**PRÁCTICA 3**

ESTADÍSTICA

**ESTADÍSTICA**

**53247378D**

**JAVIER RIVILLA ARREDONDO**

**Ejercicios**

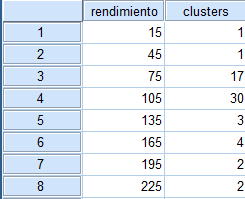
1. El rendimiento –referido a capacidad de procesamiento- de los 60 clusters de los disntos departamentos de una gran empresa es el siguiente, medido en GFLOPS (10^9 operaciones en coma flotante):

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

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

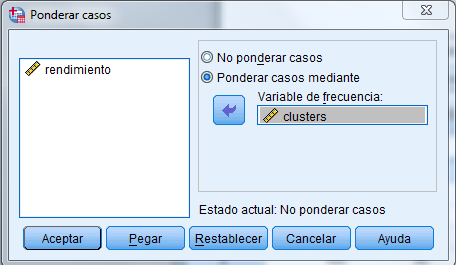
1. Construye la tabla de frecuencias completa

Introducimos los datos en el SPPS:

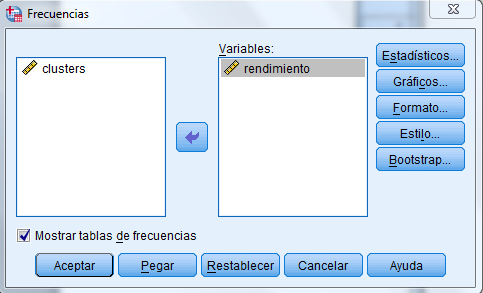


Una vez hemos introducidos los datos en el SPSS lo que haremos será irnos a

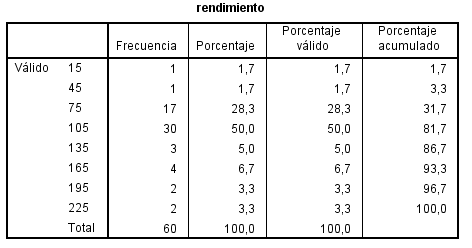
**Datos->Ponderar Casos** y seleccionamos la variable clusters.



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

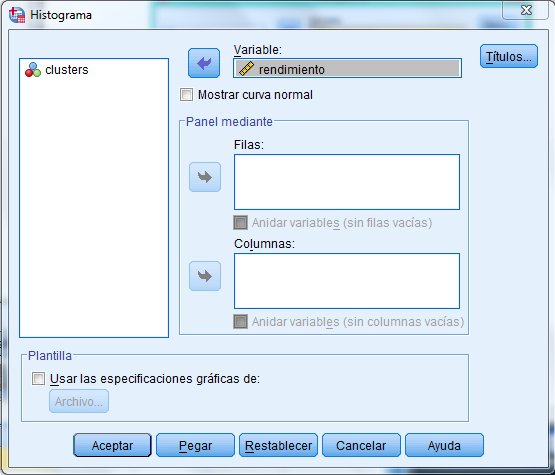


Esto nos mostrará la tabla de frecuencias de la variable rendimiento, que a continuación se mostrará:

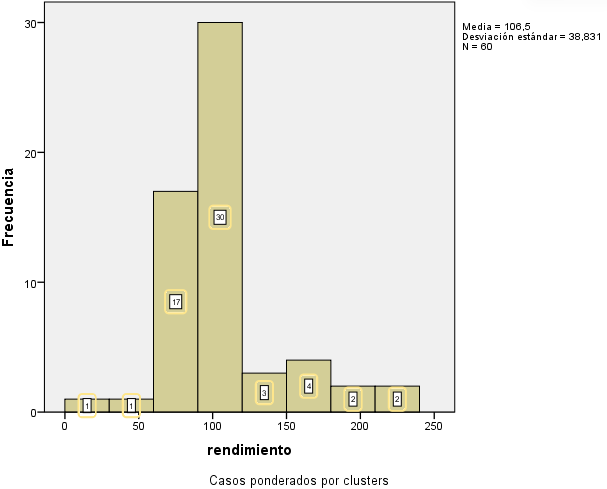


1. Representa el histograma

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



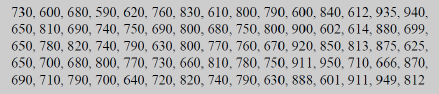
Seleccionamos la variable rendimiento y obtenemos el siguiente histograma:



1. Explica e interpreta los resultados obtenidos en los apartados anteriores

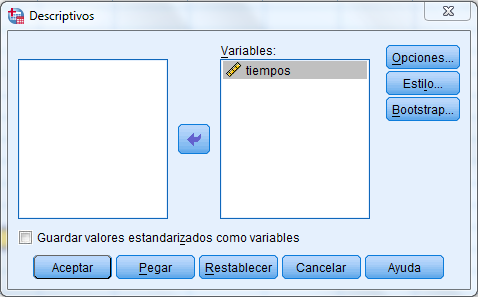
Como podemos observar tanto en la tabla de frecuencias el porcentaje más alto es el 105(de 90 a 120) con un porcentaje del 50%, siendo el que más rendimiento tendrá. Siguiéndole el valor de 75 (de 60 a 90) con un 28,3%. Los que tendrán menos rendimiento serán los valores que estén entre 0-30 y 30-60.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

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

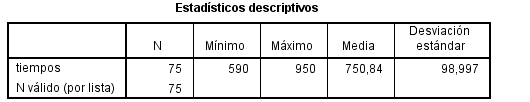


Agrupando los tiempos en intervalos de clase de longitud 50, obtener el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas. Explica e interpreta los resultados obtenidos.

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



Y nos mostrará la siguiente información de los tiempos:



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

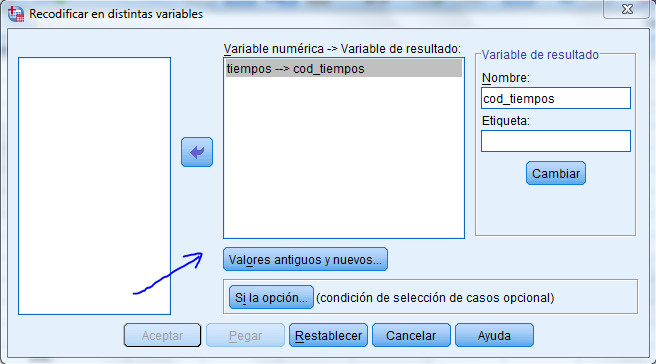
[840-890[ -> valor medio=865

[890-940[ -> valor medio=915

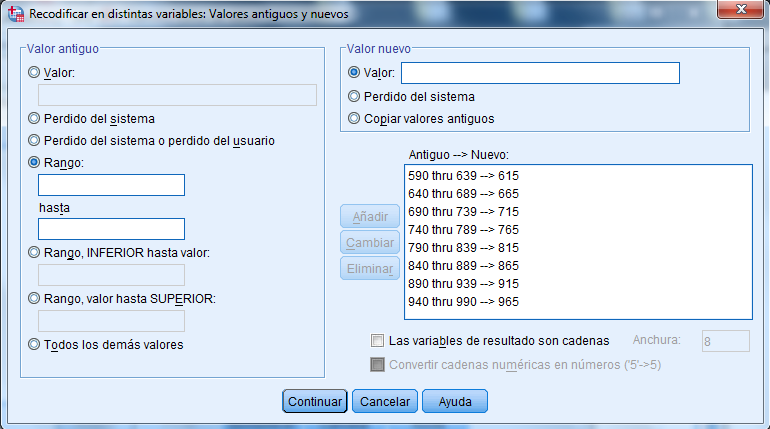
[940-990[ -> valor medio=965

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.

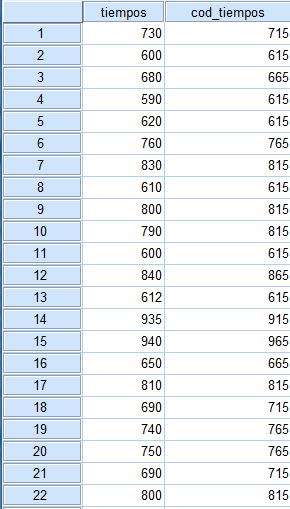
**Transformar->Recodificar en distintas variables**



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

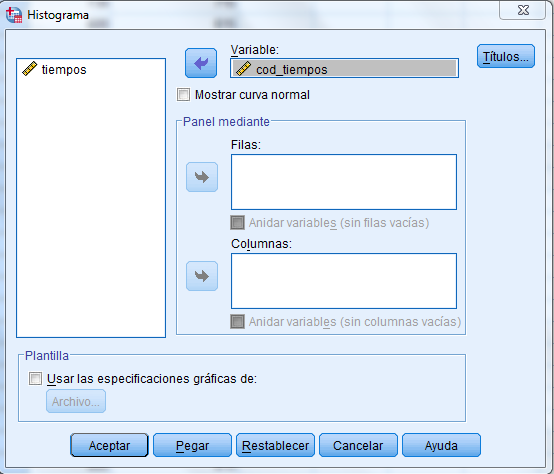


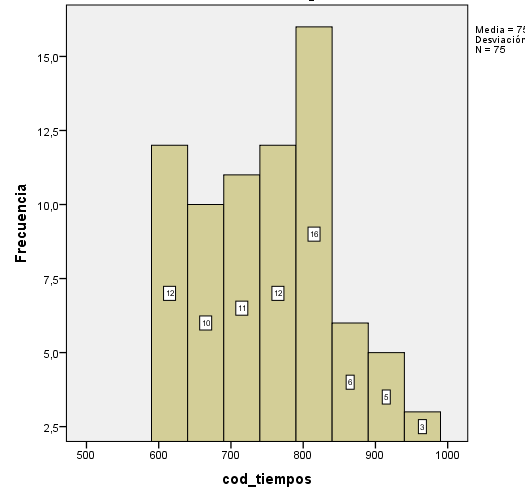
Y nos saldrá la siguiente columna “cod\_tiempos” :



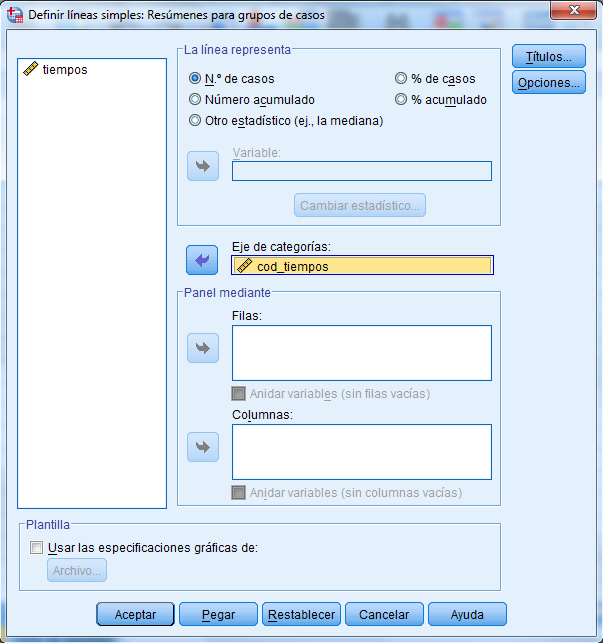
Como podemos ver el valor por ejemplo x donde este pertenece al rango [x-100[ y no al rango [10-x[ por esto, es un intervalo abierto.

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

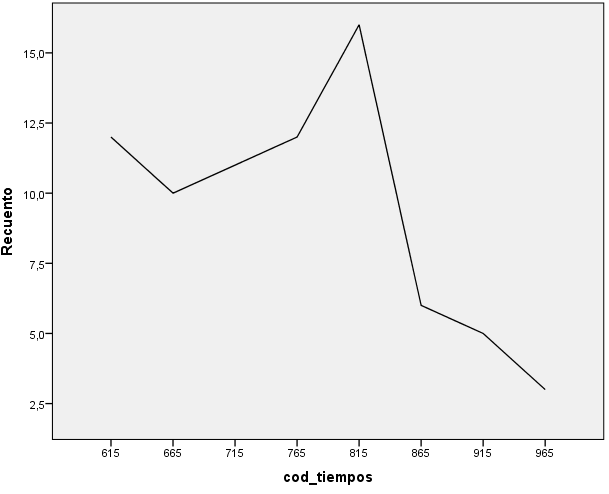




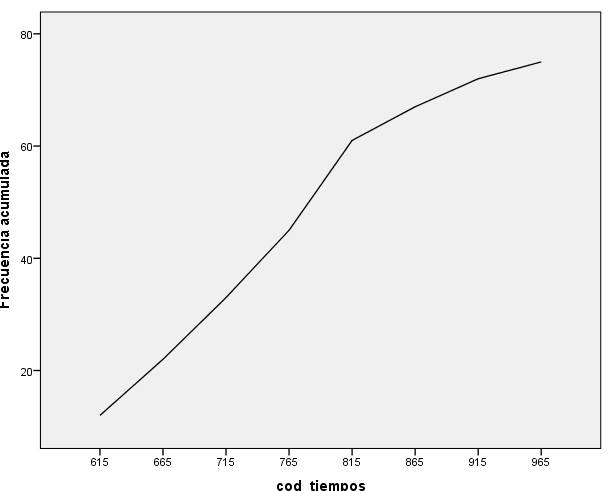
Ahora, para obtener el polígono de frecuencias iremos a “gráficos”->”Cuadros de diálogos antiguos” -> “Líneas”-> “Simples”



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.

1. 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

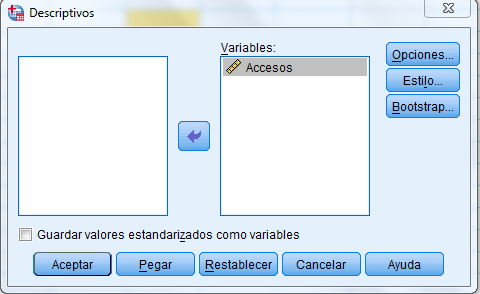
más adecuado para representar los datos. Explica e interpreta los resultados

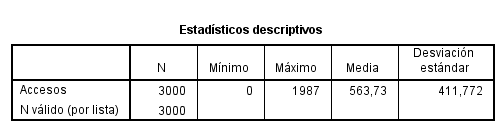
obtenidos.

d) ¿Qué conclusiones generales puedes extraer?

1. 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”





[0-400[-> valor medio = 200

[400-800[->valor medio = 600

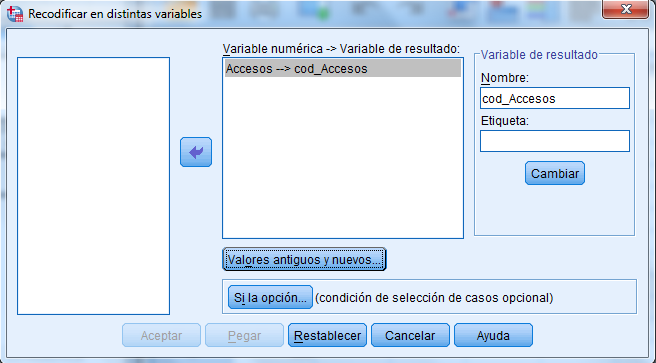
[800-1200[->valor medio = 1000

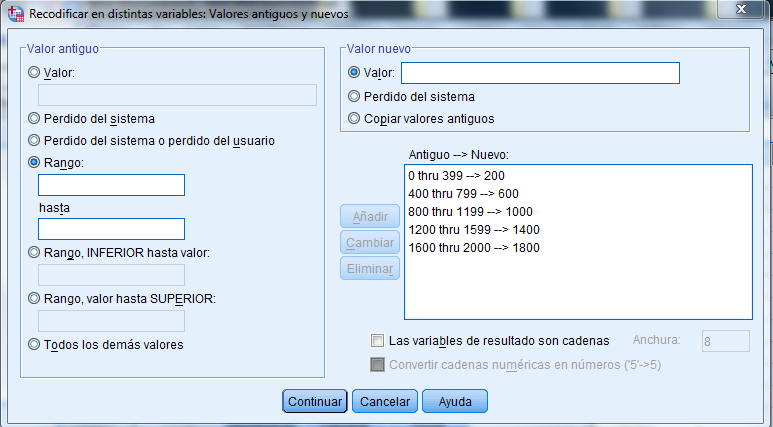
[1200-1600[->valor medio = 1400

[1600-2000[->valor medio = 1800

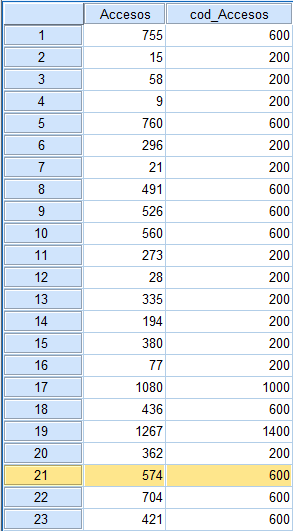
Ahora recodificamos en distintas variables:

“Transformar” -> “Recodificar en distintas variables”.



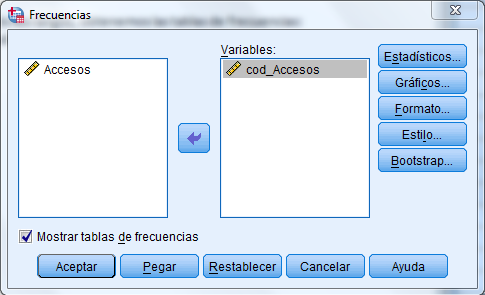


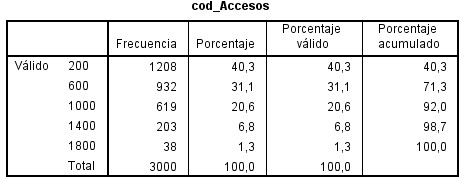
Y quedaría de la siguiente forma:



Una vez establecido los rangos, obtenemos las tablas de frecuencias:

“Analizar” -> “Estadísticos Descriptivos”-> “Frecuencias”.

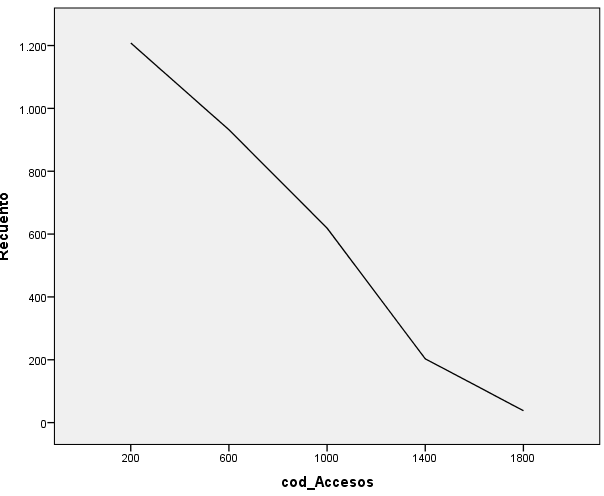




En esta tabla de frecuencias el valor que más se repite es el 200 con 40,3%. Y los que menos en el intervalo [1600-2000[ (entre 1600 y 1999).

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

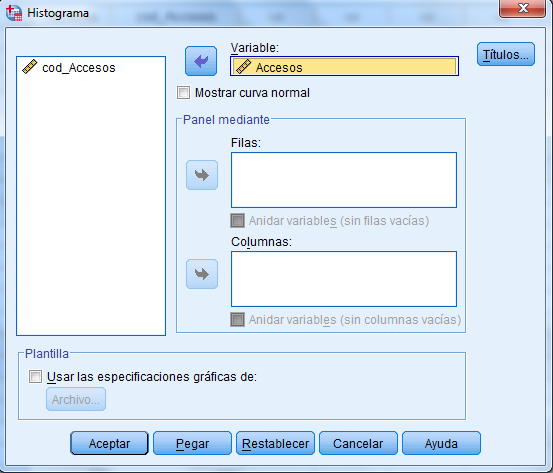
Para obtener el polígono de frecuencias iremos a “Gráficos” -> “Cuadros de diálogos antiguos” -> “Líneas” -> “SIMPLES”.

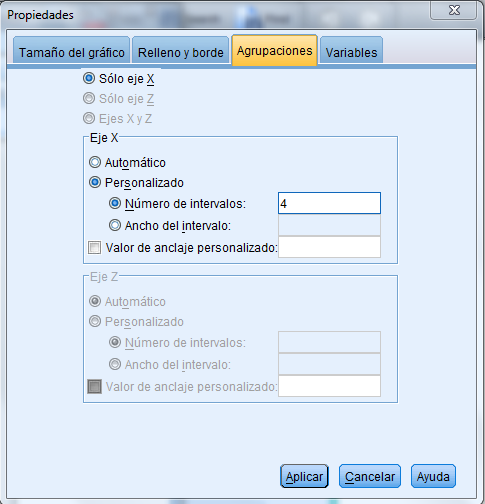
En EN 

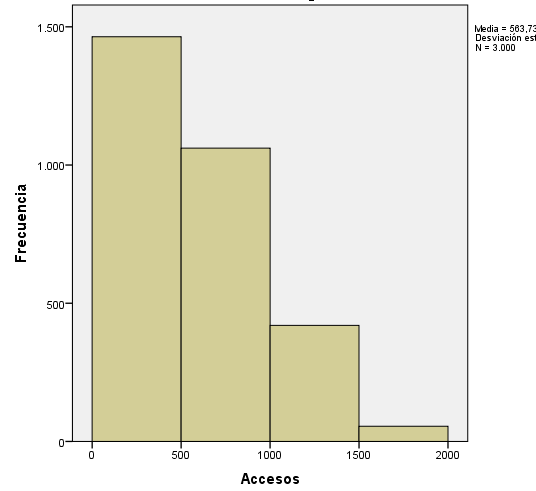
En el gráfico de líneas observamos que el valor 200, en el eje recuento marca 1208, el máximo. Mientras que el valor 1800 su recuento es 38, la más baja. Recuento representa la frecuencia.

1. 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”

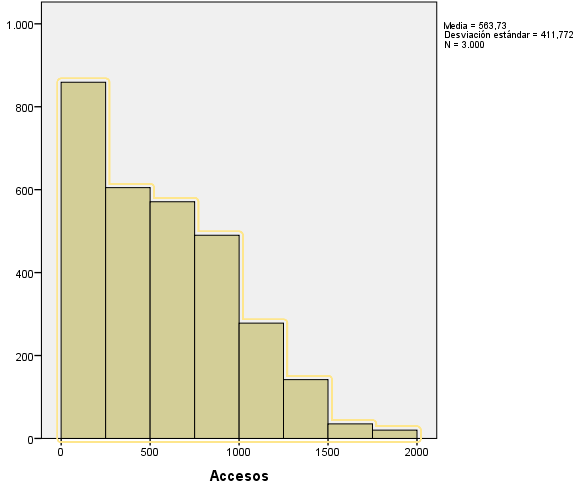






Podemos interpretar que donde más valores hay es en el intervalo 0-500 y donde menos en el 1500-200.

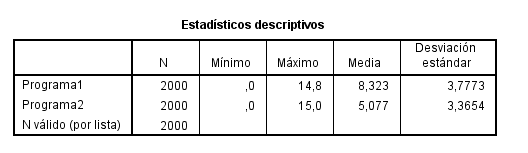
Ahora pondremos el gráfico de 8 para observar las diferencias.



1. ¿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.

1. 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.
2. Representa cada distribución mediante una tabla de frecuencias y un histograma (indica cuántos intervalos has elegido y los motivos).



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

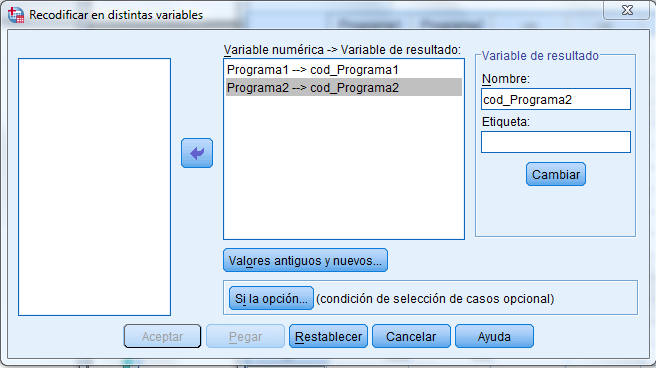
[3-6[->valor medio = 4,5

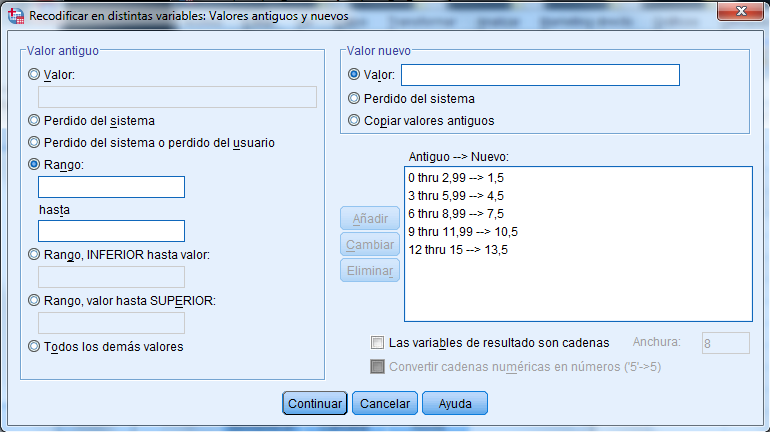
[6-9[->valor medio = 7,5

[9-12[->valor medio = 10,5

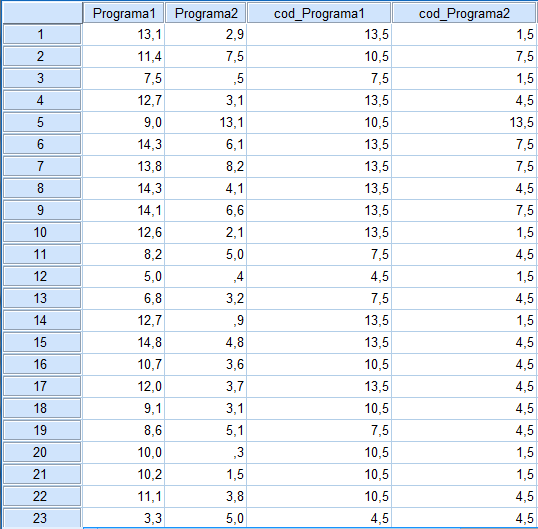
[12-15[->valor medio = 13,5

Con esto recodificamos en distintas variables Programa1 en cod\_programa1 y Pograma2 en cod\_programa2.

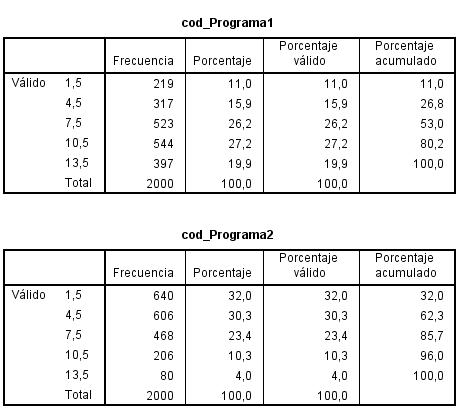




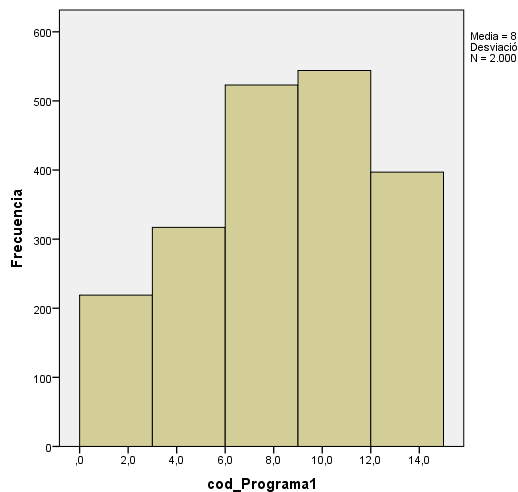
Y el resultado sería el siguiente:

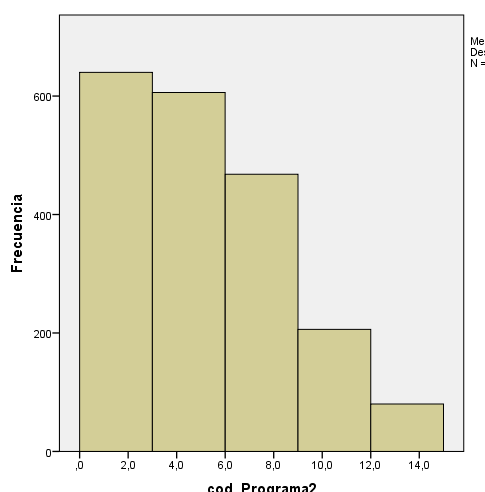


Tras hacer esto mostramos las tablas de frecuencias de las dos variables recodificadas.



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





1. 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[.