

# EL ANÁLISIS DE DATOS MEDIANTE PROCEDIMIENTOS INFORMÁTICOS

## Introducción al SPSS

### ÍNDICE DE CONTENIDOS

1. El proceso de investigación socioeducativa.....	2
2. Análisis estadístico: función y objetivos .....	2
3. Trabajando con SPSS .....	2
3.1. Descripción.....	2
3.2. Creación de variables .....	4
3.2.1. Definición de las variables.....	5
3.2.2. El nombre de las variables .....	5
3.2.3. El tipo de las variables.....	5
3.2.4. Las etiquetas de las variables .....	7
3.3. Estructura del fichero de datos .....	8
3.4. Grabación y Edición de datos.....	8
3.4.1. Introducción de los datos por casillas, por filas y por columnas .....	8
3.5. Recodificación de variables .....	9
3.6. Ordenación y selección de casos .....	11
3.6.1. Procedimiento Ordenar casos .....	11
3.6.2. Procedimiento Seleccionar casos .....	11
3.7. El Visor de Resultados .....	12
3.8. Análisis de datos.....	13
3.8.1. Estadística Descriptiva .....	14
3.8.1.1. Procedimiento Frecuencias .....	14
3.8.2. Estudio psicométrico de instrumentos: fiabilidad y validez .....	16
3.8.2.1. Análisis de consistencia interna.....	16
3.8.2.2. Discriminación de los ítems .....	17
3.8.2.3. Estudio Factorial.....	18
3.8.3. Pruebas de significación estadística.....	19
4. Bibliografía y recursos .....	20

## 1. El proceso de investigación socioeducativa

La puesta en marcha de un proceso de investigación socioeducativa implica clarificar qué entendemos por investigación, dando respuesta a las preguntas de qué es investigar, por qué, para qué, cómo, quién, a quién, cómo y cuándo. El esquema general de todo proceso de investigación incluye los siguientes elementos: Formulación del Problema, Objetivos, Hipótesis, Variables, Diseño, Instrumentos, Muestra, Análisis de datos, Interpretación de los resultados, Conclusiones y Difusión

## 2. Análisis estadístico: función y objetivos

Una característica fundamental de la sociedad es la abundancia de información disponible. El ser humano se siente continuamente bombardeado por datos que debe aprender a *leer e interpretar críticamente*.

El estudioso de la realidad social debe, por tanto, ser capaz de llevar a cabo estas dos funciones con la información, así como formular preguntas adecuadas e importantes y generar los datos que faciliten las respuestas apropiadas. En definitiva, es necesario aprender a utilizar los métodos para extraer conclusiones basadas en los datos.

La Estadística es, en este contexto, una herramienta de gran utilidad que permite alcanzar un mayor conocimiento y entendimiento de los problemas reales.

Puede decirse, por lo tanto, que el objetivo del análisis estadístico se concreta en descubrir las conclusiones que pueden extraerse de los datos y en presentar éstas de una forma simple y clara.

## 3. Trabajando con SPSS

El programa SPSS (paquete estadístico aplicado a las ciencias sociales) constituye un programa modular que implementa gran variedad de temas estadísticos orientados al ámbito de las ciencias sociales desde hace más de 30 años. Actualmente, cubre casi todas las necesidades del cálculo estadístico de los investigadores y profesionales, no sólo del campo de las ciencias sociales, sino también de las humanas y de las biomédicas y, en general, de cualquier campo de actividad en el que se precise el tratamiento estadístico de la información.

### 3.1. Descripción

El paquete SPSS, es un programa adaptado al entorno WINDOWS con lo cual la forma de ejecutarlo es a través de ventanas en las que se despliegan menús, de los que se pueden elegir distintas opciones y así sucesivamente; por tanto es a través de un entorno de tipo gráfico desde donde se solventan los problemas, y no a través de comandos (aunque también se puede hacer así) como se hacía antes en los paquetes estadísticos más usuales.

Por lo que acabamos de decir, la forma de iniciar la ejecución del programa SPSS es pinchando dos veces con el ratón ("pinchar" lo utilizaremos como sinónimo de "hacer clic" con el botón principal del ratón) en el icono de SPSS que es como el de la Figura 1 y que se suele encontrar en el escritorio en forma de enlace o en el menú de inicio dentro del apartado de programas. Una de las

primeras tareas que tendrá que hacer el usuario de SPSS será localizar la posición del icono y adaptarlo a su gusto para que la entrada al programa sea fácil



Figura 1: Icono del programa SPSS

Cuando se ha pinchado dos veces sobre el icono, se abre la ventana principal de SPSS que es el Editor de datos de SPSS, aunque también la llamaremos ventana principal de SPSS. Esta ventana tiene dos versiones o vistas: vista de datos y vista de variables. En la figura 2 podemos ver la vista de datos, incorporados los datos de un fichero de datos.

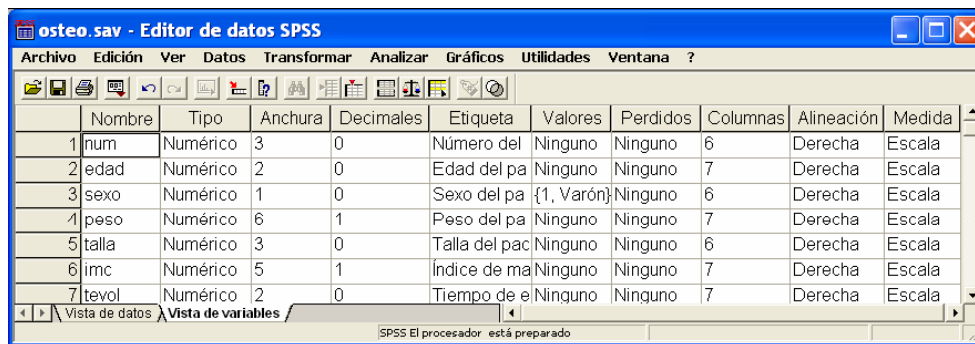


Figura 2: Vista de variables en el editor de datos

En la figura 3 aparece la vista de datos; en ella se muestran las características de todas las variables del fichero de datos. De una vista a otra se cambia pinchando con el ratón en la pestaña correspondiente en la parte inferior izquierda de la ventana. Dentro de la vista de datos se pueden distinguir varias zonas. La primera zona (parte más alta de la ventana, generalmente en color azul) está formada por la barra que contiene el nombre de la ventana, incluyendo el nombre del fichero de datos activo si existe. La segunda zona a destacar (justo debajo de la anterior) es la zona de los menús en la que aparecen los nombres de los menús desplegables que nos servirán para llevar a cabo nuestras tareas; cuando se coloca el cursor sobre uno de los rótulos y se pincha con el ratón se despliega un menú sobre el que desplazando el ratón se remarcarán las acciones que se pueden ejecutar y de la que nosotros escogeremos una; los menús que se pueden desplegar figuran en la siguiente tabla con una breve descripción de lo que se puede hacer con ellos, descripción que corresponde en muchos casos con la de cualquier programa Windows que presenta tales menús.

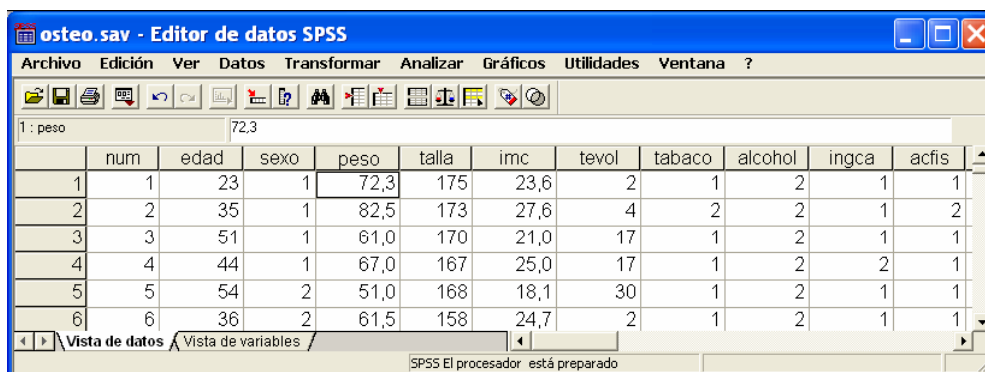


Figura 3: Vista de datos en el editor de datos

Tabla 1. Menús de la ventana principal de SPSS

Menú	Función
<u>A</u> rchivo	Todas las funciones que se pueden hacer con archivos: Abrir, cerrar, guardar, importar, exportar, imprimir, etc.
<u>E</u> dición	Realiza todas las funciones típicas de la edición como son: cortar, copiar, eliminar, buscar, reemplazar, etc...
<u>V</u> er	Permite controlar la forma en la que se ve la pantalla principal, controlando las barras que aparecen en ella así como la forma en la que se presentan los datos.
<u>D</u> atos	Contiene el conjunto de acciones que se pueden llevar a cabo con los datos: definir propiedades de las variables, seleccionar casos, ordenar casos y muchas más.
<u>T</u> ransformar	Permite realizar cualquier función conducente a crear nuevas variables a partir de otras existentes o no: transformar, recodificar, asignar rangos, etc...
<u>A</u> nalizar	Acceso al conjunto de programas de SPSS, que van desde la generación de una tabla de frecuencias a análisis multivariantes complejos.
<u>G</u> ráficos	Acceso al conjunto de gráficos estadísticos que van desde un simple histograma al dibujo de una curva ROC.
<u>U</u> tilidades	Acceso a la descripción de las variables del fichero activo, creación de grupos de variables, así como edición de los menús.
<u>V</u> entana	Acceso rápido a las ventanas de datos, de resultados, de sintaxis.
<u>?</u>	Ayuda en línea sobre todo el paquete SPSS.

Como se puede observar en la tabla 1, bajo una de las letras del rótulo del menú, aparece un subrayado, indicando tal cosa que combinando la tecla Alt con la letra subrayada se puede desplegar el menú correspondiente, pudiendo desplazarse uno por él sin más que usar las teclas de flecha presentes en el teclado.

### 3.2. Creación de variables

El programa SPSS trabaja con los datos previamente grabados en un fichero al que denominamos fichero de datos y que ya sabemos que tiene la extensión obligatoria .sav. Cuando se ejecuta el programa SPSS tal fichero deberá estar activo para que sobre él se lleven a cabo los cálculos oportunos. Esto se conseguirá o bien leyendo un archivo de datos previamente grabado o bien creando uno y grabando en él los datos. Una vez que el fichero está activo se puede ver en la ventana del editor de datos en forma de una rejilla en la que sus filas son los individuos o casos y las columnas son las variables objeto del estudio; a la presentación de la citada rejilla la denominamos matriz de datos.

De todo lo dicho queda claro que lo primero es disponer de los datos y para disponer de los datos tendremos que definir, antes de grabarlos, cada una de las variables que vamos a grabar para cada individuo (caso); lo común, y así será siempre en este manual, es que para todos los individuos se disponga de información sobre las mismas variables, aunque, evidentemente, pueda faltar en algunos información de alguna variable; la matriz de datos que obtendríamos así se denomina matriz rectangular (fichero rectangular), indicando con ello que a todos los casos se les recoge información de las mismas variables. Por tanto empezamos definiendo las variables de un nuevo fichero de datos.

### 3.2.1. Definición de las variables

Una variable es cualquier fenómeno o cualidad que pueda sumir valores numéricos. Estas se clasifican, en función de la naturaleza matemática que cumplen para la investigación, en cualitativas y cuantitativas. Una variable de tipo cualitativo es aquella que representa una cualidad (no son medibles numéricamente), mientras que la variable cuantitativa expresa una cantidad (implican el concepto de magnitud). Desde el punto de vista metodológico, estas pueden ser dependientes (el objeto de estudio) e independientes (de clasificación o aquellas que introducimos para provocar cambios en la variable dependiente).

Se definen a partir de un sistema de categorías que las representen. Han de cumplir tres condiciones para su correcta formulación (claramente definidas, exhaustivas –han de tenerse en cuenta todos y cada uno de los casos, de modo que todos los sujetos puedan estar identificados– y excluyentes mutuamente –cada caso sólo puede estar incluido en una categoría). En este sentido, podemos disponer de variables unicategoriales, dicotómicas y politómicas.

Cuando se abre SPSS aparece la matriz de datos vacía indicando en ese caso que no hay ningún fichero seleccionado para trabajar con él; es en esta situación en la que debemos estar para proceder a crear la estructura (definir las variables) de un nuevo fichero de datos de SPSS. A esa situación también se puede llegar después de haber trabajado con SPSS sin más que desplegar el menú Archivo y elegir de él la opción Nuevo y dentro de ella datos (*Archivo, Nuevo, Datos*), lo que hará que se elimine el fichero activo y se deje limpia la matriz de datos. Ahora cada columna será una nueva variable en potencia hasta que sea definida, en cuyo caso pasará a ser una variable en la nueva matriz de datos.

Una variable queda automáticamente definida si en alguna casilla de su columna se introduce algún dato: SPSS define una variable adecuada al primer dato introducido. A estas variables, SPSS les asigna por defecto un nombre constituido por las letras VAR seguidas de un código de cinco dígitos (00001, para la primera; 00002 para la segunda; ...). Siempre podremos cambiar después el nombre y las características de la(s) variable(s) así definida(s).

De todas formas, la manera más natural de crear las variables de un fichero es definir las antes de introducir dato alguno. Para ello hemos de situarnos en la vista de variables de la ventana principal de SPSS, pinchando en la pestaña correspondiente o haciendo doble clic en la cabecera de la columna. En la vista de variables, cada fila corresponde a una variable, y las columnas representan las características de la(s) variable(s). Para cada variable habrá que ir especificando cada una de sus características, empezando por el nombre (primera columna).

### 3.2.2. El nombre de las variables

Se pincha (o se hace doble clic) sobre la casilla correspondiente al nombre de la variable que se está definiendo, y se escribe el nombre que deseemos que tenga, con un máximo de 64 caracteres, sin espacios en blanco.

### 3.2.3. El tipo de las variables

Una vez escrito el nombre de la variable, si se pincha en la casilla correspondiente al tipo de variable (o en cualquier otro sitio), aparecerá, para dicha variable, “numérico” en la columna Tipo, un 8

en la columna Anchura y un 2 en la columna Decimales. Esto es lo que SPSS ofrece por defecto para esa variable. Para cambiar el tipo de la variable se pincha en la casilla de tipo de variable y se pincha sobre el pequeño recuadro gris situado en dicha casilla, aparece entonces una ventana como la de la Figura 4; en ella se puede definir el tipo de la variable de entre los que aparecen: numérico, cadena, fecha, etc. El formato de la ventana es muy simple, a la izquierda aparecen los diferentes tipos que pueden tener las variables y a la derecha aparece el tamaño o características particulares de cada tipo que también son definibles; así en la figura 4, el tipo es numérico y la variable tendrá una anchura total de 8 caracteres, dos de los cuales son decimales: esto es lo que por defecto coloca SPSS. Pero describamos los diferentes tipos.

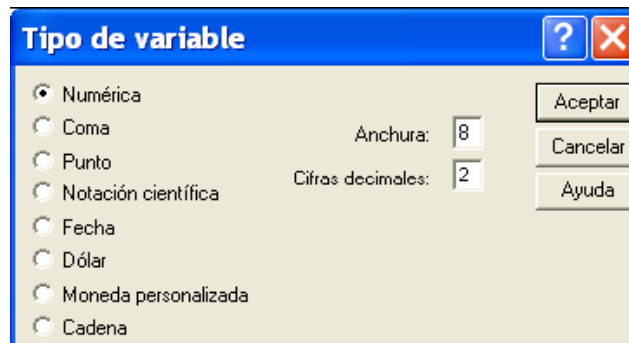


Figura 4: Tipología de variables

El tipo Numérico es para una variable cuyos valores son números. Los valores se muestran en el formato numérico estándar, utilizando como delimitador decimal el especificado en la Configuración Regional del Panel de control de Windows. El tipo Coma define una variable numérica cuyos valores se muestran con la coma de separación de miles y con un punto como separador de la parte decimal. El tipo Punto define una variable numérica cuyos valores se muestran con el punto de separador de miles y con una coma como separador de la parte decimal. El tipo Notación científica define una variable numérica cuyos valores se muestran con una E intercalada y un exponente con signo que representa una potencia de base diez. El exponente puede ir precedido de E o D con un signo opcional, o por el signo solamente. Por ejemplo, 123, 1,23E2, 1,23D2, 1,23E+2, o incluso 1,23+2. El tipo Fecha define una variable numérica cuyos valores se muestran en uno de los diferentes formatos de fecha-calendario u hora-reloj. Para ello, habrá que seleccionar un tipo de fecha de la lista desplegable que aparece a la derecha cuando se selecciona el tipo fecha. Se pueden introducir las fechas utilizando como delimitadores: barras, guiones, puntos, comas o espacios en blanco. El tipo Dólar sirve para definir una variable numérica (con diferentes formatos) cuyos valores contienen un signo de dólar y una coma para la separación de los miles. Moneda personalizada sirve para definir 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 dentro del menú Edición. Los caracteres definidos en la moneda personalizada no pueden emplearse para la introducción de datos, pero sí los mostrará el Editor de Datos. Por último el tipo Cadena define una variable cuyos valores no son numéricos y, por ello, no se utilizan en los cálculos. Pueden contener cualesquiera caracteres hasta la longitud definida. Estas variables son conocidas como variables alfanuméricas.

El tamaño y el formato de cada tipo se expresa en los campos que aparecen en la parte de la derecha de la ventana. Habrá que especificar el tamaño total y el número de decimales en los tipos Numérico, Coma, Punto y Notación Científica, colocándose sobre los huecos oportunos y escribiendo ahí un número para el tamaño global y otro para el número de decimales. La anchura máxima en todos ellos es de 40 posiciones, siendo 16 el número máximo de cifras decimales. Para el tipo Cadena habrá

que especificar la anchura total que no podrá sobrepasar los 255 caracteres. Para los tipos de Fecha, Dólar y de Moneda Personalizada habrá que elegir el formato de entre los que aparecen a la derecha en la ventana de definición de tipos.

Por último, la anchura y número de decimales se pueden cambiar igualmente en las columnas Anchura y Decimales de la vista de variables.

#### 3.2.4. Las etiquetas de las variables

Generalmente el nombre de las variables suele tener pocos caracteres, lo que hace que dicho nombre sea en muchos casos un acrónimo de difícil traducción para el que no lo ha escrito; por ello se suele poner una etiqueta de variable a la variable, que la identifique de una manera más precisa y que nos permita reconocerla cuando se presentan los resultados. Esa etiqueta puede ser como máximo de 130 caracteres. Pero no sólo se puede poner una etiqueta general para la variable, sino que en las variables que sean cualitativas y cuyas categorías suelen estar representadas por códigos numéricos, pueden establecerse etiquetas de valor que permitan identificar a las categorías con ellas (haciéndolas más explicativas) en vez de con los códigos numéricos. Estas etiquetas pueden ser de hasta 60 caracteres. Estas etiquetas no son viables para variables de cadena de más de 8 caracteres de amplitud.

La etiqueta de la variable se escribe directamente en la casilla correspondiente de la columna Etiqueta en la vista de variables. Para las etiquetas de valor hay que pinchar sobre la casilla correspondiente de la columna Valores (en la vista de variables) y luego pinchar sobre el pequeño recuadro gris que aparece en dicha casilla, se despliega entonces una ventana como la de la Figura 5. En esta ventana, la forma en que se asignan etiquetas a los códigos de las variables, es escribir un código en la zona en la que aparece "Valor" y la etiqueta correspondiente en la zona de "Etiqueta de valor" y después pulsar el botón de **Añadir**; así, en la Figura 5, al código 1 se le ha asignado la etiqueta "no", al código 2 se le ha asignado la etiqueta "moderado" y al código 3 se le añadirá la etiqueta "grave" en cuanto pulsemos el botón **Añadir**. El código al que no se le asigne etiqueta aparece como tal en todos los resultados de los análisis. Como se ve, los códigos y sus etiquetas correspondientes son guardados para que queden a la vista del usuario en una zona en la que son susceptibles de ser seleccionados. Cuando se selecciona una de estas parejas "código = etiqueta" vuelve a ser colocada en la zona de etiquetas de valor para que se pueda proceder a su modificación; si después de hacer las modificaciones oportunas se desea guardar la información bastará pulsar en **Cambiar**; si una vez seleccionada la pareja se desea borrar bastará pulsar el botón **Borrar** y la pareja será eliminada; el programa permite seleccionar una pareja cada vez. Cuando se hayan escrito todas las etiquetas bastará con pulsar el botón **Aceptar** y se volverá a la vista de variables.

Las etiquetas, tanto de las variables como de los códigos de las variables, no son obligatorias pero si son muy útiles a la hora de entender y presentar los resultados.

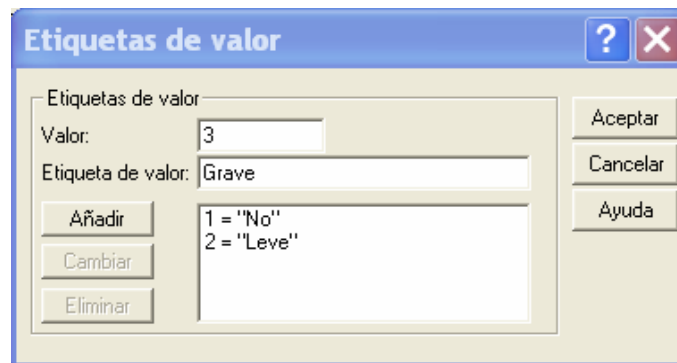


Figura 5: Etiquetas de valor

### 3.3. Estructura del fichero de datos

Con todo lo visto hasta aquí podemos definir las variables en SPSS; la definición cuidadosa y detallada de las variables ayudará mucho en los análisis y en la interpretación de los resultados, por lo que recomendamos al usuario que gaste el tiempo necesario en tales definiciones (pues no será tiempo perdido). Cada variable definida aparece en la vista de datos con su nombre en la cabecera de la columna y se puede seleccionar pulsando el botón izquierdo del ratón sobre ella una vez.

El conjunto de variables definidas, junto con las características que les hayamos asignado, forman lo que podemos llamar la estructura del fichero de datos; esto sería una de las dos partes de un fichero de datos de SPSS. Esta estructura se puede guardar en un fichero .sav que aparecerá sin los datos, pero donde se han guardado las variables y sus características, y la podemos ver resumida en una ventana mediante **Utilidades**→**Variabes**. Sobre dicha estructura se puede añadir el otro ingrediente, los datos propiamente dichos, y juntos conformarán el archivo de datos de SPSS. Para guardar el trabajo actual, efectuar **Archivo**→**Guardar** (ó usar el botón “guardar”) asignando el nombre deseado, por ejemplo, prueba.sav.

### 3.4. Grabación y Edición de datos

Una vez que el usuario ha definido las variables de su fichero de datos, tendrá una rejilla en la que por columnas aparecerán dichas variables con el nombre de cada una de ellas. Es claro que tal rejilla estará vacía y que el siguiente paso será el rellenarla a base de grabar la información de los diferentes casos (que aparecen en el Editor de Datos por filas). Por ello lo primero que explicaremos es cómo se graban los datos. Desde luego que antes de la grabación de casos se debería conocer cómo se mueve uno dentro de la rejilla de datos. El movimiento más simple es desplazar el cursor mediante el ratón a la posición deseada y pinchar en ella con el botón izquierdo. Otras formas más lentas es usando las teclas de flechas en las cuatro direcciones que funcionan como ellas indican.

#### 3.4.1. Introducción de los datos por casillas, por filas y por columnas

La primera forma en que se puede grabar un dato es situarse en una casilla (cruce de una variable, columna, con un caso, fila) pinchando en ella con el botón izquierdo del ratón y escribir el valor deseado seguido de la tecla **Intro**. Tras ello, el cursor aparecerá en la casilla inmediatamente debajo. Evidentemente esto nos permitirá la grabación de los datos por columnas (o variables): cuando



lleguemos al final de una columna pasamos al principio de la siguiente. Sin embargo, esta forma de grabar los datos, no es la más común para proceder a la grabación de un fichero de datos.

La forma más común es grabar los datos por filas, es decir, por casos. Para ello se coloca el cursor en la casilla más a la izquierda de una fila (de un caso) y se introduce el dato correspondiente a esa casilla, seguido de la pulsación de la tecla de flecha hacia la derecha, o pulsando la tecla Tab. Obsérvese que si se pulsa la tecla Intro el cursor bajará una fila abajo con lo que estará fuera del caso, habiendo que volver a él mediante la tecla hacia arriba, lo que alarga claramente la grabación. Cuando uno haya grabado un caso, puede saltar al siguiente pulsando la tecla Intro, y así aparecerá al final del nuevo caso, tras lo cual pulsando la tecla Inicio estará colocado en la primera variable del nuevo caso. Cuando se procede así, SPSS identifica cada caso con un número que es interno y que no existe como variable del fichero, por lo que si el usuario desea una identificación para cada caso debe crear, y grabar, una variable a tal efecto, lo cual es conveniente.

Hay una forma, algo más cómoda, de grabar un caso en SPSS; consiste en seleccionar el caso a grabar (pinchando con el ratón sobre el número asignado por SPSS al mismo) y proceder a la grabación de cada uno de los datos correspondientes a cada una de las variables, seguidos del Intro; en este caso, tras cada Intro el cursor no se desplazará hacia abajo sino que lo hará horizontalmente hasta llegar al final del caso desde donde saltará al principio del mismo caso; en ese instante seleccionaremos el nuevo caso para grabar. Si bien este procedimiento es más cómodo, tiene un problema, y es que si se comete un error, al intentar borrar el dato con la tecla Supr o la tecla BackSpace, se pierde toda la información del caso (aunque puede ser recuperada con Deshacer). En estos casos, para corregir el error, antes debemos situar el cursor en la línea de edición de datos, pinchando en ella con el ratón.

### *3.5. Recodificación de variables*

La creación de nuevas variables se puede llevar a cabo por el procedimiento de Recodificación que, como su nombre indica, cambia los valores de una variable por otros mediante un proceso que no es de tipo numérico sino de especificación de las reglas que producirán el cambio. La forma de invocar la recodificación es a partir del menú *Transformar, Recodificar...* Colocados sobre la opción Recodificar, aparece un nuevo menú desplegable para darnos a elegir entre: En las mismas variables... o En distintas variables... Estas dos alternativas nos permiten grabar el resultado de la recodificación en la misma variable fuente que la produce o en una variable distinta (nueva o no). La recodificación sobre la misma variable tienen el inconveniente, si es que lo es, de que el resultado de la misma destruye lo que ya había en ella, de manera que la vuelta a la codificación primitiva sólo será posible mientras se pueda aplicar la acción de Deshacer; no obstante, a veces eso es lo deseado por el usuario. La recodificación en una variable distinta conserva las dos, la codificación primitiva y la recodificación, pero a costa de agrandar el espacio ocupado por el fichero. Veremos aquí la recodificación En distintas variables..., ya que generalmente es la más aconsejable (la recodificación En las mismas variables... se hace de forma similar).

La ventana correspondiente es la de la Figura 6; dicha ventana nos permite elegir las variables que van a ser recodificadas; lo que significa que varias variables pueden ser recodificadas simultáneamente con las mismas reglas.

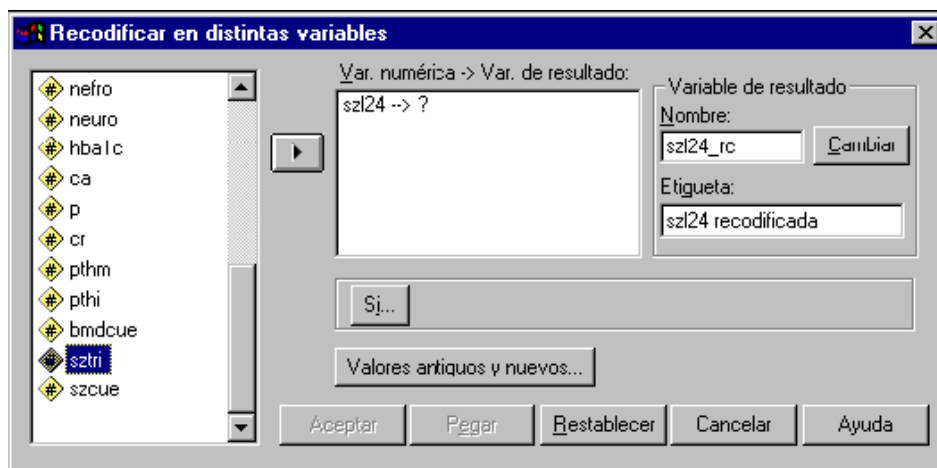


Figura 6: Ventana de recodificación en distintas variables

Según se observa en la ventana hay una caja a la izquierda que tiene todas las variables que aparecen en el fichero activo para seleccionar aquella o aquellas que van a ser recodificadas; la forma de seleccionarlas es pinchar sobre ellas con el ratón y pulsar el botón ►, para que sean incluidas en el cajón que hay a la derecha (Var. Numérica→Var. de Resultado) en el que aparecen las variables a ser recodificadas. En dicho cajón debe aparecer obligatoriamente, para cada una de ellas, una variable destino de la recodificación. Así, seleccionada una variable para recodificar y llevada al cajón de Var. Numérica, Var. de resultado, aparece un símbolo de interrogación señalando que debe escribirse la variable de destino. Para ello bastará escribir, en la zona de Nombre, el nombre de una variable de SPSS, que puede ser nueva o ya existente; en cuanto esté escrito el nombre se iluminará el botón Cambiar que nos permitirá asociar la variable destino con la variable fuente ya seleccionada. Si la variable destino existe, los valores de ella quedarán machacados por el resultado de la recodificación y SPSS nos pedirá una confirmación para ello. En el caso de que no exista, en la variable nueva aparecerá el resultado de la recodificación, teniendo en cuenta que si hay algún valor fuente que no ha sido recodificado, aparecerá en la variable de destino como un dato faltante del sistema. En la ventana se ve que se puede especificar una etiqueta para la variable resultado de la recodificación, sujetándose a las normas ya descritas para las etiquetas.

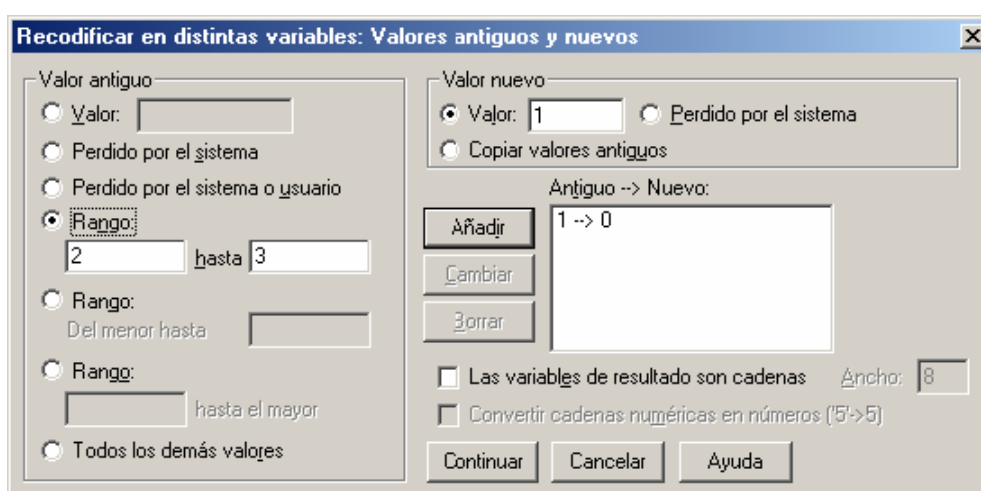


Figura 7: Especificación de valores antiguos y valores nuevos en una recodificación

A cada especificación de valores antiguos habrá que hacerle corresponder un valor nuevo (ver figura 7). Los tres valores nuevos posibles son: un único valor que especifica el usuario, o el valor

de perdido por el sistema o copiar el valor antiguo existente. En cuanto hayamos elegido una pareja valores\_antiguos/valor\_nuevo se iluminará el botón Añadir y pulsándolo añadiremos la regla de recodificación a la caja de las mismas que aparece debajo de la zona de valores nuevos. Con este proceso se van añadiendo las reglas de recodificación hasta la última. Si alguna regla no estuviera correcta se podría pulsar sobre ella y volvería a ser colocada sobre los valores antiguos y el valor nuevo para que hagamos las modificaciones oportunas pulsando el botón

### 3.6. Ordenación y selección de casos

Explicaremos a continuación una serie de procedimientos que suponen ya un manejo más sofisticado de un fichero de datos. En primer lugar explicaremos cómo se puede ordenar un fichero de datos en función de los valores que presenten una o varias variables. A continuación explicaremos como se pueden seleccionar casos de un archivo para llevar a cabo en ellos análisis estadísticos y no hacerlos con los casos no seleccionados. También veremos la ponderación de casos, lo que nos permitirá poder introducir y analizar datos con frecuencias. Por último explicaremos como se divide un fichero en subconjuntos de casos, en cada uno de los cuales se lleva a cabo el análisis elegido; a este procedimiento se le denomina Segmentación. Los tres procedimientos que vamos a explicar se encuentran dentro del menú Datos de la ventana de Edición de Datos.

#### 3.6.1. Procedimiento Ordenar casos

Una tarea muy frecuente, cuando uno ha acabado de grabar datos en un fichero, es la de ordenar los mismos por algunas variables que nos permitan revisarlos de una manera fácil. Eso se lleva a cabo con el Procedimiento Ordenar casos, que se invoca como *Datos, Ordenar casos*. Este procedimiento ordena los casos (las filas) del archivo de datos activo basándose en los valores de una o más variables de ordenación. Puede ordenar los casos en orden ascendente o descendente. Si selecciona más de una variable de ordenación, los casos se ordenarán respecto a cada variable dentro de las categorías de la variable anterior de la lista Ordenar por.

#### 3.6.2. Procedimiento Seleccionar casos

Seleccionar casos proporciona varios métodos para seleccionar un subgrupo de casos basándose en criterios que incluyen variables y expresiones complejas. También se puede seleccionar una muestra aleatoria de casos. Los criterios usados para definir un subgrupo pueden incluir: Valores y rangos de las variables, Rangos de fechas y horas, Números de caso (filas), Expresiones aritméticas, Expresiones lógicas y Funciones.

La forma de invocar la selección de casos es *Datos, Seleccionar casos*. Cuando se hace tal cosa aparece una ventana como la de la Figura 8. La ventana muestra, a la izquierda, una caja con las diferentes variables del fichero activo para que sean utilizadas en las distintas formas en que uno quiere seleccionar los casos. A la derecha aparecen las cinco formas de seleccionar casos; aquí describiremos las dos primeras. Todos los casos desactiva el filtrado y utiliza todos los casos (es realmente una no-selección de casos); ella está activa cuando se entra por primera vez a la ventana de selección de casos.

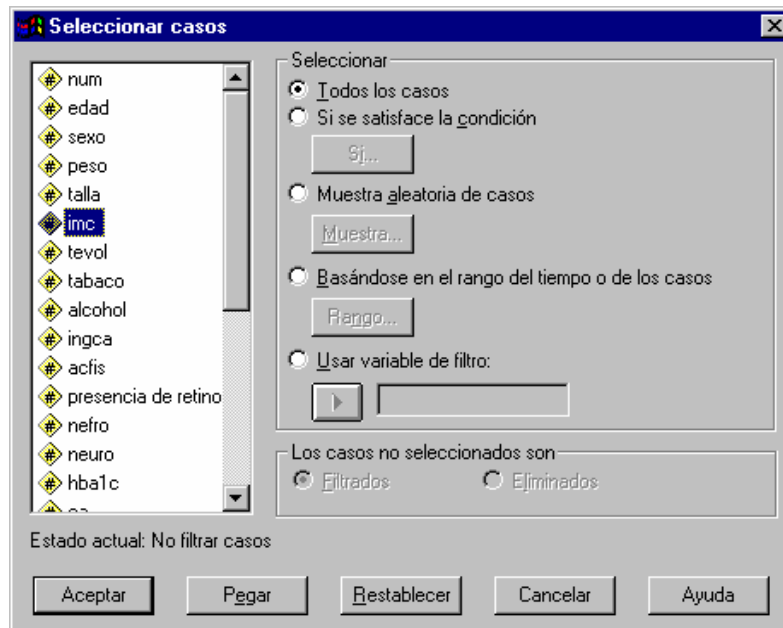


Figura 8: Selección de casos

Si se satisface la condición utiliza una expresión condicional para seleccionar los casos. Si el resultado de la expresión condicional es verdadero, el caso se selecciona. Si el resultado es falso o perdido, entonces el caso no se selecciona. La forma de escribir la expresión condicional es la misma que ya hemos descrito anteriormente y la forma de acceder a ella, una vez seleccionada la opción Si se satisface la condición, es pinchar en el botón Si... Una vez que hayamos escrito la condición y vuelto a la ventana de Seleccionar casos.. la condición aparecerá escrita a la derecha del botón Si...

### 3.7. El Visor de Resultados

Para ejecutar cualquier análisis estadístico habrá que invocar el menú análisis que se encuentra, entre otras, en la ventana del Editor de datos. Al desplegarse el menú se elegirá el análisis deseado y el resultado aparecerá en una ventana como la de la Figura 9. Esa ventana (llamada Visor de resultados de SPSS) es siempre la misma, aunque en el ejemplo que hemos descrito se refiere a los resultados de la descriptiva de unas variables de nuestro fichero activo. Lo primero que hay que decir de los resultados es que son una entidad aparte de los datos y que como tal son susceptibles de imprimirse, editarse y ser guardados en un fichero aparte, que tendrá la extensión .spo. Como toda ventana de SPSS, está compuesta por una zona de menús desplegables y barra de herramientas, y una zona dónde se presentan los resultados, zona que se divide a su vez en dos partes, una en la que aparece un índice de tales resultados y otra en la que aparecen los resultados propiamente dichos. A continuación describiremos la zona de resultados con detalle.

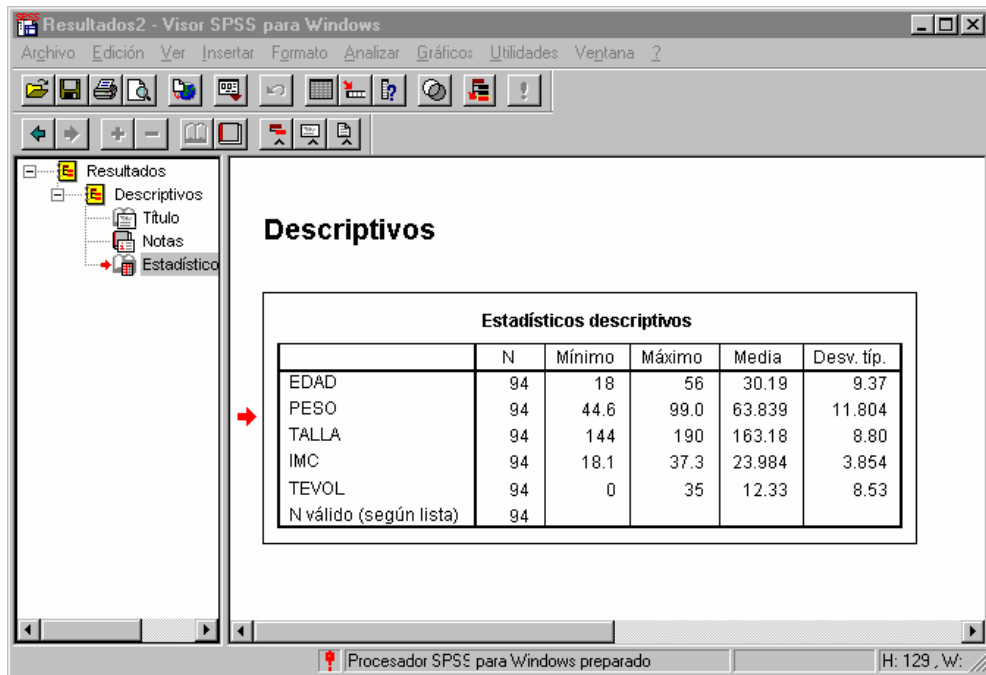


Figura 9: Visor de resultados

### 3.8. Análisis de datos

1. Estadística Descriptiva
  - Distribución de frecuencias
  - Creación de gráficos
  - Medidas de tendencia central o posición:
    - Media
    - Mediana
    - Moda
    - Centiles y cuarteles
  - Medidas de dispersión o variabilidad:
    - Recorrido
    - Desviación típica
    - Varianza
    - Asimetría (Fisher): indicador que permite establecer el grado de simetría o asimetría que presenta una distribución sin necesidad de representarla gráficamente.
    - Curtosis: estudia la distribución de frecuencias en la zona central. La mayor o menor concentración de frecuencias alrededor de la media y en la zona central de la distribución dará lugar a que ésta se más o menos apuntada. Mesocúrtica: normal, leptocúrtica: apuntamiento mayor, platicúrtica: apuntamiento menor.
  - Medidas de correlación:
    - Coeficiente de correlación de Pearson ( $r$ )
2. Estadística Inferencial (pruebas de significación estadística)
  - Ji cuadrado (cualitativa + cualitativa):
  - R de Pearson (cuantitativa + cuantitativa)
  - Prueba de t (cualitativa dicotómica + cuantitativa)
  - Análisis de Varianza (ANOVA) (cualitativa politómica + cuantitativa)
3. Análisis Multivariado
  - Regresión lineal

- Análisis Factorial
- 4. Estudio Psicométrico de instrumentos (fiabilidad y validez)
  - Análisis de consistencia interna: Alfa de Cronbach
  - Análisis de la capacidad de discriminación de los elementos: t de Student
  - Análisis Factorial

### 3.8.1. Estadística Descriptiva

A partir de ahora describiremos los procedimientos estadísticos de SPSS. Desde luego que no describiremos todos los que están disponibles en el programa, pues estas notas pretenden ser una iniciación al SPSS, pero sí daremos lo suficiente para que el usuario se enfrente a un volumen importante de problemas de Estadística.

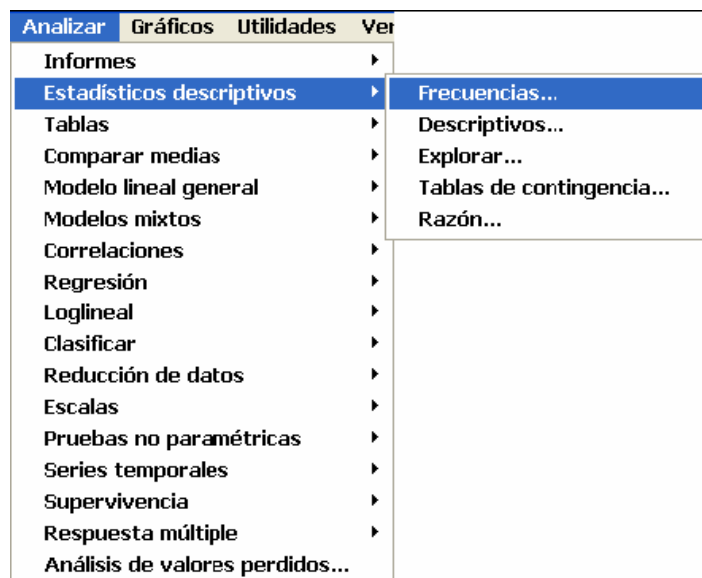


Figura 10: Menú Analizar con todos los procedimientos estadísticos disponibles en SPSS, y con el submenú de Estadísticos descriptivos desplegado.

Todos los procedimientos se invocan como menús dentro del menú *Analizar*, que al desplegarse nos da uno como el de la Figura 10. En ella se ve que hay una gran variedad de procedimientos, nosotros empezaremos por el menú *Estadísticos descriptivos*. Este menú tiene un submenú con cinco opciones: *Frecuencias...*, *Descriptivos...*, *Explorar...*, *Tablas de contingencia...* y *Razón...*. Por ahora estudiaremos únicamente la primera. A partir de ahora todos los procedimientos estadísticos que se presenten lo harán con un mismo formato; se presentará la ventana general, la ventana de opciones y el conjunto de resultados que se dan; en muchos casos las opciones no se harán en forma de una figura sino que se harán en forma de una tabla de texto en la que aparecerán las distintas alternativas de cada opción.

#### 3.8.1.1. Procedimiento Frecuencias

El procedimiento frecuencias se invoca como *Analizar, Estadísticos descriptivos, Frecuencias...*, y permite para cada una de las variables seleccionadas obtener la distribución de frecuencias de ellas, así como algunas medidas básicas de resumen a elegir y algunos gráficos también a elegir (ver figura 11)

En la ventana principal aparecen a la derecha los botones generales de SPSS. Además aparece la posibilidad de incluir la tabla de frecuencias de las variables elegidas sin más que marcar en el recuadro correspondiente a **Mostrar tablas de frecuencias**. En la parte baja de la ventana aparecen tres botones que nos permiten especificar o particularizar la descriptiva que queremos hacer.

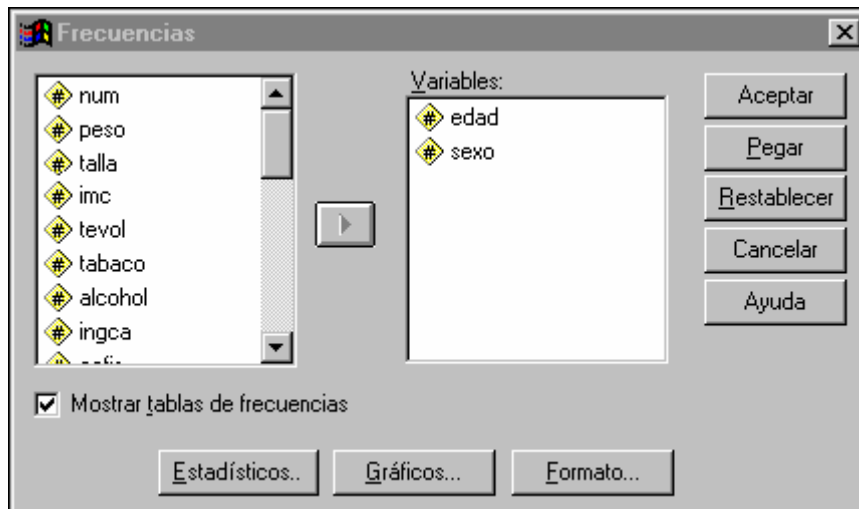


Figura 11: Análisis de frecuencias

El primero es el de **Estadísticos...**, que permite calcular estadísticos básicos de resumen para las variables seleccionadas. Pulsado dicho botón aparecerá una ventana como la de la Figura 12. En ella se ve que los estadísticos que se pueden elegir se agrupan en diferentes epígrafes que son: **Valores percentiles**, **Tendencia Central**, **Dispersión** y **Distribución** (éste último no lo veremos). Dentro de cada epígrafe aparecen las opciones que se corresponden con los estadísticos más básicos, por lo que no se requiere de explicación adicional. En cualquier caso, siempre podremos utilizar la ayuda contextual simplemente pinchando con el botón derecho del ratón sobre la opción. Por último aparece una opción que califica cómo son los datos; la marcaremos cuando tengamos datos agrupados y los valores presentes en la matriz de datos sean puntos medios de grupos (o marcas de clase). Esto afecta al cálculo de los Cuantiles, Percentiles y Mediana.



Figura 12: Opción Estadísticos en el procedimiento de frecuencias

El segundo botón opcional es el de Gráficos...; pulsándolo aparece una ventana en la que podemos optar por no presentar ningún gráfico o por presentar uno de los gráficos disponibles: barras, sectores o histograma. A este último se le puede superponer una curva Normal con la media y desviación típica de los datos. Finalmente, en la parte baja de la ventana podemos optar por hacer el gráfico para frecuencias absolutas o para porcentajes.

### 3.8.2. Estudio psicométrico de instrumentos: fiabilidad y validez

La precisión de los datos obtenidos del cuestionario y la estabilidad de la medida obtenida en diferentes aplicaciones es uno de los elementos básicos que ha de cumplir todo procedimiento de obtención de información. La calidad de los hallazgos obtenidos ha de atender a una serie de factores de los cuáles es preciso obtener información tras la implementación del instrumento. En este sentido Hyman (1984: 189) cita a Deming (1944) al advertir como factores básicos que afectan a la precisión y consistencia de la información los siguientes: la variabilidad de las respuestas, las diferencias entre los distintos tipos y grados de pedidos de opinión, parcialidades y variaciones derivadas del entrevistador, imperfecciones en el diseño del cuestionario, la falta de respuesta, errores de muestreo, errores en el procesamiento de la información y errores de interpretación. Es por ello que se considera preciso un estudio de la fiabilidad y validez del instrumento con el objeto de dotar a nuestros datos de la veracidad suficiente para no comprometer el estudio.

El objetivo de este apartado consiste en valorar el grado de consistencia del instrumento de medida a emplear (propiedades de la escala y de sus elementos constitutivos) y su adecuación al objeto de la medición.

En definitiva, para el estudio de la fiabilidad y validez hemos de realizar los siguientes análisis:

1. Análisis de Consistencia Interna, en el sentido de dotar de significación a los ítems de la prueba, es decir, conseguir que cada uno de ellos mida una porción del rasgo o característica que se desea estudiar. Para ello utilizaremos el coeficiente Alfa de Cronbach.
2. Análisis de la capacidad de discriminación de los elementos de modo que se refuerce el carácter unidimensional de la prueba. Se utilizará para ello la *t* de Student entre las medias de los grupos establecidos.
3. Estudio Factorial con el objetivo de explorar el instrumento y averiguar si existe una estructura dimensional en el cuestionario que pueda servir de base para la interpretación de los resultados.

#### 3.8.2.1. Análisis de consistencia interna

El coeficiente de fiabilidad del cuestionario vendrá determinado por el coeficiente Alfa de Cronbach basado en la correlación ínter elementos promedio. Con este tipo de análisis obtendremos la siguiente información:

- Media y desviación típica de los ítems eliminados.
- Coeficiente de Homogeneidad corregido para cada ítem.
- Coeficiente Alfa en caso de eliminación del ítem.



- Valores de Alfa para el conjunto de los sujetos que respondieron a la escala, en función de las dimensiones y en función de la Rama de Especialización.

Realizando una primera aproximación al estudio de la consistencia interna, podemos advertir cómo los valores correspondientes a cada una de las ramas de especialización descritas en la investigación (valores alfa por encima de 0.9) indican que las relaciones entre los diferentes elementos de la escala atendiendo a este criterio son muy elevados. Por su parte, el valor total de Alfa en la escala (0.9144) indica una correlación alta, un nivel elevado de estabilidad en las respuestas, por lo que el cuestionario presenta indicios de garantías de fiabilidad.

### 3.8.2.2. Discriminación de los ítems

Un ítem tiene poder discriminativo si es capaz de distinguir entre aquellos sujetos que puntúan alto en la prueba y los que puntúan bajo, es decir, si discriminan entre los que poseen un nivel alto en el rango medido y los que poseen un nivel bajo. El poder de discriminación de todos los ítems de un test reforzará el carácter unidimensional de la prueba, puesto que todos los ítems constituyen elementos que funcionan de modo análogo convirtiéndose en pequeños tests que informan sobre el mismo rasgo que el test global. Lo deseable es que la discriminación de los ítems sea elevada.

Para llevar a cabo este estudio se seleccionan los ítems cerrados de elección ordinales de manera que la suma total se recodifique en tres grupos (Bajo, Medio y Alto):

- 1 = Grupo Bajo (valor mínimo, percentil 33)
- 2 = Grupo Medio (percentil 34, percentil 66)
- 3 = Grupo Alto (percentil 67, valor máximo)

Para realizar este cálculo hemos de transformar las variables escalares y calcular su suma en una nueva variable (Transformar, Calcular), de modo que quede recodificada en los tres grupos anteriores.

De este modo la realización de la prueba de t para muestras independientes permitió establecer la existencia o no de diferencia estadística (n.s.=0.05) entre los grupos que puntúan bajo y alto en los ítems.

### Trabajo con SPSS: ANALIZAR, COMPARAR MEDIAS, PRUEBA DE T PARA MUESTRAS INDEPENDIENTES

**Hipótesis nula** (de igual probabilidad):  $H_0$ : No existen diferencias entre las medias de los grupos de (las medias de los grupos son iguales)  
 $p \geq 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy alta por lo tanto, **ACEPTAMOS LA HIPÓTESIS NULA, NO EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON IGUALES**  
 $P < 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy **BAJA** por lo tanto, **rechazamos LA HIPÓTESIS NULA, SI EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON DIFERENTES**

Todos aquellos valores de p menores que 0.05 representan un alto poder de discriminación por parte del ítem. Por su parte, aquellos valores de p iguales o mayores a 0.05 no permiten rechazar la hipótesis nula de igual discriminación y, por tanto, que el ítem discrimine, por lo que éste debería ser revisado.

### 3.8.2.3. Estudio Factorial

La técnica del análisis factorial es utilizada para reducir un amplio número de fenómenos, conceptos o variables, a un número más reducido de componentes o factores, de tal modo que sean representativos de esos conceptos. Así, lo que se busca es que todas las variables entre las que existe una relación se agrupen o saturen en un mismo factor, utilizándose como método exploratorio. Se supone que todas ellas poseen un atributo común que es etiquetado por el investigador, dando así nombre al factor.

Esta técnica se puede utilizar para estudiar la estructura de los instrumentos de manera que podamos afirmar que determinados ítems se explican mejor desde una dimensión que desde otra e incluso aconsejar la eliminación de ciertos reactivos dada su escasa consistencia interna con los demás ítems de las diferentes dimensiones identificadas.

Previa selección de los métodos de extracción (*componentes principales*) y rotación (*varimax*) ya explicados, procedemos a realizar el análisis factorial con los datos procedentes de los instrumentos.

#### Trabajo con SPSS: ANALIZAR, REDUCCIÓN DE DATOS, ANÁLISIS FACTORIAL

La primera fase, extracción de los componentes, tiene por objetivo determinar el número mínimo de factores comunes capaces de reproducir, de un modo satisfactorio, las correlaciones observadas entre las variables.

Realizaremos este proceso a partir del método de extracción de componentes principales, cuyo objetivo es maximizar la varianza explicada. El factor que mejor explique la dimensión analizada (el que represente mayor variabilidad) se convertirá en el primer componente principal y así sucesivamente. Su aplicación supone transformar directamente un conjunto de variables correlacionadas en un conjunto de variables no correlacionadas (ver figura 13).

Dado que el principal objetivo es explicar la varianza común entre las variables (comunalidad) con el menor número de factores (parsimonia) debemos, en primer lugar, comprobar a través del estudio de las comunalidades que el total de la variabilidad de nuestra matriz será explicada por todos los componentes extraídos.

A continuación el objetivo consiste en maximizar la explicación de esa varianza con el menor número de factores, aspecto que determinará el total de factores a extraer. A partir de la rotación varimax, que logra extraer de forma ortogonal el valor de la correlación de la varianza en el factor, trataremos de hacer más sencilla la interpretación de los componentes (ver figura 14).

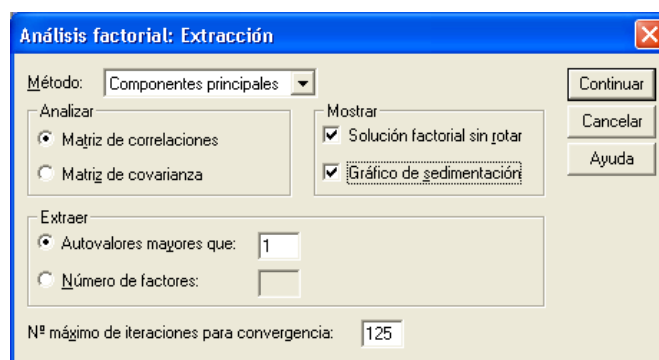


Figura 13: Métodos de extracción en el Análisis Factorial

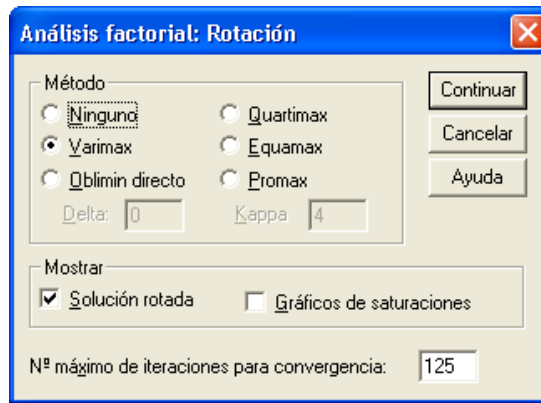


Figura 14: Métodos de rotación en el Análisis Factorial

### 3.8.3. Pruebas de significación estadística

Al realizar pruebas de hipótesis, se parte de un valor supuesto (hipotético) en parámetro poblacional. Después de recolectar una muestra aleatoria, se compara la estadística muestral, así como la media ( $\bar{x}$ ), con el parámetro hipotético, se compara con una supuesta media poblacional ( $\mu_0$ ). Después se acepta o se rechaza el valor hipotético, según proceda. Se rechaza el valor hipotético sólo si el resultado muestral resulta muy poco probable cuando la hipótesis es cierta.

Las técnicas para verificar la existencia de diferencias estadísticamente significativas entre diferentes grupos poblacionales son las siguientes, atendiendo al tipo de variables que en cada momento se empleen son:

- **Prueba de t para muestras independientes:** se emplea cuando la variable independiente es cualitativa dicotómica (dos grupos de comparación) y las variables dependientes son cuantitativas.

*Hipótesis nula (de igual probabilidad):  $H_0$ : No existen diferencias entre las medias de los grupos (las medias de los grupos son iguales)*

$p \geq 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy alta por lo tanto, ACEPTAMOS LA HIPÓTESIS NULA, NO EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON IGUALES  
 $P < 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy BAJA por lo tanto, rechazamos LA HIPÓTESIS NULA, SI EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON DIFERENTES

**Trabajo con SPSS: ANALIZAR, COMPARAR MEDIAS, PRUEBA DE T PARA MUESTRAS INDEPENDIENTES**

- **Análisis de varianza (ANOVA):** se utiliza cuando la variable independiente es cualitativa politómica (más de dos grupos de comparación) y las variables dependientes son cuantitativas.

*Hipótesis nula (de igual probabilidad):  $H_0$ : No existen diferencias entre las medias de los grupos (las medias de los grupos son iguales)*

$p \geq 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy alta por lo tanto, ACEPTAMOS LA HIPÓTESIS NULA, NO EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON IGUALES

$P < 0.05$  la probabilidad de que se cumpla la hipótesis nula es muy BAJA por lo tanto, rechazamos LA HIPÓTESIS NULA, SI EXISTEN DIFERENCIAS SIGNIFICATIVAS ENTRE LAS MEDIAS DE LOS GRUPOS, LAS MEDIAS SON DIFERENTES

#### Trabajo con SPSS: ANALIZAR, COMPARAR MEDIAS, ANOVA DE UN FACTOR

- Ji cuadrado: Prueba no paramétrica que establece la existencia de diferencias significativas en los valores obtenidos al comparar dos variables cualitativas. Esta prueba de bondad de ajuste compara las frecuencias observadas y las esperadas en cada categoría para contrastar si todas las categorías contienen la misma proporción de valores o si cada categoría contiene una proporción de valores especificadas por el investigador o investigadora.

#### Trabajo con SPSS: ANALIZAR, PRUEBAS NO PARAMÉTRICAS, CHI CUADRADO

## 4. Bibliografía y recursos

- Calvo Gómez, F. (2006). *Técnicas estadísticas multivariantes aplicadas con SPSS-Windows*. Bilbao: Universidad de Deusto.
- Camacho, J. (2005). *Estadística con SPSS para Windows, versión 12*. Madrid: Ra-ma.
- Gardner, R.C. (2003). *Estadística para psicología usando SPSS para Windows*. México: Pearson Educación.
- Gil Pascual, J.A. (2006). *Estadística e informática SPSS en la investigación descriptiva e inferencial*. Madrid: UNED.
- Gondar, J.E. (2003). *AF, análisis factorial y de varianza con SPSS*. Madrid: Data Mining Institute.
- Gondar, J.E. (2003). *IS. Metodología de la investigación estadística con SPSS*. Madrid: Data Mining Institute.
- Lizasoain, L. y Joaristi, L. (2003). *Gestión y análisis de datos con SPSS*. Madrid: Thomson Paraninfo.
- Martín, Q. (2007). *Tratamiento de datos con SPSS: prácticas resueltas y comentadas*. Madrid: Paraninfo
- Pardo, A. y Ruiz, M.A. (2005). *Análisis de datos con SPSS 13 base*. Madrid: Mc-Graw-Hill.
- Pérez López, C. (2004). *Técnicas de análisis multivariante de datos con SPSS*. Madrid: Pearson
- Pérez, C. (2001). *Técnicas estadísticas con SPSS*. Madrid: Prentice Hall.

Pérez, C. (2005). *Técnicas estadísticas con SPSS 12: aplicaciones al análisis de datos*. Madrid: Pearson Alambra.

Portilla, M.L. (2008). *Manual práctico del paquete estadístico SPSS para Windows*. Pamplona: Universidad Pública de Navarra.

Rodríguez, C.; Gallardo, M.A.; Pozo, T. y Gutiérrez, J. (2006). *Iniciación al análisis de datos cuantitativos en educación. Análisis descriptivo básico: teoría y práctica mediante SPSS*. Granada: Grupo Editorial Universitario.

Rodríguez, C.; Pozo, T. y Gutiérrez, J. (2007). *Fundamentos conceptuales y desarrollo práctico con SPSS de las principales pruebas de significación estadística en el ámbito educativo*. Granada: Grupo Editorial Universitario.

Visauta, B. (2007). *Análisis estadístico con SPSS 14: estadística básica*. Madrid: McGraw-Hill

SPSS. Guía para el análisis de datos

<http://www2.uca.es/serv/ai/formacion/spss/Inicio.pdf>

Introducción al SPSS. Aula Virtual de Bioestadística

[http://e-estadistica.bio.ucm.es/web\\_spss/indice\\_spss.html](http://e-estadistica.bio.ucm.es/web_spss/indice_spss.html)

Guía de Ayuda SPSS

<http://psico.fcep.urv.es/spss/index0.htm>