

# 1A. INTRODUCCIÓN A LA METODOLOGÍA DE LA INVESTIGACIÓN Y AL MANEJO DE DATOS

## INTRODUCCIÓN A LA METODOLOGÍA DE LA INVESTIGACIÓN

### Estudio sobre factores de riesgo cardiovascular

La enfermedad cardiovascular es una de las principales causas de morbilidad y de mortalidad. Su desarrollo y expresión clínica se ve favorecido por la presencia de los factores de riesgo cardiovasculares. Se quiere contrastar la hipótesis de la posible existencia de asociación entre algunos de ellos. En concreto existe interés en verificar estas dos hipótesis:

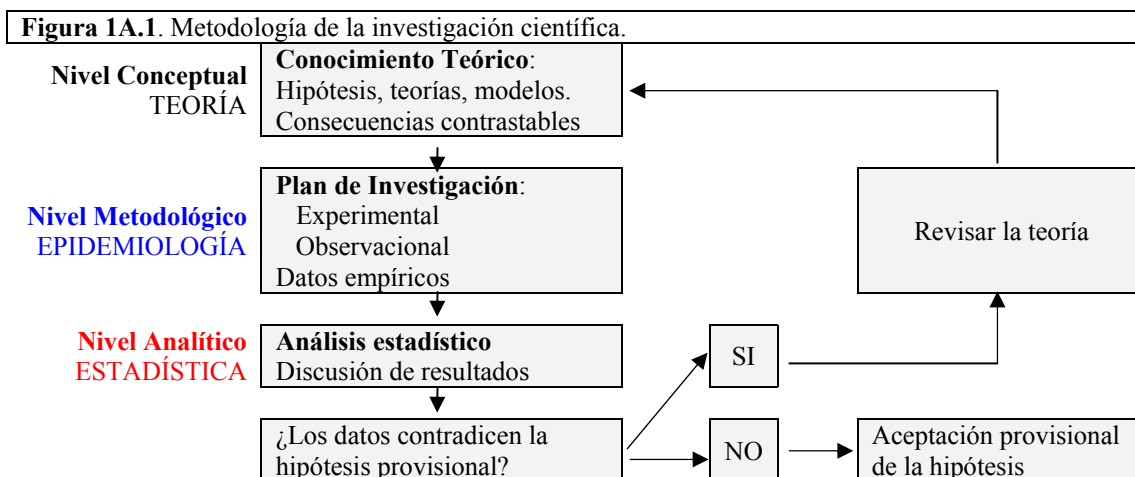
- La presencia de obesidad condiciona la presencia de diabetes: en los obesos es más frecuente la diabetes que en los no obesos.
- El tabaquismo condiciona la presencia e hipertensión arterial: en los fumadores es más frecuente la hipertensión arterial que en los no fumadores.

Para comprobar estas hipótesis se ha realizado un estudio. Los datos necesarios de 112 pacientes recogidos en cinco hospitales de la comunidad de Madrid se encuentran en un archivo Excel *Factores.xls*. El archivo se presenta como una *Matriz de datos* en forma de tabla donde las filas representan los casos o sujetos y las columnas las variables. La primera fila corresponde a los nombres de las variables. Los datos son reales aunque se han realizado ligeras modificaciones por motivos didácticos. Se han registrado las siguientes variables, todas numéricas excepto *Genero* y *Origen* que son alfanuméricas o cadena:

- **Genero:** Mujer o Varón
- **FNac:** fecha de nacimiento
- **FIntro:** fecha de introducción en el estudio
- **Origen:** hospital de la comunidad de Madrid de procedencia (Ramón y Cajal, Príncipe de Asturias, Gregorio Marañón, Clínico San Carlos y 12 de Octubre).
- **Tabaco:** tabaquismo expresado en número de cigarrillos fumados al día.
- **Peso:** peso en Kg.
- **Talla:** talla en cm.
- **Pas:** presión arterial sistólica en mmHg.
- **Pad:** presión arterial diastólica en mmHg.
- **Glucosa:** glucosa plasmática en mg/dl.

### Metodología científica

La **Investigación científica** es un proceso continuo e iterativo de aprendizaje. En base a conocimientos teóricos, se establece una hipótesis inicial y se formula una *pregunta de investigación* de la que se deducen consecuencias que deben ser contrastadas con los datos empíricos de una investigación. Esta investigación puede ser experimental u observacional (Figura 1A.1). Se diseña y se define el proceso de recogida de datos lo más adecuado posible para responder a la pregunta formulada. Se obtienen unos datos empíricos que se someten a análisis estadístico para obtener unos resultados. La discusión de los resultados permite valorar si contradicen la hipótesis inicial. Si no la contradicen se acepta provisionalmente la hipótesis. Si la contradicen se rechaza la hipótesis inicial, se revisa la teoría y servirá de ayuda para formular una nueva hipótesis teórica, manteniendo el proceso iterativo. En todo este proceso son necesarios los conocimientos de dos disciplinas: la epidemiología clínica y la estadística.



## Estadística y Epidemiología Clínica

- La **Epidemiología clínica** estudia la frecuencia de los eventos clínicos y sus relaciones con características del individuo y del medio. Realiza predicciones precisas sobre pacientes concretos mediante el estudio, con métodos científicos, de grupos de pacientes similares.
- La **Estadística** se ocupa de la recopilación, organización, presentación y análisis de datos, con el fin de deducir conclusiones y tomar decisiones razonables. Es una disciplina especializada con una terminología propia (hipótesis nula, potencia, varianza, regresión, modelización, otras) con la que con frecuencia no se está familiarizado. En concreto, el significado del valor “p” frecuentemente es malentendido y todavía sigue siendo clave en la interpretación de los estudios científicos.
  - La **estadística descriptiva** se ocupa de mostrar información concreta sobre la muestra del estudio.
  - La **estadística inferencial** es la encargada de obtener información de la población origen de la muestra a partir de los datos encontrado en la muestra del estudio. Según el número de variables que intervienen se habla de estadística *univariable*, *bivariable* o *multivariable* (o *multivariante*), si intervienen una, dos o más de dos variables, respectivamente.
  - Se habla de **estadística analítica** o *no analítica* si hay o no hay hipótesis previa, respectivamente.

En el presente curso se describen, a partir de un sencillo ejemplo, la mayoría de los términos y conceptos de estadística y epidemiología clínica que se pueden encontrar en la literatura médica.

## Población y Muestra

- La **Población de referencia** o *diana* de un estudio es el conjunto completo de individuos no seleccionados a los que se pretende extrapolar las conclusiones del estudio. Es de gran tamaño y suele estar definida por características que pueden ser geográficas (p. ej. población de Madrid), demográfica (p. ej. mayores de 65 años), o tener una enfermedad (p. ej. diabéticos) o una determinada gravedad (p. ej. neumonía que ha precisado hospitalización), entre otras. Población *elegible* es el conjunto de sujetos que cumplen los criterios de inclusión y no cumplen los criterios de exclusión para poder ser elegidos en la investigación.
- Se denomina **Muestra** o población de estudio al subgrupo seleccionado y representativo de la población elegible en el que se llevará a cabo el estudio. Trabajar con una muestra de tamaño adecuado reduce el coste y aumenta la calidad de los datos. La investigación científica consiste en inferir a una población los resultados encontrados en una muestra.
- Se entiende por **Variable** cada uno de los caracteres o aspectos que se registran en los individuos de una muestra. Las variables pueden tener Valores perdidos (*missing*) que son los casos en que se desconoce el valor de la variable. Pueden ser valores perdidos declarados por el sistema (*system missing*) o declarados por el usuario (*user missing*). Estos últimos son aquellos valores desconocidos por una causa concreta que se quiere especificar. Es el caso de *Tabaco*, en el que dos sujetos no quisieron contestar a la pregunta sobre tabaquismo. Su respuesta se ha registrado con el código 99.

## Tipos de variables

- **Variables según el papel en el estudio.** Todo estudio de investigación se origina para responder una determinada pregunta de investigación. En la primera hipótesis del ejemplo podría ser: ¿la obesidad influye en el desarrollo de diabetes? En esencia, la respuesta a esa pregunta se reduce a analizar la relación entre dos variables, en este caso la *obesidad* y la *diabetes*. Habitualmente las dos variables juegan un papel asimétrico.

Tabla 1A.1. Variables según el papel en el estudio.			
Variable X Independiente Predictora Explicativa Covariable		Variable Y Dependiente Resultado Efecto Criterio	Covariables: -De confusión -De interacción
<b>EXPOSICIÓN</b>	→	<b>RESPUESTA</b>	<b>DE CONTROL</b>
Factor de riesgo Factor pronóstico Intervención terapéutica Intervención preventiva Prueba diagnóstica		Enfermedad Curación Recidiva Evento	Edad Sexo Otras

- Una es la **variable exposición o independiente**, también llamada explicativa, predictora, covariable o variable X (Tabla 1A.1). Es toda variable (factor, característica, tratamiento o exposición, tanto ambiental

como personal) que se utiliza para explicar o predecir un efecto o que puede desarrollar o tener influencia sobre una respuesta. En este caso sería la *obesidad*.

- Otra es la **variable respuesta o dependiente**, también llamada efecto, resultado, criterio o variable Y (Tabla 1A.1). Es toda variable que mide el final de un proceso (efecto, respuesta o resultado) dentro de la historia natural de la enfermedad (p. ej., morbilidad, mortalidad, coste) y que puede estar influido o depender de uno o más factores. En este caso sería la *diabetes*.
- **Variables según la escala de medida.** Se distinguen tres **tipos de variables**: nominal, ordinal (ambas son categóricas, no métricas) y cuantitativas o métricas.
  - Una variable **categórica nominal** se define porque solo se pueden establecer relaciones de igualdad y desigualdad entre las categorías de la variable. No hay categorías ordenadas. Si tienen dos categorías se denominan variables dicotómicas o binarias, como *Genero (Mujer y Varón)*. Si tienen más de dos categorías se denominan politómicas, como *Origen*.
  - Una variable **categórica ordinal** se define porque además de relaciones de igualdad-desigualdad, también se pueden establecer relaciones de orden entre las categorías de la variable. Hay categorías ordenadas. La variable *ObesidadOr* con tres categorías (*Imc Normal, Sobrepeso y Obesidad*) es ordinal porque entre sus tres categorías, además de relaciones de igualdad-desigualdad, se puede establecer una gradación, un orden: *Imc Normal < Sobrepeso < Obesidad*. Permite ordenar la posición de cada sujeto. Las variables binarias siempre tiene escala de medida nominal, nunca pueden ser ordinales.
  - Las variables **cuantitativas** contienen datos métricos, medibles o contables. Según el número de variaciones entre valores, pueden ser discretas o continuas. Las *discretas* son *recuentos*, como *Tabaco*, en el que el número de variaciones entre valores es finito. Las *continuas* son *medidas*, en donde el número de variaciones entre valores es infinito. Pueden dar la apariencia de discretas debido a la falta de precisión del instrumento de medida. Se muestran *Redondeadas* al centro del intervalo como la *Glucosa* o la *Talla*, o bien *Truncadas* al inicio del intervalo como la *Edad* que se calculará posteriormente.

Dependiendo de su escala de medida, la asignación de valores en una variable consiste en *clasificar* (nominal), *ordenar* (ordinal), *contar* (cuantitativa discreta) o *medir* (cuantitativa continua).

En este ejemplo partimos de las variables cuantitativas exposición *Peso* y *Talla* (a partir de ellas se obtiene el *Imc*) y de la variable respuesta *Glucosa*. A partir de ellas se crean variables categóricas con dos o más de dos categorías: *Obesidad Sí/No*, *Nivel de Imc Normal/Sobrepeso/Obesidad* y *Diabetes Sí/No*. Esto se hace por motivos didácticos, para permitir explicar de forma clara la mayoría de las pruebas estadísticas. Pero siempre que sea posible registrar una misma característica con diferentes escalas de medida se debe utilizar la que aporte mayor información, que es la cuantitativa. En este caso son *Imc* y *Glucosa*.

- La variable de **supervivencia o tiempo a un evento**, es un tipo de variable respuesta especial utilizada en los estudios de seguimiento en los se sigue a los sujetos hasta el desarrollo de un evento final (estudios de supervivencia). Está compuesta por dos variables: la respuesta binaria (*Evento Sí/No*) y el *Tiempo* de seguimiento del sujeto, que suele ser diferente para cada sujeto: hasta que ocurre el evento final buscado (tiempo que se denomina completo o no censurado) o hasta el final del seguimiento si no ocurre el evento (tiempo que se denomina incompleto o censurado). En el ejemplo, en el caso de que se monitorizara el tiempo hasta el desarrollo de diabetes, sería la variable *Tiempo Diabetes*, compuesta por el evento *Diabetes Sí/No* y el *Tiempo* de seguimiento de cada sujeto.

## INTRODUCCIÓN AL MANEJO DE DATOS

### Programas estadísticos

- **SPSS**

El programa SPSS esta diseñado para el manejo y análisis estadístico de datos. Permite capturar y leer archivos en diversos formatos, modificar los datos y generar nuevas variables, mostrar listados de casos, generar tablas y gráficos y realizar una gran variedad de pruebas y de análisis estadísticos univariados y multivariados. El acrónimo SPSS procede de *Statistical Package for Social Sciences* o de *Statistical Products Software Solutions*. Actualmente se denomina IBM SPSS y la versión más reciente es la 22.0. La versión 18 se llamó *PASW Statistics 18*. PASW de *Predictive Analytics Software*. En su página Web [www.spss.com/es](http://www.spss.com/es) se puede descargar una versión de evaluación para Windows, Macintosh y Linux, se puede instalar y utilizar libremente durante unos 15-20 días. Tiene tres tipos de archivos:

- **.sav**. Archivos de datos. Consta de *Vista de datos* (matriz con casos en filas y variables en columnas) y *Vista de variables* (matriz con variables en filas y sus propiedades en columnas).

- **.sps.** Archivos de sintaxis que contiene las instrucciones que generarán los resultados. Es recomendable que el tipo de letra o Fuente sea *Courier New*, Estilo *Regular* y Tamaño *12*.
- **.spv (.spo** en versión 15 y anteriores). Archivos de resultados donde aparecen las tablas y gráficos realizados por el programa. Para poder abrir ambos tipos con cualquier versión es necesario instalar un módulo de compatibilidad que se puede obtener libremente de la página web del programa.

### **Videotutorial 1A1Configuración.avi**

Se muestra la descripción general del programa y las configuraciones del ordenador y del programa previas al inicio del curso. En el ordenador hacer lo siguiente:

- Crear una carpeta llamada CURSO que dependa del directorio raíz del disco C.
  - Situarse en Mi PC | Disco C. Con el botón derecho del ratón se abre el menú contextual y se selecciona Nuevo | Carpeta. Se crea la carpeta de nombre Nueva carpeta, que cambiaremos por CURSO. Su ruta es C:\CURSO.
  - Copiar en dicha carpeta los archivos de datos necesarios para seguir el curso: *Factores.xls*, *IntroCurso.sav*, *Factores.sav*, *FactoresIntraF.sav*, *CA19.9ROC.sav*, *Graves.sav*, *Diabetes.sav*, *Kappa.sav*, *MedidasAyB.sav*, *Orinas.sav*, *Pólipos5.sav*.
  - Crear la carpeta SOLUCIONES donde se guardarán los archivos resultantes de los ejercicios mostrados en los videotutoriales.
- Que todos los archivos de su ordenador muestren su extensión.
  - Situarse en cualquier carpeta. La recién creada “CURSO” puede valer.
  - En Windows XP, en la barra de menús al seleccionar Herramientas | Opciones de carpeta... se abre el cuadro Opciones de carpeta. En Windows Vista y 7, Opciones de carpeta esta en Organizar | Opciones de carpeta y de búsqueda...
  - En el cuadro Opciones de carpeta situarse en la pestaña Ver y en Configuración avanzada: buscar Ocultar extensiones para los tipos de archivos conocidos que está hacia el final de la lista. Está activada y hay que desactivarla.

#### ▪ **PSPP**

Es de distribución libre y gratuita, de código abierto y desarrollada por programadores voluntarios como alternativa libre al SPSS, con el que presenta una interfaz muy parecida. Para los investigadores con dificultades de acceso a SPSS y que no requieren procedimientos estadísticos complejos resulta una excelente alternativa. Comparte gran parte de la funcionalidad de su homólogo propietario, pero sin coste. Permite crear e importar bases de datos para el análisis estadístico básico (estadística descriptiva e inferencial, tablas de contingencia, pruebas de hipótesis estadísticas, análisis de varianza, pruebas no paramétricas, regresión lineal, clasificación de datos, etc.).

#### ▪ **Epidat**

Es un programa de epidemiología de libre distribución desarrollado por el Servicio de Información sobre Salud Pública de la Dirección General de Salud Pública de la Consejería de Sanidad de la Xunta de Galicia, en colaboración con la Unidad de Análisis de Salud y Sistemas de Información de Salud de la Organización Panamericana de la Salud (OPS-OMS). Esta dirigido a profesionales de la salud para el análisis de datos tabulados y facilitar la docencia en epidemiología y estadística. Para el manejo del programa no es necesario tener amplios conocimientos epidemiológicos ni estadísticos. Dispone de un excelente manual de ayuda con ejercicios prácticos. La última versión es la 4.0. Se puede descargar libremente de la página <http://www.sergas.es/EPIWB/DescargaEpidat.aspx?IdPaxina=62714&idv=2&lng=es>

#### ▪ **Stata**

Es un potente y eficiente software de análisis estadístico creado en 1985 por StataCorp. El nombre *Stata* es una "palabra baúl" que combina las palabras *statistics* (estadística) y *data* (datos). No es un acrónimo y por lo tanto no debe ser escrito en mayúsculas: no es "STATA" sino "Stata". Es utilizado principalmente por instituciones académicas y empresariales dedicadas a la investigación, especialmente en economía, sociología, ciencias políticas, biomedicina y epidemiología. Es un software completo e integrado que permite, entre otras funcionalidades, la gestión o manejo de datos, el análisis estadístico, el trazado de gráficos y las simulaciones. No se presenta por módulos por lo que en un solo paquete tiene todas las herramientas que se puedan necesitar. Además sus licencias son perpetuas por lo que no debe pagar incómodos mantenimientos. La última versión de Stata es la 13 (Junio de 2013). Los tipos de archivos que utiliza son *.dta* (datos), *.do* (comandos), *.ado* (programas), *.hlp* (ayuda), *.gph* (gráficos), *.dct* (diccionarios) y *.smcl* (archivos log).

### **Modalidades de trabajo con SPSS**

En la Tabla 1A.2 se muestran las tres posibles formas de trabajar con SPSS (Menús, Sintaxis y Mixto) con las ventajas (señaladas en azul) e inconvenientes (señalados en rojo) de cada modalidad.

#### 1. **Con Menús y Cuadros De Diálogo.**

Tiene la gran ventaja de ser la forma más fácil, directa e intuitiva de trabajar con el programa. Pero el resto son inconvenientes. Es necesario repetir todas las operaciones con menús y cuadros de diálogo si se añaden

nuevos casos a la matriz de datos o si ha habido errores. Si se cometen errores es difícil su identificación y probablemente pasarán desapercibidos, y si se detectan es más difícil averiguar su origen para corregirlos. No permite guardar pruebas para justificar lo realizado en caso de ser requerido por los revisores de las revistas o en otras instancias. Toda la potencialidad del programa no está implementada en los cuadros de diálogo y en los menús, por lo que es una modalidad de trabajo limitada.

<b>Tabla 1A.2. Modalidades de trabajo con SPSS.</b>	
<b>1. MENÚ y Cuadros de Diálogo</b>	<b>2. SINTAXIS</b>
<ul style="list-style-type: none"> <li>- Fácil de utilizar, directo, intuitivo</li> <li>- Evanescente y no Reproducible</li> <li>- Dificulta la detección de errores</li> <li>- No permite corrección de errores</li> <li>- No permite evaluar el trabajo realizado</li> <li>- No permite comprobar el trabajo realizado</li> <li>- Limitado: no permite todo el potencial del programa</li> </ul>	<ul style="list-style-type: none"> <li>- Difícil y engorroso: hay que conocer la sintaxis</li> <li>- Permanente y Reproducible</li> <li>- Facilita la detección de errores</li> <li>- Facilita la corrección de errores</li> <li>- Permite evaluar el trabajo realizado</li> <li>- Permite comprobar el trabajo realizado</li> <li>- Ampliado: permite todo el potencial del programa</li> </ul>
<b>3. MIXTO con botón Pegar</b>	

## 2. Con Sintaxis.

Su gran y único inconveniente es que es muy difícil y engorroso: es necesario aprender y conocer el lenguaje de sintaxis propio y preciso del programa. Sin embargo el resto son ventajas. Permite la fácil detección y corrección de errores, es permanente y reproducible con un simple clic de ratón, hace posible la comprobación o evaluación del trabajo realizado examinando el archivo de sintaxis y permite utilizar toda la potencialidad del programa, incluso las operaciones e instrucciones que no están disponibles por menús.

La estructura de una instrucción de sintaxis consta de tres partes. Siempre comienza con el nombre principal de la instrucción, es el Procedimiento o Comando y termina con un punto. Todo lo que hay entre medias es el Campo de Especificación. Desde la versión 18 se utilizan colores para jerarquizar las órdenes de una instrucción. Los colores son configurables en el cuadro *Opciones*, ficha *Editor de sintaxis*. Por defecto son:

- **Comandos** en azul.
- **Subcomandos** en verde.
- **Palabras clave** (keywords) en marrón, como:
  - Operadores lógicos: **AND OR NOT**
  - Operadores relacionales: **EQ NE LT GT LE GE**
  - Otras palabras clave: **ALL BY TO WITH MEAN...**
- **Valores** en naranja y
- **Comentarios** en gris. Lo que hay entre un asterisco (\*) y un punto (.) es comentario en gris. Se pueden escribir indicaciones libremente, el programa no lo identifica como instrucción.

También pueden aparecer Nombres de variables (en negro) y Delimitadores que pueden ser:

- Ordinarios: espacio en blanco y coma
- Especiales: paréntesis, apóstrofes, comillas, barra oblicua o “slash” (/) y el signo igual.
- Operadores aritméticos: + - \* / \*\*

Por ejemplo esta sintaxis tiene los siguientes elementos:

```

FRECUENCIES VARIABLES=Pas Pad
  /FORMAT=NOTABLE
  /STATISTICS= STDDEV MEAN
  /ORDER=ANALYSIS.
```

- Nombre del Procedimiento o Comando: **FRECUENCIES**
- Subcomandos: **FORMAT STATISTICS ORDER**
- Palabras clave: **VARIABLES NOTABLE STDDEV MEAN ANALYSIS.**
- Nombre de Variables: **Pas y Pad**
- Separador ordinario: el espacio en blanco
- Separador especial: los signos igual (=) y barra oblicua (/)
- Finalizador: el punto (.)

Sólo se necesitan los tres primeros caracteres de los nombres de las instrucciones y palabras clave:

```

FRE VAR=Pas Pad /STA=STD MEAN.
```

## 3. Mixto con el botón Pegar.

Trata de aprovechar las ventajas y evitar los inconvenientes de los dos métodos anteriores, y lo consigue. Consiste en realizar los procedimientos con menú y cuadros de diálogo y salir de ellos con el botón Pegar (Paste) que escribe la instrucción en la ventana de sintaxis. Posteriormente hay que seleccionar y ejecutar la instrucción de sintaxis. Es el método deseable y que se utilizará durante el curso.

Para ejecutar un procedimiento escrito en la ventana de sintaxis, se coloca el cursor en cualquier punto de la orden de sintaxis y se pulsa el botón Ejecutar. Para ejecutar dos o más órdenes se deben marcar todas juntas arrastrando el ratón con el botón izquierdo pulsado desde una parte de la primera orden hasta otra parte de la última y después ejecutarlas. Al ejecutar un procedimiento se envían las instrucciones al sistema. Si todo es correcto se ejecuta, pero si hay algún error, no puede ejecutarlo y escribe en el visor de Resultados un mensaje de error indicando la causa del error. Se debe leer, intentar entender y después se debe eliminar. Se soluciona este error en la ventana de sintaxis y se ejecuta de nuevo. También se puede utilizar el menú Ejecutar de la ventana de sintaxis con sus opciones Todo, Selección, Actual y Hasta el final.

## Capturar y guardar archivos

Los archivos de datos en formato .sav son los más utilizados, pero el programa tiene capacidad para leer datos de casi todos los formatos: Excel (.xls), dBase, etc.

### Capturar y Guardar archivos de datos SPSS (\*.sav)

Desde el menú Archivo | Abrir | Datos... se abre el cuadro *Abrir datos*. En Archivos de tipo se selecciona PASW Statistics (\*.sav) para que muestre los archivos .sav o Todos los archivos (\*.\*) para que muestre todos los tipos de archivos. Se busca la ubicación del archivo deseado y se sale con el botón Pegar (Paste). En el caso del archivo C:\CURSO\IntroCurso.sav se escribe la siguiente sintaxis:

```
GET FILE='C:\CursoSPSS\IntroCurso.sav'.  
DATASET NAME $Conjunto_de_datos WINDOW=FRONT.
```

El comando `DATASET NAME $Conjunto_de_datos WINDOW=FRONT` se refiere al trabajo con varios archivos de datos a la vez. Recomendamos no tener abiertos más de un archivo de datos a la vez. Es preferible borrar este comando o mejor, inactivarlo (“hacerle comentario”) anteponiéndole un asterisco (\*) con lo que se vuelve de color gris para que no se lleve a cabo en futuras ejecuciones de la sintaxis. Quedaría así:

```
*DATASET NAME $Conjunto_de_datos WINDOW=FRONT.
```

Para guardar un archivo de datos, desde el menú Archivo | Guardar, si el archivo ya tiene nombre asignado previamente, el programa guarda los cambios con el mismo nombre y ubicación. Si el archivo de datos es de nueva creación y no tiene nombre asignado, el programa pide confirmación para guardarlo. Si se responde afirmativamente, aparecerá el cuadro de diálogo *Guardar datos como*, en el que debe escribir el nombre del archivo, en este caso *IntroCurso.sav*.

### Guardar archivos de sintaxis SPSS (\*.sps)

Para guardar los archivos de sintaxis desde el menú de la ventana de sintaxis se debe abrir el cuadro *Guardar sintaxis como* desde el menú Archivo | Guardar si no tiene ningún nombre asignado. También desde Archivo | Guardar como... válido tanto para si ya tiene nombre asignado como si no lo tiene. Sólo falta escribir el nombre del archivo en Nombre de archivo:

### Salir del programa SPSS

Desde Archivo | Salir se cerrará automáticamente el programa con todas sus ventanas si estas tienen nombre asignado y no han sufrido cambios. En caso de que hayan sufrido cambios o no tengan nombre asignado, pide confirmación para guardar las tres ventanas, sintaxis, resultados y datos. Sólo nos interesa la ventana de sintaxis que debemos guardar con el nombre que deseemos. La ventana de resultados no nos interesa guardarla, ya que con los datos y la sintaxis los resultados se generan automáticamente en segundos. Sólo interesa guardar aquellos gráficos o tablas que han sido editados.

### Videotutorial 1A2Guardar.avi

Se muestra como se abre el archivo *IntroCurso.sav* y como se guardan, en la carpeta SOLUCIONES, el archivo de datos con el mismo nombre (*IntroCurso.sav*) y el archivo de sintaxis con el nombre *Guardar.sps*.

### Capturar archivos en formato Excel (\*.xls)

Desde el menú Archivo | Abrir | Datos..., se abre el cuadro *Abrir archivo*. En Archivos de tipo se selecciona Excel (\*.xls \*.xlsx \*.xlsm). Se busca la ubicación del archivo deseado, se selecciona y se sale con el botón Pegar. Aparece el cuadro *Apertura de origen de datos de Excel*:

- Leer nombre de variables de la primera fila de datos. Al marcarla toma los nombres de variables de la primera fila del archivo. En Excel se suelen poner los nombres de variables en primera fila. Es el caso del archivo *Factores.xls*.
- Hoja de trabajo. Permite seleccionar la hoja de trabajo de la que se obtendrán los datos, en este ejemplo es *FRCV*. Se indica la primera y última celda que contienen los datos, en este caso A1:J113.
- Rango. Permite seleccionar una fracción de casillas con datos a seleccionar. Es conveniente, por seguridad, marcar el rango que vamos a capturar escribiendo la posición de la 1ª casilla superior izquierda (A1), dos puntos (:) y la posición de la última casilla inferior derecha (J113). En este caso sería A1:J113.

– Anchura máxima para las columnas de cadena. Permite indicar la anchura máxima destinada a variables cadena. Conviene acortarla hasta contener el nombre mas largo de las posibles categorías a importar, ya que por defecto es muy grande.

Se pulsa el botón Aceptar y se escribe la instrucción en la ventana de sintaxis. En el caso del archivo C:\CURSO\Factores.xls y limitando a 20 el ancho máximo de variables cadena, sería:

```
GET DATA
  /TYPE=XLS
  /FILE='C:\CURSO\Factores.xls'
  /SHEET=name 'FRCV'
  /CELLRANGE=range 'A1:J113'
  /READNAMES=on
  /ASSUMEDSTRWIDTH=20.
*DATASET NAME $Conjunto de datos WINDOW=FRONT.
```

### Videotutorial 1A3Captura.avi

Se muestra la captura del archivo en formato Excel *Factores.xls* y el guardado del archivo de datos (*Captura.sav*) y de sintaxis (*Captura.sps*) en la carpeta SOLUCIONES.