**Curso de Analisis Estadistico**

**Sesion 1**

**Objetivo del curso**

**Hacer estadistica con Minitab**

**Evaluacion**

**1 parcial 40%**

**2 parcial 40%**

**Ex cortos y tareas 20%**

**Estadistica es:**

**-Un proceso de toma de decisiones que se compone de cinco partes**

**1. Tomar los datos**

**2. ordenar los datos**

**3. analizar los datos**

**4. inferir acerca de los datos**

**5. tomar decisiones**

**Durante el proceso de inferencia considera que siempre**

**se va a definir una hipotesis a la que se va a asociar una confianza de decision**

**generalmente se toma como referencia practica el 95%**

**Minitab siempre va a dar el resultado de valor p o p value que representa**

**el riesgo o peligro de rechazar la hipoteisi propuesta.**

**si el valor p es muy alto entonces no se debe rechazar la hipotesis. Alto significa**

**que sea mayor o igual a 0.05 (5%).**

**Ejemplo 1**

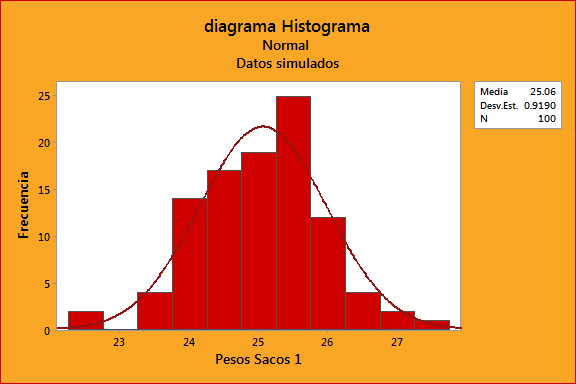
**1. Toma de datos: para esto vamos a simular estar en un beneficio de cafe**

**en el que estamos interesados en tomar la decision si el peso de los sacos de cafe**

**molido es el correcto o no? para esto vamos a simular que estamos en ese proceso**

**y tomar una muestra de 100 sacos que tiene un promedio de 25 kg y una desviacion**

**estandar de 1 kg.**



**Sesion 2**

**continuamos con el caso del beneficio de cafe.**

**media= 25 kg**

**desviacion estandar = 1kg**

**NOTA:**

**Diagrama de Cajas**

**-Escoger siempre la caja con la menor altura posible, ya que contendria la cantidad de datos menos variados.**

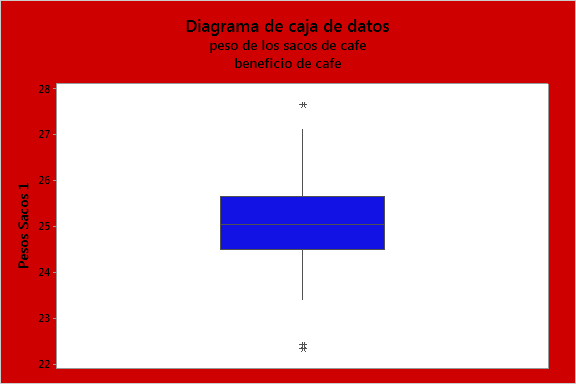
**- Existe un valor del 50% dentro de la caja , del punto de abajo hasta el superior de la caja es 75% y hasta el inferior es 25%.**

**- El \* representa los datos mas alejados.**

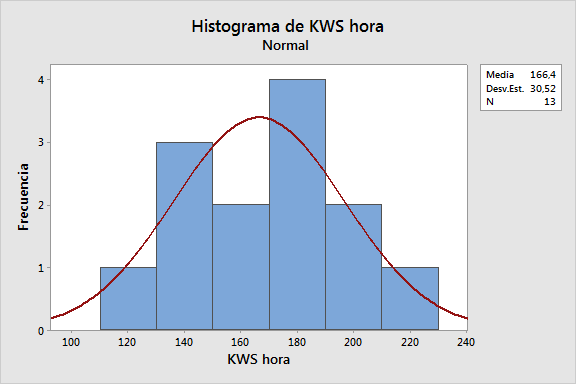
**-la Grafica mostrada es una grafica simple de un solo dato Y.**

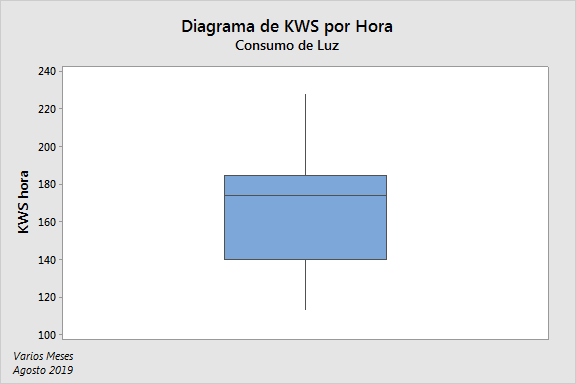
**-Los bigotes son los valores alejados moderados .**

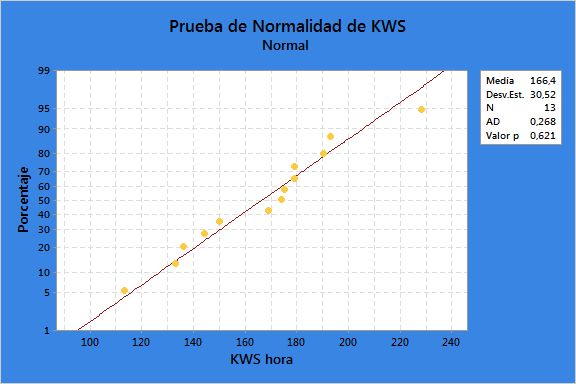
**-La mediana se ecnuentra en la linea divisora de la caja.**



**Consumo de Energia en el Hogar**







**Ho: los resultados de los consumos de energia**

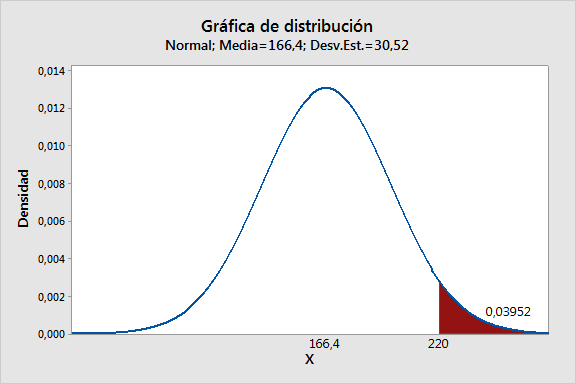
**electrica siguen un comportamiento normal.**

**Ha: No siguen un comportamiento normal.**

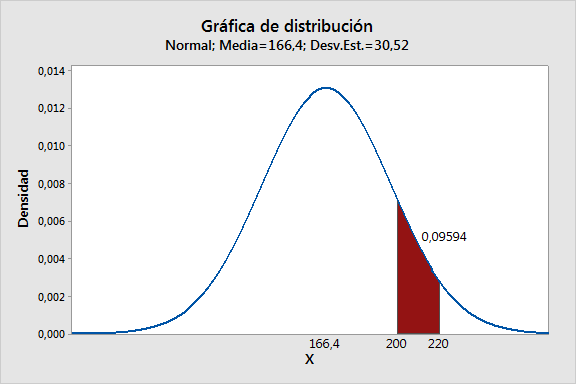
**Valor P: 0,621**

**Conclusion: como mi valor p es mayor a 0.05 entonces no**

**se debe rechazar la hipotesis nula, entonces es normal.**



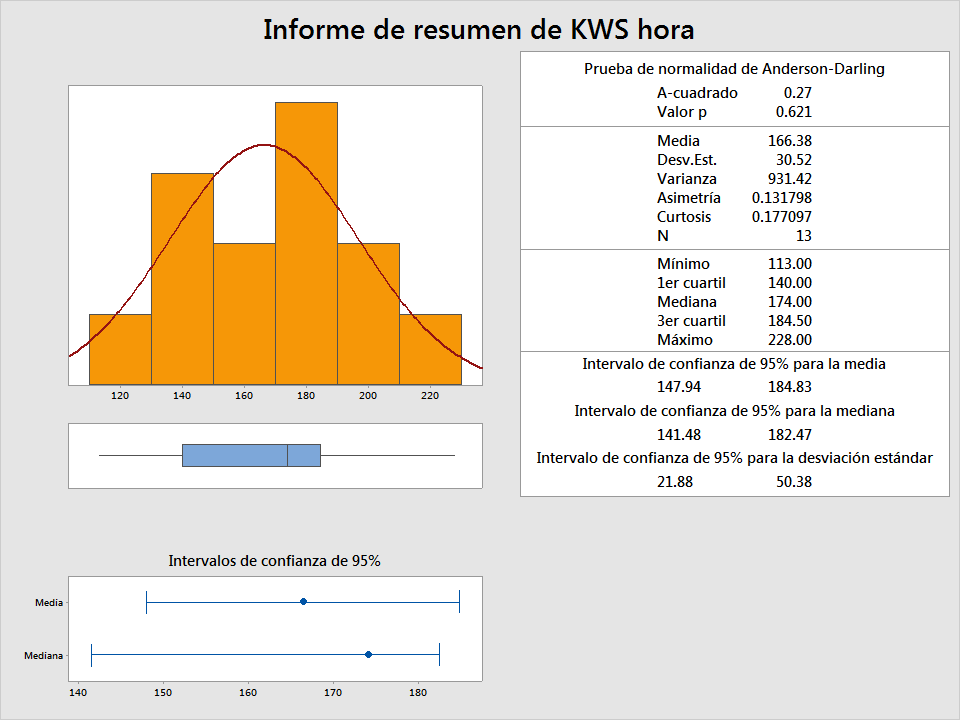
**Existe una probabilidad de 0,39 veces de que mi consumo sea de 220 en 100 meses**



**Existe una probabilidad de 0,95 veces de que mi consumo sea de entre 200 y 220 en 100 meses**

**Sesion 3**

**Resumen grafico : muestra todos los datos obtenidos un conjunto de datos**



**VALOR P**

**1. si el valor p es mayor a 0.05 por tanto sigue el comportamiento normal y**

**no se debe rechazar la hipotesis**

**2. Existe suficiente evidencia estadistica para sustentar esa decision poblacional**

**Conceptos Basicos**

**-Media: es el valor promedio que representa un conjunto de datos**

**-Desviacion Estandar: es la desviacion promedio de los datos con respecto a la media**

**-Asimetria: muestra como se agregan los datos con respecto un lado del otro**

**-Curtosis: es una medida de forma que mide cuán escarpada o achatada está una curva o distribución.**

**-Mediana: es el valor promedio de los valores centrales de un conjunto de datos tomadas 50 a 50.**

**-1 cuartil: el 25% de los datos**

**-3 cuartil: el 75% de los datos**

**-Intervalo de confianza para la media: es el rango de valores entre los que se espera se encuentre con un 95% el valor de la media poblacional (es decir de lo que siempre ocurre en esa situacion).**

**-Intervalo de confianza para la mediana: en el rango de valores en el que se encuentra la media**

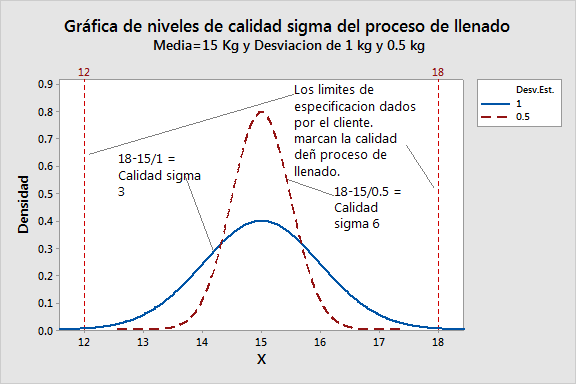
**- Intervalo de confianza para la desviacion estandar: el rango de valores en el que se espera se encuentre la desviacion estandar.**

**Nivel de Calidad**

La calidad se puede medir en terminos de sigma

si un lado es tiene mayor diferencia que el otro , se escoje el lado de menor diferencia

con respecto a los limites.



Una vez que se tiene clara la voz del cliente y a la vez la voz del proceso.

Entonces se hace necesario realizar un analisis que indique la causa de

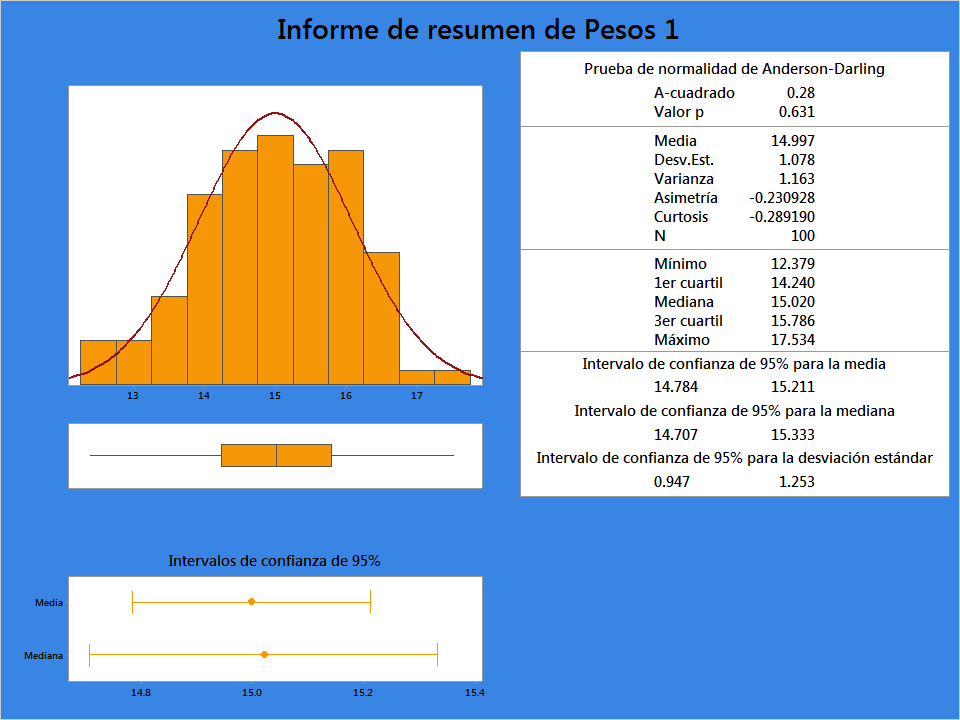
raiz que provoca que no se este logrando lo que el cliente desea.

**PRÂCTICA**

Considere estra en un proceso de llenado de sacos de cafe. simule una muestra que sigue

el comportamineto con media de 15 kg y desviacion de 1kg.

- realice el proceso de estadistica con minitab.



Describa:

a. prueba de normalidad. hacer referencia al valor p

b. el histograma

c. el grafico de caja de datos

d. la media

e. la desviacion estandar

f. la mediana

g. la curtosis

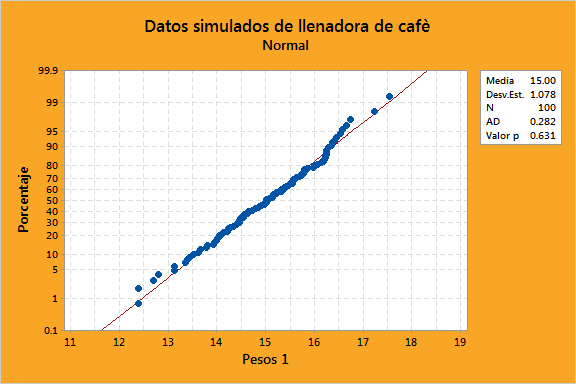
h. la asimetria

i. los intervalos de confianza: media, median y desviacion estandar.

j. calcular el nivel de calidad de sigma.

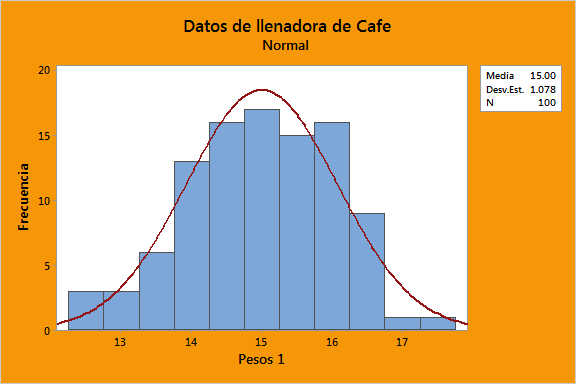
**Respuestas**

**A)** Prueba de normalidad. hacer referencia al valor p



Los resultados de la prueba indican si usted debe rechazar o no puede rechazar la hipótesis nula de que los datos provienen de una población distribuida normalmente. El valor p obtenido de la grafica muestra que es mayor al 0.05 por lo tanto no se rechaza

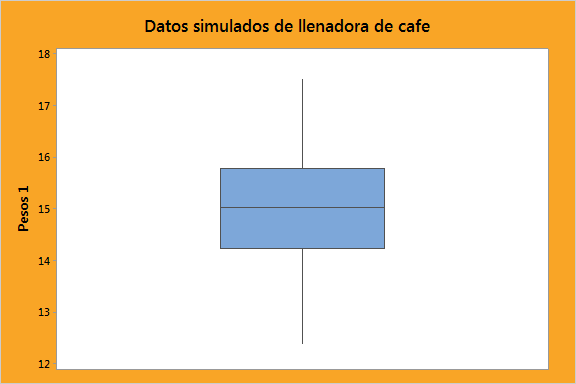
**B)** El histograma



El histograma revela que los datos se encuentran moderadamente bien distribuidos

con una media de valor de 15.

**C)** El grafico de caja de datos



**D)**

**E)**

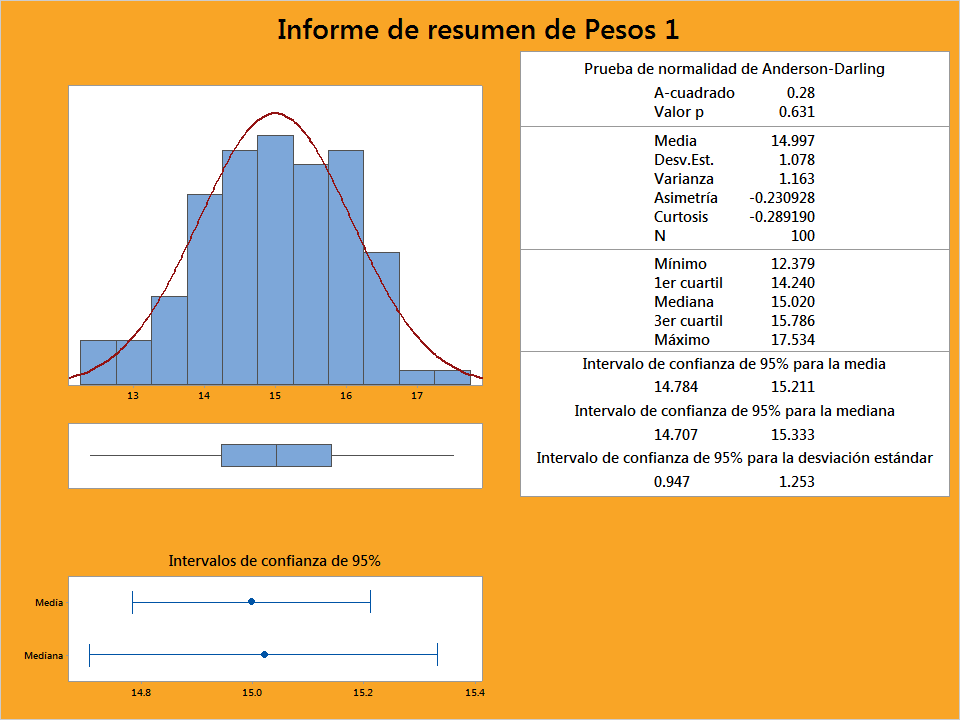
**F)**

**G)**

**H)**

**I)**

**J)**



**Tarea.**

Los hornos para postar el cafe requieren que las temperaturas se mantengan dentro de ciertos limites para evitar daños en el grano. Demuestre sus conocimientos para hacer estadistica con MINITAB, simulando obtener una muestra de 50 temperaturas del horno con un promedio de 240 grados y una desviacion estandar de 12 grados. Cuando obtenga la muestra ponga un comentario, utilice algun color para distinguirlos y trabaje con dos decimales, y describa:

DESCRIBA:

a) Prueba de normalidad, hacer referencia al valor p

b) Histograma

c) Grafico de cajas

d) la media

e) la desviacion estandar

f) la mediana

g) curtosis

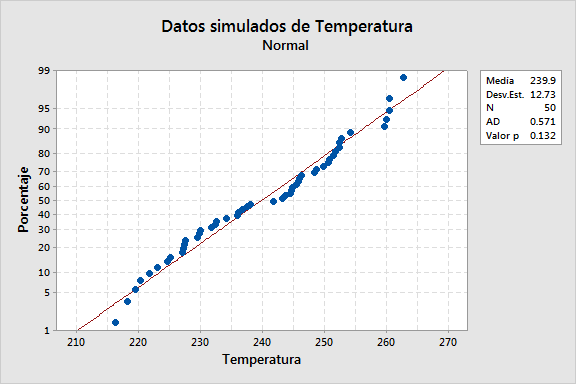
h) la asimetria

i) los intervalos de confianza: media,

mediana y desviacion estandar.

j) Nivel de Calidad Sigma

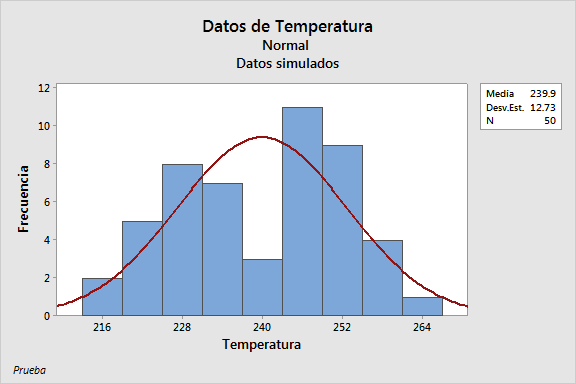
a) Prueba de normalidad, hacer referencia al valor p



- El valor de p (0.132) es mayor a 0,05 por lo tanto no se rechaza la hipotesis

, es decir que los datos siguen un comportamiento normal.

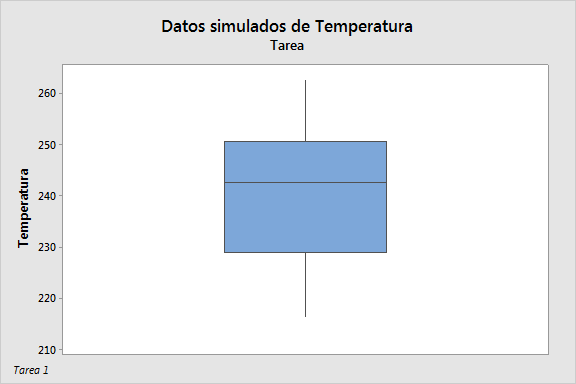
b) Histograma



En el histograma podemos observar que nuestros datos tienen una media de 239,9

con una desviacion de 12,73

c) Grafico de cajas



En el grafico de cajas podemos observar que el 50% de los datos estan contenidos en la caja

entre 230 y 250, donde la linea central nos indica la Mediana y los "bigotes" de la grafica representan valores alejados moderados y si tenemos asteriscos representan valores criticos, es decir muy alejados.

d) la media

e) la desviacion estandar

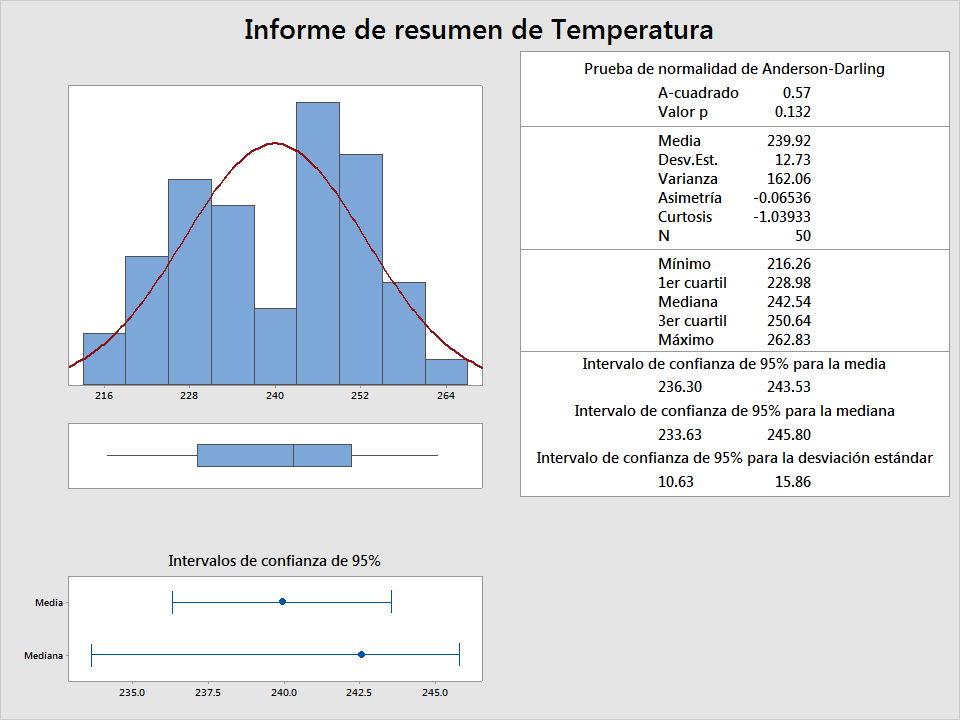
f) la mediana

g) curtosis

h) la asimetria

i) los intervalos de confianza: media,

mediana y desviacion estandar.



d) La media es el valor caracteristico de nuestra serie de datos y esta es 239,92

e) La deviacion estandar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media, en nuestros datos esta es 12.73

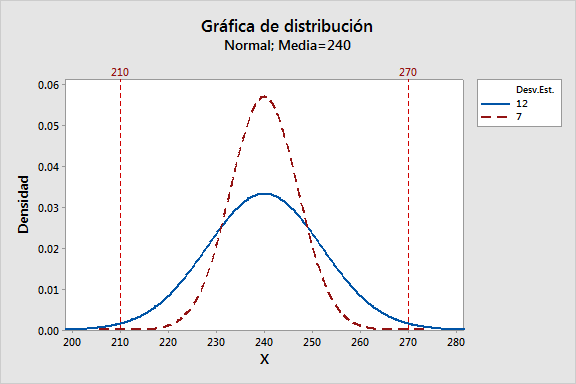
f) La mediana representa el valor de la variable de posición central en nuestro conjunto de datos ordenados.

g) La curtosis es una medida de apuntamiento que determina el grado de concentración que presentan los valores de una variable alrededor de la zona central de la distribución de frecuencias.

h) La asimetria son los indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria.

I) Los intervalos de confianza para la media, mediana y desviacion estandar son un rango de valores, en que posiblemente incluya el valor de los parámetro de la muestra.

j) Nivel de Calidad Sigma



Esta grafica nos muestra la comparacion de dos distribuciones, una con desviacion estandar de 12 y la otra de 7, una mas dispersa que la otra con respecto a la media. Tambien se pueden ver los limites de especificacion superior e inferior establecidos por el cliente, los cuales marcan el rango de temperaturas para evitar ciertos daños en el grano de cafe.

**Sesion 4**

DIARGRAMA DE ISHIKAWA

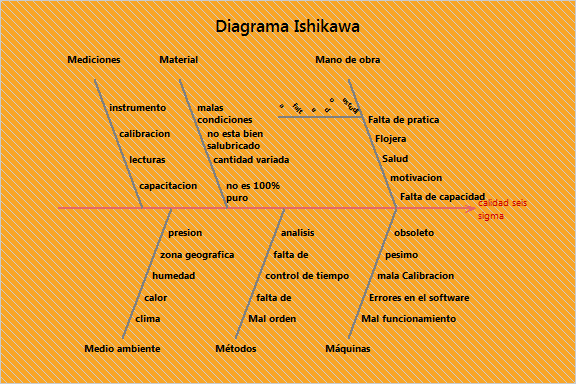


DIAGRAMA DE PARETO

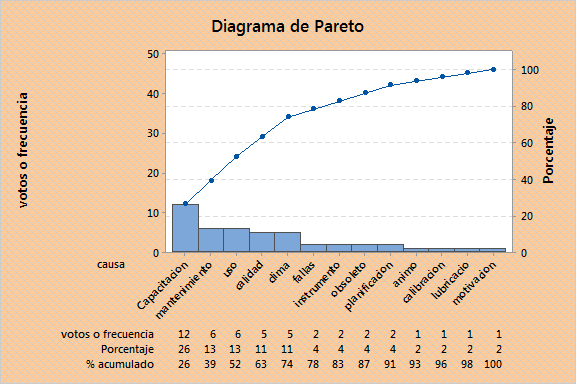
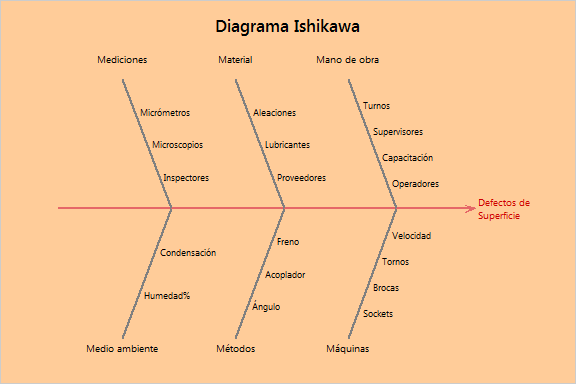


Diagrama ejercicio de ayuda



**SESION 5**

En ocasiones nos encontramos con la necesidad de evaluar si un proceso

esta dando los resultados esperados. Para esto es importante que recordemos

lo que significa Hacer Estadistica.

Hoy vamos a practicar con la generacion de una muestra aleatoria tomada

en una camara de refrigeracion de flores que debe tener siempre una temperatura

entre -3 grados a 3 grados

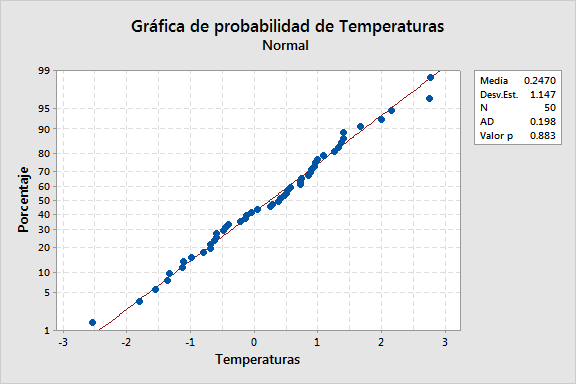
La dusa esta en si realmente las temperatura estan cumpliendo con lo solictado o es necesario que se realicen ajustes

Se pide para que puedas practicar:

1. Obtenga del software una muestra de 50 temp. con media de 0.25

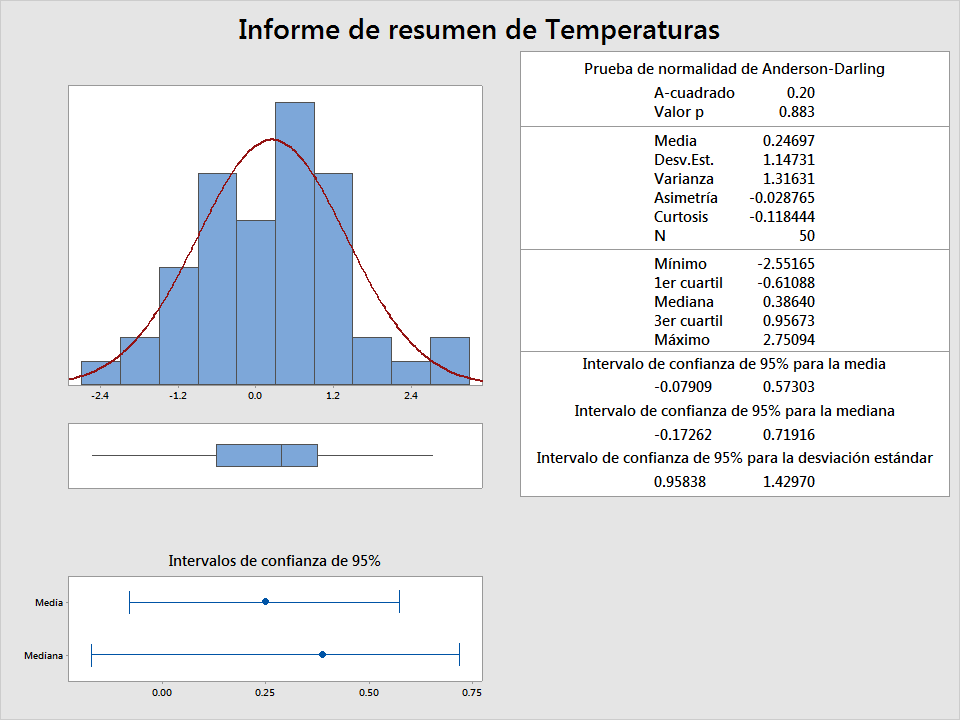
y desviacion de 0.98 grados (solo dos decimales).

2. Realice una prueba de Normalidad y compruebe. si por algun motivo los resultados esperados no siguen el comportamiento normal, entonces vuelva a tomar la muestra.



si muestra un comportamiento normal .

3. realice un resumen grafico estadistico completo y conteste los siguiente:



3.1 Es realmente el promedio u n buen representante de los datos

porque?

- Desde le origen del hombre buscar el numero que muestre el quilibrio entre ambas partes de un sistema ha sido esencial para tener una idea de una buena representacion de datos; por lo que nuestgra media de 0.25 no es aceptable para nuestro problema, porque es mayor a 5%.

x=(Desviacion/ media )\*100

si x>5% no es un buen representante.

x=460%

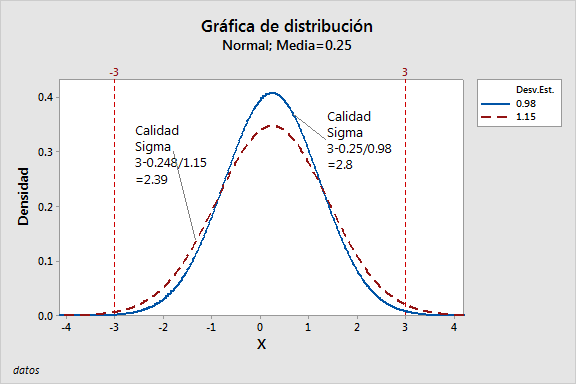
3.2 es la desviacion una buena medida de variacion de los datos?

-con respecto ala media la deviacion muestra un valor aceptable en este caso.

3.3 tiene sufieciente informacion para concluir que el proceso esta cumplinedo

con lo que se pide?

- si tenemos toda la informacion sufieciente



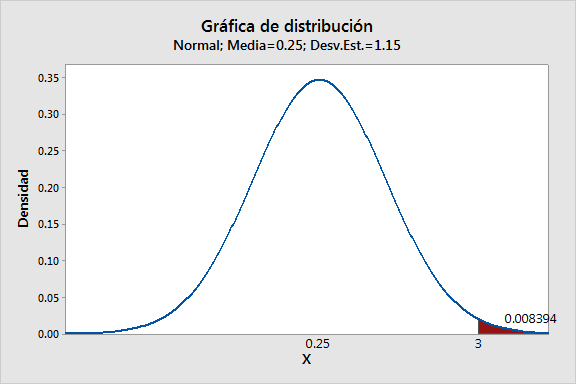
Calidad sigma azul=2.8

calidad sigma rojo=2.39

como hacer para que ese nivel de calidad se mayor?

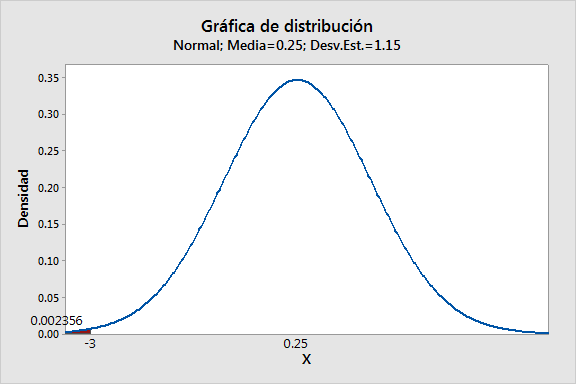
3-0.23/6=0.46

3.4 cual es la probabilidad de tener temp. superior a los 3 grados?



la probabilidad es de 0.008394 para que sea superior a 3 grados

3.5 cual es la probabilidad de tener temp. menores a los -3 grados?



la probabilidad es de 0.002356 para que sea menor a -3 grados

3.6 Cuales son las propuestas de mejoras?

- Deberian ajuster la maquinas para obtener mejores datos y con esto tener parametros

que no sobrepasen los limites que se tienen para este caso.

NOTAS

-La camara de refrigeracion tienm primero una calidad muy baja para cumplir con lo que se

pide , ademas el promedio no es un buen representante.

-no todo promedio es buen representante, el verdadero culpable es la desviacion estandar

y esto justifica que siempre debes calcularla para verificar ese promedi.

- Para mejorar la calidad de un proceso es necesario reducir la desviacion estandar.

para eso es dificl ya que se tiene que reducir a la fuerza las 6M , en el ejercicio de la camara de temperatura la que parece estar fallando es la maquinaria.

**SESION 6**

Ejemplo

-La temperatura

Correlación: TEmperatura; Calidad %

Correlaciones

|  |  |
| --- | --- |
| Correlación de Pearson | -0,934 |
| Valor p | 0,000 |

Ho: una variable no afecta a la otra

Ha: una varibale afecta a la otra

valor P: 0,000

si se rechaza la hipotesis nula ya que el valor de p es menor al 0,05

por lo que existe suficiente evidencia de correlacion entre temperatura

y calidad.

COEFICIENTE DE CORRELACION

- Este mide porcentualmente el grado de impacto de la variable independiente

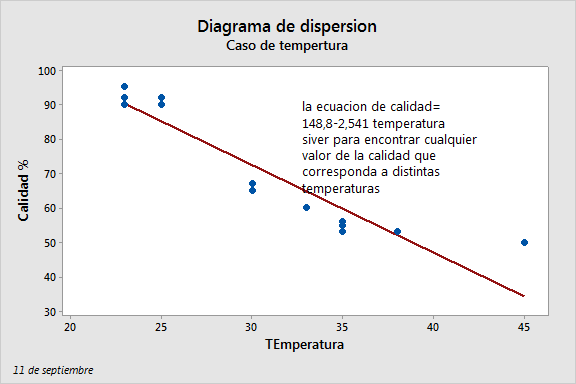
r= (-0,934)

- La r se lo eleva al cuadrado y se multiplica por 100

r2=0,87236=87,23%

- Esto significa que existe un cambio en la Calidad

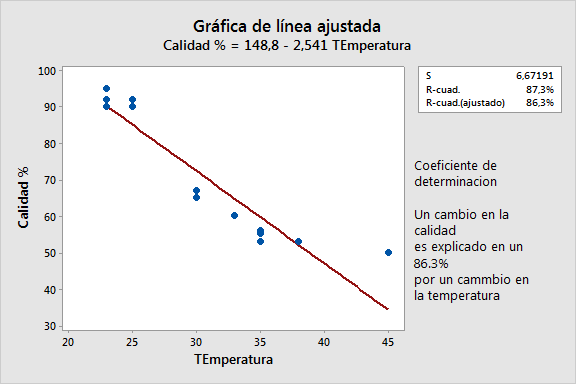
de un 87,23 % por un cambio en la tempertura.



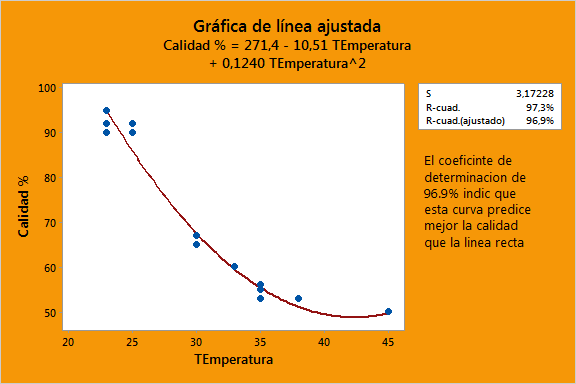
Utilizando en comando de regresion es posible adicionar el calculo del coeficiente

de determinacion , de esta forma:

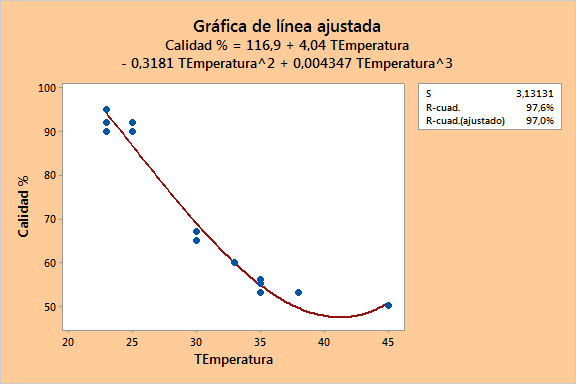
LINEAL



CUADRATICA



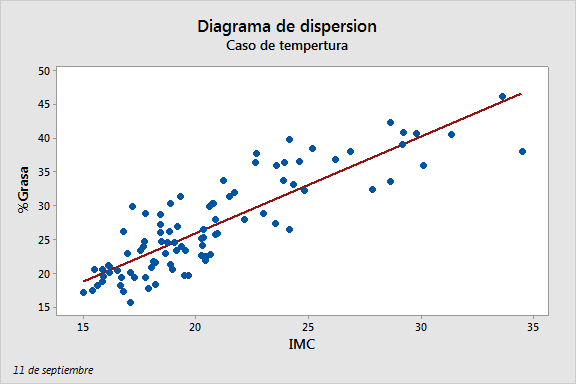
CUBICA



Ejemplo 1

Un investigador médico estudia la obesidad en las chicas adolescentes. Puesto que el porcentaje de grasa corporal es difícil y costoso de medir directamente, el investigador desea determinar si el índice de masa corporal (IMC), una medición que es fácil de tomar, es un predictor adecuado del porcentaje de grasa corporal. El investigador recopila el IMC, el porcentaje de grasa corporal y otras variables personales de 92 chicas adolescentes.

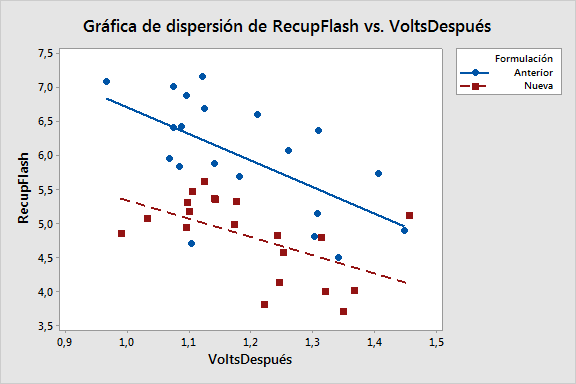
Como parte de la investigación inicial, el investigador crea una gráfica de dispersión del porcentaje de grasa corporal vs. el IMC para evaluar la relación entre las dos variables.



Ejemplo 2

Un ingeniero especializado en calidad que trabaja para un fabricante de cámaras desea acortar el tiempo de recuperación del flash. El tiempo de recuperación del flash es la menor cantidad de tiempo que se necesita entre los destellos. El ingeniero quiere determinar si existe una relación entre el voltaje que queda en la batería de la cámara inmediatamente después de un flash y el tiempo de recuperación del flash. El ingeniero también quiere determinar si existen diferencias en el tiempo de recuperación del flash entre la formulación anterior de la batería y la nueva formulación. El ingeniero recolecta muestras aleatorias de baterías hechas con la formulación anterior y la formulación nueva. El ingeniero mide los voltios que quedan inmediatamente después de un flash y el tiempo de recuperación del flash para cada una.

Como parte de la investigación inicial, el ingeniero crea una gráfica de dispersión de los voltios restantes después del flash vs. el tiempo de recuperación del flash, agrupados por formulación de batería, para evaluar la relación entre las dos variables para las dos formulaciones.



Correlación: RecupFlash; VoltsDespués

Correlaciones

|  |  |
| --- | --- |
| Correlación de Pearson | -0,478 |
| Valor p | 0,002 |

Al integrar los datos, la correlacion es baja pero siempre negativa