

**Prácticas de Estadística
con
Statgraphics Plus**

Dpto. de Estadística e I.O.
Universidad de Granada

15 de abril de 2008

Índice general

Introducción	1
1. Introducción al Statgraphics Plus	3
1.1. El entorno de Statgraphics Plus	3
1.1.1. Entorno de trabajo de Statgraphics	3
1.2. Opciones de la barra de menú principal	8
1.3. Editor de datos	17
1.3.1. Introducción directa de datos	18
1.3.2. Introducción de datos desde un fichero	19
1.3.3. Modificación de los datos	19
1.4. Ventana de análisis	22
1.5. Ejemplo de introducción de datos en Statgraphics	23
1.6. Ejercicios	27
2. Variables estadísticas unidimensionales	29
2.1. Introducción	29
2.2. Tabla de frecuencias y representación de variables discretas	29
2.2.1. Ejemplo práctico	30
2.3. Tabla de frecuencias y representación de variables continuas	34
2.3.1. Ejemplo práctico	34
2.4. Representación de variables cualitativas	39
2.4.1. Ejemplo práctico	39
3. Síntesis numérica de una variable estadística	43
3.1. Introducción	43
3.2. Medidas de posición, dispersión y forma	43
3.2.1. Ejemplo práctico	44
3.3. Ejercicios	54
4. Distribuciones estadísticas bidimensionales	55
4.1. Distribuciones bidimensionales de frecuencias	55
4.1.1. Ejemplo práctico	56

ÍNDICE GENERAL

4.2. Distribuciones condicionadas	60
4.2.1. Ejemplo práctico	60
4.3. Ejercicios	62
4.4. Análisis bidimensional de datos categóricos	63
4.4.1. Ejemplo práctico	63
4.5. Procedimiento Tabulación Cruzada	64
4.5.1. Tabla de frecuencias	66
4.5.2. Test de la χ^2	67
4.6. Procedimiento Tablas de Contingencia	68
4.6.1. Ejemplo práctico	68
4.7. Ejercicios	70
5. Regresión y Correlación	71
5.1. Introducción	71
5.2. Regresión lineal simple	72
5.2.1. Ejemplo práctico	72
5.2.2. Predicciones	75
5.2.3. Comparación de Modelos Alternativos	75
5.2.4. Residuos Atípicos	76
5.3. Regresión no Lineal	76
5.4. Opciones Gráficas	77
5.5. Ejercicios	79
6. Algunas distribuciones de probabilidad conocidas	83
6.1. Introducción	83
6.2. Funciones de distribución	85
6.3. Inversa de la función de distribución	85
6.4. Opciones gráficas	86
6.5. Ejemplos prácticos	87
6.6. Ejercicios	95
7. Estimación	97
7.1. Introducción	97
7.2. Intervalos de confianza para una muestra	97
7.2.1. Ejemplo práctico	99
7.3. Intervalos de confianza para dos muestras	101
7.3.1. Ejemplo práctico	103
7.4. Ejercicios	106

Introducción

El contenido de este guión de prácticas se orienta a la presentación de las técnicas de análisis estadístico de una forma sencilla de entender a través del programa Statgraphics Plus 5.1.

Hoy en día se utiliza la estadística en diferentes campos de la ciencia, desde la economía, las finanzas, la ingeniería, la matemática en general, hasta la sociología, la psicología, la biología, la medicina, etc. Lo esencial para empezar a trabajar en estadística, sea cual sea el campo en que se aplique, es la comprensión de los propios conceptos estadísticos, tarea que no exige el dominio del aparato matemático. Posteriormente, cuando ya sea necesaria la operatoria avanzada, los programas de ordenador permiten obtener los resultados sin necesidad de descifrar el desarrollo matemático de los algoritmos que están debajo de los procedimientos estadísticos.

Cada una de las prácticas que se incluyen comienza con una breve introducción teórica sobre los métodos estadísticos a utilizar para, a continuación, presentar las diferentes opciones que aporta el programa para abordar dichos métodos, aplicándolas a casos prácticos.

ÍNDICE GENERAL

Práctica 1

Introducción al Statgraphics Plus

1.1. El entorno de Statgraphics Plus

Hoy en día es imposible trabajar en estadística sin el apoyo de un ordenador, mínimo de una calculadora. No obstante, es estrictamente necesario conocer los conceptos estadísticos, su metodología y sus aplicaciones. La combinación de una buena metodología, prácticas y aplicaciones informáticas producen como resultado la optimización de la aplicación de los métodos estadísticos en la práctica.

El programa Statgraphics se trata de un software estadístico considerado actualmente como estándar, tanto en el entorno universitario como en el profesional. A pesar de las dificultades de los procedimientos estadísticos, Statgraphics los presenta de tal forma que resultan inteligibles a primera vista sin necesidad de conocimientos muy profundos en la materia. El programa incluye las técnicas estadísticas básicas, modelos lineales y no lineales, análisis multivariante, series temporales, etc.

Comenzando con Statgraphics en Windows

Para empezar a trabajar con el programa se tiene que seleccionar la siguiente ruta en el menú de Windows:

Inicio / Programas / Statgraphics Plus 5.1 / Sgwin

1.1.1. Entorno de trabajo de Statgraphics

La pantalla *StatWizard*

Al ejecutar el programa se presenta la pantalla *StatWizard* (Figura 1.1), cuya finalidad es ofrecer ayuda sobre el tema que el usuario seleccione. En esta ventana puede elegirse la opción de que no vuelva a aparecer al arrancar el programa, quitando la señal que aparece en el campo **Mostrar StatWizard al Iniciar**.

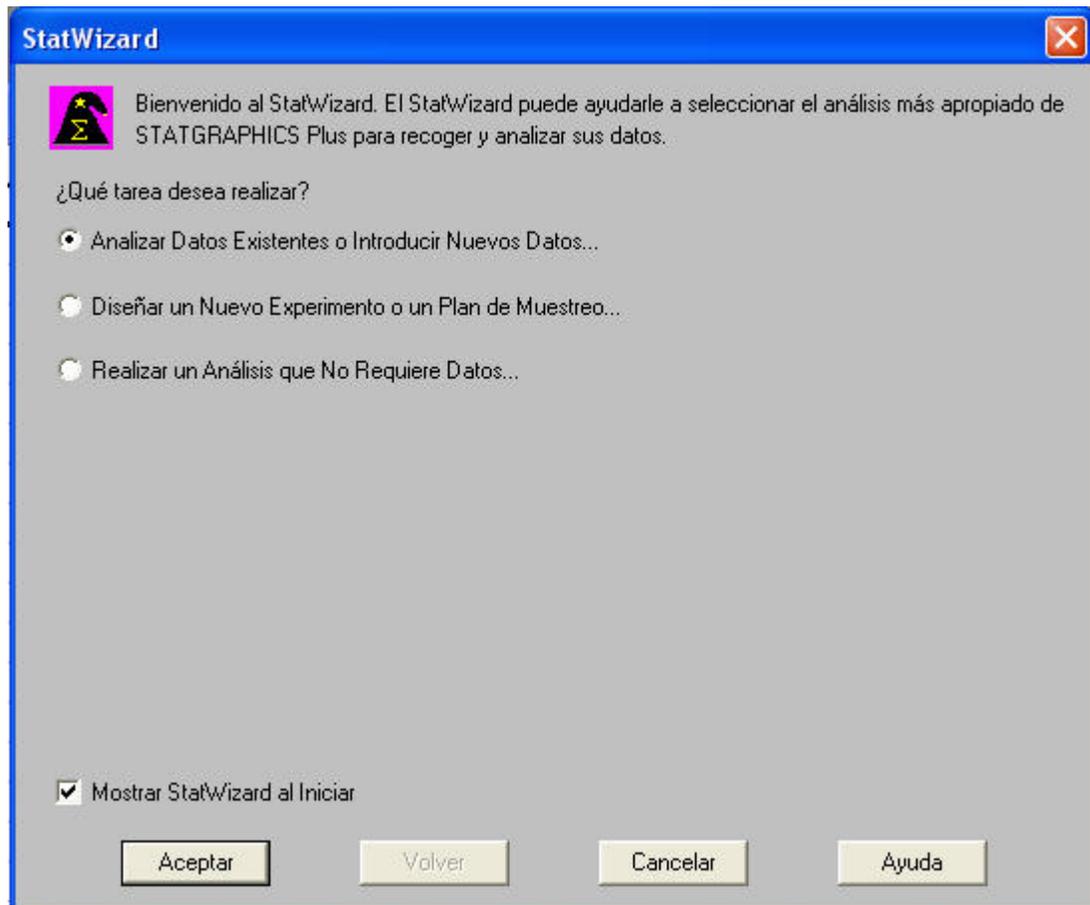


Figura 1.1: Pantalla *StatWizard*.

La ventana de aplicación

Es el marco de trabajo del programa (Figura 1.2).

En la línea superior de esta ventana está el icono de Statgraphics y el nombre del programa, seguido del nombre del fichero con el que se trabaja en ese momento (Figura 1.3).



Figura 1.3: Icono del programa y nombre del fichero abierto.

La siguiente línea corresponde a la *Barra de menú* (Figura 1.4), que contiene el menú general de la aplicación con todas sus opciones.

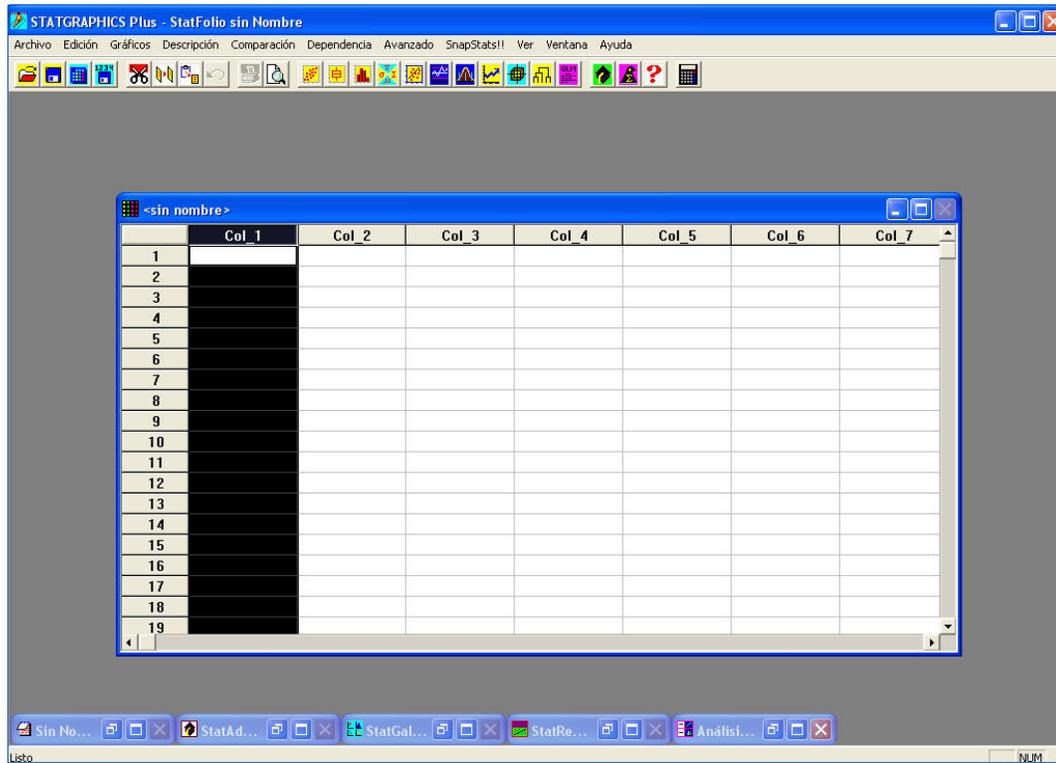


Figura 1.2: Pantalla de aplicación.

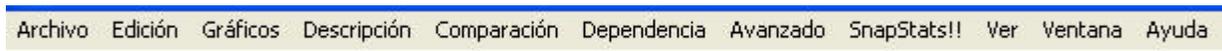


Figura 1.4: Barra de menú.

La tercera línea corresponde a la *Barra de herramientas* (Figura 1.5) que contiene diferentes iconos que permiten el acceso rápido a las opciones más comunes sin necesidad de acudir al menú general. El significado de cada icono se identifica dinámicamente situando el ratón sobre el propio icono.



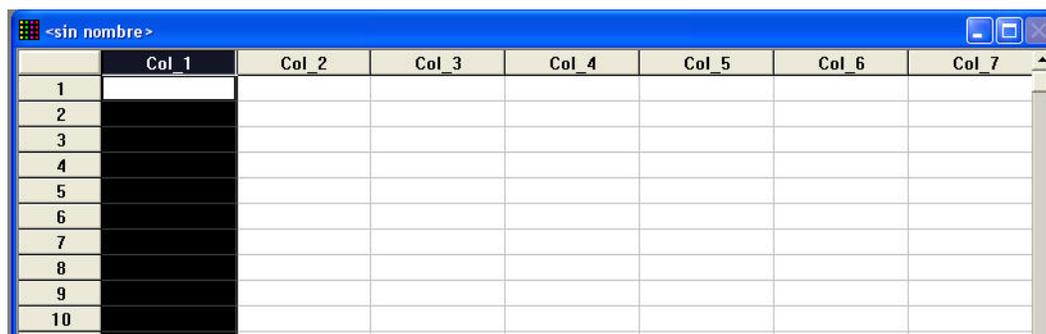
Figura 1.5: Barra de herramientas.

1.1. El entorno de Statgraphics Plus

A continuación se describen cada uno de estos iconos:

	Abrir Statfolio		Guardar Statfolio
	Abrir fichero de datos existente		Guardar fichero de datos actual
	Cortar selección		Copiar selección
	Pegar lo copiado en la selección		Deshacer la última acción
	Imprimir la pantalla actual		Vista preliminar de la impresión actual
	Gráfico de dispersión		Gráfico de caja y bigotes
	Histograma		Estadísticos: Resumen estadístico
	Regresión múltiple		Gráfico de medias y rangos
	Análisis de la capacidad		Predicción de series temporales
	Abrir fichero de diseño		Análisis cluster
	Modelo lineal general		StatAdvisor: Intérprete de resultados
	StatWizard: Ayuda estadística		Ayuda general

En el centro de la pantalla de aplicación se encuentra el *Editor de datos* (Figura 1.6), desde el que se pueden introducir las observaciones. Es similar a una hoja de cálculo cuyas columnas serán las variables utilizadas en los análisis estadísticos. El conjunto de todas las columnas (variables) forma un fichero. Cuando se tiene activo un fichero, el nombre de este aparece en la parte superior de esta ventana.



	Col_1	Col_2	Col_3	Col_4	Col_5	Col_6	Col_7
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							

Figura 1.6: Editor de datos.

En la parte inferior de la pantalla de aplicación se muestra la *Barra de tareas* (Figura 1.7) que incluye cuatro iconos de los que surgen las siguientes ventanas:



Figura 1.7: Barra de tareas.

La ventana *StatAdvisor* (Figura 1.8)

En esta ventana aparece una interpretación de los resultados obtenidos después de haber ejecutado un procedimiento.

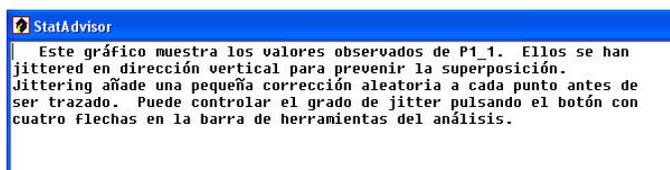


Figura 1.8: Ventana *StatAdvisor*.

La ventana *StatGallery* (Figura 1.9)

La ventana *StatGallery* se utiliza para realizar presentaciones de gráficos a medida. Los gráficos obtenidos en las salidas de los procedimientos pueden colocarse en esta ventana para una posterior presentación óptima de los mismos.

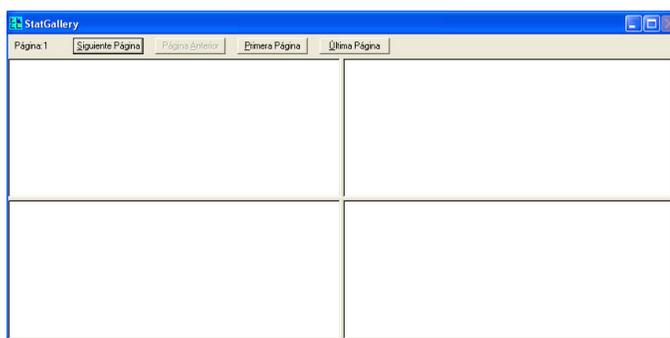


Figura 1.9: Ventana *StatGallery*.

La ventana *StatReporter* (Figura 1.10)

La ventana *StatReporter* permite generar informes que contienen texto, gráfico o análisis estadísticos

La ventana de *comentarios* (Figura 1.11)

Bloc de notas para introducir comentarios sobre el Statfolio activo.

1.2. Opciones de la barra de menú principal

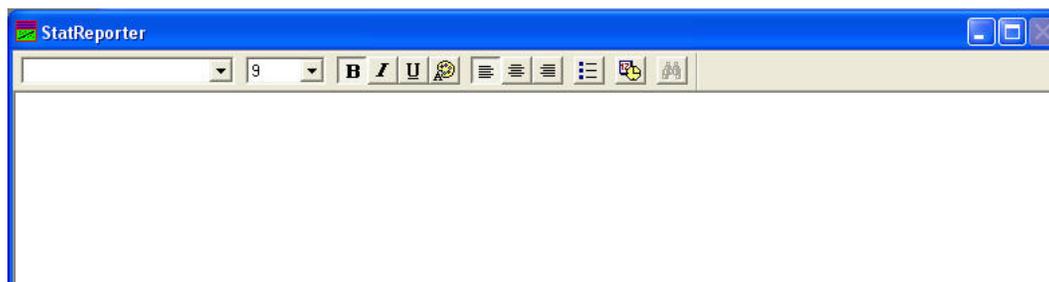


Figura 1.10: Ventana *StatReporter*.



Figura 1.11: Ventana de *comentarios*.

1.2. Opciones de la barra de menú principal

A continuación se explica la finalidad de las distintas opciones que aparecen en la *Barra de menú* del programa (Figura 1.4).

La opción **Archivo** de la *Barra de menú* principal presenta las siguientes opciones:

- **Abrir *StatFolio*...** (abrir macro)
- **Abrir Datos...** (abrir fichero de datos)
- **Abrir *StatGallery*...** (abrir *StatGallery*)
- **Abrir *StatReport*...** (abrir *StatReport*)
- **Info. Base de Datos (ODBC)...** (abrir consulta a base de datos)
- **Leer Portapapeles...** (leer el contenido actual del portapapeles)
- **Cerrar *StatFolio*** (cerrar la macro actual)
- **Cerrar Datos** (cerrar el fichero de datos actual)

- Cerrar *StatGallery* (cerrar *StatGallery*)
- Cerrar *StatReport...* (cerrar *StatReport*)
- Guardar *StatFolio* (guardar macro)
- Guardar Datos (guardar el fichero de datos actual)
- Guardar *StatGallery* (guardar *StatGallery*)
- Guardar *StatReport...* (guardar *StatReport*)
- Guardar *StatFolio Como...* (guardar macro con otro nombre)
- Guardar Datos Como... (guardar el fichero con otro nombre)
- Guardar *StatGallery Como...* (guardar *StatGallery* con otro nombre)
- Guardar *StatReport Como...* (guardar *StatReport* con otro nombre)
- *StatLink* (enlazar con un origen de datos)
- Imprimir... (imprimir el contenido de la ventana actual)
- Vista Previa... (vista preliminar)
- Configuración de impresión... (definir las características de impresión)
- Configuración de página... (definir las características de página)
- Guardar Gráfico... (guardar gráfico)
- *StatPublish...* (editar la macro actual en formato HTML)
- Vista de resultados editados... (ver macro editada)
- Combinar *StatFolios...* (combinar distintas macros)
- Combinar Datos... (combinar distintos ficheros)
- Enviar (enviar documento)
- Vínculos... (editar objetos vinculados)
- Salir de Statgraphics (salir del programa)

La opción **Edición** de la *Barra de menú* principal presenta las siguientes opciones:

- **Deshacer** (deshacer la última opción)

1.2. Opciones de la barra de menú principal

- **Cortar** (cortar la selección)
- **Copiar** (copiar la selección)
- **Pegar** (pegar la selección)
- **Pegar Vínculo** (insertar contenido del bloc de notas y un vínculo a su fuente)
- **Pegado Especial** (insertar el contenido del bloc de notas con opciones)
- **Preferencias...** (preferencias gráficas y del sistema)
- **Cambiar Fuente...** (cambiar la fuente del texto)
- **Cambiar Título de Análisis...** (cambiar el título de un análisis)
- **Copiar Análisis** (copiar un análisis)
- **Insertar** (insertar el contenido del portapapeles)
- **Borrar** (borrar la selección)
- **Actualizar Fórmulas** (actualizar fórmulas)
- **Modificar Columna...** (modificar columna)
- **Generar Datos...** (generar nuevos datos)
- **Recodificar Datos...** (recodificar datos)
- **Ordenar Datos...** (ordenar fichero)

La opción **Gráficos** de la *Barra de menú* principal presenta las siguientes opciones:

- **Gráficos de Dispersión** (gráficos de dispersión de datos)
 - **Gráfico Unidimensional...** (gráficos unidimensionales)
 - **Gráfico X-Y...** (gráfico bidimensional)
 - **Gráfico X-Y-Z...** (gráfico tridimensional)
 - **Gráfico de Matriz...** (gráficos matriciales)
 - **Gráfico X-Y Múltiple...** (gráfico múltiple bidimensional)
 - **Gráfico X-Y-Z Múltiple...** (gráfico múltiple tridimensional)
 - **Gráfico de Coordenadas Polares...** (gráfico en coordenadas polares)
- **Gráficos Exploratorios** (gráficos exploratorios de datos)

- **Gráfico de Caja y Bigotes...** (gráfico de caja y bigotes)
- **Gráfico de Caja y Bigotes Múltiple...** (gráfico múltiple de caja y bigotes)
- **Gráfico Normal de Probabilidad...** (gráfico probabilístico normal)
- **Histograma...** (histograma de frecuencias)
- **Diagrama de Puntos...** (diagrama de puntos)
- **Diagrama de Puntos Múltiple...** (diagrama de puntos múltiple)
- **Gráfico de Burbujas...** (gráfico de burbujas)
- **Gráfico Radar/Araña...** (diagrama radar/araña)
- **Diagramas de Presentación** (gráficos de presentación)
 - **Diagrama de Barras...** (gráfico de barras)
 - **Diagrama de Barras Múltiple...** (gráfico de barras múltiple)
 - **Diagrama de Sectores...** (gráfico de sectores)
 - **Gráfico Línea de Componentes...** (gráfico de líneas)
 - **Gráfico Alto-Bajo-Próximo...** (gráficos sobre tiempo)
- **Distribuciones de Probabilidad...** (distribuciones de probabilidad)
- **Superficies de Respuesta...** (gráficos de superficie de respuesta)
- **Gráficos a Medida...** (gráfico a medida del usuario)

La opción **Descripción** de la *Barra de menú* principal presenta las siguientes opciones:

- **Datos Numéricos** (datos de variables numéricas)
 - **Análisis Unidimensional...** (análisis de una variable)
 - **Análisis Multidimensional...** (análisis de varias variables)
 - **Análisis de Subgrupo...** (análisis de subconjuntos de datos)
 - **Estadísticos por Filas...** (estadísticas por filas)
 - **Transformación de Potencia...** (transformaciones de Box-Cox)
 - **Líneas de Tolerancia Estadística...** (tolerancia $\mu \pm 3\sigma$)
 - **Identificación de Valores Atípicos...** (valores atípicos en series)
- **Datos Cualitativos** (datos de variables categóricas)
 - **Tabulación...** (tabulación)
 - **Tabulación Cruzada...** (tabulación cruzada)

- **Tablas de Contingencia...** (tablas de contingencia)
- **Distribuciones** (distribuciones)
 - **Distribuciones de Probabilidad...** (distribuciones de probabilidad)
 - **Gráficos de Probabilidad...** (gráficos de probabilidad)
 - **Ajuste de Distribuciones (Datos no Censurados)...** (ajuste sin censura)
 - **Ajuste de Distribuciones (Datos Censurados)...** (ajuste con censura)
- **Datos de Vida** (tablas de vida)
 - **Tablas de Vida (Intervalos)...** (tablas de vida por intervalos)
 - **Tablas de Vida (Tiempos)...** (tablas de vida puntuales)
 - **Análisis Weibull...** (análisis de Weibull)
 - **Gráficos Arrhenius...** (gráficos de Arrhenius)
- **Contrastes de Hipótesis...** (contrastes de hipótesis)
- **Determinación del Tamaño de la Muestra...** (cálculo del tamaño de muestra)

La opción **Comparación** de la *Barra de menú* principal presenta las siguientes opciones:

- **Dos Muestras** (análisis con dos muestras)
 - **Comparación de Dos Muestras...** (comparación para dos muestras)
 - **Comparación de Muestras Pareadas...** (comparación para muestras pareadas)
 - **Contraste de Hipótesis...** (contrastes de hipótesis)
 - **Determinación del Tamaño de Muestra...** (cálculo del tamaño de muestra)
- **Muestras Múltiples** (análisis con múltiples muestras)
 - **Comparación de Varias Muestras...** (comparación para varias muestras)
 - **Comparación de Proporciones...** (comparación para proporciones)
 - **Comparación de Tasas...** (comparación para tasas de casos)
- **Análisis de la Varianza** (análisis de la varianza)
 - **Gráficos de Medias Factorial** (gráficos factoriales para medias)

- **ANOVA Simple...** (análisis de la varianza simple)
- **ANOVA Factorial...** (análisis de la varianza múltiple)
- **Componentes de la Varianza...** (componentes de la varianza)

La opción **Dependencia** de la *Barra de menú* principal presenta las siguientes opciones:

- **Regresión Simple...** (análisis de regresión simple)
- **Regresión Polinomial...** (análisis de regresión polinomial)
- **Transformaciones Box-Cox...** (transformaciones de Box-Cox de una serie)
- **Regresión Múltiple...** (análisis de regresión múltiple)

La opción **Avanzado** de la *Barra de menú* principal presenta las siguientes opciones:

- **Control de Calidad** (técnicas de control de calidad)
 - **Análisis de Pareto** (análisis de Pareto)
 - **Análisis de Capacidad de Proceso** (análisis de capacidad de un proceso)
 - **Gráficos de Control por Variables** (gráficos de control por variables)
 - **X-bar y R** (gráficos de recorridos o rangos)
 - **X-bar y S** (gráficos de desviaciones típicas)
 - **X-bar y S-cuadrado** (gráficos de varianzas)
 - **Individuales** (gráficos de valores individuales)
 - **Gráficos de Control por Atributos** (gráficos de control por atributos)
 - **Gráfico p** (gráficos de proporción de unidades defectuosas)
 - **Gráfico np** (gráficos de número de unidades defectuosas)
 - **Gráfico u** (gráficos de número de defectos por unidad)
 - **Gráfico c** (gráficos de número de defectos)
 - **Gráficos de Tiempo Ponderados** (gráficos de tiempo ponderados)
 - **Gráfico de Medias Móviles** (gráfico de medias móviles)
 - **Gráfico de Medias Móviles Individuales** (gráfico de medias móviles individuales)
 - **Gráfico EWMA** (gráfico exponencial ponderado de medias móviles)
 - **Gráfico EWMA Individuales** (gráfico exponencial ponderado de medias móviles individuales)
 - **Gráfico CuSum (V-Máscara)** (gráficos de sumas acumulativas)

- **Gráfico CuSum Individuales (V-Máscara)** (gráficos de sumas acumulativas individuales)
- **Gráfico CuSum (H-K)...** (gráficos de sumas acumulativas H-K)
- **Gráfico CuSum Individuales (H-K)...** (gráficos de sumas acumulativas individuales H-K)
- **Gráficos Multivariable** (gráficos de control multivariable)
- **Gráficos de Control con Propósito Especial** (gráficos de control especiales)
 - **Gráficos ARIMA...** (gráficos ARIMA según Box-Jenkins)
 - **Gráficos ARIMA Individuales...** (gráficos ARIMA según Box-Jenkins individuales)
 - **Gráficos Toolwear...** (gráficos de desgaste)
 - **Gráficos Toolwear Individuales...** (gráficos de desgaste individuales)
 - **Gráficos de Aceptación...** (gráficos de aceptación)
 - **Gráficos de Aceptación Individuales...** (gráficos de aceptación individuales)
- **Medida R&R** (medida de la repetibilidad y reproductividad)
 - **Configuración Datos** (instalación de datos)
 - **Método Media y Rango** (método media y rango)
 - **Método ANOVA** (método del análisis de la varianza)
 - **Método Rango** (método del rango)
- **Gráficos a Medida** (gráficos a medida)
- **Diagrama de Espina de Pescado...** (gráficos de causa-efecto)
- **Aceptación de la Muestra** (muestreo de aceptación)
 - **Atributos...** (muestreo de aceptación por atributos)
 - **VARIABLES...** (muestreo de aceptación por variables)
- **Gráfico Multi-Vari...** (gráfico multivariable)
- **Diseño Experimental** (diseño de experimentos)
 - **Crear Diseño...** (crear diseño)
 - **Optimizar Diseño...** (optimizar diseño)
 - **Abrir Diseño...** (abrir diseño)
 - **Analizar Diseño...** (analizar diseño)
 - **Aumentar Diseño...** (aumentar diseño)

- **Optimizar Respuesta Múltiple...** (optimizar diseño de respuesta múltiple)
- **Análisis Series Temporales** (análisis de series temporales)
 - **Métodos Descriptivos...** (métodos descriptivos)
 - **Suavizado...** (suavizado)
 - **Descomposición Estacional...** (descomposición estacional)
 - **Predicción...** (predicción)
 - **Predicción Automática...** (predicción automática)
- **Métodos Multivariantes** (métodos multivariantes)
 - **Componentes Principales...** (componentes principales)
 - **Análisis Factorial...** (análisis factorial)
 - **Análisis Cluster...** (análisis cluster)
 - **Análisis Discriminante...** (análisis discriminante)
 - **Correlaciones Canónicas...** (correlación canónica)
- **Regresión Avanzada** (regresión avanzada)
 - **Modelos Lineales Generales...** (ajustes a modelos lineales generales)
 - **Modelos de Calibración...** (ajustes de modelos de calibración)
 - **Comparación de Rectas de Regresión...** (comparación de modelos)
 - **Selección del Modelo de Regresión...** (selección de modelos óptimos)
 - **Regresión no Lineal...** (regresión con modelos no lineales)
 - **Regresión en Cadena...** (regresión en cadena)
 - **Regresión Logística...** (regresión logística)
 - **Regresión Probit...** (ajuste del modelo Probit)
 - **Regresión de Poisson...** (ajuste de modelos de Poisson)
 - **Regresión de Datos de Vida...** (regresión de Box-Cox)

La opción **SnapStats!!** de la *Barra de menú* principal presenta las siguientes opciones, que corresponden a los procedimientos estadísticos más comunes y sencillos que se utilizan en Statgraphics. Su finalidad es acceder de forma rápida sin tener que desdoblar menús:

- **Análisis Una Muestra**
- **Comparación Dos Muestras**

1.2. Opciones de la barra de menú principal

- **Comparación Muestras Pareadas**
- **Comparación Muestras Múltiples**
- **Ajuste de la Curva**
- **Valoración de Capacidad (Individuales)**
- **Valoración de Capacidad (Datos Agrupados)**
- **Medida R&R**
- **Predicción Automática**

La opción **Ver** de la *Barra de menú* principal presenta las siguientes opciones:

- **Barra de Herramientas** (activa y desactiva la barra de herramientas)
- **Barra de Estado** (activa y desactiva la barra de estado)
- **StatAdvisor** (activa y desactiva el intérprete de resultados)

La opción **Ventana** de la *Barra de menú* principal presenta las siguientes opciones:

- **Cascada** (cascada)
- **Mosaico** (mosaico)
- **Organizar Iconos** (organizar iconos)
- **Restablecer Ventanas** (restablecer ventanas a la forma por defecto)
- **Evaluación de Expresión** (evaluación de expresiones)

La opción **Help** de la *Barra de menú* principal presenta las siguientes opciones:

- **Índice** (contenidos)
- **Conocer el Programa** (aprendiendo el programa)
- **Accediendo y Usando Ficheros** (acceso y uso de ficheros)
- **Datos y Hojas de Datos** (datos y hojas de datos)
- **Gráficos y Opciones Gráficas** (gráficos y opciones gráficas)
- **Imprimiendo** (impresión)
- **Publicación Web** (obtener un fichero html de resultados)

- **Características Especiales** (características especiales)
- **Análisis de Edición Estándar** (edición estándar de los análisis)
- **Utilizando Análisis Avanzados** (uso de los análisis avanzados)
 - **Análisis para Control de Calidad** (análisis de control de calidad)
 - **Análisis para Diseño de Experimentos** (análisis de diseño de experimentos)
 - **Análisis para Regresión Avanzada** (análisis de regresión avanzada)
 - **Análisis para Métodos Multivariantes** (análisis de métodos multivariantes)
 - **Análisis para Series Temporales** (análisis de series temporales)
- **Información de Referencia** (información referenciada)
- **Soporte Técnico** (soporte técnico)
- **Acceso a Manuales On-Line**
- **Acceso a Fichero Notas en Leeme**
- **Adquiere STATGRAPHICS Ahora**
- **Noticias STATGRAPHICS**
- **Aprendiendo STATGRAPHICS**
- **StatWizard...** (activa la ventana *StatWizard*)
- **Acerca de...** (sobre Statgraphics 5.1.)

1.3. Editor de datos

Para realizar cualquier análisis estadístico, en primer lugar, hay que disponer de algunos datos con los que trabajar. Éstos se pueden obtener de varias formas:

1. Introducción directa de los datos.
2. Exportarlos desde un archivo.
3. Modificar algunos datos ya introducidos.

1.3.1. Introducción directa de datos

Para introducir directamente los datos, se escriben los valores en el editor de datos, como si fuera una hoja de cálculo, teniendo en cuenta que:

1. Cada fila será un individuo y cada columna una variable. Por ejemplo, si se tiene una encuesta que conste de 10 preguntas realizadas a 1000 individuos, esto será un fichero de datos en forma matricial de 10 columnas y 1000 filas. Para pasar de un individuo a otro se pulsa el botón ENTER y para pasar de una variable a otra el botón TABULACIÓN.
2. Por defecto el programa solo permite la introducción de valores numéricos.
3. Para modificar alguna de las características de las variables con las que vamos a trabajar:
 - Se selecciona la variable deseada.
 - Se pulsa el botón derecho del ratón.
 - Y se selecciona de la ventana la opción **Modificar Columna...**

Se genera un cuadro de dialogo en el que es posible cambiar el nombre de la variable, introducir un comentario, modificar el ancho de la columna así como el tipo de dato con el se va a trabajar.

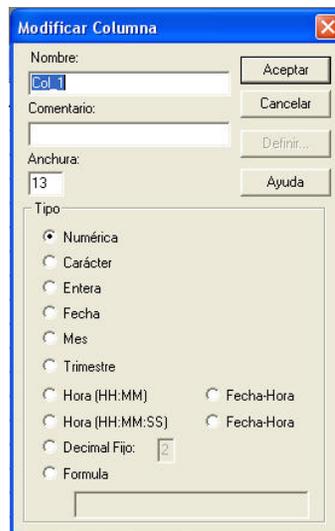


Figura 1.12: Opciones de las variables.

1.3.2. Introducción de datos desde un fichero

En este caso se elige el siguiente menú:

Archivo / Abrir / Abrir Datos

Se selecciona el fichero y la extensión del archivo que deseemos abrir:

1. Si el archivo tiene extensión SG PLUS file, se abre directamente.
2. Si se selecciona un fichero procedente de Excell, se tiene que indicar si el nombre de las variables aparece en la primera fila o las genera el programa automáticamente, y si existe algún valor que ha sido asignado a los valores perdidos.
3. Si el archivo es de tipo ASCII, además de indicar si el nombre de las variables aparece en la primera fila del archivo y si existe un código para los valores perdidos, se debe seleccionar el tipo de separación entre las variables (por tabulador, por un espacio en blanco, por comas o por algún formato).

1.3.3. Modificación de los datos

Una vez creado un fichero de datos, este se puede modificar. Para ello se selecciona la variable a modificar y se pulsa el botón derecho del ratón.

Deshacer	Ctrl+Z
Cortar	Ctrl+X
Copiar	Ctrl+C
Pegar	Ctrl+V
Pegar Enlace	Ctrl+L
Insertar	
Borrar	
Modificar Columna...	Mayús+F5
Generar Datos...	Mayús+F7
Recodificar Datos...	
Ordenar Fichero...	
Convertir Fecha-Hora desde Excel...	
Actualizar Fórmulas...	
Imprimir...	F4
Vista Previa...	Mayús+F3
Guardar Fichero de Datos	Mayús+F12
Guardar Fichero de Datos Como...	F12

Figura 1.13: Modificación de datos.

Además de realizar las tareas propias de edición (insertar, copiar o cortar las variables), Statgraphics permite realizar una serie de modificaciones. A continuación se comentan algunas de ellas:

1.3. Editor de datos

1. *Deshacer*: Deshacer la última acción
2. *Cortar*: Cortar
3. *Copiar*: Copiar
4. *Pegar*: Pegar
5. *Pegar Enlace*: Pegar con vínculo a su fuente
6. *Insertar*: Insertar
7. *Borrar*: Borrar
8. *Modificar Columna*: Elección de las opciones de la variable
9. *Generar Datos*: Generar datos. En esta nueva ventana podemos realizar operaciones que nos proporcionaran nuevas variables con los resultados de las mismas. Estas operaciones pueden ser de tipo lógico (<, >, ≤, ≥, and, or), de tipo aritmético (+, -, *, /) o mediante alguno de los operadores que tiene implementado el programa. Una vez elegido el operador, tendremos que completar los parámetros de los que depende y que aparecen con ?. Por ejemplo, si existe una variable que se llama ingresos y

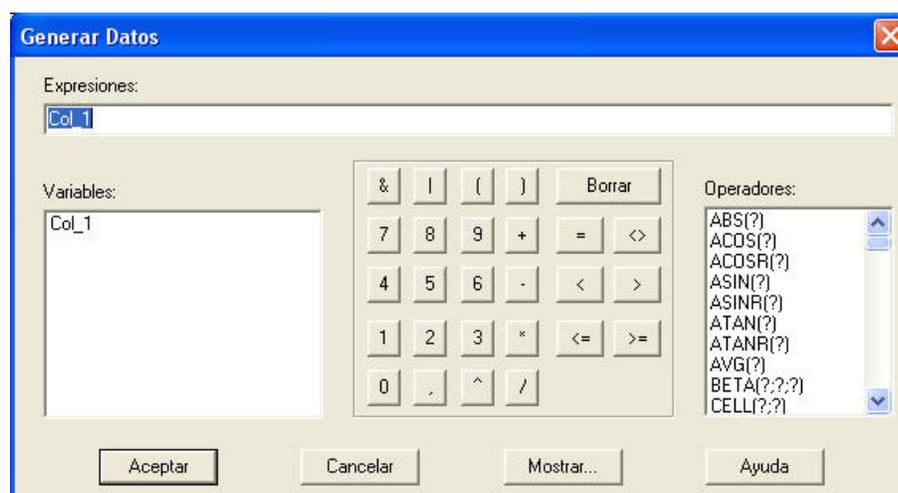


Figura 1.14: Generar datos.

se quiere calcular su valor absoluto, hay que hacer ABS(ingresos). Si existen varios signos de interrogación será que esa operación requiere de varios parámetros en cuyo caso se sustituye cada signo de interrogación por un parámetro y se separan por puntos y comas.

10. *Recodificar Datos*: Recodificar los valores de una variable. Se asignan nuevos valores a todos los comprendidos en un intervalo, teniendo que indicar las condiciones de los límites de esos intervalos, es decir, si los extremos se consideran dentro o fuera del intervalo.

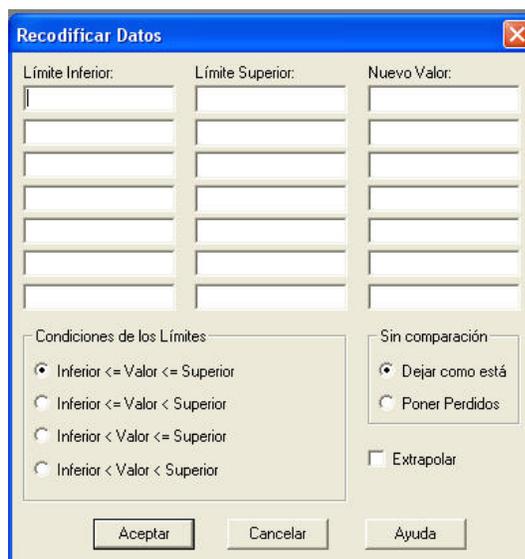


Figura 1.15: Recodificación de datos.

11. *Ordenar Fichero*: Permite ordenar los datos de modo ascendente o descendente, de modo que al ordenar el fichero según una variable, modifica el resto de variables para que no se pierda la coherencia del fichero. También permite introducir una segunda variable de orden para los casos en los que en la primera exista el mismo valor.



Figura 1.16: Ordenador datos.

1.4. Ventana de análisis

Después de ejecutar cualquier procedimiento estadístico de Statgraphics, el programa presenta una *ventana de análisis* (Figura 1.17), en la que se muestran los resultados.

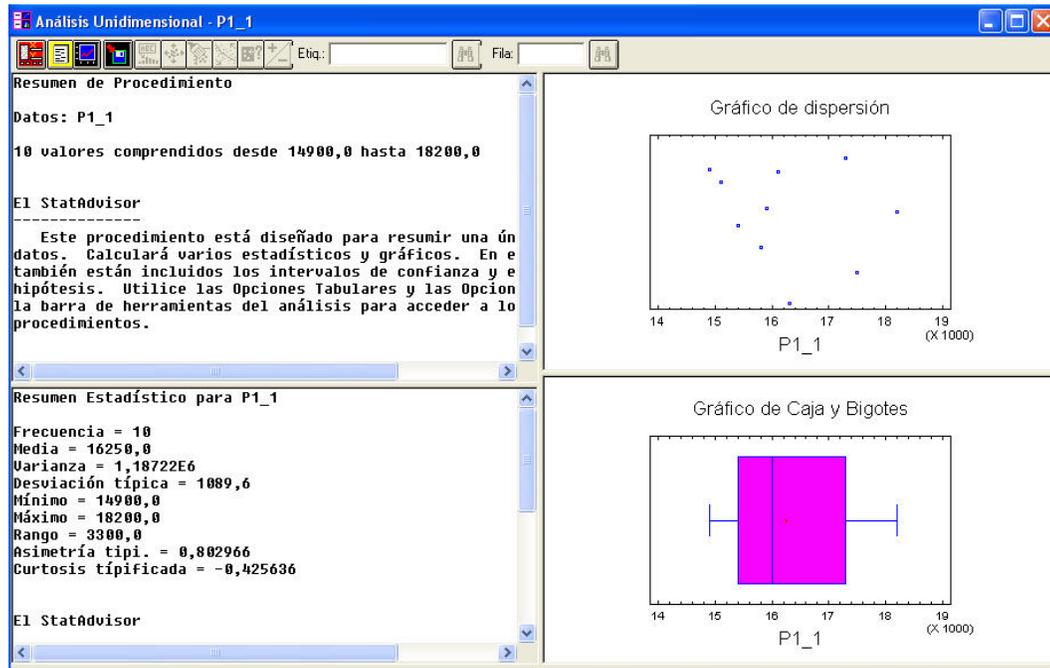


Figura 1.17: Ventana de análisis.

En esta ventana se observan tres zonas esenciales:

1. **La barra de icono y título del análisis.** Presenta el nombre del procedimiento estadístico cuyos resultados se muestran en la ventana de análisis. Además se muestra el nombre de la variable que se analiza.
2. **La barra de herramientas del análisis.** Presenta una sucesión de iconos que van a posibilitar las diferentes opciones de trabajo del análisis actual.

-  **Introducir Texto.** Reabre la ventana de dialogo que presenta las variables del fichero activo, permitiendo así la elección de nuevas variables o procedimientos para el análisis.
-  **Opciones Tabulares.** Se utiliza para presentar todas las posibles subopciones con resultados analíticos que permite el procedimiento activo.
-  **Opciones Gráficas.** Se utiliza para presentar todas las posibles subopciones con resultados gráficos que permite el procedimiento activo.

-  **Guardar Resultados.** Se utiliza para guardar los resultados numéricos del análisis estadístico en variables cuyo nombre se indica en la pantalla correspondiente.
-  **Añadir Texto.** Para añadir texto a un gráfico.
-  **Separar Puntos.** Para separar puntos en las zonas densas de un gráfico.
-  **Resaltar Puntos.** Resalta los puntos de una zona señalada.
-  **Suavizar/Rotar.** Rotación de un gráfico 3D.
-  **Identificar.** Identifica puntos de un gráfico.
-  **Excluir.** Excluye puntos de un gráfico.
-  Localiza puntos de un gráfico por nombre.
-  Localiza puntos de un gráfico por etiqueta.

3. **La ventana de salida de resultados.** Se sitúa debajo de la barra de herramientas de análisis y se divide en dos zonas. La zona de la izquierda presenta los resultados del análisis estadístico y la zona de la derecha los resultados gráficos, si existen. Pulsando dos veces con el ratón sobre alguna de ellas se maximizan. Para volver al tamaño original, se vuelve a hacer click dos veces el ratón.

A su vez, si en en la parte de resultados se pulsa el botón de la derecha aparecen dos opciones más:

1. **Opciones de Ventana.** Ofrece más opciones de ese mismo análisis.
2. **Opciones de Análisis.** Proporciona opciones generales del análisis.

1.5. Ejemplo de introducción de datos en Statgraphics

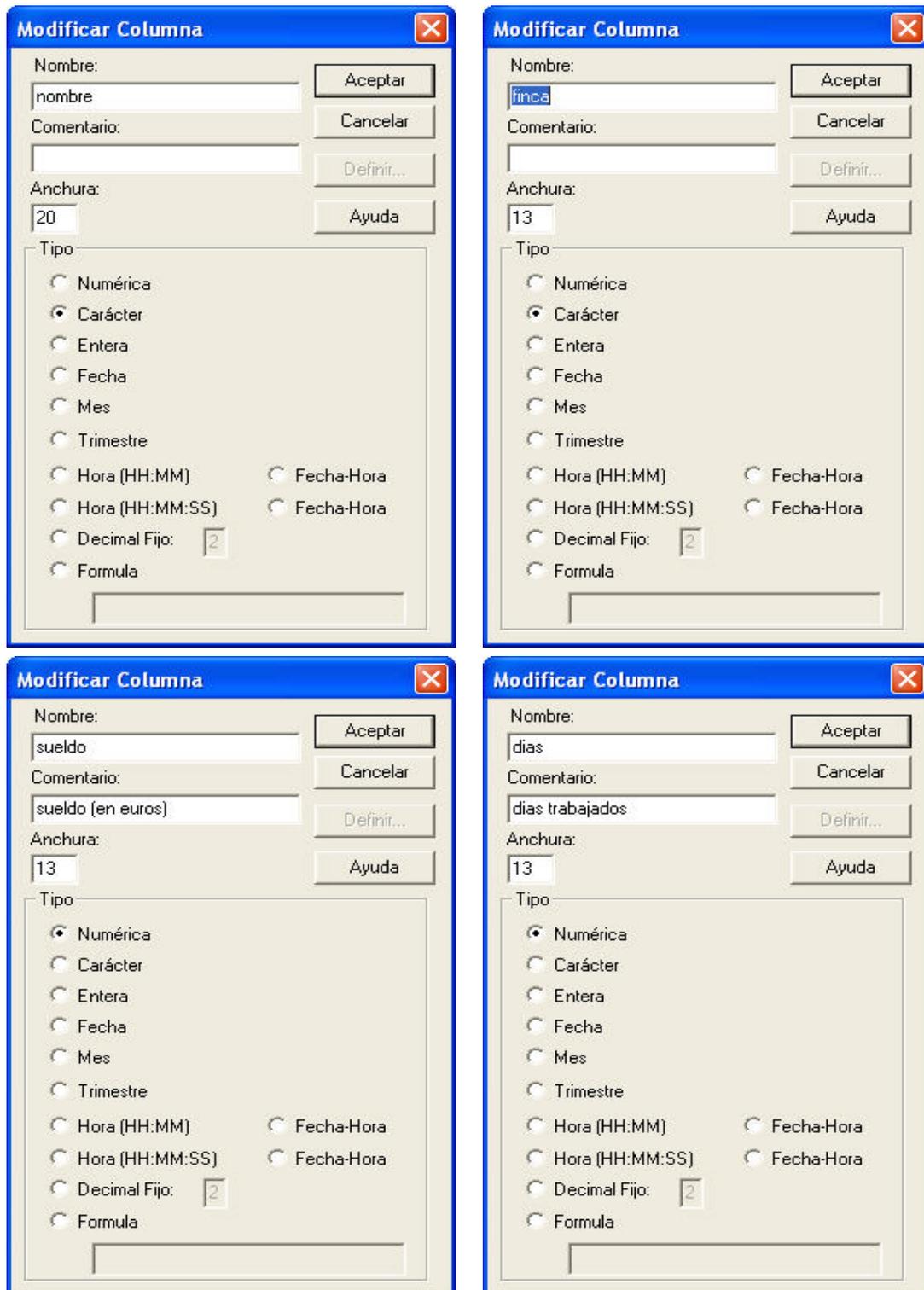
A continuación se explica, paso a paso, mediante un ejemplo, como introducir un conjunto de datos en Statgraphics.

Sea el siguiente conjunto de observaciones, que representan el *sueldo diario* y *días trabajados* de los trabajadores de distintas fincas de una cooperativa agrícola:

1.5. Ejemplo de introducción de datos en Statgraphics

Nombre	Finca	Sueldo	Días
Francisco López	Finca 1	40,5	30
Sonia Ruiz	Finca 2	30	68
Rafael León	Finca 3	30	13
Lucía Lara	Finca 4	40,5	95
Dolores Millán	Finca 1	30	30
Antonio Cano	Finca 3	30	39
Manuel Cano	Finca 2	40,5	17
Juan Ariza	Finca 2	30	68
Diego Giron	Finca 4	30	114
Julio Villar	Finca 3	40,5	91
Carmen Pérez	Finca 1	30	15
Eva Morón	Finca 1	30	15
Isabel López	Finca 2	30	153
Francisco Molina	Finca 4	30	57
Luisa Giron	Finca 4	30	76
Antonio López	Finca 2	30	34
Luis Paz	Finca 3	30	52
María Reyes García	Finca 4	30	57
Lucía Martínez	Finca 1	30	15
Rafael Trechera	Finca 3	40,5	13
Diego Delgado	Finca 1	30	30
Tamara López	Finca 4	40,5	38
Fernando Guerrero	Finca 2	30	17
Andrea García	Finca 1	30	15
Francisco Espada	Finca 4	40,5	57
Teresa Gimenez	Finca 3	30	52
Alberto Requena	Finca 3	30	65
Rocío Cobo	Finca 4	40,5	76
María Luisa Yllera	Finca 2	40,5	17
Salud Montero	Finca 2	30	17

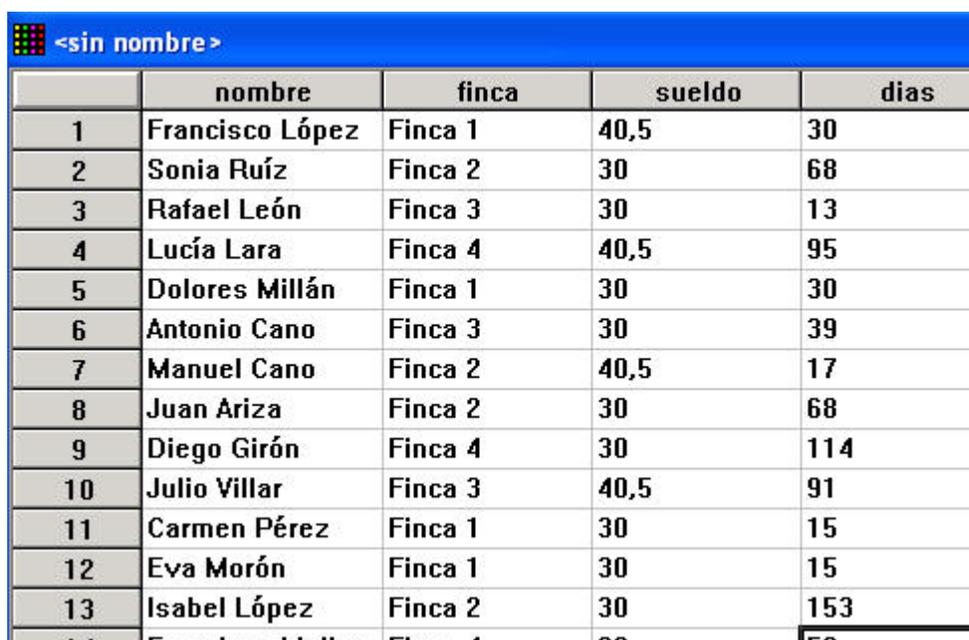
Para introducir los datos se comienza definiendo las columnas con el nombre de las variables y el tipo de dato. Para ello se pincha con el botón secundario del ratón sobre cada columna y se elige la opción **Modificar Columna...** y se introducen los nombres y tipo de las variables (Cuadro 1.1).



Cuadro 1.1: Definición de variables.

1.5. Ejemplo de introducción de datos en Statgraphics

A continuación, se introducen los datos en las casillas correspondientes (ver la Figura 1.18).



The screenshot shows a data entry window titled "<sin nombre>". It contains a table with 5 columns: an index column, "nombre", "finca", "sueldo", and "dias". The data is as follows:

	nombre	finca	sueldo	dias
1	Francisco López	Finca 1	40,5	30
2	Sonia Ruíz	Finca 2	30	68
3	Rafael León	Finca 3	30	13
4	Lucía Lara	Finca 4	40,5	95
5	Dolores Millán	Finca 1	30	30
6	Antonio Cano	Finca 3	30	39
7	Manuel Cano	Finca 2	40,5	17
8	Juan Ariza	Finca 2	30	68
9	Diego Girón	Finca 4	30	114
10	Julio Villar	Finca 3	40,5	91
11	Carmen Pérez	Finca 1	30	15
12	Eva Morón	Finca 1	30	15
13	Isabel López	Finca 2	30	153

Figura 1.18: Pantalla de edición de datos.

Para finalizar la introducción de datos se guarda el fichero en el directorio de trabajo seleccionando la opción **Archivo / Guardar Como / Guardar Datos Como...** La extensión de los ficheros de Statgraphics es *.sf3*.

1.6. Ejercicios

1. *Introducir los siguientes datos:*

peso	altura	edad	sexo
75	173	21	hombre
81	172	22	hombre
56	136	22	mujer
68	180	21	mujer
79	182	24	hombre
89	185	24	hombre
62	157	21	hombre
59	165	22	mujer
83	180	23	hombre
55	160	22	mujer
72	174	21	hombre
56	161	23	mujer

Práctica 2

VARIABLES ESTADÍSTICAS UNIDIMENSIONALES

En Estadística la información se recoge en variables. Estas variables se organizan de forma ordenada y se almacenan en ficheros. Posteriormente será posible operar con estas variables y aplicar funciones a las mismas para realizar las transformaciones y los análisis estadísticos que se desee.

2.1. Introducción

Las variables estadísticas se dividen en variables cuantitativas (contienen datos numéricos) y cualitativas (contienen datos categóricos no cuantificables). A su vez, las variables cuantitativas pueden ser discretas y continuas.

Una variable es discreta cuando toma una cantidad finita o numerable de valores aislados y es continua cuando toma una infinidad no numerable de valores.

En este capítulo se verá como organizar y representar los datos contenidos en una única variable, y se habla, por tanto, de variable estadística unidimensional. Se explicará cómo calcular su tabla de frecuencias, así como la forma de obtener distintas representaciones gráficas.

2.2. Tabla de frecuencias y representación de variables discretas

La opción **Descripción / Datos Cualitativos / Tabulación...** permite el cálculo de la tabla de frecuencias (conjunto de los valores que presenta la variable junto con sus frecuencias: frecuencia absoluta, relativa, absoluta acumulada y relativa acumulada) de una variable estadística discreta, así como su representación mediante un diagrama de barras.

2.2. Tabla de frecuencias y representación de variables discretas



Figura 2.1: Menú tabulación.

2.2.1. Ejemplo práctico

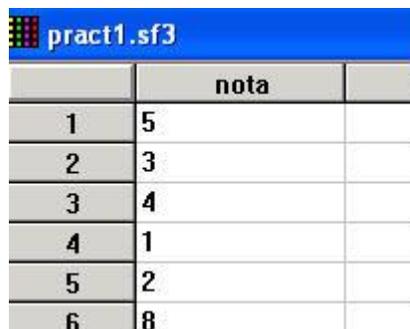
Las notas de Matemáticas de un grupo de personas han sido las siguientes:

5 3 4 1 2 8 9 8 7 6 6 7 9 8 7
7 1 0 1 5 9 9 8 0 8 8 8 9 5 7

Realizar la tabla de frecuencias y representación gráfica correspondiente a estos datos.

Se detallan los pasos a seguir:

1. En primer lugar se define la variable *nota* y se introducen los datos.



	nota	
1	5	
2	3	
3	4	
4	1	
5	2	
6	8	

Figura 2.2: Ejemplo de introducción de datos de una variable discreta.

2. Se selecciona en el *Menú principal* la opción **Descripción / Datos Cualitativos / Tabulación...**

En el *cuadro de diálogo* que aparece se introduce en el campo **Datos:** la variable que se va a analizar; para ello se selecciona con el ratón en el recuadro de la izquierda y se pulsa el icono que aparece debajo de la opción. (Figura 2.3)

Una vez seleccionada la variable se pulsa el botón **Aceptar** y se abre una nueva ventana de resultados.

Por defecto, aparecen algunos resultados del menú **Tabulación**, (ver Figura 2.4).



Figura 2.3: Cuadro de diálogo: Inserción de variable.

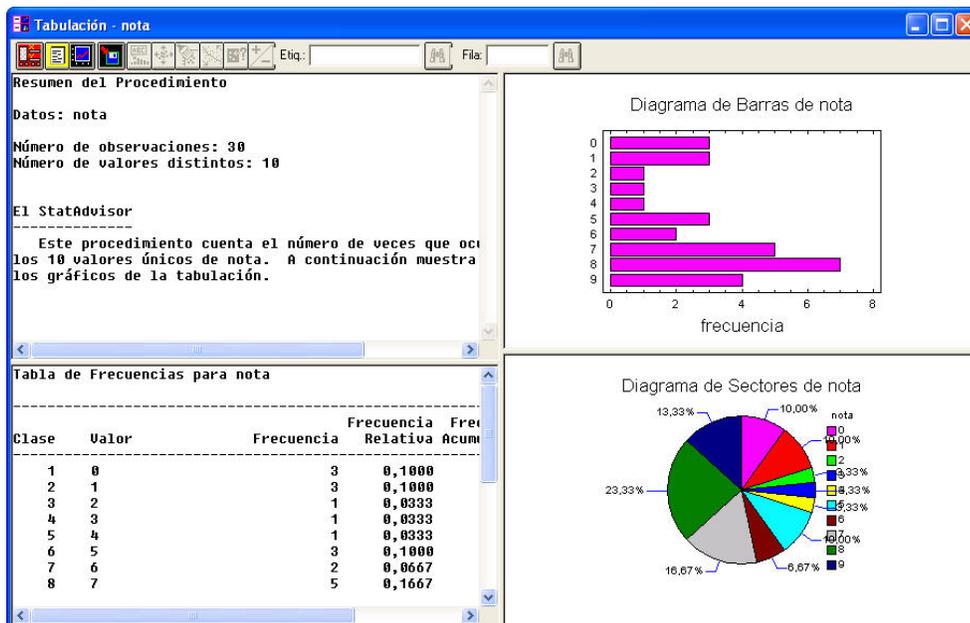


Figura 2.4: Ventana (por defecto) de resultados del análisis de la variable *nota*.

En esta ventana de análisis de resultados se observan tres zonas esenciales:

- a) La barra de icono y título del análisis. Presenta el nombre del procedimiento estadístico cuyos resultados se muestran en la ventana de análisis (en este caso se trata del procedimiento **Tabulación**). A continuación se expresa el nombre

2.2. Tabla de frecuencias y representación de variables discretas

de la variable afectada por el procedimiento estadístico (en este caso la variable *nota*).

- b) La barra de herramientas de análisis. Presenta una sucesión de iconos que van a posibilitar las diferentes opciones de trabajo en el análisis actual. Está situada inmediatamente debajo de la barra anterior (ver Figura 2.5).



Figura 2.5: Barra de herramientas de análisis.

- c) La ventana de salida de resultados. Se sitúa debajo de la barra de herramientas de análisis y se divide en dos zonas. La zona de la izquierda (zona de texto) presenta los resultados de análisis estadístico y la zona de la derecha (zona de gráficos) presenta los resultados gráficos, si existen. Si se pulsa dos veces con el ratón sobre cualquiera de las zonas se obtiene su maximización ocupando esa zona toda la pantalla. Se regresa a la situación anterior volviendo a pulsar dos veces con el ratón en cualquier parte de la pantalla maximizada. Cuando se maximiza la zona gráfica, el programa considera el gráfico como seleccionado y se activan las opciones gráficas de la barra de herramientas de análisis referentes al trabajo con gráficos.
3. Se observa la ventana de resultados obtenidos.

Se pulsa el icono de *Opciones Tabulares* y se obtiene un cuadro de diálogo (Figura 2.6)

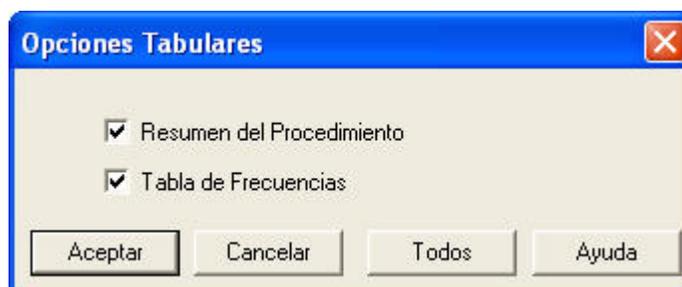


Figura 2.6: Opciones tabulares.

donde se pueden marcar y desmarcar las opciones que aparecen, según interés. En este caso, se dejan marcadas las dos opciones: **Resumen del Procedimiento** y **Tabla de Frecuencias**. Se pulsa **Aceptar**.

Se pulsa el icono de *Opciones Gráficas* y se obtiene un cuadro de diálogo (Figura 2.7)



Figura 2.7: Opciones gráficas.

donde se pueden marcar y desmarcar las opciones que aparecen, según interés. En este caso, se deja marcada la opción **Diagrama de Barras**, que es el gráfico que se utiliza para representar una variable estadística discreta. Se pulsa **Aceptar**.

Entonces, en la parte izquierda de la ventana de resultados, en la parte de abajo, aparece la tabla de frecuencias para la variable *nota*. Haciendo doble click sobre este trozo de ventana se puede maximizar el resultado para ver la tabla completa (Figura 2.8).

Tabla de Frecuencias para nota

Clase	Valor	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
1	0	3	0,1000	3	0,1000
2	1	3	0,1000	6	0,2000
3	2	1	0,0333	7	0,2333
4	3	1	0,0333	8	0,2667
5	4	1	0,0333	9	0,3000
6	5	3	0,1000	12	0,4000
7	6	2	0,0667	14	0,4667
8	7	5	0,1667	19	0,6333
9	8	7	0,2333	26	0,8667
10	9	4	0,1333	30	1,0000

Figura 2.8: Tabla de frecuencias de la variable *nota*.

Volviendo a hacer doble click sobre la pantalla anterior aparecen de nuevo todos los resultados. En la parte derecha aparece el diagrama de barras para la variable *nota*. Haciendo doble click sobre el gráfico se maximiza para verlo en detalle. Si se pulsa la opción **Opciones de Ventana** del botón derecho del ratón, existe la posibilidad de cambiar la dirección del gráfico, para que sea vertical. Obteniéndose finalmente el gráfico que se observa en la Figura 2.9.

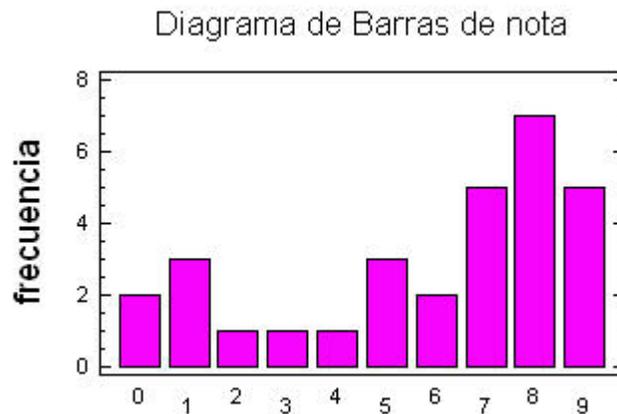


Figura 2.9: Diagrama de barras de la variable *nota*.

2.3. Tabla de frecuencias y representación de variables continuas

La opción **Descripción / Datos Numéricos / Análisis Unidimensional...** permite el cálculo de la tabla de frecuencias (conjunto de los valores que presenta la variable junto con sus frecuencias: frecuencia absoluta, relativa, absoluta acumulada y relativa acumulada) de una variable estadística continua, así como su representación mediante un histograma.

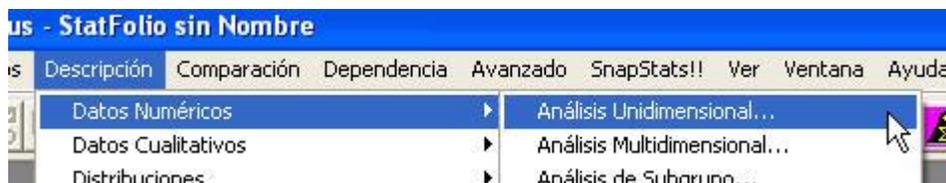


Figura 2.10: Menú **Análisis unidimensional...**

2.3.1. Ejemplo práctico

Se ha determinado el peso de 20 niños al nacer obteniéndose los siguientes resultados:

2,3 3 3,1 3,2 3,3 2,6 2,7 3,5 3,5 3,7
 4,1 4,4 3 3,2 3,3 3,3 3,1 2,8 3,6 3,4

Realizar la tabla de frecuencias y representación gráfica correspondiente a estos datos, agrupando los datos en 6 clases iguales.

A continuación, se detallan los pasos a seguir:

1. En primer lugar se define la variable *peso* y se introducen los datos.

	peso	Co
	2,3	
	3	
	3,1	
	3,2	
	3,3	
	2,6	
	2,7	

Figura 2.11: Datos de la variable *peso*.

2. Entonces se selecciona en el *Menú principal* la opción **Descripción / Datos Numéricos / Análisis Unidimensional...**

En el *cuadro de diálogo* que aparece se introduce en el campo **Datos:** la variable que se va a analizar; para ello se selecciona con el ratón en el recuadro de la izquierda y se pulsa el icono que aparece debajo de la opción. Una vez seleccionada la variable se pulsa el botón **Aceptar** y se abre una nueva ventana de resultados.

Por defecto, aparecen algunos resultados del menú **Análisis Unidimensional...**

3. Se observa la ventana de resultados obtenidos.

Se pulsa el icono de *Opciones Tabulares* y se obtiene la Figura 2.12,

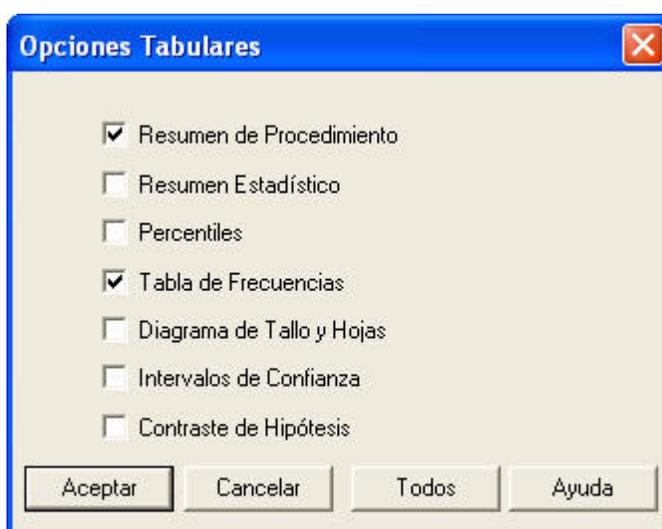


Figura 2.12: Opciones tabulares.

2.3. Tabla de frecuencias y representación de variables continuas

donde se pueden marcar y desmarcar las opciones que aparecen, según interese. En este caso, se deja marcada la opción: **Resumen del Procedimiento** y se marca **Tabla de Frecuencias**. Se pulsa **Aceptar**.

Se pulsa el icono de *Opciones Gráficas* y se obtiene la Figura 2.13,

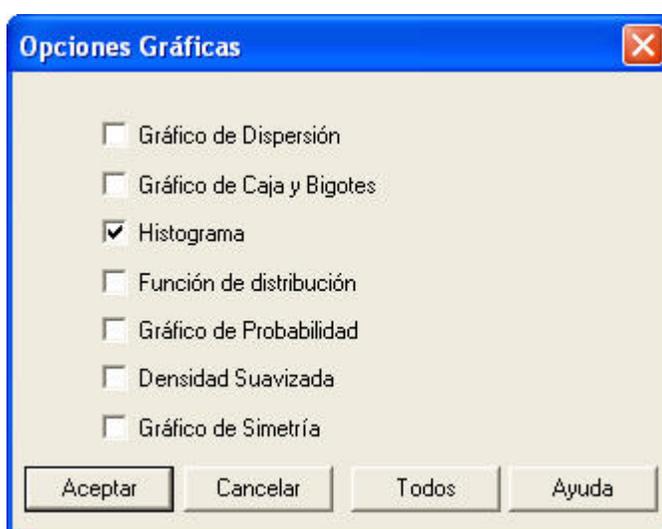


Figura 2.13: Opciones gráficas.

donde se pueden marcar y desmarcar las opciones que aparecen, según interese. En este caso, se quitan las marcas que aparecen por defecto y se marca la opción **Histograma**, que es el gráfico que se utiliza para representar una variable estadística continua. Se pulsa **Aceptar**.

Entonces, en la parte izquierda de la ventana de resultados, en la parte de abajo, aparece la tabla de frecuencias para la variable *peso*. Haciendo doble click sobre este trozo de ventana se puede maximizar el resultado para ver la tabla completa.

Tabla de Frecuencias para peso

Clase	Límite Inferior	Límite Superior	Marca	Frecuencia Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		2,1		0	0,0000	0	0,0000
1	2,1	2,51667	2,30833	1	0,0500	1	0,0500
2	2,51667	2,93333	2,725	3	0,1500	4	0,2000
3	2,93333	3,35	3,14167	9	0,4500	13	0,6500
4	3,35	3,76667	3,55833	5	0,2500	18	0,9000
5	3,76667	4,18333	3,975	1	0,0500	19	0,9500
6	4,18333	4,6	4,39167	1	0,0500	20	1,0000
mayor	4,6			0	0,0000	20	1,0000

Media = 3,255 Desviación típica = 0,488257

Figura 2.14: Tabla de frecuencias (por defecto) para la variable *peso*.

El ejercicio pide una tabla de frecuencias dividida en 6 clases. Para ello se procede a modificar los límites de los intervalos. Se pulsa sobre la tabla con el botón derecho y se selecciona **Opciones de ventana**, (ver Figura 2.16), se introducen unos valores que den una tabla de frecuencias fácil de interpretar.

Tabla de Frecuencias para peso

Clase	Límite Inferior	Límite Superior	Marca	Frecuencia Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		2,1		0	0,0000	0	0,0000
1	2,1	2,51667	2,30833	1	0,0500	1	0,0500
2	2,51667	2,93333	2,725	3	0,1500	4	0,2000
3	2,93333	3,35	3,14167	9	0,4500	13	0,6500
4	3,35	3,76667	3,55833	5	0,2500	18	0,9000
5	3,76667	4,18333	3,975	1	0,0500	19	0,9500
6	4,18333	4,6	4,39167	1	0,0500	20	1,0000
mayor	4,6			0	0,0000	20	1,0000

Media = 3,255 Desviación típica = 0,488257

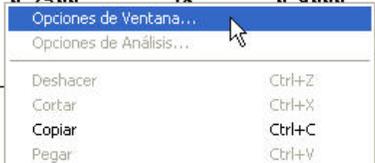


Figura 2.15: Opciones de ventana.

2.3. Tabla de frecuencias y representación de variables continuas

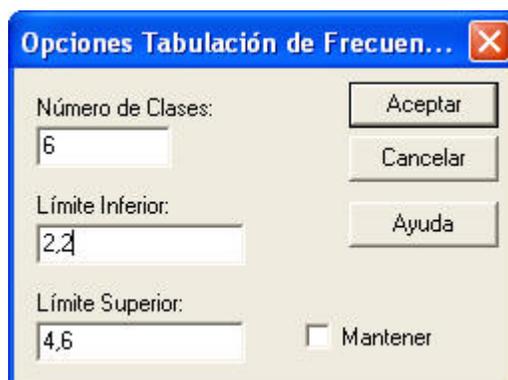


Figura 2.16: Opciones de ventana.

Por último, la tabla de frecuencias que se obtiene es la que aparece en la Figura 2.17.

Tabla de Frecuencias para peso

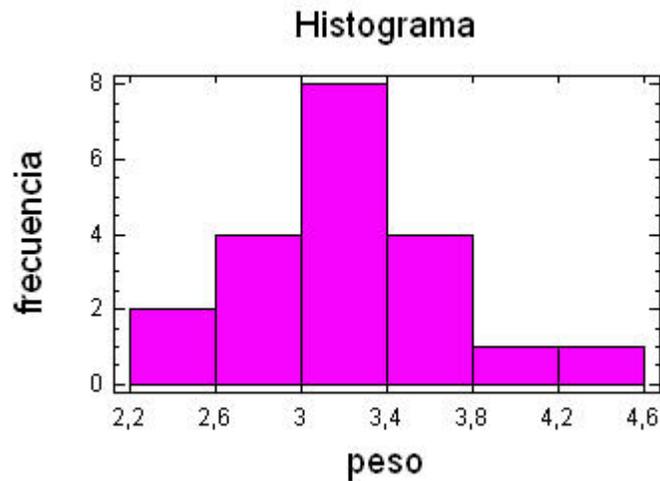
Clase	Límite Inferior	Límite Superior	Marca	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		2,2		0	0,0000	0	0,0000
1	2,2	2,6	2,4	2	0,1000	2	0,1000
2	2,6	3,0	2,8	4	0,2000	6	0,3000
3	3,0	3,4	3,2	8	0,4000	14	0,7000
4	3,4	3,8	3,6	4	0,2000	18	0,9000
5	3,8	4,2	4,0	1	0,0500	19	0,9500
6	4,2	4,6	4,4	1	0,0500	20	1,0000
mayor	4,6			0	0,0000	20	1,0000

Media = 3,255 Desviación típica = 0,488257

I

Figura 2.17: Tabla de frecuencias para la variable *peso*.

Volviendo a hacer doble click sobre la pantalla anterior aparecen de nuevo todos los resultados. En la parte derecha aparece el histograma para la variable *peso*. Haciendo doble click sobre el gráfico se maximiza para verlo en detalle (Figura 2.18).

Figura 2.18: Histograma para la variable *peso*.

2.4. Representación de variables cualitativas

La opción **Descripción / Datos Cualitativos / Tabulación...** permite la realización del gráfico de barras, así como el gráfico de sectores de una variable cualitativa, cuando los datos que se tienen son las propias observaciones.

Si en lugar de las observaciones se dispone de una tabla donde se reflejan las distintas modalidades de la variable junto con su frecuencia absoluta, entonces las opciones son **Gráficos / Diagramas de Presentación / Diagrama de Barras...** y **Gráficos / Diagramas de Presentación / Diagrama de Sectores...**, respectivamente, para realizar el diagrama de barras y el diagrama de sectores.

2.4.1. Ejemplo práctico

En la siguiente tabla se indican las cantidades en toneladas de siete especies desembarcadas en los puertos españoles así como su valor expresado en euros:

Especies	Toneladas	Euros
Almeja	4453	5640
Calamar	1644	21998
Cigala	7799	30253
Langosta	525	2568
Langostino	3548	10513
Mejillón	96253	11326
Pulpo	55205	51975

2.4. Representación de variables cualitativas

Representar los datos gráficamente, mediante un diagrama de barras y un diagrama de sectores.

Se detallan los pasos a seguir:

1. En primer lugar se definen las variables *especies*, *toneladas* y *euros*, y se introducen los datos.



	especies	toneladas	euros	
	Almeja	4453	5640	
	Calamar	1644	21998	
	Cigala	7799	30253	
	Langosta	525	2568	
	Langostino	3548	10513	
	Mejillón	96253	11326	
	Pulpo	55205	51975	

Figura 2.19: Ejemplo de datos de variables de tipo cualitativo.

2. Se va a realizar un diagrama de barras para cada una de las variables. Dado que en este caso los datos vienen dados a partir de las frecuencias de las modalidades se selecciona en el *Menú principal* la opción **Gráficos / Diagramas de Presentación / Diagrama de Barras...** (Figura 2.20).



Figura 2.20: Acceso al menú de la opción **Diagrama de Barras...**

En el *cuadro de diálogo* que aparece se introduce en el campo **Recuentos:** la variable que contiene las frecuencias; para ello se selecciona con el ratón en el recuadro de la izquierda y se pulsa el icono que aparece debajo de la opción, y en el campo **Etiquetas:** la variable que contiene los nombres de las modalidades de la variable cualitativa (ver Figura 2.21).



Figura 2.21: Introducción de variables.

Una vez seleccionadas las variables se pulsa el botón **Aceptar** y se abre una nueva ventana de resultados.

3. Se observa la ventana de resultados obtenidos.

En la parte izquierda de la ventana de resultados, aparece un resumen de la variable y a la derecha el gráfico representado. Haciendo doble click sobre esta parte de ventana se puede maximizar el resultado para ver el gráfico en detalle (ver Figura 2.22).

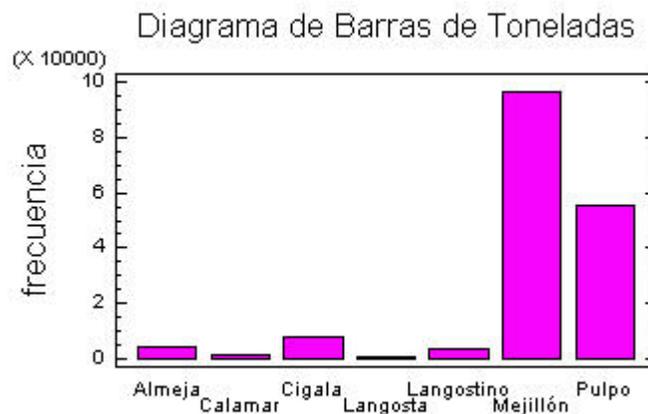


Figura 2.22: Diagrama de barras de la variable *toneladas*.

2.4. Representación de variables cualitativas

De igual forma se puede hacer el diagrama de barras para la variable *euros*.

4. Para representar las variables mediante un diagrama de sectores, se pulsa en el *Menú principal* la opción **Gráficos / Diagramas de Presentación / Diagrama de Sectores...** (Figura 2.23)



Figura 2.23: Acceso al menú **Diagrama de sectores...**

Para cada variable hay que seguir los mismos pasos que se han dado para realizar el diagrama de barras. Introducir el nombre de las variables en el lugar correspondiente y pulsar el icono **Aceptar** (ver Figura 2.24).

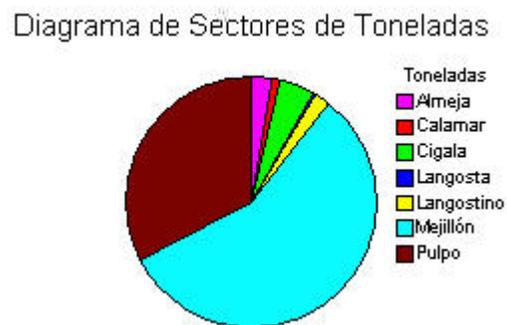


Figura 2.24: Diagrama de sectores para la variable *toneladas*.

Práctica 3

Síntesis numérica de una variable estadística

3.1. Introducción

Una vez definidos los conceptos básicos en el estudio de una distribución de frecuencias de una variable, se estudian las distintas formas de resumir dichas distribuciones mediante medidas de posición (o de centralización), teniendo presente el error cometido en el resumen mediante las correspondientes medidas de dispersión. A su vez se analizará la forma de la distribución mediante las medidas de forma. El histograma de frecuencias y el diagrama de barras, vistos en la práctica anterior, ya aportan una representación visual de las tres propiedades más importantes de los datos muestrales relativos a variables: la forma de su distribución, su tendencia central y su dispersión. Ahora se trata de cuantificar estos conceptos.

3.2. Medidas de posición, dispersión y forma

Statgraphics, mediante la opción **Descripción** de la barra de menú principal, permite analizar y resumir una distribución de frecuencias, posibilitando tabular los datos adecuadamente, hallar medidas de posición, dispersión y asimetría para los mismos. Asimismo permite realizar gráficos de barras, sectores, histogramas de frecuencias, etc, relativos a la distribución dada.

La opción **Descripción / Datos Numéricos / Análisis Unidimensional...** permite el cálculo de medias, varianzas, desviaciones típicas, medianas, modas, mínimos, máximos, rangos, cuartiles inferiores, cuartiles superiores, rangos intercuartílicos, coeficientes de asimetría y curtosis, y coeficiente de variación. La varianza y desviación típica usan como denominador $n - 1$.

3.2.1. Ejemplo práctico

Sean los siguientes datos que representan el peso en kg., la altura en cm. y el sexo de 12 individuos. Para este conjunto de datos se pide:

- a) Realizar un análisis descriptivo completo de las variables peso y altura, para todo el conjunto de individuos.
- b) A la vista de los resultados, decidir qué variable es más homogénea.
- c) Realizar un análisis descriptivo completo de las variables peso y altura, para cada grupo.

peso	altura	sexo
75	173	hombre
81	172	hombre
56	136	mujer
68	180	mujer
79	182	hombre
89	185	hombre
62	157	hombre
59	165	mujer
83	180	hombre
55	160	mujer
72	174	hombre
56	161	mujer

a) A continuación, se detallan los pasos a seguir:

1. En primer lugar, se definen las variables *peso*, *altura* y *sexo* y se introducen los datos. (Ver Figura 3.1)



peso	altura	sexo
75	173	hombre
81	172	hombre
56	136	mujer
68	180	mujer
79	182	hombre
89	185	hombre
62	157	hombre

Figura 3.1: Vista de algunos de los datos introducidos en Statgraphics.

- Entonces se selecciona en el *Menú principal* la opción **Descripción / Datos Numéricos / Análisis Unidimensional...**

En el *cuadro de diálogo* que aparece se introduce en el recuadro **Datos:** la variable que se va a analizar (en primer lugar, *peso*); para ello se selecciona con el ratón en el recuadro de la izquierda y se pulsa el icono que aparece debajo de la opción. Una vez seleccionada la variable se pulsa el botón **Aceptar** y se abre una nueva ventana de resultados.

Por defecto, aparecen algunos resultados del menú **Análisis Unidimensional...** (Ver Figura 3.2)

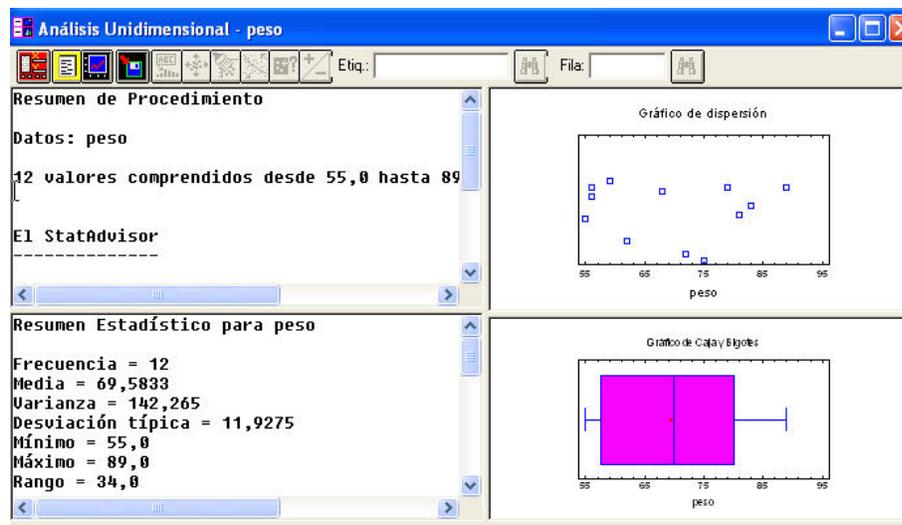


Figura 3.2: Resultados por defecto.

- Se observa la ventana de resultados obtenidos.

Pulsando el icono de *Opciones Tabulares* se obtiene la Figura 3.3,

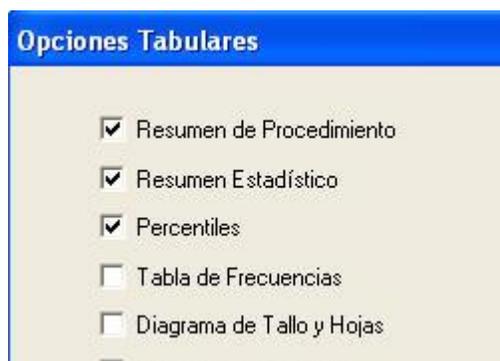


Figura 3.3: Opciones tabulares

donde se puede marcar y desmarcar las opciones que aparecen, según interés. En este caso, se dejan marcadas las opciones: **Resumen del Procedimiento** y **Resumen estadístico** y se marca **Percentiles**. Se pulsa **Aceptar**.

Entonces, en la parte central izquierda de la ventana de resultados aparecen algunas medidas para la variable *peso*. Haciendo doble click sobre esta parte de la ventana se puede maximizar esta ventana concreta de resultados. (Ver Figura 3.4)

```
Resumen Estadístico para peso
Frecuencia = 12
Media = 69,5833
Varianza = 142,265
Desviación típica = 11,9275
Mínimo = 55,0
Máximo = 89,0
Rango = 34,0
Primer cuartil = 57,5
Segundo cuartil = 80,0
Asimetría = 0,160211
Curtosis = -1,4683
```

Figura 3.4: Medidas descriptivas que aparecen por defecto.

Se observa que no aparecen, por ejemplo la mediana o la moda. Se pulsa sobre la ventana con el botón derecho y se selecciona **Opciones de ventana**, donde se marcan todas aquellas medidas que sean necesarias (ver Figura 3.5). Finalmente el resumen estadístico aparece en la Figura 3.6.



Figura 3.5: Opciones de ventana.

```
Resumen Estadístico para peso
Frecuencia = 12
Media = 69,5833
Mediana = 70,0
Moda = 56,0
Varianza = 142,265
Desviación típica = 11,9275
Mínimo = 55,0
Máximo = 89,0
Rango = 34,0
Primer cuartil = 57,5
Segundo cuartil = 80,0
Rango intercuar. = 22,5
Asimetría = 0,160211
Curtosis = -1,4683
Coef. de variación = 17,1413%
```

Figura 3.6: Resumen estadístico para la variable *peso*.

De los resultados se obtiene, por ejemplo: que el peso medio es de 69.58 kgs; la mitad de los individuos pesan menos de 70 kgs; lo más usual es un peso de 56 kgs. El coeficiente de asimetría da un valor de 0.160211, es decir, una ligera asimetría a la derecha y el coeficiente de curtosis es -1.4683, es decir, la distribución de las frecuencias es menos aplastada que la normal.

3.2. Medidas de posición, dispersión y forma

La información obtenida sobre la forma de la distribución se puede observar también si se representa un histograma de los datos. (Ver figura 3.7)

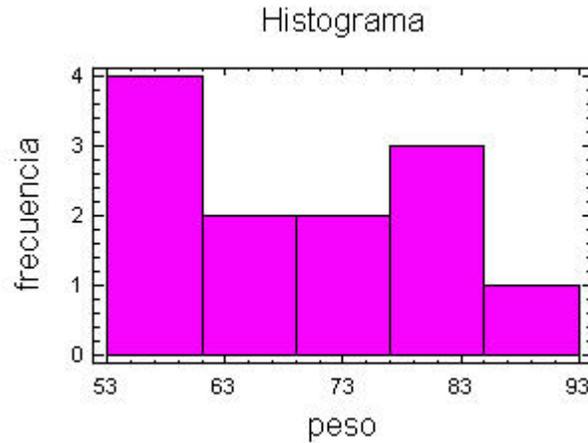


Figura 3.7: Histograma para la variable *peso*.

Volviendo a hacer doble click sobre la pantalla, ésta se maximiza apareciendo de nuevo todos los resultados. En la tercera ventana de la izquierda aparecen algunos percentiles para la variable *peso*. (Ver la Figura 3.8)

Percentiles para peso	
1,0%	= 55,0
5,0%	= 55,0
10,0%	= 56,0
25,0%	= 57,5
50,0%	= 70,0
75,0%	= 80,0
90,0%	= 83,0
95,0%	= 89,0
99,0%	= 89,0

Figura 3.8: Percentiles para la variable *peso*.

Los percentiles que se calculan se pueden modificar en el botón derecho y seleccionando la opción *Opciones de Ventana*.

4. Se repite el procedimiento con la variable *altura* y se obtienen los siguientes resultados para el resumen estadístico (ver la Figura 3.9) y los percentiles (ver la Figura 3.10).

```
Resumen Estadístico para altura
Frecuencia = 12
Media = 168,75
Mediana = 172,5
Moda = 180,0
Varianza = 191,841
Desviación típica = 13,8507
Mínimo = 136,0
Máximo = 185,0
Rango = 49,0
Primer cuartil = 160,5
Segundo cuartil = 180,0
Rango intercuar. = 19,5
Asimetría = -1,15844
Curtosis = 1,58801
Coef. de variación = 8,2078%
```

Figura 3.9: Resumen estadístico para la variable *altura*.

De los resultados obtenidos se puede resaltar que: la altura media es 168.75 cm; la mitad de los individuos miden menos de 172.5 cm y la altura que más se repite es 180 cm. El signo negativo del coeficiente de asimetría indica asimetría a la izquierda y el coeficiente de curtosis indica que la distribución es más apuntada que la normal.

```
Percentiles para altura
1,0% = 136,0
5,0% = 136,0
10,0% = 157,0
25,0% = 160,5
50,0% = 172,5
75,0% = 180,0
90,0% = 182,0
95,0% = 185,0
99,0% = 185,0
```

Figura 3.10: Percentiles para la variable *altura*.

3.2. Medidas de posición, dispersión y forma

b) La variable más homogénea es la que presente un coeficiente de variación inferior. En este caso, los valores obtenidos (en porcentaje) son:

- Variable *peso*: 17.1413 %
- Variable *altura*: 8.2078 %

Luego la variable *altura* es más homogénea. Sus valores se encuentran más concentrados en torno al valor medio de la variable.

c) A continuación se detallan los pasos a seguir:

1. En primer lugar, se selecciona en el *Menú principal* la opción **Descripción / Datos numéricos / Análisis de Subgrupo...** (Figura 3.11).

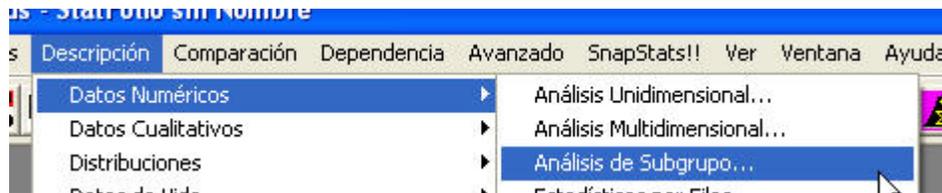


Figura 3.11: Menú **Análisis de Subgrupo**.

En el *cuadro de diálogo* que aparece hay que introducir en el recuadro **Datos:** la variable que se va a analizar (en primer lugar, *altura*); para ello se selecciona con el ratón en el recuadro de la izquierda y se pulsa el icono que aparece debajo de la opción y en el recuadro **Códigos:** la variable que diferencia los grupos, en este caso *sexo*. Una vez seleccionadas las variables se pulsa el botón **Aceptar** y se abre una nueva ventana de resultados.

Por defecto, aparecen algunos resultados del menú **Análisis de Subgrupo...** (Ver Figura 3.12)

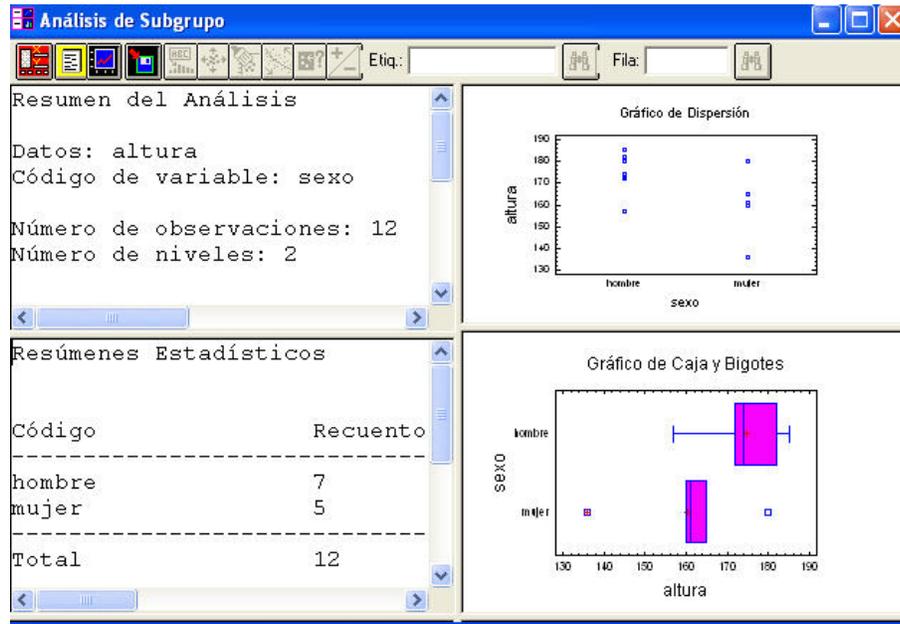


Figura 3.12: Resultados por defecto.

2. Se observa la ventana de resultados obtenidos.

Pulsando el icono de *Opciones Tabulares* se obtiene la Figura 3.13,



Figura 3.13: Opciones tabulares

donde se puede marcar y desmarcar las opciones que aparecen, según interese. En este caso, dejamos marcadas las opciones: **Resumen del Análisis**, y **Resumen estadístico**.

Entonces, en la parte inferior izquierda de la ventana de resultados aparecen algunas medidas para la variable *altura*, en cada uno de los grupos. Haciendo doble click sobre

3.2. Medidas de posición, dispersión y forma

esta parte de la ventana se puede maximizar el resultado para verlas todas. (Ver Figura 3.14)

Código	Recuento	Media	Mediana	Desviación Típica
hombre	7	174,714	174,0	9,23245
mujer	5	160,4	161,0	15,8209
Total	12	168,75	172,5	13,8507

Figura 3.14: Medidas descriptivas que aparecen por defecto.

De nuevo no aparecen, por ejemplo la mediana o la moda. Se pulsa sobre la ventana con el botón derecho y se selecciona **Opciones de ventana**, donde se marcan todas aquellas medidas que sean necesarias. Finalmente el resumen estadístico aparece en la Figura 3.15.

Código	Recuento	Media	Mediana	Moda	Desviación Típica	Asimetría
hombre	7	174,714	174,0		9,23245	-1,17458
mujer	5	160,4	161,0		15,8209	-0,725991
Total	12	168,75	172,5	180,0	13,8507	-1,15844

Código	Curtosis	Coef. de variación
hombre	1,8787	5,28431%
mujer	2,02553	9,86339%
Total	1,95033	8,48207%

Figura 3.15: Resumen estadístico para la variable *altura*, por *sexo*.

De los resultados se obtiene, por ejemplo: que la altura media en los hombres es de 174.714, mientras que en las mujeres es de 160.4; la mitad de los hombres miden más de 174 cm y la mitad de las mujeres miden más de 161 cm; destaca el hecho de que no aparece ningún valor para la moda en los grupos, lo que quiere decir que en este caso no hay ningún valor que se repita más que otro o que pueden existir más de una moda. El coeficiente de asimetría da en ambos casos valores negativos, es decir, asimetría a la izquierda y el coeficiente de curtosis, indica que la distribución de las frecuencias es más apuntada que la normal, en los dos grupos.

- Un estudio similar se realiza con la variable *peso* y el resultado que se obtiene se puede ver en la figura 3.16.

Práctica 3. Síntesis numérica de una variable estadística

Código	Recuento	Media	Mediana	Moda	Desviación Típica	Asimetría
hombre	7	77,2857	79,0		8,69318	-0,668016
mujer	5	58,8	56,0	56,0	5,35724	1,84257
Total	12	69,5833	70,0	56,0	11,9275	0,160211

Código	Curtosis	Coef. de variación
hombre	0,0769122	11,2481%
mujer	3,37469	9,11095%
Total	-1,4683	17,1413%

Figura 3.16: Resumen estadístico para la variable *peso*, por *sexo*.

3.3. Ejercicios

1. Una editorial pretende sacar al mercado una nueva revista científica. Se estudia el número de páginas de los artículos de este tipo para decidir el número de ellos que se pueden aceptar para cada volumen de la nueva revista. Se observan los artículos de una revista con una temática similar obteniéndose el número siguiente de páginas:

5	10	6	8	9	12	15	16	9	5
7	13	14	8	12	10	8	6	15	8
13	9	16	10	8	12	7	9	11	9

Determine:

- a) Tipo de variable estadística.
- b) Tabla de frecuencias, diagrama de tallo y hojas e histograma.
- c) Número medio de páginas, mediana y número de páginas más frecuente.
- d) Número máximo de páginas del 10 % de los artículos más cortos.
- e) Dispersión, asimetría y curtosis.
- f) Comentar el resultado de todas las medidas de los apartados anteriores.
- g) Compara la forma de la distribución y la dispersión con la del número de páginas de los artículos de otra revista en la que se han observado los siguientes valores:

10	16	19	15	14	13	12	10	17	13
18	11	15	10	14	19	10	12	11	15
12	14	13	19	16	19	16	15	18	12
19	18	17	16						

2. La EPA (Agencia de Protección del medio ambiente de EEUU) exige a los fabricantes de automóviles que indiquen los consumos, tanto por ciudad como por carretera, de cada uno de sus modelos. A continuación se presentan los datos de consumo por carretera de 30 modelos de automóvil de 1994, expresados en litros por cada 100 km:

12.3	9.1	10.1	10.4	9.7	11.3	11.3	10.8	9.7	10.1
10.8	10.5	11.3	9.7	10.8	10.5	12.3	11.3	9.1	11.8
11.8	14.1	10.8	18.8	10.8	10.4	10.1	10.8	12.8	10.8

Analizar descriptivamente la variable y calcular todas las medidas descriptivas estudiadas en el tema 2. Comenta el resultado de dichas medidas.

Práctica 4

Distribuciones estadísticas bidimensionales

4.1. Distribuciones bidimensionales de frecuencias

La opción **Descripción / Datos numéricos / Análisis Multidimensional...** del *Menú principal* permite realizar varios análisis estadísticos y gráficos sobre variables multidimensionales numéricas, y en particular, sobre variables bidimensionales.



Figura 4.1: Menú Análisis Multidimensional.

Como en todo análisis con Statgraphics, en la barra de herramientas de análisis se pueden elegir dos iconos esenciales, el icono **Opciones Tabulares** y el icono **Opciones Gráficas**, que permiten manejar todas las subopciones tabulares y gráficas, respectivamente, para el análisis.

La opción tabular **Resumen del Procedimiento** es la opción por defecto, y presenta un primer resumen de los datos. La opción **Resumen Estadístico** incluye por defecto varias medidas de posición, dispersión, asimetría y curtosis (número de valores de las variables, medias, cuasivarianzas, cuasidesviaciones típicas, máximos, mínimos, coeficientes de asimetría y curtosis, y sumas de los valores de las variables).

4.1.1. Ejemplo práctico

El consumo C y la renta mensual RM de 100 familias, expresadas en 10^2 euros se presentan en la siguiente tabla bidimensional de frecuencias.

$RM \setminus C$	[6,12]	(12,18]	(18,24]	(24,30]
[15,21]	10	15	0	0
(21,27]	5	20	25	0
(27,33]	0	15	5	5

Se pide:

- Obtener las distribuciones de frecuencias marginales de cada variable.
- Obtener las medias, varianzas, desviaciones típicas y coeficientes de variación, asimetría y curtosis de ambas variables conjuntamente. Y la covarianza entre ambas variables.
- Representar los datos mediante un diagrama de dispersión.

En este caso se observa que los valores de las variables vienen dados en una tabla bidimensional de frecuencias absolutas. El primer paso será introducir adecuadamente los valores de las dos variables como dos columnas de la hoja de cálculo de Statgraphics. El par de valores $RM=18$ y $C=9$ (que son las marcas de clase) se introduce 10 veces, el par de valores $RM=24$ $C=9$ se introduce 5 veces, y así sucesivamente se introduce cada par de valores tantas veces como indique su frecuencia absoluta.

a) Las distribuciones de frecuencias marginales se obtienen desde el menú **Descripción / Datos numéricos / Análisis Unidimensional...**, como ya se vió en la práctica 2. Los resultados que se obtienen son:

Tabla de Frecuencias para C

Clase	Límite Inferior	Límite Superior	Marca	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		6,0		0	0,0000	0	0,0000
1	6,0	12,0	9,0	15	0,1500	15	0,1500
2	12,0	18,0	15,0	50	0,5000	65	0,6500
3	18,0	24,0	21,0	30	0,3000	95	0,9500
4	24,0	30,0	27,0	5	0,0500	100	1,0000
mayor	30,0			0	0,0000	100	1,0000

Media = 16,5 Desviación típica = 4,62208

Figura 4.2: Tabla de frecuencias para Consumo.

En las tablas de frecuencias (Figuras 4.2, 4.3) se ha modificado el número de clases para obtener exactamente los intervalos que aparecen en el enunciado del ejercicio.

Tabla de Frecuencias para RM

Clase	Límite Inferior	Límite Superior	Marca	Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		15,0		0	0,0000	0	0,0000
1	15,0	21,0	18,0	25	0,2500	25	0,2500
2	21,0	27,0	24,0	50	0,5000	75	0,7500
3	27,0	33,0	30,0	25	0,2500	100	1,0000
mayor	33,0			0	0,0000	100	1,0000

Media = 24,0 Desviación típica = 4,26401

Figura 4.3: Tabla de frecuencias para Renta Mensual.

b) Las medidas descriptivas para ambas variables conjuntamente se obtienen desde el menú **Descripción / Datos Numéricos / Análisis Multidimensional...** En el *cuadro de diálogo* que aparece se introducen las variables en el campo **Datos:** (Figura 4.4) y se pulsa **Aceptar**.



Figura 4.4: Cuadro de diálogo del menú Análisis Multidimensional.

Por defecto aparecen algunas opciones tabulares y gráficas en el análisis resultante. Para obtener las medidas descriptivas se selecciona la opción **Resumen Estadístico** y para obtener la covarianza se selecciona la opción **Covarianzas**, ambas en el icono **Opciones Tabulares** (Figura 4.5).

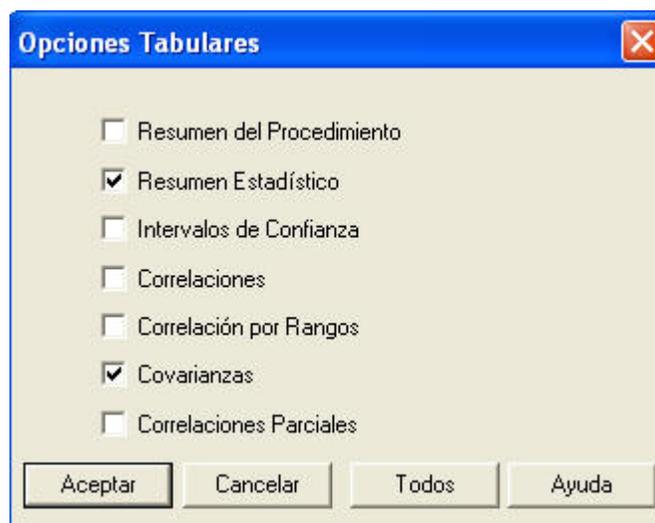


Figura 4.5: Opciones tabulares del menú Análisis Multidimensional.

El resultado del resumen estadístico se encuentra en la Figura 4.6.

Resumen Estadístico		
	C	RM
Frecuencia	100	100
Media	16,5	24,0
Varianza	21,3636	18,1818
Desviación típica	4,62208	4,26401
Asimetría	0,211374	0,0
Curtosis	-0,252738	-0,989375
Coef. de variación	28,0126%	17,7667%

Figura 4.6: Resumen estadístico.

De los resultados se puede decir que el Consumo medio son 1650 euros y la Renta Mensual media es de 2400 euros. La variable Consumo es ligeramente asimétrica a la derecha, (coeficiente > 0) y la variable Renta Mensual es simétrica (coeficiente $= 0$). Con respecto a los coeficientes de variación, la variable Renta Mensual es más homogénea con respecto a su media, ya que tiene un coeficiente de variación menor.

En la matriz de covarianzas, en la diagonal están las varianzas (divididas por $n - 1$) de cada variable y fuera de la diagonal la covarianza entre las variables. La covarianza entre las variables es 9,09091 (Figura 4.7). Lo que indica que existe relación directa entre las variables, es decir, a mayor Renta Mensual, mayor Consumo.

Covarianzas		
	C	RM
C	21,3636 (100)	9,09091 (100)
RM	9,09091 (100)	18,1818 (100)

Figura 4.7: Covarianza.

c) El diagrama de dispersión se realiza desde la opción **Gráficos / Gráficos de Dispersión / Gráfico X-Y...** (Figura 4.8).

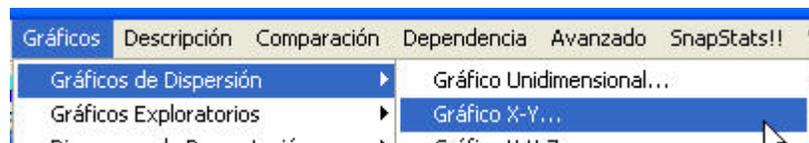


Figura 4.8: Menú Gráfico X-Y.

En el *cuadro de diálogo* que aparece se introduce una variable en el campo **Y**: y la otra en **X**:. El gráfico que se obtiene aparece en la Figura 4.9.

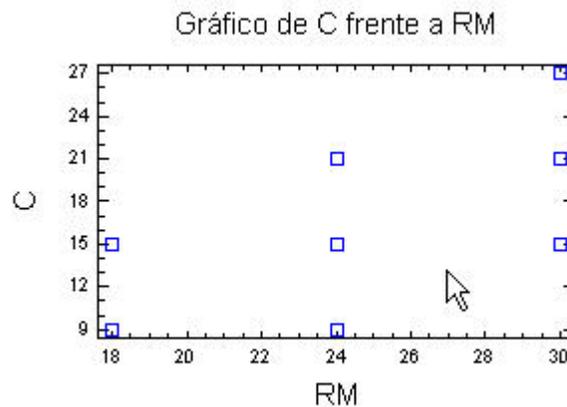


Figura 4.9: Gráfico de dispersión correspondiente a las variables Consumo y Renta Mensual.

4.2. Distribuciones condicionadas

La opción **Descripción / Datos numéricos / Análisis Unidimensional...** del *Menú principal* permite realizar resúmenes estadísticos y gráficos para una selección de un conjunto de datos. Por lo que esta opción se puede utilizar para obtener distribuciones condicionadas.

4.2.1. Ejemplo práctico

Se consideran los datos usados en el ejemplo anterior, correspondientes a la Renta Mensual y al Consumo de 100 familias. Se pide obtener la distribución de frecuencias y algunas medidas descriptivas del Consumo para aquellas familias que tienen una Renta Mensual superior a 2000 euros.

Para obtener la distribución de frecuencias se selecciona la opción **Descripción / Datos numéricos / Análisis Unidimensional...** y en el *cuadro de diálogo* que aparece se introduce en el campo **Datos:** la variable C (consumo) y en **(Selección:)** $RM > 20$. (Ver figura 4.10).



Figura 4.10: Cuadro de diálogo.

En el icono **Opciones Tabulares** se seleccionan **Resumen estadístico** y **Tabla de frecuencias**. Los resultados obtenidos se pueden ver en las Figuras 4.11 y 4.12.

```

Resumen Estadístico para C

Frecuencia = 75
Media = 17,8
Varianza = 18,8108
Desviación típica = 4,33714
Mínimo = 9,0
Máximo = 27,0
Rango = 18,0
Asimetría = 0,119259
Asimetría tipi. = 0,421643
Curtosis = -0,180802
Curtosis típificada = -0,319616
Coef. de variación = 0,244
    
```

Figura 4.11: Resumen estadístico para $C|RM>20$.

Tabla de Frecuencias para C

Clase	Límite Inferior	Límite Superior	Marca	Frecuencia Frecuencia	Frecuencia Relativa	Frecuencia Acumulativa	Frecuencia Acum.Rel.
menor o igual		6,0		0	0,0000	0	0,0000
1	6,0	12,0	9,0	5	0,0667	5	0,0667
2	12,0	18,0	15,0	35	0,4667	40	0,5333
3	18,0	24,0	21,0	30	0,4000	70	0,9333
4	24,0	30,0	27,0	5	0,0667	75	1,0000
mayor	30,0			0	0,0000	75	1,0000

Media = 17,8 Desviación típica = 4,33714

Figura 4.12: Tabla de frecuencias para $C|RM>20$.

4.3. Ejercicios

1. Sobre un conjunto de conductores se ha realizado una encuesta para analizar su edad (Y) y el número de accidentes que han sufrido (X). A partir de la misma, se obtuvieron los siguientes resultados:

$X \ Y$	[20,30]	(30,40]	(40,50]	(50,60]	(60,70]
0	74	82	78	72	7
1	7	6	5	6	5
2	3	2	2	1	1

A partir de estos datos, se pide determinar para este conjunto de conductores:

- a) Las medias y varianzas de las distribuciones marginales.
- b) ¿Qué variable es más homogénea?
- c) El número medio de accidentes que tienen los conductores con edades comprendidas entre los 40 y los 50 años.
- d) La covarianza entre las dos variables.
- e) ¿Qué edad tendría como mucho un conductor con un solo accidente para encontrarse en el grupo del 20% de los más jóvenes?
- f) Obtener el diagrama de dispersión de los datos.

4.4. Análisis bidimensional de datos categóricos

Cuando se trabaja con datos categóricos (o cualitativos), se utiliza el siguiente menú: **Descripción / Datos Cualitativos...**

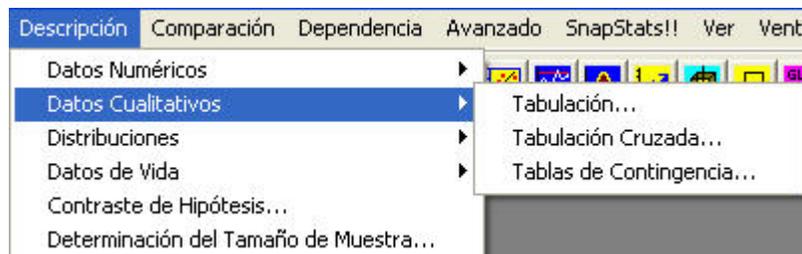


Figura 4.13: Menú Datos Cualitativos...

Dentro de esta opción se puede elegir entre varias opciones, y estudiar concretamente el procedimiento **Tabulación**. Sean los siguientes procedimientos:

- **Tabulación Cruzada**

Procedimiento que resume la distribución conjunta de dos variables discretas mediante su tabulación bidimensional. Los recuentos se efectúan midiendo el número de veces que aparece cada modalidad de la primera variable con la otra de la segunda variable.

- **Tablas de Contingencia**

Es análogo al anterior y su particularidad es que la información se introduce ya tabulada.

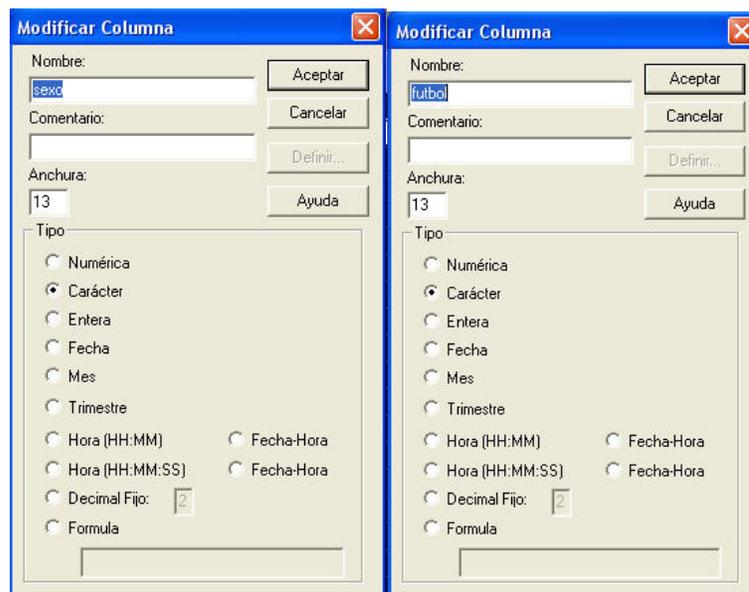
4.4.1. Ejemplo práctico

Se utilizan los datos provenientes de clasificar a 13 individuos según la variable sexo (hombre y mujer) y fútbol (deseo de ver una final de campeonato de fútbol) y que se muestran en la siguiente tabla:

4.5. Procedimiento Tabulación Cruzada

SEXO	FUTBOL
hombre	si
hombre	si
mujer	no
hombre	no
mujer	si
mujer	no
mujer	no
hombre	si
mujer	no
hombre	si
mujer	no
hombre	si
hombre	si

Antes de empezar con el análisis se introducen los datos, teniendo en cuenta que al definir la variable se tiene que indicar que son categóricos.



4.5. Procedimiento Tabulación Cruzada

Se accede mediante el menú **Descripción / Datos Cualitativos / Tabulación Cruzada** obteniendo un cuadro de diálogo en el que se tiene que decidir que variable aparecerá por filas y cual por columnas. En **Variable Fila** se introduce la variable que apare-

cerá por filas, en este caso, por ejemplo, *fútbol* y en **Variable Columna** se introduce la variable que aparecerá por columnas, en este caso *sexo*.



Figura 4.15: Tabulación Cruzada.

La primera ventana de resultados muestra el resumen del procedimiento: variables que intervienen y en que orden, el número de observaciones y el número de columnas y filas (modalidades de cada una de las variables), además de un diagrama de barras para cada una de las modalidades.

4.5. Procedimiento Tabulación Cruzada

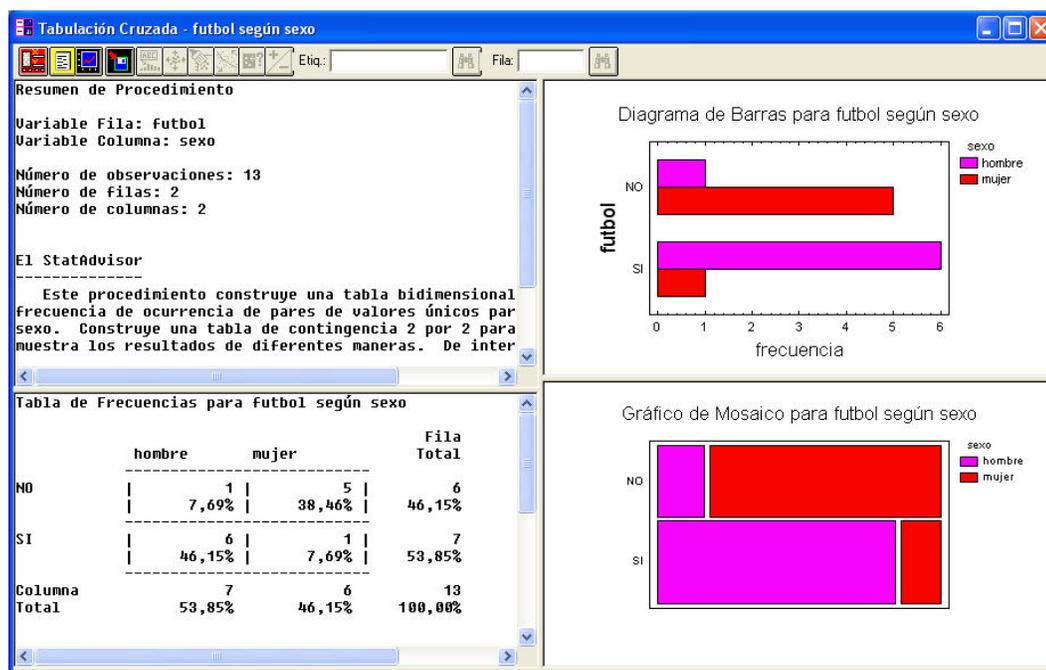


Figura 4.16: Ventana de Análisis en Tabulación Cruzada

Al pulsar **Opciones Tabulares** se obtiene una serie de opciones que son:

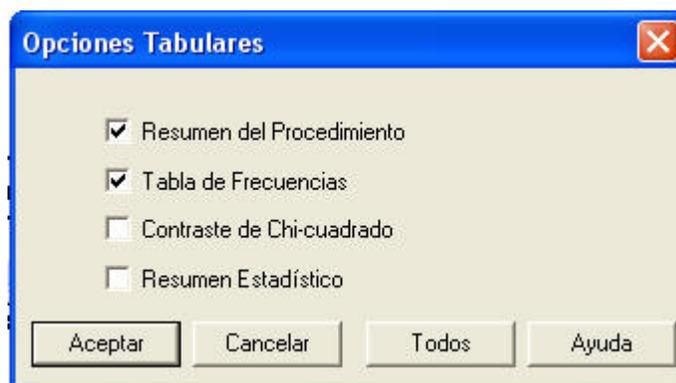


Figura 4.17: Opciones Tabulares en Tabulación Cruzada

4.5.1. Tabla de frecuencias

Al elegir la opción de **Tabla de Frecuencias**, se crea una tabla con las frecuencias absolutas para cada par de valores y la frecuencia relativa con respecto al total de la tabla. Si se pulsa el botón derecho del ratón se obtienen porcentajes con respecto a las filas (se

divide cada casilla entre el total de su fila), con respecto a su columna (se divide cada casilla respecto al total de su columna), la frecuencia esperada (resultado de multiplicar en cada casilla el total de su fila por el de su columna y dividirlo entre el total de la tabla) las desviaciones y los totales chi-cuadrado.



	hombre	mujer	Fila Total
NO	1 7,69% 16,67% 14,29%	5 38,46% 83,33% 83,33%	6 46,15%
SI	6 46,15% 85,71% 85,71%	1 7,69% 14,29% 16,67%	7 53,85%
Columna Total	7 53,85%	6 46,15%	13 100,00%

Contenido de Celda:
 Frecuencia Observada
 Porcentaje de tabla
 Porcentaje de fila
 Porcentaje de columna

Figura 4.18: Opciones de la tabla de frecuencias

En este resultado se puede comentar por ejemplo:

- Hay un 7.69% (en negro) de hombres que no quieren ver la final del campeonato (porcentaje del total, todos los individuos entran en el calculo del porcentaje).
- Del grupo de las personas que no quieren ver la final del campeonato, un 16.67% (en rojo) son hombres (porcentaje de fila, solo los individuos que presentan la modalidad de esa fila entran en el calculo del porcentaje, en este caso, solo las personas que no quieren ver la final del campeonato).
- De los hombres (solamente ese grupo), un 14.29% (en verde) no quieren ver la final del campeonato (porcentaje de columna, solo los individuos que presenten la modalidad de esa columna entran en el calculo del porcentaje, en este caso, solo los hombres).

4.5.2. Test de la χ^2

El contraste de la χ^2 proporciona un test para determinar si existe o no independencia entre las variables con las que se trabaja. Es un procedimiento inferencial en el que si el p -valor proporcionado por el test es inferior a 0.05 se concluye rechazar la hipótesis de independencia entre filas y columnas. La corrección de Yates se aplicará cuando las

4.6. Procedimiento Tablas de Contingencia

frecuencias esperadas sean menores de 5 y el contraste exacto de Fisher para tablas de 2 x 2. En este caso se obtiene:

Contraste de Chi-cuadrado		
Chi-cuadrado	GL	P-Valor
6,20	1	0,0128
3,73	1	0,0534 (con la corrección de Yates)

Precaución: La frecuencia de alguna celda es inferior a 5.

Contraste exacto de Fisher para tablas 2x2

P-Valor a una cola = 0,0250583
P-Valor a dos colas = 0,0291375

Figura 4.19: Contraste de hipótesis.

Por lo tanto, al ser el p -valor menor de 0.05, se rechaza la hipótesis de independencia para aceptar la dependencia (relación) entre ambas variables.

4.6. Procedimiento Tablas de Contingencia

Este procedimiento es análogo al procedimiento **Tabulación Cruzada**, la única diferencia está en la forma en la que se introducen los datos ya que en este caso la información está tabulada.

4.6.1. Ejemplo práctico

Se clasifican las notas de una asignatura en tres grupos distintos. Los datos aparecen en la siguiente tabla:

GRUPO	SUSPENSOS	APROBADOS	NOTABLES	SOBRESALIENTES
A	10	32	12	40
B	14	12	10	30
C	8	10	6	16

En el editor de datos se introducen los datos en cinco columnas semejantes a las del enunciado. La primera columna contiene a la variable *Grupo* y es de tipo categórica, las otras cuatro variables son numéricas y serán, respectivamente, *suspensos*, *aprobados*, *notables* y *sobresalientes* como se muestra en la imagen.

A continuación se accede al menú **Descripción / Datos Cualitativos / Tablas de Contingencia** tras el que aparece el siguiente *cuadro de dialogo*. En **Columnas** se

	Grupo	suspensos	aprobados	notables	sobresalientes
1	A	10	32	12	40
2	B	14	12	10	30
3	C	8	10	6	16

Figura 4.20:

introducen las variables que se han definido como numéricas y en **Etiquetas** la variable definida como categórica. A partir de aquí todo será igual que en el apartado anterior:

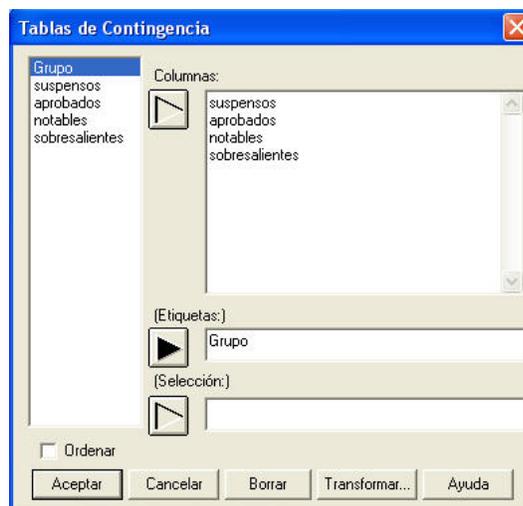


Figura 4.21: Cuadro de diálogo

4.7. Ejercicios

- La siguiente tabla clasifica a un número de personas según la frecuencia con que *leen periódicos o escuchan tertulias radiofónicas*.

Lectura prensa	tertulia sí	tertulia no
Todos los días	20	70
Alguna vez	60	70
Nunca	30	30

- Introduzca los datos y obtenga la tabla que muestre la distribución conjunta en porcentajes.
 - Obtener la tabla que exprese las distribuciones condicionadas de *LECTURA PRENSA* a las opciones de *TERTULIA*.
 - ¿Son independientes las dos variables?
- Se clasifican 6800 varones atendiendo a dos características, *el color de pelo y el color de los ojos*. Estudiar si hay o no independencia.

Ojos	pelo			
	rubio	moreno	pelirrojo	canoso
azul	1768	807	186	47
verde	946	1387	746	53
castaño	115	438	288	16

- De las personas con el pelo rubio, ¿qué porcentaje tienen los ojos azules?
 - De las personas con ojos azules, ¿qué porcentaje tienen el pelo rubio?
 - De las personas con el pelo canoso, ¿qué porcentaje tienen los ojos castaños?
 - De las personas con ojos castaños, ¿qué porcentaje tienen el pelo canoso?
 - ¿Cuál es el porcentaje que tiene el pelo canoso y los ojos castaños?
- Un centro escolar ha encargado a un psicólogo que estudie si hay relación entre el sexo y la preferencia por una carrera universitaria (ciencias, ingeniería y letras) en los alumnos de Bachiller. Para ello el psicólogo ha seleccionado una muestra aleatoria de 100 alumnos y les ha preguntado su opinión como se muestra en la siguiente tabla:

Sexo \ Carrera	Ciencias	Ingeniería	Letras
Varón	16	30	14
Mujer	14	10	16

¿Cuál será la conclusión del psicólogo con $\alpha = 0,05$?

Práctica 5

Regresión y Correlación

5.1. Introducción

Statgraphics analiza ampliamente el tema de la regresión. Aporta procedimientos para regresión simple lineal y no lineal. Los procedimientos simples a utilizar se implementan en la opción **Dependencia** de la barra de *Menú principal*.

La opción **Dependencia / Regresión Simple...** resuelve un modelo de regresión con una sola variable independiente, utilizando el ajuste por mínimos cuadrados.

Se puede representar gráficamente la línea de ajuste y los residuos para todos los modelos. El sistema, además, genera y representa predicciones para valores dados de X o Y , así como permite guardar en archivos los residuos y las predicciones para posteriores análisis.

Con este análisis se pretende conocer si una o varias variables independientes o explicativas aportan algún tipo de información sobre el comportamiento de otra variable dependiente o explicada. Una de las finalidades de este análisis es realizar predicciones para los datos. En esta práctica se considera el caso en el que se dispone de una única variable independiente (regresión simple), es decir, hay que expresar una variable independiente en función de una variable dependiente.

En regresión lineal, la función mediante la cuál se expresa esta relación será del tipo:

$$Y = a + bX$$

donde los parámetros a y b serán determinados a partir de los datos. Evidentemente el objetivo será que este ajuste sea lo mejor posible.

5.2. Regresión lineal simple

El procedimiento a utilizar es **Dependencia / Regresión Simple...**

5.2. Regresión lineal simple

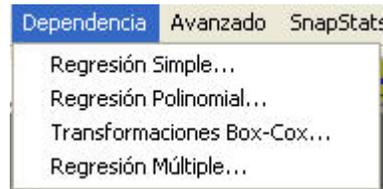


Figura 5.1: Menú **Regresión lineal...**

5.2.1. Ejemplo práctico

Sean los siguientes datos que representan el *Peso* en kg. y la *Altura* en cm. de 12 individuos. Con este conjunto de datos se procede a obtener posibles resultados en el análisis de regresión:

peso	altura
75	173
81	172
56	136
68	180
79	182
89	185
62	157
59	165
83	180
55	160
72	174
56	161

Una vez introducidos los datos se selecciona en el *Menú principal* la opción **Dependencia / Regresión Simple...**

En el *Cuadro de diálogo* que aparece se introducen las variables a analizar.



Figura 5.2: Cuadro de Entrada de datos.

En el campo **Y**: se introduce la variable dependiente, en este caso *Peso* y en **X**: la variable independiente, en este caso *Altura*. Al pulsar el botón **Aceptar** se genera la ventana de resultados correspondiente a este análisis.

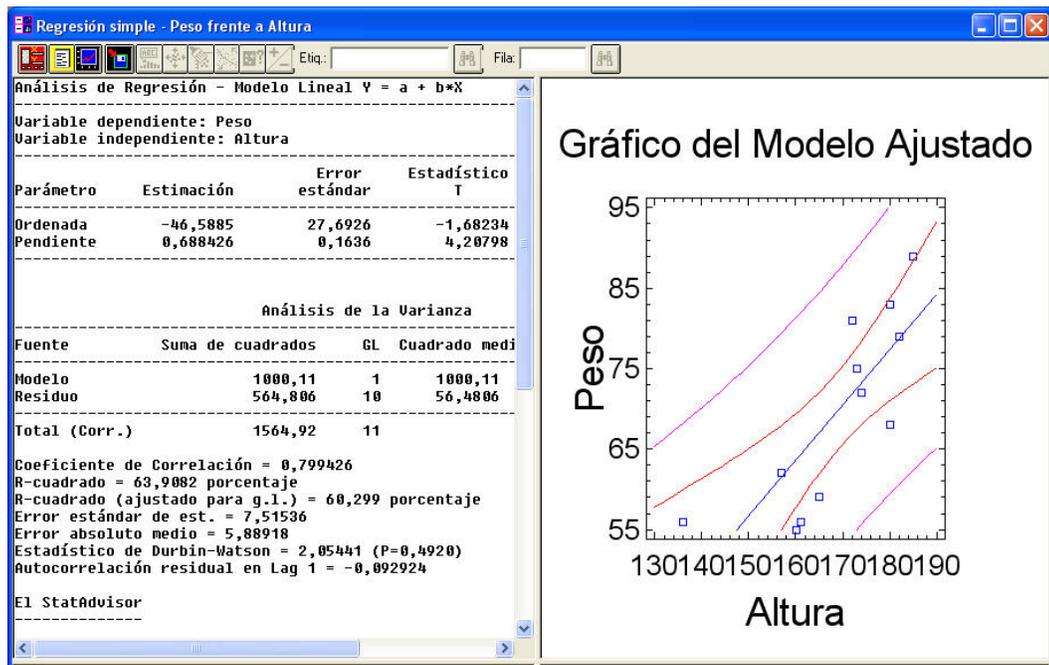


Figura 5.3: Ventana de resultados.

A la izquierda aparecen los siguientes resultados numéricos:

1. Valores estimados de los parámetros del modelo $a = -46,5885$ y $b = 0,688426$ (*Ordenada y Pendiente*), su error estándar y el siguiente contraste de hipótesis

5.2. Regresión lineal simple

sobre ellos:

$$H_0 : \text{parámetro} = 0$$

$$H_1 : \text{parámetro} \neq 0$$

si el p -valor del contraste es menor de 0.05, se rechaza la hipótesis nula (H_0) del contraste.

2. La tabla ANOVA de análisis de la varianza. En esta tabla aparecen la variabilidad explicada por nuestro modelo lineal de regresión (1000.11), la variabilidad que el modelo deja sin explicar (564.806) y la variabilidad total de la variable *Peso* que se quiere explicar (1564.92). También la tabla ANOVA resuelve el contraste de regresión:

$$H_0 : R^2 = 0$$

$$H_1 : R^2 \neq 0$$

Si el p -valor es menor de 0.05, se aceptará la hipótesis alternativa (H_1) de existencia de regresión.

3. El coeficiente de correlación $r = 0,799426$ y el coeficiente de determinación $R^2 = 0,639082$. El coeficiente de correlación lineal da una medida de la relación lineal que existe entre las variables. El coeficiente de determinación da una medida del ajuste de la regresión a los datos, siendo ésta mejor cuanto más cerca esté de 1.

Por lo tanto, la ecuación del modelo ajustado es $Peso = -46.5885 + 0.688426 \text{ Altura}$. En el icono **Opciones Tabulares** se obtienen las siguientes opciones.

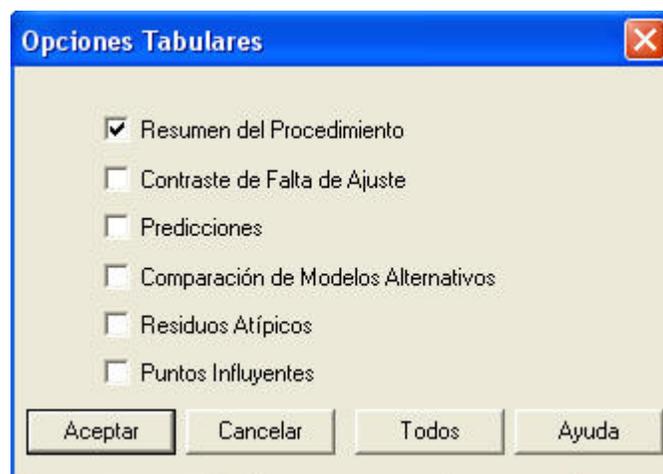


Figura 5.4: Opciones Tabulares.

5.2.2. Predicciones

Como se ha dicho, uno de los objetivos principales de la regresión es la de poder realizar predicciones sobre los datos. Statgraphics proporciona una opción que calcula dichas predicciones. Se genera una nueva ventana que contendrá una predicción y su intervalo de confianza para un valor seleccionado por el programa.

Valores predichos					
X	Predicho Y	95,00% Límites de Predicción		95,00% Límites de Confianza	
		Inferior	Superior	Inferior	Superior
136,0	47,0374	25,9118	68,163	34,1577	59,9171
185,0	80,7702	62,3621	99,1784	73,1247	88,4158

Figura 5.5: Predicciones por defecto.

En este caso realiza las predicciones para $x = 136$ y $x = 185$. Para modificar el valor para el que se realiza la predicción y el nivel al cual se hace el intervalo de confianza, se pulsa el botón derecho del ratón y se selecciona la opción **Opciones de Ventana**.

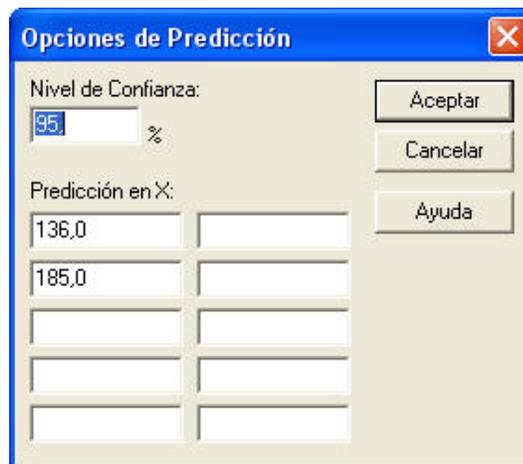


Figura 5.6: Opciones de predicción.

5.2.3. Comparación de Modelos Alternativos

La regresión que se ha realizado está basada en el modelo lineal. Ahora lo que se pretende es estudiar si existe otro modelo no lineal que mejore los resultados que se han obtenido con

5.3. Regresión no Lineal

el modelo lineal. La opción **Comparación de Modelos Alternativos** de las opciones tabulares proporciona los coeficientes de determinación (R^2) de una serie de modelos, ordenados de mayor a menor.

Comparación de Modelos Alternativos		
Modelo	Correlación	R-cuadrado
Inverso-Y	-0,8158	66,55%
Exponencial	0,8091	65,46%
Raiz cuadrada-Y	0,8046	64,74%
Lineal	0,7994	63,91%
Multiplicativo	0,7908	62,54%
Raiz cuadrada-X	0,7902	62,44%
Logarítmico-X	0,7803	60,89%
Doble inverso	0,7785	60,61%
curva-S	-0,7701	59,30%
Inverso-X	-0,7588	57,58%
Logístico		<sin ajuste>
Log Probit		<sin ajuste>

Figura 5.7: Comparación de modelos.

Observando los resultados se obtiene que el modelo para los datos es el modelo **Inverso-Y**, con un coeficiente de determinación de 0.6655.

5.2.4. Residuos Atípicos

Esta opción proporciona los valores de la variable que son anómalos o extremos. Estos puntos son aquellos que tienen un residuo estandarizado fuera del intervalo $[-2,2]$.

Residuos Atípicos					
Fila	X	Y	Predicha	Residuo	Residuo Estudentizado
3	136,0	56,0	47,0374	8,9626	2,19

Figura 5.8: Valores atípicos.

La fila 3 tiene valores anómalos o extremos ($x = 136$).

5.3. Regresión no Lineal

Para obtener un modelo de regresión simple de tipo no lineal, en la ventana de resultados inicial se pincha con el botón derecho del ratón y se elige la opción **Opciones de Análisis** que proporciona el siguiente *Cuadro de diálogo*, en el que se puede elegir el tipo de modelo

de regresión. La expresión de cada uno de estos modelos aparece en la primera línea de la ventana de resultados correspondiente.



Figura 5.9: Tipos de regresión.

5.4. Opciones Gráficas

La salida gráfica de la ventana inicial del análisis de regresión consiste en un gráfico donde están representadas la nube de puntos y la recta de regresión ajustada, así como otras curvas con los límites de confianza. Además del gráfico inicial de resultados **Gráfico del Modelo Ajustado**, existe la posibilidad de construir otros tipos de gráficos. El icono **Opciones Gráficas**,

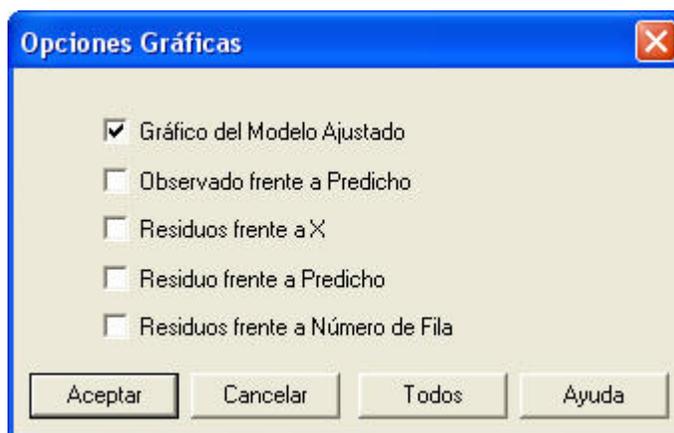


Figura 5.10: **Opciones Gráficas.**

tiene las siguientes opciones:

1. **Gráfico del Modelo Ajustado:** representa la nube de puntos y la recta de regresión y los límites de confianza de la recta. La aleatoriedad de los puntos indica que el modelo es bueno. Si se obtiene un gráfico con un patrón definido, indica que los residuos están relacionados y, por tanto, el modelo no es adecuado para los datos.
2. **Observado frente a Predicho:** es la gráfica de los valores observados frente a los predichos por el modelo ajustado incluyendo como referencia una línea con pendiente igual a 1. La aleatoriedad de los puntos indica que el modelo es bueno. Si se obtiene un gráfico con un patrón definido, indica que los residuos están relacionados y, por tanto, el modelo no es adecuado para los datos.
3. **Residuos frente a X:** residuos frente a la variable explicativa. La aleatoriedad de los puntos indica que el modelo es bueno. Si se obtiene un gráfico con un patrón definido, indica que los residuos están relacionados y, por tanto, el modelo no es adecuado para los datos.
4. **Residuos frente a Predichos:** residuos frente a la variable explicativa.
5. **Residuos frente a Número de Fila:** residuos frente al orden de aparición.

5.5. Ejercicios

- Una compañía eléctrica desarrollo un modelo de consumo de energía en función de la temperatura. Se obtuvo la siguiente información:

temperatura	0	2	4	8	13	-4	-6	-8	-11
consumo	70	79	67	66	63	97	82	90	107

- Determinar la nube de puntos.
 - Obtener la ecuación de regresión.
 - ¿Es bueno el ajuste?.
 - Obtener una predicción para el consumo de energía cuando la temperatura es 10.
- Se dispone de la información referente a 6 familias sobre el gasto en libros (Y) y la renta disponible mensual (X):

Y	0.3	0.5	0.6	0.9	1	1.4
X	6	7	8	10	15	21

Explicar el comportamiento de Y a través de X mediante una relación lineal. ¿Como de bueno es el ajuste?.

- En 1965 A. J. Lea recogió datos sobre la temperatura anual media en varias ciudades de Gran Bretaña, Noruega y Suecia y la tasa de mortalidad en un tipo de cáncer de pecho de mujeres. Los datos que obtuvo fueron los siguientes:

Temperatura anual media (°F)	Índice de mortalidad
51.3	102.5
49.9	104.5
50.0	100.4
49.2	95.9
48.5	87.0
47.8	95.0
47.3	88.6
45.1	89.2
46.3	78.9
42.1	84.6
44.2	81.7
43.5	72.2
42.3	65.1
40.2	68.1
31.8	67.3
34.0	52.5

5.5. Ejercicios

- a) Determinar la recta de regresión de mínimos cuadrados, que relaciona la variable índice de mortalidad en función de la temperatura anual media.
 - b) ¿Qué índice de mortalidad puede predecirse para una temperatura anual media de 33 °F?
 - c) ¿Es fiable esta predicción?
4. Una fábrica de cerveza quiere averiguar si existe una relación lineal entre el dinero que gasta en anuncios de televisión y sus ventas totales para lo que se anotan los siguientes datos:

Mes	Ventas (miles de euros)	Gastos en anuncios de TV (miles de euros)
Enero	50	0.5
Febrero	90	0.9
Marzo	30	0.4
Abril	90	0.7
Mayo	91	1.1
Junio	95	0.75
Julio	95	0.8

Obtener la recta de regresión y comprobar si existe relación lineal entre las variables.

5. La factura mensual del gasto telefónico de una pequeña empresa se ha incrementado notablemente en los últimos meses. Los estudios realizados por el administrador de la misma argumentan que el mayor uso de Internet dentro de la misma es la principal causa del mayor gasto en teléfono, lo que ha hecho que se estudie la posibilidad de acogerse a alguno de los múltiples bonos o tarifas especiales que ofrecen las compañías, lo que hasta la fecha no se había llevado a cabo. Las últimas cifras mensuales no hacen sino confirmar esta relación.

	Enero	Febrero	Marzo	Abril	Mayo
Cuantía de la factura telefónica (€)	55	100	118	120	142
Tiempo de conexión (minutos)	200	500	700	800	1000

- a) Suponiendo la existencia de una relación lineal entre el tiempo de conexión y gasto telefónico, obtener la recta de regresión.
- b) ¿Cuál sería la cuantía de la factura telefónica si no se conectase a Internet la empresa?
- c) ¿Cuál sería el gasto telefónico estimado según esta relación lineal si el tiempo de conexión a Internet fuera de 900 minutos?

- d)* Si se considera que un incremento del 20% en el tiempo de conexión a Internet respecto al realizado en el mes de mayo conllevaría a que la factura telefónica se elevase de forma extraordinaria, ¿cuál sería el incremento relativo en la misma si ello se produjese?.
6. Sobre el gasto en libros (Y) y la renta disponible mensual (X) se dispone de la información referente a 6 familias

Y	0.3	0.5	0.6	0.9	1	1.4
X	6	7	8	10	15	21

Explicar el comportamiento de Y por X mediante

- a)* una relación lineal.
b) una hipérbola equilátera.
c) una curva potencial.
d) ¿Qué ajuste es el más adecuado?

Práctica 6

Algunas distribuciones de probabilidad conocidas

6.1. Introducción

Statgraphics implementa hasta 25 distribuciones de probabilidad, calculando para todas ellas valores de la función de distribución, valores críticos (valores para los que la función de distribución tiene una probabilidad dada, es decir, puntos inversos de los valores de la función de distribución), secuencias de números aleatorios según una distribución dada y permite representar las correspondientes leyes de probabilidad, funciones de densidad y funciones de distribución.

La opción **Gráficos / Distribuciones de Probabilidad...** (Figura 6.1) del *Menú principal* de Statgraphics posibilita el trabajo con distribuciones de probabilidad.



Figura 6.1: Menú **Distribuciones de Probabilidad...**

El *Cuadro de diálogo* de entrada permite seleccionar la distribución que se desee, tal y como se indica en la Figura 6.2.

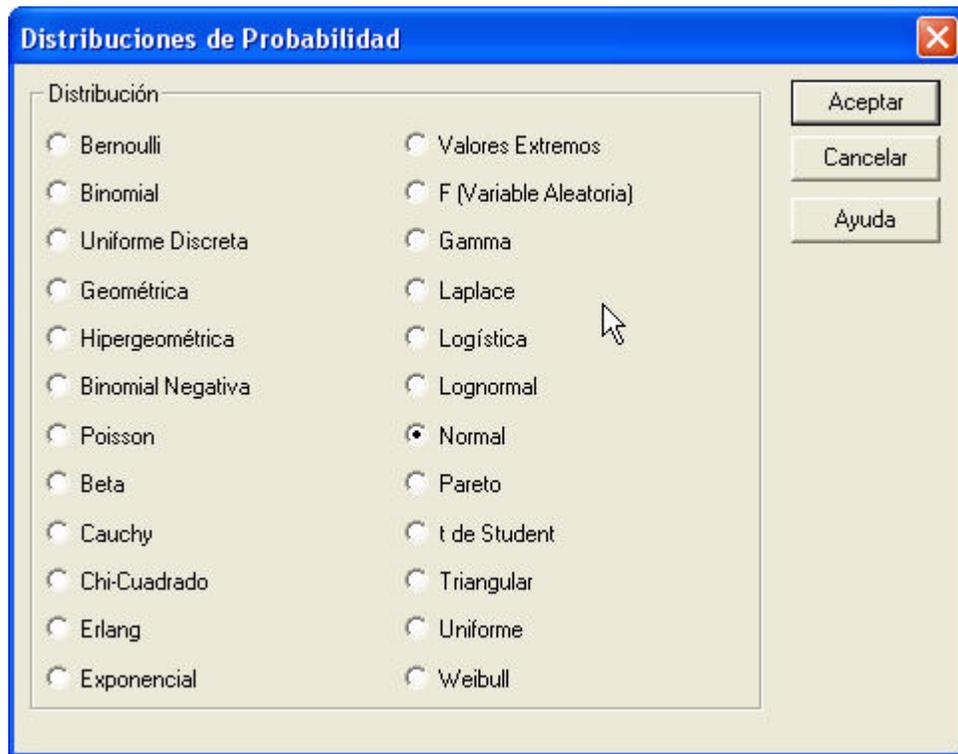


Figura 6.2: Distribuciones de Probabilidad.

Al pulsar **Aceptar** se obtiene una pantalla con información sobre la distribución elegida. Si sobre esta pantalla se pincha el icono **Opciones Tabulares** de la barra de herramientas, se obtiene el *Cuadro de diálogo* de la Figura 6.3, en la que se observan las distintas opciones de trabajo con la distribución elegida.



Figura 6.3: Opciones Tabulares.

6.2. Funciones de distribución

La opción **Distribución Acumulada** del *Cuadro de diálogo Opciones Tabulares* permite hallar el valor de la función de distribución en un punto dado para el modelo elegido. En esa pantalla se dan los valores para $F(x) = P(X < x)$, $f(x)$ (función de densidad) o $P[X = x]$ (función masa de probabilidad) y $1 - F(x) = P(X > x)$, hasta para 5 valores de parámetros distintos de la distribución en cuestión simultáneamente.

Si sobre la pantalla que se obtiene se pulsa el botón derecho del ratón y se elige la opción **Opciones de Ventana...** del menú emergente resultante, se obtiene el *cuadro de diálogo Opciones Distribución Acumulada*, en cuyo campo **Valores para la Variable:** se pueden introducir hasta 5 valores distintos de la variable para los que se halla la función de distribución, la función de densidad o función masa de probabilidad y el complementario de la función de distribución. (Ver Figura 6.4).

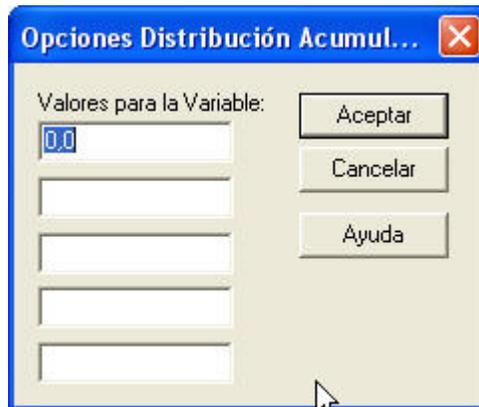


Figura 6.4: Opciones Distribución Acumulada.

Si, al pulsar el botón derecho del ratón, se elige la opción **Opciones de Análisis...** del menú emergente resultante se obtiene el *cuadro de diálogo Opciones...* de la distribución correspondiente, en cuyos campos se pueden introducir hasta 5 pares de valores distintos de los parámetros para las que se hallará los valores de la función de distribución, la función de densidad o la función masa de probabilidad y el complementario de la función de distribución.

6.3. Inversa de la función de distribución

La opción **CDF Inversa** del *Cuadro de diálogo Opciones Tabulares* permite hallar el valor de la función inversa de la función de distribución (valor crítico) en 5 puntos dados para el modelo elegido.

6.4. Opciones gráficas

Si sobre la pantalla que se obtiene se pincha con el botón derecho del ratón y se elige la opción **Opciones de Ventana...** del menú emergente resultante, se obtiene el *Cuadro de diálogo Opciones de la Función Inversa*, (Figura 6.5) en cuyo campo **CDF** se pueden introducir hasta 5 valores distintos para los que se halla la inversa de la función de distribución.



Figura 6.5: Opciones de la Función Inversa.

6.4. Opciones gráficas

Si sobre una salida cualquiera se elige el icono **Opciones Gráficas** de la barra de herramientas, se obtiene la opción **Densidad/Función Masa**, (Figura 6.6) entonces se representa la función de densidad (para variables aleatorias continuas) y la función masa de probabilidad (para variables aleatorias discretas) para el modelo de probabilidad elegido; y la opción **CDF**, representa la función de distribución para el modelo de probabilidad elegido.



Figura 6.6: Opciones Gráficas.

6.5. Ejemplos prácticos

1. Para este ejemplo se van a utilizar las distribuciones normales $\mathcal{N}(0,1)$ y $\mathcal{N}(5,10)$.

Para ello se accede al siguiente menú: **Descripción / Distribuciones / Distribuciones de Probabilidad...**

A continuación se selecciona la distribución con la que se va a trabajar, en este caso *Distribución Normal*.



Figura 6.7: Distribuciones de probabilidad.

Una vez que se ha introducido la distribución con la que trabajar, aparece una primera pantalla de resultados (Figura 6.8) donde se indica con que parámetro se está trabajando y la representación gráfica de esa distribución.

6.5. Ejemplos prácticos

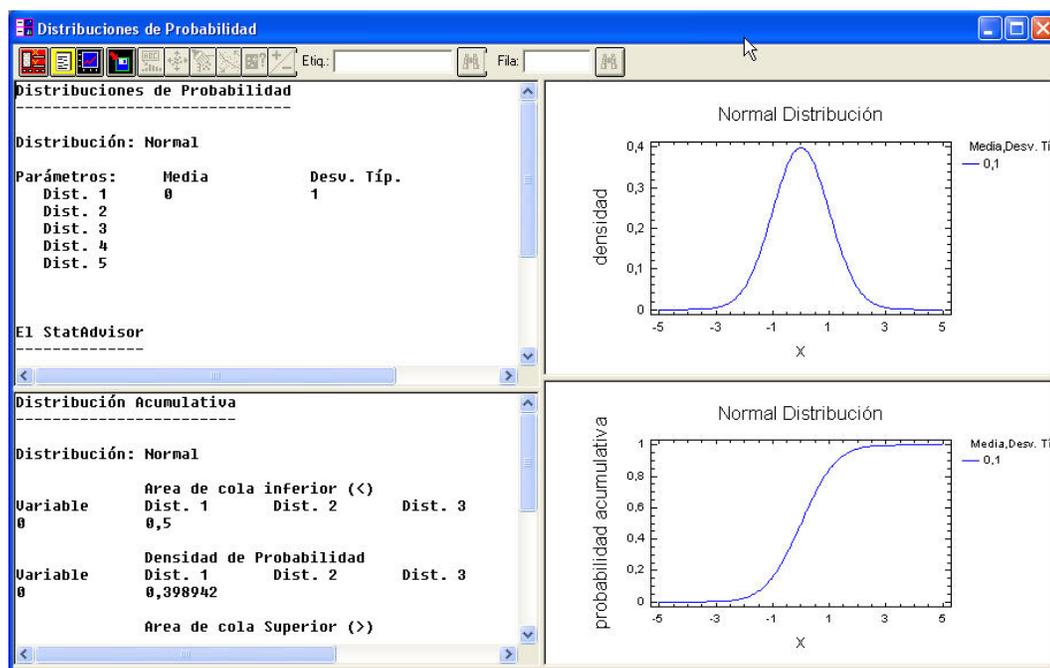


Figura 6.8: Ventana de resultados Distribuciones de Probabilidad.

Como se ha visto, el programa trabaja por defecto con la distribución seleccionada con unos parámetros aleatorios, para cambiar los parámetros con los que se trabaja o añadir más (puede trabajar hasta con 5 distribuciones con distintos parámetros) se pulsa el botón de la derecha del ratón y se selecciona la opción **Opciones de Análisis** y en esta ventana se introducen los parámetros elegidos.

The 'Opciones Normal' dialog box has two columns of input fields. The first column is labeled 'Media:' and contains five input boxes with values 0, 5, and three empty boxes. The second column is labeled 'Desv. Típ.:' and contains five input boxes with values 1, 10, and three empty boxes. On the right side, there are three buttons: 'Aceptar', 'Cancelar', and 'Ayuda'.

Figura 6.9: Selección de parámetros.

Cálculo de Probabilidades

Una vez seleccionada una distribución, para calcular probabilidades sobre ella, se selecciona en el icono **Opciones Tabulares**, la opción **Distribución Acumulati-**

va. Los resultados obtenidos se muestran a continuación:

Distribución Acumulativa					

Distribución: Normal					
	Area de cola inferior (<)				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,5	0,308536			
	Densidad de Probabilidad				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,398942	0,0352065			
	Area de cola Superior (>)				
Variable	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0	0,5	0,691464			

Figura 6.10: Algunas Probabilidades.

Es decir, la probabilidad de que la variable sea menor que 0 (0.5 en la primera distribución y 0.3085 en la segunda), el valor de la función de densidad en el cero (0.3989 y 0.035205 respectivamente) y la probabilidad de que la variable sea mayor que 0 (0.5 y 0.6914).

Para calcular alguna de estas probabilidades para otro valor que no fuese el cero, se pulsa en el botón de la derecha del ratón, la opción **Opciones de Ventana**. En la ventana resultante se introduciría el valor deseado.

Cálculo de Valores Críticos

En este caso, el objetivo es encontrar un valor que acumula una determinada probabilidad. Para realizarlo, una vez seleccionada la distribución, se pulsa el botón de **Opciones Tabulares y Inversa CDF**. Una vez realizado esto, aparece una ventana con la siguiente tabla:

Inversa CDF					

Distribución: Normal					
CDF	Dist. 1	Dist. 2	Dist. 3	Dist. 4	Dist. 5
0,01	-2,32635	-18,2635			
0,1	-1,28155	-7,81554			
0,5	0	5			
0,9	1,28155	17,8155			
0,99	2,32635	28,2635			

Figura 6.11: Valores que acumulan una determinada probabilidad a la izquierda.

Para cambiar estos valores se pulsa el botón derecho del ratón y la opción **Opciones de Ventana**. En esta ventana se sustituye alguno de los valores que aparecen por los que se quieran calcular.



Figura 6.12: Opciones Inversa CDF

Opciones Gráficas

Para obtener resultados gráficos dentro del menú de probabilidad se pulsa el icono **Opciones Gráficas**, apareciendo la siguiente ventana:



Figura 6.13: Opciones Gráficas.

En esta ventana hay las siguientes opciones:

- a)* **Densidad/Función Masa:** función de densidad de la variable si es continua o la función masa de probabilidad si es discreta.
- b)* **CDF:** función de distribución.
- c)* **Función de Supervivencia:** función que representa la probabilidad de obtener un valor mayor o igual que un valor dado.
- d)* **Función Log Supervivencia:** logaritmo de la función anterior.
- e)* **Función de Riesgo:** se obtiene dividiendo la función de densidad por la función de supervivencia y se utiliza para modelizar datos sobre tiempo de vida.

6.5. Ejemplos prácticos

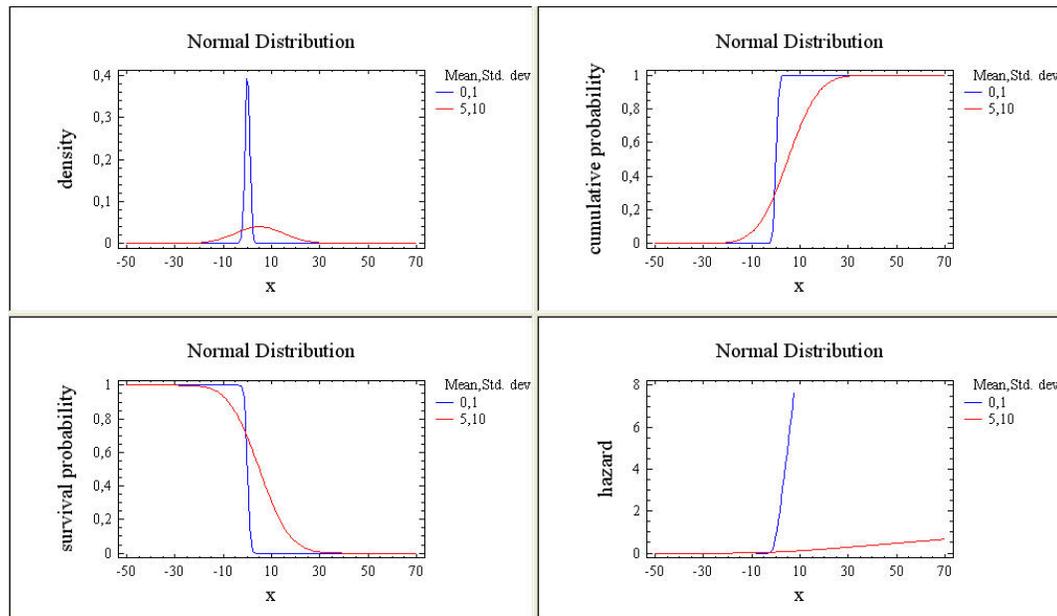


Figura 6.14: Gráficos de densidad, CDF, Supervivencia y Riesgo.

2. *El número medio de automóviles que llega a una estación de suministro de gasolina es de 210 por hora. Si dicha estación puede atender a un máximo de 10 automóviles por minutos, determinar la probabilidad de que en un minuto dado lleguen a la estación de suministro más automóviles de los puede atender.*

El número aleatorio de automóviles que llegan a la estación de servicio en un minutos puede representarse por una variable X de Poisson de parámetro $\lambda = 210/60 = 3,5$ llegadas/minuto.

La probabilidad que hay que calcular es $P[X > 10]$, ya que para que lleguen a la estación más automóviles por minuto de los que se puedan atender es necesario que lleguen más de 10 por minuto.

Para calcular dicha probabilidad se elige el menú **Gráficos / Distribuciones de Probabilidad...** y en su pantalla de entrada se selecciona la opción Poisson. Se pulsa **Aceptar** y sobre la pantalla resultante se pulsa el botón derecho del ratón y se elige la opción **Opciones de Análisis** para obtener el cuadro de diálogo **Opciones Poisson**, (ver Figura 6.15) en cuyo campo **Media** se introduce el valor 3.5. Al pulsar **Aceptar** ya se tiene la distribución $\mathcal{P}(3,5)$ como distribución de trabajo.

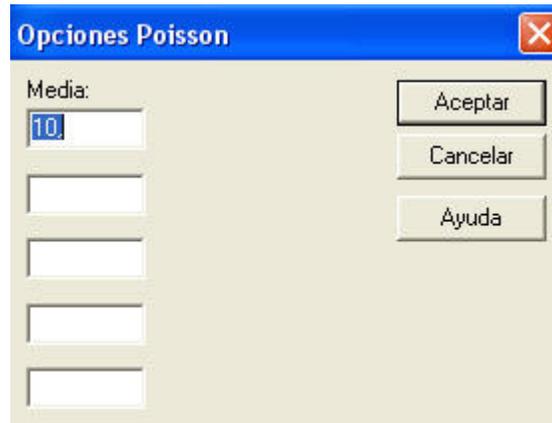


Figura 6.15: Opciones Poisson.

Para calcular la probabilidad pedida se selecciona la opción **Opciones Tabulares** de la barra de herramientas y en el menú resultante la opción **Distribución Acumulada**. Se pulsa **Aceptar** y a continuación el botón derecho del ratón sobre cualquier zona de la pantalla resultante y se elige la opción **Opciones de Ventana...** del menú emergente obtenido. Así se llega a la caja de diálogo **Opciones Distribución Acumulada**, (ver Figura 6.16 cuyo campo **Valores para la Variable** se rellena con el valor 10.

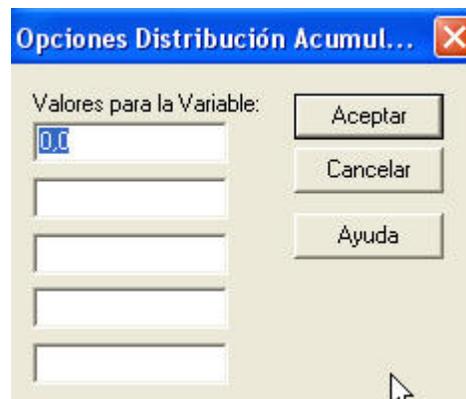


Figura 6.16: Opciones Distribución Acumulada.

Al pulsar **Aceptar** se obtiene la pantalla que se presenta a continuación, en cuya última fila, se observa que $P[X > 10] = 0,00101939$.

6.5. Ejemplos prácticos

```
Distribución: Poisson
Variable      Area de cola inferior (<)
10           Dist. 1      Dist. 2      Dist. 3
            0,996685
Variable      Probabilidad de Masa (=)
10           Dist. 1      Dist. 2      Dist. 3
            0,00229555
Variable      Area de cola Superior (>)
10           Dist. 1      Dist. 2      Dist. 3
            0,00101939
```

Figura 6.17: Resultado.

6.6. Ejercicios

1. En una biblioteca se hace un estudio de su actividad y el estado de sus fondos
 - a) Se observa que se realizan unos 7 préstamos cada media hora.
 - Calcular la probabilidad de hacer 6 préstamos en la próxima media hora.
 - Calcular la probabilidad de hacer más de 3 préstamos en la próxima media hora.
 - Calcular la probabilidad de hacer entre 5 y 8 préstamos en la próxima media hora.
 - Calcular la probabilidad de hacer 15 préstamos en la próxima media hora.
 - Calcular la probabilidad de hacer menos de 13 préstamos en la próxima media hora.
 - b) La probabilidad de que un libro devuelto esté dañado es 0.3.
 - Calcular la probabilidad de que haya 10 libros dañados entre los 50 que se han devuelto hoy.
 - Calcular la probabilidad de que haya entre 10 y 15 libros dañados entre los 50 que se han devuelto hoy.
 - Calcular la probabilidad de que haya mas de 7 libros dañados entre los 50 que se han devuelto hoy.
 - Calcular la probabilidad de que haya menos de 14 libros dañados entre los 75 que se han devuelto otro día.
 - c) Las horas que la biblioteca está abierta a la semana sigue una distribución Normal de media 45 horas y desviación típica 2.5 horas.
 - Calcular la probabilidad de que la biblioteca esté abierta entre 43 y 46 horas.
 - Calcular la probabilidad de que la biblioteca esté abierta más de 45 horas.
 - Calcular la probabilidad de que la biblioteca esté abierta menos de 41 horas.
 - Calcular la probabilidad de que la biblioteca esté abierta entre 44 y 45 horas.
2. En un año, el número medio de accidentes en una curva es 4. Determine la probabilidad de que en seis meses:
 - a) no ocurra ningún accidente
 - b) ocurran exactamente 2
 - c) menos de 5

6.6. Ejercicios

- d) más de 5
3. Un canal de comunicación recibe impulsos independientes a razón de 12 impulsos por microsegundo. La probabilidad de un error de transmisión es de 0.001 para cada impulso. Calcular las probabilidades de los siguientes sucesos.
- a) No hay ningún error en un microsegundo.
b) Hay exactamente un error en un microsegundo.
c) Hay al menos un error en un microsegundo.
d) Hay exactamente dos errores en un microsegundo.
4. En una facultad universitaria el Coeficiente Intelectual tiene una media de 108 y una desviación típica de 7 unidades. Se pide calcular los valores del Coeficiente de Inteligencia tales que
- a) Por debajo de él se encuentre el 5 % de los individuos.
b) Por encima se sitúen el 25 % de los alumnos.
c) Por debajo estén el 80 % de los estudiantes.
5. Dada una distribución χ^2 ,
- a) Representar en un mismo gráfico las funciones de distribución de una χ^2 con 10, 20, 30, 40 y 50 grados de libertad
- b) Si $X \rightsquigarrow \chi_{12}^2$, calcular $P[X \leq 32,1]$, $P[X \geq 20]$ y $P[13,8 \leq X \leq 22,4]$.
- c) Calcular a , b y c tales que:
 $P[X \leq a] = 0,85$ con $X \rightsquigarrow \chi_{20}^2$
 $P[X \geq b] = 0,975$ con $X \rightsquigarrow \chi_{18}^2$
 $P[X \leq c] = 0,05$ con $X \rightsquigarrow \chi_{10}^2$

Práctica 7

Estimación

7.1. Introducción

Esta práctica desarrolla la obtención de intervalos de confianza para una y dos muestras. Los intervalos de confianza que proporciona el programa se basan exclusivamente en poblaciones normales. Por lo tanto, un paso previo antes de calcular los intervalos de confianza es contrastar la existencia de normalidad en los datos, desde la opción **Gráficos / Gráficos Exploratorios / Gráfico Probabilístico...** del *Menú principal*. En el gráfico resultante se concluye que los datos proceden de una distribución normal si éstos se encuentran sobre la bisectriz.

7.2. Intervalos de confianza para una muestra

Cuando se utilizan directamente los datos de una única muestra de una población, Statgraphics permite el trabajo con estimación por intervalos en poblaciones normales mediante la opción **Descripción / Datos Numéricos / Análisis Unidimensional...** del *Menú principal* del programa.

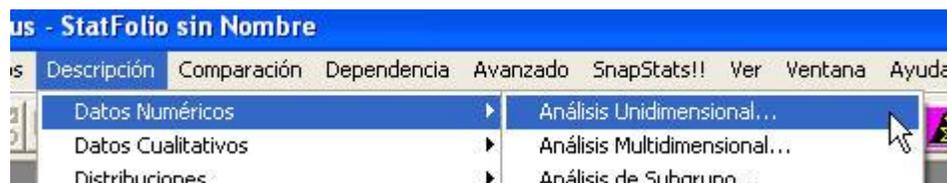


Figura 7.1: Menú **Análisis unidimensional...**

Al pulsar el icono **Opciones Tabulares** (Figura 7.2) aparece la opción **Intervalos de Confianza** que permite hallar intervalos de confianza para la media y la desviación

7.2. Intervalos de confianza para una muestra

típica de una población normal con σ desconocida y de la cual se supone que la variable dada es una muestra.



Figura 7.2: Opciones Tabulares.

Si sobre la ventana de resultados obtenida se pulsa el botón derecho del ratón y se elige la opción **Opciones de Ventana...** del menú emergente resultante, se obtiene el *Cuadro de diálogo* **Opciones de Intervalos de Confianza** de la Figura 7.3, cuyo campo **Nivel de Confianza** permite elegir el nivel de confianza.

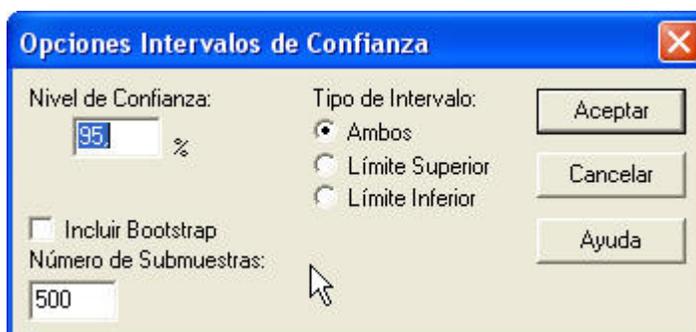


Figura 7.3: Nivel de confianza

7.2.1. Ejemplo práctico

Los siguientes datos corresponden a las calificaciones obtenidas por 25 alumnos e dos pruebas consecutivas de una misma asignatura. (Se supone que las dos variables siguen una distribución normal).

prueba1	prueba2
5.27	3.61
6.20	5.29
5.97	4.55
6.84	3.08
5.78	4.71
7.62	2.58
4.87	3.96
6.88	5.69
7.34	7.32
6.34	5.36
6.30	2.71
7.16	4.78
4.71	4.49
6.52	4.85
6.49	5.70
5.68	0.65
6.32	4.66
6.61	2.68
4.56	2.91
5.09	2.46
6.23	6.81
4.06	3.82
6.53	3.02
5.75	3.73
4.71	8.05

Utilizando el 95% de confianza obtenga un intervalo de confianza para la media y la desviación típica de cada una de las variables.

Como siempre el primer paso será la introducción de los datos en el editor, para luego elegir el menú apropiado para el análisis. En este caso será: **Descripción / Datos Numéricos / Análisis Unidimensional...**

En la ventana que aparece, se introduce en la casilla **Datos:** la variable, en este caso *prueba1*.

7.2. Intervalos de confianza para una muestra



Figura 7.4: Entrada de datos.

El programa genera la misma ventana de resultados que se ha estudiado con anterioridad. Para obtener los intervalos de confianza se pulsa el icono de **Opciones Tabulares** y se selecciona **Intervalos de Confianza**



Figura 7.5: Opciones Tabulares.

Los intervalos al 95 % para la media y la desviación típica de la variable *prueba1*, en este caso, son:

Intervalos de Confianza para prueba1

95,0% intervalo de confianza para la media: 5,9932 +/- 0,384521 [5,60868;6,37772]
 95,0% intervalo de confianza para la desviación típica: [0,727372;1,29591]

Figura 7.6: Intervalos de confianza para la variable *prueba1*.

El programa calcula el intervalo de confianza para la desviación típica, este intervalo tiene la siguiente expresión:

$$\left[\sqrt{\frac{(n-1)s^2}{\chi_{n-1,\alpha/2}^2}}; \sqrt{\frac{(n-1)s^2}{\chi_{n-1,1-\alpha/2}^2}} \right]$$

Por lo tanto, para obtener los intervalos para la varianza, se elevan los extremos del intervalo al cuadrado, resultando el intervalo de confianza al 95 % para la varianza: [0,529;1,679].

Análogamente para la variable *prueba2* se obtienen los siguientes intervalos al 95 % para la media y la desviación típica:

Intervalos de Confianza para prueba2

95,0% intervalo de confianza para la media: 4,2988 +/- 0,69191 [3,60689;4,99071]
 95,0% intervalo de confianza para la desviación típica: [1,30884;2,33187]

Figura 7.7: Intervalos de confianza para la variable *prueba2*.

Y el intervalo para la varianza es: [1.713;5.437].

7.3. Intervalos de confianza para dos muestras

Cuando se utilizan directamente los datos de dos muestras de dos poblaciones diferentes, Statgraphics permite el trabajo con estimación por intervalos para comparar medias y varianzas en poblaciones normales mediante la opción **Comparación / Dos Muestras / Comparación de Dos Muestras...** (ver Figura 7.8) del *Menú principal* del programa.

Figura 7.8: Menú **Comparación de Dos Muestras...**

7.3. Intervalos de confianza para dos muestras

La pantalla de entrada se presenta en la Figura 7.9. En el campo **Muestra 1:** de esta pantalla se introduce el nombre de la variable que contiene los datos de la primera muestra. En el campo **Muestra 2:** se introduce el nombre de la variable que contiene los datos de la segunda muestra. Statgraphics calcula un intervalo de confianza para la diferencia de medias de la Muestra 1 - Muestra 2, y para el cociente de las varianzas de las muestras. Si se pulsa el botón **Aceptar** se obtiene el resumen de los datos con el número de valores de las dos muestras y sus dominios de definición.



Figura 7.9: Comparación para dos muestras.

Como en todo análisis con Statgraphics, en la barra de herramientas del análisis se puede elegir el icono **Opciones Tabulares** que permite las opciones **Comparación de Medias** y **Comparación de Desviaciones Típicas**.

La opción tabular **Comparación de Medias** ofrece por defecto los intervalos de confianza al 95 por ciento para las medias de las dos poblaciones (supuestas normales) basados en las dos muestras. También ofrece los intervalos de confianza para la diferencia de medias suponiendo igual varianza para ambas poblaciones normales.

La opción tabular **Comparación de Desviaciones Típicas** ofrece por defecto los intervalos de confianza al 95 por ciento para las desviaciones típicas de las dos poblaciones (supuestas normales) basados en las dos muestras. También ofrece el intervalo de confianza para el cociente de varianzas.



Figura 7.10: Opciones Tabulares.

7.3.1. Ejemplo práctico

Con los datos del ejemplo anterior, ¿puede considerarse que las puntuaciones medias sean iguales en las dos pruebas?

Para responder a la pregunta hay que obtener un intervalo de confianza para la diferencia de medias de las dos variables. Dado que no se tiene ninguna información sobre las varianzas poblaciones, en primer lugar, habrá que obtener un intervalo de confianza para el cociente de varianzas y determinar si éstas se pueden considerar iguales.

Para ello, se selecciona el menú **Comparación / Dos Muestras / Comparación de Dos Muestras...**. Se introducen las dos variables como se ha indicado, *prueba1* en el campo **Muestra 1:** y *prueba2* en el campo **Muestra 2:**. (Ver Figura 7.11)

7.3. Intervalos de confianza para dos muestras

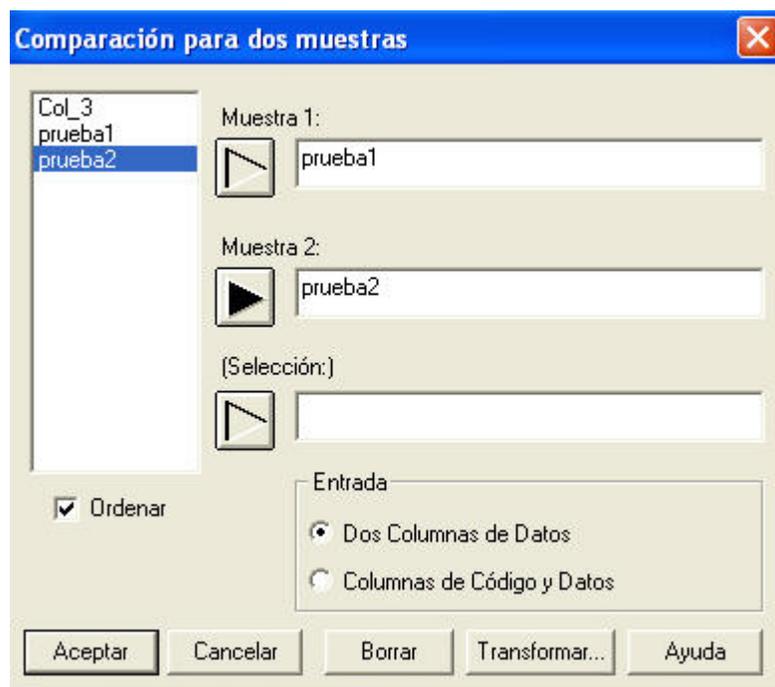


Figura 7.11: Entrada de datos.

Se pulsa **Aceptar** y se obtienen algunos resultados por defecto para este análisis. A continuación se pincha el icono de **Opciones Tabulares** y se seleccionan las opciones **Comparación de Medias** y **Comparación de Desviaciones Típicas**.

En la ventana de Comparación de Desviaciones Típicas se obtienen los siguientes resultados (ver Figura 7.12).

Comparación de Desviaciones Típicas

	prueba1	prueba2
Desviación Típica	0,931539	1,67622
Varianza	0,867764	2,8097
GL	24	24

Cociente de varianzas = 0,308846

95,0% Intervalos de Confianza

Desviación Típica de prueba1: [0,727372;1,29591]

Desviación Típica de prueba2: [1,30884;2,33187]

Cociente de varianzas: [0,136099;0,700856]

Figura 7.12: Comparación de Desviaciones Típicas.

El intervalo que se obtiene para el cociente de varianzas: $[0,136099;0,700856]$ no contiene el valor 1, luego no se puede considerar que las varianzas sean iguales. Es decir, en este caso no se puede llegar a una conclusión final sobre la diferencia de medias, ya que el programa sólo resuelve el intervalo de confianza para diferencia de medias, cuando las varianzas se suponen iguales.

En el caso de haber obtenido que las varianzas fueran iguales, se observaría ahora el resultado obtenido en la ventana Comparación de Medias.

Comparación de Medias

95,0% intervalo de confianza para la media de prueba1: 5,9932 +/- 0,384521 [5,60868,6,37772]
 95,0% intervalo de confianza para la media de prueba2: 4,2988 +/- 0,69191 [3,60689,4,99071]
 95,0% intervalos de confianza para la diferencia de medias:
 suponiendo varianzas iguales: 1,6944 +/- 0,771149 [0,923251,2,46555]

Figura 7.13: Comparación de Medias.

En cuyo caso, con el intervalo $[0,923251;2,46555]$ obtenido suponiendo varianzas iguales, se llega a la conclusión que la media de la *prueba2* es superior.

7.4. Ejercicios

1. Se ha extraído una muestra de 20 paquetes cuyos peso en gramos son:

520, 503, 538, 447, 506, 436, 478, 463, 525, 494
521, 524, 515, 550, 457, 520, 428, 519, 491, 548

Obtener la estimación del peso medio de la población mediante un intervalo de confianza a un nivel del 80 % y del 95 %.

2. Se estudian los pesos de dos grupos de niños recién nacidos según sus madres fumen o no. Los pesos son:

Fumadora	2.3 3.2 3.1 2.5 3.2 2.6 2.3 2.9 2.2 2.6 2.4 3 2.1 3 2.6 2.1 2.6 2.4 2.7 3.6
No fumadora	3.1 2.7 3.6 3.8 3.1 2.9 2.7 3.4 3.1 2.9 2.7 3.4 3.1 2.9 4.5 3.6 3.7 3.8 4.2

Suponiendo que ambos grupos provienen de poblaciones normales de medias desconocidas y con varianzas desconocidas pero iguales, determinar si influye que la madre sea fumadora en el peso del hijo.

3. Los tiempos de ejecución (en segundos) de 40 trabajos procesados por un centro de cálculo han resultado ser:

10 19 90 40 15 11 32 17 4 152
23 13 36 101 2 14 2 23 34 15
27 1 57 17 3 30 50 4 62 48
9 11 20 13 38 54 46 12 5 26

Calcular la media y la cuasidesviación típica muestrales. Obtener intervalos de confianza al 90 % para la media y la desviación típica del tiempo de ejecución de un trabajo, suponiendo que esta variable aleatoria tiene distribución normal.

4. Se quiere comparar dos métodos, A y B para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos (en calorías por gramo de masa para pasar de -0.72°C a 0°C) usando reiteradamente ambos métodos:

Método A	79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02
Método B	80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

Se supone que en ambos métodos el calor tiene una distribución normal. Obtener un intervalo de confianza al 95 % para comparar las mediciones medias obtenidas por ambos métodos. Comprobar primero si se puede suponer que las varianzas son iguales.

5. Se ha hecho un estudio para comparar los tiempos de acceso, en diferentes momentos del día, a internet desde ordenadores domésticos con modem. Para ello se cargan 8 páginas web por la tarde en el período de 14 a 15 h. y, con el mismo ordenador, las mismas 8 páginas por la noche en el periodo de 22 a 23 h. Los respectivos tiempos de acceso en minutos fueron los siguientes:

De 22 a 23 h.	2.9	1.4	1.2	3.4	1.3	2.5	1.6	1.8
De 14 a 15 h.	2.3	1.5	1	2.7	1.4	1.9	0.8	1.1

¿Puede demostrarse que el acceso es más lento en el horario nocturno?