

PRÁCTICA 2

ESTADÍSTICA

Javier Rivilla Arredondo

53247378D



ESTADÍSTICA

Objetivos

El objetivo de esta práctica es estudiar los casos de los fenómenos que se nos pueden presentar, como el tamaño muestral y el tamaño poblacional. En cualquier de los casos, el estudio estadístico se realiza sobre un conjunto de elementos pertenecientes generalmente a un colectivo mayor que se intenta describir lo más fiablemente posible.

Ejercicios

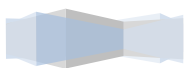
1. Explicar mediante 4 ejemplos qué diferencias hay entre una variable categórica y una medible.

Las **variables categóricas** como su nombre indica vienen de categoría mientras que las medibles vienen de medida. Ahora procedamos a los ejemplos.

Un claro ejemplo de variable **categórica nominal** sería la raza ya que como no se puede definir un orden natural entre sus categorías esta será blanca, negra. Otro claro ejemplo sería la **religión** ya puede ser católica, judía, protestante. Una variable **categórica nominal**, destaca por el *orden* como bien indica. Un ejemplo sería el rango militar (soldado, sargento, etc.) o también la clase social (alta, baja o media) que nos indican el orden o nivel que les ocupa.

Por otro lado tenemos las **variables medibles**, pueden ser *discretas* las cuáles expresan números enteros, por lo que pueden ser contados como por ejemplo **la población escolar** o bien los **nacimientos o fallecidos de una población**. También están las *continuas* que se expresan en números decimales por lo que son más exactos, como la talla de una persona (1,75cm) o bien el peso de una persona.

Con estos ejemplos podemos concluir diciendo, que las variables **categóricas** no pueden medir, pero si clasificar tanto en **nominales** como en **ordinales** y las variables **medibles** son las que *miden* y se pueden clasificar tanto en **discretas** como en **continuas**.



2. Clasificar las siguientes variables

Nombre	Categórica nominal
Edad (en años)	Medible discreta
Peso	Medible continua
Nota media del expediente(0 a 10)	Medible continua
Número gigabytes descargados de Internet en un periodo de tiempo	Medible discreta
Red social a la que perteneces	Categórica nominal
Marca personal en salto de longitud	Medible continua
Color de ojos	Categórica nominal
Compresión lectora(baja, media, alta)	Categórica ordinal
Estado de conservación de diferentes ordenadores	Categórica ordinal

3. Se desea hacer un estudio sobre uso de dispositivos móviles por parte de los estudiantes del Grado de Ingeniería Informática. Proponer de forma razonada seis variables que podrían utilizarse para realizar el estudio. ¿Qué tipo de variable es cada una de ellas? Con esas variables ¿qué tipo de información y/o conclusiones podrías obtener?

1. **Edad** (*Categórica nominal*)
2. **Sexo**(*Categórica nominal, dicotómica*)
3. **Almacenamiento**(*Medible discreta*)
4. **Pantalla**(*Medible discreta*)
5. **Megapíxeles de la cámara**(*Variable medible discreta*)
6. **Marca**(*Categórica nominal*)

La conclusión que podemos sacar con todas las variables anteriores sería que obtendríamos información de los tipos de usuarios de distintas edades, sexo e información (estas son las que les identifican) sobre cuáles son las características de sus móviles como de cuantas pulgadas es la pantallas, que marca de teléfonos móviles utilizan y como también de cuantos Megapíxeles tiene la cámara y la capacidad de almacenamiento del dispositivo.

4. En la tabla adjunta tenemos la representación de un conjunto de datos obtenidos de una población; se trata de una muestra de 20 ordenadores de una tienda informática, de los cuáles observamos varias características; para cada ordenador obtenemos datos correspondientes a las variables:



X1: Sistema Operativo preinstalado. (2)

X2: Tiempo en segundos de conexión a Internet en un determinado periodo de tiempo.

X3: Número de periféricos

X4: Marca. (1)

X5: Precio en euros.

Bueno, comenzamos a hacer la práctica:

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	X1	Númérico	8	0	SO	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
2	X2	Númérico	8	0	Tiempo_conexion	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
3	X3	Númérico	8	0	Periféricos	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
4	X4	Númérico	8	0	Marca	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada
5	X5	Númérico	8	0	Precio	Ninguna	Ninguna	8	Derecha	Desconocido	Entrada

a) Clasificar las variables. Dar una explicación razonada de la clasificación.

X1: Variable categórica nominal, ya que no hay ordenación ninguna y tampoco puede ser medible.

X2: Variable medible continua, debido a que se puede medir

X3: Variable medible discreta, como a su vez es constituido por un conjunto finito. Además se puede medir.

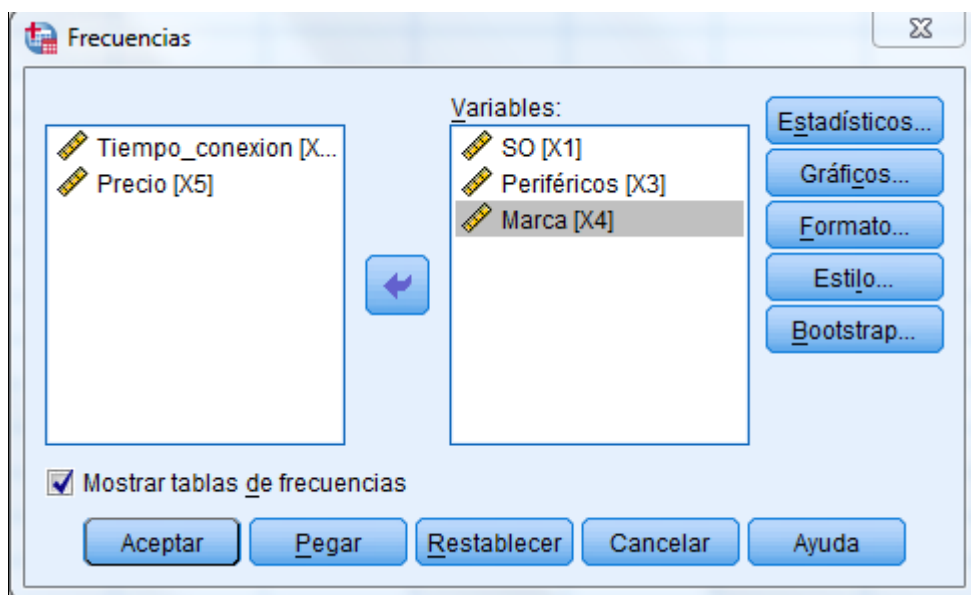
X4: No es medible, no existe orden en las marcas y es una variable categórica nominal.

X5: Es una variable que podemos medir con una unidad de medida, pudiéndose tomar cualquier valor del intervalo. Es una variable medible continua.

b) Construir las tablas de frecuencias de las variables X1, X3 y X4.
¿Qué conclusiones puedes extraer de las tablas de frecuencias?



	X1	X2	X3	X4	X5
1	2	989	6	2	1340
2	5	1900	3	6	1025
3	7	1020	5	3	795
4	1	994	3	5	904
5	3	1375	3	1	1020
6	4	3000	4	2	934
7	3	945	3	4	1170
8	4	2999	4	5	898
9	6	2184	3	7	945
10	1	3150	4	2	1999
11	4	1955	7	4	1085
12	2	1025	3	6	1120
13	3	895	5	1	1955
14	7	1170	7	4	1300
15	1	1989	2	3	1095
16	3	2900	3	1	1445
17	5	1999	5	6	999
18	3	2085	3	1	1385
19	4	3250	4	6	1120
20	6	1855	6	8	855



Le damos **Analizar->Estadísticos descriptivos->frecuencias** (como en la imagen anterior) y obtenemos las siguientes gráficas de frecuencias:



➔ Frecuencias

[Conjunto_de_datos1] C:\Users\JAVIER\Desktop\Practica2_EST.sav

Estadísticos

		SO	Periféricos	Marca
N	Válido	20	20	20
	Perdidos	0	0	0

Tabla de frecuencia

SO

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Windows 10	3	15,0	15,0	15,0
	Linux	2	10,0	10,0	25,0
	Mac Os	5	25,0	25,0	50,0
	Windows 8	4	20,0	20,0	70,0
	UNIX	2	10,0	10,0	80,0
	Windows 7	2	10,0	10,0	90,0
	FreeBSD	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Periféricos

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	2	1	5,0	5,0	5,0
	3	8	40,0	40,0	45,0
	4	4	20,0	20,0	65,0
	5	3	15,0	15,0	80,0
	6	2	10,0	10,0	90,0
	7	2	10,0	10,0	100,0
	Total	20	100,0	100,0	

Marca

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	APPLE	4	20,0	20,0	20,0
	TOSHIBA	3	15,0	15,0	35,0
	ACER	2	10,0	10,0	45,0
	SONY	3	15,0	15,0	60,0
	LENOVO	2	10,0	10,0	70,0
	HP	4	20,0	20,0	90,0
	SAMSUNG	1	5,0	5,0	95,0
	ASUS	1	5,0	5,0	100,0
	Total	20	100,0	100,0	



Aquí tenemos las 3 tablas de frecuencias de los **Sistemas operativos, los periféricos y la marca.**

Primero comentaremos los sistemas operativos, como podemos observar el sistema más utilizado es MAC OS, con un porcentaje del 25%. Posteriormente está **Windows 8** que en este caso consta de un 20%. Esto a su vez nos verifica los menos usados, que en este caso son **LINUX, UNIX, Windows 7 y FreeBSD** con un 10% cada uno.

En cuanto a los periféricos podemos observar que un 40% de los ordenadores que compra la gente tiene 3 periféricos conectados. Siendo esta la mayor configuración de periféricos usada. En cambio el otro 5% restante es la cantidad de periféricos siendo 2.

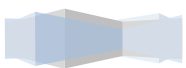
Por último comentaremos las marcas respectivas. Como podemos observar en la tabla tenemos **APPLE, TOSHIBA, ACER, SONY, LENOVO, HP, SAMSUNG, y ASUS**. En este caso la marcas más vendida es la de **APPLE** y la de **HP** cada una con un porcentaje del 20% y las menos vendidas son **SAMSUNG y ASUS** con tan solo un 5% cada una.

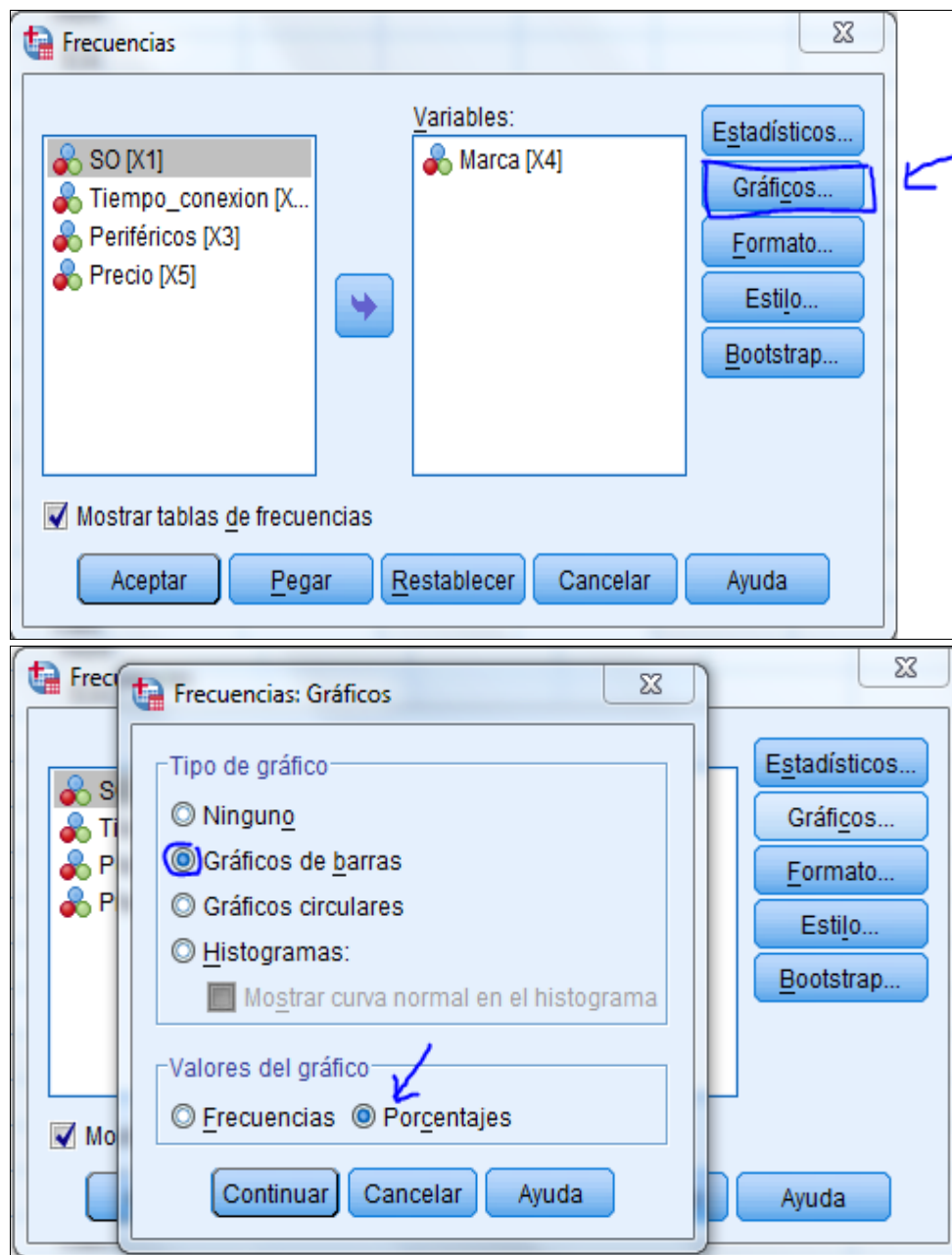
c) En general, ¿qué ventajas supone utilizar las frecuencias relativas en vez de las frecuencias absolutas?

Desde mi punto de vista, las frecuencias relativas suponen más útiles si queremos ver los porcentajes y las proporciones por información de estas. En cambio, las frecuencias absolutas son más precisas pero con menos datos.

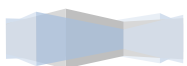
d) Representar la variable X4 de dos formas distintas: mediante un diagrama de barras y mediante un diagrama de sectores en el que se incluyan los porcentajes. ¿Qué conclusiones extraes? ¿Qué ventajas/desventajas ofrece el diagrama de sectores frente al diagrama de barras?

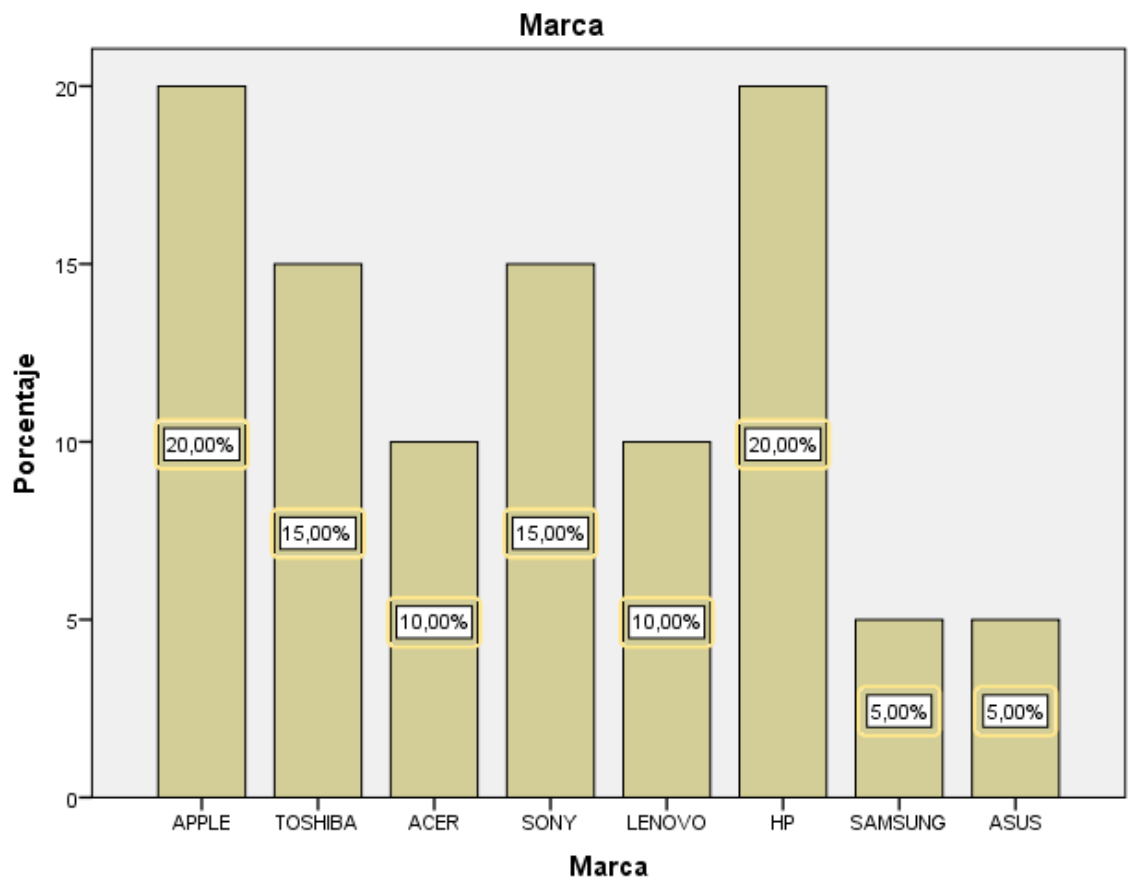
En este caso lo que tenemos que hacer es tan sencillo como lo que hemos hecho anteriormente. Nos vamos a **Analizar->Estados descriptivos->frecuencias->Gráficos->Gráficos de barras**. Eso sí, tendremos que decir si queremos el porcentaje (en este caso) y pasar de la frecuencia.





Saldría el siguiente gráfico de barras:



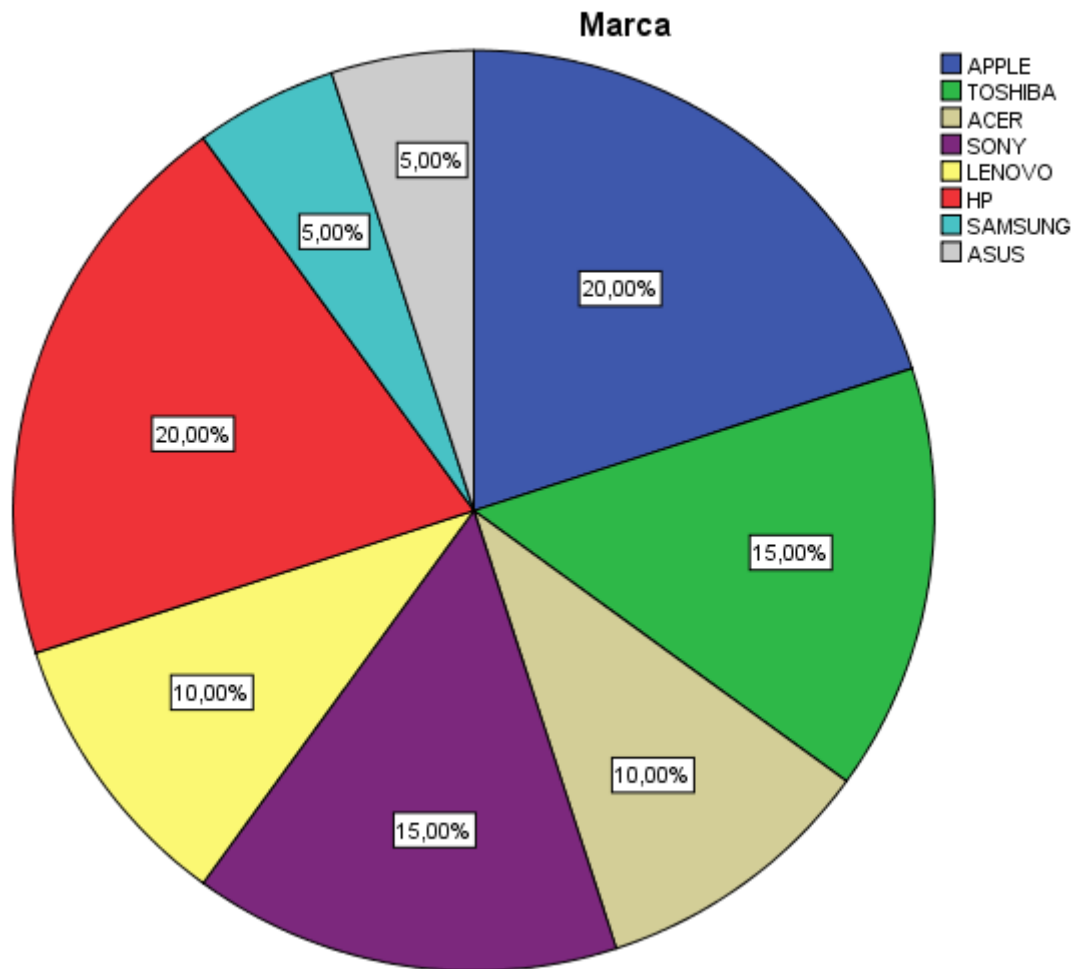


A continuación queremos el diagrama de sectores de marca, por lo que haremos lo mismo que lo anterior:

Analizar->Estados descriptivos->frecuencias->Gráficos->Gráficos circulares



Y nos haría el siguiente gráficos de sectores:



A simple vista extraemos la conclusión de que ambos diagramas se ve claramente como ocupan más superficie la marca **HP** y **APPLE**.

En cambio **ASUS** y **SAMSUNG** vemos que son las más bajas y las que menos predominan.

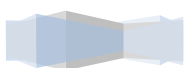
Por otro lado las ventajas/desventajas entre ambos diagramas, el diagrama de sectores ofrece más claridad junto con su leyenda, a la hora de entender quien tiene más predominio o menos.

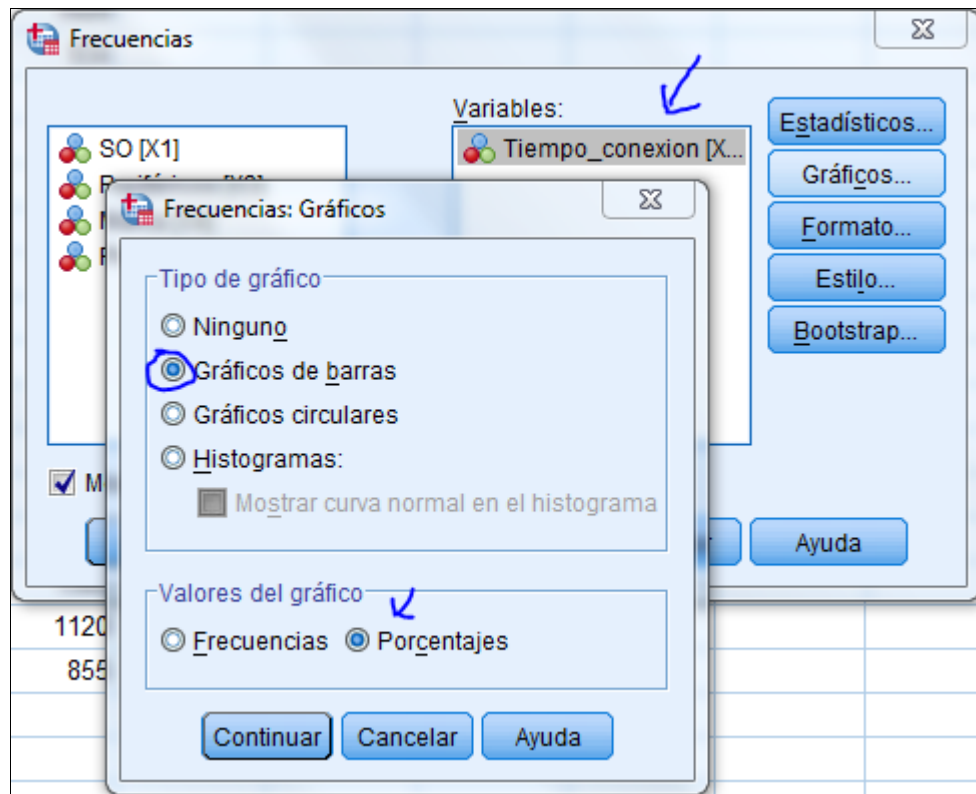
En conclusión el diagrama de barras es más preciso y el diagrama de sectores nos muestra las leyendas y podemos ver la información más proporcionada.

- e) Representar mediante un diagrama de barras la variable X2.
Interpreta el resultado.

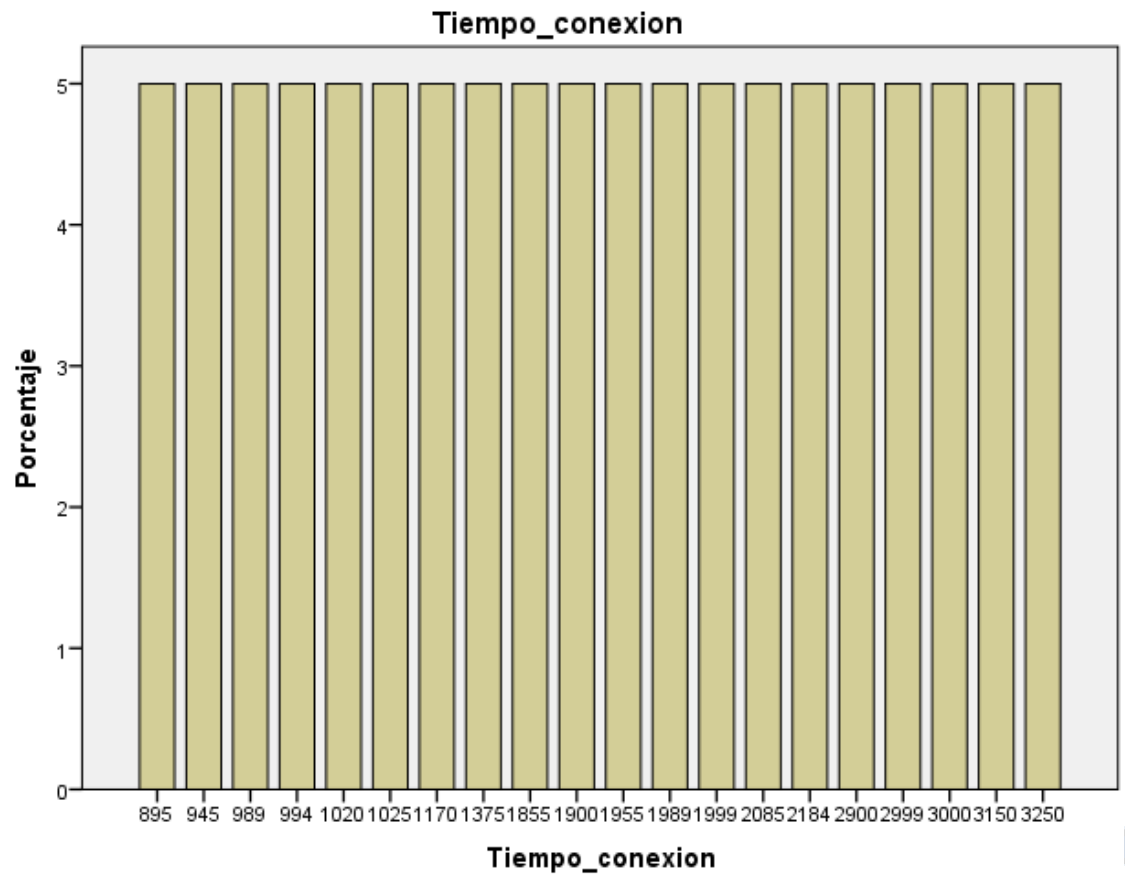
En este caso, nos vamos a **Analizar->Estados descriptivos->Frecuencias->Gráficos->Gráfico de barras**.

(lo mismo que en caso anterior solo con la variable X2).





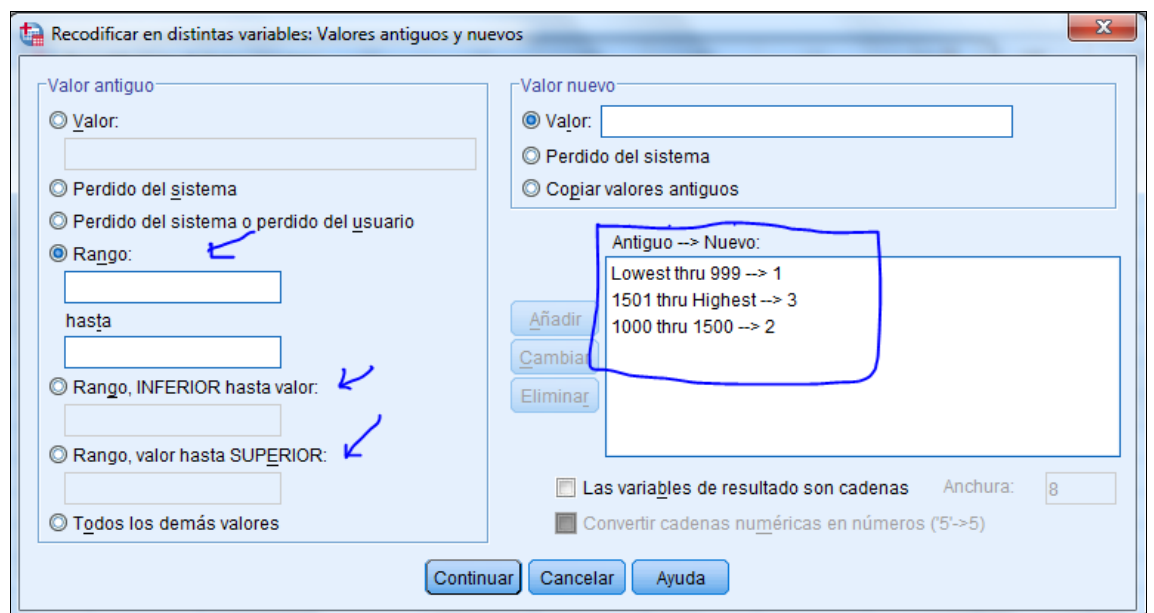
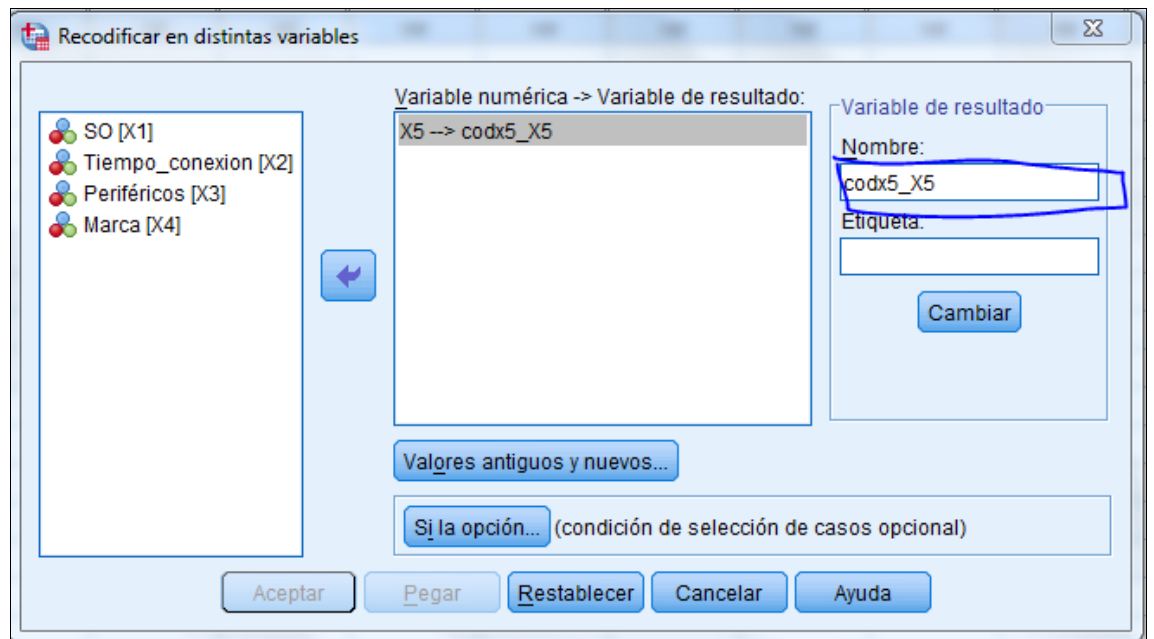
Nos quedaría el siguiente gráfico, imposible de entender desde mi punto de vista:



- f) Obtén el gráfico de sectores de la variable X5 agrupando los precios de los ordenadores en tres categorías: 1-menos de 1000 euros, 2-entre 1000 y 1500 euros, 3- más de 1500 euros. Incluye los porcentajes en el gráfico y explica lo que observas.

En este caso lo que tenemos que hacer es recodificar en distintas variables:

Transformar -> Recodificar en distintas variables

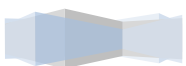
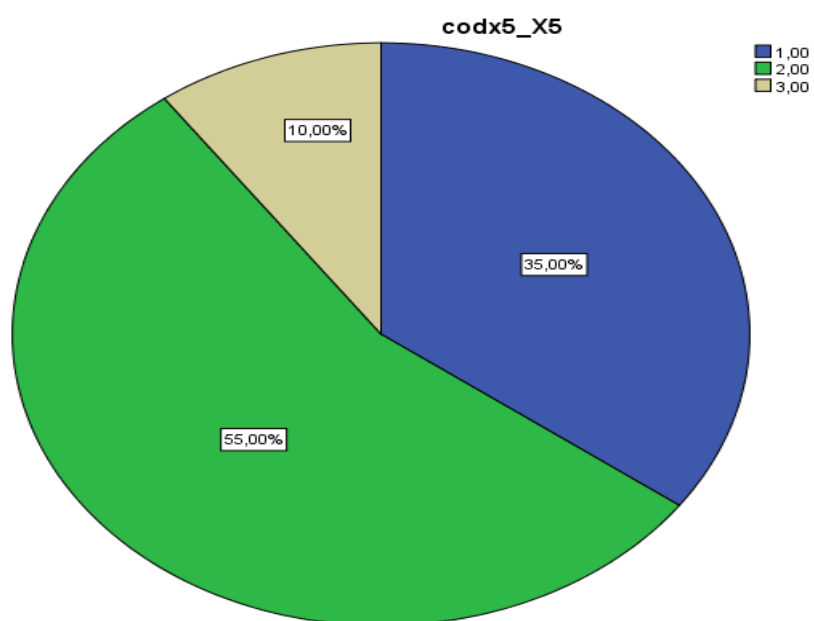


Elegimos el rango en esas casillas señaladas por las flechas.

A continuación nos sacará lo siguiente:

	X1	X2	X3	X4	X5	codx5_X5
1	2	989	6	2	1340	2,00
2	5	1900	3	6	1025	2,00
3	7	1020	5	3	795	1,00
4	1	994	3	5	904	1,00
5	3	1375	3	1	1020	2,00
6	4	3000	4	2	934	1,00
7	3	945	3	4	1170	2,00
8	4	2999	4	5	898	1,00
9	6	2184	3	7	945	1,00
10	1	3150	4	2	1999	3,00
11	4	1955	7	4	1085	2,00
12	2	1025	3	6	1120	2,00
13	3	895	5	1	1955	3,00
14	7	1170	7	4	1300	2,00
15	1	1989	2	3	1095	2,00
16	3	2900	3	1	1445	2,00
17	5	1999	5	6	999	1,00
18	3	2085	3	1	1385	2,00
19	4	3250	4	6	1120	2,00
20	6	1855	6	8	855	1,00

Y a continuación sacamos el gráfico de barras como hicimos en el apartado anterior.



Esta vez cogemos la variable creada claramente y nos sale el anterior gráfico.

Como podemos observar en el diagrama de sectores predomina con más de la mitad de ventas los ordenadores entre 1000 y 1500 euros (55%) después ordenadores con menor precio de 1000 euros y por último los que valen más de 1500 euros con tan solo un 10% de ventas.

g) ¿Qué conclusión global puede extraerse de este estudio?

La conclusión es que las configuraciones de ordenadores más demandadas son aquellas de las marcas **APPLE y HP** con un sistema operativo **MAC OS** o Windows 8, con 3 dispositivos conectados al ordenador con un rango de precio entre 100 y 1500.

Por otro lado las configuraciones de ordenadores que menos se han vendido son las marcas **SAMSUNG y ASUS** con un sistema operativo como **LINUX o Windows 7**, con 8 periféricos conectados al pc y con un precio de más de 1500 euros.

