EAN Escuela de **Administración de Negocios**

ANALIZAR

Datos NO son normales

Transformación de Datos

Si los datos no son normales, se pueden tratar de transformar con alguna función para normalizarlos utilizando el Método de Box Cox o el método de Johnson

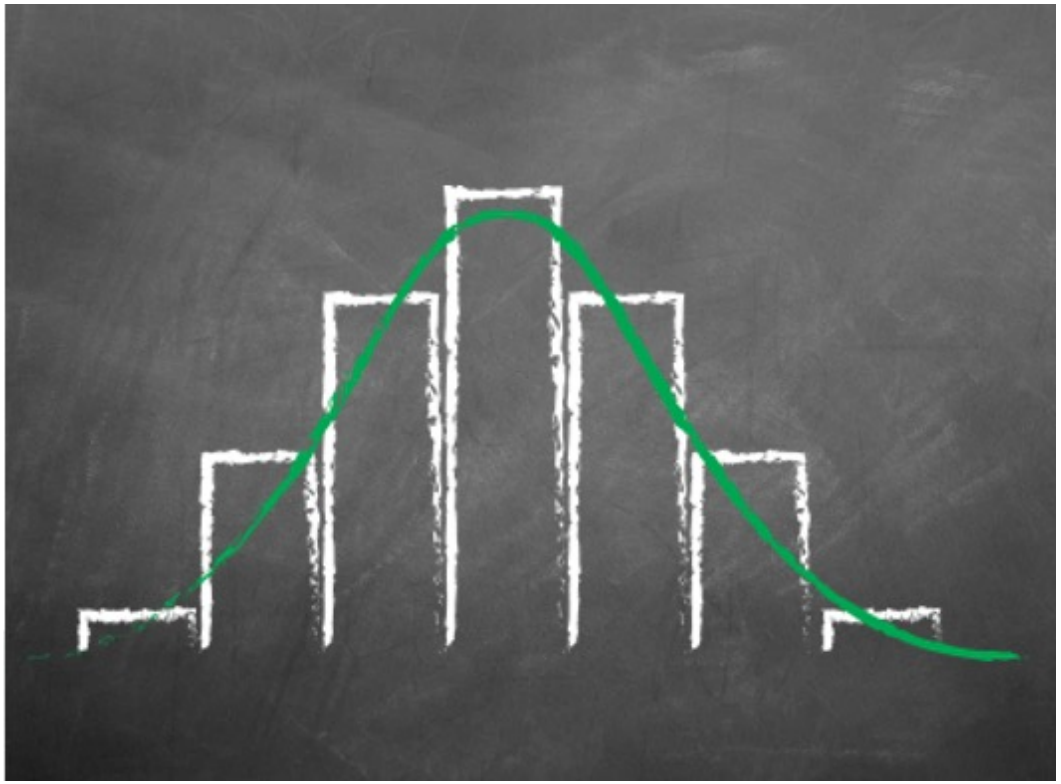
Ejercicio Call Center

En un call center se desea detarminar la capacidad del centro de atención de llamadas de servicio al cliente por lo que se desea determinar el tiempo promedio y la desviación estándar de las llamadas. A continuación se presentan los datos tomados de una muestra de 50 llamadas aleatorias para tratar de inferir los parámetros poblacionales.

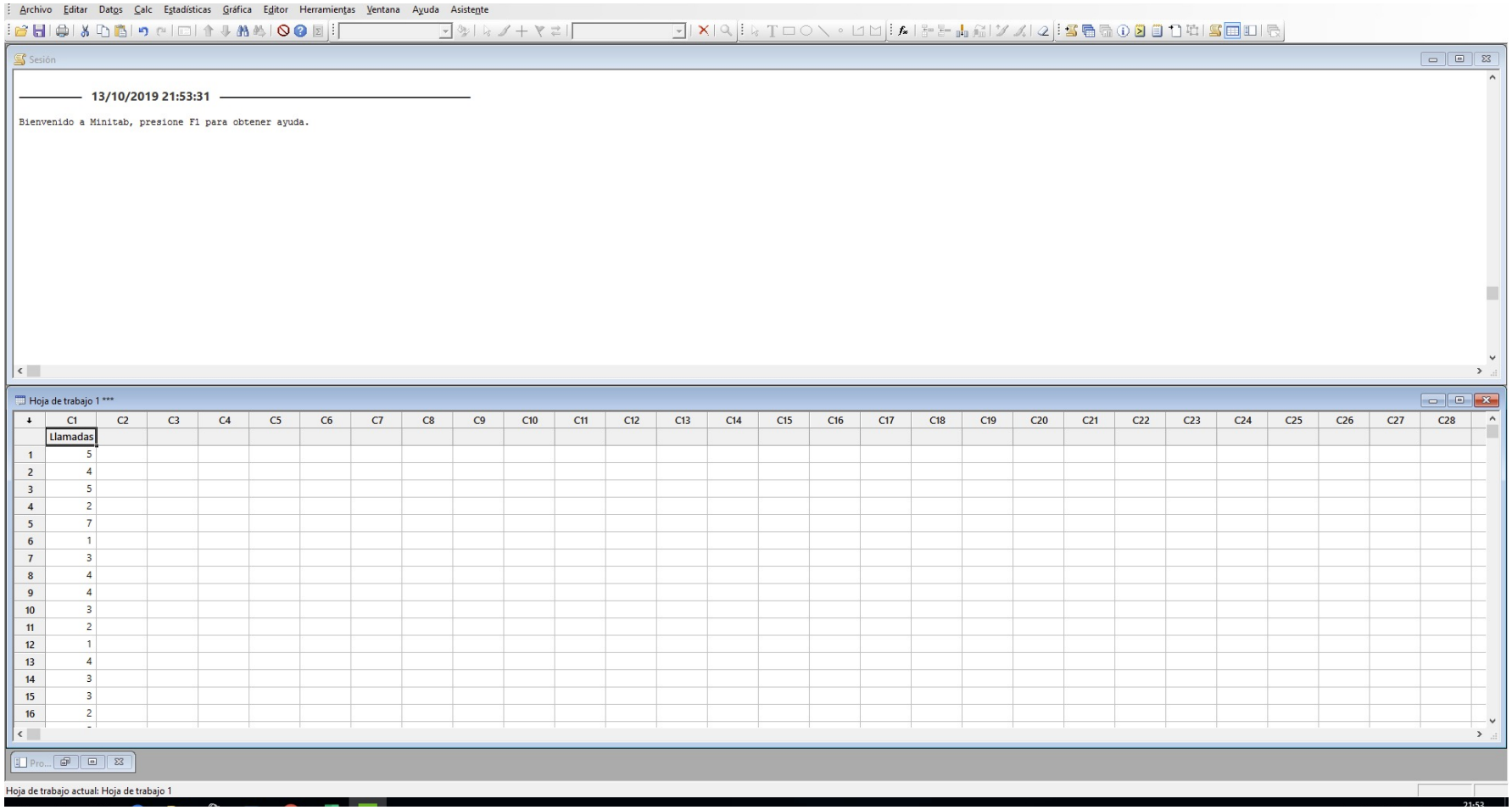
Tiempo de las llamadas				
5	2	3	2	4
4	1	4	3	2
5	4	5	5	1
2	3	3	4	3
7	3	4	4	1
1	2	6	4	1
3	5	2	1	1
4	5	2	5	2
4	2	2	2	2
3	1	6	4	4

1er paso determinar la normalidad de los datos

- Histograma
- Prueba de Bondad de Ajuste



Abrir el archivo Sesion 10 Analizar y copiar los datos a Minitab.



2do paso Transformar los datos.

- Box Cox
- Johnson

Transformación de Box Cox

El Método de Box Cox encuentra un exponente λ al que se deben elevar los datos.

Está limitada a datos positivos. Asume que los datos están en subgrupos para permitir un análisis "dentro del subgrupo." La transformación de potencia de Box-Cox está dada por:

$$x(\lambda) = \frac{x^2 - 1}{\lambda}; \lambda \neq 0$$

$$x(\lambda) = \ln(\lambda); \ \lambda = 0$$

Valor de λ	Transformación
$\lambda = -1.0,$	$x_i(\lambda) = \frac{1}{x_i}$
$\lambda = -0.5,$	$x_i(\lambda) = \frac{1}{\sqrt{x_i}}$
$\lambda = 0.0,$	$x_i(\lambda) = ln(x_i)$
$\lambda = 0.5,$	$x_i(\lambda) = \sqrt{x_i}$
$\lambda = 2.0,$	$x_i(\lambda) = x_i^2$

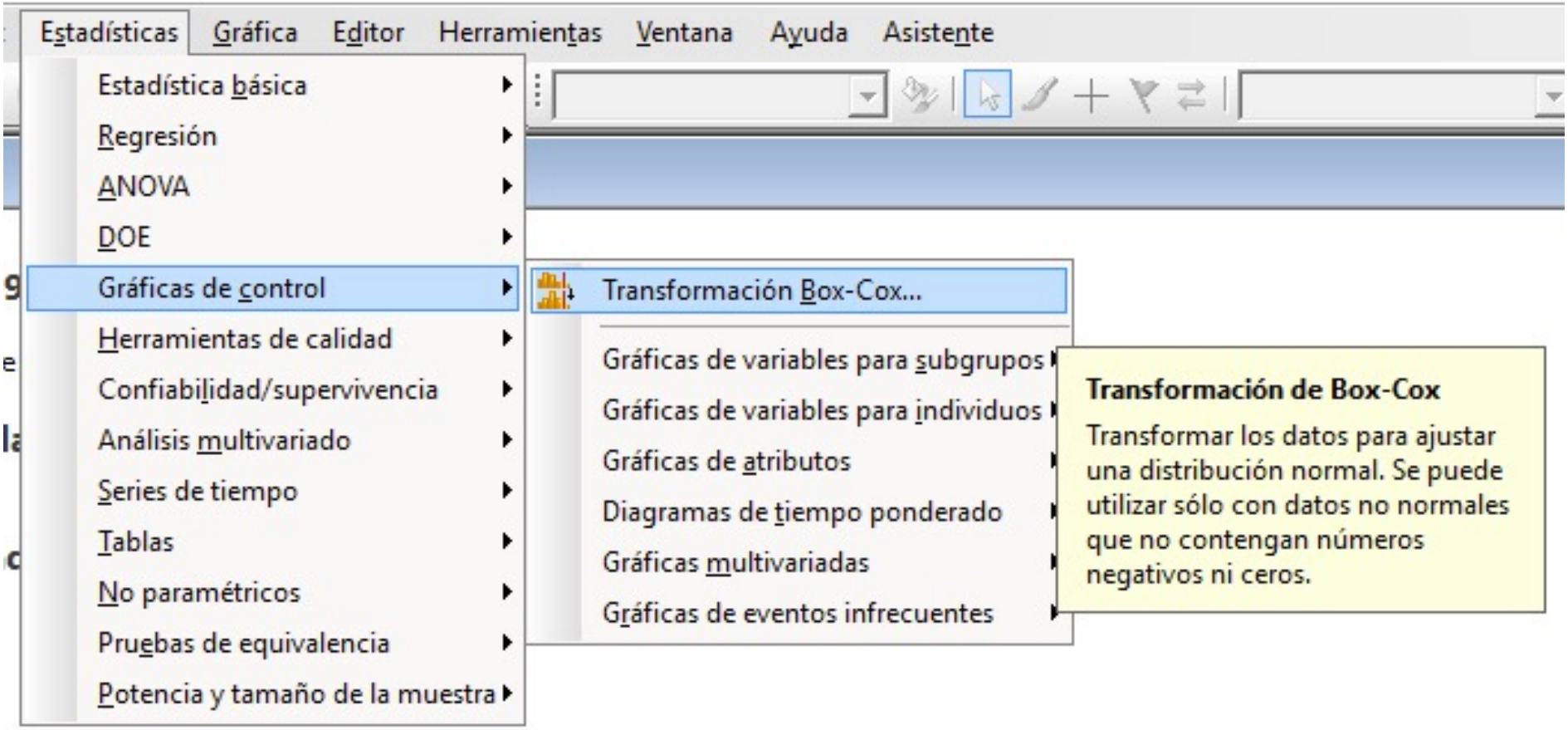
Procesos No normales

Dadas las observaciones X1, X2, …Xn, seleccionar la potencia λ que maximice el logaritmo de la función de máxima verosimilitud.

$$f(x) = -\frac{n}{2}\ln\left[\sum_{i=1}^{n}\frac{(x_i(\lambda)-\bar{x}(\lambda))^2}{n}\right]+(\lambda-1)\sum_{i=1}^{n}\ln(x_i)$$

Con la media de los datos transformados dada por:

$$\bar{x}(\lambda) = \frac{1}{n}\sum_{i=1}^{n} x_i(\lambda)$$


Estadísticas
Gráfica
Editor
Herramientas
Ventana
Ayuda
Asistente
Estadística básica
Regresión
ANOVA
DOE
Gráficas de control
Herramientas de calidad
Confiabilidad/supervivencia
Análisis multivariado
Series de tiempo
Tablas
No paramétricos
Pruebas de equivalencia
Potencia y tamaño de la muestra
Transformación Box-Cox...
Gráficas de variables para subgrupos
Gráficas de variables para individuos
Gráficas de atributos
Diagramas de tiempo ponderado
Gráficas multivariadas
Gráficas de eventos infrecuentes
Transformación de Box-Cox
Transformar los datos para ajustar una distribución normal. Se puede utilizar sólo con datos no normales que no contengan números negativos ni ceros.

Valor de λ	Transformación
$\lambda = -1.0,$	$x_i(\lambda) = \frac{1}{x_i}$
$\lambda = -0.5,$	$x_i(\lambda) = \frac{1}{\sqrt{x_i}}$
$\lambda = 0.0,$	$x_i(\lambda) = ln(x_i)$
$\lambda = 0.5,$	$x_i(\lambda) = \sqrt{x_i}$
$\lambda = 2.0,$	$x_i(\lambda) = x_i^2$

Transformación de Yeo-Johnson

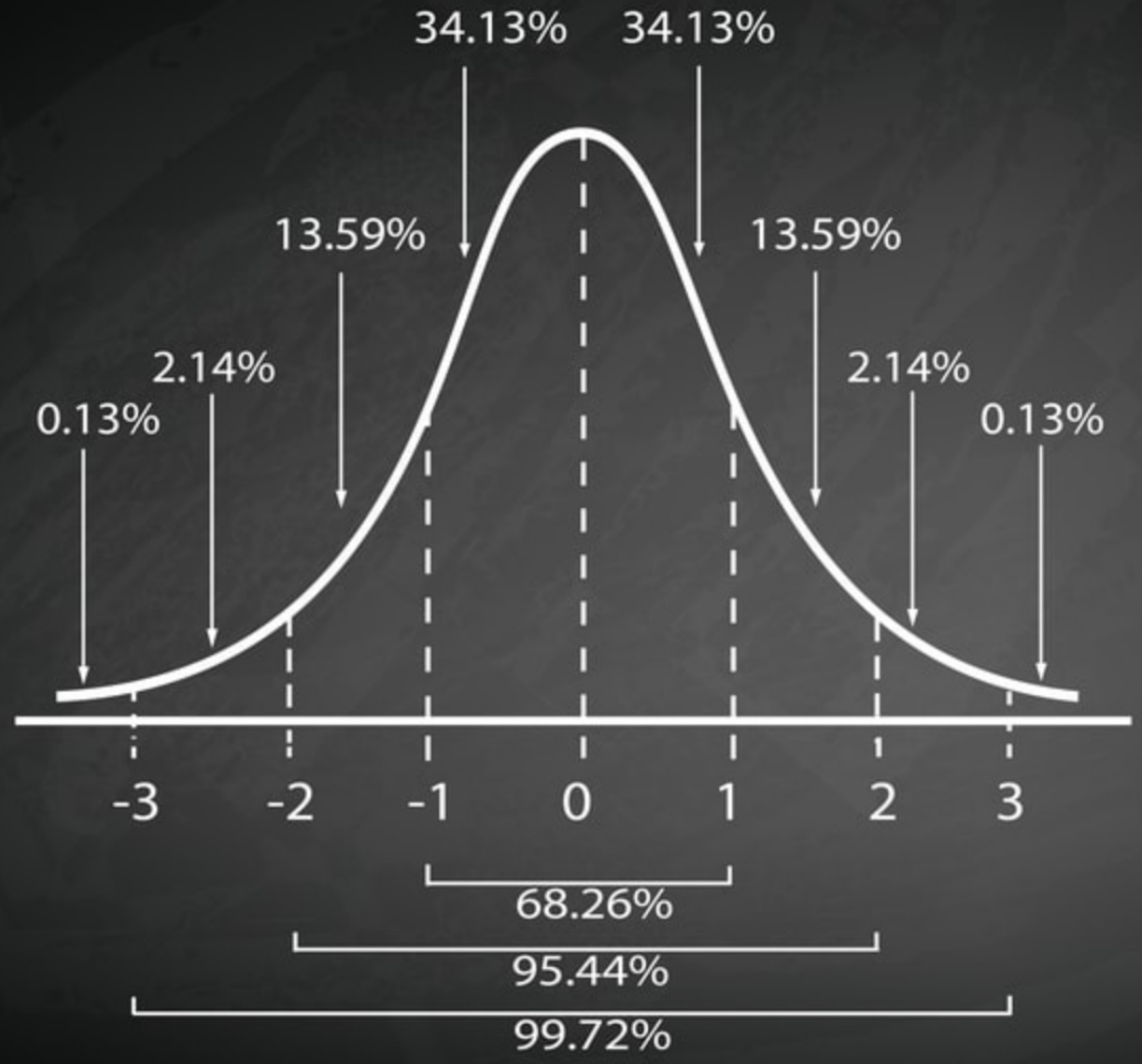
A diferencia de Box Cox, la transformación de Johnson permite también valores de cero y/o negativos.

λ puede ser cualquier número real, donde λ = 1 produce la transformación identidad. La ley de transformación es:

$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^{\lambda} - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -[(-y_i + 1)^{(2-\lambda)} - 1]/(2 - \lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

3er paso Determinar la normalidad.

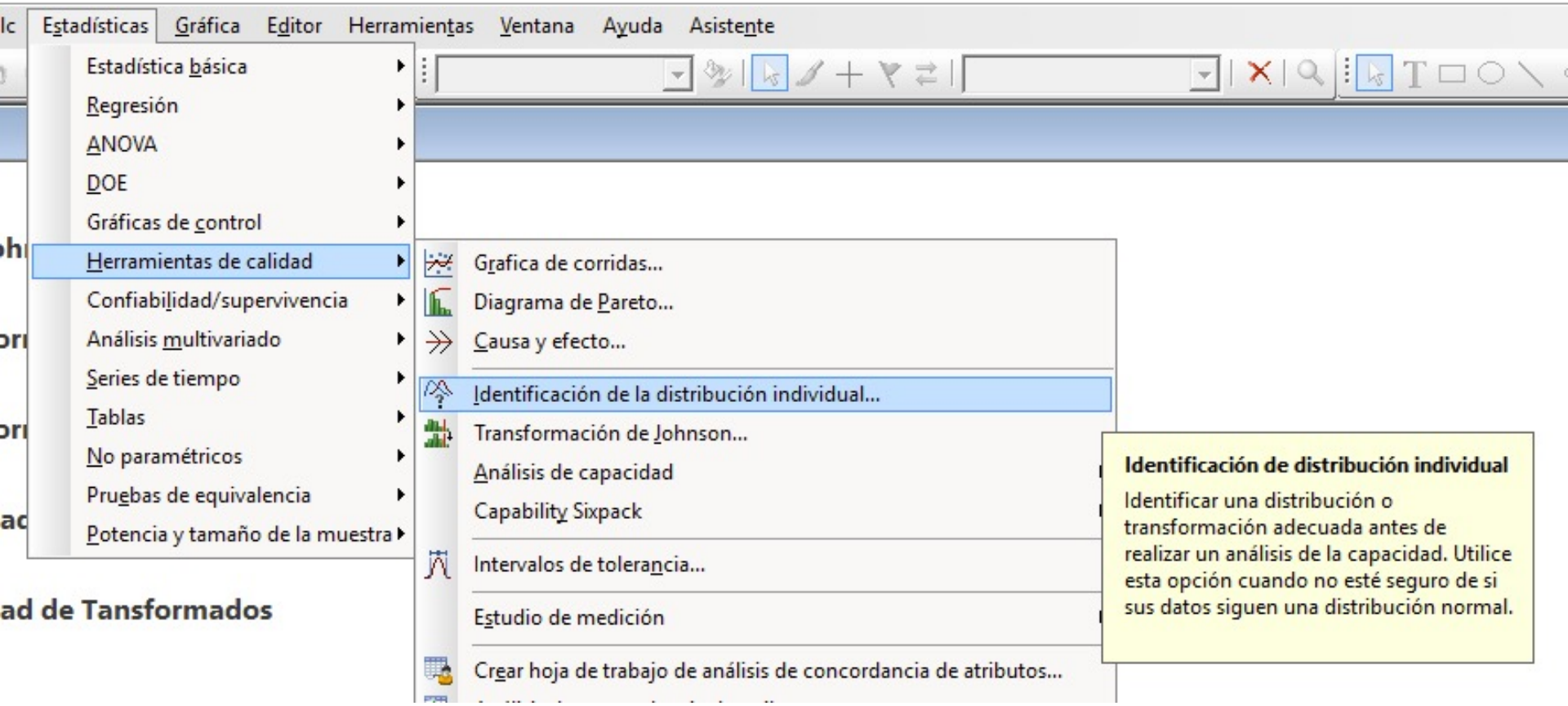
- Histograma
- Bondad de Ajuste



Datos No Normales

- Identificar Tipo de Distribución
- Pruebas No Paramétricas

Identificar el tipo de distribución que siguen los datos.

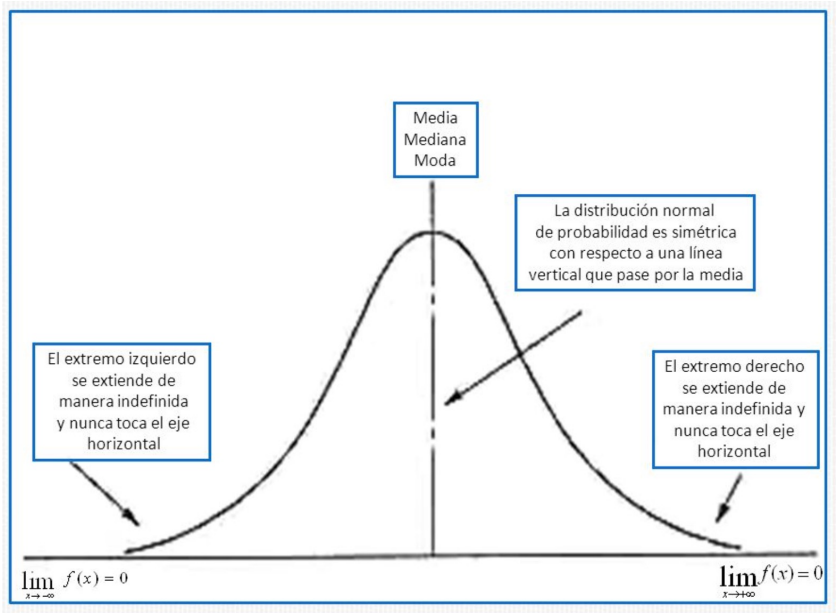


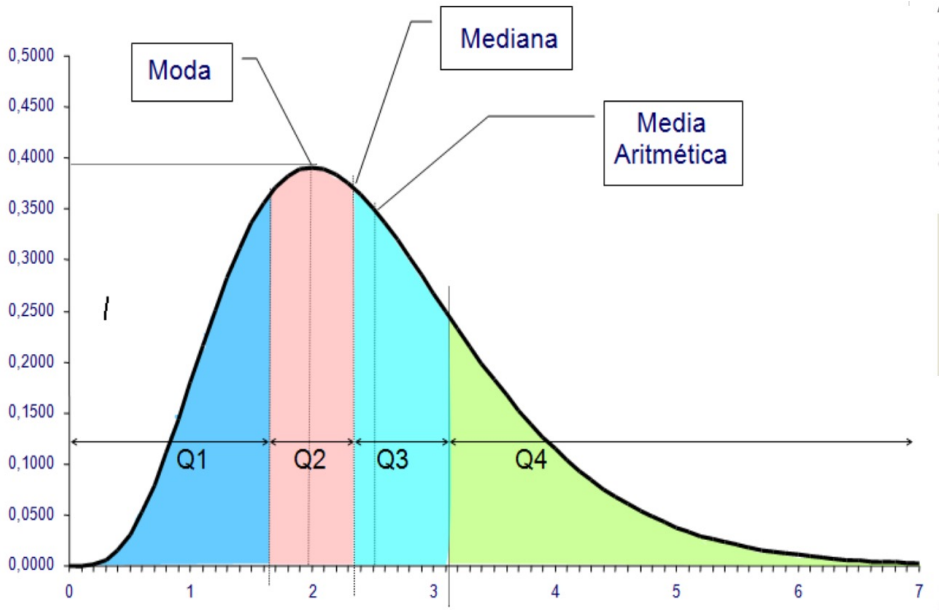
Pruebas NO paramétricas

Pruebas No Paramétricas

Existen situaciones en las que las suposiciones no se pueden justificar pues existe duda en cuanto su aplicación, como por ejemplo cuando una población es muy sesgada. Para estos casos se ha elaborado diversas pruebas y métodos que son independientes de las distribuciones poblacionales y los parámetros asociados a ellas.

¿Por qué usamos las mediana en lugar de la media en las pruebas no paramétricas?





Prueba de los Signos

Se usa para hacer pruebas de hipótesis acerca de la mediana de una población.

Ho: La Mediana poblacional es igual a un valor dado.

Ha: La Mediana es menor del valor dado.

La prueba estadística está basada en la distribución Binomial con probabilidad de éxito p=½, puesto que la probabilidad de que un dato sea mayor o menor que la mediana es ½. Para calcularla se determinan las diferencias de los datos con respecto al valor dado de la mediana y se cuentan los signos positivos y negativos.

Ejercicio tiempo de vida de una pila

Una empresa afirma que el tiempo de vida de un tipo de pila que fabrica es mayor a 250 horas. Se mide los tiempos de vida de 24 pilas, los resultados se presentan a continuación:

271	230	198	275	282	225	284	219
253	216	262	288	236	291	253	224
264	295	211	252	294	243	272	268

Suponiendo que la muestra es aleatoria determinar si la afirmación de la empresa es verdadera a un nivel de significancia de 0,05

La Prueba de Rangos con Signos de Wilcoxon

Es usada para hacer pruebas de hipótesis acerca de la mediana.

La prueba estadística se basa en el estadístico de Wilcoxon (1945), el cual se calcula de la siguiente manera:

• Se resta de cada dato el valor de la mediana que se considera en la hipótesis nula.

• Se calcula los rangos de las diferencias sin tomar en cuenta el signo de las mismas (o sea en valor absoluto). En el caso de haber empate se asigna un rango promedio a todas las diferencias empatadas es decir; se les asigna el rango:

$$\frac{\text{menor rango del grupo del empate} + \text{mayor rango del grupo del empate}}{\text{número de empates}}$$

El estadístico W de Wilcoxon será la suma de los rangos correspondientes a las diferencias positivas.

La Prueba de Mann-Withney para dos muestras independientes

Se usa cuando se quiere comparar dos poblaciones usando muestras independientes, es decir; es una prueba alterna a la prueba de t para comparar dos medias usando muestras independientes. También es conocida como la prueba de suma de rangos de Wilcoxon. La hipótesis nula es que la mediana de las dos poblaciones son iguales y la hipótesis alternativa puede ser que la mediana de la población 1 sea mayor que la mediana de la población 2. Cuando tanto n_1 como n_2 sean mayores que 10, se puede demostrar que si no hay empates, entonces W se distribuye aproximadamente como una normal con media $n1(n_1+n_2+1)/2$ y varianza $n_1n_2(n_1+n_2+1)/12$.

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1n_2(n_1+n_2+1)}{12}}} \sim N(0,1)$$

Cuando hay empates entonces la varianza es modificada y se obtiene:

$$z = \frac{W - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2}{12}\left[n_1+n_2+1-\sum_{i=1}^{g}\frac{t_i^3-t_i}{(n_1+n_2)(n_1+n_2-1)}\right]}} \sim N(0,1)$$

Donde, g y t_i tienen el mismo significado dado anteriormente.

Ejercicio resistencia de cables

En la tabla siguiente se muestra la resistencia de cables que se hicieron de dos aleaciones diferentes, I y II. En la tabla se tienen 8 cables de la aleación I y 10 de la aleación II. Se desea decidir si hay una diferencia o no entre las muestras.

ALEACIÓN 1				ALEACIÓN 2				
18.3	16.4	22.7	17.8	12.6	14.1	20.5	10.7	15.9
18.9	25.3	16.1	24.2	19.6	12.9	15.2	11.8	14.7

La Prueba de Kruskal-Wallis para comparar más de dos grupos

La prueba de Kruskal-Wallis, es una alternativa a la prueba F del análisis de varianza para diseños de clasificación simple. En este caso se comparan varios grupos pero usando la mediana de cada uno de ellos, en lugar de las medias.

Ho: La mediana de las k poblaciones consideradas son iguales

Ha: Al menos una de las poblaciones tiene mediana distinta a las otras.

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

La Prueba de Kruskal-Wallis para comparar más de dos grupos

Si hay empates en los datos entonces, se aplica la siguiente modificación a H:

$$H' = \frac{H}{1 - \frac{\sum_{i=1}^{g} t_i^3 - t_i}{n^3 - n}}$$

Se puede mostrar que si los tamaños de cada grupo son mayores que 5 entonces, H se distribuye como una Ji-Cuadrado con, k-1 grados de libertad. Luego, la hipótesis nula se rechaza si el valor es menor a 0.05.

Para hacer la prueba de Kruskal-Wallis en MINITAB, los datos de la variable cuantitativa deben ir en una columna y los niveles del factor en otra. No se permite en este caso ingresar los grupos en columnas separadas.

Ejercicio compra de máquinas

Una empresa desea comprar una de cinco máquinas: A, B, C, D o E. En un experimento diseñado para determinar si hay diferencia en el desempeño entre las máquinas, cinco operadores experimentados trabajan en ellas durante tiempos iguales. La tabla muestra el número de las unidades que produce cada máquina. Probar la hipótesis de que no existe diferencia a un nivel de significancia de 0,05.

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Ordenamos en columnas y copiamos los datos al Minitab

A	68	72	77	42	53
B	72	53	63	53	48
C	60	82	64	75	72
D	48	61	57	64	50
E	64	65	70	68	53

Datos	Máquina
68	A
72	A
77	A
42	A
53	A
72	B
53	B
63	B
53	B
48	B
60	C
82	C
64	C
75	C
72	C
48	D
61	D
57	D
64	D
50	D
64	E
65	E
70	E
68	E
53	E

Coeficiente de Correlación de Spearman

Este coeficiente es el equivalente no paramétrico del Coeficiente de Correlación, al que también se le llama Coeficiente de Pearson. Al igual que el coeficiente de correlación, el Coeficiente de Spearman puede tomar valores entre -1.0 y 1.0, un valor de -1.0 indica una correlación negativa perfecta y un valor de 1.0 indica una correlación positiva perfecta.

En donde n es el número de rangos

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

Coeficiente de Correlación de Spearman

La correlación de Spearman mide el grado de asociación entre dos variables cuantitativas que siguen una tendencia siempre creciente o siempre decreciente. Es más general que el Coeficiente de correlación de Pearson, la correlación de Spearman, en cambio se puede calcular para relaciones exponenciales o logarítmicas entre las variables.

Ejercicio Bienes Raíces

Calcular el coeficiente de Spearman para los siguientes datos y compárelo con el coeficiente de Pearson:

Años como corredor	3	4	6	7	8	12	15	20	22	26
Casas Vendidas	9	12	16	19	23	119	34	37	40	45

¿Qué diferencias encuentra?