

Curso  
**Métodos Estadísticos**  
**Guía MINITAB Módulo 3**

*Nombre del profesor: MSc. Sofía Arbuola Briceño*  
*Correo electrónico: s.arbuola@gmail.com*

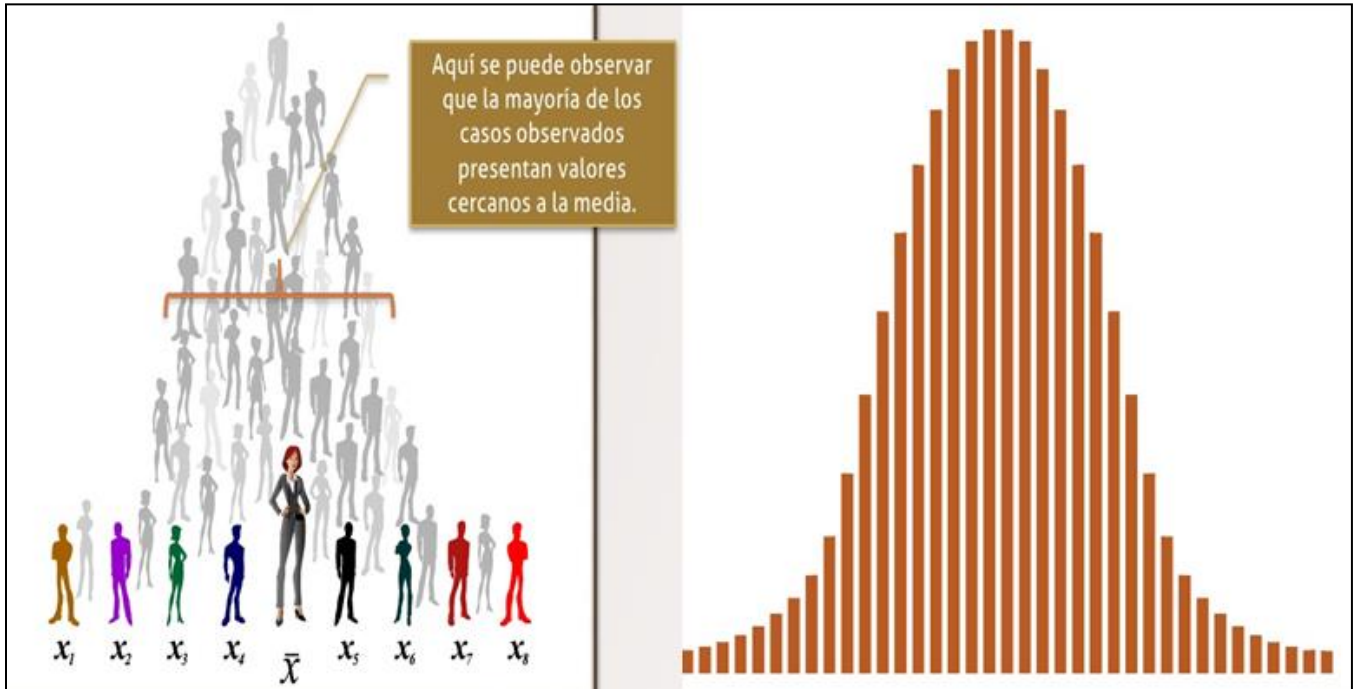


# ÍNDICE

DISTRIBUCIÓN NORMAL DE LOS DATOS .....	3
PROBABILIDADES DE OCURRENCIA DE UN EVENTO .....	4
PRUEBA DE NORMALIDAD DE LOS DATOS .....	5
▪ DATOS CON UNA DISTRIBUCIÓN NORMAL .....	5
▪ DATOS CON UNA DISTRIBUCIÓN <b>NO</b> NORMAL .....	7
▪ IDENTIFICACIÓN DE LA DISTRIBUCIÓN DE PROBABILIDAD PARA DATOS <b>NO</b> NORMALES .....	8
▪ ¿QUÉ HACER CON DATOS NO NORMALES? .....	10

## Distribución normal de los datos

La forma en cómo se distribuyen las probabilidades de ocurrencia de un determinado fenómeno ha generado modelos que facilitan mecanismos de estimación. Uno de los modelos cuyo uso se ha extendido es el de la **distribución normal**, cuya forma intuitiva se representa a continuación:

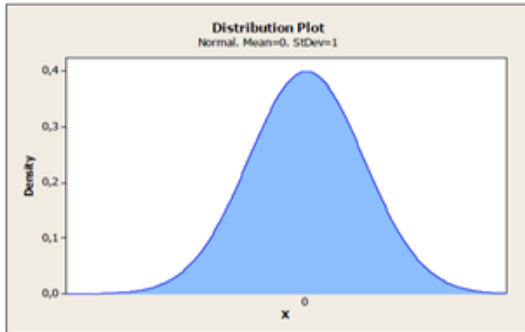


## Probabilidades de ocurrencia de un evento

El área bajo la curva es la probabilidad de que ocurra un evento. Tomando cualquier valor "X" se podrá conocer el área de probabilidad.

# Estadística Inferencial

## Unidades de desviación estándar (Z)



$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

- El área bajo la curva es la probabilidad de que ocurra un evento.
- Va de  $-\infty$  hasta  $+\infty$  la máxima probabilidad del área bajo la curva.
- Cualquier problema referente a una variable normal "X" se transfiere a una Tabla Normal Estándar.

## Prueba de normalidad de los datos

Permite determinar si los datos provienen de una distribución normal. Su importancia radica en la metodología de inferencia estadística.

# Estadística Inferencial

Datos con distribución normal  
=  
P-value  $\geq 0.05$

Si el valor p es mayor o igual que un nivel de significancia elegido (por lo general 0.05 o 0.10) entonces acepte de que los datos provienen de una distribución normal.

### Prueba de Normalidad

Hay 3 tipos de test de normalidad:

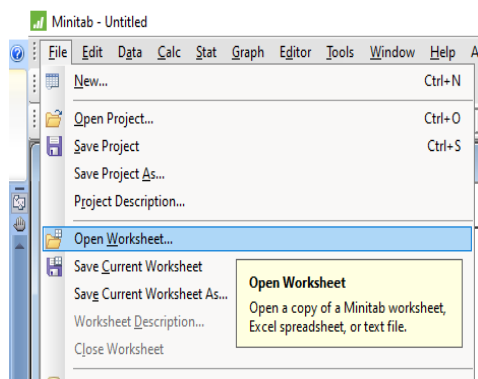
- Anderson –Darling: discrimina muy bien desde el centro de los datos hasta las colas.
- Ryan- Joiner: Discrimina muy bien los centros pero no las colas.
- Smirnov: Discrimina muy bien los centros pero no las colas, necesita muchos datos más de 100 para dar buena confiabilidad.

Diagrama de una distribución normal con una línea vertical central. Las áreas bajo la curva a la izquierda y a la derecha de la línea central están sombreadas y etiquetadas como  $\alpha/2$ . El área central no sombreada está etiquetada como  $(1 - \alpha)$  y "Zona de Aceptación". Los puntos en el eje horizontal están etiquetados como  $-Z_0$  y  $Z_0$ .

- Datos con una distribución normal

1-Cargamos la base de datos.

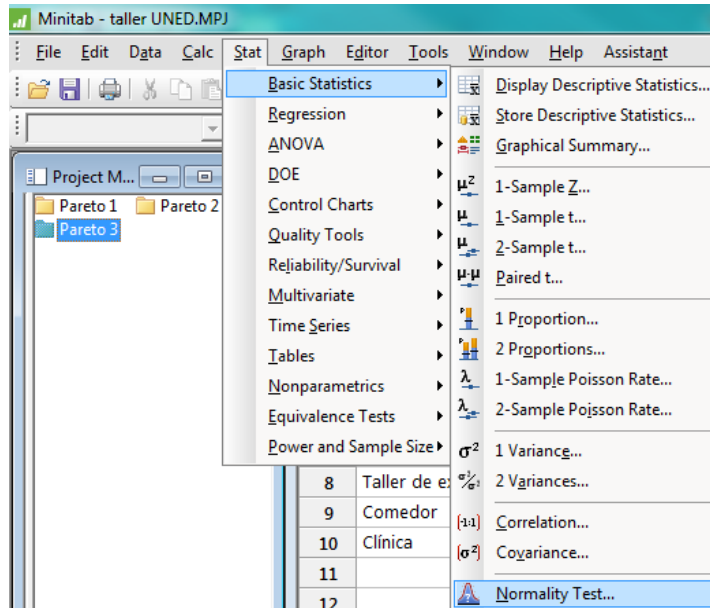
File->Open Worksheet->Seleccionar la base de datos



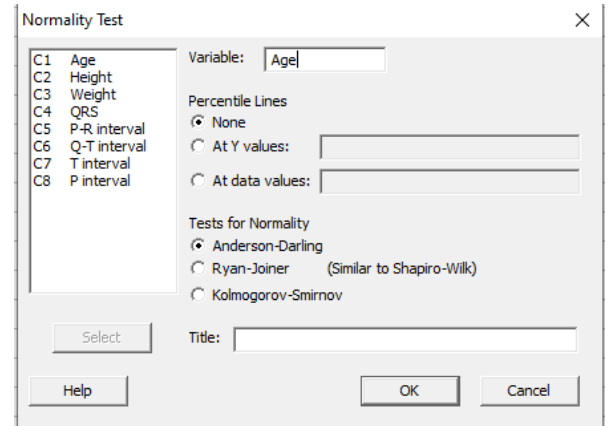
Base de datos:

	C1	C2	C3	C4	C5	C6	C7
	Age	Height	Weight	QRS	P-R interval	Q-T interval	T interval
1	75	190	80	91	193	371	174
2	56	165	64	81	174	401	149
3	54	172	95	138	163	386	185
4	55	175	94	100	202	380	179
5	75	190	80	88	181	360	177
6	13	169	51	100	167	321	174
7	40	160	52	77	129	377	133
8	49	162	54	78	0	376	157
9	44	168	56	84	118	354	160
10	50	167	67	89	130	383	156

2- Seleccionamos la función de Prueba de Normalidad  
 Stat→Basic Statistics→Normality Test

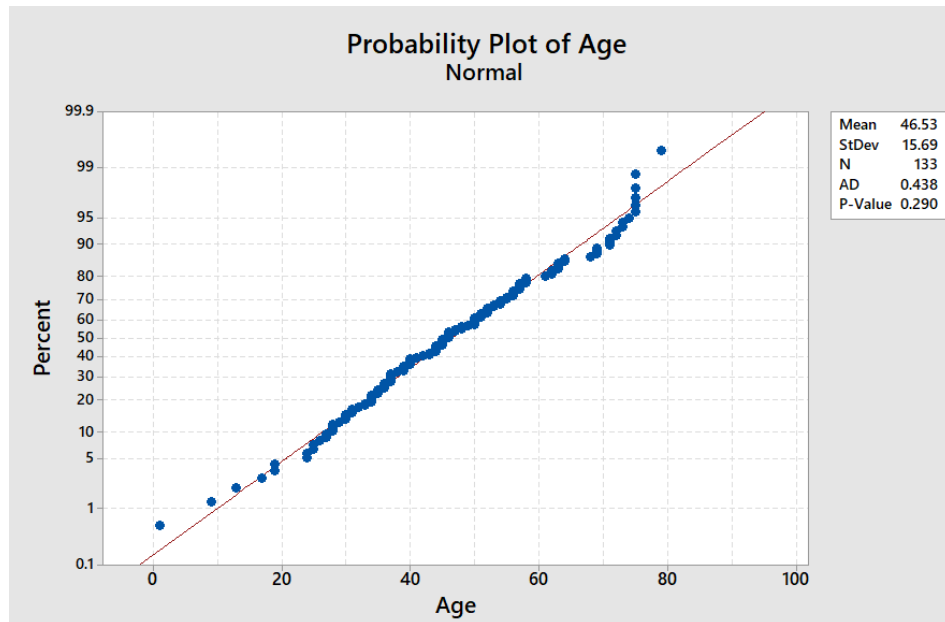


3-Se selecciona la variable en este ejemplo  
 “Age”



4-Se obtiene el gráfico de probabilidad para la variable “Age”:

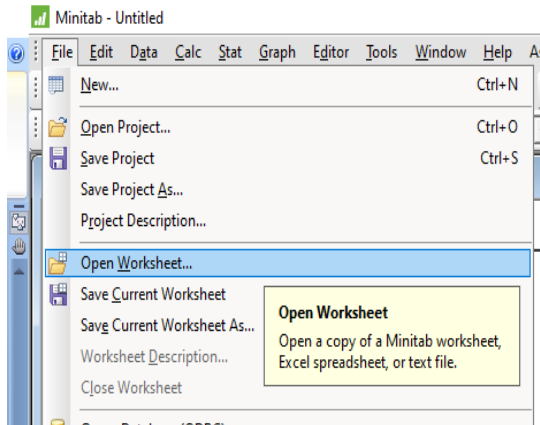
En este ejemplo el valor de p (P-value) es de 0.290 el cual es mayor a 0.05, por lo que los datos siguen una distribución normal. Esto me indica que la metodología de inferencia estadística será al de una distribución normal.



- Datos con una distribución NO normal

1-Cargamos la base de datos.

File->Open Worksheet->Seleccionar la base de datos

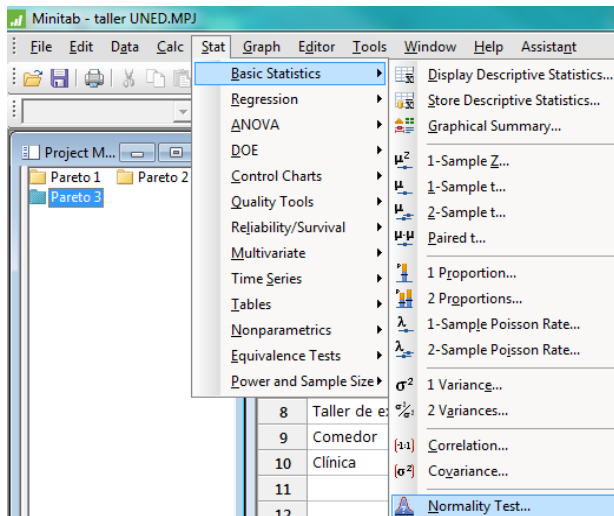


Base de datos:

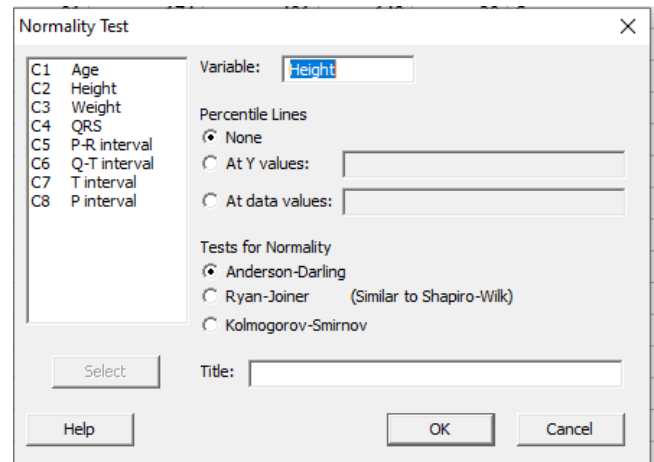
	C1	C2	C3	C4	C5	C6	C7
	Age	Height	Weight	QRS	P-R interval	Q-T interval	T interval
1	75	190	80	91	193	371	174
2	56	165	64	81	174	401	149
3	54	172	95	138	163	386	185
4	55	175	94	100	202	380	179
5	75	190	80	88	181	360	177
6	13	169	51	100	167	321	174
7	40	160	52	77	129	377	133
8	49	162	54	78	0	376	157
9	44	168	56	84	118	354	160
10	50	167	67	89	130	383	156

2- Seleccionamos la función de Prueba de Normalidad

Stat->Basic Statistics->Normality Test

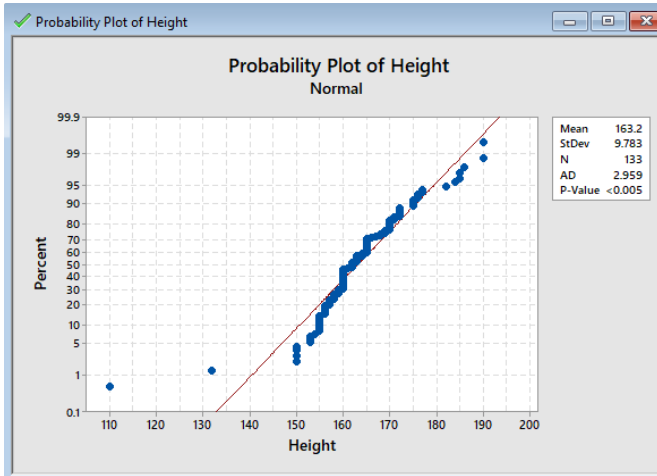


3-Se selecciona la variable en este ejemplo "Height"

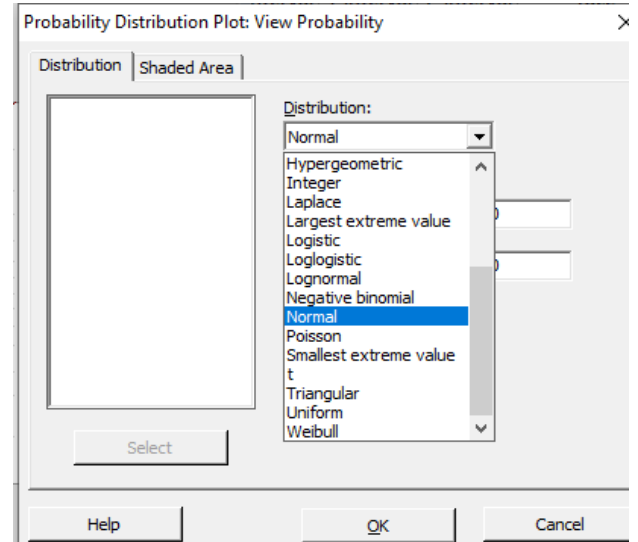


4-Se obtiene el gráfico de probabilidad para la variable "Height":

En este ejemplo el valor de p (P-value) es menor que 0.05, por lo que los datos siguen una distribución NO normal. Esto me indica que la metodología de inferencia estadística NO será al de una distribución normal.



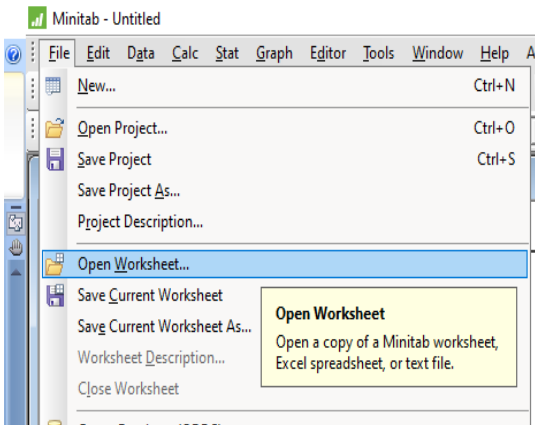
5-En este caso al tener un comportamiento de una distribución NO normal, debemos seleccionar otro tipo de distribución para el análisis de los datos:



▪ **Identificación de la distribución de probabilidad para datos NO normales**

1-Cargamos la base de datos.

File->Open Worksheet->Seleccionar la base de datos



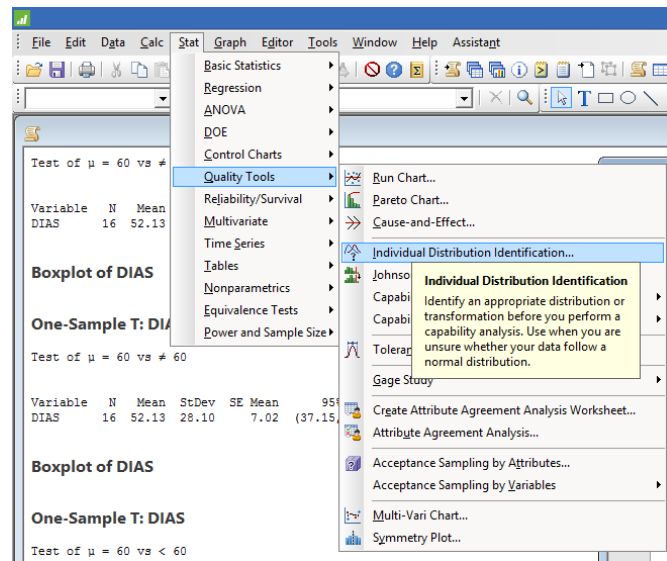
Base de datos:

	C1	C2	C3	C4	C5	C6	C7
	Age	Height	Weight	QRS	P-R interval	Q-T interval	T interval
1	75	190	80	91	193	371	174
2	56	165	64	81	174	401	149
3	54	172	95	138	163	386	185
4	55	175	94	100	202	380	179
5	75	190	80	88	181	360	177
6	13	169	51	100	167	321	174
7	40	160	52	77	129	377	133
8	49	162	54	78	0	376	157
9	44	168	56	84	118	354	160
10	50	167	67	89	130	383	156

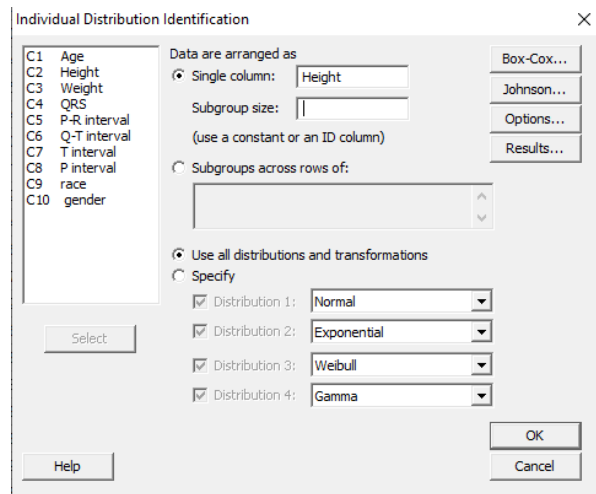


3-Seleccionamos la función de Identificación de la distribución de probabilidad

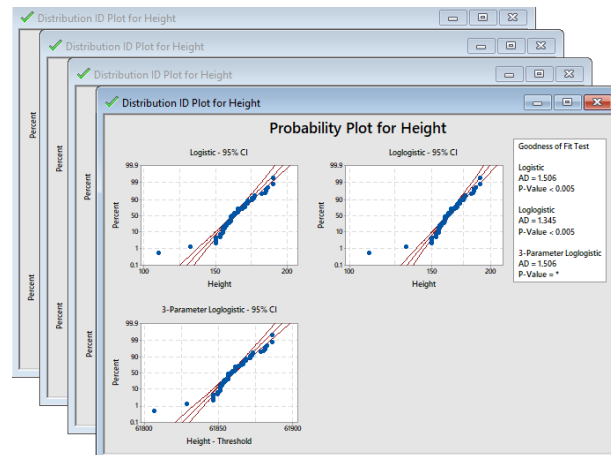
Stat → Quality Tools → Individual Distribution Identification



3-Se selecciona la variable en este ejemplo “Height”, con la opción de utilizar todas las distribuciones de probabilidad.



4-Se obtiene todos los gráficos de probabilidad para la variable “Height”:



¿Qué es el estadístico de Anderson-Darling (AD)?

El estadístico Anderson-Darling mide qué tan bien siguen los datos una distribución específica. Para un conjunto de datos y distribución en particular, mientras mejor se ajuste la distribución a los datos, menor será este estadístico.

5-En este caso al tener un comportamiento de una distribución NO normal, debemos seleccionar otro tipo de distribución para el análisis de los datos.

En la Hoja de Sesión se muestra los resultados de mejor ajuste AD= Anderson Darling. **En este caso seleccionaremos el valor más pequeño de AD**, el cual es para una distribución Logística la de mejor ajuste.

Distribution ID Plot for Height									
Descriptive Statistics									
N	N*	Mean	StDev	Median	Minimum	Maximum	Skewness	Kurtosis	
133	0	163.180	9.78313	162	110	190	-0.758267	6.90093	
Box-Cox transformation: $\lambda = 2$									
Goodness of Fit Test									
Distribution		AD	P	LRT	P				
Normal		2.959	<0.005						
Box-Cox Transformation		3.028	<0.005						
Lognormal		3.284	<0.005						
3-Parameter Lognormal		2.923	*	0.001					
Exponential		54.840	<0.003						
2-Parameter Exponential		43.278	<0.010	0.000					
Weibull		5.645	<0.010						
3-Parameter Weibull		4.768	<0.005	0.017					
Smallest Extreme Value		6.532	<0.010						
Largest Extreme Value		12.452	<0.010						
Gamma		3.086	<0.005						
3-Parameter Gamma		3.132	*	0.108					
Logistic		1.506	<0.005						
Loglogistic		1.345	<0.005						
3-Parameter Loglogistic		1.506	*	0.236					

- **¿Qué hacer con datos no normales?**

### **1- Continuar con el análisis si la muestra es lo suficientemente grande**

Aunque formalmente muchas pruebas de hipótesis se basan en el supuesto de normalidad, de todos modos, se pueden obtener resultados adecuados con datos no normales si la muestra es lo suficientemente grande. La cantidad de datos que se necesita depende del grado de no normalidad de los datos, pero un tamaño de muestra mayor a 30 suele ser adecuado. La relación entre la robustez ante la normalidad y el tamaño de la muestra se basa en el teorema del límite central. Este teorema demuestra que la distribución de la media de los datos de cualquier distribución se acerca a la distribución normal a medida que aumenta el tamaño de la muestra. Por lo tanto, si usted está interesado en hacer inferencias sobre una media de población, **el supuesto de normalidad no es fundamental siempre y cuando la muestra sea lo suficientemente grande.**

### **2-Usar una prueba no paramétrica**

**Las pruebas no paramétricas no presuponen una distribución específica para la población.** Minitab ofrece varias pruebas no paramétricas que se pueden usar en lugar de las pruebas que parten del supuesto normalidad. Estas pruebas pueden ser particularmente útiles cuando se tiene una muestra pequeña que es asimétrica o una muestra que contiene varios valores atípicos.

<b>Prueba que parte del supuesto de normalidad</b>	<b>Pruebas no paramétricas equivalentes</b>
Z de 1 muestra, t de 1 muestra	Signos de 1 muestra, Wilcoxon de 1 muestra
t de 2 muestras	Mann-Whitney
ANOVA	Kruskal-Wallis, mediana de Mood, Friedman

Las pruebas no paramétricas no están completamente libres de supuestos acerca de los datos: por ejemplo, sí requieren que los datos sean una muestra aleatoria independiente.

### **3- Transformar los datos**

A veces es posible transformar los datos mediante la aplicación de una función para que los datos se ajusten a una distribución normal, para poder terminar el análisis.

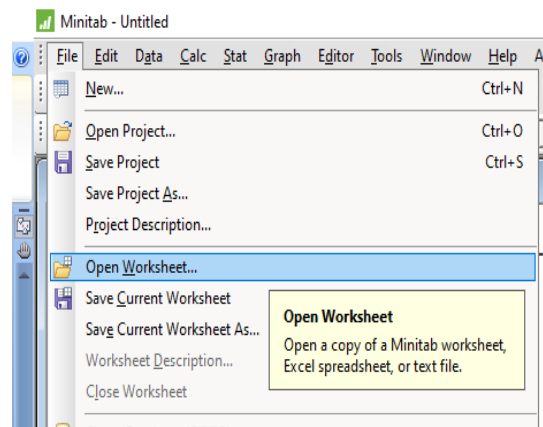
Usted puede transformar los datos usando muchas funciones, tales como la raíz cuadrada, el logaritmo, la potencia, la recíproca o el arcoseno.

- Para aplicar estas transformaciones directamente a los datos en la hoja de trabajo, utilice la Calculadora de Minitab.
- Para realizar una transformación de Box-Cox, elija **Estadísticas > Gráficas de control > Transformación Box-Cox**. Minitab determina una transformación óptima de potencia. La transformación de Box-Cox es fácil de entender, pero es limitada y, con frecuencia, no determina una transformación adecuada. Además, está disponible solo para datos que son positivos.
- Para realizar la transformación de Johnson, elija **Estadísticas > Herramientas de calidad > Transformación de Johnson**. Si el algoritmo de Box-Cox no determina una transformación adecuada, pruebe con la transformación de Johnson. La función de transformación de Johnson es más complicada, pero es muy efectiva para determinar una transformación adecuada.

Ejemplo de transformación de datos a distribución normal usando Transformación de Johnson:

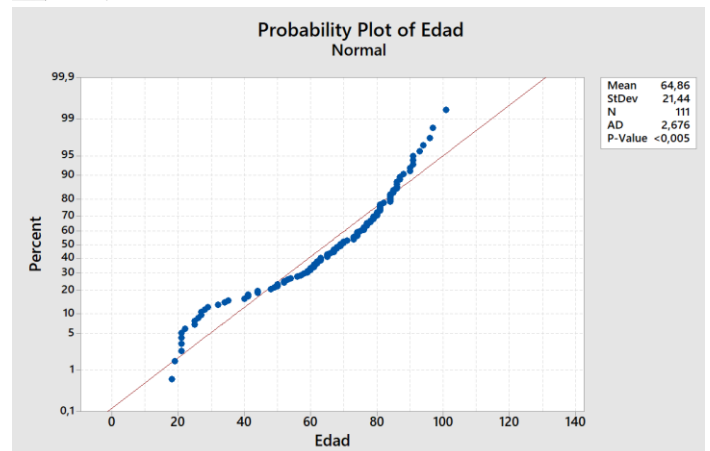
1-Cargamos la base de datos.

File->Open Worksheet->Seleccionar la base de datos

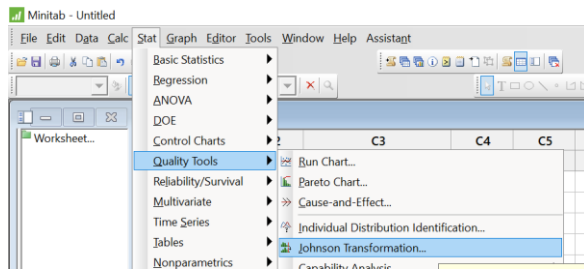


Base de datos:

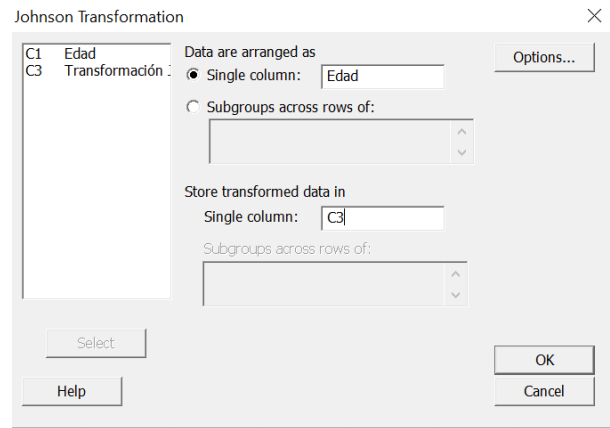
* C1	Edad
1	86
2	41
3	18
4	84
5	35
6	78
7	79
8	81
9	93
10	77
11	66
12	70
13	61
14	71
15	80
16	78
17	81
18	21
19	63
20	26
21	74



2- Seleccionamos:  
**Estadísticas > Herramientas de calidad > Transformación de Johnson**



3-En el formulario ingresamos la variable a transformar “Edad” y la columna donde se desplegará la información, en este caso “C3” columna 3.



4-En Minitab en la columna 3 se desplegarán los datos transformados:

Tomar en consideración que los datos de la columna “C3” para realizar las inferencias estadísticas, se han transformado. Ejemplo el valor de 86 es 1.02825, por lo que al seleccionar el valor de X se utilice un valor identificable en la base de datos, luego obtendremos el porcentaje de probabilidad de ocurrencia del fenómeno.

	C1	C2	C3	C4
	Edad		Transformación Johnson	
1	86		1,02825	
2	41		-1,13664	
3	18		-1,82453	
4	84		0,86647	
5	35		-1,32676	
6	78		0,45143	
7	79		0,51465	
8	81		0,64764	
9	93		1,75106	
10	77		0,39017	
11	66		-0,18575	
12	70		0,00625	
13	61		-0,40523	
14	71		0,05699	
15	80		0,58000	
16	78		0,45143	
17	81		0,64764	
18	21		-1,74000	
19	63		-0,31985	
20	26		-1,59637	
21	74		0,21687	

5-Al realizar nuevamente la prueba de normalidad obtenemos que siguen un comportamiento de distribución normal.

