

ESTADISTICA

Practica 5

Análisis de una variable medible. Gráficos caja.

Grado en ingeniería informática

Francisco Joaquín Murcia Gómez 48734281H

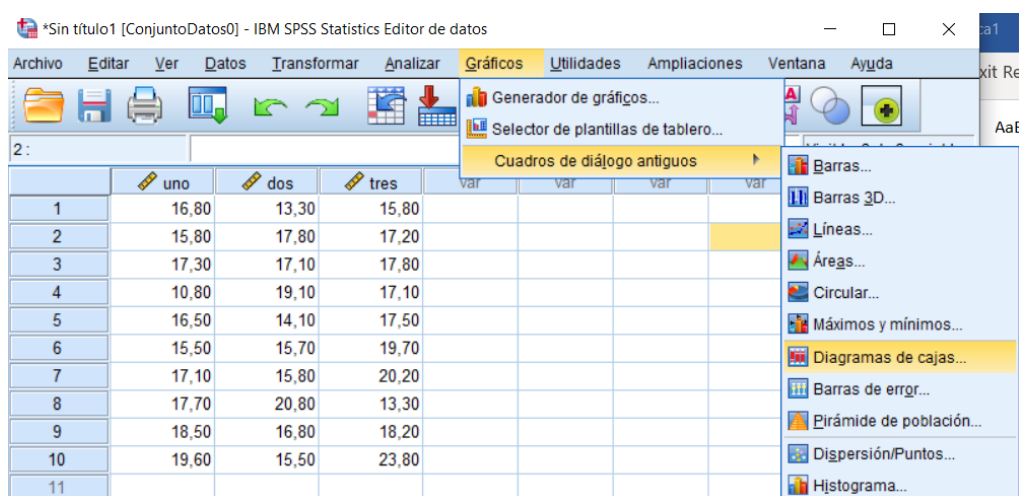
Grupo 1

- 1) En tres computadores de un determinado centro de cálculo, se ha medido durante ocho días el tiempo (en minutos) de realización de la copia de seguridad, obteniéndose los siguientes resultados:

COMPUTADOR UNO	COMPUTADOR DOS	COMPUTADOR TRES
16,8	13,3	15,8
15,8	17,8	17,2
17,3	17,1	17,8
10,8	19,1	17,1
16,5	14,1	17,5
15,5	15,7	19,7
17,1	15,8	20,2
17,7	20,8	13,3
18,5	16,8	18,2
19,6	15,5	23,8

a) **Obtener los gráficos caja para los resultados de los tres computadores**

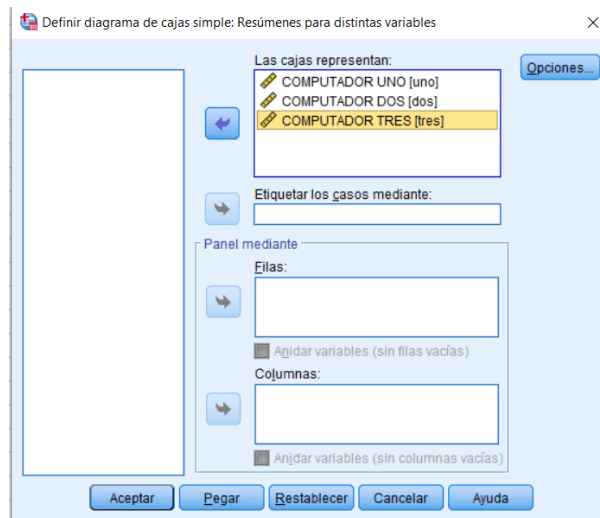
Para hacer los gráficos caja en el menú de arriba en la ruta "gráficos> gráficos antiguos> diagrama de cajas".



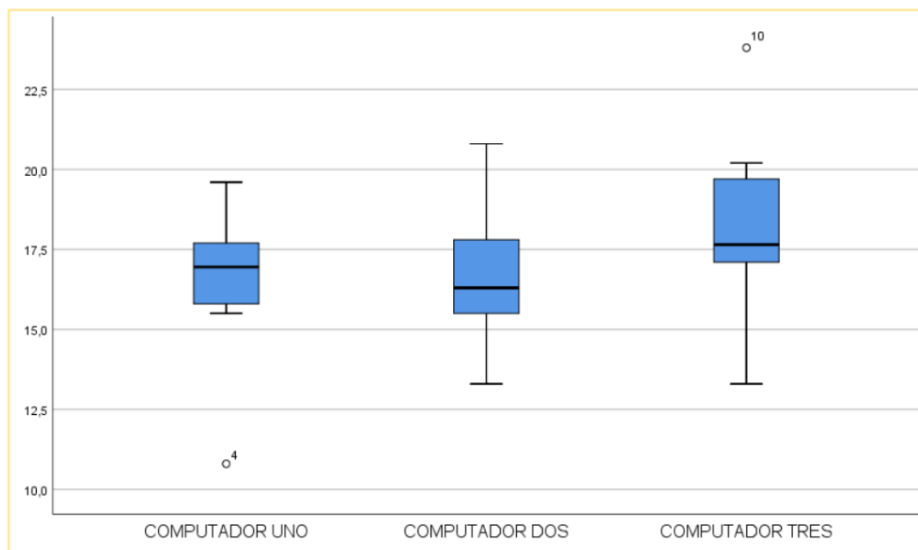
Nos aparecerá la siguiente ventana:



Allí seleccionamos "simple" y "resúmenes de las distintas variables", nos aparecerá la siguiente ventana, la cual, seleccionamos todas las variables y le damos a aceptar como se ve abajo.



El resultado es el siguiente.



b) ¿Cuál de ellos tiene mayor dispersión? ¿Son simétricos? Indica los valores extremos y atípicos si los hubiera.

El ordenador con mayor dispersión es el dos, porque las líneas que representan la dispersión son las más largas. Tampoco son simétricos porque dichas barras no son iguales arriba y abajo y los datos no se distribuyen por igual arriba y abajo. También encontramos dos valores atípicos en el computador uno en la posición 4 con un valor de 10,8 minutos por debajo del resto y en el tres en la posición 10 con un valor de 23,8 minutos por encima del resto, no se encuentran extremos.

No obstante, al hacer los cálculos a mano descubrimos vemos que el valor atípico en la posición 10 del tercer ordenador es erróneo, esta demostración lo he obtenido con los siguientes cálculos:

Q_x =cuartil número x

n=elementos de la muestra

a=posición de la muestra

M=mediana=Q2

$Q1 = (n+1)/4 = (10+1)/4 = a*2,75$

$M = (n+1)/2 = (10+1)/2 = a*5,5$

$Q3 = 3(n+1)/4 = 3(10+1)/4 = a*8,25$

IQR(intervalo intercuartílico)=Q3-Q1

Vext sup (valla exterior superior)=Q3+(3*IQR)

Vint sup (valla interior superior)=Q3+(1,5*IQR)

Vint inf (valla interior inferior)=Q1-(1,5*IQR)

Vext inf (valla exterior inferior)=Q1-(3*IQR)

Estas formulas se han introducido en un Excel junto con la muestra* dando como resultado el siguiente:

muestra						
ordenador 1	ordenador 2	ordenador 3		operaciones		
				ordenador 1	ordenador 2	ordenador 3
10,8	13,3	13,3				
15,5	14,1	15,8	Q1	15,7250	15,1500	16,7750
15,8	15,5	17,1	M	16,9500	16,3000	17,6500
16,5	15,7	17,2	Q3	17,9000	18,1250	19,8250
16,8	15,8	17,5	IQR	2,1750	2,9750	3,0500
17,1	16,8	17,8	Vext sup	24,4250	24,6500	26,3500
17,3	17,1	18,2	Vint sup	21,1625	22,5875	24,4000
17,7	17,8	19,7	Vint inf	12,4625	10,6875	12,2000
18,5	19,1	20,2	Vext inf	9,2000	6,2250	7,6250
19,6	20,8	23,8				

Como podemos observar el limite de la valla interior superior en el ordenador 3 es de 24,4, el valor de la posición 10 del tercer ordenador es 23,8, al ser dicho valor inferior a la valla interior superior, no seria un valor atípico pese a que el spss indique que sí.

*los datos de la muestra se han ordenado de menor a mayor para facilitar los cálculos

c) ¿Qué computador es más fiable? ¿Cuál proporciona una mayor velocidad media de realización de copia?

El computador 1 debido a que es el que menor dispersión con lo cual va a tener un rendimiento similar al esperado, y lo mas importante es que es el computador que mas concentrados tiene los valores.

Para hacer la media de velocidades se accede a la ruta "analizar> frecuencias> analíticos descriptivos" se accede a "estadísticos" se selecciona media y se da a aceptar, y muestra la siguiente tabla.

Estadísticos		COMPUTADOR UNO	COMPUTADOR DOS	COMPUTADOR TRES
N	Válido	10	10	10
	Perdidos	0	0	0
Media		16,5600	16,6000	18,0600
Mediana		16,9500	16,3000	17,6500
Moda		10,80 ^a	13,30 ^a	13,30 ^a

Como podemos observar, el computador 1 es el que ofrece menor tiempo tarda al copiar tiene, por lo tanto, es el que mayor velocidad ofrece.


2) Realizada una encuesta por muestreo entre usuarios de Internet, se ha obtenido los siguientes datos:


Minutos de conexión/día	Individuos de la muestra
34.5 – 39.5	10
39.5 – 44.5	15
44.5 – 49.5	25
49.5 – 54.5	18
54.5 – 59.5	22

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

Para la realización primero ponderamos los datos en la ruta "datos> ponderar casos" (como se hizo en prácticas anteriores) ponemos los minutos (el cual construimos la variable con su valor medio como se ve a la derecha), con los individuos como se ve en la foto de la derecha.

individuos	tiempo
10	37
15	42
25	47
18	52
22	57


 Ponderar casos

 Minutos de conexió...

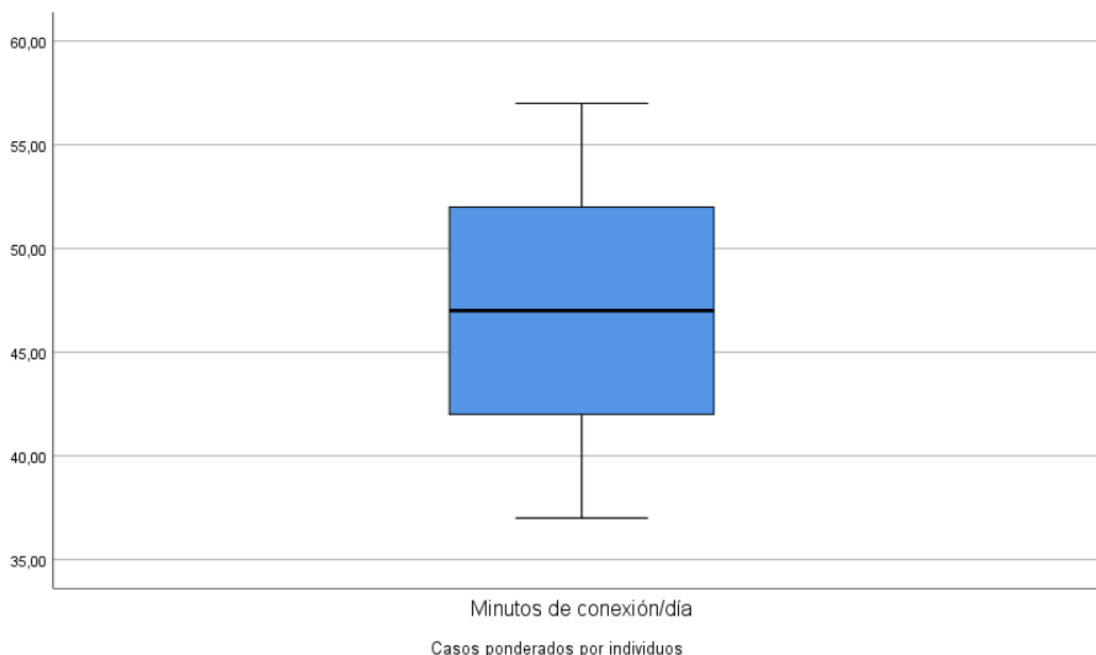
☐ No ponderar casos

☒ Ponderar casos mediante

Variable de frecuencia:

 individuos

Una vez ponderado hacemos el grafico caja de la variable tiempo con el procedimiento visto en el ejercicio 1, dando como resultado:



El resultado es un gráfico aparentemente simétrico ya que las barras que representan la dispersión son iguales y los datos se distribuyen arriba y abajo por igual, pero al hacer la moda, media y mediana, da que la media no es igual a la moda y la mediana (moda=mediana=47 y media=48.5) por lo tanto no es simétrico, pero las colas de dispersión lo son.

b) Obtén los parámetros de centralización. ¿Cuál será más adecuado para esta muestra la media o la mediana? ¿Por qué?

En la ruta "analizar> estadísticos descriptivos> frecuencias" en la parte de estadísticos seleccionamos moda, media y mediana (mismo procedimiento que en la practica 4), da como resultado los siguientes:

N	Válido	90
	Perdidos	0
Media		48,50
Mediana		47,00
Moda		47

En este caso seria mas adecuado utilizar seria la media ya que no hay valores atípicos y no puede ser influenciada por dichos valores.

3) La sección de control de calidad de una empresa ha medido la cantidad promedio en Kg. de llenado de recipientes de un determinado producto de una línea de producción:

12.00, 11.97, 12.10, 12.03, 12.01, 11.81, 11.91, 11.98, 12.30, 11.89, 12.05, 11.38, 12.89, 11.78, 12.72, 11.08, 11.78, 11.90, 11.44, 12.50, 13.75, 11.78, 11.91, 12.68, 11.39, 11.94, 13.45, 11.98, 11.27, 10.72, 11.93, 11.59, 12.79, 11.39, 13.05, 12.50, 11.58, 11.87, 11.78, 11.05, 12.02, 12.43

a) Calcular la media, mediana desviación típica y tercer cuartil.

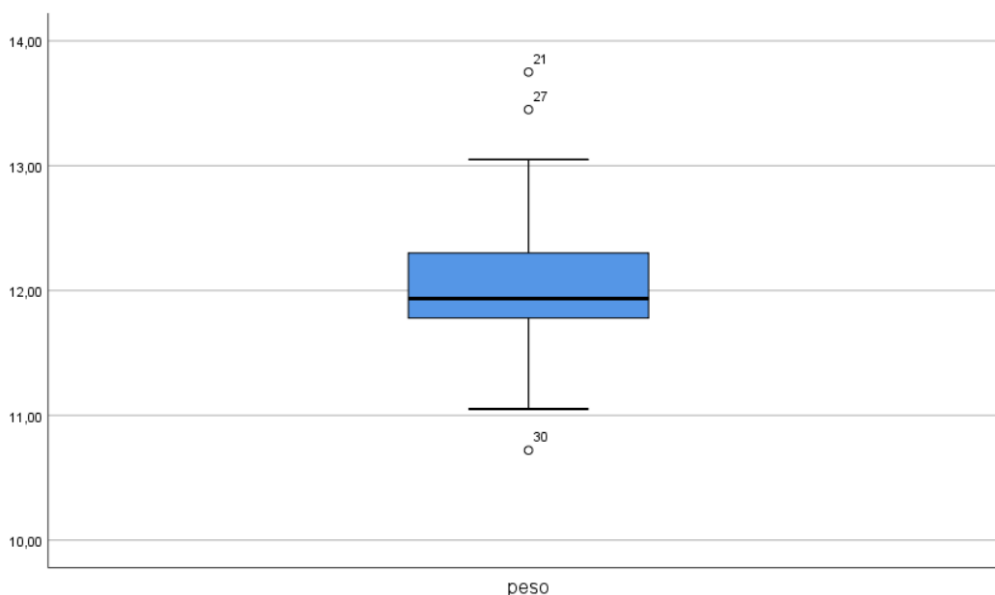
Seguimos la siguiente ruta: analizar> estadísticos descriptivos> frecuencias; y añadimos la variable, después pulsamos el botón estadístico rellenamos los siguientes valores que se ven en la foto de la derecha

El resultado es la imagen de abajo.

N	Válido	42
	Perdidos	0
<hr/>		
Media		12,0088
Mediana		11,9350
Desv. Desviación		,61390
Percentiles	75	12,3325

b) Construir el gráfico caja de los datos anteriores.

Utilizando el procedimiento del ejercicio 1 el grafico sería el siguiente:



c) Explicar los resultados obtenidos

En la siguiente muestra podemos ver en el grafico caja que existen tres valores atípicos, dos arriba en la posición 21 con un valor de 13,75 y en la posición 27 con un valor de 13,45; y uno inferior en la posición 30 con un valor de 10,72. Al existir dichos valores podemos decir que es mas adecuado utilizar la mediana para hallar el valor central porque no está influenciada por dichos valores.

Al tener la mediana con tendencia hacia abajo los valores menores están más concentrados mientras que los valores mayores están más dispersos, como se puede observar con los cuartiles 1º(11,7325) 2º(11,9350) y 3º(12,3325)

Las colas de dispersión a primera vista son simétricas, pero es un grafico asimétrico porque la moda media y mediana no coinciden y los datos no se encuentran distribuidos uniformemente arriba y debajo de la mediana.

Para comprobar que no ocurra lo mismo que en el ejercicio 1, haces las operaciones de dicho ejercicio.

Como podemos observar en la imagen de la derecha los valores atípicos de la posición 21 con un valor de 13,75 y en la posición 27 con un valor de 13,45, se encuentran entre la valla interior superior y la exterior superior; y el de la posición 30 con un valor de 10,72, se encuentra ente la valla interior inferior y la valla exterior inferior. Por lo tanto, son correctos.

Kg Llenado			
10,72	11,94		
11,05	11,97		
11,08	11,98		
11,27	11,98		
11,38	12,00		
11,39	12,01		
11,39	12,02		
11,44	12,03		
11,58	12,05		
11,59	12,10		
11,78	12,30		
11,78	12,43		
11,78	12,50		
11,78	12,50		
11,81	12,68		
11,87	12,72		
11,89	12,79		
11,90	12,89		
11,91	13,05		
11,91	13,45		
11,93	13,75		

operaciones	
	Kg Llenado
Q1	11,7325
Q2 = M	11,935
Q3	12,3325
IQR	0,6
Vext sup	14,1325
Vint sup	13,2325
Vint inf	10,8325
Vext inf	9,9325