

Esta obra esta bajo una licencia Reconocimiento-No comercial 2.5 Colombia de Creative Commons. Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by/2.5/co/> o envíe una carta a Creative Commons, 171 Second Street, Suite 300, San Francisco, California 94105, USA

STATGRAPHICS



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Autores:

**ANGELA MARIA ARAUJO
MARIA PAULA GANDUR R.**

Director Unidad Informática: Henry Martínez Sarmiento

Tutor Investigación: Brayan Ricardo Rojas

Coordinadores: Leydi Diana Rincón R.
Brayan Ricardo Rojas

Coordinador Servicios Web: Miguel Ibáñez

**Analista de Infraestructura
y Comunicaciones:** Alejandro Bolívar

**Analista de Sistemas de
Información:** Mesías Anacona O

**UNIVERSIDAD NACIONAL COLOMBIA
FACULTAD DE CIENCIAS ECONÓMICAS
UNIDAD DE INFORMÁTICA Y COMUNICACIONES
BOGOTÁ D.C.**

JUNIO 2008

STATGRAPHICS

Director Unidad Informática: Henry Martínez Sarmiento

Tutor Investigación: Brayan Rojas

Auxiliares de Investigación:

Álvaro Esneyder Roncancio	Jersson Arnulfo Guerrero Nova
Amanda Patricia Ruiz Ortiz	Jorge Alexander Ceron Sanchez
Andres Almed Yate Bejarano	Jorge Mario Cortes Cortes
Andres Mauricio Arias Toro	Jose Fernando Moreno Gutiérrez
Ángela María Araujo	Jose Luis Garcia Zapata
Benjamin Eduardo Venegas	Juan Carlos Peña Robayo
Cristian Camilo Ibañez Aldana	Juan Felipe Reyes Rodríguez
Daniel Hernán Santiago Romero	Luis Alejandro Pico Silva
David Alberto Montaña Fetecua	María Paula Gandur Ronseria
Elkin Giovanni Calderón	Rodrigo Acosta Sarmiento
Erika Zuley Guerrero Cortes	Sergio Fernando Garzón Rincón
Javier Eliécer Pirateque Niño	Zareth Manzon Garnica

Este trabajo es resultado del esfuerzo de todo el equipo perteneciente a la Unidad de Informática.

Se prohíbe la reproducción parcial o total de este documento, por cualquier tipo de método fotomecánico y/o electrónico, sin previa autorización de la Universidad Nacional de Colombia.

**UNIVERSIDAD NACIONAL COLOMBIA
FACULTAD DE CIENCIAS ECONÓMICAS
UNIDAD DE INFORMÁTICA Y COMUNICACIONES
BOGOTÁ D.C.
JUNIO 2008**

OBJETIVOS

1. GENERAL: Aproximarse de manera hábil y eficiente a las diferentes oportunidades de manejo estadístico que ofrece el programa. Con ello se pretende estudiar las particularidades, ventajas y herramientas que ofrece Statgraphics en el marco de las aplicaciones estadísticas.
2. PARTICULARES:
 1. Describir las diferentes aplicaciones que tiene el programa.
 2. Ejecutar las diferentes aplicaciones que ofrece el programa.
 3. Comparar las herramientas que ofrece el programa respecto de otros que ofrecen un conjunto similar de utilidades. Con ello, se pretenden establecer ventajas y desventajas de Statgraphics respecto de los demás productos presentes en el mercado.
 4. Diseñar una breve guía practica para el manejo del programa.
 5. Plantear los lineamientos a seguir para el respectivo curso libre del programa.

JUSTIFICACION

El desarrollo de la investigación en Statgraphics se justifica de acuerdo a los siguientes parámetros:

- Implementación de un programa de tipo estadístico en la unidad de informática de la facultad de ciencias económicas. Esta aplicación es una herramienta cotidiana en el mundo económico por lo cual es necesario que la unidad este a la orden del día con los requerimientos del estudiantado en materia de programas.
- El desarrollo de la investigación, promoverá una amplia gamma para el dominio de la unidad con ello se alimentara de aun mas herramientas logrando con ello la ejecución máxima de sus capacidades y además promoverá sus objetivos particulares.
- Extensión de conocimientos en programas de manejo estadístico y econométrico.

ALCANCES

La investigación plantea como objetivo central permear en el futuro la cotidianidad de la vida académica en la Unidad de Informática, con ello inicialmente se promoverán capacitaciones internas que de lograrse una ejecución precisa es posible sea alternada con el desarrollo de un curso libre. En general, la comunidad académica miembro de la facultad de ciencias económicas se vera alcanzada por esta investigación en la medida en que encuentren en Statgraphics una herramienta de su agrado.

CRONOGRAMA

MES	ACTIVIDAD
FEBRERO	Aproximación teórica al programa. Definición funciones principales y de interés
MARZO	Funciones principales (parte 1) Desarrollo teórico-practico de las funciones principales. Presentación PP de la metodología y aplicación de: Forma de acceder al programa – Introducción al programa – Introducción a las funciones principales.
ABRIL	Funciones principales (parte 2 Desarrollo teórico-practico de las funciones principales. Presentación PP de la metodología y aplicación de: Continuación introducción a las funciones principales. Desarrollo practico de las funciones
MAYO	Preparación curso libre y guía de manejo.
JUNIO	Entrega final

RESUMEN/ABSTRACT

El presente texto se construye como un manual de guía al usuario que esta desarrollando conocimientos primarios al programa Statgraphics. Se propone una metodología sencilla para acercar al usuario a la herramienta, de una manera sencilla y con imagines que le permitirán el fácil manejo del programa.

En este manual se promueve especialmente las aplicaciones de interés para la Facultad de Ciencias Económicas, es decir se hace especial énfasis en los análisis estadísticos y econométricos que ofrece el programa.

The following text is made as a manual user guide. It's development primary knowledge about Statgraphics by a simple and organized method. Also uses some images of the program to allow the reader the easy comprehension about the way to work of the software.

The manual presents special attention to applications of particular interests to Facultad de Ciencias Economicas like statistics and econometrics tools.

TABLA CONTENIDO

INDICE GRAFICOS	Pág. 6
INTRODUCCION	Pag. 8
1. INTRODUCCIÓN AL PROGRAMA	Pág. 9
1.1 Que es StatGraphics?	Pág. 9
1.2 ¿Cuáles son las formas de adquirir el programa?	Pág. 9
1.3 ¿Como se accede al programa?	Pág. 11
1.4 Entrando al programa: StatWizard	Pág. 11
1.4.1 Descripción del Cuadro de Dialogo StatWizard	Pág. 12
1.4.2 Selección Análisis Datos	Pág. 14
1.5 Creando reportes de trabajo: StatReporte	Pág. 17
1.5.1 Pegar Resultados	Pág. 18
1.6 Creando presentaciones de graficas: StatGallery	Pág. 18
1.6.1 Configuración	Pág. 18
1.6.2 Sobreposición de Graficas	Pág. 20
1.6.3 Copiar Gráficos	Pág. 20
1.7 Modificación de Gráficos	Pág. 20
1.7.1 Detalles Gráficos	Pág. 20
2. MANEJO DE DATOS	Pág. 23
2.1 El libro de datos	Pág. 23
2.2 Introducción de datos	Pág. 24
2.3 Manipulación de datos	Pág. 28
2.3.1 Copiar y pegar datos	Pág. 28
2.3.2 Crear Nuevas Variables basadas en columnas existentes	Pág. 28
2.3.3 Transformación de datos	Pág. 28
2.4 Organización de datos	Pág. 31
2.5 Recodificación de datos	Pág. 31
2.6 Combinación de columnas	Pág. 32
2.7 Generación de datos	Pág. 33
3. GRAFICOS	Pág. 35
3.1 Modificación de gráficas	Pág. 35
3.1.1 Opciones de Diseño	Pág. 35
3.2 Separar puntos en gráfico de dispersión	Pág. 36
3.3 Resaltar puntos en gráfico de dispersión	Pág. 36
3.4 Identificación de puntos	Pág. 37
3.5 Copiar gráficas a otras aplicaciones	Pág. 38
3.6 Guardar gráficas como archivos de imagen	Pág. 39
4. ANALISIS DE DATOS: REGRESION SIMPLE	Pág. 40
4.1 Acceso a la regresión simple	Pág. 40
4.2 Resumen del análisis	Pág. 42
4.2.1 Coeficientes	Pág. 42
4.2.2 Análisis de Varianza	Pág. 42
4.3 Opciones de análisis	Pág. 43
4.4 Grafica del análisis	Pág. 45
5. REGRESIÓN MÚLTIPLE	Pág. 49
5.1 Resumen del análisis	Pág. 52
5.1.1 Variables	Pág. 52
5.1.2 Coeficientes	Pág. 52
5.1.3 Análisis de varianza	Pág. 52
5.1.4 Estadísticos	Pág. 52
5.2 Opción de análisis	Pág. 53

5.2.1 Ajuste	Pág. 54
5.2.2 Constante en Modelo	Pág. 54
5.2.3 Transformación de Box-Cox	Pág. 54
5.2.4 Transformación de Cochrane-Orcutt	Pág. 54
5.3 Grafica del análisis	Pág. 54
5.4 Opciones de ventana	Pág. 55
5.5 Opciones de análisis	Pág. 55
6. REGRESION NO LINEAL	Pág. 57
6.1 Regresión no lineal con una variable	Pág. 57
6.2 Resumen de análisis	Pág. 60
6.2.1 Resumen de los Datos	Pág. 60
6.2.2 Función a Estimar	Pág. 60
6.2.3 Estadísticas de la Estimación	Pág. 60
6.2.4 Estimaciones de los Parámetros	Pág. 60
6.2.5 Análisis de Varianza	Pág. 61
6.2.6 Estadísticas	Pág. 61
6.3 Grafica del modelo ajustado	Pág. 61
6.4 Regresión no lineal con dos o más variables	Pág. 62
6.4.1 Opciones de Ventana	Pág. 64
6.4.2 Opciones de Análisis	Pág. 65
7. ANALISIS DE DATOS: ANALISIS ESTADISTICO	Pág. 67
7.1 Resumen estadístico y grafico de caja de bigotes	Pág. 67
7.1.1 Modificar los resultados	Pág. 70
7.1.2 Gráfico de caja y bigotes	Pág. 72
7.1.3 Tratamiento de datos atípicos	Pág. 74
7.1.4 Prueba de hipótesis	Pág. 75
7.1.5 Histograma	Pág. 77
8. COMPARACIÓN DE DOS MUESTRAS	Pág. 79
8.1 Procedimiento ejecución comparación dos muestras	Pág. 79
8.1.1 Resumen estadístico	Pág. 81
8.1.2 Histograma dual	Pág. 81
8.1.3 Gráfico de caja y bigotes	Pág. 82
8.1.4 Comparación desviaciones estándar	Pág. 82
8.1.5 Comparación de medias	Pág. 84
8.1.6 Comparación de medianas	Pág. 85
8.1.7 Gráfico de cuantiles	Pág. 86
8.1.8 Gráfico Cuantil-Cuantil	Pág. 87
8.1.9 Prueba Kolmorov-Smirnov	Pág. 88
9. COMPARACIÓN DE MÁS DE DOS MUESTRAS	Pág. 90
9.1 Procedimiento ejecución comparación de más de dos muestras	Pág. 90
9.1.1 Gráfico dispersión	Pág. 92
9.1.2 Análisis de varianza	Pág. 93
9.1.3 Comparación Medias	Pág. 94
9.1.4 Comparación Medianas	Pág. 96
9.1.5 Comparación de desviaciones estándar	Pág. 97
9.1.6 Gráfico de residuos	Pág. 98
9.1.7 Gráfico Análisis de Medias (ANOM)	Pág. 99
CONCLUSIONES	Pág. 101
BIBLIOGRAFIA	Pág. 102
ANEXOS	Pág. 103

INDICE GRAFICOS

NUMERO GRAFICO	DESCRIPCION
Gráfico 1	Icono escritorio
Gráfico 2	Ventana StatWizard
Gráfico 3	StatWizard Localización datos
Gráfico 4	StatWizard Definición Estudios
Gráfico 5	StatWizard Análisis que no requieren datos
Gráfico 6	Statwizard Selección análisis
Gráfico 7	Statwizard Selección datos
Gráfico 8	Statwizard Selección análisis/datos
Gráfico 9	Cuadro dialogo
Gráfico 10	StatWizard Selección Análisis/nombre
Gráfico 11	StatWizard Selección SnapStat
Gráfico 12	StatWizard Selección Rápida
Gráfico 13	StatReporter
Gráfico 14	StatGallery
Gráfico 15	Pestañas Navegación StatGallery
Gráfico 16	Opciones StatGallery
Gráfico 17	Cuadro Dialogo opciones StatGallery
Gráfico 18	Añadir Objeto
Gráfico 19	Libro Datos StatGraphics
Gráfico 20	Cuadro dialogo modificación columna
Gráfico 21	Cuadro dialogo origen datos
Gráfico 22	Selección opción pegar
Gráfico 23	Libro de datos StatGraphics
Gráfico 24	Ventana Abrir archivo datos
Gráfico 25	Selección Extensión archivo
Gráfico 26	Evaluador expresiones
Gráfico 27	Ubicación variable dentro de evaluador de expresiones
Gráfico 28	Cuadro dialogo ordenar datos
Gráfico 29	Cuadro dialogo recodificación datos
Gráfico 30	Cuadro dialogo estadísticas por fila
Gráfico 31	Resultados estadísticas por fila
Gráfico 32	Cuadro dialogo generación datos
Gráfico 33	Cuadro dialogo opciones gráficas
Gráfico 34	Cuadro dialogo separación
Gráfico 35	Cuadro dialogo resaltar
Gráfico 36	Identificación de puntos
Gráfico 37	Identificación de puntos
Gráfico 38	Etiqueta de punto
Gráfico 39	Cuadro dialogo guardar grafico
Gráfico 40	Menú relacionar/regresión simple
Gráfico 41	Cuadro dialogo regresión simple
Gráfico 42	Análisis regresión simple
Gráfico 43	Submenú resumen análisis
Gráfico 44	Opciones regresión múltiple
Gráfico 45	Gráfico de análisis
Gráfico 46	Cuadro dialogo opciones
Gráfico 47	Cuadro dialogo opciones de ventana
Gráfico 48	Cuadro dialogo opciones gráficas
Gráfico 49	Opciones gráfico de residuo
Gráfico 50	Menú relacionar/regresión múltiple
Gráfico 51	Cuadro dialogo regresión múltiple
Gráfico 52	Análisis regresión múltiple
Gráfico 53	Submenú opciones de análisis
Gráfico 54	Cuadro dialogo opciones regresión múltiple
Gráfico 55	Opciones de grafica
Gráfico 56	Opciones grafico efectos de componentes
Gráfico 57	Opciones regresión múltiple
Gráfico 58	Abriendo datos
Gráfico 59	Menú relacionar/regresión no lineal
Gráfico 60	Menú relacionar/regresión no lineal
Gráfico 61	Cuadro dialogo regresión no lineal
Gráfico 62	Cuadro dialogo establecimiento parámetros
Gráfico 63	Análisis regresión no lineal
Gráfico 64	Opciones grafico modelo ajustado

Gráfico 65	Cuadro dialogo regresión no lineal
Gráfico 66	Cuadro dialogo parámetros Gráfico
67	Gráfico regresión no lineal
Gráfico 68	Opciones regresión no lineal
Gráfico 69	Opciones grafico respuesta Gráfico
70	Cuadro dialogo regresión no lineal
Gráfico 71	Abriendo datos
Gráfico 72	Menú describir/análisis de una variable
Gráfico 73	Cuadro dialogo regresión no lineal
Gráfico 74	Resultados análisis una variable Gráfico
75	Resumen estadístico de variable
Gráfico 76	Opciones
Gráfico 77	Cuadro dialogo opciones resumen estadístico
Gráfico 78	Opciones grafico
Gráfico 79	Cuadro dialogo opciones grafico caja y bigotes
Gráfico 80	Menú describir identificación valores atípicos
Gráfico 81	Identificación valores atípicos
Gráfico 82	Menu describir/pruebas hipótesis
Gráfico 83	Cuadro dialogo pruebas hipótesis
Gráfico 84	Análisis pruebas hipótesis
Gráfico 85	Selección gráficas
Gráfico 86	Histograma
Gráfico 87	Menu comparar/muestras independientes
Gráfico 88	Cuadro dialogo comparación muestras independientes
Gráfico 89	Análisis comparación dos muestras
Gráfico 90	Resumen estadístico comparación dos muestras
Gráfico 91	Histograma dual Gráfico
92	Grafico caja y bigotes
Gráfico 93	Acceso selección tablas
Gráfico 94	Análisis comparación desviaciones estándar
Gráfico 95	Acceso selección tablas Gráfico
96	Análisis comparación medias
Gráfico 97	Acceso selección tablas Gráfico
98	Análisis comparación medianas
Gráfico 99	Acceso selección graficas
Gráfico 100	Gráfico cuantiles
Gráfico 101	Acceso selección tablas
Gráfico 102	Grafico cuantil-cuantil
Gráfico 103	Acceso selección tablas
Gráfico 104	Acceso selección tablas
Gráfico 105	Menú comparar/comparación varias muestras
Gráfico 106	Cuadro dialogo comparación varias muestras
Gráfico 107	Introducción variables compracion múltiples muestras
Gráfico 108	Análisis comparación varias muestras
Gráfico 109	Acceso selección tablas
Gráfico 110	Tabla anova/comparación varias muestras
Gráfico 111	Grafico anova/comparación varias muestras
Gráfico 112	Acceso selección graficas
Gráfico 113	Grafico medias/ comparación varias muestras
Gráfico 114	Pruebas multiples rangos
Gráfico 115	Resultado prueba Kruskal-Wallis
Gráfico 116	Grafico caja y bigotes
Gráfico 117	Verificación varianza/prueba Levene's
Gráfico 118	Opciones verificación varianza Gráfico
119	Grafico residuos
Gráfico 120	Grafico ANOM/ comparación varias muestras

INTRODUCCION

A continuación usted encontrara un manual para el usuario de STATGRAPHICS CENTURION V. Este viene desarrollado dada la siguiente estructura, en los primeros capítulos usted encontrara generalidades del programa, tales y como lo son El entorno del mismo, la forma de acceso, la vía para introducir los datos y el tratamiento de las graficas. Posteriormente y conforme avanza el mismo, se presentan análisis mas específicos correspondientes al área de la economía, como lo son la introducción estadística y econométrico, y el respectivo tratamiento de los datos dado el énfasis de las mismas.

Es casi evidente, el hecho que se requiere STATGRAPHICS CENTURION V, sin embargo en caso de no tener el programa con licenciamiento el manual también incluye una serie de páginas de internet a través de las cuales usted puede descargar una prueba del mismo.

En adición a lo anterior una vez inicie el estudio de análisis específicos el presente texto guía es complementado con una presentación en Macromedia lo que permitirá al lector crear una visión general del aplicativo econométrico y matemático.

El texto esta desarrollado para ser de gran utilidad a cualquier tipo de usuario, en especial a aquellos dedicados a áreas como la economía, la administración y la econometría.

I. INTRODUCCION AL PROGRAMA

- ¿Qué es STATGRAPHICS?
- ¿Cuáles son las formas para acceder al programa?
- ¿Cómo se accede al programa?
- ¿Cuáles son los componentes del entorno grafico del programa?
- ¿Cómo se introducen los datos en el programa?
 1. ¿Qué clases de datos permite introducir el programa?
 2. ¿Cuál es la extensión de cada uno de ellos?
- ¿Cómo guardar correctamente los datos?

I.1 QUE ES STATGRAPHICS

Software que está diseñado para facilitar el análisis estadístico de datos. una análisis descriptivo de una o varias variables, utilizando gráficos que expliquen su distribución o calculando sus medidas características.

El cálculo de intervalos de confianza, contrastes de hipótesis, análisis de regresión, análisis multivariantes, así como diversas técnicas aplicadas en Control de Calidad, son algunas de las herramientas que tiene el programa.

I.2 ¿CUÁLES SON LAS FORMAS DE ADQUIRIR EL PROGRAMA?

Para tener acceso al programa es importante destacar dos vías:

1. Vía Internet.

Algunas páginas ofrecen el programa por un tiempo corto como formato de licencia. Algunas a destacar son:

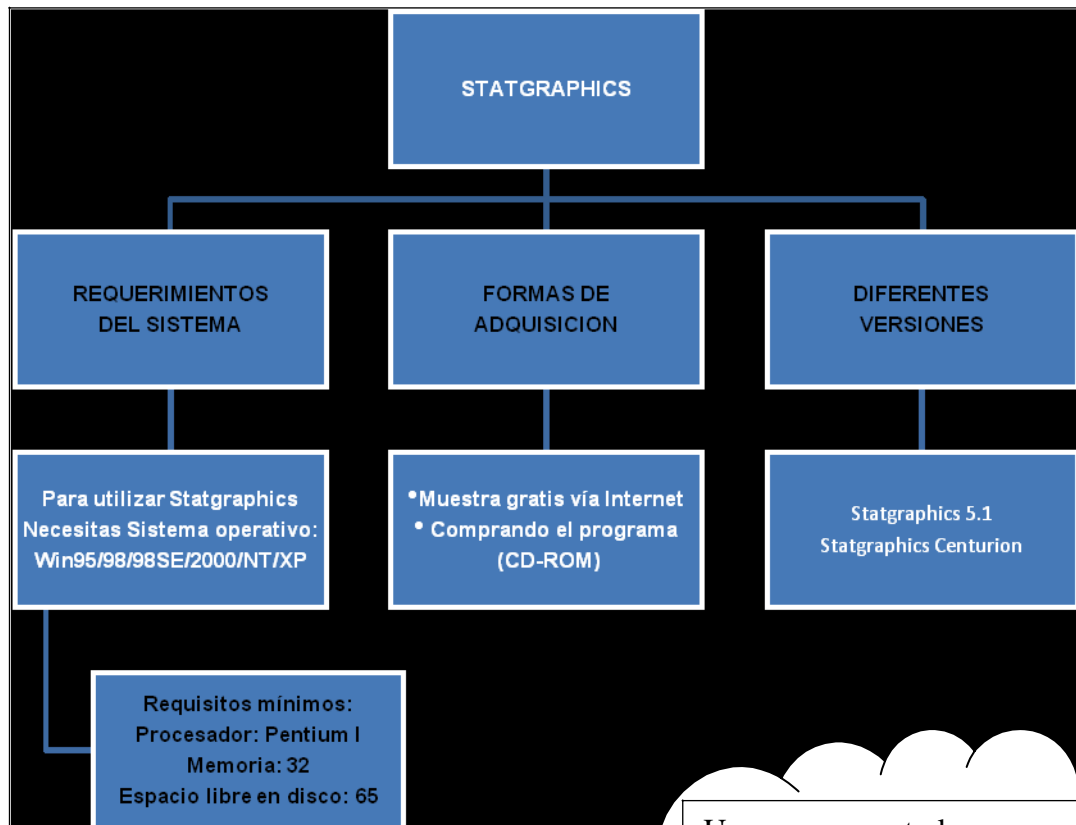
<http://statgraphics.softonic.com/>

http://www.software-shop.com/in.php?mod=ver_producto&prdID=40

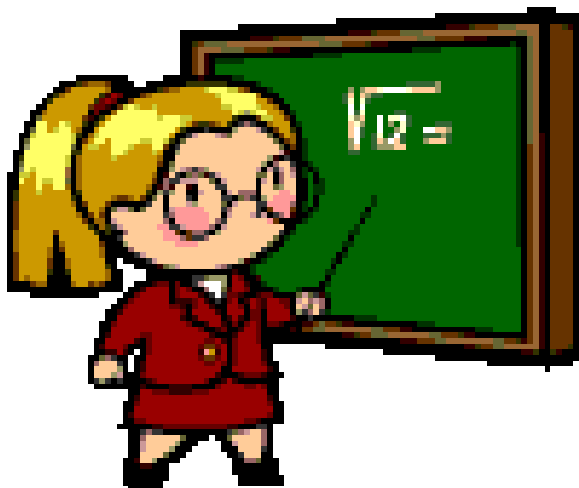
2. Adquiriendo el programa:

Al ser un programa de uso privado, para su uso permanente es necesario adquirir una licencia por ejemplo en Colombia algunas tiendas especializadas son:

Software-Shop



Una vez expuesto las generalidades del programa es necesario pasar a la interacción con el computador.



1.3 ¿COMO SE ACCEDE AL PROGRAMA?

En el escritorio de Windows usted encontrara el acceso directo para el programa identificado a través del siguiente icono:



GRAFICO 1. ICONO ESCRITORIO STATGRAPHICS

1.4 ENTRANDO AL PROGRAMA: STATWIZARD

Al iniciar el programa lo primero que encontrará el usuario es la ventana StatWizard; esta aplicación hace las veces de asistente dentro de cualquier proceso que se desee desarrollar dentro de STATGRAPHICS

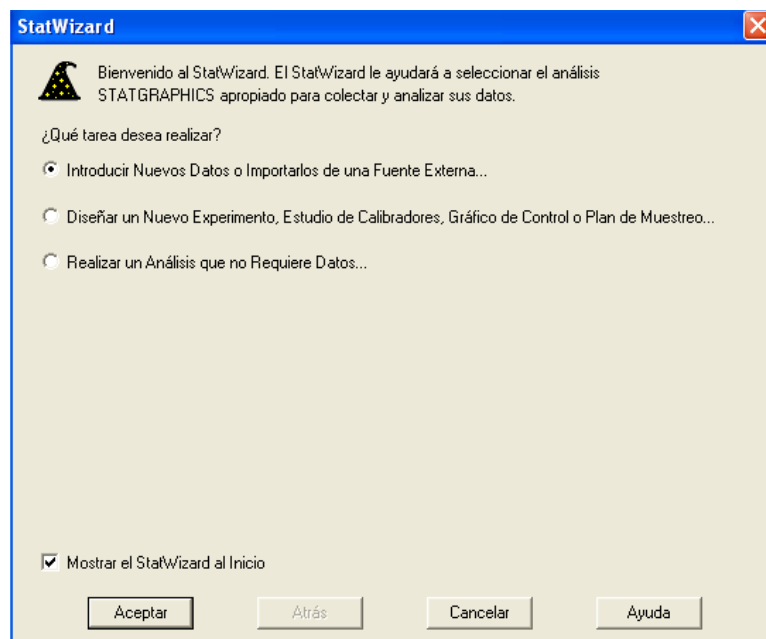


GRAFICO 2. VENTANA STATWIZARD

Esta herramienta puede ser abierta desde la barra de herramientas mediante el icono



1.4.1 Descripción del Cuadro de Dialogo StatWizard

Cuando el StatWizard es ejecutado éste nos ofrece tres opciones de trabajo:

1) Introducción o Importación de Nuevos Datos:

Al seleccionar la primera opción disponible, el asistente nos guía a través de una serie de cuadros de dialogo con el fin de establecer los parámetros del nuevo proyecto; esto es, al dar clic en aceptar luego de la primera ventana del StatWizard se despliega el siguiente cuadro de dialogo

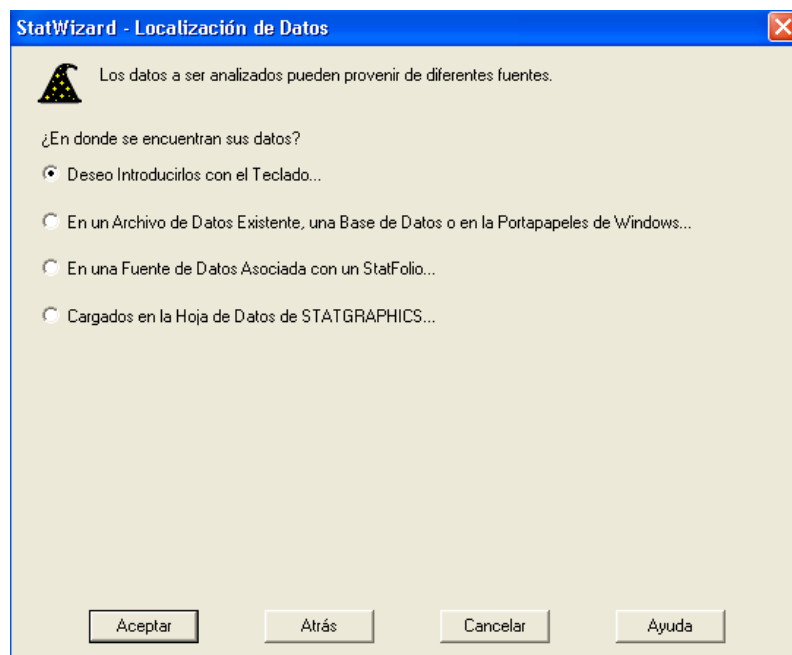


GRAFICO 3. STATWIZARD LOCALIZACION DATOS

Donde nos pregunta cual es la ubicación de los datos a trabajar; en este campo y según las diferentes necesidades del usuario se seleccionará la opción correspondiente.

En resumen las opciones existentes pueden caracterizarse como Introducción manual de datos, importación o trabajo de archivos existentes dentro de StatGraphics.

2) Diseño de experimentos

Seleccionando la segunda opción de trabajo dentro del StatWizard se despliega la siguiente ventana

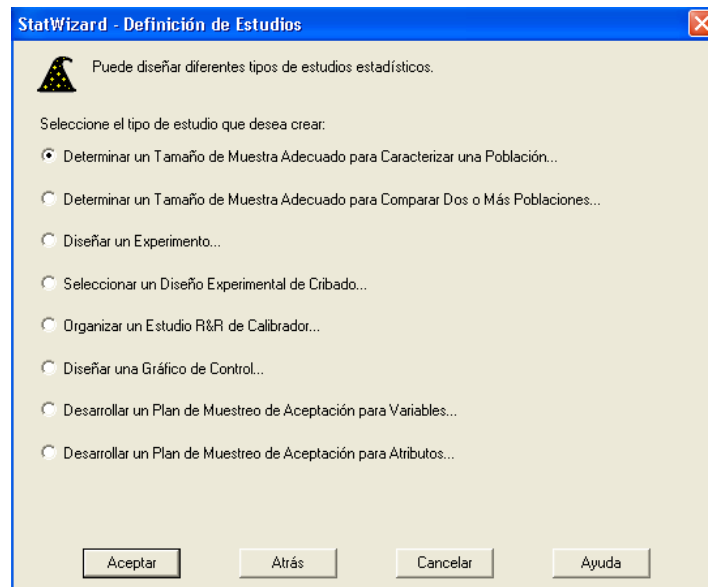


GRAFICO 4. STATWIZARD DEFINICIÒN ESTUDIOS

En ella ha de escogerse la opción que se adecue al proceso que se desee ejecutar.

3) Realizar un análisis que no requiere datos

Al seleccionar esta opción ha de desplegarse el siguiente cuadro de dialogo

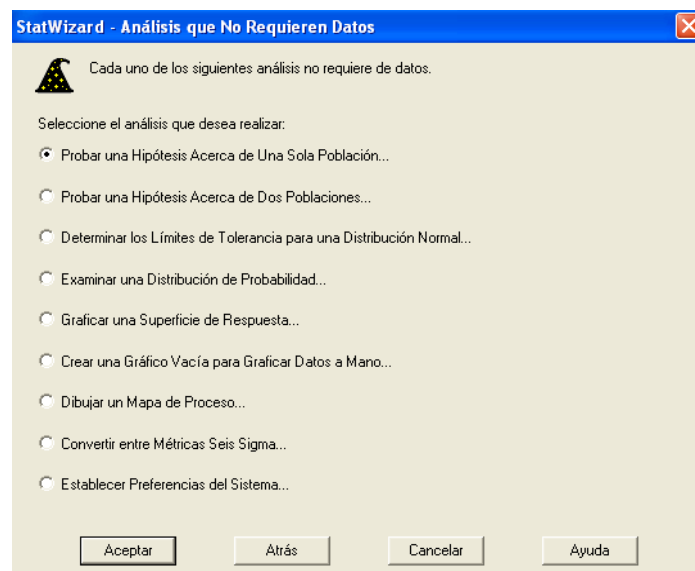



GRAFICO 5. STATWIZARD ANALISIS QUE NO REQUIEREN DATOS

Todas las opciones de trabajo disponibles en cada una de las tres subsecciones a seleccionar guían al usuario a través de una serie de ventanas que le permitirán configurar los parámetros de su estudio.

1.4.2 Selección Análisis Datos

Luego de ejecutar tanto una de las anteriores opciones como una de las referenciadas en el modulo 2 (Ver Sección 2.2) para la introducción de datos a la hoja de trabajo de StatGraphics hemos de poder seleccionar el tipo de análisis de interés a aplicar sobre las variables a trabajar.

De esta manera podemos entrar al StatWizard a través del icono  ubicado en la barra de herramientas, se desplegará la siguiente ventana

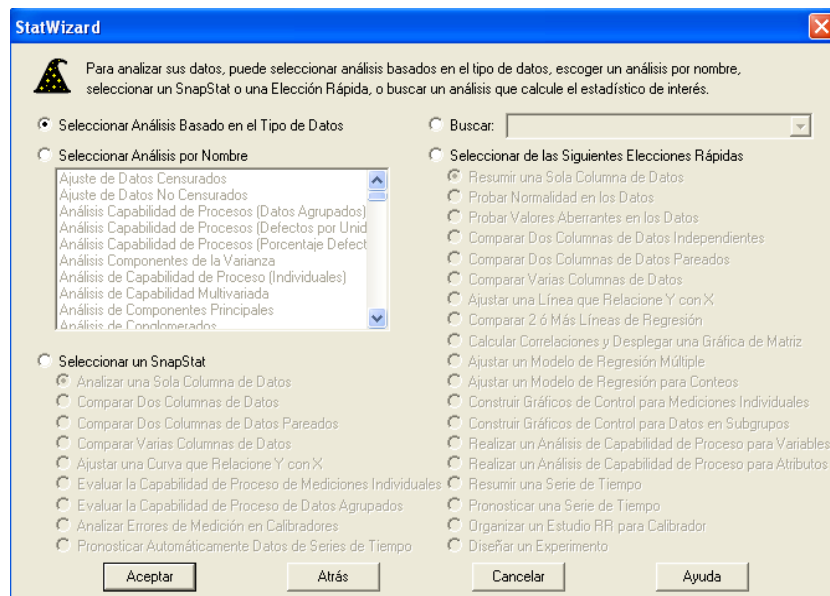


GRAFICO 6. STATWIZARD SELECCIÓN ANÁLISIS

Dentro de ella el usuario podrá seleccionar el tipo de análisis que corresponda con su estudio, esto dentro de las opciones disponibles:

- 1) Análisis Basados en Tipo de Datos: al seleccionar esta opción ha de desplegarse la ventana de selección de datos, donde el usuario podrá escoger aquellas variables que servirán como parámetros de desarrollo de estudio en curso.

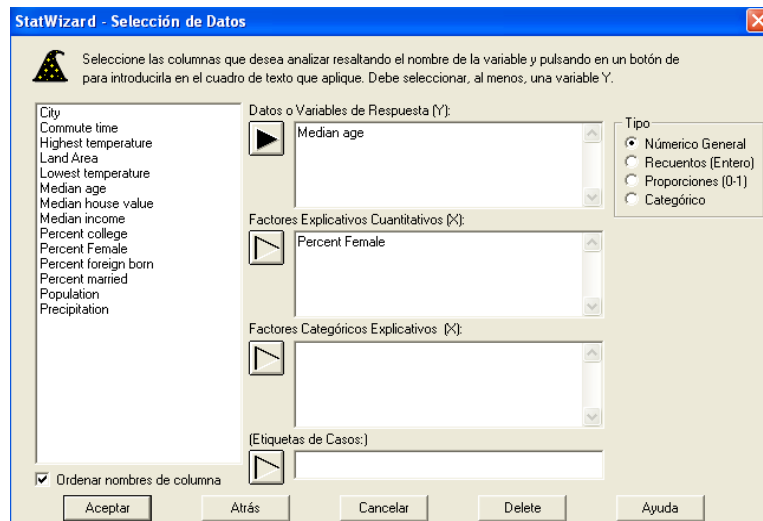


GRAFICO 7. STATWIZARD SELECCIÓN DATOS

Seguido de esto el StatWizard pedirá al usuario el especificar cuales variables analizará, esto es si se incluyen todas o por el contrario un subgrupo de las mismas.

Al seleccionar las variables que serán sujeto de estudio se desplegará la ventana de selección de Análisis

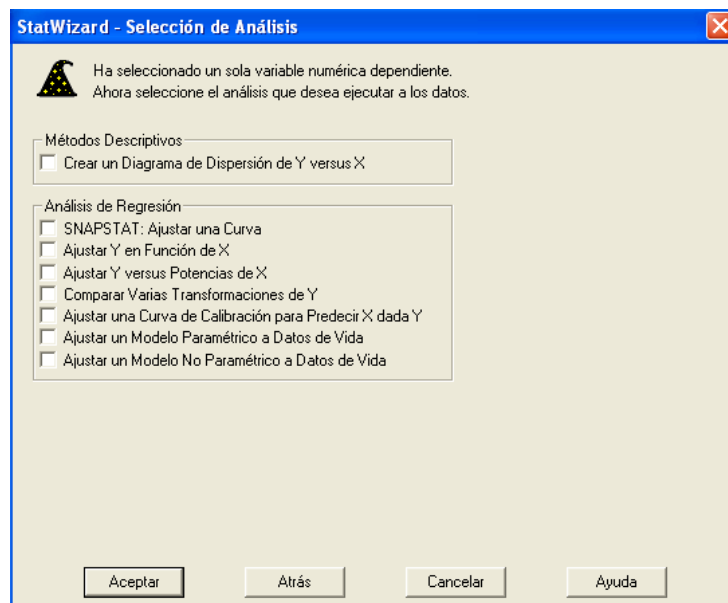


GRAFICO 8. STATWIZARD SELECCIÓN ANALISIS/DATOS

Al seleccionar el tipo de análisis de interés el StatWizard nos preguntará si ha de presentarse el resultado del análisis en una nueva ventana

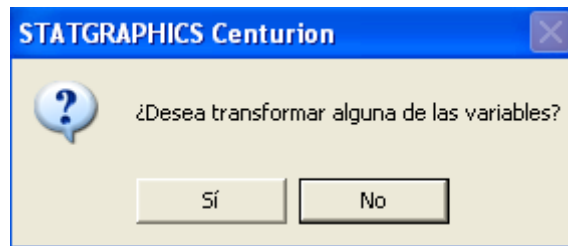


GRAFICO 9. CUADRO DIALOGO

- 2) Selección de Análisis por Nombre: se selecciona el análisis de interés de una lista desplegable disponible en el StatWizard

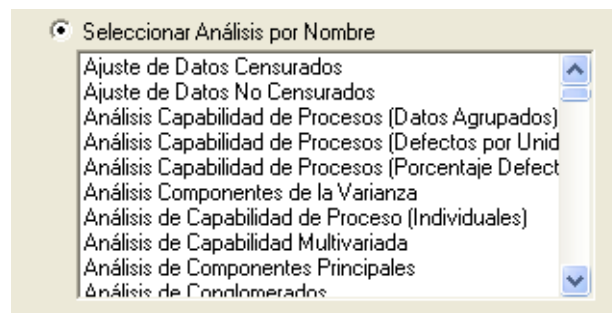


GRAFICO 10. STATWIZARD SELECCIÓN ANALISIS/NOMBRE

- 3) SnapStat: “Los SnapStats son análisis concisos que producen una sola página de salida pre-formateada.”¹ Según esto podremos ejecutar análisis sencillos sobre conjuntos de datos disponibles en la hoja de trabajo del StatGraphis, cuyos resultados serán expuestos en una nueva ventana.

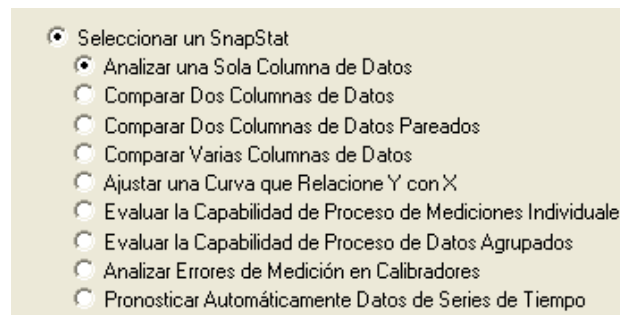


GRAFICO 11. STATWIZARD SELECCIÓN SNAPSTAT

- 4) Opción Buscar: El usuario podrá seleccionar de la lista disponible el conjunto de análisis que desee ejecutar, automáticamente las opciones existentes dentro de la opción de *Análisis por Nombre* han de cambiar haciéndose mas específicos y correspondientes con el buscado; seguido de buscar seleccionamos en el campo de *Análisis por Nombre* la tarea a ejecutar y damos clic en aceptar.

¹ Tomado de Guía del Usuario StatGraphics Centurion XV. Pág. 135

- 5) Selección Opciones Rápidas: El usuario encontrará los análisis comúnmente utilizados.

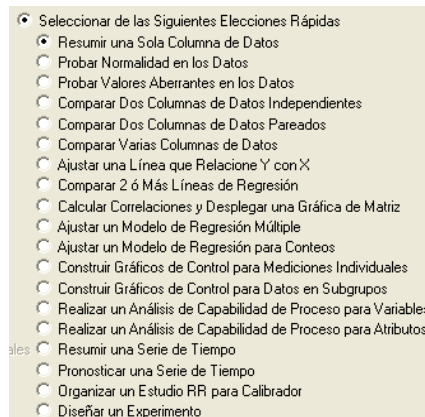


GRAFICO 12. STATWIZARD SELECCIÓN RAPIDA

1.5 CREANDO REPORTES DE TRABAJO: STATREPORTER

Esta herramienta le permite al usuario llevar un historial organizado de los procesos que ejecute dentro del StatGraphics; esta aplicación se carga predeterminadamente al abrir el programa (minimizada en el extremo derecho de la ventana de trabajo principal).

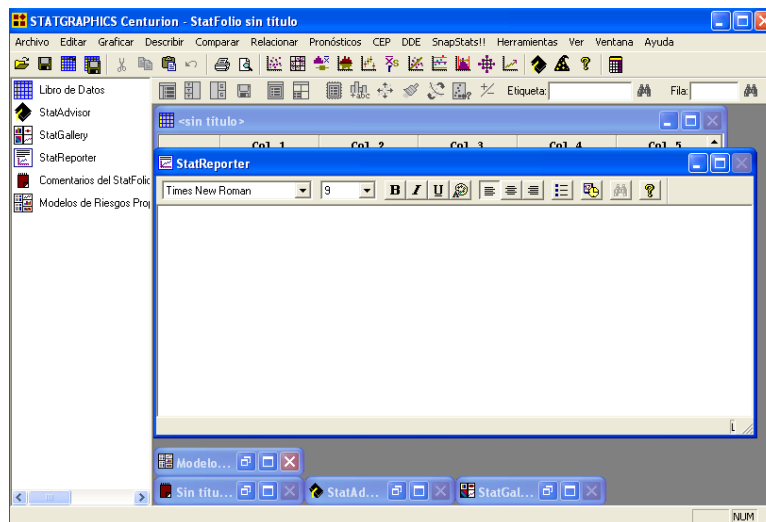


GRAFICO 13. STATREPORTER

Dentro de la ventana del StatReporter el usuario podrá editar texto referente a información relacionada con los procesos ejecutados, esto con el fin de llevar un reporte de los mismos.

1.5.1 Pegar Resultados

El StatReporter le permite al usuario copiar y pegar resultados obtenidos de análisis y procesos ejecutados dentro de la hoja de trabajo del StatGraphics.

Primer Procedimiento

- 1) Sobre el resultado de interés dar clic derecho y seleccionar la opción copiar
- 2) Maximizar el StatReporter
- 3) Sobre la hoja principal del StatReporter clic derecho y seleccionar la opción pegar.

Segundo Procedimiento

- 1) Sobre el resultado de interés dar clic doble
- 2) Clic derecho y seleccionar *Copiar Ventana al StatReporter* o en caso de querer copiar todos los resultados obtenidos se ha de seleccionar *Copiar Análisis al StatReporter*

Hay que tener en cuenta que lo que se copie dentro del StatReporter no cambiará en caso de que los parámetros de análisis sean modificados.

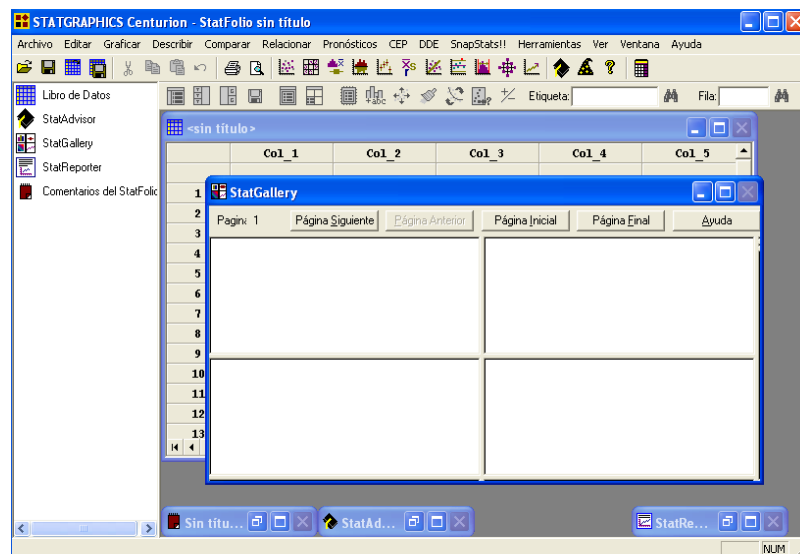
1.6 CREANDO PRESENTACIONES DE GRAFICAS: STATGALLERY

Esta aplicación consiste en una ventana especial de visualización de graficas, esto es la vista simultánea de gráficas generadas en diferentes procesos en un mismo plano.

Los archivos de esta especie se guardan en extensión sgg.

1.6.1 Configuración

El StatGallery está cargado predeterminadamente cuando se abre un nuevo proyecto, (la ventana se encuentra minimizada), se divide en cuatro secciones, cada una de las cuales es capaz de mostrar una gráfica diferente (la ventana primaria del StatGallery puede mostrar hasta nueve diferentes gráficas):



Además de la página principal se pueden crear un número mayor de páginas de navegación a las cuales podemos acceder a través de los botones que se encuentran ubicados en la parte superior de la página principal.

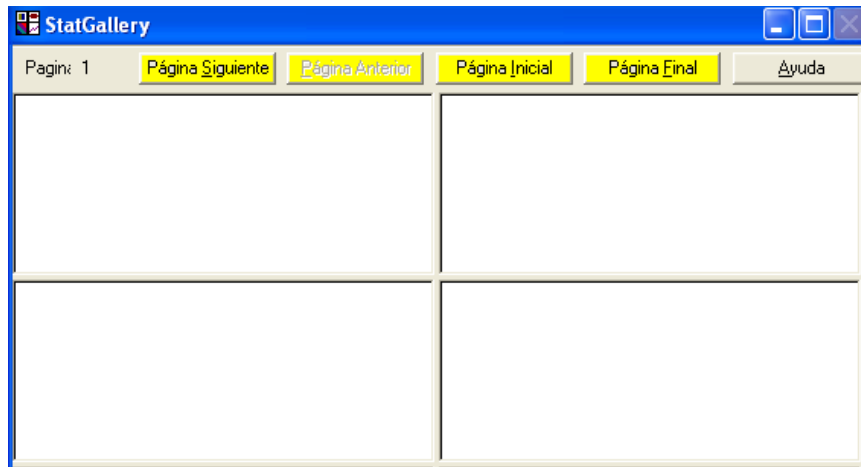


GRAFICO 15. PESTAÑAS NAVEGACIÓN STATGALLERY

A través de clic derecho sobre el StatGallery y accediendo al submenú Organizar Ventanas podemos cambiar el número de graficas a visualizar:



GRAFICO 16. OPCIONES STATGALLERY

1.6.2 Sobreposición de Graficas

Cuando una gráfica es pegada desde otra aplicación o tipo de análisis al StatGallery existen dos posibilidades, la primera es reemplazar la grafica existente o sobreponerla (esto es pegar una gráfica sobre otra sin perder la información individual) en este caso la información contenida en los ejes de la segunda gráfica serán los visibles.

1.6.3 Copiar Gráficos

Luego de haber creado la gráfica correspondiente a un análisis específico la forma como podemos ubicarla dentro del StatGallery es el siguiente:

- 1) Crear gráfica del correspondiente análisis
- 2) Sobre la gráfica dar clic derecho y sobre el submenú seleccionar Copiar (o comando rápido Ctrl+C)
- 3) Abrir la ventana del StatGallery y sobre el recuadro de interés dar clic derecho y seleccionar pegar (o comando rápido Ctrl+V)

1.7 Modificación de Gráficos

1.7.1 Detalles Gráficos

Para añadir detalles gráficos a la gráfica de interés seguimos el siguiente procedimiento:

- 1) Maximizar la gráfica, para esto damos doble clic sobre la misma
- 2) Clic derecho sobre la gráfica
- 3) Submenú **Añadir Objeto** (ver **Gráfico No 18**)

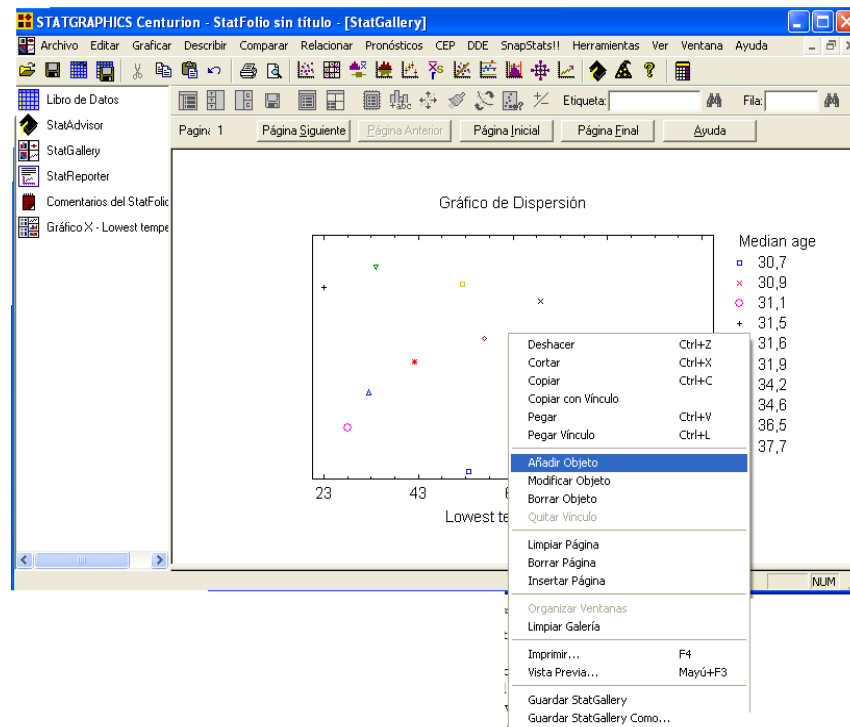


GRAFICO 17. CUADRO DIALOGO OPCIONES STATGALLERY

- 4) Se despega el siguiente cuadro de dialogo

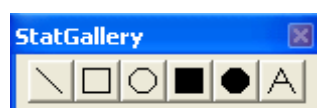



GRAFICO 18. AÑADIR OBJETO

- 5) Para añadir cualquiera de los objetos de la herramienta seleccionamos el de interés y sobre la gráfica correspondiente trazamos (con clic sostenido) el área de la nueva forma.
- 6) La opción  nos permite crear un cuadro de texto dentro del Gráfico.

Para modificar los detalles gráficos que se hayan añadido se procede de esta forma:

- 1) Maximizar el gráfico que contiene el objeto a modificar
- 2) Clic sobre el objeto (aparecen marcas alrededor del mismo, lo que indica que se ha sido seleccionado)
- 3) Clic derecho sobre el objeto y en el submenú se selecciona *Modificar Objeto*.

Este mismo procedimiento se ejecuta para borrar un detalle gráfico añadido, con la diferencia que en el submenú se selecciona *Borrar Objeto*.

2. MANEJO DE DATOS

2.1 EL LIBRO DE DATOS

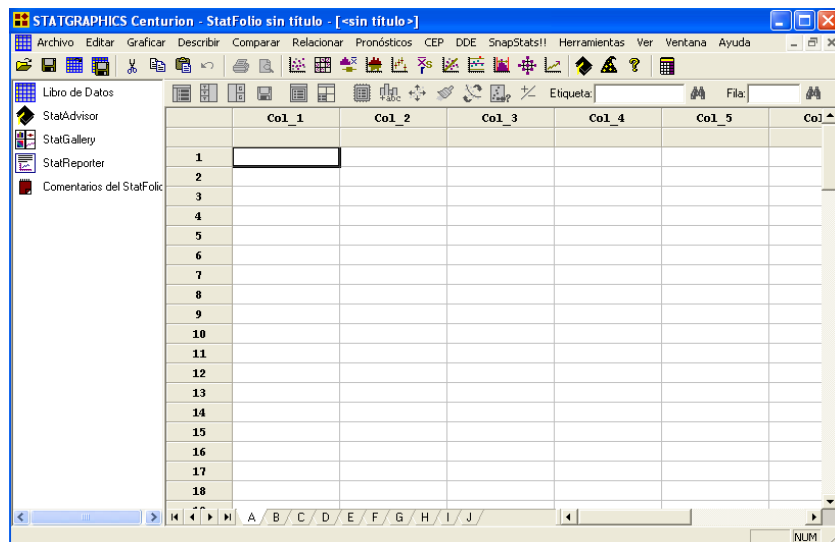


GRAFICO 19. LIBRO DE DATOS STATGRAPHICS

Como vemos cada columna ha de representar una variable, es decir el objeto sobre el cual trabajaremos diferentes mediciones. Cada variable ha de tener un nombre y un tipo determinado, los cuales podemos modificar dando doble clic sobre el nombre de columna, de manera que se despliegue el submenú de MODIFICACIÓN DE COLUMNA

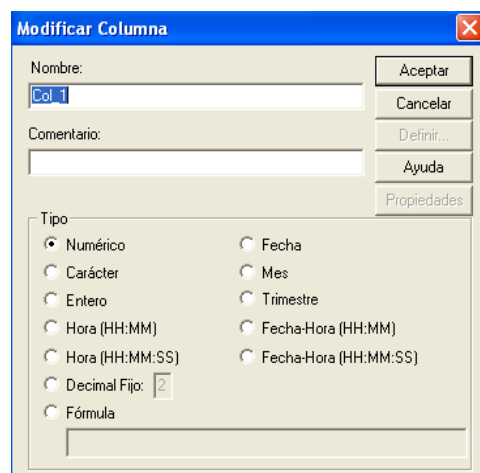


GRAFICO 20. CUADRO DIALOGO MODIFICACION COLUMNA

Donde hemos de modificar los atributos de cada variable ya especificada por columna. El **NOMBRE** permite identificar el dato dentro del análisis que ejecutemos; debe ser único para cada columna; de 1 a 32 caracteres; no puede iniciar con un número.

El (los) **COMENTARIO** provee información adicional sobre especificaciones de variables.

El tipo determina el proceso de análisis de la variable, es decir el como será tratada durante el mismo. La siguiente tabla relaciona la clasificación de tipos de variables²

TIPO	CONTENIDO
Numérico	Cualquier número válido
Carácter	Cadena alfanumérica
Integral	Número Integral
Fecha	Mes, día y año
Mes	Mes y Año
Trimestre	Trimestre y año
Hora (hh:mm)	Hora y minuto
Hora (hh:mm:ss)	Hora, minuto y segundo
Fecha-Hora (hh:mm)	Mes, día, año, hora y minuto
Fecha-Hora (hh:mm:ss)	Mes, día, año, hora, minuto y segundo
Decimal	Números de 1 a 9 posiciones decimales
Formula	Calculado de otras columnas

CUADRO I. CLASIFICACION TIPOS DE VARIABLE

De esta manera el tipo debe coincidir con la el nombre de la columna o de otra manera será rechazada.

2.2 INTRODUCCIÓN DE DATOS

Otra forma de trabajo consiste en importar bases de datos ya existentes en archivos externos a la hoja activa sobre la que ejecutemos procesos de análisis; esto puede hacerse de tres formas diferentes

- **Abriendo un archivo existente:** En la barra de menú Archivo/Abrir/Abrir Datos o utilizado comando rápido CTRL+F12; seguido de esto se desplegará un cuadro de dialogo de de origen de datos donde podremos elegir el tipo de archivo a abrir.

² Tomado de: STATGRAPHICS CENTURION XV User Manual

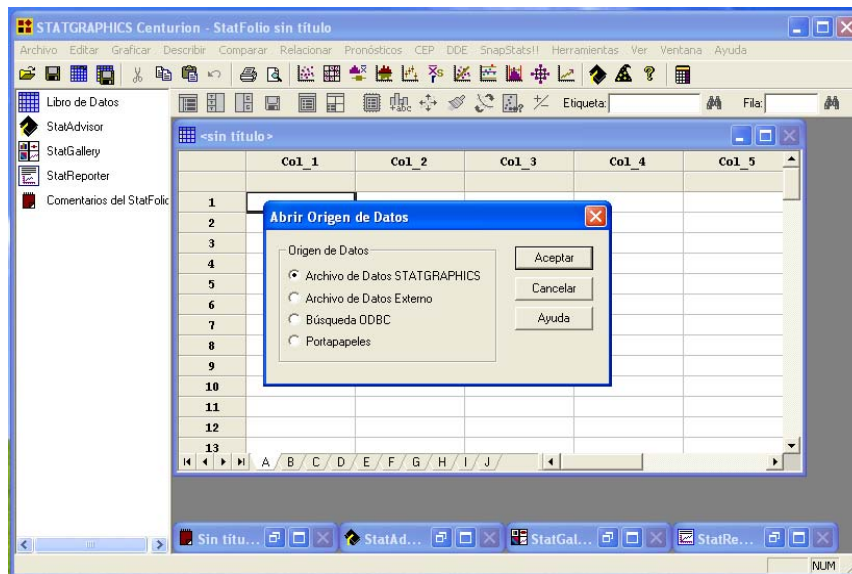


GRAFICO 21. CUADRO DIALOGO ORIGEN DATOS

- **Copiar y pegar:** copiando las hojas activas del archivo externo podemos pegarlas en el entorno de trabajo de StatGraphics por medio de (en la barra de herramientas) Editar/Pegar o utilizando el comando rápido CTRL+V.

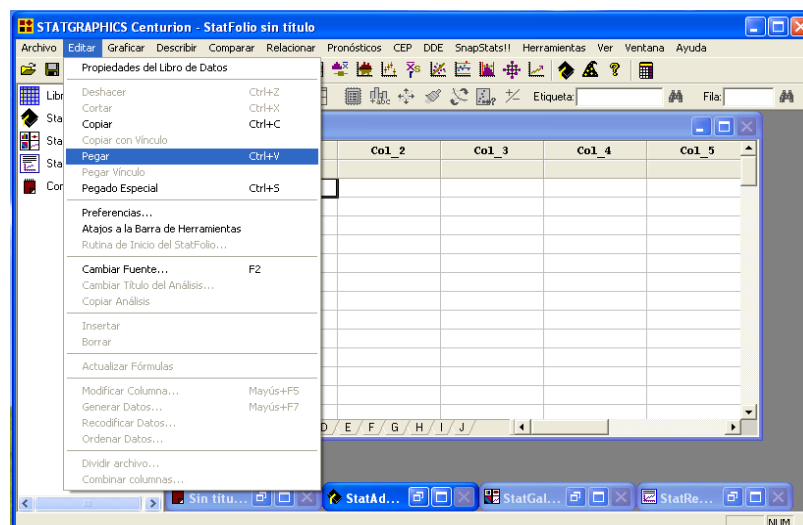


GRAFICO 22. SELECCIÓN OPCION PEGAR

Utilización de SQL para recuperación de bases de datos: Si el archivo a trabajar se encuentra en bases de datos tipo ODBC podemos importar a través de Archivo/Abrir/Abrir Datos y seleccionando Búsqueda ODBC.

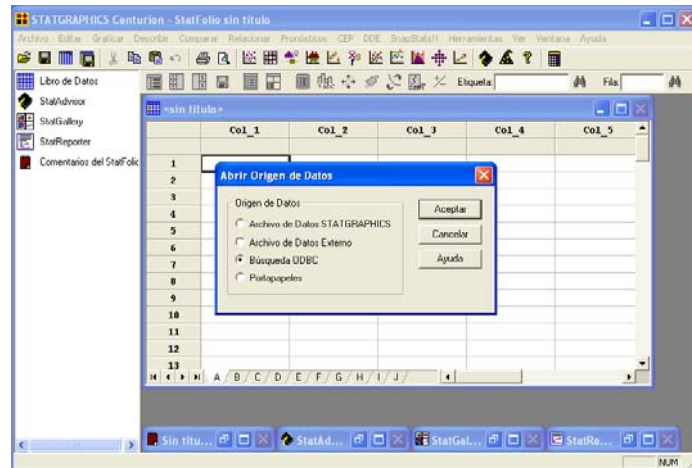


GRAFICO 23. LIBRO DE DATOS STATGRAPHICS

Otra forma de ingresar datos consiste en el trabajo sobre archivos anteriormente guardados en STATGRAPHICS; Archivo/Abrir/Abrir Datos/Archivo de datos STATGRAPHICS

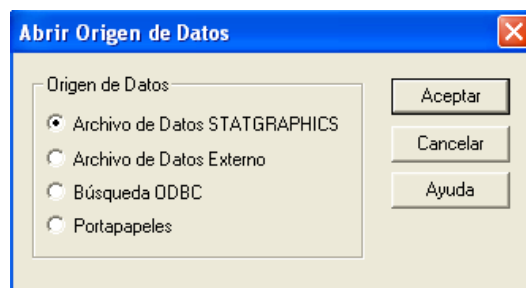


GRAFICO 24. CUADRO DIALOGO ORIGEN DATOS

Donde se desplegara un cuadro de dialogo de búsqueda de archivos

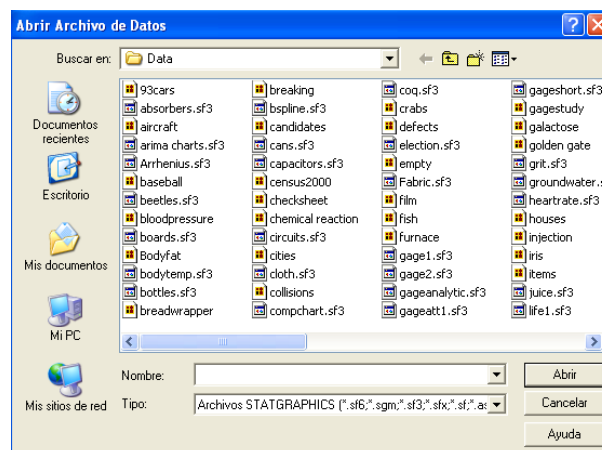


GRAFICO 25. VENTANA ABRIR ARCHIVO DATOS

Hay que tener en cuenta que al importar archivos para trabajo en STATGRAPHICS es necesario especificar la extensión del mismo, esto es ejecutando Archivo/Abrir/Abrir Datos/Archivo de datos externo

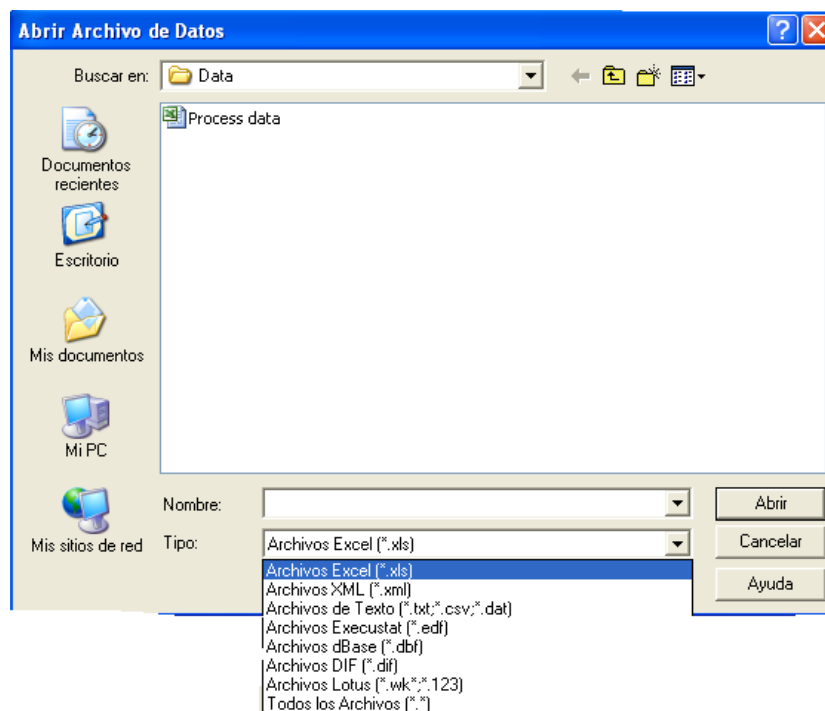


GRAFICO 26. SELECCIÓN EXTENSION ARCHIVO

En el menú desplegable de tipo seleccionamos la extensión del archivo.

Al abrir el archivo aparecerá un cuadro de dialogo donde especificaremos la información adicional del archivo de datos externo que se registrará en el libro de trabajo de STATGRAPHICS.

Podremos especificar:

- **Encabezado de Columna:** información contenida en la fila inicial.
- **Hoja Numero:** Número de hoja del archivo externo a analizar
- **Fila Inicial y final:** Rango de filas dentro del archivo externo sobre las cuales se ha de ejecutar análisis en STATGRAPHICS

2.3 MANIPULACIÓN DE DATOS

2.3.1 Copiar y pegar datos

Dentro la hoja de trabajo de STATGRAPHICS podemos copiar y pegar de una ubicación a otra diferentes rangos de datos, pero debemos tener en cuenta que (como ya se ha expuesto) cada columna ha de tener un tipo específico, por lo tanto al copiar y pegar datos sobre diferentes columnas debemos cuidar que el dato a pegar tenga el mismo formato de tipo a la ubicación donde deseamos ubicarlo, de otra manera

STATGRAPHICS cambiara el tipo de columna lo que podría generarnos problemas de resultado en la posterior ejecución de análisis.

2.3.2 Crear Nuevas Variables basadas en columnas existentes

Este procedimiento puede llevarse a cabo de dos formas diferentes (utilizando operadores de cálculo matemático):

1. Sobre los campos de datos a través de los cuadros de dialogo de introducción de datos.
2. Creando una nueva columna en la hoja de trabajo (insertando formulas dentro del dialogo de Modificación de Columnas)

2.3.3 Transformación de datos

Utilizando funciones matemáticas podemos transformar datos existentes en la hoja de trabajo de STATGRAPHICS, de la misma forma en que creamos nuevas variables, podemos transformarlas directamente sobre la base de datos (cambiando los atributos de variable dentro del cuadro de dialogo de introducción de datos) o creando nuevas columnas donde se han aplicado operadores matemáticos que relacionan los datos de las columnas precedentes.

OPERADOR	USO
+	Suma
-	Resta
/	División
*	Multipliación
^	Exponencial
ABS	Valor absoluto
AVG	Media
DIFF	Diferenciación de número precedente
EXP	Función Exponencial
LAG	Desfase por k periodos
LOG	Logaritmo Natural
LOG10	Logaritmo Base 10
MAX	Máximo
MIN	Mínimo
SD	Desviación Estándar
SQRT	Raíz cuadrada
STANDARDIZE	Conversión a valores Z

CUADRO 2. OPERADORES

Dentro de la programación de STATGRAPHICS podemos vincular formulas de manera que se realicen simultáneamente varias operaciones sobre una variable o conjunto de

variables relacionadas. Por ejemplo sobre dos variables denominadas *Peso* y *Masa* podemos elaborar la siguiente secuencia de formulas:

$$\frac{Peso}{Masa} + \{MIN(Masa) * MAX(Peso)\}$$

Debemos recordar que los paréntesis son necesarios para establecer el orden en que se ejecutarán las operaciones.

De la misma manera el programa posee una especie de tutorial para la generación de formulas, esto es:

- Al haber importado o abierto directamente de STATGRAPHICS el archivo a trabajar podemos a través de la opción “Evaluador de Expresiones” transformar los datos utilizando los operadores matemáticos/analíticos de la lista ubicada a la derecha del cuadro de dialogo

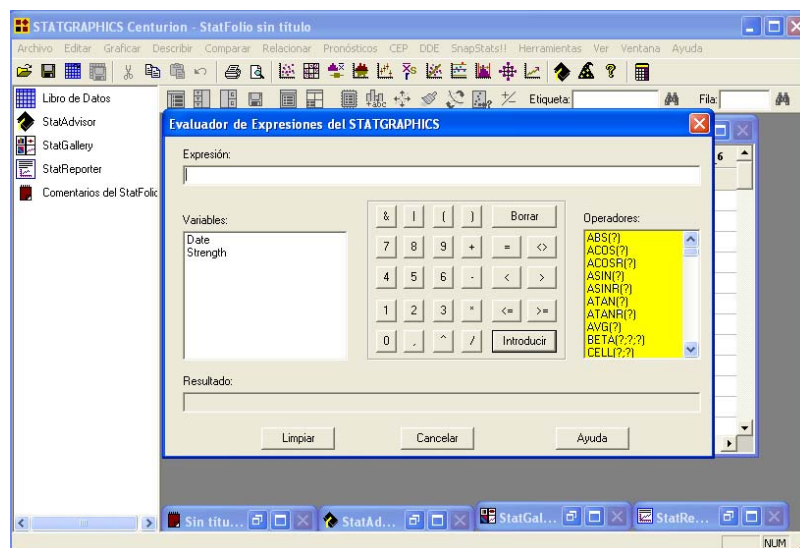


GRAFICO 27. EVALUADOR EXPRESIONES

Seleccionamos con doble clic el operador a utilizar y donde se indiquen signos de interrogación (?) ubicamos la variable a operar

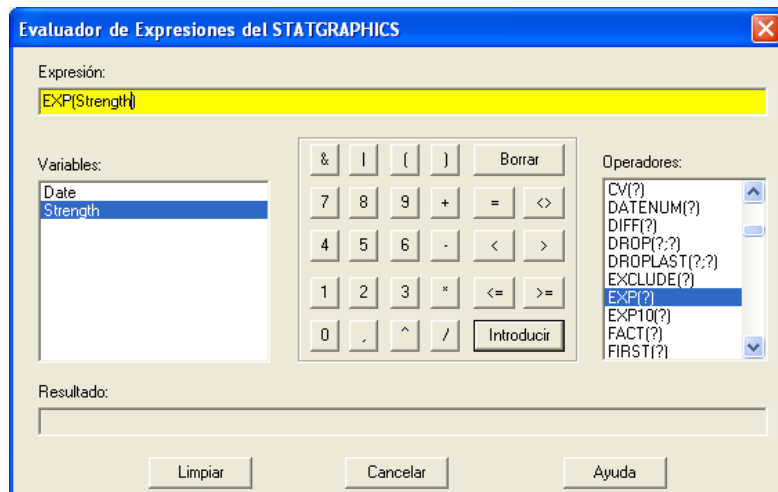


GRAFICO 27. UBICACIÓN VARIABLE DENTRO DE EVALUADOR EXPRESIONES

2.4 ORGANIZACIÓN DE DATOS

El menú Editar/Ordenar Datos nos permite acceder al siguiente cuadro de dialogo



GRAFICO 28. CUADRO DIALOGO ORDENAR DATOS

Se presenta la opción de ordenar una o dos columnas de la hoja de trabajo, donde debemos especificar la ubicación de cada una de ellas esto es si es la columna principal o la secundaria; luego de esto y según las preferencias del usuario los datos serán ordenados en forma ascendente/descendente o aleatoria.

Sin embargo es importante recordar que para la ejecución de análisis en STATGRAPHICS no es necesario organizar los datos antes de los procedimientos ya que el programa se encarga de realizar esta acción de ser necesario.

2.5 RECODIFICACIÓN DE DATOS

Esta opción nos permite agrupar datos en grupos de similares características y asignar nuevas etiquetas a las variables.

Para acceder a esta opción primero seleccionamos el rango de datos a recodificar, luego ingresamos al menú Editar/Recodificar Datos, donde nos desplegará la siguiente ventana



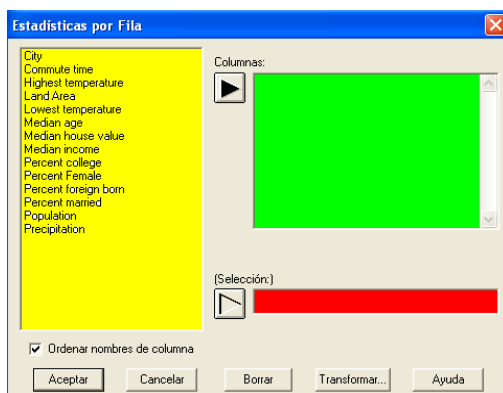
The 'Recodificar Datos' dialog box is used for recoding data. It features three columns of input fields for 'Límite Inferior', 'Límite Superior', and 'Nuevo Valor'. Below these fields are two sections of radio buttons: 'Condiciones de los Límites' with four options (the first is selected), and 'Sin condición' with two options (the first is selected). There is also an 'Extrapolar' checkbox. At the bottom are 'Aceptar', 'Cancelar', and 'Ayuda' buttons.

GRAFICO 29. CUADRO DIALOGO RECODIFICACION DATOS

En ella determinamos los valores a cambiar (límite superior e inferior) y los nuevos valores.

2.6 COMBINACIÓN DE COLUMNAS

Para analizar un rango de datos que incluye varias columnas diferentes utilizamos la Herramienta Estadísticas por Fila, entrando al menú Describir/Datos Numéricos/Estadísticas por fila



The 'Estadísticas por Fila' dialog box shows a list of variables on the left, including 'City', 'Commute time', 'Highest temperature', 'Land Area', 'Lowest temperature', 'Median age', 'Median house value', 'Median income', 'Percent college', 'Percent female', 'Percent foreign born', 'Percent married', 'Population', and 'Precipitation'. A 'Columnas:' section on the right has a green selection area. Below it is a '(Selección:)' section with a red selection area. At the bottom, there is a checked checkbox 'Ordenar nombres de columna' and buttons for 'Aceptar', 'Cancelar', 'Borrar', 'Transformar...', and 'Ayuda'.

GRAFICO 30. CUADRO DIALOGO ESTADISTICAS POR FILA

En el menú ubicado en la parte izquierda nos despliega el nombre de las columnas que contienen datos en la hoja de trabajo, seleccionamos aquellas a las cuales será aplicado el análisis y las asignamos como columna, al dar clic en aceptar el programa nos arrojará el análisis desarrollado

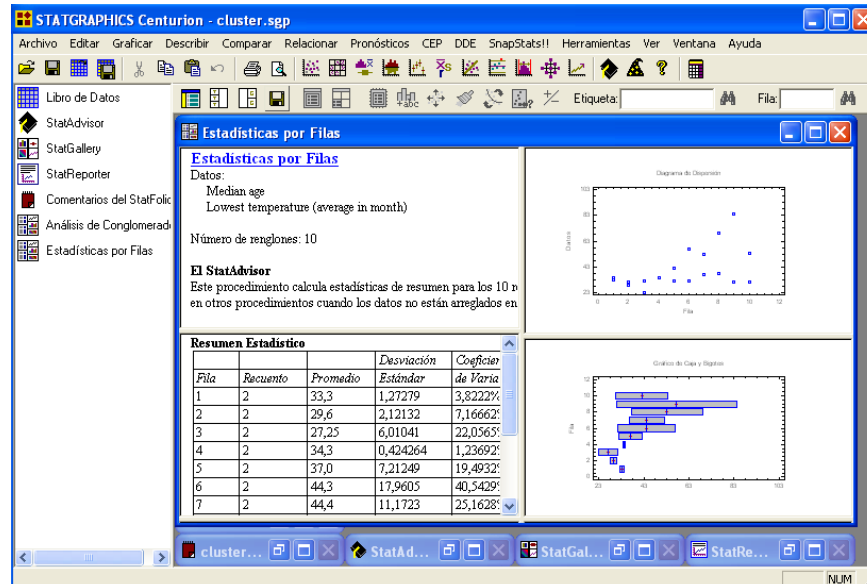


GRAFICO 31. RESULTADO ESTADISITCAS POR FILA

2.7 GENERACIÓN DE DATOS

Siguiendo la ruta Menú Editar/Generar Datos (o con el comando rápido Mayus+F/) se despliega el siguiente cuadro de dialogo

GRAFICO 32. CUADRO DIALOGO GENERACION DATOS

Se ejecuta el siguiente procedimiento para generación de datos

1. Seleccionamos la columna donde queremos queden ubicados los datos a generar


2. Entramos al cuadro de dialogo de Generación de Datos
3. De la lista de operadores seleccionamos la opción necesaria
4. Dentro de los paréntesis ubicamos la variable a calcular
5. Clic en Mostar
6. Clic en aceptar
7. En la hoja de trabajo debe aparecer en la columna anteriormente seleccionada los datos generados.

3. GRAFICOS

3.1 MODIFICACIÓN DE GRÁFICAS

3.1.1 Opciones de Diseño

Luego de crear una grafica hemos de tener la opción de modificar sus atributos ejecutando el siguiente procedimiento:

1. Doble Clic sobre la grafica de manera que se habilite sobre la barra de análisis las opciones de grafico
2. Clic sobre el icono de opciones de gráfico 
3. se ha de desplegar el siguiente cuadro de dialogo

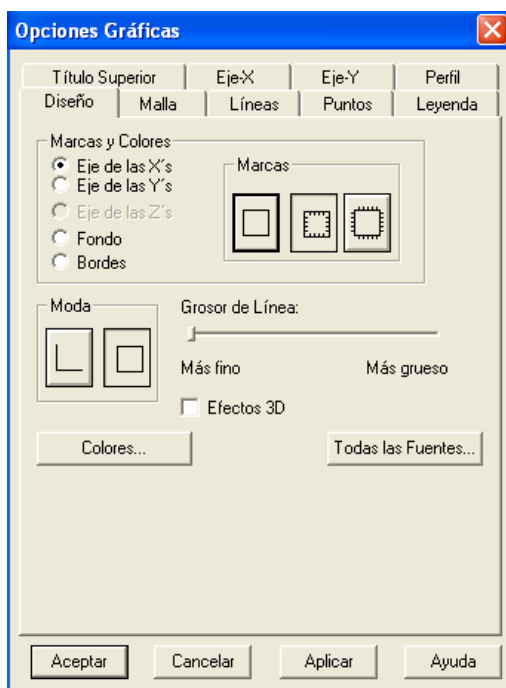



GRAFICO 33. CUADRO DIALOGO OPCIONES GRAFICAS

En el podemos cambiar configurar del gráfico ya generado. (Titulo, Aspecto de líneas y puntos, configuración de ejes)

4. Una forma de añadir texto adicional al grafico (tipo leyenda) es ingresando al cuadro de dialogo *Agregar Texto* a través del icono  de la barra de análisis de gráfico.

3.2 SEPARAR PUNTOS EN GRÁFICO DE DISPERSIÓN

Cuando una o dos de las variables de una de dispersión establecida son discretas, la oportunidad de que los puntos se ubiquen exactamente en el mismo lugar de manera que un punto se encuentre ubicado encima de otro puede ser muy grande. Para solucionar este inconveniente dentro de la barra de análisis se ofrece una herramienta cuyo trabajo consiste en asignar puntos aleatorios en dirección vertical u horizontal.

Entonces a través del icono  ingresamos al siguiente cuadro de dialogo

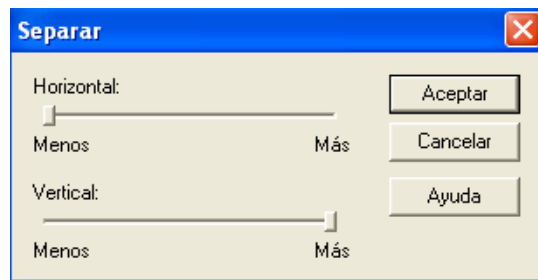



GRAFICO 34. CUADRO DIALOGO SEPARACIÓN

Donde podemos establecer la distancia y dirección sobre la cual deseamos que los puntos de dispersión de la grafica se separen.

3.3 RESALTAR PUNTOS EN GRÁFICO DE DISPERSIÓN

STATGRAPHICS ofrece la oportunidad de visualizar la relación existente entre diferentes variables, a través de la asignación de colores que resalten el valor de las variables bajo estudio.

Dentro del entorno de trabajo en la barra de análisis de gráficos se encuentra el icono  dando clic sobre el accedemos al cuadro de dialogo de “Resaltar”

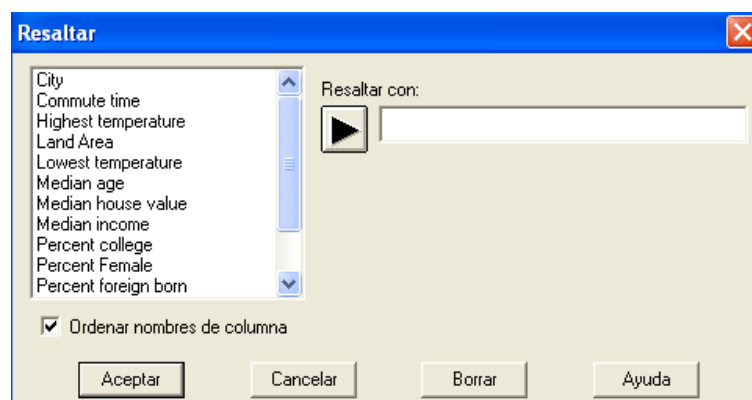


GRAFICO 35. CUADRO DIALOGO RESALTAR

Allí se selecciona la variable a resaltar (cambio de color) dentro del grafico de dispersión.


3.4 IDENTIFICACIÓN DE PUNTOS

Una forma de establecer la ubicación original de los datos de una variable consiste en que con el botón izquierdo del Mouse se de clic sobre el punto de interés de manera que en la parte superior derecha del grafico aparezcan las coordenadas del punto de interés.

Simultáneamente el numero de la fila correspondiente al punto seleccionado será colocado en el campo “fila” de la barra de herramientas de análisis.



GRAFICO 36. IDENTIFICACION DE PUNTOS

Una forma alterna para obtener información sobre un punto específico de la gráfica de dispersión consiste en seleccionar una columna del libro de trabajo y seguido de esto ingresar a la opción “Identificar” a través del icono  en la barra de herramientas para el análisis; se desplegará el siguiente cuadro de dialogo

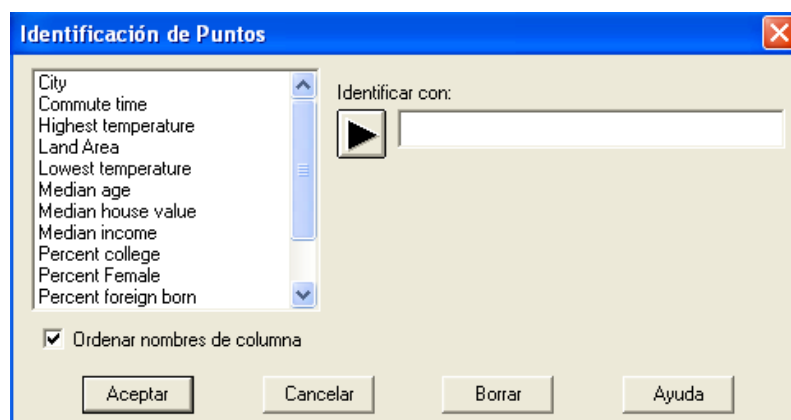


GRAFICO 37. IDENTIFICACION DE PUNTOS

Luego de seleccionar la variable correspondiente al dar clic sobre le punto de interés se desplegara su respectiva información en el campo “etiqueta” de la barra de herramientas del análisis.



GRAFICO 38. ETIQUETA DE PUNTO

Los binoculares que se encuentran a la derecha de los campos de “etiqueta” y “fila” son utilizados para ubicar puntos dentro la gráfica.

3.5 COPIAR GRÁFICAS A OTRAS APLICACIONES

Una gráfica que ha sido generada dentro del entorno de trabajo STATGRAPHICS puede ser copiada y pegada en un programa externo ejecutando el siguiente procedimiento:

1. Se maximiza la ventana que contiene el gráfico
2. A través del menú editar se selecciona la opción copiar
3. Dentro de la otra aplicación en su correspondiente menú edición seleccionamos pegar.

3.6 GUARDAR GRÁFICAS COMO ARCHIVOS DE IMAGEN

Al maximizar la ventana que contiene al gráfico se ingresa al menú archivo/Guardar Gráficos (o a través del comando rápido F3), se despliega el siguiente cuadro de dialogo

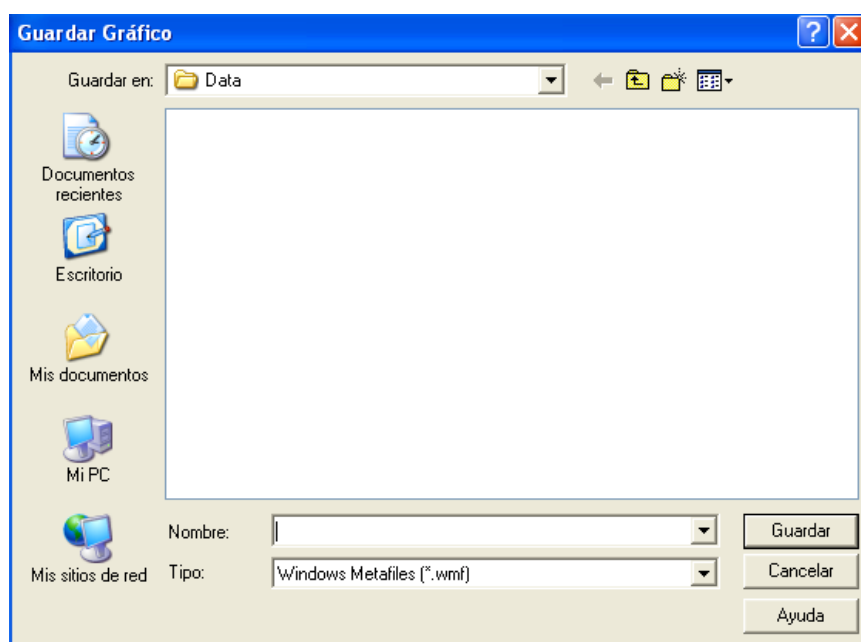


GRAFICO 39. CUADRO DIALOGO GUARDAR GRAFICO

Allí deben especificarse el nombre y tipo del archivo gráfico, para una mayor utilidad y de manera que sea posible abrir un archivo gráfico de STATGRAPHICS el tipo especificado debe coincidir con windows metafiles (*.wmf).

4. ANALISIS DE DATOS: REGRESION SIMPLE

USTED DEBE TENER EN CUENTA:

El procedimiento de **Regresión Simple** está diseñado para construir un modelo estadístico que describa el impacto de un solo factor cuantitativo X sobre una variable dependiente Y .

4.1 ACCESO A LA REGRESION SIMPLE

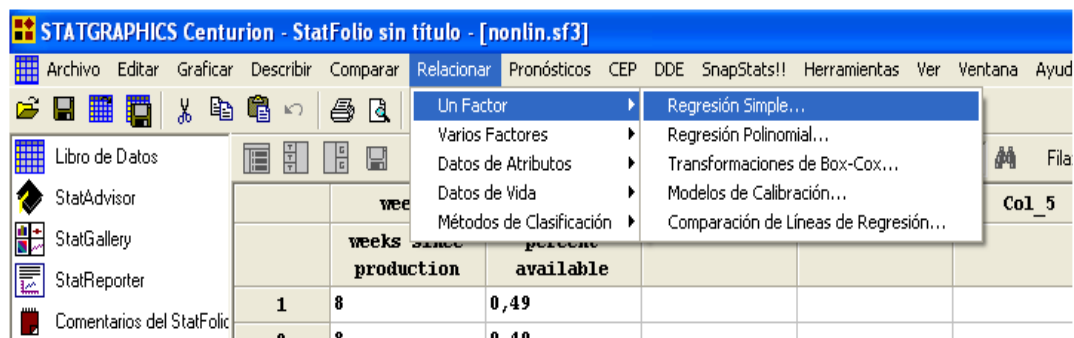


GRAFICO 40. MENU RELACIONAR/REGRESION SIMPLE

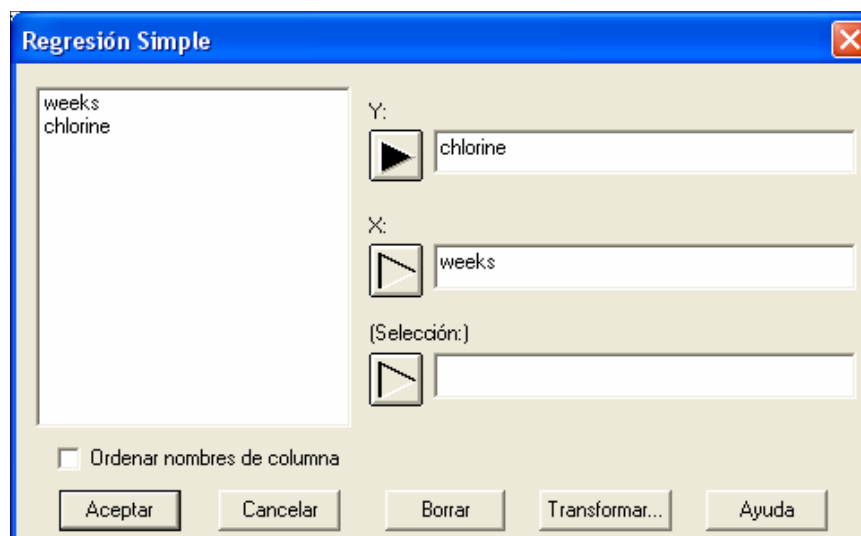


GRAFICO 41. CUADRO DIALOGO REGRESION SIMPLE

Y: columna numérica que contiene las n observaciones para la variable dependiente Y .

X: columna numérica que contiene los n valores para la variable independiente X .

NOTA TEORICA:

1. Recuerde que el modelo sobre el cual usted esta trabajando es $Y = a + b \cdot X$
2. El resultado que usted busca es especialmente respecto del valor de b (coeficiente) este le brindara la significancia, es decir que tanto la variable independiente explica a la variable dependiente.
3. Este hecho se resume en la Tabla ANOVA (análisis de varianzas)

Una vez completada la tabla de Regresión simple: Clic Aceptar. De inmediato obtendrá:

Resumen del analisis Gráfica del análisis

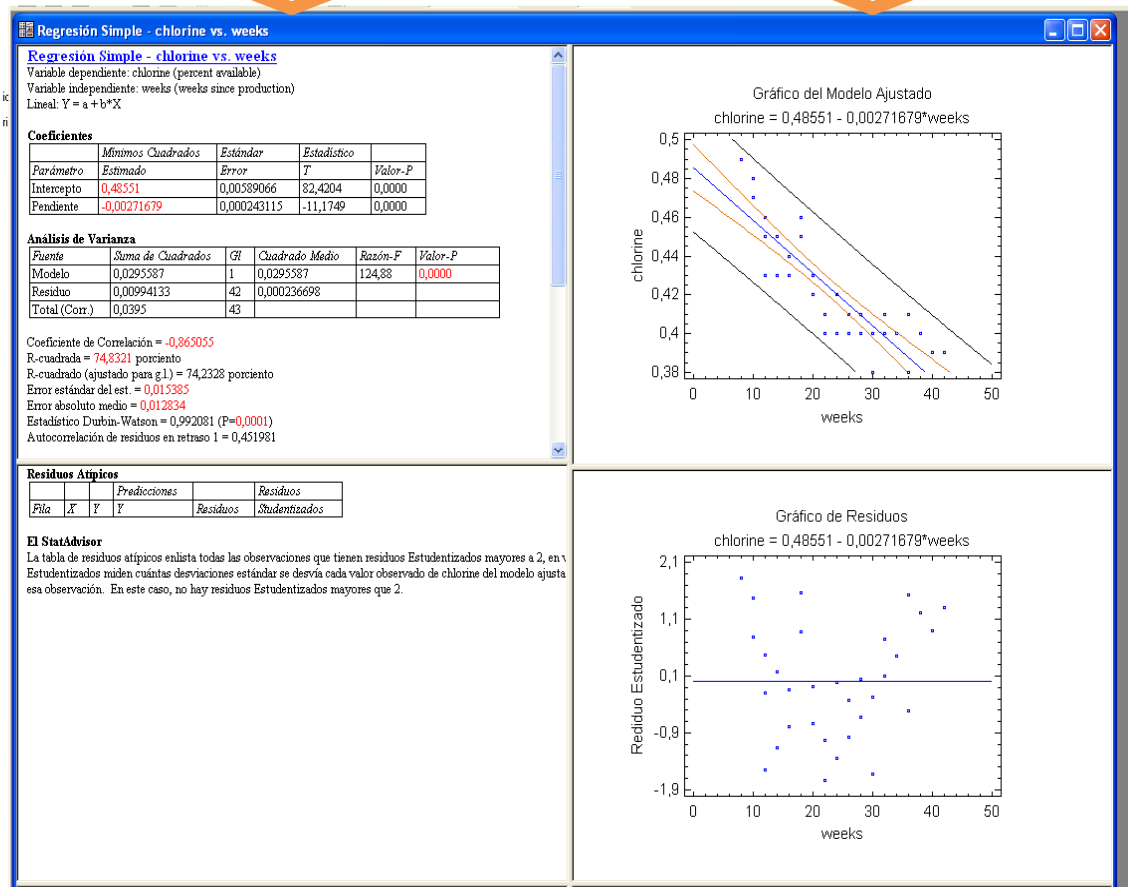


GRAFICO 42. ANALISIS REGRESION SIMPLE

4.2 RESUMEN DEL ANALISIS:

En este primer cuadro usted encontrara los resultados numéricos respecto de su análisis. Veamos cada uno de los componentes

4.2.1 Coeficientes

	<i>Mínimos Cuadrados</i>	<i>Estándar</i>	<i>Estadístico</i>	
<i>Parámetro</i>	<i>Estimado</i>	<i>Error</i>	<i>T</i>	<i>Valor-P</i>
Intercepto	0,48551	0,00589066	82,4204	0,0000
Pendiente	-0,00271679	0,000243115	-11,1749	0,0000

CUADRO 3. ANALISIS REGRESION SIMPLE/COEFICIENTES

Gracias a estos resultados usted podrá inferir el comportamiento de su modelo respecto de las variables que lo explican, además del grafico del mismo.

4.2.2 Análisis de Varianza

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Modelo	0,0295587	1	0,0295587	124,88	0,0000
Residuo	0,00994133	42	0,000236698		
Total (Corr.)	0,0395	43			

CUADRO 4. ANALISIS REGRESION SIMPLE/ANALISIS VARIANZA

RECUERDE:

Es de particular interés en éste análisis la Prueba F y su P-value asociado para probar la significancia estadística del modelo ajustado. Un pequeño P-Valor (menor a 0.05 si se opera con un nivel de significancia de 5%) indica que la relación estadística de la forma especificada existe entre Y y X. En los datos del ejemplo, el modelo es altamente significativo.

Finalmente usted encontrara información relevante para determinar la significancia:

Coeficiente de Correlación = -0,865055
 R-cuadrada = 74,8321 por ciento
 R-cuadrado (ajustado para g.l.) = 74,2328 por ciento
 Error estándar del est. = 0,015385
 Error absoluto medio = 0,012834
 Estadístico Durbin-Watson = 0,992081 (P=0,0001)
 Autocorrelación de residuos en retraso 1 = 0,451981

De este análisis usted puede inferir:

- El estadístico R-Cuadrada indica que el modelo ajustado explica 74,8321% de la variabilidad en chlorine.
- El coeficiente de correlación es igual a -0,865055, indicando una relación moderadamente fuerte entre las variables.

Para recordar como se da la construcción de estos resultados revise por favor la información contenida en ANEXO 43.1

Una vez se obtiene el resultado, es posible mostrarlo y analizarlo desde otras perspectivas, para ello dando clic derecho sobre el *Resumen del análisis* se desplegará la siguiente ventana

Opciones de Análisis...
Opciones Gráficas...
Deshacer
Seleccionar
Localizar
Zoom
Deshacer Zoom
Restablecer Escala/Perspectiva
Copiar
Copiar con Vínculo
Ctrl+C
Imprimir...
Vista Preliminar...
F4
Mayú+F3
Copiar Ventana a StatGallery...
Copiar Análisis a StatReporter...
Guardar Gráfica...

GRAFICO 43. SUB MENU RESUMEN ANÁLISIS

4.3 OPCIONES DE ANALISIS

Una vez usted da clic sobre esta opción se desplegará el siguiente cuadro de dialogo



GRAFICO 44. OPCIONES REGRESION MULTIPLE

Con esta herramienta usted podrá ajustar su modelo desde diferentes perspectivas, dados los anteriores criterios

Sobre este cuadro, usted señalará la opción que considere pertinente luego aceptar y por esta vía obtendrá un nuevo cuadro de respuestas como el descrito previamente.

La econometría ofrece la posibilidad de plantear un modelo desde diferentes énfasis, a continuación (ANEXO 4.2) usted encontrará una tabla-resumen respecto de las opciones que ofrece la literatura para plantear un modelo.

¿QUE SON LOS AJUSTES ALTERNOS?

RECUERDE:

En econometría existen diferentes métodos para calcular los Betas de la regresión (los mejores estimadores de la regresión), para ello se pueden citar:

- *Mínimos cuadrados*: Método típicamente empleado. Cuando los datos se encuentran uno en términos explicativos del otro.

También es posible encontrar métodos focalizados en el tratamiento de datos atípicos tales como:

- *Minimizar Desviaciones Absolutas* – Minimiza la suma de valores absolutos de las desviaciones alrededor del modelo ajustado.
- *Utilizar medianas de 3 grupos* – A través del método de Tukey de ajustar una línea recta, los datos son divididos en 3 grupos de acuerdo con los valores de X y se estiman las medianas para cada grupo. Finalmente una línea unirá las 3 medianas.

4.4 GRAFICA DEL ANALISIS:

En la siguiente parte del cuadro de respuesta usted encontrara el grafico del análisis.

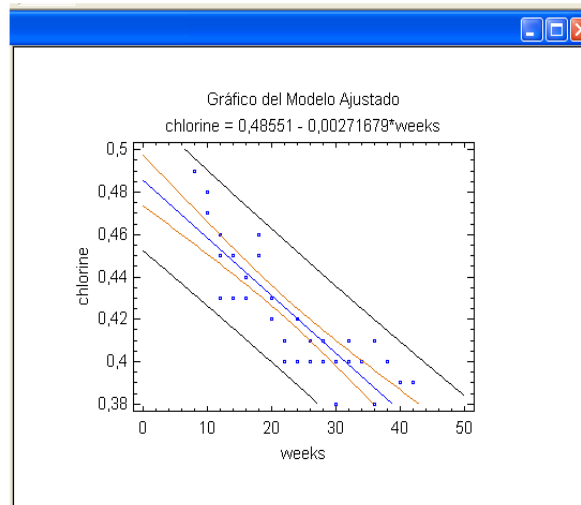


GRAFICO 45. GRAFICO DE ANÁLISIS

Recuerde que los puntos negros que aparecen en el grafico se consideran puntos de análisis de dispersión. El intercepto le permite deducir el comportamiento de la variable dependiente per se y la pendiente permite la elaboración de un grafico preciso.

Una vez usted sobre el grafico da clic derecho se despliega la siguiente ventana:



GRAFICO 46. CUADRO DIALOGO OPCIONES

OPCIONES DE VENTANA: Dando clic sobre opciones de ventana se desplegara el siguiente cuadro de dialogo:

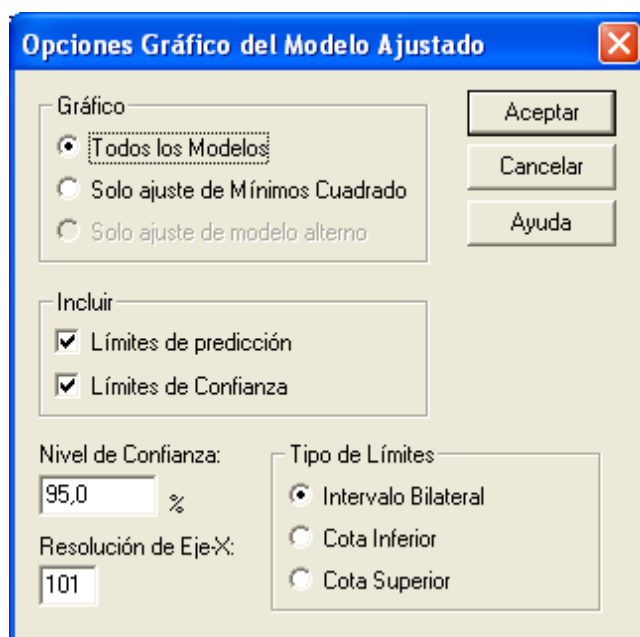


GRAFICO 47. CUADRO DIALOGO OPCIONES DE VENTANA

En esta ventana usted podrá decidir sobre los siguientes criterios:

Gráfico: Tipo de modelo o modelos para graficar.

Incluir: Límites que se van a incluir en la gráfica.

Nivel de Confianza: Porcentaje de confianza para los límites.

Resolución en el eje de las X: Número de valores de X que determinan la línea cuando ésta se grafica. Mientras más alta sea la resolución, la gráfica será más suavizada.

Tipo de Límites: Si grafica intervalos de confianza de dos o de una cola.

OPCIONES DE GRAFICO: Una vez selecciona esta opción, a través de ella usted podrá modificar la presentación de la grafica, rótulos, títulos, etc.



GRAFICO 48. CUADRO DIALOGO OPCIONES GRAFICAS

OPCIONES DE ANALISIS: Una vez usted da clic sobre esta opción se desplegará la siguiente ventana:

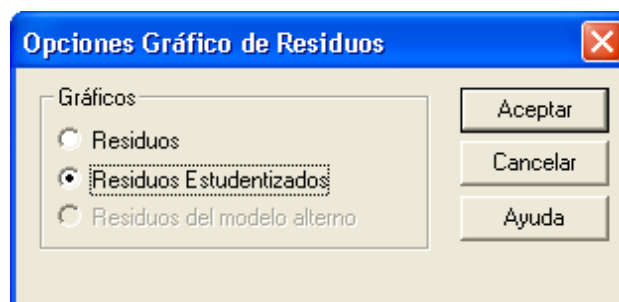


GRAFICO 49. OPCIONES GRAFICO DE RESIDUOS

Siguiendo el proceso descrito anteriormente también puede escoger como graficar su relación de residuos:

1. *Residuos* los residuos del ajuste de mínimos cuadrados.
2. *Residuos Estudentizados* la diferencia entre los valores observados y_i y los valores pronosticados \hat{y}_i cuando el modelo es ajustado usando todas las observaciones excepto la i -ésima observación, dividido por el error estándar estimado. Estos residuos son llamados algunas veces residuos eliminados externamente puesto que miden qué tan lejos cada valor se encuentra del modelo ajustado cuando el modelo se ajustó usando todos los datos excepto el punto que está siendo considerado. Esto es importante porque una observación atípica grande podría afectar el modelo tanto que no parecería estar lo suficientemente lejos de la línea.
3. *Residuos del modelo alternativo* – se refiere a los residuos cuando el modelo se estimó usando el método resistente.

RECUERDE:

En una regresión, los residuos son definidos como:

$$e = y - \hat{y}$$

Por ejemplo, los residuos son las diferencias entre los datos observados y el modelo ajustado.

Finalmente usted podrá hacer una comparación del mejor modelo para estimar a través de la ventana de Comparación de modelos alternos.

5. REGRESION MULTIPLE

El procedimiento de **Regresión Múltiple** está diseñado para construir un modelo estadístico describiendo el impacto de dos o más factores cuantitativos X sobre una variable dependiente Y . La presentación general del modelo es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Donde k es el número de variables independientes.

De allí, que el modelo colocado puede ser usado para hacer predicciones, incluyendo límites de confianza y límites de predicción. Los residuos pueden también ser graficados observando la manera en que influyen.

Para el desarrollo del presente capítulo, trabajaremos sobre el statfolio nombrado *múltiple reg.sgp*. Una vez accedemos al mismo tal y como lo hemos hecho previamente entonces seguimos la ruta descrita a continuación.

MENU: RELACIONAR: VARIOS FACTORES: REGRESION MULTIPLE.



GRAFICO 50. MENU RELACIONAR/REGRESION MULTIPLE

Una vez damos clic en regresión múltiple la *Ventana de Datos* se despliega y encontramos:

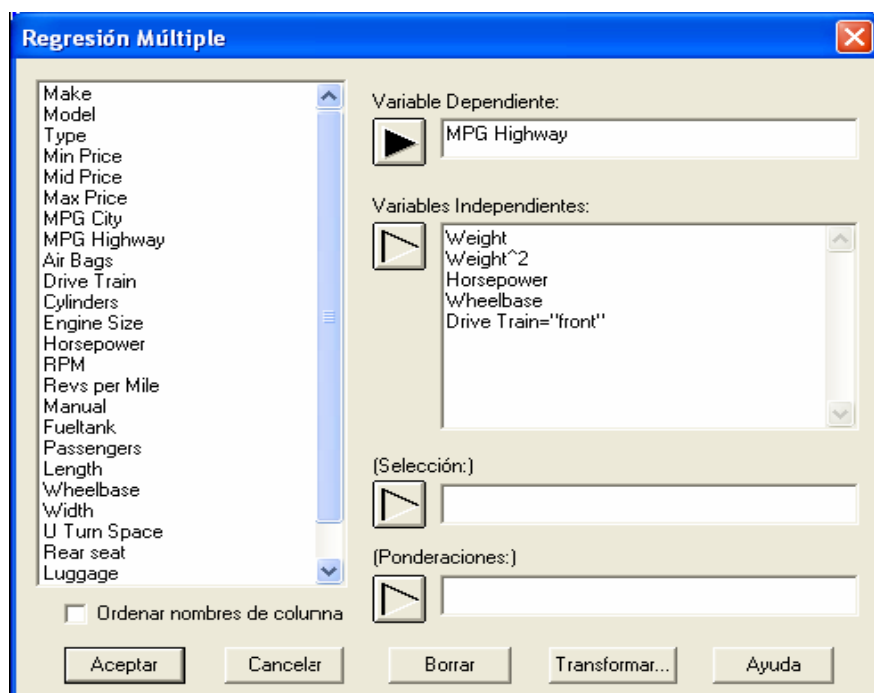


GRAFICO 51. CUADRO DIALOGO REGRESION MULTIPLE

No olvide que para completar esta tabla, es necesario tener presente que:

Variable Dependiente: columna numérica que contiene las n observaciones para la variable dependiente Y .

Variables Independientes: columnas numéricas que contienen los n valores para las variables independientes X . Pueden ser ingresados los nombres de las columnas o expresiones STATGRAPHICS.

Selección: En caso que no desee aplicar el modelo al total de datos, luego a través de esta opción podrá determinar el rango a trabajar.

Finalmente y después de dar clic en aceptar se desplegará el análisis hecho por STATGRAPHICS.

Resumen del análisis Gráfica del análisis

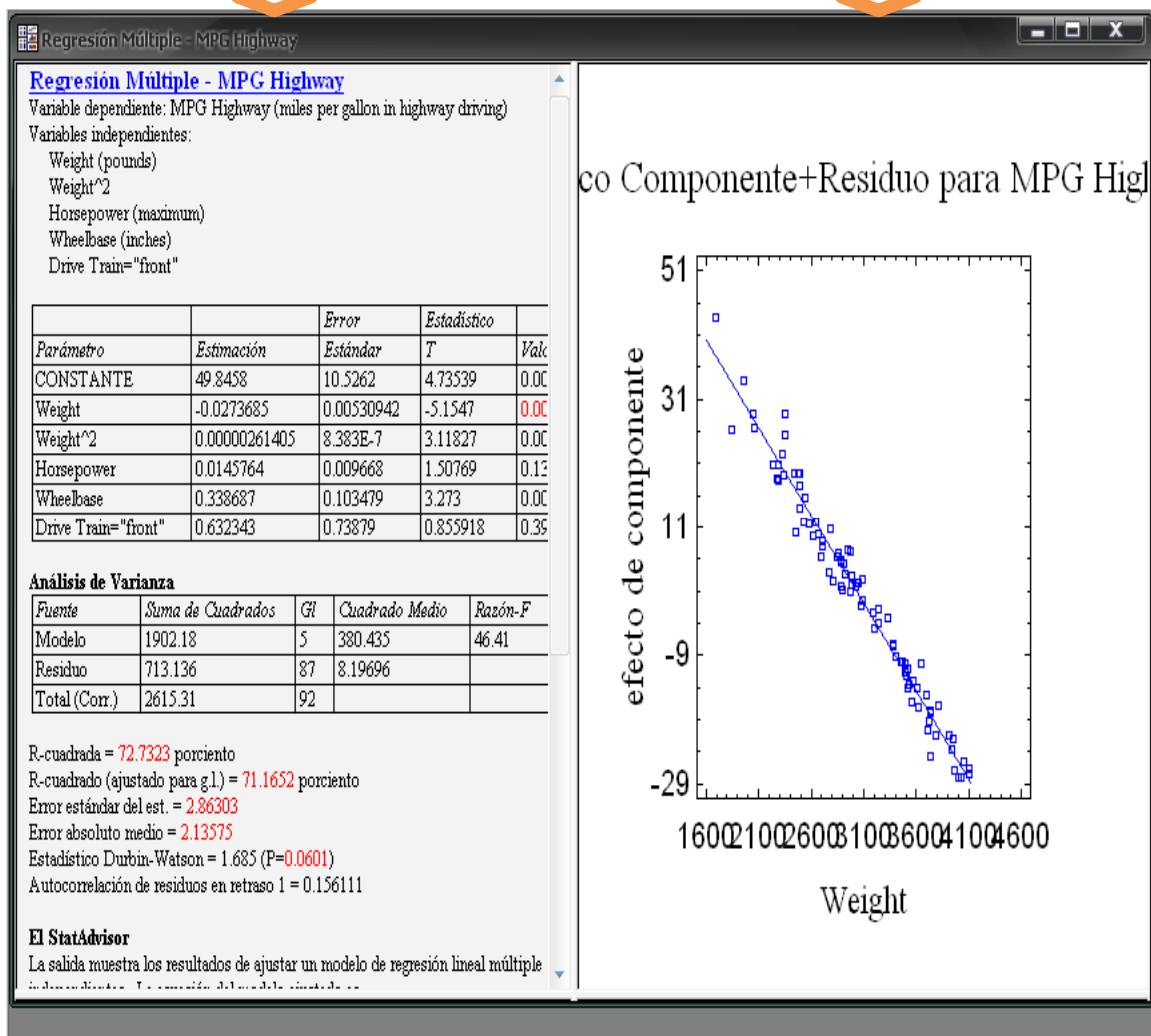


GRAFICO 52. ANALISIS REGRESION MULTIPLE

5.1 RESUMEN DEL ANALISIS

Los datos de salida que incluye en resumen de análisis son:

5.1.1 Variables: Dato gracias al cual usted identifica la variable dependiente del modelo y además conoce la forma general del mismo.

5.1.2 Coeficientes: Una vez establecido el modelo y corrido tras los parámetros establecidos el propósito fundamental es asignarle valores a los coeficientes que acompañan a las variables independientes, así en este conjunto de datos encontrará los coeficientes estimados, errores estándar, estadístico t y P-valores.

Recuerde que de forma común lo que se hace es reescribir el modelo general en términos de los coeficientes, así:

$$\text{MPG Carretera} = 49.8458 - 0.0273685*\text{Peso} + 0.00000261405*\text{Peso}^2 + 0.0145764*\text{Horsepower} + 0.338687*\text{Wheelbase} + 0.632343*\text{Drive Train}$$

5.1.3 Análisis de varianza: Descomposición de las variaciones de la variable explicada o dependiente. Además también presenta el P-valor y el F-prueba, valores empleados para determinar la significancia estadística del modelo.

5.1.4 Estadísticos: Entre los datos de salida se incluyen los estadísticos los cuales muestran el comportamiento del modelo. Entre los que calcula STATGRAPHICS encontramos.

- ☐ *R-cuadrado* - representa el porcentaje de variabilidad de Y que se ha explicado mediante el modelo ajustado de regresión, oscilando de 0% a 100%. Para los datos de muestra, la regresión ha computado alrededor de 72.7% de la variabilidad en las millas por galón.
- ☐ *R-cuadrados ajustados* – el estadístico R-cuadrado, ajustado para el número de coeficientes en el modelo. Este valor es regularmente usado para comparar modelos con diferentes números de coeficientes.
- ☐ *Error estándar de Est.* – La desviación estándar estimada de los residuos (las desviaciones alrededor del modelo). Este valor es usado para crear límites de predicciones para nuevas observaciones.
- ☐ *Error medio absoluto* – el valor absoluto del promedio de los residuos.
- ☐ *Estadístico Durbin-Watson* – una medida de correlación serial en los residuos. Si los residuos varían aleatoriamente, este valor debe ser cercano a 2. Un P-valor pequeño indica un patrón no aleatorio en los residuos

Una vez comprendidos los conceptos expuestos en el resumen del análisis, continuamos dando clic derecho sobre esta ventana, acción de la cual obtendremos la siguiente ventana:

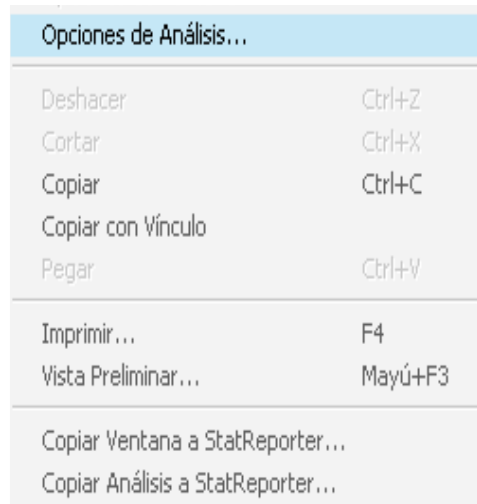


GRAFICO 53. SUBMENÚ OPCIONES DE ANALISIS

Aunque es posible que en esta imagen usted no lo note, la única opción sobre la cual usted puede trabajar es *Opción de Análisis*, la cual será descrita a continuación:

5.2 OPCION DE ANALISIS: Una vez hace clic sobre esta opción, se desplegará la siguiente ventana: *Opciones Regresión Múltiple*

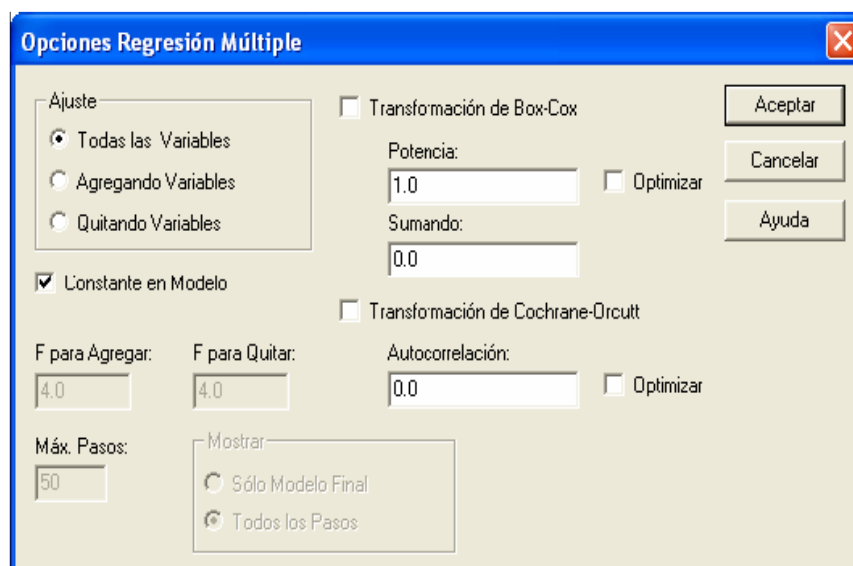


GRAFICO 54. CUADRO DIALOGO OPCIONES REGRESION MULTIPLE

A continuación se describirá en detalle cada una de las opciones que ofrece esta nueva ventana:

5.2.1 Ajuste – especifica si todas las variables independientes especificadas en el cuadro de diálogo de entrada de datos debe ser incluido en el modelo final, o si una selección de variables por pasos debe aplicarse. La selección por pasos trata de encontrar el mejor modelo que contenga sólo variables estadísticamente significativas.

5.2.2 Constante en Modelo – Si esta opción no es marcada, el término β_0 será omitido del modelo. Eliminar el término permite hacer la regresión.

5.2.3 Transformación de Box-Cox – Si se selecciona, una transformación Box-Cox será aplicada a la variable dependiente. Las transformaciones Box-Cox son un método para manejar situaciones en las que las desviaciones del modelo de regresión no tienen una varianza constante.

5.2.4 Transformación de Cochrane-Orcutt – provee un mecanismo para manejar situaciones en las que los residuos del modelo no son independientes.

5.3 GRAFICA DEL ANALISIS

En la segunda parte de respuesta dada por STATGRAPHICS tenemos la *Grafica de Análisis*, de nuevo al hacer clic derecho sobre esta, tendremos algunas opciones para cambiar la muestra grafica del resultado.

Opciones de Ventana...	
Opciones de Análisis...	
Opciones Gráficas...	
Deshacer	
Seleccionar	
Localizar	
Zoom	
Deshacer Zoom	
Restablecer Escala/Perspectiva	
Copiar	Ctrl+C
Copiar con Vínculo	
Imprimir...	F4
Vista Preliminar...	Mayú+F3
Copiar Ventana a StatGallery...	
Copiar Análisis a StatReporter...	
Guardar Gráfica...	

GRAFICO 55. OPCIONES DE GRAFICA

Las opciones de Ventana y Análisis cambian dada la regresión múltiple, sin embargo la de Opciones de Grafica permanece igual siendo el conducto a través del cual titularemos, rotularemos, etc. (Para recordar en detalle diríjase a la sección Graficas).

5.4 OPCIONES DE VENTANA

Una vez da clic sobre esta posibilidad se despliega la siguiente ventana: *Opciones Grafico Efectos de Componentes*

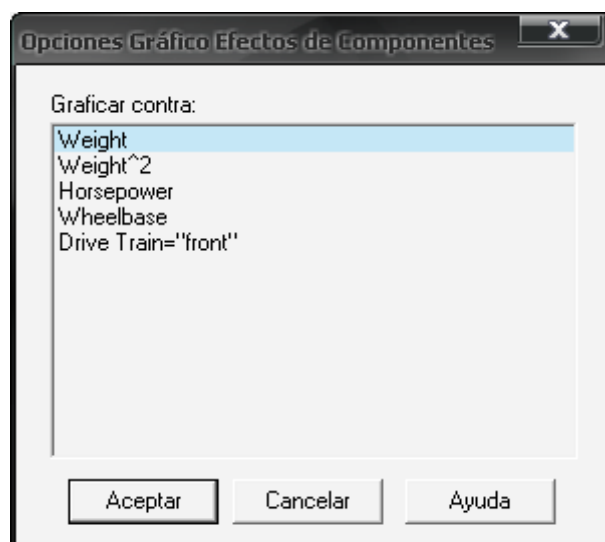


GRAFICO 56. OPCIONES GRAFICO EFECTOS DE COMPONENTES

Sobre esta ventana usted podrá modificar el contenido del eje independiente de la grafica. Se despliega el conjunto de variables independientes que explican el comportamiento de la variable dependiente y sobre este usted pobra decidir respecto de cual desea la grafica.

5.5 OPCIONES DE ANALISIS

Una vez da clic sobre esta posibilidad se despliega la siguiente ventana: *Opciones Regresión Múltiple*. Esta ventana no debe ser desconocida para usted, ya que se abordo en la sección VI.I donde se explico en detalle, es por esto que nos limitaremos a mostrarle la imagen para que recuerde las posibilidades que ofrece la misma.

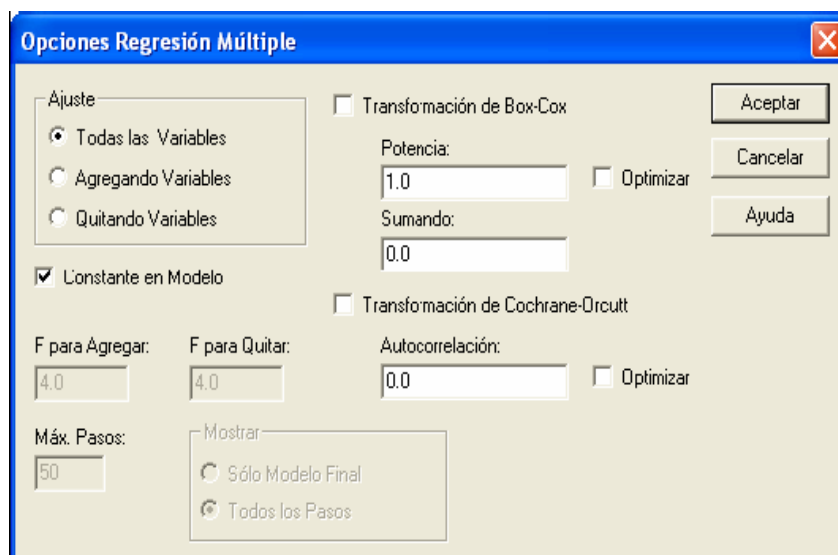


GRAFICO 57. OPCIONES REGRESION MULTIPLE

6. REGRESION NO LINEAL

6.1 REGRESION NO LINEAL CON UNA VARIABLE

Para este capítulo se trabajará sobre el StatFolio *nonlinear reg.sgp* recuerde esta es la forma de acceder al mismo

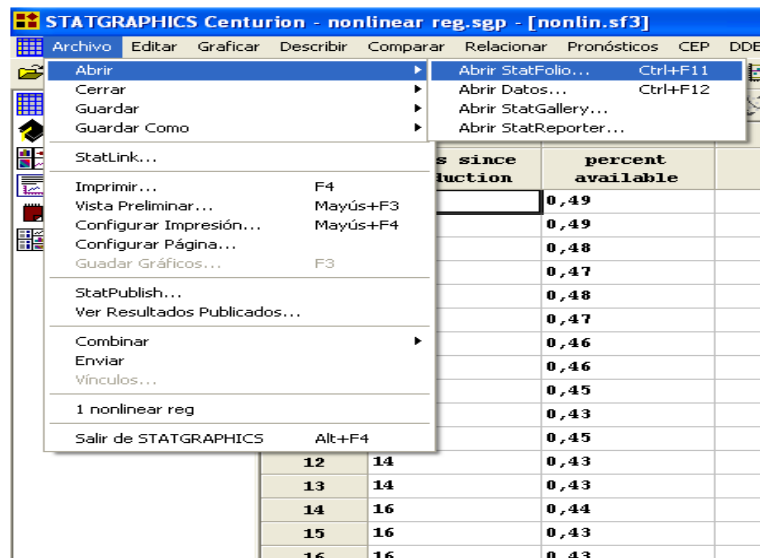


GRAFICO 58. ABRIENDO DATOS

Una vez accede al mismo lo que procede es hallar la ubicación de la regresión no lineal.

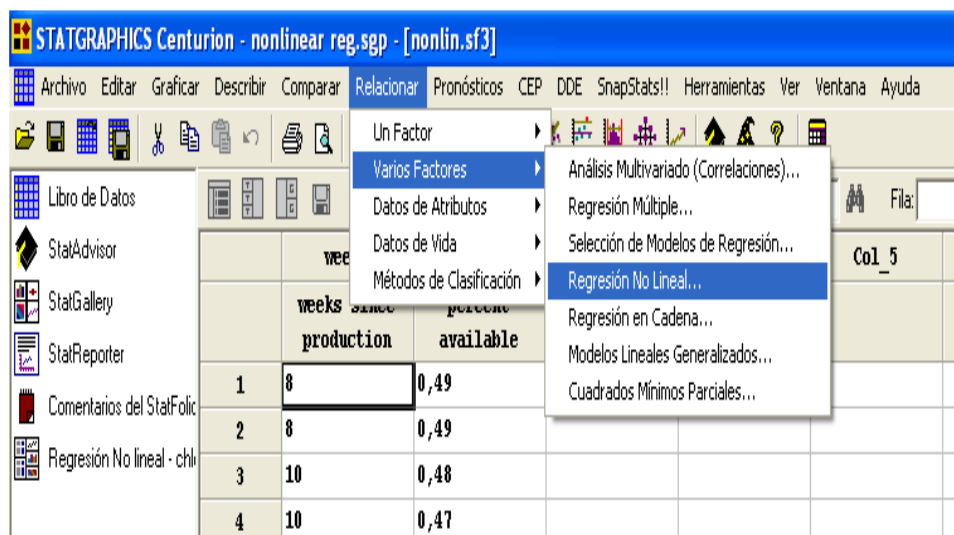


GRAFICO 59. MENU RELACIONAR/REGRESION NO LINEAL

RECUERDE: El procedimiento **Regresión No Lineal** ajusta una función especificada por el usuario relacionando una sola variable dependiente Y con una o más variables independientes X. El modelo se estima usando cuadrados mínimos no lineales.

El modelo sobre el cual trabajaremos en esta ocasión es:

$$\text{chlorine} = a + (0.49 - a)^{-b}(\text{weeks} - 8)$$

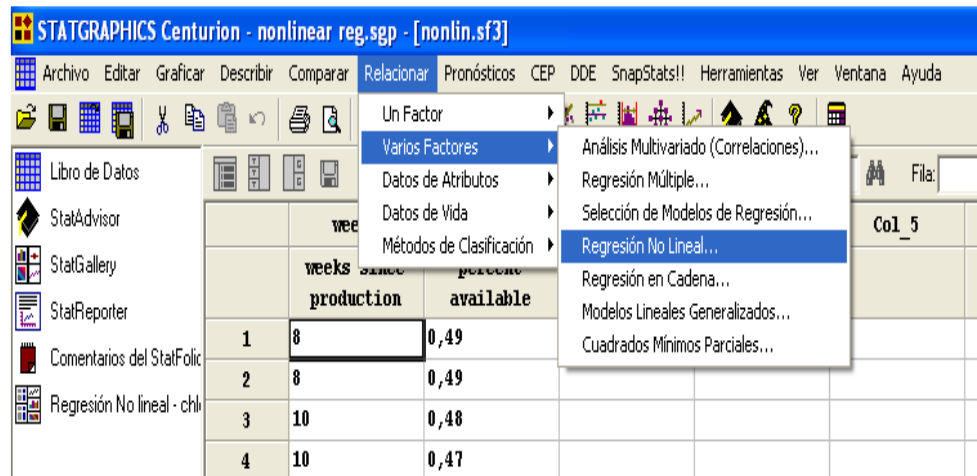


GRAFICO 60. MENU RELACIONAR/REGRESION NO LINEAL

Una vez accede al modelo no lineal se despliega la siguiente ventana, en ella debe tener presente los siguientes conceptos:

- *Variable Dependiente*: columna numérica que contiene los n valores de Y .
- *Función*: una expresión de STATGRAPHICS que representa la función a ajustar. Debe incluir uno o más nombres de columnas numéricas, que representen a las variables independientes. También puede incluir funciones tales como RAIZ o EXP. Cualquier nombre desconocido se considera que representa parámetros del modelo que tienen que ser estimados.
- *Peso*: una columna numérica opcional que contiene los pesos o ponderadores que se aplicarán al cuadrado de los residuos cuando se realice un ajuste por mínimos cuadrados ponderados.
- *Selección*: selección de un subgrupo de datos. En caso que no desee realizar el análisis con el total de datos.

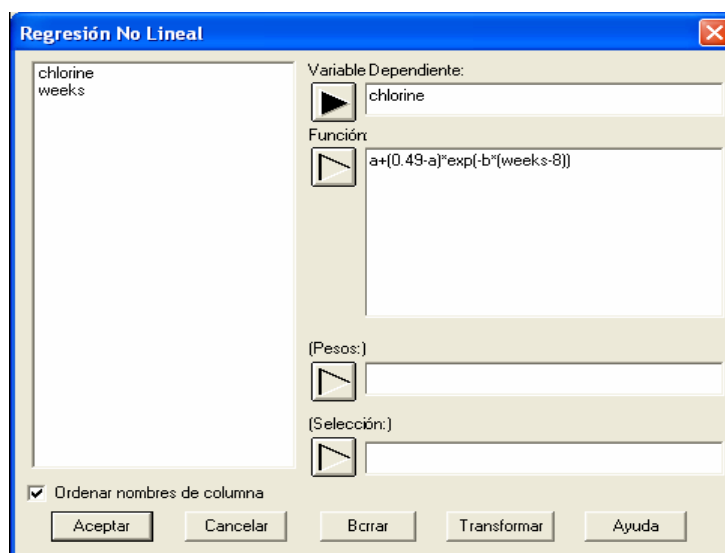


GRAFICO 61. CUADRO DIALOGO REGRESION NO LINEAL

Una vez establecidos las condiciones STATGRAPHICS, el siguiente paso es establecer los lineamientos de los parámetros establecidos en el modelo (valores presentes en la determinación del mismo pero no como un continuo de datos sino como un valor pre-establecido).

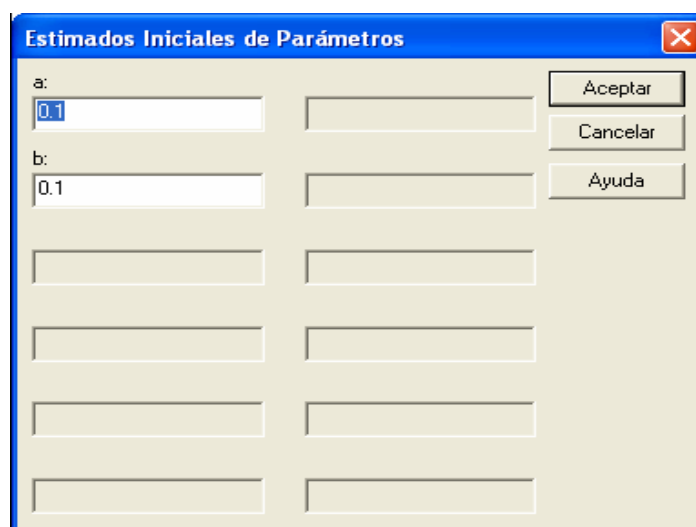


GRAFICO 62. CUADRO DIALOGO ESTABLECIMIENTO PARAMETROS

Por ejemplo en este caso, tenemos dos parámetros, a y b. Cada uno lo hemos hecho valer 0.1 (Recuerde que en este caso, como en el caso de los intervalos de confianza a mayor valor de los parámetros mayor margen de error, es decir 0.1 implica 0.9 de confiabilidad). Finalmente y tras aceptar en esta última ventana de dialogo obtenemos el análisis:

Resumen del análisis **Gráfica del análisis**



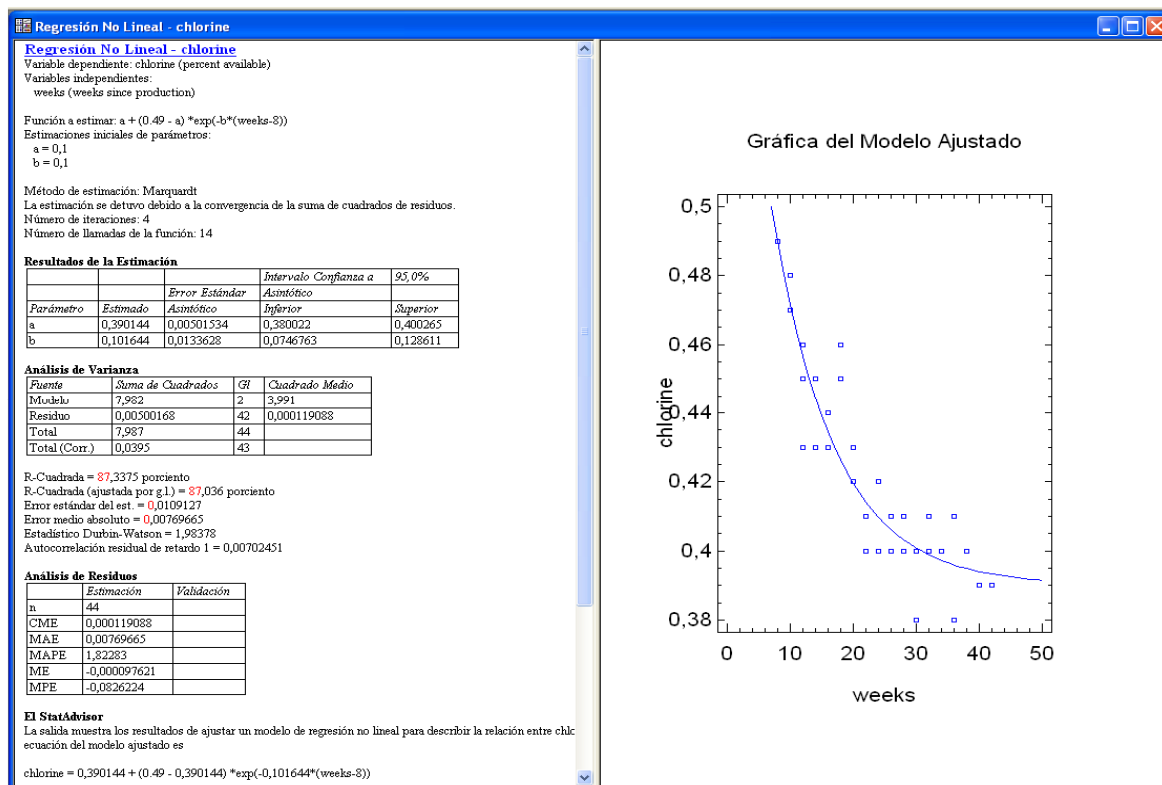


GRAFICO 63. ANALISIS REGRESION NO LINEAL

Apreciando en detalle lo que contiene cada uno de los dos componentes del análisis de STATGRAPHICS afirmamos que:

6.2 RESUMEN DE ANALISIS

Este contiene seis elementos principales de su interés:

6.2.1 Resumen de los Datos: un resumen de los datos que fueron ingresados.

6.2.2 Función a Estimar: la función que se ha de estimar y las estimaciones iniciales de los parámetros.

6.2.3 Estadísticas de la Estimación: el método empleado en la estimación así como el número de iteraciones y llamadas de la función que se llevaron a cabo.

6.2.4 Estimaciones de los Parámetros: los parámetros estimados con sus respectivos intervalos de confianza aproximados. De esta manera intervalos que no contienen al 0 indican que el parámetro del modelo es estadísticamente significativo al nivel de confianza establecido.

6.2.5 Análisis de Varianza: Este incluye descomposición de la variabilidad de la variable dependiente Y en una suma de cuadrados del modelo y una suma de cuadrados residual o del error.

6.2.6 Estadísticas: estadísticas de resumen para el modelo ajustado, incluyendo:

- ### 6.3 GRAFICA DEL MODELO AJUSTADO:

Una vez usted da clic sobre la misma tendrá:

GRAFICO 64. OPCIONES GRAFICO MODELO AJUSTADO

6.4 REGRESION NO LINEAL CON DOS O MAS VARIABLES:

Para el presente apartado se trabajara con el siguiente conjunto de datos *nlnreact.sf6*. De nuevo y de acuerdo a la forma previamente trabajado en capítulos previos y en la sección inmediatamente anterior abra los datos y realice la regresión no lineal.

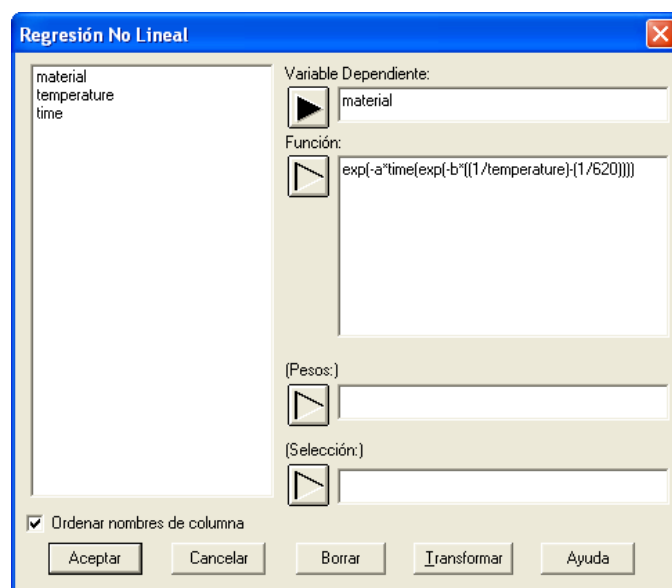


GRAFICO 65. CUADRO DIALOGO REGRESION NO LINEAL

Una vez declarado los atributos de la regresión, entonces debe determinar los valores para los parámetros a y b.

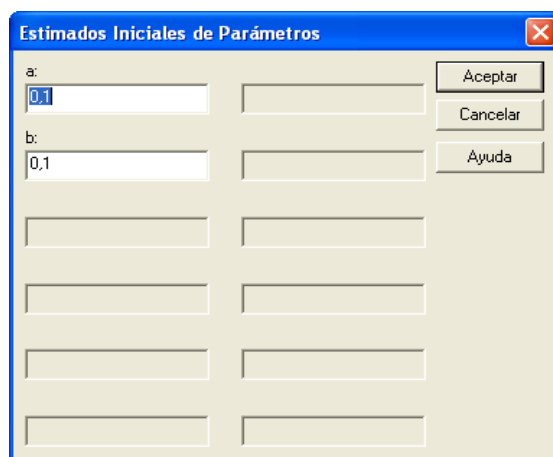


GRAFICO 66. CUADRO DIALOGO PARAMETROS

Y finalmente obtendrá.

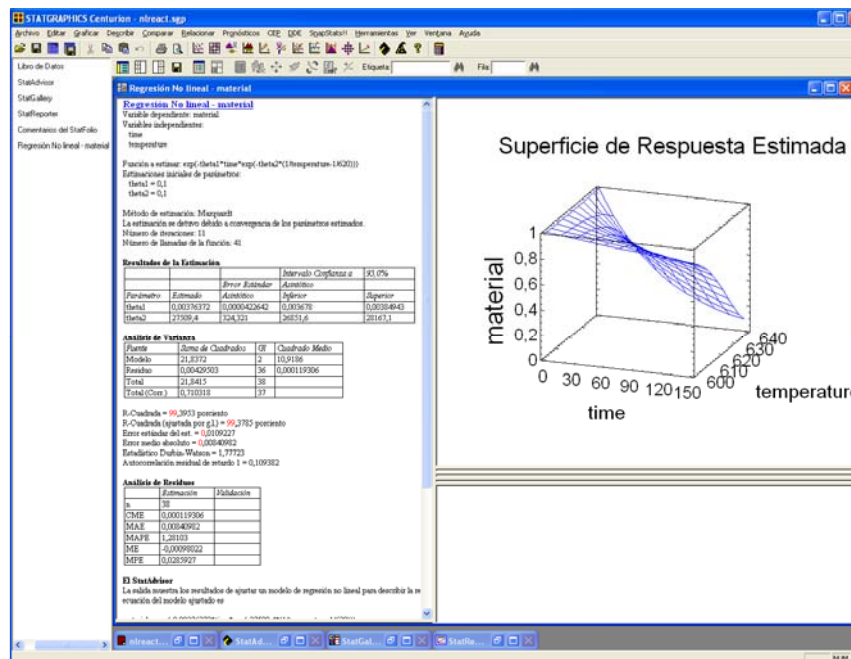


GRAFICO 67. GRAFICO REGRESION NO LINEAL

Como resulta a simple vista el grafico ha cambiado, ahora usted obtendrá un diagrama de superficie.

Opciones de Ventana...
Opciones de Análisis...
Opciones Gráficas...
Deshacer
Seleccionar
Localizar
Zoom
Deshacer Zoom
Restablecer Escala/Perspectiva
Copiar
Copiar con Vínculo
Imprimir...
Vista Preliminar...
Copiar Ventana a StatGallery...
Copiar Análisis a StatReporter...

GRAFICO 68. OPCIONES REGRESION NO LINEAL

Con este cambio también lo hacen las opciones de ventana y opciones de análisis mientras que la de grafico permanece como hasta ahora hemos visto.

6.4.1 Opciones de Ventana:

Inmediatamente da clic sobre esta se despliega el cuadro de dialogo **Opciones Grafico de Respuesta**.

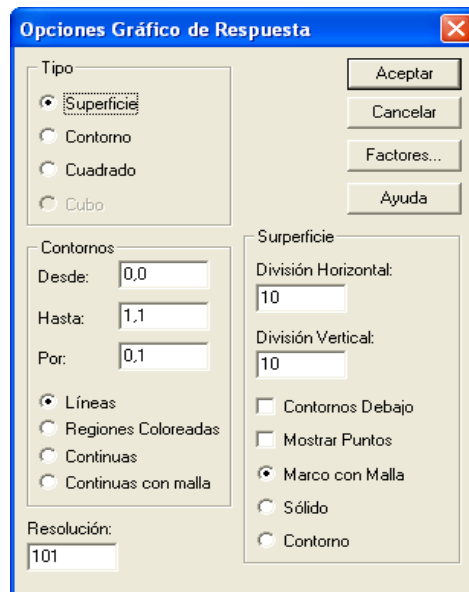


GRAFICO 69. OPCIONES GRAFICO RESPUESTA

Ahora usted podrá presentar el grafico de tres formas diferentes, gracias a la presencia de dos o más variables. A continuación se explica en detalle cada uno de los componentes de la caja de dialogo *Opciones Gráfico de Respuesta*.

- **Tipo:** escoja entre un *Gráfico de Superficie 3-D*, donde la altura de la superficie representa el valor de Y versus dos variables independientes cualesquiera; un *Gráfico de Contorno 2-D*, donde las líneas o regiones de color representan el valor de Y como una función de dos variables independientes cualesquiera; un *Gráfico Cuadrado 2-D*, donde el valor predicho de Y se muestra a diferentes combinaciones de 2 variables independientes; o un *Gráfico Cúbico 3-D*, en el cual el valor predicho de Y se muestra a diferentes combinaciones de 3 variables independientes.
- **Contornos:** los límites y espaciamiento de las líneas de contorno o regiones. Los contornos pueden dibujarse como *Líneas* sólidas representando un solo valor de Y , *Regiones Coloreadas* representando intervalos, o usando un rango *Continuo* de colores.
- **Resolución:** el número de divisiones a lo largo de cada eje sobre los cuales se grafica el valor de Y .
- **Superficie:** para un gráfico de superficie, el número de divisiones a lo largo de cada eje entre las líneas empleadas para dibujar la superficie.
- **Factores:** presione este botón para seleccionar los factores a graficar. Aparecerá una caja de diálogo similar a la descrita para el *Gráfico del Modelo Ajustado*.

6.4.2 Opciones de Análisis

Opciones de Regresión No Lineal

Estimación

Criterio de Paro 1:
0.00001

Criterio de Paro 2:
0.0001

Iteraciones Máximas:
30

Máximo de Llamadas de Función:
200

Nivel de Confianza:
95.0

Método

☒ Marquardt
☐ Gauss-Newton
☐ Descenso Máximo

Parámetro Marquardt

Valor Inicial:
0.01

Factor de Escala:
20.0

Valor Máximo:
120.0

Aceptar
Cancelar
Ayuda

GRAFICO 70. CUADRO DIALOGO REGRESION NO LINEAL

A continuación se explica en detalle cada uno de los componentes de este cuadro de dialogo *Opciones de regresión No lineal*.

- **Método:** método usado para estimar los parámetros del modelo. El método *Gauss-Newton* usa una técnica de linealización que ajusta una secuencia de modelos de regresión lineal para localizar la mínima suma de cuadrados de los residuos. El método *Descenso de Mayor Pendiente* sigue el gradiente de la superficie de la suma de cuadrados de los residuos. El método de *Marquardt*, selección por omisión, es un rápido y confiable término medio entre los otros dos.
- **Criterio de Paro 1:** Se asume que el algoritmo ha convergido cuando el cambio relativo en la suma de cuadrados de los residuos de una iteración a la siguiente es menor que este valor.
- **Criterio de Paro 2:** Se asume que el algoritmo ha convergido cuando el cambio relativo en todas las estimaciones de los parámetros de una iteración a la siguiente es menor que este valor.
- **Iteraciones Máximas:** La estimación se detiene si no se alcanza la convergencia dentro de este número de iteraciones.
- **Máximo de Llamadas de la Función:** La estimación se detiene si no se alcanza la convergencia cuando la función que se está ajustando ha sido evaluada este número de veces. Se realizan múltiples evaluaciones de la función durante cada iteración.
- **Parámetro Marquardt:** La magnitud del parámetro Marquardt controla la extensión con que participa cada método en el balance entre ambos. Para detalles sobre el algoritmo Marquardt, vea Box, Jenkins y Reinsel (1994).

- **Nivel de Confianza:** el porcentaje usado para calcular los intervalos de confianza asintóticos para los coeficientes del modelo.

7. ANALISIS DE DATOS: ANALISIS ESTADISTICO

7.1 RESUMEN ESTADISTICO Y GRAFICO DE CAJA DE BIGOTES:

Un problema muy común en estadística es el de analizar una muestra de n observaciones tomadas de una población única. Para abordar el problema se trabajara sobre el conjunto de datos *bodytemp.sf3* que reúne las temperaturas corporales del cuerpo humano. Este conjunto de datos lo ofrece Statgraphics.

Recuerde que para abrir un conjunto de Datos el procedimiento en STATGRAPHICS es el siguiente: ARCHIVO – ABRIR – ABRIR DATOS

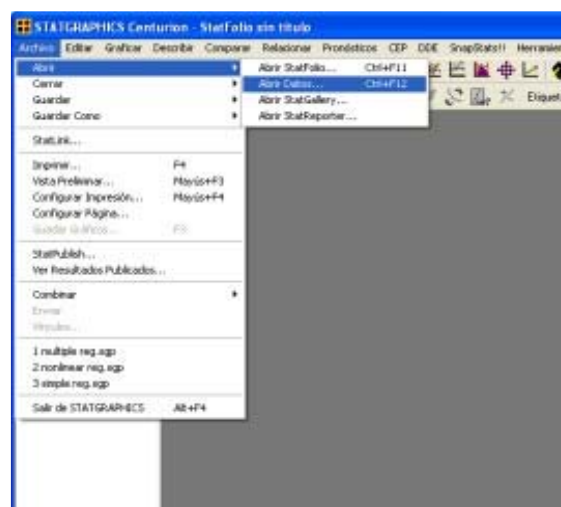


GRAFICO 71. ABRIENDO DATOS

Una vez seleccionada esta opción, se abrirá una caja de dialogo sobre la cual encontrara todos los posibles archivos de datos, sobre esta busque *bodytemp.sf3* y allí procederá a realizar el análisis de la muestra de datos.

Para realizar el análisis estadístico de una muestra usted debe proceder de la siguiente forma: DESCRIBIR – DATOS NUMERICOS – ANALISIS DE UNA VARIABLE, tal y como lo indica la figura.

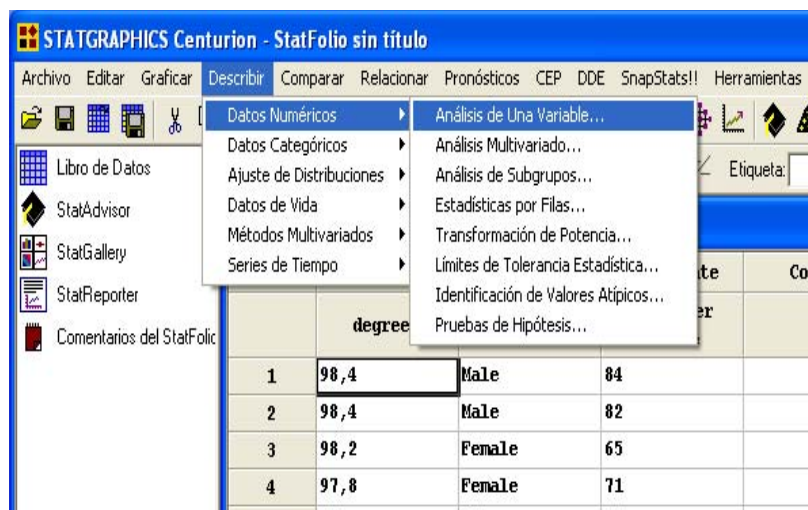


GRAFICO 72. MENU DESCRIBIR/ANALISIS DE UNA VARIABLE

Una vez escoge realizar el análisis de una variable, STATGRAPHICS le mostrara el siguiente cuadro de dialogo.

Lista de Variables Identificación

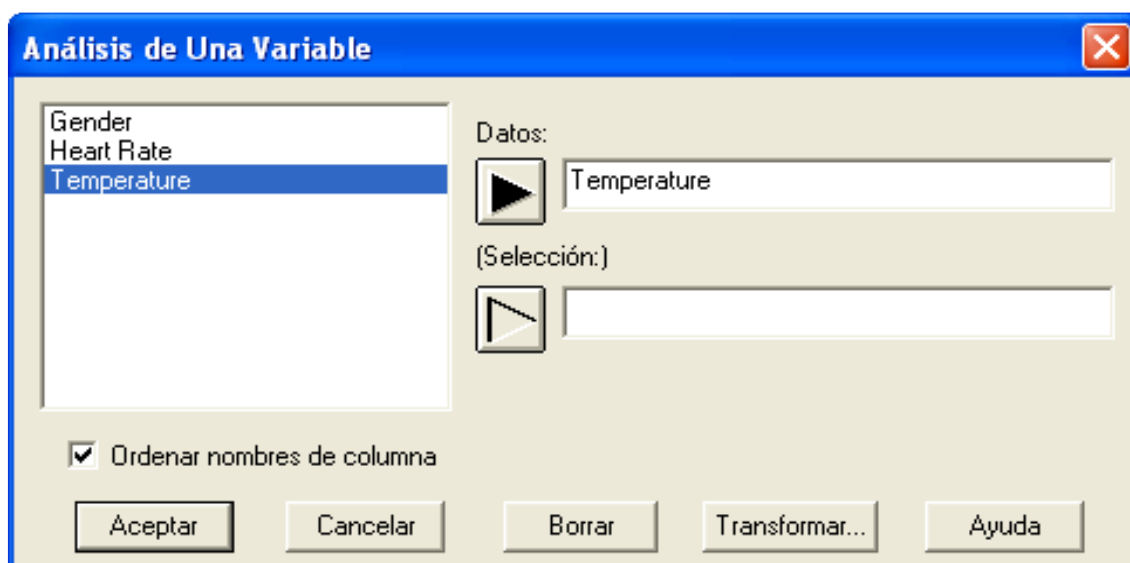


GRAFICO 73. CUADRO DIALOGO REGRESION NO LINEAL

El cuadro de dialogo “Análisis de una Variable” le muestra un listado de todas las variables involucradas en su base de datos; y además le pide que se realice un proceso de identificación respecto de la variable que usted va a trabajar. En este caso, seleccionamos la variable *Temperatura* como objeto de estudio de este análisis simple estadístico.

Una vez ha especificado su variable objeto le da clic en Aceptar, Así se desplegara el cuadro de Resultados, que para el presente caso resulto como se muestra a continuación:

Resumen del análisis Gráfica del análisis

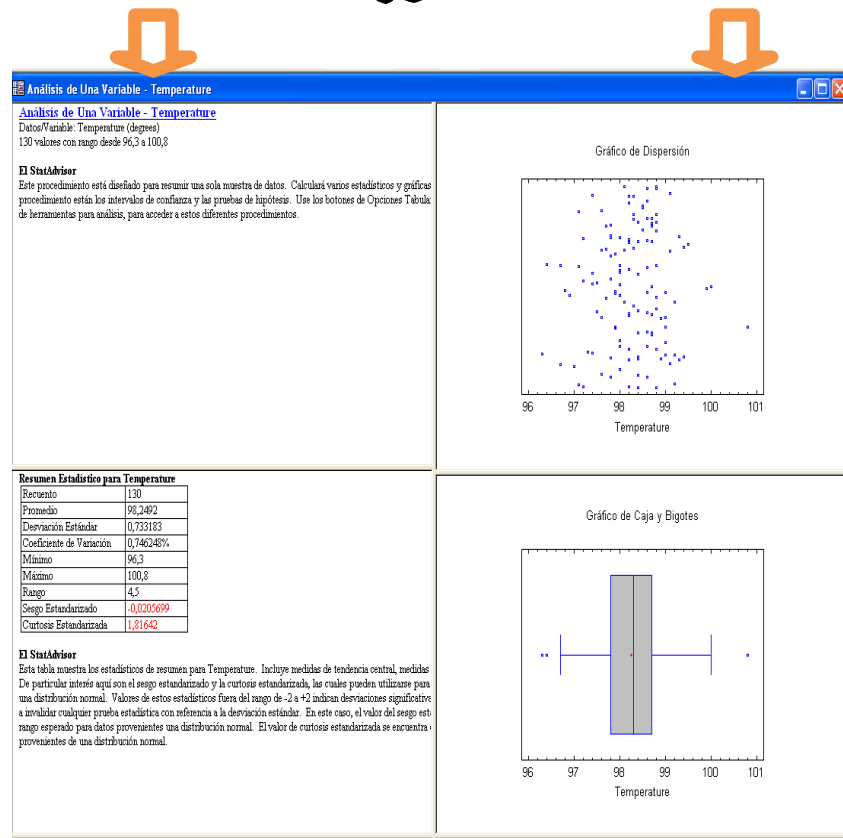


GRAFICO 74. RESULTADOS ANALISIS UNA VARIABLE

El panel superior izquierdo indica que la muestra contiene $n = 130$ valores que van desde los 96.3 hasta los 100.8 grados. El panel superior derecho muestra un gráfico de dispersión, con los puntos dispersados aleatoriamente en dirección vertical. Los paneles inferiores muestran un resumen estadístico y un gráfico de caja y bigotes.

7.1.1 Modificar los resultados

Una vez obtenidos los resultados usted podrá modificarlos a través del clic derecho sobre alguna de las cuatro ventanas de las que se compone el cuadro de resultados.

Resumen estadístico para temperatura: Este es el nombre que recibe la tabla en el panel inferior izquierdo que muestra varios estadísticos de la muestra. Los valores encontrados se muestran a continuación

Resumen Estadístico para Temperature	
Recuento	130
Promedio	98.2492
Mediana	98.3
Desviación Estándar	0.733183
Coefficiente de variación	0.746248%
Mínimo	96.3
Máximo	100.8
Rango	4.5
Cuartil Inferior	97.8
Cuartil Superior	98.7

GRAFICO 75. RESUMEN ESTADÍSTICO DE VARIABLE

NOTA TEORICA: Recordemos brevemente la definición de algunos de los estadísticos que presenta STATGRAPHICS.

Recuerde que una suposición común para medir los datos es que estos vienen de una distribución Normal o de Gauss, estas también llamadas de curva tipo campana. Los datos de una distribución normal son completamente descritos por dos estadísticos: **La**

media o promedio $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$

La desviación estándar: $S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{x})^2}{n-1}}$

La media y la desviación estándar de una muestra la describen completamente solo si viene de una distribución normal. Dos estadísticos que pueden ser usados para revisar esta suposición son el sesgo estandarizado y la curtosis estandarizada. Estas medidas estadísticas dan forma a:

- **Sesgo** \square mide el grado de simetría o la falta de esta. Una distribución simétrica como la normal tiene una nula asimetría. Las distribuciones cuyos valores tienden a caer más por *encima* del pico que por abajo de éste, tienen un sesgo positivo. Las distribuciones cuyos valores tienden a caer más por *debajo* del pico tienen un sesgo negativo.
- **Curtosis** \square mide la forma de una distribución simétrica. Una curva normal o de campana tiene cero curtosis. Una distribución con un *pico* mayor alrededor de la media que la normal tiene curtosis positiva. Una distribución más *aplanada* tiene curtosis negativa.

Si los datos provienen de una distribución normal, tanto la asimetría estandarizada como la curtosis estandarizada deben encontrarse dentro del

rango de -2 a +2. En el caso que nos compete, parece que la distribución normal es un modelo razonable para los datos.

Se pueden añadir Estadísticos adicionales maximizando haciendo clic en el botón derecho del Mouse y seleccione *Opciones de Ventana* en el menú emergente:

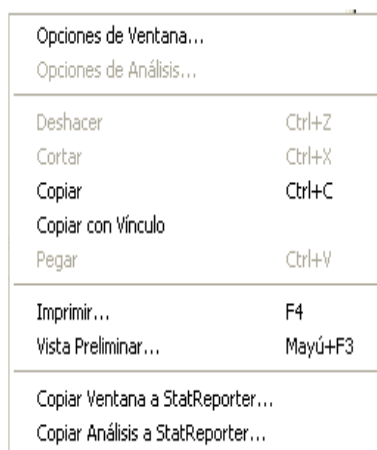


GRAFICO 76. OPCIONES

Una vez selecciona opciones de ventana se desplegara la siguiente ventana sobre la cual usted podrá seleccionar lo que mas convenga para su análisis.

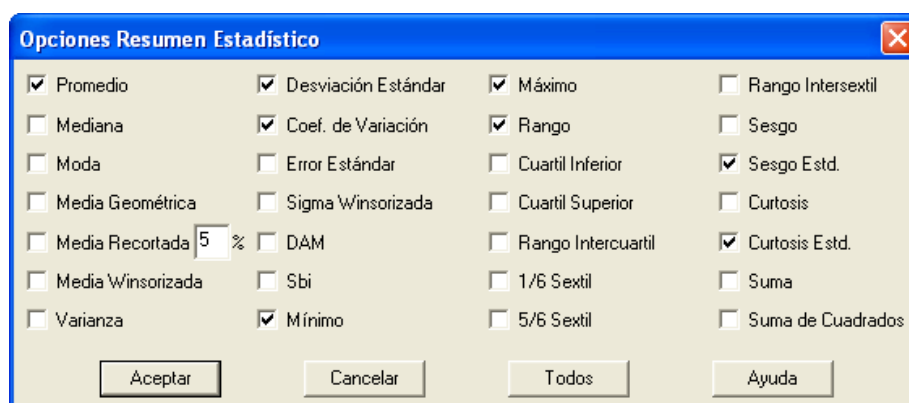


GRAFICO 77. CUADRO DIALOGO OPCIONES RESUMEN ESTADISTICO

7.1.2 Gráfico de caja y bigotes:

Recuerde que este gráfico fue creado por John Tukey, su método de elaboración es:

1. Dibujando una caja que se extiende desde el cuartil inferior hasta el cuartil superior. El 50% del total de los datos son aquellos comprendidos dentro de la caja.
2. Trazando una línea vertical en la posición de la mediana, que divide los datos a la mitad.

Si los datos vienen de una distribución normal, esta línea debe encontrarse cerca del centro de la caja.

3. Trazando un signo de más en el sitio de la media de la muestra. Cualquier diferencia substancial entre la mediana y la media indica ya sea la presencia de un dato aberrante (un valor que no proviene de la misma población que el resto) o una distribución asimétrica. En el caso de una distribución asimétrica, la media será empujada en la dirección de la cola más larga.

4. Trazando los bigotes desde cada cuartil hasta las observaciones más pequeñas y más grandes en la muestra, a menos de que algunos valores se encuentren lo bastante lejos de la caja para ser clasificados como “puntos externos”,

Es bueno tener presente que existen dos clases de puntos externos (según Turkey)

a. Puntos “muy lejanos” – puntos 3 veces mayores al rango intercuartil por arriba ó abajo de los límites de la caja. (Nota: el rango intercuartil es la distancia entre los cuartiles, equivalente al ancho de la caja). Los puntos muy lejanos son denotados mediante un símbolo (usualmente un cuadro pequeño) con un signo de más sobrepuesto. Si los datos provienen de una distribución normal y el tamaño de la muestra es como en el ejemplo, la probabilidad de que cualquier punto se encuentre lo suficientemente lejos de la caja para ser clasificado como punto muy lejano es de solo 1 en 300. A menos de que haya miles de observaciones en la muestra, los puntos muy lejanos son usualmente indicadores de datos aberrantes reales (o de una distribución no normal).

b. Puntos “externos” – puntos 1.5 veces mayores al rango intercuartil por arriba ó abajo de los límites de la caja. Los puntos externos son denotados mediante un símbolo pero sin el signo de más sobrepuesto. Incluso cuando los datos provienen de una distribución normal, la posibilidad de observar 1 ó 2 puntos externos en una muestra de $n = 100$ observaciones es del 50% y no necesariamente indica la presencia de un dato aberrante real. Estos puntos deben solo ser considerados para un análisis más detallado.

Como notara para el análisis de temperatura se muestra la existencia de un dato atípico presente en la fila #° 15. (Por favor tenga este hecho presente)

Si usted da clic derecho sobre el grafico de caja y de bigotes se desplegara el siguiente menú:

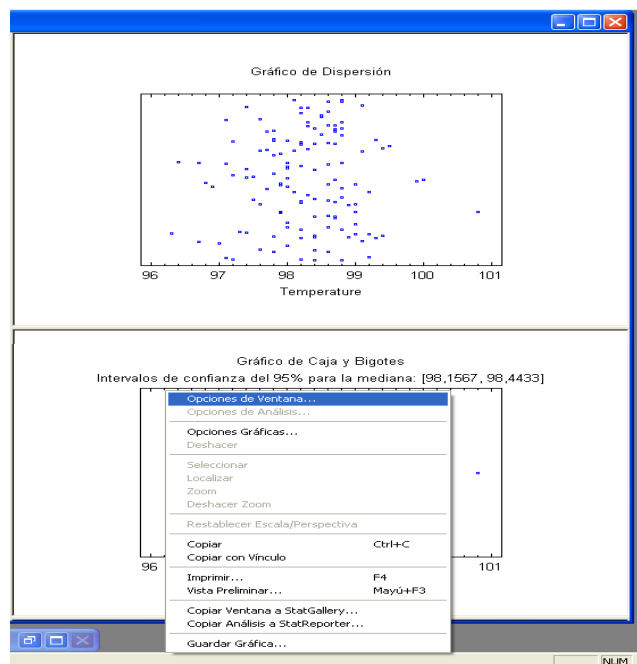


GRAFICO 78. OPCIONES GRAFICO

Seleccionando *Opciones de Ventana* se abre la siguiente caja de dialogo:



GRAFICO 79. CUADRO DIALOGO OPCIONES GRAFICO CAJA Y BIGOTE

Gracias a esa caja de dialogo: *Opciones de Grafico de Caja y Bigote* usted cambia la orientación del grafico, conforme sus necesidades.

7.1.3 Tratamiento de datos atípicos:

Recuerda en el apartado I en la sección de Grafico de Caja y Bigote, cuando se estableció la existencia de un dato atípico (ver definición en sección I), pues bien a continuación se describirá la forma adecuada como STATGRAPHICS ofrece la posibilidad de trabajar con ellos y así refinar mediante su tratamiento los datos atípicos.

El procedimiento para identificar los datos atípicos es a través del menú principal siga la ruta descrita a continuación: DATOS – DATOS NUMERICOS – IDENTIFICACION DE DATOS ATIPICOS.



GRAFICO 80. MENU DESCRIBIR/IDENTIFICACION VALORES ATÍPICOS

La respuesta que obtiene por parte de STATGRAPHICS es la siguiente:

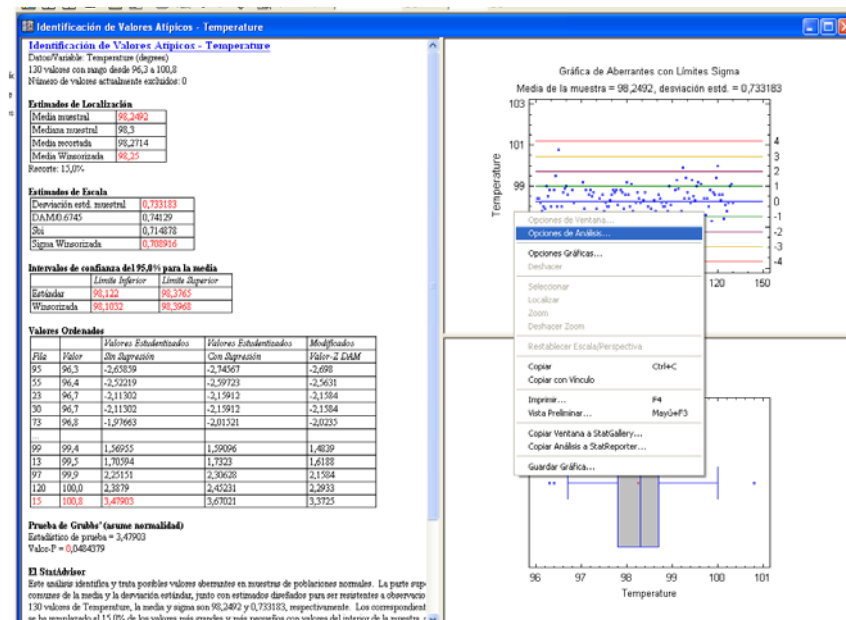


GRAFICO 81. IDENTIFICACIÓN VALORES ATÍPICOS

STATGRAPHICS identifica los valores atípicos a través del subrayado en color rojo. En este caso el valor más inusual es el de la fila #15. Este tiene un *Valor Estudentizado* de 3.479. Los valores estudentizados se calculan de acuerdo a la siguiente formula:

$$Z_i = \frac{X_i - \bar{x}}{s}$$

Un valor de 3.479 indica que una observación está 3.479 desviaciones estándar por encima de la media de la muestra, cuando la observación se incluye en el cálculo de \bar{x} y s . Los *Valores Estudentizados* indican cuantas desviaciones estándar cada observación se encuentra de la media de la muestra, cuando esa observación *no* se incluye en el cálculo.

Si no se incluye en el cálculo, la fila #15 se encuentra alejada en 3.67 desviaciones estándar.

Las observaciones a más de 3 desviaciones estándar de la muestra son inusuales, a menos que el tamaño n de la muestra sea muy grande ó que la distribución no sea normal. Una prueba formal para identificar la relación que guarda la distribución de la muestra con la distribución normal, se formaliza a través de la prueba de hipótesis.

7.1.4 Prueba de hipótesis

STATGRAPHICS ofrece la posibilidad de hacer prueba de hipótesis a la muestra para completar el análisis estadístico. A través del siguiente camino: DESCRIBIR – DATOS CATEGORICOS – PRUEBAS DE HIPOTESIS.

Recuerde que el planteamiento de las hipótesis radica en las siguientes afirmaciones:

- **Hipótesis nula:** El valor extremo más lejano proviene de la misma distribución normal que las otras observaciones.
- **Hipótesis alternativa:** El valor extremo más lejano no proviene de la misma distribución normal que las otras observaciones.



GRAFICO 82. MENU DESCRIBIR/PRUEBAS HIPOTESIS

Una vez desea plantear la prueba de hipótesis se despliega la siguiente ventana:

Pruebas de Hipótesis

Parámetro

☒ Media Normal

☐ Sigma Normal

☐ Proporción Binomial

☐ Tasa de Poisson

Aceptar

Cancelar

Ayuda

Hipótesis Nula:

0,5

Media Muestral:

0,0

Sigma Muestral:

1,0

Proporción de la Muestra:

0,5

Tasa de la Muestra:

1,0

Tamaño de Muestra:

100

GRAFICO 83. CUADRO DIALOGO PUREBAS HIPOTESIS

En ella usted debe especificar sobre que estimador desea trabajar, el nivel de confianza y el tamaño de la muestra.

RECUERDE:

La prueba de hipótesis mas común se basa en analizar la distancia que guarda la muestra con el valor de la media calculada de acuerdo a la forma normal, lo mismo ocurre para la varianza (recuerde que como desconocemos esta, la estimamos a través de sigma, su estimador para la distribución normal)

Del análisis de STATGRAPHICS usted obtiene:

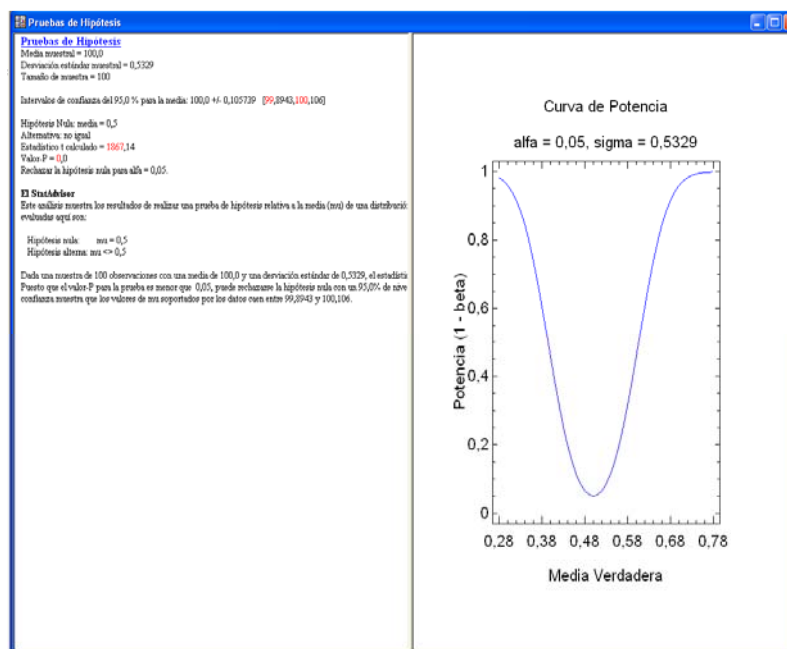



GRAFICO 84. ANALISIS PRUEBA HIPOTESIS

De este análisis, usted puede inferir que:

Si el valor-P es suficientemente pequeño, la hipótesis nula puede ser rechazada, ya que la muestra habrá sido un evento extremadamente raro. “Suficientemente pequeño” se define como menor a 0.05 y es conocido como el “nivel de significancia” de la prueba. Si existe una probabilidad menor al 5% de que la muestra tenga un resultado que indique que la hipótesis nula fué verdadera, entonces la hipótesis nula es rechazada.

En el ejemplo, el valor-P es igual a 0.0484. Debido a que el valor-P es menor a 0.05, se rechaza la hipótesis nula, concluyendo que la fila #15 es un dato aberrante comparado con el resto de la muestra.

7.1.5 Histograma

A través de este icono ubicado en la barra de herramientas  usted podrá acceder a las diferentes graficas que le ofrece el programa, una vez da clic sobre este se desplegara la siguiente ventana:

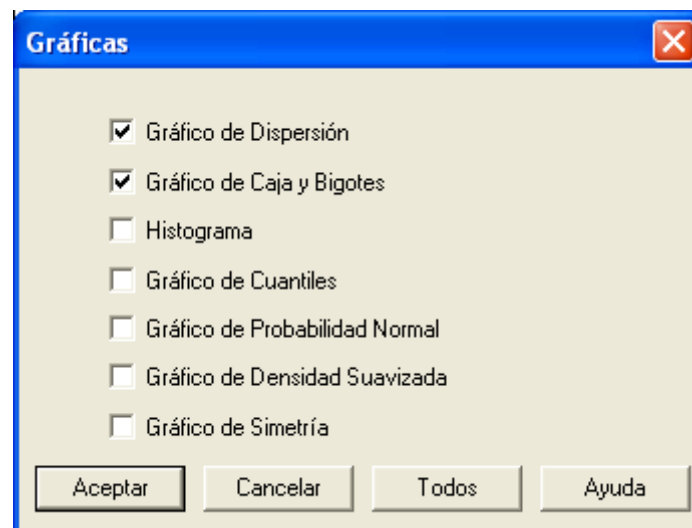


GRAFICO 85. SELECCION GRAFICAS

Seleccionando histograma obtendrá la grafica del mismo:

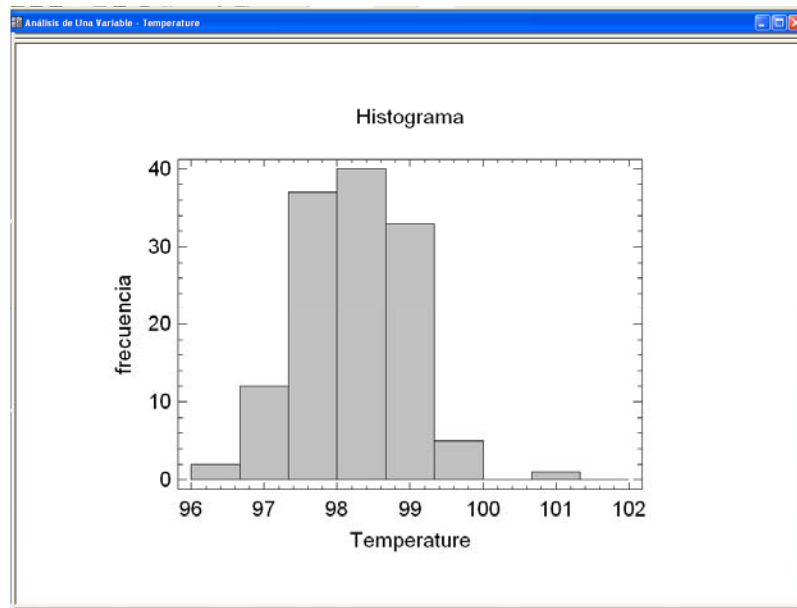


GRAFICO 86. HISTOGRAMA

8. COMPARACIÓN DE DOS MUESTRAS

Normalmente los estudios estadísticos o econométricos están constituidos por el análisis de más de una muestra, esto es grupos de datos provenientes de una población específica.

De esta manera se hace necesario establecer alguna metodología que permita establecer que posibles diferencias pueden existir entre una muestra y otra, esto a través del diseño de gráficos o planteamiento de pruebas de hipótesis.

8.1 Procedimiento ejecución comparación dos muestras

Sobre nuestros datos ya abiertos y/o exportados desde fuentes externas accedemos a través del menú principal al submenú *comparar /dos muestras /muestras independientes*

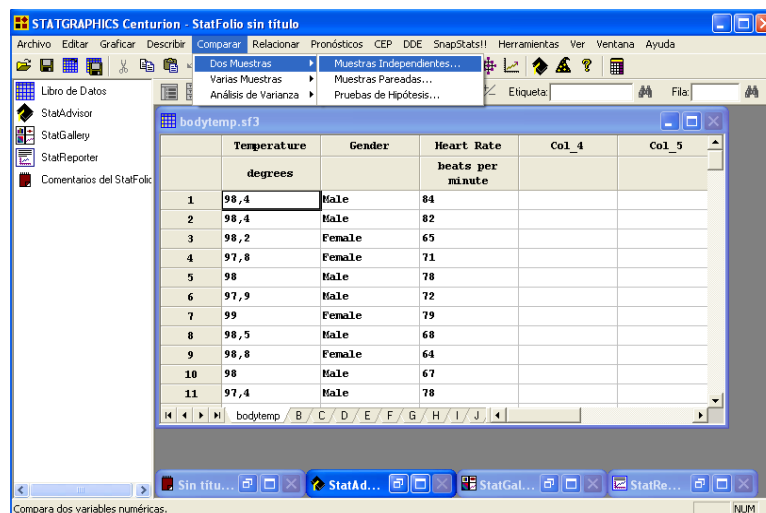


GRAFICO 87. MENU COMPARAR/MUESTRAS INDEPENDIENTES

Seguido de esto se desplegará el siguiente cuadro de dialogo

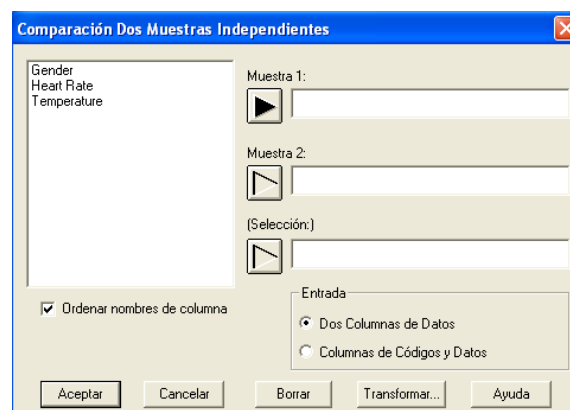


GRAFICO 88. CUADRO DIALOGO COMPARACION MUESTRAS INDEPENDIENTES

Ha de tenerse en cuenta que cada uno de los campos denominados como *Muestra* y *selección* debe contener datos numéricos de lo contrario la comparación no podrá llevarse a cabo.

Al ingresar los datos a analizar dentro de los correspondientes campos se encuentra una opción de *Entrada* la cual nos ofrece dos opciones de presentación de datos luego de ejecutar el análisis, estas son:

- Dos columnas de datos: los datos han de ser presentados en columnas separadas.
- Columnas de códigos y datos: los datos de ambas muestras se presentan en la misma columna y contigua a ella se encuentran los parámetros que las diferencian.

Luego de especificar los datos y al dar clic en aceptar para que STATGRAPHICS elabore el análisis el siguiente cuadro de desarrollo se despliega:

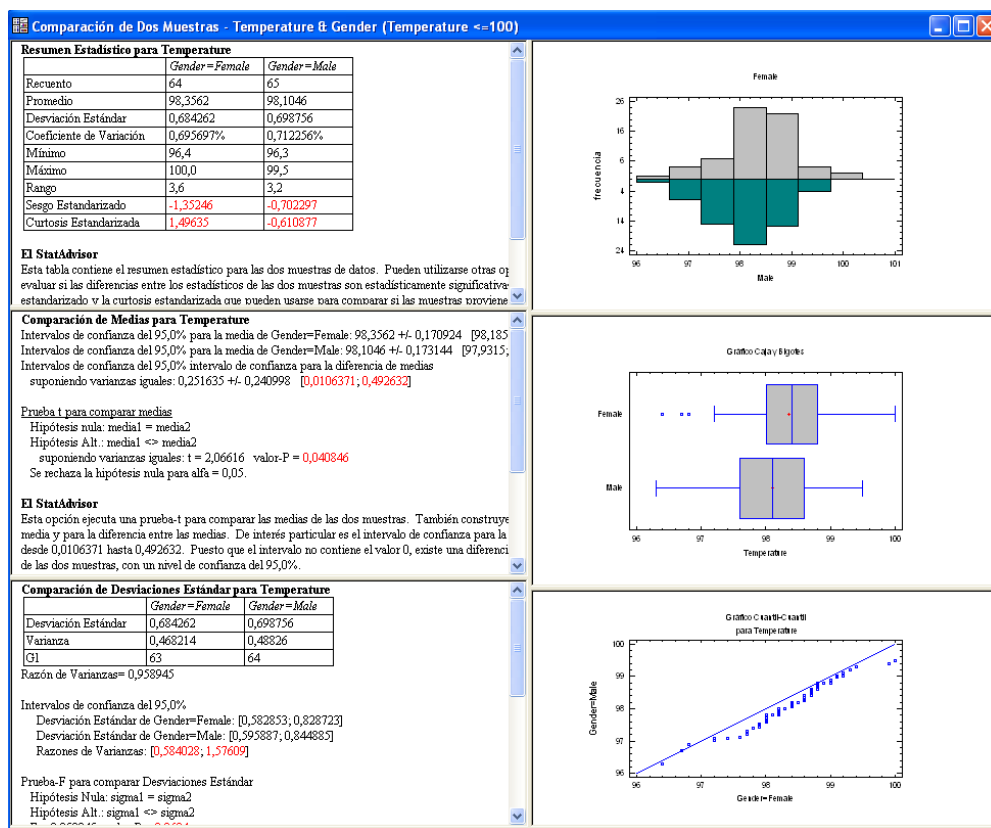


GRAFICO 89. ANALISIS COMPARACION DOS MUESTRAS

Este contiene al lado izquierdo los diferentes análisis que fueron llevado a cabo, entre ellos un resumen estadístico, una comparación de medias y desviaciones estándar; al lado derecho de la ventana de análisis se presentan las gráficas correspondientes al procedimiento desarrollado, estos son un histograma dual, un grafico de y bigotes, gráfico de cuantíles.

8.1.1 Resumen estadístico

En este recuadro se presenta un informe de los principales resultados estadísticos calculados con base en los datos ingresados, entre estos contamos con el calculo de promedio, desviación estándar, coeficiente de variación, mínimo, máximo, rango, sesgo

estandarizado, y la curtosis. El StatAdvisor funciona como asistente de explicación de procesos, esto es, ofrece una pequeña reseña sobre los resultados obtenidos y sus posibles interpretaciones.

Comparación de Dos Muestras - Temperature & Gender (Temperature <=100)		
Resumen Estadístico para Temperature		
	Gender=Female	Gender=Male
Recuento	64	65
Promedio	98,3562	98,1046
Desviación Estándar	0,684262	0,698756
Coefficiente de Variación	0,695697%	0,712256%
Mínimo	96,4	96,3
Máximo	100,0	99,5
Rango	3,6	3,2
Sesgo Estandarizado	-1,35246	-0,702297
Curtosis Estandarizada	1,49635	-0,610877

El StatAdvisor
 Esta tabla contiene el resumen estadístico para las dos muestras de datos. Pueden utilizarse otras opciones tabulares, dentro de este análisis, para evaluar si las diferencias entre los estadísticos de las dos muestras son estadísticamente significativas. De particular interés son el sesgo estandarizado y la curtosis estandarizada que pueden usarse para comparar si las muestras provienen de distribuciones normales. Valores de estos estadísticos fuera del rango de -2 a +2 indican desviaciones significativas de la normalidad, lo que tendería a invalidar las pruebas que comparan las desviaciones estándar. En este caso, ambos valores de sesgo estandarizado se encuentran dentro del rango esperado. Ambas curtosis estandarizadas se encuentran dentro del rango esperado.

GRAFICO 90. RESUMEN ESTADISTICO COMPARACION DOS MUESTRAS

8.1.2 Histograma dual

Esta gráfica nos permite comparar los resultados arrojados por el análisis, para cada una de la muestras; de esta manera sobre un eje se muestran los resultados de la muestra A, mientras que en el mismo eje invertido se encuentran los resultados de la muestra B.

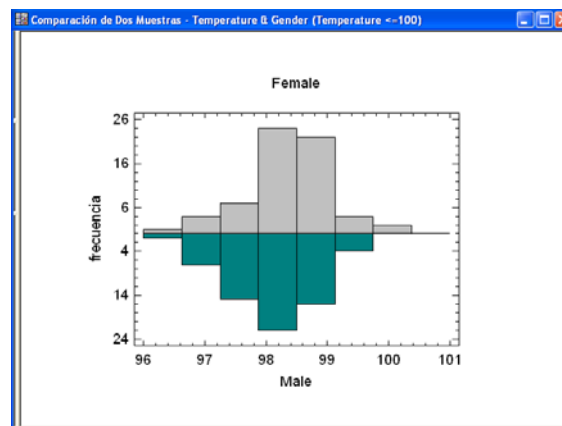


GRAFICO 91. HISTOGRAMA DUAL

8.1.3 Gráfico de caja y bigotes

Tal como se reseñó en el capítulo 7 (ver literal 7.1.2):

- **Gráfico de caja:** comprenden la mitad de los datos para cada una de las muestras.
- **Gráfico de bigotes:** valor más pequeño y más grande de cada muestra. La mediana se representa por el trazo de una línea vertical, y los signos de + señalan las media de cada muestra.

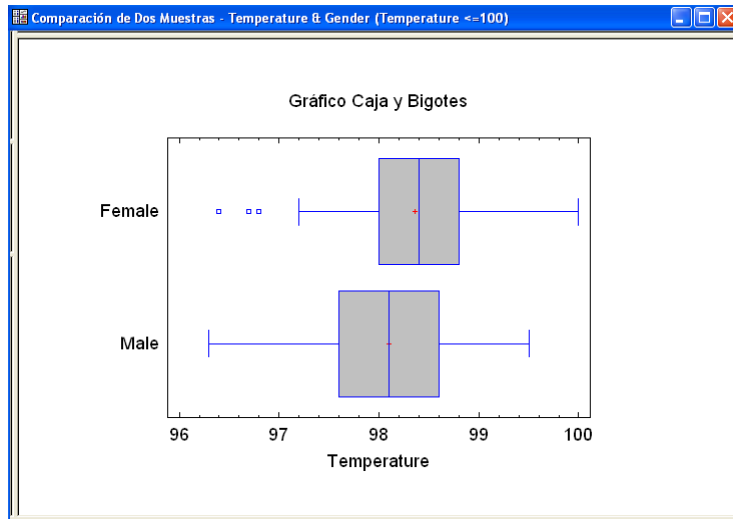


GRAFICO 92. GRAFICO CAJA Y BIGOTES

8.1.4 Comparación desviaciones estándar

Para activar esta opción, luego de haber ejecutado la comparación de muestras, sobre el icono Tablas de la barra de herramientas del análisis seleccionamos Comparación de Desviaciones Estándar

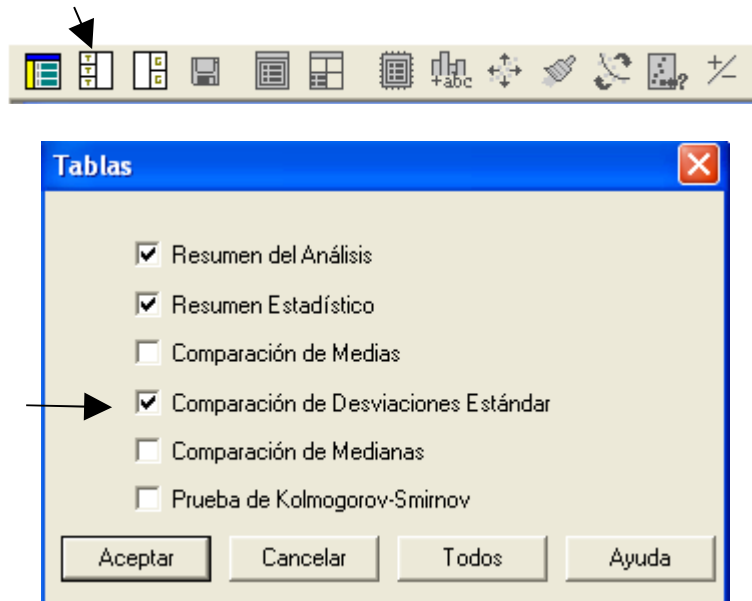


GRAFICO 93. ACCESO SELECCIÓN TABLAS

Se pretende probar la hipótesis

$$H_0 = \sigma_1 = \sigma_2$$

$$H_a = \sigma_1 \neq \sigma_2$$

Donde σ se refiere a la desviación estándar de población.

Lo cual nos permite establecer diferencias estadísticamente significativas entre las muestras estudiadas.

Comparación de Desviaciones Estándar para Temperature		
	Gender=Female	Gender=Male
Desviación Estándar	0,684262	0,698756
Varianza	0,468214	0,48826
Gl	63	64
Razón de Varianzas= 0,958945		
Intervalos de confianza del 95,0%		
Desviación Estándar de Gender=Female: [0,582853; 0,828723]		
Desviación Estándar de Gender=Male: [0,595887; 0,844885]		
Razones de Varianzas: [0,584028; 1,57609]		
Prueba-F para comparar Desviaciones Estándar		
Hipótesis Nula: $\sigma_1 = \sigma_2$		
Hipótesis Alt.: $\sigma_1 \neq \sigma_2$		
F = 0,958945 valor-P = 0,8684		
No se rechaza la hipótesis nula para $\alpha = 0,05$.		

GRAFICO 94. ANALISIS COMPARACION DESVIACIONES ESTANDAR

Razón de varianzas: intervalo de confianza al 95% para $\frac{\sigma_1^2}{\sigma_2^2}$

Valor P: Asociado a prueba F.

P-Value<0.05 : Diferencia estadísticamente significativa entre varianzas de la muestra.

P-Value>0.05: No se rechaza H_0

DEBE TENERSE EN CUENTA QUE SE ASUME DISTRIBUCIÓN NORMAL PARA CADA UNA DE LAS MUESTRAS

8.1.5 Comparación de medias

Para activar esta opción, luego de haber ejecutado la comparación de muestras, sobre el icono Tablas de la barra de herramientas del análisis seleccionamos Comparación de Medias



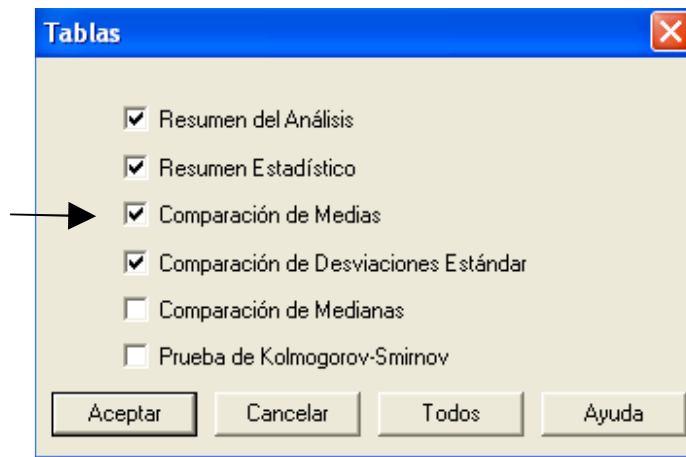


GRAFICO 95. ACCESO SELECCIÓN TABLAS

Se pretende probar la hipótesis

$$H_0 = \mu_1 = \mu_2$$

$$H_0 = \mu_1 \neq \mu_2$$

Donde μ se refiere a la media poblacional.

<p>Comparación de Medias para Temperature</p> <p>Intervalos de confianza del 95,0% para la media de Gender=Female: 98,3562 +/- 0,170924 [98,1853; 98,5272]</p> <p>Intervalos de confianza del 95,0% para la media de Gender=Male: 98,1046 +/- 0,173144 [97,9315; 98,2778]</p> <p>Intervalos de confianza del 95,0% intervalo de confianza para la diferencia de medias suponiendo varianzas iguales: 0,251635 +/- 0,240998 [0,0106371; 0,492632]</p> <p><u>Prueba t para comparar medias</u></p> <p>Hipótesis nula: media1 = media2</p> <p>Hipótesis Alt.: media1 <> media2</p> <p>suponiendo varianzas iguales: t = 2,06616 valor-P = 0,040846</p> <p>Se rechaza la hipótesis nula para alfa = 0,05.</p>
--

GRAFICO 96. ANALISIS COMPARACION MEDIAS

- *Diferencia de medias:* se suponen varianzas iguales. Intervalo de confianza para $\mu_1 - \mu_2$
- *Valor P:* Asociado a prueba t.
 P-Value<0.05: Diferencia estadísticamente significativa entre varianzas de la muestra. (En este caso se rechaza H_0)
 P-Value>0.05: No se rechaza H_0

8.1.6 Comparación de medianas

Existen casos en los que puede suponerse que los datos de la muestra han sido afectados por observaciones aberrantes; una forma de solucionar este problema radica en la realización de una comparación de medianas y no de medias

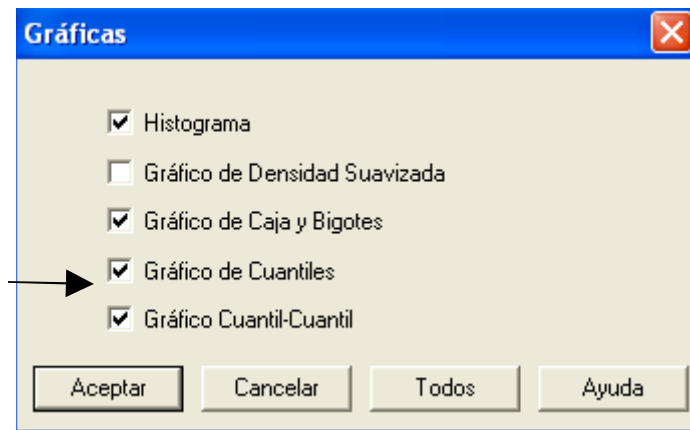


GRAFICO 99. ACCESO SELECCIÓN GRAFICAS

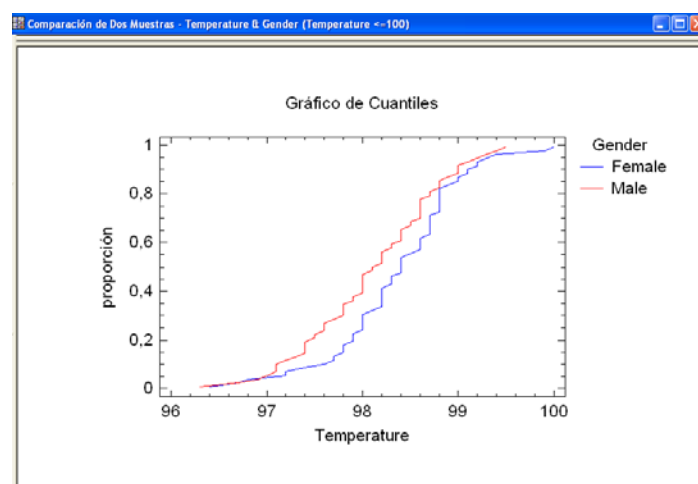


GRAFICO 100. GRAFICO CUANTILES

El gráfico de cuantiles permite establecer:

- Función de x que muestra la proporción de datos que están por debajo de un valor x .
- Separación de gráfico hacia la derecha o la izquierda evidencia diferencia de medias entre las muestras
- Pendientes disímiles indican diferencias en las desviaciones estándar entre las muestras.

8.1.8 Gráfico Cuantil-Cuantil

Para activar esta opción, luego de haber ejecutado la comparación de muestras, sobre el icono Gráficos de la barra de herramientas del análisis seleccionamos Gráfico Cuantil-Cuantil.



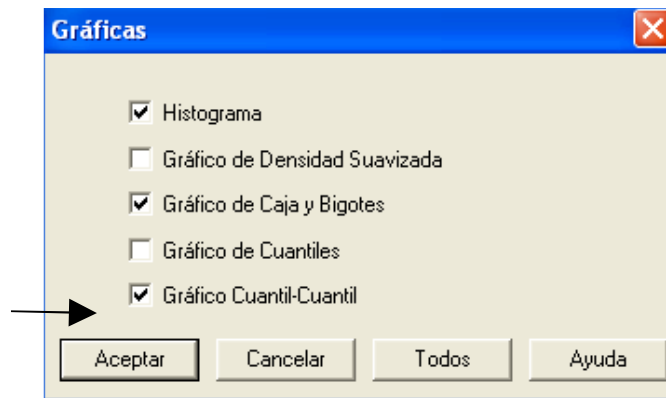


GRAFICO 101. ACCESO SELECCIÓN TABLAS

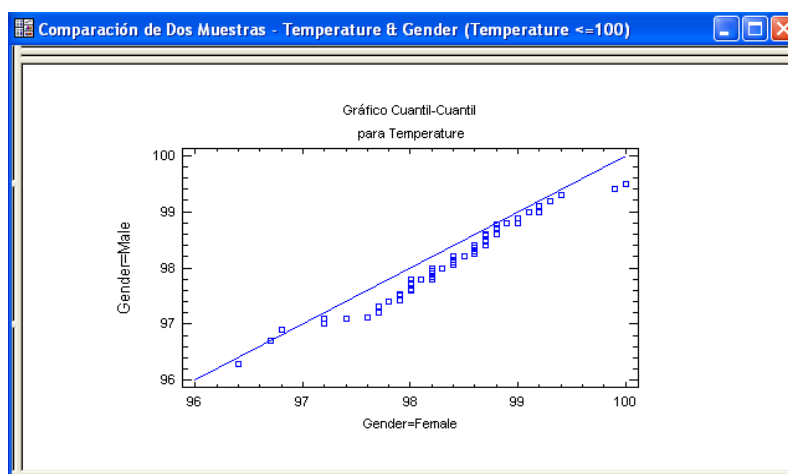


GRAFICO 102. GRAFICO CUANTIL-CUANTIL

Esta grafica contiene los puntos que corresponden con las observaciones menores de cada muestra.

INTERPRETACIONES:

- Movimiento lateral implica diferencia significativa entre los centros de las muestras.
- Muestras provenientes de población idéntica □ Puntos caen cerca de la diagonal trazada.
- Puntos divergentes sobre línea de diferente pendiente a la diagonal indican diferencia significativa entre varianzas.

8.1.9 Prueba Kolmorov-Smirnov

Esta es una prueba paramétrica que se realiza cuando no existe el supuesto de distribución normal. Consiste en establecer “la máxima distancia vertical entre las funciones de distribución acumuladas de las dos muestras, que a su vez es aproximadamente la distancia máxima entre los dos gráficos de cuantiles. Si la distancia

máxima es suficientemente amplia, entonces las dos muestras se pueden declarar provenientes de poblaciones significativamente diferentes.”³.

Para activar esta opción, luego de haber ejecutado la comparación de muestras, sobre el icono Tablas de la barra de herramientas del análisis seleccionamos Prueba Kolmorov-Smirnov.

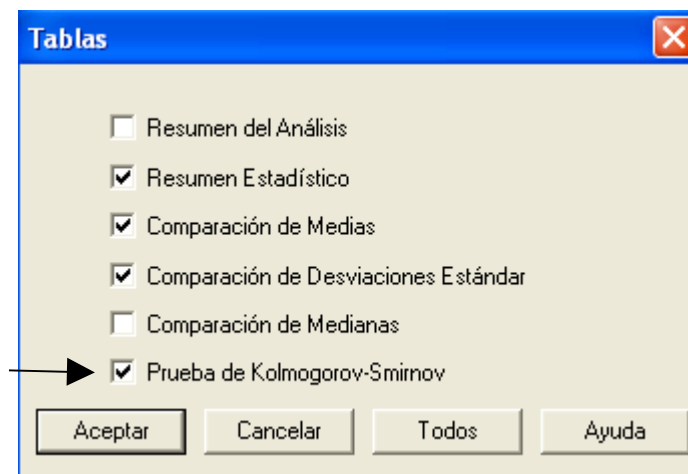


GRAFICO 103. ACCESO SELECCIÓN TABLAS

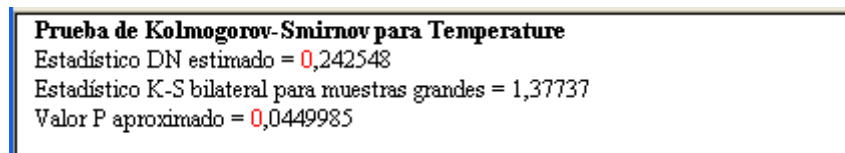


GRAFICO 104. ACCESO SELECCIÓN TABLAS

Donde:

- DN_{\square} es la máxima distancia vertical
- En este caso se evalúa de nuevo el P-Value para establecer si se rechaza o no la hipótesis nula, que se refiere a establecer si existen diferencias significativas entre las distribuciones (se rechaza la hipótesis nula).

³ Tomado de la Guía para Usuario disponible en StatGraphics Centurión XV

9. COMPARACIÓN DE MAS DE DOS MUESTRAS

En el caso en que los datos pertenezcan a mas de dos muestras su análisis conjunto será el resultado de procesos diferentes al expuesto en el capítulo anterior.

9.1 Procedimiento ejecución comparación de más de dos muestras

Sobre nuestros datos ya abiertos y/o exportados desde fuentes externas accedemos a través del menú principal al submenú *comparar /varias muestras /comparación de varias muestras*.

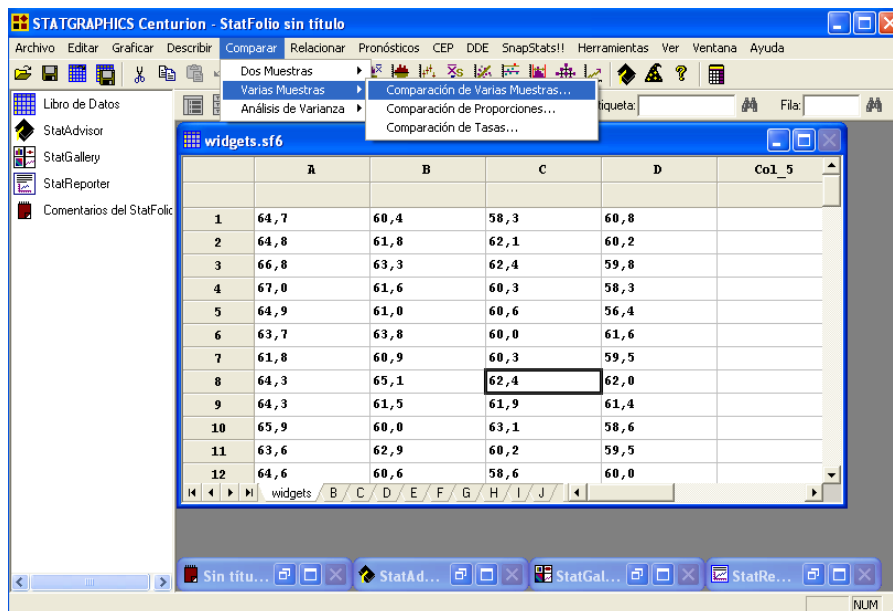


GRAFICO 105. MENU COMPARA/COMPARACION VARIAS MUESTRAS

Al seleccionar la opción indicada se despliega el siguiente cuadro de dialogo

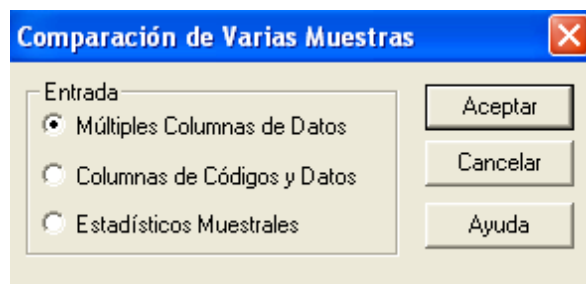


GRAFICO 106. CUADRO DIALOGO COMPARACION VARIAS MUESTRAS

Donde ha de especificarse la estructura de organización de los datos

Existen dos formas de capturar los datos de diferentes muestras dentro de la hoja de análisis de StatGraphics:

1. Datos de muestra en columna individual (*Múltiples Columnas de Datos*)
 2. Datos de todas las muestras en una única columna (*Columnas de Código y datos*)
- En este caso se selecciona *Múltiples Columnas de Datos*; seguido de esto se despliega un cuadro de dialogo que solicita el nombre de las columnas que contienen los datos.

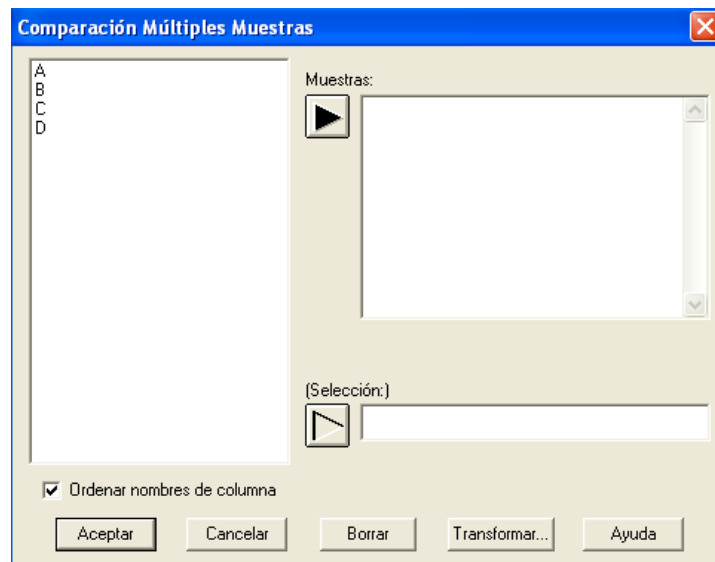


GRAFICO 107. INTRODUCCION VARIABLES COMPARACION MULTIPLES MUESTRAS

Allí seleccionamos los grupos de datos a analizar para la posterior ejecución del análisis, esto es:

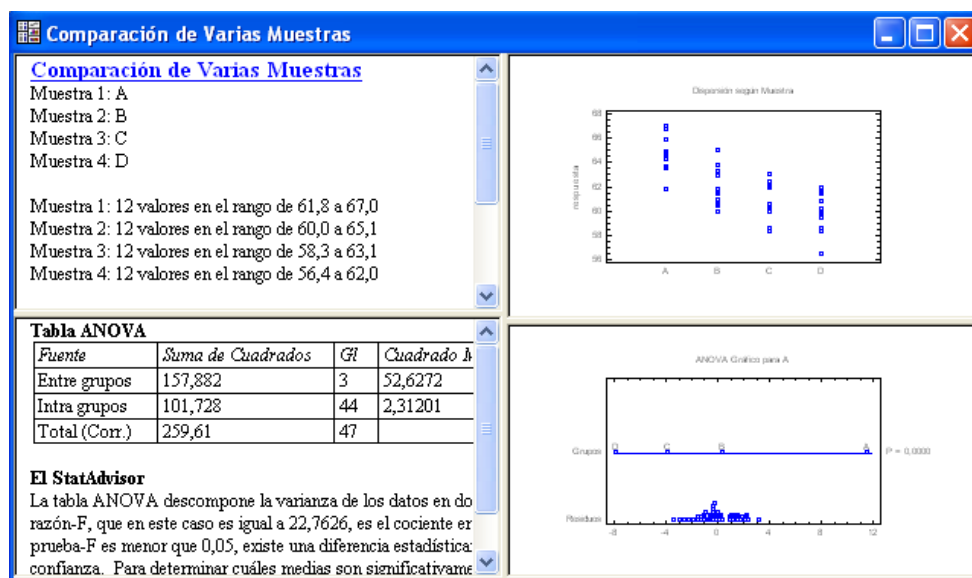


GRAFICO 108. ANALISIS COMPARACION VARIAS MUESTRAS

La tabla de análisis muestra cuatro divisiones, a saber: un resumen de las características de las muestras, una tabla ANOVA (con su correspondiente gráfico), y gráfico de dispersión para cada muestra.

9.1.1 Gráfico dispersión

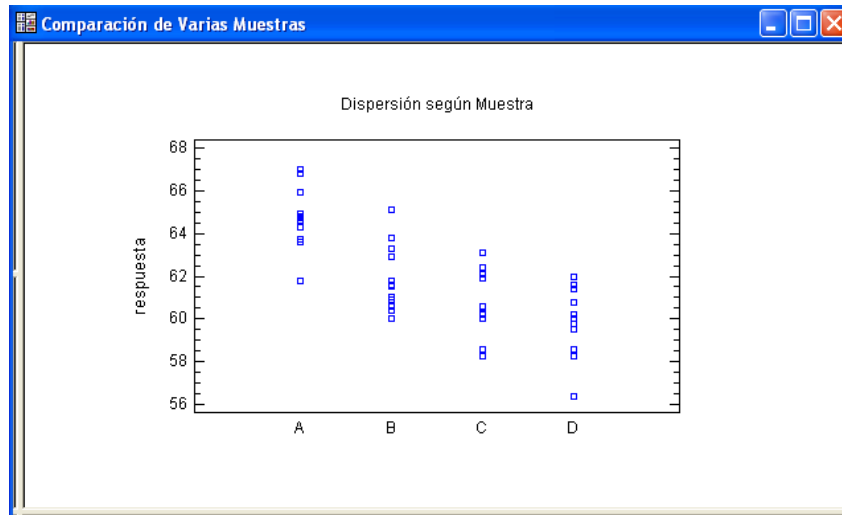


GRAFICO 109. ACCESO SELECCIÓN TABLAS

En este caso las observaciones se presentan una encima de la otra, para solucionar esta situación procedemos como se ha señalado en la sección 3.2.

DEBE TENERSE ENCUENTA QUE ESTE CAMBIO EN LA GRÁFICA NO AFECTA LOS DATOS DEL ANÁLISIS

9.1.2 Análisis de varianza

Se pretende probar la hipótesis

$$H_0 = \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_a = \exists \text{ al menos un } \mu_x \neq \mu_y$$

Donde μ se refiere a la media poblacional.

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

El StatAdvisor

GRAFICO 110. TABLA ANOVA/COMPARACION VARIAS MUESTRAS

En este caso al igual que en la sección anterior se estudia el P-Value para establecer la posibilidad de rechazo de la hipótesis nula ($P\text{-Value} < 0.05$ ☐ Rechazo H_0).

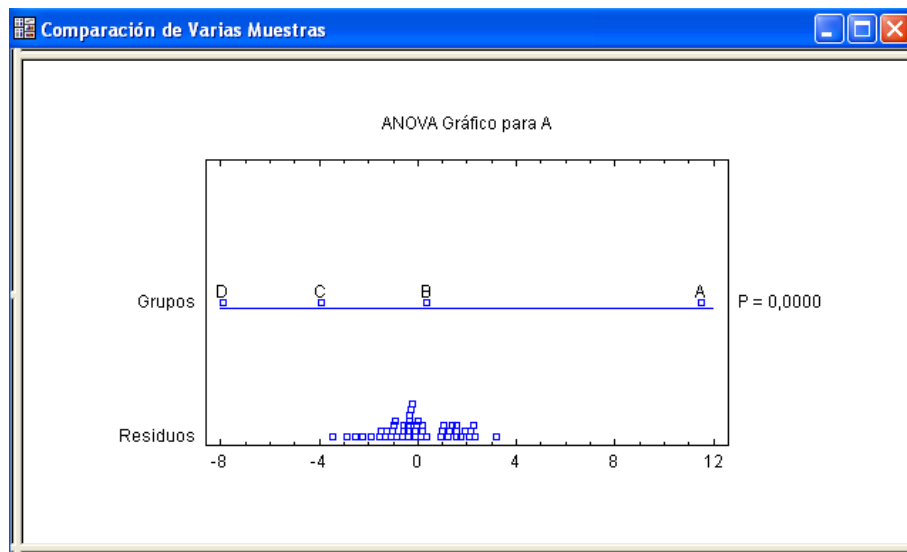


GRAFICO III. GRAFICO ANOVA/COMPARACION VARIAS MUESTRAS

Este gráfico se compone de punto que representan los residuos del análisis, esto es:

$$\text{Residuo} = \text{Observación de la muestra} - \text{Media Muestral}$$

9.1.3 Comparación Medias

Para activar esta opción, luego de haber ejecutado la comparación de muestras, sobre el icono Gráficos de la barra de herramientas del análisis seleccionamos Gráficos de Medias



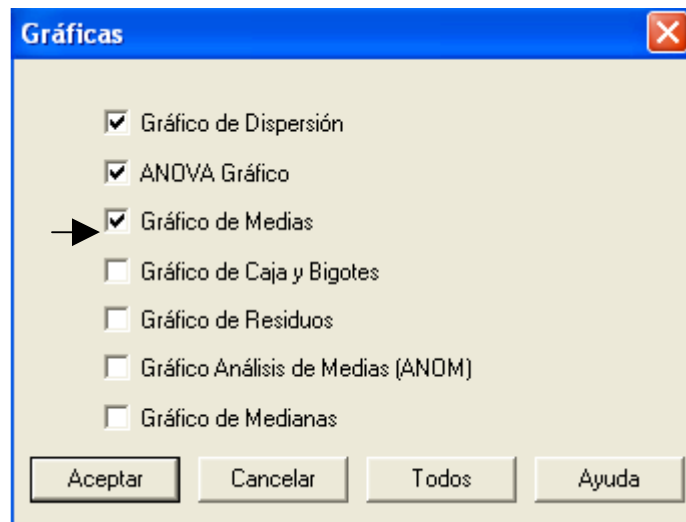


GRAFICO 112. ACCESO SELECCIÓN GRAFICAS

Al examinar el P-Value y determinar el rechazo de H_0 ha de establecerse que medias son diferentes significativamente; para esto es útil el uso del Grafico de medias.

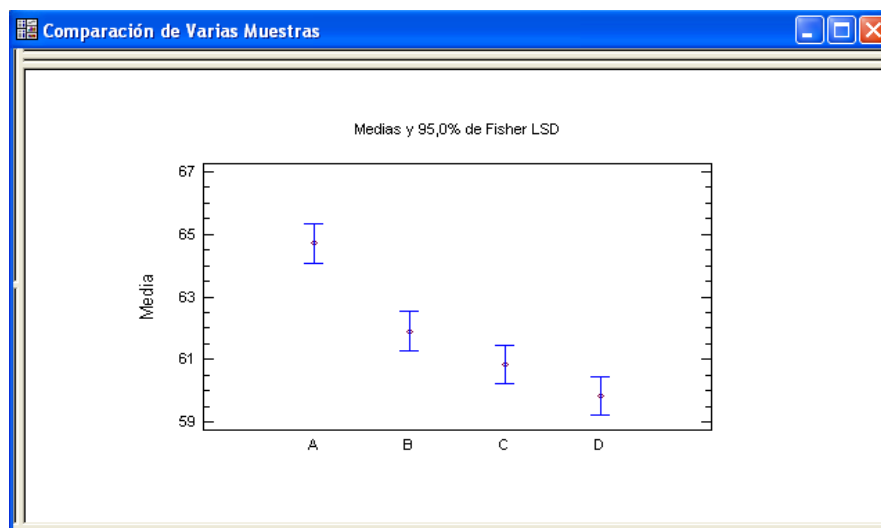


GRAFICO 113. GRAFICO MEDIAS/COMPARACIÓN VARIAS MUESTRAS

Donde se presenta la media de cada muestra rodeada por un intervalo de incertidumbre, cuya interpretación depende del tipo de intervalo elegido, esto es:

- *Intervalos LSD (Least Difference Significant) Fisher:* son escogidas un par de muestras donde se dice que sus medias son diferentes significativamente si los intervalos no se superponen en sentido vertical.
- *Intervalos HSD (Honest Difference Significant) Tuckey:* los intervalos son construidos teniendo en cuenta la tasa de error determinada a un 5%.

HA DE TENERSE EN CUENTA QUE LA SELECCIÓN DEL TIPO DE INTERVALO SE HACE SOBRE LA VENTANA DEL GRÁFICO, CLIC IZQUIERDO/OPCIONES DE VENTANA/TIPO DE INTERVALO.

De igual manera el análisis ejecutado puede mostrarse en una tabla, a la cual accedemos a través el icono Tablas de la barra de herramientas del análisis seleccionando Pruebas de Múltiples Rangos.

Comparación de Varias Muestras			
Pruebas de Múltiple Rangos			
Método: 95,0 porcentaje LSD			
	Casos	Media	Grupos Homogéneos
D	12	59,8417	X
C	12	60,85	XX
B	12	61,9083	X
A	12	64,7	X
Contraste	Sig.	Diferencia	+/- Límites
A - B	*	2,79167	1,25105
A - C	*	3,85	1,25105
A - D	*	4,85833	1,25105
B - C		1,05833	1,25105
B - D	*	2,06667	1,25105
C - D		1,00833	1,25105

* indica una diferencia significativa.

GRAFICO 114. PRUEBAS DE MULTIPLES RANGOS

La cual se interpreta de la siguiente manera:

- La primera parte de la tabla agrupa homogéneamente (no existen diferencias significativas) las muestras
- La siguiente subtabla se presenta cada par de muestras, con su respectiva diferencia y un intervalo de incertidumbre asociado a ésta. Si la diferencia excede al límite se dirá que la diferencia reportada entre pares de muestras es estadísticamente significativa.

9.1.4 Comparación Medianas

$$H_0 = \omega_A = \omega_B = \omega_C = \omega_D$$

$$H_a = \omega_A \neq \omega_B \neq \omega_C \neq \omega_D$$

Para la prueba de esta hipótesis se utilizan dos análisis (disponibles en el icono Tablas de la barra de herramientas para el análisis):

1. Prueba Kruskal-Wallis: utilizada cuando las muestras han sido ingresadas por columnas.
2. Prueba de Friedman: utilizada cuando las muestras han sido ingresadas por filas.

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

Estadístico = 27,3735 Valor-P = 0,00000491592

GRAFICO 115. RESULTADO PRUEBA KRUSKAL-WALLIS

En este caso como ya se ha venido haciendo se examina el P-Value para determinar posible rechazo de la hipótesis nula.

Un análisis gráfico radica en la construcción de un gráfico de caja y bigotes (añadiendo cambios en las gráficas que señalen las medianas, esto a través de opciones de ventana)

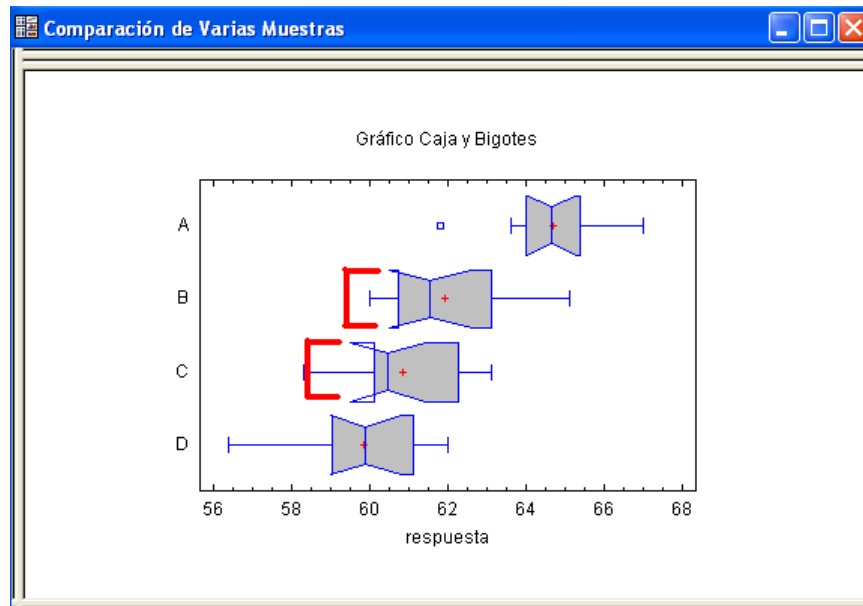


GRAFICO 116. GRAFICO CAJA Y BIGOTES

De esta manera el rango de cada cambio en la caja representa el intervalo de incertidumbre asociado a esa muestra.

Cuando los cambios en las gráficas se superponen uno encima del otro se dice que las medianas son significativamente diferentes. (Caso para la muestra b y c)

9.1.5 Comparación de desviaciones estándar

$$H_0 = \sigma_A = \sigma_B = \sigma_C = \sigma_D$$

$$H_a = \exists \text{ al menos un } \sigma_x \neq \sigma_y$$

Para activar esta opción, sobre el icono Tablas de la barra de herramientas del análisis seleccionamos Verificación de la varianza.

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

GRAFICO 117. VERIFICACIÓN VARIANZA/PRUEBA LEVENE'S

En este caso la prueba ejecutada es *Levene's*, pero pueden seleccionarse otro tipo de pruebas a través de las opciones de ventana, esto es.

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

Opciones de Ventana...
Opciones de Análisis...

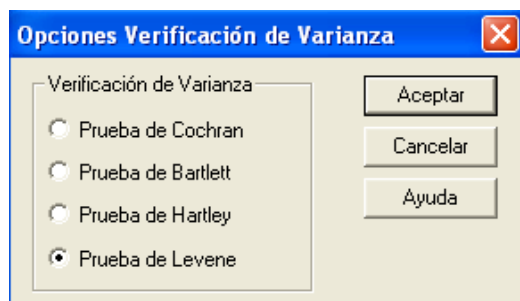


GRAFICO 118. OPCIONES VERIFICACION VARIANZA

Se debe analizar el P-Value para establecer si se rechaza o no la hipótesis nula.

9.1.5 Gráfico de residuos

Para activar esta opción, sobre el icono Gráficos de la barra de herramientas del análisis seleccionamos Gráficos de Residuos.

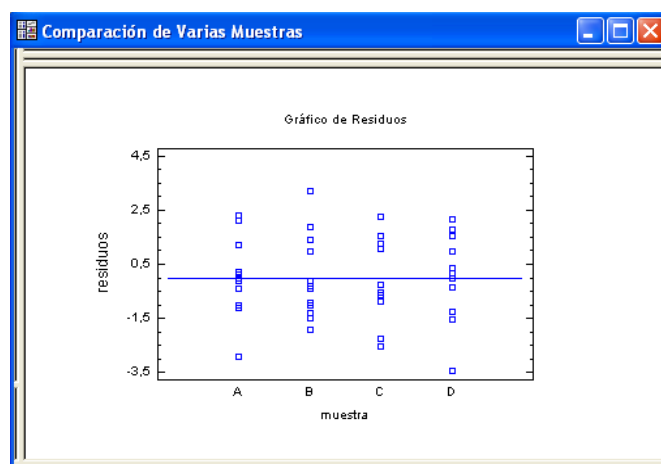


GRAFICO 119. GRAFICO RESIDUOS

Debe establecerse la presencia de los siguientes ítems en el anterior gráfico

- *Puntos Aberrantes*: residuos atípicos, esto es, ubicados a una distancia considerable de los demás residuos
- *Heterocedasticidad*: Cambios en la varianza, lo que se refleja en un aspecto de embudo sobre el gráfico; de ser así deben ejecutarse cambios sobre las variables originales (p.e, inclusión de logaritmos).

9.1.6 Gráfico Análisis de Medias (ANOM)

A través del gráfico de análisis de medias podemos comparar medias correspondientes a varias muestras.

Para activar esta opción, sobre el icono Gráficos de la barra de herramientas del análisis seleccionamos Gráficos de análisis de medias.

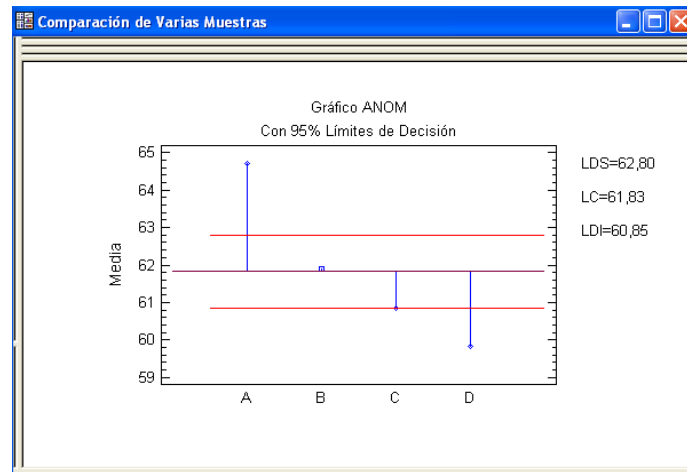


GRAFICO 120. GRAFICO ANOM/COMPARACION VARIAS MUESTRAS

Se muestran las medias de cada muestra asociada con un trazo vertical, acotados por límites de decisión, de esta manera las medias que caigan por fuera de tales límites se asumirán como significativamente diferentes.

5. CONCLUSIONES

STATGRAPHICS CENTRUION XV es un programa que permite el análisis e interpretación de datos, a través de diferentes procesos estadísticos y econométricos.

La herramienta se compone de cuatro módulos principales: editor de reportes de trabajo (Ver literal 1.5. StatReport); asistente de tareas (Ver Literal 1.4 StatWizard); y un enlace estadístico (StatLink) que enlaza el libro de análisis (Statfolio) con la fuente de datos. Adicionalmente el programa presenta una característica exclusiva respecto de los software respecto de los que compite en el mercado, el Stat Advisor herramienta que brinda orientación teórica respecto de los resultados. Mostrando en detalle las implicaciones de los resultados a nivel teórico.

El programa se caracteriza por su manejo a través de ventanas, para trabajar con el mismo no es necesaria la programación y aun a pesar de lo que se podría considerar una gran debilidad el Software se caracteriza por sus capacidades para la representación gráfica y el desarrollo de experimentos, previsiones y simulaciones en función del comportamiento de los valores.

6. BIBLIOGRAFIA

http://www.software-shop.com/in.php?mod=ver_producto&prdID=40

<http://www.statgraphics.com/>

http://www.statcon.de/statconshop/product_info.htm?cPath=9_2&products_id=59

http://www.statgraphics.net/Novedades_Centurion_XV.pdf

7. ANEXOS

ANEXO 7.1

Cálculos

Estimadores de Mínimos Cuadrados

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

donde

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Tabla de ANOVA

$$\text{Suma de cuadrados: } SSR = \hat{\beta}_1^2 S_{xx}$$

$$\text{Error de la suma de cuadrados: } SSE = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

$$\text{Error cuadrado medio: } MSE = \frac{SSE}{n-2}$$

$$\text{Radio F: } F = \frac{SSR}{MSE}$$

$$\text{Bondad de Ajuste: } SSLOF = \sum_{j=1}^c \sum_{i=1}^{n_j} (\bar{y}_j - \hat{y}_{ij})^2$$

$$\text{Error Puro: } SSPE = \sum_{j=1}^c \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$$

$$\text{Radio F para Bondad de Ajuste: } F = \frac{SSLOF / (c-2)}{SSPE / (n-c)}$$

ANEXO 7.2

Errores Standar

$$s(\hat{\beta}_0) = \sqrt{MSE \left[\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right]}$$

$$s(\hat{\beta}_1) = \sqrt{\frac{MSE}{S_{XX}}}$$

Coefficiente de Correlación

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

R-Cuadrada

$$R^2 = \frac{SSR}{SSR + SSE}$$

R-Cuadrada Ajustada

$$R_{adj}^2 = 100 \left[1 - \left(\frac{n-1}{n-2} \right) \frac{SSE}{SSR + SSE} \right] \%$$

Error Estándar de Estimación

$$\hat{\sigma} = \sqrt{MSE}$$

Predicciones

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

$$\text{Límites de Confianza: } \hat{y} \pm t_{\alpha/2, n-2} \hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

$$\text{Límites de Predicciones: } \hat{y} \pm t_{\alpha/2, n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{S_{xx}}}$$

Modelo	Ecuación	Transformación en Y	Transformación en X
Lineal	$y = \beta_0 + \beta_1 x$	ninguna	ninguna
Raíz cuadrada de Y	$y = (\beta_0 + \beta_1 x)^2$	raíz cuadrada	ninguna
Exponencial	$y = e^{(\beta_0 + \beta_1 x)}$	logaritmo	ninguna
Inverso de Y	$y = (\beta_0 + \beta_1 x)^{-1}$	inverso	ninguna
Y Cuadrática	$y = \sqrt{\beta_0 + \beta_1 x}$	cuadrado	ninguna
Raíz cuadrada de X	$y = \beta_0 + \beta_1 \sqrt{x}$	ninguna	raíz cuadrada
Raíz cuadrada doble	$y = (\beta_0 + \beta_1 \sqrt{x})^2$	raíz cuadrada	raíz cuadrada
Y Logaritmico- X Cuadrática	$y = e^{(\beta_0 + \beta_1 \sqrt{x})}$	logaritmo	raíz cuadrada
Y Inversa- raíz cuadrada de X	$y = (\beta_0 + \beta_1 \sqrt{x})^{-1}$	inverso	raíz cuadrada
Y cuadrática- raíz cuadrada de X	$y = \sqrt{\beta_0 + \beta_1 \sqrt{x}}$	raíz cuadrada	raíz cuadrada
X Logaritmico	$y = \beta_0 + \beta_1 \ln(x)$	ninguna	Logaritmo
Raíz cuadrada de Y- logaritmo de X	$y = (\beta_0 + \beta_1 \ln(x))^2$	raíz cuadrada	Logaritmo
Multiplicativo	$y = \beta_0 x^{\beta_1}$	logaritmo	Logaritmo
Y Inverso, X logaritmico	$y = \frac{1}{\beta_0 + \beta_1 \ln(x)}$	inverso	Logaritmo
Y cuadrática - logaritmo de X	$y = \sqrt{\beta_0 + \beta_1 \ln(x)}$	cuadrado	Logaritmo
Inverso de X	$y = \beta_0 + \beta_1 / x$	ninguna	inverso
Raíz cuadrada de Y -inverso de X	$y = (\beta_0 + \beta_1 / x)^2$	raíz cuadrada	inverso
Curva S	$y = e^{(\beta_0 + \beta_1 / x)}$	logaritmo	inverso
Doble inverso	$y = [\beta_0 + \beta_1 / x]^{-1}$	inverso	inverso
Y cuadrática- inverso de X	$y = \sqrt{\beta_0 + \beta_1 / x}$	cuadrado	inversa
X Cuadrática	$y = \beta_0 + \beta_1 x^2$	ninguna	cuadrado
Raíz cuadrada de Y- cuadrado de X.	$y = (\beta_0 + \beta_1 x^2)^2$	raíz cuadrada	cuadrado
Y Logaritmico- X Cuadrático	$y = e^{(\beta_0 + \beta_1 x^2)}$	logaritmo	cuadrado
Y Inverso, X cuadrático	$y = (\beta_0 + \beta_1 x^2)^{-1}$	inverso	cuadrático
Doble cuadrático	$y = \sqrt{\beta_0 + \beta_1 x^2}$	cuadrado	cuadrado
Logístico	$y = \frac{e^{(\beta_0 + \beta_1 x)}}{1 + e^{(\beta_0 + \beta_1 x)}}$	$y/(1-y)$	ninguna
Log probit	$y = \phi(\beta_0 + \beta_1 \ln(x))$	$\phi^{-1}(y)$ (inv. normal)	Logaritmo

