

Curso
Análisis de datos
Guía MINITAB Unidad 4

Nombre del profesor: MSc. Francisco Olivier Paniagua Barrantes
Correo electrónico: frasse31@gmail.com



ÍNDICE

PRUEBA DE NORMALIDAD DE LOS DATOS	3
DISTRIBUCIÓN DE PROBABILIDADES PARA VALORES MAYORES, INTERMEDIOS Y MENORES.	5
▪ COLA IZQUIERDA, VALORES DE “X” MENORES	5
▪ COLA DERECHA, VALORES DE “X” MAYORES.....	7
▪ VALORES INTERMEDIOS, ENTRE “X1” Y “X2”	8
ZONA DE ACEPTACIÓN Y RECHAZO	10

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 ≥ 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:

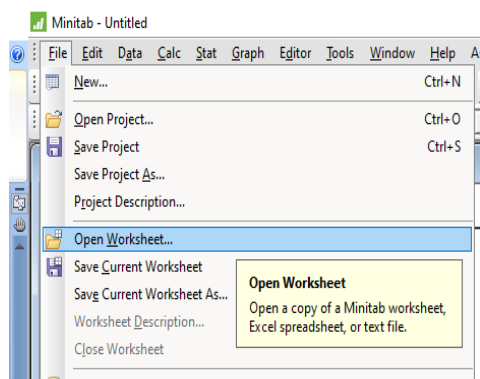
- Anderson–Darling: discrimina muy bien desde el centro de los datos hasta las colas.
- Ryan- Joiner: Discrimina muy bien los centros pero no las colas.
- Smirnov: Discrimina muy bien los centros pero no las colas, necesita muchos datos más de 100 para dar buena confiabilidad.

El diagrama muestra una curva 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 donde las líneas verticales se encuentran con la curva están etiquetados como $-Z_0$ y Z_0 .

- Datos con una distribución normal

1-Cargamos la base de datos.

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

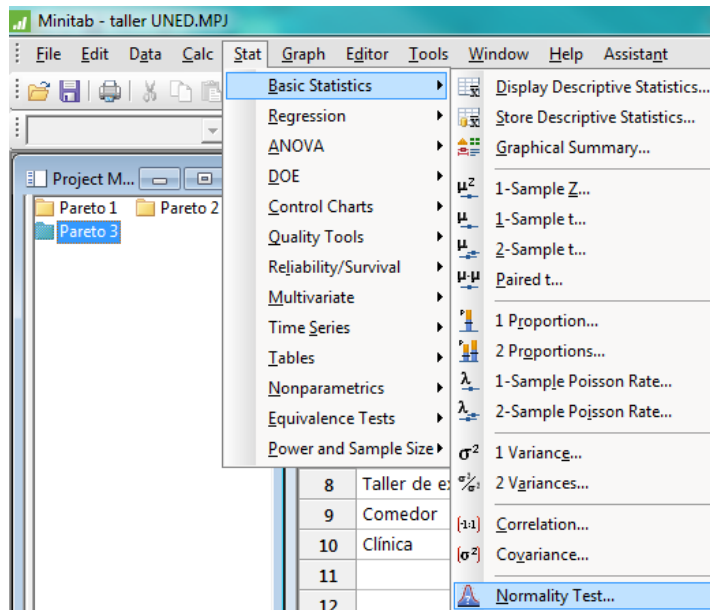


Base de datos:

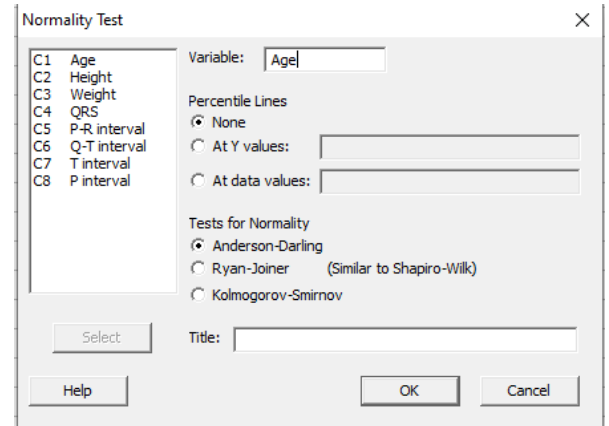
La imagen muestra una hoja de cálculo de Minitab con los siguientes datos:

	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

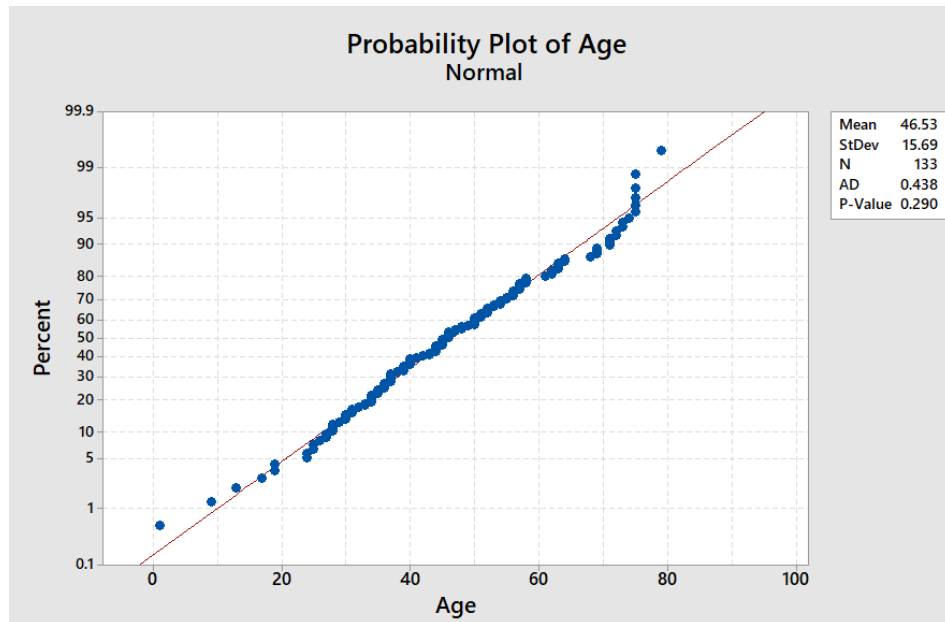


3-Se selecciona la variable en este ejemplo
 “Age”



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.



Distribución de probabilidades para valores mayores, intermedios y menores.

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, para esto calcularemos la probabilidad para cualquier valor "X".

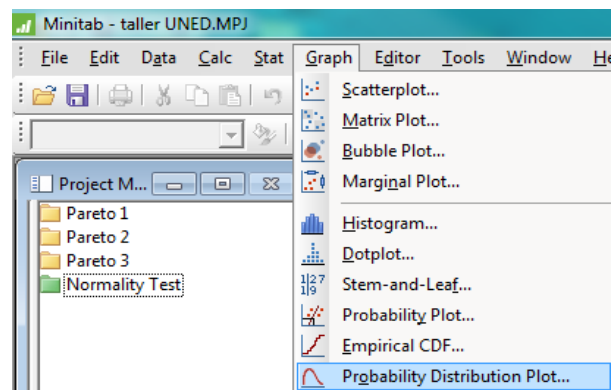
Para trabajar con la distribución de probabilidad se debe cumplir con los siguientes requerimientos:

- Realizar la prueba de normalidad para comprobar que existe distribución normal.
- Calcular el promedio y la desviación estándar de la variable.
- Cola izquierda, valores de "X" menores

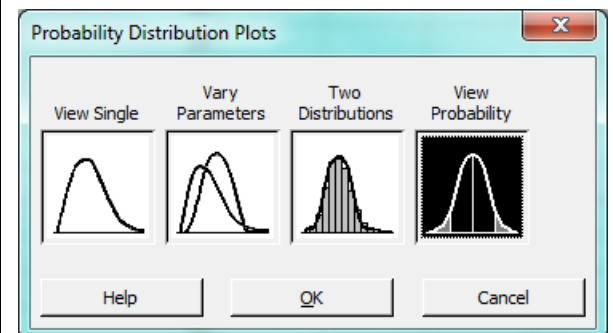
Corresponde a la probabilidad de encontrar valores menores al valor "X" seleccionado.

1-Seleccionamos la función de "Distribución de Probabilidad"

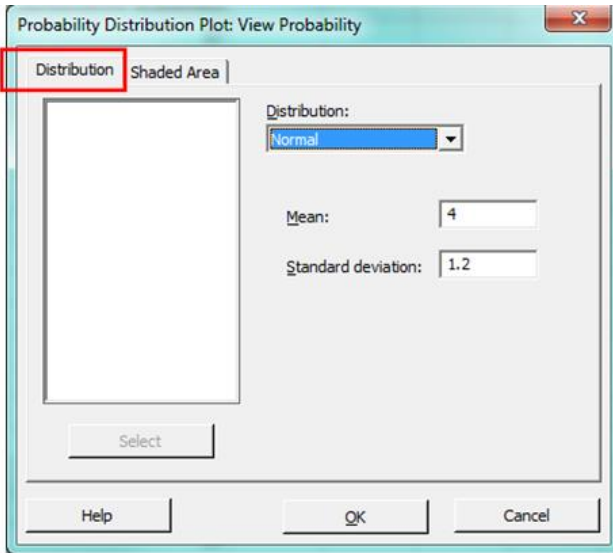
Graph → Probability Distribution Plot



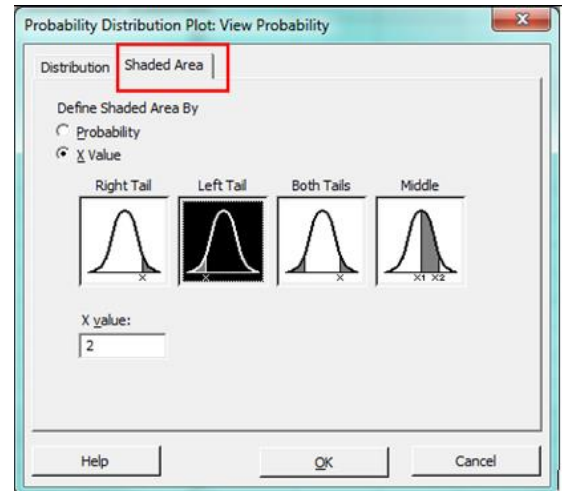
2- Se habilita un cuadro de selección. En este caso seleccionamos siempre "Ver Probabilidad"



3- Del cuadro de selección, en la pestaña de "Distribución" ingresamos los valores del promedio y la desviación estándar con una distribución normal.



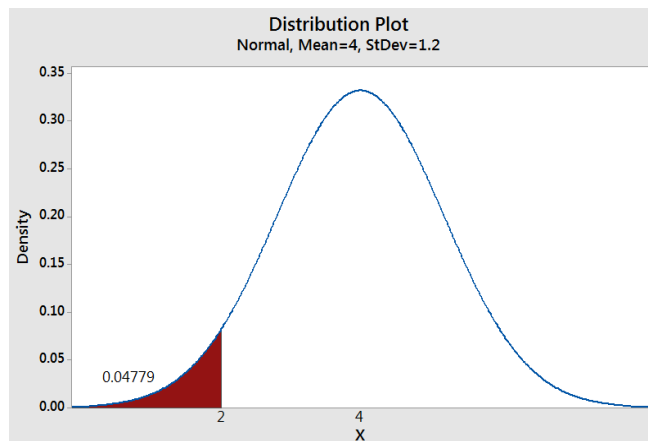
4- Del cuadro de selección, en la pestaña de "Área sombreada" seleccionamos cola izquierda.



4-Se obtiene el gráfico de probabilidad para la variable, marcando el área bajo la curva en rojo:

La probabilidad de durar menos de 2 min en la atención del servicio es de 0,04779 o 4,78%

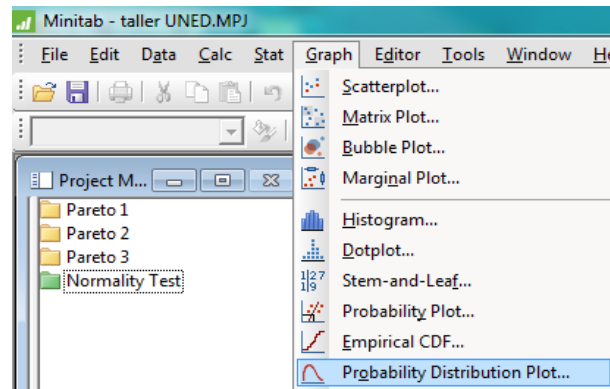
$$P(x \leq 2) = P(Z \leq -1.67) = 0,04779$$



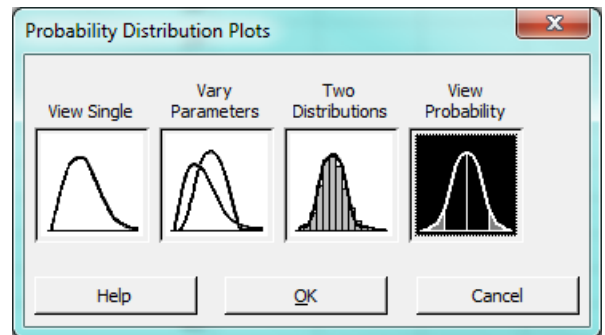
- Cola derecha, valores de “X” mayores

1-Seleccionamos la función de “Distribución de Probabilidad”

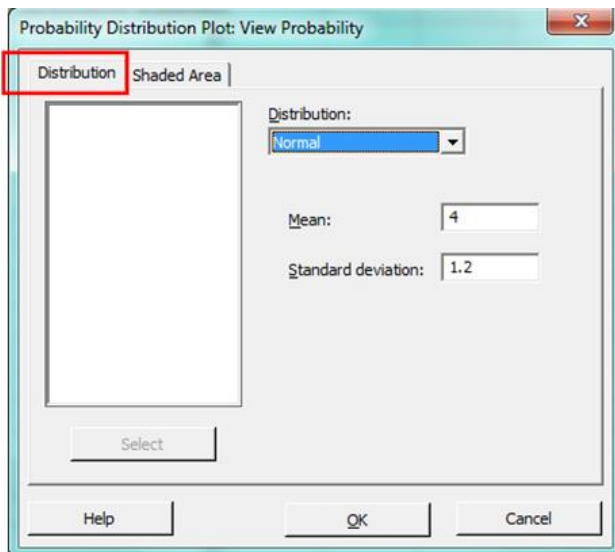
Graph→Probability Distribution Plot



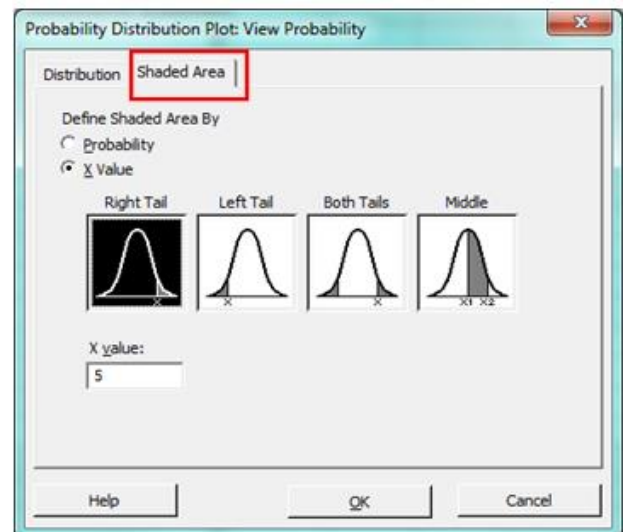
2- Se habilita un cuadro de selección. En este caso seleccionamos siempre “Ver Probabilidad”



3- Del cuadro de selección, en la pestaña de “Distribución” ingresamos los valores del promedio y la desviación estándar con una distribución normal.



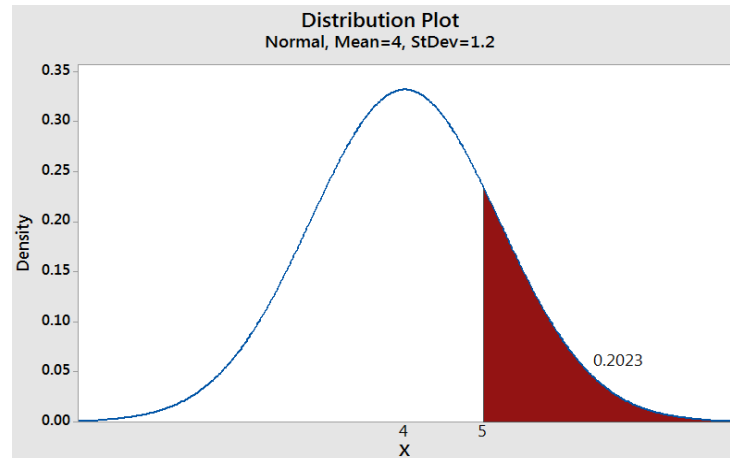
4- Del cuadro de selección, en la pestaña de “Área sombreada” seleccionamos cola derecha.



4-Se obtiene el gráfico de probabilidad para la variable, marcando el área bajo la curva en rojo:

La probabilidad es de 0,2023 o 20,23% de que dure más de 5min.

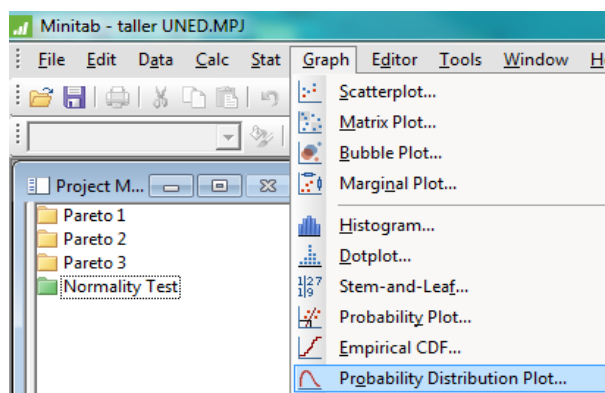
$$P(x>5) = 1-P(x\leq 5)$$



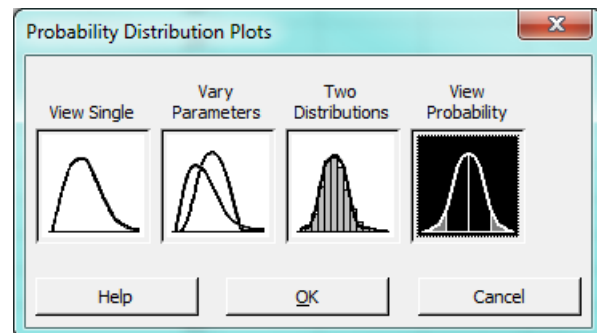
- Valores intermedios, entre “X1” y “X2”

1-Seleccionamos la función de “Distribución de Probabilidad”

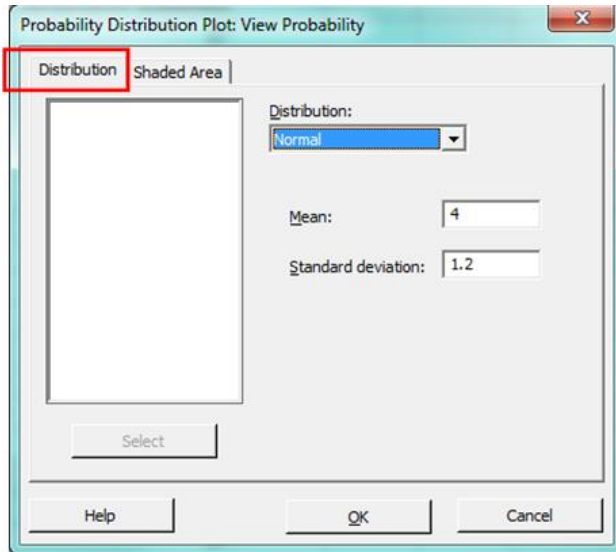
Graph→Probability Distribution Plot



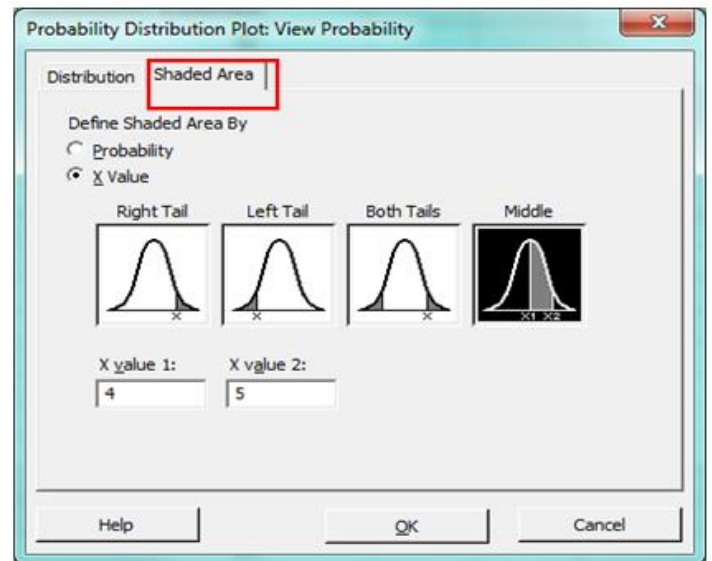
2- Se habilita un cuadro de selección. En este caso seleccionamos siempre “Ver Probabilidad”



3- Del cuadro de selección, en la pestaña de “Distribución” ingresamos los valores del promedio y la desviación estándar con una distribución normal.

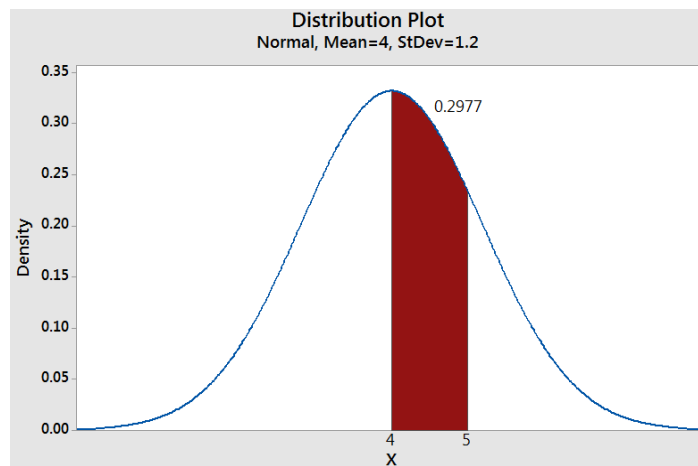


4- Del cuadro de selección, en la pestaña de “Área sombreada” seleccionamos intermedio (Middle) e ingresamos los valores “X1” y “X2”



4-Se obtiene el gráfico de probabilidad para la variable, marcando el área bajo la curva en rojo:

La probabilidad es de 0,297 o 29,7% que dure entre 4 y 5 min.



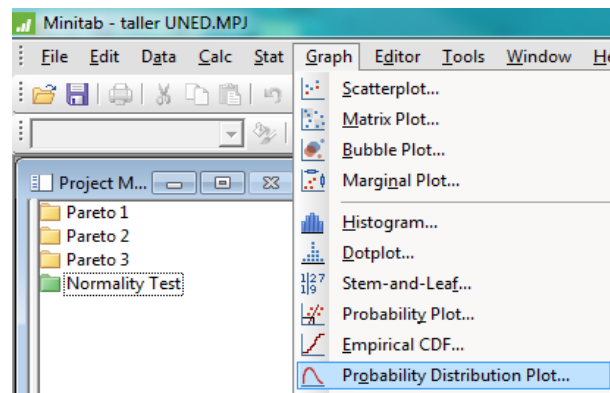
Zona de aceptación y rechazo

La zona de aceptación corresponde al intervalo de confianza de encontrar los valores de la variable estudiada con un 95% de confianza. La zona de rechazo o las colas corresponde a los valores extremos de la distribución de los datos.

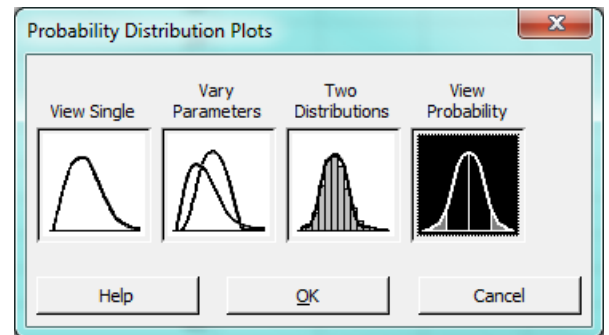


1-Seleccionamos la función de "Distribución de Probabilidad"

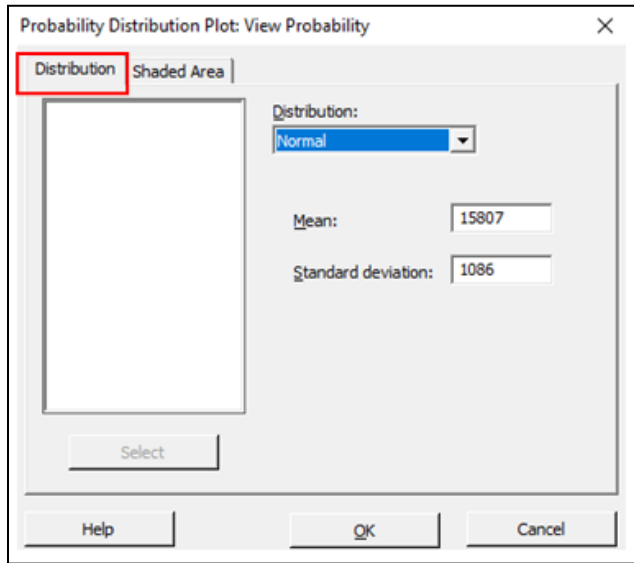
Graph → Probability Distribution Plot



2- Se habilita un cuadro de selección. En este caso seleccionamos siempre "Ver Probabilidad"

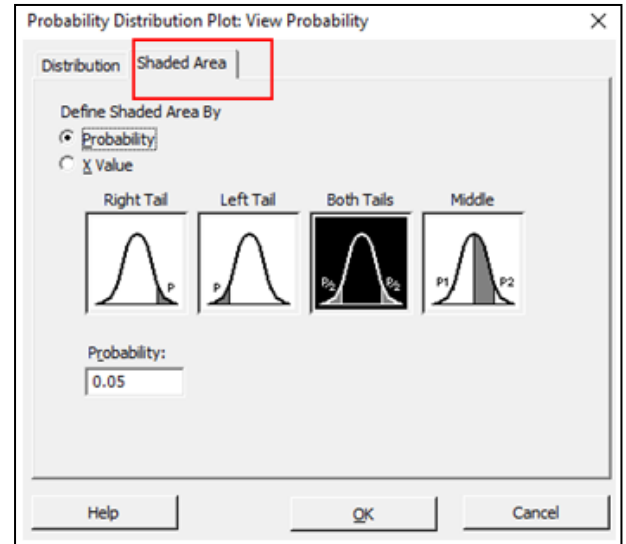


3- Del cuadro de selección, en la pestaña de “Distribución” ingresamos los valores del promedio y la desviación estándar con una distribución normal.



4- Del cuadro de selección, en la pestaña de “Área sombreada” seleccionamos:

- Probabilidad
- Ambas colas
- Ingresamos el valor de 0.05



4-Se obtiene el gráfico de probabilidad para la variable, marcando el área bajo la curva en rojo:

La probabilidad de encontrar para el X-Trimestre la cantidad de urgencias, está comprendido en el Intervalo de Confianza desde 13678 hasta 17936 urgencias.

