

ESTADÍSTICA

Práctica 5

Grupo 6 - Práctica con SPSS.

Alumno: Elvi Mihai Sabau Sabau.




DNI: 51254875L

5.4 Ejercicios 1.

Se desea comparar la calidad del servicio de fibra de tres proveedores distintos. Para ello se dispone de tres ordenadores iguales con la misma configuración sobre los que se ha realizado pruebas en distintos momentos para medir la velocidad de conexión que proporcionan los tres proveedores analizados en Mbps. Los resultados obtenidos son los siguientes:

PROVEEDOR A	PROVEEDOR B	PROVEEDOR C
18,50	15,50	16,70
18,90	14,80	21,30
21,10	15,70	19,20
17,90	17,40	19,70
16,80	16,30	18,55
17,70	17,20	19,50
21,05	16,05	21,10
18,50	19,10	15,35
19,75	15,60	11,30
20,10	15,80	18,90

Primero, creamos e introducimos los datos a las variables:

	 ProveedorA	 ProveedorB	 ProveedorC
1	16,80	14,80	11,30
2	17,70	15,50	15,35
3	17,90	15,60	16,70
4	18,50	15,70	18,55
5	18,50	15,80	18,90
6	18,90	16,05	19,20
7	19,75	16,30	19,50
8	20,10	17,20	19,70
9	21,05	17,40	21,10
10	21,10	19,10	21,30

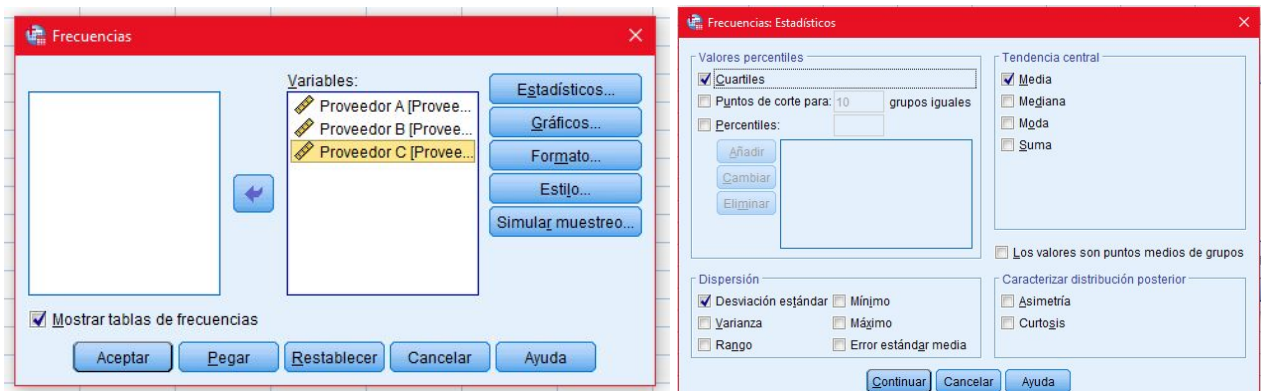
Y ordenamos de menor a mayor haciendo click derecho en cada columna, y ordenando de manera ascendente.

a) *Obtener los gráficos caja para los resultados de los tres proveedores.*

Para esto, lo primero que deberemos hacer es sacar los datos necesarios para calcular los las vallas, para esto necesitaremos los cuartiles.

Para ellos vamos a Analizar > Estadísticos Descriptivos > Frecuencias > y añadimos las 3 variables, después vamos a Estadísticos y seleccionamos Cuartiles, Media y Desviación Estándar.

(estos últimos dos datos los necesitaremos para completar los siguientes apartados (b y c)).



Estadísticos

		Proveedor A	Proveedor B	Proveedor C
N	Válido	10	10	10
	Perdidos	0	0	0
Media		19,0300	16,3450	18,1600
Desv. Desviación		1,43859	1,24241	3,00996
Percentiles	25	17,8500	15,5750	16,3625
	50	18,7000	15,9250	19,0500
	75	20,3375	17,2500	20,0500

Estos datos nos servirán para calcular las vallas, así cuando hagamos el gráfico de la caja para representar los resultados, los podremos plasmar de manera más específica.

Ahora, para calcular las vallas, usaremos las siguientes fórmulas:

Siendo n el tamaño de la muestra, se pueden calcular las posiciones de:

$$- Q1 = (n + 1) / 4$$

$$- M \text{ (Mediana)} = (n + 1) / 2$$

$$- Q3 = 3 (n + 1) / 4$$

$$\text{IQR (intervalo intercuartílico)} = Q3 - Q1$$

$$\text{Va ext sup (valla exterior superior)} = Q3 + (3 * \text{IQR})$$

$$\text{Va int sup (valla interior superior)} = Q3 + (1,5 * \text{IQR})$$

$$\text{Va int inf (valla interior inferior)} = Q1 - (1,5 * \text{IQR})$$

$$\text{Va ext inf (valla exterior inferior)} = Q1 - (3 * \text{IQR})$$

$\text{Va Ady sup (valor adyacente superior)} = \max x \leq \text{Vint sup}$. Máximo dato entre $Q3$ y dicha valla. El límite es el valor de la valla y puede coincidir con él.

$\text{Va Ady inf (valor adyacente inferior)} = \min x \geq \text{Vint inf}$.

Mínimo dato entre $Q1$ y dicha valla.

El límite es el valor de la valla y puede coincidir con él.

Valores atípicos:

Entre las dos vallas superiores ($> \text{Valla int sup}$ y $\leq \text{Valla ext sup}$)

ó entre las dos vallas inferiores ($\geq \text{Valla ext inf}$ y $< \text{valla int inf}$).

Valores extremos:

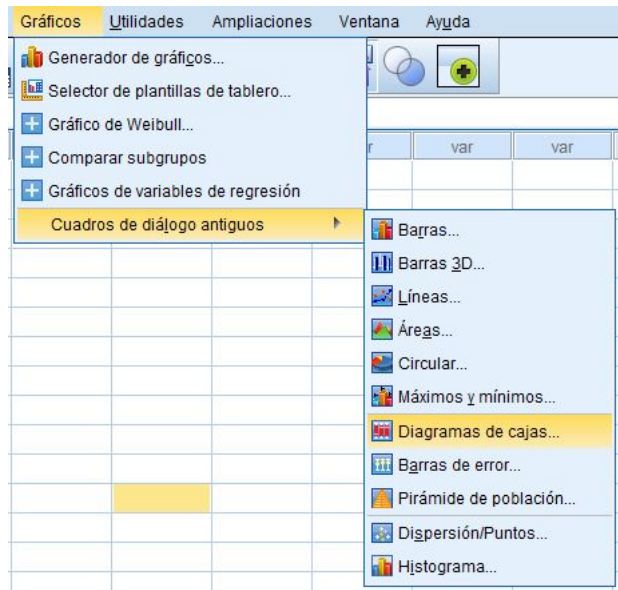
Más alejados de la valla exterior superior ($> \text{Valla ext sup}$)

o de la valla exterior inferior ($< \text{Valla ext inf}$).

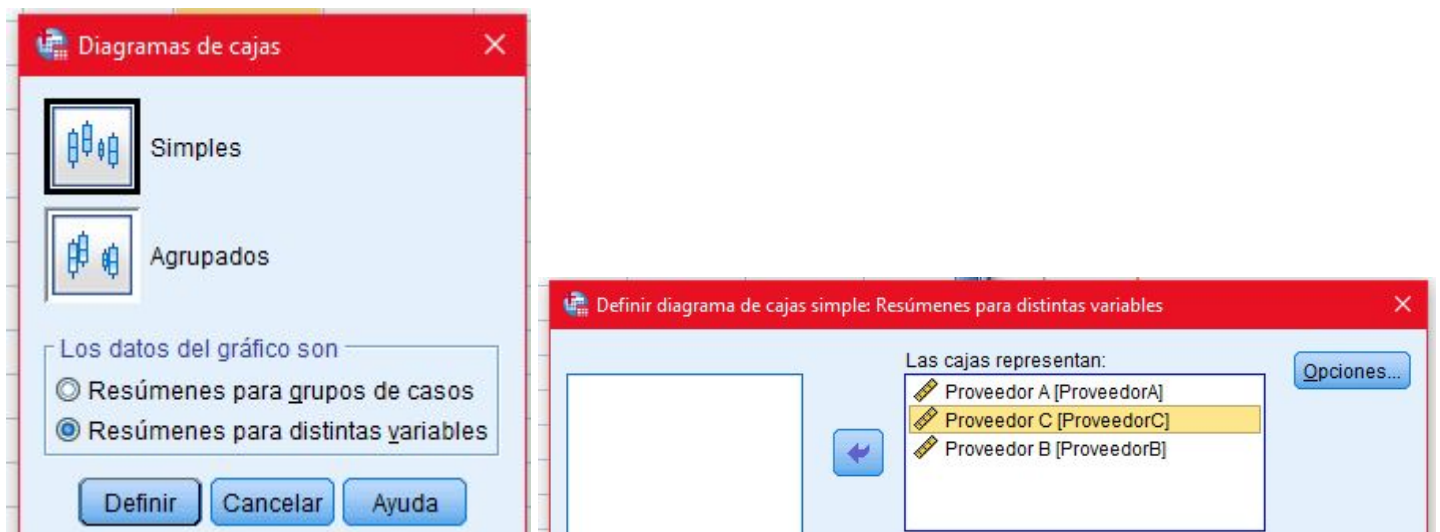
Ahora, realizaremos una tabla donde reposamos nuestros datos calculados hasta tener el gráfico de la caja hecho, yo llamaré a esta tabla “Tabla de vallas”.

	Proveedor A	Proveedor B	Proveedor C
<u>Media</u>	<u>19,03</u>	<u>16,34</u>	<u>18,16</u>
<u>IQR</u>	<u>2,48</u>	<u>1,68</u>	<u>3,69</u>
<u>Valores Atípicos</u>	<u>---</u>	<u>---</u>	<u>---</u>
<u>Valores Extremos</u>	<u>---</u>	<u>---</u>	<u>---</u>
<u>Desviación</u>	<u>1,43</u>	<u>1,24</u>	<u>3,00</u>
Va. Ext. Inf.	10,41	10,53	5,29
Va. Int. Inf.	14,13	13,05	10,82
Va. Ady. Inf.	16,80	14,80	11,30
Cuartil 1	17,85	15,57	16,36
Mediana	18,70	15,92	19,05
Cuartil 3	20,33	17,25	20,05
Va. Ady. Sup.	21,10	19,10	21,30
Va. Int. Sup.	24,05	19,77	25,58
Va. Ext. Sup.	27,77	22,29	31,12

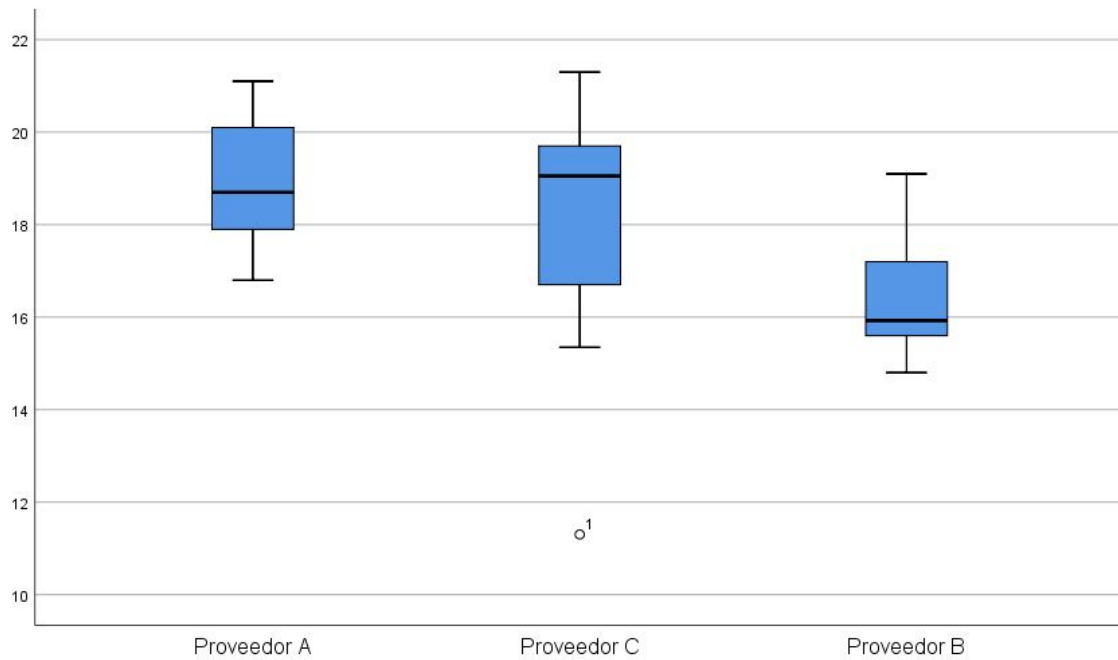
Ahora, para generar el gráfico, debemos ir a Gráficos > Cuadro de diálogos Antiguos > Diagrama de cajas.



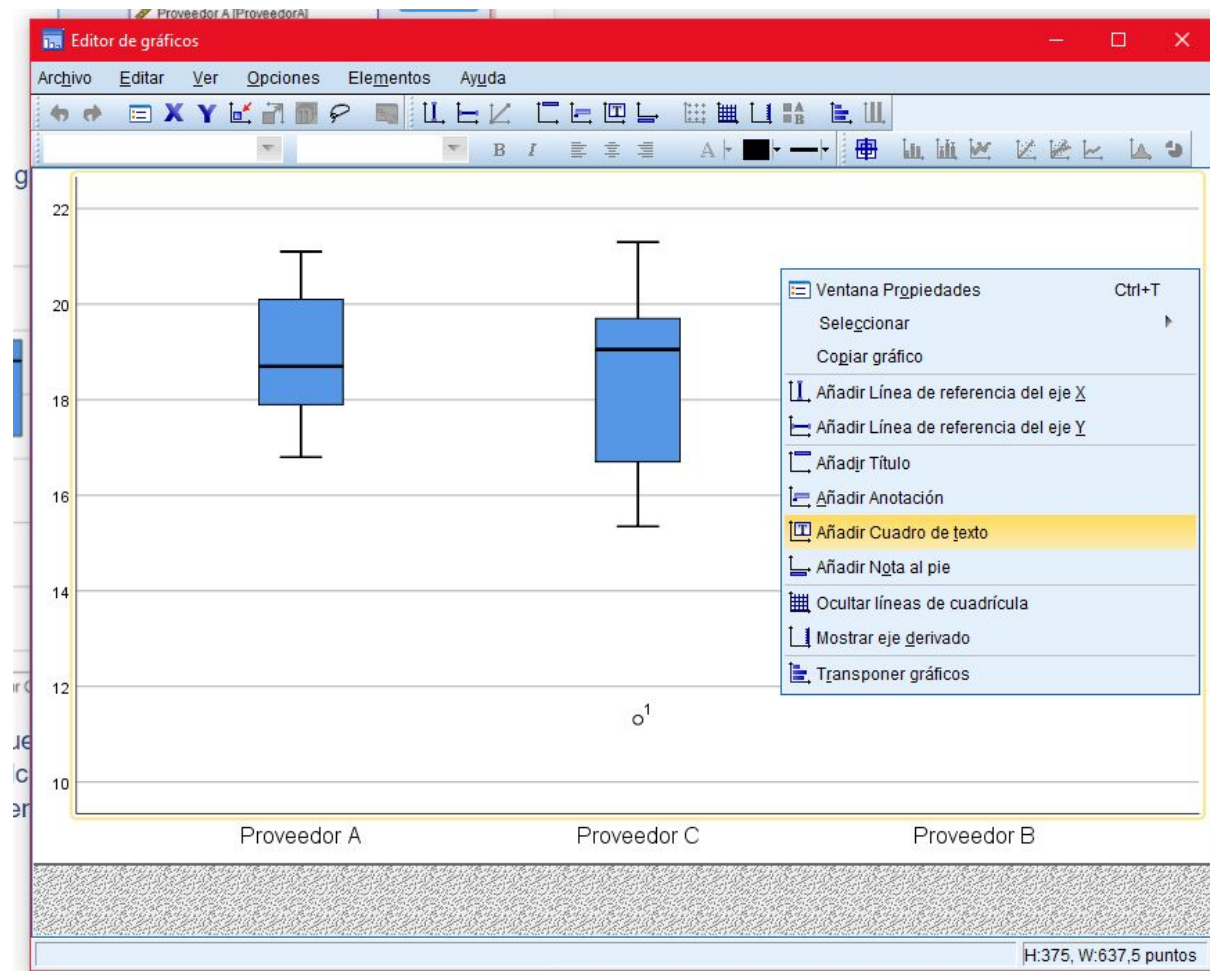
Seleccionamos “Simples” y la opción de “Resúmenes para distintas variables”, le damos a “Definir” y en el nuevo popup añadimos las variables a la lista de variables de cajas a representar.



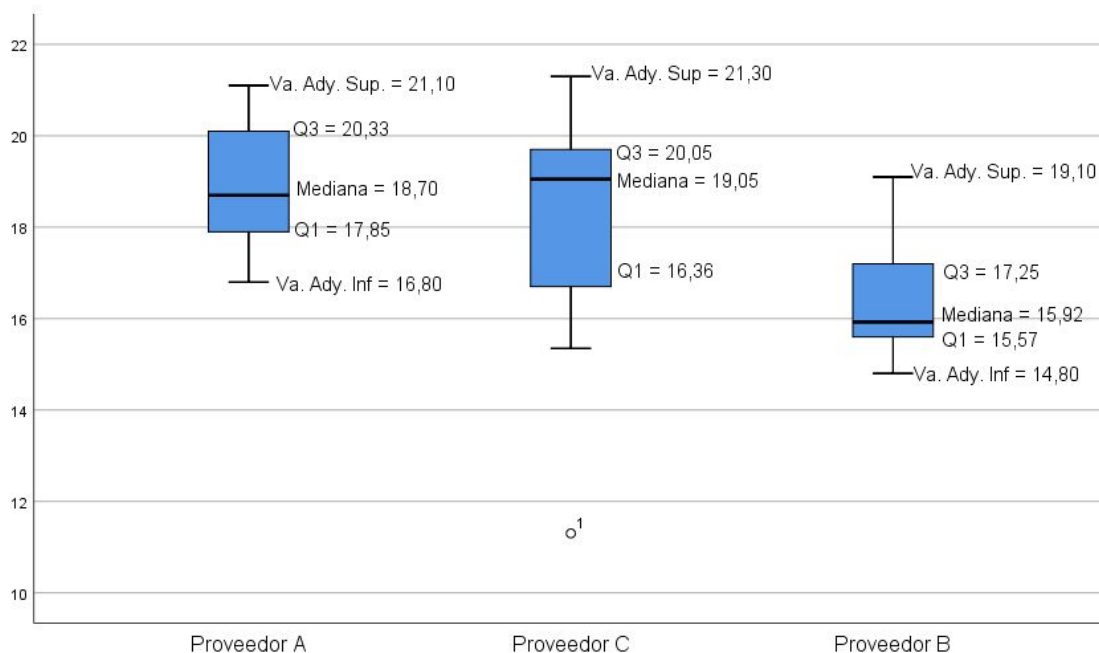
Le damos a Aceptar, y nuestro gráfico de cajas se generará.



Por desgracia, nuestro gráfico no muestra gran cosa en referente a los datos, menos mal que hemos calculado anteriormente los datos necesarios, ahora podremos condimentar nuestro gráfico manualmente, editando el grafo (doble click) y con click derecho > Añadir cuadro de texto.



Añadimos los datos, y volver a ver el gráfico.



* Los datos no tienen porqué coincidir con los del gráfico anterior ya que SPSS usa un algoritmo diferente para calcular los valores.

b) ¿Cuál de ellos tiene mayor dispersión?

Como podemos ver en la tabla del apartado anterior, el proveedor C.

¿Son simétricos?

No, tanto A como C, son sesgados a la derecha, y B es sesgado a la izquierda.

Indica los valores extremos y atípicos si los hubiera.

A primera vista, no parece haber ningún valor extremo o atípico.

c) ¿Qué proveedor es más fiable? ¿Cuál proporciona una mayor velocidad media?

El proveedor B es el más fiable ya que su IQR es el menor, y el proveedor C es el que tendría una mayor velocidad, ya que su mediana es la mayor de todas.

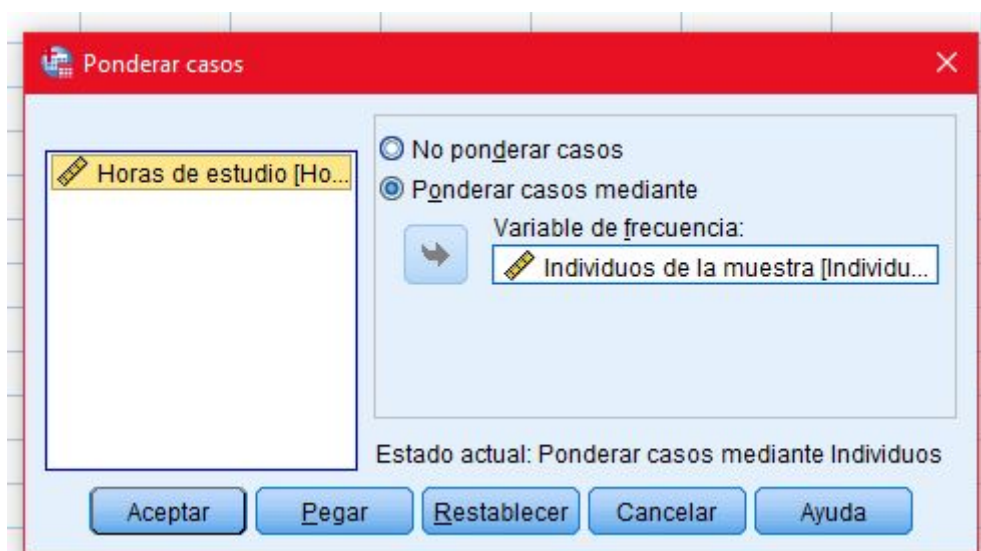
2. Las horas de estudio y trabajo semanal declaradas por parte de una muestra representativa de los alumnos de Ingeniería Informática aparecen en la siguiente tabla:

Horas de estudio/trabajo	Individuos de la muestra
Menos de 5 horas	26
De 5 a 10 horas	32
De 10 a 15 horas	50
De 15 a 20 horas	20
Más de 20 horas	12

Primero, creamos las variables, y introducimos los datos:

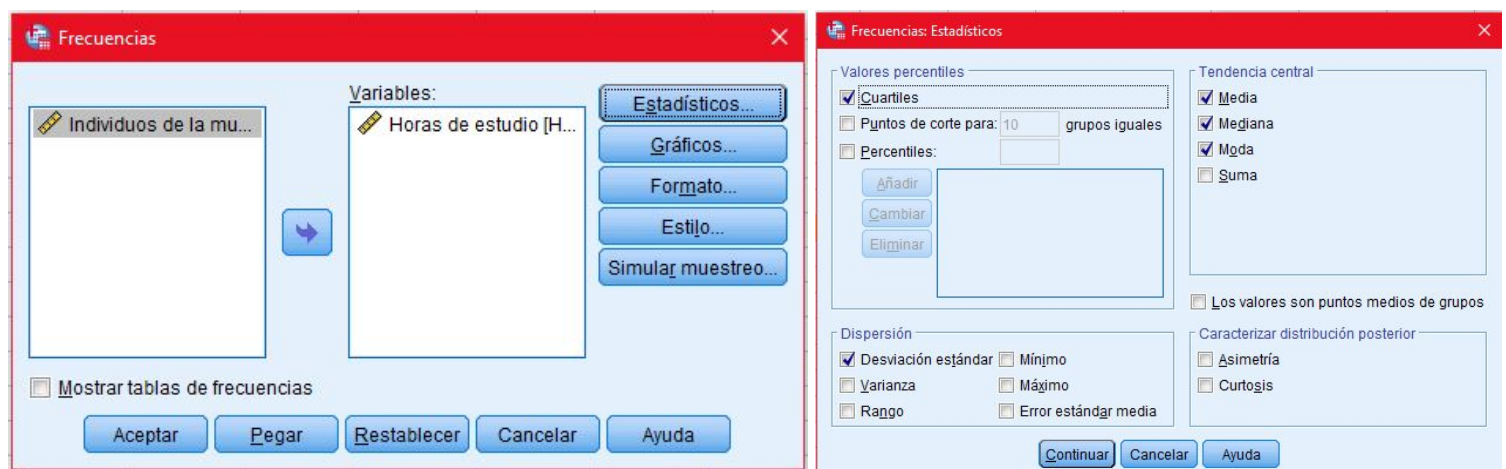
Horas	Individuos
2,5	26
7,5	32
12,5	50
17,5	20
22,5	12

Y ponderamos por la cantidad de individuos, que en nuestro caso sería la frecuencia.



a) Construir el gráfico caja y comentar los resultados (dispersión, valores extremos, simetría...)

Hacemos lo mismo que en el ejercicio anterior, primero sacamos los cuartiles, la media, mediana, desviación, etc.. para esto vamos a Analizar > Estadísticos Descriptivos > Frecuencias > Estadísticos, y seleccionamos las “Horas de Estudio” como variable.



Y Obtenemos los datos:

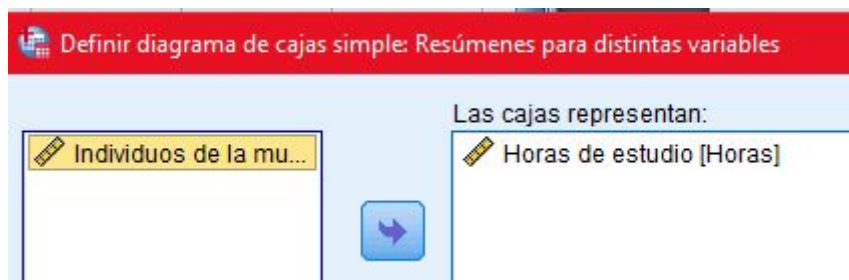
→ Frecuencias

Estadísticos		
Horas de estudio		
N	Válido	140
	Perdidos	0
Media		11,071
Mediana		12,500
Moda		12,5
Desv. Desviación		5,8852
Percentiles	25	7,500
	50	12,500
	75	12,500

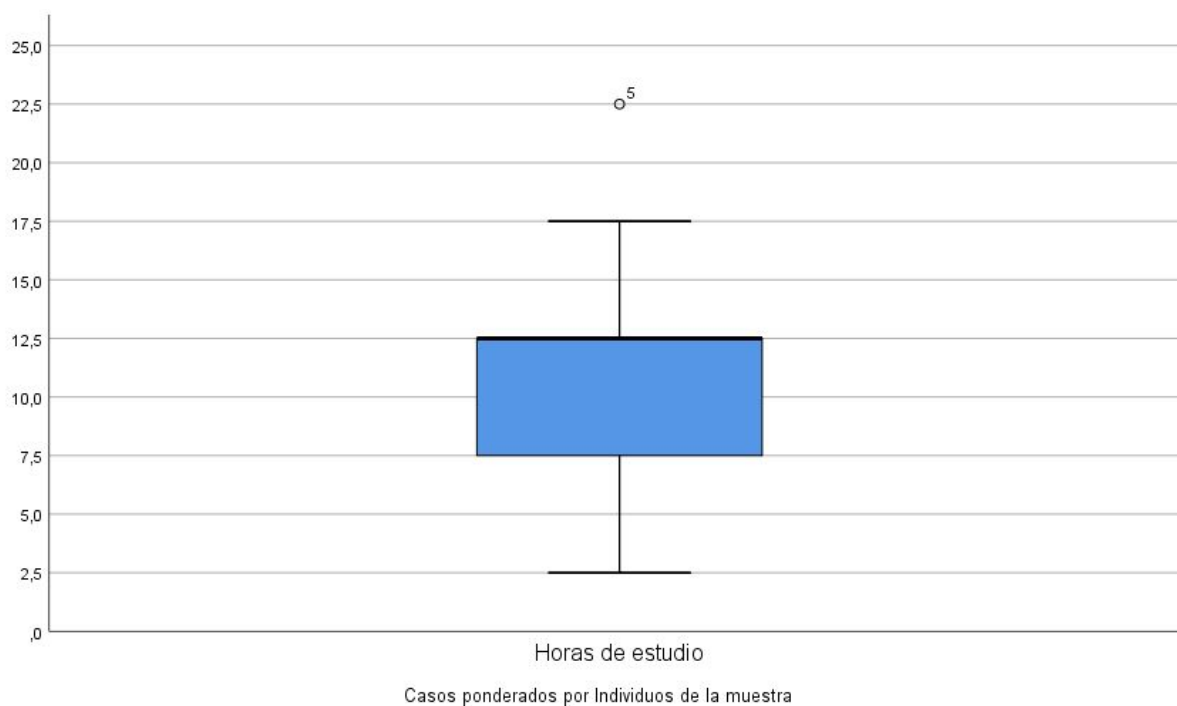
Con estos datos, empezamos a realizar nuestra Tabla de vallas.

	Horas de estudio
<u>Media</u>	<u>11,07</u>
<u>IQR</u>	<u>5</u>
<u>Valores Atípicos</u>	<u>22,5</u>
<u>Valores Extremos</u>	<u>---</u>
<u>Desviación</u>	<u>5,88</u>
Va. Ext. Inf.	-7,5
Va. Int. Inf.	0
Va. Ady. Inf.	2,5
Cuartil 1	7,5
Mediana	12,5
Cuartil 3	12,5
Va. Ady. Sup.	17,5
Va. Int. Sup.	20
Va. Ext. Sup.	27,5

Ahora, creamos el gráfico de la caja como hicimos en el apartado anterior, Gráficos > Cuadro de diálogo antiguo > Diagrama de Cajas > Simples | Resumen para distintas variables, y seleccionamos las horas de estudio.



Y ajustamos los valores para que el gráfico sea legible fácilmente.



Lo primero que podemos presenciar es que la media y el Q1 coinciden. La caja es sesgada a la derecha, aunque sus valores no tiendan a concentrarse en la misma dirección. Hay un valor atípico que es el 22,5.

b) ¿Cuál es el tamaño de la muestra?

140

Obtén los parámetros de centralización.

Cuartil 1	7,5
Mediana	12,5
Cuartil 3	12,5

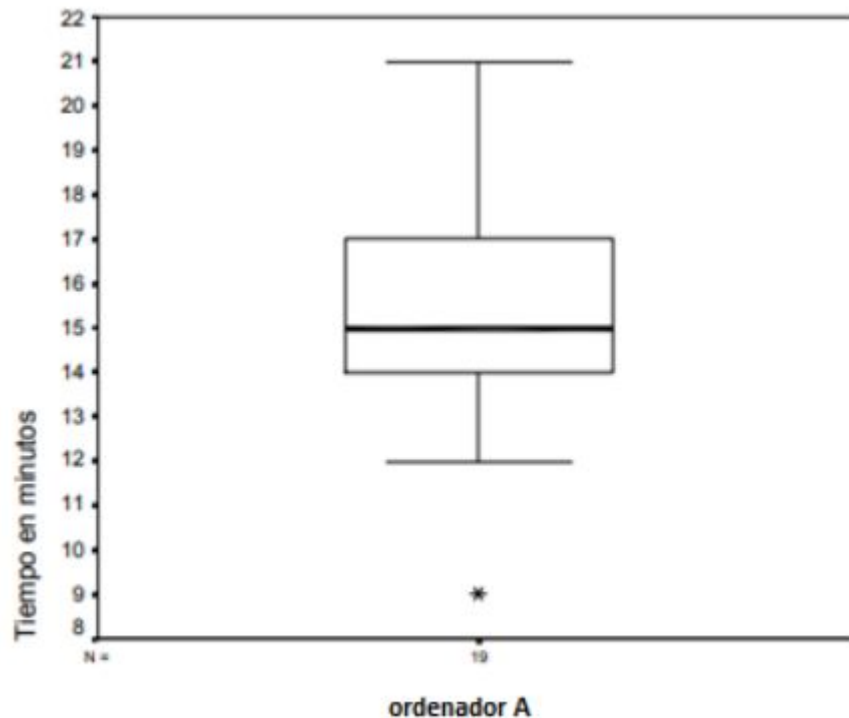
¿Cuál será más adecuado para esta muestra la media o la mediana?

La mediana

¿Por qué?

Ya que el hecho de que haya 50 personas estudiando alrededor de 12,5 horas, nos dice que más o menos la mayoría de personas estudian entorno a esa misma cantidad de horas.

3. En un determinado centro de cálculo están interesados en medir durante 19 días el tiempo (en minutos) de realización de la copia de seguridad de dos ordenadores (A y B). El gráfico muestra los tiempos medidos para el ordenador A.



a. ¿Cuál es el valor aproximado de las medidas de dispersión del tiempo para el ordenador A?

Interpretando el gráfico, podemos ver que $Q1 = 14$ y $Q3 = 17$, entonces $IQR = Q3 - Q1 \Rightarrow 17 - 14 = 3$

b. ¿En qué porcentaje de días se realiza la copia de seguridad antes de los 14 minutos con el ordenador A? Justifica la respuesta.

El 14 al ser el $Q1$, podemos asegurar que quedaría 25% del total de minutos del 14 al 9.

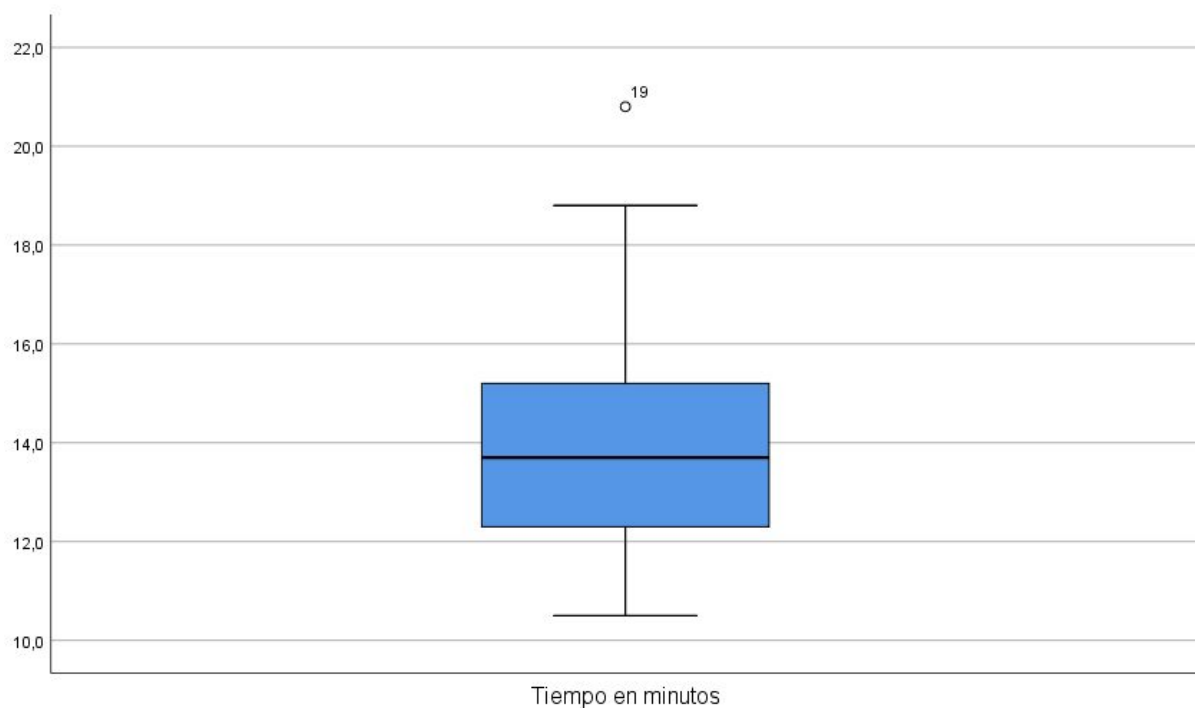
c. Los datos del ordenador B para los 19 días son los siguientes: 10.5, 11.3, 11.9, 12, 12.3, 12.3, 12.5, 12.7, 13.4, 13.7, 13.8, 14.2, 14.8, 15.1, 15.3, 16.7, 16.8, 18.8 y 20.8.

Primero creamos la variable e introducimos los datos:

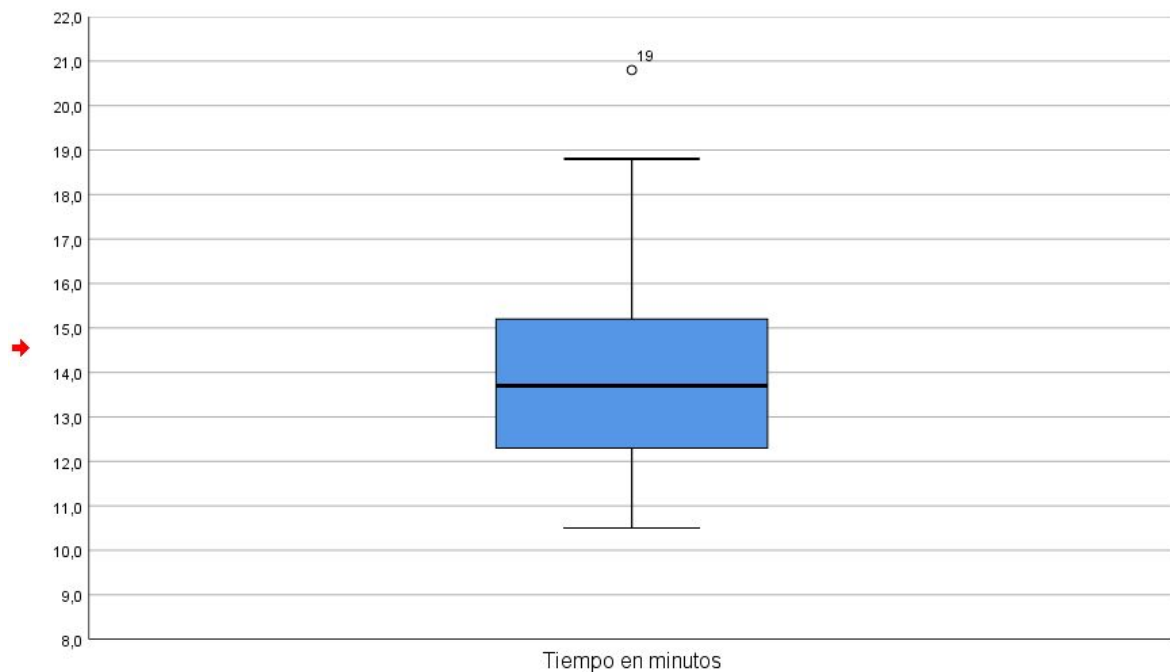
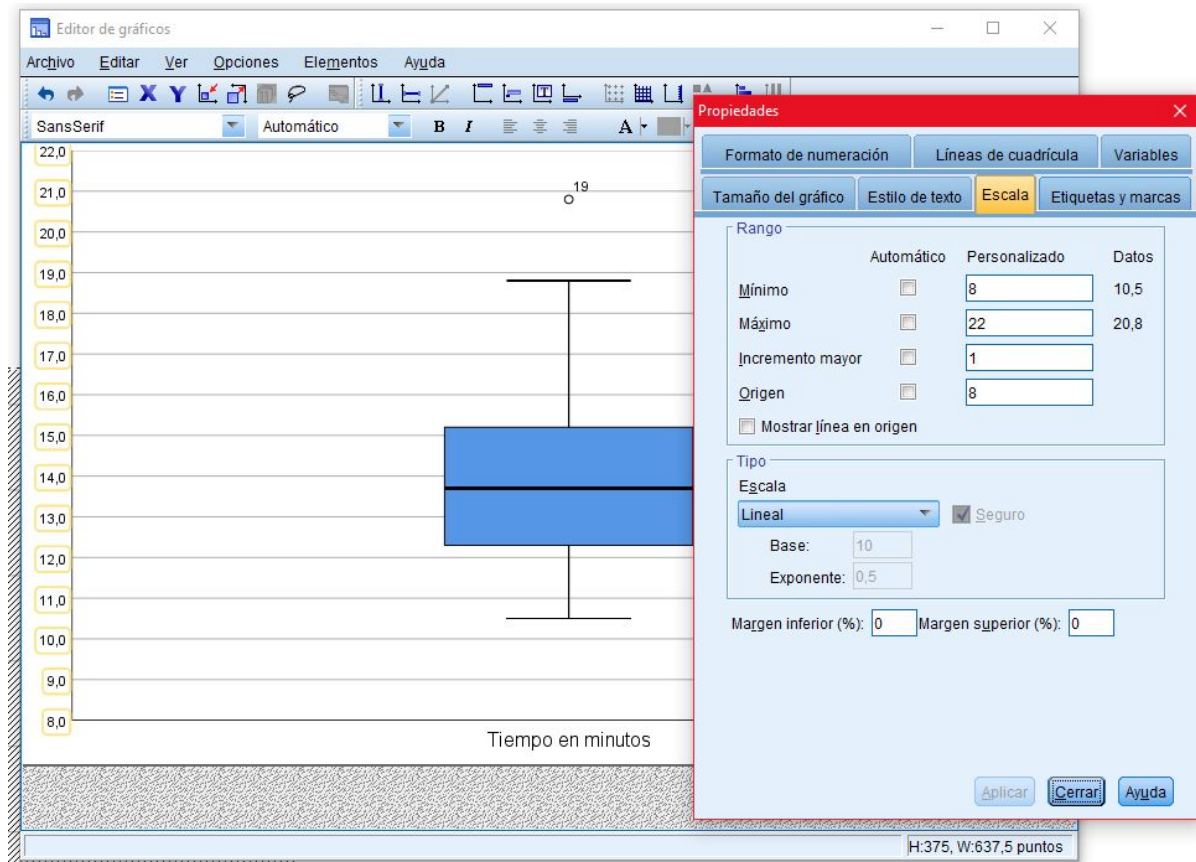
	Minutos				
1	10,5	8	12,7		
2	11,3	9	13,4		
3	11,9	10	13,7	15	15,3
4	12,0	11	13,8	16	16,7
5	12,3	12	14,2	17	16,8
6	12,3	13	14,8	18	18,8
7	12,5	14	15,1	19	20,8

Construye el diagrama de caja correspondiente a los tiempos de realización de la copia de seguridad con el ordenador B. Interpreta el resultado.

Para esto, haremos Gráficos > Cuadro de dialogo antiguo > Diagrama de Cajas > Simples | Resumen para distintas variables, y seleccionamos la variable.



Vamos a ajustarlo un poco, hacemos doble click sobre el eje Y, y ajustamos los datos, así lo veremos en la misma escala que la del otro grafo.



d. *¿Qué ordenador es más fiable?*

Sacamos los percentiles de B

Estadísticos		
Tiempo en minutos		
N	Válido	19
	Perdidos	0
Percentiles	25	12,300
	50	13,700
	75	15,300

$\text{IQR de B} = 15,3 - 12,3 = 3$

¿Cuál proporciona una mayor velocidad media de realización de copia?

La media de A, interpretado el grafo, sería $(12 + 21) / 2 = 16,5$

La media de B es 14,15.

A tiene mayor velocidad media.