

ESTADÍSTICA

Práctica 3

Grupo 6 - Práctica con SPSS.

Alumno: Elvi Mihai Sabau Sabau.

DNI: 51254875L

3.5 Ejercicios



1. Se ha observado la vida útil de 280 componentes electrónicos obteniéndose la siguiente distribución:

Vida en horas	Número de componentes
Menos de 200	6
De 201 a 400	21
De 401 a 600	107
De 601 a 800	78
De 801 a 1000	44
De 1001 a 1200	24

Primero calcularemos el intervalo del enunciado:

$(400 - 200 = 200)$, **200** es el incremento para cada variable.

Es habitual usar el valor medio de cada intervalo como marca de clase, así pues, obtenemos los valores adjuntados en la siguiente captura.

 VEH	 NDC
200	6
400	21
600	107
800	78
1000	44
1200	24

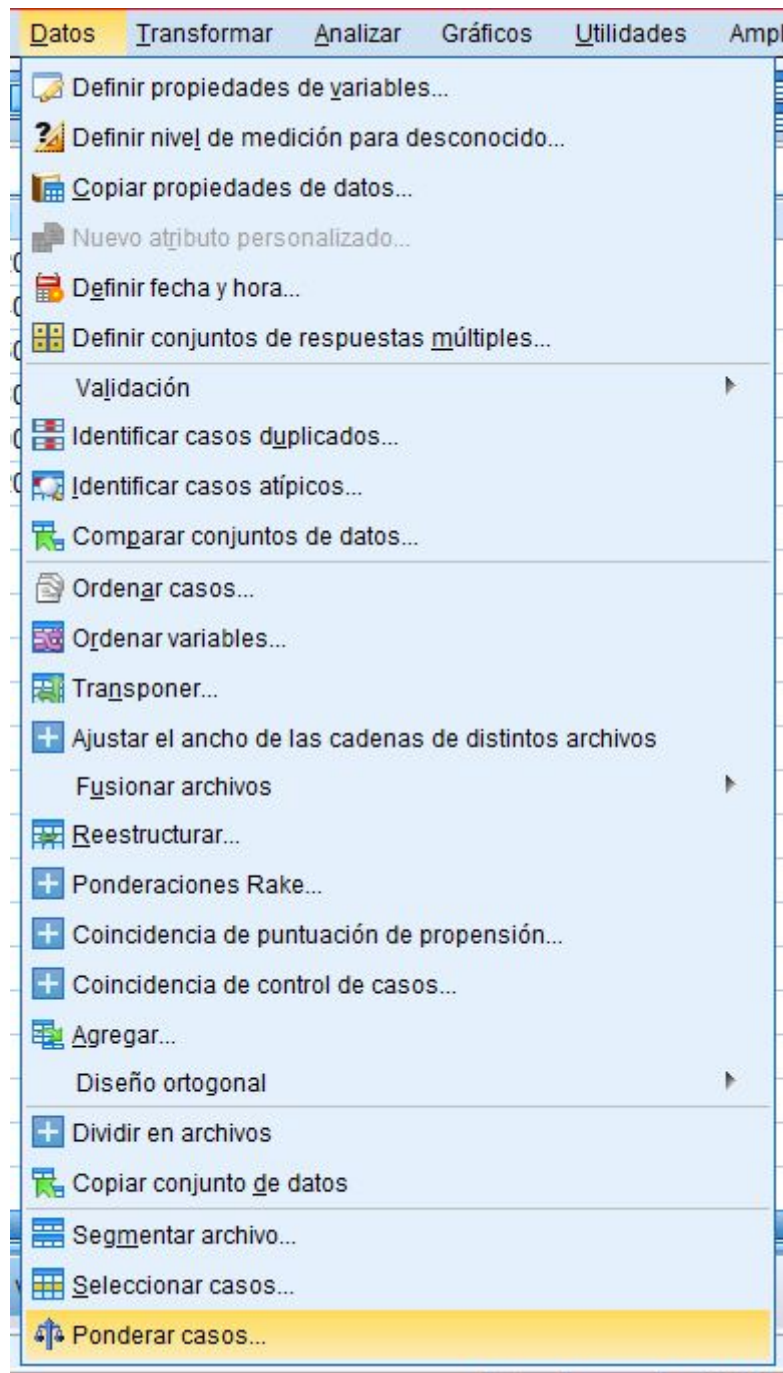
Una vez que ya tenemos los datos correctamente ajustados, ya podemos empezar a realizar el ejercicio.

Para analizar la vida útil de los componentes se pide:

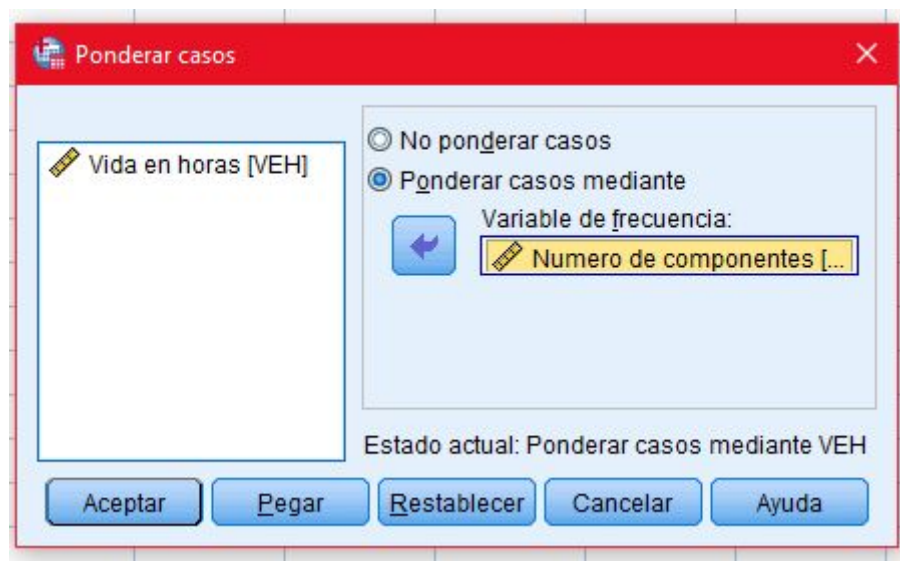
a) Construye la tabla de frecuencias completa.

Para ello, vamos a tener que primero ponderar los datos, para poder indicar que la variable VEH (Vida En Horas) aparece tantas veces como se indica en la variable NDC (Número De Componentes).

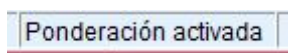
Para hacer esto, vamos al submenú Datos > Ponderar Casos.



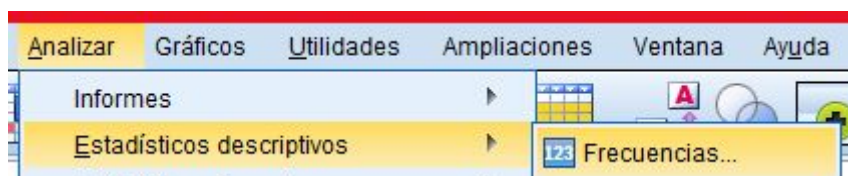
Y tendremos esta ventana:



En la cual especificaremos que queremos ponderar por dicha variable.



Ahora, para realizar la tabla de frecuencias, vamos al submenú
Analizar > Estadísticos descriptivos > Frecuencias.



Y así tendremos la tabla de frecuencias:

➔ **Frecuencias**

Estadísticos

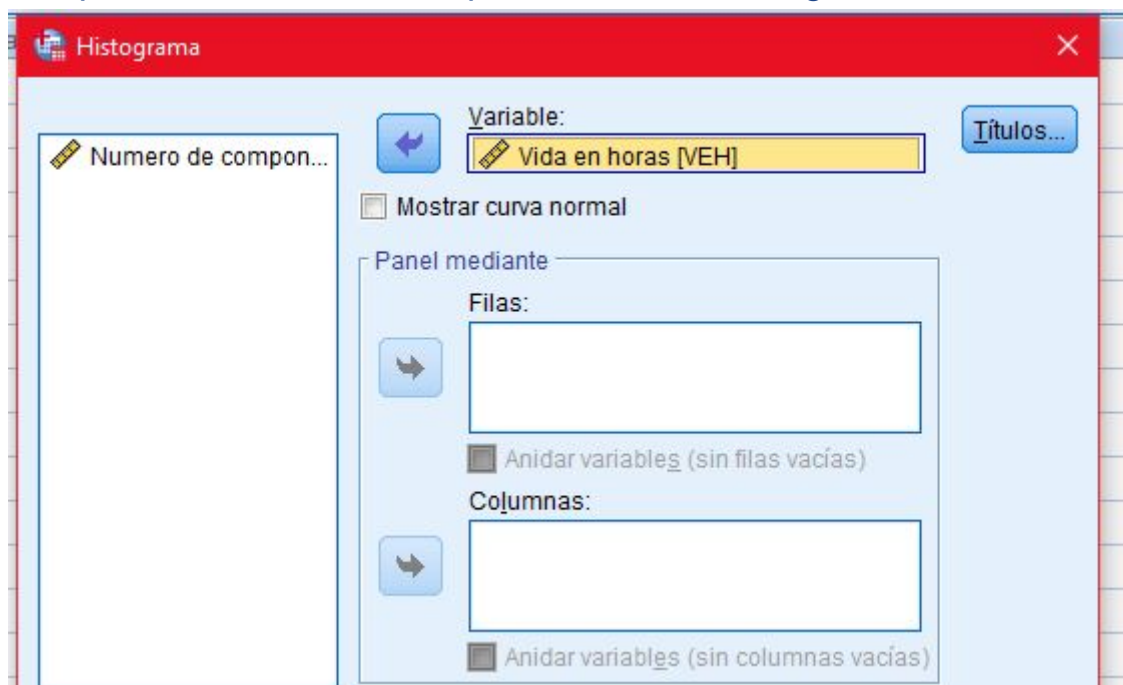
Vida en horas

N	Válido	280
	Perdidos	0

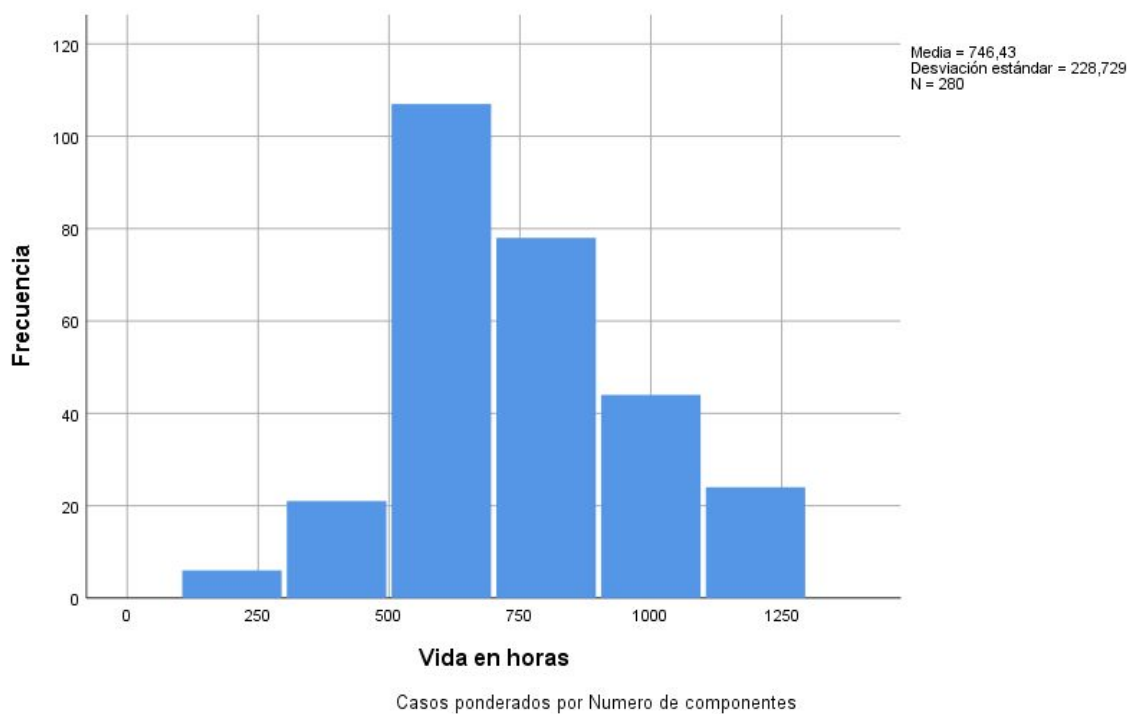
Vida en horas

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	200	6	2,1	2,1	2,1
	400	21	7,5	7,5	9,6
	600	107	38,2	38,2	47,9
	800	78	27,9	27,9	75,7
	1000	44	15,7	15,7	91,4
	1200	24	8,6	8,6	100,0
	Total	280	100,0	100,0	

Y especificamos la variable para realizar el histograma.

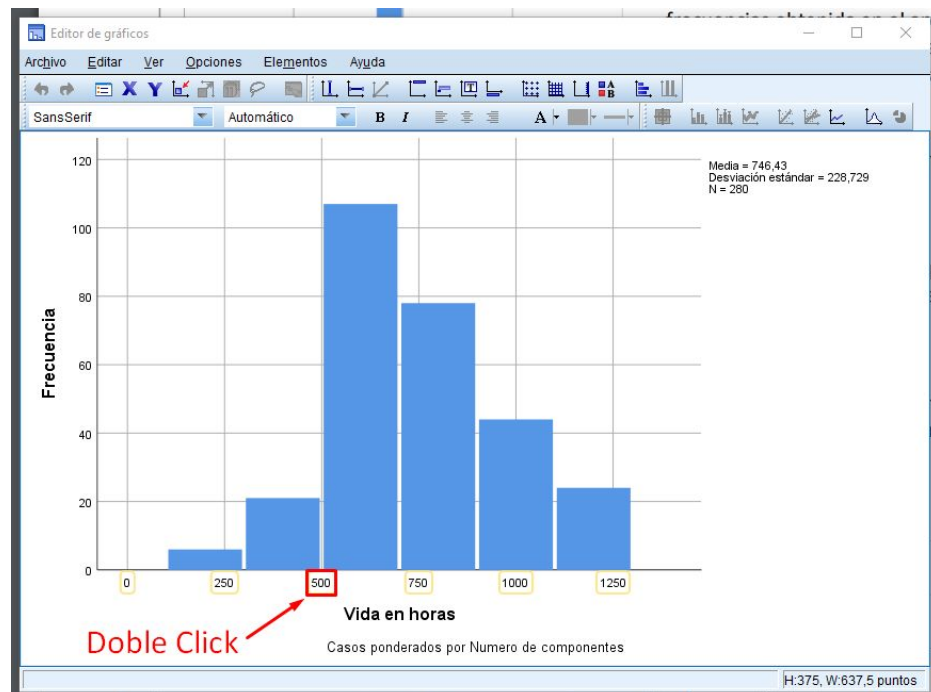


➔ Gráfico



El histograma no está mal, pero se podría mejorar la manera de representar los intervalos.

Para ello, abrimos el editor de gráficos y hacemos doble click en los valores del intervalo para abrir las propiedades, y cambiar dicho intervalo.



Vamos al apartado “Escala” y especificamos el Mínimo, el Máximo, y el incremento.

Propiedades

Formato de numeración Líneas de cuadrícula Variables

Tamaño del gráfico Estilo de texto **Escala** Etiquetas y marcas

Rango

	Automático	Personalizado	Datos
Mínimo	<input type="checkbox"/>	200	100
Máximo	<input type="checkbox"/>	1200	1300
Incremento mayor	<input type="checkbox"/>	200	
Origen	<input checked="" type="checkbox"/>	0	
<input type="checkbox"/> Mostrar línea en origen			

Tipo

Escala

Lineal ☒ Seguro

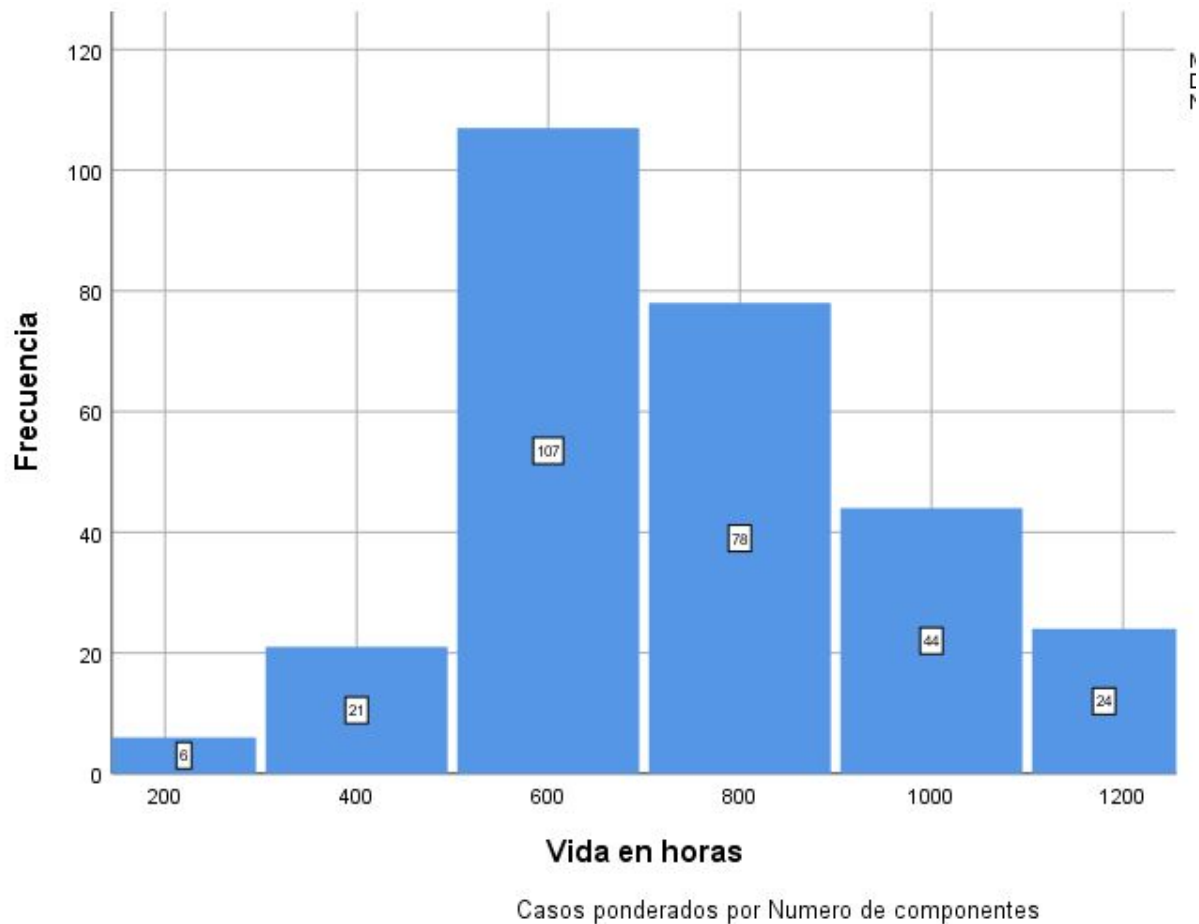
Base: 10

Exponente: 0,5

Margen inferior (%): 5 Margen superior (%): 5

Aplicar Cancelar Ayuda

Ahora sí, nuestro gráfico tiene mejor aspecto, representando los datos correctamente.



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


De los resultados podemos interpretar que la mayoría de componentes tienen una vida de 400 a 600 horas, solo unos pocos duran menos de 200 horas y a su vez, solo unos cuantos duran más de 1200 horas.

2. Se ha medido cien veces el tiempo de ejecución de un programa en un servidor, obteniéndose los siguientes resultados, en segundos, que varían dependiendo de la carga que haya tenido el servidor durante la ejecución:

85, 90, 91, 88, 91, 91, 86, 92, 90, 89 91, 87, 88, 88, 90, 90, 89, 90, 90, 89 91, 87, 90, 84, 91, 88, 90, 88, 88, 88, 92, 90, 90, 90, 93, 90, 89, 92, 91, 92 89, 88, 91, 89, 90, 90, 88, 90, 89, 86 90, 88, 88, 94, 91, 90, 92, 87, 90, 91 92, 88, 92, 92, 88, 89, 88, 91, 89, 91 91, 88, 88, 92, 89, 87, 88, 88, 91, 88 89, 90, 93, 89, 91, 92, 89, 85, 86, 91 89, 87, 88, 88, 93, 90, 95, 89, 92, 89

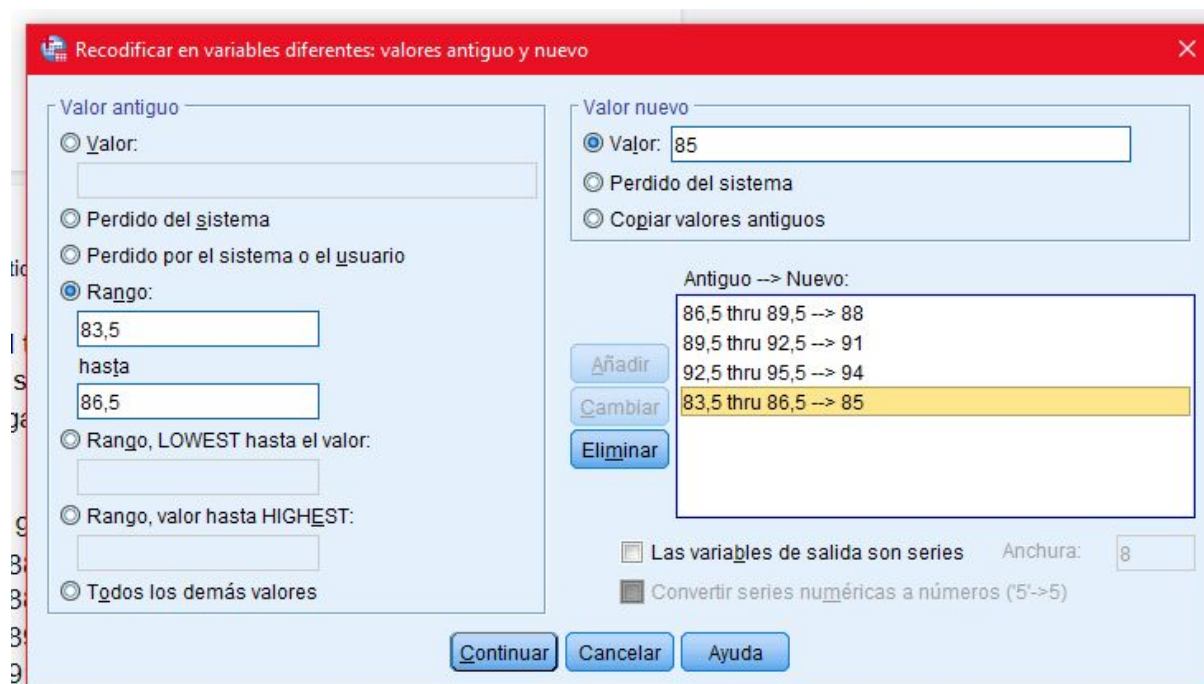
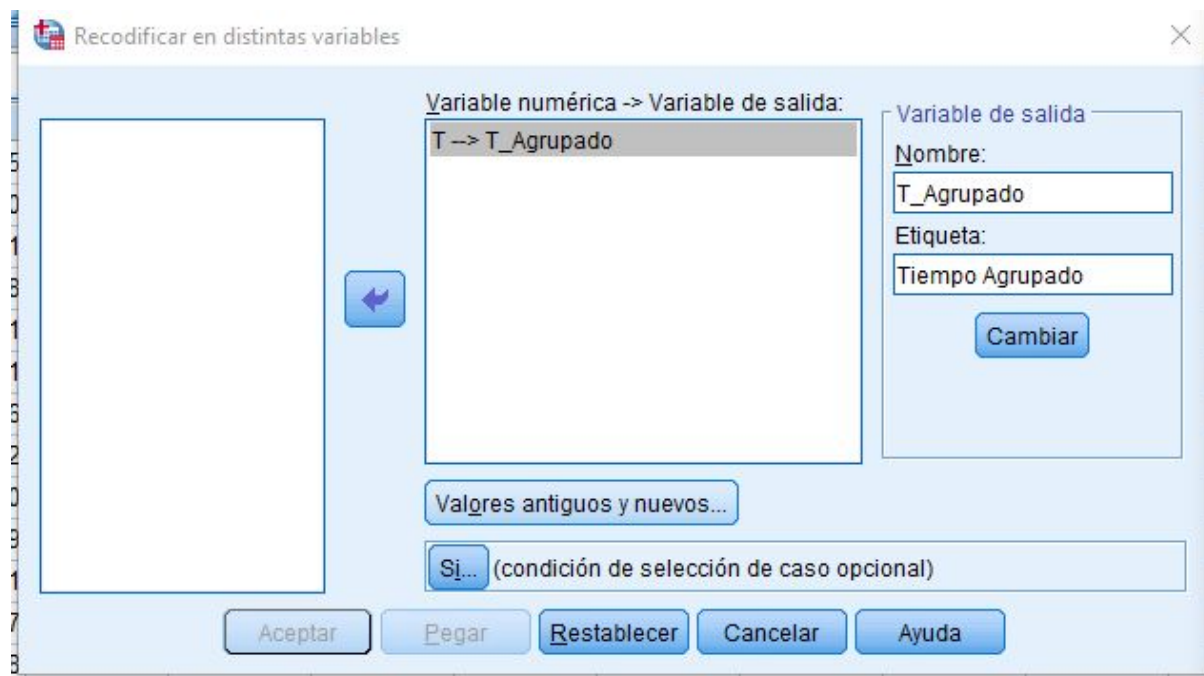
Agrupando los tiempos en los siguientes intervalos de clase 83.5-86.5, 86.5-89.5, 89.5-92.5 y 92.5-95.5, obtén el histograma, el polígono de frecuencias y el polígono de frecuencias acumuladas.

Primero, añadimos los datos:

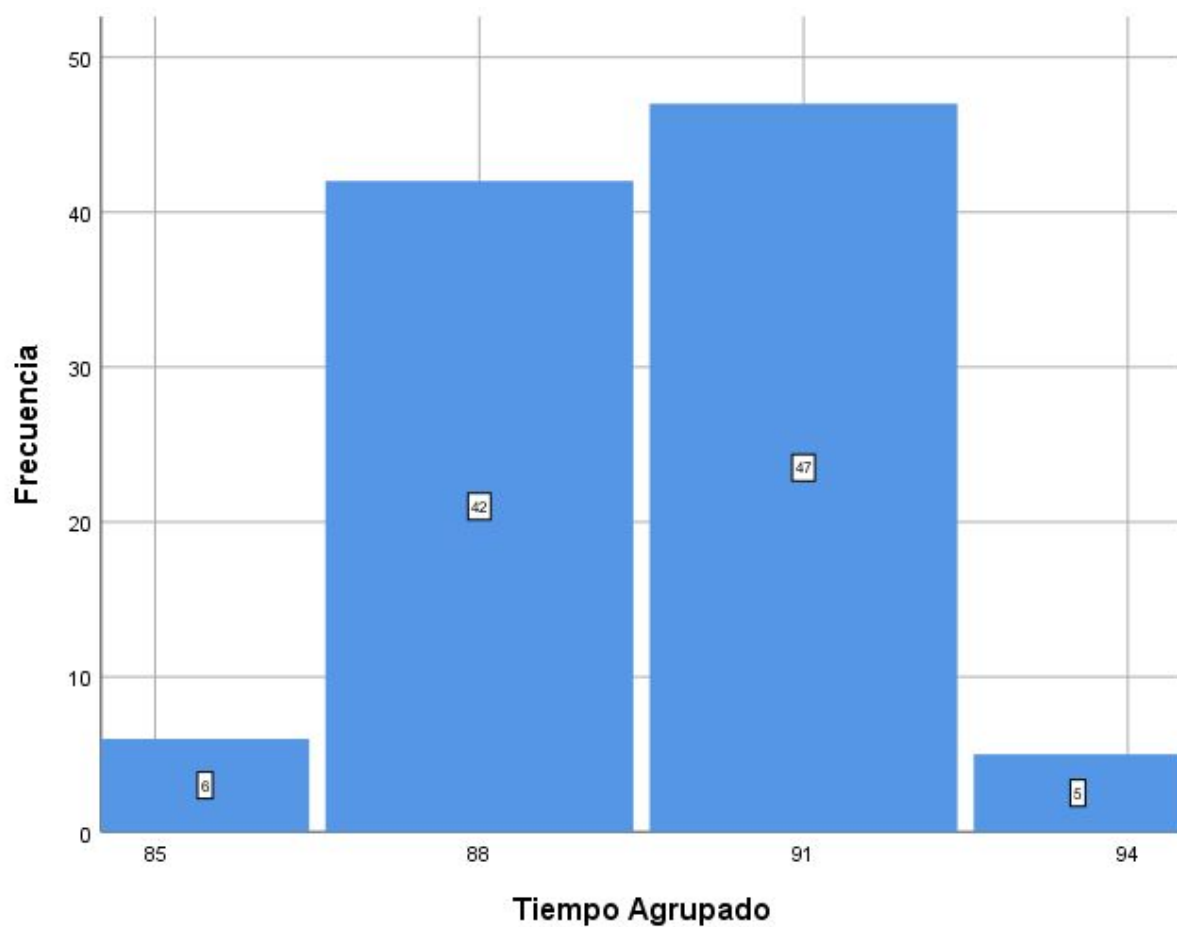
	 T
1	85
2	90
3	91
4	88
5	91
6	91
7	86
8	92
9	90
10	89
11	91

...Y así hasta el 100...

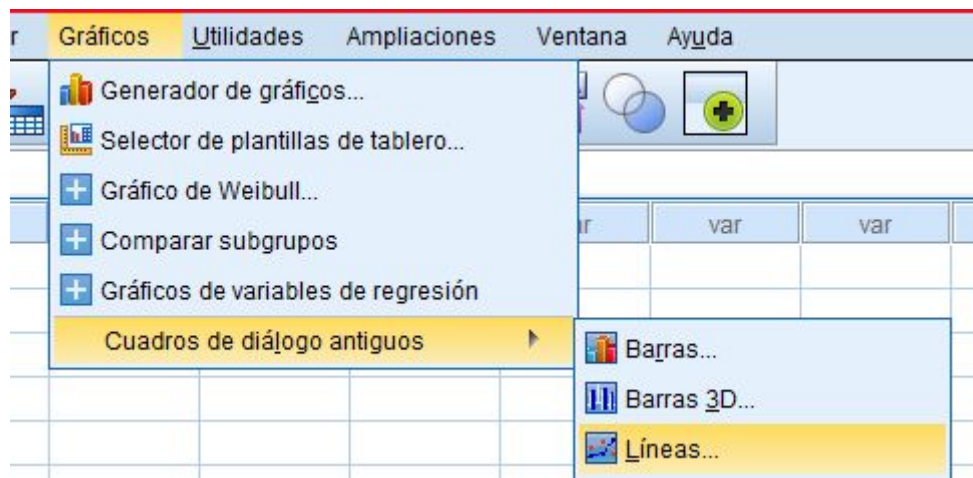
Para ello, para crear los intervalos en SPSS, deberemos agrupar en una nueva variable los intervalos. Para esto, vamos al submenú Transformar > Recodificar en distintas variables, y empezamos a agrupar.



Creamos el histograma como en el ejercicio anterior, y lo ajustamos para que se muestre de manera más correcta.



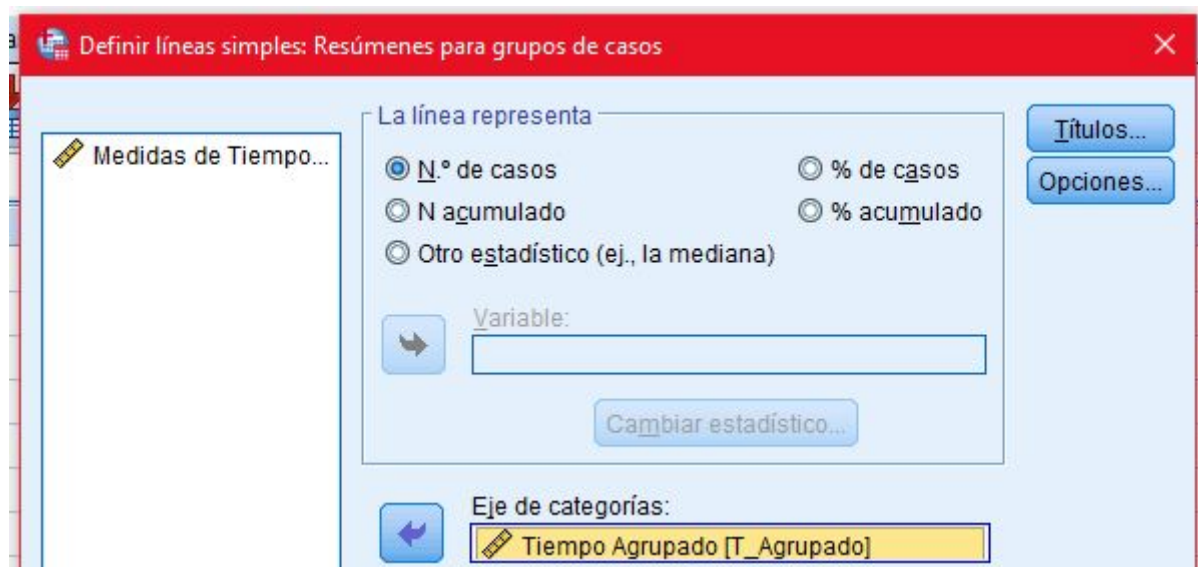
Ahora, para sacar los polígonos de frecuencias y frecuencias acumuladas, iremos al submenú Graficos > Cuadro de dialogos antiguos > Lineas.



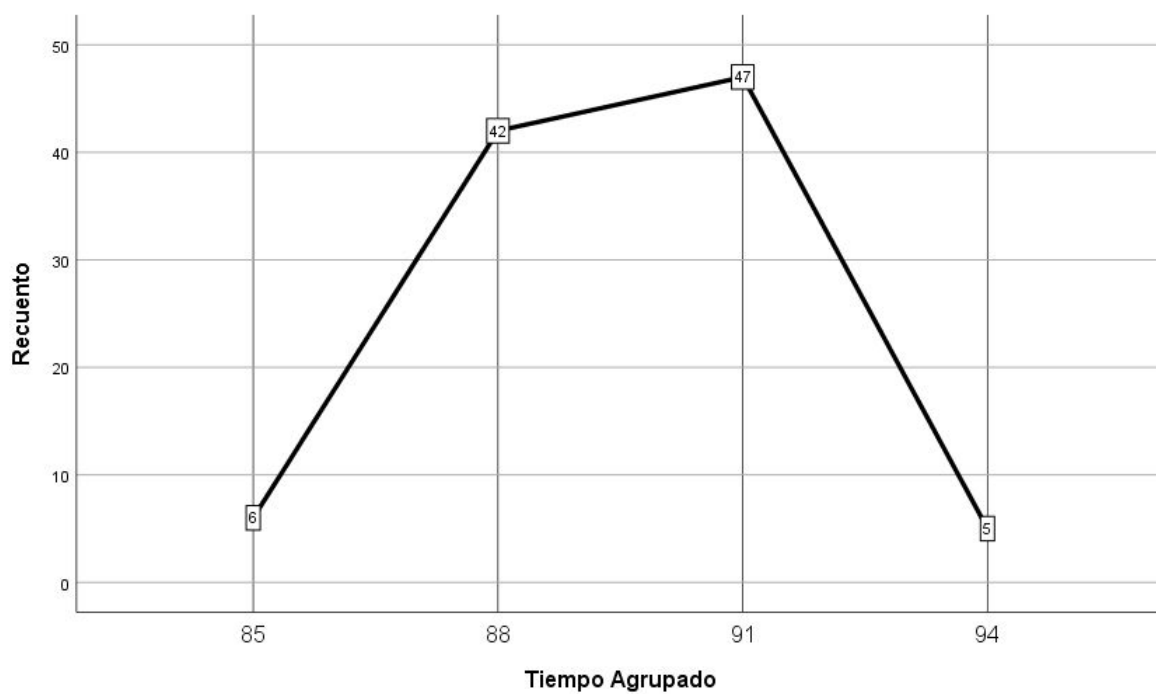
Y lo dejamos en resumen para el grupo de casos.



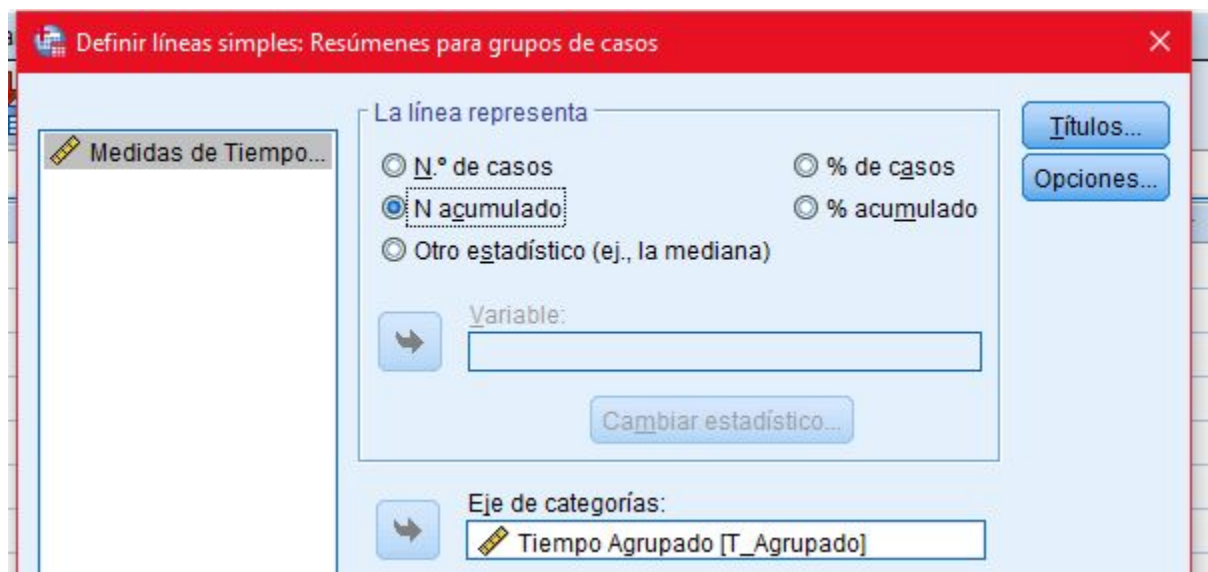
Especificamos nuevamente que lo que queremos es el N° de casos, y especificamos la variable en el eje de categorías.



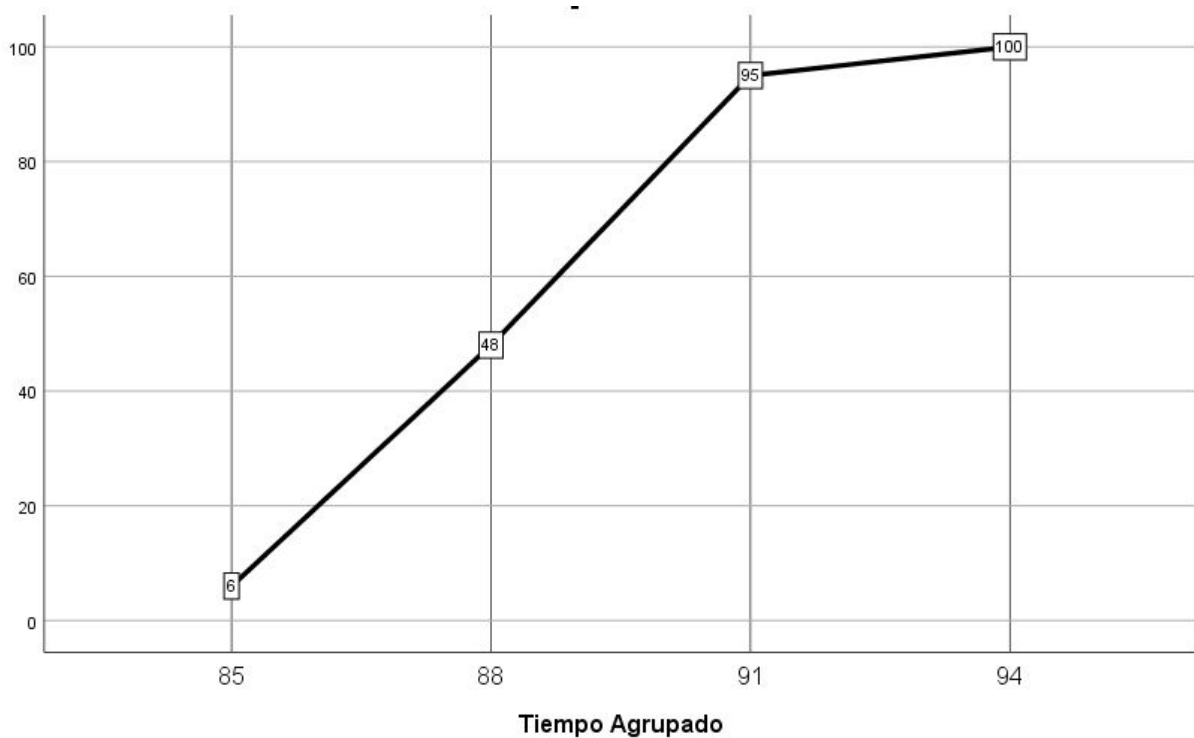
Y ya lo tenemos:



Ahora, para el gráfico de frecuencias acumuladas, haremos lo mismo, pero en vez de especificar N° de casos, especificamos N acumulado.



Y este sería el resultado.



Explica e interpreta los resultados obtenidos.

Lo que podemos deducir de los resultados es que el programa tarda de media entre 89 y 90s en responder.

3. Las dos tablas siguientes muestran datos sobre peticiones http diarias sobre un servidor web. La primera tabla representa las peticiones de URLs internas y la segunda las externas.

Peticiones HTTP de URLs internas						
345	634	456	32	666	9	671
754	399	621	43	333	71	371
234	11	887	448	452	875	121
345	353	789	594	22	943	30

Peticiones HTTP de URLs externas						
324	39	519	21	984	720	921
452	52	410	61	197	317	173
852	85	297	621	662	222	81
624	712	271	424	49	91	73

a) Utilizar alguna de las técnicas presentadas en esta práctica para estudiar ambas distribuciones. Explica e interpreta los resultados obtenidos.

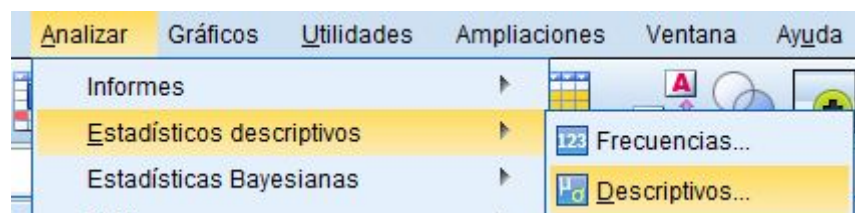
Lo primero que deberemos hacer es sacar los descriptivos, para poder saber el rango, el máximo y el mínimo, después, realizar los intervalos, analizar el resultado mediante una herramienta visual por ejemplo un histograma, y después sacar conclusiones.

Primero, para ello, añadimos los datos.

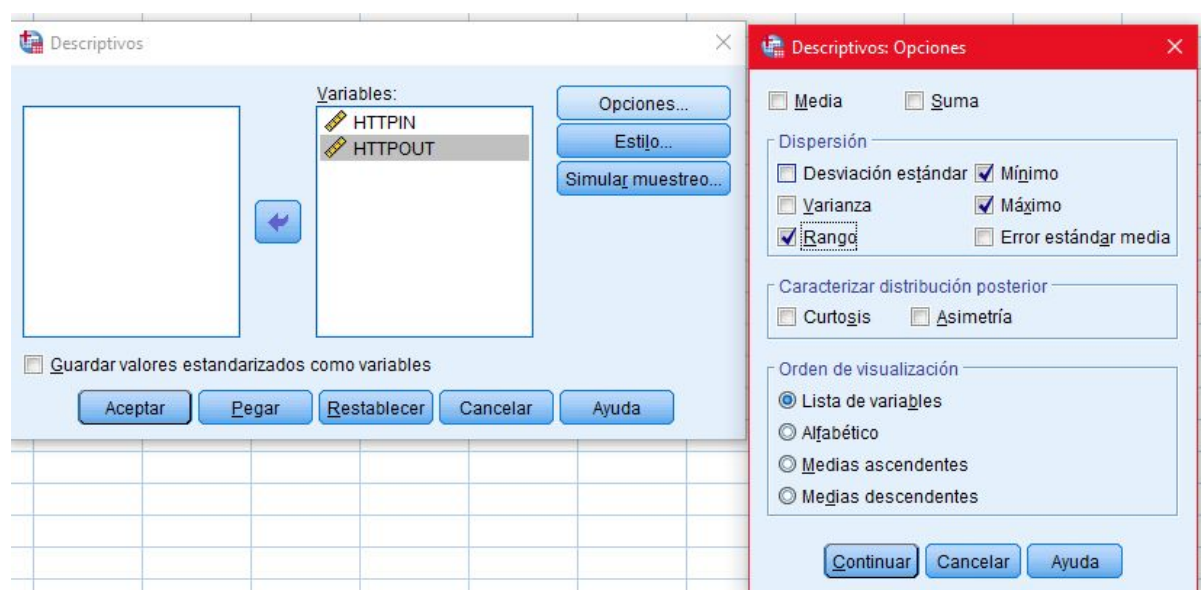
	HTTPIN	HTTPOUT
1	345	324
2	634	39
3	456	519
4	32	21
5	666	984
6	9	720
7	671	921
8	754	452
9	399	52
10	621	410

...

Ahora, vamos al submenú Analizar > Estadísticos descriptivos > Descriptivos.



Especificamos que queremos el rango, el max y el min.



Y ahora, ya tenemos con qué trabajar.

Estadísticos descriptivos

	N	Rango	Mínimo	Máximo
HTTPIN	28	878	9	887
HTTPOUT	28	963	21	984
N válido (por lista)	28			

El ancho del intervalo que podemos usar es 200, entonces.

Num. Intervalo = Rango / Ancho.

Al hacer el cálculo nos dará un resultado con decimales, alrededor de 4,9, para que el intervalo sea igual y acapare todos los valores, redondeamos por lo alto extendiendo el número de intervalos a 5.

Realizamos las agrupaciones:

Recodificar en distintas variables

Variable numérica -> Variable de salida:

HTTPOUT --> HTTPOUT_AGRUPADO
HTTPIN --> HTTPIN_AGRUPADO

Variable de salida

Nombre: HTTPOUT_AGRUPADO
Etiqueta: HTTPOUT AGRUPADO

Cambiar

Valores antiguos y nuevos...

Si... (condición de selección de caso opcional)

Aceptar Pegar Restablecer Cancelar Ayuda

Recodificar en variables diferentes: valores antiguo y nuevo

Valor antiguo

☐ Valor:

☐ Perdido del sistema

☐ Perdido por el sistema o el usuario

☒ Rango:

hasta

☐ Rango, LOWEST hasta el valor:

☐ Rango, valor hasta HIGHEST:

☐ Todos los demás valores

Valor nuevo

☒ Valor:

☐ Perdido del sistema

☐ Copiar valores antiguos

Antiguo --> Nuevo:





0 thru 200 --> 100
200 thru 400 --> 300
400 thru 600 --> 500
600 thru 800 --> 700
800 thru 1000 --> 900

Añadir
Cambiar
Eliminar

☐ Las variables de salida son series Anchura: 8

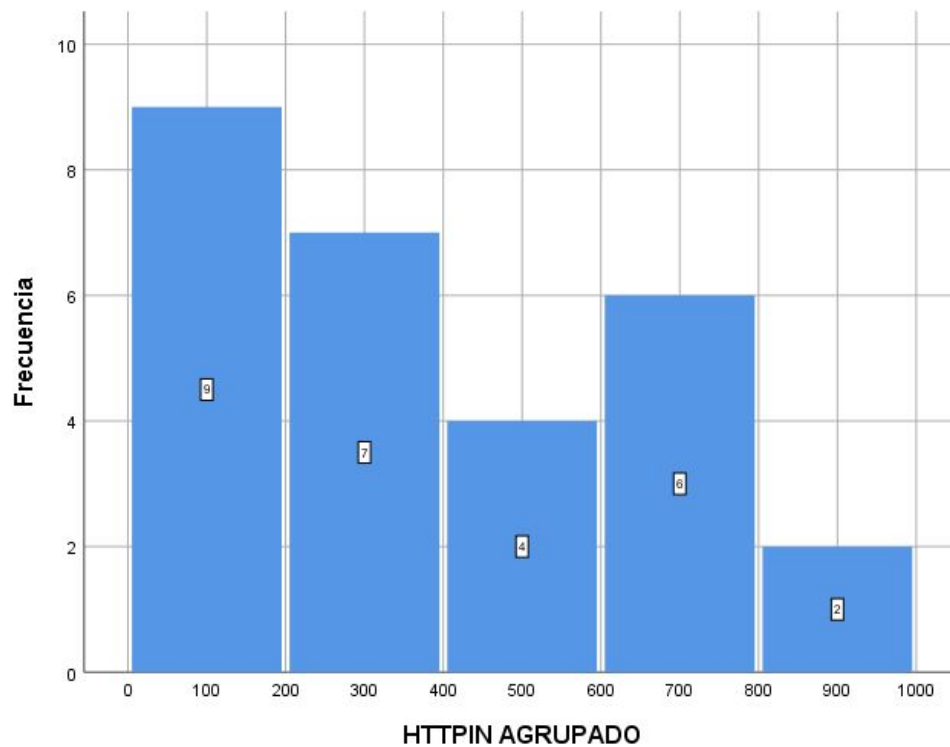
☐ Convertir series numéricas a números ('5'-->5)

Continuar Cancelar Ayuda

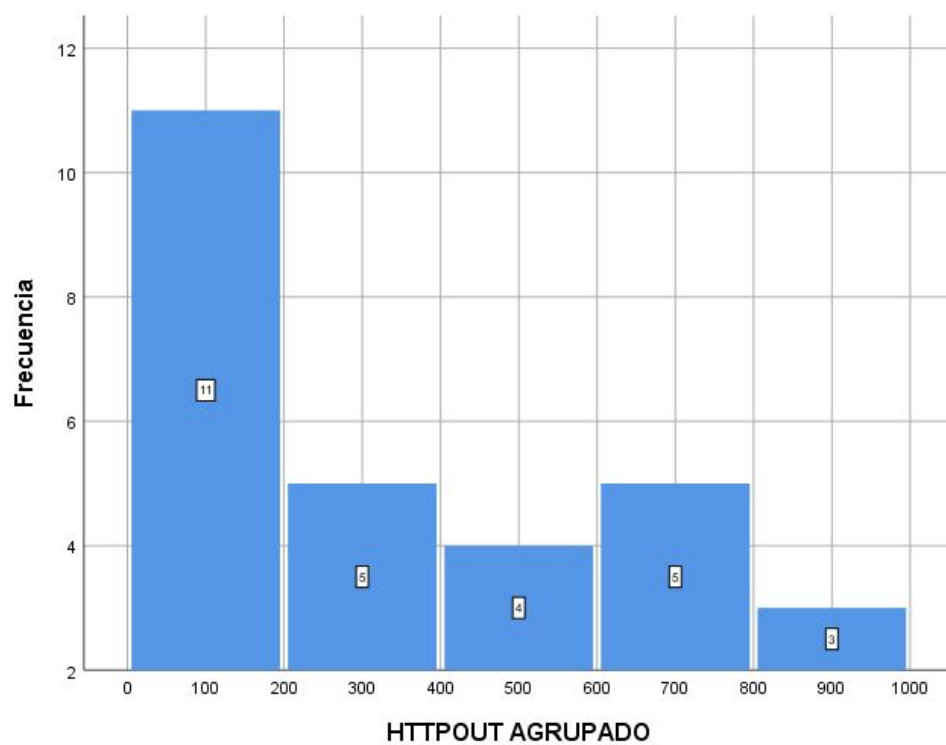
	 HTTPIN	 HTTPOUT	 HTTPOUT_AGRUPADO	 HTTPIN_AGRUPADO
1	345	324	300	300
2	634	39	100	700
3	456	519	500	500
4	32	21	100	100
5	666	984	900	700
6	9	720	700	100
7	671	921	900	700
8	754	452	500	700
9	399	52	100	300
10	621	410	500	700
11	43	61	100	100
12	333	197	100	300
13	71	317	300	100
14	371	173	100	300
15	234	852	900	300
16	11	85	100	100
17	887	297	300	900
18	448	621	700	500
19	452	662	700	500
20	875	222	300	900
21	121	81	100	100
22	354	624	700	300
23	353	712	700	300
24	789	271	300	700
25	594	424	500	500
26	22	49	100	100
27	94	91	100	100
28	30	73	100	100

Ahora, realizaremos el histograma..

Histograma de peticiones de entrada.



Histograma de peticiones de salida.



b) ¿Qué conclusiones generales pueden extraerse?

Podemos concluir que tiene sentido, por el mero hecho de que la cantidad de peticiones de salida como de entrada coinciden ligeramente, ya que por lo general, una transacción entre servidor y cliente casi siempre suele ser bidireccional.

También podemos decir que el servidor recibe más datos de los que devuelve en lo general.