

Identificación de Valores Atípicos

Resumen

El procedimiento **Identificación de Valores Atípicos** está diseñado para ayudar a determinar si una muestra de n observaciones numéricas contiene o no valores atípicos. Por “valor atípico” (outlier), queremos decir una observación que no proviene de la misma distribución que el resto de la muestra. Se incluyen ambos métodos: gráficos y pruebas estadísticas formales. El procedimiento también salva una columna en la hoja de datos identificando el valor atípico en una forma que pueda ser usado en campo *Selección* en otra caja de diálogo de ingreso de datos.

StatFolio de Ejemplo: *outlier.sgp*

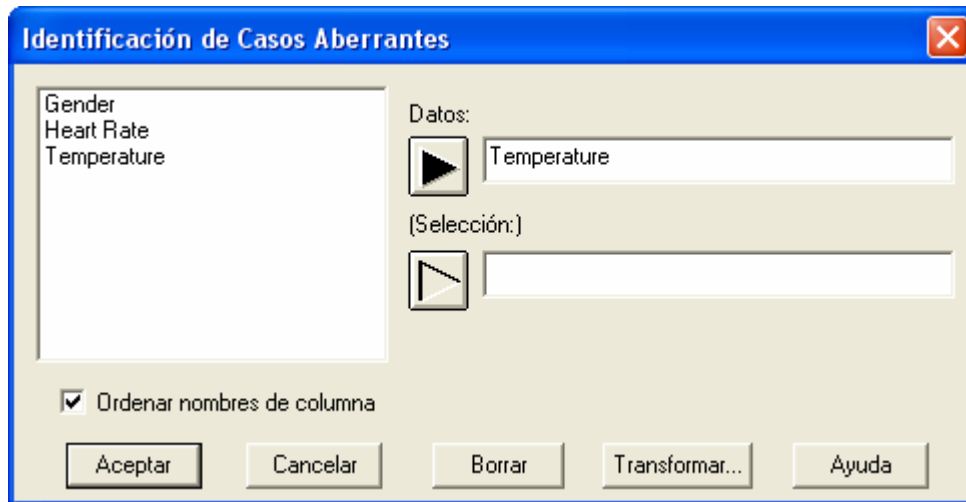
Datos de Ejemplo:

El archivo *bodytemp.sf3* contiene los datos que describen la temperatura corporal de una muestra de $n = 130$ personas. Se obtuvo del Journal of Statistical Education Data Archive (www.amstat.org/publications/jse/jse_data_archive.html) y originalmente apareció en el Journal of the American Medical Association. A continuación se muestran las primeras 20 filas del archivo.

<i>Temperature</i> (temperatura)	<i>Gender</i> (género)	<i>Heart Rate</i> (ritmo cardiaco)
98.4	Male	84
98.4	Male	82
98.2	Female	65
97.8	Female	71
98	Male	78
97.9	Male	72
99	Female	79
98.5	Male	68
98.8	Female	64
98	Male	67
97.4	Male	78
98.8	Male	78
99.5	Male	75
98	Female	73
100.8	Female	77
97.1	Male	75
98	Male	71
98.7	Female	72
98.9	Male	80
99	Male	75

Ingreso de Datos

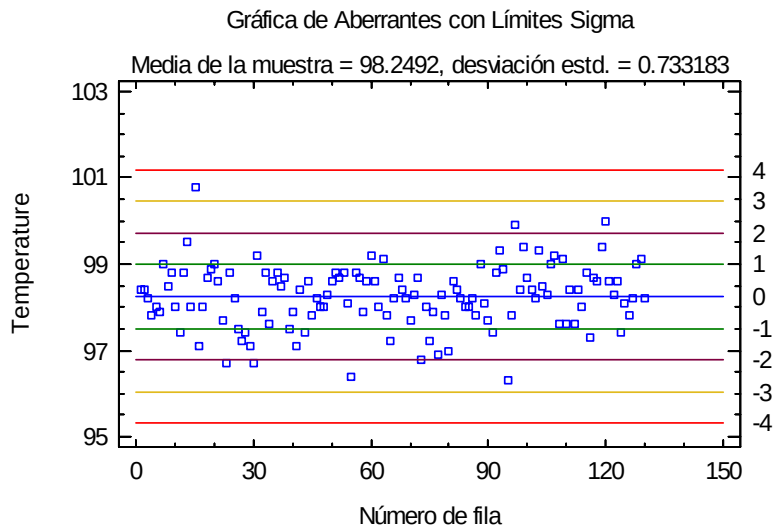
Los datos a analizar consisten de una sola columna numérica con $n = 2$ o más observaciones.



- **Datos:** columna numérica que contiene los datos a resumir.
- **Selección:** selección de un subgrupo de datos.

Gráfica de Aberrantes

Un buen lugar para comenzar cuando se considera la posibilidad de que una muestra de n observaciones contiene uno o más valores atípicos es el *Gráfico de Aberrantes*.



Este gráfico muestra cada valor de los datos junto con líneas horizontales en la media muestral más y menos 1, 2, 3, y 4 desviaciones estándar. Puntos más allá de 3 sigma, de los cuales hay uno en el gráfico anterior, se considera que son valores atípicos potenciales y dignos de investigación adicional.

Resumen del Análisis

El *Resumen del Análisis* presenta un número de estadísticas diseñadas para ser resistentes a valores atípicos, así como el resultado de varias pruebas formales para valores atípicos. A continuación se muestra la sección de arriba de la salida:

Identificación de Valores Atípicos - Temperature		
Datos/Variable: Temperature (degrees)		
130 valores con rango desde 96.3 a 100.8		
Número de valores actualmente excluidos: 0		
Estimados de Localización		
Media muestral	98.2492	
Mediana muestral	98.3	
Media recortada	98.2714	
Media Winsorizada	98.25	
Recorte: 15.0%		
Estimados de Escala		
Desviación estd. muestral	0.733183	
DAM/0.6745	0.74129	
Sbi	0.714878	
Sigma winsorizada	0.708916	
Intervalos de confianza del 95.0% para la media		
	<i>Límite Inferior</i>	<i>Límite Superior</i>
Estándar	98.122	98.3765
Winsorizada	98.1032	98.3968

Estimados de Localización

Se proporcionan cuatro estadísticas que estiman el centro o localización de la población de la cual fueron muestreados los datos, incluyendo:

1. *Media muestral* – la media aritmética de la muestra.
2. *Mediana muestral* – el centro o valor medio de la muestra.
3. *Media recortada* – el valor promedio después de eliminar un porcentaje específico de las observaciones menores y mayores.
4. *Media Winsorizada* – el valor promedio después de sustituir un porcentaje específico de las observaciones menores y mayores con los valores más extremos no incluidos dentro de ese porcentaje.

Si los datos provienen de una distribución normal, cada una de las cuatro estadísticas estima la media poblacional μ . Sin embargo, las 3 últimas estadísticas son menos sensibles a la posible presencia de valores atípicos que la media muestral ordinaria. En el presente ejemplo, hay muy poca diferencia entre las estimaciones. Aunque no siempre es ése el caso.

Estimados de Escala

También hay cuatro estimados de la dispersión de los datos, cada uno de los cuales estima la desviación estándar σ siempre que los datos provengan de una distribución:

1. *Desviación estándar muestral* – la desviación estándar usual.

2. *DAM/0.6745* – una estimación basada en la desviación absoluta mediana (la mediana de la diferencia absoluta entre cada valor de los datos y la mediana muestral).
3. *Sbi* – una estimación basada en una suma ponderada de cuadrados alrededor de la mediana muestral, donde los pesos disminuyen con la distancia a partir de la mediana.
4. *Sigma Winsorizada* – una estimación basada en las desviaciones al cuadrado alrededor de la media Winsorizada.

Los últimos 3 estimadores están diseñados para ser resistentes a valores atípicos. Para los presentes datos, las estimaciones son muy similares.

Intervalos de Confianza

Se presentan intervalos de confianza para la media μ basados en las usuales media y desviación estándar muestrales y también usando las estadísticas Winsorizadas. El hecho de que los intervalos sean tan próximos implica que los valores atípicos no son un gran problema en estos datos.

Valores Extremos

La sección media de la tabla muestra las 5 observaciones mayores y las 5 menores en los datos:

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

Las tres columnas de la derecha muestran valores estandarizados o *Puntajes Z* que pueden ayudar a identificar valores atípicos. Cada estadística mide a cuántas desviaciones estándar están los valores de los datos del centro de los datos.

Valores Estudentizados Sin Supresión – usando la media y la desviación estándar muestrales, cada valor de los datos es estandarizado por

$$t_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

Estos valores miden el número de desviaciones estándar a las que cada valor se encuentra de la media muestral y corresponde a la escala del eje de la derecha del gráfico de valores atípicos. La prueba de Grubbs, descrita más adelante, está basada en el valor Estudentizado más extremo, que en este caso es igual a 3.479.

Valores Estudentizados Con Supresión – cada valor de los datos es removido de la muestra un a la vez y la media $\bar{x}_{[i]}$ y la desviación estándar $s_{[i]}$ se calculan usando los $n - 1$ valores de los datos restantes. Cada valor de los datos se estandariza entonces con

$$t_i = \frac{x_i - \bar{x}_{[i]}}{s_{[i]}} \quad (2)$$

Estos valores miden el número de desviaciones estándar a las que cada valor se encuentra de la media muestral cuando ese valor de los datos no se incluye en la muestra. Esto es similar al cálculo de los residuos Estudentizados con supresión empleados en los procedimientos de regresión. La importancia de eliminar cada observación antes de estandarizarla es que un fuerte valor atípico, particularmente en una muestra pequeña, puede tener un impacto tan grande en la media y desviación estándar muestrales que no parezca ser inusual (atípico).

Valor Z MAD Modificado - cada valor de los datos es estandarizado por

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD} \quad (3)$$

Estos valores usan la estimación de sigma basada en la desviación absoluta mediana (MAD). Iglewicz and Hoaglin (1993) sugieren que cualquier valor de los datos para el cual $|M_i|$ sea mayor que 3.5 se etiquete como un valor atípico, que es el criterio empleado por StatAdvisor al interpretar los resultados.

Prueba de Grubbs

La sección final de la salida muestra los resultados de una o más pruebas formales para valores atípicos:

Prueba de Grubbs' (asume normalidad)

Estadístico de prueba = 3.47903

Valor-P = 0.0484379

La primera prueba se debe a *Grubbs* y se calcula si $n \geq 3$. También llamada *Prueba de la Desviación Estudentizada Extrema* (ESD), se basa en el mayor valor Estudentizado (sin eliminación) t_{max} . El estadístico de prueba T se calcula de acuerdo con

$$T = \sqrt{\frac{n(n-2)t_{max}^2}{(n-1)^2 - nt_{max}^2}} \quad (4)$$

Se obtiene un valor aproximado de P de dos colas calculando la probabilidad de exceder $|T|$ basada en una distribución t de Student con $n - 2$ grados de libertad y multiplicando el resultado por $2n$. Un pequeño valor de P lleva a la conclusión de que ese punto más extremo es de hecho un valor atípico. Para pequeñas muestras, uno puede en cambio remitirse a Iglewicz y Hoaglin (1993) quienes dan valores al 5% y 1% para t_{max} en el Apéndice A de su monografía, así como para una prueba generalizada que involucre $r > 1$ valores atípicos potenciales.

En los mismos datos, la fila 15 es el punto más extremo, con un valor Estudentizado igual a casi 3.5. Ya que el valor de P es menor que 0.05, ese punto puede ser declarado como valor atípico estadísticamente significativo a un nivel de confianza del 5%. Esta conclusión se hace sujeta al supuesto de la prueba de Grubbs de que todos los demás valores de los datos provienen de una distribución normal.

Prueba de Dixon

Para muestras pequeñas con $4 \leq n \leq 30$, también se realiza la *prueba de Dixon*. Esta prueba comienza ordenando los valores de los datos de menor a mayor. Sea $x_{(j)}$ el j-ésimo valor más pequeño de los datos, las estadísticas se calculan entonces para probar 5 situaciones potenciales:

Situación 1: **1 valor atípico a la derecha**. Calcule:

$$r = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(2)}} \quad (5)$$

Situación 2: **1 valor aberrante a la izquierda**. Calcule:

$$r = \frac{x_{(2)} - x_{(1)}}{x_{(n-1)} - x_{(1)}} \quad (6)$$

Situación 3: **2 valores aberrantes a la derecha**. Calcule:

$$r = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(2)}} \quad (7)$$

Situación 4: **2 valores aberrantes a la izquierda**. Calcule:

$$r = \frac{x_{(3)} - x_{(1)}}{x_{(n-1)} - x_{(1)}} \quad (8)$$

Situación 5: **1 valor aberrante en cualquiera de los dos lados**. Calcule:

$$r = \max \left[\frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}, \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}} \right] \quad (9)$$

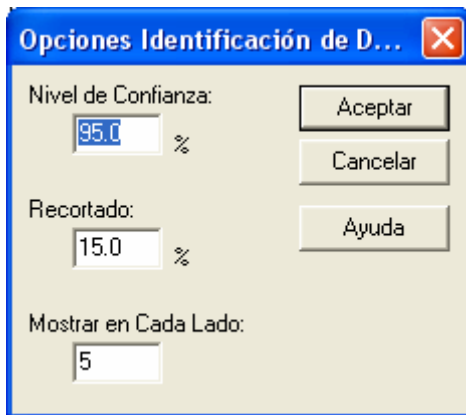
El estadístico calculado r se compara entonces con valores críticos en tablas tales como el Apéndice A.3 de Iglewicz y Hoaglin (1993). Para cada prueba, STATGRAPHICS indica si el resultado es o no estadísticamente significativo a los niveles de 5% y 1%. Un resultado significativo indica la presencia de la situación hipotética.

Por ejemplo, eligiendo arbitrariamente las primeras 30 filas del archivo de datos, se muestra la siguiente tabla:

Prueba de Dixon (asume normalidad)			
	<i>Estadístico</i>	<i>Prueba al 5%</i>	<i>Prueba al 1%</i>
1 aberrante por derecha	0.317073	Significativo	No sig.
1 aberrante por izquierda	0.0	No sig.	No sig.
2 aberrantes por derecha	0.439024	Significativo	Significativo
2 aberrantes por izquierda	0.142857	No sig.	No sig.
1 aberrante en cada lado	0.317073	Significativo	No sig.

Se obtienen resultados significativos al nivel de significancia del 5% para la hipótesis de que existe 1 valor aberrante grande a la derecha, que existen 2 valores aberrantes grandes a la derecha, y que existe 1 valor aberrante grande en cualquiera de los lados. Cuando se use esta prueba, debe seleccionar la hipótesis de interés antes de buscar en los resultados.

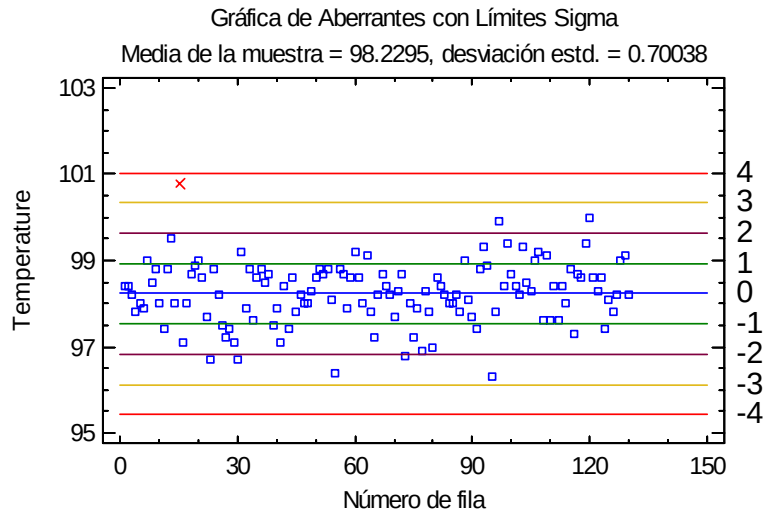
Opciones del Análisis



- **Nivel de Confianza:** nivel empleado para calcular los intervalos de confianza.
- **Recortado:** el porcentaje de datos recortados de cada lado cuando se calcula la media recortada y las estadísticas Winsorizadas.
- **Mostrar en Cada Lado:** el número de valores pequeños y grandes más extremos a incluir en la tabla.

Exclusión de Aberrantes

Los valores de los datos que se determina que son aberrantes pueden ser excluidos gráficamente haciendo clic, con el ratón, sobre los puntos en el *Gráfico de Aberrantes* y luego clic en el botón *Excluir/Incluir* en la barra de herramientas del análisis.



Los puntos excluidos se marcarán con una X y todas las estadísticas a todo lo largo del procedimiento se recalcularán sin esos datos. Por ejemplo, la *Prueba de Grubbs* muestra ahora un Valor de P muy insignificante para los valores más extremos en los datos restantes:

Prueba de Grubbs' (asume normalidad)

Estadístico de prueba = 2.75487

Valor-P = 0.676064

Resumen Estadístico

La ventana del *Resumen Estadístico* calcula un número de diferentes estadísticas que comúnmente se usan para resumir una muestra de n observaciones:

Resumen Estadístico para Temperature	
Recuento	130
Promedio	98.2492
Desviación estándar	0.733183
Coef. de variación	0.746248%
Mínimo	96.3
Máximo	100.8
Rango	4.5
Rango intercuartílico	0.9
Curtosis estandarizada	1.81642

Las estadísticas incluidas por omisión 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*. De particular interés aquí son el sesgo y la curtosis estandarizados. Ambas estadísticas deben estar entre -2 y $+2$ si los datos provienen de una distribución normal. Ya que este es un supuesto de la prueba para valores atípicos, debería revisar estos valores después de excluir los valores atípicos.

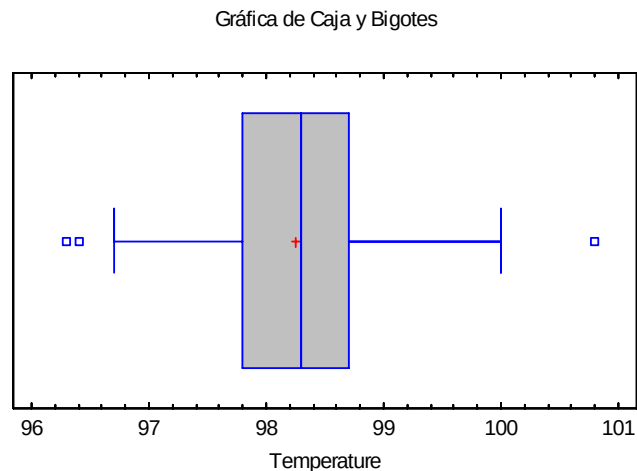
Opciones de Ventana

Opciones Resumen Estadístico			
☒ Promedio	☒ Desviación Estándar	☒ Máximo	☐ Rango Intersextil
☐ Mediana	☒ Coef. de Variación	☒ Rango	☐ Sesgo
☐ Moda	☐ Error Estándar	☐ Cuartil Inferior	☐ Sesgo Estd.
☐ Media Geométrica	☐ Sigma Winsorizada	☐ Cuartil Superior	☐ Curtosis
☐ Media Recortada 5 %	☐ DAM	☒ Rango Intercuartil	☒ Curtosis Estd.
☐ Media Winsorizada	☐ Sbi	☐ 1/6 Sextil	☐ Suma
☐ Varianza	☒ Mínimo	☐ 5/6 Sextil	☐ Suma de Cuadrados
Aceptar Cancelar Todos Ayuda			

Seleccione las estadísticas a presentar. El significado de cada estadística se describe en la documentación para el procedimiento *Análisis de Una Variable*.

Gráfico de Caja y Bigotes

Esta ventana presenta le gráfico de caja y bigotes.



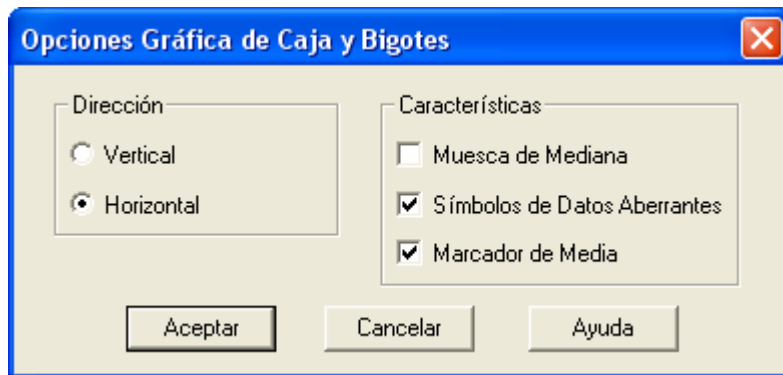
Este gráfico se construye de la siguiente forma:

- 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 los valores de los datos 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 media muestral.
- Los bigotes se dibujan desde los extremos de la caja hasta los valores mínimo y máximo de los datos, 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 los datos de temperatura corporal es muy simétrico. El signo de más para la media se encuentra muy cerca de la línea para la mediana, mientras que los bigotes son aproximadamente de igual longitud. Hay 3 puntos extremos. Cuando se muestrean 130 observaciones de una distribución normal, se puede esperar que se presenten puntos extremos tan solo por azar como la mitad de las veces, pero generalmente sólo uno o dos. Puntos extremos lejanos, de los cuales no hay, se presentan de manera extremadamente rara.

Ventana de Opciones



- **Dirección:** la orientación del gráfico, correspondiente a la dirección de los bigotes.
- **Muesca de Mediana:** si se selecciona, se agregará una muesca al gráfico que muestra un intervalo de confianza de aproximadamente $100(1-\alpha)\%$ para la 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*).
- **Símbolos de Datos Aberrantes:** si se selecciona, indica la localización de los puntos extremos o atípicos.
- **Marcador de Media:** si se selecciona, muestra la localización de la media muestral así como la mediana.

Pruebas de Normalidad

En la ventana *Pruebas de Normalidad* se realizan varias pruebas formales para normalidad y se presentan los resultados.

Pruebas de Normalidad		
Prueba	Estadístico	Valor-P
Chi-Cuadrado	54.0154	0.000424234
Estadístico W de Shapiro-Wilk	0.986473	0.821435
Puntuación Z para asimetría	0.0151112	0.987938
Puntuación Z para curtosis	1.64492	0.0999861

Cada una de las pruebas se basa en el siguiente par de hipótesis:

H₀: los datos provienen de una distribución normal

H_A: los datos no provienen de una distribución normal

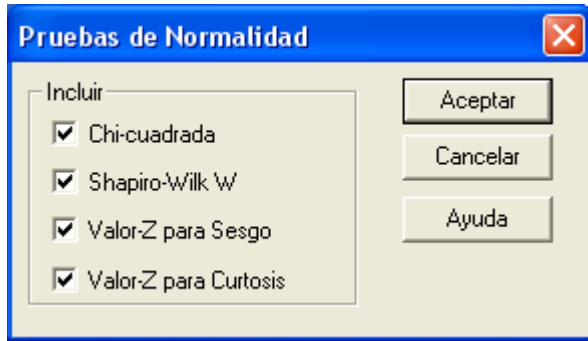
Pequeños valores de P (por debajo de 0.05 si se trabaja al nivel de significancia del 5%) conducen al rechazo de la hipótesis de normalidad.

Las cuatro pruebas, cuyos detalles se dan en la documentación de *Ajuste de Distribuciones (Datos No Censurados)*, son las siguientes:

- **Prueba Chi-Cuadrada** - divide los datos en clases no traslapadas y calcula un estadístico basado en la diferencia entre las frecuencias observadas en cada clase y las frecuencias esperadas si los datos provinieran de una distribución normal. Esta prueba *no* debe usarse si los datos están severamente redondeados, como en el presente ejemplo, ya que la naturaleza discreta de los datos puede fácilmente distorsionar los resultados.
- **W de Shapiro-Wilks** – disponible cuando $2 \leq n \leq 2000$, esta prueba compara el ajuste de mínimos cuadrados de la línea de regresión con los datos en el gráfico de probabilidad normal.
- **Valor Z para el Sesgo** – realiza una prueba basada en el sesgo estimado en los datos.
- **Valor Z para la Curtosis** – realiza una prueba basada en la curtosis estimada en los datos.

Excepto por la prueba de chi-cuadrada, cuyo comportamiento puede ser explicado por el hecho de que los datos se redondearon a la décima de grado más cercana, no hay evidencia para rechazar la hipótesis de que las temperaturas corporales siguen una distribución normal.

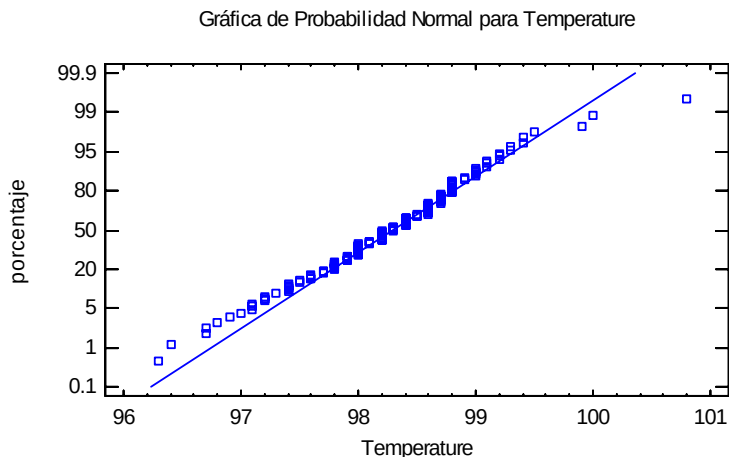
Opciones de Ventana



- **Incluir:** selecciona una o más pruebas para que se lleven a cabo.

Gráfica de Probabilidad Normal

La *Gráfica de Probabilidad Normal* muestra los datos de menor a mayor de manera que es posible juzgar si los datos provienen o no de una distribución normal.



El eje vertical se escala de tal manera que, si los datos provienen de una distribución normal, los puntos deberán caer aproximadamente a lo largo de una línea recta. Para construir el gráfico, los puntos se grafican en las coordenadas

$$\left(x_{(j)}, \Phi^{-1}\left(\frac{j-0.375}{n+0.25}\right) \right) \quad (10)$$

donde $\Phi^{-1}(u)$ representa la distribución normal estándar inversa evaluada en u . Las etiquetas a lo largo del eje vertical son iguales a $100u\%$, para valores de u que van de 0.001 a 0.999.

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 . Hay dos opciones para ajustar la línea:

1. Usando la mediana y los cuartiles muestrales:

$$\hat{\mu} = \text{median muestral} \quad (11)$$

$$\hat{\sigma} = \text{rango intercuartílico} / 1.35 \quad (12)$$

2. Ajustando una regresión por mínimos cuadrados de los cuantiles normales de los valores de los datos ordenados.

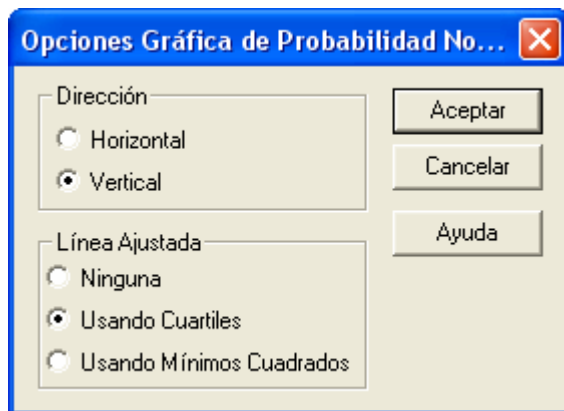
$$\hat{\mu} = - \text{intercepto} / \text{pendiente} \quad (13)$$

$$\hat{\sigma} = 1 / \text{pendiente} \quad (14)$$

El primer método es más robusto a desviaciones de la normalidad en las colas de la distribución, ya que esencialmente se apoya solo en la mitad central. Valores atípicos o colas largas tendrán una mayor influencia al usar el método de mínimos cuadrados.

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*.

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.

Salvar Resultados

El botón *Salvar Resultados* en la barra de herramientas permite que se puedan salvar los siguientes resultados en la hoja de datos:

1. **Datos Winsorizados** – los datos después de la “winsorización”. El porcentaje especificado de los valores menores y mayores habrá sido reemplazado con los valores más extremos no recortados.
2. **Banderitas de selección** – una columna que contiene un 0 para los valores que haya excluido del análisis manualmente usando la característica *Excluir* en el *Gráfica de Aberrantes*, y un 1 para todos los demás valores. En otros procedimientos, ingrese el nombre de esta columna en el campo *Selección* para excluir del análisis automáticamente los mismos valores.
3. **Valores Estudentizados (sin supresión)** – los valores estandarizados de los datos basados en estadísticas muestrales para todas las observaciones.
4. **Valores Estudentizados (con supresión)** – los datos estandarizados basados en la media y la desviación estándar calculadas después de eliminar la observación.
5. **Puntajes Z Modificados** – los datos estandarizados basados en la mediana muestral y la estimación MAD de sigma.

Cálculos**Desviación Absoluta Mediana**

$$DAM = \text{mediana}_i \left\{ |x_i - \tilde{x}| \right\} \quad (15)$$

Media Recortada en 100α%

$$T(\alpha) = \frac{1}{n(1-2\alpha)} \left[k(x_{(r+1)} + x_{(n-r)}) + \sum_{i=r+2}^{n-r-1} x_{(i)} \right] \quad (16)$$

donde $r = \lfloor \alpha n \rfloor$ y $k = 1 - (\alpha n - r)$.

Media Winsorizada en 100α%

$$T_w = \frac{1}{n} \left\{ \sum_{i=r+1}^{n-r} x_{(i)} + r[x_{(r+1)} + x_{(n-r)}] \right\} \quad (17)$$

Sbi

$$S_{bi} = \frac{\sqrt{n \sum_{i=1}^n (x_i - \tilde{x})^2 (1 - u_i^2)^4}}{\left| \sum_{i=1}^n (1 - u_i^2)(1 - 5u_i^2) \right|} \quad (18)$$

donde

$$u_i = \frac{x_i - \tilde{x}}{9DAM} \quad (19)$$

Sigma Winsorizada

$$S_w = \sqrt{\frac{n \left\{ \sum_{i=r+1}^{n-r} (x_{(i)} - T_w)^2 + r[(x_{(r+1)} - T_w)^2 + (x_{(n-r)} - T_w)^2] \right\}}{(n-2r)(n-2r-1)}} \quad (20)$$

Intervalo de confianza Winsorizado

$$T_w \pm t_{n-2r-1, \alpha/2} \frac{S_w}{\sqrt{n}} \quad (21)$$