

BREVE INTRODUCCIÓN A STATGRAPHICS CENTURION

Curso 2023/24

Departamento de Matemática Aplicada

Escuela de Ingeniería Informática de Valladolid

UNIVERSIDAD DE VALLADOLID

Prólogo

STATGRAPHICS Centurion 19.5 es un programa estadístico cuya virtud más sobresaliente es su facilidad de uso, aprendizaje y manejo. Además, este software constituye una herramienta muy potente para el análisis estadístico. Estas características lo convierten en un programa adecuado metodológicamente para la enseñanza de la Estadística en la universidad.

La utilización de esta aplicación para la realización de análisis estadísticos requiere el conocimiento de dos aspectos importantes: el entorno de trabajo del programa y los procedimientos específicos para realizar la tarea estadística deseada. El objetivo de este resumen es explicar al lector estas dos cuestiones. Por este motivo, la primera parte de esta introducción se refiere al funcionamiento general de la aplicación, mientras que el resto alude a las diferentes opciones que proporciona el programa para trabajar con los conceptos estadísticos básicos.

Las opciones del programa STATGRAPHICS Centurion 19.5 que se describen en estas notas se corresponden con el contenido de la asignatura Estadística de la titulación de Grado en Ingeniería Informática. Los lectores que deseen utilizar otras técnicas estadísticas, distintas a las aquí presentadas, y que se encuentran implementadas en la aplicación, no tendrán ninguna dificultad. Esto se debe a que todos los procedimientos se estructuran de la misma manera y, por tanto, sólo es necesario localizar donde se encuentran utilizando la ayuda del programa.

Valladolid, Septiembre de 2023

Índice general

Prólogo	III
Parte 1. El marco de trabajo de StatGraphics Centurion 19.5	1
Capítulo 1. Pantallas y ventanas de la aplicación	3
1. Ventana principal del programa	5
2. Ventana de hoja de cálculo	8
3. Opciones en los gráficos	14
4. Ventana StatGallery	17
5. Uso de la opción <i>Ayuda</i>	18
Parte 2. Estadística Descriptiva	21
Capítulo 2. Estadística Descriptiva Univariante	23
1. Descripción estadística de variables unidimensionales	23
2. Descripción gráfica de variables unidimensionales	28
Capítulo 3. Estadística Descriptiva Multivariante	31
1. Variables numéricas	31
2. Regresión simple	39
3. Variables cualitativas	43
Parte 3. Cálculo de Probabilidades	49
Capítulo 4. Modelos de probabilidad más habituales.	51
1. Opciones tabulares de <i>Distribuciones Probabilidad</i>	58
2. Opciones gráficas de <i>Distribuciones de Probabilidad</i>	60
Parte 4. Inferencia Estadística	67

Capítulo 5. Estimación puntual y confidencial	69
1. Estimación puntual	69
2. Estimación por Intervalos de Confianza	70
Capítulo 6. Contraste de hipótesis paramétricas	83
1. Una sola muestra	83
2. Dos muestras independientes	87
3. Datos pareados	89
Capítulo 7. Tests no paramétricos	91
1. Contraste de bondad de ajuste a una distribución	91
2. Test Chi-cuadrado de independencia y homogeneidad	92
3. Test de aleatoriedad	93
4. Test de los signos y de rangos con signo	95
5. Test de suma de rangos de Wilcoxon o Mann-Whitney	96

Parte 1

El marco de trabajo de StatGraphics

Centurion 19.5

CAPÍTULO 1

Pantallas y ventanas de la aplicación

En este primer capítulo se introduce el entorno de trabajo del programa STATGRAPHICS Centurion 19.5 para Windows, la gestión de los archivos de datos, la presentación de las salidas gráficas y el uso de la Ayuda.

Después de ejecutar STATGRAPHICS Centurion 19.5 nos aparece la pantalla *StatWizard* (ver la Figura 1.1), que es una ventana que ayuda a seleccionar el análisis más apropiado del programa dependiendo de la tarea que se quiera llevar a cabo.

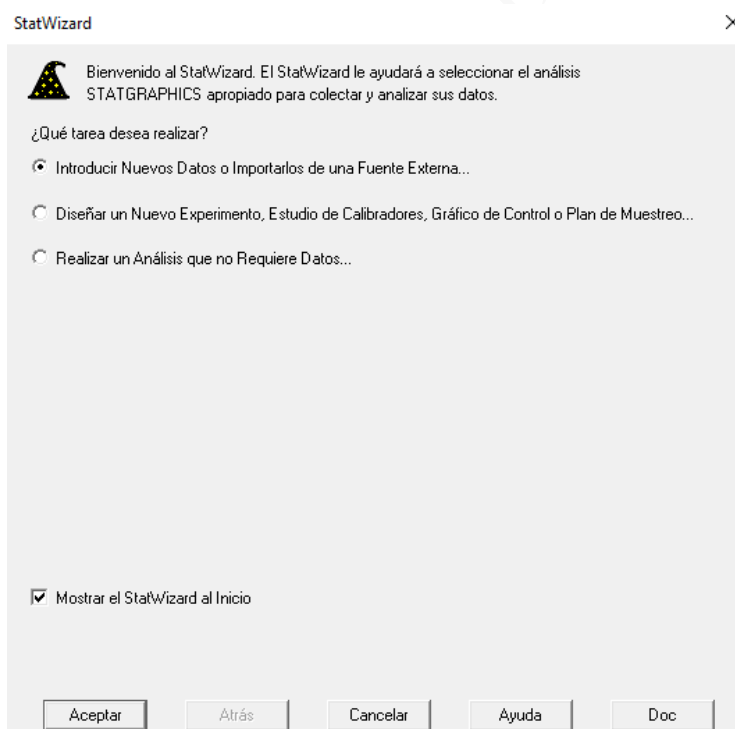


Figura 1.1 Ventana *StatWizard*

Una vez seleccionada la tarea, nos aparece una nueva ventana, en función de la elección realizada. Si la tarea elegida ha sido la primera, nos aparece la ventana de la Figura 1.2. Si, por el contrario, hubiésemos seleccionada la tercera opción (Realizar un Análisis que no Requiere Datos), nos aparecería la ventana de la Figura 1.3.

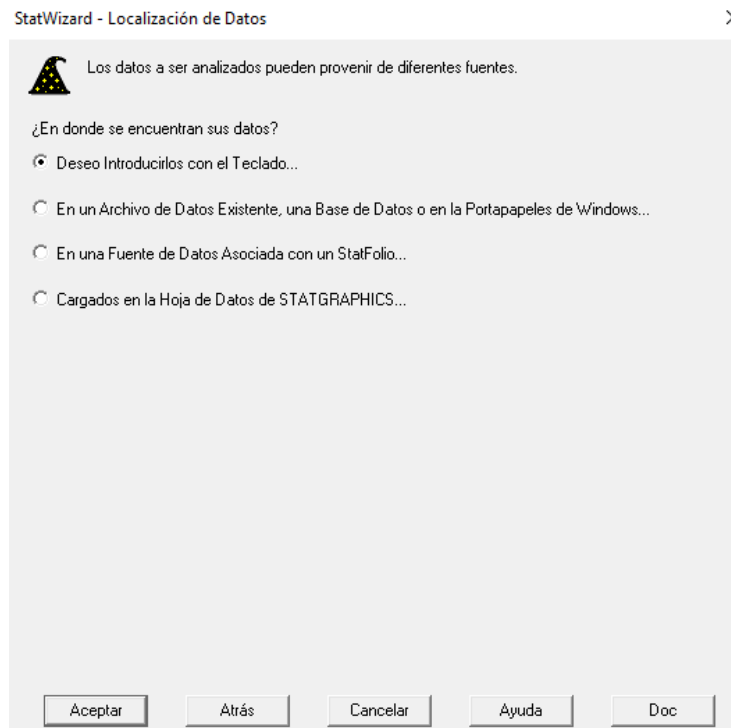


Figura 1.2 Ventana de Localización de los Datos

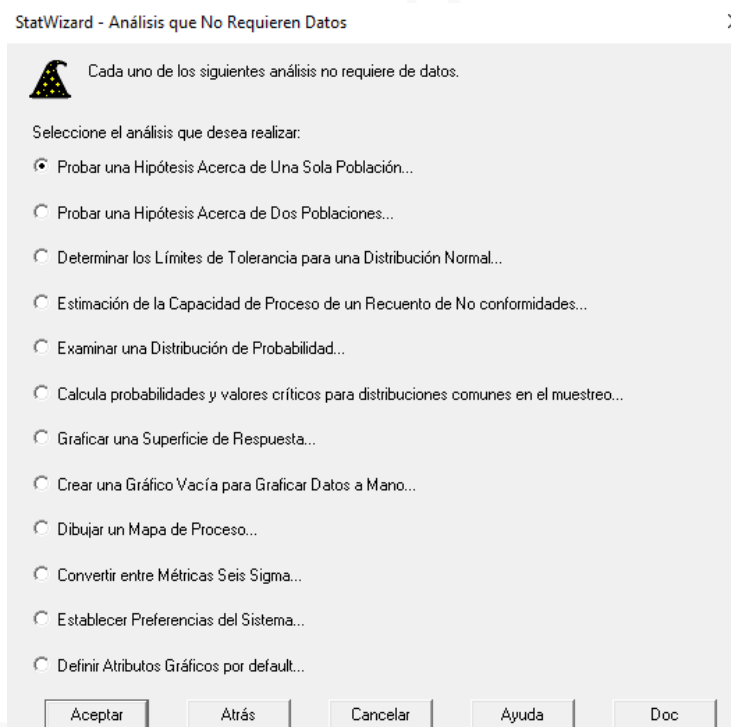


Figura 1.3 Ventana de Análisis que No Requieren Datos

1. Ventana principal del programa

Una vez cerrada la ventana *StatWizard*, nos aparece la pantalla inicial de la aplicación, que es el marco general de trabajo. Su aspecto es el habitual de cualquier aplicación Windows: tres líneas en la parte superior de la pantalla, un cuerpo central con cuatro iconos en su parte inferior izquierda y una última línea de estado en la parte inferior de la pantalla (ver la Figura 1.4).

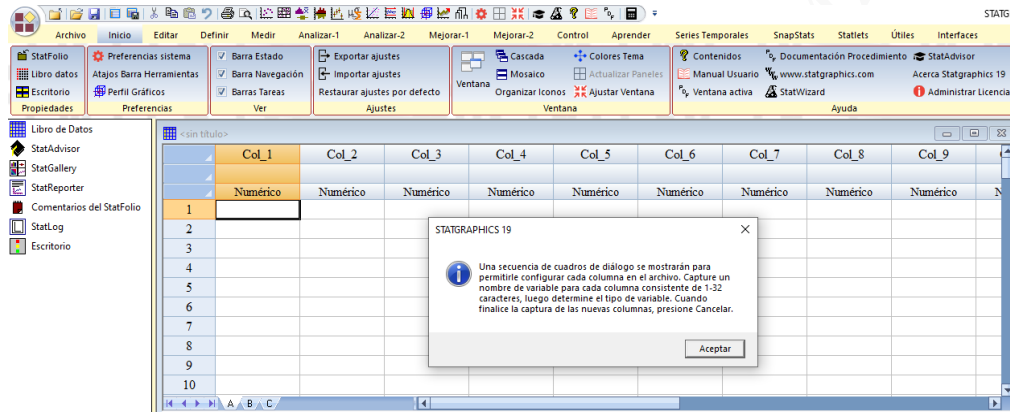



































Figura 1.4 Marco general de trabajo en STATGRAPHICS Centurion

En la línea superior de la pantalla aparece el icono de STATGRAPHICS Centurion , junto con la **barra de herramientas**, cuyo contenido son diferentes iconos que permiten acceder rápidamente a las opciones más usuales en el trabajo de la aplicación sin necesidad de pasar por menús o submenús. El significado de cada icono se obtiene posicionando el ratón sobre él. Los iconos de la barra de herramientas significan lo siguiente (de izquierda a derecha):

- Nuevo StatFolio. 
- Abrir un StatFolio. 
- Guardar el StatFolio activo. 
- Abrir un fichero de datos. 
- Guardar el archivo de datos activo. 
- Cortar la selección y pegarla en el Bloc de notas. 
- Copiar la selección y pegarla en el Bloc de notas. 
- Insertar el contenido del Bloc de notas. 
- Deshacer la última acción. 

- Imprimir. 
- Vista preliminar. 
- Gráfico de dispersión bidimensional. 
- Gráfica de matriz. 
- Análisis de una variable. 
- Comparación dos muestras independientes. 
- Comparación de varias muestras. 
- Análisis multivariable. 
- Regresión simple. 
- Gráficos X y MR(2). 
- Análisis de capacidad (individuales). 
- Crear y analizar experimento. 
- Pronósticos automáticos. 
- Análisis de conglomerados. 
- Definir preferencias del sistema. 
- Restablecer división de ventanas a su posición predeterminada. 
- Establecer ventana a valor predeterminado. 
- Abrir la ventana *StatAdvisor* (intérprete de resultados). 
- Utilizar *StatWizard* para sugerir análisis. 
- Ayuda. 
- Mostrar manual de usuario. 
- Presentar documentación para tarea actual o comando. 
- Evaluación de expresiones. 

En la línea siguiente se presenta la **barra de menú**, que contiene el menú general de la aplicación con todas sus opciones (*Archivo*, *Inicio*, *Editar*, *Definir*, *Medir*, *Analizar-1*, *Analizar-2*, *Mejorar-1*, *Mejorar-2*, *Control*, *Aprender*, *Series Temporales*, *SnapStats*, *Statlets*, *Útiles* e *Interfaces*).

En la tercera línea se presenta una **segunda barra de herramientas** que depende de la opción que esté activa en la barra de menú. En la Figura 1.5 se muestra esta barra de herramientas cuando está activa la opción *Inicio*.

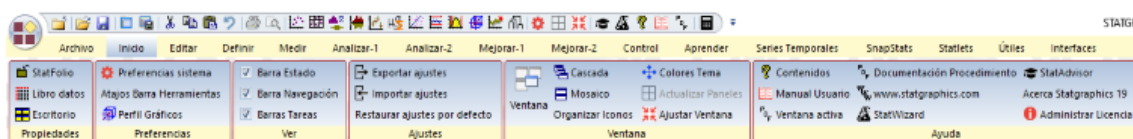


Figura 1.5 Barra de herramientas con la opción *Inicio*

En la parte central de la pantalla aparece una ventana etiquetada como *<sin título>* que presenta una especie de rejilla en su parte izquierda e informa del conjunto de datos activo. La ventana es una hoja de cálculo y forma el soporte utilizado por STATGRAPHICS para mostrar y manipular los datos.

En la parte izquierda aparece la **barra de navegación**, que contiene una lista de todas las ventanas abiertas contenidas en el *StatFolio* activo.

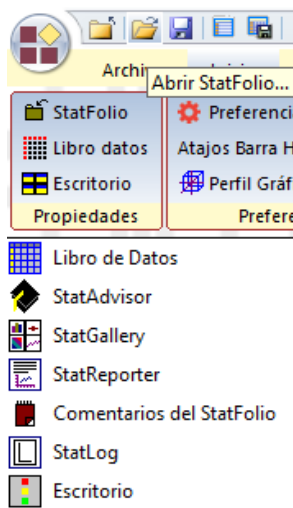


Figura 1.6 Barra de navegación

El primer icono hace referencia al *Libro de Datos* con sus correspondientes hojas activas (tres por defecto). A continuación aparece el icono etiquetado como *StatAdvisor* que se utiliza para activar el intérprete de resultados. El tercer icono que aparece es el correspondiente a *StatGallery*, que es utilizado para presentar representaciones gráficas y salidas numéricas. Debajo aparece el icono denominado *StatReporter*, que abre una ventana con un editor de texto para generar informes en formato texto enriquecido (*.rtf). El quinto icono es una especie de cuaderno abierto a su izquierda y recoge comentarios sobre el *StatFolio* activo. El sexto icono se refiere a *StatLog* que se utiliza para realizar el seguimiento de un análisis y muestra también información reactiva a cuándo se inició la sesión, qué datos fueron cargados y qué análisis

se realizaron. Se puede guardar su contenido en formato texto enriquecido (*.rtf). Por último, aparece el icono *Escritorio* que abre una ventana que está diseñada para mostrar tablas y gráficos de múltiples análisis dentro de un StatFolio. La salida de los análisis se pega en los paneles del Escritorio y se actualiza cada vez que cambian los datos o las opciones de análisis. Inicialmente, el Escritorio tiene 25 paneles dispuestos en una cuadrícula con 5 filas y 5 columnas.




Figura 1.7 Barra de ventanas abiertas en Statgraphics

2. Ventana de hoja de cálculo

En la mayoría de los procedimientos estadísticos que se realizan con STATGRAPHICS se necesitan datos sobre los que efectuar los cálculos correspondientes al procedimiento seleccionado. Dichos datos deben estar presentes en la hoja de cálculo antes de llamar al procedimiento que queremos desarrollar. Las columnas de la hoja de cálculo van a ser las variables para los análisis estadísticos y el conjunto de las columnas de toda la hoja de cálculo forman el fichero de datos.

2.1. Creación de nuevos ficheros y variables. Si comenzamos una nueva sesión con STATGRAPHICS no tendremos ningún fichero de datos en la memoria de trabajo, por lo que deberemos seleccionar y activar el icono correspondiente a la hoja de datos. Una vez que nos aparece la hoja de cálculo vacía podemos comenzar a introducir los datos del fichero. Mediante el comando *Modificar Columna* de la subopción **Columnas Datos** de la opción *Editar* de la barra de menús del programa (ver Figura 1.8) podemos definir las características de los datos (nombre de la variable que los contiene, comentario, anchura y tipo). Esta opción se activa al marcar la columna y se puede acceder a ella más rápidamente pulsando el botón derecho del ratón tras la selección de la columna.

Una vez finalizada la introducción de variables en el fichero podemos guardarlo con nombre PRACTICA1 de forma permanente seleccionando la subopción *Guardar Como* de la opción *Archivo* de la barra del menú del programa (ver Figura 1.9) o con el icono .

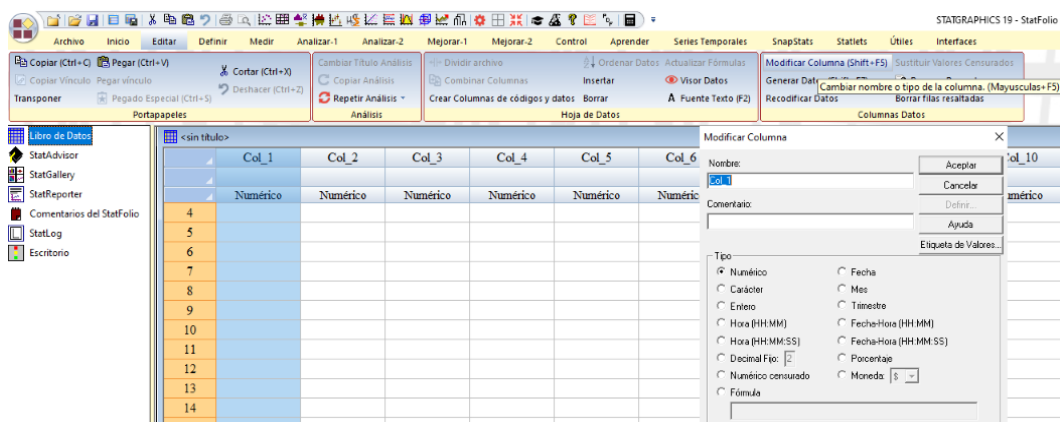


Figura 1.8 Acceso al procedimiento *Modificar Columna*

A partir de ese momento existirá ya un nuevo fichero de datos con el nombre especificado y extensión **.sgd** y sustituirá a *<sin nombre>* en la hoja de datos minimizada.

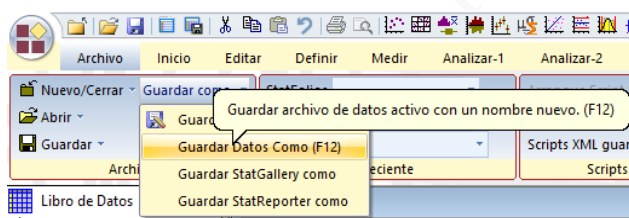


Figura 1.9 Menú de acceso a *Guardar Datos Como*

2.2. Funciones y operadores de STATGRAPHICS.

2.2.1. Operadores aritméticos:

- $x + y$ *Suma*
- $x - y$ *Resta*
- $x * y$ *Multiplicación*
- x / y *División*
- $x ^y$ *Potencia*

2.2.2. Operadores relacionales:

- $x = y$ *Igual a*
- $x <> y$ *No igual a*
- $x > y$ *Mayor que*
- $x >= y$ *Mayor o igual que*

- $x < y$ *Menor que*
- $x \leq y$ *Menor o igual que*

2.2.3. Operadores lógicos:

- $x \& y$ *AND*
- $x \mid y$ *OR*

2.2.4. Funciones matemáticas:

- ABS (x) $|x|$
- ACOS (x) Arcocoseno de x (x en grados)
- ACOSR (x) Arcocoseno de x (x en radianes)
- ASIN (x) Arcoseno de x (x en grados)
- ASINR (x) Arcoseno de x (x en radianes)
- ATAN (x) Arcotangente de x (x en grados)
- ATANR (x) Arcotangente de x (x en radianes)
- COS (x) Coseno de x (x en grados)
- COSR (x) Coseno de x (x en radianes)
- EXP (x) e^x
- EXP10 (x) 10^x
- FACT (x) Factorial de x
- FISHERZ (x) Ftransformación Z de Fisher para coeficientes de correlación,

$$Z = \frac{1}{2} \ln \frac{1+r}{1-r}$$
- LOG (x) $\ln(x)$
- LOG10 (x) $\log_{10}(x)$
- ROUND (x) Redondeo al entero más próximo
- ROUNDT0 (numvar;decimal) Redondeo al número especificado de decimales
- SIN (x) Seno de x (x en grados)
- SINR (x) Seno de x (x en radianes)
- SQRT (x) \sqrt{x}
- TAN (x) Tangente de x (x en grados)
- TANR (x) Tangente de x (x en radianes)
- TRUNCATE (x) Mayor entero menor o igual a x

2.2.5. Operadores de transformación:

- DATENUM (datevar) Convierte la variable fecha en variable de STATGRAPHICS
- DIFF (numvar) Calcula las diferencias entre valores consecutivos de la variable
- JOIN (var1; var2) Une las dos variable, una a continuación de la otra
- JOIN3 (var1; var2; var3) Une las tres variables, una a continuación de la otra
- JOIN4 (var1; var2; var3; var4) Une las cuatro variables, una a continuación de la otra
- JUXTAPOSE (charvar1; charvar2) Une las dos variables tipo carácter, una al lado de la otra, expandiendo la anchura
- LAG (numvar; n) Desplaza los valores n posiciones
- MDIFF (numvar; n) Calcula las diferencias sucesivas de orden n
- RECODE (var) Codifica a enteros la variable numérica o carácter
- REP (var; n) Repite cada valor n veces
- REPLACE (numvar; old; new) Reemplaza todas las repeticiones de un valor de la variable por otro
- RESHAPE (var; n) Expande o comprime una constante o variable a n valores
- SDIFF (numvar; n) Calcula las diferencias estacionales de orden n
- STRIPBLANKS (charvar) Elimina los dobles blancos de la variable carácter

2.2.6. Funciones de Estadística Descriptiva:

- AVR (numvar) Media
- CV (numvar) Coeficiente de variación
- GEOMEAN (numvar) Media geométrica
- IQR (numvar) Rango intercuartílico
- KURTOSIS (numvar) Coeficiente de curtosis
- MAX (numvar) Máximo
- MEDIAN (numvar) Mediana
- MIN (numvar) Mínimo
- MODE (numvar) Moda

- PERCENTILE (numvar; n) n -ésimo percentil
- Q25 (numvar) Primer cuartil
- Q75 (numvar) Tercer cuartil
- RANGE (numvar) Rango o recorrido
- RANK (numvar) Asigna rangos a las observaciones
- RUNTOT (numvar) Totales acumulados
- SD (numvar) Desviación típica
- ERROR (numvar) Error estándar
- SIZE (numvar) Número total de observaciones
- SKEWNESS (numvar) Coeficiente de asimetría
- SKURT (numvar) Coeficiente estandarizado de curtosis
- SSKEW (numvar) Coeficiente de asimetría estandarizado
- STANDARDIZE (numvar) Estandariza la variable
- SUM (numvar) Suma de las observaciones
- VARIANCE (numvar) Varianza

2.2.7. Operadores de generación:

- COUNT (start; end; step) Crea un vector que comienza en start, finaliza en end, con paso step
- FIRST (n) Genera el valor 1 para las n primeras filas y 0 para el resto
- LAST (n) Genera el valor 1 para las n últimas filas y 0 para el resto
- RANDOM (n) Genera el valor 1 para n filas seleccionadas aleatoriamente y 0 para el resto
- REXPONENTIAL (n; mean) Genera n números aleatorios de una distribución exponencial de media mean
- RGAMMA (n; alpha; beta) Genera n números aleatorios de una distribución gamma de parámetros a y b
- RINTEGER (n; lower; upper) Genera n números aleatorios de una distribución uniforme discreta
- RLOGNORMAL (n; mean; sdev) Genera n números aleatorios de una distribución lognormal

- RNORMAL (n; mean; sdev) Genera n números aleatorios de una distribución normal
- ROWS (n; m) Genera el valor 1 para las filas desde la n hasta la m y 0 para el resto
- RUNIFORM (n; lower; upper) Genera n números aleatorios de una distribución uniforme continua
- RWEIBULL (n; alpha; beta) Genera n números aleatorios de una distribución de Weibull

2.2.8. Operadores de selección:


- CELL (var; n) Selecciona el valor situado en la fila n de la variable
- COMPRESS (var; logical) Selecciona los valores de la variable que verifican la condición
- DROP (var; n) Selecciona y elimina las n primeras filas
- DROPLAST (var; n) Selecciona y elimina las n últimas filas
- EXCLUDE (x) Exclusión booleana de una sola fila
- FIRSTROWS (var; n) Selecciona los n primeros valores y reemplaza el resto por códigos de valores missing
- LASTROWS (var; n) Selecciona los n últimos valores y reemplaza el resto por códigos de valores missing
- SELECT (var; logical) Selecciona los valores que verifican la condición y reemplaza los demás por códigos de valores missing
- TAKE (var; n) Selecciona los n primeros valores
- TAKELAST (var; n) Selecciona los n últimos valores

2.2.9. Funciones de distribución:

- BETA (n; m; l) Calcula la función de distribución de una $Beta(n, m)$ en el punto l
- CHISQUARE (n; df) Calcula la función de distribución de una χ^2_{df} en el punto n
- INVBETA (n; m; l) Calcula el l -cuantil de una distribución $Beta(n, m)$
- INVCHISQUARE (n; df) Calcula el l -cuantil de una distribución χ^2_{df}

- INVNORMAL (n; mean; sdev) Calcula el l -cuantil de una distribución $N(\mu, \sigma)$
- INVSNEDECOR (n; df1; df2) Calcula el l -cuantil de una distribución $F(df1, df2)$
- INVSTUDENT (n; df) Calcula el l -cuantil de una distribución T_{df}
- NORMAL (n; mean; sdev) Calcula la función de distribución de una $N(\mu, \sigma)$ en el punto n
- SNEDECOR (n; df1; df2) Calcula la función de distribución de una $F(df1, df2)$ en el punto n
- STUDENT (n; df) Calcula la función de distribución de una T_{df} en el punto n

2.3. Selección de un fichero ya existente. En ocasiones los datos que se quieren analizar se encuentran ya grabados en un fichero tipo hoja de cálculo o en una base de datos, o se encuentran en el portapapeles de Windows.

En el primer caso su incorporación a STATGRAPHICS es directa. Basta con seleccionar la localización adecuada en la ventana *StatWizard* o pulsar en el icono . Si los datos se encuentran en el portapapeles, debemos indicar si el nombre de las variables se encuentra en la primera fila del portapapales.

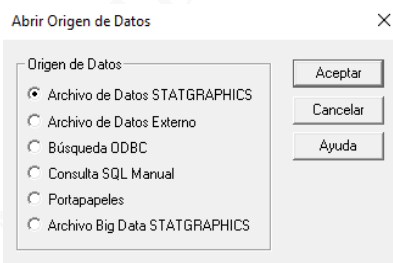


Figura 1.10 Ventana de opciones *Origen Datos*

3. Opciones en los gráficos

En esta sección vamos a analizar la forma en la que el paquete permite realizar algunas operaciones con los gráficos realizados previamente con cualquiera de los procedimientos del programa que los genere.

Como norma general, siempre que se desee realizar algún cambio sobre un gráfico, se debe maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.

3.1. Cambio de títulos y etiquetas. Para cambiar los títulos o etiquetas de los gráficos, la dirección de sus textos, borrarlos, cambiar el borde, ... se deben seguir los siguientes pasos:

1. Maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.
2. Situar el cursor sobre el título o etiqueta que se quiere modificar y se hace click simple con el botón izquierdo del ratón para seleccionarlo.
3. Pulsar el botón derecho del ratón y seleccionar la opción *Opciones Gráficas*. Introducir el nuevo texto en el campo *Título*. En la pestaña *Vertical* se puede elegir esa dirección del texto para el título. La fuente, estilo de fuente, tamaño de la fuente, alfabeto a utilizar por la fuente, efectos para la fuente y el color de la fuente se puede seleccionar pulsando el botón *Línea1 Fuente*.

Para cambiar de lugar un título o etiqueta, basta con seleccionarlo con el botón izquierdo del ratón, y cuando aparece enmarcado se pulsa sobre él en cualquier zona interior del marco y se arrastra a la zona deseada manteniéndolo pulsado. Al soltar el botón del ratón, el título se sitúa en la zona elegida.

3.2. Cambio de las escalas de los ejes. Para cambiar las escalas de los ejes se deben seguir los siguientes pasos:

1. Maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.
2. Situar el cursor sobre cualquier parte del gráfico y pulsar el botón derecho del ratón. Seleccionar *Opciones Gráficas* y pulsar sobre la pestaña del eje que se quiera modificar.
3. Introducir el valor mínimo y máximo para los valores de la variable del eje seleccionado en los campos *Desde* y *Hasta*. En el campo *Por* se introduce el valor del paso, que delimita la distancia entre dos marcas consecutivas. En

el campo *Sin Potencias* se decide si el programa usa o no potencias de 10, en el campo *Log* se indica si se va a utilizar la escala logarítmica en el eje. En el campo *Mantener* se indica si el programa debe retener la escala actual cuando se cambien los datos del análisis. En el campo *Rotar Etiquetas de los Ejes* se indica si el programa debe cambiar la posición de las etiquetas en el eje *X*. En el campo *Límite* se introduce el número de marcas en los ejes que el programa debe saltarse. La fuente, estilo de fuente, tamaño de la fuente, alfabeto a utilizar por la fuente, efectos para la fuente y el color de la fuente para el eje seleccionado se puede elegir pulsando el botón *Título Fuente* o *Marca Fuente*.

3.3. Identificación de puntos. Si se quiere identificar un punto en el gráfico por su número de fila en el fichero de datos debemos de seguir los siguientes pasos:

1. Maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.
2. Situar el cursor sobre el interior del icono *Fila* de la barra de herramientas de la ventana del análisis y se escribe el número de fila de la hoja de cálculo que contiene los datos de las variables cuyos puntos se quieren identificar.
3. Pulsar el icono representado por unos prismáticos y el punto o los puntos seleccionados resaltan sobre el gráfico.

Si lo que deseamos es identificar el número de fila para un punto en el gráfico el camino que debemos seguir es el siguiente:

1. Maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.
2. Situar el cursor sobre el punto del gráfico cuyo correspondiente número de fila se quiere mostrar. Al pulsar sobre él con el botón izquierdo del ratón se inserta en el interior del icono *Fila* de la barra de herramientas el número de la fila de la hoja de cálculo que contiene los datos relativos al punto elegido.

3.4. Añadir texto a un gráfico. Si se desea añadir texto a un gráfico se deben seguir los siguientes pasos:

1. Maximizar el gráfico en cuestión haciendo doble click con el botón izquierdo del ratón sobre cualquier parte del mismo.
2. Situar el cursor sobre el icono *Agregar texto* de la opción **Texto** de la barra de herramientas (ver Figura 1.11).



Figura 1.11 Icono Agregar Texto a un gráfico

3. Introducir el texto del subtítulo en el campo *Texto*. En el campo *Dirección* se puede elegir la dirección del texto para el subtítulo y en *Fuente...* el tipo de letra. El texto se incorpora al gráfico cuando pulsemos el botón *Aceptar*.

3.5. Guardar un gráfico. Para guardar un gráfico como fichero permanente se debe hacer doble click con el botón izquierdo del ratón sobre cualquier parte del mismo y seleccionar la opción *Guardar Gráfica....* En la ventana emergente introducir las propiedades del gráfico

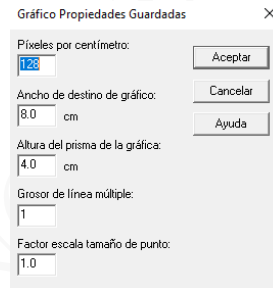


Figura 1.12 Ventana de Propiedades del gráfico que se guarda

A continuación, introducir el nombre para el fichero en el campo *Nombre de archivo* y seleccionar el tipo de gráfico (wmf, jpg, tif, png, bmp o gif) en el campo *Tipo*.

4. Ventana StatGallery

La ventana StatGallery se utiliza para realizar representaciones gráficas según se desee. El proceso que hay que seguir para copiar un gráfico a la StatGallery es el siguiente:

1. Situar sobre el gráfico y seleccionar la subopción *Copiar Ventana a StatGallery...* del menú desplegable que aparece con la opción *Copiar al* que se obtiene al pulsar el botón derecho del ratón sobre el gráfico.

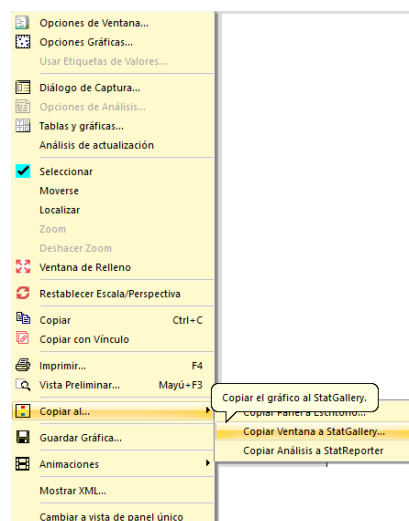


Figura 1.13 Menú de acceso a *Copiar Ventana a Statgallery*

2. Maximizar la ventana StatGallery.
3. Situarnos sobre la zona de la ventana StatGallery donde deseemos colocar el gráfico, pulsar el botón izquierdo del ratón y elegir la opción *Pegar*.

Sobre una misma cuadrícula de la ventana StatGallery se pueden observar simultáneamente varios gráficos distintos, para ello debemos seleccionar la opción *Superponer* que aparece al intentar copiar un gráfico sobre una cuadrícula ocupada. En esta situación, es interesante notar que la escala que se utiliza para todos los gráficos es la correspondiente al primer gráfico.

La subopción *Guardar StatGallery* de la opción *Guardar* del menú *Archivo* permite guardar el contenido actual de la ventana StatGallery en un fichero con extensión *.sgg*. Este fichero se puede incorporar a memoria en cualquier sesión de trabajo mediante la subopción *Abrir StatGallery* de la opción *Abrir* del menú *Archivo*.

5. Uso de la opción *Ayuda*

Cuando se elige la subopción **Ayuda** en el menú *Inicio*, nos aparecen los siguientes procedimientos:

- CONTENIDOS: proporciona el contenido completo de la ayuda para la totalidad de sus comandos y funciones, incluidos ejemplos. Al seleccionar esta opción se obtiene una pantalla de subopciones entre las que se incluyen las opciones propias del comando *Ayuda*.
- MANUAL USUARIO: proporciona el manual de usuario del programa.
- VENTANA ACTIVA: proporciona documentación para tarea actual o comando.
- DOCUMENTACIÓN PROCEDIMIENTO: proporciona información sobre los diversos procedimientos estadísticos del paquete.
- WWW.STATGRAPHICS.COM: abre en un navegador web la página oficial de Statgraphics.
- STATWIZARD: ejecuta este módulo del programa.
- STATADVISOR: abre la ventana *StatAdvisor*.
- ACERCA STATGRAPHICS 19: proporciona información sobre la versión y licencia del programa.
- ADMINISTRAR LICENCIA: proporciona información sobre el estado de la licencia del programa.

Parte 2

Estadística Descriptiva

GRADO EN INGENIERÍA INFORMÁTICA

CAPÍTULO 2


Estadística Descriptiva Univariante

En la mayoría de los procesos estadísticos hay un momento en el que nos encontramos con un conjunto de datos que hay que tratar. En este punto es fundamental el proceso de presentación, descripción numérica y representación gráfica de los datos. Supongamos que los datos los tenemos ya cargados en memoria en un fichero *.sgd*.

La descripción numérica y gráfica de los datos recogidos en una cualquiera de sus variables se realiza principalmente a través de las opciones *Analizar-1*, *Analizar-2* y *Medir* de la barra de menú principal.

1. Descripción estadística de variables unidimensionales

Mediante la opción *Analizar-1* STATGRAPHICS permite analizar y resumir los datos adecuadamente. El submenú correspondiente a esta opción distingue entre variables numéricas (cuantitativas) y cualitativas (datos categóricos).

1.1. Variables numéricas. El análisis descriptivo de una variable cuantitativa se realiza con la subopción *Análisis de Una Variable* de la opción *Datos Variable* del menú *Analizar-1*, o más directamente con el icono .

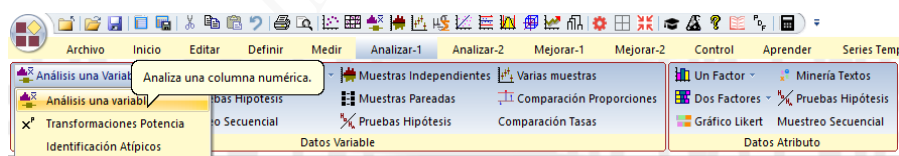


Figura 2.1 Procedimiento para el análisis descriptivo de una variable numérica

El cálculo de la varianza STATGRAPHICS lo realiza utilizando como denominador $n - 1$.

La ventana de entrada de la subopción *Análisis de Una Variable* es la habitual de cualquier procedimiento en STATGRAPHICS y aparece en la Figura 2.2. Los campos y botones de mayor interés son:

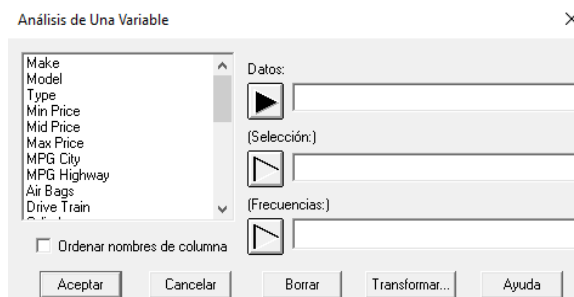


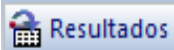
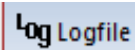


Figura 2.2 Ventana de entrada *Análisis de Una Variable*

- **DATOS:** Donde se introduce el nombre de la variable sobre la que se aplica el procedimiento (se marca la variable en el cuadro de la izquierda y se pincha el triángulo correspondiente al campo).
- **SELECCIÓN:** En este campo se puede introducir cualquier operador de selección de STATGRAPHICS o nombre de variable que acote de alguna forma los datos de la variable introducida en el campo *Datos*.
- **FRECUENCIAS:** En este campo se puede introducir el nombre de la variable que contiene las frecuencias asociadas a la variable sobre la que se aplicará el procedimiento.
- **TRANSFORMAR:** Pulsando sobre este botón se accede a una calculadora con la que se pueden generar nuevas variables a partir de las que actualmente residen en memoria.
- **ACEPTAR:** Pulsando sobre este botón se accede a la ventana de salida de resultados. Su funcionamiento es genérico en todas las aplicaciones de STATGRAPHICS:

1. En la zona izquierda se presentan los resultados del análisis estadístico (zona de texto). En la parte derecha aparecen, si existen, los resultados gráficos (zona de gráficos). Cualquier pantalla gráfica o de texto puede ser maximizada o minimizada con un doble click. Además, un click con el botón derecho sobre una pantalla gráfica o de texto activa un menú con opciones de edición y, lo que es más importante, con la posibilidad *Opciones de Ventana* que permite modificar la salida obtenida en pantalla (por ejemplo, esta opción en un histograma conduce a una ventana en

- la que es posible elegir el número de clases, si las alturas son frecuencias relativas o absolutas, etc).
- Los botones primero y tercero de la opción *Entrada* de la barra de herramientas secundaria están activos. De ellos, el tercero  (*Tablas y Gráficas*) es fundamental ya que permite seleccionar todas las subopciones numéricas y gráficas del análisis. Además, el primero  (*Diálogo de captura*) da acceso a la ventana de entrada al procedimiento.
 - Los botones primero y segundo de la opción *Salidas* de la barra de herramientas secundaria también están activos. El primero,  **Resultados** permite guardar los resultados obtenidos en variables dentro del actual fichero activo. El segundo,  **Log Logfile** guarda el análisis en el *StatLog*.

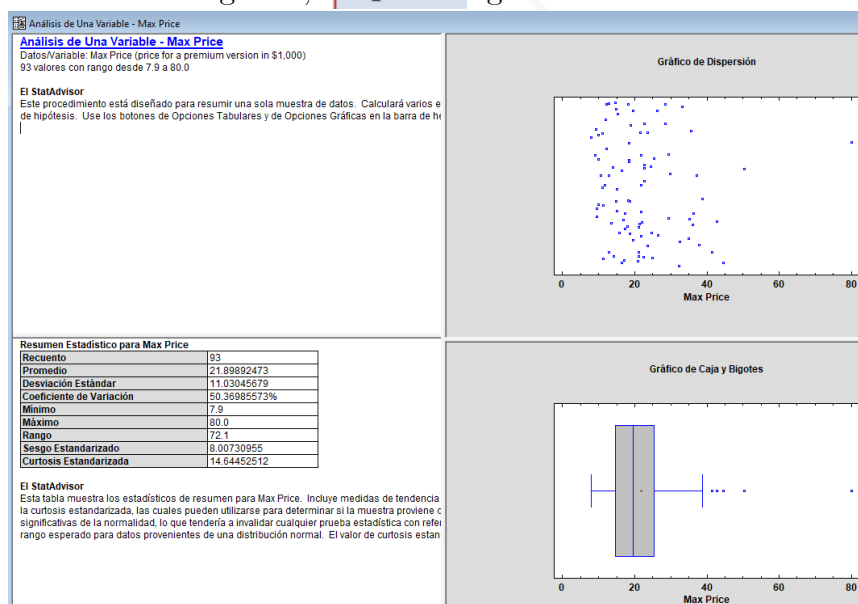


Figura 2.3 Salida del procedimiento *Análisis de Una Variable*

1.1.1. Opciones tabulares y gráficas de Análisis de Una Variable.

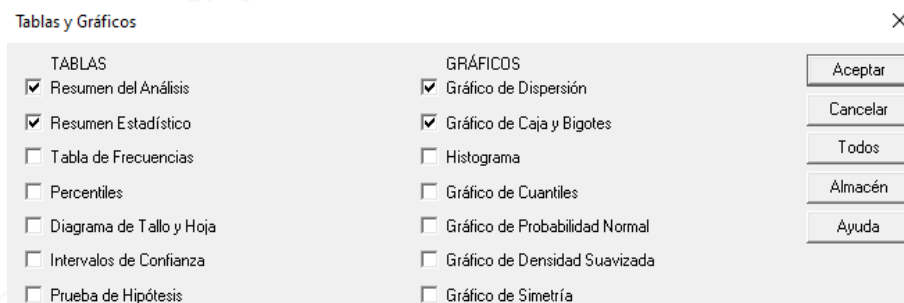


Figura 2.4 Opciones *Tablas y Gráficas* de Análisis de Una Variable

- RESUMEN DEL ANÁLISIS es la opción tabular por defecto y nos presenta, como ya hemos indicado, un resumen de los datos. Del resto de las opciones destacamos:
- RESUMEN ESTADÍSTICO: proporciona diferentes indicadores de posición, dispersión, asimetría y curtosis. Situados sobre esta salida, pulsando el botón derecho del ratón y eligiendo la opción *Opciones de Ventana* se pueden elegir las diferentes medidas que se pueden calcular sobre la variable. Si se necesita calcular otra medida que no se especifica entre las presentadas, puede utilizarse el campo FORMULA del comando *Modificar Columna* de la subopción **Columnas Datos** de la opción *Editar* de la barra de menús del programa o la opción *Generar Datos* también de la subopción **Columnas Datos**.
- TABLA DE FRECUENCIAS: presenta la tabla de frecuencias de la variable especificada. El número de clases que utiliza, por defecto, STATGRAPHIC es el determinado por la *Fórmula de Sturges*: $k = 1 + \lceil \frac{10}{3} \log n \rceil$. Situados sobre esta salida, pulsando el botón derecho del ratón y eligiendo la opción *Opciones de Ventana* se puede elegir el número de clases (NÚMERO DE CLASES), el límite inferior de la primera clase (LÍMITE INFERIOR) y el límite superior de la última clase (LÍMITE SUPERIOR). Si se señala el campo MANTENER se mantiene la definición actual en la siguiente tabla que se realice.
- PERCENTILES: proporciona hasta 10 percentiles de la variable seleccionada.
- DIAGRAMA DE TALLO Y HOJA: proporciona el diagrama de tallo y hojas para la variable seleccionada.

1.2. Variables cualitativas. El análisis descriptivo de una variable cualitativa se realiza con el procedimiento *Tabulación...* de la subopción *Un Factor*, de la opción *Datos Atributo* del menú *Analizar-1*.

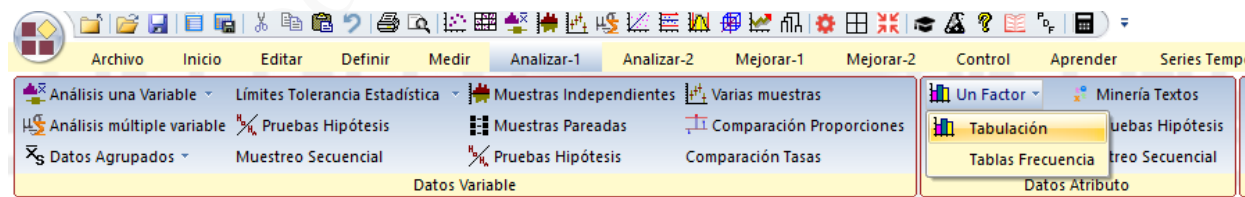


Figura 2.5 Menú de acceso al procedimiento *Tabulación*

La ventana de entrada del procedimiento *Tabulación...* es similar a la explicada anteriormente para *Análisis de una variable* y es la siguiente:

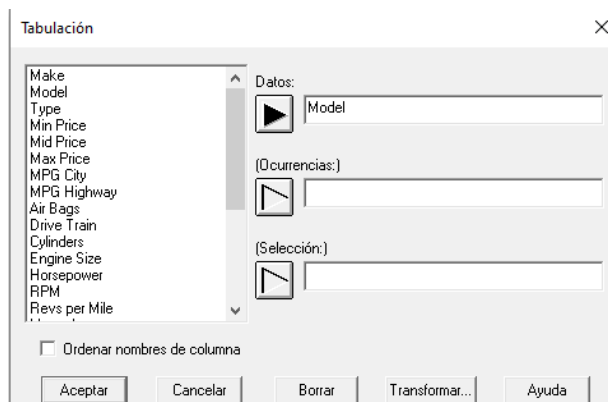


Figura 2.6 Ventana de entrada, procedimiento *Tabulación*

1.2.1. Opciones tabulares y gráficas de *Tabulación...*

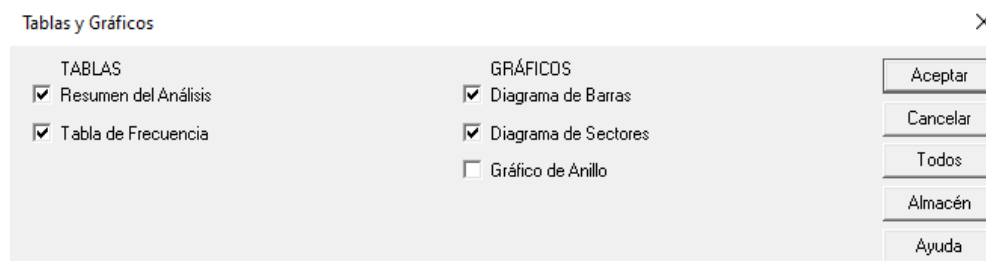


Figura 2.7 Opciones *Tablas y Gráficas*, procedimiento *Tabulación*

La opción tabular TABLA DE FRECUENCIA presenta, como su nombre indica, la tabla de frecuencias para la variable especificada.

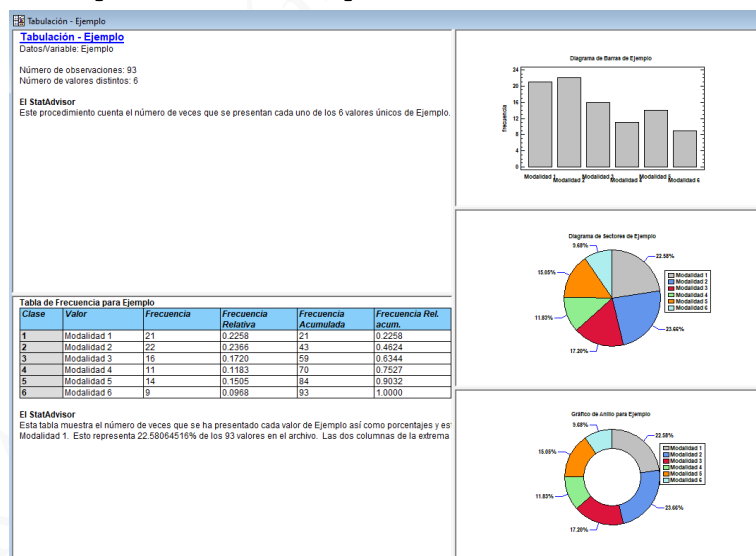


Figura 2.8 Salida del procedimiento *Tabulación*

2. Descripción gráfica de variables unidimensionales

Mediante el menú *Medir* STATGRAPHICS permite realizar diversos tipos de representaciones gráficas de variables estadísticas.

2.1. Opción *Gráficos de Dispersión*. A través de esta opción del menú *Medir*, STATGRAPHICS permite realizar diversos tipos de diagramas de dispersión. La subopción *Gráfico Univariante* proporciona el diagrama de dispersión para una sola variable estadística.

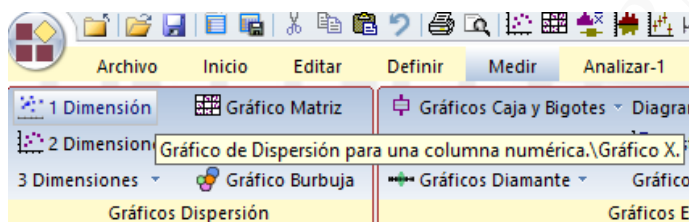


Figura 2.9 Menú de acceso al procedimiento *Gráfico X*

La pantalla de entrada es la que aparece a continuación.

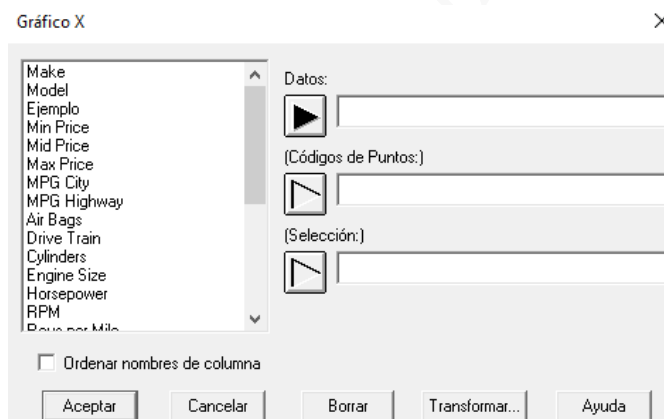


Figura 2.10 Ventana entrada de datos, procedimiento *Gráfico X*

En la opción **CÓDIGO DE PUNTOS** se debe introducir el nombre de la variable que contiene los niveles de la clasificación para el gráfico. Para cada valor distinto de esta variable se utiliza un tipo del punto diferente (ver Figura 2.11).

2.2. Opción *Gráficos Exploratorios*. A través de esta opción STATGRAPHICS posibilita el realizar gráficos de caja para una única variable unidimensional (subopción *Gráfico Caja y Bigotes*). Además en las opciones gráficas de la ventana

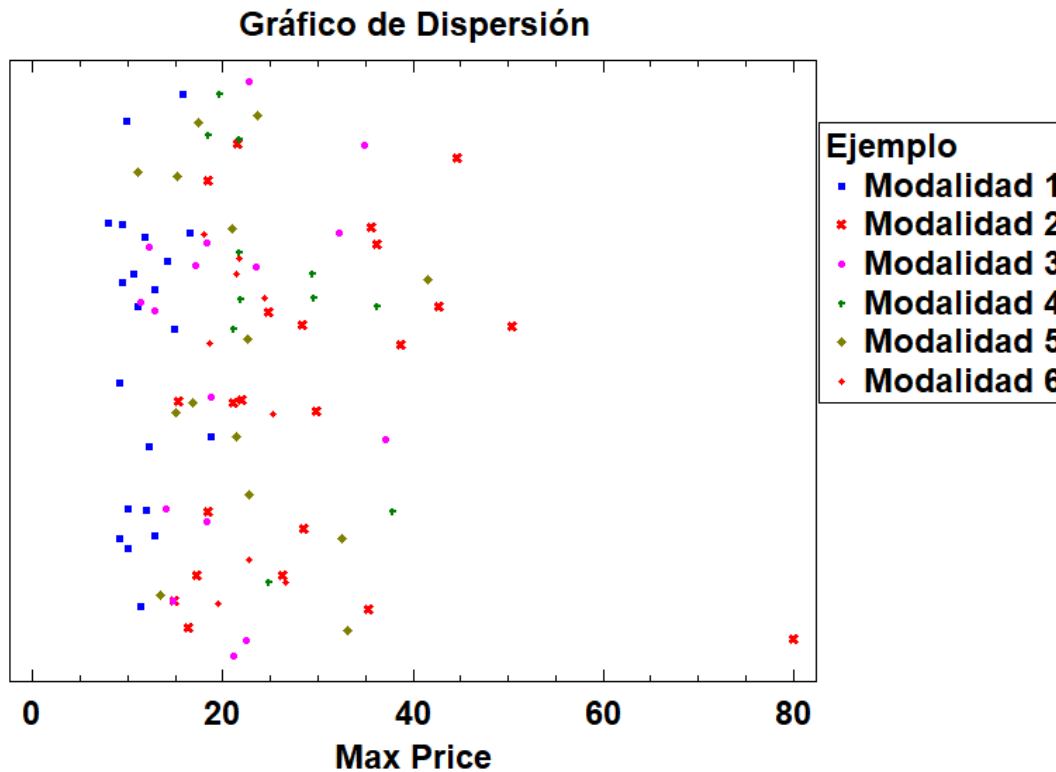


Figura 2.11 Salida del procedimiento *Gráfico X*

de resultados del procedimiento *Análisis de Una Variable* también se permite realizar este diagrama de caja.

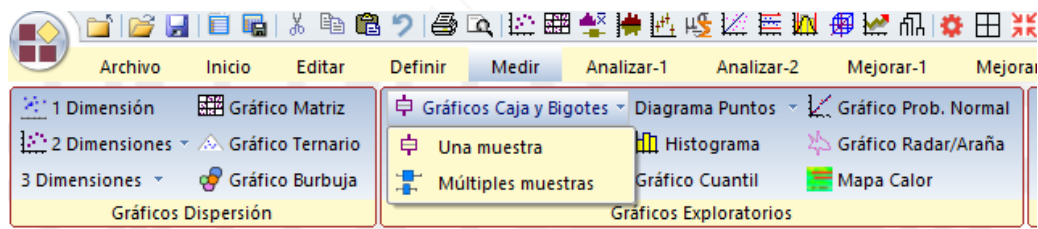


Figura 2.12 Menú de acceso al procedimiento *Gráficos Caja y Bigotes*

La subopción *Histograma* permite realizar histogramas y polígonos de frecuencias absolutas y relativas, que también pueden ser acumulados, para una determinada variable.

2.3. Gráficos para variables categóricas. Mediante el procedimiento *Gráfico Tarta/Donut* de la opción **Gráficos de Negocios** del menú *Analizar-1*, STATGRAPHICS posibilita el realizar gráficos de sectores (o anillos) a partir de las frecuencias absolutas de la variable.

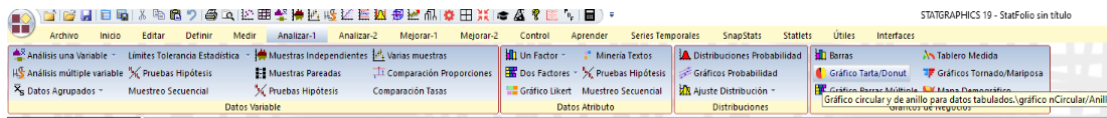


Figura 2.13 Menú de acceso al procedimiento *Gráfico Circular/Anillo*

Es interesante hacer notar que en la ventana de acceso a este procedimiento el campo DATA, datos, es sustituido por el campo CONTEOS. Si no se dispusiera de las frecuencias absolutas deberíamos recurrir a las opciones gráficas de la ventana de resultados del procedimiento *Tabulación...* de la subopción *Un Factor*, de la opción *Datos Atributo* del menú *Analizar-1*.

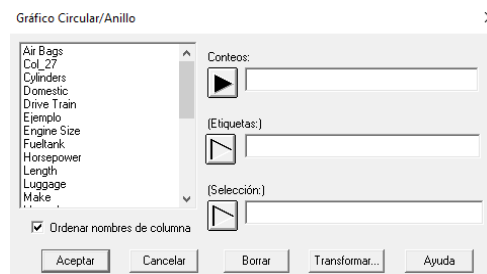



Figura 2.14 Ventana de entrada para *Gráfico Circular/Anillo*

CAPÍTULO 3

Estadística Descriptiva Multivariante

Mediante el menú *Analizar-1* STATGRAPHICS permite también analizar y representar gráficamente variables estadísticas multivariantes. La opción *Medir* es igualmente válida para realizar gráficos de variables estadísticas multidimensionales. Para saber qué tipo de procedimiento podemos utilizar debemos distinguir en primer lugar el tipo de las variables.

1. Variables numéricas

El análisis descriptivo de una variable multidimensional cuantitativa se realiza con el procedimiento *Análisis múltiple variable* de la opción *Datos Variable* del menú *Analizar-1* o con el icono .

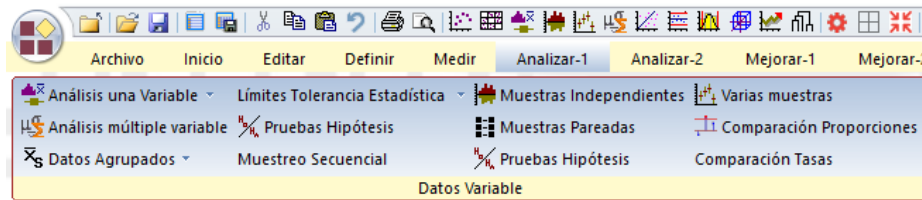


Figura 3.1 Menú de acceso al procedimiento *Análisis múltiple variable*

La ventana de entrada del procedimiento es la habitual de STATGRAPHICS, aunque aquí vamos a señalar que en el campo DATOS debemos introducir el nombre de todas las variables que contienen los datos que van a ser examinados (en estas variables los datos se encuentran listados uno a uno). Este procedimiento tiene dos opciones de análisis:

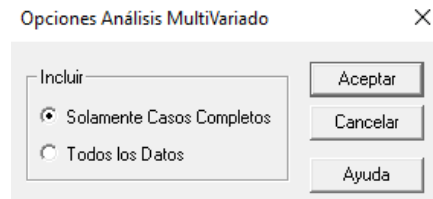


Figura 3.2 Ventana de entrada, procedimiento *Análisis múltiple variable*

- SOLAMENTE CASOS COMPLETOS: excluye del análisis a cualquier fila en la cual una o más de las columnas de datos de entrada contiene un valor faltante. En este caso, todos los análisis estarán basados en el mismo número de filas.
- TODOS LOS DATOS: utiliza todos los datos cuando sea posible.

1.1. Opciones tabulares y gráficas de *Análisis múltiple variable*. Como en todo análisis con STATGRAPHICS en la barra de herramientas aparece el icono *Tablas y Gráficos*, que permite realizar todas las subopciones tabulares y gráficas, entre las que destacamos:

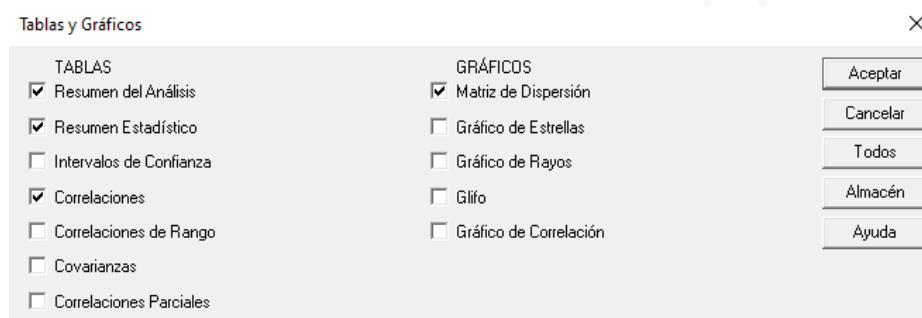


Figura 3.3 Opciones *Tablas y Gráficos*, procedimiento Análisis múltiple variable

1.1.1. Opciones tabulares.

- RESUMEN ESTADÍSTICO.
- CORRELACIONES: proporciona la matriz de coeficientes de correlación para las variables dadas. En la celda (i, j) de la matriz se presentan tres valores, el primero de ellos es el coeficiente de correlación r_{ij} entre las variables X_i y X_j .
- COVARIANZAS: proporciona la matriz de covarianzas para cada par de variables dadas. En la celda (i, j) de la matriz se presentan dos valores, el primero de ellos es la covarianza S_{ij} entre las variables X_i y X_j .

1.1.2. Opciones tabulares. De las cuatro subopciones gráficas, conviene destacar:

- MATRIZ DE DISPERSIÓN: permite realizar una matriz de gráficos de dispersión para todos los pares posibles de variables. Esta representación gráfica está también disponible en la subopción *Gráfico Matriz* de la opción *Gráficos de Dispersión* del menú *Medir*.

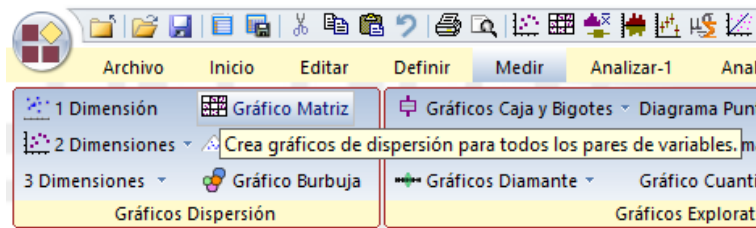


Figura 3.4 Procedimiento *Gráfico Matriz* en el menú **Medir**

1.2. Opciones del menú *Medir*. Como hemos indicado anteriormente, la opción *Medir* puede ser utilizada también cuando se desee realizar representaciones gráficas de variables multidimensionales.

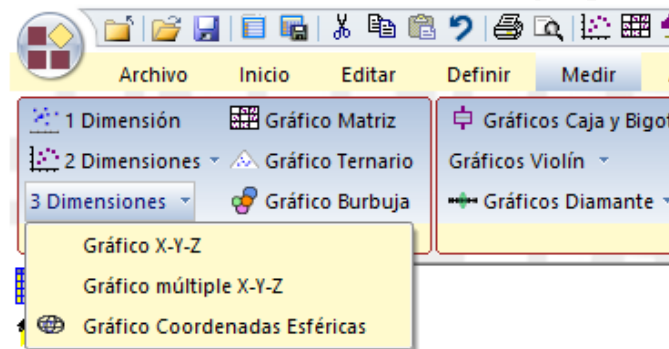


Figura 3.5 Menú de acceso a gráficos tridimensionales

La opción *Gráficos de Dispersión* permite realizar, además de los diagramas de dispersión comentados anteriormente, diagramas en tres dimensiones mediante, por ejemplo, el procedimiento *Gráfico X-Y-Z* (ver Figura 3.6) o el procedimiento *Gráfico múltiple X-Y-Z* (ver Figura 3.7)

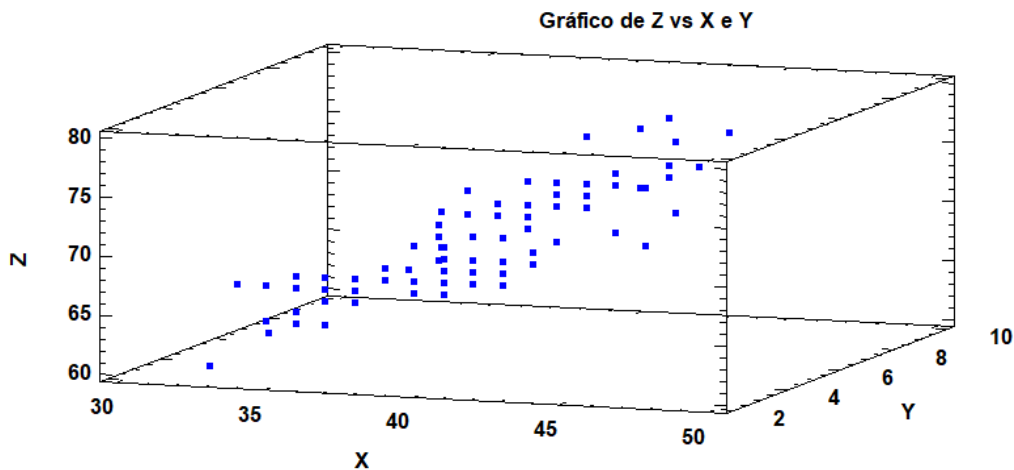


Figura 3.6 Salida del procedimiento *Gráfico X-Y-Z*

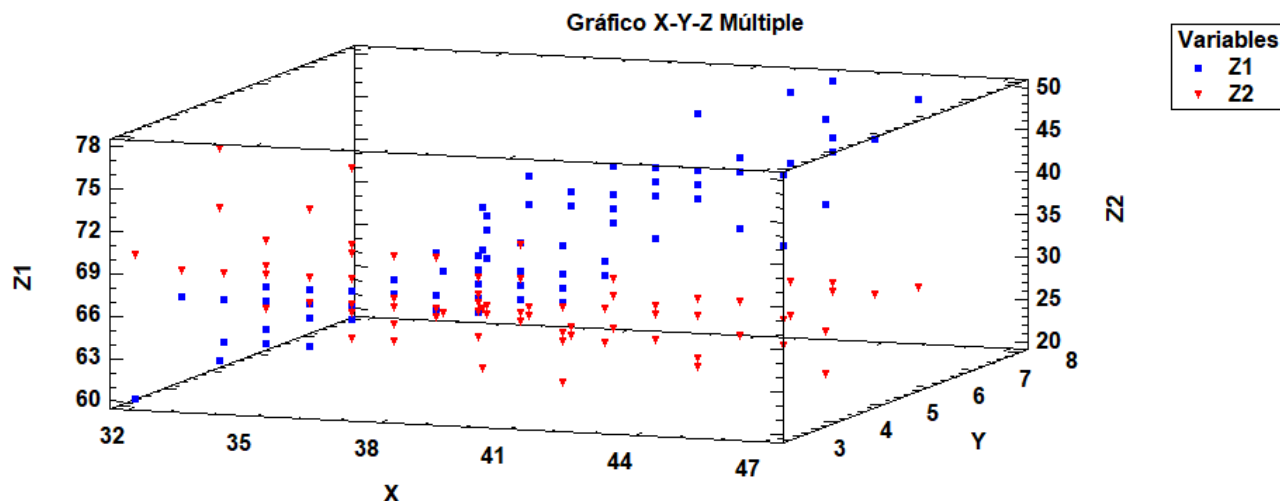


Figura 3.7 Salida del procedimiento *Gráfico múltiple X-Y-Z*

La opción *Gráficos Exploratorios* con la subopción *Gráfico de Caja y Bigotes* y el procedimiento *Múltiples Muestras* permite realizar conjuntamente los diagramas de caja de varias variables estadísticas. La ventana de entrada de datos se muestra en la Figura 3.8.



Figura 3.8 Ventana de entrada de datos del procedimiento *Gráfico Múltiple de Caja y Bigotes*

En el campo DATOS se introduce el nombre de la variable que se va a representar y en el campo CÓDIGOS POR NIVEL el nombre de la variable que contiene a los identificadores de grupo.

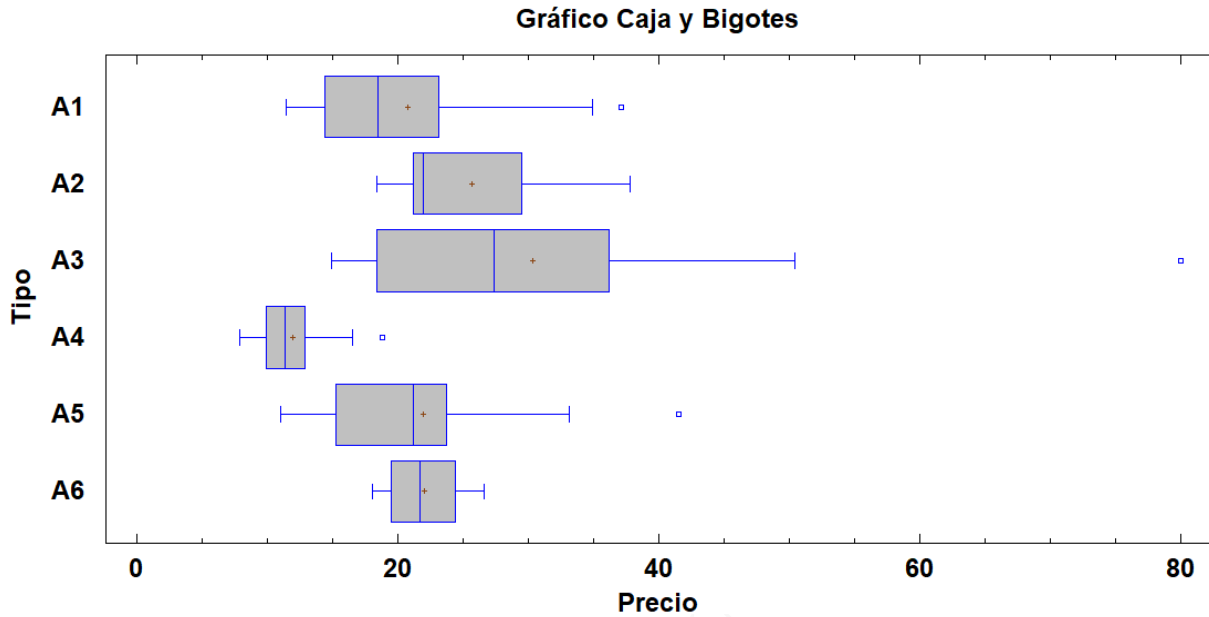



Figura 3.9 Salida del procedimiento *Gráfico Múltiple de Caja y Bigotes*

1.3. Comparación de dos variables. Cuando se desean comparar dos variables estadísticas a nivel descriptivo se puede utilizar también el procedimiento *Muestras Independientes* de la opción *Datos Variable* del menú *Analizar-1* de STAT-GRAPHICS o con el icono .

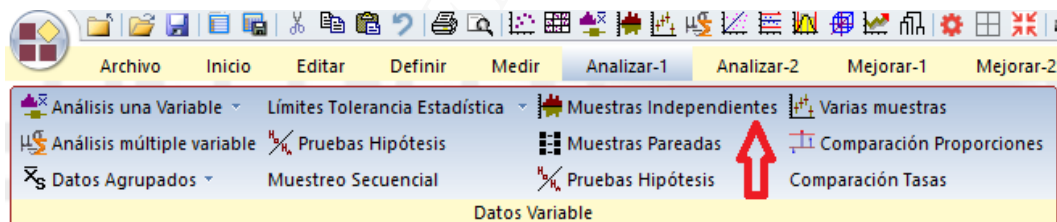


Figura 3.10 Menú de acceso al procedimiento *Muestras Independientes*

En la pantalla de entrada (ver Figura 3.11) aparecen los campos MUESTRA 1: donde se introduce el nombre de la variable que contiene los datos de la primera muestra, MUESTRA 2: donde escribiremos el nombre de la variable que contiene los datos de la segunda muestra y ENTRADA: donde se elige la forma de introducir los datos; con la opción *Dos Columnas de Datos* los datos de las dos muestras se introducen en dos columnas de la hoja de cálculo y con la opción *Columnas de Código y Datos* se introducen los datos en una columna y los códigos de nivel en otra.

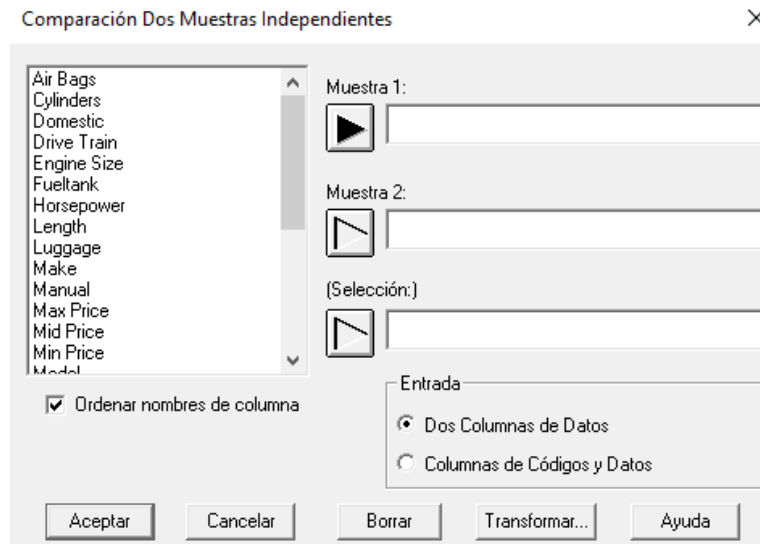


Figura 3.11 Ventana de entrada de datos, procedimiento *Muestras Independientes*

La opción tabular *Resumen Estadístico* proporciona diferentes indicadores de posición, dispersión, asimetría y curtosis para las dos variables. Mediante la opción *Opciones de Ventana* se pueden elegir los indicadores que deseemos calcular para ambas variables.

Las opciones gráficas de mayor interés son las siguientes:

- HISTOGRAMAS: dibuja conjuntamente los histogramas de las dos variables, el de una variable en la parte de arriba y el de la otra en la parte de abajo.

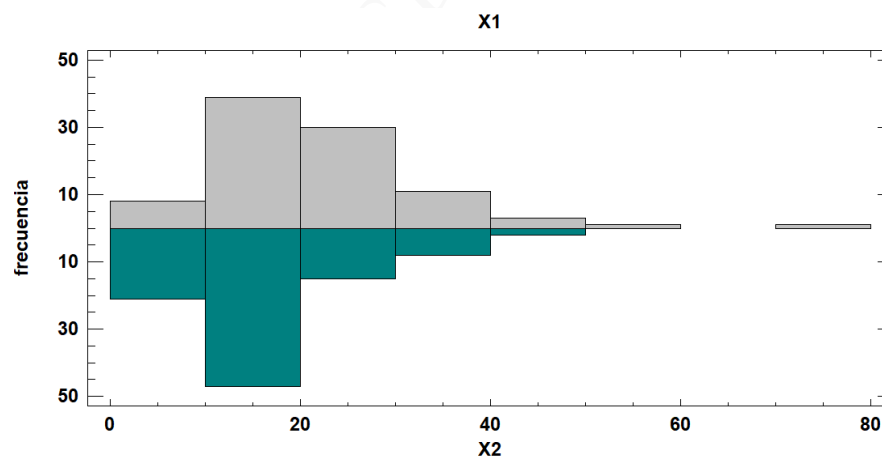


Figura 3.12 Histograma, procedimiento *Muestras Independientes*

- GRÁFICOS DE CAJA Y BIGOTE: realiza los dos diagramas de caja, uno para cada variable.

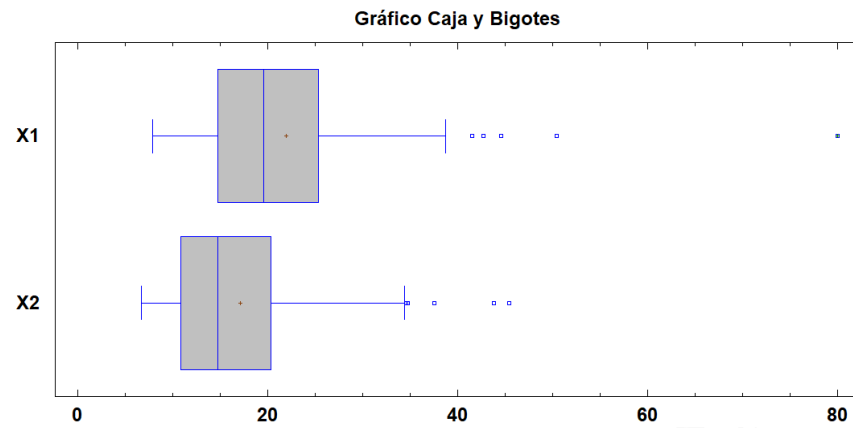


Figura 3.13 Gráfico Caja y Bigotes, procedimiento *Muestras Independientes*

- GRÁFICO DE CUANTILES: muestra las distribuciones acumulativas de los datos.

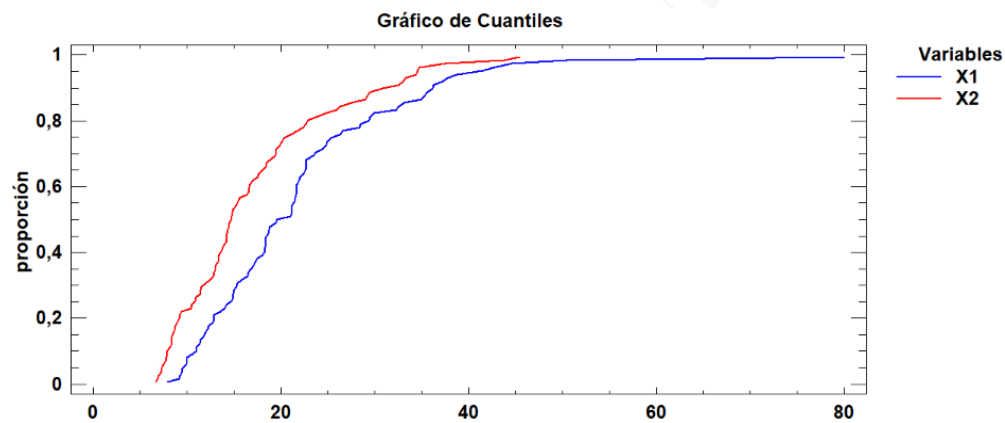


Figura 3.14 Gráfico cuantiles, procedimiento *Muestras Independientes*

- GRÁFICO CUANTIL-CUANTIL: dibuja los cuantiles de una variable frente a los cuantiles de la otra.

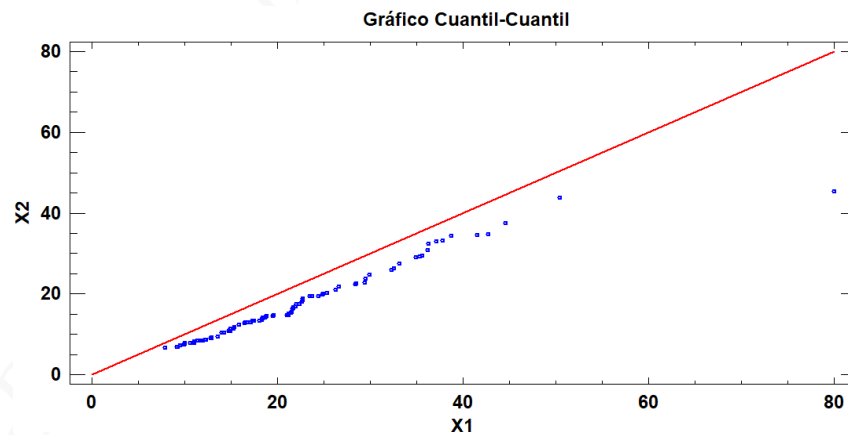


Figura 3.15 Gráfico cuantil-cuantil, procedimiento *Muestras Independientes*

1.4. Análisis de variables por subconjuntos de datos. El análisis de una variable estadística separado para subconjuntos de sus datos se puede realizar mediante el procedimiento *Análisis Subgrupos* de la subopción *Datos Agrupados* de la opción *Datos Variable* del menú *Analizar-1*.

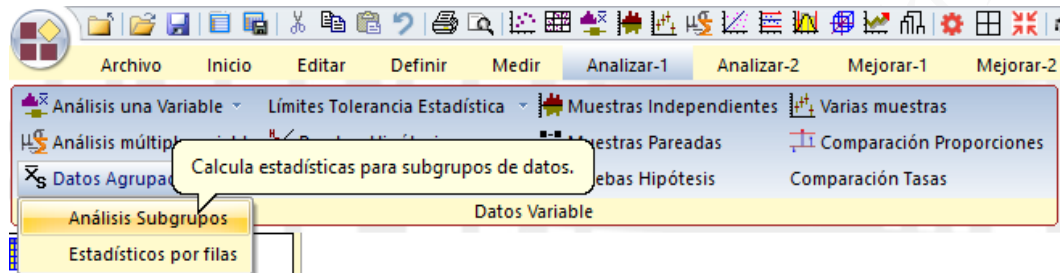


Figura 3.16 Menú de acceso al procedimiento *Análisis Subgrupos*

La ventana de entrada de la subopción contiene los campos DATOS: donde debemos introducir el nombre de la variable que contiene los datos que van a ser examinados, CÓDIGOS: donde se introduce el nombre de la variable que contiene los códigos de clasificación a utilizar para definir los subconjuntos de los valores de la variable indicada en el campo anterior, ETIQUETAS: donde se puede introducir el nombre de la variable que contiene las etiquetas para cada código de clasificación y SELECCIÓN: donde se puede introducir un operador de selección.

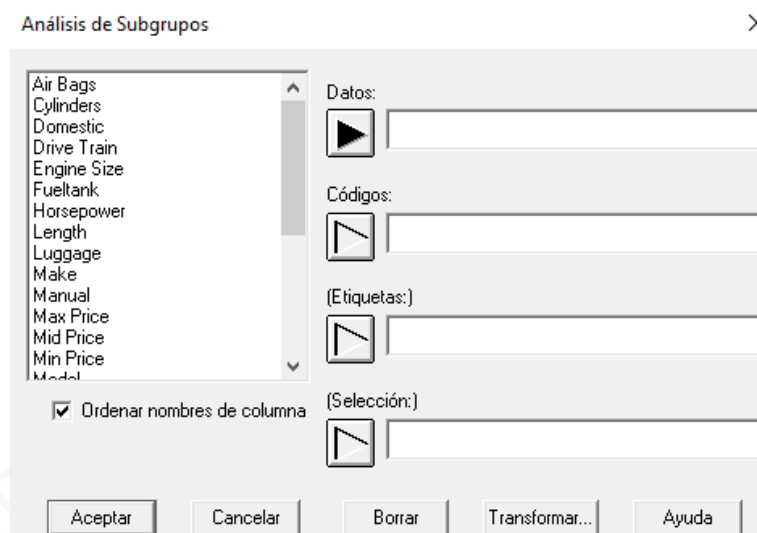



Figura 3.17 Ventana de entrada de datos, procedimiento *Análisis Subgrupos*

2. Regresión simple

El procedimiento *Regresión Simple* de la subopción *Modelos Reg. Simple* de la opción *Regresión Simple* del menú *Mejorar-1* sirve para obtener la regresión de una sola variable independiente, utilizando el criterio de mínimos cuadrados. También se puede acceder pulsando sobre el icono .

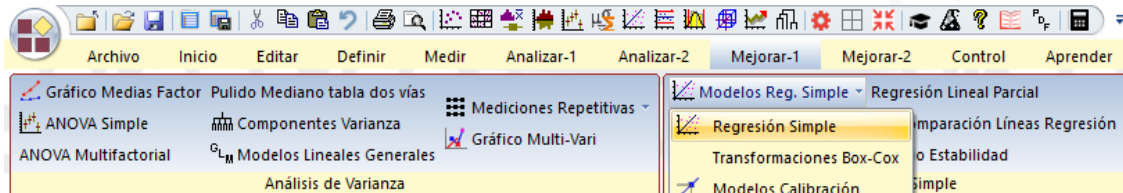


Figura 3.18 Menú de acceso al procedimiento *Regresión Simple*

La ventana de entrada de la opción tiene los campos Y: donde debemos introducir el nombre de la variable dependiente Y, X: donde se debe introducir el nombre de la variable independiente X.

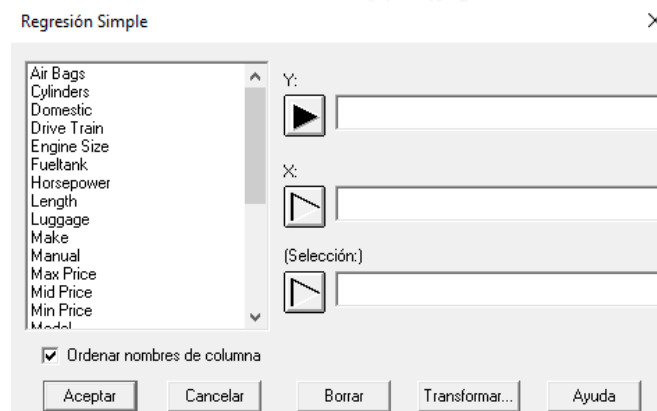


Figura 3.19 Ventana de entrada de datos, procedimiento *Regresión Simple*

Presenta las siguientes opciones de análisis:

- TIPO DE MODELO: se refiere al modelo que va a ser estimado. Todos los modelos mostrados pueden ser linealizados mediante la transformación ya sea de X o de Y, o de ambos. Cuando ajusta un modelo lineal, STATGRAPHICS primero transforma los datos, después ajusta el modelo y finalmente invierte la transformación para mostrar los resultados.
- AJUSTES ALTERNOS: se refiere a un procedimiento alternativo de estimación.

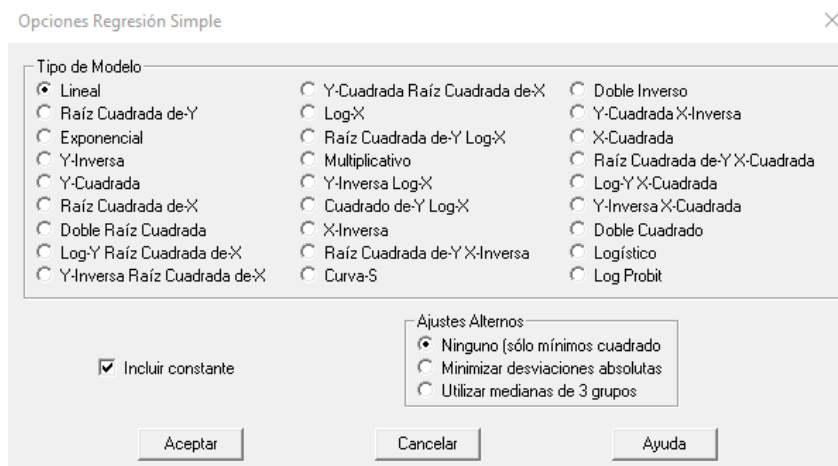


Figura 3.20 Opciones de análisis, procedimiento *Regresión Simple*

2.1. Opciones tabulares de *Regresión Simple*. Si pulsamos sobre el icono *Opciones Tablas y Gráficos* de la barra de herramientas del análisis obtenemos todas las subopciones tabulares, entre las que destacamos:

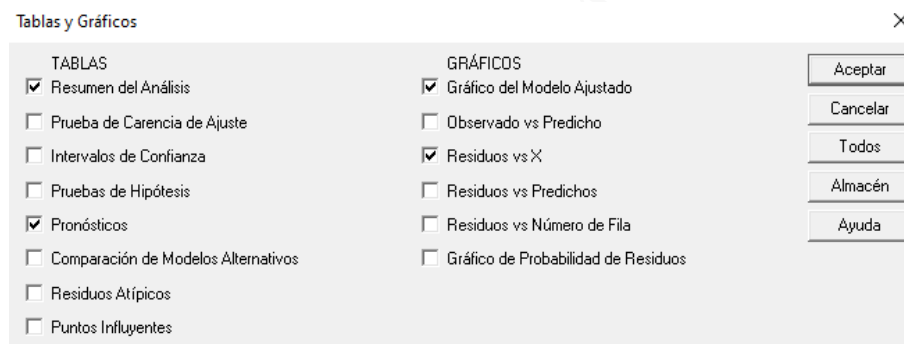


Figura 3.21 Opciones *Tablas y Gráficos*, procedimiento *Regresión Simple*

- **RESUMEN DEL ANÁLISIS:** presenta la ecuación de la recta de regresión, el coeficiente de correlación y el coeficiente de determinación R^2 (ver Figura 3.22). Si estando sobre esta pantalla pulsamos el botón derecho del ratón y en el menú desplegable seleccionamos la opción *Opciones de Análisis* se obtiene una ventana de selección del tipo de modelo para el ajuste.
- **PRONÓSTICOS:** calcula los valores de Y basados en valores de X y muestra las predicciones realizadas por el modelo ajustado para las abscisas especificadas (ver Figura 3.23) en la ventana que resulta de activar *Opciones de Ventana*. Si activamos *Opciones de Análisis* se obtiene una ventana de selección del tipo de modelo para el ajuste.

Regresión Simple - Mid Price vs. Min Price

Regresión Simple - Mid Price vs. Min Price

Variable dependiente: Mid Price (average of min and max prices in \$1,000)
 Variable independiente: Min Price (price for basic version in \$1,000)
 Lineal: $Y = a + bX$
 Número de observaciones: 93

Parámetro	Mínimos Cuadrados Estimado	Estándar Error	Estadístico T	Valor-P
Intercepto	1.151375227	0.5352718882	2.151010081	0.0341
Pendiente	1.071967165	0.02786650978	38.46793781	0.0000

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	8086.723881	1	8086.723881	1479.78	0.0000
Residuo	497.2974089	91	5.464806691		
Total (Corr.)	8584.02129	92			

Coefficiente de Correlación = 0.9706014018
 R-cuadrada = 94.20670811 por ciento
 R-cuadrado (ajustado para g.l.) = 94.14304557 por ciento
 R-cuadrado (predicho) = 93.44650933 por ciento (PRESS = 562.5530341)
 Error estándar del est. = 2.3376926
 Error absoluto medio = 1.571171861
 Estadístico Durbin-Watson = 1.742333399 (P=0.0974)
 Autocorrelación de residuos en retraso 1 = 0.1269023772

El StatAdvisor
 La salida muestra los resultados de ajustar un modelo para describir la relación entre Mid Price y Min Price. La ecuación del modelo ajustado es

Mid Price = 1.151375227 + 1.071967165*Min Price

Puesto que el valor-P en la tabla ANOVA es menor que 0.05, existe una relación estadísticamente significativa entre Mid Price y Min Price con un nivel de confianza del 95.0%.

El estadístico R-Cuadrada indica que el modelo ajustado explica 94.20670811% de la variabilidad en Mid Price. El coeficiente de correlación es igual a 0.9706014018, indicando una relación relativamente fuerte entre las variables. El error estándar del estimado indica que la desviación estándar de los residuos es 2.3376926. Este valor puede usarse para construir límites de predicción para nuevas observaciones, seleccionando la opción de Pronósticos del menú de texto.

El error absoluto medio (MAE) de 1.571171861 es el valor promedio de los residuos. El estadístico de Durbin-Watson (DW) examina los residuos para determinar si hay alguna correlación significativa basada en el orden en el que se presentan en el archivo de datos. Puesto que el valor-P es mayor que 0.05, no hay indicación de una autocorrelación serial en los residuos con un nivel de confianza del 95.0%.

Figura 3.22 Salida del procedimiento *Regresión Simple*

Valores Predichos

X	Pronosticado Y	Inferior 95% Límite Pred.	Superior 95% Límite Pred.	Inferior 95% Límite Conf.	Superior 95% Límite Conf.
7.4	9.083932249	4.384546713	13.78331779	8.361655251	9.806209247
61.9	67.50614275	62.22060541	72.79168008	64.98139235	70.03089314

El StatAdvisor
 Esta tabla muestra los valores predichos para Mid Price usando el modelo ajustado. Además de las mejores predicciones, la

(1) intervalos de previsión del 95.0% para las nuevas observaciones

(2) intervalos de confianza del 95.0% para la media de varias observaciones

Los intervalos de predicción y de confianza corresponden a las cotas internas y externas en la gráfica del modelo ajustado.

Figura 3.23 *Pronósticos*, procedimiento Regresión Simple

- COMPARACIÓN DE MODELOS ALTERNATIVOS: calcula y muestra los coeficientes de correlación y valores R^2 de los ajustes a cada uno de los modelos alternativos seleccionados (ver Figura 3.25) con la opción *Opciones de Análisis* ya comentada.
- RESIDUOS ATÍPICOS: muestra las observaciones cuyo residuo estudentizado tiene valor absoluto mayor que dos (ver Figura 3.24).

Residuos Atípicos

Fila	X	Y	Predicciones Y	Residuos	Residuos Studentizados
28	18.5	25.8	20.98276778	4.817232218	2.11
59	43.8	61.9	48.10353706	13.79646294	8.26
78	20.3	28.7	22.91230868	5.787691321	2.57

El StatAdvisor
 La tabla de residuos atípicos enlista todas las observaciones que tienen residuos Estudentizados mayores a 2, en valor absoluto.

Figura 3.24 *Residuos Atípicos*, procedimiento Regresión Simple

Comparación de Modelos Alternos		
Modelo	Correlación	R-Cuadrada
Multiplicativa	0.9752	95.11%
Raiz Cuadrada Doble	0.9751	95.08%
Lineal	0.9706	94.21%
Raiz Cuadrada de Y	0.9700	94.09%
Logarítmico-Y Raiz Cuadrada-X	0.9688	93.86%
Doble Inverso	0.9676	93.62%
Raiz Cuadrada-Y Log-X	0.9639	92.91%
Raiz Cuadrada de X	0.9580	91.78%
Cuadrado de X	0.9545	91.11%
Exponencial	0.9464	89.57%
Cuadrado Doble	0.9437	89.05%
Curva S	-0.9422	88.78%
Inversa-Y Log-X	-0.9414	88.62%
Logaritmo de X	0.9298	86.45%
Raiz Cuadrada-X Cuadrado-X	0.9219	84.99%
Inversa-Y Raiz Cuadrada-X	-0.9042	81.76%
Raiz Cuadrada-Y Inversa de X	-0.9003	81.06%
Cuadrado de Y	0.8992	80.85%
Log-Y Cuadrado-X	0.8693	75.57%
Cuadrado-Y Raiz Cuadrada-X	0.8568	73.41%
Inversa de X	-0.8388	70.36%
Cuadrado-Y Log-X	0.8024	64.38%
Inversa-Y Cuadrado-X	-0.7351	54.03%
Cuadrado-Y Inversa de X	-0.6761	45.72%
Inversa de Y	<sin ajuste>	
Logístico	<sin ajuste>	
Log probit	<sin ajuste>	

Figura 3.25 *Comparación de Modelos Alternativos*, procedimiento Regresión Simple

- PUNTOS INFLUYENTES: lista todas las observaciones que tienen valores de influencia superior a 3 veces la del punto medio de los datos.

Puntos Influyentes					
Fila	X	Y	Predicciones Y	Residuos Studentizados	Influencia
11	37.5	40.1	41.35014392	-0.55	0.0697389951
48	45.4	47.9	49.81868452	-0.88	0.1243507193
59	43.8	61.9	48.10353706	8.26	0.1118577613

Influencia Media de un punto = 0.02150537634

Figura 3.26 *Puntos Influyentes*, procedimiento Regresión Simple

2.2. Opciones gráficas de Regresión Simple. Entre las opciones gráficas de este procedimiento destacamos:

- GRÁFICO DEL MODELO AJUSTADO: dibuja el modelo ajustado sobre la nube de puntos. Para clarificar el gráfico se deben desmarcar las opciones *Límites de Predicción* y *Límites de Confianza* que resultan al seleccionar *Opciones de Ventana*. Se puede visualizar la influencia que ejerce cualquier punto sobre el ajuste. Para ello, se debe en primer lugar marcar el punto con el botón derecho del ratón (click sobre él) y después activar el décimo botón de la barra de herramientas del procedimiento; de esta forma se obtiene el ajuste que resulta de eliminar en la nube el punto marcado.
- RESIDUOS FRENTE A X: dibuja los residuos estandarizados frente a los valores de la variable independiente.

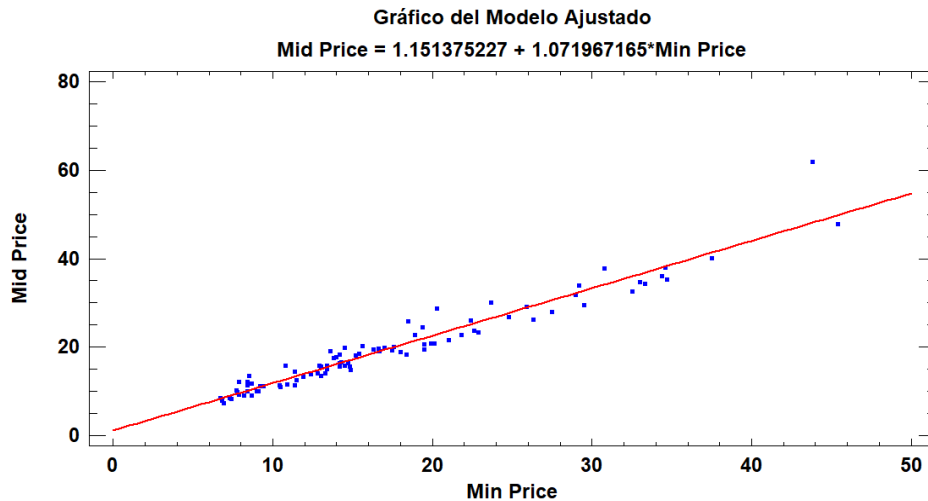


Figura 3.27 *Gráfico del Modelos Ajustado*, procedimiento Regresión Simple

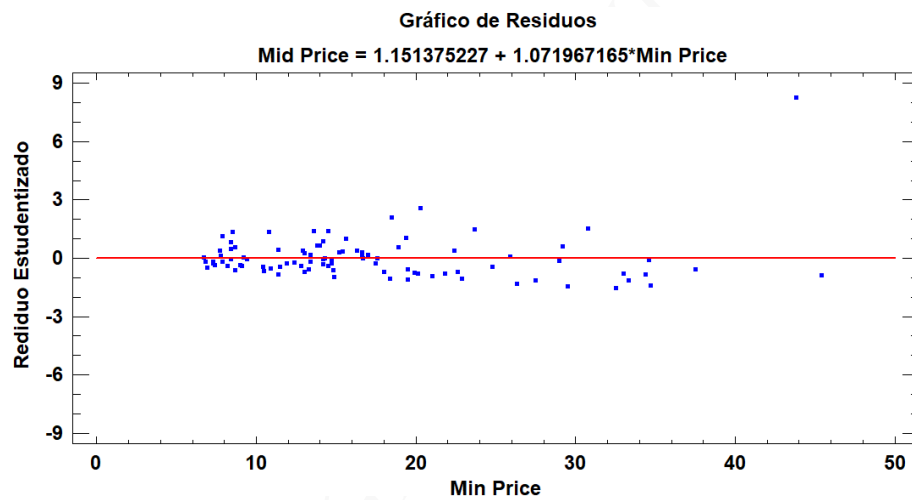


Figura 3.28 *Gráfico de Residuos vs X*, procedimiento Regresión Simple

3. Variables cualitativas

El análisis descriptivo de una variable cualitativa bidimensional se realiza con el procedimiento *Tabulación Cruzada* de la subopción *Dos Factores* de la opción *Datos Atributo* del menú *Analizar-1*.

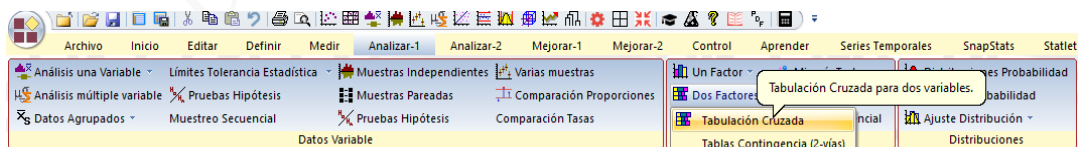


Figura 3.29 Menú de acceso al procedimiento *Tabulación Cruzada*

El procedimiento tiene dos opciones de entrada, habitualmente hay que seleccionar la primera, *Dos columnas de código*.

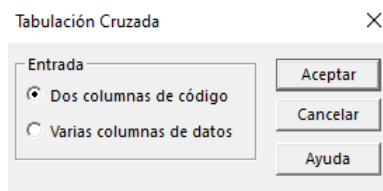


Figura 3.30 Opciones de entrada al procedimiento *Tabulación Cruzada*

La ventana de entrada de la subopción tiene los campos VARIABLE FILA: donde se introduce el nombre de la variable que contiene los valores que van a aparecer en las filas, VARIABLE COLUMNA: donde se introduce el nombre de la variable cuyos valores van a aparecer en las columnas. Es interesante notar que en las variables especificadas en los campos anteriores deben aparecer los valores concretos de las variables; es decir, si la modalidad C_1 de la variable X aparece n_1 veces, en la columna correspondiente a X debemos escribir n_1 veces el valor C_1 .

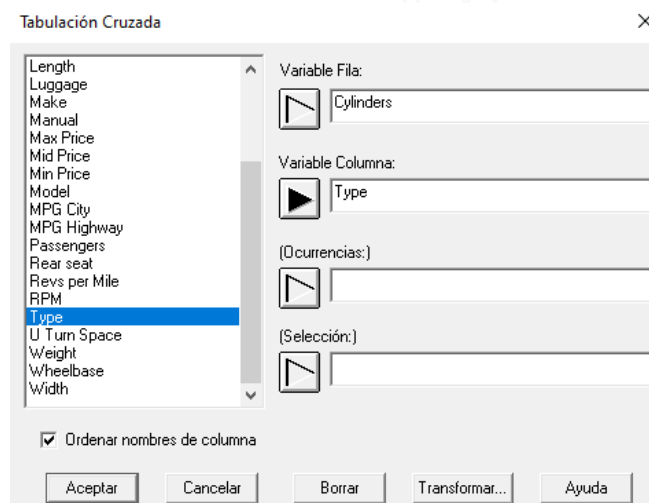


Figura 3.31 Entrada de datos, procedimiento *Tabulación Cruzada*

3.1. Opciones tabulares de *Tabulación Cruzada*. Dentro de las subopciones tabulares destacamos:

- TABLA DE FRECUENCIAS: presenta la tabla bidimensional de frecuencias de las dos variables especificadas. Situados sobre esta tabla, pulsando el botón

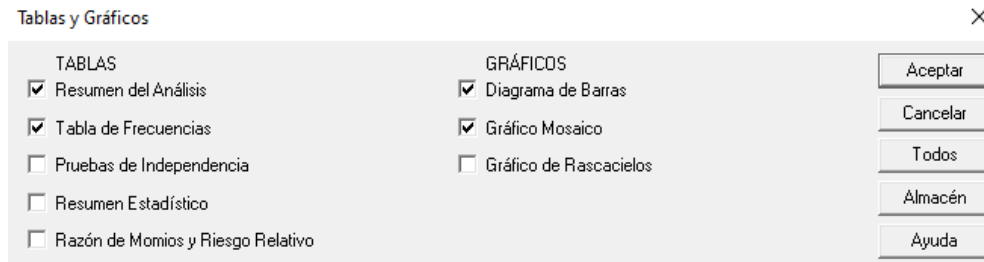


Figura 3.32 Opciones *Tablas y Gráficos*, procedimiento Tabulación Cruzada

derecho del ratón y eligiendo la opción *Opciones de Ventana* se pueden obtener las frecuencias de las celdas en porcentajes respecto al total de la fila (*Porcentajes de fila*), de la columna (*Porcentajes de columna*) o de la tabla (*Porcentajes de la tabla*).

	Compact	Large	Midsize	Small	Sporty	Van	Total por Fila
3	0	0	0	3	0	0	3
	0.00%	0.00%	0.00%	3.26%	0.00%	0.00%	3.26%
4	15	0	7	18	8	1	49
	16.30%	0.00%	7.61%	19.57%	8.70%	1.09%	53.26%
5	0	0	1	0	0	1	2
	0.00%	0.00%	1.09%	0.00%	0.00%	1.09%	2.17%
6	1	7	12	0	4	7	31
	1.09%	7.61%	13.04%	0.00%	4.35%	7.61%	33.70%
8	0	4	2	0	1	0	7
	0.00%	4.35%	2.17%	0.00%	1.09%	0.00%	7.61%
Total por Columna	16	11	22	21	13	9	92
	17.39%	11.96%	23.91%	22.83%	14.13%	9.78%	100.00%

Contenido de las celdas:
Frecuencia Observada
Porcentaje de la Tabla

Opciones Tabla de Frecuencias

Incluir

☐ Porcentajes de la Tabla

☒ Porcentajes por Fila

☒ Porcentajes por Columna

☐ Frecuencias Esperadas

☐ Desviaciones

☐ Valores de Chi-Cuadrada

☐ Residuos Ajustados

Aceptar Cancelar Ayuda

Figura 3.33 Opciones *Tabla de Frecuencias*, procedimiento Tabulación Cruzada

3.2. Opciones gráficas de Tabulación Cruzada. Este procedimiento permite realizar las siguientes opciones gráficas:

- **DIAGRAMA DE BARRAS:** permite realizar para las variables seleccionadas un gráfico de barras bidimensional. Situados sobre este gráfico, pulsando el botón derecho del ratón y eligiendo la opción *Opciones de Ventana* se permite seleccionar un formato para las barras (*Tipo de Gráfica*), utilizar frecuencias absolutas o relativas (*Escala*), dibujar las barras en sentido vertical u horizontal (*Dirección*) e indicar el punto de partida de las barras (*Línea Base*).
- **GRÁFICO DE MOSAICO:** permite realizar una variación del diagrama de barras en el que lo ancho y la altura de cada barra son escalados para representar las frecuencias de las celdas correspondientes en la tabla.

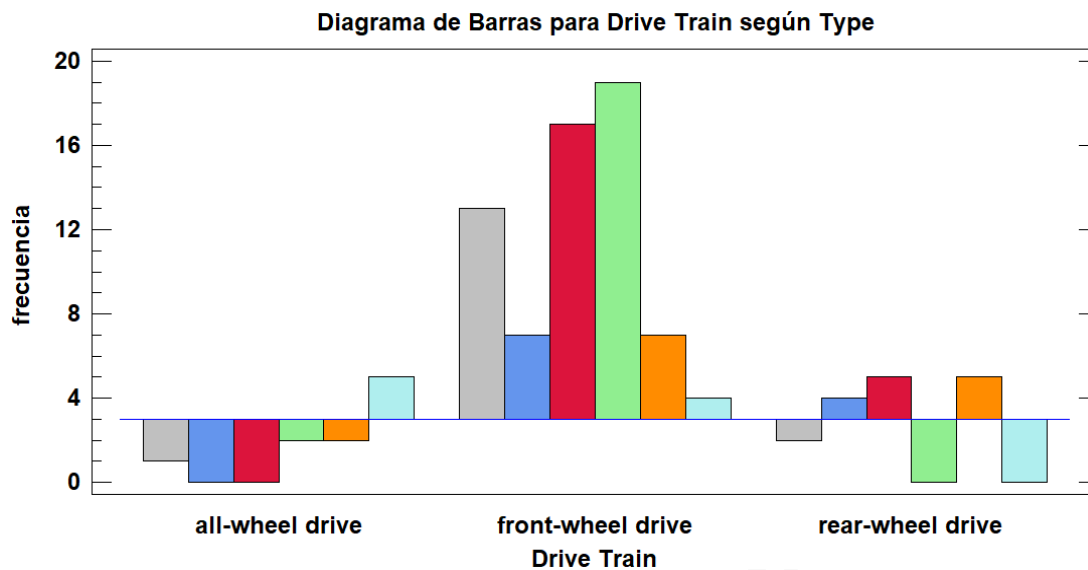


Figura 3.34 Diagrama de Barras, procedimiento *Tabulación Cruzada*

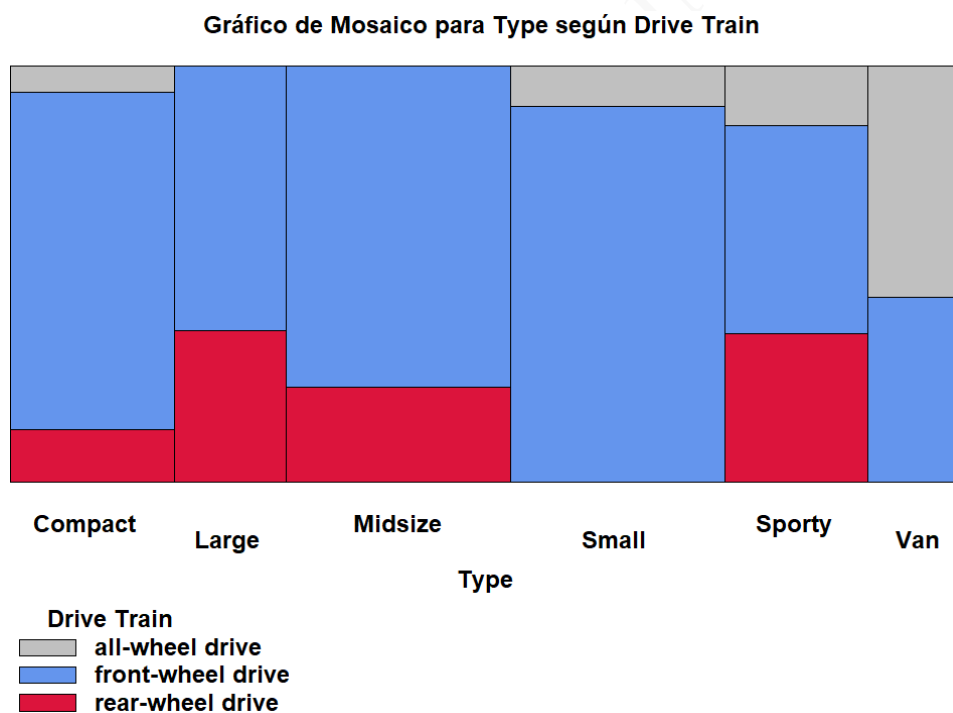


Figura 3.35 Gráfico Mosaico, procedimiento *Tabulación Cruzada*

- **GRÁFICO DE RASCACIELOS:** permite dibujar un diagrama de barras tridimensional para las variables seleccionadas. Situados sobre este gráfico, pulsando el botón derecho del ratón y eligiendo la opción *Opciones de Ventana* se permite utilizar las frecuencias absolutas o relativas.

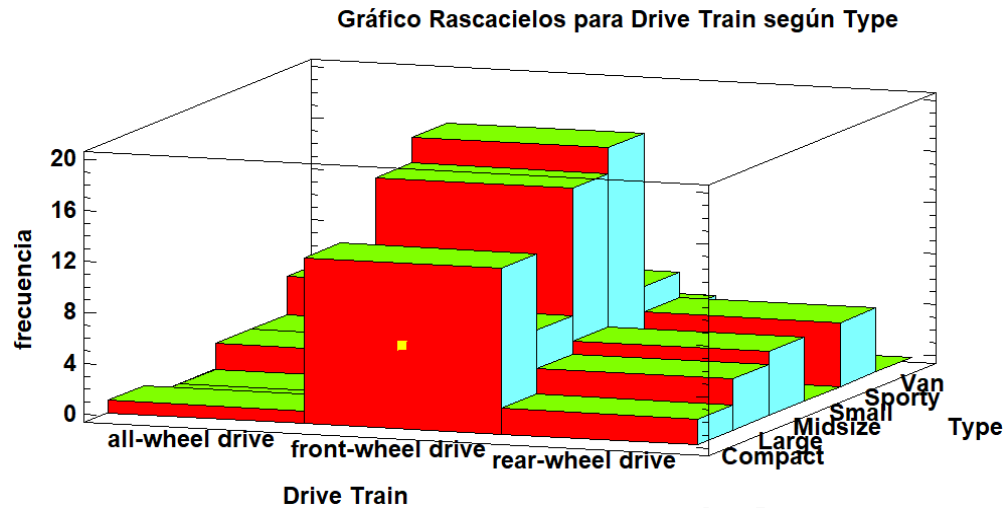


Figura 3.36 Gráfico de Rascacielos, procedimiento *Tabulación Cruzada*

3.3. Tablas de contingencia. El procedimiento *Tablas Contingencia* de la subopción *Dos Factores* de la opción *Datos Atributo* del menú *Analizar-1* permite realizar el análisis descriptivo de una variable cualitativa bidimensional cuando se introducen las frecuencias de la tabla bidimensional; es decir, cada columna de frecuencias de la tabla de contingencia se introduce como una variable en la hoja de cálculo de STATGRAPHICS. Recordemos que si en vez de introducir las frecuencias de la tabla se introduce uno a uno los valores de las variables tenemos que realizar el análisis mediante el procedimiento *Tabulación Cruzada*.

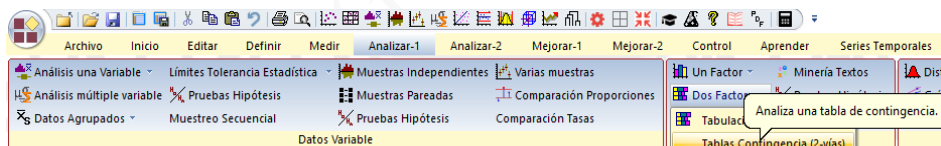


Figura 3.37 Menú de acceso a *Tablas Contingencia*

La ventana de entrada de la subopción tiene el campo COLUMNAS: donde se introducen los nombres de las variables cuyos valores son las columnas de la tabla de contingencia. También posee el campo opcional ETIQUETAS, donde se introducen las etiquetas que se asignan a cada fila en la tabla. Nótese que las etiquetas de columnas son automáticamente generadas de los nombres de las éstas (ver Figura 3.39).

Las opciones tabulares y gráficas son las mismas que las especificadas anteriormente para el procedimiento *Tabulación Cruzada* (ver Figuras 3.32 a 3.36).

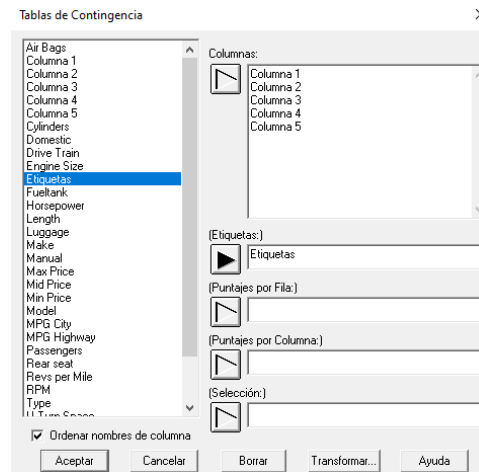


Figura 3.38 Entrada de datos, procedimiento *Tablas Contingencia*

Tabla de Frecuencias						
	Columna 1	Columna 2	Columna 3	Columna 4	Columna 5	Total por Fila
Compact	0	15	0	1	0	16
	0.00%	16.30%	0.00%	1.09%	0.00%	17.39%
Large	0	0	0	7	4	11
	0.00%	0.00%	0.00%	7.61%	4.35%	11.96%
Midsized	0	7	1	12	2	22
	0.00%	7.61%	1.09%	13.04%	2.17%	23.91%
Small	3	18	0	0	0	21
	3.26%	19.57%	0.00%	0.00%	0.00%	22.83%
Sporty	0	8	0	4	1	13
	0.00%	8.70%	0.00%	4.35%	1.09%	14.13%
Van	0	1	1	7	0	9
	0.00%	1.09%	1.09%	7.61%	0.00%	9.78%
Total por Columna	3	49	2	31	7	92
	3.26%	53.26%	2.17%	33.70%	7.61%	100.00%

Figura 3.39 Salida, procedimiento *Tablas Contingencia*