**Nombre**: Javier Rivilla Arredondo

**DNI**: 53247378D

**Email**: [javin.javn@gmail.com](mailto:javin.javn@gmail.com)

**Práctica 3: ESTADÍSTICA DESCRIPTIVA**

***Ejercicio 3.1:*** Resuelve los siguientes ejercicios atendiendo al fichero 1ordenadores.sav.

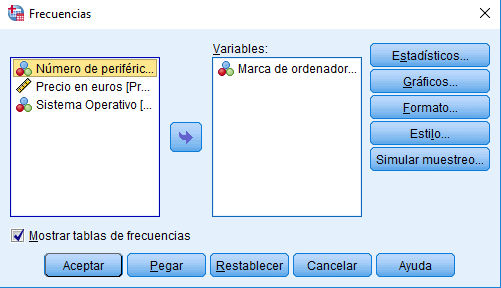
Construye la tabla de frecuencia de la variable que indica la marca de ordenador, adjúntala a la práctica y

explica lo que observas.

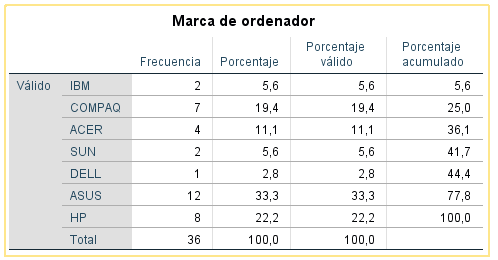
***Solución:***

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:



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…

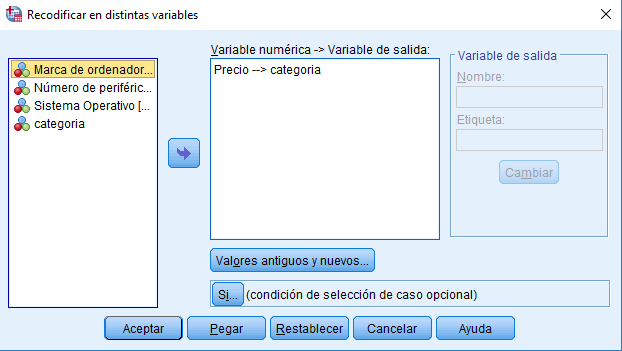
Obtén el gráfico de sectores agrupando los precios de los ordenadores en tres categorías: 1- menos de 800

euros, 2- entre 800 y 1200 euros, 3- más de 1200 euros. Incluye los porcentajes en el gráfico, adjúntalo en la

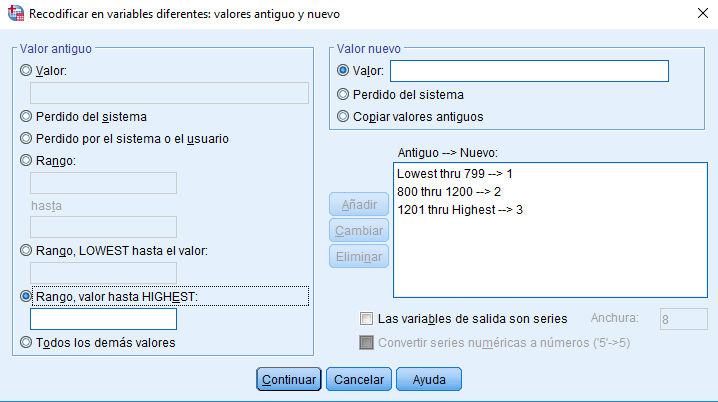
práctica y explica lo que observas.

***Solución:***

Primero agrupamos los precios de los ordenadores en una variable que se llame **categoria1** 🡪 **Recodificamos en distintas variables** seleccionamos la variable que necesitemos en este caso el precio.



**🡪 Valores antiguos y nuevos…**

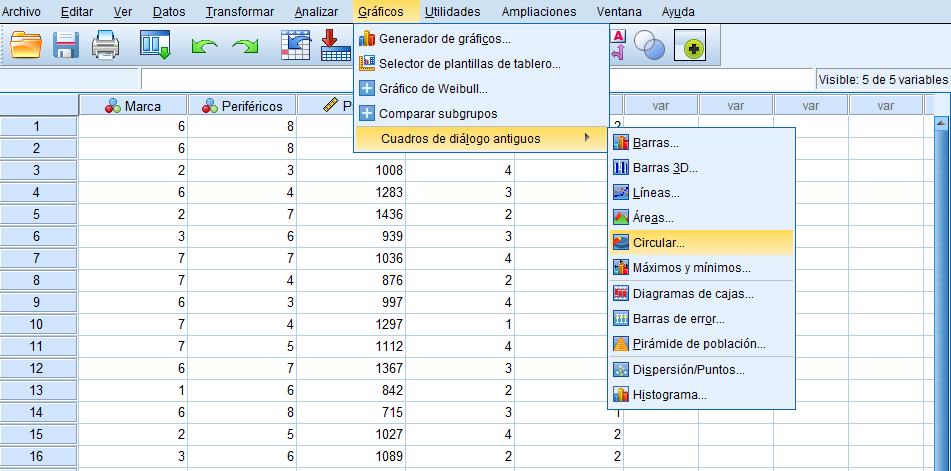


En rango ponemos las carácteristicas de cada categoría:

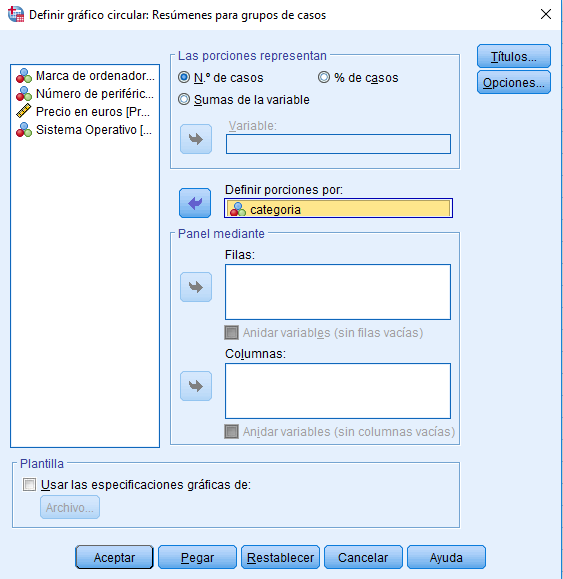
- **Categoría 1**: Menores que 799.

- **Categoria 2**: Rango de 800 y 1200.

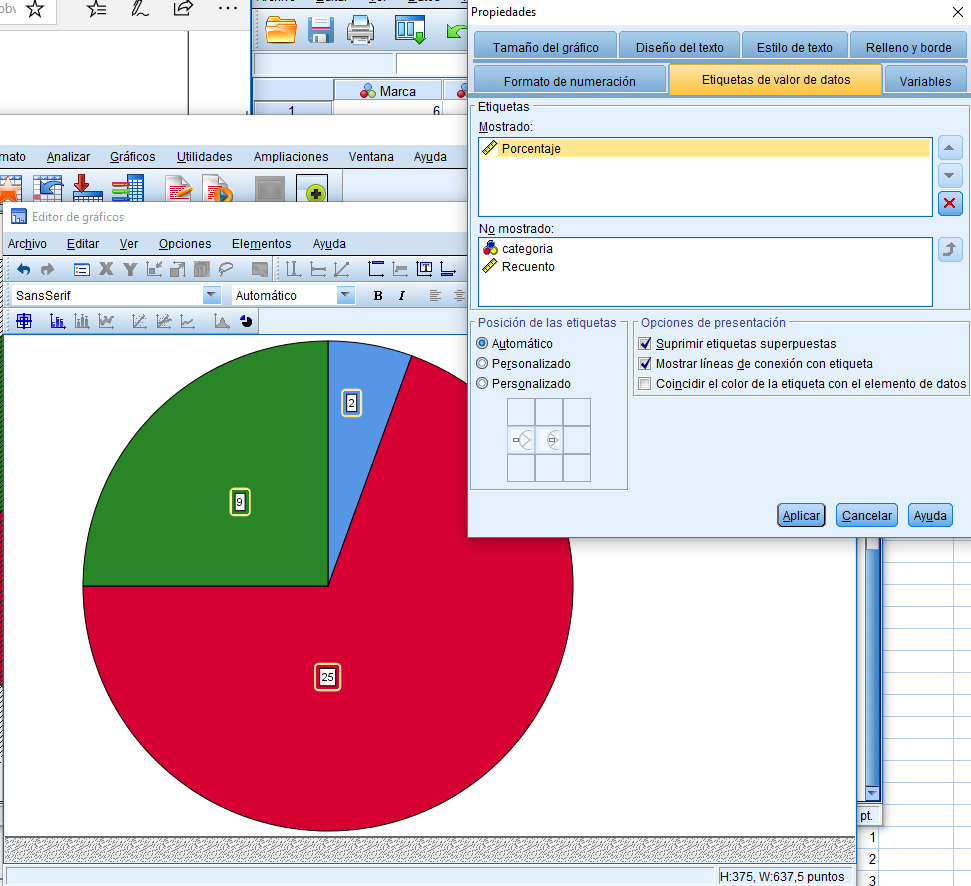
- **Categoria 3**: Más de 1201 (Porque se incluye el número, y son mayores de 1200).



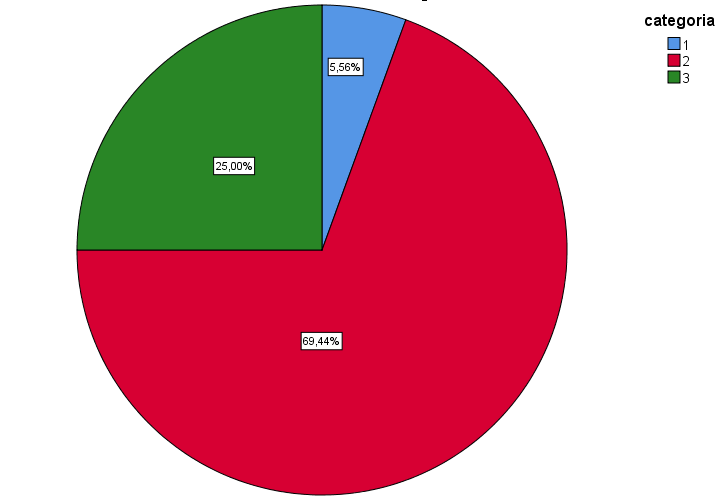
Una vez hemos sacado las categorías creamos el grafico de sectores, para ello nos vamos a **Gráficos 🡪 Cuadros de diálogos antiguos 🡪 Circular…**



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

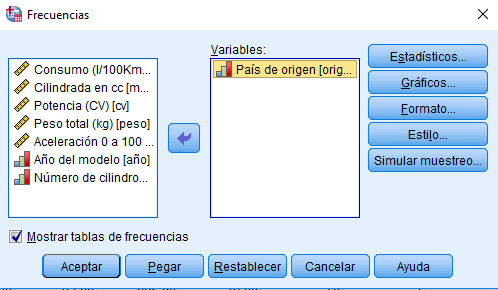
fabricados en Europa y el **19,2** por ciento en Japón.

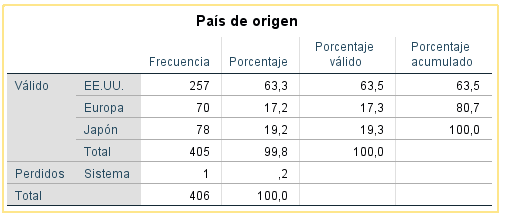
La aceleración media de dichos coches es de **15,4951** segundos de 0 a 100 km/h. Los coches con menor

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 tiempomedio 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.

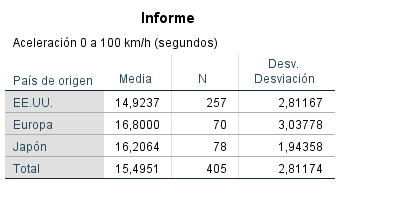
**Coches fabricados:**

****

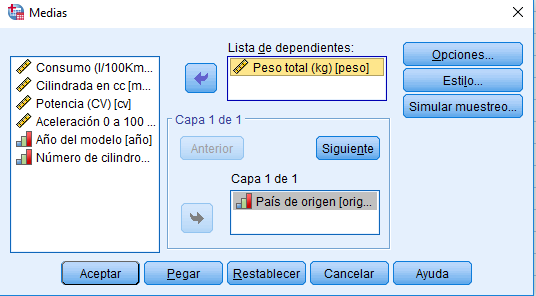


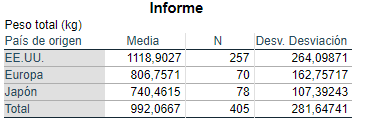
**Aceleración media:**

****

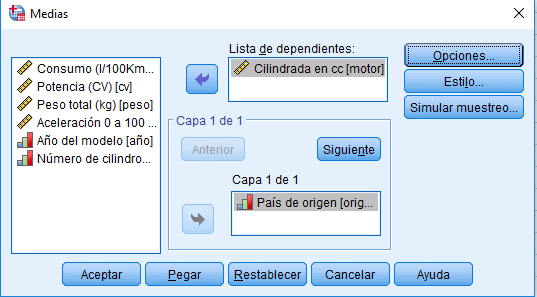
****

**Peso medio:**

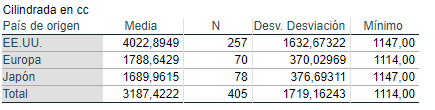
****



**Cilindrada:**

****





***Ejercicio 3.3:*** Contesta a las siguientes preguntas, basándote en el cálculo de percentiles y atendiendo al

fichero 1ordenadores.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.

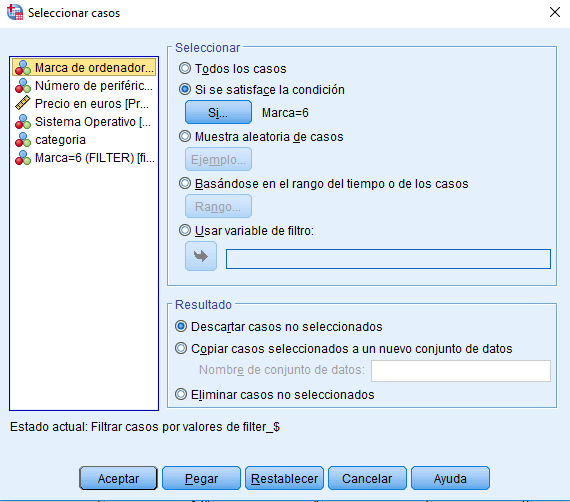
Explicación de la resolución del ejercicio:

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**.

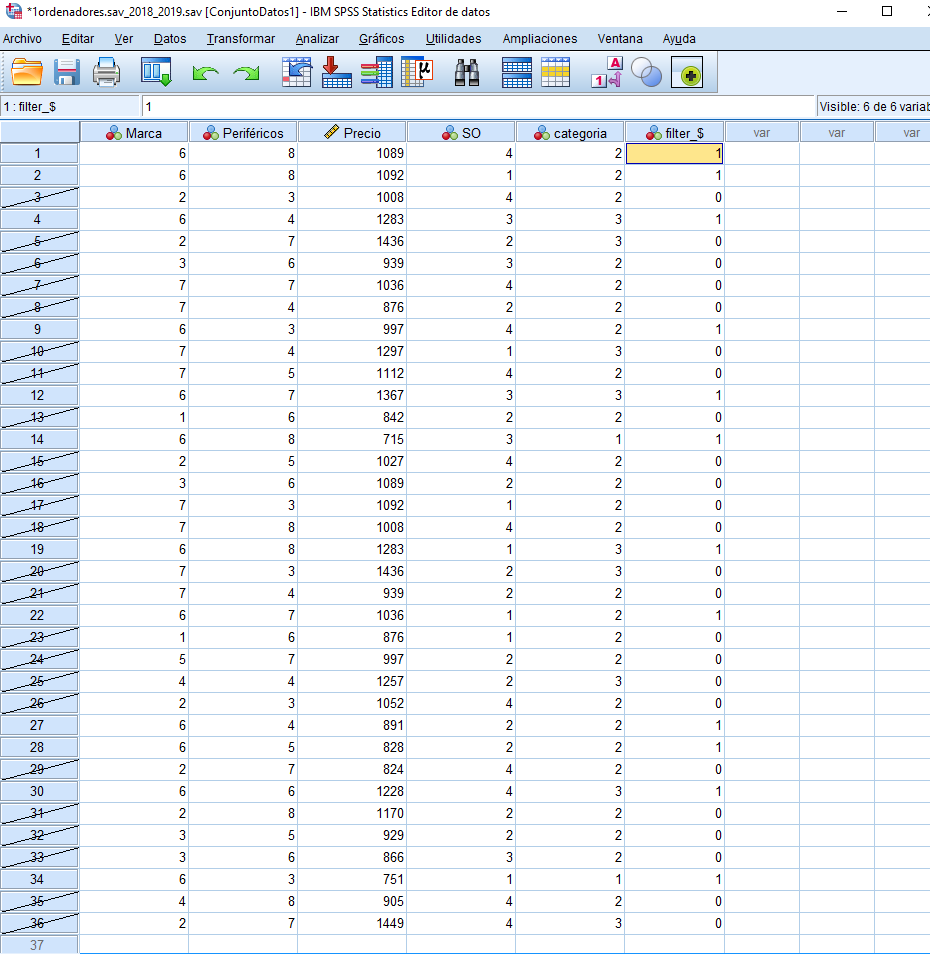


Y en el siguiente apartado nos pide **el 40% de los ordenadores que cuestan menos de**, por lo que tendremos que utilizar el **P40** ya que el percentil nos dará el valor menor o igual a ese porcentaje.

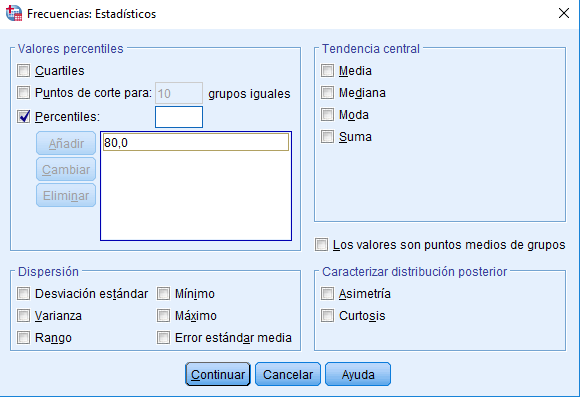
Como bien sabemos, nos piden la marca ASUS para ello tendremos que **filtrar** los casos en los que la **marca sea igual a 6** de esta forma nos dará los ordenadores que corresponden solamente a dicha opción:



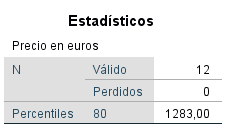
Como observamos en la siguiente foto nos saldrán tachados los ordenadores cuya marca no es igual a 6:



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



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.



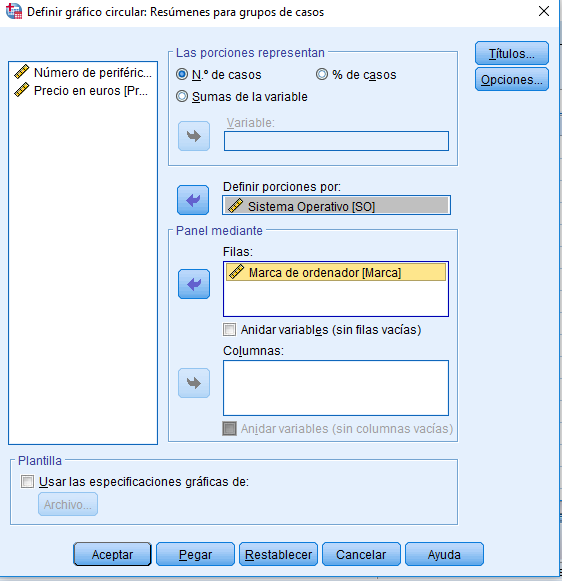
El precio por lo tanto será **1283 euros** corresponde al **percentil 80**, es decir, nos va a dar el precio que cuesta ese **precio o menos** de ese precio.

***Ejercicio 3.4:*** Atendiendo al fichero 1ordenadores.sav, obtén los gráficos de sectores que expliquen cómo se

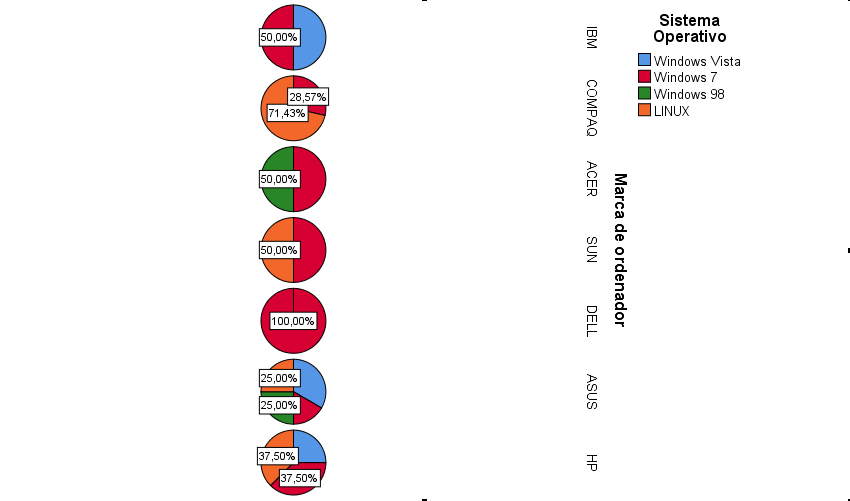
distribuyen los porcentajes de los sistemas operativos instalados atendiendo a cada marca de ordenador.

Explica lo que observas.

***Solución:***

******

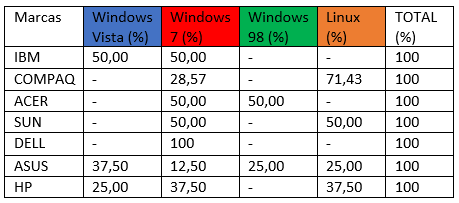
Como nos pone atendiendo a cada marca de ordenador por sistema operativo seleccionamos a la hora de definir nuestro gráfico circular dos opciones 🡪 por un lado las porciones que estarán compuestas por los sistemas operativos por marca ya que cada marca tiene x sistemas operativos, y por otro lado las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



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.

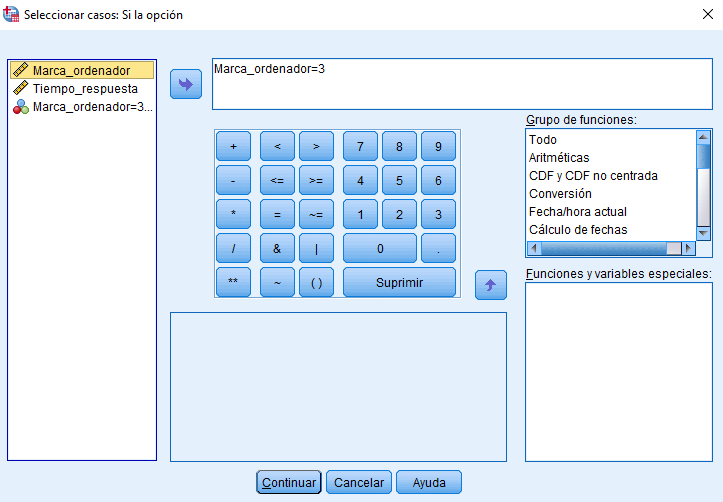
***Ejercicio 3.5:*** Realiza las siguientes cuestiones relativas al fichero 1tiemporespuesta.sav.

Calcula el gráfico caja del tiempo de respuesta de los ordenadores de la marca C con ayuda del SPSS.

Interpreta los resultados.

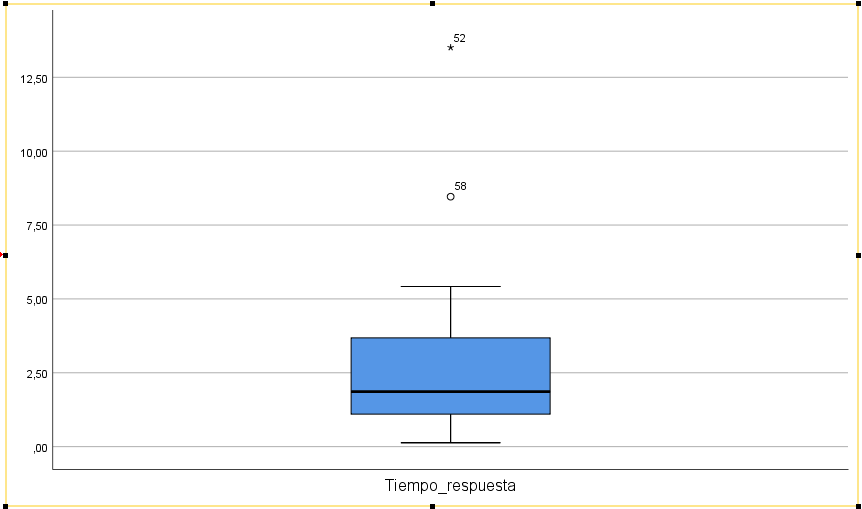
***Solución:***

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 8,50 que corresponde al valor “o”.

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

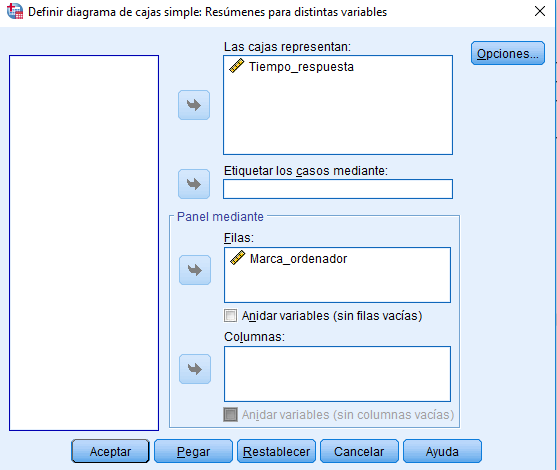
En este diagrama de cajas obtenemos valores atípicos extremos en este caso “\*” nos indica que es un valor atípico extremo y el “o” que simplemente es un valor atípico.

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

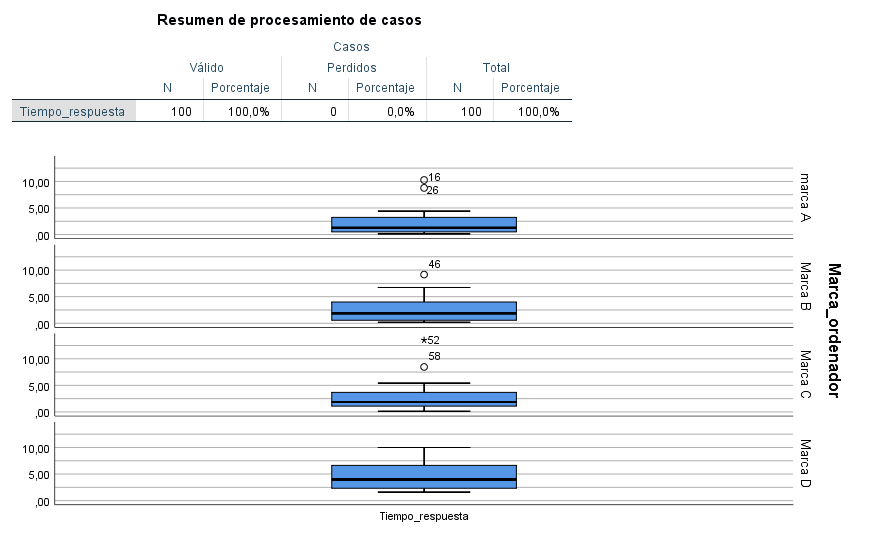
Adjúntalos e interpreta los resultados.

***Solución:***

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



Seleccionamos tanto la marca para las filas como el tiempo respuesta para lo que queremos representar y podemos observar como se muestra una caja por cada marca:



En este caso la marca A tiene dos valores atípicos, que no son extremos, en comparación a la marca C como antes hemos explicado tenia 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.

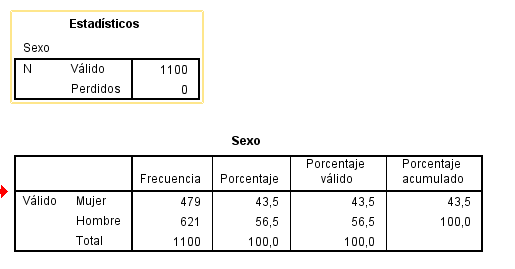
El sueldo de las personas muestreadas varía de **7200** a **65500** euros, con una media de **26053,11** euros. Las personas con menor sueldo de media son los que finalizaron la titulación en el año **2005**. Las personas con mayor sueldo de media son las que estudiaron la titulación de **I. Industriales**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

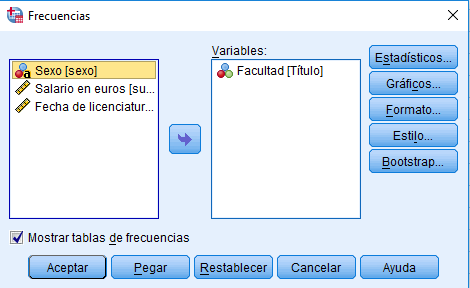
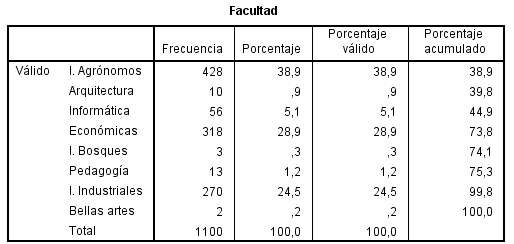
Para saber cuántas personas se han muestreado lo que haremos será sacar por sexo una tabla de estadísticos de esta forma obtendremos el número total de personas que hemos muestreado:





N = 1100 personas que hemos muestreo.

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

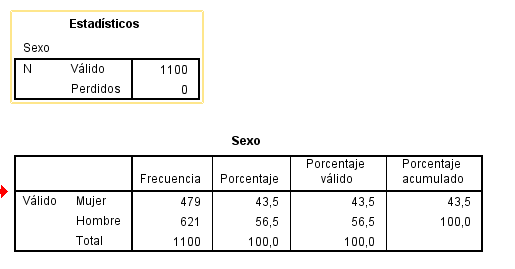
:

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:

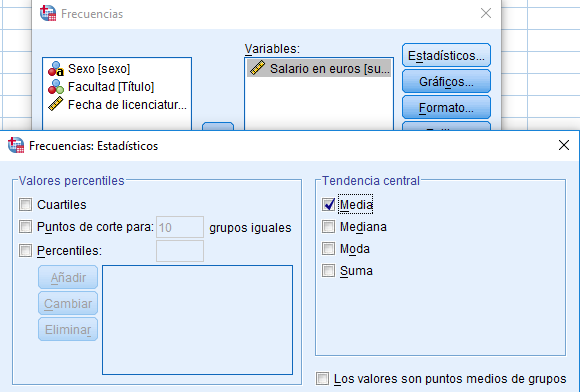
**100 – 38,9 - 24,5 – 28,9 = 7,7% restante**

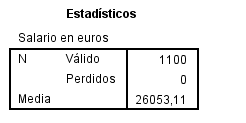
El porcentaje de mujeres y hombre lo hemos sacado anteriormente con esto:



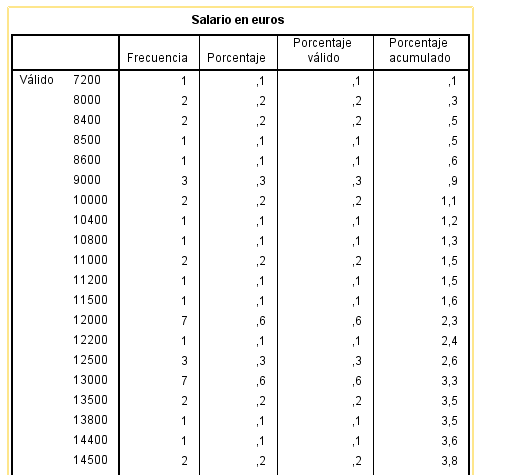
Es una tabla de estadísticos de frecuencia por sexo.

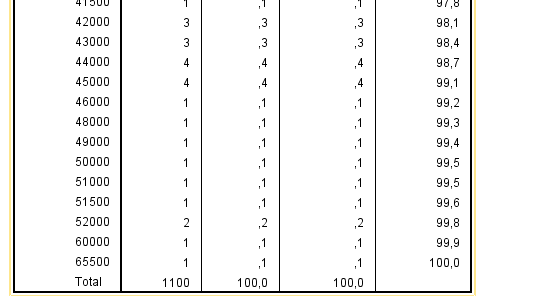
Para sacar el sueldo de las personas haremos lo mismo, pero de la variable sueldo:

La media del sueldo es el siguiente:

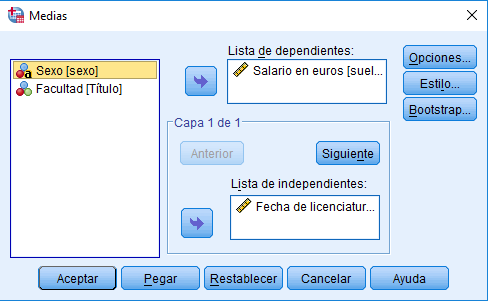


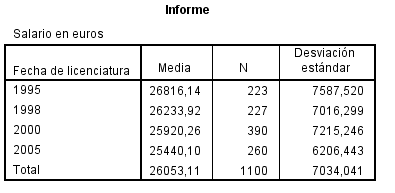
El salario varía de inicio a fin es decir de los valores **7200** al valor **65500.**



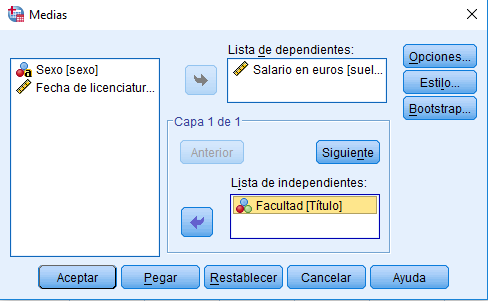


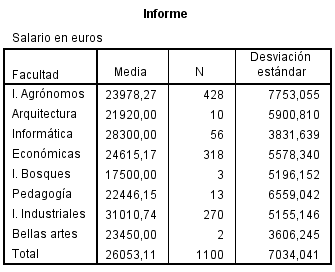
Las personas con menos sueldo de media por titulación deberemos comparar medias entre el salario y la licenciatura para sacar el año que menos sueldo tiene:





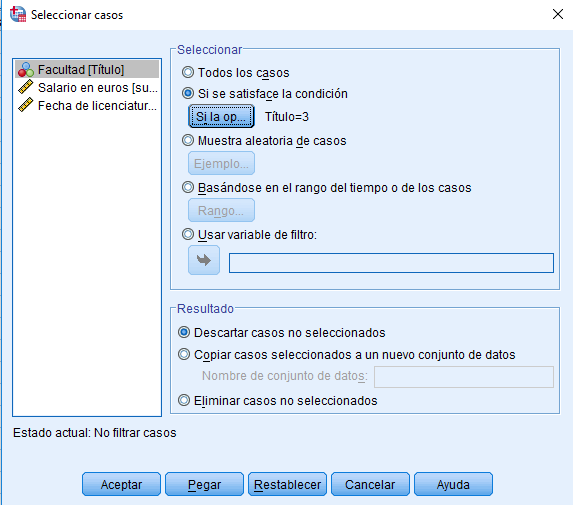
Ahora nos piden el salario de las personas que estudiaron una titulación para ello haremos lo mismo pero titulación por salario:



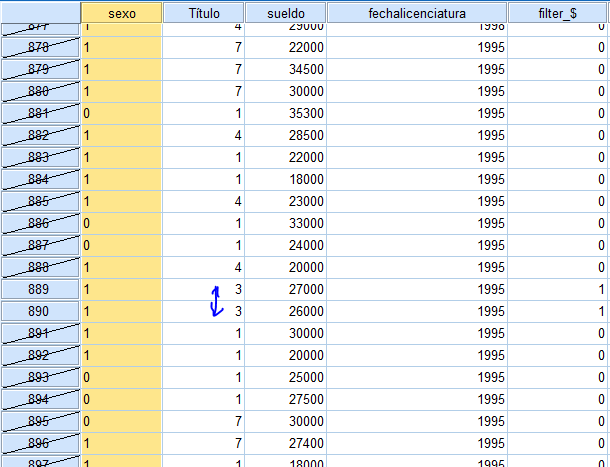


Ahora nos piden mujeres que estudiaban ingeniería informática el sueldo medio, para ello tendremos que filtrar las mujeres que corresponden a esa titulación y comparar las medias de salario por sexo:

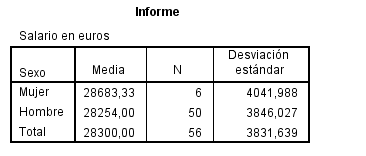
Filtramos:



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



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



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

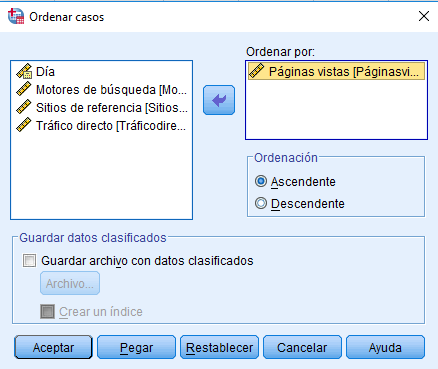
visitas fue el mes de **marzo** con un porcentaje del **9,9** por ciento de las visitas de todo el año. Y

los meses con menos visitas fueron los meses de **febrero** y **agosto** con un porcentaje, cada uno, del **7,3**

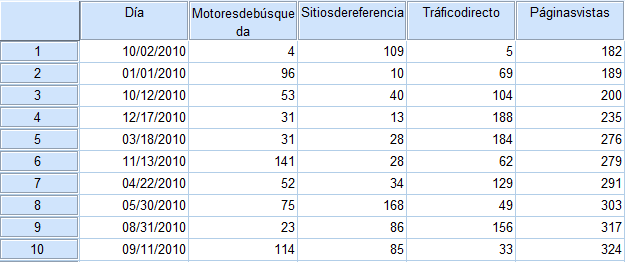
por ciento de las visitas de todo el año

**Explicación del ejercicio:**

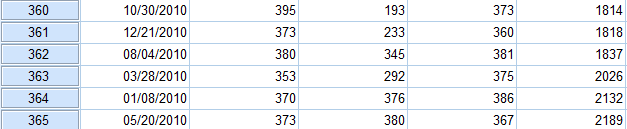
En primer lugar, ordenaremos los casos para visualizar bien los datos:



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:

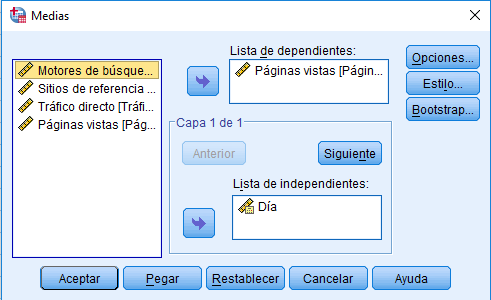


El día que menos páginas visitadas se vieron fue el **10/02/2010** con un total de **182**.

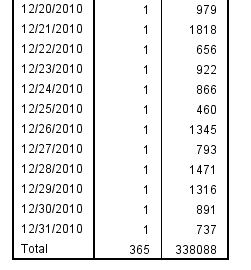


El día que más páginas se vieron fue el **05/20/2010** con un total de **2189**.

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



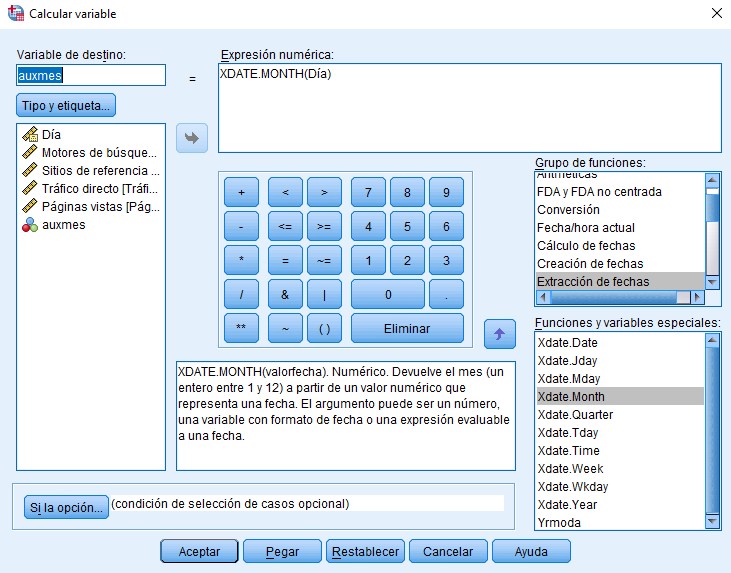
Obteniendo el siguiente resultado:



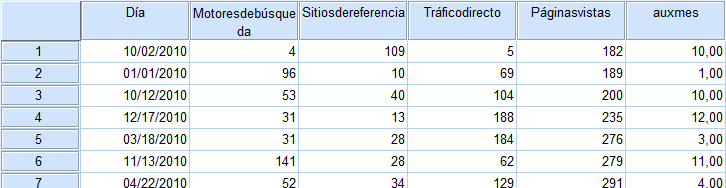
Por lo que el total de páginas vistas durante todo el año fue de **338088** páginas.

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 →

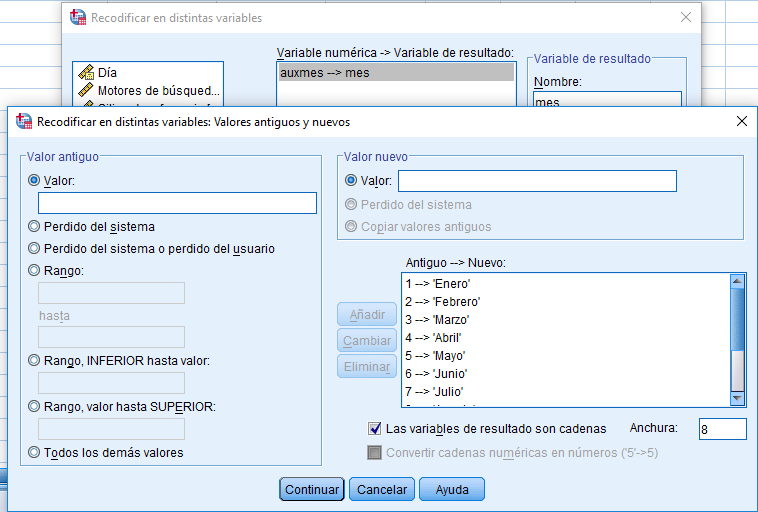


Esto nos va a crear una variable llamada **auxmes** que nos va a dar el día del mes de la fecha:

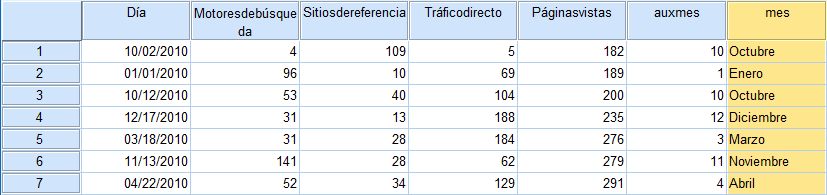


Auxmes toma los valores del formato fecha.

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

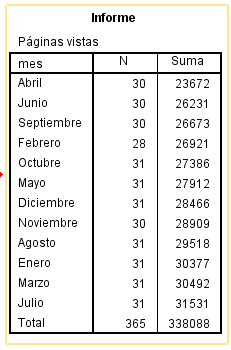


El resultado final sería el siguiente:



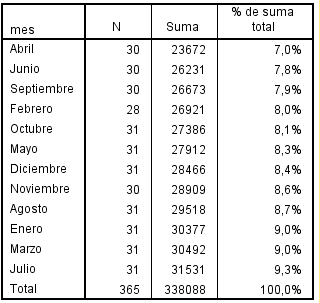
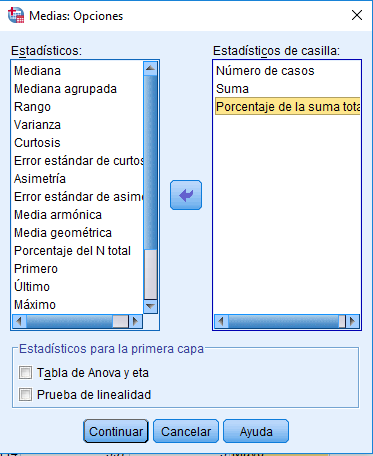
Con esto se nos quedaría todo más claro a la hora de interpretar los resultados.

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:



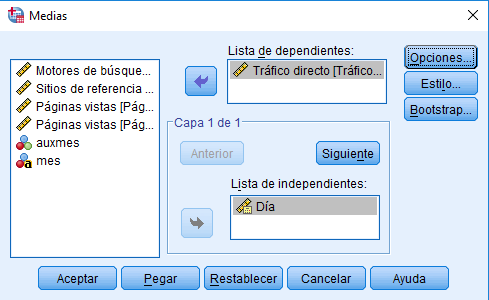
Como podemos observar el número de páginas vistas es menor en el mes de abril con un total de 23672 y el que más en Julio con 31531.

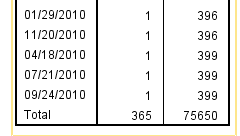
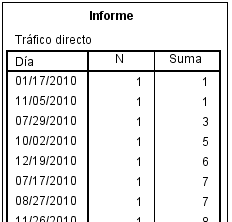
Para saber el porcentaje marcamos la opción de “Suma total de porcentaje”:



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:

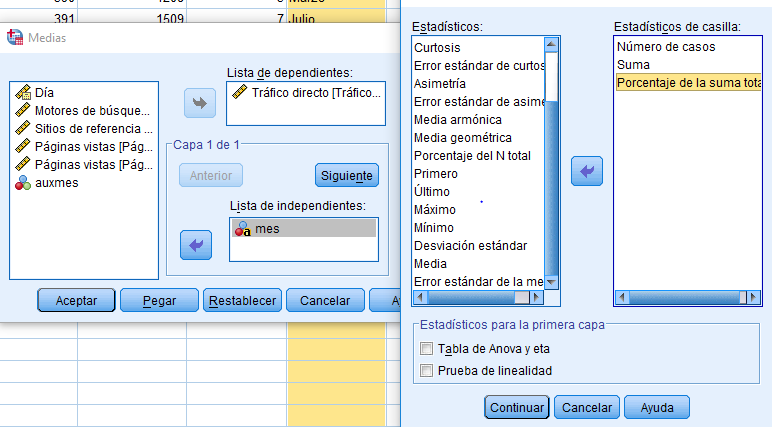


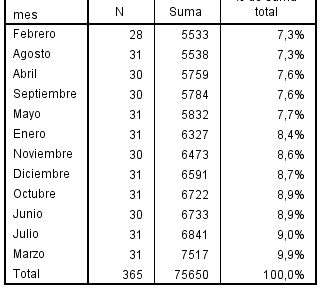


Como podemos observar 2 días tienen el mismo número de visitas el día **01/17/2010** y el día **11/05/2010** con 1 visita obteniendo menos visitas.

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:





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%.**