

STATGRAPHICS PLUS.

Aspectos esenciales

1. Introducción
2. Aspectos especiales del Statgraphics
 - 2.1. Visión global del StatAdvisor
 - 2.2. Visión global de StatFolios
 - 2.3. Visión global del StatGallery
 - 2.4. StatReporter
 - 2.5. StatWizard
3. Ventanas del Statgraphics
 - 3.1. Ventana de Aplicación
 - 3.2. Ventana de los Comentarios
 - 3.3. Ventana de datos (Spreadsheet windows)
 - 3.4. Ventana de análisis
 - 3.5. Ventana de asesoría (StatAdvisor)
 - 3.6. Ventana del StatGallery
 - 3.7. Ventana de StatReporter
 - 3.8. Ventana de Vista previa
4. Datos para los análisis (Variables) y su definición
5. Menús principales
 - 5.1. Menú File
 - 5.2. Menú Edit (Edición).
 - 5.3. Menú Plot (Gráficos).
 - 5.4. Menú Describe (Estadística Descriptiva)
 - 5.5. Menú Compare (Comparar)
 - 5.6. Menú Relate (Relación o correlación).
 - 5.7. Menú Special (Estudios especiales)
 - 5.8. Menú View (Ver)
 - 5.9. Menú Windows (Ventana)
 - 5.10. Menú Ayuda
6. Barras de Herramientas
 - 6.1. Barra de herramientas de Aplicación
 - 6.2. Barra de Herramientas de Análisis
 - 6.3. Uso de los comandos
 - 6.4. Paneles de texto y gráficos
7. Procedimiento general de análisis
8. Uso de gráficos (menú Plot)
 - 8.1. Uso de gráficos univariados
 - 8.2. Diagrama de Cajas y Bigote (Box and Whisker Plot)
 - 8.3. Diagrama múltiple de Cajas y Bigotes
 - 8.4. Gráfico probabilístico
 - 8.5. Histograma de Frecuencia
 - 8.6. Distribuciones de Probabilidad
9. Análisis usando el menú DESCRIBE
 - 9.1. Análisis de una variable
 - 9.2. Análisis de varias muestras (multivariado)
 - 9.3. Análisis de Estadísticas organizadas por filas
 - 9.4. Tabulación
10. Algunos resultados del Statgraphics
 - 10.1. Entrada de datos
 - 10.2. Gráfico de series de tiempo
 - 10.3. Diagrama de tallos y hojas.
 - 10.4. Diagrama de cajas y bigotes
 - 10.5. Estadísticas resumen
 - 10.6. Percentiles para ENE
 - 10.7. Distribución de frecuencia
 - 10.8. Histograma de frecuencia (Frequency Histogram – ENE)
 - 10.9. Histograma de frecuencia para el logaritmo de los datos.
 - 10.10. Estadísticas resumen para el logaritmo de ENE
11. Otros aspectos del Statgraphics
 - 11.1. Pruebas de hipótesis
 - 11.2. Pruebas de bondad de ajuste

STATGRAPHICS PLUS

Aspectos esenciales

1. Introducción

Con el fin de simplificar los cálculos estadísticos, principalmente cuando se manejan grandes volúmenes de información, y para liberar al analista de cálculos tediosos y a veces complicados y poder dedicarle un poco más de tiempo a las labores de análisis, se han desarrollado una serie de programas (software) de tipo estadístico. Entre los diferentes programas desarrollados, existen unos de aplicación general, que desarrollan las principales herramientas estadísticas, y existen otros de aplicación más específica, orientados hacia algunos aspectos especializados. En la mayoría de los programas de aplicación general se trata de establecer un balance adecuado entre la facilidad de uso del programa y los análisis estadísticos que contienen. En general, mientras más especializado sea el programa, más complicado es su manejo. Entre los programas de uso general se encuentran Statgraphics, Minitab, Statistica, mientras que entre los especializados merecen mencionarse el SCA y el SPSS.

La siguiente es una lista parcial del software estadístico desarrollado:

- Statgraphics
- Minitab
- Statistica
- SPSS (Statistical Package for the Social Sciences)
- Shazam
- Epiinfo
- SCA
- Eviews

Entre los programas de aplicación general que establecen un buen balance entre la facilidad de uso del programa y los análisis estadísticos que desarrolla se encuentra el Statgraphics. En este documento se presentarán los aspectos esenciales del programa, su modo de manejo y la entrada de la información, y las principales opciones gráficas y los análisis disponibles desde el punto de vista de la estadística descriptiva. También se plantea, en una forma breve, la construcción de los intervalos de confianza, la realización de las pruebas de hipótesis y los análisis de “Bondad de Ajuste”, para determinar cual es la distribución que mejor se ajusta a un conjunto de datos. Se espera que con estos breves conceptos, el estudiante esté en capacidad de usar y explorar el programa, y a través de ensayo y error, aprenda por su cuenta, el manejo de los demás aspectos que requiera del Statgraphics. Para usar opciones mas especializadas del programa, es indispensable que el estudiante consulte los respectivos manuales de usuario, que le darán toda la ayuda requerida. En este documento se supondrá que el estudiante tiene un conocimiento básico del manejo de Windows, y que conoce las principales herramientas estadísticas que acá se presentarán.

2. Aspectos especiales del Statgraphics

El Statgraphics Plus para Windows es un paquete para análisis de datos estadísticos. El diseño del Statgraphics es intuitivo y provee un conjunto de aspectos que lo hacen atractivo para profesionales que trabajan en cualquier industria. Entre los principales aspectos del programa merecen destacarse el StatAdvisor, que da una interpretación de los resultados; StatFolio, que permite guardar y reutilizar los análisis realizados previamente; gráficos interactivos; StatGallery que permite combinar textos y gráficos en múltiples páginas; StatWizard, que guía en la selección de los datos y los análisis, y StatReporter que permite organizar reportes del STATGRAPHICS Plus.

El Statgraphics contiene varios aspectos únicos que facilitan su uso y la generación de informes: El StatAdvisor, Statfolios, StatGallery, StatReporter y StatWizard, los cuales examinaremos brevemente.

2.1. Visión global del StatAdvisor

El StatAdvisor proporciona una asesoría estadística que ayuda, por lado a interpretar un análisis para determinar si los resultados son significativos, y por otro resaltar las áreas donde pueda haber problemas. La interpretación es una explicación simple de los resultados y varía según las variables que se estén usando en el análisis. Por ejemplo, cuando se usa el StatAdvisor después de realizar un análisis de Regresión Múltiple, la interpretación indica qué variables parecen ser significativas para el modelo ajustado.

Por lo tanto, el StatAdvisor da una asesoría estadística que

- Ayuda a interpretar los resultados de un análisis.
- Resalta posibles fallas y áreas donde puede haber problemas.
- Determina si los resultados son estadísticamente significativos
- Da credibilidad a los reportes.

Se puede acceder al StatAdvisor de una de tres maneras siguientes:

- Pulsando el ícono del StatAdvisor en la ventana de Aplicaciones.
- Pulsando el botón del StatAdvisor de la barra de tareas de la ventana de Aplicación.
- Pulsando el comando del StatAdvisor en el menú Ver (esta acción lleva el StatAdvisor a la ventana de Análisis).

2.2. Visión global de StatFolios

Los StatFolios eliminan la necesidad de rehacer un análisis que se usará posteriormente. Se puede guardar un análisis completo en un StatFolio, lo cual incluye guardar las variables, parámetros de los gráficos, y las opciones del análisis. Cuando se necesite repetir el análisis con un nuevo juego de datos, se puede usar el StatFolio guardado, seleccionar nuevas variables o nuevos datos, si es necesario, y ejecutar de nuevo el análisis. Cuando se imprime un StatFolio, se puede imprimir un análisis a la vez, o todos los análisis al mismo tiempo.

2.3. Visión global del StatGallery

StatGallery permite copiar hasta un panel de 100 gráficos y un número ilimitado de paneles de texto en múltiples páginas a las ventanas de StatGallery de tal forma se puedan ver o imprimir. La ventana de StatGallery consiste de nueve paneles que están organizadas en tres columnas. Los paneles se numeran en el siguiente orden:

1	2	3
4	5	6
7	8	9

y se agrupan en las columnas de la siguiente manera:

Columna 1 = Hojas 1, 4, 7,

Columna 2 = Hojas 2, 5, 8,

Columna 3 = Hojas 3, 6, 9,

Generalmente hay cuatro paneles visibles, pero usando los comandos de organización de los paneles en el menú del StatGallery se puede cambiar el arreglo de los paneles y las columnas. Las barras que separan las nueve ventanas permiten cambiar la altura y el ancho de los paneles.

2.4. StatReporter

El StatReporter permite crear un reporte personalizado. Usar esta ventana es como usar un procesador de textos. Se pueden crear informes personalizados copiando y pegando paneles de texto y/o de gráficos, y agregando su propio texto, cambiando el estilo de los textos, su forma de presentación, etc.

2.5. StatWizard

El StatWizard está especialmente diseñado para usuarios nuevos u ocasionales que requieren ayuda para seleccionar los análisis apropiados para sus datos, utilizando una de las siguientes dos opciones: seleccionar los análisis de un menú estándar, o escogiéndolos de una lista rápida.

3. Ventanas del Stagraphics

El Statgraphics presenta las siguientes ventanas: de aplicación, comentarios, datos, análisis, asesoría (StatAdvisor), galerías (StatGallery), Reportes, y vista previa.

Como ya se mencionó, en el material que se presenta en los párrafos siguientes, se supondrá que el estudiante tiene un conocimiento básico del manejo de Windows

3.1. Ventana de Aplicación

Es la ventana inicial que aparece cuando se activa el programa, y se puede considerar que es la principal, ya que contiene los diferentes menús (opciones), las barras de aplicaciones y de tareas, los diferentes íconos y a través de ella se puede acceder a las demás ventanas. En esta ventana se muestran las siguientes opciones:

- Título
- Menú
- Barra de aplicaciones
- Barra de tareas
 - Comentarios
 - Ventana de datos ~~Botón de datos no titulado~~
 - El StatAdvisor
 - StatGallery
 - StatReport

3.2. Ventana de los Comentarios

La ventana de Comentarios aparece cuando se pulsa el botón el Untitled Comments. Esta ventana se usa para registrar información de interés sobre los StatFolios y como una libreta de apuntes para hacer las anotaciones que se consideren necesarias sobre el trabajo o los análisis que se estén realizando.

3.3. Ventana de datos (Spreadsheet windows)

Para accederla se presiona el botón Untitled data task bar o se la selecciona a través del menú "Windows". Esta ventana se usa para crear nuevas hojas de datos, modificar hojas de datos existentes, y para modificar datos mediante el comando de edición.

3.4. Ventana de análisis

La ventana de Análisis aparece después de que se selecciona un gráfico o un análisis estadístico de la barra de menú y se ha suministrado la información apropiada en el gráfico o en la caja de diálogo del análisis de interés. La ventana de análisis contiene los siguientes componentes:

- Barra de título
- Iconos
- Barra de herramientas de análisis
- Paneles de texto
- Gráficos

3.5. Ventana de asesoría (StatAdvisor)

Esta ventana aparece cuando hay disponible una interpretación de un análisis o de un gráfico.

3.6. Ventana del StatGallery

Se usa para organizar texto y gráficos que posteriormente se imprimirán o serán usados para organizar un documento

3.7. Ventana de StatReporter

Esta ventana aparece cuando se pulsa el botón de StatReporter y permite crear un reporte personalizado.

Usar esta ventana es como usar un procesador de textos. Se pueden crear informes personalizados copiando y pegando paneles de texto y/o de gráficos, agregando su propio texto, cambiando el estilo de los textos, su forma de presentación, etc.

Cuando en un análisis se hacen cambios a los gráficos, los cambios son automáticamente actualizados en el StatReporter. Además, se permite realizar anotaciones, copiar items, y trabajar alternativamente entre el procesador de texto y el Statgraphics.

El StatReporter tiene su propia barra de herramientas. Sus items permiten seleccionar la fuente, cambiar el tamaño de las palabras, o la forma de las letras, alinear el texto, agregar datos, y buscar palabras o texto en el documento que se está creando.

3.8. Ventana de Vista previa

Se usa cuando se desee saber cómo quedará la impresión final de un estudio. Se activa a través del menú de Archivo (Files) o presionando el botón derecho del mouse y escogiendo la opción “vista previa”(Print Preview).

4. Datos para los análisis (Variables)

Sin tener en cuenta el tipo de análisis estadístico que se planea realizar, lo primero que se debe hacer es organizar los datos. Organizar los datos incluye recogerlos, clasificarlos y tabularlos.

En Statgraphics, los datos se almacenan en variables que se guardan en Hojas de Datos. Una variable se define como una característica o propiedad de una unidad individual de la población. Una variable describe cualquier hallazgo (un atributo o una característica), que puede cambiar, puede variar, o puede expresarse por más de un conjunto de datos, o en varios valores o categorías. Una variable contiene las observaciones o valores de los datos, que miden cierta característica de una población.

Es sumamente importante conocer el tipo de datos o el significado de la información específica que se está trabajando porque, aunque se pueden procesar varios tipos de datos en los análisis del Statgraphics, los resultados pueden carecer parcial o totalmente de sentido si los análisis no son apropiados para el tipo particular de datos.

Como ya se mencionó, en el Statgraphics las características a estudiar reciben el nombre de Variables, y por lo general cada Variable corresponde a una muestra aleatoria. La información se define en forma matricial, y cada columna de la matriz corresponde a una variable, y las columnas pueden ser de longitud diferente. Cada variable se identifica mediante un nombre, que puede ser definido por el usuario, o asignado por el sistema (por defecto). Si el sistema asigna los nombres, éstos corresponden a Col_1, Col_2, Col_3, etc. Los nombres pueden constar desde 1 hasta 32 caracteres y deben empezar con las letras (A-Z ó a-z), o el signo número(#). Además, los nombres constan de las letras (A-Z ó a-z), los dígitos (0-9), el signo (_), el punto (.), un espacio en blanco (), y los siguientes símbolos: @, #, \$, %, y &.

La entrada de los datos para los análisis se puede realizar de las siguientes maneras:

1) Manual. Entrando a la ventana de datos (presionando el botón “untitled data”) se escriben los datos por columnas, empezando en la primera fila. Por defecto el nombre asignado a la variable corresponde a Col_1 para la primera variable, Col_2 para la segunda, y así sucesivamente. Si se quiere se le puede asignar a cada variable un nombre más nemotécnico que permita identificar más fácilmente la información; para ello se presiona el botón derecho del mouse, estando el cursor ubicado en la variable (columna) de interés y se selecciona la opción “modificar columna”, en la cual se le asigna el nuevo nombre. Una vez entrada la información, el archivo puede guardarse como un archivo de datos de Statgraphics, usando el comando “File/save as/Save Data File/ seguido del nombre del archivo con extensión FS3.

2) Archivos de Statgraphics que hayan sido creados manualmente o por otros medios y guardados bajo esta modalidad. Los archivos se abren mediante el siguiente comando: “File/open/Data File seguido del nombre del archivo con extensión FS3.

3) Archivos de excel. Archivos creados en excel u otras hojas de cálculo se pueden abrir en Statgraphics (importarse) siempre y cuando tengan básicamente la misma estructura de los archivos requeridos en el programa. Bajo el menú “File/Open/Open Data file” en la caja de diálogo se escoge el tipo excel, con extensión .XLS, y luego se selecciona el nombre del archivo. En la siguiente caja de diálogo se responde sí o no, dependiendo si el primer registro en el archivo de datos corresponde al nombre de la variable, o si se desea que el programa genere uno por defecto.

4) Hojas de cálculo de excel. Si no se desea usar todo el archivo del excel, puede seleccionarse la información requerida, copiarse y luego pegarse (usando el comando "edit") en la hoja de datos del Statgraphics (untitled data), como si estuviéramos entrando los datos en forma manual.

En la definición de los datos para los análisis, es necesario tener en cuenta cual es el separador de decimales que esté utilizando el computador (el punto o la coma). En caso de que el separador de decimales del archivo diseñado no coincida con el que usa esté usando el Statgraphics, se puede cambiar presionando el botón derecho del mouse, y usando la opción

5. Menús principales

Los programas en el ambiente Windows hacen uso extenso de menús. Los estándares para la estructura de menú en Statgraphics son consistentes con los de otros productos de Windows; es decir, la barra del Menú proporciona el acceso a la mayoría de los comandos de usuario.

El Statgraphics proporciona una barra del Menú en el tope de la ventana Aplicación que permite acceder a los archivos, las herramientas de edición, y los análisis estadísticos.

La ventana de Aplicación del Statgraphics presenta las diferentes opciones del programa y las barras de aplicaciones. El menú comprende las siguientes opciones, cuyas funciones principales se describen a continuación:

FILE	EDIT	PLOT	DESCRIBE	COMPARE	RELATE	SPECIAL	VIEW	AYUDA
------	------	------	----------	---------	--------	---------	------	-------

5.1. Menú File

Este comando, común a todas las aplicaciones de Windows, se usa para el manejo de archivos (abrir o guardar archivos, imprimir, etc.).

El Menú Files (Archivos) contiene los comandos para abrir o cerrar archivos, imprimir, abrir los archivos recientemente usados, y terminar el programa. Específicamente, el menú de Archivo permite:

- Abrir, cerrar, guardar y guardar como StatFolios, StatGallery, StatReporter, y/o archivos de datos con un nombre diferente; bases de datos; y leer un Portapapeles directamente del STATGRAPHICS.
- Usar StatLink, una propiedad que permite unir StatFolios directamente a otras fuentes de datos como hojas de cálculo, bases de datos, y dispositivos de medición como los micrómetros digitales a través de software especializado
- Imprimir archivos y ver una presentación preliminar de una página antes de imprimirla; seleccionar las impresoras, y guardar un gráfico en un archivo (metafile).
- Combinar StatFolios y/o archivos de datos; enviar correo electrónico, y actualizar hojas de cálculo.

5.2. Menú Edit (Edición).

Este comando, común también a todas las aplicaciones de Windows, se emplea para la manipulación de datos (copiar, cortar, insertar valor o celda, borrar valor o celda y ordenar datos). El comando EDIT (Edición) permite:

- Usar comandos de edición tales como deshacer, cortar, copia, pegar.
- Establecer las preferencias para las opciones del sistema y las opciones de gráficos.
- Cambiar las características del texto y sus atributos, y los títulos de los análisis.
- Insertar una nueva columna o anular una existente.
- Modificar una columna en una hoja de datos, generar datos, transformar los datos, etc.

5.3. Menú Plot (Gráficos).

El comando Plot (graficar) permite realizar análisis básicos a través del uso de varios gráficos comunes.

- Scatterplots (diagramas de dispersión). Los diferentes gráficos que permite crear son: Gráficos univariados, gráficos X-Y, gráficos X-Y-Z, gráficos múltiples X-Y y X-Y-Z, y gráficos con coordenadas polares.
- Gráficos exploratorios. Contiene los análisis que se acostumbra usar para crear gráficos exploratorios que son útiles para estudiar la simetría, verificar suposiciones acerca de la distribución, y detectar valores extremos (outliers). Los análisis realizados son: Diagrama de Cajas y Bigotes (sencillo y múltiple), gráfico de probabilidad normal, histograma de frecuencia, diagrama de puntos (sencillo y múltiple), diagrama de burbuja, y gráfico de radar o estrella.
- Cartas de negocios (comerciales). Contiene análisis que se realizan a través de gráficos útiles para presentar cantidades relativas en un formato visual que sea fácilmente entendible. Los análisis se realizan principalmente a través de gráficos de barras y pastel.
- Distribuciones de probabilidad. Contiene 24 distribuciones que son útiles para generar y guardar números aleatorios, calcular probabilidades y graficar distribuciones de probabilidad y acumuladas.
- Gráficos de superficie de respuesta. Contiene análisis que crean gráficos de superficie de respuesta y gráficos de contorno basados en una función definida por el usuario. Estos gráficos son útiles cuando se necesita visualizar la relación entre los factores y una variable de respuesta.
- Gráficos personalizados. Se usa para crear cartas o gráficos útiles cuando las circunstancias requieren el uso de un gráfico personalizado.

5.4. Menú Describe (Estadística Descriptiva)

Bajo esta opción se realizan los principales análisis relacionados con la estadística descriptiva, o con el análisis de muestras de datos. Este menú permite acceder a los análisis usados para investigar y resumir datos. Las diferentes opciones son:

- Datos numéricos. Contiene análisis usados para describir y resumir un conjunto de datos. Incluye: análisis de una variable, análisis de múltiples variables, análisis de subconjuntos, transformaciones de potencia, y los límites estadísticos de tolerancias.
- Datos categóricos. Contiene análisis que trabajan con datos que contienen códigos numéricos que representan categorías discretas: Tabulaciones, tablas cruzadas y tablas de contingencia.
- Distribuciones. Permite crear distribuciones de probabilidad, producir gráficos de probabilidad para probar suposiciones acerca de una distribución, o ajustar una distribución específica a un conjunto de datos. Los análisis son: Distribuciones de probabilidad, gráficos de probabilidad y ajuste de una distribución para datos truncados y no truncados.
- Datos de vida (duración). Contiene análisis que son útiles para analizar y resumir los datos de duración (pruebas de vida). Los análisis incluyen las Tablas de Vida (intervalos y tiempo), análisis de Weibull, y gráficos de Arrhenius.
- Prueba de Hipótesis. Realiza las pruebas de la hipótesis sobre una variedad de parámetros que involucran una sola muestra.
- Determinación del tamaño de muestra. Determina el tamaño apropiado de una muestra.

5.5. Menú Compare (Comparar)

El menú compare permite acceder a los análisis usados para crear y comparar los resultados de dos o más muestras. Las opciones son las siguientes:

- Dos Muestras. Realiza análisis que comparan dos o más muestras de datos: Comparación de dos muestras, comparación de muestras apareadas, prueba de hipótesis, prueba y determinación del tamaño muestral.
- Muestras múltiples. Realiza análisis que comparan varios conjuntos de datos: Comparación de múltiples muestras, comparación de proporciones, y comparación de conteos.
- Análisis de varianza. Realiza uno de tres análisis diferentes en datos recogidos típicamente en un diseño experimental. El propósito es determinar el efecto de uno o más factores experimentales sobre los valores

de una variable de respuesta. Los análisis son: ANOVA en un sentido, ANOVA multifactor, y los componentes de varianza.

5.6. Menú Relate (Relación o correlación).

El menú Relate permite acceder a los análisis estadísticos que modelan la relación entre las variables dependientes y las variables independientes. (Análisis de regresión). Sus principales opciones son:

- La regresión simple. Ajusta un modelo que relaciona una variable dependiente a una variable independiente minimizando la suma de los cuadrados de los residuales para la línea ajustada.
- Regresión polinómica. Realiza un caso especial de regresión lineal simple cuyos modelos contienen términos de mayor orden que la variable predictora.
- Transformación de Box-Cox. Determina una transformación óptima en el contexto de ajustar un modelo de regresión simple.
- Regresión múltiple. Analiza la relación entre una variable dependiente y uno o variables más independientes.

5.7. Menú Special (Estudios especiales)

Incluye los análisis estadísticos que no están comprendidos en los temas anteriores. Incluye control de calidad, diseño de experimentos, análisis de series de tiempo, regresión avanzada y métodos multivariados.

- Análisis de control de calidad. Contiene análisis que cubren todos los aspectos del control de calidad: Control estadístico por variables y atributos, datos multivariados; identificación de defectos; determinación de la capacidad de un proceso, incluyendo control de calibres; uso de estudios especiales para propósitos de control, procesos de optimización y muestreo de aceptación.
- Diseño de experimentos. Contiene un catálogo interno de diseños que incluyen una amplia gama de opciones, incluyendo screening, superficie de respuesta, factoriales de nivel múltiple, datos categóricos de factores sencillos y múltiples, y componentes de varianza.
- Análisis avanzados de regresión. Contiene análisis que permiten explorar totalmente los datos formulando modelos complejos de regresión múltiple, y métodos de validación para escoger el mejor modelo de regresión.
- Métodos multivariados. Contiene análisis que ayudan a ordenar y agrupar datos, determinar las relaciones entre las variables, y construir pruebas de hipótesis.
- Análisis de series de tiempo. Contiene un conjunto completo de análisis para construir modelos para manejar datos que cambian con el tiempo.

5.8. Menú View (Ver)

Controla la apariencia de los ítems visibles en el Statgraphics, es decir, para seleccionar u ocultar las barras de herramientas, barra de estado y las diferentes ventanas.

5.9. Menú Windows (Ventana)

Permite organizar los íconos, actualizar los paneles, y mover las hojas (ventanas) abiertas a una posición visible.

5.10. Menú Ayuda

El menú de Ayuda accede a la ayuda en línea, y al Asistente de Estadística (StatWizard).

6. Barras de Herramientas

Las Barras de Herramientas son un conjunto de botones en la ventana de Aplicación ubicada inmediatamente debajo de la barra de Menú. Los botones corresponden a comandos del menú. Presionar un botón es equivalente a escoger el correspondiente comando del menú. Si se mueve el indicador del ratón a un botón

en una barra de herramientas y se lo deja allí por unos segundos, se despliega una etiqueta con una breve descripción del propósito del botón.

El Statgraphics tiene dos barras de herramientas: La barra de herramientas de Aplicación y la barra de herramientas de Análisis.

6.1. Barra de herramientas de Aplicación

FILE	EDIT	PLOT	DESCRIBE	COMPARE	RELATE	SPECIAL	VIEW	AYUDA
Barras de Herramientas de Aplicación								

La barra de herramientas de Aplicación despliega los botones que se acostumbra usar para realizar las tareas comunes, como abrir y guardar archivos, o realizar tareas de edición.

Esta barra también permite el acceso a las cajas de diálogo para gráficos básicos, como diagramas de dispersión, gráficos de Cajas y Bigotes. La mayoría de las funciones que se pueden acceder usando la barra de herramientas se pueden acceder también usando los menús.

6.2. Barra de Herramientas de Análisis

Esta barra de herramientas aparece debajo de la barra de herramientas de Aplicación y despliega los botones que se usan para mostrar las cajas de diálogo de análisis, acceder a las opciones tabulares y gráficas, guardar los resultados de ciertos análisis estadísticos, y cuando sea aplicable, acceder a opciones gráficas adicionales.

FILE	EDIT	PLOT	DESCRIBE	COMPARE	RELATE	SPECIAL	VIEW	HELP
Barras de Herramientas de Aplicación								
Barras de Herramientas de Análisis								

6.3. Uso de los comandos

Statgraphics permite escoger los comandos de los menús y submenús de las cajas de diálogo de los análisis y de las cajas de diálogo de las opciones. Por ejemplo, después de completar una caja de diálogo de análisis, se pueden usar los botones de comando para especificarle al programa la acción que se quiere realizar pulsando botones de comando tales como OK o Ayuda.

La mayoría de cajas de diálogo en Statgraphics contienen al menos tres órdenes que aparecen como botones: *OK*, *Cancelar* y *Ayuda*. Como éstos, la mayoría de los nombres de los comandos son auto explicativos. Una vez se realiza un análisis, aparece por lo general un panel de texto y un gráfico, como parte del análisis.

6.4. Paneles de texto y gráficos

Muestran el texto y gráficos en una o más paneles de la ventana de Análisis. El texto y los gráficos se pueden desplegar en forma independiente en paneles separados, o se pueden desplegar combinados, con texto a la izquierda y gráficos a la derecha, en paneles contiguos minimizados.

7. Procedimiento general de análisis

El procedimiento general para realizar una análisis se puede resumir en los siguientes puntos:

- Definir las variables (datos) que se van a analizar. Para ello se requiere abrir o crear un archivo de datos que contenga la(s) variable(s) a analizar (Ver lo definido en el numeral 4).
- Escoger el menú adecuado (Plot, Describe, Compare, Relate, Special) según el análisis que se desea realizar.
- Escoger las opciones (submenús) según el análisis de interés.
- Seleccionar las Variables (Datos) que sean objeto de interés en el análisis. Para ello, en la caja de diálogo que se abre, se requiere "marcar" en la columna de la izquierda, la variable o variables de interés y posteriormente se seleccionan en las columnas de la derecha, dando click en la flecha y cuando todas las

selecciones estén hechas se confirma la selección con OK (aceptar). Si es necesario efectuar una transformación a los datos, se selecciona la opción “*Transformation*” y en la nueva caja de diálogo se selecciona la transformación apropiada, se da doble click sobre dicha transformación, se borra el signo de interrogación en la fila “expresión”, y luego con el cursor se selecciona la variable de interés, y se da doble click, y finalmente se acepta (OK.).

- e) Una vez que el Statgraphics presenta el gráfico o análisis requerido, se pueden modificar uno o varios de sus parámetros, de la siguiente manera: Se posiciona el cursor en el área de análisis o del gráfico, se presiona el botón derecho del mouse, se selecciona la opción requerida (Pane Options, la mayoría de las veces), se completa la caja de diálogo abierta, y finalmente se acepta.
- f) Si se desea agregar el análisis realizado o la correspondiente gráfica al “StatReporter”, de nuevo se ubica el cursor en el área de análisis o del gráfico, se presiona el botón derecho del mouse, se selecciona la opción requerida (Copy Análisis to StatReporter) y finalmente se acepta.
- g) Si se desea realizar otros análisis sobre la misma variable, se puede usar el ícono adecuado de la barra de tareas.

El uso del botón derecho del mouse es clave en la realización de los diferentes análisis del Statgraphics, ya que permite modificar los parámetros de los análisis, la forma de las gráficas y la elaboración del reporte a través del StatReporter.

A continuación se presentan algunas instrucciones para aprender los elementos esenciales para realizar algunos análisis en el sistema base

8. Uso de gráficos (menú Plot)

El programa contiene las herramientas necesarias para crear seis tipos de gráficos: Diagramas de dispersión, gráficos exploratorios, diagramas de negocios, distribuciones de probabilidad, gráficos de superficie de respuesta y gráficos personalizados.

8.1. Uso de gráficos univariados

Los análisis univariados se aplican a métodos para agrupar datos que se refieren a una característica de los miembros de una población o muestra, conocidos como datos univariados. Por ejemplo, el estudio podría referirse a la edad de los estudiantes del grupo de estadística, pero sin considerar el sexo, la estatura y otras características. El análisis genera diagrama de dispersión de los valores.

Los pasos para crear un gráfico univariado son:

- 1) De la barra de menú se selecciona PLOT/SCATTERPLOTS/UNIVARIATE y el sistema despliega la caja de diálogo que le permite escoger o entrar los datos a graficar.
- 2) Se completa la caja del diálogo y se pulsa el botón OK. El sistema presenta el resumen del análisis, el diagrama de dispersión y un gráfico de series de tiempo.

Los gráficos X-Y permiten crear un diagrama de dispersión de una variable contra otra para examinar su relación, es decir, si existe alguna relación de dependencia entre ellas.

Un gráfico X-Y puede mostrar los puntos unidos (Lineplot) o desplegar solamente los puntos de la muestra, sin ninguna línea que los una (Scatterplot) (presionando el botón derecho del mouse y pane options).

Para crear un gráfico X-Y:

- 1) De la barra de menú se selecciona PLOT/SCATTERPLOTS/X-Y PLOT y el sistema despliega la caja de diálogo que le permite escoger o entrar los datos a usar en el gráfico X-Y.
- 2) Se completa la caja del diálogo y se pulsa el botón OK. El sistema presenta el resumen del análisis.

El análisis mediante el gráfico X-Y contiene una opción tabular y una opción gráfica: el resumen del análisis y el diagrama de dispersión, respectivamente.

8.2. Diagrama de Cajas y Bigote (Box and Whisker Plot)

El diagrama de cajas y bigotes resume un conjunto de observaciones univariadas. Es una herramienta de análisis exploratorio de datos que es útil para estudiar la simetría, verificar suposiciones sobre una distribución, y detectar puntos extremos (outliers). Este gráfico es particularmente útil para comparar lotes paralelos de datos (comparar muestras).

Para crear una diagrama de Cajas y Bigotes se requiere:

- 1) De la barra de menú se selecciona PLOT/EXPLORATORY PLOTS/ BOX AND WHISKER PLOT, el sistema despliega la caja de diálogo que le permite escoger o entrar los datos que se desean graficar.
- 2) Se completa la caja del diálogo, se pulsa el botón OK y el sistema presenta el resumen del análisis.

El Análisis de Cajas y Bigotes contiene una opción tabular y una opción gráfica: El Resumen del análisis y el diagrama de Cajas y Bigotes, respectivamente. Debe tenerse en cuenta que si se presiona el botón derecho del mouse se presentan algunas opciones para modificar el diagrama, o para incluir el análisis realizado en el Reporte del Statgraphics.

8.3. Diagrama múltiple de Cajas y Bigotes

El diagrama múltiple de Cajas y Bigotes es una extensión del diagrama sencillo, y se usa para producir diagramas separados pero conjuntos para subgrupos formados a partir de la misma muestra o población, o para analizar muestras independientes. Esto facilita comparaciones de medianas, rangos y valores extremos para cada grupo.

Para crear el Diagrama múltiple de Cajas y Bigotes se requiere:

- 1) De la barra de menú se selecciona PLOT/EXPLORATORY PLOTS/ MÚLTIPLE BOX AND WHISKER PLOT, el sistema despliega la caja de diálogo que le permite escoger o entrar los datos que se desean graficar.
- 2) Se completa la caja del diálogo, se pulsa el botón OK y el sistema presenta el resumen del análisis.

El análisis de Cajas y Bigotes contiene una opción tabular y una opción gráfica: El Resumen del análisis y el diagrama múltiple de Cajas y Bigotes, respectivamente.

8.4. Gráfico probabilístico

El análisis mediante el gráfico probabilístico permite generar un gráfico de probabilidad normal para determinar si un conjunto de datos podría provenir razonablemente de una distribución normal. Un gráfico de probabilidad normal consta de un eje horizontal con escala aritmética y un eje vertical con una escala tal que la función de distribución de una distribución normal se grafique como una línea recta. El gráfico incluye una regresión de mínimos cuadrados para efectos comparativos.

Para crear el gráfico se requiere:

- 1) Seleccionar, de la barra de menú, PLOT/EXPLORATORY PLOTS/NORMAL PROBABILITY PLOTS. El sistema despliega la caja de diálogo que le permite escoger o entrar los datos que se desean graficar.
- 2) Completar la caja de diálogo y pulsar el botón OK. El sistema presenta el resumen del análisis.

El análisis contiene una opción tabular y una opción gráfica: El Resumen del análisis y el gráfico probabilístico, respectivamente.

8.5. Histograma de frecuencia

El análisis de histograma de frecuencia crea un gráfico que despliega las clases en el eje horizontal y las frecuencias de las clases en el eje vertical. La frecuencia de cada clase se representa por una barra vertical cuya altura es igual a la frecuencia de la clase. El programa divide el rango de los datos en los intervalos no traslapados y se cuenta el número de valores que caen en cada intervalo. El número de valores que caen en cada intervalo se denomina frecuencia del intervalo.

Para crear el histograma de frecuencia se requiere:

- 1) Seleccionar PLOT/EXPLORATORY PLOTS/FREQUENCY HISTOGRAM de la barra de menú; el sistema despliega la caja de diálogo que le permite escoger o entrar los datos que se desean graficar.
- 2) Completar la caja de diálogo, y pulsar el botón OK; el sistema presenta el resumen del análisis.

El análisis presenta una opción tabular y una opción gráfica: El Resumen del análisis y el histograma de frecuencia, respectivamente. Presionando el botón derecho del mouse, bajo la opción "Pane Options", se puede escoger el número de intervalos de clase, el valor mínimo, el tamaño del intervalo, si se desea el gráfico de frecuencia relativa y/o acumulada, y si se desea un histograma o un polígono de frecuencia.

8.6. Distribuciones de Probabilidad

El Análisis de distribuciones de probabilidad contiene funciones que permiten realizar tres operaciones básicas para cada uno de 24 distribuciones de probabilidad diferentes:

- Cálculo de probabilidades
- Gráficos de probabilidad y distribuciones acumulativas
- Generación de números aleatorios.

Las 22 distribuciones son:

- Distribuciones discretas: Bernoulli, Binomial, Uniforme discreta, Geométrica, Hipergeométrica, Binomial Negativa y Poisson.
- Distribuciones continuas: Beta, Chi-cuadrado, Erlang, Exponential, Valores Extremos, F(Relación de Varianzas), Proporción, Laplace, Gamma, Logística, Lognormal, Pareto, Normal, t o de Student, Triangular, Weibull y Uniforme

Las siete primeras distribuciones son apropiadas para datos discretos; es decir, datos que pueden tomar sólo valores enteros. Para las variables discretas, las probabilidades son las frecuencias relativas.

Las otras distribuciones son apropiadas para variables continuas que pueden tomar cualquier valor en un intervalo continuo.

Creación de una Distribución de Probabilidad

- 1) De la barra de menú se selecciona PLOT/PROBABILITY DISTRIBUTIONS; el sistema despliega la caja de diálogo que le permite escoger la distribución a graficar.
- 2) Se selecciona la distribución que se desea usar y se pulsa el botón OK. El sistema presenta la ventana de análisis que despliega un resumen de los resultados y la gráfica de la distribución.

Presionando el botón derecho del mouse y bajo "Analysis Options", se le especifican al sistema los parámetros de hasta cinco opciones para la misma distribución de probabilidad.

Estos análisis contienen cuatro opciones tabulares, cinco opciones gráficas, y numerosas opciones de cajas de diálogos.

Las opciones tabulares son: Resumen del análisis, distribución acumulada, distribución inversa, y números aleatorios.

Las opciones gráficas son: Función de densidad o probabilidad, función de distribución, función de supervivencia, logaritmo de la función de supervivencia y la función de riesgo (o de tasa de falla, para estudios de confiabilidad).

9. Análisis usando el menú DESCRIBE

El menú "DESCRIBE" permite realizar análisis de una sola variable (datos univariados) o múltiples variables (análisis multivariados). En general, permite realizar todos los análisis relacionados con la estadística descriptiva.

9.1. Análisis de una variable

El análisis univariado permite analizar los datos de una variable numérica realizando varios análisis y gráficos estadísticos. Por ejemplo, las opciones de este análisis permiten calcular las estadísticas resumen tales como media y desviación estándar, percentiles, e intervalos de confianza.

Para usar el análisis univariado se requiere:

- 1) De la barra de menú seleccionar DESCRIBE/NUMERICAL DATA/ONE VARIABLE ANALYSIS; el sistema despliega la caja de diálogo que le permite escoger la variable a analizar.
- 2) Completar la caja de diálogo, pulsar el botón OK y el sistema despliega el resumen del análisis.

Estos análisis contienen siete opciones tabulares y siete opciones gráficas. Las opciones tabulares son: Percentiles, tabulaciones de frecuencia, diagrama de tallos y hojas (Stem and Leaf display), intervalos de confianza y prueba de hipótesis.

Las opciones gráficas son: Diagramas de dispersión, Gráficos de Cajas y Bigotes, histograma de frecuencia, gráfico de cuantiles, gráfico de probabilidad normal, función de densidad y gráfico de simetría.

9.2. Análisis de varias muestras (multivariado)

El Análisis multivariado permite analizar la relación entre varias variables numéricas realizando varios análisis estadísticos y gráficos. Por ejemplo, las opciones permiten calcular las estadísticas resumen tales como medias y desviaciones estándares, intervalos de confianza, correlaciones y covarianzas.

Para usar el Análisis multivariado se requiere:

- 1) Seleccionar DESCRIBE/NUMERICAL DATA/MÚLTIPLE VARIABLE ANALYSIS; el sistema despliega la caja de diálogo que le permite escoger las variables a analizar.
- 2) Completar la caja de diálogo y pulsar el botón OK; a continuación se despliega el resumen del análisis.

Estos análisis contienen siete opciones tabulares y cuatro opciones gráficas. Las opciones tabulares son: Resumen del análisis, estadísticas resumen, intervalos de confianza, correlación de rangos, covarianzas y correlaciones parciales.

Las opciones gráficas son: Matriz de dispersión, Gráficos de estrellas, gráficos de rayos, y Key Glyph.

9.3. Análisis de Estadísticas organizadas por filas

Este análisis proporciona la capacidad para calcular y guardar estadísticas para datos que están en dos o más columnas. El análisis calcula una sola estadística resumen para cada fila de un conjunto de columnas. Está diseñado para guardar las estadísticas de tal forma que se puedan usar en otros análisis cuando los datos no están organizados en la forma tradicional por columnas.

Para usar el análisis con datos organizados por filas se debe:

- 1) Seleccionar DESCRIBE/NUMERICAL DATA/ROW WISE ANÁLISIS de la barra de menú; el sistema despliega la caja de diálogo que le permite escoger la distribución a graficar.
- 2) Completar la caja de diálogo y pulsar el botón OK; Se despliega el resumen del análisis.

Estos análisis contienen tres opciones tabulares y seis opciones gráficas. Las opciones tabulares son: Medias, medianas, desviaciones, rangos y de Cajas y Bigotes.

Las opciones gráficas son: Diagramas de dispersión, Gráficos de Cajas y Bigotes, histograma de frecuencia, gráfico de cuantiles, gráfico de probabilidad normal, función de densidad y gráfico de simetría.

9.4. Tabulación

El análisis de tabulación resume la distribución de una sola variable categórica mediante una tabulación de frecuencia. La tabulación cuenta el número de veces que ocurre cada valor, o el número de observaciones que caen dentro de intervalos específicos, dependiendo del tipo de datos que se están clasificando.

Los datos que se analizan pueden estar en una de dos formas: (1) observaciones que deben contarse antes de que se desplieguen en la forma tabular o gráfica, y (2) Conteos que representan los resultados de una tabulación previa, y que están listos para el despliegue inmediato.

Para usar el análisis de Tabulación se requiere:

- 1) Seleccionar DESCRIBE/CATEGORICAL DATA/TABULATION de la barra de menú; el sistema despliega la caja de diálogo que le permite escoger los datos a resumir.
- 2) Completar la caja de diálogo y pulsar el botón OK; Se despliega el resumen del análisis.

Estos análisis contiene dos opciones tabulares y dos opciones gráficas. Las opciones tabulares son Resumen del análisis y tabla de frecuencia. Las opciones gráficas el diagrama de Barras y el gráfico de Pastel.

10. Algunos resultados del Statgraphics

En los párrafos siguientes se realizarán varios análisis usando un mismo conjunto de datos. Se presentarán diferentes gráficos y análisis gráficos, y se hará la traducción textual del informe del StatAdvisor (se presentará en letra cursiva).

Ejemplo. Considere los siguientes datos que corresponden a los caudales medios mensuales, (m³/seg) para el mes de enero de diferentes años. Obtenga las principales estadísticas descriptivas, el diagrama de Tallos y Ramas y el diagrama de cajas y bigotes (Box and Whisker Plot).

19.2	16.5	13.2	20.1	18.1	11.9	20.3	27.9	26.5	16.6	23.2	23.1	32.0
29.6	14.9	7.9	18.0	21.6	12.2	11.6	13.8	13.6	11.1	17.2	15.7	27.3
22.5	26.6	15.3	30.7	55.5	25.0	23.5	20.1	27.2	25.0	35.9	23.7	27.4

10.1. Entrada de datos

Lo primero que debe hacerse es definir un archivo con la información de los caudales. Para ello, se abre (ejecuta) el Statgraphics usando el siguiente comando: Inicio/programas/statgraphics/sgwin, y aparece la ventana de datos (untitled). Se ingresan los valores en una columna, teniendo cuidado de usar el separador de decimales adecuado, según la configuración del computador. Por defecto, la variable se denomina Col_1. Para cambiarle el nombre, se ubica el curso en el encabezamiento de la variable, se presiona el botón derecho del mouse (o el comando edit), y luego "modify column", y se le asigna a la variable el nombre deseado. Para nuestro caso se denominará ENE. Antes de realizar cualquier análisis es conveniente, por seguridad, guardar los datos en un archivo, que denominaremos "ensayo". Para ello usaremos el siguiente comando: "File/ save as/Save Data File as/directorio/ensayo". El programa le asigna por defecto la extensión sf3.

Una vez definido el archivo de datos, se procederá a realizar varios análisis gráficos y numéricos.

10.2. Gráfico de series de tiempo Plot/Scatterplots/Univariate Plots

Dado que la información se refiere a datos recogidos en orden cronológico, lo primero que debe realizarse es un gráfico de serie de tiempo o un diagrama de dispersión. Este gráfico se realiza usando la opción "Plot", si es del caso, las opciones gráficas.

Univariate Plot – ENE:

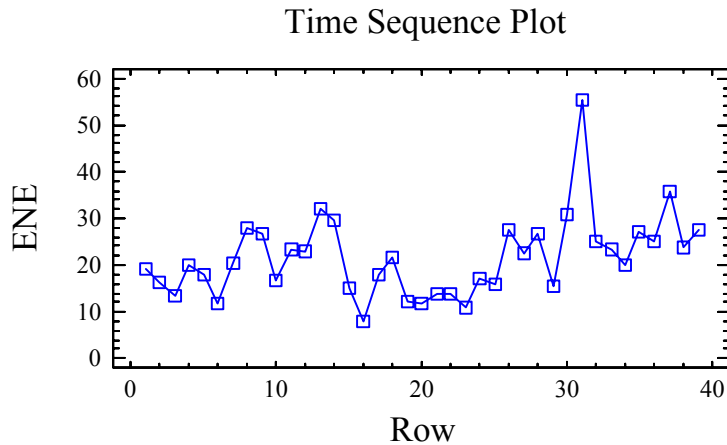
Analysis Summary

Data variable: ENE

39 values ranging from 7.9 to 55.5

The StatAdvisor.

Este procedimiento despliega un gráfico de dispersión de una sola columna de datos. Se pueden crear muchos otros gráficos y estadísticos para los datos seleccionando Describe - Numeric Data - One-Variable Analysis en el menú principal.



Al analizar la gráfica, se observa que no hay tendencias ni ciclos estacionales, y que los datos tienen un comportamiento aproximadamente aleatorio.

10.3. Diagrama de tallos y hojas.

Describe/Numeric Data/One Variable Analysis, y luego la opción tabular

Stem-and-Leaf Display for ENE: unit = 1.0 1|2 represents 12.0

```

1  0|7
9  1|11123334
17 1|55667889
(10) 2|0001233334
12 2|56677779
4  3|02
2  3|5

```

HI|55.48

The StatAdvisor

Este despliegue muestra una tabulación de frecuencia para ENE. El rango de datos ha sido dividido en 7 intervalos (llamados tallos), cada uno representado por una fila de la tabla. Los tallos son rotulados usando uno o más dígitos iniciales para los valores que caen dentro de ese intervalo. En cada fila, los valores individuales de los datos son representados por un dígito (llamado hoja) a la derecha de la línea vertical. Esto resulta en un histograma de los datos de los cuales se pueden recuperar al menos dos dígitos significativos para cada valor. Si hay cualesquier puntos que queden bastante lejos de los otros (llamados puntos extremos) son colocados en tallos separados altos y bajos. En este caso existe un unto extremo. Los puntos extremos se ilustran gráficamente en el diagrama de cajas y bigotes, opción que se puede acceder a través de la lista de opciones gráficas. Los números en el extremo izquierdo corresponden a la frecuencia, que da los conteos acumulados desde arriba y desde el fondo de la tabla, parando el conteo en la fila que contenga la mediana.

El diagrama presenta una distribución que no es simétrica, con un ligero sesgo hacia la derecha. Aunque se presenta un punto extremo, esta situación no es preocupante ya que se trata de la información de caudales de un río, y por lo general en este tipo de información se presentan valores que son o demasiado altos (caudales húmedos) o demasiado bajos (caudales secos).

10.4. Diagrama de cajas y bigotes

Plot/Exploratory Plots/Box and Whisker Plot (Opción gráfica)

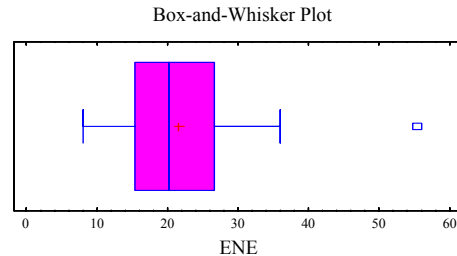
Box-and-Whisker Plot - ENE

Analysis Summary
Data variable: ENE

39 values ranging from 7.9 to 55.5

The StatAdvisor

Este procedimiento muestra un gráfico de Cajas y Bigotes para una sola variable. Se pueden crear muchos otros gráficos y estadísticos para los datos seleccionando Describe - Numeric Data - One-Variable Analysis en el menú principal.



Este diagrama simplemente confirma lo encontrado en el diagrama de tallos y hojas, con la presencia de un valor alto, muy por encima de los demás valores de la muestra.

10.5. Estadísticas resumen

Describe/Numeric Data/One Variable Análisis, y luego la opción tabular (summary Statistics)

One-Variable Analysis - ENE

Summary Statistics for ENE

Count (número de observaciones) = 39
 Average (promedio) = 21.5749
 Variance (varianza) = 74.5615
 Standard deviation (desviación estándar) = 8.6349
 Minimum (mínimo) = 7.9
 Maximum (máximo) = 55.5
 Range (rango) = 47.6
 Std. skewness (coeficiente de simetría estandarizada) = 3.98863
 Std. kurtosis (coeficiente de curtosis estandarizada) = 6.35373

The StatAdvisor

Esta tabla muestra las estadísticas resumen para ENE. Incluye medidas de tendencia central, medidas de variabilidad y medidas de forma. De interés particular acá son el coeficiente estandarizado de simetría y el coeficiente estandarizado de curtosis, que pueden usarse para determinar si la muestra proviene de una distribución normal. Valores de estos estadísticos fuera del rango -2 a +2 indican una desviación significativa de la normalidad, lo cual tendería a invalidar cualquier prueba relacionada con la desviación estándar. En este caso, el coeficiente estandarizado de simetría no está dentro del rango esperado para la distribución normal. El coeficiente estandarizado de curtosis no está dentro del rango esperado para la distribución normal.

El caudal medio (21.6 m³/seg) está por encima de la mediana (20.3 m³/seg, ver percentiles), lo cual indica que la distribución es asimétrica, con un sesgo hacia la derecha, lo cual se confirma también al analizar el valor de desviación estándar, y los valores mínimos y máximos.

10.6. Percentiles para ENE

Describe/Numeric Data/One Variable Análisis, y la opción tabular Percentiles

1.0% = 7.91
 5.0% = 11.1
 10.0% = 11.88
 25.0% = 15.32
 50.0% = 20.3
 75.0% = 26.64
 90.0% = 30.68
 95.0% = 35.91
 99.0% = 55.48

The StatAdvisor

Este panel presenta los percentiles muestrales para ENE. Los percentiles son valores debajo de los cuales se encuentran porcentajes específicos de los datos de la muestra. Se pueden ver los percentiles gráficamente seleccionando Quantile Plot de la lista de Opciones Gráficas.

10.7. Distribución de frecuencia

Describe/Numeric Data/One Variable Análisis, y luego la opción tabular Frequency Tabulation

One-Variable Analysis – ENE. Frequency Tabulation for ENE.

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
	at or below	7.0		0	0.0000	0	0.0000
1	7.0	13.125	10.0625	5	0.1282	5	0.1282
2	13.125	19.25	16.1875	12	0.3077	17	0.4359
3	19.25	25.375	22.3125	11	0.2821	28	0.7179
4	25.375	31.5	28.4375	8	0.2051	36	0.9231
5	31.5	37.625	34.5625	2	0.0513	38	0.9744
6	37.625	43.75	40.6875	0	0.0000	38	0.9744
7	43.75	49.875	46.8125	0	0.0000	38	0.9744
8	49.875	56.0	52.9375	1	0.0256	39	1.0000
above	56.0			0	0.0000	39	1.0000

Mean = 21.5749 Standard deviation = 8.6349

The StatAdvisor

Esta opción realiza una tabulación de frecuencia dividiendo el rango de ENE en intervalos iguales y contando el número de valores en cada intervalo. Las frecuencias muestran el número de datos en cada intervalo, mientras que las frecuencias relativas muestran las proporciones en cada intervalo. Se puede cambiar la definición de los intervalos presionando el botón derecho del mouse y seleccionando la opción Pane Options. Se pueden ver gráficamente los resultados de la tabulación, seleccionando Histograma de Frecuencia de la lista de opciones gráficas.

10.8. Histograma de frecuencia (Frequency Histogram – ENE)

Plot/Exploratory Plots/ Frequency Histogram (Opción gráfica)

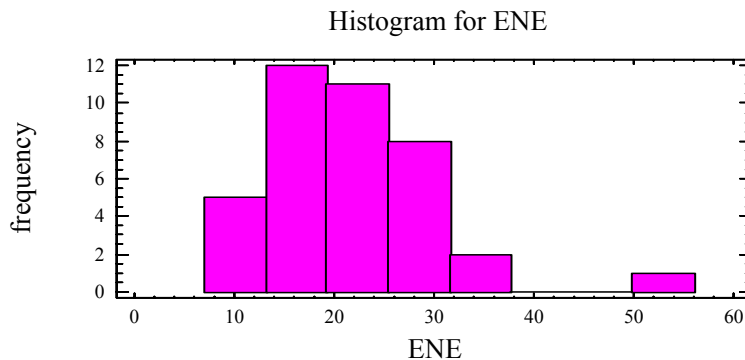
Analysis Summary

Data variable: ENE

39 values ranging from 7.9 to 55.5

The StatAdvisor

Este procedimiento muestra un histograma de frecuencia para una sola columna de datos. Se pueden crear muchos otros gráficos y estadísticos para los datos seleccionando Describe - Numeric Data - One-Variable Analysis a partir del menú principal.



Al analizar la distribución de frecuencia se observa que ésta no es simétrica, con un sesgo hacia la derecha. Los datos pueden seguir una distribución lognormal (por tratarse de unos caudales) o bien una distribución gama. Una forma de verificar si los caudales siguen una distribución logarítmica normal es obteniendo un nuevo histograma, pero tomando como variable no el caudal sino el logaritmo (decimal) del caudal.

10.9. Histograma de frecuencia para el logaritmo de los datos. Frequency Histogram – LOG(ENE) Plot/Exploratory Plots/ Frequency Histogram (Opción gráfica)

Analysis Summary

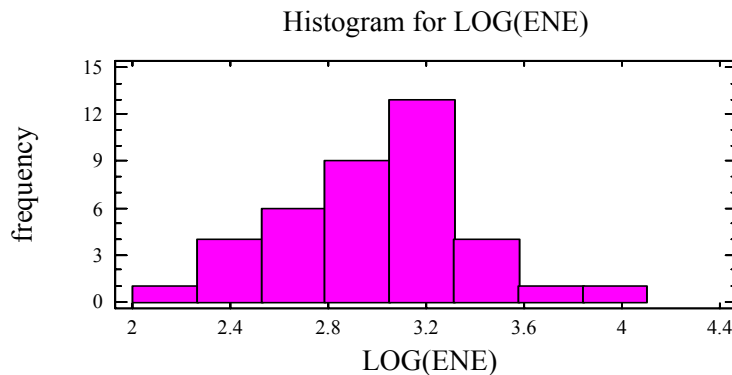
Analysis Summary

Data variable: LOG(ENE)

39 values ranging from 2.06813 to 4.01602

The StatAdvisor

Este procedimiento muestra un histograma de frecuencia para una sola columna de datos. Se pueden crear muchos otros gráficos y estadísticos para los datos seleccionando Describe - Numeric Data - One-Variable Analysis a partir del menú principal.



10.10. Estadísticas resumen para el logaritmo de ENE

Describe/Numeric Data/One Variable Análisis, y opción tabular (summary Statistics)

One-Variable Analysis - ENE

Summary Statistics for LOG(ENE)

Count (número de datos) = 39

Average (promedio) = 3.00053

Variance (varianza) = 0.145593

Standard deviation (desviación estándar) = 0.381566

Minimum (mínimo) = 2.06813

Maximum (máximo) = 4.01602

Range (rango) = 1.94789

Std. skewness (coeficiente estandarizado de simetría) = -0.0595865

Std. kurtosis (coeficiente estandarizado de curtosis) = 0.563699

The StatAdvisor

Esta tabla muestra las estadísticas resumen para ENE. Incluye medidas de tendencia central, medidas de variabilidad y medidas de forma. De interés particular acá son el coeficiente estandarizado de simetría y el coeficiente estandarizado de curtosis, que pueden usarse para determinar si la muestra proviene de una distribución normal. Valores de estos estadísticos fuera del rango -2 a +2 indican una desviación significativa de la normalidad, lo cual tendería a invalidar cualquier prueba relacionada con la desviación estándar. En este caso, el coeficiente estandarizado de simetría está dentro del rango esperado para datos de una distribución normal. El coeficiente estandarizado de curtosis está dentro del rango esperado de datos para una distribución normal.

Al observar el histograma del logaritmo se observa que esta distribución se asemeja un poco más a una distribución normal, lo cual puede comprobarse al analizar los coeficientes estandarizados de simetría y de

curtosis de los logaritmos, los cuales caen en el rango de ± 2 . Por lo tanto, los caudales pueden seguir una distribución logarítmica normal. Para verificarlos, se puede realizar una prueba de “bondad de ajuste”.

11. Otros aspectos del Statgraphics

11.1. Prueba de Hipótesis

Las pruebas de hipótesis son un enfoque clásico para establecer la significancia estadística de los resultados muestrales. Básicamente, la técnica involucra la comparación de los resultados empíricos observados en una muestra con los resultados teóricos esperados, bajo la suposición de que la hipótesis nula es cierta. Esta comparación permite calcular la probabilidad de que los resultados observados puedan ser debidos exclusivamente al azar.

El análisis del Statgraphics permite probar cuatro tipos de parámetros:

- La Media Normal
- La desviación estándar Normal
- La Proporción Binomial
- La tasa de Poisson.

Para realizar los análisis de pruebas de hipótesis se pueden entrar directamente al programa los estadísticos muestrales, o en su lugar se puede entrar los valores muestrales, usando el análisis de variables univariadas.

Cuando se están probando los parámetros correspondientes a la media y desviación normales, el programa calcula intervalos de confianza para la media y varianza usando las distribuciones t y chi-cuadrado, respectivamente, y supone que los datos provienen de una distribución normal. El nivel de significancia se puede establecer libremente para los intervalos de confianza.

EL Statgraphics permite tres hipótesis alternativas:

- Not Equal (diferente de \neq)
- Less than (menor que $<$)
- Greater than (mayor que $>$).

11.1.1. Prueba de hipótesis para una sola muestra

a) Prueba de hipótesis para una sola muestra, usando los estadísticos muestrales

Cuando se tiene una sola muestra, la prueba de hipótesis puede realizarse entrando directamente los estadísticos al programa, mediante la secuencia de comandos “**Describe/ Hypothesis Tests**”, y completando la caja de diálogo, según el tipo de parámetro a probar (la media normal, la desviación estándar normal, la proporción binomial o la tasa de Poisson) y presionando el botón aceptar se despliega el resumen del análisis. Si se desea cambiar el nivel de significancia o el tipo de hipótesis alternativa, se presiona el botón derecho del mouse, y luego **Análisis option**.

El análisis de prueba de hipótesis contiene una opción tabular y una gráfica. La opción tabular es el Resumen del análisis y la opción gráfica es la curva o función de potencia.

Ejemplo. El departamento de seguridad de una fábrica desea saber si el tiempo promedio real que requiere el celador para realizar su ronda nocturna es de 30 minutos. Si en una muestra tomada al azar de 32 rondas, el celador promedió 30.8 minutos con una desviación estándar de 1.5 minutos, determine con $\alpha = 0.01$ si ésta es evidencia suficiente para rechazar la hipótesis nula $\mu = 30$ minutos en favor de la hipótesis alternativa $\mu \neq 30$ minutos.

Solución. Usando el comando **Describe/ Hypothesis Tests**, e introduciendo la información adecuada, se obtienen los resultados mostrados a continuación:

Hypothesis Tests

Sample mean (media muestral) = 30.8

Sample standard deviation (Desviación estándar muestral) = 1.5

Sample size (Tamaño de muestral) = 32

Intervalo de confianza al 90% para la media: 30.8 +/- 0.727625 [30.0724;31.5276]

Null Hypothesis (Hipótesis nula): mean = 30.0

Alternative (Alternativa): not equal (diferente de 30.0)

Computed t statistic (Estadístico t calculado) = 3.01699

P-Value (Valor P) = 0.00506578

Reject the null hypothesis for alpha (Rechace la hipótesis nula para α) = 0.01.

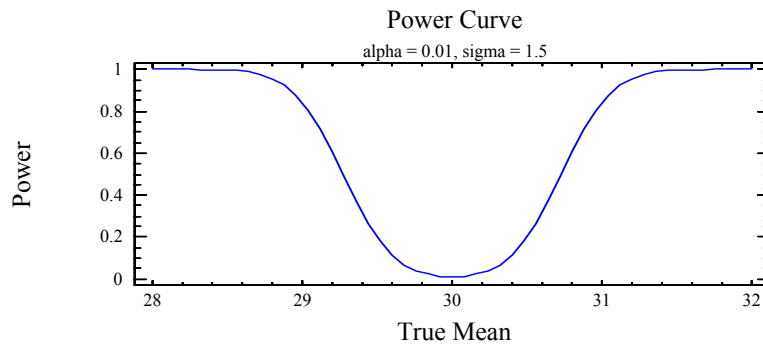
The StatAdvisor

Este análisis muestral los resultados de realizar una prueba de hipótesis sobre la media (μ) de una distribución normal. Las dos hipótesis a ser probadas son:

Hipótesis nula: $\mu = 30.0$

Alternative hypothesis: $\mu \neq 30.0$

Dado un tamaño de muestral de 32 observaciones con una media de 30.8 y una desviación estándar de 1.5, el estadístico t calculado es 3.01699. Dado que el valor P para la prueba es menor de 0.01, se rechaza la hipótesis nula con un nivel de confianza del 99.0%. El intervalo de confianza muestra que los valores de μ soportados por los datos caen entre 30.0724 and 31.5276.



b) Prueba de hipótesis para una sola muestra, usando la muestra aleatoria

Cuando se tiene una sola muestra, la prueba de hipótesis puede realizarse también usando toda la información muestral, mediante la secuencia de comandos "**Describe/Numeric Data/One Variable Analysis**" y usando luego la opción tabular se puede escoger, entre otras, las opciones de pruebas de hipótesis y de intervalos de confianza. Una vez se presenta el resumen del análisis, presionando el botón derecho del mouse y la opción Pane Options se pueden cambiar el nivel de confianza en los intervalos de confianza, o la hipótesis nula, el tipo de hipótesis alternativa y el nivel de significancia cuando se trata de una prueba de hipótesis.

11.1.2. Prueba de hipótesis para dos muestras

a) Prueba de hipótesis para dos muestras, usando los estadísticos muestrales

Al igual que en el caso de una sola muestra, la prueba de hipótesis puede realizarse entrando directamente los estadísticos al programa, mediante la secuencia de comandos "**Compare/Two Samples/Hypothesis Tests**", y completando la caja de diálogo, según el tipo de parámetro a probar (la media normal, la desviación estándar normal, la proporción binomial o la tasa de Poisson) y presionando el botón aceptar se despliega el resumen del análisis. Si se desea cambiar el nivel de significancia o el tipo de hipótesis alternativa, se presiona el botón derecho del mouse, y luego **Análisis option**.

b) Prueba de hipótesis para dos muestras, usando las muestras aleatorias

Si se tiene la información de las dos muestras, la prueba de hipótesis se realiza mediante la secuencia de comandos "**Compare/Two Samples/Two Samples comparison**" y usando luego la opción tabular se puede escoger, entre otras, las opciones de pruebas de hipótesis y de intervalos de confianza. Una vez se presenta el resumen del análisis, presionando el botón derecho del mouse y la opción Pane Options se pueden cambiar el nivel de confianza en los intervalos de confianza, o la hipótesis nula, el tipo de hipótesis alternativa y el nivel de significancia cuando se trata de una prueba de hipótesis.

Ejemplo. Los siguientes son valores de dureza Brinell obtenidos para muestras de dos aleaciones de magnesio. Con un nivel de significancia de 0.05, verifique si la aleación No 2 tiene una mayor dureza que la aleación No 1.

Aleación 1	66.3	63.5	64.9	61.8	64.3	64.7	65.1	64.5	68.4	63.2
Aleación 2	71.3	60.4	62.6	63.9	68.8	70.1	64.8	68.9	65.8	66.2

Solución. Inicialmente se deben organizar los datos en un archivo con dos variables, luego verificar si las varianzas poblacionales son iguales o diferentes, y finalmente se debe realizar la comparación de las medias.

Una vez organizados los datos y usando los comandos **Compare/Two Samples/Two Samples comparison** y luego la opción tabular "comparison of Standard Deviations" se verifica la igualdad de las varianzas.

Comparison of Standard Deviations

	Aleacion 1	Aleacion 2
Standard deviation	1.78702	3.48355
Variance (Varianza)	3.19344	12.1351
Df (Grados de libertad)	9	9

Ratio of Variances (Relación de varianzas) = 0.263157

Intervalos de confianza al 95%:

Standard deviation of Aleacion 1: [1.22918;3.2624]

Standard deviation of Aleacion 2: [2.39611;6.3596]

Ratio of Variances: [0.0653646;1.05947]

F-test to Compare Standard Deviations

Null hypothesis: $\sigma_1 = \sigma_2$

Alt. hypothesis: $\sigma_1 \neq \sigma_2$

F = 0.263157 P-value = 0.059593

The StatAdvisor

Esta opción corre una prueba F para comparar las varianzas de dos muestras. También construye intervalos o límites de confianza para cada desviación estándar y para la relación de varianzas. De particular interés es el intervalo de confianza para la relación de varianzas, que se extiende de 0.0653646 a 1.05947. Dado que el intervalo contiene el valor 1.0, no existe una diferencia significativa entre las desviaciones estándares de las dos muestras a un nivel de confianza del 95.0%

Una prueba F también puede usarse para probar una hipótesis específica acerca de las desviaciones estándares de las poblaciones de las cuales provienen las muestras. En este caso, la prueba ha sido construida para determinar si la relación de desviaciones estándares es igual a 1.0 contra la hipótesis alternativa que la relación no es igual a 1.0. Dado que el valor P calculado no es menor que 0.05, no podemos rechazar la hipótesis nula.

NOTA IMPORTANTE: Las pruebas F y los intervalos de confianza mostrados acá dependen de que las muestras provengan de distribuciones normales. Para probar esta suposición, seleccione "Summary Statistics" de la lista de "opciones tabulares" y verifique los valores de los coeficientes estandarizados de simetría y curtosis.

Una vez confirmado que las varianzas son iguales, se hará el análisis para la comparación de las medias.

Usando los comandos **Compare/Two Samples/Two Samples comparison** y luego la opción tabular "comparison of means" y "Kolmogorov-Smornov Test" se obtienen los resultados que se presentan a continuación.

Two-Sample Comparison - Aleacion 1 & Aleacion 2

Comparación de medias

Límite de confianza al 95.0% para la media de aleación 1: 64.67 + 1.0359 [65.7059]

Límite de confianza al 95.0% para la media de aleación 2: $66.28 + 2.01935 [68.2994]$

Límite de confianza al 95.0% para la diferencia entre medias suponiendo varianzas iguales: $-1.61 + 2.14692 [0.536924]$

Prueba t para comparar medias

Null hypothesis: $mean1 = mean2$

Alt. hypothesis: $mean1 < mean2$

assuming equal variances: $t = -1.30039$ $P\text{-value} = 0.104935$

The StatAdvisor

Esta opción corre una prueba t para comparar las medias de dos muestras. También construye intervalos o límites de confianza para cada media y para la diferencia entre las medias. De particular interés es el intervalo de confianza para la diferencia entre las medias, que se extiende desde 0.536924 hacia arriba.. Esto indica que los datos soportan la hipótesis de que la segunda media es mayor

Una prueba t también puede usarse para probar una hipótesis específica acerca de la diferencia entre las medias de las poblaciones de las cuales provienen las muestras. En este caso, la prueba ha sido construida para determinar si la diferencia entre las medias es igual a 0.0 contra la hipótesis alternativa de que la diferencia es menor que 0.0. Dado que el valor P calculado no es menor que 0.05, no podemos rechazar la hipótesis nula.

NOTA IMPORTANTE: estos resultados suponen que las varianzas de las dos muestras son iguales. En este caso, esta suposición parece razonable con base en los resultados de una prueba F para comparar desviaciones estándares. Se pueden ver los resultados de esta prueba seleccionando Comparison of Standard Deviations del menú de opciones tabulares.

Kolmogorov-Smirnov Test

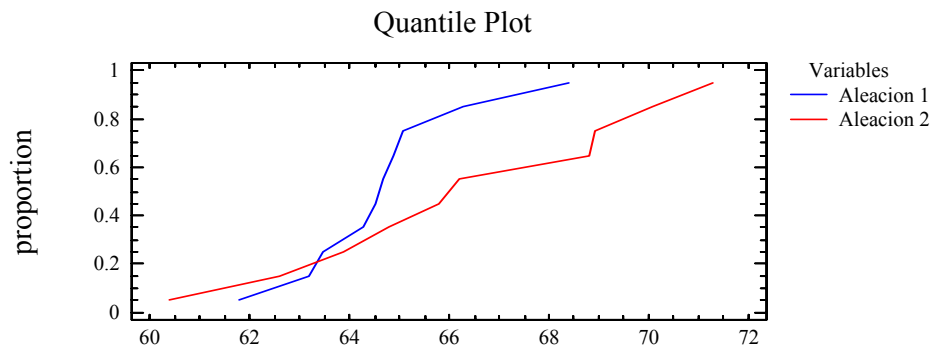
Estimated overall statistic (Estadístico global estimado) DN = 0.4

Two-sided large sample K-S statistic (estadístico S-K de dos colas para muestras grandes) = 0.894427

Approximate P value = 0.405454

The StatAdvisor

Esta opción realiza una prueba de Kolmogorov-Smirnov para comparar las distribuciones de dos muestras. Esta prueba se realiza calculando la máxima distancia entre las distribuciones acumuladas de las dos muestras. En este caso, la máxima distancia es 0.4, que se puede ver visualmente seleccionando Quantile Plot de la lista de opciones gráficas. De especial interés es el valor aproximado de P para la prueba. Dado que el valor P calculado no es mayor de 0.05, no existe diferencia estadística significativa entre las distribuciones a un nivel de confianza del 95.0%.



11.2. Pruebas de bondad de ajuste

Los datos puros rara vez son apropiados para realizar cualquier análisis, de tal forma que es necesario convertirlos en una forma adecuada que pueda soportar análisis significativos. El objetivo es entender la variabilidad aleatoria que existe en cada medición de los datos. Datos no truncados son datos que se han

recogido a través de la duración completa de un experimento, o datos que no tienen que ser excluidos de un experimento por cualquier razón.

El "Uncensored Data Analysis" (análisis de datos no truncados) permite ajustar una entre 24 distribuciones de probabilidad en Statgraphics Plus a un conjunto de datos. El análisis suministra un procedimiento para determinar si los datos siguen una distribución normal (por defecto) o cualquier otro tipo de distribución. El análisis calcula y muestra los parámetros estimados para cada distribución que se escoja.

Para realizar el análisis Distribution Fitting (Uncensored Data) se siguen los siguientes pasos:

1. Se escoge Describe/Distributions/Distribution Fitting (Uncensored Data),
2. Se completa la caja de diálogo y se presiona OK para mostrar el resumen del análisis y el trazo de la función de densidad en la ventana de análisis. Por defecto se ajusta inicialmente una distribución normal.
3. Presionando el botón alterno (botón derecho del mouse) y las opciones Pane Options y Analysis option se seleccionan las opciones para las pruebas de bondad de ajuste y el tipo de distribución a ajustar.
4. Usando el ícono de la opción tabular se selecciona "Goodness of Fit Tests" y usando el ícono de la opción gráfica se selecciona "Frequency Histogram".

A continuación se presentan diferentes análisis realizados al conjunto de datos sobre el caudal mensual de enero para un río, datos que fueron usados en la sección 10. Para el primer ajuste (distribución normal) se presentarán los resultados completos del Statgraphics, y para los demás sólo se presentarán las estadísticas globales

a) Ajuste de una distribución normal

Uncensored Data - ENE

Analysis Summary

Data variable: ENE

39 values ranging from 7.9 to 55.5

Fitted normal distribution:

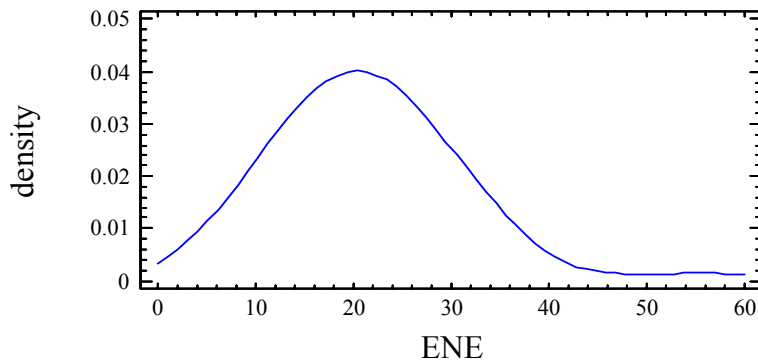
mean = 21.5749

standard deviation = 8.6349

The StatAdvisor

Este análisis muestra los resultados de ajustar una distribución normal a los datos de la variable ENE. Los parámetros estimados de la distribución ajustada se muestran en la parte superior. Se puede probar si la distribución normal se ajusta adecuadamente a los datos seleccionando "Goodness-of-Fit Tests" de la lista de opciones tabulares. También se puede evaluar visualmente que tan bien se ajusta la distribución normal seleccionando "Frequency Histogram" de la lista de opciones gráficas. Otras opciones dentro del procedimiento permiten calcular y mostrar las áreas de las colas y los valores críticos para la distribución. Para seleccionar una distribución diferente, presione el botón alterno del mouse y selecciones Analysis Options.

Density Trace for ENE



Uncensored Data - ENE Uncensored Data - ENE
Goodness-of-Fit Tests for ENE

Chi-Square Test

	Lower Limit	Upper Limit	Observed Frequency	Expected Frequency	Chi-Square
at or below		11.0347	1	4.33	2.56
	11.0347	14.9717	8	4.33	3.10
	14.9717	17.8556	5	4.33	0.10
	17.8556	20.3685	6	4.33	0.64
	20.3685	22.7813	2	4.33	1.26
	22.7813	25.2942	6	4.33	0.64
	25.2942	28.1781	6	4.33	0.64
	28.1781	32.115	3	4.33	0.41
above		32.115	2	4.33	1.26

Chi-Square = 10.6157 with 6 d.f. P-Value = 0.101006

Estimated Kolmogorov statistic DPLUS = 0.103723

Estimated Kolmogorov statistic DMINUS = 0.0869072

Estimated overall statistic DN = 0.103723

Approximate P-Value = 0.795488

EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0.103723	0.660831	>0.10*
Anderson-Darling A^2	0.655467	0.669042	0.0807*

*Indica que el valor P ha sido comparado con tablas de valores críticos construidas especialmente para ajustar la distribución actualmente seleccionada.

Otros valores de P están basados en tablas generales y pueden ser muy conservadores.

The StatAdvisor

Este panel muestra los resultados de pruebas realizadas para determinar si ENE puede modelarse adecuadamente por una distribución normal. La prueba chi cuadrado divide el rango de ENE en intervalos no traslapados y compara el número de observaciones en cada clase con el número esperado, basado en la distribución ajustada. La prueba Kolmogorov-Smirnov calcula la distancia máxima entre la distribución acumulada de ENE y la función de distribución de la distribución normal ajustada. En este caso, la distancia máxima es 0.103723. Los otros estadísticos EDF comparan la función de distribución empírica a la función de distribución ajustada de diferentes maneras.

Dado que el valor más pequeño de P entre las pruebas realizadas es mayor que o igual a 0.10, no podemos rechazar la idea de que ENE viene de una distribución normal, con una confianza del 90% o superior.

b) Ajuste de una distribución gama

Goodness-of-Fit Tests for ENE

Chi-Square = 1.84616 with 6 d.f. P-Value = 0.933289

Estimated Kolmogorov statistic DPLUS = 0.0714619

Estimated Kolmogorov statistic DMINUS = 0.0573575

Estimated overall statistic DN = 0.0714619

Approximate P-Value = 0.988537

EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0.0714619	0.456114	>0.10
Anderson-Darling A^2	0.24214		>0.10*

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution.

Other P-values are based on general tables and may be very conservative.

c) Ajuste de una distribución lognormal

Goodness-of-Fit Tests for ENE

Chi-Square = 3.69238 with 6 d.f. P-Value = 0.718224

Estimated Kolmogorov statistic DPLUS = 0.0667284

Estimated Kolmogorov statistic DMINUS = 0.0783689

Estimated overall statistic DN = 0.0783689

Approximate P-Value = 0.970315

EDF Statistic	Value	Modified Form	P-Value
Kolmogorov-Smirnov D	0.0783689	0.500199	>0.10
Anderson-Darling A^2	0.260598	0.260598	>0.10

*Indicates that the P-Value has been compared to tables of critical values specially constructed for fitting the currently selected distribution.

Other P-values are based on general tables and may be very conservative.

La tabla siguiente resume los valores obtenidos al analizar el ajuste de las distribuciones normal, gama y lognormal para el conjunto de datos. Se observa que para ambas pruebas el mejor ajuste para los datos corresponde a la distribución gama, aunque el ajuste de la distribución logarítmica normal es casi tan bueno como el de la distribución gama.

Resumen de las pruebas de bondad de ajuste			
Criterio	Distribución		
	Normal	Gama	Lognormal
Prueba chi cuadrado			
Valor chi cuadrado	10.6157	1.84616	3.69238
Valor P	0.101006	0.93339	0.718224
Prueba Smirnov kolmogorov			
Diferencia positiva (dplus)	0.10372	0.07146	0.066728
Diferencia negativa (-dminus)	0.08691	0.05736	0.078369
Estadístico global DN	0.10373	0.07146	0.078369
Valor P	0.79549	0.98854	0.970315