



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



BIG DATA: Aplicaciones a Inteligencia Artificial, Ciberseguridad y Knowledge discovery:

*¿Y si lo analizamos?*

# I. INTRODUCCIÓN AL MANEJO DE STATGRAPHICS CENTURIÓN

## 1 ARRANQUE DEL PROGRAMA

El arranque del programa se realiza abriendo la carpeta DEIOAC del escritorio y pulsando dos veces el botón izquierdo del ratón sobre el icono.

## 2 FINALIZACIÓN DEL PROGRAMA

Para finalizar el programa pueden utilizarse cualquiera de las dos siguientes opciones: en el menú principal se pulsa *File* y luego *Exit STATGRAPHICS*, o bien se pulsa el botón de cierre de la ventana del programa (☒), situado en la esquina superior derecha.

## 3 VENTANA DE EJECUCIÓN DEL STATGRAPHICS

Como todos los programas de Windows, el STATGRAPHICS se ejecuta en una ventana, y los resultados de las operaciones se presentan también en ventanas. El conjunto de los datos y resultados obtenidos reciben el nombre de *StatFolio*, y pueden así almacenarse y recuperarse fácilmente.

## 4 FICHEROS DE DATOS

El STATGRAPHICS permite realizar las operaciones básicas habituales con un fichero: crear, guardar, recuperar y modificar.

### 4.1 CREAR UN FICHERO DE DATOS

Nada más abrir el programa se presenta una hoja de datos en blanco, de aspecto muy similar a las hojas de Excel. Si se estaba trabajando con el programa, y ya existe un fichero abierto en la hoja actual, es necesario cerrarlo antes de crear el nuevo en esa misma hoja, utilizando la orden *File/Close/Close data file*.

En la hoja de datos se presenta una primera fila en la que se crean las variables, y una primera columna en la que aparece la numeración de las filas.

Si se selecciona una columna pulsando sobre su nombre (por ejemplo *Col\_1*) con el botón izquierdo del ratón, y se pulsa luego el botón derecho, se abre un menú de ventana que permite trabajar con las variables. Entre otras opciones hay que destacar la de modificar la columna (*Modify column*) y generar datos (*Generate data*).

- La opción modificar columna (*Modify column*) permite dar nombre a la variable (*name*), comentar su contenido (*comment*), así como establecer su tamaño (*Width*) y tipo (*Type*).
- La opción generar datos (*Generate data*) permite utilizar diversas funciones matemáticas **para generar nuevas variables a partir de las ya existentes.**

## 4.2 GUARDAR UN FICHERO DE DATOS

Una vez creadas todas las variables e introducidos los datos, se debe guardar el fichero de datos mediante la orden *File/Save as/Save data file as*, y en la ventana de guardar fichero (habitual de windows), se selecciona el subdirectorio y se le da nombre al fichero. Si no es la primera vez que va a ser guardado el fichero, la orden es *File/Save/Save data file*.

El formato en que se guardan los datos es propio del STATGRAPHICS, y tienen extensión *sgd*, *sf3* o *sf*.

## 4.3 RECUPERAR UN FICHERO DE DATOS

Para abrir un fichero de datos del STATGRAPHICS se realiza la siguiente operación: *File/Open/Open data source*, se selecciona el tipo de fichero, y en la ventana habitual de windows se selecciona el subdirectorio donde se encuentra el fichero, se busca y se abre.

## 4.4 FICHEROS DE ANÁLISIS

Los análisis realizados pueden guardarse en forma de *StatFolio* para ser recuperados más tarde. Para guardar los análisis se utilizan las siguientes ordenes:

- Si es la primera vez se utiliza *File/Save as/Save StatFolio as* y aparece la ventana habitual de windows para guardar un fichero.
- Si no es la primera vez entonces se selecciona *File/Save/Save StatFolio*

Para recuperar un *StatFolio* se debe utilizar la siguiente opción *File/Open/Open StatFolio*, y aparece la ventana habitual de windows para recuperar un fichero. El fichero en cuestión tendrá extensión *sgp*.

## 4.5 ANÁLISIS DE LOS DATOS

Los análisis estadísticos disponibles en el STATGRAPHICS aparecen en la línea de menú, agrupados de forma conveniente, según diversas categorías.

Una vez seleccionado el tipo de análisis, el programa presenta una ventana de introducción de datos en la que se debe poner el nombre (s) de la variable(s) a estudiar. Tras pulsar *[OK]* el programa realiza el cálculo deseado, aunque es posible que no lo muestre por pantalla. Dado que el análisis solicitado puede incluir resultados numéricos y gráficos, es necesario indicar al programa que deseamos conocerlos.

Existen cuatro iconos en la esquina superior derecha de la ventana de resultados del análisis que permiten ciertas acciones con la variable y los cálculos deseados:



- El primer icono permite ir (volver) a la ventana de introducción de variables para modificar las opciones del análisis de la variable.
- El segundo icono, si está disponible, permite elegir entre cierto número de opciones para transformar la variable.

- El tercer icono permite seleccionar los resultados numéricos y gráficos del análisis deseado que se calcularán y mostrarán por pantalla.
- El cuarto icono, cuando esté disponible, permite modificar ciertos aspectos del gráfico representado.
- El quinto y último icono, con la imagen del disco, permite guardar en el fichero de datos (en la hoja que se desee), algunos de los resultados numéricos del análisis realizado.

## II. ANÁLISIS DE DATOS CON STATGRAPHICS CENTURION. EJEMPLO

A partir de un fichero con datos históricos sobre teléfonos móviles desde el año 2007 hasta la actualidad (2022), podría interesarnos plantearnos diferentes cuestiones sobre la evolución de varias características de este tipo de producto.

En el fichero se puede observar que de cada *smartphone* se han registrado las características o variables siguientes:

AÑO	Año de compra
SOPER	Sistema operativo
CRAM	Tamaño memoria interna
TRAM	Tipo de memoria RAM
PESO	Peso del terminal en gramos
TDISP	Tamaño de la pantalla en pulgadas
RDISP	Resolución de la pantalla
CTRAS	Resolución cámara trasera
CDEL	Resolución cámara delantera

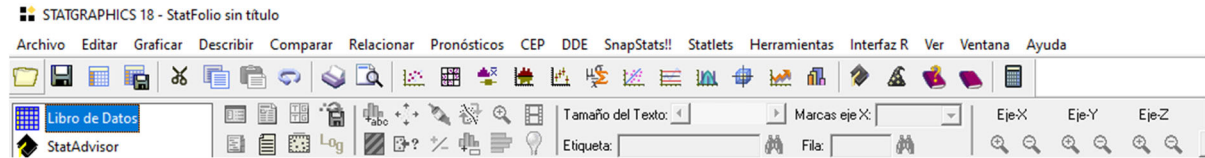
Si alguna de estas características no está disponible en alguno de los móviles registrados, el campo correspondiente en el fichero aparece en blanco. Se trataría de un “dato faltante”.

	AÑO	AÑO_T	SOPER	CRAM	TRAM	PESO	TDISP	RDISP	CTRAS	CDEL
22	2015	2015	IOS	2 GB	LPDDR4	143	4,7	1334x750	12MP	5MP
23	2015	2015	IOS	2 GB	LPDDR4	192	5,5	1920x1080	12MP	5MP
24	2015	2015	ANDROID	3 GB	LPDDR4	138	5,1	2560x1440	16MP	5MP
25	2015	2015	ANDROID	3 GB	LPDDR4	132	5,1	2560x1440	16MP	5MP
26	2015	2015	ANDROID	4 GB	LPDDR4	153	5,7	2560x1440	16MP	5MP
27	2015	2015	ANDROID	4 GB	LPDDR4	171	5,7	2560x1440	16MP	5MP
28	2015	2015	ANDROID	3 GB	LPDDR4	154	5,2	1920x1080	23MP	5MP
29	2014	2014	ANDROID	2 GB		160	5	1920x1080	4MP	5MP
30	2014	2014	ANDROID	3 GB		149	5,5	2560x1440	13MP	2, 1MP
31	2014	2014	IOS	1 GB	LPDDR3	129	4,7	1334x750	8MP	1, 2MP
32	2014	2014	IOS	1 GB	LPDDR3	172	5,5	1920x1080	8MP	1, 2MP
33	2014	2014	ANDROID	2 GB	LPDDR3	145	5,1	1920x1080	16MP	16MP
34	2014	2014	ANDROID	3 GB	LPDDR3	176	5,7	2560x1440	16MP	3, 7MP
35	2014	2014	ANDROID	3 GB	LPDDR3	152	5,15	BRAVIA	7MP	
36	2013	2013	ANDROID	2 GB		143	4,7	1080p	4MP	
37	2013	2013	ANDROID	2GB	LPDDR3	130	4,95	1920x1080	8MP	
38	2013	2013	IOS	1 GB	LPDDR3	112	4	1136x640	8MP	1, 2MP
39	2013	2013	BLACKBERRY	2 GB		137.5	4.2	1280x768	8MP	2MP

A partir de los datos registrados, por ejemplo, podría interesarnos conocer:

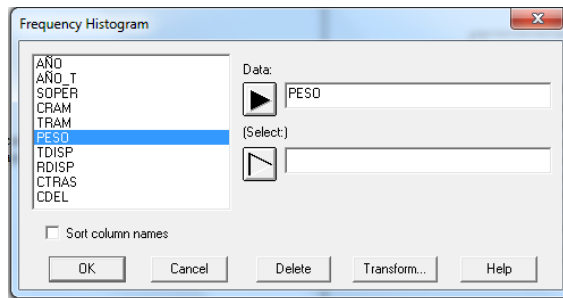
¿Cuál ha sido el peso de los smartphones en este periodo?

Podemos representar un histograma de frecuencias para la característica o variable PESO. Para ello en Statgraphics seguimos la secuencia de opciones:

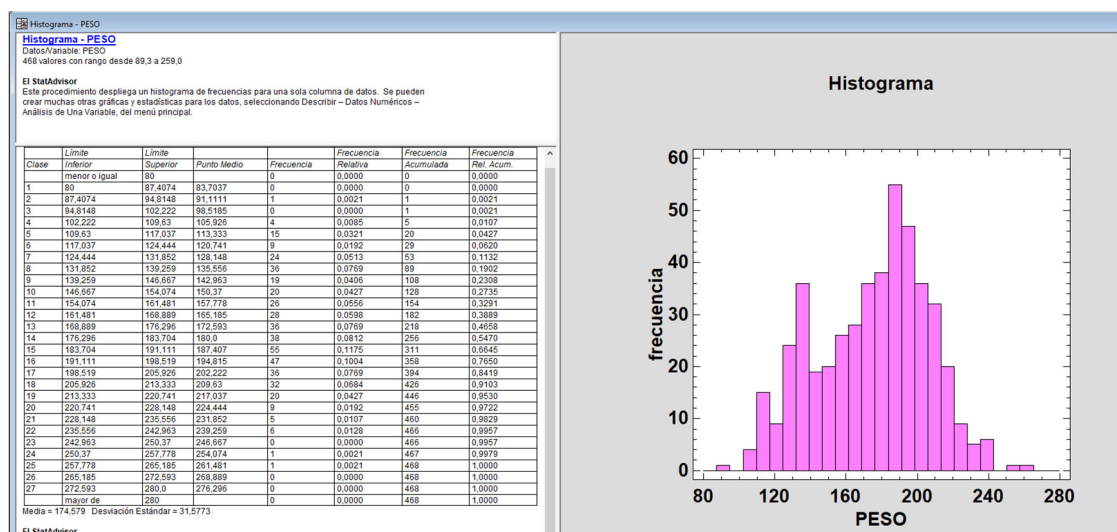


Plot → Exploratory Plots → Frequency Histogram

Indicamos en el campo Data el nombre de la variable que contiene los datos, PESO en este caso.

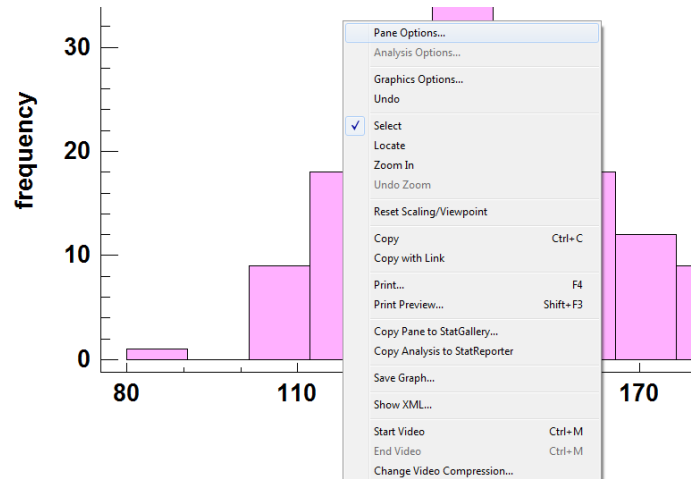


Tras seleccionar OK, indicamos que esta variable contiene datos de tipo continuo y aceptamos el resto de opciones por defecto. La salida proporcionada por el programa está dividida en tres zonas o paneles:

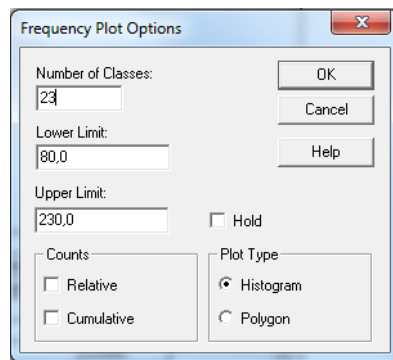


El gráfico representado en la parte derecha es un **Histograma de frecuencias**. Se trata de una representación descriptiva muy utilizada en el análisis estadístico de datos.

Haciendo doble click con el ratón sobre el histograma, podemos variar el aspecto del histograma. Para ello mediante el botón derecho del ratón, seleccionar *Pane Options* (*Opciones de ventana*).



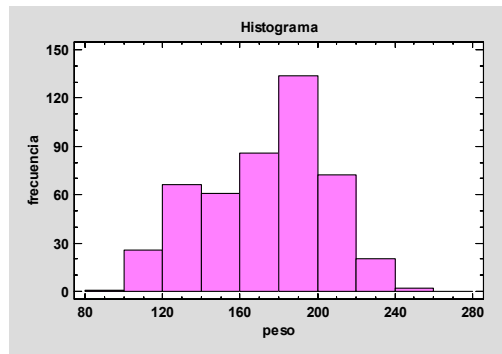
Tras seleccionar *Pane Options* aparece la siguiente ventana:



Desde la ventana *Frequency Plot Options* podemos indicar:

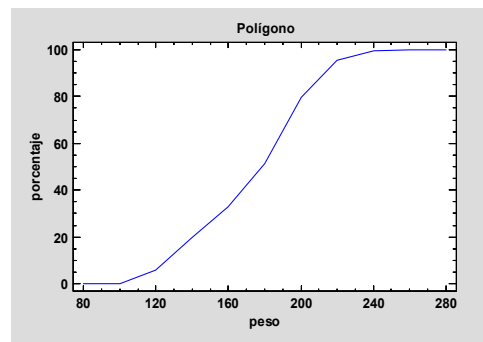
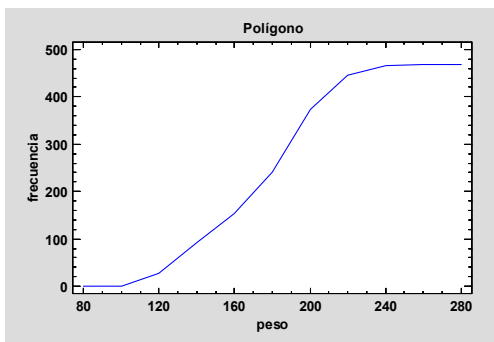
- El número de tramos o barras del histograma, por ejemplo, cambiar 23 (valor por defecto) por 10.
- Límite inferior y superior de la representación.
- Tipo de frecuencia considerada en el eje Y, por defecto, absoluta.
- Tipo de representación: barras o polígono.

El nuevo histograma de frecuencias es el siguiente:



**¿Qué información podemos extraer de esta representación?**

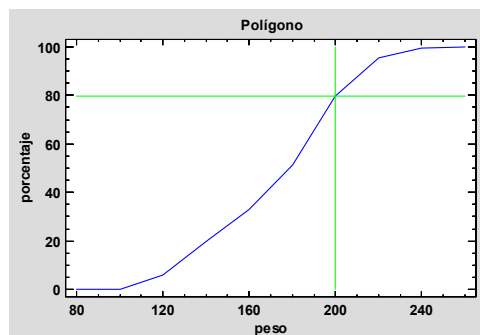
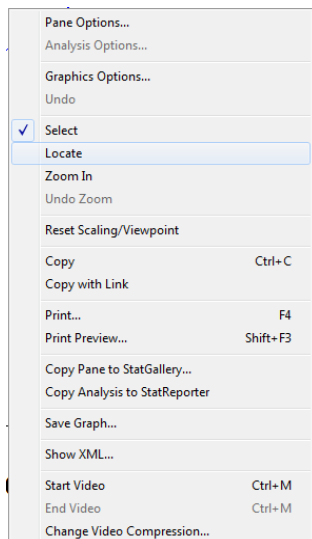
Cambiando las frecuencias absolutas por frecuencias acumuladas (absolutas o relativas) y las barras por una línea quebrada (polígono) obtenemos los siguientes diagramas de frecuencias acumuladas:



A partir de la información proporcionada por los gráficos es posible contestar a la pregunta:

**¿Qué porcentaje de los *smartphones* analizados tienen un peso no superior a 200 gramos?**

Para responder a esta pregunta puede resultar útil la opción *Locate* del menú de ventana desplegable con el botón derecho del ratón.



También se puede cambiar la configuración del propio gráfico mediante el botón derecho del ratón, seleccionando *Graphics Options*. Podemos especificar:

- El color de relleno de las barras.
- Tipo de fuente y tamaño para el texto del gráfico.
- Título del gráfico.
- Texto y escalas de los ejes X e Y
- .....

La información proporcionada a partir del gráfico puede verse también de forma numérica a partir de la **Tabla de Frecuencias**. Por ejemplo, la tabla de frecuencias obtenida para 10 tramos es la siguiente:

Tabla de Frecuencias para peso

Clase	Limite Inferior	Limite Superior	Punto Medio	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. Acum.
	menor o igual	80		0	0,0000	0	0,0000
1	80	100,0	90,0	1	0,0021	1	0,0021
2	100	120,0	110,0	26	0,0556	27	0,0577
3	120	140,0	130,0	66	0,1410	93	0,1987
4	140	160,0	150,0	61	0,1303	154	0,3291
5	160	180,0	170,0	86	0,1838	240	0,5128
6	180	200,0	190,0	134	0,2863	374	0,7991
7	200	220,0	210,0	72	0,1538	446	0,9530
8	220	240,0	230,0	20	0,0427	466	0,9957
9	240	260,0	250,0	2	0,0043	468	1,0000
10	260	280,0	270,0	0	0,0000	468	1,0000
	mayor de	280		0	0,0000	468	1,0000

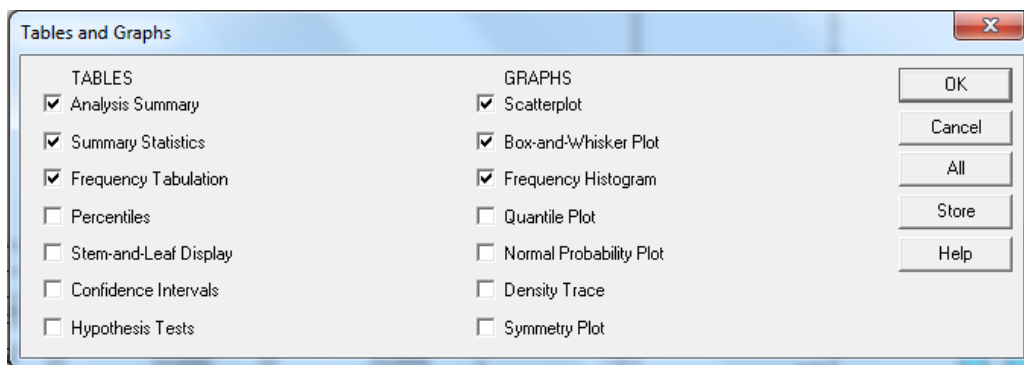
Media = 174,579 Desviación Estándar = 31,5773

**¿Qué información proporciona la tabla de frecuencias para PESO?**

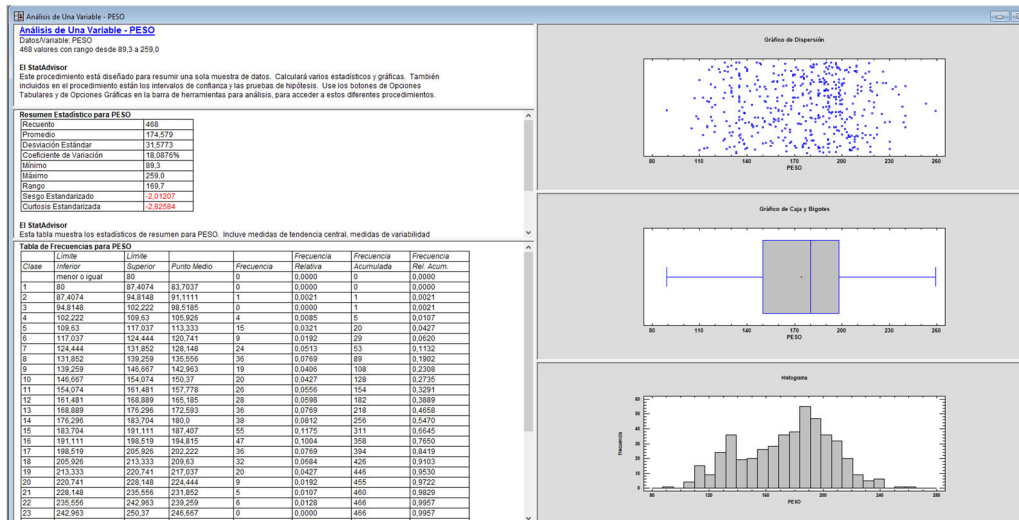
Quizá podamos completarla con el cálculo de parámetros numéricos, como es el caso de la media, varianza, desviación típica, mediana..... Para ello seguimos la secuencia:

*Describe → Numeric Data → One Variable Analysis.....*

El programa solicita el nombre de la variable que contiene los datos. Introducimos PESO en el campo *Data*. En el siguiente menú, marcamos las opciones siguientes:

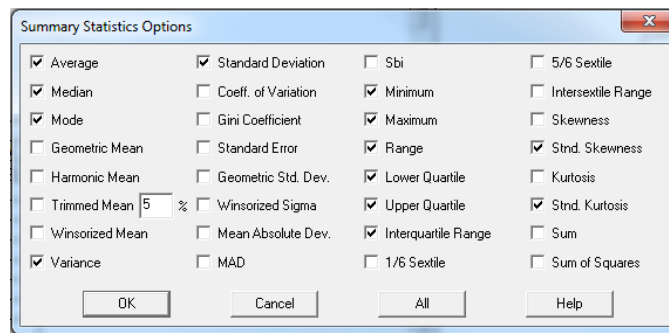


La información proporcionada por el programa está organizada en una serie de paneles:



Los paneles de la izquierda proporcionan información numérica, mientras que los de la derecha muestran gráficos.

En la parte izquierda, el panel *Summary Statistics*, proporciona el valor de los parámetros descriptivos: media, moda, mediana, cuartiles... para la variable analizada. Para conocer la lista de parámetros que puede proporcionar satatgraphics, situados en dicho panel, seleccionar *Pane Options* y seleccionar los parámetros desde el menú de *Summary Statistics Options*.

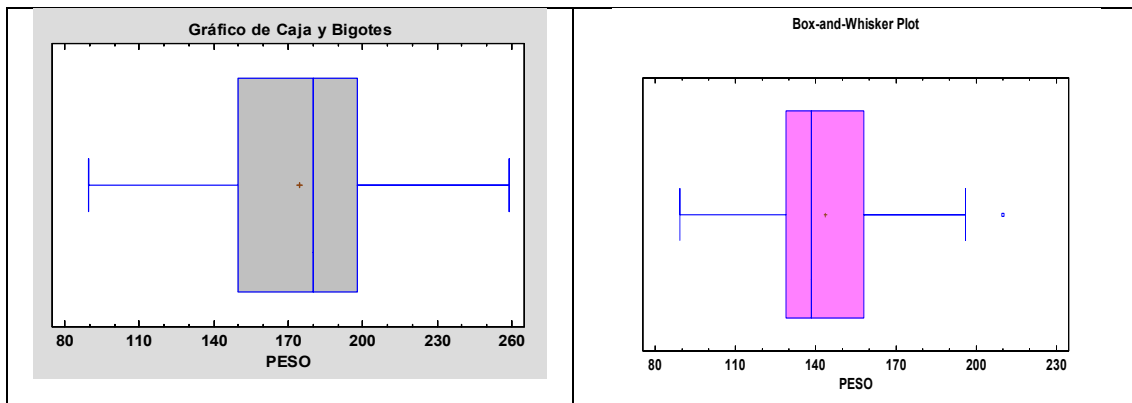


Los valores de los parámetros para la variable PESO son:

Resumen Estadístico para PESO	
Recuento	468
Promedio	174,579
Mediana	180,0
Moda	135,0
Varianza	997,125
Desviación Estándar	31,5773
Mínimo	89,3
Máximo	259,0
Rango	169,7
Cuartil Inferior	150,0
Cuartil Superior	198,0
Rango Intercuartílico	48,0
Sesgo Estandarizado	-2,01207
Curtosis Estandarizada	-2,82584

Entre los gráficos mostrados por defecto para la variable PESO encontramos el **Box-and-Whisker Plot** (Gráfico de caja y bigotes). Este tipo de gráfico nos proporciona

directamente el valor mínimo y máximo valor de la variable, su media, mediana, primer y tercer cuartil. Además, se trata de una herramienta gráfica descriptiva muy útil, sobre todo para la detección de datos anómalos.



En el “hipotético” Box-whisker de la variable PESO, situado a la derecha de la figura (*smartphones* comercializados hasta 2017) , el punto aislado de valor 210 gramos se señala como atípico (anormalmente grande) . Se debería revisar esta observación para confirmar que es correcta o bien, se trata de un error. En este último caso se debería sustituir esta observación por la correcta, y analizar de nuevo la variable PESO.

Con el gráfico en la pantalla, abrimos el menú con el botón derecho del ratón y podemos cambiar con *Pane Options*: la orientación del gráfico, mostrar la media, datos anómalos....; mediante *Graphics Options* podemos definir el aspecto de la representación gráfica: título, fuente, escala, color de relleno, color de las líneas y puntos.... Estas opciones están disponibles para cualquier gráfico que vayamos a representar.

También podemos representar el gráfico Box-and-Whisker desde el menú principal de *statgraphics*, a partir de *Plot* siguiendo la secuencia:

*Plot* → *Exploratory Plots* → *Box-and-Whisker Plots* → *One sample..*

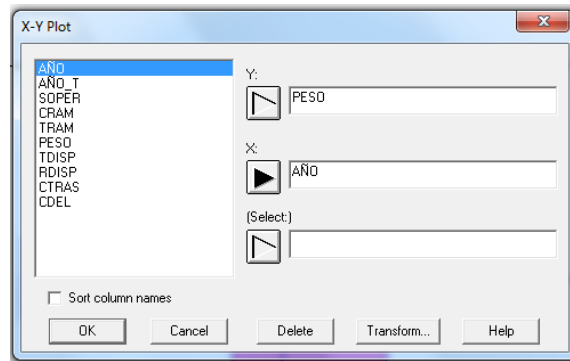
Solicita el nombre de la variable. Debemos introducir PESO en el campo *Data*

Podría interesarnos estudiar cuál ha sido la tendencia experimentada por el peso de los teléfonos móviles a lo largo del tiempo. Para ello, podemos representar un **Diagrama de dispersión**. Este tipo de gráfico nos permite estudiar la posible relación existente entre dos características cuantitativas estudiadas sobre un mismo individuo, en nuestro caso, cada *smartphone*.

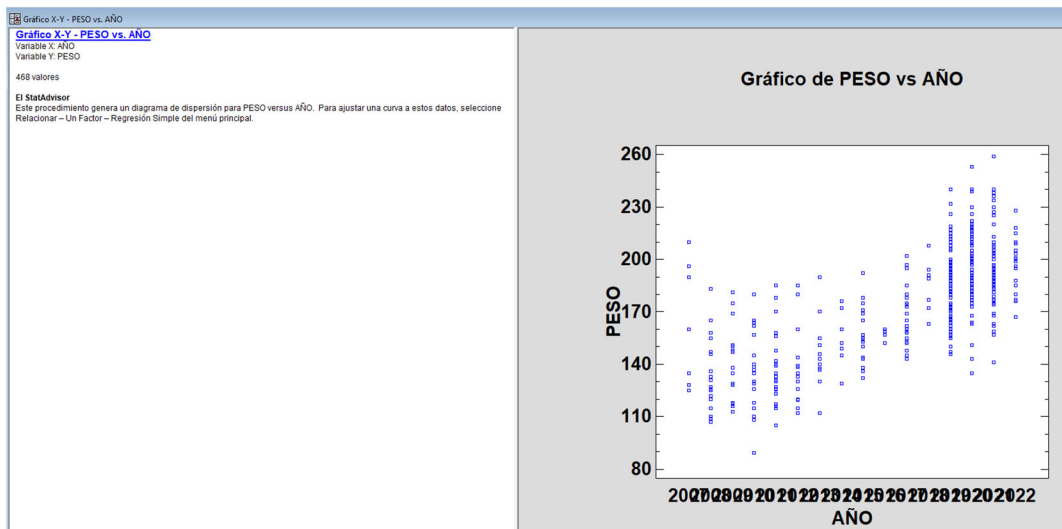
Para representar un diagrama de dispersión seguimos la secuencia:

*Plot* → *Scatterplots* → *X-Y Plot*

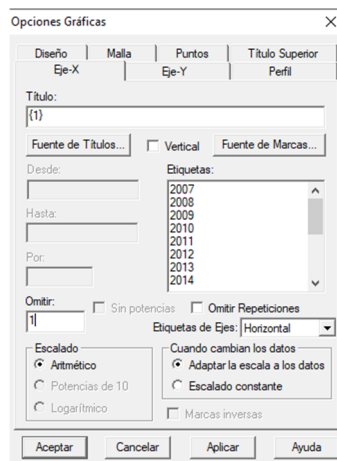
Seleccionamos la variable PESO en el campo *Y*, AÑO en el campo *X*, tal como se indica a continuación:



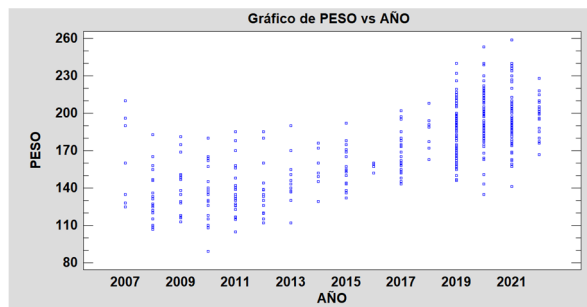
Hacemos click en el botón OK. El resultado se muestra a continuación:



En el panel de la derecha se muestra el diagrama de dispersión solicitado. Haciendo doble click con el ratón el diagrama se amplía ocupando toda la pantalla. Con la opción *Graphics Options* del menú desplegable con el botón derecho del ratón, podemos cambiar la escala en el eje X con el fin de que sea legible la escala en este eje. En el campo *Omitir*: sustituimos el valor "0" por "1".

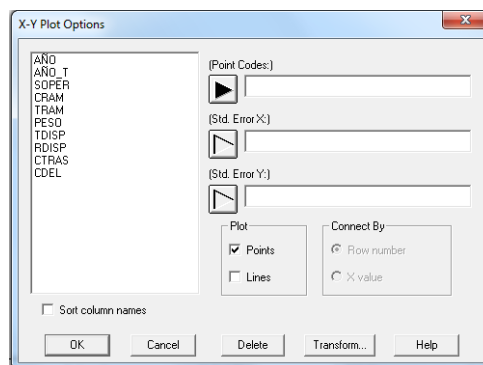


El nuevo diagrama de dispersión, más legible, es el siguiente:

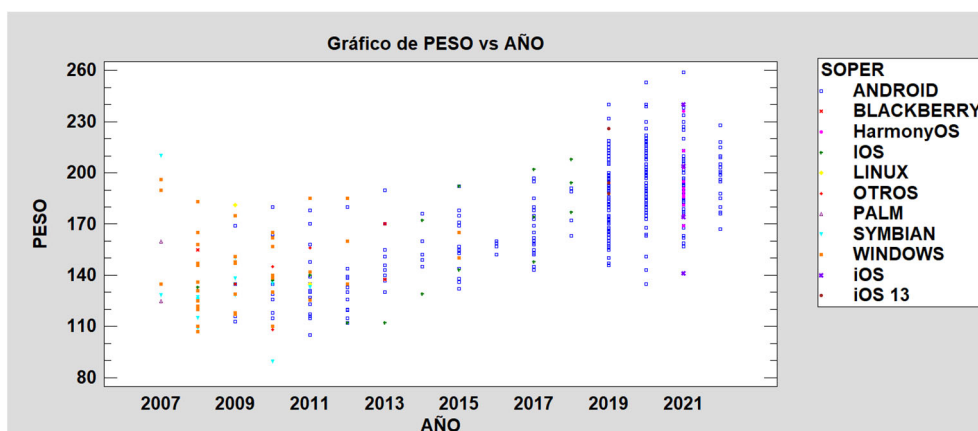


**¿Qué información se desprende del gráfico respecto a la evolución del PESO desde 2007 hasta 2022?**

Podemos obtener más información a partir del diagrama si disponemos de alguna variable cualitativa que nos permita diferenciar los puntos representados respecto a los valores de esta nueva variable. En nuestro caso podemos utilizar la variable SOPER. Para ello con el botón derecho del ratón y *Pane Options*, seleccionamos la variable SOPER de la lista y la introducimos en el campo *Point Codes*.



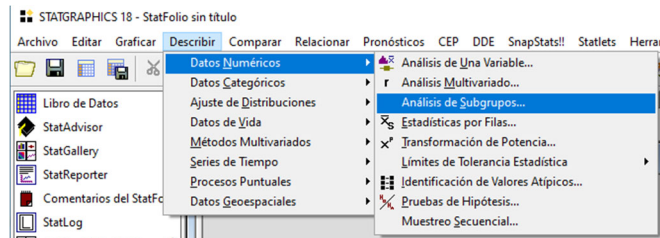
Tras hacer click al botón de Aceptar, el gráfico mostrado es el siguiente:



**¿Añade esta nueva variable información adicional a la evolución del peso de los smartphones en el periodo 2007-2022?**

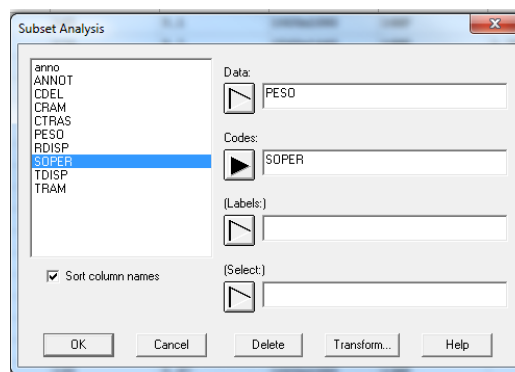
Considerando PESO y SOPER, podemos analizar, por ejemplo, el peso medio de los móviles según el tipo de sistema operativo, sea cual sea el año. Establecemos un grupo de móviles por cada tipo de sistema operativo (valor de la variable SOPER) y de ellos analizamos su peso. En esta ocasión utilizamos la secuencia de opciones:

*Describe* → *Numeric Data* → *Subset Analysis*



En el campo *Data*: PESO, que es la variable que contiene los datos.

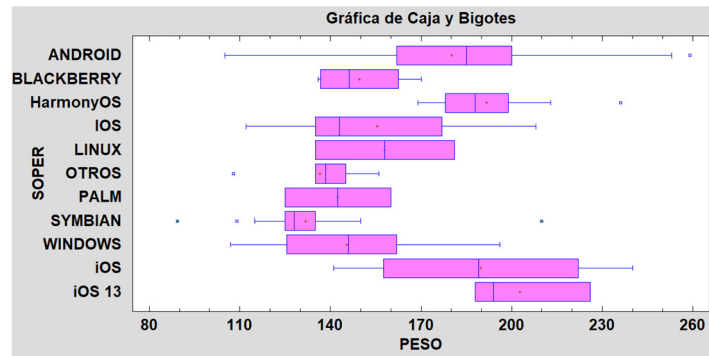
En el campo *Code*: SOPER, que es la variable que establece los grupos de móviles. Un grupo por cada tipo diferente de sistema operativo.



Hacemos click en el botón *OK*. Aceptamos por defecto las opciones del menú de *Tables and Graphs*.



Una vez más y como resultado, los paneles de la izquierda corresponden a análisis numéricos, mientras que los de la derecha corresponden a gráficos. Por ejemplo podemos observar un gráfico **Box-and\_Whisker múltiple**. En este caso representa un Gráfico de caja y bigotes de la variable PESO para cada grupo de móviles con el mismo sistema operativo, en total once gráficos de caja y bigotes. Esta representación por defecto en horizontal se puede cambiar en vertical mediante la opción *Pane Options* del menú desplegable con el botón derecho del ratón.



Los parámetros descriptivos del PESO de los móviles para los diferentes sistemas operativos se muestran en el panel de la izquierda. Podemos solicitar los que más nos interesen situándonos encima del panel y seleccionando *Pane Options* a partir del menú desplegable con el botón derecho del ratón. Por ejemplo, media, desviación típica, mínimo y máximo valor y parámetros de asimetría y curtosis. Este es el valor de los parámetros mencionados:

	Recuento	Promedio	Desviación		Coficiente			Sesgo		Curtosis	
			Estándar	de Variación	Mínimo	Máximo	Rango	Estandarizado	Estandarizada		
SOPER	367	180,18	29,2061	16,2093%	105,0	259,0	154,0	-2,88436	-1,18025		
ANDROID	4	149,625	16,0902	10,7537%	136,0	170,0	34,0	0,562123	-0,856917		
BLACKBERRY	12	191,667	18,5978	9,70321%	169,0	236,0	67,0	1,82073	1,35738		
HarmonyOS	17	155,471	31,0627	19,9798%	112,0	208,0	96,0	0,59185	-1,00881		
IOS	2	158,0	32,5269	20,5867%	135,0	181,0	46,0				
LINUX	5	136,5	17,8255	13,059%	108,0	156,0	48,0	-1,01506	0,923386		
OTROS	2	142,5	24,7487	17,3675%	125,0	160,0	35,0				
PALM	13	131,792	27,744	21,0513%	89,3	210,0	120,7	2,65175	4,15513		
SYMBIAN	39	145,428	23,6612	16,27%	107,0	196,0	89,0	1,00587	-0,763619		
WINDOWS	4	189,75	42,2404	22,2611%	141,0	240,0	99,0	0,0723221	-0,327692		
iOS	3	202,667	20,4287	10,08%	188,0	226,0	38,0	1,10696			
iOS 13	468	174,579	31,5773	18,0876%	89,3	259,0	169,7	-2,01207	-2,82584		

**¿Qué información propocionan los dos últimos análisis realizados?**

Resaltar que en los diferentes tipos de análisis realizados fundamentalmente se ha trabajado con los datos de la variable PESO, es decir, sobre datos de naturaleza cuantitativa continua.

**¿Con qué otra variable del fichero Datos\_Moviles.sf3 podríamos realizar este tipo de análisis?**

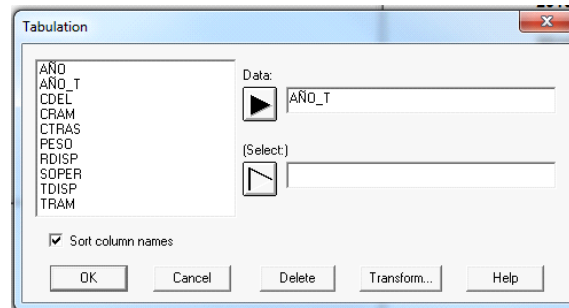
Con TDISP, puesto que contiene valores de naturaleza cuantitativa. Se trata de una variable numérica y de tipo continuo.

**¿Qué información puedes obtener del análisis descriptivo del Tamaño de pantalla (TDISP)?**



Vamos a continuar con otro tipo de análisis especialmente orientado a variables de naturaleza cualitativa (también podría ser el caso de variables cuantitativas discretas con pocos valores diferentes). Por ejemplo, podría interesarnos analizar la cantidad de móviles que se han registrado por año, en el periodo de 2007-2022. Trabajamos con los datos de la variable AÑO\_T. Para ello seguimos la secuencia de opciones:

*Describe* → *Categorical Data* → *Tabulation*



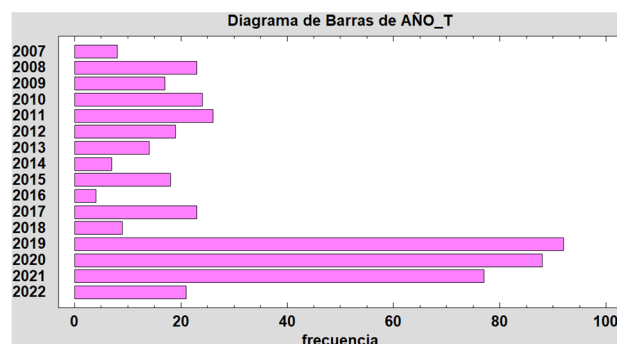
Seleccionamos la variable AÑO\_T en el campo *Data:* y posteriormente hacemos click en el botón de OK.

Marcamos la opción de Gráfico de Anillo del menú de Tablas y Gráficos.



Una vez más los paneles de la derecha corresponden a análisis gráficos (Diagrama de Barras y Diagrama de Sectores), mientras que a la izquierda muestra una Tabla de Frecuencias.

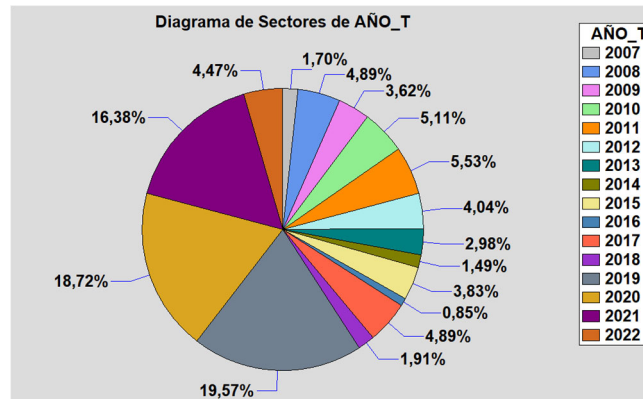
El **Diagrama de Barras** es similar al histograma de frecuencias, pero al tratarse de una variable discreta no agrupa los valores en tramos, simplemente traza una barra con una longitud/altura proporcional a la frecuencia con la que aparece cada valor de la variable.



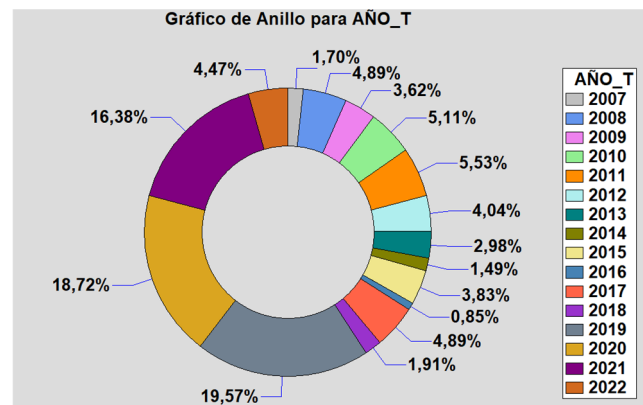
Con el botón derecho del ratón y *Pane Options* es posible modificar algunas características del gráfico, como el tipo de frecuencia representada (absoluta/relativa),

el tipo de barras representadas, la dirección.....

Esta misma información que proporciona el diagrama de barras se puede obtener a partir del **Diagrama de Sectores** o Diagrama de Tarta.



Como siempre, con el botón derecho del ratón y Pane Options es posible modificar algunos de los parámetros del diagrama. Por ejemplo, frecuencia absoluta, absoluta y relativa, desplazamiento de algún sector.....



**¿Qué conclusiones se obtienen a partir de estos gráficos?**

Observa la **Tabla de Frecuencias** de la variable AÑO\_T representada en el panel de la izquierda. Responde a las siguientes cuestiones:

- \*.- ¿Qué porcentaje de los *smartphones* analizados son anteriores a 2015?
- \*.- ¿Cuántos móviles corresponden al año 2010?
- \*.- ¿Qué porcentaje de los *smartphones* analizados corresponden al año pasado?
- \*.- ¿Qué porcentaje de los móviles analizados corresponden a los últimos tres años?

Tabla de Frecuencia para AÑO\_T

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. acum.
1	2007	8	0,0170	8	0,0170
2	2008	23	0,0489	31	0,0660
3	2009	17	0,0362	48	0,1021
4	2010	24	0,0511	72	0,1532
5	2011	26	0,0553	98	0,2085
6	2012	19	0,0404	117	0,2489
7	2013	14	0,0298	131	0,2787
8	2014	7	0,0149	138	0,2936
9	2015	18	0,0383	156	0,3319
10	2016	4	0,0085	160	0,3404
11	2017	23	0,0489	183	0,3894
12	2018	9	0,0191	192	0,4085
13	2019	92	0,1957	284	0,6043
14	2020	88	0,1872	372	0,7915
15	2021	77	0,1638	449	0,9553
16	2022	21	0,0447	470	1,0000

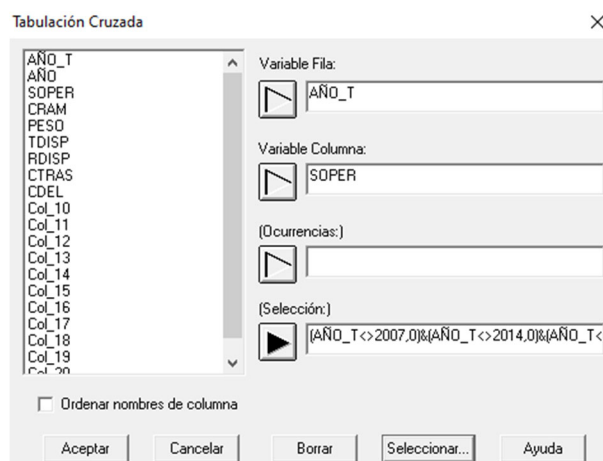
Respecto al resto de variables incluidas en el fichero de datos, ¿podrías hacer estos análisis con la variable SOPER? En caso afirmativo realiza los análisis que consideres más relevantes y responde a las siguientes preguntas:

- \*.- ¿Qué sistema operativo ha sido el más frecuente en los móviles en el periodo analizado?
- \*.- ¿Qué porcentaje de los móviles han incorporado el sistema operativo LINUX?
- \*.- ¿Cuál de los sistemas operativos ha sido el “menos exitoso” en el periodo analizado?
- \*.- A partir de los datos analizados, ¿parece el sistema WINDOWS una buena opción para los *smartphones*?

El análisis anterior podemos completarlo mediante una Tabla de doble entrada o **Tabla de frecuencias cruzada**, en la que estudiaremos la frecuencia de la variable SOPER en función del año (variable AÑO\_T). Para la construcción de la tabla seguimos la secuencia de opciones de statgraphics:

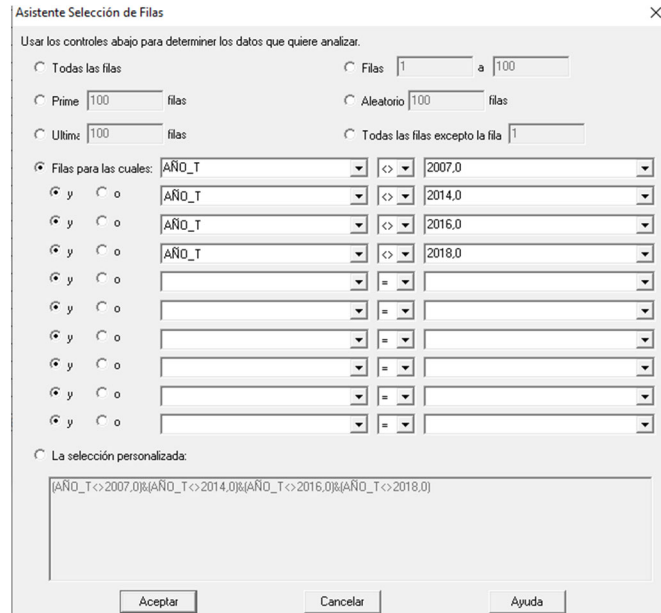
*Describe* → *Categorical data* → *Crosstabulation*

A continuación indicamos el nombre de las variables que se sitúan en las filas (AÑO\_T) y en las columnas (SOPER) de la tabla, tal como aparece en el siguiente cuadro:



Como para los años 2007, 2014, 2016 y 2018 hay muy pocas observaciones en relación al resto de años, no se han considerado en el análisis las observaciones correspondientes a estos cuatro años. De esta manera el análisis queda más resumido y compacto, siendo posible captar toda la información en un golpe de vista. Para ello en el campo *Select*, indicamos las observaciones a tener en cuenta en el análisis. Serán aquellas que cumplan la condición:

$(AÑO\_T <> 2007) \& (AÑO\_T <> 2014) \& (AÑO\_T <> 2016) \& (AÑO\_T <> 2018)$



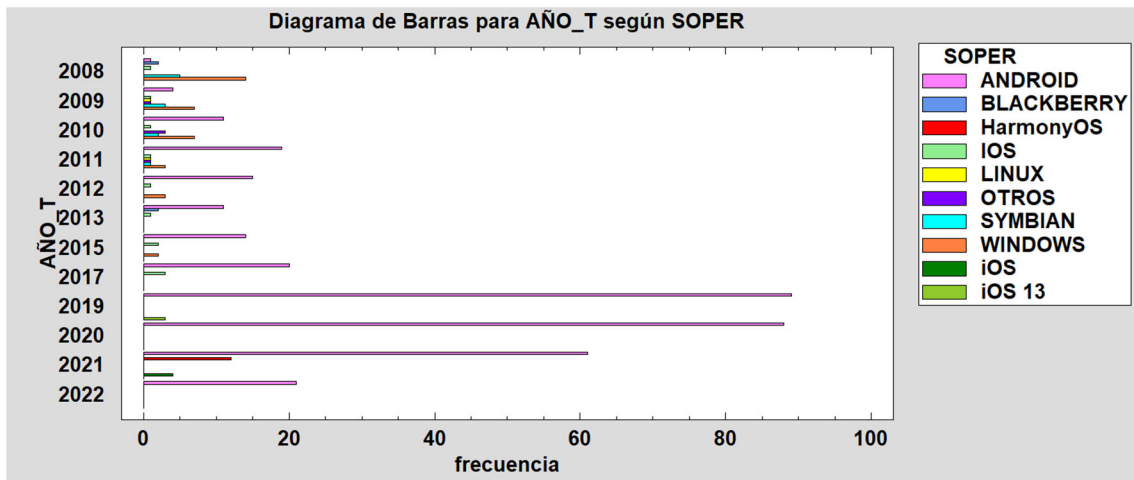
Hacemos click en el botón OK. Como resultado, esta opción proporciona dos paneles de gráficos a la parte derecha de la pantalla: Diagrama de Barras y Gráfico de Mosaico. El segundo panel a la izquierda muestra la Tabla de frecuencias cruzadas o Tabla de contingencia para las variables seleccionadas: AÑO\_T y SOP.

Tabla de Frecuencias para AÑO\_T por SOPER

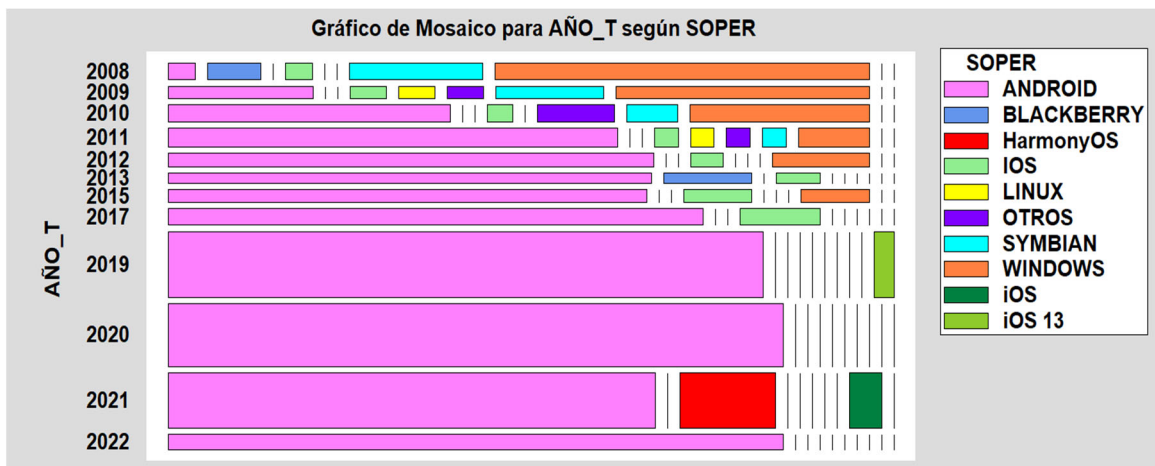
	ANDROID	BLACKBERRY	HarmonyOS	IOS	LINUX	OTROS	SYMBIAN	WINDOWS	iOS	iOS 13	Total por Fila
2008	1 0,23%	2 0,45%	0 0,00%	1 0,23%	0 0,00%	0 0,00%	5 1,13%	14 3,17%	0 0,00%	0 0,00%	23 5,20%
2009	4 0,90%	0 0,00%	0 0,00%	1 0,23%	1 0,23%	1 0,23%	3 0,68%	7 1,58%	0 0,00%	0 0,00%	17 3,85%
2010	11 2,49%	0 0,00%	0 0,00%	1 0,23%	0 0,00%	3 0,68%	2 0,45%	7 1,58%	0 0,00%	0 0,00%	24 5,43%
2011	19 4,30%	0 0,00%	0 0,00%	1 0,23%	1 0,23%	1 0,23%	1 0,23%	3 0,68%	0 0,00%	0 0,00%	26 5,88%
2012	15 3,39%	0 0,00%	0 0,00%	1 0,23%	0 0,00%	0 0,00%	0 0,00%	3 0,68%	0 0,00%	0 0,00%	19 4,30%
2013	11 2,49%	2 0,45%	0 0,00%	1 0,23%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	14 3,17%
2015	14 3,17%	0 0,00%	0 0,00%	2 0,45%	0 0,00%	0 0,00%	0 0,00%	2 0,45%	0 0,00%	0 0,00%	18 4,07%
2017	20 4,52%	0 0,00%	0 0,00%	3 0,68%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	23 5,20%
2019	89 20,14%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	3 0,68%	92 20,81%
2020	88 19,91%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	88 19,91%
2021	61 13,80%	0 0,00%	12 2,71%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	4 0,90%	0 0,00%	77 17,42%
2022	21 4,75%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	0 0,00%	21 4,75%
Total por Columna	354 80,09%	4 0,90%	12 2,71%	11 2,49%	2 0,45%	5 1,13%	11 2,49%	36 8,14%	4 0,90%	3 0,68%	442 100,00%

Contenido de las celdas:  
Frecuencia Observada  
Porcentaje de la Tabla

El gráfico superior es un **Diagrama de barras** para los diferentes años considerados el análisis. Para cada año muestra la frecuencia de cada uno de los sistemas operativos en un color diferente, según corresponde en la leyenda.



El **Gráfico de Mosaico** muestra, para cada año, una barra por cada categoría de SOP con una anchura más o menos grande dependiendo de su frecuencia. La altura de cada barra depende del número de observaciones o frecuencia para cada año. A mayor número de observaciones, mayor altura.



**¿Qué información proporciona este nuevo gráfico respecto a la evolución de los sistemas operativos de los smartphones en el periodo considerado?**

