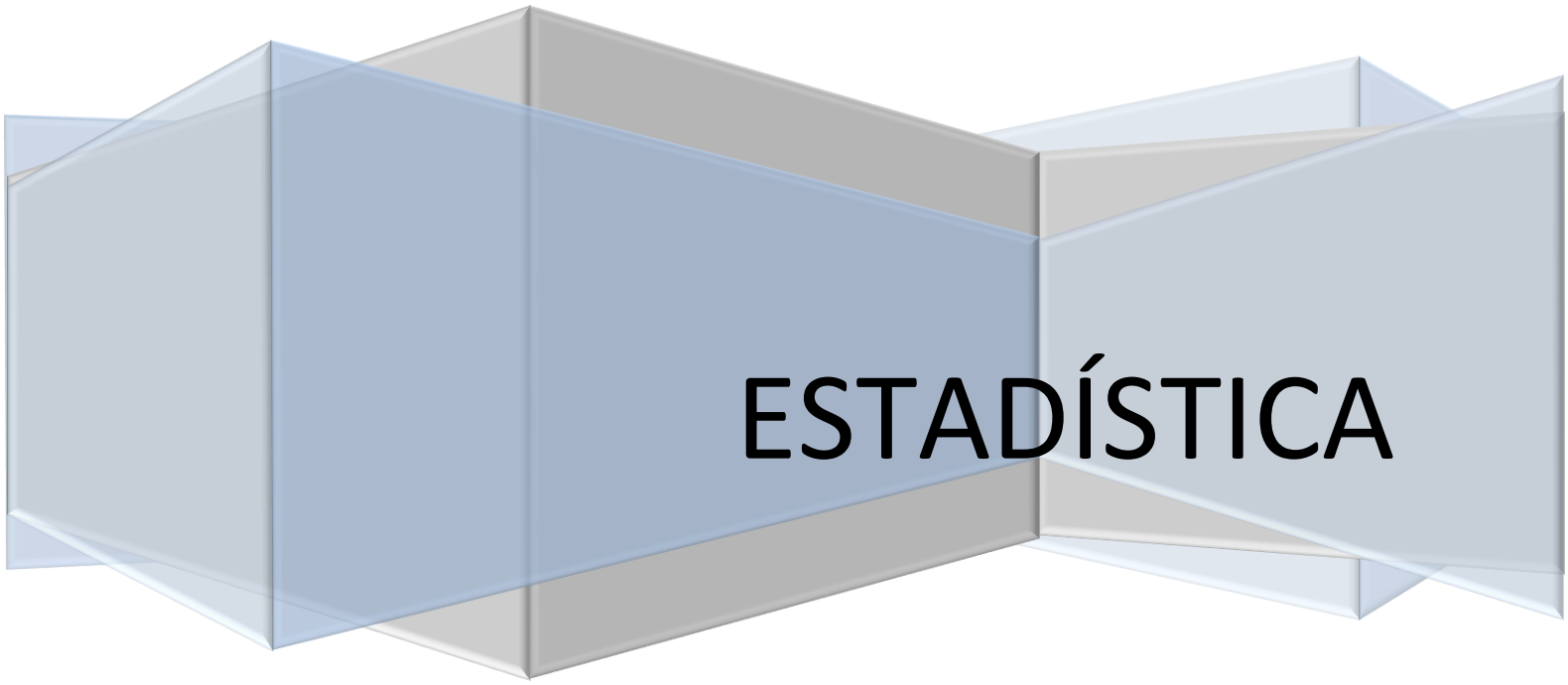
PRÁCTICA 3

ESTADÍSTICA

53247378D

JAVIER RIVILLA ARREDONDO



Ejercicios

1. El rendimiento –referido a capacidad de procesamiento- de los 60 clusters de los disntos departamentos de una gran empresa es el siguiente, medido en GFLOPS (10^9 operaciones en coma flotante):

Rendimiento (GFlops)	Número de clusters
Menos de 31	1
De 31 a 60	1
De 61 a 90	17
De 91 a 120	30
De 121 a 150	3
De a 151 a 180	4
De 181 a 210	2
Más de 210	2

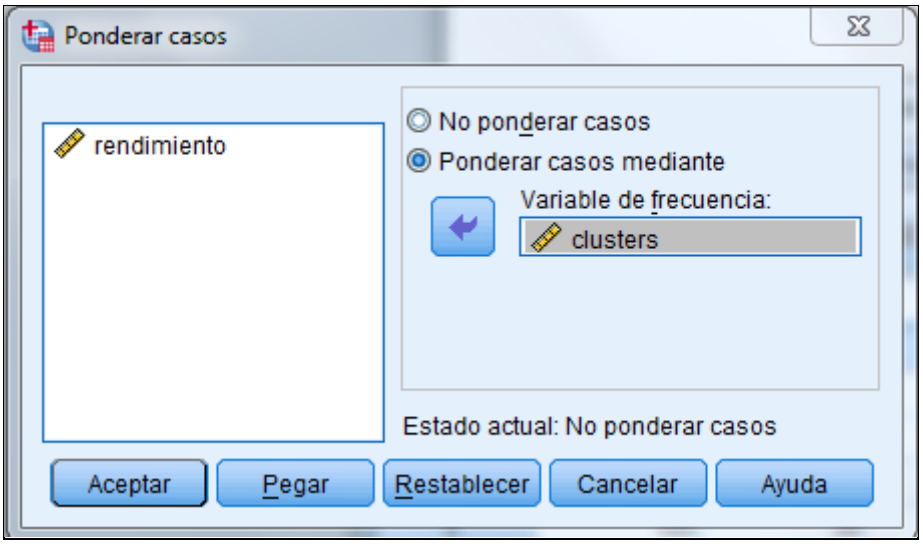
Para analizar la distribución de la capacidad de procesamiento disponible en la empresa, se pide:

a) Construye la tabla de frecuencias completa

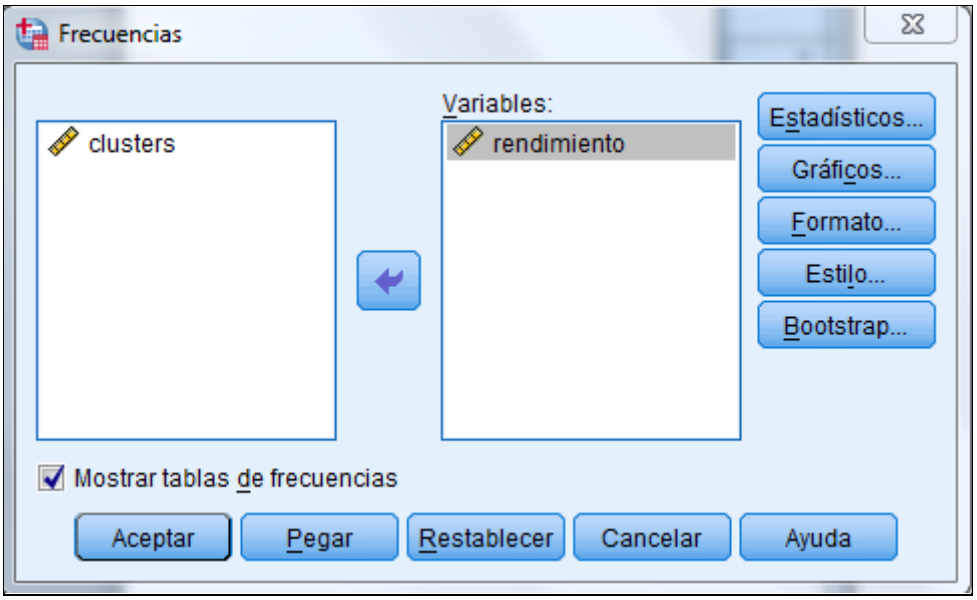
Introducimos los datos en el SPPS:

	rendimiento	clusters
1	15	1
2	45	1
3	75	17
4	105	30
5	135	3
6	165	4
7	195	2
8	225	2

Una vez hemos introducidos los datos en el SPSS lo que haremos será irnos a **Datos->Ponderar Casos** y seleccionamos la variable clusters.



Aceptamos la ponderación e iremos a **Analizar -> Estadísticos Descriptivos -> Frecuencias** y seleccionamos la variable *rendimiento*.



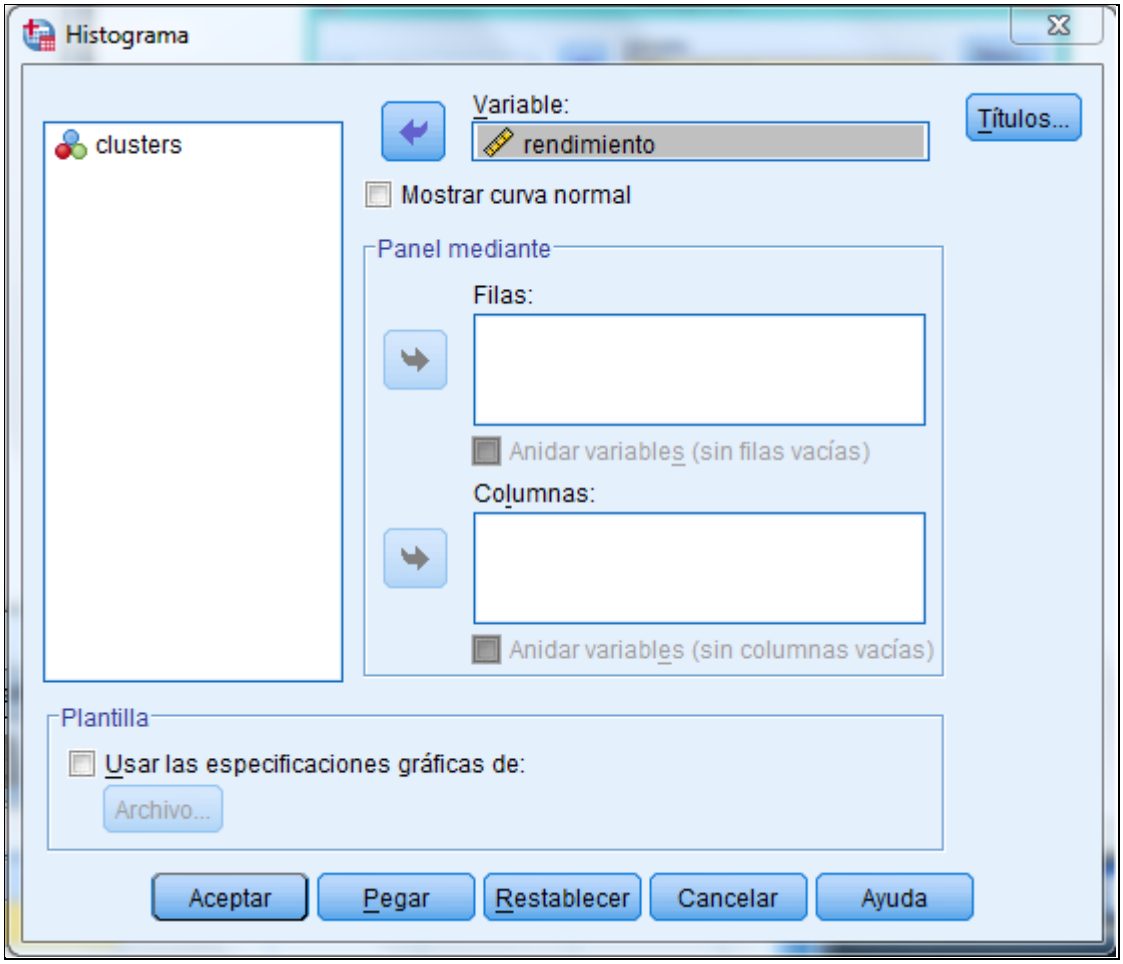
Esto nos mostrará la tabla de frecuencias de la variable rendimiento, que a continuación se mostrará:

rendimiento

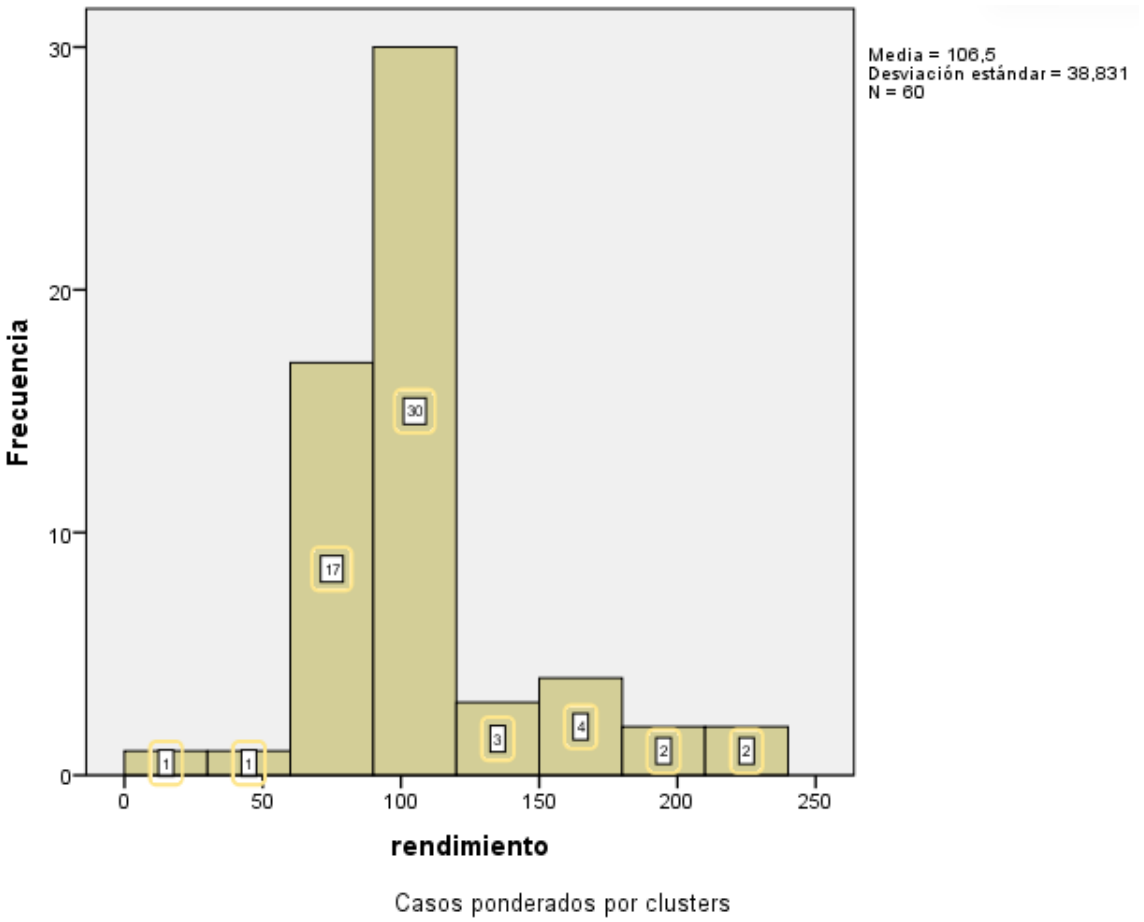
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	15	1	1,7	1,7	1,7
	45	1	1,7	1,7	3,3
	75	17	28,3	28,3	31,7
	105	30	50,0	50,0	81,7
	135	3	5,0	5,0	86,7
	165	4	6,7	6,7	93,3
	195	2	3,3	3,3	96,7
	225	2	3,3	3,3	100,0
	Total	60	100,0	100,0	

b) Representa el histograma

Para mostrar este, tenemos que seleccionar **Gráficos->Cuadros de diálogos antiguos:**



Seleccionamos la variable rendimiento y obtenemos el siguiente histograma:



c) Explica e interpreta los resultados obtenidos en los apartados anteriores

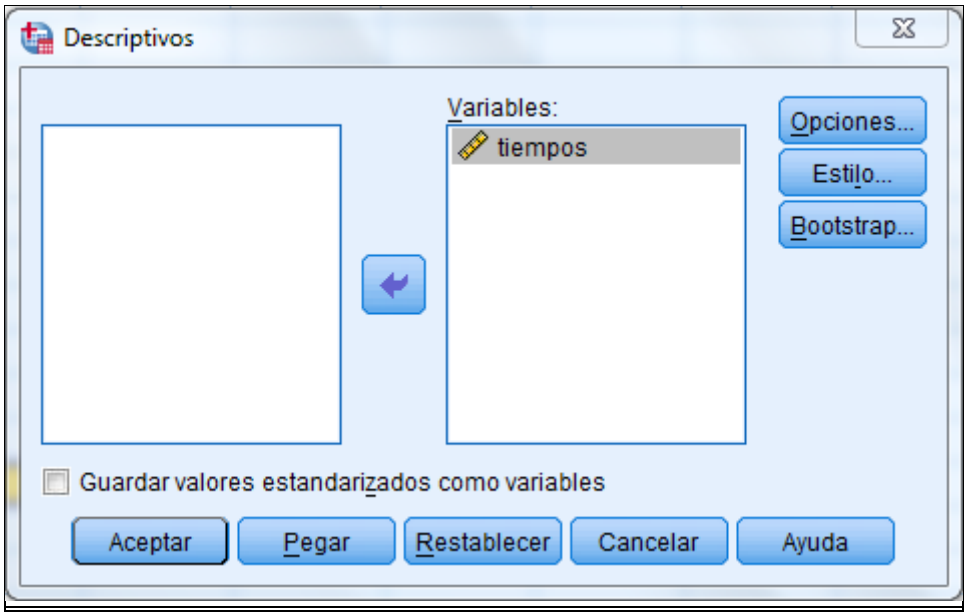
Como podemos observar tanto en la tabla de frecuencias el porcentaje más alto es el 105(de 90 a 120) con un porcentaje del 50%, siendo el que más rendimiento tendrá. Siguiéndole el valor de 75 (de 60 a 90) con un 28,3%. Los que tendrán menos rendimiento serán los valores que estén entre 0-30 y 30-60. También podemos observar esto en el histograma. Viendo la frecuencia de cada valor más fácil de intuir.

2. De la misma empresa, se ha contabilizado la capacidad de almacenamiento de los clusters medida en GB obteniendo:

730, 600, 680, 590, 620, 760, 830, 610, 800, 790, 600, 840, 612, 935, 940,
650, 810, 690, 740, 750, 690, 800, 680, 750, 800, 900, 602, 614, 880, 699,
650, 780, 820, 740, 790, 630, 800, 770, 760, 670, 920, 850, 813, 875, 625,
650, 700, 680, 800, 770, 730, 660, 810, 780, 750, 911, 950, 710, 666, 870,
690, 710, 790, 700, 640, 720, 820, 740, 790, 630, 888, 601, 911, 949, 812

Agrupando los tiempos en intervalos de clase de longitud 50, obtener el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas. Explica e interpreta los resultados obtenidos.

Introducimos los datos en el SPPS y una vez introducimos haremos **analizar->estadísticos descriptivos->descriptivos.**



Y nos mostrará la siguiente información de los tiempos:

Estadísticos descriptivos

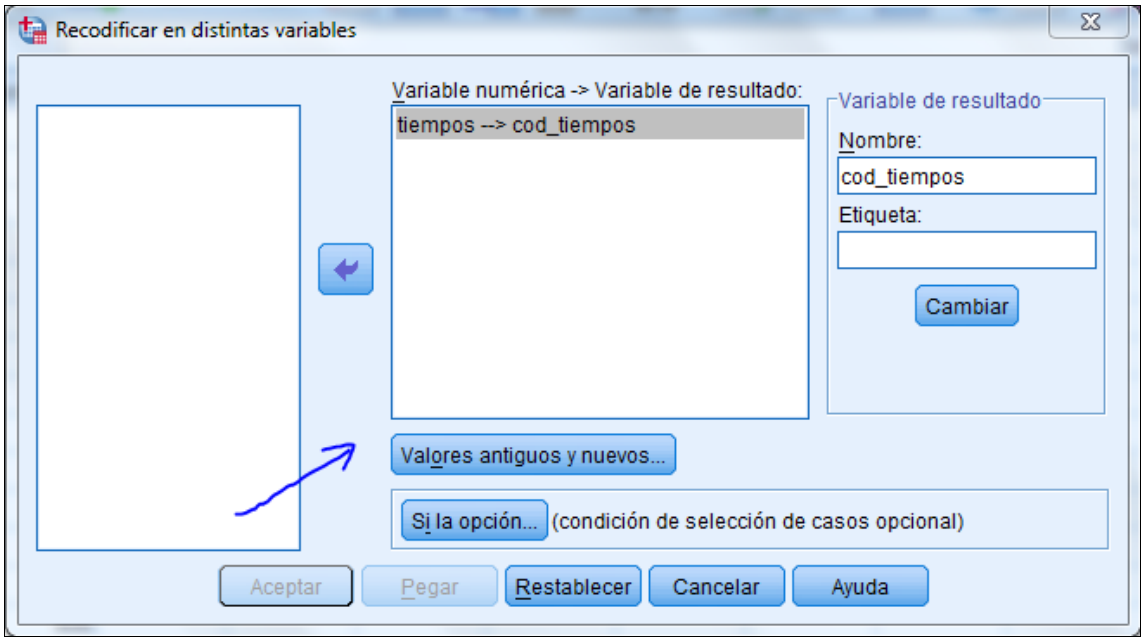
	N	Mínimo	Máximo	Media	Desviación estándar
tiempos	75	590	950	750,84	98,997
N válido (por lista)	75				

Cuando tengamos la siguiente información haremos los rangos con longitud 50. Sacando su valor medio.

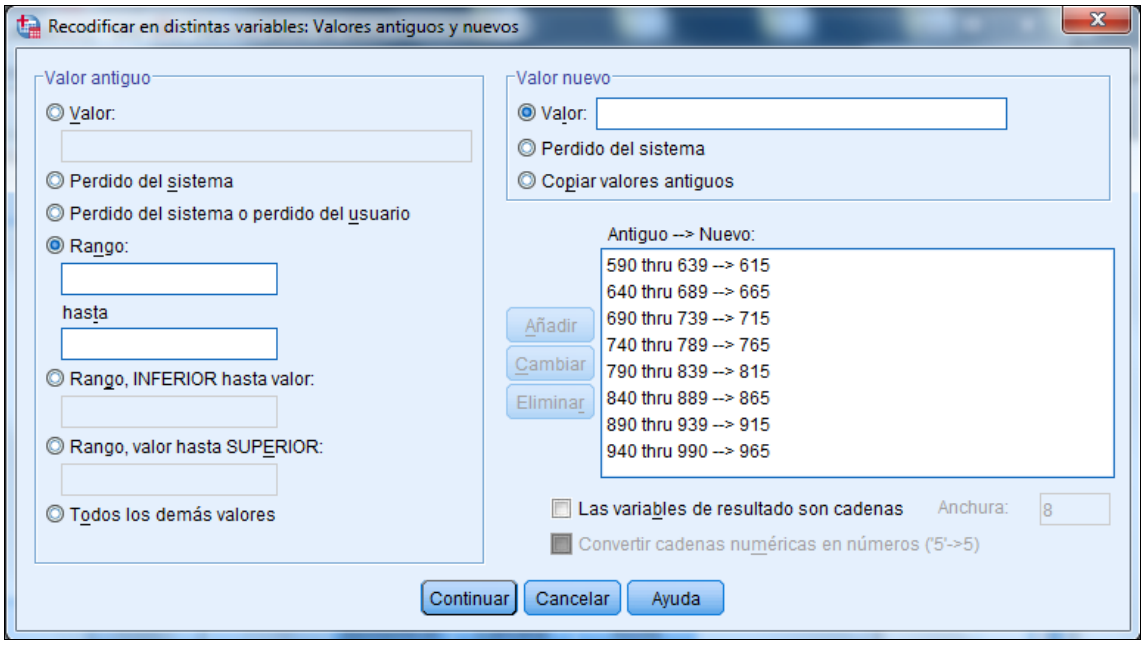
[590-640[-> valor medio =615
[640-690[->valor medio=665
[690-740[-> valor medio=715
[740-790[-> valor medio=765
[790-840[-> valor medio=815
[840-890[-> valor medio=865
[890-940[-> valor medio=915
[940-990[-> valor medio=965

Una vez realizado los rangos con esa longitud y obtenemos el valor medio, recodificaremos la variable tiempos en otra variable, para ellos crearemos una nueva variable llamada cod_tiempos, utilizando los rangos que hemos obtenido anteriormente.

Transformar->Recodificar en distintas variables



Le damos a valores antiguos y nuevos… y hacemos lo siguiente:

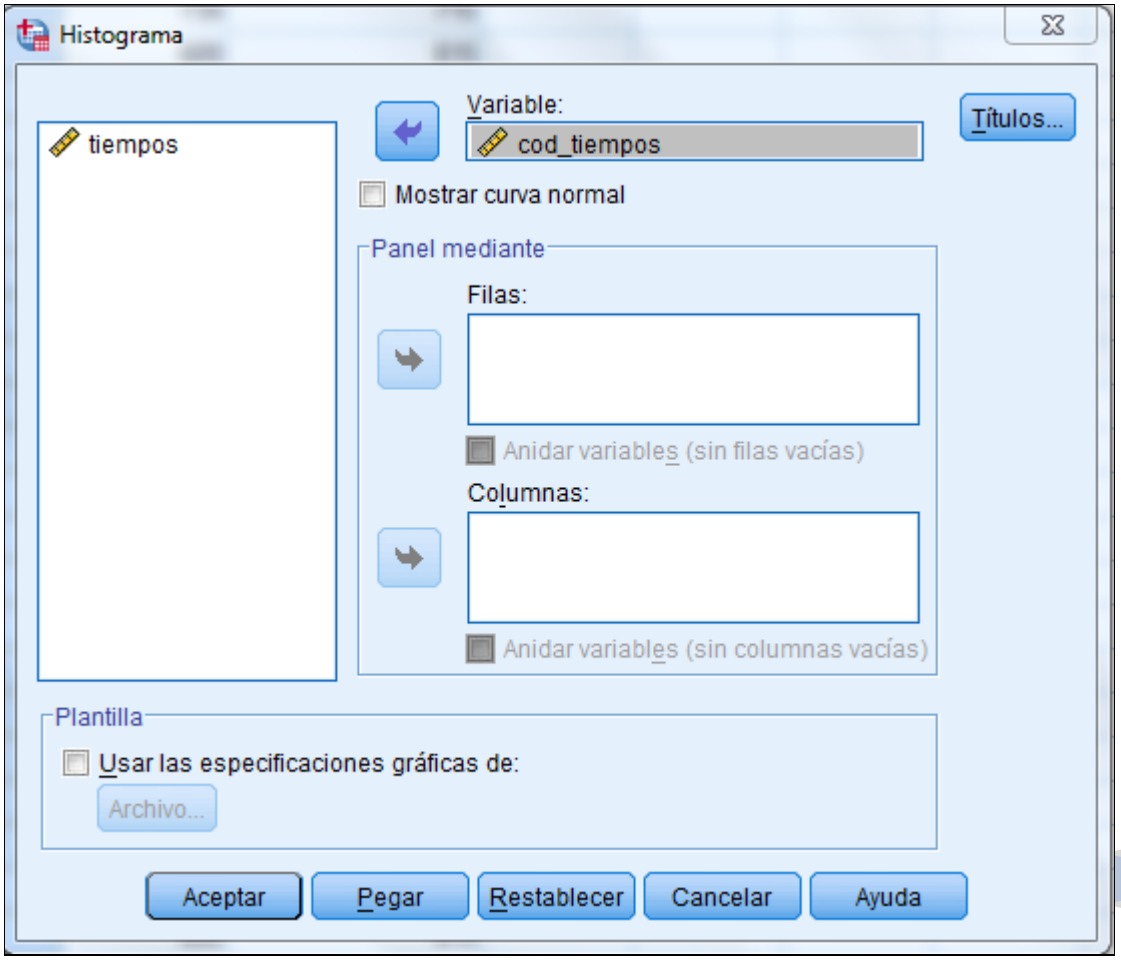


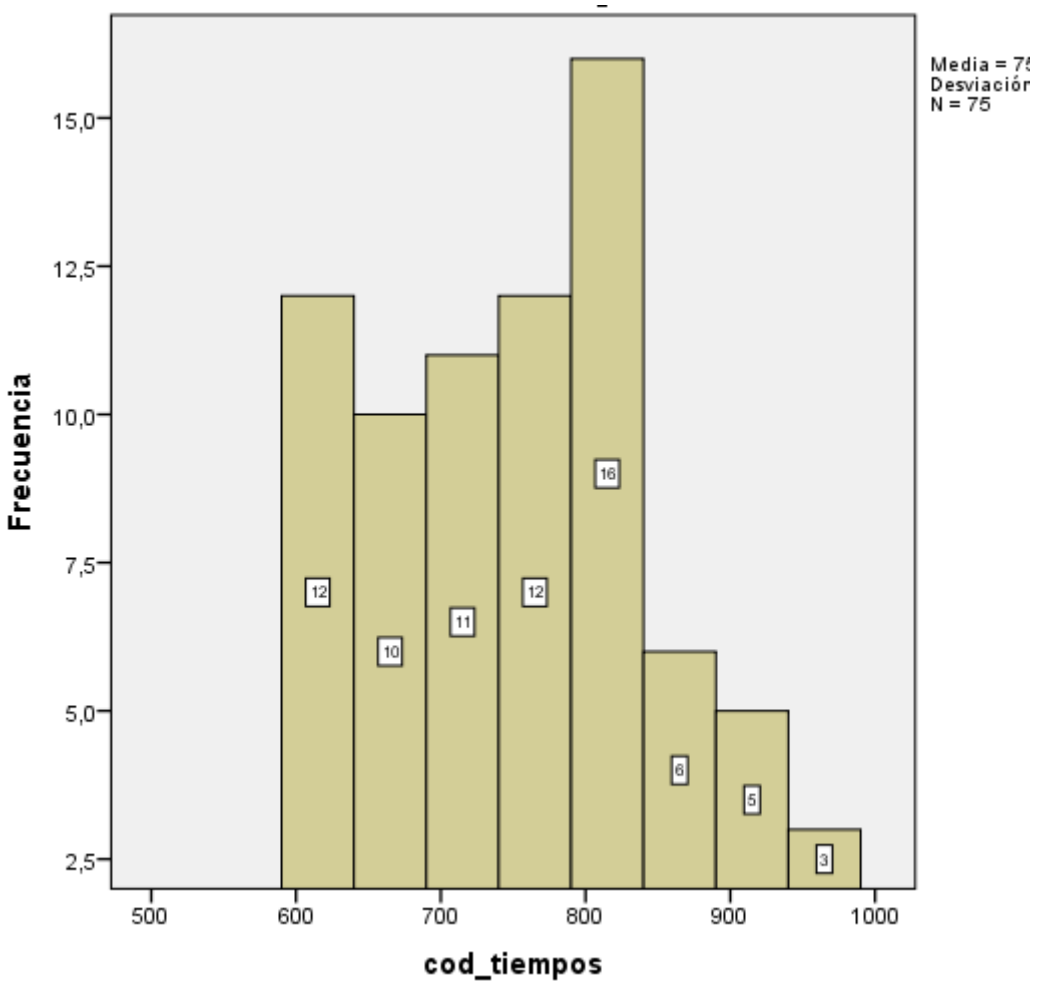
Y nos saldrá la siguiente columna “cod_tiempos” :

	tiempos	cod_tiempos
1	730	715
2	600	615
3	680	665
4	590	615
5	620	615
6	760	765
7	830	815
8	610	615
9	800	815
10	790	815
11	600	615
12	840	865
13	612	615
14	935	915
15	940	965
16	650	665
17	810	815
18	690	715
19	740	765
20	750	765
21	690	715
22	800	815

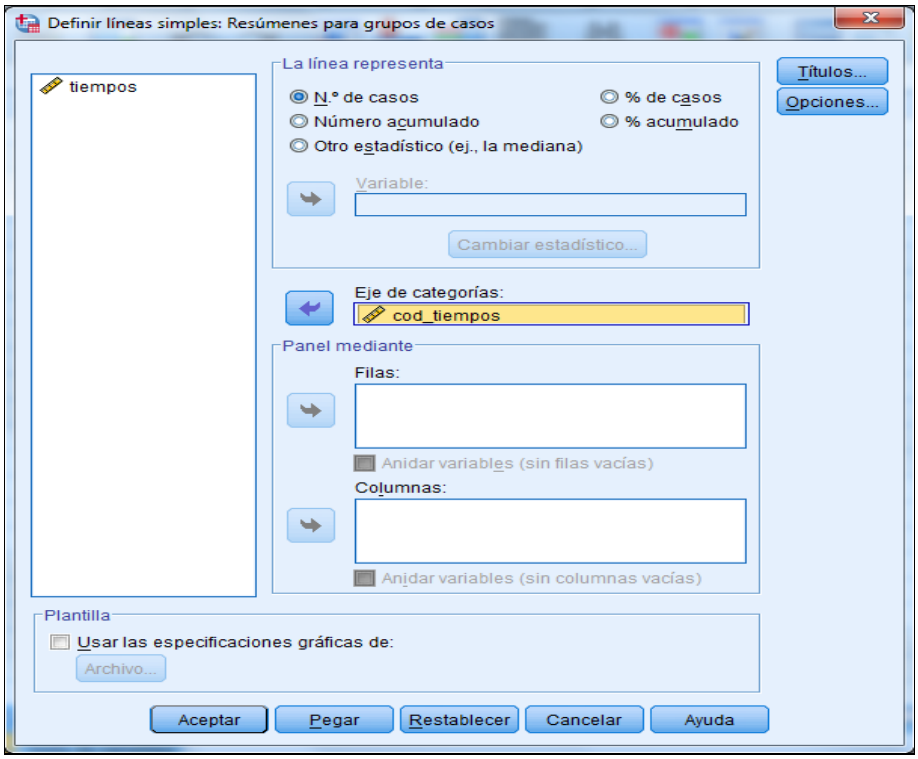
Como podemos ver el valor por ejemplo x donde este pertenece al rango [x-100[y no al rango [10-x[por esto, es un intervalo abierto.

Para obtener el histograma iremos a “Gráficos” y luego a “Cuadros de diálogos antiguos” -> “Histograma”.

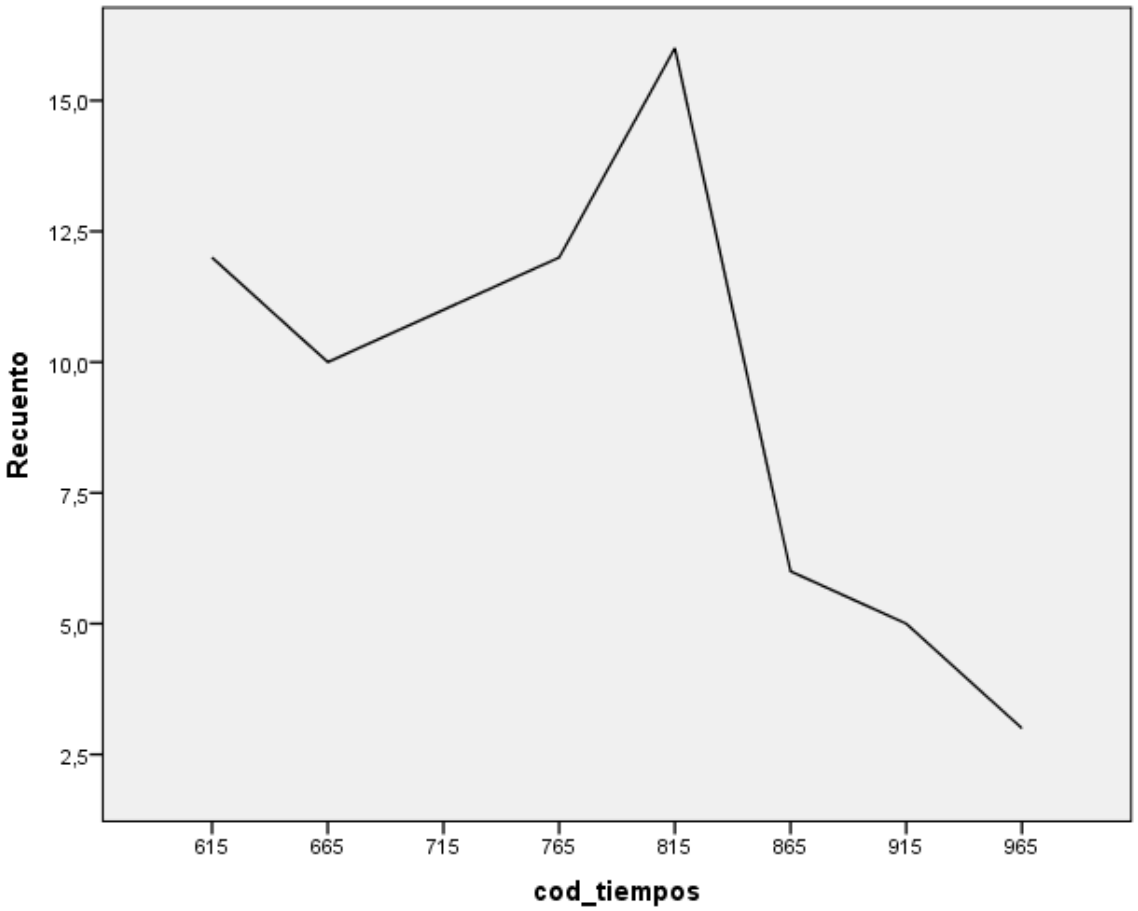




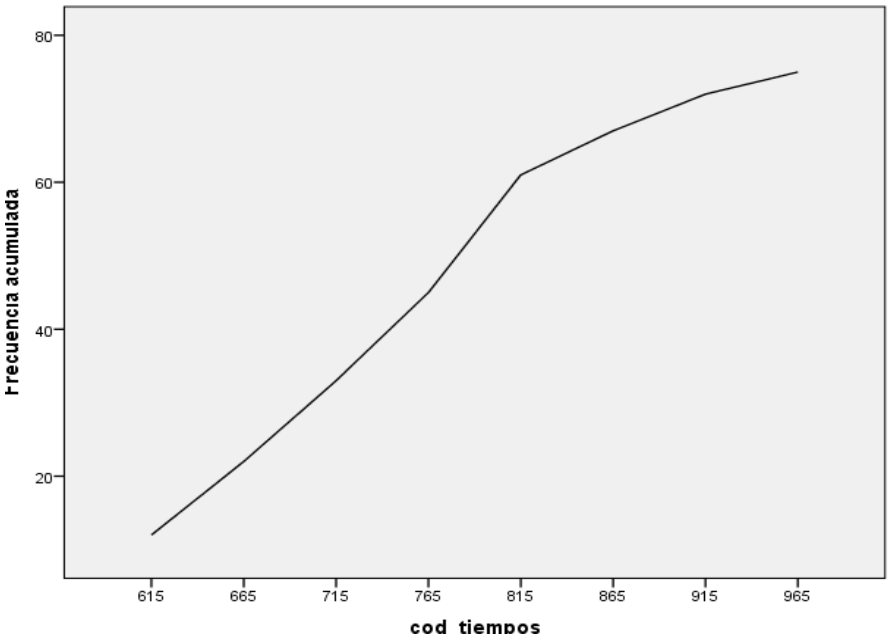
Ahora, para obtener el polígono de frecuencias iremos a “gráficos”->”Cuadros de diálogos antiguos” -> “Líneas”-> “Simples”



Y nos sale el siguiente gráfico:



El anterior era con el número de casos, en el siguiente para obtener el polígono de frecuencias acumuladas realizaremos lo mismo pero eligiendo nº acumulado.



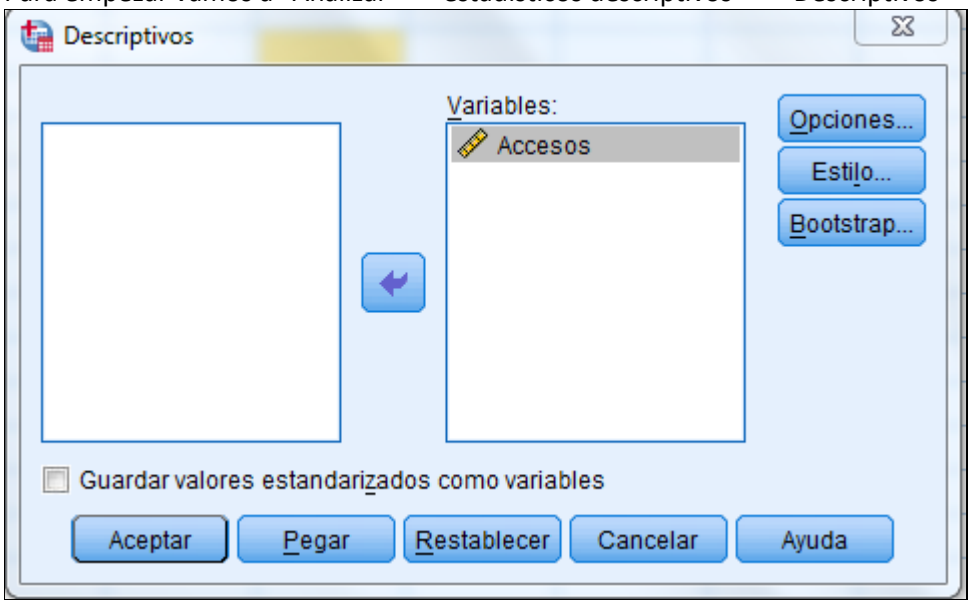
Podemos observar en el histograma que los valores que más se repiten son los que están en el intervalo 790 y 840.

3. Durante un tiempo se registra el número de peticiones por minuto a determinado sitio web. Es necesario analizar las diferencias que se producen en el rendimiento del servidor entre los momentos de mayor carga y los de menor. Para ello se debe realizar un primer análisis sobre los accesos (peticiones http). El fichero datos-pr3-ejer3 con los datos de accesos se puede encontrar en Campus Virtual.

a) Agrupa los datos en intervalos de la misma amplitud y forma la correspondiente tabla de frecuencias. Explica e interpreta los resultados obtenidos.
b) Obtén el polígono de frecuencias. Explica e interpreta los resultados obtenidos.
c) Dibuja dos histogramas de 4 y 8 intervalos y razona cuál de ellos sería el más adecuado para representar los datos. Explica e interpreta los resultados obtenidos.
d) ¿Qué conclusiones generales puedes extraer?

a) Agrupa los datos en intervalos de la misma amplitud y forma la correspondiente tabla de frecuencias. Explica e interpreta los resultados obtenidos.

Para empezar vamos a "Analizar" ->"estadísticos descriptivos" -> "Descriptivos"

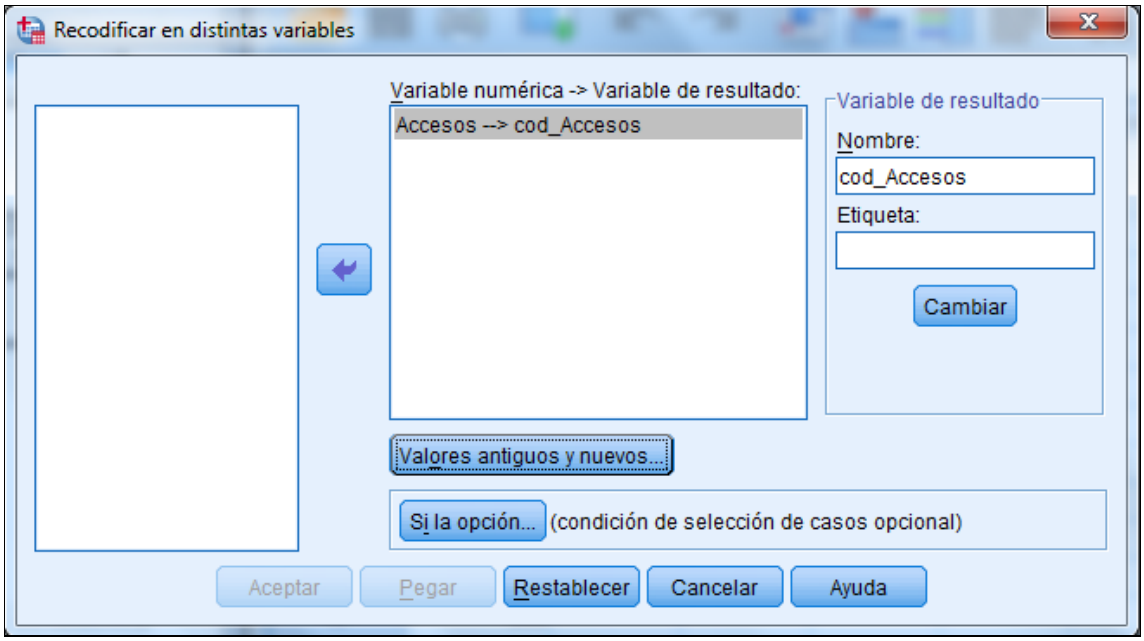


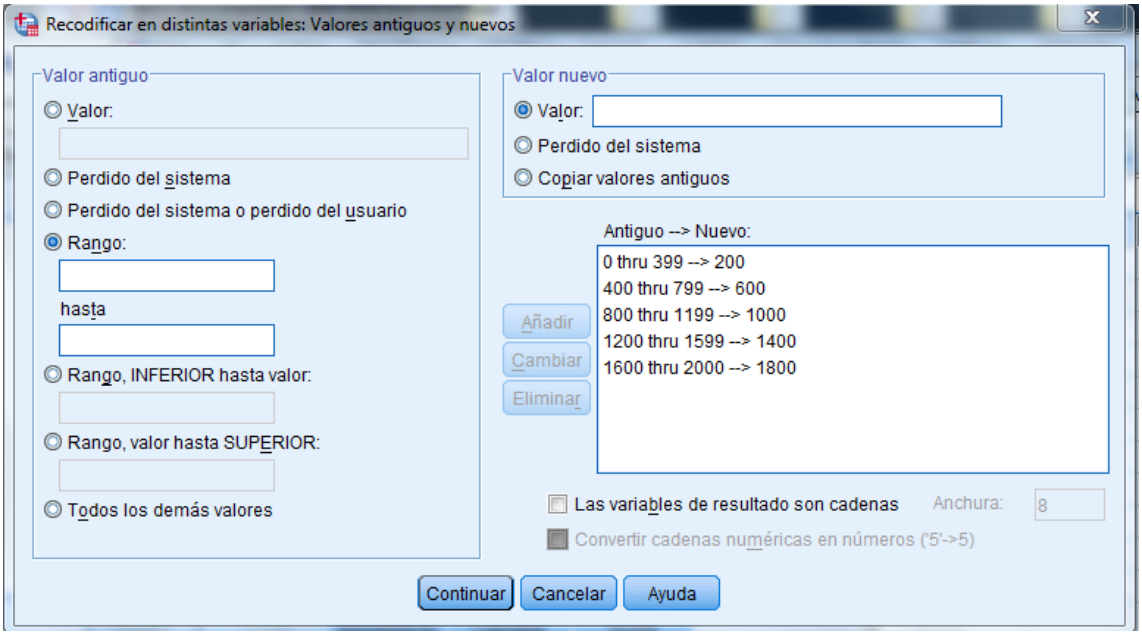
Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desviación estándar
Accesos	3000	0	1987	563,73	411,772
N válido (por lista)	3000				

[0-400[-> valor medio = 200
[400-800[->valor medio = 600
[800-1200[->valor medio = 1000
[1200-1600[->valor medio = 1400
[1600-2000[->valor medio = 1800

Ahora recodificamos en distintas variables:
"Transformar" -> "Recodificar en distintas variables".

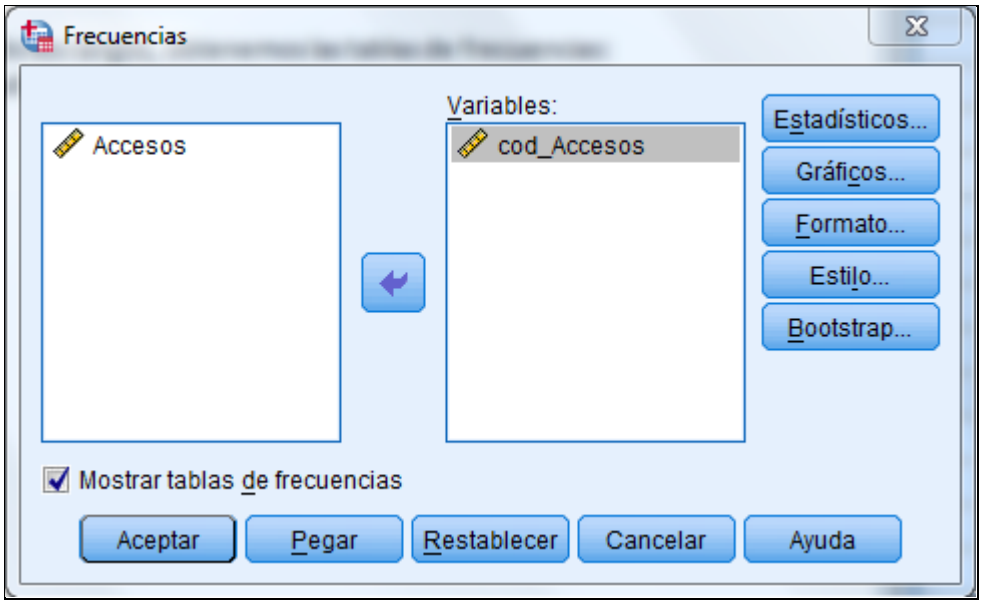




Y quedaría de la siguiente forma:

	Accesos	cod_Accesos
1	755	600
2	15	200
3	58	200
4	9	200
5	760	600
6	296	200
7	21	200
8	491	600
9	526	600
10	560	600
11	273	200
12	28	200
13	335	200
14	194	200
15	380	200
16	77	200
17	1080	1000
18	436	600
19	1267	1400
20	362	200
21	574	600
22	704	600
23	421	600

Una vez establecido los rangos, obtenemos las tablas de frecuencias: "Analizar" -> "Estadísticos Descriptivos"-> "Frecuencias".



cod_Accesos

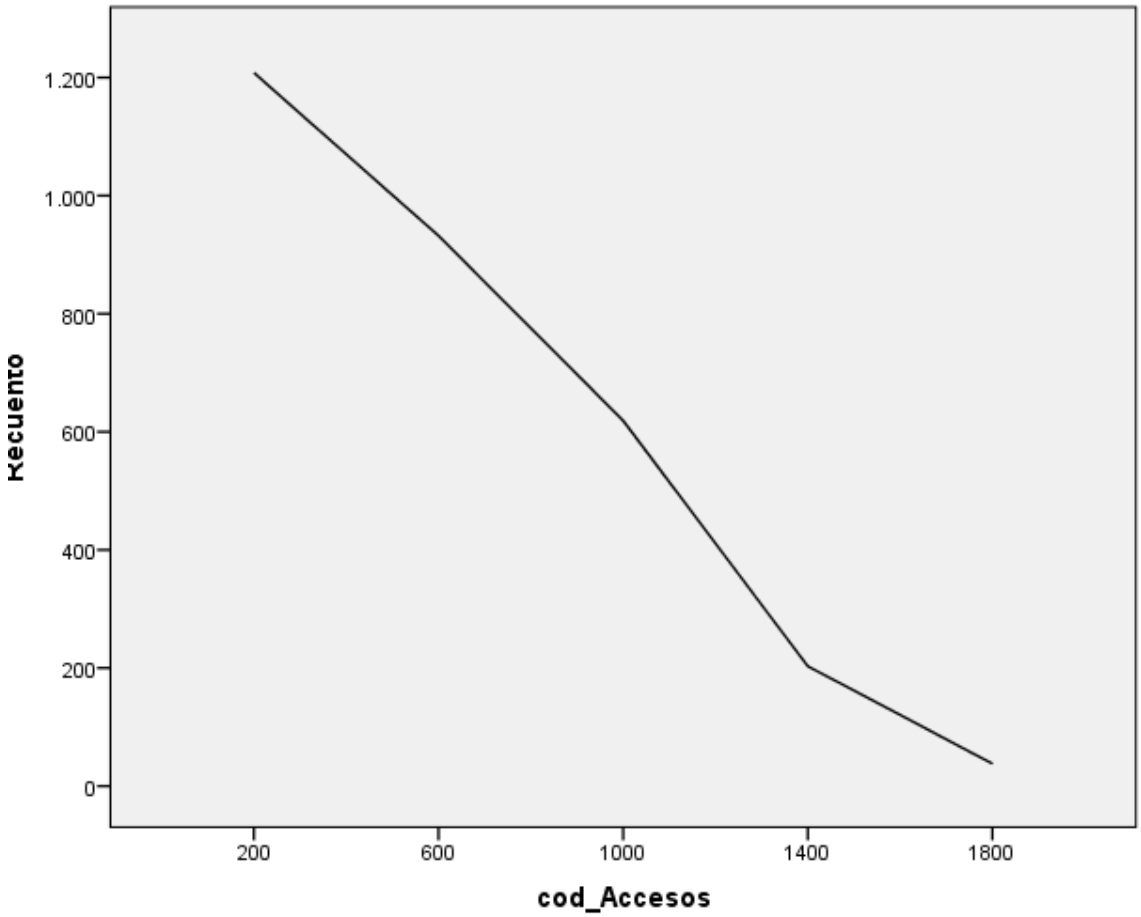
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	200	1208	40,3	40,3	40,3
	600	932	31,1	31,1	71,3
	1000	619	20,6	20,6	92,0
	1400	203	6,8	6,8	98,7
	1800	38	1,3	1,3	100,0
	Total	3000	100,0	100,0	

En esta tabla de frecuencias el valor que más se repite es el 200 con 40,3%. Y los que menos en el intervalo [1600-2000[(entre 1600 y 1999).

b) Obtén el polígono de frecuencias. Explica e interpreta los resultados obtenidos.

Para obtener el polígono de frecuencias iremos a “Gráficos” -> “Cuadros de diálogos antiguos” -> “Líneas” -> “SIMPLES”.

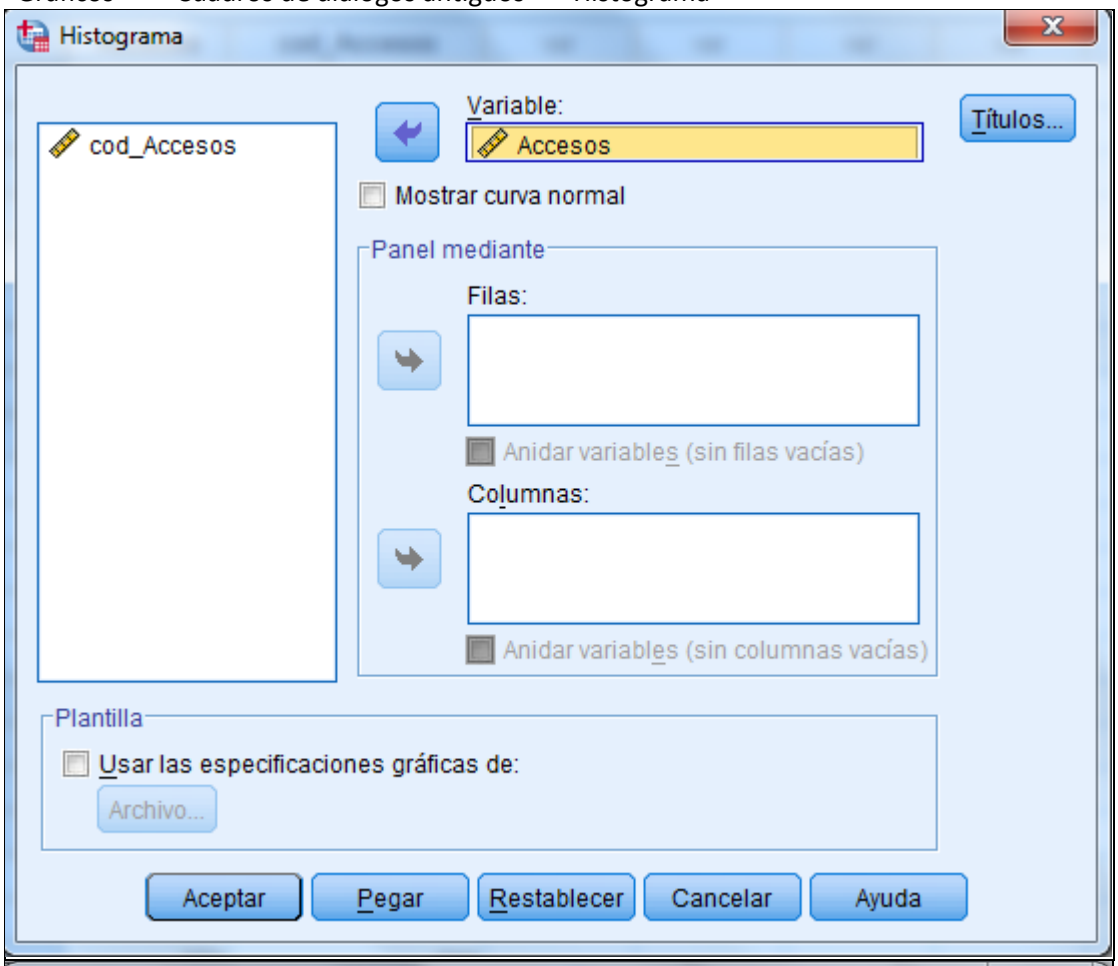
En EN

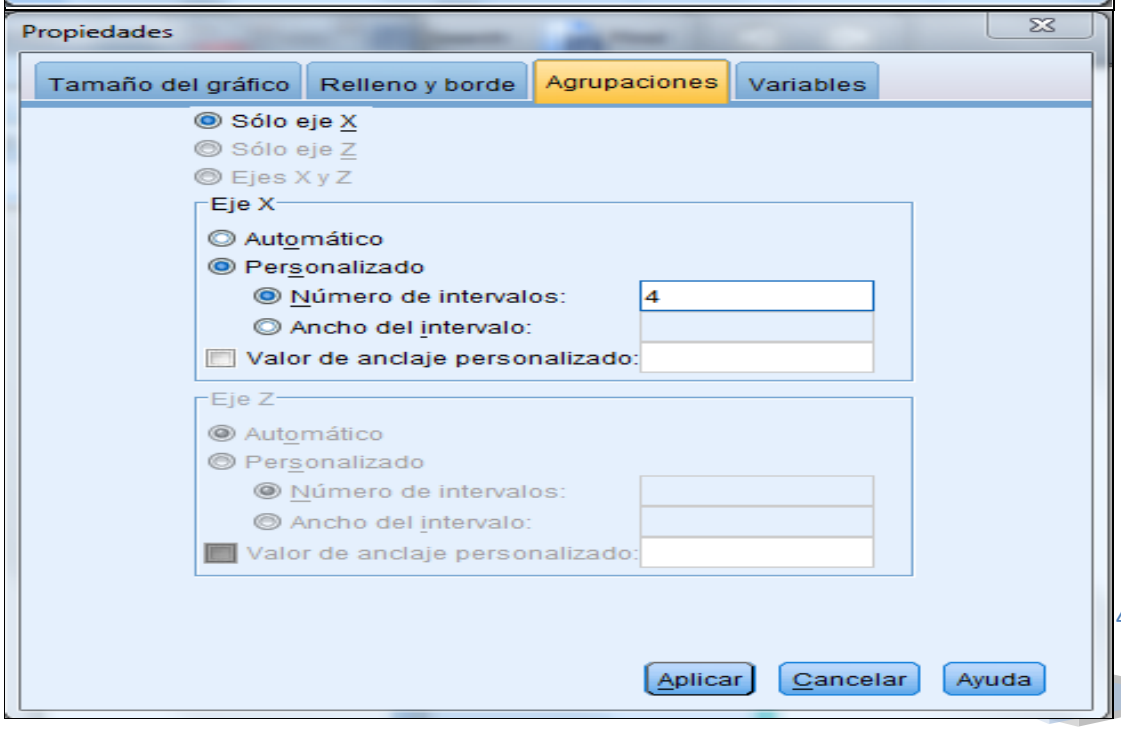


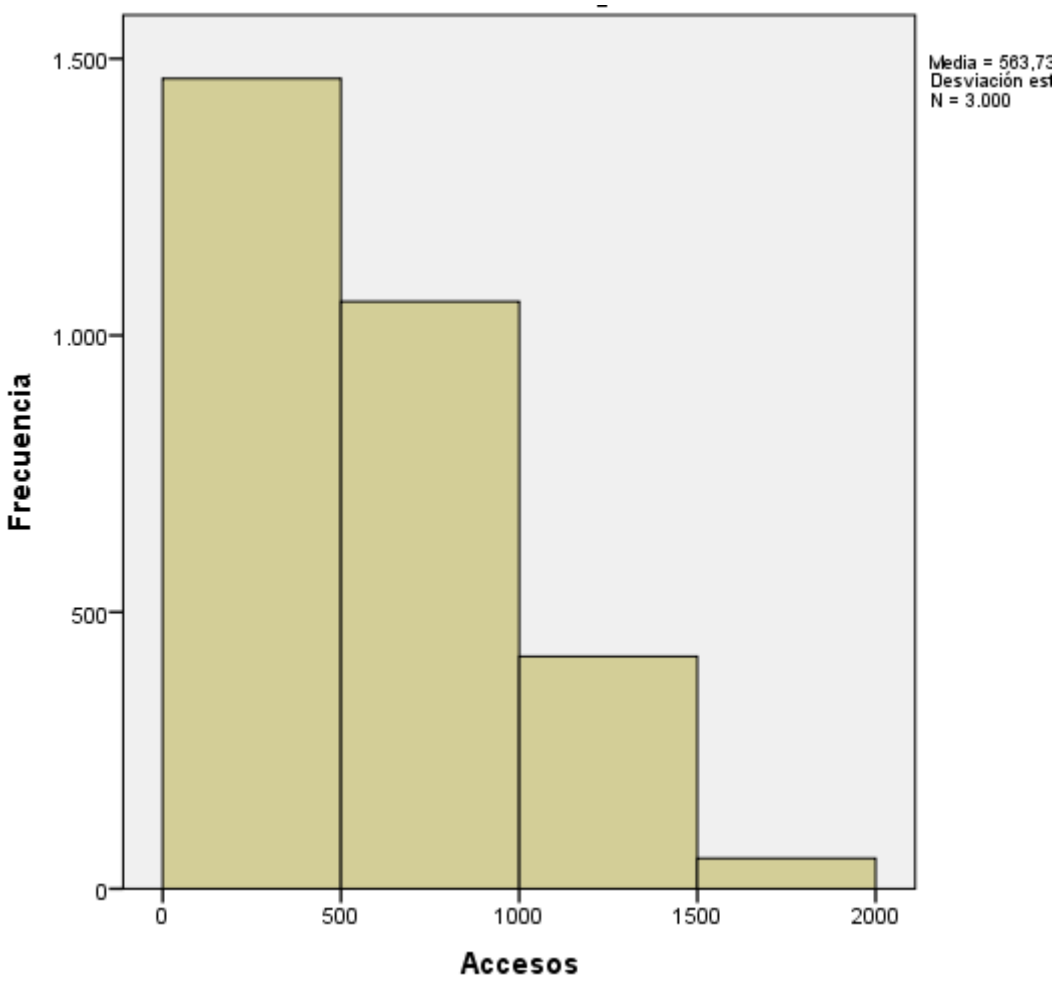
En el gráfico de líneas observamos que el valor 200, en el eje recuento marca 1208, el máximo. Mientras que el valor 1800 su recuento es 38, la más baja. Recuento representa la frecuencia.

c) Dibuja dos histogramas de 4 y 8 intervalos y razona cuál de ellos sería mas adecuado para representar los datos. Explica e interpreta los resultados obtenidos.

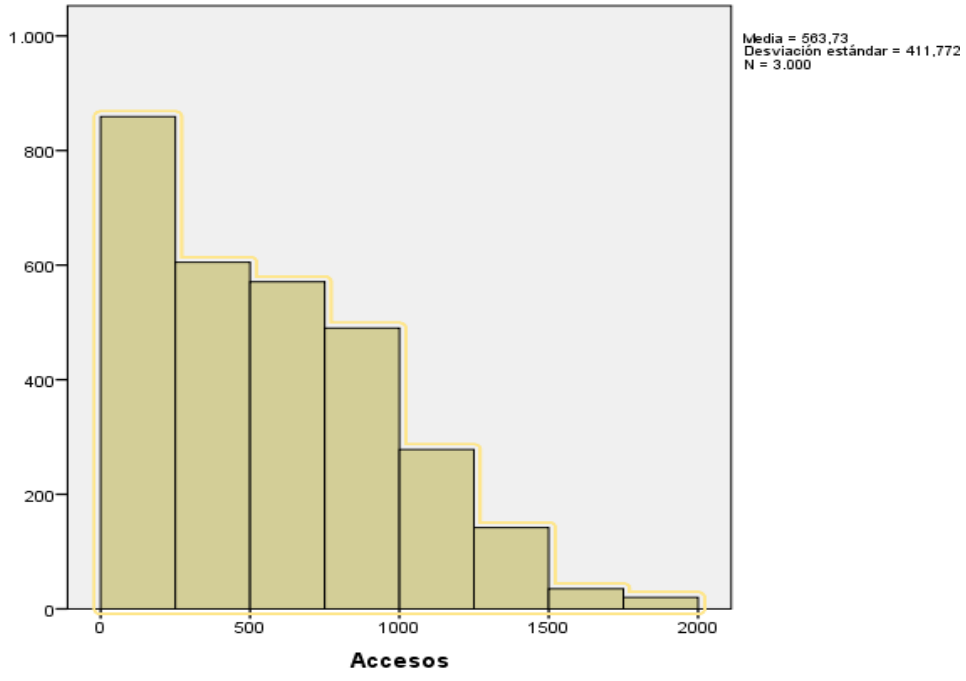
"Gráficos" -> "Cuadros de diálogos antiguos"->"Histograma"







Podemos interpretar que donde más valores hay es en el intervalo 0-500 y donde menos en el 1500-200.
Ahora pondremos el gráfico de 8 para observar las diferencias.



d) ¿Qué conclusiones generales puedes extraer?

Si te fijas en la tabla de frecuencia podemos ver que el valor es 200 es decir los que pertenecen al intervalo [0-400[son los que más se repiten.

4. Se han ejecutado dos programas, 2000 veces cada uno, en un servidor en diferentes momentos y condiciones de carga. En Campus Virtual puedes encontrar el fichero datos-pr3-ejer4 con los tiempos de ejecución (en centésimas de segundo) de ambos programas.

a) Representa cada distribución mediante una tabla de frecuencias y un histograma (indica cuántos intervalos has elegido y los motivos).

Estadísticos descriptivos

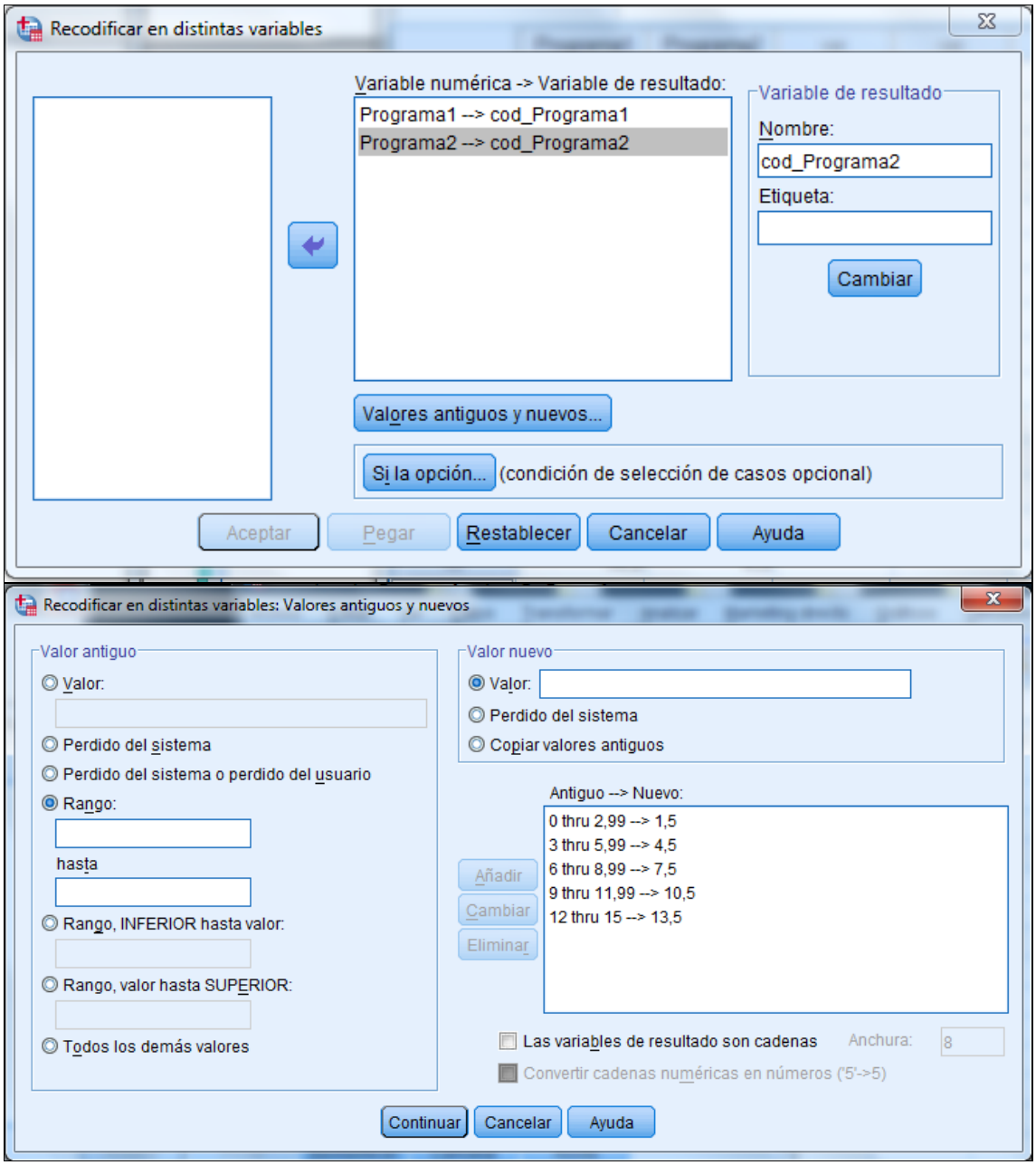
	N	Mínimo	Máximo	Media	Desviación estándar
Programa1	2000	,0	14,8	8,323	3,7773
Programa2	2000	,0	15,0	5,077	3,3654
N válido (por lista)	2000				

Una vez tenemos esto, sacamos los rangos para poder recodificar en distintas variables:

Sabiendo que el mínimo es 0, y el máximo es 15, utilizamos los 5 intervalos de longitud 3.

[0-3[->valor medio = 1,5
[3-6[->valor medio = 4,5
[6-9[->valor medio = 7,5
[9-12[->valor medio = 10,5
[12-15[->valor medio = 13,5

Con esto recodificamos en distintas variables Programa1 en cod_programa1 y Pograma2 en cod_programa2.



Y el resultado sería el siguiente:

	Programa1	Programa2	cod_Programa1	cod_Programa2
1	13,1	2,9	13,5	1,5
2	11,4	7,5	10,5	7,5
3	7,5	,5	7,5	1,5
4	12,7	3,1	13,5	4,5
5	9,0	13,1	10,5	13,5
6	14,3	6,1	13,5	7,5
7	13,8	8,2	13,5	7,5
8	14,3	4,1	13,5	4,5
9	14,1	6,6	13,5	7,5
10	12,6	2,1	13,5	1,5
11	8,2	5,0	7,5	4,5
12	5,0	,4	4,5	1,5
13	6,8	3,2	7,5	4,5
14	12,7	,9	13,5	1,5
15	14,8	4,8	13,5	4,5
16	10,7	3,6	10,5	4,5
17	12,0	3,7	13,5	4,5
18	9,1	3,1	10,5	4,5
19	8,6	5,1	7,5	4,5
20	10,0	,3	10,5	1,5
21	10,2	1,5	10,5	1,5
22	11,1	3,8	10,5	4,5
23	3,3	5,0	4,5	4,5

Tras hacer esto mostramos las tablas de frecuencias de las dos variables recodificadas.

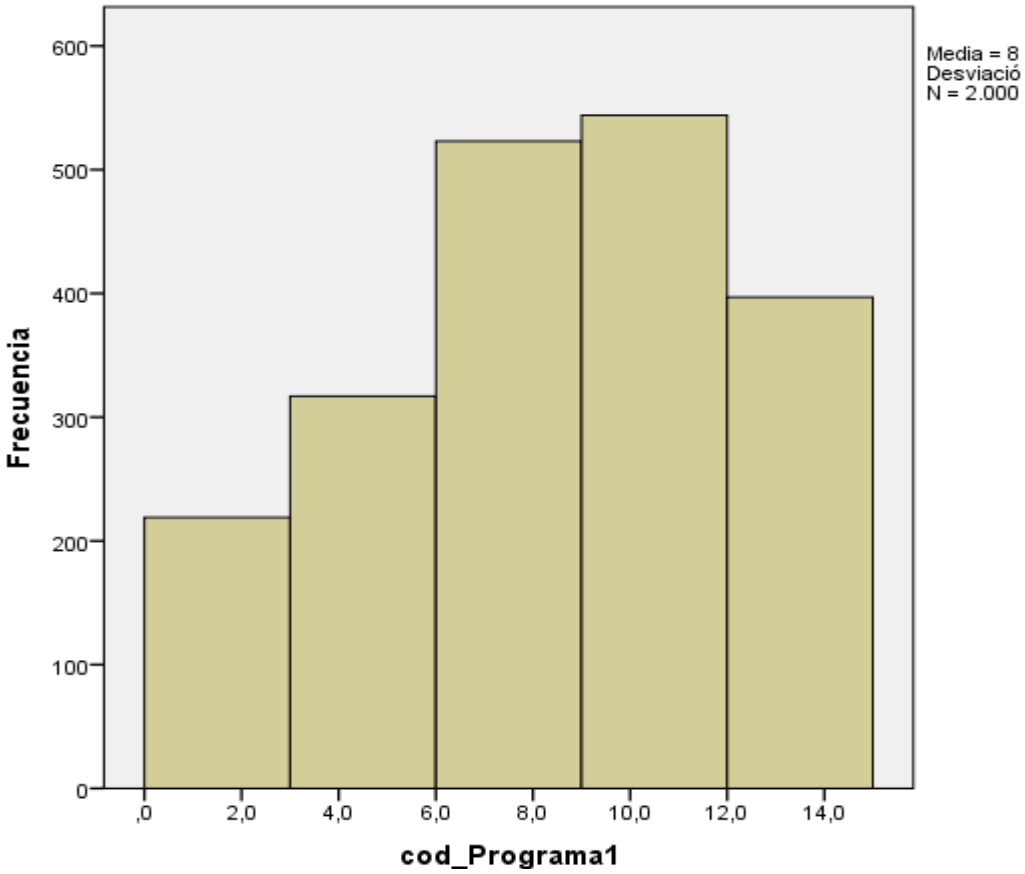
cod_Programa1

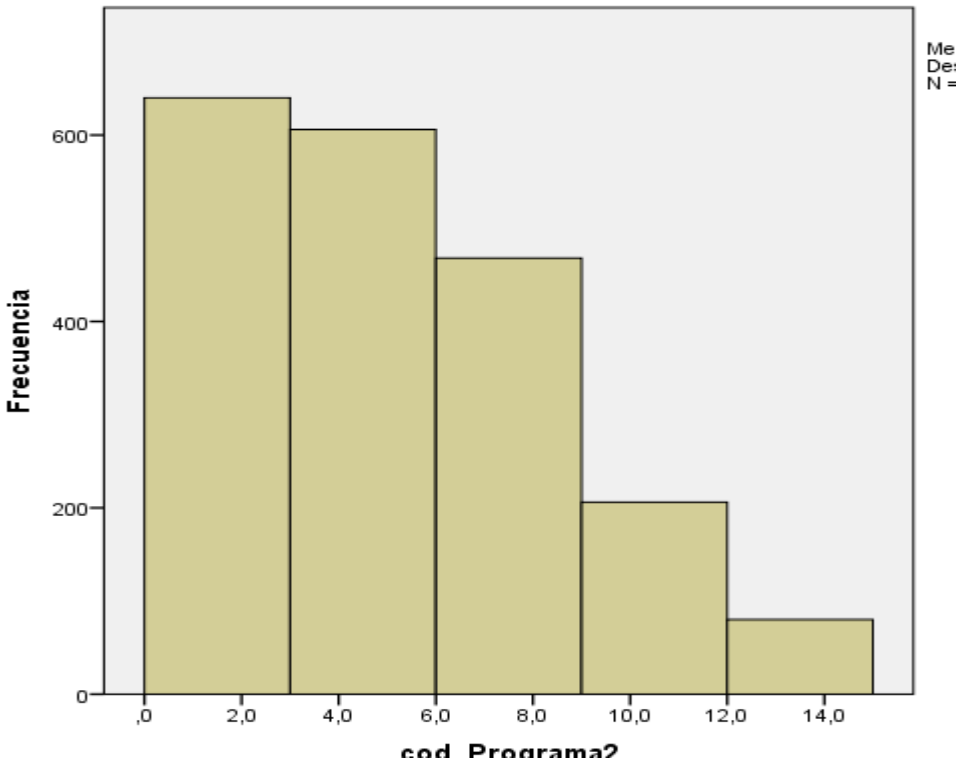
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,5	219	11,0	11,0	11,0
	4,5	317	15,9	15,9	26,8
	7,5	523	26,2	26,2	53,0
	10,5	544	27,2	27,2	80,2
	13,5	397	19,9	19,9	100,0
	Total	2000	100,0	100,0	

cod_Programa2

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1,5	640	32,0	32,0	32,0
	4,5	606	30,3	30,3	62,3
	7,5	468	23,4	23,4	85,7
	10,5	206	10,3	10,3	96,0
	13,5	80	4,0	4,0	100,0
	Total	2000	100,0	100,0	

Ahora obtenemos el histograma, nos vamos a “Gráficos”->”Cuadros de diálogos antiguos” ->”Histograma”.





b) Compara los resultados de ambos programas y coméntalos.

Mediante los histogramas podemos apreciar que el programa 1 el valor que más se repite es el 10,5 cuyo intervalo es [9-12[mientras que en el programa 2 el repetido es el 1,5 es decir el intervalo [0-3[.