

62  
2e1



**UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO**

**FACULTAD DE CIENCIAS**

**STATISTICA: Manual auxiliar para los cursos básicos de Probabilidad y Estadística**

**T E S I S**

QUE PARA OBTENER EL TITULO DE:

**A C T U A R I O**

P R E S E N T A N:

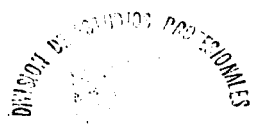
**MA. ANGELICA MENDOZA GOMEZ**

**MA. EUGENIA RIZO RIVERA**

**DIRECTOR DE TESIS:  
MAT. MARGARITA E. CHAVEZ CANO**



MEXICO D.F.



**FACULTAD DE CIENCIAS  
SECCION ESCOLAR**

1997

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL  
AVENIDA DE  
MEXICO

**M. en C. Virginia Abrín Batule**  
Jefe de la División de Estudios Profesionales de la  
Facultad de Ciencias  
Presente

**Comunicamos a usted que hemos revisado el trabajo de Tesis:**  
"Statistica: Manual auxiliar para los cursos básicos de Probabilidad y Estadística"

realizado por Mendoza Gómez María Angélica  
Rizo Rivera María Eugenia  
con número de cuenta 8938092-9 , pasante de la carrera de Actuaría  
9251650-4

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis	Mat. Margarita Elvira Chávez Cano
Propietario	M. en C. Beatriz Eugenia Rodríguez Fernández
Propietario	Dr. José Rodolfo Mendoza Blanco
Suplente	Mat. Norma Edith Alamilla López
Suplente	Mat. Vinicio Pérez Fonseca

*M. E. Abrin Batule*  
*Beatriz R. Rodríguez del.*  
*José Rodolfo Mendoza Blanco*  
*Norma Edith Alamilla López*

*V. Pérez Fonseca*

Consejo Departamental de Matemáticas  
Act. Académico Román Aguilar

FACULTAD DE CIENCIAS  
CONSEJO DEPARTAMENTAL  
DE  
MATEMÁTICAS

**Deseamos expresar nuestro más profundo y sincero agradecimiento a nuestra maestra y asesora: Mat. Margarita E. Chávez Cano; ya que sin su confianza, motivación, apoyo incondicional y prestancia inmediata no hubiera sido posible la realización de este trabajo.**

A la Universidad Nacional Autónoma de México, por habernos brindado la oportunidad de formar parte de su comunidad.

A la Facultad de Ciencias; en particular a todos los profesores que nos transmitieron sus valiosos conocimientos de manera abierta y generosa y nos dieron su apoyo durante nuestra formación.

Al personal del Centro de Cómputo de la Facultad de Derecho, particularmente al Ing. Jorge León Martínez por hacernos sentir como en casa y habernos permitido usar el equipo de cómputo para la elaboración de esta tesis.

A Mónica y Alvaro por aguantarnos tanto y quitarnos el tiempo. Gracias a su amistad pasamos momentos muy agradables .

A Dios, por haberme dado la vida y la certeza de su presencia en cada instante y circunstancia de ella. Por ser mi fortaleza y la luz que ilumina mi camino.

A mis padres: Fernando T. Mendoza y Ma. Dolores Gómez, quienes bajo cualquier circunstancia e independientemente de mi estado de ánimo, siempre me han brindado su apoyo y amor incondicionales. Gracias por su ejemplo de vida que es y será mi mejor herencia.

A mis hermanos Fernando e Isaac, por compartir conmigo éste y tantos otros momentos felices y tristes de la vida. Gracias por el cariño, apoyo y confianza que siempre me han demostrado. Por permanecer siempre unidos.

A la familia Gómez Ramírez, por la dicha de tenerlos a mi lado y contar siempre con su cariño.

A Maru, cuya amistad y apoyo hicieron muy agradable el camino que decidimos recorrer juntas.

A mis demás familiares y amigos con los que he compartido gratos momentos.

Finalmente, deseo agradecer a todas las personas que a lo largo de mi vida han contribuido directa o indirectamente a la realización de esta meta.

**Ma. Angélica Mendoza Gómez**

A mi madre: Elena Rivera, mi amor y eterno agradecimiento por su cariño, apoyo y confianza. Mi respeto y admiración por todos sus esfuerzos y sacrificios que ha hecho para darme la oportunidad de obtener una mejor educación académica.

A mi familia por estar siempre al cuidado de nosotras.

A mis amigos mi agradecimiento por su amistad y apoyo en todo momento sobre todo a Angélica por aguantarme todo este tiempo (igual que *Cerebro* aguanta a *Pinky*).

A todos aquellos que creen en mi y de quienes he recibido enseñanzas, consejos y cariño.

**Ma. Eugenia Rizo Rivera**

**La sabiduría es la fusión perfecta de la inteligencia y del amor.  
El amor sin la inteligencia puede hacer involuntariamente mucho daño  
y la inteligencia sin amor puede resultar en la crueldad refinada.**

**“... y aunque posea el don de la profecía  
y la ciencia de todos los misterios,  
teniendo aún la fe para mover montañas,  
si no tengo amor, no soy nada.”**

**I Corintios 13,2**





# INDICE

<b>INTRODUCCION</b> .....	v
<b>CAPITULO 1: FUNDAMENTOS</b> .....	1
<b>1.1 COMO INICIAR</b> .....	1
1.1.1 RECURSOS DE HARDWARE.....	1
1.1.2 RECURSOS DE SOFTWARE.....	1
1.1.3 INSTALACION DE STATISTICA.....	1
1.1.4 INICIANDO STATISTICA.....	2
1.1.5 PANTALLA DE INICIO.....	3
<b>1.2 CONVENCIONES GENERALES</b> .....	4
1.2.1 EL SPREADSHEET O EDITOR DE DATOS.....	4
1.2.2 EL SCROLLSHEET.....	5
1.2.3 VENTANA TEXT/OUTPUT.....	5
1.2.4 TIPOS DE ARCHIVOS.....	5
1.2.5 ABRIR Y GUARDAR ARCHIVOS.....	7
1.2.6 IMPRESION.....	7
<b>CAPITULO 2: MANEJO DE DATOS</b> .....	9
<b>2.1 DESCRIPCION GENERAL</b> .....	9
2.1.1 VARIABLES CLASIFICADORAS.....	10
2.1.2 VARIABLES CUANTITATIVAS.....	11
2.1.3 SELECCION DE VARIABLES.....	11
2.1.4 VARIABLES DE PESO.....	13
2.1.5 DATOS FALTANTES.....	14
2.1.6 SELECCION DE CASOS.....	14
<b>2.2 OPCIONES DEL MODULO</b> .....	16
2.2.1 CREAR UN NUEVO ARCHIVO DE DATOS.....	17
2.2.2 IMPORTAR UN ARCHIVO DE DATOS.....	19
2.2.3 EXPORTAR A OTRO FORMATO DE ARCHIVO.....	22
2.2.4 UNIR DOS ARCHIVOS DE DATOS.....	25
2.2.5 CREAR UN SUBCONJUNTO DE UN ARCHIVO DE DATOS.....	28
2.2.6 ORDENAR LOS CASOS.....	29
2.2.7 MODIFICAR VARIABLES.....	30
2.2.8 MODIFICAR CASOS.....	41
2.2.9 VERIFICAR NOMBRES Y FORMATOS DE LAS VARIABLES.....	43



2.2.10 VERIFICAR VALORES DE DATOS.....	44
2.2.11 ESTANDARIZAR VARIABLES; REEMPLAZAR DATOS FALTANTES POR MEDIAS.....	46
<b>CAPITULO 3: MENUS PRINCIPALES.....</b>	<b>49</b>
<b>3.1 SPREADSHEET.....</b>	<b>49</b>
3.1.1 FILE.....	49
3.1.2 EDIT.....	56
3.1.3 VIEW.....	60
3.1.4 ANALYSIS.....	63
3.1.5 OPTIONS.....	64
3.1.6 MACRO.....	66
3.1.7 WINDOW.....	68
3.1.8 HELP.....	69
3.1.9 BARRA DE HERRAMIENTAS.....	69
<b>3.2 SCROLLSHEET.....</b>	<b>70</b>
3.2.1 FILE.....	70
3.2.2 EDIT.....	72
3.2.3 VIEW.....	74
3.2.4 BARRA DE HERRAMIENTAS.....	75
<b>3.3 VENTANA TEXT/OUTPUT.....</b>	<b>75</b>
3.3.1 FILE.....	75
3.3.2 EDIT.....	76
3.3.3 OPTIONS.....	76
3.3.4 BARRA DE HERRAMIENTAS.....	76
<b>CAPITULO 4: GRAFICAS.....</b>	<b>77</b>
<b>4.1 CLASIFICACION GENERAL.....</b>	<b>77</b>
4.1.1 GRAFICAS PERSONALIZADAS.....	77
4.1.2 GRAFICAS ESTADISTICAS SG.....	82
4.1.3 GRAFICAS ESTADISTICAS RAPIDAS QSG.....	83
4.1.4 OTRAS GRAFICAS ESPECIALIZADAS.....	87
<b>4.2 ELEMENTOS PRINCIPALES.....</b>	<b>88</b>
<b>4.3 MENUS Y BARRA DE HERRAMIENTAS.....</b>	<b>94</b>
4.3.1 FILE.....	95
4.3.2 EDIT.....	97
4.3.3 VIEW.....	102
4.3.4 LAYOUTS.....	104
4.3.5 GRAPHS.....	106
4.3.6 OPTIONS.....	107
4.3.7 BARRA DE HERRAMIENTAS.....	108

<b>CAPITULO 5: REGRESION LINEAL</b> .....	109
<b>5.1 DESCRIPCION GENERAL</b> .....	109
<b>5.2 OPCIONES DEL MODULO</b> .....	111
5.2.1 PANEL DE INICIO.....	111
5.2.2 REVISAR ESTADISTICAS DESCRIPTIVAS.....	115
5.2.3 DEFINICION DEL MODELO.....	117
5.2.4 RESULTADOS DE REGRESION MULTIPLE.....	120
<b>5.3 EJEMPLOS</b> .....	127
5.3.1 EJEMPLO 1.....	128
5.3.2 EJEMPLO 2.....	134
<b>CAPITULO 6: ESTADISTICA NO PARAMETRICA</b> .....	139
<b>6.1 DESCRIPCION GENERAL</b> .....	139
<b>6.2 OPCIONES DEL MODULO</b> .....	141
6.2.1 PANEL DE INICIO.....	141
6.2.2 TABLAS DE $2 \times 2$ $\chi^2/N^2$ /PHI, McNEMAR, PROBABILIDAD EXACTA DE FISHER.....	144
6.2.3 PRUEBA $\chi^2$ PARA UNA MUESTRA.....	146
6.2.4 COEFICIENTES DE CORRELACION.....	148
6.2.5 PRUEBA DE RACHAS DE WALD-WOLFOWITZ PRUEBA U DE MANN WHITNEY Y PRUEBA DE DOS MUESTRAS DE KOLMOGOROV-SMIRNOV.....	151
6.2.6 ANALISIS DE VARIANZA DE KRUSKAL-WALLIS, PRUEBA DE LA MEDIANA.....	155
6.2.7 PRUEBA DE LOS SIGNOS.....	159
6.2.8 PRUEBA DE WILCOXON.....	160
6.2.9 ANALISIS DE VARIANZA DE FRIEDMAN Y COEFICIENTE DE CONCORDANCIA DE KENDALL.....	162
6.2.10 PRUEBA Q DE COCHRAN.....	165
6.2.11 ESTADISTICAS DESCRIPTIVAS ORDINALES.....	167
<b>BIBLIOGRAFIA</b> .....	171
<b>TECLAS DEL METODO ABREVIADO</b> .....	173
<b>INDICE DE ILUSTRACIONES</b> .....	175

# INTRODUCCION

En la actualidad la estadística ha tenido un gran desarrollo debido en gran parte a que las computadoras han evolucionado de forma sorprendente en los últimos años y su capacidad permite hacer cálculos muy complicados o trabajar con archivos de datos enormes. Diversas empresas se han dedicado a desarrollar software para la solución de problemas estadísticos específicos existiendo hoy en día una amplia variedad de paquetes en el mercado; de entre ellos destaca STATISTICA, ya que es una poderosa herramienta para procesar, analizar y desplegar información.

Este paquete fue desarrollado por la empresa StatSoft Inc., la primera versión llamada CSS:STATISTICA (Complete Statistical System) salió en 1986 y corría en todas las computadoras personales compatibles con IBM y con sistema operativo DOS Y OS/2. En 1993 quedó lista la versión 4.2 llamada simplemente STATISTICA, diseñada para correr bajo ambiente WINDOWS. Esta versión aprovecha todas las facilidades que el ambiente gráfico ofrece para dar al usuario una interfaz que le permita familiarizarse rápidamente con el programa y explotar todas sus características.

STATISTICA combina el manejo de bases de datos, análisis estadístico, generación de gráficas y reportes en un solo sistema de alto desempeño con múltiples aplicaciones en la ciencia, ingeniería y economía.

El sistema está integrado por un conjunto de módulos independientes cada uno de los cuales está diseñado para llevar a cabo un procedimiento específico de análisis estadístico. La selección de facilidades para el manejo de datos incluye prácticamente todos los procedimientos y técnicas disponibles en los sistemas manejadores de bases de datos que existen en el mercado, además de un buen número de características propias para facilitar el análisis estadístico y gráfico.

STATISTICA puede tomar los datos directamente vía "spreadsheet" (editándolos en pantalla) o importarlos de otros archivos incluso con formatos distintos ya que cuenta con una interfaz que permite convertir archivos de LOTUS, EXCEL, D'BASE, DIF, SYLK, etc.

Prácticamente no hay límites en cuanto al tamaño de los archivos de datos, cada registro puede contener un nombre y hasta 300 variables "dobles", es decir, variables que contienen simultáneamente valores numéricos y alfanuméricos.

Una de las características que distinguen a STATISTICA de otros paquetes con el mismo propósito es que ofrece facilidades para manejar, editar y controlar directamente los datos en cada paso del análisis, inclusive a la mitad de éste. Dondequiera que sea creada una gráfica el usuario puede acceder y editar los datos que la generan e interactivamente examinar los cambios que resultan.



Las opciones a elegir aparecen directamente en pantalla o en menús desplegables escritos en un inglés muy sencillo, además de que la mayoría de las veces, una explicación más extensa de la opción a utilizar aparece junto a ésta.

El uso del ratón facilita en gran medida todas las operaciones, pues si antes se requería pulsar una cierta combinación de teclas para llevar a cabo una tarea, ahora simplemente con un clic o doble clic en uno de los botones del ratón apuntando en el lugar indicado conduce directamente al resultado deseado.

STATISTICA ofrece una buena cantidad de procedimientos para el análisis, desde estadísticas básicas de propósito general hasta aplicaciones muy especializadas para la investigación en laboratorios científicos, mercadotecnia, economía e ingeniería.

Debido a que STATISTICA fue pensado como una herramienta de análisis de datos para todo propósito, muchas características del paquete sólo son necesarias para problemas complejos.

Este trabajo ha sido realizado para satisfacer la necesidad de un manual de referencia accesible y sencillo para utilizar el paquete en los módulos necesarios para dar apoyo a los cursos básicos de probabilidad y estadística que se imparten en la Facultad de Ciencias. Cuando en este manual se describe una instrucción en particular, se encontrará toda la información referente a ella y, aunque se presentan los objetivos y características de las técnicas tratadas por cada módulo y se buscó que los ejemplos cumplieran los supuestos y presentaran resultados coherentes, no se profundiza en ellas, ya que cada técnica podría ser por sí misma tema para un libro.

Por lo tanto, se requiere que los lectores de este manual y usuarios de STATISTICA cuenten al menos con conocimientos estadísticos básicos para poder apreciar la capacidad y resultados generados por cada módulo, ya que, una vez que se aprende a usar el paquete, lo más importante será la interpretación de sus resultados. Para profundizar en alguna o todas las técnicas estadísticas, se recomienda consultar la amplia bibliografía existente.

Esperamos que en un futuro próximo este manual se complemente para que otros cursos y usuarios diversos tengan el apoyo que se espera.

## **ORGANIZACION DEL MANUAL**

En la primera parte de este trabajo se describe la forma de iniciar el programa y las características generales de éste que será necesario recordar durante la ejecución de cualquier módulo.

Cada uno de los capítulos 2, 5 y 6 están dedicados a la descripción de un módulo de STATISTICA con base en el siguiente esquema:

- Descripción general: Contiene una breve introducción a los métodos de análisis de datos incluidos en el módulo respectivo, sus aplicaciones y su relación con otros métodos.



- **Opciones:** Se da una referencia completa de todas las opciones del módulo, organizada jerárquicamente de acuerdo a como son presentadas en pantalla las ventanas con dichas opciones.
- **Ejemplos de aplicaciones típicas del módulo paso por paso.**

En el capítulo tres se describen todos los menús de comandos disponibles en cualquier módulo siguiendo el orden en el que aparecen en la barra de menús.

Debido a la gran variedad de gráficas que STATISTICA puede crear, fue necesario dedicar el capítulo 4 a este tema. Aunque la mayoría de las opciones para crear gráficas se encuentran disponibles en la barra de menús que se describe en el capítulo 3; en éste se da una referencia general de todos los tipos de gráficas que se pueden manejar.

## CONVENCIONES USADAS EN EL MANUAL

Para facilitar la lectura, se usaron las siguientes convenciones en el libro:

**Hacer clic:** Presionar y soltar rápidamente el botón izquierdo del ratón.

**Hacer doble clic:** Presionar y soltar dos veces en forma consecutiva y rápidamente el botón izquierdo del ratón.

**Arrastrar:** Hacer clic y , manteniendo oprimido el botón izquierdo del ratón, moverlo hasta seleccionar el dato u objeto deseado; entonces soltar el botón.

**Tecla1 + Tecla2:** Es una combinación de teclas que indica que se debe pulsar la tecla1 y sin soltarla, inmediatamente presionar también la tecla2 para ejecutar una acción. Por ejemplo: Para salir del paquete se debe presionar la combinación: Alt + F4.

Los nombres de teclas específicas aparecen encerradas entre los signos < >, v.g. <Control>, <Tab>, etc.



# CAPITULO 1

## FUNDAMENTOS

### 1.1 COMO INICIAR

Este capítulo contiene instrucciones, paso a paso, para iniciar con STATISTICA. Como con cualquier programa, existe un hardware y un software específicos necesarios para correrlo.

#### 1.1.1 RECURSOS DE HARDWARE

Para usar STATISTICA 4.2 para Windows se necesita:

- Cualquier PC compatible con IBM, con un procesador 80386 o superior.
- Un disco duro con espacio libre de al menos 13 megabytes.
- Una tarjeta gráfica compatible con la versión 3.1 o posterior de Windows, tal como EGA o VGA.
- Por lo menos 4 megabytes de memoria RAM.
- Un ratón.
- Una impresora (opcional).

#### 1.1.2 RECURSOS DE SOFTWARE

La versión 3.1 o posterior de MS-DOS y la versión 3.1 o posterior de Windows.

#### 1.1.3 INSTALACION DE STATISTICA

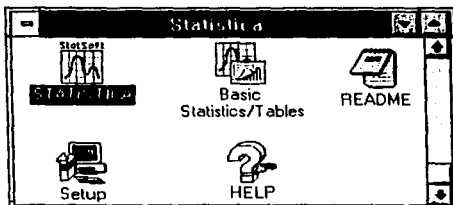
Para instalar STATISTICA 4.2 para Windows:

- 1.- Iniciar Windows.
- 2.- Insertar el disco No. 1 de STATISTICA (STATISTICA SETUP) en la unidad A (o B).
- 3.- Elegir la opción *Ejecutar (Run)* del menú *Archivo (File)* en el *Administrador de Programas (File Manager)*.
- 4.- En el cuadro de texto escribir A:SETUP (o B:SETUP) y presionar <ENTER>.
- 5.- Seguir las indicaciones de instalación del programa que aparecen en la pantalla.

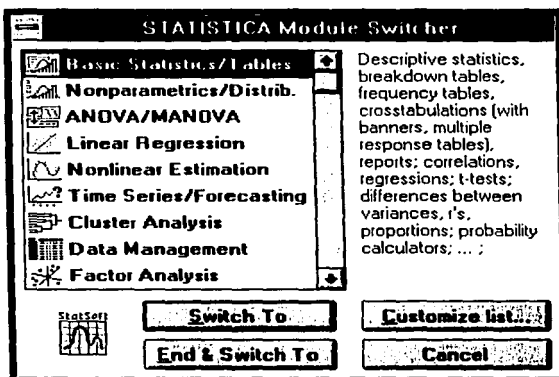


### 1.1.4 INICIANDO STATISTICA

STATISTICA se inicia desde Windows.



Para entrar al programa, en el grupo de programas STATISTICA se debe hacer doble clic en el icono STATISTICA, éste desplegará el menú principal del programa (*STATISTICA Module Switcher*) conteniendo el nombre de cada uno de los módulos independientes que integran el paquete:



Al seleccionar un módulo, una breve descripción de los procedimientos que realiza aparece junto al menú. Esta ventana además, puede ser llamada desde cualquier módulo del paquete haciendo doble clic en cualquier punto del área en blanco de la ventana de trabajo o presionando el siguiente botón de la barra de herramientas:



(En la sección denominada: Pantalla de Inicio, se describe la barra de herramientas).

Se puede acceder directamente a cualquier módulo haciendo doble clic en el área donde se despliega su nombre.

El menú principal (*STATISTICA Module Switcher*) contiene además las siguientes opciones:



**Switch to:** Ejecuta el módulo que se ha seleccionado. Si se estaba trabajando en otro módulo, permanecerá abierto, pudiendo regresar a él en cualquier momento seleccionando su ventana correspondiente.

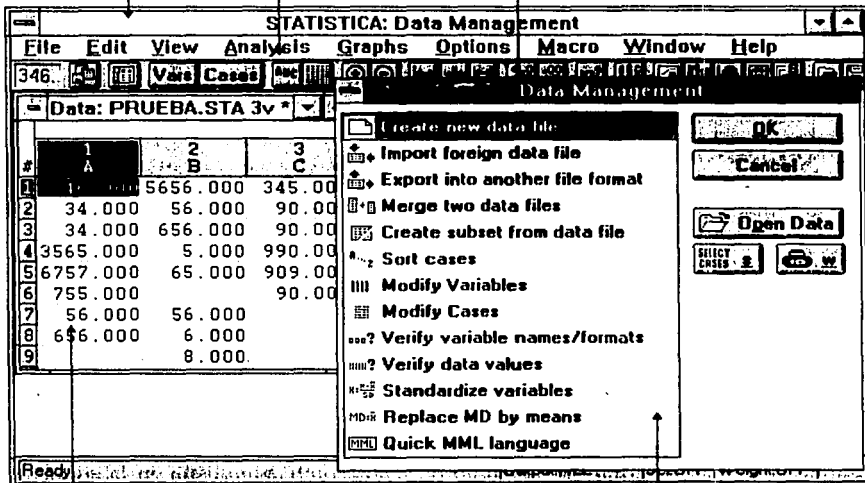
**End & Switch to:** Cierra el módulo que se estaba usando y ejecuta el que se ha seleccionado en el menú.

**Customize list:** Este botón abrirá la ventana *Customize list of modules*, en la cual se puede personalizar (reordenar, ampliar o reducir) la lista de módulos. Esta ventana permite reemplazar la lista entera de módulos por los que se han seleccionado o agregar los módulos seleccionados a la lista que ya se tiene.

### 1.1.5 PANTALLA DE INICIO

Como STATISTICA corre bajo Windows, la interfaz del usuario tiene los mismos elementos como la de cualquier otro programa de Windows, por lo que al iniciar cualquier módulo se debe exhibir una ventana similar a la que se muestra en la siguiente figura:

barra de título    barra de menú    barra de herramientas



spreadsheet

barra de estado

panel de inicio



**BARRA DE TITULO:** Muestra el título de la ventana, éste está formado por el nombre del programa y el módulo que se está manejando.

**BARRA DE MENUS:** Contiene los nombres de los menús de comando disponibles. Las opciones de cada uno de estos menús varían de acuerdo a la herramienta que se esté utilizando (spreadsheet, scrollsheet, ventana text/output, gráficas. Consultar la sección: Convenciones Generales para obtener información sobre estas herramientas). Si está disponible una combinación de teclas equivalente a los comandos, ésta se muestra a la derecha del comando.

**BARRA DE HERRAMIENTAS:** Contiene algunas de las opciones más utilizadas del menú para que puedan ejecutarse directamente con un clic del ratón.

**BARRA DE ESTADO:** Exhibe mensajes acerca de los menús (al momento de seleccionarlos) y del estado general del archivo de datos activo. Mientras se ejecute una tarea, *STATISTICA* la usará para indicar el progreso de ésta. (Una descripción mas completa de la barra de estado se da en la opción *Status bar* del menú *View* en el capítulo 3).

**SPREADSHEET:** Despliega el contenido del archivo de datos que se va a analizar. Al iniciar la sesión mostrará el último archivo de datos utilizado. (En la sección Convenciones Generales se dan otras características importantes de esta ventana).

**PANEL DE INICIO:** Despliega un menú con las opciones de análisis disponibles en el módulo activo. (Para obtener más información, consultar la opción *Startup panel* del menú *Analysis* en el capítulo 3).

## 1.2 CONVENCIONES GENERALES

*STATISTICA* tiene integradas tres herramientas muy importantes que pueden ser utilizadas en cualquier módulo del paquete.

### 1.2.1 EL SPREADSHEET O EDITOR DE DATOS

Es la ventana más importante de *STATISTICA*, puesto que proporciona el acceso directo a los datos que se van a analizar; además de que ahí se pueden llevar a cabo todas las operaciones de edición y transformación que se quieran realizar sobre los mismos.

Como ya se ha dicho anteriormente, al iniciar cualquier módulo siempre aparecerá esta ventana abierta con un determinado conjunto de datos como se muestra en la siguiente figura:



Data MOTHERS STA 20v 1.0						
Ranking of mothers by 13 judges						
Case	1 MOTH 1	2 MOTH 2	3 MOTH 3	4 MOTH 4	5 MOTH 5	6 MOTH 6
H. Smith		2.000	3.000	4.000	5.000	6.000
B. Jones	5.000	1.000	16.000	8.000	9.000	2.000
T. Hill	3.000	2.000	7.000	5.000	14.000	9.000
W. Miller	8.000	3.000	10.000	11.000	4.000	2.000
B. Robertson	2.000	1.000	15.000	8.000	14.000	4.000
C. Hansson	16.000	17.000	5.000	13.000	15.000	11.000
S. Mann	12.000	9.000	14.000	6.000	7.000	2.000
F. Jordan	11.000	2.000	13.000	10.000	7.000	3.000
B. Frank	9.000	2.000	15.000	6.000	5.000	7.000

## 1.2.2 EL SCROLLSHEET

Es un editor similar al spreadsheet, con la salvedad de que no trabaja directamente con el archivo de datos, puede contener los mismos datos que el spreadsheet pero las modificaciones que se hagan a éstos no afectarán al archivo original (desplegado en el spreadsheet), además, es el medio que STATISTICA emplea para desplegar las tablas de datos que resultan de los análisis realizados.

Nota: De aquí en adelante para nombrar a estas dos herramientas se utilizarán los términos en inglés (spreadsheet y scrollsheets).

## 1.2.3 VENTANA TEXT/OUTPUT

Toda la información generada en un análisis será dirigida a esta ventana (siempre que se encuentre activa) al igual que todo lo que se quiera imprimir. En ella se puede editar texto, así como agregar alguna información adicional al análisis que se acaba de realizar (ya sea de STATISTICA o cualquier otra aplicación de Windows). Al correr múltiples análisis toda la información nueva que se genere puede ser enviada a esta ventana, agregándose automáticamente al final del archivo de texto asociado (con extensión \*.txt) que almacena la información que la ventana despliega.

Así, al abrir esta ventana, toda la salida es direccionada a ella y de ahí se puede elegir si se manda directamente a la impresora o a un archivo en disco, mediante las opciones de impresión. También es factible usar esta utilidad como un block para guardar registros y notas de lo que se quiera.

## 1.2.4 TIPOS DE ARCHIVOS

STATISTICA utiliza diferentes extensiones de archivo para designar el tipo de información que contienen.

Dondequiera que se abra o guarde un archivo de STATISTICA se tiene la opción de aceptar la extensión de archivo por omisión o usar una diferente. A continuación se



proporciona una lista con las diferentes extensiones que STATISTICA maneja indicando el tipo de archivo al que pertenecen.

**Archivos de STATISTICA:**

**STA** Spreadsheets (archivos de datos del sistema).

**SCR** Scrollsheets.

**STG** Archivos de gráficas.

**TXT** Archivos de texto (para salida e impresión).

**SEL** Archivos de condiciones de selección de casos.

**CSM** Macro archivos.

**MFM** Archivos de Megafile Manager.

**QML** Archivos de programas en Quick MML.

**MML** Archivos de programas en MML.

**SCL** Archivos del lenguaje de comandos de STATISTICA (STATISTICA Command Lenguaje).

**HLP** Archivos de ayuda.

**INI** Archivos de configuración.

**PAL** Archivos de la paleta de colores.

**Archivos de otras aplicaciones:**

**XLS** Hoja de trabajo de Excel.

**WKS, WK1, WKR, WK3, WQI** Hoja de trabajo de Lotus.

**DBF** Archivo de Dbase.

**DB** Archivos de Paradox.

**POR** Archivos portables de SPSS.

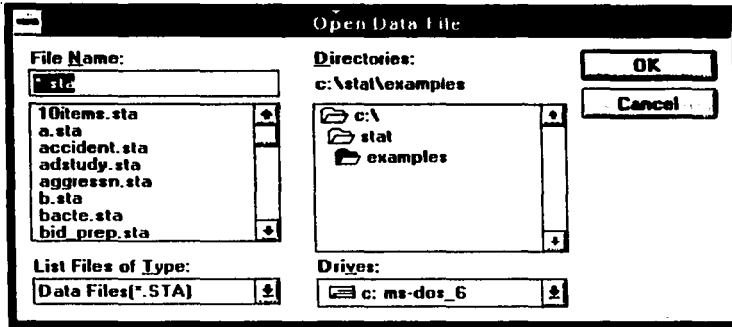
**FRE** Archivos en código ASCII (Formato de texto libre).

**FIX** Archivos en código ASCII (Formato de texto fijo).



## 1.2.5 ABRIR Y GUARDAR ARCHIVOS

Dentro de las diferentes opciones de cada módulo, con frecuencia se desplegarán ventanas que permiten abrir o guardar los diferentes tipos de archivos que STATISTICA maneja:



Las opciones que despliegan estas ventanas son las siguientes:

**File name:** Especificar el nombre del archivo que se desea abrir/guardar, seleccionándolo de la lista que aparece debajo de donde se pide éste.

**List files of type:** Permite especificar el tipo de archivos que se desean ver en la lista.

**Directories:** Indicar el directorio donde está localizado el archivo.

**Drive:** Especificar la unidad de disco donde está localizado el archivo.

## 1.2.6 IMPRESION

En STATISTICA es posible imprimir (o guardar en un archivo) los resultados de cualquier análisis realizado, de dos formas:

**Automáticamente:** Los contenidos de cada scrollsheet que se genere serán enviados directamente a la impresora (archivo) y/o a la ventana text/output.

**Manualmente:** El usuario elige los scrollsheets (o parte de ellos) que desea imprimir y el momento de imprimirlos.

Antes de imprimir los resultados de cualquier análisis es necesario indicar el tipo de salida deseado, es decir, se debe especificar si los resultados serán enviados a un archivo en disco, a la impresora o a la ventana text/output. Esta elección se lleva a cabo a través de la ventana *Page/Output Setup* (descrita en la sección File del capítulo 3) donde también es



posible elegir la cantidad de información suplementaria (*suppl. info*) que se desea imprimir junto con los scrollsheets.

Para llevar a cabo una impresión automática se debe elegir la opción *Automatically print all scrollsheets* en la ventana *Page/Output Setup*. Así, toda la salida generada en un análisis será enviada a la impresora (archivo en disco y/o ventana *text/output*) de forma automática. Esto es útil cuando se desean tener todos los resultados del análisis en el orden en que fueron generados.

La única diferencia entre impresión manual (la opción *Automatically print all scrollsheets está desactivada*) y automática es que, en la manual lo único que se imprime automáticamente es la información suplementaria (*suppl. info*) mientras que los scrollsheets (o sus partes) se imprimirán únicamente al elegir la opción *Print* del menú *File* (o su botón asociado).

Para dar una mayor rapidez al proceso, STATISTICA almacena internamente la salida de cualquier análisis en un buffer (memoria intermedia). El usuario puede controlar la forma en que esta información almacenada va a ser impresa mediante la opción *Automatically eject each filled page* de la ventana *Page/Output Setup* (descrita en la sección *File* del capítulo 3).

Todas las opciones de impresión se encuentran disponibles en el menú *File*, por lo que se describen con detalle en el capítulo 3: Menús Principales.



## CAPITULO 2

### MANEJO DE DATOS

#### 2.1 DESCRIPCION GENERAL

En este capítulo se introducirá la organización básica de los archivos de datos, así como las operaciones elementales para su manejo.

En STATISTICA los datos están organizados como una base de datos. Se puede pensar que cada archivo es una tabla donde los renglones representan registros y cada registro consta del mismo número de campos.

**Casos (registros):** Son los renglones en el archivo de datos y representan unidades elementales de análisis. Por ejemplo: pacientes, sujetos en el experimento, etc.

**Variables (campos):** Son las columnas en el archivo de datos. Cada variable contiene información específica aplicable a cada caso. En otras palabras, cada caso consta de un conjunto de valores, uno para cada variable definida en el archivo.

Ejemplo:

Supóngase que los casos representan pacientes, y para cada paciente se registran cuatro características: género, edad, educación y cantidad de colesterol.

Entonces se tiene un archivo de datos con cuatro variables.

Nombre del caso	Género	Edad	Educación	Colesterol
Juan Pérez	Masculino	21	Alto	190
Mary Blanco	Femenino	19	Medio	205
Adán Ruiz	Masculino	32	Alto	217
Julia López	Femenino	31	Bajo	231

Al introducir los datos en el spreadsheet el archivo queda de la siguiente forma:

Data: NONAME STA 4v * 4c				
Case	1 GENERO	2 EDAD	3 EDUCACION	4 COLESTEROL
Juan Pérez	M	21.000	ALTO	190.000
Mary Blanco	F	19.000	MEDIO	205.000
Adán Ruiz	M	32.000	ALTO	217.000
Julia López	F	31.000	BAJO	231.000



La estructura de los archivos de STATISTICA ofrece un buen número de características adicionales que al ser habilitadas, todos los procedimientos estadísticos y gráficos del paquete las aprovechan. Así se tiene que:

- A cada variable se le puede asignar un nombre corto (*short name*), un nombre largo, comentario o etiqueta (*long name*), un valor para datos faltantes (*missing data value*) y un formato de despliegue (*display format*).
- Cada valor de una variable puede llevar un comentario o etiqueta (*long label*).

Estas características se definen en la opción *Current specs* descrita en la sección Modificar Variables de este capítulo.

- Cada caso puede tener su propio nombre.
- Se puede incluir un comentario o nota como encabezado para identificar al archivo (*file header*).

STATISTICA soporta la doble notación para cada uno de los valores de las variables: numérico y alfanumérico (variables dobles). En el spreadsheet se puede cambiar de un formato a otro. Cuando se escriba un valor alfanumérico en una celda vacía del spreadsheet, el programa le asignará internamente un valor numérico (un entero consecutivo). Si dicha celda ya desplegaba un valor numérico al momento de editarla (con un alfanumérico) éste no se perderá y será asignado automáticamente como su valor numérico equivalente y todas las demás celdas (de esa variable) que tengan el mismo valor numérico tendrán su equivalente alfanumérico. Así, la correspondencia entre valores numéricos y alfanuméricos es manejada automáticamente por el programa al editar los datos en el spreadsheet.

nombre del caso                      variables                      título o encabezado (file header)

Case	1 GENERO	2 EDAD	3 EDUCACION	4 COLESTER
Juan Pérez	M	21.000	ALTO	190.000
Mary Blanco	F	Area de datos		205.000
Adán Ruiz	M			217.000
Julia López	F	31.000	BAJO	

## 2.1.1 VARIABLES CLASIFICADORAS

El archivo de datos del ejemplo anterior ilustra los diferentes tipos de datos que se pueden almacenar en las variables. En términos técnicos, las variables se diferencian dependiendo del tipo de escala de medición que sus valores representan. Por ejemplo, la variable género no contiene mediciones "reales" (como la cantidad de colesterol) sino simples





códigos o nombres de categorías (Masculino, Femenino) que indican a qué grupo pertenece cada uno de los casos.

En el ejemplo, el primer caso pertenece al grupo Masculino y el segundo al Femenino. La variable educación (Alta, Media, Baja) tiene exactamente las mismas características.

Este tipo de variables son llamadas variables clasificadoras o categóricas porque permiten identificar los casos que pertenecen a cada grupo para poder posteriormente realizar análisis donde los grupos van a ser comparados. Técnicamente son variables cualitativas discretas que toman valores de acuerdo a una escala de medición nominal u ordinal.

En este ejemplo, ambas variables clasificadoras (género, educación) contienen valores alfanuméricos, pero sus valores también podrán ser numéricos (v.g. 1=Masculino, 2=Femenino). En efecto, debido a la doble notación de variables en STATISTICA, ambos valores (numérico y texto) pueden ser usados de forma intercambiable para referenciar valores en la misma variable.

### 2.1.2 VARIABLES CUANTITATIVAS

Las otras variables del archivo de datos del ejemplo (edad y colesterol), contienen valores cuantitativos. A diferencia de las variables clasificadoras, éstas son las típicas variables que pueden usarse para calcular medias, desviaciones estándar y otras estadísticas. Por ejemplo, en un análisis muy simple se podrían comparar las cantidades de colesterol en hombres y mujeres. Para realizar esto, el programa primero identificaría a hombres y mujeres en el conjunto de datos (basado en los valores de la variable clasificadora género) y calcularía la media de la variable colesterol en cada grupo.

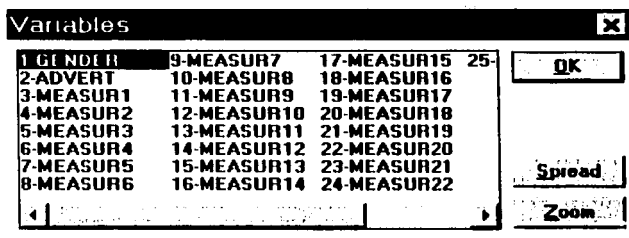
La diferencia entre estos dos tipos de variables radica únicamente en la manera en que se instruye al programa para usarlas e interpretarlas. Además, las variables cuantitativas también pueden usarse para indicar pertenencia a un determinado grupo. La forma de hacer esto es recodificándolas usando transformaciones de datos (ver opción *Recode*). Por ejemplo, basándose en los valores de la cantidad de colesterol, se podría crear una nueva variable (v.g. nivel) en la cual el valor Alto sería asignado a cada caso donde la cantidad de colesterol se encuentra por encima de 200 y Bajo si es menor e igual que 200.

### 2.1.3 SELECCION DE VARIABLES

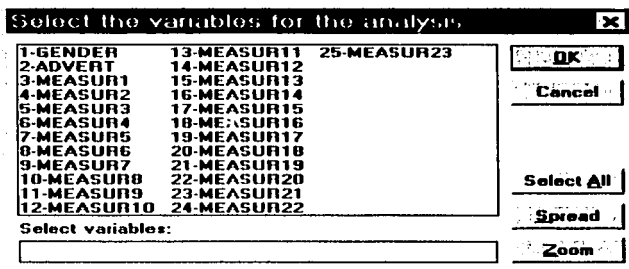
Siempre que se requiera seleccionar una o más variables para un análisis, STATISTICA despliega una ventana para llevar a cabo dicha selección.

Existen tres tipos de ventanas para la selección de variables dependiendo del análisis que se quiere realizar:

1. *Selección de una sola variable de una lista.* En este tipo de ventana, la selección se lleva a cabo haciendo doble clic en el nombre de la variable deseada que aparece en la lista.



2. *Selección de varias variables de una lista.* En esta ventana se puede elegir más de una variable de la lista. La selección se lleva a cabo resaltando sus nombres o escribiendo su número en el cuadro de texto abajo de la lista.



Para elegir un bloque continuo de variables se debe arrastrar el ratón desde la primera hasta la última variable deseada dentro de la lista, o bien hacer clic en la primera y presionando la tecla <Shift> hacer clic en la última.

Si se desean seleccionar bloques discontinuos se debe hacer clic en cada variable deseada presionando simultáneamente la tecla <Control>. Este mismo procedimiento sobre las variables ya elegidas, anula la selección.

3. *Seleccionar variables de varias listas.* En este caso aparecen varias listas de variables, pudiendo elegir una o varias de cada lista. El procedimiento es el mismo que el anterior.

Cabe señalarse que la mayoría de las veces (cuando es posible) al resaltar un bloque de datos en el spreadsheet y abrir una ventana de selección de variables por primera vez, las variables cuyos datos están resaltados aparecen "seleccionadas".



Select one or two variable lists


1-GENDER	12-MEASUR10	1-GENDER	12-MEASUR10
2-ADVERT	13-MEASUR11	2-ADVERT	13-MEASUR11
3-MEASUR1	14-MEASUR12	3-MEASUR1	14-MEASUR12
4-MEASUR2	15-MEASUR13	4-MEASUR2	15-MEASUR13
5-MEASUR3	16-MEASUR14	5-MEASUR3	16-MEASUR14
6-MEASUR4	17-MEASUR15	6-MEASUR4	17-MEASUR15
7-MEASUR5	18-MEASUR16	7-MEASUR5	18-MEASUR16
8-MEASUR6	19-MEASUR17	8-MEASUR6	19-MEASUR17
9-MEASUR7	20-MEASUR18	9-MEASUR7	20-MEASUR18
10-MEASUR8	21-MEASUR19	10-MEASUR8	21-MEASUR19
11-MEASUR9	22-MEASUR20	11-MEASUR9	22-MEASUR20

First variable list:   
 Second variable list (optional):

### 2.1.4 VARIABLES DE PESO

A veces se requiere que la contribución relativa de cada caso al resultado del análisis sea diferente. Esto se puede llevar a cabo dando un peso específico a cada caso, estos valores (pesos) se pueden guardar en una variable específica dentro del archivo de datos (variable de peso) para después poder indicarle al programa que le dé un peso a cada caso en proporción a los valores guardados en esa variable. A menos que se especifique otra cosa, los valores de la variable de peso son redondeados al entero más cercano y usados como "multiplicadores de casos". Es decir, cada caso es leído un número de veces equivalente al valor actual de su peso.

Esta opción se encuentra disponible:

- Haciendo clic en el campo WEIGHT de la barra de estado.
- Presionando el botón *Weight*: disponible en el panel de inicio y otras ventanas de análisis: 
- Seleccionando la opción *Weight* del menú *Options*.

Al seleccionarla se abre la siguiente ventana:

Define Weight

Weight variable:

Status:  ON  OFF

Double-click on edit window to choose from list of all variables. Values of selected variable will be used as case multipliers.



**Weight Variable:** Escribir el nombre de la variable que se va a utilizar como variable de peso.

**Status:** El estado de esta ventana se desplegará en la barra de estado indicando si la variable de peso está en uso (**on**) o no (**off**).

### 2.1.5 DATOS FALTANTES

Con frecuencia ocurre que el conjunto de datos no está completo y en algunos casos (renglones) no existen datos disponibles para algunas variables (v.g. no existe el dato de la cantidad de colesterol para uno de los pacientes), en el spreadsheet estas celdas aparecen en blanco. Estos datos faltantes (*missing data*) son marcados en el archivo de datos por un valor reservado que puede ser especificado separadamente para cada variable (por omisión es -9999). Los valores de una variable a los que se les asigna tal código de falta de dato son ignorados en todo el análisis. Cuando se crea un nuevo archivo de datos, inicialmente todos sus valores son nulos (técnicamente son llenados con código de falta de dato).

### 2.1.6 SELECCION DE CASOS

Por omisión STATISTICA supone que todos los casos (todos los renglones) del archivo de datos van a ser incluidos en el análisis. Sin embargo, a veces se requiere analizar únicamente un determinado subconjunto de casos.

Las condiciones para la selección de casos (*Case selection conditions*) son condiciones lógicas que determinan si un caso particular será incluido o no en el análisis respectivo. Por ejemplo, si se desea realizar un análisis únicamente para el sexo masculino, entonces, si la variable género entró como primera variable (V1) y sus códigos o valores son: Masculino=1, Femenino=2, se puede indicar al programa que incluya en el análisis solamente los casos que tienen el valor 1 en la variable 1.

INCLUDE IF:

V1=1

o

INCLUDE IF:

V1='MASCULINO'

El siguiente ejemplo excluirá del análisis todos los casos que sean hombres menores de 20 años.

Suponiendo que la variable edad es la número 10 se tiene:

EXCLUDE IF:

V1='MASCULINO' AND V10 < 20

Cabe resaltar que si se especifican condiciones que involucren valores alfanuméricos, la selección se basará en los valores numéricos subyacentes (más que en un orden




alfabético). Por ejemplo: Si se tiene una variable (V1) con valores: Carlos=1, Bruno=2, Alberto=3 y se especifica la condición:

INCLUDE IF:  
V1<'BRUNO'

Se seleccionarán los casos que tengan el valor Carlos (1) ya que Alberto (aunque A es primero que C) tiene el valor de 3.

Para poder llevar a cabo lo anterior, STATISTICA cuenta con la herramienta *Case selection conditions* cuya ventana se puede desplegar de diferentes maneras:

- Haciendo doble clic en la barra de estado (en el campo marcado con SEL).
- Presionando el botón *Select cases* que aparece en el panel de inicio o en algunas ventanas de análisis: 
- Mediante la opción *Select* del menú *Options*.

**Edit/enter selection conditions:** En este cuadro de texto se pueden introducir nuevas condiciones de selección o editar unas creadas previamente (ver botones *Open* y *Save as*, abajo).

**Include if:** Los casos que cumplan las condiciones lógicas establecidas serán incluidos en el análisis.

**Exclude if:** Los casos que cumplan con las condiciones lógicas serán excluidos del análisis.

Sintaxis:

Las variables se deben especificar ya sea por su número (v1,v2...) o por su nombre (género, edad,...), en este caso, las mayúsculas son equivalentes a las minúsculas ('SEXO'='sexo'). La columna que contiene el número de cada caso se identifica como v0.



Se deben usar comillas simples para hacer referencia a los valores alfanuméricos (v1='masculino' o v1='MASCULINO').

Se pueden usar paréntesis para especificar condiciones lógicas complejas o para cambiar la precedencia de los operadores.

A continuación se muestra la lista de operadores disponibles para la selección de casos:

=	igual
#, <>, ><	distinto
<	menor que
>	mayor que
<=	menor o igual
>=	mayor o igual
NOT, ~	negación
AND, &	y
OR,	o

**Header:** Introducir una línea (hasta 77 caracteres) de comentario acerca del archivo que se obtiene de la selección de casos (será guardada junto con las condiciones de selección).

**Status:** El estado de esta ventana se desplegará en la barra de estado indicando si las condiciones de selección se encuentran activas (on) o no (off).

**Variables:** Este botón abrirá la ventana *Variables* que permite revisar los nombres de todas las variables del archivo de datos así como los atributos generales de éstas.

**Open:** Este botón despliega la ventana *Open Selection Conditions File* que permite abrir un archivo de condiciones de selección de casos creado previamente.

**Save as:** Este botón permite guardar en un archivo las condiciones de selección de casos editadas en esta ventana para poder utilizarlas posteriormente (mediante el botón *Open*). El nombre del archivo sugerido por el programa es *archivo.sel*, donde *archivo* es el nombre del archivo de datos actual. Una ventaja de esta convención es que asocia el archivo de condiciones de selección con su respectivo archivo de datos; esto previene errores que podrían ocurrir cuando se tratan de usar condiciones de selección basadas en variables que no existen en el actual archivo de datos.

## 2.2 OPCIONES DEL MODULO

Entre las principales herramientas del módulo *Data Management* se encuentran la importación y exportación de archivos de una gran variedad de formatos estándar.

El spreadsheet tiene integradas herramientas para la edición y transformación de datos, ésta incluye no solo introducir o cambiar valores individuales; con frecuencia se necesita realizar operaciones más globales como: reordenar variables o casos, mover, copiar,



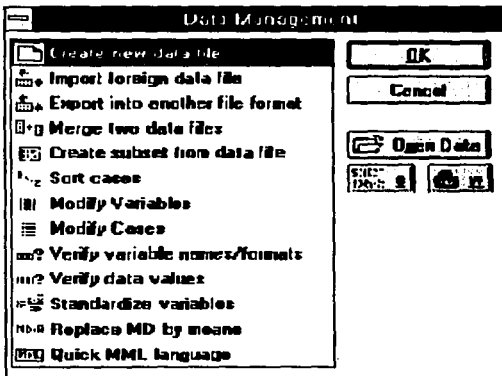
transponer bloques o rangos de datos, cambiar el tamaño de los archivos de datos ya sea por agregar o por borrar casos y variables; todo esto es llamado reestructuración de los datos.

La transformación de datos considera operaciones para definir variables por medio de fórmulas, verificar que los valores de las variables cumplan condiciones específicas, recalcular variables, definir rangos con base en los valores de una variable, etc.

El principal propósito de todas estas operaciones es facilitar la preparación de los archivos de datos para análisis específicos.

STATISTICA además cuenta con un lenguaje de programación especial para la transformación de datos: Quick MML, este lenguaje es similar a Basic o Pascal en términos de convenciones generales de sintaxis. Se recomienda consultar la ayuda en línea asociada a esta herramienta para un mejor aprovechamiento.

Como ya se mencionó en el Capítulo 1, al iniciar cualquier módulo se desplegará su correspondiente panel de inicio. En el módulo *Data Management*, el panel de inicio contiene un menú con las opciones de edición y transformación de datos. Para acceder a cualquier opción se debe hacer doble clic o bien, seleccionarla y presionar **OK**.

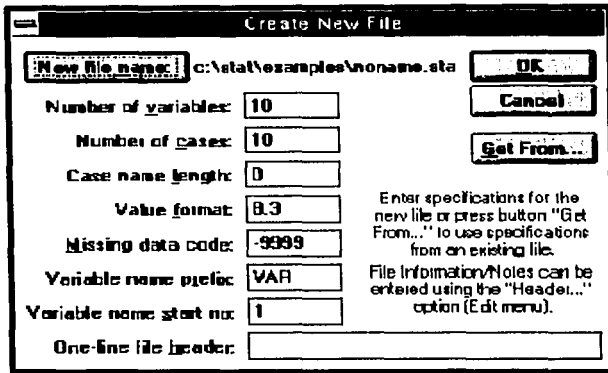


### 2.2.1 CREAR UN NUEVO ARCHIVO DE DATOS (CREATE NEW DATA FILE)

Los archivos de datos pueden ser creados por importación de otros formatos o ser introducidos directamente en el spreadsheet. Cada archivo de datos tiene un número fijo de variables y de casos (aunque puede ser cambiado fácilmente en el spreadsheet).

Una manera rápida y fácil para crear un archivo de datos se incluye en el menú *File* de todos los módulos; la ventaja de crearlo dentro del módulo *Data Management* es que ofrece opciones adicionales.

Todas las especificaciones del archivo que se determinen en esta ejecución podrán ser modificadas después de que el archivo se haya creado.



**Create New File**

New file name: c:\stat\examples\lname.sta

Number of variables: 10

Number of cases: 10

Case name length: 0

Value format: 8.3

Missing data code: -9999

Variable name prefix: VAR

Variable name start no: 1

One-line file header:

OK

Cancel

Get From...

Enter specifications for the new file or press button "Get From..." to use specifications from an existing file.

File Information/Notes can be entered using the "Header..." option (Edit menu).

**New file name:** Escribir el nombre para el nuevo archivo de datos. Como se muestra en la imagen, se debe incluir la ruta donde será creado el archivo. Por omisión el programa le asignará la extensión \*.STA y lo creará en el directorio C:\STAT\EXAMPLES.

**Number of variables:** Indicar el número de variables que serán incluidas en el nuevo archivo de datos. El límite es de 300 variables por archivo.

**Number of cases:** Indicar el número de casos que serán incluidos. STATISTICA prácticamente no impone límite en el número de casos en un archivo.

**Case name length :** Especificar la longitud de los nombres de los casos, hasta un máximo de 20 caracteres. Si la longitud varía de caso a caso entonces especificar la longitud del nombre más largo.

**Value format:** El formato de despliegue de valores se define de la forma X.Y donde X es la extensión total del número (puede ser de hasta 30 caracteres) y Y es el número de dígitos decimales (el formato por omisión es 8.3; 8 caracteres de extensión con 3 dígitos al lado derecho del punto decimal, v.g. 1234.567). El formato afecta sólo la manera en que se presentan los valores en el spreadsheet y no afecta la precisión almacenada del número.

**Missing data code:** Por omisión asigna el valor -9999 para representar falta de dato. Este valor puede ser cambiado para variables individuales mediante la opción *Current Specs* del menú *Edit Variables*, STATISTICA despliega en blanco la casilla en el spreadsheet siempre que falta un dato.





**Variable name prefix:** Definir el prefijo que llevarán por nombre todas las variables al que se le agregará un número secuencial al final (ver abajo). Los últimos caracteres del prefijo serán reemplazados por el número si es necesario para mantener la extensión total del nombre de 8 caracteres (var1, var2,...).

**Variable name start no:** Indicar el número inicial del prefijo de la variable. Se puede iniciar con un número distinto de 1 si se piensa unir el archivo a crear con otro.

**One-line file header:** Incluir una línea (77 caracteres máximo) de comentario que será salvada como parte del archivo y desplegada en el área de título del spreadsheet.

**Get from:** Al presionar este botón se obtendrá el conjunto de todas las especificaciones (número de variables y casos, nombres de variables, formatos y códigos de falta de datos) de un archivo de datos de STATISTICA ya existente y serán utilizadas en el archivo de datos por crear. Este botón abre la ventana *Select File to Get Specs From* en la que se debe indicar el nombre del archivo que contiene las especificaciones que se desean.

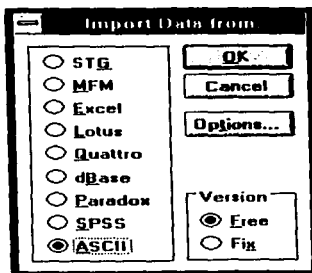
Una vez especificados todos los requerimientos de esta ventana, presionar el botón **OK** para desplegar el scrollsheet donde se editarán los datos del nuevo archivo.

## 2.2.2 IMPORTAR UN ARCHIVO DE DATOS (IMPORT FOREIGN DATA FILE)

La importación de archivos de datos es una de las formas con las que STATISTICA cuenta para tener acceso a la información creada en otras aplicaciones, otra forma sería mediante el establecimiento de ligas para el intercambio dinámico de los datos (ver opción *Paste link* descrita en la sección Edit del capítulo 3).

Al momento de importar, STATISTICA reconoce la estructura del archivo, las etiquetas de valores, las variables numéricas y alfanuméricas, las celdas en blanco y los convierte a sus equivalentes dentro del formato de archivo de STATISTICA.

Al seleccionar esta opción aparece primero la ventana *Select File to Import* donde se debe indicar el nombre del archivo a importar y después la siguiente:





Cada opción dentro del cuadro representa un grupo de formatos (v.g. con Lotus se pueden importar archivos de Lotus versiones 1A, 2, 3, Symphony, etc).

**STG:** Es el formato por omisión para los archivos gráficos de STATISTICA. Es útil cuando se desean importar gráficas desde CSS/3 (versión anterior a STATISTICA 4.2).

**MF:** Archivos de Mega File Manager (Mega File Manager es una herramienta disponible en la versión CSS/3 para DOS y sirve para manejar archivos de datos muy grandes).

**Excel:** Hojas de cálculo de Excel versión 4.0 y anteriores.

**Lotus:** Worksheets de Lotus 123 versión 1A, 2.01/2.2, 3, Lotus for Windows (ver 1 y 1.1) y Symphony.

**Quattro:** Worksheets de Quattro Pro versión para DOS.

**Dbase:** Archivos de Dbase versión III y IV.

**Paradox:** Archivos de Paradox.

**SPSS:** Archivos portables de SPSS, así como cualquier archivo que tenga formato PFF (Portable File Format).

**ASCII:** Archivos en código ASCII ya sea en versión libre (Free) o fija (Fix).

Cuando se importan archivos, STATISTICA tomará los primeros 8 caracteres de los nombres de las variables/columnas/campos del archivo fuente y los demás los ignorará.

**Options:** Este botón abrirá una ventana para elegir lo que se tomará en cuenta en la importación de datos dependiendo del formato elegido.

A continuación se explicarán las opciones de importación para los formatos STG, Excel, Lotus y SPSS. (Los demás formatos se pueden consultar en la ayuda en línea (*Help*)).

**STG:** Cada gráfica será importada como un par (o terna) de variables conteniendo los valores X,Y (o Z) de cada gráfica. Los nombres por omisión para estas variables (si la gráfica no tiene etiquetas) serán Plot 1X, Plot 1Y, (Plot 1Z), Plot 2X y Plot 2Y, (Plot 2Z), etc.

**Excel, Lotus y SPSS:** STATISTICA puede importar valores numéricos, alfanuméricos y tipo fecha desde estas aplicaciones. Si un valor alfanumérico contiene más de 8 caracteres, al importarlo STATISTICA lo trunca hasta 8 y lo guarda como el valor alfanumérico de una variable doble. Los valores tipo fecha de las variables de Excel y Lotus son importados en el formato numérico conocido como Julian (este formato despliega las fechas contando el número de días que han transcurrido desde el primero de enero de 1900; por ejemplo, 24858 representa el día 21 de enero de 1968).



Las variables de SPSS conservan sus nombres cortos y largos en STATISTICA (el nombre largo puede tener hasta 40 caracteres).

El botón *Options* contiene alternativas similares para los tres formatos:

**Import text columns:** STATISTICA convierte las columnas con valores alfanuméricos (del archivo de Excel, Lotus o SPSS) en variables dobles asignando a cada valor alfanumérico un número entero consecutivo; si no se selecciona esta opción no se importarán dichas columnas.

**Import case names:** Los nombres de los casos pueden ser importados de la primera columna del rango de datos especificado en el archivo de Excel y Lotus, o bien de la primera variable de SPSS.

**Automatic numeric/text match:** Si se elige esta opción, STATISTICA combinará columnas consecutivas del archivo a importar (de Excel o Lotus) para construir variables dobles de acuerdo a las siguientes reglas:

- Una columna con valores numéricos puede aparearse con una o dos columnas de texto consecutivas, las cuales se convertirán en los valores alfanuméricos y etiquetas de valores respectivamente para la variable doble de STATISTICA.
- Las tres columnas consecutivas (una numérica y dos alfanuméricas) deben llevar el mismo nombre de máximo 8 caracteres. Para la primera columna de valores alfanuméricos se debe sustituir el segundo carácter (si el nombre es de 8 caracteres) por un \$ o se debe agregar al final de éste (si es menor). Para la segunda columna se debe sustituir el segundo carácter (si el nombre es de 8 caracteres) por una @ o se debe agregar al final de éste (si es menor).

La correspondencia uno a uno entre los valores numéricos y alfanuméricos será revisada durante el proceso de creación del archivo en STATISTICA y si se encuentra más de un valor alfanumérico apareado con un valor numérico dado, entonces se emitirá un aviso.

**Range:** Por omisión, el área completa de la hoja de trabajo es importada. Esta opción permite seleccionar un rango específico de casos y variables a ser importado.

**Variables (Cases):** Especificar (**From**) a partir de qué variable (caso) y (**To**) hasta qué variable (caso) se va a importar.

**Sheets:** Esta opción está disponible únicamente para Lotus 1,2,3 versión 3. Sirve para seleccionar el grupo de hojas de trabajo que se importarán. (**From**) - (**To**)

**Review ranges:** Este botón abre la ventana *Review ranges* donde se puede seleccionar un rango determinado de la lista de todos los rangos en el archivo seleccionado.

**Variable names:** Especificar el origen (en Excel y Lotus) de los nombres de las variables de STATISTICA:



**From Spreadsheet:** STATISTICA usará los nombres de las columnas del archivo origen como los nombres de las variables del archivo importado.

**From first row of selected range:** STATISTICA usará las etiquetas del primer renglón del rango que se especificó para importar, tomando únicamente los primeros 8 caracteres.

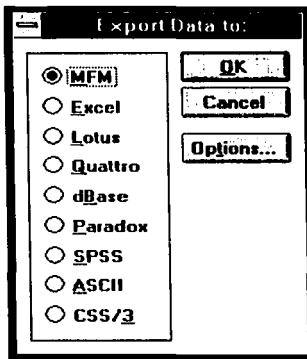
**Create (var1, var2,...):** Los nombres de las variables se formarán con un prefijo (var) y un número secuencial.

**Import value labels:** Al seleccionar esta opción STATISTICA importará las etiquetas de valores de SPSS (hasta 40 caracteres). Si una etiqueta de valor de SPSS contiene caracteres ilegales (" #&' ()+.,/;:><=[ ] \ { } ^~) entonces se le asigna un "\_".

### 2.2.3 EXPORTAR A OTRO FORMATO DE ARCHIVO (EXPORT INTO ANOTHER FILE FORMAT)

Esta opción permite exportar los resultados de análisis estadísticos y todas las transformaciones de datos a otras aplicaciones con una adaptación automática a sus formatos.

Las ventanas que se despliegan son similares a las que se utilizan para importar: Primero se debe elegir el nombre del archivo en la ventana *Select File to Export* y después seleccionar el formato al que se va a exportar dicho archivo:



**MFM:** Es el formato de los archivos que maneja Mega File Manager (ver opción anterior).

**Excel:** STATISTICA exporta archivos a cualquier versión de Excel.

**Lotus:** Exporta a Lotus 123 versión 1A, 2.01/2.2, 3, Lotus for Windows y Symphony.



**Quattro:** Exporta a cualquier versión de Quattro.

**Dbase:** Exporta a Dbase III+ y IV.

**Paradox:** Exporta a cualquier versión de Paradox.

**SPSS:** Exporta al formato portable de SPSS (Portable File Format).

**ASCII:** Exporta a código ASCII ya sea en versión libre (Free) o fija (Fix).

**CSS/3:** Exporta a las versiones de STATISTICA para DOS (CSS: STATISTICA o STATISTICA/DOS).

**Options:** Este botón abrirá una ventana para elegir lo que se tomará en cuenta para la exportación de datos, como es, el nombre de los casos, valores alfanuméricos, valores numéricos y etiquetas de valores.

A continuación se describirán las opciones de exportación que despliega el botón *Options* para los formatos Excel, SPSS, Lotus y CSS/3.

**Excel:** Los archivos de Excel pueden contener hasta 256 columnas y 16384 renglones. Si el archivo de STATISTICA es de mayor dimensión, se pueden crear subconjuntos (mediante la opción *Create subset from data file*) para exportarlos por separado.

**SPSS:** Cuando se exporta a SPSS, las etiquetas de valores de STATISTICA también serán etiquetas de valores para SPSS. Los nombres de las variables de STATISTICA deben adecuarse a los requerimientos de SPSS (para verificar esto es recomendable usar la opción *Verify names/formats* de este módulo).

**Ventana Excel (SPSS) Export Options:**

**Case names:** Seleccionar esta opción para exportar los nombres de los casos a la primera columna del archivo de Excel (SPSS).

**Variable names:** Si esta opción está activada, STATISTICA exportará los nombres de las variables al primer renglón del archivo de Excel, de lo contrario; los nombres de las variables no serán incluidos en el archivo.

**Variable with text values:** Las variables dobles de STATISTICA pueden ser expandidas para convertirse en 3 columnas de Excel (representando los valores numéricos, alfanuméricos y etiquetas de valores respectivamente).

Las siguientes opciones controlan la creación de estas columnas, permitiendo volver a comprimirlas para crear la variable original de STATISTICA (de la cual provienen) cuando el archivo sea importado de nuevo. Si todas las opciones están desactivadas, entonces sólo las variables que no tienen valores alfanuméricos serán exportadas.



**Text values:** Excel creará una columna con valores alfanuméricos para cada variable doble. El nombre de esta columna será el mismo que el de la variable doble excepto por el símbolo \$ que se agrega de la siguiente manera: Si el nombre de la variable de STATISTICA es más corto que 8 caracteres, entonces el nombre de la columna de texto en Excel tendrá un \$ al final; si el nombre es de 8 caracteres, entonces el segundo será cambiado por \$. Esto es para facilitar la recombinación de los valores numéricos y alfanuméricos si el archivo es importado de nuevo.

Para SPSS los valores alfanuméricos de las variables dobles de STATISTICA se conservan como simples variables alfanuméricas.

**Numeric values:** Excel creará columnas de valores numéricos asociados con las columnas alfanuméricas de la opción anterior. Si no se activa esta opción, entonces únicamente los valores numéricos de las variables simples (sin valores alfanuméricos) aparecerán en el archivo exportado.

Para SPSS los valores numéricos de las variables dobles de STATISTICA se conservan como simples variables numéricas.

**Long value labels:** Excel creará una tercera columna de valores alfanuméricos representando las etiquetas de valores de la variable. El nombre de esta columna será el mismo que el de la variable doble excepto por una @ que se agrega de la siguiente forma: Si el nombre de la variable de STATISTICA es más corto que 8 caracteres, entonces el nombre de la columna de texto en Excel tendrá una @ al final; si el nombre es de 8 caracteres, entonces el segundo será cambiado por @. Esto es para facilitar la recombinación de los valores numéricos y alfanuméricos si el archivo es importado de nuevo.

En SPSS las etiquetas de valores se conservan o pueden ser tomadas como variables alfanuméricas (ver abajo).

Es posible hacer combinaciones con las opciones anteriores:

**Text/Numeric:** Crea variables numéricas usando los valores alfanuméricos como sus etiquetas de valores.

**Text/labels:** Crea variables alfanuméricas con etiquetas de valores.

**Numeric/Labels:** Crea variables numéricas con etiquetas de valores.

Si se eligen las tres opciones al mismo tiempo (text/numeric/labels) se crearán dos columnas: Numérica y alfanumérica con etiquetas de valores. Si no se activa ninguna, entonces sólo aquellas variables de STATISTICA que no contengan valores alfanuméricos serán exportadas.

**Lotus:** Las opciones de exportación a Lotus 123 son las mismas que las descritas anteriormente para Excel.

Las de la versión 3 y para Windows se describen en seguida:

**How many:** Por omisión todas las variables en el archivo de STATISTICA son exportadas como una hoja de trabajo (worksheet) a Lotus. Pero si se desea, las variables pueden repartirse equitativamente en varias hojas especificando el número en esta opción.

**By names:** Las variables de STATISTICA pueden ser asignadas por nombres a hojas de trabajo específicas; es decir, desde la variable A-A hasta la variable A-IV serán colocadas en la hoja de trabajo A, las variables B-A hasta B-IV en la hoja B, etc.

**CSS/3:** Los archivos de CSS/3 soportan la mayoría de las especificaciones de formato de STATISTICA 4.2 y todas ellas son transferidas automáticamente al exportar salvo las siguientes excepciones:

Si el nombre largo de una variable tiene más de 40 caracteres de longitud será truncado.

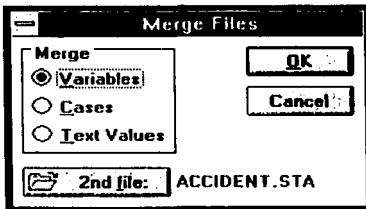
CSS/3 no soporta los comentarios y notas de archivo (file notes/comments) únicamente el título o encabezado (one-line file header).

La información referente al color, estilo y fuentes del spreadsheet no es exportada.

## 2.2.4 UNIR DOS ARCHIVOS DE DATOS (MERGE TWO DATA FILES)

Existen muchas circunstancias donde es necesario unir archivos. Por ejemplo, agregar datos de sujetos adicionales al final de un archivo de datos existente ( es decir, unir un archivo con otro que contiene las mismas variables. *Merge Cases*). Se pueden realizar más pruebas ( crear variables adicionales) a los casos de un archivo existente. Un archivo que contiene esas nuevas variables puede unirse al final (lado derecho) de un archivo existente ( es decir, unir un archivo con otro que contiene los mismos casos. *Merge Variables*). También se pueden unir los valores numéricos y alfanuméricos de dos archivos ( *Merge Text Values*).

Al seleccionar esta opción, se abrirá la siguiente ventana, donde se debe especificar el tipo de unión requerida así como el nombre del archivo que se va a unir con el actual.



**Merge:** Seleccionar una de las tres opciones para unir:



- **Variables:** Agrega las variables del segundo archivo después de las variables del primero.

**Merge Variables**

File I: c:\stat\examples\noname.sta

File II: c:\stat\examples\la.sta

**Mode**

**Not Relational**

**Relational**

**Relational Hierarchical**

**Unmatched Cases**

**Fill with MD**

**Delete**

**Abort Merge**

**Relational Merge Options**

1

1

**Key Comparison**

**Numeric**  **Text**

Variables from the second file will be appended after the last variable of the first file.

The two files should be sorted (by the respective key variables) before a relational merge can be performed.

Las opciones de esta ventana sirven para alinear los casos del segundo archivo con los del primero.

**Mode:** Si el número y orden de los casos en los dos archivos no es el mismo (es decir, si falta al menos un caso en el segundo archivo o no existe un caso equivalente en el primero) entonces, al hacer la unión podría haber casos con datos que no les corresponden. Para evitar este problema, se puede elegir una de las tres formas de unión que STATISTICA proporciona.

Si se escoge una unión relacional (simple o jerárquica) es necesario escoger una variable en cada archivo cuyos valores van a ser tomados como criterio de unión, esta variable se conoce como llave (*Key*). Así, para cada caso, el programa revisará los valores de esta llave en ambos archivos y unirá los casos sólo si los respectivos valores de las llaves coinciden.

**Not relational:** En este modo, las variables del segundo archivo simplemente se agregarán a los del primero (éste es el modo por omisión).

**Relational:** Si se elige este modo, los casos del segundo archivo serán igualados con los del primero, con base en los valores de la llave elegida.

**Relational hierarchial:** Este modo difiere del anterior ya que maneja varios registros con la misma llave, sea en el primero o segundo archivo. En el modo relacional estándar (anterior) los registros sucesivos con valores llave idénticos serán unidos. Si existe un número desigual de registros con valores llave idénticos en los dos archivos, entonces STATISTICA asignará el código de falta de dato al archivo con menor número de





registros. En contraste, si se escoge el modo *Relational Hierarchical*, entonces el archivo será completado con los valores del registro de la última llave que pudo ser igualada.

**Ejemplo:** Supóngase que se van a unir los siguientes archivos:

Archivo I		Archivo II	
Llave 1	Var1	Llave 2	Var2
1	1	1	1
2	4	1	2
2	5	1	3
3	2	2	4
3	3	2	5

Si se escoge el modo *Relational* la unión resultaría de la siguiente forma:

1	1	1	1
MD	MD	1	2
MD	MD	1	3
2	4	2	4
2	5	2	5
3	2	MD	MD
3	3	MD	MD

En cambio, en el modo *Relational hierarchical* el resultado es:

1	1	1	1
1	1	1	2
1	1	1	3
2	4	2	4
2	5	2	5
3	2	MD	MD
3	3	MD	MD

**Relational merge options:** Cuando se elige alguno de los dos modos: *Relational* o *Relational hierarchical*, STATISTICA espera que los dos archivos hayan sido ordenados (mediante la opción *Sort data*) con base en sus respectivas variables llave antes de realizar la unión.

**Key in file I (II):** Especificar el número de la variable que va a ser usada como llave en el primer (segundo) archivo.

**Key comparison (Numeric/Text):** Elegir si se van a comparar los valores numéricos o alfanuméricos de las variables llave.

**Unmatched cases:** Existen tres formas de tratar los casos que no cumplan con el criterio de unión relacional o que sobran porque el número de casos en cada archivo es diferente:

1. Llenar las celdas vacías con el código de falta de dato (**Fill with MD**).
2. Abortar la ejecución mandando un mensaje de error siempre que se encuentre un caso (en cualquiera de los archivos) que no se pueda unir (**Abort merge**).



### 3. Borrar los casos no apareados (**Delete**).

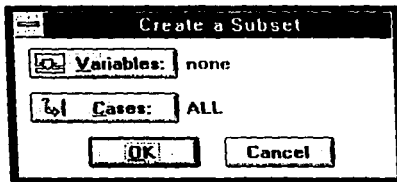
- **Cases:** Une los casos del segundo archivo después de los casos del primero. El número de variables en los dos archivos debe ser igual. Las listas de valores alfanuméricos y etiquetas de valores de los dos archivos serán concatenadas. Si los valores alfanuméricos o las etiquetas de valores de una variable del segundo archivo no coinciden con los del primero, STATISTICA toma las de éste último.
- **Text values:** Esta opción se utiliza cuando se desean unir los valores alfanuméricos (y etiquetas de valores) de aquellas variables que comparten el mismo nombre tanto en el archivo actual como en el segundo. Si dos variables (de los dos archivos) comparten el mismo nombre pero no se desea unir sus valores alfanuméricos, se debe cambiar el nombre de una de ellas temporalmente hasta que el proceso de unión haya terminado. Al presionar **OK** se abrirá la ventana *Merge Text Values* donde se debe seleccionar una de las tres posibles opciones para unir los valores alfanuméricos:

**Replace text values from file I with text values from file II:** Reemplaza todas las listas de valores alfanuméricos del primer archivo con las del segundo.

**Concatenate get duplicated text values from file I (file II):** Combina las listas de valores alfanuméricos del segundo archivo con las del primero, y si se detecta un valor duplicado, entonces STATISTICA usa los valores alfanuméricos del primer archivo (segundo).

**2nd file:** Al presionar este botón se abrirá la ventana *Open File* en la que se selecciona el archivo a unir con el archivo actual. No se puede unir un archivo consigo mismo, es decir, no se puede especificar el nombre del archivo actual como el nombre del segundo archivo. Si desea duplicar un archivo, ir a la opción *Save As* del menú *File* para crear una segunda copia del archivo actual con otro nombre.

## 2.2.5 CREAR UN SUBCONJUNTO DE UN ARCHIVO DE DATOS (CREATE SUBSET FROM DATA FILE)



Esta opción permite extraer un subconjunto de datos del archivo actual y guardarlo como un archivo independiente.

**Variables:** Indicar las variables que se incluirán en el subconjunto de datos.



**Cases:** Este botón abre la ventana *Case Selection Conditions* donde se pueden especificar los criterios de selección que deben cumplir los casos que se incluirán en el subconjunto.

## 2.2.6 ORDENAR LOS CASOS (SORT CASES)

STATISTICA realiza un ordenamiento jerárquico de los datos con base en los valores de una o varias variables llave especificadas por el usuario (si dos casos tienen valores idénticos en una llave dada, los valores de la siguiente llave serán utilizados para determinar el orden).

**Sort Options**

**Key 1**  
 Var: Y\_1983  
 Case Name  
 Ascen Numeric

**Key 2**  
 Var:   
 Case Name  
 Ascen Numeric

**Key 3**  
 Var:   
 Case Name  
 Ascen Numeric

OK  
 Cancel  
 More Keys...

Select one or more sorting keys. Double click on variable field to select from list of all variables, or sort by Case Names.

Ascen = ascending.  
 Desc = descending sort order. Sort by Numeric or Text values.

**Key #:** Seleccionar la llave que se utilizará para ordenar los casos, ésta puede ser:

**Var:** Una variable, o bien

**Case name:** La primera columna del archivo (la cual contiene los nombres de los casos).

**Ascen/Desc:** Esta opción sirve para indicar el orden que se desea, ya sea ascendente o descendente.

**Numeric/Text:** Seleccionar qué valores - (Text) alfanuméricos o (Numeric) numéricos - de la variable llave van a ser usados como criterio de orden. Cuando se ordena por valores alfanuméricos, si algún código numérico no tiene equivalente alfanumérico, entonces STATISTICA supone que los valores alfanuméricos son más pequeños que los numéricos.

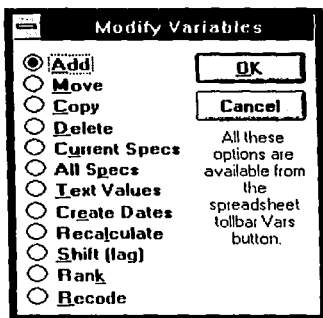
**More keys:** Este botón sirve para incrementar el número de llaves hasta un total de 7.

## 2.2.7 MODIFICAR VARIABLES (MODIFY VARIABLES)

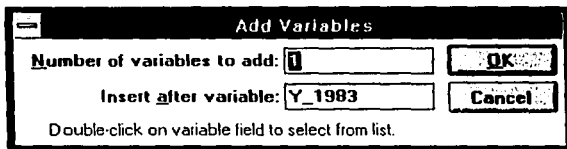
En esta opción además de las operaciones elementales de edición (agregar, mover, copiar, borrar) STATISTICA ofrece herramientas para facilitar la transformación y manipulación de las variables. Cada una de estas opciones despliega una ventana donde se especifican los requisitos necesarios para ejecutar la operación.

Botón asociado: **Vars**

También se encuentra disponible en el menú *Edit*.



- **Add:** Agrega nuevas variables al archivo de datos.



**Number of variables to add:** Indicar el número de variables que se agregarán en el archivo de datos actual. Se pueden incluir hasta 300 variables.

**Insert after which variable:** Las nuevas variables serán adheridas a la derecha de la variable especificada aquí.

- **Move, Copy, Delete:** Mueve, Copia o Borra el bloque de variables que se indique en la opción correspondiente.

Las ventanas de requisitos para estas tres opciones son similares:



**Move Variables**

From variable:

To variable:

Insert after:

Double-click on variable field to select from list.

**From variable:** Designar la primera variable incluida en el bloque de variables que será utilizado en la opción determinada. (Si se quiere utilizar una sola variable, entonces se debe designar la misma variable en *From variable* y *To variable*).

**To variable:** Designar la última variable incluida en el bloque de variables a ser utilizado.

**Insert after:** El bloque de variables será movido (copiado) a la derecha de la variable especificada aquí.

Nota: Para insertar antes de la variable 1; primero se debe mover (copiar) el bloque de variables después de la variable número 1 y luego mover esta variable a la derecha del bloque movido (copiado).

- **Current specs:** Esta opción despliega una ventana con las especificaciones de la variable seleccionada (resaltada) en el spreadsheet.

**Variable 1**

Name:

MD code:

**Format**

Width:

Decimal:

Date:

**Long name (label, formula, or link):**

Examples. Label: Gross income in 1991  
 Link: @Excellc:\file.xls!r2c2:r4c4  
 Formulas: = v1 + v2 ; comment  
 = {v1>0}\*INCOME + {v3<0}\*PROFIT

Esta ventana también se encuentra disponible haciendo doble clic en el nombre de la variable en el spreadsheet.



**Name:** Aquí se puede editar el nombre de la variable. Acepta como máximo 8 caracteres, no debe empezar con los caracteres . - + (de lo contrario el nombre será considerado inválido) o los dígitos (0-9) y no debe contener espacios, caracteres no imprimibles o caracteres utilizados en transformaciones de datos y lenguajes de programación de STATISTICA ("#&+\*,.-:<=>[]^{}). Si cualquiera de estos caracteres aparece en el nombre de la variable, entonces STATISTICA lo reemplazará con un subguión (\_). El programa convierte todas las letras en mayúsculas.

**MD code:** Se puede editar el valor que se utilizará como código de falta de dato. (Para obtener más información, ver la primera sección de este capítulo).

**Format:** Este recuadro contiene opciones para especificar el formato de despliegue de los valores de la variable:

**Width:** Editar la extensión total de los valores de la variable. La extensión total por omisión es de 8 caracteres. Este formato sólo afecta la manera en que los valores serán desplegados en el spreadsheet sin afectar la precisión almacenada del número.

**Decimal:** Especificar el número de espacios decimales. El formato por omisión reserva 3 espacios. Al especificar el formato decimal para variables tipo fecha (ver abajo), las desplegará de acuerdo al formato Julian (este formato despliega las fechas como un simple número entero que representa el número de días que han transcurrido desde el primero de enero de 1900; por ejemplo, 24858 representa el día 21 de enero de 1968).

**Date:** Seleccionar este formato cuando se tengan variables cuyos valores van a ser interpretados como fechas. Existen varias opciones a elegir dentro de este formato. Para obtener más información acerca de las variables tipo fecha, consultar la opción *Create dates* descrita más adelante dentro de la presente sección.

**Long name (label, formula or link):** En este cuadro se puede editar una etiqueta, fórmula o liga para la variable:

**Label:** A cada variable se le puede asignar una etiqueta de hasta 128 caracteres. La etiqueta puede contener notas, incluyendo cualquier carácter imprimible.

**Formula:** Las fórmulas matemáticas deben empezar con el signo "=". Si la fórmula se considera correcta, se tiene la oportunidad de recalcular la variable en cualquier momento del análisis (por medio de la ventana *Recalculate* descrita más adelante en esta sección). Las fórmulas son útiles para verificar los datos, transformar una variable, recodificar una variable o crear valores de la variable basados en condiciones lógicas. Puede agregarse un comentario a la fórmula después de un punto y coma ";" (ejemplo:  $= (v1+v2+v3)/3$ ; promedio de las primeras tres variables). Para obtener información detallada acerca de la sintaxis y funciones permitidas, consultar la ayuda en línea del paquete.


**Link:** Se puede ligar una variable (o bloque de datos) de STATISTICA con una variable o bloque en una ventana de edición de otro programa, permitiendo actualizar automáticamente el spreadsheet siempre que el archivo fuente cambie. La manera más



común de establecer una liga entre otra aplicación y STATISTICA es por medio de las opciones *Paste Link* y *Links* del menú *Edit*. Una liga también puede ser creada dentro de esta opción (*Long Name*). Ejemplo: Para ligar un bloque de datos de STATISTICA con un bloque de datos de Excel se debe escribir: @Excel |c:\excel\file.xls!r2c2:r4c4. Este ejemplo establece una liga dinámica con el bloque "desde renglón número 2, columna número 2 al renglón número 4, columna número 4" en el archivo file.xls de Excel. Una explicación detallada acerca de la sintaxis para establecer ligas se encuentra en la sección *Edit* del capítulo 3.

**All specs:** Al presionar este botón se cerrará esta ventana (guardando los cambios) y se abrirá la ventana *View Variable Specs* (ver abajo).

**Text values:** Al presionar este botón se cerrará esta ventana (guardando los cambios) y se abrirá la ventana *Text Values Manager* (ver abajo).

**Values/stats:** Este botón abre la ventana *Variable #* que presenta los valores de la variable así como el número de datos, media y desviación estándar. Es posible copiar el nombre de la variable y estas estadísticas descriptivas al portapapeles de Windows presionando el botón:  para pegarlos en una gráfica, ventana text/output o en otras aplicaciones de Windows.

**Graphs:** Este botón cerrará esta ventana (guardando los cambios) y abrirá el menú de *Quick Stats Graphs* (ver el capítulo 4).

- **All specs:** Esta opción desplegará en un scrollsheet las especificaciones de todas las variables: número, nombre, código de falta de dato, formato de valores y nombre largo.

Variables: ACCIDENTI_STA 2v * 12c				
Var	Name	MD Code	Format	Long Name (label)
1	Y_1983	-9999	8.0	Number of accidents in 1983
2	Y_1985	-9999	8.0	Number of accidents in 1985

Botón asociado: 

Esta opción también se encuentra disponible presionando el botón *All specs* de la ventana que despliega la opción *Current specs*.

- **Text values:** Esta opción permite editar los valores alfanuméricos (y sus equivalentes numéricos) de la variable seleccionada (resaltada) en el spreadsheet, dentro de la ventana *Text Values Manager*:

Para crear una nueva asignación "numérica-alfanumérica" se deben introducir los dos valores respectivos (texto, número) en los dos primeros campos del renglón (ejemplo: masculino,1); presionando la tecla <tab> es posible moverse al siguiente campo (o renglón); para cambiar un campo es necesario sobrescribirlo; para poder editarlo, se debe



presionar la tecla <F2> o hacer doble clic en él. La tecla <Insert> permite insertar una nueva línea; la tecla <Supr> borra la línea actual.

**Text value:** **Numeric:** **Value label:**

--	--	--

To create a new 'text-numeric' assignment, enter the respective two values (text, numeric) into the first two fields of a row, e.g., MALE and 1 (Value label is optional).

Press Tab to move to the next field (or row); the contents of the last row are automatically copied (they can be edited or typed over). To change a field - type over; to edit a field - press F2 or double-click.

Insert - inserts a new line;  
Delete - deletes the current line.

**OK**  
**Cancel**  
**Fill Down**  
**Clear All**

**Sort by**  
**Text-incr**  
**Text-decr**  
**Num-incr**  
**Num-decr**

**Copy from**  
**Variable**  
**Case Names**

**Text value:** Esta columna permite editar los valores alfanuméricos que pueden tener hasta 8 caracteres de longitud; no deben iniciar con los caracteres . - + (si no serán considerados inválidos) o dígitos (0-9) y no pueden contener espacios, caracteres no imprimibles o caracteres utilizados en transformaciones de datos y lenguajes de programación de STATISTICA. ("#&^()+\*,-.;<=>[]^{}). Si cualquiera de estos caracteres aparecen en un valor alfanumérico, STATISTICA los reemplazará con un subguión (\_). No hay diferencia entre letras mayúsculas y minúsculas (todas las letras se convierten en mayúsculas).

**Numeric:** Esta columna permite editar los valores numéricos. Si una variable con valores alfanuméricos es seleccionada para un análisis que requiere datos numéricos, entonces sus valores numéricos son utilizados automáticamente.

**Value label:** A cada valor alfanumérico se le puede asignar una etiqueta; ésta puede ser de hasta 40 caracteres de largo y contener notas incluyendo cualquier carácter imprimible. Las etiquetas de valores son utilizadas automáticamente en la salida (*output*) de todo análisis que involucre valores individuales.

**Fill down:** Este botón asigna automáticamente valores numéricos (enteros consecutivos) a los alfanuméricos, iniciando con el número asociado al primer valor alfanumérico de la lista.

**Clear all:** Al presionar este botón, automáticamente se limpian todas las entradas en todos los campos de la ventana.





### Sort by.

**Text-incr (Text-decr):** Ordena los valores numérico, alfanumérico y etiquetas de valores en orden alfabético creciente (decreciente) utilizando los valores alfanuméricos como llave.

**Num-incr (Num-decr):** Ordena los valores numérico, alfanumérico y etiquetas de valores en orden numérico creciente (decreciente) utilizando los valores numéricos como llave.

**Copy from:** Agrega los valores numérico y alfanumérico de otra variable doble (**Variable**) o el nombre y número de los casos (**Case names**).

Botón asociado:

Esta opción también se encuentra disponible presionando el botón *Text values* de la ventana que despliega la opción *Current specs*.

- **Create dates:** Esta opción permite crear una variable tipo fecha (mes/día/año; mes/año; día/mes) a partir de dos o tres variables (v.g. V1=mes, V2=día, V3=año) o descomponer una variable tipo fecha en dos o tres variables.

Internamente, STATISTICA almacena los valores de las variables tipo fecha en formato decimal o Julian. Este formato muestra las fechas como un simple número entero que representa el número de días que han transcurrido desde el primero de enero de 1900; por ejemplo, 24858 representa el día 21 de enero de 1968. Se puede cambiar entre éste y los diferentes formatos de despliegue disponibles con la opción *Dates* de la ventana *Current specs*.

Si se escribe una fecha en el formato mes-año (v.g. Mar-1997) STATISTICA interpreta que se trata del día primero. En el caso del formato día-mes (v.g. 6-Mar) va a considerar el año en curso; esto es para poder almacenar las fechas en formato Julian.

Al escribir una fecha sin separadores (v.g. 2855 en lugar de 2/8/55) STATISTICA la considerará escrita en formato Julian y la convertirá al formato especificado dando como resultado una fecha incorrecta (v.g. 2855 lo interpretaría como 10/25/07 en lugar de 2/8/55).

**Date Operations**

**Create Date from 2 or 3 Variables**

**Split Date into 2 or 3 Variables**

**Text Values / Dates**

<p><b>Source Variables</b></p> <p>Day: <input type="text"/></p> <p>Month: <input type="text"/></p> <p>Year: <input type="text"/></p>	<p><b>Destination Variable</b></p> <p>Name: <input type="text" value="Y_1983"/></p> <p>Format: <input type="text" value="3/6/87"/> <input type="button" value="↓"/></p>
--	---

Double-click on variable field to select from list.



**Create date from 2 or 3 variables:** Seleccionar esta opción para crear una variable tipo fecha con los valores de mes, día y año a partir de 3 variables existentes o con los valores de mes y año o día y mes de 2 variables existentes.

**Split date into 2 or 3 variables:** Seleccionar esta opción para descomponer los valores tipo fecha de una variable existente en valores separados de día, mes y año que serán guardados en 3 nuevas variables (creadas previamente) o en valores separados para mes y año o día y mes en 2 nuevas variables (creadas previamente).

**Text values/dates:** Este botón abre la ventana *Text Value/Dates* donde se pueden transformar los valores tipo fecha de una variable en los valores alfanuméricos de otra (puede ser la misma) y viceversa (**Date variable to variable with text values, Variable with text values to date variable**). Se debe indicar el nombre de la variable que va a ser transformada o variable origen (**Source variable**) y el de la variable destino (**Destination variable**) así como el formato de la fecha (en caso de haber elegido la segunda opción).

**Source variables:** De las variables que se seleccionen en esta opción dependerá el resultado de crear una variable tipo fecha o la descomposición de una variable tipo fecha.

Si se eligió la opción *Create date from 2 or 3 variables* es necesario especificar como variables origen o fuente, las 3 variables que contienen los valores de mes, día y año o las 2 variables que contienen día y mes o mes y año.

Para la opción *Split Date into 1, 2 or 3 variables* se debe especificar como variable fuente, la que contiene los valores tipo fecha.

**Destination variables:** Las variables que se seleccionen aquí tendrán los valores que resulten al crear o descomponer valores tipo fecha.

Si se eligió la opción *Create date from 2 or 3 variables* es necesario especificar como variable destino, el nombre de una variable en la que se introducirán los nuevos valores tipo fecha. También se debe especificar el formato de la fecha eligiendo uno de los que se encuentran disponibles en la opción *Format*.

En caso de haber elegido *Split date into 1, 2 or 3 variables* se deben especificar como variables destino, variables previamente creadas en las que los valores de mes, día y año, día y mes o mes y año se introducirán.

- **Recalculate:** Es posible recalcular una variable (o grupo de variables) con base en las fórmulas introducidas en la ventana *Current Specs* para cada variable.



**All variables:** Esta opción sirve para recalcular simultáneamente todas las variables que contienen fórmulas.

**Variable:** Indicar el nombre de la variable que se va a recalcular.

**Subset:** Por omisión, todos los casos serán recalculados. Si un bloque de casos está marcado en el spreadsheet, sólo los casos incluidos en el bloque serán recalculados. Si se quiere recalcular un bloque diferente de casos se debe introducir el número de los casos en las opciones: **From Case, To Case**.

Botón asociado:

- **Shift (lag):** Esta opción sirve para desplazar una variable seleccionada un número específico de casos, ya sea hacia arriba o hacia abajo, en relación a otras variables en el archivo.

STATISTICA agregará casos en blanco a las variables restantes del archivo ya sea antes del primer caso ( si la variable seleccionada es movida hacia arriba) o después del último caso ( si la variable seleccionada es movida hacia abajo).

**Variable:** Seleccionar la variable que será desplazada.

**Lag:** Especificar la magnitud del movimiento (es decir, el número de casos que se moverá la variable).



**Direction:** Elegir el desplazamiento de la variable: **Forward** (hacia abajo) o **Backward** (hacia arriba).

- **Rank:** Esta opción permite ordenar los valores de las variables seleccionadas asignándoles un rango. Este rango va a remplazar los valores originales de la variable. Si se desea conservar estos valores, será necesario hacer una copia de la misma (con la opción *Copy* descrita en esta sección).

The screenshot shows the 'Rank Order Values' dialog box. It has a title bar 'Rank Order Values'. Below the title bar, there are several sections:

- Variables:** A list box containing 'Y\_1983'.
- Cases:** A list box containing 'ALL'.
- Weight:** A list box containing 'OFF'.
- Assign Rank 1 to:** Two radio buttons: 'smallest value' (selected) and 'largest value'.
- Ranks for Ties:** Four radio buttons: 'mean' (selected), 'sequential', 'low', and 'high'.
- Type of Ranks:** Three radio buttons: 'regular' (selected), 'fractional', and 'fractional as Z'.

At the top right of the dialog, there are 'OK' and 'Cancel' buttons.

**Variables:** Se abrirá la ventana donde se deben indicar las variables que serán ordenadas.

**Cases:** Al presionar este botón se abrirá la ventana *Case selection conditions* donde se deben seleccionar por medio de condiciones lógicas los casos que serán ordenados (consultar la sección Selección de Casos al inicio de este capítulo).

**Weight:** Este botón abre la ventana *Define weight* descrita al inicio del capítulo.

**Assign rank 1 to:** Se pueden ordenar los valores en forma ascendente (al valor más pequeño se le asigna el número 1) o descendente (al valor más grande se le asigna el número 1).

**Rank for ties:** Existen 4 posibles opciones para dar un rango a los valores empatados:

**Mean:** Asigna la media de los rangos de los valores empatados.

**Sequential:** Ignora los empates y asigna un valor secuencial a cada valor empatado.

**Low:** Asigna el rango más pequeño de los valores empatados a cada uno de ellos.

**High:** Asigna el rango más grande de los valores empatados a cada uno de ellos.

**Type of ranks:** Elegir el tipo de rango que se quiere asignar:



**Regular:** Asigna rangos de 1 a n.

**Fractional:** Asigna rangos entre 0 y 1.

**Fractional %:** Los rangos son porcentajes basados en el rango fraccional.

- **Recode:** En esta opción se pueden traducir o convertir los valores originales de la variable seleccionada (resaltada en el spreadsheet) en otros valores nuevos ( que van a depender de los valores de algunas variables en el conjunto de datos). Una aplicación de esta herramienta podría ser el crear una variable para agrupar datos y así identificar subconjuntos de datos que serían comparados en análisis subsecuentes. Ejemplo: Se podría asignar el valor 1 a los sujetos menores de 17 años y 2 a los sujetos mayores de 18 o bien, asignar el valor 1 a los hombres entre 18 y 25 años con nivel de colesterol menor que 200; 2 a los hombres entre 18 y 25 años con nivel de colesterol mayor que 200; y asignar el código de falta de dato a todos los demás.

**Recode Values of Variable 1: GENDER**

<b>Category 1</b> <input type="checkbox"/> Include If <input type="checkbox"/> Exclude If <input type="button" value="Open/Save"/>	<b>New Value 1</b> <input checked="" type="radio"/> value: 1 <input type="radio"/> MD code	<input type="button" value="OK (Recode)"/> <input type="button" value="Cancel"/> <input type="button" value="Review Vars"/> <input type="button" value="Open All"/> <input type="button" value="Save All"/> <input type="button" value="Print"/> Other If no conditions are met, set values to: <input type="radio"/> MD code <input type="radio"/> value: <input type="text"/> <input checked="" type="radio"/> unchanged
<b>Category 2</b> <input type="checkbox"/> Include If: <input type="checkbox"/> Exclude If: <input type="button" value="Open/Save"/>	<b>New Value 2</b> <input checked="" type="radio"/> value: 2 <input type="radio"/> MD code	
<b>Category 3</b> <input type="checkbox"/> Include If: <input type="checkbox"/> Exclude If: <input type="button" value="Open/Save"/>	<b>New Value 3</b> <input checked="" type="radio"/> value: 3 <input type="radio"/> MD code	
<b>Category 4</b> <input type="checkbox"/> Include If: <input type="checkbox"/> Exclude If: <input type="button" value="Open/Save"/>	<b>New Value 4</b> <input checked="" type="radio"/> value: 4 <input type="radio"/> MD code	

**Category:** Se pueden especificar hasta 16 categorías para recodificar una variable.

**Include if/Exclude if:** Cuando se elige *Include if*, los casos que cumplan la condición de recodificación (ver abajo) serán reemplazados con el nuevo valor que se especifique en *New value*. *Exclude if* asignará el nuevo valor a los casos que no cumplan la condición especificada.

En el recuadro en blanco se debe escribir la condición de recodificación siguiendo las mismas reglas de sintaxis que se usan en las condiciones de selección de casos (descritas al inicio del capítulo), es decir:

Las variables se deben especificar ya sea por su número (v1,v2...) o por su nombre (género, edad,...), en este caso, las mayúsculas son equivalentes a las minúsculas ('SEXO'='sexo'). La columna que contiene el número de cada caso se identifica como v0.



Se deben usar comillas simples para hacer referencia a los valores alfanuméricos (v1='masculino' o v1='MASCULINO').

Se pueden usar paréntesis para especificar condiciones lógicas complejas o para cambiar la precedencia de los operadores.

A continuación se muestra la lista de operadores disponibles para la selección de casos:

=	igual
#,<>,><	distinto
<	menor que
>	mayor que
<=	menor o igual
>=	mayor o igual
NOT, ~	negación
AND, &	y
OR,	o

Ejemplo:

Category 1: Include if: v1>=0

New value: 1

Category 2: Include if: v1<0

New value: -1

Este par de condiciones de recodificación convertirá los valores de la variable seleccionada (resaltada) en el spreadsheet a 1 (si el valor de la primera variable (v1) es mayor o igual que cero) o -1 (si el valor de la primera variable (v1) es menor que cero). Notar que las condiciones de recodificación pueden ser definidas de tal manera que los nuevos valores de la variable resaltada en el spreadsheet no dependan de los anteriores, sino de otra variable del archivo, es decir, no es necesario incluir en el enunciado de las condiciones de recodificación a la variable seleccionada en el spreadsheet. Por ejemplo, en este caso, la variable resaltada en el spreadsheet podría ser la tercera (v3) y sus nuevos valores dependerán únicamente de la primera (v1). Si la variable resaltada en el spreadsheet fuera v1 entonces sus valores serían transformados (los positivos se convertirían en 1 y los negativos en -1).

**Open/save:** Este botón despliega la ventana *Case Selection Conditions* que permite abrir (o guardar) cada una de las condiciones de recodificación en un archivo de condiciones de selección de casos (con extensión \*.sel).

**New value:** Especificar el nuevo valor que reemplazará al valor original. Se puede elegir entre un valor específico (**Value**) o el valor de falta de dato (**MD code**).

**Review vars:** Abre la ventana *Variables* que muestra la lista de variables disponibles en el archivo actual.



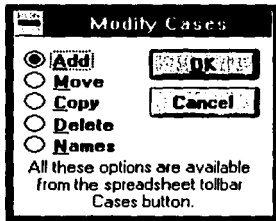
**Open all:** Este botón abre la ventana *Open Recode Conditions File* que permite abrir archivos que contienen varias condiciones de recodificación creadas previamente. Estos archivos tienen formato tipo texto (extensión \*.txt).

**Save all:** Este botón abre la ventana *Save Recode Conditions File* para guardar en un archivo todas las condiciones de recodificación que se editaron en esta ventana. Este archivo es guardado con el formato de archivo de texto (extensión \*.txt).

**Print:** Al presionar este botón se enviará el contenido de esta ventana ya sea a un archivo (para imprimir posteriormente) o directamente a la impresora dependiendo de la selección hecha en la ventana *Page/Output Setup* (descrita en el capítulo 3).

**Other:** Si un caso (observación) no cumple las condiciones se puede elegir entre sustituir los valores originales de la variable por el código de falta de dato (**MD code**), designar un valor distinto (**Value**) o dejar los valores originales (**Unchanged**).

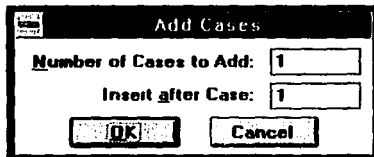
## 2.2.8 MODIFICAR CASOS (MODIFY CASES)



En esta opción se encuentran disponibles las operaciones elementales de edición de casos. Cada una de estas operaciones despliega una ventana donde se deben especificar los requisitos necesarios para que se ejecute.

Botón asociado: **Cases** También se encuentra disponible en el menú *Edit*.

- **Add:** Agrega nuevos casos al archivo de datos.



**Number of cases to add:** Indicar el número de casos que se agregarán al archivo de datos actual.



**Insert after case:** Indicar el número del caso después del cual se agregarán los nuevos casos.

- **Move, Copy, Delete:** Mueve, copia o borra el bloque de casos que se indique en la opción correspondiente.

Las ventanas de requisitos para estas tres opciones son similares:

Move Cases

From Case: 0

To Case: 1

Insert after:

OK Cancel

**From case:** Especificar el primer caso a incluir en el bloque que será utilizado en la opción determinada. (Si se quiere utilizar un solo caso, entonces se debe designar el mismo caso en ambos: *From case* y *To case*).

**To case:** Designar el último caso a incluir en el bloque.

**Insert after:** El bloque de casos será movido (copiado) después del caso que se especifique aquí.

Nota: Para insertar antes del caso 1: Primero de debe mover (copiar) el bloque de casos después del caso número 1 y luego mover el caso número 1 después del bloque de casos movido (copiado).

- **Names:** Esta opción enlista los nombres de los casos en una tabla donde se pueden revisar o editar.

Case Name Manager

1:	January
2:	February

OK Cancel Options

**Options:** Este botón abrirá la ventana *Case Names Options* donde se puede ajustar la longitud de los nombres de los casos.

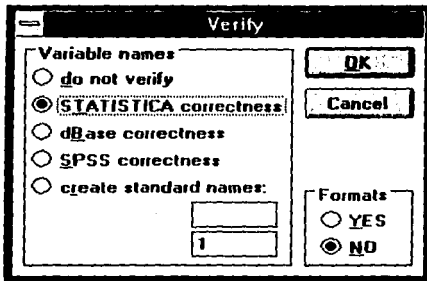




Se puede acceder a esta ventana haciendo doble clic en el nombre del caso que se quiere editar.

### 2.2.9 VERIFICAR NOMBRES Y FORMATOS DE LAS VARIABLES (VERIFY VARIABLE NAMES / FORMATS)

Esta opción permite verificar los nombres y formatos de las variables. Puede utilizarse para revisar y ajustar el formato de las columnas a los datos, revisar si existen nombres de columnas duplicados o verificar su compatibilidad con estándares seleccionados.



**Variable names:** Las opciones dentro de este recuadro sirven para verificar los nombres de las variables según los requisitos de STATISTICA, dBASE, SPSS, o crear nombres estandarizados. Si se encuentra un nombre no válido o duplicado, se abrirá la ventana *Name Verification* para corregirlo:

**Do not verify:** Seleccionar esta opción si únicamente se quieren verificar los formatos de las variables y no sus nombres.

**STATISTICA correctness:** Revisa si existen nombres duplicados, ya sea para cumplir con los requerimientos de Quick MML o antes de exportar el archivo.

**dBase correctness:** Revisa conforme los requerimientos de dBase. Los nombres válidos de las variables pueden contener las letras A-Z, a-z, los dígitos 0-9, y el subguión (\_). Cada nombre debe iniciar con una letra y todos los nombres de las variables deben ser únicos (lo que se puede verificar con la opción anterior).

**SPSS correctness:** Revisa los nombres de acuerdo con los requerimientos de SPSS. Los nombres válidos pueden contener los siguientes caracteres: A-Z, a-z, 0-9, (\_), (.), @, #, y \$; cada nombre debe empezar con cualquiera de las letras, \$ o @ y no puede ser ninguna palabra reservada de SPSS (all, and, by, eq, ge, le, lt, en, not, or, to, with).

**Create standar names:** Esta opción asigna nombres estándar (prefijo y número secuencial) a todas las variables por medio de las siguientes opciones:



**Prefix:** Especificar un prefijo estándar de 5 letras como máximo.

**Start number:** Indicar el número inicial que formará parte del nombre de las variables.

**Formats:** YES: Esta opción examina todas las variables para determinar si los formatos de despliegue están ajustados a los datos. Si el formato es más grande que cualquiera de los valores en la columna, la longitud del formato se ajustará a la longitud del valor más largo.

## 2.2.10 VERIFICAR VALORES DE DATOS (VERIFY DATA VALUES)

Una vez que los datos (introducidos o importados) están listos en el archivo, se puede verificar su integridad (consistencia lógica) y completez. La verificación puede ser tan simple como revisar si los valores de las variables son legales ( por ejemplo en la variable sexo sólo son válidos los valores : femenino y masculino); o si están incluidos dentro del rango de valores permitidos ( por ejemplo la variable edad debe encontrarse entre 0 y 200 años). También puede ser tan complejo como revisar condiciones lógicas que algunos valores deben cumplir en relación a otros valores ( por ejemplo: si una persona es de sexo masculino o es menor que una cierta edad entonces el número de embarazos no puede ser mayor que cero).

Verify Data

A case will be considered valid if:

All conditions are met  At least one condition is met

Condition 1

Valid If: [ ] Open/Save

Condition 2

Valid If: [ ] Open/Save

Condition 3

Valid If: [ ] Open/Save

Condition 4

Valid If: [ ] Open/Save

OK (Verify)

Cancel

Review Vars

Open All

Save All

Print

Range

From case: 1

To case: 24

Los datos serán examinados secuencialmente ( un caso cada vez) en su consistencia con las condiciones especificadas, cuando un caso no coincida con las condiciones, el renglón



respectivo en el spreadsheet será marcado o resaltado y se abrirá la ventana *Data Verification* permitiendo ignorar (y continuar o detener la verificación) o editar (corregir) el caso.

**A case will be considered valid if:** Seleccionar una de las siguientes opciones para definir cuándo un caso será considerado válido.

**All conditions are met:** Un caso será considerado correcto sólo si todas las condiciones se cumplen.

**At least one condition is met:** Un caso será considerado correcto si al menos una de las condiciones definidas se cumple.

**Condition:** Se pueden definir hasta 16 diferentes condiciones de verificación.

**Valid/invalid if:** Seleccionar si el valor será considerado válido o inválido cuando cumpla las condiciones de verificación.

**Open/save:** Este botón abrirá la ventana *Case Selection Conditions* que permite abrir un archivo con una condición de selección de casos previamente guardada para ser utilizada como condición de verificación o bien, guardar la condición actual.

En el recuadro en blanco se debe introducir la condición de selección de casos que será utilizada como criterio de verificación. Las reglas de sintaxis son las mismas que se usan en las condiciones de selección de casos (descritas al inicio del capítulo), es decir:

Las variables se deben especificar ya sea por su número (v1,v2...) o por su nombre (género, edad,...), en este caso, las mayúsculas son equivalentes a las minúsculas ('SEXO'='sexo'). La columna que contiene el número de cada caso se identifica como v0.

Se deben usar comillas simples para hacer referencia a los valores alfanuméricos (v1='masculino' o v1='MASCULINO').

Se pueden usar paréntesis para especificar condiciones lógicas complejas o para cambiar la precedencia de los operadores.

A continuación se muestra la lista de operadores disponibles para la selección de casos:

=	igual
#,<>, ><	distinto
<	menor que
>	mayor que
<=	menor o igual
>=	mayor o igual
NOT, ~	negación
AND, &	y
OR, !	o



Ejemplo:

Invalid if: (v1='masculino' or edad<14) and embarazo <> 0

**Review vars:** Este botón abre la ventana *Single Variable List* en donde se pueden revisar los nombres de las variables.

**Open all:** Este botón abre la ventana *Open Verify Conditions File* que permite abrir un archivo creado previamente con varias condiciones de selección de casos para utilizarlas como condiciones de verificación.

**Save all:** Este botón abre la ventana *Save Verify Conditions File* para guardar en un archivo las condiciones de verificación que se editaron en esta ventana.

**Print:** Al presionar este botón se enviará el contenido de esta ventana ya sea a un archivo (para imprimir posteriormente) o directamente a la impresora dependiendo de la selección hecha en la ventana *Page/output Setup* (descrita en el capítulo 3).

**Range:** Aquí se define el rango de casos a verificarse. De qué caso (**From case**) a qué caso (**To case**).

## 2.2.11 ESTANDARIZAR VARIABLES; REEMPLAZAR DATOS FALTANTES POR MEDIAS (STANDARDIZE VARIABLES; REPLACE MD BY MEANS)

**Estandarizar variables.**

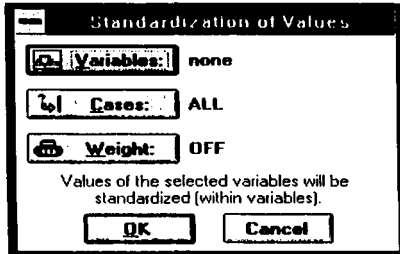
Esta opción permite estandarizar los valores de las variables seleccionadas de acuerdo a las siguiente fórmula:

Valor estándar = (Valor original - Media) / Desviación estándar

**Reemplazar datos olvidados por sus medias.**

Esta opción reemplaza permanentemente los valores faltantes de las variables seleccionadas con las medias de esas variables (esta opción también está disponible en otros módulos de STATISTICA donde el reemplazamiento es temporal, es decir, se aplica sólo en el análisis actual o vigente).

Las ventanas de requerimientos para estas dos opciones son similares:



**Variables:** Seleccionar las variables que se utilizarán en la opción determinada.

**Cases:** Indicar el rango de casos que serán considerados en la opción determinada (consultar la sección Selección de Casos al inicio de este capítulo).

**Weight:** Se explicó en la sección Variables de peso al inicio del capítulo.



## CAPITULO 3

### MENUS PRINCIPALES

Como se mencionó en la sección Convenciones Generales del capítulo 1, STATISTICA tiene integradas tres herramientas muy importantes para el manejo de los datos que son accesibles desde cualquier módulo del paquete:

- El spreadsheet
- El scrollsheet
- La ventana text/output

Al activar cualquiera de estas tres utilerías en algún módulo del paquete, automáticamente desplegarán su propio menú y barra de herramientas.

La mayor parte de las opciones de menús para las tres ventanas, son las mismas que se despliegan cuando el spreadsheet está activo, y éste por ser el más importante, es el que se activa por omisión al iniciar la sesión, por lo que se comenzará describiendo la barra de menús del spreadsheet y después sólo se agregarán aquellas opciones exclusivas del scrollsheet y de la ventana text/output.

### 3.1 SPREADSHEET

#### 3.1.1 FILE

Contiene opciones referentes al acceso y almacenamiento de los diferentes archivos con los que STATISTICA trabaja.

**NEW DATA:** Permite crear un nuevo archivo de datos con 10 casos y 10 variables.

**OPEN DATA:** Abre un archivo de datos almacenado en disco. Al seleccionar esta opción, se despliega la ventana *Open Data File*, donde se elige el archivo que se quiere abrir.

Botón asociado:

**IMPORT DATA:** Importa archivos de datos de otras aplicaciones compatibles con STATISTICA (las opciones de importación se explican en la sección Importar Archivo de Datos del capítulo 2).

**OPEN OTHER:** Permite abrir otros archivos aparte del que se tiene en uso.



**GRAPH FILE** Despliega la ventana *Open Graph File*, donde se selecciona el archivo de gráficas (con extensión \*.stg) que se quiere abrir.

Los archivos de gráficas no sólo contienen la gráfica (en el formato de STATISTICA) sino además todos los objetos que se le quieran agregar (por ejemplo: dibujos, líneas, etc.) y los datos representados en la gráfica. Los datos pueden ser editados o manipulados una vez que el archivo ha sido abierto. Para obtener información acerca de los diferentes formatos de archivos gráficos consultar la sección Menús y Barra de Herramientas del capítulo 4.

**SCROLLSHEET:** Esta opción permite abrir scrollsheets (archivos con extensión \*.scr). Al seleccionarla despliega la ventana *Open Scrollsheet* de donde se elige el archivo requerido.

Los archivos tipo scrollsheet contienen conjuntos de datos (obtenidos de análisis estadísticos realizados) con la información referente a su formato y todas las características con las que han sido editados. Más información referente al scrollsheet, se encuentra en la sección Convenciones Generales del capítulo 1.

**TEXT/OUTPUT:** Despliega la ventana *Open Text File* donde se elige el archivo de texto que se quiere abrir (\*.txt). El archivo de texto seleccionado se desplegará en la ventana *Text/Output Window*, la cual se describe con más detalle en la sección Convenciones Generales del capítulo 1.

**SAVE:** Guarda un archivo de datos. Si el archivo no ha sido guardado antes se abrirá la ventana *Save File As*, en la cual se debe especificar el nombre del archivo (\*.sta) y la ruta donde se va a guardar, una vez que se han especificado, cada vez que se elige esta opción, el archivo será guardado automáticamente sobrescribiendo lo que se había guardado la vez anterior. Si se quiere cambiar el nombre o localización del archivo, seleccionar la opción *Save as*.

Botón asociado: 

**SAVE AS:** Despliega la ventana *Save File As*, en ésta, por omisión aparecen el nombre del archivo y la ruta actuales. Se puede guardar el archivo de datos bajo un nuevo nombre o dejar el que tenía así como también especificar otra ruta donde guardarlo.

**MAKE SCROLLSHEET:** Crea un scrollsheet del actual archivo de datos (spreadsheet), al seleccionar esta opción se abrirá la ventana *Select Subset for Scrollsheet*, en ésta se debe especificar el rango de datos que se van a incluir en el scrollsheet (por omisión toma todo el archivo o un bloque que se haya seleccionado previamente). Los correspondientes nombres de variables y de casos del archivo de datos serán ahora las columnas y nombres de renglones del scrollsheet. El encabezado (*file header*), será ahora el nombre del archivo (*file name*) y el rango o subconjunto (*subset*) de datos seleccionados indicará el título del scrollsheet.

Opciones de la ventana *Select Subset for Scrollsheet*:

**First/Last case:** Aquí se especifica el número correspondiente al primer/último caso que va a ser incluido en el scrollsheet.



**First/Last variable:** Se debe indicar el nombre o número de la primera/última variable que va a ser incluida en el scrollsheets.

**PRINT/EJECT CURRENT PAGES:** Generalmente cuando se ejecuta un análisis y la salida está direccionada hacia la impresora o a un archivo (esto se especifica en la ventana *Page/Output Setup*) STATISTICA almacenará internamente esta salida hasta que se llene una página (si la opción *Automatically Eject Each Filled Page* es seleccionada) o hasta que esté completo todo el análisis (si la opción *Automatically Eject Each Filled Page* es desactivada) antes de enviarla a la impresora o al archivo. Esta opción no hará caso de estas condiciones y al ejecutarla, inmediatamente imprimirá (o enviará a un archivo) toda la salida que tenga acumulada (técnicamente se trata de una forma continua). Cabe resaltar que esta opción no es aplicable en la ventana *text/output*.

**PRINT:** Imprime el spreadsheet activo, ya sea que lo envíe a un archivo (para imprimirlo posteriormente), a la impresora o a la ventana *text/output* según se haya establecido previamente en la ventana *Page/Output Setup*. Esta opción despliega la siguiente ventana:

**Print:** Permite seleccionar las partes del spreadsheet que se quieren imprimir (cualquier elección que se haya hecho previamente en la ventana *Page/Output Setup*, será ignorada si causa algún conflicto con lo que se seleccione aquí).

**Data values:** Imprime el contenido del spreadsheet, es decir, los valores de los datos.

**Variable specs:** Esta selección incluirá en la impresión, una tabla con todas las especificaciones de las variables, es decir, nombre, código de falta de dato, formato y nombre largo (etiqueta, fórmula o liga) tal como se especificaron en la ventana *Current Specs* descrita en la sección *Modificar Variables* en el capítulo 2.





**Text value assignment:** Esta opción incluirá en la impresión, una tabla separada conteniendo los valores alfanuméricos de los datos, su equivalente numérico y etiquetas, tal como se especificaron en la ventana *Text Values Manager* descrita en la sección Modificar Variables en el capítulo 2.

**File info/notes:** Esta opción incluirá el encabezado del archivo y las notas e información establecidas en la opción *Header* del menú *Edit*.

**Variables:** Este botón permite seleccionar las variables que se desean imprimir (por omisión imprimirá todas las variables o el bloque actualmente seleccionado o resaltado).


**Print data value options:** Este recuadro contiene opciones para la impresión de los valores de los datos:

**Cases from-to:** Permite definir un subconjunto de casos para ser impreso (del caso n al caso m).

**Case names:** Incluye el nombre de los casos en la impresión.

**Text values:** Imprime los valores alfanuméricos de las variables (en lugar de su equivalente valor numérico).

**Column width:** Define el ancho de las columnas del spreadsheet, éste puede ser definido de acuerdo al formato de los valores (**By format**) o determinando un ancho mínimo de 8 caracteres (**Min 8 char**), este último permite imprimir los nombres completos de las variables.

**Print/eject pages after this printout:** Si esta opción está activa, cuando se presione el botón **OK** en esta ventana (*Print Data*), o el botón de impresión de la barra de herramientas: , la información automáticamente será dirigida a la impresora o al archivo de salida (dependiendo de las elecciones hechas en *Page/Output Setup*). Cada impresión empezará en una nueva página. Si la opción se encuentra desactivada, la salida será almacenada en el buffer interno y enviada a la impresora dependiendo del estado actual de la opción *Automatically Eject Each Filled Page* que también está disponible en *Page/Output Setup*. Cabe señalar que esta opción sólo es aplicable localmente, es decir, durante la sesión y módulo actualmente en uso, mientras que en la ventana *Page/Output Setup*, la opción *Print/eject pages after each printout* es global, es decir, se aplica permanentemente en cada módulo.

**All data:** Este botón selecciona automáticamente todas las variables y casos del archivo para imprimirlos.

**Setup:** Al presionar este botón se despliega la ventana *Text Output Printer Setup* donde se puede seleccionar la calidad de la impresión así como el número de copias.

Botón asociado: 

**PRINT FILE:** Permite imprimir en lotes o paquetes, gráficas y scrollsheets guardados previamente en un archivo. Es muy útil cuando se desean imprimir posteriormente las



gráficas y scrollsheets que se acaban de crear, simplemente se deben guardar (mediante la opción *Save*) después de haberlas creado, tiempo después, seleccionar esta opción para abrir la ventana *Print Graph and Scrollsheet Files*, elegir los scrollsheets y gráficas deseados y pulsar el botón **OK** para empezar el proceso de impresión por lotes. Al pulsar **OK** se imprimirán primero las gráficas elegidas (una en cada página, en el orden en que fueron seleccionadas) y después los scrollsheets juntos (en el orden en que fueron seleccionados).

**Ventana *Print Graph and Scrollsheet Files*:**

**List of files to be printed:** Despliega la lista de los archivos que se pueden imprimir, al momento de imprimir uno, su nombre es borrado de la lista.

**Add graph/scrollsheet:** Al presionar estos botones se abre la ventana *Select File to be Printed* que permite agregar nombres a la lista de archivos a imprimir.

**Remove:** Este botón borra el archivo seleccionado (resaltado) en la lista de archivos a imprimir.

**Cancel:** Al presionar este botón, se interrumpe el proceso de impresión por lote, por lo que sólo se imprimirá la última gráfica o scrollsheet completo que se haya enviado al buffer.

**PAGE/OUTPUT SETUP:** Antes de imprimir cualquier spreadsheet o scrollsheet se debe elegir el destino de salida así como otras opciones de impresión que se despliegan al abrir la ventana *Page/Output Setup*, la cual también se encuentra disponible desde la opción *Print* del menú *Options* o haciendo doble clic en el campo *Output* de la barra de estado. Todas las opciones que se activen en esta ventana se tomarán por omisión y se aplicarán al iniciar STATISTICA y abrir cualquier módulo. Ventana *Page/Output Setup*:

**Page/Output Setup**

**Output**

Off

File (Text)

Printer

Window

**Suppl. Info**

Brief

Short

Medium

Long

**Page #**

10

**OK** **Cancel**

**Lines**

1

**Font** **Margins**

**Output File (text):** C:\STAT\EXAMPLES\JEJEMPLO3.TXT

**Output Header:**

Date & Time  Center Titles

Automatically Print/Eject Pages after Each Printout

Automatically Eject Each Filled Page

Automatically Print All Scrollsheets (Auto-report)

Automatically Print All Graphs



**Output:** Seleccionar hacia dónde será dirigida la salida:

**Off:** La información permanecerá en memoria y no será impresa o enviada a un archivo de salida.

**File (Text):** Al seleccionar esta opción y presionar **OK** se abrirá la ventana *Save Output File As* donde se debe determinar el nombre del archivo de texto donde va a ser enviado el contenido del spreadsheet (para imprimirlo posteriormente si se desea).

**Printer:** Envía el contenido del scrollsheet o el spreadsheet directamente a la impresora.

**Window:** Al elegir esta opción, independientemente del destino de la salida seleccionado en las opciones anteriores (i.e. *File* o *Printer*), ésta será direccionada también a la ventana text/output.

**Supplementary Information:** Seleccionar la cantidad de información suplementaria que se desea imprimir.

**Brief:** Da un formato "económico" a la salida ya que permite imprimir sólo los contenidos de los spreadsheets o scrollsheets seleccionados, sin encabezados de página, de trabajo, o cualquier información acerca de las variables o condiciones específicas del análisis.

**Short:** Este estilo de salida incluirá además del contenido del spreadsheet o scrollsheet, el encabezado de página (nombre del módulo, fecha/hora, número de página), nombre del actual archivo de datos, información relativa a las condiciones de selección de casos o variables de peso (si es que hay), el encabezado de salida (*output header*, ver abajo), una lista con los nombres (cortos) de todas las variables seleccionadas para cada análisis y el código de falta de dato.

**Medium:** Este formato incluye además de lo anterior, los nombres largos o etiquetas de las variables, reservando una línea de salida para cada variable.

**Long:** Además de toda la información dada por el formato medio, este estilo incluye además, para cada variable seleccionada, una lista completa de todos sus valores dobles, es decir, valores numéricos y alfanuméricos y/o valores con etiquetas o comentarios.

(Recordar que las elecciones hechas con la opción *Print* (descrita anteriormente) tienen prioridad sobre las opciones aquí establecidas si es que llegan a entrar en conflicto).

**Output file (text):** Al presionar este botón, se abrirá la ventana *Specify Output File*. Aquí se debe indicar el nombre del archivo donde se va a guardar la información del spreadsheet o scrollsheet que ha sido direccionada previamente a un archivo de salida en la opción *Output* (ver arriba). Si actualmente se encuentra abierto un archivo, esta opción lo cerrará y dirigirá la salida al nuevo archivo que se le indique.



**Page#:** Indicar el número de la página desde donde se quiere que comience la impresión, las páginas siguientes serán numeradas secuencialmente tomando como base este número. Las páginas no podrán ser numeradas secuencialmente si se está corriendo simultáneamente más de un módulo de STATISTICA.

**Lines:** Permite escoger diferentes bordes para las tablas de salida.


**Output header:** Especificar el encabezado que será impreso en cada página.

**Date and time:** Hacer clic en esta opción si se quiere incluir la fecha y hora en el encabezado de salida.

**Center titles:** Centra el encabezado de salida.

**Font:** Permite elegir el tipo y tamaño de la letra con la que será impresa la información de salida.

**Margins:** Aquí se especifican los márgenes de página para el texto de salida del spreadsheet, scrollsheet o la ventana text/output.

**Automatically print/eject pages after each printout:** Si esta opción está activa, la información automáticamente será dirigida a la impresora o al archivo de salida (dependiendo de las elecciones anteriores) cuando se presione el botón **OK** en las ventanas: *Print Data* o *Print Scrollsheet*, o el botón de impresión de la barra de herramientas:  Cada impresión empezará en una nueva página. Si la opción se encuentra desactivada, la salida será almacenada en el buffer interno y enviada a la impresora dependiendo del estado actual de la opción *Automatically Eject Each Filled Page*.

**Automatically Eject Each Filled Page:** Cuando esta opción está activa, STATISTICA almacenará los resultados del análisis en su buffer interno hasta que se tenga la suficiente información para llenar una página. Cuando ésta esté llena, será enviada al destino de salida elegido anteriormente (*File (text)*, *Printer*, *Window*). STATISTICA continuará almacenando los resultados y enviándolos a la impresora o al archivo una página a la vez. Se puede usar la opción *Print/eject current pages* en cualquier punto de un análisis para mandar a imprimir todo lo que se haya acumulado en el buffer hasta ese momento (sin importar si se ha completado o no la página). Cuando esta opción se encuentra desactivada, la impresión tomará menos tiempo en promedio: STATISTICA almacenará todos los resultados de un análisis y no los imprimirá o enviará a un archivo hasta que se haya cerrado el módulo en uso o se seleccione la opción *Print/eject current pages*.

**Automatically print all scrollsheets (Auto-report):** Si este modo de impresión se encuentra activo, entonces, además de la información suplementaria (dada en la opción *Supplementary information*) toda la salida de un análisis será enviada automáticamente a la dirección señalada previamente (impresora, archivo y/o ventana text/output). Si la opción es desactivada, entonces, los scrollsheets (o sus partes) serán impresos sólo en el momento que se requiera, seleccionando la opción *Print* o presionando el botón de impresión de la barra de herramientas.



**Automatically print all graphs:** Al seleccionar esta opción, STATISTICA automáticamente imprimirá (o guardará en un archivo) las gráficas tal como se vayan creando. Si la opción está inactiva, se tendrá que usar la opción *Print graph* o el correspondiente botón de la barra de herramientas de gráficas para imprimir únicamente las gráficas elegidas (las opciones de impresión de gráficas se explican en el capítulo correspondiente).

**PRINTER SETUP:** Al seleccionar esta opción se abrirá la ventana *Printer Setup*, donde se elige la impresora (**Printer**), la orientación de las páginas que se van a imprimir (**Orientation**) y el tamaño del papel así como la fuente de alimentación del mismo (**Paper**). Hay que notar que se pueden establecer opciones diferentes para la impresión de gráficas y texto.

**EXIT:** Finaliza la sesión con STATISTICA.

### 3.1.2 EDIT

Contiene opciones referentes a la edición y formato de cada uno de los datos.

**CELL:** Edita la celda seleccionada (resaltada) en el spreadsheet. Otra manera de editar una celda es escribiendo directamente en ella cuando esté resaltada o bien, haciendo doble clic en ella.

**VARIABLES:** Despliega todas las opciones disponibles para la edición de variables las cuales se explican en la sección Modificar Variables del capítulo 2.

Botón asociado: 

**CASES:** Despliega todas las opciones disponibles para la edición de los casos que fueron descritas en la sección Modificar Casos del capítulo 2.

Botón asociado: 

**HEADER:** Despliega la ventana *Data File Header and Info*, donde se especifican encabezados, notas o comentarios del archivo de datos:

**On line header:** Permite escribir una línea de hasta 77 caracteres que aparece como el encabezado del archivo de datos y es guardada como parte de éste, pudiéndose imprimir al seleccionar la opción *File Info/Notes* de la ventana *Print data* o al hacer un análisis y elegir un estilo de reporte de la ventana *Page/Output Setup*. Ambas ventanas fueron descritas anteriormente en las opciones de impresión.

**File Information/Notes:** Permite escribir notas o comentarios acerca de la información que contiene el archivo. Estas notas pueden tener hasta 512 caracteres de longitud y pueden incluir cualquier carácter imprimible.



Las notas y comentarios se pueden imprimir con el archivo de datos al seleccionar la opción *File Info/Notes* de la ventana *Print Data* (descrita en las opciones de impresión).

**CUT:** Borra el contenido de las celdas seleccionadas (resaltadas) y las reemplaza con un código de falta de dato hasta que se les den nuevos valores, también borra texto y gráficas copiando en el portapapeles (clipboard) de Windows todo lo que elimina.

Botón asociado:

**COPY:** Copia el contenido de las celdas seleccionadas (resaltadas), texto de la ventana *text/output* u objetos gráficos, al portapapeles (clipboard) de Windows.

Para el *scrollsheet* y *spreadsheet* copiará no sólo el contenido de las celdas seleccionadas sino también los nombres de renglón y columna asociados a esas celdas. (Para copiar sólo las celdas seleccionadas sin los nombres de renglones y columnas asociados, usar la opción *Copy raw*).

Botón asociado:

**COPY RAW:** Copia el contenido de las celdas seleccionadas (resaltadas) al portapapeles (clipboard) de Windows .

**PASTE:** Pega (inserta) el contenido actual del portapapeles (clipboard) de Windows al *spreadsheet*, *scrollsheet*, gráfica o a la ventana *text/output* en el lugar donde se encuentre el cursor. Esta opción no está disponible si el portapapeles de Windows está vacío o contiene un formato incompatible. Por ejemplo, pegar una gráfica en una celda del *scrollsheet*.

Botón asociado:

**CLEAR:** Borra texto previamente seleccionado o el contenido de las celdas seleccionadas (resaltadas) en el *spreadsheet* o *scrollsheet* reemplazando los valores eliminados con un código de falta de dato. A diferencia de *cut*, esta opción no copia los datos borrados en el portapapeles de Windows.

**PASTE LINK:** Esta opción permite establecer una liga para el intercambio dinámico de datos *DDE (Dynamic Data Exchange)* entre otra aplicación (servidor) y *STATISTICA* (cliente) de tal forma que los cambios hechos al archivo fuente (servidor) sean actualizados automáticamente en el archivo cliente (*spreadsheet* de *STATISTICA*).

Se pueden copiar celdas de una aplicación de windows (por ejemplo Excel) a un bloque de celdas del *spreadsheet* de *STATISTICA* ligando permanentemente los datos del *spreadsheet* con el archivo fuente o servidor.

Al usar esta opción, la liga se crea y guarda en el espacio reservado para el nombre largo (*long name*) de la variable actual (resaltada). El nombre largo se especifica al elegir la opción *Current Specs*. (ver sección *Modificar Variables* en el capítulo 2).



Para establecer una liga DDE:

- 1.- Abrir el archivo fuente y copiar en el portapapeles de Windows el bloque de datos (o texto) que se quiere ligar con el archivo cliente. Dejando abierto el archivo fuente, activar el módulo de STATISTICA que se va a manejar.
- 2.- En el spreadsheet seleccionar (resaltar) la celda a partir de la cuál, hacia abajo y a la derecha va a ser insertado el bloque .
- 3.- Elegir la opción Paste Link para crear la liga.

Las ligas también se pueden crear y editar desde el manejador de ligas (*Links manager*), el cual se describe en la última opción del menú *Edit*.

**FIND:** Permite buscar un valor específico de la variable seleccionada (resaltada) o un nombre de un caso en el spreadsheet o un texto o número específico en la ventana text/output.

**REPLACE:** Permite buscar y reemplazar un valor específico de la variable seleccionada (resaltada) o un nombre de un caso en el spreadsheet o un texto o número específico en la ventana text/output.

La ventana que despliegan estas dos opciones son similares:

**Find what:** Se debe especificar el valor de la variable (numérico o alfanumérico) o el nombre del caso que se busca en el spreadsheet o el texto o número que se quiere encontrar en la ventana text/output.

**Direction:** Se debe especificar la dirección de búsqueda: hacia arriba o hacia abajo a partir de la celda seleccionada (resaltada). La búsqueda se detendrá cuando llegue al final del archivo (en ambas direcciones).

**Search:** Esta opción es válida sólo en la búsqueda de valores de variables o nombres de casos en el spreadsheet.

**Find next:** Ya que se han especificado las opciones anteriores, hacer clic en este botón para iniciar la búsqueda.

**Replace:** Cuando STATISTICA encuentre el valor, nombre del caso o texto especificado aparecerán las siguientes opciones a elegir:

**Yes:** Aceptar el reemplazo.

**No:** Rechazar el reemplazo.

**Yes to all:** Reemplaza todo lo que concuerde con lo especificado, todas las veces que lo encuentre y sin pedir confirmación.

**Find:** Este botón muestra la siguiente localización del valor buscado cuando hay varios lugares donde se encuentra y ya se ha encontrado por lo menos el primero. La búsqueda se puede repetir pulsando Ctrl + F.



**Replace with:** Se debe especificar el nuevo valor con que se va a reemplazar lo que se busca.

**Replace all:** Este botón reemplaza todo lo que concuerde con lo especificado, todas las veces que lo encuentre y sin pedir confirmación.

**TRANSPOSE BLOCK:** Transpone un bloque cuadrado de datos del spreadsheet previamente seleccionado. (No puede transponer el archivo completo).

**TRANSPOSE DATA FILE:** Transpone el archivo completo transformando los casos en variables y viceversa. Los valores alfanuméricos de las variables se perderán. Los nombres de los casos (si hay) del archivo original serán ahora los nombres largos de las variables (*Long variable names*) y los nombres de los casos del archivo transpuesto estarán formados por los nombres de las variables (8 caracteres) y hasta 12 caracteres tomados del nombre largo (en total 20 caracteres).

**FILL BLOCK:** Rellena un bloque de celdas.

**FILL DOWN:** Copia los contenidos de las celdas del primer renglón de un bloque seleccionado en el spreadsheet en las celdas restantes del bloque.

**FILL RIGHT:** Copia los contenidos de las celdas de la primera columna de un bloque seleccionado en el spreadsheet en las celdas restantes del bloque.

**SCREEN CATCHER:** Esta utilidad permite seleccionar cualquier área rectangular de la pantalla (o la pantalla completa) y copiarla como un mapa de bits en el portapapeles de Windows. El área copiada puede tener gráficas, scrollsheets, spreadsheets o incluso ventanas de otras aplicaciones (si están en pantalla en el momento de activar esta utilidad). El mapa de bits puede ser insertado en cualquier aplicación que soporte este formato gráfico. Para obtener información acerca de éste y otros formatos gráficos, consultar la opción *Save bitmap* descrita en la sección Menús y Barra de Herramientas del capítulo 4.

**LINKS :** Esta opción abre la ventana *Links Manager* la cual permite la edición de las ligas que han sido establecidas entre otra aplicación y STATISTICA (ver *Paste link*).

Opciones de la ventana *Links Manager*:

**Links:** Despliega todas las ligas que se han creado hasta el momento. Se puede editar cualquier liga seleccionándola (resaltándola) y dando un clic al botón *Change Link* (descrito al final).

**Update:** Despliega el estado de la liga resaltada en la opción anterior. Este puede ser automático (el símbolo @ precede al enunciado) o manual (el símbolo ^ precede al enunciado). Estado automático significa que cualquier cambio en los datos ligados del archivo fuente (servidor) será efectuado automáticamente en los datos ligados del archivo de STATISTICA (cliente). Estado manual significa que la liga está deshabilitada temporalmente y se tendrán que actualizar manualmente los datos del archivo cliente usando el botón *Update Now* o cambiando al estado automático con el botón *Change Link*.





**Update now:** Presionando este botón se actualizan los datos ligados del archivo de STATISTICA (cliente). Cuando existen varias ligas, se actualizan los datos de la liga resaltada en la ventana *Links*.

**Cancel link:** Este botón deshabilita temporalmente la liga seleccionada (resaltada) en la ventana *Links*; para habilitarla nuevamente se presiona el botón *Update Now*.

**Nota:** Es recomendable usar esta opción cuando se está corriendo un análisis para deshabilitar la liga y evitar cambios imprevistos en los archivos ligados.

**Change Link:** Este botón remite a la ventana *Current Specs* donde se puede editar la liga dentro del cuadro *Long Name*.

Las ligas tienen la siguiente sintaxis:

1.- **@<Aplicación fuente> |**: Especifica el nombre de la aplicación fuente o servidor. Esta debe estar separada del resto del enunciado por un |. El primer símbolo indica el estado de la liga (@: automático, ~: manual). Estado automático significa que cualquier cambio en los datos ligados del archivo fuente (servidor) será efectuado automáticamente en los datos ligados del archivo de STATISTICA (cliente). Estado manual significa que la liga está deshabilitada temporalmente y se tendrán que actualizar manualmente los datos del archivo cliente usando el botón *Update now* o cambiando al estado automático con el botón *Change link*.

2.- **Ruta!**: Designa el nombre y ruta del archivo ligado de la aplicación fuente o servidor, debe separarse del resto del enunciado con un !.

3.- **Item;**: Indica la celda, rango o algunos otros datos específicos de la aplicación fuente o servidor, finaliza con un ";".

4.- **Columna; renglón**: Esta parte es opcional, indica la celda a partir de la cual serán insertados los datos en el archivo cliente cuando sea creada la liga. Los datos se insertan desde esa celda, hacia abajo y a la derecha. STATISTICA toma por omisión la celda que se encuentra resaltada en el spreadsheet al momento de elegir la opción *Paste link* y no lo indica en el enunciado de la liga.

Ejemplo: @Excel | c:\Excel\libro1.xls!l2c2:l4c4;c1;r1

### 3.1.3 VIEW

Contiene opciones que permiten ver características tanto de la pantalla como de los archivos de datos.

**TEXT VALUES:** Muestra/oculta los valores alfanuméricos de las variables en el spreadsheet (al ocultarlos mostrará su equivalente numérico).

Botón asociado: 



**CASE NAMES:** Muestra/oculta los nombres de los casos en el spreadsheet. Si la opción está precedida por una "✓" muestra los nombres, de lo contrario, muestra únicamente los números de los casos.

Botón asociado: 

**FULL VAR NAMES:** Cuando el formato de despliegue de valores de una variable es menor que 8 caracteres (la máxima longitud para el nombre de una variable) el nombre de la variable aparece cortado. Esta opción permite desplegar los nombres completos de todas las variables en el spreadsheet (cuando está precedida por una "✓") ajustando automáticamente el ancho de las columnas que lo requieran.

**FONT:** Despliega la ventana *Font* que permite elegir el tipo (**Font**), estilo (**Font style**) y tamaño de letra (**Font size**) que se usará en el spreadsheet, scrollsheets y otras utilerías de STATISTICA, mostrando un ejemplo a medida que se elige cada tipo. Además, dichas selecciones pueden guardarse como predeterminadas mediante el botón **Save as default**. La opción **Apply**: permite establecer un tamaño de letra no definido en el tipo elegido, pudiendo revisar los resultados en la ventana de muestra (*View box*) antes de ser aceptados.

**COLORS:** Personaliza los colores del spreadsheet y scrollsheets a través de la ventana *Customize Colors*. Cada una de las opciones disponibles en esta ventana abrirá la paleta de colores donde se pueden seleccionar los que se deseen e incluso personalizar un color determinado (con el botón *Define custom colors*). Los cambios efectuados se pueden revisar en la ventana muestra que se encuentra abajo de las opciones.

Opciones:

**Screen element:** Seleccionar el elemento del scrollsheets/spreadsheet que se desea personalizar:

- Info Box (La esquina superior izquierda de la hoja)
- Titles (Títulos)
- Toolbar (Barra de herramientas)
- Column Names (Nombres de columnas)
- Row Names (Nombres de renglones)
- Data Area (Área de datos)
- Highlight (Resaltado)

2:3456 Toolbar					
Info Box	Titles				
	Col1	Col2	Col3	Col4	Col5
Row1	Aa	1.2	0.02	4.10	5.10
Row2	Bb	2.3	0.03	4.20	5.20
Row3	Cc	3.4	0.04	4.30	5.30

Para cada uno de estos elementos se puede personalizar el color del texto con el botón **Text** o el color de fondo mediante el botón **Backgr**.

**Lines:** Personaliza los colores de las líneas del spreadsheet/ scrollsheets.

**Main:** Cambia el color de las líneas que separan los nombres de los renglones y columnas.



**Grid Vert (Horz):** Cambia el color de las líneas verticales (horizontales) que separan los valores de los datos.

**Defaults:** Si se desea que los colores personalizados sean guardados como los colores por omisión o predeterminados, hacer clic en la opción **Save as default**. Para restaurar los colores que STATISTICA tiene por omisión, presionar el botón **Use System Defaults**.

**TEXT/OUTPUT WINDOW:** Esta opción permite abrir la ventana text/output descrita en la sección Convenciones Generales del capítulo 1. Otra forma de abrirla es seleccionando la opción *Window* en la ventana *Page/Output Setup* que se despliega en las opciones de impresión.

**TITLES:** Permite mostrar/ocultar el título del spreadsheet o el scrollsheet que se definió en la ventana *Data File Hider and Info* como el encabezado del archivo. Este título se puede editar haciendo doble clic en esa área del spreadsheet para desplegar la ventana antes mencionada.

**TOOLBAR:** Muestra/Oculta la barra de herramientas.

**CHANGE SHOW:** Despliega el contenido de la celda seleccionada o resaltada (o la posición del cursor en las gráficas) en el cuadro de muestra que aparece en la esquina superior izquierda de la barra de herramientas:

2.99691763054292

Cada vez que se selecciona esta opción, el cuadro de muestra cambia al siguiente de los tres tamaños que tiene definidos.

En el spreadsheet (o scrollsheet), cuando el formato de despliegue de los valores de una variable (o el ancho de las columnas del scrollsheet) no permite observar todas las cifras significativas, se puede usar esta opción para cambiar a otro tamaño del cuadro de muestra que permita ver el valor completo.

En el caso de las gráficas, el cuadro de muestra indica en qué coordenadas se encuentra posicionado el cursor. Dependiendo del modo gráfico en el que se esté trabajando: Dinámico o fijo (el cual se describe en el capítulo dedicado a gráficas), la ventana mostrará las coordenadas cartesianas (x,y) partiendo del origen y la escala de la gráfica trazada (modo fijo), o bien, mostrará coordenadas basadas en el tamaño de la ventana que despliega la gráfica como porcentajes del largo y ancho de la ventana, así, la esquina inferior izquierda tiene los valores de (0%,0%) y la esquina superior derecha (100%,100%). En cualquier caso, esta opción se puede activar directamente haciendo clic en el cuadro de muestra de la barra de herramientas.

**STATUS BAR:** Muestra/Oculta la barra de estado (la barra en la parte inferior de la pantalla). Mientras se ejecuta una tarea, STATISTICA la usará para indicar el progreso de ésta (ver barra de progreso, abajo).



La Barra de estado consta de las siguientes partes:

**Comentario:** Esta área está reservada para dar una breve descripción (en una línea) del actual comando seleccionado (resaltado). Cuando no hay nada seleccionado (los datos no están siendo procesados y STATISTICA está esperando el siguiente comando) aparece la palabra Ready (Listo).

**Output:** Indica hacia qué salida (*Off*, *File* o *Printer*) serán direccionados los datos de un análisis. Haciendo doble clic en esta área de la barra, se despliega la ventana *Page/Output Setup* descrita en las opciones de impresión. Ahí se puede seleccionar el tipo de salida además de otras opciones.

**Sel:** Despliega el estado actual (*On*:Activas, *Off*:Inactivas) de las condiciones de selección de casos. Haciendo doble clic en esta área de la barra se despliega la ventana donde se especifican estas condiciones (*Case Selection Conditions*) descrita en la sección Selección de Casos del capítulo 2.

**Weight:** Despliega el estado actual (*On*:Activas, *Off*:Inactivas) de las condiciones de peso de las variables. Haciendo doble clic en esta área de la barra se despliega la ventana *Define Weight* (donde se especifican las condiciones) descrita en la sección Variables de Peso en el capítulo 2.

La barra de progreso consta de las siguientes partes:

**Progress:** Esta área despliega el avance en la tarea que se está llevando a cabo.

**Elapsed Time:** Indica el tiempo transcurrido en segundos desde que se inició la tarea.

**Cancel:** Presionando este botón se interrumpe la tarea que se está ejecutando.

**SHORT/FULL MENUS:** Oculta/Muestra las combinaciones de teclas equivalentes a cada una de las opciones de los menús que aparecen después del nombre de la opción (*Full menus* es la opción activa por omisión).

### 3.1.4 ANALYSIS

Las opciones contenidas en este menú permiten mostrar todos los procedimientos de análisis disponibles en STATISTICA y en el módulo activo.

**STARTUP PANEL:** Despliega el panel de inicio. Siempre que se inicia un módulo se despliega su correspondiente panel de inicio; éste permite seleccionar un procedimiento de análisis de los que se encuentran disponibles en el módulo activo (cuando el módulo contiene varios métodos a elegir) o bien, permite definir las condiciones iniciales para el análisis. El panel de inicio también se encuentra disponible presionando las teclas Control+S.

**RESUME ANALYSIS:** Cuando el método de análisis que se ha ejecutado se encuentra activo todavía pero su ventana se encuentra minimizada en la parte inferior izquierda de



la pantalla, esta opción permite abrirla nuevamente para realizar otro análisis. También es factible hacer doble clic en el ícono de la ventana minimizada para abrirla nuevamente.

**OTHER STATISTICS:** Despliega el menú principal del programa (*STATISTICA: Module Switcher*) descrito en el capítulo 1.

Botón asociado: 

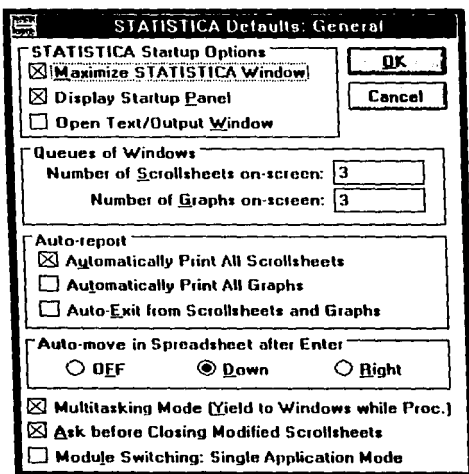
### 3.1.5 OPTIONS

Este menú agrupa un conjunto de opciones que permiten personalizar la ventana en la que se está trabajando, así como también la forma de impresión, la creación de gráficas y las condiciones de selección de variables y casos.

**SELECT:** Permite establecer condiciones de selección de casos para el archivo de datos que se está manejando, estas condiciones se especifican en la ventana *Case Selection Conditions* descrita en la sección Selección de Casos del capítulo 2.

**WEIGHT:** Permite ajustar la contribución de ciertos casos individuales al resultado de un análisis, ponderando aquellos casos en proporción a los valores de una variable seleccionada. Esta variable se define en la ventana *Define Weight* descrita en la sección Variables de Peso del capítulo 2.

**GENERAL:** Permite personalizar la ejecución y despliegue general de *STATISTICA* activando las diferentes opciones que aparecen en la ventana *STATISTICA Defaults: General*. Las selecciones hechas aquí, serán tomadas como las condiciones de operación por omisión siempre que se inicie *STATISTICA*:



**STATISTICA Defaults: General**

**STATISTICA Startup Options**

- Maximize STATISTICA Window
- Display Startup Panel
- Open Text/Output Window

**Queues of Windows**

Number of Scrollsheets on-screen: 3

Number of Graphs on-screen: 3

**Auto-report**

- Automatically Print All Scrollsheets
- Automatically Print All Graphs
- Auto-Exit from Scrollsheets and Graphs

**Auto-move in Spreadsheet after Enter**

OFF     Down     Right

- Multitasking Mode (Yield to Windows while Proc.)
- Ask before Closing Modified Scrollsheets
- Module Switching: Single Application Mode

OK Cancel

**STATISTICA startup options.**

**Maximize STATISTICA window:** Seleccionar esta opción si se quiere que la ventana de STATISTICA se maximice al iniciar la sesión.

**Display startup panel:** Seleccionar esta opción si se desea que el panel de inicio sea desplegado siempre que se abra un módulo.

**Open text/output window:** Seleccionar esta opción si se desea que la ventana de texto se abra al mismo tiempo que se inicia un módulo.

**Queues of windows:** Permite definir la cantidad de ventanas que pueden permanecer abiertas en pantalla, es decir, el tamaño de la cola. En STATISTICA, los scrollsheets y gráficas son manejados automáticamente en colas separadas con la estructura FIFO (*First In-First Out*) primero en entrar-primero en salir. Se debe especificar un entero entre 0 y 20, sin embargo, si se especifica 0, el tamaño de la cola será ilimitado (se necesitará cerrar manualmente algunas ventanas para evitar problemas de saturación de memoria).

**Number of scrollsheets (graphs) on-screen:** Especificar el máximo número de scrollsheets (gráficas) que pueden permanecer abiertos en pantalla.

En las opciones *Sscrollsheet manager* y *Graph manager* del menú *Window* hay mas información referente a este tópico.

**Auto-Report:** Las siguientes opciones permiten imprimir automáticamente scrollsheets o gráficas:

**Automatically print all scrollsheets (graphs):** En estas opciones STATISTICA automáticamente imprimirá (o guardará en un archivo, dependiendo de la elección hecha en la ventana *Page/Output setup* de las opciones de impresión) el contenido de los scrollsheets (gráficas) en el orden que fueron creados, de lo contrario, únicamente se imprimirán los scrollsheets (gráficas) seleccionados mediante la opción *Print*.

**Auto-exit from scrollsheets and graphs:** Cuando los scrollsheets o gráficas son creados en paquete como resultado de un análisis, esta opción suprime el botón *Continue* cuando la cola está llena, impidiendo de esta forma que se despliegan todas las ventanas que generó el análisis. Sin embargo, al momento de imprimir, cuando se seleccionan las opciones de impresión automática de scrollsheets y gráficas (vistas arriba). Las gráficas o scrollsheets serán impresos en paquete; sin interrupción.

**Auto-move in spreadsheet after Enter:** Al introducir o editar datos en el spreadsheet, esta opción permite controlar el movimiento del cursor después de presionar la tecla <Enter>.

**Down:** Mueve el cursor a la siguiente celda de la columna.

**Right:** Mueve el cursor a la siguiente celda del renglón.

**Off:** El cursor no se mueve después de haber presionado <Enter>.



**Multitasking mode (yield to windows while processing):** Esta opción permite cambiar a otras aplicaciones de Windows o a diferentes módulos de STATISTICA, mientras se estén procesando los datos de un análisis (presionando las teclas Ctrl + Esc que nos muestran la lista de tareas o Alt + Tab para pasar de una aplicación a otra). Si se está procesando un archivo de datos muy grande que requiere un poco más de tiempo que el promedio, esta opción permitirá usar otro programa o aplicación mientras se está esperando. Sin embargo, si la opción está inactiva los datos son procesados un poco más rápido.

**Ask before closing a modified scrollsheet:** Seleccionar esta opción si se desea que aparezca un mensaje de advertencia cada vez que se va a cerrar un scrollsheet que se ha modificado previamente. El mensaje da la oportunidad de guardar los cambios antes de cerrarlo.

**Module switching: Single application Mode:** Al seleccionar esta opción, cada vez que se quiera cambiar a otro módulo (vía *Module Switcher*) no se abrirá una nueva ventana de aplicación. Cada nuevo módulo será abierto en la misma ventana, reemplazando el módulo anterior. Si la opción está desactivada, entonces, se podrá cambiar a los otros módulos sin cerrar los anteriores.

**DISPLAY:** Permite personalizar la pantalla de STATISTICA, mediante la ventana *STATISTICA Defaults: Display*. Las selecciones hechas aquí serán usadas por omisión siempre que se inicie una sesión. Cabe señalar que el color y tipo de letra que se elijan aquí no afectarán el spreadsheet o scrollsheet actualmente en uso (para esto, ver las opciones respectivas del menú *View*), solamente afectarán los nuevos que sean creados (vía *New Data*). Entre las opciones que se pueden activar se encuentran: La barra de estado (**Status bar**), menús completos (i. e. menús con teclas de método abreviado: **Fulls menus**), iconos en menús (**Icons in menus**), iconos en ventanas (**Icons in dialog boxes**); además permite cambiar el tipo de letra (**Fonts**) y los colores (**Colors**) predeterminados en el spreadsheet, scrollsheet y gráficas.

**PRINT:** Despliega la ventana *Page/Output Setup* descrita en las opciones de impresión.

**STATS GRAPHS:** Abre la ventana *STATISTICA Defaults: Stats Graphs*, la cual se describe en el capítulo 4.

**USER GRAPHS:** Esta opción despliega la ventana *Modify the List of User-defined Graphs*, que al igual que la ventana anterior, es descrita en el capítulo 4.

### 3.1.6 MACRO

Este menú contiene opciones relativas al manejo de macros.

**RECORD:** Permite grabar una macro, al seleccionar esta opción se abre la ventana *Define New Macro* (también se puede abrir presionando las teclas Control + F3) donde se debe especificar el nombre y una tecla llamada *hot-key* que al presionarla junto con la tecla <Control> permitirá ejecutar la macro cada vez que se pulse esta combinación de teclas.



Una vez que se han definido las opciones que aparecen en esta ventana, **STATISTICA** empezará a grabar las acciones que se lleven a cabo y terminará al presionar nuevamente las teclas Control + F3.

Ventana *Define New Macro*:

**Name:** Especificar un identificador o comentario de máximo 40 caracteres de longitud.

**Key: Ctrl + letter:** Escoger una letra (que será usada junto con la tecla <Ctrl>) para crear la hot-key de la macro que servirá para ejecutarla en cualquier momento dentro del programa.

**Fast playback:** Esta opción permite definir la velocidad de ejecución de la macro, si se encuentra activa, ésta será muy rápida, de lo contrario, se ejecutará a la misma velocidad en que fue grabada.

**Record mouse:** Al estar creando la macro, permite grabar movimientos del ratón junto con teclas, o si está desactivada, graba únicamente las teclas que se pulsen. Debe tomarse en cuenta que las macros que implican uso del ratón dependen del contexto en que se grabó dicho movimiento, por lo tanto, se pueden producir resultados diferentes si el tamaño o posición de la ventana en la pantalla cambia. Es recomendable crear macros sin usar ratón para que éstas sean ejecutadas de una manera más confiable.

**RUN:** Si se conoce la combinación de teclas que ejecutan la macro (Ctrl + letra), ésta será ejecutada automáticamente presionando dichas teclas. De lo contrario, esta opción despliega la ventana *Run Macro* con una lista de todas las macros grabadas hasta este momento, pudiendo seleccionar la que se desee ejecutar de acuerdo al modo que se indique:

**Default:** Se ejecuta a la velocidad definida en la ventana *Define New Macro* (en la opción *Fast playback*).

**Normal:** Se ejecuta a la misma velocidad en que fue grabada.

**Fast:** Se ejecuta a una velocidad superior a la que fue grabada.

**REVIEW:** Esta opción permite revisar información específica acerca de una macro a través de la ventana *Review Macros*:

**List of macros:** Seleccionar la macro que se quiere revisar.

**Macro Info:** Al presionar este botón se despliega una ventana con la información relativa a la macro que se ha seleccionado. Esta información incluye el nombre o comentario de la macro, la letra que se combina con la tecla <Ctrl> para ejecutar la macro, el modo de ejecución y si se grabaron o no movimientos del ratón.





### 3.1.7 WINDOW

Este menú contiene opciones relativas al despliegue de las diferentes ventanas de STATISTICA.

**SCROLLSHEET MANAGER / GRAPH MANAGER:** Esta opción permite especificar el número de scrollsheets (gráficas) que pueden permanecer abiertas en pantalla mediante la ventana *Scrollsheet Manager (Graph Manager)*. En STATISTICA, scrollsheets y gráficas son manejados automáticamente en colas separadas con la estructura FIFO (*First In-First Out*) primero en entrar-primero en salir. Por ejemplo: Si el tamaño de la cola para los scrollsheets se definió como 3, al crear el 4º, el primero se cerrará automáticamente.

Ventana *Scrollsheet (Graph) Manager*:

**Queue:** Especificar el número máximo (0 a 20) de ventanas que pueden permanecer abiertas en pantalla. Si se indican cero, el tamaño de la cola será ilimitado, pudiendo ocasionar problemas de saturación de memoria al tener muchas ventanas abiertas (se deberán cerrar manualmente algunas ventanas para evitar esto).

**Discardable windows:** Contiene una lista de todas las ventanas abiertas en pantalla.

**Locked windows:** Contiene una lista de todas las ventanas minimizadas en pantalla. Para borrar de pantalla una ventana minimizada debe cerrarse manualmente.

**Lock:** Este botón permite minimizar las ventanas en pantalla y borrarlas de la cola de ventanas abiertas. Primero se selecciona (de la lista que aparece en *Discardable windows*) la ventana que se quiere cerrar y después se presiona este botón. La ventana minimizada permanecerá en pantalla hasta que sea cerrada manualmente o se salga del módulo o de STATISTICA.

**CASCADE:** Esta opción acomoda las ventanas abiertas de STATISTICA sobreponiéndolas de tal forma que la barra de título de cada ventana permanezca visible.

**TILE HORIZONTALLY (VERTICALLY):** Esta opción acomoda las ventanas abiertas de STATISTICA en un patrón horizontal (vertical) ajustando su tamaño de acuerdo al espacio de que disponga. Las gráficas conservarán su tamaño normal a no ser que se haya desactivado la opción *Fixed Proportions* en las opciones de gráficas (que se describen en el capítulo 4).

**ARRANGE ICONS:** Esta opción acomoda en una fila o renglón todas las ventanas minimizadas.

**CLOSE ALL WINDOWS:** Esta opción cerrará todas las ventanas abiertas en STATISTICA menos el spreadsheet. Esta opción es de utilidad cuando se necesita limpiar pantalla para empezar un nuevo análisis.



En la parte inferior del menú *Window* se despliegan los nombres de las ventanas que se encuentran abiertas o minimizadas en la pantalla de STATISTICA, el nombre de la ventana activa aparece señalado con una "✓".

### 3.1.8 HELP

Como su nombre lo indica, este menú contiene opciones que permiten desplegar la ayuda en línea de que dispone el paquete.

### 3.1.9 BARRA DE HERRAMIENTAS

2.99691763054292



Vars

Cases



CUADRO DE MUESTRA

MENU PRINCIPAL

TODAS LAS ESPECIFICACIONES

MENU MODIFICAR VARIABLES

MENU MODIFICAR CASOS

VER VALORES ALFANUMERICOS

VER NOMBRES DE CASOS

ACERCAMIENTO

ALEJAMIENTO

INCREMENTAR ANCHO DE COLUMNA

DECREMENTAR ANCHO DE COLUMNA

ANCHO GLOBAL

AUMENTAR DECIMALES

DISMINUIR DECIMALES

RECALCULAR VARIABLES

MANEJADOR DE VALORES ALFANUMERICOS

OPCIONES PARA GRAFICAS

ABRIR ARCHIVO



GUARDAR ARCHIVO



IMPRIMIR



CORTAR



COPIAR



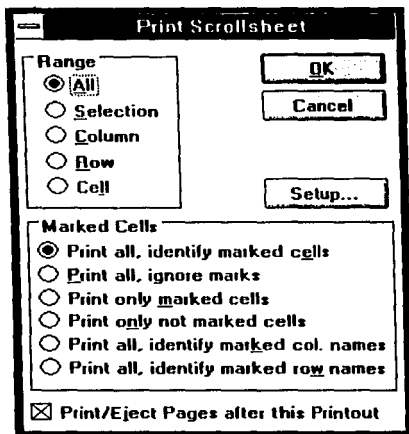
PEGAR

## 3.2 SCROLLSHEET

### 3.2.1 FILE

**SAVE AS DATA:** Despliega la ventana *Save Scrollsheet As* que permite guardar el contenido del scrollsheet como un archivo de datos (con extensión \*.sta) lo cual permite emplear los datos de salida generados en el scrollsheet como los datos de entrada para un nuevo análisis o exportarlos a otro formato de archivo. Cuando STATISTICA crea un archivo de datos del scrollsheet, los nombres de renglón se convertirán en nombres de casos, los encabezados de las columnas (primera y segunda líneas) serán ahora los nombres de las variables (primera línea) y nombres largos o etiquetas de las variables (segunda línea) y por último, la línea de título del scrollsheet corresponderá al encabezado del archivo de datos.

**PRINT:** Imprime el scrollsheet activo o lo direcciona a otra salida, dependiendo de la elección hecha en la ventana *Page/Output Setup*. Al elegir esta opción, se abrirá la ventana *Print Scrollsheet*:





**Range:** Indicar la parte del scrollsheet que será impresa.

**All:** Imprime todo el scrollsheet activo.

**Selection:** Imprime el bloque seleccionado (resaltado) de celdas del scrollsheet activo.

**Column/Row/Cell:** Imprime sólo la columna/renglón/celda en la que está posicionado el cursor.


**Marked cells:** Las siguientes opciones sólo son aplicables a las celdas marcadas (mediante la opción *Mark selection* del menú *View* o por un procedimiento de análisis, por ejemplo, en un análisis de correlación, las correlaciones significativas aparecen marcadas automáticamente).

**Print all, identify marked cells:** Imprime la parte del scrollsheet que se haya especificado en la opción *Range* colocando además un asterisco a la derecha del valor de cada celda marcada.

**Print all, ignore marks:** Imprime la parte del scrollsheet que se haya especificado en la opción *Range* sin indicar cuales celdas fueron marcadas.

**Print only marked cells:** Unicamente imprime las celdas marcadas del scrollsheet, dejando el resto de ellas en blanco (los encabezados de renglón y columna si se imprimirán).

**Print all, identify marked column (row) names:** Cuando existen celdas marcadas en el rango seleccionado (ver *Range*), elegir esta opción para marcar en la impresión unicamente el nombre de la columna (renglón) asociado con las celdas marcadas, en lugar de marcar cada celda individual.

**Print/eject pages after this printout:** Si esta opción está activa, cuando se presione el botón **OK** en esta ventana (*Print Scrollsheet*), o el botón de impresión de la barra de herramientas:  la información automáticamente será dirigida a la impresora o al archivo de salida (dependiendo de las elecciones hechas en *Page/Output Setup*). Cada impresión empezará en una nueva página. Si la opción se encuentra desactivada, la salida será almacenada en el buffer interno y enviada a la impresora dependiendo del estado actual de la opción *Automatically Eject Each Filled Page* que también está disponible en *Page/Output Setup*.

**Setup:** Al presionar este botón se abrirá la ventana *Text Output Printer Setup* en la cual, se puede seleccionar la calidad de la impresión así como el número de copias.

Botón asociado: 



### 3.2.2 EDIT

**CELL:** Permite editar la celda actualmente resaltada en el *scrollsheet*. Después de seleccionarla, la opción permanece activa hasta que se concluye la edición.

**COLUMN SPECS:** Permite editar las especificaciones de la columna actualmente resaltada mediante la ventana *Scrollsheet Column Specs*. Estas especificaciones son particulares para cada columna, excepto las que se refieren al ancho global y margen las cuales se aplican a todo el *scrollsheet*.

*Ventana Scrollsheet Column Specs:*

**First line:** Escribir la primera línea del encabezado de la columna, la longitud de cada línea está limitada al ancho global de la columna (ver abajo) y puede contener cualquier carácter imprimible. Si el texto es muy largo y la segunda línea está vacía, el encabezado se desplegará también en esta última, pero se podrá seguir editando sólo en la primera.

**Second line:** Escribir la segunda línea del encabezado de la columna, otra vez, la longitud de cada línea está limitada al ancho global de la columna y puede contener cualquier carácter imprimible.

**Format:** Da formato a los valores desplegados en las columnas:

**Automatic:** Alinea los valores de acuerdo al número más largo, colocando el punto decimal de cada celda en la misma posición a lo largo de la columna.

**Optimal:** Aquí los valores no son alineados de acuerdo a la posición del punto decimal, sino que cada celda será ajustada individualmente para permitir un despliegue "óptimo" de valores. Los valores llenarán o no el ancho de la columna de acuerdo al tamaño que se especifique en el ancho global y de margen (ver abajo).

**Float:** Esta opción permite especificar el número de decimales para los valores de la columna específica. El número mínimo de decimales es 1 y el máximo es 15.

**Integer:** Esta opción redondea cualquier valor decimal a entero. Desde 1 hasta el valor que permita desplegar el ancho de la columna actual.

**Global width:** El valor escrito aquí (mínimo=1, máximo=20) determina el ancho de cada columna del *scrollsheet*. Es una operación global, es decir, el valor especificado se aplica a todas las columnas. Esta opción también está disponible en la ventana *Customize Width*, que se describe más adelante en el menú *View*.

**Margin Width:** Los valores de cada columna pueden tener un margen predeterminado (mínimo=1, máximo=10) en ambos lados para distinguir qué valores pertenecen a cada columna. También es una operación global y está disponible en la ventana *Customize Width* de la opción *Width* del menú *View*.



**ROW NAME:** Permite editar el nombre del renglón actualmente resaltado (opción disponible al hacer doble clic en el nombre del renglón en el scrollsheet). Al seleccionarla aparece la ventana *Scrollsheet Row Name* con las siguientes opciones:

**Name:** Aquí se edita el nombre del renglón, éste puede ser de hasta 40 caracteres (incluye cualquier caracter imprimible).

**Width:** Aquí se especifica el número (1 a 40) que indica el ancho global para el nombre de cada uno de los renglones.

**TITLES:** Es posible editar el título del scrollsheet através de la ventana *Scrollsheet Titles* disponible en esta opción o haciendo doble clic en el área de título del scrollsheet.

*Ventana Scrollsheet Titles:*

**Window title:** Por omisión el título de la ventana contendrá el nombre del análisis realizado así como el del archivo de datos. Aquí se puede personalizar.

**First line / Second Line:** Además del título de la ventana mencionado arriba, se permiten 2 líneas más de hasta 63 caracteres imprimibles cada una.

**Alignment:** Aquí se indica cómo se quiere la alineación de los títulos: centrados (**center**), alineados a la izquierda (**left**) o derecha (**right**).

**INFO BOX:** Es el área a la izquierda de los títulos (en la esquina superior izquierda del scrollsheet) que permite personalizar más el scrollsheet con la información adicional que aquí se proporciona.

*Ventana Scrollsheet Info Box:*

Permite editar 4 líneas de información general acerca del scrollsheet, cuya longitud depende del ancho definido para el nombre del renglón en la opción *Row Name* (ver atrás).

En la línea 1, STATISTICA guarda por omisión el nombre del módulo en el cual se está trabajando y si es necesario, ocupa también la línea siguiente, la línea 3 por omisión aparece en blanco y la 4 está reservada para el encabezado de los nombres de renglón.

Estas líneas pueden ser editadas con la información que se desee.

**APPEND COLUMNS/ROWS:** Estas opciones permiten agregar columnas/renglones en blanco después de la última columna o renglón en el scrollsheet. Al seleccionarlas se despliega una ventana preguntando el número de renglones/columnas que se desean agregar al scrollsheet.



### 3.2.3 VIEW

**WIDTH:** Abre la ventana *Customize Width* en la cual se puede indicar el ancho que se desea para todas las columnas, renglones y márgenes del *scrollsheet*.

Ventana *Customize Width*:

**Column Width:** Escribir un número comprendido entre 1 y 20 para especificar el ancho de las columnas.

**Row name width:** Dar un número del 1 al 40 para establecer el ancho de la columna que contiene los nombres de los renglones.

**Margin width:** Los valores de cada columna pueden tener un margen predeterminado (min=1, máx=10) en ambos lados para distinguir qué valores pertenecen a cada columna.

Estas opciones también están disponibles en las ventanas *Scrollsheet Column Specs* y *Scrollsheet Row Names* descritas en el menú *Edit*.

Botón asociado: 

**ZOOM IN/OUT:** Esta opción permite aumentar/disminuir el tamaño de la fuente o tipo de letra que se está usando en el *scrollsheet* (en incrementos definidos por el tipo de letra usado). Puede suceder que al seleccionar por vez primera la opción aparezca una advertencia indicando que el tipo de letra usado no permite aumentar/disminuir más su tamaño e invitando a cambiar por un tipo de letra que si lo permita.

Botones asociados:  Zoom in

 Zoom out

**MARK SELECTION:** Permite marcar o desmarcar valores individuales o en bloque con el fin de poderlos distinguir del resto del *scrollsheet*. Primero se debe seleccionar (resaltar) la celda o bloque que se quiere marcar y después se elige la opción. Los valores marcados aparecerán en un color diferente al del resto y cuando sean impresos serán señalados por omisión con un asterisco (\*); esta marca se puede cambiar en la ventana *Print Scrollsheet* de las opciones de impresión.

Cuando se lleva a cabo algún análisis, celdas o bloques específicos ya aparecen marcados en muchos *scrollsheet* (en el momento en que son generados) para resaltar resultados estadísticamente significantes. Por ejemplo: coeficientes de correlación; tales *scrollsheets* generalmente ofrecen una opción que permite cambiar el criterio de selección de valores (v.g. el nivel de significancia).

Botón asociado: 



### 3.2.4 BARRA DE HERRAMIENTAS

Además de los botones indicados en el spreadsheet contiene el siguiente:

#### **Options**

Este botón permite cambiar o seleccionar una opción específica (dependiendo del módulo que se está usando) que afectará los resultados en el spreadsheet. Por ejemplo, en algunas matrices de correlación este botón abrirá una ventana permitiendo establecer el nivel de significancia para los coeficientes de correlación el cual será marcado en el spreadsheet (el nivel por omisión es  $\alpha=.05$ ).

## 3.3 VENTANA TEXT/OUTPUT

### 3.3.1 FILE

**NEW:** Borra todo el texto desplegado en la ventana para crear un nuevo archivo de texto, eliminando la información que no ha sido guardada previamente.

**OPEN TEXT:** Permite abrir un archivo de texto y lo despliega en la ventana, si ésta ya estaba abierta con otro archivo, lo cierra preguntando antes si se desean guardar los cambios.

Botón asociado:

#### **OPEN OTHER:**

**NEW DATA:** Crea un nuevo archivo de datos con las características por omisión preestablecidas en la ventana *Create New File*. Al seleccionar la opción despliega la ventana *New Data*, donde se especifica el nombre que va a tener el archivo y su localización.

**DATA FILE:** Abre otro archivo de datos (\*.sta) para reemplazar al que está actualmente en uso.

**SAVE TEXT AS:** Permite guardar el texto editado en la ventana como un archivo de texto (con extensión \*.txt).

Botón asociado:

**PRINT:** A diferencia de las opciones de impresión disponibles cuando el spreadsheet o el spreadsheet están activos, esta opción enviará automáticamente el texto de la ventana a la impresora (es decir, no habrá ventanas intermediarias con distintas opciones de impresión).

Botón asociado:





### 3.3.2 EDIT

**UNDO:** Invalida la última instrucción ejecutada. Esta opción se puede aplicar en las siguientes instrucciones:

**CUT, PASTE, CLEAR:** Corta, pega o borra el último texto escrito en la ventana.

**SPLIT:** Divide la ventana en dos partes a partir de la posición actual del cursor. La parte superior permanecerá fija (sin poder moverse ni editarse) , mientras que la inferior se desplazará normalmente de acuerdo al movimiento de la barra de desplazamiento del lado derecho de la ventana y podrá ser editada. Esto permite hacer comparaciones entre el texto congelado en la parte superior y el resto del texto contenido en la otra parte. Se puede regresar al estado normal de la ventana haciendo nuevamente clic en la opción.

### 3.3.3 OPTIONS

**APPEND RESULTS:** Cuando la opción se encuentra activa (una "✓" precede el nombre), los resultados de cualquier análisis que se realice se agregarán automáticamente al final del texto que se encuentre desplegado en la ventana en ese momento. Esta opción es muy útil cuando se ejecuta continuamente un conjunto de análisis. Sin embargo, a veces se requiere usar esta ventana sólo como un block de notas, si éste es el caso, se puede deshabilitar temporalmente la opción (haciendo clic nuevamente en ella para quitar el símbolo o "✓" que precede al nombre).

### 3.3.4 BARRA DE HERRAMIENTAS

**Start-up Panel**

PANEL DE INICIO

**Resume Analysis**

RESUMIR ANALISIS

Las opciones de menú asociadas a estos botones se describen en la sección Analysis de este capítulo.



## CAPITULO 4

### GRAFICAS

#### 4.1 CLASIFICACION GENERAL

Además de las gráficas especializadas que están disponibles en las ventanas de todos los procedimientos estadísticos, existen dos categorías generales de gráficas que se encuentran disponibles en el menú y la barra de herramientas del scrollsheet y spreadsheet.

- Custom Graphs. (Gráficas Personalizadas)
- Stats (y Quick Stats) Graphs. (Gráficas Estadísticas)

Estas gráficas difieren en términos de la fuente de datos que visualizan.

##### 4.1.1 GRAFICAS PERSONALIZADAS (CUSTOM GRAPHS)

Están disponibles en el menú *Graphs*; permiten visualizar cualquier combinación de valores del scrollsheet o spreadsheet (de renglones, columnas, o ambos, y/o de subconjuntos de los mismos). Cuando se selecciona uno de los cuatro tipos de estas gráficas, se abrirá la ventana respectiva para especificar los datos que serán graficados. La distribución de esta ventana depende del tipo global de gráfica personalizada que se haya seleccionado (ejemplo: *Custom 2D Graphs*). La selección inicial de los datos a ser graficados (sugeridos en la ventana) depende del bloque resaltado o de la posición del cursor en el actual scrollsheet o spreadsheet. Cada una de las ventanas de definición de gráficas personalizadas permite seleccionar tipos específicos de éstas (dentro del tipo global). Sin embargo, el tipo de gráfica también puede ser ajustado después de que ésta ha sido creada (por medio de las ventanas *General Layout* o *Plot Layout* las cuales se explican en la sección Elementos Principales de este capítulo). Es posible trazar más gráficas dentro de la misma ventana uniendo una gráfica previamente guardada a la actual mediante la opción *Merge graph* del menú *File* o bien, agregando más series de datos al editor de datos para gráficas: *Graph Data Editor* (disponible en la opción *Edit data* del menú *Layouts*).

Existen 4 tipos:

**1. CUSTOM 2D GRAPHS.**

Esta opción permite definir una gráfica bidimensional hasta para cuatro grupos diferentes de datos (i.e., se pueden desplegar hasta cuatro gráficas en la misma ventana). Cuando se selecciona esta opción se abrirá la ventana *2D Graphs* donde se especifica el tipo de cada una de las cuatro gráficas así como la fuente de valores a graficarse.

Nota: si se definen dos o más gráficas, éstas serán desplegadas una sobre otra en la misma ventana.

Botón asociado:

Descripción de la ventana *Custom 2D Graphs*:

**Plot 1 (2,3,4):** En este recuadro se debe seleccionar el tipo de gráfica de la lista disponible para cada una de las cuatro gráficas que esta ventana puede definir.

Existen 14 tipos :



Scatterplot



Line Plot



Step Plot



Min-Max X



Min-Max Y



Bar X



Bar Left Y



Bar Right Y



Bar Top



Bar Dev



Pie



Mpatt Bar



Voronoi



Ignore

**Spreadsheet Custom 2D Graphs**

**Plot 1**  
 Line Plot  
 X: Row Names  
 Y: GENDER

**Plot 2**  
 Ignore  
 X: Row Names  
 Y: ADVERT

**Plot 3**  
 Ignore  
 X: Row Names  
 Y: MEASUR1

**Plot 4**  
 Ignore  
 X: Row Names  
 Y: MEASUR2

OK  
 Cancel

Values From:  
 1  
 Lo:  
 50

Double-click on the field to select from list.



**X y Y:** Haciendo doble clic en cualquiera de estos campos se abre la ventana *Get Plotted Values From* donde se deben seleccionar los valores fuente (valores de columnas (*Column*), renglones (*Row*) o sus nombres (*Col names, Row names*) para graficarse.

**Values:** Designar el rango de valores a graficar (de qué valor **-From-**, a qué valor **-To-**). Si hay un bloque de valores resaltado en el spreadsheet, STATISTICA por omisión incluirá el rango de dicho bloque. Por ejemplo, si se designa el rango From 10 To 15; y en la opción anterior se especificó que los valores para X se obtuvieron de la columna 3 y para Y de la columna 4, entonces, STATISTICA va a graficar del 10° al 15° valor de la columna 3 vs los valores 10° al 15° de la 4.

## 2. CUSTOM 3D SECUENTIAL GRAPHS.

Esta opción permite definir una gráfica tridimensional utilizando cualquier combinación de valores de renglones y/o columnas del actual scrollsheet o spreadsheet, o subconjuntos de los mismos, para visualizar una o mas secuencias de valores en forma tridimensional. En general, estas gráficas representan cada serie de datos (v.g. cada columna seleccionada o su subconjunto) como una gráfica. Los valores consecutivos de las series de datos son representados a lo largo del eje X, y los datos como las alturas de las gráficas (como medida del eje Z). Los tipos de perímetro (*contour*) y superficie (*surface*) integran todas las gráficas individuales en una sola estructura.

Botón asociado:

Descripción de la ventana *Custom 3D Sequential Graphs*:

**Graph type:** Elegir el tipo de gráfica haciendo clic en el nombre.

Existen 8 tipos:



Box Plot



Block Plot



Ribbons



Lines



Spikes



Contour/Discrete



Surface Plot



Contour Plot

Spreadsheet Custom 3D Sequence Graphs

**Graph Type:**

- Box Plot
- Block Plot
- Ribbons
- Lines
- Spikes
- Contour/Discrete
- Surface Plot
- Contour Plot

**Plots**

1: GENDER	Z:	
2: ADVERT	8:	
3: MEASUR1	9:	
4:	10:	
5:	11:	
6:	12:	

Double-click on the field to select from list.

OK

Cancel

Values

From: 15

To: 24

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA



**Plots:** Se pueden trazar hasta 12 gráficas dentro de la misma ventana. Los datos a graficarse se obtienen de las variables (columnas), para esto, se deben escribir sus nombres en los campos de edición correspondientes o hacer doble clic en ellos para abrir la ventana *Get Plotted Values From* donde aparece la lista de todas las variables del archivo para elegir la que se desee, esta ventana también permite graficar los valores por renglón (*row*) en lugar de columnas (*variables*).

**Values:** Se explica en la opción anterior.

### 3. CUSTOM 3D SCATTERPLOTS AND SURFACES.








Esta opción permite definir una gráfica tridimensional de dispersión (scatterplot) o superficie (surface) utilizando cualquier combinación de valores de renglones y/o columnas del actual scrollsheet o spreadsheet, o subconjuntos de los mismos, para visualizar relaciones interactivas complejas entre variables. A diferencia de *Custom 3D Sequential Graphs* las cuales simplemente ofrecen una manera alternativa de desplegar series individuales de datos, en estas gráficas, cada punto se define por 3 valores ( $x, y, z$ ) y es visualizado en el espacio tridimensional de acuerdo a sus coordenadas  $x$ - $y$ - $z$ .

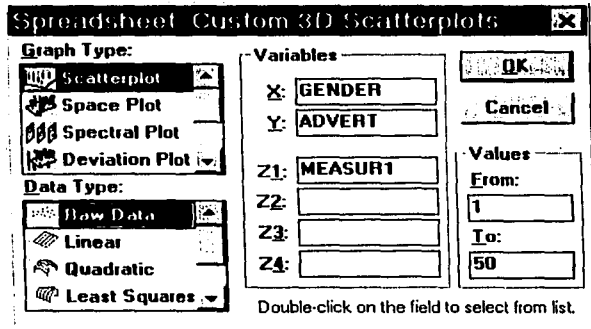
Botón asociado:

Descripción de la ventana *Custom 3D Scatterplot and Surfaces*:

**Graph type:** Elegir el tipo de gráfica tridimensional de la lista que se muestra.

Existen 7 tipos:

-  Scatterplot
-  Space Plot
-  Spectral Plot
-  Deviation Plot
-  Trace Plot
-  Surface Plot
-  Contour Plot



**Data type:** Realizando una transformación de datos se puede desplegar una superficie (o contorno) que presente una imagen más uniforme de éstos; en lugar de graficar los datos originales (por omisión).



Las transformaciones de datos disponibles son las siguientes:



Linear



Quadratic



Least Squares



Negative Exponential



Spline Smooth



Custom Function

**Variables:** Seleccionar las variables que se van a graficar ya sea escribiendo su nombre o haciendo doble clic en el campo de edición de la variable para abrir la ventana *Get Plotted Values From* donde aparece la lista de todas las variables del archivo para elegir la que se desee, esta ventana también permite graficar los valores por renglón (*row*) en lugar de columnas (*variables*). Se pueden crear hasta 4 gráficas XYZ con los mismos valores en X y Y; cambiando únicamente los valores en Z.

**Values:** Se describe en la ventana de *Custom 2D Graphs*.

#### 4. CUSTOM MATRIX AND ICON GRAPHS.

Esta opción permite definir una matriz de gráficas o una gráfica de iconos.

**Matriz de gráficas:** Es el resultado de cruzar los valores de dos renglones, dos columnas, o un renglón y una columna del spreadsheet o scrollsheets. Las matrices de gráficas representan relaciones entre variables. Se pueden seleccionar hasta 12 renglones o columnas para cruzar. STATISTICA despliega histogramas en la diagonal de estas matrices que representan las distribuciones de frecuencia de los valores graficados. Por omisión, las escalas verticales de estos histogramas son independientes; específicamente, son ajustadas a los valores de cada renglón o columna para proporcionar las mejores condiciones para examinar la forma de cada distribución y aprovechar el área total de la gráfica o de cada gráfica pequeña (se puede forzar a todos los histogramas a tener una escala común seleccionando la opción *Histograms-Common* en la ventana *General Layout* (para Matriz de gráficas) descrita en la sección Elementos Principales de este capítulo).

**Gráficas de iconos:** Representan unidades individuales de observación (casos) como objetos gráficos individuales (ejemplo: caras de Chernoff, estrellas, rayos de sol, polígonos, pays, columnas, líneas, y contornos) donde los valores de hasta 12 variables son asignados a características o dimensiones específicas de los objetos (caso uno=objeto uno). Examinar dichos iconos puede ayudar a descubrir grupos específicos de relaciones simples e interacciones entre variables.



Botón asociado: 

Descripción de la ventana *Custom Matrix / Icon Graphs*:

**Graph type:** Elegir el tipo de gráfica ya sea matriz o icono haciendo clic en el nombre.

Existen 11 tipos:



Scatterplot (matrix)



Line (matrix)



Column (matrix)



Chernoff Faces



Stars



Sun Rays



Polygons



Pies



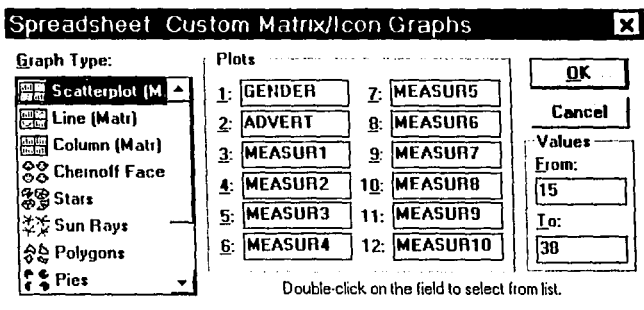
Columns



Lines



Profiles



**Plots:** Seleccionar las variables que se van a graficar ya sea escribiendo su nombre o haciendo doble clic en el campo de edición de la variable para abrir la ventana *Get Plotted Values From* donde aparece la lista de todas las variables del archivo para elegir la que se desee, esta ventana también permite graficar los valores por renglón (*row*) en lugar de columnas (*variables*).

**Values:** Se describe en la ventana *Custom 2D Graphs*.

#### 4.1.2 GRAFICAS ESTADISTICAS SG (STATS GRAPHS)

Las gráficas estadísticas están disponibles en el menú *Graphs* y ofrecen cientos de tipos de representaciones gráficas y resúmenes analíticos de datos. A diferencia de las gráficas personalizadas, las gráficas estadísticas están pre-definidas y no están limitadas a los valores del scrollsheet activo ni de los bloques resaltados o la posición actual del cursor, en lugar de ello, procesan los datos directamente del archivo de datos actual. Ellas representan ya sea métodos estándares para resumir los datos gráficamente (ejemplo:




gráficas de dispersión, histogramas, o gráficas de tendencia central tales como medianas) o técnicas estándar de análisis gráfico (ejemplo: gráficas en papel normal, gráficas de intervalo de confianza de regresión lineal, etc.).

En el momento que se genera una gráfica estadística, STATISTICA tomará en cuenta la selección de casos y las condiciones de peso de las variables seleccionadas para graficarse.

Nota: Debido a la extensión de las gráficas *Stat Graphs*, la explicación se limita sólo a las gráficas *Custom* y *Quick Stat Graphs*, ya que son las más utilizadas dentro de las ventanas de análisis.

### 4.1.3 GRAFICAS ESTADISTICAS RAPIDAS QSG (QUICK STATS GRAPHS)

Incluyen los tipos de gráficas estadísticas comúnmente utilizadas. QSG no ofrece tantas opciones como SG, sin embargo, son más rápidas de seleccionar porque a diferencia de SG:


- QSG pueden ser llamadas con un botón de la barra de herramientas:  o presionando el botón derecho del ratón.
- QSG no requiere que el usuario seleccione la variable a graficar (la variable seleccionada es determinada por la posición actual del cursor dentro del scrollsheet o spreadsheet).
- QSG no requiere que el usuario seleccione opciones de alguna ventana intermedia (automáticamente se producen los formatos por omisión de la gráfica respectiva).

Al seleccionar la opción de QSG, se abrirá un menú donde se puede elegir una de las gráficas estadísticas aplicable a la variable actual, es decir, la variable resaltada en el spreadsheet o scrollsheet.

Siempre que se genera una gráfica estadística rápida, STATISTICA tomará en cuenta la selección de casos y las condiciones de peso de las variables seleccionadas para graficarse.

Existen 10 tipos:

**1. QSG - VALUES OF.** Esta opción permite revisar los valores de la variable seleccionada y sus especificaciones (nombre, código de falta de dato, formato y nombre largo) así como algunas estadísticas descriptivas (n, media y desviación estándar) para esa variable.

Las estadísticas descriptivas pueden copiarse al portapapeles de Windows (para poder pegarlas posteriormente en otra aplicación) haciendo clic en el botón 





**2. QSG - HISTOGRAM OF.** Esta opción despliega un histograma (representación gráfica de una tabla de frecuencia) de los valores de la variable actualmente resaltada ya sea con (seleccionar *Normal Fit*) o sin (seleccionar *Regular*) una curva normal ajustada a los datos. Si la variable seleccionada clasifica los datos (usualmente hasta 10 categorías), entonces STATISTICA utilizará estas categorías como los intervalos del histograma. Si las variables tienen más de 10 categorías o los valores no son enteros (continuos), entonces STATISTICA dividirá los valores en 10, 11 ó 12 intervalos basados en el rango de los valores.

**3. QSG - BOX-WHISKER OF.** Este tipo de gráficas representan la tendencia central y la dispersión de la variable actualmente resaltada. Como medidas de tendencia central se puede elegir entre graficar las medias y medianas con sus respectivas medidas de variabilidad: desviaciones estándar o errores estándar (asociados con las medias), y rangos (asociados con las medianas). Seleccionar uno de los cuatro tipos:

**Median/Quart/Range:** Describe la tendencia central de la variable en términos de la mediana (representada por el cuadrado). La dispersión de los valores de la variable está indicada por los cuartiles (percentiles 25° y 75°) representados con el cuadro más grande y por los valores mínimo y máximo de la variable (los "whiskers").

**Mean/SE/SD:** Aquí el cuadrado representa la media (tendencia central), mientras que la dispersión (variabilidad) está representada por  $\pm 1$  veces el error estándar (cuadro grande) y  $\pm 1$  veces la desviación estándar alrededor de la media ("whiskers").

**Mean/SD/1.96\*SD:** Esta gráfica mostrará la media (cuadrado) de la variable, rodeada por el cuadro grande que representa  $\pm 1$  veces la desviación estándar. Si la distribución es normal, los "whiskers" representan un intervalo del 95% de confianza definido como la media  $\pm 1.96$  veces la desviación estándar de la variable.

**Mean/SE/1.96\*SE:** Esta gráfica mostrará la media (cuadrado) de la variable, rodeada por el cuadro grande que representa  $\pm 1$  veces el error estándar. Si la distribución es normal, los "whiskers" representan un intervalo del 95% de confianza definido como la media  $\pm 1.96$  veces el error estándar de la variable.

**4. QSG PROBABILITY PLOT OF...** Este tipo de gráficas son útiles para verificar visualmente la normalidad en la distribución de la variable actualmente resaltada. En estadística, la hipótesis de normalidad es muy importante ya que muchos procedimientos de análisis dan por hecho que la distribución de la variable analizada es normal.

STATISTICA proporciona tres tipos diferentes de gráficas de probabilidad normal:

**Normal Probability Plot:** Es la gráfica estándar de probabilidad normal. Para construirla, primero se ordenan (asignándoles un rango) los valores de las desviaciones de la media (residuales); de estos  $z$  rangos asignados se calculan  $z$  valores que se supone van a provenir de una distribución normal estándar. Estos  $z$  valores son colocados en el eje Y. Si los residuales observados (colocados en el eje X) se distribuyen normal, entonces, todos los puntos de la gráfica deben seguir una tendencia lineal, de lo contrario, no se puede



confirmar una distribución normal. Estas gráficas también son útiles para detectar la presencia de outliers.

**Half - Normal Probability Plot:** Esta gráfica es construida del mismo modo que la anterior, excepto que sólo se considera la mitad positiva de la curva normal; por lo tanto, sólo se graficarán valores  $z$  positivos. Esta gráfica se usa cuando se está interesado principalmente en la distribución de los residuales absolutos sin importar el signo.

**Detrended Normal Probability Plot:** Esta gráfica es similar a la primera, salvo que antes de que se genere, la tendencia lineal es eliminada. Esto hace que los puntos de la gráfica se dispersen permitiendo al usuario detectar fácilmente patrones de desviaciones.

**5. QSG SCATTERPLOT BY...** Las gráficas de dispersión visualizan la relación entre 2 variables  $X$  y  $Y$  (ejemplo: peso y altura). Al seleccionar esta opción, se puede elegir entre una gráfica de dispersión estándar (*Regular*) o una gráfica de dispersión con una línea de regresión lineal o polinomial (por omisión se ajusta un polinomio de quinto grado) ajustada a los datos con un intervalo del 95% de confianza. Una vez que se realiza la selección, se abre la ventana *Select Variable* donde se debe indicar la variable que va a ser graficada contra la variable actualmente resaltada. Los 3 tipos de gráficas son los siguientes:

**Regression, 95% conf:**  
**Polynomial, 95% conf.**  
**Regular**

Nota: El valor de la probabilidad del intervalo de confianza así como otro tipo de funciones a ser ajustadas pueden ser elegidos dentro de la ventana *Plot Layout* accesible haciendo doble clic en el fondo de la gráfica.

**6. QSG 2D HISTOGRAM BY...** Esta opción ofrece una extensión a la opción *Histogram of*, ya que permite examinar la distribución de los valores de la variable actualmente resaltada clasificados de acuerdo a los valores de otra variable independiente (variable clasificadora). Por ejemplo, graficar la distribución de los resultados de un examen clasificados por el sexo de los alumnos.

En algunos scrollsheets, la selección de la variable clasificadora está predeterminada por una variable clasificadora global, es decir, de antemano se toma la variable clasificadora designada en la ventana del análisis que se esté realizando. Cuando éste es el caso, STATISTICA incorporará el nombre de la variable clasificadora predeterminada en la opción del menú de QSG (ejemplo: *2D Histogram by gender* (por género)). En la mayoría de los otros casos, STATISTICA permite seleccionar la variable clasificadora (por medio de la ventana *Select Variable*) que será utilizada para clasificar los valores de la variable resaltada. Una vez que se selecciona ésta; se abrirá la ventana *Codes* donde se deben especificar los valores (códigos) a utilizar para la clasificación. (Si la lista de códigos se deja vacía, todos los valores de la variable clasificadora serán utilizados automáticamente). Después de que se introducen los códigos, STATISTICA despliega un histograma para cada categoría de la variable clasificadora.



**7. QSG 3D HISTOGRAM BY...** Esta opción permite crear diferentes tipos de histogramas bivariados tridimensionales. Estos varían dependiendo de los tipos de datos (continuos o discretos) que contiene el archivo. En esta opción se abrirá la ventana *Select One Variable* donde se debe seleccionar la variable que se graficará junto con la variable actualmente resaltada. Ambas variables serán divididas en categorías (basadas en el rango de los datos) para definir los intervalos del histograma. Cuando una variable seleccionada es categórica por naturaleza (generalmente hasta 10 categorías), entonces STATISTICA utilizará estas categorías como los intervalos del histograma para esa variable. Si una variable tiene más de 10 categorías o sus valores son continuos (no enteros), entonces STATISTICA dividirá los valores en 10, 11 ó 12 intervalos de acuerdo al rango de los mismos.

**Continuous by continuous:** Seleccionar esta opción si las dos variables que se van a graficar son continuas y/o se desea que STATISTICA cree automáticamente los intervalos del histograma de acuerdo al procedimiento descrito anteriormente.

**Discrete by discrete:** Esta opción se aplica cuando las dos variables a graficar son discretas. En la ventana *Codes* se deben especificar qué valores de las variables van a definir los intervalos del histograma.

**Continuous by discrete:** Esta opción permite graficar una variable continua (la variable actualmente resaltada) contra una discreta. La primera variable automáticamente es dividida en categorías para crear los intervalos del histograma de acuerdo a las convenciones señaladas anteriormente mientras que para la segunda se pueden especificar los valores en la ventana *Codes*.

**Discrete by continuous:** Esta opción es la inversa de la anterior.

**8. QSG BOX- WHISKER BY...** Las gráficas que resultan de esta opción representan la tendencia central y la dispersión para cada categoría (definida por una variable clasificadora) de la variable actualmente resaltada. Como medidas de tendencia central, se puede elegir entre graficar las medias o medianas de los valores en cada categoría con sus respectivas medidas de variabilidad: desviaciones estándar o errores estándar (asociados con las medias), y rangos (asociados con las medianas).

Median / Quart / Range  
Mean / SE / SD  
Mean / SD /  $1.96 \cdot SD$   
Mean / SE /  $1.96 \cdot SE$

Estos tres tipos de gráficas se explicaron en la opción *QSG Box-whisker of...* la única diferencia es que en este caso se tiene una gráfica para cada categoría.

En algunos scrollsheets, la selección de la variable clasificadora está predeterminada por alguna variable clasificadora global (i. e., la variable clasificadora designada en alguna ventana de análisis). Si éste es el caso, STATISTICA incorpora el nombre de dicha variable en la opción del menú (v.g. *Box-whisker by gender*).



\*Nota: En muchos casos, STATISTICA pide al usuario que seleccione la variable clasificadora a través de una ventana de selección de variables y posteriormente también pide que se elija (en la ventana *Codes*) qué valores de esta variable van a ser usados como códigos para la clasificación. Una vez especificado lo anterior, se producirán (dentro de una sola ventana) varias gráficas correspondientes a las distintas categorías de la variable clasificadora.

**9. QSG PROBABILITY PLOT BY...** Cuando se selecciona esta opción, se puede elegir uno de los tres tipos diferentes de gráficas de probabilidad normal para examinar la normalidad de los valores de una variable clasificados en categorías (por medio de una variable clasificadora). STATISTICA produce una gráfica para cada categoría.

**Normal Probability Plot**  
**Half - Normal Probability Plot**  
**Detrended Normal Probability Plot**

Estos tres tipos de gráficas se explicaron en la opción *QSG Probability plot of...*

En algunos scrollsheets, la selección de la variable clasificadora está predeterminada por alguna variable clasificadora global (i.e., la variable clasificadora designada en alguna ventana de análisis). Si éste es el caso, STATISTICA incorpora el nombre de dicha variable en la opción (v.g. *Probability plot by gender*). (Revisar \*Nota, arriba).

**10. MATRIX SCATTERPLOT...** Esta opción produce una representación (gráfica de dispersión) de las correlaciones entre varias variables en formato de matriz. Pueden seleccionarse una o dos listas de variables en la ventana *Select Variables From Matrix*, y una gráfica de dispersión será generada para cada par de variables. Si se seleccionan todas las variables en la primera lista, entonces se producirá una matriz cuadrada de scatterplots. Se puede generar una matriz rectangular de scatterplots seleccionando grupos distintos de variables en ambas listas.

Cuando se selecciona esta opción es posible escoger dos métodos de supresión de falta de dato (MD): *Casewise MD Deletion*: Los casos que tienen código de falta de dato en al menos una de las variables seleccionadas, serán excluidos de cada scatterplot. Todos los scatterplots se basarán en el mismo grupo de observaciones. Si se selecciona *Pairwise MD Deletion*, cada scatterplot en la matriz se basará en los casos que tienen datos válidos en ese par de variables, sin importar si existe falta de dato en otra variable.

#### 4.1.4 OTRAS GRAFICAS ESPECIALIZADAS

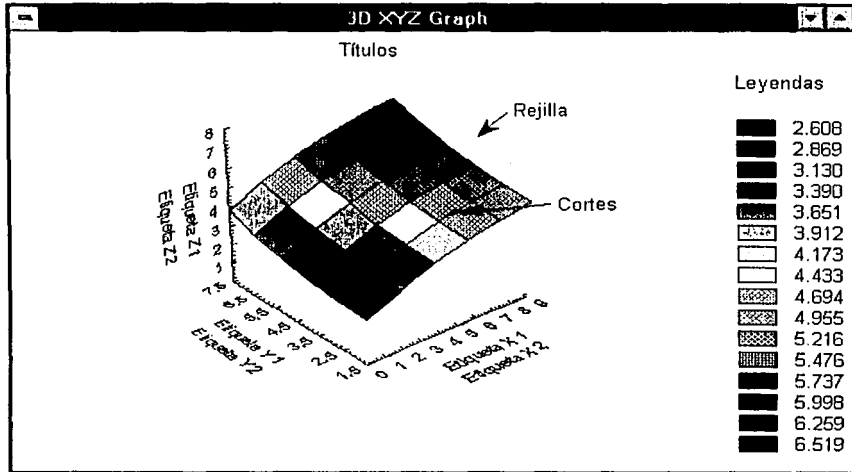
Además de la selección estándar de QSG, algunos scrollsheets también ofrecen otras gráficas estadísticas más especializadas (ejemplo: gráficas de residuales). También se encuentran las gráficas especiales relacionadas a un tipo de análisis estadístico (ejemplo: resultados de estimación no lineal) las cuales son accesibles directamente de las ventanas de resultados (es decir, ventanas que contienen opciones de salida del análisis).



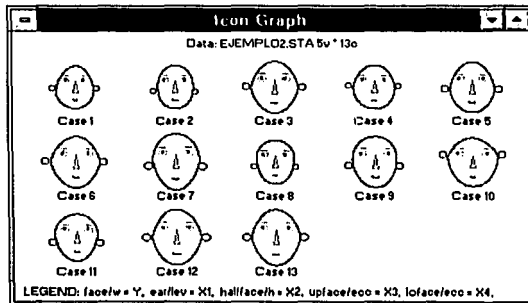
## 4.2 ELEMENTOS PRINCIPALES

Los elementos principales de la ventana de despliegue de una gráfica se muestran en las siguientes figuras:

Para gráficas 2D y 3D:



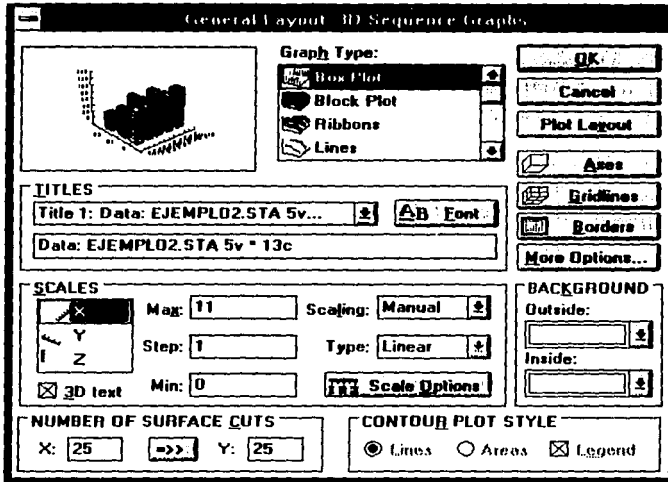
Para gráficas de iconos:



Estos elementos junto con otras características se pueden controlar mediante la ventana *General Layout* que se encuentra disponible al hacer doble clic en el fondo de la gráfica o seleccionando la opción *General Layout* del menú *Layouts*. Las opciones de esta ventana



varían dependiendo del tipo de gráfica elegida. La imagen que se muestra a continuación corresponde a la ventana *General Layout* para las gráficas personalizadas: *3D Sequence Graphs*.



A continuación se describen todas las opciones que esta ventana despliega para los dos tipos globales de gráficas (personalizadas y estadísticas):

El icono de la gráfica en la esquina superior izquierda de la ventana muestra previamente cualquier cambio que se realice en esta ventana.

**Titles:** Se pueden editar o introducir hasta 11 líneas de texto en la gráfica activa. Cada línea puede ser de un máximo de 80 caracteres de largo que a su vez se posicionan centrados automáticamente en el área de la gráfica. Estas 11 líneas de texto incluyen hasta 5 títulos (situados arriba de la gráfica) y 6 etiquetas distribuidas en los ejes. El texto de los títulos puede ser personalizado utilizando caracteres de control (ver en la ayuda en línea (menú *Help*) *Control Characters*, donde se enlistan los códigos disponibles).

**Font:** Abre la ventana *Font* donde se puede elegir una fuente diferente para los títulos y las etiquetas (la elección afectará solo el título que está actualmente seleccionado).

**Scales:** Permite cambiar la escala de los ejes indicando el eje que se modificará y el tipo de escala requerida:

**3D text:** Cuando se selecciona esta opción, las etiquetas de los ejes aparecerán a lo largo del eje del primer plano en la perspectiva tridimensional. Si no se selecciona, las etiquetas de los ejes aparecerán como en las gráficas bidimensionales.

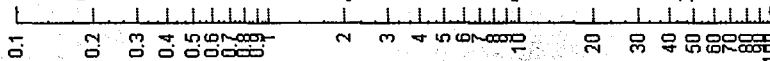


**Max:, Step:, Min:** Estos campos determinan los valores mínimo y máximo y la unidad de la escala (*step*) para el eje especificado.

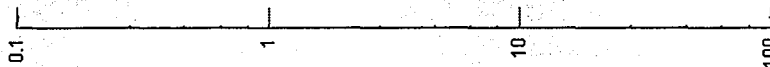
**Scaling:** Proporciona cuatro métodos para definir la escala. El método que se elija determina los valores *Min*, *Max* y la unidad de la escala de cada eje. La escala será fijada automáticamente por el programa (**Auto** o **Auto/0**) o manualmente (**Manual** o **Manual/0**). Cuando se selecciona *Auto* o *Manual*, los pasos serán determinados contando a partir del valor mínimo actual. La opción *Auto/0* o *Manual/0*, determinarán los pasos a partir del 0, aún si el 0 no se encuentra dentro del rango de valores.

**Type:** Elegir entre los tres tipos de escala (**linear**, **logarithmic**, o **log-decimal**). Si se selecciona la escala *logarithmic* o *log-decimal* el valor mínimo en la escala para el eje debe ser mayor que 0 y la unidad de escala, es decir, el ancho de los pasos en la gráfica, será determinada por STATISTICA.

Escala logarítmica:



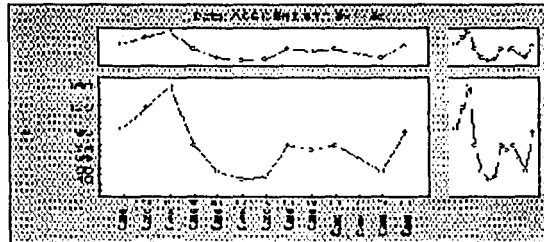
Escala log-decimal:



**Scale options:** Abre la ventana *Scale Options* que contiene muchas otras opciones de escalas para el eje seleccionado así como la posibilidad de cambiar los valores de escala numérica a valores de escala de texto, cambiar las marcas del eje o las líneas de escala, etc.

**Graph type:** La lista de gráficas disponibles en el tipo global actual, se despliega en esta opción. Si hay más de una gráfica desplegada en la ventana actual, entonces el tipo que se seleccione se aplicará a todas las gráficas. Si se quiere cambiar el tipo de cada gráfica individualmente, entonces se debe presionar el botón *Plot Layout*.

**Layered:** Esta opción reduce el área de la gráfica principal dejando espacio en la parte superior y lado derecho de la ventana para desplegar gráficas marginales (y una gráfica pequeña en la esquina) que representan vertical y horizontalmente imágenes comprimidas de la gráfica principal. Esto también llamado *Layered Compression* es una técnica de análisis exploratorio de datos que puede facilitar la identificación de tendencias ocultas y patrones en el conjunto de datos.

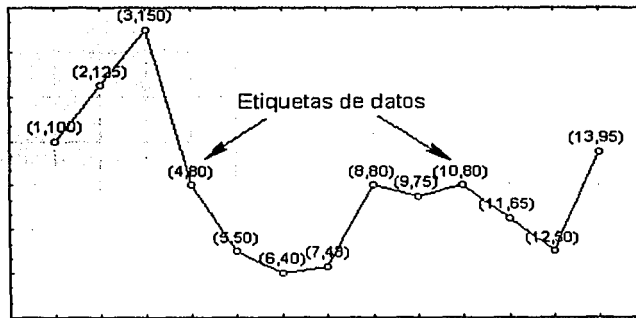


**Categorization:** Abre la ventana *Categorization Options*, sirve para editar el tipo de clasificación, etiquetas de categorías, y la escala de las gráficas de categorías.

**Plot layout:** Abre la ventana *Plot Layout* donde se puede editar la leyenda de la gráfica, escoger otro tipo (para cada gráfica individual), cambiar los patrones de línea o el estilo de barra, y otras opciones.

**Gridlines:** Abre la ventana *Gridlines* que permite elegir uno de los cuatro tipos de rejilla o red (*ninguna, X, Y, X+Y*).

**Data labels:** Abre la ventana *Data Point Labels* donde es posible determinar si se desea que los valores de los datos (coordenadas) o sus etiquetas sean desplegados arriba del punto respectivo en la gráfica. Para cambiar las etiquetas de una gráfica individual es necesario hacer doble clic en la etiqueta del dato correspondiente para abrir la ventana *Data Point Labels*.



**Borders:** Abre la ventana *Line Pattern - Borders* que permite escoger un borde dibujado al rededor de la gráfica de un color y grosor determinado.

**Background:** Hacer clic en el color deseado para cambiar el color del fondo de la gráfica.

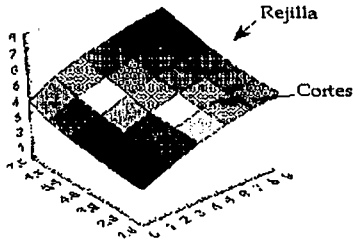
**Outside:** Hacer clic en el color para especificar el fondo exterior de la gráfica.

**Inside:** Seleccionar el color para el fondo interior de la gráfica.

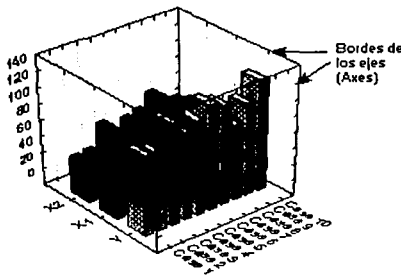




**Number of cuts:** Esta opción se aplica si el tipo de gráfica es de superficie o contorno. Se debe especificar el número de cortes para cada eje, lo que afectará la suavidad de la superficie de la gráfica.



**Axes:** Abre la ventana *Display Axes* que sirve para especificar exactamente cuales de los 12 bordes en la gráfica 3D (cuatro para cada eje) serán desplegados.



**More options:** Abre la ventana *3D Graphs / More Options* donde se puede ajustar la perspectiva y las proporciones de los ejes de la gráfica, así como el desplegado de las líneas y el contorno de las superficies graficadas.

**Contour plot:** Controla el desplegado de la gráfica de contorno. Puede desplegar estas gráficas con o sin leyenda y puede escoger desplegar el contorno como series de líneas o áreas.

**Fit (surface & contour):** Las siguientes opciones permiten ajustar una superficie al conjunto de datos. Se aplican solamente a gráficas de superficie (*Surface plot*) o de contorno (*Contour plot*) donde los datos son representados con las coordenadas cartesianas X-Y-Z. Cuando se selecciona uno de estos tipos de gráficas, es posible especificar el tipo de superficie a ajustar: Plano, superficie cuadrática, ajuste por mínimos cuadrados, exponencial negativo, etc.

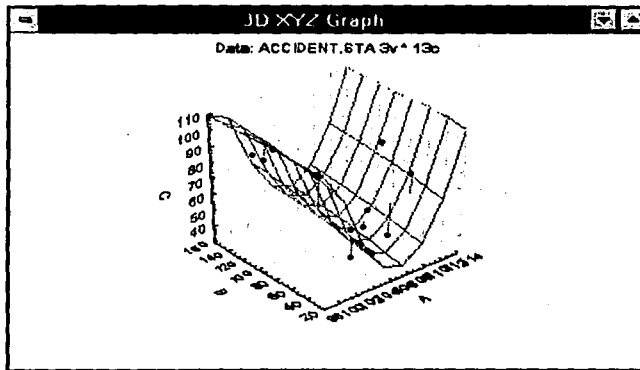


**Custom function:** En esta opción es posible definir una expresión matemática personalizada para la superficie que se desea ajustar.

**Fit options:** Al hacer clic en este botón se abrirá la ventana *Fitting Options* que permite escoger el nivel de rigidez, es decir, la sensibilidad de la curva a las variaciones locales de los datos, el orden del polinomio que define a la superficie, así como otras opciones.

**Number of Cuts:** Se explicó en la página anterior.

**Raw data spikes:** Seleccionar esta opción si se quiere que los puntos que representan los datos originales sean desplegados en la gráfica de superficie. (Si los puntos no se localizan directamente en la superficie, entonces serán conectados a ésta mediante una línea punteada) como se muestra en la siguiente figura:



Las siguientes opciones están disponibles únicamente para las gráficas de iconos y matrices de gráficas:

**Layout:** Modifica la distribución y la resolución de los iconos de la gráfica cambiando el número de renglones y columnas en los cuales se despliegan. También es posible modificar la asignación de las variables a los componentes consecutivos de cada icono en la gráfica mediante la ventana *Order of Variables* que se despliega al hacer clic en el botón *Var Order* (en el caso de las gráficas tipo *Chernoff Faces*, se abrirá la ventana *Chernoff Faces Definition* (Botón *Definition*) que permite cambiar las dimensiones de las características faciales de los iconos).

**Legend:** Permite desplegar la gráfica con (On) o sin (Off) leyenda.

**Case names:** En el caso de las gráficas de iconos, por omisión los nombres de los casos se despliegan bajo sus iconos respectivos. En esta opción se puede elegir no desplegarlos.



**Selection:** Hacer clic en este botón para seleccionar el patrón del marco utilizado para marcar los iconos seleccionados (esta opción es aplicable si la selección del icono fue definida cuando la gráfica fue especificada).

**Points:** Sirve para cambiar el patrón y tamaño de los puntos de la gráfica por medio de la ventana *Points Pattern* (esta opción es aplicable a las gráficas de iconos *Sun Ray* y *Star*).

**Lines:** Abre la ventana *Lines Pattern* donde es posible seleccionar un patrón de línea diferente. En el caso de Chernoff Faces se abrirá la ventana *Chernoff Faces Pattern* donde se puede cambiar los patrones de cada línea facial.

**Areas:** Abre la ventana *Area Pattern* donde se puede cambiar el patrón y el color de relleno de las áreas de los iconos.

**Icon info:** Este botón despliega una breve descripción de la actual gráfica de icono (v.g. número de casos, variables, condiciones de selección de iconos, etc).

**Histograms:** Frecuentemente los histogramas se colocan en la diagonal de matrices cuadradas de gráficas y en el primer renglón y columna de una matriz rectangular de gráficas. En esta opción se puede elegir el tipo de escala para los histogramas: **Common:** La escala se basará en un rango común para todos los histogramas lo que permite compararlos. Cuando se elije **Independent:** STATISTICA graficará cada uno de los histogramas basándose en rangos individuales; en este caso la comparación no es recomendable porque las escalas de los histogramas son diferentes.

**Matrix Type:** Despliega una matriz cuadrada de gráficas ya sea con formato de matriz **Full** o **Half**. En el formato *Full* cada relación entre variables X y Y se representa dos veces (una como "XY" y una como gráfica "YX"), en un arreglo simétrico con histogramas en la diagonal. Cuando una función es ajustada a los datos de una matriz cuadrada de gráficas, las funciones ajustadas en las mismas gráficas son diferentes dependiendo de cual variable es vertical y cual horizontal.

**Frame:** Abre la ventana *Line Pattern - Borders* donde se puede elegir tener un marco alrededor de la gráfica así como el color y el grosor de dicho marco.

**Patterns:** Sirve para cambiar los colores y patrones de paredes específicas de las gráficas.

### 4.3 MENUS Y BARRA DE HERRAMIENTAS

Algunas opciones del menú y algunos botones de la barra de herramientas en la ventana de gráficas se encuentran también disponibles al activar el spreadsheet, scrollsheets, o la ventana text/output y ya se han descrito en el capítulo anterior, por lo tanto, en esta sección únicamente se describirán las opciones exclusivas de la ventana de gráficas.



### 4.3.1 FILE

**OPEN GRAPH:** Permite abrir archivos gráficos creados en STATISTICA (\*.stg). Estos archivos contienen no sólo la gráfica, sino también los objetos y dibujos que se le hayan agregado, además de los valores representados en dicha gráfica.

Botón asociado:

**MERGE GRAPH:** Despliega la ventana *Open Graph File* que permite seleccionar una gráfica previamente guardada para unirla con la actual.

**SAVE:** Esta opción guarda la gráfica actual en el formato de STATISTICA (formato nativo, con extensión \*.stg). A diferencia de otros formatos tales como metafile y bitmap (ver abajo) este formato no sólo guarda la imagen sino también toda la información necesaria para continuar posteriormente la personalización de la gráfica o el análisis de los datos. Si la gráfica no ha sido guardada antes se abrirá la ventana *File Save* para especificarle un nombre y una ruta donde será guardada.

Botón asociado:

**SAVE AS:** Permite guardar la gráfica bajo un nombre diferente o en una unidad de disco o ruta distintos a los establecidos previamente.

**SAVE METAFILE:** Permite guardar la gráfica en formato metafile (\*.wmf, Windows metafile format), este formato guarda la imagen como un conjunto de descripciones o definiciones de todas las componentes de la gráfica y sus atributos (v.g. segmentos de línea, relleno, texto y sus atributos, etc.), por lo tanto, comparado con el formato bitmap (ver siguiente opción), metafile ofrece opciones más flexibles para la modificación y personalización de las gráficas en otras aplicaciones de Windows. Sin embargo, a diferencia del formato estándar de STATISTICA (\*.stg) esta opción guarda sólo la imagen de la gráfica e ignora cualquier objeto ligado o pegado adicionalmente, los datos y cualquier otra información analítica específica.

**SAVE BITMAP:** Permite guardar la gráfica en formato bitmap (\*.bmp) el cual guarda únicamente una representación pasiva de los píxeles que forman la imagen. Los archivos con este formato pueden ser abiertos en otras aplicaciones de Windows; pero las opciones de edición y personalización de la gráfica estarán limitadas únicamente a cortar, pegar, ampliar, reducir y dibujar sobre ella.

**PRINT GRAPH:** Al seleccionar esta opción se desplegará la ventana *Print Graphs* con las siguientes opciones:

**Printer:** Muestra el nombre de la impresora en uso.

**Use dithered color to fill background:** STATISTICA usará un color determinado para rellenar el fondo de la gráfica así como el de los demás objetos y texto personalizado que se le agregue.



**Use dithered color for solid lines:** Los colores de las líneas sólidas de la gráfica aparecerán sombreados en la impresión. (Esta opción puede ser útil cuando las líneas son gruesas).

**Min line thickness (printer pixels):** Permite especificar el ancho mínimo de las líneas de la gráfica (por omisión es 1 píxel).

**Print Quality:** Despliega los niveles de resolución para la impresora actual (alto, medio, bajo).

**Print to file:** Direcciona la salida de impresión hacia un archivo (este archivo ya tendrá todas las especificaciones necesarias para imprimir posteriormente la gráfica). Al elegir esta opción, se abrirá la ventana *Imprimir en un archivo* pidiendo el nombre del archivo de salida. Después se podrá imprimir la gráfica con el comando COPY de DOS (COPY archivo PRN, copiar el archivo a la impresora). Esta opción también permite crear archivos gráficos para ser usados por otras aplicaciones como archivos HPGL o Postscript.

**Copies:** Indicar el número de copias que se desea imprimir.

**Collate copies:** Cuando se desea imprimir varias páginas, permite imprimir cada una con su correspondiente número de copias.

**Setup...:** Se describió en la sección Scrollsheet del capítulo 3.

Botón asociado: 

**PRINT PREVIEW:** Al elegir esta opción se abrirá la ventana *Print Preview* desplegando la gráfica tal y como aparecerá impresa en el papel y con los siguientes botones de opciones:

**Margins:** Permite ajustar interactivamente la posición de la gráfica en la página mediante los márgenes que se definen en la ventana *Margins* que aparece al presionar este botón (también está disponible en la opción *Page setup*). El ancho de los márgenes puede ser cambiado directamente colocando el puntero del ratón en las líneas que los definen y arrastrándolo hasta el sitio deseado (simultáneamente su posición será desplegada en la ventana *Margins*).

**Printer setup:** Se describió en la sección Scrollsheet del capítulo 3.

**Print:** Imprime automáticamente la gráfica con las características establecidas previamente.

**Close:** Cierra la ventana *Print Preview*.


**PAGE SETUP:** Al igual que el botón *Margins* (ver arriba) esta opción despliega la ventana *Margins* que permite ajustar interactivamente la posición de la gráfica en la página mediante los márgenes.

**EXIT:** Finaliza la sesión con STATISTICA.



### 4.3.2 EDIT

**LINK/EMBED:** Permite ligar o incrustar objetos (con formato nativo, metafile o bitmap) a la gráfica actual. Los objetos gráficos que se van a ligar o incrustar no pueden contener a su vez objetos ligados o incrustados, pues serán ignorados.

Botón asociado: 

Al seleccionar la opción se abre la ventana *Embed/Link Files* que también se despliega haciendo doble clic en el objeto ligado o incrustado.

#### *Mode.*

**Embed:** Al activar esta opción, una copia del objeto que se quiere incrustar será colocada en la gráfica actual (formará parte de la gráfica actual).

**Link:** Esta opción únicamente ligará el objeto seleccionado a la gráfica actual, es decir, sólo una referencia del objeto ligado formará parte de la gráfica actual y no el objeto mismo (que es el caso anterior). Cualquier cambio realizado sobre el objeto original podrá efectuarse también sobre su imagen que aparece en la gráfica, para esto es necesario:

1. Guardar los cambios efectuados en el objeto original.
2. Guardar la gráfica de STATISTICA con la que está ligado el objeto.
3. Cerrar y volver a abrir la gráfica de STATISTICA para que se actualicen los cambios hechos en el objeto original.

**Format:** Seleccionar el formato (graph file, metafile, bitmap) del objeto que se va a ligar o incrustar. (Los formatos se explicaron en las opciones *Save* de la sección anterior).

**Dynamic:** Al activar esta opción, las coordenadas del objeto que se va a ligar o incrustar serán fijadas de acuerdo a la escala de los ejes X y Y (en las gráficas de dos dimensiones), de tal forma que si se cambian algunos aspectos de la gráfica como tamaño, escala o hay una rotación, el objeto incrustado no pierda su posición relativa y su tamaño respecto a ésta. Si se desactiva la opción, el objeto será asignado a una posición fija de la ventana sin ninguna relación con la escala de los ejes cartesianos de tal forma que si hay un cambio en la gráfica, el objeto permanecerá sin cambios.

Esta opción no está disponible para las gráficas de más dimensiones.

Es posible pasar de un modo fijo a un modo dinámico y viceversa mediante el correspondiente botón de la barra de herramientas. Primero se debe seleccionar (resaltar) el objeto haciendo clic en él y después presionar el botón:



**Modo fijo** (Botón presionado).



**Modo dinámico** (Botón sin presionar).



**Filled background:** El relleno del objeto incrustado se desplegará en la gráfica actual cubriendo una parte de ésta.

**Original proportion:** Permite cambiar el tamaño del objeto incrustado conservando su proporción original.

**Frame:** Permite colocar un marco alrededor del objeto incrustado seleccionando la forma, ancho y tamaño de éste mediante el botón *Frame Pattern*.

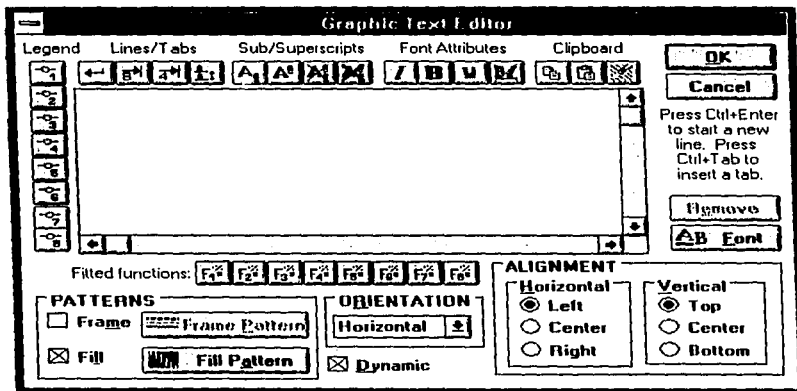
**Remove:** Este botón se puede presionar hasta que se tiene un objeto incrustado y se hace doble clic en su fondo para abrir nuevamente la ventana. Permite borrar dicho objeto de la gráfica actual. Notar que también se puede borrar seleccionando el objeto y presionando la tecla <Supr>.

Una vez que ya se han especificado las opciones anteriores, al presionar el botón **OK** se abrirá la ventana *Open Graph File* para escoger el archivo gráfico que se quiere ligar o incrustar; ya seleccionado el archivo, en la ventana de la gráfica actual se debe arrastrar el ratón para dibujar un cuadro donde será puesto el nuevo objeto.

**ENTER/DRAW:** Estas opciones permiten editar texto y dibujos.

**TEXT:** La ventana de gráficas de STATISTICA permite editar texto aparte del que por omisión aparece al generar cualquier gráfica, mediante el *Graphic Text Editor* (editor de texto para gráficas). Este editor contiene diversas opciones relativas a la posición, formato y despliegue del texto en la gráfica.

Botón asociado:



Haciendo doble clic en el texto que se acaba de insertar abre nuevamente el editor.



Los botones permiten personalizar el texto mediante los caracteres de control que aparecen al presionar cualquiera de ellos (estos caracteres también pueden insertarse manualmente) y que indican que el texto en la gráfica se desplegará con una determinada característica dependiendo del carácter de control que se haya activado.

Los caracteres de control que definen una opción de personalización específica siempre comienzan con el símbolo @ seguido por una letra que designa la opción y entre paréntesis rectangulares los valores disponibles para esa opción. Por ejemplo: @T[4] indica una tabulación de 4 espacios.

En la mayoría de los casos, los caracteres de control se colocan antes del texto que se quiere personalizar.

#### Legend:



Al hacer clic en cualquiera de estos botones, el texto editado aparece como leyenda de la gráfica, es decir, el patrón de dibujo de la gráfica respectiva (*plot 1* hasta *plot 8*) será colocado en un rectángulo junto al texto.

#### Fitted functions:



Cuando se tiene una función ajustada a los puntos de una gráfica, al hacer clic en cualquiera de estos botones, la ecuación de la función ajustada a la gráfica correspondiente (de la 1 a la 8) será desplegada en la parte del texto donde se coloque el carácter de control correspondiente al presionar el botón.

#### Lines/tabs:



Mueve el cursor del editor al principio de una nueva línea, esto también es posible presionando Ctrl +Enter.



Inserta una tabulación de 8 caracteres en el punto donde se encuentra el cursor.



Inserta la tabulación por omisión (4 caracteres) en el punto donde se encuentra el cursor.



Determina el espacio entre las líneas de texto. Por omisión el espacio es de 4 píxeles.



**Sub/superscripts:**

Estos botones permiten poner texto como subíndice texto subíndice o superíndice texto <sup>superíndice</sup> de otro.



Estos botones deshabilitan la opción de texto como subíndice o superíndice en un nivel (primer botón) o todos los niveles (segundo botón). Es decir, el texto que se escriba después de haber presionado el botón estará al mismo nivel del texto escrito antes de presionar los botones de subíndice o superíndice. El segundo botón vuelve automáticamente al nivel normal sin importar cuantos niveles haya habido antes.

**Fonts attributes:**

Estos botones permiten modificar el estilo de la letra a cursiva, negrita y subrayada respectivamente.



Deshabilita las opciones del estilo de letra y regresa al tipo definido por omisión.

**Clipboard:**

Estos botones permiten copiar, pegar y despegar respectivamente, mientras se está agregando o editando texto (alternativamente, presionar: Ctrl+C, Ctrl+V, y Ctrl+Z) Para cortar texto resaltado, presionar Ctrl+C.

**Alignment: Horizontal:** Permite alinear el texto a la izquierda, centro o derecha del punto de partida. **Vertical:** Permite posicionar el texto arriba, al centro o abajo del punto de partida.

**Frame:** Coloca un marco alrededor del texto. Se debe presionar el botón *Frame pattern* para abrir la ventana *Line Pattern* en la cual se selecciona el patrón, ancho, y color del marco.

**Filled Background:** Rellena el fondo del texto con el color que se desee. Se debe presionar el botón *Frame pattern* para abrir la ventana *Area Pattern* en la cual se selecciona el color y patrón del fondo del texto.

**Remove:** Este botón permanece desactivado hasta que se vuelve a abrir el editor haciendo doble clic en el texto insertado. Permite borrar dicho texto de la gráfica actual. Notar que también se puede borrar resaltando el texto y presionando la tecla <Delete> (<Supr>).

**Font:** Permite cambiar el tipo de letra mediante la ventana *Font* donde se define tipo, estilo y tamaño de la letra.






**Orientation:** Seleccionar la orientación del texto: Horizontal, Vertical izquierda, Vertical derecha.



**Dynamic:** Se puede elegir entre dos estados de texto: Fijo y Dinámico. El estado fijo hace que la posición del texto permanezca fija dentro de la ventana aún cuando se cambie la escala y tamaño de la gráfica. El estado dinámico cambia la posición del texto cada vez que se hace un cambio en la escala y tamaño de la gráfica para mantener la misma posición relativa respecto a ésta.

### RECTANGLE, ROUND RECTANGLE, OVAL, POLYLINE, ARROW

Seleccionar cualquiera de estas opciones en el menú, equivale a presionar el botón correspondiente de la barra de herramientas que permite dibujar la figura elegida (rectángulo, rectángulo redondo, óvalo, polilínea, flecha). Una vez presionado el botón, se debe hacer clic en el punto de la gráfica donde se quiere insertar el dibujo y arrastrar el ratón hasta que la figura tenga el tamaño deseado.

-  Rectángulo
-  Rectángulo redondo
-  Ovalo
-  Polilínea
-  Flecha

En el caso de las flechas, al presionar el botón aparece la ventana *Arrow Type Selection* con diferentes opciones para personalizarlas, pudiendo ver los efectos de cada opción (en el cuadro donde se dibuja la flecha) antes de aceptarlas.


Ya que fueron insertados, los objetos pueden cambiarse de tamaño o de lugar dentro de la ventana de gráficas y pueden tener un estado fijo o dinámico (ver la opción anterior: *Dynamic*) este estado se puede cambiar seleccionando el objeto con el ratón y presionando el botón: **Dyn**.

**Para cambiar el tamaño de un objeto:** Posicionarse en cualquier lugar de éste, hacer un clic para que aparezcan pequeños cuadros negros en las esquinas del dibujo, colocar el puntero en cualquiera de estos cuadros y cuando se vea una flecha de dos cabezas arrastrar el puntero del ratón hasta que el dibujo tenga el tamaño deseado.

**Para mover un objeto:** Posicionarse en cualquier lugar de éste, hacer un clic para que aparezca una flecha de 4 puntas y arrastrar el ratón hasta la posición deseada.


**MODIFY OBJECT:** Modifica los atributos del objeto seleccionado.

**LINE/FILL ATTRIBUTES:** Seleccionar estas opciones equivale a presionar los siguientes botones de la barra de herramientas:

 **Atributos de línea:** Permite cambiar el color, ancho y patrón de las líneas de los objetos insertados en las gráficas. Para esto, se debe seleccionar previamente el objeto



haciendo clic en cualquiera de sus líneas. Si no se ha seleccionado un objeto, los atributos definidos serán aplicados al nuevo objeto que se inserte.

 **Atributos de relleno:** Permite rellenar el fondo del objeto con un patrón bicolor o un color sólido, mediante la ventana *Area Pattern*. Para esto, se debe seleccionar previamente el objeto haciendo clic en cualquiera de sus líneas. Si no se ha seleccionado un objeto, los atributos definidos serán aplicados al nuevo objeto que se inserte.

Notar que si las ventanas correspondientes a estos botones son abiertas por vez primera, se debe cambiar el estado de despliegue (*Display*) a **ON** para poder hacer las elecciones requeridas.

**REMOVE OBJECT:** Borra el objeto seleccionado en la ventana de gráficas.


**PASTE SPECIAL:** Permite seleccionar el formato de despliegue del objeto que se desea pegar en una gráfica. Esta es la diferencia con la opción *Paste*, la cual elige el formato de acuerdo a una prioridad ya establecida; para objetos es: Nativo, Metafile, Bitmap; para texto es: Nativo, Metafile, Texto, Bitmap (los formatos se explicaron en las opciones *Save* de la sección anterior).


Generalmente los diversos formatos serán desplegados en el menú dependiendo de los que estén disponibles (i.e. dependiendo del contenido actual del portapapeles de Windows).


### 4.3.3 VIEW

**FIXED PROPORTION, CHANGE PROPORTION, ADJUST MARGINS:** Estas opciones permiten cambiar el tamaño de la gráfica ya sea en forma proporcional o desproporcional a sus dimensiones originales, así como ajustar su posición en la ventana mediante márgenes.

Seleccionar estas opciones equivale a presionar los siguientes botones de la barra de herramientas:

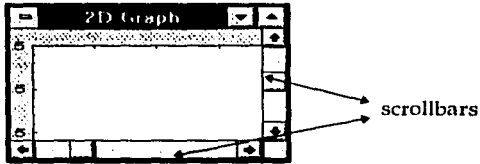
 **Proporción fija:** Al cambiar el tamaño de la ventana, la gráfica junto con todos sus objetos también cambiará de tamaño manteniendo la proporción con sus dimensiones originales.

 **Cambiar proporción:** Al cambiar el tamaño de la ventana, la gráfica junto con todos sus objetos también cambiará de tamaño pero sin mantener la proporción con sus dimensiones originales.

 **Ajustar márgenes:** Al presionar este botón aparecerán en la ventana de gráficas las llamadas scrollbars, que permiten repositionar la gráfica, moviéndola hacia arriba, abajo,



izquierda o derecha de la ventana permitiendo liberar espacio para pegar otros objetos si se desea.




Para restaurar las proporciones y márgenes originales de la gráfica usar la opción *Restore defaults*, que se describe más adelante.

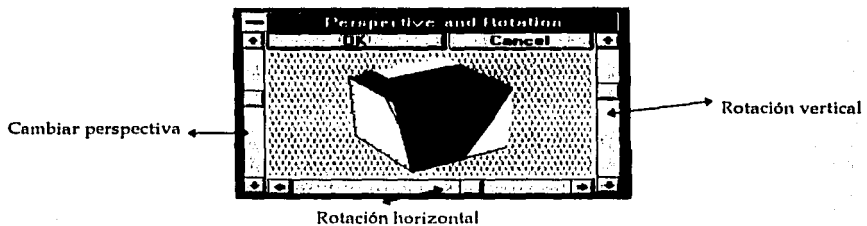
**PAGE/PRINTER PROPORTION:** Esta opción ajustará automáticamente las proporciones de la gráfica para que se adecúe al tamaño de la hoja de impresión que se definió en las ventanas *Print Preview*, *Margins* y *Printer Setup*. Es particularmente útil cuando se desea pegar más de una gráfica en una hoja en blanco (ver *Blank graph* en la sección *Graphs*) para imprimir más de una gráfica en una página.

**RESTORE DEFAULTS:** Cuando se hacen cambios a las proporciones y márgenes de la gráfica mediante las opciones *Fixed proportions*, *Change proportions*, *Adjust graph margins*, *Zoom in*, y *Zoom out* esta opción reestablecerá las proporciones y márgenes originales de la gráfica.

#### IGNORE MARKED POINTS:

 Este botón de la barra de herramientas permite seleccionar puntos individuales en la gráfica para marcar o resaltar (mediante un color diferente a los demás) los correspondientes valores de sus coordenadas en el *Graph Data Editor* (GDE, descrito en el menú *Layouts*). Al activar la opción *Ignore Marked Points*, los puntos marcados serán borrados temporalmente de la gráfica y sólo permanecerán marcados en el GDE (al imprimirlos aparecerá un asterisco junto a su valor). Si la opción se desactiva, entonces los puntos marcados no se borrarán de la gráfica y seguirán resaltados en el GDE.

**ROTATE:** Al seleccionar esta opción se abre la ventana *Perspective and Rotation* que permite cambiar la perspectiva o punto de vista así como la rotación de gráficas de 3 dimensiones.





Al presionar **OK** la gráfica original adoptará las modificaciones hechas en esta ventana.

**ANIMATED STRATIFICATION:** Esta opción sólo se encuentra disponible para las gráficas secuenciales de 3 dimensiones (*3D Sequential Graphs*). Genera la gráfica en capas que representan secciones o cortes determinados por los valores en X y Y que se definieron en la opción *Number of cuts* de la ventana *General Layout: 3D Sequence Graphs* (disponible en el menú *Layouts* o haciendo doble clic en el fondo de la gráfica). Después de generar cada corte hay una pausa donde **STATISTICA** espera que se presione cualquiera de los tres botones que aparecen en una pequeña ventana arriba de la gráfica:

**Continue:** Despliega el siguiente corte de la gráfica.

**Break:** Muestra la gráfica hasta el último corte que se desplegó con el botón anterior.

**Cancel:** Cancela la estratificación y despliega la gráfica original.

Botón asociado: 

**OBJECT SELECTION:** Esta opción habilita el botón izquierdo del ratón para poder seleccionar objetos en la gráfica haciendo clic en ellos o para abrir varias ventanas de opciones haciendo doble clic en las diferentes componentes de una gráfica (i.e. títulos, escalas, cuadrícula, fondo, etc.).

**DISPLAY FILTERS:** Cuando los puntos o el texto (valores de la escala o etiquetas de los puntos) de una gráfica se superponen y no se entiende lo que dicen, esta opción permite establecer un radio donde los caracteres superpuestos que se encuentren serán borrados de la gráfica, dejando sólo los puntos o el texto que se puedan leer. Esto se establece en la ventana *Display Filters* donde se indica el radio o nivel (*level*) del filtro (0=off hasta 3) tanto para puntos como para texto.

**ALIGNMENT GRID:** Despliega sobre la gráfica actual una cuadrícula de 16x16 (no imprimible) que sirve como referencia para colocar y alinear los objetos que se quieran agregar o incrustar en la gráfica. Esta cuadrícula es particularmente útil al usar la opción *Blank Graph* del menú *Graphs* ya que ahí aparece una hoja en blanco donde la cuadrícula servirá para alinear las gráficas y objetos que se coloquen en ella.

#### 4.3.4 LAYOUTS

**GENERAL LAYOUT:** En **STATISTICA** existen dos ventanas importantes que permiten dar un formato personalizado a las gráficas según las preferencias de cada usuario:

*General Layout:* Controla los atributos y componentes globales de las gráficas (títulos, cuadrículas, escalas, etc.). Ver sección *Elementos Principales* de este capítulo.

*Plot Layout:* Controla los atributos de gráficas individuales.



Al seleccionar esta opción se abrirá la ventana *General Layout* correspondiente a la gráfica actual.

**PLOT No...:** Despliega la ventana *Plot No...(Plot Layout)* de la gráfica actual.

**NEW PLOT:** Permite especificar el número de gráficas nuevas que se agregarán al editor de datos (ver opción *Edit Data* en esta sección) de gráficas para ser dibujadas así como el número de variables a graficar.

Ventana *New Plot(s)* :

**No of plots:** Indicar el número de gráficas.

**No. of variables:** Esta opción permite especificar el número de variables que serán usadas para definir los puntos de las nuevas gráficas. (2 para gráficas 2D, 3 para gráficas 3D). Al presionar **OK** se abre el editor de datos para gráficas para especificar ahí los valores de las variables. Se pueden usar los datos que se encuentran en el spreadsheet o scrollsheets resaltando el nombre de la columna de la cual se van a tomar y seleccionando la opción *Copy*, después en el editor de gráficas resaltar el nombre de la columna donde se quieren poner los datos y usar la opción *Paste*, de esta forma se pueden transferir datos para graficar no sólo desde el spreadsheet o scrollsheets sino desde otras aplicaciones de Windows (v.g. Excel).

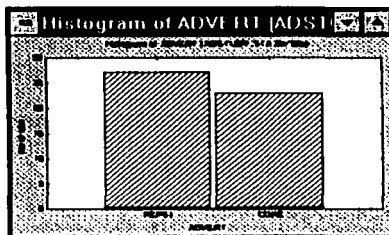
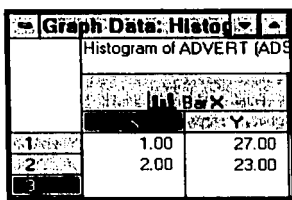
La opción *Weight* agrega una columna donde se puede indicar el peso de cada punto en la gráfica.

**No. of points:** Especificar el número de puntos a trazar en la nueva gráfica.



**Plot Layout:** Este botón abre la ventana de opciones para el diseño personalizado de las nuevas gráficas.


**EDIT DATA:** Despliega el editor de datos para gráficas o *Graph Data Editor* (GDE). Este editor es un scrollsheets especial que contiene los datos representados en la gráfica actual. Los datos para las gráficas de dos dimensiones están organizados en parejas (X,Y), y para las de 3 dimensiones en ternas (X,Y,Z) que representan las coordenadas de los puntos en la gráfica. Cada pareja o terna de columnas en el editor representa una gráfica cuyo nombre y tipo se despliegan en el encabezado de columna.


Botón asociado: 



Los menús y barra de herramientas de este editor permiten ejecutar diferentes operaciones de edición sobre los datos:

- Se pueden editar valores individuales haciendo doble clic en la celda correspondiente.
- Las opciones *Copy* y *Paste* permiten transferir datos de otras aplicaciones de Windows (v.g. Excel) directamente a este editor para graficarlos en STATISTICA con todas las ventajitas que este paquete proporciona.
- Se pueden agregar variables o columnas de datos mediante la opción *Add Plot* la cual abre la ventana *New Plot(s)* (descrita en la opción anterior) así como agregar nuevos renglones al final mediante *Add Rows*. Ambas opciones pertenecen al menú *Edit*.
- Haciendo doble clic en el encabezado de la columna o en el botón *Plot layout*  se abre la ventana *Plot No...* al igual que un doble clic en el área de título o en el botón  abre la ventana *General Layout* permitiendo personalizar el formato de despliegue de la gráfica.

Una vez que se hayan realizado todas las operaciones de edición deseadas, es conveniente presionar el botón  para desplegar la gráfica con los cambios (el editor permanece abierto).

 Cierra el editor de datos y vuelve a trazar la gráfica con las últimas modificaciones.

### 4.3.5 GRAPHS

Las primeras opciones de este menú corresponden a los tipos de gráficas explicados en la primera sección del capítulo.

**BLANK GRAPH:** Cuando se elige esta opción del menú *Graphs*, se abre una ventana de gráfica en blanco donde es posible agregar objetos gráficos nuevos o ya existentes



(ejemplo: flechas, dibujos a mano libre, gráficas previamente guardadas, ligar objetos, agregar texto, etc).

**DEFAULT PROPORTION:** Se abrirá una ventana de una gráfica en blanco que es proporcional al tamaño default de la ventana de gráfica. La proporción de la gráfica puede ajustarse mas adelante con la opción *Use Print Proportion* del menú *View* (la gráfica llenará el tamaño de una página de impresión).

**PRINTER / PAGE PROPORTIONS:** Abrirá una ventana de gráfica en blanco que será proporcional al tamaño de una página de impresión (como se especifica en las ventanas *Print Preview*, *Margins* y *Printer Setup*). La proporción de la gráfica en blanco será ajustada automáticamente al tamaño de una página de impresión. La proporción puede ajustarse a la proporción default en la opción *Restore Defaults* del menú *View*.

### 4.3.6 OPTIONS

**GLOBAL DEFAULTS:** Define las características generales de la ventana de gráficas mediante un conjunto de opciones que se aplican de manera global a cualquier gráfica que se genere:

**Background:** Define el color de fondo interior y exterior de la gráfica.

**Titles:** Define el tipo de letra de los títulos de la gráfica.

**Axes:** Establece la escala de medición para los ejes.

**Copy options to current graph:** Aplica los cambios a la gráfica actual de una gráfica previamente guardada.

**Get options from current graph:** Establece las características generales de la gráfica actual como las opciones por omisión para aplicarse en las gráficas subsecuentes.

**Restore System Defaults:** Reestablece las opciones por omisión que STATISTICA tenía definidas originalmente.

La lista de botones que comienza con *Border* y termina con *Specific Plots* despliegan una serie de ventanas que en general definen el color y la forma de cada uno de los elementos de las gráficas.

Los siguientes 4 botones de esta ventana son equivalentes a las opciones que se despliegan en el menú.

**Save:** Guarda en el archivo de inicialización de STATISTICA (*statist.ini*) los cambios hechos para que se apliquen por omisión a todas las gráficas que se generen posteriormente.





**Save as:** Permite guardar las opciones por omisión en un archivo distinto al *statist.ini*. Estas opciones van a poder ser establecidas nuevamente presionando el botón *Retrieve From*.

**Retrieve:** Reestablece las opciones por omisión guardadas en *statist.ini* (ver arriba) y las aplica a las gráficas que se generen posteriormente. Notar que si esta opción se elige directamente del menú, entonces las características por omisión que contiene *statist.ini* serán aplicadas automáticamente a la gráfica desplegada actualmente.


**Retrieve from:** Este botón abre la ventana *Retrieve defaults* donde se puede seleccionar un archivo distinto al *statist.ini* que contenga las opciones por omisión que se quieren establecer.

**GENERAL LAYOUT DIALOG:** En la ventana *General Layout* de cualquier gráfica aparece un recuadro (icono de la gráfica) que permite observar los cambios que se le hacen antes de activarlos al presionar **OK**. Esta opción permite definir los elementos (títulos, leyendas, escala) que se desea que aparezcan en el recuadro, sabiendo que mientras más elementos haya, el desplegado es más lento, también brinda la posibilidad de no desplegar la gráfica en dicho recuadro: desactivando la opción *Enable graph icon*.


**STAT GRAPHS y USER GRAPHS:** Son opciones que se aplican exclusivamente a las Gráficas Estadísticas SG por lo que no se describen en este manual. Si se desea obtener información acerca de estas opciones se recomienda consultar la ayuda.


### 4.3.7 BARRA DE HERRAMIENTAS

La mayoría de los botones ya han sido descritos con su correspondiente opción del menú sólo restan los siguientes:

 **Primer plano:** Cuando en una gráfica se tienen objetos adicionales (i.e óvalos, rectángulos, flechas, objetos ligados o incrustados, etc.) permite colocar en primer plano el objeto seleccionado. Es útil por ejemplo, cuando no se puede ver completamente porque hay otros objetos que lo cubren.

 **Ultimo plano:** Coloca el objeto seleccionado en último plano.

 **Incrementar fuente:** Incrementa globalmente y proporcionalmente el tamaño de la letra de todo el texto que aparece en la gráfica. Alternativamente se puede cambiar el estilo y tamaño de la letra de cualquier gráfica haciendo doble click en el texto deseado para abrir la ventana correspondiente donde aparece el botón *Font*.

 **Decrementar fuente:** Decrementa globalmente y proporcionalmente el tamaño de la letra de todo el texto que aparece en la gráfica. Alternativamente se puede cambiar el estilo y tamaño de la letra de cualquier gráfica haciendo doble click en el texto deseado para abrir la ventana correspondiente donde aparece el botón *Font*.



## CAPITULO 5

# REGRESION LINEAL

### 5.1 DESCRIPCION GENERAL

La regresión lineal múltiple es una técnica estadística que permite analizar la relación entre una variable dependiente o de respuesta y un conjunto de variables independientes o de predicción. Puede verse como un instrumento descriptivo por el cual la dependencia lineal de una variable con otras es resumida y analizada, o como un instrumento inferencial por el cual las relaciones en la población son evaluadas de datos de la muestra, tratando de responder a la pregunta de inferencia estadística, si se pueden generalizar los datos de una muestra observada al universo.

Los usos más importantes de esta técnica como instrumento descriptivo son: (1) Encontrar la mejor ecuación lineal y evaluar su exactitud en la predicción, (2) encontrar relaciones estructurales y proporcionar explicación a relaciones multivariadas aparentemente complejas.

El problema general de cálculo a resolver en el análisis de regresión lineal es ajustar una línea recta a un conjunto de puntos. En el caso simple (una variable independiente y una dependiente) se pueden visualizar fácilmente los puntos en una gráfica de dispersión o scatterplot donde se tiene una variable independiente "X" y una dependiente "Y". Estas variables pueden representar por ejemplo: Coeficiente Intelectual (CI) y Rendimiento Escolar (RE) respectivamente. Así, cada punto en la gráfica se refiere a las características (CI,RE) de un determinado estudiante. El objetivo es encontrar la mejor línea que se ajuste a esos puntos. Específicamente el programa hará cálculos para definir una línea para la cual la suma de los cuadrados de las distancias entre los puntos observados y ella, sea mínima. Este procedimiento es conocido como estimación por mínimos cuadrados.

En general, se estima una ecuación lineal de la forma:  $Y=B_0+B_1X_1+...+B_pX_p$

En esta ecuación, los coeficientes de regresión ( $B_i$ ) representan las contribuciones individuales de cada variable independiente X a la predicción de la variable dependiente Y. La línea de regresión ajustada expresa la mejor predicción de la variable dependiente Y dadas las variables independientes X.

**Linealidad:** El modelo supone linealidad en los parámetros. Si esto no se cumple, es posible hacer una transformación de variables (vía *Quick MML* o mediante fórmulas específicas que se editan en la opción *Current specs* del menú *Edit Variables*) o incluir



componentes no lineales en el modelo (ver la opción *Fixed nonlinear regression* del panel de inicio; además STATISTICA cuenta con el módulo *Nonlinear Regression*).

**Normalidad:** Las pruebas de significancia estadística generalmente son construidas basadas en suposiciones acerca de la distribución de las variables involucradas. En regresión múltiple se supone que las variables siguen una distribución normal  $N(\mu, \sigma^2)$ . Aunque la mayoría de las estadísticas de prueba son robustas a las violaciones de esta suposición, se recomienda revisar las distribuciones de las variables de interés. En este módulo se pueden producir histogramas con curvas normales sobrepuestas para ver las distribuciones de las variables a partir del *scrollsheet* que despliega las estadísticas descriptivas y que está disponible en cualquier momento en el programa.

**Limitaciones:** La mayor limitación conceptual radica en el hecho de que el análisis de regresión sólo descubre una asociación entre la variable de respuesta y las variables de predicción, en lugar de detectar una relación causa-efecto. La causalidad implica que un cambio en las X causará uno correspondiente en la variable de respuesta Y. Por ejemplo, cuando se calienta un metal éste se expande; en este caso no existe ninguna duda de que establecer una relación causa-efecto es muy importante. Pero desafortunadamente, lo anterior generalmente no puede establecerse con base en un análisis estadístico, a menos que se efectúe un experimento rigurosamente controlado.

**Selección del número de variables:** La mayoría de los autores recomiendan tener al menos de 10 a 20 veces más observaciones (casos) que variables. De otra forma, los estimadores de los coeficientes de la línea de regresión probablemente resultarán muy inestables e irrepetibles si se volviera a hacer nuevamente el estudio.

**Multicolinealidad y matrices mal condicionadas:** Este es un problema común en los análisis de correlación. Supóngase que se tienen dos variables de predicción para la altura de una persona: 1) Peso en libras y 2) Peso en onzas. Obviamente estos dos predictores son completamente redundantes. Tratar de decidir cual de las dos medidas es un mejor predictor de la altura sería bastante tonto; sin embargo, esto es exactamente lo que se trataría de hacer si se realizara un análisis de regresión tomando la altura como la variable dependiente (Y) y las dos medidas del peso como las variables independientes (X). STATISTICA mandaría un mensaje de matriz mal condicionada. Cuando se tienen muchas variables involucradas no es inmediato ver que este problema exista y se puede llegar a manifestar sólo después de que varias variables ya han sido incluidas en la ecuación de regresión. Por lo tanto, cuando ocurre este problema, significa que al menos una de las variables de predicción es completamente redundante. Este módulo contiene muchos indicadores estadísticos de este tipo de redundancia (tolerancias, semi-parcial R, etc., ver la sección correspondiente a *Model Definition*) así como algunas soluciones (v.g. *Ridge regression*).

**Importancia del análisis de residuales:** Aunque la mayoría de las suposiciones en análisis de regresión no pueden ser probadas explícitamente, importantes violaciones pueden ser detectadas y deben ser corregidas apropiadamente. En particular, los outliers (observaciones extremas) pueden afectar seriamente los resultados cambiando la dirección de la línea de regresión y proporcionando estimadores sesgados de B. Por lo tanto, una de



las metas en el diseño de este módulo fue facilitar al usuario de una manera rápida y eficaz, herramientas para el análisis de residuales que le permitan identificar fácilmente las violaciones a las suposiciones del modelo.

## 5.2 OPCIONES DEL MODULO

En esta sección se describen cada una de las ventanas que STATISTICA despliega a lo largo del análisis en el orden en que éstas aparecen. En cada una de ellas se tiene el botón **OK** que permite pasar a la siguiente ventana y el botón **Cancel** que regresa a la anterior.

### 5.2.1 PANEL DE INICIO

Al activar por primera vez el módulo *Multiple Regression* se despliega el correspondiente panel de inicio que permite definir las condiciones iniciales para el análisis:

**Variables:** Este botón permite desplegar la ventana de selección de variables donde se eligen las variables independientes y dependientes para el análisis (STATISTICA calcula una matriz de correlación para todas las variables seleccionadas). Se pueden elegir hasta 100 variables.

**Input file:** Indicar si el archivo de entrada va a contener los datos originales (**Raw Data**) tomados directamente del spreadsheet o va a contener una matriz de correlación (**Correlation matrix**) creada previamente en éste o cualquier otro módulo de STATISTICA que lo permita.

Para que STATISTICA reconozca un archivo con formato de matriz de correlación, éste debe cumplir las siguientes condiciones:

- El número de casos debe ser igual al número de variables+4.
- La matriz debe ser cuadrada y los nombres de los casos iguales a los de las variables.



- Los últimos cuatro casos deben contener los siguientes nombres e información:

*Means:* Este renglón debe contener la media de cada variable.

*Std. Dev:* Este renglón debe contener la desviación estándar de cada variable.

*No. Cases:* Indicar el número de casos del archivo de datos original del cual se creó la matriz.

*Matrix:* 1 indica que es una matriz de correlación.

Las matrices de correlación son cuadradas y simétricas y contienen los coeficientes de correlación por parejas de todas las variables especificadas.

Ejemplo de archivo con formato de matriz de correlación:

	VAR1	VAR2	VAR3
VAR1	1.00	.20	.30
VAR2	.20	1.00	.10
VAR3	.30	.10	1.00
MEANS	12	11	10
STD. DEV.	3	5	2
NO. CASES	50		
MATRIX	1		

Se pueden crear matrices de correlación salvando las correlaciones generadas en las ventanas correspondientes (ver la opción *Save correlations* de la ventana *Review Descriptive Statistics*) o de forma manual metiendo las correlaciones en un spreadsheet en el módulo *Data Management* e incluyendo en el archivo los últimos cuatro casos descritos arriba.

**MD deletion:** Esta opción se activa solamente si se ha especificado *Raw Data* en la anterior, proporciona 3 formas para tratar los casos con falta de dato (*missing data*):

**Casewise:** Excluye de todo el análisis cualquier caso que tenga un dato faltante.

**Mean substitution:** Los datos faltantes serán reemplazados por las medias de las respectivas variables (sólo para este análisis, no en el archivo de datos original).

**Pairwise:** Los casos con valores faltantes serán excluidos de los cálculos que involucren variables para las cuales ellos tienen datos faltantes. En el caso de las correlaciones, la correlación entre cada par de variables es calculada con los casos que tienen datos válidos para esas dos variables.

Se recomienda tener precaución al interpretar los resultados de la regresión si se ha elegido esta opción ya que STATISTICA realizará el análisis tratando la matriz de correlación como si estuviera basada en casos completos y usará el número mínimo de casos válidos de las variables para determinar los grados de libertad en pruebas de significancia estadística y estimar sumas de cuadrados. Por lo tanto, será siempre recomendable comparar los resultados obtenidos usando esta opción con los que se



obtengan usando las otras (*Casewise o Mean substitution*), para poder detectar discrepancias.

Si se llegaran a seleccionar más de 60 variables para el análisis, se tendrían las opciones **Pairwise short** (los cálculos son optimizados para obtener velocidad) y **Pairwise long** (los cálculos son optimizados para obtener precisión). En ambos casos la matriz de correlación, las medias y las desviaciones estándar son almacenadas en doble precisión.

### **Mode.**

**Standard:** El análisis de regresión se ejecutará en el modo estándar (todas las variables independientes seleccionadas formarán parte de la ecuación de regresión) calculando la matriz de correlación.

**Fixed non-linear:** Este modo permite transformar variables. Al activarlo, se abre una ventana similar a la actual (panel de inicio) donde se deben especificar las variables que van a ser transformadas; éstas se eligen en la ventana de selección de variables que se despliega al presionar el botón *Variables* (el análisis posterior se restringirá a las variables que aquí se seleccionen). Una vez realizada la elección, al presionar **OK** se abre la ventana *Non-linear Components Regression* que contiene las posibles transformaciones con sus rangos válidos. Se puede seleccionar más de una transformación para las variables elegidas y cada una se aplicará a todas las variables.

Para cada variable seleccionada, se crearán nuevas variables temporales (en memoria, el archivo de datos original no será modificado) que contendrán los resultados de las respectivas transformaciones y se incluirán en la matriz de correlación pero no podrán graficarse. Los nombres de las nuevas variables serán una combinación del número de la variable original y el símbolo de la transformación elegida. Por ejemplo:  $V4^{**2}$  indica que la variable número 4 se eleva al cuadrado.

El código de falta de dato por omisión (-9999) será asignado a las nuevas variables para un caso si:

La variable original tiene un dato faltante para el mismo caso.  
El valor de la variable transformada no cae dentro del rango válido.

Una aplicación común de esta técnica es estimar las componentes lineales, cuadráticas, cúbicas, etc. de la relación de una variable independiente con la dependiente. Es preciso señalar que STATISTICA incluye el módulo *Nonlinear Estimation* para estimar modelos de regresión no lineal.

**Perform default analysis:** Al activar esta opción y presionar **OK**, el programa aceptará las selecciones que por omisión tiene (el análisis se realizará bajo un modelo de regresión estándar con ordenada al origen o intercepción) la ventana *Model Definition* (descrita más adelante), y se desplegará la ventana de resultados del análisis (*Regression Results*). Si no se activa la opción y se pulsa **OK**, entonces se desplegará la ventana *Model Definition*.



**Review desc. stats, corr. matrix:** Al seleccionar esta opción y presionar **OK**, se desplegará la ventana *Review Descriptive Statistics* que permite visualizar las estadísticas descriptivas de las variables involucradas (la descripción completa de esta ventana se da en la siguiente sección). Pulsando **OK** en esta ventana (o si no se selecciona esta opción) el análisis sigue adelante y se despliega la ventana *Model Definition o Regression Results* (dependiendo de *Perform default analysis*).

**Extended Precision Computations:** El programa usará el algoritmo de precisión extendida al realizar los cálculos para generar la matriz de correlación que es usada en los cálculos del análisis de regresión múltiple. Aunque por omisión los cálculos se realizan en doble precisión ya que ofrece una exactitud superior y para casi todos los conjuntos de datos obtenidos de mediciones reales (i.e., no creados artificialmente) produce resultados idénticos a los que se obtienen usando la opción extendida.

**Batch processing and printing:** Esta opción se encuentra disponible solamente si se activó anteriormente *Perform Default Analysis* y si la impresora, archivo en disco, y/o ventana text/output están seleccionados actualmente en la ventana *Page/Output Setup* (ver opciones de impresión en el capítulo 3).

Esta opción afecta la forma en que la salida del análisis es impresa. Por omisión, el programa realiza un proceso interactivo (*Batch processing and printing* está desactivada) que permite determinar los scrollsheets que se desean imprimir. Cuando la opción se activa, los resultados se imprimen automáticamente conforme se van generando (en la impresora, archivo en disco y/o ventana text/output), es decir, si la opción está activa, el programa ejecuta el análisis de regresión específico, imprime los resultados y regresa a la ventana de definición del modelo: *Model Definition* (descrita posteriormente).

**Print residual analysis:** Si esta opción se encuentra activa, la salida impresa incluirá un scrollsheet con varias estadísticas residuales para cada caso:

Valor observado, valor estimado, residual, valor estimado estandarizado, residual estandarizado, error estándar del valor estimado, distancia de Mahalanobis, residual eliminado, distancia de Cook (ver botón *Display residuals & pred.* de la ventana *Residual Analysis*).

Esta opción se encuentra disponible solamente si se activaron anteriormente *Perform Default Analysis* y *Batch processing and printing* y si la impresora, archivo en disco, y/o ventana text/output están seleccionados actualmente en la ventana *Page/Output Setup* (ver opciones de impresión en el capítulo 3).



## 5.2.2 REVISAR ESTADISTICAS DESCRIPTIVAS (REVIEW DESCRIPTIVE STATISTICS)

Review Descriptive Statistics

missing data were casewise deleted

27 cases were processed  
27 valid cases accepted

---

<input type="checkbox"/> Means & SD	<input type="checkbox"/> Box & wh.	<input checked="" type="checkbox"/> SD=Sums of Squares/N	<input type="button" value="OK"/>
<input type="checkbox"/> Correlations	<input type="checkbox"/> Graph	<input type="checkbox"/> Save correlations	<input type="button" value="Cancel"/>
<input type="checkbox"/> Covariances			

Esta ventana intermedia en el análisis de regresión permite hacer una revisión de las estadísticas descriptivas específicas de las variables (originales y transformadas) seleccionadas para el análisis. En la parte superior se indica la forma en que se trataron los casos con falta de dato (missing data, ver en el panel de inicio la opción *MD deletion*).

**Means & SD:** Este botón despliega un scrollsheet con las medias y desviaciones estándar así como el número de casos involucrados en el cálculo de estas medidas para cada variable. Por omisión, la desviación estándar es calculada como la raíz cuadrada de la suma de las desviaciones cuadradas de las observaciones alrededor de la media dividida entre  $n-1$  (un estimador insesgado de  $\sigma^2$ ). Esta estadística puede ser usada para hacer inferencias sobre la población de la que fue extraída la muestra.

La gráfica por omisión para este scrollsheet es un histograma (de distribución de frecuencias) para la variable seleccionada con la gráfica de la distribución normal sobrepuesta. Esta gráfica es útil para verificar rápidamente si la suposición de normalidad ha sido o no violada por la variable respectiva.

**Box & whisker:** Este botón permite crear una gráfica *Box & whisker* para las variables seleccionadas. Este tipo de gráficas representan la tendencia central y la dispersión de cada una de las variables. Se puede elegir entre cuatro tipos de gráficas:

**Median/Quart/Range:** Describe la tendencia central de la variable en términos de la mediana (representada por el cuadrado). La dispersión en los valores de la variable está indicada por los cuartiles (percentiles 25° y 75°) representados con el cuadro más grande y por los valores mínimo y máximo de la variable (los "whiskers").

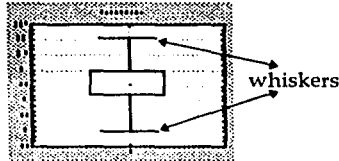
**Mean/SE/SD:** Aquí el cuadrado representa la media (tendencia central), mientras que la dispersión (variabilidad) está representada por  $\pm 1$  veces el error estándar (cuadro grande) y  $\pm 1$  veces la desviación estándar ("whiskers").





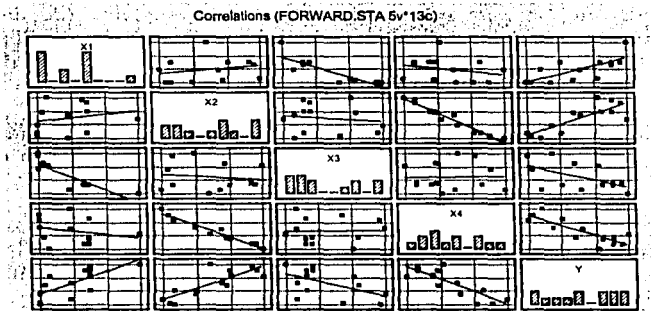
**Mean/SD/1.96\*SD:** Esta gráfica mostrará la media (cuadrado) de la variable, rodeada por el cuadro grande que representa  $\pm 1$  veces la desviación estándar. Si la distribución es normal, los "whiskers" representan un intervalo del 95% de confianza definido como la media  $\pm 1.96$  veces la desviación estándar de la variable.

**Mean/SE/1.96\*SE:** Esta gráfica mostrará la media (cuadrado) de la variable, rodeada por el cuadro grande que representa  $\pm 1$  veces el error estándar. Si la distribución es normal, los "whiskers" representan un intervalo del 95% de confianza definido como la media  $\pm 1.96$  veces el error estándar de la variable.



**Correlations:** Este botón despliega un scrollsheet con las correlaciones de todas las variables. Si se desean graficar dos variables que se encuentran en la matriz, resaltar la celda donde se cruzan dichas variables y seleccionar la gráfica *Scatterplot/conf.* del menú *Quick stat graphics*. La gráfica indicará la línea de regresión para las dos variables así como el intervalo de confianza del 95% para dicha línea.

**Graph:** Este botón produce una representación visual de las correlaciones entre las diferentes parejas de variables (mediante gráficas de dispersión o scatterplots) en un formato matricial. Al presionarlo, primero aparece la ventana para seleccionar las variables (mínimo 2) que se deseen graficar.



**Covariances:** Despliega un scrollsheet con las covarianzas entre cada pareja de variables involucradas en el análisis. Si se desean graficar dos variables que se encuentran en la matriz, resaltar la celda donde se cruzan dichas variables y seleccionar la gráfica



*Scatterplot/conf.* del menú *Quick stat graphics*. La gráfica indicará la línea de regresión para las dos variables así como el intervalo de confianza del 95% para dicha línea.

**Save correlations:** Permite guardar la matriz de correlación como un archivo estándar de STATISTICA (extensión \*.sta) para que pueda ser usada posteriormente como los datos de entrada, en lugar de los datos originales (ver la opción *Input file* del panel de inicio).

**Pairwise n:** Despliega un scrollsheets con el número de casos válidos para cada par de variables en la matriz de correlación. Este botón sólo está disponible si en el panel de inicio se eligieron las opciones *Raw data* en *Input file* y *Pairwise* en *MD deletion*.

**SD=sums of squares/n:** Esta opción afecta la forma en que son calculadas las desviaciones estándar así como las varianzas y covarianzas en las opciones *Means & SD* y *Covariances*. Si está activada, las estadísticas resultantes serán únicamente una descripción de la desviación estándar, varianza y covarianza de la muestra (no se deberán hacer inferencias acerca de la población de la cual se extrajo la muestra). Esta opción sólo afecta los valores respectivos en los scrollsheets, no influye en ningún cálculo posterior.

### 5.2.3 DEFINICION DEL MODELO (MODEL DEFINITION)

Esta ventana permite definir las características generales con las que se realizará el análisis.

**Model Definition**

**Variables**

Independent: X1 X7  
 Dependent: Y

Method: Standard

Intercept: Include in model

Tolerance: .01000 (Enter 0.0 to set to minimum=1.e-25)

Ridge regression; lambda: .100

Stepwise Multiple Regression:

1.0000 0.0000

7

Summary only

Batch processing/printing

Print residual analysis

**Variables:** Este botón despliega la ventana de selección de variables donde se eligen las variables independientes y dependientes para el análisis. Si se selecciona más de una variable dependiente, entonces el análisis de regresión será realizado consecutivamente para cada variable en la lista de dependientes.

**Method:** Permite elegir el método de selección de variables independientes para el modelo:

**Standard:** Todas las variables independientes entran en la ecuación de regresión en un solo bloque o paso.

**Forward stepwise:** Este método empieza con cero variables independientes en el modelo y trata de insertar una a cada paso de la regresión pudiendo ser borrada posteriormente al entrar otras variables, hasta que se obtiene el "mejor" modelo. La entrada o salida de cada variable depende de si el valor de una estadística de prueba (F-parcial) que STATISTICA calcula internamente es mayor o menor que los valores especificados en las opciones *F to enter* y *F to remove* respectivamente (ver abajo).

**Backward stepwise:** Este método empieza con un modelo que incluye todas las variables independientes en la ecuación y las va eliminando una a cada paso de la regresión hasta que el "mejor modelo" es obtenido. La salida de cada variable depende de si el valor de una estadística de prueba (F-parcial) que STATISTICA calcula internamente es menor que un valor predeterminado que se especifica en la opción *F to remove* (ver abajo).

Nota: STATISTICA lleva un registro automático de los análisis realizados durante una sesión. Así, cuando se selecciona el método *Standard*, el programa automáticamente calcula el incremento en  $R^2$  (coeficiente de determinación) que resulta de realizar diferentes análisis para la misma variable dependiente. Por ejemplo, si en un primer análisis se eligen las variables 1, 2 y 3 como independientes y en el siguiente se eligen 1, 2, 3, 4 y 5 (sin cambiar la variable dependiente) entonces el programa automáticamente calculará el incremento en  $R^2$  debido a las variables 4 y 5. Este resultado se observa al presionar el botón *Stepwise (summary)* de la ventana *Multiple Regression Results* (ver siguiente sección).

**Intercept:** Permite ajustar una ecuación de regresión con intercepción (**Include in model**) o un modelo de regresión a través del origen (**Set to zero**).

**Tolerance:** La tolerancia de una variable independiente se define como el porcentaje de variabilidad de la misma que no es explicada por las variables independientes que ya forman parte de la ecuación de regresión, y se calcula como 1 menos el cuadrado del coeficiente de correlación múltiple de esa variable con todas las demás variables independientes en la ecuación de regresión ( $1 - R^2$ ). Mientras más pequeña sea la tolerancia de una variable, más redundante será su contribución a la ecuación de regresión, provocando que ésta no pueda ser evaluada ya que la matriz estará mal condicionada y no podrá ser invertida.



La tolerancia mínima aceptable por STATISTICA es de  $1.00E-25$ . Sin embargo, no se recomienda poner un valor tan extremadamente pequeño ya que si la tolerancia de una variable que va a entrar en la ecuación es menor que el valor por omisión 0.01, significa que es 99% redundante con las variables que ya están en la ecuación.

**Ridge regression:** Este tipo de análisis de regresión se emplea cuando las variables independientes están altamente correlacionadas y no se pueden obtener estimadores estables para los coeficientes de regresión por el método ordinario de mínimos cuadrados. Específicamente, una constante ( $\lambda$ ) es sumada a la diagonal de la matriz de correlación la cual es entonces estandarizada nuevamente para que todos los elementos de la diagonal sean iguales a 1 (y los elementos fuera de la diagonal sean divididos entre esa constante). En otras palabras, *Ridge regression* decrementa artificialmente los coeficientes de correlación para que sean calculados estimadores (beta) más estables (todavía sesgados). Notar que los errores estándar para B (estimadores de los coeficientes de regresión) y beta (estimadores estandarizados de los coeficientes de regresión) son calculados usando las fórmulas para regresión estándar. La constante  $\lambda$  es proporcionada por el usuario.

**Stepwise multiple regression:** Las opciones dentro de este recuadro sólo están disponibles si en *method* se eligió *Forward* o *Backward stepwise*:

**F to enter, F to remove:** El valor que se especifique en *F to enter* determina que tan significativo debe ser la contribución de una variable a la regresión para que pueda entrar en el modelo. El valor que se especifique en *F to remove* determina que tan insignificante debe ser la contribución de una variable en la ecuación de regresión para que pueda ser eliminada del modelo. *F to enter* siempre debe ser mayor que *F to remove*.

Si en *method* se eligió *Backward stepwise* el programa incluirá todas las variables en la ecuación de regresión (paso 0) y entonces buscará la variable que tenga el valor más pequeño de la estadística F-parcial y lo comparará con *F to remove*, si  $F\text{-parcial} < F\text{ to remove}$  esa variable sale de la ecuación. Así, a cada paso se van eliminando variables hasta que ninguna variable en la ecuación de regresión tenga una F-parcial  $< F\text{ to remove}$ . Por lo tanto, si se quisieran borrar todas las variables de la ecuación, una a cada paso, sería necesario especificar un valor muy grande para *F to remove* y por lo tanto también para *F to enter*, por ejemplo 998 y 999.

Si se eligió *Forward stepwise*, el programa revisará a cada paso la lista de variables independientes que todavía no se han incluido en la ecuación de regresión y meterá aquella que tenga el mayor valor de F-parcial y que cumpla que  $F\text{-parcial} > F\text{ to enter}$ . El proceso termina cuando ninguna variable fuera de la ecuación cumple con este criterio. A cada paso, el programa también revisa las F-parciales de las variables que entraron anteriormente y si  $F\text{-parcial} < F\text{ to remove}$ , la variable vuelve a quedar fuera del modelo. Por lo tanto, en cada paso puede entrar o salir una variable. Si se quisiera que todas las variables entraran en la ecuación, una en cada paso, sería necesario especificar un valor muy pequeño para *F to enter* (v.g., 0.0001) y *F to remove* tendría que ser 0.0 ya que  $F\text{ to enter} > F\text{ to remove}$  siempre.

**Number of steps:** Seleccionar el número máximo de pasos en que STATISTICA realizará el análisis.

**Displaying results:** Permite elegir entre revisar el resultado final del análisis (Summary only) o revisar los resultados después de cada paso (At each step).

**Batch processing and printing:** Esta opción se explica en la sección Panel de Inicio de este capítulo.

**Print residual analysis:** Esta opción se explica en la sección Panel de Inicio de este capítulo y se encuentra disponible solamente si se activó anteriormente *Batch processing and printing* y si la impresora, archivo en disco, y/o ventana text/output están seleccionados actualmente en la ventana *Page/Output Setup* (ver opciones de impresión en el capítulo 3).

**Review correlation matrix / means / SD:** Este botón abre la ventana *Review Descriptive Statistics* descrita anteriormente.

## 5.2.4 RESULTADOS DE REGRESION MULTIPLE (MULTIPLE REGRESSION RESULTS)

Después de que se han especificado las condiciones iniciales del análisis, los resultados se mostrarán en la siguiente ventana:

Multiple Regression Results		
Multiple Regression Results (Step 3)		
Dep. Var. : Y	Multiple R : .84464576	F = 19.08621
	R <sup>2</sup> : .71342647	df = 3, 23
No. of cases: 27	adjusted R <sup>2</sup> : .67604731	p = .000002
	Standard error of estimate: .963966515	
Intercept: 3.503452287	Std.Error: .9151498	t( 23) = 3.8283 p < .0009
X6 beta=.595      X7 beta=.212      X1 beta=-.22 (significant beta's are highlighted)		
<input type="checkbox"/> Regression summary	<input type="checkbox"/> Partial correlations	<input type="button" value="OK"/>
<input type="checkbox"/> Analysis of variance	<input type="checkbox"/> Predict dependent var.	<input type="button" value="Cancel"/>
<input type="checkbox"/> Covar. of reg. coefficients	<input type="checkbox"/> Redundancy	<input type="checkbox"/> Residual analysis
<input type="checkbox"/> Current swagep matrix	<input type="checkbox"/> Stepwise (summary)	<input type="checkbox"/> Correlations and desc.
Alpha (display): <input type="text" value=".0500"/> <input type="button" value="Appl"/>		

En la primera sección se desplegarán los resultados generales del análisis en el siguiente orden:



**Step:** Número de etapas o pasos en los que se llevó a cabo el análisis.

**Dep. var.:** Indica el nombre de la variable que se eligió como dependiente.

**No. of cases:** Indica el tamaño de la muestra (N) (dependiendo de la opción *MD Deletion* del panel de inicio).

**Multiple R:** Proporciona el coeficiente de correlación múltiple que es una medida de la asociación lineal entre Y y las variables independientes ( $X_1, X_2, \dots, X_k$ ).

**R<sup>2</sup>:** Muestra el coeficiente de determinación múltiple que mide la variación total de la variable dependiente debida a las variables independientes.

**Adjusted R<sup>2</sup>:** El coeficiente de determinación ajustado es una corrección de R<sup>2</sup> para eliminar la influencia del azar incluido en el cálculo de dicho coeficiente. Se calcula con la siguiente fórmula:

$$R^2 \text{ ajustado} = R^2 - k(1 - R^2) / (N - k - 1)$$

Donde:

k: Número de variables independientes que han intervenido.

N: Tamaño de la muestra.

**Standard error of estimate:** Despliega el error estándar de la estimación, el cual mide la dispersión de los valores observados respecto a la línea de regresión.

**Intercept :** Indica el valor de la intercepción si es que se seleccionó el análisis de regresión que incluye intercepción (dentro de la ventana *Model Definition*).

**Std. error :** Muestra el error estándar de la intercepción.

El valor t y su respectivo nivel p se utilizan para probar la hipótesis de que la intercepción es igual a cero. Donde t(gl) es la estadística de prueba que se distribuye como una t de student con gl grados de libertad.

**F, df y p:** El valor F y su nivel p se utilizan para probar la hipótesis de que todos los coeficientes de regresión ( $B_i$ ) son iguales a cero. Donde F es la estadística de prueba que se distribuye como una F de Fisher con df grados de libertad.

En la siguiente sección de la ventana se muestran los coeficientes de regresión estandarizados (beta) resaltando los más significativos. El criterio para determinar la significancia estadística (alpha " $\alpha$ ") puede cambiarse del .05 por omisión mediante la opción *Alpha (display)* (ver abajo).

**Regression summary:** Al oprimir este botón se abre un scrollsheet con los coeficientes de regresión estandarizados (beta) y no estandarizados (B), sus errores estándar, y



significancia estadística. Se despliega en el encabezado el resumen del análisis de regresión ( $R$ ,  $R^2$ , etc.).

**Analysis of variance:** Despliega un scrollsheet con un análisis completo de la tabla de análisis de varianza de la ecuación de regresión actual.

**Covar. of reg. coefficients:** Esta opción da como resultado:

1. Un scrollsheet con las correlaciones de los coeficientes de regresión, y
2. Un scrollsheet con las varianzas y covarianzas de los coeficientes de regresión.

**Current sweep matrix:** Despliega en un scrollsheet una matriz con todas las variables (independientes y dependientes) elegidas para el análisis. La submatriz de las variables independientes que se encuentran actualmente en la ecuación de regresión es menos una vez la inversa de la matriz de correlación de esas variables (esto es, el signo invertido de cada elemento de la matriz). Los elementos de la diagonal para aquellas variables que no están en la ecuación pueden interpretarse como los valores  $(1-R^2)$ , tratando la variable respectiva como variable dependiente, y usando todas las variables independientes actuales.

**Partial correlations:** Despliega un scrollsheet con:

1. La beta in: Coeficiente de regresión estandarizado para la variable respectiva si fuera a entrar en la ecuación de regresión como variable independiente.
2. La correlación parcial: Correlación entre la variable respectiva y la variable dependiente, eliminando la influencia de las otras variables independientes ya incluidas en la ecuación.
3. La correlación semi-parcial: Correlación entre la variable dependiente sin ajustar con la variable respectiva, eliminando la influencia de las otras variables independientes ya incluidas en la ecuación.
4. La tolerancia de la variable independiente respectiva definida como uno menos el cuadrado de la correlación múltiple entre la variable respectiva y todas las variables independientes en la ecuación de regresión (ver sección Definición del modelo en este capítulo).
5. El coeficiente de determinación entre la variable respectiva y la variable dependiente.
6. La tolerancia mínima, es decir la tolerancia más pequeña entre todas las variables independientes si la variable respectiva entrara como una variable independiente adicional.
7. El valor de la estadística  $t$  de student que permite analizar la significatividad de cada parámetro beta mediante el nivel  $p$ .



Estas estadísticas serán desplegadas de manera separada para las variables que no se encuentran actualmente en la ecuación de regresión y para las variables dentro de la ecuación de regresión.

**Predict dependent var.:** Esta opción abre una ventana donde se introducen los valores para cada variable independiente con el fin de calcular el valor estimado para la variable dependiente basado en la ecuación de regresión actual.

**Redundancy:** Despliega un scrollsheet con varios indicadores de la redundancia de las variables independientes (incluidas o no en la ecuación actual). Específicamente, para cada variable, el scrollsheet mostrará:

1. La redundancia o tolerancia ( $1-R^2$ ).
2. El coeficiente de determinación  $R^2$  entre la variable respectiva y todas las otras variables independientes incluidas en la ecuación de regresión.
3. La correlación parcial.
4. La correlación semi-parcial.

**Stepwise (summary):** Este botón se encuentra activo si:

1. El usuario selecciona una regresión de tipo forward o backward (ver sección Definición del modelo de este capítulo).
2. Si en el análisis de regresión estándar previo, la misma variable dependiente fue analizada y la lista de variables independientes era un subconjunto de la lista actual de variables independientes.

Despliega un scrollsheet con la siguiente información para cada variable independiente:

El paso en el que entró o salió de la ecuación.

El coeficiente de correlación múltiple.

El coeficiente de determinación múltiple.

El cambio en  $R^2$  causado por remover o agregar variables en un paso.

El valore de F to enter o F to remove (ver sección Definición del modelo de este capítulo) y su respectivo nivel p.

El número de variables incluidas en la ecuación en cada paso.

**Correlations and descr.:** Abre la ventana *Review Descriptive Statistics* (descrita en la sección Revisar estadísticas descriptivas de este capítulo) donde puede elegir revisar las medias y desviaciones estándar, la matriz de correlación y la matriz de covarianza, o guardar la matriz de correlación en un archivo tipo matriz.

**Alpha (display):** El valor por omisión es .05. Este valor es usado como el nivel de significancia estadística, es decir, si el nivel p es menor que alpha , entonces las betas significativas son resaltadas. En esta ventana se puede editar este valor ( $.0001 < \alpha < .05$ ) y luego hacer clic en el botón *Apply* para aplicar este cambio.

**Residual analysis:** Este botón despliega la ventana *Residual Analysis* que contiene distintas opciones para llevar a cabo el análisis de residuales.





Residual Analysis		
Dep. Var. : Y	Multiple R : .84464576	F = 19.08621
	R <sup>2</sup> : .71342647	df = 3, 23
No. of cases: 27	adjusted R <sup>2</sup> : .67604731	p = .000002
	Standard error of estimate: .963966515	
Intercept: 3.503452287	Std.Error: .9151498	t( 23) = 3.8283 p < .0009

Statistics	Scatter Plots	Probability Plots
<input checked="" type="checkbox"/> Correlations & descr. (1)	<input checked="" type="checkbox"/> Pred. & residuals (D)	<input type="checkbox"/> Normal plot of reside (M)
<input checked="" type="checkbox"/> Regression summary (2)	<input checked="" type="checkbox"/> Pred. & squared reside (E)	<input type="checkbox"/> Half-normal plot (N)
<input checked="" type="checkbox"/> Display residuals & pred. (3)	<input checked="" type="checkbox"/> Pred. & observed (F)	<input type="checkbox"/> Detrended normal plot (P)
<input checked="" type="checkbox"/> Durbin-Watson stat (4)	<input checked="" type="checkbox"/> Obs. & residuals (G)	<input type="checkbox"/> Bivariate Scatterplots
<input checked="" type="checkbox"/> Save residuals & pred. (5)	<input checked="" type="checkbox"/> Obs. & squared reside (H)	<input checked="" type="checkbox"/> Bivariate correlation (Q)
<b>Casewise Plots</b>	<input checked="" type="checkbox"/> Resids & del. reside (I)	<input checked="" type="checkbox"/> Resids & indep. var. (R)
<input checked="" type="checkbox"/> Plots of residuals (A)	<b>Histograms</b>	<input checked="" type="checkbox"/> Pred. & indep. var. (S)
<input checked="" type="checkbox"/> Plots of outliers (B)	<input checked="" type="checkbox"/> Graph of observed (J)	<input checked="" type="checkbox"/> Partial resid. plot (T)
<input checked="" type="checkbox"/> Plots of predicted (C)	<input checked="" type="checkbox"/> Graph of predicted (K)	
	<input checked="" type="checkbox"/> Graph of residuals (L)	

Esta ventana sólo está disponible si el tipo de dato se especificó como *Raw Data* en el panel de inicio.

En la primera sección se exhiben nuevamente los resultados básicos del análisis de regresión.

### Statistics.

Los botones **Correlations and descr.** y **Regression summary** se describen en la sección anterior.

**Display residuals & pred.:** Despliega un scrollsheets con varias estadísticas para cada observación.

- Observed value: Es el valor observado de la variable dependiente.
- Predicted value: Es el valor estimado como resultado de la ecuación de regresión actual.
- Residual value: El valor observado menos el valor estimado.
- Standard predicted value: Es el valor estimado estandarizado de la variable dependiente.
- Standard residual value: Es el valor residual estandarizado (valor observado menos valor estimado entre la raíz cuadrada del cuadrado medio del error).
- Standard error of predicted value: Es el error estándar del valor estimado no estandarizado.
- Mahalanobis distance: La distancia de Mahalanobis es la distancia de un caso al centriode (punto que representa las medias de todas las variables independientes en el



espacio multidimensional) de todos los casos en el espacio definido por las variables independientes. Sirve para indicar si existen casos que son outliers para las variables independientes.

- **Deleted residual:** El residual eliminado es el valor residual estandarizado del caso respectivo si no hubiera sido incluido en el análisis de regresión. Si el residual eliminado difiere en gran medida de su respectivo residual estandarizado, es posible que este caso sea un outlier porque su exclusión cambia la ecuación de regresión.
- **Cook's distance:** Es otra medida del impacto del caso respectivo en la ecuación de regresión. Indica la diferencia entre los valores B calculados y los valores que uno habría obtenido si ese caso hubiera sido excluido. Todas las distancias deben ser de igual magnitud, si no, hay una fuerte razón para creer que el caso respectivo sesga la estimación de los coeficientes de regresión (B).

El propósito de todas estas estadísticas es identificar los outliers ya que una observación extremosa puede influenciar en gran medida las estimaciones finales. Es recomendable revisar siempre estas estadísticas y repetir el análisis después de eliminar cualquier outlier. Otra alternativa es repetir el análisis utilizando desviaciones absolutas en lugar de mínimos cuadrados para disminuir los efectos de los outliers. *Nonlinear Estimation* permite estimar dichos modelos.

**Durbin-Watson stat:** Este botón despliega un scrollsheet con la estadística D de Durbin-Watson y la correlación serial (correlación entre residuales adyacentes). La estadística de Durbin-Watson es útil para evaluar la presencia de una correlación serial de residuales, indicando así que las observaciones o casos en el archivo de datos no son independientes. Notar que toda prueba de significancia estadística en regresión múltiple supone que los datos provienen de una muestra aleatoria de observaciones independientes. De lo contrario, los coeficientes de regresión estimados (B) pueden ser más inestables de lo que se muestra en los niveles de significancia.

**Save residuals & pred.:** Este botón abre la ventana *Save Pred/Resid As* donde se indica el nombre del archivo en el que se guardarán los valores estimados y/o los residuales así como las variables que se especifiquen.

#### **Casewise plots.**

**Plots of residuals:** Abre una ventana donde se elige graficar: Los residuales originales (raw residuals), residuales estandarizados (standard residuals), distancias de Mahalanobis (Mahalanobis distances), residuales eliminados (deleted residuals), y distancias de Cook (Cook's distances). A continuación se abrirá un scrollsheet donde la opción elegida es graficada en la primera columna y en las columnas subsecuentes aparecen todas las estadísticas de residuales.

**Plots of outliers:** Se abrirá una ventana donde se puede elegir:

1. Graficar sólo los casos para los que el valor absoluto del residual estandarizado es mayor que +2, o



2. Graficar los 100 casos más extremos ordenados en forma descendente de acuerdo a la opción elegida: Standard predicted, standard residual, Mahalanobis distances, deleted residuals, Cook's distances.

Esta opción permite identificar outliers en archivos de datos muy grandes. En el scrollsheet, los valores seleccionados son graficados en la primera columna y en las siguientes son desplegadas todas las estadísticas de residuales.

**Plots of predicted:** Después de presionar este botón, se puede elegir graficar los valores estimados originales o los valores estimados estandarizados. En el scrollsheet los valores seleccionados son graficados en la primera columna y en las siguientes se despliegan todas las estadísticas de residuales.

### *Scatterplots.*

**Pred. & residuals:** Esta opción da como resultado un scatterplot de los valores estimados originales (en el eje X) contra los residuales originales (en el eje Y). Esta gráfica es útil para representar la tendencia lineal que existe entre las variables independientes y la variable dependiente.

**Pred. & squared resids:** Esta opción crea un scatterplot de los valores estimados originales (en el eje X) contra el cuadrado de los residuales originales (en el eje Y).

**Pred. & observed:** Esta opción permite crear un scatterplot de los valores estimados (en el eje X) contra los valores observados (en el eje Y). Esta gráfica es útil para identificar grupos potenciales de casos que no están bien estimados.

**Obs. & residuals , obs. & squared resids:** Estas gráficas son muy útiles para detectar outliers o grupos de observaciones que están sobre o por debajo de la línea estimada, graficando en el eje X los valores observados y en el eje Y los residuales.

**Resids & del. resids:** Esta opción crea un scatterplot de los residuales originales (eje X) contra los residuales eliminados (eje Y). Los residuales eliminados son los residuales estandarizados que se obtendrían si el caso respectivo fuera excluido de la estimación de la regresión múltiple. Así, si existe gran diferencia entre los residuales eliminados y los residuales estandarizados regulares, se concluye que los coeficientes de regresión no son muy estables.

### *Histograms.*

**Graph of observed:** Esta opción crea un histograma de los valores observados de la variable dependiente, mostrando también la gráfica de la distribución normal.

**Graph of predicted:** Se puede elegir graficar un histograma de los valores estimados originales o de los valores estimados estandarizados.



**Graph of residuals:** Después de presionar este botón aparecerá un submenú donde se elige graficar en un histograma: Los residuales originales (raw residuals), residuales estandarizados (standard residuals), distancias de Mahalanobis (Mahalanobis distances), residuales eliminados (deleted residuals), y distancias de Cook (Cook's distances).

### *Probability plots.*

**Normal plot of resids:** Las hipótesis del análisis de regresión establecen que las variables se distribuyen de forma normal y que la función de regresión es lineal, por lo que se puede esperar que los valores residuales se distribuyan de manera normal. Esta opción permite visualizar el patrón de residuales siguiendo una distribución normal. Sirve para ver si los residuales siguen o no una distribución normal, para identificar outliers y carencias de ajuste.

**Half-normal plot:** Esta gráfica se construye de la misma manera que la gráfica estándar de probabilidad normal (anterior), excepto que solo se considera la mitad positiva de la curva normal. Esta gráfica se utiliza cuando se quiere ignorar el signo de los residuales.

**Detrended normal plot:** Esta gráfica se construye en la misma forma que la gráfica estándar de probabilidad normal, excepto que la tendencia lineal es descartada. Esto con frecuencia extiende la gráfica, de ese modo le permite al usuario detectar patrones de desviación mas fácilmente.

### *Bivariate scatterplots.*

**Bivariate correlation:** Esta opción permite producir una scatterplot de cualesquiera dos variables del archivo de datos, sin importar si están incluidas o no en el análisis de regresión.

**Resids & indep. var:** Esta opción permite producir un scatterplot. Al presionar este botón se abre una ventana donde se especifican los valores que irán en el eje X: Los residuales originales (raw residuals), residuales estandarizados (standard residuals), distancias de Mahalanobis (Mahalanobis distances), residuales eliminados (deleted residuals), y distancias de Cook (Cook's distances). Después se debe indicar la otra variable para graficar (eje Y).

**Pred. & indep. var.:** Esta opción permite producir un scatterplot de los valores estimados (eje X) con cualquier variable en el archivo de datos (eje Y). Al presionar el botón, se abre una ventana donde se selecciona graficar los valores estimados originales o los valores estimados estandarizados. Después se indica la otra variable que se graficará.

**Partial resid. plot:** Esta opción produce una gráfica de residuales parciales para cualquier variable que se encuentre actualmente en la ecuación de regresión. En una gráfica de residuales parciales, el residual mas la contribución de la variable independiente respectiva ( $\text{residual} + B_i * X_{ii}$ ) es graficado contra los valores de esa variable independiente.



## 5.3 EJEMPLOS

A continuación se desarrollarán dos problemas; el primero enfocado a regresión lineal simple y el segundo a regresión lineal múltiple. Ambos se encuentran desarrollados en [Montgomery/Peck, 1992]. En cada uno de ellos se indicarán los pasos a seguir para obtener los resultados más comunes ( parámetros B y Beta, análisis de varianza, gráficas, etc.).

### 5.3.1 EJEMPLO 1.

#### Planteamiento:

Un ingeniero desea realizar una investigación acerca del uso de un molino de viento para generar electricidad. Para ésto ha realizado un conjunto de 25 mediciones de la velocidad del viento (millas por hora) y la corriente eléctrica generada por el molino (amperes).

Los datos obtenidos se encuentran en la siguiente tabla:

Observación <i>i</i>	Velocidad del viento <i>X</i>	Corriente generada <i>Y</i>
1	5.00	1.582
2	6.00	1.822
3	3.40	1.057
4	2.70	0.500
5	10.00	2.236
6	9.70	2.386
7	9.55	2.294
8	3.05	0.558
9	8.15	2.166
10	6.20	1.866
11	2.90	0.653
12	6.35	1.930
13	4.60	1.562
14	5.80	1.737
15	7.40	2.088
16	3.60	1.137
17	7.85	2.179
18	8.80	2.102
19	7.00	1.800
20	5.45	1.501
21	9.10	2.303
22	10.20	2.310
23	4.10	1.194
24	3.95	1.144
25	2.45	0.123

Parece lógico señalar que la corriente eléctrica generada (*Y*) guarda una relación con la velocidad del viento (*X*).

Específicamente se plantea la siguiente pregunta: ¿Qué cambios se producirán en la variable *Y* cuando se alteran los valores de la variable *X*? ¿Hasta qué punto pueden ser explicados los cambios que se producen en la variable *Y* por los cambios que se han producido en la variable *X*?



Antes de realizar el análisis para responder estas preguntas, es necesario crear un nuevo archivo de datos vía spreadsheet. Existen dos posibilidades:

1. La opción *New data* del menú *File* proporciona un spreadsheet que permite editar 10 casos y 10 variables. Como en este ejemplo se requieren 25 casos y 2 variables, se pueden agregar casos mediante la opción *Add* del menú *Edit-cases* (botón *cases*) y las variables restantes (8) se pueden eliminar en la opción *Delete* del menú *Edit-variables* (botón *variables*).
2. La opción *Create new file* del módulo *Data Management* proporciona un spreadsheet de un tamaño definido por el usuario. Así, para este ejemplo se debe especificar que se tienen 25 casos y 2 variables.

Así, una vez introducidos los datos, el archivo queda de la siguiente forma:

#	1 K	
1	5.000	
2	6.000	1.822
3	3.400	1.057
4	2.700	.500
5	10.000	2.236
6	9.700	2.386
7	9.550	2.294
8	3.050	.558
9	8.150	2.166
10	6.200	1.866
11	2.900	.653

Para obtener más información acerca de la entrada y edición de datos en el spreadsheet consultar los capítulos 2 y 3.

#### Especificación del análisis:

En el panel de inicio del módulo se debe seleccionar (botón *variables*) a la variable dependiente (Y) y a la independiente (X).

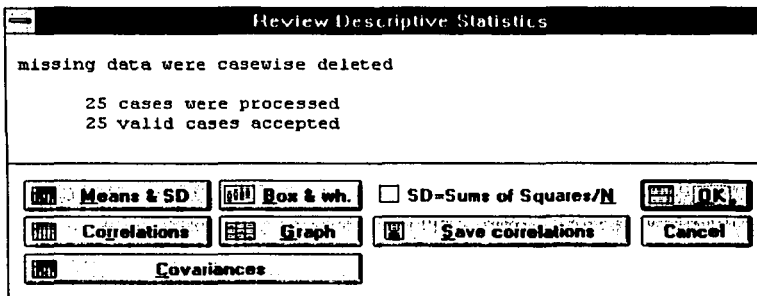
Puesto que los datos para el análisis se van a leer directamente del archivo de datos que se acaba de crear, en *Input file* seleccionar *Raw data*.

Como los 25 casos están completos (no existen celdas vacías) cualquier elección hecha en *MD deletion* no afectará el resultado del análisis.

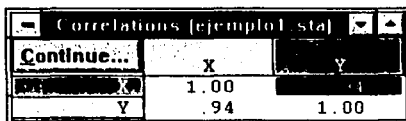
En *Mode* seleccionar el modo *Standard*.

Se realizará el análisis por omisión (*Perform default analysis*).

Por último, es conveniente seleccionar la opción *Review descr. stats. corr. matrix*, así, al presionar **OK** en el panel de inicio se abrirá la ventana que permite revisar algunas estadísticas descriptivas así como las correlaciones y covarianzas entre las variables.

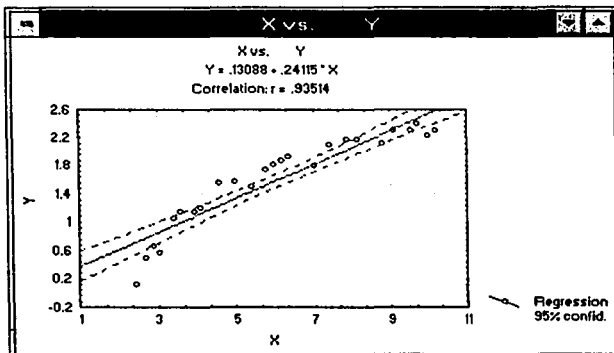


En este punto es conveniente dar un vistazo a la matriz de correlación de las variables presionando el botón *Correlations*.



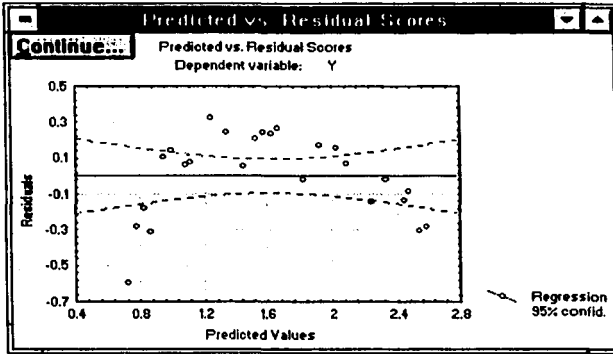
	X	Y
X	1.00	.94
Y	.94	1.00

A partir de este scrollsheet el programa permite obtener una gráfica de dispersión (scatterplot) de la pareja de variables que se encuentra resaltada (X,Y) para esto, en el menú *Graphs* se debe elegir *Quick stats graphs* y dentro de sus opciones *Scatterplot/conf.* para generar la siguiente gráfica:









La gráfica indica que la relación lineal no ha capturado toda la información de la variable X. Cabe resaltar que la curvatura observada en la gráfica X vs Y hecha anteriormente es amplificada notablemente en esta gráfica de residuales. Por lo tanto, es necesario proponer otro modelo.

De acuerdo a [Montgomery/Peck, 1992], un modelo razonable para estos datos sería:

$$Y = B_0 + B_1 (1/X) + \epsilon$$

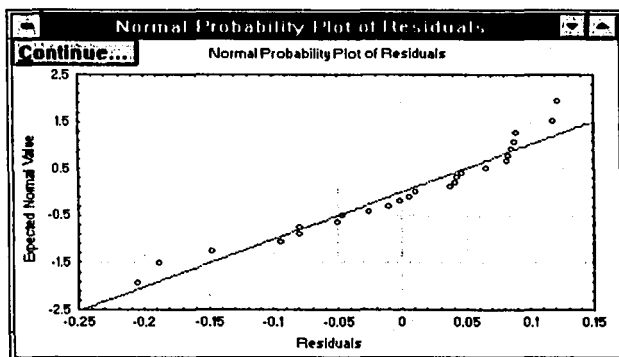
Para realizar el análisis de acuerdo a este nuevo modelo se debe activar nuevamente el panel de inicio del módulo. (Nota: Observar que las ventanas están consecutivas, es decir, si se presiona **OK** se despliega la siguiente ventana y si se presiona **Cancel** se regresa a la ventana anterior). En la opción *Mode* seleccionar *Fixed non-linear*. En la ventana que se despliega al presionar **OK** se deben elegir nuevamente las dos variables que van a intervenir en el análisis (botón *variables*). En la ventana *Non-linear Components Regression* seleccionar la transformación  $1/X$ . Esta transformación afectará a ambas variables (X,Y) por lo que automáticamente se crearán dos nuevas ( $1/V1$ ,  $1/V2$ , estas variables no pueden graficarse); a continuación se abre nuevamente la ventana *Model Definition* donde se debe indicar como variable dependiente a Y y como independiente a ( $1/V1$ ).

Los resultados de este nuevo modelo se muestran a continuación:





La gráfica de los residuales contra los valores estimados no revela ninguna tendencia en la distribución de los mismos.



Finalmente, al graficar en papel normal (botón *Normal plot of resids*) se observa una buena aproximación a esta distribución.

Por lo tanto, se concluye que el modelo transformado es el adecuado.

### 5.3.2 EJEMPLO 2.

#### Planteamiento:

Se desea investigar acerca de la relación que existe entre el calor (calorías/gramo) generado por el cemento y la cantidad (gramos) con que intervienen en la mezcla cuatro aditivos. Para esto se definen las siguientes variables:

Y: Calor generado

X1: Tricalcio de aluminio

X2: Tricalcio de silicio

X3: Tetracalcio de aluminio ferroso

X4: Dicalcio de silicio

Como en el ejemplo anterior, los datos se introducen en el spreadsheet para crear un nuevo archivo.



Data: F:\JMPLO2\STAT5V * 13c					
#	1 Y	2 X1	3 X2	4 X3	5 X4
1		7.000	25.000	6.000	60.000
2	74.300	1.000	29.000	15.000	52.000
3	104.300	11.000	56.000	8.000	20.000
4	87.600	11.000	31.000	8.000	47.000
5	95.900	7.000	52.000	6.000	33.000
6	109.200	11.000	55.000	9.000	22.000
7	102.700	3.000	71.000	17.000	6.000
8	72.500	1.000	31.000	22.000	44.000
9	93.100	2.000	54.000	18.000	22.000
10	115.900	21.000	47.000	4.000	26.000
11	83.800	1.000	40.000	23.000	34.000
12	113.300	11.000	66.000	9.000	12.000
13	109.400	10.000	68.000	8.000	12.000

Ahora se tienen más de dos variables por lo que en el panel de inicio se selecciona a Y como variable dependiente y a X1, X2, X3 y X4 como independientes.

En las demás opciones se toman los valores por omisión:

*Input file:* raw data  
*MD deletion:* casewise  
*Mode:* standard  
*Perform default analysis*

Y se presiona **OK** para acceder a la ventana de resultados; pero al estar ejecutando el análisis STATISTICA manda un mensaje diciendo que la tolerancia mínima ha sido rebasada por lo que la matriz está mal condicionada y no se puede seguir adelante en los cálculos.

Como se mencionó en la descripción general del módulo al inicio del capítulo, este problema se presenta cuando una o más variables independientes son redundantes. Para poder resolverlo, el programa despliega la ventana *Model Definition* donde es posible elegir un método de selección de variables que elimine la o las posibles variables redundantes. En la sección correspondiente a Definición del Modelo se describen los métodos disponibles: Standard, Forward stepwise y Backward stepwise.

El método Standard se eligió al principio, por lo que ahora se ejecutarán los dos siguientes y se compararán los resultados.

En primer lugar, en esta ventana (*Model Definition*) se debe elegir nuevamente a Y, como variable dependiente y a X1, X2, X3 y X4 como independientes. En la opción *Method* elegir *Forward stepwise* y en *Displaying results : at each step*; las demás opciones quedan con sus valores por omisión.



A continuación se presenta un resumen de los resultados del análisis. En cada paso se despliega la ventana *Multiple Regression Results* (para tener acceso a las ventanas consecutivas presionar Next).

Multiple Regression Results			
Multiple Regression Results (Step 1)			
Dep. Var. : Y	Multiple R : .82130504	F = 22.79852	
	R <sup>2</sup> : .67454196	df = 1, 11	
No. of cases: 13	adjusted R <sup>2</sup> : .64495487	p = .000576	
	Standard error of estimate: 8.963901935		
Intercept: 117.56793118	Std.Error: 5.262207	t( 11) = 22.342	p < .0000
X4 beta=-.82			
Multiple Regression Results (Step 2)			
Dep. Var. : Y	Multiple R : .98613947	F = 176.6270	
	R <sup>2</sup> : .97247105	df = 2, 10	
No. of cases: 13	adjusted R <sup>2</sup> : .96696526	p = .000000	
	Standard error of estimate: 2.734266120		
Intercept: 103.09738164	Std.Error: 2.123984	t( 10) = 48.540	p < .0000
X4 beta=-.68		X1 beta=.563	
Multiple Regression Results (step 3, final solution)			
minimum tolerance less than specifid limit			
Dep. Var. : Y	Multiple R : .99112837	F = 166.8317	
	R <sup>2</sup> : .98233545	df = 3, 90	
No. of cases: 13	adjusted R <sup>2</sup> : .97644727	p = .000000	
	Standard error of estimate: 2.308744955		
Intercept: 71.648306974	Std.Error: 14.14239	t( 9) = 5.0662	p < .0007
X4 beta=-.26		X1 beta=.568 X2 beta=.430	

Por último, los estimadores para el modelo final se pueden observar (botón *Regression summary*) en el siguiente scrollsheets:

Regression Summary for Dependent Variable						
MULTIPLE R = .99112837 R <sup>2</sup> = .98233545 Adjusted R <sup>2</sup> = .97644727						
REGRESS. F(3,9) = 166.83 p < .00000 Std. Error of estimate: 2.3087						
N=13	BETA	St. Err. of BETA	B	St. Err. of B	t(9)	p-level
Intercept			71.64831	14.14239	5.06621	.000675
X4	-.263183	.192806	-.23654	.17329	-1.36501	.205395
X1	.567737	.045748	1.45194	.11700	12.40998	.000001
X2	.430414	.191991	.41611	.18561	2.24184	.051687

El método Forward stepwise encontró a la variable X3 altamente redundante por lo que se eliminó del modelo final. Las demás estadísticas muestran que el modelo es aceptable (R<sup>2</sup> = 0.98, F = 166.83 es significativa con  $\alpha = 0.01$ ).



Ahora se realizará el análisis con el método Backward. En la ventana *Model Definition* es necesario definir una tolerancia menor que 0.01 (tolerancia por omisión) lo que indica que existe al menos una variable muy redundante (más del 99%), las demás opciones toman sus valores por omisión.

En seguida se muestra la forma en que se despliegan los resultados después del último paso.

**Stepwise Multiple Regression**

Backward Stepwise Regression; dependent variable: Y

Step: 2 F-to remove: 1.86 multiple R: .9893

Variables (R)emoved/(E)ntered:  
1 (R)X3            2 (R)X4

---

No other F to remove is less than specified limit.

Al presionar OK se tiene la ventana *Multiple Regression Results* con el resumen final del análisis. Para tener la información de los principales resultados en cada paso se debe presionar el botón *Stepwise (summary)*.

variable	Step in/out	Multiple R	Multiple R-Square	R-Square change	F - to entr/rem
X3	1	.991128	.982335	-.000040	.018233
X4	-2	.989282	.978678	-.003657	1.863262

El resumen de resultados con este método se encuentra en el siguiente scrollsheets (botón *Regression (summary)*).

Continue...						
R= .98928175 R²= .97867837 Adjusted R²= .97441405						
F(2, 10)=229.50 p<.00000 Std. Error of estimate: 2.4063						
N=13	BETA	St. Err. of BETA	B	St. Err. of B	t(10)	p-level
X1	.574137	.047431	1.46831	.121301	12.10465	.000000
X2	.685017	.047431	.66225	.045855	14.44236	.000000

El método Backward además de la variable X3 también elimina a X4. El valor de F aumenta considerablemente en este método (F= 229.50) lo que hace más confiables estos resultados.



## CAPITULO 6

# ESTADISTICA NO PARAMETRICA

### 6.1 DESCRIPCION GENERAL

Los procesos inferenciales que realizan estimaciones con respecto a los parámetros de la población de interés son conocidos como métodos paramétricos. En general estos métodos son más sensibles a las suposiciones para muestras de tamaño pequeño y, para muchos de ellos, su aplicación se encuentra limitada a aquellas muestras con distribución normal y a observaciones que tienen carácter cuantitativo (observaciones definidas sobre un intervalo o sobre una escala de proporciones).

La mayor parte de los métodos paramétricos como son la construcción de intervalos de confianza, las pruebas de hipótesis estadísticas y el ajuste de ecuaciones son aplicables a todas aquellas observaciones que se encuentran definidas, por lo menos, sobre una escala de intervalo.

Sin embargo, en muchas situaciones lo que se observa tiene un carácter cualitativo, es decir, los números no tienen ningún significado físico, más allá que el de representar con un número más grande la respuesta más favorable. Tal es el caso en las ciencias sociales y encuestas de mercado.

Las observaciones de este tipo pueden definirse sobre una escala ordinal, dado que las distancias entre dos puntos no es de consecuencia y sólo es importante el orden o rango de los números. En algunas ocasiones, las observaciones sólo pueden definirse sobre una escala nominal, donde el principio de orden no es de consecuencia.

Se han desarrollado procedimientos inferenciales llamados métodos no paramétricos o métodos de distribución libre que no se encuentran sujetos a la forma de la distribución de la población de interés y no requieren, en forma necesaria, que las observaciones se definan por lo menos en una escala de intervalo.

Los métodos no paramétricos pueden aplicarse en aquellas situaciones para las que las observaciones se definan, por lo menos, en una escala ordinal o nominal. Pero si las observaciones se definen al menos en escala de intervalo y la distribución de la población de interés es normal, los métodos no paramétricos son menos eficientes comparados con los procedimientos paramétricos que se basan en la suposición de normalidad.

El módulo *Non parametric Statistic & Distribution Fitting* contiene un amplio conjunto de estadísticas descriptivas e inferenciales para el análisis de datos ordinales y nominales.



Básicamente, existe al menos una prueba no paramétrica equivalente para cada tipo de prueba paramétrica.

En general, estas pruebas caen dentro de las siguientes categorías:

Pruebas de diferencias entre grupos (muestras independientes).

Pruebas de diferencias entre variables (muestras relacionadas).

Pruebas de relaciones entre variables.

Diferencias entre grupos independientes:

En el caso paramétrico, cuando se tienen dos muestras y se quieren comparar sus medias, se emplea la prueba  $t$  para muestras independientes; sus alternativas no paramétricas serían: La prueba de las rachas de Wald- Wolfowitz, La prueba  $U$  de Mann- Whitney y la prueba de Kolmogorov-Smirnov. Si se tienen más de dos muestras entonces, en el caso paramétrico se usaría el ANOVA mientras que en el no paramétrico se tiene el análisis de varianza de una clasificación por rangos de Kruskal-Wallis y la prueba de la mediana.

Diferencias entre variables:

Si se desea comparar dos variables medidas en la misma muestra, la prueba  $t$  para muestras relacionadas sería la prueba paramétrica apropiada, sus equivalentes no paramétricos son: La prueba del signo y la prueba de Wilcoxon. Si las variables de interés son clasificatorias, (v.g. masculino vs femenino) entonces la prueba de Mc'Nemar es la apropiada. Si se tienen más de dos variables medidas en la misma muestra, es conveniente la prueba de Friedman y la  $Q$  de Cochran; ésta es particularmente útil para medir los cambios de las frecuencias a través del tiempo.

Relaciones entre variables:

Para expresar relación entre dos variables, por lo regular se calcula el coeficiente de correlación de Pearson. En el caso no paramétrico se cuenta con varias medidas equivalentes: La  $\rho$  de Spearman, la  $\tau$  de Kendall y el coeficiente Gamma ( $\gamma$ ). Si las dos variables de interés son clasificatorias (v.g. pasó vs falló y hombre vs mujer) las estadísticas apropiadas para probar la relación entre las dos variables son : La prueba  $\chi^2$ , el coeficiente  $\phi$ , y la prueba exacta de Fisher. Además también se cuenta con una prueba de relación entre múltiples variables: El coeficiente de concordancia de Kendall.

Estadísticas descriptivas:

Cuando los datos no están normalmente distribuidos y las medidas a lo más están en una escala de medición ordinal, entonces el cálculo de las estadísticas descriptivas más comunes (v.g. media y desviación estándar) a veces no es la forma más apropiada para describir y resumir los datos. Este módulo permite calcular una amplia variedad de medidas de localización (media, mediana, moda, etc) y dispersión (varianza, desviación estándar, rango intercuartil, etc.) que permiten dar una descripción más completa de los datos.





## Ajuste de distribuciones:

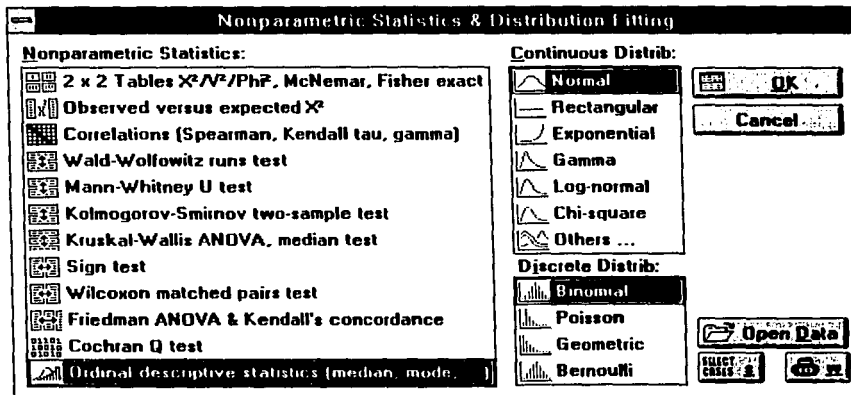
En algunas aplicaciones de investigación uno puede formular hipótesis acerca de una distribución específica de una variable de interés. Por ejemplo, la variable cuyos valores están determinados por un número infinito de eventos aleatorios independientes seguirá la distribución normal: Se puede pensar que la altura de una persona es el resultado de muchos factores independientes, tales como numerosas predisposiciones genéticas específicas, enfermedades de la infancia, nutrición, etc. y como resultado, la altura de la población tiende a seguir la distribución normal. Por otro lado, se tiene que si los valores de la variable son el resultado de eventos raros, entonces la variable seguirá la distribución Poisson. Por ejemplo, los accidentes industriales pueden ser vistos como el resultado de la intersección de una serie de eventos desafortunados y poco probables. El módulo *Nonparametrics and Distributions* permite probar la bondad de ajuste de los datos con respecto a la distribución hipotética.

Para obtener una explicación detallada de cada una de las pruebas, se recomienda consultar [Siegel/Castellan, 1988].

## 6.2 OPCIONES DEL MODULO

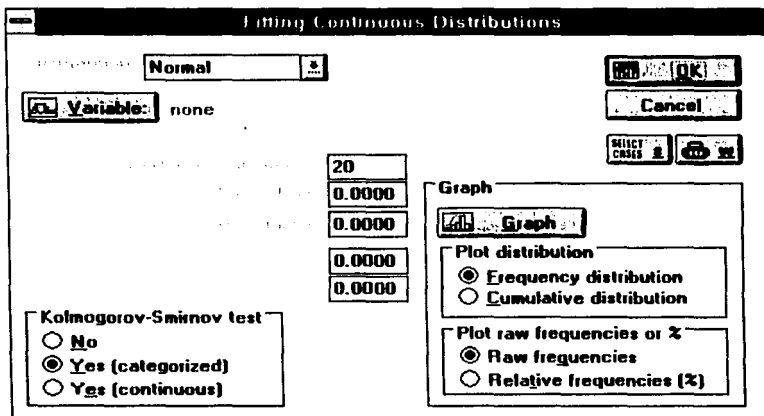
Al inicializar el módulo se despliega el panel de inicio con el menú que da acceso (haciendo doble clic en la opción correspondiente) a todas las pruebas estadísticas no paramétricas y distribuciones incluidas en el módulo:

### 6.2.1 PANEL DE INICIO



Cuando se desea verificar si los datos observados de una variable siguen una distribución de probabilidad determinada, se debe seleccionar el nombre de dicha distribución

(haciendo doble clic) dentro de los recuadros **Continuous distrib** o **Discrete distrib** para abrir la siguiente ventana:



**Fitting Continuous Distributions**

DISTRIBUTION: Normal

Variable: none

Number of categories: 20, 0.0000, 0.0000, 0.0000, 0.0000

**Kolmogorov-Smirnov test**

No

Yes (categorized)

Yes (continuous)

**Graph**

Graph

**Plot distribution**

Frequency distribution

Cumulative distribution

**Plot raw frequencies or %**

Raw frequencies

Relative frequencies (%)

En esta ventana se calculan dos pruebas de bondad de ajuste:  $\chi^2$  y Kolmogorov-Smirnov.

**Prueba  $\chi^2$ :** Se utiliza en datos clasificados en diferentes categorías para probar la existencia de una diferencia significativa entre un número observado de objetos o respuestas de cada categoría (frecuencia observada) y un número esperado (frecuencia esperada), basado en la distribución teórica específica. Las categorías donde la frecuencia esperada es menor que 5 son compactadas para formar una categoría más grande. Si esta prueba es significativa, se rechaza la hipótesis de que los datos observados siguen la distribución establecida.

**Prueba de Kolmogorov-Smirnov:** Esta prueba lleva consigo la especificación de la distribución de frecuencia acumulativa que ocurriría bajo la distribución teórica y su comparación con la distribución de frecuencia acumulativa observada. Se aplica a observaciones en escala ordinal.

Descripción de la ventana:

**Distribution:** Indica la distribución teórica a la que se ajustarán los datos.

**Variable:** Especificar la variable que será ajustada a la distribución indicada.

**Number of categories:** En esta opción se especifica el número de categorías en la tabla de frecuencia o histograma. En el campo **Lower limit** se controla el límite inferior y en el campo **Upper limit** el límite superior.



En la siguiente parte de la ventana se muestran los parámetros con valores por omisión para la distribución correspondiente. El usuario puede personalizar dichos valores.

**Kolmogorov-Smirnov test:** Esta prueba puede calcularse de dos formas:

**Yes (categorized):** La estadística D es calculada en base a las categorías establecidas en el scrollsheet de resultados (ver abajo).

**Yes (Continuous):** En este caso el programa utiliza los datos originales tal y como se encuentran en el spreadsheet; los ordena y calcula la frecuencia acumulada esperada en cada punto y posteriormente calcula la estadística D.

En ambos casos D es la diferencia absoluta más grande entre la distribución acumulada observada y la esperada.

**Graph:** Después de que las variables han sido seleccionadas, se producirá una gráfica con las siguientes opciones:

**Plot distribution:** Indicar la distribución que se graficará: La distribución de frecuencia (**Frequency distribution**) o la distribución acumulativa (**Cumulative distribution**).

**Plot raw frequencies or %:** Si se elige graficar la distribución de frecuencia, se puede seleccionar entre graficar las frecuencias absolutas (**Raw frequencies**) o las frecuencias relativas (**Relative frequencies (%)**).

Al presionar **OK** se desplegará el siguiente scrollsheet donde además de los resultados de las pruebas se muestran las frecuencias absolutas, absolutas acumuladas, relativas y relativas acumuladas de los puntajes observados y esperados, utilizadas para el cálculo de  $\chi^2$ .

Variable GENDER : distribution: Normal (aggress sta)									
Kolmogorov-Smirnov d = .3361964, p < .01									
Chi-Square: 25.22503, df = 1, p = .0000005									
Upper Boundary	Observed freq	cumulative observe	percent observe	cumul. observe	expected freq	cumulative expected	percent expected	cumul. expected	observed - expected
<= .90000	0	0	0.00	0.0	2.881	2.88	12.01	12.0	-2
1.00000	12	12	50.00	50.0	1.050	3.93	4.38	16.4	10
1.10000	0	12	0.00	50.0	1.271	5.20	5.30	21.7	-1
1.20000	0	12	0.00	50.0	1.481	6.68	6.17	27.8	-1
1.30000	0	12	0.00	50.0	1.661	8.34	6.92	34.8	-1
1.40000	0	12	0.00	50.0	1.793	10.14	7.47	42.2	-1
1.50000	0	12	0.00	50.0	1.863	12.00	7.76	50.0	-1



## 6.2.2 TABLAS DE $2 \times 2$ $\chi^2/N^2/PHI$ , McNEMAR, PROBABILIDAD EXACTA DE FISHER ( $2 \times 2$ TABLES $\chi^2/N^2/PHI$ , McNEMAR, FISHER EXACT)

Esta opción despliega una ventana que permite al usuario representar frecuencias en una tabla de contingencia de  $2 \times 2$  y así poder calcular varias estadísticas que permitan evaluar la relación existente entre dos variables dicotómicas:

		Variable 2	
		Categ I	Categ II
Variable 1	Categ I	A	B
	Categ II	C	D

**Prueba  $\chi^2$  para dos muestras independientes.**

Esta prueba se aplica cuando se tienen dos muestras independientes clasificadas en dos categorías, se utiliza para determinar la significancia de las diferencias entre las dos muestras. Acepta una escala de medición nominal.

		Categ I	Categ II
Muestra I		A	B
Muestra II		C	D

**Estadística  $V^2$ .**

Esta es una corrección de  $\chi^2$  para el tamaño de la muestra. Si las frecuencias en la tabla son pequeñas (menores que 10 en promedio) entonces, la estadística  $V^2$  será siempre más pequeña que la estadística  $\chi^2$  y por lo tanto, mas confiable en el resultado.

**Coefficiente  $Phi^2$ .**

Es una medida de correlación entre dos variables clasificadoras, basada en el hecho de que  $\chi^2$  es directamente proporcional a  $N$  (total de datos de la muestra). Sus valores van de 0 a 1.

**Prueba de McNemar para la significancia de los cambios.**

Esta prueba es apropiada para los diseños experimentales de "antes y después" en los que cada sujeto es su propio control. Sirve para probar la significancia de cualquier cambio observado en el experimento. Son de interés los cambios en las celdas A y D. Es aplicable a muestras relacionadas con escala de medición nominal y ordinal.

		Después	
		resp II	resp I
Antes	resp I	A	B
	resp II	C	D



**Probabilidad exacta de Fisher.**

Esta prueba calcula la probabilidad exacta bajo la hipótesis nula de observar el conjunto actual de frecuencias en la tabla de 2 x 2 o uno más extremo cuando los totales marginales se consideran fijos.

Ejemplo:

Se realiza un estudio para saber la preferencia que tienen hombres y mujeres en relación a la nueva o vieja fórmula de una soda. De 50 hombres, 41 prefieren la nueva fórmula sobre la vieja; de 50 mujeres, sólo 27 prefieren la nueva fórmula.

**Especificación del análisis:**

Los datos se introducen como sigue:

En esta ventana se introducen las frecuencias dentro de la tabla de 2 x 2.

Al presionar **OK** los resultados de esta opción se desplegarán en un scrollsheet con la siguiente información:

2 x 2 Table {adstudy.sta}			
	Column 1	Column 2	Row Totals
Percent of total	41	9	50
Percent of total	41.000%	9.000%	50.000%
Frequencies, row 2	27	23	50
Percent of total	27.000%	23.000%	50.000%
Column totals	68	32	A= 100
Percent of total	68.000%	32.000%	
Chi-square (df=1)	9.01	p= .0027	
V-square (df=1)	8.92	p= .0028	
Yates corrected Chi-square	7.77	p= .0053	
Phi-square	.09007		
Fisher exact p, one-tailed		p= .0025	
two-tailed		p= .0049	
McNemar Chi-square (A/D)	4.52	p= .0336	
Chi-square (B/C)	8.03	p= .0046	



A) Esta parte del scrollsheets contiene las frecuencias absolutas y relativas de cada entrada de la tabla, así como las frecuencias marginales de los renglones y columnas. Enseguida se despliegan los valores de las estadísticas de prueba:

Chi-square (df=grados de libertad)

V-square (df= grados de libertad)

Yates corrected Chi-square: Es una corrección por continuidad necesaria para disminuir el error ocasionado por utilizar datos con distribución discreta y aproximarlos a una distribución continua ( $\chi^2$ ).

Phi-square

Fisher-exact: Calcula la probabilidad exacta de Fisher para una y dos colas.

McNemar: Calcula dos valores de  $\chi^2$ : A/D prueba la hipótesis que las frecuencias en las celdas A y D son idénticas B/C prueba la hipótesis que las frecuencias en las celdas B y C son idénticas.

Se puede concluir para este ejemplo lo siguiente:

El valor de  $\chi^2$  para estos datos (9.01) es altamente significativo, lo que indica que las preferencias expresadas por los hombres son diferentes a las de las mujeres (la nueva fórmula es más aceptada por los hombres).

Phi<sup>2</sup>: Muestra que la correlación entre el género y la preferencia resultó ser muy pequeña (.09).

Fisher exact p: Suponiendo que en la población hombres y mujeres no difieren en sus preferencias, la probabilidad exacta de obtener dicha distribución de frecuencias en la tabla dadas las frecuencias marginales (50, 50, 68, 32) , es pequeña (.0025 para una cola y .0049 para dos colas).

McNemar  $\chi^2$ : Es necesario recordar que esta prueba es aplicable a muestras relacionadas y a modelos de "antes y después" por lo que en este ejemplo, los resultados no tienen significado.

### 6.2.3 PRUEBA $\chi^2$ PARA UNA MUESTRA (OBSERVED VERSUS EXPECTED $\chi^2$ )

Esta opción permite evaluar el ajuste de frecuencias observadas a un grupo de frecuencias esperadas mediante la estadística  $\chi^2$ .

Ejemplo:

El archivo de datos utilizado es: C:\STAT\EXAMPLES\ACCIDENT.STA



Case	Number of accident	
	1 Y_1983	2 Y_1985
January		85
February	150	80
March	80	85
April	50	40
May	40	45
June	43	40
July	80	42
August	75	43
September	80	50
October	65	36
November	50	78
December	95	83

Este ejemplo se basa en un conjunto de datos ficticios de accidentes de tránsito en una avenida en particular. Los datos fueron registrados para cada mes durante dos años: 1983 y 1985. Supóngase que se implantó un sistema de seguridad para disminuir los accidentes en 1984. Si el sistema fue un fracaso (la hipótesis nula), entonces se estima que el número y distribución de los accidentes en 1985 son los mismos que en 1983. Así, los datos para 1983 servirán como los valores esperados, y los datos para 1985 como los valores observados.

#### Especificación del análisis:

Al seleccionar esta opción se despliega la siguiente ventana:

Observed vs. Expected Frequency	
<input type="button" value="Variables"/>	<input type="button" value="OK"/>
with observed: Y_1985	<input type="button" value="Cancel"/>
with expected: Y_1983	<input type="button" value="SELECT CASES"/>
	<input type="button" value="HELP"/>

**Variables:** Este botón permite seleccionar el nombre de las variables que serán utilizadas. Introducir var1 (Y-1985) como la variable con las frecuencias observadas (**with observed**) y la var2 (Y-1983) como la variable con las frecuencias esperadas (**with expected**).

Los resultados de esta prueba se desplegarán en un scrollsheet con la siguiente información:



Observed vs Expected Frequencies [accident sta]				
Continue...	Chi-Square = 121.7010    df = 11    p < .00000		NOTE: Unequal sums of obs. & exp. frequencies	
Case	observed Y_1985	expected Y_1983	O - E	(O-E)**2 /E
January	100.00	125.00	-40.0	12.80
February	80.00	150.00	-70.0	32.67
March	85.00	80.00	5.0	.31
April	40.00	50.00	-10.0	2.00
May	45.00	40.00	5.0	.63
June	40.00	43.00	-3.0	.21
July	42.00	80.00	-38.0	18.05
August	43.00	75.00	-32.0	13.65
September	50.00	80.00	-30.0	11.25
October	36.00	65.00	-29.0	12.94
November	78.00	50.00	28.0	15.68
December	83.00	95.00	-12.0	1.52
Sum	707.00	933.00	-226.0	121.70

En la parte superior de la ventana se muestra el valor de la estadística  $\chi^2$ , los grados de libertad y el nivel p.

Notar que el scrollsheet contiene los totales anuales, las diferencias de los valores observados y esperados, y el cuadrado de la diferencia dividida entre los valores esperados (componentes de la estadística  $\chi^2$ ).

Con base en el valor de  $\chi^2$  (121.701,  $p < 0.0000$ ) se puede concluir que existe una diferencia significativa en el número de accidentes entre 1983 y 1985.

## 6.2.4 COEFICIENTES DE CORRELACION (CORRELATIONS (SPEARMAN, KENDALL TAU, GAMMA))

Esta opción permite calcular tres alternativas diferentes al coeficiente de correlación de Pearson:  $\rho$  de Spearman,  $\tau$  de Kendall y Gamma ( $\gamma$ ).

### Coefficiente de correlación de rangos de Spearman.

Es una medida de asociación aplicable a variables con escala de medición al menos ordinal. Arregla a las N observaciones individuales en estudio en dos series ordenadas de acuerdo a dos variables. Esta medida de correlación de rangos, determina la relación entre los valores de las dos variables.

### Coefficiente de correlación de Kendall.

Es una medida de correlación conveniente para la misma clase de datos que utiliza Spearman. A cada pareja de valores la clasifica como concordante o discordante  $\tau$  dará una medida del grado de asociación o correlación entre los dos conjuntos de valores.





Generalmente la magnitud de  $\tau$  es diferente a la de  $\rho$  ya que tanto la lógica como la fórmula empleadas para su cálculo difieren. Por lo tanto, no son directamente comparables entre sí.

### Gamma ( $\gamma$ )

Este coeficiente es preferible a  $\rho$  de Spearman o a  $\tau$  de Kendall cuando los datos contienen muchas observaciones empatadas. Gamma es básicamente equivalente a  $\tau$ , excepto que los empates son tomados en cuenta explícitamente.

Ejemplo:

Data		
Correlate		
#	1	2
	AUTH	STRIV
1		42
2	98	46
3	87	39
4	40	37
5	116	65
6	113	88
7	111	86
8	83	56
9	85	62
10	126	92
11	106	54
12	117	81

El archivo de datos para este ejemplo es: C:\STAT\EXAMPLES\STRIVING.STA

Se desea buscar información acerca de la correlación entre los puntajes de autoritarismo y de los esfuerzos por alcanzar posición social de 12 estudiantes.

### Especificación del análisis:

Después de seleccionar esta opción, se desplegará la siguiente ventana donde el usuario puede seleccionar las variables, y un tipo específico de correlación:

Nonparametric Correlations

**Variables**

Variable list 1: AUTHORIT  
Variable list 2: STRIVING

Correlation: Spearman R

Compute: Detailed report

**Matrix plot**

**Variables:** Este botón permite seleccionar al menos dos variables para las que se calcularán las correlaciones (cada variable en la primera lista será correlacionada con cada variable en la segunda lista).

Seleccionar la variable *Authorit* como la primera variable, y la variable *Striving* como la segunda variable.

**Correlation:** Permite elegir entre el coeficiente  $\rho$  de Spearman (**Spearman R**),  $\tau$  de Kendall (**Kendall tau**) o **Gamma**.

**Compute:** Se puede elegir entre calcular correlaciones de manera individual o en matrices:

**Detailed report:** Despliega las correlaciones entre las variables indicadas y también el número válido de observaciones (N), el valor de la estadística t de Student (o Z) para evaluar la significancia de la correlación con su respectivo nivel p.

**Matrix of two lists:** Esta opción calculará las correlaciones entre las variables especificadas en la primera lista contra las variables de la segunda lista.

**Square Matrix:** Esta opción calculará las correlaciones de una lista de variables (resultando la matriz cuadrada de correlaciones). Notar que si se especificaron dos listas de variables; serán consideradas como una sola.

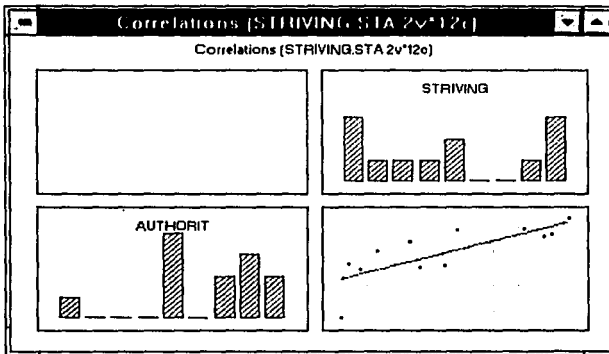
**Matrix plot:** Produce una matriz de gráficas de dispersión o scatterplots para todas las variables seleccionadas.

La siguiente ventana muestra los resultados después de haber seleccionado la opción *Detailed Report* para el coeficiente de Spearman.

Spearman Rank Order Correlations [striving sta]				
NONPAR STATS		MD pairwise deleted		
Pair of Variables	Valid N	Spearman R	t(N-2)	p-level
AUTHORITARY STRIVING	1	.818182	4.500000	.001143

Se desplegará para cada par de variables (var. primera lista, var. segunda lista), el número de observaciones (N), el valor del coeficiente de correlación, el valor de la estadística t de Student (o Z) con su respectivo nivel p.

La correlación entre las dos escalas es altamente significativa ( $R=.81$ ,  $p=.00114$ ), y se puede concluir que los individuos autoritarios también buscan alcanzar una posición social. Se puede visualizar la correlación desplegando la gráfica por omisión para este scrollsheet.



### 6.2.5 PRUEBA DE RACHAS DE WALD-WOLFOWITZ (WALD-WOLFOWITZ RUNS TEST) PRUEBA U DE MANN WHITNEY (MANN WHITNEY U TEST) Y PRUEBA DE DOS MUESTRAS DE KOLMOGOROV-SMIRNOV (KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST)

Estas pruebas son aplicables cuando se desea saber si dos muestras independientes provienen de la misma población. La prueba de Wald-Wolfowitz y la de Kolmogorov-Smirnov descubren cualquier clase de diferencia (tendencia central, dispersión, simetría, etc.) en la forma de la distribución de las dos muestras, mientras que la de Mann-Whitney únicamente descubre diferencias de localización (tendencia central); sin embargo es la más poderosa. Todas las pruebas suponen que la variable considerada tiene distribución continua y su medida es al menos ordinal.

En muestras grandes, para evaluar la significancia estadística se utiliza la aproximación a la distribución normal ( $Z$ ). Para muestras pequeñas (menos de 20 elementos) se recomienda consultar las tablas de [Siegel/Castellan, 1988] o [Conover, 1980].

Ejemplo:

En este ejemplo se observaron a 12 niños y 12 niñas de 4 años de edad durante dos sesiones de juego de 15 minutos; el juego de cada niño durante ambos periodos fue calificado por incidencia y grado de agresión. La hipótesis nula a probar es que la incidencia y grado de agresión son iguales en los niños de 4 años de ambos sexos.

Se utilizó el archivo: `C:\STAT\EXAMPLES\AGGRESSN.STA`.



Data AGGRESSN		
Aggression score		
#	GENDER	AGGRESSN
1	BOYS	86
2	BOYS	69
3	BOYS	72
4	BOYS	65
5	BOYS	113
6	BOYS	65
7	BOYS	118
8	BOYS	45
9	BOYS	141
10	BOYS	104
11	BOYS	41
12	BOYS	50
13	GIRLS	55

### Especificación del análisis:

Estas tres pruebas (Wald-Wolfowitz, Mann-Whitney y Kolmogorov-Smirnov) tienen la misma ventana para la introducción de los datos.

Wald-Wolfowitz Runs Test

Variables

Grouping variable: GENDER

Dependent variables: AGGRESSN

Code for Group 1: BOYS

Code for Group 2: GIRLS

Box & whisker

Categorized histogram

OK Cancel

SELECT CRISIS

Nota: El archivo de datos debe contener una variable clasificadora (variable independiente) con al menos dos valores distintos (códigos) que identifiquen a los miembros de cada muestra o grupo.

**Variables:** Despliega la ventana *Two-variable lists selection*. En la primera lista, se debe seleccionar la variable que contiene los códigos de clasificación (**grouping variable**) y en la segunda, la lista de variables que van a compararse en los dos grupos (**dependent variables**).

En este caso, seleccionar la variable *Gender* como variable clasificadora o independiente, y la variable *Aggressn* como la variable dependiente. (Notar que pueden realizarse simultáneamente varias pruebas seleccionando una lista de variables dependientes).



Luego se seleccionan los códigos que serán utilizados únicamente para identificar los sujetos de cada género. Seleccionar 1 (*boys*) y 2 (*girls*) (ver siguiente opción).

**Code for group 1, 2:** Especificar los códigos que serán usados en la variable clasificadora para identificar los miembros de cada grupo.

**Box & whisker:** Este botón despliega gráficas para comparar la distribución de la variable dependiente en los diferentes grupos definidos por la variable independiente o clasificadora. Pueden graficarse medianas y cuartiles, o medias y otras medidas de dispersión (consultar el capítulo 4 para una descripción más detallada de las opciones en este tipo de gráficas).

**Categorized histogram:** Esta opción producirá un histograma de la variable dependiente para cada clase o grupo definido por la variable clasificadora.

Los resultados de estas tres pruebas se desplegarán en un scrollsheet con la siguiente información:

En la parte superior de la ventana se indica la variable clasificadora que se utilizó en la prueba, así como los códigos utilizados para crear el grupo 1 y 2.

Wald-Wolfowitz Runs Test [aggressn.sta]										
By variable GENDER										
Group 1: 1-BOYS      Group 2: 2-GIRLS										
Variable	Valid BOYS	Valid GIRLS	Mean BOYS	Mean GIRLS	Z	p-level	Z adj	p-level	No. o Runs	No. o ties
AGGRESSN	12	81	27	-4	.0	3.5	.0	4	0	

La prueba de Wald-Wolfowitz despliega para cada variable dependiente:  
 El número válido de observaciones (n1 y n2) y la media para cada grupo.  
 El valor de la estadística de prueba Z, Z ajustada y su nivel p.  
 El número de rachas o corridas y el número de empates.

Mann-Whitney U Test [aggressn.sta]										
By variable GENDER										
Group 1: 1-BOYS      Group 2: 2-GIRLS										
variable	Rank BOYS	Rank GIRLS	U	Z	p-level	Z adjust	p-level	Valid BOYS	Valid GIRLS	
AGGRESS	84.0	6.00	-3.8	.00	-3.8	.00	12	12		

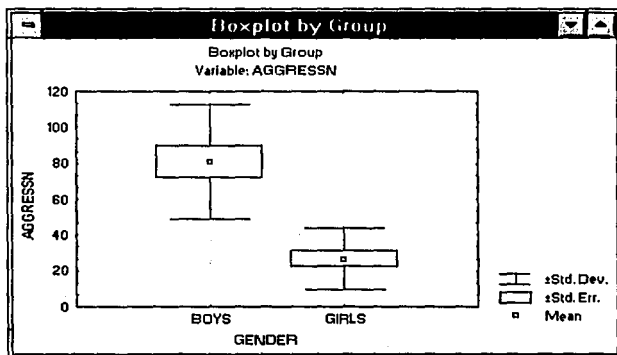
La prueba U de Mann-Whitney despliega para cada variable dependiente:  
 La suma de rangos para cada grupo.  
 El valor de la estadística de prueba U.  
 El valor de Z, de Z ajustada y su respectivo nivel p.  
 El número válido de observaciones para cada grupo.

Kolmogorov-Smirnov Test [aggresso.sta]									
By variable GENDER									
Group 1: 1-BOYS					Group 2: 2-GIRLS				
variable	Max. N	Max. Diff.	Pos. Level	Mean	Mean	Std. Dev.	Std. Dev.	Valid	Valid
				BOYS	GIRLS	BOYS	GIRLS	BOYS	GIRLS
AGGRESSN	0.0	.8	.001	81.	27.	32.	17.	12	

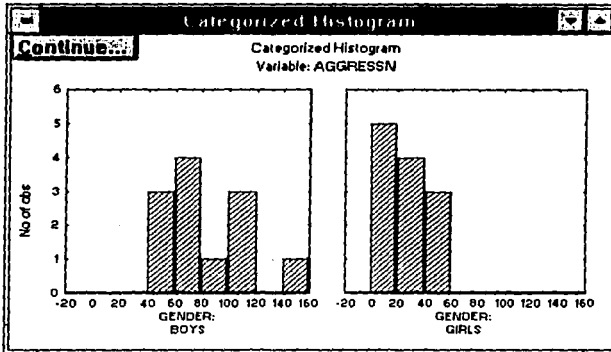
La prueba de Kolmogorov-Smirnov despliega para cada variable dependiente la máxima diferencia positiva y negativa, el nivel p, y para cada grupo: La media, la desviación estándar y el número válido de observaciones.

La diferencia entre niños y niñas en este estudio es altamente significativa (sin tomar en cuenta qué prueba fue utilizada. Para Wald-Wolfowitz  $Z = -3.756$  con  $p = .0001$ , en Mann-Whitney  $Z = -3.810$  con  $p = .0001$  y en Kolmogorov-Smirnov  $p = .001$ ).

La gráfica por omisión para estos scrollsheets es la *Box & whisker*, que marca las medias, desviaciones estándar y errores estándar para la variable dependiente en cada grupo, en esta gráfica se puede apreciar que los niños fueron más agresivos que las niñas.



La otra gráfica disponible en la ventana de introducción de datos (*Categorized histogram*) también muestra que los niños son más agresivos. El eje X mide la agresividad y el eje Y la frecuencia (número de niños en cada intervalo).



## 6.2.6 ANÁLISIS DE VARIANZA DE KRUSKAL-WALLIS, PRUEBA DE LA MEDIANA (KRUSKAL-WALLIS ANOVA, MEDIAN TEST)

### Análisis de varianza de Kruskal-Wallis.

El análisis de varianza de una clasificación por rangos de Kruskal-Wallis se utiliza para decidir si  $k$  muestras (grupos) independientes provienen de poblaciones diferentes. Esta técnica establece la hipótesis nula de que las  $k$  muestras proceden de la misma población o de poblaciones con la misma mediana. La prueba supone que la variable en estudio tiene como base una distribución continua y requiere por lo menos una escala de medición ordinal.

### Prueba de la mediana.

La prueba de la mediana es una versión más simple de la prueba de Kruskal-Wallis ya que realiza los cálculos en términos de una tabla de contingencia. Específicamente, el programa va a contar el número de casos en cada muestra (grupo) que caen por encima o por debajo de la mediana común y calcula el valor de una estadística con distribución  $\chi^2$  para la tabla de contingencia resultante de  $2 \times k$ . Bajo la hipótesis nula: Todas las muestras (no necesariamente de igual tamaño) provienen de poblaciones con medianas idénticas, se espera que aproximadamente el 50% de todos los casos de cada muestra caiga arriba o abajo de la mediana común. Al igual que en Kruskal-Wallis esta prueba requiere que la variable en estudio haya sido medida por lo menos en una escala ordinal.

### Ejemplo:

El conjunto de datos para este ejemplo está contenido en el archivo  
C:\STAT\EXAMPLES\KRUSKAL.STA



Correct	
1	2
COND	PERF
1	6
2	FORM 11
3	FORM 12
4	FORM 20
5	FORM 24
6	FORM 21
7	FORM 18
8	FORM 15
9	FORM 14
10	FORM 10
11	FORM 8
12	FORM 14
13	COLO 31

Los datos se obtuvieron de un conjunto de bebés que fueron asignados aleatoriamente a uno de tres grupos experimentales. A cada bebé se le mostró una serie de pares de objetos de diferente forma, color y tamaño. Su tarea consistía en escoger uno de esos objetos y si era el correcto, recibía un premio. En un grupo, lo que los niños debían detectar para que su elección fuera correcta era la forma, en un segundo grupo, el color y en el tercer grupo, el tamaño. Esta información está contenida en la variable clasificadora *Conditi*n. La variable de interés (dependiente) es el número de aciertos que tuvieron, es decir, el número de veces que detectaron la dimensión que estaba siendo premiada (*Perfmnc*).

#### Especificación del análisis:

Al seleccionar esta opción se desplegará la siguiente ventana que permite especificar la información inicial requerida para llevar a cabo la prueba:

Kruskal Wallis ANOVA and Median Test

**Variables** [OK] [Cancel]

Grouping variable: CONDITN  
Dependent variables: PERF MNC

**Codes:** 3

Batch processing/printing

**Box & whisker**

**Categorized histogram** [SELECT CASES] [Help]

**Variables:** Este botón abre la ventana que permite seleccionar las variables involucradas en la prueba. Específicamente el archivo de datos debe contener una variable clasificadora con al menos dos distintos valores (códigos) que indicarán a que grupo pertenece cada caso del archivo. Así, estos códigos van a determinar las  $k$  ( $2 \leq k \leq 10$ ) muestras (grupos) que se van a comparar (ver botón *Codes*, abajo).





En la ventana de selección de variables, la primera lista permite elegir la variable clasificadora (**grouping variable**). En la segunda lista se eligen las variables en las cuales los grupos van a ser comparados (**dependent variables**). STATISTICA realiza una prueba para cada variable dependiente seleccionada.

Para el ejemplo, seleccionar la variable *Conditn* como la variable independiente o clasificadora y la variable *Perfrmnc* como la variable dependiente.

**Codes:** Este botón permite especificar los valores (códigos) de la variable clasificadora que se van a emplear para determinar las k muestras (grupos) que se van a comparar.

Los valores de la variable *Conditn* son 3 (1-forma, 2-color, 3-tamaño) por lo tanto en la ventana que despliega este botón se eligen todos los códigos.

**Batch processing/printing:** Esta opción es aplicable si la impresora, archivo en disco, y/o ventana text/output están seleccionados actualmente en la ventana *Page/Output Setup* (ver opciones de impresión en el capítulo 3).

Esta opción afecta la forma en que la salida del análisis es impresa. Por omisión, el programa realiza un proceso interactivo (*Batch processing and printing* está desactivada) que permite determinar los scrollsheets que se desean imprimir. Cuando la opción se activa, los resultados se imprimen automáticamente conforme se van generando (en la impresora, archivo en disco o ventana text/output), es decir, si la opción está activa, el programa ejecuta la prueba, imprime los resultados y regresa a la ventana de información inicial.

**Box & whisker:** Este botón despliega gráficas para comparar la distribución de una variable en los diferentes grupos. Pueden graficarse medianas y cuartiles, o medias y algunas medidas de dispersión (ver capítulo 4 para obtener una descripción detallada de las opciones en este tipo de gráficas).

**Categorized Histograms:** Esta opción producirá un histograma de distribución de frecuencias para cada clase o grupo de una variable seleccionada.

Una vez especificada la información requerida por esta ventana, al presionar el botón **OK**, STATISTICA comienza a ejecutar el análisis e inmediatamente después despliega los scrollsheets con los resultados correspondientes:

Kruskal Wallis ANOVA by Ranks [kruskal.sta]			
NONPAR STATS	Kr <b>A</b>	Independent (grouping) variable: CONDITN	
Depend <b>B</b> PERFRMNC	<b>D</b>	-Wallis test: $H(2, N=36) = 13.84438$ $p = .0010$	
<b>FORM</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>COLOR</b>	<b>C</b>	Valid N	Sum of Ranks
<b>SIZE</b>			
	2	12	139.0000
	3	12	200.0000
		12	327.0000



Median Test, Overall Median = 19.50000 [kruskal_walsh]					
NONPAR STATS		Independent (grouping) variable: CONDITN			
Dependent: PERFRMNC		Chi-Square = 8.666667, df = 2, p = .0131			
		FORM	COLOR	SIZE	Total
<=	Median observed		7.00000	2.00000	18.00000
	expected	6.00000	6.00000	6.00000	
	obs - exp	3.00000	1.00000	-4.00000	
>	Median observed	3.00000	5.00000	10.00000	18.00000
	expected	6.00000	6.00000	6.00000	
	obs - exp	-3.00000	-1.00000	4.00000	
	Total observed	12.00000	12.00000	12.00000	36.00000

G

A) La parte superior de cada scrollsheet contiene:

- El nombre de la variable clasificadora (independiente) que se utilizó para dividir el archivo de datos en las k muestras que se van a comparar.
- El valor de la estadística de prueba:  
H (grados de libertad, número de casos) para Kruskal-Wallis,  
Chi-Square para la mediana (df = grados de libertad).
- El correspondiente nivel p.

B) Nombre de la variable dependiente en la que se van a comparar los k grupos.

C) Nombres de las muestras o grupos que se van a comparar.

D) Códigos (valores numéricos) asociados a cada muestra o grupo (definidos mediante el botón *Codes* de la ventana de información inicial).

E) Número de casos válidos en cada muestra.

F) Suma de los rangos en cada muestra.

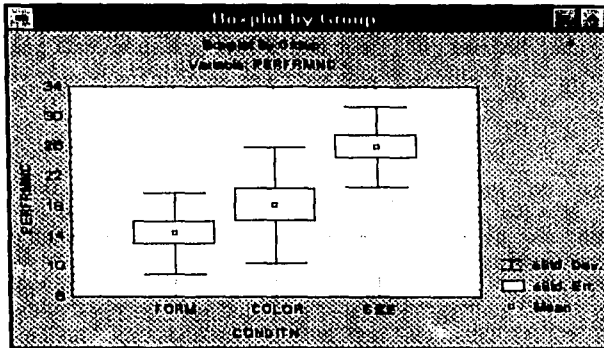
G) Esta tabla contiene, para cada grupo:

- La frecuencia de los valores observados menores o iguales que la mediana común (y después de los mayores).
- El correspondiente valor esperado bajo la hipótesis nula.
- La diferencia entre el valor observado y el esperado.
- El total de valores observados por encima y por debajo de la mediana común.

En cuanto a la prueba de Kruskal-Wallis, el valor de H resultó ser altamente significativo:  $H=13.8443$ ,  $p=0.001$ . Se concluye que el desempeño de los tres grupos experimentales fue diferente. En la última columna, la suma de rangos más grande aparece en la condición *size* (tamaño), es decir, en la condición donde los niños fueron premiados por identificar el tamaño como la dimensión crítica discriminante. Esto se puede apreciar gráficamente con



una gráfica tipo *Box & whisker* con media, desviación estándar y error estándar de la variable *Perfrmnc* en cada grupo.



La media en la condición *size* (tamaño) es mayor que en las otras; por lo tanto, se concluye que los niños identificaron más fácilmente el tamaño cuando éste fue premiado.

La prueba de la mediana también es significativa, sin embargo es menos que la de Kruskal ya que es más "cruda" y menos sensible. De nuevo se observa que el mayor número de sujetos por encima de la mediana común se encuentra en la condición tamaño. Así, esta prueba confirma que el tamaño fue reconocido más fácilmente por los bebés.

### 6.2.7 PRUEBA DE LOS SIGNOS (SIGN TEST)

La prueba de los signos es aplicable a situaciones donde el investigador tiene dos medidas (dos condiciones) para cada sujeto y desea establecer si las dos medidas (o condiciones) son diferentes. Esta prueba debe su nombre al uso de los signos más y menos en la medición en lugar de cantidades. El único supuesto subyacente de la prueba es la continuidad de la variable de interés. No hay ningún supuesto acerca de la forma de la distribución ni pide que todos los sujetos se tomen de la misma población. Las diferentes parejas pueden provenir de poblaciones distintas con respecto a edad, sexo, inteligencia, etc. La prueba simplemente calcula el número de veces (en todos los sujetos) en que el valor de la primera variable (A) es más grande que el de la segunda variable (B). Bajo la hipótesis nula (que establece que no hay diferencia entre las dos variables) se espera que esto suceda aproximadamente el 50 % de las veces. Con base en la distribución binomial, se puede calcular un valor Z para el número de casos donde se observa que  $A > B$ , y calcular la probabilidad asociada a ese valor. En muestras grandes se utiliza la aproximación de la binomial a la normal para evaluar la significancia estadística. Cuando se tienen muestras pequeñas (menos de 20 elementos) es preferible usar las tablas de [Siegel/Castellan, 1988] o [Conover, 1980].

## 6.2.8 PRUEBA DE WILCOXON (WILCOXON MATCHED PAIRS TEST)

La prueba de rangos señalados y pares igualados de Wilcoxon es similar a la de los signos pero más poderosa, ya que considera la magnitud relativa así como la dirección de las diferencias de las parejas, es decir, da mayor peso al par que muestra una diferencia grande entre las dos condiciones que el par que exhibe una diferencia pequeña. Así, el investigador puede hacer el juicio de mayor que entre las ejecuciones de cualquier par, y también entre los puntajes de dos diferencias cualesquiera procedentes de dos pares.

Ejemplo:

Se llevó a cabo un estudio para averiguar si los bebés de 10 a 16 semanas de edad ya son conscientes de la estrecha relación que existe entre la palabra hablada y los correspondientes movimientos de los labios (de quien habla). Con ese objetivo, se colocó a los bebés en un cuarto donde podían observar a una persona leyendo un discurso detrás de una ventana. Primero, las palabras llegaban directamente al cuarto (en sincronía) y después entraron con 400 milisegundos de retardo (fuera de sincronía). La variable dependiente fue el tiempo que el bebé permaneció observando la cara detrás de la ventana.

Los datos para este análisis se encuentran en el archivo :  
C:\STAT\EXAMPLES\SYNCHRON.STA

Data: SYNCHRON.STA		
Case	Sensitivity of ir	
	1 IN_SYNC	2 OUT_SYNC
DC		50.4
NK	17.0	87.0
VH	6.5	25.1
JH	25.0	28.5
SB	5.4	26.9
NH	29.2	36.6
RH	2.9	1.0
DJ	6.6	43.8
JD	15.8	44.2
ZC	8.3	10.4
CW	34.0	29.9
AF	8.0	27.7

Cada caso representa un bebé al cual se le calcula el tiempo que duró observando la cara detrás de la ventana cuando había sincronía con las palabras que escuchaba (*In\_sync*) y cuando no (*Out\_sync*).



### Especificación del análisis:

La prueba de los signos y la de Wilcoxon despliegan la misma ventana para introducir los datos:

**Variables:** Despliega la ventana de selección de variables donde se van a elegir dos listas de éstas. Cada variable de la primera lista será comparada con cada una de las variables de la segunda lista.

En este caso, sólo hay una variable para cada lista, así, en la primera se elige la variable *In\_sync* y en la segunda *Out\_sync*.

**Box & whisker:** Este botón despliega gráficas para comparar la distribución de las variables de interés. Pueden graficarse medianas y cuartiles, o medias y algunas medidas de dispersión (ver capítulo 4 para una descripción detallada de las opciones en este tipo de gráficas).

Los resultados de estas dos pruebas se desplegarán en un scrollsheet con la siguiente información para cada pareja de variables:

Sign Test (synchron.sta)			
Pair of Variables	Non-ties	Z	p-level
IN_SYNC & OUT_SYNC	83.33334	2.020726	.043308

El número de valores no empatados (*Non-ties*) para cada pareja de variables.  
 El porcentaje de veces en que la primera variable es menor que la segunda.  
 El valor de la estadística de prueba Z.  
 El nivel p.

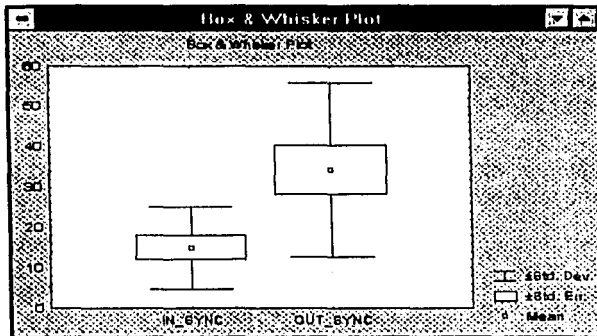
Wilcoxon Matched Pairs Test (synchron.sta)			
Pair of Variables	T	Z	p-level
IN_SYNC & OUT_SYNC	5.000000	2.667179	.007653

Número de casos válidos.  
 El valor de la estadística de prueba T.  
 El valor de la estadística de prueba estandarizada (Z).  
 El nivel p.



El nivel  $p$  para la prueba del signo es significativo con  $\alpha=.05$ , mientras que la prueba de Wilcoxon es significativa al nivel  $\alpha=.01$ . Esto refleja el hecho de que la prueba del signo sólo usa información acerca del signo de la diferencia entre las dos variables mientras que la de Wilcoxon también toma en cuenta la magnitud relativa de esas diferencias. Por lo tanto, de acuerdo a estos resultados, se concluye que aún a edades tempranas los niños son capaces de distinguir entre el discurso sincronizado con el movimiento de los labios y el que no lo está (asíncrono).

Es conveniente crear una gráfica tipo *Box & whisker* con media, desviación estándar y error estándar para las variables *In\_sync* y *Out\_sync*.



Aparentemente, el discurso asíncrono tuvo una atención media mayor, pero al mismo tiempo la variabilidad más grande.

### 6.2.9 ANALISIS DE VARIANZA DE FRIEDMAN Y COEFICIENTE DE CONCORDANCIA DE KENDALL (FRIEDMAN ANOVA & KENDALL CONCORDANCE)

La naturaleza de estas dos pruebas es diferente, sin embargo requieren una entrada de datos similar.

#### Análisis de varianza de Friedman.

El análisis de varianza por rangos de Friedman establece la hipótesis nula de que las diferentes columnas de datos (variables de STATISTICA) contienen muestras extraídas de la misma población, o más específicamente de poblaciones con medianas idénticas. Además, esta prueba supone que las variables bajo consideración fueron medidas al menos en una escala ordinal.



**Coefficiente de concordancia de Kendall.**

El coeficiente de concordancia de Kendall ( $W$ ) proporciona una medida del grado de asociación o relación entre  $K$  variables ( $K$  muestras correlacionadas). Se supone que las variables fueron medidas al menos en una escala ordinal.

Ejemplo:

Veinte madres con hijos sordos asistieron a un seminario ideado para dar adiestramiento introductorio en el tratamiento y manejo de los niños sordos.

Un cuerpo de 13 psicoterapeutas trabajaron con las madres y los niños durante sesiones de dos semanas. Al final del seminario se les pidió a los 13 psicoterapeutas que elaboraran un orden de las 20 madres basado en la probabilidad existente en cada madre de criar a su hijo de tal manera que sufriera inadaptación personal.

El conjunto de datos para este ejemplo está contenido en el archivo :  
 C:\STAT\EXAMPLES\MOTHERS.STA

Data: MOTHERS.STA 20v * 13c						
Case	Ranking of mothers by 13 judges					
	MOTR 1	MOTR 2	MOTR 3	MOTR 4	MOTR 5	MOTR 6
H. Smith	1.000	2.000	3.000	4.000	5.000	6
B. Jones	5.000	1.000	16.000	8.000	9.000	2
T. Hill	3.000	2.000	7.000	5.000	14.000	9
W. Miller	8.000	3.000	10.000	11.000	4.000	2
B. Robertson	2.000	1.000	15.000	8.000	14.000	4
C. Hansson	16.000	17.000	5.000	13.000	15.000	11
S. Mann	12.000	9.000	14.000	6.000	7.000	2
F. Jordan	11.000	2.000	13.000	10.000	7.000	3
B. Frank	9.000	2.000	15.000	6.000	5.000	7

El propósito del análisis es doble:

En primer lugar, se desearía determinar si existen diferencias significativas entre las madres con respecto a su habilidad para criar a sus hijos de acuerdo al juicio emitido por los 13 psicoterapeutas. Esta pregunta puede responderse aplicando el Análisis de varianza por rangos de Friedman.

En segundo lugar, ¿Se puede creer en el dictamen de los terapeutas? en otras palabras, ¿Concuerdan las ordenaciones que estableció cada uno de los terapeutas (están correlacionadas)? De no ser así, no se puede confiar mucho en el juicio de los terapeutas acerca de la capacidad de las madres para criar a sus hijos. Por lo tanto, la hipótesis de que los terapeutas están de acuerdo más allá de lo que se esperaría por casualidad será probada vía el coeficiente de concordancia de Kendall.

**Especificación del análisis:**

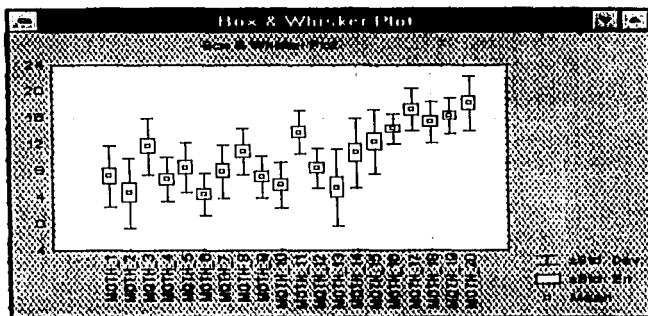
Al seleccionar esta opción se desplegará la siguiente ventana:



**Variables:** Este botón abre la ventana que permite seleccionar las variables que van a ser analizadas, cada variable representa una muestra.

Para este ejemplo, es necesario seleccionar todas las variables del archivo (botón *Select all*)

**Box & whisker:** Este botón despliega gráficas para comparar la distribución de las variables de interés. Pueden graficarse medianas y cuartiles, o medias y algunas medidas de dispersión (ver capítulo 4 para una descripción detallada de las opciones en este tipo de gráficas).



Esta gráfica indica qué madres tienen menos probabilidad de éxito en el cuidado óptimo de sus niños sordos.

Al presionar **OK** se despliega el siguiente scrollsheet con los resultados de la prueba:





Friedman ANOVA and Kendall Coeff. of Concordance (mothers sta)				
NONPAR STATS		ANOVA Chi Sqr. (N = 13, df = 19) = 139.9814 p < .00000		
		Coeff. of Concordance = .56673 Aver. rank r = .53062		
Variable	Rank	Sum of Ranks	Mean	Std. Dev.
MOTH_1		94.5000	7.15385	4.597937
MOTH_2	4.69231	61.0000	4.69231	5.266001
MOTH_3	11.92308	155.0000	11.69231	4.230536
MOTH_4	6.69231	87.0000	6.69231	3.351234
MOTH_5	8.57692	111.5000	8.46154	3.755338
MOTH_6	4.38462	57.0000	4.38462	3.150092
MOTH_7	8.00000	104.0000	7.92308	3.988766
MOTH_8	11.07692	144.0000	10.92308	3.475187
MOTH_9	7.19231	93.5000	7.07692	3.174417
MOTH_10	5.92308	77.0000	5.84615	3.436232
MOTH_11	14.03846	182.5000	13.84615	3.262058
MOTH_12	8.53846	111.0000	8.46154	3.017045

La parte superior contiene:

El valor de la estadística de prueba de Friedman: ANOVA Chi-Sqr (número de casos, grados de libertad) con su respectivo nivel p.

El valor del coeficiente de concordancia de Kendall y el rango promedio (promedio de la p de Spearman).

Rango promedio =  $(K \times \text{coef. de Kendall} - 1) / (k-1)$

Además, para cada variable se tiene:

El rango promedio, suma de rangos, media y desviación estándar.

De acuerdo a estos resultados se puede apreciar que existen diferencias significativas entre las madres ya que el valor de la estadística de prueba (139.9814) es altamente significativo ( $p < .00000$ ) Además, los 13 psicoterapeutas muestran acuerdo en sus juicios ya que el coeficiente de concordancia de Kendall es igual a .56673.

### 6.2.10 PRUEBA Q DE COCHRAN (COCHRAN Q TEST)

La prueba Q de Cochran es una extensión de la prueba de McNemar para la significación de los cambios en las frecuencias o proporciones de k ( $k > 2$ ) muestras relacionadas. Específicamente, la prueba Q de Cochran proporciona un método para examinar si tres o más conjuntos de frecuencias o proporciones relacionadas difieren significativamente entre sí. Esta prueba requiere solamente una escala nominal o en su lugar que los datos hayan sido dicotomizados artificialmente, ya que supone que las variables contienen únicamente dos valores o códigos (0 y 1) de tal manera que si se tuvieran otros valores, STATISTICA los recodificaría a ceros y unos (sin modificar el archivo de datos original).

Ejemplo:

Se desea realizar un estudio para conocer la influencia que un entrevistador amistoso tiene en las respuestas de las amas de casa al hacer una encuesta de opiniones. Se puede adiestrar a un entrevistador para que conduzca tres clases de entrevistas: Entrevista 1, mostrando interés, amistad y entusiasmo; entrevista 2, mostrando formalidad, reserva y cortesía; entrevista 3, mostrando desinterés, precipitación y formalidad austera. Al entrevistador se le pedirá que visite tres grupos de 18 casas, y use la entrevista 1 en un



grupo, la entrevista 2 en otro y la entrevista 3 con el tercero. Así se habrán obtenido 18 conjuntos de amas de casa, con las amas de casa igualadas de 3 en 3 por cada conjunto. Así se tendrán tres muestras relacionadas ( $k=3$ ) con 18 miembros en cada una ( $N=18$ ). Entonces se podrá probar si las diferencias brutas entre los tres estilos de entrevistas influyeron en el número de respuestas "sí" dadas a un ítem particular por los tres grupos igualados.

El conjunto de datos para este ejemplo está contenido en el archivo C:\STAT\EXAMPLES\INTRVIEW.STA

Data INTRVIEW.STA N*11			
Yes and no responses by h			
#	INTRV 1	INTRV 2	INTRV 3
1	NO	NO	NO
2	YES	YES	NO
3	NO	YES	NO
4	NO	NO	NO
5	YES	NO	NO
6	YES	YES	NO
7	YES	YES	NO
8	NO	YES	NO
9	YES	NO	NO
10	NO	NO	NO
11	YES	YES	YES
12	YES	YES	YES
13	YES	YES	NO

### Especificación del análisis:

Al seleccionar esta opción en el panel de inicio del módulo se abrirá la siguiente ventana:

Cochran's Q Test

Variables: none

Code 1 (recoded to 0):

Code 2 (recoded to 1):

**Variables:** Este botón permite seleccionar las variables que se van a analizar. STATISTICA requiere que las variables elegidas contengan únicamente dos valores o códigos enteros ya que la prueba supone que las variables elegidas sólo contienen ceros y unos; si éstas tienen diferentes valores, éstos se pueden especificar en la siguiente opción, (**Code 1 (recoded to 0)**, **Code 2 (recoded to 1)**) así, el programa internamente recodificará esos valores a ceros y unos (sólo temporalmente, el archivo de datos no se modifica).

Para el ejemplo, seleccionar las tres variables que contiene el archivo, éstas ya tienen valores de ceros (No) y unos (Yes).



Al presionar OK se despliega el siguiente scrollsheets con los resultados de la prueba:

Cochran's Q Test (interview sta)			
NONPAR STATS	Number of valid cases: 18		
	Q = 16.66667, df = 2, p < .000241		
Variable	Sum	Percent 0's	Percent 1's
INTRVW_1	12.00000	27.77778	72.22222
INTRVW_2	13.00000	27.77778	72.22222
INTRVW_3	3.00000	83.33334	16.66667

La parte superior contiene el número de casos tomados en cuenta para el análisis, así como el valor de la estadística de prueba Q, los grados de libertad y el nivel de significancia. Además, para cada variable se tiene: La suma de sus valores, el porcentaje de ceros y el porcentaje de unos.

Se puede observar que la prueba Q resultó altamente significativa. La columna Percent 1's (porcentaje de unos) muestra el porcentaje de casas que revelaron información personal. Resulta interesante notar que no hay diferencia en la efectividad de los dos primeros estilos de entrevistas. Así, parece ser que el interés del entrevistador no tiene importancia. Sin embargo, las respuestas negativas se incrementaron cuando el entrevistador actuó mostrando desinterés, precipitación y formalidad austera.

### 6.2.11 ESTADÍSTICAS DESCRIPTIVAS ORDINALES (ORDINAL DESCRIPTIVE STATISTICS)

Esta opción permite calcular varias estadísticas descriptivas ordinales (mediana, percentiles, rango, rango intercuartil) y otras estadísticas descriptivas (media, media armónica, media geométrica, desviación estándar, coeficiente de asimetría, curtosis, varianza, desviación promedio, suma) para las variables seleccionadas.

Al seleccionar esta opción se despliega la siguiente ventana:

Descriptive Statistics	
<input checked="" type="checkbox"/> Variables:	MEASURI-MEASUR5
Compute percentile boundaries:	First: 25.000
	Second: 75.000
<input checked="" type="checkbox"/> Box & whisker	
<input type="button" value="OK"/> <input type="button" value="Cancel"/>	

**Variables:** Despliega la ventana donde se eligen las variables a las que se les van a calcular las estadísticas descriptivas.

**Compute percentile boundaries:** Calcula los percentiles que se le indiquen; uno inferior (**first**) y otro superior (**second**) y los despliega en el *scrollsheet* de resultados. Por omisión STATISTICA calcula los cuartiles, es decir el 25° y 75° percentiles respectivamente.

**Box & whisker:** Este botón permite crear una gráfica *Box & whisker* para las variables seleccionadas; representando la tendencia central y la dispersión de cada una de ellas.

Además de las estadísticas descriptivas más comunes (valores mínimo y máximo, media y número de casos válidos) el *scrollsheet* de resultados despliega las siguientes:

**Median:** La mediana es el valor que “corta” los valores de la variable por la mitad, es decir, el 50% de los casos se encuentra por debajo de la mediana y el 50% por encima de su valor.

**Mode:** La moda es el valor que ocurre con mayor frecuencia, esta frecuencia también es desplegada en el *scrollsheet*, si hay un empate (2 o más valores con la misma frecuencia) éste se indica con una marca o etiqueta.

**Geometric mean:** La media geométrica es el producto de todos los valores o puntajes a la potencia  $1/n$  (uno entre el número de casos válidos). Esta medida es útil en situaciones donde se conoce que la escala de medición no es lineal. STATISTICA calcula la media geométrica vía logaritmo ( $\log$ ):

$$\log(\text{media geométrica}) = \log(\prod X_i)^{1/n} = [\sum (\log(X_i))]/n$$

$X_i$  es el  $i$ -ésimo valor o puntaje.  
 $n$  es el número de casos válidos.

Si una variable contiene valores negativos o cero, la media geométrica no puede ser calculada.

**Harmonic mean:** La media armónica a veces es usada para promediar frecuencias. Se calcula de la siguiente manera:

$$HM = n / \sum(1/X_i)$$

Si una variable contiene el valor cero, la media armónica no puede ser calculada.

**Variance y Standar deviation:** La varianza y desviación estándar son medidas de la variabilidad. STATISTICA calcula la varianza como la suma de las desviaciones cuadradas alrededor de la media dividida entre  $n-1$ . La desviación estándar es la raíz cuadrada de este valor.

**Average deviation:** Es otra medida de variabilidad. Se calcula como la suma de las desviaciones absolutas dividida entre  $n$ .



**Range:** También es un indicador de variabilidad. Se calcula como valor máximo menos valor mínimo.

**Quartile range:** El rango intercuartil es la diferencia de valores entre el 75° y 25° percentiles. Es el rango alrededor de la mediana que incluye el 50% de los casos.

**Skewness:** Calcula el coeficiente de asimetría que indica qué tan sesgada hacia la izquierda (valor negativo) o a la derecha (valor positivo) se encuentra la distribución de probabilidad de la variable en relación con la distribución normal estándar. Si la distribución es simétrica, entonces esta medida es igual a cero.

$$\text{Skewness} = 1/n * [\Sigma\{((X_i - \text{media})/\sigma)^3\}]$$

$\sigma$  es la desviación estándar.

**Kurtosis:** La kurtosis es una medida de qué tan puntiaguda es la distribución de probabilidad de la variable. Si ésta sigue la distribución normal estándar, su kurtosis es cero.

$$\text{Kurtosis} = 1/n * [\Sigma\{((X_i - \text{media})/\sigma)^4\}] - 3$$

$\sigma$  es la desviación estándar.

**BIBLIOGRAFIA**

- [Blalock, 1981] Blalock Hubert M.  
SOCIAL STATISTICS.  
Second Edition.  
McGraw-Hill 1981.
- [Conover, 1980] Conover W. J.  
PRACTICAL NONPARAMETRIC STATISTICS.  
Second Edition.  
John Wiley & sons, N.Y. 1980.
- [Draper/Smith, 1981] Draper N.R. and Smith H.  
APPLIED REGRESSION ANALYSIS.  
Second Edition.  
John Wiley & sons, N.Y. 1981.
- [Montgomery/Peck, 1992] Montgomery D.C. and Peck E.A.  
INTRODUCTION TO LINEAR REGRESSION ANALYSIS.  
Second Edition.  
John Wiley & sons, N.Y. 1992.
- [Noether, 1990] Noether G.E.  
INTRODUCTION TO STATISTICS, THE NON  
PARAMETRIC WAY.  
Springer-Verlag 1990.
- [Siegel/Castellan, 1988] Siegel Sidney and Castellan N. John.  
NONPARAMETRIC STATISTICS FOR THE  
BEHAVIORAL SCIENCES.  
Second Edition.  
McGraw-Hill 1988.
- [StatSoft, Inc. 1991] CSS: STATISTICA.  
Volume I (Quick CSS).  
StatSoft, Inc. 1991.



## TECLAS DEL METODO ABREVIADO

Open Data	Ctrl + F12
Open Graph	Shift + F3
Open Scrollsheet	Ctrl + F11
Save	Shift + F12
Save As	F12
Make Scrollsheet	F11
Print	F4
Page/Output Setup	Shift+F4
Exit	Alt+F4
Edit Cell	F2
Column Specs	Ctrl+F2
Row Name	Alt + F2
Cut	Ctrl + X o Shift + Del
Copy	Ctrl + C o Ctrl + Ins
Paste	Ctrl + V o Shift + Ins
Clear	Del
Screen Catcher	Alt+F3
Font dialog	F9
Customize Colors	Shift+F9
Weighting Variable	F7
Case Selection Conditions	F8
Record macro	Ctrl + F3
Run macro	Ctrl + letter
Cascade	Shift + F6
Tile Horizontally	Alt + F6
Tile Vertically	Alt + Shift + F6
Close All	Ctrl + L
Help	F1
Close	Ctrl + F4
Switch to	Ctrl + Esc



## INDICE DE ILUSTRACIONES

### CAPITULO 1

Grupo de programas .....	2
STATISTICA Module Switcher .....	2
Pantalla de inicio .....	3
Ejemplo de spreadsheet.....	5
Open Data File .....	7

### CAPITULO 2

Representación de un archivo creado....	9
Elementos del spreadsheet.....	10
Ventanas de selección de variables:	
• Variables .....	12
• Select the Variables for the Analysis..	12
• Select one or two Variable List .....	13
Define Weight .....	13
Case Selection Conditions .....	15
Data Management .....	17
Create New File .....	18
Import Data From:.....	19
Export Data To:.....	22
Merge Files .....	25
• Merge Variables.....	26
Create a Subset.....	28
Sort Options.....	29
Modify Variables .....	30
• Add Variables .....	30
• Move Variables .....	31
• Current Specs .....	31
• All Specs.....	33
• Text Values Manager .....	34
• Data Operations.....	35
• Recalculate.....	37
• Shift Variable.....	37
• Rank Order Values .....	38
• Recode Values of Variable.....	39
Modify Cases.....	41
• Add Cases.....	41

• Move Cases .....	42
• Case Name Manager .....	42
Verify .....	43
Verify data.....	44
Standardization of Values.....	47

### CAPITULO 3

Print Data .....	51
Page/Output Setup.....	53
Elementos del scrollsheet.....	61
STATISTICA Defaults: General.....	64
Print Scrollsheet.....	70

### CAPITULO 4

Custom 2D Graphs.....	78
Custom 3D Sequence Graphs.....	79
Custom 3D Scatterplots.....	80
Custom Matrix/Icon Graphs.....	82
Elementos de una gráfica 3D.....	88
Ejemplo de una gráfica de iconos .....	88
General Layout .....	89
Escala logarítmica .....	90
Escala log-decimal.....	90
Efecto de la opción Layered.....	91
Efecto de la opción Data Labels .....	91
Efecto de la opción Number of Cuts ...	92
Efecto de la opción Axes .....	92
Efecto de las opciones: Raw Data Spikes y Surface-Lines (Show Hidden) .....	93
Graphic Text Editor .....	98
Efecto de ajustar márgenes .....	103
Perspective and Rotation .....	103
Ejemplo de la asociación directa entre Graph Data y su gráfica respectiva ...	106





## CAPITULO 5

Multiple Regression .....	111
Representación de un archivo con formato de matriz de correlación .....	112
Review Descriptive Statistics .....	115
Elementos de la gráfica Box & Whiskers .....	116
Gráfica de correlaciones .....	116
Model Definition .....	117
Multiple Regression Results .....	120
Residual Analysis .....	124
<b>Ejemplo 1</b>	
• Tabla de datos .....	128
• Spreadsheet de los datos .....	129
• Review Descriptive Statistics .....	130
• Correlations .....	130
• Scatterplot X vs. Y .....	130
• Multiple Regression Results .....	131
• Analysis of Variance .....	131
• Gráfica Predicted vs. Residual Scores .....	132
• Multiple Regression Results .....	133
• Analysis of Variance .....	133
• Gráfica Predicted vs. Residual Scores .....	133
• Normal Probability plot of Residuals .....	134
<b>Ejemplo 2</b>	
• Spreadsheet de los datos .....	135
• Multiple Regression Results .....	136
• Regression Summary for Dependent Variable .....	136
• Stepwise Multiple Regression .....	137
• Summary of Stepwise Regression .....	137
• Regression Summary for Dependent Variable .....	137

## CAPITULO 6

Non Parametric Statistics & Distribution Fitting .....	141
Fitting Continuous Distribution .....	142
• Resultados .....	143

Interpretación general de la tabla de 2x2 .....	144
Interpretación de la tabla de 2x2 para la prueba $\chi^2$ .....	144
Interpretación de la tabla de 2x2 para la prueba de McNemar .....	144
2x2 Tables .....	145
• Resultados .....	145
Spreadsheet: Accident.sta .....	147
Observed vs. Expected Frequency .....	147
• Resultados .....	148
Spreadsheet: Striving.sta .....	149
Nonparametric Correlations .....	149
• Resultados .....	150
Gráfica de correlaciones de Striving.sta .....	151
Spreadsheet: Aggressn.sta .....	152
Wald-Wolfowitz Runs Test .....	152
• Resultados .....	153
Resultados de Mann-Whitney U Test .....	153
Resultados de Kolmogorov-Smirnov .....	154
Boxplot de Aggressn.sta .....	154
Histograma por categorías de Aggressn.sta .....	155
Spreadsheet: Kruskal.sta .....	156
Kruskal-Wallis ANOVA and Median Test .....	156
• Resultados .....	157,158
Boxplot de Kruskal.sta .....	159
Spreadsheet: Synchron.sta .....	160
Sign Test .....	161
• Resultados .....	161
Resultados de Wilcoxon Matched Pairs Test .....	161
Box & Whisker de Synchron.sta .....	162
Spreadsheet: Mothers.sta .....	163
Friedman ANOVA by ranks .....	164
Box & Whisker de Mothers.sta .....	164
Resultados de Friedman ANOVA .....	165
Spreadsheet: Intrview.sta .....	166
Cochran's Q test .....	166
• Resultados .....	167
Descriptive Statistics .....	167