



STATPOINT
TECHNOLOGIES, INC.

STATGRAPHICS[®] Centurion XVI

Manual de usuario

STATGRAPHICS® CENTURION XVI

MANUAL DE USUARIO

© 2010 StatPoint Technologies, Inc.
www.STATGRAPHICS.com

Todos los derechos reservados. Ninguna parte de este documento puede ser reproducida, de ninguna forma y por ningún medio, sin el consentimiento expreso y por escrito de StatPoint Technologies, Inc.

Referencia a: *STATGRAPHICS® Centurion XVI User Manual*

STATGRAPHICS es una marca registrada. STATGRAPHICS Centurion XVI, StatPoint, StatFolio, StatGallery, StatReporter, StatPublish, StatWizard, StatLink, y SnapStats son marcas registradas. Todos los productos o servicios mencionados en este libro son marcas registradas o marcas de servicio de sus respectivos propietarios.

Impreso en los Estados Unidos de América.

Tabla de Contenidos

Tabla de Contenidos	iii
Introducción	vii
Comenzando.....	1
1.1 Instalación.....	1
1.2 Ejecutando el programa.....	8
1.3 Introduciendo datos	14
1.4 Leyendo un archivo de datos guardado	18
1.5 Analizando los datos	20
1.6 Utilizando la barra de herramientas de análisis	24
1.7 Difundiendo los resultados	29
1.8 Guardando su trabajo	30
Administración de datos	33
2.1 El libro de datos.....	34
2.2 Accediendo a los datos	36
2.2.1 Leyendo datos de una archivo de datos de STATGRAPHICS Centurion.....	36
2.2.2 Leyendo datos de un archivo Excel, ASCII, XML, u otro archivo externo de datos ...	38
2.2.3 Transfiriendo datos utilizando Copiar y Pegar.....	39
2.2.4 Consultando una base de datos ODBC	40
2.3 Manipulando Datos.....	41
2.3.1 Copiando y pegando datos	41
2.3.2 Creando nuevas variables de columnas existentes.....	41
2.3.3 Transformando datos.....	45
2.3.4 Ordenando datos	48
2.3.5 Recodificando datos	50
2.3.6 Combinando múltiples columnas.....	51
2.4 Generando datos	53
2.4.1 Generando datos con patrones.....	54
2.4.2 Generando números aleatorios.....	56
2.5 Propiedades del libro de datos.....	57
2.6 Visor de datos	59
Ejecutando análisis estadísticos	61
3.1 Cuadros de diálogo de entrada de datos.....	63
3.2 Ventana de análisis	65
3.2.1 Botón Entrada de diálogo	67
3.2.2 Botón Opciones de análisis.....	67

3.2.3 Botón Tablas y Gráficos.....	68
3.2.4 Botón Opciones de panel.....	71
3.2.5 Botón Guardar resultados	73
3.2.6 Botones gráficos	74
3.2.7 Botón Excluir.....	75
3.3 Imprimiendo los resultados	76
3.4 Publicando los resultados	78
Gráficos.....	79
4.1 Modificando gráficos	80
4.1.1 Opciones de trazado	81
4.1.2 Opciones de rejilla (malla).....	83
4.1.3 Opciones de líneas.....	85
4.1.4 Opciones de puntos	87
4.1.5 Opciones del título superior.....	89
4.1.6 Opciones de escalado de ejes	91
4.1.7 Opciones de relleno.....	93
4.1.8 Texto, Opciones de etiquetas y leyendas.....	94
4.1.9 Añadiendo texto nuevo	94
4.2 Separando un gráfico de dispersión.....	96
4.3 Resaltando un gráfico de dispersión	98
4.4 Alisando un gráfico de dispersión (tendencia)	100
4.5 Identificando Puntos.....	101
4.6 Copiando gráficos a otras aplicaciones.....	105
4.7 Guardando gráficos en archivos de Imagen.....	105
StatFolios	107
5.1 Guardando su sesión.....	107
5.2 Código de StatFolio.....	109
5.3 Apilando orígenes de datos	112
5.4 Publicando datos en formato HTML.....	113
Utilizando StatGallery	117
6.1 Configurando una página de StatGallery	117
6.2 Copiando gráficos a StatGallery	119
6.3 Solapando gráficos.....	120
6.4 Modificando un gráfico en StatGallery	121
6.4.1 Añadiendo ítems	121
6.4.2 Modificando ítems.....	122
6.4.3 Borrando ítems.....	122
6.5 Imprimiendo StatGallery	123
Utilizando StatReporter.....	125

7.1 La ventana StatReporter	125
7.2 Copiando Salidas a StatReporter	126
7.3 Modificando la salida de StatReporter.....	127
7.4 Guardando StatReporter	127
Utilizando StatWizard	129
8.1 Accediendo a Datos o Creando un nuevo estudio	130
8.2 Seleccionando análisis para sus datos	134
8.3 Buscando los estadísticos o test deseados.....	139
Preferencias del sistema	143
9.1 Comportamiento general del sistema	143
9.2 Imprimiendo.....	146
9.3 Gráficos.....	146
Tutorial #1: Analizando una muestra simple.....	149
10.1 Ejecutando el procedimiento Análisis de una variable	150
10.2 Resúmenes estadísticos.....	153
10.3 Gráfico de caja y bigotes	156
10.4 Contrastando valores atípicos.....	158
10.5 Histograma	162
10.6 Gráficos de cuantiles y percentiles.....	167
10.7 Intervalos de confianza.....	168
10.8 Test de hipótesis	170
10.9 Límites de tolerancia	172
Tutorial #2: Comparando dos muestras	175
11.1 Ejecutando el procedimiento Comparación de dos muestras	175
11.2 Estadísticos resumen.....	177
11.3 Histograma dual.....	178
11.4 Gráfico dual de caja y bigotes.....	179
11.5 Comparando desviaciones típicas	181
11.6 Comparando medias	182
11.7 Comparando medianas	183
11.8 Gráfico de cuantiles	184
11.9 Test de Kolmogorov-Smirnov para dos muestras.....	185
11.10 Gráfico cuantil-cuantil.....	186
Tutorial #3: Comparando más de dos muestras.....	187
12.1 Ejecutando el procedimiento comparación de varias muestras.....	188
12.2 Análisis de la varianza	192
12.3 Comparando medias	194
12.4 Comparando Medianas.....	196
12.5 Comparando desviaciones típicas	198

12.6 Gráficos de los residuos	198
12.7 Análisis de gráficos de medias (ANOM)	200
Tutorial #4: análisis de la regresión.....	201
13.1 Análisis de la correlación	202
13.2 Regresión simple.....	206
13.3 Ajustando un modelo no lineal	209
13.4 Examinado los residuos.....	211
13.5 Regresión múltiple.....	213
Tutorial #5: Analizando datos de atributos	221
14.1 Resumiendo datos de atributos	222
14.2 Análisis de Pareto	223
14.3 Tabulación cruzada	226
14.4 Comparando dos o más muestras.....	233
14.5 Tablas de contingencia.....	236
Tutorial #6: Análisis de la capacidad de un proceso	239
15.1 Graficando los datos	240
15.2 Procedimiento Análisis de la capacidad	242
15.3 Tratando con datos no normales	245
15.4 Índices de capacidad	252
15.5 Calculadora Seis Sigma	255
Tutorial #7: Diseño de experimentos (DOE).....	257
16.1 Creando un diseño	258
Paso 1: Definir respuestas.....	259
Paso 2: Definir factores experimentales.....	260
Paso 3: Seleccionar diseño	261
Paso 4: Especificar el modelo.....	268
Paso 5: Seleccionando ejecuciones	269
Paso 6: Evaluar el diseño	269
Paso 7: Guardar experimento.....	271
16.2 Analizando los resultados.....	272
Paso 8: Analizar datos.....	272
Paso 9: Optimizar respuestas	284
Paso 10: Guardar resultados	287
16.3 Experimentación posterior	287
Paso 11: Aumentar diseño	288
Paso 12: Extrapolar.....	289
Lecturas sugeridas.....	291
Conjuntos de datos.....	292
Índice.....	293

Introducción

Este libro está diseñado para introducir a los usuarios en STATGRAPHICS Centurion XVI en lo referente a las operaciones básicas del programa y su utilización en el análisis de datos. Aporta una introducción comprensiva al uso del sistema, incluyendo instalación, manejo de datos, creación de análisis estadísticos e impresión y publicación de resultados. Ya que el libro está concebido para introducir a los usuarios rápidamente, concentra las características más importantes del programa, en lugar de intentar la cobertura con todo detalle. El menú Ayuda de STATGRAPHICS Centurion XVI da acceso a gran cantidad de información adicional, incluyendo archivos PDF para cada uno de los aproximadamente 160 procedimientos estadísticos.

Los primeros nueve capítulos de este libro cubren el uso básico del programa. Aunque probablemente deberá utilizar otro material adicional mientras usa el programa, la lectura de estos capítulos le ayudará a introducirse rápidamente y le asegurará no fallar en las características más importantes.

Los últimos siete capítulos incluyen tutoriales enfocados a:

1. Introducirle en algunos de los análisis estadísticos más comúnmente utilizados.
2. Ilustrar cómo las características únicas de STATGRAPHICS Centurion XVI facilitan el proceso de análisis de datos.

Es recomendable que explore los tutoriales, ya que aportarán una buena idea de cómo STATGRAPHICS Centurion XVI puede utilizarse de la mejor forma cuando se analizan datos reales.

NOTA: una copia de este manual en formato PDF se incluye con el programa y puede accederse a ella mediante el menú *Ayuda*. En el documento PDF todos los gráficos están en color. Los archivos de datos y StatFolios referenciados en el manual se aportan también con el programa.

StatPoint Technologies, Inc.
Agosto de 2009

Comenzando

Instalando STATGRAPHICS Centurion XVI, ejecutando el programa y creando un archivo de datos simple.

1.1 Instalación

STATGRAPHICS Centurion XVI se distribuye de dos formas: a través de Internet en un solo archivo que se descarga a su ordenador, y como un conjunto de archivos en un CD-ROM. Para ejecutar el programa, debe ser instalado previamente en el disco duro. Como en la mayoría de los programas Windows, la instalación es extremadamente simple:

Paso 1: si recibe el programa en un CD, inserte el CD en su unidad de CD-ROM. Después de unos momentos, el programa de instalación comenzará automáticamente. Si no es así, abra el Explorador de Windows y ejecute el archivo **sgcinstall.exe** en el directorio raíz del CD-ROM.

Si ha descargado el programa de Internet, localice el archivo, haga doble clic sobre él y comenzará el proceso de instalación.

Paso 2: Se mostrarán sucesivos cuadros de diálogo. Si usted está ejecutando el programa desde un CD, el primer cuadro de diálogo le pregunta por la especificación del idioma o idiomas que van a ser instalados:

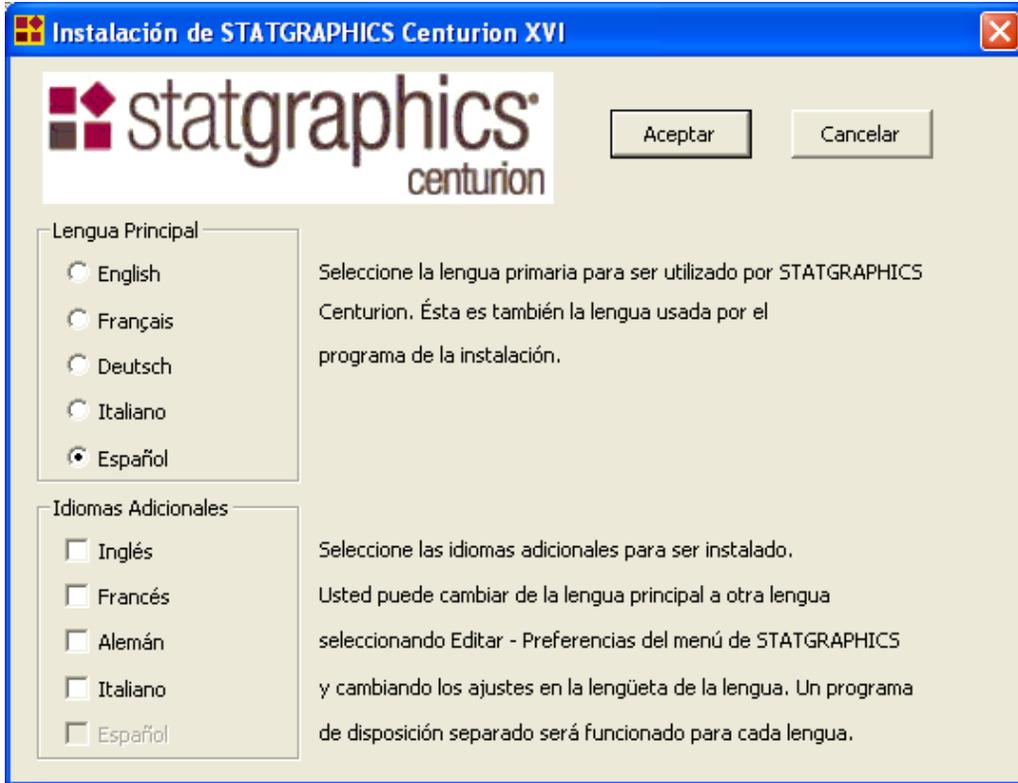


Figura 1-1. Cuadro de diálogo de selección de idioma

Seleccione un idioma principal y uno o más idiomas adicionales. El idioma principal será utilizado durante la instalación y también como el idioma por defecto cuando el programa se ejecute por primera vez. Si instala idiomas adicionales, puede cambiar entre ellos desde el programa seleccionado *Editar – Preferencias* en el menú principal.

Si ha descargado el programa desde Internet, necesitará ejecutar un programa de instalación por separado para cada idioma que descargue.

NOTA: Durante el periodo de evaluación los usuarios pueden acceder a cualquiera de los idiomas disponibles en STATGRAPHICS Centurion XVI. En la compra se le pedirá que defina el idioma principal y los adicionales (si hay alguno). Por favor, es necesario tener en cuenta que sólo los idiomas especificados estarán disponibles para su uso en STATGRAPHICS Centurion XVI.

Paso 3: STATGRAPHICS Centurion XVI utiliza InstallShield para instalar el programa en su ordenador. El asistente InstallShield controla la instalación por medio de una serie de cuadros de diálogo. El primer cuadro de diálogo da la bienvenida a STATGRAPHICS Centurion XVI:

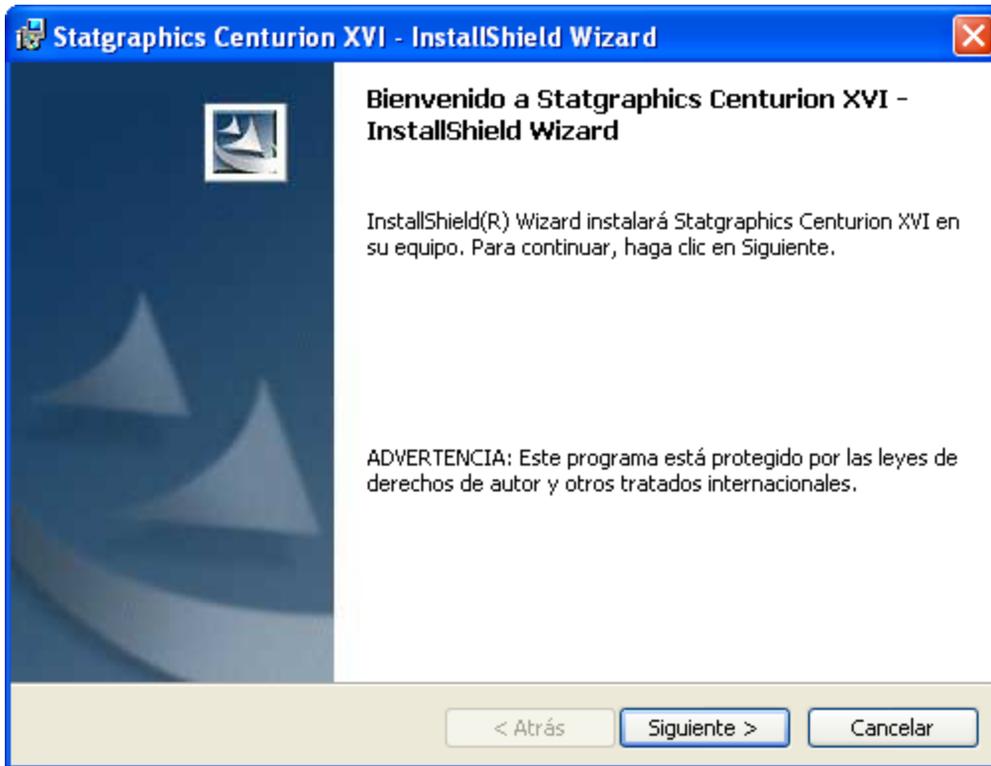


Figura 1-2. Cuadro de diálogo de bienvenida

Presione el botón *Siguiente*.

NOTA: En orden a instalar y activar STATGRAPHICS Centurion XVI debe tener derechos de administrador en su ordenador. Por si lo necesita, el administrador del sistema debe estar presente durante el proceso de instalación. Recomendamos encarecidamente que el administrador esté presente durante el proceso de instalación y activación del software.

Paso 4: El segundo cuadro de diálogo muestra el contrato de licencia del software:

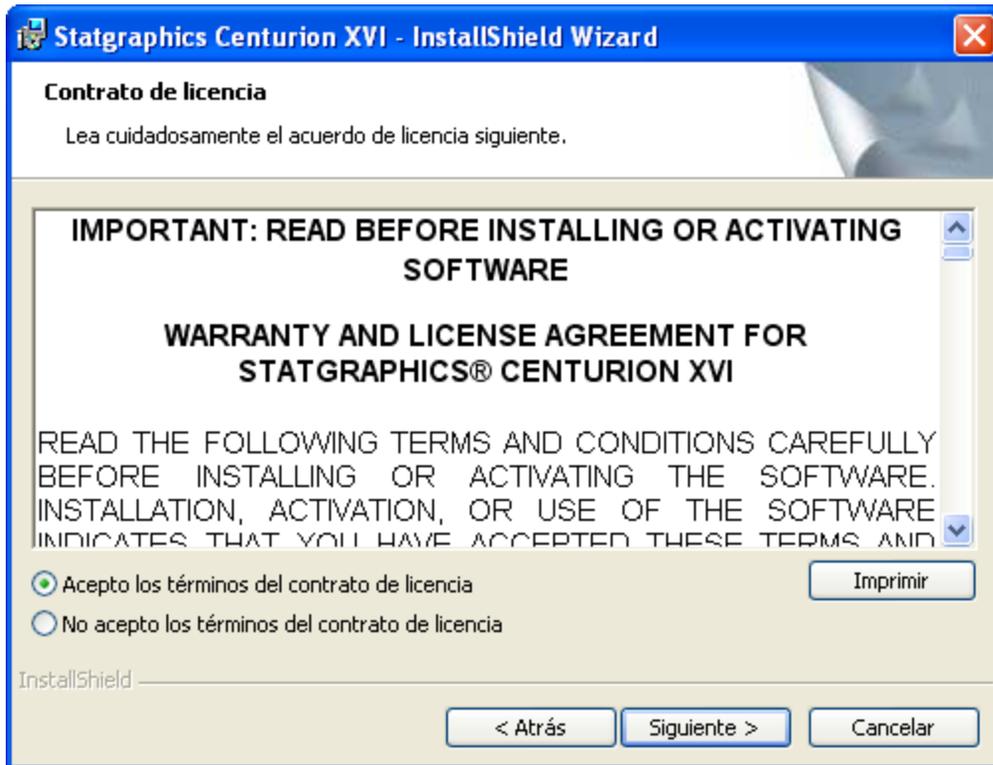


Figura 1-3. Cuadro de diálogo del contrato de licencia

Lea cuidadosamente el contrato de licencia. Si acepta los términos, haga clic en el botón indicado y presione *Siguiete* para continuar. Si no está de acuerdo, presione *Cancelar*. Si no está de acuerdo con los términos del contrato, no debe utilizar el programa.

Paso 5: El siguiente cuadro de diálogo requiere información acerca de la persona que utilizará el programa:

Statgraphics Centurion XVI - InstallShield Wizard

Información del cliente

Por favor, introduzca la información.

Nombre de uusuario:
Your name

Organización:
Your company

Instalar esta aplicación para:

Cualquiera que utilice este equipo (todos los usuarios)

Sólo para mí (Test Platform X64)

InstallShield

< Atrás Siguiete > Cancelar

Figura 1-4. Cuadro de diálogo de información del cliente

Introduzca la información requerida. Si quiere permitir a alguien que utilice el ordenador para tener acceso a STATGRAPHICS Centurion XVI, seleccione el botón apropiado.

Paso 6: El siguiente cuadro de diálogo indica el directorio en el que se instalará el programa:

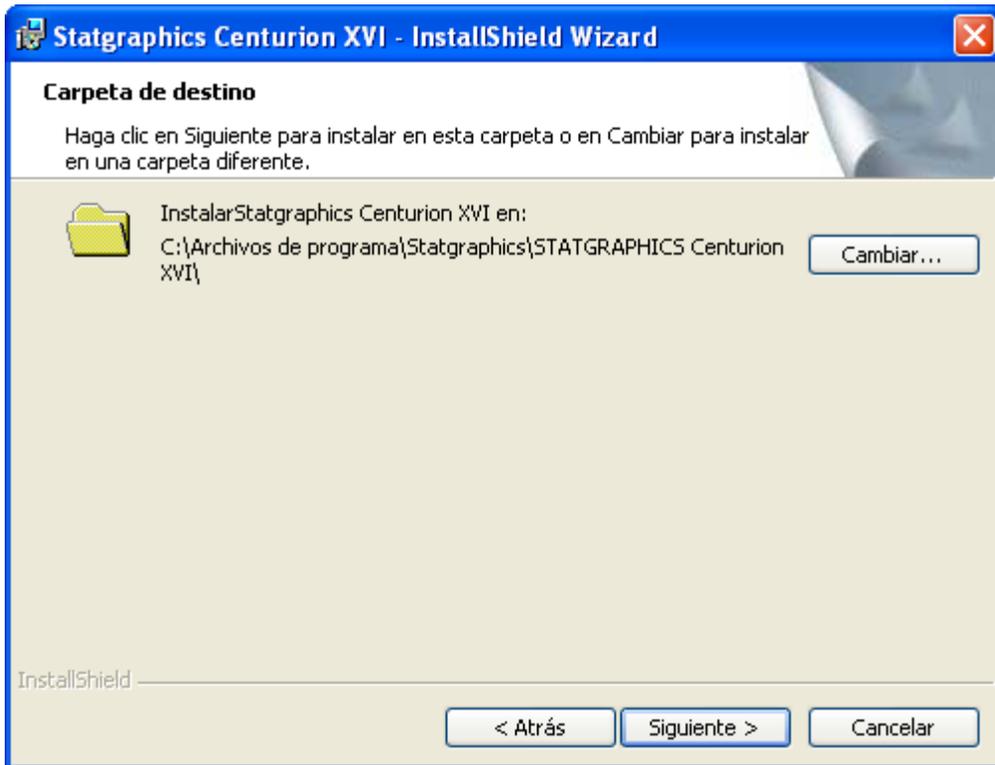


Figura1-5. Cuadro de diálogo de carpeta de destino

Por defecto, STATGRAPHICS Centurion XVI se instala en el subdirectorio de *Archivos de programa* denominado *STATGRAPHICS Centurion XVI*. Si está instalando el programa en un servidor de red, instálelo en una localización en la que los potenciales usuarios tengan acceso a la red. No se requiere acceso de escritura para los usuarios. Consulte el archivo *léame.txt* del CD de STATGRAPHICS Centurion XVI o descargue el archivo para los detalles de instalación de red.

Paso 7: El siguiente cuadro de diálogo permite especificar el tipo de instalación que se va ejecutar:

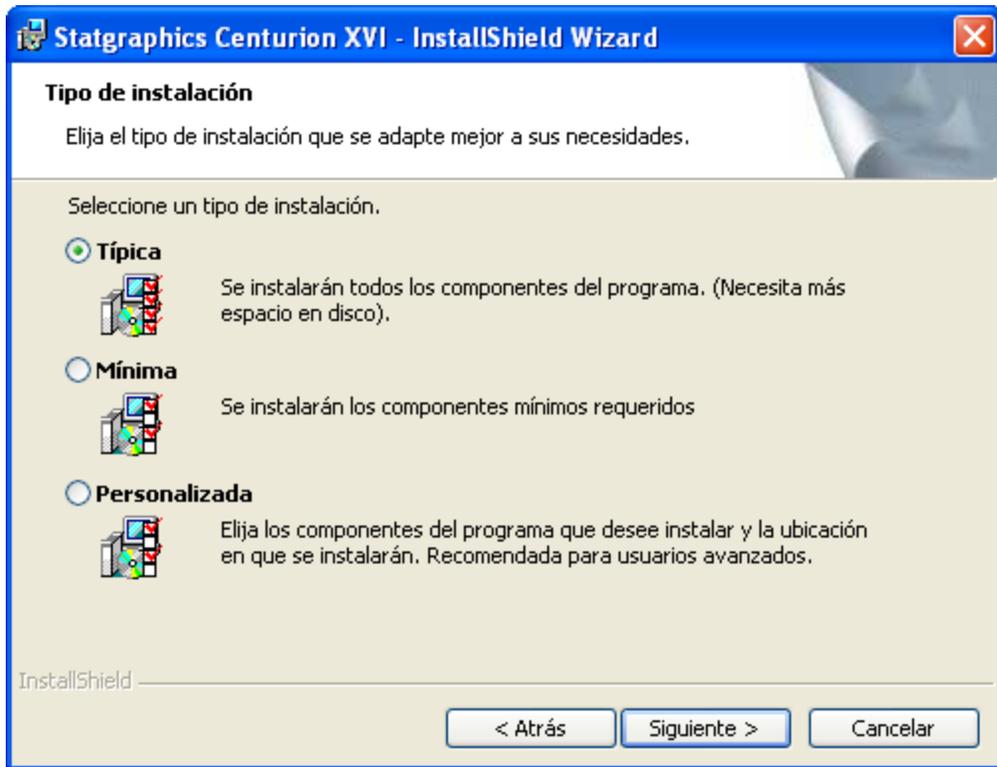


Figura1-6. Cuadro de diálogo Tipo de instalación

Seleccione una de las opciones siguientes:

Típica– instala el programa, archivos de ayuda, documentación y los archivos de datos. Requiere como mínimo 60MB de espacio en su disco duro.

Mínima – instala sólo el programa y los archivos de ayuda. Requiere como mínimo 30MB de espacio en su disco duro.

A medida – instala sólo los componentes que seleccione.

Puede ahorrar espacio en disco ejecutando una instalación mínima, sin embargo no tendrá acceso a la documentación en línea ni a los archivos de datos.

Paso 8: Continúe las instrucciones restantes para ejecutar la instalación. Cuando se termina la instalación, se mostrará el cuadro de diálogo final:

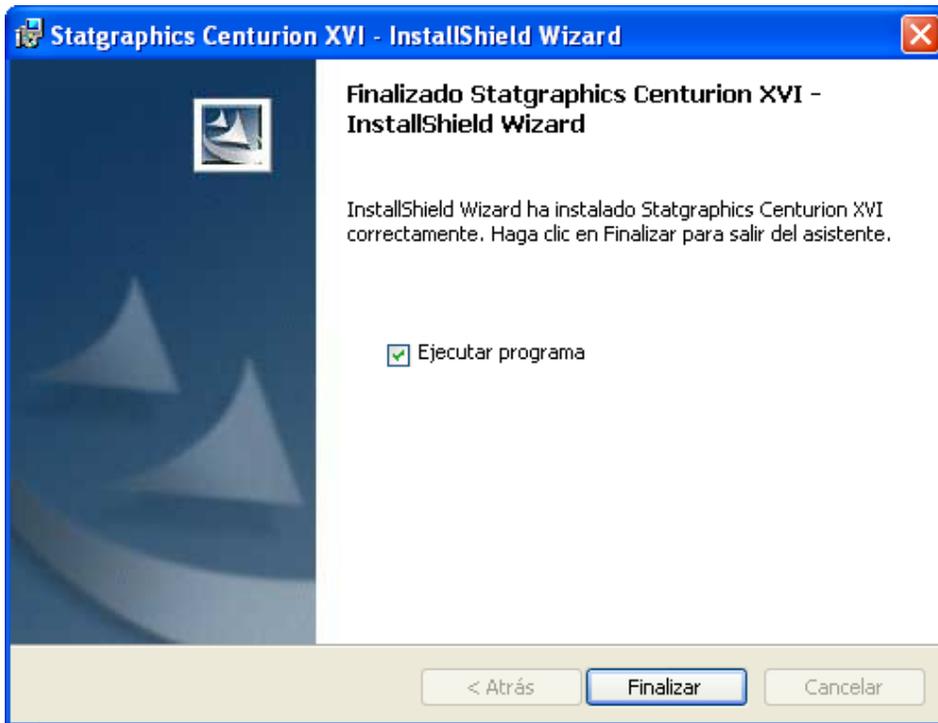


Figura 1-7. Cuadro de diálogo Final de instalación

Haga clic en *Finalizar* para completar la instalación. Marque el botón *Acceder al programa* si quiere iniciar STATGRAPHICS Centurion XVI inmediatamente, o continúe con las instrucciones.

1.2 Ejecutando el programa

Como parte del proceso de instalación, se añadirá un acceso directo a STATGRAPHICS Centurion XVI en el menú *Inicio* de Windows y también en su escritorio. Para iniciar el programa:

Paso 1: haga clic en el acceso directo del escritorio, presione el botón Inicio de Windows situado en la esquina inferior izquierda de su pantalla y haga clic en el icono de *Statgraphics*. También puede seleccionar *Archivos de programa – Statgraphics - STATGRAPHICS Centurion XVI* utilizando el explorador de Windows y haciendo clic en el icono de la aplicación *sgwin* para ejecutar el programa.

Paso 2: Cuando STATGRAPHICS Centurion XVI se inicia, se abrirá una nueva ventana. La primera vez que se ejecuta el programa, se muestra el cuadro de diálogo de *Bienvenida*:

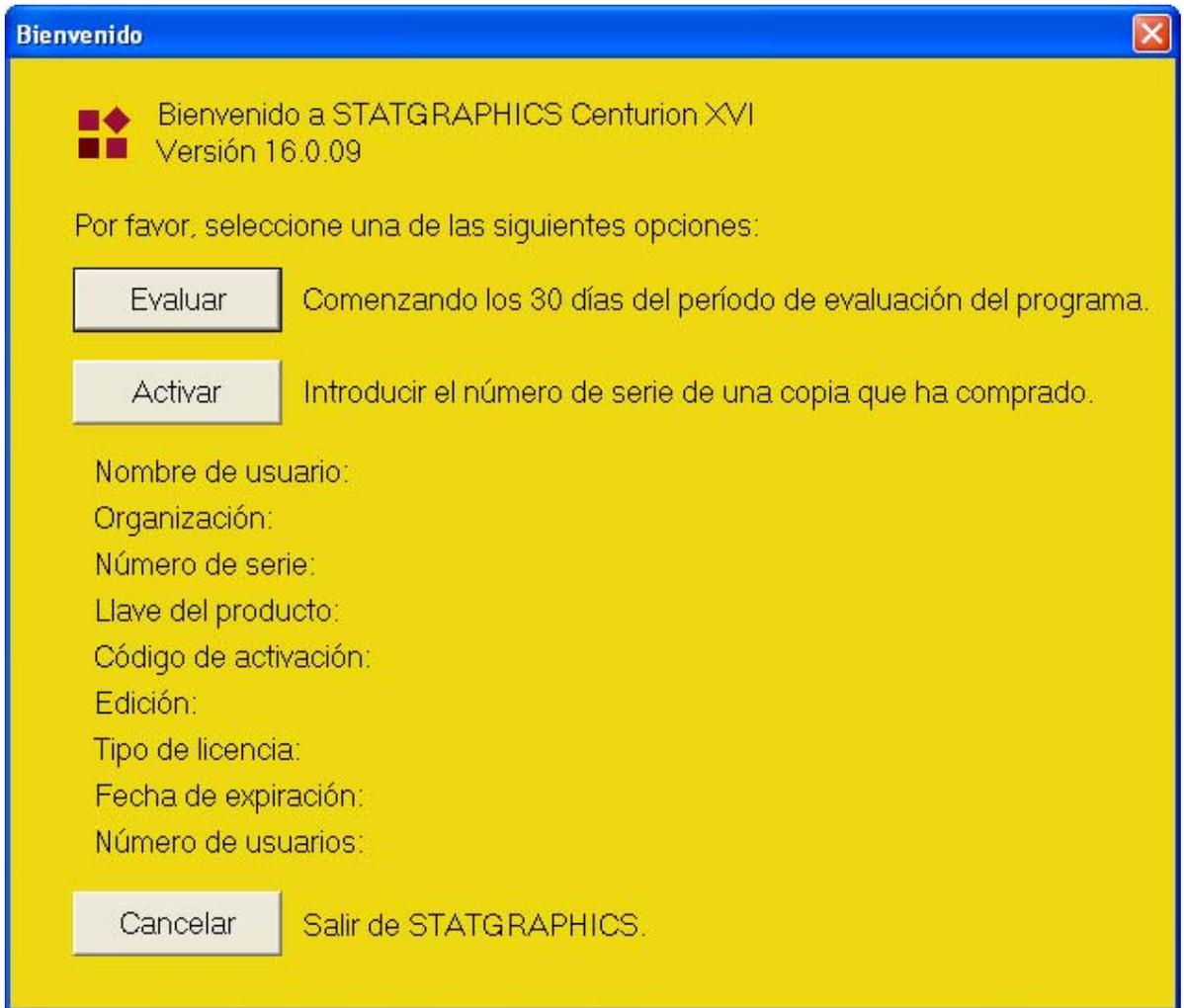


Figura1-8. Cuadro de diálogo de *Bienvenida*

Tiene dos posibilidades:

1. Para comenzar con una versión de entrenamiento de 30 días antes de comprar el programa, pulse el botón *Evaluación*.

2. Si ha comprado ya el programa y ha recibido un número de serie, presione el botón *Activar*.

Si ha pulsado el botón *Evaluar*, aparecerá el siguiente cuadro de diálogo:

Período de evaluación y activación

StatPoint Technologies, Inc.

Paso 1: Para comenzar sus 30 días de período de pruebas, primero introduzca la siguiente información (* requerido):

*Apellido: *Nombre:

*Organización:

*E-Mail:

Dirección:

Dirección:

Ciudad: Estado:

*País: Código Postal:

*Teléfono: Fax:

Llave del producto: Requerido:

Paso 2: Utilice uno de los dos caminos siguientes para requerir el código de activación:

1. Presione aquí Para presentar la solicitud por medio de Internet (recomendado - respuesta automática).

2. Presione aquí Para requerir un código de activación via E-mail (requiere respuesta manual).

Paso 3: Cuando reciba su código de activación por e-mail, introdúzcalo abajo y presione el botón Activar:

Figura 1-9. Cuadro de diálogo *Activación del período de evaluación*

El cuadro de diálogo muestra una *Clave de producto* de 16 caracteres que es única para su ordenador. Para comenzar el periodo de evaluación, debe introducir un *Código de activación* correcto. Para recibir un código de activación, presione uno de los dos botones siguientes:

1. El botón etiquetado *1. Presione aquí* automáticamente envía un mensaje a StatPoint Technologies mediante Internet solicitando un Código de activación. Un servicio Web responderá inmediatamente a este requerimiento, enviando el Código de activación a la dirección de e-mail aportada en la solicitud.
2. El botón etiquetado *2. Presione aquí* accede a un programa de e-mail por defecto, situando la información en un nuevo e-mail que puede enviar a StatPoint. El email requerido se procesará durante el horario normal de trabajo.

Para evitar retrasos, utilice el primer método siempre que sea posible.

Paso 3: Una vez procesado su requerimiento, recibirá un e-mail conteniendo el Código de activación. Introduzca el código en el campo provisto en el *paso 3* y presione el botón *Activar*. Si el código concuerda con la clave de producto, verá el mensaje siguiente:



Figura 1-10. Mensaje de activación

Presione *Aceptar* para acceder a la sección principal del programa.

NOTA #1: Si está ejecutando Microsoft Vista, cuando haga doble clic en el icono STATGRAPHICS en el escritorio para iniciar el programa puede que no tenga éxito. Si es éste el caso, debe hacer clic en el icono del programa con el botón derecho del ratón y elegir la opción *Ejecutar como administrador* en la lista de opciones que aparecen.

NOTA #2: Si posteriormente instala STATGRAPHICS Centurion XVI en un ordenador diferente tendrá que repetir el proceso de obtención del código de activación, ya que la clave de producto es única para cada máquina.

Paso 4: La primera vez que ejecute el programa, se le preguntará por el tipo de menú del sistema que quiere utilizar:



Figura 1-11. Ventana Selección de menú

Puede elegir entre el menú clásico de STATGRAPHICS Centurion XVI, que organiza los procedimientos estadísticos en los encabezamientos *Gráficos, Describir, Comparar, Relación, Predicción, SPC, y DOE*, o el menú Seis Sigma, que organiza los procedimientos en los encabezamientos *Definir, Medir, Analizar, Mejorar, Controlar y Predicción*. Ambos menús incluyen los mismos procedimientos. Sólo es diferente la organización. Puede cambiar su elección inicial posteriormente seleccionando *Preferencias* en el menú *Editar* del programa.

Paso 5: Se obtendrá la ventana principal de STATGRAPHICS Centurion XVI:

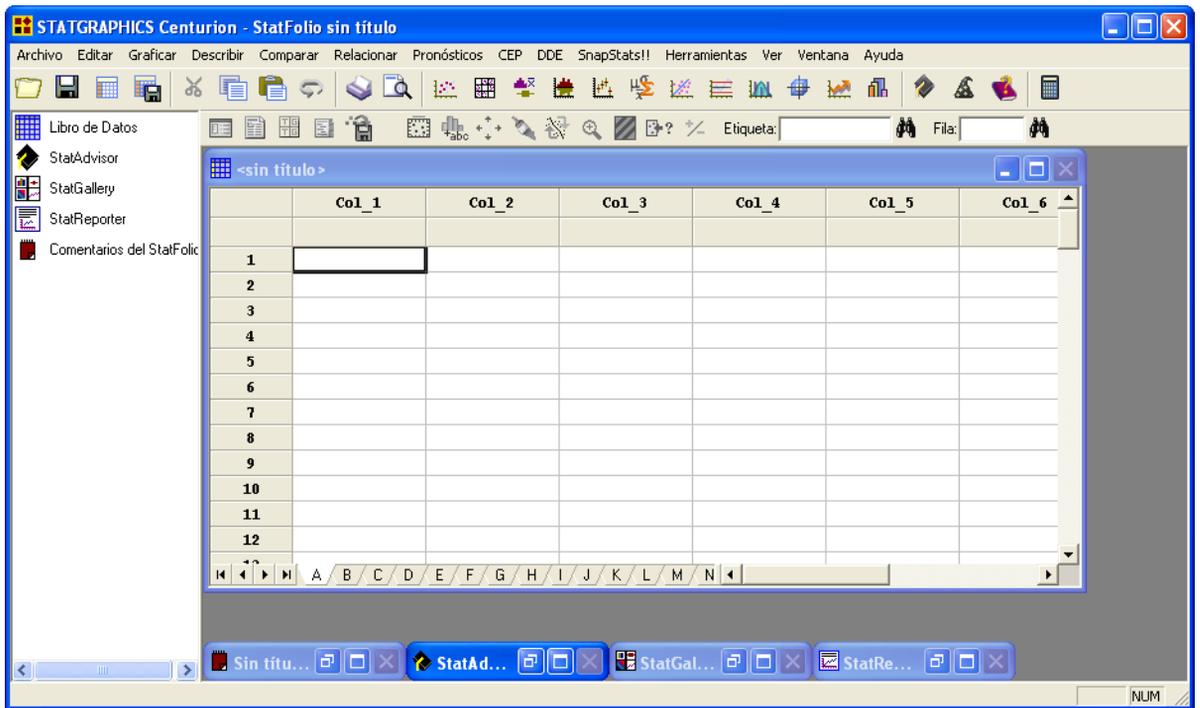


Figura 1-12. Ventana principal de STATGRAPHICS

Las secciones que siguen ilustran cómo se crean archivos de datos conteniendo información del Censo de Estados Unidos de 2000.

1.3 Introduciendo datos

En orden a analizar datos en STATGRAPHICS Centurion XVI, es necesario situarse en el *libro de datos* de STATGRAPHICS. El libro de datos está formado por hasta 26 hojas, indicadas por las letras desde la A a la Z, que contienen un cuadro rectangular de filas y columnas:

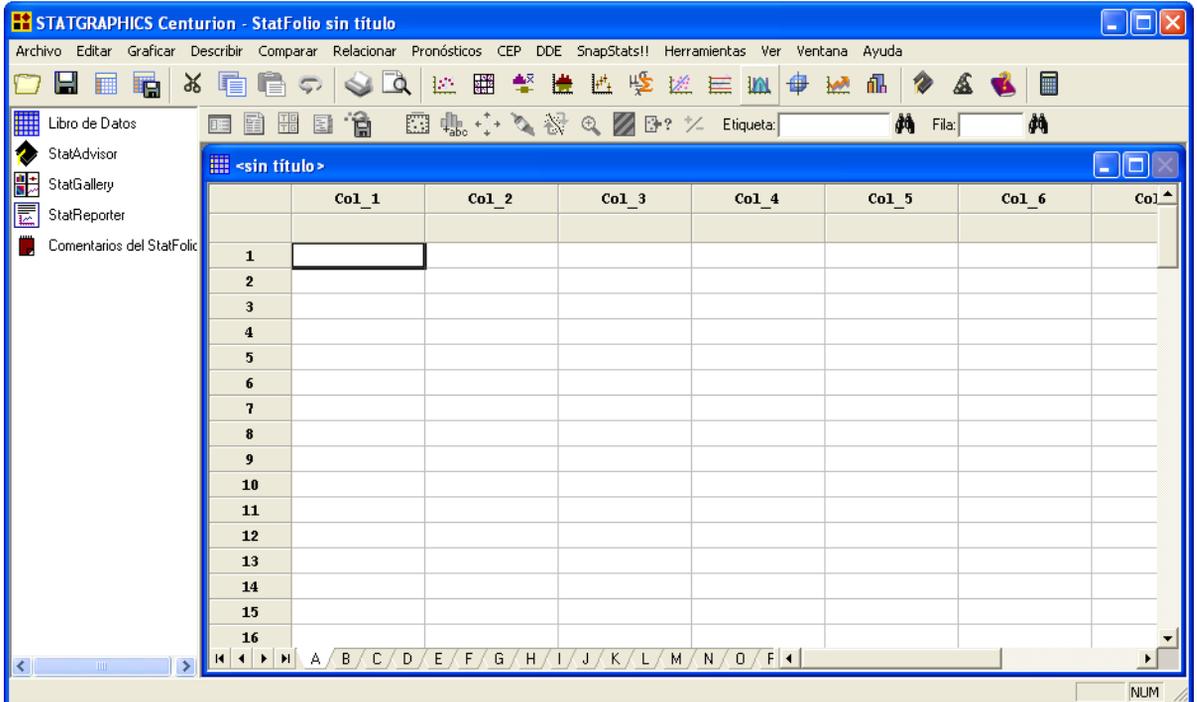


Figura 1-13. El libro de datos de STATGRAPHICS

En una típica hoja, cada fila contiene información sobre un individuo simple, caso u observación, mientras que cada columna representa una variable.

Por ejemplo, supongamos que queremos utilizar STATGRAPHICS Centurion XVI para analizar los datos del censo de Estados Unidos de 2000. Una pequeña sección de los resultados de este censo se muestra a continuación:

Estado	Población	Edad media	% Mujeres	Ingresos per cápita
Alabama	4,447,100	35.8	51.7	\$18,819
Alaska	626,932	32.4	48.3	\$22,660
Arizona	5,130,632	34.2	50.1	\$20,275
Arkansas	2,673,400	36.0	51.2	\$16,904
California	33,871,648	33.3	50.2	\$22,711
Colorado	4,301,261	34.3	49.6	\$24,049

Figura 1-14. Datos del Censo de E.U.A. de 2000

Cuando introducimos los datos en una hoja de STATGRAPHICS Centurion XVI, la información referente a cada estado se situará en filas diferentes. Se crearán cinco columnas para situar los nombres de los estados y el contenido de la información del censo.

Para introducir datos tales como los que se ven en la tabla anterior STATGRAPHICS Centurion XVI ofrece dos posibilidades:

1. Teclar directamente los datos en el libro de datos de STATGRAPHICS Centurion XVI.
2. Introducir los datos en otro programa como Excel y leerlos o copiarlos al libro de datos de STATGRAPHICS Centurion XVI.

En esta sección, tomaremos la primera opción. Para comenzar, haga doble clic en el encabezamiento de la primera columna donde aparece el nombre *Col_1*. Esto mostrará un cuadro de diálogo que puede utilizar para cambiar propiedades importantes para cada columna:



Figura 1-15. Cuadro de diálogo utilizada para definir columnas

Cada columna en una hoja de STATGRAPHICS Centurion XVI tiene un nombre, comentario y tipo asociados con ella:

- *Nombre*– Dar a cada columna un único nombre conteniendo de 1 a 32 caracteres. Estos nombres son utilizados por el programa para identificar las variables a analizar cuando se selecciona un procedimiento estadístico. También sirven como etiquetas por defecto en la mayoría de los gráficos. Los nombres pueden contener cualquier tipo de caracteres y no son sensible a letras mayúsculas. Se permiten los espacios. El programa presentará un mensaje de error si intenta utilizar el mismo nombre para más de una columna en una hoja, aunque las columnas de distintas hojas pueden tener nombres idénticos.
- *Comentario*– Introducir un comentario que identifique los datos en la columna. Los comentarios pueden tener hasta 64 caracteres y son opcionales. Si se introducen, aparecerán en la segunda línea de la cabecera de cada columna.
- *Tipo* – Especificar el tipo de datos a introducir en la columna. En este caso, la primera columna contiene los nombres de los estados que deben ser de tipo *Carácter*. Las otras columnas puede situarse como *Numéricas* o como *Enteras* o *Decimales de punto fijo* si quiere restringir el tipo de datos que van a ser introducidos. Para información detallada de la columna *Tipos*, ver el capítulo 2.

Después de la definición de cada columna, presione *Aceptar*. Se crean 5 columnas como se muestra a continuación:

	Estado	Población	Edad Media	Porcentaje de Mujeres	Renta per cápita
	Nombre del Estado		Años		Media
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					

Figura 1-16. Hoja de datos con nombres de columnas de STATGRAPHICS Centurion XVI

Ahora se introducen los datos tal y como se haría en una hoja de cálculo, utilizando las teclas de flechas para moverse de celda en celda. No introduzca comas en números grandes. Cuando finalice, la hoja debe tener la siguiente apariencia:

	Estado	Población	Edad Media	Porcentaje de Mujeres	Renta per cápita
	Nombre del Estado		Años		Media
1	Alabama	4447100	35.8	51.7	18819
2	Alaska	626932	32.4	48.3	22660
3	Arizona	5130632	34.2	50.1	20275
4	Arkansas	2673400	36	51.2	16904
5	California	33871648	33.3	50.2	22711
6	Colorado	4301261	34.3	49.6	24049
7					
8					
9					
10					
11					

Figura 1-17. Hoja de datos de STATGRAPHICS Centurion XVI después de introducir 6 filas de datos

Finalmente, necesitará guardar los datos en un archivo. Elija *Archivo – Guardar – Guardar archivo de datos* en el menú principal. Seleccionar un nombre de archivo en el cual se guardan los datos:

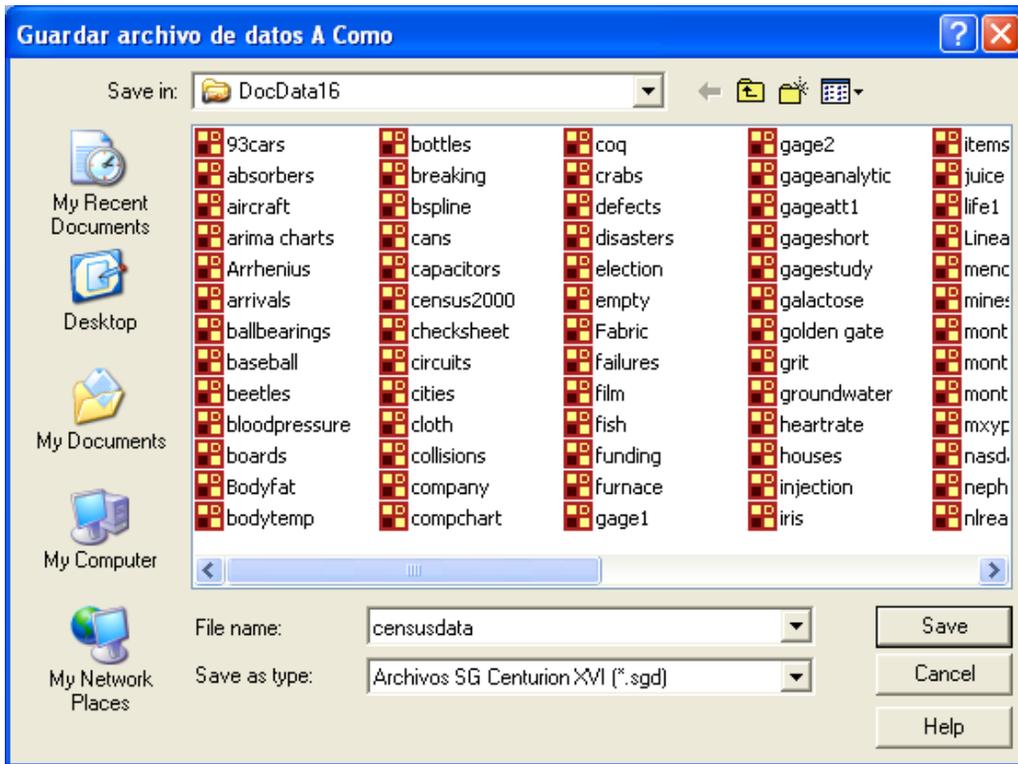


Figura 1-18. Cuadro de diálogo Selección de archivos de datos para guardar

Los archivos de datos en STATGRAPHICS Centurion XVI se guardan en disco con la extensión *.sgd* por defecto, la cual almacena los datos en formato XML. Cuando se guarda el archivo, puede cambiar sus características a diferente formato de archivo en el campo *Guardar como* si lo desea.

1.4 Leyendo un archivo de datos guardado

Una vez que los datos han sido introducidos en la hoja, están listos para el análisis. Para tomar un ejemplo más interesante, recuperaremos los datos del censo para 50 estados y el distrito de Columbia, que son provistos en STATGRAPHICS Centurion XVI en una archivo de nombre *census2000.sgd*. Para abrir el archivo de datos, seleccione *Archivo – Abrir – Abrir orígenes de datos* en el menú principal. Primeramente se preguntará por la ruta de localización en el disco duro de los datos a los que queremos acceder:

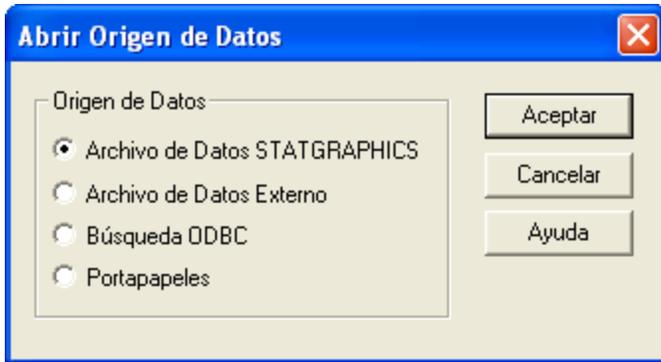


Figura 1-19. Cuadro de diálogo Abrir origen de datos

La selección por defecto es correcta en este caso. A continuación, seleccione el nombre del archivo que contiene los datos:

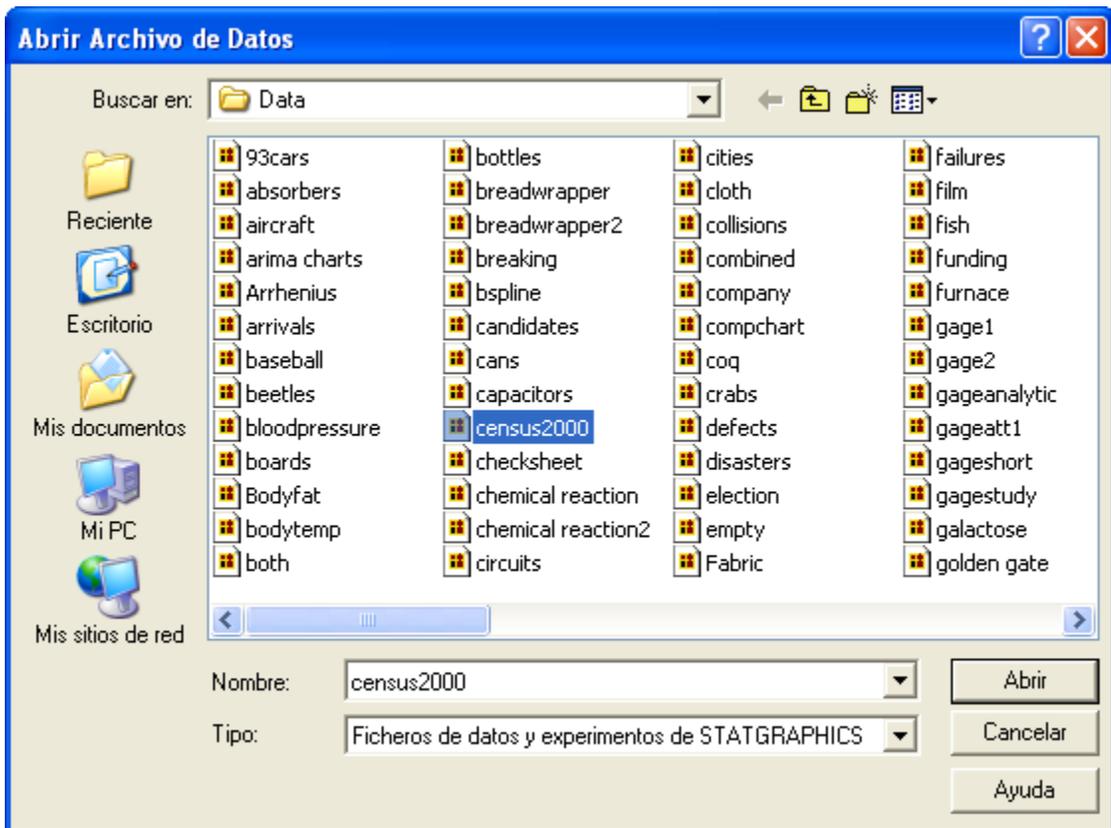


Figura 1-20. Cuadro de diálogo Abrir archivo de datos

El archivo de ejemplo se localiza en el directorio de datos por defecto (usualmente *c:\Archivos de programa\Statgraphics\STATGRAPHICS Centurion XVI\Data*). Abriendo el archivo se leen las 51 filas de datos en la hoja:

	State	Population	Median Age	Percent Female	Per Capita Income	Col_6	Col_7	Col_8
1	Alabama	4447100	35,8	51,7	18819			
2	Alaska	626932	32,4	48,3	22660			
3	Arizona	5130632	34,2	50,1	20275			
4	Arkansas	2673400	36	51,2	16904			
5	California	33871648	33,3	50,2	22711			
6	Colorado	4301261	34,3	49,6	24049			
7	Connecticut	3405565	37,4	51,6	28766			
8	Delaware	783600	36	51,4	23305			
9	D.C.	572059	34,6	52,9	28659			
10	Florida	15982378	38,7	51,2	21557			
11	Georgia	8186453	33,4	50,8	21154			
12	Hawaii	1211537	36,2	49,8	21525			
13	Idaho	1293953	33,2	49,9	17841			
14	Illinois	12419293	34,7	51	23104			
15	Indiana	6080485	35,2	51	20307			

Figura 1-21. Hoja mostrando el contenido del archivo *Census2000.sgd*

1.5 Analizando los datos

Una vez que los datos han sido leídos en el libro de datos de STATGRAPHICS Centurion XVI, cualquiera de los más de 160 procedimientos estadísticos pueden ser utilizados de varias formas:

1. Por selección del procedimiento deseado en el menú principal.
2. Presionando uno de los botones de la barra de herramientas.
3. Invocando el StatWizard (asistente estadístico) presionando el botón en la barra de herramientas mostrado como *wizard's cap*.

Comenzamos resumiendo la variabilidad en los ingresos per cápita a través de los estados. El mayor procedimiento para resumir una columna simple de datos numéricos es *Análisis de una variable*. Este procedimiento calcula resúmenes estadísticos tales como la media muestral y la desviación típica. También crea varios gráficos, incluyendo el histograma y el gráfico de caja y bigotes.

La localización del procedimiento *Análisis de una variable* depende del menú que se esté utilizando:

1. **Menú clásico:** Seleccione *Describir – Datos numéricos – Análisis de una variable*.
2. **Menú Seis Sigma:** Seleccione *Analizar – Datos de variable – Análisis de una variable*.

Como todos los procedimientos estadísticos, *Análisis de una variable* comienza mostrando los datos en el cuadro de diálogo de entrada:



Figura 1-22. Cuadro de diálogo de entrada de datos en *Análisis de una variable*

La lista de la parte izquierda del cuadro muestra los nombres de todas las columnas en la hoja que contienen datos. Para analizar los datos en la columna *Per Capita Income*, haga clic en su nombre y presione el botón con la flecha negra del campo *Datos*. Se colocará el nombre de la columna que contiene los datos de ingresos en el campo *Datos*. Dejar el campo *Selección* en blanco (se utiliza sólo cuando se quiere analizar un subconjunto en la hoja en lugar de todas sus filas).

Cuando se pulse *Aceptar*, aparece el cuadro de diálogo *Tablas y Gráficos*. Este cuadro de diálogo muestra las tablas y gráficos que son posibles en el procedimiento *Análisis de una variable*. Por ahora, serán aceptadas las características por defecto:

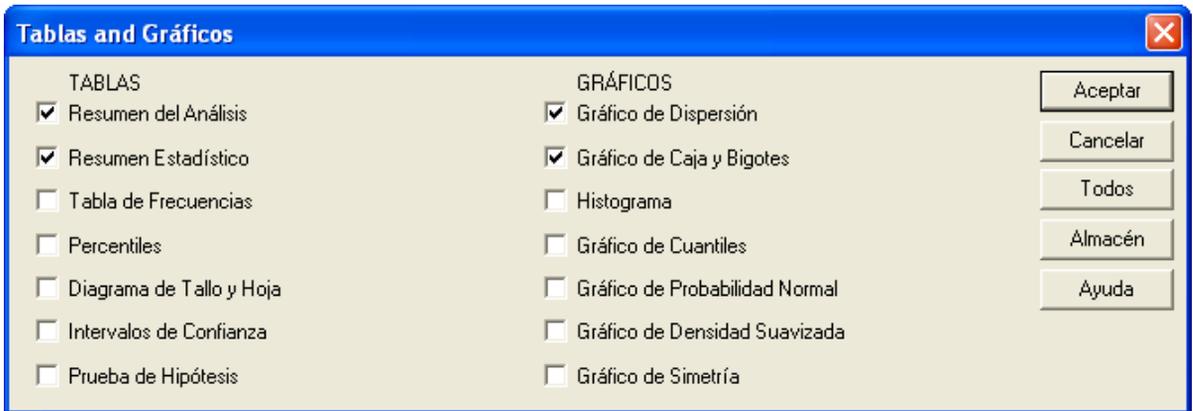


Figura 1-23. Cuadro de diálogo Tablas y Gráficos

Cuando se presiona *Aceptar* otra vez, se creará una nueva ventana de análisis:

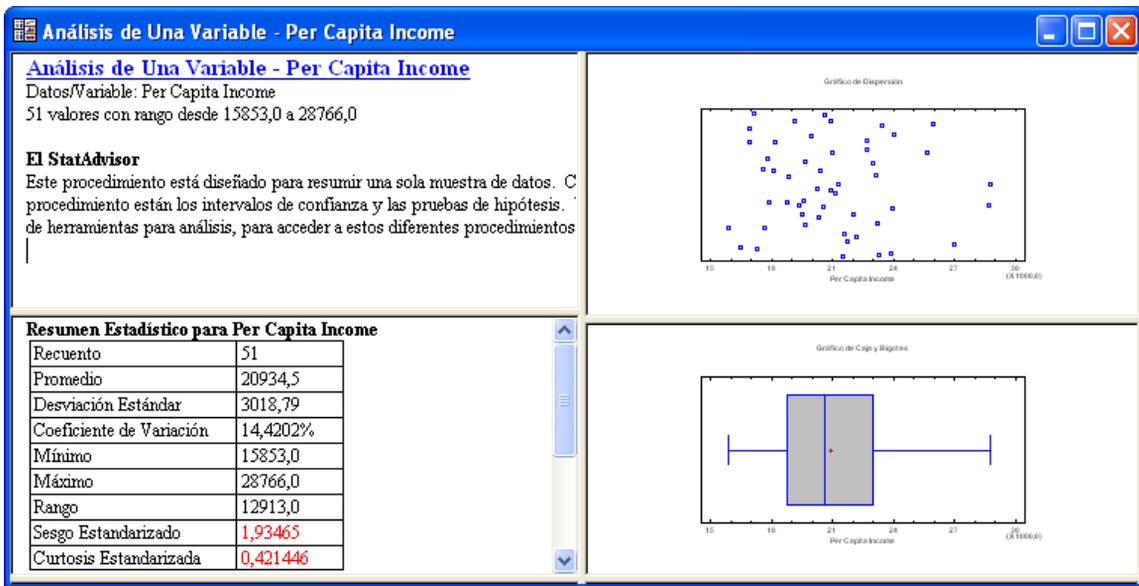


Figura 1-24. Ventana de Análisis de una variable

Esta ventana contiene 4 paneles, divididos por barras móviles. Los dos paneles de la izquierda muestran salida tabular, mientras que los dos paneles de la parte derecha muestran salida gráfica. Si se hace doble clic en cualquier zona del panel izquierdo, la tabla de resúmenes estadísticos se maximizará:

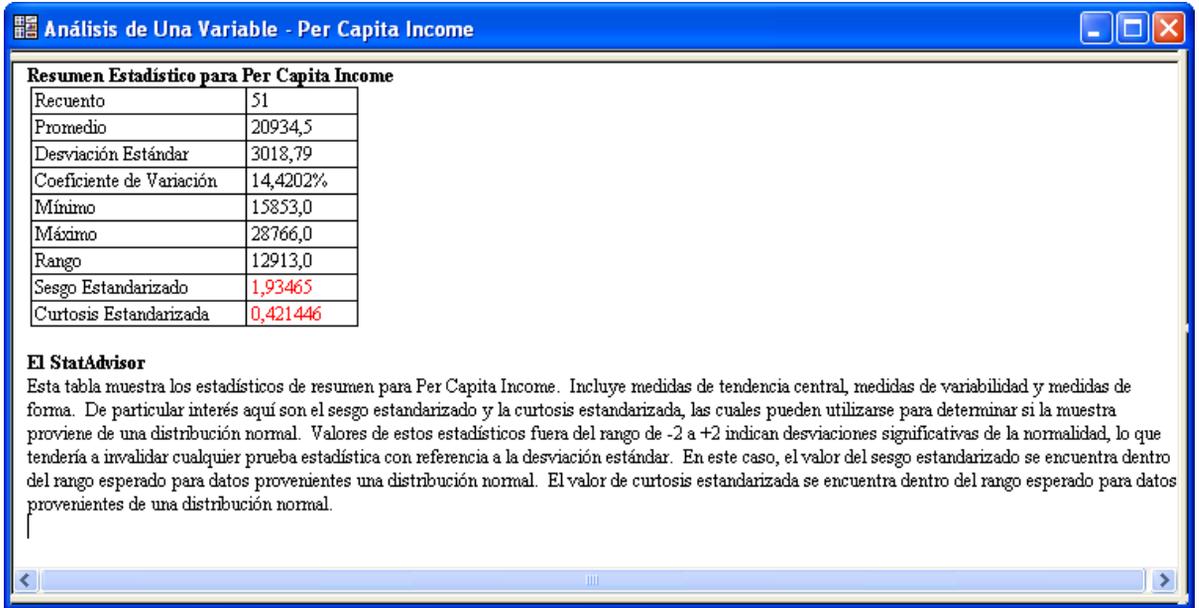


Figura 1-25. Panel de estadísticos resumen maximizado

En la tabla se ofrecen varias estadísticas interesantes. Para los $n = 51$ estados más D.C., los ingresos per cápita varían en los rangos \$15,853 a \$28,766. Los ingresos medios per cápita son \$20,934.50.

Debajo de la tabla se encuentra la salida del StatAdvisor, que ofrece una interpretación corta de los resultados. En este caso, el StatAdvisor se concentra en los dos estadísticos que miden la asimetría y la curtosis en los datos. Como explica el StatAdvisor, los datos que provienen de una distribución normal o Gaussiana deben tener la asimetría y curtosis estandarizadas entre -2 y $+2$. En este caso, ambos estadísticos están dentro del rango, indicando un modelo de ajuste razonable a la curva normal para las observaciones, aunque la asimetría es muy cerrada para ser estadísticamente significativa.

Haciendo doble clic otra vez en la tabla de estadísticos resumen se restaurará la división original de la salida. Un doble clic en el panel derecho maximiza el gráfico de caja y bigotes:

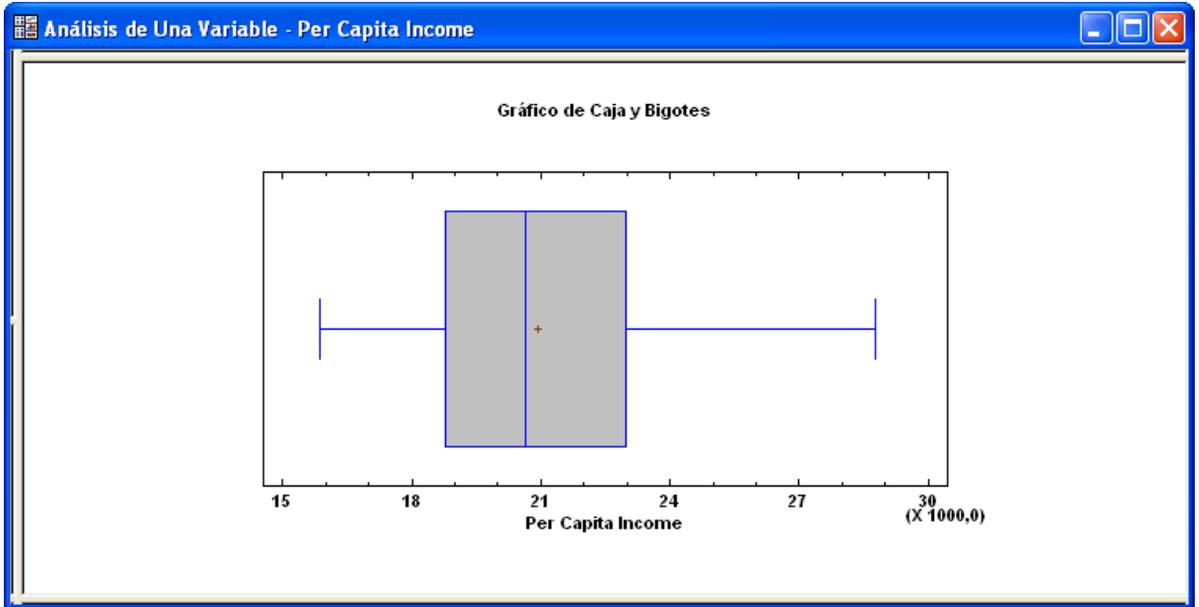


Figura 1-26. Panel Gráfico de caja y bigotes maximizado

El gráfico de caja y bigotes, inventado por John Tukey, aporta 5-números resumen de la muestra de datos. La caja central cubre la mitad de los datos, extendiéndose desde el cuartil inferior hasta el cuartil superior. Las líneas extendidas a izquierda y derecha de la caja (los bigotes) muestran la localización del mayor y menor de los datos. La mediana de los datos está indicada por la línea vertical en el interior de la caja, mientras que el signo (+) muestra la localización de la media muestral. El hecho de que el bigote superior es algo más largo que el inferior, a la vez que la media es algo mayor que la mediana, es indicativo de asimetría positiva en los datos.

1.6 Utilizando la barra de herramientas de análisis

Cuando una ventana de análisis como la de *Análisis de una variable* se muestra por primera vez, sólo se incluyen algunas de las tablas y gráficos posibles. Para mostrar salida adicional, debe pulsar el botón apropiado en la *Barra de herramientas de análisis*, que se muestra inmediatamente encima del título del análisis:



Figura 1-27. La Barra de herramientas de análisis

Los botones de la barra de herramientas de análisis son muy importantes. A continuación se resumen las acciones de sus primeros seis botones de la izquierda:

	Nombre	Función
	Entrada de diálogo	Muestra el cuadro de diálogo de entrada de datos que permite cambiar las columnas de datos seleccionadas para el análisis.
	Opciones de análisis	Selecciona opciones que pueden ser aplicadas a todas las tablas y gráficos en el análisis actual.
	Tablas y Gráficos	Muestra una lista de otras tablas y gráficos que pueden ser creados.
	Opciones de panel	Selecciona opciones para aplicar solo a la tabla o gráfico actualmente maximizado.
	Guardar resultados	Permite calcular estadísticos sobre las columnas de la hoja para ser guardados.
	Opciones gráficas	Permite cambiar títulos, escalas y otras características del gráfico actualmente maximizado.

Figura 1-28. Botones importantes en la barra de herramientas de análisis

Los botones adicionales de la derecha de la barra permiten otras acciones cuando un gráfico está maximizado, tal y como se explica en el capítulo 5.

Por ejemplo, si se presiona el botón *Tablas y Gráficos* , un cuadro de diálogo mostrará una lista de otras opciones gráficas posibles en el procedimiento *Análisis de una variable*:

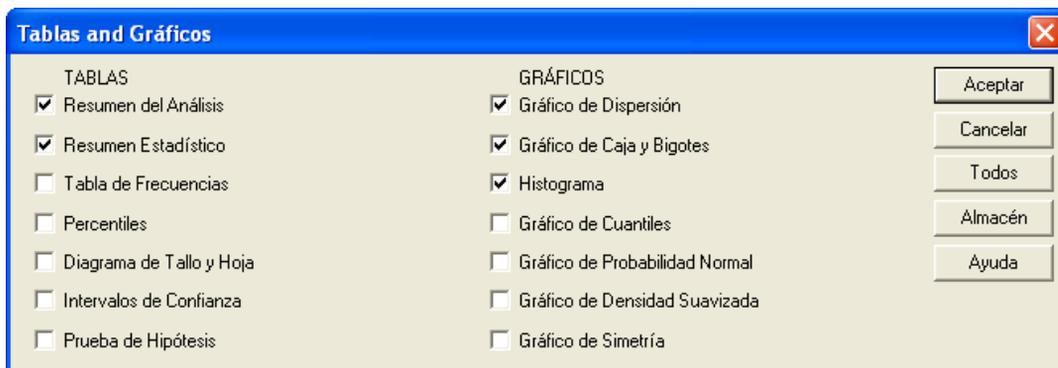


Figura 1-29. Lista de opciones para Tablas y Gráficos

Marcando el cuadro relativo a *Histograma de frecuencias* y presionando *Aceptar* se añade un tercer panel en el lado derecho de la ventana de análisis:

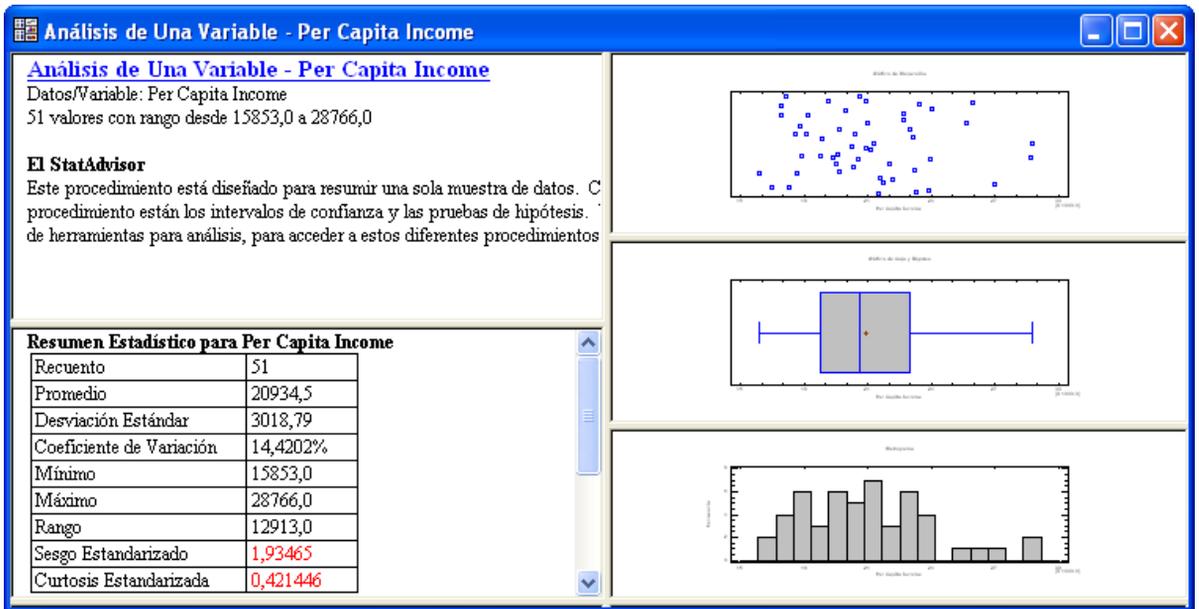


Figura 1-30. Ventana de Análisis de una variable con el Histograma de frecuencias añadido

Si hace doble clic en el histograma para maximizarlo y presiona el botón *Opciones de panel*, se muestra un cuadro de diálogo con opciones específicas para el histograma:



Figura 1-31. Cuadro de diálogo de Opciones de panel para el histograma de frecuencias

Utilizando este cuadro, pueden cambiarse el número de barras en el histograma, así como el rango que cubre. Si *Número de clases* se sitúa en 15 y se presiona *Aceptar*, el histograma cambiará para reflejar la nueva selección:

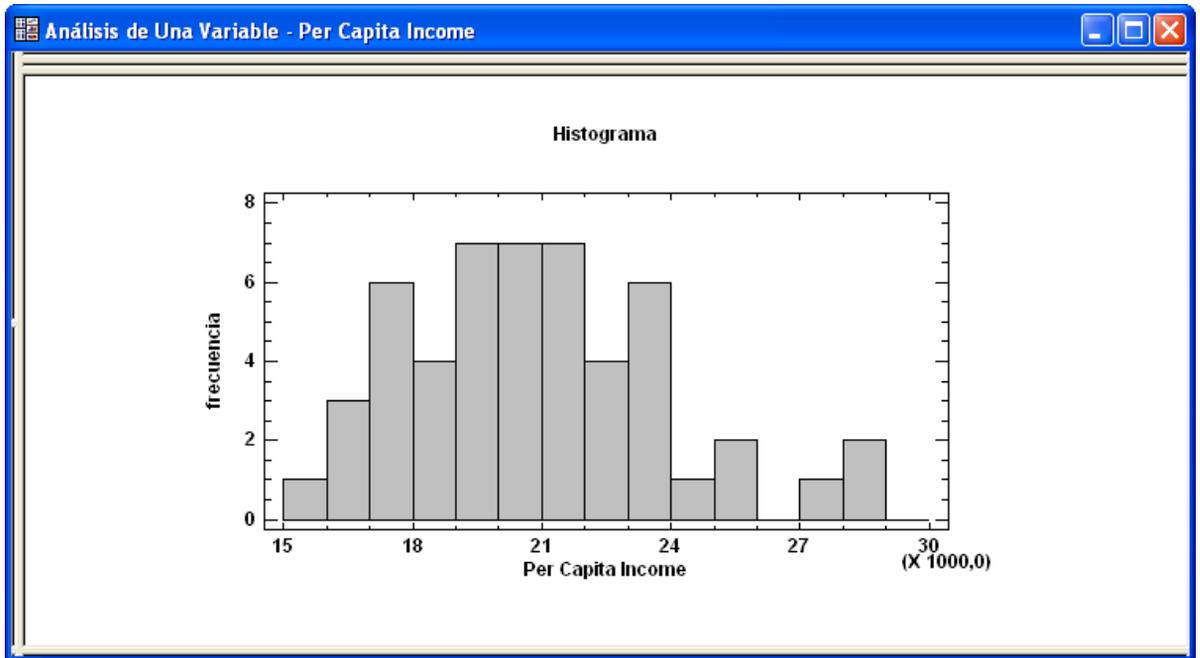


Figura 1-32. Histograma de frecuencias después de cambiar el número de clases

Puede también cambiar el patrón de relleno y/o el color de las barras en el histograma presionando el botón *Opciones gráficas*. Se muestra un cuadro de diálogo que permite cambiar la mayoría de las características del gráfico. Si hace clic en la solapa *Relleno*, se mostrará lo siguiente

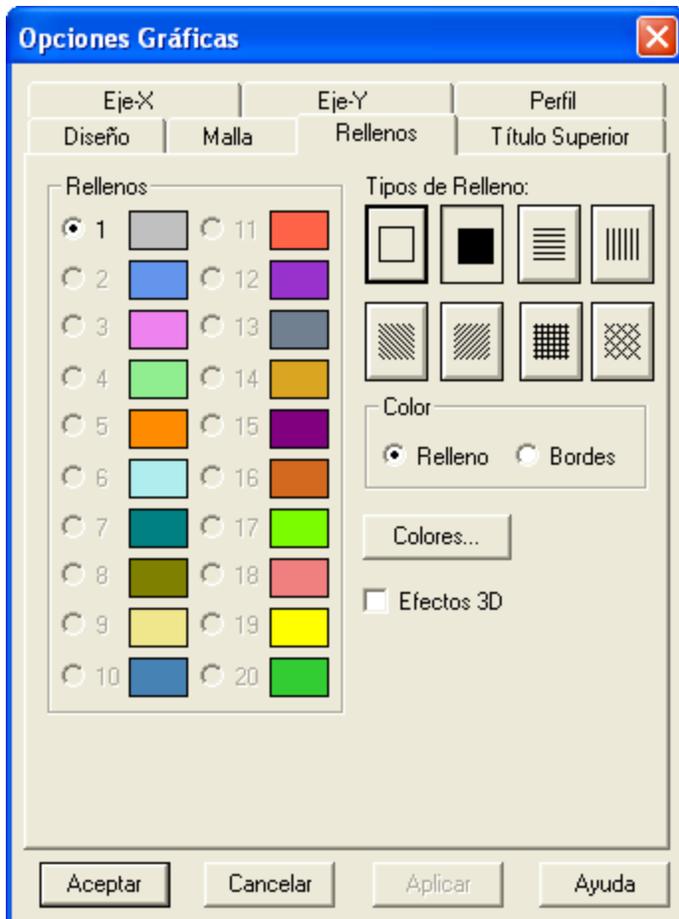


Figura 1-33. Cuadro de diálogo de Opciones gráficas

Haciendo clic en el botón #1 y seleccionando un nuevo *Tipo de relleno* o *Color* se cambiarán las barras en el histograma.

NOTA: Las operaciones de la mayoría de los botones en la barra de herramientas de análisis pueden obtenerse haciendo clic con el botón derecho del ratón en el panel que contiene la tabla o gráfico. Se mostrará un menú emergente conteniendo las opciones disponibles.

1.7 Difundiendo los resultados

Una vez que el análisis ha sido ejecutado, los resultados pueden ser difundidos por varios caminos. Estos incluyen:

<i>Acción</i>	<i>Método</i>
Imprimir la salida.	Presione el botón de la impresora en la barra de herramientas principal para imprimir todas las tablas y gráficos, o haga clic en un panel simple con el botón derecho del ratón y elija <i>Imprimir</i> en el menú emergente resultante para imprimir una tabla o gráfico simple.
Publicando la salida para mostrarla en un navegador.	Seleccione <i>StatPublish</i> en el menú <i>Archivo</i> . Se mostrará un cuadro de diálogo para especificar la localización de la salida HTML.
Copiar la salida en otra aplicación.	Haga clic en la tabla o gráfico que va a ser copiado y seleccione <i>Copiar</i> en el menú <i>Edición</i> . Active otra aplicación y seleccione <i>Editar– Pegar</i> .
Guardar los resultados en un informe.	Presione el botón derecho del ratón y seleccione <i>Copiar análisis a StatReporter</i> . El StatReporter, descrito en el capítulo 7, puede guardarse como un archivo en formato RTF que puede ser importado en programas como Microsoft Word.
Guardar un gráfico en un archivo imagen.	Maximice el gráfico a guardar y seleccione <i>Guardar gráfico</i> en el menú <i>Archivo</i> .

Figura 1-34. Métodos para difundir los resultados del análisis

Cada una de estas operaciones se describe en capítulos posteriores.

1.8 Guardando su trabajo

Puede guardar los resultados de la sesión actual de STATGRAPHICS Centurion XVI seleccionando *Guardar StatFolio* en el menú *Archivo* e introduciendo el nombre del archivo:

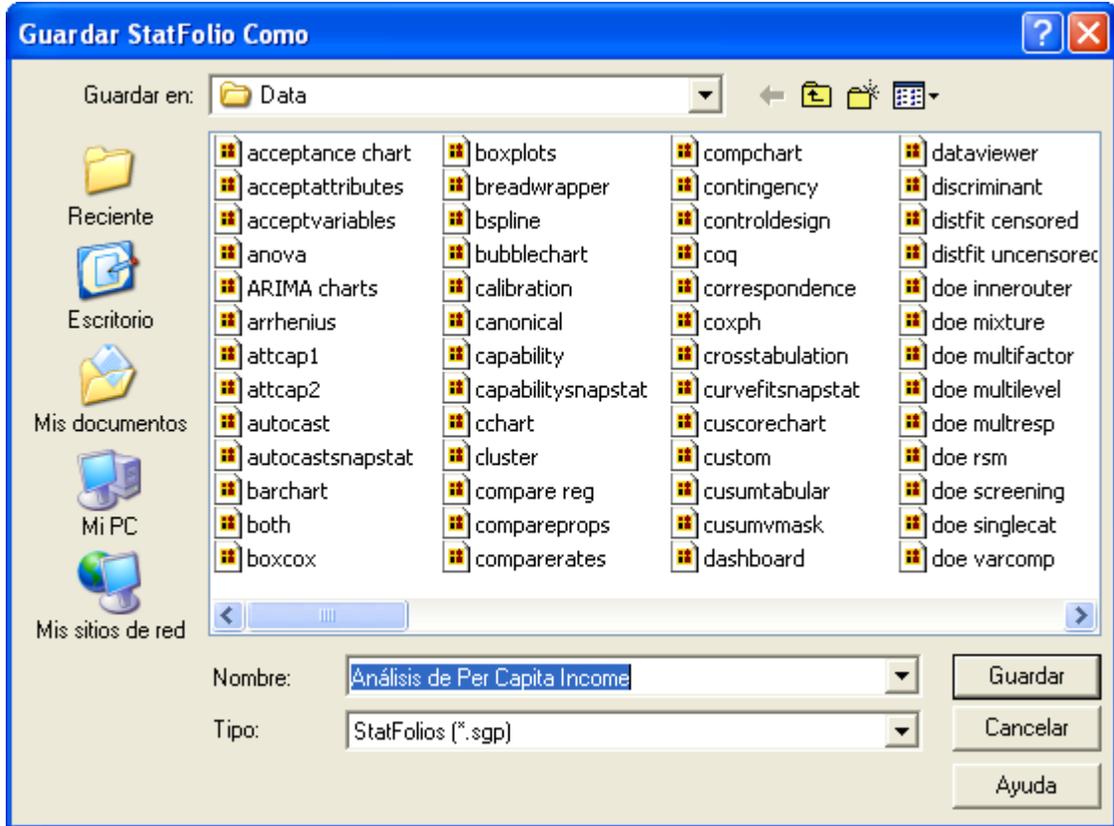


Figura 1-35. Cuadro de diálogo para Guardar StatFolio

Un StatFolio contiene instrucciones de cómo crear cada uno de los análisis de la sesión actual, con punteros a los archivos o bases de datos que contiene los datos. Si se vuelve a leer el StatFolio en un momento posterior, automáticamente se volverán a leer los datos y se ejecutarán los análisis. Se retendrán las opciones seleccionadas para el análisis.

NOTA #1: Si los datos en los orígenes de datos cambian durante el tiempo que el StatFolio está guardado y éste vuelve a leerse, los análisis cambiarán para reflejar los nuevos valores. Esto aporta un método simple para ejecutar análisis que necesitan repetirse en un período determinado sin tener que ser creados de nuevo.

NOTA #2: Los datos y el StatFolio se almacenan en diferentes archivos. Si necesita mover un StatFolio de un ordenador a otro, esté seguro de mover también los archivos de datos.

Administración de datos

Accediendo a datos de archivos y bases de datos, transformando valores de datos y generando patrones de datos.

Para analizar datos en STATGRAPHICS Centurion XVI, en primer lugar debe situar en memoria el libro de datos. El libro de datos consiste en una ventana en forma de tabla formada por 26 hojas. Una hoja es una matriz rectangular formada por filas y columnas. Cada columna en una hoja representa una variable. Cada fila representa una observación o caso. Por ejemplo, la hoja de datos que se presenta a continuación contiene información de un número de diferentes marcas y modelos de automóviles.

	Make	Model	Type	Min Price price for basic version in \$1,000	Mid Price average of min and max prices in \$1,000
1	Acura	Integra	Small	12,9	15,9
2	Acura	Legend	Midsize	29,2	33,9
3	Audi	90	Compact	25,9	29,1
4	Audi	100	Midsize	30,8	37,7
5	BMW	535i	Midsize	23,7	30
6	Buick	Century	Midsize	14,2	15,7
7	Buick	LeSabre	Large	19,9	20,8
8	Buick	Roadmaster	Large	22,6	23,7
9	Buick	Riviera	Midsize	26,3	26,3
10	Cadillac	DeVille	Large	33	34,7

Figura 2-1. Muestra de hoja de datos

El capítulo describe todo lo necesario que hay que conocer acerca de STATGRAPHICS Centurion XVI, incluyendo cómo se accede, cómo se manipula y cómo se utiliza en los análisis estadísticos.

2.1 El libro de datos

Cada columna de una hoja de STATGRAPHICS Centurion XVI representa una variable diferente. Las variables son habitualmente atributos o mediciones asociadas con los ítems que definen las filas de una hoja. Por ejemplo, en la hoja *93cars*, hay una columna identificando la marca de cada automóvil, una columna identificando su tipo, columnas conteniendo las millas por galón consumidas conduciendo en ciudad y en carretera, columnas conteniendo la longitud de los automóviles, altura y anchura e información similar. Cada columna tiene un *nombre* y un *tipo* asociado con ella. El nombre se utiliza para identificar los datos a utilizar en el análisis estadístico. El tipo afecta a cómo será analizada la variable. También, asociado con cada columna, puede introducirse un *comentario* adicional, que se utilice para aportar información adicional acerca de los contenidos de la columna. NOTA: los datos fueron obtenidos del Journal of Statistical Education Data Archive (www.amstat.org/publications/jse/jse_data_archive.html) y son utilizados con permiso.

Para mostrar o cambiar las propiedades de una columna en la hoja, haga doble clic en el nombre de la columna y se mostrará el cuadro de diálogo *Modificar Columna*:

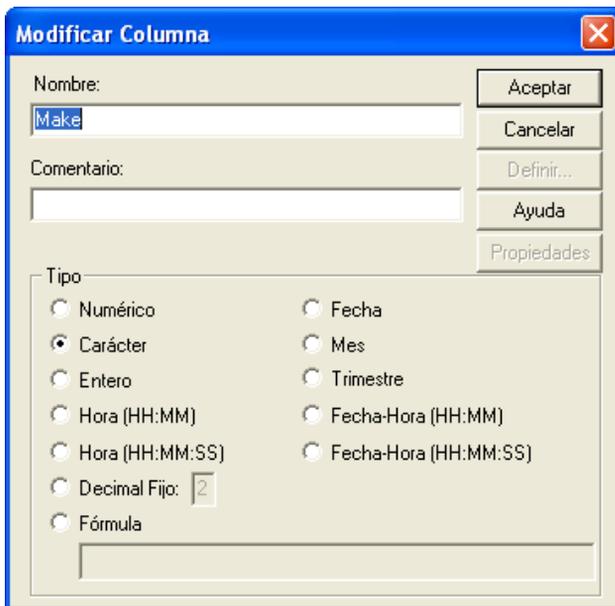


Figura 2-2. Cuadro de diálogo utilizada para cambiar propiedades de las columnas

Puede especificar:

1. **Nombre:** de 1 a 32 caracteres. Cuando se muestran análisis estadísticos, las columnas se identifican utilizando estos nombres. Cada columna en una hoja debe tener un único nombre, sin embargo puede tener diferentes nombres a lo largo de las distintas hojas de un libro de datos. Los nombres pueden incluir cualquier carácter, incluyendo espacios. Los nombres de las variables no son sensibles a mayúsculas/minúsculas.
2. **Comentario:** de 0 a 64 caracteres, aportando información adicional acerca de los contenidos de las columnas.
3. **Tipo:** el tipo de datos permitido para la columna. Se pueden especificar los siguientes tipos:

<i>Tipo</i>	<i>Contenido</i>	<i>Ejemplo</i>
Numérico	Cualquier número válido	3.14
Carácter	Una cadena alfanumérica	Chevrolet
Entero	Un número entero	105
Fecha	Mes, día y año	4/30/05
Mes	Mes y año	4/05
Trimestre	Trimestre y año	Q2/05
Hora (HH:MM)	Hora y minuto	3:15
Hora (HH:MM:SS)	Hora, minuto y segundo	3:15:53
Fecha-Hora (HH:MM)	Mes, día, año, hora y minuto	4/30/05 3:15
Fecha-Hora (HH:MM:SS)	Mes, día, año, hora, minutos y segundos	4/30/05 3:15:53
Decimal fijo	Números de 1 a 9 caracteres	34.10
Fórmula	Calculada de otras columnas	MPG City/MPG Highway

Figura 2-3. Tipos de columna

Cuando se introducen datos en una hoja, tienen que ser conformes con el tipo de columna en la cual se insertan. Por ejemplo, si se intenta introducir un nombre en una columna numérica resultará un error. Cuando se introducen datos, su formato debe coincidir con las características actuales de Windows. En particular, STATGRAPHICS Centurion XVI asume las características actuales de Windows para:

1. Separador decimal para valores numéricos
2. Formato de hora y separador de hora para valores con horas

3. Formato de fecha corta y separador para fechas

Para comprobar las características actuales de su ordenador, acceda al Panel de Control de Windows. Cuando introduzca una fecha, debe utilizar el formato especificado en el cuadro de diálogo *Editar - Preferencias*. Pueden utilizarse 4 dígitos en los años (como en 4/30/2005) o 2 dígitos para los años (como en 4/30/05). Si se utilizan dos dígitos en los años, se asume la problemática con los años de 1950 hasta 2049.

Más información acerca de las fórmulas en las columnas puede encontrarse en una sección posterior de este capítulo titulada *Manipulando datos*.

2.2 Accediendo a los datos

El capítulo 1 muestra cómo pueden introducirse los datos de modo manual en la hoja de datos. Además, los usuarios accederán a los datos ya existentes de otro archivo o aplicación. Hay tres caminos básicos para situar datos existentes en una hoja de datos de STATGRAPHICS Centurion XVI:

1. **Leer un archivo ya existente:** Si los datos han sido introducidos previamente en un archivo, puede leerlos en la hoja de datos seleccionando *Archivo – Abrir – Abrir origen de datos*. Esto le permitirá leer datos almacenados en varios formatos de archivo, incluyendo archivos Excel, archivos de texto ASCII delimitado, archivos XML, archivos de STATGRAPHICS, y archivos de otros paquetes estadísticos.
2. **Copiar y pegar utilizando el portapapeles de Windows:** Si tiene los datos leídos en un programa tal como Excel, puede fácilmente copiarlos en el portapapeles y pegarlos en STATGRAPHICS Centurion XVI seleccionando *Editar – Pegar*.
3. **Emitir una consulta SQL para recuperar los datos de una base de datos:** Si los datos residen en una base de datos compatible ODBC, tal como Oracle o Microsoft Access, pueden ser recuperados seleccionando *Archivo – Abrir – Abrir origen de datos* y seleccionando entonces la *Consulta ODBC*.

2.2.1 Leyendo datos de un archivo de datos de STATGRAPHICS Centurion

Para leer datos que han sido ya guardados en un archivo STATGRAPHICS Centurion, seleccionar una de las 26 hojas de datos en el libro de datos haciendo clic en su pestaña correspondiente. Seleccionar entonces *Archivo – Abrir – Abrir origen de datos* y especificar *Archivos de datos STATGRAPHICS* en el cuadro de diálogo que se muestra a continuación:



Figura 2-4. Cuadro de diálogo Abrir origen de datos

Después presionar *Aceptar* y seleccionar el archivo STATGRAPHICS deseado:

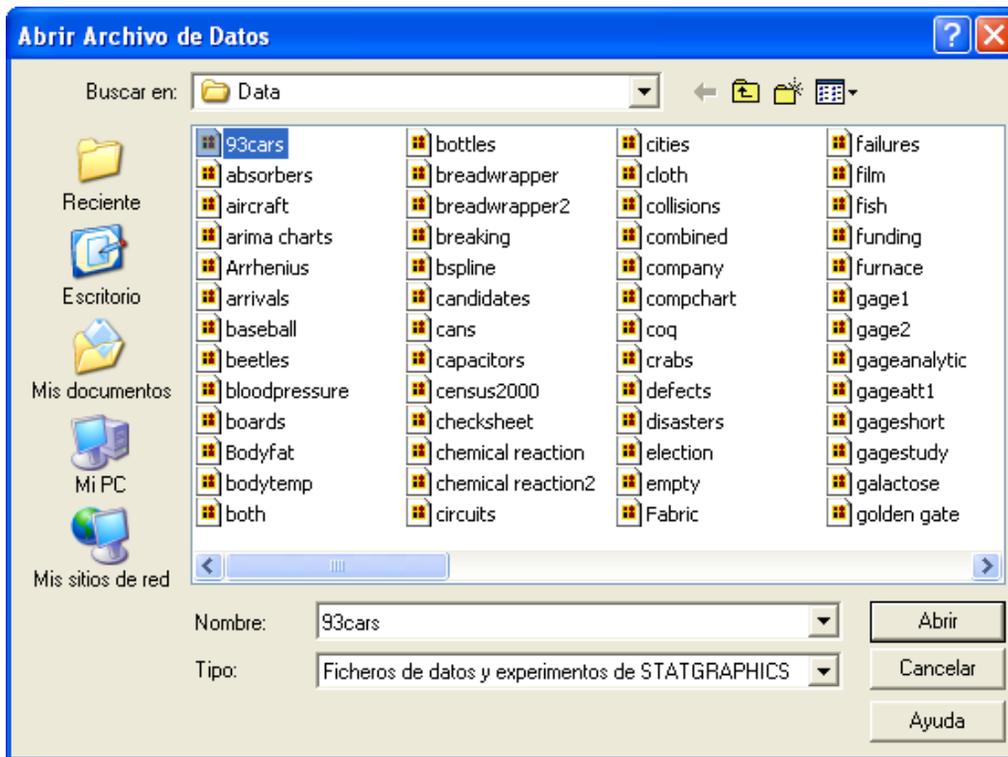


Figura 2-5. Seleccionando un archivo de datos STATGRAPHICS

Puede leer datos de STATGRAPHICS Centurion XVI o de una versión anterior de STATGRAPHICS, incluyendo STATGRAPHICS *Plus*. Los datos del archivo reemplazarán el contenido de la hoja de datos actualmente seleccionada.

2.2.2 Leyendo datos de un archivo Excel, ASCII, XML, u otro archivo externo de datos

Para leer datos que han sido guardados en un archivo de datos creado por otra aplicación, seleccione una de las 26 hojas de datos en la hoja de datos haciendo clic sobre su pestaña. Entonces seleccione *Archivo – Abrir – Abrir origen de datos* y especifique *Datos de archivo externo* en el cuadro de diálogo que se muestra a continuación:

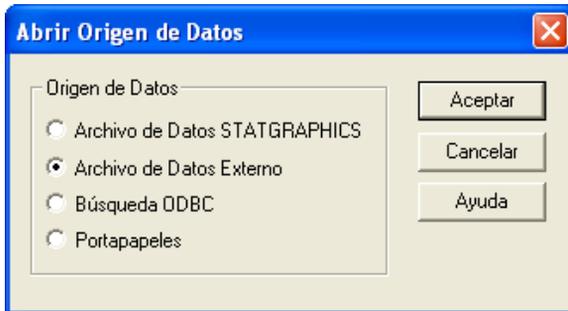


Figura 2-6. Caja de diálogo *Abrir origen de datos*

Después de presionar *Aceptar*, se mostrará un cuadro de diálogo en la cual se especificará el archivo que va a ser importado y otra información relevante:

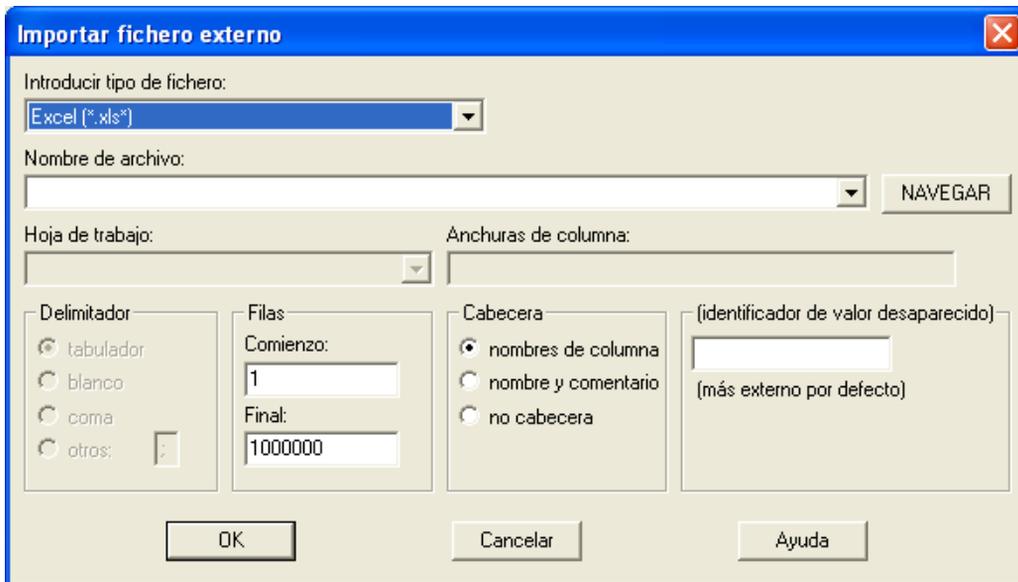


Figura 2-7. *Seleccionando un archivo de datos externo*

Los campos en el cuadro de diálogo incluyen:

1. **Tipo de archivo de entrada** – Tipo de archivo a importar. STATGRAPHICS Centurion XVI puede importar datos de muchas otras aplicaciones, incluyendo Excel, Matlab, Minitab, JMP, SPSS, SAS, y de la mayoría del resto de los paquetes estadísticos.
2. **Nombre de archivo** – nombre del archivo que va a ser importado. Presione el botón *MOSTRAR* para seleccionar el archivo deseado.
3. **Hoja de cálculo** – nombre de la hoja de cálculo a importar (si procede). Sólo puede ser leída una hoja a la vez.
4. **Anchuras de columnas** – anchura de cada columna, separadas por comas (sólo para archivos ASCII con formato).
5. **Delimitador** – delimitador de columna (sólo para archivos ASCII delimitados).
6. **Filas** - el rango de filas con el que será leída la hoja de cálculo. Este rango incluye los nombres de las variables, si están presentes.
7. **Cabecera** - información contenida en las dos primeras filas del rango especificado (para hojas de cálculo de programas tales como Excel). Las dos filas inmediatamente por encima de los datos se leerán como nombres de columna y/o comentarios. Si los nombres no están contenidos en el archivo, entonces se generarán nombres por defecto.
8. **Identificador de valor desaparecido** - Algún símbolo especial utilizado en el archivo externo para indicar valor desaparecido, tal como *NA*. Las celdas que contengan este valor especificado se convertirán en celdas vacías cuando se sitúan en una hoja de datos de STATGRAPHICS Centurion XVI.

Cuando se presiona *ACEPTAR*, los datos del archivo externo se leerán en STATGRAPHICS Centurion XVI. Cada columna será observada en una columna apropiada con su tipo asignado. Los datos serán entonces leídos para ser analizados.

2.2.3 Transfiriendo datos utilizando Copiar y Pegar

El camino más fácil para transferir datos de otra aplicación a STATGRAPHICS Centurion XVI es a través del portapapeles de Windows. Por ejemplo, si los datos residen en un archivo Excel, se puede iniciar Excel y copiar los datos al portapapeles seleccionando los datos en Excel y eligiendo *Copiar* e el menú *Edición de Excel*. Volviendo a STATGRAPHICS, los datos pueden ser pegados directamente en la hoja de datos de STATGRAPHICS Centurion XVI seleccionando

Pegar en el menú *Edición* de STATGRAPHICS. Cuando los datos son pegados en una columna de la hoja de datos, STATGRAPHICS Centurion XVI automáticamente chequea los datos y selecciona un tipo apropiado para la columna.

Cuando se copian y pegan datos, los nombres de las columnas y los comentarios pueden ser transferidos también. Incluir los nombres de columna y comentarios en Excel cuando se copien los datos al portapapeles. En STATGRAPHICS Centurion XVI, haga clic en la fila de la cabecera de la hoja de datos de STATGRAPHICS Centurion XVI antes de seleccionar *Pegar*. La información del principio del portapapeles se pegará entonces en la fila(s) de la cabecera.

2.2.4 Consultando una base de datos ODBC

STATGRAPHICS Centurion XVI también permite leer datos de Oracle, Access, u otra base de datos utilizando ODBC. Para acceder a los datos de la base de datos, primero seleccione *Archivo – Abrir – Abrir origen de datos*. Entonces seleccione *Consulta de base de datos* en el cuadro de diálogo inicial:



Figura 2-8. Cuadro de diálogo *Abrir origen de datos*

Se mostrará una secuencia de cuadros de diálogo adicionales en las cuales es posible:

1. Seleccionar el nombre de la base de datos a leer.
2. Seleccionar los campos que van a ser transferidos.
3. Especificar un filtro para limitar los registros que serán recuperados.
4. Especificar un orden de salida de los resultados.

Se construye una consulta SQL y los resultados se situarán en la hoja de datos activa de STATGRAPHICS Centurion XVI. Se puede encontrar información detallada sobre la construcción de consultas ODBC en el documento PDF titulado *Archivos de datos y StatLink*.

2.3 Manipulando Datos

Una vez que los datos han sido situados en la hoja de datos de STATGRAPHICS Centurion XVI, pueden ser manipulados por varios caminos distintos:

1. Los datos pueden ser copiados y pegados en otras localizaciones.
2. Pueden ser creadas columnas adicionales a partir de columnas existentes.
3. Los datos pueden ser transformados mediante expresiones algebraicas o funciones matemáticas.
4. La hoja de datos puede ser ordenada según una o más columnas.
5. Los valores de los datos pueden ser grabados para formar grupos o por otras razones.
6. Los datos que se extienden a lo largo de múltiples columnas pueden ser recolocados en una columna simple si se requiere para un procedimiento estadístico.

Estas importantes operaciones se describen a continuación.

2.3.1 Copiando y pegando datos

La hoja de datos de STATGRAPHICS Centurion XVI soporta la mayoría de las operaciones típicas en las hojas de cálculo, incluyendo *cortar*, *copiar*, *pegar*, *insertar* y *borrar*. El único hecho importante a recordar cuando se usan estas operaciones es que cada columna tiene un tipo especificado. Si por descuido se pegan datos carácter en columnas numéricas, STATGRAPHICS Centurion XVI cambiará el tipo de la columna para acomodarla a los nuevos datos. Si hay alguna duda acerca de los tipos de las columnas, haga clic en la cabecera de la columna para mostrar el cuadro de diálogo *Modificar Columna*. Puede cambiar el tipo de la columna utilizando este cuadro de diálogo.

2.3.2 Creando nuevas variables de columnas existentes

STATGRAPHICS Centurion XVI tiene una amplia variedad de operadores para ayudar en los cálculos y transformaciones matemáticas. Uno de los usos más importantes de estos operadores de análisis de datos es crear nuevas variables basadas en columnas existentes. En STATGRAPHICS Centurion XVI pueden crearse nuevas variables:

1. “Sobre la marcha” directamente con los campos de datos en las entradas de datos de los cuadros de diálogo, sin salvar la variable en la hoja de datos.
2. Creando una nueva columna en una de las 26 hojas de datos en el libro de datos.

Por ejemplo, supongamos que deseamos información acerca del ratio millas por galón conduciendo en ciudad contra millas por galón conduciendo en carretera para cada automóvil de los 93 del archivo de datos. Este archivo contiene 2 columnas separadas, una llamada *MPG City* y otra llamada *MPG Highway*. Para resumir la distribución de los ratios, puede seleccionar el procedimiento *Análisis de una variable* y especificar el ratio directamente en el campo *Datos* del cuadro de diálogo de entrada de datos:

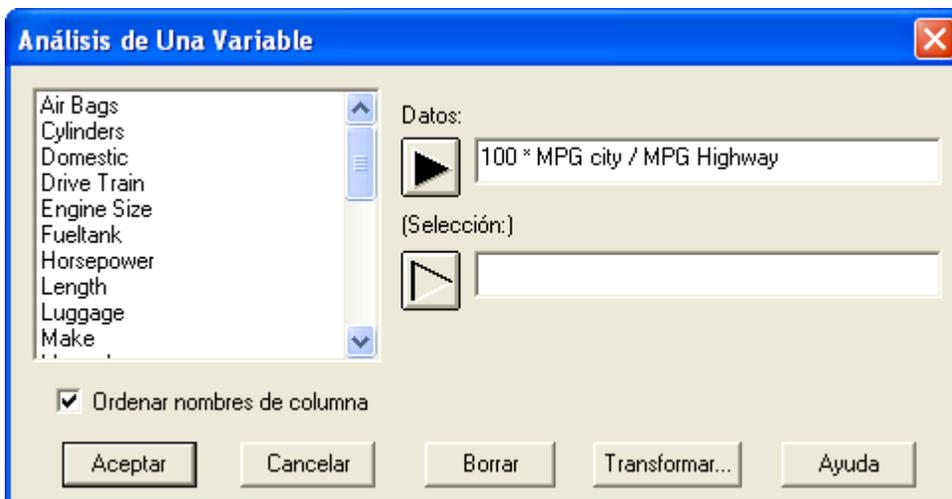


Figura 2-9. Creando una transformación “Sobre la marcha”

Cuando se presiona ACEPTAR, se generará un análisis para 100 veces el ratio, sin cambiar los datos en la hoja de datos:

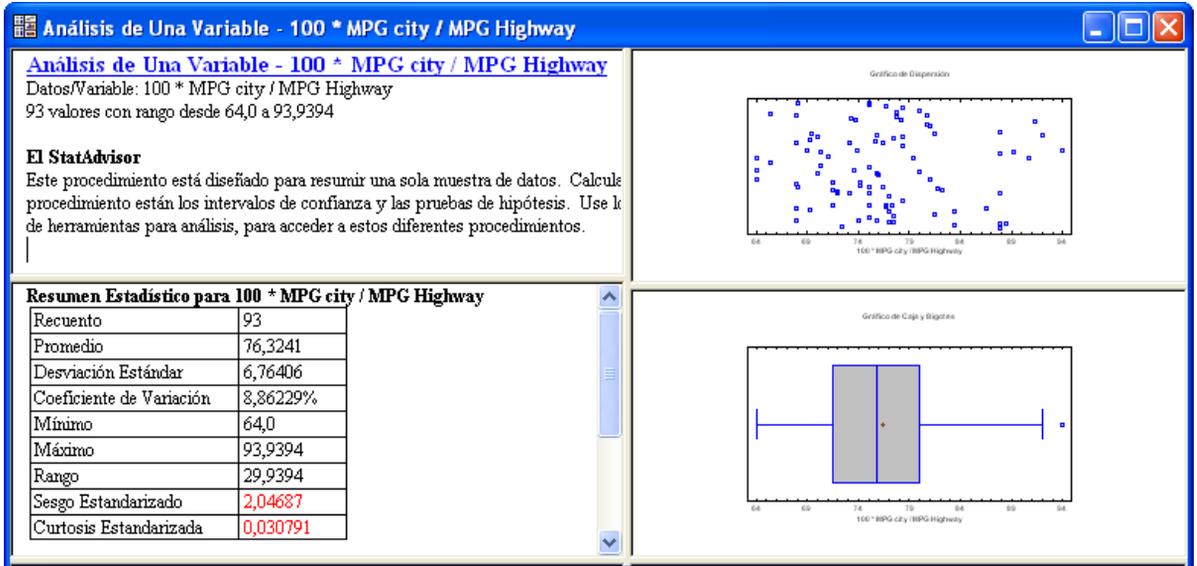


Figura 2-10. Transformando datos en Análisis de una variable

El ratio medio es aproximadamente 76.3%, variando desde el 64.0% por abajo hasta el 93.9% por arriba. La posibilidad de realizar análisis sin modificar las hojas de datos es muy importante para facilitar la exploración de datos.

Si lo desea, puede crear una nueva columna en la hoja de datos conteniendo los valores transformados. Por ejemplo, puede volver a la ventana que contiene los datos de los 93 coches y hacer doble clic en la cabecera de la columna etiquetada *Col_27*. El cuadro de diálogo *Modificar columna* puede ser utilizado para definir una nueva variable de tipo *fórmula* con la transformación deseada:



Figura 2-11. Creando una columna mediante fórmula

Esto creará una nueva columna cuyos valores son calculados a partir de las dos columnas originales que contienen los datos de las millas por galón. Las columnas de *Fórmula* se muestran en la hoja de datos utilizando una escala de grises, siempre que se calculen automáticamente a partir de otras columnas:

	Luggage	Weight	Domestic	MPG Ratio	Col_28
	cu. ft.	pounds	1=U.S. manufacturer		
1	11	2705	0	80,6451612903	
2	15	3560	0	72	
3	14	3375	0	76,9230769231	
4	17	3405	0	73,0769230769	
5	13	3640	0	73,3333333333	
6	16	2880	1	70,9677419355	
7	17	3470	1	67,8571428571	
8	21	4105	1	64	
9	14	3495	1	70,3703703704	
10	18	3620	1	64	

Figura 2-12. Apariencia de una columna Fórmula en una hoja de datos

Si los valores de las columnas *MPG City* o *MPG Highway* cambian, *MPG Ratio* se recalculará automáticamente para reflejar los cambios.

NOTA: El recálculo de columnas *Fórmula* no ocurre normalmente hasta que los datos de estas columnas se necesiten para un cálculo o se guarden o impriman. Puede especificar un recálculo para que ocurra inmediatamente seleccionando *Actualizar fórmulas* en el menú *Edición*.

2.3.3 Transformando datos

STATGRAPHICS Centurion XVI también contiene un gran número de funciones matemáticas que pueden ser utilizadas para transformar datos existentes. Como en la creación de variables, las transformaciones pueden ser hechas directamente en los campos de entrada de datos del cuadro de diálogo o creando nuevas columnas en la hoja de datos.

Por ejemplo, supongamos que deseamos representar las millas por galón consumidas por un automóvil frente al logaritmo natural del peso de los vehículos. Seleccionando el procedimiento *Gráfico X-Y* del menú principal se muestra el siguiente cuadro de diálogo de entrada de datos:

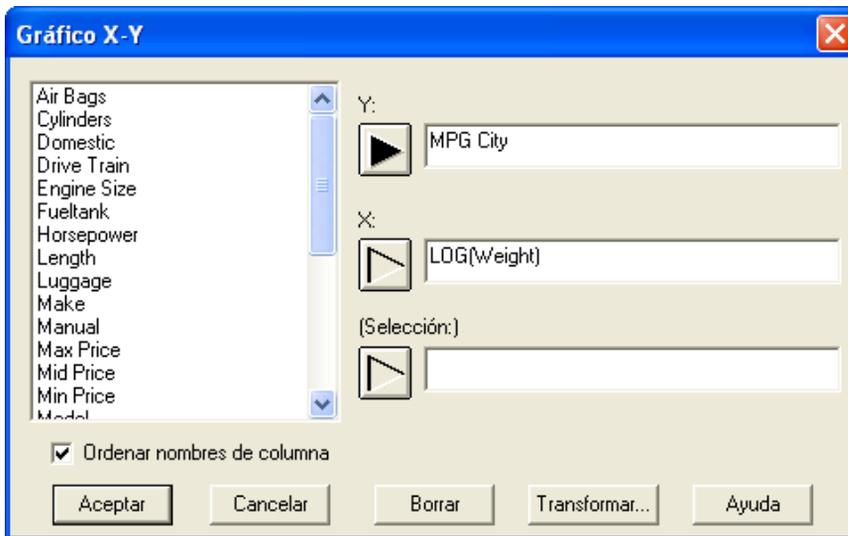


Figura 2-13. Transformando datos en un cuadro de diálogo de entrada de datos

En lugar de teclear el nombre de una columna en un campo de datos, puede teclear una expresión STATGRAPHICS Centurion. Las expresiones de STATGRAPHICS Centurion son fórmulas que operan en los datos utilizando símbolos algebraicos y operadores especiales. Está

disponible una amplia variedad de operadores, como se describe en el documento PDF titulado *STATGRAPHICS Operators*. La tabla siguiente muestra los operadores más utilizados habitualmente:

<i>Operador</i>	<i>Uso</i>	<i>Ejemplo</i>
+	Suma	X+100
-	Resta	X-100
/	División	X/100
*	Multipliación	X*100
^	Exponenciación	X^2
ABS	Valor absoluto	ABS(X)
AVG	Media	AVG(X)
DIFF	Diferencia hacia atrás	DIFF(X)
EXP	Función exponencial	EXP(10)
LAG	Paso para <i>k</i> periodos	LAG(X,k)
LOG	Logaritmo natural	LOG(X)
LOG10	Logaritmo en base 10	LOG10(X)
MAX	Máximo	MAX(X)
MIN	Mínimo	MIN(X)
SD	Desviación típica	SD(X)
SQRT	Raíz cuadrada	SQRT(X)
STANDARDIZE	Conversión a puntuaciones Z	STANDARDIZE(X)

Figura 2-14. Operadores de STATGRAPHICS utilizados habitualmente

Cuando se construye una expresión de STATGRAPHICS Centurion, pueden combinarse múltiples operadores utilizando las reglas formales de precedencia algebraica. Por ejemplo, la siguiente expresión convierte cada valor de la columna de nombre *Weight* en una fracción de la distancia entre el máximo y el mínimo para todos los automóviles:

$$(Weight - MIN(Weight)) / (MAX(Weight) - MIN(Weight))$$

El paréntesis es necesario para asegurar que las restas se hacen antes que la división. Las expresiones no son sensibles a mayúsculas y la inclusión de espacios en blanco no es relevante.

Cada cuadro de diálogo de entrada de datos incluye un botón etiquetado *Transformar*, como en la Figura 2-13. Este botón puede ser utilizado para ayudar a crear expresiones STATGRAPHICS Centurion, si no recuerda los operadores a utilizar. Si sitúa el cursor en un campo de datos y presiona *Transformar*, se muestra un cuadro de diálogo similar al siguiente:

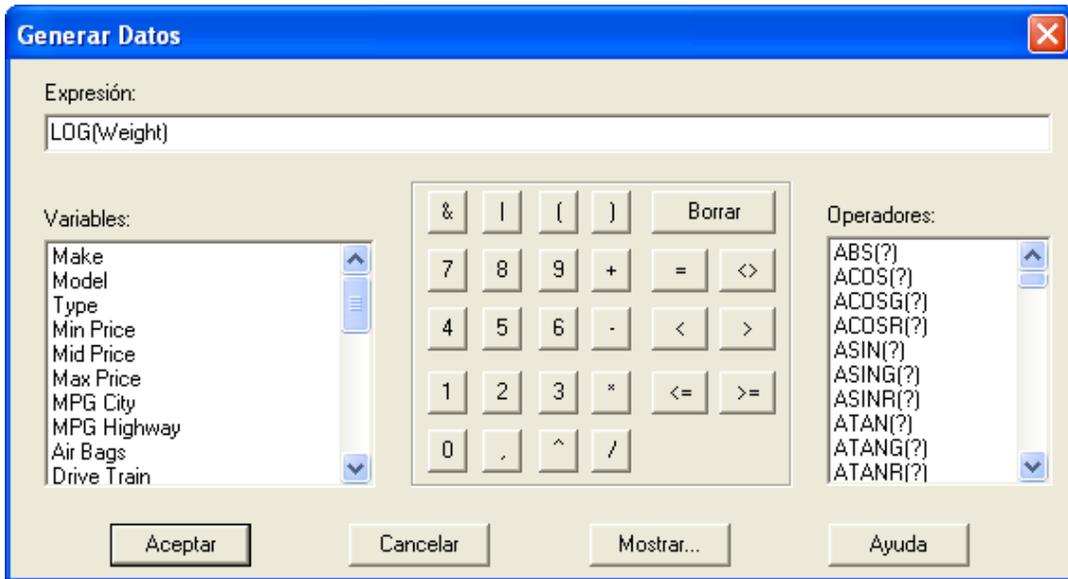


Figura 2-15. Cuadro de diálogo mostrado por el botón transformar

A la izquierda de la pantalla hay una lista con todos los operadores de STATGRAPHICS Centurion, con una indicación del número de argumentos que deben ser cumplimentados. Haciendo clic en el nombre de un operador, este se sitúa en el campo *Expresión*. Después puede reemplazar los espacios entre comillas con nombres de columnas o números. Es posible presionar el botón *Mostrar* para ver el primero de los valores generados por la expresión, o presionar el botón *Aceptar* para tener la expresión introducida en el cuadro de diálogo de entrada de datos.

NOTA: No necesita utilizar el botón *Transformar* si quiere teclear la expresión directamente en el cuadro de diálogo de entrada de datos.

Una vez que la transformación ha sido especificada en el cuadro de diálogo de entrada de datos, como se muestra en la *figura 2-13*, será utilizada cuando se ejecute el procedimiento:

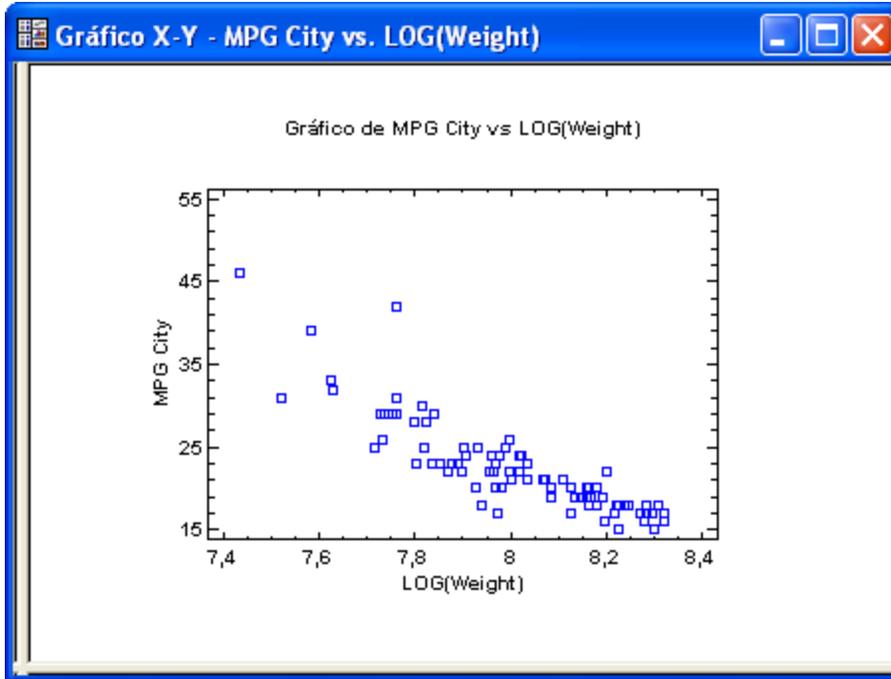


Figura 2-16. Procedimiento Gráfico X-Y utilizando valores transformados de Weight

Los operadores de STATGRAPHICS Centurion pueden también ser utilizados cuando se crean columnas *fórmula*, similar a la ilustración en la sección precedente.

2.3.4 Ordenando datos

El contenido de la hoja de datos puede ser ordenado realizando la columna o columnas que van a ser utilizadas para definir el orden y seleccionando *Ordenar datos* en el menú *Edición*. Por ejemplo, para ordenar los datos de los 93 coches del archivo de acuerdo a las millas por galón, realizar las columnas de nombres *MPG City* y *MPG Highway* y seleccionar *Ordenar datos*. Se muestra el siguiente cuadro de diálogo:



Figura 2-17. Cuadro de diálogo Opciones de ordenación

Debe especificar cada una de las dos columnas en que se basa la ordenación, así como la dirección del orden. Ordenado por *MPG City* y después por *MPG Highway* se ordena primero por millas por galón conduciendo en ciudad y después se ordena por millas por galón conduciendo en carretera para automóviles con el mismo valor de *MPG City*:

	Mid Price	Max Price	MPG City	MPG Highway
	average of min and max prices in \$1,000	price for a premium version in \$1,000	miles per gallon in city driving	miles per gallon in highway driving
1	16,6	18,6	15	20
2	19,9	25,3	15	20
3	23,7	24,9	16	25
4	34,7	36,3	16	25
5	40,1	42,7	16	25
6	19	24,4	17	21
7	19,7	22,7	17	21
8	47,9	50,4	17	22
9	19,1	21,5	17	23
10	38	41,5	17	25
11	32,5	32,5	17	25
12	18,8	19,6	17	26
13	34,3	35,3	17	26
14	22,7	26,6	18	22

Figura 2-18. Archivo 93cars.sgd después de la ordenación

NOTA: El procedimiento estadístico no requiere que ordene los datos antes, ya que dicha ordenación será automática si es necesario. Además, el archivo de datos en disco no cambia cuando se ejecuta una ordenación hasta que no vuelva a guardar los datos. La ordenación sólo afecta a las filas de la hoja de datos que se muestra actualmente (y no al resto de las hojas).

2.3.5 Recodificando datos

En algunas ocasiones es conveniente recodificar los datos, bien sea agrupando en grupos similares o asignando nuevas etiquetas. Para recodificar una columna de datos, primero haga clic en la cabecera de la columna que va a ser recodificada. Después seleccione *Recodificar datos* en el menú *Edición*. Se mostrará el siguiente cuadro de diálogo:

Límite Inferior:	Límite Superior:	Nuevo Valor:
0	0	Extranjero
1	1	U.S.

Condiciones de los Límites

- Inferior <= Valor <= Superior
- Inferior <= Valor < Superior
- Inferior < Valor <= Superior
- Inferior < Valor < Superior

Sin condición

- Dejar igual
- Establecer como faltante

Extrapolar

Aceptar Cancelar Ayuda

Figura 2-19. Cuadro de diálogo para Recodificación de datos

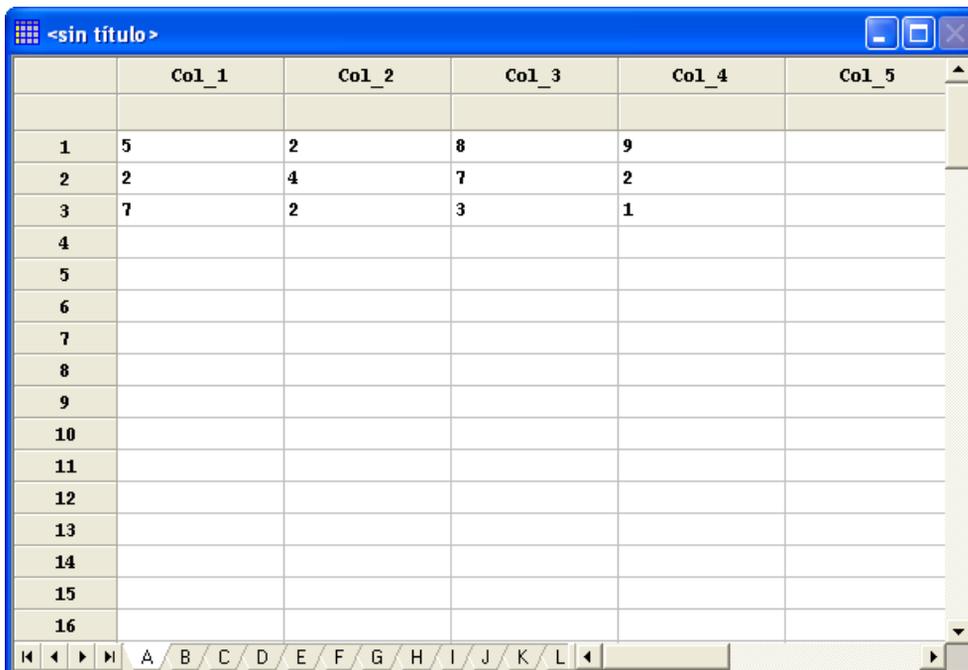
Por ejemplo, la columna de nombre *Domestic* en el archivo *93cars* contiene un 1 para cada coche fabricado en U.S. y un 0 para el resto de los coches. Para cambiar todos los ceros de la columna

a “Extranjero” y todos los unos a “U.S.”, se utilizará el cuadro de diálogo mostrado arriba. Hasta 7 rangos de valores se pueden especificar a la vez para recodificar.

El documento PDF titulado *Menú Editar* tiene una discusión detallada de los dos ejemplos de recodificación.

2.3.6 Combinando múltiples columnas

Muchos procedimientos estadísticos en STATGRAPHICS Centurion XVI esperan los datos para analizar en una columna simple. Algunas veces no se dispone de los datos en tal formato. Como ejemplo simple, supongamos que tenemos una muestra de 12 observaciones, preparada en cuatro columnas como sigue:



	Col_1	Col_2	Col_3	Col_4	Col_5
1	5	2	8	9	
2	2	4	7	2	
3	7	2	3	1	
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					

Figura 2-20. Muestra de datos en múltiples columnas

Para situar los datos en una columna única, deberán ejecutarse múltiples operaciones de cortar y pegar. Una solución simple es utilizar el procedimiento *Combinar columnas*, definido bajo *Editar* en el menú principal. Este procedimiento muestra primero el cuadro de diálogo de entrada de datos requiriendo los nombres de las columnas que contienen los datos:

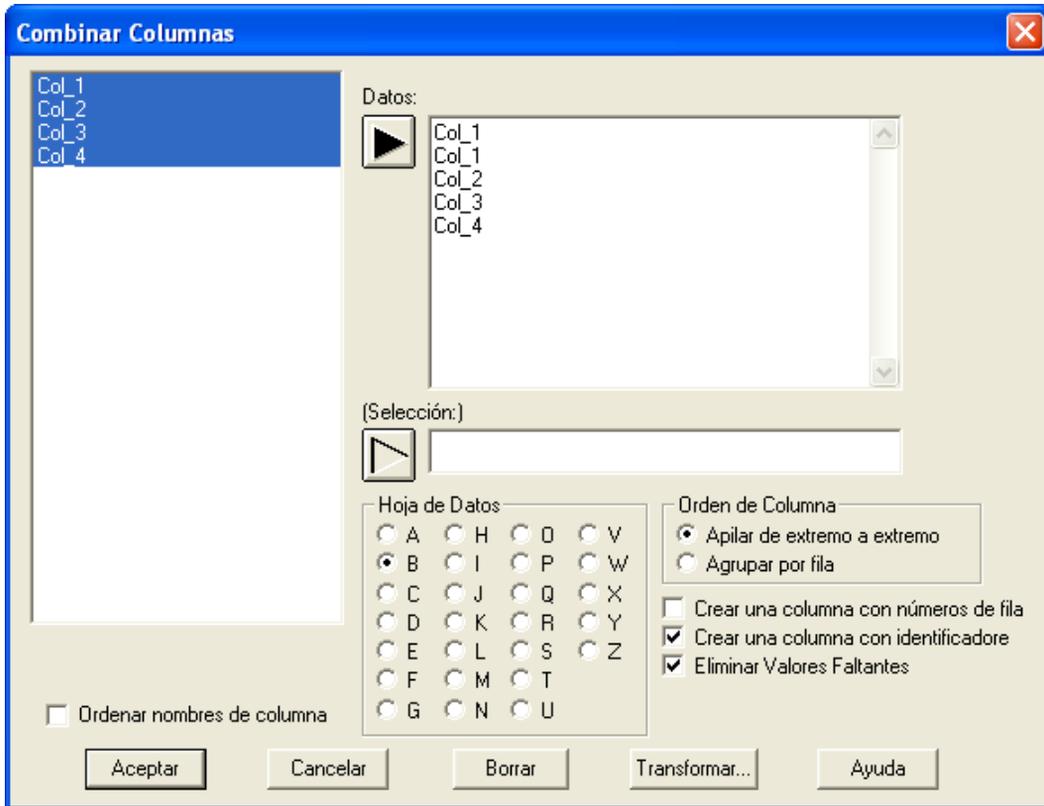


Figura 2-21. Cuadro de diálogo de entrada de datos para combinar columnas

Contiene los siguientes campos:

1. **Datos** – columnas que van a ser combinadas.
2. **Seleccionar** – subconjunto estándar de campos de selección para elegir un subconjunto de filas.
3. **Situar en la hoja de datos** – datos de la hoja de destino para la combinación de datos.
4. **Columna de ordenación** – cuando se apilan los datos por columnas (una columna después de otra) o por filas.
5. **Crear columna con números de fila** – cuando se crea una segunda columna identificando la fila original que contiene cada valor de dato.

6. **Crear columnas con identificadores** – cuando se crea una segunda columna identificando la columna original conteniendo cada valor de datos.
7. **Cambiar valores desaparecidos** – cuando se saltan todas las celdas blancas en vez de dejarlas con su marcador de posición.

Cuando se presiona ACEPTAR, los datos se combinarán en una columna simple como se muestra a continuación:

	Datos	Identificador	Col_3	Col_4
1	5	Col_1		
2	2	Col_1		
3	7	Col_1		
4	2	Col_2		
5	4	Col_2		
6	2	Col_2		
7	8	Col_3		
8	7	Col_3		
9	3	Col_3		
10	9	Col_4		
11	2	Col_4		
12	1	Col_4		
13				
14				
15				
16				

Figura 2-22. Datos combinados en una columna simple

2.4 Generando datos

STATGRAPHICS Centurion XVI tiene la posibilidad de generar datos y situarlos en columnas de la hoja de datos. Esta sección describe dos ejemplos importantes:

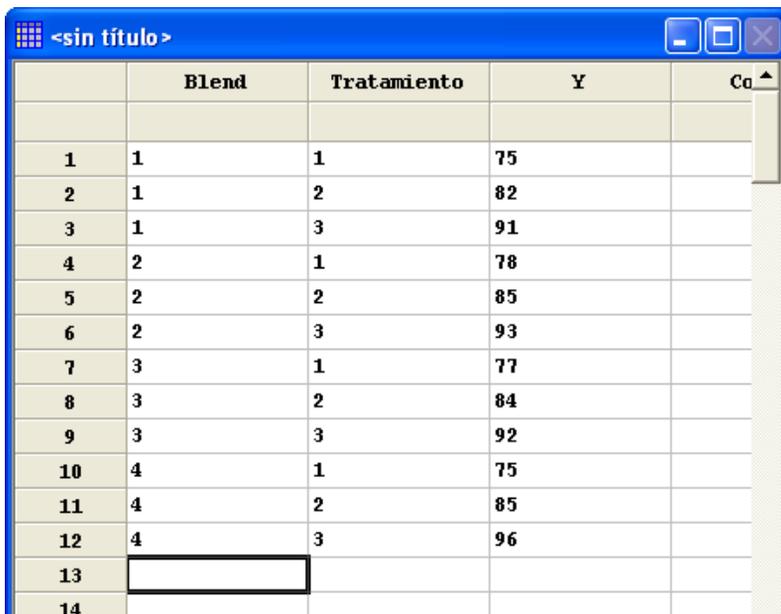
1. Generando datos con patrones simples.
2. Generando números aleatorios.

2.4.1 Generando datos con patrones

Varios procedimientos en STATGRAPHICS Centurion XVI, particularmente los que ejecutan un análisis de la varianza, esperan que los datos para ser analizados se sitúen en una columna simple de la hoja de datos, junto con una o más columnas de códigos identificando los factores explicativos. Por ejemplo, considerar los datos de la siguiente tabla de doble entrada:

<i>Blend</i>	<i>Treatment 1</i>	<i>Treatment 2</i>	<i>Treatment 3</i>
1	75	82	91
2	78	85	93
3	77	84	92
4	75	85	96

Para analizar los datos utilizando el procedimiento *ANOVA Multifactorial*, se necesita situarlos en una hoja de datos con el formato siguiente:



The screenshot shows a window titled "<sin título>" with a grid icon and window controls. The table inside has the following structure:

	Blend	Tratamiento	Y	Co
1	1	1	75	
2	1	2	82	
3	1	3	91	
4	2	1	78	
5	2	2	85	
6	2	3	93	
7	3	1	77	
8	3	2	84	
9	3	3	92	
10	4	1	75	
11	4	2	85	
12	4	3	96	
13				
14				

Figura 2-23. Estructura deseada de los datos

Las dos primeras columnas indican los niveles de los factores correspondientes a cada nivel de datos. La tercera columna contiene todas las observaciones.

Para crear tal fichero, la solución más fácil es a menudo teclear las dos primeras columnas. Sin embargo, si las columnas siguen patrones simples, puede generar tales columnas utilizando

operadores especiales de STATGRAPHICS Centurion. Por ejemplo, los números de la columna “blend” pueden ser generados haciendo clic en la cabecera de la columna #1 y seleccionando *Generar Datos* en el menú *Edición*. Se muestra el siguiente cuadro de diálogo, en el cual se ha introducido una expresión:

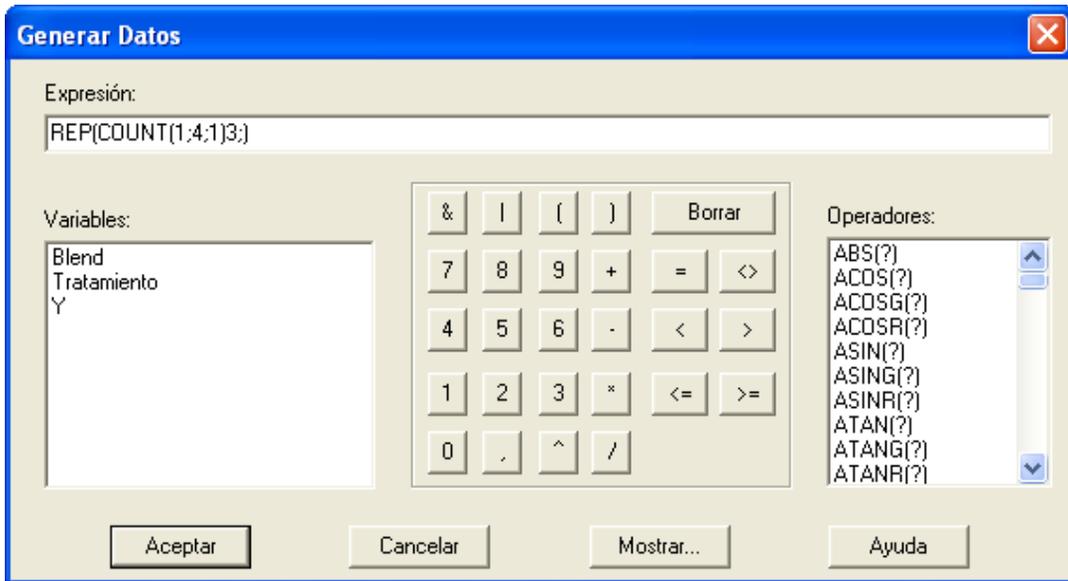


Figura 2-24. Generando números Blend

La opción *Generar datos* evalúa una expresión STATGRAPHICS Centurion y sitúa los resultados en la columna seleccionada. En la expresión que se muestra arriba, se utilizan dos operadores importantes:

COUNT(from, to, by) – genera valores comenzando en *from* y finalizando en *to*, en intervalos iguales a *by*. *COUNT(1,4,1)* genera los enteros 1, 2, 3 y 4.

REP(X, repetitions) – repite cada valor en *X repetitions* veces, en grupos. En este caso, cada entero entre 1 y 4 se repite 3 veces.

Los números de tratamiento pueden ser generados de una forma similar haciendo clic en la cabecera de la columna #2, seleccionando *Generar Datos* del menú *Edición* e introduciendo lo siguiente:

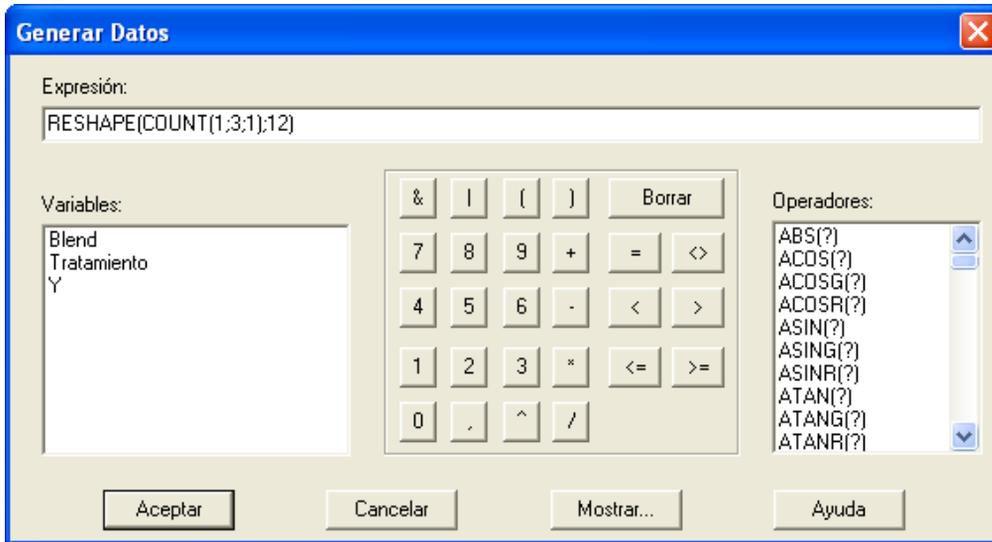


Figura 2-25. Generando números de tratamiento

Estas expresiones utilizan un operador condicional:

$RESHAPE(X, size)$ – repite los valores en X en forma circular hasta que $size$ valores han sido generados. En este caso, la secuencia 1, 2, 3 se repite 4 veces.

Estos generadores de patrones pueden ayudar cuando los archivos de datos son muy grandes.

2.4.2 Generando números aleatorios

Los números aleatorios pueden ser generados en STATGRAPHICS Centurion XVI por dos caminos:

1. Si los números provienen de una distribución exponencial, gamma, lognormal, normal, uniforme, o Weibull, pueden ser generados en una hoja de datos haciendo clic en una cabecera de columna, seleccionando *Generar datos* en el menú *Edición* e introduciendo las expresiones apropiadas de STATGRAPHICS Centurion.
2. Para otras distribuciones, los números aleatorios deben ser generados con el procedimiento *Distribuciones de probabilidad*.

Como ejemplo, supongamos que se necesitan 100 números aleatorios de una distribución normal de media 20 y desviación típica igual a 2. Hacer clic en la cabecera de una columna vacía

en una hoja de datos para seleccionar la columna. Seleccionar *Generar datos* en el menú *Edición* y completar el cuadro de diálogo como se muestra a continuación:

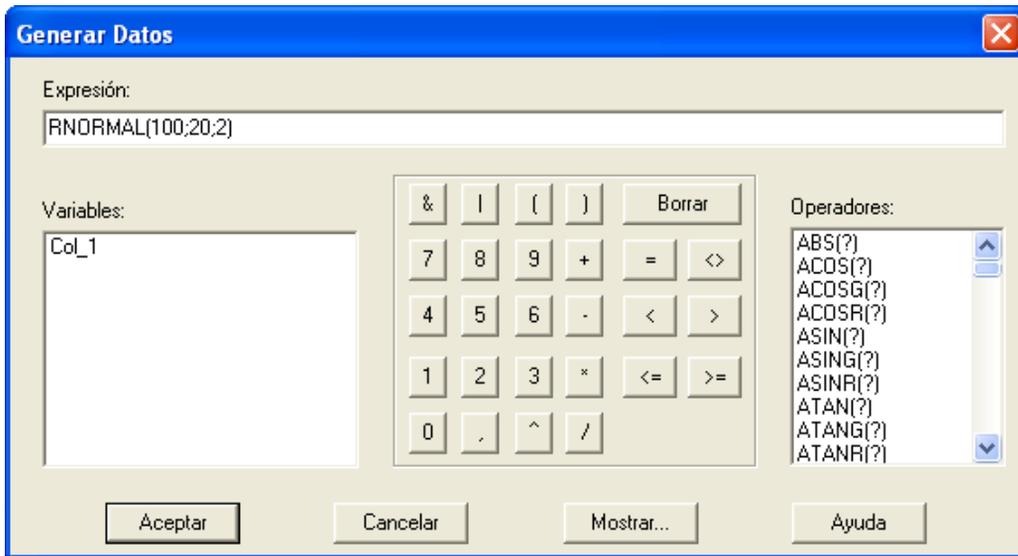


Figura 2-26. Generando números aleatorios para una distribución normal

La sintaxis del operador RNORMAL es la siguiente:

$RNORMAL(n, \mu, \sigma)$ – genera n números pseudo-aleatorios de una distribución normal con media μ y desviación típica σ .

Presionar *Aceptar* para generar los números aleatorios y situarlos en la columna seleccionada.

La sintaxis de otros generadores de números aleatorios está contenida en el documento PDF titulado *STATGRAPHICS Centurion Operators*.

2.5 Propiedades del libro de datos

Este capítulo describe aspectos importantes del manejo de datos con STATGRAPHICS Centurion XVI. En particular, se muestra como leer datos de archivos y bases de datos y cómo manipular los datos una vez que han sido situados en la hoja de datos de STATGRAPHICS Centurion XVI. En un momento dado, puede mostrarse el estado de las hojas de datos activando la ventana del libro de datos y seleccionando *Propiedades del libro de datos* en el menú *Edición* o seleccionando *StatLink* en el menú *Archivo*:

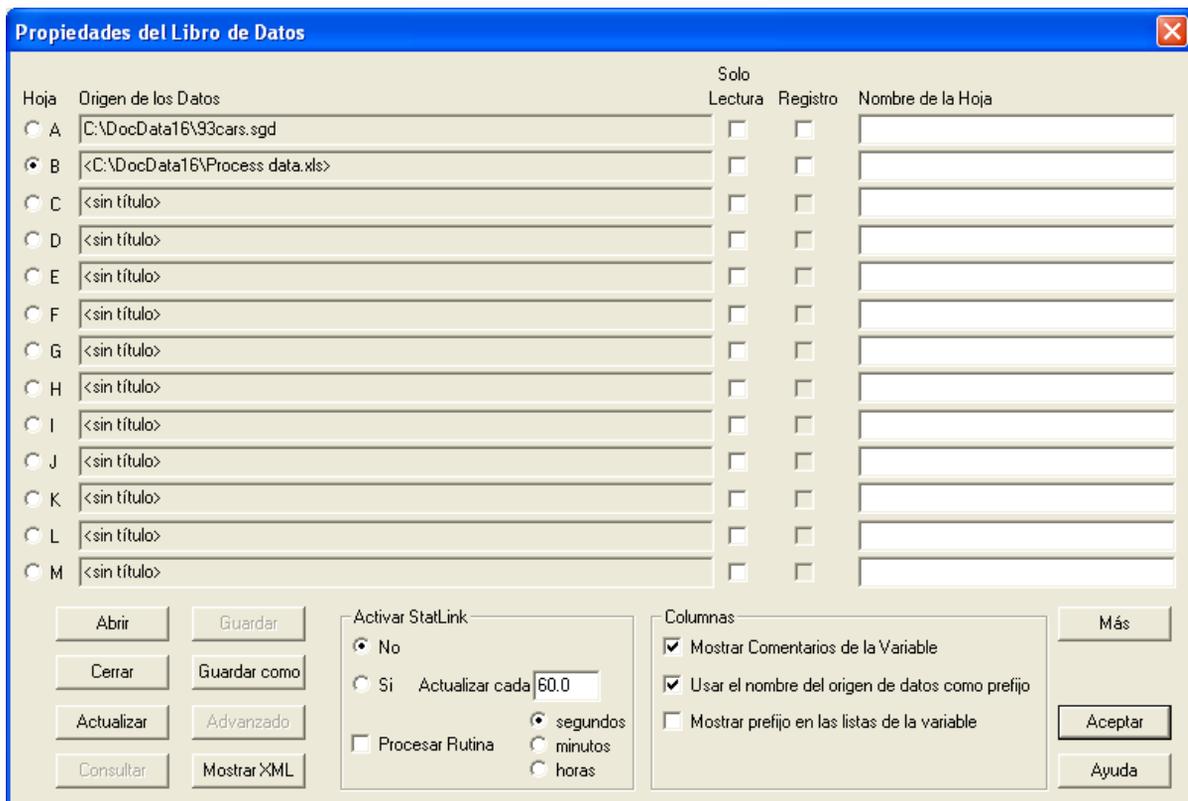


Figura 2-27. Cuadro de diálogo de Propiedades del libro de datos

Este cuadro de diálogo muestra el origen actual de los datos de cada hoja. Si se desea, las hojas de datos pueden hacerse de sólo lectura para no ser cambiadas sin darnos cuenta. Es también posible observar el origen de datos (releerlo) en intervalos regulares y tener los procedimientos estadísticos actualizados automáticamente. Estas importantes características se describen en el capítulo 5.

2.6 Visor de datos

Se ha añadido un Nuevo procedimiento para ver archivos de datos en STATGRAPHICS Centurion XVI. El procedimiento, accesible seleccionando *Visor de datos* en el menú *Herramientas*, produce un resumen del número de valores únicos no desaparecidos y del máximo y el mínimo de las variables seleccionadas:

Visor de Datos						
Número de columnas: 26						
Número de filas: 93						
Número de casos completos: 82						
Columna			Nodesaparecido	Unico		
Nombre	Comentario	Tipo	Valores	Valores	Mínimo	Máximo
Air Bags	0=none, 1=driver only, 2=driver and passenger	Numérico	93	3	0	2,0
Cylinders		Numérico	92	5	3,0	8,0
Domestic	1=U.S. manufacturer	Numérico	93	2	0	1,0
Drive Train		Texto	93	3		
Engine Size	liters	Numérico	93	25	1,0	5,7
Fuel tank	gallons	Numérico	93	38	9,2	27,0
Horsepower	maximum	Numérico	93	57	55,0	300,0
Length	inches	Numérico	93	51	141,0	219,0
Luggage	cu. ft.	Numérico	82	16	6,0	22,0
Make		Texto	93	32		
Manual	0=no, 1=yes	Numérico	93	2	0	1,0
Max Price	price for a premium version in \$1,000	Numérico	93	79	7,9	80,0
Mid Price	average of min and max prices in \$1,000	Numérico	93	81	7,4	61,9
Min Price	price for basic version in \$1,000	Numérico	93	79	6,7	45,4
Model		Texto	93	93		
MPG City	miles per gallon in city driving	Numérico	93	21	15,0	46,0
MPG Highway	miles per gallon in highway driving	Numérico	93	22	20,0	50,0
Passengers	persons	Numérico	93	6	2,0	8,0
Rear seat	inches	Numérico	91	24	19,0	36,0
Revs per Mile	revs per mile in highest gear	Numérico	93	78	1320,0	3755,0
RPM	revs per minute at maximum horsepower	Numérico	93	24	3800,0	6500,0
Type		Texto	93	6		
U Turn Space	feet	Numérico	93	14	32,0	45,0
Weight	pounds	Numérico	93	81	1695,0	4105,0
Wheelbase	inches	Numérico	93	27	90,0	119,0
Width	inches	Numérico	93	16	60,0	78,0

Figura 2-28. Cuadro de diálogo de propiedades del libro de datos

Ejecutando análisis estadísticos

Generando un análisis, seleccionando tablas y gráficos adicionales, seleccionando opciones, cambiando la entrada de datos y guardando los resultados.

Hay alrededor de 160 procedimientos estadísticos en el menú principal de STATGRAPHICS Centurion XVI. Cada selección accede a un procedimiento estadístico. Todos los procedimientos trabajan según el mismo camino básico:

1. Cuando se selecciona un análisis del menú, se muestra su *cuadro de diálogo de entrada de datos*. Los campos en este cuadro de diálogo se utilizan para especificar las variables que va a ser analizadas.
2. Si el procedimiento seleccionado tiene opciones que afectan a tablas y gráficos, se muestra el cuadro de diálogo *Opciones de análisis* para seleccionar las características deseadas.
3. Si el procedimiento seleccionado tiene más de una tabla simple y más de un gráfico simple, se muestra el cuadro de diálogo *Tablas y Gráficos* en la cual se puede seleccionar la salida deseada.
4. Se leen y analizan los datos especificados, y se crea una nueva *ventana de análisis*.
5. Las opciones seleccionadas pueden cambiarse utilizando el botón *Opciones de análisis* de la barra de herramientas de análisis y todas las tablas y gráficos de la ventana de análisis podrán actualizarse.
6. Si lo desea, pueden requerirse tablas y gráficos adicionales con el botón *Tablas y Gráficos* de la barra de herramientas de análisis.

7. Se pueden modificar tablas y gráficos individuales maximizando el panel correspondiente seleccionando *Opciones de panel* en la barra de herramientas de análisis.
8. Para gráficos, el título por defecto, escala, tipos de puntos, fuentes, etc. pueden cambiarse haciendo doble clic en el gráfico para maximizarlo seleccionando *Opciones gráficas* en la barra de herramientas de análisis.
9. Las tablas y gráficos pueden ser impresos, publicados como archivos HTML, copiados a otras aplicaciones tales como Microsoft PowerPoint, o guardados en StatReporter.
10. Los resultados numéricos pueden ser guardados en columnas de una hoja de datos utilizando el botón *Guardar resultados* en la barra de herramientas de análisis.
11. El análisis completo puede ser guardado en disco como un *StatFolio* para recuperarlo después.

En este capítulo se describe en detalle un análisis típico. La finalidad del análisis es construir un modelo estadístico que explique las millas por galón recorridas conduciendo en ciudad para $n = 93$ automóviles del archivo *93cars.sgd* en función de su peso. A continuación se muestra un gráfico de dispersión de los datos:

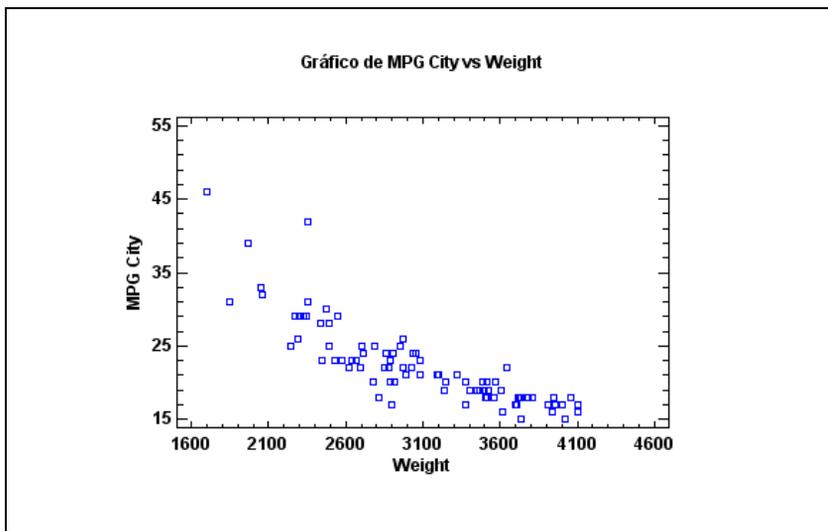


Figura 3-1. Gráfico X-Y de las millas por galón conduciendo en ciudad contra el peso en Weight en libras

Como era de esperar, las millas por galón recorridas están correlacionadas negativamente con el peso. Son evidentes algunas no linealidades en la relación, y al menos un punto parece ser un potencial atípico.

El procedimiento primario en STATGRAPHICS Centurion XVI para ajustar un modelo estadístico que relaciona dos variables es el procedimiento *Regresión Simple*. Este procedimiento ajusta modelos lineales y no lineales. Los modelos simples relacionando una variable dependiente Y con una variable independiente X consisten en una línea recta de la forma

$$Y = a + b X$$

donde b es la pendiente de la línea y a es la ordenada en el origen. Los modelos curvilíneos tales como el modelo exponencial

$$Y = \exp(a + b X)$$

pueden utilizarse si la relación es no lineal.

3.1 Cuadro de diálogo de entrada de datos

El procedimiento *Regresión Simple* está localizado en el menú principal:

1. Si utiliza el menú clásico, bajo *Relacionar – Un Factor*.
2. Si utiliza el menú Seis Sigma, bajo *Mejorar – Análisis de la regresión – Un Factor*.

Se comienza mostrando el cuadro de diálogo de entrada de datos típico:

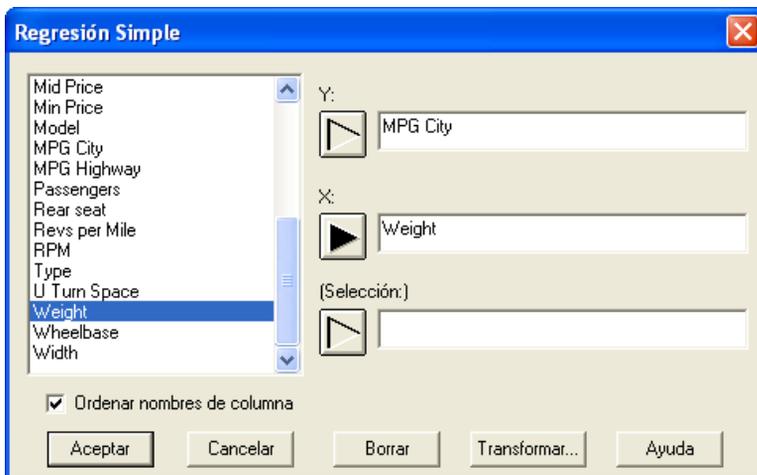


Figura 3-2. Cuadro de diálogo de entrada de datos de *Regresión simple*

Se requieren los dos primeros campos de entrada:

Y: Variable dependiente o respuesta.

X: Variable independiente o predictora.

En los campos de entrada de datos, puede introducir el nombre de la columna como *MPG City* o una expresión de STATGRAPHICS Centurion como *LOG(MPG City)*. Si hay más de una hoja de datos que contenga una columna de nombre indicado, será necesaria una indicación de la hoja de datos deseada para trabajar. Por ejemplo, si las hojas de datos A y B contienen una columna de nombre *Weight* y se quiere utilizar la columna en la hoja de datos A, se debe introducir su nombre como *A.Weight*.

El campo *Seleccionar* puede utilizarse para seleccionar un subconjunto de filas en la hoja de datos. Por ejemplo, si se introduce una sentencia tal como *FIRST(50)* en este campo, sólo se utilizarán las primeras 50 filas de la hoja de datos. Entradas típicas en el campo *Seleccionar* son:

<i>Entrada</i>	<i>Uso</i>	<i>Ejemplo</i>
FIRST(k)	Selecciona las primeras <i>k</i> filas.	FIRST(50)
LAST(k)	Selecciona las primeras <i>k</i> filas.	LAST(50)
ROWS(start,end)	Selecciona filas entre <i>start</i> y <i>end</i> , inclusive.	ROWS(21,70)
RANDOM(k)	Selecciona un conjunto aleatorio de <i>k</i> filas.	RANDOM(50)
column < value	Selecciona sólo las filas para las cuales <i>column</i> es menor que <i>value</i> .	Passengers < 5
column <= value	Selecciona sólo las filas para las cuales <i>column</i> es menor o igual que <i>value</i> .	Passengers <= 5
column > value	Selecciona sólo las filas para las cuales <i>column</i> es mayor que <i>value</i> .	Passengers > 5
column >= value	Selecciona sólo las filas para las cuales <i>column</i> es mayor o igual que <i>value</i> .	Passengers >= 5
column = value	Selecciona sólo las filas para las cuales <i>column</i> es igual a <i>value</i> .	Cylinders = 6
column <> value	Selecciona sólo las filas para las cuales <i>column</i> no es igual a <i>value</i> .	Cylinders <> 4
condition1 & condition2	Selecciona sólo las filas que cumplen ambas condiciones.	Cylinders = 6 & Make = "Ford"
condition1 condition2	Selecciona sólo las filas que cumplen al menos una de las dos condiciones.	Cylinders = 6 Make = "Ford"
binarycolumn	Selecciona sólo las filas para las cuales el valor en <i>binarycolumn</i> no es igual a 0.	Domestic

Figura 3-3. Entradas permitidas para el campo *seleccionar*

Cuando especificamos una condición envolviendo una variable no numérica, *value* debe ser incluida entre comillas dobles y es *sensible a mayúsculas*. Pueden ser combinadas condiciones múltiples combinando los símbolos AND (&) y OR (|).

Cada una de las entradas permitidas en el campo *Seleccionar* generan una secuencia de ceros y unos Booleanos, donde cero representa *FALSO* y uno representa *VERDADERO*. Cuando utilizamos el campo *Seleccionar* del cuadro de diálogo de entrada de datos, el resultado es la selección de todas las filas para las cuales la condición es *VERDADERA* y la exclusión de todas las filas para las cuales la condición es *FALSA*.

3.2 Ventana de análisis

Una vez que se han especificado los datos, se crea una nueva *ventana de análisis*:

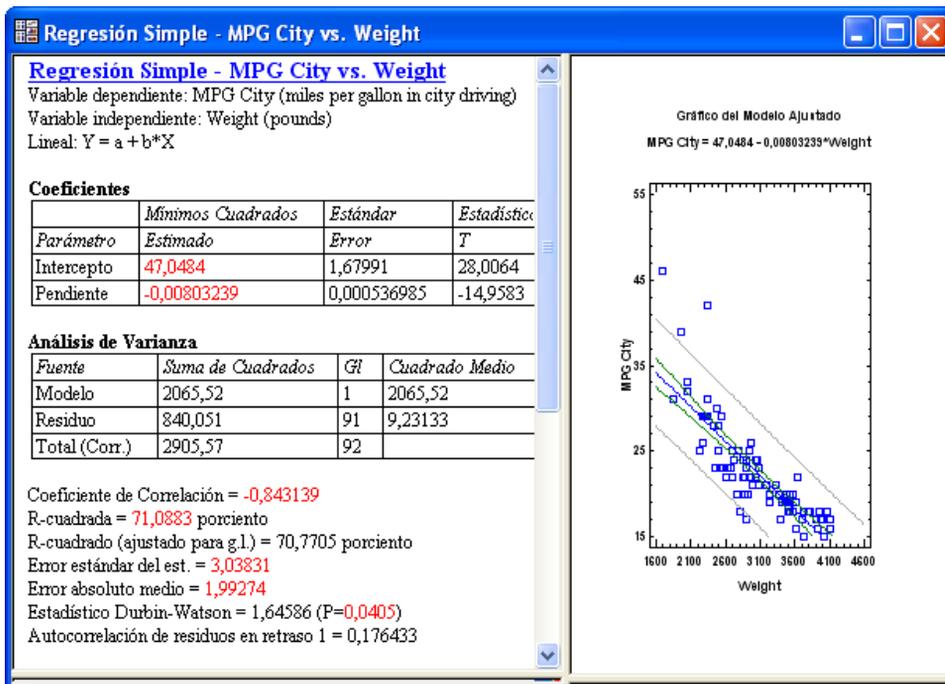


Figura 3-4. Ventana de análisis de Regresión simple

La ventana es una “ventana desgajada”, con múltiples *paneles* divididos por una barra de separación móvil. Las tablas se localizan a lo largo del lado izquierdo de la ventana, mientras que los gráficos se localizan a lo largo del lado derecho.

Puede maximizar la tabla o gráfico en un panel haciendo doble clic sobre él, en cuyo caso ocupará toda la ventana:

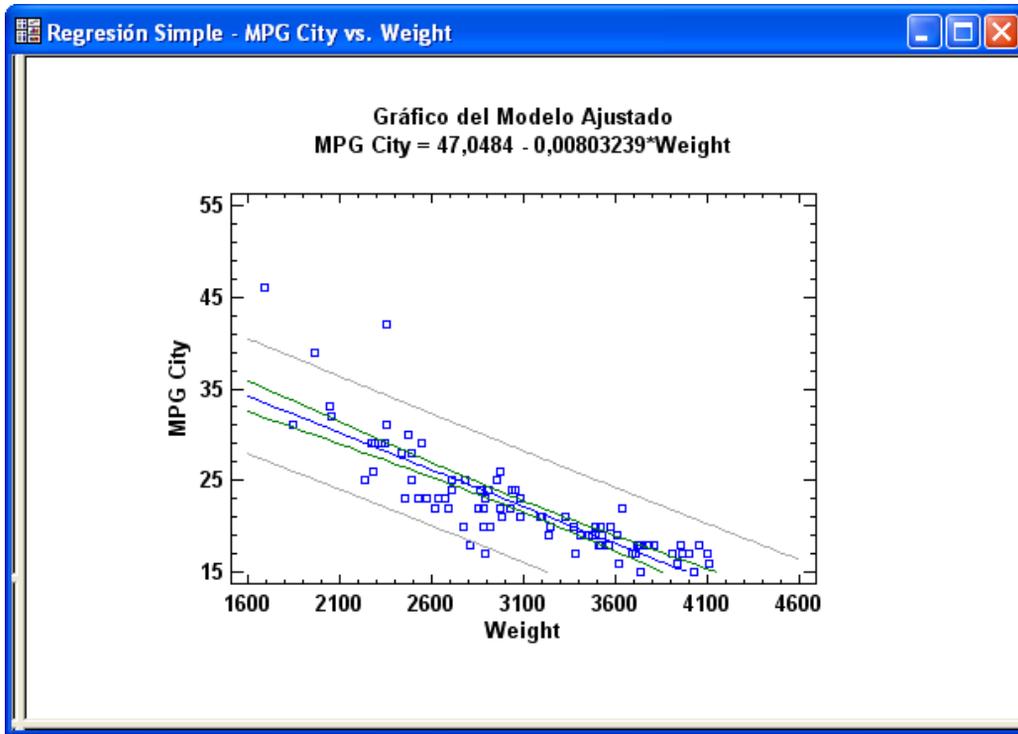


Figura 3-5. Ventana de análisis de Regresión simple con el panel maximizado

Haciendo doble clic en el panel una segunda vez, se restaura el panel múltiple.

Cuando se ha elegido una ventana de análisis, se activa directamente una segunda barra de herramientas bajo la barra de herramientas principal de STATGRAPHICS Centurion XVI. La barra de herramientas de análisis se muestra a continuación:



Cada uno de los botones en esta barra de herramientas ejecuta una operación importante.

3.2.1 Botón Entrada de diálogo



Cuando se presiona este botón, se muestra el cuadro de diálogo de entrada de datos utilizado originalmente para especificar las variables de datos, tal y como se muestra en la *figura 3-2*. Si cambia las variables de datos y presiona *ACEPTAR*, los análisis cambiarán para reflejar las nuevas selecciones. Esto posibilita intentar diferentes combinaciones de datos sin tener que comenzar un Nuevo análisis.

3.2.2 Botón Opciones de análisis



La mayoría de los análisis tienen múltiples opciones. Cuando se ejecuta un análisis por primera vez, se seleccionan los valores por defecto para estas opciones, que son suficientes a menudo. Sin embargo, presionando el botón *Opciones de análisis* en un procedimiento se permitirán cambiar estas opciones básicas. Para los cuadros de diálogo de *Regresión simple*, el cuadro de diálogo *Opciones de análisis* especifica el tipo de modelo que se va a ajustar y el método de estimación para los coeficientes desconocidos:

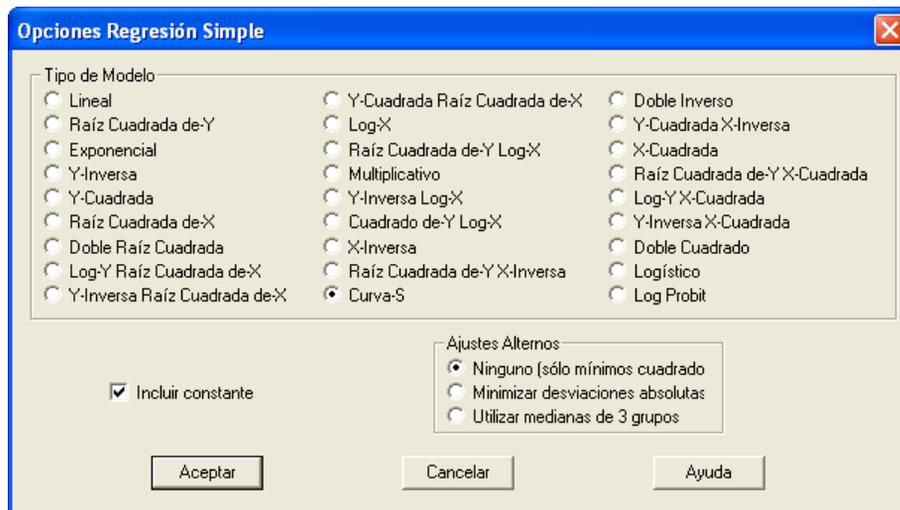


Figura 3-6. Cuadro de diálogo de opciones de análisis en regresión simple

Si examina la salida de la *figura 3-9*, puede observar en la tabla de modelos alternativos que varios modelos curvilíneos tienen un valor de R-cuadrado más alto que en el modelo lineal. Al principio de la lista está el modelo *S-Curva*. Si se selecciona este modelo en el cuadro de diálogo *Opciones de análisis* y se presiona *ACEPTAR*, cambiará el análisis completo para reflejar el nuevo modelo. Como podemos ver examinando el gráfico del modelo ajustado, una curva S captura la curvatura en los datos:

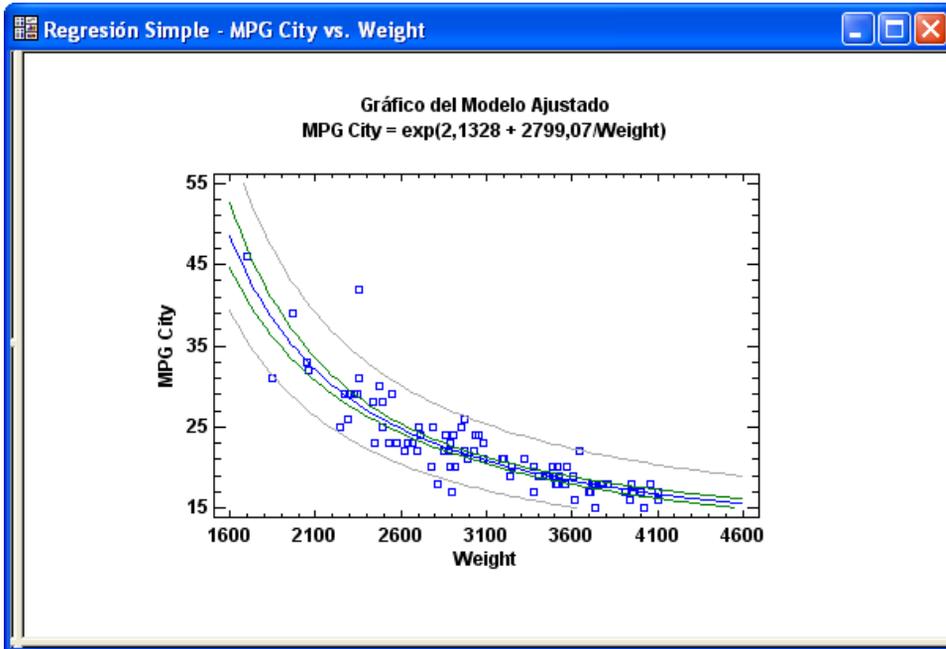


Figura3-7. Modelo en curva S ajustado

3.2.3 Botón tablas y gráficos



Este botón muestra una lista de tablas y gráficos adicionales que pueden ser añadidas a la ventana de análisis. Para *Regresión simple*, las tablas y gráficos disponibles son:

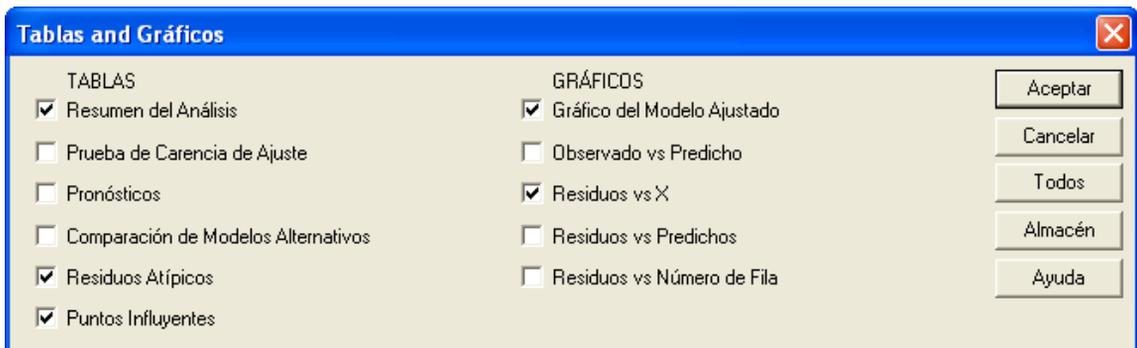


Figura 3-8. Cuadro de diálogo Tablas y Gráficos en Regresión simple

Por ejemplo, si selecciona añadir tablas mostrando modelos alternativos y residuos atípicos, se añadirán nuevos paneles a la ventana de análisis:

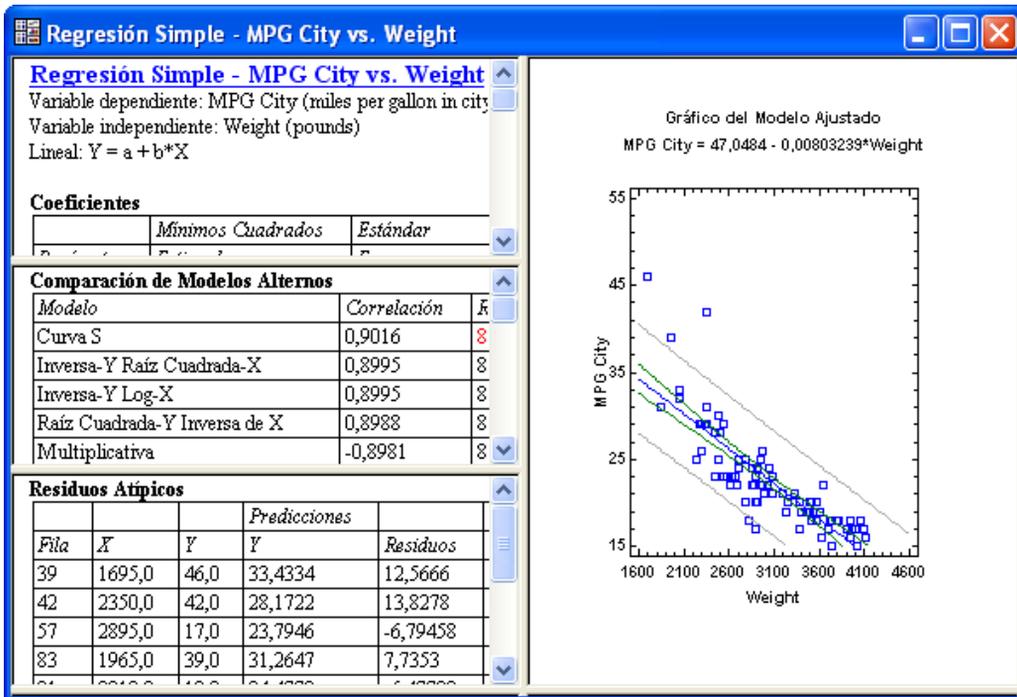


Figura 3-9. Ventana de análisis de Regresión simple con tablas añadidas

Añadiendo un gráfico de residuos se sitúa un gráfico adicional en la ventana de análisis:

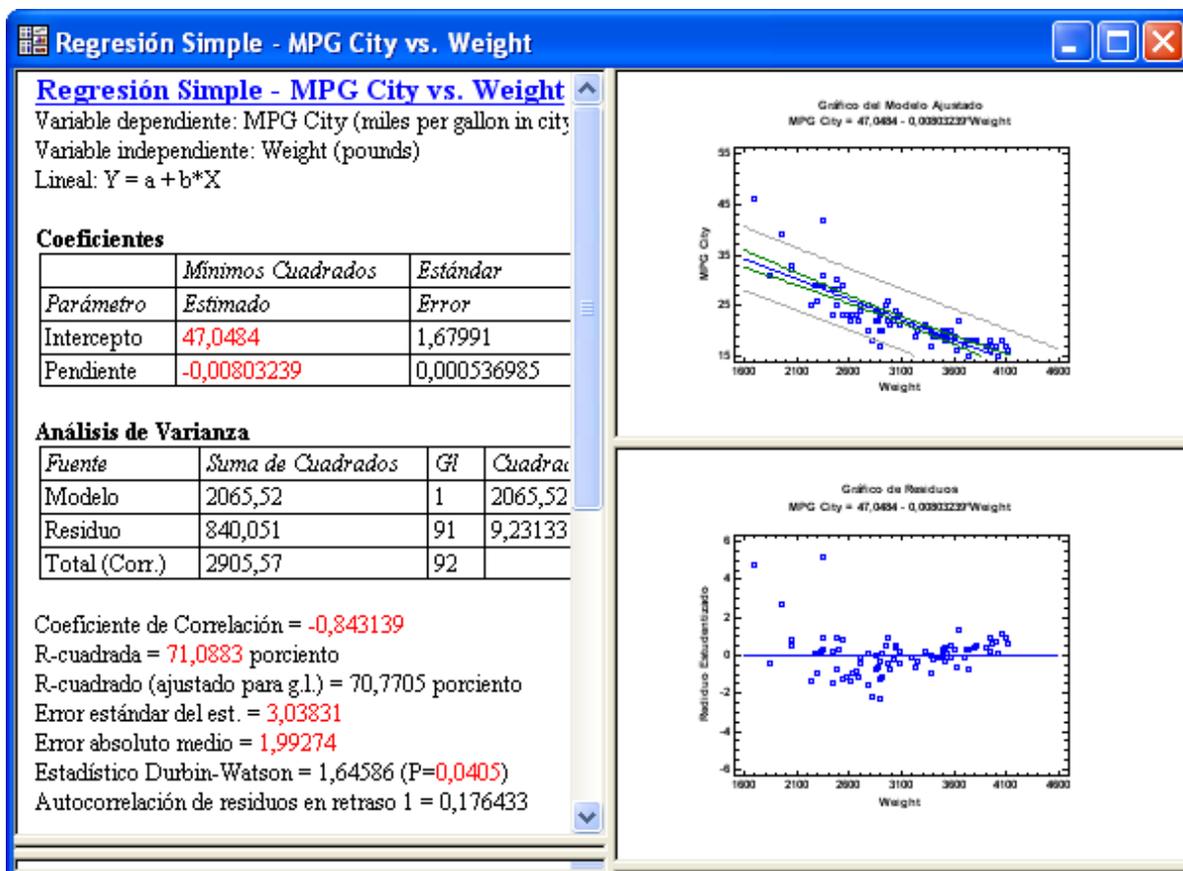


Figura 3-10. Ventana de análisis de Regresión simple con un gráfico añadido

3.2.4 Botón Opciones de panel

 En suma a las opciones que se aplican a la ventana entera del análisis, muchas tablas y gráficos individuales tienen opciones que se aplican sólo a ellos. Estas opciones son accesibles maximizando la tabla o gráfico y presionando *Opciones de panel*. Para un *Gráfico de modelo ajustado*, las opciones de panel son:

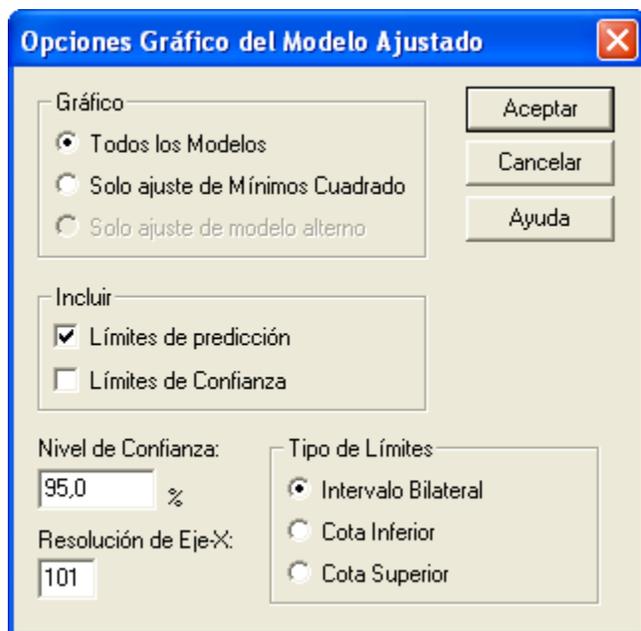


Figura 3-11. Cuadro de diálogo Opciones de panel para el gráfico del modelo ajustado

Por ejemplo, quitando las marcas correspondientes a los *límites de confianza* y presionando *Aceptar* se redibujará el gráfico sin los límites interiores:

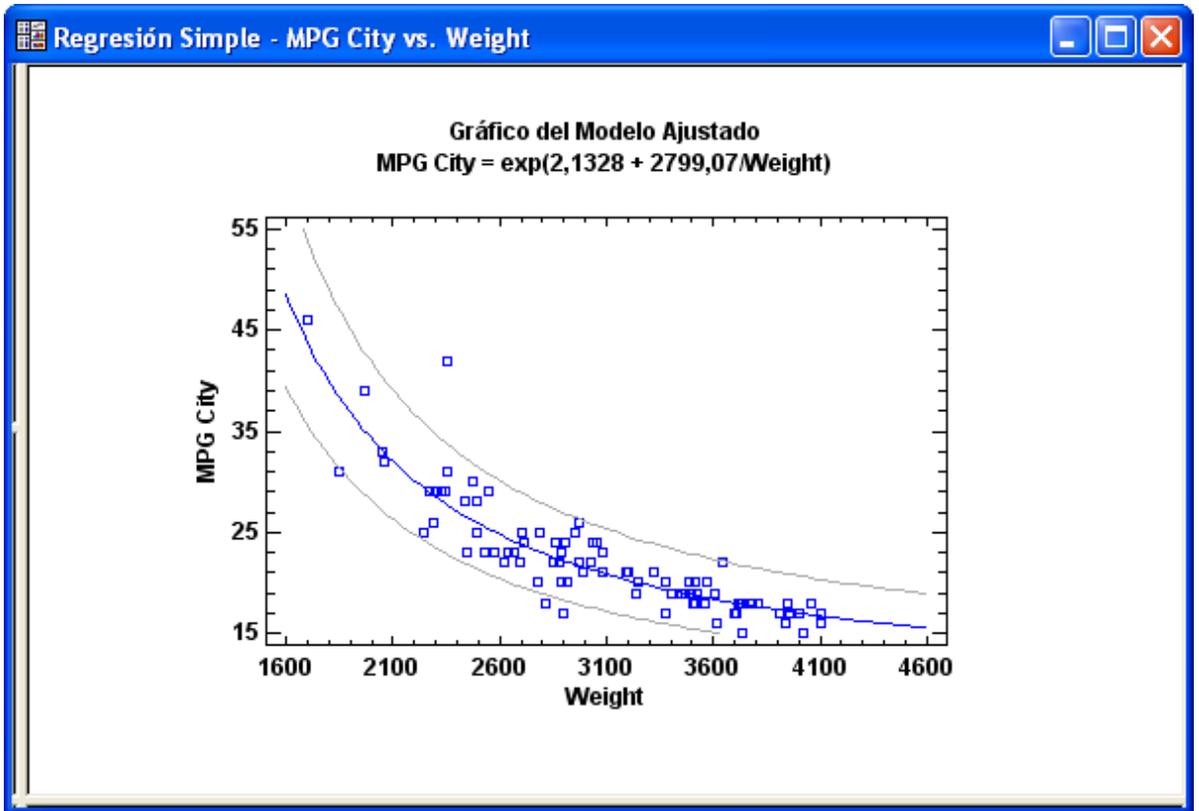


Figura 3-12. Gráfico del modelo ajustado sin los límites de confianza

3.2.5 Botón Guardar resultados

 Este botón permite guardar resultados numéricos calculados por el análisis estadístico en las columnas de la hoja de datos. Para *Regresión simple*, se muestran las siguientes posibilidades:

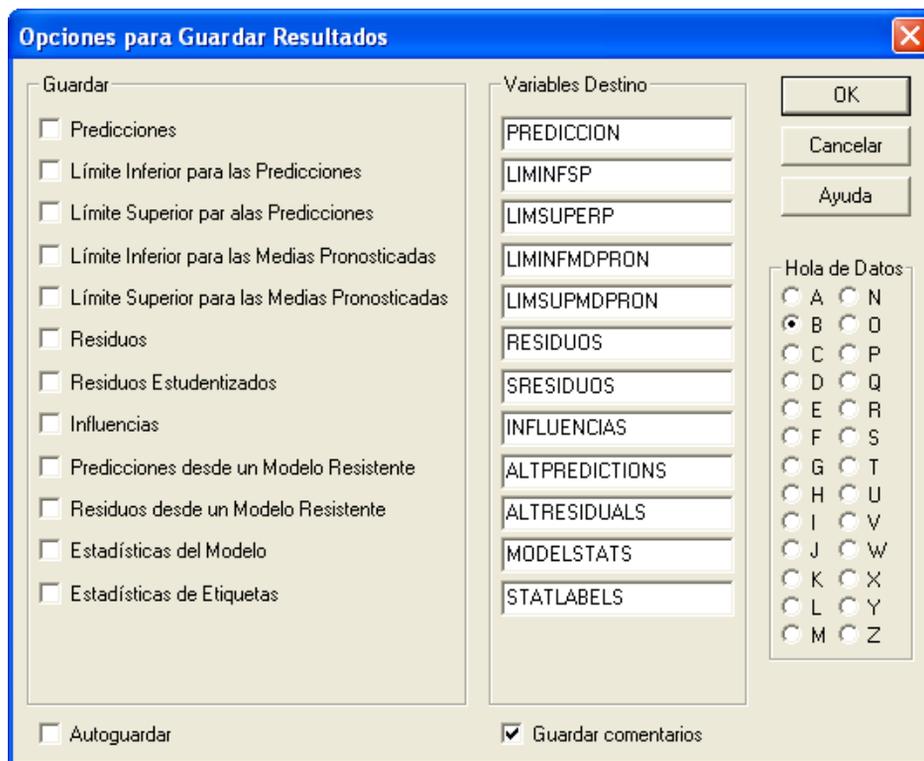


Figura 3-13. Cuadro de diálogo Guardar resultados en Regresión simple

Para guardar información, marque los ítems de interés en el campo *Guardar*. Para cada ítem a guardar, se asigna un nombre de columna bajo *Variables de destino* y se indica la hoja de datos deseada para el almacenamiento. Si quiere guardar un comentario con los datos, marque *Guardar comentarios*.

La casilla *Autoguardar* se utiliza para volver a guardar automáticamente el ítem seleccionado cuando el análisis se vuelva a ejecutar. Se utiliza si se intenta guardar el análisis en un StatFolio, análisis que será vuelto a ejecutar cuando se lea el StatFolio. Marcando la casilla *Autoguardar*, puede configurar un StatFolio para realizar cálculos automáticamente y guardar los estadísticos deseados. Cuando se combinan los StatFolios con las capacidades de código que se muestran en el capítulo 5, se posibilita la automatización de tareas.

3.2.6 Botones gráficos

Cuando se maximice un gráfico en la ventana de análisis, se habilitan varios botones adicionales. Estos botones incluyen:



Opciones gráficas – muestra un cuadro de diálogo utilizado para cambiar colores, etiquetas, escalado de ejes y otras características similares.



Añadir texto – se utiliza para añadir texto adicional al gráfico.



Separar – se utiliza para separar puntos en dirección horizontal o vertical para prevenir su solapamiento.



Resaltar – se resaltan puntos de colores en un gráfico de dispersión de acuerdo con los valores de una variable seleccionada.



Alisado/Rotación – alisa gráficos en dos dimensiones, o rota gráficos tridimensionales.



Pan o zoom – estrecha o agranda los gráficos X-, Y-, o Z.



Explorar – explora dinámicamente superficies de respuesta o gráficos de contorno.



Identificar – muestra la etiqueta identificativa de un punto cuando se hace clic en él con el ratón.



Localizar por nombre – ilumina los puntos cuyos valores son iguales a los introducidos en el campo *Localizar* (usado en conjunción con el botón *Identificar*).



Localizar por fila – ilumina los puntos correspondientes al número de fila introducido en el campo *Fila*.

Cada uno de estos botones se describe en el capítulo 4.

3.2.7 Botón Excluir

 Algunos procedimientos estadísticos permiten excluir interactivamente supuestos valores atípicos de un análisis, haciendo clic en el punto supuesto y presionando el botón *Excluir*. Por ejemplo, el gráfico de la *figura 3-12* muestra un punto que está fuera de los límites de predicción. Haciendo clic en este punto y presionando el botón *Excluir*, el modelo se reajusta sin el punto. El gráfico del modelo ajustado muestra el nuevo modelo, indicando con una X qué punto (o puntos) han sido removidos:

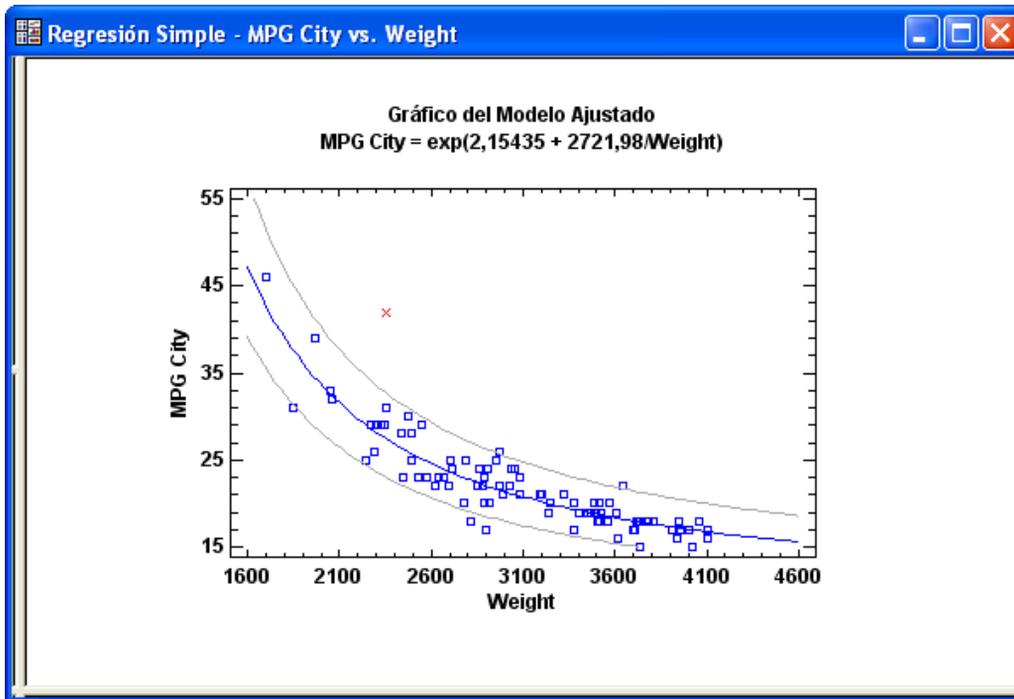


Figura 3-14. Modelo en curva S ajustado después de excluir un supuesto atípico

El resto de tablas y gráficos en la ventana de análisis se cambiarán también para reflejar el nuevo modelo.

Pueden excluirse múltiples puntos de un modelo haciendo clic en ellos de uno en uno y presionando el botón *Excluir*. Haciendo clic en un punto que ha sido removido se reintegrará en el modelo.

3.3 Imprimiendo los resultados

Para imprimir los resultados de un análisis estadístico son posibles dos opciones:

1. Para imprimir todas las tablas y gráficos de la ventana de análisis, presione el botón *Imprimir* en la barra de herramientas de análisis o seleccione *Imprimir* en el menú *Archivo*.
2. Para imprimir una tabla o gráfico simples, haga clic en su panel con el botón derecho del ratón y seleccione *Imprimir* en el menú emergente resultante.

Cuando imprima el análisis completo, se mostrará el siguiente cuadro de diálogo:



Figura 3-15. Cuadro de diálogo para imprimir un análisis

Bajo *Rango de impresión*, especifique los paneles a imprimir. Puede simultáneamente imprimir la salida en otra ventana de análisis eligiendo *Todos los análisis*.

Se utilizan opciones adicionales cuando la impresión se refiere a un cuadro de diálogo accesible seleccionando *Configurar página* en el menú *archivo*:



Figura 3-16. Cuadro de diálogo *Configurar página*

En este cuadro de diálogo es posible:

1. Especificar *márgenes* para imprimir páginas.
2. Indicar información de *cabecera* a imprimir al principio de cada página.
3. Indicar cuando cada panel (tabla o gráfico) debe ser mostrado en una página separada, o cuando *múltiples paneles* deben situarse en una página si ellos se ajustan.
4. Especificar el tamaño relativo de los gráficos como un porcentaje de las dimensiones de la página.
5. Elegir graficar la salida en blanco y negro, cuando la impresora tiene capacidades de color.
6. Imprimir el color del fondo (si existe) o sus gráficos.
7. Graficar *líneas vacías* usando 2 píxeles en lugar de 1. Esta última opción puede hacer aparecer los gráficos mejor marcados en una impresora de alta resolución.

Otras opciones, tales como imprimir la salida en modo retrato o en modo paisaje, se sitúan seleccionando *Opciones de impresión* en el menú *Archivo*, que accede al cuadro de diálogo ofrecido con el controlador de su impresora.

3.4 Publicando los resultados

La salida de un análisis estadístico puede ser publicada en formato HTML para verla con un navegador Web seleccionando *StatPublish* en el menú *Archivo*. Esto posibilita hacer la salida disponible para todo el mundo en la organización, cuando no se tiene STATGRAPHICS Centurion XVI en los ordenadores. La publicación se describe en el capítulo 5.

Puede también copiar el análisis a StatReporter, que permite anotar la salida y guardarla en un archivo RTF (formato de texto enriquecido), de modo que pueda ser leída en otros programas tales como Microsoft Word. La utilización de StatReporter se describe en el capítulo 6.

Gráficos

Modificando gráficos, guardando perfiles gráficos, interactuando con gráficos, guardando gráficos en archivos imagen y copiando gráficos a otras aplicaciones.

Juntos, los 160 procedimientos estadísticos de STATGRAPHICS Centurion XVI crean cientos de tipos diferentes de gráficos. Para facilitar el proceso de análisis de datos, los títulos por defecto, escalado y otros atributos se seleccionan en cualquier momento cuando se crea un nuevo gráfico. Para los análisis habituales, las opciones por defecto suelen ser suficiente. Sin embargo, cuando llega el momento de publicar los resultados finales los gráficos son importantes para crear publicaciones de calidad.

Este capítulo describe todo lo que necesita conocer para trabajar con gráficos en STATGRAPHICS Centurion XVI. Se muestra cómo preparar los gráficos para una publicación y cómo copiarlos en aplicaciones como Microsoft Word y PowerPoint. También muestra cómo interactuar con gráficos. Por ejemplo, necesita observar un punto interesante y conocer algo más de él, o necesita rotar un gráfico 3D teniendo presente la relación entre las variables proyectadas sobre los ejes X, Y y Z.

Como ejemplo, consideramos otra vez los datos del archivo *93cars.sgd*. Para comenzar, el gráfico del modelo ajustado relaciona millas por galón conduciendo en ciudad y el peso de los vehículos y nos servirá para ilustrar algunas de las operaciones más importantes con gráficos.

4.1 Modificando gráficos

El procedimiento *Regresión simple* se utiliza habitualmente para ajustar curvas relacionando una variable respuesta Y y una segunda variable explicatoria X. Como se ilustra posteriormente en este capítulo, un modelo en curva S proporciona un buen ajuste a la relación entre la columna *MPG City* y la columna *Weight* en el archivo *93cars.sf6*.

La primera vez que se crea, se obtiene un gráfico del modelo en Curva S como se muestra a continuación:

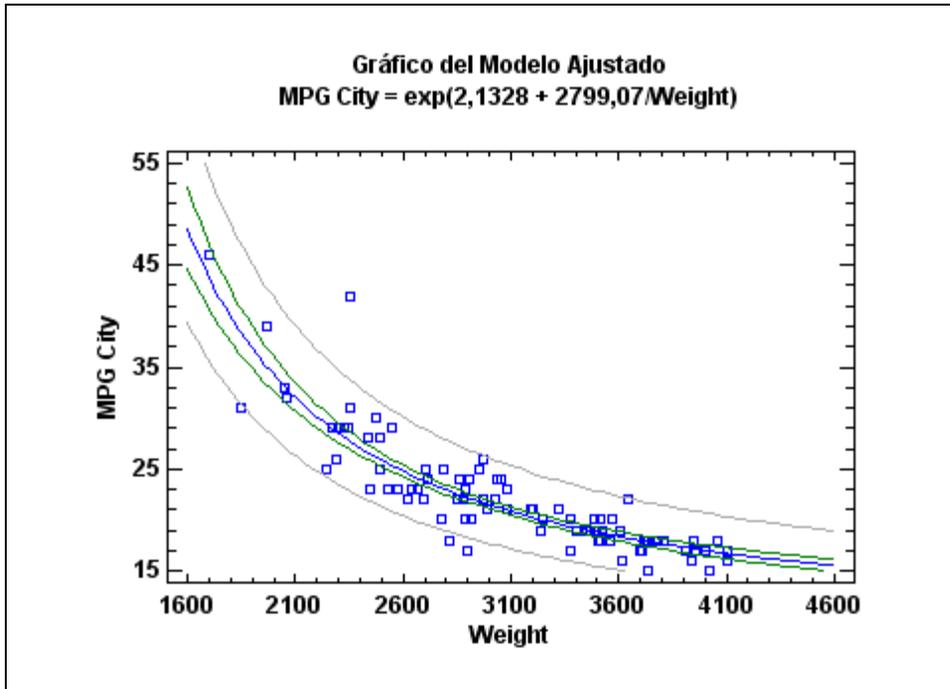


Figura 4-1. Gráfico del modelo ajustado con título y escalado por defecto

Los títulos, escalado, puntos, tipos de línea, colores y otros atributos gráficos se generarán automáticamente.

4.1.1 Opciones de trazado

Para modificar un gráfico una vez que ha sido creado, primero haga doble clic sobre él para que ocupe la ventana de análisis. Haga clic en el botón *Opciones gráficas*  localizado en la barra de herramientas de análisis. Se mostrará un cuadro de diálogo con varias pestañas correspondientes a diferentes elementos gráficos.

La pestaña *Diseño* del cuadro de diálogo *Opciones gráficas* se utiliza para cambiar algunas de las características básicas del gráfico:

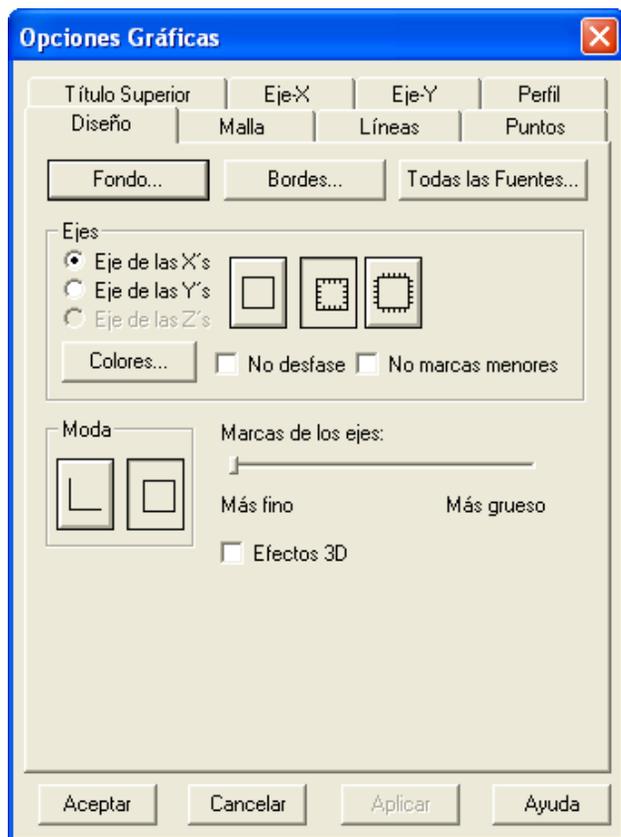


Figura 4-2. Pestaña *Trazado* en el cuadro de diálogo *Opciones gráficas*

Incluye la orientación y las marcas de los ejes, el grosor de los ejes y el color del fondo y bordes de los gráficos. Por ejemplo, cambiando el color del *Fondo* a amarillo y añadiendo *Efectos 3D* se modifica el gráfico como se muestra a continuación:

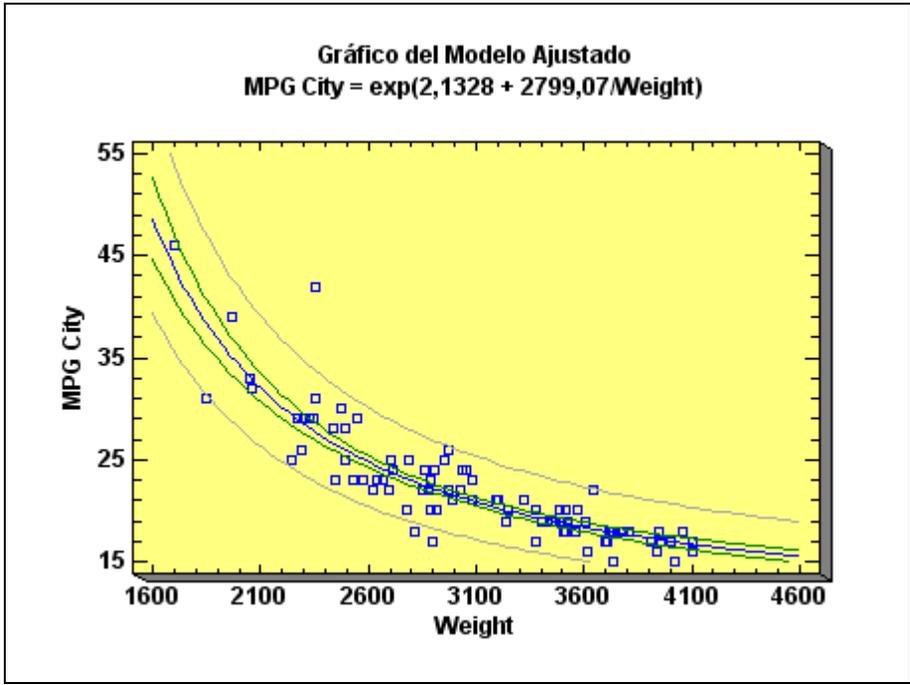


Figura 4-3. Gráfico después de modificar el color del fondo y seleccionar efectos 3D

NOTA: El proceso de cambio de color puede verse en la documentación de ayuda que aporta el software haciendo clic en *Ayuda – Guía de usuario*.

4.1.2 Opciones de rejilla (malla)

La pestaña *Malla* se utiliza para añadir una rejilla al gráfico:



Figura 4-4. Pestaña *Rejilla* en el cuadro de diálogo *Opciones gráficas*

Añadiendo un gris, los bordes de la rejilla en *Ambas* direcciones producen el siguiente gráfico:

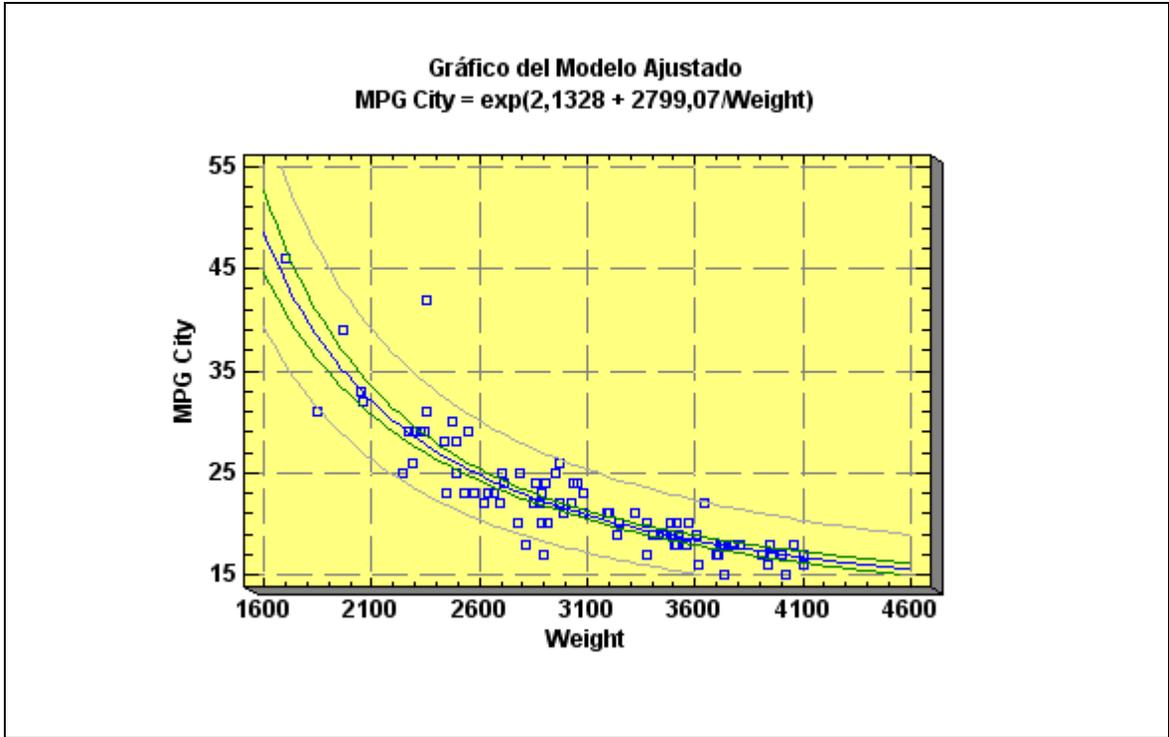


Figura 4-5. Gráfico después de añadir una rejilla

4.1.3 Opciones de líneas

La pestaña *Líneas* se utiliza para especificar el tipo, color y grosor de las líneas de un gráfico:

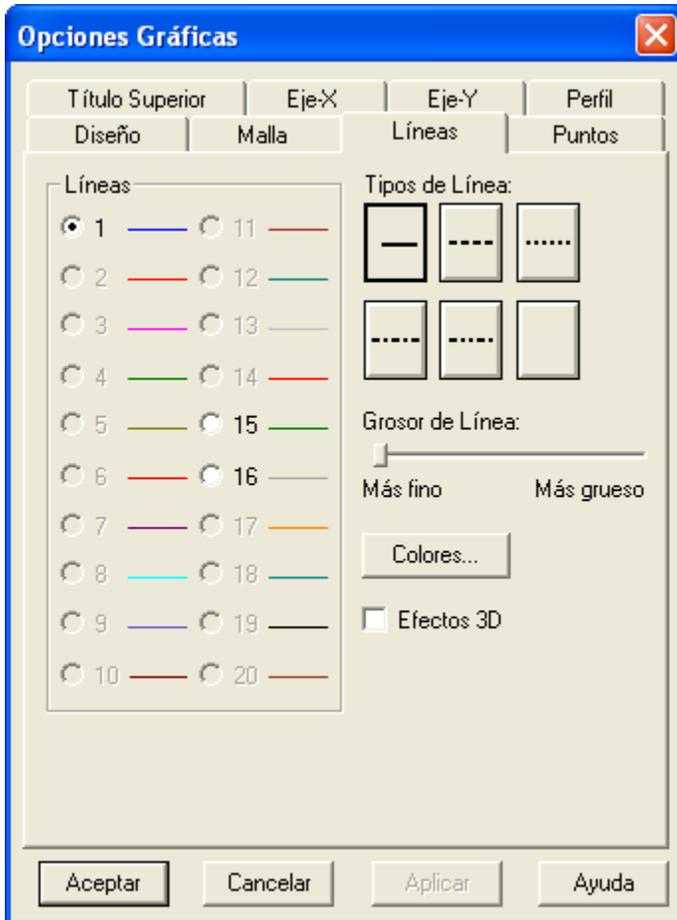


Figura 4-6. Pestaña *Líneas* en el cuadro de diálogo *Opciones gráficas*

Un gráfico tal como el del modelo ajustado tiene tres conjuntos de líneas: la línea del mejor ajuste, los límites de confianza interiores y los límites de predicción exteriores. Para cambiar alguno de estos tipos, haga clic en los botones #1, #2 o #3 y seleccione los atributos deseados. Aumentando el grosor de la línea central y cambiando otros tipos de línea se tiene:

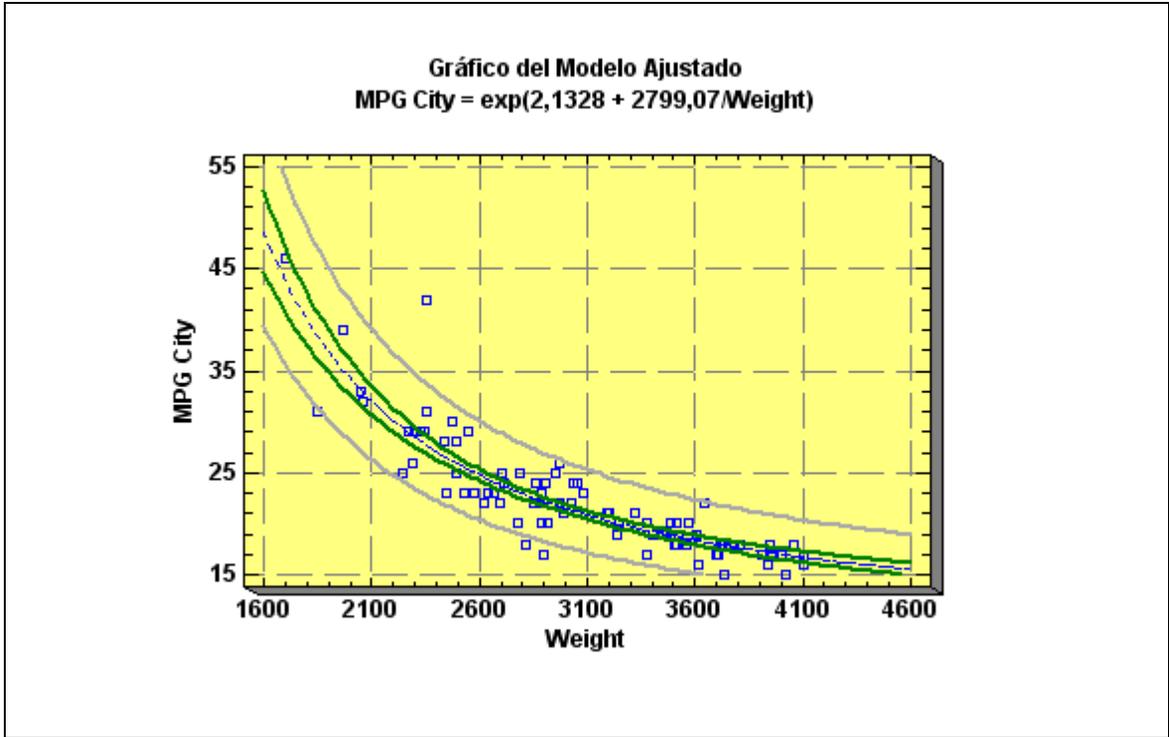


Figura 4-7. Gráfico después de modificar los Tipos de línea

NOTA: Solamente puede cambiar la anchura de líneas sólidas.

4.1.4 Opciones de puntos

La pestaña *Puntos* se utiliza para especificar el tipo, color y tamaño de los puntos en un gráfico:

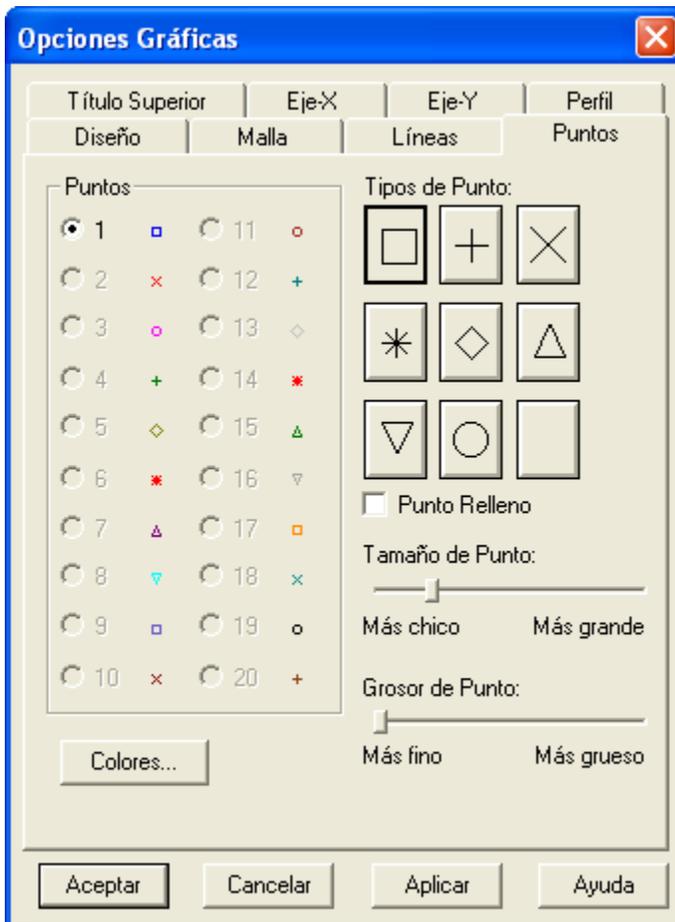


Figura 4-8. Pestaña *Puntos* en el cuadro de diálogo *Opciones gráficas*

El botón #1 controla los atributos del primer conjunto de puntos de un gráfico. En el ejemplo actual, hay solo un conjunto. Cambiando los puntos a diamantes sólidos se crea el gráfico siguiente:

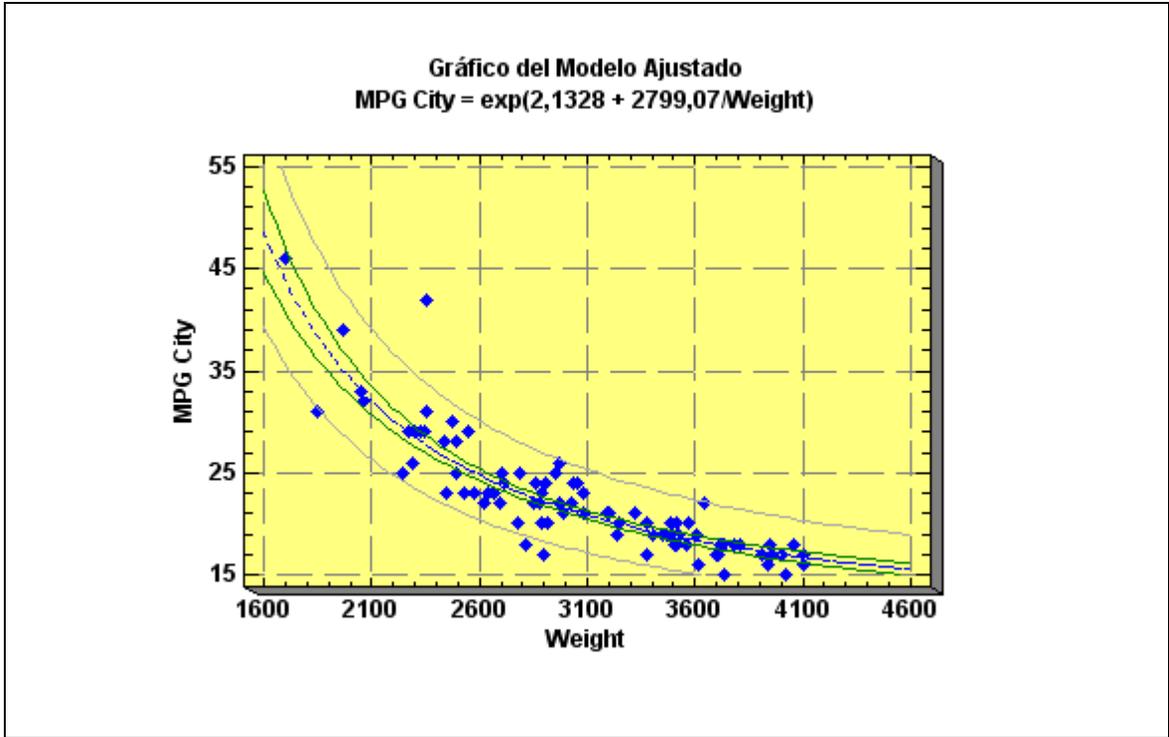


Figura 4-9. Gráfico después de modificar el tipo de puntos

4.1.5 Opciones del título superior

La pestaña *Título superior* se utiliza para especificar el tipo de texto y fuente para la información mostrada en la parte superior del gráfico:

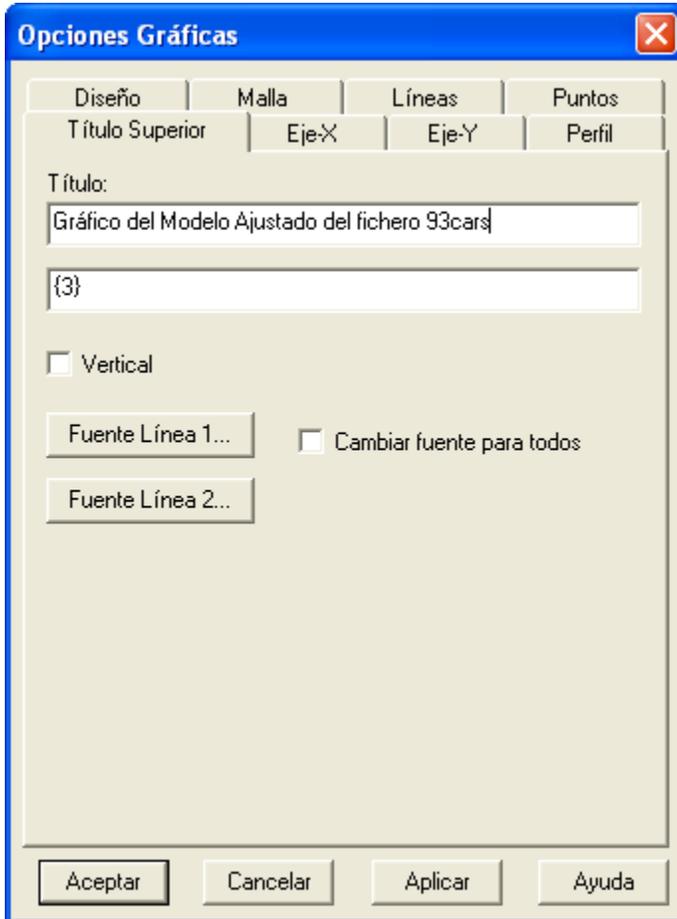


Figura 4-10. Pestaña *Título Superior* en el cuadro de diálogo *Opciones Gráficas*

Los gráficos tienen hasta dos líneas de título. Una entrada tal como “{3}” en un campo título indica que el texto se genera automáticamente por el procedimiento de análisis, conteniendo habitualmente nombres de variable o estadísticos calculados. Puede cambiar cualquier título, incluyendo los que se han generado automáticamente. Puede también arrastrar el título con el ratón a una nueva localización:

Gráfico del Modelo Ajustado del fichero 93cars

$$\text{MPG City} = \exp(2,1328 + 2799,07/\text{Weight})$$

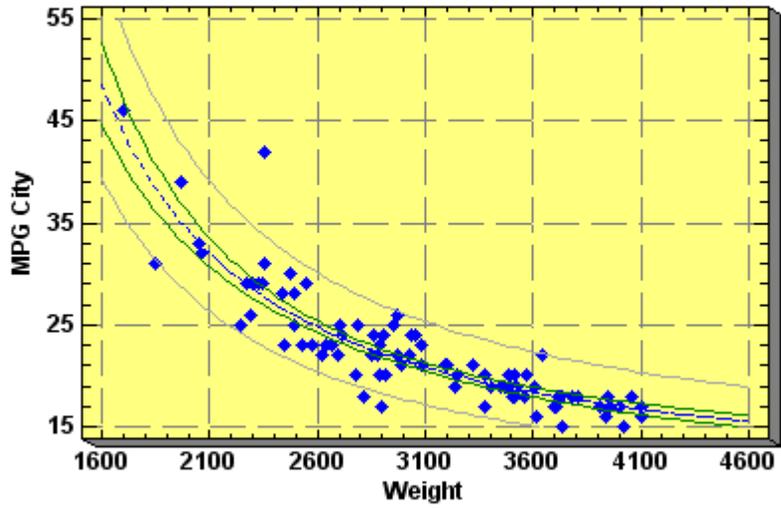


Figura 4-11. Gráfico después de modificar el Título superior

4.1.6 Opciones de escalado de ejes

El cuadro de diálogo *Opciones gráficas* contiene pestañas que permiten modificar los títulos de los ejes y su escala:

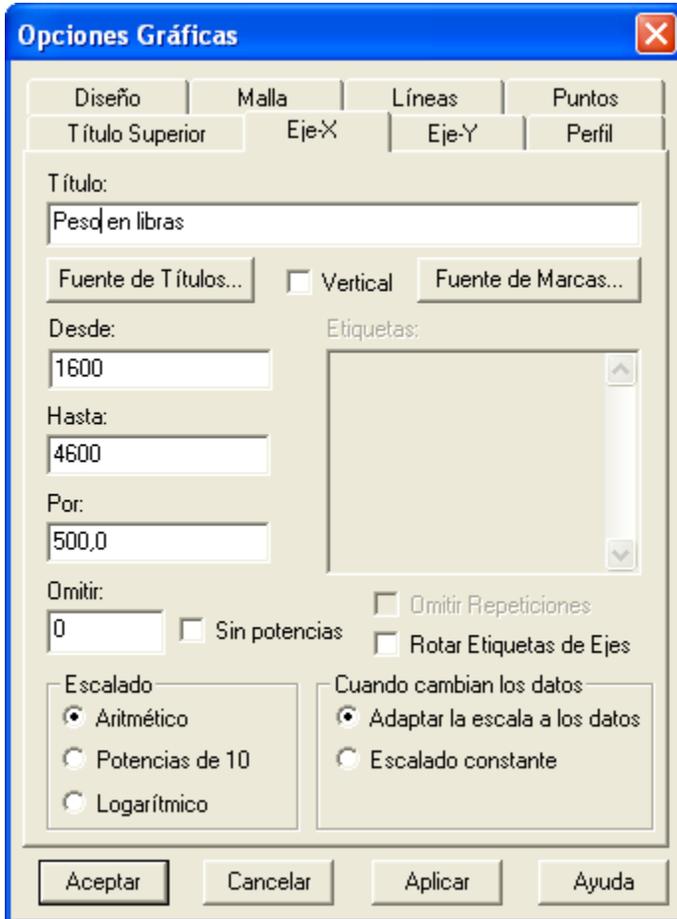


Figura 4-12. Pestaña Eje X en el cuadro de diálogo *Opciones Gráficas*

Hay varios campos importantes en este cuadro de diálogo:

1. *Título*: Título graficado a lo largo de los ejes.
2. *Desde*, *Hasta*, *Por* y *Omitir*: conjunto de marcas de escalado. El valor de *Salto* se utiliza para prevenir que se muestren ciertas marcas solapándose unas con otras. Por ejemplo, un valor de 1 en el campo *Omitir* mostrará un salto en cada marca.

3. *Rotar etiquetas de los ejes:* cambia las etiquetas de las marcas a vertical.
4. *No Potencia:* suprime el mostrar números grandes y pequeños utilizando etiquetas tales como (X 1000).
5. *Escalado:* dibuja los ejes utilizando dos escalas logarítmicas en base 10 diferentes.
6. *Cuando cambian los datos:* especifica cuándo el escalado será constante o cambiará cuando se grafiquen nuevos datos.
7. *Fuente de marca:* presiones estos botones para cambiar el color, tamaño o estilo del título y marcas.

La salida generada por el cuadro de diálogo anterior realiza los cambios que se muestran a continuación:

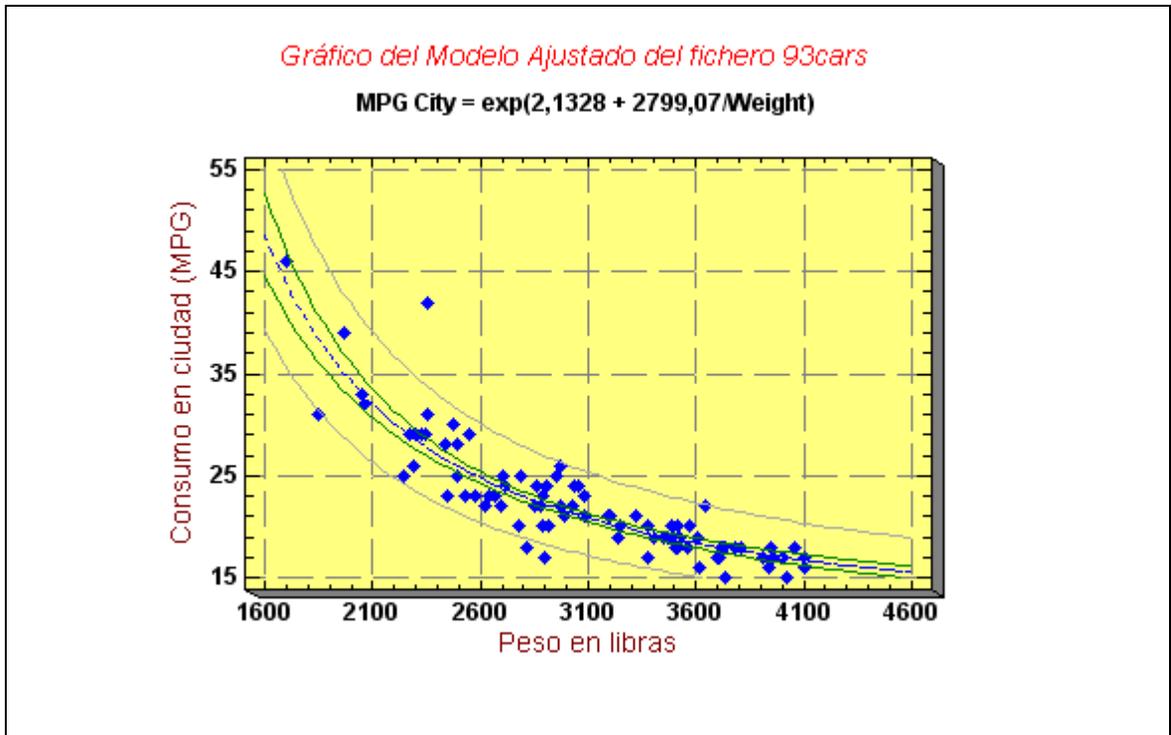


Figura 4-13. Gráfico después de modificar los títulos de los ejes y escalado

4.1.7 Opciones de relleno

Algunos gráficos, tales como los histogramas, contienen áreas sólidas. La pestaña *Relleno* en el cuadro de diálogo *Opciones gráficas* controla el color y tipo de relleno de barras, polígonos y trozos de los sectores:

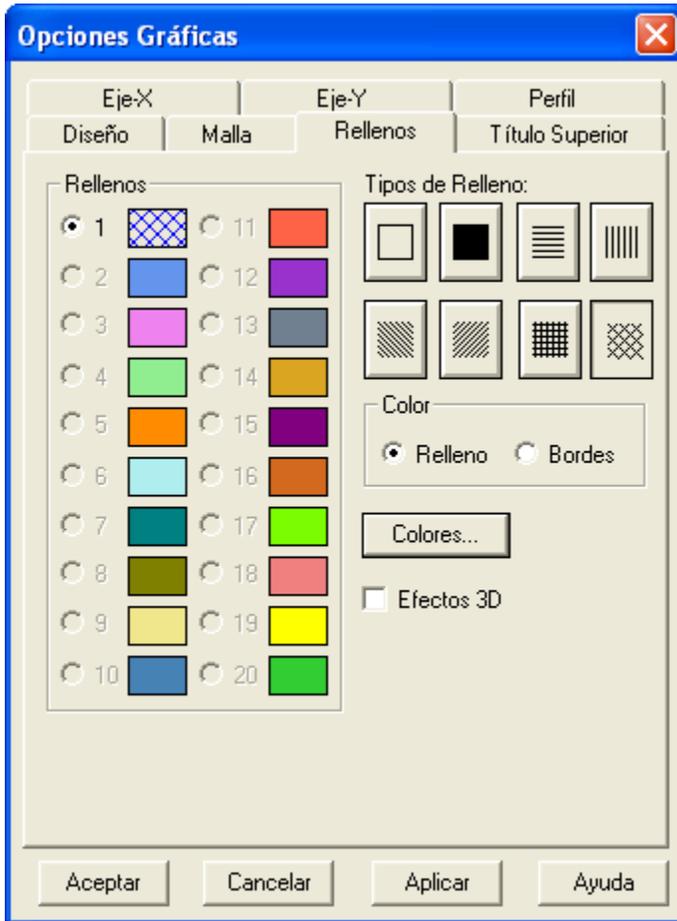


Figura 4-14. Pestaña Relleno en el cuadro de diálogo Opciones Gráficas

El botón #1 controla el primer tipo de relleno en un gráfico. En un histograma, todas las barras utilizan el primer tipo de relleno. En algunos gráficos, tales como gráficos de sectores, se utiliza más de un tipo de relleno. En estos casos, los botones #2 hasta #20 controlan los otros tipos de relleno.

Para gráficos tales como histogramas, situar el tipo de relleno en “no sólido” es una buena idea cuando se imprimen los resultados en blanco y negro:

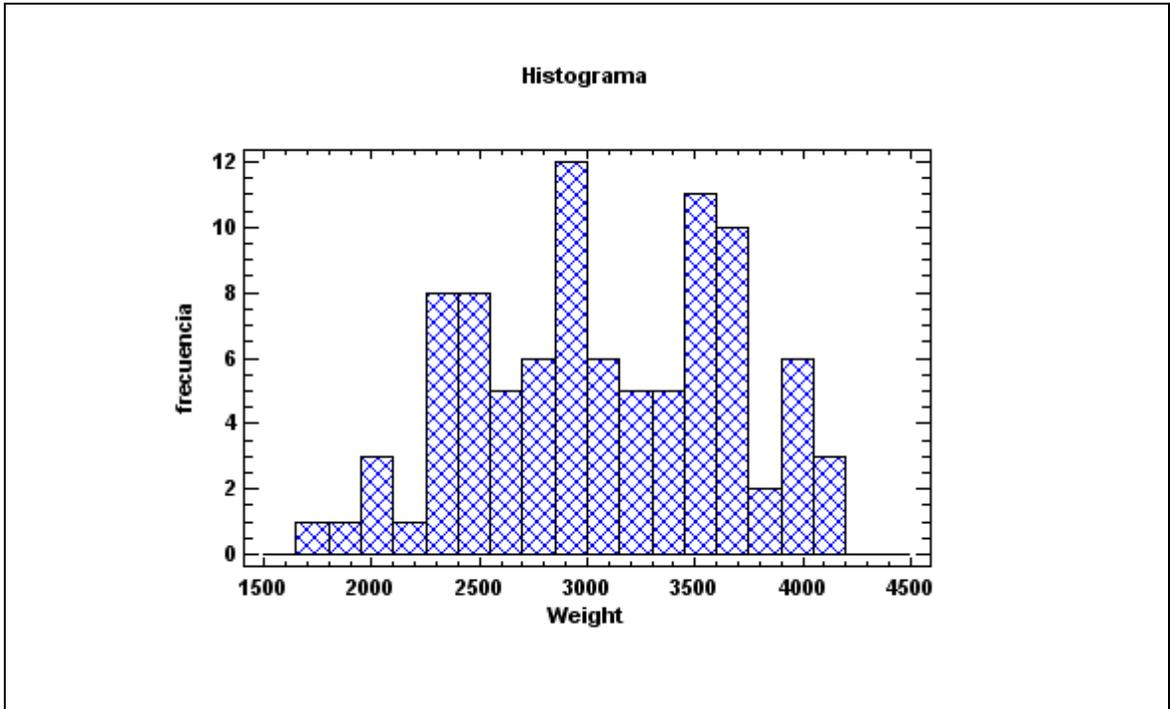


Figura 4-15. Histograma de frecuencias con tipo de relleno modificado

4.1.8 Texto, Opciones de etiquetas y leyendas

Para gráficos que contienen leyendas adicionales o etiquetas, puede cambiar texto y fuentes por medio de las pestañas del cuadro de diálogo *Opciones gráficas*.

4.1.9 Añadiendo texto nuevo

Puede añadirse texto adicional a un gráfico presionando el botón *Añadir texto*  en la barra de herramientas de análisis. Se generará un cuadro de diálogo en el cual se introduce el nuevo texto:



Figura 4-16. Cuadro de diálogo para añadir Nuevo texto

El texto de la cadena estará posicionado inicialmente bajo el título superior, sin embargo, puede ser arrastrado con el ratón a cualquier otra localización:

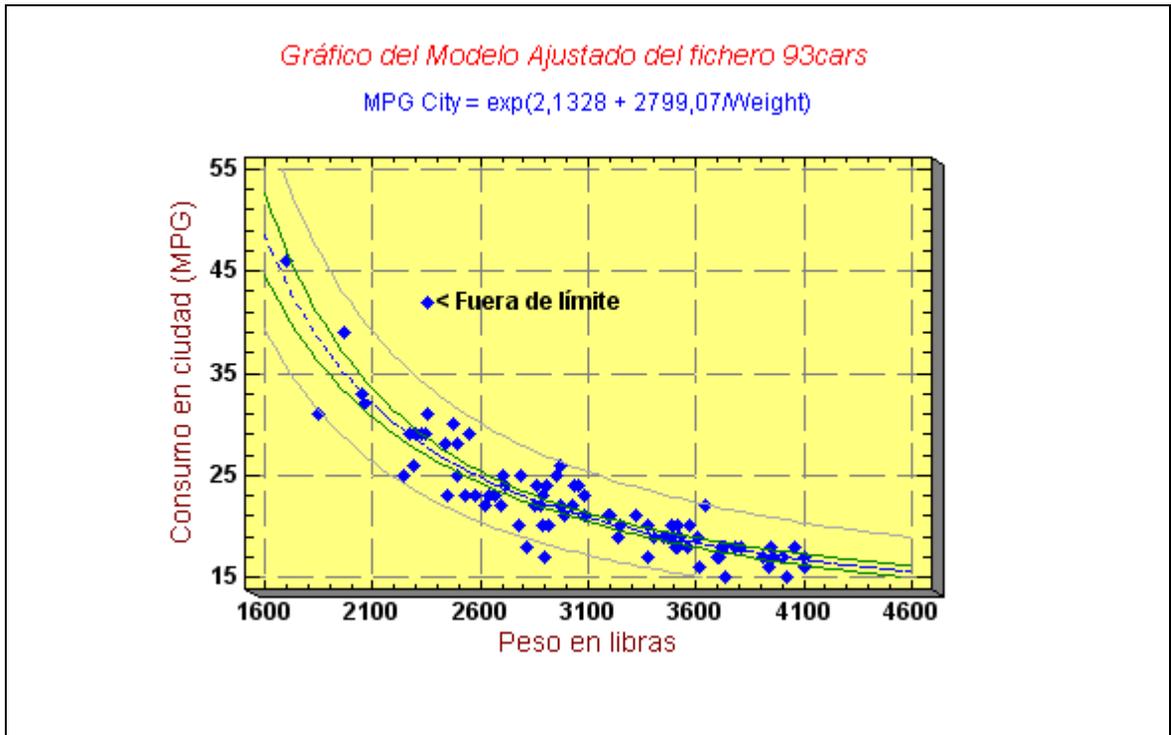


Figura 4-17. Gráfico después de añadir una nueva cadena de texto

Después de añadir el texto, haga clic en él y presione el botón *Opciones gráficas* si necesita realizar cambios.

4.2 Separando un gráfico de dispersión

Cuando una o ambas de las variables de un gráfico de dispersión es discreta, puede haber muchos puntos coincidentes prácticamente en la misma posición que oscurecen el gráfico. La barra de herramientas tiene el botón *Separar* que soluciona este problema excluyendo aleatoriamente puntos en la dirección de los ejes horizontal y vertical. Por ejemplo, considerar el siguiente gráfico con los datos del archivo *93cars.sgd*:

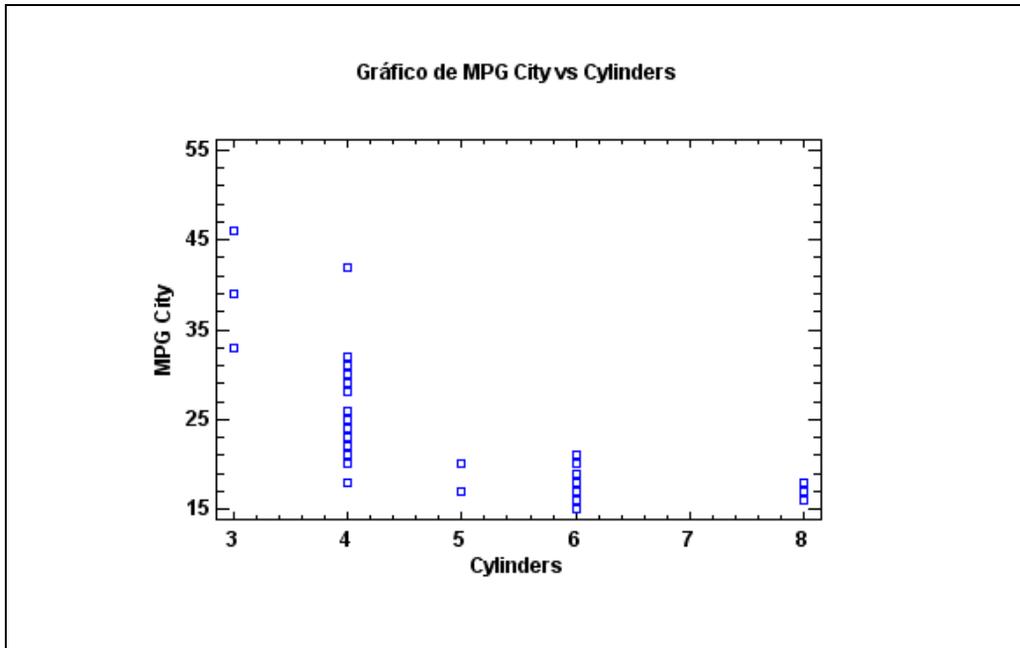


Figura 4-18. Gráfico de dispersión de millas por galón contra cilindros

Aunque hay 93 filas en la hoja de datos, hay menos puntos en el gráfico.

Si se presiona el botón *Separar* aparece el cuadro de diálogo siguiente para añadir una pequeña separación (aleatoriamente) en los puntos:

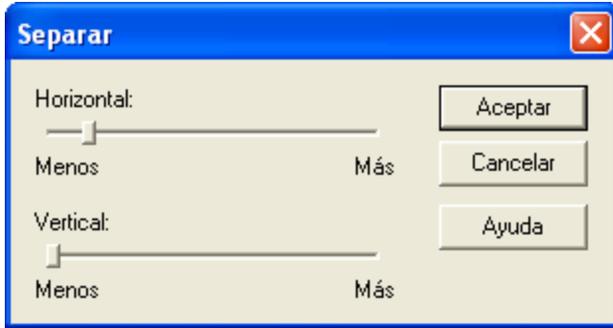


Figura 4-19. Cuadro de diálogo de Separación

En este caso, añadiendo una pequeña cantidad de separación horizontal se ve mejor el gráfico y la localización de los puntos:

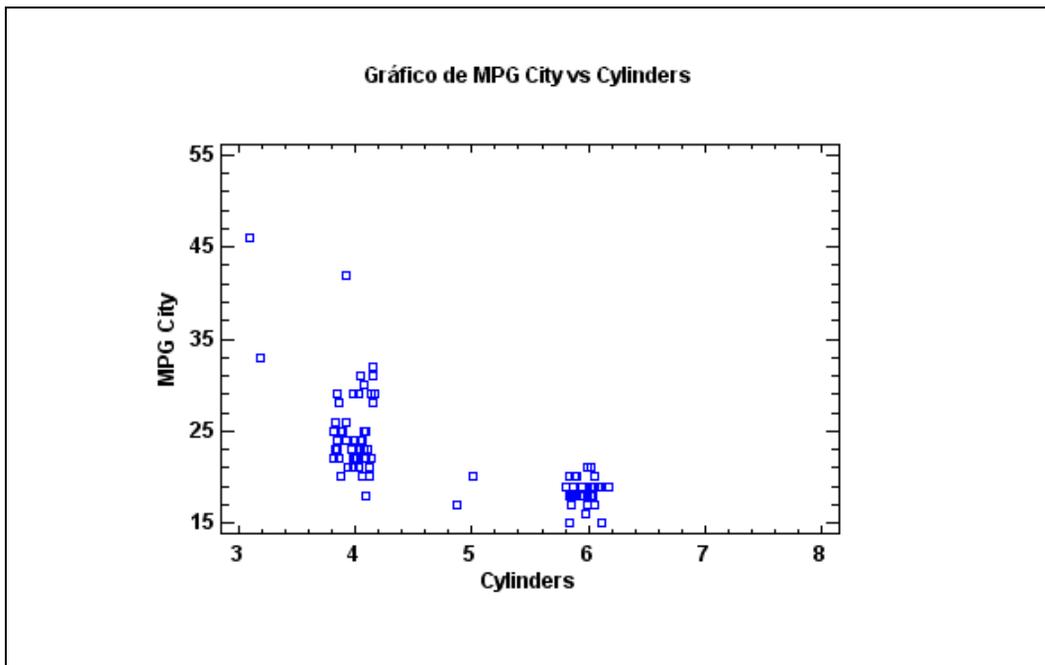


Figura 4-20. Gráfico de dispersión para Separación horizontal

Cada punto ha sido desplazado una pequeña cantidad a lo largo del eje X. La separación de puntos afecta solamente al aspecto del gráfico. No afecta a los datos de la hoja de datos y a ningún cálculo hecho con ellos.

4.3 Resaltando un gráfico de dispersión

Un método interesante de visualización de relaciones entre variables consiste en colorear los puntos del gráfico de dispersión según los valores de otra variable. Por ejemplo, considerar el siguiente *Gráfico matricial* para variables del archivo *93cars.sgd*:

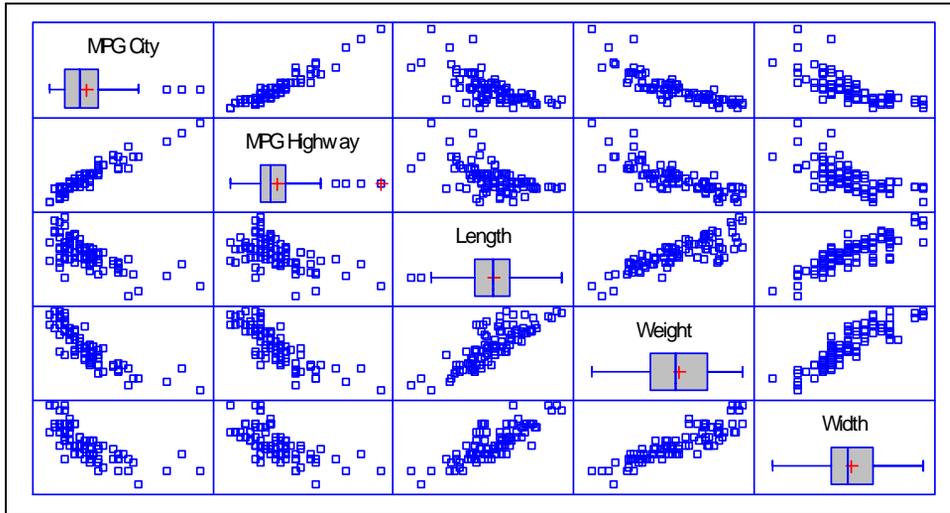


Figura 4-21. Gráfico matricial para datos del archivo *93cars*

El gráfico de dispersión de cada celda del gráfico matricial cruza los valores de las variables correspondientes a su fila y columna.

Supongamos que queremos visualizar cómo la potencia de los automóviles está relacionada con las 5 variables representadas. Si presiona el botón *Resaltar*  en la barra de herramientas de análisis, se mostrará el cuadro de diálogo siguiente:

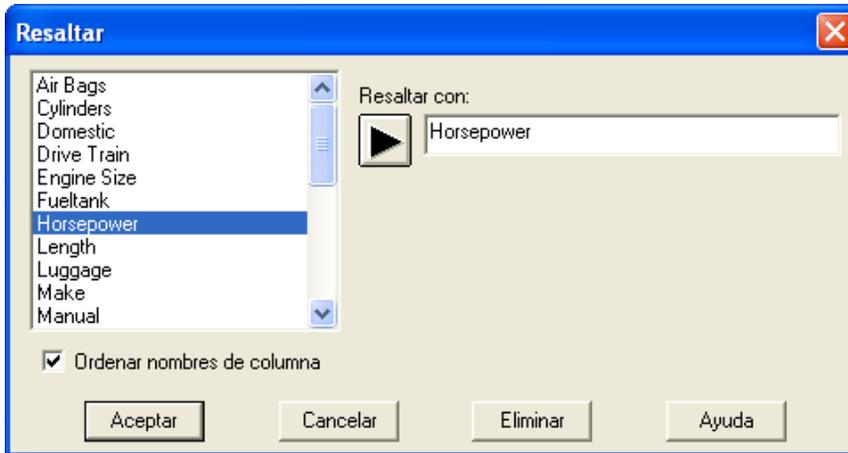


Figura 4-22. Cuadro de diálogo para seleccionar la variable de resaltado

Seleccione una variable cuantitativa para utilizar el código de puntos de resaltado. Después de la selección de la variable de resaltado, aparecerá un cuadro de diálogo flotante:

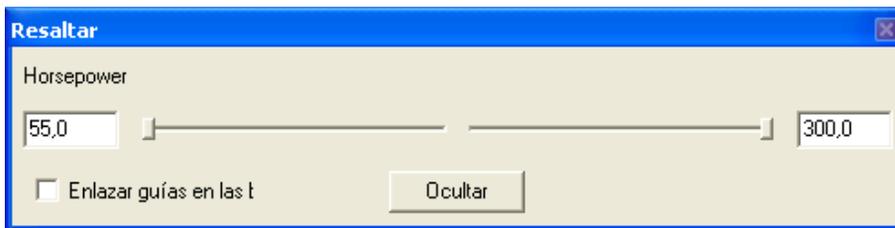


Figura 4-23. Cuadro de diálogo flotante para la selección del intervalo de resaltado

Las dos barras deslizantes se utilizan para especificar los límites inferior y superior para la variable de resaltado. Todos los puntos en el gráfico serán coloreados de azul claro si caen en el intervalo especificado. Por ejemplo, en el gráfico siguiente, todos los automóviles con potencia entre 55.0 y 121.15 se colorean de azul claro:

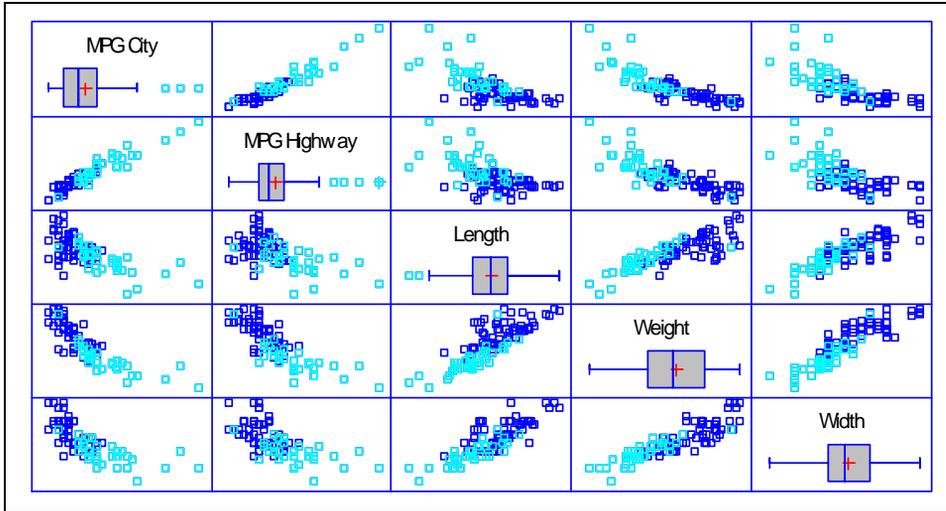


Figura 4-24. Gráfico matricial después del resaltado de puntos

Es evidente en el gráfico anterior que la *Potencia* está fuertemente correlacionada con las otras variables.

4.4 Alisando un gráfico de dispersión (tendencia)

Para ayudar a visualizar las relaciones entre las variables en un gráfico de dispersión, se puede añadir un alisado o tendencia. Para alisar un gráfico de dispersión, presione el botón

Alisar/Rotars  en la barra de herramientas de análisis. Se mostrará el siguiente cuadro de diálogo:

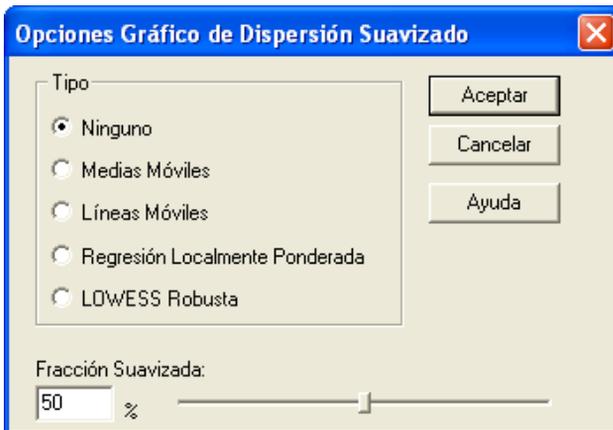


Figura 4-25. Cuadro de diálogo Alisando gráfico de dispersión

El alisamiento de un gráfico de dispersión se realiza seleccionando un conjunto de localizaciones a lo largo del eje X graficando cada localización como una media ponderada de la fracción especificada de los puntos que estarán contenidos en la localización. Uno de los mejores métodos de alisado es el llamado LOWESS (LOcally WEighted Scatterplot Smoothing), usualmente con una fracción de alisado entre el 40% y el 60%. El resultado del alisado del gráfico matricial de los datos de automóviles se muestra a continuación:

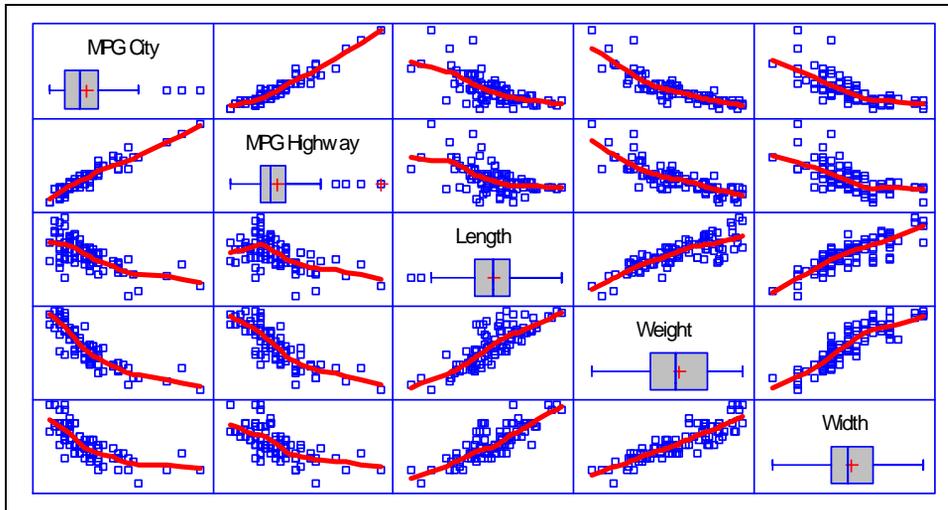


Figura 4-26. Alisado del gráfico matricial utilizando fracción de alisado inferior al 50%

El alisado ayuda a ilustrar el tipo de relación entre las variables.

4.5 Identificando Puntos

Para mostrar el número de fila y coordenadas correspondientes a un punto del gráfico, debe de presionar con el botón del ratón en el punto. Se mostrará un pequeño cuadro de diálogo en la esquina superior derecha del gráfico, mostrando el número de fila y coordenadas del punto:

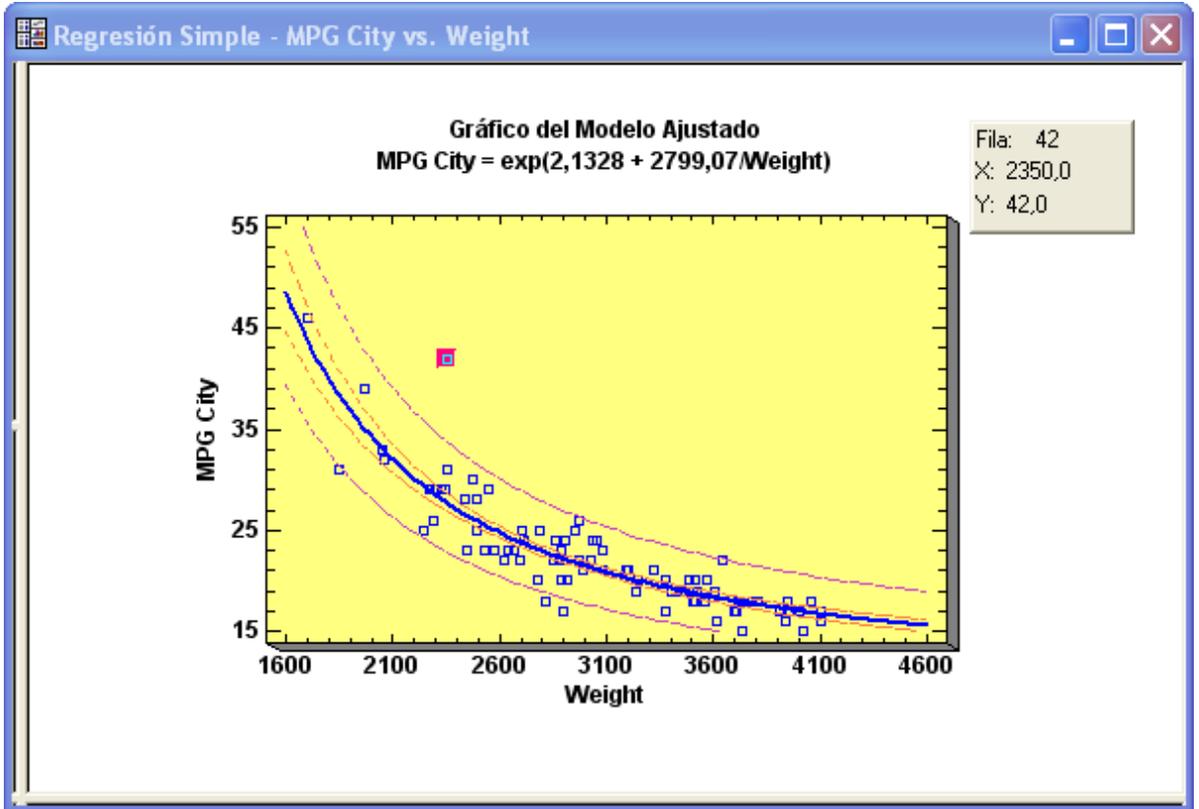


Figura 4-27. Mostrando información acerca de puntos seleccionados

Al mismo tiempo, el número de fila del punto se situará en el campo *Fila* de la barra de herramientas de análisis:



Figura 4-28. Barra de herramientas de análisis mostrando número de fila del punto seleccionado

puede obtenerse información adicional acerca del punto presionando el botón *Identificar*  y seleccionando una columna del libro de datos:

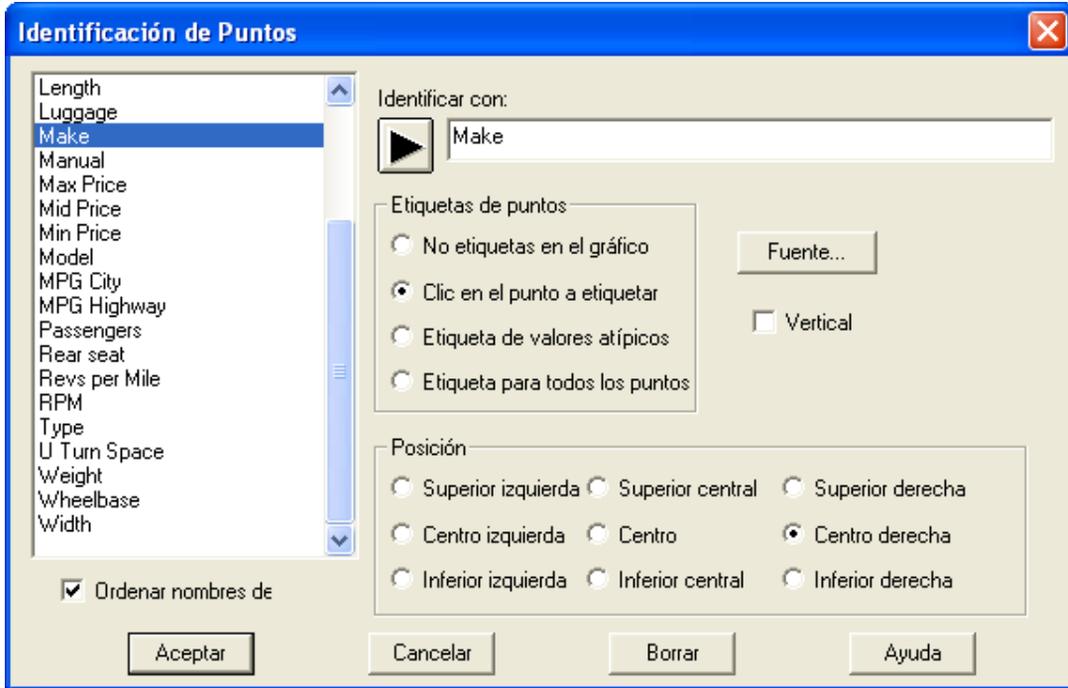


Figura 4-29. Cuadro de diálogo de Identificación de puntos

después de seleccionar una variable, haciendo clic en un punto se añadirá el valor correspondiente de la variable al campo etiquetado *Etiqueta* de la barra de herramientas de análisis:



Figura 4-30. Barra de herramientas de análisis mostrando una selección de punto

El botón de los binoculares  situado a la izquierda de *Etiqueta* y *Fila* puede ser utilizado para localizar puntos en un gráfico. Si introduce un valor en cualquier campo al lado del binocular y hace clic sobre el citado binocular, todos los puntos del gráfico coincidentes con el valor introducido en el campo se iluminarán. Por ejemplo, el gráfico siguiente colorea de rojo todos los puntos correspondientes a la marca de automóvil Honda:

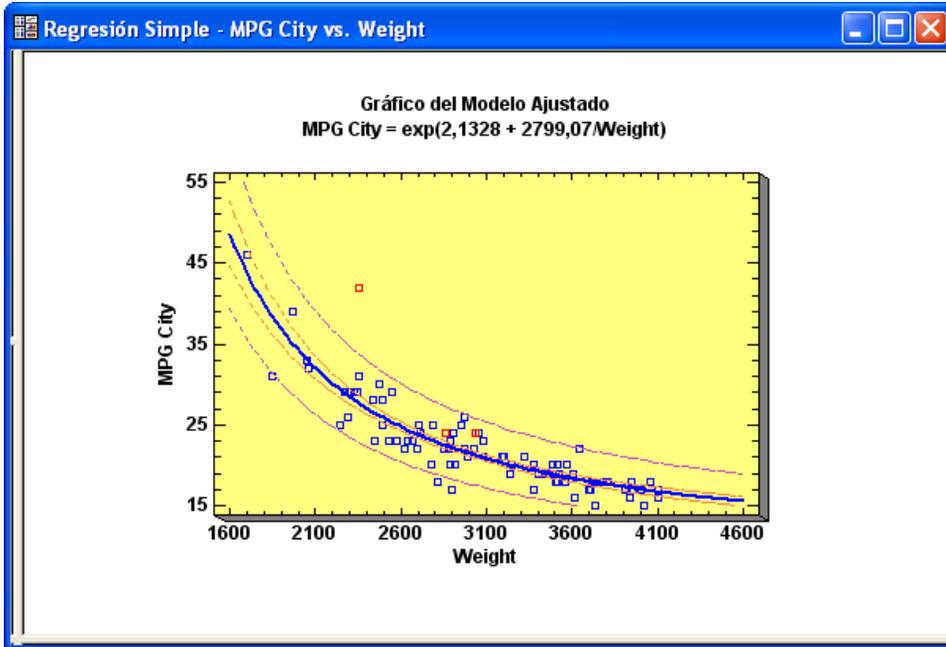


Figura 4-31. Gráfico iluminando todos los puntos de la marca Honda

Esta técnica es también efectiva en los gráficos matriciales. En el gráfico que se muestra a continuación, todos los puntos correspondientes a la fila #42 han sido iluminados:

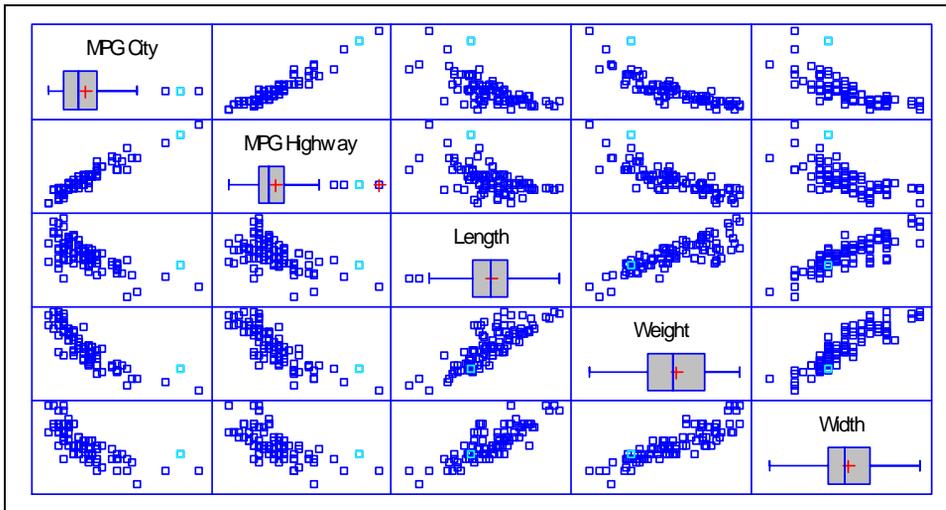


Figura 4-32. Gráfico matricial con la fila #42 iluminada

Localizar un punto en un gráfico matricial puede ayudar a decidir cuándo se trata de un atípico respecto a más de una variable.

NOTA: el color utilizado para iluminar los puntos se especifica en la pestaña *Gráficos* del cuadro de diálogo *Preferencias*, accesible en el menú *Edición*.

4.6 Copiando gráficos a otras aplicaciones

Una vez creado un gráfico en STATGRAPHICS Centurion XVI, puede copiarse fácilmente a otros programas como Microsoft Word o PowerPoint mediante los pasos siguientes:

1. Maximizando el panel que contiene el gráfico.
2. Seleccionando *Copiar* en el menú *Edición* de STATGRAPHICS Centurion XVI.
3. Seleccionando *Pegar* en la otra aplicación.

Por defecto, los gráficos se pegan en formato “Picture”, correspondiente a Windows metafile. En raros casos se realiza el pegado en otro formato. Se puede seleccionar *Pegado especial* en lugar de *Pegar*.

Para copiar un análisis completo en otra aplicación, incluyendo todas las tablas y gráficos, primero se copia el análisis a StatReporter utilizando el menú emergente del botón derecho del ratón, y después se copia desde StatReporter a la otra aplicación. Esta técnica se ilustra en el capítulo 7.

Para copiar gráficos incluidos en su ventana, como en la *figura 4-31*, se recomienda utilizar una herramienta de captura de terceras partes. Para realizar este manual, se ha utilizado un programa llamado *SnagIt*, disponible en la dirección www.techsmith.com. Si usa *SnagIt*, le recomendamos que sitúe la opción *Input* a “Window” y la opción *Output* a “Clipboard”. Entonces podrá pegar imágenes directamente en el documento.

4.7 Guardando gráficos en archivos de Imagen

Los gráficos individuales pueden también guardarse en archivos de Imagen maximizándolos y seleccionando *Guardar gráfico* en el menú *archivo*. Se mostrará un cuadro de diálogo en el cual se especificará el nombre del archivo y el formato de imagen:

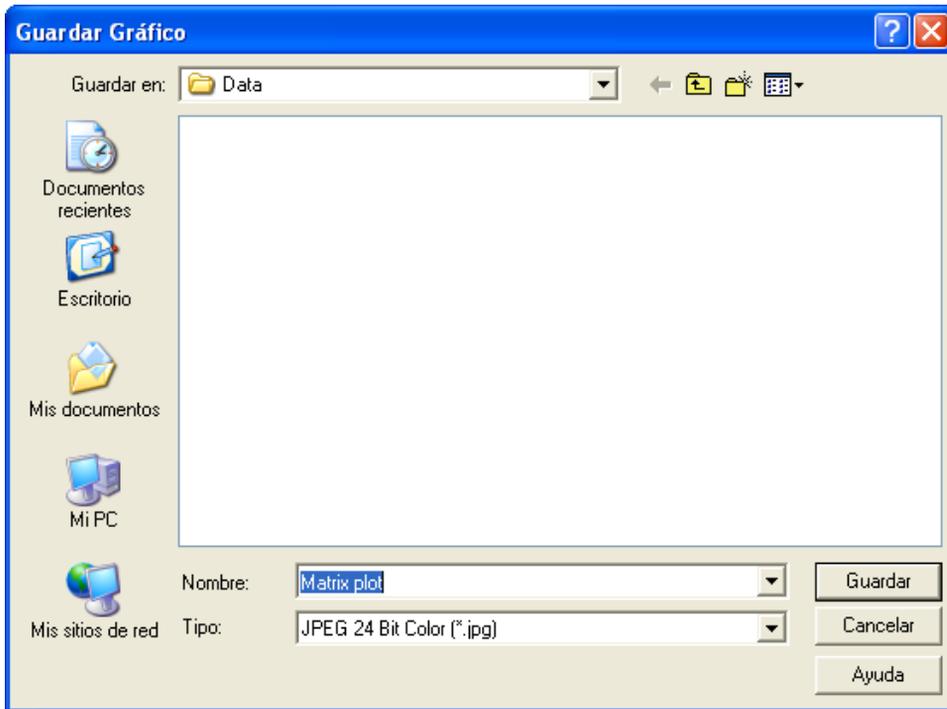


Figura 4-33. Cuadro de diálogo de selección de archivo para guardar gráficos en fichero imagen

Para guardar gráficos que posteriormente serán leídos por Word o PowerPoint, debe utilizarse el formato Windows metafile que da más flexibilidad. Si el gráfico hay que mostrarlo en una página Web, es conveniente guardarlo en formato JPEG.

StatFolios

Guardando su sesión, publicando resultados en formato HTML y automatizando el análisis utilizando código.

Cada vez que selecciona un análisis estadístico del menú de STATGRAPHICS Centurion XVI, se crea una nueva ventana de análisis. Puede guardar todas las ventanas de análisis de una vez creando un *StatFolio*. Un StatFolio es un archivo que contiene la definición de todos los análisis estadísticos que han sido creados, con punteros a los datos utilizados para ellos. Salvando un StatFolio y reabriéndolo posteriormente, efectivamente se guarda y se recupera la sesión actual de STATGRAPHICS Centurion XVI.

Cuando se guarda una sesión en un StatFolio, es la definición de los análisis lo que se ha guardado, no la salida. Cuando se reabre un StatFolio, se releen los datos de los orígenes de datos asociados y se recalculan todos los análisis. Los StatFolios constituyen un método simple para repetir análisis anteriores con datos diferentes.

Puede también crear código que se ejecuta cuando se lee el StatFolio. Detalles de esta y otras características del StatFolio se describen en este capítulo.

5.1 Guardando su sesión

Para guardar el estado actual de su sesión de STATGRAPHICS Centurion XVI, seleccionar *Archivo – Guardar – Guardar StatFolio* del menú principal. Introduzca un nombre para el StatFolio en el cuadro de diálogo que se muestra a continuación:

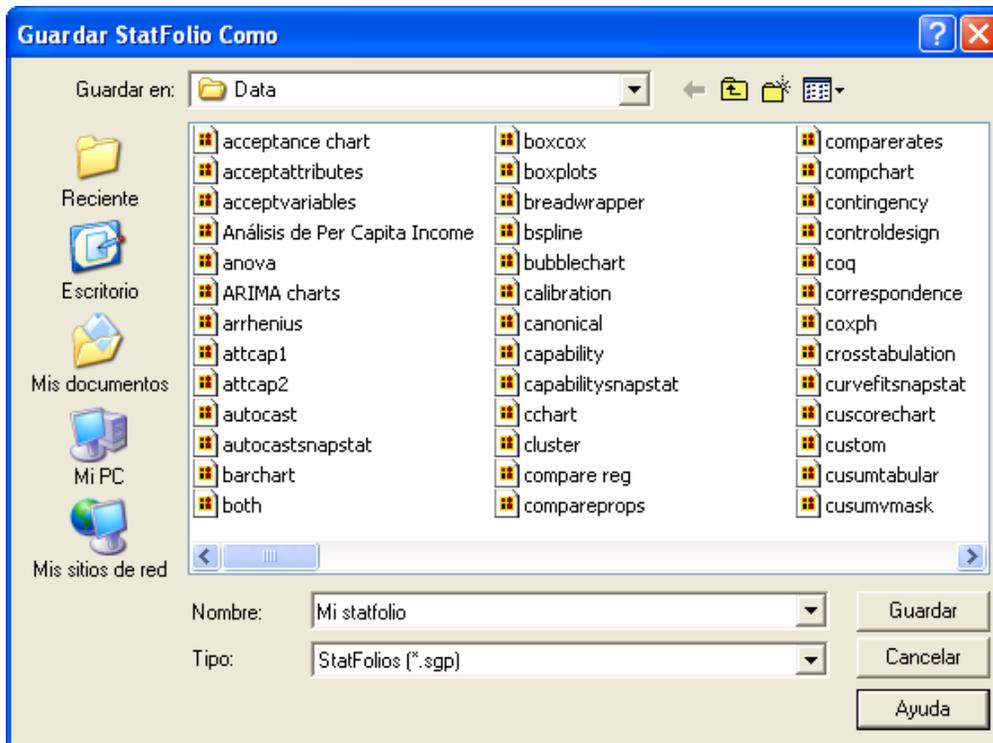


Figura 5-1. Cuadro de diálogo de selección de archivo para guardar StatFolio

Los StatFolios se guardan en archivos con extensión *.sgp*. Contienen:

1. Una definición de todos los análisis que han sido creados, incluyendo las variables de entrada, las tablas y gráficos, características de todas las opciones, cambios hechos en los gráficos, etc. Cuando se reabre un StatFolio, los análisis se recalculan y todas las tablas y gráficos se actualizan.
2. Enlaza los orígenes de datos contenidos en el libro de datos. Si los datos cambian desde el momento en que el StatFolio fue guardado hasta cuando se reabre, la ventana de análisis reflejará los cambios.
3. Enlaza el archivo de StatGallery y StatReporter, si se ha situado material en ellos antes de que el StatFolio fuese guardado. El programa le preguntará por los nombres ofrecidos para el StatGallery y el StatReporter cuando fue guardado el StatFolio.

5.2 Código de StatFolio

Cuando se lee un StatFolio, se restauran todas las ventanas de análisis a sus situaciones previas. STATGRAPHICS Centurion XVI procesa el código que ha sido salvado con el StatFolio y lo ejecuta. El código puede ser creado seleccionando *Rutina de Inicio del StatFolio* en el menú *Edición*. Se muestra un cuadro de diálogo con campos para definir una secuencia de acciones para ser ejecutadas:

Paso:	Operación:	Argumento:	Objetivo:
1	Ejecutar	Regresión Simple - MPG City vs. Weight	
2	Asignar	100*RESIDUOS/MPG City	PERROR
3	Ejecutar	Análisis de Una Variable - PERROR	
4	Imprimir	Todos los Análisis	
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			

Buttons: Aceptar, Ejecutar (F10), Insertar, Borrar, Cancelar, Ayuda

Figura 5-2. Código de Rutina de Inicio de un StatFolio simple

Las operaciones deseadas se especifican en el orden en que deben ser ejecutadas. Las operaciones disponibles son:

<i>Operation</i>	<i>Argument</i>	<i>Target</i>	<i>Description</i>
Execute	Título de análisis		Actualiza el análisis indicado.
Assign	Expresión de STATGRAPHICS Centurion	Nombre de columna	evalúa la expresión y le asigna la columna especificada
Print	Para imprimir ventana(s)		Imprime los contenidos de las ventanas indicadas.
Publish			Ejecuta StatPublish para publicar el contenido del StatFolio en formato HTML.
Shell	Comando Windows a ejecutar	Argumento del comando	Causa la ejecución de un comando Windows.
Delay	Número de segundos		Realiza una pausa por el tiempo especificado.
Load	Nombre del StatFolio		Especifica el StatFolio a leer después de ejecutar el código. Esto permite ejecutar StatFolios en cadena.
Exit			Sale de STATGRAPHICS Centurion XVI

Figura 5-3. Operadores de código de Rutina de Inicio

En el ejemplo que se muestra en la *figura 5-2*, se ejecuta una *Regresión Simple*. En el análisis se asume que *Guardar resultados* ha sido automáticamente configurado para guardar los residuos del modelo ajustado en la columna de nombre *RESIDUALS*. Los residuos se dividen por los valores originales de los datos y se multiplican por 100 para crear errores en porcentaje, que serán asignados a la nueva variable llamada *PERROR*. Los valores en *PERROR* son resumidos utilizando el procedimiento *Análisis de una variable* y después se imprimen los resultados de ambos análisis.

Obsérvese que los StatFolios pueden ser encadenados juntos utilizando el operador *LOAD* en un script (código) para leer y comenzar el script en otro StatFolio. Se puede también salir automáticamente de STATGRAPHICS Centurion XVI utilizando el operador *EXIT*.

NOTA: Puede suprimir la ejecución de los scripts seleccionado *Deshabilitar Rutina de Inicio* en la pestaña *General* del cuadro de diálogo *Preferencias*, accesible desde el menú *Edición*:

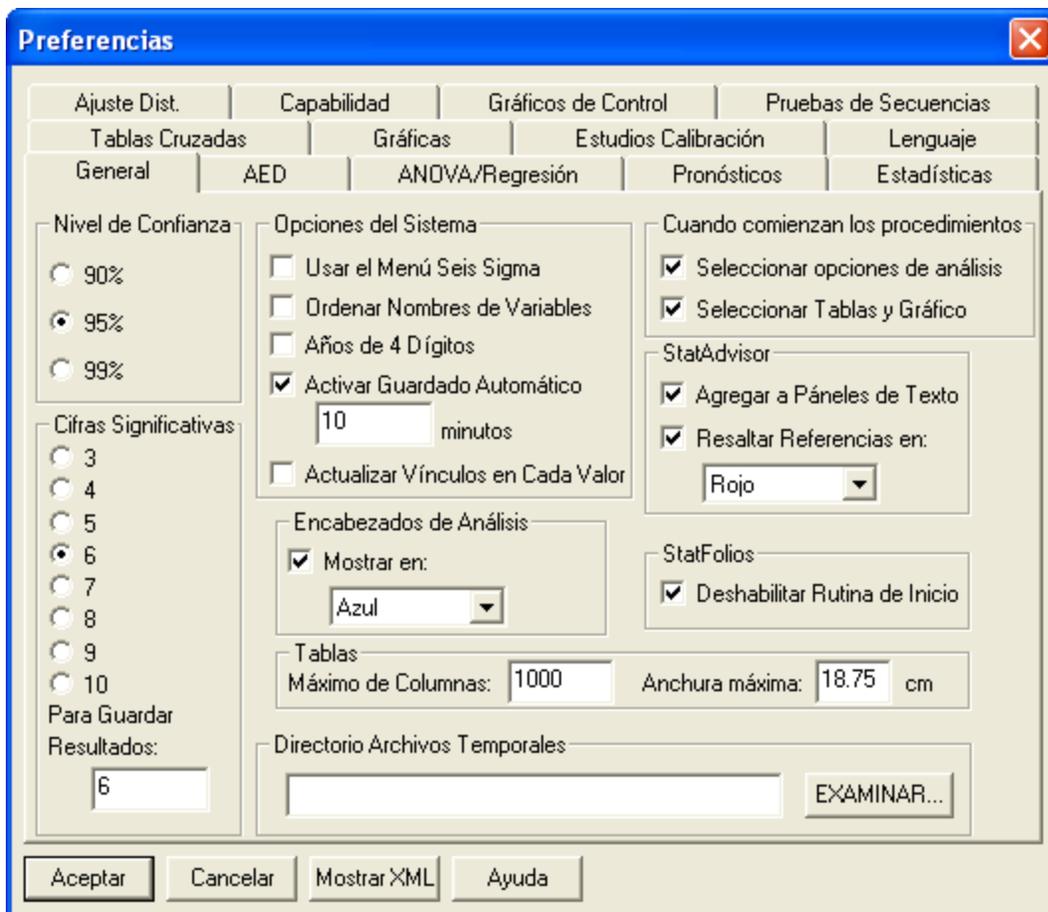


Figura 5-4. Deshabilitando Rutina de Inicio

5.3 Apilando orígenes de datos

Una vez que se ha creado el StatFolio conteniendo varios análisis, los datos de los orígenes de datos pueden ser releídos en un intervalo fijo de tiempo y todos los análisis actualizados. Esto se ha logrado utilizando el cuadro de diálogo *Propiedades del libro de datos* en el menú *Edición* o seleccionando *StatLink* en el menú *Archivo*:

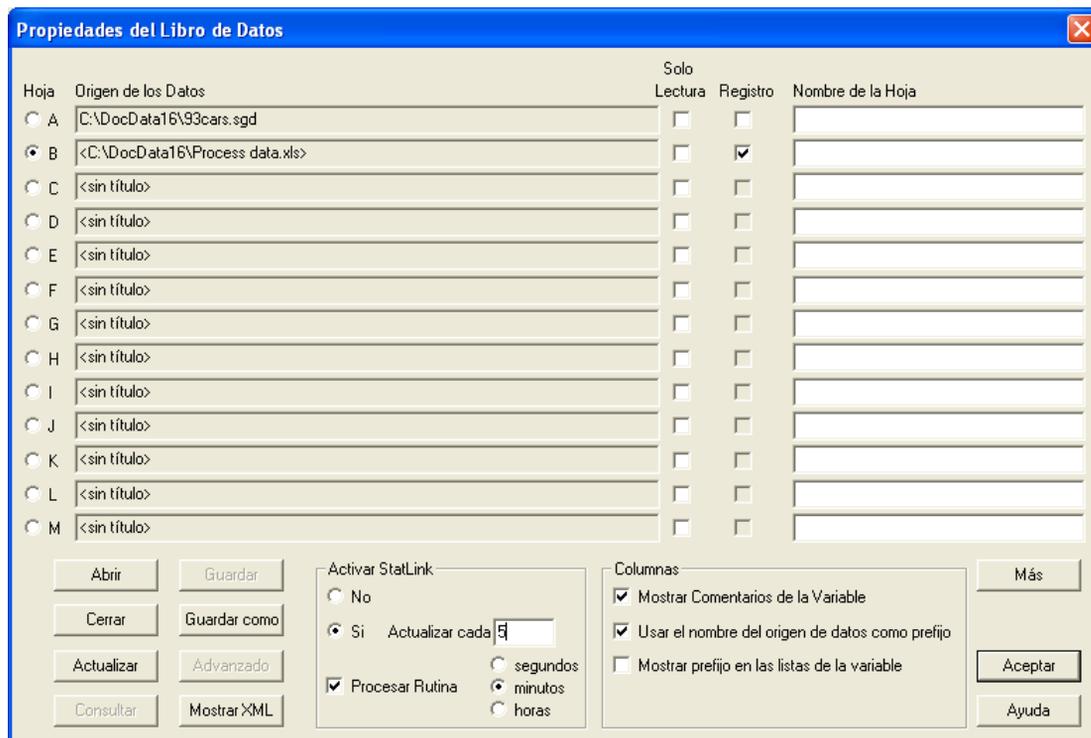


Figura 5-5. Cuadro de diálogo *Propiedades del libro de datos* para orígenes de datos apilados

para consultar los orígenes de datos repetitivamente:

1. Sitúe una marca en la caja *Poll* para cada origen de datos a releer.
2. Sitúe los botones del campo *Polling* en *On*.
3. Especifique la frecuencia de requerimiento de cada origen de datos.
4. Seleccione *Ejecutar Script* si quiere ejecutar el script del StatFolio cada vez que se releen los datos.

Incluyendo el paso *Publicar* en el script, puede tener actualizada automáticamente la salida de STATGRAPHICS Centurion XVI a un servidor de red.

5.4 Publicando datos en formato HTML

La salida de un StatFolio puede ser publicada en un formato que es visible utilizando solamente un navegador Web estándar seleccionando *StatPublish* en el menú *Archivo*. Se muestra el cuadro de diálogo para especificar la salida a publicar y dónde se va a ser situada:

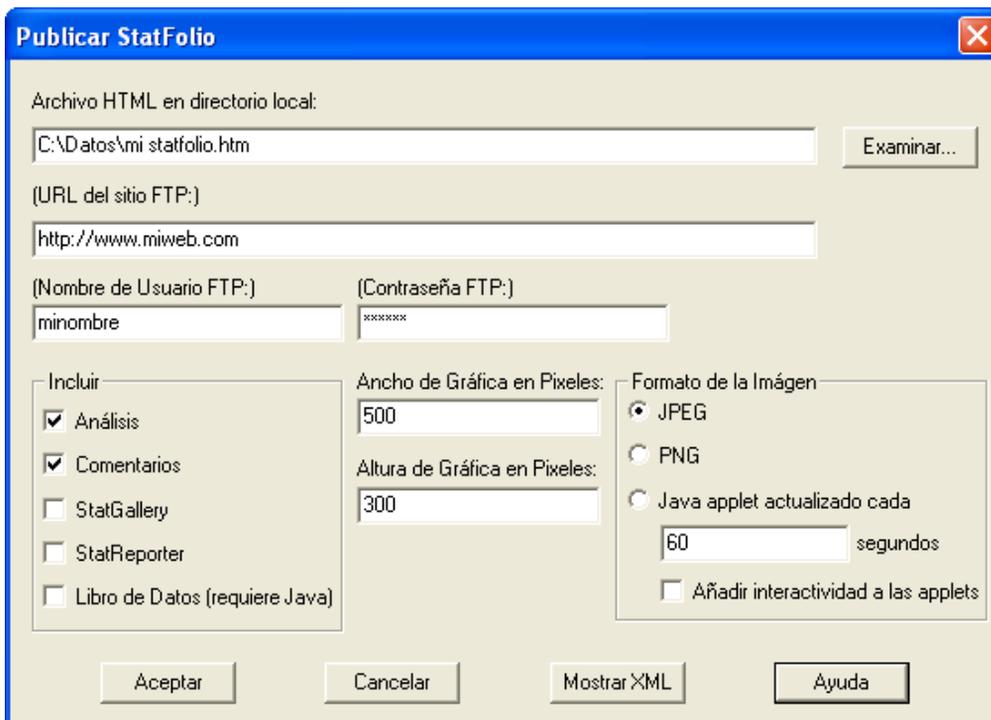


Figura 5-6. Cuadro de diálogo *StatPublish* para crear salida HTML

Los campos en este cuadro de diálogo se utilizan para especificar:

- **Archivo HTML en directorio local:** es el nombre del archivo HTML que se situará en Tabla de contenidos para el StatFolio. Listará el contenido del StatFolio y aportará los enlaces a otros archivos HTML correspondientes a cada ventana del StatFolio. Por defecto, se sitúa en el mismo directorio que el propio StatFolio, con el mismo nombre que el

StatFolio pero con la extensión *.htm* en vez de *.sgp*. Para ver un StatFolio publicado, puede abrirse el archivo con un navegador normal.

- **URL del Sitio FTP:** Todas las salidas publicadas se sitúan en primer lugar en el directorio local indicado arriba. Se incluyen archivos HTML, archivos imagen conteniendo los gráficos y otros archivos de soporte. Si se sitúa una entrada en el campo *URL del sitio FTP*, todos los archivos también se subirán a la localización de referencia de la dirección URL, que comúnmente será un directorio de un servidor. Tome nota de que es necesario tener acceso de escritura a la URL indicada en el sitio FTP, que tiene que ser concedido por el administrador de red.
- **Nombre de usuario FTP:** nombre de usuario para el acceso FTP a la URL indicada.
- **Contraseña FTP:** contraseña para el acceso FTP a la URL indicada.
- **Incluir:** Marcar toda la ventana del StatFolio que va a ser publicado.
- **Ancho de gráfico y altura en píxeles:** tamaño de los gráficos cuando son embebidos en archivos HTML.
- **Formato de imagen:** Los gráficos pueden ser embebidos en archivos HTML en uno de los tres formatos siguientes:
 1. *JPEG* – imagen estática guardada en formato JPEG. Los archivos se crean con nombres tales como *pubexample_analysis1_graph1.jpg*.
 2. *PNG* – imagen estática guardada en formato PNG. Los archivos son creados con nombres tales como *pubexample_analysis1_graph1.png*.
 3. *Applets Java* – salida dinámica que puede ser actualizada mientras se ve en el navegador. En el navegador, los gráficos serán actualizados con los incrementos especificados leídos de archivos auxiliares con un nombre tal como *pubexample_analysis1_graph1.sgz*. Esta opción se ha diseñado para ser utilizada en conjunto con el apilado de datos en tiempo real utilizando las características de StatLink, como se describe en el documento PDF titulado *Dynamic Data Processing and Analysis*. Nota: no todos los gráficos pueden ser publicados utilizando esta opción. Si uno o más gráficos no se muestran correctamente en la salida publicada, seleccione una opción diferente.

- **Añadir applets interactivamente:** Para gráficos publicados como applets, seleccionar las características permitidas para ver la información acerca de valores de datos haciendo clic en un punto con el ratón en el navegador Web.

Después de completar los campos de entrada, presione *Aceptar* para publicar el StatFolio.

Para ver un StatFolio publicado, arranque un navegador Web y utilice su menú *Archivo* para abrir el fichero especificado en el campo de la parte superior de la *figura 5-6*. Puede también ver la salida seleccionando *Ver resultados publicados* del menú *Archivo* de STATGRAPHICS Centurion XVI.

NOTA: Las tablas y gráficos son embebidos en archivos de salida HTML con nombres que son automáticamente generados por StatPublish. Mientras está en el navegador Web, puede ver el código HTML origen y determinar fácilmente los nombres de los archivos. Estos archivos pueden ser embebidos en su propia página Web si lo prefiere.

Utilizando StatGallery

Mostrado gráficos juntos y solapados.

StatGallery es una ventana especial en STATGRAPHICS Centurion XVI en la que se pegan gráficos creados con otros procedimientos uno al lado de otro (juntos) o uno encima del otro (solapados). Las comparaciones de gráficos juntos son una potente herramienta para comparar dos conjuntos de datos, dos modelos estadísticos, o dos niveles de un gráfico de contorno. Superponiendo gráficos se crea una única imagen imposible de conseguir con el sistema.

La salida de StatGallery se guarda en archivos con extensión `.sgg`. Si sitúa la salida en StatGallery, se creará un puntero al archivo StatGallery al guardar el StatFolio actual. Cuando se reabre el StatFolio posteriormente, automáticamente se leerá el StatGallery asociado.

6.1 Configurando una página de StatGallery

StatGallery se sitúa en una galería separada que se crea en STATGRAPHICS Centurion XVI la primera vez que se abre. Consiste en una o más páginas capaces de mostrar hasta 9 gráficos. Por defecto, cada página de la galería está configurada para 4 gráficos, como se muestra a continuación:

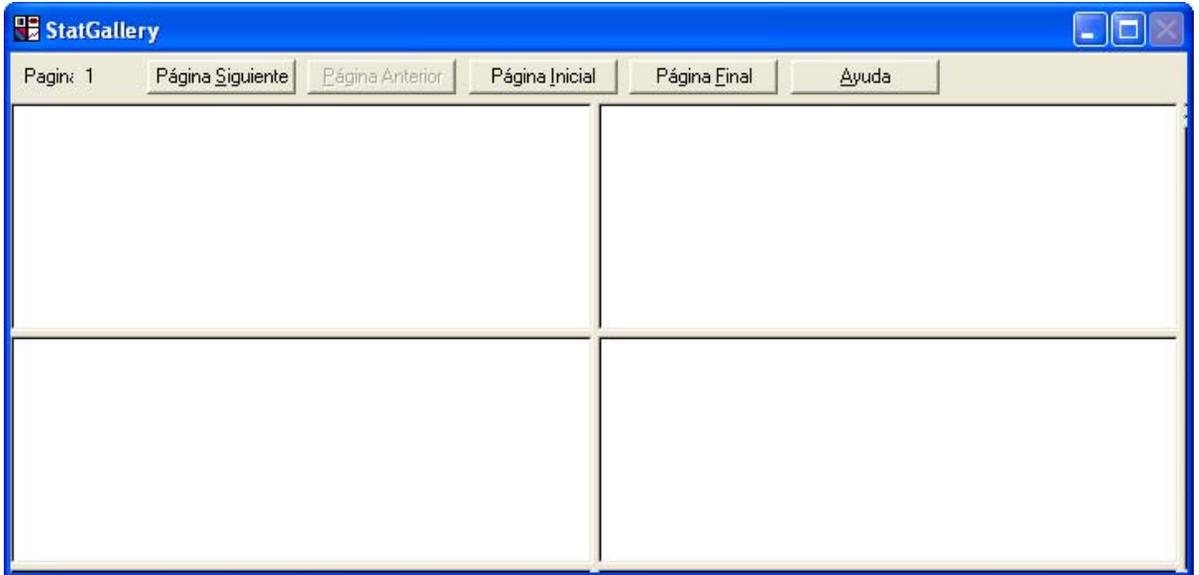


Figura 6-1. Ventana StatGallery

Los botones de la parte superior e la ventana permiten navegar por otras páginas de la galería. Si quiere cambiar el número de gráficos mostrados en una página, presione el botón derecho del ratón y seleccione *Organizar Ventanas*. Pueden seleccionarse arreglos conteniendo hasta 9 gráficos para una página simple:



Figura 6-2. Alternativas de configuración de la página de StatGallery

Los siete arreglos de la izquierda corresponden a conjuntos rectangulares de filas y columnas. La opción *Por columnas* permite crear un arreglo con diferente número de filas en cada una de las tres columnas.

Puede también utilizar barras deslizantes en la ventana StatGallery para mover los paneles obteniendo el arreglo que desee.

6.2 Copiando gráficos a StatGallery

Para situar un gráfico en StatGallery, debe primero copiarlo en el portapapeles de Windows desde la ventana de análisis donde ha sido creado. Por ejemplo, supongamos que se quiere mostrar gráficos de contorno creados con el procedimiento DOE *Analizar diseño* para dos niveles diferentes de un factor experimental seleccionado. Los pasos son los siguientes:

1. Configure la página seleccionada de StatGallery para mostrar gráficos en un formato *Izquierda y Derecha*.
2. Genere un gráfico de contorno con *Analizar diseño* para un nivel de un factor experimental y cópielo en el portapapeles de Windows.
3. Active la ventana StatGallery. Haga clic en el panel situado más a la izquierda con el botón derecho del ratón y seleccione *Pegar* en el menú emergente resultante para colocar el gráfico de contorno en StatGallery.
4. Vuelva a la ventana *Analizar diseño* y genere un Segundo gráfico de contorno para un nivel diferente del factor experimental. Cópielo en el portapapeles de Windows.
5. Vuelva a la ventana StatGallery. Haga clic en el panel situado más a la derecha con el botón derecho del ratón y seleccione *Pegar* en el menú emergente resultante. El Segundo gráfico de contorno se situará en StatGallery al lado del primero.

La galería resultante es similar a la que se muestra a continuación.

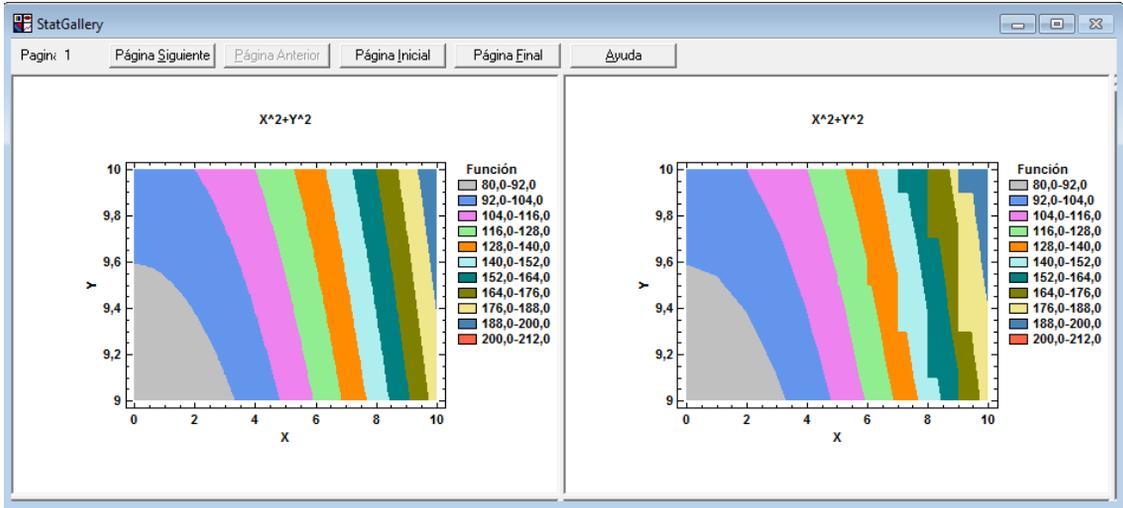


Figura 6-3. Gráficos en StatGallery situados uno al lado del otro (juntos)

En el gráfico anterior, la progresión de colores de un gráfico respecto del siguiente muestra un decrecimiento en intensidad con el decrecimiento del polietileno.

Cuando se pega un gráfico en StatGallery, debe seleccionar *Pegar enlace* en el menú emergente en lugar de *Pegar*. Cuando pega enlace, el gráfico de la galería es enlazado con la ventana de análisis en la que se creó originalmente y cambiará en StatGallery cuando cambie en la ventana original.

6.3 Solapando gráficos

Cuando se pega un gráfico en un panel de StatGallery que ya contiene otro, hay que elegir entre reemplazar el gráfico ya existente o solaparlo con el nuevo gráfico. El solapamiento de gráficos suele ser utilizado cuando se ajustan dos modelos estadísticos diferentes:

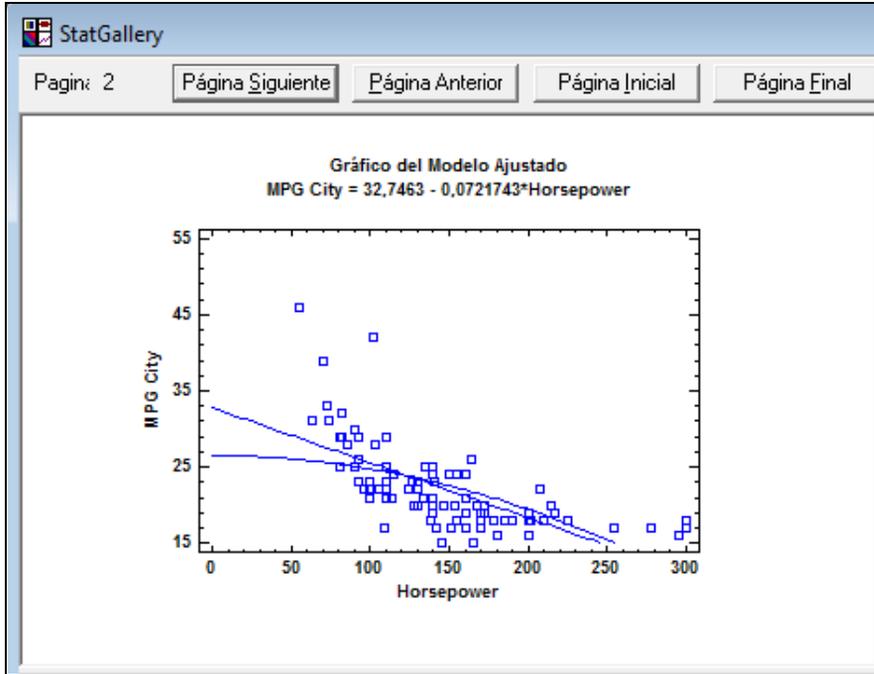


Figura 6-4. Gráficos solapados en StatGallery

Cuando un gráfico se solapa con otro que ya existe en StatGallery, sólo se añaden los contenidos del segundo no coincidentes con el primero. No se incluye el texto del segundo gráfico.

NOTA: Si el escalado del Segundo gráfico es diferente del escalado del primero, el Segundo gráfico ajustará su escala a la del primero.

6.4 Modificando un gráfico en StatGallery

Ciertos aspectos de un gráfico pueden cambiarse después de pegarlo en StatGallery.

6.4.1 Añadiendo ítems

Para añadir un ítem a un gráfico:

1. Haga doble clic en el gráfico deseado para maximizarlo en el panel.

2. Presione el botón derecho del ratón y seleccione *Añadir ítem en el menú emergente resultante*. Aparecerá el siguiente cuadro de diálogo flotante:



Figura 6-5. Cuadro de diálogo para Añadir un ítem

3. Seleccione el tipo de ítem que quiere añadir al gráfico.

Los primeros 5 botones del cuadro de diálogo de la *figura 6-5* trabajan presionando el botón del ratón sobre ellos y estirando la línea o figura hasta cubrir el área deseada sobre la que se utilizarán. El último botón activa el modo texto, de manera que el texto introducido en el cuadro de diálogo se muestra la siguiente vez que se haga clic en el gráfico. El texto añadido puede ser arrastrado a la localización deseada.

6.4.2 Modificando ítems

Para modificar un ítem en StatGallery:

1. Haga doble clic en el gráfico deseado para maximizarlo en su panel.
2. Haga clic con el ratón en el ítem que va a ser modificado para cambiarlo. Un pequeño bloque rectangular se situará alrededor del ítem que ha sido marcado.
3. Presione el botón derecho del ratón y elija *Modificar ítem* en el menú emergente resultante.

Aparecerá un cuadro de diálogo correspondiente al tipo de ítem marcado, en la cual se realizarán los cambios deseados.

6.4.3 Borrando ítems

Para borrar un ítem en StatGallery:

1. Haga doble clic en el gráfico deseado par maximizarlo en su panel.
2. Haga clic con el ratón en el ítem que va a ser borrado para marcarlo.

3. Presione el botón derecho del ratón y seleccione *Borrar ítem* en el menú emergente resultante.

6.5 Imprimiendo StatGallery

Para imprimir ítems de StatGallery:

1. Active la ventana de StatGallery haciendo clic sobre ella con el ratón.
2. Presione el icono *Imprimir* en la barra de herramientas principal, o presione el botón derecho del ratón y seleccione *Imprimir* en el menú emergente resultante.

Puede imprimir todas las páginas o seleccionar un conjunto de páginas.

Utilizando StatReporter

Copiando análisis a StatReporter, realizando anotaciones en la salida y guardando los resultados en un archivo RTF para importarlo en Microsoft Word.

StatReporter es una ventana en la cual pueden ser integradas las salidas de diferentes procedimientos estadísticos en un informe formal. Es una versión autónoma de WordPad, ejecutándose con STATGRAPHICS Centurion XVI. StatReporter permite:

1. Crear un informe completo en STATGRAPHICS Centurion XVI, sin necesidad de utilizar otra aplicación. Suele utilizarse mucho cuando los recursos son muy limitados, así como en planta de producción.
2. Guardar el contenido de StatReporter en un archivo RTF (Rich Text Format), que puede ser leído directamente en programas como Microsoft Word.

7.1 La ventana StatReporter

StatReporter está formado por una ventana separada en STATGRAPHICS Centurion XVI, creada automáticamente cuando el programa arranca. Consiste en un editor enriquecido, junto con una barra de herramientas:

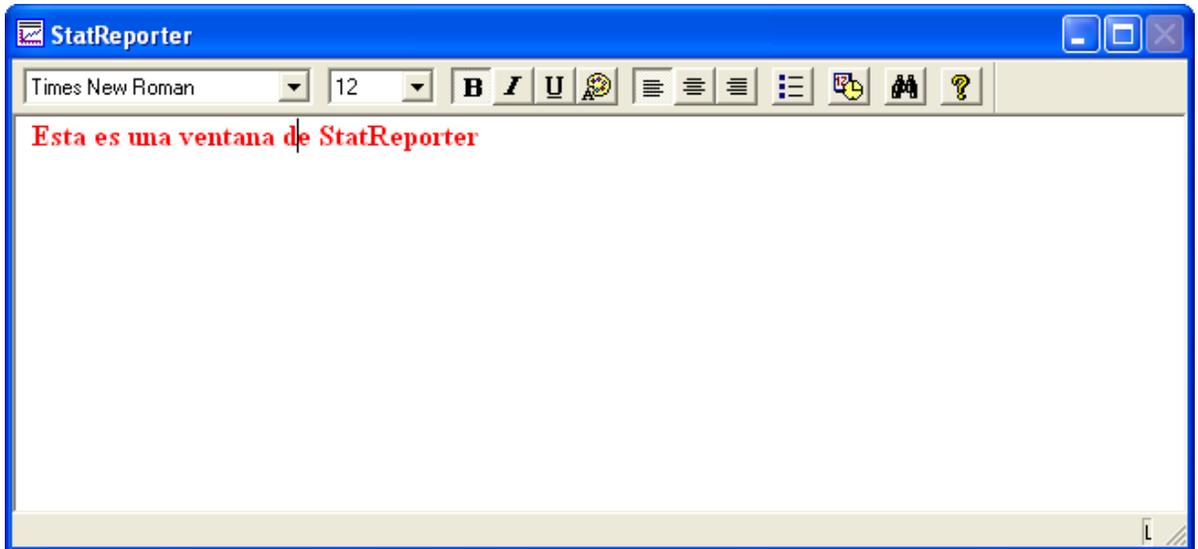


Figura 7-1. La ventana de StatReporter

Puede teclear texto en la ventana o pegar la salida creada en otro lugar de STATGRAPHICS Centurion XVI.

7.2 Copiando Salidas a StatReporter

STATGRAPHICS Centurion XVI proporciona tres métodos para copiar la salida a StatReporter:

1. Para copiar una tabla simple o un gráfico a StatReporter, en primer lugar se copia al portapapeles de Windows maximizando su panel y seleccionando *Copiar* en el menú *Edición*. A continuación nos situamos en la ventana de StatReporter, se coloca el cursor en la situación deseada, y se selecciona *Editar – Pegar*.
2. Alternativamente, se maximiza el panel que contiene la tabla o gráfico a mover haciendo doble clic en él. Presione el botón derecho del ratón y seleccione *Copiar panel a StatReporter* en el menú emergente resultante. Automáticamente la tabla o gráfico se pega en StatReporter en el lugar en el que el cursor está actualmente localizado.
3. Para copiar toda la salida de una ventana de análisis, presione el botón derecho del ratón y seleccione *Copiar análisis a StatReporter* en el menú emergente resultante. Todas las tablas y gráficos de la ventana de análisis se pegarán en StatReporter.

Cada una de las operaciones anteriores se realiza con un pegado estático (la salida en StatReporter nunca cambiará). Puede enlazar la tabla o gráfico a su origen, mediante el método #1 visto anteriormente seleccionando *Pegar enlace* en lugar de *Pegar*. La tabla o gráfico pegada en StatReporter será dinámico, en cuanto a que cambiará automáticamente cuando la salida origen cambie en la ventana de análisis de cualquier tabla o gráfico copiados.

7.3 Modificando la salida de StatReporter

La barra de herramientas de StatReporter permite modificar la salida una vez que ha sido situada en la ventana. Para cambiar texto, seleccione el texto que va a ser cambiado y pulse cualquiera de los botones de la barra de herramientas de StatReporter. También puede insertar la fecha y hora actuales presionando el botón *Fecha/hora*.

7.4 Guardando StatReporter

Para guardar la salida de StatReporter, seleccione *Archivo – Guardar – Guardar a StatReporter* en el menú principal e introduzca un nombre para el archivo que va a ser guardado. Todos los archivos guardados en StatReporter tienen extensión *.rtf* y pueden ser leídos directamente por Microsoft Word.

Siempre que se abre un StatFolio, automáticamente lee el StatReporter que estaba presente cuando el StatFolio fue guardado. Puede también abrir StatReporter independientemente utilizando el menú *Archivo – Abrir*.

Utilizando StatWizard

Seleccionando un análisis estadístico, buscando estadísticas y test deseados y generando ventanas múltiples para niveles de factor.

StatWizard es una característica especial de STATGRAPHICS Centurion XVI diseñada como un asistente en varias vías:

1. Puede ayudar a crear una nueva hoja de datos o leer un origen de datos existente.
2. Puede sugerir análisis basados en el tipo de datos que se están analizando.
3. Puede buscar los estadísticos que se deseen o test y considerar los procedimientos de análisis que los calculan.
4. Puede ayudar definiendo transformaciones de datos o seleccionando subconjuntos de datos.
5. Puede repetir análisis para cada valor único en una columna de datos.

Se puede acceder a StatWizard directamente presionando el botón  en la barra de herramientas principal.

8.1 Accediendo a Datos o Creando un nuevo estudio

Si el libro de datos está vacío cuando se active StatWizard, se muestra un cuadro de diálogo preguntando por sus necesidades de datos:

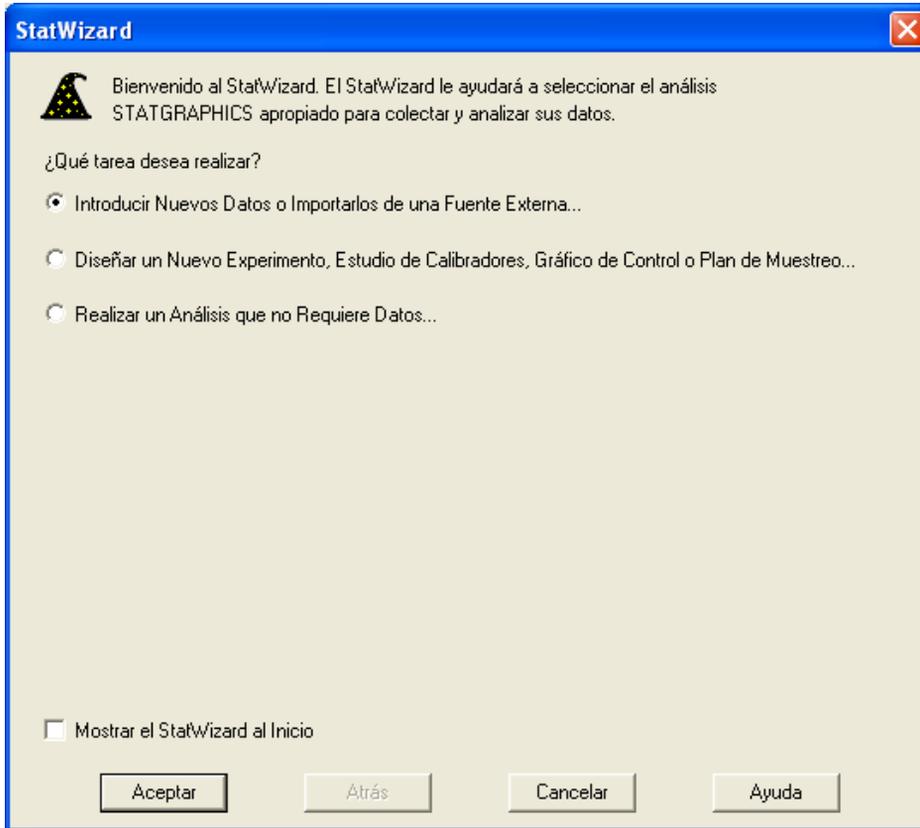


Figura 8-1. Cuadro de diálogo de entrada de datos en StatWizard

Existen tres opciones:

1. Puede necesitar leer los datos de un libro de datos de STATGRAPHICS Centurion XVI. El asistente los obtendrá por medio de una secuencia adicional de cuadros de diálogo en orden a definir las columnas de las hojas de datos o seleccionar un origen de datos, como se describe en capítulos posteriores de este manual.
2. Puede necesitar diseñar un nuevo estudio antes de obtener los datos. En este caso, el asistente preguntará que especifique el tipo de estudio a crear y los pasos a través de la secuencia de cuadros de diálogo en los cuales se define el estudio que va a ser creado.

3. Puede necesitar ejecutar un análisis que no requiere datos. En este caso, el asistente listará todos los análisis de este tipo, preguntará que seleccione uno, y ofrecerá el análisis.

Por ejemplo, supongamos que quiere realizar un nuevo estudio de medición para estimar la repetibilidad y reproductibilidad de los procesos de medida. Seleccionando el segundo botón de la *figura 8-1* y presionando *Aceptar* se obtienen las opciones que se muestran a continuación:

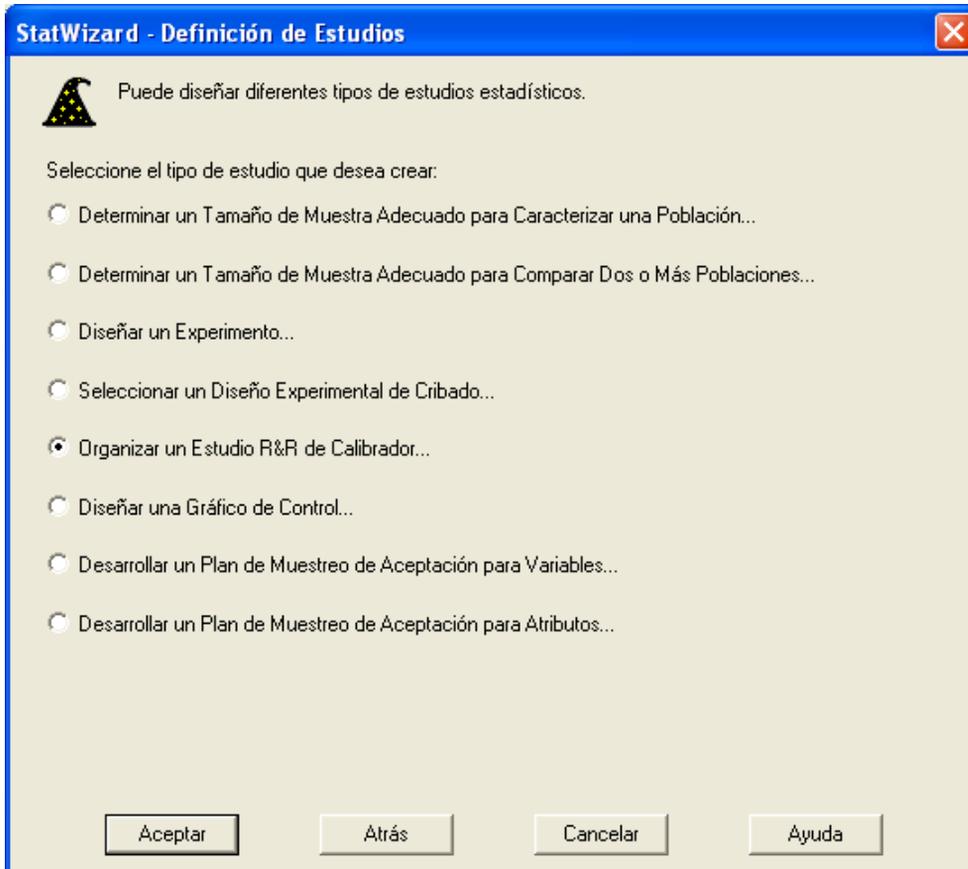


Figura 8-2. Cuadro de de diálogo de Definición del estudio en StatWizard

Seleccione *Realizar un estudio de Medición R&R* y presione *Aceptar* para mostrar un tercer cuadro de diálogo requiriendo información acerca del estudio:

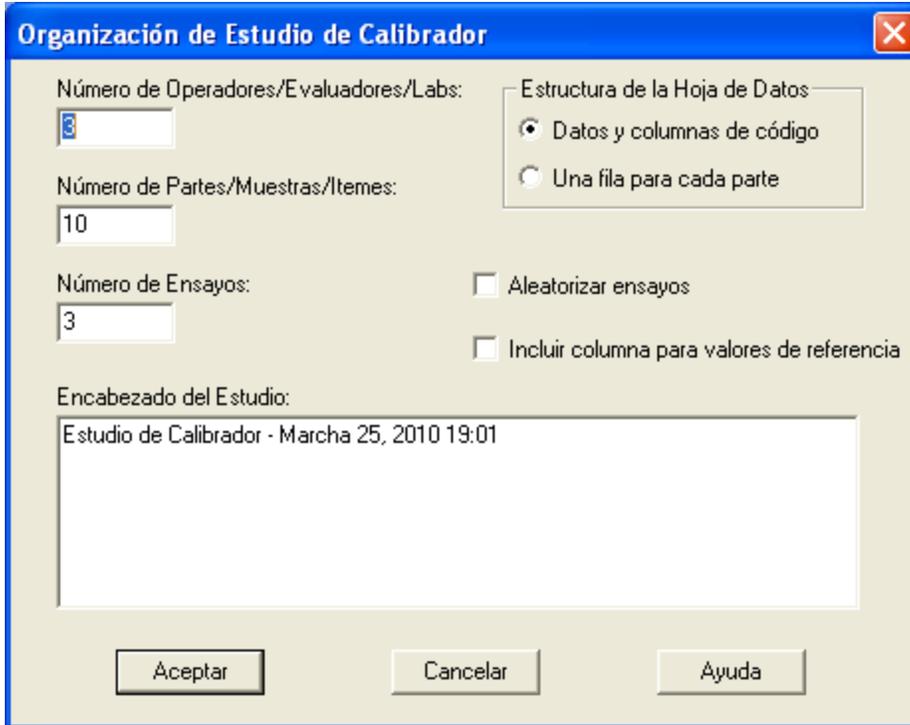


Figura 8-3. Cuadro de diálogo de Definición de estudio de medición en StatWizard

En el cuadro de diálogo, introduzca el número de operadores que serán involucrados en el estudio, el número de partes que serán medidas, y el número de veces que cada operador medirá cada parte. Puede especificar también una cabecera para el estudio.

El cuadro de diálogo final requiere nombres para los operadores, valoraciones, o laboratorios que realicen las mediciones:

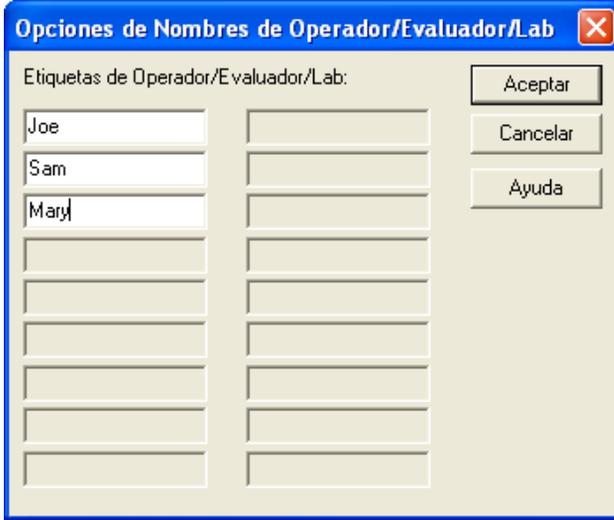


Figura 8-4. Cuadro de diálogo para especificar nombres de operadores

StatWizard crea el estudio deseado y lo sitúa en la hoja de datos del libro de datos:

	Operadores	Partes	Ensayos	Medidas	Encabezado
1	Joe	5	1		Estudio de Calibrado
2	Joe	4	1		
3	Joe	1	1		
4	Joe	10	1		
5	Joe	8	1		
6	Joe	2	1		
7	Joe	3	1		
8	Joe	6	1		
9	Joe	9	1		
10	Joe	7	1		
11	Joe	6	2		
12	Joe	1	2		

Figura 8-5. Estudio de medición creado por StatWizard

El estudio ejecutará las mediciones y las introducirá en la hoja de datos. Se podrá acceder a StatWizard otra vez al seleccionar un procedimiento de análisis (o podrá hacer directamente el análisis relevante en el menú principal).

8.2 Seleccionando análisis para sus datos

Si los datos han sido leídos de un libro de datos, tecleando el botón StatWizard se muestra el cuadro de diálogo del que se seleccionan uno o más análisis a ejecutar:

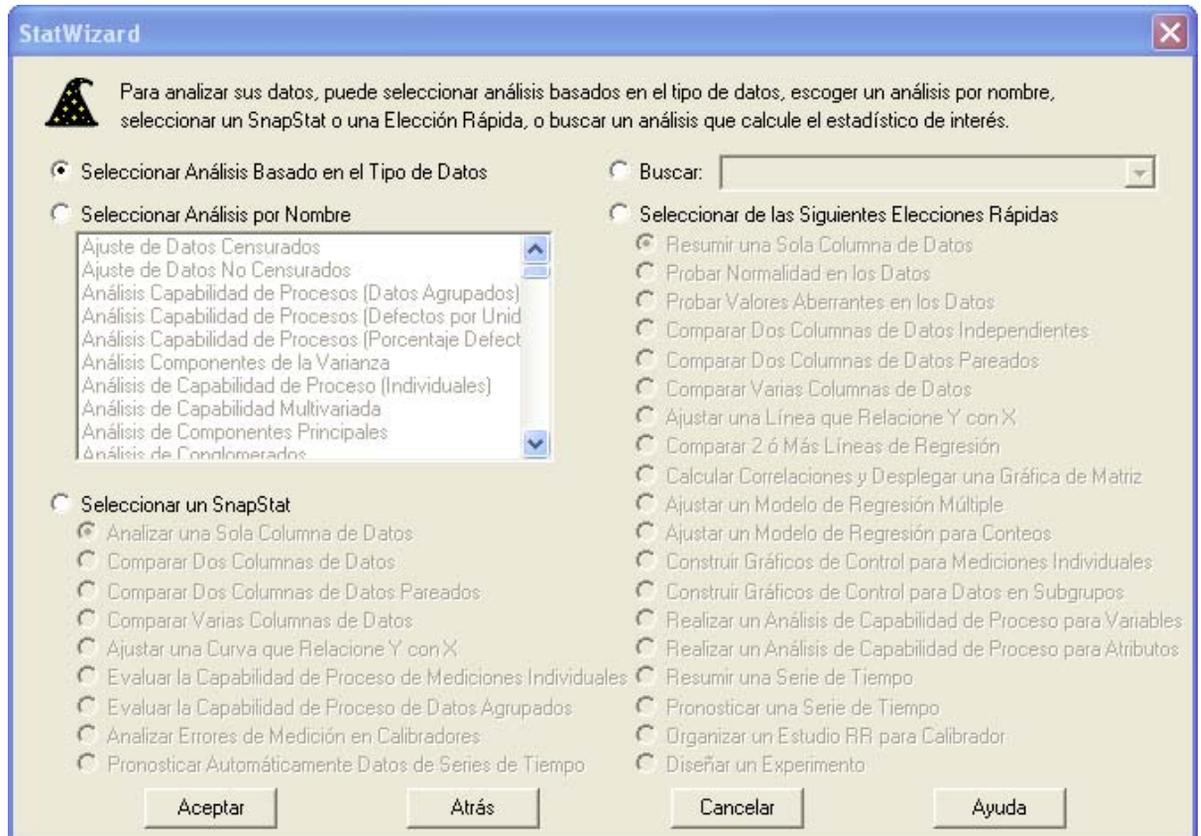


Figura 8-6. Cuadro de diálogo de StatWizard para Seleccionar análisis

Existen 5 opciones:

1. **Seleccionar un análisis basado en un tipo de datos:** Muestra cuadros de diálogo adicionales requiriendo información acerca de los datos a analizar, después de lo cual presenta una lista de procedimientos relevantes.
2. **Selecciona análisis por nombre:** Muestra todos los análisis en orden alfabético. Seleccionando un análisis por nombre y presionando *Aceptar* se introducen directamente los datos en el cuadro de diálogo de entrada de datos del análisis pasando los menús habituales.

3. **Seleccionando SnapStat:** permite seleccionar un SnapStat. SnapStats son análisis racionalizados que producen una página simple de salida pre formateada. Hay menos opciones que en otros análisis, pero son muy fáciles de crear.
4. **Búsqueda:** Muestra una lista de test, gráficos y otras salidas que pueden ser creadas en STATGRAPHICS Centurion XVI. Seleccionando un ítem de la lista cambia lo mostrado en el campo *Seleccionar análisis por nombre* para listar sólo los análisis que calculan el ítem deseado.
5. **Seleccionar rápido:** Lista algunos de los análisis más utilizados actualmente. Seleccione un análisis y presione *Aceptar* para ir directamente al cuadro de diálogo de entrada de datos para el análisis.

Si selecciona opción #1, el StatWizard mostrará a continuación un cuadro de diálogo en el que se indican los datos que van a ser analizados. Por ejemplo, si se lee el archivo *93cars.sgd* en el libro de datos, el cuadro de diálogo tomará la siguiente forma:

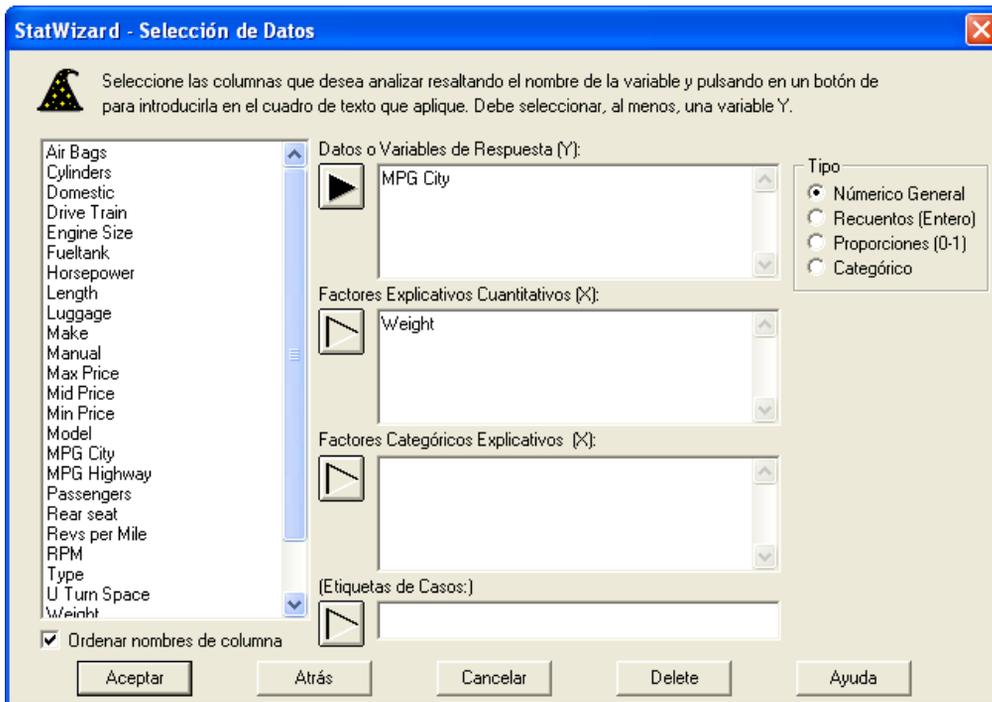


Figura 8-7. Cuadro de diálogo Selección de datos en StatWizard

Los campos de este cuadro de diálogo son:

- *Datos o variables respuesta (Y)*: una o más variables respuesta conteniendo los valores que van a ser analizados. Si solo hay una columna de datos para analizar, debe introducirse aquí.
- *Tipo*: el tipo de datos contenido en las variable(s) respuesta. Los análisis mostrados en los cuadros de diálogo siguientes dependen de esta elección.
- *Factores explicativos cuantitativos (X)*: factores cuantitativos que van ser usados para predecir las variables respuesta. En una regresión, aquí irán las variables independientes.
- *Factores Categóricos/Explicativos (X)*: factores no cuantitativos que van a ser utilizados para predecir la(s) variable(s) respuesta. En un ANOVA, aquí van los factores explicativos.
- *Etiquetas de casos*: una columna conteniendo etiquetas para cada una de las observaciones (filas).

Los procedimientos ofrecen los consiguientes cuadros de diálogo dependiendo de los datos introducidos

Figura 8-7.

El siguiente cuadro de diálogo pregunta qué filas del archivo se analizan:

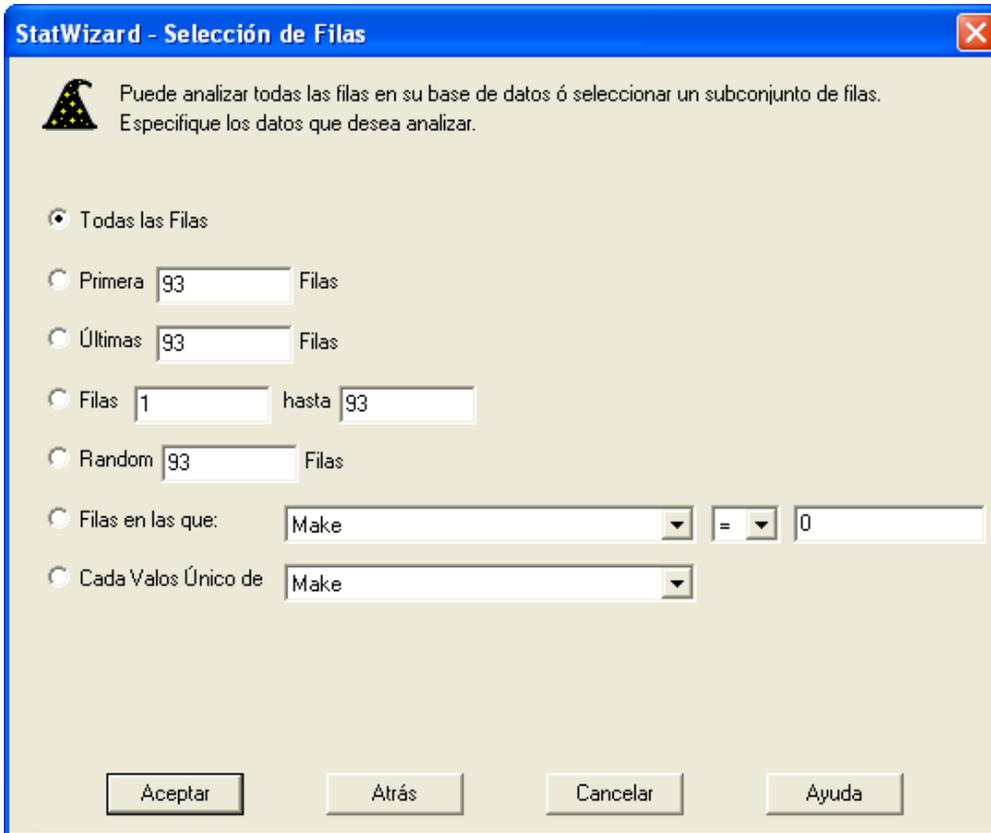


Figura 8-8. Cuadro de diálogo de Selección de filas en StatWizard

Las primeras seis opciones asumen que se quiere crear sólo un análisis simple. La última opción creará ventanas de análisis múltiples, uno para cada valor único contenido en la columna indicada. Este es un camino fácil para especificar una variable “BY” para el análisis.

A continuación se preguntará si quiere transformar alguna de las variables indicadas. Si responde afirmativamente, se muestra el siguiente cuadro de diálogo:

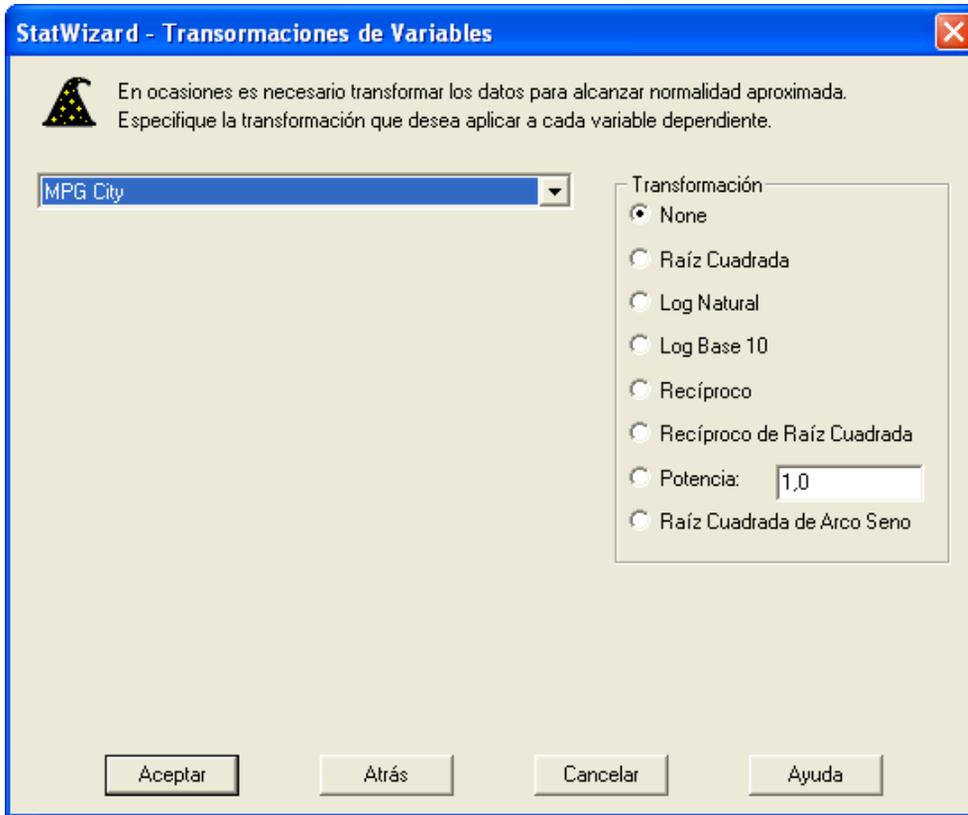


Figura 8-9. Cuadro de diálogo de Transformación de variable en StatWizard

Puede seleccionar una transformación para una o más variables. Si se requiere la transformación, se creará la expresión adecuada. Por ejemplo, si se requiere la raíz cuadrada de *MPG City* se creará la expresión $\text{SQRT}(\text{MPG City})$ para utilizar por los procedimientos de análisis.

El cuadro de diálogo final se mostrará listando todos los análisis apropiados para el tipo de datos que se han especificado:

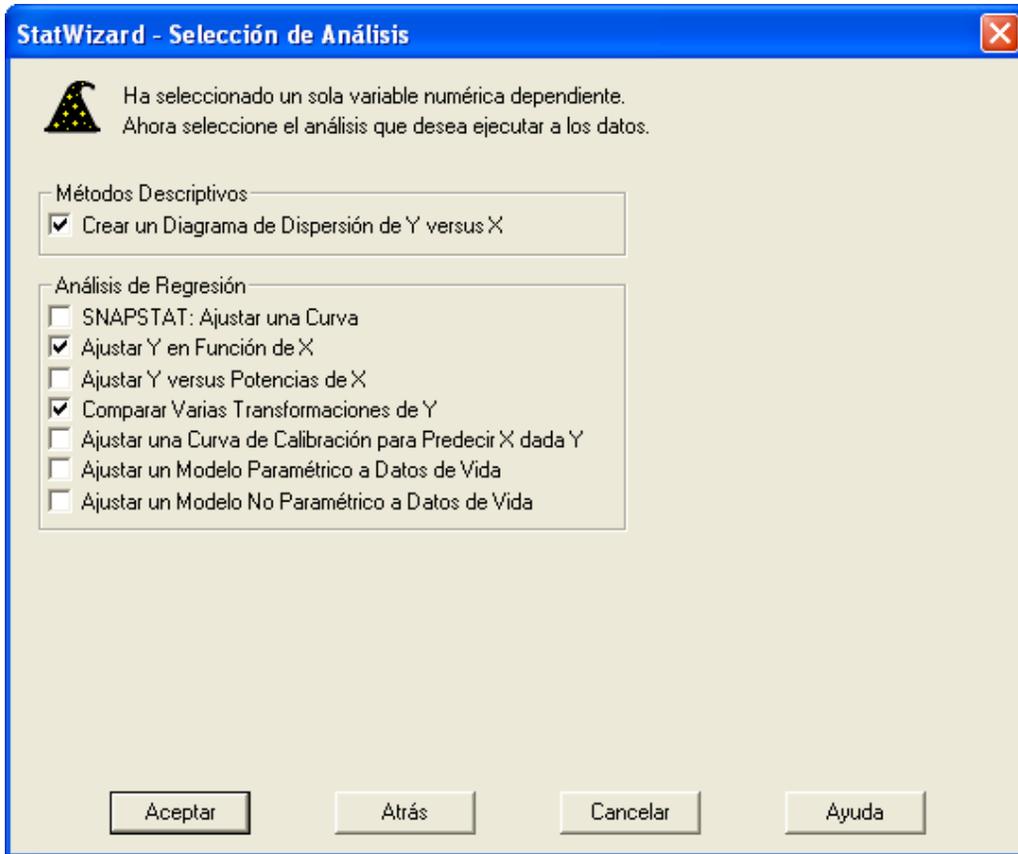


Figura 8-10. Cuadro de diálogo de Selección de análisis en StatWizard

Seleccione uno o más análisis de la lista. Cuando presione *Aceptar*, se creará una ventana de análisis para cada análisis seleccionado.

8.3 Buscando los estadísticos o test deseados

Si se quiere calcular un estadístico particular o realizar un test y no está seguro de qué análisis lo calcula, puede introducir sus datos en la hoja de datos y presionar el botón *StatWizard* en la barra de herramientas principal. En el cuadro de diálogo inicial de StatWizard, seleccione *Buscar* y descienda por la lista. Se mostrará una lista de todos los estadísticos, test y otros cálculos ejecutados por STATGRAPHICS Centurion XVI:

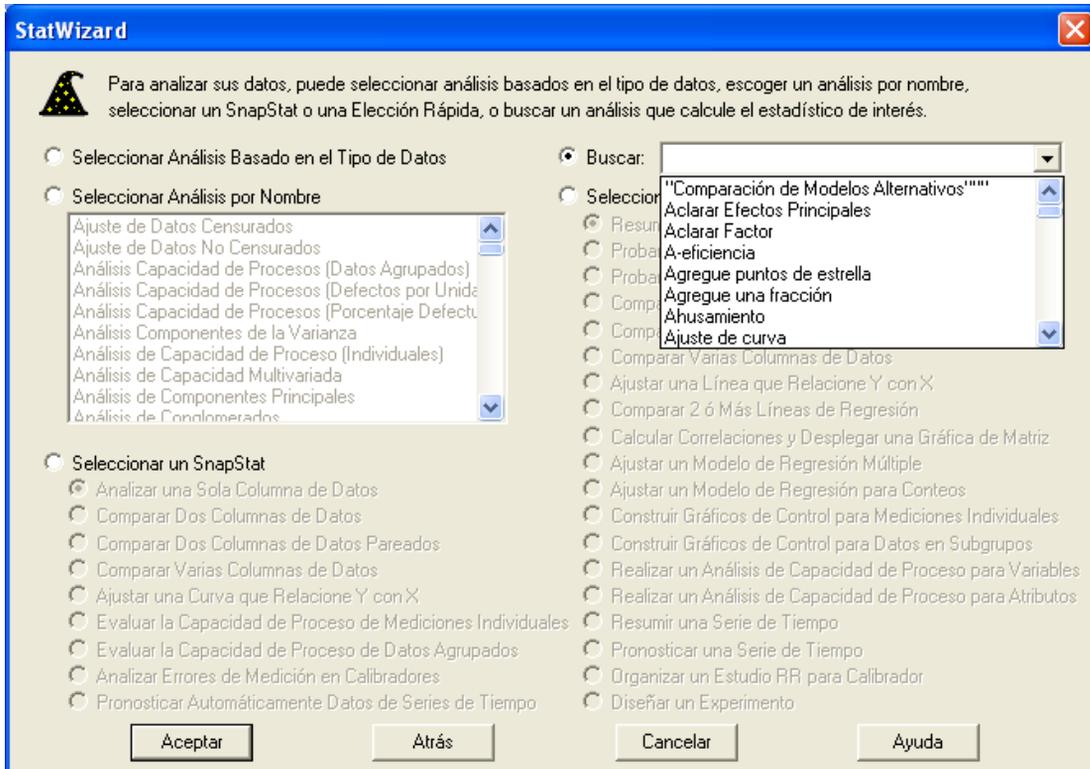


Figura 8-11. Utilizando la opción *Buscar* en StatWizard

Si selecciona un ítem de la lista, todos los análisis que calcula el ítem seleccionado se mostrarán en el campo *Seleccionar análisis por nombre*:

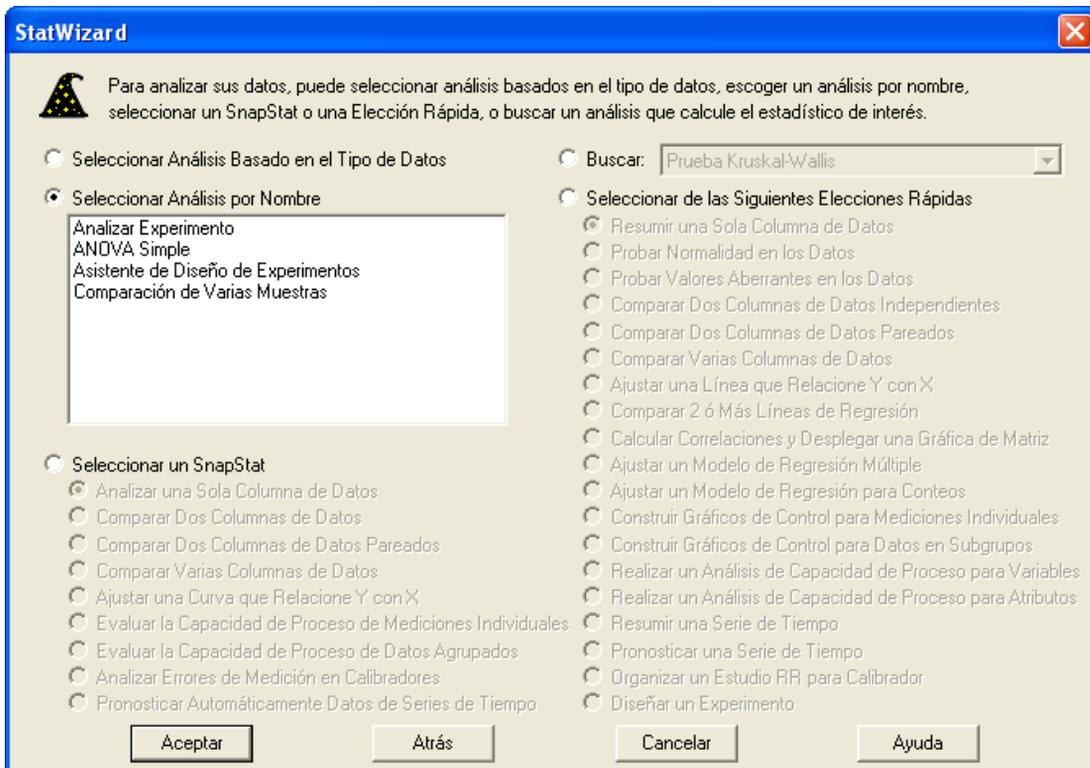


Figura 8-12. Lista de todos los análisis coincidentes con la opción buscada

Para ejecutar un análisis seleccionado:

1. Haga clic en el botón *Seleccione análisis por nombre*.
2. Ilumine un análisis.
3. Presione *Aceptar*.

Los datos se tomarán directamente del cuadro de diálogo de entrada de datos para el análisis seleccionado, pasando por los menús habituales.

Preferencias del sistema

Situando las preferencias para el comportamiento del sistema.

STATGRAPHICS Centurion XVI contiene cientos de opciones, cada una de las cuales tiene un valor por defecto que ha sido seleccionado para cubrir las necesidades de la mayoría de los usuarios. Si lo desea, puede situar nuevas opciones por defecto para la mayoría de las características. Hay 3 partes principales en el programa:

1. **Comportamiento general del sistema:** se sitúan en el cuadro de diálogo *Preferencias* accesible desde el menú *Edición*.
2. **Opciones de impresión:** se sitúan en el cuadro de diálogo *Configurar página* accesible en el menú *Archivo*.
3. **Gráficos:** se sitúan mientras se ve un gráfico seleccionando *Opciones gráficas*. La pestaña *Perfil* del cuadro de diálogo *Opciones gráficas* permite guardar múltiples conjuntos de atributos gráficos.

9.1 Comportamiento general del sistema

Los valores por defecto para el comportamiento general del sistema y los procedimientos estadísticos seleccionados pueden ser cambiados seleccionando *Preferencias* en el menú *Edición*. Se muestra un cuadro de diálogo con varias pestañas. Existe una pestaña *General* para el total de comportamientos del sistema y otras pestañas para opciones por defecto de análisis estadísticos:

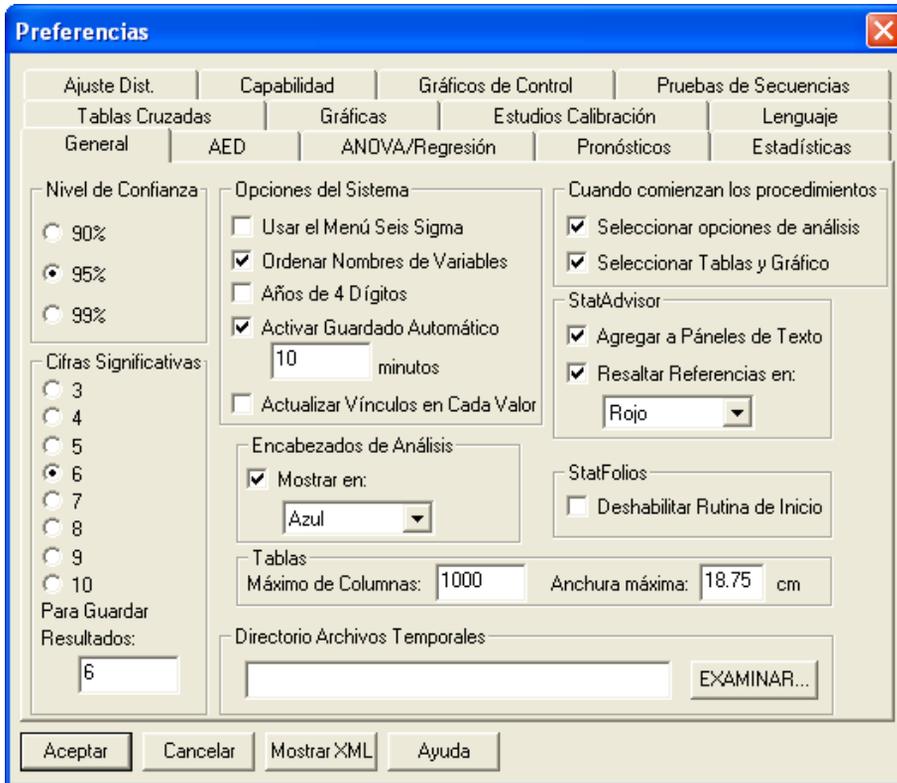


Figura 9-1. Cuadro de diálogo Preferencias

Algunas de las opciones más importantes a configurar son las siguientes:

- **Nivel de confianza:** porcentaje por defecto utilizado para los límites de confianza, límites de predicción, test de hipótesis e interpretación de P -valores por el StatAdvisor.
- **Dígitos significativos:** número de dígitos significativos utilizado cuando se muestran resultados numéricos. Se mostrará el número indicado de dígitos, excepto en el caso de completar con ceros que serán borrados. Se dispone de una entrada separada para guardar resultados numéricos de respaldo en la hoja de datos.
- **Opciones del sistema:** opciones que se aplican a todo el sistema.
 - **Utilizar menú Seis Sigma:** muestra las selecciones del menú bajo las cabeceras correspondientes a las fases de la metodología Seis Sigma DMAIC (Definir, Medir,

Analizar, Mejorar, Controlar). Las mismas selecciones son posibles con el menú clásico, excepto que se reordenarán bajo distintas cabeceras de menú.

- **Ordenar nombres de variables:** cuando se listan nombres de columnas en orden alfabético en los cuadros de diálogo de entrada de datos. En otro caso, los nombres de las columnas se listarán en el mismo orden en el que aparecen situados en la hoja de datos.
- **4-Dígitos para los años:** cuando las fechas deben mostrar 4 dígitos en los años en lugar de dos. Por defecto, se asumen dos dígitos para los años (2/1/05) para representar fechas entre 1950 y 2049. los cambios en esta opción no tendrán efecto hasta que no se reinicie el sistema.
- **Habilitar autoguardar:** cuando hay que guardar automáticamente en segundo plano el *StatFolio* actual y los archivos de datos y fijar el intervalo de tiempo entre dos acciones de guardado automático. Si se habilita esta opción y hay un mal funcionamiento de programa en el ordenador, será posible restaurar el estado actual del StatFolio y de las hojas de datos la próxima vez que se inicie el programa.
- **Actualizar enlaces en cada valor:** cuando se recalculan todos los estadísticos siempre que cambien los valores de datos en una de las hojas de datos. Normalmente, los estadísticos no se recalculan hasta que no se enfoca el análisis, se imprime o publica, o se guarda el StatFolio.
- **StatAdvisor:** sitúa el comportamiento por defecto del StatAdvisor.
 - **Añadir a paneles de texto:** cuando la salida del StatAdvisor se debe añadir automáticamente al final de los paneles de texto. La salida del StatAdvisor está siempre disponible mostrando su botón en la barra de herramientas principal (muestra un gorro de graduación).
 - **Iluminando referencias en ...:** cuando se iluminan con colores especiales valores destacados en los paneles de texto que se refieren al StatAdvisor.
- **Cabeceras de los análisis:** cuando se usa una fuente azul para mostrar el título de los análisis en la parte superior del panel de *resumen de análisis*.
- **StatFolios:** marque *Deshabilitar código* para prevenir que el código no se ejecute al leer StatFolios.

- **Directorio de archivos temporales:** Si se especifica, StatFolios, archivos de datos y otros archivos se escribirán en este directorio antes de ser copiados en su localización final. Especificando una unidad local, se puede aumentar la velocidad para salvar ficheros cuando se trabaja en red, a la vez que se reduce el número de requerimientos de la red.

Para una descripción detallada de las opciones de otras pestañas, referirse al documento PDF titulado *Preferencias*.

9.2 Imprimiendo

Pueden realizarse dos selecciones de control de la salida de impresión en el menú *Archivo*:

1. *Configurar impresión:* accede al cuadro de diálogo estándar de opciones de impresión ofrecido por el controlador de su propia impresora. Este cuadro configura el tamaño del papel y permite elegir entre modos *apaisado* y *normal* para la salida.
2. *Configurar página:* un cuadro de diálogo específico de STATGRAPHICS Centurion XVI permite configurar márgenes, cabeceras y otras opciones. Este cuadro de diálogo se trató en la sección 3.3.

9.3 Gráficos

Maximizando un panel que contiene un gráfico en una ventana de análisis se activa el botón *Opciones gráficas* en la barra de herramientas de análisis. Este botón muestra un cuadro de diálogo con pestañas que le permite cambiar la apariencia del gráfico, como se describe en el capítulo 4. También incluye en este cuadro de diálogo una pestaña etiquetada *Perfil*, que habilita la posibilidad de guardar conjuntos de atributos gráficos en un perfil de usuario y cambiar el perfil por defecto que se utiliza cuando se crea un nuevo gráfico:

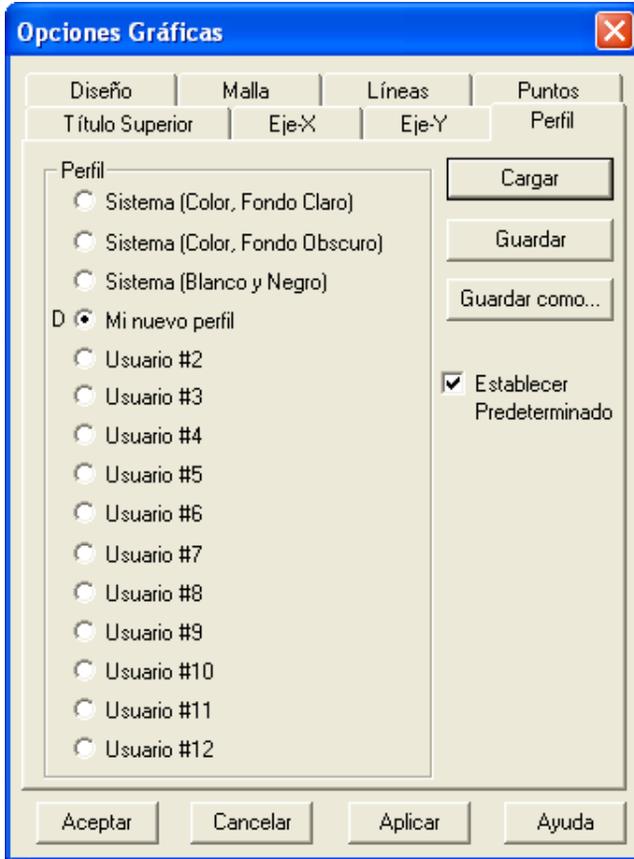


Figura 9-2. Pestaña Perfil en el cuadro de diálogo Opciones Gráficas

Para cambiar las opciones del sistema por defecto:

1. Modifique las características de un gráfico en una ventana de análisis. Sitúe colores, fuentes y otras opciones que quiera reflejar en gráficos futuros.
2. Seleccione *Opciones gráficas* de la barra de herramientas de análisis y vaya a la pestaña *Perfil*.
3. Marque *Hacer por defecto*.
4. Seleccione uno de los 12 perfiles de usuario y presione el botón *Guardar como* (los perfiles del sistema son de sólo lectura).
5. Introducir un nombre para el perfil que va a ser guardado:

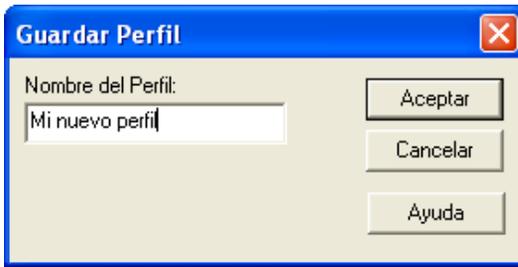


Figura 9-3. Cuadro de diálogo Guardar perfil

6. Presione *Aceptar* para guardar las opciones actuales de atributos gráficos (colores, fuentes, estilos de líneas y puntos, etc.) en un nuevo perfil.

El gráfico siguiente se creará utilizando el perfil guardado recientemente.

puede también aplicar otros perfiles guardados a un nuevo gráfico creando el gráfico con las opciones por defecto y a continuación:

1. Seleccione *opciones gráficas* de la barra de herramientas de análisis y vaya a la pestaña *Perfil*.
2. Seleccione uno de los 15 perfiles y presione el botón *Leer*.

El gráfico actual se actualizará automáticamente para reflejar las características del perfil seleccionado.

Tutorial #1: Analizando una muestra simple

Estadísticos resumen, histogramas, gráficos de caja y bigotes, intervalos de confianza y contrastes de hipótesis.

Un problema común en estadística es analizar una muestra de n observaciones tomada de una población simple. Por ejemplo, considere las siguientes temperaturas corporales tomadas para $n = 130$ individuos:

98.4	98.4	98.2	97.8	98	97.9	99	98.5	98.8	98
97.4	98.8	99.5	98	100.8	97.1	98	98.7	98.9	99
98.6	97.7	96.7	98.8	98.2	97.5	97.2	97.4	97.1	96.7
99.2	97.9	98.8	97.6	98.6	98.8	98.5	98.7	97.5	97.9
97.1	98.4	97.4	98.6	97.8	98.2	98	98	98.3	98.6
98.8	98.7	98.8	98.1	96.4	98.8	98.7	97.9	98.6	99.2
98.6	98	99.1	97.8	97.2	98.2	98.7	98.4	98.2	97.7
98.3	98.7	96.8	98	97.2	97.9	96.9	98.3	97.8	97
98.6	98.4	98.2	98	98	98.2	97.8	99	98.1	97.7
97.4	98.8	99.3	98.9	96.3	97.8	99.9	98.4	99.4	98.7
98.4	98.2	99.3	98.5	98.3	99	99.2	97.6	99.1	97.6
98.4	97.6	98.4	98	98.8	97.3	98.7	98.6	99.4	100
98.6	98.3	98.6	97.4	98.1	97.8	98.2	99	99.1	98.2

Los datos fueron obtenidos del archivo de datos del Journal of Statistical Education (www.amstat.org/publications/jse/jse_data_archive.html) y son utilizados con permiso. Han

sido situados en un archivo de nombre *bodytemp.sgd*, en una columna de nombre *Temperatura* que contiene 130 filas (una fila para cada individuo en el estudio).

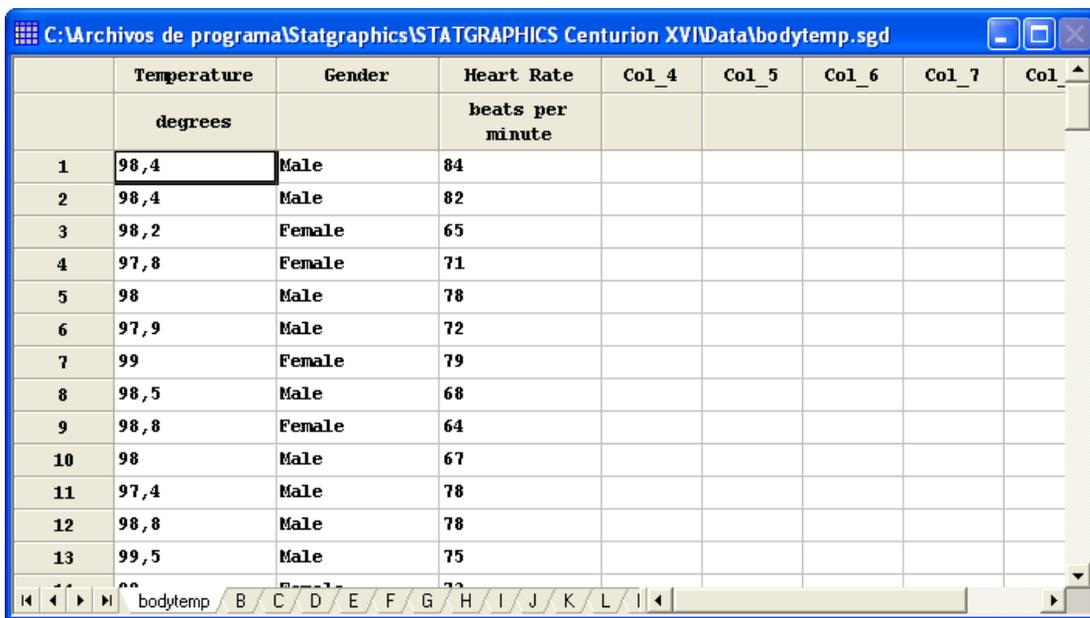
El primer procedimiento en STATGRAPHICS Centurion XVI para resumir la muestra tomada de la población es el procedimiento *Análisis de una variable*. Este procedimiento resume los datos en forma numérica y gráfica y realiza contrastes de hipótesis acerca de la media de la población, mediana y desviación típica.

10.1 Ejecutando el procedimiento Análisis de una variable

Para analizar los datos de la temperatura corporal, primero se lee el archivo *bodytemp.sgd* llevando a cabo las acciones siguientes:

1. Seleccione *Archivo – Abrir – Abrir origen de datos* del menú principal.
2. En el cuadro de diálogo *Abrir origen de datos*, indique que quiere abrir un archivo de datos de *STATGRAPHICS*.
3. Seleccione *bodytemp.sgd* de la lista de archivos en el cuadro de diálogo *Abrir archivo de datos*.

Los datos deben aparecer como se muestra a continuación:



	Temperature	Gender	Heart Rate	Col_4	Col_5	Col_6	Col_7	Col_8
	degrees		beats per minute					
1	98,4	Male	84					
2	98,4	Male	82					
3	98,2	Female	65					
4	97,8	Female	71					
5	98	Male	78					
6	97,9	Male	72					
7	99	Female	79					
8	98,5	Male	68					
9	98,8	Female	64					
10	98	Male	67					
11	97,4	Male	78					
12	98,8	Male	78					
13	99,5	Male	75					

Figura 10-1. Hoja de datos con los datos de temperatura corporal

Las temperaturas corporales están en la columna de la izquierda, medidas en grados Fahrenheit.

Se accede al procedimiento *Análisis de una variable* del menú principal como sigue:

1. Si usa el menú clásico, seleccione *Describir – Datos numéricos – Análisis de una variable*.
2. Si usa el menú Seis Sigma, seleccione *Análizar – Datos de variables – Análisis de una variable*.

En el cuadro de diálogo de entrada de datos, indique la columna a analizar:



Figura 10-2. Cuadro de diálogo de entrada de datos de *Análisis de una variable*

Deje en blanco el campo *Seleccionar* para analizar las 130 filas. Presione *Aceptar*.

Cuando se presiona *Aceptar*, aparecerá la ventana *Tablas y Gráficos* que muestra las tablas y gráficos disponibles. Por ahora, se aceptarán las características por defecto.

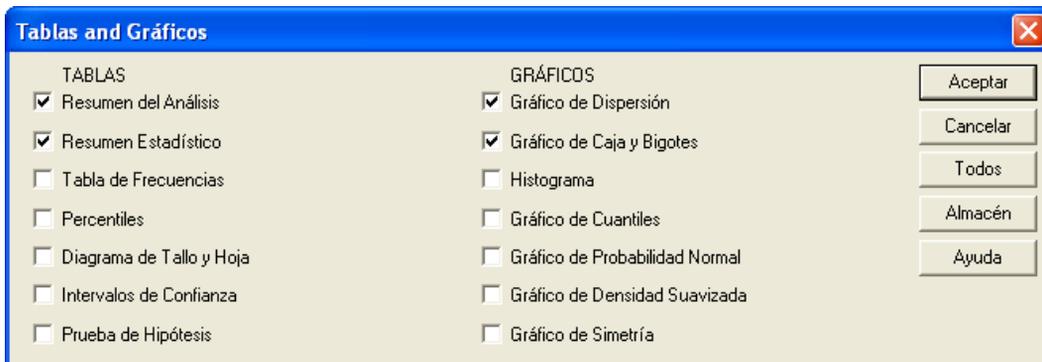


Figura 10-3. Cuadro de diálogo *Tablas y Gráficos*

Aparecerá la ventana de análisis con cuatro paneles:

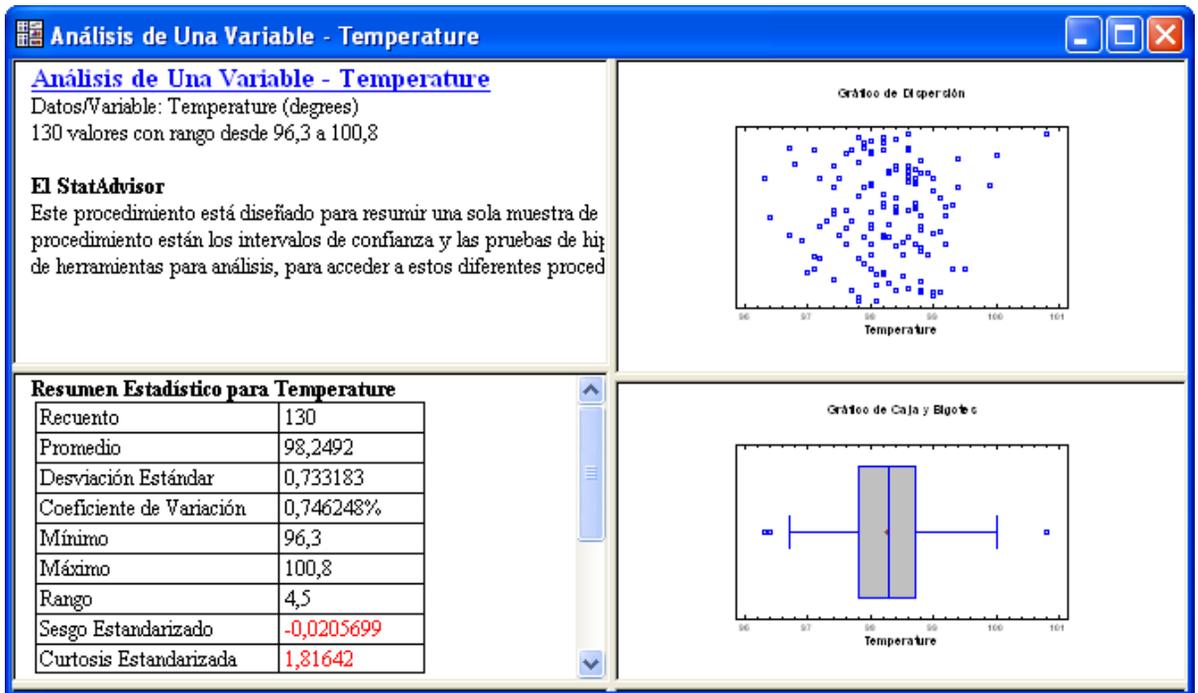


Figura 10-4. Ventana Análisis de una variable

La parte superior izquierda del panel indica que la muestra tiene $n = 130$ valores situados entre 96.3 y 100.8 grados. La parte superior derecha del panel muestra un gráfico de dispersión de los datos, con los puntos aleatoriamente dispersos en la dirección vertical. Tome nota de que los puntos tienen más densidad entre 98 y 99 grados, disminuyendo fuera. Este tipo de comportamiento es típico de datos que son muestreados de una población cuya distribución tiene bien definido un pico central.

Los paneles inferiores muestran resúmenes estadísticos y el gráfico de caja y bigotes, descrito en las secciones siguientes.

10.2 Resúmenes estadísticos

La tabla de la parte inferior izquierda del panel muestra estadísticos simples. Adicionalmente pueden añadirse estadísticos maximizando el panel (doble clic en él con el ratón) y seleccionando *Opciones de ventana*:

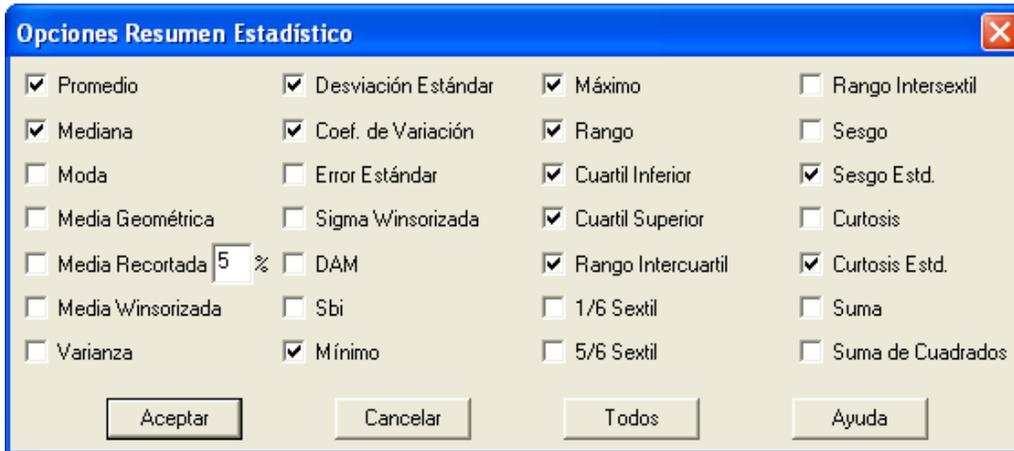


Figura 10-5. Cuadro de diálogo Opciones de resumen estadístico

Incluyendo la mediana simple, cuartiles y el rango intercuartílico resulta:

Resúmenes estadísticos para Temperatura	
Tamaño	130
Media	98.2492
Mediana	98.3
Desviación típica	0.733183
Coeficiente de variación	0.746248%
Mínimo	96.3
Máximo	100.8
Rango	4.5
Cuartil inferior	97.8
Cuartil superior	98.7
Rango intercuartílico	0.9
Asimetría estandarizada	-0.0205699
Curtosis estandarizada	1.81642

Figura 10-6. Tabla de resúmenes estadísticos

Una asunción común para medidas de datos es que provienen de una distribución normal. Los datos de una distribución normal están completamente descritos por dos estadísticos:

1. La media (*mean* o *average*) $\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 98.25$, que estima el centro de la distribución.

2. La desviación típica muestral (*standard deviation*) $s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = 0.733$, que explica la variabilidad de la distribución.

Para una distribución normal, aproximadamente el 68% de los valores se cubrirán con una desviación típica de valor uno, aproximadamente el 95% con una desviación típica de valor dos y aproximadamente el 99.73% con una desviación típica de valor tres.

La media muestral y la desviación típica describen completamente la muestra sólo si provienen de una distribución normal. Dos estadísticos que se pueden utilizar para chequear la asunción de normalidad son la asimetría y la curtosis normalizadas. Estos estadísticos miden la forma:

1. *Asimetría* mide la simetría o la ausencia de la misma. Una distribución simétrica, tal como la normal tiene asimetría cero. Distribuciones para las cuales los valores de su asimetría estén por encima de cero presentan asimetría positiva. Distribuciones para las cuales los valores de su asimetría estén por debajo de cero presentan asimetría negativa.
2. *Curtosis* mide el apuntamiento de la distribución. Una distribución normal tiene curtosis nula. Una distribución más apuntada que la normal tiene curtosis positiva. Una distribución menos apuntada que la normal tiene curtosis negativa.

Si los datos provienen de una distribución normal, los coeficientes de asimetría y curtosis estandarizados deben de estar entre -2 y +2. En este caso, la distribución normal parece ser un modelo razonable para los datos.

Otro resumen habitual de los datos lo aportan los 5 números resumen de John Tukey:

Mínimo (valor más pequeño de los datos) = 96.3

Cuartil inferior (percentil 25) = 97.8

Mediana (percentil 50) = 98.3

Cuartil superior (percentil 75) = 98.7

Máximo (valor más alto de los datos) = 100.8

Estos cinco números dividen la muestra en cuartos y forman el gráfico básico de caja y bigotes, descrito en la sección siguiente.

NOTA: Seleccione estadísticos de resumen adicionales utilizando *Opciones de panel* sólo cambia la selección de los análisis actuales. Para cambiar los estadísticos por defecto para futuros análisis, vaya al menú *Edición* y seleccione *Preferencias*. La pestaña *Estadísticos* del cuadro de diálogo permite cambiar los estadísticos a calcular por defecto cuando se ejecuta *Análisis de una variable* (así como para varios otros procedimientos que muestran resúmenes estadísticos):

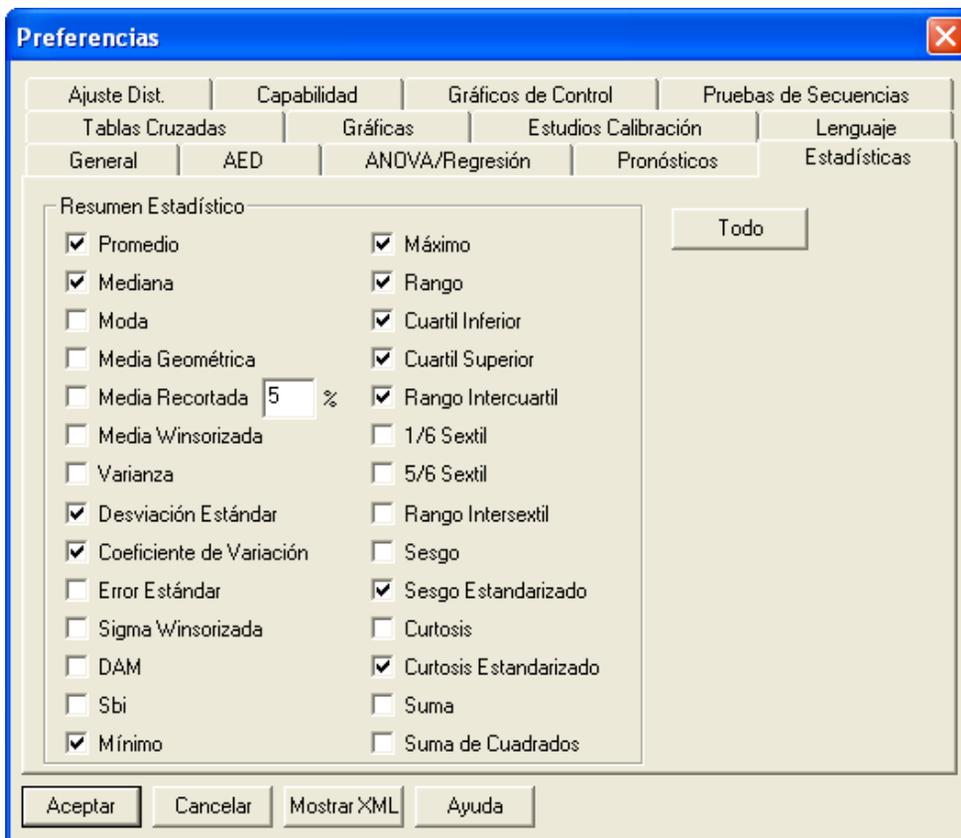


Figura 10-7. Cuadro de diálogo Preferencias utilizada para seleccionar estadísticos por defecto

10.3 Gráfico de caja y bigotes

Un gráfico usual utilizado para resumir datos, descubierto por John Tukey, es el gráfico de caja y bigotes mostrado en la esquina inferior derecha de la *figura 10-4* y aumentado a continuación:

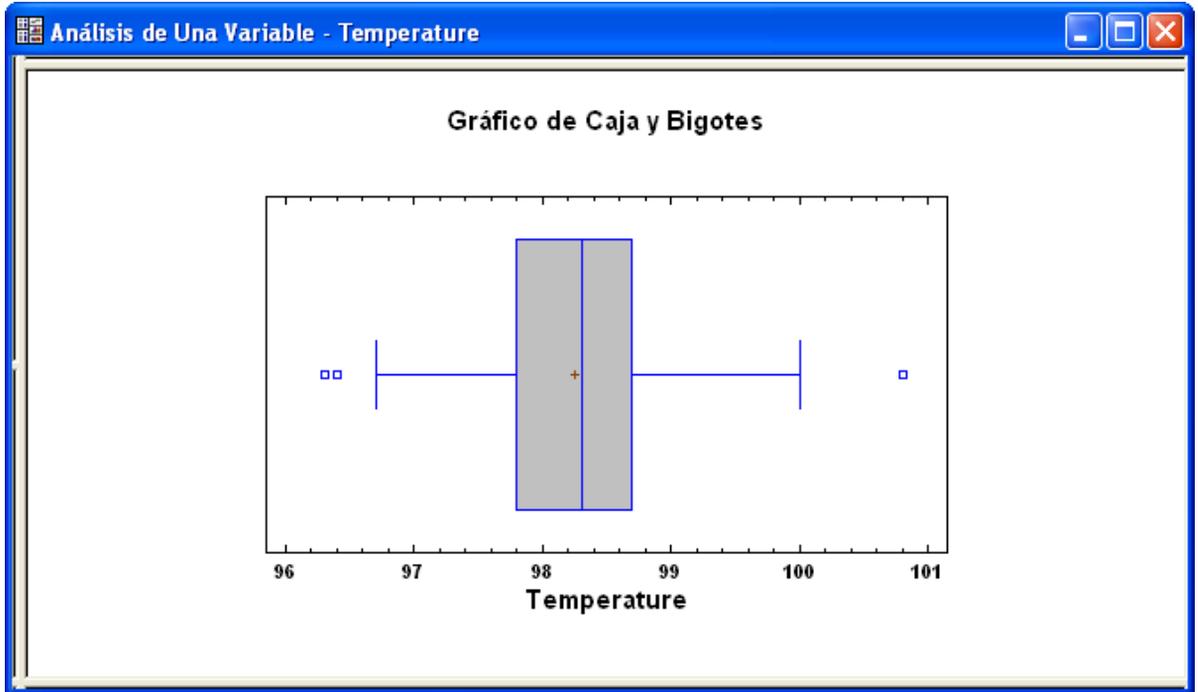


Figura 10-8. Gráfico de caja y bigotes para temperaturas corporales

El gráfico de caja y bigotes se construye como sigue:

1. Se dibuja una caja cuyos extremos se extienden desde el cuartil inferior al cuartil superior de la variable. El 50% central de los valores de datos están cubiertos por esta caja.
2. Se dibuja una línea vertical en el interior de la caja en la localización de la mediana simple, que divide la caja en dos partes. Si los datos provienen de una distribución simétrica, esta línea vertical debe pasar por el centro de la caja.
3. Se dibuja un signo más en la localización de la media simple. Una diferencia sustancial entre la media y la mediana indica habitualmente la presencia de valores atípicos (datos con valores que no provienen de la misma distribución que el resto) o distribución asimétrica. En este caso de distribución asimétrica, la media se situará en la dirección de la mitad mayor de la caja y existirá asimetría hacia esa dirección.

4. Los bigotes se extienden desde los cuartiles hasta el máximo y el mínimo de las observaciones en la muestra, a no ser que algunos valores estén lo suficientemente lejos de la caja para clasificarse como “valores atípicos”, en cuyo caso los bigotes se extienden hasta el más extremo de los puntos no calificado como “atípico”. STATGRAPHICS Centurion XVI sigue a Tukey tomando dos tipos de valores atípicos:
 - a. “Atípicos lejanos” – puntos situados a más de 3 veces el rango intercuartílico por arriba o por debajo de los límites de la caja. (Nota: el rango intercuartílico es la distancia entre los cuartiles primero y tercero y coincide con la anchura de la caja.) Los valores fuera de estos límites se denotan por un punto (habitualmente un pequeño cuadrado) con un signo más superpuesto en él. Si los datos provienen de una distribución normal, la posibilidad de que un punto esté lo suficientemente lejos de la caja para clasificarlo como un punto atípico es sólo de 1 entre 300 en una muestra del tamaño actual. A menos que haya cientos de observaciones en la muestra, los puntos fuera de estos límites son indicativos de valores atípicos (o de una distribución no normal).
 - b. “Atípicos dudosos” - puntos situados a más de 1.5 veces el rango intercuartílico por arriba o por debajo de los límites de la caja. Estos valores se denotan por puntos y no se le superpone el signo más. Si los datos provienen de una distribución normal, la posibilidad de observar 1 o 2 puntos atípicos en una muestra de $n = 100$ observaciones está cercana al 50% y no indica necesariamente la presencia de un valor atípico verdadero. Estos puntos deben ser considerados simplemente buenos para una investigación posterior.

El gráfico de caja y bigotes de la *figura 10-8* es razonablemente simétrico. Los bigotes están cerca de ser de la misma longitud y la media y la mediana de la muestra son similares coincidiendo prácticamente en la mitad de la caja. Se han marcado tres puntos atípicos dudosos, sin embargo no hay atípicos lejanos. Haciendo clic con el ratón en el atípico dudoso situado más a la derecha se observa que corresponde a la fila #15 en el archivo.

Si selecciona *Opciones de panel* en la barra de herramientas de análisis, puede añadir una muesca a la mediana en el gráfico:

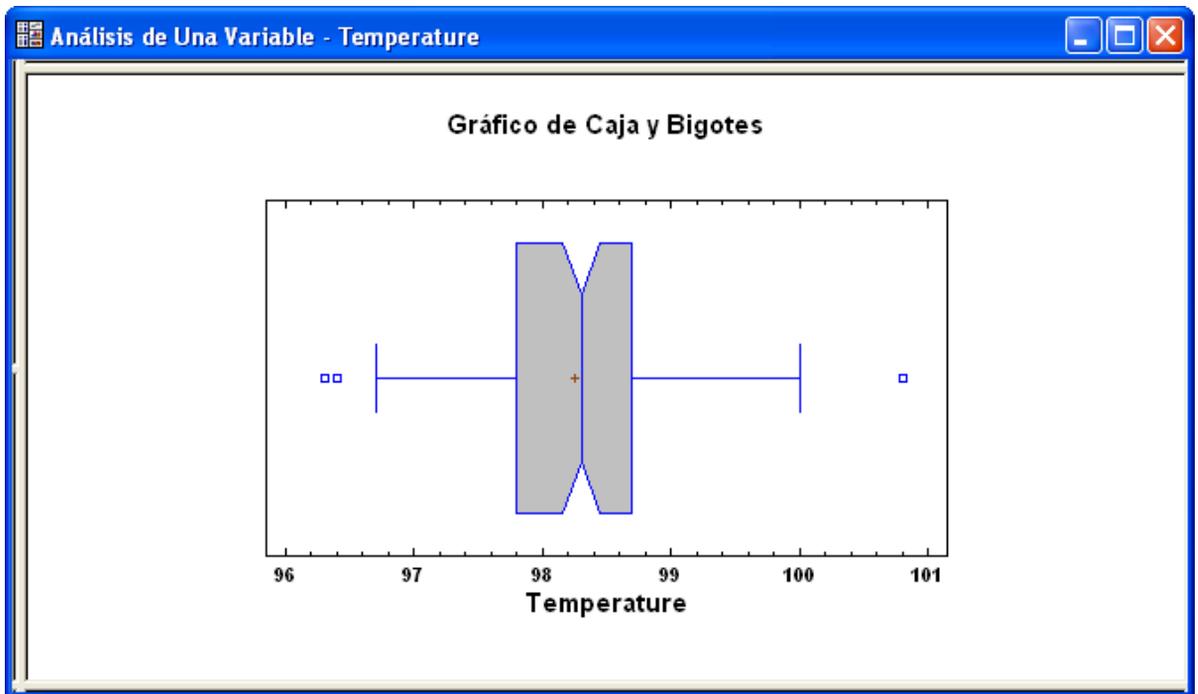


Figura 10-9. Gráfico de caja y bigotes con muesca en la mediana al 95%

La muesca de la mediana añadida cubre un intervalo de confianza para la mediana de la población, al 95% de confianza por defecto. Muestra el margen de error con que se estima la temperatura mediana de la población con la muestra que se ha tomado. En este caso, el error de muestreo es cercano a 0.15 grados en cada dirección. Una muestra de mayor tamaño presentará un menor margen de error.

10.4 Contrastando valores atípicos

Antes de estimar estadísticos adicionales, vale la pena tomarse un tiempo cuando la fila #15 debe de ser considerada como un atípico verdadero a remover potencialmente del archivo de datos. STATGRAPHICS Centurion XVI incluye un procedimiento que ejecuta un contraste formal para determinar cuándo una observación puede razonablemente provenir de una distribución no normal. El contraste es accesible en el menú principal seleccionando:

1. Si usa el menú clásico, seleccione *Describir – Datos numéricos – Identificación de Valores atípicos*.
2. Si usa el menú Seis Sigma, seleccione *Analizar – Datos de variable – Identificación de Valores atípicos*.

Especificando *Temperatura* en el campo *Datos* de la ventana *Opciones* se genera la ventana *Tablas y Gráficos*. Después de seleccionar todas las opciones deseadas, se genera una amplia tabla de estadísticos que se muestran en la mitad inferior del panel izquierdo. De particular interés en la tabla son los cinco valores mayores y los cinco valores menores 5 en la muestra:

Valores Ordenados				
		<i>Valores Studentizados</i>	<i>Valores Studentizados</i>	<i>Modificados</i>
<i>Fila</i>	<i>Valor</i>	<i>Sin Supresión</i>	<i>Con Supresión</i>	<i>Valor-Z DAM</i>
95	96.3	-2.65859	-2.74567	-2.698
55	96.4	-2.52219	-2.59723	-2.5631
23	96.7	-2.11302	-2.15912	-2.1584
30	96.7	-2.11302	-2.15912	-2.1584
73	96.8	-1.97663	-2.01521	-2.0235
...				
99	99.4	1.56955	1.59096	1.4839
13	99.5	1.70594	1.7323	1.6188
97	99.9	2.25151	2.30628	2.1584
120	100.0	2.3879	2.45231	2.2933
15	100.8	3.47903	3.67021	3.3725

Prueba de Grubbs (asume normalidad)
 Estadístico de prueba = 3.47903
 Valor-P = 0.0484379

Figura 10-10. Salida seleccionada del procedimiento *Identificación de valores atípicos*

El valor más atípico está en la fila #15, que se destaca en rojo. Tiene un valor estudentizado sin supresión (*Studentized Value Without Deletion*) de 3.479. Los valores estudentizados se calculan mediante:

$$z_i = \frac{x_i - \bar{x}}{s}$$

Un valor de 3.479 indica que la observación está a 3.479 desviaciones típicas sobre la media muestral, cuando la observación se incluye en el cálculo de \bar{x} y s . El valor estudentizado con supresión (*Studentized Values With Deletion*) es 3,67 e indica la misma interpretación pero ahora sin incluir la observación de la fila #15 en el cálculo de \bar{x} y s .

Las observaciones a más de 3 desviaciones típicas de la media se consideran atípicas, a menos que el tamaño de la muestra n sea muy grande o la distribución no sea normal. Puede ejecutarse un test formal bajo las siguientes hipótesis:

Hipótesis nula: El valor más extremo proviene de una distribución normal al igual que las otras observaciones.

Hipótesis alternativa: El valor más extremo no proviene de una distribución normal.

Un test ampliamente utilizado es el test de Grubbs, también llamado test de *Desviaciones extremas estudentizadas*. STATGRAPHICS Centurion XVI muestra el *P-valor* de este test. En general, un *P-valor* cuantifica la probabilidad de obtener un estadístico como inusual o más inusual que el observado en la muestra, si la hipótesis nula fuera cierta. Si el *P-valor* es lo suficientemente pequeño, la hipótesis nula puede ser rechazada, ya que la muestra ha tenido un evento extremadamente raro. “Suficientemente pequeño” es definido habitualmente como menor que 0.05, valor que se denomina “nivel de significación” o “riesgo alfa” del test. Si es menor del 5% la hipótesis nula se rechaza.

En este ejemplo, el test estadístico equivale al criterio del valor estudentizado sin supresión (*Studentized Value Without Deletion*), ya que el *P-valor* es igual a 0.0484. Como el *P-valor* es menor que 0.05, rechazamos la hipótesis nula, concluyendo de este modo que la fila #15 es un atípico comparado con el resto de los valores de la muestra.

Se puede quitar la fila #15 presionando el botón *Cuadro de diálogo de entrada* en la barra de herramientas de análisis e introduciendo una expresión en el campo *Seleccionar* como la que se muestra a continuación:

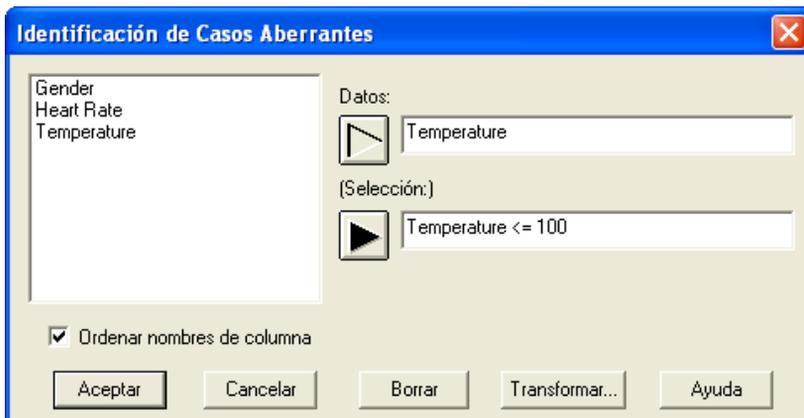


Figura10-11. Cuadro de diálogo de identificación de atípicos removiendo valores atípicos

Ya que la fila #15 es la única observación que excede 100 grados, el campo *Seleccionar* introducido arriba seleccionará sólo $n = 129$ filas. La salida modificada de *Identificación de valores atípicos* se muestra a continuación:

Valores ordenados				
Fila	Valor	Valores estudentizados Sin supresión	Valores estudentizados Con supresión	Modificado MAD Z-Score
95	96.3	-2.75487	-2.85205	-2.698
55	96.4	-2.61209	-2.6956	-2.5631
23	96.7	-2.18375	-2.23455	-2.1584
30	96.7	-2.18375	-2.23455	-2.1584
73	96.8	-2.04097	-2.08332	-2.0235
...				
119	99.4	1.6713	1.69652	1.4839
99	99.4	1.6713	1.69652	1.4839
13	99.5	1.81408	1.84516	1.6188
97	99.9	2.3852	2.44992	2.1584
120	100.0	2.52798	2.60411	2.2933

Test de Grubbs (asume normalidad)
 Test estadístico = 2.75487
 P-Valor = 0.676064

Figura 10-12. Salida de identificación de valores atípicos después de quitar la fila #15

El valor más extremo de las observaciones restantes está en la fila #95. Ya que el P -valor del test de Grubbs es ahora mayor que 0.05, todas las observaciones restantes parecen provenir de la misma población.

Idealmente, se debe volver al estudio original y encontrar una causa asignable para el valor anormal de la fila #15. Ya que esto es imposible de hacer ahora, aceptaremos el resultado del test de Grubbs y eliminaremos la fila #15 para todos los cálculos subsecuentes. Modificando el cuadro de diálogo de entrada de datos de *Análisis de una variable* como se muestra en la figura 10-11, los estadísticos resumen son ahora los que se muestran a continuación:

Estadísticos resumen para Temperatura	
Tamaño	129
Media	98.2295
Mediana	98.3
Desviación típica	0.70038
Coefficiente de variación	0.713004%
Mínimo	96.3
Máximo	100.0
Rango	3.7
Cuartil inferior	97.8
Cuartil superior	98.7
Rango intercuartílico	0.9
Asimetría estandarizada	-1.40217
Curtosis estandarizada	0.257075

Figura 10-13. Resumen de estadísticos después de eliminar la fila #15

10.5 Histograma

Otro gráfico habitual a mostrar que ilustra una muestra de datos es el histograma de frecuencias. Volviendo al procedimiento *Análisis de una variable*, se puede crear un histograma presionando el botón *Tablas y Gráficos*  en la barra de herramientas de análisis y seleccionando *Histograma de frecuencias*. El histograma por defecto se muestra a continuación:

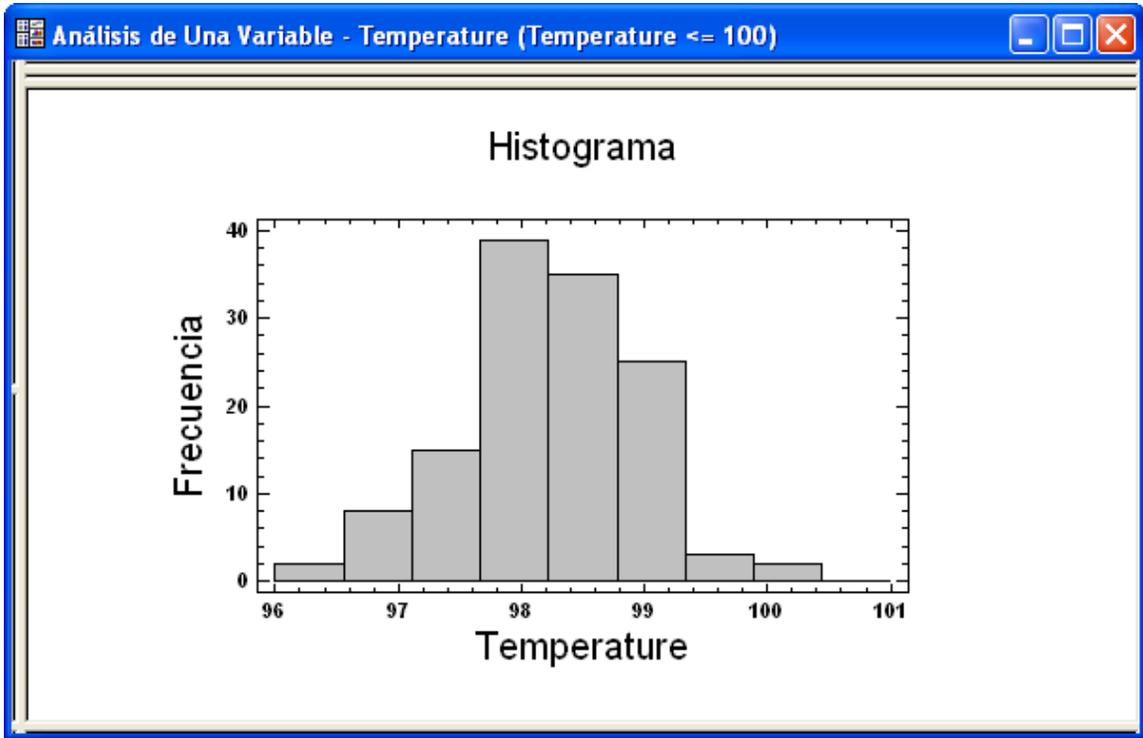


Figura 10-14. Histograma de frecuencias con clases por defecto

La altura de cada barra en el histograma representa el número de observaciones que caen en el intervalo de *temperatura* correspondiente a la barra. El número de barras en el rango se sitúa por defecto en función de n , utilizando la regla seleccionada en la pestaña *AED* (Análisis de Datos Exploratorio) del cuadro de dialogo *Editar - Preferencias*:

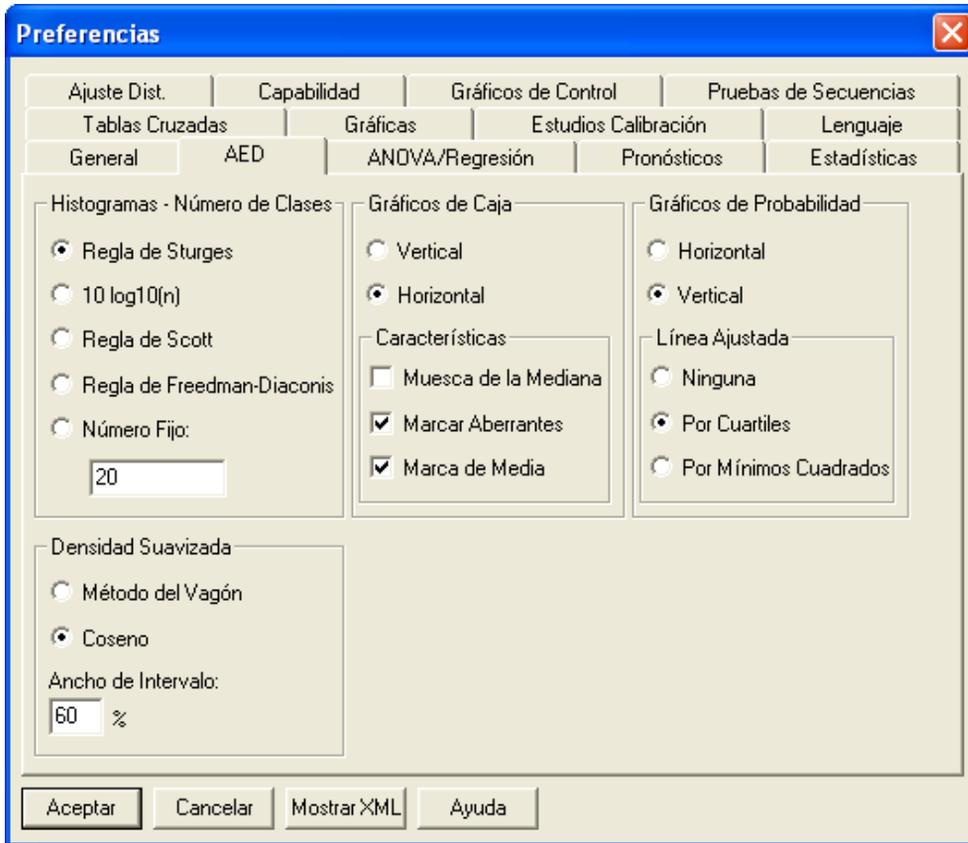


Figura 10-15. Pestaña AED del cuadro de diálogo Preferencias

Utilizando la regla de Sturges, el número de barras es la parte entera de $(1 + 3.322 \log_{10}(n))$. Otras reglas, tales como la regla $10 \log_{10}(n)$, tienden a producir más barras por defecto y pueden ser preferibles cuando se trabaja con grandes conjuntos de datos.

Puede hacerse caso omiso del número temporal de barras ya que es posible modificar un histograma después de haber sido creado haciendo clic sobre él para maximizar su panel y seleccionando *Opciones de panel*.



Figura 10-16. Cuadro de diálogo de Opciones de panel para Histograma de frecuencias

Cuando se deciden las clases, hay que tener en cuenta el número de dígitos significativos en los datos. Por ejemplo, las temperaturas corporales fueron medidas sólo en las cercanías de 0.1 grados. La anchura de los intervalos correspondientes a las barras deben ser enteros múltiplos de 0.1. Por esta vía, cada barra cubrirá el mismo número de posibles mediciones. El gráfico siguiente muestra 25 intervalos entre 96 y 101 grados, cubriendo cada uno un intervalo de 0.2 grados:

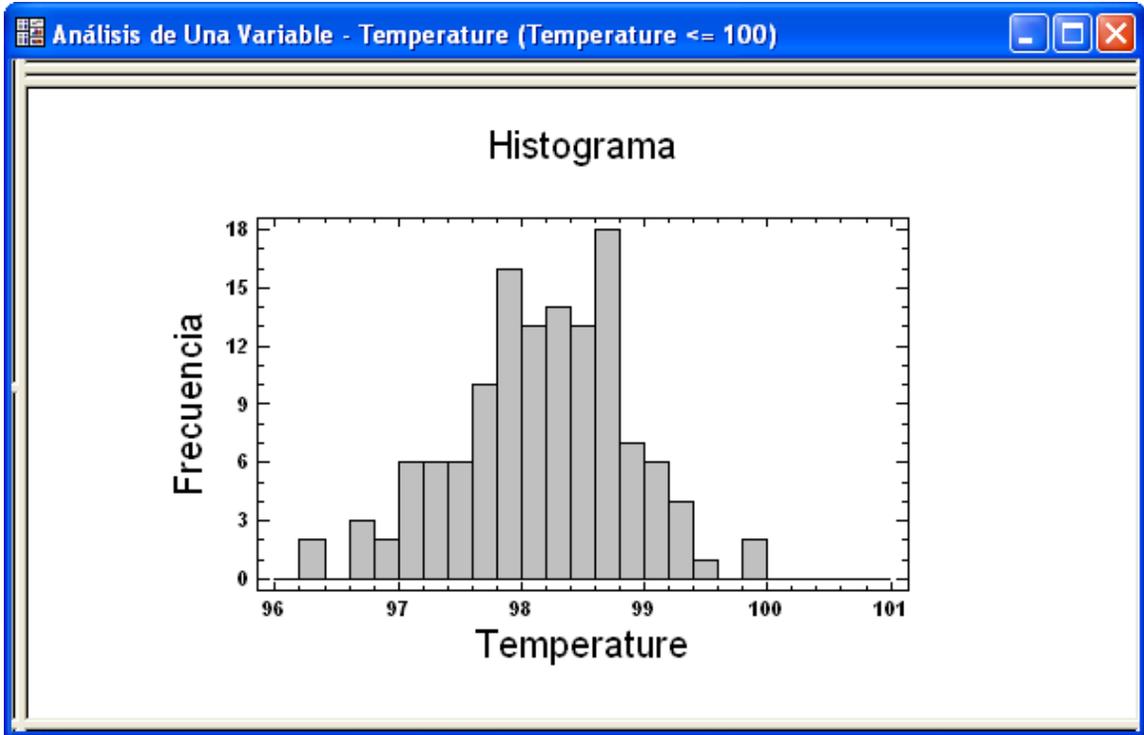


Figura 10-17. Histograma de frecuencias con clases redefinidas

Con el mayor número de clases se aparenta más detalle. La apariencia general de la distribución es similar a la de la curva normal ajustada al histograma.

Los datos mostrados en el histograma pueden ser obtenidos de forma tabular presionando el botón *Tablas y Gráficos*  en la barra de herramientas de análisis y seleccionando *Tabulación de frecuencias*:

Tabulación de frecuencias para Temperatura							
<i>Clase</i>	<i>Límite inferior</i>	<i>Límite superior</i>	<i>Punto Punto</i>	<i>Frecuencia</i>	<i>Frecuencia relativa</i>	<i>Frecuencia acumulada</i>	<i>Frecuencia relat. acum.</i>
	menor de	96.0		0	0.0000	0	0.0000
1	96.0	96.2	96.1	0	0.0000	0	0.0000
2	96.2	96.4	96.3	2	0.0155	2	0.0155
3	96.4	96.6	96.5	0	0.0000	2	0.0155
4	96.6	96.8	96.7	3	0.0233	5	0.0388
5	96.8	97.0	96.9	2	0.0155	7	0.0543
6	97.0	97.2	97.1	6	0.0465	13	0.1008
7	97.2	97.4	97.3	6	0.0465	19	0.1473
8	97.4	97.6	97.5	6	0.0465	25	0.1938
9	97.6	97.8	97.7	10	0.0775	35	0.2713
10	97.8	98.0	97.9	16	0.1240	51	0.3953
11	98.0	98.2	98.1	13	0.1008	64	0.4961
12	98.2	98.4	98.3	14	0.1085	78	0.6047
13	98.4	98.6	98.5	13	0.1008	91	0.7054
14	98.6	98.8	98.7	18	0.1395	109	0.8450
15	98.8	99.0	98.9	7	0.0543	116	0.8992
16	99.0	99.2	99.1	6	0.0465	122	0.9457
17	99.2	99.4	99.3	4	0.0310	126	0.9767
18	99.4	99.6	99.5	1	0.0078	127	0.9845
19	99.6	99.8	99.7	0	0.0000	127	0.9845
20	99.8	100.0	99.9	2	0.0155	129	1.0000
21	100.0	100.2	100.1	0	0.0000	129	1.0000
22	100.2	100.4	100.3	0	0.0000	129	1.0000
23	100.4	100.6	100.5	0	0.0000	129	1.0000
24	100.6	100.8	100.7	0	0.0000	129	1.0000
25	100.8	101.0	100.9	0	0.0000	129	1.0000
	above	101.0		0	0.0000	129	1.0000

Media = 98.2295 Desviación típica = 0.70038

Figura 10-18. Tabla de tabulación de frecuencias

Tome nota de que las observaciones se cuentan pertenecientes a un intervalo si son mayores que el límite inferior del intervalo y menores o iguales que el límite superior (intervalos cerrados por la derecha y abiertos por la izquierda).

La última columna de la derecha es también de considerable interés, ya que me muestra la probabilidad acumulada de que un individuo caiga en la clase seleccionada o en clases anteriores. por ejemplo, el 89.92% de los valores de datos son menores o iguales que 99.0 grados.

10.6 Gráficos de cuantiles y percentiles

Otro camino para mostrar probabilidades acumuladas es seleccionar *Gráfico de cuantiles* de la lista de *Gráficos* en el procedimiento *análisis de una variable*:

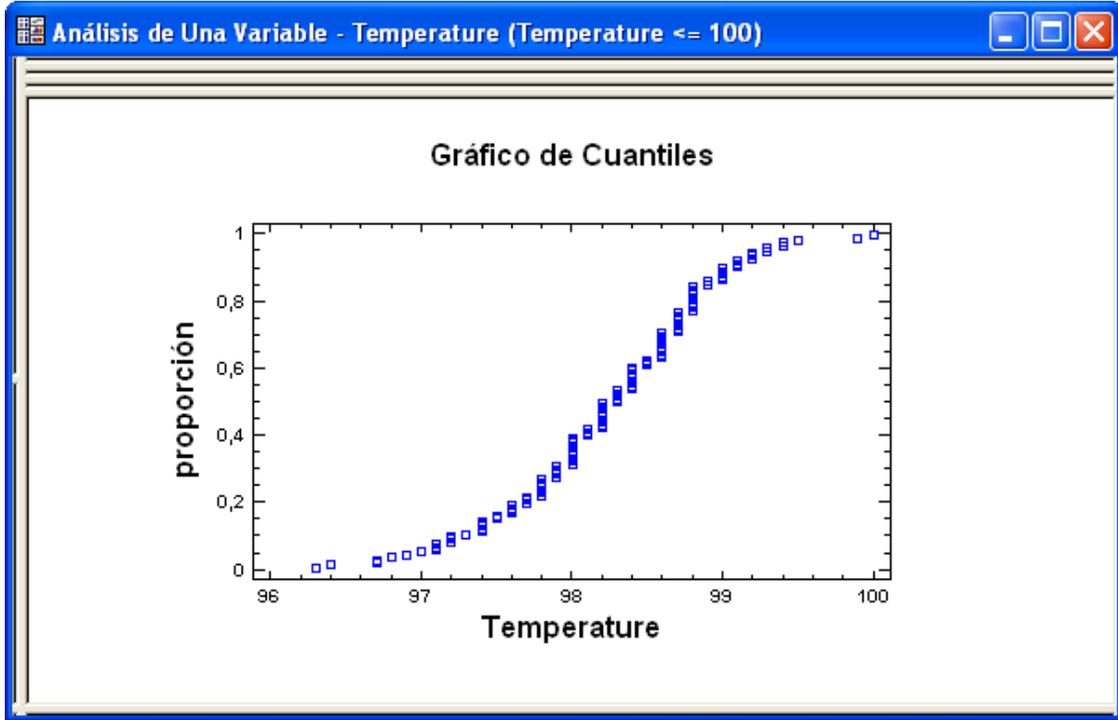


Figura 10-19. Gráfico de cuantiles

En este gráfico, los datos son primeramente ordenados de menor a mayor. El j^{th} mayor de los valores de datos se grafica como $Y = (j+0.5)/n$. Esto estima la proporción de la población por debajo de la temperatura observada. Como muestra la columna más a la derecha en la tabla de frecuencias, la curva representa la probabilidad acumulada de un individuo teniendo una temperatura menor o igual que la que se muestra en el eje horizontal. Ya que la temperatura fue solo medida en las cercanías de 0.1 grados, hay saltos verticales en la gráfica que se muestra.

La *figura 10-19* también muestra un conjunto de cursores cruzados. Estos se han creado presionando con el botón derecho del ratón en un punto mientras se está viendo el gráfico y seleccionando *Localizar* en el menú emergente resultante. Puede utilizar el ratón para arrastrar las cruces a otra localización. Los números pequeños cerca de las líneas cruzadas indican la posición del punto en la nueva localización. En el gráfico anterior, las líneas cruzadas han sido utilizadas

para localizar la mediana o *percentil 50*, que es el valor de *temperatura* para el que la proporción mostrada en el eje vertical es igual a 0.5.

Puede crearse también una tabla de percentiles seleccionando *Percentiles* en la lista *Tablas*:

Percentiles para Temperatura	
	<i>Percentiles</i>
1.0%	96.4
5.0%	97.0
10.0%	97.2
25.0%	97.8
50.0%	98.3
75.0%	98.7
90.0%	99.1
95.0%	99.3
99.0%	99.9

Salida incluyendo límites de confianza normales al 95.0%.

Figura 10-20. Tabla de percentiles

El percentil de orden p estima el valor de la temperatura por debajo del cual se encuentran el $p\%$ de los valores de datos. Se ha utilizado *Opciones de panel* añadiendo límites de confianza al 95% a estos percentiles, basados en la asunción de que la muestra proviene de una población normal.

Por ejemplo, el percentil 90 temperaturas el valor de la temperatura excedida por sólo el 10% de los individuos en la población. El mejor estimador del percentil basado en la muestra de datos es 99.1 grados. Sin embargo, dado el tamaño limitado de la muestra, el percentil 90 debe caer entre 98.98 y 99.31 grados, con un 95% confianza.

10.7 Intervalos de confianza

Eliminados los valores atípicos de la muestra, podemos proceder a establecer los estimadores finales para los parámetros de la distribución provenientes de los datos. Seleccionando *Intervalos de confianza* del cuadro de diálogo de *Tablas y Gráficos* tenemos:

Intervalos de confianza para Temperatura	
95.0% intervalo de confianza para la media:	98.2295 +/- 0.122015 [98.1074,98.3515]
95.0% intervalo de confianza para la desviación típica:	[0.624081,0.798114]

Figura 10-21. Intervalos de confianza al 95% para la media y la desviación típica

Los intervalos de confianza aportan una frontera del error potencial al estimar la media y la desviación típica de una población. Dadas las restantes $n = 129$ observaciones, concluimos con

el 95% de confianza que la temperatura media en la población está entre 98.11 y 98.35 grados. Del mismo modo, la desviación típica de la población está entre 0.624 y 0.798 grados.

Seleccionando *Opciones de panel*, pueden ser requeridos intervalos de confianza adicionales utilizando el método bootstrap:

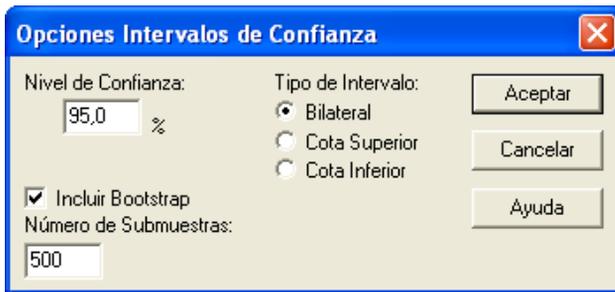


Figura 10-22. Cuadro de diálogo de opciones de intervalos de confianza

Los intervalos Bootstrap, diferentes de los intervalos de la *figura 10-21*, no cuentan con la asunción de que la población es normal. En su lugar, se toman muestra aleatorias de $n = 129$ observaciones, mediante muestreo con reposición (las mismas observaciones pueden ser seleccionadas más de una vez). Este muestreo se repite 500 veces, se calculan los estadísticos simples y el 95% de los resultados más centrales se utilizan para calcular los intervalos de confianza. La tabla siguiente muestra los intervalos bootstrap para la media de la población, desviación típica y mediana:

intervalos de confianza para Temperatura

95.0% intervalos de confianza para la media: 98.2295 +/- 0.122015 [98.1074,98.3515]

95.0% intervalos de confianza para la desviación típica: [0.624081,0.798114]

Intervalos Bootstrap

Media: [98.1132,98.3519]

Desviación típica: [0.621373,0.785949]

Mediana: [98.1,98.4]

Figura 10-23. Intervalos de confianza Bootstrap al 95% de confianza

NOTA: Sus resultados pueden variar ligeramente de los obtenidos aquí.

Los intervalos anteriores, calculados utilizando la distribución t de Student y la distribución chi-cuadrado, son mejorados por los intervalos bootstrap. Esto no es inesperado, ya que los datos no muestran a simetría o curtosis significativa.

10.8 Test de hipótesis

Es posible ejecutar también los test de hipótesis formales habituales. Por ejemplo, se asegura habitualmente que la temperatura humana es de 98.6 grados Fahrenheit. Para contrastar si los datos provienen de una distribución normal con una media dada, se puede considerar el test de hipótesis siguiente:

Hipótesis nula: $\mu = 98.6$ grados

Hipótesis alternativa: $\mu \neq 98.6$ grados

Para ejecutar este test con el procedimiento *Análisis de una variable*, seleccione *Test de hipótesis* de la lista de *Tablas y Gráficos*. Antes de examinar los resultados, seleccione *opciones de panel* y especifique los atributos deseados para el test:

The image shows a dialog box titled "Opciones Pruebas de Hipótesis". It has a blue title bar with a close button (X) on the right. The dialog is divided into two main sections: "Localización" and "Dispersión".

Localización section:

- There are three checkboxes: "Prueba t" (checked), "Prueba de los Signos" (unchecked), and "Prueba de Rangos con Signo" (checked).
- A text box labeled "Media/Mediana" contains the value "98.6".
- A text box labeled "Alfa:" contains "5.0" followed by a percent sign.
- Under "Hipótesis Alternativa", there are three radio buttons: "Diferente de" (selected), "Menor Que", and "Mayor Que".

Dispersión section:

- There is one checkbox: "Prueba Chi-Cuadrada" (unchecked).
- A text box labeled "Desviación Estándar:" contains the value "1.0".
- A text box labeled "Alfa:" contains "5.0" followed by a percent sign.
- Under "Hipótesis Alternativa", there are three radio buttons: "Diferente de" (selected), "Menor Que", and "Mayor Que".

On the right side of the dialog, there are three buttons: "Aceptar", "Cancelar", and "Ayuda".

Figura 10-24. Opciones de panel para Test de hipótesis

El valor introducido para *Media* representa la hipótesis nula. Como *Alt. Hipótesis*, puede seleccionar cualquiera de las hipótesis alternativas siguientes:

1. *Distinto*: $\mu \neq 98.6$
2. *Menor que*: $\mu < 98.6$
3. *Mayor que*: $\mu > 98.6$

Aún cuando la muestra sugiere una temperatura menor, hemos seleccionado una hipótesis alternativa de dos lados. Creando un test de un lado con una hipótesis alternativa $\mu < 98.6$ puede ser considerado “fisgón de datos”, ya que se formula la hipótesis después de haber mirado los datos.

Los resultados del test se muestran a continuación:

<p>Test de hipótesis para Temperatura Media muestral = 98.2295 Mediana muestral = 98.3 Desviación típica muestral = 0.70038</p> <p><u>t-test</u> Hipótesis nula: media = 98.6 Alternativa: distinto</p> <p>Estadístico t = -6.00896 P-Valor = 1.81264E-8 Rechaza la hipótesis nula para alfa = 0.05.</p> <p><u>test de los signos rangos</u> Hipótesis nula: mediana = 98.6 Alternativa: distinto</p> <p>Rango medio de valores por debajo de la media de la hipótesis: 67.7099 Rango medio de valores por encima de la media de la hipótesis: 43.5658</p> <p>Test estadístico para muestras grandes = 5.07771 (aplicada corrección de continuidad) P-Valor = 3.82663E-7 Rechazo de la hipótesis nula para alfa = 0.05.</p>
--

Figura 10-25. Resultados del test de hipótesis

Los resultados de dos test son los siguientes:

1. Un test estándar de la t , que asume que los datos provienen de una distribución normal (aunque no es demasiado sensible a esta asunción).

2. Un test no paramétrico de los signos rangos, basado en los rangos de las distancias de cada observación a la mediana supuesta. Este test no asume normalidad y es menos sensible a valores atípicos que el test de la t .

En ambos casos, el P -valor es menor que 0.05, rechazándose la hipótesis de la que la muestra proviene de una población con media 98.6 grados.

NOTA: La notación E-8 después de un número significa que el número está multiplicado por 10^{-8} . El P -valor se muestra como 1.81264E-8 que por consiguiente es igual a 0.0000000181264.

Se debe hacer hincapié en que el intervalo de confianza para la media, dado en la sección 10.8, no incluye el valor 98.6. Los valores no incluidos en el intervalo de confianza para la media serán rechazados por el test al mismo nivel de confianza. Se puede decir que el intervalo de confianza contiene todos los posibles valores para la media de la población que son aceptables a través de los datos de la muestra.

10.9 Límites de tolerancia

Se suele utilizar un análisis adicional para los datos de la temperatura corporal que crea límites de tolerancia normal. Se trata de límites dentro de los cuales se estima que caiga un porcentaje seleccionado de la población para un nivel de confianza dado. Los límites de tolerancias son accesibles desde el menú principal mediante:

1. Si usa el menú clásico, seleccione *Describir – Datos numéricos – Límites estadísticos de tolerancia*
2. Si usa el menú Seis Sigma, seleccione *Analizar – Datos de variable – Límites estadísticos de tolerancia*

El procedimiento comienza mostrando un cuadro de diálogo en el cual se introduce el tamaño de la muestra n y la media y desviación típica de la muestra. Usando los resultados de la *figura 10-13*, las entradas apropiadas son:



Figura 10-26. Cuadro de diálogo para Límites estadísticos de tolerancia

Cuando se presiona Aceptar, aparece el menú *Opciones* y el cuadro de diálogo de *Tablas y Gráficos*. La salida se muestra a continuación:

Límites estadísticos de tolerancia

Tamaño de la muestra = 129

Media de la muestra = 98.2295

Desviación típica de la muestra = 0.70038

95.0% intervalo de confianza para el 99.0% de la población

Xbar +/- 2.88436 sigma

Upper: 100.25

Lower: 96.2093

StatAdvisor

Asumiendo que los datos provienen de una distribución normal, los límites de tolerancia indican que podemos afirmar con una confianza del 95% que el 99.0% de la distribución se encuentra entre 96.2093 y 100.25. Este intervalo se computa tomando la media de los datos +/-2.88436 veces la desviación típica.

Figura 10-27. Resumen de análisis para Límites estadísticos de tolerancia

La interpretación del StatAdvisor resume los resultados sucintamente. El nivel de confianza y el porcentaje de población cubierto pueden cambiarse en *Opciones de panel*.

El procedimiento *Límites de tolerancia estadística* también crea el *Gráfico de tolerancia*, que muestra los límites de tolerancia:

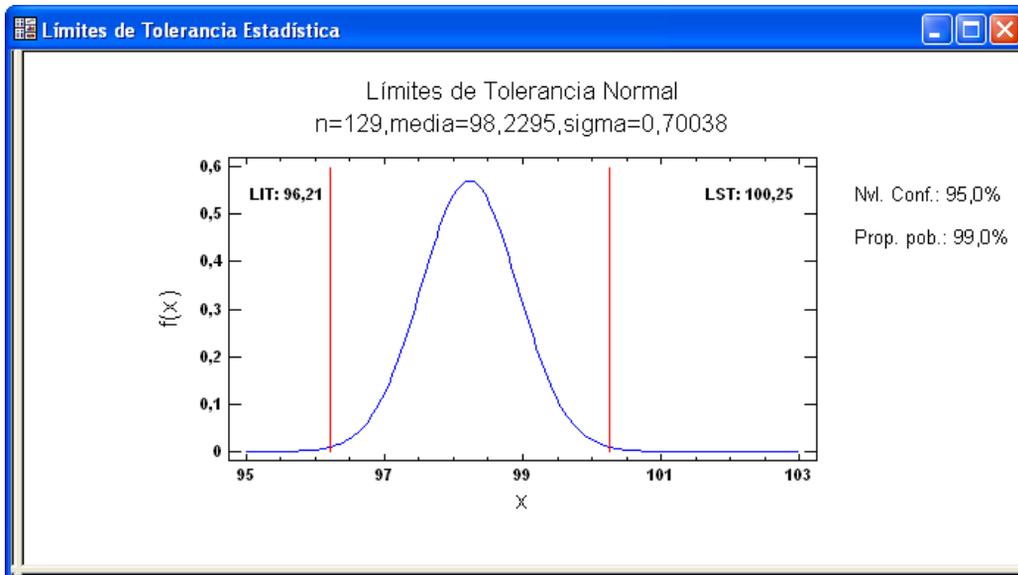


Figura 10-28. Gráfico de tolerancia

No es admisible más de un individuo sobre 100 fuera de los límites de tolerancia.

Tutorial #2: Comparando dos muestras

Comparaciones gráficas y test de hipótesis.

A menudo se tienen dos muestras para analizar, posiblemente de poblaciones diferentes. En tales casos es usual:

1. Mostrar los datos para realizar comparaciones visuales.
2. Contrastar hipótesis para determinar cuándo hay (o no) diferencias significativas entre las dos muestras.

El tutorial #1 del último capítulo analizaba un conjunto de temperaturas corporales tomadas de 130 sujetos. De entre estos sujetos 65 eran mujeres y 65 eran hombres. En este tutorial, compararemos los datos de las mujeres y los hombres.

Para analizar las temperaturas corporales, se abre el archivo de datos *bodytemp.sgd* utilizando *Abrir origen de datos* en el menú *Archivo – Abrir*.

11.1 Ejecutando el procedimiento Comparación de dos muestras

El procedimiento principal para comparar datos de dos muestras es el procedimiento *Comparación de dos muestras*, accesible desde el menú principal como sigue:

1. Si usa el menú clásico, seleccione *Comparar – Dos muestras– Muestras independientes*.
2. Si usa el menú Seis Sigma, seleccionar *Analizar – Datos de variable – Comparación de dos muestras – Muestras independientes*.

El cuadro de diálogo de entrada de datos para el procedimiento es el siguiente:



Figura 11-1. Cuadro de diálogo Comparación de dos muestras

La caja *Entrada* indica la forma en que se introducirán los datos para las dos muestras:

1. *Datos en dos columnas* – los datos para cada muestra están en columnas diferentes.
2. *Columnas de datos y códigos* – los datos para ambas muestras están en la misma columna, y una segunda columna contiene códigos que diferencian los datos de las dos muestras.

El archivo *bodytemp.sgd* tiene el segundo tipo de estructura, con las $n = 130$ observaciones en una misma columna de nombre *Temperatura*, mientras una segunda columna de nombre *Género* contiene las etiquetas “Mujer” u “Hombre”. En el campo *Seleccionar*, se elegirán solo las filas para las cuales la *Temperatura* es menor o igual que 100. Así se excluye la fila #15 del análisis, cuya observación había sido identificada en el capítulo 10 como atípica.

Después del cuadro *Tablas y Gráficos* se presenta una ventana que contiene 4 paneles: un resumen de los datos, un histograma dual, un resumen estadístico de cada grupo y un gráfico de caja y bigotes dual, como se muestra a continuación.

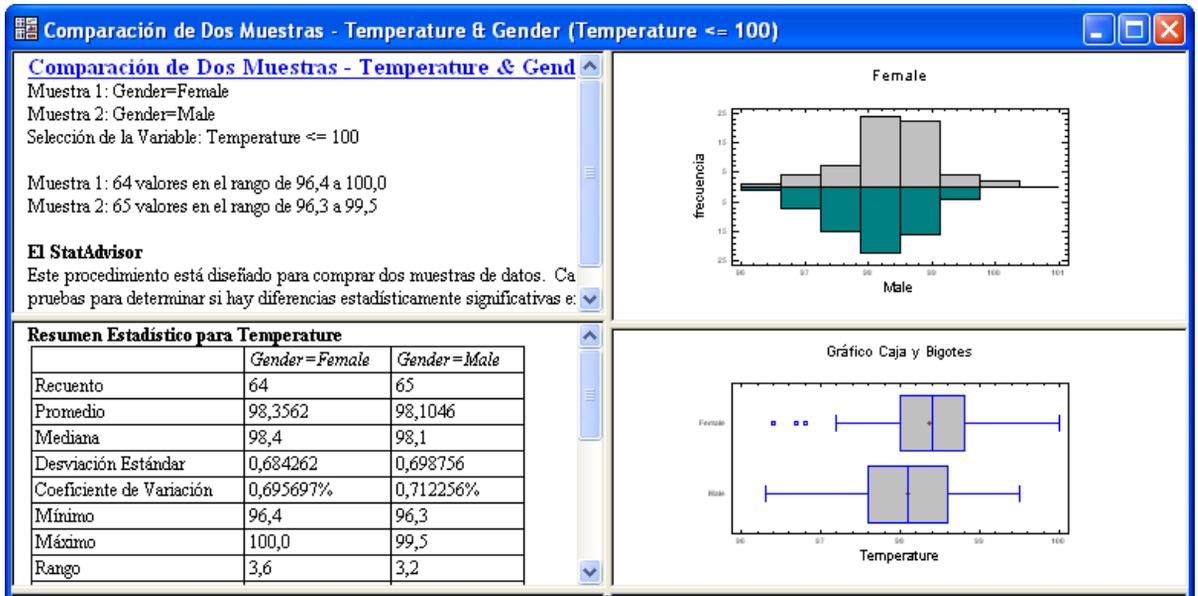


Figura 11-2. Ventana de análisis de Comparación de dos muestras

Después de eliminar el valor atípico, hay $n_1 = 64$ observaciones para mujeres, con rangos de temperaturas entre 96.4 y 100.0 grados, y $n_2 = 65$ observaciones para mujeres, con rangos de temperaturas entre 96.3 y 99.5 grados.

11.2 Estadísticos resumen

La tabla *Estadísticos resumen* muestra los estadísticos calculados para cada muestra:

Estadísticos resumen para Temperatura		
	<i>Género=Mujer</i>	<i>Género=hombre</i>
Tamaño	64	65
Media	98.3562	98.1046
Mediana	98.4	98.1
Desviación típica	0.684262	0.698756
Coefficiente de variación	0.695697%	0.712256%
Mínimo	96.4	96.3
Máximo	100.0	99.5
Rango	3.6	3.2
Cuartil inferior	98.0	97.6
Cuartil superior	98.8	98.6
Rango intercuartílico	0.8	1.0
Asimetría estandarizada	-1.35246	-0.702297
Curtosis estandarizada	1.49635	-0.610877

Figura 11-3. Estadísticos resumen para la muestra

Varias facetas son particularmente interesantes:

1. La temperatura media de las mujeres es cerca de 0.25 grados más alta que la de los hombres. La diferencia entre las medianas es 0.30 grados.
2. La desviación típica de las mujeres es ligeramente menor que la de los hombres, indicando que las temperaturas corporales de las mujeres pueden ser menos variables que las de los hombres.
3. Ambas muestras tienen valores de asimetría y curtosis estandarizadas en el rango entre -2 y 2. Como se ha explicado en el capítulo 10, valores en dicho rango son consistentes con la hipótesis de que los datos provienen de distribuciones normales.

Resta por determinar si la diferencia entre las temperaturas corporales entre los hombres y las mujeres es estadísticamente significativa.

11.3 Histograma dual

El histograma dual de frecuencias compara las dos muestras simultáneamente. Utilizando *Opciones de panel* para reescalar los intervalos de clase de tal modo que haya 25 intervalos entre 96 y 101 grados, se genera el gráfico siguiente:

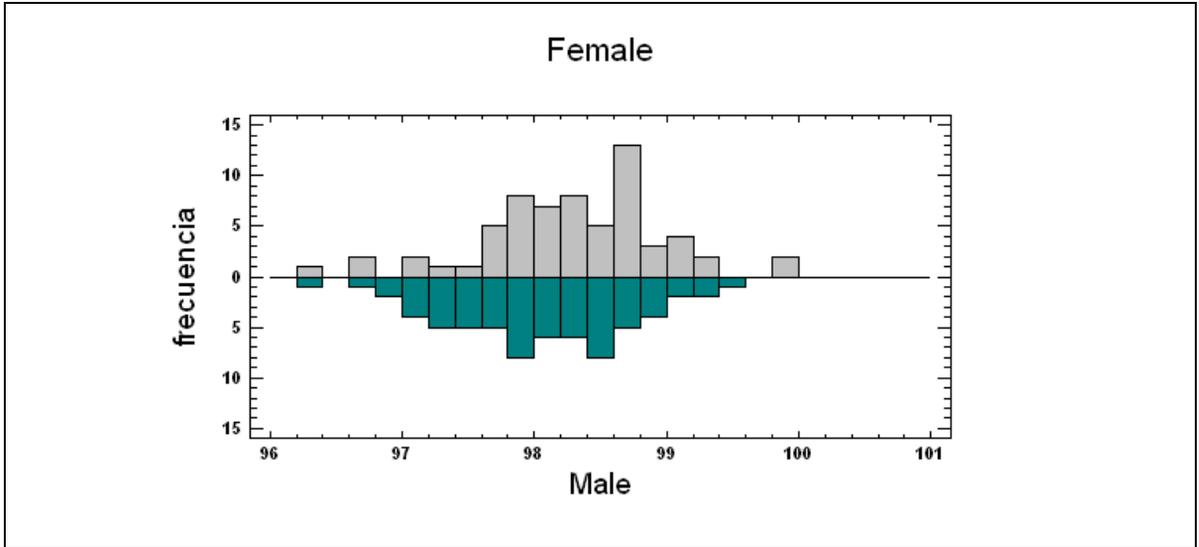


Figura 11-4. Histograma de frecuencias dual

El histograma para las mujeres se muestra por encima de la línea horizontal. El histograma para los hombres se muestra invertido por debajo de la línea horizontal. Las formas de las distribuciones son similares, con un posible desplazamiento de distribución de las mujeres a la derecha de la de los hombres.

11.4 Gráfico dual de caja y bigotes

La ventana de análisis también muestra el gráfico de caja y bigotes para las dos muestras. Como se ha explicado en el capítulo 10, la caja central cubre la mitad de las observaciones de cada muestra. Los bigotes se extienden entre los valores máximo y mínimo de cada muestra, excepto para algunos puntos que se sitúan inusualmente lejanos a las cajas. Se dibuja una línea vertical en cada caja a la altura de la mediana, mientras que los pequeños signos más situados en el interior de las cajas indican las localizaciones de las medias muestrales.

En este caso, es usual añadir muescas al gráfico accediendo al *Panel de Opciones*. Los resultados se grafican como sigue:

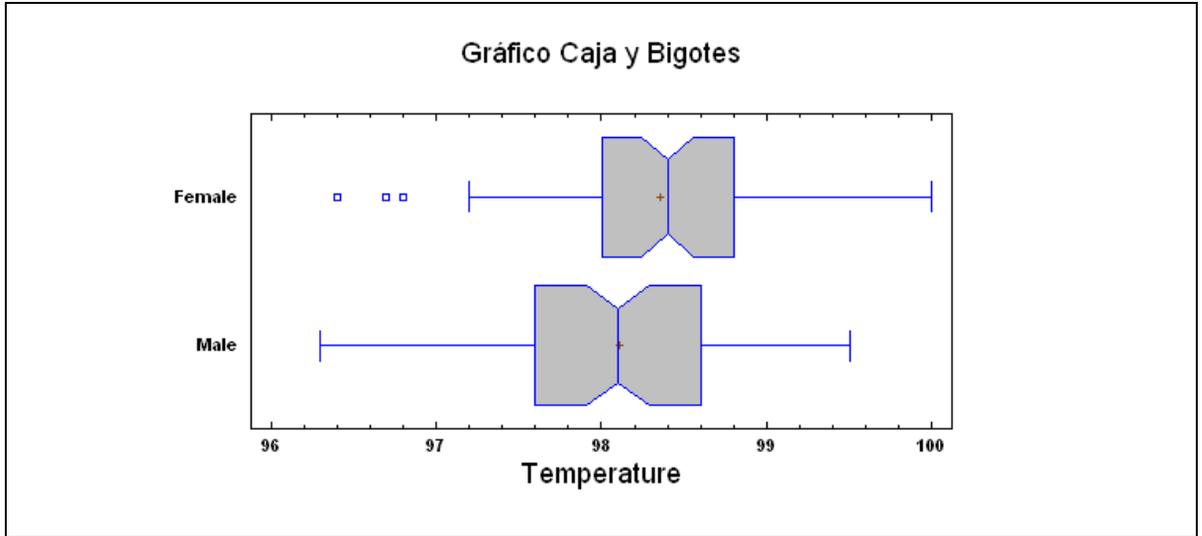


Figura 11-5. Gráfico dual de caja y bigotes con muescas para las medianas

En el gráfico se observa lo siguiente:

1. Un aparente desvío del centro de la distribución de las mujeres hacia la derecha respecto del centro de la distribución de los hombres. Medias y medianas muestran una diferencia similar.
2. El rango cubierto por las mujeres es más ancho que el rango cubierto por los hombres, pero solamente si se incluye el punto más pequeño.
3. La muesca de la mediana de las mujeres coincide en parte con la de los hombres (es un poco más estrecha). Si las muescas de las medianas no son coincidentes en parte, las medianas de hombres y mujeres serán significativamente diferentes al nivel por defecto del sistema (actualmente del 5%). Una comparación más formal se describe en la sección siguiente.

Basándose en este gráfico, parece haber una diferencia en el centro de las dos muestras, aunque la significación estadística de la diferencia está indeterminada.

11.5 Comparando desviaciones típicas

La primera comparación formal entre las dos muestras debe de ser el contraste de hipótesis de que la desviación típica (σ) de la población de la cual provienen los datos es igual para las dos muestras contra la hipótesis alternativa de que es diferente:

Hipótesis nula: $\sigma_1 = \sigma_2$

Hipótesis alternativa: $\sigma_1 \neq \sigma_2$

Este contraste nos permitirá determinar si la diferencia aparente entre las variabilidades de las temperaturas corporales de hombres y mujeres es estadísticamente significativa, o si está dentro del rango de la variabilidad aleatoria normal para muestras del tamaño actual.

Para ejecutar este test, presione el botón *Tablas y Gráficos*  en la barra de herramientas de análisis y seleccione *Comparación de desviaciones típicas*. Los resultados se muestran a continuación:

Comparación de desviaciones típicas para Temperatura		
	<i>Género=Mujer</i>	<i>Género=Hombre</i>
desviación típica	0.684262	0.698756
Varianza	0.468214	0.48826
Grados de libertad	63	64

Ratio de varianzas = 0.958945

95.0% Intervalos de confianza
 Desviación típica de Género=Mujer: [0.582853,0.828723]
 Desviación típica de Género=Hombre: [0.595887,0.844885]
 Ratio de varianzas: [**0.584028,1.57609**]

F-test para Comparar Desviaciones típicas
 Hipótesis nula: $\sigma_1 = \sigma_2$
 Alt. hipótesis: $\sigma_1 \neq \sigma_2$
 F = 0.958945 P-valor = **0.8684**
 No se rechaza la hipótesis nula para $\alpha = 0.05$.

Figura 11-6. Comparación de desviaciones típicas de dos muestras

Lo más importante de la salida de esta tabla está iluminado en rojo:

1. *Ratio de Varianzas*: muestra un intervalo de confianza al 95% para el ratio de varianzas (varianza de la población de mujeres σ_1^2 dividida por la varianza de la población de hombres σ_2^2). La *Varianza* es una medida de la variabilidad calculada como el cuadrado de la desviación típica. (NOTA: comparaciones de variabilidad entre más de una muestra están basadas habitualmente en las varianzas en lugar de en las desviaciones típicas, ya

que tienen propiedades matemáticas más atractivas.) El intervalo para σ_1^2 / σ_2^2 está entre 0.58 y 1.58. Esto indica que la varianza de las mujeres puede estar entre aproximadamente el 58% y el 158% de la varianza de los hombres. Esta falta de precisión es típica cuando se intentan las comparaciones de variabilidades con muestras pequeñas.

2. El *P-valor* asociado con el estadístico de la *F* se observa arriba. Un *P-valor* menor que 0.05 indicará diferencia estadísticamente significativa entre la varianza de las temperaturas corporales de las mujeres y los hombres al 5% de nivel de significación. Ya que *P* no es menor que 0.05, no hay evidencia suficiente para rechazar la hipótesis de igualdad de varianzas (o lo que es lo mismo, igualdad de desviaciones típicas).

Por consiguiente no hay evidencia estadísticamente significativa para concluir que la variabilidad de la temperatura corporal de las mujeres es diferente que la de los hombres.

Se debe hacer notar que este contraste es bastante sensible a la asunción de que las muestras provienen de poblaciones normales, una asunción que ya mostró ser razonable observando los valores de las asimetrías y las curtosis de las dos muestras (ambas entre -2 y 2).

11.6 Comparando medias

La segunda comparación entre las dos muestras contrasta la hipótesis de que la media (μ) de las dos poblaciones es la misma:

Hipótesis nula: $\mu_1 = \mu_2$

Hipótesis alternativa: $\mu_1 \neq \mu_2$

Para ejecutar este test, presione el botón *Tablas* después de seleccionar *Comparación de medias*. Los resultados son los siguientes:

Comparación de medias para Temperatura

95.0% intervalo de confianza par la media de Género=Mujer: 98.3562 +/- 0.170924 [98.1853,98.5272]

95.0% intervalo de confianza par la media de Género=Hombre: 98.1046 +/- 0.173144 [97.9315,98.2778]

95.0% intervalo de confianza para la diferencia entre las medias

asumiendo varianzas iguales : 0.251635 +/- 0.240998 [0.0106371,0.492632]

t test para comparar medias

hipótesis nula: media1 = media2

Alt. hipótesis: media1 NE media2

asumiendo varianzas iguales: t = 2.06616 P-valor = 0.040846

Rechazamos la hipótesis nula para alfa= 0.05.

Figura 11-7. Comparación de medias para dos muestras

Lo más importante de la salida está iluminado en rojo en la tabla:

1. *diferencia entre las medias (asumiendo varianzas iguales)*: muestra un intervalo de confianza al 95% para la media de temperaturas corporales de la población de mujeres menos la media de la población de los hombres. El intervalo para $\mu_1 - \mu_2$ varía desde 0.01 a 0.49, indicando que la media de la temperatura corporal de las mujeres está entre 0.01 y 0.49 más alta que la temperatura corporal media de los hombres.
2. El *P-valor* asociado con el test de la t se muestra en la tabla anterior. Ya que el *P-valor* es menor que 0.05, hay evidencia significativa para rechazar la hipótesis de igualdad de medias y declarar las medias de las dos poblaciones estadísticamente diferentes al 5% de nivel de significación.

Tome nota de que el test ha sido construido suponiendo que las varianzas de las dos poblaciones son desiguales, lo cual ya fue validado con el estadístico de la *F* en la sección previa. Si las varianzas hubiesen sido iguales, aproximadamente hubiera utilizado un test de la t adecuado accediendo al *Panel de opciones* y marcando la opción etiquetada *Asumir sigmas iguales*.

Esto refrenda el hecho de que las mujeres provienen de una población con temperatura corporal más alta que la de los hombres.

11.7 Comparando medianas

Si se sospecha que los datos contienen valores atípicos inevitables, se puede ejecutar un test no paramétrico para comparar las medianas en lugar de las medias. Los test no paramétricos no asumen que los datos provengan de una distribución normal y tienden a ser menos afectados por la presencia de valores atípicos.

Seleccionando *Comparación de medianas* del cuadro de diálogo *Tablas y Gráficos* se genera el test del estadístico *W* de Mann-Whitney (Wilcoxon). En este test, las dos muestras han sido primero combinadas. Los datos combinados se han reordenado de 1 hasta $n_1 + n_2$, y los valores de los datos originales han sido reemplazados por sus respectivos rangos. El test estadístico *W* se construye comparando los rangos medios de las observaciones en las dos muestras:

Comparando las medianas para Temperatura

Mediana de la muestra 1: 98.4

Mediana de la muestra 2: 98.1

W-test de comparación de medianas de Mann-Whitney (Wilcoxon)

Hipótesis nula: mediana1 = mediana2

Alt. hipótesis: mediana1 NE mediana2

Rango medio de la muestra 1: 71.9219

Rango medio de la muestra 2: 58.1846

W = 1637.0 P-valor = 0.0368312

Se rechaza la hipótesis nula para alfa = 0.05.

Figura 11-8. Comparación de medianas de dos muestras

La interpretación del test de Mann-Whitney (Wilcoxon) es paralela a la del test de la t descrito en la última sección, con un P -valor pequeño que permite concluir que las medianas de las muestras son significativamente diferentes.

11.8 Gráfico de cuantiles

Para ilustrar la diferencia entre las dos distribuciones, se realizan simultáneamente gráficos de cuantiles para cada muestra que pueden ser mostrados seleccionando *Gráfico de cuantiles* de la caja de diálogo *Gráficos*:

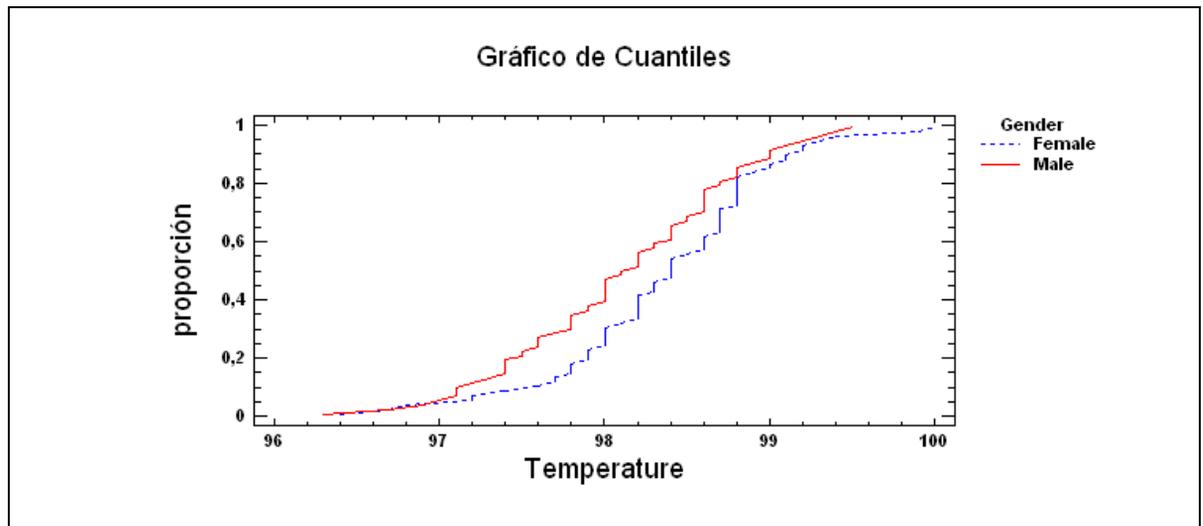


Figura 11-9. Gráficos de cuantiles simultáneos

El gráfico de cuantiles ilustra la proporción de datos en cada muestra que queda por debajo de un valor dado de X , como una función de X . Si las muestras provienen de la misma población, los gráficos de cuantiles deben superponerse. Gráficos situados uno a la izquierda o la derecha del otro indican diferencia entre las dos medias muestrales. Una diferencia entre las pendientes de las curvas indica diferencia entre las desviaciones típicas.

En el gráfico anterior, es evidente que la distribución de las mujeres está situada a la derecha de la de los hombres. Las pendientes, sin embargo, son similares.

11.9 Test de Kolmogorov-Smirnov para dos muestras

Un test no paramétrico adicional que puede ser ejecutado si la asunción de la normalidad de las distribuciones no es sostenible es el test de Kolmogorov-Smirnov para dos muestras. Este test está basado en el cálculo de la distancia vertical máxima entre las funciones de distribución acumuladas de las dos muestras, que aproximadamente es la máxima distancia entre los dos gráficos de cuantiles de la *figura 11-9*. Si la máxima distancia es lo suficientemente grande, las dos muestras pueden ser declaradas provenientes de poblaciones significativamente diferentes.

Seleccionando *Test de Kolmogorov-Smirnov* en el cuadro de diálogo de *Tablas y Gráficos* se muestra lo siguiente:

Test de Kolmogorov-Smirnov para Temperatura Estadístico global estimado DN = 0.242548 Estadístico K-S de dos lados para muestras grandes = 1.37737 P valor aproximado = 0.0449985

Figura 11-10. Salida del test de Kolmogorov-Smirnov

La máxima distancia vertical, denotada por DN, es igual aproximadamente a 0.24 para los datos de las temperaturas corporales.

El P -valor es utilizado para determinar cuándo las distribuciones son o no son significativamente diferentes. Un pequeño P -valor nos lleva a la conclusión de que hay una diferencia significativa. Ya que el P -valor para esta muestra de datos es menor o igual que 0.05, hay una diferencia significativa entre las distribuciones de la temperatura corporal de los hombres y las mujeres al 5% de nivel de significación.

Peligro: Si los datos se redondean, el test puede no ser fiable ya que la función de distribución acumulativa empírica (CDF) puede tener saltos en pasos largos. Cuando sea

posible, es mejor confiar en una comparación de parámetros seleccionados de distribuciones tales como la media, la desviación típica o la mediana.

11.10 Gráfico cuantil-cuantil

Un gráfico final, disponible seleccionando *Gráfico cuantil-cuantil* de la caja de diálogo *Gráficos*, representa los cuantiles estimados en una muestra contra los cuantiles de la otra muestra:

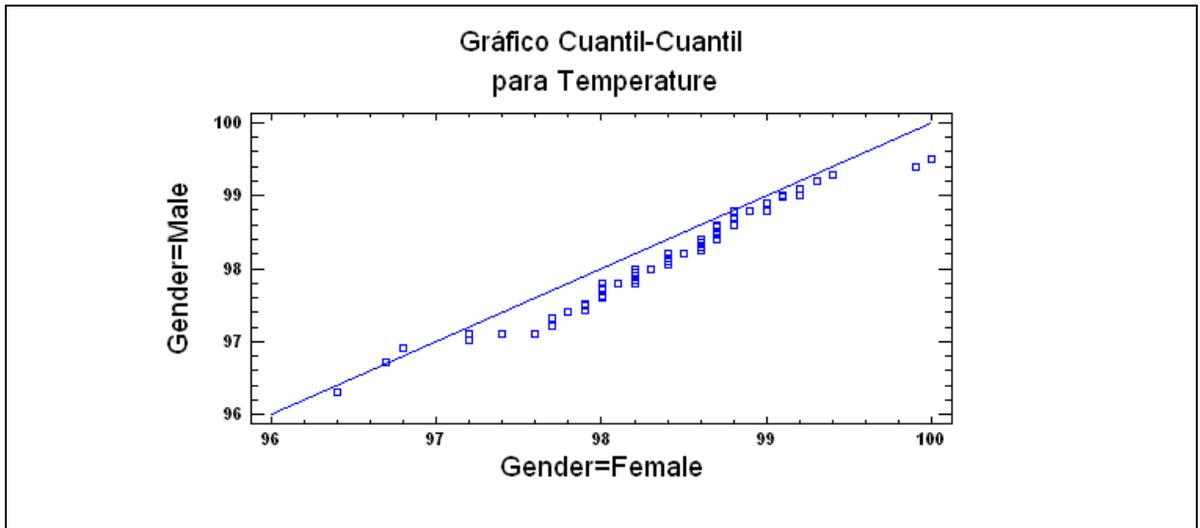


Figura 11-11. Gráfico Q-Q para los datos de las temperaturas corporales

Hay un punto en este gráfico correspondiendo a cada observación en la menor de las dos muestras. En el otro eje se representa el cuantil estimado de la muestra mayor. Si las muestras provienen de poblaciones idénticas, los puntos del gráfico Q-Q deben estar muy cercanos a la línea diagonal. Una desviación constante hacia la izquierda o hacia la derecha es síntoma de diferencia significativa entre los centros de las dos distribuciones. Puntos divergentes en la línea con pendiente diferente a la de la diagonal indican una diferencia significativa en variabilidad. En este caso, la diferencia entre las poblaciones puede ser más complicada de observar que en el caso del cambio en la media, ya que los puntos cierran la línea en temperaturas altas y bajas. Se observa que la distribución de las temperaturas para las mujeres está más concentrada en el centro que la distribución de los hombres.

Tutorial #3: Comparando más de dos muestras

Comparando medias y desviaciones típicas, ANOVA de un factor, ANOM, y métodos gráficos.

Cuando los datos caen en más de dos grupos, se necesita utilizar técnicas diferentes a las usadas en el capítulo anterior. Por ejemplo, supongamos que queremos comparar la resistencia de diferentes aparatos fabricados con 4 materiales distintos. En un experimento típico, construimos 12 aparatos de cada uno de los 4 materiales para compararlos. Los datos siguientes representan los resultados del experimento:

<i>Material A</i>	<i>Material B</i>	<i>Material C</i>	<i>Material D</i>
64.7	60.4	58.3	60.8
64.8	61.8	62.1	60.2
66.8	63.3	62.4	59.8
67.0	61.6	60.3	58.3
64.9	61.0	60.6	56.4
63.7	63.8	60.0	61.6
61.8	60.9	60.3	59.5
64.3	65.1	62.4	62.0
64.3	61.5	61.9	61.4
65.9	60.0	63.1	58.6
63.6	62.9	60.2	59.5
64.6	60.6	58.6	60.0

Es de considerable interés determinar qué materiales construyen los aparatos más resistentes, así como qué materiales son estadísticamente diferentes entre sí.

Hay dos caminos diferentes para introducir datos de múltiples muestras en la hoja de datos:

1. Usar una columna separada para cada muestra.
2. Usar una columna simple para todos los datos y crear una segunda columna con códigos identificando de qué muestra proviene cada observación.

Para este ejemplo se ha seleccionado el primero de los caminos. Los datos para los aparatos han sido situados en cuatro columnas del archivo de nombre *widjets.sgd*, que se puede abrir seleccionando *Abrir – Abrir origen de datos* del menú *Archivo*.

12.1 Ejecutando el procedimiento comparación de varias muestras

El procedimiento *Comparación de varias muestras* es accesible en el menú principal de las dos formas siguientes:

1. Si usa el menú clásico seleccione: *Comparar – Varias muestras – Comparación de varias muestras*.
2. Si usa el menú Seis Sigma seleccione: *Analizar – Datos de variables – Comparación de varias muestras – Comparación de varias muestras*.

El cuadro de diálogo inicial se utiliza para estructurar los datos:

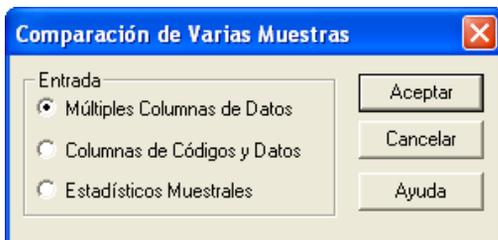


Figura 12-1. Cuadro de diálogo inicial para comparación de varias muestras

En este caso, los datos han sido situados en varias columnas de la hoja de datos.

El segundo cuadro de diálogo requiere los nombres de las columnas que contienen los datos:

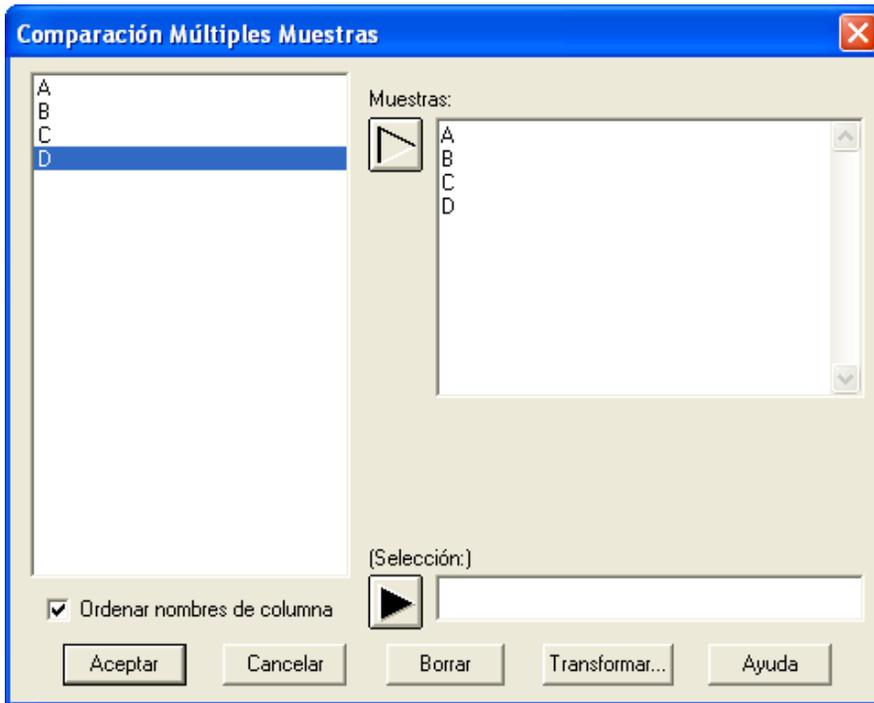


Figura 12-2. Cuadro de diálogo de entrada de datos de Comparación de varias muestras

En el archivo de la muestra de datos, las observaciones han sido situadas en cuatro columnas de nombres A, B, C y D.

Cuando se presiona *Aceptar*, aparece el cuadro de diálogo *Tablas y Gráficos*. Se aceptan las características por defecto en este tutorial.

Cuando se abre la ventana de análisis, aparecerán cuatro paneles:

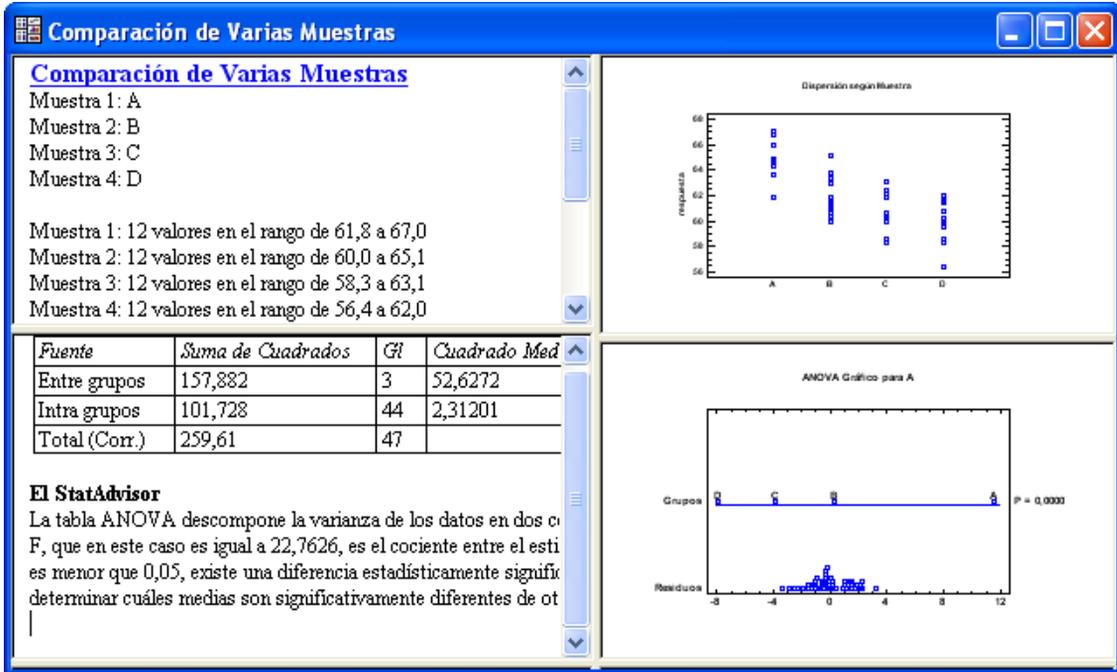


Figura 12-3. Ventana de análisis de Comparación de varias muestras

El panel superior izquierdo resume el tamaño de cada muestra y su rango. El panel superior derecho muestra un gráfico de dispersión de los datos, ampliado a continuación:

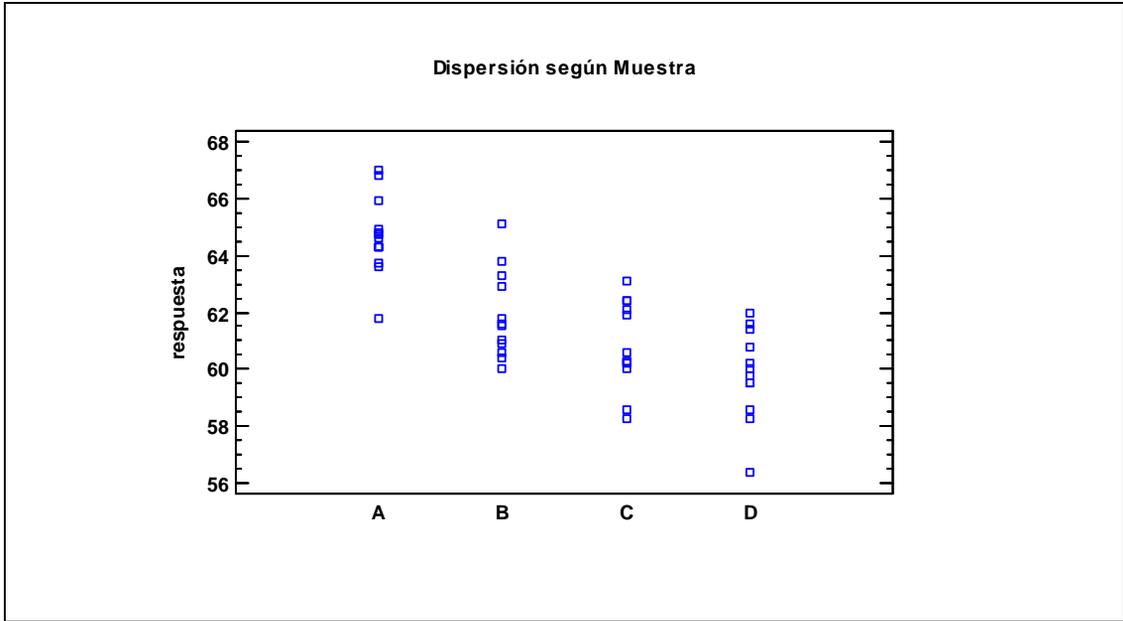


Figura 12-4. Gráfico de dispersión de Resistencia contra material

Observad que muchas de las observaciones aparecen una encima de otra en líneas verticales. Para aliviar este problema, se hace doble clic en el panel gráfico para maximizarlo y se presiona el botón *Separar*  de la barra de herramientas de análisis y se añade una pequeña cantidad de separación horizontal moviendo el botón deslizante un poco hacia la derecha:

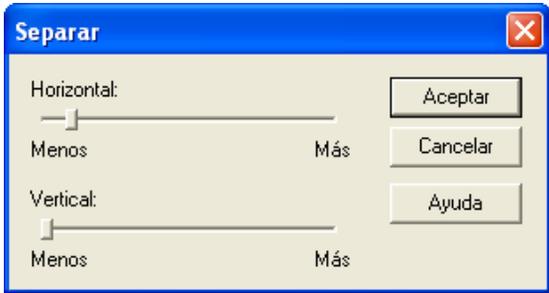


Figure 12-5. Cuadro de diálogo de Separación de puntos

Se desplaza una pequeña cantidad de puntos aleatoriamente en la dirección horizontal, haciendo que los puntos individuales se observen más fácilmente:

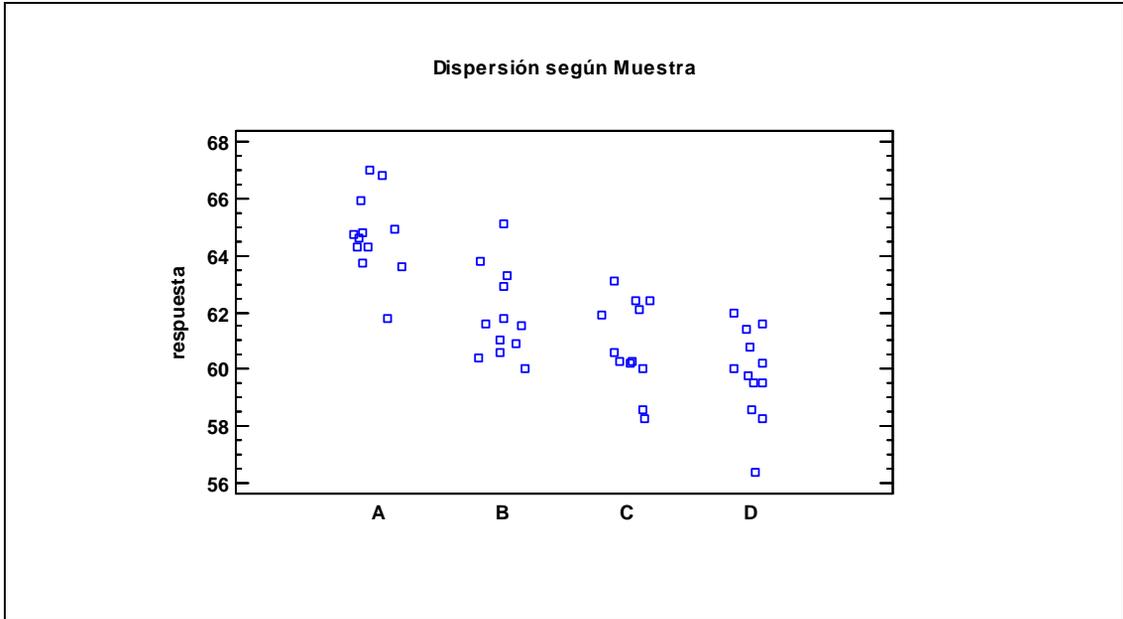


Figura 12-6. Diagrama de dispersión después de la separación de puntos

La separación afecta solo a la visión de los puntos, pero no a los datos ni a los cálculos realizados con ellos.

12.2 Análisis de la varianza

El primer paso cuando se comparan varias muestras es habitualmente ejecutar un análisis de la varianza simple (ANOVA). El análisis ANOVA se utiliza para contrastar la hipótesis de igualdad de medias poblacionales eligiendo entre las dos hipótesis siguientes:

Hipótesis nula: $\mu_A = \mu_B = \mu_C = \mu_D$

Hipótesis alternativa: Las medias no son iguales

donde μ_j representa la media de la población de la cual hemos tomado la muestra j -ésima. El rechazo de la hipótesis nula indicará que las muestras provienen de poblaciones con medias diferentes.

La salida del procedimiento ANOVA está contenida en la tabla ANOVA, que se muestra inicialmente en el panel inferior izquierdo de la ventana de análisis:

Tabla ANOVA

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	157,882	3	52,6272	22,76	0,0000
Intra grupos	101,728	44	2,31201		
Total (Corr.)	259,61	47			

Figura 12-7. Tabla del análisis de la varianza

El análisis de la varianza descompone la variabilidad de los datos observados en dos componentes: una componente entre grupos, que cuantifica las diferencias entre aparatos hechos de diferentes materiales, y una componente dentro de grupos, que cuantifica las diferencias de los aparatos hechos con el mismo material. Si se estima la variabilidad entre grupos y es significativamente mayor que la variabilidad dentro de grupos, es evidente que las medias de los grupos no son similares.

El valor clave en la figura 12-7 es el *P-valor*. *P*-valores pequeños (menores que 0.05 operando al 5% de nivel de significación) llevan al rechazo de la hipótesis de igualdad de medias. En el ejemplo actual, hay una pequeña duda de si las muestras son significativamente diferentes.

En la última edición de *Statistics for Experimenters* de Box, Hunter y Hunter (John Wiley and Sons, 2005), los autores presentan una nueva salida diseñada para mostrar los resultados de un modelo ANOVA en formato gráfico. Este *Gráfico ANOVA* se muestra por defecto en el panel inferior derecho:

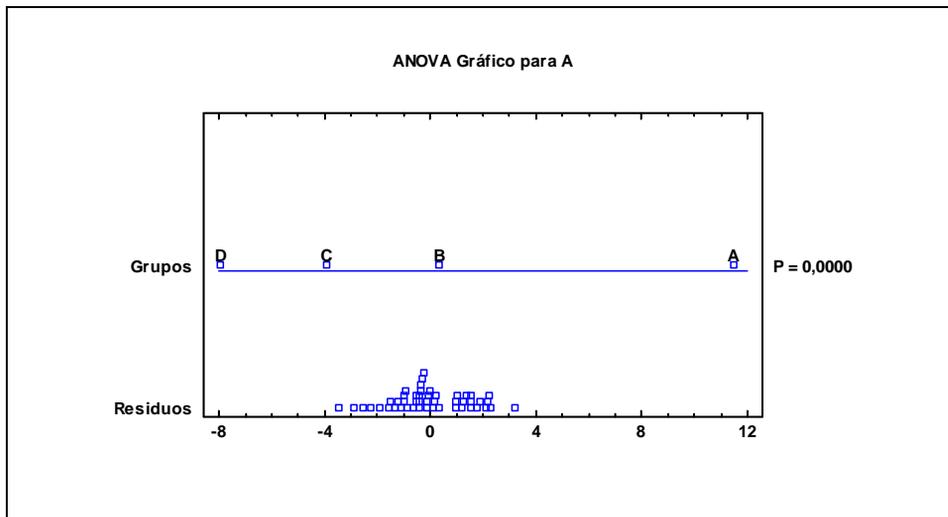


Figura 12-8. Gráfico ANOVA

A lo largo de la parte inferior del gráfico hay un diagrama de puntos de los *residuos* del modelo. En el ANOVA simple, los residuos son iguales a las diferencias entre cada observación y la media de todas las observaciones de ese grupo. En el ejemplo actual, la variabilidad observada en los residuos es indicativa de la variabilidad natural entre los aparatos hechos del mismo material. Representados por encima de la línea central están escaladas las desviaciones de las medias de los grupos respecto de la media total de las $n = 48$ observaciones. Este grupo de desviaciones está escalado de modo que su variabilidad pueda ser comparada con la de los residuos. Grupos cuyos puntos están demasiado lejos probablemente provengan de una distribución con extensión similar a la de los residuos como corresponde a poblaciones diferentes.

En la *figura 12-8*, el grupo A parece estar bastante separado de los otros grupos. La separación de las otras tres medias es menos clara. Una comparación más formal de las medias de las cuatro muestras se describe en la sección siguiente.

12.3 Comparando medias

Si el P -valor en la tabla ANOVA es pequeño, entonces la media de la muestra debe ser examinada, para determinar qué medias son significativamente diferentes unas de otras. Un gráfico habitual para esta finalidad es *Gráfico de medias*, disponible en el cuadro de diálogo *Tablas y Gráficos*:

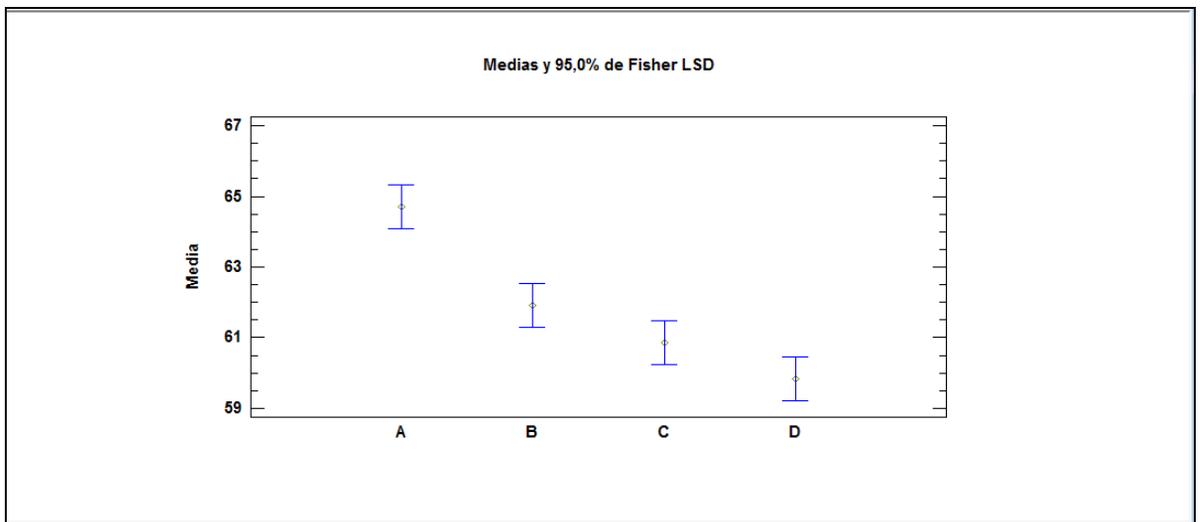


Figura 12-9. Gráfico de medias

El gráfico de medias muestra cada media de la muestra, junto con un intervalo de confianza a su alrededor. La interpretación de los intervalos depende del tipo de intervalo representado, el cual

puede ser cambiado utilizando *Opciones de ventana*. Los dos intervalos más habitualmente utilizados son:

1. *Intervalos LSD de Fisher LSD (Least Significant Difference)*: Estos intervalos están escalados de modo que un par de muestras tiene medias significativamente diferentes si los intervalos no se solapan en la dirección vertical. Mientras la posibilidad de declarar incorrectamente dos muestras con media diferente con este método se fija en el 5%, haciendo comparaciones de más de dos pares de muestras la probabilidad de error es considerablemente superior.
2. *Intervalos HSD de Tukey (Honestly Significant Difference)*. Estos intervalos están escalados para controlar el error del experimento como mucho con una tasa del 5%. Usando el método de Tukey, no se declararán incorrectamente que *ningún* par de medias sea significativamente diferente cuando realmente no los son en más de 5% de los análisis que se hagan.

Los intervalos de la *figura 12-9* utilizan el método de Tukey. Ya que el intervalo para la muestra A no solapa ningún otro intervalo, la media de la muestra A difiere significativamente de la de las otras tres muestras. La muestra B también es significativamente diferente de la muestra D, ya que sus intervalos no se solapan. La muestra C, sin embargo, no es significativamente diferente de las muestras B o D.

El mismo análisis puede ser mostrado en forma tabular seleccionando *Pruebas de Múltiples Rangos* del cuadro de diálogo *Tablas y Gráficos*:

Pruebas de Múltiple Rangos			
Método: 95,0 porcentaje LSD			
	<i>Casos</i>	<i>Media</i>	<i>Grupos Homogéneos</i>
D	12	59,8417	x
C	12	60,85	xx
B	12	61,9083	x
A	12	64,7	x

<i>Contraste</i>	<i>Sig.</i>	<i>Diferencia</i>	<i>+/- Límites</i>
A - B	*	2,79167	1,25105
A - C	*	3,85	1,25105
A - D	*	4,85833	1,25105
B - C		1,05833	1,25105
B - D	*	2,06667	1,25105
C - D		1,00833	1,25105

* indica una diferencia significativa.

Figura 12-10. Test de Rangos Múltiples

La sección inferior de la salida muestra cada par de medias. La columna *Diferencia* muestra la media simple del primer grupo menos la del segundo. La columna *+/- Límites* muestra un intervalo de confianza para la diferencia. Cualquier par de medias para el que el valor absoluto de la diferencia exceda el límite presenta diferencia estadísticamente significativa al nivel de confianza seleccionado y es representado por un * en la columna *Sig.* En el ejemplo actual, cuatro de los seis pares de medias muestran diferencias significativas.

La sección superior de la salida presenta las muestras en grupos homogéneos, presentando la letra X en columnas. Un grupo homogéneo es aquél para el que no hay diferencias significativas. En este caso, la muestra A es un grupo homogéneo en sí mismo, ya que es significativamente diferente de todos los demás (sólo hay una X en su columna). La muestra C cae en dos grupos, en uno con B (hay una X en la misma columna para C y B) y en otro con D (hay una X en la misma columna para C y D).

12.4 Comparando Medianas

Cuando existen valores atípicos, deben utilizarse procedimientos no paramétricos como una alternativa a los análisis estándar de la varianza seleccionando los test de *Kruskal-Wallis* y *Friedman* en la caja de diálogo *Tablas*. Estos test comparan las medianas de las muestras en lugar de las medias:

Hipótesis nula: las medianas son todas iguales

Hipótesis alternativa: las medianas no son todas iguales

Este tipo de test puede ser utilizado seleccionando *Opciones de ventana*. Hay dos tipos de test:

1. *Test de Kruskal-Wallis*— apropiado cuando cada columna contiene una muestra aleatoria de su población. En tal caso, las filas no tienen significado intrínseco.
2. *Test de Friedman*— apropiado cuando cada fila representa un bloque. Variables típicas de bloque son día de la semana, turnos, o localización de la producción.

En este ejemplo, las filas no tienen significado, por lo que es apropiado el test de *Kruskal-Wallis*:

Prueba de Kruskal-Wallis		
	Tamaño de Muestra	Rango Promedio
A	12	40,7917
B	12	25,7917
C	12	19,25
D	12	12,1667

Estadístico = 27,3735 Valor-P = 0,0000491592

Figura 12-11. Test de Rangos Múltiples

La entrada importante de la tabla anterior es el P -valor. Ya que el P -valor es pequeño (menor que 0.05), la hipótesis de igualdad de medianas se rechaza).

Se pueden comparar también pares de medianas seleccionando *Gráfico de caja y bigotes* del cuadro de diálogo *Tablas y Gráficos* y utilizando *Opciones de ventana* para añadir muescas:

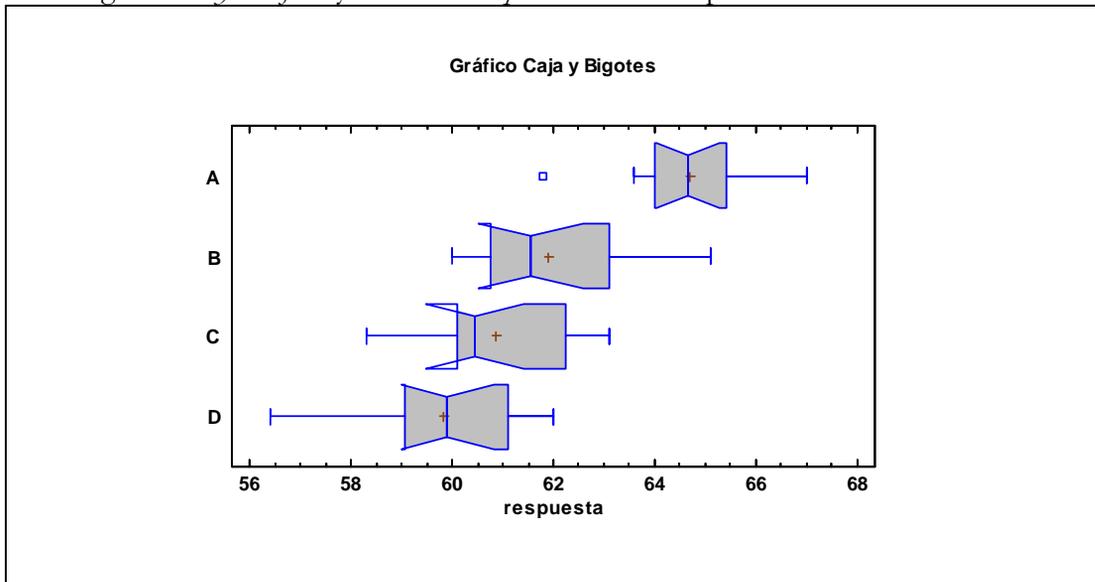


Figura 12-12. Gráfico de caja y bigotes con muescas en la mediana

El rango cubierto por cada muesca muestra el intervalo de confianza estimado para la mediana de cada grupo. Las muescas son escaladas de modo que dos muescas no solapadas tienen medianas diferentes significativamente al nivel de significación por defecto (usualmente 5%). En el gráfico anterior, las muescas para las muestras B, C y D se solapan, sin embargo, la mediana para la muestra A es significativamente superior a la de las otras tres muestras.

NOTA: El comportamiento observado en la *figura 12-12* ocurre cuando una muesca se extiende más allá del eje de la caja.

12.5 Comparando desviaciones típicas

También es posible contrastar la hipótesis de igualdad de desviaciones típicas:

Hipótesis nula: $\sigma_A = \sigma_B = \sigma_C = \sigma_D$

Hipótesis alternativa: las desviaciones típicas no son iguales

Esto se lleva a cabo seleccionando *Verificación de la varianza* en el cuadro de diálogo *Tablas y Gráficos*:

Verificación de Varianza		
	Prueba	Valor-P
Levene's	0,143286	0,933432

Figura 12-13. Comparación de varianzas de muestras

Se mostrará uno de entre cuatro test, dependiendo de las características para *Opciones de ventana*. Tres de los cuatro test, incluyendo el test de Levene, muestran P-valores. Un P-valor menor que 0.05 lleva al rechazo de la hipótesis nula de igualdad de desviaciones típicas al 5% de nivel de significación. En este caso, las desviaciones típicas no son significativamente diferentes la una de la otra, ya que el P-valor es bastante superior a 0.05.

En resumen, se observa que la *Resistencia* media es diferente para distintos materiales. Sin embargo, la variabilidad entre aparatos hechos de la misma materia está cercana a ser la misma a través de los cuatro materiales.

12.6 Gráficos de los residuos

Siempre que se ajusta un modelo estadístico a los datos, es importante examinar los residuos del modelo ajustado. En este análisis, hay un residuo correspondiente a cada uno de los $n = 48$ aparatos, definidos como la diferencia entre la resistencia de los aparatos y la resistencia media de todos los aparatos fabricados del mismo material.

La caja de diálogo *Gráficos* contiene una entrada para generación automática de gráficos de residuos. En la sección *Opciones de ventana*, se pueden representar residuos por grupos, contra valores predichos, o un orden de fila definido en la hoja de datos. El gráfico siguiente muestra los residuos contra valores predichos de *resistencia*:

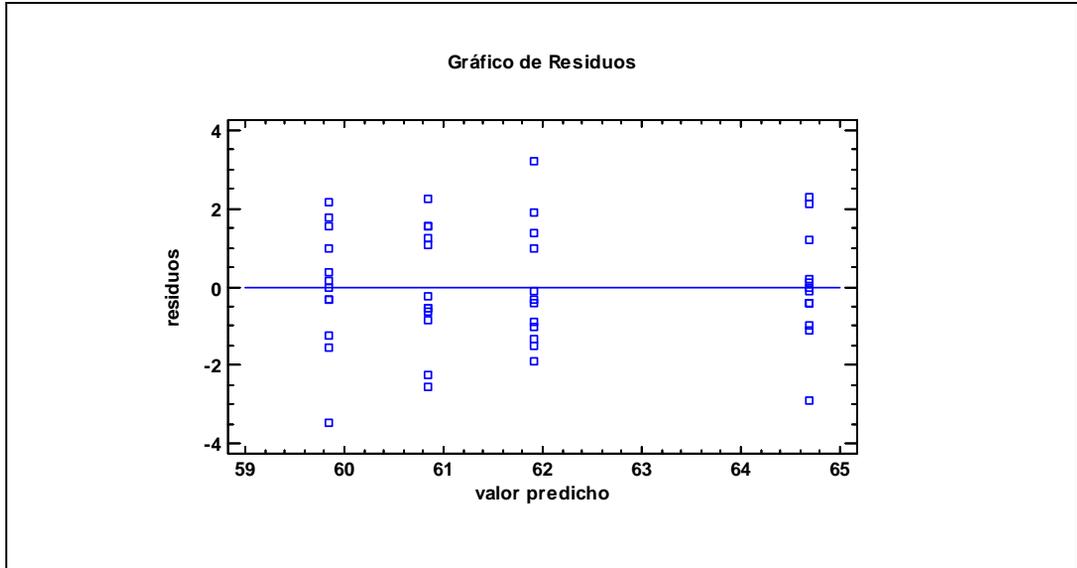


Figura 12-14. Gráfico de residuos contra resistencias predichas

En este tipo de gráficos, debe observar lo siguiente:

1. *Outliers* – residuos aislados respecto de los demás. Tales puntos necesitarán ser investigados posteriormente para determinar cuándo existe una causa asignable que explique su comportamiento inusual.
2. *Heteroscedasticidad* – un cambio sistemático en las varianzas de los valores predichos creciente o decreciente. Esta condición aparece típicamente mediante una apariencia en forma de embudo en el gráfico necesitando una transformación de las observaciones originales tomando logaritmos de los datos antes de ejecutar el análisis. Procedimientos tales como *Test de los rangos múltiples* no trabajarán adecuadamente cuando la variabilidad dentro de grupos difiera significativamente entre los grupos.

Si se desea, los residuos pueden ser guardados como una columna de la hoja de datos presionando el botón *Guardar resultados*  en la barra de herramientas de análisis.

12.7 Análisis de gráficos de medias (ANOM)

Un camino diferente para comparar varias muestras consiste en utilizar *Análisis del gráfico de medias*, también disponible en el cuadro de diálogo *Tablas y Gráficos*:

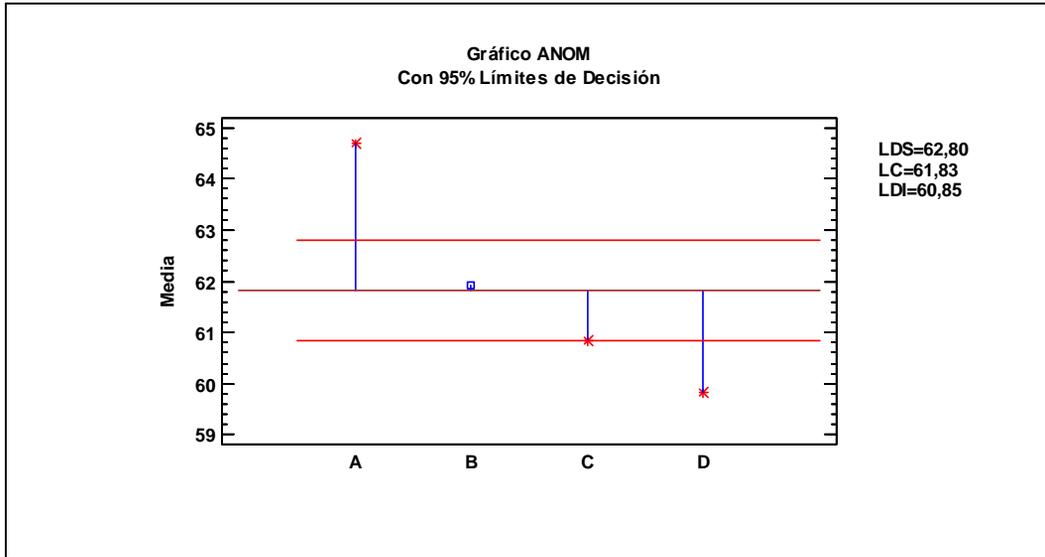


Figura 12-15. Análisis del gráfico de medias

Con un diseño similar al de un gráfico de control, este gráfico muestra cada media muestral junto con una línea vertical dibujada en la gran media de todas las observaciones. Los límites de decisión están incluidos por arriba y por debajo de la gran media. Cualquier media simple que caiga fuera de los límites puede ser declarada como significativamente diferente de la gran media.

En este caso, la interpretación dice que los aparatos provenientes de la muestra A son significativamente más resistentes que la media, mientras que los aparatos de las muestras C y D son significativamente más débiles que la media. Este tipo de interpretación puede ser algunas veces muy usual.

Tutorial #4: análisis de la regresión

Ajustando modelos lineales y no lineales, seleccionando el mejor modelo, representando residuos y mostrando resultados.

Una de las secciones más amplias de STATGRAPHICS Centurion XVI es el conjunto de procedimientos que ajustan modelos de regresión estadística. En un modelo de regresión, una variable respuesta Y se expresa en función de una o más variables predictoras X , más un ruido (o error). En la mayoría de los casos (sin embargo no en todos), la forma funcional en los coeficientes desconocidos es lineal, de modo que el modelo se expresa como sigue:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \dots + \beta_k X_{k,i} + \varepsilon_i$$

donde el subíndice i representa la i -ésima observación en la muestra de datos, los β son los coeficientes desconocidos del modelo y ε es una desviación aleatoria, habitualmente con distribución normal de media 0 y desviación típica σ .

Dado un conjunto de datos con una variable respuesta Y y una o más posibles variables predictoras, la finalidad del análisis de la regresión es construir un modelo que:

1. describa las relaciones que existen entre las variables de tal manera que sea posible predecir Y para valores conocidos de las X .
2. contiene las X necesarias para generar buenas predicciones.

La última consideración se denomina *parsimonia*. Habitualmente, los modelos que envuelven un pequeño conjunto de buenos predictores son los mejores en la práctica.

Este capítulo considera varios tipos de modelos de regresión. Como ejemplo, las millas por galón que recorre un automóvil en ciudad para los coches del archivo *93cars.sgd* servirán como variable respuesta Y. La finalidad es construir un modelo de las otras columnas del archivo que pueda predecir con éxito las millas por galón para un automóvil.

13.1 Análisis de la correlación

Una herramienta habitual para comenzar el análisis de la regresión es el procedimiento *Análisis de variables múltiple*. Este procedimiento se obtiene del menú principal de dos formas:

1. Si usa el menú clásico, seleccione *Describir – Datos numéricos – Análisis de variables múltiples*.
2. Si usa el menú Seis Sigma, seleccione *Análizar – Datos de variable – Métodos multivariantes – Análisis de variables múltiples*.

El análisis comienza mostrando el siguiente cuadro de entrada de datos:

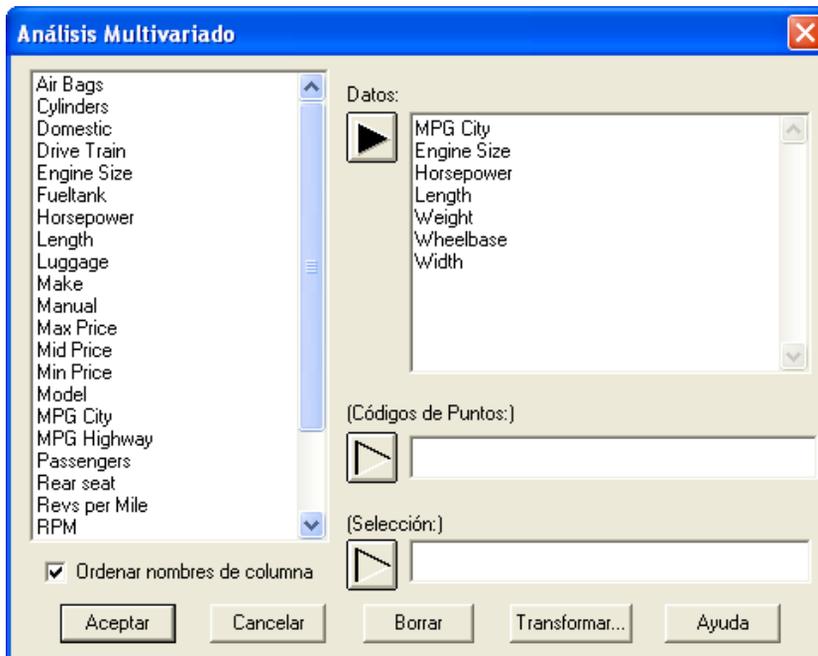


Figura 13-1. Cuadro de diálogo de entrada de datos de *Análisis de variables múltiple*

Se seleccionan seis posibles predictores, en suma a *MPG City*. Los predictores potenciales son:

- X_1 : *Tamaño del motor* (litros)
- X_2 : *Potencia* (máxima)
- X_3 : *Longitud* (pulgadas)
- X_4 : *Peso* (libras)
- X_5 : *Base del volante* (pulgadas)
- X_6 : *Anchura* (pulgadas)

Presionando *Aceptar* se muestra el menú *Opciones*, el cuadro de diálogo *Tablas y Gráficos* y la ventana de análisis:

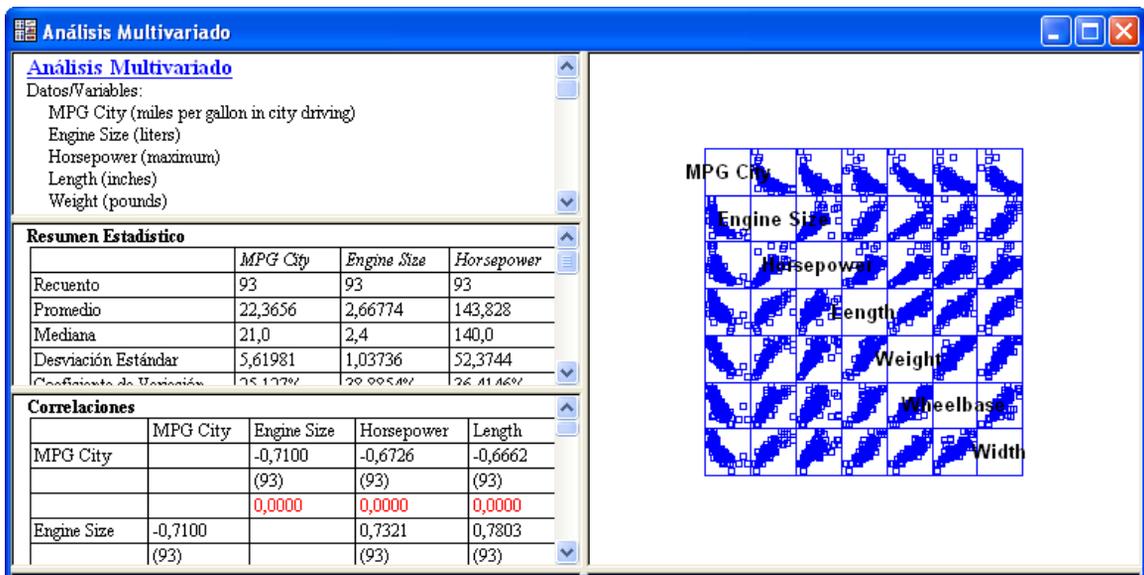


Figura 13-2. Ventana de Análisis de variables múltiples

El panel superior izquierdo lista las variables de entrada, mientras que el panel central izquierdo muestra el resumen de estadísticos. Hay un total de 93 filas en el archivo de datos que tienen información completa en todas las variables a analizar.

El gráfico matricial de la derecha muestra los gráficos X-Y para cada par de variables:

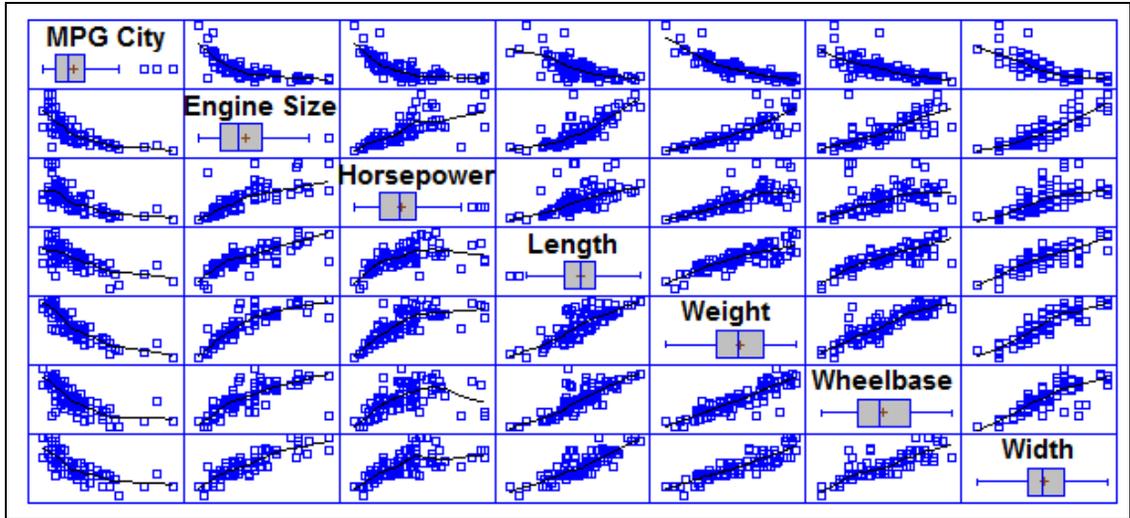


Figura 13-3. Gráfico matricial con alisado añadido

Para interpretar el gráfico, observe la etiqueta de una variable, tal como *MPG City*. La variable indicada se muestra en el eje vertical de todos los gráficos de su fila y en el eje horizontal de todos los gráficos de su columna. Cada par de variables se muestra de este modo dos veces, una vez por encima de la diagonal y otra vez por debajo.

En el gráfico anterior se han añadido alisados robustos LOWESS maximizando el panel y utilizando el botón *Alisado/Rotación* de la barra de herramientas de análisis. Del mayor interés es la fila superior de gráficos, que muestra *MPG City* graficada contra cada una de las 6 variables predictoras potenciales. Todas las variables están claramente correlacionadas con las millas por galón, algunas de forma no lineal. Hay también mucha correlación entre las variables predictoras y, por lo tanto, presencia de multicolinealidad, lo cual sugiere que algunas combinaciones diferentes de variables pueden ser igualmente buenas para predecir *Y*.

La tabla siguiente muestra la matriz de coeficientes de correlación estimados para cada par de variables en el análisis:

Correlaciones							
	MPG City	Engine Size	Horsepower	Length	Weight	Wheelbase	Width
MPG City		-0.7100	-0.6726	-0.6662	-0.8431	-0.6671	-0.720
		(93)	(93)	(93)	(93)	(93)	(93)
		0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Engine Size	-0.7100		0.7321	0.7803	0.8451	0.7325	0.8671
	(93)		(93)	(93)	(93)	(93)	(93)
	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000
Horsepower	-0.6726	0.7321		0.5509	0.7388	0.4869	0.6444
	(93)	(93)		(93)	(93)	(93)	(93)
	0.0000	0.0000		0.0000	0.0000	0.0000	0.0000
Length	-0.6662	0.7803	0.5509		0.8063	0.8237	0.8221
	(93)	(93)	(93)		(93)	(93)	(93)
	0.0000	0.0000	0.0000		0.0000	0.0000	0.0000
Weight	-0.8431	0.8451	0.7388	0.8063		0.8719	0.8750
	(93)	(93)	(93)	(93)		(93)	(93)
	0.0000	0.0000	0.0000	0.0000		0.0000	0.0000
Wheelbase	-0.6671	0.7325	0.4869	0.8237	0.8719		0.8072
	(93)	(93)	(93)	(93)	(93)		(93)
	0.0000	0.0000	0.0000	0.0000	0.0000		0.0000
Width	-0.7205	0.8671	0.6444	0.8221	0.8750	0.8072	
	(93)	(93)	(93)	(93)	(93)	(93)	
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	

Correlación
(Tamaño de la muestra)
P-Valor

Figura 13-4. Matriz de correlaciones

La tabla muestra los coeficientes de correlación para cada par de variables, el número de observaciones utilizadas en la estimación y un P -valor. Un coeficiente de correlación r es un número entre -1 y $+1$, que mide la intensidad de la relación lineal entre las dos variables. Los valores de máxima correlación son -1 (correlación negativa) y $+1$ (correlación positiva). El signo de la correlación indica su dirección. Un valor positivo indica que Y aumenta cuando X aumenta. Una correlación negativa indica que Y disminuye cuando X aumenta.

Para determinar cuándo un par de variables está efectivamente correlacionado, se calcula el P -valor de su coeficiente de correlación (test del coeficiente de correlación). Si el P -valor es menor o igual que 0.05 la correlación lineal de las dos variables es estadísticamente significativa al 5% de nivel de confianza.

La fila superior muestra la correlación entre *MPG City* y los 6 predictores. La correlación más fuerte se produce con *Peso* y vale -0.8431 . El signo negativo implica que las millas por galón y el peso varíen en sentido contrario, lo cual no sorprende.

13.2 Regresión simple

El primer modelo estadístico que se ajustará será la línea recta de la forma:

$$MPG\ City = \beta_0 + \beta_1 Peso + \varepsilon$$

En la ecuación anterior, β_1 es la pendiente de la línea en unidades de millas por galón por libra, mientras β_0 es la ordenada en el origen de Y. Para ajustar el modelo:

1. Si usa el menú clásico, seleccione *Relacionar – Un Factor – Regresión simple*.
2. Si usa el menú Seis Sigma, seleccione *Mejorar – Análisis de regresión – Un Factor – Regresión simple*.

El cuadro de diálogo de entrada de datos se cumplimenta como sigue:

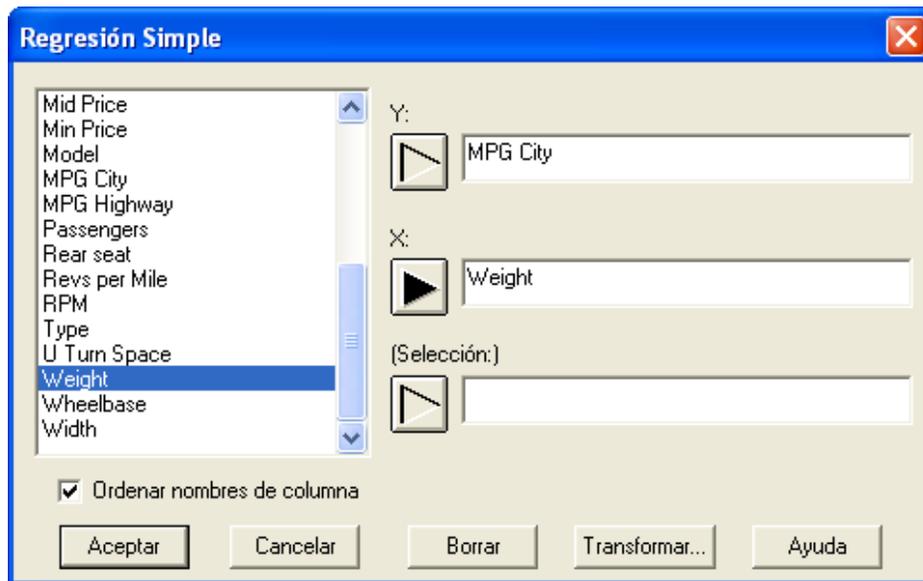


Figura 13-5. Cuadro de diálogo de entrada de datos de Regresión simple

Después del menú *Opciones* y del cuadro de diálogo *Tablas y Gráficos*, la ventana inicial tiene cuatro paneles mostrando información acerca del modelo ajustado y de los residuos:

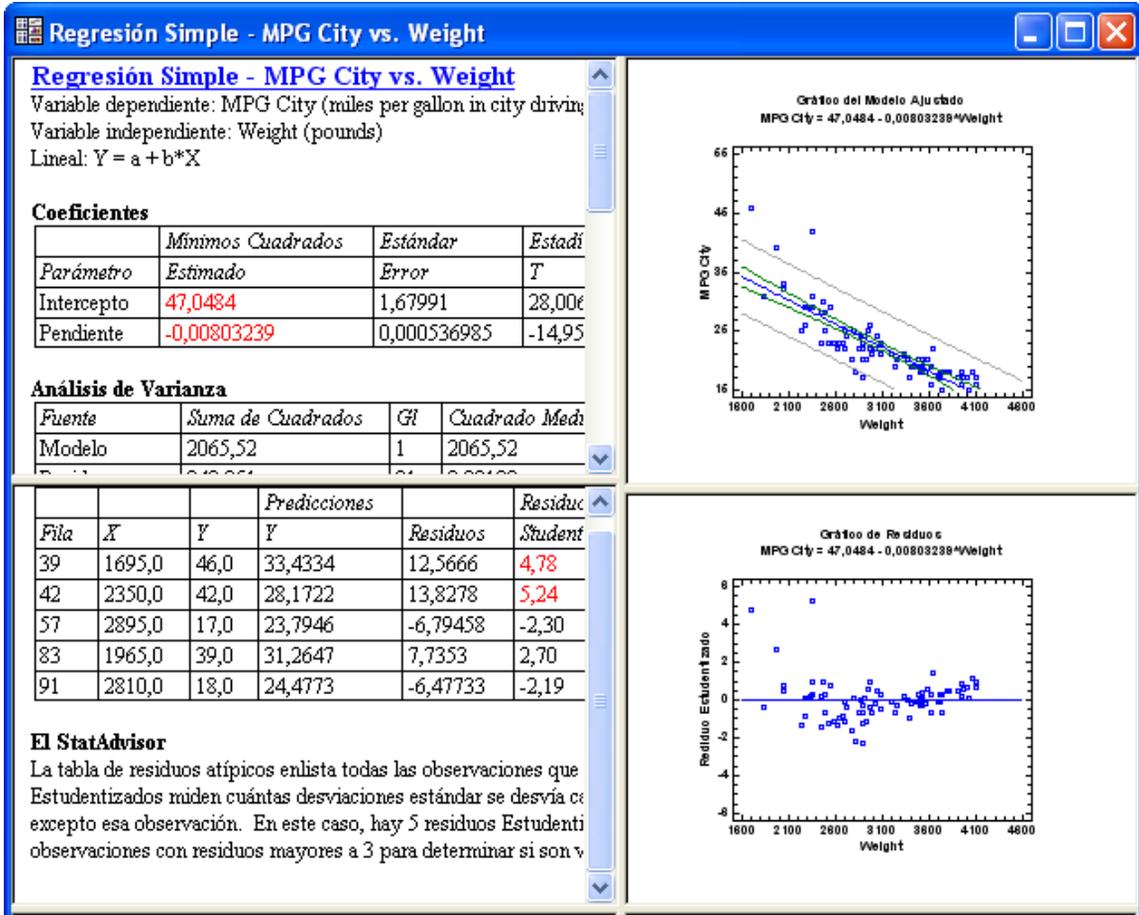


Figura 13-6. Ventana de análisis de Regresión simple

El Resumen de análisis en el panel superior izquierdo resume el ajuste:

Regresión Simple - MPG City vs. Weight

Variable dependiente: MPG City (miles per gallon in city driving)

Variable independiente: Weight (pounds)

Lineal: $Y = a + b \cdot X$

Coefficientes

	<i>Mínimos Cuadrados</i>	<i>Estándar</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimado</i>	<i>Error</i>	<i>T</i>	<i>Valor-P</i>
Intercepto	47,0484	1,67991	28,0064	0,0000
Pendiente	-0,00803239	0,000536985	-14,9583	0,0000

Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	2065,52	1	2065,52	223,75	0,0000
Residuo	840,051	91	9,23133		
Total (Corr.)	2905,57	92			

Coefficiente de Correlación = -0,843139

R-cuadrada = 71,0883 por ciento

R-cuadrado (ajustado para g.l.) = 70,7705 por ciento

Error estándar del est. = 3,03831

Error absoluto medio = 1,99274

Estadístico Durbin-Watson = 1,64586 (P=0,0405)

Autocorrelación de residuos en retraso 1 = 0,176433

Figura 13-7. Resumen de análisis de Regresión simple

Entre los muchos estadísticos de la tabla anterior, los más importantes son los siguientes:

1. **Coefficientes:** coeficientes del modelo estimado. El modelo ajustado que se utilizará para la predicción es:

$$MPG\ City = 47.0484 - 0.00803239 \textit{peso}$$

2. **R-cuadrado:** el porcentaje de variabilidad en Y que ha sido explicado por el modelo. En este caso, la regresión lineal contra *Peso* explica cerca del 71.1% de la variabilidad en *MPG City*.
3. **P-Valor del modelo :** Un *P*-valor inferior a 0.05, como en el ejemplo actual, indica que *Peso* es un buen predictor para *MPG City*.

En el gráfico del panel superior derecho muestra el modelo ajustado:

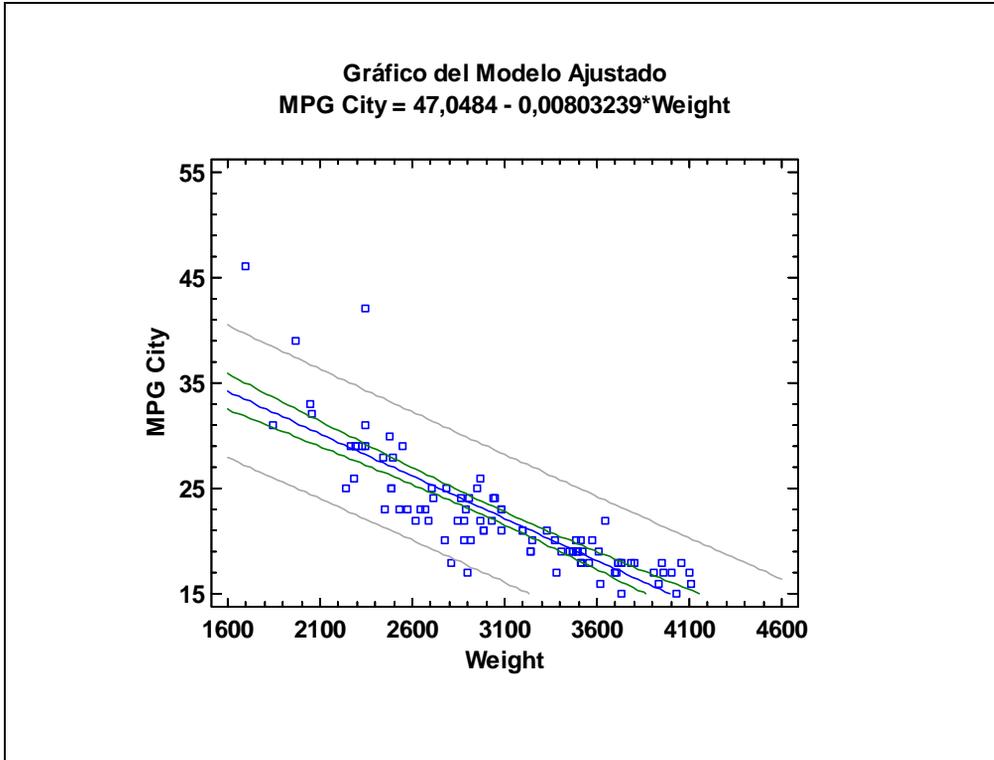


Figura 13-8. Gráfico del modelo lineal ajustado

El gráfico muestra la línea de regresión por mínimos cuadrados y dos conjuntos de límites. Los límites interiores son intervalos de confianza al 95% para el valor medio de Y dado un X. Esto indica la calidad de la estimación de los puntos de la línea de regresión, supuesto que la relación es lineal. A mayor tamaño de muestra para la estimación, intervalos más estrechos. Las líneas exteriores son límites de predicción al 95% para nuevas observaciones. Se estima que el 95% de observaciones adicionales, similares a las de este ejemplo, caerán entre las bandas.

Es permisible que 3 observaciones de valores bajos de *Peso* caigan más allá de los límites de predicción del 95%. Esto puede ser indicativo de la presencia de valores atípicos o de un fallo en el modelo de no linealidad en la relación actual entre *MPG City* y *Peso*.

13.3 Ajustando un modelo no lineal

El procedimiento *Regresión simple* incluye la posibilidad de ajustar una amplia variedad de modelos no lineales. Para evaluar la mejora relativa que pueden aportar varios modelos, seleccione *Comparación de modelos alternativos* del cuadro de diálogo *Tablas y Gráficos*. Esto ajustará los posibles modelos y los listará en orden decreciente de R cuadrado:

Comparación de Modelos Alternos		
<i>Modelo</i>	<i>Correlación</i>	<i>R-Cuadrada</i>
Curva S	0,9016	81,29%
Inversa-Y Raíz Cuadrada-X	0,8995	80,92%
Inversa-Y Log-X	0,8995	80,90%
Raíz Cuadrada-Y Inversa de X	0,8988	80,78%
Multiplicativa	-0,8981	80,65%
Inversa de Y	0,8969	80,44%
Logarítmico-Y Raíz Cuadrada-X	-0,8919	79,54%
Doble Inverso	-0,8896	79,14%
Inversa de X	0,8888	79,00%
Raíz Cuadrada-Y Log-X	-0,8879	78,83%
Inversa-Y Cuadrado-X	0,8852	78,35%
Exponencial	-0,8833	78,03%
Raíz Cuadrada Doble	-0,8784	77,16%
Logaritmo de X	-0,8705	75,78%
Raíz Cuadrada de Y	-0,8668	75,14%
Log-Y Cuadrado-X	-0,8611	74,15%
Raíz Cuadrada de X	-0,8577	73,56%
Cuadrado-Y Inversa de X	0,8472	71,77%
Lineal	-0,8431	71,09%
Raíz Cuadrada-X Cuadrado-X	-0,8393	70,44%
Cuadrado-Y Log-X	-0,8146	66,35%
Cuadrado de X	-0,8106	65,71%
Cuadrado-Y Raíz Cuadrada-X	-0,7957	63,31%
Cuadrado de Y	-0,7758	60,18%
Cuadrado Doble	-0,7346	53,96%
Logístico	<sin ajuste>	
Log probit	<sin ajuste>	

Figura 13-9. Modelos no lineales alternativos

Los modelos al principio de la lista explican el mayor porcentaje de la variación en la variable respuesta. R-cuadrado es sólo un criterio que puede ayudar a elegir el modelo. Modelos con valores de R-cuadrado más bajo que le modelo del principio de la lista pueden ser preferibles si son más sensibles en el contexto de los datos.

En el contexto actual, un modelo atractivo cercano al principio de la lista es el modelo *Recíproco-Y*. Este modelo toma la forma:

$$\frac{1}{MPG_{City}} = \beta_0 + \beta_1 \text{Peso} + \varepsilon$$

En él, el recíproco de las millas por galón está expresado como una función lineal del peso. Es frecuente que transformaciones de Y, X, o ambas puedan aventajar a los mejores modelos. Para ajustar el modelo *Recíproco-Y*, presione el botón *Opciones de análisis* y seleccione *Recíproco-Y* en el cuadro de diálogo. El ajuste resultante se muestra a continuación:

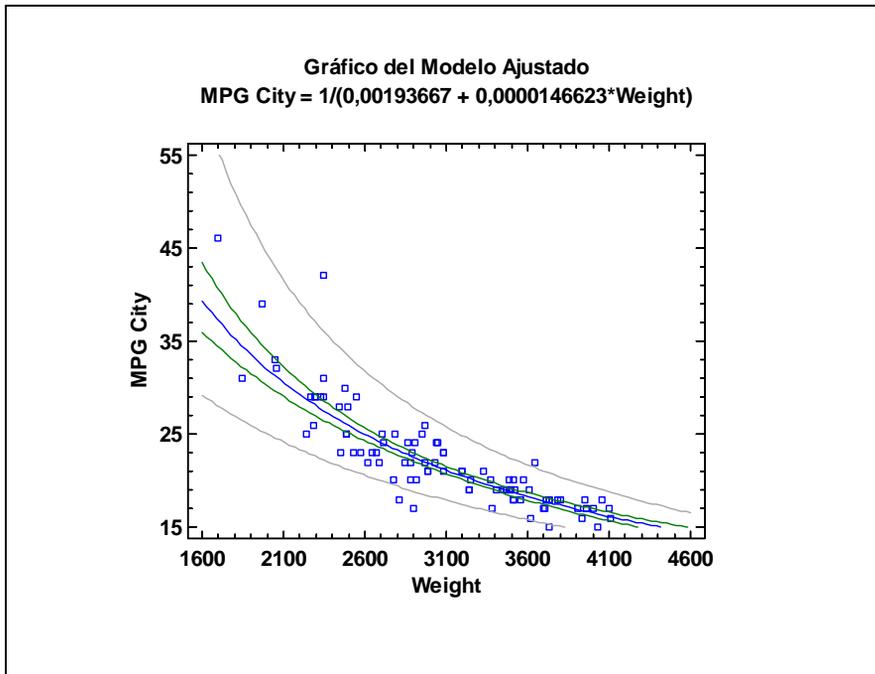


Figura 13-10. Modelo *Recíproco-Y* ajustado

Aunque el modelo es lineal en el recíproco de *MPG City*, el modelo es no lineal en la métrica original. Hay que notar también que los límites de predicción de *Peso* se hacen más largos. Esto tiene sentido en el contexto de los datos, ya que ello implica que hay más variabilidad entre los coches ligeros que entre los coches pesados.

13.4 Examinando los residuos

Una vez que se ha ajustado un modelo razonable, hay que examinar los residuos del ajuste. En general, un residuo puede ser observado a través de la diferencia entre el valor observado de Y y el valor predicho por el modelo:

$$\text{residuo} = Y \text{ observado} - Y \text{ predicho}$$

El análisis de *Regresión Simple* automáticamente grafica los residuos frente a la variable X:

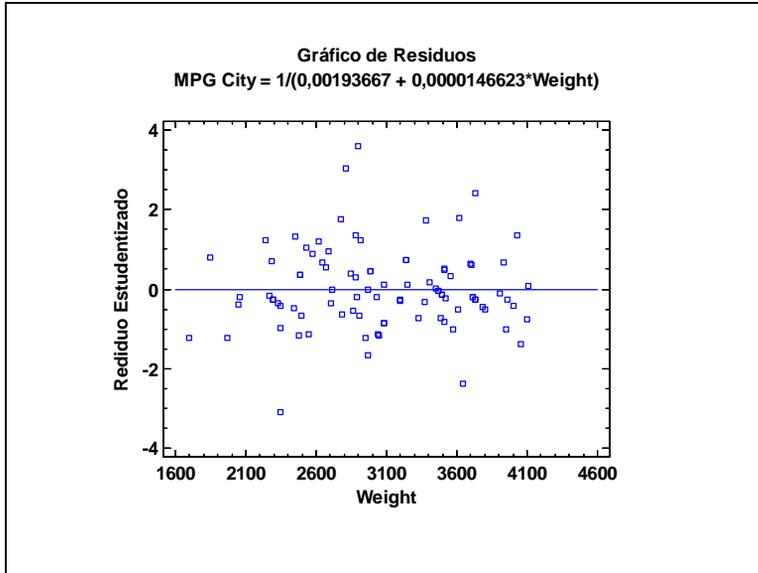


Figura 13-11. Gráfico de los residuos estudentizados

Usando *Opciones de ventana*, puede elegir entre graficar residuos simples o residuos estudentizados. Los residuos estudentizados se obtienen dividiendo los residuos ordinarios por sus errores estándar estimados. Un residuo estudentizado indica qué cantidad de error estándar de los datos proviene del modelo ajustado.

STATGRAPHICS Centurion XVI calcula actualmente residuos estudentizados *borrados*. Los residuos *borrados* se calculan eliminando una observación, reajustando el modelo, y determinando el número de errores estándar que se separa del nuevo modelo ajustado. Así se observan los valores atípicos que tienen un gran impacto en el modelo cuando se calculan los residuos.

La selección de *Residuos atípicos* en el cuadro de diálogo *Tablas y Gráficos* lista todos los residuos estudentizados que son mayores que 2 en valor absoluto:

Residuos Atípicos					
			Predicciones		Residuos
Fila	X	Y	Y	Residuos	Studentizados
5	3640,0	22,0	18,0808	3,91924	-2,38
36	3735,0	15,0	17,6366	-2,63658	2,41
42	2350,0	42,0	27,4778	14,5222	-3,11
57	2895,0	17,0	22,5306	-5,53064	3,60
91	2810,0	18,0	23,1816	-5,18157	3,04

Figura 13-12. Tabla de residuos atípicos

Residuos estudentizados mayores que 3, como el de la fila #57, son atípicos potenciales que parecen no pertenecer al resto de los datos. La fila #57 corresponde al Mazda RX-7 que se registra en el archivo con sólo 17 millas por galón conduciendo en ciudad, aunque el modelo predice 22.5 mpg. En la sección siguiente se añaden variables adicionales al modelo, lo que debe ayudar a su capacidad predictiva para tales coches deportivos y la fila #57 no se excluirá del modelo para el ajuste, aunque haya que prestarle mucha atención.

13.5 Regresión múltiple

Para mejorar el modelo, es necesario añadir otras variables predictoras. Esto se logra más fácilmente utilizando análisis de *Regresión múltiple*, que se define en el menú principal bajo:

1. Si usa el menú clásico, seleccione *Relacionar – Factores múltiples – Regresión múltiple*.
2. Si usa el menú Seis Sigma, seleccione *Mejorar – Análisis de la regresión – Factores múltiples – Regresión múltiple*.

El cuadro de diálogo de entrada de datos tiene la siguiente forma:

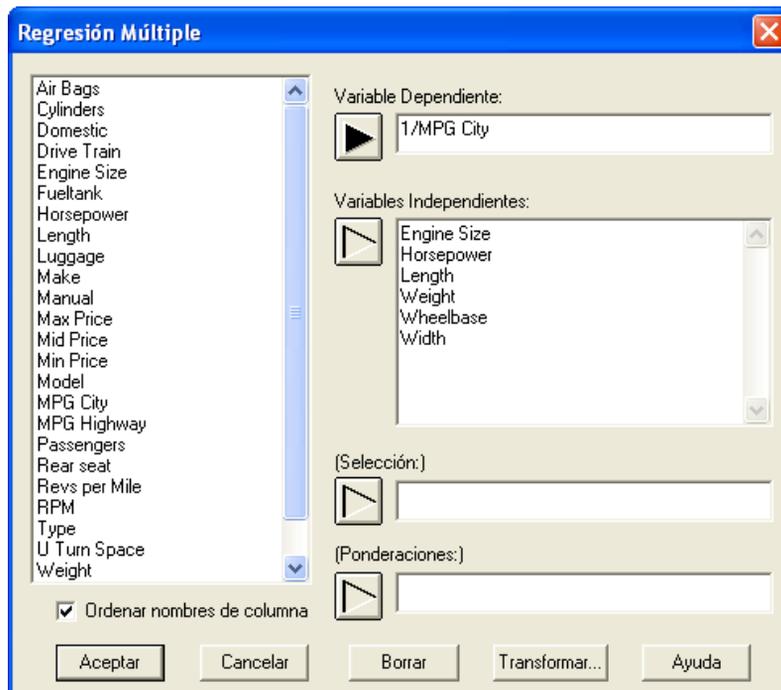


Figura 13-13. Cuadro de diálogo de entrada de datos de Regresión múltiple

Para comenzar, los 6 predictores considerados en el procedimiento *Análisis de múltiples variables* discutidos anteriormente se introducirán como variables independientes. La variable dependiente es el recíproco de *MPG City*, que identifica las millas por galón. A continuación se utiliza el menú *Opciones* y se muestra el cuadro de diálogo de *Tablas y Gráficos*. El resumen de análisis resultante se muestra a continuación:

Regresión Múltiple - 1/MPG City

Variable dependiente: 1/MPG City

Variables independientes:

- Engine Size (liters)
- Horsepower (maximum)
- Length (inches)
- Weight (pounds)
- Wheelbase (inches)
- Width (inches)

		Error	Estadístico	
Parámetro	Estimación	Estándar	T	Valor-P
CONSTANTE	0,0155897	0,0177088	0,880334	0,3811
Engine Size	0,00072849	0,000980504	0,742974	0,4595
Horsepower	0,0000132632	0,000014911	0,889485	0,3762
Length	-0,000101355	0,0000608857	-1,66468	0,0996
Weight	0,0000149727	0,00000242804	6,1666	0,0000
Wheelbase	-0,000148122	0,000163073	-0,908321	0,3662
Width	0,000223526	0,00028967	0,771658	0,4424

Análisis de Varianza

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Modelo	0,00705967	6	0,00117661	67,64	0,0000
Residuo	0,001496	86	0,0000173954		
Total (Corr.)	0,00855567	92			

R-cuadrada = **82,5145** por ciento

R-cuadrado (ajustado para g.l.) = **81,2946** por ciento

Error estándar del est. = **0,00417077**

Error absoluto medio = **0,00304978**

Estadístico Durbin-Watson = 1,6264 (P=**0,0306**)

Autocorrelación de residuos en retraso 1 = 0,186005

El StatAdvisor

La salida muestra los resultados de ajustar un modelo de regresión lineal múltiple para describir la relación entre 1/MPG City y 6 variables independientes. La ecuación del modelo ajustado es

$$1/\text{MPG City} = 0,0155897 + 0,00072849 * \text{Engine Size} + 0,0000132632 * \text{Horsepower} - 0,000101355 * \text{Length} + 0,0000149727 * \text{Weight} - 0,000148122 * \text{Wheelbase} + 0,000223526 * \text{Width}$$

Puesto que el valor-P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables con un nivel de confianza del 95,0%.

Figura 13-14. Resumen de análisis de regresión múltiple con 6 variables predictoras

Observar que el estadístico R cuadrado se ha elevado hasta el 82.5%. Sin embargo, el modelo se ha complicado innecesariamente. Cerca de la parte superior de la salida está la columna de P -valores. Estos P -valores contrastan la hipótesis de que el coeficiente correspondiente a la variable seleccionada es igual a 0 (coeficiente no significativo), en presencia de las variables restantes del modelo. P -valores mayores que 0.05 indican que la variable no contribuye significativamente al ajuste, en presencia del resto de las variables.

Excepto *Peso*, todos los predictores tienen P -valores superiores a 0.05. Esto implica que al menos una de estas variables predictoras debe eliminarse para mantener el modelo significativo.

NOTA: es erróneo asumir en este punto que las 5 variables predictoras con P -valor por encima de 0.05 deben eliminarse. Debido a la alta multicolinealidad en los datos, los P -valores pueden cambiar drásticamente si alguna de las variables se elimina del modelo.

Un método habitual para simplificar el modelo es ejecutar la regresión paso a paso. En cada paso de la regresión, se añaden o eliminan variables de la regresión una cada vez, con la finalidad de obtener un modelo que contiene sólo predictores significativos. La regresión paso a paso está disponible en el cuadro de diálogo de *Opciones de análisis*:

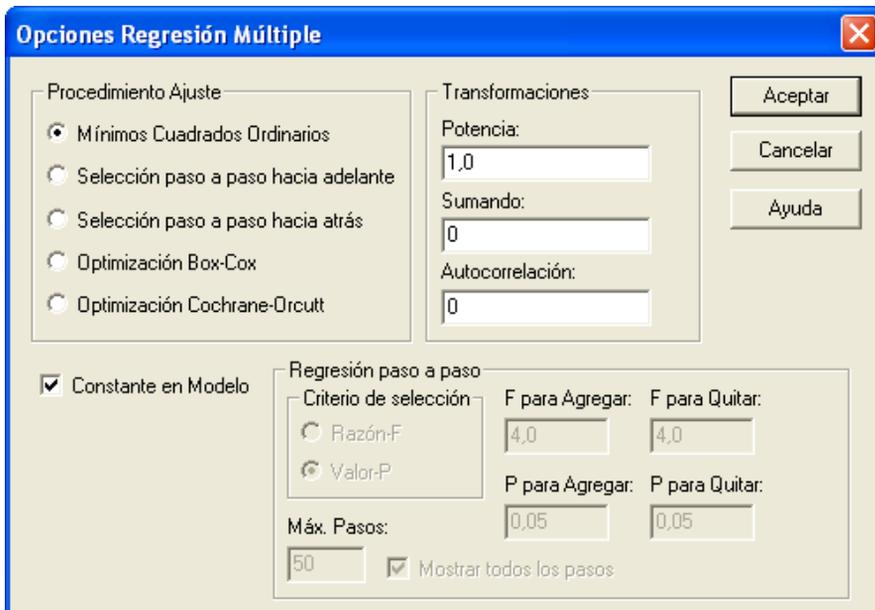


Figura 13-15. Cuadro de diálogo de Opciones de análisis de regresión múltiple

Hay dos opciones paso a paso:

1. *Selección hacia adelante* – comienza con un modelo que contiene sólo la constante e introduce variables de una en una que mejoran la significatividad del ajuste.
2. *Selección hacia atrás* – comienza con todas las variables del modelo y las va eliminando de una en una hasta que el modelo resulta significativo.

En ambos métodos, las variables eliminadas pueden ser introducidas en un paso posterior si deben ser utilizadas como predictores, y las variables introducidas pueden ser eliminadas posteriormente si no aportan suficiente significatividad al modelo.

Ejecutando una regresión hacia atrás se obtienen los siguientes resultados:

Regresión Múltiple - 1/MPG City				
Variable dependiente: 1/MPG City				
<i>Parámetro</i>	<i>Estimación</i>	<i>Error Estándar</i>	<i>Estadístico T</i>	<i>Valor-P</i>
CONSTANTE	0,0034427	0,00243602	1,41325	0,1610
Horsepower	0,0000260839	0,0000124356	2,09752	0,0388
Weight	0,0000129513	0,0000011041	11,7302	0,0000

Análisis de Varianza					
<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	0,00696044	2	0,00348022	196,35	0,0000
Residuo	0,00159524	90	0,0000177249		
Total (Corr.)	0,00855567	92			

R-cuadrada = **81,3546** por ciento
R-cuadrado (ajustado para g.l.) = **80,9403** por ciento
Error estándar del est. = **0,00421009**
Error absoluto medio = **0,00313061**
Estadístico Durbin-Watson = 1,62892 (P=**0,0338**)
Autocorrelación de residuos en retraso 1 = 0,184113

El StatAdvisor
La salida muestra los resultados de ajustar un modelo de regresión lineal múltiple para describir la relación entre 1/MPG City y 6 variables independientes. La ecuación del modelo ajustado es

$$1/MPG \text{ City} = 0,0034427 + 0,0000260839 * \text{Horsepower} + 0,0000129513 * \text{Weight}$$

Puesto que el valor-P en la tabla ANOVA es menor que 0,05, existe una relación estadísticamente significativa entre las variables con un nivel de confianza del 95,0%.

Figura 13-16. Resumen de análisis de Regresión múltiple después de la Selección hacia atrás

Sólo dos variables se han mantenido en el modelo: *Potencia* y *Peso*. Ambas tienen P-valores menores que 0.05.

Una vez que se ha definido la ecuación matemática, se utiliza para graficar la ecuación. Cuando el modelo contiene 2 variables predictoras, la ecuación representa una superficie en tres dimensiones, habitualmente conocida como *superficie de respuesta*. En este caso, la ecuación ajustada corresponde a un plano, ya que *Potencia* y *Peso* se introducen en el modelo de forma lineal.

Para graficar el modelo, puede:

Utilizar el procedimiento *Gráficos de superficie y contorno* copiando la función a graficar y definiendo sus propios títulos y escalas-

1. Si usa el menú clásico, seleccione *Gráficos – Gráficos de superficie y contorno*.
2. Si usa el menú Seis Sigma, seleccione *Herramientas – Gráficos de superficie y contorno*.

En el cuadro de diálogo de entrada de datos, introduzca el modelo, expresando las dos variables predictoras *X* e *Y*. El camino más fácil es pegar la ecuación generada por el procedimiento *regresión múltiple*, cambiando *Potencia* por *X* y *Peso* por *Y*:

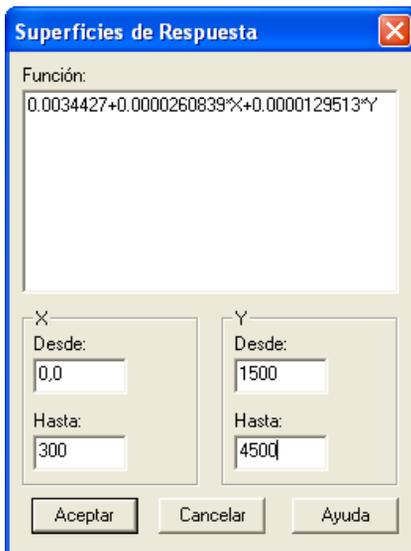


Figura 13-17 Cuadro de diálogo de entrada de datos para *Superficie de respuesta* y *Gráfico de contorno*

El escalado de X e Y deben cambiarse también para representar los datos utilizados en el ajuste del modelo.

Cuando presione *Aceptar*, aparece el cuadro de diálogo *Tablas y Gráficos* y a continuación se genera la superficie de respuesta. El gráfico inicial toma la forma de una superficie:

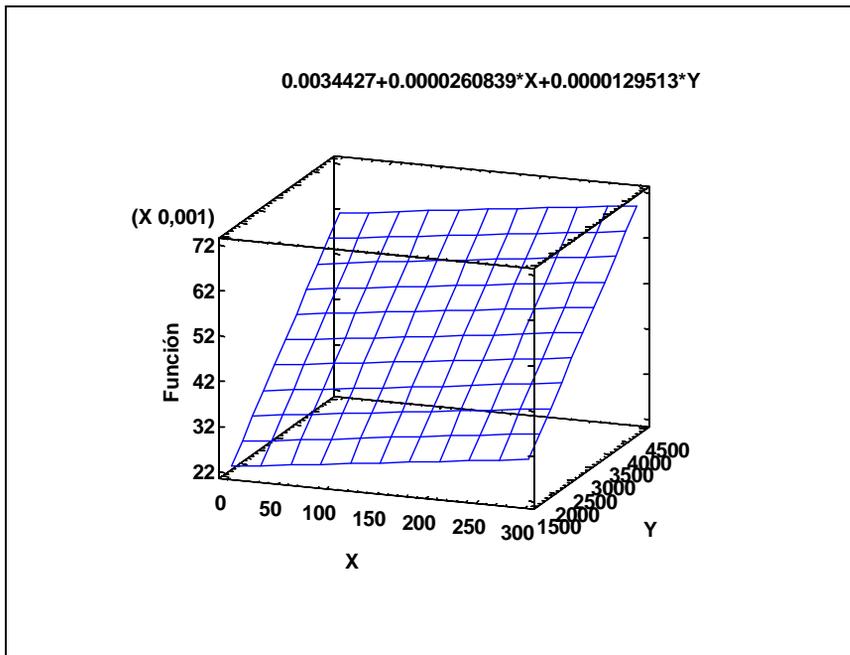


Figura 13-18. Gráfico de superficie con etiquetas y escala por defecto

Puede mejorar el gráfico como sigue:

Seleccionando *Opciones gráficas* de la barra de herramientas de análisis y cambiando las etiquetas y escalas en *Título superior*, *Eje-X*, *Eje-Y* y *Eje-Z*. En particular:

- Cambiar el título del eje X a *Potencia*.
- Cambiar el título del eje Y a *Peso*.
- Cambiar la escala del eje Y de 1500 a 4500 por 1000.
- Cambiar el título del eje Z a *1/MPG City*.

Seleccionando *Opciones de ventana* y cambiando el tipo de gráfico mostrado:

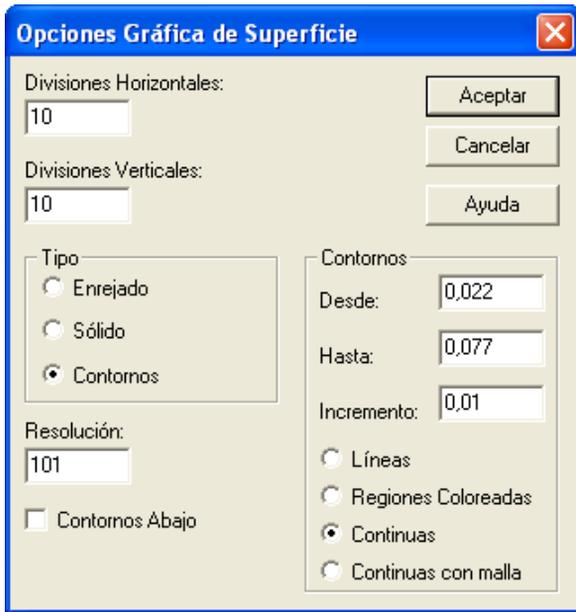


Figura 13-19. Opciones de ventana de superficie de respuesta

En el cuadro de diálogo superior, *Tipo* se ha situado en *Contorneado* y el campo *Contorno* en *Continuo*. El gráfico final se muestra a continuación:

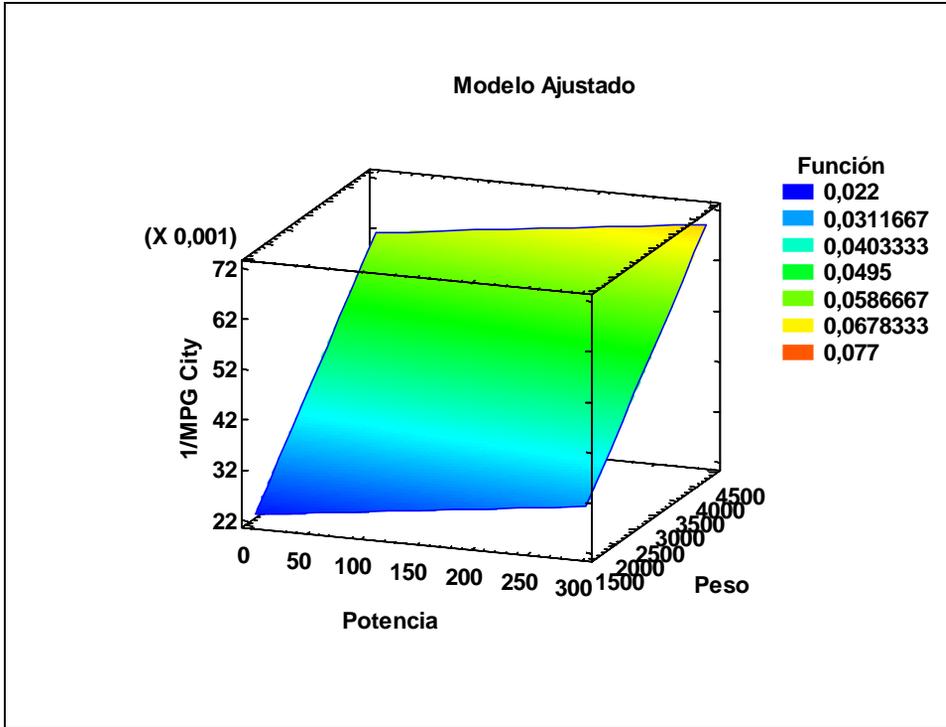


Figura 13-20. Gráfico del modelo ajustado

Los coches que usan más combustible están en la esquina inferior derecha del gráfico: coches grandes con motores grandes.

Tutorial #5: Analizando datos de atributos

Tablas de frecuencias, tablas de contingencia y análisis de Pareto

Cada uno de los primeros cuatro tutoriales trabajan con datos de variables, con las observaciones numéricas en una escala continua. Este tutorial examina un conjunto de datos de atributos, en los cuales una observación representa una categoría en la que se clasifica el atributo, en vez de una medición numérica.

Como ejemplo, consideramos los datos contenidos en el archivo *defects.sgd*. Una parte de este archivo se muestra a continuación:

<i>Defecto</i>	<i>Instalación</i>
Desalineado	Virginia
Contaminado	Texas
Contaminado	Virginia
Contaminado	Texas
Partes desaparecidas	Texas
Desalineado	Virginia
Contaminado	Texas
Filtrando	Texas
Dañado	Virginia
Contaminado	Texas

Los datos tienen $n = 120$ filas, correspondientes cada una a un defecto que fue observado en el proceso de fabricación. El archivo también indica el tipo de defecto producido y las instalaciones en las que fue producido.

14.1 Resumiendo datos de atributos

Ignorando por un momento la instalación en la que cada ítem fue producido, los datos de tipo de defecto pueden ser resumidos mediante:

1. Si usa el menú clásico, seleccione *Describir – Datos categóricos – Tabulación*.
2. Si usa menú Seis Sigma, seleccione *Analizar – Datos de atributos – Un Factor - Tabulación*.

El cuadro de diálogo de entrada de datos espera una columna simple conteniendo datos de atributos:

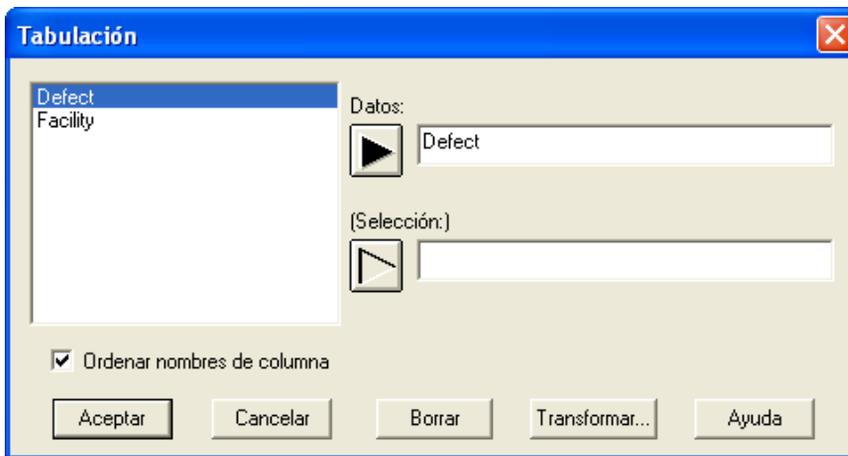


Figura 14-1. Cuadro de diálogo de entrada de datos de tabulación

El procedimiento observa la columna, identificando cada valor único. Aparece el cuadro de diálogo *Tablas y Gráficos* y se genera una ventana de análisis similar a la siguiente:

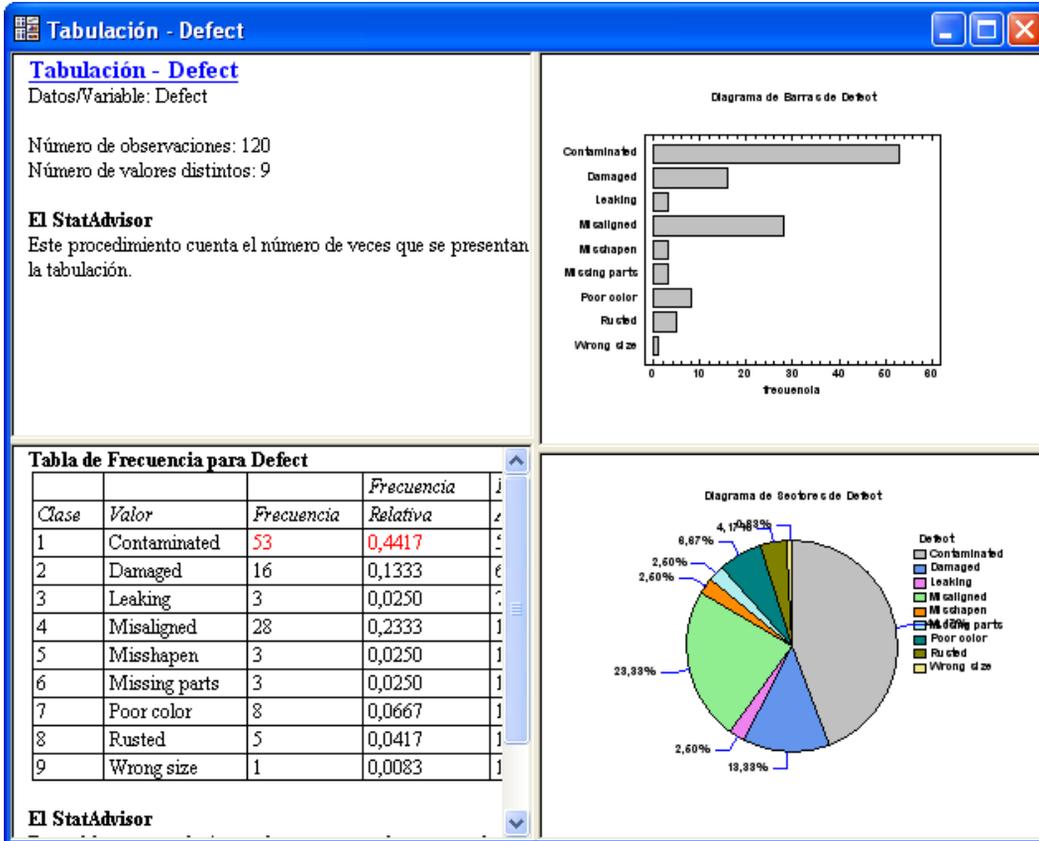


Figura 14-2. Ventana de tabulación de análisis

El panel superior izquierdo muestra que hay 9 valores distintos en las $n = 120$ filas. El gráfico de barras y el gráfico de sectores de la derecha ilustran las frecuencias observadas de cada tipo de defecto, las cuales se tabulan en el panel inferior izquierdo. El tipo más común de defecto es “Contaminación”, que representa cerca del 44% de todos los defectos.

14.2 Análisis de Pareto

El procedimiento *Tabulación de frecuencias* ordena los tipos de defecto en orden alfabético. Para ordenarlos de más a menos frecuentes, se utilice en su lugar el procedimiento *Análisis de Pareto*. Este procedimiento es accesible mediante:

1. Si usa el menú clásico, seleccione *CEP – Evaluación de la calidad – Análisis de Pareto*.

2. Si usa el menú Seis Sigma, seleccione *Analizar – Datos de atributos – Un Factor – Análisis de Pareto*.

El cuadro de diálogo de entrada de datos puede ser completado como sigue:

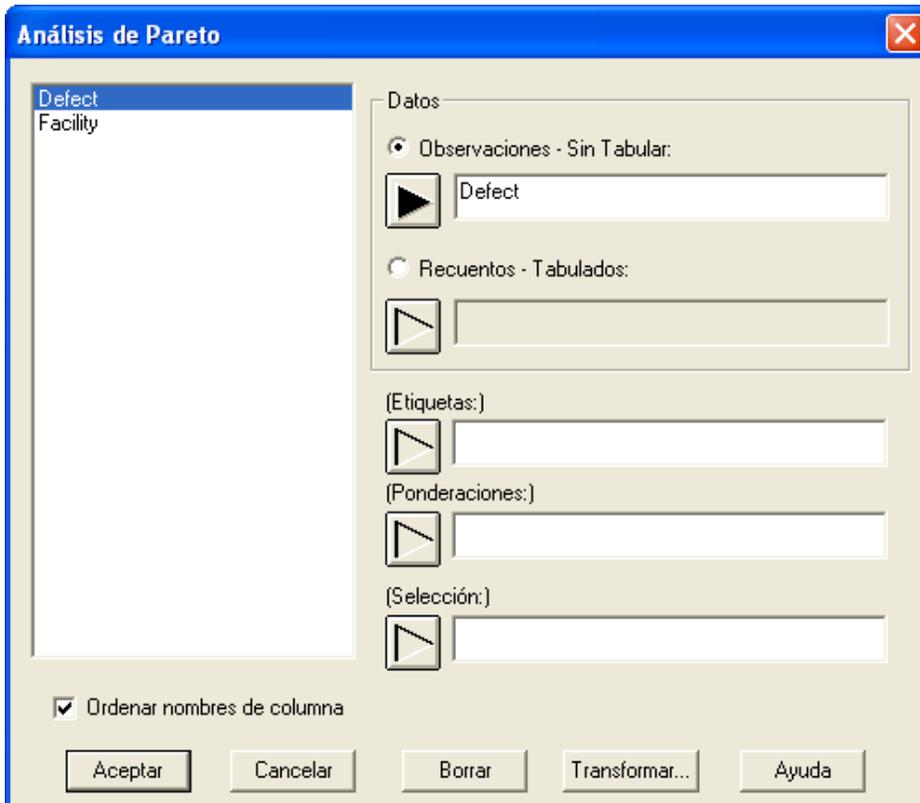


Figura 14-3. Cuadro de diálogo de entrada de datos de Análisis de Pareto

El procedimiento *Análisis de Pareto* acepta datos en dos formatos:

1. Datos no tabulados en una columna, como en el ejemplo actual.
2. Datos que han sido agrupados por tipos de defectos. Aplicable si se tienen dos columnas, una identificando los tipos de defectos y otra conteniendo el número de veces que ocurre cada tipo de defecto.

La ventana de análisis muestra ambas tablas resumen y el gráfico de Pareto:

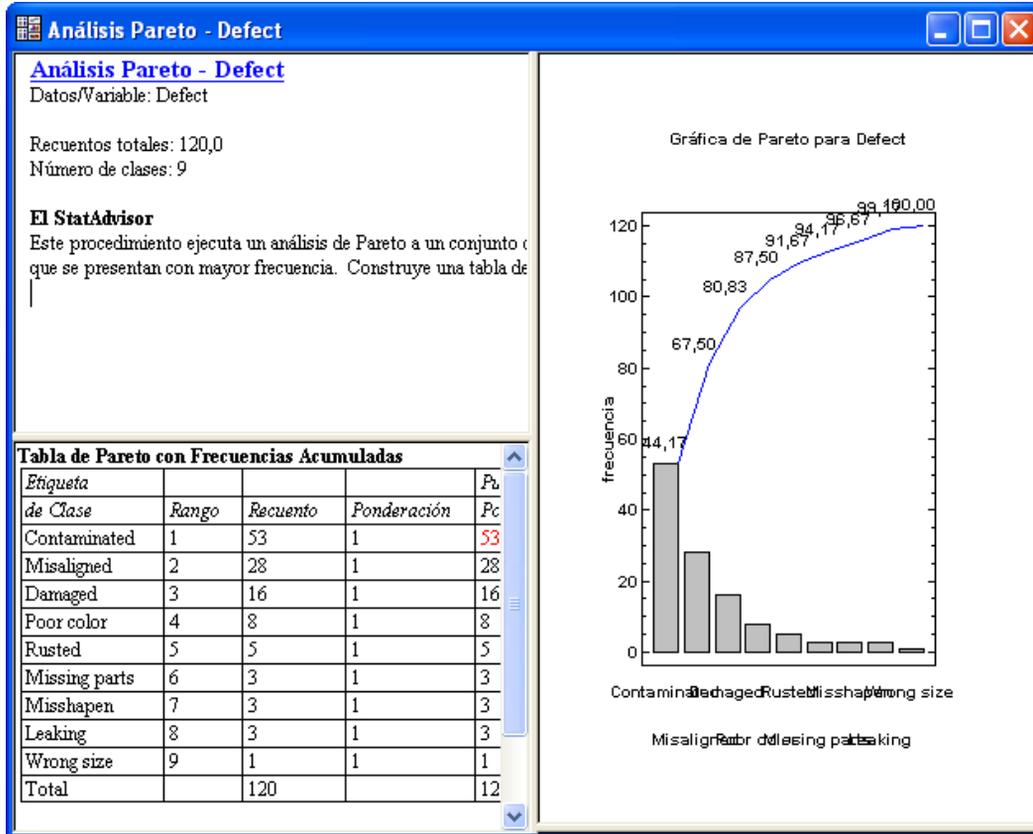


Figura 14-4. Ventana de Análisis de Pareto

De particular interés es el gráfico de Pareto de la derecha, que representa las frecuencias de cada tipo de defecto de los más comunes a los menos comunes. Inicialmente, las etiquetas de las barras se solapan debido a su número y longitud. Este puede resolverse mediante:

1. Doble clic en el gráfico con el ratón para maximizar el panel en la ventana de análisis.
2. Presionando el botón *Opciones gráficas* en la barra de herramientas de análisis, haciendo clic en la solapa *Eje-X*, y marcando la casilla *Rotar etiquetas de los ejes*.
3. Después de salir del cuadro de diálogo *Opciones gráficas*, pueden ajustarse o no completamente a la pantalla. En caso negativo, haga clic con el ratón en la parte principal del gráfico y arrastrando sin levantarlo aumente su altura o anchura, o arrastre el eje X para reducir el tamaño del eje vertical.

Cuando finalice, el gráfico de Pareto debe mostrarse como se indica a continuación:

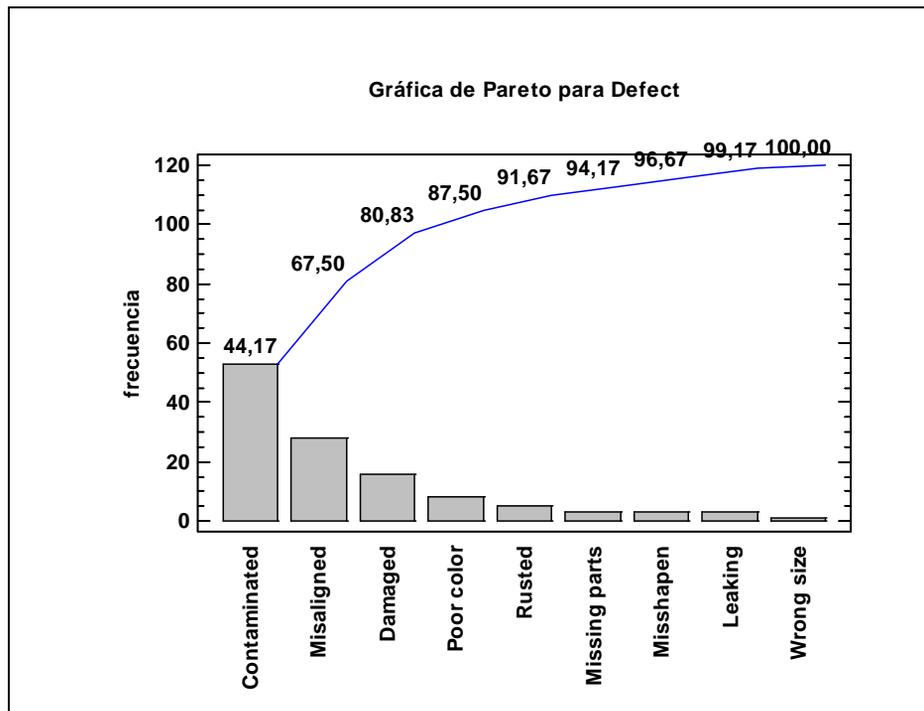


Figure 14-5. Gráfico de Pareto alargado

Las barras verticales en el gráfico de Pareto están dibujadas con altura proporcional al número de veces que ocurre cada defecto. La línea por encima de las barras es la frecuencia acumulada de izquierda a derecha. Encima de cada barra se muestra el porcentaje de defectos ocurridos en una determinada clase a clases lejanas de la izquierda.

El principio básico de Pareto establece que la mayoría de los defectos son habitualmente debidos a un número pequeño de causas posibles. En este caso, los 3 tipos de defectos más frecuentes sobrepasan el 80% de todos los defectos.

14.3 Tabulación cruzada

El archivo de datos *defects.sgd* también contiene una identificación de qué establecimientos producen cada ítem definitivo. Para resumir los datos para tipos de defectos y establecimientos:

1. Si usa el menú clásico, seleccione *Describir – Datos categóricos – Tabulación cruzada*.

2. Si usa el menú seis Sigma, seleccione *Analizar – Datos de atributos – Factores múltiples - Tabulación cruzada*.

El cuadro de diálogo de entrada de datos espera dos columnas, una definiendo las filas o las dos vías de frecuencias de la *tabla de contingencia* y la otra definiendo las columnas:

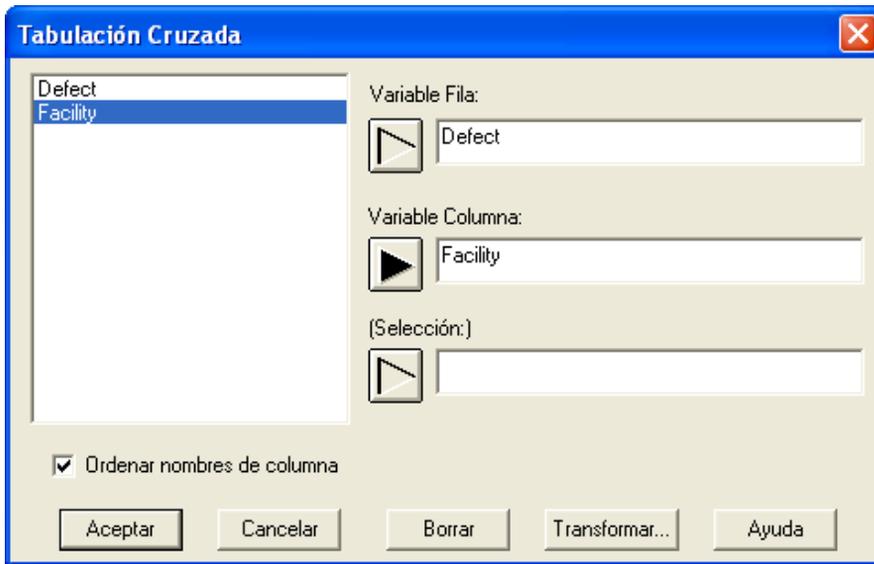


Figura 14-6. Cuadro de diálogo de entrada de datos de tabulación cruzada

Después de los cuadros de diálogo de *Opciones* y de *Tablas y Gráficos*, se genera la siguiente ventana de análisis:

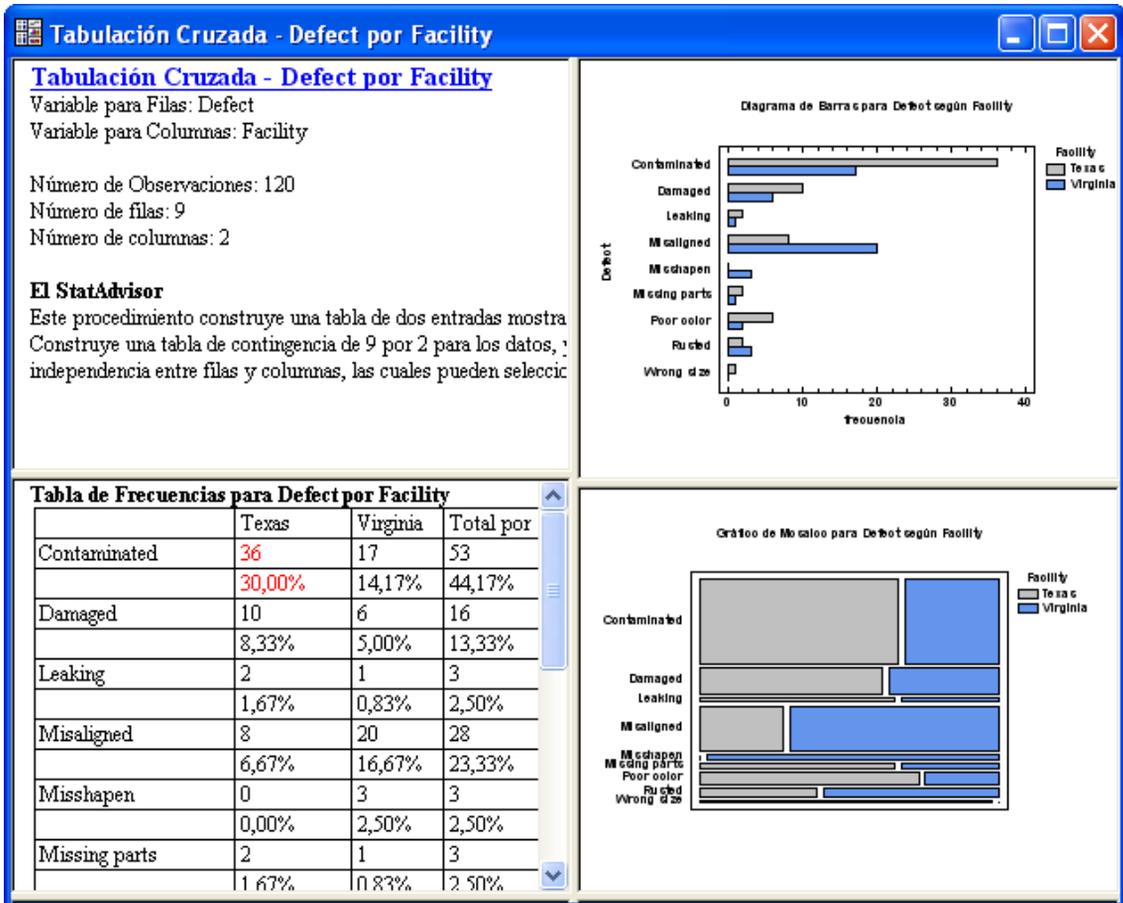


Figura 14-7. Ventana de análisis de tabulación cruzada

La tabla del panel inferior izquierdo tabula los datos para tipos de defectos y establecimientos:

Tabla de Frecuencias para Defect por Facility

	Texas	Virginia	Total por Fila
Contaminated	36	17	53
	30,00%	14,17%	44,17%
Damaged	10	6	16
	8,33%	5,00%	13,33%
Leaking	2	1	3
	1,67%	0,83%	2,50%
Misaligned	8	20	28
	6,67%	16,67%	23,33%
Misshapen	0	3	3
	0,00%	2,50%	2,50%
Missing parts	2	1	3
	1,67%	0,83%	2,50%
Poor color	6	2	8
	5,00%	1,67%	6,67%
Rusted	2	3	5
	1,67%	2,50%	4,17%
Wrong size	1	0	1
	0,83%	0,00%	0,83%
Total por Columna	67	53	120
	55,83%	44,17%	100,00%

Contenido de las celdas:

Frecuencia Observada

Porcentaje de la Tabla

Figura 14-8 Tabla de doble entrada con porcentajes de tabla

Como se muestra inicialmente, cada celda de la tabla muestra el número de filas en el archivo de datos correspondiente a una combinación particular fila-columna. También indica el porcentaje de toda la tabla representado por esta celda. Por ejemplo, hay 36 ítems contaminados producidos en la instalación de Texas, representando el 30 por ciento de todos los ítems defectuosos en la muestra.

Opciones de ventana permite seleccionar otros ítems para representar en cada celda:



Figura 14-9 Cuadro de diálogo de Opciones de ventana para tabulación cruzada

Una interesante elección para los datos actuales es mostrar *Porcentajes de filas* en vez de *Porcentajes de tablas*:

Tabla de Frecuencias para Defect por Facility			
	Texas	Virginia	Total por Fila
Contaminated	36	17	53
	67,92%	32,08%	44,17%
Damaged	10	6	16
	62,50%	37,50%	13,33%
Leaking	2	1	3
	66,67%	33,33%	2,50%
Misaligned	8	20	28
	28,57%	71,43%	23,33%
Misshapen	0	3	3
	0,00%	100,00%	2,50%
Missing parts	2	1	3
	66,67%	33,33%	2,50%
Poor color	6	2	8
	75,00%	25,00%	6,67%
Rusted	2	3	5
	40,00%	60,00%	4,17%
Wrong size	1	0	1
	100,00%	0,00%	0,83%
Total por Columna	67	53	120
	55,83%	44,17%	100,00%

Contenido de las celdas:
 Frecuencia Observada
 Porcentaje de la Fila

Figura 14-10 Tabla de doble entrada con porcentajes de fila

El porcentaje tabulado ahora indica el porcentaje que cada celda representa en su fila. Por ejemplo, el 67.92% de todos los ítems contaminados fueron producidos en Texas, mientras que el 71.43% de todos los ítems mal alineados fueron producidos en Virginia. Esto sugiere que algunos tipos de defectos pueden ocurrir más frecuentemente en un establecimiento que en otro, una hipótesis que será contrastada formalmente en la sección siguiente.

Se muestran varios gráficos que son también de ayuda. Por ejemplo, el gráfico de barras siguiente muestra los datos para defectos y establecimientos:

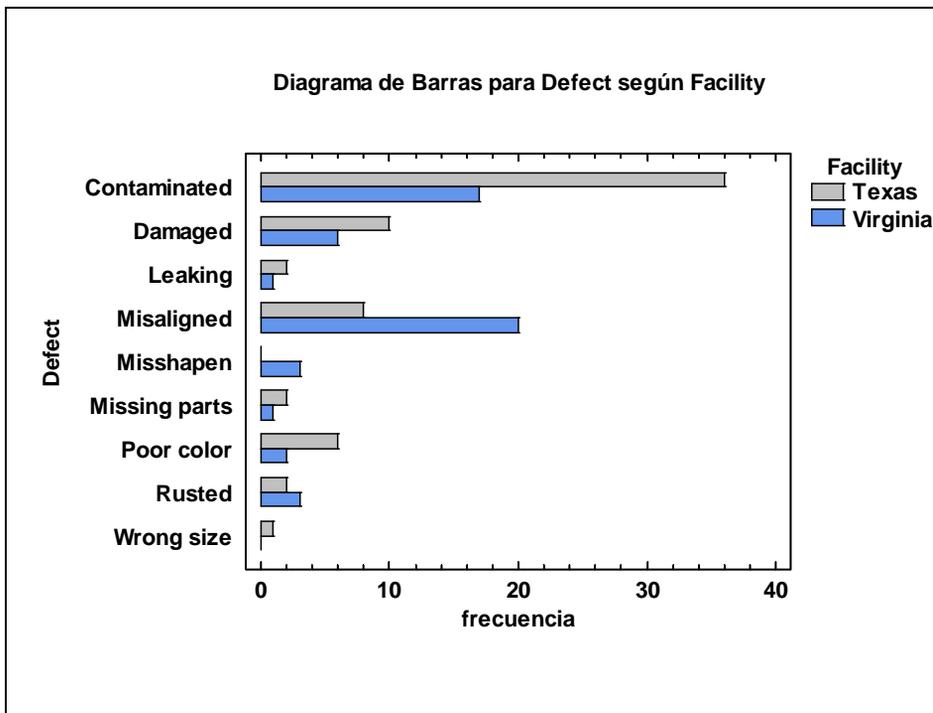


Figura 14-11. Gráfico de barras agrupado

La diferencia entre los establecimientos es evidente. Un gráfico, denominado *Gráfico de mosaico*, es también muy informativo:

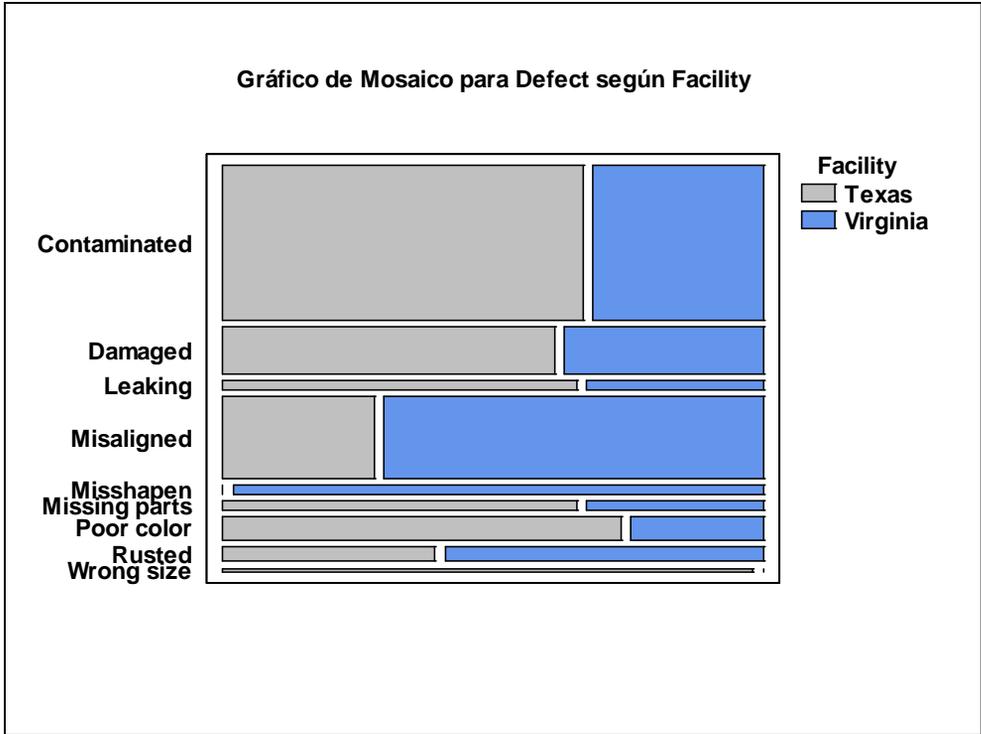


Figura 14-12. Gráfico de mosaico

En este gráfico, la altura de cada barra es proporcional al número total de defectos de cada tipo. La anchura de las barras es proporcional al porcentaje relativo de cada tipo de defecto en cada localización. Consecuentemente, el área total de cada rectángulo es proporcional a la frecuencia de la correspondiente celda en la tabla de doble entrada.

Si lo desea, las frecuencias de celda pueden mostrarse también en tres dimensiones seleccionando Gráfico de rascacielos (*Skychart*) en el cuadro de diálogo de *Tablas y Gráficos*:

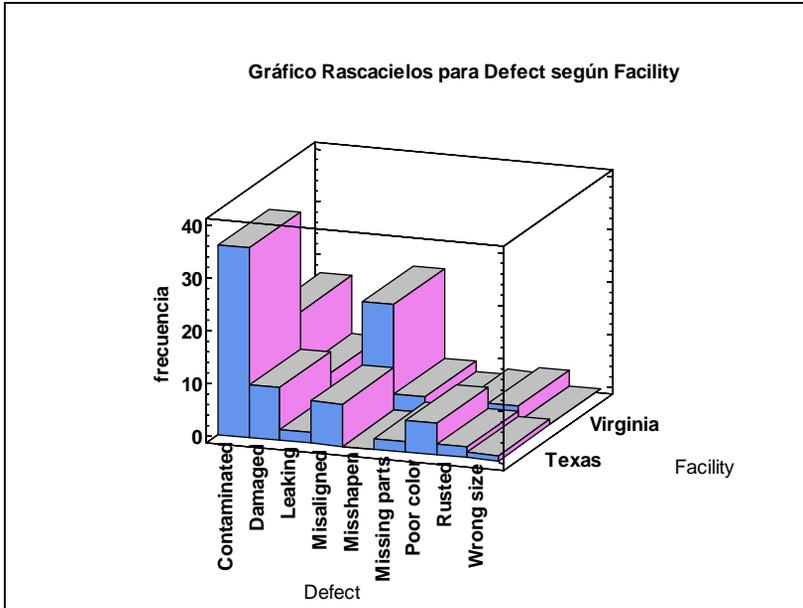


Figura 14-13. *Skychart tridimensional*

En un *Skychart*, la altura de cada barra representa la frecuencia de una celda en la tabla de contingencia.

14.4 Comparando dos o más muestras

Para determinar cuándo o no las diferencias aparentes entre los establecimientos de Texas y Virginia son estadísticamente significativas, seleccione *Test de Independencia* del cuadro de diálogo *Tablas y Gráficos*. Para una tabla de este tamaño, el procedimiento muestra los resultados del contraste de la chi-cuadrado:

Pruebas de Independencia

Prueba	Estadístico	Gl	Valor-P
Chi-Cuadrada	18,438	8	0,0182

Advertencia: algunas celdas contienen menos de 5 casos.

Figura 14-14. *Test de la Chi-cuadrado de independencia*

El test de la chi-cuadrado de independencia se utiliza para decidir entre dos hipótesis:

Hipótesis nula: clasificaciones de filas y columnas son independientes.

Hipótesis alternativa: clasificaciones de filas y columnas no son independientes.

La independencia debe implicar que el tipo de defecto definido en un ítem no tiene nada que ver con el establecimiento en el cual este ítem fue fabricado.

Para el test de la chi-cuadrado, un pequeño P -valor indica que las clasificaciones de filas y columnas no son independientes. En este caso, el P -valor es menor que 0.05, indicando al 5% de nivel de significación que la distribución de tipos de defectos es diferente en la instalación de Texas que en la instalación de Virginia. También se muestra un mensaje de peligro, ya que algunas frecuencias de celdas en la tabla de doble entrada son menores que 5. (Técnicamente, el peligro ocurre si la frecuencia esperada en alguna celda es menor que 5 asumiendo que la hipótesis nula es cierta). Con celdas con frecuencias pequeñas, el P -valor puede ser poco formal. Una solución de este problema es agrupar todos los tipos de defectos infrecuentes en una clase única y reejucutar el test. Esto se hace fácilmente en STATGRAPHICS Centurion XVI de la siguiente forma:

1. Volver a la hoja de datos y hacer clic en la cabecera de la columna *Defectos* para seleccionarla.
2. Presionar el botón derecho del ratón y seleccionar *Recodificar datos* en el menú emergente.
3. Completar el cuadro de diálogo *Recodificar datos* como se muestra a continuación para combinar los tipos de defectos menos comunes en una clase única etiquetada “Otros”:



Figura 14-15. Recodificando los tipos de defectos menos frecuentes

Las entradas del cuadro de diálogo *Recodificar datos* instruyen al programa para buscar los valores en la columna *Defectos* que caen en cada intervalo definido. Una etiqueta que cae alfabéticamente entre los límites muestra cómo una fila dada es recodificada a los valores especificados en la columna *Nuevos valores*.

Después de ejecutar la operación de recodificación, vuelva a la ventana de análisis *Tabulación cruzada*. En respuesta al cambio en el conjunto de datos, el análisis será automáticamente actualizado. La nueva clase *Otros* tiene ahora una frecuencia razonable, como muestra el *Gráfico de mosaico* revisado:

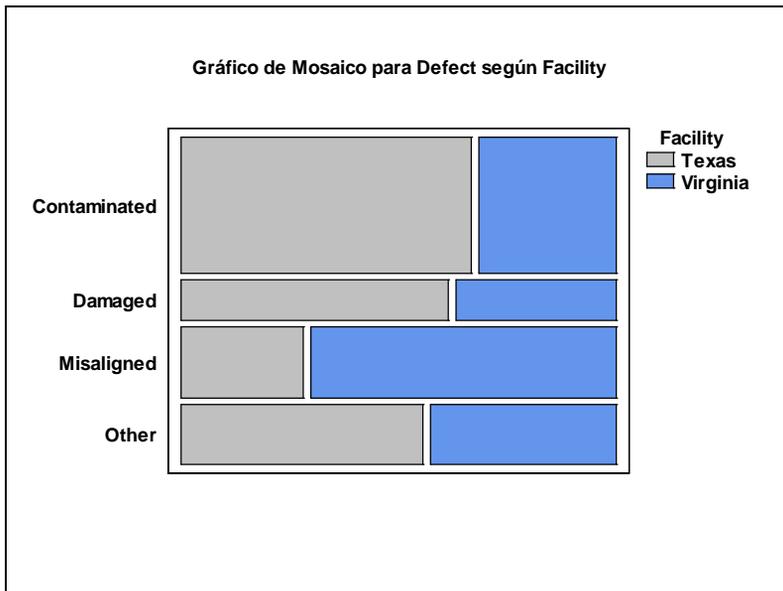


Figura 14-16. Gráfico de mosaico para datos recodificados

Después de la recodificación, el test de la chi-cuadrado muestra aún una diferencia significativa entre los establecimientos de Texas y Virginia:

Pruebas de Independencia			
Prueba	Estadístico	Gl	Valor-P
Chi-Cuadrada	11,874	3	0,0078

El StatAdvisor
 Esta tabla muestra los resultados de la prueba de hipótesis ejecutada para determinar si se rechaza, o no, la idea de que las clasificaciones de fila y columna son independientes. Puesto que el valor-P es menor que 0,05, se puede rechazar la hipótesis de que filas y columnas son independientes con un nivel de confianza del 95,0%. Por lo tanto, el valor observado de Defect para un caso en particular, está relacionado con su valor en Facility.

Figura 14-17. Test de la chi-cuadrado después de la recodificación de datos

Si de este modo aparece que el tipo de defecto está efectivamente relacionado con el establecimiento en el cual un ítem fue producido.

Se debe notar que el test anterior compara las *distribuciones* de tipos de defectos entre los dos establecimientos. No compara los números o porcentajes de ítems defectuosos de cada localización. Tal comparación requiere un contraste diferente, como el explicado en la sección siguiente.

14.5 Tablas de contingencia

Para determinar cuándo una instalación produce más ítems defectuosos que otra, necesitamos conocer la producción total de cada instalación. Supongamos que la siguiente tabla describe producciones mensuales:

<i>Establecimiento</i>	<i>Número de defectos</i>	<i>Número de ítems producidos</i>
Texas	67	6,237
Virginia	53	7,343

Sea θ_1 la proporción de ítems defectuosos producidos en Texas. Sea θ_2 la proporción de ítems defectuosos producidos en Virginia. Las proporciones estimadas vienen dadas por:

$$\hat{\theta}_1 = \frac{67}{6237} = 0.0107 \qquad \hat{\theta}_2 = \frac{53}{7343} = 0.0072$$

Basado en estos datos, se muestra que el porcentaje de ítems defectuosos producidos en Texas puede ser mayor que el porcentaje de ítems defectuosos producidos en Virginia. Para determinar cuándo esta diferencia aparente es estadísticamente significativa, se crea una hoja de datos como la siguiente:

	Attribute	Texas	Virginia	C
1	Defective	67	53	
2	Not defective	6170	7290	
3				
4				
5				
6				
7				
8				
9				
10				
11				

Figura 14-18. Hoja de datos para comparación de dos proporciones

Las filas contienen frecuencias de ítems defectuosos y no defectuosos. Entonces seleccione *Tablas de contingencia* del mismo menú que *Tabulación cruzada*. Introducir:

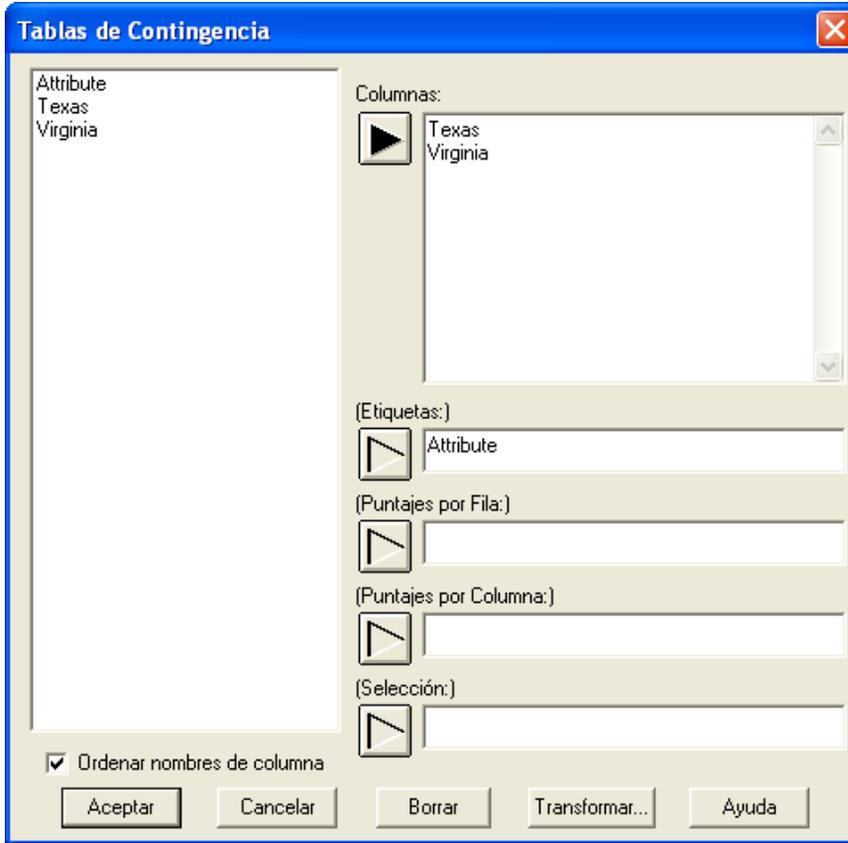


Figura 14-19. Cuadro de diálogo de entrada de datos de tablas de contingencia

El análisis mostrará un test de la chi-cuadrado de tablas 2 por 2:

Pruebas de Independencia			
<i>Prueba</i>	<i>Estadístico</i>	<i>Gl</i>	<i>Valor-P</i>
Chi-Cuadrada	4,783	1	0,0287

Figura 14-20. Test de la chi-cuadrado de tablas 2 por 2

Recordar que este contraste de la chi-cuadrado determina cuándo o no las clasificaciones de filas y columnas son independientes. En este caso, la independencia implicará que cuando un ítem fue defectuoso o no nada tiene que ver con el establecimiento en el cual fue producido.

Ya que el P -valor en la tabla anterior es menor o igual que 0.05, la hipótesis de independencia se rechaza al 5% de nivel de significación. Podemos concluir por tanto que las proporciones de defectos en las dos instalaciones son significativamente diferentes.

Tutorial #6: Análisis de la capacidad de un proceso

Determinando los defectos por millón o porcentaje más allá de los límites de especificación.

STATGRAPHICS Centurion XVI es generalmente utilizado por individuos cuyo trabajo es asegurar que los productos y servicios producidos sean de alta calidad. Una tarea común en tal trabajo es recoger datos de un proceso y compararlos con los límites de especificación establecidos. La salida de este tipo de *análisis de la capacidad* es una estimación de lo capaz que es el proceso para producir dentro de estas especificaciones. Seis Sigma es una metodología extensamente practicada para conseguir calidad, objetivos y tasas de defectos de 3.4 por millón.

Como ejemplo, consideramos un producto cuya resistencia se requiere que caiga entre 190 y 230 psi. Supongamos que se toman $n = 100$ muestras del proceso de producción cuyas resistencias medidas se muestran en la tabla siguiente:

213.5	203.3	191.3	197.1	205.7	215.6	193.7	201.7	201.5	207.1
207.0	200.4	197.2	202.4	205.2	211.0	214.5	201.5	200.9	206.8
205.8	200.3	196.1	205.9	195.1	203.9	192.9	199.0	195.5	203.1
197.4	194.8	201.0	202.5	199.0	200.7	197.6	198.5	205.3	197.1
202.8	201.6	197.4	200.9	203.3	209.4	201.4	199.5	207.8	204.9
205.5	203.0	208.1	200.2	218.2	202.0	209.3	201.2	200.4	201.0
195.7	229.5	199.9	208.1	210.3	202.0	202.6	213.6	198.0	197.8
196.7	216.0	211.6	208.7	199.4	200.8	201.1	195.3	206.8	211.3
201.5	200.0	211.8	195.6	201.9	199.0	200.3	197.8	200.8	194.8
199.5	195.5	201.0	206.0	215.3	202.6	199.9	200.6	197.6	207.4

Este capítulo describe cómo realizar un análisis de la capacidad para este tipo de variables de datos.

15.1 Graficando los datos

El primer paso en el examen de un nuevo conjunto de datos es su representación gráfica. Para un conjunto de datos tal como el anterior, el *análisis de una variable* descrito en el capítulo 10 aporta varias herramientas de uso. Para analizar estos datos:

1. Abra el archive de nombre *items.sgd*.
2. Ejecute el procedimiento *Análisis de una variable* utilizando la columna de nombre *resistencia (Strength)*.

La ventana de análisis inicial se muestra a continuación:

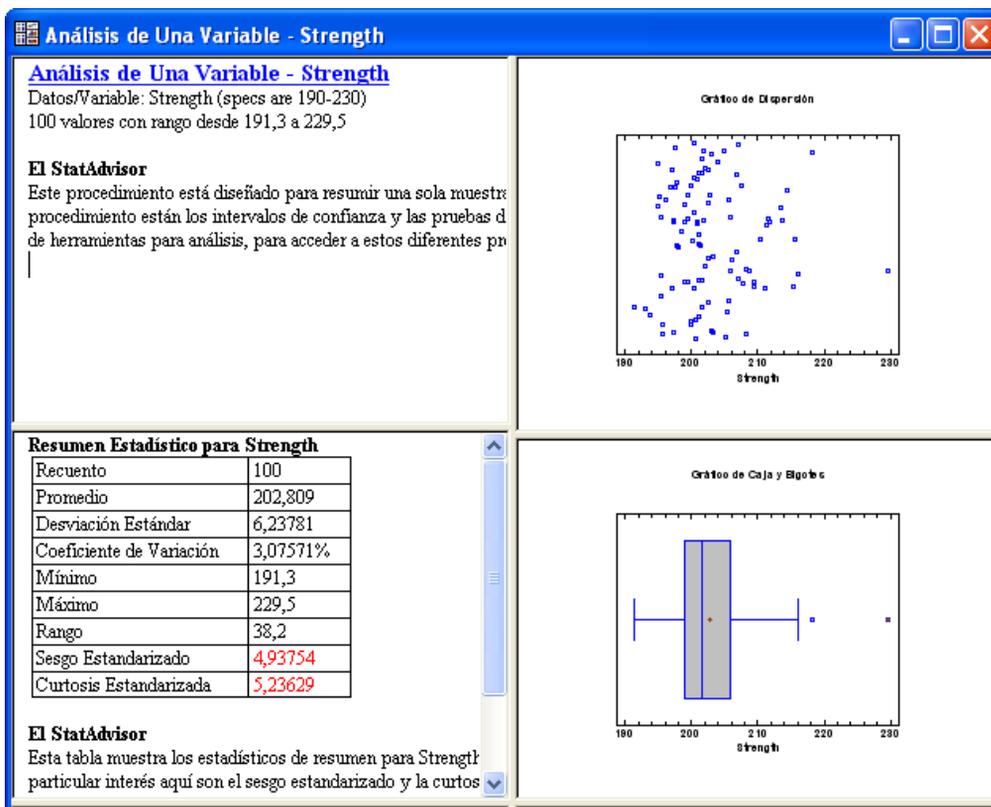


Figura 15-1. Ventana de Análisis de una variable

Varios factores interesantes son evidentes inmediatamente:

1. Los datos están todos dentro de los límites de especificación, registrados entre 191.3 y 229.5.
2. El gráfico de caja y bigotes muestra un punto extremo lejano (un cuadrado pequeño con un signo más rojo dentro de él). Tales puntos son considerados habitualmente como atípicos, si el resto de los datos provienen de una distribución normal. En este caso, sin embargo, incluso descontando el aparente atípico, la forma de la caja no es muy simétrica. El bigote superior es más largo que el bigote inferior y la caja es más amplia por encima de la mediana (la línea vertical dentro de la caja) que por debajo.
3. Si expande el panel *Estadísticos resumen*, verá que la desviación típica estandarizada es igual a 4.94. Si los datos provienen de una distribución normal, la asimetría y curtosis estandarizadas deben caer entre -2 y +2. Eliminando el mayor valor sólo se reduce la asimetría en 2.81.

También se puede presentar un histograma de frecuencias presionando en el botón *Tablas y Gráficos* en la barra de herramientas de análisis y seleccionando *Histograma de frecuencias* en la caja de diálogo *gráficos*:

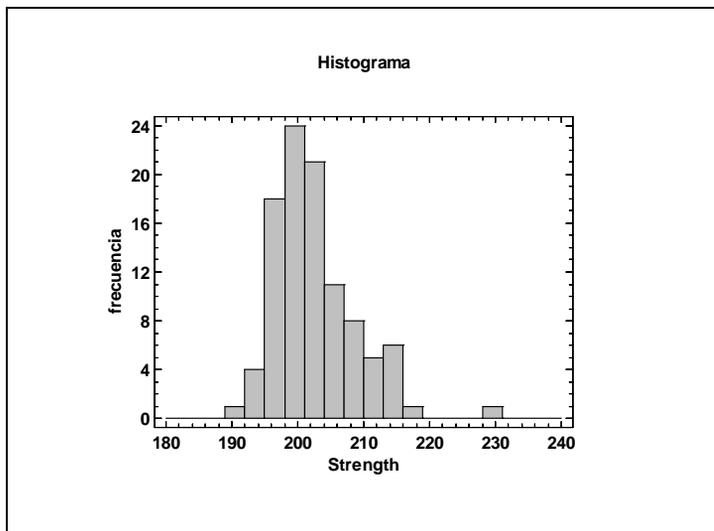


Figura 15-2. Histograma de frecuencias

Los datos presentan claramente asimetría positiva, extendiéndose más a la derecha de la caja que a la izquierda.

Datos no normales como los que acabamos de mostrar son comunes. Una típica aproximación para trabajar con tales datos, desafortunadamente, es simplificar ignorando la no linealidad y calculando índices tales como C_{pk} utilizando formulas para datos con distribución normal. Como se verá en este tutorial, ignorando la no normalidad pueden obtenerse resultados incorrectos, a menudo sobreestimando o infraestimando significativamente el porcentaje de productos que queda fuera de los límites de especificación.

15.2 Procedimiento Análisis de la capacidad

STATGRAPHICS Centurion XVI contiene procedimientos para ejecutar análisis de la capacidad en datos recogidos uno a uno (datos individuales) o en subgrupos (tales como 5 observaciones cada hora). Asumiendo que la muestra es de datos individuales, un análisis de la capacidad del proceso puede realizarse como sigue:

1. Si se usa menú clásico, seleccione *CEP – análisis de la capacidad – Variables – Individuales*.
2. Si se usa el menú Seis Sigma, seleccione *Analizar – Datos de variables – Análisis de la capacidad – Individuales*.

El cuadro de diálogo de entrada de datos requiere el nombre de la columna simple que contiene los datos. Los datos de la muestra pueden definirse en una columna llamada *Resistencia (Strength)* en el archivo de nombre *items.sgd*:

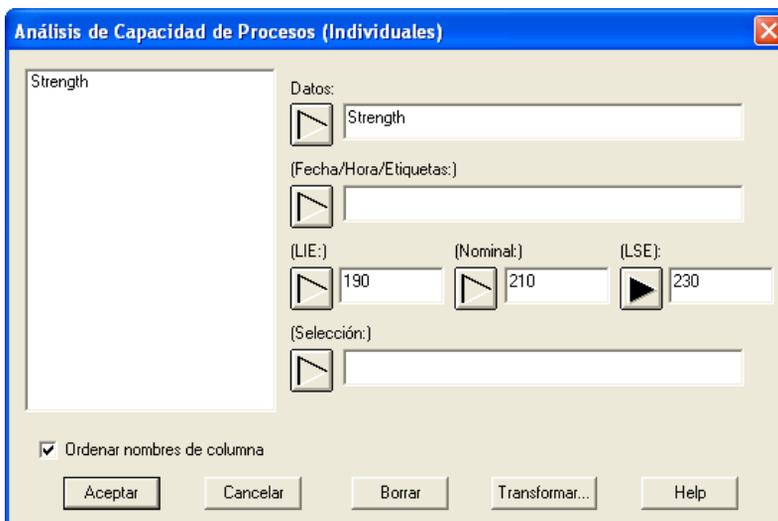


Figura 15-3. Cuadro de diálogo de Análisis de la capacidad de un proceso

Los límites superiores e inferiores de especificación han sido también indicados, así como el valor nominal o valor objetivo.

Cuando se presiona *Aceptar* y el menú *Opciones*, aparece el cuadro de diálogo de *Tablas y Gráficos*. Utilice los valores por defecto en ambos menús para el beneficio de este tutorial.

La ventana inicial de análisis muestra un resumen de los datos, una tabla de índices de capacidad y un gráfico de capacidad:

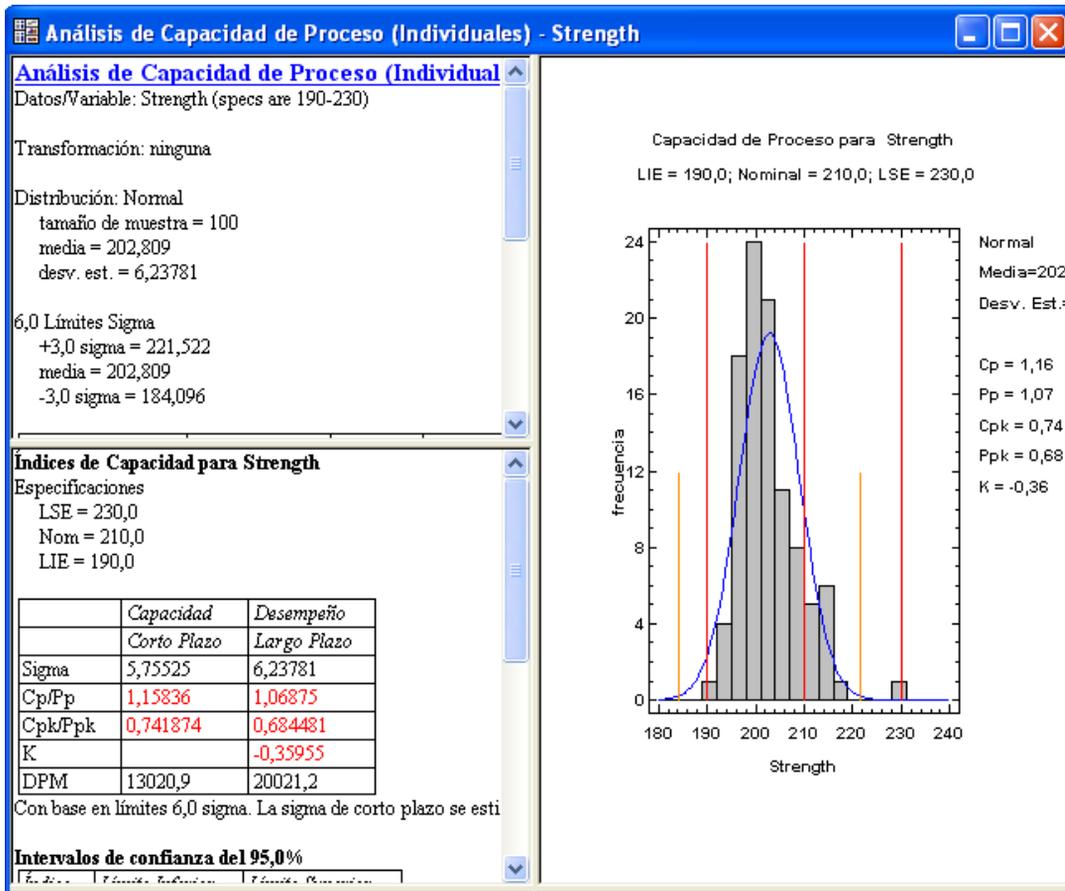


Figura 15-4. Ventana de Análisis de la capacidad de un proceso

Cuando se ejecuta por primera vez un análisis de la capacidad, una distribución normal ajusta los datos. El *Gráfico de capacidad* muestra un histograma de los datos, junto con el mejor ajuste a la distribución normal:

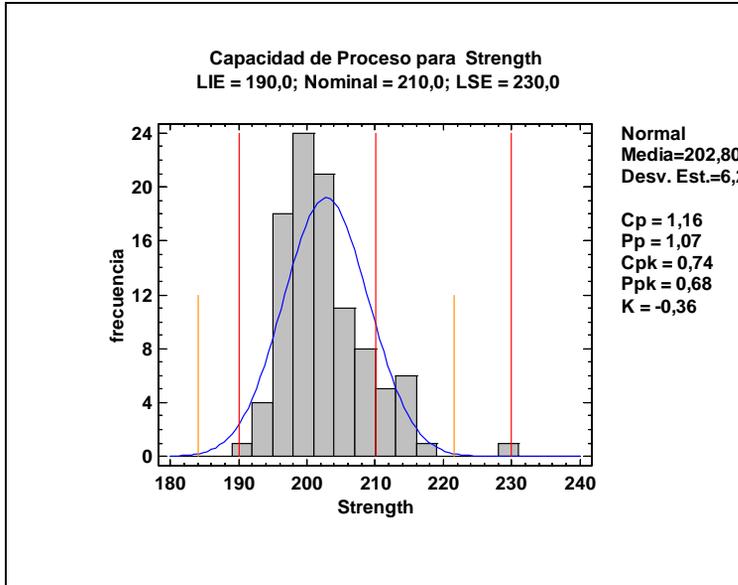


Figura 15-5. Gráfico de capacidad con distribución normal

La altura de las líneas verticales en el gráfico muestra la localización de los límites de especificación y los valores nominales. Las líneas verticales más bajas están localizadas en la media de la muestra más y menos 3 desviaciones típicas. Particularmente destacables en el gráfico anterior son:

1. La distribución normal ajustada no coincide muy bien con nuestros datos. Aunque la forma de campana de la curva normal tenga la misma media y desviación típica que los datos, la asimetría de los datos causa que la curva coincida de forma pobre con las barras del histograma.
2. La media de la muestra está localizada en 202,8, lo cual es considerablemente menor que el valor nominal de 210.
3. Aunque ninguna de las observaciones sea menor o igual que el límite inferior de especificación, una cantidad de la cola inferior de la distribución normal está por debajo de tales límites.
4. Las líneas de más y menos 3 sigma están ajustadas lo suficiente con las especificaciones. Sin embargo están cambiadas hacia la izquierda.

El *Resumen de análisis* en el panel superior izquierdo cuantifica lo siguiente:

Análisis de la capacidad de un proceso (Individuales) - Resistencia

Variable de datos: Resistencia (las especificaciones son 190-230)

Transformación: ninguna

Distribución: Normal

tamaño de la muestra = 100

media = 202.809

desviación típica = 6.23781

6.0 límites Seis Sigma

+3.0 sigma = 221.522

media = 202.809

-3.0 sigma = 184.096

	<i>Observados</i>		<i>Estimados</i>	<i>Defectos</i>
<i>Especificaciones</i>	<i>Fuera de las especific.</i>	<i>Z-Score</i>	<i>Fuera de las especific.</i>	<i>Por Millón</i>
USL = 230.0	0.000000%	4.36	0.000654%	6.54
Nominal = 210.0		1.15		
LSL = 190.0	0.000000%	-2.05	2.001465%	20014.65
Total	0.000000%		2.002119%	20021.19

Figura 15-6. Resumen de análisis de la capacidad

De primordial interés es la tabla inferior, que estima el tanto por ciento del producto fuera de las especificaciones. Basado en el ajuste a la distribución normal, el porcentaje estimado de producto fuera de los límites de especificación está cercano al 2%, siendo igual a 20,021 defectos por millón (DPM).

15.3 Tratando con datos no normales

El DPM estimado calculado anteriormente depende de la asunción de que los datos provienen de una distribución normal. Un contraste formal de esta hipótesis se puede realizar eligiendo *Test para Normalidad* en el cuadro de diálogo *Tablas y Gráficos*:

Tests de Normalidad para Resistencia		
<i>Test</i>	<i>Estadístico</i>	<i>P-Valor</i>
Shapiro-Wilks W	0.931784	0.0000321356

Figura 15-7. Test de normalidad

Dependiendo de las preferencias de su sistema, pueden mostrarse uno o más test de normalidad. Cada uno de los test disponibles está basado en los siguientes conjuntos de hipótesis:

Hipótesis nula: los datos provienen de una distribución normal.

Hipótesis alternativa: los datos no provienen de una distribución normal.

Un P -valor por debajo de 0.05 significa el rechazo de la hipótesis de normalidad al 5% de nivel de significación.

En la tabla anterior, el test de Shapiro-Wilks rechaza sólidamente la hipótesis de que los datos provengan de una distribución normal. Por lo tanto, el valor estimado de DPM o los valores basados en índices de capacidad con la asunción de normalidad no son válidos.

Cuando los datos son no normales, pueden tenerse en cuenta una de las dos siguientes aproximaciones:

1. Seleccionar una distribución distinta de la normal a la cual se ajusten los datos.
2. Transformar los datos con una transformación métrica para que sigan una normal.

Para ayudar en la selección de diferentes distribuciones, STATGRAPHICS Centurion XVI aporta una opción llamada *Comparación de distribuciones alternativas* en el cuadro de diálogo *Tablas y Gráficos*. Esta opción ajusta a otras distribuciones y las lista en orden de su bondad de ajuste. Utilizando la selección de distribuciones por defecto tenemos la siguiente salida:

Comparación de distribuciones alternativas			
<i>Distribución</i>	<i>Est. Parámetros</i>	<i>KS D</i>	<i>A²</i>
Mayor valor extremo	2	0.0675422	0.372613
Loglogística	2	0.0913779	1.15081
Laplace	2	0.0920985	1.68399
Logística	2	0.0941708	1.27599
Lognormal	2	0.13213	1.66564
Gamma	2	0.134136	
Normal	2	0.138628	1.90094
Weibull	2	0.177886	5.67166
Menor valor extremo	2	0.189989	6.28546
Exponencial	1	0.61064	43.3327
Pareto	1	0.628084	45.3859

Figura 15-8. Distribuciones ajustadas en orden de bondad de ajuste

Las distribuciones han sido listadas de acuerdo con el valor del estadístico de Kolmogorov-Smirnov de la bondad de ajuste, que mide la máxima distancia entre la distribución acumulada de los datos y la de la distribución ajustada. En este caso, la mejor distribución de ajuste es la del *mayor valor extremo*.

Es posible cambiar a la distribución del mayor valor extremo mediante *Opciones de análisis*:

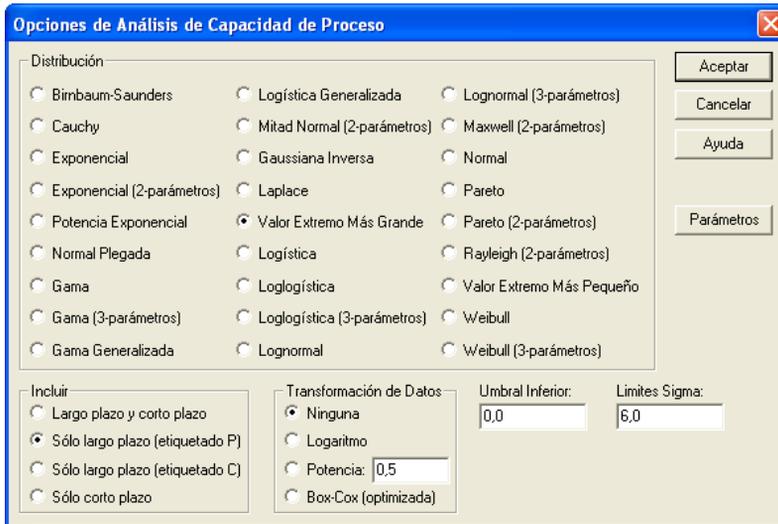


Figura 15-9. Cuadro de diálogo de opciones de la capacidad de un proceso

El resultado del nuevo ajuste se muestra a continuación:

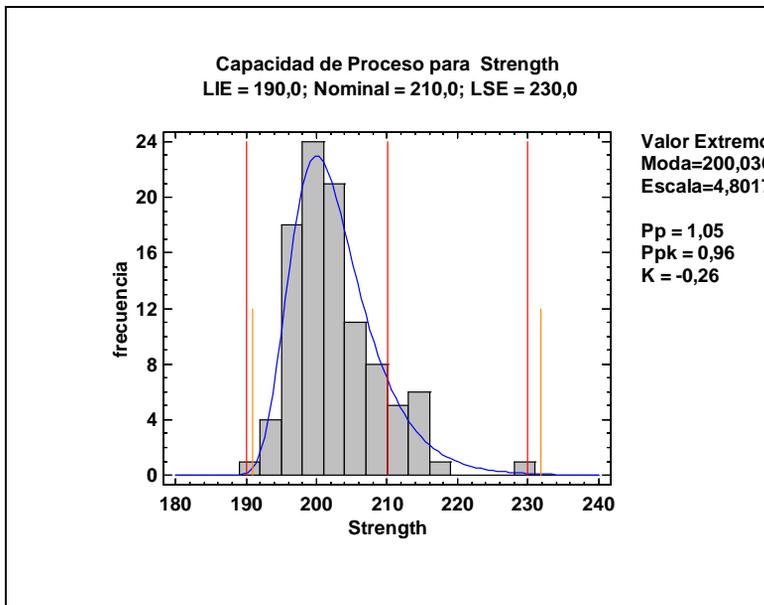


Figura 15-10. Ajuste a la distribución del mayor valor extremo

Hay que notar que la distribución es asimétrica la derecha, coincidiendo los datos observados mejor que para la distribución normal. Las líneas cortas verticales han sido posicionadas equivalentes a los límites 3 sigma, límites dentro de los cuales se localiza el 99.73% de la distribución ajustada (media más menos 3 sigma para una distribución normal). Hay que notar que estos límites no están simétricamente espaciados alrededor del máximo de la distribución, debido a su asimetría positiva.

El *Resumen de análisis* muestra una diferencia dramática en el porcentaje estimado del producto que va estar fuera de las especificaciones, comparado con el ajuste de la distribución normal:

<u>Análisis de la capacidad de un proceso (Individuales) – Resistencia</u>				
Variable de datos: Resistencia (las especificaciones son 190-230)				
Transformación: ninguna				
Distribución: Mayor valor extremo				
tamaño de la muestra = 100				
moda = 200.036				
escala = 4.80179				
(media = 202.808)				
(sigma = 6.15853)				
Equivalente a 6.0 límites sigma				
99.865 percentil = 231.761				
mediana = 201.796				
0.134996 percentil = 190.969				
	<i>Observado</i>		<i>Estimado</i>	<i>Defectos</i>
<i>Especificaciones</i>	<i>Fuera de Spec.</i>	<i>Z-Score</i>	<i>Fuera de Spec.</i>	<i>Por Millón</i>
USL = 230.0	0.000000%	2.89	0.194758%	1947.58
Nominal = 210.0		1.19		
LSL = 190.0	0.000000%	-3.42	0.030805%	308.05
Total	0.000000%		0.225563%	2255.63

Figura 15-11. Resumen de análisis después de ajustar los datos a la distribución del mayor valor extremo

El porcentaje estimado fuera de las especificaciones es ahora sólo del 0.23 por ciento, o 2,256 DPM, un décimo del resultado de distribución normal. En este caso, la asunción incorrecta de distribución normal hace que el proceso parezca peor que lo que realmente es.

NOTA: Dependiendo de los límites de especificación y de la distribución verdadera, la asunción incorrecta de normalidad puede hacer que el proceso parezca significativamente peor o significativamente mejor que cuando se usa la distribución adecuada.

Una alternativa a la selección de una distribución diferente es la transformación de los datos. El cuadro de diálogo *Opciones de análisis* permite seleccionar varias *Transformaciones de datos*:

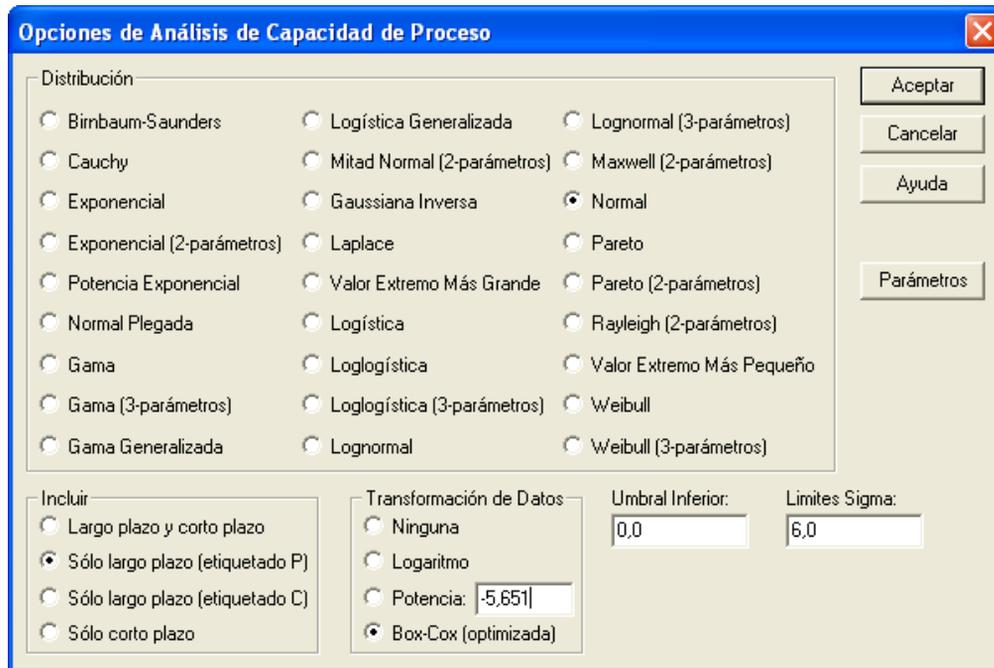


Figura 15-12. Cuadro de diálogo *Opciones de análisis* para selección de transformación

La selección incluye el logaritmo natural, elevando cada valor a la potencia especificada, o seleccionando una transformación según los métodos de Box y Cox. La última aproximación considera una variedad de transformaciones de la forma Y^p utilizando los métodos de Box y Cox y seleccionando un valor óptimo para p .

Si se selecciona una transformación, se ajusta una distribución normal a los datos transformados. El gráfico siguiente muestra los resultados de tomar la transformación de Box-Cox:

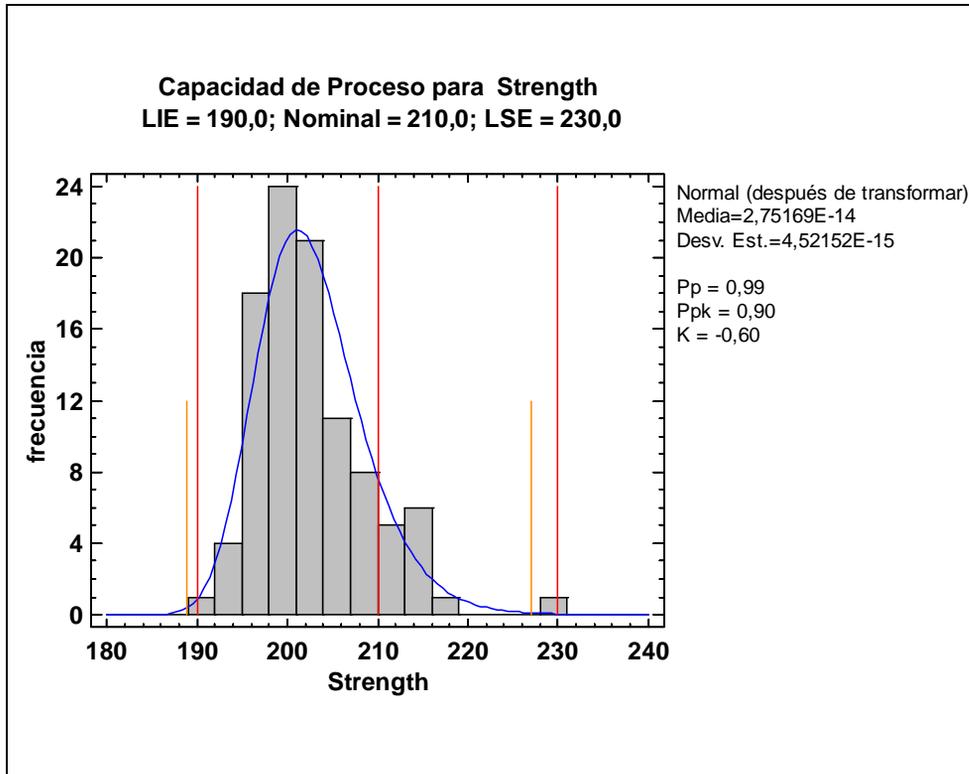


Figure 15-13. Gráfico de capacidad después de una transformación Box-Cox

Para el gráfico, se ha aplicado una transformación inversa mostrando el ajuste en la métrica original. La transformación ha tenido un efecto similar en la forma de la distribución, aunque no es tan fuerte como asumiendo la distribución del mayor valor extremo. El DPM estimado es 4,353, que es cerca de dos veces mayor que cuando se utilizaba la distribución del mayor valor extremo, sin embargo aún mucho mayor que cuando se consideraba distribución normal.

NOTA: la media y la desviación típica mostradas en el gráfico correspondiente a los datos transformados no son habitualmente muy utilizadas. STATGRAPHICS Centurion XVI convierte automáticamente cada valor a las unidades originales.

Para comparar dos aproximaciones, se selecciona *Gráfico de probabilidad* en el cuadro de diálogo de *Tablas y Gráficos* para cada aproximación y se pegan lado a lado en StatGallery:

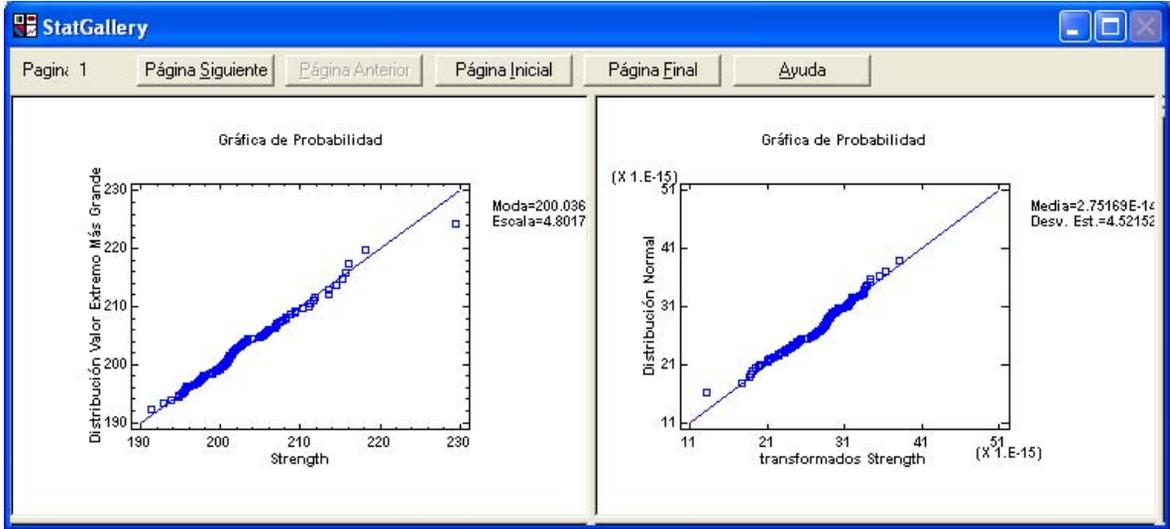


Figure 15-14. Gráficos de probabilidad en StatGallery

Si se asume que la distribución es correcta, los puntos deben situarse a lo largo de la línea diagonal cuando se muestran en el gráfico. Ambos métodos parecen tratar bien la no normalidad, presentando dificultades de elección entre ellos. Sea cual sea el método que se use, es importante establecer un protocolo para cómo manejar una variable en particular (tal como *Resistencia*) y aplicar el mismo protocolo cada vez que se analizan los datos. Es un error realizar todos los tipos de análisis exploratorio explicados en este capítulo cada vez que se recoja un conjunto similar de datos. En su lugar, este tipo de análisis debe ser hecho una vez para determinar cómo manejar la variable seleccionada, y entonces la aproximación seleccionada debe ser aplicada a la variable cuando sea analizada.

15.4 Índices de capacidad

La esencia de un análisis de la capacidad radica en la estimación del porcentaje de la producción que cae fuera de los límites de especificación (o equivalentemente DPM, los defectos por millón). Para resumir la capacidad del proceso, los investigadores han habilitado varios índices de capacidad. El índice más generalmente utilizado es C_{pk} , definido como:

$$C_{pk} = \min\left(\frac{\hat{\mu} - LSL}{3\hat{\sigma}}, \frac{USL - \hat{\mu}}{3\hat{\sigma}}\right)$$

Simplemente, C_{pk} es la distancia de la media del proceso estimado al límite de especificación más cercano, dividida por 3 veces el valor de sigma estimado para el proceso.

El procedimiento *Análisis de la capacidad de un proceso* en STATGRAPHICS Centurion XVI muestra los índices de capacidad en el *Gráfico de capacidad* y también en la tabla de *Índices de capacidad*. Si se asume distribución normal, se calculan dos tipos de índices (*Short-Term o de capacidad* y *Long-Term o de rendimiento*):

índices de capacidad para resistencia		
Especificaciones		
USL = 230.0		
Nom = 210.0		
LSL = 190.0		
	<i>Short-Term</i>	<i>Long-Term</i>
	<i>Capacidad</i>	<i>Rendimiento</i>
Sigma	5.75525	6.23781
Cp/Pp	1.15836	1.06875
Cpk/Ppk	0.741874	0.684481
Cpk/Ppk (upper)	1.57485	1.45302
Cpk/Ppk (lower)	0.741874	0.684481
K		-0.35955
DPM	13020.9	20021.1
Nivel de calidad Sigma	3.73	3.55
95.0% Intervalos de confianza		
<i>Índice</i>	<i>Límite inferior</i>	<i>límite superior</i>
Cp	0.997149	1.31931
Pp	0.920008	1.21725
Cpk	0.619618	0.864129
Ppk	0.568904	0.800059
Cpm	0.61885	0.777645

Basado en límites 6 sigma. Short-term sigma se estima de los rangos de medias móviles. El nivel de calidad Sigma incluye una distancia de 1.5 sigma de la media.

Figura 15-15. Tabla de índices de capacidad

Los índices short-term, se calculan utilizando una estimación del sigma obtenido de todas las observaciones en un momento dado y describe lo que el proceso es capaz de hacer si la media se mantiene constante. Los índices long-term, que se calculan utilizando un estimador del sigma obtenido de la variabilidad total entre las observaciones en todo el periodo de muestreo y describe como se ejecuta el proceso actualmente. Un proceso fuera de control en el cual la media ha cambiado significativamente en el transcurso de la recolección de datos puede mostrar peor representación. Por defecto, STATGRAPHICS Centurion XVI etiqueta los índices de capacidad utilizando la letra “C” y los índices de rendimiento utilizando la letra “P”.

La pestaña *Capacidad* del cuadro de diálogo *Preferencias*, accesible bajo *Editar* en el menú principal de STATGRAPHICS Centurion XVI, especifica los índices que van a ser calculados por defecto, así como otras opciones importantes:

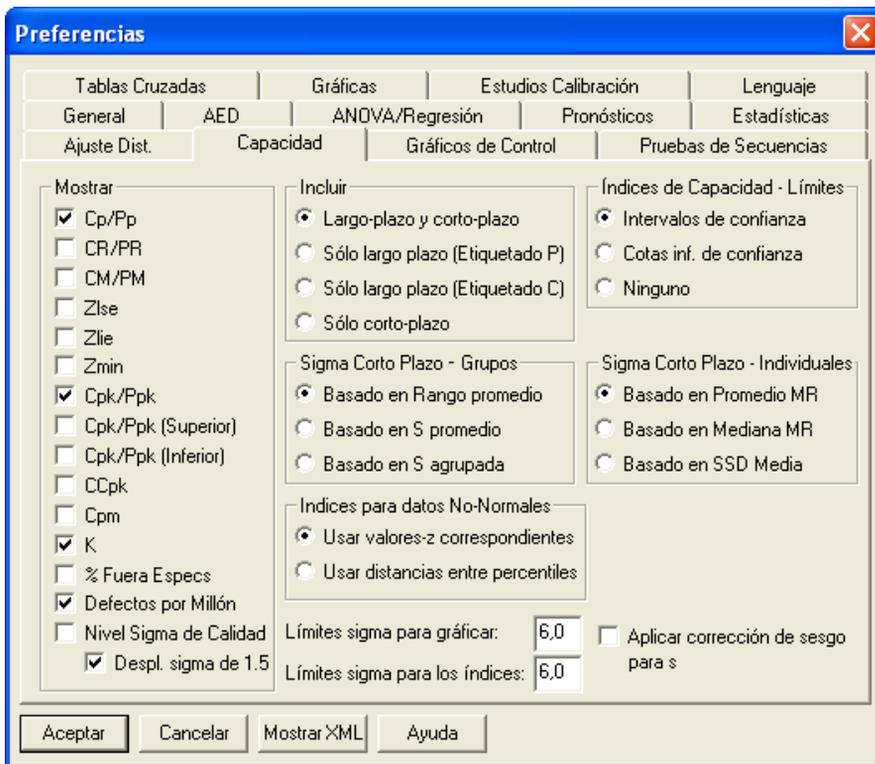


Figura 15-16. Índices de capacidad en Preferencias del sistema

El lado izquierdo del cuadro de diálogo lista los índices que pueden ser calculados. En suma a C_{pk} , los índices posibles incluyen:

1. C_p – índice de capacidad de dos lados calculado como

$$C_p = \frac{USL - LSL}{6\hat{\sigma}}$$

El índice mide la distancia entre los límites de especificación relativos a la distancia cubierta por los límites seis sigma. C_p es siempre mayor o igual que C_{pk} . Una diferencia sustancial entre los dos índices indica que el proceso no está bien centrado.

2. K – es una medida del alejamiento del centro del proceso. K se calcula mediante

$$K = \frac{\hat{\mu} - NOM}{(USL - LSL)/2}$$

donde NOM es el valor nominal objetivo. Un valor de K cercano a 0 es indicativo de un proceso bien centrado.

3. *Nivel de calidad Sigma* – un índice utilizado en Seis Sigma para indicar el nivel de calidad asociado con un proceso. Un *nivel de calidad Sigma* de 6 es habitualmente asociado con una tasa de defectos de 3.4 por millón.

El cuadro de diálogo *Preferencias* también afecta a los índices que se muestran en el *Gráfico de capacidad* y como son etiquetados. Una discusión detallada de varios índices puede ser encontrada en el documento PDF titulado *Análisis de la capacidad – Variables de datos*.

Adicionalmente a los índices de capacidad, la tabla de la *figura 15.15* incluye intervalos de confianza que muestran el margen de error en la estimación de estos índices. Por ejemplo, la tabla anterior muestra un C_{pk} de 0.74. El intervalo de confianza del 95% se extiende desde 0.62 a 0.86. Esto indica que el verdadero C_{pk} en el proceso del que los datos fueron muestreados puede estar en cualquier lugar entre 0.62 y 0.86.

Cuando los datos no siguen una distribución normal, los índices de capacidad necesitan ser modificados. La opción por defecto en el cuadro de diálogo *Preferencias* calcula índices no normales computando primeramente las puntuaciones Z equivalentes para la distribución no normal ajustada. Para una distribución normal, la puntuación Z mide el número de desviaciones típicas de la media del proceso como un límite de especificación y está directamente relacionado con la probabilidad de que una observación caiga fuera de los límites. Para una distribución no normal, una puntuación Z equivalente se calcula primero determinando la probabilidad de exceder el límite y entonces encuentra la puntuación Z que iguala la probabilidad. Después de

calcular la puntuación Z equivalente para los límites superior e inferior de especificación, C_{pk} puede ser calculado como

$$C_{pk} = \min(Z_{lsl}, Z_{usl}) / 3$$

NOTA: A través del cuadro de diálogo *Preferencias* se nos ofrece la opción de calcular índices de capacidad de percentiles en lugar de puntuaciones Z , eliminando la relación habitual entre los índices de capacidad y DPM.

15.5 Calculadora Seis Sigma

Un índice, C_{pk} resume habitualmente la capacidad de un proceso. Una vez calculado puede ser relacionado con DPM. El menú *Herramientas* de STATGRAPHICS Centurion XVI *contiene una calculadora Seis Sigma* que hará conversiones adecuadas teniendo en cuenta:

1. Los datos provienen de una distribución normal.
2. Puntuaciones Z equivalentes se utilizan para calcular los índices.

El cuadro de entrada de datos de la *Calculadora Seis Sigma* se muestra a continuación:

Índices Seis Sigma

Entrada

Valor-Z: 4,5

DPM: 10,0

Defectos (%): 0,01

Rendimiento (%): 99,99

Cpk: 1,33

Nivel Sigma: 6,0

Desplazamiento Sigma: 1,5

Especificaciones

Bilateral

Límite inferior sólo

Límite superior sólo

Aceptar

Cancelar

Ayuda

Figura 15-17. Calculadora Seis Sigma

Para utilizar el procedimiento:

1. Seleccione los botones de entrada e introduzca un valor para el estadístico correspondiente.
2. Si sólo quiere calcular valores en especificaciones cercanas a los límites, seleccione cualquiera de los botones *límite inferior sólo* o *límite superior sólo*.
3. Indique el valor que quiere asumir para el cambio long-term en la media del proceso. En Seis Sigma, se asume a veces que la media del proceso oscilará alrededor de su valor long-term en 1.5 sigma.
4. Presione el botón *Calcular* para mostrar los valores asociados de los otros estadísticos.

Índice	Valor
Valor-Z	3,99
DPM	33,0518
Defectos	0,00330518
rendimiento	99,9967
Cpk	1,33
SQL	5,49

Figura 15-18. Valores equivalentes de índices de calidad

Asumiendo que la media del proceso no cambia, un C_{pk} de 1.33 equivale a 33 defectos por millón fuera de las especificaciones.

Tutorial #7: Diseño de experimentos (DOE)

Diseñando experimentos para ayudar a la mejora de los procesos.

No todos los datos se crean igual. A menudo, un pequeño y planificado estudio aporta más información que uno largo mal diseñado. El final de este tutorial examina algunas de las capacidades de STATGRAPHICS Centurion XVI para crear y analizar diseños de experimentos.

Considerar el caso de un ingeniero que quiere determinar qué variables de un proceso tienen mayor impacto en el producto final. Intenta investigar el impacto cambiando 5 factores: temperatura de entrada, tasa de flujo, concentración, tasa de agitación y porcentaje de catalizador. En la práctica, este problema puede ser aproximado por varios caminos, incluyendo:

1. *Entrenamiento y error:* se seleccionan arbitrariamente combinaciones diferentes de los factores y cada vez se ejecuta el experimento. Tal aproximación raramente da información útil.
2. *Un factor en el momento de la experimentación:* participando todos los factores con un factor constante para determinar el efecto de este factor. Esta aproximación es extremadamente ineficiente y puede ser errónea si uno de los factores presenta interacción.
3. *Utilizando un diseño estadístico de experimentos:* configurando una secuencia de experimentos a ejecutar que produzcan la mayor información acerca de los factores y sus interacciones con los menos experimentos posibles.

Este tutorial describirá cómo un diseño de experimentos puede ser construido utilizando la tercera aproximación y cómo el resultado puede ser analizado.

16.1 Creando un diseño

STATGRAPHICS Centurion XVI contiene un Asistente de diseño de experimentos que guía a los usuarios a través de la construcción y análisis de un diseño de experimentos. Para acceder al asistente DOE:

1. Si usa el menú clásico, seleccione *DOE – Asistente de diseño de experimentos*.
2. Si usa el menú Seis Sigma, seleccione *Mejorar – Asistente de diseño de experimentos*.

Se creará una nueva ventana conteniendo una barra de herramientas que nos guiará a través de una secuencia de 12 pasos:

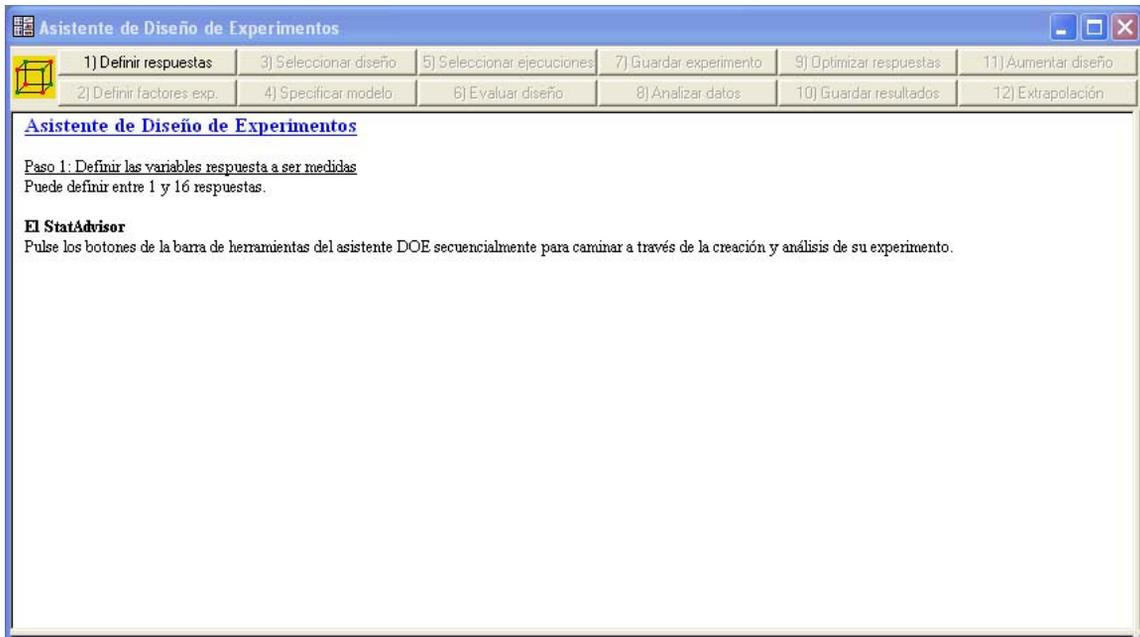


Figura 16-1. Ventana principal de Asistente de diseño de experimentos con una barra de herramientas de 12 pasos

Los primeros 7 pasos de la secuencia construyen el diseño experimental y son ejecutados antes de que el experimento sea realizado. Los últimos 5 pasos son ejecutados después de que el experimento ha sido completado y se refieren a análisis de los resultados.

Paso 1: Definir respuestas

El primer paso en la creación de un diseño de experimentos es especificar las variables respuesta que serán medidas durante la ejecución del experimento. Presionando el botón *Paso 1* muestra el siguiente cuadro de diálogo:

Asistente de diseño de experimentos - Definir Respuestas

Archivo de diseño: <sin título>

Comentario: Tutorial #7

Número de respuestas: 2

Respuesta	Nombre	Unidades	Analizar	Objetivo	Destino	Impacto (1-5)	Sensibilidad	Mínimo	Máximo
1	Rendimiento	gramos	Media	Maximizar	0,5	3,0	Medio	80,0	90,0
2	Resistencia	psi	Media	Objetivo alcar	250	5,0	Medio	200	300
3	Var_3		Media	Maximizar	0,5	3,0	Medio		
4	Var_4		Media	Maximizar	0,5	3,0	Medio		
5	Var_5		Media	Maximizar	0,5	3,0	Medio		
6	Var_6		Media	Maximizar	0,5	3,0	Medio		
7	Var_7		Media	Maximizar	0,5	3,0	Medio		
8	Var_8		Media	Maximizar	0,5		Medio		
9	Var_9		Media	Maximizar	0,5	3,0	Medio		
10	Var_10		Media	Maximizar	0,5	3,0	Medio		
11	Var_11		Media	Maximizar	0,5	3,0	Medio		
12	Var_12		Media	Maximizar	0,5	3,0	Medio		
13	Var_13		Media	Maximizar	0,5	3,0	Medio		
14	Var_14		Media	Maximizar	0,5	3,0	Medio		
15	Var_15		Media	Maximizar	0,5	3,0	Medio		
16	Var_16		Media	Maximizar	0,5	3,0	Medio		

Aceptar Cancelar Ayuda

Figura 16.2. Definición de variables respuesta

En este ejemplo, hay dos variables respuesta: *rendimiento* en gramos y *resistencia* en libras por pulgada cuadrada. La finalidad del experimento es maximizar el *rendimiento* manteniendo la *resistencia* en 250 si es posible. Las cuatro columnas de la derecha son utilizadas para balancear los requerimientos de las dos respuestas, que pueden entrar en conflicto. *Impacto* especifica la importancia de cada respuesta en una escala de 1 a 5, con 5 significando la mayor importancia. El *mínimo* y *máximo* valores especifican el rango deseable para cada respuesta, con *sensibilidad* indicando lo importante que puede ser acertar con la mejor posición en el rango. En el ejemplo actual, *resistencia* es más importante que *rendimiento* y de este modo se le asigna mayor impacto. La sensibilidad de ambas respuestas se sitúa en “Medio”, lo cual significa que la deseabilidad de cada respuesta crece de forma lineal en el rango especificado.

Paso 2: Definir factores experimentales

El botón *Paso 2* se utiliza para introducir información acerca de los factores experimentales que será cambiada durante el transcurso del experimento. Muestra el cuadro de diálogo que se presenta a continuación:

Asistente de diseño de experimentos - Definir Factores

Archivo de diseño: <sin título>
Comentario: Tutorial #7

Número de factores controlables del proceso: 5 Número de componentes controlables de mezcla: 0 Número de factores de ruido: 0

Factor	Nombre	Unidades	Tipo	Función	Bajo	Alto	Niveles
A	temperatura	grados C	Continuo	Controlable	150	180	1,2,3,4
B	tasa de flujo	litros/minuto	Continuo	Controlable	10	12	1,2,3,4
C	concentración	%	Continuo	Controlable	5	8	1,2,3,4
D	tasa de agitación	rpm	Continuo	Controlable	125	150	1,2,3,4
E	catálisis	%	Continuo	Controlable	1	1,5	1,2,3,4
F	Factor_F		Continuo		-1,0	1,0	1,2,3,4
G	Factor_G		Continuo		-1,0	1,0	1,2,3,4
H	Factor_H		Continuo		-1,0	1,0	1,2,3,4
I	Factor_I		Continuo		-1,0	1,0	1,2,3,4
J	Factor_J		Continuo		-1,0	1,0	1,2,3,4
K	Factor_K		Continuo		-1,0	1,0	1,2,3,4
L	Factor_L		Continuo		-1,0	1,0	1,2,3,4
M	Factor_M		Continuo		-1,0	1,0	1,2,3,4

Total para componentes controlables de mezcla: 1,0

Factores A-M Factores N-Z

Aceptar Atrás Cancelar Ayuda

Figura 16-3. Definición de los factores experimentales

En el ejemplo, varían 5 factores controlables del proceso. Introducir el nombre de cada factor, sus unidades y el rango sobre el cual variarán. Todos los factores son *continuos*, ya que pueden ser configurados con un valor entre los niveles más bajo y más alto indicados.

Paso 3: Seleccionar diseño

El tercer paso en la creación de un experimento es seleccionar el tipo de diseño que se va a ejecutar. Cuando se presiona el botón del *paso 3*, el primer cuadro de diálogo obtenido se muestra a continuación:

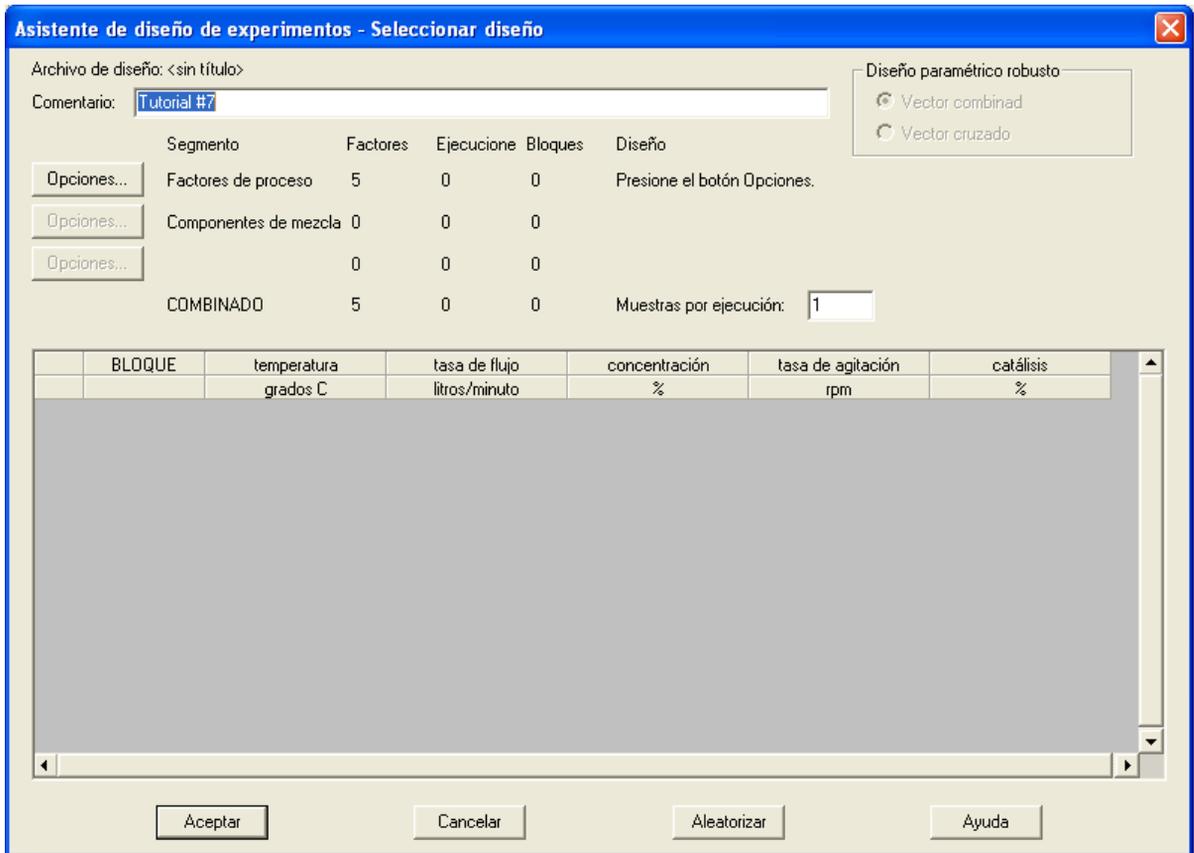


Figura 16-4. Seleccionar el cuadro de diálogo del diseño

Para crear un diseño para los 5 factores del proceso 5, presione el botón *Opciones*. Muestra una lista de tipos de diseños que puede ser apropiada para 5 factores continuos:

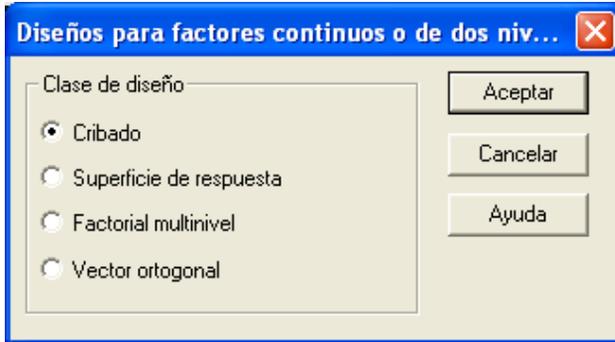


Figura 16.5. Cuadro de diálogo mostrando los tipos disponibles de diseños

Ya que queremos crear un diseño de pantalla, presionamos *Aceptar*.

El siguiente cuadro de diálogo se utiliza para elegir el diseño deseado de un catálogo de diseños de pantalla apropiados para los 5 factores:

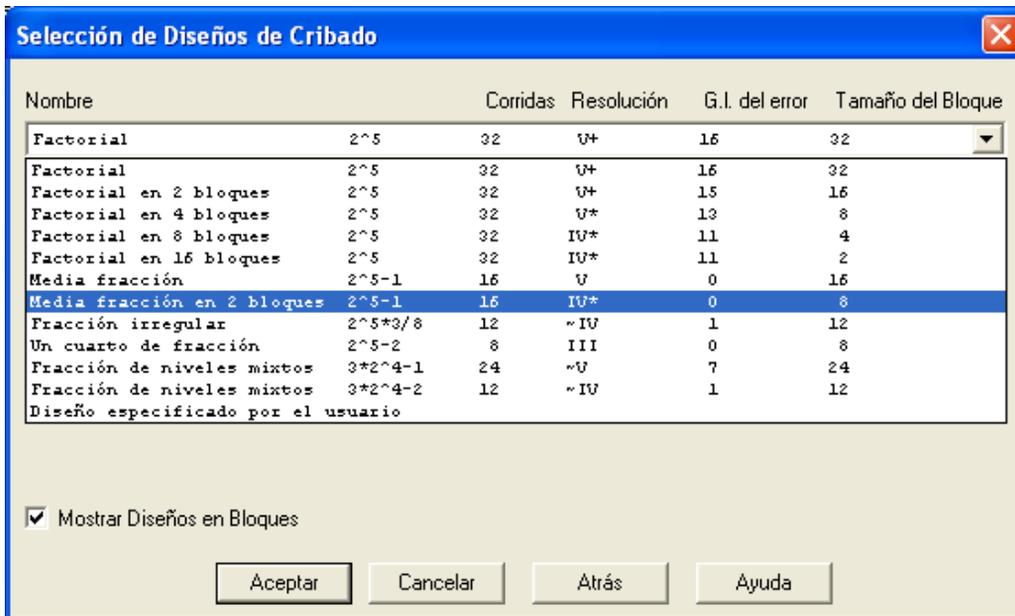


Figura 16-6. Selección de diseño

Para ver la lista de diseños de pantalla disponibles para 5 factores, haga clic en la flecha que muestra la lista. En la lista tenemos:

1. **Nombre:** el nombre de cada diseño disponible.

2. **Ejecuciones:** el número de ejecuciones en el diseño base, antes de que se añada la ejecución de cualquier punto central o replicación.
3. **Resolución:** la resolución del diseño. Un diseño de resolución V puede estimar todos los efectos principales y todas las interacciones de dos factores. Un diseño de resolución IV puede estimar todos los efectos principales, sin embargo las interacciones de dos factores se confundirán entre ellos mismos o con efectos de bloque. Resolución III confunde interacciones de dos factores con efectos principales.
4. **Error d.f.:** el número de grados de libertad disponible para estimar el error experimental. La potencia de este test estadístico está relacionada con el número de grados de libertad, así como con el número total de ejecuciones en el experimento. Normalmente, al menos deben estar disponibles 3 grados de libertad, aunque siempre es preferible más.
5. **Tamaño de bloque:** el número de ejecuciones en el bloque más largo.

En este caso, el ingeniero ha seleccionado una fracción un medio en dos bloques de 8 ejecuciones cada uno.

El cuadro de diálogo final se utiliza para añadir puntos centrales o replicar ejecuciones:

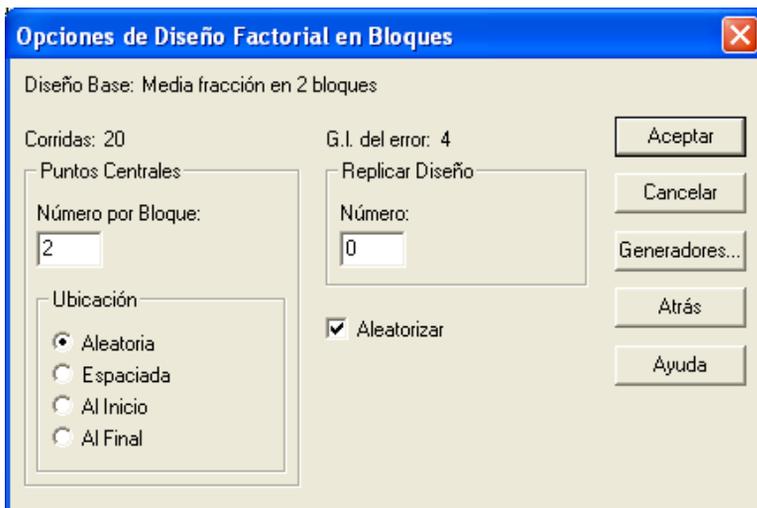


Figura 16-7. Opciones de diseño de pantalla bloqueado

Los campos de entrada especifican:

1. *Puntos centrales*: el número de ejecuciones a realizar en el centro de la región del experimento. Añadir puntos centrales es una buena idea para aumentar los grados de libertad para el error experimental.
2. *Ubicación*: La ubicación de los puntos centrales. Las elecciones más comunes son *Aleatoria*, en la cual los puntos centrales se extienden aleatoriamente a través de otras ejecuciones, y *Espaciada*, en la cual los espacios de los puntos centrales se extienden uniformemente a través del diseño.
3. *Diseño replicado*: el número de veces adicionales que cada conjunto de condiciones experimentales es ejecutado. Replicando el diseño completo se puede aumentar el número de ejecuciones rápidamente.
4. *Aleatorizar*: cuando las ejecuciones se listan en orden aleatorio. La aleatorización debe ser hecha siempre que sea posible para prevenir variables externas acechantes (tales como cambios en el proceso en el tiempo) que sesgan los resultados.

En el experimento actual, han sido requeridos cuatro puntos centrales, tomando hasta 20 ejecuciones para el diseño final. Ha sido requerido también que el diseño sea realizado en orden aleatorio, lo que significa que el orden de las 10 ejecuciones en cada bloque se genere aleatoriamente.

Después del cuadro de diálogo final, la ventana *Seleccionar diseño* se rellena con las ejecuciones del experimento a realizar:

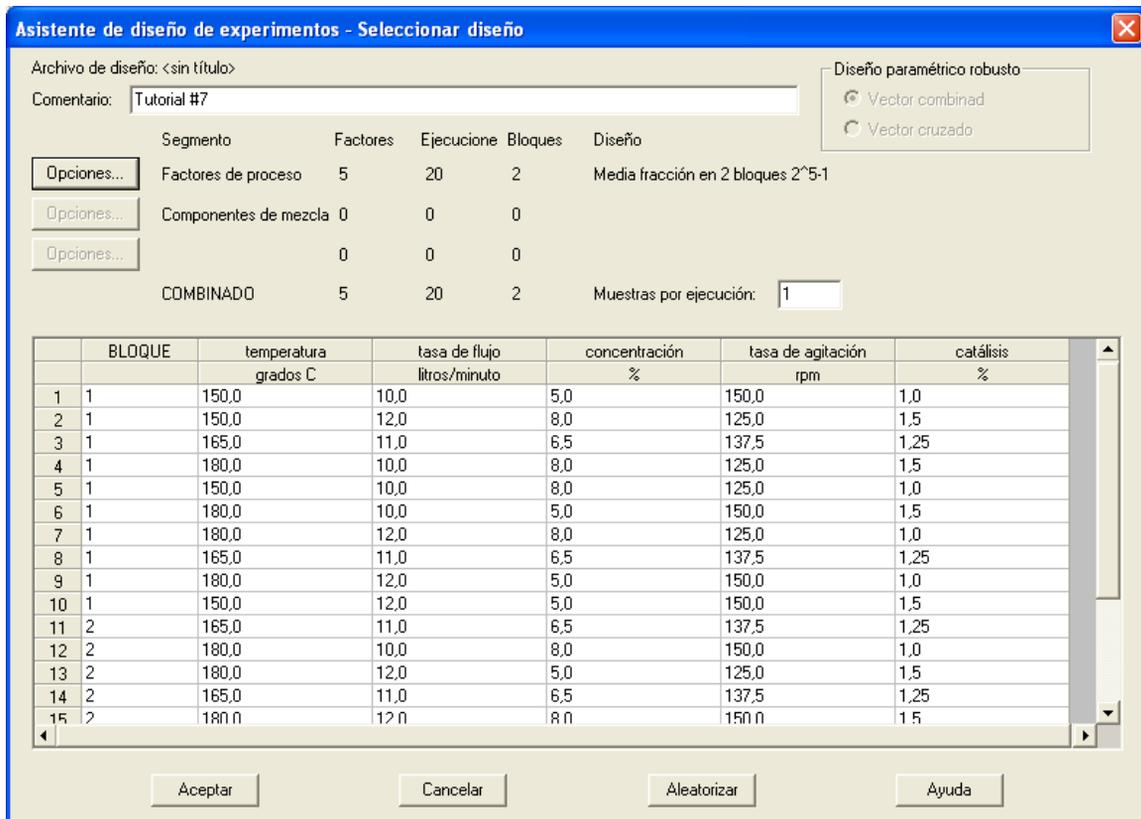


Figura 16-8. Seleccionar ventana del diseño con las ejecuciones a realizar

Si está satisfecho con el diseño, presione *Aceptar* una vez más para ejecutar el *Asistente de diseño de experimentos* en la ventana que resumirá las selecciones hechas hasta el momento:

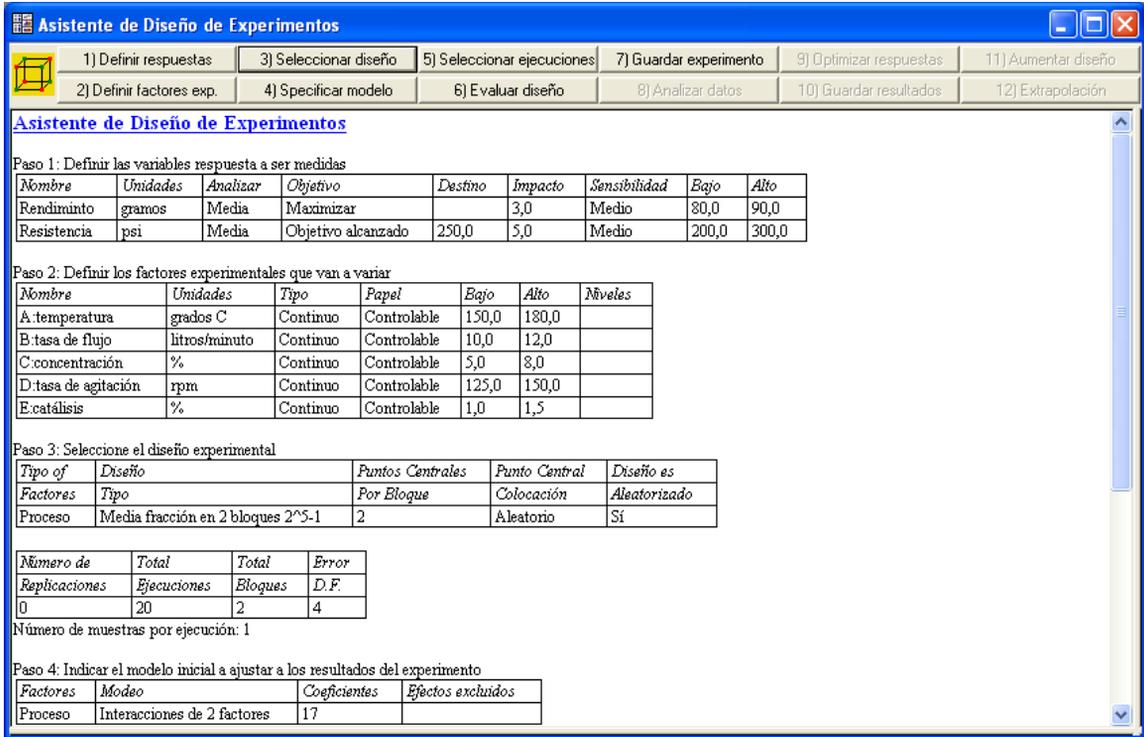


Figura 16-9. Ventana del asistente de diseño de experimentos después de seleccionar un diseño

Al mismo tiempo, el diseño ha sido leído en la hoja de datos en un libro de datos de STATGRAPHICS Centurion XVI:

	BLOQUE	temperatura	tasa de flujo	concentración	tasa de agitación	catálisis	Rendimiento	Resistencia
		grados C	litros/minuto	%	rpm	%	gramos	psi
1	1	150,0	10,0	5,0	150,0	1,0		
2	1	150,0	12,0	8,0	125,0	1,5		
3	1	165,0	11,0	6,5	137,5	1,25		
4	1	180,0	10,0	8,0	125,0	1,5		
5	1	150,0	10,0	8,0	125,0	1,0		
6	1	180,0	10,0	5,0	150,0	1,5		
7	1	180,0	12,0	8,0	125,0	1,0		
8	1	165,0	11,0	6,5	137,5	1,25		
9	1	180,0	12,0	5,0	150,0	1,0		
10	1	150,0	12,0	5,0	150,0	1,5		
11	2	165,0	11,0	6,5	137,5	1,25		
12	2	180,0	10,0	8,0	150,0	1,0		
13	2	180,0	12,0	5,0	125,0	1,5		
14	2	165,0	11,0	6,5	137,5	1,25		
15	2	180,0	12,0	8,0	150,0	1,5		
16	2	150,0	12,0	8,0	150,0	1,0		
17	2	180,0	10,0	5,0	125,0	1,0		
18	2	150,0	10,0	5,0	125,0	1,5		
19	2	150,0	12,0	5,0	125,0	1,0		
20	2	150,0	10,0	8,0	150,0	1,5		

Figure 16-10. Diseño final

La hoja de datos contiene una columna con números de bloque, 5 columnas con las configuraciones de los factores experimentales y 2 columnas para introducir las respuestas una vez que las ejecuciones del diseño experimental han sido realizadas.

Paso 4: Especificar el modelo

El Asistente de diseño de experimentos evaluará el diseño que ha creado respecto a un modelo estadístico especificado. Si ha presionado el botón etiquetado *Paso 4*, se mostrará el siguiente cuadro de diálogo:

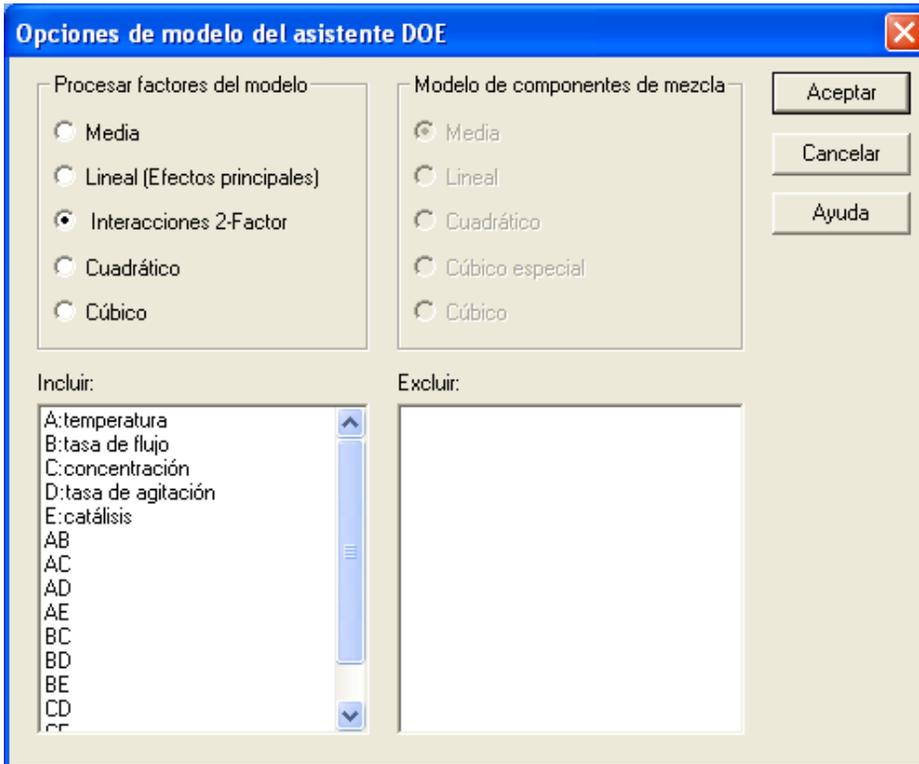


Figura 16-11. Cuadro de diálogo de selección del modelo

Debe seleccionar el modelo más complicado que quiera utilizar para sus datos. En este caso de dos niveles factoriales, el modelo más complicado que puede ser ajustado es el modelo de dos factores con interacción definido como sigue:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 + \beta_{15} x_1 x_5 + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{25} x_2 x_5 + \beta_{34} x_3 x_4 + \beta_{35} x_3 x_5 + \beta_{45} x_4 x_5$$

Está formado por cada factor experimental aislado (efectos principales) y términos envolviendo cada par de factores (interacciones de dos factores). Los términos individuales pueden ser

excluidos del modelo seleccionado haciendo doble clic en ellos con el ratón, a los cuales mueve a su campo en el cuadro de diálogo. En este caso, seleccionaremos el modelo con interacciones completo de dos factores.

Paso 5: Seleccionando ejecuciones

Para diseños más complicados, puede ser deseable ejecutar solo un subconjunto de ejecuciones de las que fueron creadas en el paso 3. Si se presiona el botón *paso 5*, se accede a un algoritmo de selección de ejecución que crea un subconjunto de las ejecuciones que es *D-óptimo*. En este caso, todas las ejecuciones se realizarán y se omitirá el paso 5.

Paso 6: Evaluar el diseño

Si presionamos el botón etiquetado *Paso 6*, se mostrará un cuadro de diálogo con una lista de tablas y gráficos que pueden ser añadidos a la ventana del Asistente de diseño de experimentos:

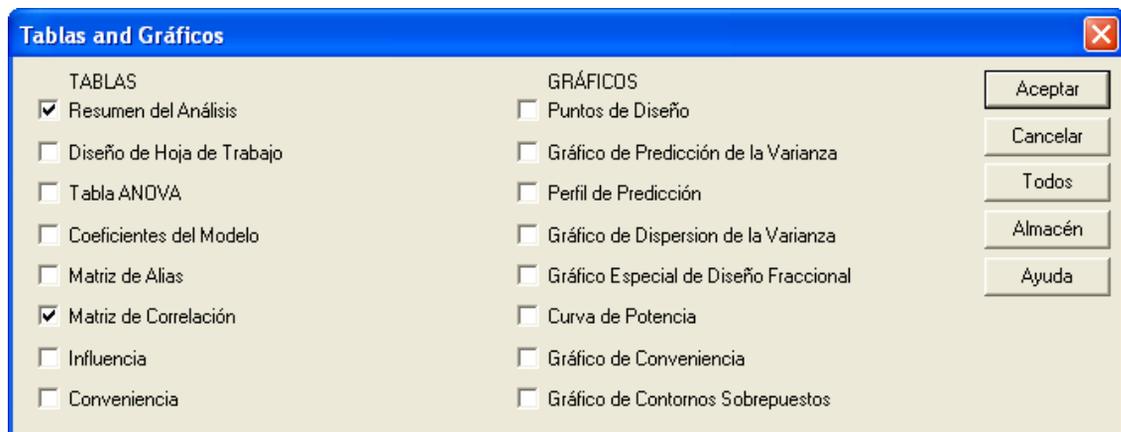


Figura 16-12. Tablas y Gráficos para evaluar el diseño de experimentos seleccionado

Una opción habitual para los diseños de pantalla es la *Matriz de correlaciones*, que muestra cuando hay una confusión entre los términos del modelo que va a ser ajustado:

Matriz de Correlación

	bloque	A	B	C	D	E	AB	AC	AD	AE	BC	BD	BE
bloque	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
A	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
B	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
C	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
D	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
E	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
AB	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
AC	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000	0,0000
AD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000	0,0000
AE	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000	0,0000
BC	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000	0,0000
BD	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000	0,0000
BE	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	1,0000
CD	0,8944	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
CE	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
DE	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

	CD	CE	DE
bloque	0,8944	0,0000	0,0000
A	0,0000	0,0000	0,0000
B	0,0000	0,0000	0,0000
C	0,0000	0,0000	0,0000
D	0,0000	0,0000	0,0000
E	0,0000	0,0000	0,0000
AB	0,0000	0,0000	0,0000
AC	0,0000	0,0000	0,0000
AD	0,0000	0,0000	0,0000
AE	0,0000	0,0000	0,0000
BC	0,0000	0,0000	0,0000
BD	0,0000	0,0000	0,0000
BE	0,0000	0,0000	0,0000
CD	1,0000	0,0000	0,0000
CE	0,0000	1,0000	0,0000
DE	0,0000	0,0000	1,0000

Figura 16-13. Matriz de correlación para el diseño seleccionado

Un valor no cero en una celda de la diagonal de la tabla indica que los efectos de esta fila y columna están confundidos y que no pueden ser separados claramente. En el diseño actual, la interacción CD tiene una correlación grande con los bloques. Notar que el diseño tiene arbitrariamente sacrificada la posibilidad de estimar la interacción entre los factores C y D, que son *concentración* y *tasas de agitación*. Si esta es una interacción que el ingeniero cree importante

debe cambiar el orden de las variables C y D correspondientes a los dos variables de la interacción.

Paso 7: Guardar experimento

Presionando el botón etiquetado *Paso 7* es posible guardar el experimento en un archivo. Se utiliza el cuadro de diálogo que se muestra a continuación:

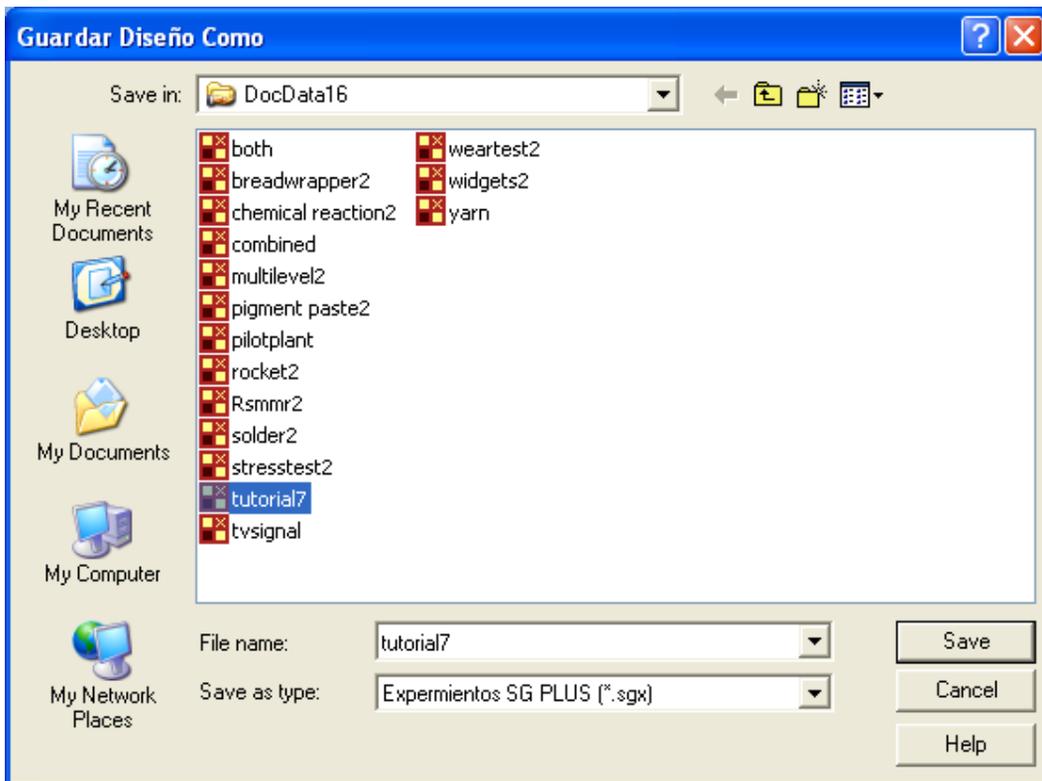


Figura 16-14. Cuadro de diálogo para Guardar experimento

El diseño de experimentos creado por el Asistente DOE se guardará en un archivo con la extensión *.sgx*. Es similar a los archivos de datos estándar, con la excepción de que contiene información adicional acerca del diseño del experimento y el modelo estadístico seleccionado.

16.2 Analizando los resultados

Después de diseñar el experimento, el ingeniero realiza las 20 ejecuciones indicadas. Reinicia el programa, abre el archivo de diseño de experimentos e introduce los valores medidos de *rendimiento* y *resistencia* en la hoja de datos del experimento. Para repicar el análisis, puede leer el archivo *tutorial7.sgx* de la misma manera que se puede leer un archivo de datos de STATGRAPHICS, seleccionando *Abrir origen de datos* en el menú *archivo*. Abriendo un archivo de diseño de experimentos automáticamente se abre la ventana principal del Asistente DOE.

Paso 8: Analizar datos

Para analizar los datos del experimento, presione el botón etiquetado *Paso 8*. El análisis comienza mostrando el siguiente cuadro de diálogo:

Respuesta	Transformación	Potencia	Sumando
Rendimiento	Ninguno	1.0	0
Resistencia	Ninguno	1.0	0

Aceptar Cancelar Ayuda

Figura 16-15. Cuadro de diálogo de entrada de *Analizar datos*

Si es necesario, se puede especificar una transformación para una o más variables. Ya que estamos analizando variables respuesta continuas, las transformaciones serán necesarias si la varianza de la respuesta crece con la media. En el ejemplo actual, no se necesitan transformaciones.

Cuando se presiona el botón *Aceptar*, se abre una nueva ventana de análisis para cada respuesta. La ventana de análisis para *Rendimiento* muestra inicialmente la siguiente salida:

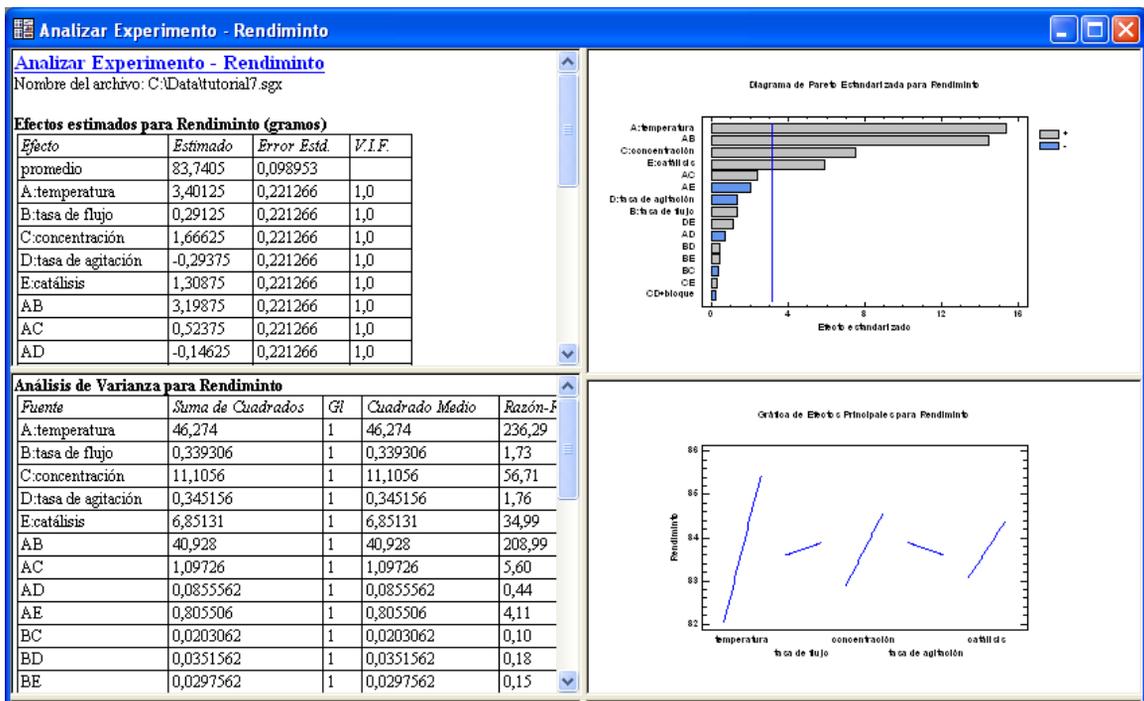


Figura 16-16. Ventana de análisis *Analizar diseño* para rendimiento

La ventana muestra cuatro paneles:

1. *Resumen de análisis*: lista los efectos principales estimados y sus interacciones.
2. *Tabla ANOVA*: contiene *P*-valores que pueden ser utilizados para contrastar la significación estadística de cada efecto.
3. *Gráfico de Pareto estandarizado*: muestra los efectos en orden decreciente de significación, con una línea para determinar qué efectos son estadísticamente significativos.

4. *Gráfico de efectos principales*: representa el cambio estimado en la respuesta cuando cada uno de los factores se mueve de su nivel bajo a su nivel alto.

El gráfico de Pareto estandarizado en la esquina superior derecha puede ser utilizado para determinar rápidamente qué efectos son los más importantes:

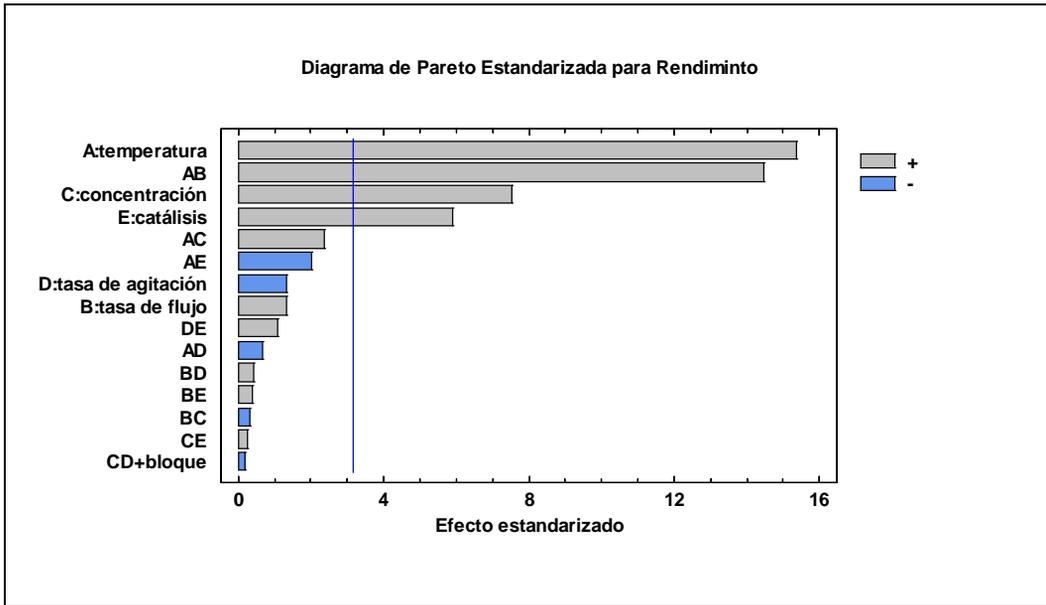


Figura16-17. Gráfico de Pareto estandarizado

La longitud de cada barra es proporcional al valor del estadístico t calculado para el correspondiente efecto. Barras fuera de las líneas verticales son estadísticamente significativas al nivel de confianza seleccionado, situado por defecto al 5%. En este caso, hay tres efectos principales significativos: *temperatura*, *concentración*, y *catálisis*. Hay también una interacción significativa entre *temperatura* y *tasa de flujo*.

El panel *Gráfico de efectos principales* en la parte inferior derecha muestra cómo cada factor afecta a *rendimiento*:

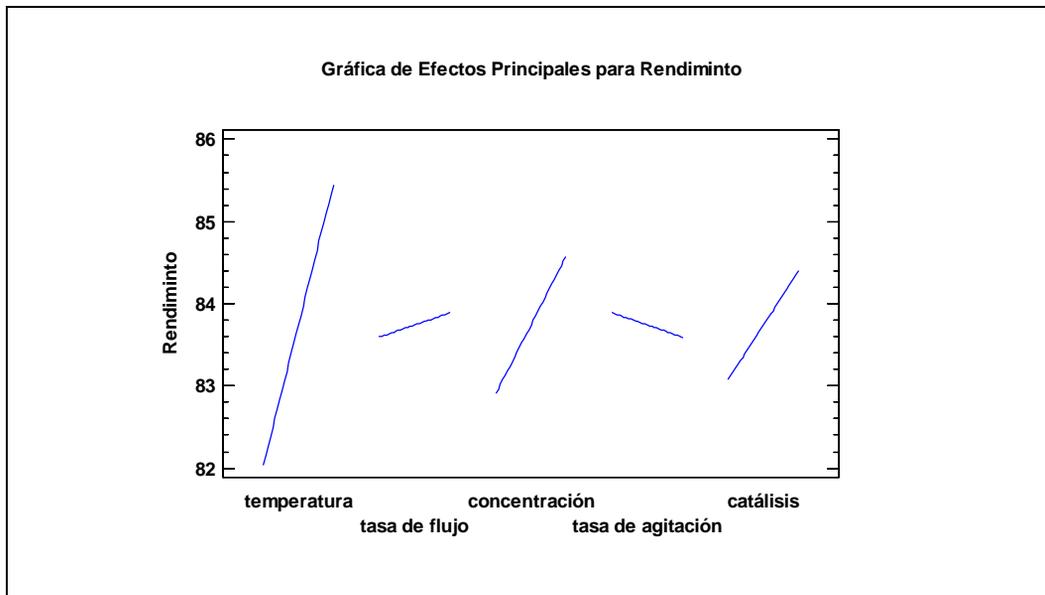


Figura 16-18. Gráfico de efectos principales

Las líneas indican los cambios estimados en *rendimiento* cuando cada factor se mueve de su nivel bajo a su nivel alto, con todos los otros factores constantes en un valor a mitad de camino entre sus valores bajos y altos. Notar que los tres factores con efectos principales significativos tienen mayor impacto en unas respuestas que en las otras. Por ejemplo, el rendimiento medio de las temperaturas bajas es aproximadamente 82, mientras que el rendimiento medio de las temperaturas altas es aproximadamente 85.4. La diferencia de 3.4 se llama “efecto principal” de temperatura.

Para el gráfico de interacción entre *temperatura* y *tasa de flujo*, seleccione primero *Gráfico de interacción* de la caja de diálogo *Gráficos*. Entonces utilice *Opciones de panel* para seleccionar sólo estos dos factores:



Figura 16-19. Cuadro de diálogo Opciones de panel para gráfico de interacción

El gráfico resultante muestra el rendimiento medio con el cambio de la temperatura, para cada nivel de tasa de flujo:

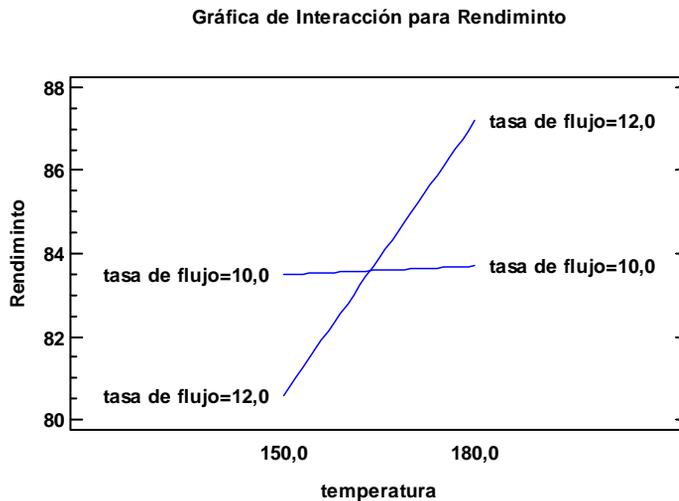


Figura 16-20. Gráfico de interacción par tasa de flujo y temperatura

Notar que en tasa de flujo baja, temperatura tiene un efecto pequeño si existe. En tasa de flujo alta, temperatura es un factor muy importante. Antes de utilizar el modelo estadístico subyacente a este análisis, es importante eliminar efectos no significativos. Para eliminar efectos:

1. Presione el botón *Opciones de análisis* en la barra de herramientas de análisis.
2. Presione el botón *Excluir* en el cuadro de diálogo *Opciones de análisis*.
3. En el cuadro de diálogo *Opciones de excluir efectos*, haga doble clic en el efecto que quiere excluir, el cual se moverá de la columna *Incluir* a la columna *Excluir*:

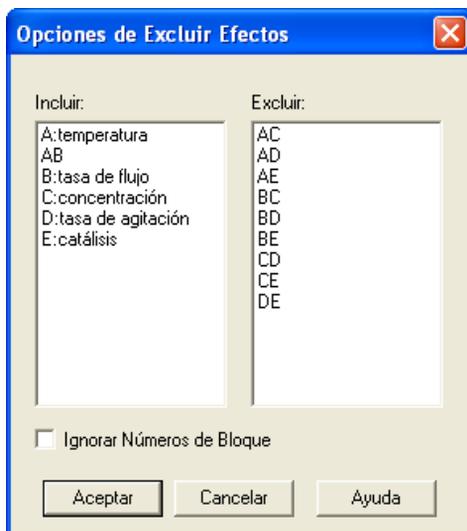


Figura 16-21. Cuadro de diálogo para excluir efectos

La regla que se sigue en la exclusión de efectos es:

1. Excluir interacciones no significativas de dos factores.
2. Excluir efectos principales no significativos que no están envueltos en interacciones significativas.

En este caso, estas medias se remueven cada vez que no son significativas en el gráfico de Pareto, excepto para el efecto principal de B. Este efecto principal se retiene porque está envuelto en una interacción significativa con el factor A.

Una vez que los efectos han sido removidos, aparecerá el gráfico de Pareto como se muestra a continuación:

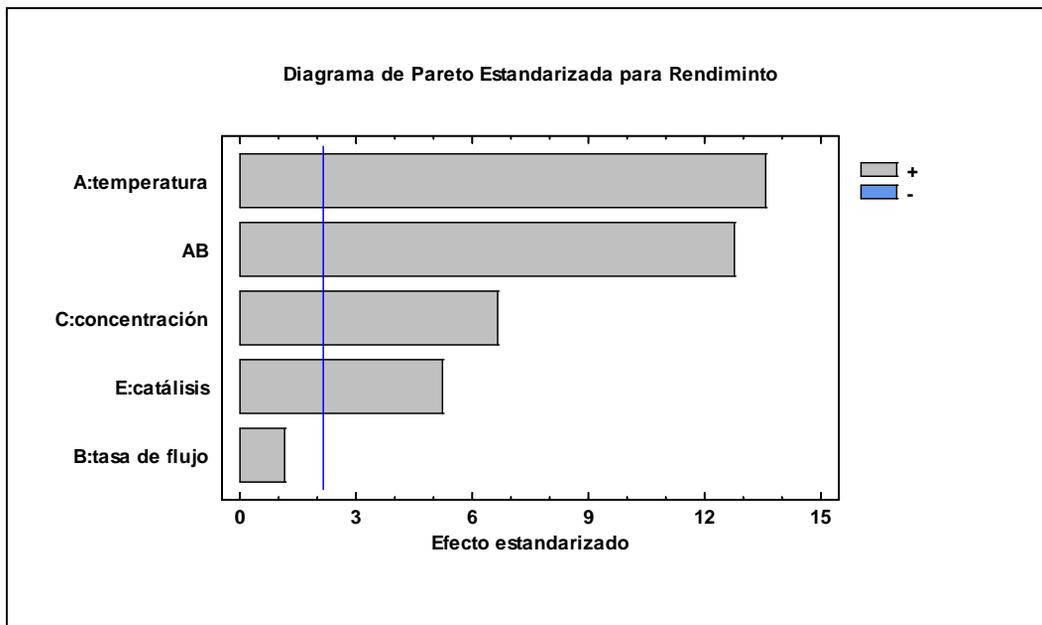


Figura 16-22. Gráfico de Pareto estandarizado para efectos removidos

Excepto para el efecto principal del factor B, todos los efectos restantes son estadísticamente significativos. El modelo final puede ser observado seleccionando *Coefficientes de regresión* del la caja de diálogo *Tablas*:

Coef. de regresión para Rendimiento	
Coeficiente	Estimado
constante	250,074
A:temperatura	-1,0595
B:tasa de flujo	-17,4475
C:concentración	0,555417
E:catálisis	2,6175
AB	0,106625

El StatAdvisor
 Esta ventana despliega la ecuación de regresión que se ha ajustado a los datos. La ecuación del modelo ajustado es

$$\text{Rendimiento} = 250,074 - 1,0595 \cdot \text{temperatura} - 17,4475 \cdot \text{tasa de flujo} + 0,555417 \cdot \text{concentración} + 2,6175 \cdot \text{catálisis} + 0,106625 \cdot \text{temperatura} \cdot \text{tasa de flujo}$$

Figura 16-23. Modelo de regresión ajustado para rendimiento

Notar que los modelos toman la forma de un modelo lineal de regresión múltiple. Cada efecto principal retenido está incluido en el modelo por sí mismo, mientras que las interacciones de dos factores están representadas por los productos cruzados de *temperatura* y *tasa de flujo*.

Para comprender completamente el modelo ajustado, lo mejor es representarlo gráficamente. Pueden crearse varios tipos de gráficos seleccionando *gráficos de respuesta* del cuadro de diálogo *Tablas y Gráficos*. Por defecto, se muestra una superficie como marco:

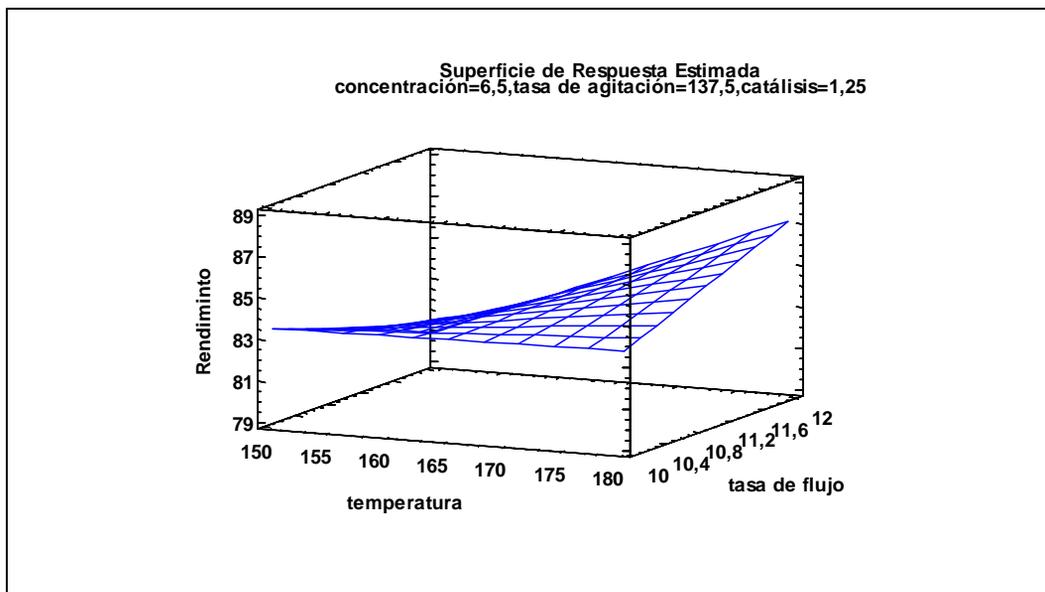


Figura 16-24. Gráfico de superficie de respuesta

En este gráfico, la altura de la superficie representa los valores predichos de *rendimiento* sobre el espacio de *temperatura* y *tasa de flujo*, con los otros tres factores constantes en sus valores medios. Los rendimientos más altos se obtienen temperaturas altas y tasa de flujo altas.

El tipo de gráfico y los factores sobre los que se representa la respuesta pueden ser cambiados utilizando *Opciones de panel*.



Figura 16-25. Opciones de panel para gráficos de respuesta

Los tipos de gráficos que pueden ser creados son:

1. *Superficie*: gráficos que ajustan una ecuación como una superficie 3-D con respecto a los dos factores experimentales. La superficie puede tener marco, color sólido, o mostrar niveles de contorno para las respuestas. *Contornos inferiores* incluye contornos en la cara inferior del gráfico.
2. *Contorno*: crea un gráfico de contorno 2-D con respecto a los dos factores experimentales. Los contornos pueden mostrar líneas, como un mapa topográfico, regiones pintadas, o rampas de color continuo.
3. *Cuadrado*: grafica la región experimental para cualesquiera dos factores experimentales y muestra la respuesta predicha en cada esquina del cuadrado.
4. *Cubo*: grafica la región experimental para 3 factores experimentales y muestra la respuesta predicha en cada esquina del cubo. Para crear este gráfico, debe primero presionar el botón *Factores* y seleccionar un tercer factor.

5. *Contornos 3-D*: dibuja contornos para la respuesta respecto a 3 factores experimentales simultáneamente.
6. *Malla 3-D*: crea un gráfico de mallas mostrando el valor de la variable respuesta a través de una región experimental tridimensional.

El botón *Factores* se usa para seleccionar los factores que definen los ejes de los gráficos y los valores de cada uno de los otros factores:

	Bajo	Alto	Mantener
<input checked="" type="checkbox"/> temperatura	150,0	180,0	165,0
<input checked="" type="checkbox"/> tasa de flujo	10,0	12,0	11,0
<input type="checkbox"/> concentración	5,0	8,0	6,5
<input type="checkbox"/> tasa de agitación	125,0	150,0	137,5
<input type="checkbox"/> catálisis	1,0	1,5	1,25
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0
<input type="checkbox"/>	0,0	1,0	0,0

Figura 16-26. Cuadro de diálogo de opciones de factor en gráficos de respuesta

Para crear el gráfico siguiente, el campo *Contornos* ha sido configurado como *Pintado*, el campo *Superficie* como *Sólido* con *Contornos inferiores* y los contornos han sido escalados en rangos de 81 a 86 con paso 1:

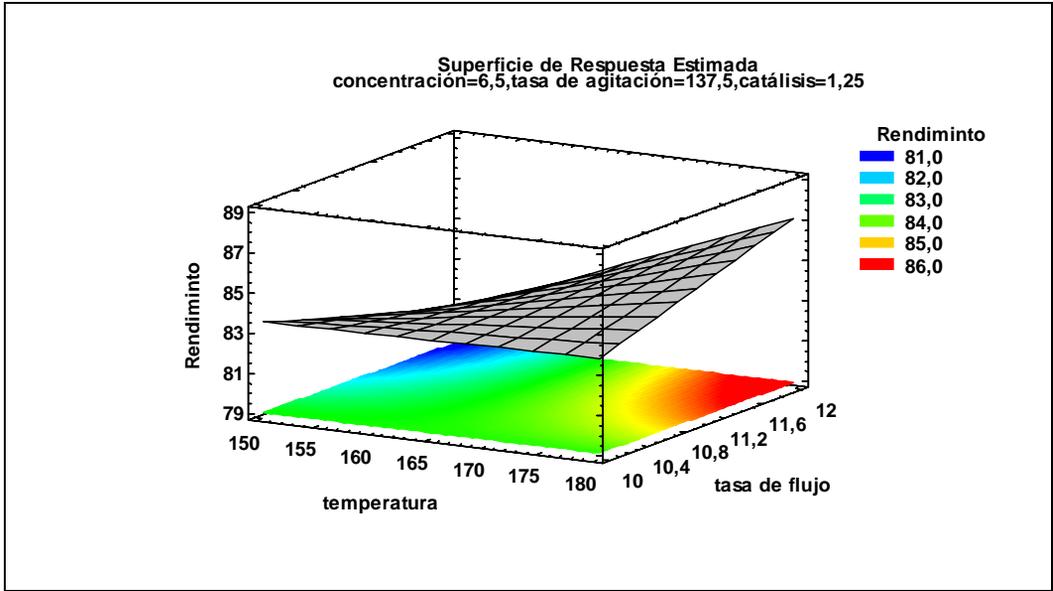


Figura 16-27. Gráfico de superficie de respuesta con contornos inferiores

El mismo gráfico puede ser mostrado como un gráfico de contorno en vez de cómo un gráfico de superficie:

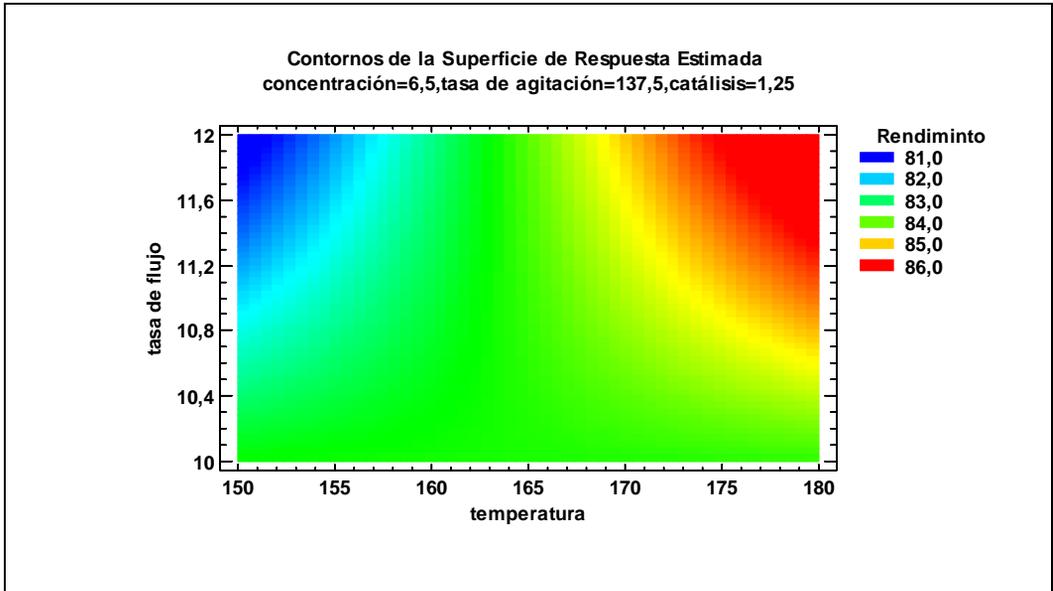


Figura 16-28. Gráfico de contorno de superficie de respuesta para Rendimiento

Valores altos de *rendimiento* se obtienen en la esquina superior derecha.

La segunda variable respuesta medida durante el experimento fue *resistencia*. La ventana de análisis para *Resistencia* muestra el siguiente gráfico de Pareto:

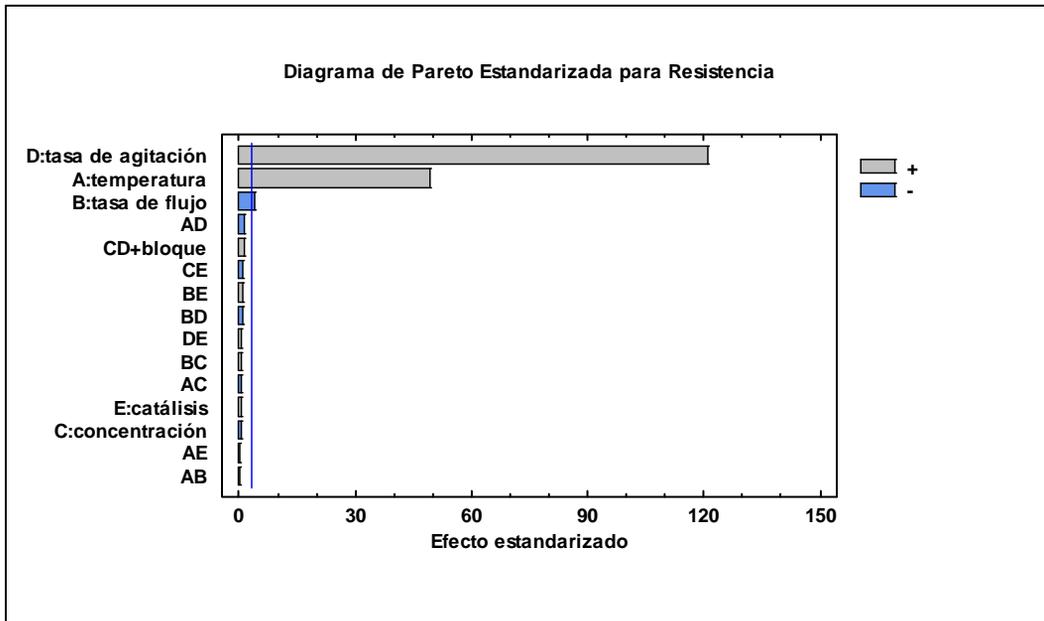


Figura 16-29. Gráfico de Pareto estandarizado para Resistencia

Después de eliminar los efectos no significativos, el modelo ajustado es:

$$\text{resistencia} = -317.288 + 1.02083 \cdot \text{temperatura} - 1.3125 \cdot \text{tasa de flujo} + 3.005 \cdot \text{tasa de agitación}$$

Notar que la *tasa de agitación* impacta en *resistencia*, aunque no tenga un efecto significativo en *rendimiento*. El gráfico de contorno para los dos factores más fuertes se muestra a continuación:

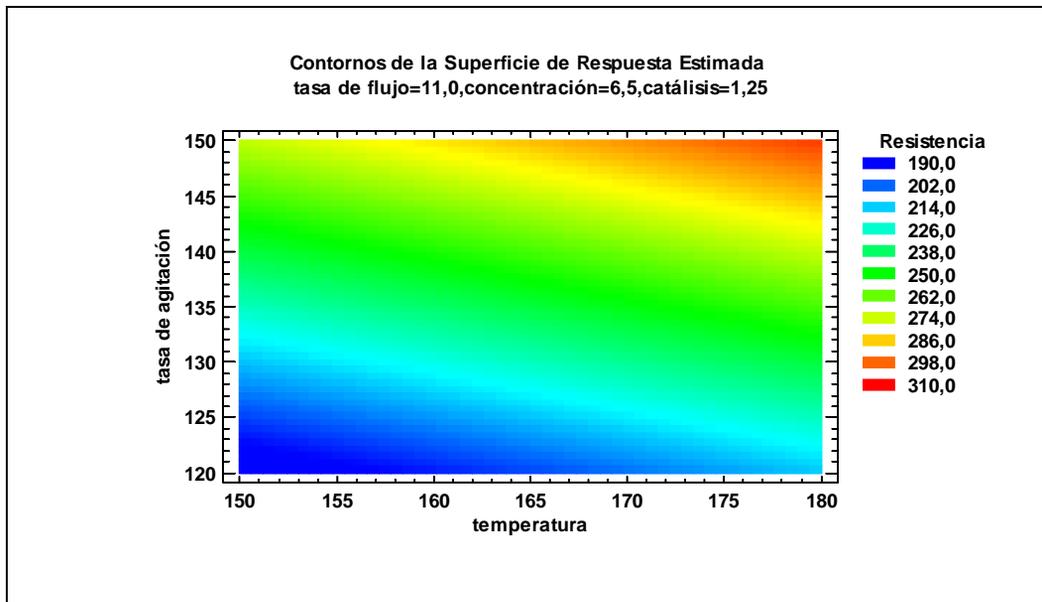


Figura 16-30. Gráfico de contorno para las superficies de respuesta de resistencia

Paso 9: Optimizar respuestas

Habiendo construido modelos significativos para ambas respuestas, las características óptimas de los factores pueden ser ahora determinadas. Recordar que la finalidad del experimento era maximizar *rendimiento* mientras se mantiene *resistencia* hasta 250 p.s.i. como posible. Si se presiona el botón etiquetado *paso 9*, se muestra el cuadro de diálogo siguiente:

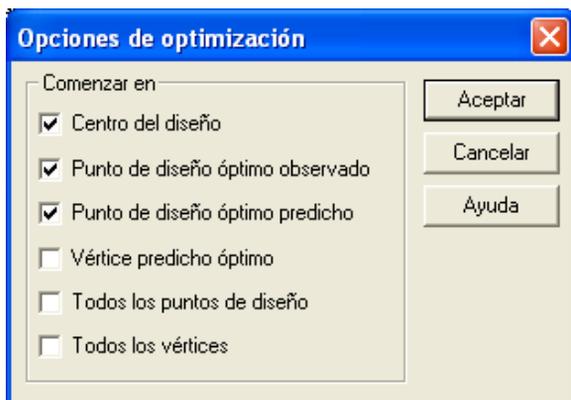


Figura 16-31. Cuadro de diálogo de opciones de optimización

Ya que el programa utiliza búsqueda numérica para encontrar la mejor localización en la región experimental, es buena idea comenzar la búsqueda de varios puntos posibles para encontrar un punto local óptimo.

Presione *Aceptar* para comenzar la búsqueda. Después de pocos momentos, se mostrará el siguiente mensaje:



Figura 16-32. Mensaje mostrado después de que la optimización ha sido completada

Al mismo tiempo, la siguiente salida se añadirá a la ventana principal del Asistente DOE:

<u>Paso 9: Optimizar las respuestas</u>				
Valores de respuesta y Óptimo				
Respuesta	Predicción	Inferior 95,0% Limite	Superior 95,0% Límite	Deseabilidad
Rendimiento	88,6734	78,5663	98,7804	0,867338
Resistencia	250,0	212,56	287,44	1,0
Sobredeseabilidad = 0,948027				
Factores establecidos y Óptimo				
Factor	Establecimiento			
temperatura	180,0			
tasa de flujo	12,0			
concentración	8,0			
tasa de agitación	132,875			
catálisis	1,5			

Figura 16-33. Resumen de optimización añadido a la ventana principal del Asistente DOE

En las características indicadas de los factores, se estima que *rendimiento* será igual a 88.67 gramos mientras *resistencia* será igual a 250 p.s.i. El rendimiento resultante tiene una “deseabilidad” cociente igual a 0.867, ya que es un 86.7% del rango especificado de 80 a 90 gramos. *Resistencia* tiene una deseabilidad cociente igual a 1, ya que está exactamente en el objetivo. La deseabilidad general es igual a 0.948, la cual es calculada tomando la deseabilidad de cada respuesta, elevándola a la potencia especificada por su *impacto*, multiplicando los resultados juntos y elevando el producto a una potencia 1 dividida por la suma de los impactos. El resultado es un número entre 0 y 1, con más peso en la respuesta que en el impacto alto.

Si se presiona el botón *Tablas y Gráficos* en la barra de herramientas de análisis, se pueden crear dos gráficos adicionales. El *Gráfico de contorno revestido* muestra los contornos de dos variables respuestas solapadas una en la otra:

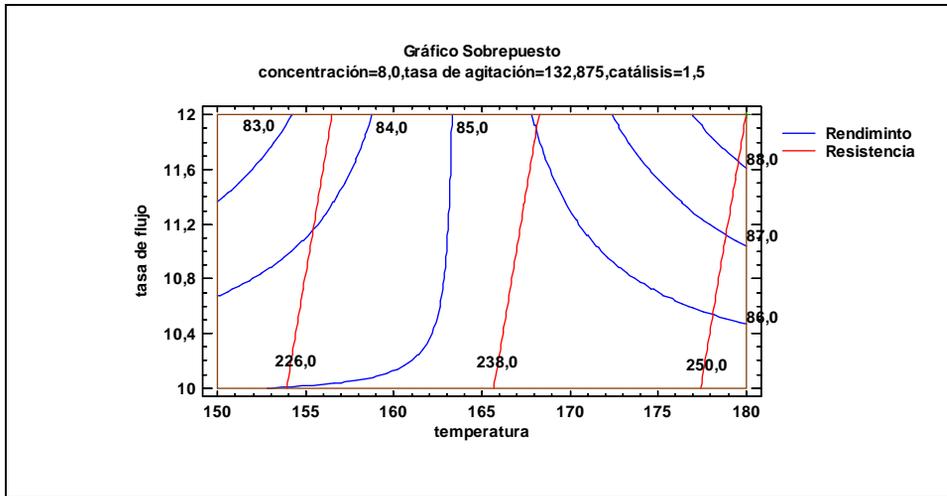


Figura 16-34. Gráfico de contorno solapado para las dos respuestas

El punto óptimo está en la esquina superior derecha, donde se maximiza *rendimiento* a través de la línea de *resistencia* = 250. El *Gráfico de deseabilidad* puede ser utilizado para mostrar la deseabilidad total para dos o tres factores a la vez. Seleccionando un gráfico de malla 3-D se crea el gráfico siguiente:

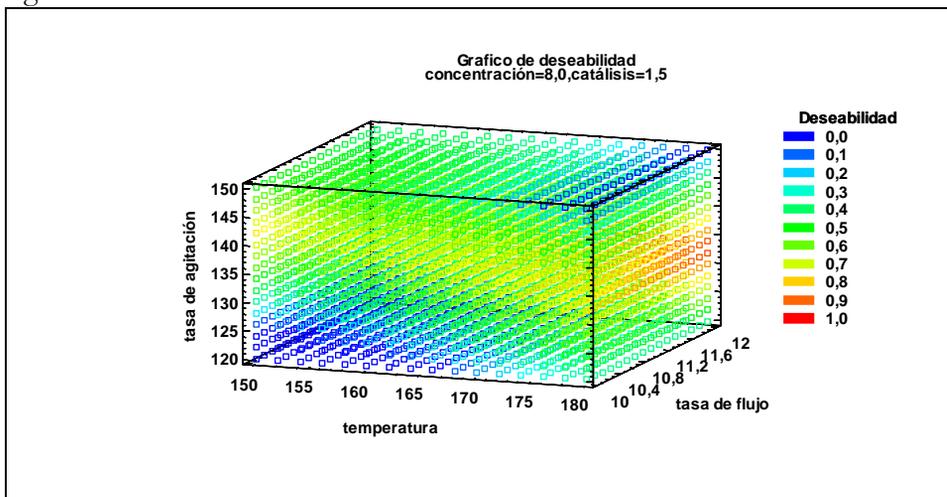


Figura 16-35. Gráfico de malla 3-D para la deseabilidad total

La mejor localización se muestra en rojo, donde *temperatura* y *tasa de flujo* son más altas, mientras *tasa de agitación* tiene un valor medio.

Paso 10: Guardar resultados

Para guardar los resultados del análisis y optimización, presione el botón etiquetado *paso 10* para guardar los resultados en un StatFolio:

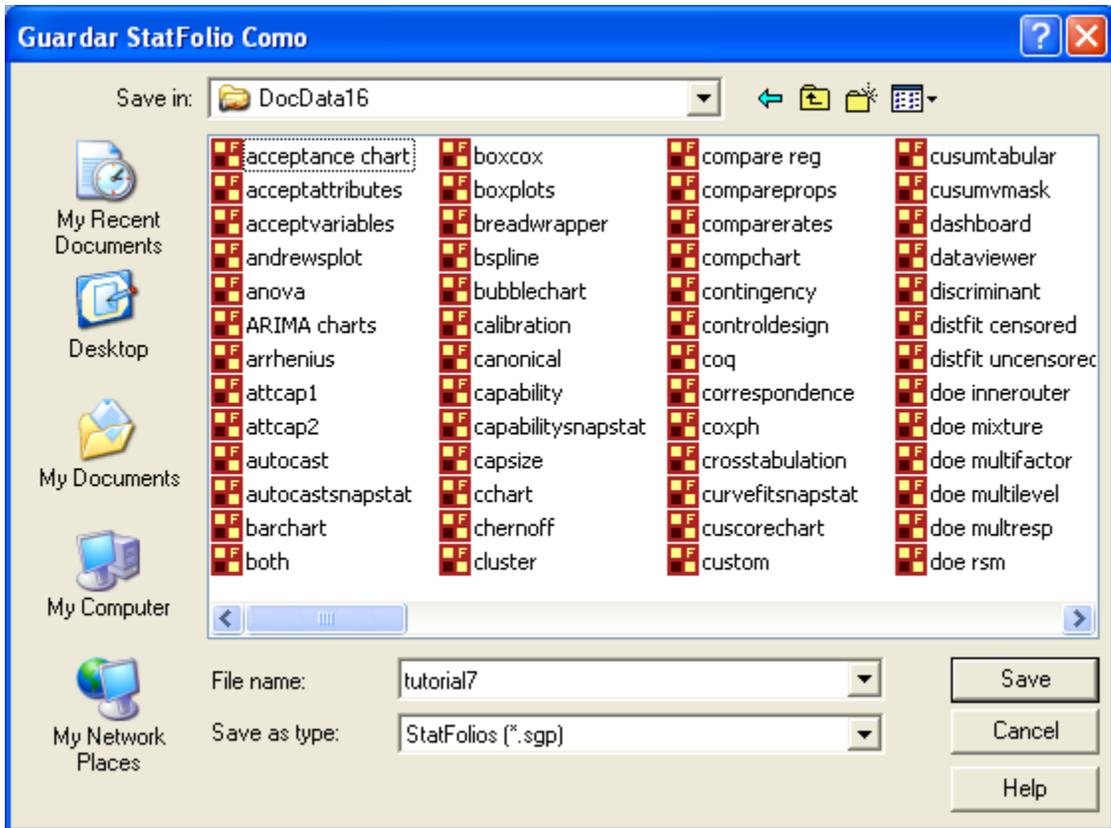


Figura 16-36. Cuadro de diálogo Guardar resultados

16.3 Experimentación posterior

Si se desea una experimentación posterior, STATGRAPHICS Centurion XVI puede ayudar aumentando el diseño existente o generando puntos a lo largo del camino de paso ascendente.

Paso 11: Aumentar diseño

Si presiona el botón etiquetado *Paso 11*, puede añadir ejecuciones adicionales al experimento actual. Comienza presentando el cuadro de diálogo que se muestra a continuación:

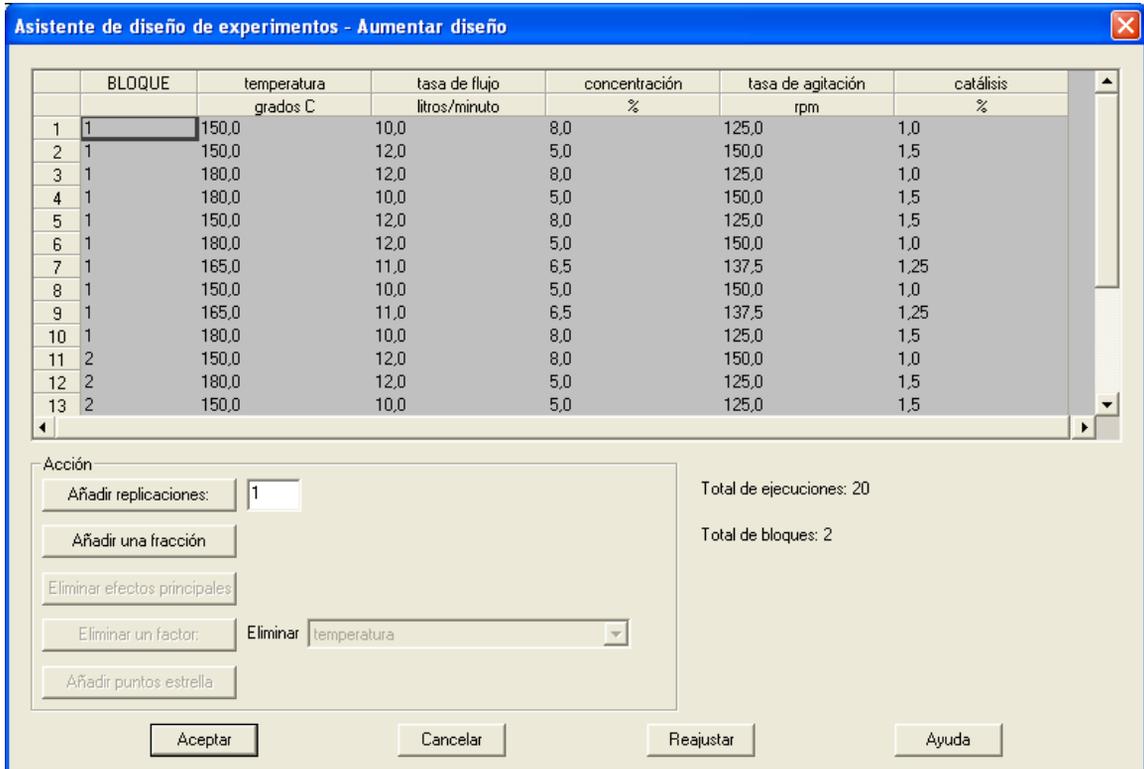


Figura 16-37. Cuadro de diálogo aumentar diseño

Son posibles tres opciones:

1. *Añadir réplicas*: añade otras 20 ejecuciones al diseño idénticas a las primeras 20. Esto permite añadir más grados de libertad para estimar el error experimental.
2. *Colapso del diseño*: remueve un factor experimental especificado del diseño y los análisis resultantes.
3. *Añadir una fracción*: añade 20 o más ejecuciones para hacer el diseño factorial completo.

Paso 12: Extrapolar

Puede también generar puntos a lo largo del *camino de paso ascendente* en un intento de moverse rápidamente a las regiones de más alto rendimiento. Desde un punto específico en la región experimental nos movemos en la dirección que presenta el mayor cambio en la respuesta estimada para el más pequeño cambio en los factores experimentales. Seguir este camino puede ser muy efectivo para la obtención rápida de mejoras muy fuertes.

Cuando se presiona el botón etiquetado *Paso 12*, se muestra el siguiente cuadro de diálogo:

Cambio	Comienzo	Bajo	Alto
<input checked="" type="checkbox"/> temperatura	180,0	135,0	195,0
<input checked="" type="checkbox"/> tasa de flujo	12,0	9,0	13,0
<input checked="" type="checkbox"/> concentración	8,0	3,5	9,5
<input checked="" type="checkbox"/> tasa de agitación	132,874	112,5	162,5
<input checked="" type="checkbox"/> catálisis	1,49999	0,75	1,75
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0
<input type="checkbox"/>	0,0	0,0	1,0

Figura 16-38. Cuadro de diálogo Extrapolar

Las características de este cuadro de diálogo instruyen al programa para comenzar en la derivada óptima y dejar variar 5 factores entre valores *bajo* y *alto* que dobla la anchura de la región experimental en cada dimensión. Muestra combinaciones de los factores siempre que la deseabilidad estimada sea al menos 0.5%. Después de presionar *Aceptar*, se añade la tabla siguiente a la ventana de análisis del Asistente DOE:

Paso 12: Extrapolar el modelo			
Valores de respuesta extrapolados			
Paso	Deseabilidad	Rendimiento	Resistencia
0	0,948022	88,6736	249,999
1	0,953451	88,8107	249,985
2	0,958935	88,9483	249,979
3	0,96487	89,094	250,012
4	0,97038	89,2326	250,01
5	0,975828	89,376	250,025
6	0,981255	89,519	250,035
7	0,987743	89,6839	250,023
8	0,993077	89,8332	250,051
9	0,99884	89,9783	249,972
10	0,999491	90,0045	249,959

Factores establecidos para la interpolación					
paso	temperatura	tasa de flujo	concentración	tasa de agitación	catálisis
0	180,0	12,0	8,0	132,874	1,49999
1	180,051	12,05	8,04282	132,874	1,5056
2	180,109	12,1	8,0836	132,874	1,51093
3	180,206	12,15	8,12264	132,874	1,51605
4	180,268	12,2	8,16179	132,874	1,52117
5	180,347	12,25	8,20118	132,874	1,52633
6	180,421	12,3	8,24027	132,874	1,53145
7	180,486	12,36	8,28549	132,874	1,53737
8	180,578	12,41	8,32475	132,874	1,5425
9	180,578	12,47	8,36554	132,874	1,54784
10	180,578	12,48	8,37467	132,874	1,54904

Figura 16-39. Resumen de extrapolación añadido a la ventana principal del Asistente DOE

Se estima que el rendimiento puede ser elevado a su valor objetivo de 90 gramos manteniéndose una Resistencia igual a 250 aumentando la temperatura a 180.6 grados, aumentando la tasa de flujo a 12.48 litros por minuto, aumentando la concentración a 8.37% y aumentando la catálisis a 1.55%. Ya que esto es una extrapolación del modelo ajustado estadísticamente fuera de la región experimental original, es necesario realizar ejecuciones confirmatorias para verificar el resultado.

Lecturas sugeridas

Los libros siguientes son excelentes y constituyen orígenes amenos de información acerca de las técnicas que se describen en esta guía:

Basic statistics: Applied Statistics and Probability for Engineers, 4th edition, by Douglas C. Montgomery and George C. Runger (2006). John Wiley and Sons, New York.

Analysis of variance: Applied Linear Statistical Models, 5th edition, by Michael H. Kutner, Christopher J. Nachtsheim, and John Neter (2004). McGraw Hill.

Regression analysis: Applied Linear Regression, 3rd edition, by Sanford Weisberg (2005). John Wiley and Sons, New York.

Statistical process control: Introduction to Statistical Quality Control, 6th edition, by Douglas C. Montgomery (2008). John Wiley and Sons, New York.

Design of experiments: Statistics for Experimenters: Design, Innovation and Discovery, 2nd edition by George E. P. Box, William G. Hunter, and J. Stuart Hunter (2005). John Wiley and Sons, New York.

Conjuntos de datos

93cars.sgd

Descargados del Journal of Statistical Education (JSE) Data Archive. Fue compilado por Robin Lock del Mathematics Department at St. Lawrence University permission es usado con su permiso. Un artículo asociado a este conjunto de datos aparece en el *Journal of Statistics Education*, Volume 1, Number 1 (July 1993).

bodytemp.sgd

Descargados del Journal of Statistical Education (JSE) Data Archive. Fue compilado por Allen Shoemaker del Psychology Department en el Calvin College permission es usado con su permiso. Los datos fueron extraídos de un artículo del *Journal of the American Medical Association* (1992, vol. 268, pp. 1578-1580) titulado "A Critical Appraisal of 98.6 Degrees F, the Upper Limit of the Normal Body Temperature, and Other Legacies of Carl Reinhold August Wunderlich" by P. A. Mackowiak, S. S. Wasserman, and M. M. Levine. U artículo asociado con este conjunto de datos aparece en el *Journal of Statistics Education*, Volume 4, Number 2 (July 1996).

Journal of Statistical Education (JSE) Data Archive web site:

http://www.amstat.org/publications/jse/jse_data_archive.htm

Índice

- ABS, 46
- actualizar enlaces, 145
- Actualizar fórmulas*, 45
- aleatorizar, 264
- alisando un gráfico de dispersión, 101
- análisis de correlación, 202
- análisis de la capacidad, 239
- análisis de la regresión, 201
- análisis de la varianza, 192
- análisis de medias, 200
- Análisis de Pareto*, 223
- Análisis de una variable*, 21, 151, 240
- AND, 65
- ANOM, 200
- ANOVA, 192
- archivos, 18
- archivos de datos
 - lectura, 36
- archivos HTML, 114
- ASCII, 38
- asimetría, 154
- Asistente de diseño de experimentos, 257
- atípicos lejanos, 157
- aumentar diseño, 288
- Autoguardar*, 73, 145
- average, 154
- AVG, 46
- barra de herramientas, 66
- barra de herramientas de análisis, 24
- bondad de ajuste, 246
- buscando los estadístico o test, 139
- BY variables, 137
- cabeceras de los análisis, 145
- Calculadora Seis Sigma*, 255
- camino de paso ascendente, 289
- capacidad de un proceso, 239
- coeficientes de regresión, 278
- columna de datos
 - comentario, 16, 35
 - nombre, 16, 35
 - tipo, 16, 35
- Comparación de dos muestras*, 175
- Comparación de varias muestras*, 188
- Configurar impresión*, 146
- Configurar página*, 76
- contrato de licencia, 4
- COUNT, 55
- C_p , 254
- C_{pk} , 252
- cuartil, 154
- curtosis, 154
- datos
 - accediendo, 36
 - borrar, 41
 - combinando columnas, 51
 - con patrones, 54
 - copiar, 41
 - cortar, 41
 - generando, 53
 - hoja, 14
 - insertar, 41
 - introduciendo, 14
 - nuevas variables, 41
 - ordenado, 48
 - pegar, 41
 - recodificar, 234
 - releerlo, 58
 - transformaciones, 45
- datos de atributos, 221
- desviación típica, 154
- diálogo de entrada de datos, 63, 67
- DIFF, 46

dígitos, 145
 dígitos significativos
 defecto, 144
 directorio de archivos
 temporales, 146
 diseños de cribado, 262
 diseños de experimentos, 257
 distribución mayor valor extremo, 246
 distribución normal, 154, 244
 DPM, 248, 252
 eliminar efectos, 277
 estudio de Medición R&R, 131
 Excel, 38, 39
Exclair, 75
 EXP, 46
 Expresiones Booleanas, 65
 extrapolar, 289
 FIRST, 64
 fórmulas
 conversión a puntuaciones Z, 46
 desviación típica, 46
 diferencia hacia atrás, 46
 función exponencial, 46
 logaritmo en base 10, 46
 logaritmo natural, 46
 máximo, 46
 media, 46
 mínimo, 46
 paso para k periodos, 46
 raíz cuadrada, 46
 valor absoluto value, 46
 FTP, 114
Generar Datos, 47, 55
 grafica los residuos, 211
 gráfico ANOVA, 193
 gráfico cuantil-cuantil, 186
 gráfico de barras, 223, 231
 gráfico de caja y bigotes, 24, 156, 179, 197
 gráfico de capacidad, 243, 250
 gráfico de contorno, 280
 gráfico de contorno revestido, 286
 gráfico de cuantiles, 167, 184
 gráfico de efectos principales, 274
 gráfico de interacción, 275
 gráfico de malla, 287
 gráfico de medias, 194
 gráfico de mosaico, 231
 gráfico de Pareto, 274
 gráfico de Pareto estandarizado, 273
 gráfico de probabilidad, 250
 gráfico de rascacielos, 232
 gráfico de sectores, 223
 gráfico de superficie, 280
 gráfico de superficie de respuesta, 280
 gráfico de tolerancia, 174
 gráfico del cuadrado, 280
 gráfico del cubo, 280
 gráficos
 añadirse texto adicional, 94
 aparición del defecto, 146
 botones, 74
 copiando, 105
 escalado logarítmicas, 92
 excluya los puntos, 75
 fondo, 81
 fuente de marca, 92
 guardando en archivos de imagen, 105
 identificando puntos, 101
 modificando, 80
 rotar, 100
 rotar etiquetas de los ejes, 92
 títulos de los ejes, 91
 gráficos de residuos, 198
Gráficos de superficie y contorno, 217
 gráficos matriciales, 104, 203
Guardar Resultados, 73
 heteroscedasticidad, 199
 histograma, 243

histograma de frecuencias, 162
 histograma dual de frecuencias, 178
 imprimiendo
 análisis, 76
 cabecera, 77
 fondo, 77
 líneas vacías, 77
 márgenes, 77
 índices de capacidad, 252
 iniciar el programa, 8
 instalación, 1
 intervalos bootstrap, 169
 intervalos de confianza
 desviación típica, 168
 media, 168
 mediana, 169
 intervalos HSD, 195
 intervalos LSD, 195
 K, 254
 Kolmogorov-prueba Smirnov, 246
 LAG, 46
 LAST, 64
 lecturas sugeridas, 291
 libro de datos, 14, 33
Límites de tolerancia, 172
Límites estadísticos de tolerancia, 172
 línea de regresión, 209
 LOG, 46
 LOG10, 46
 LOWESS, 101, 204
 matriz de coeficientes de correlación, 205
 matriz de correlación, 269
 MAX, 46
 máximo, 154
 media, 154
 mediana, 154
 menú del sistema, 12
 menú Seis Sigma, 12, 144
 métodos del nonparametric
 prueba de rangos con signos, 170
 prueba Friedman, 196
 prueba Kolmogorov-Smirnov, 246
 prueba Kolmogorov-Smirnov test, 185
 prueba Kruskal-Wallis, 196
 prueba Mann-Whitney (Wilcoxon), 183
 MIN, 46
 mínimo, 154
 modelo no lineal, 209
Modificar Columna, 34
 muesca a la mediana, 157
nivel de calidad Sigma, 254
 nivel de confianza
 defecto, 144
 números aleatorios, 56
 ODBC, 40
Opciones de análisis, 67
Opciones de panel, 26, 71
Opciones Gráficas
 perfil, 146
Opciones Gráficas, 27
 diseño, 81
 Efectos 3D, 81
 ejes, 91
 etiquetas y leyendas, 94
 líneas, 85
 malla, 83
 puntos, 87
 relleno, 93
 título superior, 89
 operadores
 división, 46
 exponenciación, 46
 multiplicación, 46
 resta, 46
 suma, 46
 optimización, 284
 OR, 65
Ordenar datos, 48

ordenar nombres de variables, 145
orígenes de datos
 registro, 112
paneles, 65
parsimonia, 202
percentiles, 168
Preferencias, 110, 143
 AED, 162
 Capacidad, 253
 Estadísticas, 155
probabilidad acumulada, 166
Propiedades del libro de datos, 57
prueba de rangos con signos, 170
prueba F, 181
prueba Friedman, 196
prueba Kolmogorov-Smirnov, 185
prueba Kruskal-Wallis, 196
prueba Levene, 198
prueba Mann-Whitney (Wilcoxon), 183
prueba Shapiro-Wilks, 245
prueba t, 170, 182
pruebas de hipótesis
 coeficiente de correlación, 205
 comparando desviaciones típicas, 181
 comparando distribuciones, 185
 comparando medianas, 183
 comparando medias, 182
 comparando proporciones, 238
 comparando varias desviaciones típicas,
 198
 comparando varias medianas, 196
 comparando varias muestras, 192
 distribución normal, 245
 media, 170
 mediana, 170
 regresión, 208
 tabla de este tamaño, 233
 valores atípicos, 160
pruebas de múltiples rangos, 195
puntuación Z, 255
P-valor, 160
RANDOM, 64
R-cuadrado, 208, 210
Recodificar datos, 50
recodificar los datos, 50
regla de Sturges, 163
Regresión múltiple, 213
regresión paso a paso, 215
Regresión Simple, 63, 206
REP, 55
resaltando un gráfico de dispersión, 98
RESHAPE, 56
residuos, 198, 211
residuos estudentizados, 212
Resumen Estadístico, 23, 153, 177
RNORMAL, 57
ROWS, 64
SD, 46
Seis Sigma, 239
seleccionando análisis, 134
Seleccionar, 64
separando un gráfico de dispersión, 96, 191
sgcinstall.exe, 1
SQRT, 46
STANDARDIZE, 46
StatAdvisor
 defecto, 145
StatFolios
 guardando, 107
 guardar, 30
 publicando, 113
 rutina de inicio, 109, 112, 145
StatGallery, 250
 configurando, 117
 copiando gráficos, 119
 imprimiendo, 123
 modificando gráficos, 121
 solapando gráficos, 120

Statistics for Experimenters, 193

StatLink, 57, 112

StatPublish, 113

StatReporter, 125

copiando salidas a, 126

guardando, 127

modificando, 127

StatWizard, 129

tabla ANOVA, 273

tabla de doble entrada, 229

tablas de contingencia, 236

Tablas y Gráficos, 68

Tabulación, 222

Tabulación cruzada, 226

Tabulación de frecuencias, 165

test de Desviaciones extremas

estudentizadas, 160

test de Grubbs, 160

test de la chi-cuadrado, 233, 235

transformación Box-Cox, 249

transformaciones, 138

ubicación, 264

valor estudentizado, 159

valores atípicos, 158, 199

ventana de análisis, 22

XML, 38