

## Introducción

El texto docente que se desarrolla en los siguientes capítulos constituye una guía en formato multimedia, eminentemente práctica, diseñada para el aprendizaje del análisis de datos mediante el paquete estadístico *IBM SPSS Statistics*, desde una perspectiva integrada del diseño de la investigación y el análisis de datos. Está dirigido fundamentalmente a estudiantes del Grado en Psicología y ciencias afines (Educación, Trabajo Social, Ciencias Sociales y de la Salud), en los que la perspectiva práctica y aplicada de la investigación supone el eje principal de su formación en metodología. Adicionalmente, también puede ser de interés para estudios de postgrado de estas mismas disciplinas, como material de nivelación o repaso, o como material introductorio a la metodología de investigación.

A lo largo de sus seis capítulos, el texto combina los fundamentos teóricos necesarios, con supuestos prácticos aplicados en diferentes áreas de la Psicología, enlaces a vídeos demostrativos con opciones de subtítulos, y modelos de informes de resultados siguiendo el formato APA. Se cierra con un apartado final de bibliografía complementaria y recursos web.

A continuación, se presenta un resumen de los seis capítulos que componen la obra:

- **Capítulo 1: Introducción al SPSS. Análisis iniciales y exploratorios de datos.** Se centra en el manejo básico del programa (entorno, tipos de archivos y edición de datos) y en las etapas iniciales de depuración y transformación de la información. Incluye procedimientos de estadística descriptiva como frecuencias, exploración de datos, tablas de contingencia y representaciones gráficas fundamentales.
- **Capítulo 2: Análisis de datos en diseños con dos muestras independientes y relacionadas.** Aborda los diseños de investigación que comparan dos grupos, ya sean independientes (sujetos distintos) o relacionados (mismas personas evaluadas en dos momentos distintos). Explica tanto pruebas paramétricas (*T de Student*) como no paramétricas (*U de Mann-Whitney*, *Wilcoxon*, *Chi-cuadrado* y *McNemar*), introduciendo además el concepto de tamaño del efecto.
- **Capítulo 3: Análisis de datos en diseños con más de dos grupos independientes.** Se enfoca en los diseños unifactoriales multigrupo. El eje central es el *ANOVA* de un factor, detallando sus condiciones de aplicación y alternativas como la prueba de *Brown-Forsythe* (cuando no hay igualdad de varianzas) o la prueba de *Kruskal-Wallis* (opción no paramétrica).
- **Capítulo 4: Análisis de datos en diseños factoriales con grupos independientes.** Trata sobre investigaciones donde se estudian dos o más variables independientes de forma simultánea. Explica cómo analizar los efectos principales de cada factor y, sobre todo, el efecto de interacción, que permite determinar si el impacto de una variable depende de los niveles de la otra.
- **Capítulo 5: Análisis de datos en diseños de investigación de 1 Factor con medidas repetidas.** Analiza situaciones donde un mismo grupo es evaluado en tres o más ocasiones o condiciones. Presenta

el *ANOVA* de medidas repetidas, haciendo hincapié en el supuesto de esfericidad y sus correcciones, además de alternativas no paramétricas como la *Prueba de Friedman* y la *Q de Cochran*.

- **Capítulo 6: Correlación, dependencia y predicción entre dos variables.** Se dedica al estudio del grado de relación entre dos variables y a la capacidad de predicción entre ellas. Cubre índices para distintos niveles de medida (como *Pearson*, *Spearman*, *Tau-b de Kendall*) y finaliza con la introducción al modelo de regresión lineal simple para predecir el comportamiento de una variable a partir de otra.

Este texto docente ha sido elaborado como resultado del Proyecto de Innovación Docente “Creación de material docente en formato accesible multimedia para asignaturas de metodología de investigación en Psicología, Ciencias Sociales y de la Salud” con código 24\_25\_1\_81C, coordinado por la profesora M<sup>a</sup> Leticia de la Fuente Sánchez, concedido por el Vicerrectorado de Grados e Innovación Docente de la Universidad de Almería, en la Convocatoria de Proyectos de Innovación Docente (PID). Bienio 2024-2025.

## Capítulo 1

### Introducción al SPSS. Análisis iniciales y exploratorios de los datos

1. El paquete estadístico SPSS ([vídeo Introducción SPSS](#))
  - 1.1. Acceso y salida del SPSS
  - 1.2. Ayudas y guías en SPSS
  - 1.3. Etapas para analizar los datos
    - 1.3.1. El editor de datos
    - 1.3.2. Seleccionar el procedimiento
    - 1.3.3. Seleccionar variables
    - 1.3.4. Evaluación de resultados
  - 1.4. Tipos de archivos
  - 1.5. Supuesto práctico 1 ([vídeo supuesto práctico 1](#))
2. Análisis iniciales y exploratorios de los datos ([vídeo Estadística descriptiva](#))
  - 2.1. Depuración de datos
  - 2.2. Transformación de datos
  - 2.3. Análisis descriptivos-exploratorios de datos.
    - 2.3.1. El procedimiento Frecuencias
    - 2.3.2. El Procedimiento Explorar
    - 2.3.3. Tablas de contingencia
    - 2.3.4. Gráficos
      - 2.3.4.1. Representación gráfica de una variable
      - 2.3.4.2. Gráficos para la representación conjunta de variables
  - 2.4. Supuesto práctico ([vídeo supuesto práctico 2](#))
  - 2.5. Apéndice I. Notación científica
  - 2.6. Apéndice II. Propiedades de la distribución de frecuencias y evaluación de índices

## 1. El paquete estadístico SPSS

La primera versión del paquete estadístico SPSS (Statistical Package for the Social Science) fue creada en los años 60 por Norman H. Nie, C. Hadlay y Dale Bent. A partir de ahí hasta nuestros días se han ido desarrollando múltiples versiones, actualmente el programa ha pasado a llamarse IBM SPSS Statistics. La versión con la que trabajaremos está implementada tanto para entornos Windows como Mac. El programa es capaz de leer unos datos, manipularlos, realizar distintos análisis estadísticos y presentar los resultados obtenidos. Una ventaja de este programa es que puede utilizarse en diferentes áreas de trabajo: encuestas, marketing y análisis de ventas, mejora de la calidad y, por supuesto, en los diferentes campos en los que se realice investigación aplicada.

La versión de IBM SPSS Statistics para Windows contiene los procedimientos estadísticos más conocidos para el análisis de datos básico. Entre estos procedimientos se encuentran: tablas de frecuencias, tablas de contingencia, estadísticos descriptivos, análisis factorial, correlación, regresión, pruebas no paramétricas, ANOVA, etc. Además, existen nuevos procedimientos estadísticos agregados en módulos complementarios para mejorar el sistema de base de IBM SPSS Statistics. Es un programa que se actualiza regularmente, presentando versiones actualizadas en las que se incluyen nuevos análisis de datos.

### [\(Vídeo Introducción SPSS\)](#)

#### 1.1. Acceso y salida de IBM SPSS Statistics

Comenzaremos con lo primero que se necesita conocer, cómo entrar y salir del programa. Para acceder al programa, el usuario deberá hacer clic con el botón derecho del ratón en el icono del programa, dentro del menú de programas de su ordenador, provocando que el programa se active y aparezca la pantalla de bienvenida con las diferentes opciones para abrir y/o crear una base de datos.

El usuario podrá crear una base de datos nueva, podrá abrir una base de datos propia ya existente y/o podrá abrir una de las bases de datos que ofrece el programa como archivos de muestra. Si queremos abrir una base de datos existente, debemos seleccionar el nombre del archivo, buscando la carpeta en la que se encuentre y posteriormente pulsando el botón **Abrir**. Para comenzar a introducir datos seleccione la opción **Nuevo conjunto de datos**, una vez seleccionada la opción haga clic con el ratón en **Abrir**.

Para salir del programa simplemente debe hacer clic sobre el menú **Archivo**, situado en la barra de menús, y elegir la opción **Salir**. Otra opción es con el procedimiento de habitual de Windows de cerrar la ventana (**pulsa X**).

En la pantalla principal (o editor de datos) se pueden ver los tres elementos básicos del programa: la barra de menús, la barra de herramientas y la barra de estado. La barra de herramientas y de estado pueden ser manipuladas desde el menú **Ver** o el botón **Personalizar Barra de herramientas** en la barra de herramientas y la barra de menús puede ser modificada desde el menú **Utilidades**.

Además, SPSS va a trabajar con diferentes tipos de ventanas que de forma genérica se pueden resumir en:

**Editor de datos.** Esta ventana muestra el contenido del archivo de datos. Con el Editor de datos puede crear nuevos archivos de datos o modificar los existentes. La ventana Editor de datos se abre automáticamente cuando se inicia una sesión de *IBM SPSS Statistics*.

**Visor.** Todas las tablas, los gráficos y los resultados estadísticos se muestran en el Visor. La ventana del Visor se abre automáticamente la primera vez que se ejecuta un procedimiento que genera resultados.

**Visor de borrador.** Los resultados pueden mostrarse como texto normal (en lugar de como tablas pivote interactivas) en el Visor de borrador.

**Editor de tablas pivote.** Con el Editor de tablas pivote es posible modificar los resultados mostrados en este tipo de tablas de diversas maneras. Puede editar el texto, intercambiar los datos de las filas y las columnas, añadir colores, crear tablas multidimensionales y ocultar y mostrar los resultados de manera selectiva.

**Editor de gráficos.** Puede modificar los gráficos y diagramas de alta resolución en las ventanas de los gráficos. Es posible cambiar los colores, seleccionar diferentes tipos de fuentes y tamaños, intercambiar los ejes horizontal y vertical, rotar diagramas de dispersión 3-D e incluso cambiar el tipo de gráfico.

**Editor de resultados de texto.** Los resultados de texto que no aparecen en las tablas pivote pueden modificarse con el Editor de resultados de texto. Puede editar los resultados y cambiar las características de las fuentes (tipo, estilo, color y tamaño).

**Editor de sintaxis.** Puede pegar las selecciones del cuadro de diálogo en una ventana de sintaxis, donde aparecerán en forma de sintaxis de comandos. A continuación, puede editar ésta sintaxis para utilizar las funciones especiales de SPSS que no se encuentran disponibles en los cuadros de diálogo. También puede guardar los comandos en un archivo para utilizarlos en sesiones de SPSS posteriores.

**Editor de procesos.** El procesamiento y la automatización OLE permiten personalizar y automatizar muchas tareas en SPSS. Utilice el Editor de procesos para crear y modificar los procesos básicos.

## **1.2. Ayudas y guías en SPSS**

*IBM SPSS Statistics* ofrece un menú específico de **Ayuda** en el que el usuario puede resolver dudas sobre los diferentes análisis de datos que se pueden realizar con el programa.

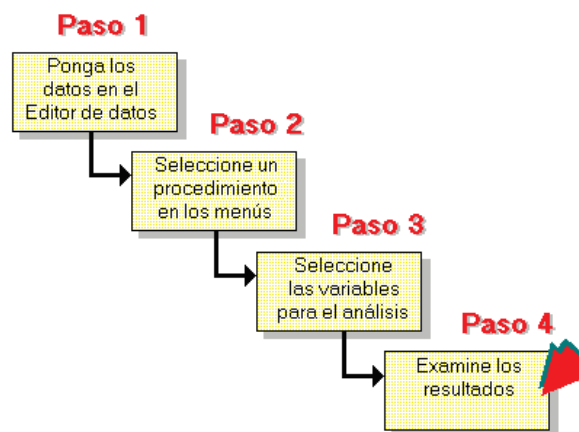
Una forma de acceder en cualquier momento a la información es desplegar el menú **Ayuda** y el usuario tendrá acceso a las diferentes opciones de Ayuda. En concreto, tendrá disponible un buscador en el que al indicar el análisis que se quiere realizar, nos indicará el itinerario para llegar a él y pinchando en la cadena indicada nos abrirá la ventana específica para solicitar dicho análisis de datos. Además de dicha opción también tendrá acceso

a diferentes documentos de SPSS referidos a los análisis que se pueden realizar con el programa, foros de dudas, sintaxis que aplicar con el programa.

El buscador de Ayuda se encuentra disponible tanto en el menú **Ayuda** como en la barra de herramientas. Además, en muchos menús y submenús de los programas seleccionados se incorpora un botón de ayuda.

### 1.3. Etapas para analizar los datos

En este programa se establecen cuatro pasos o etapas para analizar los datos:



En la primera etapa, el editor de datos, hay que introducir los datos que pretendemos analizar. O bien abriendo un archivo de datos previamente guardado con éste u otros programas. (menú **Archivo, Abrir, Datos**). Para ello debemos conocer cómo es la estructura básica de un archivo de datos, y cómo vamos a introducirlos, teniendo en cuenta el tipo de datos con los que trabajamos y que podemos designar etiquetas para los distintos valores de las variables a introducir. En la segunda etapa, debemos seleccionar un procedimiento de la barra de menús para calcular estadísticos o crear gráficos. En tercer lugar, se seleccionan las variables con las que vamos a trabajar. Por último, examinaremos los resultados de nuestros análisis en el visor. Veamos más detenidamente cómo realizar estas operaciones.

#### 1.3.1. El editor de datos

El editor de datos proporciona un método práctico para la creación y edición de archivos de datos. La ventana del editor de datos se abre automáticamente cuando se inicia una sesión o podemos hacerlo a través del menú **Archivo, Nuevo, Datos**.

Muchas funciones del Editor de datos son similares a las que se encuentran en aplicaciones de hojas de cálculo. En nuestro caso en el editor los datos debemos introducirlos siguiendo una serie de características:

1. Las filas en la matriz se denominan *casos*. Cada fila en la matriz de datos representa un caso o una observación.
2. Las columnas son *variables*. Cada columna representa una variable o una característica que se mide. Por ejemplo, cada ítem de un cuestionario será una variable.
3. Las casillas contienen *valores o modalidades*. Cada casilla contiene un valor único de una variable para cada caso.
4. El archivo de datos es rectangular. Las dimensiones del archivo de datos vienen determinadas por el número de casos y de variables.

Antes de empezar a introducir los datos, debemos definir las variables de la investigación o estudio. Para acceder a definir la variable debemos pulsar sobre el cuadro donde pone **vista de variables** en la parte inferior derecha.

Al definir la variable debemos nombrar la variable sustituyendo el nombre que el programa asigna por defecto (por ejemplo, VAR00001). A los nombres de las variables se le deben aplicar las reglas siguientes:

- El nombre debe comenzar por una letra. El resto de caracteres pueden ser letras, números, un punto o los símbolos @, #, \_, \$, ¿ o ¡.
- Los nombres de variables no pueden acabar en un punto.
- Se deben evitar nombres de variable que terminen con subrayado, aunque si lo acepta.
- La longitud del nombre no debe ser mayor de 64 caracteres.
- Los espacios en blanco y algunos caracteres especiales (por ejemplo “\*”) no se pueden utilizar. El espacio se puede simular con un guion bajo.
- Cada nombre será único, no se permiten duplicados. Los nombres no distinguen entre mayúsculas y minúsculas.

Al definir la variable podemos seleccionar la opción **Tipo** que permite definir el tipo de variable en función de los datos que incluya. En concreto, los tipos de variable a elegir son los siguientes: Numérico, Coma, Punto, Notación científica, Fecha, Dólar, Moneda personalizada, Cadena y Numérico restringido (entero con ceros iniciales). El contenido del cuadro de diálogo **Definir tipo** depende del tipo de datos seleccionado. Para algunos tipos de datos, hay cuadros de texto para el ancho y el número de decimales; para otros, simplemente puede seleccionar un formato de una lista desplegable de ejemplos.

Dependiendo del formato, la presentación de valores en el **Editor de datos** puede ser diferente, según las siguientes normas generales:

- Para formatos numéricos, de coma y de punto, se pueden introducir valores con cualquier número de dígitos decimales (hasta 16) y el valor completo se almacena internamente. El editor muestra sólo el número

definido de dígitos decimales y redondea los valores con más decimales. Sin embargo, el valor completo se utiliza en los cálculos.

- Para variables de cadena, todos los valores se rellenan por la derecha hasta el máximo de ancho.
- Para formatos de fecha, se pueden utilizar guiones, barras, espacios, comas o puntos como separadores entre los valores de día, mes y año
- Para formatos de hora, puede utilizar punto y coma, puntos o espacios como separadores de horas minutos y segundos.

Después puede seleccionar *etiquetas* para la variable, escribiendo una descripción más amplia de dicha variable. Y a continuación puede definir las etiquetas descriptivas para los distintos valores que puede tener una variable, seleccionando en la celda concreta el cuadro con ..., y genera la ventana de la **Etiquetas de Valor**. Esas etiquetas se emplean en los gráficos y las salidas del programa. Así, por ejemplo, podría asignar las etiquetas ‘Hombre’ y ‘Mujer’ a los valores numéricos 1 y 2. Para ello escriba el valor y la etiqueta descriptiva y pulse **Añadir**. Al finalizar todos simplemente pulsar **Aceptar**.

A continuación, tiene la opción de definición de los valores perdidos que se activa igual que el anterior, pinchando en el extremo derecho de la celda de valores perdidos de la variable seleccionada. Las siguientes opciones se centran en el número de caracteres que queremos tener visibles en las celdas (columnas), la alineación de los datos en la celda y el nivel de medida de la variable, **nominal, ordinal o de escala** (intervalo o razón).

Finalmente podemos definir el **Rol** de la variable dentro de la base de datos. A continuación, se indican los posibles roles y su significado.

- Entrada: La variable se utiliza como predictor o variable independiente. Por defecto, la mayoría de las variables nuevas se establecen con este rol.
- Objetivo: La variable se utiliza como variable dependiente o de salida.
- Ambos: La variable se usará tanto como entrada como objetivo.
- Ninguno: Se omite el rol de la variable en los análisis.
- Partición: Se usa para dividir los datos en muestras diferentes.
- División: Este rol se incluye por compatibilidad con IBM® SPSS® Modeler; las variables con este rol no se usan en IBM® SPSS® Statistics para segmentar datos.

Después de definir las opciones más comunes de las **variables** puede comenzar a **introducir datos**. Se pueden introducir datos en cualquier orden. Asimismo, se pueden introducir datos por caso o por variable, para áreas seleccionadas o para casillas individuales.

- La casilla que está activa se resalta con borde grueso



- El nombre de la variable y el número de fila de la casilla activa se muestran en la esquina superior izquierda del Editor de datos.
- Cuando seleccione una casilla e introduzca un valor de datos, el valor se muestra en el editor de casillas situado en la parte superior del Editor de datos.
- Los valores de datos no se registran hasta que pulse Intro o seleccione otra casilla.
- Para introducir datos distintos de los numéricos, en primer lugar, se debe definir el tipo de variable.

Igualmente, con el Editor de datos es posible modificar un archivo de datos de muchas maneras: cambiar valores de datos, cortar, copiar y pegar valores de datos, añadir y eliminar casos, añadir y eliminar variables, cambiar el orden de las variables y cambiar las definiciones de variables.

Por último, antes de seleccionar un procedimiento debemos guardar el archivo de datos que hemos creado. Para ello seleccione la opción **Guardar como** en el menú **Archivo** y proporcione un nombre con el que denominar al recién creado archivo de datos.

### **1.3.2.- Seleccionar el procedimiento**

Una vez que se han introducido los datos, la segunda etapa es seleccionar el procedimiento (o análisis de datos) que queremos para nuestros datos, bien eligiendo una opción del menú **Analizar** o bien del menú **Gráficos**.

El procedimiento que elegiremos depende de nuestros intereses de análisis. Debemos seleccionar un procedimiento estadístico o gráfico adecuado a los objetivos que pretendemos cubrir.

### **1.3.3.- Seleccionar variables**

Posteriormente a la selección del análisis, debemos elegir qué variables son las que van a ser analizadas y definir las opciones, si las hubiera, de selección de variables que en todos los casos depende del tipo de procedimiento elegido.

El cuadro anterior, es el que obtenemos al seleccionar en el menú **Estadística** la opción **Resumir** y después **Frecuencias**. En el cuadro de la izquierda aparecen todas las variables de nuestro estudio y debemos seleccionarlas trasladándolas al cuadro de la derecha haciendo clic sobre las variables con el ratón y posteriormente pinchando en la flecha que hay entre ambos cuadros. Una vez seleccionadas las variables, se pincha sobre el botón aceptar y el programa ejecuta los resultados.

### **1.3.4.- Evaluación de resultados**

La última etapa al analizar los datos es la de evaluación de los resultados. En función del análisis estadístico y de los gráficos, así como de las variables seleccionadas obtendremos una salida diferente. En general, cuando se ejecuta un procedimiento, los resultados se muestran en una ventana llamada visor. Desde esta ventana puede desplazarse con facilidad a cualquier parte de los resultados que desee ver. También puede

modificar los resultados y crear un documento que contenga exactamente los resultados que desee, organizados y con formatos adecuados.

Los resultados se muestran en el visor, donde podemos examinar los resultados, mostrar u ocultar tablas y gráficos seleccionados, cambiar el orden de presentación de los resultados moviendo los elementos seleccionados y también mover los elementos entre el visor y otras aplicaciones.

El visor está dividido en dos paneles. El panel de la izquierda, denominado **Panel de Titulares** y el de la derecha, **Panel de Contenido**. El panel de titulares proporciona una tabla de contenidos del documento del visor. Utilice este panel para navegar por los resultados y controlar su presentación. La mayoría de las acciones en dicho panel tienen su efecto correspondiente en el panel de contenidos:

- Si selecciona un elemento en el panel de titulares (pinchándolo con el ratón) se seleccionará y mostrará el elemento correspondiente en el panel de contenidos.
- Si mueve un elemento en el panel de titulares (pinchando y arrastrando con el ratón), también se moverá el elemento correspondiente en el panel de contenidos.
- Si contrae la vista de titulares (pulsando dos veces en el icono del libro) se ocultarán los resultados de los elementos en los niveles contraídos. Puede expandirlos de nuevo haciendo doble clic.
- Si suprime un elemento en el visor (seleccionándolo y pulsando la tecla **Supr**) desaparecerá del panel de contenido.

Los resultados, una vez que los ha diseñado de la forma deseada puede guardarlos en un fichero seleccionando la opción **Guardar** del menú **Archivo**. Escriba el nombre que desee dar al archivo y seleccione la opción Guardar. Si desea guardar los resultados en otros formatos externos (por ejemplo, HTML o texto), utilice el comando **Exportar** del menú **Archivo**. Finalmente, también puede optar por **Imprimir** los contenidos seleccionados o parte de ellos.

#### **1.4. Tipos de archivos**

El programa SPSS lee diferentes formatos de archivos, entre los que destacamos entre los más habituales los siguientes: **SPSS Statistics (\*.sav, \*.zsav)**, **SPSS/PC+ (\*.sys)**, **Excel (\*.xls, \*.xlsx, \*.xlsm)**, **CSV (\*.csv)**, **Texto (\*.txt, \*.dat, \*.csv, \*.tab)**, **SAS (\*.sas7bdat, \*.sd7, \*.sd2, \*.ssd01, \*.ssd04, \*.xpt)** y **Stata (\*.dta)**.

Si queremos importar datos desde una base de datos, podemos seleccionar la ruta Archivo, Importar datos, Base de datos, Nueva consulta, y aparecerá un asistente para guiarnos en la importación.

#### **1.5 Supuesto Práctico**

En este supuesto práctico el objetivo es la creación de un fichero de datos en IBM SPSS Statistics, introduciendo los datos de las variables de forma manual. Para facilitar la interpretación del fichero de datos se añadirán etiquetas a cada variable y para la variable nominal se añadirán etiquetas de valor para poder interpretar con claridad las categorías.

Lea el siguiente enunciado e introduzca los datos de las variables del ejemplo siguiendo las instrucciones proporcionadas:

“Como parte de las actuaciones dirigidas a reducir el gasto en un centro sanitario, el equipo de dirección decide evaluar la eficacia de determinados servicios que se prestan actualmente en dicho centro, y eliminar aquellos que no muestren una efectividad contrastada. Uno de los servicios que se someten a evaluación son los cursos de preparación al parto. Como parte del equipo de salud mental encargado de dichos cursos, usted mantiene la hipótesis de que las mujeres que asisten a tales cursos manifiestan un menor nivel de ansiedad ante el parto respecto de aquellas que no asisten, lo que favorece a su vez una actitud positiva y colaboradora ante la situación del parto y aumenta la eficacia global de los procedimientos. A continuación, se presentan las puntuaciones en ansiedad ante el parto de 20 participantes, la mitad de las cuales recibieron el curso”.

No reciben curso	7	13	10	7	8	13	8	10	15	15
Reciben curso	5	2	1	3	1	2	3	1	5	1

La resolución del supuesto práctico se encuentra disponible en el vídeo [Supuesto práctico 1](#).

## 2. Análisis iniciales y exploratorios de los datos

A continuación, vamos a tratar sobre aquellos aspectos que se relacionan con un primer nivel de análisis de nuestros datos, que van desde la evaluación de la calidad de nuestros datos, hasta un primer nivel de descripción de los mismos, pasando por determinados aspectos que se refieren a la manipulación de nuestros datos originales. En este capítulo se utilizará el fichero de datos: [Actitud Ley Tabaco.sav](#). El vídeo con el desarrollo de todos los procedimientos del punto 2 es el siguiente: [vídeo Estadística descriptiva](#)

### 2.1. Depuración de datos

Consiste en la inspección de la información contenida en las respuestas de los sujetos o en los campos de una base de datos con el fin de comprobar su consistencia, veracidad o actualidad (Santesmases, 1997). Teniendo en cuenta que la primera técnica de depuración de datos es la propia prevención de los errores comentada en el tema introductorio.

1. *Detección de errores en los datos en la fase de codificación e introducción de datos*, se pueden plantear algunas técnicas como:

- *Control de grabación*. Una técnica es la conocida como “Ciega”, que hace que dos personas introduzcan los datos. Otra posible estrategia es la grabación inteligente (se detectan los errores mientras se graba).
- *Comprobación de códigos y rangos*. Consiste en ver si existen valores de las variables que se establecen fuera del rango de valores permitido. Para ello basta con estudiar la distribución de frecuencias de las

variables introducidas. En *IBM SPSS Statistics* la ruta a seguir sería: Analizar, Estadísticos Descriptivos, Frecuencias.

En la tabla que presenta la distribución de frecuencias para la/s variable/s se puede comprobar si hay algún valor de la/s variable/s que pueda ser un error.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Hombre	18	60,0	60,0	60,0
	Mujer	11	36,7	36,7	96,7
	11	1	3,3	3,3	100,0
	Total	30	100,0	100,0	

Una vez identificado el valor o valores erróneos, podemos localizarlos y, en su caso, modificarlo o indicarlo como un valor perdido (missing). Para ello nos situamos en la variable correspondiente y nos vamos a *Edición, Buscar* y le indicamos el valor a buscar.

- *Comprobación de inconsistencias lógicas.*

A veces en muchas investigaciones tenemos variables que están condicionadas entre sí. Por ejemplo, en encuestas donde se nos dice, si ha respondido afirmativamente conteste la pregunta siguiente. En este caso para comprobar esas inconsistencias, se pueden estudiar las tablas de frecuencias individuales o mejor estudiar la distribución de frecuencias conjunta.

Para esta última opción utilizaremos el procedimiento **Analizar, Estadísticos Descriptivos, Tablas cruzadas**. Analizaremos si las distribuciones de frecuencias conjuntas son congruentes y en su caso realizaremos las comprobaciones pertinentes para subsanar el error.

		¿Ha consumido tabaco alguna vez?		Total
		Sí	No	
¿Consume tabaco?	Sí	17	0	17
	No	3	10	13
Total		20	10	30

## 2. Detección de los datos perdidos (*missing*).

Una vez detectados y corregidos los errores, puede ser que tengamos una serie de datos erróneos, que no hemos podido corregir, y sobre los que tendremos que tomar decisiones. La primera operación es cuantificar ese error. Una regla empírica ampliamente contrastada es esperar entre un 2% y un 5% de observaciones con errores (Peña, 2001, p.31). Entre ellos se incluyen los comentados en el apartado anterior y que ya habríamos corregido. Lo deseable es calcular el tanto por ciento de error cometido, aunque no siempre es posible realizarlo dado el número de variables y casos, en este caso se estima dicho error, calculando el error en una muestra de nuestros datos. Nosotros nos centraremos en la forma de localizar todos los datos erróneos una vez corregidos los datos detectados en la fase anterior, que serán considerados como datos faltantes, perdidos o ausentes (*missing*), siendo su evaluación y tratamiento clave para una correcta utilización e interpretación de los análisis de datos que realicemos. En líneas generales buscaremos si los datos perdidos responden a un patrón aleatorio o tienen algún elemento sistemático.

### 1. Detección y diagnóstico.

Para detectar los valores perdidos, podemos proceder como se ha comentado anteriormente sobre la distribución de frecuencias, para la detección de errores. Para ver como los errores se distribuyen por los casos, se puede trasponer la matriz de casos x variables, en una de variables x casos, y analizar las distribuciones de frecuencias de los casos. El procedimiento sería **Datos, Trasponer**<sup>1</sup>. Con este procedimiento los casos pueden ser tratados como variables y se puede estudiar si existen o no algún tipo de patrón en los datos perdidos, mediante análisis exploratorios como los que introduciremos más adelante.

Otra forma de estudiar este fenómeno, es analizando si existen diferencias estadísticamente significativas, entre los sujetos con valores perdidos y los que no tienen valores perdidos, en alguna variable de interés. Para ello se recodificaría la variable y se aplicaría la prueba estadística correspondiente. Por el mismo procedimiento de recodificación, se puede estudiar las correlaciones entre distintas variables con datos ausentes. Tanto si encontramos diferencias estadísticamente significativas, como correlaciones de cierta magnitud, se puede concluir que el patrón de datos perdidos no es aleatorio.

2. *Tratamiento de datos perdidos.* En términos generales y en muchos procedimientos, IBM SPSS Statistics contempla dos formas de tratar los datos perdidos, una es eliminando todos los casos con algún dato perdido (*listwise*) y otra es eliminando los casos que están implicados en algún análisis concreto por pares de variables (*pairwise*).

Cuando tenemos muy pocos datos perdidos, puede ser razonable eliminar esos casos, pero muchas veces nos interesa mantener esos datos y puede ser útil una técnica de imputación. El paquete IBM SPSS Statistics dispone de las siguientes:

Media de la serie. Sustituye los valores perdidos con la media de la serie completa.

---

<sup>1</sup> Es conveniente guardar el fichero original antes de ejecutar esta acción.

Media de puntos adyacentes. Sustituye los valores perdidos por la media de los valores válidos circundantes. La amplitud de los puntos adyacentes es el número de valores válidos, por encima y por debajo del valor perdido, utilizados para calcular la media.

Mediana de puntos adyacentes. Sustituye los valores perdidos por la mediana de los valores válidos circundantes. La amplitud de los puntos adyacentes es el número de valores válidos, por encima y por debajo del valor perdido, utilizados para calcular la mediana.

Interpolación lineal. Sustituye los valores perdidos utilizando una interpolación lineal. Se utilizan para la interpolación el último valor válido antes del valor perdido y el primer valor válido después del valor perdido. Si el primer o el último caso de la serie tiene un valor perdido, el valor perdido no se sustituye.

Tendencia lineal en el punto. Reemplaza los valores perdidos de la serie por la tendencia lineal en ese punto. Se hace una regresión de la serie existente sobre una variable índice escalada de 1 a n. Los valores perdidos se sustituyen con sus valores pronosticados.

El procedimiento a seguir es **Transformar, Reemplazar valores perdidos**, e introduciremos la variable o variables correspondientes, y elegimos el método de imputación. La salida genera una nueva variable con los datos imputados.

### 3. Detección de datos alejados o aislados (*outliers*).

En determinadas ocasiones existen datos que se consideran que no son representativos de la muestra considerada, y por tanto pueden adulterar los resultados de nuestros análisis. Para ello se analizan las puntuaciones que pueden ser consideradas como alejadas y extremas, para ello una de las herramientas más útiles es el diagrama de caja, en el que profundizaremos más adelante, pero que para nuestros intereses en este punto se ejecuta en **Gráficos, Diagrama de cajas**.

## 2.2. Transformación de datos

Esta fase es considerada por algunos autores como un primer nivel de análisis estadístico (Aparicio, 1991). En este contexto del tratamiento de datos son fundamentales las cuestiones en torno a las escalas de medida, sus transformaciones admisibles y la problemática que ha generado. Los tipos de transformaciones más usuales son:

### 1. Transformaciones lógicas.

Responde a transformaciones del tipo asignar un valor si satisface alguna condición concreta de tipo *Si ..., entonces*. El procedimiento en SPSS es **Transformar, Recodificar**, aquí nos permite elegir entre, **En las mismas variables o En distintas**. Nuestro criterio personal es seleccionar siempre en distintas, con lo que se crea una variable nueva, para así conservar la variable original, a menos que trabajemos con distintos ficheros, y en uno tengamos las variables originales y en otro las recodificaciones. Se introduce la variable y se nombra y etiqueta la nueva variable, se pulsa el botón de *cambiar* y se indican los valores a recodificar.

## 2. Transformaciones aritméticas.

Las más frecuentes son la creación de variables por adición o promedio, recodificación, inversión y estandarización. Habrá que tener en cuenta la transformación admisible según el nivel de medida (nominal, ordinal, etc) que supongamos a las variables que estemos utilizando. Aquí también se incluyen aquellos aspectos de transformación de variables conducentes a cumplir determinados supuestos de técnicas estadísticas concretas y que comentaremos más adelante, en la línea de lo que el Análisis Exploratorio de Datos entiende por transformación. Todos estos procedimientos se establecen con **Transformar**, **Recodificar** o **Transformar, Calcular variable**.

IBM SPSS Statistics posee diferentes procedimientos de manipulación de datos y transformación de los mismos, que van desde la ordenación de datos, la selección, recodificación, creación de variables, y que tienen relación con lo comentado en los anteriores apartados.

### 2.3. Análisis descriptivos-exploratorios de datos

Dentro del programa IBM SPSS Statistics existen varios procedimientos con los que abordar un estudio descriptivo de los datos. Hay diversas opciones que debemos conocer para establecer una primera aproximación a los datos. Los procedimientos fundamentales para esta descripción se encuentran en la opción **Estadísticos Descriptivos** del menú **Analizar**, algunos de ellos ya los hemos mencionado:

- Frecuencias
- Descriptivos
- Explorar
- Tablas cruzadas

El procedimiento **Descriptivos** aporta la misma información que **Frecuencias** y **Explorar**, que son más genéricos, por lo que obviaremos su presentación, salvo en la única función adicional que posee que es generar puntuaciones típicas, es decir convertir las variables originales a puntuaciones  $Z = \frac{X - \bar{X}}{S_x}$ .

Comencemos por la caracterización de la distribución de frecuencias, tanto desde el punto de vista de la tabla de frecuencias como de sus estadísticos descriptivos.

#### 2.3.1. El procedimiento Frecuencias

Proporciona tablas, estadísticos y representaciones gráficas que resultan útiles para describir muchos tipos de variables. Es un buen procedimiento para una inspección inicial de los datos.

La tabla de distribución de frecuencias ofrece información sobre Frecuencia, Porcentaje, Porcentaje válido y Porcentaje acumulado. Frecuencia, es la frecuencia absoluta (número de veces que se repite un valor concreto) de cada nivel de la variable. Porcentaje es la proporción de sujetos multiplicado por cien (frecuencia relativa x 100). Porcentaje válido, eliminado los valores perdidos, y el acumulado de los valores inferiores.

En el procedimiento **Frecuencias** podemos establecer una serie de opciones. Así, si en la ventana de Frecuencias elegimos el botón de **Estadísticos** podemos elegir un conjunto de ellos sobre las variables

seleccionadas (valores percentiles, estadísticos de tendencia central, de posición, de dispersión o sobre la forma de la distribución). Igualmente, en el procedimiento frecuencias existe una opción para dibujar gráficos, se puede elegir entre: gráfico de barras, gráfico de sectores e histograma (con/sin curva normal),

Además, en el procedimiento **Frecuencias** podemos abrir un cuadro de diálogo para establecer las opciones de formato, que son tres:

- Ordenar por: la tabla de frecuencias se puede ordenar respecto a los valores o respecto a la frecuencia de aparición. Por defecto muestra los datos por valores de forma ascendente.
- Múltiples variables: Si desea generar tablas de estadísticos para múltiples variables podrá mostrarlas, bien por separado, bien conjuntamente.
- Suprimir tablas con más de  $n$  categorías: esta opción impide que se muestren tablas que contengan más valores que el número especificado.

### 2.3.2. El Procedimiento Explorar

Explorar genera estadísticos de resumen y representaciones gráficas, bien para todos los casos o bien de forma separada para grupos de casos. Por ejemplo, si nosotros quisiéramos saber las diferencias de edad para hombres y mujeres de la muestra con la que venimos trabajando este sería el procedimiento adecuado.

En el cuadro de **Dependientes**, se deben incluir las variables a las que pretendemos calcular los estadísticos. En el cuadro **Factores** debemos incluir las variables que utilizaremos para definir los grupos de casos. Por ejemplo, ahora el procedimiento calculará los estadísticos separadamente para los grupos realizados en la variable género, esto es, hombres y mujeres. En el cuadro **Etiquetar los casos mediante** se pueden etiquetar los casos con sus valores en una variable, como puede ser una variable identificadora de caso. Por defecto, los casos se identifican mediante su número secuencial dentro del archivo de datos. Además, la salida proporciona gráficos, de tallo y hojas (stem and leaf) y de caja (box) para cada grupo.

Dentro del cuadro de diálogo Explorar hay varios botones con los que podemos acceder a nuevas ventanas.

Pinchando en el botón **Estadísticos** nos aparecen varias opciones:

- **Descriptivos:** Por defecto se muestran los estadísticos que han aparecido en la salida anterior: media, intervalo de confianza al 95%, media recortada, mediana, varianza, desviación típica, mínimo y máximo, rango, amplitud intercuartil, asimetría y curtosis.
- **Estimadores robustos centrales:** Permite calcular alternativas robustas a la mediana y a la media muestral para calcular el centro de localización. Se muestran los siguientes: estimador  $M$  de *Huber*, estimador en onda de *Andrew*, estimador  $M$  redescendente de *Hampel* y estimador bponderado de *Tukey*.
- **Valores atípicos:** Muestra los cinco valores mayores y los cinco menores, con las etiquetas de caso.
- **Percentiles:** Muestra los valores de los percentiles.

En el botón **Gráficos** también existen varias opciones:

- **Diagramas de caja:** Estas alternativas controlan la presentación de los diagramas de caja cuando existe más de una variable dependiente. Niveles de los factores juntos genera una presentación para cada variable dependiente, en cada una se muestran diagramas de caja para cada uno de los grupos definidos por una variable de factor. *Dependientes juntas* genera una presentación para cada grupo definido por una variable de factor, en cada una se muestran juntos los diagramas de caja de cada variable dependiente.
- **Descriptivos:** permite seleccionar gráficos de tallo y hojas e histogramas
- **Gráficos con pruebas de normalidad:** Muestra los diagramas de probabilidad normal y de probabilidad sin tendencia. Se muestra el estadístico de *Kolmogorov-Smirnov* con un nivel de significación de *Lilliefors* para contrastar la normalidad. El estadístico de *Shapiro-Wilk* se calcula para las muestras con 50 o menos observaciones.
- **Dispersión por nivel con prueba de Levene:** Controla la transformación de los datos para los diagramas de dispersión por nivel. Para todos los diagramas de dispersión por nivel se muestra la pendiente de la línea de regresión y las pruebas robustas de Levene sobre la homogeneidad de la varianza. Si selecciona una transformación, las pruebas de Levene se basarán en los datos transformados. Si no selecciona ninguna variable de factor, no se generará ningún diagrama de dispersión por nivel.

Por último, en el procedimiento Explorar puede seleccionarse el botón **Opciones**. En esta ventana podemos ver **Excluir casos según la lista** (listwise); **Excluir los casos según pareja** (pairwise); **Mostrar valores:** los valores perdidos para las variables de factor se tratan como una categoría diferente. Todos los resultados se generan para esa categoría adicional. Estos procedimientos se han comentado en el apartado del tratamiento de datos perdidos.

### 2.3.3. Tablas de contingencia

El procedimiento *tablas de contingencia* crea tablas de clasificación múltiple y además proporciona pruebas y medidas de asociación para tablas de doble clasificación. La estructura de la tabla y el hecho de que las categorías estén ordenadas o no, determinan las pruebas o medidas que se utilizan. Los estadísticos de tablas de contingencia y las medidas de asociación sólo se calculan para las tablas de doble clasificación. Si se especifica una fila, una columna y un factor de capa (variable de control), el procedimiento tablas de contingencia crea un panel de medidas y estadísticos asociados para cada valor del factor capa.

El procedimiento Tablas de contingencia tiene tres posibilidades. Como siempre para acceder a ellas debemos pulsar los botones de la parte inferior de la ventana de Tablas de contingencia. La primera opción es **Estadísticos**. Esta opción proporciona distintos estadísticos que hemos de seleccionar en función del tipo de variables con las que estamos trabajando (dicotómicas, nominales, ordinales, etc.).

Además, el procedimiento permite acceder a la ventana **Casillas**. Así permite mostrar las frecuencias esperadas y tres tipos de residuos que miden la diferencia entre las frecuencias observadas y las esperadas. Cada casilla de la tabla puede contener cualquier combinación de recuentos, porcentajes y residuos seleccionados. Por último, la opción **Formato** permite, como en otros procedimientos del programa, elegir las filas en orden ascendente o descendente.

Si repitiéramos el análisis anterior de sexo x consumo de alcohol, pero ahora marcando en **Casillas** porcentajes de fila, columna y total. Lo que nos llevaría a interpretar diferentes porcentajes en función del total de fila, de columna y total.

Una opción interesante del SPSS, es que nos permite reproducir tablas de contingencia sin necesidad de disponer de los datos originales. Para ello hay que pinchar en **Datos, Ponderar casos**, e indicar la variable de ponderación.

### 2.3.4. Gráficos

En SPSS existen multitud de tipos de gráficos, comentaremos los gráficos más adecuados para las variables consideradas categóricas y continuas. Muchos procedimientos incorporan los gráficos más adecuados a los mismos. No obstante, en SPSS existe el modulo **Gráficos**, donde se incorporan los distintos gráfico y opciones del programa. Tomaremos la clasificación de las variables desde el punto de vista estadístico como guía:

#### 2.3.4.1. Representación gráfica de una variable

a) Variables Cualitativas.

Gráfico de columnas, rectángulos o barras, representa una serie de valores utilizando barras verticales. La altura representa la frecuencia, en el ejemplo la absoluta, y en este caso las columnas suelen ser de igual área, pero de forma arbitraria. **Gráficos, Barras**, y después *Definir* El gráfico representa la variable sexo, y es la que introducimos en eje de categorías.

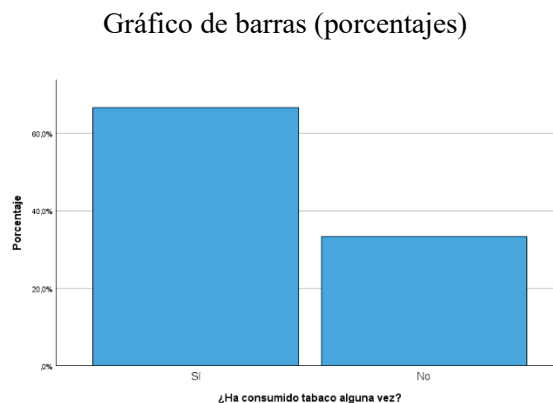
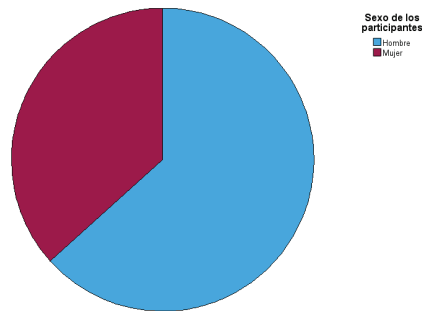


Diagrama de sectores o circular, representa una única serie de datos dónde cada valor de la serie se muestra como un sector del círculo.



b) Variables Ordinales o Cuasi-cuantitativas.

A diferencia del tipo de variable anterior aquí se establecen relaciones de orden. Por tanto, aquí podemos introducir gráficos con frecuencia acumulada, por lo demás son similares a los anteriores. Así podríamos representar gráficos de barras, columnas, sectores, y pictogramas.

Diagrama de rectángulos o columnas acumulado. Del siguiente gráfico de columnas generado a partir de la variable posición que ha obtenido un alumno al ser evaluado en diferentes asignaturas. Surge el siguiente diagrama de columnas acumulado, en los cuales se comienza por la categoría considerada inferior. Se procedería pinchando en *acumulado*, del menú **Gráficos, Barras**.

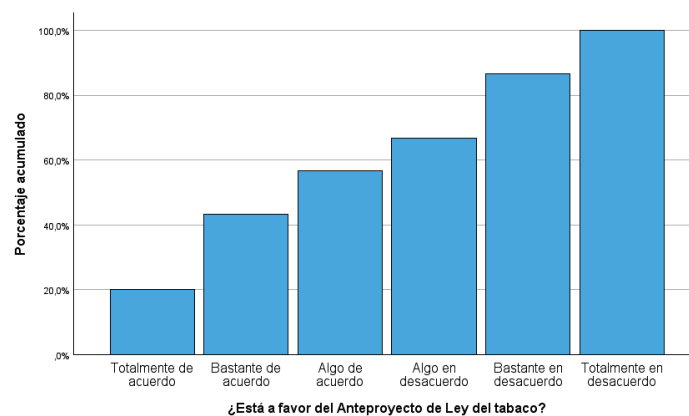


Diagrama de tallo y hojas, que hemos visto en el procedimiento explorar (Freixa et al., 1992), se construye del modo siguiente:

- Se elige el intervalo de unidades para representar el tronco, teniendo en cuenta que hay que cubrir todos los datos. Se ha de indicar en algún lugar en el diagrama, la unidad que se utiliza en el tronco para su fácil comprensión.
- Las hojas del diagrama se escriben separadas del tronco por una línea vertical y corresponden a las unidades de cada uno de los datos que se han obtenido.

- Si las hojas se han escrito a la derecha del tronco, a la izquierda, y separadas por una línea vertical, se escriben las frecuencias absolutas, es decir, el número de veces que se repite cada dato.

Suponga, por ejemplo, los datos, 12, 12, 14, 15, 16, 20, 21, 23, 24, 24, 27, 34, 34, 34, 44, 45, 46, 48, 48. El tallo o tronco son los valores que representan las decenas de los datos, y a la derecha aparece cada hoja, o valor de las unidades de los mismos:

Gráfico tallo y hojas.

1	2	2	4	5	6	
2	0	1	3	4	4	7
3	4	4	4			
4	4	5	6	8	8	

### c) Variables Cuantitativas Discretas

Diagrama de barras, representa una serie de valores utilizando barras verticales. La altura representa la frecuencia, en el ejemplo se utilizan frecuencias absolutas. Se obtiene en el menú **Gráficos, Barras**.

Polígono de frecuencias, sería el resultado de unir mediante líneas las partes superiores de las columnas. Se obtiene en el menú **Gráficos, Líneas**.

### d) Variables Cuantitativas Continuas

Histograma y Polígono de frecuencias. El histograma es una sucesión de rectángulos, uno asociado a cada modalidad de la variable, de forma que la superficie de cada rectángulo sea proporcional a la frecuencia de cada modalidad. El polígono de frecuencias es el resultado de unir los puntos medios de las barras superiores de los rectángulos en un histograma con rectángulos de igual base. En el gráfico se representa un Histograma con curva normal. Se obtiene en el menú **Gráficos, Histograma**.

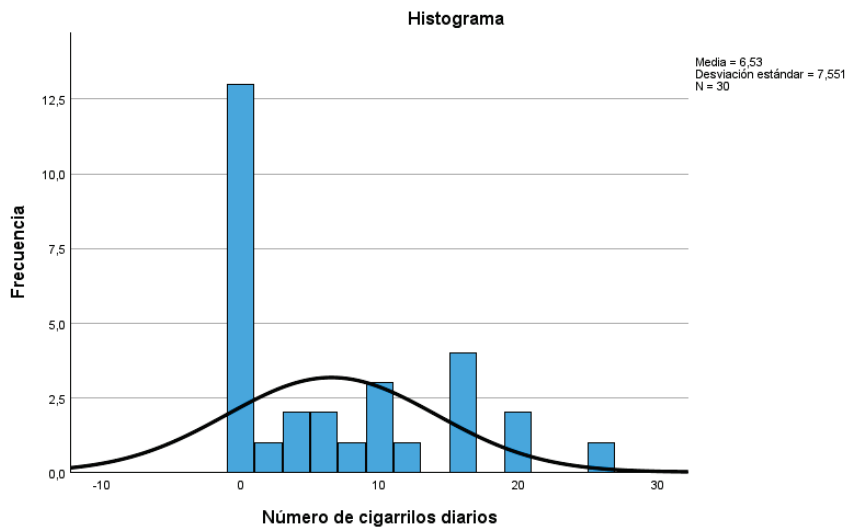


Diagrama de Caja.

Dentro del enfoque del análisis exploratorio de datos se han desarrollado una serie de gráficos que se han hecho muy populares, entre ellos está el tronco y hojas que ya vimos anteriormente en el apartado de organización de datos, ya que también puede ser considerado como una técnica de tabulación. Y por otra parte el diagrama de caja, con sus extensiones de patillas, bigotes, etc.

Para la construcción del diagrama se pueden seguir las siguientes etapas de Batista y Valls (1985):

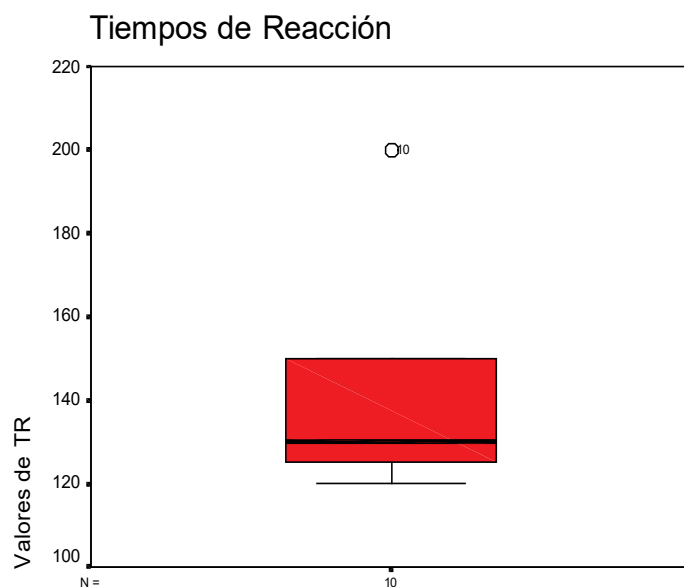
- 1.- En primer lugar, seleccionaremos una escala conveniente que cubra el recorrido de la variable.
- 2.- Trazaremos segmentos perpendiculares al eje que señalarán las posiciones de los cuartiles ( $Q_1$ ,  $Q_2$  y  $Q_3$ ).
- 3.- La caja se completará uniendo con líneas paralelas al eje, los extremos de las marcas del paso anterior.
- 4.- A partir de ambos cuartiles se trazarán líneas a trazos denominadas patillas (whiskers), que se prolongará hasta el punto más alejado que no sea anomalía (adyacente). Estableciendo los límites superior e inferior como:

$$Ls/i = Q_{3/1} \pm 1,5 ( Q_3 - Q_1 )$$

5.- Añadiremos, por último, anomalías moderadas como puntos vacíos ( $o$ ) y las extremas como puntos sombreados. Veamos como la variable tiempo de reacción podría representarse mediante este gráfico. **Gráficos,**

**Diagramas de Cajas**

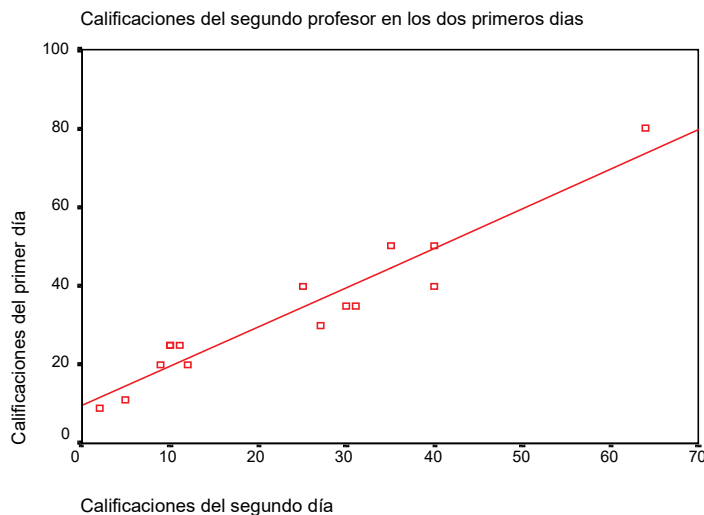
Sujeto	TR (X)
1	150
2	120
3	130
4	130
5	140
6	150
7	120
8	130
9	125
10	200
Md	130
$Q_1$	123,75
$Q_3$	150
Ls	189,38
Li	84,37



### 2.3.4.2. Gráficos para la representación conjunta de variables

#### Diagrama de dispersión.

En este gráfico se representan en cada eje una variable y los puntos representan la intersección de un valor de X con un valor de Y. Así para la variable calificaciones de un profesor en dos días de corrección, tenemos el siguiente gráfico. Nótese que hemos incluido el ajuste de la recta a esos datos. **Gráficos, Dispersión, Simple.**



## 2.4. Supuesto Práctico

El objetivo de este supuesto práctico es que el alumno sea capaz de realizar análisis iniciales con el programa IBM SPSS Statistics, tales como, cálculo e interpretación de estadísticos descriptivos, construcción de tablas de frecuencias y elaboración de gráficos básicos.

- Utilizando el procedimiento **Frecuencias** estime para la variable *Ansiedad* los siguientes estadísticos descriptivos: Media, Mediana, Moda, Varianza, Desviación típica, cuartiles, rango de la variable, el máximo y el mínimo de la variable. Además de representar gráficamente dicha variable.
- Representar gráficamente la variable *Asistencia curso*, esta vez seleccionando el **Gráfico de sectores**.
- Obtener los mismos estadísticos descriptivos de la variable *Ansiedad*, esta vez utilizando el procedimiento **Descriptivos**. Además, generar una nueva variable con las puntuaciones típicas de la variable *Ansiedad*.
- Mediante el procedimiento **Explorar** estimar estadísticos descriptivos para la variable *Ansiedad* atendiendo a los grupos formados por la variable *Asistencia al curso*.

La resolución del supuesto práctico se encuentra disponible en el vídeo [Supuesto práctico 2](#). Se ha utilizado la base de datos creada en el supuesto práctico 1, denominada [Datos Supuesto 1.sav](#)

## 2.5. Apéndice I. Notación científica

En muchas ocasiones las tablas de resultados de SPSS, al igual que otros programas o calculadoras, incluyen datos en la conocida como *notación científica*, debido al menor espacio que ocupan con respecto a la notación tradicional.

La notación científica se utiliza para señalar la posición del primer valor decimal. Por ejemplo, un número escrito en notación científica del siguiente modo: 1.225E-03 significa 0.001225 (0,001225) en notación tradicional. Del mismo modo, el número 4325+03 es igual a 4325000.

## 2.6. Apéndice II. Propiedades de la distribución de frecuencias y evaluación de índices

Podemos caracterizar la distribución de frecuencias entre sus tres propiedades: (a) *Tendencia central*. Las observaciones se concentran en la parte central de la distribución. (b) *Variabilidad*. Se detecta el grado de concentración de las observaciones alrededor del promedio. (c) *Forma*. Las dos características que resumen la forma de la distribución son: *Sesgo*. En las distribuciones que presentan simetría, el rango de puntuaciones por debajo de la media que contiene el 50% de las primeras observaciones es aproximadamente igual al rango de puntuaciones por encima de la media que contiene el 50% de las observaciones restantes. Cuando estos rangos son diferentes, las distribuciones están sesgadas (son asimétricas). *Apuntamiento o curtosis*. Estrechamente relacionada con la dispersión, la forma señala el mayor o menor apuntamiento (elevación) de la parte central de la distribución. Por tanto, una dispersión media provocará distribuciones llamadas normales; una dispersión pequeña provocará distribuciones muy apuntadas en el centro; y una dispersión alta provocará distribuciones achatadas en su parte central.

Con carácter general, el estadístico de tendencia central más adecuado para representar la distribución de frecuencias es la media aritmética, dado que en su cálculo intervienen todas y cada una de las puntuaciones de la distribución. Por otro lado, la media, como estadístico muestral es un mejor estimador de su parámetro poblacional, que la moda o la mediana lo son del suyo. Por otro lado, la mediana es preferible a la media, en los casos donde no es posible calcular la media, como es al realizar agrupaciones con intervalos abiertos. Así mismo, la mediana es preferible cuando las variables están medidas en una escala ordinal y cuando la distribución de frecuencias posee valores extremos o presenta una acentuada asimetría. Finalmente, la moda es el único índice que se puede obtener cuando la variable está medida en una escala nominal, o en agrupaciones con intervalos abiertos, cuando el intervalo mediano es el intervalo abierto. Las medidas de posición general, como los cuantiles (percentiles y cuartiles), requieren de un nivel de medida al menos ordinal, para que su utilización sea correcta.

Con respecto a los estadísticos de dispersión y con carácter general, el estadístico más adecuado para representar la distribución de frecuencias es la varianza o desviación típica (por las mismas razones que esgrimíamos para la media aritmética), dado que en su cálculo intervienen todas y cada una de las puntuaciones de la distribución y son mejores estimadores de su parámetro poblacional, que otros estadísticos de dispersión

La interpretación de este índice (siempre que trabajemos con distribuciones aproximadamente normales), donde se puede establecer que fuera del intervalo  $\bar{X} \pm k.S_x$  se encuentra al menos el  $[(1/k^2) \times 100]$  por ciento de las observaciones, sea cual sea la forma de la distribución de frecuencias, siempre que  $k > 1$ . Por otro lado, en los casos donde no es posible calcular la varianza, como es al realizar agrupaciones con intervalos abiertos, así como cuando las variables están medidas en una escala ordinal o la distribución de frecuencias posee valores extremos, es más adecuado calcular la mediana de las desviaciones absolutas o el rango semi-intercuartil.

La amplitud total es muy fácil de calcular, pero es muy sensible al tamaño de la muestra y solo tiene en cuenta los datos extremos. Finalmente, la desviación media plantea el problema de trabajar con los valores absolutos.

Si la distribución es simétrica, los índices de simetría son iguales a cero, si son menores que cero existe asimetría negativa en los datos, y a su vez, si son mayores que cero existe asimetría positiva.

Los estadísticos de apuntamiento nos indican si la distribución de los datos es más apuntada o menos que un grupo de datos con distribución normal. Una distribución se considera *mesocúrtica* si tiene el mismo apuntamiento que la distribución normal; es *leptocúrtica* si es más apuntada que la normal; y la consideraremos *platicúrtica* si es menos apuntada que la normal. Si la distribución del grupo de datos que se maneja, es una Normal, estos índices son iguales a cero, si son menores que cero la distribución es platicúrtica, a su vez, si son mayores que cero la distribución es leptocúrtica.