

Nombre: Javier Rivilla Arredondo

DNI: 53247378D

Email: javin.javvn@gmail.com

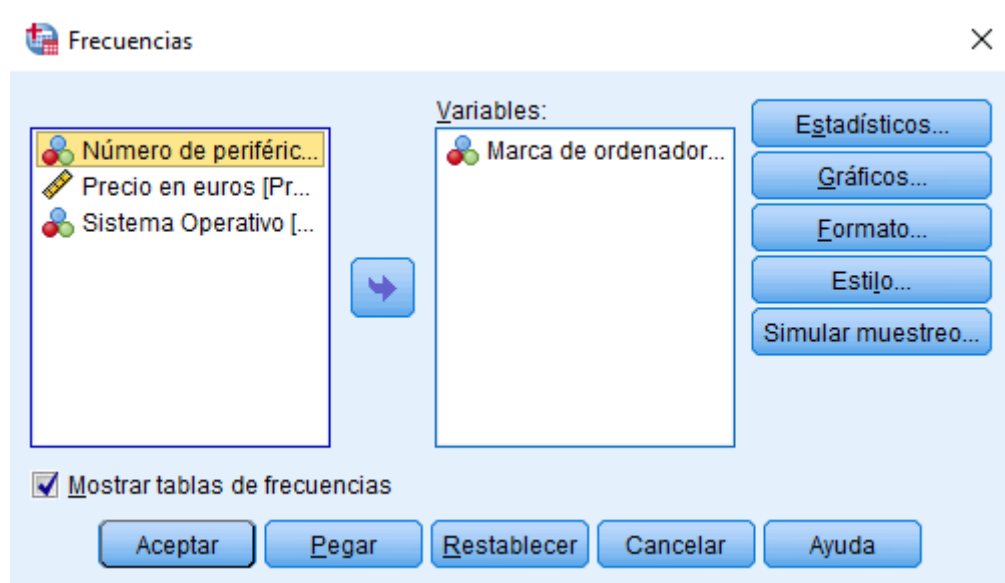
PRÁCTICA 3: ESTADÍSTICA DESCRIPTIVA

Ejercicio 3.1: Resuelve los siguientes ejercicios atendiendo al fichero lordenadores.sav.
Construye la tabla de frecuencia de la variable que indica la marca de ordenador, adjúntala a la práctica y explica lo que observas.

Solución:

Primero realizaremos las siguientes operaciones para obtener dicha tabla de frecuencia:

Analizar → Estadísticos descriptivos → Frecuencias



Como observamos ponemos la marca de ordenador, y nos mostrará la siguiente tabla de frecuencias:

Marca de ordenador					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	IBM	2	5,6	5,6	5,6
	COMPAQ	7	19,4	19,4	25,0
	ACER	4	11,1	11,1	36,1
	SUN	2	5,6	5,6	41,7
	DELL	1	2,8	2,8	44,4
	ASUS	12	33,3	33,3	77,8
	HP	8	22,2	22,2	100,0
	Total	36	100,0	100,0	

Podemos observar con qué frecuencia aparece las marcas, con esto podemos saber que tenemos un total de 36 marcas de las cuales 2 corresponden a IBM ($2/36 * 100$ es el porcentaje = 5,6), el porcentaje válido sería como el porcentaje, cuantos ordenadores hay de IBM en total? Pues la solución sería 2 de un total de 36 ($2/36$) y el acumulado sería hasta el momento cuantos llevamos. Claramente todos tienen que sumar el 100% porque todos los ordenadores disponen de una marca, no hay ninguna suelta en este caso. Por lo que de IBM hay 2 ordenadores, COMPAQ hay 7 ordenadores, ACER hay 4 ordenadores y así sucesivamente...

Obtén el gráfico de sectores agrupando los precios de los ordenadores en tres categorías: 1- menos de 800 euros, 2- entre 800 y 1200 euros, 3- más de 1200 euros. Incluye los porcentajes en el gráfico, adjúntalo en la práctica y explica lo que observas.

Solución:

Primero agrupamos los precios de los ordenadores en una variable que se llame **categoría1** →

Recodificamos en distintas variables seleccionamos la variable que necesitamos en este caso el precio.

Recodificar en distintas variables

Variable numérica -> Variable de salida:
Precio --> categoria

Variable de salida
Nombre:
Etiqueta:
Cambiar

Valores antiguos y nuevos...

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

Aceptar Pegar Restablecer Cancelar Ayuda

→ Valores antiguos y nuevos...

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:

Lowest thru 799 --> 1

800 thru 1200 --> 2

1201 thru Highest --> 3

Añadir

Cambiar

Eliminar

☐ Las variables de salida son series Anchura: 8

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

Continuar Cancelar Ayuda

En rango ponemos las características de cada categoría:

- **Categoría 1:** Menores que 799.
- **Categoría 2:** Rango de 800 y 1200.
- **Categoría 3:** Más de 1201 (Porque se incluye el número, y son mayores de 1200).

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Generador de gráficos...
Selector de plantillas de tablero...
Gráfico de Weibull...
Comparar subgrupos
Cuadros de diálogo antiguos

Visible: 5 de 5 variables

	Marca	Periféricos							
1	6	8							
2	6	8							
3	2	3	1008	4					
4	6	4	1283	3					
5	2	7	1436	2					
6	3	6	939	3					
7	7	7	1036	4					
8	7	4	876	2					
9	6	3	997	4					
10	7	4	1297	1					
11	7	5	1112	4					
12	6	7	1367	3					
13	1	6	842	2					
14	6	8	715	3					
15	2	5	1027	4	2				
16	3	6	1089	2	2				

Barras...
Barras 3D...
Líneas...
Áreas...
Circular...
Máximos y mínimos...
Diagramas de cajas...
Barras de error...
Pirámide de población...
Dispersión/Puntos...
Histograma...

Una vez hemos sacado las categorías creamos el grafico de sectores, para ello nos vamos a **Gráficos** → **Cuadros de diálogo antiguos** → **Circular...**

Definir gráfico circular: Resúmenes para grupos de casos

Las porciones representan

☒ N.º de casos ☐ % de casos

☐ Sumas de la variable

Variable:

Definir porciones por:

categoria

Panel mediante

Filas:

Anidar variables (sin filas vacías)

Columnas:

Anidar variables (sin columnas vacías)

Plantilla

☐ Usar las especificaciones gráficas de:

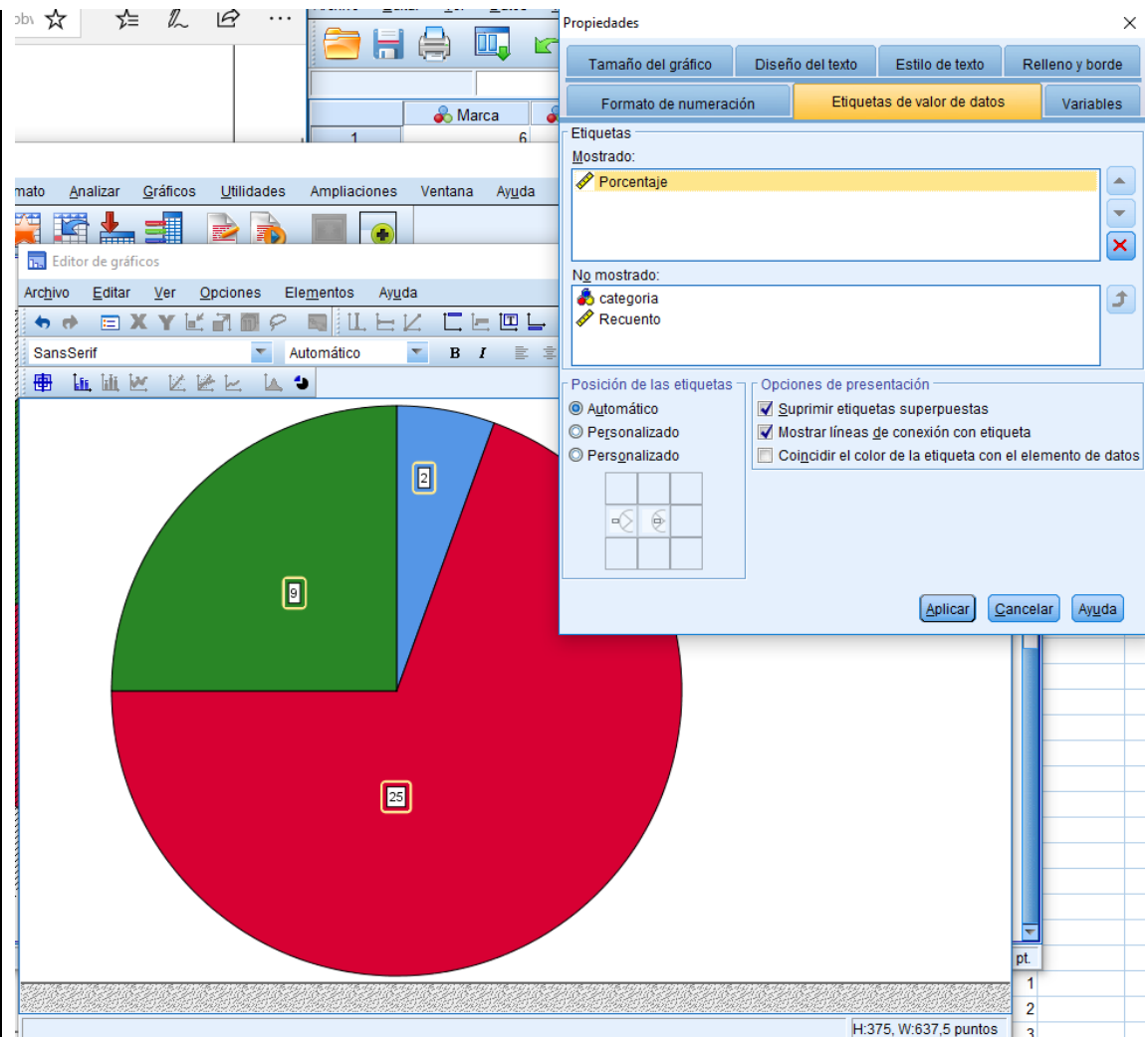
Archivo...

Acceptar Pegar Restablecer Cancelar Ayuda

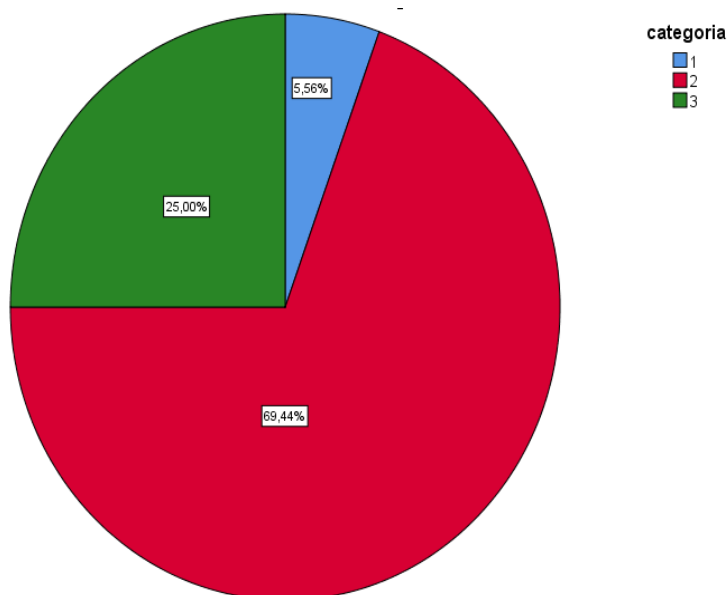
Títulos...

Opciones...

Seleccionamos por categoría y aceptamos, nos saldrá el siguiente gráfico circular:



Así se ven el total de marcas por categoría, pero visualmente el porcentaje representa un significado mayor, por eso vamos a poner el porcentaje:



Visualmente podemos observar que tenemos tres categorías donde cada porción del sector se refiere a una de estas, el **5,56%** pertenece a la **categoría 1**, es decir, el porcentaje de ordenadores con menor precio de 800, el rojo corresponde a la **categoría 2**, porcentaje de ordenadores donde el precio está comprendido entre 800 y 1200 correspondiente a un **69,44%** y por último la **categoría 3** que esta representada por el color verde donde el porcentaje es mayor a 1200 euros por un **25%**.

Ejercicio 3.2: Completa las siguientes frases, atendiendo al fichero 1car.sav.

Solución:

De los coches muestreados, el **63,3** por ciento han sido fabricados en EEUU, el **17,2** por ciento fabricados en Europa y el **19,2** por ciento en Japón.

La aceleración media de dichos coches es de **15,4951** segundos de 0 a 100 km/h. Los coches con menor tiempo de aceleración media son los de **EE.UU** con una media de **14,9237** segundos y los de mayor tiempo de aceleración los de **Europa** con un tiempo medio de **16,8** segundos. Respecto al peso, los coches con mayor peso medio son los de **EEUU** y los de menor peso medio los de **Japón**. El número de cilindros varía de **66** a **7456**, con una media de **3179,73** cilindros y una mediana de **2433,5** cilindros. La cilindrada media inferior corresponde a los coches de **Japón** con una media de **1689,96** cc.

Coches fabricados:

The screenshot shows the SPSS 'Frecuencias' (Frequencies) dialog box. On the left, a list of variables includes 'Consumo (l/100Km...)', 'Cilindrada en cc [m...]', 'Potencia (CV) [cv]', 'Peso total (kg) [peso]', 'Aceleración 0 a 100 ...', 'Año del modelo [año]', and 'Número de cilindro...'. The 'País de origen [orig...]' variable is selected and moved to the 'Variables:' box on the right. Below the variable lists, the checkbox 'Mostrar tablas de frecuencias' is checked. At the bottom, there are buttons for 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda'.

País de origen					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	EE.UU.	257	63,3	63,5	63,5
	Europa	70	17,2	17,3	80,7
	Japón	78	19,2	19,3	100,0
	Total	405	99,8	100,0	
Perdidos	Sistema	1	,2		
Total		406	100,0		

Aceleración media:

Medias
 ×

Consumo (l/100Km...)
 Cilindrada en cc [m...
 Potencia (CV) [cv]
 Peso total (kg) [peso]
 Año del modelo [año]
 Número de cilindro...

→

Lista de dependientes:
 Aceleración 0 a 100 k...

Capa 1 de 1

Anterior

Siguiente

↶

Lista de independientes:
 País de origen [orig...

Opciones...

Estilo...

Simular muestreo...

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Informe

Aceleración 0 a 100 km/h (segundos)

País de origen	Media	N	Desv. Desviación
EE.UU.	14,9237	257	2,81167
Europa	16,8000	70	3,03778
Japón	16,2064	78	1,94358
Total	15,4951	405	2,81174

Peso medio:

Medias
 ×

Consumo (l/100Km...)
 Cilindrada en cc [m...
 Potencia (CV) [cv]
 Aceleración 0 a 100 ...
 Año del modelo [año]
 Número de cilindro...

↶

Lista de dependientes:
 Peso total (kg) [peso]

Capa 1 de 1

Anterior

Siguiente

→

Capa 1 de 1
 País de origen [orig...

Opciones...

Estilo...

Simular muestreo...

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Informe

Peso total (kg)

País de origen	Media	N	Desv. Desviación
EE.UU.	1118,9027	257	264,09871
Europa	806,7571	70	162,75717
Japón	740,4615	78	107,39243
Total	992,0667	405	281,64741

Cilindrada:

Estadísticos

Cilindrada en cc

N	Válido	406
	Perdidos	0
Media		3179,7340
Mediana		2433,5000
Varianza		2972220,092

Cilindrada en cc

País de origen	Media	N	Desv. Desviación	Mínimo
EE.UU.	4022,8949	257	1632,67322	1147,00
Europa	1788,6429	70	370,02969	1114,00
Japón	1689,9615	78	376,69311	1147,00
Total	3187,4222	405	1719,16243	1114,00

Ejercicio 3.3: Contesta a las siguientes preguntas, basándote en el cálculo de percentiles y atendiendo al fichero lordenadores.sav.

Solución:

Se estima que el 20 por ciento de los ordenadores cuestan al menos **1272,60** euros.

Se estima que el 40 por ciento de los ordenadores cuestan menos de **997** euros.

Se estima que de los ordenadores de la marca ASUS, el 80 por ciento cuestan menos de **1283** euros.

euros.


Explicación de la resolución del ejercicio:


Cuando nos dicen al menos nos están dando a entender que precio **al menos tiene ese 20 por ciento**, es decir, el **P20 nos dará el precio menor o igual** pero no nos está dando el que podría tomar para ello tendremos que escoger el **P80**, este nos dará el **precio que cuesta del sobrante 80% menos o iguales a este**.


Estadísticos		
Precio en euros		
N	Válido	36
	Perdidos	0
Percentiles	40	997,00
	80	1272,60


Y en el siguiente apartado nos pide **el 40% de los ordenadores que cuestan menos de**, por lo que tendremos que utilizar el **P40** ya que el percentil nos dará el valor menor o igual a ese porcentaje.


Como bien sabemos, nos piden la marca ASUS para ello tendremos que **filtrar** los casos en los que la **marca sea igual a 6** de esta forma nos dará los ordenadores que corresponden solamente a dicha opción:


 Selecccionar casos ✕


 Marca de ordenador...

 Número de periféric...

 Precio en euros [Pr...

 Sistema Operativo [...]

 categoria

 Marca=6 (FILTER) [fi...

Selecccionar

☐ Todos los casos

☒ Si se satisface la condición

☐ Muestra aleatoria de casos

☐ Basándose en el rango del tiempo o de los casos


☐ Usar variable de filtro:

Sí...

Marca=6

Ejemplo...

Rango...



Resultado

☒ Descartar casos no seleccionados

☐ Copiar casos seleccionados a un nuevo conjunto de datos

☐ Eliminar casos no seleccionados

Nombre de conjunto de datos:

Estado actual: Filtrar casos por valores de filter_\$

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Como observamos en la siguiente foto nos saldrán tachados los ordenadores cuya marca no es igual a 6:


*ordenadores.sav_2018_2019.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

1 : filter_\$ 1 Visible: 6 de 6 variat

	Marca	Periféricos	Precio	SO	categoria	filter_\$	var	var	var
1	6	8	1089	4	2	1			
2	6	8	1092	1	2	1			
3	2	3	1008	4	2	0			
4	6	4	1283	3	3	1			
5	2	7	1436	2	3	0			
6	3	6	939	3	2	0			
7	7	7	1036	4	2	0			
8	7	4	876	2	2	0			
9	6	3	997	4	2	1			
10	7	4	1297	1	3	0			
11	7	5	1112	4	2	0			
12	6	7	1367	3	3	1			
13	1	6	842	2	2	0			
14	6	8	715	3	1	1			
15	2	5	1027	4	2	0			
16	3	6	1089	2	2	0			
17	7	3	1092	1	2	0			
18	7	8	1008	4	2	0			
19	6	8	1283	1	3	1			
20	7	3	1436	2	3	0			
21	7	4	939	2	2	0			
22	6	7	1036	1	2	1			
23	1	6	876	1	2	0			
24	5	7	997	2	2	0			
25	4	4	1257	2	3	0			
26	2	3	1052	4	2	0			
27	6	4	891	2	2	1			
28	6	5	828	2	2	1			
29	2	7	824	4	2	0			
30	6	6	1228	4	3	1			
31	2	8	1170	2	2	0			
32	3	5	929	2	2	0			
33	3	6	866	3	2	0			
34	6	3	751	1	1	1			
35	4	8	905	4	2	0			
36	2	7	1449	4	3	0			
37									

Una vez realizado dicho filtro procedemos a calcular el percentil correspondiente:

 Frecuencias: Estadísticos

Valores percentiles

☐ Cuartiles

☐ Puntos de corte para: 10 grupos iguales

☒ Percentiles:

Añadir

Cambiar

Eliminar

Tendencia central

☐ Media

☐ Mediana

☐ Moda

☐ Suma

☐ Los valores son puntos medios de grupos

Dispersión

☐ Desviación estándar

☐ Varianza

☐ Rango

☐ Mínimo

☐ Máximo

☐ Error estándar media

Caracterizar distribución posterior

☐ Asimetría

☐ Curtosis

Continuar Cancelar Ayuda

Como bien sabemos el percentil es una medida de posición, en este caso queremos obtener el precio del 80% de los ordenadores de marca ASUS, el percentil 80 (**P80**) nos dará el precio de los ordenadores **menores o iguales a 80** por lo que escogeremos ese valor para nuestro percentil puesto que nosotros sabemos que el 80% costaran ese precio o menos.

Estadísticos

Precio en euros

N	Válido	12
	Perdidos	0
Percentiles	80	1283,00

El precio por lo tanto será **1283 euros** corresponde al **percentil 80**, es decir, nos va a dar el precio que cuesta ese **precio o menos** de ese precio.

Ejercicio 3.4: Atendiendo al fichero lordenadores.sav, obtén los gráficos de sectores que expliquen cómo se distribuyen los porcentajes de los sistemas operativos instalados atendiendo a cada marca de ordenador. Explica lo que observas.

Solución:

Definir gráfico circular: Resúmenes para grupos de casos

☒ Número de periféricos...
☒ Precio en euros [Pr...

Las porciones representan

☒ N.º de casos ☐ % de casos
☐ Sumas de la variable

Variable:

Definir porciones por:

☒ Sistema Operativo [SO]

Panel mediante

Filas:

☒ Marca de ordenador [Marca]

☐ Anidar variables (sin filas vacías)

Columnas:

☐ Anidar variables (sin columnas vacías)

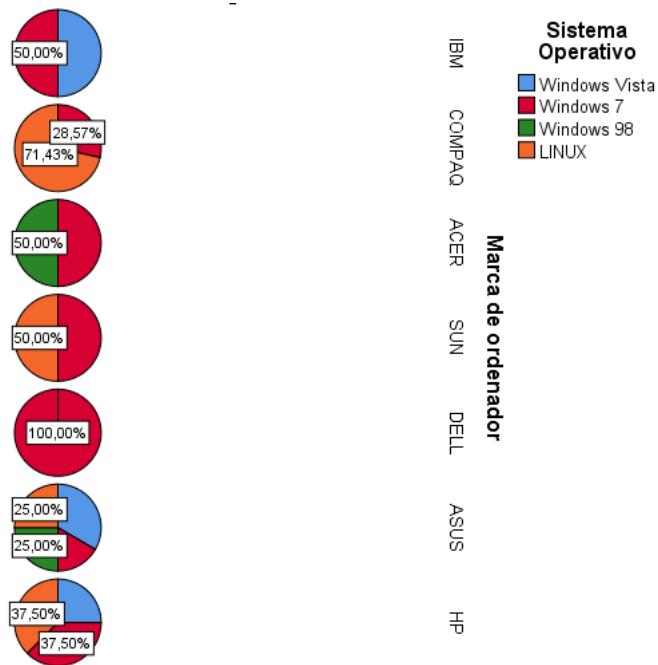
Plantilla

☐ Usar las especificaciones gráficas de:
 Archivo...

Títulos...
 Opciones...

Aceptar Pegar Restablecer Cancelar Ayuda

Como nos pone atendiendo a cada marca de ordenador por sistema operativo seleccionamos a la hora de definir nuestro gráfico circular dos opciones → por un lado las porciones que estarán compuestas por los sistemas operativos por marca ya que cada marca tiene x sistemas operativos, y por otro lado las marcas que corresponden a nuestras filas.



Como podemos observar tenemos un diagrama de sector por cada marca de ordenador, para la marca **IBM** tenemos que está formado por dos sistemas operativos (visualmente dos colores) que corresponden a **Windows vista** y el otro a **Windows 7** con un 50% cada sistema operativo.

Por otro lado, tenemos la marca **COMPAQ** formado por dos sistemas operativos, en este caso **Linux** con un **71,43%** y **Windows 7** con un **28,57%**.

Y así con las demás marcas obteniendo la siguiente tabla:

Marcas	Windows Vista (%)	Windows 7 (%)	Windows 98 (%)	Linux (%)	TOTAL (%)
IBM	50,00	50,00	-	-	100
COMPAQ	-	28,57	-	71,43	100
ACER	-	50,00	50,00	-	100
SUN	-	50,00	-	50,00	100
DELL	-	100	-	-	100
ASUS	37,50	12,50	25,00	25,00	100
HP	25,00	37,50	-	37,50	100

Para **ACER** está compuesto por un **50%** de **Windows 7** y un **50%** de **Windows 98**.

SUN está compuesto por **50%** de **Windows 7** y un **50%** de **Linux**.

DELL solo está formado por el sistema operativo **Windows 7**.

ASUS por **Windows vista** con un **37,50%**, **Windows 7** con un **12,50%**, **Windows 98** con un **25%** y **Linux** **25%** también.

HP por **Windows vista** con un **25%**, **Windows 7** con un **37,50%** y **Linux** con un **37,50%**.

Como conclusión podemos saber con estos diagramas de sectores como están compuestos las diferentes marcas por sus respectivos sistemas operativos, dándonos una información más concisa y clara mediante el uso de los colores.

Ejercicio 3.5: Realiza las siguientes cuestiones relativas al fichero 1tiemporespuesta.sav. Calcula el gráfico caja del tiempo de respuesta de los ordenadores de la marca C con ayuda del SPSS. Interpreta los resultados.

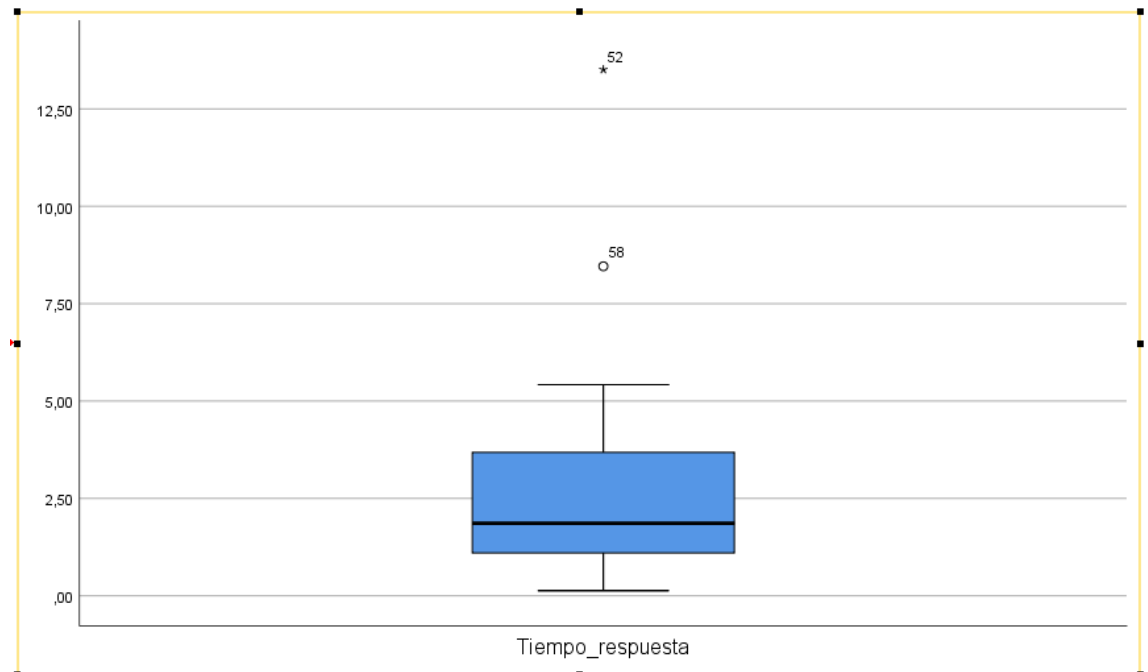
Solución:

En primer lugar, nos fijamos en la etiqueta que corresponda a la marca C, obtenemos que es la 3 y filtramos por esta:



Una vez tenemos esto y filtremos la marca que es la que nos piden creamos el diagrama de caja:

Gráficos → Cuadros de diálogos antiguos → Diagrama de cajas → simple.



Como solo la queremos de la variable tiempo respuesta ponemos la opción de una variable y listo. Procedemos a explicar el diagrama de cajas obtenido:

Como podemos apreciar en la imagen:

- **Q1** corresponde al valor 3,5 (25% de los datos).
- **Q2** el valor de la mediana corresponde a 2,50 (50% de los datos).
- **Q3** corresponde al valor 1 (75% de los datos).

Por otro lado, tenemos los llamados “bigotes” que son aquellas líneas que extienden desde la caja hasta los valores máximo y mínimos. Esto significa que hay valores atípicos en la serie. Los últimos valores que no son atípicos son los extremos de los bigotes, es decir, el 0’1 y el 5,5.

Los valores que estén fuera del intervalo 0,1 – 5,5 serán valores atípicos, estos como podemos ver son “*” que corresponde al número 13,50 y el “o” que corresponde al valor “o”.

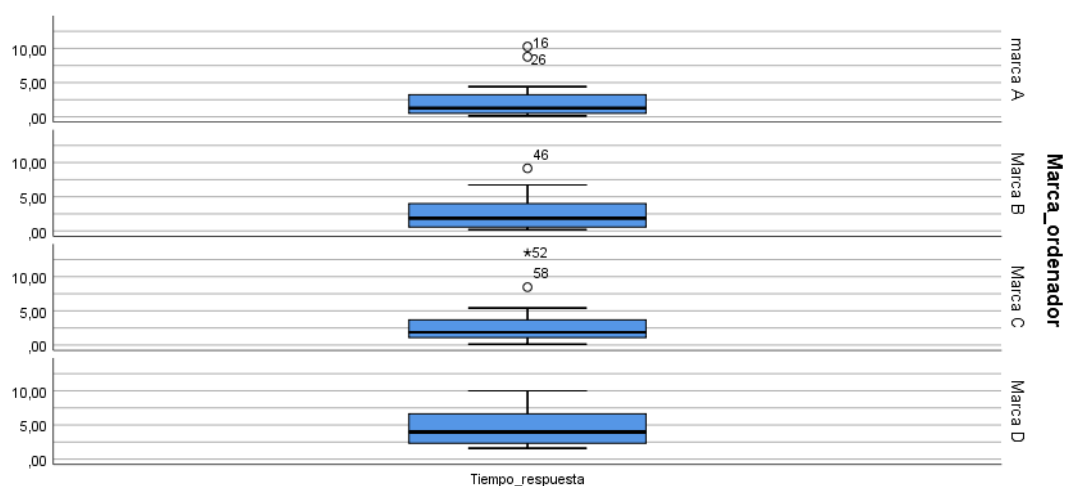
La distribución es asimétrica por la izquierda ya que **Q1** toma un valor más diferenciado.

En este diagrama de cajas obtenemos valores atípicos extremos en este caso “*” nos indica que es un valor atípico extremo y el “o” que simplemente es un valor atípico.

Obtén mediante el SPSS los gráficos caja del tiempo de respuesta de los ordenadores atendiendo a su marca. Adjúntalos e interpreta los resultados .

Solución:

Como esta vez nos pide el tiempo de respuesta por marca de ordenador tendremos que marcar la opción “**para distintas variables**”:



En este caso la marca A tiene dos valores atípicos, que no son extremos, en comparación a la marca C como antes hemos explicado tenía tanto valores extremos atípicos como valores atípicos. A la marca B le pasa lo mismo solo que esta tiene solamente un valor atípico y la marca D en cambio no tiene ningún valor atípico. También se puede observar que todas son asimétricas por la izquierda fijándote bien en los bigotes.

Ejercicio 3.6: Completa las siguientes frases, atendiendo al fichero 1salarios.sav.

Solución:

Se han muestreado a **1100** personas. De las personas muestreadas, el **38,9** por ciento ha estudiado Agrónomos, el **28,9** por ciento Económicas, el **24,5** por ciento Industriales y el **7,7** restante otras titulaciones. Un **43,5** por ciento son mujeres y un **56,5** por ciento son hombres.

El sueldo de las personas muestreadas varía de **7200** a **65500** euros, con una media de **26053,11** euros. Las personas con menor sueldo de media son los que finalizaron la titulación en el año **2005**. Las personas con mayor sueldo de media son las que estudiaron la titulación de **I. Industriales** y las de menor sueldo de media son las que estudiaron la titulación de **I. Bosques**.

El sueldo medio de las mujeres que estudiaron Ingeniería Informática es de **28683,33** euros y el de los hombres que estudiaron esa misma titulación de **28254** euros.

Explicación del ejercicio:

Para saber cuántas personas se han muestreado lo que haremos será sacar por sexo una tabla de estadísticos de esta forma obtendremos el número total de personas que hemos muestreado:



Estadísticos		
Sexo		
N	Válido	1100
	Perdidos	0

Sexo

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Mujer	479	43,5	43,5	43,5
	Hombre	621	56,5	56,5	100,0
	Total	1100	100,0	100,0	

N = 1100 personas que hemos muestreo.

Para saber cuántos han estudiado Agrónomos, eco.... Tendremos que hacer un estadístico por título

Frecuencias

Variables:

Sexo [sexo]
Salario en euros [su...]
Fecha de licenciatur...

Facultad [Título]

☒ Mostrar tablas de frecuencias

Estadísticos...
Gráficos...
Formato...
Estilo...
Bootstrap...

Aceptar Pegar Restablecer Cancelar Ayuda

Facultad

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	I. Agrónomos	428	38,9	38,9	38,9
	Arquitectura	10	,9	,9	39,8
	Informática	56	5,1	5,1	44,9
	Económicas	318	28,9	28,9	73,8
	I. Bosques	3	,3	,3	74,1
	Pedagogía	13	1,2	1,2	75,3
	I. Industriales	270	24,5	24,5	99,8
	Bellas artes	2	,2	,2	100,0
	Total	1100	100,0	100,0	

Aquí tenemos todos los porcentajes de todos los títulos estudiados en la facultad.

Para sacar el restante o bien como sabemos que tiene que sumar todo 100 se lo restamos al acumulado con las otras titulaciones o bien sumamos los % de las demás para ello haremos:

$$100 - 38,9 - 24,5 - 28,9 = 7,7\% \text{ restante}$$

El porcentaje de mujeres y hombre lo hemos sacado anteriormente con esto:

Estadísticos		
Sexo		
N	Válido	1100
	Perdidos	0

Sexo					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Mujer	479	43,5	43,5	43,5
	Hombre	621	56,5	56,5	100,0
	Total	1100	100,0	100,0	

Es una tabla de estadísticos de frecuencia por sexo.

Para sacar el sueldo de las personas haremos lo mismo, pero de la variable sueldo:

The image shows two screenshots from the SPSS software interface. The top screenshot is the 'Frecuencias' (Frequencies) dialog box, where the variable 'Salario en euros [su...]' is selected in the 'Variables:' list. The bottom screenshot is the 'Frecuencias: Estadísticos' (Frequencies: Statistics) sub-dialog box. In the 'Valores percentiles' (Percentile values) section, 'Puntos de corte para: 10 grupos iguales' is selected. In the 'Tendencia central' (Central tendency) section, 'Media' (Mean) is checked, while 'Mediana', 'Moda', and 'Suma' are unchecked. The checkbox 'Los valores son puntos medios de grupos' is also unchecked.

La media del sueldo es el siguiente:

Estadísticos		
Salario en euros		
N	Válido	1100
	Perdidos	0
Media		26053,11

El salario varía de inicio a fin es decir de los valores **7200** al valor **65500**.

	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido 7200	1	,1	,1	,1
8000	2	,2	,2	,3
8400	2	,2	,2	,5
8500	1	,1	,1	,5
8600	1	,1	,1	,6
9000	3	,3	,3	,9
10000	2	,2	,2	1,1
10400	1	,1	,1	1,2
10800	1	,1	,1	1,3
11000	2	,2	,2	1,5
11200	1	,1	,1	1,5
11500	1	,1	,1	1,6
12000	7	,6	,6	2,3
12200	1	,1	,1	2,4
12500	3	,3	,3	2,6
13000	7	,6	,6	3,3
13500	2	,2	,2	3,5
13800	1	,1	,1	3,5
14400	1	,1	,1	3,6
14500	2	,2	,2	3,8
41500	1	,1	,1	97,8
42000	3	,3	,3	98,1
43000	3	,3	,3	98,4
44000	4	,4	,4	98,7
45000	4	,4	,4	99,1
46000	1	,1	,1	99,2
48000	1	,1	,1	99,3
49000	1	,1	,1	99,4
50000	1	,1	,1	99,5
51000	1	,1	,1	99,5
51500	1	,1	,1	99,6
52000	2	,2	,2	99,8
60000	1	,1	,1	99,9
65500	1	,1	,1	100,0
Total	1100	100,0	100,0	

Las personas con menos sueldo de media por titulación deberemos comparar medias entre el salario y la licenciatura para sacar el año que menos sueldo tiene:

Medias

Lista de dependientes:

Salario en euros [suel...]

Lista de independientes:

Fecha de licenciatur...

Capa 1 de 1

Anterior

Siguiente

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Opciones...

Estilo...

Bootstrap...

Informe

Salario en euros

Fecha de licenciatura	Media	N	Desviación estándar
1995	26816,14	223	7587,520
1998	26233,92	227	7016,299
2000	25920,26	390	7215,246
2005	25440,10	260	6206,443
Total	26053,11	1100	7034,041

Ahora nos piden el salario de las personas que estudiaron una titulación para ello haremos lo mismo pero titulación por salario:

Medias

Lista de dependientes:

Sexo [sexo]
Fecha de licenciatur...

Salario en euros [suel...]

Capa 1 de 1

Anterior Siguiete

Lista de independientes:

Facultad [Título]

Aceptar Pegar Restablecer Cancelar Ayuda

Opciones...
Estilo...
Bootstrap...

Informe

Salario en euros

Facultad	Media	N	Desviación estándar
I. Agrónomos	23978,27	428	7753,055
Arquitectura	21920,00	10	5900,810
Informática	28300,00	56	3831,639
Económicas	24615,17	318	5578,340
I. Bosques	17500,00	3	5196,152
Pedagogía	22446,15	13	6559,042
I. Industriales	31010,74	270	5155,146
Bellas artes	23450,00	2	3606,245
Total	26053,11	1100	7034,041

Ahora nos piden mujeres que estudiaban ingeniería informática el sueldo medio, para ello tendremos que filtrar las mujeres que corresponden a esa titulación y comparar las medias de salario por sexo:

Filtramos:

Seleccionar casos

Facultad [Título]
 Salario en euros [su...
 Fecha de licenciatur...

Seleccionar

☐ Todos los casos

☒ Si se satisface la condición

Si la op... Título=3

☐ Muestra aleatoria de casos

Ejemplo...

☐ Basándose en el rango del tiempo o de los casos

Rango...

☐ Usar variable de filtro:

Resultado

☒ Descartar casos no seleccionados

☐ Copiar casos seleccionados a un nuevo conjunto de datos

Nombre de conjunto de datos:

☐ Eliminar casos no seleccionados

Estado actual: No filtrar casos

Aceptar Pegar Restablecer Cancelar Ayuda

Y nos sacara solo los que pertenezcan a ing. Informática:

	sexo	Título	suelo	fechalicenciatura	filter_\$
877	1	4	29000	1996	0
878	1	7	22000	1995	0
879	1	7	34500	1995	0
880	1	7	30000	1995	0
881	0	1	35300	1995	0
882	1	4	28500	1995	0
883	1	1	22000	1995	0
884	1	1	18000	1995	0
885	1	4	23000	1995	0
886	0	1	33000	1995	0
887	0	1	24000	1995	0
888	1	4	20000	1995	0
889	1	3	27000	1995	1
890	1	3	26000	1995	1
891	1	1	30000	1995	0
892	1	1	20000	1995	0
893	0	1	25000	1995	0
894	0	1	27500	1995	0
895	0	7	30000	1995	0
896	1	7	27400	1995	0
897	1	1	18000	1995	0

Una vez esto comparamos medias y nos saca la siguiente tabla:

Informe

Salario en euros

Sexo	Media	N	Desviación estándar
Mujer	28683,33	6	4041,988
Hombre	28254,00	50	3846,027
Total	28300,00	56	3831,639

Donde el sueldo medio de las mujeres que estudiaron Ingeniería Informática es de **28683,33 euros** y la de hombres de una media de **28254 euros**.

Ejercicio 3.7: Completa las siguientes frases, atendiendo al fichero 1sitioweb.sav

Solución:

El día que menos páginas se vieron de este sitio fue el **02 de octubre**, concretamente el número de páginas vistas fue **182**. Por otro lado, el día que más páginas se vieron fue el **20 de junio**, concretamente el número de páginas vistas fue **2189**.

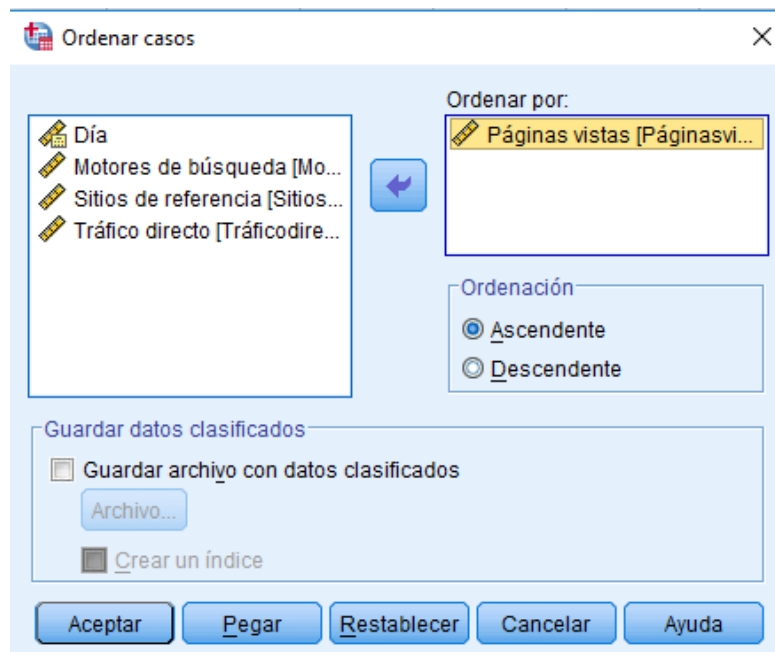
El número total de páginas vistas durante todo el año fue de **338088** páginas. De las cuales, el mes que más páginas se vieron fue el mes de **julio** con un porcentaje del **9,3** por ciento de las páginas vistas en todo el año. Y el mes que menos páginas se vieron fue el mes de **abril** con un porcentaje del **7** por ciento de las páginas vistas en todo el año.

Respecto a las visitas a dicho sitio, el día que menos visitas tuvo este sitio fue el **17 de enero** y el **05 de noviembre**, concretamente el número de visitas fue **1**. Por otro lado, el día que más visitas tuvo fue el **24 de septiembre**, concretamente el número de visitas fue **399**.

El número total de visitas durante ese año fue de **75650** visitas. De las cuales el mes con más visitas fue el mes de **marzo** con un porcentaje del **9,9** por ciento de las visitas de todo el año. Y los meses con menos visitas fueron los meses de **febrero** y **agosto** con un porcentaje, cada uno, del **7,3** por ciento de las visitas de todo el año

Explicación del ejercicio:

En primer lugar, ordenaremos los casos para visualizar bien los datos:



Una vez los tenemos ordenados procedemos a visualizar los datos obtenidos una vez que hemos ordenado y vemos los días que menos páginas y el que más:

	Día	Motoresdebusqueda	Sitiosdereferencia	Tráficodeirecto	Páginasvistas
1	10/02/2010	4	109	5	182
2	01/01/2010	96	10	69	189
3	10/12/2010	53	40	104	200
4	12/17/2010	31	13	188	235
5	03/18/2010	31	28	184	276
6	11/13/2010	141	28	62	279
7	04/22/2010	52	34	129	291
8	05/30/2010	75	168	49	303
9	08/31/2010	23	86	156	317
10	09/11/2010	114	85	33	324

El día que menos páginas visitadas se vieron fue el **10/02/2010** con un total de **182**.

360	10/30/2010	395	193	373	1814
361	12/21/2010	373	233	360	1818
362	08/04/2010	380	345	381	1837
363	03/28/2010	353	292	375	2026
364	01/08/2010	370	376	386	2132
365	05/20/2010	373	380	367	2189

El día que más páginas se vieron fue el **05/20/2010** con un total de **2189**.

Para sacar el número total de páginas vistas comparamos media y obtenemos la suma, al final nos mostrará un total con la suma de todos los valores de las páginas de días.

The screenshot shows the 'Medias' software interface. On the left, there is a list of variables: 'Motores de búsqueda', 'Sitios de referencia...', 'Tráfico directo [Tráfico]', and 'Páginas vistas [Páginas]'. In the center, there is a 'Lista de dependientes:' (List of dependents) box containing 'Páginas vistas [Páginas]'. Below this is a 'Capa 1 de 1' (Layer 1 of 1) section with 'Anterior' (Previous) and 'Siguiente' (Next) buttons. To the right of this is a 'Lista de independientes:' (List of independents) box containing 'Día'. On the far right, there are three buttons: 'Opciones...' (Options...), 'Estilo...' (Style...), and 'Bootstrap...'. At the bottom, there are five buttons: 'Aceptar' (Accept), 'Pegar' (Paste), 'Restablecer' (Reset), 'Cancelar' (Cancel), and 'Ayuda' (Help).

Obteniendo el siguiente resultado:

12/20/2010	1	979
12/21/2010	1	1818
12/22/2010	1	656
12/23/2010	1	922
12/24/2010	1	866
12/25/2010	1	460
12/26/2010	1	1345
12/27/2010	1	793
12/28/2010	1	1471
12/29/2010	1	1316
12/30/2010	1	891
12/31/2010	1	737
Total	365	338088

Por lo que el total de páginas vistas durante todo el año fue de **338088** páginas.

Ahora nos piden que saquemos cuantas páginas hay por mes, para ello necesitaremos extraer del día que es un formato tipo fecha el mes de la siguiente forma:

Nos vamos a calcular variable →

Calcular variable

Variable de destino: = Expresión numérica:

Tipo y etiqueta...

☐ Día
☐ Motores de búsqueda...
☐ Sitios de referencia...
☐ Tráfico directo (Tráfico...)
☐ Páginas vistas [Pág...]
☒ auxmes

XDATE.MONTH(valorfecha). Numérico. Devuelve el mes (un entero entre 1 y 12) a partir de un valor numérico que representa una fecha. El argumento puede ser un número, una variable con formato de fecha o una expresión evaluable a una fecha.

(condición de selección de casos opcional)

Grupo de funciones:
 Aritméticas
 FDA y FDA no centrada
 Conversión
 Fecha/hora actual
 Cálculo de fechas
 Creación de fechas
 Extracción de fechas
 Funciones y variables especiales:
 Xdate.Date
 Xdate.Jday
 Xdate.Mday
 Xdate.Month
 Xdate.Quarter
 Xdate.Tday
 Xdate.Time
 Xdate.Week
 Xdate.Wkday
 Xdate.Year
 Yrmoda

Esto nos va a crear una variable llamada **auxmes** que nos va a dar el día del mes de la fecha:

	Día	Motoresdebusqueda	Sitiosdereferencia	Tráficodeirecto	Páginasvistas	auxmes
1	10/02/2010	4	109	5	182	10,00
2	01/01/2010	96	10	69	189	1,00
3	10/12/2010	53	40	104	200	10,00
4	12/17/2010	31	13	188	235	12,00
5	03/18/2010	31	28	184	276	3,00
6	11/13/2010	141	28	62	279	11,00
7	04/22/2010	52	34	129	291	4,00

Auxmes toma los valores del formato fecha.

Ahora vamos a darle una recodificación para que nos salgan los meses para ello nos vamos a Transformar → Recodificar en distintas variables:

Recodificar en distintas variables

Variable numérica -> Variable de resultado:
auxmes -> mes

Variable de resultado
Nombre:
mes

Recodificar en distintas variables: Valores antiguos y nuevos

Valor antiguo

☒ Valor:

☐ Perdido del sistema

☐ Perdido del sistema o perdido del usuario

☒ Rango:

hasta

☐ Rango, INFERIOR hasta valor:

☐ Rango, valor hasta SUPERIOR:

☒ Todos los demás valores

Valor nuevo

☒ Valor:

☐ Perdido del sistema

☒ Copiar valores antiguos

Antiguo -> Nuevo:

1 -> Enero
2 -> Febrero
3 -> Marzo
4 -> Abril
5 -> Mayo
6 -> Junio
7 -> Julio

☒ Las variables de resultado son cadenas Anchura: 8

☐ Convertir cadenas numéricas en números ('5'->5)

Continuar Cancelar Ayuda

El resultado final sería el siguiente:

	Día	Motoresdebusqueda	Sitiosdereferencia	Tráficodeirecto	Páginasvistas	auxmes	mes
1	10/02/2010	4	109	5	182	10	Octubre
2	01/01/2010	96	10	69	189	1	Enero
3	10/12/2010	53	40	104	200	10	Octubre
4	12/17/2010	31	13	188	235	12	Diciembre
5	03/18/2010	31	28	184	276	3	Marzo
6	11/13/2010	141	28	62	279	11	Noviembre
7	04/22/2010	52	34	129	291	4	Abril

Con esto se nos quedaría todo más claro a la hora de interpretar los resultados.

Ahora para obtener en qué mes se vieron más/menos páginas comparamos medias de mes por páginas vista, de esta forma nos dará la suma de todas páginas vistas por cada mes:

Informe		
Páginas vistas		
mes	N	Suma
Abril	30	23672
Junio	30	26231
Septiembre	30	26673
Febrero	28	26921
Octubre	31	27386
Mayo	31	27912
Diciembre	31	28466
Noviembre	30	28909
Agosto	31	29518
Enero	31	30377
Marzo	31	30492
Julio	31	31531
Total	365	338088

Como podemos observar el número de páginas vistas es menor en el mes de abril con un total de 23672 y el que más en Julio con 31531.

Para saber el porcentaje marcamos la opción de “Suma total de porcentaje”:

Medias: Opciones

Estadísticos:

Mediana

Mediana agrupada

Rango

Varianza

Curtosis

Error estándar de curtos

Asimetría

Error estándar de asim

Media armónica

Media geométrica

Porcentaje del N total

Primero

Último

Máximo

Estadísticos de casilla:

Número de casos

Suma

Porcentaje de la suma total

Estadísticos para la primera capa

☐ Tabla de Anova y eta

☐ Prueba de linealidad

Continuar

Cancelar

Ayuda

mes	N	Suma	% de suma total
Abril	30	23672	7,0%
Junio	30	26231	7,8%
Septiembre	30	26673	7,9%
Febrero	28	26921	8,0%
Octubre	31	27386	8,1%
Mayo	31	27912	8,3%
Diciembre	31	28466	8,4%
Noviembre	30	28909	8,6%
Agosto	31	29518	8,7%
Enero	31	30377	9,0%
Marzo	31	30492	9,0%
Julio	31	31531	9,3%
Total	365	338088	100,0%

Y sacamos que el mes con menor % tiene es Julio con 9.3% y el que menos abril con un 7%

Respecto a las visitas, para obtener que día tiene más y cual menos comparamos media ordenamos ascendentemente y observamos:

Medias

Motores de búsqueda...

Sitios de referencia...

Páginas vistas [Pág...

Páginas vistas [Pág...

auxmes

mes

Lista de dependientes:

Tráfico directo [Tráfico...

Capa 1 de 1

Anterior

Siguiente

Lista de independientes:

Día

Aceptar

Pegar

Restablecer

Cancelar

Ayuda

Informe		
Tráfico directo		
Día	N	Suma
01/17/2010	1	1
11/05/2010	1	1
07/29/2010	1	3
10/02/2010	1	5
12/19/2010	1	6
07/17/2010	1	7
08/27/2010	1	7
11/26/2010	1	0

01/29/2010	1	396
11/20/2010	1	396
04/18/2010	1	399
07/21/2010	1	399
09/24/2010	1	399
Total	365	75650

Como podemos observar 2 días tienen el mismo número de visitas el día **01/17/2010** y el día **11/05/2010** con 1 visita obteniendo menos visitas.

Y los días que más visitas obtuvieron son el **04/18/2010**, **07/21/2010** y el **09/24/2010** con **399** visitas.

El número total de visitas para calcularlo comparamos medias por mes que también nos saldrá el total de las visitas, marcamos porcentaje de la suma porque en el apartado de después nos lo pedirán:

mes	N	Suma	total
Febrero	28	5533	7,3%
Agosto	31	5538	7,3%
Abril	30	5759	7,6%
Septiembre	30	5784	7,6%
Mayo	31	5832	7,7%
Enero	31	6327	8,4%
Noviembre	30	6473	8,6%
Diciembre	31	6591	8,7%
Octubre	31	6722	8,9%
Junio	30	6733	8,9%
Julio	31	6841	9,0%
Marzo	31	7517	9,9%
Total	365	75650	100,0%

El número total de visitas de este año fueron **75650** visitas. El mes con más visitas fue **marzo** con un porcentaje de **9,9%**. Los meses con menos visitas fueron los meses de **febrero** y **agosto** con un porcentaje del **7,3%**.