

Comparación de Muestras Pareadas

Resumen

El procedimiento **Comparación de Muestras Pareadas** está diseñado para comparar datos en 2 columnas numéricas donde los valores en cada fila están pareados, i.e., corresponden al mismo sujeto o unidad experimental. La razón principal para tal comparación típicamente es determinar si el factor que diferencia las columnas tiene o no efecto en los datos.

Nota: el procedimiento *Comparación de Muestras Pareada* es análogo al procedimiento *Análisis de Una Variable* en términos de salida y opciones, ya que opera sobre una sola columna de diferencias pareadas. Una discusión detallada de las bases matemáticas de los cálculos realizados pueden encontrarse en la documentación del *Análisis de Una Variable*.

StatFolio de Ejemplo: *paired.sgp*

Datos de Ejemplo:

El archivo *93cars.sf3* contiene información de 26 variables para $n = 93$ marcas (*Make*) y modelos (*Model*) de automóviles, tomada de Lock (1993). La tabla a continuación muestra una lista parcial de 4 columnas de ese archivo:

<i>Make</i>	<i>Model</i>	<i>MPG Highway</i>	<i>MPG City</i>
Acura	Integra	31	25
Acura	Legend	25	18
Audi	90	26	20
Audi	100	26	19
BMW	535i	30	22
Buick	Century	31	22
Buick	LeSabre	28	19
Buick	Roadmaster	25	16
Buick	Riviera	27	19
Cadillac	DeVille	25	16
Cadillac	Seville	25	16
Chevrolet	Cavalier	36	25

En este archivo, la unidad experimental es un modelo específico de automóvil. Es de interés comparar las millas por galón (*MPG*) alcanzadas por un modelo particular durante su manejo en carretera (*Highway*) con el desempeño durante su manejo en la ciudad (*City*).

Ingreso de Datos

Los datos a analizar consisten de dos columnas numéricas de observaciones pareadas.



- **Muestra 1:** columna numérica que contiene el primer valor para cada unidad experimental.
- **Muestra 2:** columna numérica que contiene el segundo valor para cada unidad experimental.
- **Selección:** selección de un subgrupo de datos.

Resumen del Análisis

El *Resumen del Análisis* muestra el número de observaciones pareadas en la columna de datos.

Muestras Pareadas - MPG Highway & MPG City

Datos/Variable: MPG Highway-MPG City
93 valores con rango desde 2.0 a 11.0

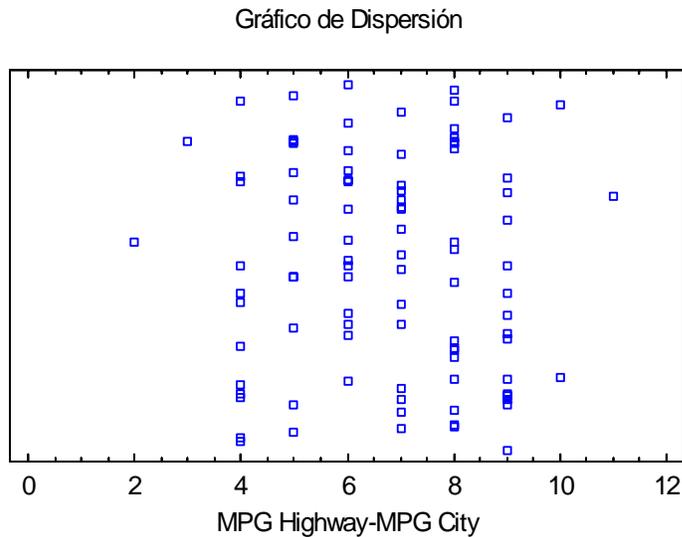
En el análisis que sigue, se presta principal atención a las diferencias pareadas

$$d_i = x_{i,1} - x_{i,2} \quad (1)$$

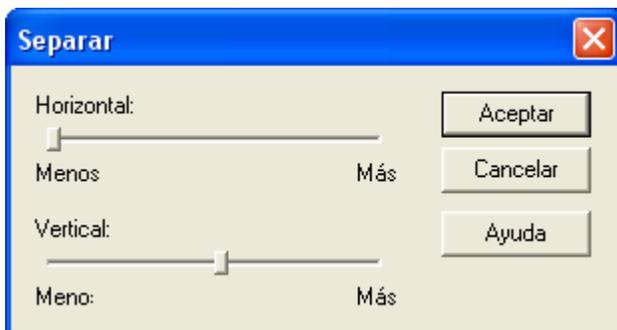
que representan las diferencias entre cada par de observaciones. En los datos muestrales, d_i representa el millaje adicional alcanzado por el modelo i durante el manejo en carretera con respecto al alcanzado en el manejo en la ciudad.

Gráfico de Dispersión

El *Gráfico de Dispersión* grafica cada diferencia pareada.

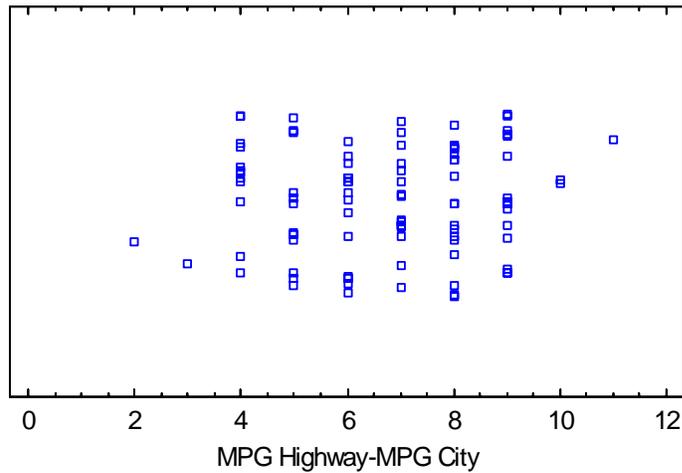


Las diferencias se grafican a lo largo del eje horizontal. A lo largo del eje vertical, los puntos se separan aleatoriamente hacia arriba o hacia abajo. Esto se hace para evitar que puntos con igual valor se traslapen. La cantidad de separación se controla con el botón *Separar* en la barra de herramientas de análisis:



Reduciendo la cantidad de separación *Vertical* se reducirá la cantidad de distanciamiento aleatorio:

Gráfico de Dispersión



Resumen Estadístico

La ventana del *Resumen Estadístico* calcula un número de diferentes estadísticas para las n diferencias pareadas:

Resumen Estadístico para MPG Highway-MPG City	
Recuento	93
Promedio	6.72043
Desviación Estándar	1.85546
Coefficiente de Variación	27.6092%
Mínimo	2.0
Máximo	11.0
Rango	9.0
Rango Intercuartílico	3.0
Sesgo Estandarizado	-0.686851
Curtosis Estandarizada	-1.33445

La mayoría de las estadísticas caen en una de tres categorías:

1. Medidas de *tendencia central* – estadísticas que caracterizan el “centro” de los datos.
2. Medidas de *dispersión* – estadísticas que miden la dispersión de los datos.
3. Medidas de *forma* – estadísticas que miden la forma de los datos con respecto a una distribución normal.

Las estadísticas incluidas en la tabla se controlan por las definiciones de configuración en la ventana *Estadísticas* de la caja de diálogo *Preferencias*. Dentro del procedimiento, se puede cambiar la selección usando la *Ventana de Opciones*. El significado de cada estadística se muestra a continuación.

De particular interés para el presente ejemplo es el promedio o la mediana, que muestran que los automóviles promedian casi más de 7 millas por galón manejados en carretera que conducidos en la ciudad. El rango de la diferencia fue de 2 a 11. El sesgo y la curtosis estandarizados están

ambos entre -2 y +2, indicando que no hay desviación significativa en forma de una distribución normal.

Opciones de Ventana

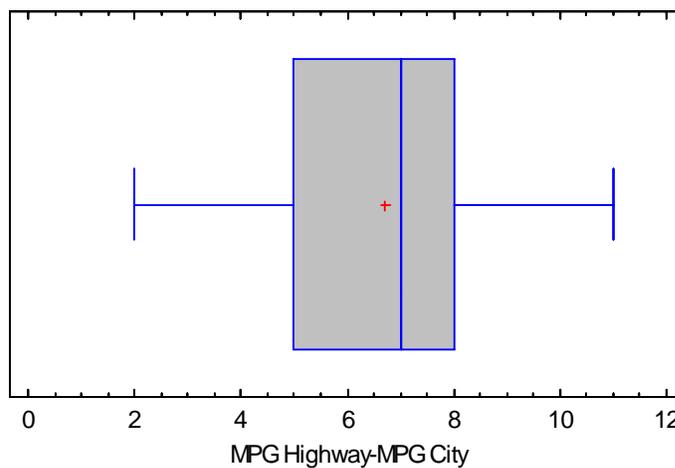


Seleccione las estadísticas deseadas.

Gráfico de Caja y Bigotes

Esta ventana presenta el gráfico de caja y bigotes de las diferencias pareadas.

Gráfico de Caja y Bigotes



Este gráfico se construye de la siguiente forma:

- Se dibuja una caja que se extienda desde el *cuartil inferior* de la muestra hasta el *cuartil superior*. Este es el intervalo cubierto por el 50% central de las diferencias cuando se ordenan de menor a mayor.

- Se dibuja una línea vertical en la *mediana* (el valor de en medio).
- Si se solicita, un signo de más se coloca en el lugar de la diferencia media muestral.
- Los bigotes se dibujan desde los extremos de la caja hasta las diferencias mínima y máxima, a menos que haya valores inusualmente muy alejados de la caja (a los cuales Tukey llama *puntos extremos*). Los puntos extremos, que son puntos a más de 1.5 veces el rango intercuartílico (ancho de la caja) por arriba o por debajo de la caja, se indican por símbolos de señalamiento. Cualesquiera puntos a más de 3 veces el rango intercuartílico por arriba o por debajo de la caja se les llama *puntos extremos lejanos*, y se indican por símbolos de señalamiento con signos de más superpuestos por arriba de ellos. Si hay presentes puntos aberrantes (extremos o extremos lejanos), los bigotes se dibujan a los valores máximo y mínimo que no sean puntos aberrantes.

El gráfico anterior para las diferencias pareadas es muy simétrico. Aunque la línea vertical para la mediana se encuentra algo a la derecha del signo de más para la media, esto probablemente se debe al hecho de que las millas por galón se mide solo al entero más próximo. No hay puntos aberrantes.

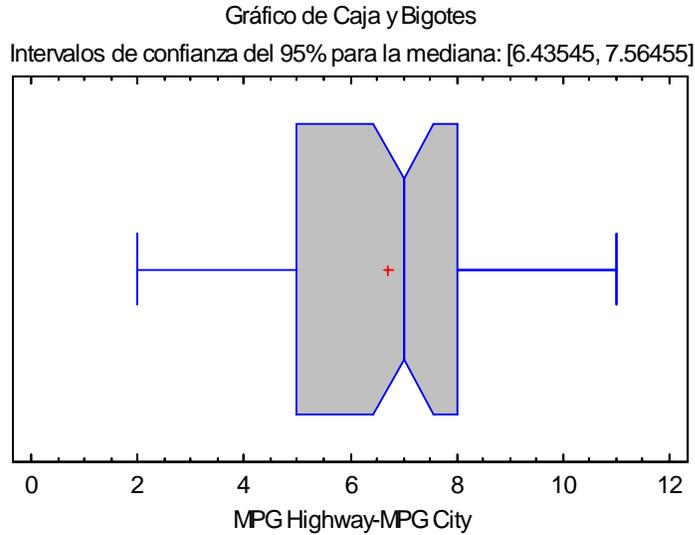
Opciones de Ventana



- **Dirección:** la orientación del gráfico, correspondiente a la dirección de los bigotes.
- **Muesca sobre la Mediana:** si se selecciona, se agregará una muesca al gráfico que muestra un intervalo de confianza de aproximadamente $100(1-\alpha)\%$ para la diferencia mediana al nivel de confianza por omisión del sistema (establecido en la pestaña *General* de la caja de diálogo de las *Preferencias* en el menú *Editar*).
- **Mostrar aberrantes:** si se selecciona, indica la localización de los puntos extremos.
- **Mostrar Media:** si se selecciona, muestra la localización de la media muestral así como la mediana.

Ejemplo – Gráfico de Caja y Bigotes con Muecas

El siguiente gráfico muestra la adición de unas muescas a la mediana a un nivel de confianza del 95%.



La muesca, que va aproximadamente de 6.5 a 7.5, provee de una indicación del potencial error de muestreo en la mediana, suponiendo que las diferencias son una muestra aleatoria de una población normal.

Tabla de Frecuencias

Un método común de resumir datos cuantitativos es construir k intervalos que cubran el rango de los datos y luego calcular el número de observaciones que caen dentro de cada intervalo. STATGRAPHICS presenta este tipo de tabla en la ventana *Tabla de Frecuencias*:

Tabla de Frecuencias para MPG Highway-MPG City							
Clase	Límite Inferior	Límite Superior	Punto Medio	Frecuencia	Frecuencia Relativa	Frecuencia Acumulada	Frecuencia Rel. Acum.
	menor o igual	0.5		0	0.0000	0	0.0000
1	0.5	1.5	1.0	0	0.0000	0	0.0000
2	1.5	2.5	2.0	1	0.0108	1	0.0108
3	2.5	3.5	3.0	1	0.0108	2	0.0215
4	3.5	4.5	4.0	12	0.1290	14	0.1505
5	4.5	5.5	5.0	12	0.1290	26	0.2796
6	5.5	6.5	6.0	15	0.1613	41	0.4409
7	6.5	7.5	7.0	16	0.1720	57	0.6129
8	7.5	8.5	8.0	18	0.1935	75	0.8065
9	8.5	9.5	9.0	15	0.1613	90	0.9677
10	9.5	10.5	10.0	2	0.0215	92	0.9892
11	10.5	11.5	11.0	1	0.0108	93	1.0000
12	11.5	12.5	12.0	0	0.0000	93	1.0000
	mayor de	12.5		0	0.0000	93	1.0000

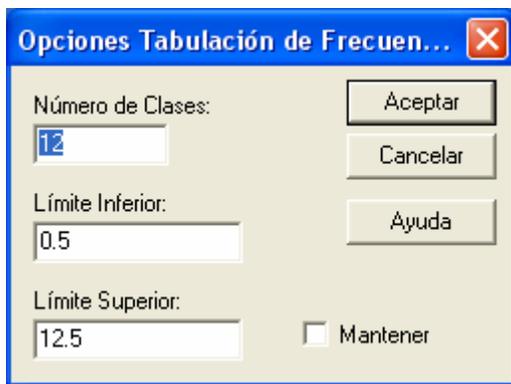
Media = 6.72043 Desviación Estándar = 1.85546

Esta tabla está vinculada con el *Histograma de Frecuencias* y presenta la siguiente información para cada intervalo o “clase”:

- **Límite Inferior** – el límite inferior de la clase.

- **Límite Superior** - el límite superior de la clase
- **Punto Medio** – el punto medio de la clase (a la mitad del recorrido entre los límites superior e inferior).
- **Frecuencia** – el número de observaciones f_j que son mayores que el límite inferior de la clase y menores o iguales al límite superior.
- **Frecuencia Relativa** - la proporción de observaciones que caen en cada clase, dada por f_j/n .
- **Frecuencia Acumulada** - el número de observaciones que caen en la clase actual o en las previas.
- **Frecuencia Relativa Acumulada** – la proporción de observaciones que caen en la clase actual o en las previas.

Opciones de Ventana

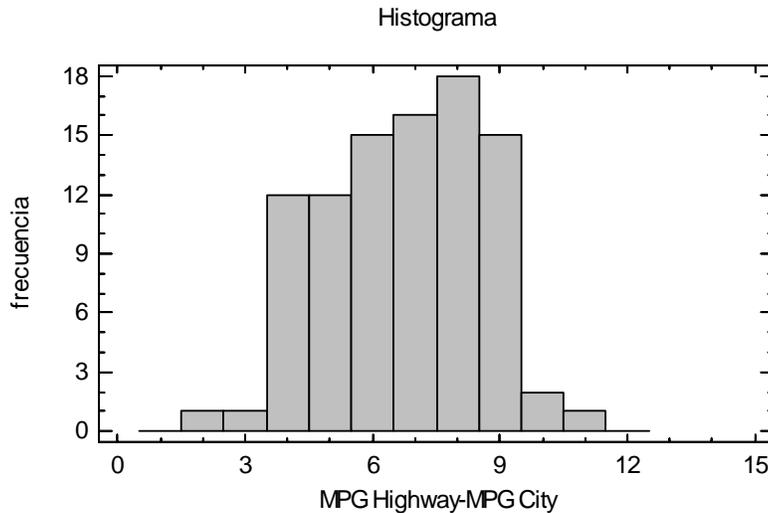


- **Número de Clases:** el número de intervalos en los que se dividirán los datos. Los intervalos son adyacentes unos a otros y de la misma amplitud.
- **Límite Inferior:** límite inferior del primer intervalo.
- **Límite Superior:** límite superior del último intervalo.
- **Mantener:** conserva el número de intervalos y límites seleccionados aún cuando cambien los datos fuente. Por omisión, el número de clases y los límites se recalculan siempre que cambien los datos. Esto es necesario para que todas las observaciones se exhiban aun cuando algunos de los nuevos datos cayeran fuera de los límites originales.

El número de intervalos en los cuales los datos son agrupados por omisión se establece por el criterio especificado en la pestaña *AED* de la caja de diálogo de *Preferencia* en el menú *Editar*. Para datos muy redondeados tales como los del presente ejemplo, se debe tener cuidado de asegurarse que los puntos medios de las clases se alineen con los valores redondeados. De otro modo, se podría presentar una visión distorsionada.

Histograma de Frecuencias

La ventana *Histograma de Frecuencias* presenta el resultado de la tabla de frecuencias en la forma de un diagrama de barras o un gráfico de líneas, dependiendo de las definiciones de configuración de las *Opciones de Ventana*.



La altura de cada barra en el gráfico anterior representa el número de diferencias iguales a cada número entero.

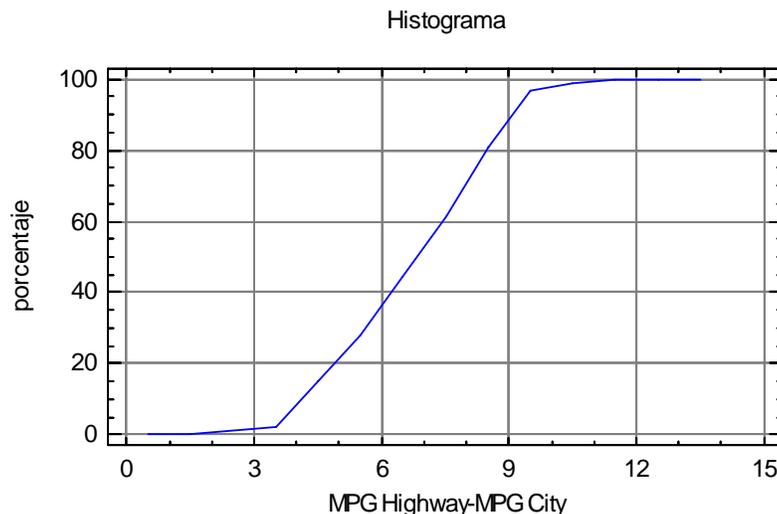
Opciones de Ventana

- **Número de Clases:** el número de intervalos en los que se dividirán los datos. Los intervalos son adyacentes unos a otros y de la misma amplitud.
- **Límite Inferior:** límite inferior del primer intervalo.
- **Límite Superior:** límite superior del último intervalo.

- **Mantener:** conserva el número de intervalos y límites seleccionados aún cuando cambien los datos fuente. Por omisión, el número de clases y los límites se recalculan siempre que cambien los datos. Esto es necesario para que todas las observaciones se exhiban aun cuando algunos de los nuevos datos cayeran fuera de los límites originales.
- **Frecuencia:** si es *Relativa*, la altura representa la proporción de las observaciones con respecto al total en la muestra, y de no señalarse la altura representa el número de observaciones. Si es *Acumulada*, la altura representa las observaciones en el intervalo indicado y en todos los intervalos a su izquierda, y de no señalarse la altura representa las observaciones en un solo intervalo.
- **Tipo de Gráfico:** si es *Histograma*, las frecuencias de las clases se mostrarán como un diagrama de barras. Si es *Polígono*, las frecuencias de las clases se mostrarán usando un gráfico de líneas conectadas.

Ejemplo – Polígono de Frecuencias Acumuladas

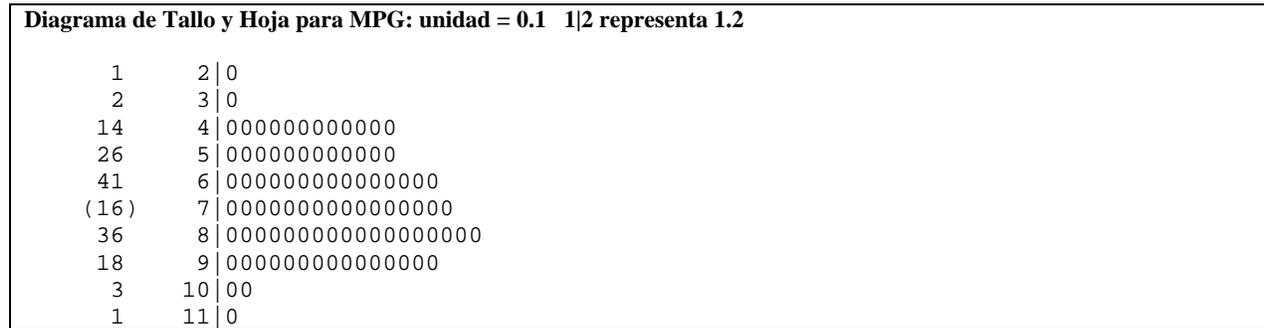
Estableciendo el *Tipo de Gráfico* como *Polígono* y señalando los cuadros de *Acumulada* y *Relativa* da una presentación de la distribución acumulada de las diferencias pareadas:



El gráfico anterior muestra un incremento casi uniforme 3.5 y 9.5.

Diagrama de Tallo y Hojas

El diagrama de tallo y hojas también presenta una tabla de las diferencias.



Este diagrama, debido a John Tukey (1977), toma cada valor de los datos y lo divide en un *tallo* y una *hoja*. Por ejemplo, la diferencia entre las millas por galón en carretera y en la ciudad para el automóvil en la primera fila es igual a 6.0. Llamemos al primer dígito (“6”) el tallo, y el segundo dígito (“0”) la hoja. Cada fila del diagrama de tallo y hojas corresponde a valores con el mismo tallo, mostrado a la izquierda de la línea vertical. A la derecha de la línea vertical, se muestra un solo dígito presentado la hoja para cada valor de los datos. Por ejemplo, la fila que muestra

7|0000000000000000

indica que 16 automóviles rindieron 7.0 millas por galón más en la carretera que en la ciudad. Puntos extremos, definidos de igual forma que para el gráfico de caja y bigotes, se grafican en tallos especiales HI y LO (alto y bajo).

Los números en la columna de hasta la izquierda, llamados *profundidades (depths)*, dan una cuenta acumulada de las observaciones de arriba y abajo hacia el centro del diagrama. En la fila que contiene la mediana, en cambio se muestra puesto entre paréntesis el número de observaciones en esa fila.

Aunque similar a un histograma volcado en su costado, Tukey pensó que el gráfico de tallo y hojas era preferible a un diagrama de barras ya que los valores de los datos podían recuperarse a partir del diagrama. Él usaba las profundidades para localizar la mediana y los cuartiles cuando tabulaba los datos a mano.

Opciones de Ventana



- **Marcar Aberrantes:** si se selecciona, los puntos aberrantes se pondrán en tallos separados HI (de *high*, alto) y LO (de *low*, bajo). De otro modo, se incluirán en la parte principal del gráfico.

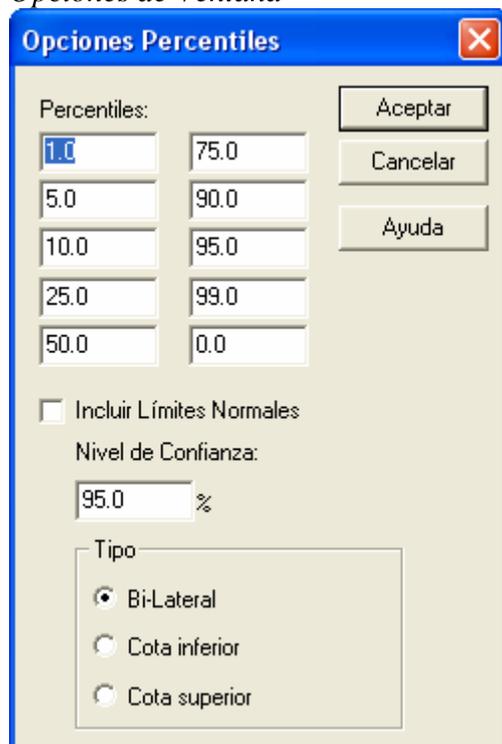
Percentiles

El *p-ésimo percentil* de una distribución de probabilidad continua se define como el valor de X para el cual la probabilidad de ser menor o igual a X es de $p/100$. Por ejemplo, el 90º percentil es el valor por debajo del cual está el 90% de la población. La ventana de *Percentiles* presenta una tabla de percentiles seleccionados con base en los datos muestrales.

	Percentiles
1.0%	2.0
5.0%	4.0
10.0%	4.0
25.0%	5.0
50.0%	7.0
75.0%	8.0
90.0%	9.0
95.0%	9.0
99.0%	11.0

Por ejemplo, el 90º percentil de las diferencias es igual a 9.0, lo que implica que 90% de todos los automóviles tienen diferencias de 9 millas por galón o menores. Si se solicitan usando las *Opciones de Ventana*, también se pueden incluir los límites inferior y superior de confianza o cotas de confianza unilaterales, asumiendo que las diferencias siguen una distribución normal.

Opciones de Ventana

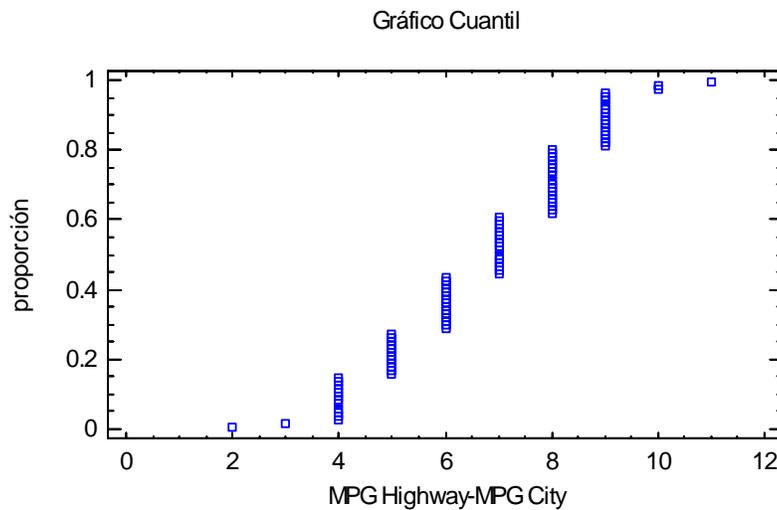


- **Percentiles:** los porcentajes a los que se deberán calcular los percentiles. Poner en 0 para eliminar el cálculo.

- **Incluir Límites Normales:** señalar para incluir límites de confianza o cotas con base en el supuesto de que los datos son muestras aleatorias de una distribución normal.
- **Nivel de Confianza:** nivel para los límites o las cotas.
- **Tipo:** seleccione *Bi-Lateral* para un intervalo de confianza, o una cota unilateral, *Cota inferior* o *Cota superior*, para calcular una cota inferior o superior, respectivamente, para el percentil.

Gráfico de Cuantiles

Esta ventana grafica los cuantiles (percentiles) de las diferencias.



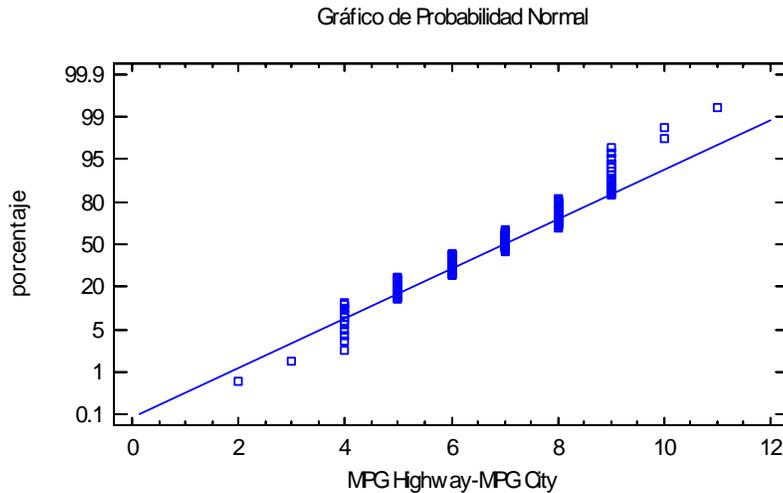
En este gráfico, las diferencias se ordenan de menor a mayor y se grafican en las coordenadas

$$\left(x_{(j)}, \frac{j-0.5}{n} \right) \quad (2)$$

En el ejemplo anterior, el considerable redondeo de los datos provoca que las diferencias se apilen verticalmente en los valores enteros.

Gráfico de Probabilidad Normal

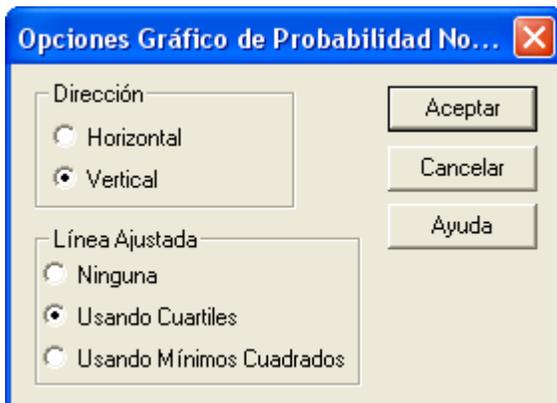
A semejanza del *Gráfico de Cuantiles*, el *Gráfico de Probabilidad Normal* muestra las diferencias pareadas de menor a mayor. Sin embargo, lo hace de manera que es posible juzgar si las diferencias provienen o no de una distribución normal.



El eje vertical se escala de tal manera que, si las diferencias provienen de una distribución normal, los puntos deberán caer aproximadamente a lo largo de una línea recta. Para ayudar a determinar que tan cercanamente los puntos corresponden a una línea recta, se puede superponer una línea de referencia en el gráfico que corresponda a una distribución normal con media μ y desviación estándar σ . Nota: establezca el método por omisión para ajustar líneas en el gráfico de probabilidad normal usando la ventana *AED* en la caja de diálogo de las *Preferencias*, del menú *Editar*. Se puede encontrar una discusión más detallada en la documentación para el procedimiento *Gráfico de Probabilidad Normal*.

En el presente ejemplo, los puntos se desvían de la línea en el extremo superior, no extendiéndose tanto como se esperaría de datos de una distribución normal. La cola superior corta también es evidente en el *Histograma de Frecuencias* ilustrado anteriormente.

Opciones de Ventana



- **Dirección:** la orientación del gráfico. Si es *Vertical*, el *Porcentaje* se presenta en el eje vertical. Si es *Horizontal*, el *Porcentaje* se presenta en el eje horizontal.
- **Línea Ajustada:** el método usado para ajustar la línea de referencia a los datos. Si es *Usando Cuartiles*, la línea pasa por la mediana cuando el *Porcentaje* es de 50 con una pendiente determinada a partir del rango intercuartílico. Si es *Usando Mínimos Cuadrados*, la línea se ajusta con la regresión por mínimos cuadrados de los cuantiles normales de las estadísticas de orden. El primer método basado en los cuartiles da más peso a la forma de los datos cerca del centro y frecuentemente permite mostrar desviaciones de la normalidad en las colas que no serían evidentes usando el método de mínimos cuadrados.

Intervalos de Confianza

La ventana *Intervalos de Confianza* muestra intervalos de confianza para la media y la desviación estándar de las diferencias pareadas. Si se le solicita, también incluye intervalos para la media, mediana y desviación estándar calculados por el método bootstrap.

Intervalos de Confianza para MPG Highway-MPG City

Intervalos de confianza del 95.0% para la media: 6.72043 +/- 0.382128 [6.3383, 7.10256]

Intervalos de confianza del 95.0% para la desviación estándar: [1.62175, 2.1685]

Intervalos Bootstrap

Media: [6.32258, 7.10753]

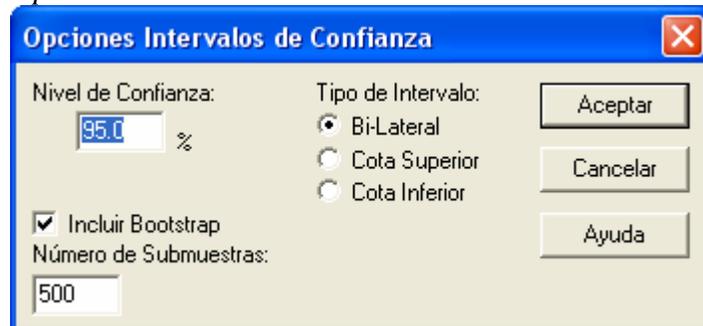
Desviación Estándar: [1.63027, 2.04918]

Mediana: [6.0, 7.0]

Los intervalos de confianza al 95% se construyen de tal manera que, en repetidos muestreos, 95% de tales intervalos contendrán el verdadero valor del parámetro que se estima. También puede ver un intervalo de confianza como especificando el “margen de error” de la misma forma como se enuncia cuando se hace una encuesta de opinión. En el ejemplo anterior, aunque la diferencia media en la muestra fue de 6.72, la diferencia media en la población de la cual los datos fueron muestreados bien puede diferir de esa estimación por 0.38 millas por galón en cualquier dirección.

Los intervalos de confianza para la media y la desviación estándar descansan en el supuesto de que los datos provienen de una distribución normal. Si esto no se puede sostener, entonces una alternativa es construir intervalos usando el método *bootstrap*. En este método, se forman q submuestras seleccionando aleatoriamente con reemplazo (i.e., la misma observación puede ser seleccionada más de una vez) m observaciones de la muestra original. Para cada una de las q submuestras, se calculan la media, la mediana y la desviación estándar. Luego se obtienen intervalos de confianza bilaterales o unilaterales usando percentiles de la distribución observada de las estadísticas de las submuestras. Si los datos no provienen de una distribución normal, los intervalos bootstrap pueden diferir considerablemente de los obtenidos analíticamente. También, a causa de la naturaleza aleatoria de este procedimiento, se obtendrán diferentes resultados cada vez que el método bootstrap se lleve a cabo.

Opciones de Ventana



- **Nivel de Confianza:** nivel para los intervalos o las cotas.
- **Tipo de Intervalo:** seleccione *Bi-Lateral* para un intervalo de confianza o para una cota de confianza seleccione *Cota Superior* o *Cota Inferior*.
- **Incluir Bootstrap:** incluye intervalos bootstrap en la salida.

- **Número de Submuestras:** el número de submuestras q en las que se basarán los intervalos.
Nota: cada submuestra tendrá $m = n$ observaciones, muestreadas con reemplazo.

Pruebas de Hipótesis

Frecuentemente surgen circunstancias donde es necesario determinar si la muestra proviene de una distribución con una media o desviación estándar particulares. Por ejemplo, frecuentemente es de interés probar si la diferencia media pudiera ser igual a cero. Para determinar si esta es o no una aseveración razonable dados los datos que han sido colectados, son posibles dos enfoques:

1. Construir un *intervalo de confianza* para la diferencia media y determinar si 0 está o no dentro del intervalo de confianza.
2. Realizar una *prueba de hipótesis* estadística formal.

La ventana de las *Pruebas de Hipótesis* da soporte al último enfoque.

Prueba t para la Media

A continuación se muestra la parte superior de la salida:

<p>Prueba de Hipótesis para MPG Highway-MPG City Media Muestral = 6.72043 Mediana Muestral = 7.0 Desviación Estándar de la Muestra = 1.85546</p> <p><u>Prueba t</u> Hipótesis Nula: media = 0.0 Alternativa: no igual</p> <p>Estadístico t = 34.9291 Valor-P = 0.0 Se rechaza la hipótesis nula para alfa = 0.05.</p>

Para correr una prueba de hipótesis, se formulan dos hipótesis que entran en competencia:

- **Hipótesis Nula:** una hipótesis tal como $\mu = 0$ a la que se le dará el beneficio de la duda. El valor especificado por la hipótesis nula se etiqueta μ_0 .
- **Hipótesis Alternativa:** una hipótesis tal como $\mu \neq 0$ que conducirá al rechazo de la hipótesis nula si hay suficiente evidencia en contra de la nula.

El enfoque estadístico estándar a este problema es construir una prueba t usando:

$$t = \frac{\bar{d} - \mu_0}{s / \sqrt{n}} \quad (3)$$

y comparándola con una distribución t de Student con $v = n - 1$ grados de libertad.

La tabla anterior muestra los resultados de esta prueba:

- *Estadístico t calculado* – el valor calculado $t = 34.9291$
- *Valor de P* – un valor que puede ser usado para rechazar la hipótesis nula si es lo suficientemente pequeño. Al nivel de significancia $\alpha = 5\%$, la hipótesis nula se rechazará si $P < 0.05$.

En este caso, hay una evidencia extremadamente fuerte de que los datos no provienen de una población en la cual la diferencia media sea igual a 0.

Prueba para la Mediana

Si la distribución de la que provienen las diferencias no es normal, tal vez sea de mayor interés probar una hipótesis sobre la mediana poblacional más que sobre la media. STATGRAPHICS realiza dos de tales pruebas: una prueba de los signos y una prueba de rangos con signo.

Prueba de los signos

Hipótesis Nula: mediana = 0.0

Alternativa: no igual

Número de valores menores a la mediana hipotética: 0

Número de valores mayores a la mediana hipotética: 93

Estadístico para Grandes Muestras = 9.53996 (aplicada la corrección por continuidad)

Valor-P = 0.0

Se rechaza la hipótesis nula para $\alpha = 0.05$.

Prueba de rangos con signo

Hipótesis Nula: mediana = 0.0

Alternativa: no igual

Rango medio de valores menores a la mediana hipotética: 0.0

Rango medio de valores mayores a la mediana hipotética: 47.0

Estadístico para Grandes Muestras = 8.39782 (aplicada la corrección por continuidad)

Valor-P = 0.0

Se rechaza la hipótesis nula para $\alpha = 0.05$.

La *Prueba de los Signos* se basa en la comparación del número de diferencias por debajo de la mediana hipotética con el número de diferencias por arriba de la misma. Una gran discrepancia conduce al rechazo de la hipótesis nula. La *Prueba de Rangos con Signo* le da rango a las diferencias absolutas entre las diferencias pareadas y la mediana hipotética de menor a mayor y compara el rango promedio de las diferencias pareadas por debajo de la media hipotética con el rango promedio de las de por arriba.

De primordial importancia en la tabla anterior son los Valores de P. Valores pequeños (por debajo de 0.05 si se trabaja al nivel de significancia del 5%) conducen al rechazo de la hipótesis nula. En el presente ejemplo, ambas pruebas rechazan la idea de que la median de las diferencias sea igual a 0.

Opciones de Ventana

Opciones Pruebas de Hipótesis

Localización

Prueba t Media/Mediana: 0.0

Prueba de los Signos

Prueba de Rangos con Signo

Alfa: 5.0 %

Hipótesis Alternativa

Diferente de

Menor Que

Mayor Que

Dispersión

Prueba Chi-Cuadrada Desviación Estándar: 1.0

Alfa: 5.0 %

Hipótesis Alternativa

Diferente de

Menor Que

Mayor Que

Aceptar

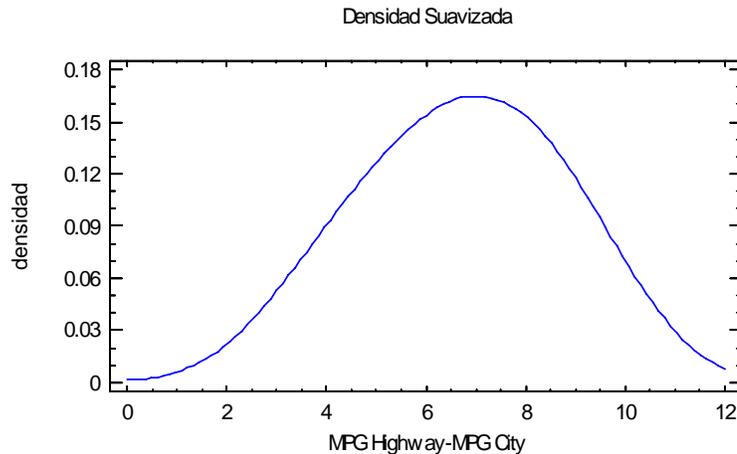
Cancelar

Ayuda

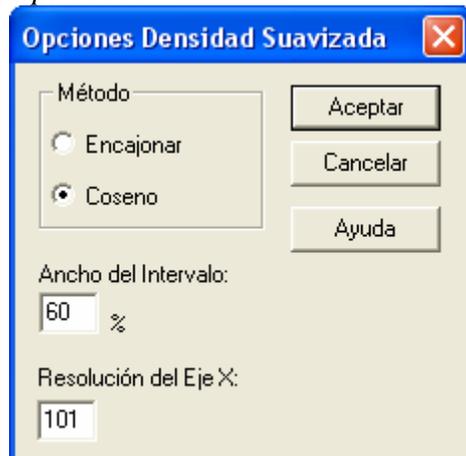
- **Media/Mediana:** μ_0 , el valor de la media o mediana especificada por la hipótesis nula.
- **Alfa:** el nivel de significancia de la prueba, generalmente establecido en 0.01, 0.05 ó 0.10. Éste es igual a la probabilidad de rechazar la hipótesis nula siendo cierta. No afecta al Valor de P, solo las conclusiones expuestas inmediatamente a continuación del Valor de P.
- **Hipótesis Alternativa:** la hipótesis alternativa puede ser de dos colas (“Diferente de”) o de una cola (tal como $\mu > 0$ si se especifica “Mayor que”).

Densidad Suavizada

La *Densidad Suavizada* provee una estimación no paramétrica de la función de densidad de probabilidad de la población de la cual fueron muestreadas las diferencias. Se crea contando el número de diferencias que caen dentro de una ventana de ancho fijo que se mueve a través del rango de los datos.



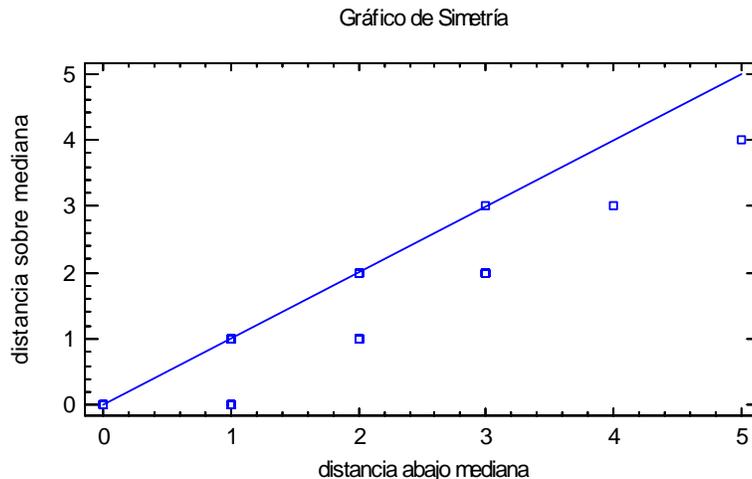
Opciones de Ventana



- **Método:** la función ponderadora deseada. La función encajonar pondera todos los valores dentro de la ventana de igual forma. La función coseno da pesos decrecientes a las observaciones más alejadas del centro de la ventana. La selección por omisión está determinada por las definiciones de configuración en la pestaña *AED* de la caja de diálogo de las *Preferencias* del menú *Editar*.
- **Ancho del Intervalo:** el ancho h de la ventana dentro de la cual las observaciones afectan la densidad estimada, como un porcentaje del rango cubierto por el eje x . $h = 60\%$ no es irrazonable para una muestra pequeña pero puede no dar tanto detalle como lo haría un valor menor en muestras más grandes.
- **Resolución del Eje X:** el número de puntos en los cuales se estimará la densidad.

Gráfico de Simetría

El gráfico de simetría se usa para ayudar a juzgar si las diferencias provienen de una distribución simétrica, i.e., una distribución que tiene una función de densidad con la misma forma a cada lado de la mediana.



Para crear este gráfico, las diferencias se ordenan y luego se hacen pares con base en su localización con respecto a la mediana. Por ejemplo, con 93 diferencias, los puntos ordenados se aparean así:

$$(d_{(46)}, d_{(48)}), (d_{(45)}, d_{(49)}), (d_{(44)}, d_{(50)}), \dots, (d_{(1)}, d_{(93)})$$

Se grafica la distancia de cada par por arriba y por debajo de la mediana. Si los datos provienen de una distribución simétrica, los puntos deberán caer cerca de una línea a 45 grados. Si no, los puntos se desviarán de la línea en una dirección particular. Este gráfico no es muy útil para una muestra con datos muy redondeados.

Salvar Resultados

Se pueden salvar los siguientes resultados en la hoja de datos:

1. **Resumen Estadístico** – los valores de las estadísticas mostradas en la ventana *Resumen Estadístico*.
2. **Etiquetas de las Estadísticas** – las etiquetas para las estadísticas mostradas en la ventana *Resumen Estadístico*.
3. **Percentiles** – los valores de los percentiles exhibidos en la ventana *Percentiles*.
4. **Frecuencias** – las frecuencias de clase exhibidas en la ventana *Tabla de Frecuencias*.
5. **Frecuencias Acumuladas** – las frecuencias acumuladas de clase exhibidas en la ventana *Tabla de Frecuencias*.
6. **Frecuencias Relativas** – las frecuencias relativas de clase exhibidas en la ventana *Tabla de Frecuencias*.
7. **Frecuencias Relativas Acumuladas** – las frecuencias relativas acumuladas de clase exhibidas en la ventana *Tabla de Frecuencias*.

Cálculos

Consulte la documentación para el procedimiento *Análisis de Una Variable* para detalles sobre los cálculos realizados.