

ESTADÍSTICA – 1º AMBIENTALES

PRIMERA PRÁCTICA: ESTADÍSTICA DESCRIPTIVA (GUIADA & RESUELTA)

El objetivo de esta práctica es hacer un análisis descriptivo de una colección de datos con el programa SPSS. Para ello, se plantean dos ejercicios: uno sobre el cálculo de estadísticos básicos y representaciones gráficas para una única variable, y otro para hallar los estadísticos correspondientes a dos variables y a la relación entre ambas.

Un paso previo a la resolución de los ejercicios es introducir los datos al programa SPSS: los datos pueden introducirse manualmente, o bien se facilitan en un archivo, éste puede venir dado ya en formato SPSS (extensión **.sav**), o en otros como texto (**.txt** ó **.csv**) o Excel (**.xlsx**)

PRIMERA PARTE: ESTADÍSTICA DESCRIPTIVA PARA UNA VARIABLE

Descripción de los datos: El archivo **cobre .txt** muestra la concentración del nutriente Cu (en mg/kg) en una muestra de suelo agrícola estudiada por 27 laboratorios diferentes. Los datos se obtuvieron mediante la extracción con agua regia. (Ref.: Ellison, S. L. R., Barwick, V. J. y Duguid Farrant, T. J. (2009). *Practical Statistics for the Analytical Scientist*. RSC Publishing, Cambridge.)

EJERCICIO 1:

- Calcular los estadísticos : media, mediana, varianza, desviación típica y cuartiles.
- Obtener el histograma y el diagramas de cajas (*boxplot*).
- Comentar las conclusiones que se obtienen: diferencias entre media y mediana, a qué se deben, conclusiones del histograma y del *boxplot*...

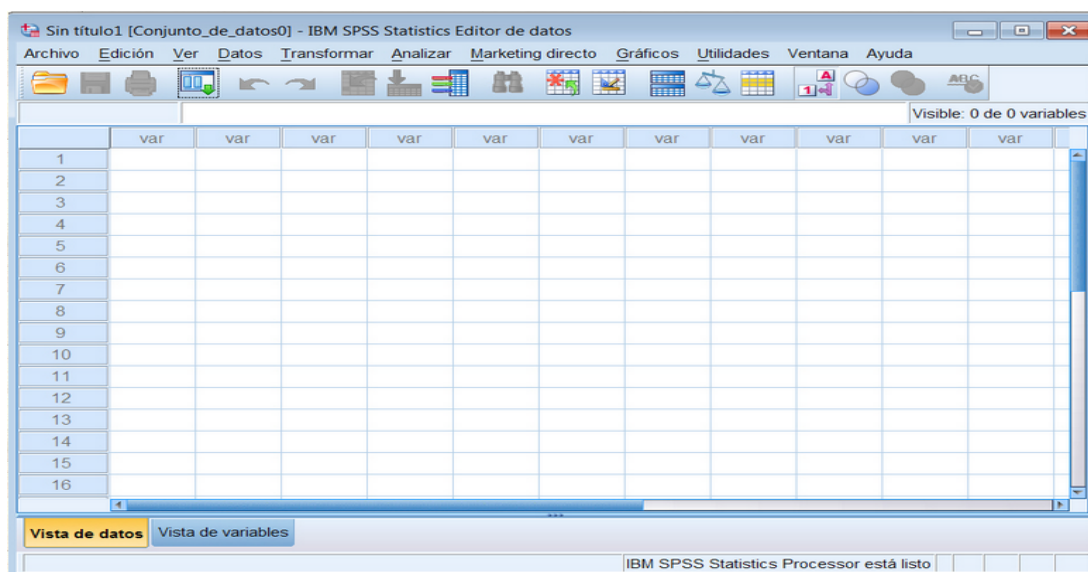
INTRODUCCIÓN DE DATOS EN EL PROGRAMA SPSS

1. DESCARGA DE DATOS DE MOODLE AL ORDENADOR (p.e., al escritorio):

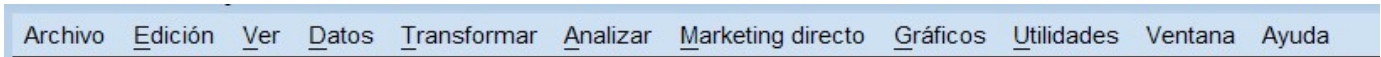
Se trata del fichero de texto **cobre .txt** en el que aparecen las concentraciones de Cu determinadas por los distintos laboratorios. Para ello posicionar el ratón sobre **cobre .txt** en *moodle*, pinchar con el botón derecho y marcar *Guardar* (ó *File*).

2. ARRANQUE DEL PROGRAMA SPSS:

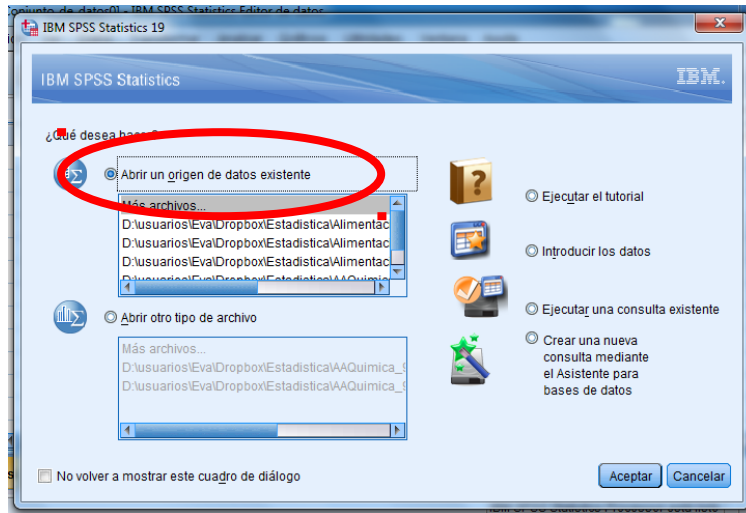
INICIO → PROGRAMAS → IBM SPSS STATISTICS → IBM SPSS STATISTICS 23.



Esta ventana: el **Editor de datos de SPSS**; es el marco de trabajo inicial que se utiliza para introducir los datos y elegir el procedimiento adecuado para el análisis. Esta ventana está formada por la *Barra de menú* que contiene el *Menú principal* de SPSS con todas sus opciones y que usaremos en la resolución de la práctica



- Si se abre el siguiente cuadro de diálogo:



elegir: **ABRIR UN ORIGEN DE DATOS EXISTENTE** (seguramente venga marcado por defecto)

- Si no aparece dicho recuadro, en la ventana de Editor de Datos ir a:

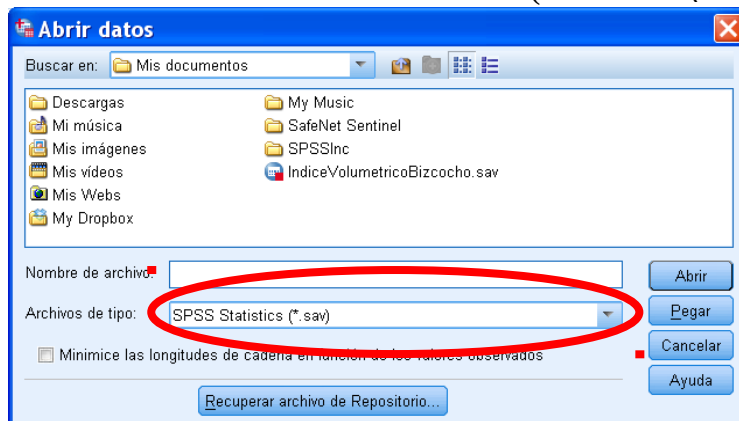
ARCHIVO → ABRIR → DATOS

Por cualquiera de los dos caminos se abrirá el cuadro de diálogo de más abajo. Por defecto viene marcada la extensión de SPSS:

ARCHIVOS DE TIPO: SPSS Statistics (*.sav)

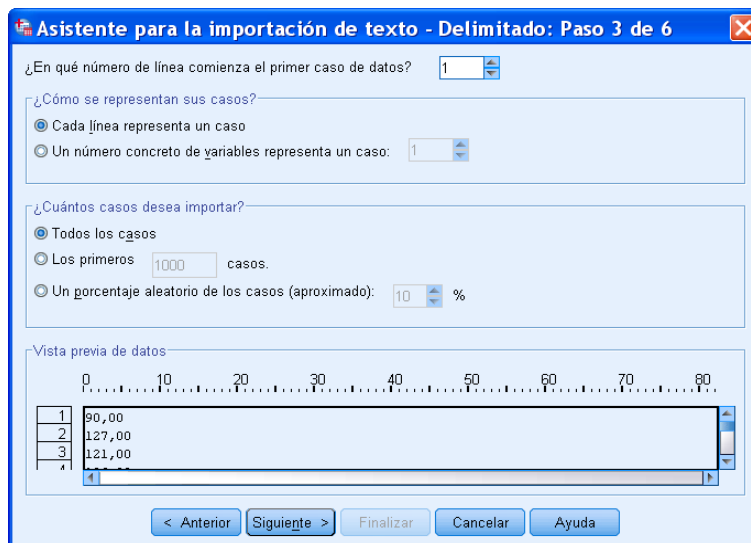
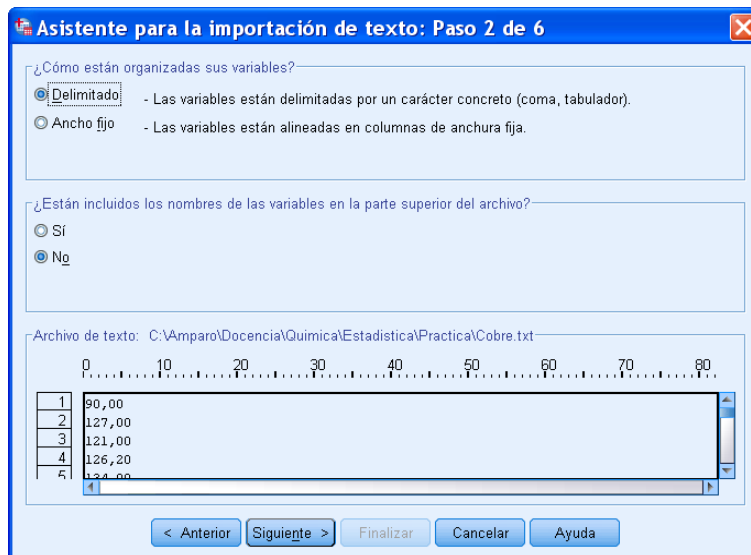
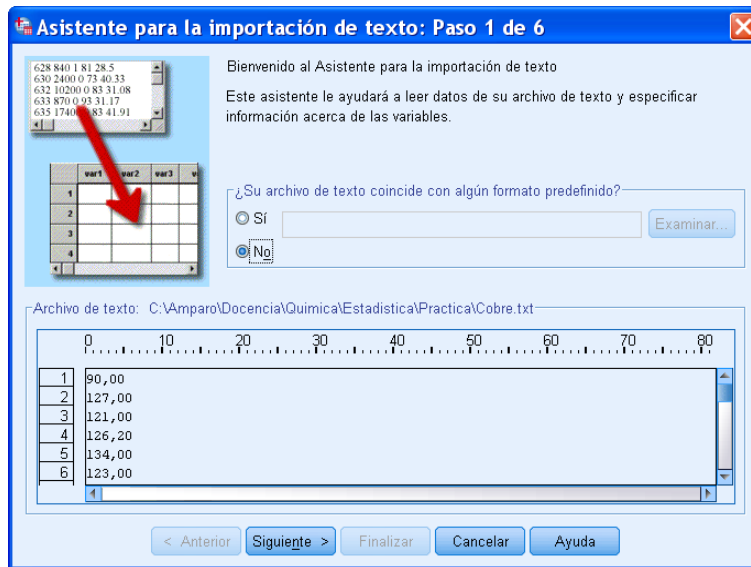
Debe cambiarse, ya que el archivo cobre.txt no es .sav. Para ello, se pincha la pestaña y se selecciona

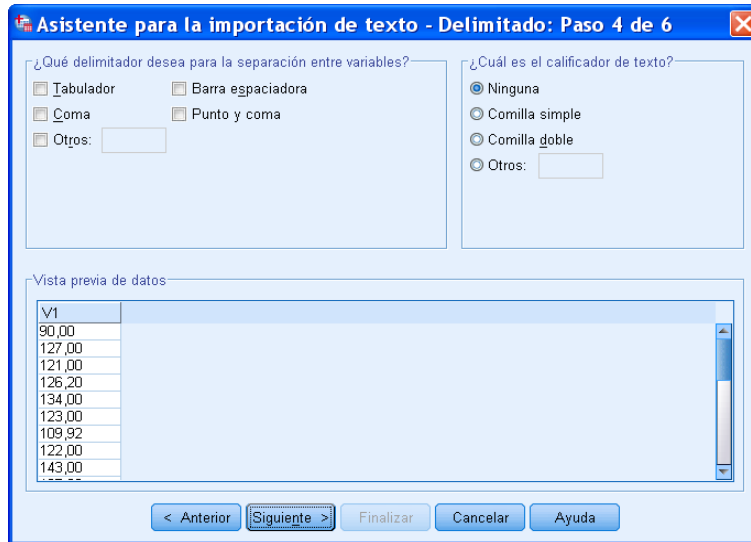
ARCHIVOS DE TIPO: TODOS LOS ARCHIVOS (O TEXTO (*.txt,*.dat))



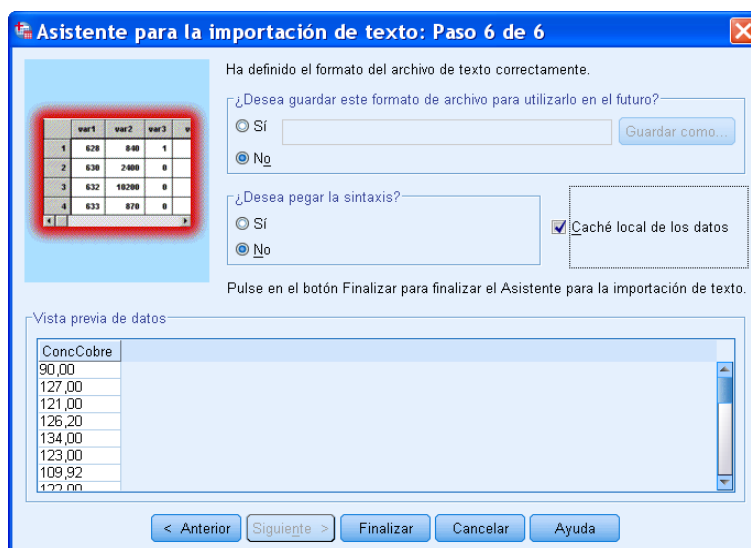
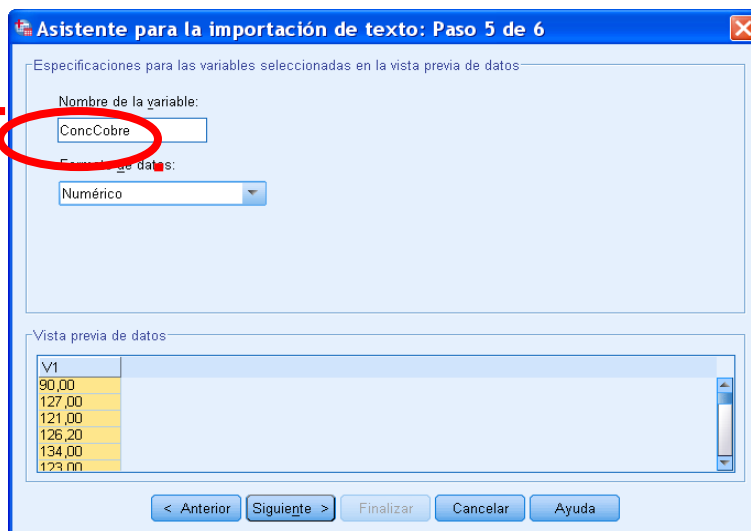
finalmente se abre el archivo **cobre.txt**.

A continuación aparecen unas ventanas donde se pregunta por algunas características del archivo de texto: deben leerse con detenimiento (para entenderlas) y seguirse los pasos que se muestran:





Por defecto SPSS nombra a la variable como **V1**. Debe cambiarse a un nombre más apropiado. El nombre de las variables debe ser *sencillo* (sin espacios, sin acentos...), p. e. **ConcCobre**.



3. Aparecen dos pantallas que se mantendrán durante todo el manejo de SPSS:

❖ La primera pantalla es, como ya hemos dicho, EL EDITOR DE DATOS:

	ConcCobre	var	var	var	var	var	var	var
1	90,00							
2	127,00							
3	121,00							
4	126,20							
5	134,00							
6	123,00							
7	109,92							
8	122,00							
9	143,00							

en el que es conveniente seleccionar ARCHIVO → GUARDAR COMO ... y guardarlo con extensión .sav (recuérdese, ese es el formato de SPSS) por si se tuviera que rescatar.

En la parte inferior del editor aparecen dos pestañas: VISTA DE DATOS y VISTA DE VARIABLES. La pestaña VISTA DE DATOS corresponde a la imagen anterior y nos muestra los nombres y valores de las variables.

La pestaña VISTA DE VARIABLES corresponde a la siguiente imagen y muestra las características de cada variable (Nombre, Tipo, Anchura,...).

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	ConcCobre	Numérico	6	2		Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
2											
3											

❖ La segunda pantalla es el VISOR DE RESULTADOS:

```

GET DATA
  /TYPE=TEXT
  /FILE="C:\Amparo\Docencia\Quimica\Estadistica\Practica\Cobre.txt"
  /DELCASE=LINE
  /DELIMITERS=""
  /ARRANGEMENT=DELIMITED
  /FIRSTCASE=1
  /IMPORTCASE=ALL
  /VARIABLES=
  ConcCobre F6.2.
DATASET NAME Conjunto_de_datos1 WINDOW=FRONT.

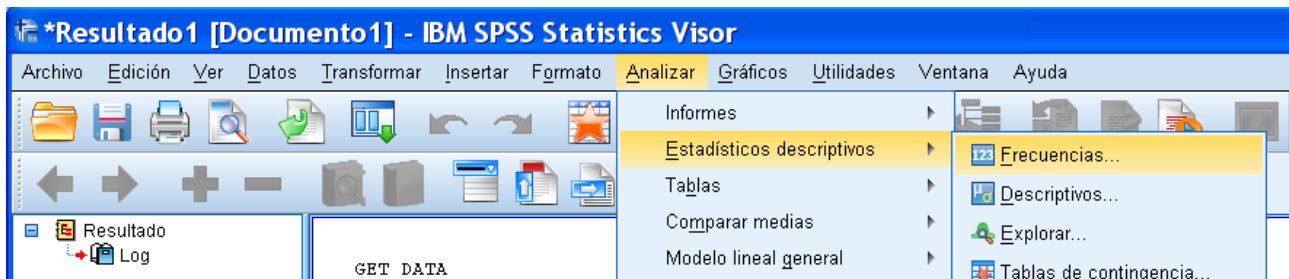
```

en el que irán apareciendo los resultados del análisis estadístico (tablas, gráficos,...). Una vez terminado el ejercicio **los resultados se guardan en un fichero de extensión .spv**. Obsérvese que esta ventana VISOR también contiene la *Barra de menú*, como la ventana EDITOR.

GUÍA PARA REALIZAR EL EJERCICIO 1 DE LA PRÁCTICA:

4. Para calcular los **ESTADÍSTICOS DESCRIPTIVOS** de los datos (*apartado (a)*) se selecciona en el Editor de datos o en el Visor de resultados (es indiferente):

ANALIZAR → ESTADÍSTICOS DESCRIPTIVOS → FRECUENCIAS.

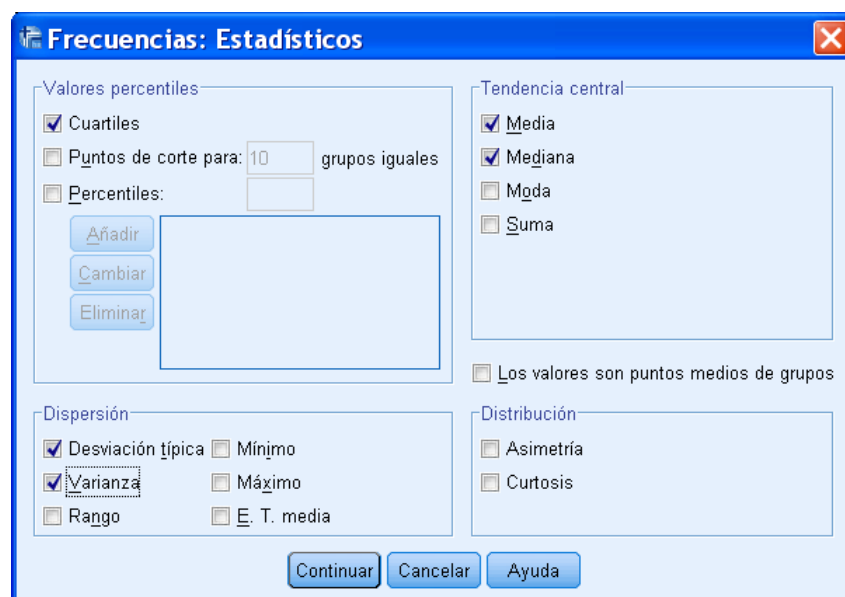


En el cuadro de diálogo



se pasa ConcCobre a la columna VARIABLES pinchando en la flecha.

Indicación: Conviene quitar la selección de la casilla **MOSTRAR TABLAS DE FRECUENCIAS** En el botón **GRÁFICOS** se selecciona **NINGUNO** (seguramente aparece así por defecto) y en **ESTADÍSTICOS** se marcan los que interesan: media, mediana, varianza, des. típica y cuartiles.



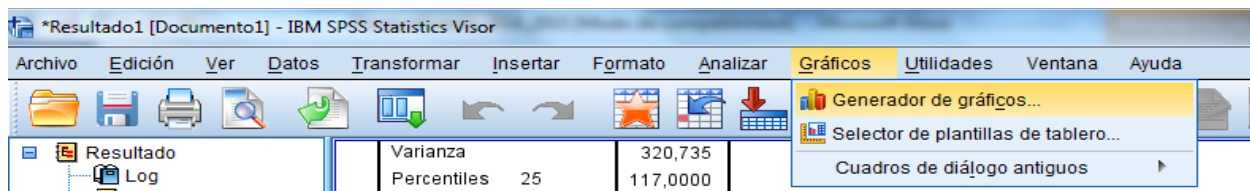
Se hace clic en ACEPTAR y en la ventana de VISOR debe aparecer la siguiente tabla con todos los estadísticos pedidos:

Estadísticos		
ConcCobre		
N	Válidos	27
	Perdidos	0
Media		123,4781
Mediana		124,1600
Desv. típ.		17,90908
Varianza		320,735
Percentiles	25	117,0000
	50	124,1600
	75	135,0000

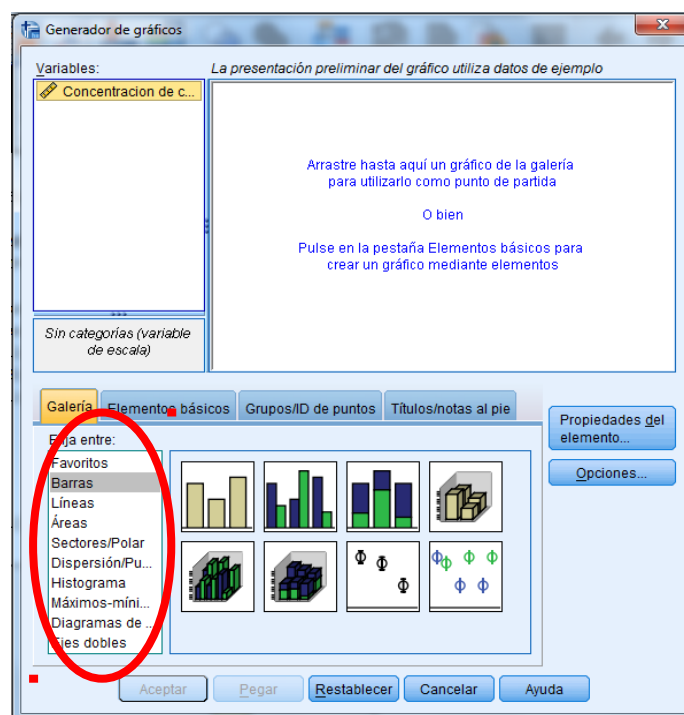
Observación: Percentil 25 = Primer Cuartil, Percentil 50 = Mediana y Percentil 75 = Tercer Cuartil

5. Para las **REPRESENTACIONES GRÁFICAS** (apartado (b)) se pinchar en

GRÁFICOS → GENERADOR DE GRÁFICOS

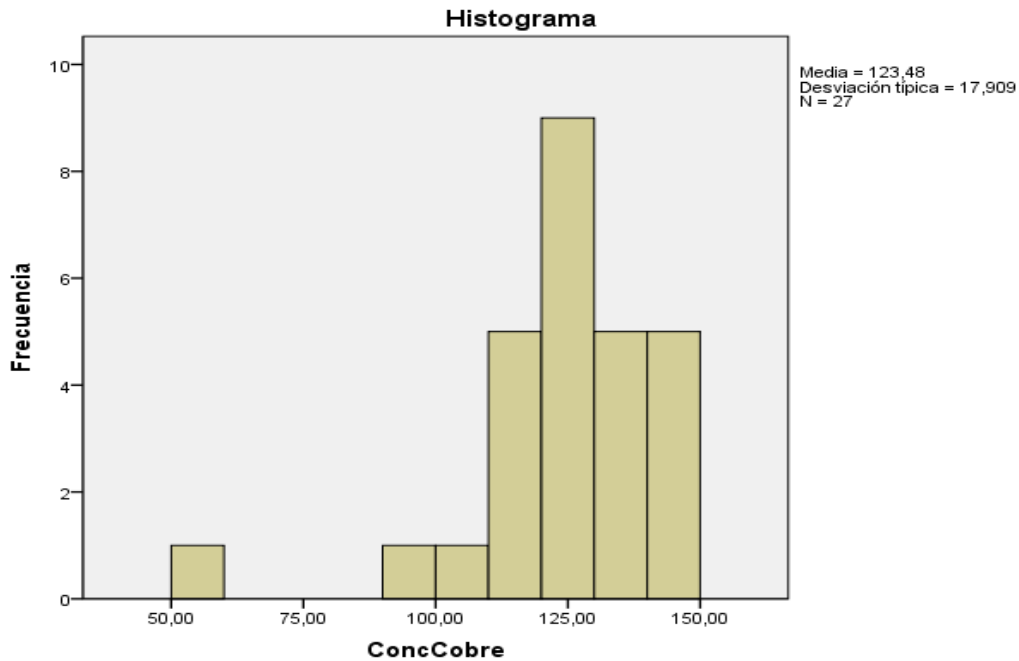


Aparecerá la siguiente ventana:



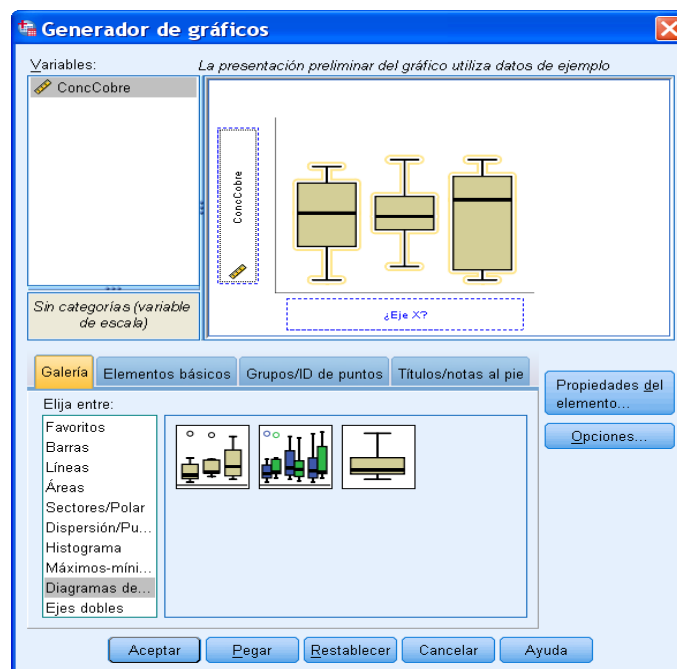
se elige el que interesa:

- ❖ **HISTOGRAMA:** Se pincha en HISTOGRAMA y se arrastra el icono de HISTOGRAMA SIMPLE a la esquina superior derecha y la variable ConcCobre al Eje X (En el Eje Y se muestran las frecuencias). Se pincha en ACEPTAR y entonces aparece, dibujado en el Visor, el histograma:

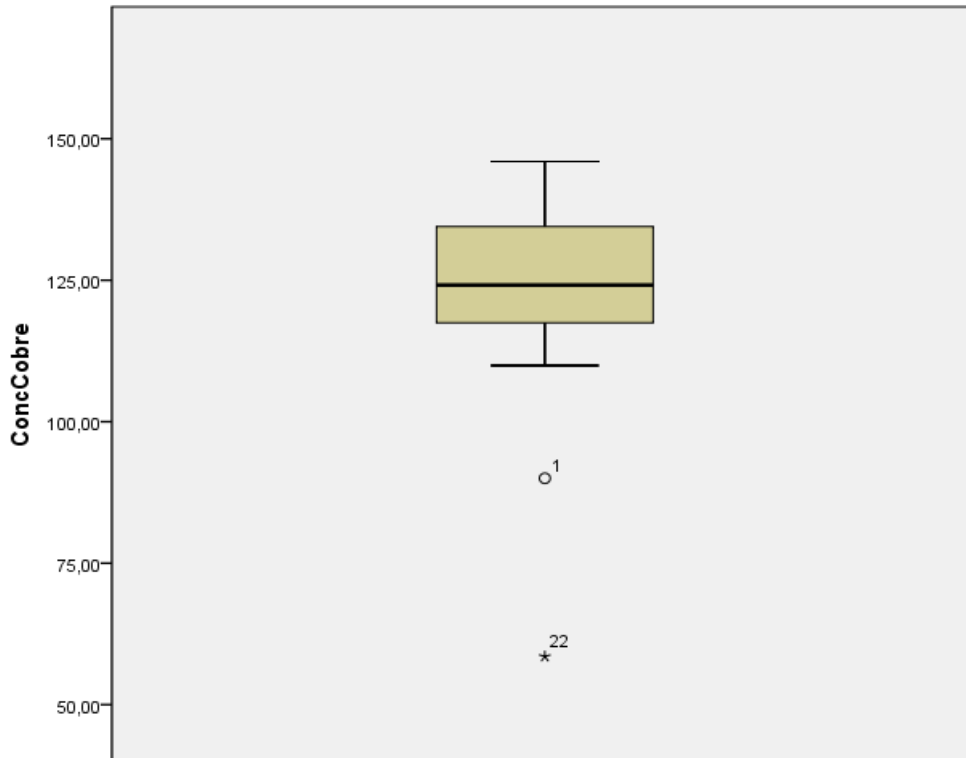


Un doble clic en el dibujo lo muestra en una pantalla en la que se puede modificar el gráfico con otro doble clic en aquello que se quiera modificar (color, tamaño de fuente, etc). Por ejemplo, con un doble clic sobre el dibujo del histograma aparece la siguiente ventana, en la que se puede seleccionar un número distinto de intervalos para el histograma:

- ❖ **DIAGRAMA DE CAJAS:** Se pincha en DIAGRAMAS DE CAJA, se arrastra el icono correspondiente a una única variable a la esquina superior derecha y la variable ConcCobre al Eje Y. Luego pinchar en ACEPTAR.



Se selecciona el diagrama de cajas adecuado para el análisis planteado y debe verse:



CONCLUSIONES: Al analizar la media y la mediana, se comprueba que son ligeramente distintas, esto es debido a los “datos atípicos” que arrastran “hacia abajo” el valor de la media, ya que ésta es muy sensible a ellos.

En el histograma se observa que la mayoría de los datos se concentran en la parte derecha de éste (en los valores altos), es decir, casi todos los valores se concentran en una única zona dejando otros valores (menores) aislados, por ello puede afirmarse que existe una alta dispersión y que su desviación típica es alta.

En el *boxplot* se comprueba que Q2 (la mediana) está más próximo a Q1 que a Q3, es decir: entre Q2 y Q3 los datos se encuentran más dispersos, mientras que entre Q1 y Q2 se encuentran más próximos, por lo que los valores están más concentrados. No obstante, la caja es más o menos simétrica.

Además, las observaciones 1 y 22 corresponden a datos que quedan fuera del *boxplot*, es decir, son valores atípicos.

SEGUNDA PARTE: ESTADÍSTICA DESCRIPTIVA PARA DOS VARIABLES

Descripción de los datos: Se utilizará el archivo SPSS: metabolismo.sav que ya tiene una estructura de base de datos SPSS. Las variables son: la *Tasa metabólica* (Tasa) y la *Masa corporal magra* (Masa) de una muestra de 7 hombres y 12 mujeres (variable Sexo).

La tasa metabólica basal, o TMB, se refiere a la cantidad de calorías se queman diariamente a través de una actividad mínima. Este número se basa en factores como la altura, el peso, el sexo y la edad.

EJERCICIO 2:

- Comparar los diagramas de cajas de la variable Tasa en hombres y en mujeres ¿cuál de las dos poblaciones tiene su centro más alto?
- Obtener el diagrama de dispersión de las variables Masa y Tasa (sin distinguir entre sexos), ¿Para qué sirve este gráfico de los datos? ¿Qué se observa en él?
- ¿Cuál es la recta de regresión estimada $Y=a+bX$?
- ¿Parece bueno el ajuste de la recta de regresión a los datos? ¿Por qué?

INTRODUCIR DATOS EN EL PROGRAMA SPSS

6. Descargar los datos de Moodle al disco duro del ordenador (por ejemplo, al escritorio): se posiciona el ratón sobre metabolismo.sav en moodle, se pincha con el botón derecho y se elige *Guardar*.

7. Para empezar a realizar el ejercicio se puede pinchar sobre el propio archivo metabolismo.sav una vez guardado. Si no se abre, debe utilizarse el desplegable Archivo.

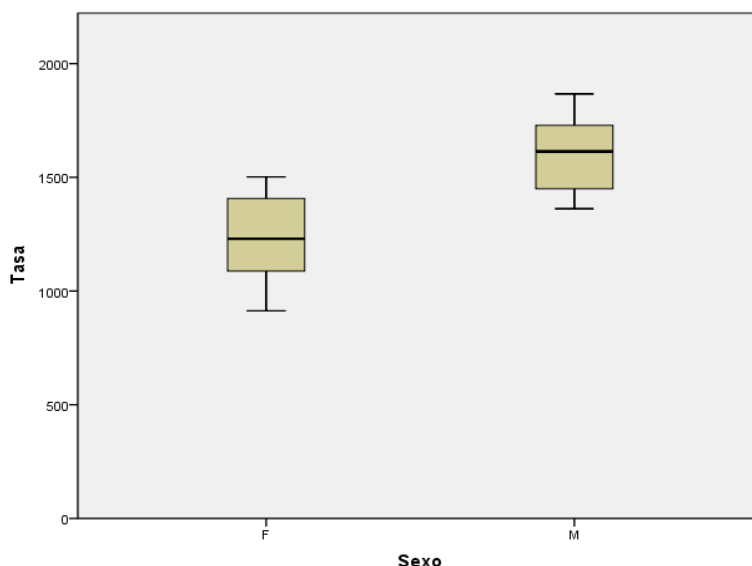
GUÍA PARA REALIZAR EL EJERCICIO 2 DE LA PRÁCTICA:

8. Para las **REPRESENTACIONES GRÁFICAS** (apartados (a y b)) se selecciona

GRÁFICOS → GENERADOR DE GRÁFICOS

y se elige el que interese:

- Para dibujar el **DIAGRAMA DE CAJAS**: se pincha en **DIAGRAMAS DE CAJA**, se arrastra el icono **DIAGRAMA DE CAJAS SIMPLE** a la esquina superior derecha, la variable Tasa al Eje Y y Sexo al Eje X y, finalmente, se pincha **ACEPTAR**:

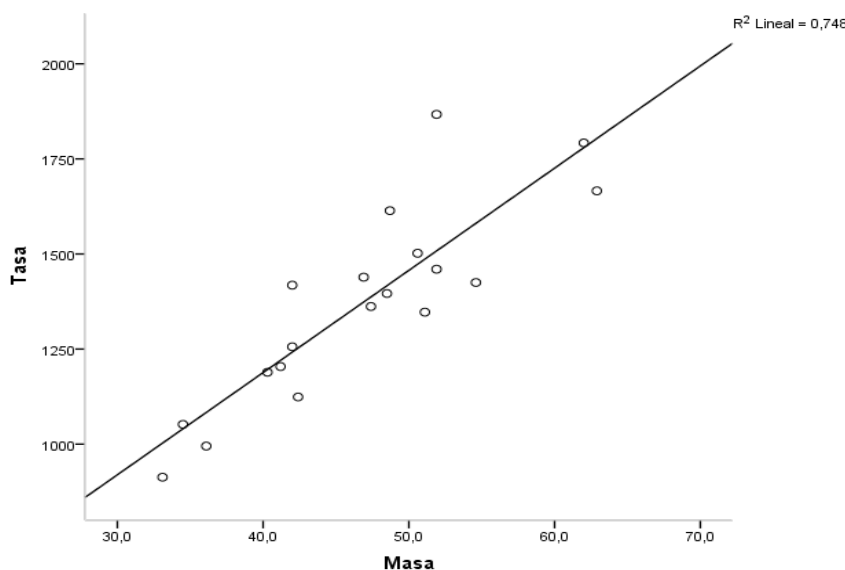
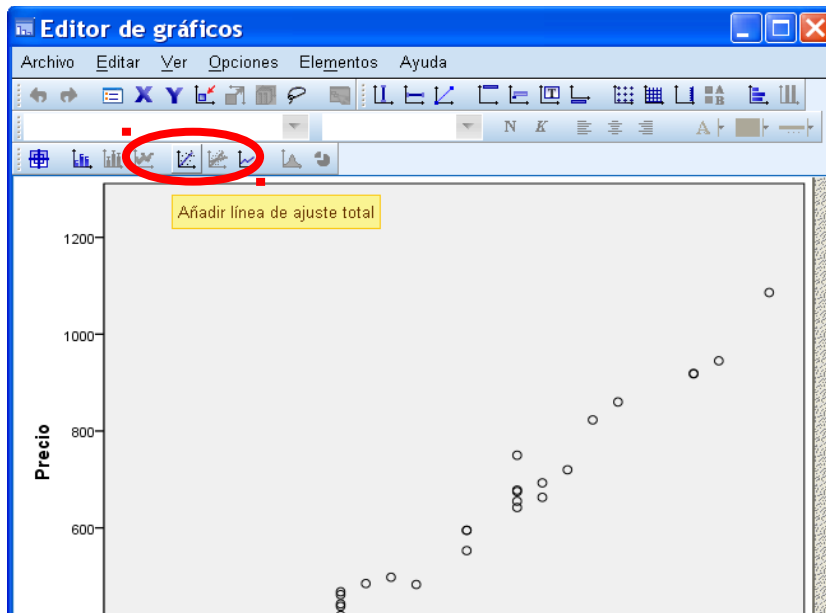


Se observamos que la caja de la derecha (“hombres”), está por encima de la caja de la izquierda (“mujeres”), por lo tanto gran parte de los datos correspondientes a los hombres está por encima de gran parte de los datos correspondientes a las mujeres. También se observa una cierta diferencia en la dispersión de los cuartiles aunque, dado el escaso número de datos, no tiene por qué ser significativa.

- Para dibujar un **DIAGRAMA DE DISPERSIÓN** de las variables Masa y Tasa (sin tener en cuenta ya el SEXO), se pincha en **DIAGRAMA DE DISPERSIÓN**, se arrastra el icono de **DISPERSIÓN SIMPLE** y luego se colocan las variables Masa y Tasa de la izquierda en el eje de las x y de las y, respectivamente. Se hace clic en **ACEPTAR**:
Haciendo doble clic en el gráfico resultante puede cambiarse el aspecto del gráfico.

- Para **DIBUJAR LA RECTA DE REGRESIÓN** se hacer doble click en el diagrama de dispersión.

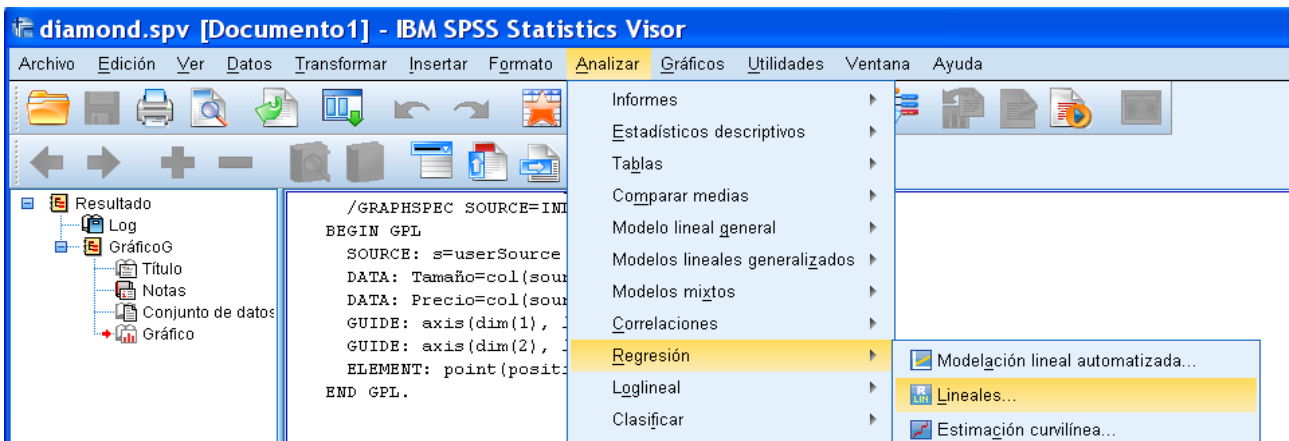
➤ Aparece el dibujo en el editor de gráficos. Se pincha en el botón “**AÑADIR LÍNEA DE AJUSTE TOTAL**”. Al cerrar el editor de gráficos aparecerá dibujada la recta de regresión.



Este gráfico sirve para visualizar una posible relación lineal entre las dos variables. Puede observarse que la mayoría de los datos se ajustan a la línea recta.

➤ Para hallar la expresión explícita de la recta pinchar en

ANALIZAR → REGRESIÓN → LINEALES



En el cuadro de diálogo resultante se elige la variable dependiente (Tasa) y la independiente (Masa) y se pincha en ACEPTAR.

Aparecen varias tablas de datos, de todas ellas la necesaria para escribir la ecuación de la recta es la tabla COEFICIENTES. La columna de B ofrece los parámetros estimados de la recta de regresión

Coeficientes^a

Modelo	Coeficientes no estandarizados		Coeficientes tipificados		t	Sig.
	B	Error típ.	Beta			
1 (Constante)	113,165	179,587			,630	,537
Masa	26,879	3,786	,865		7,099	,000

a. Variable dependiente: Tasa

$$Y = 113,165 + 26,879X$$

9. Para estudiar la **CORRELACIÓN** entre ambas variables se pincha en

ANALIZAR → CORRELACIONES → BIVARIADAS

y se obtiene una matriz donde aparece el valor del coeficiente de correlación (de Pearson).

correlación

		Masa	Tasa
Masa	Correlación de Pearson	1	,865**
	Sig. (bilateral)		,000
	N	19	19
Tasa	Correlación de Pearson	,865**	1
	Sig. (bilateral)	,000	
	N	19	19

** . La correlación es significativa al nivel 0,01 (bilateral).

Al ser la correlación 0,865, puede afirmarse que existe una fuerte correlación lineal positiva entre las variables.