprensión del concepto de potencia es difícil, en línea con lo sugerido por Batanero (2000), Castro Sotos *et al*. (2007) y Díaz, Batanero y Wilhelmi (2008).

Ítem 6. Supongamos que conocemos la „verdad absoluta" sobre la eficacia de dos tratamientos (A y B), y sabemos que existen diferencias en la efectividad de ambos para curar la depresión. Un investigador que realice un estudio y parta de la hipótesis „no existen diferencias en la efectividad de los tratamientos A y B para curar la depresión" cometerá un error tipo II cuando:

a. Concluya que A y B no son efectivos para curar la depresión
b. **Concluya que A y B no difieren en su efectividad para curar la depresión**
c. Concluya que A y B difieren en su efectividad para curar la depresión

Tabla 4.6.6. Distribución de respuestas al ítem 6

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 7 | 3,1 |
| **b** | **145** | **64,7** |
| c | 39 | 17,4 |
| sin contestar | 33 | 14,7 |
| Total | 224 | 100,0 |

Para este ítem se evalúa si el alumno es capaz de diferenciar los errores que se podrían cometer al realizar un contraste de hipótesis, y de interpretar el enunciado de un problema (distractor (a)). Se tiene un 64,7% (Tabla 4.6.6) de respuestas correctas, mientras que en las pruebas piloto encontramos un 65,2% (n=93). Vallecillos (1994) plantea dos ítems relacionados con este. En el primero, donde se pide identificar la definición correcta de error tipo I, un 6,4% de estudiantes confunden las definiciones de error tipo I y tipo II (11,4% en Psicología). En el segundo ítem (que pregunta cuándo se comete error tipo I, una vez tomada una decisión) obtiene un 20% de estudiantes que dan erróneamente las condiciones para cometer error tipo II (24,3% en Psicología).

Un 17,4% responde la opción c. y por tanto confunde potencia con el nivel de significación (probabilidad de rechazar la hipótesis nula siendo verdadera); esta misma confusión la tienen un 18,3%) en la prueba piloto.

Sólo un 3,1% no interpretan enunciado (opción a), ya que el problema afirma que se conoce que ambos medicamentos han sido probados separadamente, resultando efectivos; en tanto que en esta confusión incurren un 1,1% (n=93) para la prueba piloto. Hay un alto porcentaje, 14,7%, de estudiantes que no responden al ítem, lo que conlleva

el desconocimiento sobre la diferencia entre hipótesis estadísticas e hipótesis de investigación; y entre error tipo I y tipo II (16,3 % lo deja en blanco para la prueba piloto).

**Ítem 7.** El valor crítico correspondiente a un $\alpha = 0{,}01$ en un contraste unilateral derecho, suponiendo normalidad de los datos es:
a. -2,33
b. **2,33**
c. 3,10

Tabla 4.6.7. Distribución de respuestas al ítem 7

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 52 | 23,2 |
| **b** | **144** | **64,3** |
| c | 10 | 4,5 |
| sin contestar | 18 | 8,0 |
| Total | 224 | 100,0 |

Con este ítem se evalúa la comprensión de la relación entre el nivel de significación y la región crítica, la diferencia entre contraste unilateral y bilateral, y el concepto de puntuación crítica. El porcentaje de aciertos es del 64,3% (Tabla 4.6.7), mayor que para la prueba piloto (34,8%; n=93) y menor en n=130, con 90,8%. Vallecillos no tiene algún ítem equivalente, resultando este entonces otro aporte de nuestro trabajo.

Un 23,2% ha decidido escoger el distractor (a) (46,7%; n=93 en la muestra piloto). Toman el contraste como unilateral izquierdo o lo que es lo mismo confunden hipótesis nula y alternativa o bien la región crítica y de aceptación. Se confirma lo expuesto por Vallecillos (1994), quien observa una confusión en la forma para construir las regiones de aceptación y rechazo al tener un test unilateral ó bilateral. Tauber (2001) y Batanero, Tauber y Sánchez (2001, 2004) en su investigación sobre la construcción de significados acerca de la distribución Normal, nos muestra un ítem similar, pero con el fin de observar el conocimiento de la gráfica de la función de densidad.

Un 4,5% elige el distractor (c) (estarían pensando en $\alpha = 0.001$), mientras que en las pruebas piloto lo elige un 5,4% (n=93) y 0,8% (n=130). Es bajo el porcentaje que lo deja en blanco, 8%, para las piloto 14% (n=93) y 4,6% (n=130).

**Ítem 8.** Cuando realizamos un contraste, la regla de decisión nos lleva a rechazar la hipótesis nula siempre que:
a. El estadístico de contraste caiga en la región de rechazo
b. La probabilidad asociada al estadístico de contraste (el nivel de significación) sea menor que el valor de alfa
c. **a) y b) son correctas**

Evaluamos a través de este ítem la comprensión del criterio de toma de decisión en un contraste. De la Tabla 4.6.8, observamos un 58% de respuestas correctas; en nuestras piloto, los porcentajes de respuestas correctas fueron 59,1% (n=93) y 68,5% (n=130). Además un 1,8% (2,2% (n=93), 3,1% (n=130) en las pruebas piloto) elige el distractor (b) y un 29,5% el distractor (a) (un 25,7% (n=93) y 20% (n=130) en las

pruebas piloto). En consecuencia, el concepto de región de rechazo ha sido comprendido por la mayoría de los estudiantes, pero no todos relacionan el hecho de que el estadístico caiga en dicha región con la necesidad que el valor *p* correspondiente sea menor que el nivel de significación. Estos últimos muestran dificultad de comprensión de la idea de distribución muestral, uno de los conceptos más relevantes en inferencia (Harradine, Batanero y Rossman, 2011).

Tabla 4.6.8. Distribución de respuestas al ítem 8

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 66 | 29,5 |
| b | 4 | 1,8 |
| **c** | **130** | **58,0** |
| sin contestar | 24 | 10,7 |
| Total | 224 | 100,0 |

Vallecillos (1994) plantea un ítem relacionado en que se da una hipótesis nula y dos alternativas diferentes, preguntando por la hipótesis que se debe aceptar. La autora encuentra un 33,7% de respuestas correctas (42,9% en el grupo de estudiantes de Psicología), resultado peor que el nuestro. Sin embargo, hay que tener en cuenta que el ítem planteado por ella es más difícil, pues los fallos podrían producirse por un error aritmético en el cálculo del estadístico de contraste o de lectura de las tablas de la distribución.

### 4.6.3. MODELO DE ANÁLISIS DE VARIANZA. SELECCIÓN Y SUPUESTOS

El resto de los ítems se refieren específicamente al análisis de varianza, teniendo en cuenta diferentes dimensiones en su comprensión y proporcionando resultados originales, en cuanto no hemos encontrado estudios previos de evaluación de la comprensión de este objeto.

El primer paso para realizar correctamente un problema es elegir un modelo adecuado. Sobre este punto se ha incluido cuatro ítems que estudian la elección de un modelo entre varios posibles, los supuestos que deben cumplir los datos para aplicar dicho modelo y la descomposición de la varianza total en diferentes componentes de acuerdo con el modelo seleccionado para aplicar a los datos.

**Ítem 9.** Para mejorar la psicomotricidad de los niños de primaria, una maestra cree que ayudarán unas nuevas actividades físicas. La maestra divide aleatoriamente su grupo de trabajo en tres partes iguales. A cada grupo le aplica un tipo de ejercicio diferente, pues desea saber qué tipo de ejercicios le dará mejores resultados. De las técnicas estadísticas que siguen, ¿cuál debería aplicar la maestra para comprobar si los métodos que aplica son diferentes?

a. Contraste de hipótesis *t* sobre medias independientes.
b. Contraste de hipótesis *t* sobre medias relacionadas.
c. **Análisis de varianza de un factor completamente aleatorizado.**

El primero de los ítems relacionados directamente con el análisis de varianza evalúa la comprensión de los estudiantes acerca de elementos estadísticos básicos del modelo elemental de análisis de varianza, y las diferencias entre ese modelo y el que se usa para una prueba *t*. Notamos (Tabla 4.6.9) que el 50,9% de los estudiantes contestan

correctamente al ítem, lo que nos dice que poseen una capacidad media de escoger el modelo que mejor se adecua al problema. En las pruebas piloto 81,7% (n=93) y 72,3% (n=130) de los estudiantes contestan correctamente el ítem; por tanto han mostrado un nivel ligeramente más alto.

Tabla 4.6.9. Distribución de respuestas al ítem 9

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 53 | 23,7 |
| b | 17 | 7,6 |
| **c** | **114** | **50,9** |
| sin contestar | 40 | 17,9 |
| Total | 224 | 100,0 |

Un 23,7% de la muestra elige el distractor (a); aunque este grupo es capaz de diferenciar en el problema entre muestras independientes y relacionadas dificultad señalada por Rubin y Rosebery (1990), no elige el procedimiento adecuado. En la muestra piloto lo elige el 16,9% (n=130) y 3,2% (n=93). Otro 7,6 %, además de elegir incorrectamente el contraste *t*, confunde muestras independientes y relacionadas. En la muestra piloto lo elige un 7,7% (n=130) y 3,2% (n=93).

Un 17,9% parece no conocer la diferencia entre estos supuestos, ya que no contesta el ítem, 11,8% (n=93) y 3,1% (n=130) proceden en forma análoga en las pruebas piloto, estando también por debajo de los resultados en el cuestionario definitivo.

**Ítem 10.** Un investigador utilizará un análisis de varianza de dos factores, efectos fijos y completamente aleatorizado cuando:
a. En el estudio haya dos variables dependientes.
b. En el estudio haya una variable independiente, con dos niveles seleccionados al azar.
c. **En el estudio haya dos variables independientes, cada una con dos o más niveles.**

Tabla 4.6.10. Distribución de respuestas al ítem 10

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 11 | 4,9 |
| b | 22 | 9,8 |
| **c** | **143** | **63,8** |
| sin contestar | 48 | 21,4 |
| Total | 224 | 100,0 |

Este ítem permite evaluar si el alumno comprende cuando un investigador aplicará un análisis de varianza de dos factores con efectos fijos, así como también la diferencia entre variables dependientes e independientes y el rol que estas juegan, dependiendo del modelo que se escoja. Hemos encontrado un 63,8% de respuestas correctas (Tabla 4.6.10), lo que supone una dificultad de baja a moderada. Rubin y Rosebery (1990) advierten sobre la dificultad de distinguir entre variables dependientes e independientes en un modelo de análisis de varianza. Para las pruebas piloto encontramos un 78,5% (n=130) de respuestas correctas, ligeramente superior al conseguido en esta prueba.

Otro 4,9% de sujetos (distractor (a)) desconoce los supuestos del modelo de análisis de varianza de dos factores, confundiéndolo con la prueba *t* de muestras pareadas, que podría aplicarse frente a esta tipología de datos, alcanzando un 3,8% (n=130) los resultados para las pruebas piloto. Un 9,8% (distractor (b)) confunde el modelo de análisis de varianza de dos factores con el de un factor con efectos fijos completamente aleatorizado; en las pruebas piloto un 8,5% (n=130) tiene esta confusión el cual casi se equipara a nuestros resultados. Es bastante alto el porcentaje de estudiantes que no responden el ítem (21,4%), mientras que es bastante menor este porcentaje para las pruebas piloto 9,2% (n=130).

**Ítem 11**. Los supuestos de aplicación del análisis de varianza de dos factores, efectos fijos y completamente aleatorizado son:

a. Independencia de las observaciones, normalidad de las distribuciones, y aditividad
b. Independencia de las observaciones, igualdad de varianza y aditividad.
c. **Independencia de las observaciones, normalidad de las distribuciones e igualdad de varianzas.**

Tabla 4.6.11. Distribución de respuestas al ítem 11

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 25 | 11,2 |
| b | 9 | 4,0 |
| **c** | **97** | **43,3** |
| sin contestar | 93 | 41,5 |
| Total | 224 | 100,0 |

En esta pregunta se quiere evaluar la comprensión de los supuestos que deben cumplir las observaciones para aplicar el análisis de varianza. Nos encontramos con un 43,3% de respuestas correctas (Tabla 4.6.11), lo que muestra una dificultad moderada baja en el ítem. En las pruebas piloto encontramos un 51,6% (n=93) de respuestas correctas; mientras que un ítems similar en otra prueba piloto, pero en que se preguntaba por los supuestos de aplicación de los modelos para un factor con efectos fijos encontramos un 50% (n=130) de respuestas correctas.

Un 11,2% olvida que para aplicar este modelo es necesario que las varianzas tomadas para cada nivel deben ser estadísticamente iguales (principio de homocedasticidad); en la pruebas piloto un 12,9% (n=93), y un 6,9% (n=130). Su importancia dependerá del fin del análisis pues si se viola este supuesto, (homogeneidad de varianzas), la prueba *F* sólo se verá afectada ligeramente en el modelo balanceado (el mismo tamaño de muestra en todos los tratamientos) con efectos fijos (Montgomery, 2005), pero es grave en caso de modelo de efectos aleatorios.

También un 4% olvida el supuesto de normalidad, mientras en las pruebas piloto lo hace un 5,4% (n=93) y un 3,1% (n=130). A la hora de decidir la importancia del cumplimiento de este supuesto se debe tener en cuenta que una desviación moderada de la normalidad no es motivo de gran preocupación para los modelos de efectos fijos (Montgomery, 2005). El ítem presenta un alto porcentaje de no respuesta (41,5%), 30,1% (n=93) y 40% (n=130), respetivamente en las pruebas pilotos. No hemos encontrado investigaciones que evalúen la comprensión por parte de los estudiantes de los supuestos en las pruebas de análisis de varianza.

**Ítem 12.** El análisis de varianza de dos factores, con efectos fijos, descompone la variabilidad total en los siguientes componentes:
a. Variabilidad total = V. entre grupos + V. error
b. Variabilidad total = V. entre grupos + V. entre sujetos + V. error
c. **Variabilidad total = V. factor A + V. factor B + V. interacción + V. error**

Tabla 4.6.12. Distribución de respuestas al ítem 12

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 13 | 5,8 |
| b | 28 | 12,5 |
| **c** | **155** | **69,2** |
| sin contestar | 28 | 12,5 |
| Total | 224 | 100,0 |

En este ítem queremos evaluar la capacidad de los estudiantes para asociar un modelo estadístico, de acuerdo con el estudio que se desea realizar de los datos. Obtuvimos un 69,2% (Tabla 4.6.12) de respuestas correctas; en las pruebas piloto, en un ítem similar, pero donde se pregunta utilizando simbología algebraica para los distractores se obtuvo un 52,7% (n=93); y para otro ítem análogo en otra de las pruebas piloto obtuvimos un 83,1% (n=130) de respuestas correctas.

Un 5,8% de la muestra confunde la descomposición de la variabilidad total de los datos con la de un modelo de análisis de la varianza de un factor (3,8%; n=130 en la prueba piloto y un 3,2% (n=93) en un ítem similar cuando se pregunta incluyendo las interacciones). También un 12,5% de la muestra confunde la descomposición de la variabilidad de un modelo de dos factores con el que corresponde a un modelo de un factor y medidas repetidas (6,9%; n=130) y 11,8% (n=93) en las pruebas piloto). Un 12,5% de la muestra no responde; en las pruebas piloto 32,3% (n=93) y 6,2% (n=130).

Si bien más de la mitad de la muestra evaluada es capaz de asociar un modelo estadístico adecuado, al análisis de varianza de dos factores, un 21% no está asociando la existencia de interacción con el modelo completo de análisis de la varianza de dos factores fijos. Al respecto, varios autores señalan a la interacción entre factores como el resultado peor interpretado entre los estudiantes de Psicología (Rosnow y Rosenthal, 1991; Pardo *et al*., 2007), así como dificultades para comprender su relación con la interpretación de los efectos principales (Green, 2007). Somos conscientes, en consecuencia, que se debe continuar tratando el tema con mayor profundidad, sobre todo al tratarse de estudiantes de la carrera Psicología, los cuales no tienen costumbre de manejar en su vida académica modelos matemáticos.

### 4.6.4. CÁLCULO EN ANÁLISIS DE VARIANZA

Una vez elegido un modelo de análisis de varianza es necesario realizar una serie de cálculos a partir de los datos para estimar los diferentes componentes de la tabla de análisis de varianza y realizar los contrastes de hipótesis requeridos. Los siguientes ítems evalúan la comprensión de los estudiantes en dichos cálculos.

**Ítem 13.** Si en un Análisis de varianza de un factor, y medidas repetidas encuentro que la *F* empírica u observada toma un valor de 8,16 esto quiere decir que:
a. **CM entregrupos / CM error = 8,16**
b. CM entresujetos / CM error = 8,16
c. CM entregrupos / CM intrasujetos = 8,16

A través de este ítem evaluamos la comprensión de algunos de los cálculos realizados para interpretar una tabla de análisis de varianza en el diseño de un factor con medidas repetidas. Más concretamente, se estudia si el alumno comprende los conceptos de cuadrado medio entre e intra sujetos y del error, así como el modo en que intervienen en el cálculo del estadístico *F*. Este ítem resultó con un nivel medio de dificultad, ya que un 57,1% (Tabla 4.6.13) lo responde correctamente. En la primera prueba piloto los resultados fueron mucho peores, pues se obtuvo un 7,6% (n=93) de respuestas correctas, mientras que para la segunda encontramos un 76,9% (n=130) de respuestas correctas. Quizá esta diferencia se deba a la gran variabilidad de las muestras, al nivel de preparación de la prueba y/o de la comprensión grupal e individual de las intervenciones.

Tabla 4.6.13. Distribución de respuestas al ítem 13

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| **a** | **128** | **57,1** |
| b | 16 | 7,1 |
| c | 49 | 21,9 |
| sin contestar | 31 | 13,8 |
| Total | 224 | 100,0 |

Un 7,1% elige el distractor (b), confundiendo la varianza entre grupos con la varianza entre sujetos, mientras en las pruebas piloto un 63% (n=93) y 3,8% (n=130) tuvieron este mismo error. Un 21,9% piensa que en el análisis de varianza de medidas repetidas se debe observar la interacción entre las variables sujeto-grupo; 16,3% (n=93) y 4,6% (n=130) en las pruebas pilotos, muestran una confusión, puesto que en estos modelos no se tiene en cuenta la interacción. No contamos con antecedentes para esta pregunta, por lo que se trata también de un aporte de nuestra investigación.

A partir de la información de la tabla contesta a las preguntas de los ítems 14, 15 y 16 se quiere estudiar el efecto de ciertas variables motivacionales sobre el rendimiento en tareas de logro. Se manipularon dos variables: "tipo de entrenamiento motivacional" (A1: instrumental; A2: atribucional y A3: control) y "clima de clase" (B1: cooperativo; B2: competitivo y B3: individual). Se seleccionaron a 45 sujetos y se dividieron en grupos para cada condición experimental. A continuación presentamos la tabla del Análisis de varianza incompleta.

Tabla 1

| Fuente de variación | SC | GL | CM | F |
|---|---|---|---|---|
| Factor A | 70 | | | |
| Factor B | | | 20 | |
| Interacción AB | | | | 3,91 |
| Error | 46 | | 1,278 | |
| Total | 176 | 44 | | |

**Ítem 14.** El valor del sumatorio cuadrado para el factor B (Tabla 1) es:

a. 15,65
b. 35
c. **40**

En los ítems 14 y 15 se realizan preguntas sobre el cálculo de los valores que forman una tabla de análisis de varianza de dos factores. En el primero, sobre la suma de cuadrados, se observa un 72,3% de respuestas correctas (Tabla 4.14), mientras en las pruebas piloto encontramos un 63,8% (n=47) y un 70% (n=130) para un ítem similar.

Tabla 4.14. Distribución de respuestas al ítem 14

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 8 | 3,6 |
| b | 9 | 4,0 |
| **c** | **162** | **72,3** |
| sin contestar | 45 | 20,1 |
| Total | 224 | 100,0 |

Un 3,6% elige el distractor (a) que da el valor del *F* observado para ese factor, confundiéndolo con la suma de cuadrados. En las pruebas piloto, para la muestra de tamaño n=47, nadie toma este distractor como el correcto, y en la de 130 estudiantes lo elige un 5,4% (n=130). Otro 4% confunde el valor del sumatorio cuadrado para el factor B con el de la media cuadrática para el otro factor 6,4% (n=47) y 2,3% (n=130) de las muestras piloto. Un alto porcentaje que no responde a la pregunta (20,1%), al igual que en las pruebas piloto, donde no contenta un 29,8% (n=47) y un 22,3% (n=130).

**Ítem 15.** El valor de la media cuadrática para el factor A (ver Tabla 1) es:
a. 5
b. **35**
c. 15,65

Tabla 4.6.15. Distribución de respuestas al ítem 15

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 6 | 3,1 |
| **b** | **162** | **72,3** |
| c | **7** | 3,1 |
| sin contestar | 49 | 21,4 |
| Total | 224 | 100,0 |

En este ítem que pide el valor de la media cuadrática tenemos un 72,3% de respuestas correctas (ver Tabla 4.6.15), y en las muestras piloto encontramos un 73,9% (n=46) y 70% (n=130). Por tanto resulto muy sencillo para la muestra.

Un 3,1% (distractor a.) confunde la media cuadrática de la interacción con la del factor A, mientras que esta confusión no la presenta ningún alumno en la prueba piloto de la muestra de tamaño n=46 y el 5,4% (n=130) en la segunda muestra. También encontramos un bajo porcentaje, 3,1%, de estudiantes que confunden la media cuadrática del factor A con el *F* observado del factor B (distractor c.) ; para las pruebas piloto un 4,3% (n=93) tienen esta confusión, mientras que un 2,3% (n=130) lo tiene en una unidad similar. El porcentaje de estudiantes que no responde fue 21,4%; 21,7% (n=46) y 22,3% (n=130) en las pruebas piloto. Es decir, en estos dos ítems la mayoría de estudiantes recuerda y ha comprendido el cálculo, pero una proporción importante no lo recuerda, puesto que no lo resuelve.

### 4.6.5. INTERPRETACIÓN DE RESULTADOS

Una vez finalizado los cálculos, el estudiante ha de interpretarlos para tomar una decisión adecuada sobre el efecto de los posibles factores y/ o interacción. Los siguientes ítems tratan de evaluar este conocimiento.

**Ítem 16.** Una de las conclusiones del estudio sería (alfa = 0,05) (usar Tabla 1 completa)
a. **Hay efecto del factor A ("entrenamiento") sobre el rendimiento en tareas de logro**
b. No hay efecto del factor B ("clima de clase") sobre el rendimiento en tareas de logro
c. No hay interacción de los factores

Tabla 4.6.16. Distribución de respuestas al ítem 16

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| **a** | **74** | **33,0** |
| b | 14 | 6,3 |
| c | 28 | 12,5 |
| sin contestar | 108 | 48,2 |
| Total | 224 | 100,0 |

En este ítem se quiere evaluar la interpretación de los valores obtenidos en la tabla completa, para tomar una decisión sobre la hipótesis. Encontramos un bajo porcentaje de respuestas correctas 33% (ver Tabla 4.6.16), y un alto porcentaje de no respuestas (48,2%). En las pruebas piloto las respuestas correctas fueron 25,8% (n=93) y 43,1% (n=130), y la no respuestas (62,4% (n=93) y 41,5% (n=130)).

Estos resultados nos indican la dificultad en este ítem encontrada por los estudiantes. Hay que tener en cuenta que, para dar respuesta al ítem, el alumno tiene primeramente que finalizar la tabla de análisis de varianza, donde obtendrá tres valores *F* que son todos estadísticamente significativos. En consecuencia hay efecto de los dos factores y de la interacción. El estudiante debe también interpretar correctamente un resultado significativo, interpretación que resultó difícil en la investigación de Vallecillos (1994) y Lecoutre (1999) y Cañadas (2012). También en las investigaciones sobre intervalos de confianza como la de Olivo (2008) la interpretación fue mucho más difícil que el cálculo.

Un 6,3% de la muestra (distractor b.) interpreta que no hay efecto del factor B, afirmación incorrecta, ya que el valor observado de *F* está muy por encima del valor del percentil; mientras que en las muestras piloto un 6,5% (n=93) y un 7,7% (n=130) en la segunda toma este distractor o uno similar como correcto. Un 12,5% (distractor c.) afirma que no existe efecto de interacción entre los factores, aunque también aquí el valor del *F* observado es mayor que el percentil obtenido de la tabla. En las pruebas piloto se encontró que eligen este distractor un 5,4% (n=93) y un 7,7% (n=130).

En resumen, hay un porcentaje apreciable de estudiantes que realizan una mala interpretación del resultado, en coincidencia con lo señalado en otros trabajos (Rosnow y Rosenthal, 1991; Green, 2007; Pardo *et al*., 2007).Mientras que a la hora de realizar los cálculos para determinar las puntuaciones de la tabla (ítems 14 y 15), parece que los estudiantes comprenden el tema, cuando se trata de interpretar (entrar en el proceso de argumentación) se encuentra una dificultad mayor. Por ello recomendamos insistir en los aspectos interpretativos en el estudio del tema.

La incorporación de la mujer al mercado de trabajo es un fenómeno que se verifica en todos los países de nuestro entorno geográfico y cultural. La clave de este proceso ha sido el interés creciente de la mujer por adquirir mayor formación. Pero el proceso es lento, con dificultades, sobre todo cuando se trata de integrar la vida familiar y laboral. En este año hemos estudiado la influencia que tienen diferentes variables sobre la satisfacción laboral. Se ha decidido estudiar si hay efecto del "tipo de trabajo" (A1: trabajos de cualificación baja; A2: trabajos de cualificación media o A3: trabajos de cualificación alta) sobre la "satisfacción laboral". También se ha tenido en cuenta, al mismo tiempo, si la "flexibilidad en el horario" (B1: si; B2: no) influye en la "satisfacción laboral".

Tabla 2. Prueba de los efectos Intra-sujetos

| Fuente | Sum.de Cuadr. | GL | Media Cuadrática | F | Significación |
|---|---|---|---|---|---|
| Modelo corregido | 465,614(a) | 5 | 93,123 | ,852 | ,516 |
| Intersección | 17192,763 | 1 | 17192,763 | 157,355 | ,000 |
| Tipo de trabajo | 49,120 | 2 | 24,560 | ,225 | ,799 |
| Flexibilidad horaria | 270,878 | 1 | 270,878 | 2,479 | ,008 |
| Tipo * Flexibilidad | 34,383 | 2 | 17,191 | ,157 | ,855 |
| Error | 12564,981 | 115 | 109,261 | | |
| Total | 33298,000 | 121 | | | |
| Total corregido | 13030,595 | 120 | | | |

**Ítem 17.** El tipo de análisis aplicado en este estudio (Tabla 2) es:

a. Contraste de dos medias independientes.
b. Análisis de varianza de un factor, con medidas repetidas.
c. **Análisis de varianza de dos factores completamente aleatorizado**

Tabla 4.6.17. Distribución de respuestas al ítem 17

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 11 | 4,9 |
| b | 21 | 9,4 |
| **c** | **167** | **74,6** |
| sin contestar | 25 | 11,2 |
| Total | 224 | 100,0 |

A través de este ítem, el primero de la serie de ítems que le siguen, se quiere evaluar si el grupo ha comprendido en general, el análisis de las salidas que ofrece un software estadístico; es decir si comprende los resultados arrojados desde un ordenador, luego de aplicar un modelo adecuado a un conjunto de datos. El porcentaje de respuestas correctas (74,6% de los estudiantes, ver Tabla 4.6.17) y 68,8% (n=93) y 81,5% (n=130) para las pruebas piloto es alto, lo que sugiere que el grupo posee una capacidad buena de interpretación de las salidas del software.

Un 4,9% (0,8% (n=130) y 7,5% (n=93) en las pruebas piloto) elige como correcto el distractor (a), contraste de dos medias independientes, sin tener en cuenta que se quiere estudiar si influyen o no dos variables independientes sobre una tercera ("tipo de trabajo" y "flexibilidad en horario" sobre la "satisfacción laboral"), además de estudiar si interactúan.

También un 9,4% (10% (n=130) y 5,4% (n=93) en las pruebas piloto), elige el modelo medidas repetidas, el cual no sería correcto por la independencia de las dos variables presentadas. También es bajo el porcentaje de no respuesta (11,2%), mientras que para las piloto es 18,3% (n=93) y 7,7% (n=130). No hay antecedentes para compararlo con otras investigaciones.

**Ítem 18**. Entre las posibles conclusiones del estudio se encuentra, considerando un $\alpha$=0,05 (ver Tabla 2):

a. Hay diferencia estadísticamente significativas en la "satisfacción laboral" en función del "tipo de trabajo".
b. **Hay efecto del factor "flexibilidad horaria" sobre la "satisfacción laboral".**
c. Hay interacción entre la "flexibilidad horaria" y "tipo de trabajo".

Tabla 4.6.18. Distribución de respuestas al ítem 18

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 31 | 13,8 |
| **b** | **60** | **26,8** |
| c | 37 | 16,5 |
| sin contestar | 96 | 42,9 |
| Total | 224 | 100,0% |

Este ítem también evalúa si el grupo es capaz de dar una respuesta correcta a la toma de decisión sobre la hipótesis a partir de las salidas de un software estadístico. Se evidencia la dificultad para los estudiantes, ya que sólo hubo un 26,8% de respuestas correctas (ver tabla 4.6.18) y 40,8% (n=93) y 30,8% (n=130) en las pruebas piloto.

Un 13,8% eligen como correcto el distractor a. (10,8% (n=93) y 21,5% (n=130) en las pruebas piloto), creyendo que el tipo de trabajo incide significativamente en la satisfacción laboral. Este resultado es incorrecto; basta con observar la significación de este factor en la Tabla 2. Un 16,5% elige como correcto del distractor (c) creyendo que existe interacción entre los dos factores, a pesar de que en la Tabla 2 encontramos en este nivel una significación 0,855. Para las pruebas piloto encontramos un 23,7% (n=93) y un 14,6% (n=130) que eligen un distractor similar como correcto.

Destacamos que es alto el porcentaje de no respuesta, 42,9%, y para las pruebas piloto 24,7% (n=93) y 33,1% (n=130). No existen investigaciones empíricas de este tipo para que puedan ser comparados los resultados.

**Ítem19**. Una vez analizados los supuestos de aplicación del modelo análisis de varianza, vamos a estudiar si la "estabilidad en el trabajo" influye sobre la "satisfacción laboral". Estos son los resultados (Tabla 3).

La hipótesis nula que se somete a contraste en la prueba de Levene es la siguiente:

a. Ho: $\mu_{SI} = \mu_{NO}$.
b. **Ho: $\sigma^2_A = \sigma^2_B$**
c. Ho: $\sigma^2_A \neq \sigma^2_B$

Tabla 3. Prueba de muestras

| | Prueba de Levene | | Prueba T para la igualdad de medias | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | F | Sig. | T | gl | Sig. (bilateral) | Diferencia medias | Error típ. Diferencia | 95% IC para la diferencia | |
| | | | | | | | | Inferior | Superior |
| Varianzas iguales | 3,391 | ,168 | 2,690 | 119 | ,008 | 5,820 | 2,164 | 1,535 | 10,104 |
| Varianzas distintas | | | 4,548 | 110,174 | ,000 | 5,820 | 1,280 | 3,284 | 8,356 |

En este ítem se solicita alguna conclusión acerca de la tabla de análisis de varianza obtenida con SPSS. Como en el ítem 16, el estudiante debe también interpretar correctamente un resultado significativo, interpretación que resultó difícil en la investigación de Vallecillos (1994) y Lecoutre (1999). Específicamente se quiere evaluar si el grupo ha comprendido como plantear una hipótesis estadística para comprobar la homocedasticidad, uno de los supuestos del modelo de análisis de la varianza, que analiza a través de la Prueba de Levene. Encontramos un 50,9% (Tabla 4.6.19) de respuestas correctas, mientras en la prueba piloto se encontró un 57,7% (n=130) para un ítem muy similar.

Tabla 4.6.19. Distribución de respuestas al ítem 19

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| a | 49 | 21,9 |
| **b** | **114** | **50,9** |
| c | 14 | 6,3 |
| sin contestar | 47 | 21,0 |
| Total | 224 | 100,0 |

Un 21,9% elige como correcto el distractor (a) confundiendo el contraste con una prueba para igualdad de dos media (en la prueba piloto un 26,2% (n=130) elige este mismo distractor). Un 6,3% elige el distractor (c) como correcto y un 2,3% (n=130) en la piloto sin tener en cuenta que no es posible esa sea una hipótesis nula adecuada (Vera, Batanero, Díaz, 2011). Hay un alto porcentaje de estudiantes que no responden 21%, mientras que en la piloto no lo hacen un 13,8% (n=130).

**Ítem 20**. De acuerdo con la prueba de Levene (Tabla 3):
a. **Se cumple el supuesto de homocedasticidad**
b. No se cumple el supuesto de homocedasticidad
c. No se puede concluir sobre la homocedasticidad

Tabla 4.6.20. Distribución de respuestas al ítem 20

| Respuesta | Frecuencia absoluta | Porcentaje |
|---|---|---|
| **a** | **79** | **35,3** |
| b | 27 | 12,1 |
| c | 14 | 6,3 |
| sin contestar | 104 | 46,4 |
| Total | 224 | 100,0 |

A través de este ítem, se quiere evaluar si el grupo ha comprendido las salidas de un software estadístico; se particulariza sobre la interpretación de la homocedasticidad, cuyo no cumplimiento es grave en modelos desbalanceados (Montgomery, 2005). Observamos también aquí un bajo porcentaje de respuestas correctas, 35,3% (Tabla 4.6.20), mientras que en las pruebas piloto hubo un porcentaje de respuestas correctas considerablemente más alto, 62,4% (n=93) y 61,5% (n=130).

Un 12,1% ha elegido el distracto (b) como correcto, evidenciando que no puede interpretar correctamente cuales son los valores en la significación de la prueba para no rechazar la igualdad de varianzas (en las pruebas piloto un 10,7 % (n=93) y un 16,2% (n=130)). Del mismo modo, un 6,3% elige como correcto el distractor c), 2,2% (n=93) y 6,9% (n=130) en las pruebas piloto, evidenciando que no interpreta esa salida del software. También presenta un alto porcentaje de no respuesta (46,4%); 24,7% (n=93) y 15,4% (n=130) para las pruebas piloto.

## 4.7. INDICE DE DIFICULTAD DE LA PRUEBA

En la Tabla 4.7.1 se presentan los índices de dificultad de los ítems del cuestionario, junto con los intervalos de confianza correspondientes. Un índice de dificultad cercano al 1, indica que el ítem ha resultado fácil pues la mayoría de los estudiantes lo ha resuelto correctamente.

Supongamos que con $X$ medimos la cantidad de estudiantes que ha contestado correctamente al ítem, entonces se tiene que en forma aleatoria, $X$ tomará valores entre 0 y 224. Por ello se puede asumir que $X$ distribuye en forma binomial con parámetros $n$=224, y probabilidad de éxito estimada por $\hat{p} = \frac{X}{n}$ , de donde cada intervalo de confianza se determina utilizando la aproximación normal (Vera, 2009):

$$LI = \hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{224}} \qquad LS = \hat{p} + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{224}}$$

Se trata de intervalos asintóticos ya que la distribución se aproxima a la normal por el tamaño muestral y aplicando el Teorema Central del Límite (Alvarado, 2007). Hemos usado un nivel de confianza del 95%. Los valores marcados en negrita en la Tabla 4.7.1, indican aquéllos ítems cuyo nivel de dificultad es alto ó bajo.

Tabla 4.7.1. Índice de Dificultad e Intervalo de Confianza para cada ID

| Item | ID | Int. Conf(ID) | Item | ID | Int. Conf(ID) |
|---|---|---|---|---|---|
| 1 | 0,6116 | (0,497;0,726) | 11 | 0,433 | (0,317;0,549) |
| 2 | 0,2723 | (0,168;0,377) | 12 | 0,692 | (0,584;0,8) |
| 3 | 0,9330 | (0,874;0,992) | 13 | 0,571 | (0,455;0,688) |
| 4 | 0,8482 | (0,764;0,932) | 14 | 0,723 | (0,618;0,828) |
| 5 | 0,5089 | (0,392;0,626) | 15 | 0,723 | (0,618;0,828) |
| 6 | 0,6473 | (0,535;0,759) | 16 | 0,330 | (0,220;0,441) |
| 7 | 0,6429 | (0,530;0,755) | 17 | 0,746 | (0,643;0,848) |
| 8 | 0,5804 | (0,465;0,696) | 18 | 0,268 | (0,164;0,372) |
| 9 | 0,5089 | (0,392;0,626) | 19 | 0,509 | (0,392;0,626) |
| 10 | 0,6384 | (0,526;0,751) | 20 | 0,353 | (0,241;0,465) |

*Ítems sencillos y difíciles.*

El más sencillo, con ID mayor ó igual que 0,9 fue el ítem 3, que evalúa el conocimiento del planteamiento de hipótesis estadísticas y en el cuál los estudiantes mostraron un buen conocimiento. Obtuvimos mejores resultados que los presentados por Vallecillos (1994) y los de Cañadas (2012), cuyos estudiantes en ítems similares cometieron más errores en el planteamiento de las hipótesis.

Fue también muy sencillo el ítem 4 con ID de casi 0,85 que pregunta, entre varias posibles hipótesis estadísticas, cuál no cumple las reglas de planteamiento de hipótesis. También en este caso se obtuvieron mejores resultados que los de Vallecillos (1994), aunque los distractores son distintos en nuestra investigación.

Finalmente queremos comentar el ítem 12 con casi 0,7 de valor en el ID; trata de la descomposición de la variabilidad total en un modelo de análisis de varianza de dos factores fijos, que los alumnos parecen haber comprendido. Aparentemente este resultado contradice las dificultades descritas tanto de estudiantes (Estepa, 2013) como de profesores (Sánchez *et al*., 2011) en torno a la comprensión del concepto de variabilidad aleatoria.

Respecto a los ítems difíciles (con ID menor que 0,2) no aparece ninguno, no obstante destacamos el ítem 2 (0,27) sobre la interpretación de un intervalo de confianza, coincidiendo por lo denunciado tanto por Behar (2001) como por Olivo (2008), quienes destacan la dificultad, al interpretar este objeto estadístico.

En cuanto a los ítems 18 (ID=0,268) y 16 (ID=0,33), ambos relacionados con la interpretación en contexto del resultado de un contraste de hipótesis (uno de los contrastes incluidos en el análisis de la varianza) destacamos, por un lado, que el criterio de decisión en un contraste resulta difícil a los estudiantes (Batanero, Díaz y Wilhelmi, 2008). En investigaciones previas se han encontrado fallos en este sentido: Falk (1996), Vallecillos (1994), Lecoutre (1999), Haller y Kraus (2002) encuentran deficiencias en estos conceptos entre profesores de metodología e investigación en Psicología. Además, en estos ítems es necesario también realizar el último paso en el proceso de modelización, interpretando un resultado, que resultó difícil en Cañadas *et al.* (2012).

## 4.8. ESTUDIO DE LA PUNTUACIÓN TOTAL

Además de estudiarse la puntuación de cada ítem, es interesante estudiar el número de respuestas correctas de cada estudiante, que nos da una idea de la proporción de elementos de significado adquiridos en relación con los elementos pretendidos al elaborar el instrumento de evaluación. Se trata, por cada alumno, del número de respuestas correctas al cuestionario. La media esperada sería 10 puntos, pues el cuestionario consta de 20 preguntas.

En la Figura 4.1 se presenta un diagrama de barras de la puntuación total, cuyos valores oscilan entre 0 y 20; aunque un solo caso alcanza el máximo y ninguno el mínimo. Se observa que el valor con mayor frecuencia en el puntaje es el 5, con un 11% de la muestra y el que le sigue es el 13 con un 9,4%. La mayoría de las frecuencias se concentran en el intervalo [10,13] valores que alcanzan el 20% de los estudiantes. El gráfico cuenta con dos modas bien diferenciadas, una correspondiente a 5 respuestas y la otra a 13. Observamos que aproximadamente la mitad de los alumnos se sitúa sobre la media teórica, lo que es usual en las pruebas de evaluación.

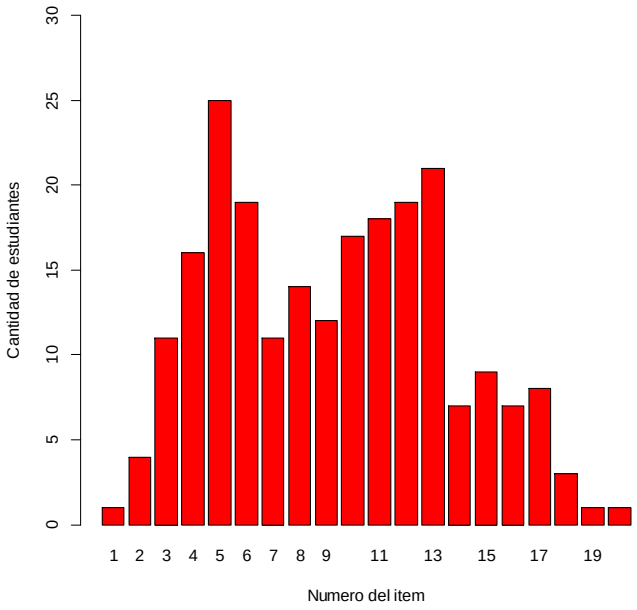



Figura 4.1. Diagrama de barras de la puntuación total en el cuestionario

Tabla 4.8.1. Estimación de estadísticos descriptivos de la puntuación total

| | Estadístico | Error típico |
|---|---|---|
| Media | 11,54 | 0,233 |
| Mínimo | 1 | |
| Máximo | 20 | |
| Rango | 19 | |
| Amplitud intercuartil | 5 | |
| Primer cuartil 25% | 9 | |
| Tercer cuartil 75% | 14 | |

En la Tabla 4.8.1, mostramos los estadísticos de la puntuación total, que refuerzan los comentarios anteriores. La media alcanza el valor 11,54 (cerca del valor esperado) con una desviación típica y varianza altas, pues hay bastante dispersión de los resultados. Un 50% de la muestra contestó correctamente entre 9 y 14 ítems (observar primer y tercer cuartil en Tabla 4.8.1); este resultado significa que las tres cuartas partes ha resuelto más de 9 ítems y menos de 14 ítems. Estos mismos resultados se deducen del gráfico de cajas (Figura 4.2).

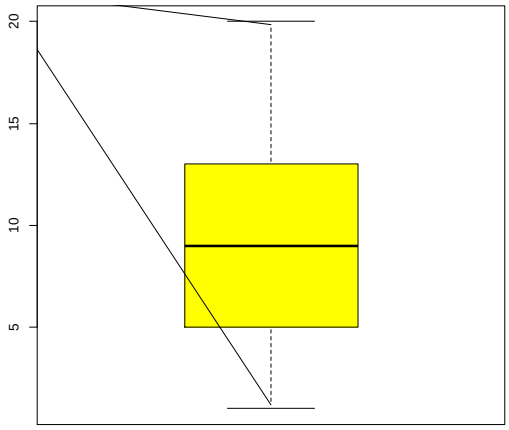



Figura 4.2. Diagrama de caja de la puntuación total

## 4.9. ESTUDIOS DE VALIDACIÓN

Para finalizar el Estudio 2 y cumplir su segundo objetivo se analizó la validez y fiabilidad que permiten controlar tanto los errores sistemáticos (sesgos) y aleatorios de las conclusiones obtenidas con el mismo (Thorndike, 1989). Seguimos parte de la metodología que se describe ya en Díaz (2007). Primeramente realizamos un estudio de validación del instrumento, desde el punto de vista de la validez de contenido, de la validez discriminante y la validez de constructo. Además se realizó el estudio de fiabilidad (Meliá, 2001). La medida siempre produce un cierto error aleatorio, siendo la fiabilidad la tendencia a la consistencia o precisión del instrumento en la población medida; para ello se partió de la teoría clásica de los test donde la fiabilidad se define como correlación entre las puntuaciones verdaderas y observada (Martínez Arias, 1995). Se utilizan dos procedimientos:

- Estimaremos el coeficiente de fiabilidad de consistencia interna Alfa, y el coeficiente $\boldsymbol{\theta}$, que se basan en los resultados de un análisis factorial cuya matriz de

análisis contiene las respuestas dicotómicas a cada ítem respecto de cada estudiante (Samapieri, Collado y Lucio, 2006). Podremos constatar también con este análisis estadístico que el constructo "comprensión del análisis de varianza" (por la diversidad de unidades de contenido descriptas en la definición semántica de la variable, Capitulo 3) es multifactorial.

- Haremos otra aproximación al problema de fiabilidad desde la teoría de la generalizabilidad (Martínez Arias, 1995), calculando un coeficiente de generalizabilidad, para otros ítems y sujetos, teniendo en cuenta entonces, estas dos posibles fuentes de error en el muestreo.

### 4.9.1. VALIDEZ DE CONTENIDO

Siguiendo a León y Montero (2002, p. 56), consideramos que un cuestionario tiene validez de contenido "*si los elementos que lo configuran son representativos de todo el universo que forman los aspectos que se pretende medir".* Al construir el cuestionario, siguiendo la metodología de Díaz (2007), hemos mostrado cuatro tipos de evidencia de validez de contenido:

1. Delimitamos el contenido a evaluar, a partir del análisis de significado de referencia que se hizo en el Capitulo 1 y del estudio de las dificultades previsibles de los estudiantes, analizados en los antecedentes del Capítulo 2.
2. Se seleccionaron posibles ítems tomando algunos de las investigaciones previas y otros de diferentes libros de texto (Manzano, 1997; García, 2004; Peers, 2006; Ferré y Ruiz, 2008 y Pagano, 2010) ó de exámenes anteriores en las asignaturas que se pasaron a los estudiantes previamente. Con el conjunto obtenido se realizó un estudio de cómo los mismos contribuyen a evaluar el constructo que subyace. De esta forma se llegó a una tabla de especificaciones de las unidades de medida de la variable objeto de medición (Tabla 1.8.2 en Capítulo 1) que cubría el significado de referencia.
3. Se utilizó una evaluación de ítems mediante el juicio a expertos (Lozada y López, 2003). En tal evaluación, participaron 11 expertos que nos informarán acerca del grado en que el contenido del instrumento es relevante y representativo del universo que se desea medir. Este es el método habitualmente utilizado en la validación de contenido.
4. Para la selección final de ítems se tuvo también en cuenta las pruebas piloto descriptas en el Capítulo 3. Finalmente se realiza un análisis para mostrar el contenido cubierto por los ítems del cuestionario, comparando con las especificaciones del contenido.

Todo lo descrito asegura una buena validez de contenido del instrumento.

### 4.9.2. FIABILIDAD DE CONSISTENCIA INTERNA

Se tomó el coeficiente Alfa de Cronbach (que se reduce al de Kuder- Richardson para ítems dicotómicos), como estimador del coeficiente de fiabilidad (Tabla 4.9.1). Este coeficiente refleja el grado en el que covarían los ítems que constituyen el test y estima la fiabilidad en el acto (Martínez Arias, 1995). El valor obtenido con el total de la muestra es de 0,79, el obtenido con los elementos tipificados difiere muy poco de este valor. Se considera que los valores son razonables, pues el cuestionario cubre una

amplia gama de conceptos, aunque no hay regla fija para cual debe ser el valor mínimo de este coeficiente (Morales, 2011). Diversos autores han dado distintas reglas (Santisteban, 1990), que indican como límite mínimo general 0,50.

Tabla 4.9.1. Estadísticos de fiabilidad

| Alfa de Cronbach | basado en los elementos tipificados | N de elementos |
|---|---|---|
| 0,790 | 0.788 | 20 |

### 4.9.3. VALIDEZ DISCRIMINANTE DE LOS ÍTEMS

También se ha analizado la validez discriminante, es decir la potencia de los ítems del cuestionario para discriminar a los estudiantes según su conocimiento. Por ello se espera que los estudiantes con puntuación total alta tengan mayor facilidad para contestar cada uno de los ítems. Para evaluar este tipo de validez (Tabla 4.9.2), se complementa el cálculo de la fiabilidad suprimiendo sucesivamente los ítems, para ver la variación del coeficiente alpa. Podemos observar que aumenta la fiabilidad, aunque no mucho, si suprimimos algún ítem de la lista. Por ello decidimos conservarlos a todos en el análisis. También es posible ver las correlaciones de cada ítem con la puntuación total en la prueba, si se suprime el ítem; que constituye un *índice de discriminación del ítem.* Todas las correlaciones excepto para el ítem 8 resultaron positivas indicando que casi todos los ítems contribuyen a la puntuación total. Sólo las correspondientes a los ítems 9 y 18 indican correlaciones cercanas 0, se dice entonces que ellas son las que menos contribución dan a la puntuación global.

Tabla 4.9.2. Estadísticos total-elemento

| | Media de la escala si se elimina el ítem | Varianza de la escala si se elimina el ítem | Correlación ítem-total corregida | Alfa de Cronbach si se elimina el ítem |
|---|---|---|---|---|
| Item 1 | 8,82 | 16,294 | 0,430 | 0,776 |
| Item 2 | 9,10 | 17,407 | 0,209 | 0,789 |
| Item 3 | 8,66 | 15,527 | 0,673 | 0,761 |
| Item 4 | 8,74 | 15,388 | 0,682 | 0,759 |
| Item 5 | 8,95 | 16,374 | 0,431 | 0,776 |
| Item 6 | 8,79 | 16,418 | 0,399 | 0,778 |
| Item 7 | 8,67 | 17,632 | 0,105 | 0,796 |
| Item 8 | 8,73 | 18,621 | -0,136 | 0,811 |
| Item 9 | 8,80 | 17,964 | 0,017 | 0,802 |
| Item 10 | 8,67 | 17,729 | 0,080 | 0,798 |
| Item 11 | 9,03 | 16,887 | 0,323 | 0,783 |
| Item 12 | 8,62 | 17,224 | 0,221 | 0,789 |
| Item 13 | 8,91 | 15,942 | 0,536 | 0,769 |
| Item 14 | 8,80 | 15,300 | 0,698 | 0,758 |
| Item 15 | 8,80 | 15,318 | 0,693 | 0,759 |
| Item 16 | 9,03 | 16,744 | 0,363 | 0,781 |
| Item 17 | 8,73 | 16,412 | 0,406 | 0,778 |
| Item 18 | 8,98 | 17,641 | 0,108 | 0,796 |
| Item 19 | 8,98 | 16,224 | 0,487 | 0,773 |
| Item 20 | 9,08 | 16,662 | 0,418 | 0,778 |

### 4.9.4. ANÁLISIS FACTORIAL Y VALIDEZ DE CONSTRUCTO

Puesto que evaluamos muchos conocimientos diferentes, debemos asumir que el cuestionario evalúa un constructo multidimensional. Por ello la fiabilidad se calcula más exactamente con un coeficiente basado en el análisis factorial, que usa la contribución del primer factor al total de la fiabilidad del instrumento (Barbero, 2003).

$$\theta = \frac{n}{n-1}\left(1 - \frac{1}{\lambda_1}\right) = 0{,}85$$

En la expresión anterior *n* representa el número de ítems del instrumento y $\boldsymbol{\lambda_1}$ es el primer autovalor en el análisis factorial (Tabla 4.9.4). Hemos obtenido el valor 0,85 para este coeficiente que sube respecto de Alfa y además es lo suficientemente alto para realizar un estudio de evaluación (Santisteban, 1990).

Tabla 4.9.3. KMO y prueba de Bartlett

| | | |
|---|---|---|
| Medida de adecuación muestral de Kaiser-Meyer-Olkin. | | 0,828 |
| Prueba de esfericidad de Bartlett | Chi-cuadrado aproximado | 1318,042 |
| | Gl | 190 |
| | Sig. | 0,000 |

En la Tabla 4.9.3 hacemos una prueba de Kaiser-Meyer-Olkin (KMO) y la prueba de Bartlett que comprueban los supuestos para aplicar el análisis factorial.

Tabla 4.9.4. Varianza total explicada

| | Autovalores iniciales | | | Sumas de las saturaciones al cuadrado de la extracción | | |
|---|---|---|---|---|---|---|
| Componente | Total | % de la varianza | % acumulado | Total | % de la varianza | % acumulado |
| 1 | 5,341 | 26,707 | 26,707 | 5,341 | 26,707 | 26,707 |
| 2 | 1,480 | 7,402 | 34,110 | 1,480 | 7,402 | 34,110 |
| 3 | 1,279 | 6,395 | 40,505 | 1,279 | 6,395 | 40,505 |
| 4 | 1,239 | 6,197 | 46,702 | 1,239 | 6,197 | 46,702 |
| 5 | 1,161 | 5,806 | 52,507 | 1,161 | 5,806 | 52,507 |
| 6 | 1,022 | 5,109 | 57,616 | 1,022 | 5,109 | 57,616 |
| 7 | ,959 | 4,795 | 62,411 | | | |
| 8 | ,895 | 4,476 | 66,887 | | | |
| 9 | ,843 | 4,216 | 71,103 | | | |
| 10 | ,768 | 3,839 | 74,942 | | | |
| 11 | ,733 | 3,667 | 78,609 | | | |
| 12 | ,719 | 3,597 | 82,206 | | | |
| 13 | ,687 | 3,433 | 85,639 | | | |
| 14 | ,612 | 3,061 | 88,700 | | | |
| 15 | ,540 | 2,698 | 91,397 | | | |
| 16 | ,492 | 2,458 | 93,855 | | | |
| 17 | ,482 | 2,411 | 96,266 | | | |
| 18 | ,438 | 2,192 | 98,458 | | | |
| 19 | ,202 | 1,012 | 99,469 | | | |
| 20 | ,106 | ,531 | 100,000 | | | |

Vemos (Tabla 4.9.4) que es posible extraer 6 componentes independientes del análisis con autovalor mayor que 1, y que la primera explica ya el 26,7% de la

variabilidad total, mientras que los siguientes entre el 7% y el 5%, indicando la gran importancia del primer factor. Hemos seguido también una sugerencia dada por Santisteban (1990) donde el número máximo que factores que aconseja extraer nunca supere la mitad del número total de variables menos 1.

Los porcentajes de variabilidad explicados por estas componentes destacadas otorgan una fuerte evidencia de la validez del constructo. Sabemos que no existe un acuerdo sobre el porcentaje que debe explicar la primera componente, pero sí que se debe destacar del explicado por las restantes; en nuestro caso es casi el cuádruple, mientras que el aporte de las otras cinco componentes se mantiene constante alrededor del 5% y 7%.

También con el gráfico de sedimentación de la Figura 4.3 es posible observar la gran diferencia entre el porcentaje aportado por la primera componte y el resto, además se constata la homogeneidad de los restantes cinco aportes.

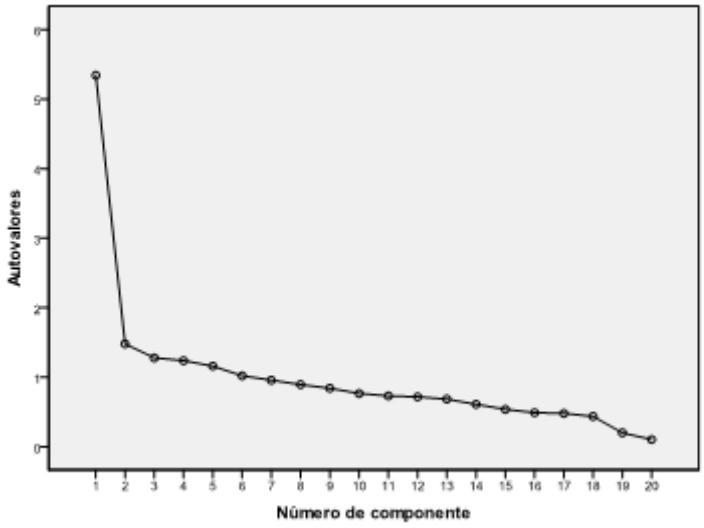



Figura 4.3. Gráfico de sedimentación

## 4.10. GENERALIZABILIDAD

Finalmente se realiza el estudio de la fiabilidad desde la teoría de la generalizabilidad, que permite por medio del análisis de varianza extender la teoría clásica de la medición para analizar diferentes fuentes de error en un proceso de medida. Se trata de generalizar la medición más allá del momento en que se hizo la medida, y eso solo se puede lograr cuando el instrumento es fiable, de ahí que este concepto sea una cualidad de la generalizabilidad que resulta una exigencia clave de la medición (León y Montero, 2002).

El cociente entre la varianza verdadera en la puntuaciones de la prueba y la observada, que es la suma de la verdadera más la debida al error resulta ser el coeficiente de generalizabilidad.

$$G = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_R^2} \quad (1)$$

La varianza del error depende de la forma en que definamos el universo de las puntuaciones verdaderas (Thorndike, 1989). En este trabajo diferenciamos dos fuentes y hemos obtenido la generalizabilidad de los mismos: generalizabilidad respecto de los sujetos (inter-personas) y la generalizabilidad de los ítems (inter-elementos). Para poder estimar las varianzas, hemos adoptado un modelo de análisis de varianzas de medidas repetidas que toma al ítem como factor fijo y a los sujetos como factor aleatorio; usamos el software SPSS con el modelo de estimación de Dunn y Clark (1987). Obtenemos las componentes de la varianza que se muestran en la Tabla 4.10.1 que estiman la varianza entre sujeto, entre ítems y la residual. Con esos elementos podremos determinar las dos medidas para la generalizabilidad (respecto a los ítems, respecto a los estudiantes).

Tabla 4.10.1. Análisis de varianza de medidas repetidas

| | | Suma de cuadrados | GL | Media cuadrática | F | Sig. |
|---|---|---|---|---|---|---|
| Inter-personas | | 203,887 | 223 | 0,914 | | |
| Intra-personas | Inter-elementos | 97,338 | 19 | 5,123 | 26,686 | 0,000 |
| | Residual | 813,412 | 4237 | 0,192 | | |
| | Total | 910,750 | 4256 | 0,214 | | |
| Total | | 1114,637 | 4479 | 0,249 | | |

Media global = 0,47

De la tercera columna de la tabla última anterior obtenemos los cuadrados medios: entre sujetos, entre los elementos y del error aleatorio, que estiman los parámetros que se indican abajo, de acuerdo al modelo de análisis de varianza que se aplico.

$CM_S$ = 0,914 que es un estimador de $b\sigma_S^2 + \sigma_R^2$, siendo $b = 20$, el número de ítems,

$CM_I$ = 5,123 que es un estimador de $a\sigma_I^2 + \sigma_R^2$ siendo $a = 224$, el número de sujetos,

$CM_R$ = 0,192 que es un estimador de $\sigma_R^2$.

De donde, despejando obtenemos las siguientes estimaciones:

Varianza estimada dentro de los sujetos, $\widehat{\sigma_S^2}$ = 0,0361

Varianza estimada dentro de los ítems $\widehat{\sigma_I^2}$ =0,022

Varianza estimada residual $\widehat{\sigma_R^2}$ = 0,192

Sustituyendo ahora estas varianza estimadas en la fórmula (1) y teniendo en cuenta los tamaños de muestra (20 ítems y 224 estudiantes), según si consideramos como fuente de variación los ítems o los estudiantes, obtenemos las siguientes estimaciones paran ambos coeficientes de generalizabilidad:

*Generalizabilidad respecto a los ítems:*

$$G_I = \frac{\sigma_s^2}{\sigma_s^2 + \frac{\sigma_R^2}{20}} = 0{,}789 \quad (2)$$

Observamos que se obtiene un coeficiente próximo en valor al coeficiente Alfa, lo que es lógico pues como ya dijimos estamos midiendo la generalizabilidad a otros ítems diferentes sobre el mismo contenido.

*Generalizabilidad respecto a los estudiantes:*

$$G_S = \frac{\sigma_I^2}{\sigma_I^2 + \frac{\sigma_R^2}{224}} = 0{,}961 \ (3)$$

De (3) vemos que es muy alta la medida de generalizabilidad a otros estudiantes de esta prueba, de donde utilizando este mismo instrumento que se ha diseñado se puede tomar a otros estudiantes similares en formación a los que se tomo la prueba con muy alta chance de generalizar nuestros resultados, claro que se debe conservar tanto las características sociológicas como educativas de los mismos.

## 4.11. CONCLUSIONES SOBRE LA EVALUACIÓN CON CUESTIONARIO

Para finalizar el capítulo exponemos a continuación nuestras conclusiones sobre los objetivos e hipótesis del Estudio 2.

### Conclusiones sobre los objetivos

El *objetivo principal del Estudio 2 es evaluar la comprensión del objeto estadístico análisis de varianza por parte de estudiantes de Psicología*.

*Un segundo objetivo es obtener evidencias de fiabilidad y generalizabilidad de la prueba, que podría ser utilizada en investigaciones posteriores.* Como vemos, se ha cumplido de sobra este objetivo. Se ha mostrado tanto la validez de contenido, de constructo y discriminante. La fiabilidad fue calculada por dos vías: el coeficiente Alfa de Cronbach (0,79) y basándonos en el análisis factorial por la multidimensionalidad del constructo que se evalúa (0,85), resultó aún más alto que el Alfa. Con este último análisis obtuvimos también seis componentes independientes, siendo que la primera explica ya casi el 30% de la variabilidad total. En cuanto a la generalizabilidad, la calculamos tanto para los ítems (0,79) como para los sujetos (0,96) resultando ambos índices muy buenos indicadores para recomendar la aplicación del instrumento a otras muestras de características similares como la posibilidad de ampliarlo con otras unidades de medida con semejanzas a los utilizados.

Respecto al primer objetivo hemos presentado los resultados detallados que pasamos a comentar con una muestra de 224 estudiantes de la Licenciatura en Psicología de la Universidad de Huelva. A la vez comentaremos las conclusiones sobre las hipótesis

### Conclusiones sobre las hipótesis

H1: Una primera hipótesis es que *se esperaba confirmar los errores y dificultades ya encontrados en la investigación previa realizada, en relación con algunos de los prerrequisitos al análisis de varianza*. Esta hipótesis se ha confirmado sólo en parte, como vemos a continuación.

*Contenido Intervalo de Confianza*: Hubo un alto índice de respuestas correctas (61,2%) en la construcción del mismo, pero se redujo casi a su tercera parte (21%) aquella que se relacionaba con su interpretación. Por otro lado hay mayor compresión del proceso de algoritmación que del de argumentación, confirmando lo denunciado tanto por Behar (2001) como por Olivo (2008) en cuanto a lo difícil que resulta al estudiante, y aún al investigador experto (Behar, 2001) interpretar los resultados obtenidos a parir de un intervalo de confianza.

*Contenido Contraste de Hipótesis*: Nuestro instrumento presenta seis ítems: los relacionados con la comprensión de las hipótesis estadísticas han tenido un altísimo índice de aciertos (93,3% y 84,8%), mejorando los resultados de Vallecillos (1994). Los relacionados con la potencia del contraste, la definición de error tipo II, el cálculo del percentil y la probabilidad de error tipo I y la regla de decisión ya sea por comparación entre zona de rechazo y percentil ó entre percentil, zona de rechazo y valor p, oscilan entre 50,9 % (potencia del contraste) y 64,7% para la probabilidad de error tipo II, que los transforman en unidades con dificultad moderada. Confirmamos lo denunciado en la investigación (Batanero, 2000; Castro Sotos *et al*., 2007 y Díaz, Batanero y Wilhelmi, 2008) que la comprensión del concepto de potencia es difícil. Tanto el que pregunta directamente por la potencia como el que pregunta por el percentil no los hemos encontrado en otros investigaciones.

H2: Una segunda hipótesis es que *se espera la existencia de dificultades a la hora de interpretar los resultados de los cálculos respecto a los contextos de los problemas planteados*.

H3. Una tercera hipótesis es que *observaremos dificultad por parte del estudiante a la hora de interpretar un resultado proveniente del cálculo que realiza un software estadístico*.

Además hemos presentado en el cuestionario doce ítems en relación al contenido de análisis de varianza, para los cuáles no tenemos antecedentes y comentamos en lo que sigue, viendo que confirman en parte las hipótesis:

*Modelo de Análisis de varianza, selección y supuestos* (4 ítems). La mayor cantidad de respuestas correctas la encontramos en la descomposición de la variabilidad total en un modelo de análisis de varianza de dos factores fijos completamente aleatorizado (69,2%), aunque el 20% no tuvo en cuenta el papel fundamental de la interacción coincidiendo con Pardo *et al.* (2007). El porcentaje más bajo (43,3%) está en relación con los supuestos de aplicación, con un porcentaje muy alto de no respuesta (41,5%). Respecto la elección de un modelo en un problema contextualizado el principal problema fue la interpretación de la diferencia entre muestras independientes y dependientes; un 50,9% que respondió exitosamente.

*Cálculo en análisis de varianza* (3 ítems)*:* resultó más difícil (57,1%) interpretar un modelo de medidas repetidas que uno de dos factores (72,3%), debido a la cantidad de cálculos que se deben cumplimentar para obtener el valor de una *F* empírica a partir de una suma de cuadrados o la media de cuadrados de alguna fuente de variación.

*Interpretación de resultados* (5 ítems); Con la misma tabla de análisis de varianza que completó para responde a otros dos ítems, se pidió observar la relación entre los valores de *F* empírica y la teórica para decidir si un factor produce o no efecto. Para responder a los últimos cuatro ítems del cuestionario pasado, se da una salida de

ordenador con SPSS, que hay que interpretar para ver cuál es el tipo de análisis que se aplicó (74,6%), si hay efecto de una de las fuentes de variación sobre la respuesta (26,8%), si se cumple la hipótesis de homocedasticidad (50,9%) según una prueba de Levene, y si esta prueba estudia lo que se pretende (35,3%). La fluctuación en estos porcentajes (casi siempre bajos, excepto un ítem) podría deberse a la dificultad del proceso interpretativo, coincidiendo con la apreciación de Olivo (2008) y Cañadas (2012) que la interpretación en inferencia es de un nivel de dificultad mucho mal alto que los cálculos.

# CAPITULO 5.

# ESTUDIO 3. ANALISIS SEMIÓTICO DE RESPUESTAS A PROBLEMAS ABIERTOS


5.1. Introducción
5.2. Objetivos del estudio semiótico
5.3. Método
5.4. Análisis a priori de los problemas propuestos
5.4.1. Análisis a priori del problema 1. Contraste de hipótesis
5.4.2. Análisis a priori del problema 2. Análisis de varianza de medidas repetidas.
5.4.3. Síntesis de contenido evaluado en los problemas abiertos
5.5. Resultados en el problema 1: Contraste de hipótesis
5.5.1. Definición de las hipótesis estadísticas
5.5.2. Elección del contraste
5.5.3. Cálculo de las regiones y decisión sobre las hipótesis
5.5.4. Interpretación de resultados
5.5.5. Cálculo de la potencia del contraste
5.5.6. Síntesis de resultados en el problema 1
5.6. Resultados en el problema 2: Análisis de varianza de medidas repetidas
5.6.1. Definición de las hipótesis estadísticas
5.6.2. Variable dependiente e independiente
5.6.3. Elección de un contraste
5.6.4. Cálculo de las regiones y decisión sobre las hipótesis
5.6.5. Interpretación de resultados
5.6.6. Síntesis de resultados en el problema 2
5.7. Comparación de resultados en los dos problemas
5.8. Conclusiones sobre los problemas abiertos


## 5.1. INTRODUCCIÓN

Como se ha indicado en los capítulos anteriores, en la prueba de evaluación pasada a los estudiantes se incluyeron algunos problemas abiertos, con la finalidad de evaluar su capacidad de argumentación y las estrategias utilizadas al resolver los problemas. Para completar el estudio de ítems de opción múltiple, en este capítulo realizamos el estudio semiótico de las respuestas de los estudiantes de la muestra en dichos problemas abiertos.

Godino, Font y Wilhelmi (2008) destacan la existencia de cinco niveles del análisis didáctico, sobre el segundo de los cuáles nos centraremos en este apartado. Este nivel de análisis se centra en los objetos y procesos que intervienen en la realización de las prácticas matemáticas y tiene como objetivo describir su complejidad onto-semiótica como un factor explicativo de los conflictos semióticos que en su realización se producen. Entenderemos por análisis semiótico de un texto matemático a su descomposición en unidades, la identificación de las entidades puestas en juego y de las funciones semióticas que se establecen (Godino, 2002, 2003).

Seguimos el método usado en investigaciones previas, por ejemplo, Alvarado (2007), Olivo (2008), Arteaga (2011), Cañadas (2012) y Gea (2014). De este modo caracterizamos de una forma más profunda, la comprensión lograda por los estudiantes de los diferentes elementos de significado del contraste de hipótesis (problema 1) y del análisis de la varianza (problema 2).

## 5.2. OBJETIVOS E HIPÓTESIS DEL ESTUDIO SEMIÓTICO

El *objetivo principal* al construir las tareas abiertas fue completar la evaluación realizada con el cuestionario y descrita en el Capítulo 4 con indicadores de la capacidad de aplicación de lo aprendido a la resolución de problemas relacionados con el contraste de hipótesis y el análisis de varianza. De este objetivo principal se deducen otros:

- O1: *El primero de ellos consiste en estimar la proporción de estudiantes en la muestra que comprenden y particularizan los objetos matemáticos incluidos en la enseñanza en un contexto de aplicación de los mismos.* Para ello, se descompone cada problema en diferentes apartados, clasificando y categorizando las respuestas en correctas, parcialmente correctas e incorrectas, que se vuelven a clasificar para mostrar la variedad de las mismas. Esta categorización nos permite analizar las estrategias y argumentos de los estudiantes, evaluando su comprensión de una forma más profunda que la llevada a cabo con los ítems de opción múltiple (Capítulo 4).
- O2: *Por otro lado, se desea identificar y clasificar los conflictos semióticos* (Godino, Batanero y Font, 2007) presentados respecto a los diferentes objetos que intervienen en el análisis de varianza y analizar en nuestra muestra la presencia de otros conflictos descritos anteriormente (Alvarado, 2007; Cañadas, 2012) respecto al contraste de hipótesis.

Todo ello permitirá completar la caracterización del significado personal de los estudiantes sobre el análisis de varianza realizada en el Capítulo 4 y compararlo con el significad institucional de referencia y el pretendido en la enseñanza (Capítulo 1).

### Revisión de la tercera hipótesis inicial

En el Capítulo 1 se formuló, de manera provisional, la siguiente hipótesis sobre los resultados que esperaríamos encontrar en el análisis cualitativo de las respuestas abiertas a los dos problemas planteados. La experiencia previa obtenida por miembros del equipo de investigación como Alvarado (2007) y Cañadas (2012) nos ayudó.

> Hipótesis 3. *El análisis semiótico de problemas abiertos permitirá explicar parte de las dificultades detectadas en la evaluación con el cuestionario en términos de conflictos semióticos.*

Más concretamente, ahora podemos desglosar esta hipótesis en las siguientes, algo más específicas:

> Hipótesis 3.1. *Esperamos explicar las dificultades de los estudiantes al plantear hipótesis estadísticas por conflictos consistentes en la confusión entre estadístico y parámetro, el contraste bilateral y unilateral o no reconocer que la hipótesis nula es puntual.*

Estas dificultades aparecen tanto en la investigación de Vallecillos (1999), como en la de Cañadas (2012) donde se denuncia que los estudiantes al contrastar hipótesis estadísticas utilizan el estadístico muestral en lugar de usar parámetros., confunden contrastes bilaterales y unilaterales o plantean hipótesis que no cubren todo el espacio paramétrico. Asimismo en la investigación de Schuyten (1991) se indica que los estudiantes confunden los estadísticos con los parámetros. Por otro lado Castro Soto *et al*. (2007) informan que los estudiantes tienen dificultad en identificar la población bajo estudio lo cual los lleva a no poder discriminar entre contrastes unilaterales y bilaterales. Seguramente estos conflictos se reproduzcan en la determinación de las hipótesis en un problema de análisis de varianza.

Hipótesis 3.2. *Pensamos que algunos estudiantes no podrán tomar decisiones correctas con sus contrastes o en el análisis de varianza ya que, tendrán dificultades para determinar tanto el estadístico de prueba como su distribución o pueden confundir los criterios para aceptar y rechazar hipótesis estadísticas*

Estos estudiantes llegan a una decisión equivocada, pues tienden a confundir el criterio de decisión al aplicar un contraste, con una prueba deductiva matemática, dificultad que también la encontramos denunciada en Vallecillos (1994). Esta dificultad se relaciona con la incorrecta comprensión de la lógica que subyace a las pruebas de hipótesis (Harradine, Batanero y Rossman, 2011). Por otro lado, pueden llegar a confundir los supuestos de aplicación de las diferentes distribuciones muestrales, como ocurrió en el cálculo de intervalos de confianza en la investigación de Olivo (2008).

Hipótesis 3.3. *Esperamos que algunos estudiantes que realicen correctamente todos los pasos del contraste de hipótesis o del análisis de varianza tengan dificultad en contextualizar los resultados, relacionándolos con la pregunta de investigación planteada.*

Nos basamos en los resultados del trabajo de Arteaga (2011) cuyos estudiantes resuelven correctamente en gran medida los pasos de un proyecto estadístico, pero se limitan a presentar finalmente los resultados matemáticos obtenidos sin contextualizar. Al igual que este autor pensamos que los alumnos fallan en el último paso del proceso de modelización que según Dantal (1997) incluye: (1) Observación de la realidad, (2) Descripción simplificada de la realidad, (3) Construcción de un modelo, (4) Trabajo matemático con el modelo, (5) Interpretación de resultados en la realidad.

## 5.3. MÉTODO

Como se ha indicado, en este capítulo se analizan dos problemas abiertos y las respuestas a los mismos por parte de los estudiantes. A continuación se justifica la elección de los problemas y su elaboración y se describe el método de análisis de los mismos.

### Elaboración de los problemas

En el proceso de elaboración de los problemas y al igual que se hizo en la construcción del cuestionario se han seguido una serie de recomendaciones psicométricas habituales (Martínez Arias, 1995; Osterlind, 1989; Thorndike, 1989):

- En primer lugar, se delimitó el contenido a evaluar con los problemas, a partir del análisis del significado de referencia llevado a cabo en el Capítulo 1 y del estudio de las dificultades previsibles de los estudiantes, identificados en el estudio de los antecedentes (Capítulo 2).
- Se seleccionaron posibles problemas, de diferentes libros de texto o de exámenes previos en la asignatura realizados en cursos anteriores. Se consultaron, entre otros los textos de Manzano (1997), Peers (2006), García (2004) y Pagano (2011).
- Con todo ello se llegó a un conjunto potencial de problemas que eran interesantes para evaluar el contenido pretendido. Se procedió a homogeneizar la redacción y a cambiar el contexto cuando éste no fuese familiar al estudiante. Asimismo se añadieron preguntas o se modificaron para lograr cubrir el contenido pretendido.
- A partir de esta colección inicial, mediante una serie de revisiones y analizando en cada una los contenidos evaluados, se seleccionaron, los que se plantearon a los estudiantes.
- Se hicieron pruebas de comprensión del enunciado, con algunos estudiantes de la misma especialidad y curso que los participantes en el estudio, modificando la redacción en los casos que fue necesario. Se tuvieron también en cuenta las siguientes recomendaciones: Evitar detalles innecesarios, relevancia de las preguntas formuladas para el estudio, nivel de lectura adecuado, brevedad, claridad y nivel apropiado de abstracción.

**Muestra y condiciones de evaluación**

La muestra que respondió a los problemas es la misma que completó el cuestionario analizado en el Capítulo 4, es decir, se trata de 224 alumnos de segundo año de la Licenciatura en Psicología de la Universidad de Huelva, que cursaban una asignatura de Análisis de Datos II.

Completaron los problemas como parte de la misma evaluación en la que respondieron el cuestionario. Se les entregó además una lista de fórmulas estadísticas para evitar tuvieran que memorizarlas. Los alumnos pudieron utilizar la calculadora. Todos respondieron con interés; agradecemos su ayuda para esta investigación.

**Análisis de contenido**

Como se ha indicado, cada uno de los problemas estaba formado por varios apartados. Una vez recogidas las respuestas de los estudiantes, que son datos textuales, se utilizó el análisis de contenido, una técnica muy utilizada en la investigación educativa (véase, por ejemplo, Cañadas, 2012 o Gea, 2014). Nosotros la aplicamos a las producciones escritas de los estudiantes.

Esta técnica se basa en la idea de que las unidades de un texto pueden clasificarse en un número reducido de categorías (Krippendorff, 1991, 2013) y permite evaluar las ideas, concepciones o conocimientos del autor del texto mediante el análisis de su discurso escrito. Se utiliza para efectuar inferencias mediante la identificación sistemática y objetiva, de las características específicas de un texto (Ghiglione y Matalón, 1991). Permite utilizar técnicas cualitativas y cuantitativas que se complementan y enriquecen mutuamente durante el desarrollo de la investigación (Zapico, 2006).

Su objetivo final es la búsqueda del significado implícito en el texto, a partir de un estudio sistemático del mismo. Para realizar el análisis, hemos recurrido a nuestro conocimiento previo sobre el tema, apoyado en la revisión de antecedentes, que se recogen en el Capítulo 2. Los pasos en este análisis fueron los siguientes:

- Una vez recogidas las respuestas escritas de los estudiantes, para cada estudiante y cada apartado de los dos problemas, analizamos los tipos de respuestas diferenciados. Para ello se efectuaron varias lecturas cuidadosas para determinar la semejanza o diferencias de respuestas al mismo apartado. Cada apartado de cada problema es la primera unidad de análisis
- Mediante un proceso cíclico e inductivo, se compararon las respuestas del mismo apartado hasta llegar a un sistema de categorías (unidad secundaria de análisis). El equipo de investigación formado por directoras y doctorando, junto con algunos colegas en caso de dudas ayudaron a la depuración de las categorías.
- Primero se clasifican según sean correctas, parcialmente correctas e incorrectas; y dentro de cada una de ellas, cuando se consideró interesante, se hacen subcategorías.
- En cada una de las categorías de respuestas encontradas, reproducimos un ejemplo realizado por un estudiante de la muestra, para el cual se realiza un análisis semiótico, dividiéndolo en unidades de análisis y destacando para cada unidad las principales funciones semióticas establecidas por el alumno.
- Paralelamente, se realizó una clasificación de los posibles conflictos semióticos observados, en las respuestas incorrectas y parcialmente correctas.
- Finalmente presentamos la tabla de frecuencias de respuestas diferentes en el apartado y para cada uno de los problemas se presenta una síntesis de los conflictos semióticos identificados.

## 5.4. ANÁLISIS A PRIORI DE LOS PROBLEMAS PROPUESTOS

A continuación se analizan dos problemas abiertos, que fueron resueltos por los estudiantes junto con el cuestionario, identificando su contenido y las respuestas esperadas de los estudiantes. Cada uno se componen de un enunciado inicial y varios apartados, en los que se realizan preguntan encaminadas a evaluar diferentes conocimientos de los estudiantes.

### 5.4.1. ANÁLISIS A PRIORI DEL PROBLEMA 1: CONTRASTE DE HIPÓTESIS

**Problema 1.** Se sabe por diversos trabajos de investigación que los niños de seis años tienen una velocidad lectora media de 40 palabras por minuto, con varianza igual a 16. Un profesor quiere saber si los niños de su clase se sitúan o no en la media de palabras por minuto. Para ello mide la velocidad de lectura en los 25 niños de su clase, obteniendo una media de 43 palabras por minuto:

1. Define las hipótesis estadísticas.
2. ¿Qué tipo de contraste de hipótesis vas a utilizar?
3. ¿Qué decisión debes tomar sobre las hipótesis, con un nivel de confianza del 95%?
4. ¿Qué puede el profesor concluir sobre la media en velocidad lectora de sus alumnos?
5. ¿Cuál es la potencia del contraste si partimos de que el valor que toma la hipótesis alternativa es igual a 42 ($H_1: \mu = 42$)?

En primer lugar se presentó un problema sobre contraste de hipótesis donde se plantea una situación práctica. Con los datos aportados deberá completar cinco apartados, donde realizará todos los pasos de un contraste de hipótesis de dos colas para una media con una población de varianza conocida. Pretendemos evaluar mediante la respuesta a esta tarea en qué medida los estudiantes son capaces de cumplimentar los siguientes puntos:

1. Identificar la hipótesis nula y la alternativa correcta respecto al parámetro poblacional presente en el problema, decidiendo si se trata de un contraste unilateral o bilateral. Plantear las hipótesis en forma adecuada y utilizando la notación conveniente.
2. Reconocer que se trata de un contraste sobre el valor medio de una variable para una sola muestra, donde la desviación estándar es conocida.
3. Diferenciar entre el valor teórico de la media poblacional o parámetro ($\mu_0 = 40$), y el valor del estadístico media muestral observado ($\bar{x} = 43$).
4. Elegir adecuadamente un estadístico de contraste; en este caso, la tipificación de la media muestral: $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$.
5. Reconocer que la variable $Z$ seguiría una distribución Normal (con desviación estándar conocida) en forma aproximada, debido al tamaño de la muestra, aplicando el teorema central del límite y teniendo en cuenta ciertas condiciones.
6. Construir correctamente la región crítica, estableciendo sus valores límites, identificando que su complemento será la región de aceptación y comprender el significado de estas regiones así como la acción que debe tomarse según el valor del estadístico calculado.
7. Tomar una decisión correcta de acuerdo con los valores calculados y con el nivel de significación propuesto en el problema.
8. Responder en términos del enunciado de la situación planteada, contextualizando su respuesta al fenómeno.
9. Determinar la potencia del contraste para un valor dado de la media poblacional bajo la hipótesis alternativa.

Seguidamente se realiza un análisis de las soluciones correctas previstas en cada apartado del problema.

**Identificación de datos**

Se espera del estudiante que identifique los datos emergentes del enunciado para poder resolver la tarea. Es decir, se espera que identifique el valor de la media y la varianza poblacional (deduzca de esta la desviación estándar), el tamaño muestral, el valor de la media muestral, reconozca la distribución Normal y el nivel de significación.

De la identificación de datos se toman los valores para calcular el estadístico de prueba, es decir la media poblacional, $\mu_0 = 40$, media muestral, $\bar{x} = 43$, la desviación estándar, $\sigma = 4$, y el tamaño muestral $n = 25$.

**Resolución del apartado 1. Planteamiento de las hipótesis**

Frente a una prueba de hipótesis, el primer paso es plantear las hipótesis nulas y alternativa. De la lectura del enunciado, y puesto que se quiere comprobar si los niños la clase tienen una velocidad lectora media de 40 palabras por minuto, se deduce que la hipótesis nula es, $H_0: \mu = 40$. Puesto que la complementaria es la hipótesis alternativa, se deduce que la hipótesis alternativa es $H_1: \mu \neq 40$ . Por tanto, las hipótesis estadísticas del contraste son: $H_0: \mu = 40, vs\ H_1: \mu \neq 40$.

**Resolución del apartado 2: Elección del tipo de contraste**

Una vez establecidas las hipótesis estadísticas, deberemos dar cuenta que se trata de una prueba bilateral, el texto dice: "….se sitúan o no…", entonces la alternativa se escribirá en función de una desigualdad que no indica dirección de la misma, involucrando solamente a una población, la de los niños de la clase del profesor.

Es un contraste sobre la media de una población, ya que se desea contestar a la pregunta de si los niños de la clase (población) se sitúan o no (prueba de dos colas o contraste bilateral) en "la media" de palabras por minuto. Del texto del problema "…varianza igual a 16…", se deduce que la varianza es conocida, y de la misma se puede identificar la desviación estándar, calculando su raíz cuadrada. De todo lo explicado, se desprende que la respuesta correcta a este apartado es: contraste bilateral para una media con $\sigma = 4$.

**Resolución del apartado 3: Decisión a nivel de confianza del 95%**

Para este apartado, debemos calcular el estadístico muestral usando los datos proporcionados; traducidos desde el lenguaje coloquial al simbólico. Sabemos que $\sigma = 4$. Además el nivel de significación de la prueba es $\alpha = 0{,}05$ ya que el nivel de confianza propuesto es del 95%, lo que genera valores críticos -1,96 y 1,96 para la distribución normal N(0,1) (Figura 1). Con esos elementos podemos definir las zonas de rechazo y aceptación de la prueba.

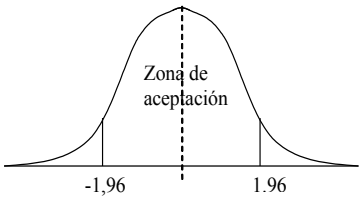



Figura 5.1. Gráfico de distribución para el apartado 3

El estadístico es: $z = \frac{43-40}{4/\sqrt{25}} = \frac{3}{4/5} = 3{,}75$ . Ya que los valores críticos son -1,96 y 1,96 por ser la distribución elegida la Normal y el nivel de significación 95%; resulta que la zona de rechazo es el intervalo real: $(-\infty; -1{,}96) \cup (1{,}96; +\infty)$. Como el valor *z* del estadístico observado cae en la zona de rechazo, más exactamente en el intervalo de la derecha, se tomará la decisión de "rechazar la hipótesis nula", con un nivel de confianza igual a 95%.

**Resolución del apartado 4: Interpretación de los resultados**

Para este apartado, debemos dar las soluciones en términos del problema, es decir

responder a la pregunta de investigación que se hace ese profesor acerca de los niños de su clase. Como decidió, de acuerdo con el apartado último anterior, rechazar la hipótesis nula, pues el estadístico de prueba ha caído en la zona de rechazo, diremos que *el profesor debe concluir que sus alumnos tienen una velocidad media de lectura que no se sitúa en la media de la velocidad de lectura conocida de palabras por minuto*, ya que hemos rechazado la hipótesis nula.

**Resolución del Apartado 5: Cálculo de la potencia**

Para calcular la potencia del contraste, simbolizado con la letra griega $\pi$, es conveniente primero determinar la probabilidad de Error tipo II, simbolizado con la letra griega $\beta$, puesto que $\pi = 1-\beta$. También se debe tener en cuenta que la $\mathrm{P(Error\ Tipo\ II)} = \mathrm{P(no\ rechazar\ Ho, siendo\ Ho\ falsa)}$ es variable, depende del valor que tome la media poblacional en la alternativa. Decidimos no rechazar dicha hipótesis nula cuando el estadístico observado de la prueba caiga en la zona de aceptación, es decir cuando $|Z| < 1{,}96$ $(\alpha = 0{,}05)$. Esto último equivale a decir que:

$$-1.96 < \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < 1{,}96 \Leftrightarrow \underbrace{-1.96 \cdot \frac{4}{\sqrt{25}} + 40}_{L_1} < \bar{X} < \underbrace{1{,}96 \cdot \frac{4}{\sqrt{25}} + 40}_{L_2}.$$

Los extremos $L_1$ y $L_2$ en la expresión anterior son $L_1$ =38,432 y $L_2$ =41,568, que se obtienen al operar los extremos en la doble desigualdad última anterior. Para calcular la potencia se necesita poner en juego un valor de la media poblacional que corresponda a la hipótesis alternativa. El propuesto en la tarea es $\mu = 42$. Determinaremos primeramente $\beta(42)$, es decir la probabilidad de no rechazar la hipótesis nula, cuando el valor de la media en la población es 42, procediendo:

$$\beta(42) = P(\,38{,}432 < \bar{X} < 41{,}568 | \mu = 42) = P\left(\frac{38{,}432-42}{4/5} < \underbrace{\frac{\bar{X}-42}{4/5}}_{Z} < \frac{41{,}568-42}{4/5} \mid \mu = 42\right)$$

$$= P(-4{,}46 < Z < -0{,}54) = P(Z < -0{,}54) - P(Z < -4{,}46) \sim 0{,}2946.$$

Finalmente, la potencia de la prueba se calculada mediante la expresión $\pi(42) = 1 - \beta(42)$, es decir: $\pi(42) = 1 - 0{,}2946 = 0{,}7054$.

## 5.4.2. ANÁLISIS A PRIORI DEL PROBLEMA 2: ANALISIS DE VARIANZA DE MEDIDAS REPETIDAS

**Problema 2.** Se ha llevado a cabo una investigación para estudiar si las técnicas de reducción de estrés tienen efecto sobre la ansiedad precompetitiva en atletas. Para ello se seleccionaron 5 atletas con alta ansiedad y les enseñaron estas técnicas. Se tomaron tres medidas de ansiedad: A1: antes de la enseñanza; A2: durante la enseñanza; A3: después de la enseñanza. Los resultados obtenidos se muestran en la tabla que sigue.

| Sujeto | A1 | A2 | A3 |
|---|---|---|---|
| 1 | 8 | 7 | 5 |
| 2 | 7 | 7 | 5 |
| 3 | 7 | 6 | 4 |
| 4 | 6 | 5 | 3 |
| 5 | 9 | 7 | 4 |

1. Define las hipótesis estadísticas que se someten a contraste.
2. Define la variable dependiente y variable independiente.
3. ¿Qué técnica estadística tienes que aplicar para comprobar el efecto de la variable independiente?
4. Suponiendo que se cumplen los supuestos necesarios para el análisis, y que el valor de α=0,05, ¿Qué decisión tomarías respecto a las hipótesis estadísticas?
5. ¿Qué conclusión extraes del análisis?

En segundo lugar se planteó un problema sobre análisis de varianza, donde se recogen medidas de ansiedad en atletas para tres momentos distintos: antes, durante y después de la enseñanza de una técnica de reducción de estrés, y se quiere saber el efecto de este factor con tres niveles sobre las respuestas. Se aportan datos para completar cinco apartados, en los que se pide realicen todos los pasos de un análisis de la varianza de un factor con medidas repetidas. Pretendemos evaluar mediante las respuestas a los diferentes apartados de esta tarea en qué medida los estudiantes son capaces de cumplimentar los siguientes puntos:

1. Identificar la hipótesis nula y alternativa de interés para realizar un contraste. Formularla en términos de los parámetros poblacionales presentes en el problema, y utilizar la notación conveniente.
2. Reconocer que el contraste se debe modelar como un problema de análisis de varianza de un solo factor con medidas repetidas (se toman diferentes mediciones de un mismo atributo sobre los mismos sujetos).El cálculo sería similar a un modelo de dos factores, donde el primero es el tiempo y el segundo el sujeto.
3. Reconozca que en este modelo se cuenta con una observación por celda. Por tanto, no es posible estudiar el efecto de la interacción. Sólo interesa el nivel de significación del factor inter sujeto (tiempo).
4. Que identifique el factor (variable independiente), constituido por las medidas repetidas de cada sujeto y la variable dependiente (ansiedad).
5. Reconocer como se descompone para este modelo la varianza, representada por la suma de cuadrados total, usando las sumas de cuadrados tanto inter sujeto, como intra sujeto, como del error, teniendo en cuenta que cada sumando representa una parte de la variabilidad total. Es decir, que comprenda la expresión:

$$SC_{Total} = SC_{Inter} + SC_{sujeto} + SC_{error}$$

6. Que elija adecuadamente un estadístico de contraste; en este caso, el estadístico de prueba $F_0$.Que pueda determinar la distribución de $F_0$ bajo la hipótesis nula y sus grados de libertad.
7. Que sea capaz de completar la tabla de análisis de varianza con sus diferentes componentes, a saber: fuentes de variación, grados de libertad, cuadrados medios y valores de los estadísticos.
8. Que logre comparar el valor obtenido de $F_0$ con el percentil de la distribución correspondiente y tome una decisión adecuada sobre la hipótesis nula.
9. Que interprete los resultados del análisis estadístico en el contexto del problema, obteniendo las conclusiones del análisis sobre la efectividad de la técnica de reducción del estrés. De este modo se completa el último paso en el ciclo de modelización.

Seguidamente se realiza un análisis de las soluciones correctas previstas en cada apartado del problema con la finalidad de identificar el significado evaluado en el mismo.

**Resolución del apartado 1: Planteamiento de las hipótesis**

Al igual que en el problema 1, el primer paso para responder a la pregunta de interés es plantear las hipótesis nulas y alternativa. Para esta apartado en las hipótesis se debe establecer las posibles conclusiones a obtener al comparar las medias poblacionales correspondientes a la ansiedad precompetitiva en los atletas antes, durante y después de la enseñanza de las técnicas de reducción al estrés.

Puesto que la hipótesis nula es siempre la hipótesis de no efecto, hay que suponer que las técnicas en ningún momento producen cambios sobre la ansiedad, entonces la respuesta media para los tres estadios de los sujetos no varía. Expresado en notación simbólica:

$H_0: \mu_{antes} = \mu_{durante} = \mu_{despues}$ ,

La hipótesis alternativa es el complemento de la nula, entonces lo expresamos en forma general, sin indicar cuál de las tres medias es diferente de las otras:

$H_1: algún\ par\ de\ medias\ son\ diferentes.$

**Resolución del apartado 2: Definición de variables**

En esta pregunta se trata de ver si el estudiante discrimina el concepto de variable dependiente (la respuesta del tratamiento) de la independiente (el factor que hacemos variar en el experimento). Como el cambio en la ansiedad va a depender de las mediciones para cada estadío, se tendrán dos tipos de variables. La variable dependiente, en este caso es "ansiedad", además podemos considerarla de tipo continua. La variable independiente o factor son los tres momentos en los que se hacen las mediciones, es decir: "antes, durante y después", la cual es de tipo categórica.

**Resolución del apartado 3: Elección de la técnica estadística**

Una vez establecidas las hipótesis y los tipos de variables en juego, debemos elegir la técnica estadística para responder. En esta pregunta se trata de ver, en primer lugar, que los estudiantes sepan diferenciar el análisis de varianza del contraste para diferencia de dos medias. Además se trata de ver si elige el modelo adecuado al tipo de problema planteado. Como se hacen mediciones de un mismo atributo sobre los mismos sujetos (son tres grupos de variables dependientes) en tres momentos distintos, la técnica requerida para comparar las medias es un análisis de varianza con medidas repetidas.

**Resolución del apartado 4: Decisión sobre la hipótesis al nivel de significación 5%**

Para tomar la decisión sobre si se debe o no rechazar la hipótesis planteada se deben completar los pasos en el análisis de varianza de medidas repetidas. Éste tipo de análisis de varianza sigue un procedimiento de cálculo similar al realizado para el caso de dos factores fijos, interpretando un factor como "tratamiento", y el otro como "sujetos" (Montgomery, 2002). Una diferencia con el modelo bifactorial es que no interesa estudiar el efecto interacción. El modelo se puede escribir del siguiente modo:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad \begin{cases} i = 1,2,\dots,a \\ j = 1,2,\dots,b \end{cases},$$

Se tiene una observación por celda (tratamiento, sujeto). Para simplificar los cálculos se aconseja confeccionar la Tabla 5.4.1, donde por filas enumeramos cada sujeto (son en total cinco los que participan). Por columnas listamos el valor observado de ansiedad en tres momentos $A_1$=Antes, $A_2$=Durante y $A_3$=Después del tratamiento, y junto a cada valor el de su cuadrado que anotamos $(A_i)^2$ (i=1,2,3).

En la columna final escribimos los totales de nivel de ansiedad observados en cada estadío por sujeto, mientras en la última fila tomamos los totales por tratamiento y los totales de los cuadrados de los valores observados. Este método permitirá utilizar algunas de las fórmulas dadas junto con el enunciado del problema.

Tabla 5.4.1. Análisis medidas repetidas. Simplificación de cálculos

| Sujeto | $A_1$ | $(A_1)^2$ | $A_2$ | $(A_2)^2$ | $A_3$ | $(A_3)^2$ | Total |
|---|---|---|---|---|---|---|---|
| 1 | 8 | 64 | 7 | 49 | 5 | 25 | 20 |
| 2 | 7 | 49 | 7 | 49 | 5 | 25 | 19 |
| 3 | 7 | 49 | 6 | 36 | 4 | 16 | 17 |
| 4 | 6 | 36 | 5 | 25 | 3 | 9 | 14 |
| 5 | 9 | 81 | 7 | 49 | 4 | 16 | 20 |
| | 37 | 279 | 32 | 208 | 21 | 91 | 90 |

Así la suma total de observaciones es 90 (columna Total), mientras que la suma total de los cuadrados de las observaciones (valores obtenidos de sumar $(A_1)^2$ , $(A_2)^2$ y $(A_3)^2$) es $\sum y_{ij}^2 = 279 + 208 + 91 = 578$; y el total de observaciones por cada nivel es 37, 32 y 21 respectivamente, hacen $\sum y_{ij} = 90$. $N = 15$, cantidad total de observaciones, $n = 5$ cantidad total de sujetos estudiados y $j = 3$ es la cantidad de momentos en que se toman las observaciones. Para calcular la $SC_{Total}$ , conviene usar la fórmula:

$SC_{Total} = \sum y_{ij}^2 - \frac{[\sum y_{ij}]^2}{N} = 578 - \frac{90^2}{15} = 578 - 540 = 38$; en $y_{ij}$ guardamos el valor de cada observación. Para $SC_{Inter}$ , usamos:

$SC_{Inter} = \frac{\sum_{j=1}^{3} t_{nivelj}^2}{n} - \frac{[\sum y_{ij}]^2}{N} = \frac{37^2+32^2+21^2}{5} - \frac{90^2}{15} = 566{,}8 - 540 = 26{,}8$; en $t_{nivelj}$ $(1 \leq j \leq 3)$ indicamos la suma de observaciones por cada nivel. Para $SC_{sujeto}$ , usamos:

$SC_{sujeto} = \frac{\sum_{i=1}^{5} t_{sujetoi}^2}{j} - \frac{[\sum y_{ij}]^2}{N} = \frac{20^2+19^2+17^2+14^2+20^2}{3} - \frac{90^2}{15} = 548{,}6 - 540 == 8{,}6$ ; en $t_{sujetoi}$ $(1 \leq i \leq 5)$ indicamos la suma de observaciones por cada sujeto.

Finalmente, la $SC_{Error} = SC_{Total} - SC_{Inter} - SC_{sujeto}$ , entonces se tiene:

$SC_{Error} = 38 - 26{,}8 - 8{,}8 = 2{,}6$.

Tabla 5.4.2. Análisis de varianza para el apartado 4

| Fuente de variación | Suma de cuadrados | Grados de libertad | Cuadrados medios | $F_0$ |
|---|---|---|---|---|
| Inter sujetos | $SC_{Inter} = 26{,}8$ | 2 | $CMI = 13{,}4$ | 41,23 |
| Sujetos | $SC_{sujeto} = 8{,}6$ | 4 | | |
| Error | $SC_{Error} = 2{,}6$ | 8 | *CME*=0,325 | |
| Total | $SC_{Total} = 38$ | 14 | | |

$F_0 = \frac{CMI}{CME}$ Tiene distribución de *F* de Fisher con 2 y 8 grados de libertad bajo la hipótesis nula, entonces el percentil correspondiente para $\alpha = 0{,}05$ es: $f_{0{,}05;2;8} = 4{,}46$

que es la *F* teórica. Finalmente como el valor observado, $F_0$ (Tabla 5.4.2) es mayor que el de la *F* teórica $(f_{0,05;2;8})$, tomamos la decisión de rechazar la hipótesis nula con un nivel de significación $\alpha = 0{,}05$.

**Resolución del apartado 5: Conclusiones obtenidas del análisis estadístico**

Del análisis estadístico concluimos que al rechazar la hipótesis nula para la igualdad de medias (antes, durante y después de la enseñanza de las técnicas) que las técnicas de reducción del estrés tienen efecto sobre la ansiedad precompetitiva en atletas. De acuerdo con esto acordamos que la ansiedad cambia cuando se enseña a los sujetos esas técnicas.

### 5.4.3. SINTESIS DE CONTENIDO EVALUADO POR LOS PROBLEMAS ABIERTOS

En los problemas abiertos hemos evaluado tanto contenido en relación con los prerrequisitos de nuestro objeto de estudio como los que están en relación directa con el análisis de varianza.

En relación con los prerrequisitos destacamos la lectura de tablas de distribución y cálculo de probabilidades, la diferencia entre muestras independientes y relacionadas. Como otro prerrequisito destacamos el intervalo de confianza y el contraste de hipótesis. En cuanto al primero evaluamos la construcción del mismo como su interpretación. Destacamos para el segundo la asignación de hipótesis, diferenciar entre hipótesis nula y alternativa, el papel que juega el valor p en la toma de decisión, su relación con el nivel de significación y la probabilidad de error tipo I, el cálculo de la potencia del contraste y su relación con la probabilidad de error tipo II. Igualmente con qué criterio decidimos usar un determinado contraste, si este resulta unilateral o bilateral, las diferencias entre las regiones críticas y de aceptación; para finalizar con la regla de decisión.

En relación con el objeto de estudio propiamente dicho: análisis de varianza destacamos los siguientes contenidos evaluados: el análisis de los supuestos de independencia, normalidad y homocedasticidad y la comprobación de los mismos, el reconocimiento del modelo lineal que se corresponde con este objeto, el cálculo del estadístico *F* empírico, la interpretación de la tabla y de gráficos de análisis de varianza, los distintos tipos de factores, el análisis de la interacción en los modelos, el estudio, la interpretación y diferencia entre modelos de efectos fijos, aleatorios, de medidas repetidas y de dos factores para finalizar con las conclusiones de modelos aplicados.

En ambos problemas destacamos sus cinco apartados con 1.1 hasta 1.5 para el primero y desde 2.1 hasta 2.5 en el segundo, aclarando en que apartado está presente cada contenido evaluado usando una x (Tabla 5.4.3) Todo ello conforma el significado evaluado en los problemas abiertos.

Tabla 5.4.3. Contenido evaluado en los problemas

| | | Problemas | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Contenido evaluado | 1.1 | 1.2 | 1.3 | 1.4 | 1.5 | 2.1 | 2.2 | 2.3 | 2.4 | 2.5 |
| PRERREQUISTOS | Lecturas de tablas de distribución y cálculo de probabilidades | | | x | | x | | | | x | x |
| | Diferenciar muestras independientes y relacionadas | | x | | | | | x | x | x | |
| | Distribución muestral. Variabilidad del estadístico en diferentes muestras | | | x | | x | | | | x | |
| | ***Intervalo de confianza*** | | | | | | | | | | |
| | Construcción | x | | x | x | x | x | | | x | x |
| | Interpretación | x | x | | x | | x | x | x | x | x |
| | ***Contraste de hipótesis*** | | | | | | | | | | |
| | Hipótesis; Asignación de hipótesis | x | | | x | x | x | | x | x | x |
| | Hipótesis nula y alternativa | x | | x | x | x | x | | x | x | x |
| | Valor p; nivel de significación, probabilidad error I | | | x | x | x | | | | x | x |
| | Potencia, probabilidad error II | | | | | | | | | | |
| | Selección de un contraste | x | x | | x | x | x | | x | | x |
| | Región crítica y de aceptación | | | x | x | x | | | | x | x |
| | Contraste unilateral y bilateral | x | x | | x | x | | | | | |
| | Regla de decisión | | | x | x | | | | | x | x |
| ANOVA[1] | Supuesto de independencia | | | | | | | x | x | x | |
| | Normalidad y homocedasticidad | | | x | | x | | | | x | |
| | Comprobación de supuestos | | | x | x | | | x | x | x | x |
| | Modelo lineal correspondiente a un análisis de varianza | | | | | | | x | | x | |
| | Cálculo del estadístico *F* | | | | | | | | | x | x |
| | Interpretación de la tabla y de gráficos de análisis de varianza | | | | | | | | | x | x |
| | Factor y sus tipos | | | | | | x | x | x | x | x |
| | Interacción | | | | | | x | x | x | x | x |
| | Modelo de efectos fijos | | | | | | x | x | x | x | x |
| | Modelo de efectos aleatorios | | | | | | x | x | x | x | x |
| | Modelo de un factor | | | | | | x | x | x | x | x |
| | Modelo de medidas repetidas | | | | | | x | x | x | x | x |
| | Modelo de dos factores | | | | | | x | x | x | x | x |
| | Conclusiones de modelos aplicados | | | | x | | | | | | x |

[1] Análisis de Varianza

## 5.5. RESULTADOS EN EL PROBLEMA 1: CONTRASTE DE HIPÓTESIS

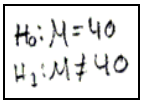
### 5.5.1. DEFINICIÓN DE LAS HIPOTESIS ESTADISTICAS

A continuación, se describen las categorías encontradas para el primer apartado del problema, en el que el alumno ha de interpretar el enunciado y formula las hipótesis estadísticas adecuadas, que se han publicado en Vera, Díaz y Batanero (2011). Presentamos, las respuestas correctas, seguidas por aquellas que, aunque incorrectas muestran una discriminación entre parámetro poblacional y estadístico muestral, finalmente por las que no lo discriminan, en ese orden. Luego se presentan y discuten las frecuencias de resultados.

**Respuestas correctas**

*C. Hipótesis estadísticas y notación correctas.* En esta categoría el alumno logra definir claramente tanto la hipótesis nula como la alternativa. Elige el parámetro adecuado sobre el cual se plantean las hipótesis (la media poblacional); usa también la notación aceptada al respecto, $\mu$. Puede diferenciar entre el valor concreto de la media poblacional (40 palabras por minuto, que se conoce de trabajos anteriores de investigación) y la media de la muestra en estudio 43. Finalmente, puede establecer que se trata de una prueba bilateral, interpretando la frase del enunciado "quiere saber si los niños de su clase se sitúan o no en la media" como indicación de que se desea saber si se puede admitir que la media de la población es o no diferente del valor hipotético, sin indicación del signo de la diferencia. En la Tabla 5.5.1 se detalla el análisis semiótico de un ejemplo.

Tabla 5.5.1. Análisis semiótico de un ejemplo en la categoría C

| Expresión | Contenido |
|---|---|
|  | - El alumno lee el enunciado e identifica correctamente el parámetro para contrastar (concepto y procedimiento).<br>- Identifica correctamente el valor hipotético del parámetro (particularización de un concepto)<br>- Traduce la expresión "se sitúan en la media" a notación matemática μ=40 (interpretación y cambio de representación); concepto "igualdad matemática".<br>- Traduce la expresión "no se sitúan en la media" a notación matemática μ≠40 (interpretación y cambio de representación); concepto "desigualdad matemática".<br>- Discrimina entre hipótesis nula y alternativa y las expresa en notación adecuada (concepto y representación); la hipótesis nula es puntual (propiedad) y es la contraria a la que el investigador quiere probar (propiedad).<br>- Reconoce en la situación un contraste bilateral (campo de problemas).<br>- Plantea una hipótesis nula y alternativa adecuada al enunciado (procedimiento). |

Como observamos en el análisis semiótico, el estudiante ha de discriminar el papel de la hipótesis nula y alternativa. El alumno ha de comprender que la hipótesis nula es puntual; y en el ejemplo, puesto que la media de la muestra es superior a 40, tomamos como hipótesis nula lo contrario de lo que esperamos mostrar; es decir, la hipótesis nula se plantea para ser rechazada (Batanero, 2000).

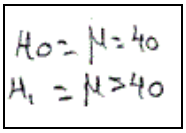
**Categorías para respuestas incorrectas (*I*) que usan el parámetro poblacional**

La discriminación entre parámetro poblacional y estadístico muestral, discriminación que no siempre se consigue según Schuyten (1991). Dentro de los que identifican correctamente el parámetro, los errores encontrados se refieren a: confundir el contraste unilateral y bilateral; confundir el problema con otro referido a dos poblaciones (tomando bien un contraste bilateral o unilateral), plantear la hipótesis alternativa como puntual, no cubrir el espacio paramétrico al definir las dos variables o confundir la variable en estudio.

*I1. Plantea la hipótesis sobre la media de la población, pero con contraste unilateral.* Estos estudiantes emplean una prueba de hipótesis de tipo unilateral, tomando la hipótesis alternativa a la derecha o hacia la izquierda de la nula. A pesar de este conflicto, usan simbología adecuada, el parámetro elegido es el correcto, y el valor sobre el cual se basa la conjetura también. En la Tabla 5.5.2, se muestra un ejemplo de respuesta de este tipo. Aunque el alumno reconoce el parámetro y lo diferencia del

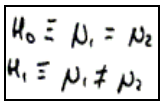
estadístico, no interpreta correctamente el enunciado, pues no considera la expresión "igual o distinto a la media" como correspondiente a un contraste bilateral. Ello le lleva a considerar una prueba unilateral a derecha, utilizando igualdad para la hipótesis nula, es decir, reconociendo que tiene que ser puntual.

Tabla 5.5.2. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
|  | – Identifica correctamente el valor hipotético del parámetro (particularización de un concepto)<br>– Traduce la expresión "se sitúan en la media" a notación matemática $\mu=40$ (interpretación y cambio de representación); concepto "igualdad matemática"<br>– Discrimina las dos hipótesis y las expresa en notación adecuada (concepto y expresión)<br>– Reconoce que la hipótesis nula es puntual (propiedad) y es la contraria a la que el investigador quiere probar (propiedad)<br>– Aparece un *conflicto* no traducir la expresión "no se sitúan en la media" a notación matemática $\mu\neq40$; ello le lleva al considerar un contraste unilateral, por fallo en la interpretación del enunciado (fallo al reconocer el campo de problemas) |

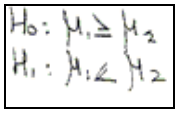
*I2. Plantea la hipótesis de diferencia de dos medias poblacionales, bilateral*. En esta categoría se encuentran los estudiantes que, habiendo dado una notación correcta, han especificado la prueba de hipótesis bilateral, confundiendo el campo de problemas planteado. Han usado como parámetros las medias poblacionales para las poblaciones supuestas. En la Tabla 5.5.3 analizamos un ejemplo. El alumno realiza una interpretación incorrecta, asumiendo que existen dos poblaciones, la hipotética y aquella de donde se tomó la muestra de alumnos. Ello le lleva a plantear un contraste de comparación de dos medias. Se observan varios conflictos, en particular una posible confusión entre población y muestra (por tanto entre estadístico y parámetro), citada por Schuyten (1991).

Tabla 5.5.3. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
|  | – El alumno realiza una interpretación incorrecta del enunciado, asumiendo que existen dos poblaciones, la hipotética y aquella de donde se tomó la muestra de alumnos (conceptos).<br>– Un posible *conflicto* es la confusión entre población y muestra. Relacionado con él, otro *conflicto* es confundir la media muestral que es un estadístico con un parámetro (conceptos).<br>– Todo ello causa un *conflicto* al plantear un contraste de diferencia entre dos poblaciones (confusión del campo de problemas).<br>– La notación sería adecuada, y el alumno asimismo discrimina la hipótesis nula como puntual y la hipótesis alternativa como aquella que quiere probar. |

*I3. Plantea la hipótesis de diferencia de dos medias poblacionales, unilateral*. Estos estudiantes, habiendo usado notación correcta, han establecido hipótesis como si se tratara de una prueba de comparación de dos muestras unilateral, tomando la hipótesis alternativa, ya sea para la izquierda como la derecha. En la Tabla 5.5.4, se analiza un ejemplo donde se repiten los conflictos descritos en la respuesta anterior y, además, el estudiante plantea una hipótesis nula no puntual.

Tabla 5.5.4. Análisis semiótico de un ejemplo en la categoría I3

| Expresión | Contenido |
|---|---|
|  | – *Conflictos:* confusión entre población y muestra; confusión entre estadístico y parámetro (conceptos)<br>– No traduce la expresión "no se sitúan en la media" al no usar la notación "≠" bajo la alternativa. Ello le lleva a considerar un contraste unilateral (confunde el campo de problemas).<br>– Otro *conflicto* es no reconocer que la hipótesis nula es puntual (propiedad). |

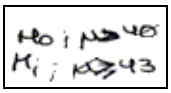
*I4. La hipótesis alternativa es puntual y las hipótesis no cubren el espacio paramétrico.* En esta categoría, también encontrada por Cañadas (2012), agrupamos los estudiantes que han elegido el parámetro adecuado para establecer la conjetura de la prueba de hipótesis, pero han escogido un valor puntual como conjunto posible de valores de la variable en la hipótesis alternativa. Por tanto, no llegan a cubrir el espacio paramétrico al establecer hipótesis. En la Tabla 5.5.5, presentamos un ejemplo donde se observan varios conflictos: además de considerar una hipótesis alternativa puntual; no reconoce que ha de ser complementaria a la nula. Toma como valor de la hipótesis alternativa el valor de la media muestral; por tanto no reconoce que las hipótesis se plantean en función de los parámetros y no de los estadísticos. Otro conflicto aparece al ordenar las hipótesis estadísticas.

Tabla 5.5.5. Análisis semiótico de un ejemplo en la categoría I4

| Expresión | Contenido |
|---|---|
| $H_0 = 40$<br>$H_1 \mu = 43$<br>$H_0 < H_1$ | – El alumno identifica correctamente el valor hipotético del parámetro (particularización de un concepto)<br>– Utiliza una notación adecuada para la hipótesis nula (lenguaje, expresión)<br>– *Conflicto:* Considera una hipótesis alternativa puntual; no reconoce que ha de ser complementaria a la nula, (propiedad).<br>– *Conflicto:* toma como valor de la hipótesis alternativa el valor de la media muestral; no reconoce que las hipótesis se plantean en función de los parámetros y no de los estadísticos (propiedad); Podría haber un *conflicto suby*acente entre población y muestra o estadístico-parámetro (conceptos)<br>– Otro *conflicto* aparece al ordenar las hipótesis estadísticas (concepto, procedimiento, propiedad). |

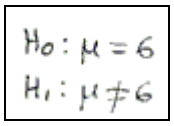
*I5. Plantea la hipótesis nula y alternativa que no cubren el espacio paramétrico.* Estos estudiantes han tomado valores posibles del parámetro que no son los correctos, ni logran cubrir el espacio paramétrico, ya que eligen intervalos no disjuntos para los valores hipotéticos del parámetro. En la Tabla 5.5.6, realizamos un análisis semiótico de un ejemplo.

Tabla 5.5.6. Análisis semiótico de un ejemplo en la categoría I5

| Expresión | Contenido |
|---|---|
|  | – El alumno identifica correctamente el valor hipotético del parámetro (particularización de un concepto).<br>– Presenta un *conflicto* al no reconocer que la hipótesis nula es puntual (propiedad).<br>– También aparece un *conflicto* al considerar dirección en la desigualdad para la hipótesis alternativa (cambio de representación).<br>– Podría haber un *conflicto* entre población y muestra o estadístico-parámetro al usar los valores 40 y 43(conceptos).<br>– De éste se deduce otro *conflicto* al no cubrir el espacio paramétrico (propiedad). |

*I6. Confunde la variable al identificar la hipótesis*. En esta categoría concentramos los estudiantes que usan notación adecuada para ambas hipótesis, el parámetro sobre el cual se desea hacer la prueba también es correcto. Se elige un contraste bilateral, pero confunde la variable, pues proponen un contraste sobre la edad de los alumnos en la muestra, que es un dato constante en el problema; por tanto no necesita ser contrastado. En la Tabla 5.5.7 realizamos un análisis semiótico de un ejemplo. No hemos encontrado este conflicto en los antecedentes.

Tabla 5.5.7. Análisis semiótico de un ejemplo en la categoría I6

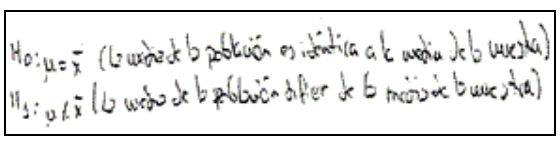
| Expresión | Contenido |
|---|---|
|  | – El alumno interpreta incorrectamente el enunciado (*conflicto*) para definir las hipótesis (proceso de interpretación),<br>– Aparece un *conflicto* al confundir la variable bajo estudio: número de palabras leídas en algún tiempo" con la edad de los participantes (fallo en la particularización del concepto variable)<br>– La notación sería adecuada, y el alumno asimismo discrimina la hipótesis nula como puntual y la hipótesis alternativa como aquella que quiere probar, reconociendo así un contraste bilateral (campo de problemas); y las hipótesis cubren el espacio paramétrico (conceptos, propiedades y procedimientos) |

**Respuestas incorrectas que usan la media de la muestra como parámetro**

Vallecillos (1994) encuentra que muchos estudiantes responden que "…*la hipótesis puede referirse indistintamente a un parámetro poblacional o a un estadístico muestral…"* (pp. 596). La autora indica que los estudiantes no alcanzan a discriminar una propiedad importante que diferencia a estos dos conceptos: que la media de la población es una constante, mientras que la media muestral es una variable aleatoria. Este error se describe asimismo en Vallecillos y Batanero (1997) y Castro Sotos *et al*. (2007). A continuación describimos algunas categorías de respuesta en que subyace este conflicto.

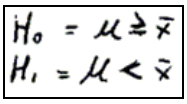
*I7. Plantea como hipótesis la igualdad entre media muestral y poblacional, en contraste bilateral.* Estos estudiantes plantean las hipótesis estadísticas igualando la media poblacional a la muestral ($\bar{x}$). En el ejemplo de la Tabla 5.5.8 el estudiante ha utilizado notación adecuada y también ha establecido correctamente que se trata de una prueba bilateral. Además, identifica que se trata de un contraste sobre el parámetro ($\mu$), pero confunde la media muestral y poblacional.

Tabla 5.5.8. Análisis semiótico de ejemplos en la categoría I7

| Expresión | Contenido |
|---|---|
|  | – El estudiante usa notación adecuada (lenguaje).<br>– Establece correctamente una prueba bilateral (campo de problemas, procedimiento)<br>– Usa lenguaje coloquial para explicar el contenido de las hipótesis (lenguaje).<br>– *Conflicto* ya que define las hipótesis relacionando la media de la población y la de la muestra (lenguaje y conceptos). |

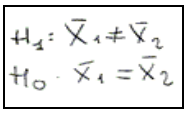
*I8. Plantea como hipótesis la igualdad entre media muestral y poblacional, en contraste unilateral*. Se utiliza una prueba unilateral. Si bien tanto los símbolos usados para la hipótesis nula y la alternativa son correctos, y el estudiante identifica el parámetro a contrastar, no utilizan el valor numérico adecuado para el planteamientos de las hipótesis. En su lugar, conjetura si el parámetro es o no igual al estadístico, comparación que no tiene sentido, pues al ser el estadístico una variable aleatoria pocas veces coincidirán exactamente estos valores (Batanero, 2000). La hipótesis planteada en estos términos, además, no permite continuar el problema, pues para fijar la distribución muestral, habrá que dar un valor numérico supuesto al parámetro, lo que no hacen estos estudiantes. En la Tabla 5.5.9, realizamos un análisis semiótico de un ejemplo.

Tabla 5.5.9. Análisis semiótico de un ejemplo en la categoría I8

| Expresión | Contenido |
|---|---|
|  | – El estudiante interpreta incorrectamente el enunciado.<br>– Plantea la hipótesis nula y alternativa, con notación adecuada (concepto y lenguaje).<br>– Aparece un *conflicto*, al establecer una prueba (discriminación de campos de problemas) pues define las hipótesis relacionando la media de la población y el de la muestra (lenguaje y conceptos).<br>– No reconoce que la media muestral es una variable aleatoria (propiedad)<br>– Otro *conflicto* es no reconocer que la hipótesis nula es puntual (propiedad). |

*I9. Plantea la hipótesis de diferencia de dos medias muestrales*. Estos estudiantes, habiendo usado notaciones correctas, han planteado hipótesis como si se tratase de una prueba de comparación de las medias de dos muestras. Además, no diferencian entre estadístico y parámetro ni aprecian que el estadístico es una variable aleatoria. En la Tabla 5.5.10, realizamos un análisis semiótico de un ejemplo.

Tabla 5.5.10. Análisis semiótico de un ejemplo en la categoría I9

| Expresión | Contenido |
|---|---|
|  | – El alumno realiza una interpretación incorrecta del enunciado, asumiendo que existen dos poblaciones (conceptos) .Escribe sus hipótesis utilizando una notación adecuada (representación)<br>– *Conflicto* al plantear un contraste de diferencia entre dos poblaciones (confusión del campo de problemas).<br>– Discrimina la hipótesis nula como puntual y la hipótesis alternativa como aquella que quiere probar (propiedades)<br>– Genera *conflicto* al plantear las hipótesis en término de un estadístico y no de un parámetro (concepto). Un posible *conflicto* es la confusión entre población y muestra (conceptos)<br>– Relacionado con el último anterior, presenta el *conflicto* consistente en confundir la media muestral que es un estadístico con un parámetro (conceptos) |

*I10. Respuestas no relacionadas con el problema*. Se contemplan en esta categoría todas las respuestas que ofrecen los estudiantes y no están relacionadas con la tarea, o cuyas respuestas son erróneas con poca incidencia, o son de muy difícil interpretación.

En la Tabla 5.5.11 presentamos los resultados en este apartado, donde sólo un tercio de los estudiantes las plantea correctamente, habiendo sido, por tanto capaces de interpretar el enunciado, identificar el campo de problemas, reconocer el valor hipotético del parámetro, escribir las hipótesis adecuadas con notación correcta. Este porcentaje es preocupante, pues este primer apartado va a condicionar el resto del problema. Además nuestro porcentaje es un poco menos alentador que el obtenido en una pregunta similar en Cañadas *et al*. (2012), que fue el 46,6%, pero se acerca más al encontrado por Vallecillos (1994) que fue de un 26% también con estudiantes de Psicología.

Tabla 5.5.11. Frecuencias (y porcentajes) de respuestas al Problema 1: Establecer hipótesis

| | Categorías | Frecuencia | % |
|---|---|---|---|
| | Correctas | 74 | 33,0 |
| Media Población | I1. Contraste unilateral | 13 | 5,8 |
| | I2. Diferencia medias poblacionales. Bilateral | 23 | 10,3 |
| | I3. Diferencia medias poblacionales. Unilateral | 5 | 2,2 |
| | I4. Hipótesis alternativa puntual | 17 | 7,7 |
| | I5. Intervalos no disjuntos; hipótesis no complementarias | 2 | 1 |
| | I6. Confusión en variable | 1 | 0,4 |
| Media muestra | I7. Compara media poblacional y muestral. Bilateral | 42 | 18,7 |
| | I8. Comparación media poblacional y muestral. Unilateral | 3 | 1,3 |
| | I9. Diferencia medias muestrales | 1 | 0,4 |
| | I10. No relacionadas con la tarea | 6 | 2,7 |
| | Sin responder | 37 | 16,5 |
| Total | | 224 | 100,0 |

No hemos encontrado para este apartado respuestas parcialmente correctas, aunque sí una variedad de respuestas incorrectas, que se han clasificado a las respuestas incorrectas en dos grandes grupos.

En primer lugar, hemos considerado aquellas para las cuales el estudiante usa el parámetro poblacional para establecer las hipótesis estadísticas (27,2% de estudiantes), estos al menos comprenden que las hipótesis se plantean en términos de parámetros y discriminan adecuadamente los conceptos de estadístico y parámetro; que unidos al 33% que dio la respuesta correcta suponen casi el 60% de los estudiantes. Otro segundo grupo menor no realiza esta discriminación y usan el estadístico muestral al establecer las hipótesis (21%). En Vallecillos (1994) hay un 18,7% que tiene esta confusión al plantear las hipótesis estadísticas, mientras que en Cañadas et al. (2012) la tiene un 16.4%.

El porcentaje total de alumnos que responden incorrectamente al apartado alcanza el 50,4%. Este alto porcentaje de respuestas incorrectas podría ser decisivo a la hora de la resolución del problema, ya que podría representar un bloqueo para los estudiantes. Hemos señalado que Vallecillos (1994) presenta un problema similar a este en su investigación con una muestra de 375 estudiantes de diversas especialidades (n=70 Psicología), en cuyo primer apartado realiza una pregunta sobre el planteamiento de la hipótesis obteniendo 41,9% de respuestas incorrectas en total. También Cañadas et al (2012) en una pregunta sobre formulación de hipótesis estadística en una prueba Chi cuadrado para estudiantes de Psicología con una muestra de tamaño 92, encuentra un 45,8% de respuestas que no son correctas. De esta forma vemos que los porcentajes se mantienen parejos con las investigaciones revisadas.

**Conflictos semióticos al establecer las hipótesis**

En el análisis semiótico de las respuestas incorrectas que se ha ido detallando anteriormente, se han encontrado los siguientes conflictos semióticos (Godino, 2002):

- *Confusión en la utilización del parámetro* (aparece en las categorías I7 a I9). Un 20,4% presenta un conflicto a la hora de determinar sobre qué parámetro se contrasta; eligen la media muestral y no la poblacional; conflicto también presente en un 12% en la investigación de Cañadas et al. (2012).
- *Confusión entre contraste unilateral y bilateral* (que aparece en las categoría I1, I3, I5, I8). Un 10,2% de los estudiantes tiene un conflicto con respecto al tipo de contraste que debe aplicar, ya que plantea hipótesis estadísticas como si la tarea se tratase de una prueba de tipo unilateral. Por tanto no discriminan las situaciones en que deben plantearse contrastes unilaterales o bilaterales (confunde el campo de problemas).
- Estos mismos alumnos *no reconocen que la hipótesis nula es puntual* (que aparece en las categorías I1, I3, I5, I8) presenta el conflicto de no usar la igualdad para definir la hipótesis nula. Este error provocará otros nuevos al momento de definir la distribución del estadístico de prueba, pues dicha distribución depende de un único valor del estadístico.
- *Hipótesis que no cubren el espacio paramétrico* (que aparece en las categoría I4, I5; en total 8,4%). Supone un conflicto al interpretar erróneamente que se pueden utilizar *hipótesis no complementarias.* Este formato para el planteo de hipótesis estadísticas la encontramos en un 19,6% en el trabajo de Cañadas *et al*. (2012).
- *Usan un contraste de comparación de dos medias* (que aparece en las categorías I2, I3, I9). 13% presenta este error que indica una confusión entre *estadístico y parámetro* pues confunden la media muestral y poblacional. El mismo conflicto se repite en los estudiantes que usan estadístico muestral en lugar del parámetro correspondiente a la media poblacional (que aparece en las categorías I7, I8 e I*9*; en total 20,5%). Ob*s*ervamos que el porcentaje que se corresponde con los estudiantes que usan el parámetro muestral para establecer las hipótesis es bastante más bajo que el de los que usan el parámetro poblacional, aunque alto. Vallecillos (1994) encuentra que un 18,7% confunde parámetro muestral con parámetro poblacional.

Muy pocos alumnos que no contestan el ítem, sólo del 16,5%, mientras que la investigación de Vallecillos no lo hace un 32% de la muestra. Esta diferencia de porcentajes podría deberse a una mejor enseñanza o bien a que en nuestro caso el cuestionario se pasó en una situación de examen, para asegurarnos que los estudiantes habían estudiado a fondo el tema.

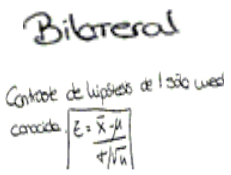
### 5.5.2. ELECCIÓN DEL CONTRASTE

En el segundo apartado del problema el estudiante decide qué tipo de contraste empleará. Este es un punto básico pues cada tipo de dato y pregunta de investigación requiere un método estadístico apropiado para responderla. Ni Vallecillos (1994) ni Cañadas (2012) realizaron esta pregunta; por tanto este apartado es un punto original de nuestro estudio.

**Respuestas correctas**

*C. Contraste de hipótesis de una sola media con desviación típica conocida, bilateral*. Los estudiantes que han sido clasificados en esta categoría, han interpretado correctamente el enunciado e identificado que, para resolver el problema, se ha de llevar a cabo una prueba de hipótesis de tipo bilateral, sobre una sola media poblacional. Además se dan cuenta que la desviación estándar de la poblaciones es conocida, es decir forma parte de la información que se brinda en la tarea. Finalmente también han especificado correctamente el estadístico de prueba que usarían para la toma de decisión. Dicho estadístico sería el valor observado de la media de la muestra, previamente tipificado. En la Tabla 5.5.12 analizamos un ejemplo.

Tabla 5.5.12. Análisis semiótico de un ejemplo en la categoría C

| Expresión | Contenido |
|---|---|
|  | – El alumno ha interpretado correctamente que trata de una prueba bilateral (campo de problemas).También que el contraste pone en juego una sola población (campo de problemas).<br>– Reconoce y aplica en este contexto los conceptos de población, media y desviación típica interpreta que la desviación estándar es conocida. Usa un símbolo adecuado (representación).<br>– Especifica el estadístico de prueba (particularización).<br>– Reconoce la media muestral (concepto) y la representa en forma correcta; además aplica la fórmula de la desviación típica de la media muestral (concepto, procedimiento, propiedad). |

**Respuestas parcialmente correctas que hacen referencia al parámetro**

Las siguientes respuestas son parcialmente correctas pues aunque identifican el contraste de hipótesis adecuado, olvidan o confunden alguna de las condiciones.

*PC1. Contraste de hipótesis de una sola media con desviación estándar conocida.* Estos estudiantes, si bien consideran correctamente que se trata de un contrataste para una sola media con desviación estándar conocida, no establecieron que se trata de una prueba bilateral. Un error similar encuentra Vallecillos (1994) observando una confusión entre región de rechazo y aceptación y en la forma como se determinan dependiendo de qué tipo de prueba se trate. En la Tabla 5.5.13 se analiza un ejemplo donde el alumno reconoce el campo de problemas al afirmar que se trata de un contraste de una media, pero no especifica que se estará dentro de los contrastes bilaterales. Además no queda claro si identifica el estadístico o su distribución.

Tabla 5.5.13. Análisis semiótico de un ejemplo en la categoría PC1

| Expresión | Contenido |
|---|---|
| *Contraste de una media con σ conocida* | - El alumno reconoce que el contraste pone en juego una sola población (campo de problemas). Reconoce y aplica los conceptos de población, media y desviación típica.<br>- No indica si se trata de una prueba bilateral o unilateral (posible conflicto semiótico respecto al campo de problemas).<br>- Interpreta que la desviación estándar es conocida. Usa el símbolo adecuado (representación).<br>- No declara la formula ni el estadístico que usaría para la toma de decisiones en el contraste, no queda claro si identifica el estadístico y su distribución. |

*PC2. Contraste de una media*. En esta categoría agrupamos a los estudiantes que solamente indican que se trata de un contraste de una media. Su respuesta es demasiado concisa, pues no indican que se trata de una prueba bilateral, o si es o no conocida la desviación estándar. En la Tabla 5.5.14, realizamos un análisis semiótico de una respuesta en esta categoría. Al igual que el caso anterior hay un posible conflicto de confusión en el campo de problemas. Además no queda claro si identifica el estadístico o su distribución pues no lo especifica.

Tabla 5.5.14. Análisis semiótico de un ejemplo en la categoría PC2.

| Expresión | Contenido |
|---|---|
| *Contraste de una media* | - Reconoce se trata de una prueba de hipótesis referida a la media.<br>- Podemos suponer que reconoce que el contraste pone en juego una sola población, al responder que se trata de "contraste de una media" (campo de problemas).<br>- No alude a la desviación estándar, no queda claro si interpreta que la misma es conocida.<br>- No declara la formula ni el estadístico que usaría para la toma de decisiones en el contraste, no queda claro si identifica el estadístico y su distribución. |

**Respuestas parcialmente correctas que no refieren al parámetro**

*PC3. Contraste de hipótesis bilateral*. En esta categoría agrupamos a los estudiantes que refieren a la respuesta del apartado solamente con la frase "contraste bilateral". En la Tabla 5.5.15, realizamos un análisis semiótico de una tarea realizada por un alumno, de este modo se pueden observar una gran cantidad de omisiones, relacionadas con campos de problemas que contienen la tarea, a saber: que la desviación estándar es conocida y que se trata de una prueba para una sola media.

Tabla 5.5.15. Análisis semiótico de un ejemplo en la categoría PC3

| Expresión | Contenido |
|---|---|
| *Contraste bilateral* | - Reconoce se trata de una prueba bilateral (campo de problemas); por tanto diferencia el contraste bilateral y unilateral (propiedad).<br>- No especifica que el contraste pone en juego alguna población, al no responder sobre que parámetro se contrasta (concepto); puede enmascarar alguna dificultad confusión entre estadístico y parámetro (conflicto).<br>- No hace alusión alguna a la desviación estándar, no queda claro si puede interpretar correctamente sobre si es o no conocida la desviación típica. |

Tabla 5.5.16. Análisis semiótico de un ejemplo en la categoría PC4

| Expresión | Contenido |
|---|---|
| *Contraste de hipótesis con σ conocida* | - El estudiante muestra un conflicto ya que no reconoce se trata de una prueba bilateral (campo de problemas).<br>- No indica el parámetro sobre el cual se realiza el contraste.<br>- Aclara que la desviación estándar es conocida. (representación-materialización-concepto, particularización). |

*PC4. Contraste de hipótesis con desviación estándar conocida*. En esta categoría agrupamos a los estudiantes que indican solamente "contraste de hipótesis con la desviación estándar conocida". No indican si se trata de una prueba para una sola media, si es una prueba bilateral o unilateral. En la Tabla 5.5.16 mostramos un ejemplo.

El estudiante podría presentar dos conflictos en relación con el campo de problemas: no reconoce se trata de un contraste bilateral, ni sobre que parámetro se hace.

**Respuestas incorrectas**

Los errores encontrados se refieren al uso de pruebas unilaterales, considerar dos o más poblaciones, o no hacer mención de ellas considerando pruebas unilaterales. Las agrupamos en las que usan un solo parámetro, más de un parámetro y ninguno.

*I1. Contraste de hipótesis de una sola media con desviación estándar conocida, unilateral*. En esta categoría situamos los estudiantes que han indicado que se trata de un contraste de hipótesis para una media, que la desviación estándar es conocida, pero han deducido que se trata de un contraste unilateral derecho. En la Tabla 5.5.17, hacemos un análisis semiótico de un ejemplo. Se desprende un conflicto al trabajar en un campo de problemas que no se ajusta a la tarea, el de los contrastes unilaterales. Con todo, reconoce que la prueba es para una media y que la desviación estándar es conocida.

Tabla 5.5.17. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
| *"Contraste de hipótesis para una media con sigma conocida. Contraste unilateral derecho"* | - El alumno ha interpretado correctamente el enunciado y reconoce que el contraste pone en juego una sola población (campo de problemas).<br>- Reconoce y aplica en este contexto los conceptos, media y desviación típica.<br>- El alumno hace una interpretación incorrecta del enunciado, asumiendo un "…contraste unilateral derecho…" (*conflicto* entre campo de problemas).<br>- Aclara que la desviación estándar es conocida (particularización correcta de un concepto) materializa su conocimiento, no usa símbolos, solo la describe.<br>- No especifica el estadístico de prueba que ocupará en la solución. |

*I2. Contraste de dos medias relacionadas*. Del análisis semiótico del ejemplo en esta categoría se desprende esencialmente un conflicto al trabajar en un campo de problemas que no se ajusta a la tarea, el de los contrastes unilaterales. Este conflicto es esencial al momento de tomar decisión pues en ese concepto de basan las zonas de rechazo y aceptación de la prueba. Con todo, reconoce que la prueba es para una media y que la desviación estándar es conocida. A partir del análisis realizado en Tabla 5.5.18, se desprende otro conflicto, que las muestras que comparará estarían relacionadas, como si los datos viniesen por pares. Por tanto hay un problema en la comprensión de la idea de independencia (también aparece en Cañadas, 2012).

Tabla 5.5.18. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
| *"Contraste de medias relacionadas"* . | - El alumno aquí genera un *conflicto*, ya que con la sentencia "contraste de medias", hace referencia a la comparación de dos muestras (conflicto en la identificación del campo de problemas).<br>- Posiblemente ese *conflicto* enmascare otro, consistente en confundir estadístico (media muestral) con parámetro (media poblacional),<br>- Con la palabra "relacionadas" aparece otro *conflicto*, pues hace referencia al menos a dos variables de un solo grupo (propiedad y concepto). |

*I3. Análisis de varianza*. En esta categoría agrupamos a los alumnos que han respondido que utilizarían un modelo análisis de la varianza. Este método sería apropiado si en el problema hubiese presentes tres o más variables para comparar, pues esa es una condición mínima para pensar en aplicar este modelo estadístico (además de independencia, normalidad y varianzas estadísticamente iguales). En la Tabla 5.5.19 realizamos un análisis semiótico de un ejemplo. Del análisis llevado a cabo emergen dos conflictos: uno el modelo elegido (análisis de la varianza) que no es el adecuado, unido a éste, enfatiza sobre la presencia de dos factores para estudiar (propiedades).

Tabla 5.5.19. Análisis semiótico de un ejemplo en la categoría I3

| Expresión | Contenido |
|---|---|
| *"Anova de 2 factores"* | - El estudiante realiza una interpretación inadecuada del enunciado, generando un conflicto al afirmar que la prueba a aplicar es un análisis de la varianza (procedimiento), no reconoce que del problema se desprende solo la presencia de una variable (campo de problemas).<br>- Otro conflicto aparece con la frase "2 factores" ya que supone habría dos variables independientes (concepto). |

*I4. Contraste unilateral*. En esta categoría hemos clasificado a todos los estudiantes que sobre la tarea han respondido se trata de una prueba de tipo unilateral sin indicar el parámetro en estudio. En la Tabla 5.5.20 hemos realizado un análisis semiótico de la tarea resuelta por uno de los estudiantes a modo de ejemplo. Emerge fundamentalmente un conflicto, pues enfoca la tarea sostenida en un campo de problemas que no es el correcto, ya que al expresar en el problema "…se sitúan o no…"se deduce se trata de un contraste bilateral. No reconoce que se trata de una prueba de comparación de una media de una población, ni que se conoce la varianza.

Tabla 5.5.20. Análisis semiótico de un ejemplo en la categoría I4

| Expresión | Contenido |
|---|---|
| *Unilateral izquierdo* | - El estudiante presenta un *conflicto*, ya que afirma se trata de una prueba unilateral izquierda, no reconoce en la frase "…saber si los niños de su clase se sitúan o no en la media de palabras por minuto…"que la prueba es de dos colas (confusión en campo de problemas).<br>- No indica cual es el parámetro bajo estudio (*conflicto* para la particularización de un concepto). |

Tabla 5.5.21. Análisis semiótico de un ejemplo en la categoría I5

| Expresión | Contenido |
|---|---|
| *Potencia del contraste* | - El estudiante muestra una confusión entre los conceptos de contraste, potencia y muestra; posiblemente quiere indicar se trata de un contraste de dos poblaciones (conceptos).<br>- No indica un parámetro bajo prueba (fallo en la particularización de un concepto).<br>- No parece identificar los datos que emergen de la tarea: varianza conocida, contraste bilateral (conflicto en la interpretación). |

*I5. Potencia de contrastes*. Hemos clasificado en esta a los estudiantes que han respondido al apartado indicando que hay que realizar un contraste de potencia. En el ejemplo de la Tabla 5.5.21 vemos que la respuesta dada por este alumno no es la

adecuada. Su conflicto esencial se basa en no reconocer que la potencia de una prueba de hipótesis es una probabilidad, que específicamente se calcula cada vez que se quiere controlar el porcentaje de veces que se detecta la hipótesis alternativa (para un valor dado fijo del parámetro) cuando esta es cierta. Además no identifica los datos del problema.

Para resumir este apartado presentamos la Tabla 5.5.22. Hemos encontrado pocas respuestas totalmente correctas (en porcentaje mínimo 1,3%), que probablemente indiquen dificultad de verbalización de los conceptos, pues posteriormente los alumnos continúan resolviendo el problema y, aunque no muchos, como veremos en el siguiente apartado son alrededor de un 20% los que logran finalizar correctamente y bien expresado el contraste. Deducimos que algunas de las respuestas parcialmente correctas han llegado a identificar el tipo de contraste correcto, pero lo expresan incorrectamente en sus respuestas (conflicto de representación).

Tabla 5.5.22. Frecuencias (y porcentajes) de respuestas al Problema 1: Elegir contraste

| Parámetro | | Categorías | Frecuencia | % |
|---|---|---|---|---|
| | C. | Correcta y completa | 3 | 1,3 |
| Indican | PC1. | De una media con $\sigma$ conocida | 46 | 20,5 |
| uno sólo | PC2. | De una media | 43 | 19,2 |
| | PC3. | Bilateral | 4 | 1,8 |
| | PC4. | Con $\sigma$ conocida | 1 | 0,4 |
| | I1. | De una media , $\sigma$ conocida, unilateral | 2 | 0,9 |
| Más de uno | I2. | De dos medias | 41 | 18,4 |
| | I3. | ANOVA | 1 | 0,4 |
| No indican | I4. | Unilateral | 1 | 0,4 |
| | I5. | Potencia de contrastes | 12 | 5,4 |
| | I6. | No relacionados con la tarea | 13 | 5,8 |
| | S/R | | 57 | 25,4 |
| Total | | | 224 | 100,0 |

El porcentaje de respuestas parcialmente correctas llega casi a 42 puntos, clasificadas en las que refieren al parámetro sobre el que se hace la prueba y las que no. La mayoría (PC1 y PC2) (39,75% del total) identifica el parámetro sobre el que se hace el contraste, mientras que es muy bajo el porcentaje en PC3 y PC4 (2,2%) el que no lo identifica. Un alto porcentaje (41%, al contabilizar el porcentaje de correctas) reconoce que un contraste de hipótesis para este tipo de problemas debe referir siempre a un parámetro de al menos una población. El porcentaje de respuestas incorrectas (31,3%) se divide entre los que usan un solo parámetro, más de un parámetro y no se refieren al parámetro cuando responden. También las hemos clasificado en las que consideran el problema como contraste de una sola media, de dos medias (son los que más contribuyen al porcentaje total) y los que suponen se comparan más de dos medias.

**Conflictos semióticos en la elección del contraste**

En el análisis semiótico de las respuestas a este apartado se han encontrado los siguientes conflictos, al elegir un contraste adecuado al problema:

- *No reconoce que se trata de una prueba bilateral* (que aparece en las categorías PC1, PC2 y PC4, además de en I1 hasta I5, en total 67% de los estudiantes). Por tanto supone una confusión entre los conceptos de prueba unilateral y prueba bilateral. También aparece en las categorías I1 e I4; en total 1,3% de los estudiantes. Como

hemos visto este conflicto ya apareció en el planteamiento de las hipótesis y en Vallecillos (1994).

- *No reconoce que la desviación estándar es conocida* (aparece en las categorías PC2 y PC3, además de en I2, I4 e I5; en total un 45,2% de estudiantes). Por tanto se *confunde la desviación típica de la población y de la muestra*, lo que supone un conflicto de *discriminación entre estadístico y parámetro,* ya señalado anteriormente, al igual que por Schuyten (1991).
- *No identifica el parámetro sobre el que se contrasta* (aparece en las categorías PC3 y PC4, además de I3 hasta I5, en total en 8,4% de estudiantes). Hay un conflicto en la interpretación de datos del enunciado.
- *Considera que hay más de una población presente en el problema* (aparece en un 18,8% de estudiantes, distribuidos en las categorías I2 e I3). Subyace una *confusión entre población y muestra.*

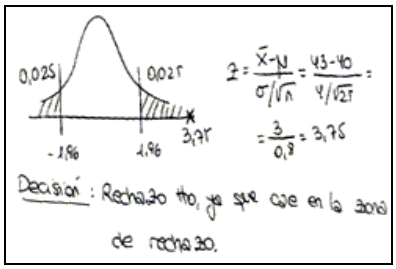
### 5.5.3. CÁLCULO DE LAS REGIONES Y DECISIÓN SOBRE LAS HIPÓTESIS

En esta sección analizamos las categorías de respuestas al tercer apartado del problema, donde el estudiante he de calcular las regiones críticas y de aceptación y tomar una decisión acerca de, rechazar o no la hipótesis nula, de acuerdo con un nivel de confianza también establecido en el enunciado.

#### Respuestas correctas

*C. Calcula correctamente el estadístico, el percentil, y las regiones y rechaza la hipótesis nula*. Estos estudiantes que determinan el valor correcto del estadístico, leen correctamente el percentil de la tabla de la distribución normal, realizan una representación gráfica ajustada a la tarea determinando las zonas de rechazo y aceptación y toman la decisión correcta de rechazar la hipótesis nula. Para ello debe distinguir entre el valor de la media muestral y la poblacional, además de la desviación estándar y el tamaño muestral.

Tabla 5.5.23. Análisis semiótico de un ejemplo en la categoría C

| Expresión | Contenido |
|---|---|
|  | - El alumno lee el enunciado (proceso de interpretación) y elige correctamente la fórmula para el estadístico (procedimiento).<br>- Reemplaza los valores de la media muestral y la poblacional (particulariza los conceptos de parámetro y estadístico).<br>- Distingue la varianza poblacional de la desviación estándar (conceptos) y reemplaza su valor en la fórmula (particulariza un concepto).<br>- Calcula correctamente el estadístico y determina correctamente los percentiles (procedimientos)<br>- Grafica la densidad ubicando tanto los percentiles como el estadístico de prueba (cambio de representación).<br>- Reconoce la zona de rechazo (aplica una propiedad)<br>- Toma en consecuencia la decisión de rechazar la hipótesis nula, argumentando que el estadístico de prueba cae en la zona de rechazo de rechazo (argumento). |

Deben particularizar correctamente todos estos conceptos y asignarle el valor adecuado. Debe completar los cálculos y finalmente ver a qué zona (aceptación o rechazo) pertenece el valor calculado del estadístico. En general, los alumnos se ayudan con una gráfica para tomar la decisión final, donde se especifican los percentiles y el valor del estadístico. En la Tabla 5.5.23, efectuamos un análisis semiótico de una respuesta dada por un estudiante de la muestra para esta tarea.

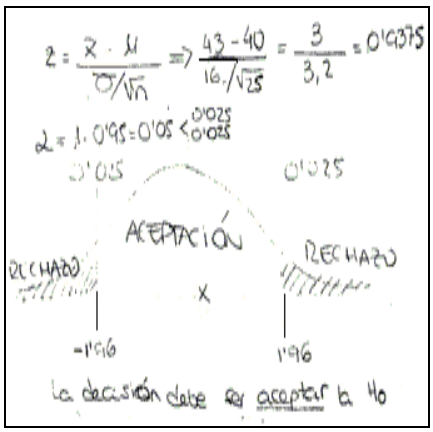
Observamos la gran complejidad del razonamiento, pues el alumno finalmente ha de reconocer que la región de rechazo le sirve para rechazar la hipótesis nula. En el ejemplo, el alumno marca gráficamente las dos regiones y utiliza correctamente otra serie de representaciones simbólicas, de la operación de tipificación, la distribución normal y los percentiles.

**Respuestas parcialmente correctas**

La gran complejidad de la tarea lleva a algunos alumnos a fallar en alguno de los pasos descritos, obteniéndose los siguientes tipos de respuesta:

*PC1 Cálculo correcto del percentil con estadístico incorrecto, dibuja las regiones y toma decisión en consecuencia*. En esta categoría los estudiantes, tras haber tenido un fallo en el cálculo del estadístico (ya sea por confundir varianza con desviación estándar o intercambiar media poblacional con muestral, etc.) han encontrado correctamente el percentil de acuerdo con el nivel de significación propuesto para la tarea.

Tabla 5.5.24. Análisis semiótico de un ejemplo en la categoría PC1.

| Expresión | Contenido |
|---|---|
|  | - El alumno lee el enunciado (proceso de interpretación) y elige correctamente la fórmula para el estadístico (procedimiento).<br>- Reemplaza los valores de la media muestral y la poblacional (particulariza conceptos).<br>- No distingue la varianza poblacional de la desviación estándar presentando un *conflicto* (conceptos).<br>- Calcula correctamente el estadístico (procedimiento). No llega al resultado correcto debido al *conflicto* indicado.<br>- Determina en forma correcta los percentiles de la distribución normal de acuerdo con el nivel de confianza propuesto (particularización de un concepto)<br>- Grafica la densidad ubicando tanto los percentiles como el estadístico de prueba (cambio de representación).<br>- Reconoce gráficamente la zona de rechazo y aceptación (aplica una propiedad)<br>- Toma en consecuencia la decisión de aceptar la hipótesis nula (procedimiento), mostrando gráficamente (forma de representación) que el estadístico de prueba cae en la zona de aceptación (propiedad). |

También realizaron una interpretación gráfica adecuada, mostrando las zonas de rechazo y aceptación del contraste, y finalmente han tomado la decisión de acuerdo con el resultado encontrado al calcular el estadístico. Cuando el error se manifiesta en la no determinación de la desviación estándar, el contenido encontrado para el estadístico cae en la zona de aceptación, lo que lleva a una decisión equivocada. En la Tabla 5.5.24

realizamos un análisis semiótico, destacando los conflictos que consisten en confundir varianza y desviación típica en la fórmula de cálculo del estadístico.

*PC2. Contraste unilateral, con cálculo correcto del estadístico, cálculo incorrecto del percentil, dibuja las regiones correctamente y, toma la decisión al respecto*. Estos estudiantes han realizado un cómputo correcto para el estadístico de la prueba, un gráfico adecuando, pero han determinado en forma errónea el percentil por haber supuesto que se trataba de un contraste unilateral en el apartado anterior. Para la respuesta presentada el estudiante elige un contraste unilateral izquierdo, a pesar de un correcto cálculo del estadístico, resulta en la zona de aceptación. En la Tabla 5.5.25, hacemos un análisis semiótico de un ejemplo.

Tabla 5.5.25. Análisis semiótico de un ejemplo en la categoría PC2.

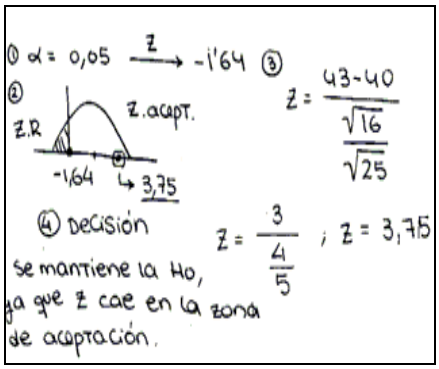
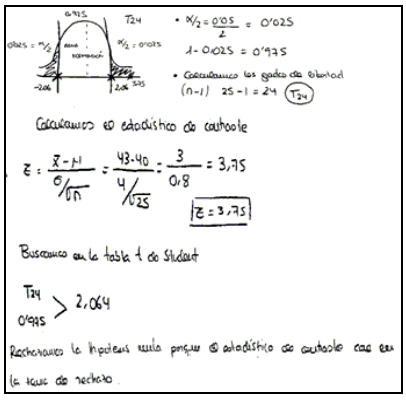
| Expresión | Contenido |
|---|---|
|  | - Lee el enunciado (proceso de interpretación) y elige correctamente la fórmula para el estadístico (campo de problemas).<br>- Reemplaza los valores de la media muestral y la poblacional (particulariza al problema).<br>- Calcula correctamente el estadístico (procedimiento).<br>- No considera una prueba de dos colas, presentando un *conflicto* al considerar una de cola inferior (campo de problemas).<br>- Determina correctamente el percentil de acuerdo al campo de problema elegido (particularización).<br>- Grafica la densidad ubicando tanto los percentiles como el estadístico de prueba (cambio de representación).<br>- Reconoce gráficamente la zona de rechazo y aceptación (aplica una propiedad)<br>- Toma en consecuencia la decisión de aceptar la hipótesis nula, mostrando gráficamente (forma de representación) que el estadístico de prueba cae en la zona de aceptación (propiedad). |

Tabla 5.5.26. Análisis semiótico de un ejemplo en la categoría PC3

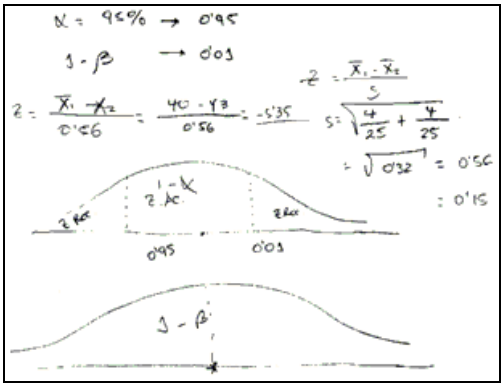
| Expresión | Contenido |
|---|---|
|  | - Interpreta el enunciado (proceso de interpretación) elige una fórmula incorrecta para el estadístico; aparece un *conflicto* (campo de problemas) ya que utiliza una distribución que no se corresponde con ese estadístico.<br>- Son correctas las expresiones de la media muestral y poblacional y la desviación estándar (particulariza al problema los conceptos).<br>- Calcula correctamente el estadístico (procedimiento).<br>- Grafica la densidad ubicando los percentiles y estadístico de prueba (cambio de representación). Da cuenta que realiza una prueba de dos colas (campo de problemas).<br>- Determina correctamente el percentil (procedimiento).<br>- Reconoce gráficamente la zona de rechazo y aceptación (aplica una propiedad).<br>- Rechaza la hipótesis nula, mostrando gráficamente que el estadístico de prueba cae en la zona de rechazo (conceptos, propiedad). |

*PC3. Confunde la distribución normal con la distribución t*. Los estudiantes que han sido clasificados en esta categoría han confundido la distribución a utilizar para la resolución, probablemente no han tenido en cuenta que la desviación estándar es conocida. Han utilizado la distribución t de Student en vez de la distribución Normal. Pese a ello han podido determinar correctamente los grados de libertad, los percentiles y el estadístico de prueba, finalmente han tomado una decisión en consecuencia. En la Tabla 5.5.26 analizamos un ejemplo.

**Respuestas incorrectas**

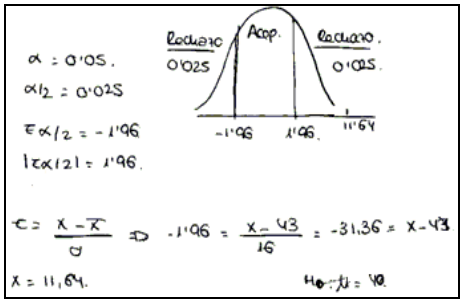
*I1. Incorrecto el estadístico y percentil, no toma decisión.* Estos estudiantes han elegido o calculado mal el estadístico, y o el percentil y no han tomado una decisión acorde con los cálculos. En esta categoría, se involucra también en la respuesta la probabilidad de error tipo II, $\beta$. En la Tabla 5.5.27 hacemos un análisis de un ejemplo, mostrando los conflictos que a pesar de marcarse correctamente la zona de rechazo de la prueba, no se determinan los percentiles ni se llega a una decisión. Una explicación es que algunos estudiantes no comprenden la lógica del contraste de hipótesis y confunden los criterios para aceptar o rechazar una hipótesis (Vallecillos, 1994).

Tabla 5.5.27. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
|  | - Genera un *conflicto*, ya que elige el estadístico como si fuese prueba de comparación de dos medias (campo de problemas).<br>- Distingue la varianza y desviación estándar (concepto), pero no distingue media poblacional y muestral, aparece así un *conflicto* (particularización).<br>- Interpretación correcta para una prueba de dos colas (campo de problemas); lo vemos en el gráfico al indicar las zonas de rechazo posibles (propiedad-representación).<br>- *Conflicto* al involucrar como parte de su respuesta el error de tipo 2, β(uso inadecuado de una propiedad).<br>- Aparece otro *conflicto* con la ausencia de los percentiles en su respuesta (particularización).<br>- No toma una decisión con la prueba resuelta (procedimiento no completo- argumentación). |

*I2. No toma una decisión, a pesar de realizar cálculos al respecto*. Los estudiantes enmarcados bajo esta categoría han calculado o elegido incorrectamente el estadístico, han encontrado correctamente el percentil, pero no han llegado a tomar una decisión. En el trabajo de Cañadas (2012) el 32,7% de estudiantes no fueron capaces de tomar una decisión en un contraste Chi- cuadrado. En algunos casos han realizado una interpretación gráfica adecuada, marcando zonas de rechazo y aceptación, percentiles y el resultado del contenido asignado al estadístico. En la Tabla 5.5.28 analizamos un ejemplo.

Tabla 5.5.28. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
|  | – En la interpretación se genera un *conflicto,* ya que elige el estadístico como si fuese prueba de comparación de dos medias (campo de problemas).<br>– Determina correctamente los percentiles, las zonas de rechazo y aceptación a través de una gráfica (representación). Interpretación correcta para una prueba de dos colas (campo de problemas).<br>– *Conflicto* al seleccionar un estadístico de prueba inadecuado (campo de problemas).<br>– Distingue entre la media poblacional y la muestral al utilizar como contenido de $\bar{x}$ el valor 43 (discriminación adecuada de conceptos).<br>– No diferencia entre varianza y desviación estándar, aparece este *conflicto* (propiedad-definición).<br>– Genera un *conflicto* al utilizar como estadístico de prueba el límite inferior de un intervalo de confianza para la media poblacional (confusión de conceptos). $-1{,}96 * 16 + 43 = 11{,}64$. |

Como resumen de este apartado presentamos la Tabla 5.5.29. Hemos encontrado un 20,5% de respuestas totalmente correctas, un porcentaje bastante alto, teniendo en cuenta la complejidad semiótica de la misma. Además igualmente se confirma que parte de las respuestas parcialmente correctas en el apartado 2 se deben a dificultad de verbalización pues si hubiesen elegido allí un contraste incorrecto, ese error se hubiese arrastrado en este apartado.

Tabla 5.5.29. Frecuencias (y porcentajes) de respuestas al Problema 1: Regiones y decisión

| Distribución | Categorías | Frecuencia | % |
|---|---|---|---|
| Normal | Correctas | 46 | 20,5 |
| | PC1. Estadístico incorrecto con percentil correcto | 68 | 30,4 |
| | PC2. Contraste unilateral y percentil incorrecto | 6 | 2,7 |
| *t*-Student | PC3. Usa la distribución *t* de Student | 5 | 2,2 |
| Normal | I1. Incorrecto estadístico y percentil. Toma o no decisión | 30 | 13,4 |
| | I2. No toma decisión | 9 | 4 |
| | I3. No relacionados con la tarea | 9 | 4 |
| | S/R | 51 | 22,8 |
| Total | | 224 | 100,0 |

Hay bastantes parcialmente correctas (35,3%), entre las cuales un 33,1% usan la distribución normal en el procedimiento, mientras que un 2,2% usa la distribución *t*. En cuanto a las respuestas incorrectas las hemos encontrado en un 21,4%, en todos los casos cuando utilizan alguna distribución, en su procedimiento y argumentación, deciden usar la normal. Sin embargo no llegan a una decisión correcta, bien por error de cálculo o porque confunden la lógica del proceso, al igual que ocurrió en la investigación de Vallecillos (1994). Igualmente en el trabajo de Cañadas (2012) el 32,7 de estudiantes no fueron capaces de tomar una decisión en un contraste Chi- cuadrado. En nuestro caso el porcentaje de respuestas correctas es menor que en las

investigaciones anteriores, pero hemos englobado varios pasos y no sólo la toma de decisión.

**Conflictos en el cálculo del estadístico y toma de decisión**

En el análisis del apartado podemos resumir se han encontrado los siguientes conflictos:

- *Confusión entre varianza y desviación estánda*r en la fórmula del estadístico, que puede ser debida a confusión en los conceptos (que aparece en las categorías PC1, PC3 e I2), en total 35,6% de los estudiantes. Este mismo conflicto apareció en la investigación de Olivo (2008) durante el cálculo de intervalos de confianza.
- *Confusión entre contraste bilateral y unilateral* que ya apareció en los apartados anteriores, (para las categorías PC2, I1 e I2) en total 20,1% de estudiantes) y como hemos dicho, en las investigaciones de Vallecillos (1994) y Cañadas (2012).
- *Confusión en la elección del estadístico o de su distribución* (aparece en I1, PC3, I2) en un 29,6% de estudiantes. Dicha confusión implica que los estudiantes no comprenden los supuestos de aplicación de las diferentes distribuciones que han estudiado a lo largo del curso.
- No es capaz de tomar una decisión final sobre la hipótesis nula, (aparece en I1 hasta I3) en un 21,4% del total de estudiantes. Puede subyacer una confusión entre región de aceptación y rechazo que implica no comprender la lógica del contraste de hipótesis y que se dio igualmente en Vallecillos (1994) y Cañadas (2012).
- *Confusión* en el cálculo de los percentiles para determinar la región de aceptación y rechazo (aparece en I1), en total 13,4% de estudiantes. Puede ser debida a dificultad de lectura de la tabla de la distribución normal y también aparece en la investigación de Olivo (2008).

### 5.5.4. INTERPRETACIÓN DE RESULTADOS

Describimos ahora las respuestas encontradas para el cuarto apartado del problema, en el cual el estudiante he de interpretar los resultados obtenidos en términos del fenómeno que se estudia. El estudiante ha de completar el último paso en la modelización según Henry (1997), que con frecuencia resulta difícil para los estudiantes y que pocos alumnos lograron completar en la investigación de Arteaga (2011).

**Respuestas correctas**

*C. Realizan una interpretación correcta de los resultados, argumentando por qué los alumnos de la clase del profesor no se sitúan en la media de palabras por minuto*. Es decir, los estudiantes han completado el último paso de la modelización, consistente en traducir las implicaciones que se deducen del trabajo con el modelo matemático a la realidad de donde surgió el problema. En esta categoría hemos agrupado a los estudiantes que dan una argumentación precisa a partir de rechazar la hipótesis nula. Indican cuál es la conclusión que el profesor puede obtener con respecto al comportamiento de sus alumnos respecto al nivel de lectura. En la Tabla 5.5.30, se analiza un ejemplo.

Tabla 5.5.30. Análisis semiótico de una respuesta correcta

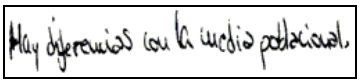
| Expresión | Contenido |
|---|---|
| | – El estudiante argumenta qué decisión debe tomar (proceso de argumentación con símbolos, $H_0$ adecuados).<br>– Reconoce que la decisión debe ser rechazar la hipótesis nula (concepto).<br>– Traduce la decisión al contexto del problema: los alumnos no se sitúan en la media de palabras (contextualización de resultados). |

**Respuestas parcialmente correctas**

Hemos encontrado tres categorías; en la primera interpretan los resultados de su análisis en términos matemáticos, sin contextualizarlos; en la segunda y tercera, aunque contextualizan los resultados, interpretan el contraste como unilateral en una, en otra han llegado a no rechazar la hipótesis nula, debido a errores en los pasos anteriores.

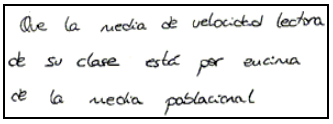
*PC1. Aunque se interpreta correctamente que no hay diferencia de medias, no se relaciona la respuesta con el contexto del problema.* En esta categoría agrupamos a los estudiantes que si bien responden en términos de rechazar la hipótesis nula, no interpretan ese resultado en el contexto del problema; por tanto no completan el último paso del proceso de modelización (Henry, 1997). No aparece en la respuesta la conclusión que debe sacar el profesor con respecto al nivel de lectura de los alumnos de su clase comparados con los de la población. En la Tabla 5.5.31 se analiza un ejemplo.

Tabla 5.5.31. Análisis semiótico de un ejemplo en la categoría PC1.

| Expresión | Contenido |
|---|---|
|  | – El estudiante reconoce que debe rechazar la hipótesis nula (propiedad, está en la zona de rechazo).<br>– Sin embargo, no es capaz de contextualizar. Aparece un *conflicto* ya que no utiliza el fenómeno presente en la tarea ("*el estudio del nivel de lectura de un grupo de estudiantes).* |

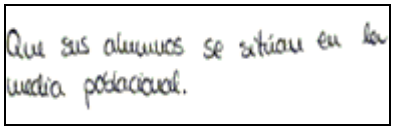
*PC2. Interpreta el contraste como unilateral y hace una interpretación correcta para este.* Decidimos agrupar en esta categoría a todos los estudiantes que respondieron a la tarea como si fuese un contraste unilateral, es decir han rechazado la hipótesis nula y, dieron una interpretación parcial en el contexto del problema. Como vimos en el apartado 1 algunos alumnos plantearon contrastes unilaterales, quizá porque el valor de la media muestral ha influenciado. Por ello ahora hacen una interpretación consistente, pero es parcialmente correcta para el problema planteado. En la Tabla 5.5.32, analizamos un ejemplo.

Tabla 5.5.32. Análisis semiótico de un ejemplo en la categoría PC2

| Expresión | Contenido |
|---|---|
|  | – El estudiante realiza el contraste en el que decide rechazar la hipótesis nula (proceso de interpretación).<br>Concluye que cae en la zona de rechazo de la prueba (propiedad) usando los términos del problema que se estudia, es decir, toma la decisión correcta.<br>– Al haber planteado un contraste unilateral (*conflicto)* interpreta el rechazo al decir *"está por encima"* confunde el campo de problemas, responde como prueba unilateral derecha. |

*PC3. Interpretan que la media de la población es igual a la de la muestra*. En esta categoría hemos incluido a todos los estudiantes que, aunque contextualizan sus resultados en términos del problema, completando el ciclo de modelización, concluyen que los alumnos de su clase tienen igual velocidad lectora que los de la población. Ello se debe a que están asumiendo que hay que aceptar la hipótesis nula, por una confusión entre región de aceptación y rechazo, es decir, porque arrastran un error del apartado anterior. En la Tabla 5.5.33, realizamos el análisis semiótico de un ejemplo.

Tabla 5.5.33. Análisis semiótico de un ejemplo en la categoría PC3

| Expresión | Contenido |
|---|---|
|  | – El estudiante decide no rechazar la hipótesis nula (proceso de interpretación), genera un *conflicto*, puesto que de acuerdo con los datos de la tarea, debería rechazarla (cae en la zona de rechazo: propiedad).<br>– Al interpretar los resultados del estudio inferencial (concepto), deduce que los alumnos se sitúan en *la media de velocidad lectora*, (proceso de argumentación). La interpretación sería correcta para la decisión tomada |

**Respuestas incorrectas**

*I1. Concluye, correcta o incorrectamente, pero no interpreta.* Para esta categoría hemos seleccionado a todos los estudiantes que han respondido al apartado, ya sea rechazando o no rechazando la hipótesis nula; pero sin interpretar los resultados En la Tabla 5.5.34, se incluye un ejemplo, donde se usan símbolos y terminología que no coincide con la contextualización de la tarea.

De la Tabla 5.5.35 se desprende que hemos encontrado casi un 25% de respuestas totalmente correctas para el apartado 4. En alto el porcentaje de respuestas parcialmente correctas (39,3%), de las cuales solo casi un 5% corresponden a estudiantes que si bien responden rechazando la hipótesis nula, no lo hacen usando términos del enunciado del problema; y un 3,1% es para los que hace una interpretación correcta, pero como contraste unilateral. El grueso interpreta que no rechaza $H_0$.

Tabla 5.5.34. Análisis semiótico de un ejemplo en la categoría I1

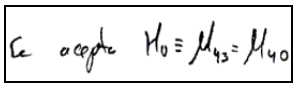
| Expresión | Contenido |
|---|---|
|  | – El estudiante decide no rechazar la hipótesis nula (proceso de interpretación). Responde en términos de aceptarla (concepto: proceso lógico del contraste de hipótesis). Genera *conflictos* uno al no rechazarla y otro respondiendo en términos de "aceptarla".<br>– Utiliza simbología no institucionalmente aceptada, *conflicto* (atributo expresión-contenido desarticulado),<br>– *Conflicto,* no interpreta los resultados. |

Tabla 5.5.35. Frecuencias (y porcentajes) de respuestas al Problema 1-Interpretación

| Categorías | | Frecuencia | % |
|---|---|---|---|
| Correctas | | 55 | 24,6 |
| PC1. | Interpretación correcta del resultado, pero sin relacionar con el contexto. | 11 | 4,9 |
| PC2. | Interpreta contraste unilateral correctamente. | 7 | 3,1 |
| PC3. | Interpretan, aceptando la igualdad de media de población y muestra. | 60 | 31,3 |
| I1. | No interpreta los resultados en el contexto del problema. | 9 | 4,0 |
| I2. | No relacionadas con la tarea | 5 | 2,2 |
| S/R | | 67 | 29,9 |
| Total | | 224 | 100,0 |

Si consideramos las respuestas correctas, los que han interpretado, de acuerdo a sus resultados (PC2 y PC3) obtenemos un 59% de estudiantes que han sabido contextualizar los resultados del estudio estadístico, completando el ciclo de modelización descrito por Henry (1997). Este es un porcentaje muy alto, en comparación con estudios como el de Arteaga (2011), en que menos del 25% de futuros profesores de educación primaria eran capaces de completar este paso. En el trabajo de Cañadas *et al*. (2012) un 31,5% no alcanza a interpretar el resultado en relación con el contexto del problema. Es verdad que una gran parte de estos estudiantes han confundido los pasos previos, o bien han usado un contraste unilateral, o llegaron a no rechazar la hipótesis nula. Pero la interpretación que hacen es correcta.

Hay además un pequeño porcentaje que no llega a contextualizar (I1) y alrededor del 30% no responde este apartado.

**Conflictos en la interpretación de resultados**

En el análisis del apartado podemos resumir se han encontrado los siguientes conflictos:

- *No contextualiza los resultados del análisis estadístico al fenómeno presente en la tarea*, que aparece en las categorías PC1, además de en I1 e I2 (11,1% de estudiantes)
- *Confunde la prueba bilateral y unilateral* (aparece en PC2, 3,1% de estudiantes).
- *Confunde las regiones de aceptación y rechazo y no rechaza la hipótesis nula cuando debería hacerlo* (aparece en las categorías PC3 e I1; en total un 35,3% de estudiantes). Esta dificultad también la encuentra Cañadas *et al*. (2012).

- *Responde en términos de aceptar la hipótesis nula* (aparece en I1; representa el 4% de estudiantes.
- *Utiliza símbolos incorrectos* (aparece en la categoría I1, en el 4%).
- *Confunde media muestral con poblacional* (aparece en I1, 4% de estudiantes).

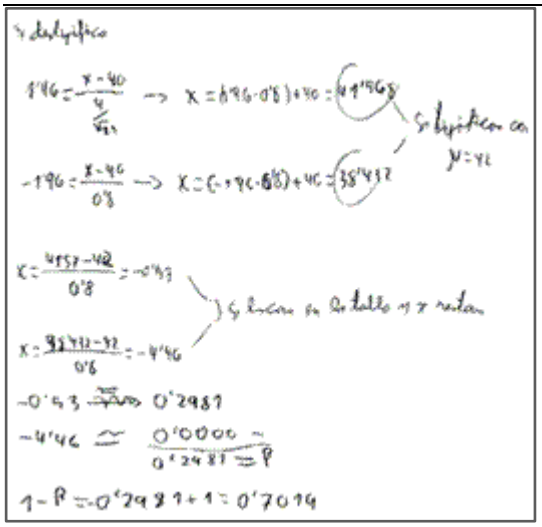
### 5.5.5. CÁLCULO DE LA POTENCIA DEL CONTRASTE

En las respuestas correspondientes al último apartado de este problema, el estudiante he de determinar la potencia del contraste, dado un valor de la media poblacional en la hipótesis alternativa y un nivel de significación para el contraste. A continuación se analizan los resultados.

**Respuestas correctas**

*C. Calcula los extremos del intervalo para obtener la probabilidad de no rechazar la hipótesis nula y estandariza correctamente. Calcula el valor correcto de la potencia. Re*producimos un ejemplo en la Tabla 5.5.36. Observamos, que el estudiante ha de poner en relación gran variedad de conceptos diferentes, propiedades y procedimientos con su argumentación, y aplicarlos, mediante un proceso de particularización al contexto del problema.

Tabla 5.5.36. Análisis semiótico de una respuesta correcta

| Expresión | Contenido |
|---|---|
|  | − El alumno identifica los percentiles para calcular los extremos del intervalo de la zona de no rechazo (particularización de un concepto). Encuentra los extremos del intervalo (procedimiento).<br>− Identifica la media poblacional $\mu = 42$ y la desviación estándar de la media muestral. Los usa para tipificar (procedimiento; conceptos).<br>− Calcula las probabilidades acumuladas a izquierda para -0,53 y -4,46 y resta los valores obtenidos en el orden correcto (procedimientos)<br>− Iguala el resultado usando símbolo adecuado para la probabilidad de error tipo II, $\beta$ (interpretación y cambio de representación).<br>− Resta de la unidad el valor de $\beta$, naturaleza complementaria de la potencia (propiedad, definición). Obtiene así la potencia de la prueba (procedimiento). |

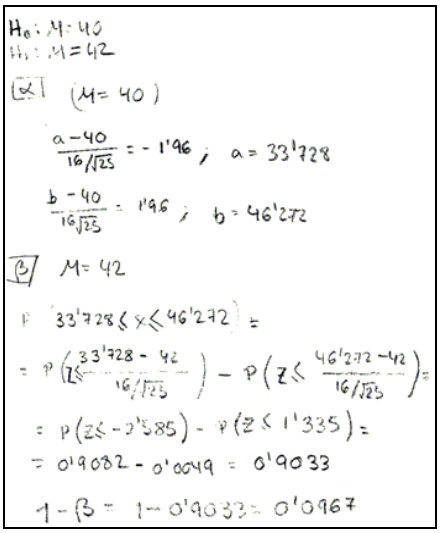
En particular debe establecer la relación entre la probabilidad de error tipo II y la potencia de un contraste de hipótesis, y reconocer que ésta también es variable. El alumno ha de comprender que la potencia del contraste es una función cuyos valores cambian a medida que lo hace el valor de la media poblacional bajo la hipótesis alternativa. Por ello no es posible controlar los dos errores a la vez; estos errores, junto con la potencia del contraste tienen una naturaleza condicional. En la interpretación de esa probabilidad condicional es donde encontramos más errores y concepciones erróneas respecto al contraste de hipótesis (Batanero, 2000).

**Respuestas parcialmente correctas**

Las respuestas parcialmente correctas, en general consisten en usar la varianza en lugar de la desviación estándar a la hora de tipificar, error también encontrado por Olivo (2008), incorrecta lectura de la tabla para obtener el percentil o en el cálculo de probabilidades para un intervalo. La complejidad de la tarea da lugar a observar varias categorías de respuestas parcialmente correctas, que agrupamos según cometen o no algún error en la determinación de los extremos del intervalo para el cálculo de la probabilidad de no rechazar la hipótesis nula.

*PC1. Usa la varianza, en vez de la desviación típica en las fórmulas de tipificación.* En esta categoría hemos agrupado a todos aquellos estudiantes que no determinan correctamente los extremos del intervalo para calcular la zona de aceptación, usando para estandarizar el valor de la varianza en lugar del de la desviación estándar, error que ya apareció en apartados anteriores, así como en la investigación de Olivo (2008). Arrastran este fallo hasta la finalización de la tarea. Calcula la probabilidad de error tipo II arrastrando ese error, coloca la diferencia de probabilidades en forma inadecuada realizando bien el cálculo de las mismas. Calcula la potencia restando de 1 el valor de beta (probabilidad de error tipo II). En la Tabla 5.5.37 realizamos un análisis semiótico de un ejemplo, donde el resto del procedimiento y las operaciones realizadas han sido correctos.

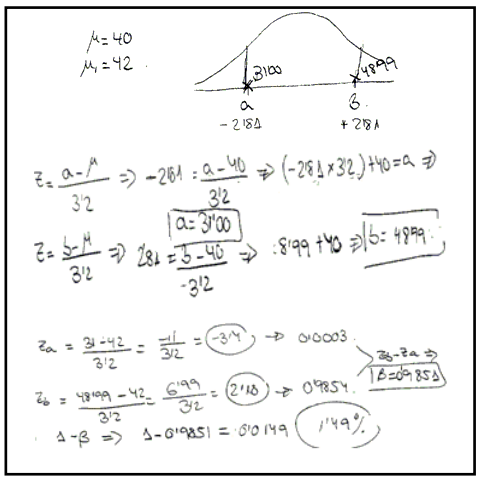
Tabla 5.5.37. Análisis semiótico de un ejemplo en la categoría PC1

| Expresión | Contenido |
|---|---|
|  | − El alumno indica los valores que usara para la media poblacional como hipótesis nula y alternativa y los extremos del intervalo para calcular la probabilidad de error Tipo II (particularización de conceptos).<br>− Iguala ambas expresiones para el cálculo de a y b usando adecuadamente los percentiles y despejando correctamente a y b (concepto y procedimiento)<br>− Presenta un *conflicto* al utilizar la varianza en lugar de la desviación estándar al estandarizar (concepto-procedimiento).<br>− Tipifica respecto a la segunda media poblacional dada en la hipótesis alternativa, para calcular la probabilidad de error tipo II (procedimiento y argumentación). Presenta el mismo *conflicto* al utilizar la varianza en lugar de la desviación estándar (procedimiento).<br>− Calcula correctamente las probabilidades para los valores obtenidos (procedimiento).<br>− Aparece un *conflicto* al calcular la probabilidad en un intervalo, ya que no resta el mayor del menor valor (concepto, *propied*ad). Invierte los valores de las pro*babilidad*es, quedando ahora correcto el cálculo (técnica del cálculo).<br>− Calcula la potencia restando de 1 del valor hallado para $\beta$ (concepto y procedimiento). |

*PC2. Usa la varianza, en vez de la desviación típica en la tipificación. Además, hace un error en la determinación del percentil*. En esta categoría hemos agrupado a todos aquellos alumnos que, además del error anterior, no logran calcular adecuadamente los extremos del intervalo para la probabilidad de no rechazar la

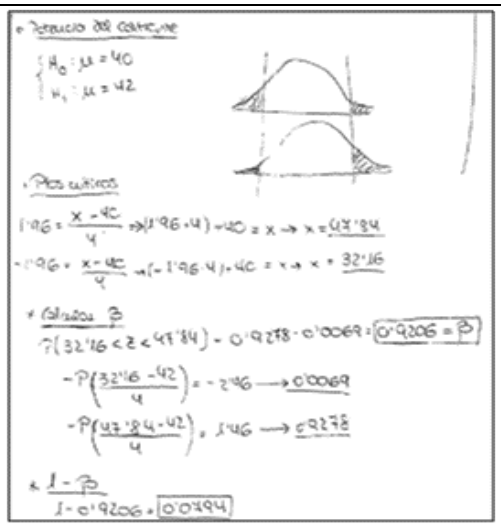
hipótesis nula, pues el percentil no es el adecuado. Este fallo lo arrastra a lo largo de toda la tarea y llega a un resultado erróneo. En la Tabla 5.5.38 realizamos un análisis semiótico de un ejemplo.

Tabla 5.5.38. Análisis semiótico de un ejemplo en la categoría PC2

| Expresión | Contenido |
|---|---|
|  | − Indica los valores que usara para la media poblacional bajo hipótesis nula como alternativa (particularización de un concepto).<br>− Representa la función de densidad, e indica un valor erróneo (*conflicto)* de los percentiles. Toma como $\alpha/2$ 0,0025 en lugar de 0,025 (procedimiento).<br>− Otro *conflicto* es utilizar la varianza en lugar de la desviación estándar al estandarizar (concepto-procedimiento).<br>− Confunde otra vez la varianza con la desviación estándar al tipificar (procedimiento). Las probabilidades acumuladas que encuentra son correctas (operaciones).<br>− Arrastrando los conflictos señalados llega a un valor posible para $\beta$ (concepto).<br>− Calcula la potencia restando de 1 del valor hallado para $\beta$ (definición).Da una respuesta final usando notación en porcentajes (cambio de representación) |

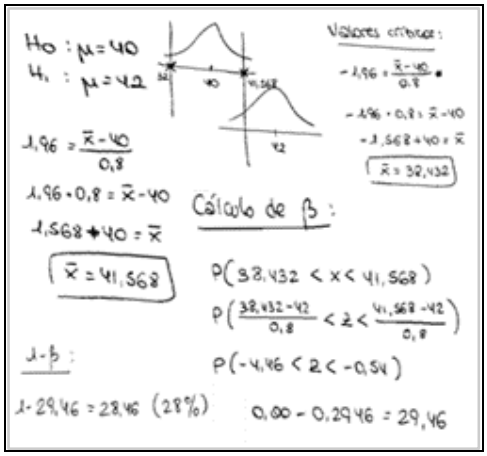
*PC3. Error al tipificar al inicio de la tarea y en la desviación típica del estadístico*. Estos estudiantes han cometido un fallo solamente al tipificar y el resto de la tarea la resuelve correctamente. En la Tabla 5.5.39, analizamos un ejemplo, en que emerge este conflicto, y además los estudiantes no dividen la desviación estándar por la raíz cuadrada del tamaño muestral, error que también cita Alvarado (2007).

Tabla 5.5.39. Análisis semiótico de un ejemplo en la categoría PC3

| Expresión | Contenido |
|---|---|
|  | − Indica los valores que usara para la media poblacional bajo hipótesis nula y alternativa (particularización de un concepto) y representa la función de densidad.<br>− Presenta un *conflicto* al estandarizar; no divide la desviación estándar por la raíz cuadrada del tamaño de la muestra (propiedad-procedimiento).<br>− Iguala el cálculo de los extremos del intervalo, usando los percentiles correctos, obteniendo valores de los extremos correctos de acuerdo con conflicto señalado (procedimientos).<br>− Repite el *conflicto* al tipificar, ya que no divide la desviación estándar por la raíz cuadrada del tamaño muestral (procedimiento).<br>− El valor de las probabilidades acumuladas que encuentra son correctos (operaciones).<br>− Arrastrando el conflicto señalado, arriba así a un valor posible para $\beta$ (concepto).Calcula la potencia restando de 1 del valor hallado para $\beta$ (definición). |

*PC4 Calcula la probabilidad en un intervalo cambiando el orden de los extremos, por lo que obtiene una probabilidad mayor que uno*. En esta categoría los estudiantes, obtienen los extremos del intervalo para la probabilidad de no rechazar la hipótesis nula y de estandarizar para determinar la probabilidad de error tipo II correctamente. Al ordenar los valores de las probabilidades lo hace en forma invertida, generando para la probabilidad un valor más grande que la unidad, de donde la potencia del contraste se determina a partir de una cifra incorrecta. En la Tabla 5.5.40, realizamos un análisis semiótico de un ejemplo. Se observa que los conflictos están relacionados con el cálculo de probabilidad, o sea es de tipo procedimental. Se muestra desconocimiento de las propiedades fundamentales de la probabilidad, lo que también ocurrió en el trabajo de Contreras (2011).

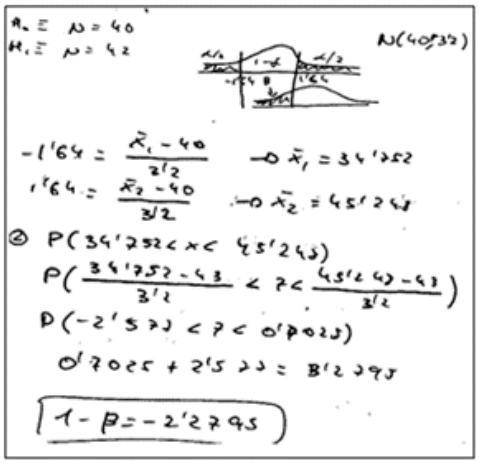
Tabla 5.5.40. Análisis semiótico de un ejemplo en la categoría PC4

| Expresión | Contenido |
|---|---|
|  | − Indica los valores que usara para la media poblacional en la hipótesis nula y alternativa y representa la función de densidad (densidad)<br>− Iguala para el cálculo de los extremos del intervalo, y estandariza correctamente, El valor de las probabilidades acumuladas que encuentra son correctos (procedimientos).<br>− *Conflicto* al calcular la probabilidad para el intervalo (procedimiento), invierte los valores de los extremos, resta mal y obtiene una probabilidad mayor que 1 (definición).<br>− Designa con $\beta$ la probabilidad de error tipo II (cambio de representación). Calcula la potencia restando de 1 del valor hallado para $\beta$ (definición). Fuerza un valor positivo tiene un *conflicto* ya que el resultado es mayor que uno (definición-propiedad). |

**Categorías para respuestas incorrectas**

*I1. Error en el cálculo del percentil y estandarización.* En esta categoría hemos clasificado a los estudiantes que han cometido los siguientes fallos: el percentil no es el que corresponde; en la estandarización para el cálculo de la probabilidad de no rechazar la hipótesis nula usan la varianza en lugar de la desviación estándar; al calcular la probabilidad de error tipo II, no usan el valor consignado en la tarea, sino la media muestral. Finalmente calculan incorrectamente la probabilidad para el intervalo hallado, produciendo un valor mayor que la unidad. De este resultado, luego de restar de uno, obtiene una cantidad negativa para la potencia de la prueba. En la Tabla 5.5.41, realizamos un análisis semiótico de esa respuesta

Tabla 5.5.41. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
|  | − Especifica los valores supuestos de la hipótesis nula y alternativa en $\mu = 40$ y $\mu = 42$ y representa la función de densidad. Aparece un *conflicto*, ya que confunde el percentil (particularización de conceptos).<br>− Indica mediante el símbolo "N(40,3´2)" la distribución que usará, generando un *conflicto*, pues no es correcta la desviación estándar (propiedad, procedimiento).<br>− Identifica los percentiles para calcular los extremos del intervalo de la zona de no rechazo (particularización de un concepto). *Conflicto* a pesar que encuentra los extremos de tal intervalo sin fallos (procedimiento), pero arrastrando error.<br>− Otro *conflicto* al especificar que tipificara con el valor de $\mu = 43$ para la probabilidad error tipo II (procedimiento-concepto). En la estandarización, usa $\sigma^2$ en lugar de $\sigma$ (propiedad).<br>− *Conflicto*, al obtener una probabilidad mayor que la unidad (procedimiento, propiedad).<br>− Presenta un *conflicto*, ya que obtiene como resultado una potencia con valor menor que 1 (propiedad). |

*I2. Fórmula del estadístico de prueba y del intervalo para la probabilidad de error tipo II incorrectas*. Estos estudiantes no usan la fórmula adecuada del estadístico y estandarizan cometiendo doble fallo para los extremos del intervalo que daría la probabilidad de no rechazar la hipótesis nula. No es adecuado el procedimiento para determinar la probabilidad de error tipo II, todo lo que redunda en obtener una potencia errónea de la prueba. En la Tabla 5.5.42, realizamos un análisis semiótico de la misma.

Tabla 5.5.42. Análisis semiótico de un ejemplo en la categoría I2

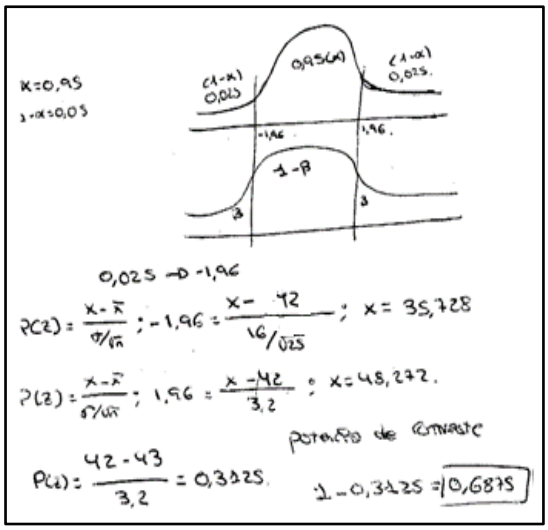
| Expresión | Contenido |
|---|---|
|  | − *Conflicto*, al usar como probabilidad de error tipo I, su complemento (propiedad-concepto).<br>− Representa la función de densidad en gráficos comparativos (cambio de representación).<br>− Indica la zona de aceptación y rechazo correctamente, separadas por los percentiles (concepto, propiedad y particularización).<br>− *Conflicto* al igualar una probabilidad con el estadístico de prueba (propiedad).<br>− *Conflicto*, ya que no usa la media muestral en la estandarización (procedimiento).<br>− *Conflicto* pues usa la varianza para tipificar (propiedad-particularización). Indica tanto la zona de aceptación y rechazo correctamente, separadas por los percentiles (concepto, propiedad y particularización).<br>− Genera una probabilidad de error tipo II denotando un *conflicto*, que se propaga hacia la potencia mostrada de la prueba (concepto, propiedad y particularización). |

Tabla 5.5.43. Frecuencias (y porcentajes) de respuestas al Problema 1-Potencia del contraste

| | Categorías | Frecuencia | % |
|---|---|---|---|
| C. | Correcta | 32 | 14,3 |
| PC1. | Error en tipificación, usando la varianza en lugar de desviación al estandarizar. | 32 | 14,3 |
| PC2. | Igual error que PC1., además incorrecto el percentil. | 15 | 6,7 |
| PC3. | Error en la tipificación y no divide por el tamaño de muestra | 4 | 1,8 |
| PC4. | Intercambia los extremos al calcular probabilidad en un intervalo. | 6 | 2,7 |
| I1. | Errado percentil, mal la estandarización, vuelve a tipificar usando $\bar{x}$ (43). | 18 | 8,0 |
| I2. | Error en formula de estadístico de prueba y del intervalo para la P(ETII). | 7 | 3,1 |
| I3. | No relacionadas con la tarea. | 19 | 8,5 |
| S/R | | 91 | 40,6 |
| | Total | 224 | 100,0 |

En el apartado 5 hemos encontrado pocas respuestas totalmente correctas, apenas en un 14,3%, se desprende de la Tabla 5.5.43 que probablemente indiquen la complejidad semiótica de esta tarea para cuya solución entran en juego una gran cantidad de conceptos, procedimientos, argumentos; además debe utilizarse un lenguaje muy preciso para arribar a una respuesta correcta en su totalidad.

Hay un más respuestas parcialmente correctas (22,8%), clasificadas en las que confunden varianza con desviación estándar, cometen algún error en la tipificación de variables y tienen algún fallo en cuanto al cálculo de probabilidades que están en juego en la tarea; e incorrectas (19,6%). Los errores están relacionados con la estandarización (al igual que en Olivo, 2008 y Cañadas, 2012), la confusión entre media muestral y poblacional (pues tipifica usando el valor de la media muestral y no el de la media poblacional en la alternativa), que también cita Schuyten (1991), la elección de la fórmula para el estadístico de prueba y también en el cálculo de probabilidad (error similar al que señala Contreras, 2011). Todos estos casos implican llegar a respuestas incorrectas o no dar respuesta.

Un alto porcentaje (casi 40%, al contabilizar el porcentaje de correctas y parcialmente correctas) conoce el procedimiento para la determinación de la potencia de un contraste, logran luego de seguir una serie de pasos coherentes una respuesta, siendo menos el porcentaje (19,6%) que llegan a una respuesta incoherente que entre otras están en relación con exhibir una probabilidad cuya valor sea mayor que la unidad o negativa, error señalado por Contreras (2011).

**Conflictos en el cálculo de la potencia**

En el análisis del apartado se han encontrado los siguientes conflictos semióticos que han llevado a los alumnos a un cálculo erróneo de la potencia.

- *Confunde la varianza y desviación estándar en la tipificación* (que aparece en las categorías PC1, PC2 y PC3; I1 e I2, en total 33,9% de los estudiantes). Como hemos dicho Olivo (2008) también cita este conflicto en el cálculo de intervalos de confianza.
- *Confusión en el cálculo de probabilidades* (que aparece en las categorías PC1, PC2 y PC4; I1 e I2; en total 34,8% de los estudiantes). Lleva a probabilidades incorrectas, a veces mayores que la unidad, al igual que en Contreras (2011).

- *Confusión para el cálculo de percentiles* (aparece en las categorías PC2 e I1; en total un 14,7% de estudiantes). No hemos encontrado descrito este conflicto.
- *No usan la distribución de la media muestral en el cálculo del estadístico de prueba* (aparece en las categorías PC3, en total en 1,8% de estudiantes).
- *Mantiene una confusión entre media muestral y poblacional*, al igual que indica Schuyten (1991) (aparece en las categorías I1 e I2, en total un 11,1% de estudiantes).

### 5.5.6. SÍNTESIS DE RESULTADOS EN EL PROBLEMA 1

Para finalizar el estudio del primer problema presentamos en la Tabla 5.5.44 los porcentajes de respuestas correctas, parcialmente correctas e incorrectas en cada apartado.

Tabla 5.5.44. Porcentaje de tipos de respuestas en cada apartado

| Apartado | Correctas | Parcialmente correctas | Incorrectas | No responde |
|---|---|---|---|---|
| 1. Definición de las hipótesis estadísticas | 33,0 | 0,0 | 50,5 | 16,5 |
| 2. Elección del contraste | 1,3 | 41,9 | 31,4 | 25,4 |
| 3. Cálculo de regiones y decisión sobre las hipótesis | 20,5 | 35,3 | 21,4 | 22,8 |
| 4. Interpretación de resultados | 24,6 | 39,3 | 6,2 | 29,9 |
| 5. Cálculo de la potencia del contraste | 14,3 | 25,5 | 19,6 | 40,6 |

Destacamos que el mayor porcentaje dentro de las correctas (33%) corresponde a la definición de las hipótesis estadísticas, lo cual es lógico, pues un error en este apartado suele arrastrar errores posteriores. No obstante, celebramos esta alta valoración ya que el establecimiento correcto de las mismas es esencial para los pasos sucesivos (Vallecillos, 1994). Este porcentaje es un poco menor que el obtenido en una pregunta similar en Cañadas *et al*. (2012), que fue el 46,6%, pero se acerca más al encontrado por Vallecillos (1994) que fue de un 26% también con estudiantes de Psicología. No encontramos respuestas parcialmente correctas en el apartado. En las respuestas incorrectas los estudiantes plantean hipótesis unilaterales en vez de bilateral, coincidiendo también con Vallecillos (1994) y Cañadas (2012). En otros casos se plantean hipótesis utilizando el estadístico muestral o no cubren el espacio paramétrico.

El porcentaje más bajo para las correctas (1,3%) lo tiene la elección del contraste, la mayoría de la valoración en este apartado se la llevan las de tipo parcialmente correctas (41,9%), puesto que la precisión en el lenguaje a usar requiere muchas palabras con significados diversos que se ensamblan para formar la respuesta. No obstante hacemos notar que muchos de ellos eligieron un contraste adecuado para las hipótesis establecidas; es decir, el bajo porcentaje de éxito en este apartado en parte se debe a haber establecido hipótesis inadecuadas. En las respuestas parcialmente correctas, aunque los estudiantes identifican el contraste, olvidan algunas de las condiciones (por ejemplo, que la desviación típica es conocida). Además, siguiendo el razonamiento en el apartado anterior, la mitad de la muestra elige acertadamente la prueba a aplicar. Es decir, plantean una hipótesis incorrecta, pero luego eligen un contraste adecuado para contrastarla.

Una vez elegido el contraste, en el siguiente paso (cálculo de regiones y decisión) mejora los resultados, pues algunos alumnos en el punto anterior, no supieron verbalizar correctamente el contraste elegido, pero llevan correctamente a cabo los cálculos y la toma de decisión. Son también bastantes las respuestas parcialmente correctas en este apartado, donde hay algún error en el cálculo de estadísticos, por ejemplo al tipificar, pero luego toman una decisión coherente con los resultados obtenidos.

El porcentaje más bajo entre las incorrectas (19,6%) corresponde al cálculo de la potencia, lo que es razonable, debido a la complejidad del profesor y la cantidad de conceptos y procedimientos que se deben aplicar. Ello ocasiona un porcentaje alto de parcialmente correctas, debido a fallo en alguno de los pasos. También el mayor porcentaje de no respuesta se relaciona con el cálculo de la potencia del contraste (40,6%).

Finalmente en el apartado 4 de interpretación de resultados, una cuarta parte contextualiza correctamente los resultados y otra parte importante también contextualiza, aunque usando sus resultados (por ejemplo, aceptación de la hipótesis) incorrectos debido a fallo en algún paso anterior. No obstante, la proporción de alumnos que llega al último paso de la modelización (Henry, 2001) es mayor que lo encontrado en la investigación de Arteaga (2011). En resumen, los resultados ponen de manifiesto la alta complejidad de aplicación correcta de todos los pasos en el contraste de hipótesis, aunque aún una parte de la muestra los completa correcta o parcialmente correctamente.

## 5.6 RESULTADOS EN EL PROBLEMA 2: ANÁLISIS DE VARIANZA DE MEDIDAS REPETIDAS

El segundo problema abierto es específico del análisis de varianza y por ello, no tenemos muchas posibilidades de comparar con estudios previos. Nuestros resultados serán, en su mayoría originales. Trataremos de comparar en lo posible con otras investigaciones sobre contraste de hipótesis, así como con el primer problema analizado en las secciones anteriores.

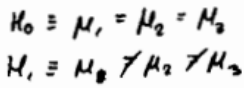
### 5.6.1 DEFINICIÓN DE LAS HIPÓTESIS ESTADÍSTICAS

A continuación, se describen las categorías encontradas para el primer apartado del problema 2, en el que el alumno ha de interpretar el enunciado y formular las hipótesis estadísticas adecuadas.

**Respuestas correctas**

*C. Hipótesis estadísticas y notación correctas.* En esta categoría el alumno logra definir claramente tanto la hipótesis nula, referida a la igualdad entre las medias poblacionales, como la alternativa, que debe indicar la contraria. Elige los parámetros adecuados sobre los que se plantean las hipótesis (las medias poblacionales); usa también la notación aceptada al respecto, $\mu_i$ $(i = 1,2,3)$. Puede discriminar que tratará de probar si existe diferencia entre los niveles medios en tres momentos (factor) distintos de medición a los mismos sujetos. En la Tabla 5.6.1 se muestra un ejemplo de esta categoría, desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.1. Análisis semiótico de un ejemplo en la categoría C

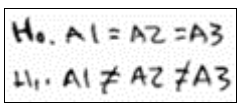
| Expresión | Contenido |
|---|---|
|  | − El alumno identifica correctamente los parámetros para contrastar (concepto y procedimiento).<br>− Plantea una hipótesis nula y alternativa adecuada al enunciado (conceptos y procedimiento). Identifica correctamente que son tres momentos distintos para comparar entre si las medias poblacionales (particularización de un concepto).<br>− Traduce la expresión "antes, durante y después" a notación matemática (interpretación y cambio de representación), usa la idea de igualdad (concepto) al establecer la hipótesis nula y la de desigualdad (concepto) al establecer la alternativa.<br>− Discrimina entre hipótesis nula y alternativa y las expresa en notación adecuada (concepto y representación); en la hipótesis alternativa plantea la que el investigador quiere probar y en la nula la contraria (propiedades). |

Como observamos en el análisis semiótico, el estudiante ha de discriminar el papel de la hipótesis nula y alternativa. Ha de comprender que la hipótesis nula plantea la igualdad entre las medias poblacionales, que se desprenden de medir el nivel de ansiedad en tres momento diferentes; mientras que en la alternativa se plantea que al menos un par de ellas sean distintas, es decir son complementarias. La hipótesis nula se plantea para ser rechazada, ya que expresamos en ella lo contrario de lo que esperamos mostrar (Batanero, 2000).

**Respuestas parcialmente correctas**

*PC. Respuesta parcialmente correcta:* Sólo se ha encontrado un tipo de respuesta parcialmente correcta, en la que el alumno usa una simbología no aceptada naturalmente en la literatura especializada. Consiste en la letra A con respectivo subíndice para indicar seguramente el nivel de ansiedad medido en cada momento. No la consideramos correcta, pues con esa notación no se hace referencia al nivel medio poblacional a través del tiempo, que es sobre lo que se desea hacer inferencia. En la Tabla 5.6.2 se detalla un análisis semiótico usando un ejemplo. Estos estudiantes dan una respuesta parcial. Si bien consideran que se trata de una prueba estadística, discriminando tanto hipótesis nula como alternativa correctamente, (usando la igualdad en la primera y la desigualdad en la segunda), no queda claro si reconoce que compara medias poblacionales en cada tiempo de medición.

Tabla 5.6.2. Análisis semiótico de un ejemplo en la categoría PC

| Expresión | Contenido |
|---|---|
|  | − No identifica claramente los parámetros sobre los que debe contrastar, *conflicto* (concepto y procedimiento).<br>– Identifica correctamente la igualdad para la hipótesis nula (propiedad) y lo contrario en la hipótesis alternativa (particularización de un concepto).<br>– Traduce del enunciado que las comparaciones son en tres momentos distintitos (interpretación y cambio de representación).<br>– Discrimina las dos hipótesis, pero no puede expresarlas con la notación adecuada (concepto y expresión).<br>– Aparece un *conflicto* al no usar la notación aceptada institucionalmente, no generaría las inferencias del caso (función semiótica: plano de expresión-plano de contenido). |

**Respuestas incorrectas**

Seguidamente se describen las respuestas incorrectas, que se han clasificado en siete categorías. Los errores encontrados son plantear hipótesis como si se tratara de comparar sólo dos muestras, no usar la simbología adecuada, tendiendo a verbalizar las hipótesis y hacer referencia a efectos o comparar con una variable independiente (que denota: VI). En otros casos utiliza la media muestral como símbolo para las comparaciones (confusión descrita por Schuyten, 1991), realiza pruebas unilaterales estudiando relaciones entre técnicas utilizadas y nivel de ansiedad, aunque las primeras nunca se especifican en el problema; también expresan las hipótesis estadísticas como una combinación lineal entre parámetros, o consideran un solo momento de análisis comparando una única media poblacional.

*I1. Plantea las hipótesis como si se tratase de una prueba de dos muestras pareadas.* Los estudiantes que han sido agrupados en esta categoría son los que han considerado para el análisis sólo dos momentos, identificándolos como antes y después de aplicar algún tratamiento. Deberían comparar tres momentos, siendo eso lo que debe especificarse a través de las hipótesis. A pesar de este conflicto usan simbología adecuada, y el parámetro elegido de comparación es correcto. Realiza además una prueba unilateral. En la Tabla 5.6.3 se detalla un análisis semiótico, describiendo un ejemplo de respuesta de este tipo producido por uno de los estudiantes.

Tabla 5.6.3. Análisis semiótico de un ejemplo en la categoría I1

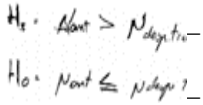
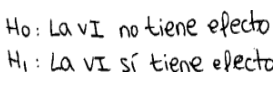
| Expresión | Contenido |
|---|---|
|  | – El alumno realiza una interpretación incorrecta del enunciado, aparece un *conflicto*, pues asume que existen dos poblaciones, interpretándolas como la existencia de dos momentos (campo de problemas), antes y después de aplicar el tratamiento).<br>– Discrimina las dos hipótesis y las expresa en notación adecuada usando el parámetro poblacional (concepto y expresión).<br>– Reconoce la igualdad para la hipótesis nula, aunque no la considera puntual (propiedad).<br>– Considera un contraste unilateral *conflicto* (confunde el campo de problemas).<br>– No traduce la expresión "tienen efecto sobre la ansiedad" al no usar la notación "≠" bajo la alternativa (campo de problemas). |

Tabla 5.6.4. Análisis semiótico de un ejemplo en la categoría I2

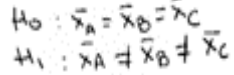
| Expresión | Contenido |
|---|---|
|  | – El estudiante establece correctamente una prueba bilateral (campo de problemas, procedimiento).<br>– Usa el término *efecto* para explicar el contenido de las hipótesis (concepto y lenguaje).<br>– Establece igualdad para la hipótesis nula (propiedad) haciendo referencia a un efecto nulo.<br>– *Conflicto* ya que define las hipótesis relacionando con efectos sin referir cuantos son y qué relación tienen con la media muestral. (lenguaje y conceptos). No reconoce los tres momentos en que se mide la variable. |

*I2. Plantea las hipótesis en términos de efectos sin utilizar parámetros*. Los estudiantes que han respondido de este modo no usan los parámetros adecuados y no hacen referencia a medias poblacionales; en lugar de ello las plantean en términos de

"efectos", sin indicar los momentos en que se mide la ansiedad. En la Tabla 5.6.4, se muestran un ejemplo y se detalla un análisis semiótico.

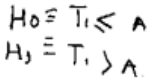
*I3. Utiliza la media muestral para plantear las hipótesis estadísticas.* Los estudiantes que han sido agrupados en esta categoría plantean las hipótesis utilizando los resúmenes estadísticos muestrales. Por tanto se confunde los niveles de análisis que se corresponde con media poblacional y media global (Schuyten, 1991). Este tipo de error en el planteamiento de hipótesis ha sido reportado en las investigaciones de Vallecillos (1994), en otras incluidas en el resumen de Castro Soto et al. (2007) y en la de Cañadas *et al*. (2012) sobre otros tipos de contraste, así como en el primer problema, pero no lo hemos encontrado citado en relación al análisis de varianza. En la Tabla 5.6.5, se muestra un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.5. Análisis semiótico de un ejemplo en la categoría I3

| Expresión | Contenido |
|---|---|
|  | – El alumno realiza una interpretación correcta del enunciado, asumiendo que existen tres poblaciones (conceptos).<br>– Los símbolos que utiliza tanto en la hipótesis nula (igualdad) como en la alternativa (desigualdad) son los adecuados (concepto y representación).<br>– El alumno asimismo, discrimina la hipótesis nula como puntual y la hipótesis alternativa como aquella que quiere probar (propiedades).<br>– Genera *conflicto* al plantear las hipótesis en término de estadísticos muestrales y no de parámetros (concepto).<br>– Un posible *conflicto* es la confusión entre población y muestra (conceptos)<br>– Relacionado con el último anterior, presenta el *conflicto* consistente en confundir la media muestral que es un estadístico con la media poblacional que es un parámetro (conceptos). |

*I4. Compara técnicas con nivel de ansiedad.* Los estudiantes que han sido agrupados en esta categoría confunden la variable independiente (llamándola técnica de reducción de estés) y dependiente (ansiedad) del enunciado. Por ello plantean las hipótesis comparando estas dos variables. Usan además una prueba unilateral, en vez de bilateral y no usan los parámetros adecuados (medias de las variables); por otro lado, reconoce solo dos comparaciones. En la Tabla 5.6.6, se muestra un ejemplo de respuesta y se detalla un análisis semiótico. Hacemos notar que la confusión entre variables dependiente e independiente aparece en la investigación de Arteaga (2011) en relación con la elaboración de gráficos estadísticos.

Tabla 5.6.6. Análisis semiótico de un ejemplo en la categoría I4

| Expresión | Contenido |
|---|---|
|  | – Reconoce la igualdad para la hipótesis nula, aunque no la considera puntual, sino plantea un contraste unilateral (campo de problemas) *conflicto.*<br>– *Conflicto* al plantear la comparación de la variable dependiente e independiente (conceptos); también referido al lenguaje y argumentación. Posiblemente otro consiste en confundir los dos tipos de variables.<br>– *Conflicto* no logra identificar algún parámetro adecuado para contrastar (lenguaje).<br>– Parece desea poner de manifiesto la comparación de dos momentos en el análisis. *Conflicto:* No reconoce debe usar algún modelo de análisis de varianza (campo de problemas). |

*I5. Expresa las hipótesis estadísticas con una combinación lineal de los parámetros*. Los estudiantes que han sido agrupados en esta categoría tienden a utilizar los contrastes a posteriori, que se suelen aplicar en análisis de varianza una vez que se ha rechazado la hipótesis nula (con la salvedad que la suma de los coeficientes del contraste debe ser siempre la unidad). La finalidad que buscan estos contrastes es descubrir cuáles son los niveles que causan las diferencias (Montgomery 2002). Estos estudiantes no comparan la igualdad de medias poblacionales entre sí, sino la diferencia de dos de ellas con respecto a la tercera. En la Tabla 5.6.7, se detalla el análisis semiótico de un ejemplo.

Tabla 5.6.7. Análisis semiótico de un ejemplo en la categoría I5

| Expresión | Contenido |
|---|---|
| $H0\ \mu_1-\mu_2-\mu_3=0$<br>$H_1\ \mu_1-\mu_2-\mu_3\neq 0$ | – Identifica correctamente los símbolos para el parámetro (particularización de un concepto) y reconoce que debe realizar una comparación donde intervienen tres distintos (campo de problemas).<br>– Reconoce una igualdad para la hipótesis nula y desigualdad para la hipótesis alternativa (propiedad).<br>– *Conflicto* No compara los tres momentos del análisis, usa signos inadecuados (establece una función semiótica incorrecta) tanto en la hipótesis nula como en la hipótesis alternativa.<br>– Plantea la hipótesis de los contrastes no ortogonales *conflicto* (sub-campo de problemas). |

*I6. Considera un solo momento de análisis y las técnicas de reducción del estrés como la única variable.* Concentramos en esta categoría todos aquellos estudiantes que comparan una media poblacional de los atletas, una vez sometidos a las técnicas de reducción de estrés con el valor nulo. No comparan tres momentos en el análisis. En la Tabla 5.6.8, se muestra un ejemplo de respuesta con su análisis semiótico.

Tabla 5.6.8. Análisis semiótico de un ejemplo en la categoría I6

| Expresión | Contenido |
|---|---|
| $H_0: \mu_{TIE} = 0$ $H_1: \mu_{TIE} \neq 0$ | – Genera un *conflicto* puesto que no puede identificar los tres momentos de análisis que debe comparar (particularización de un concepto).<br>– Utiliza una simbología adecuada cuando usa el símbolo: $\mu_{TIE}$, suponemos con esto quiere representar la "media poblacional de la respuesta a las técnicas aplicadas", lo cual produce un *conflicto* con la variable independiente (concepto-expresión-interpretación).<br>– Toma como variable dependiente las "TIE", *conflicto* en relación con las definiciones de factor-variables.<br>– Reconoce que en la hipótesis nula debe haber una igualdad (propiedad) y es la contraria a la que el investigador quiere probar (propiedad) en este caso: "si existen diferencias significativas en el nivel de ansiedad entre las diferentes mediciones".<br>– Aparece un *conflicto* al no traducir la expresión "mediciones antes, durante y después" en el nivel de ansiedad entre las diferentes mediciones a notación matemática $\mu_1 = \mu_2 = \mu_3$ en $H_0$; $\mu_1 \neq \mu_2 \neq \mu_3$ para $H_1$ razón que lo le lleva a considerar un contraste unilateral, por fallo en la interpretación del enunciado (fallo al reconocer el campo de problemas). |

En la Tabla 5.6.9 se presentan las frecuencias por categoría. Hemos encontrado una gran variedad de respuestas incorrectas para este apartado, pero una sola

parcialmente correcta. Únicamente la cuarta parte de estudiantes completa correctamente el apartado, aunque Sumando las respuestas parcialmente correctas, donde el único problema es la simbología usada llegamos casi al 30 %. Observamos que el porcentaje de estudiantes que responde correctamente es similar al obtenido en la primera pregunta, Se encuentra también equilibrado con los que no dan una respuesta (25%).

Tabla 5.6.9. Frecuencias (y porcentajes) de respuestas al Problema 2. Hipótesis

| Categorías | Frecuencia | % |
|---|---|---|
| Correctas | 56 | 25,0 |
| Parcialmente correctas (simbología inadecuada) | 11 | 4,9 |
| I1. Prueba de dos muestras | 26 | 11,6 |
| I2. Plantea la hipótesis en términos de efectos | 45 | 20,1 |
| I3. Plantea las hipótesis usando las medias muestrales | 4 | 1,8 |
| I4. Compara las variables dependiente e independiente | 10 | 4,5 |
| I5. Combinación lineal de parámetros | 6 | 2,7 |
| I6. Compara un solo momento de análisis | 4 | 1,8 |
| I7. Respuesta imprecisa | 6 | 2,7 |
| Sin responder | 56 | 25,0 |
| | 224 | 100,0 |

Hay un 34,4% de estudiantes que reconocen la existencia de tres momentos de comparación para el análisis, ya sea usando signos y símbolos adecuados (C), usando signos pero no símbolos (I1 e I3) y los que usan signos incorrectos pero bien los símbolos (I5). El resto o no responde o bien confunde el análisis de varianza con otro contraste de comparación de una o dos muestras.

En un porcentaje pequeño (1,8%) encontramos los que no pueden discriminar adecuadamente los conceptos de estadístico y parámetro, confusión que se ve muy acentuada a la hora de plantear hipótesis para pruebas unilaterales o bilaterales de una o dos muestras, lo denuncian tanto Vallecillos (1994) como Cañadas (2012). Otro pequeño porcentaje no diferencia entre variable dependiente e independiente, al igual que en Arteaga (2011). El mayor porcentaje de respuestas incorrectas son los que no pueden reconocer el análisis de varianza como adecuado para este tipo de tareas.

**Conflictos semióticos al establecer las hipótesis**

En el análisis semiótico de las respuestas incorrectas en los registros escritos de los estudiantes que se han ido detallando a lo largo del parágrafo 5.6, se han encontrado los siguientes conflictos semióticos (Godino, 2002):

- *Usan un contraste de comparación de tres medias muestrales*. Aparece en la categoría I3, en un 1,8%. Este conflicto implica una confusión entre *estadístico y parámetro*, que ya apareció en el primer problema, pues confunden la media muestral y poblacional, aunque reconocen la comparación de los tres momentos de análisis. El mismo conflicto se repite en los estudiantes que usan el estadístico muestral en lugar del parámetro para plantear sus hipótesis. Como hemos dicho, apareció en la investigación de Cañadas (2012).
- *Confusión al no reconocer que la hipótesis nula es puntual*, también presente en el primer problema Es el caso de las categorías I1 e I4 para un 16,1% de la muestra.

No usan la igualdad para definir la hipótesis nula, error que podría incidir en la generación de otros conflictos al momento de definir la distribución del estadístico de prueba.

- *Plantea las hipótesis en término de efectos, confundiendo variables dependientes e independientes* (aparece en la categoría I2). Este conflicto sólo lo hemos encontrado en la investigación de Arteaga (2011) al comparar gráficos estadísticos, pero no en relación al contrate de hipótesis y es específico del análisis de varianza, Reconocemos un 20,1% de la muestra que tiene esta confusión. En algunos casos cuando establece hipótesis, compara un efecto sobre la ansiedad (si bien las hipótesis en un análisis de varianza, pueden generarse usando efectos, pero en esta categoría no describe cuales son, ni usa lenguaje adecuado para referirse a ellos). Asimismo en otros casos compara el efecto de una VI (que suponemos es una variable independiente) asumiendo que no lo tiene (en $H_0$) y que lo tiene (para $H_1$).
- *Confusión en el campo de problemas* (aparece en las categorías I1, I4, I6). Reconocemos un 17,9% de la muestra que al establecer las hipótesis confunde el campo de problemas, generando hipótesis como si se tratase de una prueba de hipótesis para comparar dos medias (I1 e I4, 16,1%), asumiendo en I1 un contraste unilateral; como de una media (I6 1,8%). No asumen la existencia de tres niveles del factor para ser comparados. La confusión del campo de problemas también se describe en Olivo (2008) en el cálculo de intervalos de confianza.

En consecuencia, estos errores se unen a los ya citados por otros autores sobre el planteamiento de hipótesis estadísticas y sugiere la dificultad de este concepto, de acuerdo con Chow (1996) y Batanero (2000). Aunados a estos conflictos hay muchos problemas *la utilización de símbolos para el parámetro* (aparece en las categorías PC, I2, I4). Un 29,5% utiliza símbolos inadecuados para anotar ambas hipótesis; algunos eligen letras imprenta mayúsculas, indicando los momentos en que se toman los datos (caso PC 4,9%); otros como en I2 en un 20,1%, expresan las hipótesis en forma coloquial, mientras que en I4 (4,5%) lo hacen usando otra simbología alejada de la usada en la bibliografía para referirse a las medias poblacionales ($\mu$).

### 5.6.2 VARIABLE DEPENDIENTE E INDEPENDIENTE

Para este segundo apartado del problema, el estudiante debe definir cuál es la variable dependiente y cual la independiente en el estudio descrito. Deberá reconocer se mide la ansiedad (variable dependiente), y que ésta depende del momento en que se la mide (variable independiente), pasando por tres distintos, a saber: antes, durante y después de enseñadas algunas técnicas no especificadas. Este apartado se relaciona con las hipótesis estadísticas, puesto que estas se plantean respecto a la variable dependiente.

**Respuestas correctas**

*C. Variables identificadas correctamente.* En esta categoría el alumno logra identificar claramente que el "nivel de ansiedad", es la dependiente, pues puede cambiar según el momento en que se la mide, logrando discernir que la independiente es el "momento de la medición". En la Tabla 5.6.10 se muestra un ejemplo. Tal como observamos en el análisis semiótico, el estudiante ha de discriminar la existencia de dos

variables en el problema, una independiente y la otra dependiente. El alumno ha de comprender que la independiente plantea los niveles del factor, es decir los momentos en que se toman los datos; mientras que la dependiente es el nivel de ansiedad para cada momento.

Tabla 5.6.10. Análisis semiótico de un ejemplo en la categoría C

| Expresión | Contenido |
|---|---|
| VD: nivel de ansiedad<br>VI: momento de la medición | – El alumno ha identificado correctamente ambos tipos de variables (conceptos-propiedad).<br>– Reconoce que el contraste pone en juego una variable independiente, factor, que tiene tres niveles (momentos de la toma de los datos) (campo de problemas).<br>– Para la expresión VD le ha asignado el contenido "nivel de ansiedad"(definición).<br>– Para la expresión VI le ha asignado el contenido "momentos de la medición" (definición).<br>– Con la asignación realizada a VI reconoce que el modelo de análisis de varianza que se aplica pone en juego tres momentos para una misma población (campo de problemas). |

**Respuestas parcialmente correctas**

Seguidamente se describen las categorías de respuestas que, no siendo correctas, no llegan a ser del todo incorrectas, son las parcialmente correctas. Encontramos dos categorías.

*PC1. Variable dependiente correctamente identificada, pero se toma como variable independiente las técnicas para combatir el estrés.* Estos estudiantes contestan siempre correctamente lo que se mide en la variable dependiente, pero sistemáticamente confunden la variable independiente (tiempo de medida) con el tratamiento que se dio a los atletas, las técnicas de reducción de estrés. En la Tabla 5.6.11 se muestra un ejemplo de esta categoría, y su análisis semiótico.

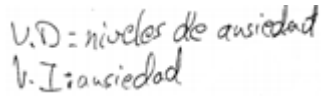
Tabla 5.6.11. Análisis semiótico de un ejemplo en la categoría PC1

| Expresión | Contenido |
|---|---|
| VD: Ansiedad precompetitiva en atletas<br>VI: Técnicas de Inoculación de Estrés | – Identifica correctamente la variable dependiente (concepto-definición), asignando a la expresión VD el contenido "Ansiedad precompetitiva en atletas" (particularización de un concepto).<br>– *Conflicto* al identificar la expresión VI con "Técnicas de Inoculación de estrés" (concepto-definición-particularización).<br>– Podría acarrear otro *conflicto* relacionado con los tres momentos en que se comparan las respuestas medias para establecer las hipótesis estadísticas (campo de problemas). |

*PC2. Está correctamente definida solo una de las variables.* Estos estudiantes tienen una respuesta parcial, puesto que definen correctamente sólo una de las dos

variables, ya sea la dependiente o la independiente, con respuestas diversas. Sería parecido al caso anterior, aunque ahora no se hace referencia a las técnicas de reducción de estrés y puede que la variable que no se identifica sea la dependiente. También incluimos acá los que sólo definen con acierto la independiente, o nunca la definen o consideran a la independiente representada por los atletas. En la Tabla 5.6.12 se analiza un ejemplo.

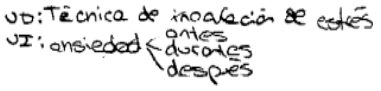
Tabla 5.6.12. Análisis semiótico de un ejemplo en la categoría PC2

| Expresión | Contenido |
|---|---|
|  | – Identifica correctamente la variable dependiente (concepto-definición), asignando a la expresión VD el contenido "niveles de ansiedad" (particularización de un concepto).<br>– *Conflicto* al identificar la expresión VI con el contenido "ansiedad" (concepto-definición-particularización).<br>– *Conflicto* en el contenido asignado a la expresión VI reconoce un único momento de medición, un único nivel (campo de problemas). |

**Respuestas incorrectas**

Sólo hemos encontrado una categoría de respuesta incorrecta. Los estudiantes definen la variable independiente como la ansiedad medida en los tres momentos, y la dependiente la definen como las técnicas que se enseñan para disminuir la ansiedad en atletas. Es decir sería la confusión entre variable dependiente e independiente encontrada, entre otros trabajos en el de Arteaga (1011). En la Tabla 5.6.13, se muestra un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.13. Análisis semiótico de un ejemplo en la categoría I

| Expresión | Contenido |
|---|---|
|  | – *Conflicto*: No identifica correctamente la variable dependiente (definición-particularización de un concepto).<br>– Con el contenido "Técnicas de inoculación de estrés" identifica la expresión VD, genera un *conflicto* puesto que VD mide la ansiedad de cada individuo para cada momento (concepto-definición).<br>– *Conflicto*: No identifica correctamente la variable independiente (definición-particularización de un concepto)<br>– Aclara para la expresión VI que la ansiedad se mide en tres momentos distintos, a saber: antes-durante-después. Genera un *conflicto* al considerar que para esta variable mide la ansiedad (conceptualización).<br>– *Conflicto* que se genera al confundir la VI con el contenido que se le debería dar a la expresión VD (definición-concepto). |

En la Tabla 5.6.14 se presenta una tabla de frecuencia y de porcentaje para las categorías de este apartado. Observamos que el porcentaje de estudiantes que responde correctamente es muy bajo (21,9%) y muy alto el porcentaje de no respuesta (17,6%). Es de esperar que con estos niveles no sea alto el porcentaje de estudiante que analice

correctamente los datos y aplique el modelo de análisis de varianza adecuado, ya que es requisito esencial la identificación de las variables y sus tipos.

Tabla 5.6.14. Frecuencias (y porcentajes) de respuestas al Problema 2. Variable dependiente e independiente

| Categorías | Frecuencia | % |
|---|---|---|
| Correctas | 49 | 21,9 |
| PC1. La dependiente es correcta y la independiente errónea | 110 | 48,7 |
| PC2. Define correctamente solo una de las dos variables | 8 | 4,0 |
| Incorrectas. Confunde las variables. | 17 | 7,6 |
| Sin responder | 40 | 17,9 |
| | 224 | 100,0 |

La mayoría de los estudiantes eligieron la categoría PC1 (48,7%), es decir describe correctamente la variable dependiente, pero indica que la independiente son las técnicas de inoculación de estrés. Resultó de una gran complejidad describir la variable independiente.

**Conflictos semióticos al identificar las variables del problema**

En el análisis semiótico de las respuestas incorrectas en los registros escritos de los estudiantes que se han ido detallando a lo largo del parágrafo 5.6.2, se han encontrado los siguientes conflictos semióticos (Godino, 2002):

- *Confusión de la variable tiempo de medición con el tratamiento técnico de reducción de estrés* (aparece en las categorías PC1 e I). Un 56,3% presenta un conflicto a la hora de determinar las variables tanto dependiente como independiente ya que, en (PC1) identifica a las técnicas de inoculación de estrés como la variable independiente y en (I) lo hace para la variable dependiente.
- *Confusión de la variable dependiente e independiente* (aparece en las categorías PC2 e I). Un 11,6% genera conflicto ya que reconoce en el contenido de la variable independiente el que se corresponde para la dependiente. En el caso de PC2 la identifica con la ansiedad, mientras que en I identifica con la ansiedad que la mediría en tres momentos. Como hemos indicado, ha sido encontrado en investigaciones previas, como la de Arteaga (2011).
- *Confunde el modelo a aplicar para resolver problema* (aparece en PC2, 4%). Genera conflicto ya que en el modelo de análisis de varianza de medidas repetidas, debería identificarse mínimo un factor con al menos tres niveles para la comparación de medias poblacionales. En el caso de esas respuestas confundiría con prueba de hipótesis de comparación de dos medias.

### 5.6.3 ELECCIÓN DE UN CONTRASTE

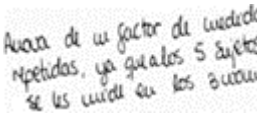
En este tercer apartado del problema, se pide al estudiante que elija cual es el contraste que debe aplicar para responder a la pregunta del enunciado. Deberá descubrir que el contraste a aplicar es un análisis de varianza, con un solo factor, que mide el nivel de ansiedad en tres momentos distintos a los mismos cinco sujetos, luego de aplicar ciertas técnicas previstas. Como todos los niveles del factor se aplican a los

mismos sujetos, se elige aplicar un modelo de análisis de varianza de un factor con medidas repetidas. Este punto es fundamental a la hora de explicar los resultados numéricos con los que se completa la tabla de análisis de varianza.

**Respuestas correctas**

*C. Elige el contraste adecuado para el problema en estudio.* En esta categoría el estudiante logra elegir el campo de problemas en el que debe incluir el suyo. Se trata de un análisis de varianza de un factor de medidas repetidas, además justifica su elección con la frase "ya que a los 5 sujetos se les mide en los tres momentos". Debe discriminar entre todos los modelos estudiados cual es el que debe aplicar para responder a la pregunta del enunciado. En la Tabla 5.6.15 se muestra un ejemplo de cuyo análisis se deduce que el estudiante ha interpretado correctamente la situación del enunciado, las variables dependiente e independiente y el tipo de análisis. El alumno argumenta por qué elige un modelo de medidas repetidas.

Tabla 5.6.15. Análisis semiótico de un ejemplo en la categoría C

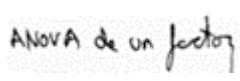
| Expresión | Contenido |
|---|---|
|  | – El alumno ha interpretado correctamente que se trata de un análisis de varianza de un solo factor (campo de problemas y definición).<br>– Reconoce debe utilizar un modelo de medidas repetidas (particularización).<br>– También explicita por que debe asignar este modelo, pues los sujetos son los mismos para recolectar los datos en los tres momentos (argumentación). |

**Respuestas parcialmente correctas**

Seguidamente se describen las categorías de respuestas que, aunque identifican gran parte de la respuesta adecuada, olvidan otras o confunden condiciones. Encontramos tres categorías para esta tipología.

*PC1. No específica que se debe utilizar el modelo de medidas repetidas.* Estos estudiantes tienen una respuesta parcial, ya que consideran que se trata de una prueba de análisis de varianza de un factor. En la Tabla 5.6.16 se muestra un ejemplo desarrollado por uno de los estudiantes y su análisis semiótico.

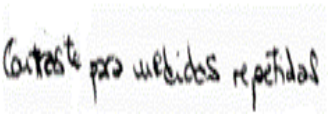
Tabla 5.6.16. Análisis semiótico de un ejemplo en la categoría PC1

| Expresión | Contenido |
|---|---|
|  | – El alumno ha interpretado correctamente que trata de un análisis de varianza (definición-concepto).<br>– También ha especificado que elige el modelo de un solo factor (campo de problemas).<br>– No puede discriminar que debe usar un modelo de medidas repetidas, razón por lo cual aparece un *conflicto* (particularización). |

*PC2. No especifica que se deba utilizar un modelo de análisis de varianza de un factor.* Consideramos también que estos estudiantes presentan una respuesta parcial. Aunque consideran que se trata de un análisis de varianza, no especifican que debe ser un modelo de un factor, limitándose a decir que es para medidas repetidas. También

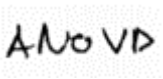
como en el último anterior hay una parte de la respuesta que se obvia, la respuesta también es parcial puesto que el análisis de varianza lo consideramos un tipo de contraste de hipótesis. En la Tabla 5.6.17 se muestra un ejemplo.

Tabla 5.6.17. Análisis semiótico de un ejemplo en la categoría PC2

| Expresión | Contenido |
|---|---|
|  | – El alumno ha interpretado correctamente que debe utilizar un contraste (definición-concepto).<br>– Elige correctamente utilizar un modelo de medidas repetidas (particularización).<br>– No puede especificar que el contraste elegido deberá ser un análisis de varianza, ocasiona un *conflicto* (campo de problemas).<br>– El *conflicto* descripto se relaciona con otro, pues no puede establecer que el análisis de varianza que se llevará a cabo es de un solo factor (campo de problemas). |

*PC3. Considera solamente que debe utilizar un modelo ANOVA.* La respuestas es parcial, pues si bien consideran que se trata de una prueba estadística de análisis de varianza, no especifican que es aquel que tiene un solo factor, ni que debe ser de medidas repetidas por la forma en que se toman las mediciones de la ansiedad. En la Tabla 5.6.18 se muestra un ejemplo de esta categoría.

Tabla 5.6.18. Análisis semiótico de un ejemplo en la categoría PC3

| Expresión | Contenido |
|---|---|
|  | – Ha interpretado correctamente que debe utilizar para su análisis algún modelo "ANOVA" (concepto-definición).<br>– No indica si se trata de un modelo con un factor o más (campo de problemas), aparece un *conflicto.*<br>– Tampoco declara si es algún modelo completamente aleatorizado o de medidas repetidas, no puede determinarlo, *conflicto* también en el campo de problemas |

**Respuestas incorrectas**

Seguidamente se describen las respuestas incorrectas. Los errores encontrados se refieren a utilizar ya sea un análisis de varianza de dos factores o un contraste de una media.

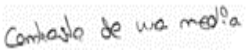
*I1. Especifica un modelo de análisis de varianza de dos factores.* Estos estudiantes han considerado un modelo de análisis de varianza de dos factores. El fallo recae en no tener en cuenta que existe para el problema una única fuente de variación, a saber: los momentos en que se toman las medidas de ansiedad. En la Tabla 5.6.19, se muestra un ejemplo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.19. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
| | – Ha interpretado correctamente el enunciado y reconoce que se trata de un modelo de análisis de varianza, utiliza la sigla "Anova" (particularización de un concepto).<br>– El estudiante genera un *conflicto* ya que con la sentencia "dos factores", alude a un campo de problemas que no es el que se corresponde con el dado.<br>– No puede reconocer que los datos se toman a los mismos sujetos en distintos tres momentos, *conflicto* por omisión (conceptualización). |

*I2. Especifica un modelo de contraste de una media.* Los estudiantes que han sido agrupados en esta categoría son los que han considerado utilizar un modelo de contraste de una sola media. Deberán comparar respuestas para tres momentos, siendo eso lo que debe especificarse a través de las hipótesis, pero lo confunden con el test de hipótesis de una sola muestra. En la Tabla 5.6.20, se analiza un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes.

Tabla 5.6.20. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
|  | – Identifica que su tarea se relaciona con la realización de un contraste (particularización de un concepto).<br>– No reconoce la presencia de un factor con niveles, ya que genera un *conflicto* al mencionar que se contraste sobre una sola media (campo de problemas).<br>– *Conflicto*: No reconoce que los niveles estarían relacionados (particularización).<br>– Entendemos realiza una interpretación inadecuada del enunciado, puesto que genera un *conflicto* al afirmar que la prueba a aplicar en un "contraste de una media".(campo de problemas).<br>– *Conflicto*: no puede deducir que se deben comparar los datos generados en tres momentos distintos (se trata de una situación-problema que conlleva otro algoritmo para su resolución). |

Seguidamente, se presenta la Tabla 6.5.21 con las frecuencias de respuestas por categoría en la muestra y sus respectivos porcentajes. El porcentaje de respuestas correctas ha sido de un 35,3%; si se añaden las parcialmente correctas (25,9%), donde el estudiante al menos ha identificado el análisis de varianza, aunque no llega a verbalizar el modelo concreto llegamos a un 61,2% de respuestas. Pensamos esto indica un buen reconocimiento del campo de problemas planteado. Estos resultados son originales, puesto que no disponemos de investigaciones previas para realizar una comparación.

Dentro de las respuestas parcialmente correctas, la mayoría opta por contestar que se trata de un problema de "Anova de medidas repetidas" (15,2%), da cuenta que los estudiantes se encuentran ubicados en cuanto al campo de problemas y discriminan entre factor inter sujeto y factor intra sujeto. La mayoría de las categorías para las incorrectas la tienen los que respondieron se trata de un "Anova de dos factores" (4,5%), aunque supone un número pequeño respecto al total. Alrededor del 30% no responde el problema.

Tabla 5.6.21. Frecuencias (y porcentajes) de respuestas al Problema 2-Elección de un contraste

| Categorías | Frecuencia | % |
|---|---|---|
| Correctas | 79 | 35,3 |
| PC1. Análisis de varianza de un factor | 15 | 6,7 |
| PC2. Contraste de medidas repetidas | 34 | 15,2 |
| PC3. Análisis de varianza | 9 | 4,0 |
| I1. Análisis de varianza de dos factores | 10 | 4,5 |
| I2. Contraste de una media | 5 | 2,2 |
| I3. No relacionadas con la tarea | 5 | 2,2 |
| Sin responder | 67 | 29,9 |
| | 224 | 100,0 |

**Conflictos semióticos al elegir un contraste**

En el análisis semiótico de las respuestas incorrectas en los registros escritos de los estudiantes que se han ido detallando a lo largo del parágrafo 5.6.2, se han encontrado los siguientes conflictos semióticos (Godino, 2002):

- *Confusión al no determinar que debe usar un modelo de medidas repetidas* (aparece en las categorías PC1, PC3, I1 e I2). Un 17,4% presenta un conflicto al no reconocer que debe utilizar un modelo de medidas repetidas por lo que no se diferencia un factor intra sujeto de otro intersujeto. En algunas respuestas confunde esta técnica con la del modelo completamente aleatorizado para un factor; confusión razonable porque tiene la misma técnica de cálculo, aunque diferente interpretación.
- *Confusión al no reconocer se trata de un contraste de análisis de varianza* (aparece en las categorías PC2 e I2). Un 17,4% de estudiantes muestra un conflicto ya que no puede reconocer que debe utilizar un análisis de varianza para la resolución de la tarea. Estas respuestas confunden con los contrastes de hipótesis para una o dos muestras.
- *Confusión al no establecer que el análisis de varianza para el caso es de un solo factor* (PC2, PC3, I1 e I2). Un 25,9% de estudiantes analizan el problema simplemente como un análisis de varianza, sin identificar el campo de problemas en el que estaría inserto, es decir: es de un factor, dos o más, completamente aleatorizado o de medidas repetidas. También dentro de este porcentaje están aquellos estudiantes que lo identifican con un análisis de varianza de dos factores.

## 5.6.4 CÁLCULO DE LAS REGIONES Y DECISIÓN SOBRE LAS HIPÓTESIS

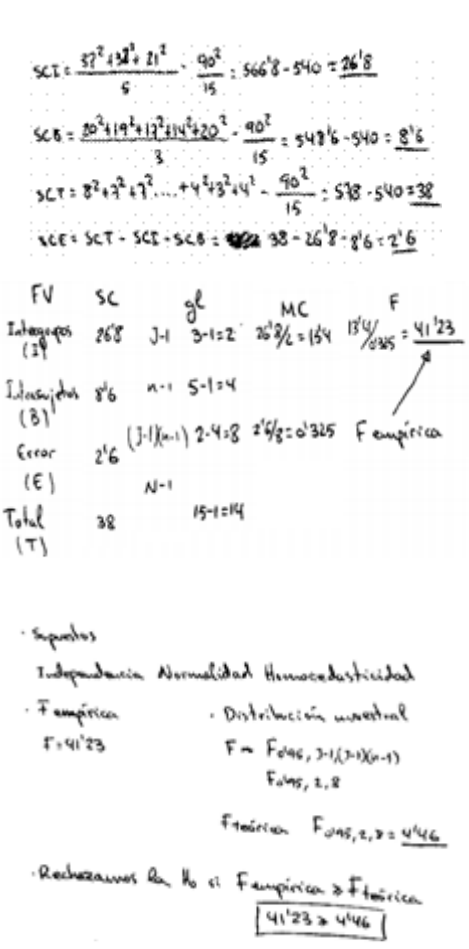
A continuación, se describen las categorías encontradas para el cuarto apartado considerado del problema 2, en el que el alumno ha de realizar una serie de pasos con la finalidad de tomar una decisión. En primer lugar deben completar una tabla de análisis de varianza, que responderá a las características del análisis de varianza para medidas repetidas. En segundo lugar y con el valor de un percentil teórico correspondiente a la distribución *F* de Fisher, según el nivel de la prueba y ciertos grados de libertad, GL. Finalmente determinará si el *F* empírico que emerge de la tabla, pertenece o no a la región de rechazo. De esta comparación se desprenderá la decisión que tome sobre las hipótesis.

**Respuestas correctas**

*C. Cálculos correctos, presentación de la tabla de análisis de varianza, fundamento de la toma de decisión.* Hemos encontrado sólo un tipo de respuesta correcta, donde el estudiante logra calcular claramente las cuatro sumas de cuadrados necesarias. Las sumas de cuadrados total, inter grupos y por grupos se determinan utilizando la fórmulas presentadas con el enunciado del problema, mientras que la del error mediante diferencia con estas. Deducen los grados de libertad de cada suma, las medias de cuadrados y el único valor *F* empírico que se debe analizar en este modelo. Encuentran correctamente el valor de la *F* teórica usando los grados de libertad adecuados para finalmente compararlos tomando la decisión correcta.

Del análisis semiótico pormenorizado de un ejemplo en la Tabla 5.6.22 emerge que el estudiante ha de encadenar una gran cantidad de *funciones semióticas,* pasando por la selección de fórmulas que utiliza para encontrar luego los procedimientos de cálculo de las sumas de cuadrado de cada fuente de variación, con las que determina: los grados de libertad, las medias de cuadrados para finalizar con el valor *F* empírico.

Tabla 5.6.22. Análisis semiótico de un ejemplo en la categoría C

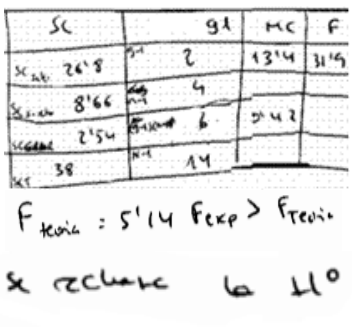
| Expresión | Contenido |
|---|---|
|  | – Identifica las fórmulas que debe utilizar para la tarea (particularización de un concepto).<br>– Obtiene la suma de cuadrados del error por diferencia entre la suma de cuadrados total y la suma de las sumas de cuadrados inter grupo y entre grupo (concepto-procedimiento).<br>– Confecciona correctamente la tabla ANOVA, especifica en columnas las fuentes de variación (FV), sumas de cuadrados (SC), grados de libertad (GL), media de cuadrados (MC) y el valor F empírico (F) (conceptos y representación).<br>– Particulariza en la primera columna de la tabla las fuentes de variación: inter grupos, inter sujetos, del error y total. En la segunda particulariza el contenido de cada suma de cuadrados. En la tercera columna genera los grados de libertad de cada fuente de variación (particularización de conceptos).<br>– En MC realiza el cociente de cada suma de cuadrados con sus respectivos GL para las dos necesarias, inter grupos y del error. En F, indica la *F* empírica como cociente entre las MC calculadas (procedimientos).<br>– Indica los supuestos necesarios para aplicar este modelo; independencia, normalidad y homocedasticidad (propiedades).<br>– Indica la distribución muestral, el nivel de confianza y GL que usa en el cálculo de la *F* teórica (particularización de conceptos). Determina correctamente el valor de la *F* teórica (procedimiento).<br>– Compara ambos valores de *F* hallados tanto empírico como teórico (procedimiento-argumentación).<br>– Utiliza la argumentación última anterior para tomar la decisión de rechazar la hipótesis nula (definición-concepto). |

Completada de este modo la tabla ANOVA, deberá encontrar el valor del percentil teórico de la distribución para compararlo con el valor *F* objetivo de la tabla confeccionada. Con todas estas funciones ensambladas, deberá tomar una decisión entre rechazar o no rechazar la hipótesis nula.

**Respuestas parcialmente correctas**

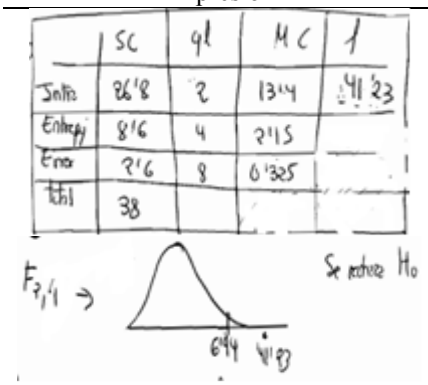
*PC1. Error en los grados de libertad de la suma de cuadrados del error.* Si bien el resto del procedimiento ha sido correcto, e incluso los ha llevado a tomar una decisión coherente con sus cálculos, no les ha quedado claro la forma en que se determinan los GL, lo que como vemos es esencial. Este fallo les trae como consecuencia errores en el cálculo del percentil de la distribución *F* teórica para compararlo con la empírica. En la Tabla 5.6.23 se detalla un análisis semiótico de la respuesta de un estudiante, quien también realiza los cálculos en su hoja de trabajo y así completa la tabla ANOVA que se presenta para ser analizada.

Tabla 5.6.23. Análisis semiótico de un ejemplo en la categoría PC1

| Expresión | Contenido |
|---|---|
|  | − Presenta una tabla de análisis de varianza adecuada (particularización de un procedimiento).<br>− Presenta un *conflicto* para el cálculo de los GL del error (procedimiento) pues suma los anteriores en lugar de hacer su producto.<br>− El error anterior ocasiona otro *conflicto* para la media de cuadrados del error (en vez de dividir la suma de cuadrados por 6, debería hacerlo por 8) (procedimiento).<br>− Encuentra un valor incorrecto para la *F* empírica (el valor 0,42 no es el correcto para MC del error) (procedimiento correcto para obtener *F*).<br>− De acuerdo con conflictos anteriores se obtiene el valor correcto del percentil (utiliza 2 y 6 GL, nivel 0,05).<br>− Compara ambos valores de *F*, argumenta con eso su decisión correcta de rechazar la hipótesis nula. |

*PC2. Error en el cálculo del percentil de la distribución F.* Estos estudiantes tampoco han podido dar una respuesta exitosa, ya que su fallo proviene de la confusión en los GL para determinar el valor de *F* teórico. Para obtener este valor son necesarios tres elementos a saber: los grados de libertad para el numerador, los correspondientes para el denominador y el nivel de la prueba. Los del denominador deben ser tomados de los de la suma de cuadrados del error, pero en este caso los toman desde los de la suma de cuadrados entre grupos, generando un valor incorrecto. El resto del procedimiento es el correcto, incluso toman una decisión coherente con los resultados, además muestra un gráfico de análisis para rechazar la hipótesis nula, ubicando los valores de las *F´*s (empírica y teórica). En la Tabla 5.6.24 se detalla un análisis semiótico usando un ejemplo de esta categoría desarrollado por uno de los estudiantes.

Tabla 5.6.24. Análisis semiótico de un ejemplo en la categoría PC2

| Expresión | Contenido |
|---|---|
|  | – Presenta una tabla ANOVA adecuada, especificando los contenidos en cada columna (campo de problemas).<br>– Los contenidos de la tabla son todos correctos (definición-conceptos).<br>– Aparece un *conflicto* al no usar para la *F* teórica como GL del denominador los de la suma de cuadrados del error, en cambio usa los de la entre grupos (concepto y procedimiento). Muestra un análisis gráfico adecuado ubicando los datos correctamente sobre el mismo (cambio de representación).<br>– En consecuencia decide rechazar la hipótesis nula (definición-campo de problemas). |

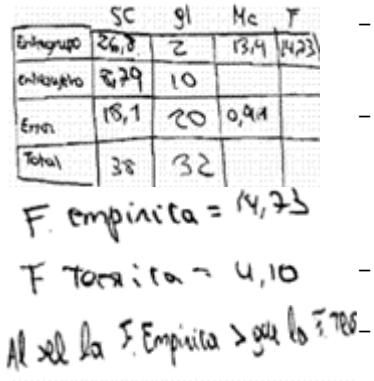


**Respuestas incorrectas en las que aparece la tabla de análisis de varianza**

Los errores encontrados en las respuestas incorrectas los clasificamos según incluyen o no la tabla de análisis de varianza. Describimos a continuación cada categoría, dentro de los que la incluyen.

*I1. Calcula incorrectamente la suma de cuadrados del error y la mayoría de los grados de libertad*. Estos estudiantes no obtienen la suma de cuadrados del error como diferencia, usando todas las otras sumas de cuadrados que son correctas; arrastran error de los GL entre sujetos. A pesar de esos fallos, los restantes cálculos son coherentes, pudiendo arribar a una decisión en consecuencia. En la Tabla 5.6.25, se muestra un ejemplo y se detalla un análisis semiótico.

Tabla 5.6.25. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
|  | – Presenta una tabla adecuada, especificando los contenidos en cada columna (procedimiento). Usa notación apropiada (lenguaje).<br>– *Conflicto* en suma de cuadrados del error, pues lo obtiene como diferencia entre suma de cuadrados entre grupos y entre sujetos (concepto y procedimiento).<br>– *Conflicto* para el cálculo de los GL de la suma de cuadrados entre sujetos (procedimiento), posiblemente lo obtenga como producto entre GL de suma de cuadrados entre grupos y cantidad de sujetos (2*5=10).Arrastra *conflicto* en el resultado de la media de cuadrados del error.<br>– El error anterior ocasiona otro *conflicto para* la *F* empírica (el valor 0,91 no es el correcto para MC del error)<br>– *Conflicto* para obtener el percentil (*utiliza 2* y 10 GL, nivel 0,05), los GL del denominador no son los de la suma de cuadrados del error (20) (concepto-definición-propiedad).<br>– Compara ambos valores de *F*, argumenta con eso su decisión correcta de rechazar la hipótesis nula. |



*I2. Resuelve como un análisis de varianza de un factor completamente aleatorizado.* Estos estudiantes no discriminan entre suma de cuadrados entre grupos e intra grupos. Eso se refleja tanto en la suma de cuadrados del error y en sus GL. Determinan la suma de cuadrados para el modelo elegido como la suma de cuadrados intra grupo más la del error (no tiene en cuenta el efecto grupo). Tampoco son correctos

los GL de la suma de cuadrados total (aunque no afecta al resultado). No calculan ninguno de los valores de las *F*. Toman una decisión sin coherencia ni argumentos.

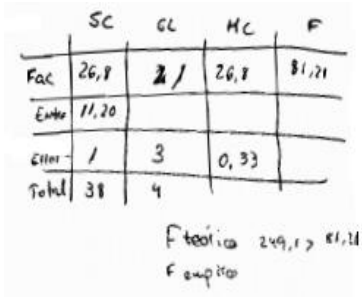
Tabla 5.6.26. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
| | – Considera un modelo de análisis de varianza completamente aleatorizado conflicto (confunde el campo de problemas).<br>– Otro *conflicto* es que no asume que existe un factor grupo (campo de problemas), debe discriminar las sumas de cuadrados en: entre grupos, intra grupos, error y total (conceptos).<br>– No usa las fórmulas facilitadas para la resolución junto con el problema (Expresión-contenido).<br>– Aparece un *conflicto* en el cálculo de los GL de la suma de cuadrados total (definición).<br>– No calcula ninguna medida para la *F*, ni empírica ni teórica, generando un *conflicto* (procedimiento).<br>– Tiene un *conflicto* pues toma una decisión sin basarse en la comparación entre la *F* empírica y la teórica (concepto-argumentación). |

En la Tabla 5.6.26 se detalla el análisis semiótico de un ejemplo para I2.

*I3. Lo resuelve como completamente aleatorizado y sólo es correcta la suma de cuadrados total.* Estos estudiantes pueden reconocer que deben aplicar un modelo de análisis de varianza de una vía, pero lo resuelven como modelo completamente aleatorizado. En la tabla presentada sólo es correcta la suma de cuadrados total, a pesar de estar completa. No utiliza las fórmulas dadas con el enunciado para su resolución, encuentra sus sumas de cuadrados siempre como sumas de cuadrados de diferencias. En la Tabla 5.6.27, se muestra un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.27. Análisis semiótico de un ejemplo en la categoría I3

| Expresión | Contenido |
|---|---|
|  | – Presenta una tabla ANOVA no adecuada, pues es la del modelo completamente aleatorizado de una vía. *Conflicto* con la elección del campo de problemas.<br>– No es correcto ninguno de los GL, *conflicto* (concepto-definición).<br>– Para determinar *F* teórica para sus GL usa 10 y 1 incoherente con los valores de su tabla (técnicas de determinación), $\alpha = 0.05$ es correcto.<br>– Si bien el procedimiento para la *F* empírica es correcto, su contenido no es el adecuado (expresión-contenido) genera *conflicto*.<br>– *Conflicto* al no materializar una decisión (ejemplar-argumentación). |

*I4. Modelo completamente aleatorizado, tabla de análisis de varianza incompleta.* Los estudiantes que han sido agrupados en esta categoría plantean el problema tratando de calcular una suma de cuadrados total, como si fuese una suma de cuadrados entre

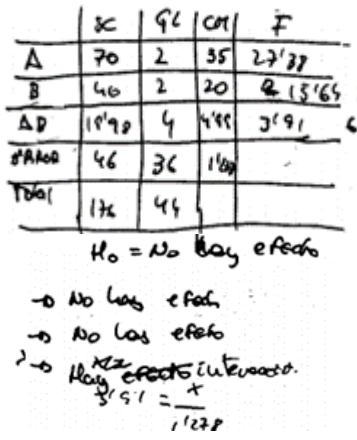
para el modelo completamente aleatorizado. Los GL que emergen de su tabla ANOVA son correctos, pero no completan la tabla, ni toman decisión. En la Tabla 5.6.28, se muestra y analiza un ejemplo de respuesta.

*I5. Aplica un modelo de análisis de varianza de dos factores con interacción.* Estos estudiantes plantean el problema como modelo de análisis de varianza de dos factores, incluyendo además la interacción, es decir un modelo completo en ese campo de problemas. Dentro de este modelo no resultan correctos sus cálculos que se reflejan en la tabla de análisis de varianza presentada. En la Tabla 5.6.29, se muestra un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.28. Análisis semiótico de un ejemplo en la categoría I4

| Expresión | Contenido |
|---|---|
| | – *Conflicto* resuelve como modelo de un factor completamente aleatorizado (campo de problemas).<br>– Intenta calcular la suma de cuadrados total, como suma de cuadrados dentro para el modelo que se eligió.<br>– A cada dato le resta la media de su grupo, eleva el resultado al cuadrado y suma, *conflicto* (definición). No usa las formulas brindadas con el problema.<br>– Presenta una tabla incompleta *conflicto* (procedimiento), solo dos sumas de cuadrado, los grados de libertad y una media de cuadrados; esta última incorrecta.<br>– No calcula valores para $F$ ni empírica ni teórica, *conflicto*.<br>– El *conflicto* anterior arrastra otro, no puede tomar alguna decisión sobre las hipótesis (argumentación-materialización). |

Tabla 5.6.29. Análisis semiótico de un ejemplo en la categoría I5

| Expresión | Contenido |
|---|---|
|  | – *Conflicto* puesto que no presenta una tabla ANOVA adecuada (campo de problemas) resuelve como un modelo de dos factores con interacción.<br>– Los GL son incorrectos, los cuadrados medios correctos (procedimientos y conceptos),<br>– Los valores $F$ empíricos son incorrectos *conflicto* (expresión-contenido), los cálculos son coherentes con los errores (procedimiento).<br>– *Conflicto* (definición-concepto) generado al calcular los valores $F$ teóricos. El 3,35 corresponde a GL 2; 27. El 3,63 a GL 2; 16 y el 6,39 a GL 4; 4. En todos los casos sólo coincide uno de los GL, pero para todos usa el nivel $\alpha = 0{,}05$.<br>– *Conflicto* conceptual, decide no rechazar $H_o$ cuando debería aceptarla y viceversa (argumentación) tanto para los efectos principales que presenta como para la interacción. |

**Respuestas incorrectas en las que no aparece la tabla de análisis de varianza**

Entre las respuestas incorrectas para las que el estudiante no presenta una tabla de análisis de varianza hemos encontrado dos categorías.

*I6. Sólo intenta calcular la suma de cuadrados total*. Los estudiantes que han sido agrupados en esta categoría responden como los de la categoría I3, pero no presentan una tabla de análisis de varianza relacionada. Observamos que intenta solamente obtener la suma de cuadrados total, incluso con errores de cálculo. Esta es toda la respuesta que brindan a este apartado. En la Tabla 5.6.30, se muestra un ejemplo y un análisis semiótico.

Tabla 5.6.30. Análisis semiótico de un ejemplo en la categoría I6

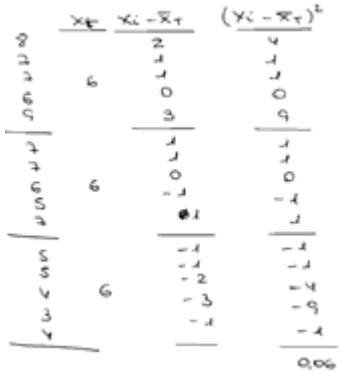
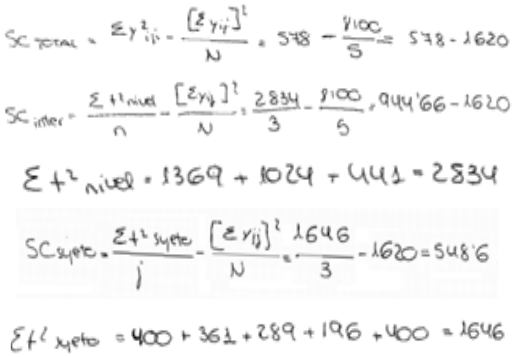
| Expresión | Contenido |
|---|---|
|  | – Determina correctamente la media global, cuyo valor es 6 (concepto-procedimiento).<br>– Intenta calcular la suma de cuadrados total, a cada valor le resta la media global, los eleva al cuadrado y los suma con errores en los cálculos *conflicto semiótico* (propiedad-procedimiento).<br>– *Conflicto semiótico* pues no reconoce que al elevar al cuadrado un número negativo, su resultado se transforma en positivo (particularización).<br>– No intenta presentar ninguna de las otras sumas de cuadrados, tampoco las medias de cuadrados con sus grados de libertad. |

Tabla 5.6.31. Análisis semiótico de un ejemplo en la categoría I7

| Expresión | Contenido |
|---|---|
|  | – Intenta adecuadamente utilizar las fórmulas dadas con el problema para la suma de cuadrados total ($SC_{Total}$), aunque utiliza como medida total de datos el de los sujetos (5 en lugar de 15) *conflicto* (procedimiento).<br>– En la determinación de la suma de cuadrados inter sujetos ($SC_{Inter}$) presenta un nuevo *conflicto* (confunde el número de niveles con número de observaciones por nivel). Para este cálculo arrastra el *conflicto* último anterior en esta categoría.<br>– Para la medida suma de cuadrados entre sujetos $\left(SC_{Sujetos}\right)$ considera la suma de los cuadrados de la suma de observaciones por cada sujeto. Arrastra error realizado en el cálculo de $SC_{Total}$<br>– No calcula ni $SC_{Error}$, ni compone la tabla para el análisis de varianza, *Conflicto* (procedimiento). |

*I7. Sólo calcula la suma de cuadrados utilizando las fórmulas dadas y arrastrando errores.* Concentramos en esta categoría todos aquellos estudiantes que intentan obtener las sumas de cuadrados para completar la tabla y utilizan las fórmulas que se dieron con el enunciado. Para la suma de cuadrados total cometen error en donde debe considerar la cantidad total de datos obtenidos (sujetos x medidas), utilizando en

vez la cantidad de sujetos, este fallo es arrastrado para el resto de las sumas de cuadrados, no logrando algún resultado certero. En la Tabla 5.6.31, se muestra un ejemplo.

Seguidamente se presenta la tabla de frecuencias por categoría. El porcentaje de respuestas correctas según la Tabla 5.6.32 ha sido de un 14,3% lo consideramos bajo, es muy similar al encontrado para las categorías parcialmente correctas, 14,7%. Contando las dos tenemos un 29% de respuestas aceptables, teniendo en cuenta la complejidad de la tarea. Tampoco en este apartado tenemos antecedentes de comparación.

Tabla 5.6.32. Frecuencias (y porcentajes) de respuestas al Problema 2: Regiones y decisión

| Categorías | Frecuencia | % |
|---|---|---|
| Correctas | 32 | 14,3 |
| PC1. Grados de libertad de la suma de cuadrados del error | 24 | 10,7 |
| PC2. Percentil de la distribución *F* | 9 | 4,0 |
| CON TABLA ANOVA | | |
| I1. Suma de cuadrados del error y grados de libertad | 15 | 6,7 |
| I2. Completamente aleatorizado | 20 | 8,9 |
| I3. Completamente aleatorizado; solo correcta la SCT | 5 | 2,2 |
| I4. Completamente aleatorizado, tabla ANOVA incompleta | 4 | 1,8 |
| I5. Análisis como dos factores con interacción | 1 | 0,4 |
| SIN TABLA ANOVA | | |
| I6. Solo intenta calcular la suma de cuadrados total | 29 | 12,9 |
| I7. Calcula suma de cuadrados con fórmulas dadas y arrastra errores | 10 | 4,5 |
| Sin responder | 75 | 33,5 |
| | 224 | 100,0 |

Las categorías incorrectas han sido clasificadas según los que presentan una tabla de análisis de varianza y los que no; los primeros representan un 20% de la muestra, mientras que los otros un 17,4%. Tengamos en cuenta que una tercera parte de la muestra no responde a este apartado, lo consideramos alto.

De los que presentan una tabla análisis de varianza entre las incorrectas, el más alto porcentaje (12,9%) son los que responden como si se tratase de un problema completamente aleatorizado, es decir confunden el campo de problemas. Un alto porcentaje (12,9%) todo lo que da por respuesta a este apartado es el cálculo de la suma de cuadrados total, parece no poder discriminar la presencia de otros tipos de variaciones en el problema. Veamos a continuación los conflictos encontrados en este apartado.

### 5.6.1. Conflictos semióticos en el cálculo de las regiones y decisión sobre las hipótesis

En el análisis semiótico de las respuestas incorrectas (Godino, 2002) de los registros escritos de los estudiantes que se han ido detallando a lo largo del parágrafo 5.6.4, se han encontrado los siguientes conflictos semióticos:

- *Confusión al determinar los grados de libertad (GL) de la suma de cuadrados del error* (aparece en las categorías PC1, I1, I2, I3, I5, I6). Un 41,8% presenta este conflicto, ya sea porque suma los grados de libertad de las sumas de cuadrados entre grupos e inter grupos, o arrastra error habiendo calculado previamente

incorrectamente los GL de la suma de cuadrados entre sujetos, o no lo determina. También en el trabajo de Cañadas (2012) son bastantes los estudiantes que confunden los grados de libertad en el estadístico Chi-cuadrado. Deducimos que el concepto de grados de libertad no es tan sencillo como aparenta.

- *Confusión en el cálculo de la media de cuadrados del error* (aparece en las categorías PC1, I1 e I5). Un 17,8% de estudiantes presenta un conflicto al mostrar el resultado de este cálculo, ya que arrastran el error cometido en los GL de la $SC_{Error}$ o para el cálculo mismo de la $SC_{Error}$ .
- *Confusión en el cálculo de la F empírica* (aparece en las categorías PC1, I2, I3, I4, I5 e I6). Un 36,9% de estudiantes no logran calcular correctamente la *F* empírica ya sea por que arrastran error desde el cálculo de GL, directamente no la obtienen o los cálculos son coherentes al ensamblar con los anteriores (GL y medias de cuadrados).
- *Confusión en el cálculo de la F teórica* (aparece en las categorías PC1, PC2, I1, I2, I4 e I5). Un 32,5% de estudiantes no obtienen la *F* teórica correctamente pues arrastran error por un mal cálculo de los GL, toman los GL de otra suma de cuadrados y no del error, directamente no la obtienen, usan grados de libertad que no son los correctos, o también es correcto solamente el del denominador.
- *Confusión en el campo de problemas* (aparece en las categorías I2, I3, I4, I5 e I6). Un 26,2% de estudiantes no consideran el problema como un modelo de análisis de varianza con medidas repetidas, yerran en el campo de problemas elegido, ya sea por considerarlo como un análisis de varianza completamente aleatorizado, sin tener en cuenta la presencia de un factor grupo, o lo resuelven como modelo de dos factores con interacción.
- *Confusión en el cálculo de la suma de cuadrados* (aparece en las categorías I1, I6 e I7). Un 24,1% de estudiantes calculan algunas de las sumas de cuadrados incorrectamente, ya sea porque al determinar la $SC_{Error}$ lo hacen como la diferencia entre $SC_{Entregrupos}$ y $SC_{Entresujetos}$ , o solo intentan determinar la suma de cuadrados total y comenten errores en cálculos elementales aritméticos o no aplican correctamente las fórmulas facilitadoras.
- *Confusión al momento de tomar la decisión sobre si rechazar o no la hipótesis nula* (aparece en las categorías desde I1 hasta I5). Un 20% de estudiantes toma decisión sin algún fundamento, o no la materializa, o decide rechazar la hipótesis nula cuando en realidad de acuerdo con los resultados de sus cálculos no debería hacerlo. Como indicamos en el problema 1 esta confusión también aparece en nuestro estudio en el contraste de hipótesis y en trabajos previos de Vallecillos (1994) y Cañadas (2012).

### 5.6.5 INTERPRETACIÓN DE RESULTADOS

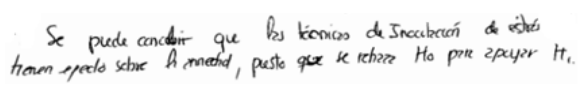
Describimos ahora las respuestas encontradas para el quinto y último apartado del problema, en el cual el estudiante he de interpretar los resultados obtenidos en términos del fenómeno que se estudia. Del análisis estadístico que realizó, en el apartado último anterior, debería rechazar la hipótesis nula para la igualdad de medias (antes, durante y después de la enseñanza de las técnicas). De acuerdo con esa decisión, en este apartado deberá responder que *la ansiedad cambia cuando se enseña a los sujetos esas técnicas.*

**Respuestas correctas**

Decidimos presentar tres categorías para las respuestas correctas ya que, varían en su redacción, y de ello depende además la calidad de la respuesta. Los del primer caso responden en términos del efecto de las técnicas aplicadas sobre la ansiedad, los del segundo refieren que existen diferencias en el nivel de ansiedad en las distintas mediciones, mientras que los últimos responden utilizando algunas claves. En todos los casos se completa el ciclo de modelización propuesto por Henry (1997).

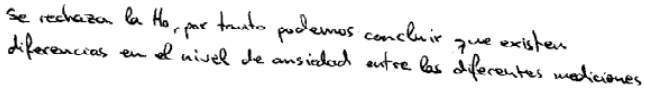
*C1. Las técnicas tienen efecto sobre la ansiedad.* En esta categoría el estudiante logra enmarcar la decisión que tomo al rechazar la hipótesis nula. Concluye que las técnicas de inoculación del estrés tienen efecto sobre la ansiedad, afirmando esa respuesta luego de rechazar la hipótesis nula en apoyo de la alternativa. En la Tabla 5.6.33 se muestra un ejemplo de esta categoría, desarrollado por uno de los estudiantes y se detalla un análisis semiótico. No obstante, como advertimos anteriormente, puede haber una confusión en esta respuesta entre la técnica utilizada y el momento de medición. Como acá solo valoramos la interpretación, la tomamos como correcta.

Tabla 5.6.33. Análisis semiótico de un ejemplo en la categoría C1

| Expresión | Contenido |
|---|---|
|  | – El estudiante argumenta la decisión a tomar (proceso de argumentación con símbolos adecuados, $H_0$ y $H_1$).<br>– Reconoce que la decisión debe ser rechazar la hipótesis nula para apoyar la alternativa (concepto).<br>– Traduce la decisión al contexto del problema: "las técnicas de inoculación de estrés tienen efecto sobre la ansiedad (contextualización de resultados). |

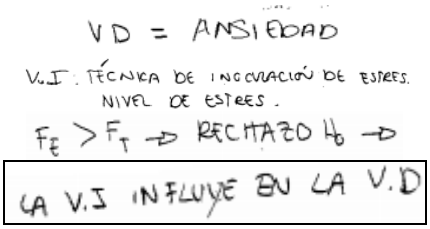
*C2. Existen diferencias en el nivel de ansiedad.* En esta categoría agrupamos aquellos estudiantes que, luego de rechazar la hipótesis nula concluyen que existen diferencias en el nivel de ansiedad entre las diferentes mediciones que se tomaron, relaciona las hipótesis estadísticas. En la Tabla 5.6.34 se muestra un ejemplo de esta categoría, desarrollado por uno de los estudiantes y se detalla un análisis semiótico. Vemos que el estudiante hace referencia a las mediciones que se hicieron.

Tabla 5.6.34. Análisis semiótico de un ejemplo en la categoría C2

| Expresión | Contenido |
|---|---|
|  | – El estudiante argumenta la decisión a tomar (proceso de argumentación, con símbolos, $H_0$).<br>– Reconoce que la decisión debe ser rechazar la hipótesis nula en forma explícita con la frase: "*se rechaza* $H_0$*"* (concepto), argumentado de ello su conclusión.<br>– Traduce la decisión al contexto del problema: "existen diferencias en el nivel de ansiedad en diferente niveles (contextualización de resultados). |

*C3. La VI influye en la VD.* En esta categoría el estudiante define mediante los símbolos VI y VD respectivamente cuales son, a su entender tanto la variable independiente como la dependiente respectivamente. Utiliza una doble expresión condicional indicando que la *F* empírica es mayor que la *F* teórica como consecuencia de ello, decide rechazar la hipótesis nula y de esto se desprende que la variable independiente (a su entender, las técnicas de inoculación de estrés) influye sobra la variable dependiente (la ansiedad). La respuesta está cargada de simbolismo, pero responde de acuerdo con la expectativa esperada. En la Tabla 5.6.35 se muestra un ejemplo de esta categoría, desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.35. Análisis semiótico de un ejemplo en la categoría C3

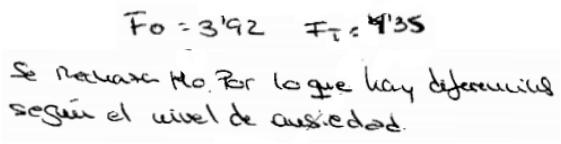
| Expresión | Contenido |
|---|---|
|  | – El estudiante ha definido la variable dependiente y la independiente, a su entender. *Conflicto* con la elección de la independiente (concepto) ya que mezcla las técnicas con los niveles de estrés.<br>– Simboliza ambos tipos de variables presentes en el problema, VD para dependiente y VI para independiente (proceso de significación).<br>– Interpreta correctamente que debe tomar una decisión argumentando con los valores de *F* calculados (proceso de argumentación, definición-concepto).<br>– Deduce a partir de lo último anterior que se rechaza la hipótesis nula (proceso de argumentación utilizando símbolos, $H_0$).<br>– Concluye en términos del problema utilizando los símbolos creados (proceso de comunicación). |

**Respuestas incorrectas**

No hemos podido encontrar respuestas parcialmente correctas. Hemos agrupado en cinco categorías las respuestas incorrectas.

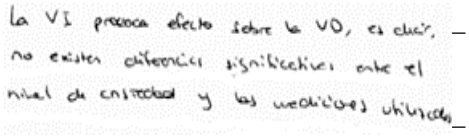
*I1. Incoherente de acuerdo con tabla ANOVA presentada.* Los estudiantes que han sido agrupados en esta categoría son los que han encontrado desde su análisis estadístico un valor para la *F* empírica más pequeño que el de la teórica, pero han decidido rechazar la hipótesis nula. Esta decisión resulta incoherente con los resultados del análisis estadístico. En la Tabla 5.6.36 se muestra un ejemplo de respuesta de este e implica una falta de comprensión del razonamiento subyacente en un contraste de hipótesis (Batanero, 2000).

Tabla 5.6.36. Análisis semiótico de un ejemplo en la categoría I1

| Expresión | Contenido |
|---|---|
|  | – El estudiante ha obtenido valores incorrectos de *F, conflicto*.<br>– *Conflicto,* decide rechazar la hipótesis nula, aunque obtuvo una *F* teórica más grande que la *F* empírica (argumentación, concepto).<br>– *Conflicto,* en la conclusión, no puede relacionar la variable dependiente con la independiente, solo aparece el nivel de ansiedad y no el momento en que se aplica la técnica. |

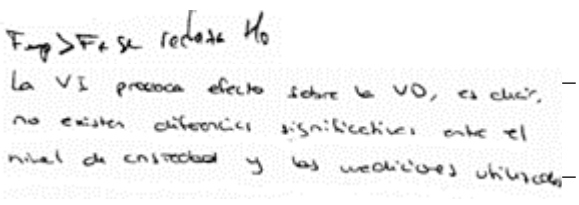
*I2. Hay efecto pero no hay diferencias en el nivel de ansiedad.* Los estudiantes que han sido agrupados en esta categoría son los que luego de obtener resultados que le permiten rechazar la hipótesis nula, de indicar cuál es la VI (variable independiente) y cual la VD (variable dependiente), deciden que hay efecto de una variable sobre la otra, pero concluyen que *no* hay diferencias significativas en el nivel de ansiedad. Presentan un contrasentido puesto que si hay efecto es porque existen diferencias significativas que los han llevado a rechazar $H_0$. Los estudiantes no comprenden el concepto de significación estadística (Vallecillos, 1994; Batanero, 2000). En la Tabla 5.6.37, se muestra un ejemplo de respuesta de este tipo desarrollado por uno de los estudiantes y se detalla un análisis semiótico.

Tabla 5.6.37. Análisis semiótico de un ejemplo en la categoría I2

| Expresión | Contenido |
|---|---|
|  | – Como en C3, simboliza la variable independiente y la dependiente, asigna una expresión a un contenido (proceso de significación).<br>– Afirma que la variable independiente provoca efecto sobre la dependiente (las técnicas aplicadas provocan efecto sobre la ansiedad) (particularización).<br>– Deduce que no hay diferencia significativa entre el nivel de ansiedad y las mediciones realizadas, *conflicto,* no se trata de las mediciones, sino de los momentos en que se mide (argumentación).<br>– Se genera un *conflicto* entre el antecedente y el consecuente de la proposición, ya que no se deduce el segundo del primero (proceso de interpretación-argumentación). |

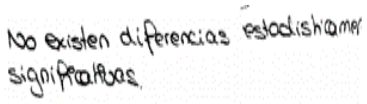
*I3. Las técnicas no tienen efecto sobre la ansiedad, a pesar de rechazar la hipótesis nula.* Estos estudiantes, a pesar de completar correctamente su tabla de análisis de varianza y concluir que se rechaza la hipótesis nula, finalizan decidiendo que las técnicas de inoculación de estrés no tienen efecto sobre la ansiedad precompetitiva en atletas. Nuevamente como en I2, se presenta una contradicción y falta de comprensión de la lógica del contraste de hipótesis. En la Tabla 5.6.38, se muestra un ejemplo de respuesta.

Tabla 5.6.38. Análisis semiótico de un ejemplo en la categoría I3

| Expresión | Contenido |
|---|---|
|  | – Realiza una tabla de análisis de varianza, obtiene que la $F$ empírica es mayor que la $F$ teórica (procedimientos).<br>– Deduce que rechaza la hipótesis nula (concepto y proceso de argumentación utilizando símbolos, $H_0$)<br>– *Conflicto* concluye en términos del problema pero es incoherente con los resultados que ha deducido de su tabla ANOVA (particularización de conceptos). |

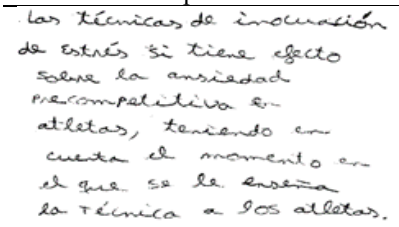
*I4. No rechaza Ho, no existen diferencias significativas.* Los estudiantes que han sido agrupados en esta categoría son los que luego de haber presentado errores en el proceso de cálculo de la tabla de análisis de varianza, los resultados los llevan a no rechazar la $H_0$, y de allí concluyen coherentemente que no existen diferencias significativas. Si bien la decisión es consecuente con sus resultados, esta es errónea puesto que proviene de un error anterior. En la Tabla 5.6.39, se muestra un ejemplo.

Tabla 5.6.39. Análisis semiótico de un ejemplo en la categoría I4

| Expresión | Contenido |
|---|---|
|  | – Admite no haber encontrado diferencias estadísticamente significativas (propiedad), se debe a errores en el proceso de cálculo al completar la tabla de análisis de varianza, *conflicto* (procedimiento).<br>– *Conflicto*, pues no explicita entre que variables no hay las diferencias significativas (proceso de comunicación).<br>– *Conflicto,* no logra responder en términos del contexto del problema (proceso de argumentación). |

*I5. Responde sin realizar el análisis estadístico.* Los estudiantes que han sido agrupados en esta categoría son los que no han considerado un análisis estadístico para responder el apartado, solo de observar los datos dan alguna respuesta, en algunos casos pertinente en otros no. En la Tabla 5.6.40, se muestra un ejemplo.

Tabla 5.6.40. Análisis semiótico de un ejemplo en la categoría I5

| Expresión | Contenido |
|---|---|
|  | – Responde en términos del problema (contextualización de resultados).<br>– Este resultado no se desprende del análisis estadístico, *conflicto* del proceso de análisis (procedimiento). |

Se presenta en la Tabla 5.6.41 un resumen con todas las categorías encontradas para este apartado con las frecuencias respectivas. El porcentaje de respuestas correctas ha sido de un 25%; muy parecido al que dieron una interpretación correcta en el problema 1; por tanto asumimos que la capacidad para completar el último paso del proceso de modelización (Henry, 1997) no depende del problema en sí. Se podrían añadir el 3,6% de la categoría I4, que hicieron errores de cálculo pero interpretan correctamente sus resultados.

Hubo un gran índice de no respuesta que fue del 58,5%, mucho mayor que en el primer problema, de lo que deducimos mayor dificultad de interpretación en este caso. Las respuestas incorrectas representan el 16,5%. De estas el porcentaje más alto se corresponde con los estudiantes que responden que las técnicas de inoculación de estrés utilizadas no tienen efecto sobre la ansiedad, creemos que haber conseguido este resultado es natural ya que exactamente es la contraria a la respuesta correcta, pues la aplicación de técnicas produce cambios en el nivel de ansiedad. Se deduce la confusión entre variable dependiente e independiente ya descrita.

Tabla 5.6.41. Frecuencias (y porcentajes) de respuestas al Problema 2-Apartado 5

| Categorías | Frecuencia | % |
|---|---|---|
| C1. Las técnica tienen efecto sobre la ansiedad | 21 | 9,4 |
| C2. Existen diferencias en el nivel de ansiedad | 18 | 8,0 |
| C3. La VI influye en la VD | 17 | 7,6 |
| I1. Incoherente de acuerdo con tabla ANOVA presentada | 7 | 3,1 |
| I2. Hay efecto pero no hay diferencias en el nivel de ansiedad | 3 | 1,3 |
| I3. Las técnicas no tienen efecto sobre la ansiedad | 14 | 6,3 |
| I4. No rechaza Ho, no existen diferencias significativas | 8 | 3,6 |
| I5. Responde sin realizar el análisis estadístico | 5 | 2,2 |
| Sin responder | 131 | 58,5 |
| | 224 | 100,0 |

**Conflictos semióticos en la interpretación de resultados**

A lo largo del parágrafo 5.6.6, hemos venido desarrollando un análisis semiótico de las respuestas incorrectas en los registros escritos de los estudiantes (Godino, 2002). Seguidamente presentamos en detalle el elenco de conflictos semióticos encontrados:

- *Confusión al completar la tabla de análisis de varianza* (aparece en las categorías en las categorías desde I1 hasta I4). Un 14,3% presenta errores, relacionados en con los grados de libertad para las sumas de cuadrados, o en el cálculo de la suma de cuadrados, lo que se transmite a las medias de cuadrados y para determinar tanto la *F* empírica como la teórica. Se generan así incoherencias en la toma de decisiones.
- *Confusión al definir variables* (aparece en las categorías I2 e I4). Un 4,9% de estudiantes presenta errores al interpretar los resultados pues confunde las mediciones con los momentos en que se las hace y no puede definir claramente la variable independiente.
- *Confusión al relacionar las variables* (aparece en las categorías I1 e I4). Un 6,7% de estudiantes cuando comunica su interpretación no puede relacionar la variable dependiente e independiente, o no las puede explicitar claramente. En general confunden nivel de ansiedad, técnicas de inoculación al estrés y momentos en que hacen las mediciones.
- *Confusión al interpretar, pues usan solo los valores de los datos* (aparece en la categoría I5). Un 2,2% no realiza análisis estadístico, pero si responde a este apartado usando la intuición que se desprende de los valores medidos para la ansiedad sin realizar el cálculo para completar la tabla de análisis de varianza.
- *No usa el contexto para responder,* aparece en la categoría I4 en un 3,6%. Este porcentaje de estudiante presenta errores, puesto que responden sin tener en cuenta la fenomenología que da contexto a la tarea.
- *Confusión deducir efecto si no hay diferencias entre los factores* (aparece en la categoría I2). Se trata de un 1,3% de estudiantes que si bien deducen, con errores, que no existe diferencia significativa (lo cual implicaría que las técnicas no producen efecto sobre la ansiedad, es decir el nivel medio de la ansiedad será estadísticamente el mismo en los tres momentos), finalizan respondiendo que las técnicas provocan efecto sobre la ansiedad.

### 5.6.6 SÍNTESIS DE RESULTADOS EN EL PROBLEMA 2

Para finalizar el estudio del segundo problema presentamos en la Tabla 5.6.42 los porcentajes de respuestas correctas, parcialmente correctas e incorrectas en cada apartado.

Tabla 5.6.42. Porcentaje de tipos de respuestas en cada apartado

| Apartado | Correctas | Parcialmente correctas | Incorrectas | No responde |
|---|---|---|---|---|
| 1. Definición de las hipótesis estadísticas | 25,0 | 4,9 | 45,1 | 25,0 |
| 2. Variable dependiente e independiente | 21,9 | 52,7 | 7,6 | 17,9 |
| 3. Elección del contraste | 35,3 | 25,9 | 8,9 | 29,9 |
| 4. Calculo de las regiones y decisión sobre las hipótesis | 14,3 | 14,7 | 37,4 | 33,5 |
| 5. Interpretación de resultados | 25,0 | 0,0 | 16,5 | 58,5 |

Los valores de los porcentajes entre las respuestas correctas vemos se encuentran bastante equilibrados aunque bajos, en especial los que corresponden con el primero, el segundo y el último apartado; esto es razonable porque los errores en estos apartados se arrastran en el resto del problema. La baja mayor se produce en el apartado 4 (14,3%) por la gran cantidad de funciones semióticas que conlleva su resolución. El mayor número de respuestas correctas en el apartado tres se debe a que el alumno reconoce fácilmente el campo de problemas, aunque plantee mal las hipótesis correspondientes o no diferencie la variable dependiente e independiente. Del mismo modo, algunos estudiantes que han fallado en el aspecto de cálculo (apartado 4) realizan una contextualización correcta de sus resultados, por lo que el porcentaje de correctas en el último apartado es el mismo que en el primero.

El porcentaje más alto entre las del tipo parcialmente correcto se corresponde con el apartado 2 (52,7%) más de la mitad de la muestra puede discriminar entre los dos tipos de variables del problema. Esta respuesta los a lleva pensar que los cambios en el nivel de ansiedad podrían deberse a los distintos momentos en los que se hace la medición, sin embargo al definirlas algunos confunden la variable momento de medición con la técnica experimental aplicada. Otro porcentaje importante tiene respuestas parcialmente correctas en la elección del contraste, debido a que olvidan mencionar alguno de los supuestos, por ejemplo, el tratarse de un solo factor o ser de medidas repetidas.

El desequilibrio en valores grandes para las respuestas incorrectas se percibe en el valor para el apartado 4 (37,4%) resultados que se esperaban por la gran complejidad entre las funciones semióticas que emergen y se ensamblan entre los procedimientos y argumentos para resolverlo y, en el apartado 1 (45,1%), debido a la gran cantidad de errores de interpretación de las hipótesis estadísticas ya descritos por Vallecillos (1994) y Cañadas (2012).

El apartado 5 tienen un gran nivel de no respuesta (58,5%), le sigue no tan lejos el 4 y luego el 3. Los apartados 1 y 2 tienen los más bajos valores de porcentajes de no respuesta los que se equilibran con los valores para las respuestas correctas.

## 5.7 COMPARACIÓN DE RESULTADOS EN LOS DOS PROBLEMAS

Tanto el problema 1, que se relaciona con los prerrequisitos para el análisis de varianza, como el problema 2 que abarca ese contenido estadístico para uno de sus modelos estadísticos, comparte cuatro de los cinco apartados a saber: definición de las hipótesis estadísticas, elección del contraste adecuado, cálculo de las regiones y decisión sobre las hipótesis e interpretación de resultados. Para finalizar, en la Tabla 5.6.43 comparamos los resultados en los apartados comunes a los dos problemas.

Tabla 5.6.43. Porcentaje comparativo de tipos de respuestas en cada apartado

| | Correctas | | P. Correctas | | Incorrectas | | No responde | |
|---|---|---|---|---|---|---|---|---|
| Problema / Apartado | P1 | P2 | P1 | P2 | P1 | P2 | P1 | P2 |
| - Definición de las hipótesis estadísticas | 33,0 | 25,0 | 0,0 | 4,9 | 50,5 | 45,1 | 16,5 | 25,0 |
| - Elección del contraste | 1,3 | 35,3 | 41,9 | 25,9 | 31,4 | 8,9 | 25,4 | 29,9 |
| - Calculo de las regiones y decisión sobre las hipótesis | 20,5 | 14,3 | 35,3 | 14,7 | 21,4 | 37,4 | 22,8 | 33,5 |
| - Interpretación de resultados | 24,6 | 25,0 | 39,3 | 0,0 | 6,2 | 16,5 | 29,9 | 58,5 |

En la definición de las hipótesis hay un mayor porcentaje de respuestas correctas en el contraste de hipótesis, frente al análisis de varianza; sin embargo estas diferencias se equilibran si se tiene en cuenta las particularmente correctas. En conjunto suponen alrededor de la tercera parte de alumnos en ambos problemas. Esto pone de nuevo de manifiesto el hecho de que el concepto de hipótesis no es trivial para el alumno y debería tenerse en cuenta en la enseñanza.

En el apartado donde se pide la elección de contraste es donde se observan las diferencias más grandes de respuestas correctas (P1 1,3%; P2 35,3%) y parcialmente correctas (P1 41,9%; P2 25,9%) e incorrectas (P1 31,4%; P2 8,9%), mientras que los porcentajes de no respuesta están equilibrados. Estos resultados demuestran que resultó más fácil elegir el modelo de contraste en el problema 2, lo que podría deberse a que se enseñaran en el curso menos modelos de análisis de varianza y por el contrario los alumnos tienen un abanico mucho más grande de posibilidades para las pruebas de hipótesis. Por ejemplo, en el análisis de la varianza no tienen que elegir entre prueba unilateral y bilateral que resultó una fuente de conflictos en el problema 1.

En el tercer punto, cálculo de las regiones y decisión sobre las hipótesis, por el contrario son menos las respuestas correctas y parcialmente correctas en el análisis de varianza. Nuestra explicación es que es más difícil para el estudiante el proceso de cálculo que se lleva a cabo para completar la tabla de análisis de varianza (se requieren para la toma de decisión) que en la prueba de hipótesis.

Cuando se interpretan los resultados obtenidos en términos de la fenomenología del problema (último punto comparado) los resultados son similares cuando se miran las respuestas correctas y parcialmente correctas en su conjunto. Es de destacar que los mayores porcentajes de no respuesta se corresponde con el apartado interpretación de resultados, resultado más complejo para el problema 2 (P1 29,9%; P2 58,5%), posiblemente porque no se han completado los pasos anteriores.

## 5.8 .CONCLUSIONES SOBRE LOS PROBLEMAS ABIERTOS

Para finalizar el capítulo exponemos nuestras conclusiones del estudio semiótico realizado de las respuestas de los estudiantes a los dos problemas abiertos. Parte de estos resultados se publicaron en Vera y Díaz (2008), Vera (2009), y Batanero, Vera y Díaz (2012). A continuación exponemos estas conclusiones comparando con los objetivos e hipótesis del estudio.

### Conclusiones respecto a los objetivos

O1: *El primer objetivo fue estimar la proporción de estudiantes en la muestra que comprenden y particularizan los objetos matemáticos incluidos en la enseñanza en un contexto de aplicación de los mismos.*

Pensamos que en este estudio se ha cumplido razonablemente este objetivo. Para ello, se han analizado las respuestas de una muestra amplia de estudiantes a dos problemas abiertos, uno sobre contraste de hipótesis y otro sobre análisis de la varianza.

Se descompone cada uno de estos problemas en diferentes apartados, clasificando y categorizando las respuestas en correctas, parcialmente e incorrectas, que a su vez se vuelven a clasificar, para mostrar la variedad de las mismas. De cada categoría así obtenida se realiza un análisis semiótico siguiendo la metodología propuesta en diversos trabajos del enfoque onto-semiótico y se van describiendo la trama de funciones semióticas establecidas por el estudiante en su respuestas.

Por tanto este análisis nos permite describir las estrategias y argumentos de los estudiantes, evaluando su comprensión de una forma más profunda que la llevada a cabo con los ítems de opción múltiple (Capítulo 4**).** Además las diferentes tablas de frecuencia presentadas y comentadas a lo largo del capítulo nos permiten estimar la proporción de estudiantes que completa de forma correcta, parcialmente correcta e incorrecta los apartados.

O2: *El segundo objetivo fue identificar y clasificar los conflictos semióticos* (Godino, Batanero y Font, 2007) encontrados respecto a los diferentes objetos que intervienen en el análisis de varianza, y analizar en nuestra muestra la presencia de otros conflictos descritos anteriormente (Alvarado, 2007; Cañadas, 2012) respecto al contraste de hipótesis.

Igualmente este objetivo se cumple de forma adecuada. Nuestro análisis semiótico permite identificar una serie de conflictos semióticos, que, en parte coinciden con otros ya descritos en investigaciones previas, pero también se proponen otros nuevos no descritos por otros autores. Del mismo modo, se estima la proporción de estudiantes que muestran los principales tipos de conflictos semióticos referidos. Este listado y clasificación de conflictos semiótico es una guía importante que el profesor puede tener en cuenta en la enseñanza y evaluación del tema.

**Conclusiones respecto a las hipótesis**

A continuación discutimos cada una de las hipótesis que se plantearon en este estudio.

Hipótesis 3.1. *Esperamos explicar las dificultades que tienen los estudiantes al plantear hipótesis estadísticas por conflictos consistentes en la confusión entre estadístico y parámetro, el contraste bilateral y unilateral o no reconocer que la hipótesis nula es puntual.*

Esta hipótesis se ha confirmado en las respuestas de nuestros estudiantes. En primer lugar hemos encontrado dificultades que aparecen tanto en la investigación de Vallecillos (1999), como en la de Cañadas (2012). Entre ellas destacamos el uso del estadístico muestral en lugar de usar parámetros al plantear las hipótesis del primer problema; la confusión entre contrastes bilaterales y unilaterales o plantea hipótesis que no cubren todo el espacio paramétrico, también en el primero.

Hipótesis 3.2. *Pensamos que algunos estudiantes no podrán tomar decisiones correctas con sus contrastes o en el análisis de varianza ya que, tendrán dificultades para determinar tanto el estadístico de prueba como su distribución o pueden confundir los criterios para aceptar y rechazar hipótesis estadísticas*

En los dos problemas encontramos errores en la determinación de las regiones críticas y de aceptación (coincidiendo con Vallecillos, 1994), es decir los estudiantes llegan a una decisión equivocada; en otros casos tienden a confundir el criterio de decisión al aplicar un contraste, con una prueba deductiva matemática, que también la encontramos denunciada en Vallecillos (1994). Esta dificultad se relaciona con la incomprensión de la lógica que subyace a las pruebas de hipótesis (Harradine, Batanero y Rossman, 2011). Los supuestos de aplicación de las diferentes distribuciones muestrales son confundidos, como ocurrió en el cálculo de intervalos de confianza en la investigación de Olivo (2008). Un error nuevo no tratado en la investigación previa es la confusión entre diferentes modelos de análisis de varianza que lleva a errores en el resto del proceso. El cálculo de la tabla de análisis de varianza resulta complejo y por este motivo los estudiantes llegan a una decisión final equivocada.

Hipótesis 3.3. *Esperábamos que algunos estudiantes que realizan correctamente todos los pasos del contraste de hipótesis o del análisis de varianza tuviesen dificultad en contextualizar los resultados, relacionándolos con la pregunta de investigación planteada.*

Nos basamos en los resultados que obtiene Arteaga (2011), cuyos estudiantes resuelven correctamente en gran medida los pasos de un proyecto estadístico, pero finalmente sólo presentan los resultados matemáticos obtenidos sin contextualizar. Muchos estudiantes de nuestra muestra, al igual que en la descrita por este autor, fallan en el último paso del proceso de modelización que según Dantal (1997) incluye: (1) Observación y (2) Descripción simplificada de la realidad; (3) Construcción de un modelo, (4) Trabajo matemático con el modelo, (5) Interpretación de resultados en la realidad. Los estudiantes fallan en todos ellos, pero incluso si completan los primeros correctamente, todavía tienen dificultades en el último. Ello se debió en parte a la confusión entre variable dependiente e independiente y también a la confusión de esta con la técnica aplicada en el experimento.

# CAPITULO 6.
# CONCLUSIONES

6.1. Introducción
6.2. Conclusiones respecto de los objetivos de la investigación
6.3. Conclusiones respecto a las hipótesis iniciales
6.4. Idoneidad del cuestionario de evaluación
6.5. Aportaciones y limitaciones del estudio
6.6. Problemas de investigación abiertos

## 6.1. INTRODUCCIÓN

En esta investigación se analizaron los significados personales de una muestra amplia de estudiantes de Licenciatura en Psicología, en relación con los objetos estadísticos incluidos en el análisis de varianza elemental (analizados en el Capítulo 1).

En este capítulo y para finalizar este trabajo presentaremos las conclusiones principales relacionadas con objetivos e hipótesis generales planteadas en el Capítulo 1. A continuación, nos ocupamos de la valoración del cuestionario desde el punto de vista de los criterios de idoneidad (siguiendo la propuesta de Batanero y Díaz, 2005). Destacaremos las aportaciones del trabajo, especialmente poniendo énfasis en los puntos en que se realiza alguna contribución, sin dejar de mencionar coincidencias ó disidencias con la investigación previa. Del mismo modo reconoceremos las limitaciones del estudio, que no se pueden evitar por las restricciones inherentes al tiempo y al espacio. Finalmente sugeriremos algunas líneas que podrán servir para, en el futuro completar y continuar esta investigación, dando así posibilidades al surgimiento de otras nuevas.

## 6.2. CONCLUSIONES RESPECTO DE LOS OBJETIVOS DE LA INVESTIGACIÓN

El objetivo general de nuestra investigación que se definió en al Capítulo 1 fue *realizar un estudio de evaluación de la comprensión que muestran los estudiantes de Psicología sobre algunos objetos básicos del análisis de varianza y su aplicación en la resolución de problemas relacionados*. Tal como se ha mencionado, en esta tesis continuamos nuestro trabajo iniciado para una tesis de fin de Máster (Vera, 2008).

Nuestro trabajo cumple ampliamente con este objetivo, puesto que la muestra es lo suficientemente grande y representativa de la población que se estudia, razón que nos permitirá inferir sobre los resultados y contribuir a la mejora de la enseñanza de este tema en Psicología y otras carreras relacionadas con las Ciencias Sociales, y por tanto a la mejor aplicación de la estadística en estas áreas, como investigadores.

Toda la investigación que hemos llevado a cabo la orientamos a cumplir este objetivo, pues se ha analizado en forma detallada la comprensión de las definiciones, propiedades, lenguaje, procedimientos, problemas y argumentos relacionada con el objeto de estudio, en la forma siguiente:

- Limitando los contenidos de la variable de investigación (Capitulo 1), describiendo el significado institucional de referencia, basándonos en libros de texto del área de Psicología para la enseñanza de la inferencia, y el análisis de las evaluaciones tomadas al mismo tipo de estudiantes previamente. También hemos tenido en cuenta los errores y dificultades que se han señalado en las investigaciones previas sobre inferencia y sus prerrequisitos.
- Construyendo un instrumento de evaluación válido y fiable (Capitulo 3) utilizando el análisis descrito en Capitulo 1, contando con ayuda de la investigación previa y nuestro marco teórico (Capítulo 2). Este instrumento nos proporciona datos cuantificables sobre la comprensión del estudiante, pudiendo así obtener desde la respuesta a cada ítem significados parciales del análisis de varianza y sus prerrequisitos.
- Realizando un estudio de evaluación (Capítulo 4) con dicho cuestionario, y analizando su validez y fiabilidad y las características psicométricas de los ítems. Con ello se ha mostrado evidencias de que nuestro cuestionario puede ser utilizado en otros estudiantes (con características sociales y educativas similares).
- Estudiando los conflictos semióticos cualitativamente (explicitando los que se registran en la respuesta a cada apartado de las tareas luego de identificar las funciones semióticas que emergen), y cuantitativamente (especificando cantidades de respuestas correctas, parcialmente correctas e incorrectas) en Capitulo 5.

Como argumentamos anteriormente la "comprensión del análisis de varianza" es un constructo psicológico y su caracterización ha de hacerse a través de las respuestas de los estudiantes a tareas cuyas respuestas proporcionen indicadores empíricos de dicha comprensión. Para ello desdoblamos el objetivo central en otros tres específicos.

*Objetivo 1: Realizar un análisis de los objetos matemáticos básicos que intervienen en el análisis de varianza que nos fundamente la definición semántica de la variable objeto de medición, delimitando las unidades de contenido semántico de dicha variable.*

Para llevarlo a cabo, presentamos en el Capítulo 1, Sección 1.7 y 1.8 un análisis de contenido matemático del objeto de estudio desde el punto de vista elemental (respetando necesidades del público al que va dirigido) para fijar parte del significado de referencia. Un análisis curricular nos da una semblanza de los contenidos de las asignaturas en relación con la estadística cursadas por los estudiantes de la muestra, para asegurarnos posteriormente la validez de contenido (Barbero, 2003; Morales, 2011).

*Objetivo 2: Construir un instrumento para evaluar la comprensión de los elementos básicos del análisis de la varianza, siguiendo unos criterios metodológicos rigurosos en la selección de los ítems que lo compongan.*

Una vez delimitado el contenido objeto de evaluación se realizó un proceso de construcción de un instrumento compuesto de dos partes: (a) un cuestionario formado por ítems de opción múltiples que permitan evaluar en un tiempo limitado la mayor cantidad posible de elementos de significado del análisis de varianza elemental; y (b) algunos problemas abiertos que permitan profundizar mejor en los razonamientos y la comprensión de los estudiantes.

El proceso de construcción del cuestionario lo hemos descrito en el Capítulo 3 y tiene también en cuenta el estudio de los antecedentes (Capítulo 2) que permite, por un lado, seleccionar algunos de los ítems del cuestionario y por otro, realizar un análisis a priori de las posibles dificultades esperadas en los estudiantes. Como parte del proceso hemos construido un banco de ítems de acuerdo con normas psicométricas, realizamos pruebas empíricas de dichos ítems en muestras de estudiantes de Psicología y evaluamos mediante un panel de expertos y nos permitió establecer un concenso sobre la tabla de especificaciones presentada en Capitulo 1. Con todo ello se finalizó con la selección de los ítems y elaboración del cuestionario que nos proponíamos en este objetivo.

*Objetivo 3: Llevar a cabo un estudio de evaluación con el citado cuestionario en una muestra de estudiantes de Psicología.*

Como se indicó, esta evaluación, tuvo dos partes diferenciadas que nos permitieron cumplir con este objetivo. En primer lugar, mediante el cuestionario previamente elaborado, con el que llevamos a cabo un estudio cuantitativo de la comprensión en una muestra de estudiantes sobre un amplio conjunto de ítems que cubren los diferentes contenidos incluidos (Capítulo 4). Aunque de una forma más superficial, este estudio (descrito en el Capítulo 4) permitió informar sobre algunos errores y dificultades de los estudiantes en relación al tema. Igualmente permite informar de las características psicométricas del cuestionario: dificultad, discriminación, validez y fiabilidad. En segundo lugar (Capítulo 5) realizamos un análisis semiótico de las respuestas a problemas abiertos realizados por los mismos estudiantes. Ello nos permitió evaluar de una forma más profunda el razonamiento y la comprensión de los estudiantes sobre elementos básicos del análisis de varianza. Utilizamos el enfoque ontosemiótico de la cognición matemática (EOS) propuesto por Godino y colaboradores para identificar los conflictos semióticos.

Hemos presentado los resultados de cada ítem, analizando las respuestas y comparándolas tanto con resultados de pruebas piloto como de la investigación previa cuando correspondía, dando lugar a la dimensión cuantitativa de esta investigación. También describimos los conflictos semióticos encontrados en las tareas desarrolladas por los estudiantes, que han dado lugar a la dimensión cualitativa de esta investigación.

A continuación informaremos sobre la comprensión de los significados personales en torno a nuestro objeto de estudio, de acuerdo con las unidades de contenido definidas en el cuestionario que, a grandes rasgos tiene dos secciones bien diferenciadas, la de los prerrequisitos y la del análisis de varianza.

**Prerrequisitos**

1. *Construcción de un intervalo de confianza* (Ítem 1). Encontramos un alto índice de respuestas correctas (61,2%), mientras en Olivo (2008) se informa sobre una amplia gama de posibilidades la construcción, según el campo de problemas.
2. *Interpretación del significado de un intervalo de confianza* (Ítem 2). Hemos tenido un bajo índice de respuesta correcta (27,2%) coincidiendo con la investigación previa (Behar, 2001, Olivo, 2008) que denuncia la gran dificultad tanto en estudiantes como en investigadores al momento de interpretar este objeto estadístico, siendo el proceso de interpretación inferencial esencial para el análisis de varianza.

3. *Contraste bilateral y unilateral e hipótesis nula y alternativa: asignación de hipótesis* (Ítem 3-4). Hemos tenido un alto índice de respuestas correctas en establecimiento de hipótesis (93,3% y 84,8%), mejor que en la investigación previa (Vallecillos, 1994; Cañadas, 2012). Sin embargo, en el *análisis semiótico* de las respuestas a problemas abiertos, encontramos solo un 33% (problema 1) y 25% (problema 2) de respuestas correctas en este punto. Los principales conflictos fueron: uso de estadístico en lugar de parámetro; contrastar dos poblaciones cuando se tenía una sola, uso del test unilateral cuando se pedía bilateral, hipótesis que no cubren el espacio paramétrico (50% en el problema 1) y usar una combinación lineal de los parámetros para contrastar (42,5% en el problema 2).
4. *Errores tipo I y II, nivel de significación y potencia* (Ítem 5-6). Encontramos un 50,9% y 64,7% de aciertos, mejorando considerablemente los resultados de Vallecillos (1994). Elaborar una respuesta adecuada a preguntas en relación con la potencia de un contraste de hipótesis es de gran complejidad, por lo que sólo un 14% de estudiantes responden correctamente. Los conflictos que generan la dificultad están relacionados con: la tipificación la variable aleatoria, el cálculo de probabilidades, determinación de percentiles, utiliza como probabilidad de error tipo I su complemento, e igualar probabilidades con estadísticos. Coincidimos con lo Batanero (2000), Castro Sotos *et al.* (2007) y Díaz, Batanero y Wilhelmi (2008) quienes afirman que la comprensión del concepto de potencia de una prueba de hipótesis es muy difícil.
5. *Cálculo de valor crítico, dado el nivel de significación* (Ítem 7). Se obtiene un 64,3% de aciertos. En el análisis semiótico del apartado 3 en problema 1, encontramos que algunos estudiantes no calculan correctamente el percentil, razón que los puede llevar a una decisión equivocada.
6. *Regla de decisión en un contraste* (Ítem 8). Encontramos 58% de respuestas correctas, aunque no son tantos los estudiantes que pueden relacionar el hecho de que el estadístico caiga en la región crítica y el valor *p* correspondiente sea menor que el nivel de significación. Vallecillos (1994) plantea un ítem similar, aunque algo más difícil que el nuestro, lo que quizá sea la causa por la cual la autora encuentra sólo 33,7% de respuestas correctas. En el problema 1 encontramos un 20,5% de respuestas correctas, mientras que en del problema 2 un 14,3%. Las dificultades encontradas están relacionadas con el cálculo incorrecto del estadístico, seleccionar un estadístico inadecuado, o cálculo incorrecto del percentil (problema 1). En el mismo sentido, la confusión en la determinación de los cálculos para encontrar algunas sumas de cuadrados y/o sus grados de libertad, en el cálculo de la media de cuadrados, del valor de la *F* lleva a los estudiantes a la toma de una decisión equivocada acerca de las hipótesis en el problema 2 (52,1%).

**Análisis de varianza**

Todos los resultados en este punto son originales, pues no se han desarrollado investigaciones previas. Hemos encontrado los siguientes:

- *Selección de un modelo de acuerdo con las características de un problema* (Ítem 9-10). Se encontró un 50,9% (ítem 9) de respuestas correctas al elegir un modelo de acuerdo con el problema propuesto en el ítem 9. No obstante, algunos lo confunde con el contraste *t* (14,7%) o no pueden reconocer la diferencia entre variables dependiente e independiente. También en el ítem 10 con 63,8% de respuestas correctas encontramos casi un 30% de respuestas correctas. Los conflictos

semióticos serían confundir la variable independiente con tipo de tratamiento aplicado (56,3%), o la variable dependiente con la independiente. Del mismo modo en el problema 2 algunos no puede determinar que deben usar un modelo de medidas repetidas, no reconoce se trata de un modelo de análisis de varianza, o que debe responder a efectos de un solo factor.

- *Supuestos de aplicación del análisis de varianza, comprobación de supuestos en un modelo y modelo lineal asociado a un tipo de análisis de varianza* (Ítem 11-12). Hemos obtenido un 43,3% de respuestas correctas (ítem 11) al evaluar la comprensión de supuestos de las observaciones para aplicar un modelo de análisis de varianza. Sin embargo, un 11,5% olvida la homocedasticidad, aunque este supuesto no es de gran importancia cuando se trata de diseños balanceados, pero si cuando alguno de los factores no es fijo. El ítem 12 evalúa la descomposición de la variabilidad total en un modelo de análisis de varianza de dos factores fijos. Encontramos un 69,2% de respuestas correctas, pero el 18,4% no comprende la necesidad del estudio del efecto de la interacción en estos modelos, según lo denunciado en la investigación por Green (2007) y Pardo *et al*. (2007).

- *Cálculo en el análisis de varianza, medidas repetidas.* En el ítem 13 hemos obtenido un 57,1% de respuestas correctas en el cálculo de *F* empírica. Un alto porcentaje elige estudiar la interacción en este modelo (21,9%) donde no es necesario. En el apartado 4 del problema 2, donde se debe ensamblar una gran cantidad de funciones semióticas para completar la tabla análisis de varianza y tomar una decisión, es muy bajo el porcentaje que lo hace correctamente (14,3%) y (14,7%) lo contesta parcialmente.

- *Cálculo en el análisis de varianza. Modelo de dos factores. Interpretación de tabla de análisis de varianza* (Ítem 14-15, 16 y 17). Para responder a estos ítems se les presenta un problema en contexto y una tabla de análisis de varianza. Luego se pide la complete usando algunos datos y responda al ítem (ítem 14 y 15), o la tabla completa desde una salida de software (ítem 16 y 17). Para los ítems 14 y 15, tuvimos un alto porcentaje de respuestas correctas (72,3%), su resolución comprendía sólo cálculos. También fue alto en el ítem 17 (74,6%) pues debía deducir un modelo. que se ha aplicado. Es decir si debe deducir cual es el modelo aplicado y el contenido de una tabla de análisis de varianza, resultan porcentajes correctos altos, mientras que si se le pide haga *interpretaciones* del resultado de la misma tabla a través de un p-valor bajan esos porcentajes a menos de la mitad (ítem 16, 33%).

- *Interpretación de salida de software. Resultados* (Ítem 18). El porcentaje de respuestas correctas es del 26,8%, bajo, suponemos por el proceso de resolución que se emplea (debe responder interpretando un p-valor). También hay un alto porcentaje de estudiantes que no lo resuelven, casi el 50% de la muestra. Hay un 16,5% que no puede interpretar la no existencia del efecto de una interacción y otro 13,8% creyó que el "tipo de trabajo" (factor) influía sobre la respuesta, mientras ocurre lo contrario. En resumen un 79,2% no puede interpretar la implicación que significa el resultado de un p-valor en una tabla de análisis de varianza. Coincidimos también con la investigación previa (Cañadas *et al*, 2012) que afirma que los resultados significativos son los peor interpretados.

- *Interpretación de salida de software. Supuestos del modelo* (Ítem 19-20). En ambos reactivos encontramos porcentajes de respuestas correctas en el orden de 50,9 y 35,3 respectivamente. Podemos suponer que el primero tiene un porcentaje más alto

puesto que se proponen, mediante fórmulas posibles hipótesis nulas y debe el estudiante indicar cuál es la que se somete a contraste en la prueba de Levene. Este tipo de problemas ya los abordó en los ítems 3 y 4 yen los primeros apartados de ambos problemas abiertos. Se reduce mucho el porcentaje de aciertos para el ítem 20, puesto que debe interpretar un resultado de una salida de software.

En vista de estos resultados se recomienda prestar especial atención para la enseñanza de estos modelos, a los conflictos que se proponen, en particular cuando se trata de discutir el significado personal de los siguientes objetos:

- Proceso de interpretación de los resultados de una tabla de análisis de varianza, sea o no obtenida de salida de software la del problema.
- Relación entre valor *p,* región crítica y nivel de significación.
- La descomposición de la variabilidad total en análisis de varianza.
- Diferencias entre variables aleatorias dependientes e independientes y la relación entre ellas.
- Los supuestos tanto de las observaciones, a priori; como los de diagnóstico.

## 6.3. CONCLUSIONES RESPECTO A LAS HIPÓTESIS INICIALES

En este apartado analizamos las hipótesis formuladas inicialmente, que se han estudiado únicamente en forma exploratoria y fueron las siguientes:

> Hipótesis 1. *Esperábamos encontrar dificultades en algunos pre-requisitos del tema, en particular, en la interpretación de intervalos de confianza, el nivel de significación, la asignación de hipótesis en un contraste y el cálculo de la potencia.*

Esta hipótesis se fundamentaba en los estudios de los antecedentes, en los cuáles se han descrito estas dificultades en otros estudiantes (Vallecillos, 1994; Olivo, 2008; Díaz, Batanero y Wilhelmi, 2008 y Harradine, Batanero y Rossman, 2011) en relación a contraste de hipótesis, intervalos de confianza y con varios objetos en relación con la inferencia estadística, respectivamente.

En el Capítulo 4 (Estudio 2) observamos que esta hipótesis se confirma. Con respecto al prerrequisito interpretación de un intervalo de confianza, en nuestra muestra sólo lo interpreta correctamente el 27,2% (Ítem 2). Si tenemos en cuenta los resultados en los ítems 3 y 4, los estudiantes han alcanzado un alto índice de acierto, desde donde se desprendería que no se cumple la segunda parte de la hipótesis 1, pero cuando se deben establecer hipótesis estadísticas dado un problema contextualizado, los porcentajes de respuestas correctas descienden notoriamente (estudio semiótico, Capitulo 5). Especialmente del análisis de respuestas en el apartado 5 del problema 1 se advierte la enorme dificultad que les ha acarreado el cálculo de la potencia de un contraste. En consecuencia, esta hipótesis se cumple casi en su totalidad.

Hipótesis 2. *Se preveía que también existan dificultades en relación con el análisis de varianza, por ejemplo, en la selección del modelo, la interpretación de la tabla o salida de software y la comprensión de los supuestos.*

Aunque no tenemos antecedentes en este punto, pero puesto que el análisis de varianza es un tipo particular de contraste de hipótesis, esperábamos se reprodujeran algunos problemas descritos respecto al contraste de hipótesis, por ejemplo en las investigaciones de Vallecillos (1994) y Cañadas (2012). Efectivamente en el Estudio 2 (Capítulo 4) en el ítem 9 casi un 50% de estudiantes que eligen alguno de los distractores o no lo contesta. Análogo resultado presenta el ítem 11 (supuestos que deben cumplir las observaciones para aplicar el modelo).

En el ítem 13 (cálculos para interpretar una tabla de análisis de varianza) 43% comenten fallos; en el ítem 16 (interpretación de valores obtenidos al completar una tabla de análisis de varianza) un 67%, en el ítem 18 (interpretación de salida de software tabla análisis de varianza) un 79,2% y en el ítem 20 (análisis de supuestos del modelo con salida de software) 64,7% de respuestas incorrectas podemos evaluar el cumplimiento de esta hipótesis. Pero, en el resto de los ítems: 10 (aplicación de un modelo de análisis de varianza de dos factores y efectos fijos), 14 y 15 (cálculo de valores que forman la tabla de análisis de varianza de dos factores), y 17 (tipo de modelo aplicado según tabla de análisis de varianza obtenido desde una salida de software) dan mejores resultados, por lo que esta hipótesis se ha cumplido parcialmente.

Hipótesis 3. *El análisis semiótico de problemas abiertos permitirá explicar parte de las dificultades anteriores en términos de conflictos semióticos.*

Esta hipótesis se apoya en la experiencia previa obtenida por otros miembros del equipo de investigación, como Alvarado (2007), Olivo (2008), y Cañadas *et al.* (2012) que usaron el mismo método para identificar conflictos semióticos relacionados con el contraste de hipótesis. Se confirma dada la gran cantidad de conflictos semióticos que hemos encontrado en nuestro Estudio 3 (Capitulo 5). Los hemos identificado tanto para el problema de contraste de hipótesis, confirmando los denunciados por los autores mencionados, como para el problema de análisis de varianza donde los conflictos emergen desde el establecimiento de las hipótesis estadísticas, la identificación del tipo de variables, elección de un contraste, cálculo de las regiones, decisión sobre la hipótesis e interpretación de resultados.

## 6.4. IDONEIDAD DEL CUESTIONARIO DE EVALUACIÓN

En este apartado utilizamos algunas ideas sobre la idoneidad didáctica de Godino y sus colaboradores, adaptadas en particular al caso de los cuestionarios de evaluación. En el nuestro, no queremos limitar el interés del estudio al significado declarado por los estudiantes, puesto que suponemos que las respuestas dadas se presumen una muestra representativa de las prácticas de los estudiantes. Por ello, es nuestro interés indagar acerca del significado personal de los estudiantes respecto de las tareas. Aspiramos también a obtener un conocimiento generalizable en relación a las capacidades y dificultades de otros estudiantes con características similares a los de nuestra muestra. En este sentido, se aplicó el concepto *idoneidad* y sus tipos al caso de la evaluación (descrito en el Apartado 2.2.6).

La *dificultad* de un ítem o tarea daría una medida de su *idoneidad cognitiva*; es decir del grado de representatividad de los significados evaluados respecto a los significados personales. Hemos obtenido ítems con un grado de dificultad que varía desde el más fácil con ID=0,93 (ítem 3) hasta los más difíciles con ID=0,27 (ítem 2 y 18); el 50% de los ítems varían entre los valores 0,49 y 0,70 y podemos aceptar ampliamente una distribución normal para los índices de dificultad (p-valor 0,731). Por lo tanto, tenemos ítems que evalúan gran cantidad de conocimientos de los estudiantes y que van desde los más fáciles a los más difíciles, con la gran mayoría de dificultad intermedia.

La *discriminación* de un ítem valoraría su *idoneidad evaluador;* un ítem puede ser adecuado cognitivamente, pero no diferenciar (por ser demasiado fácil) los alumnos que tienen un mayor o menor conocimiento del concepto. Esta idoneidad podría ser un componente de la idoneidad *instruccional*, en cuanto uno de los objetivos de la instrucción es la función evaluadora. Evaluamos la discriminación calculando las correlaciones de cada ítem con la puntuación total, encontrando solo uno (ítem 8) con correlación negativa y dos (ítem 9 y 18) con correlaciones cercanas a 0. Estos resultados arrojan una muy adecuada capacidad de discriminación de los ítems para el cuestionario en todo su conjunto.

La *validez de contenido* de un cuestionario indicaría una *idoneidad epistémica*, o grado de representatividad del instrumento en cuanto al significado objeto de evaluación. Mediante el análisis de contenido detallado que nos llevó a la definición semántica de la variable objeto de medición, el juicio de expertos, las pruebas piloto y el análisis de la investigación previa nos hemos asegurado la validez de contenido para nuestro instrumento.

La *fiabilidad* o *generalizabilidad a otros ítems* daría una medida de la estabilidad de la respuesta, es decir sería otro componente de la *idoneidad evaluadora.* El análisis factorial que realizamos con la tabla dicotómica de respuesta a los ítems, dio como resultado la presencia de al menos seis componentes, lo que pone de manifiesto la multidimensionalidad del cuestionario, a pesar de ello el coeficiente de generalizabilidad a ítems fue de 0,79, o sea el mismo que el valor para Alfa de Cronbach (fiabilidad de consistencia interna), ello nos permitió calificarlo como fiable.

La *generalizabilidad a otros estudiantes,* sugeriría una *idoneidad generalizadora o externa* en cuanto a que los resultados se generalizarían a otros estudiantes. En nuestro estudio si hemos encontrado una alta generalizabilidad a estudiantes con la mismas características (0,961) dando valores muy cercanos a la unidad, pero en este sentido reconocemos límites del instrumento ya que fue pasado a estudiantes de Psicología solamente.

## 6.5.APORTACIONES Y LIMITACIONES DEL ESTUDIO

Una vez discutidos los objetivos e hipótesis pasamos a resaltar nuestras aportaciones, la principal de las cuáles es la c*onstrucción de un cuestionario válido*, con alta consistencia interna, que proporciona información acerca del significado personal de los estudiantes a los que se pasó. Es un instrumento que puede ayudar a abrir investigación futura en la enseñanza y el aprendizaje del análisis de varianza a nivel elemental para otros estudiantes, así como en la evaluación del tema dentro de la enseñanza. En particular el cuestionario contiene *ítems originales* (9-20)*,* al no haber encontrado en la investigación previa trabajos de evaluación de la comprensión del análisis de varianza. También los ítems 5, 6 y 7 son originales, aunque se refieren a los requisitos.

El *análisis semiótico* profundo de las respuestas de los estudiantes a dos problemas abiertos es otra aportación. Por ejemplo, en el problema 1, aportamos los resultados del análisis de la potencia del contraste, que resulto con alta complejidad, por la cantidad de funciones semióticas en juego. Mostramos así una gran cantidad de conflictos que llevan al estudiante a un cálculo erróneo (apartado 5). En el problema 2, analizamos las dificultades en la comprensión de los cálculos para completar la una tabla de análisis de varianza en un modelo de medidas repetidas, toma de decisión e interpretación de resultados en términos del problema (apartados 4 y 5) que son originales en este estudio. Registrando una gran cantidad de conflictos semióticos que se describen en 5.6.1. Estos resultados permitirán mejorar la enseñanza actual sobre el tema.

**Limitaciones del estudio**

Como en otras investigaciones una limitación es la muestra empleada intencional; pero al mismo tiempo nos da posibilidad de ampliar el estudio para generalizarlos a estudiantes de otras carreras universitarias, o en otros contextos socio-educativos.

Respecto a la *delimitaciones del significado de referencia*, elaborado en el Capítulo 1, reconocemos que el objeto análisis de varianza tiene un significado mucho más completo. Por ejemplo no hemos tratado modelos con más de dos factores ó con distribución libre. Por ello, hacemos notar que presentamos en este trabajo un significado de referencia parcial.

Debido al trabajo implicado en el análisis cualitativo, sólo se realizó el análisis semiótico de un problema de análisis de varianza de medidas repetidas. Sería interesante proponer otros problemas; por ejemplo, de análisis de varianza de factores fijos o aleatorios con uno o dos factores y analizar si se mantienen los conflictos semióticos identificados en nuestro estudio.

## 6.6. PROBLEMAS DE INVESTIGACIÓN ABIERTOS

Las limitaciones señaladas en el trabajo abren nuevas vías de investigación, En primer lugar sería posible ampliar el significado del concepto análisis de varianza, teniendo en cuenta tanto el currículo diferente, como los posibles enfoques distintos dados al mismo. Con esa información sería posible construir un nuevo cuestionario basándose en el nuestro y estudiar comparativamente resultados a distintas carreras universitarias.

Por otro lado, hemos limitado exclusivamente el significado que adquieren estudiantes de Psicología. De este modo una extensión de esta investigación podría estar dirigida a abordar desde otras áreas de enseñanza como ciencias exactas, ciencias naturales, o ciencias de la salud, analizando el significado del análisis de varianza en esos estudiantes.

Igualmente es posible indagar en relación al significado institucional en la enseñanza y la formación de profesores. Muchas investigaciones indican carencias tanto del conocimiento estadístico como profesional requerido para la enseñanza del análisis de varianza, especialmente en carreras en relación con ciencias sociales: Sociología, Psicología, Antropología etc., que son esos profesionales quienes imparten las clases y no alcanzar a tener los conocimientos profundos de estadística. También inversamente, cuando son matemáticos, ingenieros y estadísticos los que las imparten sin tener conocimiento alguno de la especialidad.

Estos resultados podrían sentar las bases para el diseño de propuestas de enseñanza sobre el análisis de varianza a estudiantes de Psicología. La propuesta también podría contener apoyo de software con enfoque didáctico y el desarrollo de guías didácticas para apoyar el proceso de estudio.

Se podría completar el análisis de los datos obtenidos con nuestros estudios (Capítulos 4 y 5) para completar con otras técnicas estadísticas que den más información sobre las tendencias, variabilidad de errores y dificultados específicos en la muestra total. Se podría complementar estos datos con entrevistas clínicas, y observar el desempeño de los estudiantes en clase, que permiten analizar casos específicos (estudio de casos) para determinar tipología y niveles de la comprensión del análisis de varianza.

*Análisis semiótico de respuestas que se hacen sobre la salida de software estadístico*. Sería posible también seguir indagando desde otra perspectiva el significado personal de estudiantes (con características similares) realizando un análisis semiótico acerca de las argumentaciones que utilizan estos en la elección de respuestas desde la salida de software estadísticos que incluyan el análisis de varianza.

Finalizamos el trabajo indicando que la importancia de los errores descritos justifica la necesidad de continuar con las investigaciones sobre el tema desde el marco de la Didáctica de la Matemática, puesto que solo un mejor y profundo conocimiento de las dificultades y errores en el aprendizaje puede producir el que se tengan en cuenta en la enseñanza y ello redunde, de manera indirecta en un uso más correcto del mismo en las aplicaciones:

> .resulta evidente la importancia de que los profesores conozcan los errores y dificultades en los que incurren sus estudiantes, ya que solo así podrán tenerlos en cuenta en la planificación y puesta en práctica de la enseñanza en las mejores condiciones (Vallecillos, 1996, p.74).

# REFERENCIAS


Alvarado, H. (2007). *Significados del teorema central del límite en la enseñanza de la estadística en ingeniería.* Tesis Doctoral. Universidad de Granada.

American Psychological Association, American Educational Research Association y National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Andrés, A. M. y Luna, J. (2004). *Bioestadística para ciencias de la salud*. Madrid: Capitel.

ANECA (2005). *Libro blanco del título de grado en psicología*. Madrid: Autor.

Arteaga. P. (2011). *Evaluación de conocimientos sobre gráficos estadísticos y conocimientos didácticos de futuros profesores*. Tesis Doctoral. Universidad de Granada.

Aron, A., Coups, E. y Aron, E. (2012). *Statistics for psychology*. New York: Pearson.

Backhoff, E., Larrazolo, N. y Rosas, M. (2000). Nivel de dificultad y poder de discriminación del Examen de Habilidades y Conocimientos Básicos (EXHCOBA). *Revista Electrónica de Investigación Educativa*, *2* (1). Online; redie.uabc.mx/vol2no1/.

Balluerka, N. y Vergara, A. I.. (2002). *Diseño de investigación experimental en psicologia.* Madrid: Pearson Educación.

Barbero, M. (2003). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, *2*(1-2), 75-98.

Batanero, C. (2001) (Ed.). *Training researchers in the use of statistics*. Granada: International Statistical Institute.

Batanero, C. y Díaz, C. (2005). Análisis del proceso de construcción de un cuestionario sobre la probabilidad condicional. Reflexiones desde el marco de la TSFS. En A. Contreras, L. Ordóñez y C. Batanero (Eds.). *Investigación en Didáctica de las Matemática. I Congreso Internacional sobre Aplicaciones y Desarrollos de la Teoría de las Funciones Semióticas.* (pp. 15-39). Jaén. Universidad de Jaén.

Batanero, C. y Díaz, C. (2008). *Análisis de datos con Statgraphics*. Granada: Departamento de Didáctica de la Matemática.

Batanero, C., Tauber, L. M. y Sánchez, V. (2001). Significado y comprensión de la distribución normal en un curso introductorio de análisis de datos. *Quadrante*, *10* (1), 59–92.

Batanero, C., Tauber, L. M. y Sánchez, V. (2004). Students' reasoning about the normal distribution. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 257–276). Dordrecht, The Netherlands: Kluwer.

Batanero, C., Vera, O. y Díaz, C. (2012). Dificultades de estudiantes de psicología en la comprensión del contraste de hipótesis. *Números, 80,* 91-10..

Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística.* Tesis doctoral. Universidad Politécnica de Cataluña.

Belia, S., Fidler, F. y Cumming, G. (2005). Researchers misunderstand confidence intervals and standar error bars. *Psychological Methods, 4,* 389-396.

Ben-Zvi, D. y Garfield, J. B. (Eds.) (2004). Research on reasoning about variability. *Statistics Education Research Journal*, *3*(2). Online: iase-web.org/Publications.php?p=SERJ.

Birnbaum, I. (1982). Interpreting statistical significance. *Teaching Statistics, 4*, 24–27.

Bisquerra, R. (1989). *Métodos de investigación educativa.* Barcelona: Ceac.

Bono, C., Roser, J. y Vallejo, G. (2008). Técnicas de análisis aplicadas a datos longitudinales en psicología y ciencias de la salud: período 1985-2005. *Papeles del Psicólogo*, *29,* 136-146.

Box G. E. P., Hunter, W. G. y Hunter J .S. (2008). *Estadística para experimentadores.* Barcelona: Reverté.

Brousseau, G. (1997). *Theory of didactical situations in mathematics*. Dordrecht, The Netherlands: Kluwer.

Carpintero, H. (1990). Sobre el aprendizaje de la estadística. *Estadística española*, *31* (122), 476-479.

Castle, R. (2002). Using interactive visualisation to develop statistical understanding. En B, Phillips /Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Online: http://iase-web.org/Conference_Proceedings.php.

Castro Sotos, A. E., Vanhoof, S., Van den Nororgate, W., Onghena, P. (2007). Student´s misconceptions of statistical inference: A review of the empirical evidence form research on statistical education. *Educational Research Review, 2* (2), 98-113.

Cañadas, G. (2012). *Comprensión intuitiva y aprendizaje formal de las tablas de contingencia en alumnos de psicología*. Tesis Doctoral. Universidad de Granada.

Cañadas. G., Batanero, C., Díaz, C.y Roa, R. (2012). Psychology students' understanding of the chi-squared test. *Statistique et Enseignement, 3* (1), 3-18.

Chance, B. L., delMas, R. C., y Garfield, J. B. (2004). Reasoning about sampling distributions. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-324). Dordrecht, The Netherlands: Kluwer.

Chiocca, C.M. (2006). An Instrumental obstacle example. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. International Association for Statistical Education. Online: Online: http://iase-web.org/Conference_Proceedings.php.

Chow, L. S. (1996). *Statistical significance: Rationale, validity and utility.* London: Sage.

Cohen, J. (1994). The earth is round (p<.05). *American Psychologist, 49* (12), 997 – 1003.

Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional.* Tesis Doctoral. Universidad de Granada.

Cumming, G. y Fidler, F. (2005). Interval estimates for statistical communication: problems and possible solutions. Trabajo presentado en la *IASE Satellite Conference on Communication of Statistics.* Sydney: International Association for Statistical Education.

Cumming, G., Williams, J. y Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics, 3,* 299-311.

Dantal, B. (1997). Les enjeux de la modélisation en probabilité. En *Enseigner les probabilités au lycée* (pp. 57-59). Reims: Commission Inter-IREM.

De la Fuente, E. I. y Díaz, C. (2003). Reflexiones sobre los métodos inferenciales en psicología. *Libro de resúmenes del VIII Congreso de Metodología de las Ciencias Sociales y de la Salud* (pp. 326 – 327). Valencia: Asociación Española de Metodología de las Ciencias del Comportamiento.

delMas, R. (2001). *What makes the standard deviation larger or smaller?* Statistics Teaching and Resource Library (STAR). Online: www.causeweb.org/.

delMas, R. C., Garfield, J. B. y Chance, B. L. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, *7* (3). On line: www.amstat.org/publications/jse.

Díaz, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en psicología*. Tesis Doctoral. Universidad de Granada.

Díaz, C. y Batanero, C. (2006). ¿Cómo puede el método bayesiano contribuir a la investigación en psicología y educación? *Paradigma*, *27* (2), 35-53.

Díaz, C., Batanero, C. y Wilhelmi, M. R. (2008). Errores frecuentes en el análisis de datos en educación y Psicología. *Publicaciones.* 35, 109-133.

Díaz, C. y de la Fuente, E. I. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento,* Volumen especial 2004, 161-167.

Dunn, O. J. y Clark, V.A. (1987). *Applied statistics: Analysis of variance and Regression*. New York: John Wiley.

Dunn, D. S., Smith, R. A., & Beins, B. C. (Eds.). (2012). *Best practices in teaching statistics and research methods in the behavioral sciences*. New York: Routledge.

Eco, U. (1979). *Tratado de semiótica general.* Barcelona: Lumen.

Estepa, A. (2013). Los fenómenos de cambio. *Probabilidad Condicionada 2,* 47-53.

Estepa, A. y del Pino J. (2013). Elementos de interés en la investigación didáctica y enseñanza de la dispersión estadística. *Números,* 83, 43-63.

Falk, R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning, 9*, 83–96.

Ferré, J. y Rius, F. X. (2008). *Introducción al diseño estadístico de experimentos.* Tarragona: Universitat Rovira i Virgili., On line: www.quimica.urv.es/quimio/general/dis.pdf.

Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th International Statistics Institute Session* [CD-ROM]. Sidney: International Statistical Institute.

Fox, D. J. (1981). *El proceso de investigación en la educación*. Pamplona: Eunsa.

García, R. M (2004). *Inferencia estadística y diseño de experimentos*. Buenos Aires: EUDEBA.

Garfield, J. y Ben-Zvi, D. (Eds.) (2005). Reasoning about variation. *Statistics Education Research Journal*, *4* (1). Online: iase-web.org/Publications.php?p=SERJ.

Garfield, J. y Ben-Zvi, D. (2008). *Developing students' statistical reasoning. Connecting research and teaching practices*. New York: Springer.

Garfield, J. B., delMas, R. C., y Chance, B. L. (2004). *Tools for teaching and assessing statistical inference.* On line: www.gen.umn.edu/research/stat_tools.

Gea, M. M. (2014). *La correlación y regresión en Bachillerato: análisis de libros de texto y del conocimiento de los futuros profesores*. Tesis doctoral. Universidad de Granada.

Ghiglione, R. y Matalón, B. (1991). *Les enquêtes sociologiques. Théorie et practique* París: Armand Colin.

Gigerenzer, G. y Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats (pp. 129-161). *Psychological Review, 102*, 684-704.

Godino, J. D. (1996). Mathematical concepts, their meanings and understanding. En L. Puig y A. Gutiérrez (Eds.)*, Proceedings of the 20th PME Conference* (v.2, 417- 424). Universidad de Valencia.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathematiques, 22* (2 y 3), 237-284.

Godino, J. D. (2003). Teoría de las funciones semióticas. Un enfoque ontológico-semiótico de la cognición e instrucción matemática. *Departamento de Didáctica de la Matemática. Universidad de Granada*. Online: www.ugr.es/local/jgodino/.

Godino, J. D. (2011). Indicadores de la idoneidad didáctica de procesos de enseñanza y aprendizaje de las matemáticas. Conferencia presentada en el *XIII CIAEM-IACME,* Recife, Brasil.

Godino, J. D. (2012). Origen y aportaciones de la perspectiva ontosemiótica de investigación en Didáctica de la Matemática. En A. Estepa, A. Contreras, J. Deulofeu, M. C. Penalva, F. J. García y L. Ordóñez (Eds.), *Investigación en Educación Matemátic*a XVI, 49 - 68.

Godino, J. D. (2014). *Síntesis del enfoque ontosemiótico del conocimiento y la instrucción matemática: motivación, supuestos y herramientas teóricas.* Universidad de Granada. Online: www.ugr.es/local/jgodino/.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathematiques*, *14* (3), 325-355.

Godino, J. D. y Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. En A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics education a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Godino, J. D. Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education, 39* (1-2), 127-135.

Godino, J. D., Batanero, C., Rivas, H. y Arteaga, P. (2013). Componentes e indicadores de idoneidad de programas de formación de profesores en didáctica de las matemáticas. *Revemat: RevistaEeletrônica deEeducação Matemática, 8*(1), 46-74.

Godino, J. D., Batanero, C**.** y Roa, R. (2005). An onto-semiotic analysis of combinatorial problems and the solving processes by university students. *Educational Studies in Mathematics, 60*( 1), 3-36.

Godino, J. D., Font, V. y Wilhelmi, M. R. (2008). Análisis didáctico de procesos de estudio matemático basado en el enfoque ontosemiótico. *Publicaciones, 38,* 25-48.

Green, K. E. (2007). Assessing understanding of the concept of iteraction in analysis of variance. Trabajo presentado en la *IASE/ISI-Satellite Conference on Assessing Student Learning in Statistics.* Guimarães, Portugal: International Statistical Institute.

Haller, H. y Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research, 7* (1), 1–20.

Harlow, L. L., Mulaik, S. A. y Steiger, J. H. (1997*). What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum.

Hernández, R., Fernández, C. y Baptista, C. (1998). *Metodología de investigación.* México: McGraw-Hill.

Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 235-246). New York: Springer.

Henry, M. (1997). Notion de modéle et modélization en l'enseignement. En *Enseigner les probabilités au lycée* (pp. 77-84). Reims: Commission Inter-IREM.

Hjelmslev, L. (1971). *Prolegómenos a una teoría del lenguaje*. Madrid: Gredos.

Kahneman, D., Slovic, P. y Tversky, A. (1982) *Judgment under uncertainty:Heuristics and biases*. New York: Cambridge University Press.

Kirk, R. E. (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement*, *61* (2), 213–218.

Kline, P. (2013). *Handbook of psychological testing*. Routledge.

Koan, N. C. (2001). Importancia del avance en la investigación psicométrica. Trabajo presentado en el *XXVIII Congreso Interamericano de Psicología*. Santiago de Chile: USAL. Online: www.salvador.edu.ar/psic/ua1-9pub02-7-01.htm.

Krippendorff, K. (2013). *Content analysis: an introduction to its methodology*. Los Angeles / London, Sage.

Kurji, P. (2002). Statistics made alive. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Cape Town: International Association for Statistical Education. Online: Online: http://iase-web.org/Conference_Proceedings.php.

Lann, A. y Falk, R. (2003).What are the clues for intuitive assessment of variability? En C. Lee (Ed.), *Proceedings of the Third International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-3).* [CDROM] Mount Pleasant, MI: Central Michigan University.

Lecoutre, B. (1996). *Traitement statistique des données experimentales.* París: CISIA.

Lecoutre, B. (1999). Beyond the significance test controversy: Prime time for Bayes? *Bulletin of the International Statistical Institute: Proceedings of the Fifty-second Session of the International Statistical Institute* (Tome 58, Book 2) (pp. 205 – 208). Helsinki: International Statistical Institute.

Lecoutre, B., Lecoutre M. P., y Poitevineau J. (2001). Uses, abuses and misuses of significance tests in the scientific community: Won't the Bayesian choice be unavoidable? *International Statistical Review*, *69,* 399-418.

León, O. G. y Montero, I. (2002). *Métodos de investigación en psicología y educación*. Madrid: McGraw-Hill.

Lipson, K. (1997). What do students gain from computer simulation exercices? En J. B. Garfield. y G. Burril (Eds.), *Proceedings of the IASE 1996 Round Table Conference*. (pp. 131-144). Vooburg, Holanda: International Statistical Institute e International Association for Statistical Education.

Lipson, K. (2002). The role of computer based technology in developing understanding of the concept of sampling distribution. En B, Phillips /Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Online: http://iase-web.org/Conference_Proceedings.php.

Livacic-Rojas, P., Vallejo, G., Fernández, P. (2006). Procedimientos estadisticos alternativos para evaluar la robustez mediante diseños de medidas repetidas. *Revista Latinoamericana de Psicología 38,* 579-598.

López Sánchez, J., Pérez de Vargas, A., Zamora Romero, J., Murciano Cespedosa, A., Alonso Fernández, J., Reviriego Eiros, M. y Lahoz Beltrá, R. (2004). *Aula virtual de Bioestadística.* Online: http://e-stadistica.bio.ucm.es/.

Losada, J. L. y López Feal, R. (2003). *Métodos de investigación en ciencias humanas y sociales*. Madrid: Thomson.

Manzano, V. (1997). *Inferencia estadística. Aplicaciones con SPSS/PC+*. Madrid: RA-MA.

Martínez Arias, R. (1995). *Psicometría*. Madrid: Síntesis.

Meliá, J. L. (2001). *Teoría de la fiabilidad y la validez*. Valencia: Cristóbal Serrano.

Méndez, H. (1991). *Understanding the central limit theorem*. Tesis Doctoral. Universidad de California.

Millman, J. y Greene, J. (1989). The specification and development of test of achievement and ability. En R. L. Linn (Ed.), *Educational Measurement* (pp. 335 –366). London*:* Macmillan.

Molinero, L (2003). *Análisis de varianza.* Asociación española contra la hipertensión arterial. Online: http://www.seh-lelha.org/anova.htm.

Montgomery, D. (2005). *Diseño y análisis de experimentos*. Mexico: Limusa.

Moore, D. S. (1995). *The basic practice of statistics*. New York: Freeman.

Morales, P. (2011). *Guía para construir cuestionarios y escalas de actitudes*. Madrid: Universidad Pontificia Comillas.

Moses, L. E. (1992). The reasoning of statistical inference. In D. C. Hoaglin y D. S. Moore (Eds.), *Perspectives on contemporary statistics* (pp. 107-122). Washington, DC: Mathematical Association of America.

Oliver, S. P. (2011). Metodología de la investigación social. Librería-Editorial Dykinson.

Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México.* Tesis Doctoral. Universidad de Granada.

Osterlind, S. J. (1989). *Constructing test items*. Boston: Kluwer.

Pardo, A., Garrido, J., Ruiz, M.A. y San Martín, R. (2007). La interacción entre factores en el análisis de la varianza: error de interacción. *Psicothema 19* (2), 343-349.

Pagano, R. R. (2011). *Estadística para las ciencias del comportamiento* (9ª. ed.). (O. A. Velasco, Trad.) Madrid: Paraninfo Thomson Learning.

Peers, I. (2006). *Statistical analysis for education and psychology researchers.* London: Falmer Press.

Reading, C. y Reid, J. (2005). Consideration of variation: A model for curriculum development. En G. Burrill y M. Camden (Eds.), *Curricular Development in Statistics Education: International Association for Statistical Education 2004 Roundtable* (pp. 36-53). Voorburg, The Netherlands: International Statistical Institute.

Reading, C., & Shaughnessy, J. M. (2004). Reasoning about variation. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201–226). Dordrecht, The Netherlands: Kluwer.

Rivadulla, A. (1991). *Probabilidad e inferencial científica.* Barcelona: Anthropos.

Rosnow, R. L. y Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110*, 574-576.

Rubin, A. y Rosebery, A. S. (1990). Teachers' misunderstandings in statistical reasoning; evidence from a field test of innovative materials. En A. Hawkins (Ed.) *Training teachers to teach Statistics* (pp. 72-89). Voorburg, The Netherlands: International Statistical Institute.

Sánchez, J. C. (2012). *Los métodos de investigación*. Ediciones Díaz de Santos.

Sánchez, E., da Silva, C. B., y Coutinho, C. (2011). Teachers' understanding of variation. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 211-221). New York: Springer.

Samapieri, R. H, Collado, C. F., y Lucio, P. B. (2006). *Metodología de la Investigación*. McGraw-Hill.

Santisteban, C. (1.990). *Psicometría: Teoría y técnica de construcción de tests*. Madrid. Norma.

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation.* Belmont, CA: Wadsworth.

Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.). *Proceeding of the Third International Conference on Teaching Statistics* (pp. 486-490).Voorburg, The Netherlands: International Statistical Institute. Online: http://iase-web.org/Conference_Proceedings.php.

Tauber, L. (2001). *Significado y comprensión de la distribución normal a partir de actividades de análisis de datos*. Tesis Doctoral. Universidad de Sevilla.

Tejedor, F. J. (1999). *Análisis de varianza*. Schaum. Madrid: La Muralla S.A.

Thorndike, R. L. (1989). *Psicometría aplicada.* Mexico: Limusa.

Torok, R. y Watson, J. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal, 12* (2), 147-169.

Trigo, M. E., López, J., Martínez, R. y Moreno, R. (2005). Propiedades psicométricas de un conjunto de pruebas en función del contenido, tipo y número de opciones de respuesta. *Iberpsicología, 10*, 8-14.

Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis Doctoral. Universidad de Granada.

Vallecillos, A. (1996). *Inferencia estadística y enseñanza: Un análisis didáctico del contraste de hipótesis.* Granada: Comares.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. In *Proceedings of the 52 session of the International Statistical Institute* (pp. 201–204). Helsinki: International Statistical Institute. Tome 58, Book 2.

Vallecillos, A. (2002). Empirical evidence about understanding of the level of significance concept in hypotheses testing by university students. *Themes in Education*, *3*(2), 183–198.

Vallecillos, A. y Batanero, C. (1997). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitarios. *Recherches en Didactique des Mathematiques*, *17*, 29–48.

Vallejo, G., Lozano, L. (2006). Modelos de análisis para los diseños multivariados de medidas repetidas. *Psicothema 18* (2), 293-299.

Vaquero, E., Cardoso, Mª J., Vázquez,M. (2001). Psicología on line. Formación, autoayuda y consejo on line. *Congreso Internacional Online de Psicología Aplicada:* On line: http://psicologia-online.com/ciopa2001/index.shtml.

Vera, O. (2008). *Dificultades de estudiantes universitarios en algunos conceptos de diseño experimental***.** Tesis de Master. Universidad de Granada.

Vera, O. (2009). Dificultades de estudiantes universitarios en algunos conceptos de diseño experimental. Trabajo presentado en el *X Simposio de Educación Matemática.* Buenos Aires: Universidad de Lujan.

Vera, O. y Batanero, C. (2008). Análisis comparativo de ítems en estudios de evaluación. Intervalos de confianza, interpretación por estudiantes universitarios. Trabajo presentado en el *VIII Congreso Latinoamericano de Sociedades de Estadística*. Montevideo, Uruguay.

Vera, O. y Díaz, C. (2008). Algunas dificultades de estudiantes de psicología en relación al contraste de hipótesis. Trabajo presentado en el *I Encuentro de docentes e Investigadores de Estadística en Psicología*. Buenos Aires: Universidad de Buenos Aires.

Vera, O. y Díaz, C. (2010). Errores y dificultades de estudiantes de Psicología en relación a los intervalos de confianza y contraste de hipótesis. Trabajo presentado en el *Simposio de la SEIEM, Grupo de Trabajo "Didáctica de la Estadística, Probabilidad y Combinatoria"*. Universidad de Lleida. Barcelona, España.

Vera, O. y Díaz, C. (2013). Dificultades de estudiantes de psicología en relación al contraste de hipótesis. En J. M. Contreras (Ed.), *Actas de las I Jornadas Virtuales de Didáctica de la Estadística, Probabilidad y Combinatoria*. Granada: SEIEM.

Vera, O. y Díaz, C. (2014). Elements of variance analysis, evaluation of difficulties by questionnaire. En K. Makar (Ed.), *Proceedings of the Ninth International Conference on Teaching Statistcs.* Flagstaff, AR: International Association for Statistical Education. Online: http://iase-web.org/Conference_Proceedings.php.

Vera. O., Díaz, C. y Batanero, C. (2011). Dificultades en la formulación de hipótesis estadísticas por estudiantes de Psicología. *Unión, 27,* 41- 61.

Vera, O., Olivo, E. y Díaz, .C. (2009). Intervalos de confianza. Interpretación por estudiantes universitarios. Trabajo presentado en el *VI Congreso Iberoamericano de Educación Matemática*. Puerto Montt, Chile: Universidad de los Lagos.

Vera. O., Olivo, E., Alvarado, H. y Batanero, C. (2007). Estadística y competencias en la formación del ingeniero. En M. Molina, P. Pérez-Tyteca y M. A. Fresno (Eds.). *Jornadas de Investigación en el Aula de Matemáticas. Competencias matemáticas.* Granada. Sociedad Thales y Departamento de Didáctica de las Matemáticas. CD- ROM.

Viladrich, M. C. (2012). *Psicometría.* Barcelona: Universidad Autónoma.

Wang, Z. y Osterlind, S. J. (2013). Classical test theory. *Handbook of Quantitative Methods for Educational Research* (pp. 31-44). New York: Sense Publishers.

Watson, J. M. y Kelly, B. A. (2003). The vocabulary of statistical literacy. En *Educational Research, Risks, and Dilemmas: Proceedings of the Joint Conferences of the New Zealand Association for Research in Education and the Australian Association for Research in Education* [CD-ROM]. Auckland.

Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review , 67 (3)*, 223-265.

Zapico, M. (2006). Interrogantes acerca de análisis de contenido y del discurso en los textos escolares. *Primer seminario internacional de textos escolares* (pp. 149-155). Santiago: Ministerio de Educación.

Zuckerman, M., Hodgins, H. S., Zuckerman, A. y Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science, 4,*49-52.