



# Práctica 1

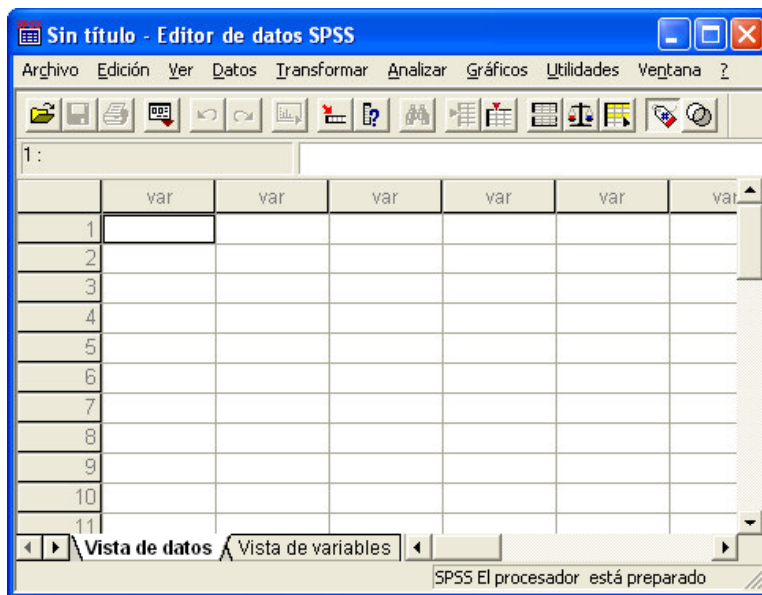
## Introducción al SPSS

### Objetivos

En esta práctica se pretende familiarizar al alumno con el entorno de trabajo del programa de análisis estadístico SPSS. Para ello se verán diferentes tipos de ventanas, menús y cuadros de diálogo que pueden surgir en una sesión de trabajo. Además se pretende que el alumno aprenda a crear sus propios ficheros de datos y a trabajar con variables, ya sea creándolas nuevas o transformando algunas ya existentes.

### Introducción

El programa estadístico que utilizaremos es el SPSS para Windows. Para ejecutar el programa pulsar dos veces con el ratón sobre el icono correspondiente al SPSS. Aparece la ventana principal del programa.



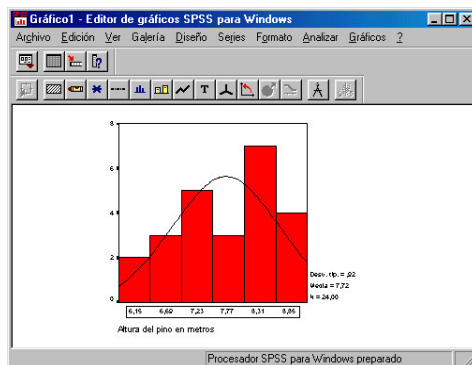
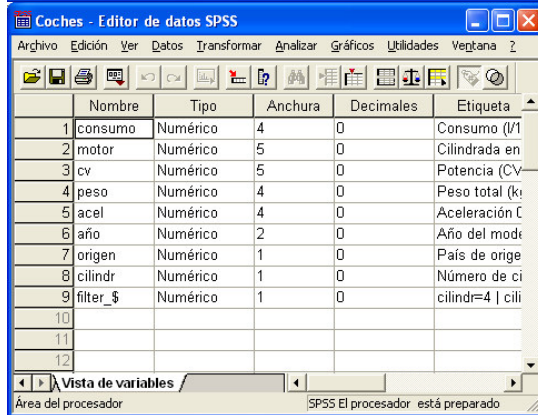
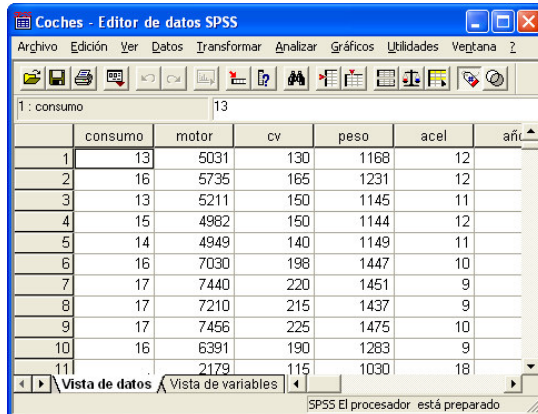
Esta ventana nos muestra:

- **Menú principal**, que permite acceder a todos los comandos y tipos de análisis que puede realizar el programa
- **Barra de herramientas**. Conjunto de iconos para realizar tareas específicas
- **Editor de datos**, que contiene el conjunto de datos que estudiamos.
- **Barra de estado**, que nos muestra información sobre el estado del programa. Además contiene dos pestañas que permiten seleccionar *Vista de datos* o *Vista de variables*.

El menú principal permite acceder a todas las opciones de que dispone el programa. Estas opciones aparecen agrupadas en varios encabezados:

<b>Archivo</b>	Permite abrir, guardar, imprimir ficheros de datos
<b>Edición</b>	Contiene opciones para copiar, cortar, recuperar datos o texto, buscar datos, etc. Además contiene las opciones de configuración del programa.
<b>Ver</b>	Contiene diferentes opciones referentes al aspecto de la ventana principal, tales como diferentes barras de herramientas, etc
<b>Datos</b>	Se utiliza para realizar cambios en los ficheros de datos tales como crear variables, combinar ficheros, trasponer variables y casos, ponderar casos
<b>Transformar</b>	Se utiliza para realizar cambios en variables del fichero de datos y crear nuevas variables basadas en valores de otras variables que ya existen.
<b>Analizar</b>	Contiene las opciones que dan paso a los diferentes análisis estadísticos que podemos hacer con el programa.
<b>Gráficos</b>	Se utiliza para crear gráficos.
<b>Utilidades</b>	Contiene opciones para ver información sobre el contenido del fichero de datos y de las variables, para definir conjuntos de variables, etc
<b>Ventana</b>	Permite cambiar la ventana activa en una sesión de trabajo.
<b>?</b>	Ayuda sobre el SPSS

## Ventanas del SPSS



### Editor de datos

Está en la ventana principal del SPSS. Contiene el fichero de datos. Además permite crear ficheros o modificar uno existente, crear variables, seleccionar variables y casos. Si hay un fichero de datos abierto y en la barra de estado se ha seleccionado *Vista de datos* vemos el contenido del fichero (los valores de las variables). Si por el contrario se ha seleccionado *Vista de variables* vemos las características de las variables del fichero.

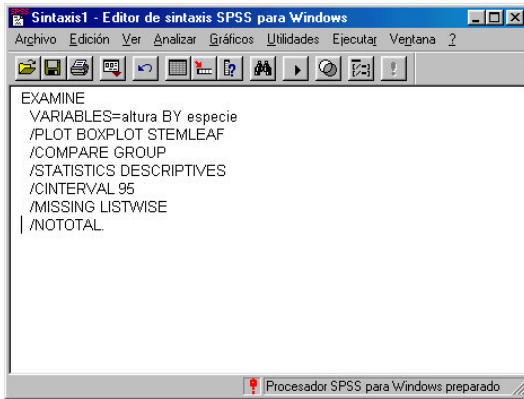
Los ficheros de datos de SPSS llevan como extensión SAV.

### Visor de resultados

Contiene los resultados de la ejecución de comandos del SPSS, tanto de lectura de datos como de transformación y análisis de datos, con salidas de texto y gráficas. Además permite modificar los resultados, añadir comentarios y exportar los resultados. Los resultados de una sesión de trabajo se pueden guardar en un fichero que tiene extensión SPO.

### Editor de gráficos

Este editor aparece cuando queremos modificar las características de un gráfico. Se puede activar pulsando dos veces seguidas con el ratón sobre cualquier gráfico del visor de resultados.



### Editor de sintaxis

Se utiliza para ejecutar comandos del SPSS. Estos comandos pueden estar en un fichero, pueden escribirse directamente en ese editor o pueden proceder de la opción *Pegar* de alguna de las ventanas de diálogo del programa. Estos comandos se pueden guardar en un fichero para utilizarlos en otro momento de la sesión de trabajo. Para ejecutar los comandos que muestra la ventana se pincha en el icono ▶. Los archivos de sintaxis de SPSS llevan por extensión SPS.

### Ventanas de diálogo.

Además de las ventanas que ya hemos explicado, existen otras que aparecen cuando se ejecuta algún procedimiento del SPSS: las ventanas de diálogo. Estas ventanas contienen distintos campos que hay que completar. El acceso a estos campos se puede realizar con el ratón. Si accedemos al menú *Edición* y elegimos *Opciones* en el submenú obtenemos una de estas ventanas de diálogo. Existen diferentes formas de introducir información a través de estas ventanas: listas donde se pueden elegir todas las opciones, listas que sólo admiten una de las opciones, cuadros que debemos completar, etc.

Elegir sólo una opción

Completar

Al pulsar en la flecha se abre un cuadro con varias opciones

Podemos elegir más de una opción o no elegir ninguna

## Introducción de datos

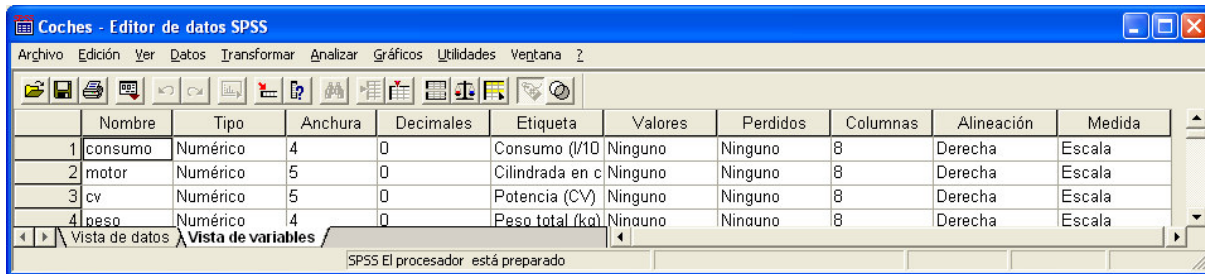
Los datos de la muestra se pueden organizar en una matriz de datos de dimensión  $n \times m$ , donde  $n$  es el tamaño de la muestra y  $m$  es el número de variables estudiadas. Cada fila de dicha matriz corresponde a un **caso**, es decir, son los datos asociados a un individuo de la muestra. Cada columna de la matriz corresponde a una **variable**. Cada casilla de esta matriz corresponde con un **dato**.

	$X_1$	$X_2$	...	$X_m$
Caso 1				
Caso 2				
⋮				
Caso n				

El editor de datos del SPSS está organizado de esta forma. Al pulsar con el ratón sobre una casilla se selecciona el dato concreto. Por otra parte los bordes del editor son sensibles, es decir, si pulsamos con el ratón sobre el nombre de una variable se selecciona dicha variable, mientras que al pulsar sobre el número de un caso se selecciona todo el caso.

Cuando introducimos una variable, además del nombre y los propios datos, el programa nos permite introducir información adicional: nombre de la variable, tipo de datos que contiene, formato de presentación de estos datos en el editor, etiqueta para la variable, etiquetas para los valores de la variable (variables cualitativas) y definición de valores perdidos.

Para definir una nueva variable podemos hacer doble *click* sobre el encabezamiento de una columna vacía o seleccionar en la barra de estado la pestaña *Vista de variables*.



Cada fila del editor de variables se corresponde con las características de las variables, que en el SPSS son 10: nombre, tipo, anchura, decimales, etiqueta, valores, perdidos, columnas, alineación y medida.

- **Nombre de la variable.** Como máximo podemos usar 8 caracteres siendo el primero de ellos una letra. No se pueden usar símbolos como &, /, \$ ni espacios en blanco. Tampoco pueden ser nombre de variables expresiones como ALL, AND, BY, EQ, GE, GT, LE, NE, NOT, OR, TO o WITH son palabras reservadas.
- **Tipo.** Tipo de variable que puede ser: numérico (se puede especificar el número de decimales), coma (para usar el punto como separador de decimales y la coma de separación cada tres posiciones), punto (el contrario del anterior), fecha, dólar, moneda personalizada, cadena (podemos especificar el ancho)
- **Etiqueta.** Permite introducir información sobre la variable con un máximo de 120 caracteres. Es altamente recomendable utilizar esta opción ya que con el nombre de la variable no siempre queda claro cuál es el contenido u otros aspectos tales como las unidades de medida.
- **Valores.** Esta opción permite añadir una etiqueta para cada uno de los diferentes valores de la variable. La etiqueta puede tener como máximo 60 caracteres. Una práctica habitual consiste en definir variables cualitativas como numéricas y asignarle etiquetas a los números.
- **Perdidos.** Se utiliza para especificar la ausencia de datos o valores perdidos. A menudo es útil para saber por qué se pierde información. Por ejemplo, podemos querer distinguir el dato perdido correspondiente a un entrevistado que se niega a responder del dato perdido debido a que la pregunta no afectaba a dicho entrevistado. Los valores de datos especificados como perdidos por el usuario aparecen marcado ara un tratamiento especial.
- **Columna.** Permite establecer el ancho de la columna que ocupará la variable en el editor
- **Alineación.** Permite establecer la alineación de los datos dentro de las casillas.

- **Medida.** Nos permite especificar el tipo de escala de medida de la variable: *escala*, *ordinal* o *nominal*. Utilizaremos *escala* para variables cuantitativas, mientras que *nominal* y *ordinal* se utiliza para variables cualitativas. Si los datos representan categorías con un cierto orden intrínseco (bajo, medio, alto o suspenso, aprobado, notable, sobresaliente) la variable es de escala *ordinal*. Si entre las categorías no existe ese orden (hombre, mujer) la variable es *nominal*.

Una vez definidas unas variables, podemos introducir los datos de las mismas. Estos datos los podemos introducir por variables (dato + ENTER) o por casos (seleccionamos primero el caso y luego dato + ENTER).

Una vez que disponemos de un conjunto de datos los podemos almacenar en un fichero para un uso posterior

Menú *Archivo* → Opción *Guardar o Guardar como*

En la ventana de diálogo que aparece podemos elegir el directorio que contendrá el archivo, el nombre del archivo y el tipo de archivo. Entre los diferentes tipos tenemos: archivos SPSS (\*.SAV), archivos ASCII fijo (\*.DAT) que guardan sólo la matriz de datos, ficheros de la hoja de cálculo EXCEL (\*.XLS) o ficheros de base de datos dBASE (\*.DBF).

## Ejercicio 1:

Una empresa de talado de árboles para la elaboración de madera ha tomado una serie de características de 24 pinos elegidos al azar. Estas características son: especie de pino, diámetro del pino (medido a 1.33 metros del pie y expresado en centímetros), altura del pino (en metros), volumen de madera obtenido (en decímetros cúbicos) y calidad de la madera. Los datos se presentan en la tabla 1.

Crear un fichero de datos SPSS que contenga los datos de los pinos estudiados definiendo las variables con el nombre, tipo y formato convenientes.

Pino	Especie	Diámetro	Altura	Volumen	Calidad
1	pinea	20.05	8.52	130.57	normal
2	pinea	19.90	7.15	126.79	normal
3	pinea	18.64	6.43	105.39	mala
4	pinea	20.46	6.21	100.47	regular
5	pinea	18.04	8.13	109.76	normal
6	pinea	19.33	6.12	91.18	mala
7	pinaster	19.34	6.45	198.56	superior
8	pinaster	25.08	8.73	204.87	superior
9	pinaster	24.71	8.52	201.70	normal
10	pinaster	24.91	8.83	187.06	extra
11	pinaster	24.56	8.53	193.80	normal
12	pinaster	23.91	8.94	191.07	extra
13	pinaster	23.46	8.40	182.25	superior
14	pinaster	25.07	8.91	212.97	superior
15	pinaster	23.07	8.19	195.70	superior
16	pinaster	24.03	8.56	203.45	regular
17	laricio	19.86	7.55	115.62	regular
18	laricio	23.45	7.74	113.69	superior
19	laricio	21.37	7.84	130.98	normal
20	laricio	19.48	7.41	108.75	superior
21	laricio	18.83	7.40	115.88	mala
22	halapensis	20.06	6.54	96.34	mala
23	halapensis	19.57	7.07	98.73	normal
24	halapensis	20.12	7.17	103.44	superior

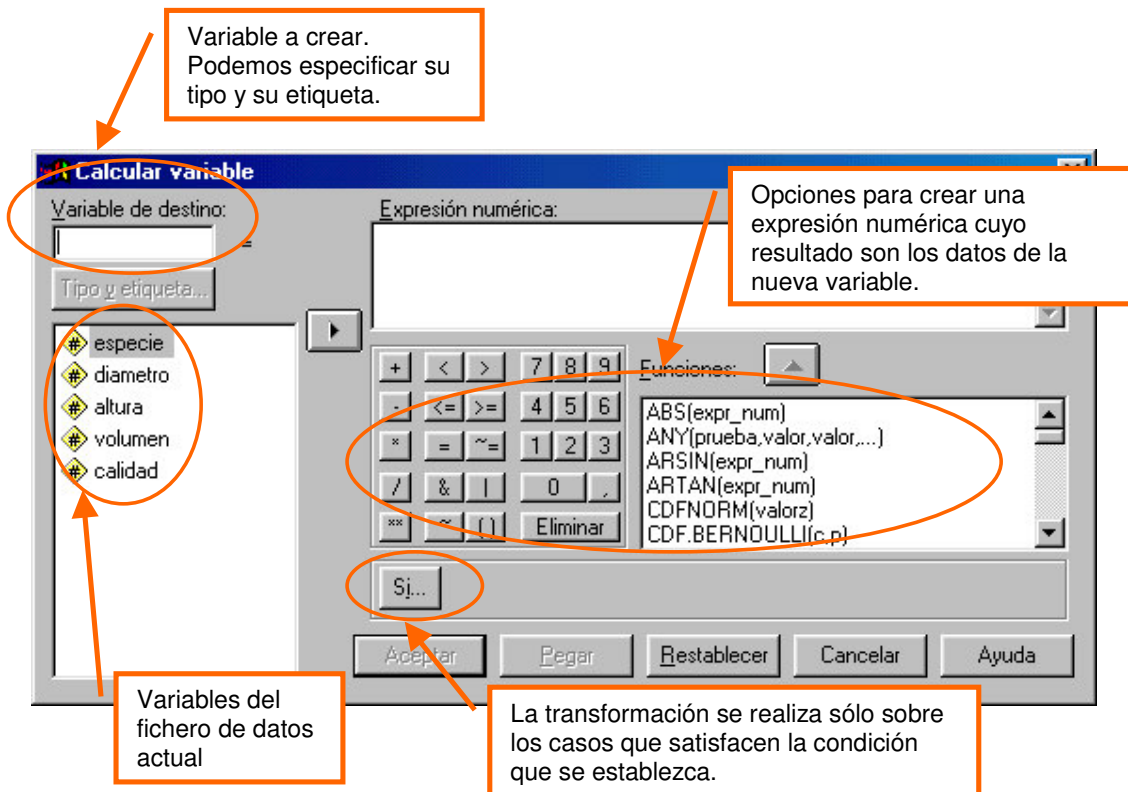
## Transformación de los datos.

A lo largo de una sesión de trabajo con el SPSS es posible modificar los datos originales. Tales cambios pueden ser a nivel de variables, casos o a todo el fichero. Puede interesarnos tanto la recodificación de los valores de una variable como la generación de nuevas variable mediante la transformación de algunas ya existentes en el archivo de datos.

### Crear variables

Podemos crear nuevas variables usando transformaciones de la siguiente manera

Menú *Transformar* → Opción *Calcular*



En expresión numérica debemos escribir una expresión válida para el SPSS cuyo resultado sean los valores de la variable destino. Para construir estas expresiones pueden usarse nombres de variables (podemos escribirlos o seleccionarlos del listado que aparece), constantes, operadores y funciones. Existen operadores aritméticos como +, -, \*, /, funciones matemáticas como SQRT, LN, ABS, estadísticas como MEAN o SD, de probabilidad, etc.

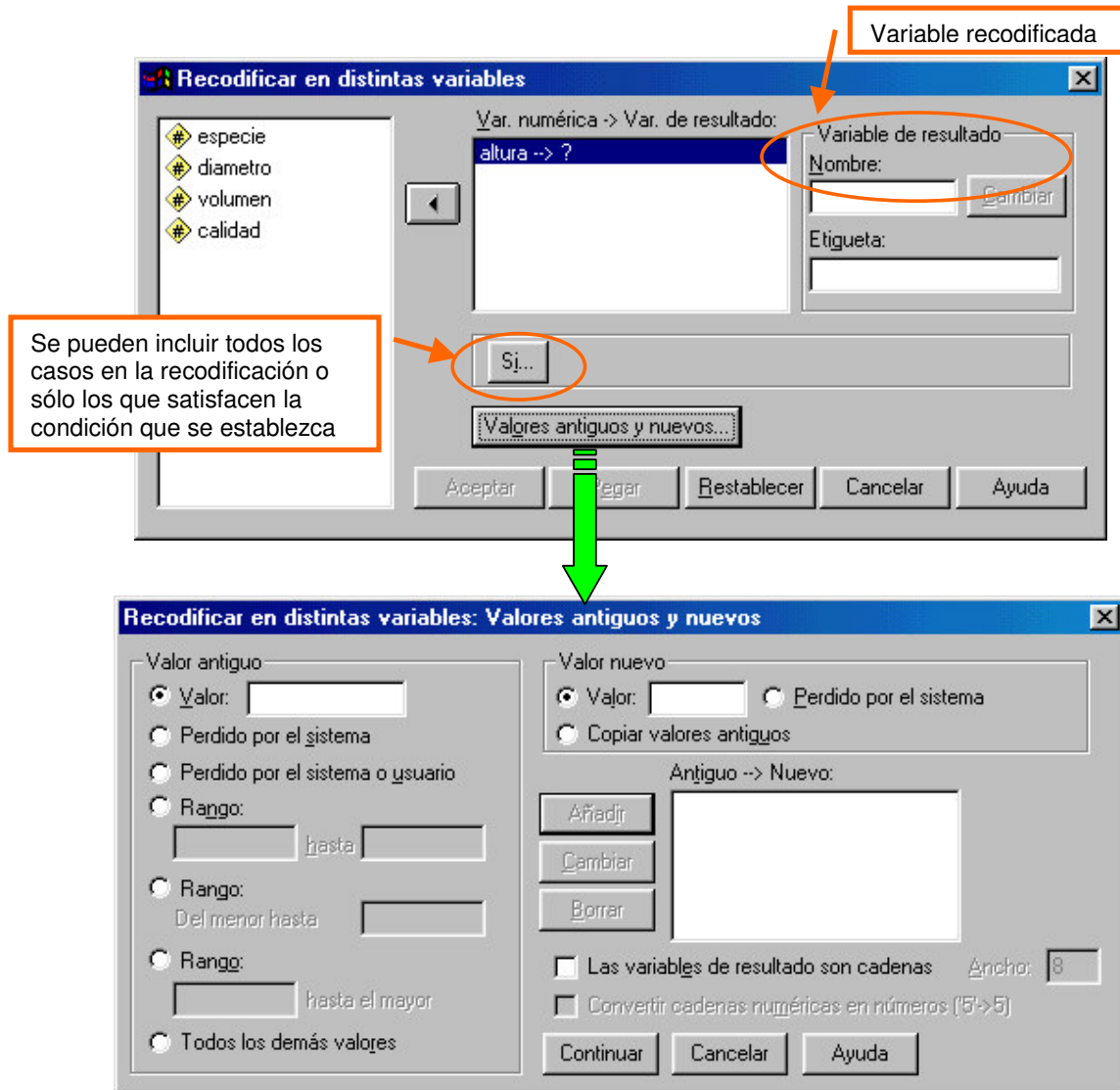
La expresión se calcula para todos los valores de las variables implicadas en ella, y los resultados constituirán los valores de la variable destino. A veces no interesa que la transformación afecte a todos los casos lo cual puede conseguirse a través de expresiones lógicas definidas en el cuadro de diálogo que surge al pulsar en el botón con la expresión *Si...*

### Recodificación de variables

Podemos volver a codificar variables de la siguiente manera

Menú *Transformar* → Opción *Recodificar*

Mediante la recodificación podemos discretizar una variable continua o podemos unir modalidades de variables cualitativas o discretas. La recodificación puede realizarse sobre la misma variable o se puede crear una nueva. Con la opción *Valores antiguos y nuevos* se aparece una ventana de diálogo donde se recodifica la variable. Los valores antiguos se pueden expresar como valores concretos o como rangos y los valores nuevos como valores concretos.

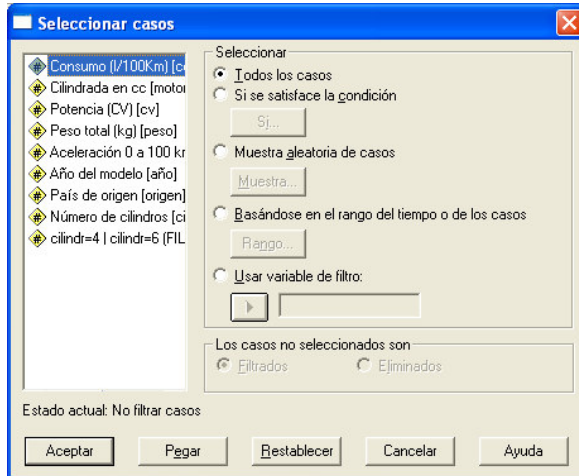


## Filtrado de datos

El programa SPSS permite seleccionar determinados casos para así poder aplicar algún análisis sobre los individuos seleccionados. La selección se realiza sobre la base de un criterio lógico o de una decisión aleatoria. Podemos realizar tales selecciones siguiendo los pasos

Menú *Datos* → Opción *Seleccionar casos*

Aparece el cuadro de diálogo:



La selección de casos puede ser temporal (*filtrados*) o permanente (*eliminados*). Las diferentes opciones son:

- Si se satisface la condición. Quedarán seleccionados sólo aquellos casos que cumplan una condición lógica que especificaremos pulsando en el botón *Si...* El programa añade una variable *filter\_\$* para controlar el filtrado de los datos.
- Muestra aleatoria de los casos. El programa selecciona aleatoriamente un porcentaje de casos especificado por el usuario.
- Basándose en el rango de tiempo o de los casos. Permite seleccionar casos consecutivos o basados en una variable fecha previamente definida.

También es posible usar una variable filtro previamente creada por el usuario. Aquellos casos que tengan valor 0 o perdido en la variable filtro no serán seleccionados

## Ejercicio 2:

Consideremos el fichero que contiene información acerca de los pinos del ejercicio 1.

- Crear una variable que contenga el volumen de madera expresado en metros cúbicos.
- Crear una nueva variable de calidad de la madera agrupando los grupos mala y regular en calidad baja, y los grupos superior y extra en calidad alta. El grupo calidad normal será ahora calidad media.

# Distribuciones de probabilidad

## Generación de muestras aleatorias

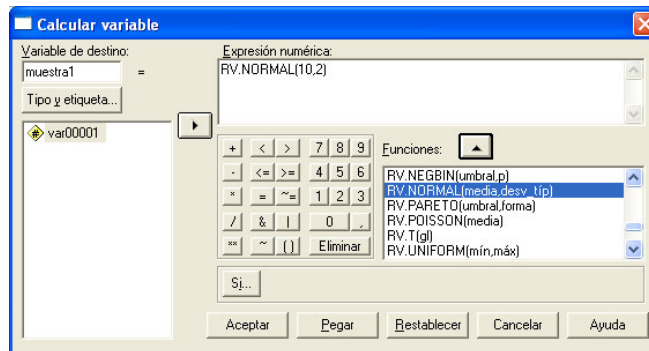
El programa SPSS permite generar muestras aleatorias procedentes de diversas distribuciones conocidas. El programa genera los datos de una distribución y los sitúa en una columna (variable) con la longitud que tenga el fichero de datos, es decir, genera tantos valores aleatorios como casos existan en el fichero. Si queremos obtener un muestra y el editor de datos está vacío, debemos, en primer lugar, dimensionar el conjunto de datos. Por ejemplo, para generar 100 datos nos situamos primero en la casilla 100 (*Datos / Ir a caso...*) de la primera columna e introducimos un número cualquiera.

El programa dispone de funciones específicas para cada distribución. Algunas de ellas son las siguientes

<i>RV.BINOM</i> ( <i>n, prob</i> )	Devuelve un valor aleatorio de la distribución binomial donde <i>n</i> es el número de pruebas y <i>prob</i> la probabilidad de éxito en cada prueba
<i>RV.EXP</i> ( <i>forma</i> )	Devuelve un valor aleatorio de una distribución exponencial con el parámetro de forma especificado
<i>RV.NORMAL</i> ( <i>media, desv_tip</i> )	Devuelve un valor aleatorio de una distribución normal con media y desviación típica especificadas
<i>RV.UNIFORM</i> ( <i>min, max</i> )	Devuelve un valor aleatorio de una distribución uniforme en el intervalo [ <i>min, max</i> ]

**Ejemplo:** Deseamos crear una muestra aleatoria de tamaño 100 procedente de una distribución normal de media 10 y desviación típica 2 y guardarla en una variable de nombre *muestra1*.

Dimensionamos adecuadamente el editor de datos tal y como se comentaba en párrafos anteriores. Seleccionamos *Transformar / Calcular* y completamos la ventana que nos aparece de la forma siguiente. Pulsamos con el ratón el botón *Aceptar* y el programa genera la muestra que se añade en la primer columna libre del editor de datos.



## Función de distribución.

En este apartado veremos como obtener los valores de una función de distribución conocida, es decir, dada una variable aleatoria con distribución conocida queremos obtener el valor  $F(x) = P(X \leq x)$ . Para ello seleccionamos *Transformar / Calcular*, en el vamos variable destino escribimos el nombre de la variable que va a contener la probabilidad y en la lista de funciones elegimos alguna de las que comienzan con CDF y especificamos sus parámetros.

Por ejemplo, para obtener la probabilidad de que una variable aleatoria con distribución normal de media 15 y desviación típica 1 sea menor que 17 en el campo Expresión numérica debemos escribir *CDF.NORMAL(17, 15, 1)*.

Es importante notar que el valor  $P(X \leq a)$  coincide con el valor  $P(X < a)$  en el caso de distribuciones continuas pero no así en el caso de distribuciones discretas.

El programa también tiene funciones para calcular percentiles, es decir los inversos de la función de distribución, es decir, dada una variable aleatoria  $X$  con distribución conocida y dado un valor de probabilidad  $p$ , encontrar el valor  $a$  tal que se verifica  $P(X \leq a) = p$ . Para calcular estos valores se procede de forma similar a la anterior pero se utilizan las funciones SPSS que comienzan por *IDF*. Por ejemplo, la expresión *IDF.NORMAL(0.95, 0, 1)* nos da el valor de una normal estándar que deja por debajo de él 0.95 de probabilidad.

## Función de densidad.

SPSS no permite obtener directamente una función de densidad. En este apartado veremos como aplicar los cálculos del apartado anterior para poder obtener una representación gráfica de una función de densidad.

**Ejemplo:** Obtener una representación de la función de densidad de una distribución normal  $N(2, 1)$  en el intervalo  $[-5, 5]$ .

Generamos en primer lugar una serie de valores en dicho intervalo. En concreto utilizaremos 100 valores comenzando por el  $-4.9$  y con incrementos de  $0.1$ . Por tanto dimensionamos el editor de datos a 100 casos y elegimos *Transformar / Calcular*. En el campo variable de destino ponemos  $x$  y en expresión numérica

$$0.1 * \$casenum - 5$$

donde  $\$casenum$  es una variable interna del programa que da para cada caso el número correspondiente. Una vez obtenido estos valores entramos de nuevo en *Transformar / Calcular* y ponemos en expresión numérica

$$(CDF.NORMAL(x, 2, 1) - CDF.NORMAL(lag(x), 2, 1)) / 0.1$$

donde  $lag(x)$  es una función que proporciona para cada caso el valor de  $x$  en el caso anterior y el cociente sobre  $0.1$  nos da el promedio sobre la longitud del intervalo. En variable destino ponemos *densidad* como nombre de la variable que contendrá los valores del cálculo anterior.

Finalmente, seleccionamos *Gráficos / Líneas* y la opción *Valores individuales de los casos*. Introducimos la variable  $x$  en *Etiquetas de las categorías* y la variable *densidad* en el campo *La línea representa*.

## Ejercicio 3:

- Generar muestras de tamaño 50 para las distribuciones  $N(2, 4)$ ,  $t$  de Student con 10 grados de libertad,  $Beta(3, 7)$
- Simular 50 lanzamientos de una moneda no trucada.
- Obtener el valor de  $P(X \leq 3)$ ,  $P(X \leq 5)$  y  $P(X = 9)$  para una distribución binomial de parámetros  $n = 20$  y  $p = 0.3$
- Obtener los valores  $P(X \leq -1)$  y  $P(X > 2)$  para una  $N(2, 3)$ .
- Obtener los percentiles 90, 5 y 50 de una distribución  $N(10, 3)$
- Obtener la representación gráfica de la función de densidad de una distribución  $Beta(2,5)$ .
- Obtener la representación gráfica de la función de probabilidad de una distribución binomial de parámetros  $n = 20$  y  $p = 0.4$ . (NOTA: Usar el tipo de gráfico de entre los que ofrece el SPSS que consideres más adecuado para una función de probabilidad)