

ESTADÍSTICA APLICADA. PRÁCTICAS CON SPSS. TEMA 1

1.- INTRODUCCIÓN A SPSS

El SPSS (Statistical Package for the Social Sciences) es un conjunto de programas orientados a la realización de análisis estadísticos aplicados a las ciencias sociales. Con más de 30 años de existencia es, en la actualidad, el paquete estadístico con más difusión a nivel mundial.

El SPSS implementa una gran variedad de temas estadísticos en los distintos módulos del programa. Nosotros utilizaremos solo el **módulo SPSS base**, que incluye, entre otros, análisis y presentación de datos, comparación de medias y análisis de la varianza.

En los siguientes apartados se hace una breve introducción a los conceptos básicos de este programa. Para una mejor exposición de los mismos puede consultarse la “**Guía breve de SPSS**”.

1.1.- PASOS BÁSICOS EN EL ANÁLISIS

- **Introducir los datos en SPSS.**

Es posible:

- **Introducir nuevos datos.**
- **Abrir un archivo creado anteriormente.**

Los archivos de datos tienen extensión *.sav y formato SPSS. Para abrir un archivo de datos de este formato, seleccione en el menú principal: *Archivo/Abrir/Datos*. Por defecto, SPSS dará una relación de los archivos en su directorio con extensión *.sav. Busque y seleccione el archivo que se desee abrir. Además de los archivos con este formato, SPSS puede abrir archivos de EXCEL, LOTUS 1-2-3, dBASE sin necesidad de convertirlos a un formato intermedio ni de introducir información sobre la definición de los datos.

Desde aplicaciones como Microsoft Excel también puede leer los encabezados de las columnas como nombres de variables. Para ello elija en los menús: *Archivo/Abrir/Datos...seleccione Excel*

(*.xls) de la lista desplegable *Archivos de tipo*, aparecerá el cuadro de diálogo *Apertura de origen de datos de Excel*, que permite especificar si se incluyen los nombres de las variables en la hoja de cálculo, así como las casillas que se desea importar. En Excel 5 o posterior, también se pueden especificar las hojas de cálculo que se desea importar. Si los encabezados de las columnas no cumplen las normas de denominación de variables de SPSS, se convertirán en nombres de variables válidos y los encabezados originales de las columnas se guardarán como etiquetas de variable.

- **Seleccionar un procedimiento estadístico para analizar** los datos con el sistema de menús.

- **Seleccionar las variables** para el análisis.

Las variables que podemos usar en cada procedimiento se muestran en un cuadro de diálogo del que se seleccionan.

- **Ejecutar el procedimiento y ver resultados.**

Los resultados aparecen en una ventana de resultados y se pueden guardar como archivos de extensión *.spo. Los gráficos se pueden modificar en la ventana del editor de gráficos.

1.2.- ENTORNO DE TRABAJO

Existen diversos **tipos de ventanas** en SPSS. Nosotros, principalmente, utilizaremos dos:

- **Editor de datos.** Es la ventana que se abre automáticamente cuando se inicia una sesión de SPSS. Muestra el contenido del archivo de datos actual. Con él, se pueden crear nuevos archivos o modificar los ya existentes.
- **Visor de resultados.** Todas las tablas, gráficos y los resultados estadísticos se muestran en el visor. Puede editar resultados y guardarlos. Esta ventana se abre automáticamente la primera vez que se ejecuta un procedimiento.

Ejemplo 1.-

- Abrir archivo de datos “Datos de empleados.sav”
- Realizar el procedimiento:
Analizar/ Estadísticos Descriptivos/ Descriptivos con la variable “Salario Actual”
- Realizar lo mismo con la variable “Meses desde el contrato”.

Además de las distintas ventanas, existen un conjunto de elementos, comunes a todas ellas que configuran la apariencia general del SPSS:

- **Barra de títulos.** Muestra el icono de SPSS, sigue con el nombre del fichero que estamos utilizando y el nombre de la ventana activa y termina con los botones de minimizar, maximizar y cerrar ventana.
- **Barra de menús.** Recoge las denominaciones de los menús de SPSS a través de los cuales se pueden ejecutar todos los posibles comandos que proporciona el paquete.
- **Barra de herramientas.** Proporciona un acceso rápido y fácil a las tareas más comunes de cada ventana de SPSS. El significado de cada icono puede verse situando el ratón sobre el propio icono.
- **Barra de estado.** Suministra información sobre el estado en que se encuentra SPSS.

1.3.- EDITOR DE DATOS

El editor de datos proporciona dos vistas:

- **Vista de los datos.** Muestra los valores de datos reales o las etiquetas de valor definidas. Las funciones de la vista de datos son similares a las que se encuentran en aplicaciones de hojas de cálculo, sin embargo, existen algunas diferencias:
 - Cada fila representa un caso u observación (atención en ejercicio 1).
 - Las columnas son variables.
 - Las casillas contienen valores numéricos o de cadena. A diferencia de una hoja de cálculo, las casillas del editor de datos no pueden contener fórmulas.
- **Vista de variables.** Contiene descripciones de los atributos de cada variable del archivo de datos. Aquí:
 - Las filas son variables.
 - Las columnas son características de las variables.

Ejemplo 2.- Analizar el archivo de datos “Datos de Empleados.sav”.

Una vez que nos encontramos en la ventana de edición de datos de SPSS podemos abrir un archivo de datos creado con anterioridad o crear un nuevo archivo. En el caso en que queramos **crear un archivo de datos nuevo**, el primer paso es el de definir las variables que formarán el archivo. Para **definir una variable** se pueden seguir dos procedimientos:

- **Haciendo doble clic** con el botón izquierdo del ratón cuando el puntero del mismo se encuentra situado **en la palabra *var*** del extremo superior de la columna (en vista de datos)
- **Pulsando la pestaña de vista de variables.**

Para la especificación del **nombre de las variables** se debe tener en cuenta:

- El nombre debe comenzar por una letra. Los demás caracteres pueden ser letras, dígitos, puntos o los símbolos @, #, _ o \$.
- Los nombres de variable no pueden terminar en punto.
- Se debe evitar acabar los nombres de variable con subrayado (para evitar conflictos con las variables creadas automáticamente por algunos procedimientos).
- La longitud del nombre no debe exceder los 64 bytes. Normalmente, 64 bytes suelen equivaler a 64 caracteres en idiomas de un solo byte (por ejemplo, inglés, francés, alemán, español, italiano, hebreo, ruso, griego, árabe, tailandés) y a 32 caracteres en los idiomas de dos bytes (por ejemplo, japonés, chino, coreano).
- No se pueden utilizar espacios en blanco ni caracteres especiales (por ejemplo, !, ?, ' y *).
- Cada nombre de variable debe ser único; no se permiten duplicados.
- Las palabras reservadas no se pueden utilizar como nombres de variable. Las palabras reservadas son: ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.
- Los nombres de variable se pueden definir combinando de cualquier manera caracteres en mayúsculas y en minúsculas, esta distinción entre mayúsculas y minúsculas se conserva en lo que se refiere a la presentación.
- Cuando es necesario dividir los nombres largos de variable en varias líneas en los resultados, SPSS intenta dividir las líneas aprovechando los subrayados, los puntos y los cambios de minúsculas a mayúsculas.

Una vez que se haya determinado el nombre de la variable, hay que definir sus especificaciones:

- **Tipo.** Especifica el tipo de datos que contiene la variable. Si se pulsa el botón tipo y luego los puntos suspensivos que aparecen en la casilla aparece un cuadro de dialogo. Los tipos más usuales son numérico y cadena.

Tipo de variable especifica los tipos de datos de cada variable. Por defecto se asume que todas las variables nuevas son numéricas. Se puede utilizar Tipo de variable para cambiar el tipo de datos. El contenido del cuadro de diálogo Tipo de variable depende del tipo de datos seleccionado. Para algunos tipos de datos, hay cuadros de texto para la anchura y el número de decimales; para otros tipos de datos, simplemente puede seleccionar un formato de una lista desplegable de ejemplos.

Los **tipos de datos** disponibles son los siguientes:

- **Numérico.** Una variable cuyos valores son números. Los valores se muestran en formato numérico estándar. El Editor de datos acepta valores numéricos en formato estándar o en notación científica.
- **Coma.** Una variable numérica cuyos valores se muestran con comas que delimitan cada tres posiciones y con el punto como delimitador decimal. El Editor de datos acepta valores numéricos para este tipo de variables con o sin comas, o bien en notación científica. Los valores no pueden contener comas a la derecha del indicador decimal.
- **Punto.** Una variable numérica cuyos valores se muestran con puntos que delimitan cada tres posiciones y con la coma como delimitador decimal. El Editor de datos acepta valores numéricos para este tipo de variables con o sin puntos, o bien en notación científica. Los valores no pueden contener puntos a la derecha del indicador decimal.
- **Notación científica.** Una variable numérica cuyos valores se muestran con una E intercalada y un exponente con signo que representa una potencia de base diez. El Editor de datos acepta para estas variables valores numéricos con o sin el exponente. El exponente puede aparecer precedido por una E o una D con un signo opcional, o bien sólo por el signo (por ejemplo, 123, 1,23E2, 1,23D2, 1,23E+2 y 1,23+2).
- **Fecha.** Una variable numérica cuyos valores se muestran en uno de los diferentes formatos de fecha-calendario u hora-reloj. Seleccione un formato de la lista. Puede introducir las fechas

utilizando como delimitadores: barras, guiones, puntos, comas o espacios. El rango de siglo para los valores de año de dos dígitos está determinado por la configuración de las opciones (en el menú Edición, seleccione Opciones y, a continuación, pulse en la pestaña Datos).

- **Dólar.** Una variable numérica que se muestra con un signo dólar inicial (\$), comas que delimitan cada tres posiciones y un punto como delimitador decimal. Se pueden introducir valores de datos con o sin el signo dólar inicial.
 - **Moneda personalizada.** Una variable numérica cuyos valores se muestran en uno de los formatos de moneda personalizados que se hayan definido previamente en la pestaña Moneda del cuadro de diálogo Opciones. Los caracteres definidos en la moneda personalizada no se pueden emplear en la introducción de datos pero sí se mostrarán en el Editor de datos.
 - **Cadena.** Una variable cuyos valores no son numéricos y, por lo tanto, no se utilizan en los cálculos. Los valores pueden contener cualquier carácter siempre que no se exceda la longitud definida. Las mayúsculas y las minúsculas se consideran diferentes. Este tipo también se conoce como variable alfanumérica.
-
- **Nivel de medida.** Puede especificar el nivel de medida como Escala (datos numéricos de una escala de intervalo o de razón), Ordinal o Nominal. Los datos nominales y ordinales pueden ser de cadena (alfanuméricos) o numéricos.
-
- **nominal.** Una variable puede ser tratada como nominal cuando sus valores representan categorías que no obedecen a una ordenación intrínseca. Por ejemplo, el departamento de la compañía en el que trabaja un empleado. Son ejemplos de variables nominales: la región, el código postal o la confesión religiosa.
 - **ordinal.** Una variable puede ser tratada como ordinal cuando sus valores representan categorías con alguna ordenación intrínseca. Por ejemplo los niveles de satisfacción con un servicio, que vayan desde muy insatisfecho hasta muy satisfecho. Son ejemplos de variables ordinales: las puntuaciones de actitud que representan el nivel de satisfacción o confianza y las puntuaciones de evaluación de la preferencia.
 - **escala.** Una variable puede ser tratada como de escala cuando sus valores representan categorías ordenadas con una métrica con significado, por lo que son adecuadas las comparaciones de distancia entre valores. Son ejemplos de variables de escala: la edad en años y los ingresos en dólares.

Nota: Para variables de cadena ordinales, se asume que el orden alfabético de los valores de cadena indica el orden correcto de las categorías. Por ejemplo, en una variable de cadena cuyos valores sean bajo, medio, alto, se interpreta el orden de las categorías como alto, bajo, medio (orden que no es el correcto). Por norma general, se puede indicar que es más fiable utilizar códigos numéricos para representar datos ordinales.

- **Anchura.** Número de dígitos de los valores de esa variable.
- **Columnas.** Anchura de las columnas. Se puede especificar un número de caracteres para el ancho de la columna. Los anchos de columna también se pueden cambiar en la Vista de datos pulsando y arrastrando los bordes de las columnas.
Los formatos de columna afectan sólo a la presentación de valores en el Editor de datos. Al cambiar el ancho de columna no se cambia el ancho definido de una variable. Si el ancho real y definido de un valor es más ancho que la columna, aparecerán asteriscos (*) en la ventana Vista de datos.
- **Decimales.** N° de decimales de los datos.
- **Etiqueta.** Puede asignar etiquetas de variable descriptivas de hasta 256 caracteres de longitud (128 caracteres en los idiomas de doble byte). Las etiquetas de variable pueden contener espacios y caracteres reservados que no se admiten en los nombres de variable.
- **Valores.** Puede asignar **etiquetas de valor** descriptivas a cada valor de una variable. Este proceso es especialmente útil si el archivo de datos utiliza códigos numéricos para representar categorías que no son numéricas (por ejemplo, códigos 1 y 2 para hombre y mujer). Las etiquetas de valor se guardan con el archivo de datos. No es necesario volver a definir las etiquetas de valor cada vez que se abre un archivo de datos. Las etiquetas de valor pueden ocupar hasta 120 bytes. Las etiquetas de valor no están disponibles para las variables de cadena larga (variables de cadena de más de 8 caracteres).
- **Perdidos.** Valores perdidos define los valores de los datos definidos como perdidos por el usuario. Por ejemplo, es posible que quiera distinguir los datos perdidos porque un encuestado se niegue a responder de los datos perdidos porque la pregunta no afecta a dicho encuestado. Los valores de

datos que se especifican como perdidos por el usuario aparecen marcados para un tratamiento especial y se excluyen de la mayoría de los cálculos.

- Las especificaciones de valores perdidos definidos por el usuario se guardan junto con el archivo de datos. No es necesario volver a definir los valores definidos como perdidos por el usuario cada vez que se abre un archivo de datos.
 - Se pueden introducir hasta tres valores perdidos (individuales) de tipo discreto, un rango de valores perdidos o un rango más un valor de tipo discreto.
 - Sólo pueden especificarse rangos para las variables numéricas.
 - No se pueden definir valores perdidos para variables de cadena larga (variables de cadena de más de ocho caracteres).
 - Se considera que son válidos todos los valores de cadena, incluidos los valores vacíos o nulos, a no ser que se definan explícitamente como perdidos. Para definir como perdidos los valores nulos o vacíos de una variable de cadena, escriba un espacio en blanco en uno de los campos debajo de la selección Valores perdidos discretos.
-
- ***Alineación.*** Alineación de los datos (Izquierda, derecha o centro)

Una vez definidas las variables, para la introducción de los datos (en la pestaña vista de datos) habrá que situar el cursor en la primera celda de la columna y comenzar a escribir los distintos valores, pulsando ENTER o moviéndonos con el cursor.

También podemos **modificar datos ya creados**:

- **Insertar un nuevo caso** entre los casos existentes.

Seleccionar en la vista de datos, cualquier casilla debajo de la posición donde se desea insertar el nuevo caso y

- Elegir los menús: *Datos/Insertar Caso*
- O el botón de la barra de herramientas
- O con el botón derecho del ratón elegir *Insertar caso*.

- **Insertar una nueva variable** entre las variables existentes.

Seleccionar en la vista de datos, una casilla de la variable a la derecha de la posición donde se desea insertar la nueva variable y

- Elegir los menús: *Datos/Insertar variable*
- O el botón de la barra de herramientas

- O con el botón derecho del ratón elegir *Insertar variable*

- **Mover variables.**

Si queremos mover una variable que está entre otras dos, en la vista de datos, podemos insertar una nueva variable en el lugar donde la queramos copiar, luego cortar de donde estaba y por último pegar en la nueva variable insertada.

- **Borrar algún caso o variable.**

Seleccionar previamente en la vista de datos las filas a borrar, las columnas o el área y pulsar SUPR o *Edición/Borrar* o con el botón derecho del ratón elegir *Eliminar*.

- **Ir a un caso en el editor de datos.**

Elegir los menús: *Datos/Ir al caso* e introducir el número de fila o con el botón de la barra de herramientas.

Para **guardar un archivo de datos creado** tendremos que situarnos en *Archivo/Guardar como*. Nos aparecerá un cuadro de diálogo en el cual debemos indicar el nombre del archivo y el lugar donde queremos guardarlo. En el caso de que se trate de **cambios en un archivo que ya ha sido guardado con anterioridad**, solo tendremos que situarnos en *Archivo/Guardar* o con el botón de la barra de herramientas y el archivo se guardará con el mismo nombre y ubicación que tenía con anterioridad.

EJERCICIOS

1. La siguiente tabla nos muestra la edad de 16 clientes que compraron un determinado producto en una semana determinada. La variable edad es cuantitativa y mostramos sus valores, la variable sexo es cualitativa y utilizamos una variable numérica (escala nominal: 1, hombre; 2, mujer). Los datos son los siguientes:

Hombres	32	32	42	55	37	61	48	43
Mujeres	50	80	61	49	30	21	37	34

Se pide:

- a. Crea un archivo con la definición anterior de las variables y los datos y guárdalo con el nombre *Edad.sav*

- b. Crea una nueva variable denominada *Unidades compradas* y dale los siguientes valores

Hombres	1	1	1	2	1	1	2	1
Mujeres	1	5	1	1	2	2	1	1

Sitúala entre las variables anteriores.

2. La siguiente tabla nos muestra los datos de los representantes de 16 empresas:

SEXO	REGION DE PROCEDENCIA	MESES COMO REPRESENTANTE	INGRESOS MENSUALES en €
Hombre	Andalucía (1)	60	1950
Mujer	Cataluña (2)	72	1235
Mujer	Madrid (3)	48	2251
Hombre	País Valenciano (4)	36	3581
Mujer	Galicia (5)	60	1500
Mujer	Cataluña (2)	24	2500
Hombre	País Vasco (6)	36	5890
Hombre	Andalucía (1)	48	3510
Hombre	Madrid (3)	84	2456
Mujer	Andalucía (1)	84	2474
Hombre	País Vasco (6)	48	3000
Mujer	Madrid (3)	36	2958
Mujer	Galicia (5)	24	1354
Hombre	Cataluña (2)	12	1100
Hombre	Andalucía (1)	16	3581
Hombre	Galicia (5)	10	2456

Se pide:

- a. Crear un archivo de datos con el nombre *Representantes*, en el que debes incluir los datos de la tabla anterior, definiendo las variables de forma adecuada.
- b. Inserta un nuevo caso entre los existentes con estos valores:

SEXO	REGION DE PROCEDENCIA	MESES COMO REPRESENTANTE	INGRESOS MENSUALES en €
Hombre	Cataluña	48	1500

- c. Inserta una nueva variable que será el estado civil de los representantes con los siguientes valores.

soltero
soltero
casado
soltero
divorciado
casado
casado
casado
soltero
soltero
viudo
casado
casado
casado
soltero
soltero
divorciado

- d. Obtenga el número medio de meses como representante y los valores máximo, mínimo y mediano de los ingresos.

2.- CONTRASTES PARAMÉTRICOS

En SPSS, los contrastes paramétricos que se pueden estudiar son aquellos que en las hipótesis hacen referencia a la media de una población normal, cuando tenemos una muestra, dos independientes, o dos relacionadas. Se denominan **prueba T**, y nosotros estudiaremos:

- **Prueba T para una muestra.**
- **Prueba T para dos muestras independientes.**
- **Prueba T para muestras relacionadas** (apareadas)

Estas pruebas se encuentran en el menú *Analizar/Comparar medias*.

- **Analizar/Comparar medias/Medias**

Este procedimiento calcula los estadísticos: media, desviación típica, varianza, suma y número de individuos, para una o más variables (Dependientes), para los distintos valores, niveles o categorías de una o más variables (Independientes). Dentro de los cuadros de variables **Dependientes** e **Independientes** podemos insertar una o más variables. En **Opciones**, podemos dejar por defecto: media, desviación típica y nº de casos, o seleccionar otros.

Ejemplo 3.-

Archivo de datos: “representantes.sav”

Variable dependiente: “Ingresos mensuales”

Variable independientes: “Sexo” y “Estado civil”

Resumen del procesamiento de los casos

	Casos					
	Incluidos		Excluidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
Ingresos mensuales * Sexo	17	100,0%	0	,0%	17	100,0%
Ingresos mensuales * Estado civil	17	100,0%	0	,0%	17	100,0%

Ingresos mensuales * Sexo

Ingresos mensuales

Sexo	Media	N	Desv. típ.
hombre	2902,40	10	1364,08
mujer	2038,86	7	670,50
Total	2546,82	17	1186,23

Ingresos mensuales * Estado civil

Ingresos mensuales

Estado civil	Media	N	Desv. típ.
casado	2794,71	7	1604,64
divorciado	1500,00	2	,00
soltero	2533,29	7	840,30
viudo	3000,00	1	,
Total	2546,82	17	1186,23

2.1- Analizar/Comparar medias/Prueba T para una muestra

Con este procedimiento tratamos de comprobar si la media de una sola variable difiere o no de una constante, es decir, tratamos de contrastar la hipótesis nula $H_0 : \mu = \mu_0$, bajo hipótesis de normalidad, con varianza desconocida y partiendo de una muestra aleatoria simple de la población bajo estudio. En **Contrastar variables** introducimos la variable para la que queremos pedir que contraste si la muestra pertenece a una población normal de media μ_0 , que es el valor que hay que introducir en **Valor de prueba**.

En **Opciones** se puede:

- cambiar el nivel de confianza (por defecto es del 95%) con el que se desea obtener el intervalo de confianza para la diferencia entre la media muestral y la media poblacional (valor de prueba).
- Optar por dos formas diferentes de tratar los casos de valores perdidos:
 - *Excluir casos según análisis*. Esta opción excluye del análisis (de cada prueba T) los casos con valor perdido en la variable concreta que se está contrastando.
 - *Excluir casos según lista*. Esta opción excluye de todos los análisis los casos con algún valor perdido en una cualquiera de las variables seleccionadas en la lista **Contrastar variables**.

Como **resultado** se obtienen dos tablas:

○ **Tabla 1:** Los estadísticos para una muestra: n° de casos, media, desviación típica y error típico de la media (desviación típica de la media muestral).

○ **Tabla 2:** Prueba para una muestra.

- $t \rightarrow$ valor del estadístico usado en el contraste para los datos de la muestra.
- $gl \rightarrow$ grados de libertad de la distribución t-Student que sigue el estadístico de contraste.
- Sig (bilateral) \rightarrow valor que permite decidir la aceptación o no de la hipótesis nula. Es la significación muestral de la hipótesis nula, es decir, el p-valor.

Si $p \geq \alpha$, se acepta la hipótesis nula.

Si $p \leq \alpha$, se rechaza la hipótesis nula.

Si realizamos un contraste unilateral, nuestro valor de p será la mitad del que aparezca en el cuadro (que siempre es bilateral).

- Diferencia de medias \rightarrow entre la observada en la muestra y el valor que aparece en la hipótesis nula.
- Intervalo de confianza para la diferencia (de medias $\mu - \mu_0$) \rightarrow es otra forma de comprobar si se verifica la hipótesis nula, ya que si 0 está en dicho intervalo, aceptamos la hipótesis nula.

Ejemplo 4.- Con el archivo de datos “representantes.sav” contrastar si la media de los ingresos mensuales es igual a 2500.

Tabla 1

Estadísticos para una muestra

	N	Media	Desviación típ.	Error típ. de la media
Ingresos mensuales	17	2546,82	1186,23	287,70

Tabla 2

Prueba para una muestra

	Valor de prueba = 2500					
	t	gl	Sig. (bilateral)	Diferencia de medias	95% Intervalo de confianza para la diferencia	
					Inferior	Superior
Ingresos mensuales	,163	16	,873	46,82	-563,08	656,73

Notas:

$t_{16;0,975}=2,12$.

$(46'82 - [2'12 \times 287'7], 46'82 + [2'12 \times 287'7]) = (-563'08, 656'73)$, salvo errores de redondeo.

Intervalo de confianza para la media μ : $(2500 - 563'08, 2500 + 656'73)$

2.2- Analizar /Comparar medias/ Prueba T para muestras independientes

Con este procedimiento se quiere comprobar si las medias poblacionales de dos poblaciones normales e independientes son iguales o no ($H_0 : \mu_x = \mu_y$)

Para contrastar esta hipótesis **existen distintos estadísticos según las varianzas poblacionales sean iguales o no**, esto se estudia con la prueba de Levene, cuyo estadístico sigue una distribución F-Snedecor. Todo se realiza a la vez con sólo definir el contraste.

Al seleccionar este procedimiento, aparece un cuadro de diálogo en el que en *Contrastar variables* hay que colocar la/s variables cuantitativas que intervienen en el contraste, y en *Variable de agrupación* seleccionamos la variable que nos sirve para **definir los dos grupos en la población**. Dicha variable aparecerá seguida de dos interrogantes, para que definamos los grupos de la variable que queremos que realice. Para ello pulsamos el botón **Definir grupos**. Hay dos opciones excluyentes:

- *Usar valores especificados*. Tendremos que escribir un valor para el grupo 1 y otro para el 2. Los casos con otros valores quedan excluidos del análisis.
- *Punto de Corte*. Podemos escribir un número que divida los valores de la variable de agrupación en dos conjuntos.

En **Opciones** se puede:

- Cambiar el nivel de confianza (por defecto es del 95%) con el que se desea obtener el intervalo de confianza para la diferencia de medias.
- Optar por dos formas diferentes de tratar los casos de valores perdidos:
 - *Excluir casos según análisis*. Esta opción excluye del análisis (de cada prueba T) los casos con valor perdido en la variable agrupación o en la variable concreta que se está contrastando en ese análisis.
 - *Excluir casos según lista*. Esta opción excluye de todos los análisis los casos con algún valor perdido en la variable agrupación o en una cualquiera de las variables seleccionadas en la lista *Contrastar variables*.

Como **resultado** se obtienen dos tablas:

- **Tabla 1**: Los estadísticos para cada muestra: nº de casos, media, desviación típica y error típico de la media.
- **Tabla 2**: Como ya hemos comentado, este procedimiento utiliza distintos estadísticos según se verifique la igualdad o no de las varianzas, y para comprobarlo está la prueba de Levene, en la que F es el valor del estadístico de la prueba y sig es el p-valor. Esto aparece en las tres primeras columnas. Las restantes son la prueba T, que se interpreta igual que para una muestra.

Ejemplo 5.- Con archivo de datos “representantes.sav”, contrastar si los ingresos mensuales medios para los hombres y las mujeres son iguales o no.

Tabla 1

Estadísticos de grupo

Sexo		N	Media	Desviación típ.	Error típ. de la media
Ingresos mensuales	hombre	10	2902,40	1364,08	431,36
	mujer	7	2038,86	670,50	253,42

Tabla 2

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia	
									Inferior	Superior
Ingresos mensuales	Se han asumido varianzas iguales	1,662	,217	1,539	15	,145	863,54	561,08	-332,36	2059,45
	No se han asumido varianzas iguales			1,726	13,816	,107	863,54	500,30	-210,83	1937,91

Nota:

$$t = \frac{863'54}{561'08} = 1'539$$

$$t = \frac{863'54}{500'30} = 1'726$$

2.3- Analizar /Comparar medias/ Prueba T para muestras relacionadas

La prueba T para dos muestras relacionadas permite contrastar hipótesis referidas a la diferencia entre dos medias relacionadas.

Ahora se dispone de una población de diferencias con media μ_D obtenida al **restar las puntuaciones del mismo grupo de casos en dos variables diferentes o en la misma variable medida en dos momentos diferentes** (de ahí que se hable de muestras relacionadas). De esa población de diferencias se extrae una muestra aleatoria de tamaño n y se utiliza la media \bar{Y}_D de esas n diferencias para contrastar la hipótesis nula de que la media μ_D de la población de diferencias vale cero. **Desde el punto de vista estadístico, este contraste es idéntico al presentado en el apartado Prueba T para una muestra.**

Al seleccionar este procedimiento, aparece un cuadro de diálogo en el que en *Variables relacionadas* hay que trasladar el par de variables cuyas medias se desea comparar. Las variables deben trasladarse a esas listas por pares, es decir, es necesario marcar dos variables de la lista para que el botón flecha esté activo.

En *Opciones* se puede:

- Cambiar el nivel de confianza (por defecto es del 95%) con el que se desea obtener el intervalo de confianza para la diferencia de medias.
- Optar por dos formas diferentes de tratar los casos de valores perdidos:
 - *Excluir casos según análisis*. Esta opción excluye de cada análisis (de cada prueba T) los casos con valor perdido en cualquiera de las dos variables que están siendo contrastadas.
 - *Excluir casos según lista*. Esta opción excluye de todos los análisis los casos con algún valor perdido en cualquiera de las variables seleccionadas en la lista *Variables relacionadas*.

Como **resultado** se obtienen tres tablas:

- **Tabla 1:** Los estadísticos para cada muestra: nº de casos, media, desviación típica y error típico de la media.
- **Tabla 2:** Contiene el coeficiente de correlación de Pearson entre ambas variables junto con el nivel crítico bilateral que le corresponde bajo la hipótesis de independencia. **El concepto**

de correlación se refiere al grado de parecido o variación conjunta de dos o más variables. El coeficiente de Pearson estudia el grado de relación lineal existentes entre dos variables. Toma valores entre -1 y 1: un valor de 1 indica relación lineal perfecta positiva, un valor de -1 indica relación lineal perfecta negativa (en ambos casos los puntos del correspondiente diagrama de dispersión se encuentran dispuestos en una línea recta), un valor de 0 indica relación lineal nula. **Un coeficiente de correlación alto no implica causalidad, es decir, dos variables pueden estar linealmente relacionadas sin que una sea la causa de la otra.**

- **Tabla 3:** En la primera mitad incluye tres estadísticos referidos a las diferencias entre cada par de puntuaciones: media, desviación típica y error típico de la media. La siguiente columna contiene el intervalo de confianza para la diferencia de medias. La segunda mitad de la tabla informa sobre el valor del estadístico *t*, sus grados de libertad y el nivel crítico bilateral.

Ejemplo 6.- Con el archivo de datos “Datos de empleados.sav”, contrastar si el salario medio inicial y el salario medio actual difieren o no.

Tabla 1

Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. De la media
Par 1	Salario actual	\$34,419.57	474	\$17,075.661	\$784.311
	Salario inicial	\$17,016.09	474	\$7,870.638	\$361.510

Tabla 2

Correlaciones de muestras relacionadas

	N	Correlación	Sig.
Par 1 Salario actual y Salario inicial	474	,880	,000

Tabla 3**Prueba de muestras relacionadas**

	Diferencias relacionadas					t	gl	Sig. (bilateral)
	Media	Desviación típ.	Error típ. de la media	95% Intervalo de confianza para la diferencia				
				Inferior	Superior			
Par 1 Salario actual - Salario inicial	\$17,403.481	\$10,814.620	\$496.732	\$16,427.407	\$18,379.555	35,036	473	,000

Nota:

$$t = \frac{17403'481}{496'732} = 35'036$$

EJERCICIOS

1. **(T, una muestra)** Los valores sobre las longitudes en milímetros de 50 filamentos de la producción de una máquina (que se supone normal) son los siguientes:

102	98	93	100	98	105	115	110	99	120
115	130	100	86	95	103	105	92	99	134
116	118	89	102	128	99	119	128	110	130
112	114	106	114	100	116	108	113	106	105
120	106	110	100	106	117	109	108	105	106

- Crea un archivo con estos datos denominado *filamentos.sav*.
- Hallar un intervalo de confianza para la media de la producción basado en esta muestra al 90% de confianza.
- Contrastar la hipótesis nula de que la longitud media de los filamentos de la producción es de 100 al nivel $\alpha = 0.10$.

2. **(T, dos muestras independientes)** Una central de productos lácteos recibe diariamente la leche de dos granjas X e Y. Deseando estudiar la calidad de los productos recibidos se extraen dos muestras, una de cada granja (independientes), y se analiza el contenido de materia grasa, obteniendo los siguientes resultados:

X	0.32	0.29	0.30	0.28	0.33	0.31	0.30	0.29	0.33	0.32	0.30
	0.29										
Y	0.28	0.30	0.32	0.29	0.31	0.29	0.33	0.32	0.29	0.32	0.31
	0.29	0.32	0.31	0.32	0.33						

- Crea un archivo con estos datos denominado *granjas.sav*.
- Suponiendo normal la variable estudiada (contenido de materia grasa) en ambas poblaciones independientes, realizar un contraste de igualdad de medias para $\alpha = 0.01$
- Al nivel de confianza del 99%, determinar un intervalo de confianza para la diferencia de medias.

3. **(T, dos muestras independientes)** Con el archivo de datos *Representantes.sav*, comparar los ingresos mensuales entre los hombres y las mujeres.

4. **(T, dos muestras independientes)** Con el archivo de datos *Edad.sav*, comparar la edad entre los hombres y las mujeres.

5. **(T, dos muestras relacionadas)** Un bioquímico quiere comparar dos métodos de determinación de la concentración de testosterona en la orina. Sean A y B dichos métodos. Para la comparación, dispone de 9 muestras de orina (los valores vienen dados en miligramos contenidos en la orina):

Muestra nº	1	2	3	4	5	6	7	8	9
A (mg/muestra)	0,47	1,02	0,33	0,70	0,94	0,85	0,39	0,52	0,47
B (mg/muestra)	0,41	1,00	0,46	0,61	0,84	0,87	0,36	0,52	0,51

Crea un archivo con estos datos denominado *testosterona.sav*. Compara la concentración media de testosterona determinada por ambos métodos.