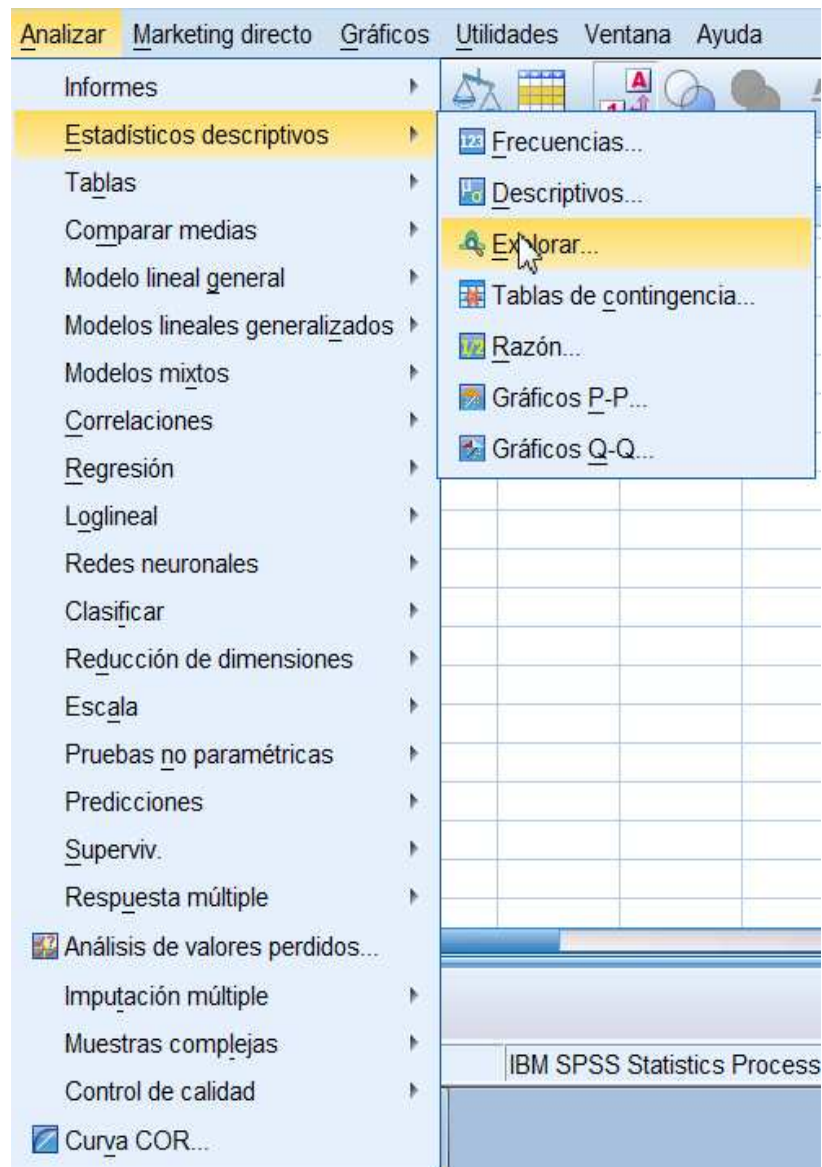


## Procedimiento "Explorar..."

El procedimiento Explorar genera estadísticos de resumen y representaciones gráficas como Diagrama de cajas, gráficos de tallo y hojas, histogramas, diagramas de normalidad y diagramas de dispersión por nivel con pruebas de Levene.

Para explorar los datos se elige en el *Menú principal* **Analizar/Estadísticos descriptivos/Explorar...**



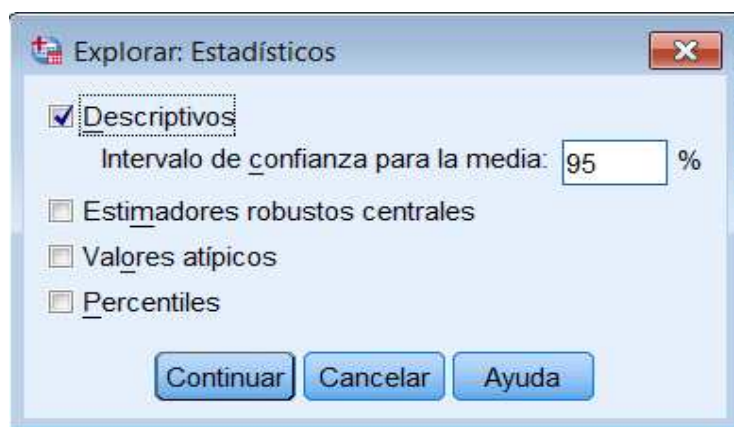
y se muestra el siguiente *Cuadro de diálogo*



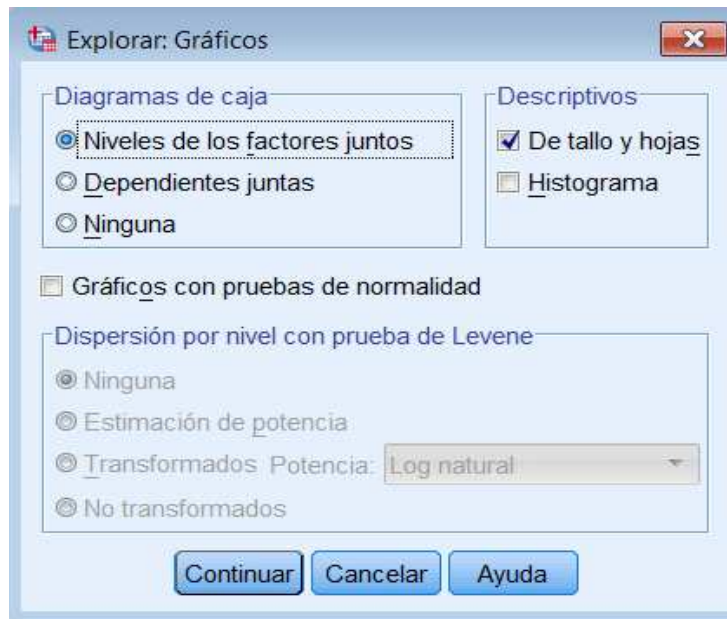
donde se selecciona una o más variables dependientes (**Lista de dependientes:**) Si se desea, es posible seleccionar una o más variables de factor (**Lista de Factores:**) cuyos valores definirán grupos de casos; o bien seleccionar una variable de identificación para etiquetar casos (**Etiquetar los casos mediante:**).

En dicho cuadro de diálogo se presentan tres botones de comandos: **Estadísticos...** (para: Intervalos de confianza para la media, Estimadores robustos centrales, Valores atípicos y Percentiles), **Gráficos...** (para Diagramas de cajas, de tallo y hojas, histogramas, pruebas y gráficos de probabilidad normal y diagramas de dispersión por nivel con la prueba de Levene), **Opciones...** (para el tratamiento de valores perdidos) y **Bootstrap...** (para realizar muestreo bootstrap). Si en **Visualización** se marca la opción **Ambos** muestra resultados estadísticos y gráficos; si se marca **Estadísticos** muestra sólo los resultados estadísticos y si se marca **Gráficos** muestra sólo resultados gráficos.

Pinchando en **Estadísticos...** se muestra el siguiente *Cuadro de diálogo*

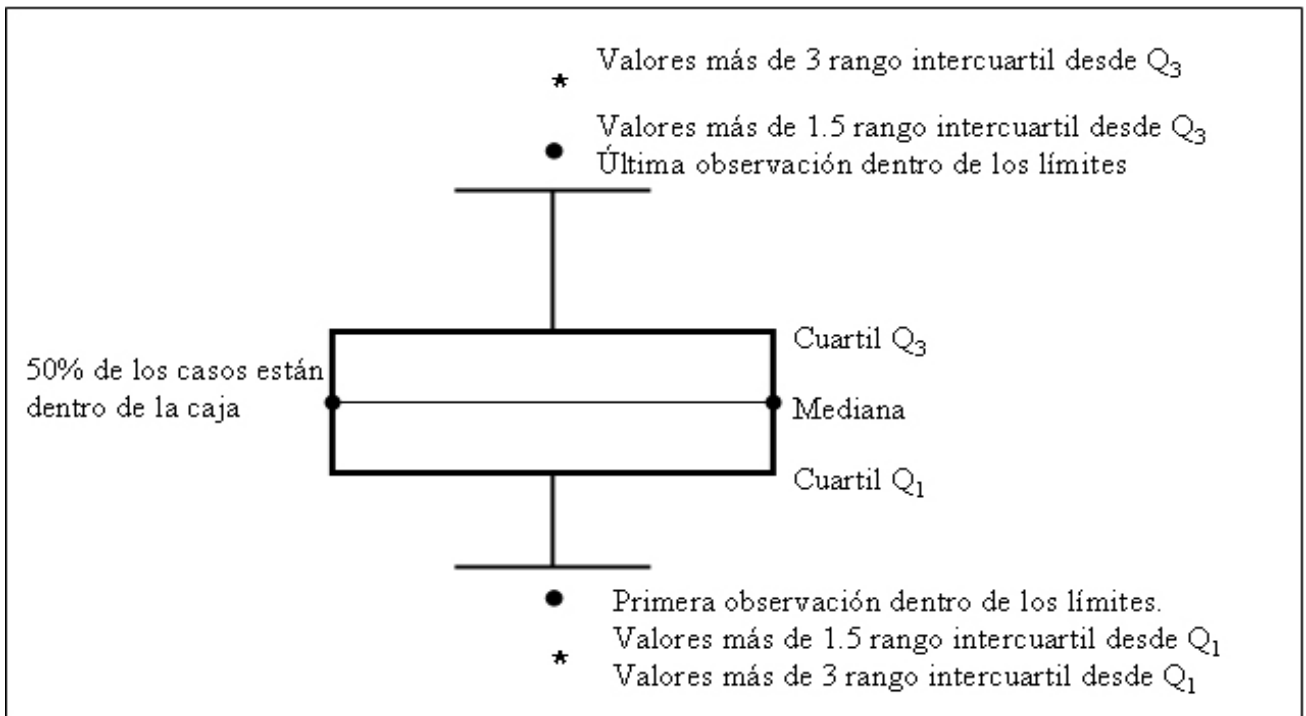


Pinchando en **Gráficos...** se obtiene el siguiente *Cuadro de diálogo*



El **Diagrama de cajas** es una forma de representación gráfica para resumir la distribución de los valores de una variable. En esta representación en vez de visualizar los valores individuales, se representan estadísticos básicos de la distribución: la mediana, el percentil 25, el percentil 75 y los extremos de la distribución. Esta representación gráfica está basada en 5 medidas estadísticas. La caja representa el 50% de las observaciones correspondientes a cada nivel del factor y se construye desde el primer cuartil al tercer cuartil; dentro de ella se destaca la posición de la mediana. Se construye un bigote desde  $Q_1$  hasta el mínimo y otro bigote desde  $Q_3$  hasta el máximo, siempre que sus correspondientes longitudes no superen 1.5 veces la longitud de la caja. En caso contrario, el correspondiente bigote se construye hasta la observación más alejada del cuartil que diste como máximo 1.5 veces la longitud de la caja; si todas las observaciones distan del correspondiente cuartil más de 1.5 veces la longitud de la caja se representan como valores aislados.

En la siguiente figura se muestra cómo se representan en un gráfico de cajas los estadísticos anteriormente mencionados



- El diagrama de caja viene determinado por los valores de los cuartiles, y la línea que aparece dentro del diagrama representa el valor de la mediana
- Los límites, a partir de los cuales se dibujan los bigotes del diagrama de caja y se marcan los valores extremos, se calculan del siguiente modo:

$$LI = Q_1 - 1.5(Q_3 - Q_1) \quad \text{y} \quad LS = Q_1 + 1.5(Q_3 - Q_1)$$

donde  $Q_1$  y  $Q_3$  son el primer y el tercer cuartil, respectivamente.

- Los valores fuera de los límites están etiquetados (en SPSS por el número de caso al que corresponden, cuando no se ha especificado ninguna variable para etiquetar). En la figura anterior se representan con un círculo aquellos valores que sobrepasan los límites  $LI$  y  $LS$  y con un asterisco los que sobrepasan los límites  $LI^*$  y  $LS^*$ , que se calculan del siguiente modo:

$$LI^* = Q_1 - 3(Q_3 - Q_1) \quad \text{y} \quad LS^* = Q_1 + 3(Q_3 - Q_1)$$

En resumen:

- Se consideran dos categorías de casos extremos, en función de cuánto se alejan con respecto del 50% central de la distribución. Aquellos casos con valores alejados más de 3 veces el rango intercuartílico desde el extremo superior o inferior de la caja (casos más extremos, representados con un (\*)) y aquellos valores que están alejados entre 1.5 y 3 veces dicho rango, representados con un círculo.
- Los valores más pequeño y más grande que estén dentro de los límites primer cuartil -1.5 y tercer cuartil +1.5 veces el rango intercuartílico constituyen los

bigotes del gráfico y aparecen representados mediante las dos líneas horizontales dibujadas a ambos extremos de la caja central.

Información que podemos obtener de este tipo de gráfico:

- La posición de la mediana nos determina la tendencia central
- El ancho de la caja nos da idea de la variabilidad de las observaciones. Si la mediana no está en el centro de la caja, podemos deducir que la distribución es asimétrica (si está próxima al límite inferior de la caja, asimétrica positiva, y si está próxima al límite superior, asimétrica negativa)
- Estos gráficos son especialmente útiles para comparar la distribución de los valores entre diferentes grupos.

**Diagrama de tallos y hojas:** Una técnica que se utiliza para la observación de la forma de la distribución de frecuencias es el diagrama de tallo y hojas, que es una representación gráfica en la que los datos se colocan en dos niveles de tal modo que podemos visualizar la forma de la distribución. Un diagrama de tallos y hojas consiste en una serie de filas horizontales de números. En la primera columna se escriben los llamados tallos, junto a ella se traza una línea vertical y, a su derecha, en cada fila se escriben las correspondientes hojas. El número utilizado para designar una fila es su tallo, el resto de números de la fila se denominan hojas. El tallo es la mayor porción del número. Por ejemplo, en los números 1.0, 2.3, 1.4, 1.1, 3.5, 4.6, 5.1, 5.3, 5.7 y 5.9. Los primeros dígitos servirán como nombre de los tallos (1, 2, 3, 4, 5). Las hojas dan una información secundaria acerca del número, en nuestro ejemplo sería la cifra decimal.

**Gráficos con prueba de normalidad:** Mediante este procedimiento se comprueba si los datos proceden de una población normal y para ello se sirve de dos gráficos y un contraste analítico.

Pinchando en **Opciones...** se obtienen las siguientes opciones para el tratamiento de **Valores perdidos**

