

**UNIVERSIDAD DE GRANADA**

**Programa de Doctorado en Ciencias de la Educación**



**COMPRENSIÓN DEL MUESTREO Y LA DISTRIBUCIÓN  
MUESTRAL EN ESTUDIANTES DE EDUCACIÓN  
SECUNDARIA OBLIGATORIA Y BACHILLERATO**

**Tesis Doctoral**

**Nuria Begué Pedrosa**

**Dirigida por:**

**Dra. Carmen Batanero Bernabeu**

**Dra. M<sup>a</sup> Magdalena Gea Serrano**

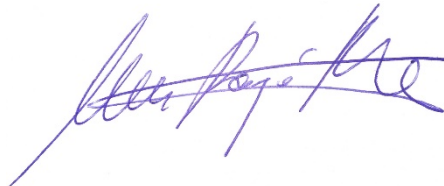
**Granada, 2019**



**UNIVERSIDAD DE GRANADA**  
**Programa de Doctorado en Ciencias de la Educación**

**COMPRENSIÓN DEL MUESTREO Y LA DISTRIBUCIÓN  
MUESTRAL EN ESTUDIANTES DE EDUCACIÓN  
SECUNDARIA OBLIGATORIA Y BACHILLERATO**

Memoria de TESIS DOCTORAL, realizada bajo la dirección de la Dra. Carmen Batanero Bernabeu y la Dra. M<sup>a</sup> Magdalena Gea Serrano en el Departamento de Didáctica de la Matemática de la Universidad de Granada, que presenta **Nuria Begué Pedrosa** para optar al grado de Doctor en el Programa de Doctorando de Ciencias de Educación.



Fd. Nuria Begué Pedrosa

V<sup>o</sup>B<sup>o</sup> de las Directoras,



Dra. Carmen Batanero Bernabeu



Dra. M<sup>a</sup> Magdalena Gea Serrano



Trabajo realizado en el marco del Proyecto EDU2016-7848-P, dentro del grupo de investigación FQM126, *Teoría de la Educación Matemática y Educación Estadística* (Junta de Andalucía).



El doctorando / The doctoral candidate [ **Nuria Begué Pedrosa** ] y los directores de la tesis / and the thesis supervisor/s: [ **Carmen Batanero Bernabeu y M<sup>a</sup> Magdalena Gea Serrano** ]

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date:

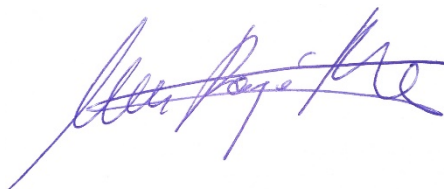
Granada, 8 de Octubre de 2019

Carmen Batanero Bernabeu M<sup>a</sup> Magdalena Gea Serrano  
Director/es de la Tesis / Thesis supervisor/s;



Firma / Signed

Nuria Begué Pedrosa  
Doctorando / Doctoral candidate:



Firma / Signed





## AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a la Dra. Carmen Batanero Bernabeu, por haberme dado la oportunidad de trabajar a su lado, lo que me ha permitido empezar a desarrollarme como profesora e investigadora. Ha sabido comprender mi situación profesional y personal caracterizándose por su paciencia y ayuda en todo momento del proceso.

También me gustaría agradecer a la Dra. María Madgalena Gea Serrano por su comprensión y apoyo durante este tiempo. Así mismo, quiero destacar su ayuda en la elaboración de este trabajo.

En segundo, quiero agradecer a todas aquellas personas que han mostrado no solamente su disponibilidad sino también la ilusión de formar parte de este trabajo, permitiéndome aplicar el instrumento de evaluación o, mostrando su preocupación en la elaboración y desarrollo del mismo. En concreto, quiero nombrar a Isabel, Cristina, Pilar, Sonia y Pablo, los cuales siempre me han contestado a esas llamadas que estaban envueltas en un favor, sin el cual no hubiera sido posible una relevante parte de este trabajo. Por otro lado, estas líneas no tendrían sentido sin nombrar a aquellos alumnos que mostraron una actitud receptiva a la participación, donde se les pedía confianza y sinceridad durante la resolución del cuestionario y no solamente respondieron al mismo, sino que algunos mostraran interés o, a veces, cierto desconcierto.

Para finalizar, aunque realmente sería el motivo impulsor que me llevó a plantearme la posibilidad de embarcarme en esta aventura, es mi pareja. Gracias Javier por encontrar el Máster que tenía todas las características que buscaba para continuar con mi carrera profesional y académica, pero, sobre todo, gracias por impulsarme a realizar la Tesis y por confiar en mí durante todo el proceso. Gracias por comprenderme, apoyarme y, como siempre digo, por proyectarme. También agradecer a mi familia y amigos más cercanos por mostrar su apoyo e sus mejores intenciones durante todo el proceso, puesto que nace desde el desinterés de que todo salga bien.



## RESUMEN

La tesis doctoral se centra en la evaluación de la comprensión de ideas elementales sobre el muestreo por parte de los estudiantes de educación secundaria obligatoria y bachillerato. El interés del tema se ha justificado por la relevancia que está alcanzando la inferencia en la actualidad y por ser el muestreo un tema fundamental para comprender otros conceptos y métodos asociados a la inferencia. Se utiliza como marco teórico el enfoque ontosemiótico de la cognición e instrucción matemática e igualmente nos apoyamos en el análisis de los documentos curriculares vigentes en Aragón y algunas orientaciones internacionales. Además, nos basamos en una serie de investigaciones previas sobre la comprensión de conceptos asociados al muestreo. Se realizan dos estudios empíricos que parten de dichos fundamentos:

- En el Estudio 1 evaluamos la comprensión que una muestra de 302 estudiantes de segundo y cuarto curso de educación secundaria obligatoria. Para ello se analizan sus respuestas de un cuestionario constituido por cuatro tareas que demandan proporcionar cuatro valores probables del número de éxitos en una distribución binomial. De ello deducimos la comprensión del valor esperado y la variabilidad en el muestreo.
- En el Estudio 2 realizamos el mismo análisis con una muestra de 234 estudiantes del último curso de bachillerato. Además, a una submuestra de 127 estudiantes se les demanda argumentar los valores dados para cada ítem. Esta ampliación nos ha permitido, mediante un análisis de contenido, identificar una serie de categorías de razonamientos de los estudiantes sobre la aleatoriedad y el muestreo. También identificamos también una serie de conflictos semióticos que se describen de manera detallada.

Las principales aportaciones del trabajo son: a) informar sobre la comprensión del valor esperado y la variabilidad en el muestreo y de las diferencias observadas en función de las variables de tarea de los ítems y del grupo de estudiantes; b) el cuestionario elaborado para la investigación y c) analizar el razonamiento sobre el muestro identificado mediante la justificación de una muestra de estudiantes. Finalmente, las publicaciones derivadas de la tesis son otras aportaciones que también se recogen en esta Memoria.



## ABSTRACT

This thesis is focused on the evaluation of secondary and high school students' understanding of elementary ideas about sampling. The interest of the topic is justified by the relevance that the inference is currently reaching, and because sampling is a fundamental subject to understand other concepts and methods associated to inference. The ontosemiotic approach to mathematical cognition and instruction is used as a theoretical framework and we also base on the analysis of curricular documents compulsory in Aragon and some international orientations. In addition, we base on previous research on understanding the concepts associated with sampling. Two empirical studies are carried out based on these foundations:

- In Study 1 we assess the understanding of a sample of 302 students in high school second and fourth grades. We analyze their responses to a questionnaire consisting of four tasks that require providing four probable values of the number of successes in a binomial distribution. From this we deduce the understanding of the expected value and the variability in sampling.
- In Study 2 we performed the same analysis with a sample of 234 students from the last year of high school. In addition, a subsample of 127 students is required to argue the values given for each item. This extension served, through a content analysis, to identify a series of categories of students' reasoning about randomness and sampling. We also identify a series of semiotic conflicts that are described in detail.

The main contributions of the work are: a) reporting on the understanding of the expected value and the variability in sampling and on the differences observed according to the task variables of the items and the group of students; b) the questionnaire prepared for the investigation and c) the analysis of reasoning about sampling in a sample of students. Finally, the publications derived from the thesis are other contributions that are also included in this report.



# ÍNDICE

INTRODUCCIÓN	5
CAPÍTULO 1. PLANTEAMIENTO DE LA INVESTIGACIÓN	9
1.1. Introducción	9
1.2. Justificación del tema	10
1.2.1. La educación estadística en la actualidad	10
1.2.2. El conocimiento estadístico como cultura	12
1.2.3. Inferencia estadística y muestreo	14
1.3. Marco teórico	17
1.3.1. Prácticas matemáticas, objetos y significado	18
1.3.2. Tipología de objetos matemáticos	22
1.3.3. La comprensión en el enfoque ontosemiótico	26
1.3.4. Función semiótica y conflictos semióticos	27
1.4. Objetivos e hipótesis de la investigación	29
1.4.1. Objetivos	29
1.4.2. Hipótesis iniciales	31
1.5. Marco curricular	32
1.5.1. El muestreo en los Decretos de Enseñanza Mínima	33
1.5.2. El muestreo en el Currículo Básico	41
1.5.3. Orientaciones curriculares internacionales	46
1.5.4. Conclusiones del estudio curricular	49
1.6. Organización del trabajo y resumen de la metodología	49
CAPÍTULO 2. ANTECEDENTES	53
2.1. Introducción	53
2.2. Niveles de uso de los conceptos de inferencia	55
2.3. Heurísticas y sesgos en el razonamiento sobre muestreo	58
2.4. Comprensión del concepto de muestra y de sesgo en el método de muestreo	62
2.5. Comprensión de las ideas de representatividad y variabilidad muestral	67
2.6. Comprensión de la distribución muestral	69
2.7. Comprensión del Teorema Central del Límite	74
2.8. Muestreo en el razonamiento inferencial informal de los estudiantes	76
2.9. Comprensión de la probabilidad desde el punto de vista frecuencial	80
2.10. Generación y reconocimiento de la aleatoriedad	86

2.11. Otras investigaciones	88
2.12. Conclusiones del estudio de los antecedentes	90
<b>CAPÍTULO 3. ESTUDIO 1. GENERACIÓN DE MUESTRAS ALEATORIAS POR ESTUDIANTES DE EDUCACIÓN SECUNDARIA OBLIGATORIA</b>	93
3.1. Introducción	93
3.2. Objetivos e hipótesis del Estudio 1	94
3.3. Contenidos sobre muestreo que se consideran en el estudio	97
3.4. Contexto educativo	99
3.5. Metodología	100
3.5.1. Descripción de la muestra	101
3.5.2. Método de recogida de datos	103
3.6. Análisis a priori del cuestionario y conocimientos evaluados	103
3.6.1. Análisis del ítem 1	104
3.6.2. Análisis del ítem 2	109
3.6.3. Análisis del ítem 3	111
3.6.4. Análisis del ítem 4	112
3.7. Resultados	115
3.7.1. Resultados en el ítem 1	116
3.7.1.1. Análisis global	116
3.7.1.2. Comparación por grupo	119
3.7.2. Resultados en el ítem 2	124
3.7.2.1. Análisis global	124
3.7.2.2. Comparación por grupo	126
3.7.3. Resultados en el ítem 3	130
3.7.3.1. Análisis global	130
3.7.3.2. Comparación por grupo	132
3.7.4. Resultados en el ítem 4	135
3.7.4.1. Análisis global	135
3.7.4.2. Comparación por grupo	138
3.8. Síntesis de resultados	141
3.8.1. Comparación de resultados según el tamaño de muestra	143
3.8.2. Comparación de resultados en experimentos con sucesos equiprobables o no	144
3.9. Análisis de casos atípicos	144
3.9.1. Estudio de casos con ausencia de variabilidad	144
3.9.2. Estudio de casos con variabilidad excesiva	147



3.10. Conclusiones del Estudio 1	148
3.10.1. Conclusiones respecto a los objetivos	148
3.10.2. Conclusiones respecto a las hipótesis	151
CAPÍTULO 4. ESTUDIO 2. JUSTIFICACIÓN DE ESTUDIANTES DE BACHILLERATO EN TAREAS DE GENERACIÓN DE MUESTRAS	155
4.1. Introducción	155
4.2. Objetivos e hipótesis del Estudio 2	156
4.3. Contexto educativo	159
4.4. Metodología	159
4.4.1. Cuestionario utilizado	160
4.4.2. Descripción de la muestra	163
4.4.3. Método de recogida de datos	164
4.5. Análisis cuantitativo	165
4.5.1. Método de análisis	165
4.5.2. Resultados en el ítem 1	168
4.5.3. Resultados en el ítem 2	175
4.5.4. Resultados en el ítem 3	181
4.5.5. Resultados en el ítem 4	186
4.5.6. Resultados comparados por ítem	192
4.5.7. Relación entre diversas respuestas	193
4.5.8. Puntuaciones totales y parciales en el cuestionario	195
4.6. Análisis cualitativo de las argumentaciones	199
4.6.1. Método de análisis	200
4.6.2. Categorías de análisis de las argumentaciones	201
4.6.3. Resultados en el ítem 1	225
4.6.4. Resultados en el ítem 2	227
4.6.5. Resultados en el ítem 3	229
4.6.6. Resultados en el ítem 4	231
4.6.7. Síntesis de argumentos en diferentes ítems	233
4.6.8. Estudio de casos con argumentos recurrentes	236
4.7. Conclusiones del Estudio 2	238
4.7.1. Conclusiones respecto a los objetivos	238
4.7.2. Conclusiones respecto a las hipótesis	241
CAPÍTULO 5. CONCLUSIONES	248

5.1. Introducción	248
5.2. Conclusiones respecto a los objetivos	249
5.3. Conclusiones respecto a las hipótesis iniciales	253
5.4. Principales aportaciones del trabajo	256
5.5. Limitaciones del trabajo	258
5.6. Líneas abiertas de investigación	259
REFERENCIAS	261
ANEXO	277
A1. Publicaciones derivadas de las Tesis	277

# INTRODUCCIÓN

La inferencia estadística es hoy día una herramienta básica en muchas de las actividades humanas, relacionadas con la ciencia, economía, gestión y política. De hecho, el gran progreso experimentado en las ciencias y la tecnología a lo largo del siglo XX y en el siglo actual se deben, sin duda, a la aplicación de métodos de inferencia, que han permitido matematizar ramas de la actividad humana que se rigen por leyes aleatorias.

Se puede pensar en la inferencia como una colección de herramientas que nos ayudan a generalizar las conclusiones obtenidas de una muestra a la población de donde se ha recogido. Según Batanero y Díaz (2015), las técnicas de inferencia se desarrollaron para fundamentar las formas de obtener un conocimiento general a partir del análisis de casos particulares, y por este motivo adquieren una gran importancia en la investigación en todas las ciencias empíricas. Efron y Hastie (2016) sugieren que la inferencia es una disciplina muy amplia, que se apoya en las matemáticas, las ciencias empíricas y la filosofía. Aunque su inicio data desde la época de Bayes, ya que su teorema, que sienta las bases de la inferencia bayesiana, realmente ha avanzado mucho en los últimos 200 años, gracias a que los ordenadores facilitan su aplicación a los no estadísticos.

No es entonces extraño que la inferencia sea una de las partes de la matemática más ampliamente estudiadas. Aunque su estudio formal se retrasa a la universidad o al bachillerato (en este caso sólo para los alumnos de Ciencias Sociales), encontramos en la educación secundaria obligatoria algunas ideas elementales sobre muestreo, cuya comprensión es básica para el estudio posterior de la inferencia. Por tanto, la elección del muestreo como tema de investigación se fundamenta en que constituye un contenido fundamental para el aprendizaje posterior y aplicación futura de la Estadística (Burril y Biehler, 2011). De hecho, el muestreo también forma parte de muchas de nuestras actividades cotidianas, ya que gran parte de nuestro conocimiento está basado en información obtenida a partir de muestras.

Aunque aparentemente este tema sea sencillo, la investigación previa, tanto con niños, como con sujetos adultos resumida, por ejemplo, en Castro Sotos, Vanhoof, Noortgate y Onghena (2007) o Harradine, Batanero y Rossman (2011) indica la existencia de creencias erróneas sobre el muestreo, que llevan a la toma de decisiones

equivocadas en diversas situaciones de la vida diaria o del trabajo. De ello se deduce la necesidad de analizar mediante la investigación dichos razonamientos y ver la forma de mejorarlos con la enseñanza.

En este trabajo nos centramos específicamente en la comprensión de algunas ideas elementales de muestreo en los estudiantes de educación secundaria obligatoria y bachillerato. Más concretamente, llevaremos a cabo dos estudios de evaluación con muestras de estudiantes de estos niveles educativos, que se describen en los capítulos 3 y 4. Esta Memoria tiene como finalidad presentar de manera sintetizada los estudios realizados y su fundamentación. A continuación, se presenta una síntesis de cada uno de los capítulos que configuran dicha Memoria.

En primer lugar, el Capítulo 1 se dedica al planteamiento del problema. Las primeras secciones se dedican a justificar la importancia del tópico elegido y seguidamente se expone un resumen de las principales ideas del marco teórico que lo fundamenta, que es el enfoque ontosemiótico de la cognición e instrucción matemática, elaborado fundamentalmente por Godino y sus colaboradores (Godino, 2002; Godino y Batanero, 1994; Godino, Batanero y Font, 2007). Descritos los elementos que se usan de este marco teórico, se presentan los objetivos e hipótesis del problema de investigación planteado. Además, en este primer Capítulo 1 se analiza la presencia de contenidos relacionados con el muestreo en los documentos curriculares de Educación Secundaria Obligatoria y Bachillerato. Igualmente se resume la organización de la Memoria y los principales aspectos metodológicos.

Posteriormente, el trabajo se centra en revisar y analizar en la literatura existente aquellas investigaciones centradas en la comprensión del muestreo o relacionadas con él. Por tanto, el Capítulo 2 presenta una síntesis de aquellas investigaciones que se han centrado en la comprensión de las propiedades del muestreo, la distribución muestral y la probabilidad, desde su significado frecuencial, así como sobre el Teorema Central del Límite y el concepto de aleatoriedad.

Este análisis previo fundamenta el diseño del instrumento de evaluación utilizado en los siguientes capítulos. Cada uno de los capítulos 3 y 4 comienza describiendo con detalle el contexto escolar, la metodología seguida y la muestra empleada. Previamente al análisis de las respuestas elaboradas por los estudiantes, se analiza el cuestionario comenzando con la reflexión sobre la respuesta esperada para cada uno de los ítems que lo constituye. Finalmente, se analizan los resultados obtenidos, detallando los

correspondientes a cada ítem y realizando tanto una síntesis como comparación de algunas variables del cuestionario.

El primero de estos estudios se presenta en el Capítulo 3, el cual está orientado a evaluar la comprensión de los estudiantes de segundo y cuarto curso de Educación Secundaria Obligatoria acerca de la relación entre el valor de una proporción en la población binomial y el valor esperado del número de sucesos de interés, en muestras tomadas de dicha población. También nos interesamos por su comprensión de la variabilidad en el muestreo de la proporción muestral y del efecto del tamaño de la muestra sobre dicha variabilidad. Por otro lado, se quiere estudiar las posibles diferencias en las respuestas de los estudiantes dependiendo de si se trabaja con sucesos equiprobables o no, la existencia de posibles sesgos y conflictos semióticos en su razonamiento y el efecto de la edad sobre todas estas variables.

Los resultados obtenidos se completan y profundizan en el Estudio 2, que se realiza con estudiantes de segundo curso de Bachillerato. En este caso, el cuestionario empleado anteriormente se completa pidiendo a los estudiantes justificar las respuestas. Esta ampliación permite completar el análisis cuantitativo con otro cualitativo y profundizar en la comprensión mostrada por los estudiantes, identificando algunos de sus conflictos semióticos.

Finalmente, el último capítulo expone las conclusiones del trabajo, donde se reflexiona sobre los objetivos previstos y las hipótesis descritas en relación con los resultados analizados. Adicionalmente se describen las aportaciones y limitaciones del trabajo así como las posibles vías para continuar investigación. Así mismo, se adjuntan las referencias bibliográficas que han sido consideradas para la elaboración del presente trabajo.

Este trabajo me ha permitido comprender y valorar la tarea investigadora que se desarrolla desde esta área, la didáctica de la matemática. La reflexión de los diferentes aspectos que configuran una investigación permite al investigador en formación conectar de una manera racional y reflexiva el trabajo que ha estado desarrollando, siendo una oportunidad de aprendizaje culmen para estos estudios doctorales.

Por otro lado, desde la perspectiva de docente, el tipo estudio realizado me ha permitido reflexionar sobre las intuiciones o concepciones previas que los estudiantes presentan sobre un contenido matemático concreto, valorando las dificultades asociadas que suponen una concepción errónea del mismo. Además, la identificación de errores de

comprensión de las ideas evaluadas constituye un argumento suficiente para apreciar la necesidad de adoptar una actitud crítica sobre nuestra práctica.

Finalmente destaco en los diversos capítulos los resultados parciales de la tesis que ya han sido publicados en revistas de Didáctica de la Matemática o congresos del área de conocimiento, cuyo listado se presenta como final de la Memoria.

## CAPÍTULO 1.

### PLANTEAMIENTO DE LA INVESTIGACIÓN

- 1.1. Introducción
- 1.2. Justificación del tema
  - 1.2.1. La educación estadística en la actualidad
  - 1.2.2. El conocimiento estadístico como cultura
  - 1.2.3. Inferencia estadística y muestreo
- 1.3. Marco teórico
  - 1.3.1. Prácticas matemáticas, objeto y significado
  - 1.3.2. Tipología de objetos matemáticos
  - 1.3.3. La comprensión en el enfoque ontosemiótico
  - 1.3.4. Función semiótica y conflictos semióticos
- 1.4. Objetivos e hipótesis de la investigación
  - 1.4.1. Objetivos
  - 1.4.2. Hipótesis iniciales
- 1.5. Marco curricular
  - 1.5.1. El muestreo en los Decretos de Enseñanza Mínima
  - 1.5.2. El muestreo en el Currículo Básico
  - 1.5.3. Orientaciones curriculares internacionales
  - 1.5.4. Conclusiones del estudio curricular
- 1.6. Organización del trabajo y resumen de la metodología

#### 1.1. INTRODUCCIÓN

Este primer capítulo de la memoria pretende, tanto describir y justificar el tema de investigación elegido, como resumir algunos fundamentos y la organización de nuestro trabajo. Por tanto, su lectura tiene como objetivo conseguir una primera comprensión de su contenido.

En primer lugar, se presenta un breve resumen sobre el estado actual de la investigación en educación estadística, con la finalidad de justificar el interés del tema elegido para la investigación. Luego, se exponen una serie de razones que acreditan la importancia de la cultura estadística en la sociedad actual y de la comprensión de las ideas básicas sobre el muestreo, tanto para el estudio de contenidos de estadística posteriores, como para la formación integral del estudiante.

El segundo bloque al que se dedica este capítulo es la descripción de los principales puntos del marco teórico que se utilizarán para basar la investigación. En concreto, se ha considerado el enfoque ontosemiótico, que constituye un marco teórico muy desarrollado actualmente y ofrece un gran número de herramientas, de las cuales solo se detallan las que se utilizarán en los diferentes estudios. Este marco teórico se viene utilizando en el grupo de investigación en el cual se integra esta tesis y proporciona

instrumentos que fundamentan los estudios empíricos que forman parte de nuestro trabajo. Esto nos permite plantear los objetivos e hipótesis de la investigación.

Seguidamente, se presenta un análisis curricular de los documentos oficiales vigentes durante el desarrollo de la investigación. Es decir, se centra en los contenidos correspondientes al muestreo presentes en el currículo actual y en el anterior, que puede haber seguido alguno de los estudiantes participantes en los estudios empíricos, puesto que la aplicación del instrumento de evaluación se produce en un periodo de cambio de la ley de educación. Finalmente, se presenta un breve esquema en el que se indica cómo se ha organizado esta investigación. Además, se incluye una descripción resumida de sus principales características metodológicas.

## **1.2. JUSTIFICACIÓN DEL TEMA**

La justificación del tema elegido para realizar la tesis doctoral se basa en tres argumentos que son detallados en los siguientes apartados. El primero de ellos, la consideración que actualmente adquiere la educación estadística como área de investigación y que se refleja en diferentes asociaciones, congresos y revistas específicas (Batanero, 2019; Zieffler, Garfield y Fry, 2018).

En segundo lugar, analizamos la relevancia actual de que todo ciudadano tenga una cultura estadística amplia, lo que implica la necesidad de realizar investigaciones que contribuyan a difundirla dentro y fuera de la escuela (Gal, 2002; Watson, 2006).

Finalmente, si consideramos la idea de muestreo, así como el conjunto de objetos matemáticos que se incluyen en esta parte de la estadística, las razones que apoyan su estudio se localizan tanto fuera como dentro de la estadística. Por un lado, es necesaria una adecuada comprensión del término para continuar con el estudio de otros conceptos de la estadística, en particular, de las ideas y procedimientos de inferencia. Por otro lado, el estudiante debe adquirir un razonamiento estadístico suficiente sobre muestreo para desenvolverse en múltiples situaciones de la vida personal y laboral.

### **1.2.1. LA EDUCACIÓN ESTADÍSTICA EN LA ACTUALIDAD**

El primer punto para justificar el tema de investigación es razonar la importancia que actualmente recibe la educación estadística como campo de investigación. Esta relevancia es debida al creciente interés que recibe la enseñanza de la estadística en las últimas décadas, debido a la necesidad reclamada por la UNESCO y otras instituciones, como el Instituto Internacional de Estadística (ISI), de ofrecer una formación estadística



general al ciudadano, con la finalidad de que sea competente en una sociedad dominada por la información (Batanero, 2004a; Engel, 2019; Ridgway, 2016).

Este reconocimiento ha trascendido igualmente a los planes de educación de diversos países, donde se observa el esfuerzo realizado en el diseño de los currículos y la elaboración de recursos didácticos específicos para la enseñanza de la estadística. Batanero, Godino, Green, Holmes y Vallecillos (1994) citaron, como ejemplos de este tipo de diseño curricular, el trabajo realizado por Holmes (1980) para el Schools Council Project on Statistical Education en Inglaterra y el libro sobre probabilidad de Godino, Batanero y Cañizares (1987). Más recientemente encontramos, por ejemplo, los libros de Batanero y Borovcnik (2016) y de Watson (2006), y también recomendaciones sobre la enseñanza incluidas en los diferentes capítulos de los libros de Chernoff y Sriraman (2014) y Jones (2005).

Batanero y Godino (2005) muestran el progreso que experimentó hasta el momento en que se publicó dicho artículo la educación estadística, para lo cual presentan una síntesis de los avances realizados desde la propia estadística, la psicología y la didáctica de la matemática, que continúan en la actualidad. En particular, en la psicología del desarrollo hubo un programa de investigación muy amplio sobre el razonamiento probabilístico de los niños y en la psicología del razonamiento otro sobre la toma de decisión en ambiente de incertidumbre. Batanero (2019) completa el anterior trabajo describiendo los cambios en la educación estadística en los últimos treinta años.

En relación con la propia estadística, los autores resumen la tendencia creciente a incluir la enseñanza de la estadística en todos los niveles educativos, a través de las acciones impulsadas por el Instituto Internacional de Estadística, ISI, que estableció en 1885 un Comité de Educación respaldado por la UNESCO. Uno de sus objetivos fue la creación de estudios reglados que formaran estadísticos, creando los Centros Internacionales de Educación Estadística, que también se han ocupado del diseño de materiales didácticos. El ISI también impulsa la puesta en marcha de los ICOTS, conferencias internacionales sobre la enseñanza de la estadística. En concreto, la última edición se ha celebrado en Japón en 2018.

En 1991 aparece la International Association for Statistical Education (IASE), que es una sección del ISI enfocada al ámbito educativo de la estadística. Su objetivo es promover, apoyar y mejorar la educación de la estadística en todos los niveles y con una proyección internacional. Para alcanzar sus propósitos, se encarga actualmente de organizar congresos y conferencias internacionales sobre la enseñanza de la estadística.

Otro indicador del creciente interés hacia la enseñanza de la estadística es la existencia de revistas específicas, entre las que citamos: *Journal of Statistics Education*, *Statistics Education Research Journal*, *Teaching Statistics*, *Technology Innovations in Statistics Education*, *Statistique et enseignement*, *Induzioni* y *Stochastik in der Schule*.

Desde el campo de la educación matemática, en los últimos años la estadística se ha ido incorporando en grupos de investigación en los principales congresos internacionales, como, por ejemplo, en el Congress of the European Society for Research in Mathematics Education (CERME) o el International Congress of Mathematics Education (ICME), organizados por la International Commission on Mathematical Instruction (ICMI).

Pero, además, hemos de citar los numerosos libros y capítulos dedicados a la estadística y probabilidad en los diferentes handbooks de educación matemática, o trabajos de survey, por ejemplo, Batanero, 2019; Batanero, Chernoff, Engel, Lee y Sánchez (2016); Bakker, Hahn, Kazak y Pratt (2018), Ben-Zvi, Makar y Garfield (2018); Chernoff y Sriraman (2014), Jones (2005), Jones, Lagrall y Mooney (2007), Jones y Thornton (2005), Shaughnessy (1992; 2007), Shaughnessy, Garfield y Greer (1996), o Zieffler, Garfield, y Fry (2018). Todo ello da idea de un crecimiento espectacular de la investigación sobre este tema, a la que tratamos de contribuir con nuestro trabajo.

Batanero (2019) destaca el desarrollo de la educación estadística en América Latina y la Península Ibérica, citando congresos como el *Encuentro Colombiano de Educación Estocástica* en Colombia, *Encuentro Internacional de la Enseñanza de la Probabilidad y Estadística* en México, *Encontro de probabilidades e estatística na escola* en Portugal, *Encuentro de Didáctica de la Estadística la Probabilidad y el Análisis de Datos* en Costa Rica, o el *Congreso Internacional Virtual de Educación Estadística* en España.

Los párrafos anteriores han mostrado el creciente interés en el estudio de estadística. A continuación, justificamos tomar como objeto de estudio el muestreo.

### **1.2.2. EL CONOCIMIENTO ESTADÍSTICO COMO CULTURA**

La importancia actual de la estadística es debida a su aplicabilidad en el ámbito profesional y personal, lo que requiere que se alcance una comprensión elemental para manejarse en la sociedad de la información. Además, el ISI ha popularizado la idea de

alfabetización estadística, que en inglés se denomina *statistical literacy* y que Batanero (2004a) tradujo como cultura estadística.

La autora se apoya en Ottaviani (1988), quien indicó que los estadísticos sienten la necesidad de difusión de la estadística, no solo como una técnica para tratar los datos cuantitativos sino, como una cultura, en términos de capacidad de comprender la abstracción lógica que hace posible el estudio cuantitativo de los fenómenos colectivos.

El interés por el tema quedó manifiesto ya en el Sexto Congreso Internacional sobre enseñanza de la estadística en la Ciudad del Cabo en 2002, cuyo tema fue “El desarrollo de una sociedad estadísticamente culta”, e igualmente en las sucesivas ediciones del Foro Internacional de Investigación sobre Razonamiento, Pensamiento y Cultura Estadística, (SRTL Forum) (<https://blogs.uni-paderborn.de/srtl/>) que cuenta ya con once ediciones y una serie de publicaciones y números monográficos en revistas de prestigio. Otra iniciativa importante es el blog Statistical literacy (<http://www.statlit.org/>) organizado por Milo Schield, que incluye artículos y libros, proyectos, competiciones y noticias sobre cultura estadística alrededor del mundo.

Esta preocupación ha llevado a interpretar cultura estadística (Arteaga, Batanero, Cañadas y Contreras, 2011; Batanero, 2004a; Gal, 2002; Engel, 2019; Watson, 2006) como conjunto de competencias, conocimientos y actitudes mínimas sobre la estadística que un ciudadano debe poseer en la sociedad de la información, las cuales le permiten enfrentarse a los retos que plantea la cantidad de información estadística en un mundo globalizado. Esta idea ha sido tratada por diferentes autores; Wallman (1993, p. 1) la define como:

La habilidad de comprender y evaluar críticamente los resultados estadísticos que permean nuestra vida diaria, unido a la habilidad de apreciar las contribuciones que el pensamiento estadístico puede hacer a las decisiones públicas y privadas, personales y profesionales.

Uno de los autores que más ha escrito sobre el tema es Gal (2002), quien indica que la cultura estadística sería lo que se espera de los adultos sobre su conocimiento básico del tema. Una parte de la cultura sería la capacidad de interpretar mensajes, términos y datos estadísticos en el día a día y otra parte utilizarlos para apoyar nuestros argumentos y opiniones. Los autores del proyecto GAISE (Franklin et al., 2007) sugieren que la cultura estadística es esencial para los consumidores, ciudadanos y profesionales, pero que la capacidad de razonamiento estadístico tarda mucho tiempo en desarrollarse, por

lo que es necesario incidir en su enseñanza desde los primeros niveles. Watson (2006, p. 11) indica que la cultura estadística es el:

Encuentro del azar y los datos en el currículo y la vida diaria, donde el encuentro implica contextos no investigadores y toma de decisión espontánea basada en la habilidad de aplicar útiles estadísticos, conocimiento general del contexto y competencias críticas de alfabetización.

Garfield, delMas y Zieffler (2010) asumen que la cultura estadística incluye el uso del lenguaje básico, el conocimiento del significado de los términos estadísticos, uso de símbolos sencillos y la interpretación de las representaciones de datos.

Batanero et al. (2013) indican los siguientes componentes de la cultura estadística:

- a. El desarrollo del conocimiento básico de los conceptos estadísticos y probabilísticos;
- b. La comprensión de los razonamientos y argumentos estadísticos cuando se presentan dentro de un contexto más amplio de algún informe en los medios de comunicación o en el trabajo;
- c. Una actitud crítica que se asume al cuestionar argumentos que estén basados en evidencia estadística (p.9).

Un punto importante en la cultura estadística son las actitudes positivas hacia la materia, que incluyen una serie de sentimientos y valoración de la estadística, así como una disposición a utilizar la estadística cuando sea necesario (Batanero, 2004a).

En resumen, la enseñanza de la estadística ha de contribuir a proporcionar estos conocimientos y actitudes a los estudiantes y la investigación didáctica debe colaborar a encontrar la mejor forma de conseguirlo.

### **1.2.3. INFERENCIA ESTADÍSTICA Y MUESTREO**

La inferencia estadística es una herramienta esencial, que forma parte de muchas ramas de la actividad humana, especialmente, en la investigación, la toma de decisiones, la planificación y gestión. Este hecho explica que la enseñanza de este contenido esté presente en la universidad y la formación profesional (Batanero y Díaz, 2015). La inferencia estadística se estudia actualmente en el segundo curso del Bachillerato de Humanidades y Ciencias Sociales, la mayoría de las carreras universitarias y diferentes postgrados. Además, en las pruebas de acceso a la universidad de Matemáticas Aplicadas a las Ciencias Sociales, se ha venido proponiendo con frecuencia en los últimos años un problema sobre contraste de hipótesis (López-Martín, Batanero, Díaz-Batanero y Gea, 2016).

A pesar de ser uno de los temas más enseñados, es también uno de los que peor se comprenden y aplican, posiblemente por ser poco el tiempo que se puede dedicar a su

enseñanza y por su complejidad, siendo muchos los autores que han criticado el uso de la inferencia en las revistas de investigación y la escasa comprensión de estudiantes y profesores acerca de muchos de los conceptos en que se basa la inferencia (Batanero, 2000; Castro-Sotos, Vanhoof, Noortgate y Onghena; 2007; Harradine, Batanero y Rossman, 2011). Según Batanero (2013), su justificación se halla en la gran cantidad de conceptos y procedimientos que intervienen en la inferencia, así como el tiempo que se dedica a su enseñanza.

La teoría del muestreo analiza la forma adecuada de recoger las muestras y las propiedades de las mismas para asegurar que las conclusiones que se obtienen sobre una población a partir de ellas sean válidas y fiables, es decir, no contengan sesgos y sean precisas. En Batanero, Begué y Gea (2018) sugerimos que el sentido del muestreo requiere, asimismo, el conocimiento de las siguientes ideas estadísticas fundamentales (Batanero, 2004b; Burrill y Biehler, 2011):

- *Aleatoriedad y variabilidad aleatoria.* Los estudiantes deben aprender a reconocer las diferentes fuentes de variabilidad en inferencia (Reading y Shaughnessy, 2004): a) variabilidad de resultados en un experimento aleatorio; b) variabilidad en los datos; c) variabilidad en una variable aleatoria; d) variabilidad en las muestras o la distribución muestral. En el muestreo intervienen estos diferentes tipos de variabilidad.
- *Distribución.* Este es un concepto que no se utiliza en otras ramas de las matemáticas; sin embargo, en el estudio del muestreo es preciso diferenciar los tres tipos de distribución que se trabajan en inferencia (Harradine, Batanero y Rossman, 2011): a) La distribución teórica de probabilidad, definida por uno o varios parámetros; b) La distribución del conjunto de datos que constituye una muestra aleatoria simple, en la que se obtiene el valor de un estadístico, que se corresponde con el parámetro, pero se calcula en la muestra y c) La distribución muestral de un estadístico en todas las posibles muestras del mismo tamaño y condiciones. Puesto que se pueden extraer muchas muestras de la misma población, el estadístico es una variable aleatoria y su distribución es la distribución muestral.
- *Muestreo y estimación.* El estudio de la inferencia comienza por la distinción entre población y muestra y las ideas básicas de representatividad y variabilidad muestral. La idea básica en estimación es que es posible generalizar los datos de una muestra a una población mayor. Es importante desarrollar la competencia para generalizar los valores de los estadísticos obtenidos en las muestras y obtener

información sobre el parámetro en la población. Esta información siempre va unida a cierto margen de variabilidad.

- *Probabilidad.* Aunque la probabilidad es fundamental en toda la estadística, no se destaca suficientemente la relación del muestreo con el enfoque frecuencial de la probabilidad, pues en este enfoque aparecen muchas ideas ligadas al muestreo. Recíprocamente, para poder estimar un parámetro a partir de un estadístico muestral, utilizamos la probabilidad para determinar la variabilidad de la estimación.

Como el muestreo es la base de la inferencia, es importante que los estudiantes logren una comprensión suficiente del mismo antes de continuar con el estudio de otros temas, como el contraste de hipótesis y el intervalo de confianza, pues, de lo contrario, los errores de comprensión del muestreo van a proyectarse en los contenidos posteriores. Heitele (1975) incluyó el concepto de muestreo en su lista de diez ideas estocásticas fundamentales, debido a su presencia en multitud de situaciones aleatorias, tanto en la vida cotidiana como en la profesional, en las que aparece implícitamente esta idea.

Además, la formalización de este concepto supuso un gran avance en el desarrollo de la estadística como ciencia, permitiendo superar la concepción descriptiva, en la que su estudio queda reducido a los datos disponibles. La inferencia permite planificar y predecir tanto comportamientos como sucesos en poblaciones y a lo largo del tiempo, lo que ha ayudado mucho al progreso de la mayoría de las ciencias. En relación con la enseñanza de este contenido matemático, Heitele (1975) señala que su inclusión en la enseñanza puede iniciarse desde la educación primaria, con diversos grados de formalización.

Además de enlazar la estadística y la probabilidad, todo nuestro conocimiento y juicios sobre el mundo o las personas está basado en el muestreo, ya que, usualmente, solo podemos estudiar u observar una parcela del fenómeno de interés. Actualmente, los resultados obtenidos mediante muestreo por medio de encuestas o investigaciones en campos que nos afectan (medicina, política, educación, consumo, etc.), se publican cada vez, con mayor frecuencia, en los medios de comunicación. Estas noticias, o los mismos estudios, a veces contienen sesgos no asumidos o intencionados, con fines comerciales o de otro tipo. Entonces, es relevante que los estudiantes aprendan a razonar sobre el muestreo como parte de su cultura estadística (Franklin et al., 2007) y la comprensión

adecuada de la idea de muestro es necesaria en la vida cotidiana de la persona.

Además, el muestreo recibe actualmente una gran atención en la investigación didáctica, pues las ideas asociadas al mismo subyacen en el trabajo con la simulación, cuya utilización en el aula es recomendada para mejorar la comprensión de la probabilidad y la inferencia estadística (Eichler y Vogel, 2014; Huerta, 2015).

### 1.3. MARCO TEÓRICO

Toda investigación requiere un marco teórico que la fundamente, permita describir de forma más precisa sus objetivos e hipótesis, así como discutir los resultados que se obtengan. Este marco teórico generalmente se elige teniendo en cuenta, tanto la tradición del equipo investigador, como su utilidad para el problema planteado.

Siguiendo estos criterios, el marco teórico que se ha seguido para enmarcar el estudio se corresponde con el enfoque ontosemiótico (EOS) sobre el conocimiento y la instrucción matemática (Godino, 2002; 2017; Godino y Batanero, 1994; 1998; Godino, Batanero y Font, 2007; 2019). Dicho enfoque ha servido de base durante treinta años a la investigación realizada por el grupo de educación estadística de la Universidad de Granada, en el cual me he integrado para realizar la tesis doctoral. Además, actualmente en este grupo se desarrolla un proyecto de investigación financiado, sobre el desarrollo del razonamiento inferencial de los estudiantes, basado en el enfoque ontosemiótico, en el que se inscribe mi tesis.

Este enfoque proporciona una multitud de herramientas que permiten abordar la investigación desde el currículo, el estudiante, el profesor o la enseñanza, de una forma global y relacionada, lo que nos ha parecido importante para poder continuar nuestro trabajo futuro sobre otros puntos no abordados en esta tesis. Este modelo considera que la matemática puede verse desde tres puntos de vista diferenciados (Godino y Batanero, 1998):

- Como actividad de resolución de problemas internos o externos a la propia matemática.
- Como lenguaje simbólico en el que se expresan las situaciones problemas y sus soluciones. Dicho lenguaje tiene una doble función, comunicativa e instrumental, pues permite operar con los objetos inmateriales.
- Sistema conceptual lógicamente organizado, que no se reduce a la suma de componentes aislados, porque lo que constituye un sistema son precisamente las interrelaciones entre sus componentes.

El EOS parte de la idea de situación-problema, que se interpreta en sentido amplio como cualquier actividad, pregunta o cuestión que requiera una actividad de matematización. Englobaría lo que se entiende por situaciones que no tienen una solución inmediata para el resolutor. Por tanto, podemos considerar situación-problema los enunciados incluidos, tanto en los libros de texto o pruebas de evaluación como, los que resultan de aplicar las matemáticas en múltiples profesiones o incluso en la vida cotidiana.

Además, en esta teoría, se asume que los objetos matemáticos surgen de las prácticas realizadas en la resolución de problemas (Godino, 2017; Godino y Batanero, 1994). Ello es debido a que se considera que son las situaciones en las que se debe utilizar un cierto objeto matemático las que le dan sentido, es decir, las dotan de significado.

Como se ha indicado anteriormente, se describe a continuación aquellos elementos que han sido considerados, del presente marco teórico, para el desarrollo de nuestra tesis.

### **1.3.1. PRÁCTICAS MATEMÁTICAS, OBJETO Y SIGNIFICADO**

Como hemos indicado, en el marco teórico EOS se parte de la idea de situación-problema y de las prácticas realizadas para la resolución de las mismas. Para los autores del EOS, una práctica matemática es cualquier acción o expresión usada para resolver un problema matemático, generalizarlo, o bien, comunicar la solución a otros.

Asociado a cada campo de problemas particular, Godino y Batanero (1994) asumen que existen unas prácticas prototípicas que lo definen. Los autores diferencian entre prácticas operativas y discursivas. Las primeras se realizan al operar con objetos matemáticos durante la resolución de problemas, mientras que las justificaciones y argumentos que apoyan dicha resolución o que se usan para explicar una propiedad o una solución a otra persona serían prácticas discursivas.

Además, Godino et al. (2007; 2019) diferencian entre prácticas personales o institucionales, es decir, las que son privativas de un sujeto y las realizadas por una institución de personas interesadas en resolver un mismo tipo de problemas. Algunas instituciones serían las escolares, también instituciones profesionales o matemáticas.

Muchas prácticas matemáticas son compartidas por grupos de personas que tienen interés en resolver los mismos problemas matemáticos y que la sociedad los reconoce



como institución. Estas prácticas institucionales dependen de los instrumentos disponibles en la misma, y las reglas que las regulan (Godino y Batanero, 1994; Godino, Batanero y Font, 2007; 2019). Por otra parte, el sujeto realiza prácticas personales para resolver los problemas (que podrían no coincidir con las realizadas en la institución). Sin embargo, las matemáticas no son sólo personales, sino también sociales, en el aspecto de que normalmente se aprenden en grupo y cuando se investiga o resuelven problemas se comparten los resultados con los demás.

Los objetos matemáticos que, como se ha dicho, surgen del conjunto de prácticas de la actividad matemática, se conciben como símbolos de unidades culturales que nacen de los usos que hacen de los mismos las personas e instituciones y que evolucionan con el paso del tiempo en la actividad de resolución de problemas (Godino, 2002; 2017).

En nuestro caso nos interesamos por las situaciones-problema relacionadas con el muestreo, por ejemplo, elegir una muestra de forma adecuada, calcular el número posible de muestras de una población con ciertas características o decidir cómo estimar un parámetro de una población a partir de una muestra. Algunas prácticas matemáticas para resolver estos y otros problemas serían, realizar un sorteo (para elegir la muestra), realizar un muestreo sistemático (elegir un objeto de cada  $n$ , en forma secuencial), tomar un estadístico de la muestra, para estimar el parámetro de la población, o calcular la probabilidad de que el estadístico esté a cierta distancia del parámetro.

### **Significado y sus tipos**

La reflexión sobre el significado de los símbolos matemáticos ha sido una preocupación importante en la didáctica. Godino (2002) examina los presupuestos de diferentes autores y propone su propia visión sobre la idea de significado. Señala la necesidad de estudiar la relación entre pensamiento, lenguaje y problemas, e intenta desarrollar una semiótica específica que sea útil a la educación matemática. Indica que los símbolos representan objetos matemáticos y que el fin principal de la enseñanza no es solo conseguir que los estudiantes dominen la sintaxis del lenguaje simbólico, sino todo su significado.

Godino y sus colaboradores (Godino, 2002; Godino y Batanero, 1994; 1998) conciben el *significado de un objeto matemático* como el sistema de prácticas que realiza una persona (significado personal) o se realizan en una institución (significado institucional) para resolver un tipo de situaciones-problema de las que surge el objeto.

El significado del objeto depende de la institución en la que se use, consecuentemente, cabe definir el significado de un objeto matemático (O) como “el sistema de prácticas institucionales asociadas al campo de problemas de las que emerge O en un momento dado” (Godino y Batanero, 1994, p. 338). Su modelización tiene en cuenta (Godino, 2002, p. 3):

- La variedad de objetos que forman parte de la actividad matemática.
- La diversidad de procesos de interpretación de los distintos objetos, cuando se resuelven problemas.
- Los múltiples de contextos y circunstancias que determinan el aprendizaje.

En nuestro caso particular, el interés recae sobre el concepto de muestreo y los problemas relacionados con la toma de muestras de una población o la estimación de una característica de la población a partir de las muestras. De estos problemas surgen diferentes objetos matemáticos como los de población y muestra, estimación, estimador, variabilidad y representatividad o distribución muestral. Por un lado, el significado institucional que nos interesa es el fijado en los currículos para la educación secundaria y bachillerato, mientras que el significado personal sería el conjunto de prácticas que los estudiantes realizan cuando se enfrentan a situaciones de muestreo.

Según Godino y colaboradores (Godino, 2002; Godino, Batanero y Font, 2007; 2019), en el significado institucional de un objeto matemático se distinguen, a efectos de llevar a cabo una investigación o una programación de aula, los siguientes tipos (Figura 1.3.1):

- *Referencial*: es el sistema de prácticas considerado por el profesor o el investigador para la elaboración de una investigación o de una programación de aula. Para determinarlo, se utilizarán diversas fuentes necesarias: libros, orientaciones curriculares, objetivos institucionales y experiencias. Con esto se obtiene el significado del objeto matemático según su origen, evolución, contextos en los que se utiliza, etc. Para el caso del muestreo, el significado referencial sería la unión de los diferentes significados descritos en trabajos didácticos de síntesis, como Batanero y Borovcnik (2016), en análisis curriculares, como el de Begué y Batanero (2017) o Begué, Ruiz, Gea y Batanero (2017) y en libros universitarios sobre muestreo.
- *Pretendido*: es el sistema de prácticas plasmado en la planificación del proceso de

enseñanza y aprendizaje que se realizará sobre cierto objeto matemático; o bien, el que se pretende investigar en un trabajo sobre enseñanza. Suele ser una parte del significado de referencia, y se intenta que sea representativo del mismo. La razón de trabajar con un significado pretendido es que el referencial es demasiado amplio para el tiempo disponible o demasiado complejo para el conocimiento de los estudiantes que participan en la enseñanza o la investigación.

- *Implementado*: es el sistema de prácticas efectivamente desarrollado por el profesor o el investigador en la clase de matemáticas con el estudiante, para el estudio del objeto matemático, y será fundamental para el diseño de las evaluaciones que, a futuro, deberá rendir como parte del proceso de enseñanza y aprendizaje. Generalmente tiene algunas diferencias con el significado pretendido, porque la misma dinámica del aula hace que durante el proceso de enseñanza se produzcan pequeños cambios en el significado pretendido.
- *Evaluado*: es el sistema de prácticas que selecciona el profesor, mediante un conjunto de tareas y/o pautas de observaciones, con el fin de evaluar el significado personal del estudiante sobre el objeto matemático estudiado. O bien, el significado que evalúa un investigador dentro de su trabajo.



Figura 1.3.1. Tipos de significado (Godino, Batanero y Font, 2007, p. 6)

Por otra parte, en el significado personal de un objeto matemático se distinguen los siguientes tipos (Figura 1.3.1):

- *Global*: es el sistema de prácticas que un sujeto puede ser capaz de realizar,

respecto a un determinado objeto matemático. Incluye todo lo que la persona es capaz de hacer o decir sobre el objeto en cuestión, es decir, todo su conocimiento sobre dicho objeto. Pero como es muy amplio, es difícil describir el significado global de una persona en una investigación, por lo que se consideran otros tipos.

- *Declarado*: es el sistema de prácticas que un sujeto pone de manifiesto en una evaluación, como por ejemplo, la propuesta por un profesor. En los estudios de evaluación que llevaremos a cabo será el significado que podamos deducir de las respuestas de los estudiantes y dependerá del tipo de tareas que se hayan utilizado en la evaluación. Si estas son una muestra representativa del significado pretendido del objeto, se podrá acceder a gran parte del significado global del estudiante.
- *Logrado*: es el sistema de prácticas puesto de manifiesto por un sujeto antes (inicial) o después de la enseñanza (final), que se consideran correctas por la institución. Sería la parte correcta del significado declarado por el estudiante.

En la Figura 1.3.1 se pueden observar los tipos de significado implicados en el proceso de enseñanza y aprendizaje, donde, en la enseñanza de un objeto matemático, el estudiante participa en una comunidad de prácticas en la que se apropia del significado institucional. En nuestro estudio, consideraremos el significado evaluado en los instrumentos utilizados y el significado declarado por el estudiante, diferenciando en este el logrado del no logrado. Más concretamente, nos interesamos por determinar cuáles de las prácticas matemáticas realizadas por los estudiantes coinciden con las que se consideran correctas dentro de la institución, en nuestro caso, la educación secundaria y bachillerato. En particular, nuestro interés radica en identificar la *comprensión* que una muestra de estudiantes presenta sobre el concepto de muestreo. Entenderemos la comprensión en el sentido de Godino (1996; 2002), es decir, como un rasgo psicológico que no puede ser observado directamente en el sujeto, pero sí puede evaluarse indirectamente a partir de las prácticas personales.

### **1.3.2. TIPOLOGÍA DE OBJETOS MATEMÁTICOS**

En este marco teórico, se considera como objeto todo lo que podemos indicar o a lo que se puede hacer referencia en las prácticas matemáticas. Por otro lado, en el EOS los objetos matemáticos se pueden clasificar según la función desempeñada, dando lugar a las siguientes categorías que denomina “entidades primarias” (Godino, 2002):

- *Situaciones-problemas*. Aquellas en las que surge la actividad matemática, en

nuestro caso, las situaciones que motivan la idea de muestreo; por ejemplo, determinar la composición de una muestra en el muestreo estratificado o estimar las características de una población a partir de una muestra. Pueden ser situaciones más o menos abiertas, extramatemáticas o intramatemáticas. Por ejemplo, sería una situación extramatemática determinar el tamaño necesario de una muestra en cierto estudio e intramatemática encontrar el estimador de máxima verosimilitud para la varianza.

- *Lenguaje*. Son los términos, expresiones simbólicas, tablas o gráficos usados para representar la información proporcionada en una situación problemática, las operaciones y objetos utilizados en su resolución y las soluciones encontradas. Por ejemplo, las palabras población, muestra, aleatoria, independiente, estimador, error de muestreo, así como las expresiones simbólicas y los gráficos utilizados en el estudio del muestreo.
- *Conceptos*. Son los objetos matemáticos que se utilizan implícita o explícitamente en la actividad matemática y que pueden ser definidos. Asociados a la idea de muestreo aparecen, entre otros, los de población, elemento, muestra, variable, distribución (de los datos, de la población y distribución muestral), estadístico, parámetro, estimación, precisión de la estimación.
- *Propiedades o proposiciones* que relacionan entre sí los conceptos. Un ejemplo es que la variabilidad de la distribución muestral es menor al aumentar el tamaño de las muestras. Otro, el hecho de que la muestra es un subconjunto de la población, pero que muestras diferentes pueden compartir un mismo elemento.
- *Procedimientos*. Incluyen los algoritmos, operaciones o técnicas que constituyen parte de la enseñanza; por ejemplo, las distintas técnicas de muestreo o de estimación de los parámetros. Entre ellos, tendríamos el de seleccionar muestras al azar o estratificadas, realizar cálculos de probabilidades con la distribución muestral, etc.
- *Argumentos*. Son las justificaciones empleadas para mostrar la validez de una proposición o de la solución a un problema. Estas justificaciones incluyen las habituales en matemática, como la inducción, deducción o el análisis-síntesis. También consideramos las argumentaciones informales de los estudiantes en las tareas que les presentamos.

En la enseñanza o en la actividad matemática, estos objetos se relacionan entre si formando *configuraciones*, que serán epistémicas si son propias de una institución matemática o de enseñanza y cognitivas si son específicas del estudiante. El lenguaje permite representar a todos los demás objetos (papel representacional) y operar con ellos (papel instrumental). Los problemas contextualizan la actividad matemática; los procedimientos constituyen la parte práctica de la misma, mientras que el resto de objetos surgen de la reflexión sobre la actividad matemática (Godino, 2002). En la Figura 1.3.2 se muestra las relaciones entre los diferentes tipos de objetos de una configuración, que surgen en una práctica matemática.

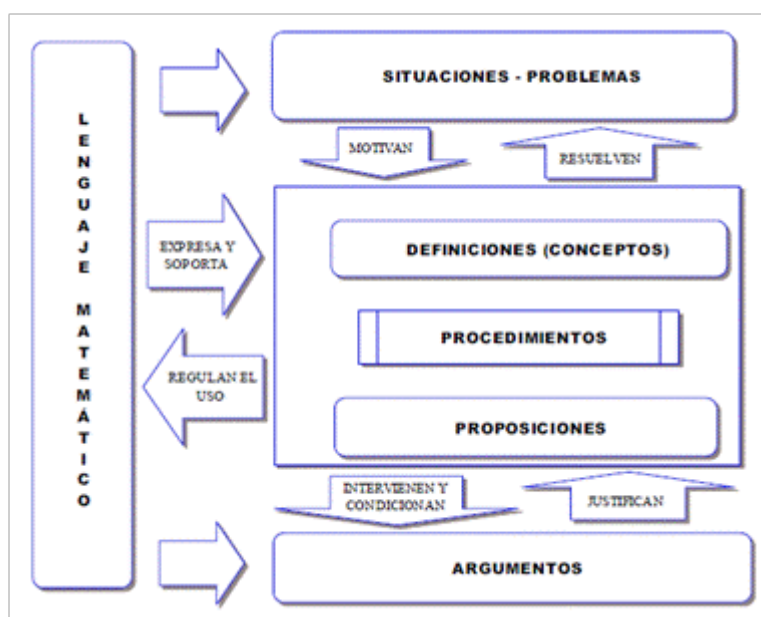


Figura 1.3.2. Configuración de objetos primarios (Godino, Batanero y Font, 2007, p. 7)

### Facetas o dimensiones del conocimiento

Los anteriores objetos matemáticos pueden ser considerados desde diferentes puntos de vista o dimensiones (Font y Godino y D'Amore, 2007):

- *Personal – institucional*. Cuando el objeto o las prácticas son característicos de una institución, diremos que el objeto es visto desde una perspectiva institucional; por ejemplo, las prácticas contenidas en libros de texto o documentos curriculares. Mientras que si son idiosincráticos de una persona, como las respuestas a una tarea, son personales. La importancia de considerar ambas facetas en la enseñanza, es su utilidad para analizar la comprensión personal.
- *Unitaria – sistémica*. Los objetos matemáticos pueden ser estudiados globalmente, como una unidad (la distribución muestral) y analizar sus componentes uno a uno

(función de densidad, media, varianza). También hay conceptos como “conjunto de datos”, que está compuesto de muchos elementos, puede tratarse como sistémico.

- *Ostensiva – no ostensiva*. Los objetos matemáticos son no ostensivos, es decir, inmateriales y no podemos visualizarlos en sí mismos; para trabajar con ellos se recurre a representaciones que tienen un carácter ostensivo y permiten mostrar de algún modo sus características. Aunque, cualquier objeto ostensivo (una expresión, un gráfico, etc.) puede también utilizarse en forma no ostensiva (por ejemplo, el signo de multiplicar en la notación algebraica). También, un cálculo puede ser realizado por una persona de manera ostensiva, o no ostensiva (Godino, 2002).
- *Expresión – contenido*. Todos los objetos pueden ser antecedentes o consecuentes de una función semiótica, que consiste en la relación de una expresión (antecedente) con un contenido (consecuente), establecida por una persona o institución con respecto a algún criterio de correspondencia. A lo largo de la resolución de un problema, se encadenan muchas funciones semióticas que forman parte de su solución.
- *Extensivo – intensivo*. El objeto se puede considerarse como una generalidad (concepto de muestra) o bien como ejemplos particulares del mismo (una muestra particular generada por el estudiante). Así, cualquier objeto puede considerarse como caso particular (extensivo) o como una clase más general o abstracta de objetos.

Los diferentes tipos de objetos matemáticos y facetas consideradas en el marco teórico se presentan en la Figura 1.3.3, donde además se resaltan los procesos matemáticos asociados, que son los siguientes:

- *Institucionalización – personalización* (faceta institucional y personal); se refiere al paso de lo institucional a lo personal, y viceversa. Un cierto contenido se puede adoptar en una institución (se institucionaliza) o bien puede ser aprendido por un estudiante (se personaliza).
- *Generalización – particularización* (faceta extensiva e intensiva). De muchos casos particulares de un mismo objeto (por ejemplo, de muchas muestras particulares) se generaliza a la idea abstracta de dicho objeto (en el mismo ejemplo, el de muestra, que se refiere a cualquiera de ellas). Recíprocamente, se puede pedir al estudiante un ejemplo (un caso particular) del concepto (general) de muestra.

- *Análisis/descomposición – síntesis/reificación* (faceta unitaria y sistémica). Esta faceta se pone de manifiesto en la resolución de problemas, donde constantemente se descomponen y recomponen los objetos matemáticos.
- *Materialización/concreción – idealización/abstracción* (faceta ostensiva y no ostensiva); cuando se pasa de lo inmaterial a la materialización del objeto, y viceversa.
- *Expresión/representación – significación* (faceta expresión y contenido). Se refieren a los componentes de la función semiótica, donde cualquier tipo de objeto puede ser expresión o contenido de la misma.

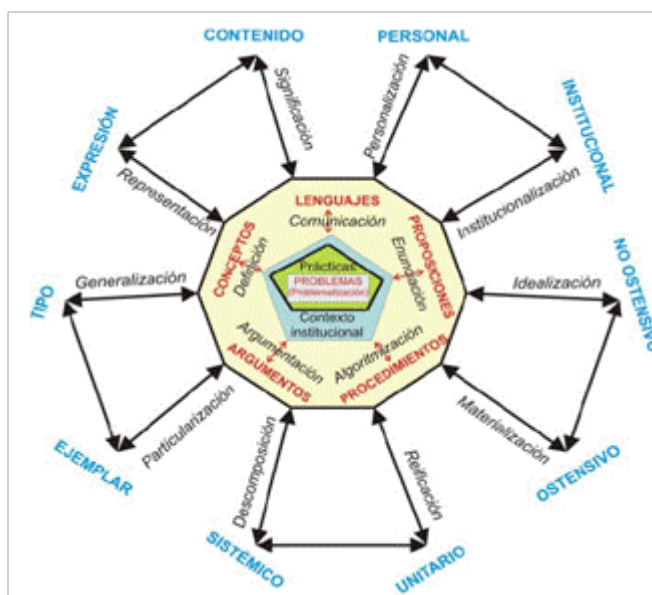


Figura 1.3.3. Objetos y facetas (Godino, Batanero y Font, 2007, p. 10)

### 1.3.3. LA COMPRENSIÓN EN EL ENFOQUE ONTOSEMIÓTICO

En los estudios empíricos que se desarrollan como parte de la tesis doctoral, nos interesamos por la comprensión que los estudiantes muestran de ciertos objetos matemáticos relacionados con el muestreo. Por tanto, es necesario preguntarnos cómo el marco teórico presentado define el concepto de comprensión.

Siguiendo el enfoque ontosemiótico (Godino y Batanero, 1994; Godino, et al. 2007), la comprensión personal será considerada como la apropiación del significado institucional del objeto matemático por el individuo, por lo que es un proceso progresivo. Puesto que los objetos han llegado a un estado de abstracción y complejidad notable dentro de la matemática, el estudiante tendrá que comenzar por adquirir una versión simplificada del significado del mismo que, progresivamente, se enriquece. Esto



implica la necesidad de un periodo largo de aprendizaje.

En el modelo de comprensión propuesto por Godino (1996), se tienen en cuenta dos ejes: uno descriptivo, que indica los aspectos o componentes de los objetos de los cuales queremos evaluar su comprensión; por ejemplo, podríamos estar interesados en evaluar la comprensión de las ideas de representatividad y variabilidad muestral o de la relación entre el estadístico y el parámetro. El segundo es el eje procesual, que indica las fases o niveles en la comprensión; en el ejemplo, podríamos utilizar algunos de los modelos que en la investigación se han propuestos para poder clasificar jerárquicamente la comprensión lograda por los estudiantes de conceptos de muestreo.

Los componentes que deseamos evaluar en este trabajo son algunos conceptos relacionados con el muestreo y sus propiedades, que se detallarán en los diferentes estudios empíricos. Podemos asumir también, que las fases o niveles están asociados a los diferentes cursos donde se realiza la evaluación, ya que a cada estudiante sólo se le evalúa una única vez. Pero al ser de diferentes edades, el utilizar tres cursos diferentes, permitirá evaluar el avance de los estudiantes en la comprensión.

La comprensión no es directamente observable, sino que debe deducirse de respuestas de los estudiantes a tareas de evaluación, debido a que es algo mental que debe deducirse cuando observamos las respuestas que nos dan. En consecuencia, la evaluación de la comprensión sería el estudio de la correspondencia entre los significados personales e institucionales de un cierto objeto matemático. Se dirá que el estudiante “comprende” si las prácticas que realiza al proponerle una tarea de muestreo son adecuadas y corresponden a las que se consideran correctas en la institución escolar. Siguiendo a estos autores, en este trabajo se evaluará la comprensión sobre el muestreo de los estudiantes que participan en el estudio mediante el análisis de las respuestas (entendidas como prácticas), que elaboran al responder a un conjunto de tareas que caracterizan el instrumento propuesto para su evaluación.

#### **1.3.4. FUNCIÓN SEMIÓTICA Y CONFLICTO SEMIÓTICO**

Dada la diversidad de objetos que se utilizan en las prácticas matemáticas, el carácter inmaterial de los objetos y la variedad de representaciones utilizadas, el trabajo matemático requiere un uso constante de procesos interpretativos, que pueden llegar a ser incorrectos cuando quienes los realizan son los estudiantes. Al trabajar con un problema, el estudiante necesita interpretar expresiones matemáticas, en las que los símbolos remiten a conceptos, propiedades o procedimientos. Sin embargo, puede

malinterpretar alguna de estas expresiones e igualmente puede utilizar una representación inadecuada cuando expone los resultados de su trabajo.

Para tener en cuenta esta realidad, el EOS adopta de otros autores la idea de función semiótica, que refleja las relaciones entre objetos y los correspondientes procesos de interpretación (Godino, 2002; Godino, Batanero y Font, 2007; 2019). En Godino y Batanero (1998), se describe la noción de función semiótica, como una "correspondencia entre conjuntos", que pone en juego tres componentes:

- Un plano de expresión (objeto inicial, considerado frecuentemente como el signo);
- Un plano de contenido (objeto final, considerado como el significado del signo, esto es, lo representado, lo que se quiere decir, a lo que se refiere un interlocutor);
- Un criterio o regla de correspondencia, esto es, un código interpretativo que relaciona los planos de expresión y contenido.

Una función semiótica sería *representacional* (un objeto se pone en lugar de otro), *instrumental* (un objeto usa a otro u otros como instrumento), o *componencial o cooperativa* (dos o más objetos componen un sistema del que emergen nuevos objetos). Estas funciones semióticas pueden producir significados sistémicos o elementales según el contexto. Por ejemplo, “la muestra tomada para esta investigación” tiene un significado elemental, mientras que “la muestra aleatoria de n elementos” tiene un significado sistémico.

La producción o interpretación de cada función semiótica implica un acto de semiosis por el estudiante y constituye un conocimiento personal, del agente interpretante. En este sentido, un conocimiento se entiende como el contenido de una función semiótica (Godino, Batanero y Font, 2007).

Podemos interpretar la comprensión de un objeto O por parte de un sujeto (institución o individuo) en términos de las funciones semióticas que X puede establecer, en unas circunstancias fijadas, en las que se pone en juego O como expresión o contenido (Godino, Batanero y Font, 2007, p. 132). La expresión y el contenido de la función semiótica pueden ser de cualquier tipo de objeto. Es decir, por ejemplo, se puede asociar una propiedad a un concepto (y la asociación puede ser correcta o incorrecta). En consecuencia, el estudiante en su trabajo lleva a cabo sus propias funciones semióticas, que pueden concordar o no con las admitidas dentro de la institución de enseñanza (Godino y Batanero, 1994). Las prácticas que desarrolla el estudiante serán consideradas correctas si se conforman a la institución e incorrectas en

caso contrario. Los autores proponen la idea de *conflicto semiótico* como cualquier disparidad entre los significados atribuidos a una expresión por dos sujetos (personas o instituciones) (Godino, 2002).

Un desajuste entre la interpretación del estudiante y del profesor corresponde a un *conflicto interaccional*, una disparidad en las prácticas de instituciones diferentes a un *conflicto epistémico* y un desajuste interno en el sujeto a un *conflicto cognitivo* (Godino, Rivas, Castro y Konic, 2008). Finalmente, uno de los propósitos de los estudios de evaluación será identificar los posibles conflictos semióticos de los estudiantes.

#### **1.4. OBJETIVOS E HIPÓTESIS DE LA INVESTIGACIÓN**

En las líneas anteriores se ha destacado el creciente interés que la enseñanza de la estadística ha experimentado. En particular, se ha subrayado la importancia de que se comprendan las ideas fundamentales sobre las que se fundamenta la inferencia, que son las relacionadas con el muestreo. En este sentido, Harradine, Batanero y Rossman (2011) indican que es importante saber relacionar las características de las muestras con las de la población, que es el principal fin de la estadística. Estos autores sugieren que es posible una comprensión informal del muestreo desde la secundaria, donde los estudiantes debieran comenzar a diferenciar entre población y muestra, y comprender ideas como la representatividad y la variabilidad de las muestras, así como el efecto sobre esta variabilidad según el tamaño de la muestra.

En este sentido, el concepto de muestreo adquiere un papel protagonista en el aprendizaje de la estadística y será necesario asegurar su comprensión por parte del alumnado (Batanero y Borovcnik, 2016). Por tanto, el *objetivo general* del trabajo es evaluar algunas ideas asociadas al concepto de muestreo en estudiantes de educación secundaria obligatoria y bachillerato, en particular. Desde este objetivo general se definen una serie de objetivos específicos que tratan de enmarcar el problema de investigación y guiar en la realización de la misma.

##### **1.4.1. OBJETIVOS**

En primer lugar, es necesaria la revisión previa de los documentos curriculares, para estar informados de la enseñanza que han recibido los estudiantes y la que se espera reciban en los cursos posteriores, lo que conduce a la caracterización del siguiente objetivo:

*Objetivo 1. Identificar, describir y comparar los contenidos matemáticos relacionados con el muestreo, presentados en los documentos curriculares a lo largo de la educación secundaria obligatoria y el bachillerato.*

La finalidad es observar cuáles son los contenidos matemáticos asociados al muestreo, contemplados por los documentos curriculares en las citadas etapas educativas y en sus diferentes niveles, así como identificar las orientaciones metodológicas consideradas en su enseñanza. Este análisis previo guía en la identificación de los contenidos que esperamos que hayan formado parte del proceso de enseñanza y aprendizaje de la etapa de secundaria obligatoria y sería nuestro significado institucional pretendido para el muestreo. Nos servirá también de base para elegir ítems de evaluación que sean asequibles a los estudiantes.

Cuando se comienza una investigación es necesario conocer qué se ha investigado con anterioridad, para no repetir el estudio y tratar de aportar nueva información. Además, el estudio de dichas investigaciones previas contribuye a enriquecer la discusión de los resultados del trabajo al relacionarlos con los anteriores. El análisis planteado en el Capítulo 2 resume los resultados que se han obtenido en investigaciones previas en relación a los contenidos matemáticos que evaluamos. De ello se deriva el siguiente objetivo:

*Objetivo 2. Realizar una síntesis de las investigaciones previas, que nos proporcione una visión de los problemas abordados en las mismas, así como ejemplos de ítems de evaluación y respuestas de estudiantes y permita enmarcar nuestro estudio con los que le anteceden.*

Este estudio nos permite identificar el tipo de tareas que han formado parte de la investigación y las respuestas típicas de los estudiantes a estas tareas. En este sentido, su realización permite mejorar posteriormente la selección de las mismas para el instrumento de evaluación que aplicamos en los estudios empíricos. Además, esta revisión nos permite valorar si nuestro trabajo puede aportar nueva información o complementar con nuevos resultados y compararlos con los de las investigaciones previas.

Además de estos dos objetivos, comunes a toda la investigación, planteamos un tercero que se desglosará en objetivos específicos en cada uno de los estudios descritos en los siguientes capítulos. Este tercer objetivo es el siguiente:

---

*Objetivo 3. Llevar a cabo una serie de estudios de evaluación de la comprensión que muestran los estudiantes de educación secundaria obligatoria y bachillerato de conceptos elementales ligados al muestreo.*

Mediante estos estudios pretendemos profundizar en la comprensión que tienen los estudiantes de estos cursos sobre algunas ideas elementales sobre muestreo. Asimismo, se desea ver la progresión de esta comprensión con el curso escolar y el efecto de ciertas variables de tarea sobre la misma. Los datos obtenidos se analizarán con métodos cuantitativos y cualitativos para extraer la tendencia de los datos y al mismo tiempo encontrar categorías de respuestas típicas. Con todo ello pensamos proporcionar información que pueda utilizarse por los profesores para la evaluación de posibles dificultades y el diseño de la enseñanza del muestreo.

#### **1.4.2. HIPÓTESIS INICIALES**

Aunque nuestro trabajo es de tipo exploratorio, los resultados obtenidos en trabajos anteriores realizados con estudiantes o futuros profesores de educación primaria (ej., Gómez, Batanero y Contreras, 2014), así como otros estudios curriculares (Batanero, Gea, Arteaga y Contreras, 2014), nos llevan a plantear una serie hipótesis. Las entendemos no como hipótesis estadísticas, sino como conjeturas sobre lo que se espera encontrar en las diferentes etapas de la investigación. Son las siguientes:

*H1. Los contenidos curriculares tanto del Decreto de Enseñanzas Mínimas como del Currículo Básico contemplan o reflejan los aspectos esenciales asociados al concepto de muestreo, que son objeto de investigación.*

En particular, se espera que en dichos documentos curriculares se introduzcan ideas sobre población y muestra, estadístico y parámetro, tipos de muestreo y nociones elementales de estimación, representatividad y variabilidad muestral. Igualmente se espera se consideren los tres niveles de distribución citados por Schuyten (1991): distribución de datos, distribución de la variable en la población y distribución muestral del estadístico. Nos apoyamos para elaborar esta hipótesis en nuestro propio conocimiento del currículo y nuestra experiencia de enseñanza.

*H2. Algunos estudiantes de educación secundaria y bachillerato presentan una comprensión insuficiente sobre algunos objetos matemáticos elementales ligados al muestreo.*

Esta hipótesis se basa en el análisis de los antecedentes de la investigación, que se describen en el Capítulo 2, donde se han analizado diferentes sesgos de comprensión referidos a la representatividad y variabilidad muestral, confusión entre estadístico y parámetro y entre la distribución de la población y la distribución muestral. Igualmente se han descrito dificultades con las ideas de independencia de ensayos sucesivos y la comprensión de la convergencia, que pensamos se repetirán en nuestros estudios empíricos.

*H3. Esperamos una mejora gradual de la comprensión en los estudiantes de mayor edad.*

Existe una diferencia de dos o cuatro años en los grupos de estudiantes que formarán parte de los trabajos empíricos, por lo que se espera que se realice la tarea mejor en los cursos superiores, ya que tanto los conocimientos matemáticos como el razonamiento de estos alumnos son más avanzados. Además, como el currículo contempla la enseñanza de ideas básicas sobre muestreo, se espera que los estudiantes mayores hayan tenido oportunidad de recibir enseñanza y aprendizaje de ciertos contenidos estadísticos, que conduzca a una mejor comprensión.

Las hipótesis relativas a los conocimientos del muestreo por parte de los estudiantes se refinarán y harán más específicas en cada uno de los estudios, en los capítulos 3 y 4.

## **1.5. MARCO CURRICULAR**

Este apartado está enfocado en realizar una revisión de los contenidos que sobre muestreo se incorporan en los documentos curriculares oficiales a nivel estatal, así como los que luego se concretan a nivel autonómico.

Como la parte empírica del trabajo se desarrolla en la comunidad autónoma de Aragón, la revisión de los documentos autonómicos está centrada en los documentos curriculares propios de esta comunidad. Además, este análisis permite realizar un estudio complementario, donde se observen las discrepancias o puntos en común que se reflejan en ambos currículos.

El foco de interés del análisis de los documentos curriculares se corresponde con la identificación de los contenidos relacionados con el muestreo durante la etapa de la educación secundaria obligatoria y bachillerato. Parte de este trabajo y otros complementarios se han publicado en Begué (2019), Begué y Batanero (2017), Begué, Ruiz, Gea y Batanero (2017) y Ruiz, Begué, Batanero y Contreras (2017).

Finalmente, queremos señalar que se realiza el estudio a lo largo de toda la etapa educativa considerada en la investigación, con el objetivo de observar la evolución que experimentan los contenidos correspondientes al bloque de contenidos asociado a este trabajo.

### **1.5.1. EL MUESTREO EN LOS DECRETOS DE ENSEÑANZAS MÍNIMAS**

#### **La enseñanza secundaria**

Este primer apartado del análisis curricular está centrado en el currículo oficial estatal previo (MEC, 2007a), por el que se establece las enseñanzas mínimas correspondientes a la etapa de educación secundaria obligatoria. El citado currículo organiza los contenidos de la materia de matemáticas en bloques de contenidos definidos para cada curso. En concreto, los aspectos asociados a la estadística quedan descritos bajo el bloque denominado: “Bloque 6. Estadística y probabilidad”. En particular, la información asociada al concepto de muestreo queda recogida en la Tabla 1.5.1.

Como vemos en la Tabla 1.5.1, el currículo estatal citado contempla explícitamente la idea de muestra, su necesidad y los métodos de muestreo, desde el tercer curso de la etapa educativa, así como el análisis de la representatividad en el muestreo en el cuarto curso. En relación con la probabilidad, a partir de tercero se incluye como contenido el cálculo de probabilidades mediante la regla de Laplace y estimación de las mismas mediante simulación y experimentación. Así, se introduce la probabilidad desde su significado clásico, además de su enfoque frecuencial, que se refuerza con la simulación como técnica para el cálculo de probabilidades.

En el último curso de esta etapa educativa se describen dos vías posibles para el estudiante. Por un lado, se considera la opción A, la cual está enfocada para aquellos estudiantes que quieren realizar una formación profesional con posterioridad, mientras que la opción B está dirigida a los que pretendan continuar con los estudios de bachillerato. La diferencia radica en una ampliación de los contenidos de matemáticas y algo más de formalización para esta segunda modalidad citada, que quedan identificados en la Tabla 1.5.1.

En relación con lo recogido en el Boletín Oficial de Aragón (BOA), elaborado desde el Departamento de Educación, Cultura y Deporte (2007), se contempla cada uno de los contenidos señalados anteriormente (Tabla 1.5.1), aunque el currículo

autonómico amplía dichos contenidos. En la Tabla 1.5.2 se muestra aquellos puntos que son añadidos.

Tabla 1.5.1. Contenidos de muestreo en el Decreto de Enseñanzas Mínimas

Curso	
1º	<p>Formulación de conjeturas sobre el comportamiento de fenómenos aleatorios sencillos y diseño de experiencias para su comprobación.</p> <p>Reconocimiento y valoración de las matemáticas para interpretar y describir situaciones inciertas.</p> <p>Diferentes formas de recogida de información. Organización en tablas de datos recogidos en una experiencia. Frecuencias absolutas y relativas (MEC, 2007a, p.753).</p>
2º	<p>Diferentes formas de recogida de información. Organización de los datos en tablas. Frecuencias absolutas y relativas, ordinarias y acumuladas.</p> <p>Medidas de centralización: media, mediana y moda. Significado, estimación y cálculo. Utilización de las propiedades de la media para resolver problemas. Utilización de la media, la mediana y la moda para realizar comparaciones y valoraciones.</p> <p>Utilización de la hoja de cálculo para organizar los datos, realizar los cálculos y generar los gráficos más adecuados. (MEC, 2007a, p.755).</p>
3º	<p>Necesidad, conveniencia y representatividad de una muestra. Métodos de selección aleatoria y aplicaciones en situaciones reales.</p> <p>Atributos y variables discretas y continuas. Agrupación de datos en intervalos.</p> <p>Media, moda, cuartiles y mediana. Significado, cálculo y aplicaciones.</p> <p>Análisis de la dispersión: rango y desviación típica. Interpretación conjunta de la media y la desviación típica. Utilización de las medidas de centralización y dispersión para realizar comparaciones y valoraciones. Actitud crítica ante la información de índole estadística.</p> <p>Utilización de la calculadora y la hoja de cálculo para organizar los datos, realizar cálculos y generar las gráficas más adecuadas.</p> <p>Experiencias aleatorias. Sucesos y espacio muestral. Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con el azar.</p> <p>Formulación y comprobación de conjeturas sobre el comportamiento de fenómenos aleatorios sencillos. Cálculo de la probabilidad mediante la simulación o experimentación.</p> <p>Utilización de la probabilidad para tomar decisiones fundamentadas en diferentes contextos. Reconocimiento y valoración de las matemáticas para interpretar, describir y predecir situaciones inciertas (MEC, 2007a, p.756).</p>
4ºA	<p>Identificación de las fases y tareas de un estudio estadístico a partir de situaciones concretas cercanas al alumnado. Análisis elemental de la representatividad de las muestras estadísticas.</p> <p>Utilización de las medidas de centralización y dispersión para realizar comparaciones y valoraciones.</p> <p>Experiencias compuestas. Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con el azar. (MEC, 2007a, p.758).</p>
4º B (se añade)	<p>Representatividad de una distribución por su media y desviación típica o por otras medidas ante la presencia de descentralizaciones, asimetrías y valores atípicos.</p> <p>Valoración de la mejor representatividad en función de la existencia o no de valores atípicos (MEC, 2007a, p.759).</p>



En términos genéricos, la lectura de las directrices recogidas en el currículo autonómico refleja un mayor nivel de especificidad en relación a los contenidos y destrezas que tienen que trabajarse bajo este bloque a lo largo de la etapa educativa.

En relación con la información resumida en la Tabla 1.5.2, identificamos la inclusión de contenidos de muestreo. Por un lado, se espera que el estudiante de tercer curso detecte la imprevisibilidad y regularidad, que constituyen una característica inherente a la probabilidad enfocada desde el significado frecuencial. Por otro lado, el currículo autonómico, en el último curso, considera la utilización de un lenguaje adecuado para describir y cuantificar informaciones asociadas a un estudio estadístico.

Tabla 1.5.2. Contenidos relacionados con el muestreo en el Decreto de Enseñanzas Mínimas de la Comunidad de Aragón

Curso	
1º	Población y muestra. Características cualitativas y cuantitativas de una población. Distribuciones discretas. (Departamento de Educación, Cultura y Deporte, 2007, p.237).
3º	Objetivos, elementos y fases de un estudio estadístico. Estadística unidimensional. Análisis de las características de una población a partir de tablas y gráficos estadísticos. Imprevisibilidad y regularidad. Frecuencia relativa y probabilidad de un suceso: estabilidad de las frecuencias (Departamento de Educación, Cultura y Deporte, 2007, pp.247-248).
4ºA 4ºB	VARIABLES discretas y continuas. Agrupación de datos: intervalos y marcas de clase. Tablas de frecuencias absolutas y relativas de datos agrupados correspondientes a una variable continua. Parámetros de centralización y de dispersión de una distribución de datos agrupados. Obtención, organización, representación e interpretación de información relevante referida a un estudio sencillo de una población. Uso de la hoja de cálculo y de la calculadora científica para obtener parámetros estadísticos correspondientes a distribuciones de datos agrupados. Experimentos aleatorios y sucesos. Probabilidad condicionada. Probabilidad total. Probabilidad estadística. Simulación. Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con estudios estadísticos de poblaciones y con el azar (Departamento de Educación, Cultura y Deporte, 2007, pp. 252-253 y 257-258).

Batanero, Gea, Arteaga y Contreras (2014) analizan los contenidos y la metodología sugeridos en las directrices curriculares españolas correspondientes a la enseñanza obligatoria. Su análisis revela que en la educación secundaria obligatoria se sigue contemplando la organización de los datos y gráficos, iniciada en la educación primaria. En el primer curso de la etapa de secundaria se amplían los contenidos, al considerar los conceptos de frecuencias absoluta y relativa, así como nociones de muestreo. Además, en este primer curso se incluye el diseño de experiencias para el

*Planteamiento de la investigación*

análisis de conjeturas relacionadas con un fenómeno aleatorio. Por tanto, se puede deducir que la idea que subyace es introducir al estudiante el significado frecuencial de la probabilidad, adoptando un proceso gradual. Puesto que en dicho enfoque subyace la idea de muestreo, entendemos que existe un primer acercamiento intuitivo al muestreo.

Tabla 1.5.3. Objetos matemáticos relacionados con el muestreo en los Decretos de Enseñanzas Mínimas

	1º	2º	3º	4ºA	4º*B
<b>Situaciones-problema</b>					
Estimar características de una población a partir de muestras		x	x	x	x
Valorar la representatividad de muestras				x	x
<b>Lenguaje matemático</b>					
Interpretar con un vocabulario adecuado la información estadística	x	x	x	x	x
Diagrama de barras, sectores, líneas	x	x	x	x	x
Histograma, polígono de frecuencias			x	x	x
Diagrama de cajas				x	x
<b>Conceptos</b>					
Población	x	x	x	x	x
Individuo					
Muestra	x		x	x	x
Espacio muestral, suceso			x	x	x
Variables estadística discreta. Variable estadística continua			x	x	x
Frecuencia absoluta, relativa	x	x	x	x	x
Frecuencia ordinaria y acumulada		x			
Distribución de datos, distribución unidimensional	x	x	x	x	x
Variabilidad muestral				x	x
Representatividad muestral			x	x	x
Medidas de posición: media, mediana, moda		x	x	x	x
Cuantiles			x	x	x
Medidas de dispersión: rango		x	x	x	x
D. típica			x	x	x
Asimetría					x
Valores atípicos					x
<b>Procedimientos</b>					
Simulación y/o experimentación	x		x	x	x
Métodos de selección de una muestra estadística			x	x	x
Diagrama de árbol (espacio muestral)				x	x
Representar la distribución de frecuencias en tablas o gráficas	x	x	x	x	x
Relacionar las características de una población y una muestra		x	x	x	x
Calcular las medidas de centralización		x	x	x	x
Calcular las medidas de dispersión		x	x	x	x
Recoger datos	x	x	x	x	x
Utilizar herramientas tecnológicas		x	x	x	x
Fases y tareas de un estudio estadístico		x	x	x	x
<b>Propiedades</b>					
Aproximación de la frecuencia relativa a la probabilidad mediante la simulación o experimentación	x		x		
Propiedades de las medidas de posición y dispersión		x			x

Con respecto a las orientaciones metodológicas, las directrices curriculares a nivel estatal y autonómico remarcan que la introducción de los contenidos correspondientes al bloque de estadística y probabilidad debe seguir un proceso gradual. También se recomienda que los primeros cursos supongan una aproximación al estudio de fenómenos aleatorios sencillos mediante la experimentación y el tratamiento de datos estadísticos por medio de tablas y gráficas.

Las orientaciones correspondientes a la comunidad de Aragón resaltan que el objetivo principal de la enseñanza de la estadística descriptiva no es el desarrollo de destrezas de cálculo o de la habilidad para elaborar gráficos estadísticos, sino que está enfocada en el análisis cualitativo de las características de la población y el fomento de la actitud crítica. En consecuencia, se concede mayor importancia a la interpretación de los datos y los resultados que a la memorización y el cálculo.

El currículo a nivel autonómico sugiere que los datos se obtengan del entorno del estudiante y contempla algunas de las ideas fundamentales descritas por Burril y Biehler (2011). En concreto, se expresa la importancia que adquieren los datos en el estudio estadístico, subrayando la importancia del contexto (Moore, 1991). Las directrices curriculares autonómicas indican la necesidad de introducir al estudiante en conceptos matemáticos asociados a la inferencia estadística desde un enfoque informal, basado en la simulación:

Además del enfoque descriptivo, en la estadística también tiene cabida el enfoque inductivo, que tiene por objetivo el conocimiento de las características de una población a partir de las observaciones realizadas sobre muestras de ella. El instrumento para este punto de vista de la estadística es la teoría de probabilidades y el estudio de las distribuciones teóricas de probabilidad, y queda fuera del alcance del nivel de la Educación secundaria obligatoria. No obstante, puede hacerse un tratamiento informal de la relación entre características de una población y de una muestra y una aproximación experimental a la idea de representatividad. (Departamento de Educación, Cultura y Deporte, 2007, p.264).

Un resultado de nuestro análisis es la Tabla 1.5.3, en que identificamos la configuración semiótica de objetos matemáticos ligados al muestreo y contenidos en los diferentes cursos según las directrices curriculares de Decretos de Enseñanzas Mínimas. Hemos incluido también los temas de estadística descriptiva que los estudiantes precisan para el trabajo con los datos de la muestra.

## **Bachillerato**

A continuación, se lleva a cabo un análisis similar del currículo con respecto a la etapa no obligatoria de la secundaria (resumido en Batanero, Begué y Gea, 2019). En este caso, los documentos se corresponden con el Real Decreto 1467/2007, de 2 de noviembre (MEC, 2007b). Los contenidos matemáticos correspondientes al bloque de estadística y probabilidad se presentan en la Tabla 1.5.4, donde dichos contenidos se organizan según el curso y la modalidad de Bachillerato. En concreto, Matemáticas I y II se localizan dentro de la modalidad denominada Ciencias y Tecnología.

El primer curso está dirigido en torno a dos grandes bloques de contenidos, los cuales quedan expuestos en el currículo autonómico: (1) Estadística descriptiva y (2) Distribuciones de probabilidad (Departamento de Educación Cultura y Deporte, 2008, p.14075 y pp.14136-14137). Por otro lado, destacar que se considera el estudio del Teorema de la probabilidad total y de Bayes como contenido de la materia Matemáticas I, luego se amplía los contenidos establecidos en el Decreto de Enseñanzas Mínimas (MEC, 2007b). Incluimos estos contenidos de probabilidad porque son necesarios para el trabajo con el muestreo y la distribución muestral, así como para la estimación.

En cuanto al segundo curso de Bachillerato, el bloque de “Probabilidad y estadística” está presente en la modalidad de Matemáticas Aplicadas a las Ciencias Sociales, donde destaca la introducción a la inferencia estadística. En este caso, es interesante considerar las directrices del currículo autonómico en relación a la enseñanza de este contenido matemático:

El trabajo central en la inferencia estadística consiste en conseguir que los estudiantes conecten los significados de tamaño de la muestra, nivel de confianza y error admisible. Estos conceptos deben aparecer en un proceso de resolución de problemas encuadrados en situaciones familiares, para así reforzar su comprensión. (Departamento de Educación Cultura y Deporte, 2008, p.14140).

El discurso seguido apoya la perspectiva que se viene desarrollando desde la investigación en educación estadística, que subraya la necesidad de que el estudiante tenga una comprensión de los conceptos, así como que se planteen situaciones familiares al estudiante para favorecer dicho aprendizaje.

El análisis de los documentos curriculares autonómicos se corresponde con una descripción más detallada de los contenidos, sin mostrar diferencias significativas con el descrito en el currículo nacional. Por otro lado, se identifica desde la lectura del documento oficial autonómico la necesidad de que el estudiante se centre en la comprensión de los conceptos más que en el cálculo, el cual debería estar apoyado en

recursos tecnológicos. De esta manera, el estudiante puede centrarse en el significado de los conceptos (Departamento de Educación Cultura y Deporte, 2008, p.14075 y pp.14136-14137).

Tabla 1.5.4. Contenidos de muestreo en el Decreto de Enseñanzas Mínimas

Curso	Contenido
Matemáticas I	Estudio de la probabilidad compuesta, condicionada, total y a posteriori. Distribuciones binomial y normal como herramienta para asignar probabilidades a sucesos. (MEC, 2007b, p.45449).
Matemáticas Aplicadas a las Ciencias Sociales I	Estadística descriptiva unidimensional. Tipos de variables. Métodos estadísticos. Tablas y gráficos. Parámetros estadísticos de localización, de dispersión y de posición. Asignación de probabilidades a sucesos. Distribuciones de probabilidad binomial y normal. (MEC, 2007b, p. 45475).
Matemáticas Aplicadas a las Ciencias Sociales II	Profundización en los conceptos de probabilidades a priori y a posteriori, probabilidad compuesta, condicionada y total. Teorema de Bayes. Implicaciones prácticas de los teoremas: Central del límite, de aproximación de la Binomial a la Normal y Ley de los Grandes Números. Problemas relacionados con la elección de las muestras. Condiciones de representatividad. Parámetros de una población. Distribuciones de probabilidad de las medias y proporciones muestrales. Intervalo de confianza para el parámetro p de una distribución binomial y para la media de una distribución normal de desviación típica conocida. Contraste de hipótesis para la proporción de una distribución binomial y para la media o diferencias de medias de distribuciones normales con desviación típica conocida. (MEC, 2007b, p. 45476)

Para completar el análisis, en la Tabla 1.5.5 se resumen los objetos matemáticos específicos que se recogen en cada curso de bachillerato en los Decretos de Enseñanzas Mínimas, siguiendo los descritos en el marco teórico del enfoque ontosemiótico.

Observamos que el contenido sobre muestreo que se concentra en el segundo curso de bachillerato de ciencias sociales constituye prácticamente un curso completo de estadística en sí mismo, similar a los impartidos en primeros cursos de universidad. La gran cantidad de objetos matemáticos implicados en su estudio, que en la Tabla 1.5.5 se presentan únicamente en forma resumida, hace inviable una enseñanza profunda del tema. Más aún, teniendo en cuenta que estos estudiantes tienen menor preparación matemática que sus compañeros del bachillerato de Ciencias y Tecnología, por lo que es complejo favorecer la comprensión de todos estos contenidos, si no se tiene el conocimiento suficiente para la aplicación adecuada de un software estadístico al resolver los problemas.

El problema se agrava por el hecho de que en las pruebas recientes de acceso a la universidad han incluido en los últimos 15 años un problema de inferencia entre los

cuatro que constituyen la prueba (López-Martín, Batanero, Díaz-Batanero y Gea, 2016). En consecuencia, con frecuencia los profesores tratan únicamente de enseñar los procedimientos necesarios para resolver estos problemas, al no tener tiempo suficiente para fomentar su comprensión.

Tabla 1.5.5 .Objetos matemáticos relacionados con el muestreo en los Decretos de Enseñanzas Mínimas

	M1	MS1	MS2
<b>Situaciones-problema</b>			
Selección de una muestra			x
Valorar la representatividad de muestras			x
Encontrar la distribución muestral de un estadístico			x
Determinar intervalos de confianza para un parámetro			x
Contrastar hipótesis sobre un parámetro			x
<b>Lenguaje matemático</b>			
Interpretar con un vocabulario adecuado la información estadística	x	x	x
Curvas de densidad	x	x	x
Tablas de la distribución	x	x	x
Representación de distribuciones muestrales			x
Representación de intervalos de confianza, regiones crítica y de rechazo			x
<b>Conceptos</b>			
Población	x	x	x
Individuo			x
Muestra			x
Variables aleatoria discreta. Variable aleatoria continua	x	x	x
Distribución de probabilidad, parámetro	x	x	x
Función de distribución. Función de densidad	x	x	x
Variabilidad muestral			x
Representatividad muestral			x
Teorema Central del Limite			x
Distribución muestral			x
Intervalo de confianza, coeficiente de confianza			x
Contraste de hipótesis, hipótesis nula y alternativa			x
Estadístico de contraste, regiones de aceptación y rechazo			x
Nivel de significación, p-valor			x
<b>Procedimientos</b>			
Simulación y/o experimentación	x	x	x
Métodos de selección de una muestra estadística			x
Diagrama de árbol (espacio muestral)		x	x
Relacionar las características de una población y una muestra			x
Calcular estadísticos a partir de las observaciones de la variable o datos	x	x	x
Tipificar una distribución normal; calcular probabilidades con la N (0,1)	x	x	x
Deducir la distribución muestral, aplicando el TCL			x
Calcular intervalos de confianza			x
Utilizar herramientas tecnológicas	x	x	x
Fases y tareas de un estudio estadístico			x
<b>Propiedades</b>			
Propiedades de las distribuciones binomial y normal	x	x	x
Teorema Central del Límite			x
Propiedades de las distribuciones muestrales			x
Propiedades del intervalo de confianza			x
Propiedades de los conceptos asociados al contraste de hipótesis			x

### 1.5.2. EL MUESTREO EN EL CURRÍCULO BÁSICO

En este apartado se realiza un análisis similar al caso anterior con respecto a la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE) que se recoge en el Currículo Básico (MECD, 2015), la cual está vigente en la actualidad. En particular, en la Tabla 1.5.6 se exponen los contenidos de muestreo descritos por el currículo para el “Bloque 5. Estadística y probabilidad”. Una de las novedades que introduce la LOMCE son las dos vías que se inician en el tercer curso: “matemáticas orientadas a las enseñanzas académicas” y “matemáticas orientadas a las enseñanzas aplicadas”. Esta diferenciación radica en que el enfoque con el cual se presenta la segunda tiene que responder a un interés por mostrar la aplicación práctica de los contenidos matemáticos en situaciones reales frente a la profundización en los aspectos teóricos, que sería el enfoque pensado para las enseñanzas académicas (MECD, 2015, p.339).

Tabla 1.5.6. Contenidos relacionados con el muestreo en el Currículo Básico

Curso	Contenido
1º y 2º ESO	Población e individuo. Muestra. Variables estadísticas. Variables cualitativas y cuantitativas. Frecuencias absolutas y relativas. Organización en tablas de datos recogidos en una experiencia. Medidas de tendencia central. Medidas de dispersión. Fenómenos deterministas y aleatorios. Formulación de conjeturas sobre el comportamiento de fenómenos aleatorios sencillos y diseño de experiencias para su comprobación. Espacio muestral en experimentos sencillos. Frecuencia relativa de un suceso y su aproximación a la probabilidad mediante la simulación o experimentación. Sucesos elementales equiprobables y no equiprobables. Espacio muestral en experimentos sencillos. (MECD, 2015, p.413).
3º curso Matemáticas orientadas a las enseñanzas aplicadas	Fases y tareas de un estudio estadístico. Población, muestra. Variables estadísticas: cualitativas, discretas y continuas. Métodos de selección de una muestra estadística. Representatividad de una muestra. Frecuencias absolutas, relativas y acumuladas. Agrupación de datos en intervalos. Parámetros de posición: media, moda, mediana y cuartiles. Cálculo, interpretación y propiedades. Parámetros de dispersión: rango, recorrido intercuartílico y desviación típica. Cálculo e interpretación. Interpretación conjunta de la media y la desviación típica (MECD, 2015, p.403).
3º curso (se añade) Matemáticas orientadas a las enseñanzas académicas	Experiencias aleatorias. Sucesos y espacio muestral. Utilización de la probabilidad para tomar decisiones fundamentadas en diferentes contextos (MECD, 2015, p.394).
4º curso Matemáticas orientadas a las enseñanzas aplicadas	Interpretación, análisis y utilidad de las medidas de centralización y dispersión. Comparación de distribuciones mediante el uso conjunto de medidas de posición y dispersión. Azar y probabilidad. Frecuencia de un suceso aleatorio. Probabilidad simple y compuesta. Sucesos dependientes e independientes (MECD, 2015, p.407).
4º curso (se añade) Matemáticas orientadas a las enseñanzas académicas	Experiencias aleatorias compuestas. Probabilidad condicionada. Utilización del vocabulario adecuado para describir y cuantificar situaciones relacionadas con el azar y la estadística. Identificación de las fases y tareas de un estudio estadístico. (MECD, 2015, p.398).

Tabla 1.5.7. Objetos matemáticos en el currículo básico

Objeto matemático	1º	2º	3ºAC	3ºAP	4ºAC	4ºAP
<b>Situaciones-problema</b>						
Estimar la característica de una población a partir de muestras de una población	x	x	x	x	x	x
Valorar la representatividad de dos muestras					x	x
<b>Lenguaje</b>						
Interpretar con un vocabulario adecuado la información estadística	x	x	x	x	x	x
Diagrama de barras	x	x	x	x	x	x
Diagrama de sectores y de líneas	x	x	x	x	x	
Histograma					x	x
Diagrama de cajas y bigotes			x	x	x	
Diagrama de dispersión					x	x
<b>Conceptos</b>						
Población, muestra	x	x	x	x	x	x
Individuo	x	x				
Espacio muestral	x	x	x		x	x
Variable estadísticas (cualitativa y cuantitativa)	x	x	x	x	x	x
Variables discretas y continuas.			x	x	x	x
Frecuencia relativa, absoluta	x	x	x	x		
Frecuencia acumulada			x	x		
Distribución de datos, distribución unidimensional	x	x	x	x	x	x
Representatividad muestral			x	x	x	x
Medidas de centralización, media, mediana, moda	x	x	x	x	x	x
Cuartiles			x	x	x	x
Medidas de dispersión. Rango		x	x	x	x	x
Recorrido intercuartílico. D. típica			x	x	x	x
<b>Procedimiento</b>						
Simulación y/o experimentación	x	x			x	x
Métodos de selección de una muestra estadística			x	x	x	
Representar la distribución de frecuencias mediante tablas o gráficas.	x	x	x	x	x	x
Relacionar características en población y muestra	x	x	x	x	x	x
Calcular las medidas de centralización	x	x	x	x	x	x
Calcular las medidas de dispersión		x	x	x	x	x
Diagrama de árbol (espacio muestral)	x	x	x		x	x
Recoger datos	x	x			x	x
Utilizar herramientas tecnológicas	x	x	x	x	x	x
Fases y tareas de un estudio estadístico					x	
<b>Propiedades</b>						
Aproximación de la frecuencia relativa a la probabilidad mediante la simulación o experimentación		x	x			
Propiedades de las medidas de posición			x	x	x	x
Propiedades de las medidas de dispersión					x	x

En particular, el currículo indica que se debe ejercitar el cálculo de probabilidades aplicando la Regla de Laplace (MECD, 2015). De modo que el estudiante que continúe la vía orientada a las enseñanzas académicas aprenda desde el tercer curso conceptos matemáticos del análisis combinatorio, necesarios para el enfoque clásico.



En comparación con el currículo oficial correspondiente a la LOE, el nuevo currículo incluye la simulación desde el primer curso de esta etapa educativa, reconociendo implícitamente su interés en el aprendizaje del enfoque frecuencial de la probabilidad (Huerta, 2015). Además, en el último curso, se contempla la introducción a la correlación, independientemente de la modalidad de enseñanza. Por otro lado, la introducción de la probabilidad desde el significado clásico se adelanta, iniciándose en los primeros cursos de la ESO.

El análisis de los documentos curriculares autonómicos se corresponde con una descripción más detallada de los contenidos, sin mostrar diferencias significativas con el descrito en el currículo nacional (Departamento de Educación Cultura y Deporte, 2016a).

Estos contenidos se desglosan en los objetos matemáticos que se presentan en la Tabla 1.5.7, que de nuevo muestra la gran complejidad del tema y explica los errores que persisten al finalizar la enseñanza, puesto que el tiempo disponible para ella es escaso.

### **Bachillerato**

La Tabla 1.5.8 presenta los contenidos relacionados con el muestreo en la estadística y la probabilidad que considera la LOMCE, a nivel estatal. La LOMCE incluye por primera vez un bloque de contenidos asociados a la estadística y probabilidad en la materia de Matemáticas II, que se cursa dentro de la modalidad denominada Ciencias.

El análisis conjunto de los contenidos que se consideran para la etapa secundaria no obligatoria revela una mayor discrepancia en el segundo curso. Como se observa desde la lectura de la Tabla 1.5.8, el contenido para el primer curso es similar. No obstante, las diferencias son notables si consideramos el contenido correspondiente al segundo curso. En particular, los contenidos descritos para el segundo curso de la opción científica tecnológica se sitúan más próximos al ámbito del cálculo de probabilidades, donde el modelo que se estudia es un modelo aleatorio, para el cual se define una distribución asociada a unos parámetros estadísticos.

Por tanto, su estudio supone un puente entre la estadística y la probabilidad, que exige que el estudiante maneje con soltura conceptos propios de la estadística, subrayando que la dificultad o el problema que se le plantea al estadístico es el desconocimiento del modelo probabilístico que rige el fenómeno aleatorio. Es decir, si

analizamos el fenómeno aleatorio caracterizado por lanzar una moneda se sabe que dicho modelo probabilístico se corresponde con una distribución binomial, donde la probabilidad de que salga cara es conocida desde el significado clásico de la probabilidad, siendo ésta  $\frac{1}{2}$ , y el número de monedas es un parámetro a determinar. Sin embargo, la complejidad aumenta cuando analizamos fenómenos en los que determinar la efectividad de un fármaco que se pretende introducir en el mercado, donde se desconoce el modelo probabilístico que rige dicha situación.

Tabla 1.5.8. El muestreo en el Currículo Básico para la etapa del Bachillerato

Curso	Contenido
Matemáticas Aplicadas a las Ciencias Sociales I	<p>Sucesos. Asignación de probabilidades a sucesos mediante la regla de Laplace y a partir de su frecuencia relativa.</p> <p>Aplicación de la combinatoria al cálculo de probabilidades. Experimentos simples y compuestos. Probabilidad condicionada. Dependencia e independencia de sucesos. Variables aleatorias discretas. Distribución de probabilidad. Media, varianza y desviación típica. Distribución binomial. Caracterización e identificación del modelo. Cálculo de probabilidades. Variables aleatorias continuas.</p> <p>Interpretación de la media, varianza y desviación típica. Distribución normal. Cálculo de probabilidades mediante la aproximación de la distribución binomial por la normal. (MECD, 2015, p.385)</p>
Matemáticas Aplicadas a las Ciencias Sociales II	<p>Profundización en la Teoría de la Probabilidad. Asignación de probabilidades a sucesos mediante la regla de Laplace y a partir de su frecuencia relativa. Experimentos simples y compuestos. Probabilidad condicionada. Dependencia e independencia de sucesos.</p> <p>Población y muestra. Métodos de selección de una muestra. Tamaño y representatividad de una muestra. Estadística paramétrica. Parámetros de una población y estadísticos obtenidos a partir de una muestra. Estimación puntual. Media y desviación típica de la media muestral y de la proporción muestral. Distribución de la media muestral en una población normal. Distribución de la media muestral y de la proporción muestral en el caso de muestras grandes.</p> <p>Estimación por intervalos de confianza. Relación entre confianza, error y tamaño muestral. Intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida.</p> <p>Intervalo de confianza para la media poblacional de una distribución de modelo desconocido y para la proporción en el caso de muestras grandes. (MECD, 2015, p.389)</p>
Matemáticas II	<p>Sucesos. Asignación de probabilidades a sucesos mediante la regla de Laplace y a partir de su frecuencia relativa. Aplicación de la combinatoria al cálculo de probabilidades. Experimentos simples y compuestos. Probabilidad condicionada. Dependencia e independencia de sucesos. Variables aleatorias discretas. Distribución de probabilidad. Media, varianza y desviación típica.</p> <p>Distribución binomial. Caracterización e identificación del modelo. Cálculo de probabilidades. Distribución normal. Asignación de probabilidades en una distribución normal. Cálculo de probabilidades mediante la aproximación de la distribución binomial por la normal. (MECD, 2015, p. 422)</p>

Tabla 1.5.9. Objetos matemáticos en el Currículo Básico

	M1	M2	MS1	MS2
<b>Situaciones-problema</b>				
Selección de una muestra				X
Valorar la representatividad de muestras				X
Encontrar la distribución muestral de un estadístico				X
Determinar intervalos de confianza para un parámetro				X
<b>Lenguaje matemático</b>				
Interpretar con un vocabulario adecuado la información estadística	X	X	X	X
Curvas de densidad y función de probabilidad	X	X	X	X
Tablas de la distribución	X	X	X	X
Representación de distribuciones muestrales				X
Representación de intervalos de confianza,				X
Representación de regiones crítica y de rechazo				X
<b>Conceptos</b>				
Población. Individuo. Muestra				X
Variables aleatoria discreta y continua	X	X	X	X
Distribución de probabilidad, parámetro	X	X	X	X
Función de distribución	X	X	X	X
Función de densidad	X	X	X	X
Variabilidad muestral				X
Representatividad muestral				X
Teorema Central del Limite				X
Distribución muestral				X
Intervalo de confianza, coeficiente de confianza				X
<b>Procedimientos</b>				
Simulación y/o experimentación		X	X	X
Métodos de selección de una muestra estadística			X	X
Relacionar las características de una población y una muestra				X
Calcular estadísticos a partir de observaciones				X
Tipificar una distribución normal; calcular probabilidades $N(0,1)$		X	X	X
Deducir la distribución muestral, aplicando el TCL		X	X	X
Calcular intervalos de confianza				X
Utilizar herramientas tecnológicas	X	X	X	X
Fases y tareas de un estudio estadístico			X	X
<b>Propiedades</b>				
Propiedades de las distribuciones binomial y normal	X	X	X	X
Teorema Central del Límite		X	X	X
Propiedades de las distribuciones muestrales				X
Propiedades del intervalo de confianza				X

En estos casos, requerimos de la inferencia estadística y de la selección de muestras asociadas al fenómeno para dar una respuesta con nivel de incertidumbre o error controlado o conocido. En concreto, este tipo de tareas son las que se enmarcan bajo el contenido que se define para el segundo curso del bachillerato de la modalidad de matemáticas aplicadas a las ciencias sociales. Esta complejidad se observa igualmente en los objetos implícitos en este currículo, presentados en la Tabla 1.5.9. Observamos en esta tabla que se han eliminado los temas referidos al contraste de hipótesis, y por

tanto, ha disminuido algo la complejidad del tema, en relación con el currículo anterior. En síntesis, se observa que el contenido de estadística tiene una presencia mayor en la modalidad de bachillerato en ciencias sociales, que culmina en el último curso con la introducción a la inferencia estadística. De hecho, se incluye como contenido matemático a enseñar conceptos propios, tales como intervalo de confianza, error de muestreo y otros que están basados en la comprensión del muestreo. El currículo que se define para la ley de educación en la comunidad de Aragón (Departamento de Educación Cultura y Deporte, 2016b) tiene contenidos similares.

### **1.5.3. ORIENTACIONES CURRICULARES INTERNACIONALES**

Tras el análisis anterior, se describen a continuación las directrices curriculares recogidas en CCSSI (2010) y el proyecto GAISE (Franklin et al., 2007). Estas directrices se comparan con las orientaciones españolas actualmente vigentes del Currículo Básico español, que también se implementan obligatoriamente en todo el territorio español (MECD, 2015). Además, comparamos con los estándares americanos NCTM (2000) y CCSSI (2010) y el proyecto GAISE (Franklin et al., 2007), ya que estos documentos son utilizado como referentes internacionales para la educación estadística.

El resultado del análisis de dichos documentos se recoge tanto en la Tabla 1.5.10 como en la Tabla 1.5.11, donde se presentan los diferentes objetos matemáticos considerando las categorías que define el marco teórico considerado. La variedad de objetos citados en dichos documentos muestra la complejidad del contenido de los currículos.

El análisis comparativo de las tablas no señala diferencias significativas, por lo que se puede afirmar que los contenidos seleccionados para los estudiantes que se sitúan en la etapa secundaria es similar. En resumen, existe cierta uniformidad en relación al contenido que se considera esencial para un estudiante que se forme en el estudio de la estadística. Las líneas siguientes se centran en cada uno de los documentos considerados en este apartado. En particular, se describe el enfoque que sugieren para la enseñanza de la estadística.

Tabla 1.5.10. Situaciones-problema, lenguaje y argumentos del muestreo en diferentes currículos

Objeto matemático	Currículo español	NCTM	CCSSI	GAISE
<b>Situaciones-problema</b>				
Estimar algunas características de una población desconocida	x	x	x	x
Determinar la composición de una población		x		
Determinar el número de muestras de un tamaño dado con y sin reemplazo	x			x
Relacionar la media de una población y las medias muestrales	x			x
Relacionar la proporción de una población y la proporción de una muestra	x		x	x
Identificar muestras aleatorias y no aleatorias	x	x	x	x
Elaboran modelos para el muestreo aleatorio	x			x
Valorar la representatividad de una muestra	x	x	x	x
<b>Lenguaje</b>				
Términos, expresiones, fórmulas y expresiones algebraicas	x	x	x	x
Gráficos	x	x	x	x
Tablas de frecuencia, listados de datos	x	x	x	x
<b>Argumentos</b>				
Verifican, utilizando herramientas tecnológicas, las conjeturas formuladas	x	x	x	x

En primer lugar, NCTM (2000) indica que uno de los objetivos del estándar denominado análisis de datos y probabilidad se corresponde con el desarrollo y evaluación de inferencias y predicciones fundamentadas en los datos. En consecuencia, los estudiantes deberían llegar a comprender los elementos básicos del análisis estadístico: a) seleccionar una muestra adecuada, b) recoger datos de esta muestra, c) describir la muestra y d) hacer inferencias razonables que relacionen la muestra y la población.

En segundo lugar, tanto el CCSSI (2010) como el proyecto GAISE (Franklin et al., 2007) consideran la estadística como una herramienta que describe la variabilidad de los datos y permite realizar decisiones razonables fundamentadas en la información proporcionada por los datos. Un aspecto que se remarca es la importancia que adquiera la forma en que se obtienen los datos para que sea posible la obtención de conclusiones válidas.

Tabla 1.5.11. Conceptos, procedimientos y propiedades del muestreo en diferentes currículos

Objeto matemático	Currículo español	NTCM	CCSSI	GAISE
<b>Conceptos</b>				
Conceptos combinatorios: variaciones, permutaciones y combinaciones	X		X	
Experimento aleatorio	X	X	X	X
Espacio muestral	X	X	X	X
Variable estadística, distribución	X	X	X	X
Muestra (Con y sin reemplazamiento, Aleatoria, sistemática, estratificada, no probabilística)	X	X	X	X
Tamaño de la muestra	X	X	X	X
Población (Elemento, Tamaño, finita o no)	X	X	X	X
Probabilidad (Clásica, Frecuencial)	X	X	X	X
Media Muestral	X	X		X
Sesgo en muestreo	X	X		X
Variabilidad	X	X	X	X
<b>Procedimiento</b>				
Realizar encuestas.	X	X	X	X
Obtener todas las posibles muestras de una población finita; calcular su número	X		X	
Estimar una proporción, media o frecuencia esperada en una muestra conocida la población	X	X		X
Estimar una proporción, media o frecuencia esperada en una población, conocida la muestra	X	X		X
Comparar datos de una muestra con una población	X	X	X	X
Decidir si un método de muestreo es adecuado o diferenciar tipos de muestreo	X	X	X	X
Dados una muestra y el tamaño de la población decidir la composición más probable de la población	X			
Calcular el promedio de todos los promedios de muestras de igual tamaño extraídas desde una población y analizar la relación	X			
Analizar las medias obtenidas de las muestras, cuando tamaño de las muestras aumenta	X			X
Usar la distribución muestral para realizar inferencias informales	X	X	X	X
Representar gráficamente datos	X	X	X	X
Fases de un estudio estadístico	X	X		X
<b>Propiedades</b>				
Representatividad de una muestra	X	X	X	X
Variabilidad de la muestra en función del tamaño	X	X	X	X

En todos los currículos se recomienda el uso de las herramientas tecnológicas para favorecer la interpretación de los valores obtenidos, en lugar de estar más interesados en realizar los cálculos algorítmicos. En conclusión, los estudiantes que finalicen la etapa secundaria deberían ser capaces de valorar de manera razonada la información estadística presente en diferentes medios de comunicación.

---

#### **1.5.4. CONCLUSIONES DEL ESTUDIO CURRICULAR**

En términos generales, se observa que la organización de los contenidos asociados al bloque de estadística y probabilidad de la educación secundaria obligatoria muestra una introducción gradual de los mismos, como se recomienda en la investigación en educación estadística (Batanero y Borovcnik, 2016). Comenzando simplemente con la idea de muestra, se va ampliando los tipos de muestreo y las ideas de representatividad y variabilidad, aunque esta solo en forma implícita. En bachillerato se introduce a los estudiantes de ciencias sociales el concepto de distribución muestral y el estudio completo de la inferencia.

Un aspecto a destacar es la importancia que el documento curricular remarca dentro del apartado de las recomendaciones metodológicas, donde se pide que el estudiante desarrolle una actitud crítica ante la información estadística, para que sea capaz de interpretar y evaluar la información estadística que recibe en distintas situaciones, como son los medios de comunicación. Por otro lado, el currículo concede importancia al lenguaje matemático, siendo un objetivo que el estudiante sea capaz de comunicarse usando correctamente este lenguaje. Estos dos aspectos guardan similitud con las dos competencias que, según Gal (2002), caracterizan la cultura estadística.

El análisis comparativo de los documentos curriculares a nivel estatal pone en relieve que las directrices del nuevo currículo se caracterizan por considerar un conjunto de contenidos más amplios en cada curso y anticipar la enseñanza de determinados contenidos, siendo este hecho más acusado o relevante en la modalidad definida como “orientada a las enseñanzas académicas” de la educación secundaria obligatoria (MECD, 2015). También la comparación con las directrices internacionales muestra que el currículo español sigue una tendencia común en países de nuestro entorno.

#### **1.6. ORGANIZACIÓN DEL TRABAJO Y RESUMEN DE LA METODOLOGÍA**

Para finalizar el primer capítulo, se describen brevemente unos elementos sobre su metodología, que luego se desarrollan en los diferentes estudios empíricos. El enfoque general de la investigación es descriptivo y exploratorio, con algunos elementos interpretativos y explicativos, puesto que evaluamos el efecto de algunas variables controladas sobre las dependientes de nuestro estudio y también porque tratamos de interpretar, bajo nuestro marco teórico, las respuestas obtenidas y relacionarlas con los resultados de otras investigaciones (Bisquerra, 1989; León y Montero, 2002). Tiene elementos de investigación experimental, puesto que no se realiza un control y

manipulación de variables independientes, sino que se engloba en la investigación cuasi - experimental (Campbell y Stanley, 2015).

De acuerdo a Bisquerra (1989) podemos clasificar nuestro trabajo como investigación aplicada, puesto que sus resultados pueden utilizarse para la mejora de la enseñanza del muestreo. Por sus fuentes, es empírica, pues se basa en la recogida y análisis de datos. Para Tamayo (2004) se trataría de un diseño transversal, puesto que tomamos los datos de varias muestras que se comparan en una sola ocasión. Es un diseño ortogonal o equilibrado, pues el número de sujetos en cada muestra es muy similar.

Una vez que se han descrito aquellos puntos que permiten enmarcar y caracterizar la investigación que llevaremos a cabo en la tesis doctoral, en este apartado describimos la organización de la Memoria, que describe los resultados de la misma, y que consta de dos partes bien diferenciadas. La primera parte de la Memoria hace referencia a los fundamentos del trabajo, mientras que la segunda parte presenta la metodología utilizada y los resultados de dos estudios empíricos de evaluación con estudiantes de diversas edades y curso escolar (Figura 1.6.1).

Los fundamentos utilizados para enmarcar el trabajo son los siguientes:

- *Análisis curricular*, que se ha presentado en este capítulo, y que engloba el estudio de dos directrices curriculares, en primer lugar, el currículo oficial estatal (MEC, 2007a y 2007b), por el que se establece las enseñanzas mínimas correspondientes a la etapa de educación secundaria obligatoria y bachillerato y en segundo el Currículo Básico (MECD, 2015), que se recoge en la Ley Orgánica para la Mejora de la Calidad Educativa (LOMCE). Todo ellos se ha complementado con el estudio del Boletín Oficial de Aragón (BOA) elaborado desde el Departamento de Educación, Cultura y Deporte (2007, 2008, 2016a, 2016b), y se ha llevado a cabo tanto en educación secundaria obligatoria como en bachillerato. Este estudio permite enmarcar el significado institucional del muestreo en nuestro trabajo, para su posterior comparación con el significado personal identificado en la evaluación realizada a los estudiantes.
- *Marco teórico*, en el que apoya la investigación y que es el enfoque ontosemiótico de la cognición e instrucción matemática (Godino, 2002; 2017; Godino y Batanero, 1994;1998; Godino, Batanero y Font, 2007; 2019), que se ha venido desarrollando y utilizando a lo largo de más de veinte años dentro de nuestro equipo de investigación. Dicho marco teórico se ha presentado en este capítulo, donde se



introduce los elementos del enfoque ontosemiótico que se utilizarán en la parte empírica. En particular, nos centramos en las ideas de objeto matemático, prácticas y significados, así como sus tipos y facetas; comprensión y evaluación, función semiótica y conflicto semiótico.

- *Antecedentes de investigación sobre muestreo*, que permiten ubicar nuestro trabajo y discutir sus resultados, validando las conclusiones obtenidas en el mismo. Su estudio nos ayuda a reflexionar en el diseño del instrumento de evaluación. En concreto, se han considerado algunos ítems presentes de otros trabajos, lo cual permite comparar nuestros resultados con estudios anteriores. Estas tareas han sido adaptadas y se han variado algunas variables de las mismas para permitir comparar su efecto sobre las respuestas de los estudiantes. La descripción de las investigaciones relacionadas con nuestro objeto de estudio se presenta con detalle en el Capítulo 2, organizado en torno a diferentes temáticas relacionadas con el muestreo como es el concepto de muestra, representatividad y variabilidad muestral, comprensión de la aleatoriedad, la probabilidad frecuencial y el teorema central del límite, entre otros puntos.

Una segunda parte de la memoria consiste en la descripción de dos estudios de evaluación con estudiantes. A continuación resumimos las características metodológicas generales de los estudios empíricos, que se describen con mayor detalle en los capítulos correspondientes.

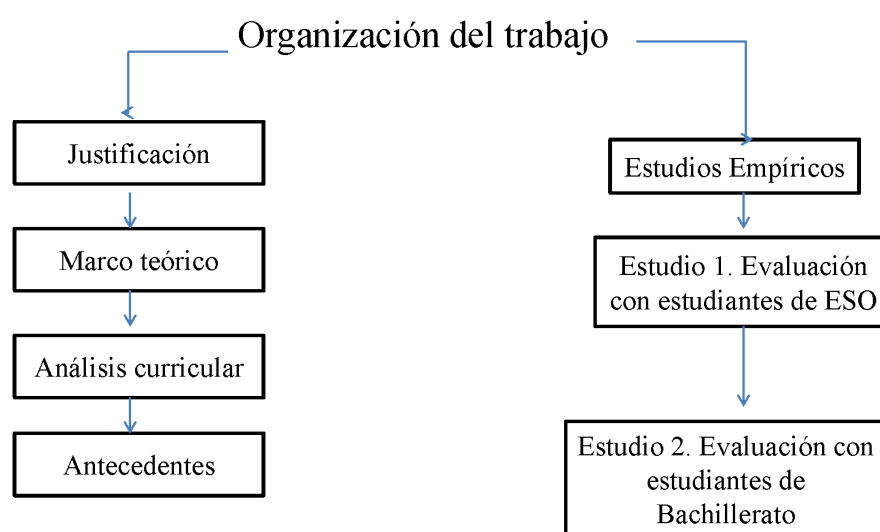


Figura 1.6.1. Organización del trabajo

El primer estudio empírico (Estudio 1, Capítulo 3) es una evaluación de la comprensión intuitiva de ideas sobre muestreo de un grupo de estudiantes de segundo y cuarto curso de educación secundaria obligatoria. Se trata de una investigación básicamente exploratoria y cuantitativa en la que se proponen cuatro tareas abiertas relacionadas con la distribución binomial, en las cuales se varía sistemáticamente el número de ensayos y el valor de la proporción.

El análisis de los datos es cuantitativo y se ha llevado a cabo con el paquete estadístico SPSS. Consiste en la realización de histogramas y gráficos de caja de la distribución de los valores medios y rangos de las cuatro respuestas de los estudiantes, tanto globalmente como por grupo. Se realiza una clasificación de los mismos en diferentes intervalos, que dan idea de su comprensión del valor medio y la variabilidad en el muestreo. Se comparan entre sí los resultados dependiendo de las variables de la tarea y del curso en que están los estudiantes. Se realizan contrastes estadísticos (test t de diferencia de medias independientes) para comparar las diferencias encontradas y analizar si los resultados son estadísticamente significativos. Se complementa con un estudio de los casos atípicos por ausencia de variabilidad o variabilidad extrema.

El segundo estudio empírico (Estudio 2, Capítulo 4) complementa la evaluación anterior con otra dirigida a estudiantes de bachillerato. Se utiliza el mismo cuestionario empleado en el Estudio 1 y se analizan las respuestas de los estudiantes, de los cuales a más de la mitad se les pidió también argumentar las respuestas. El análisis de los datos repite el realizado en el Estudio 1, con cuyos resultados se compara y además usa el análisis de varianza para comparar las diferencias de los estudiantes de Bachillerato con los dos cursos de Educación Secundaria del Estudio 1. Se completa el estudio cuantitativo con el análisis de las correlaciones de los valores medios por un lado y de los rangos por otro en los cuatro ítems, para analizar tendencias en las respuestas del mismo estudiante a los diferentes ítems. También se estudian las puntuaciones totales y parciales en el global de la muestra y se comparan por grupo.

Además, se realiza un análisis detallado cualitativo de las respuestas, clasificándolas en diferentes categorías que se infieren inductivamente por medio del análisis de contenido (Krippendorff, 2013). Dichas categorías permiten identificar diferentes concepciones correctas y algunos sesgos sobre la aleatoriedad en los estudiantes que se comparan con las descritas en trabajos previos. Se analiza la distribución de estas categorías entre los estudiantes y en función de las variables de tarea.

## **CAPÍTULO 2.**

### **ANTECEDENTES**

- 2.1. Introducción
- 2.2. Niveles de uso de los conceptos de inferencia
- 2.3. Heurísticas y sesgos en el razonamiento sobre muestreo
- 2.4. Comprensión del concepto de muestra y de sesgo en el método de muestreo
- 2.5. Comprensión de las ideas de representatividad y variabilidad muestral
- 2.6. Comprensión de la distribución muestral
- 2.7. Comprensión del Teorema Central del Límite
- 2.8. Muestreo en el razonamiento inferencial informal de los estudiantes
- 2.9. Comprensión de la probabilidad desde el punto de vista frecuencial
- 2.10. Generación y reconocimiento de la aleatoriedad
- 2.11. Otras investigaciones
- 2.12. Conclusiones del estudio de los antecedentes

#### **2.1. INTRODUCCIÓN**

La investigación sobre la comprensión del muestreo realizada tanto desde la didáctica de la matemática como desde la psicología, e incluso desde la propia estadística, describe dificultades de comprensión en este tema que han de ser tenidos en cuenta por los profesores al planificar su enseñanza. Este capítulo presenta una síntesis de las investigaciones más relevantes sobre la comprensión del muestreo y la distribución muestral, así como de otros conceptos relacionados con ellos, como es la aproximación frecuencial a la probabilidad. Dichas investigaciones han sido llevadas a cabo bajo diferentes marcos teóricos y metodologías y, nos han servido de base para formular nuestros objetivos e hipótesis, así como de apoyo para analizar y discutir los resultados de nuestra investigación.

Para realizar esta síntesis, hemos partido de los trabajos que analizan la investigación en educación estadística, como los de Batanero (2018), Batanero y Borovcnik (2016), Ben-Zvi, Bakker y Makar (2015; 2018), Castro Sotos, Vanhoof, Noortgate y Onghena (2007), Harradine et al. (2011) y Makar y Rubin (2018). Además, se ha realizado una búsqueda sistemática en revistas de didáctica de la matemática y didáctica de la estadística, servidores de educación estadística en Internet, actas de congresos y bases de datos bibliográficas.

Batanero (2018) analiza la investigación realizada sobre la didáctica de la inferencia estadística en los últimos treinta años, dando cuenta del espectacular incremento y avance de la misma. Es por ello difícil sintetizar todo el trabajo realizado, que se resume en las actas de los congresos organizados por la IASE y en revistas de educación estadística. Nosotros hemos realizado un primer avance de esta síntesis en Begué, Batanero, Ruiz y Gea (2017), que ahora retomamos.

En este capítulo, en primer lugar, se analizan los niveles de generalidad de un mismo concepto (variable, distribución y resúmenes de la misma) que caracterizan la inferencia estadística y que el estudiante ha de manejar simultáneamente para resolver problemas de inferencia. Este análisis revela la importancia y complejidad que adquiere el concepto de distribución y las posibles dificultades que pueden tener los estudiantes al pasar de un nivel del mismo a otro.

Seguidamente, se analizan las investigaciones sobre heurísticas y sesgos en el razonamiento llevadas a cabo, principalmente, desde la psicología, donde el muestreo cobra gran relevancia en la toma de decisiones y evaluación de situaciones de riesgo. En tercer lugar, el análisis se centra en la revisión bibliográfica de aquellas investigaciones enfocadas en la comprensión de los conceptos de muestra y de los diferentes tipos de muestreo aleatorio. Este análisis revela las concepciones erróneas que los individuos tienen sobre algunas propiedades de la muestra, en particular sobre los factores que determinan su variabilidad y sobre el sesgo en el muestreo. Además, se analizan las investigaciones más relevantes llevadas a cabo con estudiantes, centradas en la comprensión del concepto de muestra en entornos virtuales, que resultan al llevar a cabo un proceso de simulación realizado con herramientas de software específico.

La reflexión que se realiza en el primer apartado conduce a la identificación de la complejidad del término distribución muestral sobre el que se centra el aprendizaje de la inferencia estadística y cuya comprensión es necesaria para trabajar posteriormente los intervalos de confianza o contraste de hipótesis. Por consiguiente, el siguiente apartado se centra en presentar algunas de las investigaciones cuyo objetivo supone el análisis de la comprensión de dicho concepto y de sus propiedades por los estudiantes.

Por otra parte, la consideración del enfoque frecuencial de la probabilidad pone en relieve la estrecha relación que se establece entre la probabilidad y la estadística. De hecho, se identifica dicho vínculo cuando se pide estimar la proporción en una población a partir de la muestra, pues la proporción muestral sería un estimador de la probabilidad de ocurrencia de un suceso. Recíprocamente, estimar una probabilidad

---

teórica a partir de la frecuencia relativa en una muestra grande de ensayos es un problema de muestreo e inferencia. Por esta razón, la última de las revisiones bibliográficas se centra en mostrar aquellas investigaciones enfocadas en la comprensión de la probabilidad desde su enfoque frecuencial y el reconocimiento de la aleatoriedad.

Finalmente, se sintetizan aquellos aspectos más relevantes que se desprenden de la literatura identificada, señalando los puntos que son considerados para la elaboración del instrumento de evaluación en nuestra investigación y para el análisis de las respuestas de los estudiantes.

## **2.2. NIVELES DE USO DE LOS CONCEPTOS DE INFERENCIA**

Aunque existen diferentes metodologías de inferencia estadística (Batanero y Borovcnik, 2016; Batanero y Díaz, 2015), nos centramos en esta tesis en la metodología frecuencial, debido a que utiliza el muestreo y la distribución muestral como base para sus procedimientos; además, es la que oficialmente se enseña en el bachillerato español y también muy frecuentemente en la universidad.

La inferencia estadística frecuencial se fundamenta en el enfoque frecuencial de la probabilidad, el Teorema Central del Límite y en la aplicación de dos tipos de procedimientos o técnicas, en función de la finalidad u objetivo del problema. Por un lado, si el objetivo es proporcionar un margen de variación para un determinado parámetro desconocido de la distribución, junto con una expresión de la confianza en dicha estimación, entonces la finalidad es determinar un intervalo de confianza. Por otro lado, el objetivo puede ser confirmar o rechazar una determinada afirmación acerca de la distribución de probabilidad del fenómeno estudiado, implicando entonces la realización de un contraste de hipótesis. Un primer análisis de dichos procedimientos pone en relieve los distintos conceptos que subyacen al trabajo con la inferencia. En primer lugar, las ideas de variable y distribución, valor esperado y dispersión, población y muestra, estadístico y parámetro. Como señalan Lui y Thompson (2009), estos procedimientos exigen un esquema coordinado entre diversos conceptos, además de los citados, como son los de probabilidad, distribución de los valores de un estadístico muestral, hipótesis nula y alternativa, regiones crítica y de aceptación, tipos de error, nivel de significación (o bien, el coeficiente de confianza y el intervalo),  $p$ -valor, así como la lógica de la inferencia.

Harradine et al. (2011), así como Makar y Rubin (2018), nos recuerdan que la inferencia estadística está constituida por tres elementos fundamentales que interactúan entre ellos, los cuales son: el proceso de razonamiento, los conceptos y los cálculos asociados. En relación con los conceptos, estos autores subrayan conceptos generales, tales como la variable aleatoria y estadística, distribución, media, varianza, aleatoriedad, la probabilidad y el muestreo, y otros específicos como el intervalo de confianza o el contraste de hipótesis, que coinciden con los citados con anterioridad. Además, indican que el proceso de cálculo se encuentra actualmente resuelto por la tecnología y que la principal problemática de la enseñanza es transmitir el razonamiento estadístico y la comprensión conceptual. Siguiendo a Garfield y Gal (1999), pensamos que la enseñanza debe favorecer que el individuo sea capaz tanto de utilizar las ideas fundamentales de estadística, como de desarrollar un razonamiento estadístico adecuado, con el fin de dar sentido a la información obtenida a través de la inferencia estadística.

Todos los autores mencionados coinciden en que la finalidad de la inferencia estadística es obtener información de la población a través de las muestras. Esta finalidad exige la reflexión sobre el conjunto de datos que se van a recoger, con el fin de elaborar conclusiones con un grado de probabilidad adecuado (generalmente, una probabilidad condicional a un modelo de distribución en la población). Para conseguir este fin, es necesario trabajar con tres tipos de distribuciones asociadas a la inferencia estadística, las cuales están presentes durante todo el proceso (Harradine et al., 2011, Kadijevich, Kokol-Voljc y Lavicza, 2008):

- La *distribución teórica de probabilidad*, que modela los valores de una variable aleatoria tomada de una población o un proceso. Dicha distribución depende de algún valor del *parámetro*, que es generalmente desconocido. Con frecuencia, los problemas estadísticos relativos a una población son descritos por una variable aleatoria unidimensional con distribución teórica de tipo normal. En este caso, la distribución queda especificada por dos parámetros: su media y desviación típica, denotados por  $\mu$ ,  $\sigma$ , respectivamente. En otros casos, se trabajará con la distribución binomial, uniforme o exponencial, entre otras. En nuestro caso, se considera la distribución binomial, cuyo principal parámetro es la proporción de éxitos  $p$ .
- La *distribución del conjunto de datos*, que se obtiene a partir de la realización de varias repeticiones independientes del experimento aleatorio. Este conjunto de datos recibe el nombre de *muestra aleatoria simple*. Desde esta muestra, se definen distintos estadísticos tales como la media muestral y la desviación típica muestral,

que se denotan por  $\bar{x}$  y  $s$ , respectivamente. Estos estadísticos pueden aplicarse en el procedimiento para determinar los valores aproximados que toman los parámetros de la distribución teórica o bien para contrastar hipótesis sobre dichos parámetros. En los problemas planteados en nuestro estudio, se pide a los estudiantes pensar en cuatro muestras del mismo tamaño. En este caso, el estadístico correspondiente es la proporción de éxitos en la muestra (nosotros trabajamos con una forma equivalente que es el número de éxitos en cada muestra).

- La *distribución en el muestreo* de un estadístico, que se define como la distribución de probabilidad de todos los posibles valores que puede tomar el estadístico muestral en relación con las posibles muestras que constituyen el espacio muestral asociado al proceso o población que se desea estudiar. Aunque para una muestra particular tenemos un solo valor del estadístico, al pensar en las infinitas muestras que se pueden tomar de una población, tenemos una variable aleatoria; la distribución de esta variable aleatoria es la distribución en el muestreo. Por ejemplo, si la muestra de tamaño  $n$  se tomó de la distribución normal:  $N(\mu, \sigma)$ , la distribución en el muestreo de la media de la muestra  $\bar{x}$  tiene una distribución normal  $N(\mu, \frac{\sigma}{\sqrt{n}})$ . En nuestros Estudios 1 y 2, sólo se piden cuatro muestras, cuyo número de éxitos son elementos de la distribución muestral.

La necesidad de trabajar en el mismo problema con las tres distribuciones anteriores, para las que se define una media y desviación típica, origina una dificultad relevante para los estudiantes. En estadística descriptiva, Batanero et al. (1994) señalan que la unidad de análisis es cada dato (referido a una persona u objeto), mientras que al estudiar un estadístico en la muestra, como puede ser la media de una muestra, su análisis se refiere a la media de los datos de la muestra que constituyen. Además, al comenzar el estudio de la inferencia, el interés incide en conocer información sobre la población y estimar, continuando con nuestro ejemplo, la media teórica de la población. Este nuevo enfoque exige la obtención de distintas muestras de una población particular. En este caso, la unidad de análisis es cada una de las posibles muestras de un mismo tamaño, que se toman desde la población. Entonces, la media de cada muestra conduce al estudio de la distribución en el muestreo del estadístico denominado media muestral. De modo que, el grado de complejidad asociado a este último análisis se corresponde con un nivel de abstracción mayor.

De hecho, los estudiantes confunden con frecuencia estos diferentes niveles de concreción del concepto de distribución en estadística descriptiva e inferencia, lo que es un problema relacionado con la enseñanza del muestreo (Schuyten, 1991). Desde otro punto de vista, la complejidad de un concepto matemático dependerá del nivel de razonamiento algebraico que el trabajo con el mismo requiera del estudiante. En este sentido, si usamos la jerarquía propuesta por Godino, Neto, Wilhelmi, Aké, Etchegaray y Lasa (2015), veremos que el trabajo con la distribución de la población y la distribución muestral requiere el uso de los niveles 4 y 5, caracterizados por la presencia de parámetros y el trabajo con los parámetros, lo que aumenta su dificultad. Por ejemplo, al tipificar para calcular probabilidades de la distribución muestral, los estudiantes operan con los parámetros de la distribución.

Batanero, Díaz, Contreras y Roa (2013), al igual que Harradine et al. (2011), indican que el razonamiento inferencial requiere la coordinación de distintos conceptos: los datos, la población de donde se tomaron y las posibles muestras de la misma. Por tanto, exige conectar las tres distribuciones citadas anteriormente. A su vez, la idea de muestra y distribución muestral requieren de una adecuada comprensión del concepto de muestreo, por lo que se produce una situación circular, pues cada concepto se apoya en el resto. Esto supone una gran complejidad semiótica y explica en parte los múltiples errores que cometen los estudiantes.

Como señalan Wild y Pfannkuch (1999) o Meder y Gigerenzer (2014), uno de los modos fundamentales del razonamiento estadístico implica la percepción de la variabilidad debida a diversas fuentes como pueden ser las muestras seleccionadas mediante el proceso de muestreo. Las ideas fundamentales de distribución y variabilidad (Burril y Biehler, 2011) son la base de la comprensión de la variabilidad y representatividad muestral, que sostiene la consideración de la distribución muestral como objeto de análisis para inferir información de la distribución teórica (Ben-Zvi et al., 2015).

### **2.3. HEURÍSTICAS Y SESGOS EN EL RAZONAMIENTO SOBRE MUESTREO**

Desde el campo de la psicología se han llevado a cabo las primeras investigaciones sobre el muestreo, centradas en el contexto de la toma de decisiones en situaciones de incertidumbre (Harradine et al., 2011). Los psicólogos del razonamiento se han interesado por la forma en que se toman estas decisiones en actividades de responsabilidad, tales como el diagnóstico médico, la política, economía o el terreno



---

militar. Esta investigación mostró que cuando el sujeto se ve confrontado con una decisión en la que interviene la probabilidad, hay una tendencia a cometer errores que causa serios perjuicios. En el estudio del muestreo, la tarea típica consiste en preguntar por la probabilidad de un valor de la proporción muestral en muestras de diferente tamaño. Se realizan con estudiantes universitarios o profesionales.

Las investigaciones sobre el muestreo se realizaron dentro del programa denominado heurísticas y sesgos (Kahneman y Egan, 2011; Kahneman, Slovic y Tversky, 1982), donde el término *heurística* se concibe como una regla de conducta, de carácter generalmente inconsciente, que guía la resolución de las tareas de razonamiento complejas, porque reduce la información de las mismas. Por tanto, estos procesos cognitivos reducen el nivel de complejidad de la tarea durante el proceso de resolución de la misma. Generalmente son un buen recurso para resolver los problemas, pero en ocasiones producen sesgos en las conclusiones que se obtienen, al aplicarlas. Consideremos, por ejemplo, las siguientes secuencias de resultados de lanzar 5 monedas: CCC++, C+C+C, CC+C+. Aunque las tres tienen exactamente la misma probabilidad, cuando se pregunta a una persona hay una tendencia a indicar que las dos primeras no pueden ser resultados de un muestreo aleatorio.

En particular, en los estudios psicológicos se describen tres heurísticas fundamentales: la representatividad, la disponibilidad y el “ajuste y anclaje”. Además, se estudian los sesgos asociados, así como sus implicaciones teóricas y prácticas. Las siguientes líneas describen las dos primeras heurísticas citadas. No se considera la heurística del “ajuste y anclaje”, que se analiza en contextos experimentales bajo la probabilidad subjetiva, donde la toma de decisión se fundamenta en la información dada en la consigna, que influye de manera decisiva en la posterior estimación subjetiva de éxito (sobreconfianza o subconfianza), como producto de una fijación.

- La *heurística de representatividad* (Tversky y Kahneman, 1982) implica que la probabilidad de obtener una muestra se estima comparando la similitud que guarda la muestra con la población a la que pertenece o bien con el proceso de muestreo que la ha generado. De acuerdo a Tversky y Kahneman (1974, p. 1125): “las personas esperan que las características esenciales de un proceso aleatorio aparezcan representadas no sólo globalmente en una secuencia de resultados, sino también localmente en cada una de sus partes”. No se tienen en cuenta, por tanto, otros factores que puedan afectar a la solución del problema.

Los sesgos asociados son la *insensibilidad al tamaño de la muestra* y la *falacia del jugador* (Tversky y Kahneman, 1974). El primer sesgo implica que la persona asume que una muestra, aunque sea pequeña, siempre representa a la población, lo cual conduce a graves consecuencias en el trabajo estadístico. Sabemos que esta creencia no es cierta, pues, en primer lugar, el método de muestreo puede estar sesgado y por otro, incluso en el muestreo aleatorio, si la muestra es pequeña podemos obtener una muestra no representativa. En este sentido, Kahneman et al. (1982) indican que la persona que es insensible al tamaño de la muestra cree en la “ley de los pequeños números”, es decir, espera una convergencia estocástica a la distribución teórica incluso en pocas repeticiones de un experimento, porque generaliza indebidamente la ley de los grandes números. Ello conlleva a la expectativa de intervalos de confianza con una amplitud demasiado baja en muestras pequeñas. Además, esperan una replicación de los experimentos; es decir, esperan que los primeros resultados de un experimento sean similares a los futuros.

Por otro lado, la *falacia del jugador* supone considerar que el resultado de un experimento aleatorio afectará en la probabilidad de sucesos futuros. Por ejemplo, después de una racha de tamaño moderado de un mismo suceso, se espera que aumente la probabilidad de obtener el suceso contrario. Esta creencia también es infundada, pues, los experimentos son independientes unos de otros. Si se supone que los siguientes resultados seguirán el patrón observado, se habla de *recencia positiva* y si se piensa que se compensarán los resultados futuros con los observados, *recencia negativa*.

- La *heurística de la disponibilidad* (Tversky y Kahneman, 1974) consiste en estimar la verosimilitud de un suceso basándose en la información previa de la persona sobre la ocurrencia del suceso de interés, es decir, los recuerdos. La aplicación de dicha heurística supone la estimación de la probabilidad, en función de la facilidad por encontrar ejemplos de situaciones similares (Díaz, 2003).

Se identifican tres sesgos como consecuencia de la aplicación de esta creencia. En primer lugar, *el sesgo de equiprobabilidad* (Lecoutre, 1992), que consiste en pensar que los resultados de cualquier fenómeno aleatorio, en particular, cualquier valor de un estadístico muestral, son igualmente probables. Serrano (1996) relaciona esta creencia con una de las concepciones históricas del concepto de aleatoriedad donde, al comienzo del estudio de la teoría de la probabilidad, se relacionaba aleatoriedad y equiprobabilidad. Sin embargo, esta concepción se supera rápidamente mediante el

enfoque frecuencial, donde la probabilidad puede aplicarse a sucesos no equiprobables. El sesgo de equiprobabilidad ha sido recientemente explicado por Chernoff y Russel (2012), para lo cual describen la *falacia de la composición*, que consiste en transferir a un todo una propiedad que se cumple en una de sus partes. Según los autores, cuando los sucesos de un experimento aleatorio son equiprobables (como la cara y la cruz al lanzar una moneda) los sujetos suelen extender indebidamente la propiedad de equiprobabilidad a cualquier muestra de resultados del experimento compuesto; por ejemplo, consideran equiprobables obtener 0, 1 o 2 caras al lanzar dos monedas.

Otro de los sesgos relacionados es el *enfoque del resultado* (Konold, 1989), que implica interpretar el resultado de un experimento desde el enfoque subjetivo de la probabilidad. En otros términos, el sujeto no interpreta la situación desde un enfoque frecuencial, porque identifica en forma demasiado exacta la frecuencia relativa con la probabilidad. Cuando se pide tomar una decisión, las personas que siguen este enfoque transforman las preguntas sobre valores de probabilidad (¿Cuál es la probabilidad de que?) en una decisión dicotómica (¿ocurrirá o no?). Por ello, ante una probabilidad, digamos del 80%, espera el porcentaje exacto de ocurrencia del fenómeno, pues no asume la variabilidad del muestreo (Sánchez, García-García y Mercado, 2018). Además, cualquier pequeña desviación del experimento, respecto a sus expectativas, la explica en forma subjetiva, por lo que su respuesta supone una predicción de la probabilidad en función del grado de creencia que otorga a la obtención de dicho suceso. Konold (1989) indica que los sujetos que razonan de acuerdo al “enfoque en el resultado aislado” no ponen en relación las diferentes repeticiones de un experimento aleatorio, por lo que pueden tener dificultad para comprender el enfoque frecuencial de la probabilidad.

Finalmente, la *ilusión de control* (Langer, 1982; Yarritu, Matute y Vadillo, 2014) se corresponde con el hecho de que el sujeto no distingue la diferencia entre los juegos de habilidad y los de azar. Por tanto, como muestra Langer en una serie de experimentos, el sujeto tiene la creencia de poder controlar el azar, pues tiene una expectativa demasiado alta en la probabilidad personal de éxito en la situación. Esta concepción implica falta de comprensión de la aleatoriedad, pues uno de los rasgos de la misma es la imposibilidad de control y se muestra frecuentemente en los jugadores compulsivos. Además, esta creencia se refuerza cuando se incrementa la motivación (por ejemplo, si se espera un fuerte premio) y la persistencia (cuanto más se juega o se gana más veces). Un ejemplo de esta falacia es la necesidad que siente la gente de control de su entorno, incluso en situaciones caóticas o aleatorias (Allan y Jenkins, 1980). Explica también la

sorpresa o desengaño de la mayoría de las personas cuando afrontan una enfermedad o una pérdida, pues, aunque entiendan que esa situación ocurre un tanto por ciento de veces, no se ven formando parte de este tanto por ciento.

Serrano (1996) compara el uso de diferentes heurísticas, incluida la representatividad, en 140 estudiantes de 14 años sin enseñanza previa de la probabilidad y de otros 140 de 17 años al finalizar la secundaria, los cuales habían estudiado probabilidad en cursos anteriores. El investigador descubre que la heurística de la representatividad no es tan empleada frecuentemente como mencionan otros autores, sino que algunas respuestas erróneas a tareas sobre muestreo se deben al “enfoque en el resultado” descrito en las investigaciones de Konold (1989). En trabajos posteriores con una muestra algo mayor (n=277), se observan también pocas mejoras del razonamiento respecto a las heurísticas en el grupo de mayor edad que había tenido más instrucción en probabilidad, aunque hay una pequeña mejora con la edad en cuanto al uso de la representatividad (Batanero, Serrano y Garfield, 1996; Serrano et al., 1998).

Resultados similares son obtenidos por Guisasola y Barragués (2002) en una investigación con estudiantes universitarios, entre los que encuentran la heurística de la representatividad y el enfoque en el resultado. Describen también el *sesgo determinista* como un razonamiento que intenta explicar, mediante causas y efectos, los fenómenos aleatorios.

Pollatsek, Konold, Well y Lima (1991) observan que algunos estadísticos expertos utilizan el proceso de extracción con reemplazo de bolas de una urna como un modelo concreto para poder explicar lo que es un proceso de muestreo. En cambio, muchos estudiantes no llegan a aceptar que estos dos procesos sean isomorfos y en lugar de ello, conciben el muestreo de forma inadecuada. Por ello, no llegan a comprender la independencia de los diferentes ensayos en el muestreo, lo que provocaría el empleo de la heurística de la representatividad, incluso con muestras muy pequeñas.

#### **2.4. COMPRENSIÓN DEL CONCEPTO DE MUESTRA Y DEL SESGO EN EL MÉTODO DE MUESTREO**

Tras la identificación de los principales errores referidos al razonamiento inferencial en términos de heurísticas, las siguientes líneas investigaciones están centradas en observar cómo los estudiantes interpretan las ideas de muestra y muestreo. La toma de muestras para realizar inferencias sobre poblaciones desconocidas constituye el núcleo de la estadística y, por tanto, es esencial comprender cómo varían

---

estas muestras para realizar decisiones informadas (Ben-Zvi et al., 2015).

Mientras la psicología contribuye a la identificación de heurísticas y sesgos en el razonamiento sobre muestreo, las investigaciones en el campo de la didáctica de la matemática están centradas, principalmente, en observar el aprendizaje de los estudiantes. Algunas de estas investigaciones analizan la comprensión del concepto de muestra y del sesgo en el método de muestreo por estudiantes de varias edades. Por ejemplo Watson, en varios trabajos (Watson, 2004; Watson y Kelly, 2005; Watson y Moritz, 2000a y 2000b), analiza el desarrollo de la comprensión del concepto de muestra en estudiantes entre 8 y 15 años. La autora aplica a los estudiantes cuestionarios y entrevistas, en los que les pregunta qué significa para ellos una muestra y les pide proporcionar ejemplos de muestras.

En primer lugar, Watson y Moritz (2000a) analizan el desarrollo de la comprensión intuitiva de la idea de muestra en un conjunto de 62 estudiantes (8-9 años, 11-12 años, 14-15 años) y posteriormente (Watson y Moritz, 2000b) amplían la muestra a 3000 estudiantes de las mismas edades. A partir de los resultados de estas investigaciones, diferencian tres niveles de desarrollo en la comprensión del concepto: en el primero sólo se comprende la terminología del muestreo; en el segundo, se comprenden además las aplicaciones del muestreo en varios contextos; y en el tercero se adquiere una capacidad crítica para discutir conclusiones a partir de muestras que no sean adecuadas desde el punto de vista estadístico.

Se observa una diferencia estadísticamente significativa en el grado de comprensión del muestreo, medido por la puntuación total al cuestionario empleado con la edad. Además, algunos de los grupos participantes en el segundo estudio (Watson y Moritz, 2000b) recibieron una instrucción sobre muestreo que fue efectiva, por lo que los autores sugieren la necesidad de enseñar este tema desde el final de la educación primaria; por otro lado, también se identifica que los estudiantes tienen dificultad al realizar la transición hacia el significado formal que adquiere el término de muestra en estadística. Aunque pueden hacer generalizaciones adecuadas desde una muestra pequeña en relación a la población a la que pertenece, esta idea intuitiva no incluye aceptar la variabilidad intrínseca al concepto de muestra (Ben-Zvi et al., 2015). Desde esta observación, Watson y Moritz (2000a) indican la necesidad de hacer explícitas las diferencias del concepto en función del contexto en el que aparece, como por ejemplo, comparar las diferencias entre tomar una muestra pequeña de comida, la cual está asociada a una población homogénea, con la obtención de una muestra de una población

de sujetos para analizar una característica de los mismos.

En otro estudio, Watson (2004) lleva a cabo una investigación longitudinal en el que la muestra de estudiantes está constituida por 38 estudiantes que participaron en alguno de sus estudios previos. En particular, 19 de los estudiantes participan en la investigación de Watson y Moritz (2000a) y tres participaron en un estudio piloto anterior a dicha investigación. Los autores plantean a los estudiantes entrevistas, en las que les preguntan si han oído la palabra muestra y qué significa para ellos, así como por las razones para elegir ciertas muestras. Además, les piden comparar diferentes formas de tomar una muestra. Los autores indican que los estudiantes conceden mayor legitimidad a las muestras tomadas con algún procedimiento sesgado, como por ejemplo, muestras voluntarias que aquellos métodos que generan muestras aleatorias, como el muestreo aleatorio estratificado. Un resultado similar lo obtienen en Watson y Kelly (2005), donde se pide a 639 estudiantes entre 8 y 14 años decidir entre varios ejemplos de muestreo, aquellos que consideren adecuados. Concluyen que, con frecuencia, los estudiantes prefieren métodos sesgados y no confían en el muestreo aleatorio para producir muestras representativas.

A una conclusión similar llegan Saldanha y Thompson (2002) en un experimento de enseñanza con 26 chicos de 16 a 18 años. Estos autores también concluyen que la mayoría de los estudiantes no llegan a diferenciar claramente los tres niveles de distribución presentes en el muestreo (distribución de la población, distribución de la muestra y distribución del estadístico en el muestreo). Los investigadores llegan a estas conclusiones pidiendo a los estudiantes que extraigan muestras de 10 bolas de una población (con proporción desconocida de bolas blancas) y analicen la variabilidad de las diferentes muestras. El experimento intenta que el estudiante pase de la idea de una única muestra a visualizar la proporción muestral como variable aleatoria y conciba cómo debe ser su distribución. La última tarea es tratar de estimar la composición de la población, pero muchos estudiantes fallan, indicando que hay una “mayoría” de blancas, sin especificar la estimación de la proporción poblacional. Es decir, no son capaces de razonar desde la muestra a la población.

Como resultado de todas estas investigaciones, en Watson (2004) la autora sugiere las características de las posibles tareas que los estudiantes son capaces de llevar a cabo con éxito en los tres niveles de comprensión del muestreo, que reproducimos en forma simplificada en la Tabla 2.4.1. Como vemos, diferencia tres niveles y un total de seis categorías o subniveles, en función de la comprensión que se tenga de la terminología

dentro de un contexto o de la capacidad de cuestionamiento crítico. Además, tiene en cuenta la comprensión de diferentes tipos de muestras, según su tamaño y el tipo de selección. En cada categoría describe las competencias que se supone adquiridas por los estudiantes situados dentro de la misma. Los diferentes niveles suponen una adquisición gradual del concepto de muestra, de la capacidad para proporcionar ejemplos de muestras cada vez de mayor tamaño, así como de sugerir métodos de muestreo cada vez más sofisticados.

Tabla 2.4.1. Niveles de comprensión del concepto de muestra (Watson, 2004)

---

**Nivel 1. Comprensión de la terminología**

Pequeñas muestras sin selección (Categoría 1)

- Puede proporcionar ejemplos de muestras.
- Puede describir una muestra como una pequeña parte, o menos comúnmente como un ensayo/prueba.
- Acepta una muestra de tamaño menor que 15.
- No sugiere ningún método de selección, o un método particular.

Pequeñas muestras con un método elemental de selección (Categoría 2)

- Proporciona ejemplos de muestras.
- Describe una muestra como una pequeña parte o como un ensayo/prueba.
- Acepta una muestra de tamaño menor que 15.
- Sugiere la selección aleatoria informalmente con una simple expresión para elegir cualquiera.

**Nivel 2. Comprensión de la terminología en el contexto**

Pequeñas muestras con pre-selección de los resultados (Categoría 3)

- Proporciona ejemplos de muestras.
- Describe una muestra como tanto como una pequeña parte y como un ensayo/prueba.
- Acepta una muestra de tamaño menor que 15.
- Sugiere la selección de elementos por estratos (teniendo en cuenta el peso).

Muestras ambiguas (Categoría 4)

- Proporciona ejemplos y descripciones de las muestras.
- Puede indicar indiferencia acerca del tamaño de la muestra, a veces basado en aspectos irrelevantes.
- Puede combinar tamaños pequeños con una apropiada selección de métodos o sensibilidad parcial a sesgos, o selección de muestras de tamaño grande mediante un método inapropiado.

Muestras grandes con selección aleatoria o estratificada (Categoría 5)

- Proporciona ejemplos de muestras.
- Describe una muestra como una pequeña parte y como un ensayo/prueba.
- Puede referirse al término de promedio.
- Sugiere la selección basada en un proceso aleatorio o en un método de distribución geográfico.

**Nivel 3. Cuestionamiento crítico de afirmaciones sin justificación**

Sensible al sesgo presente en muestras grandes (Categoría 6)

- Proporciona ejemplos de muestras, a veces asociadas a una encuesta.
  - Describe una muestra como una pequeña parte y como un ensayo/prueba.
  - Puede referirse a términos como promedio o representatividad.
  - Sugiere un tamaño de la muestra de al menos 20 individuos o un porcentaje de la población.
-

- Sugiere la selección basada en un proceso aleatorio o en un método de distribución por geografía.
- Expresa preocupación por la selección de la muestras para evitar sesgos.
- Identifica muestras sesgadas en artículos de periódico que indican resultados de una encuesta.

Finalmente se adquiere la idea de sesgo en el proceso de muestreo y la competencia para identificar una muestra sesgada a partir de la descripción del método de muestreo.

Tabla 2.4.2. Niveles de razonamiento sobre muestreo (Moreno y Vallecillos. 2001)

Constructos	N1: Idiosincrásico	N2: De transición	N3: Cuantitativo	N4: Analítico
Población y muestra	Comprensión intuitiva. Confunde población y muestra.	Concepto de población; identifica población discreta. Identifica población o muestra en contexto concreto.	Concepto de población. Población de tipo discreta/continua. Identifica población y muestra sólo en ciertos contextos.	Concepto de población. Concepto de espacio muestral. Identifica siempre población y muestra .
Proceso de inferencia	Criterio subjetivo.	Criterio subjetivo y/o numérico con errores. Concepción determinista.	Criterio numérico informal. Concepción identidad.	Criterio numérico y expresión formal. Concepción correcta.
Tamaño de la muestra	No reconoce la variabilidad muestral. Indiferencia ante el tamaño de la muestra.	Importancia de la variabilidad muestral. Interés en el tamaño muestral en algunos contextos.	Reconoce la influencia del tamaño de la muestra en la estimación. Relaciona el tamaño de la muestra con la estimación en contextos numéricos.	Valora la sensibilidad del tamaño de la muestra. Relaciona el tamaño de la muestra con la estimación en todos los contextos.
Tipos de muestreo	Concepto de muestreo. No encuentra diferencias entre los tipos de muestreo aleatorios. No reconoce las fuentes de sesgo.	Métodos de muestreo aleatorios. Diferentes tipos de muestreo. Indican posibilidad de sesgos.	Tipos de muestreo aleatorio. Reconoce posibilidad de sesgos.	Métodos de muestreo y estimación. Reconoce distintos métodos adecuados. Valoran sensibilidad en el sesgo.

Moreno y Vallecillos (2001) analizan la comprensión del muestro por estudiantes de secundaria, describiendo niveles de comprensión de la taxonomía SOLO (Biggs y Collis, 1982), extendiendo el trabajo de Watson y Moritz (2000a y 2000b) sobre muestreo. Definen cuatro niveles de razonamiento en relación a la población y muestra, proceso inferencial, tamaño de la muestra y tipos de muestreo. Un resumen se presenta



---

en la Tabla 2.4.2. En el primer nivel, llamado *idiosincrásico*, el pensamiento inferencial se centra en aspectos irrelevantes en los diferentes conceptos, mostrando un pensamiento intuitivo.

El nivel *de transición* se caracteriza por abordar la tarea, centrándose en sólo un aspecto relevante de ella. El estudiante empleará alguna característica estadística para identificar la población y la muestra, realizar una inferencia estadística, justificar la influencia del tamaño de la muestra, identificar distintos tipos de muestreo y reconocer sesgos en los diferentes muestreos. A nivel *cuantitativo* se centran en varios aspectos relevantes para la resolución de la tarea. Por último, en nivel *analítico* se esperarían respuestas que integren distintas partes o distintos aspectos entre sí.

## **2.5. COMPRENSIÓN DE LAS IDEAS DE REPRESENTATIVIDAD Y VARIABILIDAD MUESTRAL**

Otros autores analizan la comprensión que tienen los estudiantes de las propiedades del muestreo, en especial, si llegan a diferenciar la representatividad de las muestras y su variabilidad, así como los factores que las determinan.

Rubin, Bruce y Tenney (1991) entrevistaron a 12 estudiantes de bachillerato antes de que hubiesen estudiado este tema y les hicieron algunas preguntas sobre el muestreo. Los autores observaron la influencia de la idea de representatividad, pues pensaban que cualquier muestra elegida con un procedimiento adecuado es representativa. Este exceso de confianza en la representatividad de las muestras lleva a los estudiantes a estimar a la baja su variabilidad. Por ello, los estudiantes subestiman la frecuencia con que se obtienen valores muestrales cerca de los extremos de la distribución muestral y conceden una frecuencia muy alta a los valores centrales. No parecen tomar en cuenta el tamaño de la muestra para analizar su variabilidad.

Son muchos los conceptos relacionados con la comprensión del muestreo, entre ellos los de aleatoriedad, independencia, distribución, centro, dispersión y sus relaciones. Los estudiantes han de relacionar todas estas ideas para comprender el muestreo (Bakker, 2004). A continuación analizamos algunas de estas investigaciones.

Moses (1992) indica que en la inferencia, una muestra proporciona “alguna” información sobre la población y de este modo aumenta nuestro conocimiento sobre la misma. Es decir, la inferencia estadística puede entenderse como una colección de métodos para aprender de la experiencia. Por ello, la comprensión del muestreo implica la armonía entre dos ideas opuestas: 1) la representatividad muestral, es decir, la

muestra tendrá frecuentemente características similares a las de la población, si ha sido elegida tomando las precauciones apropiadas; y 2) la variabilidad muestral, que nos indica que no todas las muestras son iguales entre sí. El fin de tomar una sola muestra sería cuantificar el grado de atipicidad en relación con todas las otras muestras que podrían haber sido extraídas (Saldanha y Thompson, 2002), aunque este fin no es comprendido por todos los estudiantes.

Shaughnessy, Watson, Moritz, y Reading (1999) presentan una tarea sobre una población de caramelos de dos colores, pidiéndoles formar muestras de la misma a 700 estudiantes de enseñanza secundaria de tres países. La mayoría reconocen la variabilidad en el número de caramelos de un color en las diferentes muestras, pero dan diferentes razones. Cuando han estudiado probabilidad, consideran más la variabilidad en el muestreo que sus compañeros.

Shaughnessy, Ciancetta y Canada (2004) realizan una investigación sobre la comprensión que los estudiantes tienen sobre el concepto de variabilidad en muestreo. Se trabajó con 272 estudiantes de entre 10 y 19 años utilizando un contexto de caramelos de dos colores (40% rojos y 60% negros). Los resultados se obtienen del análisis de una serie de tareas sobre muestreo, combinado con entrevistas personales. Los autores pidieron a los estudiantes dar el número de caramelos rojos en una muestra de 10 elementos y otra de 100 elementos. También les preguntaron si esperarían repetición de los resultados en una segunda muestra y qué resultados considerarían atípicos. La cuarta parte de estos estudiantes esperaban obtener el mismo resultado en las dos muestras repetidas del mismo tamaño y algunos estudiantes propusieron en sus muestras resultados muy poco probables (como 100% de negros). Por otro lado, la mayoría de los estudiantes sobreestimaron la variabilidad de la distribución muestral, independientemente del tamaño de la muestra, mientras algunos le otorgaron muy poca variabilidad.

Los autores identifican tres concepciones sobre las muestras: a) aditiva (la más frecuente) que consiste en considerar las diferentes muestras como subconjuntos disjuntos de la población; b) proporcional, en la que se comprende el valor esperado de la distribución muestral; y c) distribucional (la menos frecuente), donde se comprende tanto el valor esperado como la variación de la distribución muestral. Como resultado se observa que la mayoría de los estudiantes se sitúan en el primero de los razonamientos citados, que se relaciona con aquellas respuestas que se guían por la proporción que se presenta en la población. En relación con el concepto de variabilidad, se identifica que

muchos de los estudiantes creen en la posibilidad de resultados muy variados, independiente del tamaño de la muestra. En contraposición, otros otorgan muy poca variabilidad al conjunto de posibles resultados que se puede obtener en un muestreo.

Saldanha y Thompson (2002) llevan a cabo una experiencia de enseñanza en la que estudiantes de 16 a 18 años usan la simulación de muestras aleatorias de una población binomial cuya proporción se conoce. De este modo, los estudiantes pueden generar un gran número de muestras de la misma población y estudiar sus características y los valores de estadísticos en ellas. Los autores indican que la mayoría de los estudiantes en su experimento muestran una comprensión aditiva del muestreo al considerar la muestra como un conjunto de una población y múltiples muestras como múltiples subconjuntos disjuntos (Shaughnessy et al., 2004), es decir, no conciben que diferentes muestras puedan contener un mismo elemento. Otros presentan una concepción multiplicativa de la muestra que queda caracterizada por relacionar el concepto de muestra con la población, así como con la distribución muestral. Esta concepción conlleva una relación de conceptos más rica, favoreciendo la aceptación de muestras que sean relativamente inusuales. Por tanto, ayuda a la comprensión de la inferencia estadística.

Bakker (2004) analiza un proceso de instrucción con estudiantes de 13-14 años, apoyado en representaciones gráficas asociadas a las muestras obtenidas, donde el tamaño de las muestras va aumentando. El autor concluye que la enseñanza favoreció que los estudiantes comprendieran que un mayor tamaño de la muestra implica una menor variabilidad en el muestreo y, por tanto, permite inferir mejor características de la población. Así mismo, sugiere permitir al estudiante la formulación de conjeturas sobre las muestras, fundamentada en el ciclo de investigación descrito por Wild y Pfannkuch (1999), el cual consiste en plantear un problema y recoger y analizar los datos para tratar de darle respuesta. Dicho proceso estimula al estudiante a adquirir el esquema conceptual de distribución, lo cual beneficia el desarrollo del razonamiento sobre el muestreo (Meletiou-Mavrotheri y Papanastasiou, 2015).

## **2.6. COMPRENSIÓN DE LA DISTRIBUCIÓN MUESTRAL**

Como se ha citado anteriormente, el concepto de distribución muestral es un tópico abstracto, que requiere que el estudiante coordine distintos conceptos como: muestra, distribución, variabilidad y muestreo de un cierto estadístico en muestras repetidas de la misma población. La investigación relacionada suele utilizar poblaciones binomiales, cuya proporción conoce el estudiante. La dificultad en su comprensión conduce a que

muchos estudiantes no sean capaces de interpretar los resultados de los cálculos realizados, como veremos en esta sección.

Well, Pollatsek y Boyce (1990) investigaron la comprensión de las distribuciones muestrales para la media, con 114 estudiantes de psicología sin instrucción. En el primer experimento de una serie, comparan dos versiones del problema: uno relacionado con el grado de exactitud con el que la población es representada por la distribución de las medias muestrales, y otro en que se pide estimación del porcentaje de que un determinado valor promedio sea mayor que un valor. En el segundo experimento los estudiantes reciben un mismo problema en tres versiones, relacionadas con la estimación de porcentajes en una o dos colas de la distribución muestral y en el intervalo central, siendo las respuestas mejores en este último caso. En el último experimento los estudiantes reciben instrucción y se pide razonar sus conclusiones. En general, los sujetos parecen comprender que los promedios de muestras más grandes se acercan más a la media de la población, pero no comprenden las implicaciones de estos sobre la variabilidad de la media muestral. Esto sigue aconteciendo aún después de haberlos instruido al respecto.

La investigación realizada por Saldanha y Thompson (2002) tiene como finalidad desarrollar un análisis epistemológico de las ideas de muestra, distribución muestral y márgenes de error, en estudiantes de 16 a 18 años. Para ello, los autores llevan a cabo una experiencia de enseñanza en la que usan software que permite la generación de muestras aleatorias de una población. De este modo, los estudiantes pueden generar un gran número de muestras de la misma población y estudiar sus características, así como las de los estadísticos obtenidos de estas muestras. La investigación se centra en analizar el modo en que los estudiantes coordinan los tres niveles de la idea de distribución (distribución de la población, distribución de la muestra y distribución del estadístico en el muestreo).

Los autores identifican que algunos estudiantes tienen una imagen aditiva de la muestra, y tienden a verla como un conjunto de una población; por consiguiente, múltiples muestras se refieren a múltiples subconjuntos disjuntos de la población y diferentes muestras no puedan contener un mismo elemento. Otros estudiantes presentan una concepción multiplicativa de la muestra, en la que se relaciona el concepto de muestra con la población y la distribución muestral. En contraste con el primero, esta concepción conlleva una relación de imágenes más rica, que favorece la

---

aceptación de muestras que sean atípicas; no obstante, se detecta que la mayoría de los estudiantes se orientan hacia una imagen aditiva del concepto de muestra.

En definitiva, Saldanha y Thompson (2002) identificaron que los estudiantes no presentaban una distinción clara entre los distintos niveles asociados al concepto de distribución, anteriormente descritos. Por tanto, estos autores exponen que la utilización de software no es suficiente para conseguir una adecuada comprensión del concepto distribución muestral. Por ello, recomiendan que los estudiantes se familiaricen con el proceso de muestreo, además de permitirles que exploren los conceptos mediante actividades no guiadas. De este modo, el estudiante puede construir y evaluar su propio conocimiento. Otro aspecto relevante es proporcionar un espacio en el que los estudiantes puedan discutir las observaciones, una vez que la tarea ha concluido.

DelMas, Garfield y Chance (1999) diseñaron el software Sampling Distribution para trabajar las distribuciones muestrales. Con este software, organizaron un experimento de enseñanza con diversas actividades usando poblaciones normales u otras diversas formas, como U, asimétrica o uniforme. En la evaluación de los resultados observaron que los estudiantes podían trabajar sin problemas con varias distribuciones de probabilidad y simular la distribución muestral de diversos estadísticos con varios tamaños de muestras. Los investigadores analizaron el aprendizaje de las siguientes propiedades:

- Relación entre tamaño muestral y dispersión de la distribución muestral;
- Propiedades de los estadísticos como estimadores de los parámetros de la población;
- Comprensión del Teorema Central del Límite.

Los resultados del experimento mostraron el aprendizaje de los estudiantes, aunque un número significativo de ellos parecía no comprender las implicaciones básicas del Teorema Central del Límite. Pocos estudiantes comprenden que la variabilidad de la media muestral tiende a cero cuando el tamaño muestral crece; o que el tamaño muestral es un factor que afecta a la validez de las conclusiones de la inferencia o a la generalización de sus resultados.

Chance, delMas y Garfield (2004) presentan un proyecto de investigación que tiene como finalidad analizar si existe una mejora en el razonamiento sobre la distribución muestral en un grupo de 105 estudiantes (que van a iniciar los estudios superiores) a lo largo de un curso en que se usan herramientas de software específico para estudiar

inferencia. Los autores indican que los estudiantes no son capaces de explicar los conceptos que están en juego en los procedimientos de inferencia que aplican, a pesar de que pueden aprender cómo calcular un intervalo de confianza y lleven a cabo test de hipótesis. Es decir, aplican las técnicas como si fueran una “receta”, pero no saben describir el por qué de tales procedimientos.

En este sentido, los autores identifican que una de las razones de esta dificultad se debe a que los estudiantes no comprenden los conceptos de distribución muestral y variabilidad del estadístico en diferentes muestras, entre otros. No obstante, sí que presentan una comprensión adecuada de la variabilidad muestral en relación con el tamaño de la muestra. De acuerdo con estos autores, para lograr esta comprensión es necesario reflexionar previamente sobre los conceptos sobre los que se basa el concepto de distribución muestral, con la finalidad de interpretarlos de manera correcta en el contexto de las distribuciones muestrales. Los conceptos previos que señalan son: la muestra, el estadístico en la muestra, la variabilidad de las muestras y el efecto del tamaño sobre la variabilidad (la comprensión de algunos de estos conceptos son objeto de evaluación en nuestro trabajo). Como consecuencia, se observa también que los estudiantes no identifican las diferencias entre los distintos niveles conceptuales asociados al concepto de distribución, resultado que confirma la investigación de Saldanha y Thompson (2002).

Vanhoof, Castro Sotos, Onghena, y Verschaffel (2007) se plantean analizar la efectividad de un software denominado *Sampling Activity Distribution* (SDA), desarrollado por Chance et al. (2004) y delMas, Garfield y Chance (2004) con el fin específico de apoyar el aprendizaje de conceptos de inferencia, entre otros el de distribución muestral. Dicho software generaba muestras de tamaño dado de diferentes poblaciones (a elegir por el usuario) y representaba en diferentes ventanas la última muestra obtenida y la distribución muestral conforme se generaba. La muestra estuvo constituida por 221 estudiantes universitarios que estaban cursando el segundo curso de la titulación. Además, el plan de estudios incluye tres asignaturas en las que se imparten contenidos de estadística.

La comparación de los resultados de un pre-test y un pos-test que evalúan la comprensión de conceptos de inferencia enseñados en el curso, revela una mejora significativa de la comprensión de la distribución muestral en estos estudiantes, aunque solamente la mitad de los sujetos muestran una comprensión adecuada. No obstante, los autores señalan que esta mejora puede deberse a otros aspectos o variables que no se

controlan en el estudio, como por ejemplo, el campo de estudio de los sujetos, así como su experiencia previa, puesto que habían sido instruidos en conceptos propios de estadística.

Ko (2016) examina la comprensión de la distribución muestral en dos grupos en función a su nivel educativo. Por un lado, 65 estudiantes de 5° grado (10 a 11 años), 34 de los cuales presentan capacidad hacia las matemáticas; y por otro lado, 65 estudiantes de 8° grado (13 a 14 años), de los cuales 36 presentan capacidad hacia las matemáticas. Los estudiantes de 5° grado habían aprendido a representar y resumir conjuntos de datos, apoyándose en representaciones como tablas, el gráfico de barras, diagramas de tallos y hojas o medias. Por su parte, los estudiantes de 8° grado aprendieron gráficos circulares, tablas de frecuencias, polígonos de frecuencias, histogramas, las frecuencias relativa y acumulada y la probabilidad. El autor usa un cuestionario constituido por dos preguntas, la primera de las cuales está ligada a la variabilidad intrínseca en la distribución muestral (el estudiante tiene que determinar el resultado del tamaño medio de una familia coreana, apoyándose en el resultado de una muestra dada). En la segunda, se muestra la distribución asociada al número de lapiceros que cada estudiante lleva en su estuche. Se presentan cuatro distribuciones de la media en el muestreo y el estudiante debe determinar cuál corresponde a una muestra de tamaño 4 y a otra de 25.

El análisis de los datos se organiza en torno a dos estudios: a) El primer análisis supone la clasificación de las respuestas en términos de lenguaje y expresiones utilizadas, y las categorías identificadas emergen del propio estudio; b) El segundo análisis utiliza el modelo SOLO (Biggs y Collis, 1982) para reagrupar las categorías del primer análisis en cinco categorías para evaluar los resultados del aprendizaje. La taxonomía SOLO considera los siguientes niveles:

1. Pre-estructural: las respuestas que da el estudiante son equivocadas o no hay respuesta. No se relaciona con la pregunta sino con su experiencia personal.
2. Uni-estructural: la respuesta del estudiante se centra en un único aspecto de la pregunta, independientemente de si es relevante o no.
3. Multi-estructural: el estudiante establece determinados aspectos correctos, pero no avanza en la complejidad de la respuesta.
4. Relacional: el estudiante enumera varios aspectos correctos en la respuesta y puede establecer relaciones entre ellos.

5. Abstracto ampliado: el estudiante ha cumplido con los anteriores niveles y puede ir más allá de la pregunta para establecer relaciones con elementos externos a la pregunta.

En la Tabla 2.6.1 se describe los distintos niveles o categorías descritos por Ko (2016), que clasifican la comprensión de la distribución muestral según dicha taxonomía.

Tabla 2.6.1. Niveles de comprensión de la distribución muestral según Ko (2016)

Nivel	Comprensión del concepto de distribución muestral
0	No entiende la variabilidad muestral, Piensa que diferentes muestras tienen el mismo estadístico.
1	Confusión del conjunto de datos en una muestra con el mismo estadístico: los estudiantes que no prestan atención a las regularidades del estadístico muestral.
2	Se centran en la variabilidad del estadístico de la muestra pero no identifican la tendencia central del estadístico de la muestra.
3	Se centra en las propiedades tanto de variabilidad como en la tendencia central del estadístico en el muestreo.
4	Comprende la relación entre el tamaño de la muestral y la variabilidad muestral: los estudiantes reconocen la tendencia central y la variabilidad del estadístico y entienden, a su vez, la relación entre el tamaño de la muestra y la variabilidad muestral.

Para finalizar este apartado, citamos otras investigaciones que analizan la comprensión de la distribución muestral desde un punto de vista formal y después de haber realizado un experimento de enseñanza, principalmente, con estudiantes universitarios. Entre ellas citamos la de Alvarado, Galindo y Retamal (2013), que evalúan el aprendizaje de estudiantes de ingeniería sobre la distribución muestral y su aplicación al cálculo de intervalos de confianza mediante problemas abiertos y un cuestionario. Los autores informan del éxito de su propuesta para la comprensión de la convergencia de la distribución muestral hacia la distribución normal y el valor esperado de dicha distribución. Sin embargo, las propiedades de la varianza de dicha distribución y la relación con el tamaño de la muestra continuaron siendo complejas.

## 2.7. COMPRENSIÓN DEL TEOREMA CENTRAL DEL LÍMITE

Aunque en nuestro estudio no se pregunta directamente por este teorema, una comprensión intuitiva del mismo es importante para comprender la forma de la distribución muestral conforme aumenta el tamaño de la muestra, y percibir la tendencia aproximada a la distribución normal.



---

Méndez (1991), a partir de un análisis de 10 libros de estadística, identifica cuatro propiedades básicas en la comprensión de este teorema:

- 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 a infinito;
- La varianza de la distribución muestral es menor que la de la población;
- 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.
- La forma de la distribución muestral crece en altura y decrece en dispersión a medida que el tamaño muestral crece.

Los estudiantes que participan en la investigación de Méndez mostraron una comprensión poco profunda del teorema y desconocimiento del vocabulario técnico. En su mayor parte, usaban los datos disponibles sin tener en cuenta la población de la que provenían y sin considerar el tamaño de muestra.

Otro trabajo que se destaca como representativo de la comprensión del Teorema Central del Límite es el realizado por Alvarado (2007). Partiendo de un estudio sobre el desarrollo histórico de este teorema y utilizando el enfoque onto-semiótico, en primer lugar estudia algunos textos universitarios, llegando a las siguientes conclusiones:

- El teorema se presenta usualmente con demasiada formalidad, y predominantemente con expresión simbólica, descuidando las representaciones gráficas y simulaciones, que pueden ayudar a la comprensión.
- En la resolución de problemas, se enfatiza el cálculo algebraico y las propiedades no se demuestran adecuadamente.
- Son prácticamente inexistentes los problemas basados en datos reales o que requieran el uso de software estadístico.

El autor diseña un curso sobre el Teorema Central del Límite para ingenieros, que previamente han seguido un curso de probabilidad, y analiza con detalle las respuestas de los estudiantes a las actividades propuestas, así como el aprendizaje logrado, por medio de una evaluación final. Entre sus conclusiones, destacamos las siguientes:

- Los estudiantes adquieren la comprensión intuitiva del teorema y lo aplican en casos sencillos, por ejemplo, para la distribución binomial.

- No se comprenden todas las propiedades, en especial, no se aplica la corrección de continuidad cuando es necesario y se confunde la suma de variables aleatorias con el promedio, al aplicar el Teorema Central del Límite.
- Se define incorrectamente, en algunas ocasiones, la variable aleatoria que sirve de modelo para resolver un problema aplicando este teorema.
- Se confunde el número de ensayos en un experimento binomial con el tamaño de la muestra en una población binomial.

Todas estas conclusiones le llevan a recomendar un mayor énfasis en la enseñanza de este teorema y el uso de recursos informáticos para realizar simulaciones, que contribuyan a la adquisición de experiencia con el teorema por parte de los estudiantes.

## **2.8. MUESTREO EN EL RAZONAMIENTO INFERENCIAL INFORMAL DE LOS ESTUDIANTES**

Encontramos algunos trabajos sobre muestreo dentro de la corriente denominada inferencia informal, que es interpretada en forma diferente por distintos autores, pero que se caracteriza en general por el uso de simulación y la sustitución de la distribución muestral teórica por otra empírica obtenida mediante el ordenador. Además, no se calculan probabilidades por métodos habituales, sino que se estiman a partir de las frecuencias relativas obtenidas en la simulación.

En algunos casos se simulan las muestras de una distribución teórica, pero no de una población dada. En otros se toma una muestra inicial de la población y se toman las muestras de la muestra inicial. En ninguno de los dos casos se trata la inferencia frecuencial clásica. En el primero, porque en inferencia clásica se trabaja con una única muestra. En el segundo porque se utilizan métodos de remuestreo, donde ya no se trabaja directamente con un modelo de distribución de la población, sino que se supone que toda la información de la población está contenida en una muestra inicial, de la cual se extraen nuevas muestras por simulación para obtener una distribución muestral del estadístico denominada distribución de remuestreo.

En todo caso, como se ha indicado, se trata el tema de una forma muy elemental, sin discutir los conceptos subyacentes y de ello se derivan diferentes definiciones de lo que sería la inferencia estadística informal (IEI). Por ejemplo, Zieffler, Garfield, delMas y Reading (2008); y Makar y Rubin (2009) indican las siguientes cuatro componentes en su modelo de IEI, que son: (1) Obtener una conclusión, que va más allá de los datos

analizados; (2) Uso de los datos como evidencia para obtener la conclusión; (3) Uso del lenguaje probabilístico que exprese incertidumbre acerca de la conclusión; y (4) Utilizar e integrar el conocimiento previo (conocimiento formal e informal) en el grado en que este conocimiento esté disponible. Observamos que realmente no se trata de inferencia, en el sentido estadístico del término, pues estas conclusiones se podrían aplicar también a la interpretación de información en el análisis exploratorio de datos.

Rossman (2008) indica que una inferencia estadística es una generalización de los datos de una muestra a una población, o una conclusión sobre la relación entre un modelo probabilístico y los datos. Considera posible usar la simulación para introducir a los estudiantes a la lógica de la inferencia estadística de manera informal. Por su parte, Makar, Bakker y Ben-Zvi (2011) la describen como una generalización probabilística sobre los patrones revelados por la información (datos) disponibles.

Uno de los trabajos recientes sobre inferencia informal es el de García-Ríos (2013), que realizó un estudio con la finalidad de explorar los niveles sobre inferencia estadística informal en una muestra constituida por 16 estudiantes de bachillerato (15-17 años) de una escuela pública mexicana. Los estudiantes de la muestra no estudiaron simultáneamente a la investigación contenidos de Estadística y Probabilidad.

El instrumento de evaluación está constituido por dos problemas, que los autores interpretan informalmente como prueba de hipótesis. En concreto, las cuestiones planteadas demandan al estudiante determinar si un tratamiento es efectivo, para lo cual se les proporciona los datos de una muestra. El análisis de las respuestas se realiza en torno a las cuatro componentes descritas por Zieffler et al. (2008), desde el que identificaron una serie de subcategorías de razonamiento de los estudiantes. Por ejemplo, el análisis de la primera de las componentes conduce a la clasificación de las respuestas dependiendo de si la generalización va más allá de los datos o no. Desde el estudio de las explicaciones dadas por los estudiantes, el estudio pone en relieve cuatro tipos de razonamiento estadístico informal:

1. Considerar un modelo o distribución inicial de referencia de la población y comparar los datos de la muestra con este modelo; sería un razonamiento similar al del estadístico cuando realiza un análisis de inferencia;
2. Limitarse sólo a calcular el porcentaje de éxitos esperado si fuese cierta la hipótesis a contrastar, y decidir con base a este porcentaje; sería un razonamiento que intuitivamente se aproxima al test de significación de Fisher, pero no se profundiza en la idea de distribución muestral;

3. Apoyar el razonamiento en el uso de creencias o teorías personales, lo cual de ningún modo constituye una inferencia estadística; y
4. Combinar el tercer razonamiento con alguno de los razonamientos 1) o 2).

Finalmente, los autores indican que el contexto considerado favorece la utilización de creencias en las explicaciones de los estudiantes. Otro de los resultados de interés es que la mayoría de los estudiantes utilizan los datos de la muestra dada en el enunciado como argumento de sus inferencias. No obstante, no se utilizan de una manera adecuada, es decir, no logran establecer argumentos que apoyen en un alto grado la conclusión a la cual se llega. Por otro lado, se observa que el uso de lenguaje probabilístico también fue una gran dificultad, así como la identificación de un modelo de referencia.

Por su parte, García-Ríos (2013) y García-Ríos y Sánchez (2014) realizan un estudio exploratorio sobre el razonamiento inferencial intuitivo de un grupo de 27 estudiantes de bachillerato (15-17 años), a través de dos problemas de inferencia estadística (media y proporción), de los cuales 13 estudiantes contestaron al primer problema y 14 estudiantes al segundo problema. Además, se les demanda buscar dos ejemplos en los que aceptasen o no su hipótesis, respectivamente. La metodología que guía la investigación es la Teoría fundamentada (Birks y Mills, 2011), una metodología de investigación usada en ciencias sociales la cual permite estudiar los fenómenos de interés a partir de los datos recopilados en un contexto natural. En este caso, los autores identifican la variación como categoría central que proporciona una explicación del razonamiento inferencial informal y su relación con otras categorías que corresponden al muestreo, valor crítico e incertidumbre (Tabla 2.7.1). Estas categorías se corresponden con las ideas fundamentales del contraste de hipótesis y de la estadística en general, tal y como señalan Garfield y Ben-Zvi (2008).

Los autores señalan que los estudiantes presentan un razonamiento inferencial intuitivo y no formal, puesto que determinan valores críticos sin ninguna operación o cálculo, de forma inmediata, global y autoevidente. En relación con el muestreo, existe una pequeña muestra de estudiantes que consideran muestras de tamaño 1 o toda la población, para obtener información de la misma. Solamente seis estudiantes conceden un grado de variación al valor crítico. Finalmente, los errores creen que son debidos a errores de medición, en lugar de justificarla en el grado inherente al trabajo estadístico.

Además, no conceden incertidumbre a la respuesta aportada. Esta observación explica la dificultad de determinar un valor crítico para rechazar o aceptar una hipótesis.

Para concluir, los autores señalan que la simulación apoyada en el software puede favorecer la percepción de la variación y su implicación en la inferencia estadística. Además, se identifica también una ausencia de un lenguaje probabilístico y una dificultad para medir la significación del estadístico en el muestreo, como se identifica en la investigación de García-Ríos (2013).

Tabla 2.7.1. Categorías de razonamiento en conceptos según inferencial informal (García-Ríos y Sánchez, 2014)

Concepto	Razonamiento	Propiedades
Muestreo	Muestra	Consideran la muestra de tamaño razonable como posible representante de la población.
	Muestra pequeña	Consideran la muestra de pequeño tamaño como representante de la población.
Valor crítico	Población	Pretenden analizar toda la población a través de una muestra.
	Determinista	No consideran la variabilidad para establecer un valor crítico.
	Semi-determinista	Utilizan la variabilidad solo en un caso (bien para rechazar o aceptar la hipótesis). Utilizan valores críticos diferentes.
Incertidumbre	Probabilístico	Utilizan la variabilidad para rechazar o aceptar la hipótesis; además, utilizan valores críticos diferentes.
	Medición	Atribuyen posibles errores a factores humanos o de medición.
	Variación	Atribuyen posibles errores a factores humanos o de medición y consideran la variabilidad natural de las muestras.

Meletiou-Mavrotheris y Paparistodemou (2015) llevan a cabo una investigación donde el interés recae en caracterizar el razonamiento que siguen un grupo de estudiantes sobre las muestras y el muestreo desde una aproximación informal. El estudio exploratorio se lleva a cabo en dos fases en un instituto. La primera fase se corresponde con una encuesta dirigida a una muestra de estudiantes de 11 años (n=69), que tiene por objeto obtener información con el conocimiento inicial sobre el significado y el papel de conceptos tales como muestra, tamaño de la muestra, método de selección y sesgo muestral. Las tareas son tomadas de investigaciones previas, en concreto, la primera y segunda cuestión se toman de la investigación de Watson y Moritz (2000a y 2000b). La última etapa de la investigación se centra en el diseño y puesta en marcha de una experiencia de enseñanza guiada por una trayectoria hipotética de aprendizaje (HLT).

Este marco teórico considera que el conocimiento matemático es adquirido a lo largo de diferentes etapas del conocimiento, donde el entorno y la cultura son factores en la adquisición del aprendizaje. Se reflexiona sobre cómo será el pensamiento y la comprensión del estudiante durante esta trayectoria.

El análisis de los datos asociados al primer cuestionario apoya los resultados obtenidos en las investigaciones previas. En particular, los autores observan que la mayoría de los estudiantes (85%) habían oído la palabra muestra en contextos no escolares. Por otro lado, el 46% no dieron explicaciones acerca del concepto en la primera tarea. Los autores identifican cinco niveles de la taxonomía SOLO, de mayor nivel a menor: (1) Nivel relacional, (2) Nivel multiestructural, (3) Nivel uniestructural, (4) Nivel preestructural y (5) No hay respuesta. En esta clasificación, el 40% de los estudiantes se localizan en el nivel uniestructural, caracterizado por concebir el concepto de muestra como una idea que se usa solamente para probar propósitos. Mientras que solamente el 12% se sitúan dentro de un nivel multiestructural, donde el estudiante es capaz de identificar la validez de seleccionar una muestra en lugar de preguntar a todos los individuos de la población.

Por otro lado, el análisis del proceso de enseñanza revela que el trabajo en grupo, donde se contemplan actividades tales como el debate, favorecen la evolución del pensamiento, ayudando a reorganizar sus concepciones previas y reconstruyendo los significados. Finalmente, los autores enfatizan la necesidad de introducir en el aula tareas cuyos contextos sean próximos a los estudiantes, así como permitirles la obtención del conjunto de datos con los que guiarles en el proceso de resolución de problemas estadísticos.

En resumen, los investigadores citados son conscientes de las dificultades de un razonamiento inferencial, incluso informal, en los primeros niveles escolares. Por ello, las investigaciones más bien tratan de evaluar la existencia de algunas intuiciones correctas que luego se puedan desarrollar para ayudar a los estudiantes en su razonamiento inferencial (Meletiou-Mavrotheris y Paparistodemou, 2015).

## **2.9. COMPRESIÓN DE LA PROBABILIDAD DESDE EL PUNTO DE VISTA FRECUENCIAL**

Siguiendo a Ben-Zvi et al. (2015), la idea de variabilidad que subyace en los conceptos de muestra y muestreo se fundamenta en la aleatoriedad y el azar, o el estudio de la probabilidad. Por otro lado, la Ley de los Grandes Números garantiza que las muestras con un tamaño mayor representan mejor a la población de la que fue tomada, de modo que los estadísticos asociados a dicha muestra están más próximos a los valores de los parámetros de la población. Entonces, la Ley de los Grandes Números supone la base del significado de la probabilidad desde el enfoque frecuencial, y

también se encuentra implícita en algunas de las ideas asociadas al concepto de muestreo. De este modo, se justifica el interés por realizar una revisión sobre aquellas investigaciones centradas en el análisis de la comprensión de la probabilidad desde el enfoque frecuencial.

La investigación didáctica sobre el significado frecuencial de la probabilidad ha aumentado recientemente, debido al cambio de las directrices curriculares que recomiendan este enfoque en la enseñanza de la probabilidad desde la educación primaria, como ocurre, por ejemplo, en España (MECD, 2014) y a la disponibilidad de software que facilita esta aproximación. A continuación se describen algunas investigaciones en las que se evalúa la comprensión de la probabilidad bajo este enfoque. A partir de la revisión de las mismas se observa que muchas de las tareas usadas para la investigación sobre la comprensión de la probabilidad desde el significado frecuencial pueden reinterpretarse en término de tareas de evaluación de la comprensión de ideas básicas de muestreo. De hecho, los ítems que constituyen nuestro cuestionario están adaptados de algunas de estas investigaciones.

En primer lugar, citamos la investigación realizada por Green (1983a), fundamentada en la elaboración de un cuestionario dirigido a niños ingleses (de 11 a 16 años), donde se incluye un ítem sobre la comprensión de la estimación frecuencial de la probabilidad asociada a un experimento con sucesos no equiprobables (lanzar al aire 100 chinchetas para determinar cuántas caen con la punta hacia arriba o hacia abajo). El autor sugiere que si el niño presenta el sesgo de equiprobabilidad, puede ignorar la información frecuencial dada en el ítem y considerar los sucesos como equiprobables.

En la prueba piloto (Green, 1983b), se pregunta a 66 niños que estimen el número de chinchetas que esperan que caigan con la punta hacia arriba y cuántas con la punta hacia abajo, cuando se lanzan 100 chinchetas sobre una mesa. Además, se les proporciona información previa, puesto que se les da datos asociados a la realización del experimento (donde 68 de 100 chinchetas caen con la punta hacia arriba). Por tanto, se espera que los niños den un resultado parecido, pero no idéntico. No obstante, los resultados revelaron que el 61% de los estudiantes mostraron el sesgo de equiprobabilidad, es decir, proporcionaron respuestas en las que la mitad de las chinchetas, aproximadamente, caerían hacia arriba, sin tener en cuenta la información frecuencial. Por otro lado, el 15% de los estudiantes mostraron una preferencia a la punta hacia arriba (correcta o parcialmente correcta), pero dando una cantidad muy alejada de la esperada. Tras los diferentes estudios pilotos y entrevistas con los niños, el

ítem se cambió un poco. A pesar de estas modificaciones, no existe una mejora del sesgo citado, el cual se mantenía en todas las edades.

La investigación de Green (1983a) constituida por una muestra de 2930 estudiantes, que supone una continuación de la citada anteriormente, refleja que el 64% presenta el sesgo de equiprobabilidad, y solamente el 17% dan una estimación correcta de tipo frecuencial. Luego, se observa que las tasas de acierto y sesgo no mejoraron con esta versión definitiva.

Por otro lado, la investigación realizada por Cañizares (1997), sobre una muestra de 253 niños españoles (11-14 años), considera el ítem propuesto por Green (1983a). En este caso, los resultados reflejan que el 64,1% de los niños muestran el sesgo de equiprobabilidad, mientras que sólo el 15% responde correctamente. Además, se observa un aumento no significativo con la edad, en la consideración del principio de equiprobabilidad. Este resultado se justifica desde la consideración de que, en aquella época, no se iniciaba el estudio de la probabilidad hasta los 14 años, por lo que la mayoría de los niños nunca la había estudiado. No obstante, la revisión del currículo oficial, expuesta en el Capítulo 1, revela que algunos contenidos de probabilidad son contemplados actualmente desde los primeros cursos de primaria. Por tanto, podemos esperar que nuestros resultados fueran mejores.

Sobre el significado frecuencial de la probabilidad, se identifican investigaciones cuyo propósito es describir la comprensión de los estudiantes sobre las características de las secuencias aleatorias. Desde esta perspectiva, Falk y Konold (1997) proporcionan una clasificación de las investigaciones enfocadas en el análisis de la comprensión de la aleatoriedad, considerando dos grupos: (a) Tareas de generación, las cuales se caracterizan porque el sujeto genera una serie de secuencias como resultados de un proceso aleatorio; y (b) Tareas de reconocimiento, en las que sujeto elige aquellas secuencias que considera aleatorias. De hecho, las tareas que nosotros proponemos se pueden considerar como de generación, aunque usamos secuencias muy cortas (sólo cuatro resultados).

La investigación de Green (1983a) recoge una tarea constituida por dos sucesiones que simulan el resultado de lanzar una moneda equilibrada 150 veces. La tarea de reconocimiento demanda que el estudiante identifique cuál de las dos secuencias es una secuencia aleatoria. El estudio muestra que la mayor parte de los niños escogen la secuencia no aleatoria. Por otro lado, el análisis de los argumentos dados por los estudiantes para la no aleatoriedad conduce a una clasificación de los mismos en



correctos e incorrectos. El autor identifica como correctos: patrón demasiado regular o rachas muy cortas, e incorrectos: patrón demasiado irregular, el porcentaje de caras no es 50% o rachas muy largas. Este estudio es relevante, porque una secuencia de resultados puede verse como una muestra generada por un proceso.

Posteriormente, Green (1991) realiza una investigación sobre una muestra de 305 niños ingleses (7 a 11 años) donde compara las respuestas a las preguntas formuladas sobre secuencias aleatorias. Una de las preguntas formuladas se corresponde con una tarea de generación, donde se pide que escriba una sucesión de tamaño 50 que simulara el lanzamiento de una moneda. El autor lleva a cabo un seguimiento longitudinal del grupo de estudiantes, durante cuatro años, a través del cual detecta que no existe una mejora durante esos años. Los resultados de la investigación reflejan que los niños tienden a subestimar la variabilidad en la secuencias. Además, escriben secuencias donde la proporción del número de caras y cruces es próxima  $\frac{1}{2}$  con una dispersión más pequeña que la que se espera de una secuencia aleatoria.

Serrano (1996) diseñó un cuestionario de 10 ítems, constituido por preguntas abiertas y de selección múltiple, que describen tareas de reconocimiento y generación de secuencias aleatorias, así como la interpretación frecuencial de la probabilidad. El cuestionario fue respondido por una muestra constituida por estudiantes españoles, entre los que se distinguen 147 estudiantes de 13 años, y 130 estudiantes de 17 años. Aunque hubo un alto porcentaje de respuestas correctas, pocos argumentos se basaron en propiedades matemáticas. En relación con la primera de las tareas comentadas, asociada al reconocimiento de sucesiones aleatorias, el autor identificó tres elementos que los estudiantes tienen en cuenta para juzgar si una secuencia es aleatoria o no: variabilidad local, regularidad global y grado de discrepancia entre la distribución esperada y la observada. Por otro lado, el autor identifica la ausencia de una comprensión adecuada de los siguientes aspectos: el uso de razones frecuenciales para explicar la ocurrencia de un suceso poco probable, la estimación de la frecuencia relativa de un suceso en una serie futura de experimentos cuando se conoce su probabilidad y que la estimación frecuencial de una probabilidad no sirve para predecir la ocurrencia del resultado de un único experimento.

Más recientemente Gómez, Batanero y Contreras (2014) elaboran un cuestionario constituido por cuatro ítems para evaluar algunas componentes del conocimiento matemático para la enseñanza de la probabilidad desde el enfoque frecuencial de futuros profesores de educación primaria. El primer ítem está dirigido a evaluar el conocimiento

común del contenido (Ball, Lubinski y Mewborn, 2001), y es una adaptación del propuesto por Green (1983a). Además, será uno de los que usaremos en nuestro estudio. Dicho ítem demanda a los participantes pensar cuatro resultados probables, si se repite el experimento. Los resultados se obtienen mediante el análisis de la media de los cuatro valores y su variabilidad. En este trabajo, los autores indican que sólo una tercera parte aproximadamente de los participantes tiene una intuición simultánea de la convergencia al valor esperado y de la variabilidad muestral. Por tanto, sugieren que los sujetos presentan pobres intuiciones iniciales sobre los experimentos aleatorios. Por otro lado, se identifica que los participantes presentan diferentes sesgos, como la equiprobabilidad o la heurística de la representatividad, o piensan que no es posible realizar la predicción.

Estos resultados coinciden con las investigaciones de Green (1983a, 1991), Serrano (1996) y Cañizares (1997), aunque son mejores. Por otro lado, Gómez et al. (2014) señalan que la adaptación del ítem proporciona información sobre la comprensión de la variabilidad. En este sentido, los resultados muestran que una parte relevante de los sujetos produce muestras de variabilidad extrema o de patrón determinista.

Valdez (2016) realiza un estudio sobre el razonamiento probabilístico de 30 estudiantes de bachillerato (17 - 18 años) para analizar si los estudiantes diferencian y ponen en relación los significados clásico y frecuencial de la probabilidad. La recolección de datos se hizo a partir de cuestionarios y entrevistas. Las situaciones que componen cada cuestionario se plantean en un contexto de urnas, en el que los cálculos son sencillos. Las entrevistas tienen como propósito esclarecer y profundizar en los razonamientos que los estudiantes manifiestan de manera escrita, además de observar si dichos razonamientos se modifican al llevar a cabo la simulación de algunas situaciones que conforman los cuestionarios.

Los cuestionarios se conforman de tres tareas, cada una, a su vez, compuesta por tres preguntas. En la primera tarea, dado el número de bolas blancas y negras de una urna con 4 bolas, se les pregunta la composición de una muestra de 1000 extracciones con reemplazamiento; el resultado en la extracción 1001 y cuál fue el primer resultado. En la segunda, a partir de dos urnas con bolas y la frecuencia de blancas y negras obtenidas de 1000 extracciones de cada una, se les pregunta qué urna elegiría si quiere sacar una bola negra en la extracción 1001 y asignar una probabilidad a este suceso en cada urna. En la tercera, se da una secuencia aleatoria de 10 bolas (blancas y negras) extraídas de cada una de las urnas anteriores y se les pregunta qué urna elegiría si quiere

sacar una bola negra en la extracción 11 y asignar una probabilidad a este suceso en cada urna. Además, se pide argumentar todas las respuestas.

El autor concluye que los estudiantes aceptan la variabilidad en la muestra cuando la distribución de bolas en la urna es equiprobable, y la ignoran al hacer predicciones sobre el posible resultado de una extracción. Reconocen el modelo equiprobable en la distribución de las bolas, pero consideran a la variabilidad en la muestra como excesiva si se aparta un poco de lo predicho por el modelo. Por otro lado, asignan la frecuencia relativa para estimar la probabilidad de un suceso, aunque conozcan la composición de la urna y la frecuencia se aparte mucho del modelo. Este autor crea, a partir de las respuestas de los estudiantes, una jerarquía de clasificación con cuatro niveles del desarrollo del razonamiento probabilístico (Tabla 2.9.1) con respecto a las ideas de variabilidad, aleatoriedad e independencia. En el primero, los estudiantes no han comenzado a desarrollar un razonamiento probabilístico, aunque conocen las definiciones de probabilidad clásica y frecuencial, mientras que en el último manifiestan un razonamiento probabilístico informal consolidado.

Tabla 2.9.1. Niveles de razonamiento probabilístico en tres conceptos (Valdez, 2016)

	Nivel 1	Nivel 2	Nivel 3	Nivel 4
Variabilidad	Esperan resultados sin variabilidad. En muestras pequeñas consideran las diferencias entre las frecuencias observadas y esperadas como significativas.	Crean que pequeñas diferencias son significativas en muestras grandes.	Comprenden que diferencias pequeñas no son significativas cuando la muestra es grande.	Reconocen que la variabilidad es grande cuando la muestra es pequeña y pequeña cuando la muestra es grande.
Aleatoriedad	Realizan una predicción determinista.	Realizan una predicción determinista, matizándola con lenguaje probabilístico.	Reconocen que no se puede predecir el resultado con exactitud.	Reconocen que no se puede predecir con certeza el resultado de un ensayo, pero sí la estabilidad de las frecuencias alrededor de un valor (probabilidad frecuencial) a largo plazo.
Independencia	Asignan probabilidades con base en las frecuencias relativas, pero introducen elementos no pertinentes.	Asignan probabilidades con base en las frecuencias relativas, ignorando el modelo.	Asignan probabilidades con base en el modelo e ignoran los resultados previos (muestra).	Reconocen la independencia, por tanto, ignoran lo ocurrido y asignan probabilidades con base en el modelo.

## 2.10. GENERACIÓN Y RECONOCIMIENTO DE LA ALEATORIEDAD

Las tareas que plantearemos a los estudiantes en los estudios empíricos en nuestra investigación se basan en que ellos tengan que imaginar y proponer muestras diferentes, de un tamaño dado y de una población de partida determinada. Nos interesamos en analizar la representatividad y variabilidad que revisten dichas muestras.

La elaboración de estas muestras supone que los estudiantes apliquen sus concepciones sobre la aleatoriedad y los procesos aleatorios, y por ello conviene también tener en cuenta las investigaciones que han analizado la comprensión de estos objetos matemáticos por parte de los estudiantes. Estas tareas son clasificadas por Falk y Konold (1997) en dos tipos principales (ver también Batanero, 2016):

1. Tareas de generación de secuencias aleatorias: En ellas se pide al estudiante producir secuencias de resultados de un proceso aleatorio típico, como el lanzamiento de una moneda;
2. Tareas de reconocimiento de la aleatoriedad, donde el estudiante debe elegir entre varias secuencias de resultados las que considera aleatorias.

En los dos tipos de tareas, Falk y Konold (1997) indican que los estudiantes muestran sesgos típicos, como la *falacia del jugador*, consistente en pensar que, después de una racha de resultados repetidos, la probabilidad del suceso contrario aumenta. Otro sesgo es la *representatividad local* (pensar que la frecuencia relativa debe tender a la probabilidad teórica en muestras de tamaño pequeño). En tareas de reconocimiento, se suelen rechazar las secuencias con rachas largas, considerándolas no aleatorias. En la generación de secuencias se observa una tendencia a cambiar de suceso ante rachas largas, lo que ocasiona demasiadas alternancias y rachas muy cortas. Estos y muchos otros más sesgos se describen en la sección 2.3.

Konold et al, (1991) plantean varias situaciones a los estudiantes, preguntando si son aleatorias o no. Los autores describen las siguientes creencias subjetivas de los sujetos sobre la aleatoriedad: a) los experimentos son aleatorios si todos los sucesos tienen la misma posibilidad de ocurrir; b) la aleatoriedad se considera como lo contrario a lo causal o la aleatoriedad es la propia causa; c) aleatoriedad como incertidumbre y d) aleatoriedad en relación con los conocimientos que poseemos sobre el fenómeno.

Green (1991) compara los resultados de 305 niños ingleses, de 7 a 11 años sobre generación de secuencias aleatorias, observando la tendencia a producir igual número de caras y de cruces. Respecto a la independencia de los resultados producidos por los

estudiantes, observa la longitud de las rachas producidas por los estudiantes, notando un sesgo importante a la alternancia de resultados, pues encontró que la longitud media de la racha más larga en su muestra era inferior al valor teórico. Observa que los niños tienden a subestimar la variabilidad en las secuencias y a aproximar a  $\frac{1}{2}$  la proporción de caras y cruces, con dispersión más pequeña de lo esperado para una secuencia realmente aleatoria.

Serrano (1996) propone a los estudiantes que participaron en su investigación algunas tareas sobre generación de la aleatoriedad. Los participantes mostraron su comprensión de las características generales de imprevisibilidad de fenómenos aleatorios y de la regularidad en la frecuencia de posibles resultados. Asimismo, se vio un razonamiento combinatorio correcto sobre la secuencia de resultados e identificación de sucesos elementales en el espacio muestral producto. Pero también aparecen estrategias o ideas incorrectas como la aplicación de la regla de Laplace en un contexto inadecuado; dificultad en la aplicación de la independencia de ensayos en contexto práctico; heurística de representatividad; confusión de la magnitud relativa de la proporción con el tamaño muestral; confianza en una continuación de la tendencia o de una compensación en pequeñas muestras; estabilización en los extremos (rápida o a largo plazo); mayor fiabilidad en las muestras pequeñas frente a las grandes o influencia del resultado observado sobre lo que ocurría en siguientes experimentos.

El autor describe algunas concepciones sobre la aleatoriedad que se asemejan a las históricamente admitidas como correctas, entre ellas: aleatoriedad como contrario a lo causado, aleatoriedad como equiprobabilidad, aleatoriedad como estabilidad de frecuencias relativas, aleatoriedad como imposibilidad de predicción o aleatoriedad como modelo matemático.

Chernoff (2009) se interesa por la comprensión de la aleatoriedad en futuros profesores, a los que propone tareas de reconocimiento de la aleatoriedad con secuencias de 5 a 8 resultados. El autor señala que las dificultades de algunos futuros profesores en la percepción de la noción de aleatoriedad, no se deben a falta de razonamiento probabilístico, sino por un análisis incompleto del espacio muestral del experimento. Los futuros profesores parten de un espacio muestral “subjetivo”, donde se basaron en la alternancia de sucesos, rachas largas y ambas propiedades a la vez, en vez de simplemente en el producto de espacios muestrales de los experimentos simples.

Un tipo de investigación diferente es la llevada a cabo por Batanero, Gómez, Gea y Contreras (2014), que combina las tareas de generación y reconocimiento de secuencias

aleatorias. Los autores proponen a una muestra de 208 futuros profesores de educación primaria un proyecto estadístico en que han de generar una secuencia de 20 lanzamientos de monedas de modo que parezca aleatoria a otra persona, y posteriormente se les pide decidir qué diferencias tienen estas secuencias inventadas y en qué se diferencian de una secuencia aleatoria. El análisis de las respuestas de los futuros profesores, los gráficos producidos y sus argumentos llevan a los autores a describir diferentes concepciones de la aleatoriedad, que reproducen las que históricamente se han aceptado dentro de la matemática y que habían sido descritas por Serrano (1996). Además, se mostraron ciertos sesgos como la ilusión de control, la representatividad o la equiprobabilidad.

## **2.11. OTRAS INVESTIGACIONES**

Es también importante tener en cuenta otras investigaciones que analizan la comprensión del concepto de distribución, de las cuales citaremos las más relevantes. Konold, Robinson, Khalil, Pollatsek, Well, Wing y Mayer (2002) se centran en analizar cómo los estudiantes razonan sobre un conjunto de datos y, en particular, cómo comunican las ideas de representatividad y variabilidad en relación a dicho conjunto de datos. El análisis recae en una serie de entrevistas a grupos de estudiantes de noveno curso de dos centros públicos, los cuales había recibido previamente ocho semanas de instrucción guiadas o asociadas a un proyecto, de modo que el conjunto de datos que se les presenta había sido trabajado bajo la dirección de dicho proyecto.

Los autores observan que los estudiantes tienden a utilizar rangos en la primera pregunta, así como comunicar tanto la variación como la representatividad en términos informales. Para la segunda cuestión, se les presentan las denominadas “gráficas de puntos”. En esta cuestión, identifican que en la mayoría de los grupos el conjunto de datos generado para la segunda pregunta es bastante consistente con los resúmenes iniciales. La última de las cuestiones muestra que los grupos tienden a ser más precisos en su discurso, señalando rangos de valores dentro de la distribución, sin ahondar en promedios o en el cálculo de parámetros estadísticos.

En definitiva, se observa cómo los estudiantes tienden a dar un rango de valores como indicador de los resultados más frecuentes, lo que se traduce en el uso de la idea de intervalos modales. Además, señalan que el uso de estos intervalos permite a los estudiantes expresar simultáneamente nociones de representatividad y variabilidad, características del concepto de muestreo. Por tanto, los autores sugieren que la idea de

---

los grupos modales puede constituir un punto de partida en el aprendizaje, para resumir el conjunto de datos.

Bakker y Gravemeijer (2004) llevan a cabo una investigación de diseño constituida en tres fases: (1) el diseño de materiales instruccionales, (2) los experimentos de enseñanza y (3) un análisis retrospectivo. En esta investigación, la última fase se lleva a cabo mediante un método de comparación de las constantes.

La investigación gira en torno a la cuestión de cómo los estudiantes de un determinado curso académico podrían aprender a razonar sobre distribuciones de una manera informal. Para ello, se diseña una experiencia de enseñanza apoyada en un software específico. La última fase de la investigación supone el análisis retrospectivo, el cual revela que los estudiantes tienden a dividir la distribución unimodal en tres grupos. De este modo, los estudiantes ven más representativo el intervalo donde se localiza la mayoría de los valores que el valor de la media. Konold et al. (2002) denominan a dicho intervalo “modal clump” o grupo modal. Además, estos autores consideran que estos intervalos modales pueden ser una manera adecuada de iniciar al estudiante en el razonamiento informal sobre conceptos tales como el centro, la dispersión, perspectiva que comparten otros autores.

Finalmente, los autores señalan que permitir a los estudiantes que inventen su propio conjunto de datos podría estimularles a pensar en el conjunto de datos como un todo en lugar de verlo como un conjunto de puntos. Los estudiantes desarrollaron nociones cualitativas de características de la distribución gracias a esta tarea. Por otro lado, los autores indican que una actividad prometedora es el aumento muestral, porque permite al estudiante darse cuenta de rasgos estables de los procesos.

En la siguiente Tabla 2.11.1 se resumen las características de cada uno de los estudios realizados por Bakker y Gravemeijer (2004) en torno al razonamiento sobre distribuciones según el tamaño de la muestra considerada, el tipo de estudio, la recolección de los datos, el número de lecciones implementadas y el nivel educativo al que se dirigen.

Tabla 2.11.1. Características de los estudios realizados por Bakker y Gravemeijer (2004)

Muestra	Tipo de estudio	Recolección de los datos	Nº de lecciones	Nivel
26 estudiantes	Entrevistas exploratorias (15 minutos para cada entrevista)	Audio	-	Enseñanza Secundaria General Media, Superior y Bachillerato.
Clase A (25)	Trabajo de campo exploratorio	Trabajo del estudiante, evaluación final, notas de campo, audios.	4	Enseñanza Secundaria General Superior
Clase F (27)	1ª Experiencia de enseñanza		12	Bachillerato
Clase E (28)	2ª Experiencia de enseñanza		15	Bachillerato
Clase C (23)	3ª Experiencia de enseñanza	Además de lo citado en la fila anterior, se llevo a cabo evaluaciones iniciales y videos.	12	Enseñanza Secundaria General Superior
Clase B (23)	4ª Experiencia de enseñanza		12	Enseñanza Secundaria General Superior
12 clases	Implementación	Informes de dos profesoras vía e-mail, notas de campo.	144	Enseñanza Secundaria General Superior y Bachillerato

## 2.12. CONCLUSIONES DEL ESTUDIO DE LOS ANTECEDENTES

Para finalizar el capítulo, resumimos a continuación las conclusiones obtenidas de la revisión de los antecedentes.

En primer lugar, se ha destacado el hecho de que una misma expresión (como distribución o media) remite a objetos matemáticos que son conceptualmente diferentes, y aunque puedan parecer semejantes entre sí, los estudiantes los confunden. Ello se debe a que se trata de diferentes niveles de uso del concepto, lo que en el enfoque ontosemiótico se pudiera describir como ejemplares de diferente nivel de un tipo general o concepto más amplio (dualidad de tipo extensivo - intensivo).

Ello causa mucha confusión a los estudiantes, porque no siempre se hace explícito el nivel con que se trabaja uno de estos objetos y por consiguiente los estudiantes los confunden entre sí. Por ello, aunque identifican los tres niveles de distribución que aparecen en el muestreo (datos, muestras y población), no se observa que relacionen la distribución muestral con la distribución de la población. Sin embargo, se trata de objetos matemáticos netamente diferentes, puesto que además el nivel de algebrización requerido para el trabajo con cada uno de ellos (Godino, Aké, Gonzato y Wilhelmi, 2014) es claramente diferente. Mientras que la distribución de datos de la muestra únicamente requiere un nivel algebraico muy elemental, ya que no se opera con



---

variables sino sólo se realizan operaciones aritméticas, el trabajar con la distribución de la población hace que intervengan sus parámetros y se opera con ellos al tipificar, igual ocurre al trabajar con la distribución muestral.

Desde el campo de la psicología se han descrito, además, el uso de heurísticas en nuestro razonamiento sobre el muestreo, que reducen la complejidad de los problemas de probabilidad, por tanto, dan una solución rápida a los mismos. Pero estas heurísticas, que son inconscientes, llevan con frecuencia a sesgos en las conclusiones. Aunque esta investigación no se ha llevado a cabo con una finalidad didáctica, es importante tenerla en cuenta en el aula, pues heurísticas como la representatividad o la disponibilidad se presentarán en las tareas que proponemos a los estudiantes.

Las investigaciones centradas en identificar la imagen que los estudiantes asocian al concepto de muestra revelan que aparentemente los estudiantes visualizan las muestras como conjuntos excluyentes de la misma población y no como conjuntos con elementos comunes. Respecto a los tipos de muestreo, los estudiantes no perciben los sesgos en muestras intencionales y con frecuencia prefieren métodos sesgados a los métodos aleatorios, considerando que proporcionan muestras más representativas. Dos ideas fundamentales son las de representatividad y variabilidad muestral, pero mientras que la primera se comprende aparentemente, no se percibe la relación del tamaño de la muestra con la variabilidad en el muestreo. Tampoco se percibe el ajuste de la distribución muestral a la curva normal cuando aumenta este tamaño, es decir, no hay una comprensión intuitiva del Teorema Central de Límite.

Se han definido diferentes modelos de razonamiento sobre la muestra y el muestreo, así como sobre la distribución muestral, y los estudiantes requieren un tiempo de maduración para progresar de unos a otros niveles; por tanto, estos niveles nos ayudan a interpretar resultados de la evaluación y a organizar acciones didácticas en el aula.

Desde el enfoque frecuencial de la probabilidad, el estudio de los antecedentes nos muestra que, respecto a la idea de muestra, se comprende bien la representatividad de la misma, pero no la de variabilidad. En las distintas investigaciones se identifica que los sujetos que componen la muestra o bien no saben reconocer la variabilidad en las tareas de reconocimiento, o bien no la saben reproducir en las tareas de generación. No obstante, las tareas demandan la reproducción de secuencias de resultados aleatorios, a diferencia de este trabajo, en el que se demanda la reproducción de muestras, con excepción del trabajo de Gómez et al. (2014) del cuál tomamos un ítem. Conclusiones

semejantes se obtienen de los trabajos relacionados con la percepción de la aleatoriedad, pues en las tareas de reproducción de secuencias aleatorias no siempre estas secuencias se ajustan a lo esperado en un proceso aleatorio.

En resumen, todos estos trabajos indican que hay dificultad en la comprensión de ideas básicas en el muestreo, en las que se han usado diferentes tipos de tarea. Nosotros trabajaremos con estudiantes españoles de tres cursos diferentes y adaptaremos algunas tareas usadas en investigaciones previas para evaluar dos puntos que no han sido muy tenidos en cuenta hasta ahora: Por un lado, el efecto de ciertas variables de tarea sobre las respuestas dadas por los estudiantes en la generación de muestras. Por ello, la evaluación se sustenta en una serie de cuestiones en las que se demanda al sujeto la elaboración de muestras grandes y pequeñas en relación con el experimento descrito en el enunciado. A diferencia de las investigaciones citadas anteriormente, los enunciados que configuran el cuestionario hacen referencia a fenómenos aleatorios caracterizados por sucesos equiprobables o no. Esta particularidad permite comparar respuestas asociadas a los mismos, siendo este punto objeto de análisis en este trabajo.

Por otro lado, trataremos de analizar los argumentos en que los estudiantes apoyan las muestras que generan, análisis que no ha sido llevado a cabo en investigaciones previas. Todo ello para tratar de evaluar si los estudiantes presentan una adecuada comprensión de la representatividad y la variabilidad muestral, así como del efecto del tamaño de la muestra en la misma.

# CAPÍTULO 3.

## ESTUDIO 1. GENERACIÓN DE MUESTRAS ALEATORIAS POR ESTUDIANTES DE EDUCACIÓN SECUNDARIA OBLIGATORIA

- 3.1. Introducción
- 3.2. Objetivos e hipótesis del Estudio 1
- 3.3. Contenidos sobre muestreo que se consideran en el estudio
- 3.4. Contexto educativo
- 3.5. Metodología
  - 3.5.1. Descripción de la muestra
  - 3.5.2. Método de recogida de datos
- 3.6. Análisis a priori del cuestionario y conocimientos evaluados
  - 3.6.1. Análisis del ítem 1
  - 3.6.2. Análisis del ítem 2
  - 3.6.3. Análisis del ítem 3
  - 3.6.4. Análisis del ítem 4
- 3.7. Resultados
  - 3.7.1. Resultados en el ítem 1
    - 3.7.1.1. Análisis global
    - 3.7.1.2. Comparación por grupo
  - 3.7.2. Resultados en el ítem 2
    - 3.7.2.1. Análisis global
    - 3.7.2.2. Comparación por grupo
  - 3.7.3. Resultados en el ítem 3
    - 3.7.3.1. Análisis global
    - 3.7.3.2. Comparación por grupo
  - 3.7.4. Resultados en el ítem 4
    - 3.7.4.1. Análisis global
    - 3.7.4.2. Comparación por grupo
- 3.8. Síntesis de resultados
  - 3.8.1. Comparación de resultados según el tamaño de muestra
  - 3.8.2. Comparación de resultados en experimentos con sucesos equiprobables o no
- 3.9. Análisis de casos atípicos
  - 3.9.1. Estudio de casos con ausencia de variabilidad
  - 3.9.2. Estudio de casos con variabilidad excesiva
- 3.10 Conclusiones del Estudio 1
  - 3.10.1. Conclusiones respecto a los objetivos
  - 3.10.2. Conclusiones respecto a las hipótesis

### 3.1. INTRODUCCIÓN

Este capítulo presenta los resultados del primero de los estudios empíricos realizados como parte de la tesis doctoral, parte de los cuales han sido publicados en Begué (2016), Begué, Batanero y Gea (2018) y Begué, Batanero, Gea y Beltrán-Pellicer (2017). En este estudio se pretende analizar la comprensión de una muestra de 302 estudiantes de segundo

y cuarto curso de educación secundaria obligatoria sobre algunos conceptos elementales asociados al muestreo.

La investigación tiene carácter exploratorio y está basada en el análisis estadístico de las respuestas de una muestra de estudiantas a un cuestionario formado por cuatro ítems de respuesta abierta. A partir de este análisis se obtienen conclusiones sobre la comprensión mostrada por los estudiantas, comparando con los resultados de investigaciones previas descritas en el Capítulo 2.

El capítulo se inicia describiendo los objetivos e hipótesis específicos que han guiado este estudio, así como los contenidos matemáticos que se evalúan a través del cuestionario. También se detalla el contexto de donde se han tomado los datos. A continuación, se resume la metodología utilizada, donde se describe tanto la caracterización de la muestra como la descripción del método de análisis de los datos.

Pasamos a continuación a describir el instrumento diseñado para evaluar algunas de las ideas que se relacionan con el muestreo. El análisis previo de dicho cuestionario permite la caracterización de cada ítem, a través del análisis del razonamiento que se requiere en la determinación de su respuesta correcta asociada. Tras la aplicación del instrumento de evaluación, el capítulo se centra en analizar las respuestas al cuestionario por parte de los estudiantas. En este análisis se señalan los aspectos más relevantes que se observan en cada grupo e ítem, se comparan los grupos y con otras investigaciones previas.

### **3.2. OBJETIVOS E HIPÓTESIS DEL ESTUDIO 1**

Los objetivos específicos de este estudio desarrollan el tercer objetivo general de la investigación, descrito en el Capítulo 1, que establecía la aplicación de una serie de estudios de evaluación acerca de la comprensión del muestreo por parte de los estudiantas. Más concretamente, nos centramos en este estudio en estudiantas de educación secundaria obligatoria y nuestra intención es proponerles tareas de generación de muestras de poblaciones binomiales cuya proporción se conoce o se puede deducir desde los datos dados en el enunciado. De este modo, se puede evaluar la comprensión intuitiva del estudiante de las siguientes propiedades del muestreo:

1. Proporción muestral esperada y relación que establece el estudiante entre dicha proporción muestral y la proporción en la población.
2. Variabilidad de la proporción en el muestreo y efecto del tamaño de la muestra sobre

dicha variabilidad.

De todo lo expuesto se deducen los siguientes objetivos específicos del Estudio 1.

*Objetivo 1.1. Evaluar y analizar la comprensión de la relación entre el valor de una proporción en la población y la proporción muestral esperada en muestras tomadas de dicha población en estudiantes de educación secundaria obligatoria.*

Nos hemos centrado en la estimación de una proporción, porque es un estadístico muy sencillo que aparece en la vida diaria y en los medios de comunicación en muchas situaciones comprensibles para el alumnado. Para la evaluación de los contenidos citados hemos considerado la aplicación de un cuestionario, el cual se describe de manera más detallada en la Sección 3.4. En dicho cuestionario planteamos tareas en las cuales el estudiante ha de dar cuatro valores probables para el número de éxitos de una distribución binomial, cuya proporción se conoce y cuyo valor, así como el tamaño de la muestra, varían sistemáticamente. Puesto que en una población binomial hay una correspondencia biunívoca entre el número de éxitos en  $n$  ensayos y la proporción, trabajamos directamente con el número de éxitos para facilitar la respuesta, pues, de este modo, los estudiantes no necesitan realizar el cálculo de las proporciones, sino que les basta dar un valor entero del número de éxitos.

*Objetivo 1.2. Evaluar y analizar la comprensión de los estudiantes de segundo y cuarto curso de educación secundaria obligatoria de la variabilidad de la proporción muestral y del efecto del tamaño de la muestra sobre dicha variabilidad.*

Cada uno de los ítems del cuestionario pide al estudiante generar cuatro valores probables de la distribución binomial dada en el enunciado. Por tanto, el conjunto de los cuatro valores constituye una muestra de tamaño cuatro y nos permite analizar la variabilidad de las estimaciones del número de éxitos (e indirectamente de la proporción muestral). Además, en los ítems planteados se utilizan distribuciones binomiales formadas por 100 o por 10 elementos. De este modo, se pueden comparar las respuestas de los estudiantes en las tareas, dependiendo del tamaño de la muestra. El cálculo se realiza a partir del número de éxitos dado por los estudiantes, por las mismas razones que en el apartado anterior.

*Objetivo 1.3. Comparar las respuestas proporcionadas por los estudiantes en función de algunas variables de tarea y por curso.*

El mismo instrumento de evaluación nos permite evaluar esta comprensión a través del planteamiento de ítems en los que se demanden la generación de muestras, cuando la población de partida constituye una muestra pequeña o grande. Además, se consideran fenómenos aleatorios caracterizados por sucesos elementos equiprobables y no equiprobables. Estas diferencias nos permiten analizar las respuestas según estas variables. En relación con la muestra participante, se comparan los resultados en función a los dos cursos de estudiantes participantes.

### **Hipótesis del Estudio 1**

Como consecuencia de los objetivos planteados y, teniendo en cuenta los resultados de la investigación previa descritos en el Capítulo 2, es posible establecer algunas hipótesis sobre lo que se espera encontrar como resultado del Estudio 1, que exponemos a continuación.

*H2.1. Algunos estudiantes presentan una comprensión insuficiente de las ideas básicas asociadas al concepto de muestreo evaluadas en el cuestionario.*

Esperamos que la mayoría de estudiantes tenga una buena percepción de la proximidad entre la proporción esperada en la muestra y la proporción en la población. Esta hipótesis se sustenta en los resultados de muchos trabajos anteriores; incluso en los relacionados con la heurística de la representatividad (Tversky y Kahneman, 1982), cuyos resultados muestran una gran precisión en la estimación de los valores esperados. Igualmente, las estimaciones de la media o la proporción son muy precisas en las investigaciones sobre generación de secuencias aleatorias, como las de Green (1989) o Serrano (1996). Por tanto, esperamos que se conserve esta precisión cuando se les pide directamente el valor probable del número de éxitos en la población binomial.

Sin embargo, y, aunque la mayoría de los estudiantes proporcionen respuestas adecuadas, se espera que algunos de ellos muestren el sesgo de equiprobabilidad (Lecoutre, 1992), que consiste en estimar frecuencias esperadas cercanas al 50% independientemente de la proporción en la población, simplemente porque se trata de un fenómeno aleatorio. Serrano (1996) relaciona esta creencia con la acepción de aleatoriedad como equiprobabilidad, que supone la creencia de que los sucesos de

cualquier experimento aleatorio son equiprobables. Chernoff y Russel (2012) la atribuyen a la falacia de la composición, que consiste en transferir a un todo una propiedad que se cumple en una de sus partes; por ejemplo, al ser equiprobables el nacer niño o niña, pensar que es equiprobable cualquier número de niñas (0 a 4) entre cuatro hermanos.

En relación con el análisis de la variabilidad, esperamos más dificultad en la comprensión de la variabilidad muestral y es posible que no se perciba el efecto del tamaño de la muestra en la población binomial de partida sobre la variabilidad, como se describe en sesgos como la insensibilidad al tamaño de la muestra (Tversky y Kahneman, 1974). Estos resultados se observaron en el estudio de Gómez (2014) con futuros profesores y pensamos se repetirán en el nuestro. También se presentan en el estudio de Moreno y Vallecillos (2001) o en la investigación de Shaughnessy et al. (1999).

*H3.1. El análisis de las respuestas del instrumento de evaluación refleja una mejor comprensión y razonamiento sobre los contenidos evaluados en los estudiantes del cuarto curso.*

La muestra está constituida por estudiantes de educación secundaria obligatoria (ESO) de dos cursos diferentes: segundo y cuarto curso de ESO. El cuarto curso está constituido por estudiantes que, en promedio, tienen dos años más que sus compañeros y, además, deberían haber estudiado algunos contenidos más de probabilidad que ellos los dos años anteriores, si consideramos el marco curricular. Por tanto, consideramos relevante para la investigación estudiar si existe una diferencia significativa entre las respuestas que caracterizan a los dos grupos que configuran la muestra. Esperamos que por su mayor conocimiento y madurez, los estudiantes de cuarto tengan mejores resultados. En este sentido, la revisión de los antecedentes recoge investigaciones en las que se identifica una mejora en las respuestas según la edad (Batanero, Serrano y Garfield, 1996; Serrano et al., 1998).

### **3.3. CONTENIDOS SOBRE MUESTREO QUE SE CONSIDERAN EN EL ESTUDIO**

En capítulos anteriores se ha indicado que el concepto de muestreo nos introduce en la inferencia y supone un puente que conecta la estadística y la probabilidad. Según Batanero (2001), el conocimiento científico tiene una naturaleza pragmática, entendiendo que su construcción se debe a los resultados de las experiencias empíricas, las cuales

suelen ser limitadas. Este hecho implica que las conclusiones deben aportar una información más amplia que los datos utilizados u obtenidos a partir de las observaciones. En el caso particular de la estadística, el proceso de muestreo proporciona información sobre la población y, por tanto, aumenta nuestro conocimiento sobre la misma, siendo éste el objetivo principal del trabajo estadístico.

Como se ha señalado, Batanero, et al. (1994) indican que la aprehensión del muestreo implica el equilibrio adecuado entre dos ideas que parecen opuestas: la representatividad muestral y la variabilidad muestral. La primera idea permite señalar que la muestra, si es representativa, tendrá características similares a las de la población. Esto exige que se haya elegido a través de técnicas adecuadas. Por otro lado, la variabilidad muestral implica que no todas las muestras son iguales. Siguiendo a estos autores, el punto adecuado de equilibrio entre los extremos de información total e información nula en relación a una población es complejo y dicho punto depende de la variabilidad de la población y el tamaño de la muestra.

Los ítems que consideramos se centran en la proporción de valores en una muestra que cumplen una cierta condición. Por tanto, intuitivamente se está trabajando con una población con solo dos resultados: éxito y fracaso. En concreto, nos interesamos por la estimación que hacen los estudiantes del valor esperado de éxitos en una muestra de tamaño  $n$ , cuando se conoce la proporción de éxitos  $p$  en la población y también, cómo estiman este valor  $p$  en el caso de que conozca la composición de una muestra de tamaño  $n$ .

Por último, estamos interesados en analizar si los estudiantes comprenden la variabilidad del muestreo y cómo la reproducen cuando se les pide dar la composición de cuatro muestras del mismo tamaño  $n$  en la población de partida. Además, las diferencias en el tamaño de muestra que se demanda permiten centrar el estudio en evaluar de qué manera afecta el tamaño de la muestra en la variabilidad concedida a partir de las muestras dadas por el estudiante.

Estos contenidos son considerados en el currículo, como se presenta en el análisis curricular descrito en el Capítulo 1. De esta forma, reproducimos a continuación en la Tabla 3.3.1, una parte de la Tabla 1.5.3 (objetos matemáticos relacionados con el muestreo en el Decreto de Enseñanzas Mínimas) que recoge el contenido evaluado en el Estudio 1.



Tabla 3.3.1. Objetos matemáticos considerados en el cuestionario de evaluación

	1º	2º	3º	4ºA	4ºB
<b>Situaciones-problema</b>					
Estimar la característica de una población a partir de muestras de una población		x	x	x	x
Valorar la representatividad de muestras				x	x
<b>Lenguaje matemático</b>					
Interpretar con un vocabulario adecuado la información estadística	x	x	x	x	x
<b>Conceptos</b>					
Población	-	x	x	x	x
Individuo					
Muestra	-		x	x	x
Espacio muestral, suceso			x	x	
Variables estadísticas discretas	-		x	x	-
Frecuencia absoluta, relativa	x	x			
Distribución de datos, distribución unidimensional	x	x	x	x	x
Variabilidad muestral				x	x
Representatividad muestral			x	x	x
Parámetros de posición: media, mediana, moda		x	x	x	x
Parámetros de dispersión: rango		x	x	x	x
<b>Procedimientos</b>					
Relacionar las características de una población y una muestra		x	x	x	x
<b>Propiedades</b>					
Aproximación de la frecuencia relativa a la probabilidad	x				

### 3.4. CONTEXTO EDUCATIVO

El contexto educativo de la investigación está centrado en la etapa obligatoria de educación secundaria y, en particular, en los dos cursos considerados (segundo y cuarto). Los estudiantes de estos cursos han seguido los contenidos analizados en el Capítulo 1, relacionados con el muestreo.

Los participantes que configuran la muestra pertenecen a dos institutos (centros públicos) que se localizan en la ciudad de Huesca. Aunque los estudiantes provienen de dos centros distintos, las diferencias debidas a este hecho no son relevantes, puesto que diversos factores como, la situación socio-económica de las familias, el número de alumnado, la ratio de alumnos por clase es muy similar en ambos institutos.

El acceso a los mismos fue facilitado por los directores de los centros y por la Consejería de Educación, a los que se solicitó permiso para realizar el estudio. Igualmente, se pidió la colaboración a los profesores de los grupos, los cuales la prestaron voluntariamente y se interesaron por obtener los resultados finales del estudio. Por lo tanto, agradecemos a todos su cooperación, sin la cual no hubiese sido posible realizar esta investigación.

### **3.5. METODOLOGÍA**

El estudio que se presenta es de evaluación educativa, el cual supone un tema que ha sido una preocupación constante en educación y ha experimentado un auge creciente. En el ámbito educativo debe entenderse la evaluación como actividad crítica, que permite el aprendizaje de investigadores y profesores, debido al conocimiento que proporciona (Arredondo y Bolívar, 2002).

En concreto, se trata de un estudio exploratorio, pues aunque el primer ítem que utilizamos se ha adaptado de otro trabajo, el resto son de elaboración propia. Además, el primer ítem se aplicó a futuros profesores de educación primaria, de mayor edad y formación que nuestros estudiantes, por lo que no tenemos antecedentes de estos ítems con el mismo tipo de estudiantes. La investigación exploratoria tiene como objetivo examinar un problema de investigación poco estudiado, por lo que ayuda a entender fenómenos científicamente desconocidos, poco estudiados o nuevos, apoyando en la identificación de conceptos o variables potenciales, e identificando relaciones posibles entre ellas (Cazau, 2006). Como en nuestro caso, puede partir de algunas hipótesis previas y es flexible en la forma de interpretar los resultados.

En síntesis, el estudio se corresponde con una investigación preferentemente cuantitativa, puesto que se basa en el análisis estadístico de los datos obtenidos a partir de las respuestas dadas por los estudiantes. Se complementa con algunos aspectos cualitativos, porque analizamos algunos tipos de respuestas desde las que se vislumbra la presencia de sesgos de razonamiento. Dicho análisis se realiza a partir de las cuaternas de valores numéricos que constituyen las cuatro muestras dadas por el estudiante y en algunas de ellas se reflejan sesgos, como el de equiprobabilidad cuando el alumno concede valores próximos al 50% siendo los sucesos elementales no equiprobables. Según las clasificaciones de Bisquerra (1989), se trata de una investigación aplicada y descriptiva.

La muestra de estudiantes es intencional, aunque elegida de forma que represente adecuadamente a otros estudiantes del mismo contexto. Este tipo de muestra es habitual en estudios exploratorios, y en general en educación, donde el muestreo probabilístico resulta excesivamente costoso y se utilizan muestras intencionales, siendo conscientes de que no permiten generalizar a otros tipos de estudiantes. No obstante, para aumentar la posibilidad de extender los resultados, se ha utilizado muestras de tamaño alto, constituidas por estudiantes de diferentes grupos y centros de enseñanza.

Según Martínez-Arias, Llorreda y Lloreda (2014), el diseño de instrumentos para la medida de constructos educativos presenta algunos problemas, entre otros, que la aproximación a la medida de un constructo no suele ser única. Además, es difícil identificar el constructo y su definición operativa con independencia del investigador.

Para evitar en lo posible estos problemas, la elección de los contenidos objeto de evaluación se apoya en fundamentos adecuados para definir “el conocimiento sobre el muestreo”, que es un constructo inobservable (León y Montero, 2002). En primer lugar, se ha considerado la revisión de las investigaciones previas porque permite identificar los errores relacionados con nuestro objeto de estudio y, en concreto, aquellos que han sido investigados.

En segundo lugar, una definición más precisa de la variable se basa en el análisis del currículo, descrito en el Capítulo 1. En este sentido, la revisión del currículo permite al investigador conocer el contenido que se le ha presentado al alumno en su etapa educativa. De esta forma, podemos enmarcar qué contenidos relacionados con el objeto de estudio han sido estudiados por el alumno desde un punto de vista teórico. Finalmente, la construcción del cuestionario se apoya en la revisión y recopilación de ítems de investigaciones previas. Entre ellas, se seleccionan aquellas preguntas que pueden ayudar a la consecución de los objetivos descritos en nuestro trabajo, lo cual exige la realización de cambios y adaptaciones para elaborar la versión que finalmente se presenta a la muestra participante.

El análisis de datos es, en general, descriptivo, y se reduce a la elaboración y comparación de gráficos estadísticos y a la elaboración de algunas tablas de frecuencia, junto con la obtención de valores medios y desviaciones típicas de ciertas variables. También se usan algunos contrastes de hipótesis, siendo conscientes de las posibilidades limitadas de generalización relacionadas con la selección de la muestra. Tras la aplicación del cuestionario, se codificaron previamente los datos y se utilizó el paquete estadístico SPSS para su análisis estadístico.

### **3.5.1. DESCRIPCIÓN DE LA MUESTRA**

La muestra está constituida por un total de 302 estudiantes, de los cuales 157 son de segundo curso de la ESO (12-13 años) y 145 de cuarto curso (15-16 años). El número de grupos de estudiantes que participaron fue nueve grupos de segundo curso de la ESO y ocho grupos de cuarto, de los cuales seis grupos corresponden a la opción B. En concreto,

el cuestionario es respondido por 108 estudiantes que cursan la opción B (74,5% de los estudiantes de cuarto curso).

Tabla 3.5.1. Características de la muestra participante en el Estudio 1

Centro educativo	Materia	Tamaño de la muestra
C1	2ºESO	100
	4ºESO Opción B	68
	4ºESO Opción A	19
C2	2ºESO	57
	4ºESO Opción B	40
	4ºESO Opción A	18

Siguiendo las directrices curriculares, se considera que la opción A está enfocada para aquellos estudiantes que quieren realizar una formación profesional, aunque no es restrictiva su decisión para continuar con los estudios de Bachillerato. Mientras que la opción B está dirigida a los estudiantes que pretenden continuar con los estudios de Bachillerato. En relación con el currículo para la opción B, observamos que la diferencia con respecto a la primera opción supone la consideración de algunos contenidos nuevos.

Durante la aplicación del cuestionario se le pregunta al profesor encargado del aula sobre los conocimientos previos que sus alumnos pueden tener sobre el contenido del bloque de estadística y probabilidad. En relación con los estudiantes de segundo curso de la ESO, no habían recibido instrucción previa sobre contenidos de probabilidad y estadística, aparte de los conocimientos que pudieran haber obtenido en la educación primaria, donde en el último curso hay algunas ideas intuitivas sobre sucesos aleatorios y probabilidad simple, pero no se estudia explícitamente el muestreo.

Por su parte, algunos de los estudiantes de los grupos de cuarto curso de la ESO han recibido instrucción sobre probabilidad, desde el significado clásico de la misma o nociones de estadística descriptiva en los cursos anteriores (ver contenidos en el análisis curricular), independiente de la modalidad que estaban cursando en el momento de la evaluación. En otros grupos el profesor indica una ausencia de enseñanza de dichos contenidos en los cursos anteriores o, al menos en el curso anterior.

El principal factor que conduce a no enseñar dichos contenidos es el tiempo, que imposibilita a los profesores poder llevar al aula todos los contenidos que propone el currículo. Además, la probabilidad y la combinatoria están presentadas al final del temario, lo cual conduce a que dicho contenido no sea enseñado.

### **3.5.2. MÉTODO DE RECOGIDA DE DATOS**

Los datos fueron recogidos en todos los grupos participantes en los institutos a los que asisten regularmente y en las horas en las que los estudiantes tenían clase de matemáticas. El cuestionario se presenta como una actividad formativa de la citada clase. Además, se les explicó a los estudiantes el fin de la evaluación y se les pidió su colaboración en el proceso. Se leyeron los enunciados y se resolvieron posibles dudas sobre la tarea a completar.

Los estudiantes completaron el cuestionario por escrito en un intervalo de tiempo de 15-25 minutos. Durante la aplicación del cuestionario, el docente responsable del grupo aula estaba presente, aunque situado en un segundo plano. Es decir, el investigador es la persona que describe las indicaciones y finalidad del cuestionario. Adicionalmente, se observa que los estudiantes respondieron con interés, completando en su mayoría todas las preguntas. La recogida de datos se realizó por escrito, utilizando un formulario con el cuestionario preparado para la investigación.



### **3.6. ANÁLISIS A PRIORI DEL CUESTIONARIO Y CONOCIMIENTOS EVALUADOS**

Para la construcción del cuestionario, la primera tarea es una adaptación de otra utilizada en el trabajo de investigación de Green (1983a), que fue diseñada para evaluar la comprensión de la estimación frecuencial de la probabilidad, que, como se ha indicado, está relacionada con la comprensión del muestreo. En la versión utilizada por Green se pide a los estudiantes una única estimación. Dicho ítem fue adaptado por Gómez et al. (2014), quien pidió a los estudiantes generar cuatro respuestas diferentes al ítem, para poder obtener información sobre la comprensión de la variabilidad en el muestreo. En nuestro trabajo se ha utilizado directamente el ítem propuesto por Gómez et al. (2014), aunque el método de análisis es diferente y más completo. El resto de las tareas que constituyen el cuestionario son variantes del primer ítem. En todas ellas se demanda al estudiante escribir cuatro muestras asociadas a un experimento concreto descrito en el enunciado.

El cuestionario está constituido por cuatro ítems que tienen como finalidad evaluar tanto la comprensión de la variabilidad del muestreo como el efecto del tamaño de la muestra sobre dicha variabilidad. Además, esta herramienta permite observar si el estudiante muestra la relación existente entre el valor de una proporción en la población

y la frecuencia relativa esperada en una muestra. A continuación, se describe detalladamente cada una de las tareas del mismo, razonando sobre su solución correcta, la cual nos permite caracterizar si el estudiante presenta una comprensión adecuada de los conceptos evaluados.

### 3.6.1. ANÁLISIS DEL ÍTEM 1

**Ítem 1.** Un profesor vacía sobre la mesa un paquete de 100 chinchetas obteniendo los siguientes resultados: 68 caen con la punta para arriba  y 32 caen hacia abajo . Supongamos que el profesor pide a 4 niños repetir el experimento, lanzando las 100 chinchetas. Cada niño vacía una caja de 100 chinchetas y obtendrá algunas con la punta hacia arriba y otras con la punta hacia abajo. Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Daniel	Martín	Diana	María
Punta arriba:	Punta arriba:	Punta arriba:	Punta arriba:
Punta abajo:	Punta abajo:	Punta abajo:	Punta abajo:

Este ítem se ha tomado de la investigación de Gómez et al. (2014). En nuestro estudio, se ha conservado el valor  $n = 100$  del ítem usado por Green (1983a) y adaptado por Gómez et al. (2014). El fenómeno aleatorio que se describe en el enunciado es el lanzamiento de chinchetas y el número de chinchetas que caen con la punta hacia arriba se corresponde con una variable aleatoria que se distribuye según un modelo binomial,  $B(n, p)$ , donde  $n$  es el tamaño de la muestra. Entonces, definimos la variable  $X$  como el número de chinchetas que caen hacia arriba, donde el tamaño de la muestra,  $n$ , es 100 y,  $p$  es la probabilidad de que una chincheta caiga con la punta hacia arriba, siendo esta probabilidad desconocida.

Esta probabilidad se puede estimar a partir de los datos dados en el enunciado, los cuales permiten también obtener una estimación tanto del valor esperado,  $np$ , así como de la desviación típica,  $\sqrt{np(1-p)}$ . Para este caso, la inferencia estadística indica que la mejor estimación de la proporción en la población es la proporción muestral dada en el enunciado,  $\hat{p} = 0,68$ , por sus propiedades de insesgadez y mínima varianza. A partir de ella, se puede también estimar el número esperado de chinchetas que caerán con la punta hacia arriba en la muestra de 100 elementos,  $n\hat{p} = 68$  y su desviación típica,  $\sqrt{n\hat{p}(1-\hat{p})} = 4,66$ .

Por otro lado, en las cuatro estimaciones proporcionadas por el estudiante se tiene

una media muestral igual a  $n\hat{p} = 68$ , y su desviación típica es igual a  $\frac{\sigma}{\sqrt{4}}$ , siendo  $\sigma$  la desviación típica de la población, luego la desviación típica de la media muestral de las cuatro estimaciones proporcionadas por el estudiante, que se refieren a elementos de la distribución binomial  $B(100, \hat{p} = 0,68)$ , es igual a  $4,66/2=2,33$ .

En consecuencia, se considera que el sujeto presenta una buena comprensión del valor esperado, si el valor medio de las cuatro respuestas que proporciona es cercano a 68. Se considerarán cercanas aquellas respuestas cuyos valores, para el número de chinchetas que caen con la punta hacia arriba, se localicen en el intervalo  $[65,6 - 70,3]$  que se calcula al determinar la media más/menos la desviación típica. Este es el intervalo de valores que contiene el 68% de las observaciones en la distribución muestral de la media, que es aproximadamente normal. En los ítems 1 y 2 del cuestionario, la aproximación normal de la distribución binomial descrita en el ítem se justifica por el valor del tamaño de la muestra  $n=100$ , debido al teorema central del límite.

Puesto que la suma de valores de una distribución normal es también normal, en este ítem, se justifica la aproximación normal a la distribución muestral de la media de cuatro valores, que se obtiene a partir de una suma dividida por una constante. Utilizando dicha aproximación normal de la distribución de la media muestral, se obtiene que en el intervalo  $[63,3-72,7]$  se sitúan el 95% de los valores centrales, por lo que se considera como una respuesta aceptable aquellas medias muestrales que se localizan en dicho intervalo.

Observamos en la Figura 3.6.1 la distribución muestral empírica de las medias de cuatro valores de la distribución binomial  $B(100, 0,68)$  obtenida con 10000 simulaciones mediante el programa Fathom. Se han representado los percentiles del 2,5%, 97,5%, 16% y 84%, es decir, los intervalos que contienen el 95% y 68% de casos centrales en la distribución. Se puede apreciar que los resultados coinciden con los obtenidos de la distribución normal casi por completo. Puesto que la simulación siempre es aproximada y la diferencia es mínima, se decide utilizar los valores obtenidos directamente con la aproximación normal, lo que se hará en el resto de ítems.

Finalmente, otro aspecto que se debe reconocer es la ausencia de equiprobabilidad por la asimetría física del dispositivo.

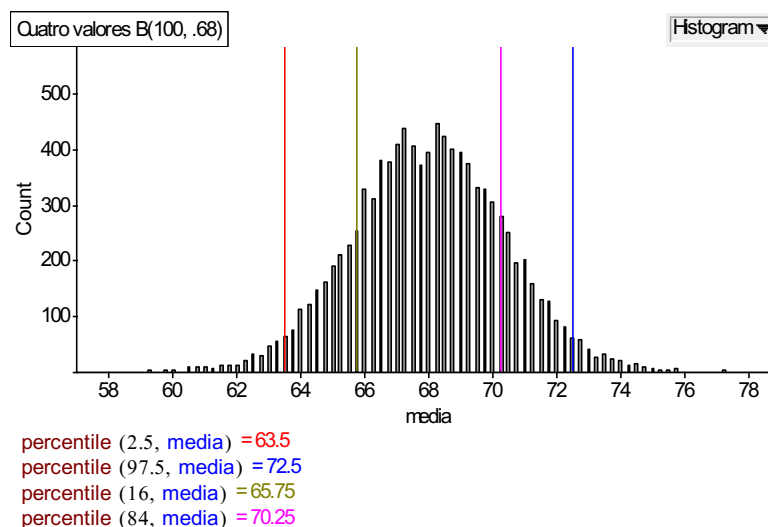


Figura 3.6.1. Distribución muestral empírica de las medias en muestras de 4 elementos de la B(100, 0,68)

Para analizar mejor las respuestas de los estudiantes en todos los ítems, se han clasificado en categorías que tienen en cuenta los valores de la media de las cuatro estimaciones proporcionadas por el estudiante. Dichos intervalos se calculan teniendo en cuenta la distribución muestral de la media de una muestra de cuatro valores del fenómeno aleatorio estudiado y también los posibles sesgos asociados a este ítem, identificados en investigaciones previas. La clasificación utilizada es la siguiente:

- *Estimación normativa del valor esperado.* Cuando el valor medio de las cuatro estimaciones proporcionadas por el estudiante en el ítem es muy próxima al valor teórico (perteneciente al intervalo central de la distribución muestral que teóricamente contendría al 68% de los valores de las medias muestrales). Los estudiantes que dan estas respuestas comprenden la propiedad de representatividad muestral.
- *Estimación aceptable del valor esperado.* Si el valor medio de las cuatro estimaciones dadas por el estudiante se encuentra fuera del intervalo central de la distribución muestral que contendría al 68% de los valores de las medias muestrales, pero dentro del intervalo que contendría al 95% de dichos valores.
- *Recencia negativa.* En los experimentos descritos en los ítems 1 y 4, el enunciado indica que se obtuvo una frecuencia mayor al 50% de uno de los sucesos. Algunos estudiantes tratan de compensar este resultado, dando estimaciones cuyo valor medio favorece al suceso contrario (como ejemplo, en el ítem 1, dan muestras con un número de chinchetas hacia arriba igual a 38 o 40). Esta tendencia muestra una recencia negativa, conducta que se ha explicado mediante la heurística de la representatividad,



e implica una pobre comprensión de la independencia de los ensayos en un experimento aleatorio (Batanero y Serrano, 1999; Fischbein y Snarch, 1991).

- *Recencia positiva.* Ante un enunciado en el que la frecuencia de los dos sucesos implicados es diferente, se tiende a proporcionar estimaciones en las que se exagera tal diferencia. Esta conducta también es explicada por la heurística de la representatividad. En el caso del ítem 1, sería, por ejemplo, proporcionar muestras con 90 o incluso con 100 chinchetas hacia arriba. Estos casos extremos también fueron informados por Fischbein y Snarch (1991) y Shaughnessy et al. (2004), aunque los últimos autores no los relacionan con los sesgos de recencia positiva o negativa.
- *Sesgo de equiprobabilidad.* Los ítems 1 y 4 describen experimentos con sucesos elementales no equiprobables. Sin embargo, algunas respuestas reflejan estimaciones cuyos valores medios se localizan muy próximos al 50% del tamaño de la muestra. En este caso, el sujeto no considera la información frecuencial del enunciado, por lo que presenta el sesgo de equiprobabilidad (Lecoutre, 1992), el cual se atribuye a la heurística de disponibilidad. Dentro de las respuestas en las que se identifica el sesgo de equiprobabilidad, algunas se caracterizan por la ausencia de variabilidad, por ejemplo, dando cuaternas de la forma (50, 50, 50, 50). Estas respuestas aparecen en el trabajo de Gómez et al. (2014).
- *No completa.* Consideramos en esta categoría a aquellos estudiantes que no responden a un ítem.

Para analizar la variabilidad de las respuestas de los estudiantes, se estudia el rango obtenido a partir de los cuatro valores dados. En este caso, la distribución muestral de estos rangos en muestras de cuatro elementos no es normal ni tampoco aproximadamente y tiene una expresión compleja. Por ello, se decidió directamente simularla mediante el programa Fathom. En la Figura 3.6.2 se muestra la simulación obtenida con 10000 muestras aleatorias de cuatro elementos de la distribución binomial  $B(100, 0,68)$ . A partir de esta distribución muestral empírica hemos calculado diferentes percentiles, por lo que observamos que el 68% de los valores centrales de los rangos en la distribución caerían entre 5 y 14 y el 95% entre 3 y 18.

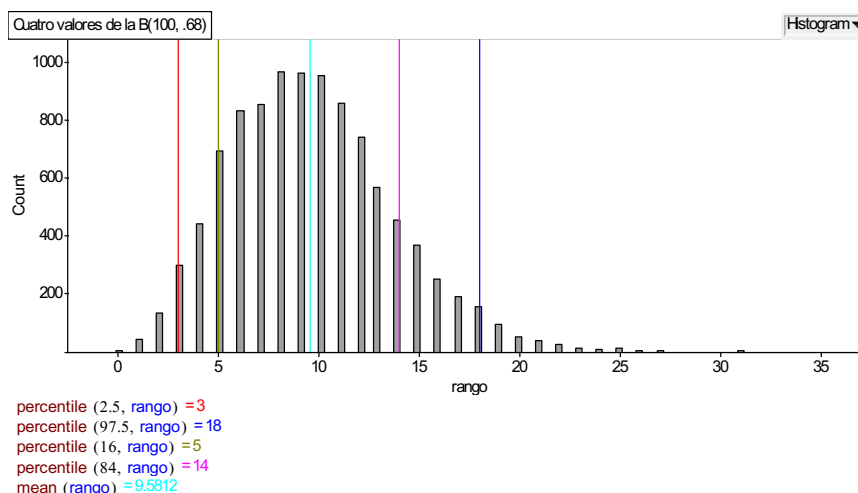


Figura 3.6.2. Distribución muestral empírica de los rangos en muestras de 4 elementos de la B(100, 0,68)

Por tanto, los valores de las chinchetas que caen con la punta hacia arriba deberían presentar una variabilidad óptima o normativa si el rango asociado a los cuatro valores proporcionados da una amplitud entre 5 y 14 (68% de los casos) y aceptable si el rango cae en el intervalo del 95%, esto es entre 3 y 18. Se ha modificado el análisis respecto al utilizado por Gómez et al. (2014), que comparan los rangos no con la distribución muestral de rangos sino con los rangos de la población de partida. Nuestra propuesta permite comparar directamente los rangos de los cuatro valores dados por el estudiante con la distribución muestral de todos los posibles rangos en una muestra de cuatro valores de la distribución B(100, 0,68).

En consecuencia, para analizar mejor la percepción de la variabilidad, por parte de los estudiantes, sus respuestas se han clasificado en las siguientes categorías:

- *Estimación normativa de la variabilidad muestral.* Es la respuesta óptima; la da el estudiante que ofrece cuatro estimaciones cuyo rango se encuentra entre los valores 5 y 14. Dicho rango corresponde a los valores más probables de los rangos de cuatro valores en la distribución binomial considerada en el enunciado.
- *Estimación aceptable de la variabilidad muestral.* El estudiante da cuatro estimaciones cuyo rango se encuentra entre 3 y 18 y fuera del rango considerado como normativo del apartado anterior. Estos valores tienen menor probabilidad en la distribución, pero no serían extremadamente raros.
- *Estimación excesiva de la variabilidad muestral.* El estudiante proporciona cuatro estimaciones cuyo rango es mayor o igual a 19. El estudiante no percibe que la variabilidad de la distribución muestral es menor que la de la población.

- *Alta concentración.* Cuando el estudiante proporciona cuatro estimaciones con el mismo valor o muy parecido, en concreto, con un rango menor o igual que 3. En este caso, subyace una concepción determinista del muestreo, esperando la replicación de los resultados en las diferentes muestras. No se muestra comprensión intuitiva de la variabilidad del muestreo.

### 3.6.2. ANÁLISIS DEL ÍTEM 2

**Ítem 2.** Un profesor vacía sobre la mesa un paquete de 100 monedas obteniendo los siguientes resultados: 53 caen con la cara hacia arriba y 47 caen con la cruz hacia arriba. Supongamos que el profesor pide a 4 niños repetir el experimento. Cada niño lanza las 100 monedas y obtendrá algunas con la cara hacia arriba y otras con cruz hacia arriba. Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Elena	Clara	Matías	Rosa
Cara:	Cara:	Cara:	Cara:
Cruz:	Cruz:	Cruz:	Cruz:

El fenómeno aleatorio descrito en este ítem se corresponde con la repetición del lanzamiento de una moneda, donde el número de ensayos es 100. Entonces, la variable aleatoria definida como el número de monedas que caen con la cara hacia arriba sigue una distribución binomial, donde  $n$  toma el valor 100. Aplicando el significado clásico de la probabilidad, se tiene que los dos resultados del experimento (obtener cara o cruz) son equiprobables, es decir, la probabilidad de que salga cara es  $\frac{1}{2}$ .

Por un lado, se realiza el análisis de la distribución, para el estudio del valor esperado medio de las muestras dadas por los alumnos. En este caso, tras la caracterización de la distribución que sigue la variable aleatoria, el cálculo de la desviación típica de la distribución muestral es  $\sigma = 2,5$ . En este caso, como el anterior, se puede aplicar la aproximación normal a la distribución binomial  $B(100, 0,5)$  y en consecuencia a la media de cuatro valores aleatorios de la misma. Además, se comprobó mediante simulación la coincidencia casi exacta de los correspondientes percentiles (Figura 3.6.3).

Por tanto, se decide aplicar la aproximación normal directamente y se consideran como correctas aquellas respuestas cuyo promedio asociado a los cuatro valores para el número de caras se acerque a 50 y oscile dentro del intervalo  $[47,5-52,5]$ . Las respuestas denominadas como aceptables se asocian con el intervalo que recoge el 95% de los casos que se corresponde con  $[45-55]$  y se encuentren fuera del intervalo considerado como correcto.

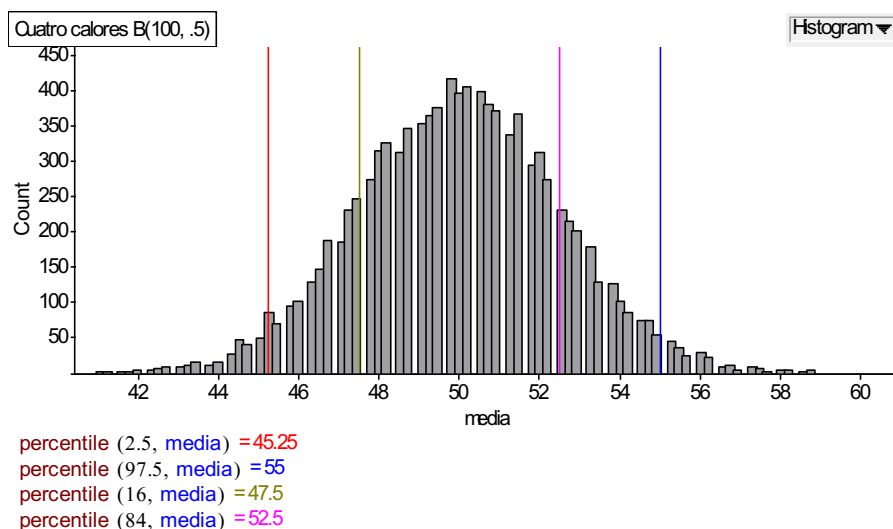


Figura 3.6.3. Distribución muestral empírica de las medias en muestras de 4 elementos de la B(100, 0,5)

Por otro lado, se considera el análisis de la comprensión de la variabilidad, fundamentado en el estudio del rango de los valores dados por los alumnos. Para ello, se obtienen los intervalos del rango [6-15] que incluyen el 68% y [3-20] el 95% de los casos, considerando los primeros como variabilidad normativa y los segundos como aceptable. Si el rango obtenido es mayor a 20, se considera que hay una variabilidad excesiva y alta concentración en caso inferior a 3. En ambos casos, se identifica que el sujeto no presenta una comprensión adecuada de la variabilidad muestral. En la Figura 3.6.2 se presenta la distribución empírica muestral de los rangos correspondiente al ítem 2, obtenida mediante 10000 simulaciones. Observamos que la distribución del rango no es simétrica, siendo infrecuentes los valores altos del rango.

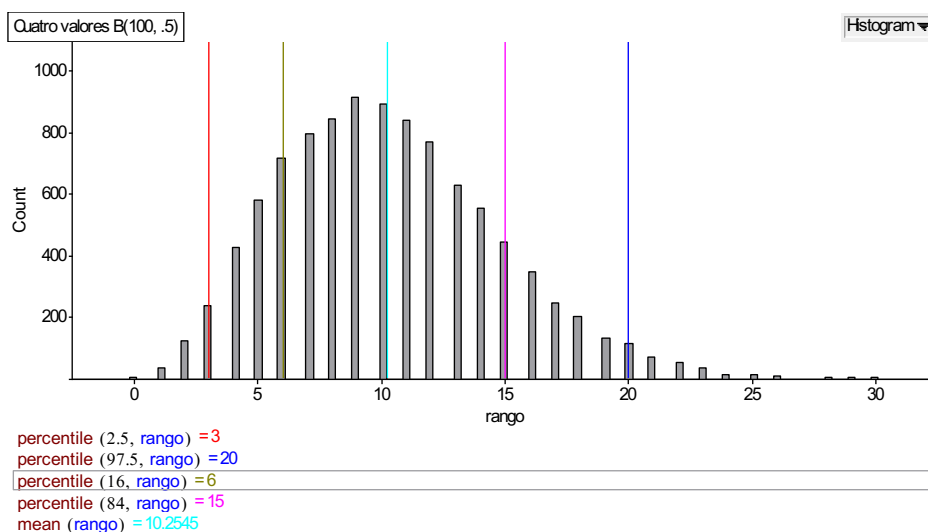


Figura 3.6.4. Distribución muestral empírica de los rangos en muestras de 4 elementos de la B(100, 0,5)

### 3.6.3. ANÁLISIS DEL ÍTEM 3

**Ítem 3.** Un profesor pide a 4 niños lanzar 10 monedas sobre la mesa y contar el número de caras y cruces obtenidos. Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Silvia	Javier	Miguel	Carmen
Cara:	Cara:	Cara:	Cara:
Cruz:	Cruz:	Cruz:	Cruz:

El fenómeno aleatorio que describe este enunciado es idéntico al presentado en el ítem anterior. La diferencia radica en que se realizan 10 ensayos del experimento aleatorio, en lugar de 100. Entonces, la variable aleatoria número de caras sigue una distribución binomial,  $B(10, 1/2)$ .

En primer lugar, se realiza el estudio del valor medio del número de caras de los cuatro valores dados por el estudiante, utilizando la distribución del estadístico de la media muestral, siguiendo el mismo procedimiento que el explicado para el primer ítem. Aunque se trata de una muestra pequeña ( $n=10$ ), la aproximación normal a la distribución normal se puede utilizar siempre que o bien  $n>30$  o que se cumpla la condición  $np$  y  $nq$  sean iguales o mayores que 5, condición que se cumple en este caso y autoriza a suponer una distribución aproximadamente normal para la media de la muestra. No, obstante, como en casos anteriores, se ha analizado la distribución muestral empírica obtenida por simulación, obteniendo como en los otros ítems una buena aproximación.

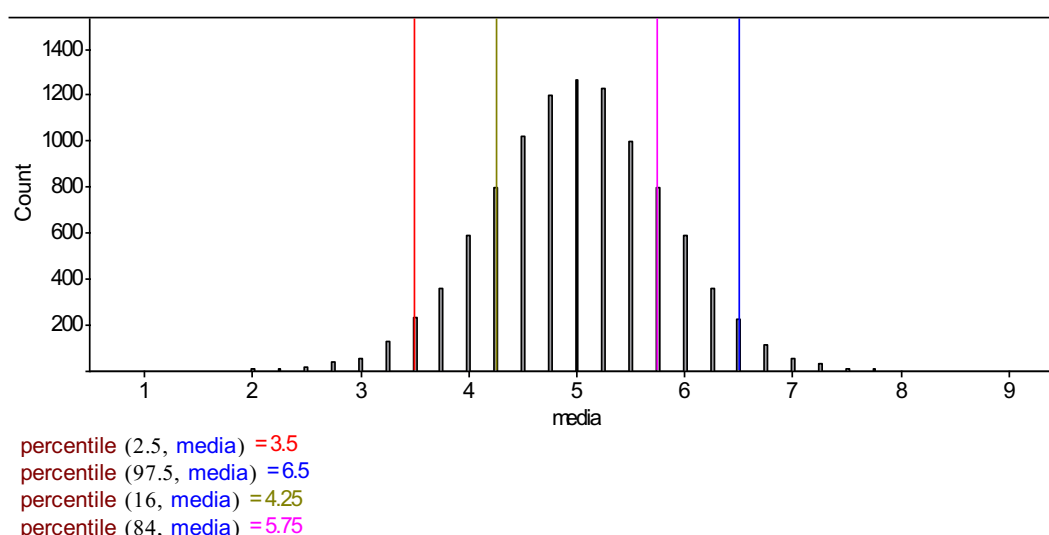


Figura 3.6.5. Distribución muestral empírica de las medias en muestras de 4 elementos de la  $B(10, 0,5)$

En este caso, se considera que una respuesta es óptima, si la media de los cuatro valores toma un valor en torno al valor medio esperado y oscila en el intervalo  $[4,2-5,8]$

y aceptable cuando el valor medio se localiza fuera del anterior rango de valores y dentro del intervalo [3,4-6,6].

Por otro lado, se simuló de nuevo la distribución muestral del rango de cuatro valores de la distribución  $B(10, 0,5)$  y se determinaron los intervalos centrales (Figura 3.6.6), luego se considera normativa la variabilidad cuando el rango se encuentra entre 2 y 5 (68% de los rangos en la distribución muestral) y aceptable si oscila fuera del anterior rango de valores, pero entre 1 y 6. Si es mayor que 6 la variabilidad sería excesiva. Si es menor que 1 se considera demasiada concentración. Estos dos últimos casos implican que no se comprende la variabilidad muestral.

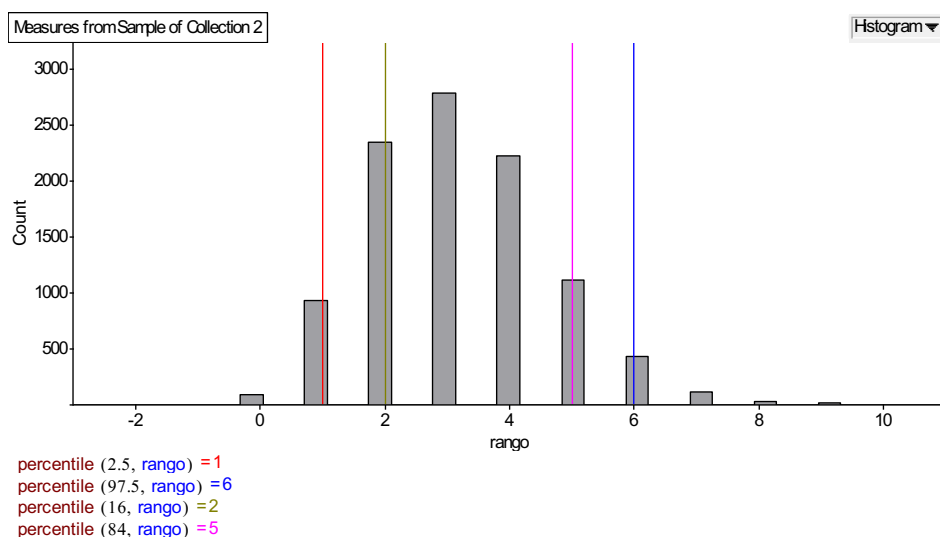


Figura 3.6.6. Distribución muestral empírica de los rangos en muestras de 4 elementos de la  $B(10, 0,5)$

Como se ha citado, el fenómeno aleatorio es análogo al ítem anterior, la diferencia se debe al número de ensayos para los que se repite el experimento. Este hecho permite analizar el efecto del tamaño de la muestra en la variabilidad de los resultados probables. Por tanto, el análisis comparativo de los dos ítems permite estudiar si existe una adecuada comprensión (al menos intuitiva) de la Ley de los Grandes Números.

### 3.6.4. ANÁLISIS DEL ÍTEM 4

**Ítem 4.** Un jugador de baloncesto suele encestar 70 de cada 100 tiros a una canasta desde la posición de tiros libres. Escribe en la siguiente tabla un resultado que sea probable para cuatro partidos en los que lanza 10 tiros desde el punto de lanzamientos personales.

Partido 1 (10 tiros)	Partido 2 (10 tiros)	Partido 3 (10 tiros)	Partido 4 (10 tiros)
Número de encestes:	Número de encestes	Número de encestes	Número de encestes
Número de fallos:	Número de fallos:	Número de fallos:	Número de fallos:

La tarea describe la frecuencia de encestes de un jugador de baloncesto. La elección de esta situación se debe a la familiaridad del contexto, lo cual puede ayudar al estudiante en la resolución del ítem. Siguiendo un razonamiento similar a los ítems anteriores, la variable aleatoria queda definida por el número de encestes, la cual sigue una distribución binomial, donde  $p$  es desconocida, pero se puede estimar por la proporción muestral dada en el enunciado,  $B(10, 0,7)$ . En este caso, la aproximación normal a la distribución de partida es algo forzada porque, aunque  $np=7$  y mayor que 5,  $nq=3$ . En consecuencia, se ha simulado la distribución de la media muestral, que se presenta en la Figura 3.6.7.

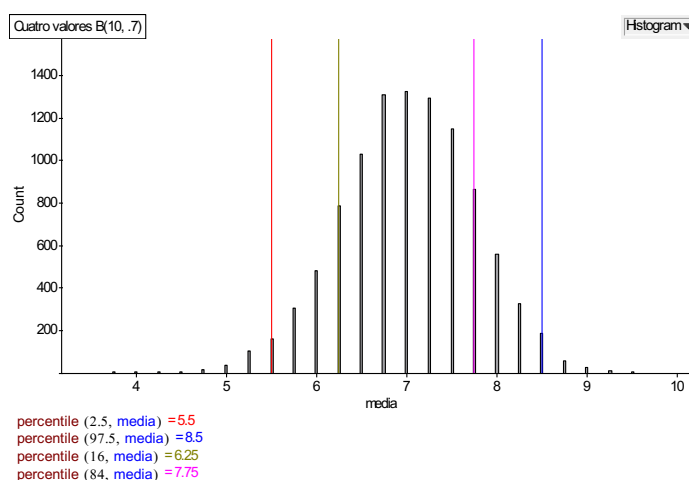


Figura 3.6.7. Distribución muestral empírica de las medias en muestras de 4 elementos de la  $B(10, 0,7)$

Si consideramos el dato del enunciado, en el caso de utilizar la distribución normal, se estima que el valor esperado es 7, y su desviación estándar es 1,45. Por tanto, el valor esperado estimado asociado a los cuatro valores escritos para el número de encestes puede oscilar en el intervalo  $[6,3-7,7]$ . Se considera aceptable el rango de valores fuera del anterior intervalo, pero entre 5,6 y 8,4. Vemos que los valores son prácticamente idénticos a los obtenidos mediante simulación, por lo que se decide utilizar los teóricos.

Se considera adecuada la variabilidad cuando el rango se encuentra entre 2 y 4 (68% de los casos). Si está entre 1 y 6 (95% de los rangos) aunque no entre 2 y 4 es aceptable. Si es mayor a 6 se considera excesiva y si es menor que 1 demasiada concentración, por lo que se deduce que no se comprende la variabilidad muestral.

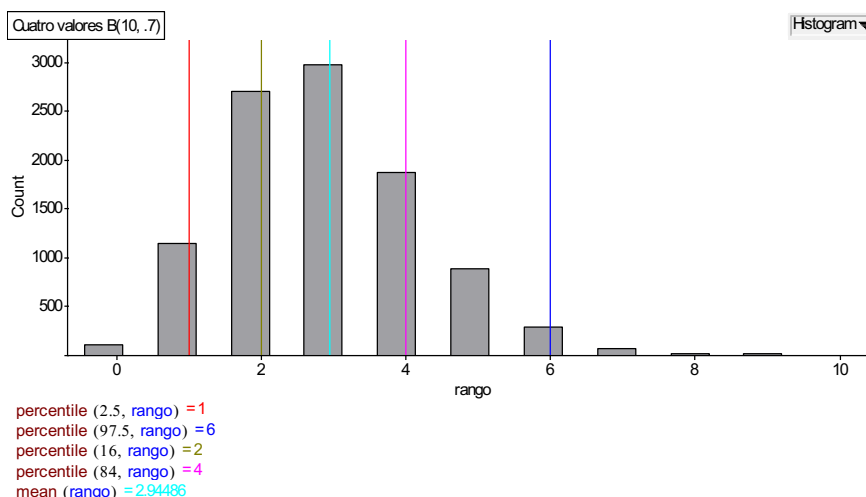


Figura 3.6.8. Distribución muestral empírica de los rangos en muestras de 4 elementos de la B(10, 0,7)

Tabla 3.6.1. Características de los ítems del cuestionario

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Tamaño de la muestra	100	100	10	10
Probabilidad del suceso de interés	0,68	0,5	0,5	0,7
Estimación de la probabilidad	Frecuencial	Clásica	Clásica	Frecuencial
Número esperado de éxitos	68	50	5	7
Desviación típica	$\sigma = 4,66$	$\sigma = 5$	$\sigma = 1,58$	$\sigma = 1,45$
Intervalo que contiene el 68% de medias muestrales	[65,7; 70,3]	[47,5-52,5]	[4,2-5,8]	[6,3-7,7]
Intervalo que contiene el 95 % de medias muestrales	[63,3;72,7]	[45-55]	[3,4-6,6]	[5,5-8,4]
Rango normativo	[5-14]	[6-15]	[2-5]	[2-4]
Rango aceptable	[3-18]	[3-20]	[1-6]	[1-6]

En resumen, el instrumento de evaluación está constituido por cuatro ítems. La Tabla 3.6.1 sintetiza las características de los ítems, donde los dos primeros enunciados demandan cuatro valores del número de éxitos en muestras grandes y los otros dos, en muestras pequeñas (3 y 4). Otro aspecto a destacar es que en el primer y último ítem se exige la aplicación del enfoque frecuencial de la probabilidad para caracterizar la probabilidad de éxito a partir de los datos dados en el enunciado. Por el contrario, los otros dos se apoyan en el significado clásico de la probabilidad a través de la regla de Laplace. Además, en la Tabla 3.6.1 incluimos el valor esperado y la desviación típica de la distribución binomial que modeliza la situación descrita en cada ítem, los intervalos que contienen el 68% y 95% de las medias muestrales y los valores del rango, considerados como normativo y aceptable.



### 3.7. RESULTADOS

Los siguientes puntos muestran los resultados de las respuestas de los estudiantes a cada uno de los ítems que constituyen el cuestionario. El estudio se centra, en primer lugar, en la muestra total de estudiantes que responden el cuestionario. Tras este análisis previo, se repite el análisis separando los grupos para identificar las posibles diferencias y estudiar si existe un progreso con la edad y/o enseñanza recibida de la probabilidad.

En cada ítem se analiza inicialmente la distribución asociada al valor medio de los cuatro ensayos dados por el estudiante. Este estudio permite observar si las muestras proporcionadas tienen una frecuencia media del suceso de interés normativa, en relación con el valor que toma la proporción en la distribución binomial que modeliza la situación descrita en cada ítem, para lo cual se considera el histograma de frecuencias o diagrama de barras asociado a dichas medias, que permite comprobar si toman el valor considerado como óptimo o al menos aceptable.

En la Tabla 3.7.1 se presentan las medias teóricas (valor esperado en la distribución binomial que modela el enunciado de cada ítem) y las obtenidas en la muestra total de estudiantes, donde vemos que, en general, la estimación es buena, pues los valores teóricos y observados son muy parecidos, excepto en el ítem 1.

Tabla 3.7.1. Media teórica y observada de las distribuciones del valor medio de las cuatro respuestas proporcionadas por cada estudiante a cada ítem, en el global de la muestra

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Media teórica	68	50	5	7
Media observada	57,9	51,2	5,1	6,6

En segundo lugar, se analiza la variabilidad esperada por los estudiantes durante el proceso de muestreo, mediante el estudio del rango asociado a los cuatro valores proporcionados, que se ha definido en el apartado anterior. Las medias teóricas y observadas de la distribución de los rangos se presentan en la Tabla 3.7.2. En este caso, hay gran diferencia en los dos primeros ítems, donde los estudiantes dan valores con un rango mucho mayor que el teórico. Además de los diagramas de barras o histogramas asociados a las distribuciones analizadas, se considera, el diagrama de caja asociado al valor medio y al rango, respectivamente. Esta representación permite observar la mediana y los cuartiles, así como la simetría de la distribución. Adicionalmente, este gráfico señala la presencia de valores atípicos, hecho que favorece la identificación de respuestas asociadas a un sesgo particular. De igual modo, el estudio se realiza considerando toda la muestra participante y, posteriormente, se realiza un análisis comparativo en relación al

curso.

Tabla 3.7.2. Media teórica y observada de las distribuciones del rango de las cuatro respuestas proporcionadas por cada estudiante a cada ítem, en el global de la muestra

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Rango teórico medio	9,6	10,3	3,2	2,9
Rango observado medio	23,6	22,1	3,5	2,9

### 3.7.1. RESULTADOS EN EL ÍTEM 1

En este ítem, la única información que se proporciona al estudiante es el número de chinchetas que el profesor obtuvo hacia arriba cuando realiza el experimento. Por tanto, los estudiantes pueden utilizar la estimación frecuencial de la probabilidad para estimar la proporción en la población, o bien razonar desde su experiencia, que es más probable que la chincheta caiga hacia arriba debido al peso de la cabeza.

#### 3.7.1.1. ANÁLISIS GLOBAL

En relación con el análisis global de la precisión de la estimación del número esperado de chinchetas que caen con la punta hacia arriba, se considera la distribución del valor medio de las cuatro respuestas proporcionadas por el estudiante, que se presenta en las Figuras 3.7.1.a y 3.7.1.c.

Además, en la Tabla 3.7.1 se muestra que el valor medio de la muestra participante se corresponde con 57,9, el cual coincide con el obtenido en el trabajo de Gómez et al. (2014). Este valor medio resulta aproximadamente 10 unidades menor que el valor medio esperado, que es 68 (línea vertical señalada la Figura 3.7.1.a). En conclusión, la muestra participante presenta una estimación del número medio de chinchetas que caen hacia arriba sesgada hacia el valor 50, debido a que algunos estudiantes no tienen en cuenta esta probabilidad frecuencial y se guían para la elaboración de su respuesta en el enfoque clásico de la probabilidad.

Además, se observan dos picos en la distribución global de la media (Figura 3.7.1.a). Por un lado, el 24,5% de la muestra proporciona valores medios que se localizan en el intervalo [45-55], es decir, centrado en el valor  $p=0,5$ . Por tanto, estos estudiantes presentan el sesgo de equiprobabilidad (Lecoutre, 1992), que supone la consideración de los resultados de un experimento aleatorio como equiprobables. Sin embargo, se puede deducir que este experimento no cumple la equiprobabilidad a partir del dato dado en el enunciado, sobre la proporción de chinchetas que caen de cierta forma en una muestra. Así mismo, el alumno puede identificar la asimetría del dispositivo y, por tanto, la no

equiprobabilidad de los resultados posibles a partir de un análisis físico del objeto.

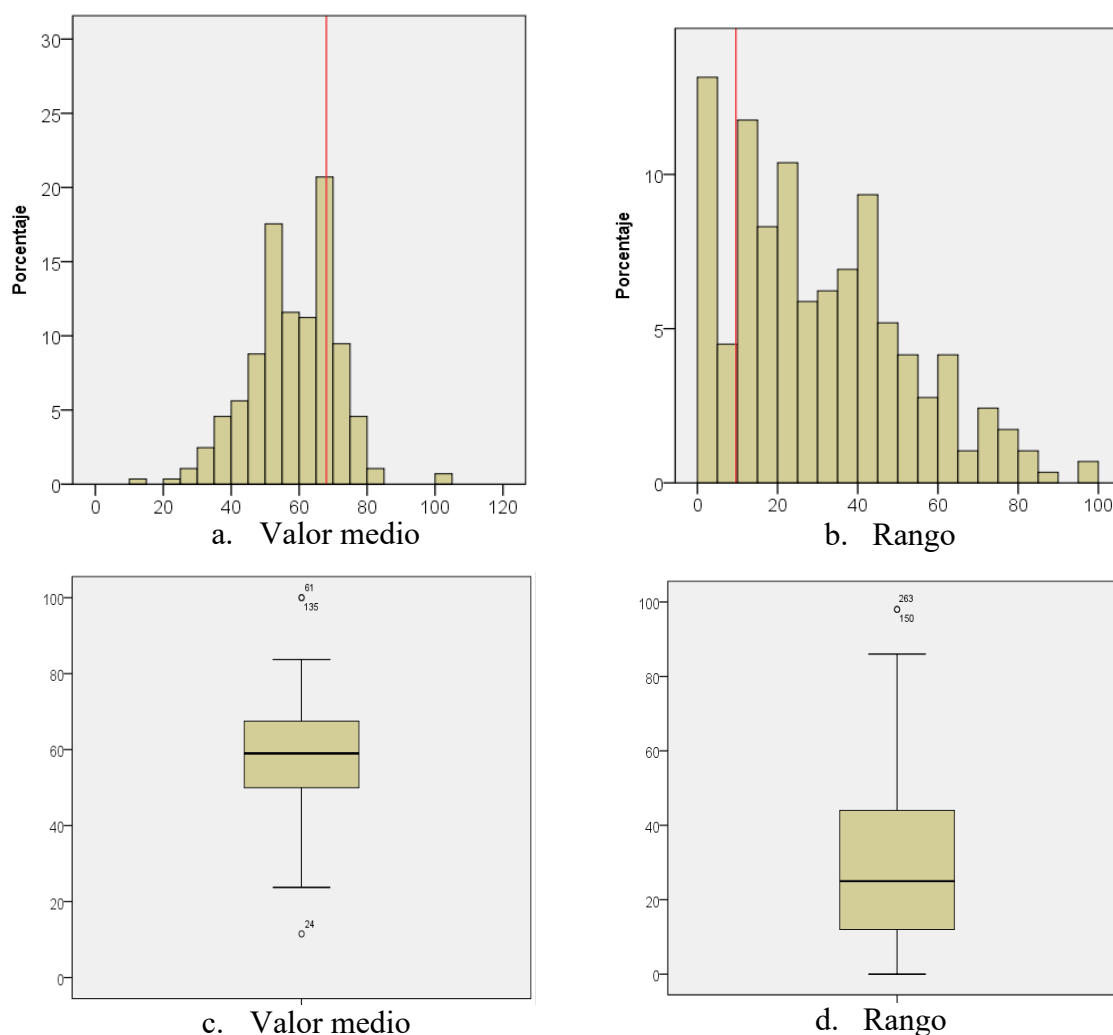


Figura 3.7.1. Distribución del valor medio y rango de cuatro valores del número de chinchetas hacia arriba en 100 chinchetas

El sesgo de equiprobabilidad ha sido recientemente explicado por Chernoff y Russel (2012), mediante lo que denominan como *falacia de la composición*, que consiste en transferir a un todo una propiedad que se cumple en una de sus partes. Según los autores, cuando los sucesos de un experimento aleatorio son equiprobables (como la cara y la cruz al lanzar una moneda), los sujetos extienden indebidamente la propiedad de equiprobabilidad a cualquier muestra de resultados del experimento compuesto; por ejemplo, consideran equiprobables obtener 0, 1 o 2 caras al lanzar dos monedas. Sin embargo, esta falacia no se sostiene en el experimento dado en el enunciado del ítem 1, ya que los resultados son claramente no equiprobables. En consecuencia, los estudiantes cuyas respuestas se localizan dentro del intervalo correspondiente a la equiprobabilidad,

estarían asignando esta propiedad a un proceso aleatorio en el que no se aplica, al igual que ocurrió en la investigación de Serrano (1996).

Por otro lado, la frecuencia de valores medios próximos al esperado [65,6-70,3] o respuestas aceptables, supone alrededor del 11% y se corresponde con aquellos alumnos que proporcionan una buena estimación del valor esperado en el experimento.

Otro grupo muestra una tendencia a dar valores medios más bajos del 50%, es decir, consideran que la proporción de chinchetas que caen con la punta hacia arriba es menor. Se observa que los estudiantes tratan de compensar el valor dado en el enunciado al producir muestras cuya proporción de chinchetas conduce a un valor medio menor que el esperado. En definitiva, estos estudiantes muestran la recencia negativa, debido a la heurística de la representatividad (Tversky et al., 1982), tal y como se ha razonado en el análisis del ítem 1, donde se explica la conducta asociada a dicha heurística y, en concreto, al sesgo recientemente explicado.

Relacionado también con la heurística de la representatividad se sitúan aquellas respuestas cuyo valor medio es superior al valor extremo del intervalo definido como aceptable, es decir, mayor que 72,7. En este caso, el grupo de estudiantes presenta la recencia positiva. Independientemente del sesgo presentado, el estudiante presenta en ambos casos una pobre comprensión de la independencia de ensayos repetidos. Esta falta de comprensión también se muestra en otros estudios como la investigación de Green (1991), Pollatsek, Konold, Well y Lima (1991) y Serrano (1996), lo que supone un bajo nivel de razonamiento, según Valdez (2016), ya que su estimación se basa en un modelo preestablecido, ignorando los datos frecuenciales.

La gráfica de caja del valor medio (Figura 3.7.1.c) muestra que aproximadamente el 25% de las respuestas se localizan en el intervalo [50, 65]. Los márgenes de dicho intervalo se corresponden con la mediana y el tercer cuartil, respectivamente. Por tanto, la muestra presenta, en general, una pobre concepción del valor esperado y, en general, de los fenómenos aleatorios en los que no se cumple la condición de equiprobabilidad. En este mismo gráfico (Figura 3.7.1.c), se identifican algunos casos atípicos que se comentan en una sección posterior de este mismo capítulo.

### **Análisis de la variabilidad**

En segundo lugar, se realiza un análisis de la variabilidad, a partir de la Figura 3.7.1.b, que representa la distribución del rango de los cuatro valores medios proporcionados por

cada estudiante en el global de la muestra. Se ha marcado el valor medio teórico (9,6) de dicha distribución, obtenida por simulación. La distribución del rango (Figura 3.7.1.b), y la media observada (23,6) se separan mucho de la media teórica (9,6), porque los estudiantes han exagerado la variabilidad de sus estimaciones. Aunque una parte de los estudiantes proporciona muestras en un rango aceptable, encontramos una gran variabilidad de rangos, llegando incluso a tomar el rango el valor de 100. Además, se obtuvieron respuestas (6,0%) con una ausencia de variabilidad total, es decir, en estos casos los estudiantes escriben una cuaterna (50, 50, 50, 50), respuesta que también aparece en el trabajo de Gómez et al. (2014).

En el gráfico de caja (Figura 3.7.1.d) se observa que una parte de la muestra proporciona rangos aceptables, mientras la otra mitad concede excesiva variabilidad o excesiva concentración a las muestras que construyen. Por tanto, siguiendo los niveles de razonamiento sobre el muestreo definidos por Moreno y Vallecillos (2001), muchos estudiantes se sitúan en uno de los niveles inferiores sobre muestreo, idiosincrático o de transición, puesto que no reconocen la variabilidad muestral y se muestran indiferentes hacia el tamaño de la muestra. Otros autores como Méndez (1991), Serrano (1996) y Shaughnessy, Ciancetta y Canada (2004) también identifican dificultades de los estudiantes para comprender la variabilidad muestral.

### **3.7.1.2. COMPARACIÓN POR GRUPO**

En este apartado vamos a comparar las respuestas según el grupo participante, utilizando los mismos estadísticos considerados anteriormente. Por tanto, se realiza un primer análisis de la distribución del valor medio (Figura 3.7.1) que se completa con la Tabla 3.7.3, en la que aparece el porcentaje de respuestas que se sitúa en cada intervalo. Los datos indican que los estudiantes de cuarto curso proporcionan un mayor número de muestras en las que el valor medio de las cuatro estimaciones se localiza, al menos, en intervalo denominado como aceptable, aunque la diferencia no resulta ser muy apreciable.

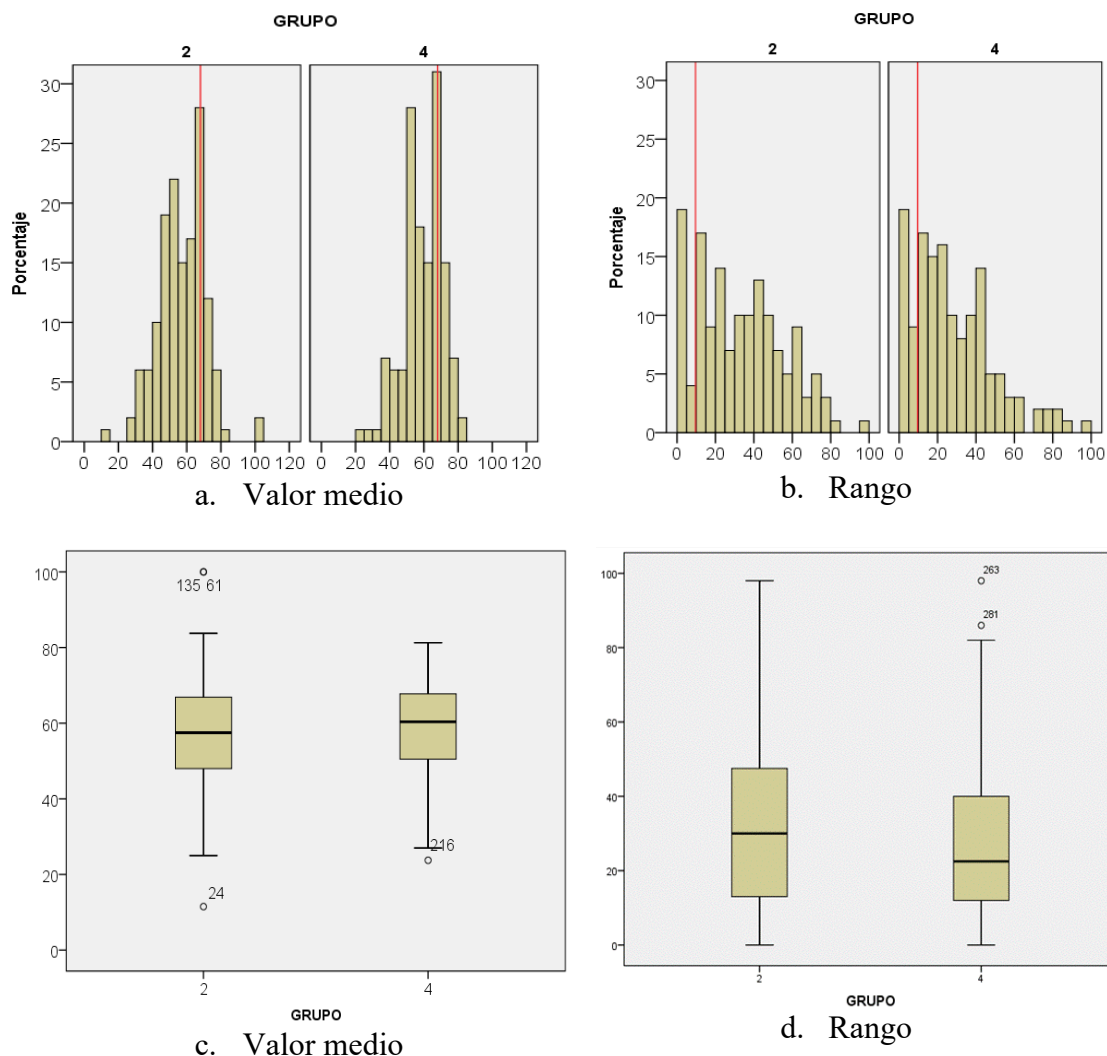


Figura 3.7.2. Distribuciones del valor medio y rango de cuatro valores del número de chinchetas hacia arriba en 100 chinchetas

Como se ha indicado en el análisis general, se han identificado algunos sesgos en las respuestas. En primer lugar, algunos estudiantes de ambos cursos presentan la recencia negativa. Por otro lado, independientemente del subconjunto estudiado, aproximadamente un cuarto de las respuestas se localiza en el intervalo definido como equiprobabilidad (ver Tabla 3.7.3). Por tanto, si consideramos la presencia conjunta de estas dos concepciones erróneas de la aleatoriedad, se deduce que una parte de la muestra presenta algún tipo de sesgo y, en consecuencia, el valor medio de las muestras producidas no se corresponde con el valor medio teórico.

Finalmente, el diagrama de cajas asociado al valor medio en cada grupo (Figura 3.7.2.c) muestra, por un lado, que la mediana es algo más baja en los estudiantes de segundo. Pensamos que esto ocurre porque, con mayor frecuencia, estos estudiantes dan valores cercanos al 50%; es decir, los estudiantes entienden que los dos resultados son

equiprobables, en vez de usar la información frecuencial. Observamos también en el gráfico de la caja (Figura 3.7.2.c) que tanto el primer cuartil como la mediana de la distribución es superior en los estudiantes de cuarto curso, lo que implica que su estimación se acerca más al valor teórico 68. Sin embargo, como se ha indicado en el párrafo anterior, la proporción de estudiantes que parece razonar de acuerdo a la equiprobabilidad es bastante parecida en los dos grupos.

Finalmente, el análisis de la Tabla 3.7.3 muestra una disminución de la proporción de estudiantes de cuarto curso que muestran recencia negativa o no completan la tarea. Este comportamiento también se observa en los histogramas (Figura 3.7.2.a) ya que, aunque en los dos gráficos siguen apareciendo dos modas, la frecuencia del grupo que da estimaciones alrededor del 50% disminuye en los estudiantes de cuarto curso.

Tabla 3.7.3. Porcentaje de estudiantes por grupo según valor medio en el ítem 1

Estimación de la media	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa del valor esperado [67,5-70,3]	10,8	11,7
Estimación aceptable del valor esperado <sup>1</sup> [63,3-72,7]	17,8	21,4
Recencia negativa (menor que 45)	16,6	11,0
Recencia positiva (mayor que 72,7)	8,3	8,3
Equiprobabilidad [45- 55]	26,8	23,4
Otros valores (55-63,3)	13,4	19,3
No completa	6,4	4,8

<sup>1</sup>Fuera del intervalo normativo

### Análisis de la variabilidad

Tras el análisis del valor medio, se presenta a continuación un estudio similar sobre el rango. En primer lugar, en las Figuras 3.7.2.b y 3.7.2.d y la Tabla 3.7.4 se representa la distribución del rango según el curso y se observa que ambos grupos conceden una variabilidad excesiva a las muestras proporcionadas. De hecho, la Tabla 3.7.2 muestra que la media teórica (9,6) y la observada (23,6) de los rangos es muy diferente, teniendo 14 unidades de diferencia.

La Tabla 3.7.4 nos informa que aproximadamente el 62,4% de los estudiantes de segundo curso de la ESO que participan en el estudio muestran una comprensión pobre acerca de la variabilidad asociada, debido a que muestran una variabilidad excesiva. Aunque este porcentaje es menor para los alumnos de cuarto curso, es también muy alto, del 56,6%. Por tanto, ambos grupos elaboran con alta frecuencia respuestas con una variabilidad extrema.

Además, el siguiente intervalo con mayor frecuencia se corresponde con valores de los intervalos que se han denominado como normativos o aceptables. En este caso, se

obtiene solo el 20% (ver Tabla 3.7.4.) de la muestra de segundo curso y el 30% para los de cuarto. Es notable el porcentaje de estudiantes con una alta concentración de los valores que estiman para el número de chinchetas que caen con la punta hacia arriba.

Tabla 3.7.4. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 1

Estimación de la variabilidad	2°ESO (n=157)	4°ESO (n=145)
Estimación normativa [5-14]	12,7	17,9
Estimación aceptable [3-18] <sup>1</sup>	7,6	11,7
Estimación de una variabilidad muestral excesiva >18	62,4	56,6
Alta concentración <3	10,8	9
No completa	6,4	4,8

<sup>1</sup>Fuera del intervalo normativo

Por otro lado, la dispersión asociada a este ítem es similar en los dos grupos. Los valores atípicos identificados se corresponden con respuestas cuyos valores para el número de chinchetas se sitúan en los valores extremos del rango de respuestas posibles, como se observa en las siguientes respuestas de los estudiantes: 2P\_A61 y 2S\_A135: (100, 100, 100, 100), 2P\_24: (22, 0, 8, 16) y 4BP\_A216: (20, 40, 20, 15). Como señala Gómez et al. (2014), estos dos últimos resultados pueden indicar una creencia en la compensación entre los resultados, mostrando la heurística de la representatividad, ya comentada.

La Figura 3.7.2.d, que corresponde a los rangos de las cuatro puntuaciones, muestra menor mediana y menor tercer cuartil en el grupo de estudiantes de cuarto curso, confirmando que éstos parecen conceder, en mayor proporción, una variabilidad menor a las muestras cuyo tamaño es grande.

Análogamente, se identifican algunos valores atípicos, entre los que destaca la siguiente respuesta, 4BS\_A263: (99, 1, 50, 60). Ésta revela que el estudiante tiene una concepción errónea de la variabilidad muestral asociada al muestreo, puesto que no tiene en cuenta el efecto del tamaño de la muestra, lo que conduce a la aceptación como una respuesta probable todas aquellas respuestas que se localizan en el rango de valores posibles [0-100]. En consecuencia, algunos estudiantes no llegan a alcanzar los niveles superiores de razonamiento sobre el muestreo descritos por Moreno y Vallecillos (2001), puesto que no aprecian el efecto del tamaño de la muestra sobre la variabilidad muestral. Estos errores son también descritos por Méndez (1991), Serrano (1996) y Shaughnessy, Ciancetta y Canada (2004).

Las medias y desviaciones típicas de los dos estadísticos considerados en este trabajo



se recogen en la Tabla 3.7.5, en función de los dos cursos. Además, se incluye el error de muestreo y el número de estudiantes que responden al ítem en cada grupo.

Tabla 3.7.5. Estadísticos de la media y rango en el Ítem 1 por curso

	Curso	N	Media	D. típica.	Error muestreo
II Media	2º	147	56,920	13,7228	1,1318
	4º	138	58,899	11,6006	0,9875
II Rango	2º	147	32,07	22,499	1,856
	4º	138	26,68	20,764	1,743

Observamos una pequeña diferencia en el valor medio, siendo el correspondiente al cuarto curso algo más alto, es decir, más próximo a la estimación teórica, aunque aún muy diferente de ella. Por el contrario, hay una diferencia notable, de casi 5 puntos en el valor medio del rango de las cuatro estimaciones, según el curso (ver Tabla 3.7.5). En concreto, su valor es menor para los estudiantes de cuarto curso, lo que indica que sus estimaciones tienen menor variabilidad. Por tanto, parece que visualizan mejor que sus compañeros la disminución de variabilidad en las muestras grandes.

Tabla 3.7.6. Prueba t de diferencias de medias

Prueba t para la igualdad de medias					Intervalo de confianza de la diferencia (95%)		
	t	Gl	Sig. (bilateral)	Diferencia de medias	Error típico diferencia	E. Inferior	E. Superior
II Media	-1,310	283	0,191	-1,98	1,51	-4,95	0,99
II Rango	2,115	287	0,035	5,39	2,55	0,37	10,41

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba t de diferencia de medias independientes (utilizando el caso de varianzas distintas, que es más conservador). Los resultados se presentan en la Tabla 3.7.6, junto con los intervalos de confianza del 95% para la diferencia de las medias. Observamos que la diferencia de medias no es estadísticamente significativa y sí lo es la diferencia de rangos, cuyo valor  $p$  es menor que 0,05 y cuyo intervalo de confianza no cubre el valor cero. Concluimos que no tenemos motivos para suponer que los dos grupos de estudiantes muestran diferente comprensión de la relación entre el valor esperado y el teórico en este ítem, pero los de cuarto comprenden mejor la variabilidad del muestreo en este ítem.

### **3.7.2. RESULTADOS EN EL ÍTEM 2**

Se presentan, a continuación, los resultados del análisis del Ítem 2, relativo al número de caras que se obtiene en el lanzamiento de 100 monedas equilibradas.

#### **3.7.2.1. ANÁLISIS GLOBAL**

De manera análoga a lo realizado en el primer ítem, el estudio del segundo enunciado se centra, en primer lugar, en el análisis de la distribución del valor medio asociado a los cuatro valores dados por el estudiante (Figura 3.7.5.a), cuya media toma el valor de 51,21, mientras que en el análisis previo del ítem se mostró que el valor medio esperado es 50. Por tanto, el valor medio teórico para este ítem es próximo al valor medio de las respuestas de toda la muestra participante.

En consecuencia, los participantes muestran una adecuada estimación de la proporción del número de caras esperada en un lanzamiento de 100 monedas. Asimismo, los estudiantes muestran una comprensión adecuada de la estimación clásica de la probabilidad. Este resultado puede deberse a que, por un lado, los estudiantes entienden y aplican el principio de equiprobabilidad y a que, por otro lado, el tipo de contexto presentado puede resultarle familiar. En definitiva, se obtiene un porcentaje mayor de respuestas correctas en relación con lo obtenido para el primer ítem.

El análisis de las frecuencias de los valores medios que se presentan en ciertos intervalos (Figura 3.7.5.a), indica, en primer lugar, que el 38,4% de los estudiantes proporcionan muestras tales que el valor medio del número de caras es próximo al valor esperado [47,5-52,5], intervalo que se ha considerado como normativo.

Por otro lado, las frecuencias que toman los distintos intervalos simétricos respecto a la media son similares. Interpretamos que los estudiantes identifican la equiprobabilidad del fenómeno aleatorio.

Este análisis se completa con el diagrama de cajas asociado (Figura 3.7.5.c). En este caso, el valor de la mediana toma un valor en torno al 50 y los valores del primer y tercer cuartil son próximos a dicho valor. De hecho, el rango intercuartílico que recoge el 50% de la muestra participante está constituido por el intervalo [45, 55] aproximadamente. Por otro lado, la longitud de los bigotes indica que los valores están concentrados en torno al valor de la mediana. Sin embargo, hay un gran número de casos atípicos que corresponden a aquellas respuestas cuya estimación media se aleja del 50% en más 10 unidades y que se comentan al final de este capítulo.

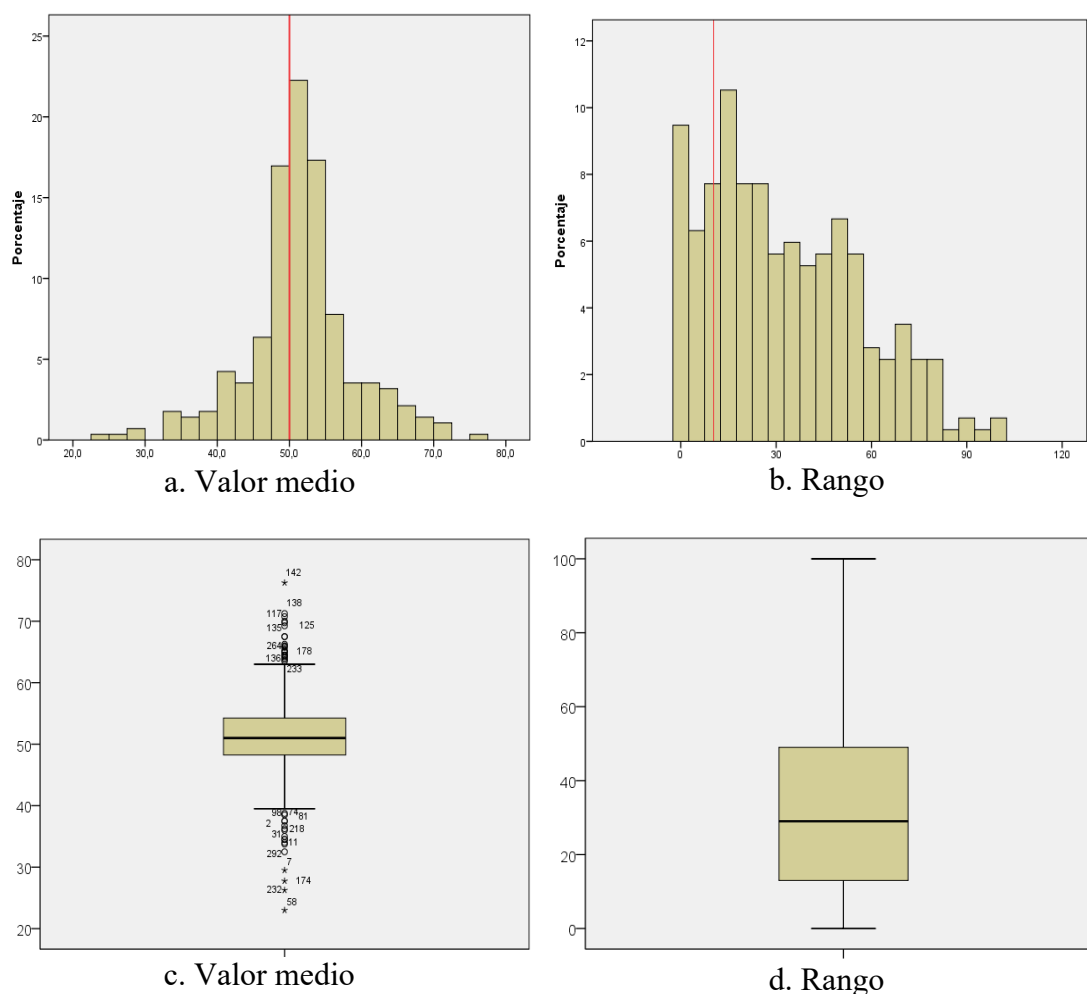


Figura 3.7.5. Distribución del valor medio y rango de cuatro valores del número de caras en 100 lanzamientos de monedas

### Análisis de la variabilidad

La Figura 3.7.5.b representa la distribución del rango asociado a la cuaterna de estimaciones proporcionadas, e indica que los estudiantes conceden una variabilidad excesiva a las respuestas aportadas. Además, este resultado se deduce también debido a la diferencia entre la media teórica (línea roja) y el centro de la distribución. De hecho, más de la mitad de la muestra (58,8%) escribe respuestas cuyo rango excede de 20, que corresponde con una dispersión excesiva. Por tanto, razonando de manera análoga que, en el ítem anterior, los estudiantes no comprenden el efecto del tamaño de la muestra, y producen valores con variabilidad excesiva, resultado que también se identifica en la investigación de Serrano (1996). En consecuencia, no se alcanzan los niveles superiores de razonamiento sobre el muestreo descritos por Moreno y Vallecillos (2001) tampoco en este ítem; los estudiantes no muestran una correcta comprensión de la variabilidad,

coincidiendo con lo reportado en las investigaciones de Méndez (1991), Serrano (1996) y Shaughnessy, Ciancetta y Canada (2004).

Por otro lado, sobre el 15% de la muestra produce valores cuyo rango es menor que 10, es decir, producen muestras una concentración excesiva. Estas mismas conclusiones se obtienen del análisis del gráfico de la caja correspondiente a los rangos (Figura 3.7.5.d), aunque en este caso no aparecen valores atípicos.

### **3.7.2.2. COMPARACIÓN POR GRUPO**

En relación al estudio de la media muestral, el análisis comparativo de los dos grupos se resume en las Figuras 3.7.6.a y 3.7.6.c. Los gráficos muestran modas próximas en los dos grupos, con frecuencia mayor cerca de la moda en el cuarto curso, mientras que la dispersión de la distribución de los valores medios es más acusada en la muestra de los estudiantes de segundo curso de la ESO, resultado que se observa también en el gráfico de caja (Figura 3.7.6.c).

Esta información se completa con los datos recogidos en la Tabla 3.7.7, donde el porcentaje de estudiantes que dan un valor medio correcto supone el 69,6% de los estudiantes de cuarto curso, mientras que dicho porcentaje es del 50,3% para el otro grupo. Por tanto, las respuestas de estudiantes de cuarto curso reflejan una mayor concentración en torno al valor esperado teórico.

Por otro lado, la tercera parte de los estudiantes de segundo curso de la ESO proporcionan muestras cuyo valor medio es normativo. Se completa con los estudiantes que dan una estimación aceptable y con los que producen sistemáticamente valores menores o mayores que los aceptables.

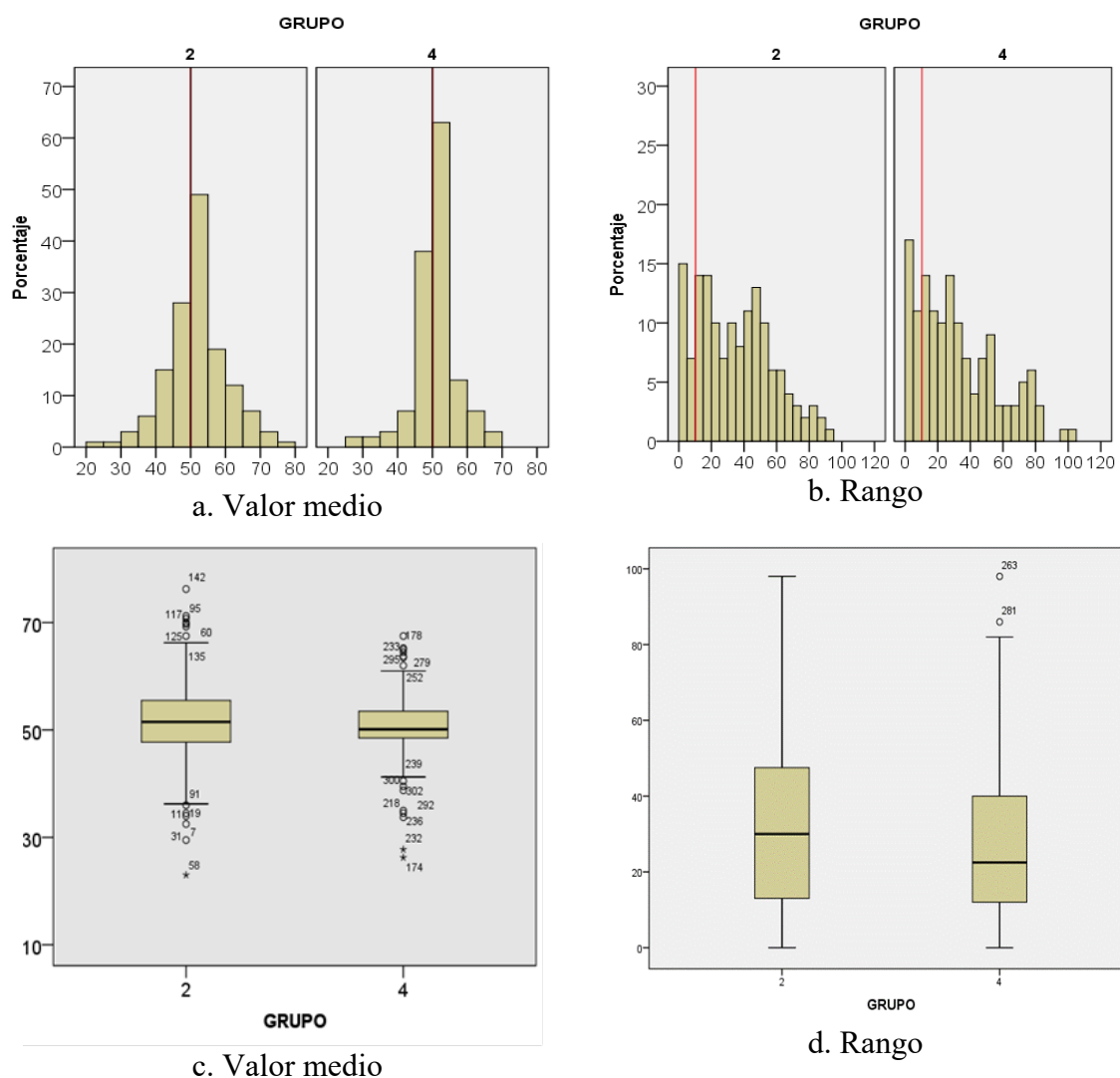


Figura 3.7.6. Distribución del valor medio y rango de cuatro valores del número de caras en 100 lanzamientos de monedas

Tabla 3.7.7. Porcentaje de estudiantes por grupo, según valor medio en el ítem 2

Estimación de la media	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa [47,5-52,5]	31,2	46,2
Estimación aceptable [45-55] <sup>1</sup>	19,1	23,4
Menores que el aceptable (< 45)	16,6	9,7
Mayores que el aceptable (>55)	25,5	15,9
No completa	7,6	4,8

<sup>1</sup>Fuera del intervalo normativo

Finalmente, el diagrama de cajas nos permite el análisis de los valores atípicos (ver Figuras 3.7.6.c), que se corresponden con respuestas que presentan una variabilidad muy baja. A continuación, se presentan dos ejemplos de respuestas de estudiantes 2S\_A142: (70, 78, 75, 82), así como la respuesta del estudiante 4BP\_174: (30, 25, 32, 18), donde

ninguno de estos conjuntos de datos incluye, además, el valor esperado.

### Análisis de la variabilidad

En segundo lugar, el estudio de la variabilidad se apoya en el cálculo del rango de las cuatro estimaciones y se fundamenta en el análisis conjunto de las Figuras 3.7.6.b y 3.7.6.d y la Tabla 3.7.8. Esta tabla revela que ambos grupos conceden una variabilidad excesiva; en concreto, más de la mitad de participantes producen respuestas que se localizan en el intervalo [20, 100]. Además, el análisis de la distribución del rango asociado a las respuestas (Figura 3.7.6.b) muestra otro intervalo, [40, 55], en el que se concentra un porcentaje significativo de respuestas.

Tabla 3.7.8. Porcentaje de estudiantes por grupo, según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 2

Estimación de la variabilidad	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa [6-15]	17,8	17,2
Estimación aceptable [3-20] <sup>1</sup>	8,3	9,7
Excesiva (>20)	58,6	59,3
Alta concentración (<3)	7,7	9,0
No completa	7,6	4,8

<sup>1</sup>Fuera del intervalo normativo

Para completar este análisis, se considera el diagrama de barras asociado a cada grupo (ver Figura 3.7.6.b). Por un lado, el diagrama de barras de la muestra de estudiantes de segundo curso toma como valores todos los del intervalo [0, 100], que se corresponde con las respuestas posibles que el alumno puede dar para el número de caras, a diferencia del grupo de cuarto, cuya amplitud es menor. Por tanto, las muestras dadas por los alumnos de segundo presentan una variabilidad mayor.

Al estudiar el diagrama de cajas (Figura 3.7.6.d), el análisis de los cuartiles revela que el 75% de las respuestas se localizan en el intervalo [15, 45] para el grupo de estudiantes de menor edad y en el intervalo [15, 40] para el otro grupo. Consecuentemente, la longitud del diagrama de caja es menor en el grupo de mayor edad y la mediana toma un valor menor también, lo cual apoya la conclusión anteriormente descrita, en la que se indica que este grupo de estudiantes producen muestras con una variabilidad menor, aunque no siempre adecuada. En resumen, el estudio de los diagramas de caja muestra menos dispersión de valores en el grupo de cuarto, puesto que dan respuestas más homogéneas (Figura 3.7.6.d).

En consecuencia, el análisis conjunto del valor medio y la variabilidad conlleva a

determinar que los estudiantes no presentan una intuición simultánea de las dos ideas antagónicas cuya coordinación es necesaria para una comprensión adecuada del concepto de muestreo. Como se ha señalado en el Capítulo 2, esas ideas fundamentales son la representatividad muestral y la variabilidad muestral (Batanero et al., 1994). Esta observación indica que, los estudiantes muestran una comprensión parcialmente correcta de la Ley de los Grandes Números, resultado que se identifica también en el trabajo de Gómez et al. (2014).

En el diagrama de cajas asociado al rango (Figura 3.7.6.d) solamente se identifican valores atípicos para la muestra de cuarto curso. En particular, las respuestas que se identifican como atípicas son las siguientes: 4BS\_A263: (1, 99, 49, 51), donde se observa que este estudiante no presenta una adecuada comprensión de la variabilidad muestral, dado que considera como probable cualquier valor que se localiza en intervalo de valores posibles definido por  $[0, 100]$ .

Para finalizar el análisis del segundo enunciado del cuestionario, se presenta la Tabla 3.7.9, donde se muestra las medias y desviaciones típicas tanto del valor medio como del rango, en función del grupo participante, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo

Tabla 3.7.9. Estadísticos de la media y rango en el ítem 2 por curso

	Curso	N	Media	D. típica.	Error muestreo
I2Media	2º	145	51,581	8,6143	0,7154
	4º	138	50,828	6,3624	0,5416
I2Rango	2º	145	33,68	23,155	1,916
	4º	138	31,20	24,548	2,082

En dicha tabla se observa una pequeña diferencia en el valor medio, siendo el correspondiente al cuarto curso algo más bajo, es decir, más próximo al valor teórico, aunque los dos son cercanos a él. En el caso del rango, hay una diferencia de unos dos puntos, siendo el valor medio del rango menor para los estudiantes de cuarto curso. Por tanto, las estimaciones de este grupo presentan una variabilidad menor, resultado que ya se ha descrito en las líneas anteriores. Por tanto, parece que visualizan mejor que sus compañeros la disminución de la variabilidad en las muestras grandes.

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba t de diferencia de medias independientes (utilizando el caso de varianzas distintas, que es más conservador). Los resultados se presentan en la Tabla 3.7.10, junto con los intervalos de confianza del 95% para la diferencia de las medias.

Observamos que ni la diferencia de medias ni la de rangos son estadísticamente significativas, por lo que en este ítem no se observan diferencias apreciables en los dos cursos.

Tabla 3.7.10. Prueba t de diferencias de medias

	Prueba t para la igualdad de medias					Intervalo de confianza de la diferencia (95%)	
	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. diferencia	E. Inferior	E. Superior
I2Media	0,833	281	0,405	0,75	0,90	-1,03	2,53
I2Rango	0,879	283	0,380	2,48	2,83	-3,08	8,04

### 3.7.3. RESULTADOS EN EL ÍTEM 3

A continuación, se estudian los resultados en el ítem 3, sobre el lanzamiento de 10 monedas equilibradas, y el primero de los dos ítems donde se tratan pequeñas muestras.

#### 3.7.3.1. ANÁLISIS GLOBAL

En primer lugar, se considera la distribución de las medias de los cuatro valores que el estudiante sugiere para el número de caras, como en los ítems anteriores. El valor medio en el global de la muestra de estudiantes, 5,12, es muy próximo al valor esperado, 5, lo que indica una estimación buena del valor esperado.

Del análisis de las frecuencias con que los valores medios se localizan en ciertos intervalos, se infieren diferentes tipos de razonamientos (Figura 3.7.11.a). En el análisis previo de la tarea se dieron como respuestas aceptables aquellas que consideran un número de caras dentro del intervalo [3,4-6,6]. En nuestro caso, el 83,7% de los estudiantes proporcionan muestras donde la proporción del número de caras esperada se localiza en dicho intervalo. Por tanto, se observa que los estudiantes reflejan una buena concepción de la frecuencia esperada para este ítem. Dentro de este grupo de estudiantes, se identifica un subgrupo que proporcionan valores medios entre [4,5-5,5]. Este grupo supone el 65,61% y proporciona unos resultados muy próximos al valor teórico, bien porque dan valores muy próximos al valor esperado o, bien porque se compensan las distintas frecuencias.



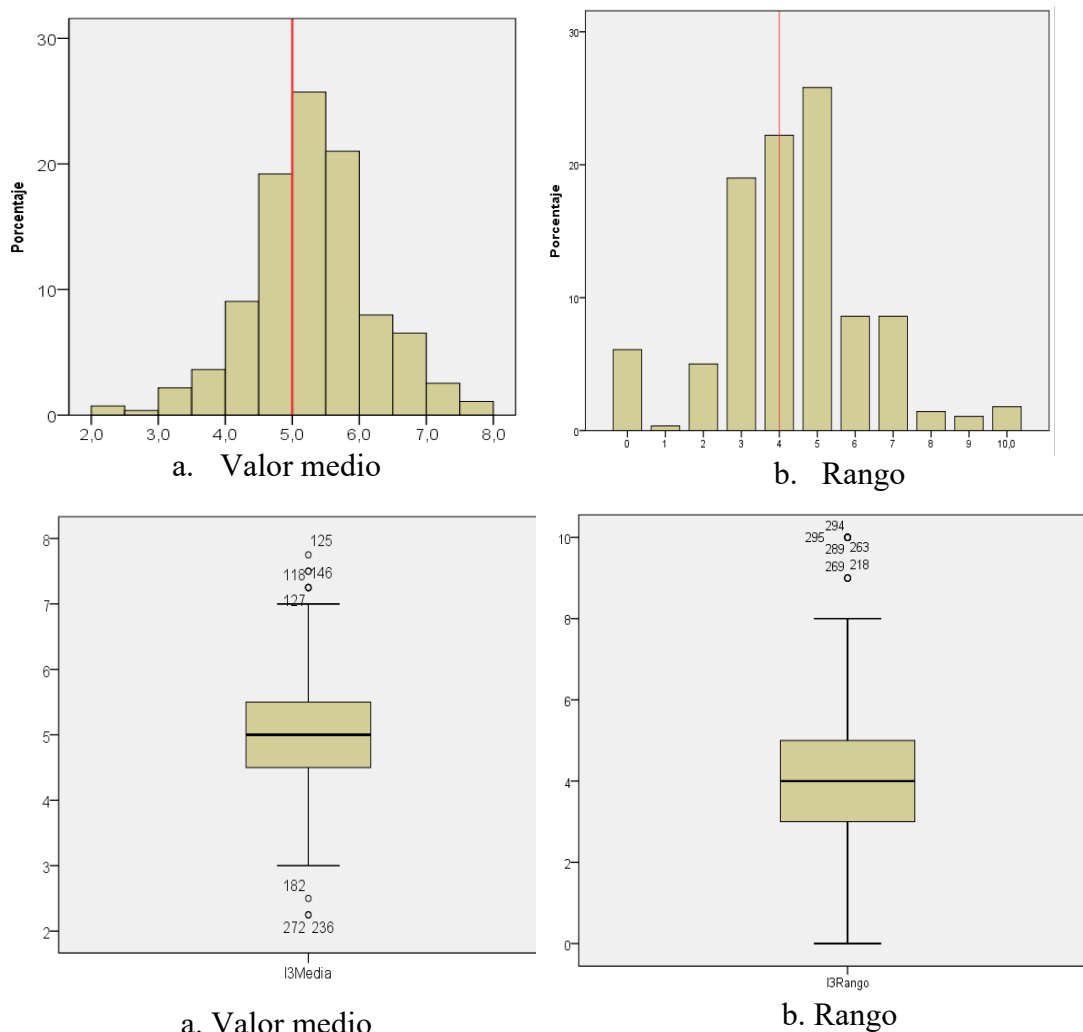


Figura 3.7.11. Distribución del valor medio y rango de cuatro valores del número de caras en 10 lanzamientos de monedas

De manera análoga al ítem anterior, se observa que los estudiantes identifican la equiprobabilidad del experimento, debido a la simetría que adopta la distribución del valor medio, tanto en el histograma (Figura 3.7.11.a) como en el gráfico de la caja (Figura 3.7.11.c). El análisis del rango intercuartílico del diagrama de caja del valor medio, (Figura 3.7.11.c) indica que el 50% de los estudiantes concentra su valor medio entre 4,5 y 5,5 aproximadamente. Si consideramos el intervalo que contiene al 95% de las respuestas, identificamos la longitud de los bigotes próxima a los extremos de dicho intervalo. Por tanto, los estudiantes proporcionan muestras adecuadas al enunciado. Finalmente, la gráfica presenta algunos casos atípicos que se comentan en una sección posterior.

### **Análisis de la variabilidad**

En cuanto al estudio de la variabilidad (Figuras 3.7.11.b y 3.7.11.d), el rango de las muestras elaboradas es ahora aceptable en el mayor número de estudiantes. Por un lado, la Figura 3.7.11.b recoge la distribución del rango de las cuatro estimaciones, mostrando que la mayoría de los estudiantes (52,31%) proporcionan muestras cuyo rango varía dentro del intervalo [2-5]. Por tanto, la proporción de estudiantes que responden a esta cuestión correctamente es mayor que en los dos ítems anteriores. Esto se traduce en que los estudiantes presentan una intuición o comprensión adecuada del muestreo en un experimento aleatorio asociado a un número de ensayos menor y cuyos sucesos elementales son equiprobables. A diferencia de los ítems anteriores, en este caso, no se proporciona los valores asociados a un ensayo previo, por lo que hay que asignar probabilidad en el significado clásico.

#### **3.7.3.2. COMPARACIÓN POR GRUPO**

El análisis de la distribución del valor medio (Figuras 3.7.13.a y 3.7.13.c) sugiere la simetría en ambos grupos, debido a que se localiza una proporción mayor de respuestas en torno al valor de la media teórica. De hecho, la Tabla 3.7.12 muestra que son más frecuentes aquellas muestras cuyo valor medio se localiza en el intervalo [4,2-5,8]. Además, el porcentaje de respuestas que se pueden considerar correctas toma valores próximos al 80% en ambos grupos. Por tanto, los estudiantes de ambos cursos presentan una adecuada comprensión del valor medio.

No obstante, el diagrama de cajas asociado a cada grupo (Figuras 3.7.13.c) indica que la distribución del estadístico para el grupo de cuarto curso presenta una simetría mayor. Por un lado, la mediana toma el valor 5 y se localiza en la zona media de la caja y, además, la longitud de los bigotes es similar por cada extremo de la caja. Por lo que, dichos estudiantes presentan una mejor comprensión del estadístico.

A pesar de que ambos grupos proporcionan muestras adecuadas en un porcentaje elevado, se observa que algunos estudiantes presentan el sesgo denominado enfoque del resultado (Konold, 1989), lo que puede deberse a una concepción errónea de la aleatoriedad o del fenómeno que se estudia, en particular. En este caso, señalamos aquellas respuestas constituidas por cuaternas en las que el número de caras es el mismo para cada ensayo.

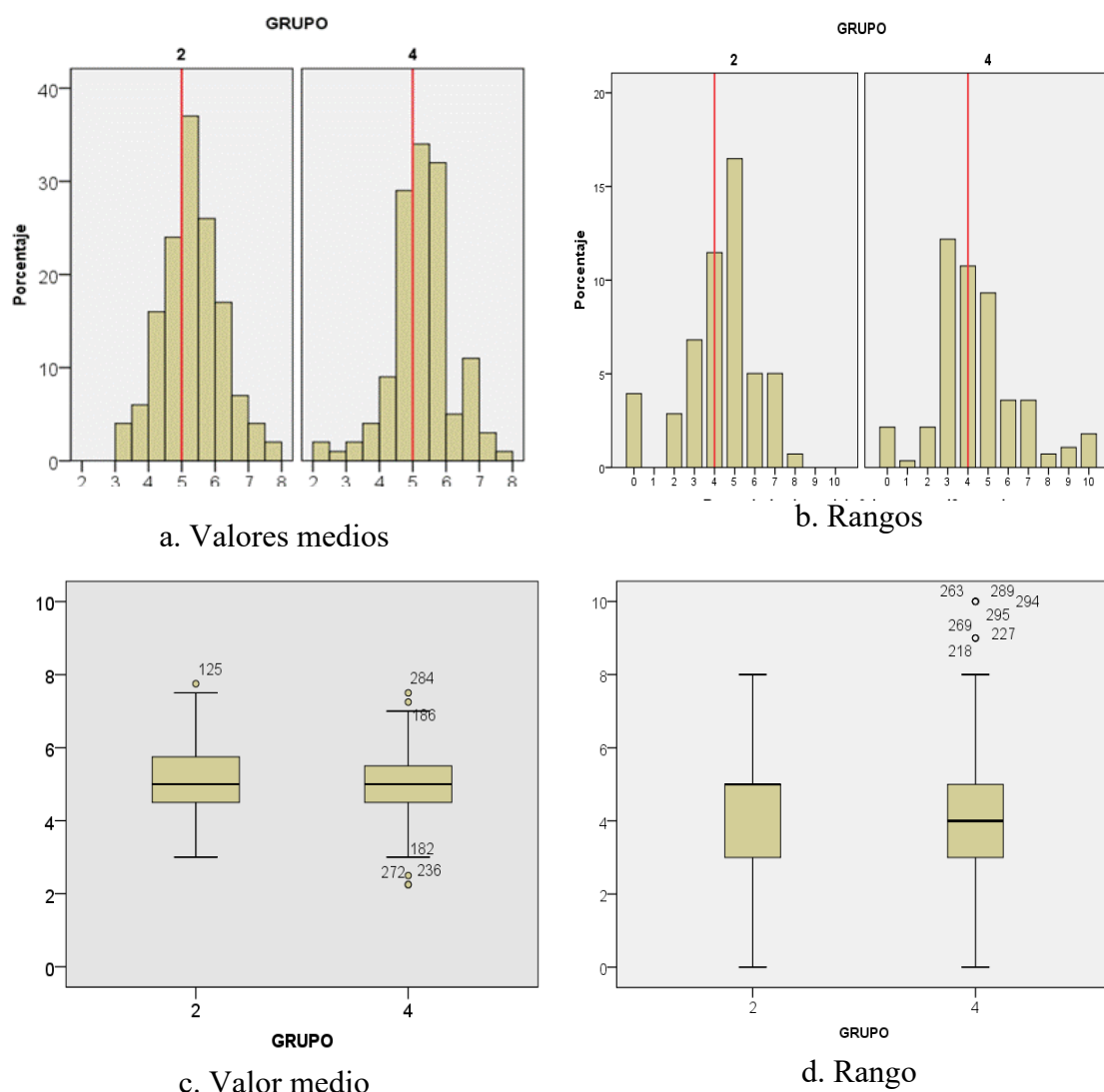


Figura 3.7.13. Distribución del valor medio y rango de cuatro valores del número de caras en 10 lanzamientos de monedas

Tabla 3.7.12. Porcentaje de estudiantes por grupo según valor medio en las cuatro estimaciones en el ítem 3

Estimación de la media	2º	4º
	ESO	ESO
Estimación normativa [4,2-5,8]	56,7	67,6
Estimación aceptable [3,4-6,6] <sup>1</sup>	26,8	16,6
Menor que el aceptable (<3,4)	2,5	3,4
Mayor que el aceptable (> 6,6)	5,1	4,1
No completa	8,9	8,3

<sup>1</sup>Fuera del intervalo normativo

El diagrama de cajas identifica una serie de valores atípicos en ambos cursos. En primer lugar, en el grupo de segundo curso, se identifica un valor atípico, que corresponde a la siguiente respuesta: 2S\_A125: (7, 6, 8, 10), donde el estudiante proporciona valores

muy extremos, mostrando una pobre intuición sobre la frecuencia esperada asociada al número de caras. Por su parte, las respuestas atípicas dadas por los estudiantes de cuarto son: A236: (3, 4, 2, 0), A272: (5, 1, 3, 0) y A284: (9, 8, 7, 6). De manera análoga, se observa que estas muestras se caracterizan por valores muy extremos y los valores se sitúan todos o sobre o bajo la media.

### Análisis de la variabilidad

En segundo lugar, el estudio de la variabilidad se basa, por un lado, en la distribución del rango (Figuras 3.7.13.b y 3.7.13.d) para cada uno de los grupos. El análisis comparativo de las frecuencias que se alcanzan en determinados intervalos, revela que la variabilidad es normativa o aceptable en aproximadamente el 70% de los estudiantes para ambos grupos (ver Tabla 3.7.14). A pesar de ello, la Tabla 3.7.14 muestra que existe un porcentaje de estudiantes que conceden una variabilidad excesiva. También, destacan aquellas respuestas caracterizadas por presentar una alta concentración que, en este ítem, se corresponde con la ausencia de variabilidad de las cuatro muestras.

Tabla 3.7.14. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 3

Estimación de la variabilidad	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa [2-5]	66,9	66,2
Estimación aceptable [1-6] <sup>1</sup>	8,9	7,6
Variabilidad excesiva (>6)	10,2	13,8
Alta concentración (0)	5,1	4,5
No completa	8,9	8,3

<sup>1</sup>Fuera del intervalo normativo

Por otro lado, el diagrama de caja (Figura 3.7.7.d.) permite la identificación de los valores atípicos. En concreto, solamente se consideran dentro de esa categoría aquellas respuestas que presentan una variabilidad excesiva, mientras que las respuestas caracterizadas por su ausencia de variabilidad dan lugar al valor mínimo que puede tomar la variable.

A continuación, se indican algunos de los casos atípicos que se visualizan en el gráfico de la caja, que son más en el grupo de cuarto curso. Así, por ejemplo, A263 produce la cuaterna (5, 10, 0, 7) y A289 la (10, 0, 5, 8). Estos estudiantes muestran una pobre concepción de la variabilidad en pequeñas muestras y desconocimiento de los casos más probables en la distribución binomial.

Las medias y desviaciones típicas del valor medio y el rango de dicho valor medio en las cuatro estimaciones de cada estudiante, en cada uno de los grupos, se presentan en

la Tabla 3.7.15, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo. Observamos una pequeña diferencia en el valor medio, siendo el correspondiente al cuarto curso algo más alto, es decir, más próximo a la probabilidad teórica, aunque diferente a ella. En relación a los valores del rango, no se identifica una diferencia notable en el rango de las cuatro estimaciones, aunque se presenta una mayor desviación típica en las respuestas de cuarto curso.

Tabla 3.7.15. Estadísticos de la media y rango en el Ítem 3 por curso

	Curso	N	Media	D. típica.	Error muestreo
I3Media	2º	143	5,142	,9441	,0790
	4º	133	5,105	,9182	,0796
I3Rango	2º	143	4,308	1,8064	,1495
	4º	133	4,421	2,1005	,1821

Tabla 3.7.16. Prueba t de diferencias de medias

Prueba t para la igualdad de medias					Intervalo de confianza de la diferencia (95%)		
	t	gl	Sig. (bilateral)	Diferencia de medias	Error típico diferencia	E. Inferior	E. Superior
I3Media	0,324	274	0,746	0,04	0,11	-0,18	0,26
I3Rango	-0,482	277	0,630	-0,11	0,23	-0,57	0,35

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba t de diferencia de medias independientes (utilizando el caso de varianzas distintas, que es más conservador). Los resultados se presentan en la Tabla 3.7.16, junto con los intervalos de confianza del 95% para la diferencia de las medias. Observamos que ni la diferencia de medias ni de rangos medios es estadísticamente significativa.

### 3.7.4. RESULTADOS EN EL ÍTEM 4

Finalmente se analiza el ítem 4, cuyo contexto es el número de 10 tiradas a una canasta de baloncesto. En este ítem, la probabilidad asignada a cada suceso elemental es diferente. Además, la identificación de la probabilidad de cada suceso debe realizarse mediante información frecuencial.

#### 3.7.4.1. ANÁLISIS GLOBAL

La media global de la distribución de las medias de los cuatro valores del estudiante del número de canastas encestandas en un partido, donde el jugador lanza 10 veces, toma un valor de 7. De hecho, el 74,2% de los estudiantes proporcionan un valor medio

localizado en el intervalo definido como respuesta correcta. Es decir, los estudiantes proporcionan, en media, muestras cuya frecuencia refleja de manera adecuada la proporción en la población.

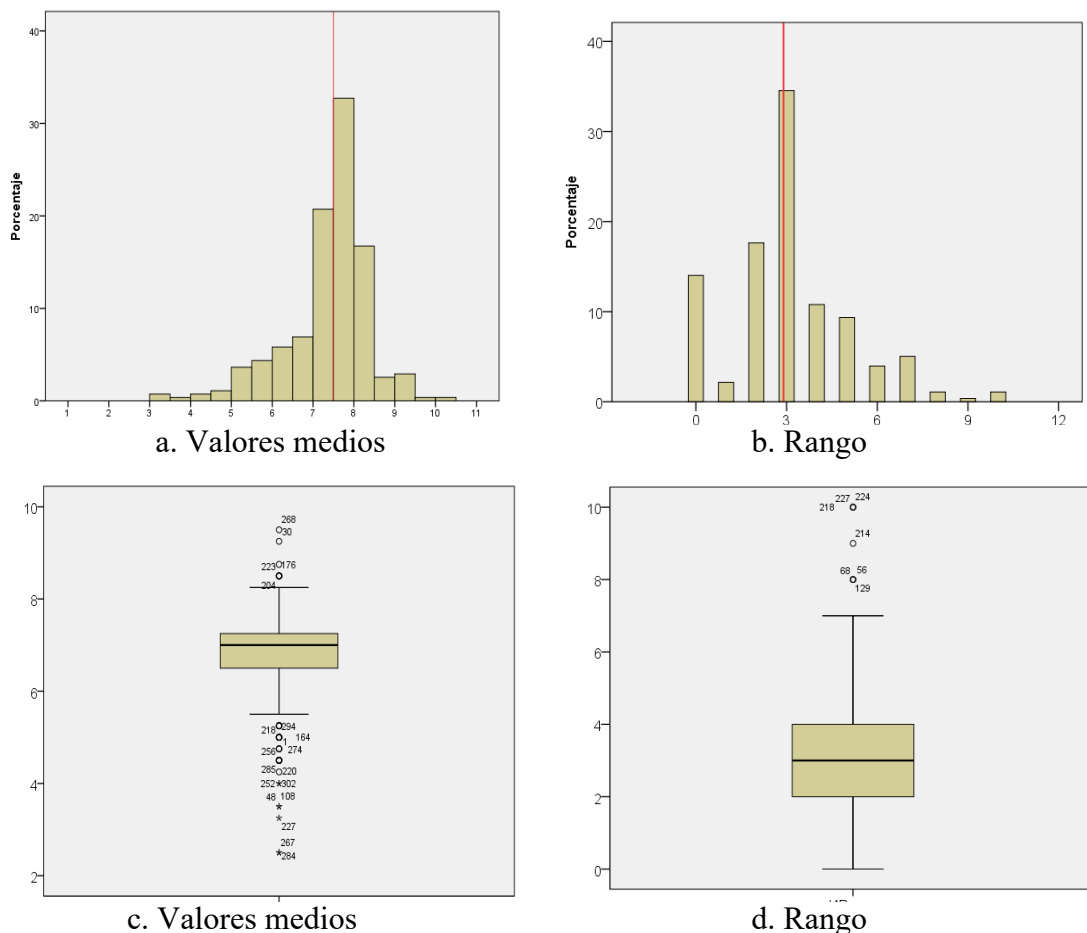


Figura 3.7.17. Distribución del valor medio y rango de cuatro valores del número de aciertos en 10 tiradas a una canasta

Por otro lado, la falta de simetría de los datos (Figura 3.7.17.a) indica que los estudiantes han identificado la no equiprobabilidad del experimento. Puede ser que su conocimiento del contexto sea mejor en este ítem que en comparación con el caso de las chinchetas. Este factor puede explicar que este hecho haya favorecido a que los estudiantes razonen correctamente sobre la probabilidad de los resultados del experimento.

La gráfica de caja del valor medio (Figura 3.7.17.c) muestra que el 50% de los estudiantes concentra su valor medio alrededor de 7, valor teórico esperado. Además, casi la totalidad de los estudiantes producen respuestas dentro del intervalo considerado como aceptable. De hecho, el resto de casos que no se localizan dentro de dicho intervalo los

señala como casos atípicos. En relación con estas respuestas, el gráfico identifica bastantes casos atípicos, tanto por encima, como por debajo de los cuartiles inferior y superior. Por ejemplo, el estudiante A284 produce la cuaterna (1, 2, 3, 4), mientras que A268 ofrece (10, 9, 9, 10), lo que muestra una pobre concepción del valor esperado y pobre comprensión de la probabilidad, desde el punto de vista frecuencial.

### **Análisis de la variabilidad**

En relación con el estudio de la variabilidad, la distribución del rango representada en la Figura 3.7.17.b alcanza un pico en torno al valor 3, correspondiente a los estudiantes que presentan una variabilidad adecuada, aunque la distribución está sesgada hacia una concentración alta de los datos. Un segundo pico en torno al valor 2 se asocia con una variabilidad demasiado baja, lo que conduce a interpretar que este grupo no identifica el efecto del tamaño de la muestra, y guía sus respuestas en relación al grado de creencia que tienen sobre el fenómeno, presentando el sesgo denominado enfoque de resultado (Konold, 1989).

Se corresponden con respuestas tales como (7, 7, 7, 7). Este grupo de estudiantes supone el 10,6%. Estos estudiantes están haciendo una interpretación determinista de un fenómeno aleatorio. Por tanto, tienen la creencia de poder controlar el azar, por lo que presentan el sesgo que Langer (1982) denomina ilusión de control. Por otro lado, puede ser que exista una dificultad para identificar la aleatoriedad del fenómeno, debido al conocimiento previo de dicho contexto, si este tipo de enunciados pueden presentarse en el libro de texto, adaptados para trabajar el porcentaje.

El análisis anterior se completa con el diagrama de caja. En primer lugar, destaca que el valor 0 del estadístico se considere como extremo del bigote inferior. No es un resultado extraño pues, como se ha razonado anteriormente, el porcentaje de alumnos que no identifican el azar en este ítem es considerable. Por otro lado, existe un mayor número de respuestas que conceden una variabilidad, puesto que la longitud del tercer cuartil al extremo es mayor que la longitud definida por el extremo inferior y el primer cuartil. Finalmente, se muestra que en este caso son atípicos los estudiantes que conceden un rango excesivamente grande, por ejemplo, el siguiente alumno, A224, quien proporciona los siguientes valores (7, 0, 5, 10).

### 3.7.4.2. COMPARACIÓN POR GRUPO

El análisis de la distribución del valor medio para cada uno de los grupos (Figura 3.7.10.a), revela que el valor esperado para el número de encestes se localiza en torno al valor teórico en los dos grupos. Concretamente, la Tabla 3.7.13, que resume la frecuencia con las que aparecen determinados intervalos, refleja que aproximadamente un 70% de los estudiantes de cada grupo proporciona una cuaterna de valores cuyo valor medio se sitúan dentro del intervalo correcto. Los gráficos de caja muestran una muy alta concentración de valores en torno al valor teórico y muchos valores atípicos, tanto por exceso como por defecto, que repiten las conductas ya analizadas en la muestra global.

Tabla 3.7.13. Porcentaje de estudiantes por grupo según valor medio en el ítem 4

Estimación de la media	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa [6,3-7,7]	64,3	66,2
Estimación aceptable [5,5-8,4] <sup>1</sup>	9,1	7,7
Recencia negativa (<4,5)	3,9	6,3
Recencia positiva (>8,4)	2,6	4,2
Equiprobabilidad (4,5-5,5)	9,1	8,5
No completa	11,0	7,0

<sup>1</sup>Fuera del intervalo normativo

Entre los valores atípicos destacan los asociados a las siguientes respuestas: 2S\_A107: (5, 3, 2, 4), 2P\_A48: (5, 4, 3, 2) y 4S\_A267: (0, 3, 1, 6). Estos estudiantes proporcionan muestras cuyo valor medio asociado al número de encestes es muy bajo. Por tanto, estas respuestas sugieren que los estudiantes no contemplan la información recogida en el enunciado y estarían en el nivel de razonamiento 3 de los cuatro descritos por Valdéz (2016), puesto que su asignación de probabilidades no tiene en cuenta los datos frecuenciales. Además, se identifican respuestas que reflejan una estimación demasiado elevada del número de encestes: 2P\_A30: (8, 9, 10, 10), 4P\_A77: (10, 10, 8, 7) y 4S\_A268: (10, 9, 9, 10). En el primer caso, los estudiantes podrían estar razonando de acuerdo a la recencia negativa y en el segundo a la recencia positiva (Kahneman et al., 1982).



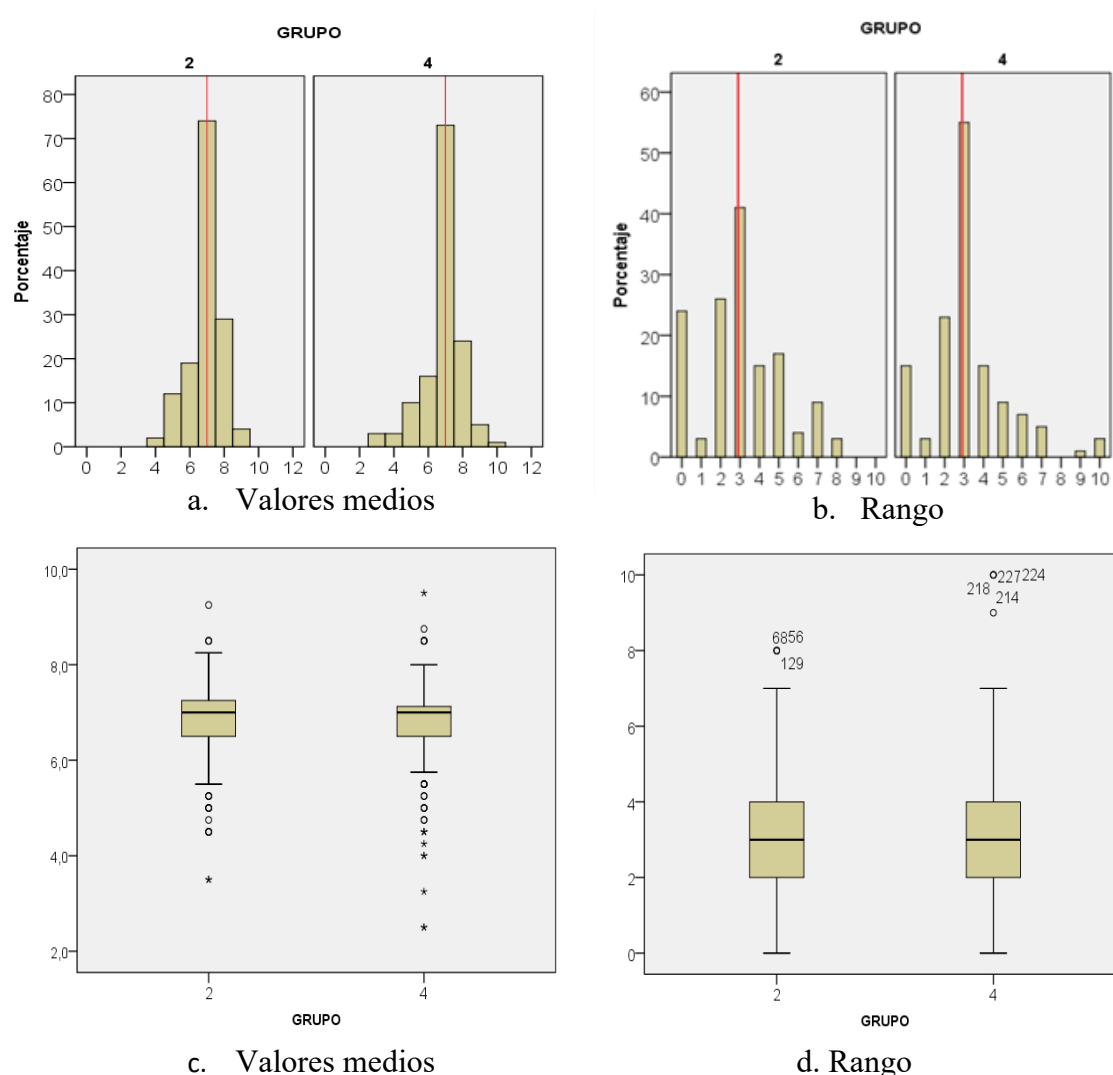


Figura 3.7.10. Distribución del valor medio y rango de cuatro valores del número de aciertos en 10 tiradas a una canasta

### Análisis de la variabilidad

Por otro lado, el estudio de la distribución del rango (Figura 3.7.10.b) revela que una amplia mayoría, en ambos grupos, producen muestras de dispersión normativa o aceptable, aunque algunos también producen muestras con una dispersión baja. La Tabla 3.7.14 indica que 14% del grupo de segundo curso de la ESO y el 9,6% del grupo de cuarto proporcionan respuestas cuyo rango corresponde con una concentración alta, lo que indica una comprensión parcialmente correcta de la Ley de los Grandes Números, puesto que el estudiante espera que los resultados estén próximos al valor teórico. Este resultado puede deberse a que el estudiante guía su razonamiento por su conocimiento previo acerca del fenómeno aleatorio. Es decir, en términos de heurísticas implica la presencia de la heurística de la disponibilidad (Tversky y Kahneman, 1974), que conlleva

que el estudiante subestime la variabilidad intrínseca a la obtención de muestras de una población dada.

Tabla 3.7.14. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 4

Estimación de la variabilidad	2ºESO (n=157)	4ºESO (n=145)
Estimación normativa [2,4]	52,2	64,1
Estimación aceptable [1,6] <sup>1</sup>	15,3	13,1
Variabilidad excesiva (>6)	7,5	6,2
Alta concentración (0)	14,0	9,6
No completa	11,0	7,0

<sup>1</sup>Fuera del intervalo normativo

En el caso del estudio del rango, en el gráfico de la caja se identifican como valores atípicos los asociados a las siguientes respuestas: 2P\_A68: (2, 10, 5, 3), 2P\_A56: (8, 4, 10, 2) y 2S\_A129: (2, 3, 5, 10), 4BP\_A227: (10, 2, 1, 0). 4BP\_A224: (7, 0, 5, 10), 4BP\_A218: (10, 0, 3, 6) y 4BP\_A214: (7, 8, 1, 10), donde el rango de dichas respuestas es mayor que 9, por tanto, la variabilidad es muy elevada. No obstante, como se identifica en la Figura 3.7.10.d, la mayoría de los estudiantes de ambos grupos proporcionan respuestas con un rango menor que 8.

En definitiva, se observa, por un lado, que existe una proporción elevada de estudiantes que producen muestras cuyo valor medio se localiza dentro del intervalo denominado como correcto, además, el estudio de la variabilidad refleja que los valores presentan una alta concentración. Por tanto, las respuestas de los estudiantes revelan que la muestra participante no tiene una comprensión profunda de las dos ideas que caracterizan al concepto de muestreo, es decir, no relacionan los conceptos de representatividad y variabilidad muestral. No obstante, los resultados para este ítem son mejores que los obtenidos para los dos primeros ítems.

Las medias y desviaciones típicas del valor medio y el rango de dicho valor medio en las cuatro estimaciones de cada estudiante en cada uno de los grupos se presentan en la Tabla 3.7.3, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo. No observamos diferencias apreciables en los dos cursos

Para comprobar si las diferencias observadas son estadísticamente significativas se ha realizado la prueba t de diferencia de medias independientes (utilizando el caso de varianzas distintas, que es más conservador). Los resultados se presentan en la Tabla 3.7.4, junto con los intervalos de confianza del 95% para la diferencia de las medias. Observamos que las diferencias no son estadísticamente significativas.

Tabla 3.7.15. Estadísticos de la media y rango en el Ítem 4 por curso

	Curso	N	Media	D. típica.	Error muestreo
I4Media	2°	140	6,716	,9432	,0797
	4°	135	6,633	1,1167	,0961
I4Rango	2°	140	3,06	2,048	,172
	4°	135	3,20	2,003	,172

Tabla 3.7.16. Prueba t de diferencias de medias

	Prueba t para la igualdad de medias				Intervalo de confianza de la diferencia (95%)		
	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. diferencia	E. Inferior	E. Superior
I4Media	,665	273	,507	,08	,12	-,16	,35
I4Rango	-,585	276	,559	-,14	,243	-,62	,34

### 3.8. SÍNTESIS DE RESULTADOS

A continuación, se lleva a cabo una síntesis de los resultados de la aplicación del cuestionario para cada uno de los ítems, presentando en la Tabla 3.8.1 las medias y desviaciones típicas de los resultados teóricamente esperados y observados en cada ítem, tanto globalmente, como por cursos.

Tabla 3.8.1. Síntesis de resultados

Ítem	Teórico		Global		2° curso		4° curso	
	Valor esperado	D. Típica	Valor esperado	D. Típica	Valor esperado	D. Típica	Valor esperado	D. Típica
1 B(100,0,68)	68	4,7	57,9	12,8	56,8	13,7	61,1	8,7
2 B(100, 0,5)	50	5	51,2	7,6	51,9	8,8	50,6	6,4
3 B(10, 0,5)	5	1,6	5,2	0,9	5,1	0,8	5,1	0,9
4 B(10, 0,7)	7	1,45	6,7	1	6,8	0,1	6,6	0,1

Globalmente el estudio de la Tabla 3.8.1 muestra que la estimación del valor esperado es buena, excepto en el primer ítem, tanto globalmente, como por grupo. Pensamos que los estudiantes tienen una buena comprensión o intuición del valor esperado en el muestreo y el fallo en el primer ítem es debido al sesgo de equiprobabilidad (Lecoutre, 1992), latente en parte de los estudiantes, y falta de experiencia con el experimento que consiste en lanzar las chinchetas. A diferencia del resto de los enunciados, en los que se presentan fenómenos aleatorios como lanzar una moneda o jugar al baloncesto, que corresponden con contextos familiares para el estudiante.

El aspecto más relevante, es la falta de comprensión de la variabilidad muestral, mostrado por las diferencias que hay entre las desviaciones típicas teóricas y las observadas. Además, coincidiendo con la investigación de Gómez et al. (2014)

observamos que no existe una intuición simultánea de la convergencia al valor esperado y de la variabilidad muestral. Por tanto, creemos que este es un punto en el que se requiere mejorar la enseñanza. Esta información se completa en la Tabla 3.8.3 y 3.8.4, en que se muestran los intervalos en que se localizan el valor esperado y rango de las cuatro estimaciones por ítem y por grupo.

Tabla 3.8.3. Porcentaje de estudiantes por grupo según intervalos en que se localiza el valor esperado de las cuatro estimaciones en cada ítem

	Ítem 1		Ítem 2		Ítem 3		Ítem 4	
	2ºESO	4ºESO	2ºESO	4ºESO	2ºESO	4ºESO	2ºESO	4ºESO
Estimación normativa del valor esperado	10,8	11,7	31,2	46,2	56,7	67,6	64,3	66,2
Estimación aceptable del valor esperado	17,8	21,4	19,1	23,4	26,8	16,6	9,1	7,7
Recencia negativa	16,6	11,0	16,6	9,7	2,5	3,4	3,9	6,3
Recencia positiva	8,3	8,3	25,5	15,9	5,1	4,1	2,6	4,2
Equiprobabilidad	26,8	23,4					9,1	8,5
Otros valores	13,4	19,3						
No completa	6,4	4,8	7,6	4,8	8,9	8,3	11,0	7,0

Tabla 3.8.4. Porcentaje de estudiantes por grupo según intervalos en que se localiza el rango de las cuatro estimaciones en cada ítem

	Ítem 1		Ítem 2		Ítem 3		Ítem 4	
	2ºESO	4ºESO	2ºESO	4ºESO	2ºESO	4ºESO	2ºESO	4ºESO
Estimación normativa de la variabilidad muestral	12,7	17,9	17,8	17,2	66,9	66,2	52,2	64,1
Estimación aceptable de la variabilidad muestral	7,6	11,7	8,3	9,7	8,9	7,6	15,3	13,1
Estimación de una variabilidad muestral excesiva	62,4	56,6	58,6	59,3	10,2	13,8	7,5	6,2
Alta concentración	10,8	12	8,3	9,7	7,0	4,5	15,3	11
No completa	6,4	2,1	7,0	4,1	7,0	8,3	9,6	6,2

El menor porcentaje de estimaciones normativas se da en el primer ítem y también de la suma de estimaciones normativas y aceptables. Es en este primer ítem donde se producen más frecuencias de sesgos y, en particular, la equiprobabilidad. En los dos primeros ítems los resultados son bastante mejores en los estudiantes de cuarto curso, no siendo tanta la diferencia en los ítems 3 y 4.

Respecto a la Tabla 3.8.4, hay una diferencia notable de estimaciones normativas en el tercer ítem, donde además se igualan los dos grupos. El porcentaje de respuestas

normativas es pequeño en todos los otros ítems e incluso cuando se suman las aceptables. Los estudiantes de cuarto curso son mejores que los de segundo curso, excepto en el último ítem. En los ítems correspondientes a grandes muestras se da una variabilidad excesiva y una excesiva concentración en el último ítem.

### **3.8.1. COMPARACIÓN DE RESULTADOS SEGÚN EL TAMAÑO DE LA MUESTRA**

El primer y segundo ítems hacen referencia a un experimento cuyo número de ensayos es de 100 elementos para cada muestra y en los otros dos las muestras son de 10 elementos. En definitiva, la diferencia en el número de ensayos permite analizar el efecto del tamaño de la muestra en relación a la variabilidad aceptada para la misma.

En este sentido, en las dos primeras tareas, los estudiantes proporcionan muestras con una variabilidad extrema, independiente del curso, siendo la frecuencia mayor en el grupo de segundo curso de la ESO. Esta observación se desprende, a su vez, de los datos resumidos en la Tabla 3.8.4, especialmente para el primer ítem. Por tanto, se deduce que los estudiantes no presentan una adecuada comprensión de la variabilidad cuando el tamaño de la muestra es elevado, mientras las muestras elaboradas para los dos últimos ítems presentan una variabilidad adecuada, con pocas diferencias en ambos cursos.

En resumen, estos resultados reflejan que los estudiantes conceden a las grandes muestras más variabilidad que la teórica y que a las muestras pequeñas, lo que es contrario a la teoría estadística y coincide con lo encontrado por Serrano (1996). Es posible que ello encubra, en algunos estudiantes, tanto la heurística de la representatividad (en su estimación en las muestras pequeñas) como la representatividad (en su respuesta a las muestras grandes).

Por otro lado, aunque la mayoría de las respuestas elaboradas proporcionan valores cuya media se localiza dentro del intervalo definido como correcto, no existe una adecuada coordinación de las dos ideas fundamentales: representatividad muestral y variabilidad muestral. En definitiva, los estudiantes no comprenden el efecto del tamaño de la muestra sobre la variabilidad en el muestreo, resultado que aparece en la investigación de Serrano (1996), Gómez (2014) y Gómez et al. (2014). Además, los resultados obtenidos se sitúan en la línea de las observaciones o resultados aportados por las investigaciones revisadas en la literatura, puesto que ponen en relieve la dificultad que los estudiantes presentan en la comprensión de la variabilidad intrínseca al proceso de

muestreo.

### **3.8.2. COMPARACIÓN DE RESULTADOS EN EXPERIMENTOS CON SUCESOS EQUIPROBABLES O NO**

Con el objetivo de examinar si los estudiantes son capaces de identificar la equiprobabilidad o no del fenómeno aleatorio, se han diseñado cuestiones que permitan analizar si se identifica la asimetría del fenómeno presentado. Por ello, los fenómenos presentados en el primer y último ítem no cumplen el principio de equiprobabilidad. Desde el análisis de los datos para el segundo y tercer ítem, se observa que la muestra percibe la equiprobabilidad del fenómeno aleatorio, que es el mismo en ambas cuestiones.

Por el contrario, el análisis de la distribución del valor medio de los datos recogidos para el primer ítem revela que en ambos grupos una proporción de estudiantes no identifican la asimetría del dispositivo, debido al sesgo de equiprobabilidad o a la heurística de la representatividad, resultado que aparece en los trabajos de Gómez (2014), Gómez et al. (2014) y Serrano (1996). Esto se muestra en la tendencia de compensar el conjunto de respuestas mediante valores donde el número de chinchetas que caen con la punta hacia arriba es menor, donde no se considera el resultado del profesor.

En el último ítem, la distribución del valor medio en el muestreo refleja que los estudiantes identifican la no equiprobabilidad del experimento. En este caso, el conocimiento o la familiaridad del contexto favorecen la comprensión del fenómeno aleatorio. Pero la escasa variabilidad indica que algunos estudiantes dan al fenómeno una interpretación prácticamente determinista.

## **3.9. ESTUDIO DE CASOS ATÍPICOS**

### **3.9.1. ESTUDIO DE CASOS CON AUSENCIA DE VARIABILIDAD**

En los apartados anteriores se ha destacado el hecho de que un porcentaje de estudiantes muestra una percepción incompleta del efecto del tamaño de la muestra sobre la variabilidad de la misma. En este apartado, analizamos con más detalle los casos de los estudiantes que proporcionan una respuesta atípica, en el sentido de no dotar a las muestras proporcionadas de alguna variabilidad, es decir, mostrar una concepción determinista del muestreo. Estos estudiantes estarían en el nivel N2 de transición entre cuatro posibles descritos por Moreno y Vallecillos (2001) sobre el muestreo, y a nivel N1, también sobre cuatro posibles sobre el concepto de aleatoriedad, descrito por Valdéz

(2016), al realizar una predicción determinista de la muestra.

Tabla 3.9.1. Casos con ausencia de variabilidad en el Ítem 1

Caso extremo.	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2°ESO)	Frec.	% (respecto al total de estudiantes de 4° ESO)
(50, 50, 50, 50)	7	4,5	9	6,2
(68, 68, 68, 68)	6	3,8	2	1,4
(70, 70, 70, 70)	1	0,6	2	1,4
(80, 80, 80, 80)			1	0,7
(100, 100, 100, 100)	2	1,3		
Total	16	10,2	14	9,7

Las tablas presentadas en esta sección reflejan aquellos estudiantes que responden a la tarea con una cuaterna caracterizada por la ausencia de variabilidad. En la Tabla 3.9.1, que corresponde al primer ítem, observamos casos que reproducen el valor dado en el enunciado (68) o uno cercano (70 u 80), pero también otros que asignan el valor 50, lo que refleja claramente el sesgo de equiprobabilidad descrito por Lecoutre (1992). Dos estudiantes asignan un valor 100, resultado muy improbable, incluso en una muestra de cuatro elementos, mostrando poca familiaridad con el muestreo.

En el Ítem 2 (Tabla 3.9.2), los casos extremos coinciden en dar un valor igual o próximo al esperado teóricamente, es decir al 50% de caras. En el primer caso se muestra una concepción determinista del muestreo, y no se diferencia entre la probabilidad teórica y un valor probable. En el caso de que el valor sea próximo, pero no idéntico al esperado, aceptan la variabilidad del muestreo respecto al valor teórico, pero no parecen comprender la variabilidad de las diferentes muestras entre sí.

Tabla 3.9.2. Casos con ausencia de variabilidad en el Ítem 2

Caso extremo. Ausencia de variabilidad	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2° de ESO)	Frec.	% (respecto al total de estudiantes de 4° de ESO)
(50, 50, 50, 50)	7	4,5	10	6,9
(53, 53, 53, 53)	4	2,6	2	1,4
(47, 47, 47, 47)	1	0,6		
Total	12	7,7	12	8.3

En el tercer ítem encontramos pocos casos con ausencia de variabilidad, por lo que parece que el experimento es más familiar para el estudiante, posiblemente porque los libros de texto presentan ejemplos de experimentos de lanzamientos de monedas con un tamaño pequeño de muestra. Los casos de ausencia de variabilidad, coinciden con el valor teórico u otro próximo (solo un estudiante), de modo que los estudiantes visualizan el proceso en forma determinista.

Tabla 3.9.3. Casos con ausencia de variabilidad en el Ítem 3

Caso extremo. Ausencia de variabilidad	2ºESO		4ºESO	
	Frec.	% (respecto al total de estudiantes de 2ºESO)	Frec.	% (respecto al total de estudiantes de 4ºESO)
(5, 5, 5, 5)	7	4,46	6	4,14
(6, 6, 6, 6)	1	0,64		

Tabla 3.9.4. Casos con ausencia de variabilidad en el Ítem 4

Caso extremo. Ausencia de variabilidad	2ºESO		4ºESO	
	Frec.	% (respecto al total de estudiantes de 2ºESO que realiza la prueba)	Frec.	% (respecto al total de estudiantes de 4ºESO que realiza la prueba)
(7, 7, 7, 7)	19	12,18	13	8,96
(6,6,6,6)	1			
(5,5,5,5)	1	0,64	1	0,68
(8,8,8,8)	1	0,64	1	0,68
Total	22	14,1	15	10,32

Por último, encontramos muchos más casos de ausencia de variabilidad en los dos grupos en el último ítem, sobre lanzamiento a una canasta de baloncesto, puesto que algunos estudiantes lo han interpretado en forma determinista. Los estudiantes que muestran ausencia de variabilidad repiten el valor del enunciado o dan valores próximos. También hay dos estudiantes que muestran la equiprobabilidad (Lecoutre, 1992), no teniendo en cuenta la información frecuencial del enunciado.

Los estudiantes que muestran una concepción determinista del muestro, suelen repetir los resultados de la primera muestra en el resto. Este resultado contradice la sugerencia de Saldahna y Thompson (2002), quienes creen que los estudiantes no son capaces de imaginar muestras que compartan los mismos elementos. Estarían, además, en el nivel más bajo de comprensión del concepto de distribución muestral, según Ko (2016), pues pensarían que diferentes muestras comparten el mismo valor del estadístico (en este caso número de éxitos en el experimento).

Los estudiantes que en los ítems 1 y 4 insisten en repetir muestras con el 50% de resultados de cada tipo, muestran una falta de comprensión y aplicación de la probabilidad frecuencial, pues no han sido capaces de deducir esta probabilidad del enunciado de las tareas propuestas. Coinciden estos casos –afortunadamente pocos- con los resultados de la investigación de Green (1983a y 1983b), Cañizares (1997) y Gómez, Batanero y Contreras (2014).



### 3.9.2. ESTUDIO DE CASOS CON VARIABILIDAD EXCESIVA

Realizamos un estudio similar para los casos de estudiantes que otorgan una variabilidad excesiva en sus estimaciones y que mostramos en las tablas que siguen.

Tabla 3.9.5. Casos con variabilidad extrema en el Ítem 1

Caso extremo.	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2°ESO)	Frec.	% (respecto al total de estudiantes de 4° ESO)
(50, 49, 99, 1)	1	0,64		
(59, 41, 3, 82)	1	0,64		
(87, 46, 3, 40)	1	0,64		
(99, 1, 50, 60)			1	0,68
(85, 60, 20, 3)			1	0,68
(78, 44, 3, 89)			1	0,68
Total	3	1,92	3	2,04

Así, en el ítem 1 (Tabla 3.8.9) encontramos estudiantes con rango superior a 80, que incluyen en sus cuaternas valores tan poco probables como 1 o 99 y, en general, valores que abarcan los extremos y el centro de la distribución. Se muestra la idea de azar como proceso equitativo, donde cualquier resultado sería igualmente probable y son menos estudiantes que los que muestran la concepción determinista de la aleatoriedad.

Es también notable el hecho de que, algunos estudiantes, después de dar un valor extremo (por ejemplo 99) proporcionan en la siguiente muestra otro extremo en sentido contrario (en el ejemplo 1) de forma que la media de los dos valores sea el 50%. Veríamos acá una conducta tipificada como la falacia del jugador (Tversky y Kahneman, 1974), donde se trata de compensar resultados por encima y debajo de la media, lo que implica falta de comprensión de la independencia de ensayos sucesivos.

Tabla 3.9.6. Casos con variabilidad extrema en el Ítem 2

Caso extremo.	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2°ESO)	Frec.	% (respecto al total de estudiantes de 4° ESO)
(32, 16, 48, 97)	1	0,64		
(51, 99, 32, 17)	1	0,64		
(27, 13, 96, 51)	1	0,64		
(55, 10, 40, 98)	1	0,64		
(90, 89, 16, 1)	1	0,64		
(50, 65, 20, 100)			1	0,68
(15, 95, 97, 26)			1	0,68
(1, 99, 49, 51)			1	0,68
(50, 100, 0, 75)			1	0,68
(0, 5, 80, 50)			1	0,68
Total	5	3,2	5	3,4

Cinco estudiantes en cada grupo proporcionan muestras demasiado variables (Tabla 3.8.10) en el ítem 2, donde en algunos casos se incluye en la cuaterna los valores 0 y 100

simultáneamente, a pesar de la poca probabilidad que tendría obtener una de estas muestras y mucho menos cuatro con estas características. Se observa de nuevo en algunos casos la falacia del jugador (Tversky y Kahneman, 1974) que supone falta de comprensión de la independencia en el muestreo. Indica también, para el caso de las muestras grandes, que no se visualiza el teorema central del límite, en el cual, los valores centrales de la distribución son más probables que los extremos (Méndez, 1991).

Sorprendentemente, en los ítems 3 y 4 sólo los estudiantes de cuarto curso tienen cuaternas con un rango igual a 10, que corresponde a una variabilidad excesiva y son los mismos estudiantes los que presentan sucesos de muy baja probabilidad, mientras que no encontramos estos casos en los estudiantes de segundo.

Tabla 3.9.7. Casos con variabilidad extrema en el Ítem 3

Caso extremo.	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2°ESO)	Frec.	% (respecto al total de estudiantes de 4° ESO)
(10,0, 5,8)			1	0,68
(10,0, 5,4)			1	0,68
(5,0,10,6)			1	0,68
(3,10, 5,0)			1	0,68
(5,10,0,7)			1	0,68
Total			5	3,4

Tabla 3.9.8. Casos con variabilidad extrema en el Ítem 4

Caso extremo.	2°ESO		4°ESO	
	Frec.	% (respecto al total de estudiantes de 2°ESO)	Frec.	% (respecto al total de estudiantes de 4° ESO)
(10, 0, 3, 6)			1	0,68
(7, 0, 5, 0)			1	0,68
(10, 2, 1, 0)			1	0,68
Total			3	2,04

### 3.10. CONCLUSIONES DEL ESTUDIO 1

#### 3.10.1. CONCLUSIONES RESPECTO A LOS OBJETIVOS

Este estudio tuvo como finalidad analizar y evaluar la comprensión que un grupo de estudiantes de la etapa obligatoria de secundaria presenta sobre algunas ideas asociadas al concepto de muestreo. Más concretamente, se propuso a una muestra de estudiantes de segundo y cuarto curso de la educación secundaria obligatoria algunas tareas en que deben estimar un valor probable en poblaciones binomiales para evaluar la comprensión intuitiva del estudiante de las siguientes propiedades del muestreo, cuando se conoce la proporción en una población:

3. Proporción muestral esperada y relación que establece el estudiante entre dicha proporción muestral y la proporción en la población.
4. Variabilidad de la proporción en el muestreo y efecto del tamaño de la muestra sobre dicha variabilidad.

Los objetivos específicos de este estudio desarrollan el tercer objetivo general de la investigación, descrito en el Capítulo 1, que pretendía llevar a cabo una serie de estudios de evaluación de la comprensión del muestreo por parte de los estudiantes. A continuación, analizamos los objetivos específicos del Estudio 1, deducidos de dicho objetivo general.

*Objetivo 1.1. Evaluar y analizar la comprensión de los estudiantes de segundo y cuarto curso de la educación secundaria obligatoria de la relación entre el valor de la proporción en una población y en una muestra.*

Pensamos que este objetivo se ha cumplido por medio del estudio realizado, donde nos hemos centrado en la estimación de una proporción, porque es un estadístico muy sencillo y aparece en muchas situaciones de la vida diaria que son comprensibles para el alumnado. Además, dicho estadístico también ha sido utilizado en otras investigaciones previas, como las de Green (1983a), Gómez et al. (2014), Moreno y Vallecillos (2001) o Shaughnessu et al. (1999), con cuyos resultados hemos podido comparar con los nuestros y nos han permitido analizar la presencia de algunos sesgos de razonamiento.

Para la evaluación de dichos contenidos hemos considerado la aplicación de un cuestionario, que se describe de manera más detallada en la Sección 3.4, indicando los criterios seguidos en su elaboración y las variables que se tienen en cuenta. Se recogieron datos de 302 estudiantes, la mitad aproximadamente de cada uno de los cursos, segundo y cuarto de educación secundaria obligatoria, y se analizaron sus respuestas. Además, se ha utilizado una muestra amplia que abarca varios grupos y escuelas en los dos cursos, lo que le da una mayor posibilidad de generalización.

Para estudiar la comprensión mostrada por los estudiantes de la muestra de la relación entre la proporción poblacional y la muestral, se calculó el valor medio de los cuatro valores probables dados por los estudiantes en sus estimaciones. La distribución de este estadístico se estudió globalmente y por grupo, por medio de histogramas o diagramas de barra y gráficos de caja, así como la comparación de las frecuencias con que los

estudiantes producen valores medios en ciertos intervalos normativos, aceptables o que indican sesgos de razonamiento. Todos estos datos se compararon con los valores teóricos dados al analizar las distribuciones teóricas de las medias en las muestras de 4 elementos, sus medias y desviaciones típicas. Los intervalos se determinaron utilizando la distribución muestral de la media de una muestra de tamaño 4, tomada de la distribución de partida. El intervalo normativo sería el que contiene el 68% de los casos centrales en dicha distribución y el aceptable el que contiene el 95% de los casos.

Se analizaron con detalle las características deducidas de estos análisis, así como los casos de respuestas atípicas, tanto por ausencia de variabilidad o concepción determinista del muestreo como por variabilidad excesiva. En consecuencia, pensamos se ha alcanzado razonablemente el objetivo propuesto.

*Objetivo 1.2. Evaluar y analizar la comprensión de los estudiantes de segundo y cuarto curso de la educación secundaria obligatoria de la variabilidad de la proporción muestral y del efecto del tamaño de la muestra sobre dicha variabilidad.*

Cada uno de los ítems del cuestionario pide al estudiante generar cuatro muestras diferentes de la población dada, lo que nos permite analizar la variabilidad de dichas estimaciones. Para estudiar la comprensión de variabilidad muestral se calculó el rango de los cuatro valores proporcionados en las estimaciones de los estudiantes. La distribución de este estadístico se estudió globalmente y por grupo, por medio de un histograma o diagrama de barras y gráficos de caja, así como analizando las frecuencias con que los estudiantes producen valores medios en ciertos intervalos normativos, aceptables o que indican sesgos de razonamiento.

Estos intervalos se calcularon simulando la distribución muestral del rango de una muestra de cuatro elementos de la distribución binomial, que se simuló para esta investigación, determinando sus percentiles y otros estadísticos. Se analizaron con detalle las características deducidas de estos análisis, así como los casos de respuestas atípicas. Por tanto pensamos se ha logrado en un grado razonable este objetivo.

*Objetivo 1.3. Comparar los resultados en función de algunas variables de tarea y por curso.*

Para lograr este objetivo, se realizaron diferentes comparaciones de los datos a lo largo y al final del capítulo. En primer lugar, para cada ítem se compararon por grupo

gráficamente la distribución de la media y del rango de las cuatro estimaciones. Además, las frecuencias de estudiantes clasificados en diferentes intervalos considerados normativos, aceptables o indicativos de sesgos se compararon informalmente. Además, las medias de dichas distribuciones se compararon mediante el contraste  $t$  de diferencias de medias. También se comparó la puntuación total en el cuestionario y las puntuaciones indicativas de comprensión del valor esperado y de la variabilidad en el muestreo.

Por otro lado, en la Sección 3.8 se comparan los valores medios teóricos y observados en cada ítem, así como el porcentaje de estudiantes en diversos intervalos, lo que permite obtener conclusiones en función de los valores  $n$  y  $p$  en la distribución binomial  $B(n, p)$ . Al final del capítulo se incluye una comparación específica de los resultados en experimentos equiprobables o no y en experimentos con pequeñas y grandes muestras.

### 3.10.2. CONCLUSIONES RESPECTO A LAS HIPÓTESIS INICIALES

Desde el interés por analizar la comprensión que un grupo de estudiantes presenta sobre diferentes ideas asociadas al concepto de muestreo, surgen una serie de hipótesis que se describen más detalladamente al inicio del capítulo. Tras la aplicación del cuestionario y el análisis de los datos obtenidos, es relevante reflexionar sobre los resultados obtenidos en relación con las hipótesis descritas previamente.

*H2.1. Algunos estudiantes presentan una comprensión insuficiente de las ideas básicas asociadas al concepto de muestreo evaluadas en el cuestionario.*

Esta hipótesis se cumple en parte. Por un lado, los resultados fueron generalmente buenos en la estimación del valor esperado, donde una parte importante de los estudiantes mostraron una buena percepción de la proximidad entre la proporción esperada en la muestra y la proporción en la población. Es decir, se observó una buena capacidad de estimar el valor esperado en la distribución muestral, ya que la media de las cuatro estimaciones producidas por los estudiantes se acerca bastante, en general, al valor teórico. Ello indica que una parte importante de los estudiantes alcanza al menos el nivel proporcional de razonamiento sobre muestreo en la categorización de Shaughnessy et al. (2004) y un nivel intermedio en la de Valdéz (2016).

Sin embargo, especialmente en los ítems 1 y 4, un grupo de estudiantes, aunque no muy numeroso, mostró el sesgo de equiprobabilidad, que consiste en estimar siempre frecuencias esperadas cercanas al 50% independientemente de la proporción en la

población. En particular en el primer ítem, aparecen dos modas en la distribución de las medias de las cuatro estimaciones. Mientras la primera corresponde al valor teórico, la segunda se centra en el 50% y nuestra interpretación es que algunos estudiantes razonan de acuerdo con el sesgo de equiprobabilidad (Lecoutre, 1992) e inconscientemente la reflejan en sus estimaciones. Además, en este ítem y en el cuarto, la muestra de estudiantes presenta tanto el sesgo de equiprobabilidad como la heurística de representatividad. Esta última se identifica en aquellos casos en los que el estudiante trata de compensar los resultados, proporcionando muestras en las que el número de chinchetas que caen con la punta hacia abajo es mayor. Esta heurística conlleva a la estimación errónea de la frecuencia. Aunque estos resultados mejoran considerablemente en la última de las cuestiones, debido a la familiaridad del estudiante con la situación problemática planteada. El análisis de los datos, revela también que la estimación es mejor, cuando el fenómeno aleatorio presentado cumple la condición de equiprobabilidad, independiente del tamaño de la muestra.

En general, se ha encontrado una pobre percepción de la variabilidad y al contrario de lo supuesto en otros estudios, la percepción es peor en lo que se refiere a las muestras de gran tamaño, en la que los estudiantes reproducen una variabilidad excesiva. Estos resultados coinciden con los obtenidos en Gómez et al. (2014) y contradicen los de Shaughnessy et al. (2004). Por tanto, el estudio realizado refleja que la mayoría de los estudiantes tiene una comprensión insuficiente de la variabilidad intrínseca al proceso de muestreo, que se acentúa en aquellas muestras cuyo tamaño es mayor. En estos ítems, la mayoría de las respuestas presentan muestras cuyos valores quedan caracterizados por una variabilidad excesiva.

Se observa mejor comprensión de la variabilidad en los ítems 3 y 4, que corresponden a muestras pequeñas, contrario a lo sugerido por Shaughnessy et al. (2004), seguramente por el tipo de tarea considerado por estos autores, que no permite analizar adecuadamente la variabilidad. También encontramos la presencia de respuestas en la que se identifica una ausencia de la misma o exceso, que hemos analizado describiendo diferentes sesgos.

En general, una proporción importante de estudiantes no llega a alcanzar los niveles superiores de razonamiento sobre el muestreo descritos por Moreno y Vallecillos (2001), puesto que no aprecian el efecto del tamaño de la muestra sobre la variabilidad muestral. En consecuencia, los estudiantes razonan a nivel proporcional, de acuerdo con Shaughnessy, Ciancetta y Canada (2004), pero no todos llegan a alcanzar el nivel de

razonamiento distribucional sobre muestreo.

Se estudiaron también las correlaciones entre las estimaciones de los valores medios, por un lado y de los rangos por otro, mostrándose una relación directa y de intensidad moderada en la mayoría de los ítems. Lo que sugiere que los estudiantes que muestran un sesgo (por ejemplo, falta o exceso de variabilidad) en un ítem, lo suelen mostrar en los restantes. Los resultados no son, por tanto, tan dependientes del instrumento de evaluación, sino que reflejan bien las concepciones de los estudiantes.

*H3.1. El análisis de las respuestas del instrumento de evaluación refleja una mejor comprensión y razonamiento sobre los contenidos evaluados en los estudiantes del cuarto curso.*

La muestra está constituida por estudiantes de educación secundaria obligatoria. Además, los individuos que conforman la muestra cursan segundo o cuarto curso de la ESO. Por tanto, se ha considerado relevante para la investigación estudiar si existe una diferencia significativa en las respuestas obtenidas a los diferentes ítems entre los dos grupos que configuran la muestra.

Se esperaba que los estudiantes de cuarto curso obtuvieran mejores resultados, debido a factores tales como la madurez y su conocimiento, y el análisis de las respuestas revela que se han encontrado algunas diferencias:

- En el primer ítem disminuye la proporción de estudiantes de cuarto curso que muestran la equiprobabilidad, recencia positiva o negativa, o que no completan la tarea, y aumenta la proporción de los que realizan una estimación normativa o aceptable de la media. También aumenta la proporción de estudiantes con variabilidad normativa o aceptable, aunque la estimación excesiva de la variabilidad sigue siendo muy grande, más de la mitad de la muestra. La diferencia en la estimación de las medias no es estadísticamente significativa, pero sí lo es la de los rangos.
- En el segundo ítem disminuye la proporción de estudiantes de cuarto curso con valores fuera del rango aceptable y aumenta la proporción de los que realizan una estimación normativa o aceptable de la media. No hay apenas variación en la estimación de la variabilidad, que es en general excesiva. No hay diferencias estadísticamente significativas, ni en la estimación de la media ni en la de los rangos.
- Tampoco se observan diferencias estadísticamente significativas en la estimación de la media o el rango en el ítem 3 y 4, aunque en este caso, la mayoría de los estudiantes

realiza una estimación aceptable e incluso óptima de los dos valores, de modo que muestran un razonamiento distribucional.



## CAPÍTULO 4.

# ESTUDIO 2. JUSTIFICACIÓN DE ESTUDIANTES DE BACHILLERATO EN TAREAS ELEMENTALES DE MUESTREO

- 4.1. Introducción
- 4.2. Objetivos e hipótesis del Estudio 2
- 4.3. Contexto educativo
- 4.4. Metodología
  - 4.4.1. Cuestionario utilizado
  - 4.4.2. Descripción de la muestra
  - 4.4.3. Método de recogida de datos
- 4.5. Análisis cuantitativo
  - 4.5.1. Método de análisis
  - 4.5.2. Resultados en el ítem 1
  - 4.5.3. Resultados en el ítem 2
  - 4.5.4. Resultados en el ítem 3
  - 4.5.5. Resultados en el ítem 4
  - 4.5.6. Resultados comparados por ítem
  - 4.5.7. Relación entre diversas respuestas
  - 4.5.8. Puntuaciones totales y parciales en el cuestionario
- 4.6. Análisis cualitativo de las argumentaciones
  - 4.6.1. Método de análisis
  - 4.6.2. Categorías de análisis de las argumentaciones
  - 4.6.3. Resultados en el ítem 1
  - 4.6.4. Resultados en el ítem 2
  - 4.6.5. Resultados en el ítem 3
  - 4.6.6. Resultados en el ítem 4
  - 4.6.7. Síntesis de argumentos en diferentes ítem
  - 4.6.8. Estudio de casos con argumentos recurrentes
- 4.7. Conclusiones del Estudio 2
  - 4.7.1. Conclusiones respecto a los objetivos
  - 4.7.2. Conclusiones respecto a las hipótesis

### 4.1. INTRODUCCIÓN

En este capítulo se presentan los resultados correspondientes a la aplicación de una versión modificada del cuestionario utilizado en el Estudio 1 a una muestra de estudiantes de segundo curso de bachillerato. Las conclusiones obtenidas a partir del análisis de los datos obtenidos en el estudio anterior, así como las diferencias evidentes en las respuestas de los participantes en el mismo, nos llevaron a realizar un análisis y reflexión crítica de dicho cuestionario. Por un lado, las respuestas dadas por los estudiantes de educación secundaria obligatoria (ESO) se restringieron a sus cuatro estimaciones del número

esperado del suceso en diferentes muestras de las poblaciones propuestas. Por tanto, sus respuestas solo permitían elaborar un estudio cuantitativo de los datos. Este aspecto limitaba nuestra capacidad de análisis sobre la comprensión que los estudiantes mostraban de los conceptos objetos de investigación.

En este nuevo estudio se decidió pasar el mismo instrumento de evaluación a estudiantes de bachillerato, con la finalidad de comprobar si existe una tendencia de mejora del razonamiento con la edad. Además, se consideró la posibilidad de demandar al estudiante la justificación de su respuesta, ya que esperábamos que los estudiantes de este curso presentasen un grado de maduración mayor, por lo que la tarea de justificación puede no resultar tan compleja como en los cursos de educación secundaria obligatoria.

En este capítulo se describen de manera detallada todos los aspectos que contextualizan el Estudio 2, realizado con una muestra de estudiantes de segundo curso de bachillerato. En las siguientes secciones se describen los elementos que nos permiten enmarcar el estudio, es decir, se detalla el contexto educativo asociado a la muestra participante. Por otro lado, se describen la composición de la muestra, la modificación realizada al cuestionario y la metodología que guía el análisis de las respuestas obtenidas a través de la aplicación del instrumento de evaluación. Finalmente, se realiza un análisis de los datos con las respuestas dadas por los estudiantes. Algunos resultados de este estudio se han publicado en Begué, Gea, Batanero y Beltrán (2017 y 2018).

#### **4.2. OBJETIVOS E HIPÓTESIS DEL ESTUDIO 2**

Los objetivos específicos de este estudio son también consecuencia del tercer objetivo general de la investigación, descrito en el Capítulo 1 y, además, completan los presentados en el Estudio 1. Más concretamente, se desea en este estudio completar el anterior, utilizando las mismas tareas de generación de muestras de poblaciones binomiales, para evaluar la comprensión intuitiva de los estudiantes de bachillerato de las siguientes propiedades del muestreo, cuando se conoce la proporción en una población:

1. Proporción muestral esperada y relación que establece el estudiante entre dicha proporción muestral y la proporción en la población.
2. Variabilidad de la proporción en el muestreo y efecto del tamaño de la muestra sobre dicha variabilidad.

Además, se desean analizar las razones que los estudiantes tienen para dar sus

muestras, por medio de sus argumentos. De ello, se deducen los siguientes objetivos específicos del Estudio 2.

*Objetivo 1.4. Evaluar y analizar la comprensión de la relación entre el valor de una proporción en la población y la proporción muestral esperada en muestras tomadas de dicha población en estudiantes de bachillerato y comparar con los resultados obtenidos previamente en estudiantes de educación secundaria obligatoria.*

Para la evaluación de dichos contenidos hemos considerado la aplicación del mismo cuestionario utilizado en el Estudio 1. A lo largo del capítulo informamos de los resultados con la nueva muestra y los comparamos con los obtenidos en los estudiantes de ESO.

*Objetivo 1.5. Evaluar y analizar la comprensión de la variabilidad en el muestreo y el efecto del tamaño de la muestra sobre dicha variabilidad en estudiantes de bachillerato y comparar con los resultados obtenidos previamente en estudiantes de educación secundaria obligatoria.*

Cada uno de los ítems del cuestionario pide al estudiante generar cuatro muestras diferentes de la población dada, o mejor dicho sus estimaciones del valor esperado en las mismas, lo que nos permite analizar la variabilidad de dichas estimaciones. Igualmente se presentan estos resultados en los estudiantes de bachillerato y se realizan comparaciones con los resultados del Estudio 1 a lo largo del capítulo.

*Objetivo 1.6. Analizar los argumentos que los estudiantes emplean para apoyar la elección de sus respuestas.*

Esta será una contribución original respecto al Estudio 1, donde no se analizaron tales argumentos. Para conseguirlo, se pidió a una parte de la muestra argumentar sus respuestas al cuestionario. Dichos argumentos se clasifican y se analizan a lo largo del capítulo, teniendo en cuenta los antecedentes de la investigación, resumidos en el Capítulo 2.

## **Hipótesis del Estudio 2**

Como consecuencia de los objetivos planteados y, teniendo en cuenta los resultados

de la investigación previa y del Estudio 1, es posible establecer algunas hipótesis sobre lo que se espera encontrar como resultado del Estudio 2. Algunas hipótesis son especificaciones de las hipótesis generales  $H2$  y  $H3$  descritas en el Capítulo 1.

*H2.2. Algunos estudiantes presentan una comprensión insuficiente de las ideas básicas asociadas al concepto de muestreo.*

Esperamos que la mayoría de estudiantes tenga una buena percepción de la proximidad entre la proporción esperada en la muestra y la proporción en la población, ya que trabajos relacionados con la heurística de la representatividad (Tversky y Kahneman, 1982) indican que los estudiantes incluso sobreestiman esta característica, como también se muestra en las investigaciones sobre comprensión de la aleatoriedad de Green (1983a y 1983b) o Serrano (1996).

También se espera encontrar estudiantes que razonan de acuerdo al sesgo de equiprobabilidad, descrito por Lecoutre (1992), donde se estiman siempre frecuencias esperadas cercanas al 50%, independientemente del valor que tome la proporción en la población. Esperamos más dificultad en la estimación de la variabilidad muestral y la comprensión del efecto del tamaño de la muestra sobre la variabilidad. Estos resultados se observaron en el Estudio 1, así como en los trabajos de Gómez et al. (2014) y en Shaughnessy et al. (2004), por lo que podríamos considerar que se repetirán en el nuestro, como ocurrió en el trabajo de Moreno y Vallecillos (2001).

*H3.2. El análisis de las respuestas del instrumento de evaluación refleja un progreso gradual de comprensión de los contenidos evaluados en los estudiantes de bachillerato, respecto a los de educación secundaria obligatoria.*

Esperamos que por su mayor conocimiento y madurez, los estudiantes de bachillerato tengan mejores resultados que los de la ESO, puesto que en el Estudio 1 se observó mejor rendimiento en los estudiantes de curso más avanzado. Este efecto de la edad en el mejor desempeño en cuestionarios relacionados con la estadística y probabilidad se ha observado en otros estudios como los de Batanero et al. (1996) y Serrano et al. (1998).

Puesto que además se incluye un estudio de los argumentos de los estudiantes, se ha planteado la siguiente hipótesis, sobre lo que se espera encontrar en el análisis de tales argumentos:

*H4. Se espera encontrar una variedad de argumentos para apoyar las respuestas de los estudiantes. En concreto, podemos identificar argumentos encontrados en investigaciones previas como la de Serrano (1996). Por otra parte, esperamos analizar la concepción de la probabilidad presente en la respuesta del estudiante.*

En relación con esto último, consideramos que las argumentaciones pueden reflejar la creencia que alumno presenta acerca de la aleatoriedad, así como la identificación de sesgos aplicados en su razonamiento, los cuales se han presentado en el Capítulo 2. Por ejemplo, podrían reflejar creencia en la equiprobabilidad de resultados aleatorios o un excesivo énfasis en el determinismo. En concreto, esperamos encontrar algunos argumentos que reflejen diferentes concepciones sobre el azar y la probabilidad descritas en trabajos como los de Batanero (2016) y Batanero y Serrano (1999).

### **4.3. CONTEXTO EDUCATIVO**

El contexto educativo de las muestras analizadas ha cambiado con respecto al Estudio 1, debido a que en este cuestionario participan nuevos centros a los considerados en el estudio anterior. En concreto, colaboran tres nuevos centros educativos en este estudio, cuyas características se describen a continuación, con la finalidad de presentar el contexto de los centros y, en definitiva, de su alumnado.

Por un lado, ha participado un centro público en el que se oferta una vía de bachillerato denominado “bachillerato internacional”, el cual se caracteriza por considerar una programación más completa, así como un nivel de profundización en el contenido mayor, que se traduce en una evaluación más exigente en relación con el resto de grupos que configuran el bachillerato.

Por otro lado, se ha contado con dos centros en régimen de concierto, por tanto, el bachillerato que ofertan es privado. En relación con uno de los colegios, es reseñable el nivel de preparación de su alumnado, que viene destacando en las pruebas de acceso a la universidad en relación con la media esperada a nivel autonómico y estatal.

### **4.4. METODOLOGÍA**

Aunque el cuestionario es una adaptación del aplicado en el Estudio 1, se trata de una investigación exploratoria porque se realiza por primera vez un estudio con estudiantes de mayor edad, así como un análisis de sus argumentos.

Este análisis de argumentos sigue una metodología cualitativa, puesto que se basa en datos no cuantitativos, y no se efectúa una medición numérica (Hernández et al., 2010). Más concretamente, el estudio de argumentos se realiza a través del análisis de contenido, que es un método específico orientado a investigar con documentos (por ejemplo, libros de texto o respuestas abiertas de los estudiantes). Zapico (2007) menciona que el análisis de contenido permite averiguar la naturaleza del discurso utilizado en documentos escritos. Según Escorcía (2010), esta metodología permite acercarse a temas poco conocidos, es decir, es aplicable en los estudios de tipo exploratorio.

Por otro lado, se repite el análisis cuantitativo realizado en el Estudio 1 ya que, se mantienen el tipo de preguntas que configuraban el primer cuestionario. Dicho análisis cuantitativo permite realizar la comparación de las respuestas de los alumnos de último curso de bachillerato con las correspondientes a las de los estudiantes de educación secundaria obligatoria. En este capítulo se presentan de manera detallada los dos tipos de análisis previamente descritos.

En conclusión, el estudio se fundamenta en un análisis estadístico de los datos obtenidos a partir de las respuestas dadas por los estudiantes que se complementa con algunos aspectos cualitativos, porque analizamos los argumentos de los estudiantes y algunos tipos de respuestas que indican sesgos de razonamiento. Entonces, realizamos una investigación mixta, siguiendo las clasificaciones de Bisquerra (1989), y también se trata de una investigación aplicada y descriptiva.



En cuanto a la selección de la muestra, dicha elección ha sido intencional, aunque pensamos que representan adecuadamente a otros estudiantes del mismo contexto.

#### **4.4.1. CUESTIONARIO UTILIZADO**

Se trata de un instrumento de medición que, por medio de las respuestas de los estudiantes a los ítems que lo conforman, permite obtener información sobre su conocimiento, al que no se podría acceder mediante observaciones y encuestas (Barbero, 1993). El cuestionario permite obtener una cantidad amplia de información sobre el tema que nos interesa en un tiempo reducido. Es un instrumento de medida centrada en el sujeto, pues los conocimientos de los estudiantes se infieren en forma indirecta, ya que los sujetos no son conscientes de estas capacidades (Kline, 2013).

En concreto, el cuestionario está constituido por cuatro ítems que, como en el Estudio

1, tienen como finalidad evaluar tanto la comprensión de la variabilidad del muestreo como el efecto del tamaño de la muestra sobre dicha variabilidad. Los ítems utilizados son idénticos a los empleados en el Estudio 1, con la única diferencia que ahora se pide al estudiante justificar la respuesta. Como ejemplo, se reproduce la nueva versión del Ítem 1, que fue adaptado por Gómez et al. (2014) de otro previamente usado por Green (1983a).

**Item 1.** Un profesor vacía sobre la mesa un paquete de 100 chinchetas obteniendo los siguientes resultados: 68 caen con la punta para arriba  y 32 caen hacia abajo .

Supongamos que el profesor pide a 4 niños repetir el experimento, lanzando las 100 chinchetas. Cada niño vacía una caja de 100 chinchetas y obtendrá algunas con la punta hacia arriba y otras con la punta hacia abajo.

Escribe en la siguiente tabla un resultado que te parezca probable para cada niño:

Daniel	Martín	Diana	María
Punta arriba:	Punta arriba:	Punta arriba:	Punta arriba:
Punta abajo:	Punta abajo:	Punta abajo:	Punta abajo:

Justifica por qué has dado esta respuesta.

Todos los ítems corresponden a una población binomial  $B(n, p)$ . En los ítems se varió sistemáticamente el tamaño de la muestra y la probabilidad de éxito. En el primer ítem, el número de chinchetas  $X$  que caen con la punta hacia arriba sigue una distribución binomial,  $B(n, p)$ , donde  $n=100$  es el tamaño de la muestra y  $p$  es la proporción de chinchetas que caen con la punta hacia arriba en la población. Esta proporción es desconocida, pero se puede estimar a partir de los datos del enunciado,  $\hat{p}=0,68$ . A partir de ella se puede también estimar el número esperado de chinchetas que caerán con la punta hacia arriba en la muestra,  $n\hat{p} = 68$  y la desviación típica de esta variable,  $\sqrt{n\hat{p}(1 - \hat{p})} = 4,66$ .

La desviación típica de la distribución muestral de la media de los cuatro valores es la mitad de esta cantidad. Por tanto, se considera que el sujeto presenta una buena comprensión intuitiva de la proporción muestral, si el valor medio del número esperado de chinchetas con la punta hacia arriba en las cuatro muestras que proporciona es cercano a 68 (media teórica). Se consideran cercanas las respuestas, si dicho valor medio se localiza dentro del intervalo  $[65,7-70,3]$ . Este intervalo, que se obtiene al sumar y restar a la media teórica (68) la desviación típica de la distribución muestral, contiene el 68% de las observaciones en la distribución normal, que aproxima a la distribución muestral

considerada. En el intervalo [63,3-72,7] se sitúan el 95% de los valores centrales, por lo que se considera aceptable la comprensión intuitiva de la proporción muestral.

Con respecto al análisis de la comprensión de la variabilidad, se estudia el rango obtenido a partir de los cuatro valores muestrales. La distribución del rango en una muestra de cuatro elementos tiene una expresión complicada, por lo que la hemos aproximado utilizando la simulación. En la Figura 3.6.1 se mostró la simulación obtenida con 10000 simulaciones mediante el programa Fathom de la distribución binomial  $B(100, 0,68)$ . A partir de esta distribución empírica hemos calculado diferentes percentiles por lo que observamos que el 68% de los valores de los rangos en la distribución caerían entre 5 y 14 y el 95% entre 3 y 18. En el primer caso se considera la respuesta óptima y en el segundo aceptable. Los valores fuera del intervalo del 95% por exceso, se asumen como de excesiva variabilidad (mayor o igual a 19) y por debajo de 3 como alta concentración.

Tabla 4.4.1. Variables consideradas en el diseño del cuestionario

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Tamaño de la muestra	100	100	10	10
Probabilidad del suceso de interés	0,68	0,5	0,5	0,7
Estimación de la probabilidad	Frecuencial	Clásica	Clásica	Frecuencial
Número esperado de éxitos	68	50	5	7
Desviación típica	$\sigma = 4,66$	$\sigma = 5$	$\sigma = 1,58$	$\sigma = 1,45$
Intervalo que contiene el 68% de medias muestrales	[65,7-70,3]	[47,5-52,5]	[4,2-5,8]	[6,3-7,7]
Intervalo que contiene el 95 % de medias muestrales	[63,3-72,7]	[45-55]	[3,4-6,6]	[5,5-8,4]
Rango normativo	[5-14]	[6-15]	[2-5]	[2-4]
Rango aceptable	[3-18]	[3-20]	[1-6]	[1-6]

Realizando un análisis similar con los otros tres ítems, obtenemos la Tabla 4.3.1, en la que se muestran las variables que se tienen en cuenta en la elaboración del cuestionario de evaluación. Los dos primeros ítems se corresponden a tamaños de muestra grande (100 elementos) y los dos últimos a muestras pequeñas (10 elementos). Por otro lado, el primero y último de los ítems describen situaciones en las que el fenómeno aleatorio presenta sucesos no equiprobables, cuya probabilidad se estima a partir de la frecuencia dada en el enunciado del ítem. Las otras dos tareas hacen referencia a sucesos equiprobables, cuya probabilidad el estudiante puede conocer a través de su experiencia o considerando la regla de Laplace (probabilidad clásica).



Además, en la Tabla 4.3.1 incluimos el valor esperado y la desviación típica de la distribución binomial que modeliza la situación descrita en cada ítem, los intervalos que contienen el 68% y 95% de las medias muestrales y los valores del rango considerados aceptables.

#### **4.4.2. DESCRIPCIÓN DE LA MUESTRA**

La muestra está constituida por un total de 234 estudiantes de 2º bachillerato de cinco centros diferentes, de los cuales 85 estudiantes cursan el bachillerato de Ciencias Sociales y 149 el bachillerato de Ciencias. De ellos, 127 justificaron la elección de las muestras proporcionadas. La forma en que se seleccionaron los estudiantes fue mediante un muestreo dirigido, es decir, se ha basado en una selección controlada y de acuerdo a determinadas características (Hernández, Fernández y Baptista, 2010). No obstante, al utilizar diferentes centros y contar con un alto número de estudiantes, se trató de aumentar la representatividad de la muestra.

La Tabla 4.4.2 muestra el número de estudiantes que participan en el estudio y la modalidad de bachillerato que están cursando, además de indicar si argumentan las muestras consideradas. Como vemos, se cuenta con cinco centros diferentes y trece grupos de estudiantes, así como con las dos especialidades de bachillerato que estudian matemáticas. La Tabla 4.4.3 presenta un resumen de las características de la muestra en relación a la modalidad de bachillerato cursada.

Los estudiantes habían recibido instrucción previa sobre contenidos de probabilidad y estadística. Si consideramos el marco curricular que regula los contenidos a enseñar, se podría suponer que a lo largo de la etapa educativa los contenidos matemáticos enseñados han tenido que ser los mismos o similares en todos los grupos participantes. No obstante, aspectos citados anteriormente como el tiempo, el volumen de temario, así como las características particulares de cada centro, la diferencia del profesor del grupo, y, en definitiva, una serie de factores cuyo control resulta complicado, ponen en duda dicha afirmación.

Aunque nos resulte complejo determinar de manera exacta los conocimientos previos de dichos estudiantes sobre probabilidad y estadística, todos los docentes que colaboraron con nosotros coincidieron en que habían recibido instrucción del tema en diferentes momentos de su etapa educativa. Por lo que podemos suponer que los contenidos más

básicos asociados al tema se han presentado al estudiante en el aula.

En el caso de estudiantes que cursaban la modalidad de Ciencias Sociales, podemos referirnos a los contenidos del curso anterior donde el bloque de probabilidad y estadística tiene un peso dentro de la programación. De hecho, los contenidos de este bloque forman parte de la prueba de acceso a la universidad. Por tanto, teniendo en cuenta el papel que juega la evaluación en las decisiones de los profesores sobre la elección de contenido, es claro que dichos estudiantes tienen conocimientos previos.

Tabla 4.4.2. Características de la muestra participante en el Estudio 2

Centro educativo	Materia	Tamaño de la muestra	Justificación de la respuesta
C1	MATII	14	No
	MATII	17	
	MATCCSSII	19	
C2	MATCCSSII	12	No
	MATII	15	
C3	MATCCSSII	11	No
	MATII	19	
C4	MATII	26	Si
	MATII	15	
	MATCCSSII	19	
C5	MATII	29	Si
	MATII	14	
	MATCCSSII	24	

Tabla 4.4.3. Resumen de la muestra participante en el Estudio 2

	Justificación	No justificación
MAT II	84	65
MAT CCSS II	43	42

#### **4.4.3. MÉTODO DE RECOGIDA DE DATOS**

Los datos fueron recogidos en todos los grupos participantes en horas en las que los estudiantes tenían clase de matemáticas, presentando el cuestionario como una actividad de la citada clase. Los estudiantes completaron el cuestionario por escrito en un intervalo de tiempo comprendido entre 25-35 minutos. Se dio tiempo suficiente a todos para que lo completaran con comodidad.

Se les explicó a los estudiantes el fin de la evaluación y se resolvieron posibles dudas sobre la forma de completarlo. Adicionalmente, se observa que la mayoría de los estudiantes respondieron con interés, completando en su mayoría todas las preguntas.

## **4.5. ANÁLISIS CUANTITATIVO**

En esta sección, se presenta un análisis cuantitativo similar al considerado en el estudio del Capítulo 3, que se desarrolla sobre las respuestas cuantitativas de los estudiantes. El tratamiento de las mismas se realiza mediante el paquete estadístico SPSS.

En primer lugar, se realiza un análisis descriptivo tanto del valor medio como del rango del conjunto de respuestas para cada ítem. Para ello, se elaboran una serie de gráficos estadísticos que se complementan con tablas de frecuencias en las que aparecen los resultados asociados a la muestra del Estudio 2. De este modo, podemos realizar un estudio comparativo de los resultados obtenidos en función del curso participante. Finalmente, se realizan también contrastes chi-cuadrado, análisis de varianza y comparaciones a posteriori de las medias para comparar los resultados obtenidos en los tres grupos de estudiantes. Se añade un análisis de la puntuación total en el cuestionario y puntuaciones parciales, así como de las correlaciones entre los valores medios y los rangos de las cuatro estimaciones en cada ítem. En lo que sigue se describe de análisis y se presentan los resultados en los diferentes ítems, comparándolos con los del Estudio 1 y los de las investigaciones previas.

### **4.5.1. MÉTODO DE ANÁLISIS**

Como en el primer estudio, para el análisis de la comprensión intuitiva del valor esperado se considera en cada ítem la distribución de la media de las cuatro respuestas proporcionadas por cada estudiante.

La Tabla 4.5.1 presenta las medias teóricas (valor esperado en la distribución binomial que modela el enunciado de cada ítem) y las obtenidas en la muestra total de estudiantes. Además, hemos incluido una fila con los valores medios observados en el Estudio 1. Si comparamos los resultados obtenidos en ambas muestras, se concluye que dichos resultados son similares, aunque un poco mejor en los estudiantes del Estudio 2. Además, ambos grupos proporcionan una estimación bastante precisa del valor esperado en los tres últimos ítems, puesto que la diferencia en valor absoluto de la media teórica y observada es como máximo de un punto aproximadamente.

Tabla 4.5.1. Media teórica y observada de las distribuciones del valor medio de las cuatro respuestas proporcionadas por cada estudiante a cada ítem en el global de la muestra

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Media teórica	68	50	5	7
Media observada en bachillerato	60,4	50,2	5,1	6,4
Media observada en ESO	57,9	51,2	5,1	6,7

Por el contrario, los resultados del primer ítem muestran una diferencia con el valor teórico. Esta diferencia es de 8 puntos para el caso de los estudiantes de bachillerato, siendo dos puntos mayor para la muestra de la ESO. Por tanto asumimos que, al igual que en dicho estudio, algunos estudiantes razonan de acuerdo a la heurística de la equiprobabilidad (Lecoutre, 1992), sin considerar la información frecuencial del enunciado.

Para analizar mejor la precisión en la estimación del valor medio, las respuestas de los estudiantes se han clasificado, al igual que se hizo en el Estudio 1, en las siguientes categorías:

- *Estimación normativa del valor esperado.* El valor medio de las cuatro estimaciones proporcionadas por el estudiante en el ítem es muy próxima al valor teórico (la distancia es menor que una desviación típica).
- *Estimación aceptable del valor esperado.* Se considera también una respuesta aceptable si el valor medio de las cuatro estimaciones dadas por el estudiante se encuentra a una distancia menor de dos desviaciones típicas del valor teórico, pues dicho intervalo contendría al 95% de los valores de las medias muestrales.
- *Recencia negativa.* En los experimentos descritos en los ítems 1, 2 y 4, el enunciado indica que se obtuvo como resultado una frecuencia mayor de uno de los sucesos. Algunos estudiantes que tratan de compensar este resultado inicial, dando estimaciones cuyo valor medio favorece al suceso contrario, muestran recencia negativa, conducta se ha explicado mediante la heurística de la representatividad (Tversky y Kahneman, 1982) e implica una pobre comprensión de la independencia de los ensayos en un experimento aleatorio (Batanero y Serrano, 1999).
- *Sesgo de equiprobabilidad.* En los ítems 1 y 4, que los sucesos elementales no son equiprobables, algunos estudiantes proporcionan estimaciones cuyos valores medios se localizan muy próximos al 50% del tamaño de la muestra, presentando el sesgo de

equiprobabilidad (Lecoutre, 1992).

- *Recencia positiva.* Ante un enunciado en el que la frecuencia de los dos sucesos implicados es diferente, se tiende a proporcionar estimaciones en que se exagera tal diferencia.
- *No completa.* No responden a un ítem, o bien proporcionan respuestas incoherentes con el enunciado.

En la Tabla 4.5.2 se presentan la media teórica de los rangos (valor esperado en la distribución muestral del rango de valores en cada ítem) y las obtenidas por estudiantes de bachillerato y de ESO. Sus resultados revelan que los estudiantes conceden muestras con una variabilidad adecuada en el caso de los dos últimos ítems, mientras que en el ítem 1 y 2 resulta la diferencia entre el valor teórico y observado de al menos 13 puntos, siendo esta diferencia mayor para la primera pregunta. Además, esta observación coincide independientemente del grupo participante, aunque los resultados son mejores en el caso de los alumnos de bachillerato.

Tabla 4.5.2. Media teórica y observada de las distribuciones del rango de las cuatro respuestas proporcionadas por cada estudiante a cada ítem en el global de la muestra

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Rango teórico medio	9,6	10,3	3,2	2,9
Rango medio observado bachillerato	23,7	23,6	3,8	3,1
Rango medio observado en ESO	29,8	32,6	4,4	3,1

Para estudiar la percepción de la variabilidad en el muestreo, se ha analizado la distribución del rango de las cuatro estimaciones proporcionadas por el estudiante en cada ítem, que se simuló en el Estudio 1, junto a los gráficos y valores de los percentiles correspondientes al 68% y 95% central de la distribución, que se dieron en el Capítulo 3. Además de utilizar los histogramas, se representan los gráficos de caja para estos rangos por curso e ítem. Las respuestas de los participantes se han clasificado en las siguientes categorías:

- *Estimación normativa de la variabilidad muestral.* Es la respuesta óptima, cuando el estudiante proporciona cuatro estimaciones con un rango incluido en el intervalo en que se sitúan el 68% de los rangos de las muestras de cuatro elementos en la distribución muestral.
- *Estimación aceptable de la variabilidad muestral.* El estudiante da cuatro

estimaciones con un rango incluido en el intervalo del 95% de la distribución muestral de rangos y fuera del intervalo anterior.

- *Estimación excesiva de la variabilidad muestral.* El estudiante proporciona cuatro estimaciones con un rango mayor que el extremo superior del intervalo del 95% en la distribución muestral de rangos.
- *Alta concentración.* El estudiante responde con cuatro estimaciones con el mismo valor o valores muy similares.

#### **4.5.2. RESULTADOS EN EL ÍTEM 1**

En este ítem la información frecuencial sobre la probabilidad del suceso de interés (punta hacia arriba) se proporciona al estudiante a partir del dato dado en el enunciado y los estudiantes también pueden razonar, desde su experiencia, que es más probable que la chincheta caiga hacia arriba debido al peso de su superficie plana.

En primer lugar, tras el análisis de las respuestas de la muestra de estudiantes para el estudio del valor medio, obtenemos un valor medio de 60,4, muy similar al obtenido con los estudiantes de ESO y en el trabajo de Gómez et al. (2014) con futuros profesores, donde el valor medio fue 57,7. Sin embargo, este valor es inferior al valor medio esperado, que es 68. La posible explicación de ese valor se debe a que un conjunto de estudiantes proporciona una estimación sesgada hacia el valor 50 del número medio de chinchetas que caen hacia arriba, lo cual sugiere que no tienen en cuenta la información frecuencial. En conclusión, pensamos que dichos estudiantes razonarían de acuerdo al sesgo de equiprobabilidad descrito por Lecoutre (1992). Además, no es posible aceptar la falacia de la composición descrita por Chernoff y Russell (2012), que consiste en transferir a la muestra una propiedad de la población. En nuestro caso, se correspondería con la equiprobabilidad, pero dicha propiedad no es cierta para el fenómeno aleatorio presentado en este primer ítem.

En la Figura 4.5.1 se presentan la distribución de la media y el rango, que se calcula a partir de los cuatro valores por los estudiantes, que es una muestra de la correspondiente distribución binomial. Además, se señala junto con cada distribución el valor teórico de las respectivas medias marcadas por una línea vertical.

En particular, consideramos la gráfica asociada a la distribución del valor medio, y se observa que los datos se concentran en torno al valor medio teórico, marcado por la

línea vertical. La asimetría de la distribución se debe a que los datos se acumulan por debajo del valor esperado, lo cual es debida a dos factores. Por un lado, existe un porcentaje de estudiantes que presentan el sesgo de equiprobabilidad, proporcionando muestras con proporción cercana al valor  $p=0,5$ . Por otro lado, aparece la recencia negativa (Kahneman et al., 1982), sugiriendo resultados que buscan compensar por defecto el dato en el enunciado y, por tanto, no tienen en cuenta el significado frecuencial de la probabilidad para la realización de la tarea. Esta última conducta es explicada por Batanero y Serrano (1999), quienes indican que se debe a una pobre comprensión de la idea de independencia en el muestreo.

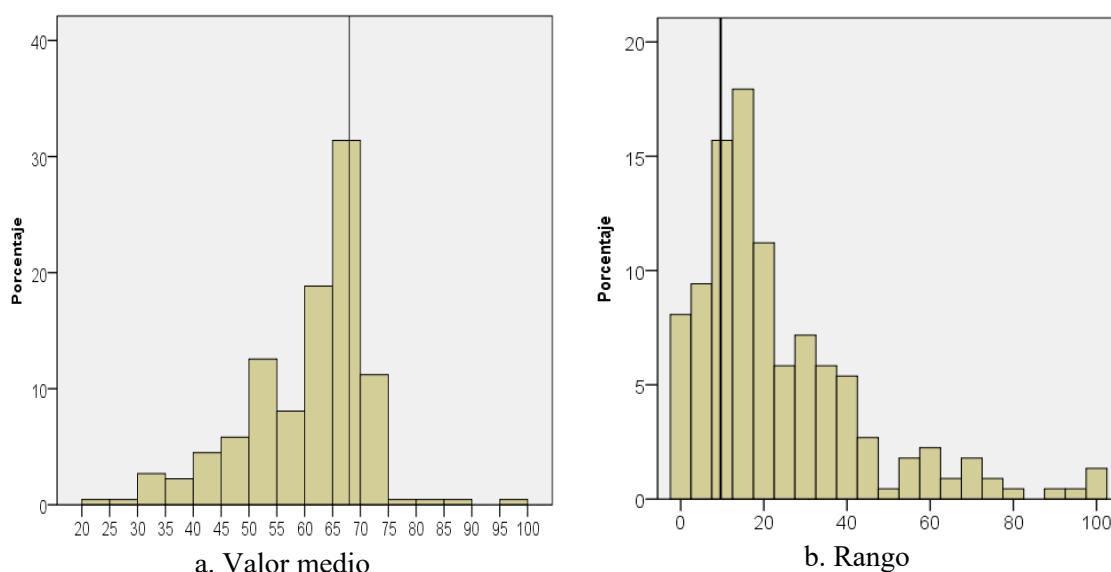


Figura 4.5.1. Distribución del valor medio y rango de cuatro valores del número de chinchetas hacia arriba en 100 chinchetas

El análisis de la distribución se complementa con la Tabla 4.5.3, en la que se presenta el porcentaje de respuestas que se localizan en cada uno de los intervalos previamente definidos en el análisis de la respuesta. Además, se presentan los valores obtenidos para los otros dos grupos que participaron en el Estudio 1. Señalamos que solo el 20,9% de los estudiantes de bachillerato da una estimación normativa de la media y el 23,4 % aceptable. En conjunto, el 44,3% da un valor aceptable, siendo estos resultados mejores que los de sus compañeros de educación secundaria obligatoria (28,6% para el grupo de 2ºESO y 33,1% para el de 4ºESO).

Como se ha deducido del análisis del gráfico, observamos una asimetría del mismo. Además, la Tabla 4.5.3 muestra que el 17,7% de los estudiantes no identifica la asimetría

del dispositivo, considerando que ambos resultados son igual de probables, es decir, razonan mostrando el sesgo de equiprobabilidad (Lecoutre, 1992), siendo este porcentaje menor que el obtenido en los cursos de segundo y cuarto de ESO. Como se ha indicado, este ítem es adaptado del trabajo de Green (1983a), en cuyo estudio el 61% de los niños mostró el sesgo de equiprobabilidad y, además, no hubo diferencia entre niños de diferentes edades. Por su parte, Cañizares (1997) obtiene que el 64,1% razonan según el sesgo de equiprobabilidad en un trabajo en el que utiliza el mismo ítem que aplicó Green.

Tabla 4.5.3. Porcentaje de estudiantes por grupo según valor medio en las cuatro estimaciones en el ítem 1

Valor medio de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa del valor esperado [67,5-70,3]	10,8	11,7	20,8
Estimación aceptable del valor esperado <sup>1</sup> [63,3-72,7]	17,8	21,4	23,4
Recencia negativa (menor que 45)	16,6	11,0	14,8
Recencia positiva (mayor que 72,7)	8,3	8,3	2,8
Equiprobabilidad [45- 55]	26,8	23,4	17,7
Otros valores (55-63,3)	13,4	19,3	15,8
No completa	6,4	4,8	4,7

<sup>1</sup>Fuera del intervalo normativo

En nuestro estudio, se obtienen mejores resultados según la edad, como se deduce de la Tabla 4.5.3, que posteriormente se analizará si puede considerarse significativa. Sin embargo, hay que considerar determinados factores que afectan a los resultados obtenidos como, por ejemplo, la formación de nuestros estudiantes, ya que la muestra del estudio de Green no había estudiado probabilidad. Otro factor que puede explicar las diferencias en los resultados obtenidos es la edad de la muestra objeto de estudio.

Por otro lado, el 14,8 % presenta recencia negativa frente al 2,8% que presenta la recencia positiva. Según Tversky et al. (1982), ambos sesgos son explicados por la heurística de la representatividad y también se han identificado en las investigaciones previas como la de Green (1991), Pollatsek, Konold, Well y Lima (1984) y Serrano (1996) y también aparecen en el trabajo de Shaughnessy et al. (2004). Siguiendo la clasificación de Valdez (2016), estos estudiantes se situarían en el nivel de razonamiento 3 con respecto a las ideas de aleatoriedad, variabilidad e independencia, puesto que su asignación de probabilidades se basa en un modelo preestablecido, ignorando los datos frecuenciales para la elaboración de su respuesta.

El estudio se complementa con el diagrama de caja asociado a la distribución del



valor medio de las cuatro respuestas (Figura 4.5.2.a). En este diagrama se han representado en rojo el intervalo teórico que debe contener al 68% de la distribución muestral de medias y en negro al intervalo que debe contener el 95%. Este gráfico muestra que el valor de la mediana se sitúa por debajo del teórico (68). Además, el extremo inferior del intervalo que recoge al 95% de las respuestas coincide con el valor de la mediana, entonces un 50% de los estudiantes quedaría por debajo de estos intervalos produciendo respuestas sesgadas hacia la equiprobabilidad.

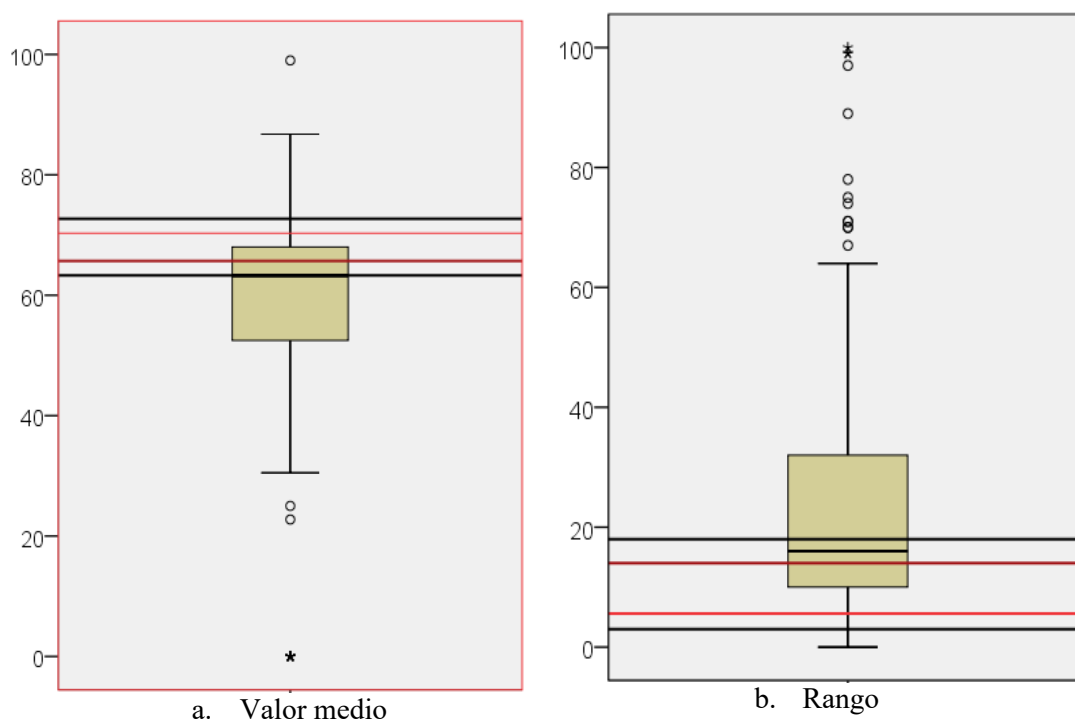


Figura 4.5.2. Diagramas de caja del valor medio y rango de cuatro valores del número de chinchetas hacia arriba en 100 chinchetas

Finalmente, si comparamos los resultados de este grupo con los obtenidos en el Estudio 1, son mejores, ya que en el Estudio 1, tan solo el 10,8% de los de 2º curso de ESO y el 11,7% de los de 4º curso proporcionan una estimación normativa. Además, disminuye la presencia tanto del sesgo de equiprobabilidad como los correspondientes a la heurística de representatividad. En consecuencia, se confirma nuestra hipótesis de mejora del razonamiento sobre muestreo con la edad en este punto.

Al realizar el test Chi-cuadrado de independencia entre curso y tipo de respuesta, teniendo en cuenta únicamente tres categorías de respuestas: optimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas), se obtiene un valor Chi=16,64 con 4 g.l. que es estadísticamente muy significativo ( $p=0,002$ ), lo que muestra

un incremento progresivo del número de respuestas normativas y aceptables con el curso y disminución de los sesgos.

### **Análisis de la variabilidad**

En cuanto al análisis de la variabilidad, la gráfica de la distribución del rango (Figura 4.5.1.b) revela cierta asimetría, que es debida a que un porcentaje de la muestra concede una variabilidad excesiva en las muestras generadas. De hecho, si consideramos la media teórica asociada a la distribución del rango de la muestra de cuatro elementos, toma un valor de 9,6 (Tabla 4.5.2), sin embargo, el valor de la observada en nuestra muestra es de 23,7.

Además, el diagrama de caja de la distribución de rangos (Figura 4.5.2.b) presenta en rojo el intervalo teórico que debe contener al 68% de la distribución muestral de rangos y en negro el intervalo que debe contener el 95%. Si consideramos dichos intervalos, observamos que un porcentaje elevado de los estudiantes concede muestras cuyo rango toma valores mayores al extremo superior del intervalo que debería contener al 95% de las respuestas. Además, la Figura 4.5.2.b marca la existencia de un número de alumnos atípicos que se corresponden con respuestas cuyas estimaciones tienen rangos superiores a 60, siendo el número de casos atípicos superior que el identificado para la muestra de la ESO. Por otro lado, la Tabla 4.5.4 muestra los porcentajes asociados a los diferentes intervalos en los que se ha clasificado la distribución del rango. Por tanto, nos permite identificar en términos de porcentaje, los alumnos que conceden una variabilidad excesiva, que se corresponde con el 42,8%.

En resumen, estos estudiantes no comprenden que el tamaño de la muestra de la distribución binomial de partida reduce la variabilidad muestral, ya que producen valores sesgados hacia el exceso de variabilidad. Por ello, se deduce que dichos alumnos llegarían solo a uno de los niveles inferiores de razonamiento sobre muestreo (idiosincrático o de transición) según Moreno y Vallecillos (2001), puesto que no reconocen el efecto del tamaño de la muestra sobre la variabilidad. Ejemplos de este mismo error se encontraron en las investigaciones de Méndez (1991) y Serrano (1996).

Por otro lado, la Tabla 4.5.4 muestra que una proporción en torno al 20% de los estudiantes proporcionan valores cuyo rango se sitúa en el intervalo que se ha definido como rango normativo y un porcentaje similar se obtiene para el intervalo del rango que

se considera aceptable. En total, el 40,7% genera muestras con una variabilidad al menos aceptable, un valor muy similar al porcentaje de respuestas que sobreestiman la variabilidad en el proceso de muestreo.

Finalmente, un 11,8% de los estudiantes (Tabla 4.5.4) proporcionan valores con una concentración elevada. Respecto a este tipo de respuesta, otros autores como Rubin et al. (1990) indican que los estudiantes conceden una variabilidad menor a muestras cuyo tamaño es pequeño, pero es menos frecuente encontrar estudios en que los estudiantes sobreestimen la variabilidad en muestras grandes. En otros casos, como en Shaughnessy et al. (2004), la mayoría de los estudiantes sobreestimaron la variabilidad de la distribución muestral, independientemente del tamaño de la muestra.

Tabla 4.5.4. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 1

Rango de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [5-14]	12,7	17,9	19,5
Estimación aceptable [3-18] <sup>1</sup>	7,6	11,7	21,2
Estimación de una variabilidad muestral excesiva >18	62,4	56,6	42,8
Alta concentración <3	10,8	9	11,8
No completa	6,4	4,8	4,7

<sup>1</sup>Fuera del intervalo normativo

Finalmente, si comparamos los porcentajes asociados a cada intervalo, según el curso, se observa una mejora con la edad. Esta observación se confirma con el test Chi-cuadrado de independencia entre curso y respuesta, teniendo en cuenta tres categorías de respuestas: optimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas) se obtiene un valor  $\chi^2=25,22$  con 4 g.l., que es estadísticamente muy significativo ( $p<0,001$ ).

### Contraste de diferencias entre grupos

Con objeto de analizar si las diferencias entre los resultados en bachillerato y los obtenidos en los cursos anteriores son estadísticamente significativas, la Tabla 4.5.5 presenta los resúmenes estadísticos de las distribuciones del valor medio y el rango de las cuatro estimaciones de cada estudiante en cada uno de los grupos, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo.

En primer lugar, observamos que el valor medio de la distribución de las medias

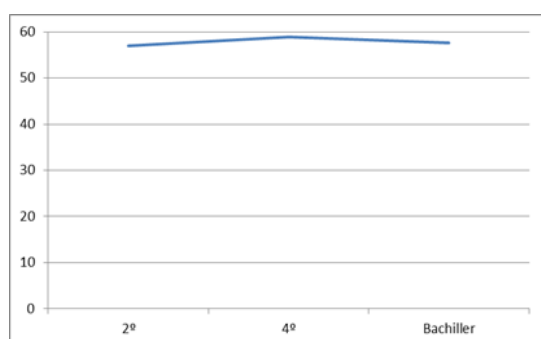
asociada a cada grupo es muy similar, siendo el correspondiente al cuarto curso algo más alto. Como se ha concluido en los análisis anteriores, estos valores son muy diferentes al valor teórico esperado del número de chinchetas con la punta hacia arriba. En conclusión, los estudiantes presentan dificultades para identificar la probabilidad de los fenómenos aleatorios que no cumplen el principio de equiprobabilidad.

En segundo lugar, el análisis de la variabilidad revela que hay una diferencia notable de la media de los rangos. En concreto, esta diferencia es de casi 10 puntos entre los estudiantes de bachillerato y los de segundo curso de ESO, lo que indica que sus estimaciones tienen menor variabilidad. Por tanto, identificamos que existe una mejora al progresar el curso sobre el concepto de la variabilidad y su relación con el tamaño de la muestras.

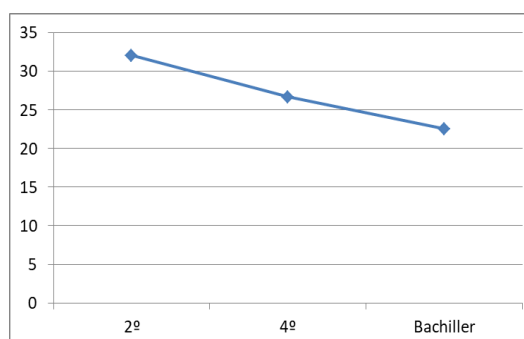
Para completar la Tabla 4.5.5, se muestran de manera gráfica los datos aportados en dicha tabla en la Figura 4.5.3, donde se ha considerado el valor de la media para los dos estadísticos considerados. De este modo, la gráfica asociada al rango se caracteriza por ser una recta que decrece en relación al aumento del curso.

Tabla 4.5.5. Estadísticos de la media y rango en las cuatro estimaciones en el Ítem 1 por curso

	Curso	N	Media	D. típica.	Error muestreo
II Media	2º ESO	147	56,920	13,7228	1,1318
	4º ESO	138	58,899	11,6006	,9875
	Bachillerato	234	57,579	16,7599	1,0956
II Rango	2º ESO	147	32,07	22,499	1,856
	4º ESO	138	26,68	20,764	1,743
	Bachillerato	234	22,55	20,680	1,352



a. Valor medio



b. Rango

Figura 4.5.3. Valores medios de la media y el rango de las cuatro estimaciones en el ítem 1

Para comprobar si las diferencias observadas son estadísticamente significativas se ha realizado la prueba del análisis de varianza tomando el grupo de estudiantes como variable independiente. Los resultados se presentan en la Tabla 4.5.6, donde observamos que la diferencia de medias no es estadísticamente significativa y sí es muy significativa la diferencia de rangos, cuyo valor p es menor que 0,05. En conclusión, no tenemos motivos para suponer que los diferentes grupos de estudiantes muestran una comprensión diferente de la relación entre el valor esperado y el teórico en este ítem. Por el contrario, sí que se identifica que algunos grupos entienden mejor la variabilidad del muestreo en muestras grandes.

Tabla 4.5.6. Resultados del análisis de varianza en el ítem 1

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	290,111	2	145,055	,672	,511
IIMedia Intra-grupos	111378,865	516	215,851		
Total	111668,976	518			
Inter-grupos	8198,057	2	4099,028	9,095	,000
IIRango Intra-grupos	234348,303	520	450,670		
Total	242546,359	522			

Desde el interés por comprobar qué grupos presentan una diferencia en la estimación del rango, se realizan las pruebas de Tukey de diferencia de medias, recomendadas una vez que el análisis de varianza da un valor significativo. Los resultados se resumen en la Tabla 4.5.7, donde solamente se obtiene diferencia estadísticamente entre bachillerato y el segundo curso de ESO en el rango, pero no con cuarto curso.

Tabla 4.5.7. Prueba post-hoc de diferencias de medias para el rango de las cuatro estimaciones en el ítem 1

	Prueba t para la igualdad de medias				Intervalo de confianza de la diferencia (95%)	
	g.l.	Diferencia de medias	Error típ. diferencia	Sig. bilateral	E. Inferior	E. Superior
Diferencia con 2º	283	-9,517	2,234	,000	-14,77	-4,27
Diferencia con 4º	287	-4,125	2,258	,162	-9,43	1,18

### 4.5.3. RESULTADOS EN EL ÍTEM 2

En este apartado se presenta el análisis de los resultados del ítem 2 relacionado con el lanzamiento de 100 monedas equilibradas. De manera análoga al caso anterior, el primer estudio se centra en el valor medio y, por tanto, se presenta la distribución asociada

al número medio de las cuatro estimaciones en la Figura 4.5.4.a. El análisis de dicha distribución muestra que el intervalo modal se sitúa próximo al valor teórico esperado, lo cual también ocurre en el Estudio 1. Por tanto, este resultado nos indica que la mayor parte de los estudiantes identificaron correctamente la equiprobabilidad de los dos sucesos que constituyen el espacio muestral del experimento.

El análisis del gráfico se completa con la Tabla 4.5.8, donde resulta destacable que el 47,7% de los estudiantes realizan una estimación normativa del valor esperado, porcentaje mayor que el obtenido en segundo curso de ESO (31,2%) y en cuarto curso (46,2%). Además, el 70% aproximadamente de las respuestas se incluye en el intervalo que contiene al 95% de los casos, que se corresponde con las respuestas que pueden considerarse aceptables. Por lo que los participantes en el estudio presentan una adecuada estimación del valor esperado.

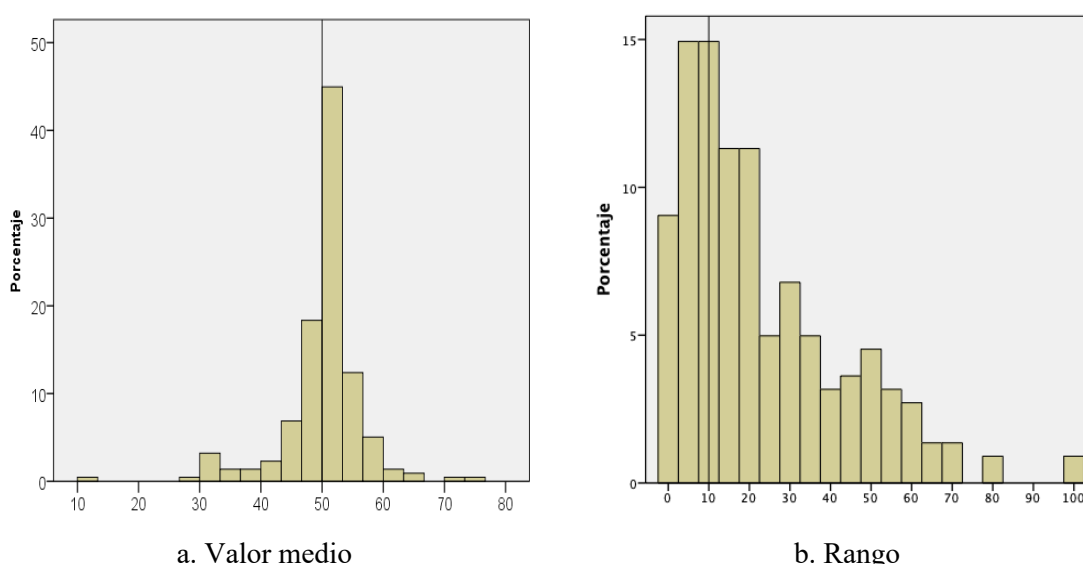


Figura 4.5.4. Distribución del valor medio y rango de cuatro valores del número de caras en 100 lanzamientos de monedas

Tabla 4.5.8. Porcentaje de estudiantes por grupo según valor medio en las cuatro estimaciones del valor esperado en el ítem 2

Valor medio de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [47,5-52,5]	31,2	46,2	47,7
Estimación aceptable [45-55] <sup>1</sup>	19,1	23,4	21,3
Menores que el aceptable (< 45)	16,6	9,7	15,2
Mayores que el aceptable (>55)	25,5	15,9	11,1
No completa	7,6	4,8	4,7

<sup>1</sup>Fuera del intervalo normativo

Finalmente, si consideramos el porcentaje de respuestas bajo las categorías normativa y aceptable, los resultados correspondientes a bachillerato y a cuarto curso son muy similares, siendo mejores que los obtenidos para el segundo curso. Todo ello se confirma al realizar el test Chi-cuadrado de independencia entre curso y respuesta, teniendo en cuenta tres categorías de respuestas: optimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas), donde se obtiene un valor  $\chi^2=23,26$  con 4 g.l. que es estadísticamente muy significativo ( $p<0,001$ ). Se confirma nuestra hipótesis de una mejor comprensión en los estudiantes de mayor curso.

El análisis del diagrama de caja del valor medio (Figura 4.5.5.a) revela la simetría de la distribución asociada a la media, donde el valor de la mediana coincide prácticamente con el valor medio teórico.

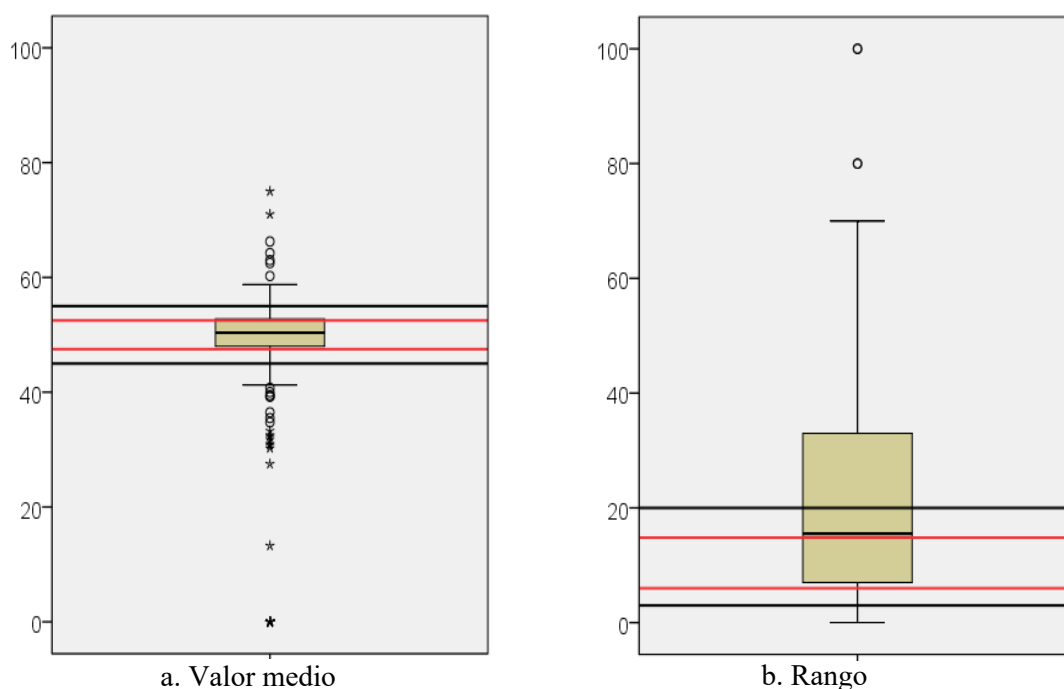


Figura 4.5.5. Diagramas de caja del número medio y rango de 4 valores del número de caras en el lanzamiento de 100 monedas

Por un lado, indica que los valores que toma el rango intercuartílico se corresponden de manera aproximada con los extremos que forman el intervalo que contiene al 68% de las respuestas consideradas como estimación normativa. No obstante, si analizamos el porcentaje de respuestas que se sitúan dentro del intervalo que debería contener al 95% de las respuestas (Tabla 4.5.8), observamos un porcentaje de estudiantes que proporcionan muestras y razonan según a la heurística de la representatividad (Kahneman

et al., 1982), tratando de compensar por exceso o defecto el dato dado en el enunciado. En concreto, un 26,3% muestran esa conducta (Tabla 4.5.8). Además, el diagrama de caja identifica valores atípicos tanto por defecto como por exceso.

### **Análisis de la variabilidad**

En primer lugar, la Figura 4.5.4.b muestra la distribución asociada al rango, donde la mayoría de los estudiantes producen rangos por encima del valor medio esperado, llegando a generar muestras cuyo valor para el rango toma incluso el valor 100. De hecho, mientras que el valor medio esperado del rango en este ítem es 10,3 (ver Tabla 4.5.2), en nuestra muestra se obtiene más del doble (23,6), aunque menos que en los cursos de la ESO. Además, observamos diferentes conductas en los estudiantes, que se precisan mejor en la Tabla 4.5.9.

Por un lado, destaca el 30,6% de estudiantes que proporcionan muestras con variabilidad normativa, en comparación con los porcentajes correspondientes a los de segundo y cuarto de ESO (aproximadamente del 17%). Se observa también una mejora considerable en el porcentaje de respuestas cuya variabilidad es aceptable (15,2% de estudiantes). De hecho, si se considera de manera conjunta los porcentajes correspondientes a ambas categorías, se obtiene como resultado que, aproximadamente, el 50% de los estudiantes conceden una variabilidad adecuada a las muestras dadas. En consecuencia, podemos concluir que mejora el razonamiento sobre la variabilidad en este ítem con el curso.

Tabla 4.5.9. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 2

Rango de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [6-15]	17,8	17,2	30,6
Estimación aceptable [3-20] <sup>1</sup>	8,3	9,7	15,2
Excesiva (>20)	58,6	59,3	37,4
Alta concentración (<3)	7,7	9,0	12,1
No completa	7,6	4,8	4,7

<sup>1</sup>Fuera del intervalo normativo

El análisis de los resultados también indica que la mitad de los estudiantes proporcionan muestras cuya variabilidad se localiza en los extremos, es decir, o la



proporción de caras se localiza muy próximo al valor medio teórico (12,1%) o las muestras presentan una variabilidad excesiva (37,4%). En particular, se observa que existe una tendencia en parte de los estudiantes en conceder una variabilidad excesiva.

Desde el diagrama de caja asociado al rango (Figura 4.5.5.b) se llega a la misma conclusión, donde se observa que el tercer cuartil toma un valor alejado al extremo superior en el que se localizarían el 95% de las respuestas. Además, el valor de la mediana se localiza entre los extremos superiores de los intervalos en los que deberían localizarse el 68% y 95% de las respuestas. Por tanto, el diagrama de caja informa que la mayoría de los estudiantes sobreestiman la variabilidad de la distribución.

En conclusión, los estudiantes de bachillerato tampoco alcanzan los niveles superiores de razonamiento sobre el muestreo descritos por Moreno y Vallecillos (2001) y presentan los mismos errores descritos por Méndez (1991), Serrano (1996) y Shaughnessy et al. (2004), los cuales se han identificado y descrito en el ítem anterior.

A pesar de que todavía se identifican respuestas inadecuadas en los alumnos de bachillerato, los resultados son mejores que los de sus compañeros de ESO, como se confirma con el test Chi-cuadrado de independencia entre curso y respuesta. En este caso, se obtiene un valor  $\chi^2=53,04$  con 4 g.l., que es estadísticamente muy significativo ( $p<0,001$ ), confirmando por tanto nuestros resultados.

### **Contraste de diferencias entre grupos**

Para analizar si las diferencias observadas son significativas, en primer lugar, presentamos la Tabla 4.5.10, donde se muestran tanto las medias como las desviaciones típicas del valor medio y el rango en las cuatro estimaciones de cada estudiante en cada uno de los grupos, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo.

Por un lado, el análisis de la media asociada a cada grupo revela que existe una pequeña diferencia en la estimación que hacen los tres grupos de estudiantes del valor medio, siendo la estimación correspondiente al bachillerato más baja. Por otro lado, desde el análisis del rango se identifica una diferencia de 15 puntos en su valor medio, si comparamos la muestra de bachillerato y segundo curso de ESO, mientras que la diferencia es de 9 puntos si comparamos los alumnos de bachillerato con los de cuarto curso. Por tanto, los estudiantes de último curso de la educación secundaria producen

muestras cuya variabilidad es menor y parece que estos estudiantes visualizan mejor que sus compañeros la disminución de variabilidad en las muestras grandes. Las tendencias observadas para ambos parámetros se resumen de manera gráfica en la Figura 4.4.6.

Tabla 4.5.10. Estadísticos de la media y rango de los cuatro valores dados en el Ítem 2 por curso

	Curso	N	Media	D. típica.	Error muestreo
I2Media	2º	145	51,581	8,6143	,7154
	4º	138	50,828	6,3624	,5416
	bachillerato	234	46,970	13,8872	,9078
I2Rango	2º	145	33,68	23,155	1,916
	4º	138	31,20	24,548	2,082
	bachillerato	234	22,00	20,333	1,329

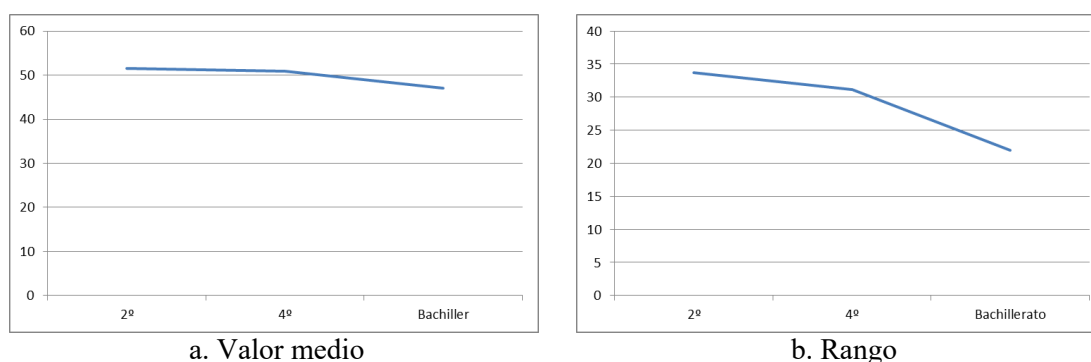


Figura.4.5.6. Valores medios de la media y el rango de las cuatro estimaciones en el ítem 2

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba del análisis de varianza tomando el grupo de estudiantes como variable independiente o factor. Los resultados se presentan en la Tabla 4.5.11 donde observamos que son estadísticamente significativas tanto la diferencia de estimaciones de las medias como la diferencia de estimaciones de los rangos, cuyo valor p es menor que 0,05. Concluimos que tenemos motivos para suponer que los diferentes grupos de estudiantes muestran una comprensión diferente de la relación entre el valor esperado y el teórico en este ítem, y que algunos grupos entienden mejor la variabilidad del muestreo en muestras grandes.

Tabla 4.5.11. Resultados del análisis de varianza en el ítem 2

		Suma de cuadrados	g.l.	Media cuadrática	F	Sig.
I2Media	Inter-grupos	2346,854	2	1173,427	9,861	,000
	Intra-grupos	61166,626	514	119,001		
	Total	63513,480	516			
I2Rango	Inter-grupos	14535,070	2	7267,535	14,579	,000
	Intra-grupos	257223,867	516	498,496		
	Total	271758,936	518			

Como el análisis de la varianza nos devuelve un valor significativo, tratamos de comprobar qué grupos presentan una mejor comprensión de ambos estadísticos. Por tanto, se realizan las pruebas de Tukey de diferencia de medias, cuyo resultado se presenta en la Tabla 4.5.12. En relación con el valor medio, se obtiene una diferencia estadísticamente significativa de la estimación media en este ítem al comparar los estudiantes de bachillerato con respecto tanto al segundo como al cuarto curso. En el caso del rango, las diferencias con los dos grupos son también estadísticamente significativas, lo que confirma que la variabilidad de la estimación es menor al avanzar el grupo escolar.

En este ítem, se observa una mejora en la comprensión del proceso de muestreo, ya que las diferencias son significativas al comparar las muestras producidas por los estudiantes de bachillerato con ambos grupos del Estudio 1.

Tabla 4.5.12. Prueba post-hoc de diferencias de medias para la media y el rango de las cuatro estimaciones en el ítem 2

		Diferencia de medias (I-J)	Error típico	Sig.
Media	Diferencia con 2º	-4,6109*	1,1529	,000**
	Diferencia con 4º	-3,8578*	1,1708	,003**
Rango	Diferencia con 2º	-11,685*	2,355	,000**
	Diferencia con 4º	-9,201*	2,391	,000**

\*\* Muy significativo

#### 4.5.4. RESULTADOS EN EL ÍTEM 3

El siguiente ítem se corresponde también con el lanzamiento de monedas equilibradas, pero en este caso se pide generar una muestra pequeña de 10 lanzamientos. La distribución del valor medio de las cuatro respuestas a este ítem se presenta en la Figura 4.5.7.a, donde la moda coincide con el valor medio esperado y la dispersión es pequeña. Como en el caso del ítem 2, los estudiantes han identificado la equiprobabilidad de los sucesos del espacio muestral, lo que lleva a la mayoría a dar una estimación

bastante precisa del valor medio.

La Tabla 4.5.13 refleja que el 71,5% de los estudiantes de bachillerato proporcionan muestras cuyo valor medio se sitúa en el intervalo normativo. Por tanto, estos alumnos proporcionan muestras donde el número de éxitos es muy aproximado al valor teórico. Este porcentaje es mayor para el caso de los estudiantes de bachillerato que en los otros cursos de la ESO.

No obstante, si consideramos el porcentaje de muestras cuyo valor medio es aceptable, observamos que el porcentaje es 84,8%. En el Estudio 1, se obtiene el 83,4% en el grupo de 2º ESO y el 83,2% para los alumnos de 4º ESO. En cuanto a la presencia de sesgos, se observa que la muestra participante muestra una tendencia en proporcionar muestras cuyo valor medio sea menor al valor medio teórico. Por tanto, estos alumnos pueden razonar según la heurística de la representatividad (Tversky y Kahneman, 1982).

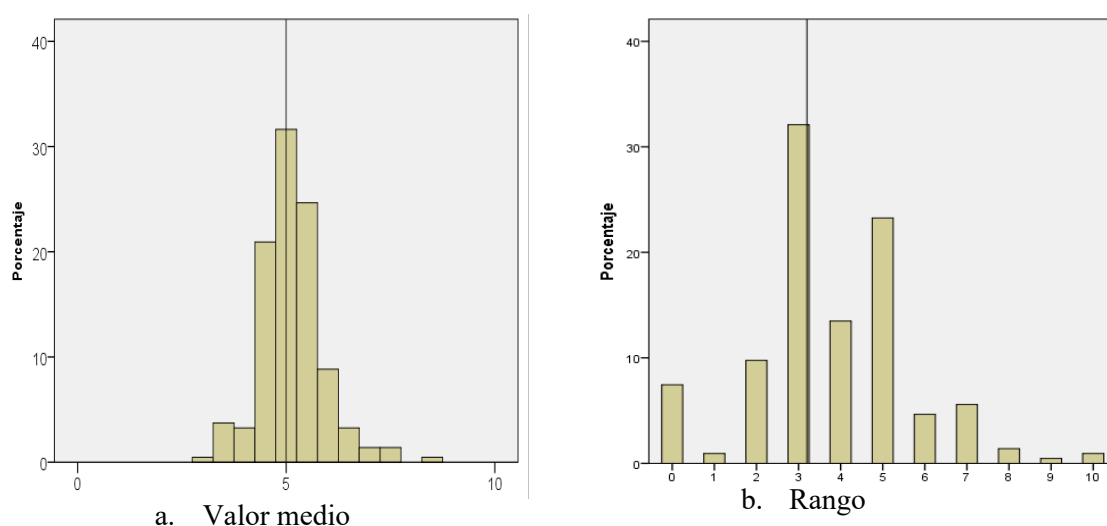


Figura 4.5.7. Distribución del valor medio y rango de cuatro valores del número de caras en 10 lanzamientos de monedas

Tabla 4.5.13. Porcentaje de estudiantes por grupo según valor medio en las cuatro estimaciones en el ítem 3

Valor medio de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [4,2-5,8]	56,7	67,6	71,5
Estimación aceptable [3,4-6,6] <sup>1</sup>	26,8	16,6	13,1
Menor que el aceptable (<3,4)	2,5	3,4	7,8
Mayor que el aceptable (> 6,6)	5,1	4,1	2,9
No completa	8,9	8,3	4,7

<sup>1</sup>Fuera del intervalo normativo

Estas diferencias de respuestas se confirman con el test Chi-cuadrado de independencia entre curso y respuesta, teniendo en cuenta tres categorías de respuestas: optimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas) se obtiene un valor  $\chi^2=16,36$  con 4 g.l., que es estadísticamente muy significativo ( $p=0,003$ ).

La Figura 4.5.8.a, muestra el diagrama de caja asociado al valor medio. Su análisis refleja la simetría de la distribución de la media de las cuatro estimaciones. Además, el rango intercuartílico que representa al 50% de los estudiantes queda contenido en el intervalo que debería contener el 68% de las respuestas. De hecho, el diagrama de caja queda contenido prácticamente en el intervalo definido por las líneas negras que hacen referencia al 95% de las respuestas. Por tanto, la muestra participante presenta una comprensión adecuada del estadístico.

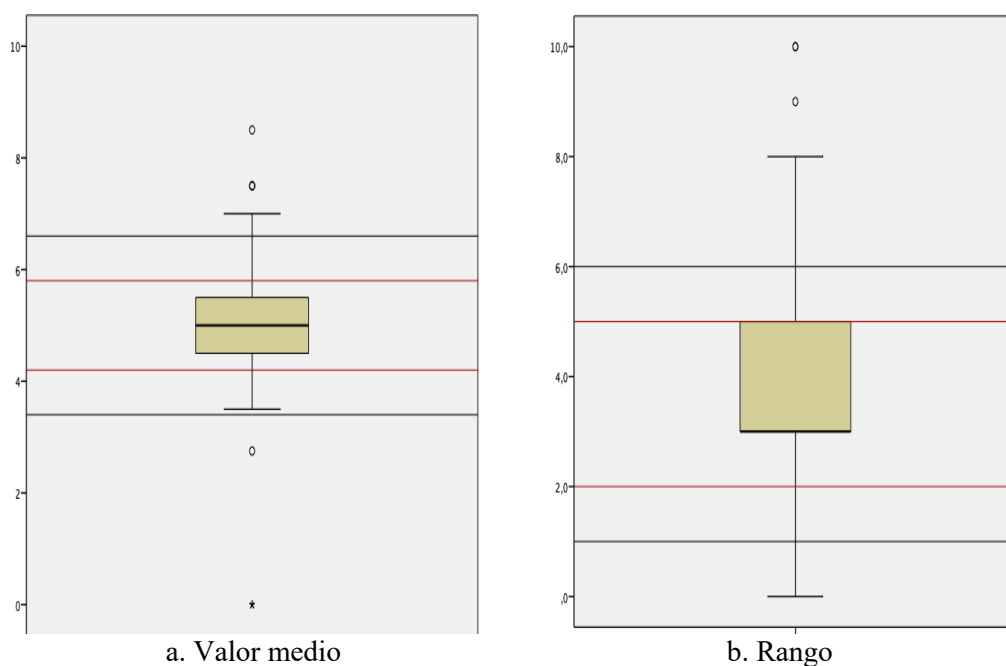


Figura 4.5.8. Diagramas de caja del valor medio y rango de cuatro valores del número de caras en 10 lanzamientos de monedas

### Análisis de la variabilidad

En relación con el análisis de la variabilidad, la distribución del rango (Figura 4.5.7.b) revela que el porcentaje de respuestas mayor se alcanza para el valor medio del estadístico, aunque también se observa un porcentaje de respuestas cuya variabilidad es nula. Con la Tabla 4.5.14 se completa el análisis del gráfico y se observa que el porcentaje

de estudiantes que responden de manera adecuada, en términos de la variabilidad dada a la muestra, es el 72,6%, siendo este porcentaje mayor que el obtenido en los grupos del Estudio 1. No obstante, disminuye un poco la proporción de estimaciones aceptables, fuera del intervalo normativo, por lo que el porcentaje de respuestas correctas es muy similar en los tres grupos (75,8%, 73,8% y 74,2% respectivamente) y el 9,8% de alumnos no conceden variabilidad a las muestras. Además, este porcentaje es mayor que el obtenido en el estudio anterior. Por tanto, estos alumnos tienen una comprensión muy pobre del proceso de muestreo y el grado de madurez de los alumnos no ayuda a eliminar este sesgo.

En cuanto al análisis del diagrama de caja (Figura 4.5.8) observamos, en primer lugar, la asimetría del diagrama, donde el primer cuartil coincide con el valor de la mediana. Por otro lado, el intervalo que representa al 95% de las respuestas incluye al 50% central de la muestra e incluye también una parte de los bigotes. Por tanto, la mayoría de los estudiantes concede una variabilidad adecuada, como también se desprende de la Tabla 4.5.14.

Tabla.4.5.14. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 3

Rango medio de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [2-5]	66,9	66,2	72,6
Estimación aceptable [1-6] <sup>1</sup>	8,9	7,6	5,1
Variabilidad excesiva (>6)	10,2	13,8	7,8
Alta concentración (0)	5,1	4,5	9,8
No completa	8,9	8,3	4,7

<sup>1</sup>Fuera del intervalo normativo

Como se ha desprendido del análisis comparativo de la Tabla 4.5.14, observamos ciertas diferencias, las cuales se confirman con el test Chi-cuadrado de independencia entre curso y respuesta, teniendo en cuenta tres categorías de respuestas: optimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas), donde se obtiene un valor Chi=16,37 con 4 g.l., que es estadísticamente muy significativo (p=0,003).

### Contraste de diferencias entre grupos

En este apartado se presenta, en la Tabla 4.5.15, las medias y desviaciones típicas del valor medio y el rango, así como el error de muestreo y el número de estudiantes que responden al ítem en cada grupo. El análisis del valor medio señala que la media tiene, en cada grupo de estudiantes, pequeñas diferencias. Así mismo, se obtiene la misma conclusión del valor medio que toma el rango en cada grupo. Estos datos se resumen de manera gráfica en la Figura 4.5.9.

Tabla 4.5.15. Estadísticos de la media y rango en el Ítem 3 por curso

	Curso	N	Media	D. típica.	Error muestreo
I3Media	2º	143	5,142	,9441	,0790
	4º	133	5,105	,9182	,01010796
	bachillerato	234	4,689	1,5449	
I3Rango	2º	143	4,308	1,8064	,1495
	4º	133	4,421	2,1005	,1821
	bachillerato	234	3,491	,1353	2,0701

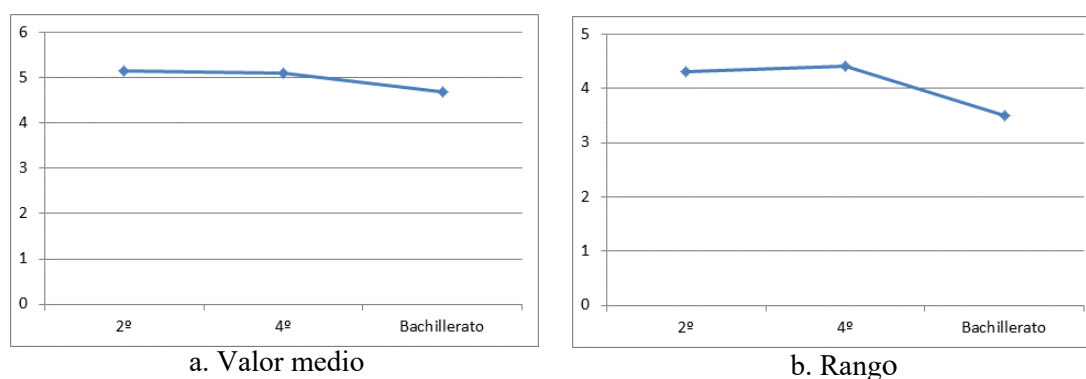


Figura 4.5.9. Valores medios de la media y el rango de las cuatro estimaciones en el ítem 3

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba del análisis de varianza tomando el grupo de estudiantes como variable independiente. Los resultados se presentan en la Tabla 4.5.16, donde observamos que las diferencias son estadísticamente significativas, lo que indican diferentes razonamientos en los distintos grupos de estudiantes.

Tabla 4.5.16. Resultados del análisis de varianza en el ítem 3

		Suma de cuadrados	g.l.	Media cuadrática	F	Sig.
I3Media	Inter-grupos	24,053	2	12,026	7,680	,001
	Intra-grupos	793,916	507	1,566		
	Total	817,969	509			
I3Rango	Inter-grupos	97,334	2	48,667	12,084	,000
	Intra-grupos	2054,034	510	4,028		
	Total	2151,368	512			

Para comprobar cuáles grupos tienen diferencia en la estimación del rango, se realizan las pruebas de Tukey de diferencia de medias, recomendadas una vez que el análisis de varianza da un valor significativo (Tabla 4.5.17). En este caso se obtiene diferencia estadísticamente significativa de la estimación media y en el rango respecto a los dos cursos. Para el rango, las diferencias con los dos grupos son estadísticamente significativas, lo que confirma que la variabilidad de la estimación es menor al avanzar el grupo escolar.

Tabla 4.5.17. Prueba post-hoc de diferencias de medias para la media y el rango de las cuatro estimaciones en el ítem 3

		Diferencia de medias	
Media	Diferencia con 2º	-,4525*	,1328 ,002**
	Diferencia con 4º	-,4162*	,1359 ,007**
Rango	Diferencia con 2º	-,8168*	,2117 ,000**
	Diferencia con 4º	-,9296*	,2179 ,000**

\*\* Muy significativo

#### 4.5.5. RESULTADOS EN EL ÍTEM 4

En primer lugar, la distribución del valor medio presentada en la Figura 4.5.10.a refleja gran concentración en torno al valor medio esperado. Además, la Tabla 4.4.18 resume, en términos de porcentajes, dicha información. Entonces, identificamos que el 62,3% de las respuestas presentan un valor medio situado en el intervalo óptimo. En el caso del Estudio 1 se obtiene el 64,3% para el grupo de 2ºESO y el 66,2% en el caso de 4ºESO. Por tanto, se obtienen mejores resultados en la estimación del valor medio en los cursos inferiores de nuestra muestra participante. Además, si consideramos el porcentaje de respuestas cuyo valor medio se localiza en el intervalo denominado aceptable, entonces obtenemos un porcentaje del 71%, que resulta inferior al que se obtiene en las otras dos muestras de estudiantes consideradas.



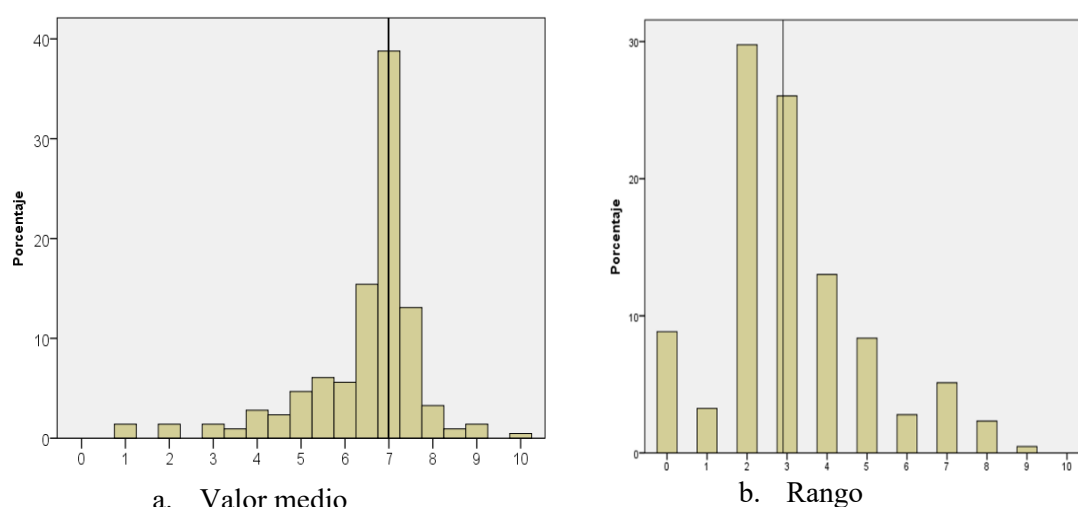


Figura 4.5.10. Distribución del valor medio y rango de cuatro valores del número de aciertos en 10 tiradas a una canasta

Tabla 4.5.18. Porcentaje de estudiantes por grupo según valor medio en las cuatro estimaciones en el ítem 4

Valor medio de las cuatro estimaciones	2°ESO (n=157)	4°ESO (n=145)	bachillerato (n=234)
Estimación normativa [6,3-7,7]	64,3	66,2	62,3
Estimación aceptable [5,6-8,4] <sup>1</sup>	9,1	7,7	8,7
Recencia negativa (<4,5)	3,9	6,3	7,7
Recencia positiva (>8,4)	2,6	4,2	2,1
Equiprobabilidad (4,5-5,5)	9,1	8,5	11,6
No completa	11,0	7,0	7,6

<sup>1</sup>Fuera del intervalo normativo

En cuanto a la presencia de sesgos, identificamos a partir de la Tabla 4.5.18 dos conductas reseñables. Por un lado, identificamos un porcentaje de alumnos que tratan de compensar el valor dado en el enunciado con muestras en las que se espera un número mayor de fallos. Estos alumnos presentan la recencia negativa y el porcentaje es ligeramente mayor al obtenido en la muestra de 4°ESO, pero alto si se compara con los alumnos de 2°ESO. Por otro lado, todavía un pequeño porcentaje de estudiantes muestra el sesgo de equiprobabilidad (Lecoutre, 1992), por lo que estaría en el nivel de razonamiento 3 descrito por Valdez (2016), ya que no consideran los datos frecuenciales.

La Figura 4.5.11.a muestra el diagrama de caja asociado al valor medio de las cuatro estimaciones. Dicho gráfico revela la asimetría de la distribución, así como la concentración de respuestas en torno al valor medio esperado. En relación con este último

aspecto, identificamos que la mediana está próxima al tercer cuartil y al valor medio esperado (7). Por tanto, más de un cuarto de los estudiantes proporcionan muestras cuyo valor medio se sitúa próximo al valor medio teórico.

Además, si observamos los intervalos que señalan al 68% y 95% de las respuestas, algunos valores de la distribución quedan por debajo de estos intervalos. Por tanto, existe en algunos estudiantes una tendencia a compensar el valor dado en el enunciado proporcionando respuestas cuyo valor medio se localiza por debajo del esperado. Estas repuestas están asociadas o bien con la heurística de la representatividad o el sesgo de equiprobabilidad. También se observan una serie de valores atípicos que se sitúan por debajo del bigote inferior. En este caso, los alumnos que proporcionan dichas muestras tratan de compensar el valor frecuencial de acierto, proporcionado en el enunciado, el cual se corresponde con la recencia negativa, que ya se había identificado desde el análisis de los porcentajes mostrados en la Tabla 4.5.18. Por tanto, los alumnos de este estudio tienen una comprensión adecuada del estadístico, con resultados ligeramente inferiores a los del Estudio 1.

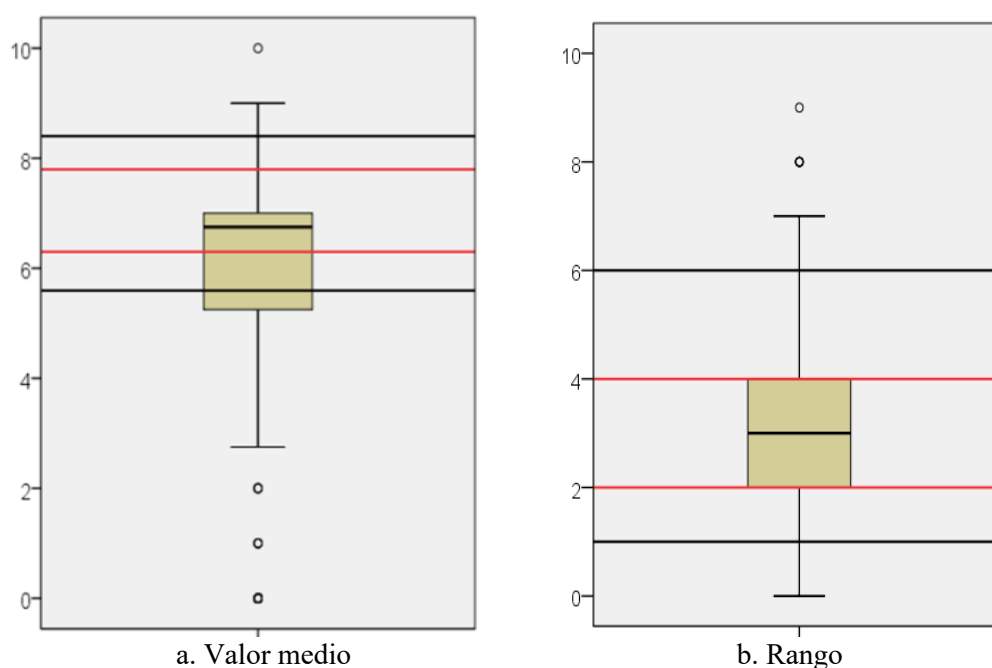


Figura 4.5.11. Diagramas de caja del valor medio y rango de cuatro valores del número de aciertos en 10 tiradas a una canasta

Finalmente, si revisamos el porcentaje de respuestas que se localizan en cada intervalo, se muestra que no existen diferencias significativas según el curso. Un estudio

más completo mediante el uso del test Chi-cuadrado de independencia entre curso y respuesta, análogo al considerado en el estudio de los anteriores ítems, nos permite confirmar las conclusiones obtenidas a partir de un análisis más descriptivo de los datos. En este caso, tras la aplicación del método se obtiene un valor  $\chi^2=2,03$  con 4 g.l., junto con un p-valor de 0,73, por lo que se deduce que no es estadísticamente significativo.

### Análisis de la variabilidad

El análisis de la variabilidad se basa en el rango asociado a las cuatro estimaciones para el número de encestes. El estudio de la distribución de dicho rango en la muestra de estudiantes (Figura 4.4.10.b) muestra que la mayoría de las respuestas se localizan dentro del intervalo normativo [2, 4] o bien aceptable [1, 6]. Esta observación se completa con la Tabla 4.5.19, en la que el porcentaje de respuestas que se sitúa en el intervalo normativo se corresponde con el 63,7%. Además, si consideramos el porcentaje de respuestas aceptables, entonces el porcentaje aumenta en un 13,3%. No obstante, como ya se ha observado desde el análisis de la gráfica, existe un porcentaje de respuestas pequeño (8,1%) con ausencia de variabilidad, por lo que tratan el fenómeno como determinista. Estos estudiantes siguen sus creencias sobre el fenómeno, presentando el sesgo denominado enfoque de resultado (Konold, 1989), ya que interpretan una pregunta probabilística en forma determinista.

En segundo lugar, el análisis se completa con el diagrama de caja asociado al rango (Figura 4.5.11.b), donde se observa que los valores centrales están en el intervalo normativo y la mayoría de los estudiantes en los valores aceptables. Desde el análisis del gráfico se observa cierta asimetría, el número de alumnos cuyas muestras presentan un valor para el rango menor al esperado es alto siendo, por tanto, la longitud del bigote inferior menor.

Tabla 4.5.19. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en el ítem 4

Rango medio de las cuatro estimaciones	2ºESO (n=157)	4ºESO (n=145)	bachillerato (n=234)
Estimación normativa [2,4]	52,2	64,1	63,7
Estimación aceptable [1,6] <sup>1</sup>	15,3	13,1	13,3
Variabilidad excesiva (>6)	7,5	6,2	7,3
Alta concentración (0)	14,0	9,6	8,1
No completa	11,0	7,0	7,6

<sup>1</sup>Fuera del intervalo normativo

En conclusión, el análisis conjunto de los dos gráficos, completado con la tabla de porcentajes (Tabla 4.5.19), refleja que los estudiantes de segundo curso de bachillerato conceden poca variabilidad a las muestras pequeñas. No obstante, el resultado es mejor que el obtenido en el Estudio 1, para los estudiantes de segundo curso, pero no para los de cuarto curso, que tienen mejores resultados que los de bachillerato.

Las diferencias se confirman con el test Chi-cuadrado de independencia entre curso y respuesta, teniendo en cuenta tres categorías de respuestas: óptimas, aceptables y otras (donde en otras se incluyen los diferentes sesgos y las no respuestas), donde se obtiene un valor  $\chi^2=6,65$  con 4 g.l., que no es estadísticamente significativo ( $p=0,156$ ).

### **Contraste de diferencias entre grupos**

Las medias y desviaciones típicas del valor medio y el rango de dicho valor medio en las cuatro estimaciones de cada estudiante en cada uno de los grupos se presentan en la Tabla 4.5.20, junto con el error de muestreo y el número de estudiantes que responden al ítem en cada grupo. Observamos una pequeña diferencia en el valor medio, siendo el correspondiente al cuarto curso algo más alto, es decir, más próximo a la probabilidad teórica. En cuanto al análisis del rango, se identifican pequeñas diferencias, siendo menor para los estudiantes de bachillerato, lo cual indica que sus estimaciones tienen menor variabilidad. Por tanto, parece que visualizan peor que sus compañeros la variabilidad en las muestras pequeñas.

Tabla 4.5.20. Estadísticos de la media y rango en el ítem 1 por curso

	Curso	N	Media	D. típica.	Error muestreo
I4Media	2º	140	6,716	,9432	,0797
	4º	135	6,633	1,1167	,0961
	bachillerato	234	5,847	2,2244	,1454
I4Rango	2º	140	3,06	2,048	,172
	4º	135	3,20	2,003	,172
	bachillerato	234	2,86	,129	1,979

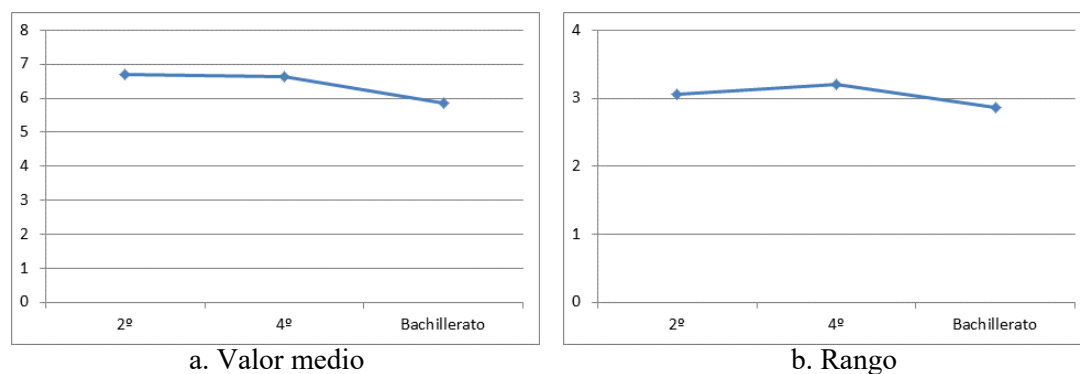


Figura 4.5.12. Valores medios de la media y el rango de las cuatro estimaciones en el ítem 4

Para comprobar si las diferencias observadas son estadísticamente significativas, se ha realizado la prueba del análisis de varianza tomando el grupo de estudiantes como variable independiente. Los resultados se presentan en la Tabla 4.5.21, donde es estadísticamente significativa la diferencia de medias. Sin embargo, no resulta significativa para el análisis del rango, cuyo valor p es mayor que 0,05.

Tabla 4.5.21. Resultados del análisis de varianza en el ítem 4

		Suma de cuadrados	g.l.	Media cuadrática	F	Sig.
I4Media	Inter-grupos	87,194	2	43,597	15,281	,000
	Intra-grupos	1443,602	506	2,853		
	Total	1530,796	508			
I4Rango	Inter-grupos	1,405	1	1,405	,342	,559
	Intra-grupos	1133,189	276	4,106		
	Total	1134,594	277			

Para comprobar los grupos que tienen una diferencia significativa para la estimación de la media, se realizan las pruebas de Tukey de diferencia de medias. Los datos obtenidos de la aplicación de dicho método se recogen en la Tabla 4.5.22. En este caso, se obtiene diferencia estadísticamente significativa de la estimación media de los estudiantes de bachillerato con los dos grupos del primer estudio.

Tabla 4.5.22. Prueba post-hoc de diferencias de medias para la media de las cuatro estimaciones en el ítem 1

		Diferencia de medias	Error típico	Sig.
Media	Diferencia con 2º	-,8688*	,1805	,000
	Diferencia con 4º	-,7861*	,1826	,000

\*significativo

#### **4.5.6. RESULTADOS COMPARADOS POR ÍTEM**

En este apartado, se presenta de manera sintetizada los resultados mostrados en el análisis de cada uno de los ítems. El objetivo es identificar las diferencias debidas a las características propias de cada enunciado.

En primer lugar, la Tabla 4.5.23 presenta el porcentaje de estudiantes que se sitúa en cada categoría, considerando el estudio del valor medio. En este sentido, se observa que el porcentaje de estudiantes que proporciona una estimación mejor del estadístico se corresponde con aquellas situaciones donde los sucesos elementales son equiprobables o enuncian contextos que resultan familiares al estudiante. Es decir, el porcentaje de respuestas normativas es mayor en los últimos tres ítems. Por el contrario, ha habido mayor dificultad en estimar este valor en los casos en que se trata de sucesos no equiprobables y la información de la probabilidad se da en forma frecuencial. Dicha dificultad se observó en los estudios previos de Cañizares (1997), Green (1993a) y Gómez et al. (2014).

Tabla 4.5.23. Porcentaje de estudiantes valor medio en las cuatro estimaciones en cada ítem

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Estimación normativa del valor esperado	20,9	47,9	71,7	59,5
Estimación aceptable del valor esperado	23,4	21,3	13,1	6,3
Menor que el aceptable	14,8	15,2	7,8	17,7
Mayor que el aceptable	2,9	11,1	2,9	2,1
Equiprobabilidad	17,7			9,7
Otros valores	15,9			
No completa	4,7	4,7	4,7	4,7

El sesgo de equiprobabilidad (Lecoutre, 1992), por el que los estudiantes consideran equiprobables sucesos que no lo son, solo se ha podido constatar en las preguntas 1 y 4 donde los sucesos no son equiprobables, pero algunos estudiantes dan estimaciones cuya media es muy próxima al 50%. En los ítems 1 y 4 aparece mayor porcentaje de estimaciones sobre o bajo el valor teórico que atribuimos a la recencia positiva o negativa.

El análisis de la variabilidad, según el tipo de pregunta, presenta una variedad más rica que la estimación del valor esperado. Al analizar la Tabla 4.5.24, se observan relativamente pocos estudiantes con variabilidad normativa en los dos primeros ítems, correspondientes a muestras grandes y un número mucho mayor en los que presentan experimentos con muestras pequeñas (ítems 3 y 4). Hay una alta proporción de

estudiantes con variabilidad excesiva en los dos primeros ítems. Serían estudiantes que no alcanzan el nivel de razonamiento distribucional en el muestreo, descrito por Shaughnessy et al. (2004) y se encuentran en los primeros niveles de razonamiento sobre aleatoriedad y probabilidad, según Valdez (2016), puesto que no comprenden el efecto del tamaño de muestra sobre dicha variabilidad.

Tabla.4.5.24. Porcentaje de estudiantes por grupo según intervalo en que se sitúa el rango de las cuatro estimaciones en cada ítem

Variabilidad	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Estimación normativa	19,6	30,2	69,3	62,3
Estimación aceptable	21,2	24,2	4,9	11,2
Estimación de una variabilidad muestral excesiva	42,8	37,1	7,3	15,0
Alta concentración	11,8	23,1	13,8	7,0
No completa	4,7	4,7	4,7	4,7

Más bien, al contrario de lo que marca la teoría estadística, conceden mayor variabilidad a las muestras grandes que a las pequeñas, un resultado también observado por Serrano (1996).

#### 4.5.7. RELACIÓN ENTRE DIVERSAS RESPUESTAS

Para completar el análisis de los datos, hemos calculado la matriz de correlaciones entre el valor medio de las cuatro estimaciones de cada estudiante en los diferentes ítems (Tabla 4.5.25), e igualmente, la matriz de correlaciones entre el rango obtenido de las cuatro estimaciones de cada estudiante en los diferentes ítems (Tabla 4.5.26). Con ello tratamos de ver si estas estimaciones están o no relacionadas.

Encontramos correlaciones estadísticamente significativas y positivas entre todos los ítems, lo que indica la tendencia de los estudiantes a producir estimaciones altas (o bien medias o bajas) en todos los ítems, independientemente del tamaño de la muestra y el contexto. No obstante, la magnitud de la correlación es moderada (I1 con I3 e I4, I2 con I3; I2 con I4). La mayor correlación se encuentra entre los valores de las estimaciones en los dos primeros ítems, lo que indica un comportamiento similar de muchos estudiantes en dichas estimaciones.

Tabla 4.5.25. Correlaciones entre valor medio de las cuatro estimaciones en cada ítem

		I1Media	I2Media	I3Media	I4Media
I1Media	Correlación de Pearson	1	,612**	,418**	,356**
	Sig. (bilateral)		,000	,000	,000
I2Media	Correlación de Pearson		1	,481**	,371**
	Sig. (bilateral)			,000	,000
I3Media	Correlación de Pearson			1	,474**
	Sig. (bilateral)				,000

\*\* . La correlación es significativa al nivel 0,01 (bilateral).

\* . La correlación es significativa al nivel 0,05 (bilateral).

En cuanto a la correlación entre los rangos de las cuatro estimaciones (Tabla 4.5.26) encontramos valores similares, pero algo más intensos, siempre positivos, en todos los ítems, lo que indica que el estudiante que tiende a dar un valor alto (o bajo) en el rango de sus cuatro estimaciones en un ítem, también lo tiende a dar en el otro. Dicho de otro modo, el estudiante tiende a dar una variabilidad adecuada o bien sobre estimarla o dar una excesiva concentración en varios ítems, independientemente del enunciado. Esta correlación es significativa en prácticamente todos los ítems, con mayor intensidad en la que corresponde a los ítems 1 y 2 (muestras grandes) e ítems 1 y 3 (lanzamiento de monedas), y menor del resto de los ítems con el ítem 4, excepto el ítem 3 (muestra pequeña).

Tabla 4.5.26. Correlaciones entre el rango de las cuatro estimaciones en cada ítem

		I1Rango	I2Rango	I3Rango	I4Rango
I1Rango	Correlación de Pearson	1	,708**	,444**	,469**
	Sig. (bilateral)		,000	,000	,000
I2Rango	Correlación de Pearson		1	,531**	,435**
	Sig. (bilateral)			,000	,000
I3Rango	Correlación de Pearson			1	,493**
	Sig. (bilateral)				,000

\*\* . La correlación es significativa al nivel 0,01 (bilateral).

Ello confirma nuestras anteriores conclusiones de sobre estimación de la variabilidad en muestras grandes, que parece ser característica del grupo de estudiantes.



#### 4.5.8. PUNTUACIONES TOTALES Y PARCIALES EN EL CUESTIONARIO

##### Puntuación total

Para tener una medida global de la respuesta del estudiante y por consiguiente de su comprensión del contenido evaluado en el cuestionario, se calificaron las respuestas obtenidas con el siguiente convenio:

- Se otorgaron dos puntos en cada uno de los ítems, cuando el valor medio obtenido en las cuatro estimaciones dada por el estudiante cae en el intervalo normativo y un punto si cae dentro del intervalo de valores aceptable pero fuera del normativo;
- Se otorgaron dos puntos en cada uno de los ítems cuando el rango obtenido en las cuatro estimaciones dada por el estudiante cae en el intervalo normativo y un punto si cae dentro del intervalo de valores aceptable pero fuera del normativo;
- En el resto de casos, se dio una puntuación cero.

En consecuencia, se obtiene una puntuación numérica que varía entre 0 y 16 puntos. Un estudiante tendría cero puntos si en todos los ítems, el valor medio y rango de las estimaciones es inadecuado; serían los casos que hemos descrito en relación a diferentes sesgos de razonamiento. Un estudiante alcanzaría 16 puntos si todos los valores medios y rangos de sus estimaciones caen en el intervalo normativo. El valor teórico medio de esta variable sería igual a 8 puntos.

En la Figura 4.5.13 se representa la distribución de la puntuación total en los tres grupos de estudiantes, representando con una línea horizontal el valor medio teórico (8 puntos). En primer lugar, hacemos notar que la mitad o más de cada grupo obtienen puntuaciones por encima de la media teórica (8); exactamente la mitad de los estudiantes en cada grupo de educación secundaria y casi las tres cuartas partes en los de bachillerato. No encontramos valores atípicos en ninguno de los grupos.

Todo ello implica que el cuestionario tiene propiedades adecuadas para evaluar los contenidos pretendidos, en el sentido que nos permite obtener una gama de valores (desde altos a bajos), estando centrado en el valor medio teórico o incluso superior.

Se observa un crecimiento de la puntuación global con el curso, estando la mediana del grupo de 2<sup>o</sup> y 4<sup>o</sup> curso situada exactamente en el valor 8, y por encima la de bachillerato. La parte central de la distribución se sitúa entre 6 y 10 puntos en segundo curso de ESO, entre 6 y 11 en cuarto curso y entre 7 y 12 en bachillerato, por lo que hay

un crecimiento progresivo por curso.

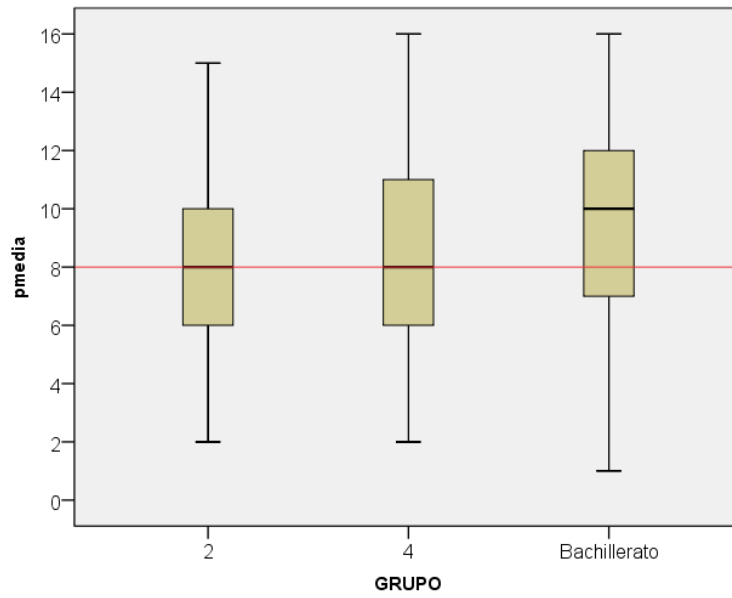


Figura 4.5.13. Distribución de la puntuación total por curso

En general, pocos estudiantes alcanzan las puntuaciones máximas 15 o 16, que supondría tener todas o casi todas las respuestas como normativas. Estos estudiantes mostrarían que han alcanzado el nivel de razonamiento distribucional definido por Shaughnessy et al. (2004), pues no sólo son capaces de estimar el valor medio en cada ítem, sino predecir una variabilidad normativa, que tendría también en cuenta el tamaño de la muestra.

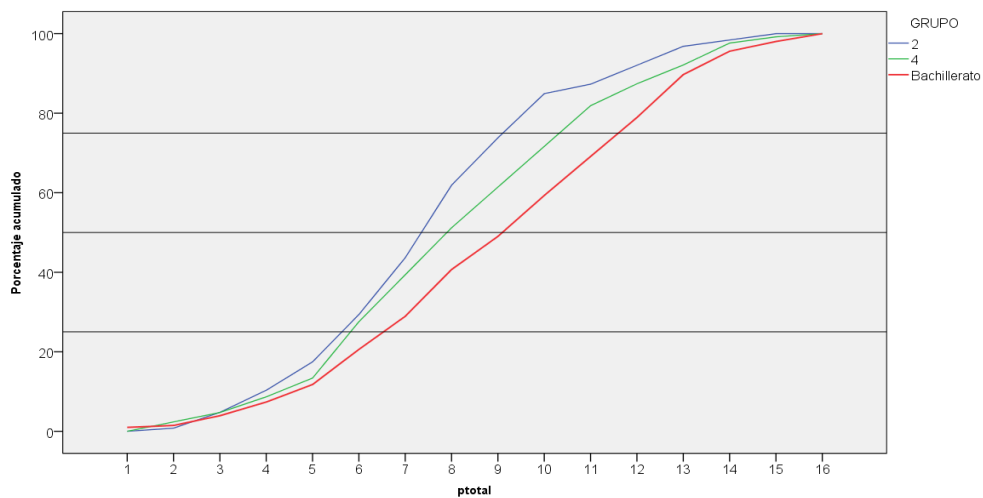


Figura 4.5.14 Distribución acumulada de la puntuación acumulada por curso

La información anterior se complementa con la representación de la distribución acumulativa de la puntuación total en los tres grupos (Figura 4.5.14), en que se observa que, para cualquier puntuación según el curso, aumenta el número de estudiantes que alcanzan al menos dicha puntuación.

### Puntuación sobre el valor esperado

Seguidamente estudiamos, primero, la parte de la puntuación correspondiente a la comprensión de la representatividad (media) y después, la correspondiente a la variabilidad (rango), usando el mismo criterio del apartado anterior. En cada una de las distribuciones resultantes se puede conseguir hasta 8 puntos con una media teórica de 4. Al analizar la distribución de la puntuación obtenida por estimación de la media por curso (Figura 4.5.15) se observa, en primer lugar, un crecimiento de la posición mediana al avanzar el curso escolar. En los tres grupos, al menos el 50% de los estudiantes se sitúan sobre la media; y tres cuartas partes en los dos cursos, superiores. Al comparar los intervalos centrales, observamos el mismo desplazamiento, aunque el primer cuartil es idéntico en 4º curso de secundaria y bachillerato, pero sube el tercer cuartil en este grupo. Por tanto, deducimos una similar comprensión creciente de la representatividad con el curso, siendo en bachillerato donde se observa mayor dispersión.

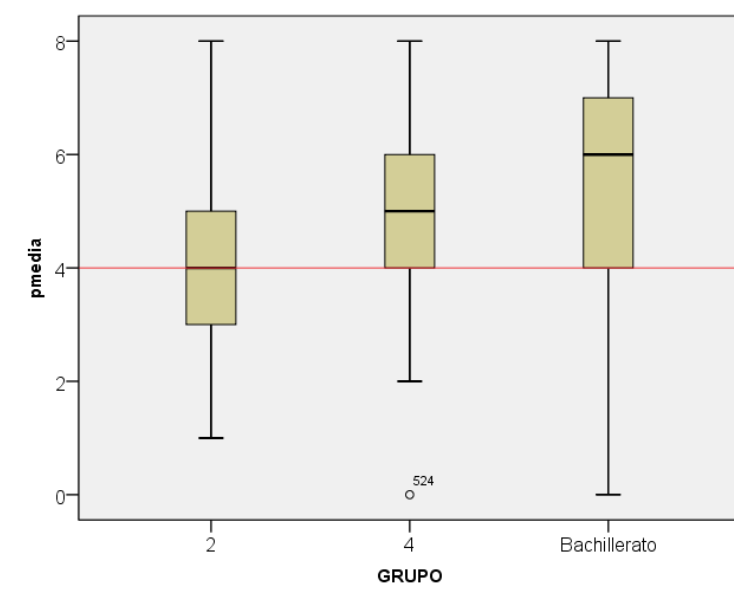


Figura 4.5.15. Distribución de la puntuación en la estimación de la media en cada curso

Como para la puntuación total, esta información se completa con el diagrama de

frecuencias acumuladas por curso, que muestra sistemáticamente mayor porcentaje de estudiantes que alcanza al menos cada una de las puntuaciones en el bachillerato y también en cuarto curso, en comparación con segundo, por lo que se confirma la mejora de la comprensión de los estudiantes al avanzar el curso escolar.

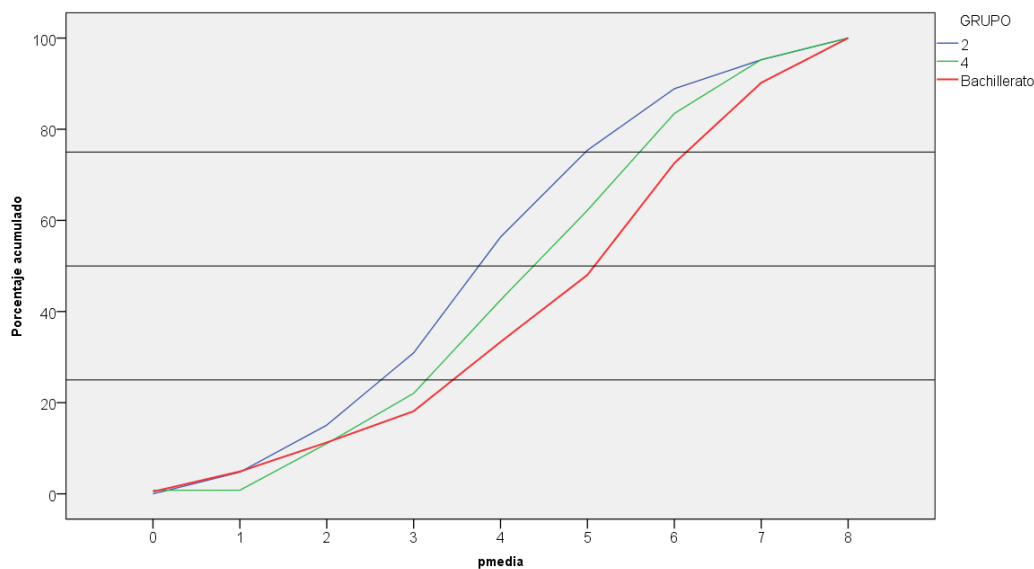


Figura 4.5.16. Distribución acumulada de la puntuación en la estimación de la media por curso

### **Puntuación sobre el rango**

Finalmente, en lo que respecta al rango, los resultados son peores, pero con más diferencia por curso. Las distribuciones de la mitad de los estudiantes en cada grupo no la supera, aunque si lo hace la otra mitad. No hay además mucha diferencia entre los grupos, especialmente entre los dos cursos de secundaria, aunque los estudiantes de bachillerato situados sobre la mediana e incluso los situados entre el primer cuartil y la mediana alcanzan puntuaciones superiores a las de los otros grupos.

Estos resultados se confirman en el diagrama acumulativo, pues son aproximadamente la mitad de los estudiantes en cada grupo los que alcanzan el valor 4 (media teórica). Los gráficos de líneas de los dos grupos de secundaria van muy igualados en todo su recorrido, indicando igualdad de competencia y se separa más el de bachillerato indicando que en cualquier puntuación hay mayor número de estudiantes de este grupo que la alcanza.

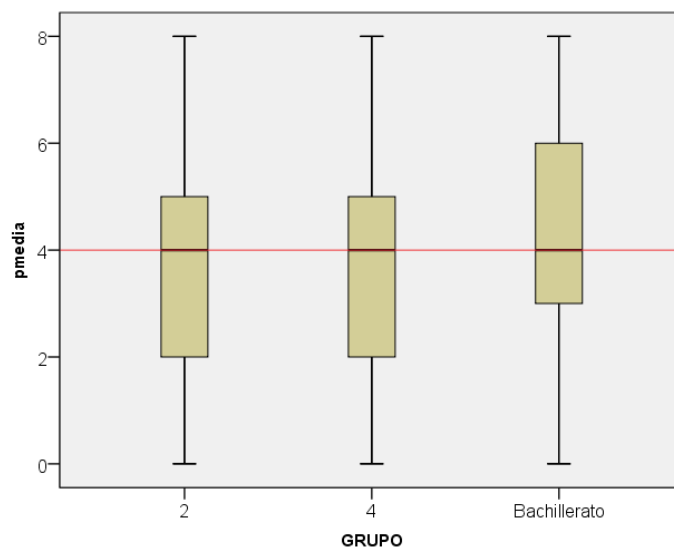


Figura 4.5.17. Distribución de la puntuación total en la estimación del rango cada curso

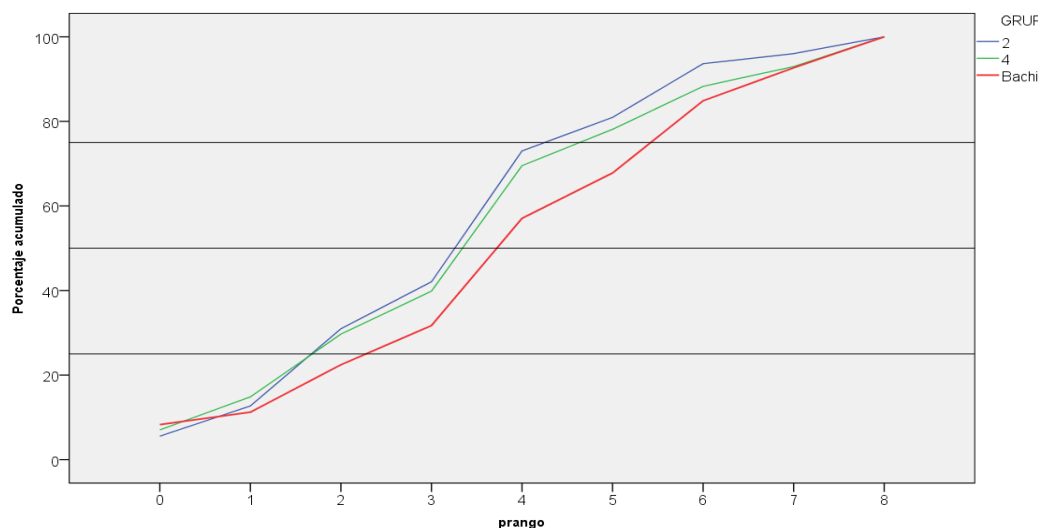


Figura 4.5.18. Distribución acumulada de la puntuación en la estimación del rango por curso

Se deduce peor comprensión de la variabilidad que de la representatividad, una comprensión similar de la variabilidad en los dos grupos de secundaria y algo mayor en los estudiantes de bachillerato.

#### 4.6. ANÁLISIS CUALITATIVO DE LAS ARGUMENTACIONES

Como se ha indicado al inicio de este capítulo, se pide a los alumnos de bachillerato la justificación de la respuesta indicada, lo cual completa el estudio cuantitativo realizado y descrito en las anteriores secciones.

En esta sección se muestra tanto el método de análisis de las argumentaciones como los resultados de dicho estudio. Este análisis se realiza sobre una parte de la muestra participante, en concreto participan 127 de los estudiantes que participan en el estudio cuantitativo. En este sentido, la Tabla 4.4.3 especifica el conjunto de alumnos que participan en este estudio.

#### **4.6.1. MÉTODO DE ANÁLISIS**

El método de análisis que se aplica sobre las justificaciones dadas por los estudiantes se corresponde con el análisis de contenido. De acuerdo con Krippendorff (2013), este método nos permite caracterizar las categorías de análisis de las argumentaciones. Además, es una técnica de investigación destinada a formular, a partir ciertos datos, inferencias reproducibles y válidas que puedan aplicarse a su contexto. Igualmente Weber (1990, p.9) indica que “el análisis del contenido es un método de la investigación que utiliza un sistema de procedimientos para hacer inferencias válidas del texto”. Porta y Silva (2003) sugieren que se trata de una técnica objetiva, sistemática, marcada por la exhaustividad y con posibilidades de generalización, y que puede apoyarse tanto en métodos cualitativos como cuantitativos.

Con la utilización del análisis de contenido, un texto puede dividirse en unidades que pueden ser clasificadas en un número reducido de categorías, en función de variables subyacentes de interés para la investigación, y que permiten realizar inferencias sobre su contenido (Krippendorff, 2013). El método contempla una serie de etapas que se describen a continuación:

- Seleccionar las *unidades de análisis* para el fenómeno que se está investigando. En nuestro caso, las unidades de análisis son cada respuesta de cada estudiante a cada una de las actividades propuestas. Por tanto, como se piden cuatro argumentos a los estudiantes, se han estudiado un total de 508 unidades de análisis.
- Transformar los datos en *unidades de registro*, es decir, porciones mínimas de contenido para su análisis por separado. Estas porciones pueden estar constituidas por palabras, frases o temas que son importantes para el estudio. En nuestro estudio, para pasar los datos a unidades de registro, en primer lugar, se transcribió a un fichero word las respuestas de todos los estudiantes a cada ítem. Al estudiante se le asignó un código para identificarlo, que se corresponde con el código asignado en el análisis

cuantitativo.

- Establecer un *sistema de variables y categorías de análisis* para desarrollar el proceso de codificación de la información, es decir, esta fase se corresponde con el paso de datos brutos a variables y categorías. Las variables se fijaron a priori y consistieron en una para el argumento de cada ítem. Para formar las categorías se compararon todas las respuestas de cada ítem entre sí, agrupando las semejantes para formar una primera lista de categorías. Luego, la investigadora y las directoras del trabajo revisaron la primera agrupación de respuestas a códigos y mediante discusión de casos discordantes y después de sucesivas revisiones se llegó a la lista final de categorías. Un aspecto a comentar es que dicha lista es la misma para todos los ítems, debido a que los argumentos se repiten, es decir, se identifican a lo largo de los cuatro ítems, aunque su frecuencia es diferente, lo cual puede esperarse por las características asociadas a cada ítem.
- Se transforman los datos transcritos en word a un fichero Excel, donde cada alumno es una fila que incluye cuatro columnas (cada una con un código para el argumento dado en los ítems 1 a 4) y otras 16 columnas adicionales (cuatro para reproducir cada una de las estimaciones dadas en los cuatro ítems. Esta información se pasa a un fichero SPSS, con el cual se calculan algunas variables auxiliares, como la media y rango de las cuatro estimaciones, que se utilizaron en la primera parte del capítulo, con el propósito de obtener las diferentes tablas y realizar los contrastes de hipótesis que se usan para obtener conclusiones.

#### **4.6.2. CATEGORÍAS DE ANÁLISIS DE LAS ARGUMENTACIONES**

Tras un estudio de las argumentaciones, mediante el método de análisis de contenido, emergen las categorías o tipos de respuesta que puede dar un estudiante. Como el argumento de un alumno en cada ítem puede apoyarse en justificaciones asociadas a más de una categoría, el porcentaje total de las tablas en las que se analiza la frecuencia de cada tipo de argumento puede ser diferente de 100.

A continuación, describimos cada categoría de respuesta, donde se incluye los ejemplos, de modo que se expliquen las unidades de registro que permiten clasificar la argumentación dada por cada estudiante.

### **TIPO 1. Aleatoriedad**

En esta categoría se incluyen las respuestas de los estudiantes que hacen referencia a términos o expresiones relacionados con el azar, para justificar su forma de realizar sus estimaciones. En el análisis de las respuestas, se puede observar que bastantes estudiantes utilizan la expresión: “*hacer a voleo*”, u otras similares, como: “*porque se me ha venido a la cabeza*”. En este caso, el diccionario de la Real Academia Española (RAE) señala que la primera expresión citada es coloquial, y significa: “Dicho de hacer algo: De una manera arbitraria o sin criterio”. Por tanto, interpretamos que el alumno considera que los resultados obtenidos al realizar un ensayo de un fenómeno aleatorio no responden a ningún juicio o criterio. De hecho, esta concepción del azar se infiere también de la segunda frase indicada. Por tanto, la asignación de los valores asociados a las cuatro muestras va a depender de su concepción de la aleatoriedad, ya que no analizan el enunciado desde la asignación de la probabilidad de un suceso como un valor que mide la verosimilitud de que suceda dicho resultado.

Por otro lado, se ha incluido dentro de esta categoría a los estudiantes que indican que el resultado es imprevisible por ser aleatorio. En este caso, los estudiantes tratan de expresar que han dado respuestas que consideran propias de una secuencia aleatoria, por lo que aplicarían una concepción de la aleatoriedad como sinónimo de un proceso que es impredecible. Esta categoría aparece en otros estudios como Briand (2005), Fischbein et al. (1991) y Savard (2010). Así mismo, se identifica en la investigación de Serrano (1996) sobre las concepciones de la aleatoriedad de los estudiantes.

Esta categoría también se corresponde con aquellas respuestas en las que el estudiante asocia la aleatoriedad con la imposibilidad de medir la probabilidad, lo que, según Serrano (1996), denotaría en los estudiantes el enfoque en el resultado, descrito por Konold (1989). En concreto, el estudiante interpreta una pregunta de probabilidad de manera no probabilística, sino determinista. En relación con nuestro estudio, se demanda: ¿cuál es el resultado más probable? Sin embargo, el alumno interpreta y responde a la pregunta: ¿qué resultado ocurrirá?, la cual exige realizar una predicción segura del resultado. En nuestro estudio, este argumento aparece en todos los ítems.

Algunos ejemplos de respuestas incluidas en esta categoría son los siguientes, las cuales se describen de manera más detallada en las siguientes líneas:

A489 en el ítem 1: “He puesto valores muy diferentes ya que puede salir cualquier resultado.”  
(47, 85, 28, 40).



A521 en el ítem 2: “Porque se me ha venido así en la cabeza.” (60, 57, 30, 45).

A442 en el ítem 3: “Como antes, en el tema de la moneda todo depende de la suerte, así que no siga un patrón, doy valores aleatorios.” (6, 4, 3, 8).

A522 en el ítem 4: “Las respuestas son aleatorias, porque un jugador puede ser malo o bueno.” (8, 5, 1, 7).

Podemos ver que A489 alude directamente a la impredecibilidad de un suceso aleatorio en el ítem 1 y, consecuentemente, produce cuatro valores muy diferentes, intentando repartir de forma equitativa los resultados en el rango de posibles valores de la variable, considerando equiprobables todos los resultados. Este estudiante muestra, por un lado, la concepción de aleatoriedad como impredecibilidad (Serrano, 1996) y también el enfoque en el resultado de Konold (1989). El razonamiento no sería totalmente correcto porque, aunque es cierto que todos los resultados pueden ocurrir, no todos son igualmente probables, por lo que se manifiesta en el estudiante el sesgo de equiprobabilidad. En particular, el estudiante ha propuesto una cuaterna de valores altamente improbables para el fenómeno presentado, por lo que tampoco parece comprender las características básicas de la distribución binomial.

Por su parte A521 simplemente indica que los resultados los ha dado sin pensar e igualmente sus propuestas son muy variadas, aunque en este caso más centradas en el valor 50. Observamos que el estudiante no da ningún valor que coincida con el esperado, pero proporciona dos estimaciones con un valor por exceso y otras dos con un valor por defecto, por lo que podría estar razonando de acuerdo a la heurística de la representatividad; más específicamente de acuerdo a la ley de los pequeños números (Kahneman et al., 1982), al tratar de compensar sus resultados.

El estudiante A442 alude al azar como razón de sus respuestas, aplicando la concepción de aleatoriedad como falta de patrón, que es otra de las concepciones sobre la aleatoriedad descritas por Batanero (2016) y Serrano (1996). De hecho, esta concepción de la aleatoriedad también es incompleta, puesto que estos autores indican que la aleatoriedad puede verse en realidad como multitud de patrones, que serían descritos por el cálculo de probabilidades. Por ejemplo, en el ítem 2 podríamos observar el patrón de la distribución binomial o de la geométrica, Al igual que el caso anterior, el estudiante produce resultados de bastante variabilidad, por encima y debajo del valor esperado, tratando de compensarlos.

A522 alude a resultados aleatorios, pero indicando desconocimiento (el jugador puede ser bueno o malo y el estudiante no lo sabe). No considera las estadísticas de los éxitos dadas por el problema, así que tampoco relaciona la probabilidad con la frecuencia, mostrando falta de comprensión de la probabilidad frecuencial. Presenta una concepción de azar como sinónimo de ignorancia, concepción que es incorrecta, según Batanero (2016), puesto que hay leyes, como las de termodinámica, que se conocen y, a pesar de ello, son aleatorias.

A522 indica que es aleatorio y proporciona resultados muy diferentes entre sí, indicando que para él, la aleatoriedad es máxima variabilidad. Por otro lado, no relaciona la probabilidad con la frecuencia.

En resumen, en las respuestas que hemos analizado como ejemplo de esta categoría observamos los siguientes conflictos semióticos de los estudiantes:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable (A489).* Se asigna a la aleatoriedad una propiedad que no le corresponde, puesto que la mayoría de los fenómenos aleatorios constan de sucesos no equiprobables. Subyace en este conflicto el sesgo de equiprobabilidad descrito por Lecoutre (1992).
- *C2. Suponer que aleatoriedad es sinónimo de impredecibilidad, incluso cuando se considera una serie de resultados (A489).* Como se ha indicado, aunque cada suceso aislado de un fenómeno aleatorio es impredecible (no se sabe si ocurrirá o no) cuando se tienen varias repeticiones del experimento podemos deducir los resultados más y menos probables, pues los teoremas de probabilidad permiten predecir la distribución de estos resultados. Por tanto, subyace el enfoque en el resultado descrito por Konold (1989) en esta concepción de la aleatoriedad.
- *C3. Creencia en la ley de los pequeños números (A521).* Estas respuestas se caracterizan por indicar que la convergencia de la frecuencia relativa observada en una serie de experimentos a la probabilidad teórica, se debe producir incluso en pequeñas muestras. Este razonamiento fue identificado por Tversky y Kahneman (1971) y encontrado en muchas otras investigaciones previas a este trabajo como, por ejemplo, Cañizares (1997) y Serrano et al. (1998).
- *C4. Aleatoriedad como falta de patrón (A442),* en vez de considerarlo como multiplicidad de patrones, como sugieren Batanero (2016) y Serrano (1996).
- *C5. Suponer que la aleatoriedad indica máxima variabilidad (A522).* Se corresponde

con aquellas respuestas que proporcionan resultados muy diferentes entre sí. Aunque el muestreo implique que las muestras obtenidas presentan cierta variabilidad, conceder una variabilidad excesiva a su generación, supone una concepción errónea de la propiedad indicada, puesto que la variabilidad asociada a las muestras generadas mediante un fenómeno aleatorio depende del tamaño de dichas muestras.

- *C6. Asimilar la aleatoriedad a la falta de información (A522)*. Batanero (2016) indica que la concepción de aleatoriedad como sinónimo de ignorancia persistió en varios filósofos hasta comienzo del siglo XX, en que Poincaré (1987/1912) la rebatió. Este autor diferencia entre fenómenos aleatorios, que pueden estudiarse mediante las leyes de la probabilidad, aunque no se conozcan sus causas, y fenómenos no aleatorios, en los que no es posible este estudio.
- *C7. No relacionar la probabilidad con la frecuencia (A522)*. En este caso, se hace referencia a aquellas respuestas que no se apoyan en la información frecuencial dada en el enunciado, por tanto, no se considera el significado frecuencial de la probabilidad y, por consiguiente, la convergencia de la frecuencia relativa a la probabilidad teórica.

## **TIPO 2. Razonamiento basado en aspectos físicos del dispositivo, sin utilizar el lenguaje probabilístico**

Bajo esta categoría se han contemplado aquellas justificaciones en las que se describe algún rasgo físico del dispositivo para argumentar que sea más frecuente la obtención de un resultado que otro, sin incluir términos o expresiones que denoten el uso de lenguaje probabilístico. Subyace en estos argumentos una interpretación de la probabilidad próxima al significado de propensión, debido a Pierce, y que describen Batanero y Díaz (2007). Dicha interpretación constituye una alternativa a la interpretación frecuencial según Roller (2002), pues es aplicable a sucesos físicos singulares.

En concreto, busca especificar el significado de las probabilidades en relación a un sistema físico, en términos de propensiones, que se conciben como las medidas de las posibilidades inherentes a los dispositivos experimentales a las que se corresponden los resultados de los experimentos. Por tanto, la propensión sería una propiedad del sistema físico en ciertas situaciones experimentales, es decir, como propiedades de tales sistemas. Por ejemplo, al realizar el lanzamiento de las chinchetas, sería una propiedad o propensión

el caer de una cierta manera debido a ciertos aspectos físicos como la distribución de la masa del objeto. Por tanto, el sistema físico, en este caso la chincheta, presenta una tendencia causal a caer de cierta manera.

En esta concepción de la probabilidad, un generador aleatorio, por ejemplo un dado, tendría una tendencia de producir cada uno de sus posibles resultados. Esta tendencia sería disposicional y estaría relacionada tanto con la frecuencia relativa a la larga, como con cada resultado singular. Batanero y Díaz (2007) indican que la propensión es diferente de la probabilidad; la propensión sería fuerte si la probabilidad del suceso es alta, pero en el caso del dado, la probabilidad de obtener un 5 al realizar un ensayo es de  $1/6$ , siendo este valor bajo. Sin embargo, para el mismo dispositivo, la propensión a la larga es fuerte, porque esta probabilidad se producirá con mucha seguridad al aumentar el número de ensayos.

Este argumento se identifica en el ítem 1, pues en el caso de las otras preguntas, se emplean rasgos físicos del dispositivo, pero se añade además algún tipo de lenguaje probabilístico. Por tanto, los estudiantes cuya respuesta se localiza en esta categoría identifican la asimetría del dispositivo o el mayor peso de la base de la chincheta para argumentar sus respuestas, y suelen argumentar que se obtendrán más chinchetas con la punta hacia arriba. Algunos ejemplos se incluyen a continuación:

A473 en el ítem 1: “Con este objeto he tenido en cuenta el peso de la chincheta, es decir, la base pesa más que la punta por lo que al lanzarlas caerán de una forma pero en ese instante se equilibrará y se decantará por el pincho hacia arriba. De allí que los porcentajes sean más hacia las chinchetas de punta hacia arriba.” (75, 70, 60, 50).

A490 en el ítem 1: “Debido a la forma de la chincheta es más fácil que caiga hacia arriba.” (69, 67, 68, 65).

A498 en el ítem 1: “La forma que posee la chincheta favorece que caiga de esta forma” (60, 57, 48, 71).

A473 identifica la no equiprobabilidad de los sucesos mediante el análisis físico del objeto. Ha tenido en cuenta que el peso de la cabeza, comparado con el de la punta, es mayor, lo que le permite imaginar el modo en que se producirá la caída, es decir, considera que la cabeza rebotará y quedará con la punta hacia arriba. Consecuentemente, produce en tres de sus valores un número mayor de chinchetas con la punta hacia arriba. Al igual que algunos estudiantes en la primera categoría, este estudiante trata de compensar los

resultados a lo largo de las cuatro muestras proporcionadas dando valores por encima y debajo del valor esperado (68) mostrando la heurística de representatividad (Kahneman et al., 1982). En concreto, se muestra la creencia en la ley de los grandes números, al tratar de mostrar la convergencia con sólo cuatro ensayos.

En las respuestas dadas, los alumnos A490 y A498 señalan que la forma del dispositivo favorece que caiga la chincheta hacia arriba, para lo cual se usa un lenguaje coloquial aplicando términos como “más fácil” o “favorecer”, sin explicar con detalle su razonamiento. Estos argumentos, aunque incompletos, son correctos. A pesar de la variabilidad presente en sus respuestas, predominan aquellas cuyas frecuencias son próximas a las dadas en el enunciado, como la respuesta de A490. También se identifican respuestas como la de A498, en las que se trata de compensar los casos por encima y debajo del valor esperado, por tanto, siguiendo la ley de los pequeños números.

De estas respuestas se han identificado los siguientes conflictos semióticos:

- *C3. Creencia en la ley de los pequeños números (A473, A490).* Se corresponde con aquellas respuestas en las que se observa que el sujeto considera que la convergencia se debe producir incluso en pequeñas muestras. Estos estudiantes realizan una generalización indebida de la aproximación frecuencial de la probabilidad.
- *C7. No relacionar la probabilidad con la frecuencia (A498).* Se observa que el estudiante identifica la no equiprobabilidad del fenómeno mediante un lenguaje coloquial no probabilístico: “favorece”. A pesar de que el sujeto considera la información frecuencial dada en el enunciado, el análisis de su respuesta numérica contradice dicha suposición. Por tanto, no se considera el enfoque frecuencial de la probabilidad para la asignación de los valores.

### **TIPO 3. Razonamiento basado en aspectos físicos del dispositivo con lenguaje probabilístico**

En esta categoría de respuesta se alude a aspectos físicos del dispositivo para argumentar la equiprobabilidad o la ausencia de la misma en el experimento. También se usa un cierto lenguaje probabilístico, lo que implica que el estudiante reconoce la aleatoriedad de la situación. Es una variante de la anterior tipología de argumentación, por lo que de nuevo podemos identificarla con una concepción de la probabilidad como propensión (Batanero y Díaz, 2007). Pero, observamos una mejora en la justificación,

puesto que utiliza expresiones que hacen referencia de manera explícita al conocimiento probabilístico del estudiante, usando, por ejemplo, términos como probabilidad, posibilidad o sus graduaciones. Makar y Rubin (2009) indican que el uso del lenguaje probabilístico acerca de una conclusión sobre una situación de incertidumbre es un componente que caracteriza el concepto de razonamiento inferencial informal o el IIEI, como lo denominan estos autores. Además, utilizan los datos, en este caso, el análisis físico del dispositivo como información para obtener una respuesta o conclusión. Por tanto, los estudiantes que dan este tipo de respuesta muestran, en cierto grado, una primera aproximación al razonamiento inferencial.

En todos los ejemplos que siguen se utilizan ideas o términos relacionados con la idea de probabilidad. En concreto, A475 utiliza las ideas de posibilidad, probabilidad y porcentaje, que incluso cuantifica, y el resto la de probabilidad. Los estudiantes A527 y A474, emplean la idea de probabilidad. Además, se identifica que los valores proporcionados son próximos al dado en el enunciado, por lo que estos estudiantes han aplicado la visión frecuencial de la probabilidad al hacer sus estimaciones. En el caso del ítem 1, se alude generalmente al peso mayor de la cabeza de la chincheta o a su forma, o a ambos, como ocurre con A527. Otros estudiantes indican la forma de colocación de las chinchetas en la mano del niño (A501), la forma plana de la base (A475), la fuerza de la gravedad (A517), la mayor base (A522) o el punto de equilibrio (A474).

A474 en el ítem 1: “Siempre habrá más probabilidad de que las chinchetas caigan con la punta hacia arriba ya que el punto de equilibrio es mayor” (65, 61, 73, 70).

A475 en el ítem 1: “En mis respuestas he puesto en todas mayoría en que caigan con la punta hacia arriba ya que creo que hay más posibilidad de que caigan con la punta hacia arriba porque tiene su base es plana. Y los resultados los he puesto parecidos al del profesor porque siempre tendrá la probabilidad parecida, en este caso el porcentaje del profesor es del 68%, los he puesto parecidos” (80, 70, 63, 74).

A517 en el ítem 1: “Es más probable que caigan boca arriba, por la fuerza de la gravedad ejercida. Pesa más la parte de abajo que la de arriba”(70, 52, 45, 82).

A522 en el ítem 1: “Es más probable que caigan con las puntas hacia arriba, porque tienen más superficie” No hay respuesta numérica.

A527 en el ítem 1: “He elegido esos números porque hay mucha probabilidad que caiga hacia arriba porque es más sencillo que caiga así gracias al peso y la forma de la chincheta” (70, 60, 75, 80).

El estudio de las cuaternas proporcionadas vuelve a mostrar la aplicación de ideas de representatividad (creencia en la ley de los pequeños números) en las respuestas de A474 y A527, quienes van alternando valores por encima y debajo del valor esperado.

En el caso de los ítems 2 y 3, se hace referencia a la forma o peso de las dos caras de la moneda para argumentar la equiprobabilidad de los dos resultados, utilizando conceptos probabilísticos como porcentaje y equidad (A473) o igual probabilidad (A485 y A534).

A473 en el ítem 2: “Las monedas se aplica más el azar ya que el peso es totalmente uniforme y cualquier lanzamiento puede provocar que 10 monedas estén de cara y 90 de cruz, puede ser. Pero como puede ser, el porcentaje es más equitativo que en el ejemplo de las chinchetas. Al haber tal cantidad de monedas (100) los porcentajes serán equitativos” (53, 47, 45, 40).

A485 en el ítem 2: “Al tener el mismo peso por ambos lado, la moneda tiene un 50% de probabilidades de caer mostrando cara y otro 50% de hacerlo mostrando cruz, por lo que el número de caras y cruces se encontrarán en torno a esos valores” (46, 51, 58, 45).

A534 en el ítem 3: “Al tener la misma forma en las dos caras tiene la misma probabilidad de caer por las dos caras” (5, 6, 5, 3).

El análisis de las respuestas proporcionadas por los estudiantes revela un mayor grado de precisión en la redacción, debido a la aplicación correcta de conceptos o ideas relacionadas con la aleatoriedad, por tanto, elevan el grado de madurez de la misma, lo cual favorece una resolución mejor de la tarea. Las cuaternas producidas en este caso no muestran un sesgo particular para los estudiantes. Sin embargo, el comentario de A473 revela la idea de equiprobabilidad, no sólo en los resultados de la moneda, sino en la distribución binomial.

Finalmente, señalar que este argumento no aparece en el ítem 4. A continuación, se han encontrado los siguientes conflictos semióticos en las respuestas:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable* (A473). En este caso, el estudiante A473 redacta: “Las monedas se aplica más el azar ya que el peso es totalmente uniforme...”, es decir, el sujeto indica que como es equiprobable entonces el fenómeno es aleatorio. Por tanto, se identifica una concepción de la aleatoriedad a la que se le asigna la propiedad de equiprobabilidad, de manera incorrecta. Esta concepción de la aleatoriedad es descrita por Batanero (2016) y aparece en el estudio

de Batanero y Serrano (1999).

- C3. *Creencia en la ley de los pequeños números* (A474, A527): pensar que la convergencia se debe producir incluso en pequeñas muestras. Dicha creencia fue descrita por Kahneman et al. (1982) y aparece en los estudios de Batanero et al. (1996), Guisasola y Barragués (2002) y Serrano (1996).

#### **TIPO 4. Razonamiento basado en los valores dados en el enunciado o asignación frecuencial de probabilidad**

Se han clasificado en esta categoría a los estudiantes que apoyan su argumento en los datos proporcionados en el enunciado del ítem sobre el experimento. Por tanto, su respuesta puede indicar una comprensión intuitiva de la probabilidad frecuencial, puesto que asignan la probabilidad del suceso de interés teniendo en cuenta los resultados al obtener una muestra.

Además, asumen explícitamente que la probabilidad de dicho suceso ha de estar cerca del valor dado en el enunciado, por lo que las muestras que generan tienen que ser representativas de dicha probabilidad. En el caso de que el tamaño de la muestra sea grande (ítem 1 y 2), implica que el estudiante comprende y aplica intuitivamente la ley de los grandes números. De acuerdo a Ben-Zvi et al. (2015), dicha ley garantiza que las muestras con un tamaño mayor representan mejor a la población de la que fue tomada. De modo que los estadísticos asociados a la muestra están más próximos a los valores de los parámetros de la población.

A continuación, se presentan y analizan algunos argumentos que se sitúan bajo esta categoría, para cada uno de los ítems.

A484 en el ítem 1: “El profesor al tirar las chinchetas ha conseguido que buena parte de ellas caigan con la punta hacia arriba, por lo que me parece probable que al repetir el experimento siga habiendo más chinchetas con la punta hacia arriba que hacia abajo. Aunque también puede ser al revés pero en mi opinión hay menos probabilidades de que sea así.” (55, 80, 48, 70).

A465 en el ítem 1: “Me he guiado por el valor del ejemplo, son resultados parecidos que se acercan a la muestra.” (72, 70, 65, 67).

La respuesta dada por A484 justifica la asignación de la probabilidad a cada suceso, debido al resultado frecuencial dado por el profesor. El estudiante espera la replicabilidad



de los resultados, aunque reconoce que podría no darse, lo cual supone que comprende la independencia de las sucesivas muestras. Esto implica que se encuentra al nivel más alto en la clasificación de Valdez (2016) para este concepto. Sin embargo, desde el análisis de su respuesta, se observa que no reconoce que la variabilidad es pequeña cuando el tamaño de la muestra es grande. Por tanto, el alumno no alcanza un nivel de razonamiento adecuado para este concepto (Valdez, 2016).

Por otro parte, en su argumento se diferencia claramente entre la posibilidad de ocurrir y probabilidad. El estudiante podría razonar de acuerdo a la ilusión de control (Langer, 1982; Yarritu et al., 2014), donde el sujeto no distingue entre los juegos de habilidad y los de azar. Por tanto, como muestra Langer en una serie de experimentos, el sujeto tiene la creencia de poder controlar el azar, pues tiene una expectativa demasiado alta en la probabilidad personal de éxito en la situación. Esta concepción implica falta de comprensión de la aleatoriedad, pues uno de los rasgos de la misma es la imposibilidad de control y se muestra frecuentemente en los jugadores compulsivos. Sin embargo, proporciona dos cuaternas con valores próximos a la equiprobabilidad, es decir, dos valores son muy cercanos al 50%. Por tanto, no es consistente su argumento con los valores asignados.

Por su parte, la respuesta de A465 para el mismo ítem resulta más consistente, debido a que proporciona una cuaterna donde se contempla, por un lado, la idea de representatividad, puesto que el valor medio de las cuatro estimaciones (68,5) se sitúa próximo al valor medio teórico (68). Y, por otro lado, la variabilidad concedida es adecuada. En este caso, siguiendo a Shaughnessy et al. (2014). el alumno alcanzaría un nivel distribucional de razonamiento.

Para el resto de preguntas del cuestionario, presentamos las siguientes argumentaciones:

A412 en el ítem 2: “Según el primer experimento, los resultados entre los que caen de cara o de cruz son mitad y mitad, pero podría variar según la forma en la que están dispuestas las monedas o como se tiran” (62, 57, 46, 52).

A484 en el ítem 2: “Al caer las monedas por primera vez casi ha salido el mismo número de caras que de cruces por lo que las siguientes veces tampoco creo que varíe mucho.” (60, 55, 70, 40).

Si analizamos la respuesta de A412 observamos, en primer lugar, que se guía por los

resultados obtenidos en el primer experimento. Además, este estudiante utiliza el mismo argumento en dos de los ítems y, en los dos casos, combina representatividad y variabilidad en sus cuaternas, con lo que alcanzaría el nivel distribucional de razonamiento descrito por Shaughnessy et al. (2014). De hecho, indica explícitamente que se basa en los resultados anteriores e incluso da una estimación para el ítem 4 de cuál sería el valor esperado (que denomina porcentaje) en un número mayor de lanzamientos. En la respuesta dada para el segundo ítem, identificamos que la variabilidad no se explica por el carácter aleatorio del fenómeno; sino que, para el alumno, su presencia en el proceso de muestreo se debe a otros factores como el modo en que están las monedas o la manera de efectuar el lanzamiento. Por tanto, su justificación se debe a las creencias subjetivas del estudiante sobre el fenómeno presentado.

La respuesta de A484 pone en relieve que ha considerado los datos dados en el enunciado e incluso nombra la variabilidad, la cual se refleja en su respuesta numérica. A pesar de que señala que la variabilidad no es excesiva, si calculamos el rango asociado a la muestra de cuatro valores se obtiene un valor de 30 puntos de diferencia, lo que supone 10 puntos por encima del extremo superior que caracteriza a aquellas respuestas con rango aceptable. Por tanto, podemos observar que el alumno no presenta una adecuada comprensión de la variabilidad asociada al proceso del muestreo.

A492 en el ítem 3: “Análogo al caso anterior.” (“En el lado de la cara hay más peso por eso es más probable que salga cara”) (6, 5, 7, 4).

A435 en el ítem 3: “Misma explicación.” (“El 53% sale cara y el 47% sale cruz, por lo que es probable que el de caras y de cruces sea similar, rondando el 50%”) (4, 5, 6, 6).

El estudiante A492 razona erróneamente, a pesar que tiene en cuenta los valores dados en el enunciado. En concreto, el estudiante considera que, como la muestra obtenida en el enunciado proporciona más número de caras, esta característica es debida al dispositivo, que no está equilibrado. Por tanto, asigna de manera incorrecta la probabilidad al suceso salir cara. Según Valdez (2016), el razonamiento dado se clasifica en el nivel 2 para el concepto de independencia, puesto que el estudiante asigna la probabilidad considerando las frecuencias relativas, pero ignorando el modelo.

Por su parte, la respuesta de A435 para el ítem 3 es la misma que elabora para el ítem anterior, por lo que el estudiante tiene en cuenta la información frecuencial para identificar la equiprobabilidad de los resultados posibles y, en base a ello, razona sobre el

valor asignado a la probabilidad, en términos de porcentaje. Además, el alumno asume la variabilidad intrínseca al proceso de muestreo, ya que indica que la proporción de respuestas toma valores próximos al promedio, mediante los términos “similar ” o “rondar”.

A429 en el ítem 4: “Si de cada 100 marca 70, sus tiros tendrán una probabilidad del 70%, 7/10, por lo que los aciertos no pueden variar mucho” (7, 6, 8, 9).

A412 en el ítem 4: “Si encesta el 70 de 100 tiros, equivale a un 70%. Si lanza 10 tiros, entonces el porcentaje son 7,2 tiros encesta de cada 10. Si en total realiza 40 tiros, debería encestar 28”. (7, 6, 8, 7).

A527 en el ítem 4: “He elegido esos resultados porque ya que la media es de una serie de lanzamientos sería algo similar al ejemplo del principio” (7, 6, 5, 8).

El estudiante A429 calcula la probabilidad frecuencial de acierto, dada por la frecuencia del enunciado, y la relaciona con el número esperado de éxitos en 10 lanzamientos. De manera análoga razona el estudiante A412 e indica además el número de encestes en 40 lanzamientos aplicando la idea de proporcionalidad. Además, da un valor aproximado del número de encestes debido a la variabilidad aleatoria que subyace en el fenómeno. Por su parte A527 realiza la misma reflexión y el mismo tipo de cuaterna. En estas respuestas identificamos que todos los estudiantes estarían en el nivel 4 de razonamiento sobre la idea de independencia (Valdez, 2016), porque asignan probabilidades en función del modelo disponible para los datos.

En resumen, en las respuestas descritas en este apartado se han encontrado los siguientes conflictos semióticos:

- *C7. No relacionar la probabilidad con la frecuencia* (A492). En esta respuesta el estudiante considera la no equiprobabilidad del fenómeno de lanzar una moneda por los resultados obtenidos en el ensayo del enunciado. A pesar de que la muestra generada es adecuada, el análisis de su justificación pone en relieve una comprensión no adecuada del significado frecuencial de la probabilidad y, en concreto, de la convergencia de la frecuencia relativa hacia la probabilidad teórica de un suceso.
- *C8. Ilusión de control* (A484). En este caso, el estudiante escribe lo siguiente: “el profesor ha conseguido”. Si analizamos el lenguaje utilizado y, en concreto, el verbo que considera para hacer referencia a la obtención del resultado del profesor, entonces vemos que el estudiante considera que los juegos de azar se pueden manipular, por lo

que razonaría de acuerdo a la ilusión de control (Langer, 1982; Yarritu, Matute y Vadillo, 2014).

- *C9. Falta de percepción de la independencia de resultados en la repetición de un experimento aleatorio (A465, A527 y A429).* Se presenta en aquellas respuestas que, tras la identificación de la probabilidad de cada suceso, espera que los resultados no varíen respecto a lo obtenido en el primer resultado. En este caso, el estudiante A465 no considera que existan muestras en las que se obtenga un ensayo en el que la proporción de chinchetas sea mayor para el caso de punta hacia abajo. Análogo al caso anterior, A527 y A429 esperan que los resultados sean similares al del enunciado. Además, estas respuestas tienen consecuencias en la adquisición adecuada de la idea de variabilidad, puesto que los estudiantes razonan dependiendo de los datos obtenidos en el experimento y, esperan, que se vuelvan a obtener en repetidos ensayos del mismo. Por tanto, la no identificación de la independencia conlleva que no se comprende la variabilidad asociada al proceso de muestreo y, por tanto, de la idea de muestreo (Valdez, 2016).
- *C10. Creencias subjetivas sobre el generador aleatorio (A412).* El sujeto identifica la equiprobabilidad del fenómeno aleatorio, pero señala que puede esperar que no se obtenga esa proporción, debido a factores que no tienen que ver con el modelo matemático subyacente. En concreto, el sujeto indica el modo en que están colocadas las monedas o la forma de tirar. Un aspecto reseñable es que estas creencias subjetivas guían la aceptación de la variabilidad en el muestreo, por tanto, el estudiante presenta una comprensión muy pobre del concepto y su relación con la generación de muestras.

#### **TIPO 5. Asignación de la probabilidad desde el punto de vista clásico.**

En esta categoría se clasifican aquellas argumentaciones en las que el estudiante indica los casos posibles del experimento descrito en el enunciado y asume que todos los sucesos elementales son equiprobables. Por tanto, interpretamos que la asignación de la probabilidad del suceso se realiza desde el enfoque clásico, apoyado en el uso de la regla de Laplace.

Este argumento es más frecuente en aquellos ítems en los que el fenómeno aleatorio se corresponde con el lanzamiento de una moneda (ítem 2 y 3) y, por tanto, el fenómeno aleatorio cumple el principio de equiprobabilidad. Sin embargo, también encontramos

algún caso en el ítem 1, como se observa en los siguientes ejemplos:

A470 en el ítem 1: “Pienso que cualquier resultado que se escriba será correcto. Pienso esto porque cada chincheta que se tira tiene un 50% de probabilidades de caer con la punta hacia arriba y 50% de probabilidades de caer con la punta hacia abajo. De esa manera el resultado de cómo caen las chinchetas, no tiene nada que ver con el resultado que daba el profesor” (73, 2, 100, 0).

A480: En el ítem 1, “Como cada chincheta tiene un 50% de probabilidades de caer un modo o de otro, pueden salir como casos: 1° que salgan más hacia arriba pero con poca diferencia, 2° que salga más hacia arriba con mucha diferencia, 3° que salgan más hacia abajo con poca diferencia, 4° que salgan más hacia abajo con mucha diferencia, 5° que salga la mitad” (54, 50, 41, 36).

A510: En el ítem 1 “He pensado en estadística. Si hay dos posibilidades, la teoría es 50%”(50, 50, 50, 50).

El estudiante A470 explícitamente asume que la probabilidad de cada suceso es del 50%. Por tanto, razona de acuerdo al sesgo de equiprobabilidad (Lecoutre, 1992), mostrando una concepción pobre de la aleatoriedad. Además, considera que aleatoriedad es sinónimo de máxima variabilidad, y en consecuencia, tiene una comprensión muy pobre de dicho concepto. De hecho su razonamiento es consistente con los datos que proporciona, que son altamente variables. Por consiguiente, podemos indicar que su respuesta es consistente con la creencia de que las muestras de mayor tamaño tienen mayor variabilidad, encontrada por Serrano (1996).

Por su parte, el estudiante A480 razona suponiendo equiprobables las posiciones de la chincheta. A diferencia del estudiante anterior, este alumno produce muestras con valores próximos al 50 %, por tanto identifica que en un proceso la variabilidad disminuye según el tamaño de la muestra. Finalmente, el estudiante A510 justifica la asignación de la probabilidad desde el significado clásico de la misma y, además, no identifica la variabilidad intrínseca al muestreo, por lo que se situaría en el nivel 1 de razonamiento para este concepto, descrito por Valdez (2016).

En conclusión, en estos casos se identifica el sesgo de equiprobabilidad (Lecoutre, 1992). Además, según Ko (2016), estos estudiantes se centran en la variabilidad del estadístico, pero no identifican la tendencia central del mismo, por lo que presentan un nivel medio de comprensión de la idea de distribución muestral. Un argumento similar es

dado por A457, que también encuentra difícil que se repita el mismo resultado, mostrando falta de comprensión de la independencia de ensayos repetidos. Consecuentemente, todos sus resultados son diferentes. A continuación, analizamos algunos argumentos en los ítems 2 y 3:

A457 en el ítem 2: “Se supone que la probabilidad de cada opción es del 50%. Sin embargo, eso sólo sucede a priori, ya que difícilmente caerán 50 de 100 cara y cruz 4 veces seguidas. Siguiendo esta regla he puesto números en torno al 50” (48, 53, 51, 50). A530 en el ítem 2: “ Los resultados que he puesto las he decidido partiendo de la base de que hay la misma probabilidad que salga cara a de que salga cruz” (55, 45, 50, 60).

A527 en el ítem 3: “He elegido esos números porque hay la misma probabilidad de que caiga de cara o de cruz”(lo mismo que en la anterior pregunta) (5, 4, 6, 7).

A477 en el ítem 3: “Al igual que en el ítem anterior cada moneda tiene un 50% de salir cara y otro 50% de salir cruz, así que lo más probable sería que los valores estuvieran más próximos al 5 que al 0”( “Hay un 50% de posibilidades de que cada moneda caiga de cara o cruz ya que la moneda es plana por ambos lados”) (5, 4, 6, 7).

La respuesta de A457 refleja que el estudiante tiene un conocimiento previo sobre el modelo probabilístico que le permite asignar la probabilidad de cada suceso. Además, identifica la variabilidad en el muestreo, puesto que considera difícil que se mantenga la proporción de monedas. Por tanto, proporciona valores cercanos al porcentaje indicado. De manera similar razona A530 para el mismo ítem.

En el caso del tercer enunciado del cuestionario, se razona de acuerdo a la identificación del modelo clásico que rige dicho fenómeno aleatorio, produciendo muestras cuyo valor medio se sitúa próximo al esperado y con cierta variabilidad. Por tanto, estos estudiantes muestran el nivel de razonamiento distribucional descrito por Shaughnessy et al. (2014).

Por otro lado, la presencia de este argumento en el ítem 4 conlleva a la presencia del sesgo de equiprobabilidad (Lecoutre, 1992). A pesar de que el estudiante A430 considera la equiprobabilidad en todas las preguntas que configuran el cuestionario, observamos que el número de encuestes de las cuatro muestras tiene una proporción similar a la proporcionada en el enunciado y en contra de su argumento, los resultados que proporcionan se acercan a la probabilidad dada en el ítem.

A430 en el ítem 4: “Lo mismo que en ítem 2” (“Todo a boleo, no sé cuánto será. Podría salir

cualquier cosa, pero es más probable que los números estén cerca de 50”) (7, 8, 6, 7)

En resumen, en las respuestas clasificadas en esta categoría se han encontrado los siguientes conflictos semióticos:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable* (A470 y A480). El análisis de los argumentos revela que ambos estudiantes razonan de acuerdo al sesgo de equiprobabilidad, descrito por Lecoutre (1992), donde identifican que el fenómeno, al ser aleatorio, tiene el 50% de ocurrencia de obtener cada suceso. Ambas respuestas presentan ciertas diferencias con respecto a la de A510, puesto que este alumno razona de acuerdo al modelo clásico del cálculo de probabilidades, es decir, aplica la regla de Laplace (“He pensado en estadística”) para la asignación de la probabilidad.
- *C3. Creencia en la ley de los pequeños números* (A477). La respuesta de este estudiante para el tercer ítem indica que espera que exista convergencia de los valores en torno a la media, independientemente del tamaño de la muestra. Dicha creencia fue descrita por Kahneman et al. (1982) y aparece en los estudios de Batanero, Serrano y Garfield (1996), Guisasola y Barragués (2002) y Serrano (1996).
- *C5. Suponer que la aleatoriedad indica la máxima variabilidad* (A470). En este sentido, la frase del estudiante: “Pienso que cualquier resultado que se escriba será correcto...”, indica que cualquier valor que esté dentro del rango 0-100 se puede considerar correcto, por tanto, el alumno no presenta una adecuada comprensión de la variabilidad, lo cual se observa también desde el análisis de la muestra generada, donde contempla como valores probables tanto el 0 como 100. En este caso, se identifica que el estudiante presenta un conflicto semiótico en relación al significado de los conceptos de probabilidad y posibilidad, que los consideran como sinónimos bajo el paradigma de los procesos donde aparece el azar.
- *C7. No relacionar la probabilidad con la frecuencia* (A470). El estudiante indica que los resultados dados no tienen que ver con la información del enunciado, por lo que el estudiante no razona de acuerdo al enfoque frecuencial. Por tanto, estas respuestas implican que no exista una comprensión del concepto de probabilidad frecuencial así como de la convergencia de la frecuencia relativa a la probabilidad teórica.
- *C11. Aplicación estricta de la proporcionalidad en relación con el tamaño de la*

*muestra, sin considerar la variabilidad del muestreo (A510). La muestra proporcionada se caracteriza por una ausencia de variabilidad, donde el valor dado se obtiene a partir de la probabilidad de ocurrencia para cada suceso.*

**TIPO 6. Referencia a la equiprobabilidad, justificada por el carácter aleatorio del experimento.**

Esta categoría se corresponde con los argumentos que indican, explícitamente, que los resultados del experimento son equiprobables debido a su naturaleza aleatoria. Es decir, el estudiante considera que puede aplicar el principio de equiprobabilidad independiente del fenómeno aleatorio que se le presenta. Por tanto, presenta el sesgo de equiprobabilidad (Lecoutre, 1992). Este tipo de concepción sobre la aleatoriedad también apareció en el trabajo de Serrano (1999) sobre concepciones de aleatoriedad de los estudiantes. A continuación, se presentan ejemplos en los cuales se justifica claramente la equiprobabilidad por el carácter aleatorio del fenómeno.

A430 en el ítem 1: “Que una chincheta caiga hacia arriba o hacia abajo 50/50. Pero como hay 100 y el mundo no es perfecto no van a ser 50/50, sino algo más y algo menos. He elegido número aleatoriamente”(60, 35, 54, 70).

A430 en el ítem 4: “Lo mismo que en ítem 2” (“Todo a boleo, no sé cuánto será. Podría salir cualquier cosa, pero es más probable que los números estén cerca de 50.”) (7, 8, 6, 7).

A479 en el ítem 1: “Las probabilidades de que caigan para arriba o para abajo son las mismas porque cada vez te saldrá un resultado diferente al inicial” (68, 70, 72, 66).

A471 en el ítem 2: “En este caso hay las mismas probabilidades de que salga cara o cruz, es cuestión de aleatoriedad” (50, 72, 39, 55).

A522 en el ítem 3: “La misma explicación que en el ítem 2” (“Hay la misma probabilidad que caiga cara o cruz, por ello todas las respuestas son aleatorias”) (2, 7, 5, 10).

A415 en el ítem 3: “Será un 50%/50% ya que es aleatorio” (8, 4, 3, 5).

El estudiante A430 explícitamente indica que ha elegido sus resultados aleatoriamente. Además, si se calcula el promedio del número de caras, dicho valor es próximo a 50, que es el valor explícito que da a la probabilidad. Por tanto, reconoce y aplica la variabilidad de la situación, pero asigna de manera incorrecta la probabilidad del suceso, ya que razona según el sesgo de equiprobabilidad. Por otro lado, si consideramos su argumento para el ítem 4, observamos que razona de igual manera, puesto que expone que el fenómeno cumple el principio de equiprobabilidad. Sin embargo, los valores dados



no presentan la equiprobabilidad, ya que se sitúan próximos al porcentaje de acierto, siendo el valor medio de los cuatro valores 7, que coincide con el valor teórico.

En el caso de la argumentación de A479 identificamos una incoherencia, puesto que supone que las probabilidades de que la chincheta caiga hacia arriba o hacia abajo son idénticas, pero proporciona valores similares a los del enunciado.

Con respecto al resto de justificaciones incluidas como ejemplo, se corresponden con aquellos ítems en los que se cumple la equiprobabilidad, Pero la asignación de la probabilidad, aunque es correcta, no se fundamenta en un razonamiento puesto que estos alumnos presentan una concepción incorrecta de la probabilidad, puesto que generalizan el significado clásico de la probabilidad a todo fenómeno aleatorio.

Del análisis de las respuestas presentadas, identificamos los siguientes conflictos semióticos:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable (A471 y A415).* Sus respuestas revelan que estos estudiantes presentan la creencia de que los sucesos elementales son equiprobables por su carácter aleatorio. En concreto, los estudiantes indican: "...hay las mismas probabilidades...,es cuestión de aleatoriedad...",o, "será un 50%/50% ya que es aleatorio" respectivamente. Por tanto, estos sujetos razonan de acuerdo al sesgo de equiprobabilidad (Lecoutre, 1992).
- *C2. Suponer que aleatoriedad es sinónimo de impredecibilidad (A522).* Este conflicto se identifica porque el sujeto utiliza expresiones para expresar la imposibilidad de poder dar un resultado, ya que no puede predecir lo que va a ocurrir, como por ejemplo: "Todo a boleo, no sé cuánto será...", la cual se identifica en A522.
- *C5. Suponer que la aleatoriedad indica máxima variabilidad (A430).* Expresiones como "...He elegido número aleatoriamente" o "Podría salir cualquier cosa" redactadas por A430, implican que el sujeto asume como cierta la propiedad de que los resultados de una serie de muestras pueden ser muy diferentes. Aunque la aleatoriedad implica variabilidad, la variabilidad en un fenómeno aleatorio puede ser grande o pequeña, por tanto, el alumno razona de acuerdo a una propiedad incorrecta.
- *C12. Inconsistencia entre la probabilidad teórica asumida y la asignación frecuencial de los datos (A479 y A430).* Como se observa en ambas respuestas, se identifica que no existe relación entre los valores elegidos para la muestra y el valor dado a la probabilidad de cada suceso, el que se fundamenta únicamente en su

significado clásico.

### **TIPO 7. Referencia a la convergencia y/o variabilidad del muestreo**

Esta categoría se corresponde con aquellas argumentaciones en las que se justifica que los resultados deben ser parecidos a los dados en el enunciado, lo que supone que el estudiante percibe la relación entre la proporción en la población y la proporción muestral. Además, el estudiante también añade que se debe esperar una cierta variabilidad y, por consiguiente, percibe igualmente la variabilidad muestral. Por tanto, aplica las propiedades de convergencia de la frecuencia relativa a la probabilidad y de variabilidad en el muestreo. En consecuencia, el estudiante se encuentra en el nivel más alto de razonamiento entre los descritos por Valdez (2016). Según Shaughnessy et al. (2014), el alumno mostraría un razonamiento distribucional que se caracteriza por comprender tanto el valor esperado como la variación de la distribución muestral. A continuación, indicamos dos ejemplos en los que se observa dicho razonamiento, que describimos de manera más detallada posteriormente.

A434 en el ítem 1: “Como el % de chinchetas que caen de cada forma en el caso del profesor: 68% hacia arriba y 32% hacia abajo. Por lo que el porcentaje que tendrá en el caso de los niños será similar. Moviendo el porcentaje mínimamente hasta un 8% por ejemplo, por lo que los márgenes serán un 60%-74% hacia arriba y 24% - 40% hacia abajo, pero siempre teniendo en cuenta que es más frecuente que las chinchetas que caigan hacia arriba, ya que la parte de abajo pesa más, por lo que tiende a ir hacia abajo” (70, 60, 72, 65).

A512 en el ítem 1: “Supongo que si la primera vez cayeron muchas punta arriba este hecho se seguirá repitiendo, aunque alguna vez puede invertirse el papel y caer más boca abajo” (52, 63, 58, 36).

En primer lugar, A434 calcula el porcentaje de chinchetas hacia arriba y hacia abajo del experimento descrito en el enunciado. Luego imagina qué ocurriría, moviendo este porcentaje un poco, en ambos sentidos de la media, y calculando los intervalos en que piensa que podrían darse los resultados. Además, es consistente en su estimación, puesto que los valores dados caen en los intervalos que ha caracterizado para cada suceso. En este caso, el estudiante, según Harradine et al. (2011), conectaría de este modo las tres distribuciones que intervienen en el proceso de muestreo: la distribución de datos en la muestra, la distribución de la población y la distribución muestral.

Por su parte, A512 no señala directamente el valor de la probabilidad, pero si el hecho de que la mayor parte de las chinchetas han caído hacia arriba y es de esperar que se repita. Además, expone que alguna vez el resultado puede variar. En consecuencia, su argumentación es algo más imprecisa, aunque su cuaterna de resultados refleja a la vez la representatividad y la variabilidad. Sin embargo, el valor medio asociado toma un valor próximo a 50, por tanto, se observa que los valores numéricos dados tratan de compensar el dato frecuencial que se indica en el enunciado.

A457 en el ítem 2: “Se supone que la probabilidad de cada opción es del 50%. Sin embargo, eso sólo sucede a priori, ya que difícilmente caerán 50 de 100 cara y cruz 4 veces seguidas. Siguiendo esta regla he puesto números en torno al 50” (48, 53, 51, 50).

En la respuesta de A457 para el segundo ítem se refleja, en primer lugar, que el estudiante razona según el Tipo 5, ya que asigna la probabilidad del suceso desde el enfoque clásico. En segundo lugar, alude explícitamente a la variabilidad que concede a sus respuestas ya que considera difícil obtener exactamente el valor medio.

A473 en el ítem 3: “En este experimento es igual que el anterior pero con un número diferente (-) por lo tanto los porcentajes pueden ser muy variados desde 7-3 a 9-1 a 7-3, debido a que al poco número de monedas hay más variedad” (6, 3, 4, 8).

En el caso de la respuesta elaborada por A473, su análisis es similar a lo indicado en las descripciones anteriores; solamente se añade que el alumno gradúa la variabilidad esperada para este caso, cuando señala que “los porcentajes pueden ser muy variados”. Por tanto, el alumno identifica que si el tamaño de la muestra es menor, entonces se puede esperar una variabilidad mayor en la muestras obtenidas. Según Ko (2016), el estudiante se sitúa en el mayor nivel de comprensión de la distribución muestral.

En las argumentaciones que presentamos para el ítem 4, observamos que se indica la idea de convergencia, es decir, tomar valores próximos al valor medio esperado.

A520 en el ítem 4: “Porque en 40 tiros tocaría marcar 28 tiros y luego esos 28 se deben repartir entre 4 días que dan a 7 pero he puesto algún caso con más y menos porque no debe ser una media” (7, 8, 6, 7).

A426 en el ítem 4: “ Porque suele encestar mucho más que fallar, así que en menos tiros, aunque lo probable es que enceste más, también puede ser que los fallos se concentren en los 10 tiros” (6, 7, 8, 6).

A429 en el ítem 4: “Si de cada 100 marca 70, sus tiros tendrán una probabilidad del 70%, 7/10,

por lo que los aciertos no pueden variar mucho” (7, 6, 8, 9).

Finalmente, del análisis de las argumentaciones para el ítem 4 identificamos un número de estudiantes reseñable que indican la idea de convergencia. En concreto, la respuesta de tanto A520 como A429, que tras calcular la probabilidad teórica estiman el número esperado en 10 lanzamientos y suponen una pequeña variabilidad.

En esta categoría de argumentación se identifica el siguiente conflicto:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable (A512).* Si analizamos conjuntamente el argumento aportado y la muestra generada por dicho estudiante, observamos que el estudiante razona de acuerdo al modelo clásico de probabilidad, puesto el valor medio de las cuatro muestras es 52,5 e indica que espera que caigan con la punta hacia arriba, pero los valores dados son próximos a 50. Por tanto, podemos pensar que este estudiante presenta el sesgo de equiprobabilidad.

### **TIPO 8. Creencias subjetivas.**

La probabilidad es un terreno en que los estudiantes, e incluso algunos sujetos adultos, muestran creencias subjetivas para explicar la ocurrencia de ciertos sucesos ligados a un fenómeno aleatorio. Así, Cañizares y Batanero (2007) describen la creencia en la suerte o en la posibilidad de controlar la ocurrencia del suceso. Por su parte, Konold (1989) señala el enfoque en el resultado, donde se interpreta una pregunta de probabilidad en forma no probabilística. Por tanto, hemos considerado como una categoría los argumentos en los que se exponen criterios que no tienen que ver con los datos del enunciado, ni con la probabilidad, sino que responden a creencias que el alumno presenta o sobre su conocimiento del contexto planteado.

A511 en el ítem 1: “He dado valores aleatorios dentro de un intervalo que a mí me ha parecido” (45, 73, 66, 50).

A453 en el ítem 1: “Suponiendo que la cabeza de la chincheta pesa más que el rabillo de esta lo lógico sería que el número de chinchetas que caerán con el rabito hacia arriba será siempre mayor que que lo que caen hacia abajo. Pero siempre existiría una excepción así que de cuatro niños tres tendrían la cabeza hacia abajo y uno con la cabeza hacia arriba, pero sería muy improbable que fuera 50 a 50” (70, 78, 39, 64).

En relación con el primer ejemplo, A511 no da ninguna razón para la cuaterna que proporciona y algunos valores de la misma están muy alejados de los datos frecuenciales

del enunciado. Por tanto, no se guía por el dato dado para la elaboración de su respuesta numérica. Además, concede una variabilidad excesiva al número de chinchetas que caen con la punta hacia arriba y asocia la aleatoriedad con la validez de cualquier tipo de respuesta, lo que conduce a un exceso de variabilidad.

En cuanto a la justificación A453, en primer lugar, da un argumento de tipo 2, pues se fija en las características físicas del objeto, mostrando la concepción de propensión descrita en Batanero y Díaz (2007) y Rolleri (2012). No obstante, supone, erróneamente, que no puede haber coincidencia de valores; es decir, tiene una concepción aditiva del muestreo, suponiendo que diferentes muestras contienen diferentes elementos (Shaughnessy et al., 2014). Además, no comprende la independencia de diferentes ensayos. Según Pollatsek, Konold, Well y Lima (1984) se debe a una concepción inadecuada del muestreo, además, de acuerdo con Bakker (2004), se debe comprender la independencia para adquirir una concepción correcta del mismo.

Este tipo de argumento también se ha identificado para el ítem 2, tal y como muestran los siguientes ejemplos:

A412 en el ítem 2: “Según el primer experimento, los resultados entre los que caen de cara o de cruz son mitad y mitad, pero podría variar según la forma en la que están dispuestas las monedas o como se tiran” (62, 57, 46, 52).

A449 en el ítem 2: “Cuando lanzamos una moneda hacia arriba hay 50% de cara y otro 50% de cruz. Puede que a un niño le salga 54 caras y 46 cruces y otro niño viceversa eso depende del azar. Luego también influye como lanzas la moneda porque si tú lanzas hacia arriba la moneda y la cruz indicaba hacia arriba tiene alguna posibilidad más” (50, 50, 50, 50).

A412 sugiere como posible causa de variación, el orden en que se tiren las monedas o, la forma de lanzarlas, por su parte, A449 hace referencia al modo en que se lanzan dichas monedas. En ambos casos los estudiantes muestran una concepción causal de la aleatoriedad (Batanero, 2016), que prevaleció históricamente hasta la Edad Media.

Los dos ejemplos que mostramos en el ítem 3 indican expresiones tales como: “creo que podría pasar” o “me ha parecido así”, por tanto no razonan su respuesta de manera justificada. En ambas expresiones el sujeto hace referencia a que su respuesta está guiada por una opinión o juicio previo sobre el fenómeno planteado.

A521 en el ítem 3: “Porque creo que podría pasar” (7, 4, 8, 5).

A529 en el ítem 3: “Porque me ha parecido así” (6, 5, 8, 3).

Y finalmente A470 hace referencia a la técnica del jugador, el día, su motivación o si se encuentra cómodo. Efectivamente todas estas circunstancias pueden influir, pero el estudiante no comprende que al realizar un estudio estadístico de los resultados del jugador, la aleatoriedad de la situación, en realidad, contempla todas estas posibles causas.

A470: En el ítem 4, Es diferente a los otros porque aquí hay una técnica y aparte del azar, influye la técnica y podrá meter más o menos dependiendo del día del jugador. De si está cómodo en el partido, si está motivado. (5, 8, 8, 7).

En las respuestas que se han descrito se identifican los siguientes conflictos semióticos:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable (A453)*. La respuesta de este estudiante indica que el alumno asigna la probabilidad desde el significado clásico. De hecho, expone una causa para explicar el resultado obtenido por el profesor.
- *C5. Suponer que la aleatoriedad indica máxima variabilidad (A453)*. Si analizamos su muestra generada, observamos el valor del rango demasiado grande. Además, el estudiante señala en su argumento que no espera que los datos estén en torno al valor medio esperado (50), concediendo, como hemos comprobado, una muestra con una variabilidad excesiva.
- *C7. No relacionar la probabilidad con la frecuencia (A453)*. Bajo este conflicto identificamos la justificación de A453, puesto que ignora la información frecuencial. Es decir, no relaciona el dato aportado para identificar la probabilidad de cada suceso, por tanto, no guía su respuesta desde el significado frecuencial de la probabilidad, el cual es necesario considerar para resolver la tarea de manera adecuada.
- *C13. Interpretar una pregunta de probabilidad en forma no probabilística (A449)*. Se corresponde con aquellas respuestas cuyas muestras no presentan variabilidad, como la indicada por el estudiante A449.
- *C14. Concepción causal de la aleatoriedad, asimilando la existencia de diferentes causas que provocan el fenómeno aleatorio (A449 y A470)*. Si analizamos las justificaciones de ambos estudiantes identificamos que indican diferentes causas ajenas al fenómeno aleatorio. En concreto, indican: “...influye como lanzas la

moneda...” o “...De si está cómodo en el partido, si está motivado...”. Su consideración se tiene en cuenta en la asignación de la probabilidad de ocurrencia del fenómeno, o de la proporción de cada resultado en la obtención de una muestra.

**TIPO 9. No justifican.**

Son los estudiantes que dan la respuesta numérica, pero no explican los criterios utilizados en la misma.

Una vez codificadas las respuestas, teniendo en cuenta las categorías de argumentos descritas, se realiza un análisis acerca de la frecuencia de cada categoría para cada uno de los ítems. Dicha información se describe en el siguiente apartado. En primer lugar, se realiza un estudio según el ítem, para lo cual nos apoyamos en las tablas en que se presenta el porcentaje de cada tipo de argumento sobre el total de estudiantes, puesto que hay argumentos en los que se identifica más de una categoría. Por tanto, la suma de los porcentajes puede ser superior a 100. Seguidamente, se realiza una síntesis global del conjunto de preguntas.

**4.6.3. RESULTADOS EN EL ÍTEM 1**

El análisis del tipo de argumento en este ítem se apoya en el cálculo de la frecuencia de cada categoría. Esta información se recoge en la Tabla 4.6.1 donde, además, se incluye el porcentaje de dicha categoría en relación al número de alumnos que justifican su respuesta de ese modo. Los tipos de argumentos más concurrentes se corresponden con las categorías 3 y 4, que describimos de manera más detallada en las siguientes líneas.

Tabla 4.6.1. Argumentos utilizados en el ítem 1

	Frecuencia	%
Tipo 1. Responder al azar	16	12,6
Tipo 2. Aspectos físicos sin lenguaje probabilístico	12	9,4
Tipo 3. Aspectos físicos con lenguaje probabilístico	42	33,1
Tipo 4. Valores dados enunciado	35	27,6
Tipo 5. Asignación de la probabilidad desde el punto de vista clásico	3	2,4
Tipo 6. Equiprobabilidad	10	7,9
Tipo 7. Variabilidad y convergencia	10	7,9
Tipo 8. Creencias subjetivas	9	7,1
Tipo 9. No justifica	18	14,2

En primer lugar, el 33,1% aportan una justificación que se apoya en aspectos físicos

del lenguaje, mediante el uso de un lenguaje probabilístico. Además, si consideramos aquellas respuestas que hacen referencia a la forma del dispositivo, es decir, si analizamos el porcentaje de respuestas que se sitúan tanto en la categoría 2 como en la 3, entonces el 42,5% de estudiantes fundamentan su justificación en relación al análisis del dispositivo. Entonces, dichos estudiantes muestran una concepción de la probabilidad como propensión (Batanero y Díaz; 2007; Rolleri, 2012), donde la probabilidad se concibe como la tendencia física de un generador aleatorio a producir un cierto resultado.

En segundo lugar, el 27,6% apoya su justificación en el dato dado al enunciado, que se corresponde con el cuarto tipo identificado. En consecuencia, diremos que el alumno presenta una adecuada comprensión del significado frecuencial de la probabilidad, si aplica de manera adecuada el dato del enunciado y, por tanto, proporciona muestras representativas. Además, el análisis de las respuestas numéricas que se clasifican dentro de esta categoría de argumento revela que los estudiantes conceden una variabilidad a los cuatro valores dados. En consecuencia, un gran porcentaje de estudiantes proporcionan una respuesta coherente, que en la mayoría de los casos corresponde a cuaternas que tienen en cuenta tanto la representatividad como la variabilidad, mostrando un razonamiento distribucional (Shaughnessy et al., 2014).

Por otro lado, el porcentaje de respuestas en las que se hace explícitamente referencia a la convergencia y variabilidad del experimento dado al enunciado y que, por tanto, reflejan un mayor conocimiento sobre el enfoque frecuencial de la probabilidad es moderado. Además, si los estudiantes son capaces de identificar la asimetría del fenómeno es debido a sus conocimientos de otras materias como la física, puesto que hacen referencia en su justificación a conceptos tales como el peso, la distribución de masa, la fuerza,...

Para concluir el análisis de la Tabla 4.6.1, señalar el porcentaje de alumnos que guían su argumento considerando el significado clásico. En concreto, únicamente tres respuestas se apoyan en el significado clásico de la probabilidad, el cual no resulta aplicable en este contexto. Además, si consideramos aquellas respuestas que aluden explícitamente a la equiprobabilidad, entonces se obtiene que el 10% aproximadamente aplican de manera explícita o implícita la suposición de la equiprobabilidad en todos los casos, mostrando el sesgo de equiprobabilidad (Lecoutre, 1992).

En relación con la validez de la justificación o corrección de la misma, existen dos



tipos de justificaciones muy diferentes en relación a la primera pregunta. Por un lado, la primera categoría, así como la 5 y 6 y la 8, muestran una comprensión pobre de la aleatoriedad y, en particular, de la probabilidad. En el caso de la lectura de argumentos que se sitúan bajo la primera y octava categoría, se desprende que los estudiantes no identifican la probabilidad como una medida del grado de incertidumbre, sino que la incertidumbre es probabilidad, por lo que aleatoriedad y probabilidad son considerados como términos similares. Este razonamiento fue descrito por Konold (1989) como enfoque en el resultado. En cuanto a los otros tipos de respuesta indicados, el alumno tiene una concepción de la probabilidad que solamente contempla uno de sus significados, lo cual conduce a la imposibilidad de asignar adecuadamente la probabilidad de aquellos fenómenos que no se localicen bajo ese significado.

Por otro lado, los argumentos que pertenecen al resto de categorías favorecen la elaboración de una respuesta adecuada, ya que reflejan la comprensión adecuada del fenómeno aleatorio que se presenta. En este sentido, encontramos respuestas muy completas, como las correspondientes a los estudiantes A410 y A427, que se presentan a continuación. De su lectura se desprende que ambos estudiantes identifican la asimetría del dispositivo, desde el análisis físico del mismo. Además, A427 presenta una comprensión del enfoque frecuencial adecuada, que lo utiliza para argumentar las muestras que proporciona, e incluso señala la variabilidad intrínseca al proceso de muestreo.

A410: “Sería más probable que las chinchetas caigan con la punta para arriba debido a que la parte circular de la base de la chincheta pesa más, por lo tanto será menos probable que las chinchetas caigan con la punta hacia abajo. Por otro lado, la base de la chincheta es más grande y esa es otra razón por la que podrá caer con la punta arriba” (70, 75, 60, 55)

A427: “Si al tirar la primera caja llena de chinchetas, 68 caen hacia arriba y 32 hacia abajo, he supuesto que las próximas veces que vacien la caja, los valores serán similares. Sin embargo, creo que es más probable que la punta vaya hacia arriba, porque la base es más pesada” (70, 69, 67, 71).

#### 4.6.4. RESULTADOS EN EL ÍTEM 2

En relación a las justificaciones del segundo ítem, la Tabla 4.6.2 muestra que la mayoría de las justificaciones se apoyan en el significado clásico de la probabilidad, a las cuales se podrían sumar las que razonan sobre la equiprobabilidad explícitamente. En este caso, como el fenómeno aleatorio cumple la equiprobabilidad, no es indicativo de sesgo.

Por un lado, el modelo presentado se corresponde con el lanzamiento de una moneda, el cual es familiar a los estudiantes, puesto que se presenta de manera frecuente en el

contexto cotidiano, por lo que tienen una experiencia previa con respecto al mismo. Por otro lado, este fenómeno aleatorio se presenta asiduamente en los libros de texto, tanto en tareas de explicación como en tareas de refuerzo de los contenidos presentados en el tema de probabilidad.

Tabla 4.6.2. Argumentos utilizados en el ítem 2

	Frecuencia	%
Tipo 1. Responder al azar	27	21,3
Tipo 2. Aspectos físicos sin lenguaje probabilístico	2	1,6
Tipo 3. Aspectos físicos con lenguaje probabilístico	11	8,7
Tipo 4. Valores dados enunciado	7	5,5
Tipo 5. Asignación de la probabilidad desde el punto de vista clásico	51	40,2
Tipo 6. Equiprobabilidad	7	5,5
Tipo 7. Variabilidad y convergencia	41	32,3
Tipo 8. Creencias subjetivas	6	4,7
Tipo 9. No justifica	33	27,6

No obstante, observamos que un cuarto de la muestra, aproximadamente, no proporcionan una respuesta argumentativa, por lo que todavía persisten ciertas dificultades que pueden deberse a la falta de familiaridad con la tarea, ya que se demanda al alumno la elaboración de una justificación. Además, el porcentaje de respuestas relacionadas con el primer tipo de argumento ha aumentado en comparación con el primer ítem.

Además, en esta pregunta hay muchos estudiantes que identifican la presencia de la convergencia y/o variabilidad en el muestro. El enunciado les demanda proporcionar cuatro valores probables, lo que permite analizar conceptos propios del proceso de muestreo, como la representatividad y variabilidad. En nuestro estudio, se obtiene que el 32,3% de los estudiantes hacen referencia a algunos de los aspectos o a ambos en sus justificaciones. En concreto, en sus justificaciones los estudiantes indican la imposibilidad de que se mantenga dicho porcentaje a lo largo de las cuatro estimaciones, por lo que señalan la necesidad de variar los datos. Siguiendo a Shaughnessy et al. (2004), estos estudiantes alcanzarían un nivel de razonamiento distribucional, que es un alto nivel de razonamiento (Valdez, 2016).

A diferencia con el ítem anterior, los porcentajes asociados a los argumentos de Tipo 4 y 5 toman valores muy diferentes. De hecho, se observa como una tendencia a escoger un tipo de argumento u otro en función a ciertas características asociadas al fenómeno

presentado. En concreto, como ambas categorías se caracterizan por el significado de la probabilidad sobre el que se apoya en la asignación de la misma. Se puede observar que la muestra escoge el significado clásico para el fenómenos que cumple la propiedad de equiprobabilidad, a pesar de que el enunciado proporciona también un dato frecuencial.

A continuación, presentamos algunos ejemplos que se corresponden con respuestas dadas por estudiantes que dan respuesta al primer ítem, pero que no responden el segundo.

A513: “Cada niño tira la caja y todo va en probabilidades de lo que puede caer con hacia arriba y cuantas hacia abajo” (50, 70, 75, 60).

A515: “Porque es azar y puede pasar cualquier cosa” (1, 78, 100, 50).

A518: “Lo he hecho inventándome el número de chinchetas con punta para arriba y desde ahí lo he relacionado con los datos que me daban (si aumentaban las chinchetas disminuye los de punta hacia abajo)” (64, 70, 75, 66).

A410: “Sería más probable que las chinchetas caigan con la punta para arriba debido a que la parte circular de la base de la chincheta pesa más, por lo tanto será menos probable que las chinchetas caigan con la punta hacia abajo. Por otro lado, la base de la chincheta es más grande y esa es otra razón por la que podrá caer con la punta arriba” (70, 75, 60, 55).

A427: “Si al tirar la primera caja llena de chinchetas, 68 caen hacia arriba y 32 hacia abajo, he supuesto que las próximas veces que vacien la caja, los valores serán similares. Sin embargo, creo que es más probable que la punta vaya hacia arriba, porque la base es más pesada” (70, 69, 67, 71).

Vemos que todos ellos tienen capacidad argumentativa y, en general, han respondido fijándose en el carácter aleatorio del experimento (A513, A515, A518), aunque también se usan los datos previos (A518, A427) o las características físicas (A410).

#### 4.6.5. RESULTADOS EN EL ÍTEM 3

El tercer ítem responde al mismo fenómeno aleatorio presentado en el segundo ítem, pero el número de repeticiones en la distribución binomial subyacente es más pequeño. Si consideramos la descripción de la respuesta correcta para esta pregunta, que se ha presentado en el capítulo anterior, se esperaría que el estudiante identificara una variabilidad mayor debido al tamaño de la muestra en la distribución binomial.

En este sentido, el análisis cuantitativo de la respuestas sobre la variabilidad indicó que alrededor del 77,7% de estudiantes, incluyendo los que no argumentaron, proporcionan muestras cuyo rango se localiza en un intervalo adecuado. Si comparamos dicho valor con el porcentaje de alumnos que argumentan según la categoría 7, entonces se obtiene que un 31,5% (Tabla 4.6.3) son los que indican de manera explícita los términos de variabilidad y/o convergencia. No obstante, el tamaño de la muestra exige aceptar una variabilidad mayor en la tarea de generar muestras. Por tanto, el número de

respuestas que, situándose bajo esta categoría indican esa propiedad en relación a la variabilidad, disminuye considerablemente, en concreto, solamente 16 de las 40 respuestas. Además, si calculamos el porcentaje que supone con respecto al total de la muestra, se obtiene el 9%. Por tanto, la relación entre el tamaño de la muestra y la variabilidad del estadístico no es bien comprendida.

Por otro lado, la Tabla 4.6.3 indica que un 43,3% se sitúan en la categoría de Tipo 5, que se corresponde con aquellas respuestas que aplican la probabilidad clásica. En conclusión, la frecuencia de los dos tipos de respuesta descritos anteriormente implica un alto nivel de razonamiento, de acuerdo con los estudios de Shaughnessy (2004) y Valdez (2016). En cuanto a cada uno de los distintos tipos de argumentaciones, observamos que la frecuencia de cada categoría toma valores similares al ítem anterior. Solamente cabe destacar el alto porcentaje de estudiantes que no justifican, a pesar de que el fenómeno es familiar al alumno.

Tabla 4.6.3. Argumentos utilizados en el ítem 3

	Frecuencia	%
Tipo 1. Responder al azar	29	22,8
Tipo 2. Aspectos físicos sin lenguaje probabilístico	0	0
Tipo 3. Aspectos físicos con lenguaje probabilístico	11	8,7
Tipo 4. Valores dados enunciado	4	3,1
Tipo 5. Asignación de la probabilidad desde el punto de vista clásico	55	43,3
Tipo 6. Equiprobabilidad	5	3,9
Tipo 7. Variabilidad y convergencia	40	31,5
Tipo 8. Creencias subjetivas	4	3,1
Tipo 9. No justifica	33	26

A continuación, indicamos dos ejemplos para este ítem, el primero de los cuales expresa un aumento de la variabilidad debido al número de ensayos, a diferencia del segundo, el cual asigna la probabilidad desde el punto de vista clásico, haciendo solamente referencia a la convergencia de los datos al valor medio esperado:

A473: “En este experimento es igual que el anterior pero con un número diferente (-) por lo tanto los porcentajes pueden ser muy variados desde 7-3 a 9-1 a 7-3, debido a que al poco número de monedas hay más variedad” (6, 3, 4, 8).

A477: “Al igual que en el ítem anterior cada moneda tiene un 50% de salir cara y otro 50% de salir cruz, así que lo más probable sería que los valores estuvieran más próximos al 5 que al 0 “(5, 4, 6, 7).

#### 4.6.6. RESULTADOS EN EL ÍTEM 4

La Tabla 4.6.4 presenta los argumentos de los estudiantes en la última pregunta referida al contexto de lanzamiento a una canasta de baloncesto. En primer lugar, se observa que el 56,7% de los estudiantes apoyan su argumento en el dato proporcionado en el enunciado. Es decir, realizan una estimación frecuencial de la probabilidad, en base a la información dada en el cuestionario.

Tabla 4.6.4. Argumentos utilizados en el ítem 4

	Frecuencia	%
Tipo 1. Responder al azar	8	6,3
Tipo 2. Aspectos físicos sin lenguaje probabilístico	1	0,8
Tipo 3. Aspectos físicos con lenguaje probabilístico	0	0
Tipo 4. Valores dados enunciado	72	56,7
Tipo 5. Asignación de la probabilidad desde el punto de vista clásico	2	1,6
Tipo 6. Equiprobabilidad	1	0,8
Tipo 7. Variabilidad y convergencia	51	40,2
Tipo 8. Creencias subjetivas	22	17,3
Tipo 9. No justifica	29	22,8

Como el número de tiros es pequeño, la variabilidad es mayor, por tanto, se esperaría que la muestra participante no solamente identifique la variabilidad en las estimaciones dadas, sino que también lo exprese en sus argumentaciones. De hecho, el 40,2% de los estudiantes aluden a aspectos relacionados con la representatividad y/o variabilidad, mostrando un razonamiento distribucional del muestreo, según Shaughnessy et al. (2004). No obstante, solamente 4 de las 51 respuestas que se localizan en el Tipo 7 redactan de manera explícita que el aumento de la variabilidad es debido al número de tiros o ensayos que constituyen cada muestra. Entonces, los estudiantes no aprecian el efecto del tamaño de la muestra sobre dicha variabilidad, en términos generales. Por otro lado, el análisis estadístico del rango revela que el 40% de las respuestas se corresponden con muestras cuya variabilidad es muy pequeña, mientras que el análisis de los argumentos revela que los estudiantes identifican la variabilidad en el muestreo. Por tanto, aunque el alumno sea capaz de identificar la variabilidad asociado a un proceso de muestreo, es también relevante que identifique el grado en que puede presentarse dicha variabilidad. Según Ko (2016), los alumnos no alcanzarían un nivel de comprensión mayor puesto que solamente se centran en las propiedades de variabilidad y tendencia central del estadístico en el muestreo, sin llegar a comprender la relación entre el tamaño de la muestra y la

variabilidad.

Además, existen argumentos en los que se identifica varias categorías para un mismo argumento. En este ítem, es reseñable que el 32,3% hacen referencia tanto a la identificación de la probabilidad desde el enfoque frecuencial (Tipo 4) como a los conceptos de representatividad y/o variabilidad (Tipo 7).

A continuación, se muestran ejemplos (A451, A452, A459) en los que se indica la variabilidad en las diferentes muestras, pero también esperan que no varíen mucho del valor medio de aciertos. No obstante, los alumnos A437 y A444 esperan que los datos estén en torno al valor medio, pero aceptan una mayor variabilidad en los diferentes partidos tal y como se identifica en la respuesta dada, donde el rango toma valores mayores o iguales que 5. Finalmente, comentar la respuesta dada por A443, en la que a pesar de que el razonamiento es correcto, la muestra dada no se corresponde con el mismo.

A451: “Comparando el número de tiros libres que suele encestar, y reduciéndolo a diez, es probable que no varíe mucho” (7, 8, 7, 6)

A452: “Si lo normal es que de 100 marque 70 pues de 10 marcaría 7 y más o menos lo he puesto entorno a esos números” (7, 8, 7, 6)

A459: “La mayor probabilidad es 7 de 10 de modo que 7 es lo más probable” (7, 8, 7, 6)

A437: “El número de aciertos va a ser casi siempre mayor al de fallos ya que acierta un 70% (un porcentaje muy elevado). Sin embargo, pueden fallar los porcentajes aunque eso pasar la minoría de veces (partidos)” (7, 5, 8, 10)

A444: “En partidos diferentes con características diferentes (puede tener mejor/peor día, estar o no confiado, estar nervioso,...) los resultados no serán siempre 70% de aciertos pero cuántos más partidos juegue, más se acercará a esa proporción de aciertos. (6, 5, 8, 2)

A443: “Porque si la media es acertar 70%, se supone que sus tiros rondarán ese número de aciertos” (7, 7, 7, 7)

La Tabla 4.6.4 revela que el 17,3% de las respuestas se corresponden con justificaciones en las que se indican aspectos subjetivos o creencias en relación con el fenómeno aleatorio presentado. A continuación, se muestran una serie de ejemplos en los que se observa como el alumno razona en base a concepciones previas:

A410: “Aumentará su número de encestes progresivamente porque tendrá más práctica conforme vaya jugando y entrenando” (6, 7, 8, 9)

A413: “Es más probable fallar tirando 100 tiros libres que 10” (9, 8, 9, 10)

A501: “La probabilidad en este caso no está predefinida, pero es cierto que si un jugador acierta su confianza aumenta y es posible que su efectividad aumente, pero si comienza a fallar en un partido se puede desmotivar y disminuir su eficacia” (9, 5, 7, 8)

#### **4.6.7. SÍNTESIS DE ARGUMENTOS EN DIFERENTES ÍTEMS**

Los apartados anteriores describen el análisis cualitativo de los argumentos empleados para cada pregunta del cuestionario. Para completar dicho análisis, se resume en la Tabla 4.6.5 los porcentajes de los diferentes argumentos utilizados en los ítems. De esta forma, podemos analizar el tipo de argumento utilizado según el enunciado, puesto que cada ítem tiene unas características (Tabla 3.6.1), es decir, podemos observar de qué modo los estudiantes han tenido en cuenta en sus argumentos las variables que hemos introducido en los ítems.

En primer lugar, el tipo de fenómeno aleatorio conlleva diferencias en relación al significado de la probabilidad sobre el que se fundamenta la asignación de la probabilidad. Por un lado, se utiliza la asignación clásica en los dos experimentos en que el fenómeno aleatorio se corresponde con el lanzamiento de una moneda equilibrada. En este caso, los sucesos elementales son equiprobables. Por tanto, se puede inferir el porcentaje de monedas que caen de una manera mediante la aplicación de la Regla de Laplace. No obstante, el alumno podría complementar su razonamiento con la información frecuencial dada en el enunciado. Por otro lado, únicamente se puede emplear la asignación frecuencial en los otros dos experimentos, que constan de sucesos no equiprobables.

Por tanto, los estudiantes muestran una competencia en discriminar estos dos tipos de experimentos al asignar la probabilidad de la forma más conveniente, que se traduce en identificar qué significado de la probabilidad me permite resolver la tarea. Además, la asignación de la probabilidad de cada suceso se apoya en aspectos físicos del dispositivo para el caso del lanzamiento de las chinchetas, siendo este tipo de argumento menos referenciado en las otras cuestiones.

El análisis de la Tabla 4.6.5 muestra que el porcentaje de justificaciones en los que se hace referencia a los conceptos de variabilidad y convergencia toma una presencia mayor en los tres últimos ítems. No obstante, como se ha indicado en el análisis de cada ítem, el porcentaje de alumnos que indican de manera explícita la relación entre la variabilidad y el tamaño de la muestra es muy pequeña.

Otro aspecto que destacamos es el porcentaje de sujetos que producen argumentos del primer tipo, puesto que presentan una concepción muy pobre de la idea de aleatoriedad, lo cual puede tener implicaciones más serias en la comprensión de la

probabilidad. Así mismo, señalamos respuestas que hacen referencia a creencias subjetivas. Además, el porcentaje más alto se obtiene en el caso del lanzamiento a la canasta de baloncesto. A pesar de ser un contexto cercano o familiar al alumno, este aspecto no tiene por qué favorecer la resolución adecuada de la tarea.

Finalmente, observamos que una cuarta parte de los alumnos no responde a aquellos enunciados que hacen referencia al lanzamiento de una moneda, a pesar de que es un modelo que se presenta como ejemplo introductorio de la probabilidad en los libros de texto.

Tabla 4.6.5. Porcentaje de argumentos utilizados en los cuatro ítems

	Ítem 1	Ítem 2	Ítem 3	Ítem 4
Tipo 1. Responder al azar	12,6	21,3	22,8	6,3
Tipo 2. Aspectos físicos sin lenguaje probabilístico	9,4	1,6	0,0	0,8
Tipo 3. Aspectos físicos con lenguaje probabilístico	33,1	8,7	8,7	0,0
Tipo 4. Valores dados enunciado	27,6	5,5	3,1	56,7
Tipo 5. Asignación de la probabilidad desde el punto de vista clásico	2,4	40,2	43,3	1,6
Tipo 6. Equiprobabilidad	7,9	5,5	3,9	0,8
Tipo 7. Variabilidad y convergencia	7,9	32,3	31,5	40,2
Tipo 8. Creencias subjetivas	7,1	4,7	3,1	17,3
Tipo 9. No justifica	14,2	27,6	26,0	22,8

A continuación, se resumen de manera gráfica los datos presentados en la Tabla 4.5.5. Por un lado, la Figura 4.6.1 muestra la presencia de cada tipo de argumento en los diferentes ítems. Por otro lado, la Figura 4.6.2 representa para cada ítem el porcentaje de cada tipo de argumento.

La asignación frecuencial se usa principalmente cuando el valor  $p$  es diferente de 0,5 y la clásica cuando es igual a dicho valor. Los aspectos físicos, principalmente en el primer ítem, y las creencias subjetivas en el último, mientras el resto de argumentos se reparten en todos los ítems.



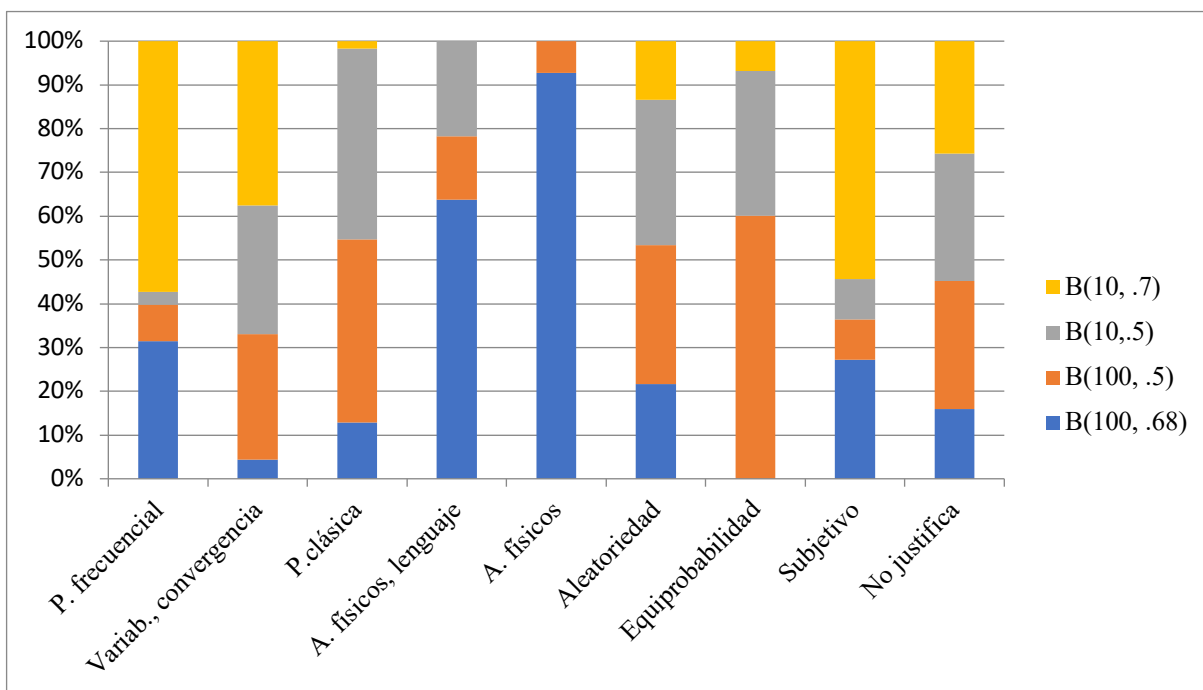


Figura 4.6.1. Porcentaje de uso del mismo argumento en diferentes ítems

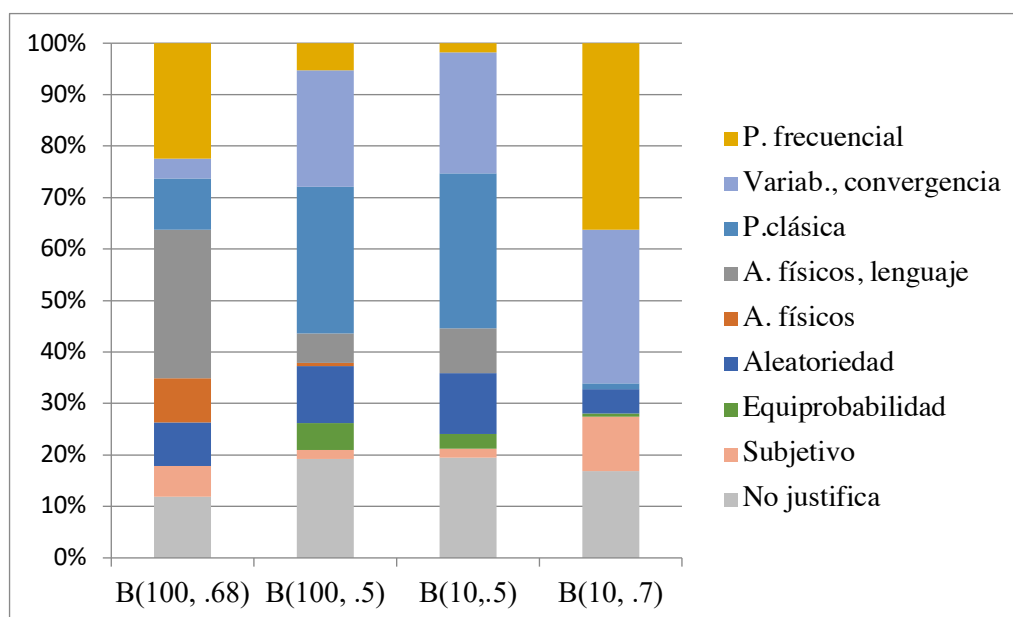


Figura 4.6.2. Porcentajes de argumentos que se presentan en cada ítem

En la Figura 4.6.2 hemos cambiado el orden de los argumentos, de modo que quedan en la parte superior los correctos (hasta argumentos de tipo físico). Los argumentos aleatoriedad, equiprobabilidad y subjetivo serían incorrectos, además de no justificar. Por tanto, obtenemos aproximadamente entre un 32% y un 40% de argumentos incorrectos o alumnos incapaces de dar un argumento en todos los ítems, con poca diferencia entre

ítems. En general, entonces, sobre el 60% de los estudiantes han sido capaces de dar argumentos adecuados para sus respuestas. Queda todavía una parte con poca capacidad de argumentación, punto que debiera tenerse en cuenta en la enseñanza de estos estudiantes.

#### 4.6.8. ESTUDIO DE CASOS CON ARGUMENTOS RECURRENTES

Una posible pregunta es, si algunos estudiantes repiten sus argumentos, independientemente de los ítems, lo que sería indicativo de concepciones específicas sobre la aleatoriedad y el muestreo. Para comprobarlo, se han buscado los alumnos que utilizan un mismo argumento en los diferentes ítems, como muestra de la estabilidad de ciertas concepciones, habiendo encontrado los siguientes casos (Tabla 4.6.6):

Tabla 4.6.6. Cuaternas producidas por alumnos con argumentos recurrentes

Arg	Sujeto	Ítem 1				Ítem 2				Ítem 3				Ítem 4						
1	A464	71	64	80	73		50	43	60	70		5	4	6	7		7	10	6	4
1	A466	72	54	32	49		52	35	90	28		4	2	7	5		8	6	5	7
1	A495	50	80	10	39		50	100	0	70		1	10	9	7		7	8	6	0
1	A532	50	80	10	39		50	45	20	15		4	2	6	5		2	5	7	8
1	A483	60	70	50	37		31	56	16	67		4	7	6	2		7	8	6	9
1	A489	70	72	80	85		70	63	34	90		6	5	7	0		7	8	6	7
1	A503	53	43	71	64		52	50	59	35		6	4	5	7		7	6	8	5
1	A506	60	70	50	37		50	50	50	50		0	7	7	4		7	6	9	5
1	A524	62	58	55	44		52	46	50	49		6	7	5	3		8	7	6	9
3	A436	65	60	75	80		60	53	45	41		6	7	4	5		7	8	6	5
3	A474	65	61	73	70		43	51	55	45		4	5	3	6		7	6	5	7
3	A485	70	66	69	67		46	51	58	45		5	4	3	7		7	8	6	2
3	A473	70	66	69	67		53	47	45	55		6	3	4	8		8	6	7	9
3	A475	80	70	63	74		50	62	47	57		5	6	5	4		7	6	5	7
3	A477	60	67	62	70		50	40	60	55		5	4	6	7		7	6	8	5
4	A447	55	70	75	80		50	43	57	55		3	7	4	5		8	7	9	6
5	A480	54	50	41	36		49	36	58	50		5	4	7	2		5	7	4	9
5	A430	60	35	54	70		60	54	36	40		7	3	6	4		7	8	6	7
5	A463	60	40	30	70		53	47	60	40		6	4	7	3		7	10	6	4
5	A479	65	77	62	68		53	35	57	50		5	4	3	6		7	10	6	5
5	A500	61	50	47	55		14	59	70	30		10	0	5	7		7	8	6	10
7	A490	69	67	68	65		54	52	48	47		7	6	3	5		7	6	8	5

- Los alumnos A464, A466, A495, A532 repiten el argumento 1 en los cuatro ítems y, en consecuencia, se proporcionan valores que no responden a ningún criterio, ya que

dichos alumnos conciben que el azar es imprevisible. Mientras que, los estudiantes A483, A489, A503 A506 y A524 lo repiten en todos excepto en el ítem 4. Mantendrían una concepción de aleatoriedad como imprevisibilidad, descrita en Batanero (2016) y Batanero y Serrano (1999). Estos estudiantes son los que siempre justifican sus muestras en base a la aleatoriedad. El análisis de las muestras aportadas en relación al cálculo de los parámetros considerados revela que estos estudiantes no presentan una adecuada comprensión de la representatividad y variabilidad, para los dos primeros ítems en particular. Por un lado, se observa el sesgo denominado enfoque del resultado (Konold, 1989), debido a que no se considera los datos aportados en el enunciado y tratan de compensar los valores. De hecho, se muestra que el valor medio se sitúa en un intervalo correspondiente a un fenómeno en el que se cumple la condición de la equiprobabilidad. En el caso del ítem 1, implica que el alumno presenta el sesgo descrito por Lecoutre (1992). Por otro lado, el análisis de la variabilidad revela que el rango concedido a las muestras se localiza en los valores extremos (A495, A532, A495, A506). Por tanto, la concepción del azar como un modelo impredecible implica asumir una variabilidad excesiva.

- Los estudiantes A436, A474, A485, A473, A475, A477 repiten el argumento 3 en los tres primeros ítems, es decir, se apoyan en elementos físicos del experimento para justificar las cuaternas proporcionadas, lo que indicaría una concepción de la probabilidad como propensión (Batanero y Díaz, 2007). Estos alumnos consideran la probabilidad como una tendencia del dispositivo a proporcionar un cierto tipo de resultado y usan ideas de probabilidad explícitamente. En este caso, se deduce a partir del análisis de las cuaternas dadas por estos estudiantes, que proporcionan muestras cuyo valor medio y rango se localiza en el intervalo caracterizado como adecuado y aceptable. Por tanto, este argumento favorece la identificación de la probabilidad (clásica o frecuencial) del fenómeno aleatorio estudiado y muestran una estimación adecuada de los parámetros, lo que implica que estos alumnos tienen una comprensión adecuada del muestreo. Siguiendo a Shaughnessy et al. (2004), dichos estudiantes presentan un razonamiento distribucional del muestreo.
- A447 repite el argumento 4 en tres ítems; es decir, este estudiante casi siempre aplica la probabilidad frecuencial para formar sus muestras (aunque también añade a veces otros argumentos). En este caso, se llega a la misma conclusión que en el punto

anterior. El uso explícito de la probabilidad frecuencial, en la mayoría de los ítems, muestra un razonamiento distribucional adecuado sobre el muestreo.

- A480 aplica el argumento 5 en las cuatro cuestiones, mientras que A430, A463, A479, A500 también consideran el enfoque clásico en tres veces. Estos estudiantes tienden a dar cuaternas cuyo valor medio es próximo al 50%, independientemente del enunciado, por tanto muestran el sesgo de equiprobabilidad (Lecoutre, 1992).
- A490 emplea tres veces el argumento tipo 7, por lo que se apoyaría en sus estimaciones en la variabilidad y convergencia, presentando una comprensión adecuada de los conceptos que subyacen a la idea de muestreo y mostrando, asimismo un razonamiento distribucional.

#### **4.7. CONCLUSIONES DEL ESTUDIO 2**

La finalidad del Estudio 2 fue evaluar la comprensión que una muestra de estudiantes de bachillerato tiene sobre el concepto de muestreo, para completar la información recogida en el Estudio 1, centrado en dos cursos de la ESO. Además, con el mismo propósito se pide justificar las estimaciones proporcionadas a una parte de la muestra de estudiantes.

Este apartado tiene como finalidad presentar una reflexión sobre los objetivos y las hipótesis que enmarcan el problema de investigación en relación con los resultados obtenidos.

##### **4.7.1. CONCLUSIONES RESPECTO A LOS OBJETIVOS**

Los objetivos que se definieron para este estudio suponen una concreción del tercer objetivo general de la investigación. A continuación se discuten los resultados obtenidos en relación con los objetivos específicos asociados a este estudio.

*Objetivo 1.4. Evaluar y analizar la comprensión de la relación entre el valor de una proporción en la población y la proporción muestral esperada en muestras tomadas de dicha población en estudiantes de bachillerato y comparar con los resultados obtenidos previamente en estudiantes de educación secundaria obligatoria.*

Este objetivo se ha cumplido razonablemente, porque se ha tomado una muestra de estudiantes de dicho curso, a la cual se le ha pedido responder el instrumento de

evaluación que ha formado parte del Estudio 1, donde el objetivo se redacta de manera similar, pero enfocado en alumnos de segundo y cuarto curso de ESO. En concreto, el análisis de la respuesta detalla el modo en que el alumno debería razonar de acuerdo a la información muestral dada en el enunciado.

La evaluación de la comprensión de la relación poblacional y muestral se apoya en el análisis de la distribución en el muestreo del valor medio de las cuatro estimaciones que cada estudiante proporciona en sus respuestas a los ítems. Dichas distribuciones se representan gráficamente mediante un histograma o diagrama de barras y un gráfico de caja, a los cuales se añaden líneas de referencia que marcan los valores considerados como normativos y aceptables. Además, se presentan tablas de frecuencia de respuestas en estos intervalos y otros que muestran ciertos sesgos de razonamiento de los estudiantes.

Como complemento, la realización del estudio previo con alumnos de cursos inferiores conduce a analizar las diferencias entre los resultados obtenidos. Para lo cual se han considerado una serie de métodos estadísticos apoyados en el contraste de hipótesis con los que poder estudiar si las diferencias son estadísticamente significativas. En concreto, se ha aplicado el contraste Chi-cuadrado (para comparar frecuencias en ciertos intervalos) y el análisis de varianza con comparaciones múltiples para comparar las medias de las distribuciones obtenidas.

Finalmente, como el instrumento se ha tomado de investigaciones previas con ciertas modificaciones o ampliaciones, podemos comparar los datos de este estudio con investigaciones descritas en el Capítulo 2, mostrando, tanto coincidencia de resultados como otros contradictorios y originales.

*Objetivo 1.5. Evaluar y analizar la comprensión de la variabilidad en el muestreo y el efecto del tamaño de la muestra sobre dicha variabilidad en estudiantes de bachillerato y comparar con los resultados obtenidos previamente en estudiantes de educación secundaria obligatoria.*

Igualmente se ha completado este objetivo, pues a partir de las cuaternas proporcionadas por los estudiantes de bachillerato, el análisis de la variabilidad se apoya en el cálculo del rango y su comparación con la distribución muestral de dicho rango de muestras de tamaño 4.

Análogamente al estudio del valor medio, el análisis de las respuestas dadas por los

estudiantes se fundamenta en la elaboración de una serie de gráficos estadísticos, como el gráfico de barras, histograma y el diagrama de caja en el cual se indican los extremos de los intervalos que se definen como normativos y aceptables. Además, se presentan los datos en una tabla de frecuencias donde se recoge el porcentaje de estudiantes cuyas respuestas se localizan en ciertos intervalos previamente definidos. De este modo, se pueden analizar tanto las respuestas que se pueden considerar adecuadas, así como la presencia de sesgos. También se comparan los resultados obtenidos para esta muestra de estudiantes con los que participaron en el primer estudio. Dicho análisis se realiza mediante la aplicación del test chi-cuadrado y el análisis de varianza con comparaciones múltiples. A partir de ello, se establece un análisis análogo al considerado en el Capítulo anterior, lo cual favorece un análisis comparativo de los diferentes grupos evaluados y con las investigaciones previas.

*Objetivo 1.6. Analizar los argumentos que los estudiantes emplean para apoyar la elección de sus respuestas.*

Una parte de la muestra de bachillerato responde de manera razonada a la respuesta dada y se han analizado los argumentos para deducir sus razonamientos. Esta será una contribución original respecto al Estudio 1, donde no se analizaron tales argumentos. Para conseguirlo, se pidió a una parte de la muestra (127 estudiantes) argumentar sus respuestas al cuestionario.

El análisis de dichos argumentos se basa en el análisis de contenido (Hernández et al., 2010; Krippendorff, 2003; Zapico, 2007) típico de la investigación cualitativa. Mediante la lectura sucesiva de las respuestas y partiendo de categorías descritas en investigaciones previas sobre comprensión de la aleatoriedad, se construye y refina un sistema de categorías. Estas categorías se han analizado en profundidad, mostrando ejemplos de cada una de ellas y estudiando los conocimientos implícitamente utilizados o los sesgos subyacentes a las mismas. Además, se compararon las categorías usadas más frecuentemente en cada ítem y se profundizó en las respuestas de los estudiantes que dan argumentos recurrentes. Se completa con una identificación de conflictos semióticos latentes en las respuestas de los estudiantes.

#### 4.7.2. CONCLUSIONES RESPECTO A LAS HIPÓTESIS

En el inicio de este capítulo se describieron una serie de hipótesis previas al estudio, que se incluyen de nuevo a continuación, para revisarlas en relación a los resultados obtenidos.

*HI.3. Algunos estudiantes presentan una comprensión insuficiente de las ideas básicas asociadas al concepto de muestreo.*

Nuestra expectativa era que la mayoría de estudiantes tuviese una buena percepción de la proximidad entre la proporción esperada en la muestra y la proporción en la población. Esta expectativa se cumplió, como muestra el análisis del valor medio. En general, los resultados revelan que la mayor parte de la muestra participante proporcionan muestras cuyo valor medio se localiza en el intervalo que se define como aceptable y/o adecuado. Por tanto, la mayor parte los estudiantes tienen una comprensión adecuada de la proporción muestral esperada, alcanzando al menos el nivel proporcional de razonamiento sobre muestreo en la categorización de Shaughnessy et al. (2004) y un nivel intermedio en la de Valdez (2016).

No obstante, el análisis del ítem 1 revela que un porcentaje de estudiantes de segundo curso de bachillerato no identifica la asimetría del dispositivo, presentando el sesgo de equiprobabilidad (Lecoutre, 1992), que consiste en estimar siempre frecuencias esperadas cercanas al 50%, independientemente de la proporción en la población. Es más, en el análisis de los argumentos, algunos estudiantes explícitamente presentan argumentos correspondientes a este sesgo. Como se ha discutido, en nuestro trabajo no es posible aceptar la teoría de Chernoff y Russell (2012), según la cual, estudiantes razonan que la existencia de la equiprobabilidad es debida a que cometen la falacia de la composición, es decir, transfiriendo a la muestra una propiedad de la población, puesto que en este ítem los sucesos son claramente no equiprobables.

Esperábamos más dificultad en la idea de variabilidad y que no se percibiese el efecto del tamaño de la muestra sobre la variabilidad. Estos resultados se observaron en el Estudio 1, así como en los trabajos de Gómez et al. (2014) y en Shaughnessy et al. (2004).

También se confirmó esta hipótesis, pues se observa una mayor dificultad en proporcionar una variabilidad adecuada a las muestras. En este caso, se observa que son minoría los estudiantes que proporcionan muestras con una variabilidad adecuada en los

ítems 1 y 2. De hecho, se observa que muchos estudiantes proporcionan muestras con una variabilidad excesiva para estos ítems. Por tanto, los alumnos tienen una comprensión intuitiva de la ley de los grandes números insuficiente, ya que conceden una variabilidad muy grande sin tener en cuenta el tamaño de la muestra.

No obstante, la percepción de la variabilidad es mejor en los ítems 3 y 4, que corresponden a muestras pequeñas. Estos resultados contradicen los de autores previos como los de Shaughnessy et al. (2004), seguramente por el tipo de tarea considerado, que en su caso, no permite analizar adecuadamente la variabilidad.

Por otro lado, el análisis de los argumentos revela una serie de creencias parcialmente correctas (pero no totalmente correctas) de los estudiantes respecto a la aleatoriedad y el muestreo, que reproducen las descritas por Batanero (2016) y Batanero y Serrano (1999), como la aleatoriedad como impredecibilidad, aleatoriedad como equiprobabilidad. Al mismo tiempo se observa un razonamiento distribucional completo (Shaughnessy et al., 2004) en una parte de la muestra, que utiliza a la vez las ideas de convergencia y variabilidad.

No obstante, dada la diversidad de objetos que se utilizan en las prácticas matemáticas, el carácter inmaterial de los objetos y la variedad de representaciones utilizadas, el trabajo matemático requiere un uso constante de procesos interpretativos, que pueden llegar a ser incorrectos cuando los que los realizan son los estudiantes. Godino (2002) propone la idea de *conflicto semiótico* como cualquier disparidad entre los significados atribuidos a una expresión por dos sujetos (personas o instituciones). Para completar el trabajo, en las respuestas que hemos analizado hemos observado los siguientes conflictos semióticos de los estudiantes en el análisis de sus argumentos:

- *C1. Confusión entre suceso aleatorio y suceso equiprobable.* Estas respuestas se caracterizan porque el sujeto asigna la equiprobabilidad a cada suceso por el carácter aleatorio del fenómeno. Por tanto, tienen una comprensión incorrecta de la aleatoriedad, ya que le asignan una propiedad que no le corresponde. En los argumentos que se sitúan en este conflicto subyace el sesgo de equiprobabilidad, descrito por Lecoutre (1992).
- *C2. Suponer que aleatoriedad es sinónimo de impredecibilidad.* Este conflicto se relaciona con aquellas respuestas en las que el alumno expresa, de cierto modo, que no es capaz de indicar un resultado incluso cuando se considera una serie de



resultados. Como se ha indicado, aunque cada suceso aislado de un fenómeno aleatorio es impredecible (no se sabe si ocurrirá o no), cuando se tienen varias repeticiones del experimento podemos deducir los resultados más y menos probables, pues los teoremas de probabilidad permiten predecir la distribución de estos resultados. Bajo este conflicto subyace el enfoque en el resultado, descrito por Konold (1989).

- *C3. Creencia en la ley de los pequeños números.* El sujeto considera que la convergencia de la frecuencia relativa observada en una serie de experimentos a la probabilidad teórica se debe producir, incluso en pequeñas muestras. Este razonamiento fue identificado por Tversky y Kahneman (1971) y encontrado en muchas otras investigaciones previas, como, por ejemplo, Cañizares (1997) y Serrano, Batanero, Ortiz y Cañizares (1998).
- *C4. Aleatoriedad como falta de patrón (A442),* en vez de considerarlo como multiplicidad de patrones, como sugieren Batanero (2016) y Serrano (1996).
- *C5. Suponer que la aleatoriedad indica máxima variabilidad.* Aparece en aquellas respuestas en las que el valor del rango de la muestra generada se considera excesivo. Además, podemos identificar que los estudiantes razonan indicando que al ser aleatorio se puede esperar cualquier resultado. Por tanto, estos estudiantes tienen una pobre comprensión de la relación de la variabilidad y el tamaño de la muestra. Así mismo, considerar como probable la ocurrencia de cualquier proporción en la obtención de una muestra implica que confunden el concepto de probabilidad y posibilidad.
- *C6. Asimilar la aleatoriedad a la falta de información.* Esta fue una concepción histórica ya descartada. Batanero (2016) indica que la concepción de aleatoriedad como sinónimo de ignorancia persistió en varios filósofos hasta comienzo del siglo XX, en que Poincaré (1987/1912) la rebatió. Este autor diferencia entre fenómenos aleatorios, que pueden estudiarse mediante las leyes de la probabilidad, aunque no se conozcan sus causas y fenómenos no aleatorios, en los que no es posible este estudio.
- *C7. No relacionar la probabilidad con la frecuencia.* En este caso el estudiante no comprende el concepto de probabilidad frecuencial ni la convergencia de la frecuencia relativa a la probabilidad teórica. Esto implica una concepción de la probabilidad muy pobre, ya que no se consideran los distintos significados de la misma y, por tanto, es

incapaz de identificar qué enfoque resuelve la tarea. Por ejemplo, el primer ítem y el último exigen que el alumno razone de acuerdo a la información frecuencial dada en el enunciado. Por lo que, identificamos alumnos que no consideran como válida la información frecuencial e intentan justificar el dato del enunciado aplicando suposiciones.

- *C8. Ilusión de control* (Langer, 1982; Yarritu, Matute y Vadillo, 2014). Los estudiantes argumentan utilizando un lenguaje que denota que el sujeto tiene la capacidad de control los resultados obtenidos. Siguiendo a los autores, el sujeto no distingue entre los juegos de habilidad y los de azar creyendo, en consecuencia, que se puede manipular el azar.
- *C9. Falta de percepción de la independencia de resultados en la repetición de un experimento aleatorio*. Los estudiantes señalan que esperan que los resultados obtenidos al repetir el experimento sean similares al obtenido en el enunciado. En este caso, observamos que la respuesta generada se caracteriza por dar valores en los que la proporción es similar al dato, por lo que estos estudiantes razonarían de acuerdo la heurística de la representatividad y, en particular, presentarían el sesgo denominado recencia positiva (Tversky et al. ,1982). Además, según Valdez (2016), sería necesaria la percepción de la independencia para adquirir un nivel avanzado de comprensión del muestreo.
- *C10. Creencias subjetivas sobre el generador aleatorio*. Se corresponde con aquellas justificaciones en las que el estudiante le atribuye características al generador que, según su opinión, influyen en los resultados obtenidos.
- *C11. Aplicación estricta de la proporcionalidad, en relación con el tamaño de la muestra, sin considerar la variabilidad del muestreo*. Estas respuestas se caracterizan porque el estudiante identifica la probabilidad en términos de porcentaje y aplica, de manera determinista, dicho porcentaje en el cálculo de las cuatro estimaciones generadas, lo cual conduce a muestras con ninguna variabilidad.
- *C12. Inconsistencia entre la probabilidad teórica asumida y la asignación frecuencial de los datos*. Se corresponde con aquellas respuestas en las que el sujeto no proporciona una muestra desde la probabilidad identificada. Es decir, el estudiante identifica que la probabilidad de que la chincheta caiga con la punta hacia arriba es mayor, pero la proporción de las cuatro muestras es equitativa.

- C13. *Interpretar una pregunta de probabilidad en forma no probabilística o enfoque en el resultado* (Konold. 1989). En este caso, el estudiante reinterpreta la pregunta formulada como la formulación de una predicción sobre si un determinado suceso se obtendrá o no. Como señalan Serrano et al. (1998), los estudiantes podrían no comprender el enfoque frecuencial de la probabilidad.
- C14. *Concepción causal de la aleatoriedad, asimilándola a la existencia de diferentes causas que provocan el fenómeno aleatorio*. Este caso se identifica en aquellas respuestas en las que el sujeto razona en base a causas que puedan justificar los efectos, es decir, el resultado obtenido al realizar un ensayo de un experimento aleatorio.

*H3.2. El análisis de las respuestas del instrumento de evaluación refleja un progreso gradual de comprensión de los contenidos evaluados en los estudiantes de bachillerato, respecto a los de educación secundaria obligatoria.*

La formación de los participantes, dado el curso en el que se encuentran, nos permite establecer un análisis comparativo que pretende responder nuestra premisa de que exista una mejora en los resultados obtenidos según el curso. Por tanto, se revisa esta hipótesis en función del estadístico analizado, debido a que en el Estudio 1 se observaba una diferencia en la dificultad que tiene la muestra participante en asignar una variabilidad adecuada, en comparación con la estimación del valor esperado.

Para estudiar las diferencias, se realiza un análisis de la varianza tanto sobre la media de las estimaciones de los estudiantes como sobre el rango, usando como variable independiente el curso escolar. Se obtienen valores estadísticamente significativos en la comparación de las frecuencias con que los estudiantes asignan valores en los intervalos normativos aceptables u otros (sesgados o no respuesta) que mejora con el curso en el nivel educativo, en el test de Chi cuadrado, en todas las comparaciones realizadas.

Respecto a la estimación de la media y en la estimación del rango en todos los ítems, la prueba de análisis de la varianza es estadísticamente significativa en la mayoría de los casos. Posteriormente, las pruebas de Tukey revelan que las diferencias siempre existen respecto a los estudiantes de 2º de la ESO y en algunas ocasiones respecto a los de 4º curso, siempre favoreciendo a los de bachillerato.

Todas estas diferencias sugieren una progresión del razonamiento sobre el muestro

hasta llegar al razonamiento distribucional de los estudiantes, desde el curso 2º de la ESO hasta llegar a 2º curso de bachillerato.

*H4. Se espera encontrar una variedad de argumentos para apoyar las respuestas de los estudiantes. En concreto, podemos identificar argumentos encontrados en investigaciones previas como la de Serrano (1996) relacionadas con la comprensión de la aleatoriedad. Por otra parte, esperamos analizar la concepción de la probabilidad presente en la respuesta del estudiante. En relación con esto último, consideramos que las argumentaciones pueden reflejar la creencia que el alumno presenta acerca de la aleatoriedad así como la identificación de sesgos aplicados en su razonamiento, los cuales se han presentado en el Capítulo 2.*

El análisis del contenido de las argumentaciones muestra que los estudiantes se apoyan en diferentes tipos de argumentaciones. Además, se observa que un alumno se puede apoyar en diferentes tipos de argumentaciones, por lo que eleva la complejidad y variedad del contenido de las mismas.

Por un lado, algunos estudiantes recurren a su visión de la aleatoriedad como proceso que no tiene patrón o que es impredecible, ideas que aparecieron en la investigación de Serrano (1996). Por otro lado, se observa que existe un conjunto de alumnos que confunden el concepto de aleatoriedad con probabilidad, lo cual dificulta elaborar su respuesta. En este sentido, los alumnos identifican que el fenómeno es aleatorio, por lo que cualquier muestra es igual de probable, por tanto, definen la probabilidad en términos de posibilidad, produciendo muestras cuya probabilidad de ocurrencia es menor. Además, se han mostrado sesgos como el enfoque en el resultado (Konold, 1989) o el sesgo de equiprobabilidad (Lecoutre, 1992). También, si observamos el porcentaje de los tipos de argumentos aplicados, se identifica una elevada presencia del Tipo 8, fundamentalmente en el ítem 4, el cual implica que los alumnos argumentan según a creencias subjetivas en relación al fenómeno presentado.

El análisis de las argumentaciones permite identificar la concepción que el alumno tiene sobre la probabilidad, que es con frecuencia correcta. La presencia del enfoque clásico para asignar la probabilidad de un suceso es adecuada si se cumple la condición de equiprobabilidad, pero su consideración, independientemente del fenómeno presentado, conduce a la presencia del sesgo de equiprobabilidad. Otros alumnos que

apoyan su argumentación en un enfoque frecuencial en el cual se considera el dato dado en el enunciado para estimar la proporción de cada suceso en la población, favorece la ausencia del anterior sesgo, ya que es válido para todo fenómeno aleatorio. Además, muchos estudiantes aluden a la convergencia y variabilidad del muestreo.

En resumen, en el Estudio 2 se han cumplido tanto los objetivos como confirmado las hipótesis planteadas. Retomaremos parte de estas conclusiones en las conclusiones finales de la tesis.

## **CAPÍTULO 5.**

### **CONCLUSIONES**

- 5.1. Introducción
- 5.2. Conclusiones respecto a los objetivos
- 5.3. Conclusiones respecto a las hipótesis
- 5.4. Principales aportaciones del trabajo
- 5.5. Limitaciones del trabajo
- 5.6. Líneas abiertas de investigación

#### **5.1. INTRODUCCIÓN**

En esta Memoria se ha resumido el trabajo realizado en torno a la comprensión de ideas elementales sobre el muestreo por parte de los estudiantes de educación secundaria obligatoria y bachillerato. El interés del tema se ha justificado por la importancia actual que la inferencia está alcanzando en muchas disciplinas, así como en la enseñanza universitaria. En concreto, nuestro estudio se centra en un concepto fundamental (Biehler y Burrill, 2011; Heitele, 1975), cuya adecuada comprensión resulta necesaria para entender otros conceptos y métodos de inferencia.

El trabajo se apoya en el enfoque ontosemiótico del conocimiento y la instrucción matemáticos (Godino, 2002; Godino y Batanero, 1994; Godino, Batanero y Font, 2007), el cual constituye el marco teórico que nos sirvió para plantear los objetivos e hipótesis de la investigación.

Basándonos en este marco, así como en las investigaciones resumidas en el Capítulo 2, se construyó un cuestionario para evaluar la comprensión de una serie de conceptos relacionados con el muestreo, que se pasó a dos muestras de estudiantes, una primera de estudiantes de educación secundaria obligatoria (Capítulo 3) y otra de estudiantes de bachillerato (Capítulo 4), para poder realizar un análisis comparativo de los resultados.

Para completar nuestro trabajo, en este último capítulo se realiza una reflexión sobre los objetivos y las hipótesis iniciales, en relación con los resultados obtenidos del análisis de los datos, según los estudios empíricos realizados. Así mismo, se analizan las limitaciones de este trabajo, y las posibles vías de investigación que quedan abiertas para continuar la misma temática de investigación abordada.

## 5.2. CONCLUSIONES RESPECTO A LOS OBJETIVOS

Como se ha reiterado en los anteriores capítulos, este trabajo tuvo como finalidad analizar y evaluar la comprensión que los estudiantes de la etapa obligatoria de secundaria y de bachillerato presentan sobre algunas ideas básicas asociadas al concepto de muestreo. A partir de este objetivo general, en el Capítulo 1 se definieron un conjunto de objetivos secundarios que enmarcan el estudio, que a su vez se desglosaron en los dos estudios empíricos realizados y que también se van a indicar. En las siguientes líneas se recoge un breve análisis del grado de consecución de los citados objetivos.

*Objetivo 1. Identificar, describir y comparar los contenidos matemáticos relacionados con el muestreo, presentados en los documentos curriculares a lo largo de la educación secundaria obligatoria y el bachillerato.*

Este objetivo se justificó por la necesidad de estar informados de la enseñanza que habían recibido los estudiantes que participaron en los estudios empíricos. Se trataba de identificar los contenidos matemáticos asociados al muestreo que son contemplados por los documentos curriculares en los diferentes cursos, así como comprender las orientaciones metodológicas consideradas en su enseñanza. Este análisis previo informaría sobre los contenidos que podríamos suponer que hubieran formado parte del proceso de enseñanza y aprendizaje de la etapa de secundaria y sería nuestro significado institucional pretendido para el muestreo.

Para cumplir con este objetivo, en el Capítulo 1 se describe un estudio detallado de los documentos curriculares actuales (MECD, 2015) y los utilizados en el anterior periodo (MEC, 2007), tanto para la educación secundaria obligatoria como para el bachillerato. Se recogen en una serie de tablas los contenidos pretendidos en cada curso y se compara con los documentos oficiales de la Comunidad de Aragón (Departamento de Educación, Cultura y Deporte, 2007). A partir de dichas tablas se realiza un análisis semiótico, identificando los objetos matemáticos considerados en nuestro marco teórico para cada uno de los cursos (situaciones-problemas, conceptos, procedimientos y propiedades). Se describen igualmente las características específicas que en las orientaciones metodológicas se sugieren para la enseñanza de la estadística.

Además, se realiza una comparación de los citados currículos con algunas

orientaciones curriculares internacionales para estos niveles educativos, concretamente con NCTM (2000) y CCSSI (2010) y el proyecto GAISE (Franklin et al., 2007). Dicho análisis comparativo se realiza desde el marco teórico citado, ya que se sigue la clasificación que el enfoque onto-semiótico realiza con respecto a los diferentes objetos matemáticos (situaciones problemas, lenguaje, conceptos, propiedades y procedimientos).

Puesto que no solamente hemos caracterizado los contenidos asociados al concepto de estudio en los documentos curriculares y otras orientaciones internacionales, sino que también se han clasificado de acuerdo al marco teórico escogido, consideramos suficientemente cumplido este objetivo al obtener una imagen detallada de los contenidos del tema y del significado institucional del muestreo en el currículo español.

*Objetivo 2. Realizar una síntesis de las investigaciones previas, que nos proporcione una visión de los problemas abordados en las mismas, así como ejemplos de ítems de evaluación y respuestas de estudiantes, los cuales permitan enmarcar nuestro estudio y poder comparar nuestros resultados con investigaciones previas.*

Un aspecto interesante en el comienzo de una investigación es revisar aquellos trabajos previos que se han interesado por analizar el objeto matemático u otros estrechamente relacionados con este. De hecho, conocer qué se ha trabajado sobre el mismo tema con anterioridad, favorece el interés de planear problemas de investigación complementarios que permitan aportar nueva información sobre el tema.

Se presenta el análisis planteado en el Capítulo 2, donde se resumen los resultados que se han obtenido en investigaciones previas en relación a los contenidos matemáticos que evaluamos. Este estudio nos permite identificar el tipo de tareas que han formado parte de las investigaciones y las respuestas típicas de los estudiantes a estas tareas. En este sentido, su consideración ha permitido reflexionar sobre la selección de las incluidas el instrumento de evaluación que aplicamos. Además, esta revisión nos permite valorar si nuestro trabajo puede aportar o complementar nuevos resultados y compararlos con los de las investigaciones previas.

En dicho capítulo se presenta una amplia síntesis de investigaciones centradas en heurísticas y sesgos en el razonamiento llevadas a cabo, principalmente, desde la psicología, donde el muestreo cobra gran relevancia en la toma de decisiones y



evaluación de situaciones de riesgo. En segundo lugar, el análisis se centra en la revisión bibliográfica de aquellas investigaciones enfocadas en la comprensión de los conceptos de muestra y de los diferentes tipos de muestreo aleatorio. Este análisis revela las concepciones erróneas que los individuos tienen sobre algunas propiedades de la muestra, en particular sobre los factores que determinan su variabilidad y sobre el sesgo en el muestreo.

Además, se analizan las investigaciones más relevantes llevadas a cabo con estudiantes, centradas en la comprensión del concepto de muestra en entornos virtuales que resultan al llevar a cabo un proceso de simulación, realizada con herramientas de software específico. Se completa con el estudio de las investigaciones sobre comprensión de la probabilidad, desde el punto de vista frecuencial y sobre concepciones de los estudiantes respecto a la aleatoriedad. Todas estas investigaciones serán posteriormente relacionadas con la nuestra al analizar las respuestas de los estudiantes en los estudios empíricos.

Además de estos dos objetivos, comunes a toda la investigación, planteamos un tercero que se desglosará en objetivos específicos en cada uno de los estudios 1 y 2 descritos en los siguientes capítulos. Este tercer objetivo es el siguiente:

*Objetivo 3. Llevar a cabo una serie de estudios de evaluación de la comprensión que muestran los estudiantes de educación secundaria obligatoria y bachillerato de conceptos elementales ligados al muestreo.*

Mediante estos estudios pretendíamos profundizar en la comprensión que tienen los estudiantes de estos cursos de algunas ideas elementales sobre muestreo. Asimismo se deseaba ver la progresión de esta comprensión con el curso escolar y el efecto de ciertas variables de tarea sobre la misma. Con todo ello pensábamos proporcionar información que pueda utilizarse por los profesores para la evaluación de posibles dificultades y el diseño de la enseñanza del muestreo. Más concretamente, este objetivo se desglosó en otros parciales, que lo concretaban en cada uno de los capítulos 3 y 4, y que pueden ser sintetizados para estas conclusiones en la forma siguiente:

*Objetivo 3.1 (y Objetivo 3.4). Evaluar y analizar la comprensión de los estudiantes de segundo y cuarto curso de la educación secundaria obligatoria (y de*

*bachillerato) de la relación entre el valor de la proporción en una población y la proporción muestral esperada en muestras tomadas de dicha población.*

En el Estudio 1 se llevó a cabo un estudio de evaluación con una muestra de 302 estudiantes de segundo y cuarto curso de la educación secundaria obligatoria y en el Estudio 2 se repitió la evaluación con de 234 estudiantes bachillerato. En los dos casos se recogieron respuestas escritas de los estudiantes a un cuestionario con cuatro tareas, en cada una de las cuales se les pidió proporcionar cuatro valores probables del número de éxito en una distribución binomial.

Para evaluar su comprensión de la relación entre la proporción en una población y la proporción muestral, se calculó la media de los cuatro valores proporcionados por cada estudiante. Se analizó la distribución de esta variable, comparándola con la que corresponde a la media de las muestras de cuatro valores en la distribución binomial, determinando el porcentaje de estudiantes cuyas medias se sitúan en intervalos normativos, aceptables y otros que sugieren la existencia de sesgos. Se comparó también la media teórica de esta distribución con la media observada.

*Objetivo 3.2 (y Objetivo 3.5). Evaluar y analizar la comprensión de los estudiantes de segundo y cuarto curso de la educación secundaria obligatoria (y de bachillerato) de la variabilidad de la proporción muestral y del efecto del tamaño de la muestra sobre dicha variabilidad.*

Para evaluar esta comprensión se calculó el rango de los cuatro valores proporcionados por cada estudiante. Se analizó la distribución de esta variable, comparándola con la que corresponde al rango de las muestras de cuatro valores en la distribución binomial, determinando el porcentaje de estudiantes cuyas medias se sitúan en intervalos normativos, aceptables y otros que sugieren la existencia de sesgos. Se comparó también la media teórica de esta distribución con la media observada.

*Objetivo 3.3. Comparar los resultados en función de algunas variables de tarea y por curso.*

Puesto que las tareas propuestas se diseñaron para variar sistemáticamente la proporción de la población y el número de ensayos, se pudo comparar los resultados obtenidos en experimentos con sucesos equiprobables y no equiprobables, así como en

muestras grandes y pequeñas.

Por otro lado, en el Estudio 1 se compararon en cada tarea los resultados correspondientes al valor medio y a la variabilidad en los dos cursos participantes, gráficamente y mediante contrastes estadísticos, habiéndose informado de las diferencias significativas y no significativas. Igualmente, en el Estudio 2 se realiza un análisis de varianza para analizar las diferencias en los tres grupos participantes, que se complementa con comparaciones múltiples y contrastes Chi-cuadrado.

*Objetivo 3.6. Analizar los argumentos que los estudiantes emplean para apoyar la elección de sus respuestas.*

El Estudio 2 finaliza con el análisis de los argumentos de 127 estudiantes de bachillerato, usando análisis de contenido (Hernández et al., 2010; Zapico, 2007).

Las categorías encontradas en el análisis se han estudiado en profundidad, mostrando ejemplos de cada una de ellas e identificando los conocimientos implícitamente utilizados o los sesgos subyacentes a las mismas. Se completa el análisis con una identificación de conflictos semióticos latentes en las respuestas de los estudiantes.

Pensamos que este tercer objetivo y los objetivos parciales en que se dividen dentro de los capítulos anteriores se han cumplido satisfactoriamente. La amplitud de las muestras analizadas y la profundidad del análisis realizado nos han permitido proporcionar una amplia información al respecto.

### **5.3. CONCLUSIONES RESPECTO A LAS HIPÓTESIS INICIALES**

Del interés inicial por analizar la comprensión que un grupo de estudiantes presenta sobre diferentes ideas asociadas al concepto de muestreo, surgieron una serie de hipótesis que se describieron con detalle en el Capítulo 1 y que ahora pasamos a analizar.

*H1. Los contenidos curriculares tanto del Decreto de Enseñanzas Mínima como del Currículo Básico contemplan o reflejan los aspectos esenciales asociados al concepto de muestreo que son objeto de investigación.*

En particular, se esperaba que dichos documentos curriculares introdujesen ideas

sobre población y muestra, estadístico y parámetro, tipos de muestreo y nociones elementales de estimación, representatividad y variabilidad muestral. Igualmente, se esperaba que se considerasen los tres niveles de distribución citados por Schuyten (1991): distribución de datos, distribución de la variable en la población y distribución muestral del estadístico. Nos apoyamos para elaborar esta hipótesis en nuestro propio conocimiento del currículo y nuestra experiencia de enseñanza.

El estudio curricular presentado en el Capítulo 1 ha mostrado una amplia variedad de problemas, conceptos, propiedades, procedimiento, lenguaje y argumentos asociados al muestreo a lo largo de la educación secundaria y bachillerato. Tanto en el Decreto de Enseñanzas Mínimas como en el Currículo Básico, se incluyen las ideas de población y muestra desde el comienzo de la secundaria y el estudio de los tipos de muestreo, representatividad de muestras y distribuciones muestrales en el segundo curso de bachillerato en la modalidad de ciencias sociales. Las distribuciones de datos se estudian desde la educación secundaria, mientras que las distribuciones de variables aleatorias (población) se incluyen en los dos bachilleratos, en primer curso. La distribución muestral y sus propiedades sólo se estudia en el segundo curso de bachillerato de la modalidad de ciencias sociales.

En consecuencia, esta hipótesis se confirma parcialmente, en el sentido de que todos los contenidos previstos se incluyen para los estudiantes de ciencias sociales en el bachillerato, pero no en el bachillerato de ciencias, y algunos contenidos no se incluyen en la educación secundaria obligatoria.

*H2. Algunos estudiantes de educación secundaria y bachillerato presentan una comprensión insuficiente sobre algunos objetos matemáticos elementales ligados al muestreo.*

Esta hipótesis se basó en el análisis de los antecedentes de la investigación, que se describen en el Capítulo 2, donde se han analizado diferentes sesgos de comprensión referidos a la representatividad y variabilidad muestral, confusión entre estadístico y parámetro y entre la distribución de la población y la distribución muestral. Igualmente se han descrito en las investigaciones previas dificultades con las ideas de independencia de ensayos sucesivos y la comprensión de la convergencia, que pensamos se repetiría en nuestros estudios empíricos.

En los capítulos 3 y 4 estas hipótesis se transformaron en las hipótesis H2.1 y H2.2 que tienen la misma formulación anterior, pero se refieren únicamente a estudiantes de secundaria, en el primer caso, y de bachillerato, en el segundo, y se discuten con detalle en dichos capítulos.

En resumen, los estudiantes, en su mayoría, mostraron una buena percepción de la proximidad de la proporción poblacional y muestral, por lo que alcanzan al menos el nivel proporcional de razonamiento sobre muestreo en la categorización de Shaughnessy et al. (2004) y un nivel intermedio en la de Valdéz (2016). Sin embargo, aparece en los ítems 1 y 4 algunos estudiantes con sesgo de equiprobabilidad (Lecoutre, 1992) o la heurística de representatividad.

Por otro lado, la percepción de la variabilidad es pobre y mejora en las pequeñas muestras, resultado contrario al encontrado por Shaughnessy et al. (2004), seguramente por el tipo de tarea considerado por estos autores. Finalmente, una proporción importante de estudiantes no llega a alcanzar los niveles superiores de razonamiento sobre el muestreo descritos por Moreno y Vallecillos (2001) puesto que no aprecian el efecto del tamaño de la muestra sobre la variabilidad muestral. Los resultados muestran también la relación en las respuestas a los diferentes ítems, indicando sesgos subyacentes en los estudiantes que no dependen del ítem.

### *H3. Esperamos una mejora gradual de la comprensión en los estudiantes de mayor edad.*

Puesto que hay una diferencia de dos o cuatro años en los grupos de estudiantes que formarán parte de los trabajos empíricos, se espera un mejor razonamiento en los cursos superiores. Además, puesto que el currículo contempla la enseñanza de ideas básicas sobre muestreo, se espera que los estudiantes mayores hayan tenido oportunidad de estudiar el tema y por tanto muestren una mejora en su comprensión. Esta hipótesis se concreta en cada uno de los Estudios 1 y 2 convirtiéndose en las hipótesis H3.1 y H3.2 respectivamente, y se discutieron con detalle en los correspondientes capítulos.

Como resumen podemos decir que la hipótesis se ha cumplido, en general, con algunas particularidades. Así, el porcentaje de estudiantes con sesgos de equiprobabilidad, recencia positiva o negativa, disminuye con el curso escolar. Por el contrario, aumenta la proporción de los que dan respuestas normativas o aceptables,

aunque sigue siendo todavía alta la de los estudiantes que dan estimaciones de variabilidad excesiva. Aunque las diferencias entre los estudiantes de segundo y cuarto curso de la educación secundaria obligatoria no es siempre estadísticamente significativa, sí lo es el análisis de varianza, cuando se incluyen los estudiantes de bachillerato. Estos muestran mejor razonamiento que sus compañeros y por tanto observamos una progresión del razonamiento sobre muestreo al avanzar el curso escolar.

Puesto que en el Estudio 2 se incluyó el análisis de los argumentos de una parte de la muestra, se incluyó la siguiente hipótesis sobre los posibles resultados de este análisis:

*H4. Se espera encontrar una variedad de argumentos para apoyar las respuestas de los estudiantes. En concreto, podemos identificar argumentos encontrados en investigaciones previas como la de Serrano (1996) relacionadas con la comprensión de la aleatoriedad. Por otra parte, esperamos analizar la concepción de la probabilidad presente en la respuesta del estudiante. En relación con esto último, consideramos que las argumentaciones pueden reflejar la creencia que el alumno presenta acerca de la aleatoriedad, así como la identificación de sesgos aplicados en su razonamiento, los cuales se han presentado en el Capítulo 2.*

Esta hipótesis se vio confirmada, como muestran las conclusiones del Capítulo 4. Como resumen, algunos estudiantes recurren a su visión de la aleatoriedad como proceso que no tiene patrón o que es impredecible, o bien confunden aleatoriedad con equiprobabilidad (Lecoutre, 1992), ideas que aparecieron en la investigación de Serrano (1996).

El análisis de las argumentaciones permitió identificar concepciones correctas sobre la probabilidad en parte de los estudiantes y su discriminación de uso, según la tarea. Además, muchos estudiantes aluden a la convergencia y variabilidad del muestreo.

#### **5.4. PRINCIPALES APORTACIONES DEL TRABAJO**

Finalizada la discusión de los objetivos e hipótesis del trabajo, conviene resaltar las principales contribuciones del mismo a la investigación en didáctica de la estadística, a lo que se dedica esta sección.

Como nuestro trabajo se fundamenta en la aplicación de un cuestionario que tiene como finalidad evaluar un objeto matemático, podemos indicar que este estudio contribuye en identificar el nivel de comprensión de los estudiantes de educación secundaria y bachillerato respecto a las ideas elementales relacionadas con el muestreo, que se han resaltado a lo largo de la Memoria. A pesar de que el primer ítem es tomado otros trabajos previos, lo hemos completado y modificado, de modo que podemos evaluar otros aspectos necesarios en nuestra investigación que no eran contemplados en dichas investigaciones anteriores. En este sentido, se parte de un ítem que aparece en la investigación de Green (1983a) y Cañizares (1997), el cual es adaptado por Gómez (2014). Como nuestro interés es analizar la comprensión del proceso de muestreo, nuestra contribución consiste en elaborar variantes de este último para poder comparar el desempeño de los estudiantes en muestras grandes y pequeñas y en experimentos con resultados equiprobables y no equiprobables. Además, en parte de la muestra de estudiantes de bachillerato, se completan los ítems pidiendo la justificación de la respuesta, lo que permite profundizar en el estudio de evaluación. Por tanto, podemos pensar que hemos realizado una aportación original fundamentada, en primer lugar, en la elaboración de un cuestionario con ítems nuevos y, en segundo lugar, la demanda de la justificación de la respuesta numérica dada, que aporta una visión más completa al estudio.

Otra aportación es la metodología de análisis de las respuestas de los estudiantes, seguida en nuestro trabajo, cuya parte cuantitativa se basa en el cálculo de la media y el rango de los cuatro valores dados por cada estudiante en cada ítem y el estudio de su distribución. En el caso de Gómez (2014), que fue la única que pidió repetir cuatro veces la estimación en nuestro ítem 1, la autora también estudia la distribución de las medias y rangos de la cuatro respuestas, que compara con los valores de la media y rango más probables en una distribución binomial (la distribución de partida en la tarea).

En nuestro caso, hemos hecho la observación de que, los cuatro valores proporcionados por cada estudiante en la tarea constituyen una muestra de tamaño 4 de la distribución binomial de partida. Por consiguiente, se ha considerado que la comparación correcta de las respuestas debe hacerse con las distribuciones muestrales correspondientes para la media y el rango de muestras de cuatro valores. La primera de

dichas distribuciones la hemos obtenido teóricamente, y la segunda mediante simulación. Nuestro análisis es mucho más detallado que el de la citada autora (Gómez, 2014), al analizar diferentes gráficas y estadísticos e incluir algunos contrastes estadísticos.

Es original, asimismo, el análisis cualitativo de los argumentos de los estudiantes de bachillerato, y las categorías identificadas en el mismo mediante el análisis de contenido. Igualmente aportamos la interpretación de estas categorías en relación con la posible existencia de sesgos o de concepciones correctas sobre muestreo y probabilidad de los estudiantes. Así mismo, es original la lista de conflictos semióticos identificados.

Finalmente, otro aporte es mostrar la mejora del razonamiento sobre muestreo con la edad, que se deduce de los resultados estadísticamente significativos de los contrastes de análisis de varianza, Chi-cuadrado y comparación de grupos, realizados en los capítulos 3 y 4, así como del análisis de las puntuaciones totales y parciales en la prueba. Además, se ha mostrado la correlación entre medias y rangos en los cuatro ítems, que indica tendencia de los mismos estudiantes a sobre estimar o estimar a la baja dichos estadísticos.

Otras aportaciones de nuestro trabajo son el estado de la cuestión presentado en el Capítulo 2, así como el análisis curricular detallado en el Capítulo 1, que pueden ser útiles a otros investigadores interesados en el muestreo e incluso a los profesores que deben enseñar el tema. Todos estos resultados se complementan con una serie de publicaciones aportadas, que incluyen artículos en JCR o Scopus, otras revistas y congresos nacionales e internacionales, que se listan en esta memoria (ver Anexo).

## **5.5. LIMITACIONES DEL TRABAJO**

Seguimos con el análisis de las limitaciones que se identifican en este trabajo y que se explican por la naturaleza del mismo. En este sentido, como se ha indicado, el trabajo de investigación es un estudio exploratorio, cuya finalidad es identificar la comprensión que los alumnos presentan acerca de diversos conceptos asociados al estudio del muestreo. Aunque en las investigaciones previas se han descrito otros trabajos relacionados, el cuestionario empleado es diferente de los utilizados por estos autores y también los estudiantes varían en edad y contexto.

Este carácter exploratorio nos llevó a trabajar con muestras intencionales, aunque



bastante amplias, pero las posibilidades de generalización son siempre menores que las de muestras aleatorias. Sería por tanto conveniente, continuar la investigación con nuevas muestras, para comprobar que se siguen conservando nuestras principales conclusiones.

Por otro lado, únicamente nos hemos limitado a un tipo de ítem, en que se dan datos sobre una distribución y se piden valores (muestras de cuatro elementos de la misma). Sería conveniente contrastar nuestros resultados con otro tipo de tarea inversa, en que se proporcionen muestras y se pide estimar las características de la población.

Además, el cuestionario se ha pasado sin una instrucción planeada previa. Sabemos que los estudiantes de segundo curso de ESO no habían recibido instrucción el año que se pasó el cuestionario sobre los contenidos referentes al bloque de probabilidad y estadística, aunque es plausible que en primer curso y en sexto curso de la educación primaria aprendieran algunas nociones, como lo requiere el currículo. Los estudiantes de cuarto curso y de bachillerato han recibido una mayor instrucción los años anteriores, en algunos casos incluso sobre probabilidad frecuencial. Pero ha sido difícil controlar el factor de enseñanza recibida, pues muchos estudiantes mostraban su falta de conocimiento sobre los mismos, a pesar de haberlos trabajado previamente.

Sería también interesante haber pedido a los estudiantes de educación secundaria obligatoria argumentar sus respuestas para poder comparar con las de sus compañeros.

## 5.6. FUTURAS LÍNEAS DE INVESTIGACIÓN

Nuestros resultados revelan una falta de comprensión adecuada sobre el concepto de muestreo, siendo más acusada para el caso de la variabilidad muestral. Tanto este resultado como las limitaciones citadas en el apartado anterior, sugieren la necesidad de plantear en el futuro una investigación más exhaustiva.

En relación con aspectos de la propia investigación, el análisis de los ítems aplicado pone en relieve ciertas limitaciones. En este sentido, sería interesante analizar la tarea inversa, que implica incluir problemas de estimación, donde al alumno se le presenten una serie de muestras. De este modo, se considera la realización de una posterior investigación, en la cual, evaluar las inferencias que realizan una muestra de estudiantes de las características de una población a partir de las muestras. Por otro lado, la identificación de una serie de sesgos conduce a la necesidad de reflexionar sobre

plantear preguntas relacionadas directamente con los sesgos detectados en esta investigación.

La importancia que adquieren la comprensión de las ideas de muestreo y distribución muestral para adquirir una adecuada comprensión de la estadística (Ben-Zvi, et al., 2015), justifica la necesidad de llevar también a cabo investigaciones que favorezcan el aprendizaje desde los primeros cursos de la secundaria, con la finalidad de no dejar este contenido a los últimos cursos del bachillerato, y principalmente, a la etapa universitaria. En este sentido, Watson y Moritz (2000) señalan que la investigación sobre el desarrollo cognitivo del concepto de muestreo tiene una presencia escasa en la literatura.

Por tanto, se nos plantea la necesidad de una investigación en la que se diseñara un proceso de instrucción para los cursos considerados en este trabajo, para introducir las ideas elementales de muestreo y analizar si se produce una mejora en la comprensión de las mismas tras la enseñanza.

Adicionalmente, se podría realizar una revisión de los libros de texto y analizar el modo en que se introduce los conceptos asociados con el bloque de contenidos indicados, así como identificar si ha habido un cambio en los últimos años. Esta última reflexión viene motivada por las conclusiones obtenidas desde el análisis curricular, puesto que se observa un cambio en la introducción de ciertos contenidos. En este sentido, se introduce el significado frecuencial de la probabilidad los últimos cursos de la educación primaria y primeros cursos de la educación secundaria. Además, las orientaciones metodológicas remarcan la necesidad de que se introduzca la simulación y experimentación en el aula.

Por último, sería necesario considerar la necesaria preparación de los profesores encargados de la enseñanza. Aunque muy posiblemente sus conocimientos matemáticos sobre el tema serán suficientes, no hay una enseñanza establecida sobre didáctica de la estadística en la formación de profesores. Será entonces plausible que estos profesores requieran reforzar sus conocimientos didácticos sobre el muestreo, por ejemplo, sobre las posibles dificultades y sesgos de razonamiento de sus estudiantes, sobre tareas de enseñanza y evaluación y sobre recursos instruccionales y tecnológicos. Aquí se abre una nueva línea de investigación sobre evaluación y desarrollo de estos conocimientos en futuros profesores.

## REFERENCIAS

- Allan, L. G. y Jenkins, H. M. (1980). The judgment of contingency and the nature of the response alternatives. *Canadian Journal of Psychology*, 34, 1-11.
- Alvarado, H. (2007). *Significados institucionales y personales del teorema central del límite en la enseñanza de estadística en ingeniería*. Tesis Doctoral. Universidad de Granada.
- Alvarado, H., Galindo, M. y Retamal, L. (2013). Comprensión de la distribución muestral mediante configuraciones didácticas y su implicación en la inferencia estadística. *Enseñanza de las Ciencias*, 31(2), 75-91. <http://dx.doi.org/10.5565/rev/ec/v31n2.803>.
- Arredondo, S. C. y Bolívar, A. (2002). *Compromisos de la evaluación educativa*. Madrid: Pearson Educación.
- Arteaga, P., Batanero, C., Cañadas, G. y Contreras, J. M. (2011). Las tablas y gráficos estadísticos como objetos culturales, *Números* 76, 55-67.
- Bakker, A. (2004). *Design research in statistics education: On symbolizing and computer tools*. Utrecht: Beta Press.
- Bakker, A. y Gravemeijer, K. P. (2004). Learning to reason about distribution. En J. Garfield y D. Ben-Zvi (Eds.). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 147-168). Dordrecht: Springer.
- Bakker, A., Hahn, C., Kazak, S. y Pratt, D. (2018). Research on probability and statistics education in ERME: Trends and directions En T. Dreyfus, M. Artigue, D. Potari, S. Predigery K. Ruthven (Eds.) *Developing research in mathematics education - twenty years of communication, cooperation and collaboration in Europe*. Oxon: Routledge.
- Ball, D. L., Lubienski, S. T. y Mewborn, D. S. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. En V. Richardson (Ed.), *Handbook of research on teaching* (pp. 433-456). Washington, DC: American Educational Research Association.

- Barbero, M. I. (1993). *Psicometría II. Métodos de elaboración de escalas*. Madrid: UNED.
- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
- Batanero, C. (2004a). Los retos de la cultura estadística. *Yupana*, 1(1), 27-37.
- Batanero, C. (2004b). Ideas estocásticas fundamentales. ¿Qué contenidos se debe enseñar en la clase de probabilidad? En J. A. Fernández (Ed.), *Actas do I Encontro de Probabilidades e Estatística na Escola* (pp. 9-30). Braga: Portugal: Universidad de Minho.
- Batanero, C. (2013). Del análisis de datos a la inferencia: Reflexiones sobre la formación del razonamiento estadístico. *Cuadernos de Investigación y Formación en Educación Matemática*, 8(11), 277-291.
- Batanero, C. (2016). Understanding randomness. Challenges for research and teaching, *Proceedings of the Ninth Congress of European Research in Mathematics Education, CERME 9* (pp. 34-49). Praga: ERME.
- Batanero, C. (2018). Treinta años de investigación didáctica sobre el análisis inferencial de datos. En A. Avila (Ed.), *Rutas de la Educación Matemática* (pp. 196-209). México: Sociedad Mexicana de Investigación y Divulgación de la Educación Matemática
- Batanero, C. (2019). Thirty years of stochastics education research: Reflections and challenges. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html)
- Batanero, C., Begué, N. y Gea, M. M. (2018). ¿Cómo desarrollar el sentido del muestreo en los estudiantes?, *3º Encuentro Colombiano de Educación Estocástica*. Poyapan, Colombia, Octubre, 2018.
- Batanero, C., Begué, N. y Gea, M.M. (2019). El muestreo en el bachillerato español En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html).
- Batanero, C. y Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publishers.
- Batanero, C., Chernoff, E., Engel, J., Lee, H. y Sánchez, E. (2016). *Research on teaching and learning probability*. ICME-13. Topical Survey series. New York: Springer.

- Batanero, C. y Díaz, C. (2007). Meaning and understanding of mathematics. The case of probability. En J.P Van Bendegem y K. François (Eds), *Philosophical Dimensions in Mathematics Education* (pp. 107-128). New York: Springer.
- Batanero, C. y Díaz, C. (2015). Aproximación informal al contraste de hipótesis. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2, 135-144.
- Batanero, C., Díaz, C., Contreras, J. M. y Roa, R. (2013). El sentido estadístico y su desarrollo. *Números* 83, 7-18.
- Batanero, C., Gea, M., Arteaga, P. y Contreras, J.M. (2014). La estadística en la educación obligatoria: Análisis del currículo español. *Revista Digital Matemática, Educación e Internet* 14(2). Disponible en: <http://revistas.tec.ac.cr/index.php/matematica/article/view/1663>
- Batanero, C. y Godino, J. (2005). Perspectivas de la educación estadística como área de investigación. En L. Blanco (Ed.), *Líneas de investigación en didáctica de las matemáticas* (pp. 203-226). Badajoz: Universidad de Extremadura.
- Batanero, C., Godino, J. D., Green, D. R., Holmes, P. y Vallecillos, A. (1994). Errores y dificultades en la comprensión de los conceptos estadísticos elementales. *International Journal of Mathematics Education in Science and Technology*, 25(4), 527-547.
- Batanero, C., Gómez, E., Gea, M. M y Contreras, J. M. (2014). Assessing and developing prospective teachers' understanding of random sequences. En U. Sproesser, S. Wessolowski y C. Wörn (Eds.), *Daten, Zufall und der Rest der Welt – Didaktische Perspektiven zur anwendungsbezogenen Mathematik*. Heidelberg: Springer Spektrum Verlag.
- Batanero, C., Gómez, E., Serrano, L. y Contreras, J. M. (2012). Comprensión de la aleatoriedad por futuros profesores de educación primaria. *Redimat* 1(3), 222-245.
- Batanero, C. y Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education*, 30(5), 558-567.
- Batanero, C., Serrano, L., y Garfield, J. B. (1996). Heuristics and biases in secondary school students' reasoning about probability. En A. Gutiérrez (Ed.). *Proceedings of the 18th PME Conference* (Vol. 2, pp. 2-51). Valencia: PME.
- Begué, N. (2016). *Comprensión de elementos básicos de muestreo en estudiantes de educación secundaria obligatoria*. Tesis de Máster. Universidad de Granada.

- Begué, N. (2019). El muestreo en el bachillerato español. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html).
- Begué, N. y Batanero, C. (2017). Significado del muestreo en el currículo de Educación Secundaria Obligatoria. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín(Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos*. Disponible en, [enfoqueontosemiotico.ugr.es/civeos.html](http://enfoqueontosemiotico.ugr.es/civeos.html).
- Begué, N., Batanero, C. y Gea, M.M.. (2018). Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria. *Enseñanza de las Ciencias*, 36(2), 63-79.
- Begué, N., Batanero, C. y Gea, M.M. (2019). Argumentos de los estudiantes de bachillerato en la generación de muestras de la distribución binomial. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html)
- Begué, N., Batanero, C., Gea, M.M. y Beltrán-Pellicer, P. (2017). Comprensión del enfoque frecuencial de la probabilidad por estudiantes de Educación Secundaria Obligatoria. En J.M. Muñoz-Escolano, A. Arnal-Bailera, P. Beltrán-Pellicer, M.L. Callejo y J. Carrillo (Eds.), *Investigación en Educación Matemática XXI* (pp. 137-146). Zaragoza: SEIEM.
- Begué, N., Batanero, C., Ruiz, K. y Gea, M.M. (2017). Understanding sampling: A summary of research. *BEIO*, 35(1), 49-78.
- Begué, N., Gea, M.M. Batanero, C., Beltrán, P. (2017). Meaning of sampling for secondary school students. En L. Gómez Chova, A. López Martínez, I. Candel Torres (Eds.), *10th annual International Conference of Education, Research and Innovation*. International Academy of Technology, Education and Development (pp. 1430-1435). Sevilla: IATED Academy.
- Begué, N., Gea, M.M. Batanero, C., Beltrán, P. (2018). Do high school students understand the sampling distribution of proportion? Trabajo invitado. En M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018)*, Kyoto, Japan: International Statistical Institute.

- Begué, N., Ruiz, K., Gea, M.M. y Batanero C. (2017). *El muestreo en el currículo de secundaria: un estudio comparado de los currículos en España y Chile. Libro de Actas del VIII Congreso Iberoamericano de Educación Matemática*. Andújar (Jaén) España: Federación Española de Sociedades de Profesores de Matemáticas.
- Ben-Zvi, D., Bakker, A. y Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, 88(3), 291-303.
- Ben-Zvi D., Makar K. y Garfield J. (2018) (Eds.), *International handbook of research in statistics education*. Nueva York: Springer.
- Biggs, J. B. y Collis, K. F. (1982). *Evaluating the quality of learning: The Solo taxonomy*. New York: Academic.
- Birks, M. y Mills, J. (2011). *Grounded theory: A practical approach*. Thousand Oaks: Sage.
- Bisquerra, R. (1989). *Métodos de investigación educativa*. Barcelona: P.P.U.
- Briand, J. (2005). Une expérience statistique et une première approche des lois du hasard au lycée par une confrontation avec une machine simple. *Recherches en Didactique des Mathématiques*, 25(2), 247–281.
- Burrill, G. y Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education – A joint ICMI/IASE study* (pp. 57-69). Dordrecht: Springer.
- Campbell, D. T. y Stanley, J. C. (2015). *Experimental and quasi-experimental designs for research*. London: Houghton Mifflin,
- Cañizares, M. J. (1997). *Influencia del razonamiento proporcional y combinatorio y de creencias subjetivas en las intuiciones probabilísticas primarias*. Tesis doctoral. Universidad de Granada.
- Cañizares, M. J. y Batanero, C. (1997). Influencia del razonamiento proporcional y de las creencias subjetivas en la comparación de probabilidades. *Uno*, 14, 99-114.
- Castro Sotos, A. E., Vanhoof, S., Noortgate, W. y Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98-113.
- Cazau, P. (2006). *Introducción a la investigación en ciencias sociales*. Buenos Aires: Red de Psicología Disponible en: [www.galeon.com/pcazau](http://www.galeon.com/pcazau).
- CCSSI. Common Core Standards Initiative. (2010). Standards for mathematical practice. Disponible en: <http://www.corestandards.org/>

- Chance, B., delMas, R. C. y Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Amsterdam: Kluwer.
- Chernoff, E. J. (2009). Sample space partitions: An investigative lens. *The Journal of Mathematical Behavior*, 28(1), 19-29.
- Chernoff, E. J. y Russell, G. L. (2012). The fallacy of composition: Prospective mathematics teachers' use of logical fallacies. *Canadian Journal for Science, Mathematics and Technology Education*, 12(3), 259-271. Disponible en: <http://dx.doi.org/10.1080/14926156.2012.704128>.
- Chernoff, E. J. y Sriraman B. (2014) (Eds.), *Probabilistic thinking: presenting plural perspectives*. New York: Springer.
- Clements, D. H., y Sarama, J. (2009). Learning trajectories in early mathematics—sequences of acquisition and teaching. *Numeracy*, 21, 21-30.
- 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). Disponible en: [www.amstat.org/publications/jse](http://www.amstat.org/publications/jse).
- delMas, R., Garfield, J., y Chance, B. (2004). Using assessment to study the development of students' reasoning about sampling distributions. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA: AERA. Disponible en: [www.gen.umn.edu/faculty\\_staff/delmas/AERA\\_2004\\_samp\\_dist.pdf](http://www.gen.umn.edu/faculty_staff/delmas/AERA_2004_samp_dist.pdf).
- Departamento de Educación, Cultura y Deporte (2007). *Orden de 9 de mayo de 2007, por el que se desarrolla el currículo de la Educación Secundaria Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad de Aragón*. Aragón: Departamento de Educación, Cultura y Deporte.
- Departamento de Educación, Cultura y Deporte (2008). *Orden de 1 de julio de 2008, del Departamento de Educación, Cultura y Deporte, por la que se aprueba el currículo del Bachillerato y se autoriza su aplicación en los centros docentes de la Comunidad autónoma de Aragón*. Aragón: Departamento de Educación, Cultura y Deporte.
- Departamento de Educación, Cultura y Deporte (2016a). *Orden ECD/489/2016, de 26 de mayo, por la que se aprueba el currículo de la Educación Secundaria*



- Obligatoria y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón.* Aragón: Departamento de Educación, Cultura y Deporte.
- Departamento de Educación, Cultura y Deporte (2016b). *Orden ECD/494/2016, de 26 de mayo, por la que se aprueba el currículo del Bachillerato y se autoriza su aplicación en los centros docentes de la Comunidad Autónoma de Aragón.* Aragón: Departamento de Educación, Cultura y Deporte.
- Díaz, C. (2003). Heurísticas y sesgos en el razonamiento probabilístico. Implicaciones para la enseñanza de la estadística. *Actas del 27 Congreso Nacional de Estadística e Investigación Operativa*. Lérida: Sociedad de Estadística e Investigación Operativa. [CD- ROM].
- Eichler, A., y Vogel, M. (2014). Three approaches for modelling situations with randomness. En E. J. Chernoff y B. Sriraman (Eds.), *Probabilistic thinking, presenting plural perspectives* (pp. 75-99). Dordrecht, Holanda: Springer.
- Engel, J. (2019). Statistical literacy and society. What is civic statistics? En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística* (pp. 1-17). Granada: Grupo de Investigación de Educación Estadística. Disponible en: <https://www.ugr.es/~fqm126/civeest/ponencias/engel.pdf>,
- Escorcía, O. (2010). *Manual para la investigación: guía para la formulación, desarrollo y divulgación de proyectos*. Bogotá: Universidad Nacional de Colombia.
- Falk, R. y Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, 104, 301-318.
- Fischbein, E., Sainati Nello, M. y Sciolis Marino, M. (1991). Factors affecting probabilistics judgements in children and adolescents. *Educational Studies in Mathematics*, 22, 523–549.
- Font, V., Godino, J.D. y D'Amore, B. (2007). An ontosemiotic approach to representations in mathematics education. *For the Learning of Mathematics*, 27(2), 3-9.
- Franklin, C. Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., y Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association. Disponible en: [www.amstat.org/education/gaise/](http://www.amstat.org/education/gaise/).

- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25.
- García-Ríos, V. N. (2013). Inferencias estadísticas informales en estudiantes mexicanos. *Probabilidad Condicionada* 2, 343-357.
- García-Ríos, V. N. y Sánchez, E. (2013). Dificultades en el razonamiento inferencial intuitivo. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2 (pp. 207-214). Granada: SEIEM.
- García-Ríos, V. N. y Sánchez, E. A. (2014). Razonamiento inferencial informal: el caso de la prueba de significación con estudiantes de bachillerato. En M. T. González, M. Codes, D. Arnau y T. Ortega (Eds.), *Investigación en Educación Matemática XVIII* (pp. 345-354). Salamanca: SEIEM.
- Garfield, J., y Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York: Springer.
- Garfield, J. y Gal, I. (1999), Teaching and assessing statistical reasoning. En L. Stiff (Ed.), *Developing mathematical reasoning in grades K-12* (pp. 207-219). Reston, VA: National Council Teachers of Mathematics.
- Garfield, J., delMas, R. y Zieffler, A. (2010). Assessing statistical thinking. En P. Bidgood, N. Hunt y F. Jolliffe (Eds.), *Assessment methods in statistical education: An international perspective* (pp. 175–186). Milton: John Wiley.
- Godino, J. D. (1996) Mathematical concepts, their meaning, and understanding. En L. Puig y A. Gutierrez (Eds.), *Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education* (v.2, pp. 417-425). Universidad de Valencia.
- Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22 (2-3), 237-284.
- Godino, J. D. (2017). Construyendo un sistema modular e inclusivo de herramientas teóricas para la educación matemática. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso Internacional Virtual sobre el Enfoque Ontosemiótico. del Conocimiento y la Instrucción Matemáticos*. Granada: Grupo FQM126.
- Godino, J. D., Aké, L. P., Gonzato, M. y Wilhelmi, M. R. (2014). Niveles de algebrización de la actividad matemática escolar. Implicaciones para la formación de maestros. *Enseñanza de las Ciencias*, 32(1), 199-219.

- Godino, J. D. y Batanero, C (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, 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. Sierpiska y J. Kilpatrick (Eds.), *Mathematics education a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.
- Godino, J.D., Batanero, M.C. y Cañizares, M.J. (1987). *Azar y probabilidad. Fundamentos didácticos y propuestas curriculares*. Madrid: Síntesis.
- 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. y Font, V. (2019). The onto-semiotic approach: Implications for the prescriptive character of didactics. *For the Learning of Mathematics*, 39(1), 38-43.
- Godino, J.D., Neto, T., Wilhelmi, M.R., Aké, L.P., Etchegaray, S. y Lasa, A. (2015). Niveles de algebrización de las prácticas matemáticas escolares. Articulación de las perspectivas ontosemiótica y antropológica. *Avances de Investigación en Educación Matemática*, 8, 117-142.
- Godino, J. D., Rivas, M., Castro, W. F., y Konic, P. (2008). Desarrollo de competencias para el análisis didáctico del profesor de matemáticas. En *Actas VI Jornadas de educación matemática región de Murcia* (pp. 25-49). Murcia: Centro de profesores y recursos de Lorca, Mar Menor, Murcia I y Murcia II.
- Gómez, E. (2014). *Evaluación y desarrollo del conocimiento matemático para enseñar la probabilidad en futuros profesores de educación primaria*. Tesis Doctoral. Universidad de Granada.
- Gómez, E., Batanero, C. y Contreras, C. (2014). Conocimiento matemático de futuros profesores para la enseñanza de la probabilidad desde el enfoque frecuencial. *Bolema*, 28(48), 209-229.
- Green, D. R. (1983a). A Survey of probabilistic concepts in 3000 pupils aged 11-16 years. En D.R. Grey et al. (Eds.), *Proceedings of the First International Conference on Teaching Statistics* (Vol.2, pp. 766-783). Universidad de Sheffield: Teaching Statistics Trust.
- Green, D. R. (1983b). From thumbtacks to inference. *School Science and Mathematics*, 83(7), 541-551.

- Green, D.R. (1989). Schools students' understanding of randomness. En R. Morris (Ed.), *Studies in Mathematics Education. v.7: The teaching of statistics* (pp. 27-39). Paris: UNESCO.
- Green, D. R. (1991). A longitudinal study of children's probability concepts. En D. Vere Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 320-328). Dunedin: Universidad de Otago.
- Guisasola, J., y Barragués, J. I. (2002). Heurísticas y sesgos de los estudiantes de primer ciclo de universidad en la resolución de problemas de probabilidad. *Enseñanza de las Ciencias, 20*(2), 285-302.
- 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). Springer.
- Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics 6, 187-205.*
- Holmes, P. (1980). *Teaching Statistics 11-16.* Sloug: Foulsham Educational.
- Huerta, M. P. (2015). La resolución de problemas de probabilidad con intención didáctica en la formación de maestros y profesores de matemáticas. En C. Fernández, M. Molina y N. Planas (Eds), *Investigación en Educación Matemática XIX* (pp. 105-119). Alicante: Sociedad Española de Investigación en Educación Matemática.
- Jones, G. A. (Ed.). (2005). *Exploring probability in school: Challenges for teaching and learning.* New York; Springer.
- Jones, G., Langrall, C. y Mooney, E. (2007). Research in probability: responding to classroom realities. En F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* Greenwich, CT: Information Age Publishing y NCTM.
- Jones, G. A. y Thornton, C. A. (2005). An overview of research into the teaching and learning of probability. En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65-92). Dordrecht: Kluwer.
- Kadijevich, D., Kokol-Voljc, V. y Lavicza, Z. (2008). Towards a suitable designed instruction on statistical reasoning: Understanding sampling distribution with technology. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table*

- Conference*. Monterrey: International Statistical Institute. Disponible en: [https://www.stat.auckland.ac.nz/~iase/publications/rt08/T4P9\\_Kadijevich.pdf](https://www.stat.auckland.ac.nz/~iase/publications/rt08/T4P9_Kadijevich.pdf)
- Kahneman, D., & Egan, P. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Ko, E. S. (2016). Development of an understanding of a sampling distribution. En D. Ben-Zvi y K. Makar (Eds.), *The Teaching and Learning of Statistics* (pp. 63-70). New York; Springer.
- Kline, P. (2013). *Handbook of psychological testing*. Londres: Routledge.
- Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.
- Konold, C., Lohmeier, J., Pollatsek, A., Well, A., Falk, R., & Lipson, A. (1991). Novice views on randomness. *Proceedings of the thirteenth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 167-173). Blacksburg: Virginia Polytechnic Institute and State University.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., & Mayr, S. (2002). Students' use of modal clumps to summarize data. En B. Phillips (Ed.). *Proceedings of the Sixth International Conference on Teaching Statistics*, Cape Town, South Africa: IASE. Disponible en [https://www.stat.auckland.ac.nz/~iase/publications/1/8b2\\_kono.pdf](https://www.stat.auckland.ac.nz/~iase/publications/1/8b2_kono.pdf).
- Krippendorff, K. (2013). *Content analysis: an introduction to its methodology*. London, Sage.
- Langer, E. J. (1982). The illusion of control. En D. Kahneman, P. Slovic, y A. Tversky (Eds.), *Judgment under uncertainty: Heuristic and biases* (pp. 231-238). New York: Cambridge University Press.
- Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, 23, 557-568.
- León, O. G. y Montero, I. (2002). *Métodos de investigación en psicología y educación*. Madrid: McGraw-Hill.
- Liu, Y. y Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies*, 4(2), 126-138.

- López-Martín, M.M., Batanero, C., Díaz-Batanero, C., y Gea, M.M. (2016). La inferencia estadística en las pruebas de acceso a la universidad en Andalucía. *Revista Paranaense de Educação Matemática*, 5(8), 33-59.
- Makar, K., Bakker, A. y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 152-173.
- Makar, K. y Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- Makar, K. y Rubin, A. (2018). Learning about statistical inference. En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International handbook of research in statistics education*. (pp. 261-294). New York: Springer.
- Martínez-Arias, R. Lloreda, M. y Lloreda, M. (2014). *Psicometría*. Alianza Editorial.
- MEC (2007a). *Real Decreto 1631/2006, de 29 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación Secundaria Obligatoria*. España: Ministerio de Educación y Cultura.
- MEC (2007b). *Real Decreto 1467/2007, de 2 de noviembre, por el que se establece la estructura del bachillerato y se fijan sus enseñanzas mínimas*. España: Ministerio de Educación y Cultura.
- MECD. (2014). *Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria*. Madrid: Autor.
- MECD (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. Madrid: Autor.
- Meder B., Gigerenzer G. (2014) Statistical Thinking: No One Left Behind. In: E., Sriraman B. (eds) Probabilistic thinking. Advances in mathematics education. Springer, Dordrecht
- Meletiou-Mavrotheris, M., y Pappastodemou, E. (2015). Developing students' reasoning about samples and sampling in the context of informal inferences. *Educational Studies in Mathematics*, 88(3), 385-404.
- Méndez, H. (1991). *Understanding the central limit theorem*. Tesis doctoral. Universidad de California. UMI 6369.
- Moore, D. S. (1991). Teaching statistics as a respectable subject. En F. Gordon y S. Gordon (Eds.), *Statistics for the twenty-first century* (pp.14-25). Mathematical Association of America.

- Moreno, A. y Vallecillos, A. (2001). Exploratory study on inferential' concepts learning in secondary level in Spain. En M. van der Heuvel (Ed.), *Proceedings of the 25 th Conference of the International Group of the Psychology of Mathematics Education (PME)* (p. 343). Utrecht: Freudenthal Institute and Utrecht University
- Moses, L.E. (1992). The reasoning of statistical inference. En D. C. Hoaglin y D. S. Moore (Eds.), *Perspectives on contemporary statistics*, (pp. 107-121). Washington: Mathematical Association of America.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston: VA.
- Ottaviani, M. G. (1998). Developments and perspectives in statistical education. *Proceedings of the Joint IASS/IAOS Conference. Statistics for Economic and Social Development*. Aguascalientes, México: IAOS (CD ROM).
- Poincaré, H. (1987). *Calcul des probabilités*. Paris, France: Jacques Gabay (Original work published in 1912).
- Pollatsek, A., Konold, C.E., Well, A.D. y Lima, S. (1991). Beliefs underlying random sampling. *Memory and Cognition*, 12, 395-401.
- Porta, L. y Silva, M. (2003). *La investigación cualitativa: el análisis de contenido en la investigación educativa*. Córdoba: CENDIE.
- Reading, C. y Shaughnessy, J. M. (2004). Reasoning about variation. En D. Ben-Zvi y J. B. Garfield (Eds), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-22 Dordrecht, Holanda: Springer.
- Ridgway, J. (2016). Implications of the data revolution for statistics education. *International Statistical Review* 84(3), 528-549. DOI: 10.1111/insr.12110.
- Rolleri, J. L. (2002). La probabilidad como grado de posibilidad. *Crítica: Revista Hispanoamericana de Filosofía*, 34, 3-26.
- Rossman, A. (2008). Reasoning about informal statistical inference: one statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- Rubin, A., Bruce, B. y Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. En D. Vere-Jones (Ed.), *Proceedings of the third international conference on teaching statistics* (Vol. 1, pp. 314-319). Otago, Nueva Zelanda: International Statistical Institute.
- Ruiz-Reyes, K., Begué, N., Batanero, C. y Contreras, J. M. (2017). Un estudio comparado de los contenidos del muestreo en la Educación Secundaria Obligatoria en Chile. *Educação Matemática Pesquisa*, 19(3), 67-83.

- Saldanha, L. y Thompson, P. (2002) Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51, 257-270.
- Sánchez, E., García-García, J. I. y Mercado, M. (2018). Determinism and empirical commitment in the probabilistic reasoning of high school students. En C. Batanero y E. Chernoff (Eds.), *Teaching and learning stochastics* (pp. 223-239). Springer, Cham.
- Savard, A. (2010). Simulating the risk without gambling: can student conceptions generate critical thinking about probability? En C. Reading (Ed.), *Proceedings of the Eighth International Conference on Teaching Statistics Ljubljana*, Slovenia: International Statistical Institute.
- 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). Otago, Nueva Zelanda: International Statistical Institute).
- Serrano, L. (1996). *Significados institucionales y personales de objetos matemáticos ligados a la aproximación frecuencial de la enseñanza de la probabilidad*. Tesis doctoral. Universidad de Granada.
- Serrano, L., Batanero, C., Ortiz, J. J. y Cañizares, M. J. (1998). Heurísticas y sesgos en el razonamiento probabilístico de los alumnos de secundaria. *Educación Matemática*, 10(1), 7-25
- Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D.A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). New York: Macmillan.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. En F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957-1009). Charlotte, NC: Information Age Publishing.
- Shaughnessy, J. M., Garfield, J. y Greer, B. (1996). Data handling. En A. J. Bishop, K. Clements, C. Keitel, J. Kilpatrick y C. Laborde (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht, Netherlands: Kluwer.
- Shaughnessy, J.M., Ciancetta, M. y Canada, D. (2004). Types of student reasoning on sampling tasks. En M.J. Høines y A.B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol.4, pp. 177-184). Bergen, Noruega: International Group for the Psychology of Mathematics Education.



- Shaughnessy, J. M., Watson, J., Moritz, J. y Reading, C. (1999, April). School mathematics students' acknowledgment of statistical variation. Trabajo presentado en *el 77th Annual NCTM Conference, San Francisco, California*.
- Tamayo, M. (2004). *El proceso de la investigación científica*. México: Editorial Limusa.
- Tversky, A y Kahneman, D. (1971). Belief in the law of small numbers. *Psychological bulletin*, 76(2), 105.
- Tversky, A. y Kahneman, D. (1974): Judgement under uncertainty: Heuristics and biases. *Science*.185, 1124-1131.
- Tversky, A. y Kahneman, D. (1982). Judgments of and by representativeness. En D. Kahneman, P.Slovic y A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 117-128). New York: Cambridge University Press.
- Valdez, J. C. (2016). *Las grandes ideas de probabilidad en el razonamiento informal de estudiantes de bachillerato*. Tesis Doctoral. Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. México.
- Vanhoof, S., Castro Sotos, A. E., Onghena, P. y Verschaffel, L. (2007). Students' reasoning about sampling distributions before and after the sampling distribution activity. *Proceedings of the 56th Session of the International Statistical Institute*. Lisboa: International Statistical Institute. Disponible en: [http://iase-web.org/documents/papers/isi56/CPM80\\_Vanhoof.pdf](http://iase-web.org/documents/papers/isi56/CPM80_Vanhoof.pdf).
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88(421), 1-8.
- Watson, J. M. (2004). Developing reasoning about samples. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 277–294). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Watson, J.M. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Watson, J. y Kelly, B. (2005). Cognition and instruction: Reasoning about bias in sampling. *Mathematics Education Research Journal* 17(1), 24-57.
- Watson, J. M. y Moritz, J. B. (2000a). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1) 44-70.
- Watson, J. M. y Moritz, J. B. (2000b). Development of understanding of sampling for statistical literacy. *The Journal of Mathematical Behavior*, 19(1), 109-136.
- Weber, R. P. (1990). Basic content analysis. Londres: Sage.

- Well, A. D., Pollastsek, A. y Boyce, S. J. (1990). Understanding the effects of the sample size on the variability of the means. *Organizational Behavior and Human Decision Processes*, 47, 289-312.
- Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 221-248.
- Yarritu, I., Matute, H. y Vadillo, M. A. (2014). Illusion of control: the role of personal involvement. *Experimental Psychology*, 61(1), 38-47.
- Zapico, M. (2007). Interrogantes acerca de análisis de contenido y del discurso en los textos escolares. En MINEDUC (Ed.), *Primer Seminario Internacional de Textos Escolares (SITE 2006)* (pp. 149-155). Santiago: MINEDUC.
- Zieffler, A., Garfield, J., Delmas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2).
- Zieffler A., Garfield J. y Fry E. (2018) What Is Statistics Education?. In: Ben-Zvi D., Makar K., Garfield J. (Eds.), En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International handbook of research in statistics education*. (pp. 261-294). New York: Springer.

## A1. PUBLICACIONES DERIVADAS DE LA TESIS

### Artículos

- Batanero, C., Begué, N., Gea, M.M. y Roa, R. (2019). El muestreo: Una idea estocástica fundamental. *Suma*, 90, 41-47.
- Batanero, C., Gea, M.M. y Begué, N. (2019). El sentido del muestreo. *Números*, 100, 121-124.
- Begué, N., Batanero, C. y Gea, M. M. (2018). Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria. *Enseñanza de las Ciencias*, 36(2), 63-79.
- Begué, N., Batanero, C., Gea, M.M y Díaz-Levicoy, D. (2019). Distribuciones muestrales: Dificultades en su comprensión y actividades de simulación en poblaciones binomiales. *UNIÓN* 100, 100-108.
- Begué, N., Batanero, C., Ruiz, K. y Gea, M.M. (2019). Understanding sampling: a summary of the research. *Beio*, 35(1), 49-78.
- Ruiz-Reyes, K., Begué, N., Batanero, C. y Contreras, J. M. (2017). Un estudio comparado de los contenidos del muestreo en la Educación Secundaria Obligatoria en Chile. *Educação Matemática Pesquisa*, 19(3).

### Congresos internacionales

- Batanero, C., Begué, N. y Gea, M. M. (2018). ¿Cómo desarrollar el sentido del muestreo en los estudiantes?, En I. Álvarez (Ed.), *Memorias del III Encuentro Colombiano de Educación Estocástica* (pp. 11-22). Bogotá, Colombia: Asociación Colombiana de Educación Estocástica.
- Begué, N. (2019). El muestreo en el bachillerato español En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html)
- Begué, N. y Batanero, C. (2017). Significado del muestreo en el currículo de Educación Secundaria Obligatoria. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín(Eds.), *Actas del Segundo Congreso*

- International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos. Disponible en, [enfoqueontosemiotico.ugr.es/civeos.html](http://enfoqueontosemiotico.ugr.es/civeos.html).
- Begué, N., Batanero, C., Díaz-Levicoy, D. Gea, M.M. (2019). Lanzamiento a una canasta de baloncesto: interpretación de una situación binomial por estudiantes de Bachillerato. Noveno Encuentro Internacional en la Enseñanza de la Probabilidad y la Estadística 2019 Puebla.
- Begué, N., Batanero, C. y Gea, M.M. (2019). Argumentos de los estudiantes de bachillerato en la generación de muestras de la distribución binomial. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), Actas del Tercer Congreso Internacional Virtual de Educación Estadística. Disponible en [www.ugr.es/local/fqm126/civeest.html](http://www.ugr.es/local/fqm126/civeest.html)
- Begué, N., Gea, M.M. Batanero, C., Beltrán, P. (2017). Meaning of sampling for secondary school students. En L. Gómez Chova, A. López Martínez, I. Candel Torres (Eds.), 10th annual International Conference of Education, Research and Innovation. International Academy of Technology, Education and Development (pp. 1430-1435). Sevilla: IATED Academy
- Begué, N., Gea, M.M. Batanero, C., Beltrán, P. (2018). Do high school students understand the sampling distribution of proportion? Trabajo En M. A. Sorto, A. White, & L. Guyot (Eds.), Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan: International Statistical Institute.
- Begué, N., Batanero, C., Gea, M.M. y Beltrán, P. (2019). Technology as a tool to understand sampling in binomial distributions. CIEAEM, Braga, Portugal.
- Begue, N., Ruiz, K., Gea, M.M. y Batanero C. (2017). El muestreo en el currículo de secundaria: un estudio comparado de los currículos en España y Chile. Trabajo presentado en el IX Cibem. Madrid.

### **Congresos Nacionales**

- Begué, N. y Gea, M.M. (2019). Dificultades de los estudiantes en la comprensión de la distribución muestral. En J. M. Marbán, M. Arce, A. Maroto, J. M. Muñoz-Escolano y Á. Alsina (Eds.), Investigación en Educación Matemática XXIII (p. 610). Valladolid: SEIEM
- Begué, N., Batanero, C., Gea, M.M. y Beltrán-Pellicer, P. (2017). Comprensión del enfoque frecuencial de la probabilidad por estudiantes de Educación Secundaria

- Obligatoria. En J.M. Muñoz-Escolano, A. Arnal-Bailera, P. Beltrán-Pellicer, M.L. Callejo y J. Carrillo (Eds.), *Investigación en Educación Matemática XXI* (pp. 137-146). Zaragoza: SEIEM.
- Begué, N. y Gea, M. (2019). Dificultades de los estudiantes en la comprensión de la distribución muestral. *Jornadas de GENAIO*, Granada
- Begué, N. y Gea, M.M. (2019). El muestreo en el currículo de educación secundaria obligatoria. Trabajo presentado en las XIII Jornadas de Aprendizaje y Enseñanza de las Matemáticas. La Coruña: Federación Española de Sociedades de Profesores de Matemáticas.
- Begué, N. y Gea, M. M. (2019). Dificultades de los estudiantes en la comprensión de la distribución muestral. En J. M. Marbán, M. Arce, A. Maroto, J. M. Muñoz-Escolano y Á. Alsina (Eds.), *Investigación en Educación Matemática XXIII* (p. 610). Valladolid: SEIEM.