

ESTADÍSTICA – 1º AMBIENTALES

PRIMERA PRÁCTICA: ESTADÍSTICA DESCRIPTIVA (GUIADA & RESUELTA)

El objetivo de esta práctica es hacer un análisis descriptivo con el programa SPSS. Para ello, se os plantean dos ejercicios: Comenzaremos con el problema de calcular estadísticos básicos y representaciones gráficas para una única variable, posteriormente hallaremos los estadísticos correspondientes a dos variables y estudiaremos la relación entre ambas.

Un paso previo a la resolución de los ejercicios es introducir los datos al programa SPSS: o bien los introducís vosotros manualmente, o bien se os facilitan en un fichero, este puede venir dado ya en formato SPSS (extensión **.sav**), o en otros como texto (**.txt**) o Excel (**.xlsx**)

PRIMERA PARTE: ESTADÍSTICA DESCRIPTIVA PARA UNA VARIABLE

Descripción de los datos: El fichero **Cobre.txt** muestra la concentración del nutriente cobre (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.
- Comentar las conclusiones que se obtienen: diferencias entre media y mediana y a qué creéis que se deben, conclusiones del histograma y del box-plot...

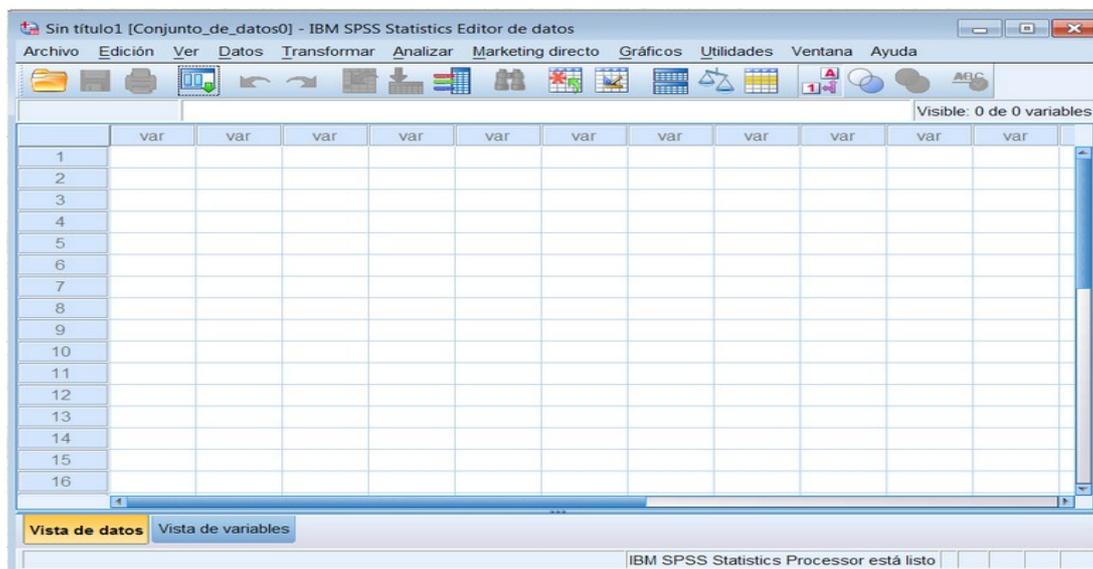
INTRODUCIR DATOS EN EL PROGRAMA SPSS

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

Se trata del fichero de texto **Cobre.txt** donde aparecen las concentraciones de cobre determinadas por los distintos laboratorios. Para ello posicionar el ratón sobre **Cobre.txt** en moodle, pinchar con el botón derecho y marcar *Guardar enlace...*

2. ABRIMOS EL PROGRAMA SPSS:

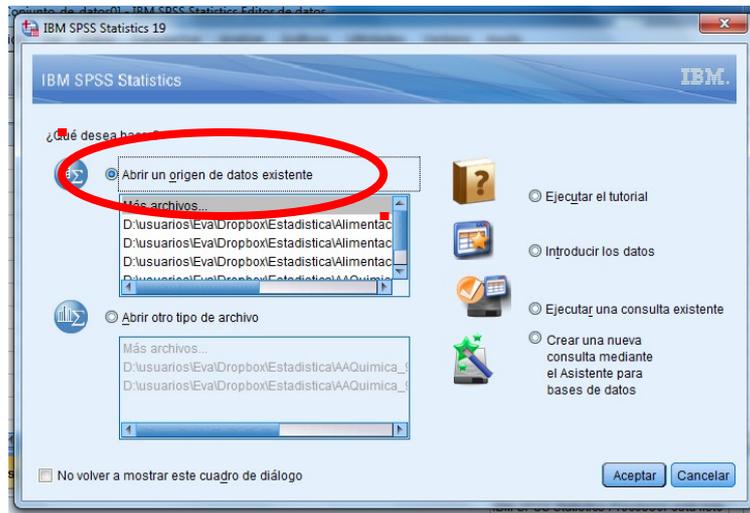
INICIO → PROGRAMAS → IBM SPSS STATISTICS → IBM SPSS STATISTICS 21.



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:

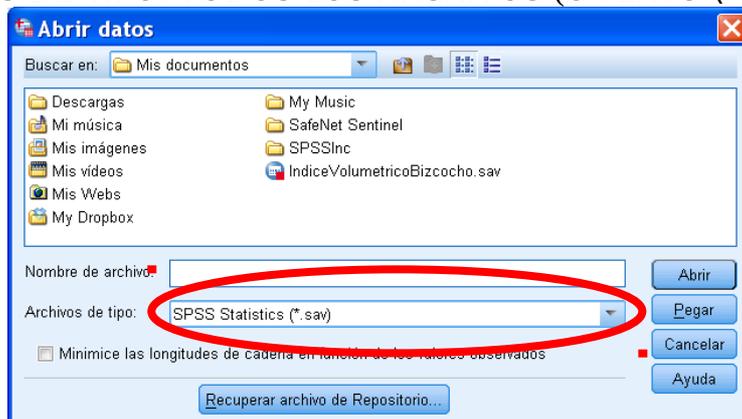
ARCHIVO → ABRIR → DATOS

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

ARCHIVOS DE TIPO: SPSS Statistics (*.sav)

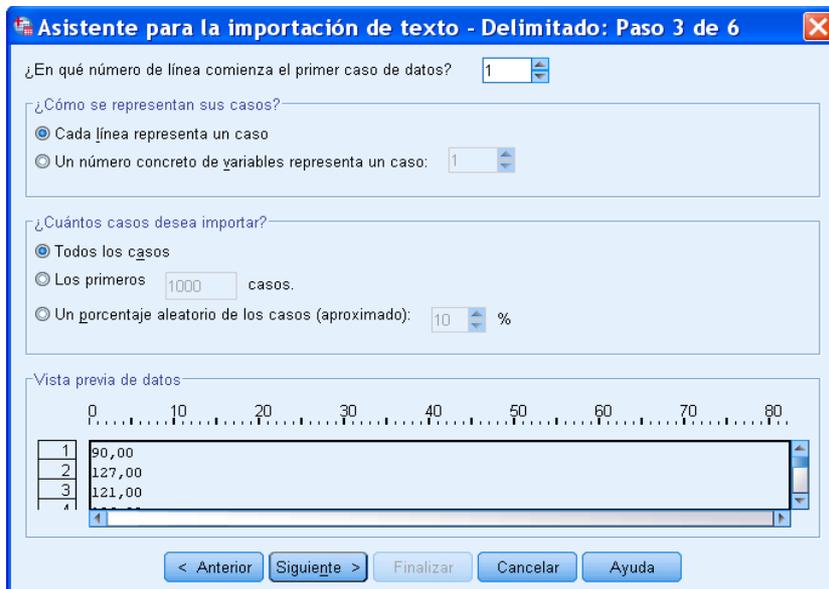
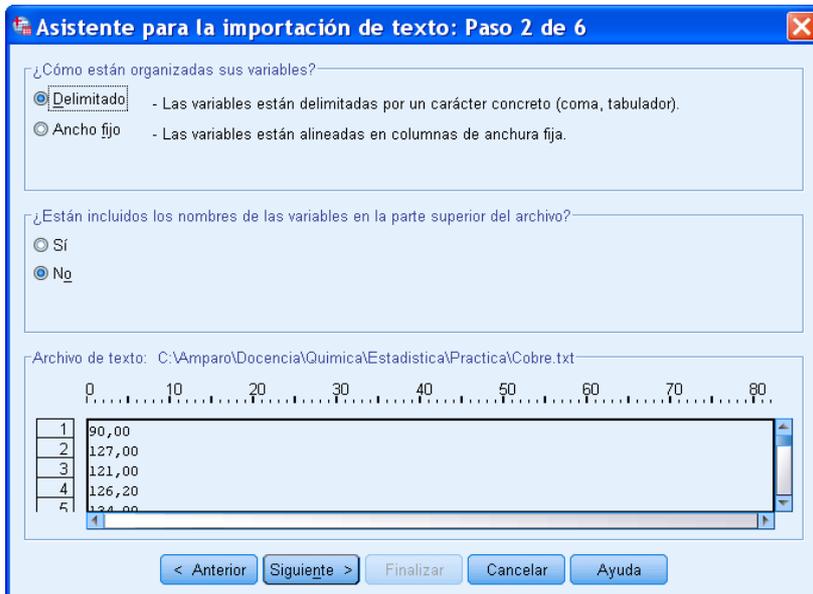
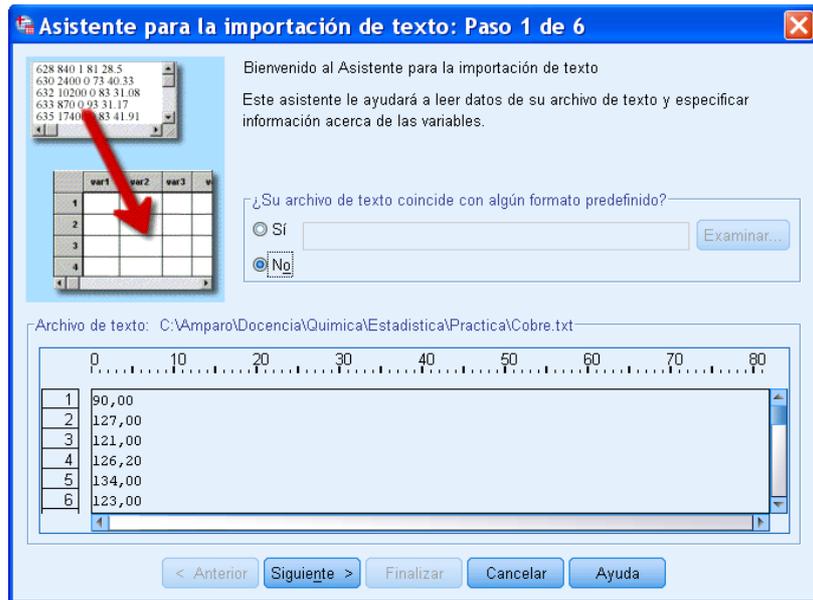
Debéis cambiarlo, ya que vuestro fichero no es .sav. Para ello, pinchar la pestaña y seleccionar

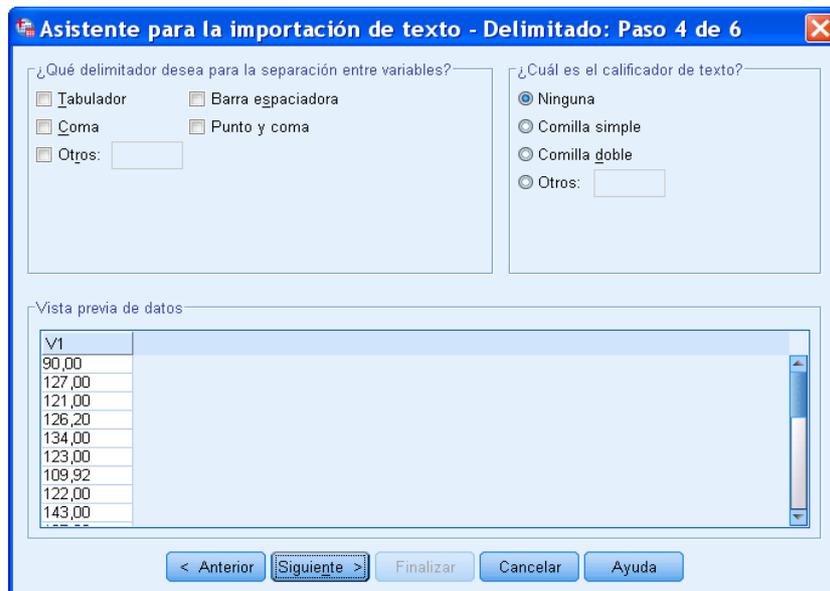
ARCHIVOS DE TIPO: TODOS LOS ARCHIVOS (O TEXTO (*.txt,*.dat))



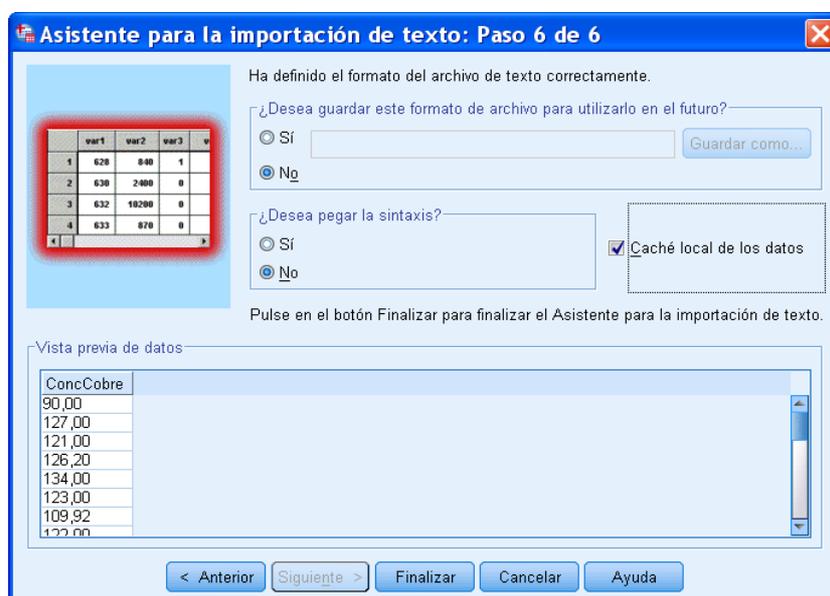
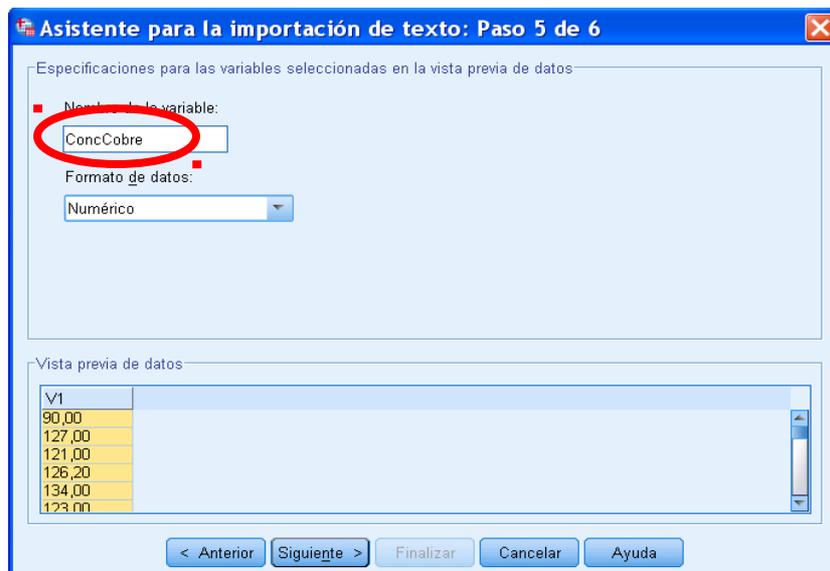
finalmente abrir el fichero **Cobre.txt**.

A continuación aparecen unas ventanas donde se os preguntan por algunas características del fichero de texto: leedlas con detenimiento (para entenderlas) y seguid los pasos que se os muestran:



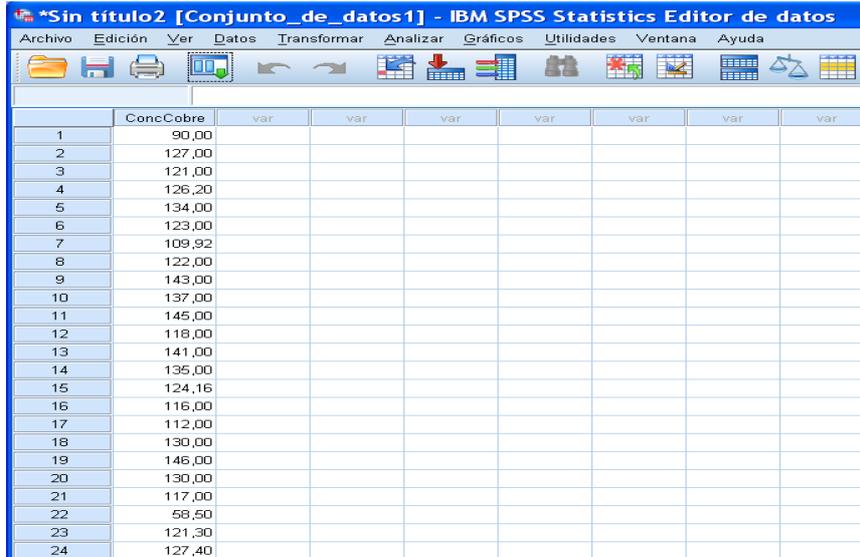


Por defecto SPSS nombra a la variable como **V1**. Debéis cambiarlo a un nombre más apropiado. El nombre de las variables debe ser *sencillo* (sin espacios, sin acentos...), p.e. **ConcCobre**.



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:



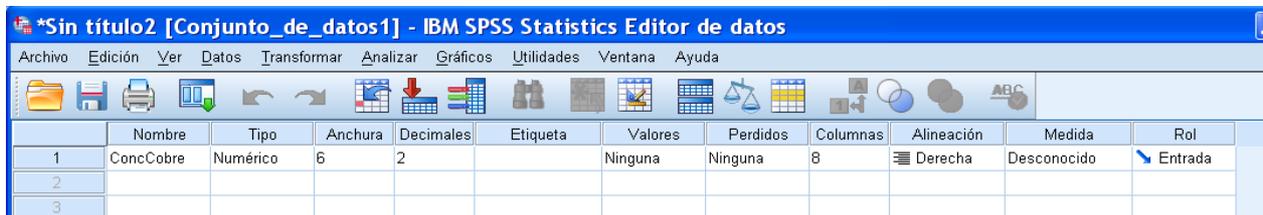
	ConcCobre	var						
1	90,00							
2	127,00							
3	121,00							
4	126,20							
5	134,00							
6	123,00							
7	109,92							
8	122,00							
9	143,00							
10	137,00							
11	145,00							
12	118,00							
13	141,00							
14	135,00							
15	124,16							
16	116,00							
17	112,00							
18	130,00							
19	146,00							
20	130,00							
21	117,00							
22	58,50							
23	121,30							
24	127,40							

en el que es conveniente seleccionar ARCHIVO → GUARDAR COMO ... y guardarlo con extensión .sav (recuerda, ese es el formato del SPSS) por si tuvierais que volver a usarlo.

En la parte inferior del editor aparecen dos pestañas: VISTA DE DATOS y VISTA DE VARIABLES.

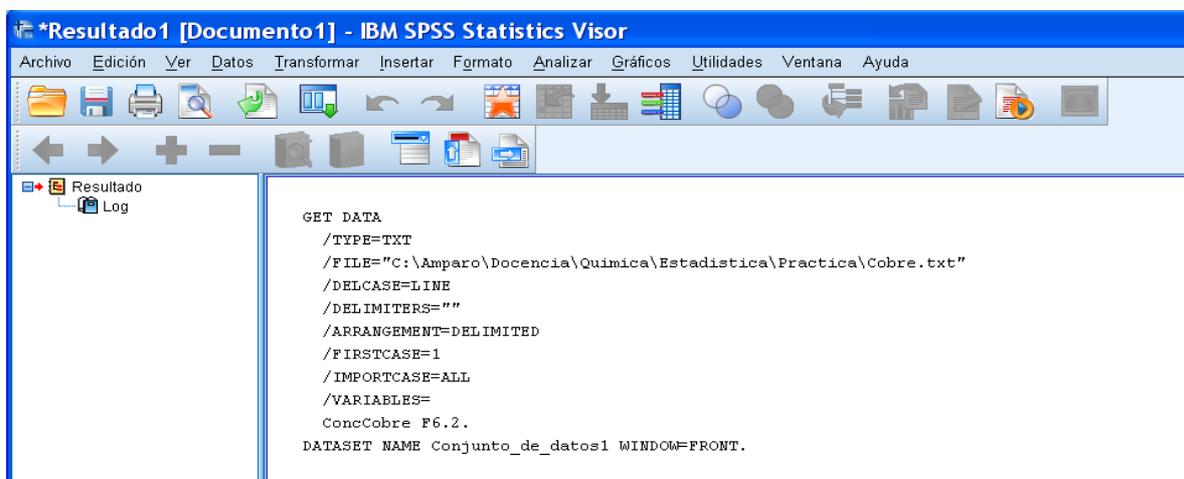
La pestaña VISTA DE DATOS corresponde a la imagen anterior y nos muestra los nombres y valores de las variables.

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	Númérico	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.
```

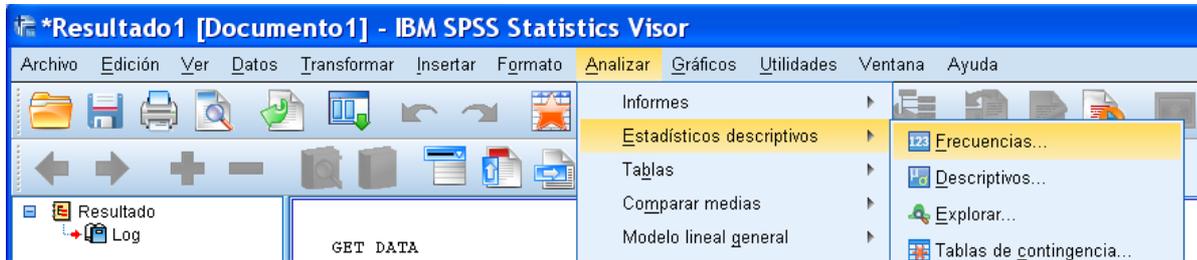
en el que irán apareciendo los resultados de nuestro análisis estadístico (tablas, gráficos,...).

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)*) seleccionar en el Editor de datos o en el Visor de resultados (da lo mismo) pinchar en:

ANALIZAR → ESTADÍSTICOS DESCRIPTIVOS → FRECUENCIAS.



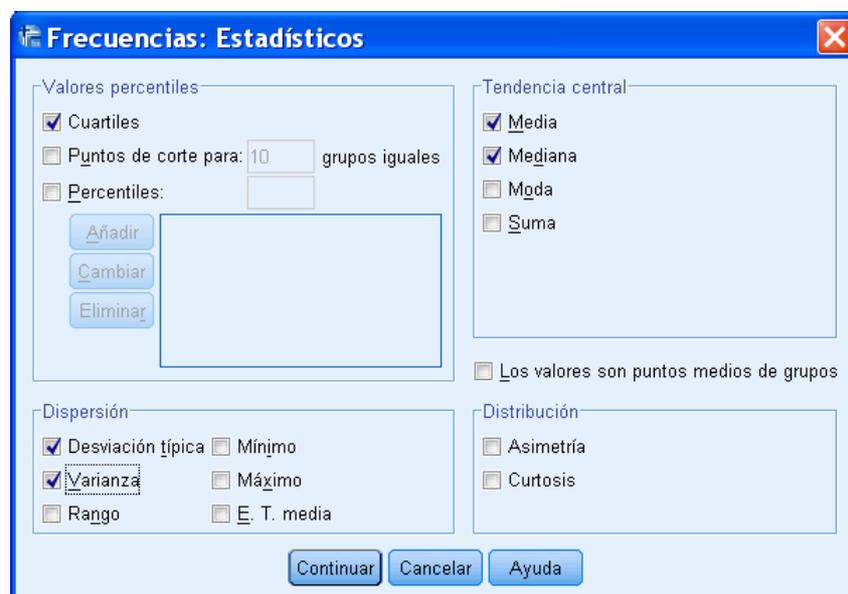
En el cuadro de diálogo



pasar ConcCobre a la columna VARIABLES pinchando en la flecha.

Indicación: Deseleccionar MOSTRAR TABLAS DE FRECUENCIAS

En el botón GRÁFICOS seleccionamos NINGUNO (seguramente seleccionado por defecto) y en ESTADÍSTICOS marcamos los que nos interesen: media, mediana, varianza, des. típica y cuartiles.



Hacer 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

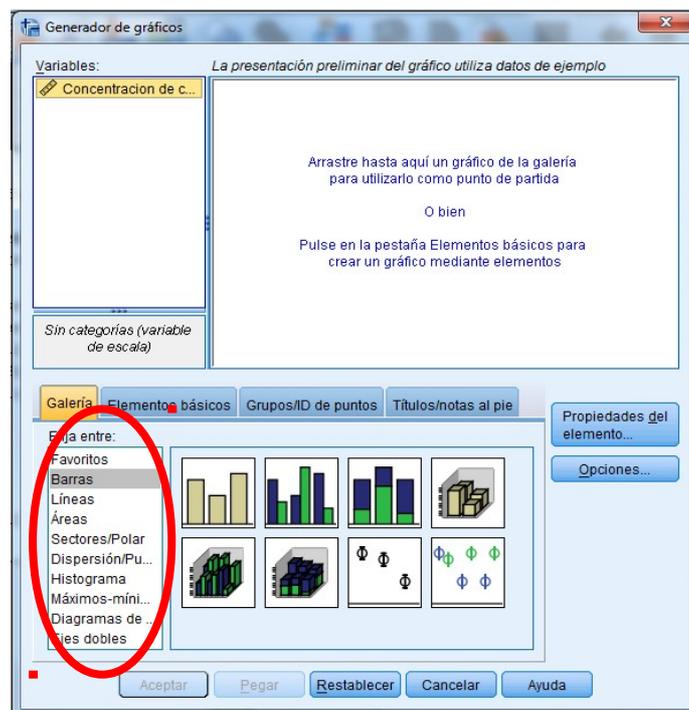
Obs: Percentil 25 = Primer Cuartil, Percentil 50 = Segundo Cuartil y Percentil 75 = Tercer Cuartil

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

GRÁFICOS → GENERADOR DE GRÁFICOS

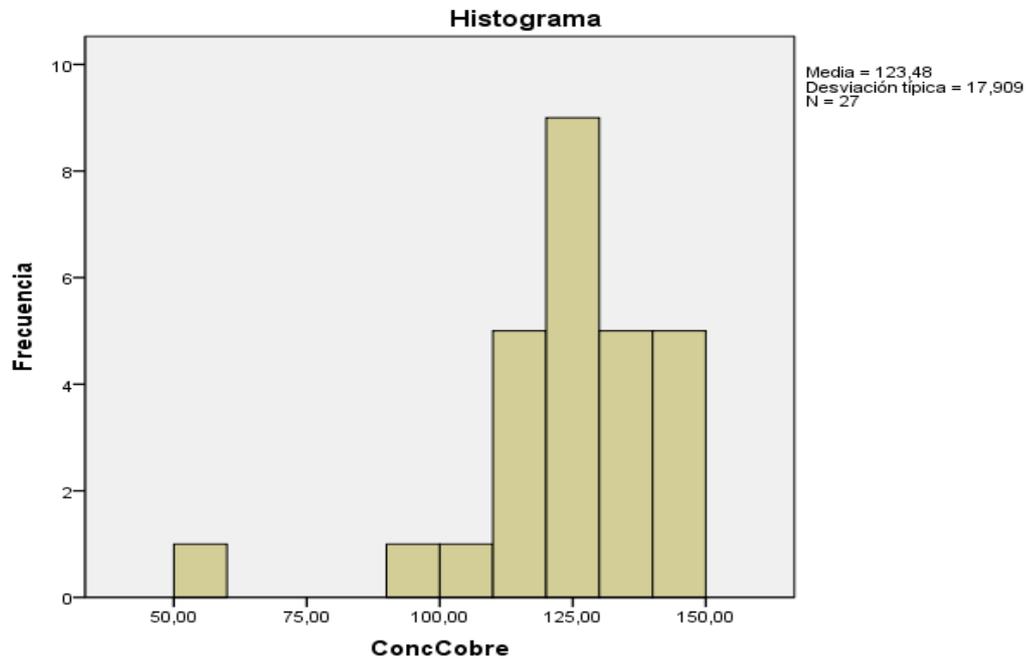


Aparecerá la siguiente ventana:



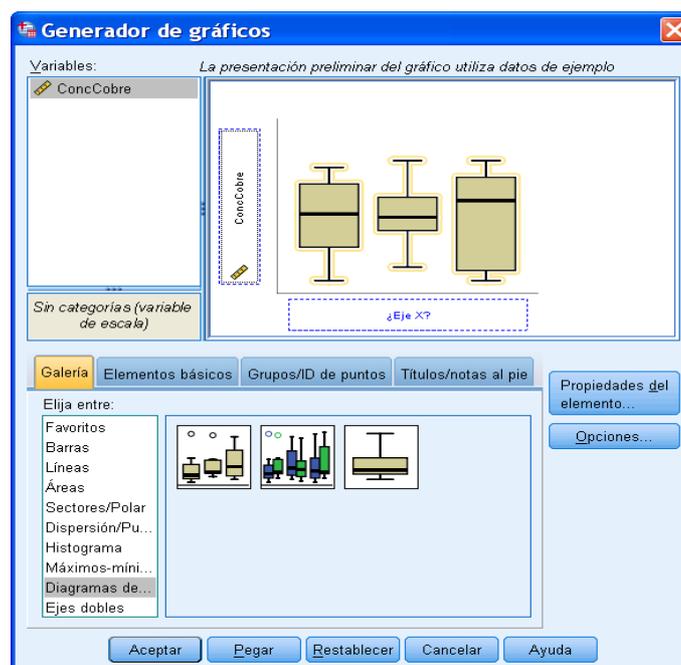
y elegimos el que nos interese:

- ❖ **HISTOGRAMA:** Pinchar en HISTOGRAMA y arrastrar 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). Pinchar en ACEPTAR y entonces aparece el histograma dibujado en el Visor:

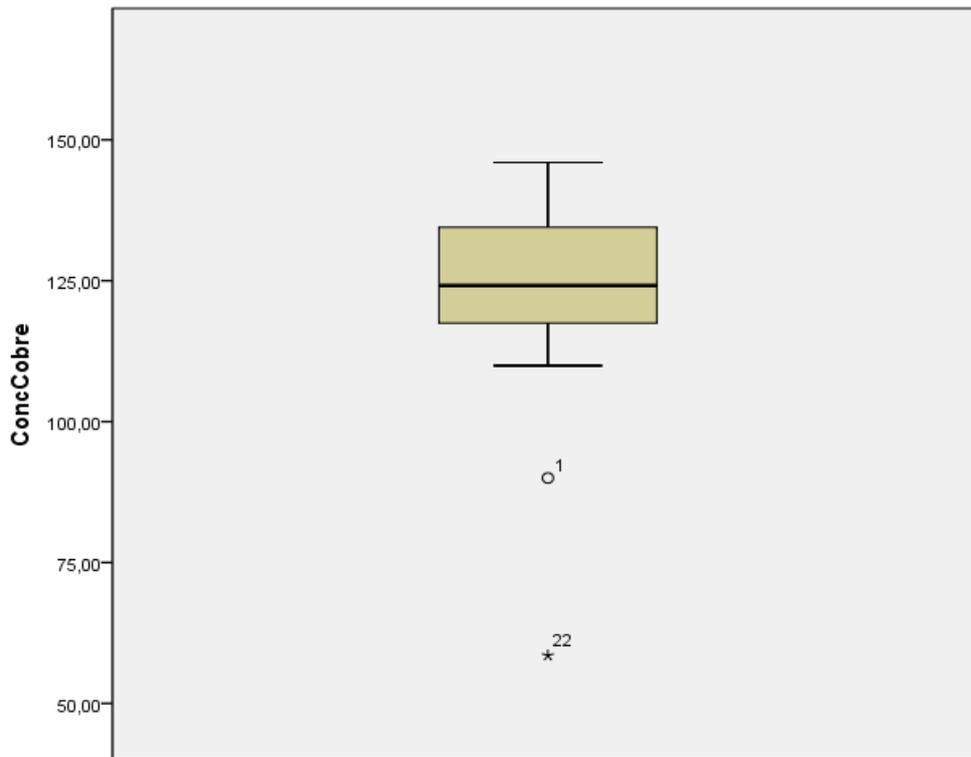


Haciendo doble clic en el dibujo, sale una pantalla con el mismo y entonces podemos modificar el gráfico. Hacemos doble clic en aquello que queremos modificar (color, tamaño de fuente, etc). Por ejemplo, haciendo doble clic sobre el dibujo del histograma aparece la siguiente ventana en la que podemos seleccionar distinto número de intervalos para el histograma:

- ❖ **DIAGRAMA DE CAJAS:** Pinchar en DIAGRAMAS DE CAJA, arrastrar 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 nuestro análisis y debe salir:



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

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

En el box-plot podemos comprobar que Q2 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, vemos que las observaciones $x(1)$ y $x(22)$ son datos que quedan fuera del box-plot, es decir, son datos atípicos.

SEGUNDA PARTE: ESTADÍSTICA DESCRIPTIVA PARA DOS VARIABLES

Descripción de los datos: Corresponde al fichero SPSS: metabolismo.sav. Las variables son: la Tasa metabólica y la Masa corporal magra; de una muestra de 7 hombres y 12 mujeres.

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

EJERCICIO 2:

- Comparar el diagrama de cajas la Tasa de hombres y la Tasa de mujeres ¿cuál de las dos poblaciones tiene la media superior?
- Obtener el diagrama de dispersión de las variables Masa y Tasa (sin distinguir entre sexos), ¿Para qué sirve este gráfico de los datos? ¿Qué observas en él?
- ¿Cuál es la recta de regresión estimada $Y=a+bX$?
- ¿Te parece bueno el ajuste de la recta de regresión a los datos? ¿Por qué?

INTRODUCIR DATOS EN EL PROGRAMA SPSS

- Descargar los datos de Moodle al disco duro del ordenador (p.e., al escritorio).
- Para empezar a realizar el ejercicio podéis pinchar sobre el propio fichero metabolismo.sav. Si no os dejar abrirlo, haced como se os indicó en la Primera Parte: posicionar el ratón sobre metabolismo.sav en moodle, pinchar con el botón derecho y marcar *Guardar enlace...*

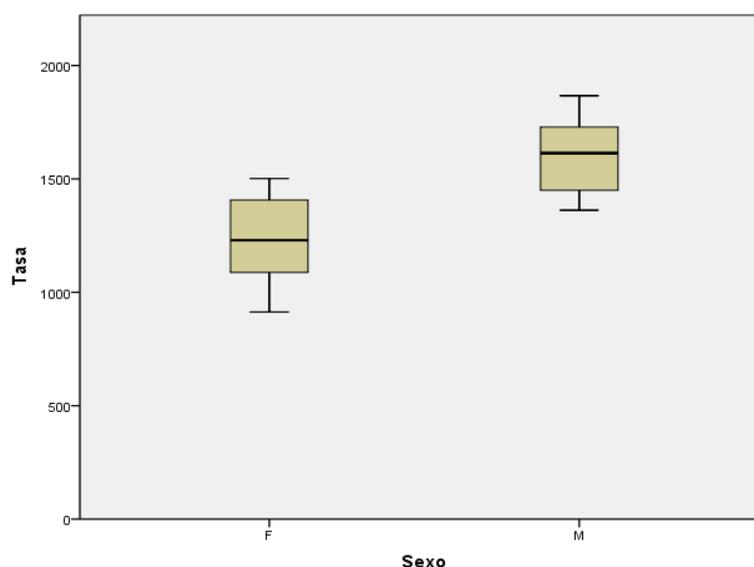
GUÍA PARA REALIZAR EL EJERCICIO 2 DE LA PRÁCTICA:

- Para las **REPRESENTACIONES GRÁFICAS** (apartados (a) y b)) pinchar en

GRÁFICOS → GENERADOR DE GRÁFICOS

y elegimos el que nos interese:

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



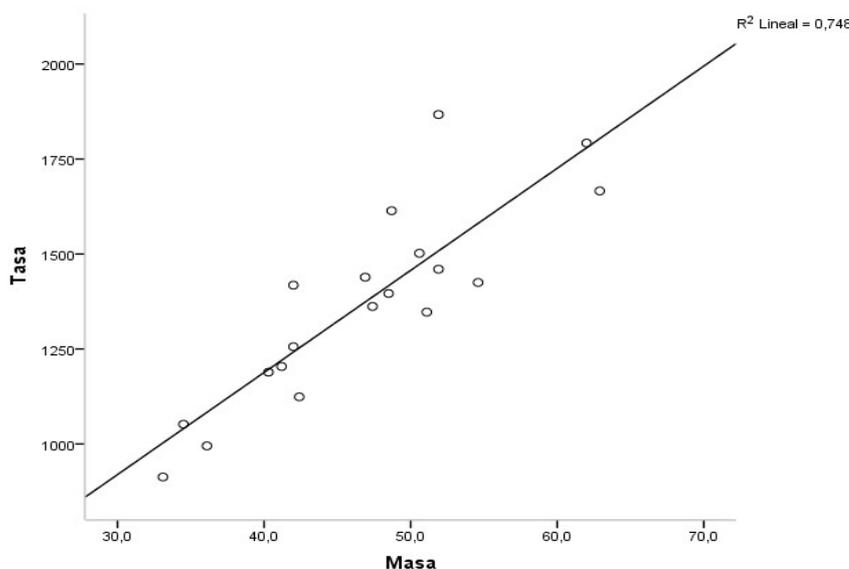
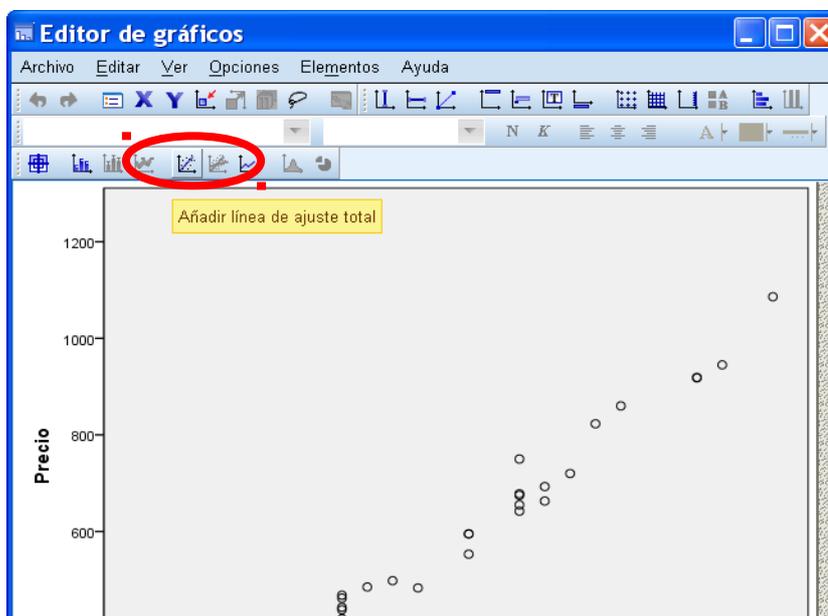
Observamos que la caja de la derecha (la de los hombres), está por encima de la caja de la izquierda (la de las mujeres), por lo tanto la media es superior en los hombres que en las mujeres, ya que se concentra en un intervalo mayor. También se refleja en los distintos cuartiles, pues las mujeres se concentran sobre todo entre el Q1 y Q2 y los hombres entre Q2 y Q3.

- Para dibujar un **DIAGRAMA DE DISPERSIÓN** de las variables Masa y Tasa (sin tener en cuenta ya el Sexo), pinchar en **DIAGRAMA DE DISPERSIÓN**, arrastrar el icono de **DISPERSIÓN SIMPLE** y luego colocar las variables Masa y Tasa de la izquierda en el eje de las x y de las y, respectivamente. Hacer clic en **ACEPTAR**:

Haciendo doble clic en el gráfico resultante podemos cambiar el aspecto del gráfico.

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

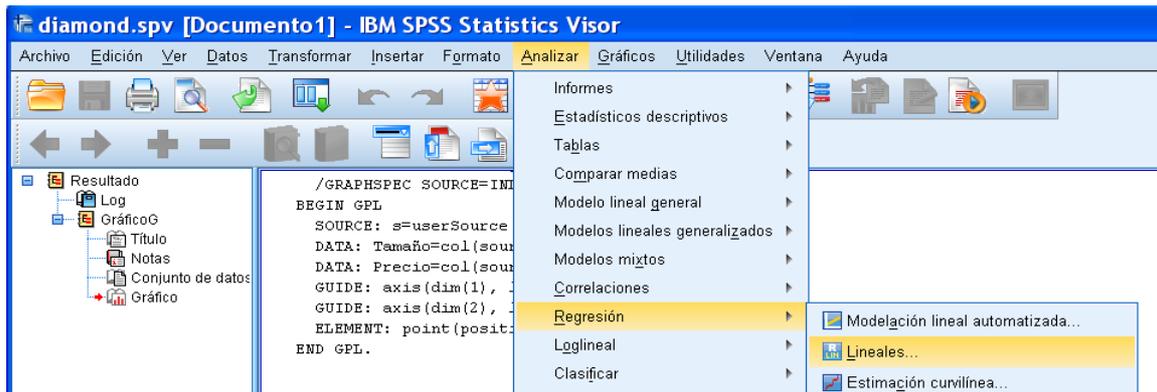
➤ Aparece el dibujo en el editor de gráficos. Pinchar en el botón “**AÑADIR LÍNEA DE AJUSTE TOTAL**”. Cerrar el editor de gráficos y aparecerá dibujada la recta de regresión.



Este gráfico sirve para estimar una cierta relación lineal entre las dos variables visualmente. Podemos comprobar que la mayoría de los datos se ciñen a la línea recta.

➤ Para hallar la expresión explícita de la recta pinchar en

ANALIZAR → REGRESIÓN → LINEALES



En el cuadro de diálogo resultante elegir la variable dependiente (Tasa) y la independiente (Masa) y pinchar en ACEPTAR.

Aparecerán varias tablas de datos, de todas ellas la necesaria para 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. Veamos la **CORRELACIÓN** entre ambas variables. Para ello pinchar en ANALIZAR → CORRELACIONES → BIVARIADAS y se obtiene una matriz donde aparece el valor del coeficiente de correlación de Pearson dado en clase.

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, afirmamos que está cerca de 1 y, por lo tanto, existe una fuerte correlación lineal positiva: podemos decir que la recta de regresión se adecua muy bien a la realidad.