MINITAB
Manual de Entrenamiento
Estadísticas Básicas
Manual del usuario
Minitab TM - Making Data Analysis Easier - Version 13

Ejemplos y ejercicios	Propósito	Pagina
Prueba de Hipótesis		4-8
Ejemplo1 Llenado de las cajas de cereal	Evaluar la diferencia entre una muestra de la media y el valor del objetivo que se usa en la prueba de hipótesis.	
Intervalos de Confianza		9-18
Ejemplo 2 Peso de las cajas de cereal Ejemplo 3 Entendiendo los intervalos de confianza	Evaluar la diferencia entre la muestra de la media y el valor del objetivo que se usa en un intervalo de confianza. Demostrar la relación entre μ y el intervalo de confianza.	
Power		19-30
Ejemplo 4 Evaluando el Power Ejemplo 5 Incrementando el Power	El Power de la prueba de hipótesis Demostrar el efecto del tamaño de le muestra en el Power.	
Ejercicio 5.1 Detectando cambios en el diámetro del Balero.	Ejecutar el análisis del Power	

Ejemplos y ejercicios
Objetivos
Prueba de Hipótesis
Objetivos

- **Prueba de la hipótesis nula utilizando t-test e intervalos de confianza.**
- **Evaluación del Power de la prueba de hipótesis utilizando el análisis del Power.**

Prueba de Hipótesis
Ejemplo 1 Llenado Cajas de cereal
El propósito de este ejemplo es de introducir los conceptos de la prueba de hipótesis. Tu usaras un one-sample t-test para analizar datos procesados para determinar sí el proceso esta en el objetivo.
Problema
El objetivo. Tu quieres determinar sí el proceso esta en el objetivo
Recolección de datos
Para evaluar el proceso de la media. Elegirás 6 cajas de cereal al azar, las pesaras, y usaras los datos de ejemplo para estimar la media de la población.
Herramientas
Stat> Estadísticas básicas>1-Sample t
Data set
CEREALBX.MPJ

Nombre	Tipo de dato	Tipo de variable
BoxWeigh	Numérico	Respuesta

One-Sample T

N	Mean	StDev	SE Mean	95% CI
6	0.365000	0.050000	0.020412	(0.312528, 0.417472)

Prueba de hipótesis

¿Qué es una prueba de hipótesis?

Una prueba de hipótesis usa datos de ejemplo para probar una hipótesis acerca de la población de cual el ejemplo es tomado. El one-sample t-test es uno de los muchos procedimientos disponibles para la prueba de una hipótesis en MINITAB.

Por ejemplo, suponga que quiere probar la medida de las ruedas del pistón es igual a la longitud deseada del objetivo. Usted medirá un numero de ruedas y usara la medida de esas ruedas de ejemplo para estimar la medida de la rueda de la población. Este es un ejemplo de *stastistical inference*, usando información acerca de un ejemplo para hacer una inferencia acerca de una población.

¿Cuándo usar una prueba de hipótesis?

Usa una prueba de hipótesis cuando tengas datos de ejemplo y quieras hacer inferencias acerca de una o más poblaciones.

¿Por qué usar una prueba de hipótesis?

La prueba de hipótesis puede ayudar a contestar preguntas como:

- ¿Esta el proceso correctamente centrado?
- ¿Es el producto de un proveedor mejor que el producto de otro?
- ¿Hay diferencias entre el tratamiento de los grupos y los experimentos?

Por ejemplo,

- ¿ Es tu surtido de tu papel en media de 8.5 pulgadas de ancho?
- ¿La gasolina del proveedor es de mejor octanaje que la del proveedor B?
- ¿El cliente prefiere una formulación de una bebida sobre otra?

Probando la hipótesis nula

Necesitas determinar si la media de un proceso de empaque difiere significativamente del peso correcto que es 365 gramos. En Términos estadísticos, el proceso de la media es también llamado la población de la media.

Hipótesis de estadística

Hay 2 posibilidades, µ es igual a 365 o no lo es. Estas alternativas pueden ser usadas como 2 hipótesis:

- La hipótesis nula (H0): µ es igual a .365 gramos
- La hipótesis alternativa(H1): µ no es igual a 35 gramos

Por que no puedes medir cada caja en la población, nunca podrás saber con exactitud cual hipótesis es correcta. Sin embargo una prueba de hipótesis apropiada pueda ayudarte a hacer un cálculo formal. Para estos datos la prueba apropiada es la one-sample t-test

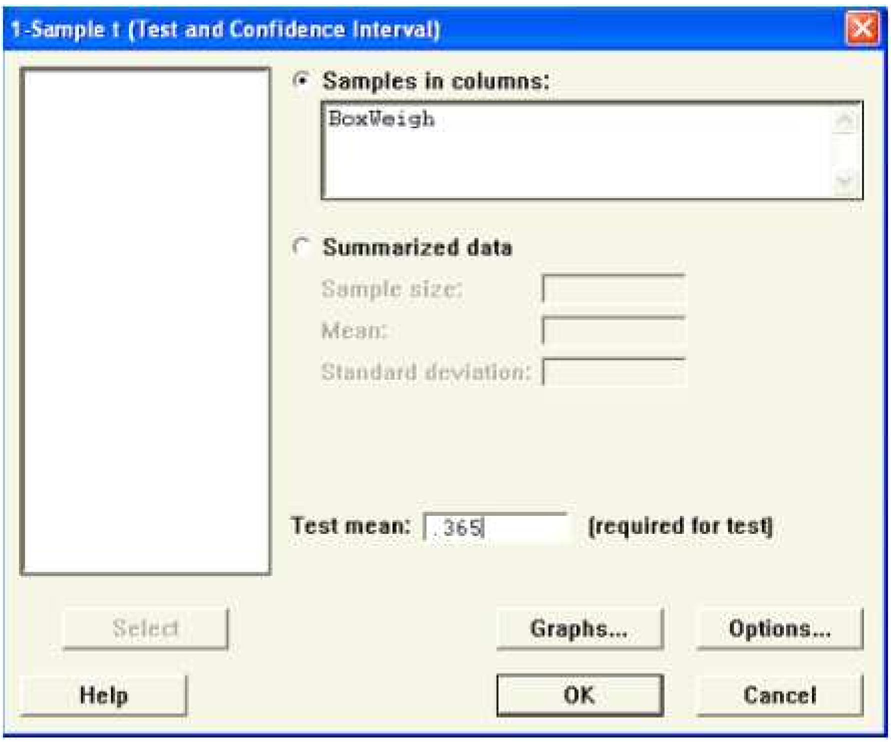
1- Sample t

1.- Abre el proyecto CEREALBX.MPJ.

2.- Escoge **STAT** > **Basic Statistics** > **1-Sample t.**

3.- Complete el recuadro como se indica a continuación:

4.- Click **OK.**



Interpretando tus resultados

La lógica de la prueba de hipótesis

Todas las pruebas de hipótesis siguen los mismos pasos:

- Asumir que H0 es verdadera.
- Determinar que tan diferente es tu muestra de lo que esperas dado que H0 es verdad.
- Si tu muestra es diferente dado que H0 es verdad, entonces descarta H0.

Por ejemplo, los resultados de t-test indican que la muestra es 366.704. De esta manera el examen contestara la pregunta, “Si µ es igual a 365, como obtendrás una muestra de 366.704(o mayor). La respuesta es dada como una probabilidad que vale (P), que para esta prueba es igual a 0.143.

Tomando una decisión

Para tomar una decisión, necesitas Escoger el nivel de importancia, α (alpha), antes de la prueba:

- Si P es menor o igual a α, rechazas H0 .
- Si P es mayor que α, si fallas al rechazar H0 (Técnicamente, nunca *aceptas* H0 , simplemente fallas al rechazarlo).

Un valor típico para α es 0.05, pero valores mayores o menores puedes ser escogidos dependiendo de la exactitud requerida para la prueba. Asumiendo que escojas un α-Nivel de 0.05 para los datos del peso de la caja no tendrás suficiente evidencia para rechazar H0. P(0.143) es mayor que α.

One-Sample T: Boxweigh

Test of mu = 365 vs not = 365

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
BoxWeigh	6	366.704	2.403	0.981	(364.183, 369.226)	1.74	0.143

Consideraciones finales

Conclusiones prácticas

Basado en tus datos de muestra, no puedes rechazar la hipótesis nula al 0.05 nivel α. No hay suficiente evidencia para sugerir que los pesos completos son diferentes a .365 gramos.

Consideraciones de estadística

Cuando es conducida una prueba de hipótesis, siempre empiezas con dos hipótesis contrarias:

La hipótesis nula(H0):

- Normalmente dice que si una propiedad de una población (tal como la media) no es diferente de un valor especifico o de otra población.
- Es asumido que es verdad hasta que tengas suficiente evidencia de lo contrario.
- Nunca es aceptado--- simplemente fallas al rechazarlo.

La hipótesis alternativa(H1):

- Dice que la hipótesis nula esta equivocada.
- También especifica la dirección de la diferencia.

Cada prueba de hipótesis esta basada en una o más suposiciones acerca de los datos que están analizando. Si esas suposiciones no son conocidas, los resultados puede que no sean precisos. Las suposiciones de cada prueba serán exploradas cuando cada prueba sea discutida.

El Power de una prueba de estadística es la probabilidad de rechazar correctamente la hipótesis nula. La tabla de abajo muestra los 4 posibles resultados de la prueba de hipótesis.

Hipótesis nula

Decisión Verdadero Falso

Decisión correcta p=1-α	Error tipo II p = B
Error tipo I p = α	Decisión correcta p = 1- α (Power)

Falla al rechazar

Rechazar

El nivel α debe ser escogido antes de conducir la prueba:

- Incrementando α incrementas tus posibilidades de detectar una diferencia (y tu Power) pero también incrementas la posibilidad de rechazar H0 cuando es verdad (error tipo I).
- Disminuyendo α disminuyes tus posibilidades de cometer el error tipo I, pero también disminuyes el poder de la prueba.

Intervalos de confianza

Ejemplo 2 peso de la caja de cereal

Problema

Recuerde que esta tratando de confirmar que el embalaje del cereal esté en un objetivo. El objetivo del peso es de 365 gramos y necesitas asegurarte que el proceso de la media esté dentro de 2.5 gramos que es el objetivo.

Recolección de datos

Seis cajas de cereal fueron elegidas al azar y pesadas.

Herramientas

Stat > Basic statistics > 1-sample t

Set de Datos

CEREALBX.MPJ

Nombre	Tipo de dato	Tipo de variable
BoxWeigh	Numérico	Respuesta

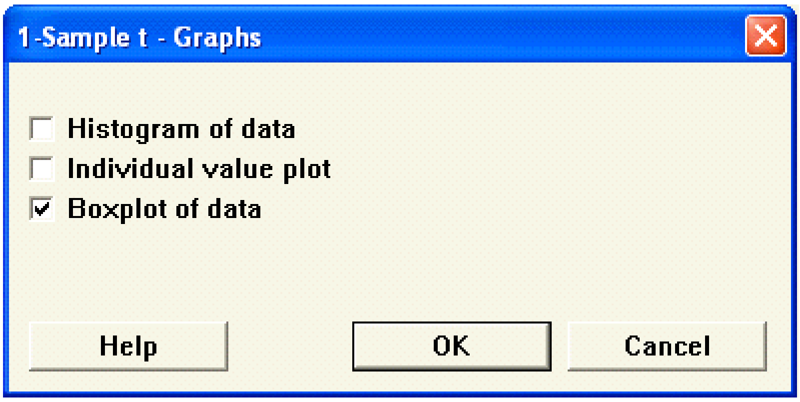
Intervalos de confianza
¿Que es un intervalo de confianza?
Un intervalo de confianza es un rango de posibles valores para un perímetro de una población (tal como µ) que esta basada en un dato de muestra. Por ejemplo, muy seguido usaras una muestra para calcular µ. Un intervalo confidencial te dirá que tan lejos esperes ese cálculo.
¿Cuándo usar el intervalo de confianza?
Usa un intervalo de confianza para hacer inferencias de una o más poblaciones de muestra de datos.
¿Por que usar intervalos de confianza?
Los intervalos de confianza te pueden ayudar a contestar muchas de las mismas preguntas de la prueba de hipótesis:

- ¿Que tan grande podría ser µ?
- ¿Qué tan grande podría ser la desviación estándar de la población?
- ¿Podría µ ser un valor cierto?

Por ejemplo,

- Es posible que la longitud de la media de los lápices sea mayor a 5.75 pulgadas?
- Podría σ para la longitud de los lápices ser tan alto como 0.25 pulgadas?

Usando el intervalo de confianza
En el ejemplo anterior, usamos una prueba de hipótesis para determinar si la media de tu proceso fuera diferente al valor del objetivo. También puedes usar un intervalo de confianza para evaluar ésta diferencia.
Esta Sesión window resulta para 1-sample t incluye valores para los fines mayor y menores del 95% del intervalo de confianza. Obtiene una grafica representativa del intervalo al seleccionar Boxplot en Graphs subdialog box.
1-**Sample t**
1.- Escoge **Stat > Basic Statistics > 1-Sample t, or press** Ctrl + E.
2.- Click **Graphs**
3.- Completa el recuadro como se indica a continuación:

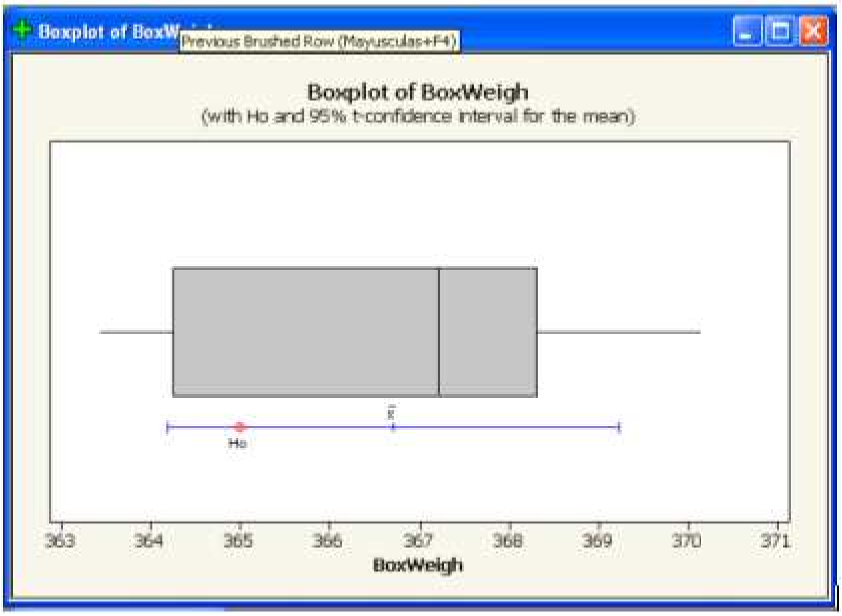


4.- Clik **OK** en cada recuadro.

Interpretando tus resultados
Intervalo de confianza
El intervalo de confianza es un rango de posibles valores para µ. Esta mostrado gráficamente como una línea roja y dos escuadras cuadradas debajo del boxplot.
Es un intervalo de confianza de 95% por que tomamos 100 muestras de la misma población, los intervalos de 95 de las muestras incluirá a µ. Por lo tanto para cualquier ejemplo que pueda ser 95% seguro que la µ está dentro del intervalo de confianza.
Prueba de hipótesis
El punto rojo de la X representa la media de la muestra y el punto azul de H0 representa la prueba de la media (365). Puedes usar el intervalo de confianza para probar la hipótesis nula:

- Si H0 está fuera del intervalo, la p-value para la prueba de hipótesis también será menor que 0.05. Puedes rechazar la hipótesis nula en α–level 0.05.
- Si H0 esta adentro del intervalo, la p-value será mayor que 0.05. No podrás rechazar la hipótesis nula en α-level 0.05.

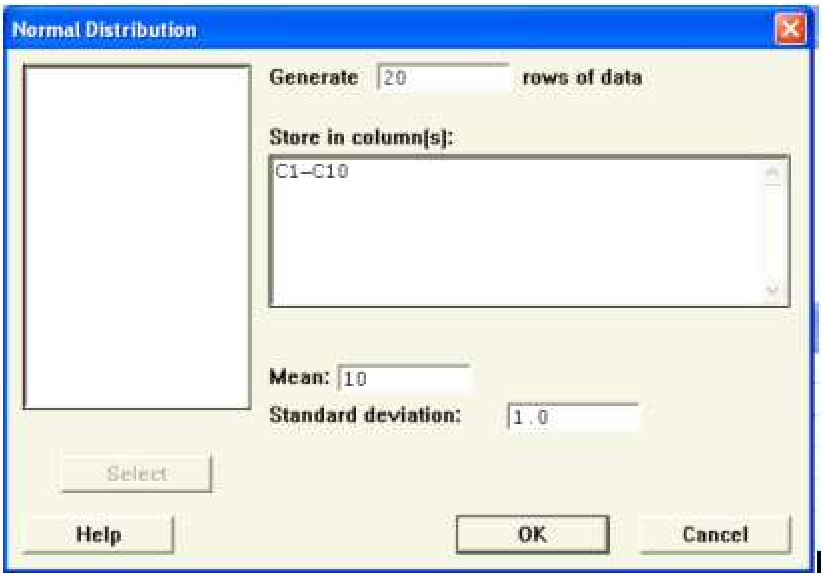
Por que H0 cae adentro del intervalo de confianza no puedes rechazar la hipótesis nula. No hay suficiente evidencia para concluir que µ no es 365 gramos, en el 0.05 nivel significante.



Consideraciones finales
Conclusiones prácticas
El intervalo de confianza de 95% (como el t-test) no provee suficiente evidencia para rechazar la hipótesis nula que la población de la media para el peso de las cajas de cereal sea de 365 gramos.
Consideraciones de Estadística
El intervalo de confianza provee un posible rango para valores de µ(u otros parámetros de población).
En muchos casos, no puedes conducir un prueba de hipótesis usando un intervalo de confianza. Por ejemplo, si el valor de la prueba no es entre un 95% de un intervalo de confianza, puedes rechazar H0 en el nivel α 0.05. Sin embargo si tu estructuras un 99% de intervalo de confianza y no tiene una prueba de la media, puedes rechazar H0 en el nivel α 0.01.
Intervalos de confianza
Ejemplo 3 Entendiendo los intervalos de confianza
Problema
Este ejemplo Explora el concepto de las intervalos de confianza. Simularas la recolección de muestras al azar para una población normal usando MINITAB's generador de números al azar.
Recolección de Datos
Tu debes generar 10 columnas de datos al azar
Herramientas
Calc > Random data > Normal.
Stat > basic Statistics > Display Descriptive Statistics.

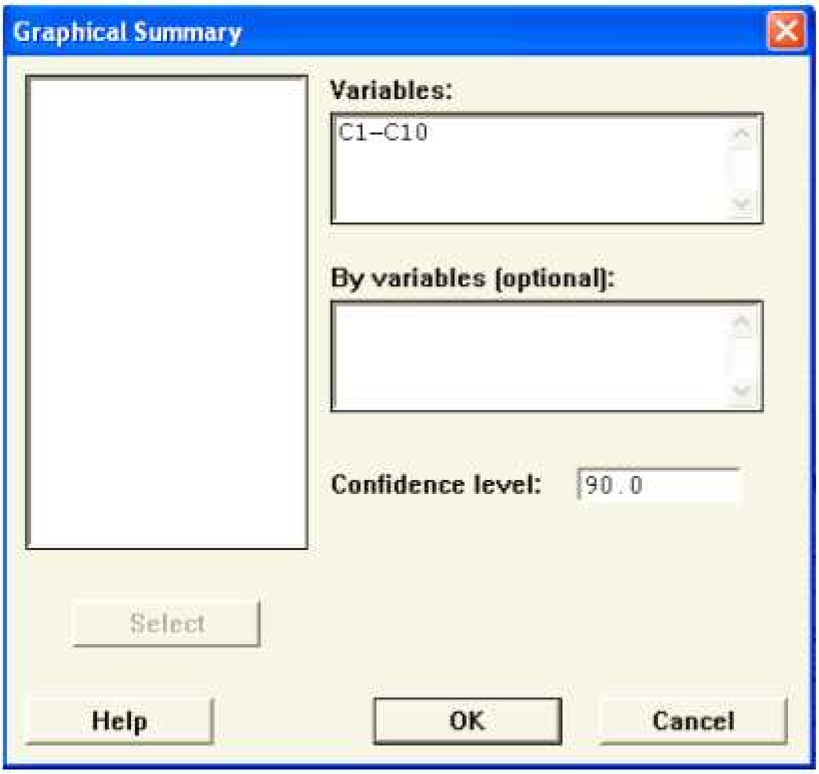
Data set
Ninguno

Generando datos normales al azar
Usando un generado de datos al azar, puedes simular la recolección de datos al azar de una población con una media dada. (Esto es una situación en la cual de hecho puedes saber el valor de µ.)
Usando el generador de datos al azar para simular la colección de 10 muestras de una población con una media(µ) de 10 y de una desviación estándar de 1. Se generan 20 observaciones para cada muestra.
Normal
1. - Escoge **File > New**
2. - Selecciona MINITAB Project.
3. - Click OK.
4. - Escoge **Calc > Random Data > Normal.**
5.- Completa el recuadro como se indica a continuación:



6.- Click **OK** en cada recuadro.

Calculando intervalos de Confianza del 90%
Usa Display Descriptive Statistics para calcular intervalos de confianza del 90% para cada muestra. Por definición, 9 de cada 10 intervalos deben contiener la µ. Desde que sabes que la µ representa muestras que son iguales a 10, puedes verificar esto directamente.
En contraste a los intervalos de confianza del 95%, los de 90% son más angostos(esto es que incluyen menos valores). Porque estos contiene menos valores, es menos probable que contengan la µ.
Para probar la hipótesis nula que la µ no es igual a un valor dado, un intervalo de confianza de 90% corresponde a un .10. nivel de α.
Display Descriptive Statistics
1.- Escoge **Stat > Basic Statistics > Graphical Summary.**
2.- En **Variables**, enter *C1-C10.*
3.- Completa el **Confidence level** como se muestra:



4.- Clik **Ok.**

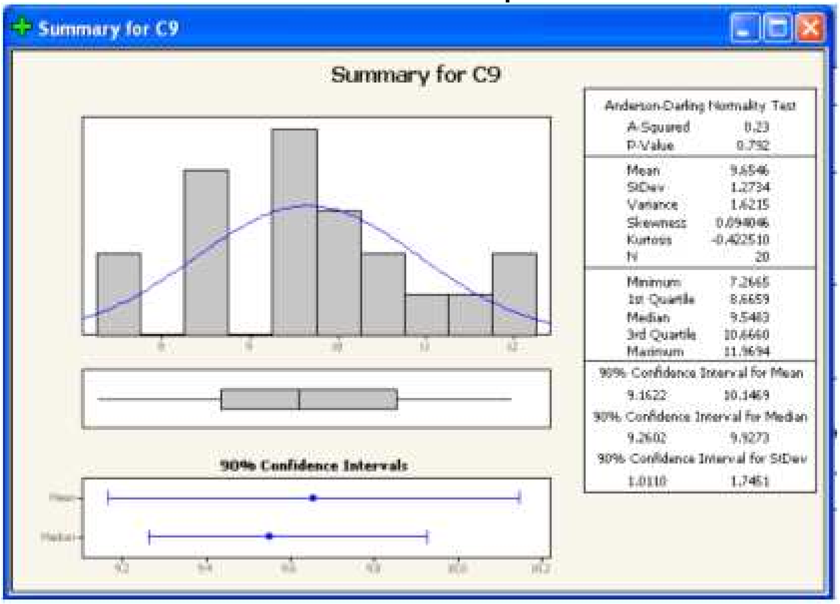
Interpretando tus resultados
90% de Intervalo de confianza para Mu (µ)
Toma un momento para repasar los intervalos de confianza para cada uno de tus muestras, Las opciones seria que uno de tus intervalos no va a contener µ(10).
Es posible que todos tus intervalos contengan µ. También es posible que ninguno lo tenga (aunque es extremadamente inusual). Sin embargo, si repites el ejercicio de la generación de muestras al azar y calculando el intervalo de confianza del 90%, encontraras ese aproximado 90% de los intervalos que contiene µ.
Resultados hipotéticos

Un ejemplo de un intervalo de confianza de 90% que no contenga µ es proveído por derecho. El intervalo se extiende de 10.0275 a 10.7894.
Date cuenta que este ejemplo en particular te llevará a un rechazo incorrecto de la hipótesis nula que µ es igual a 10 (asumiendo que escojas el nivel α de 0.10).

Intervalo de confianza de 90% para sigma

Date cuanta que la suma gráfica también incluye a un intervalo de confianza de 90% para σ (la desviación estándar de la población). El intervalo tiene en rango de 0.7882 a 1.3501. Si repites este procedimiento para un numero largo de muestras, cerca de 9 de 10 intervalos incluirá el valor para σ.

Estadísticas Descriptivas



Consideraciones finales

Conclusiones prácticas

Es probable que 1 de 10 intervalos de confianza de 90% que calcules no contengan µ. Si este procedimiento fuera repetido para un número largo de muestras, cada 10% de todos los intervalos confidenciales de 90% no tendrán µ.

Consideraciones de estadística

Este intervalo de confianza provee un rango de valores para µ(ó los parámetros de la población).
En promedio, el 90% de los intervalos de confianza de 90% calculados para muestras al azar tomado de una distribución normal de poblaciones incluirá a µ.

Power

Ejemplo 4 Evaluando el Power

Ejercicio

No estas seguro que confías en el resultado del análisis del llenado del peso (página 1-6). Vas a conducir el análisis del Power para determinar si recolectaste suficientes datos
Quieres asegurarte que el llenado de las cajas no difiera del objetivo del peso de 365 gramos no más de 2.5 gramos.

Recolección de datos

Vas a basar el análisis del Power en los resultados del t-test del ejemplo 1.

Herramientas

Stat> Power and Sample Size> 1-Sample t

Data set

Ninguno

Análisis del Power

¿Que es un análisis del Power?

Power es la habilidad de una prueba para detectar un efecto cuando existe. Cuando conduces una prueba de hipótesis, hay 4 posibles resultados:

Hipótesis nula

Decisión correcta p =1-α	Error tipo II p = β
Error tipo I p = α	**Decisión correcta** **p = 1- α** **(Power)**

Decisión Verdadero Falso

Falla al rechazar

Rechazar

El Power de la prueba es la probabilidad que rechazara la hipótesis nula correctamente, dado que la hipótesis nula es falsa. Puedes usar un análisis Power para determinar cuanto poder tiene esta prueba, o ayudar a designar una nueva prueba para que tenga el poder adecuado.

Cuando usar un análisis del Power

Usa un análisis del Power cuando estas diseñando un experimento o después de conducir una hipótesis nula. No se requieren datos. Necesitaras estimar σ (excepto por las pruebas de proporción).

¿Por qué usar un análisis del Power?

El análisis del Power te puede ayudar a responder preguntas como:

- ¿Es tu muestra lo suficiente grande?
- ¿Qué diferencia puedes detectar con tu prueba?
- ¿Deberías confiar en los resultados insignificantes de la prueba ?

Por ejemplo,

- ¿Cuántas muestras necesitas recolectar para determinar si el papel de proveedor es más delgado que el de otro por 0.0015 pulgadas?
- ¿Qué tan grande es la diferencia que puedes detectar entre la resistencia de una viga de acero y un historial de la media si reúnes 8 muestras?
- ¿Puedes confiar en los resultados de una prueba t-test que indica la resistencia de 2 fórmulas de pegamento que no tienen diferencia?

Determinando el Power

Tu meta es determinar que tan ciertos son los resultados del análisis del llenado de las cajas de cereal (pagina1-6)

Valores

Si especificas valores para cualquiera de los 2 parámetros de la prueba, MINITAB calculará el parámetro restante:

- Sample size----- el número de observaciones en la muestra
- Differences----- un significado cambio en el alejamiento del objetivo que estas interesado en detectar con alta probabilidad.
- Power values----- el poder (probabilidad de rechazar H0 cuando es falso) que te gustaría que tuviera la prueba.

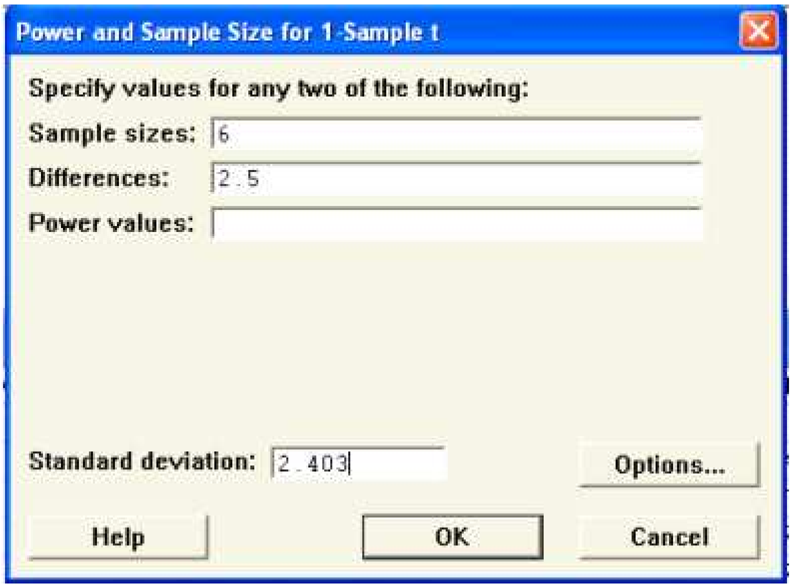
Sigma

Porque el poder de una prueba es parcialmente determinada por la variabilidad en los datos, debes proveer un estimado para σ . Usa un estimado del historial o la desviación estándar de la muestra.

1- Sample t

1.- Escoge **Stat > Power and Sample Size > 1-Sample t.**

2.- Completa el recuadro como se indica a continuación:



3.- Click **OK.**

Interpretando tus resultados

Con 6 observaciones, una desviación estándar de 2.043 y un α de 0.05, el Power solo es de .5376. Esto significa que µ esta fuera del objetivo por 2.5 gramos, solo tienes un 53.76% de oportunidad para detectarlo.

De otra manera, hay un 46.24% de probabilidad que falles al rechazar H0 e incorrectamente concluye que el proceso está en el objetivo.

¿Qué sigue?

De manera que incrementes tu probabilidad de detectar un cambio si existe, es incrementar el tamaño de la muestra. Determinar él numero de observaciones requeridas para lograr el Power adecuado.

Power and Sample Size

1-Sample t Test

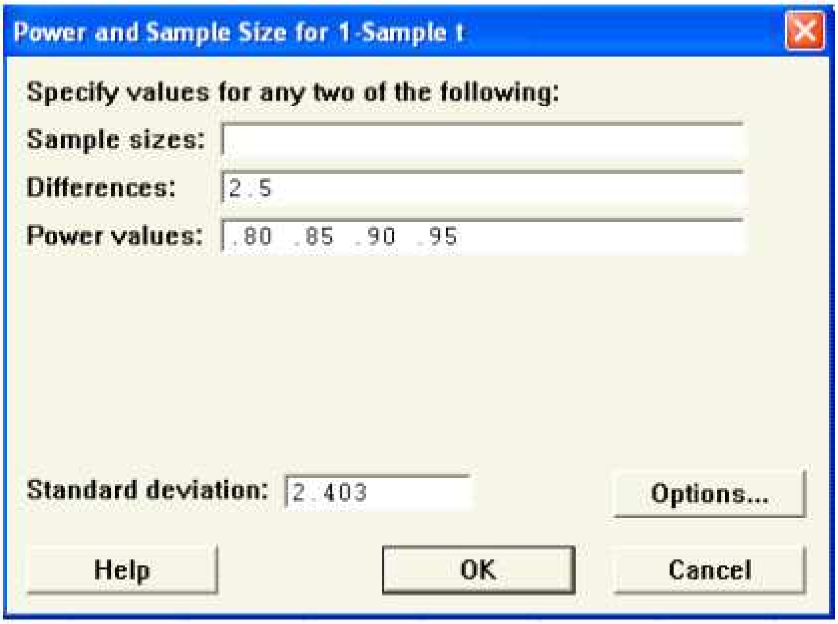
Testing mean = null (versus not = null)

Calculating power for mean = null + difference

Alpha = 0.05 Assumed standard deviation = 2.403

	Sample	
Difference	Size	Power
2.5	6	0.537662

Determinando el Power
Con 6 observaciones el Power de tu prueba fue solo de 0.5376. Para tener mejores posibilidades de detectar un efecto si es que existe, deberás incrementar el poder de tu prueba, que por lo menos sea de 0.80 (como regla general).
Calcular el tamaño de la muestra requerida para llegar los niveles de Power de 0.80, 0.85, 0.95, y 0.95.
1-Sample t
1.- Escoge **Stat > Power and Sample Size > 1-Sample t.**
2.- Completa el recuadro como se indica a continuación:



3.- Clic **OK.**

Interpretando tus resultados
Para tener un Power de al menos 0.80 (objetivo del Power) para detectar una diferencia de 2.5 gramos al nivel α de 0.05, necesitaras una muestra de tamaño 10.
Porque el tamaño de las muestras debe ser siempre un numero entero. El Power actual de la prueba con 10 observaciones (0.8327) es escasamente mayor que el objetivo Power.
Observaciones adicionales que dan mas Power:

- Con 11 observaciones, el Power es de 0.8739.
- Con 12 observaciones, el Power es de 0.9058.
- Con 15 observaciones, el Power es de 0.9625.

Al duplicar el tamaño de la muestra de 6 a 12 cajas, incrementas tus posibilidades de detectar una diferencia de 2.5 gramos (sí es que existe) de 53.76% a 90.58%.
Tal ves no quieran incrementar tu Power demasiado. Si tu Power es demasiado alto, podrías empezar a detectar cambios que son demasiado pequeños para ser parcialmente importantes.

Power and Sample Size
1-Sample t Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 2.403

	Sample	Target	
Difference	Size	Power	Actual Power
2.5	10	0.80	0.832695
2.5	11	0.85	0.873928
2.5	12	0.90	0.905836
2.5	15	0.95	0.962487

Power
Ejemplo 5 incrementando Power
Ejercicio

El resultado del análisis de tu Power sugiere que necesitas una muestra más grande para evaluar tu proceso. Con solo 6 observaciones, había muy poco Power para detectar un diferencia de 2.5 gramos

Recolección de datos

12 cajas de cereal son recolectadas al azar y pesadas

Herramientas

Stat> Basic statistic> 1-sample t

Data set

CEREALBX.MPJ

Nombre	Tipo de dato	Tipo de variable
BoxWeigh	Numérico	Respuesta

Probando la hipótesis nula

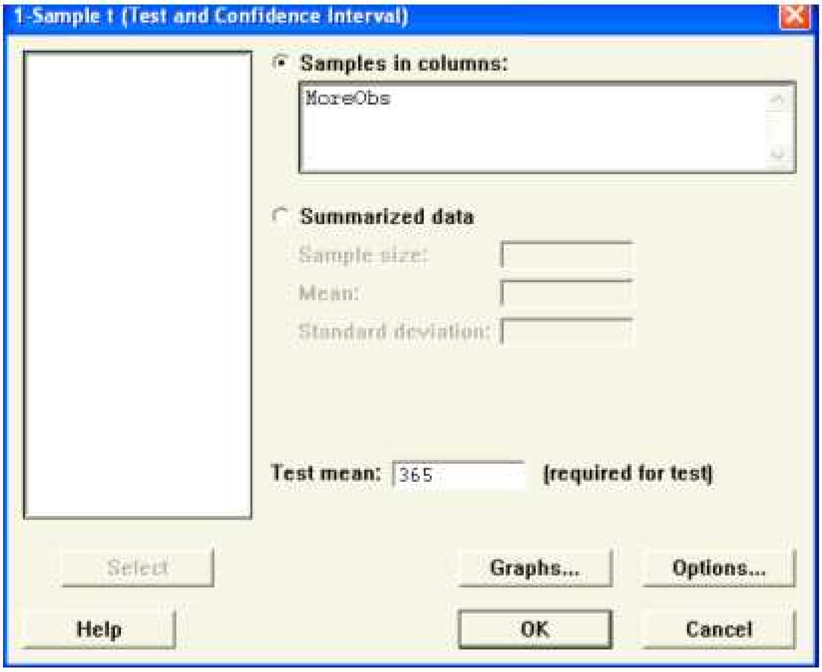
Analiza la nueva muestra para determinar si el proceso de la media es diferente a 365 gramos.

1-Sample t

1.- Abre el proyecto CEREALBX.MPJ.

2.- Escoge **Stat> Basic Statistics> 1-Sample t**

3.- Completa el recuadro como se indica a continuación:



4.- Haz clic en **Graphs.**

5.- Checa **Boxplot of data.**

6.- Click **OK** en cada recuadro.

Interpretando tus resultados

El t-test indica que la diferencia entre el proceso de la media y el objetivo de 365 gramos es significante en el nivel α 0.05

- El p-value (0.019) es menos que α (0.05).
- El intervalo de confianza de 95% no incluye el valor del objetivo.

Aparece que las cajas de cereal están siendo sobre llenadas. Se deben tomar acciones correctivas para ajustar el proceso.

One-Sample T: MoreObs

Test of mu = 365 vs not = 365

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
MoreObs	12	366.636	2.060	0.595	(365.327, 367.945)	2.75	0.019

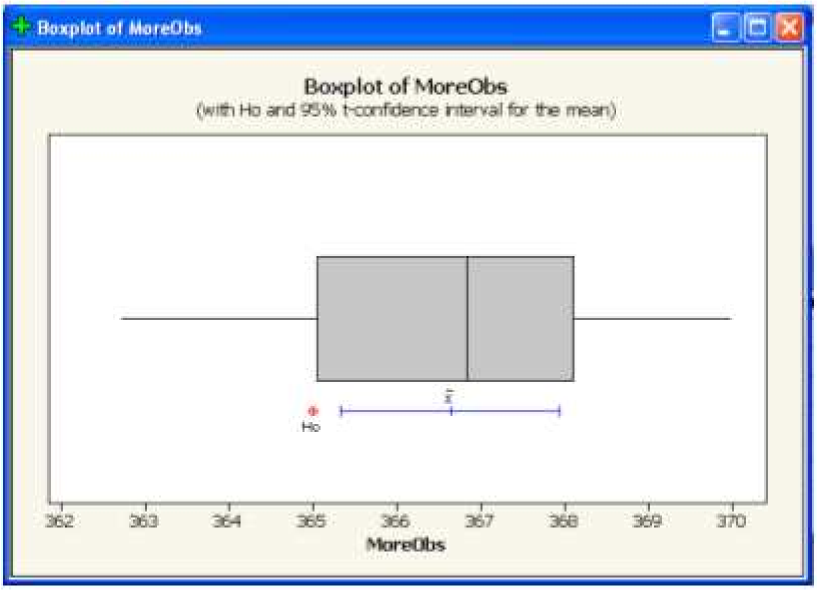
Interpretando tus resultados

El boxplot ilustra lo que encontró la prueba:

- El valor del objetivo(H0) esta afuera del intervalo de confianza.
- La muestra de la media (X) es mayor que el valor del objetivo.

Conclusión

La diferencia entre el proceso de la media y el valor del objetivo de 365 gramos es significante en el nivel α es de 0.05.



Consideraciones finales
Conclusiones prácticas
Es probable que tu primera prueba del llenado de las cajas de cereal no sea significante porque tu Power era demasiado bajo. Basado en el numero de observaciones (6), la diferencia que quieres detectar (2.5), y la variabilidad en los datos, la prueba tuvo un Power de solo 0.5376.
Usando una muestra grande te da mas Power, habilitándote para detectar la diferencia.
Consideraciones estadísticas
Para asegurar que tu prueba tenga suficiente Power, es una buena idea el conducir un análisis power para recolectar datos.
Las maneras de incrementar el Power de una prueba incluye:

- Incrementar el tamaño de la muestra.
- Disminuir la variabilidad que no esta atribuida al efecto de interés.
- Incrementar α (aunque esto también te llevara a incrementar un error del tipo I).

Mayor Power significa una mayor probabilidad de detectar los errores. Sin embargo también incrementa la probabilidad de detectar errores pequeños que puede que no sean de interés. El proceso del conocimiento ayuda a determinar el nivel optimo del Power en una prueba.

Ejercicio 5.1 Detectando posibilidades en el diámetro de un balero
Ejercicio
Una parte del Balero manufacturado está fuera de especificaciones 0.05 cm de lo correcto. Un cambio de 0.01cm es considerado lo suficientemente importante para permitir el ajuste al equipo.
La desviación estándar de los diámetros es casi siempre de 0.004 cm.
Recolección de datos
Ninguno
Instrucciones

1. Use **Stat > Power and sample size > 1-sample t** para calcular el tamaño de la muestra necesitaras detectar una diferencia de 0.01cm con el Power de 0.85 en un nivel α de 0.05
2. Calcular las diferencias puedes detectarlas con un power de 0.90 cuando recolectes 5 y 10 observaciones.

Data set
Ninguno
Power and Sample Size
1-Sample t Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.05

Difference	Sample Size	Target Power	Actual Power
0.5	3	0.85	1.00000

Power and Sample Size
1-Sample t Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.05

Sample Size	Power	Difference
5	0.9	0.0982944

Power and Sample Size
1-Sample t Test
Testing mean = null (versus not = null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.05

Sample Size	Power	Difference
10	0.9	0.0577282

PRUEBA T Y PRUEBAS DE PROPORCIÓN

Objetivos

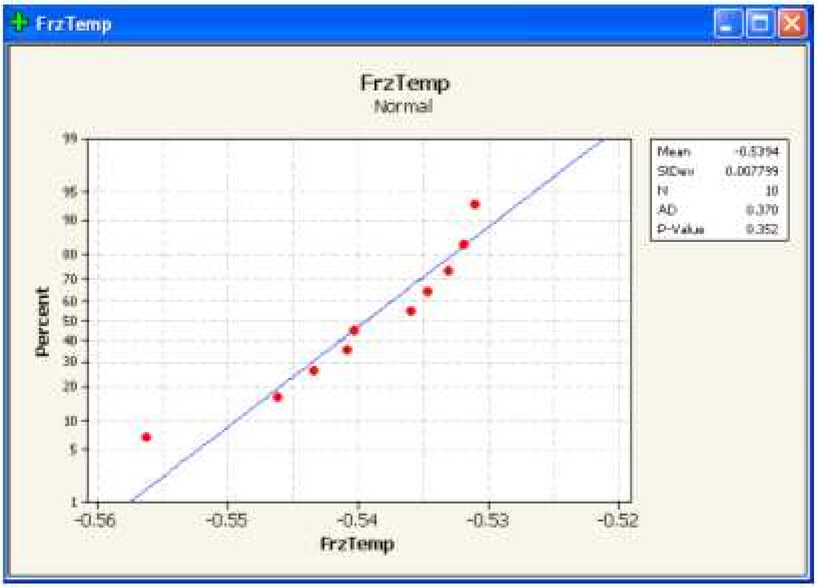
- **Evaluar la diferencia entre la media del proceso y un valor de un objetivo usando un One-Sample t-test.**
- **Evaluar la diferencia entre 2 muestras de la media usando en Two-Sample t-test.**
- **Evaluar las diferencias entre 2 observaciones usando un Paired t-test.**
- **Evaluar la diferencia entre una proporción y un valor de un objetivo usando una prueba de una proporción.**

Contenidos

One-Sample t-Test
Ejemplo 1 Problema del Gran Queso
Ejercicio
Tu compañía, El Gran Queso, Inc., sospecha que uno de tus proveedores de leche le esta añadiendo agua a su leche para incrementar sus beneficios. Añadir agua a la leche incrementa su punto de congelación, que normalmente es de –0.545º C.
Recolección de datos
El punto de congelación es medido para 10 muestras al azar de la leche del proveedor.
Herramientas
Stat> Basic Statistics> Normality Test.
Stat> Basic Statistics> 1-Sample t.

Data set
CHEESE.MPJ

Nombre	Tipo de Dato	Tipo de Variable
FrzTemp	Numérico	Respuesta



One-Sample T: FrzTemp

Variable	N	Mean	StDev	SE Mean	95% CI
FrzTemp	10	-0.539368	0.007799	0.002466	(-0.544947, -0.533790)

One-sample t-test

¿Qué es un One-Sample t-test?

Un One-Sample t-test te ayuda a determinar si µ (la población de la media) es igual a un valor d una hipótesis (la prueba de la media).

La prueba utiliza desviaciones estándar de una muestra para estimar σ (la desviación estándar de la población). Si la diferencia entre la muestra de la media y la prueba de la media es grande relativamente a la variabilidad en la muestra, entonces µ es improbable que sea igual a la prueba de la media.

¿Cuándo usar un one-sample t-test?

Usa un one-sample t-test cuando tienes datos continuos de una sola muestra al azar.

La prueba asume que la población esta distribuida normalmente. Sin embargo es muy justo a las violaciones de esta suposición, proveídas las observaciones son recolectadas al azar y los datos son continuos y racionalmente simétricos. *(ver Box, Hunter & Hunter (1978). Statistics for Experimenters, John Wiley & Sons, Inc.).*

¿Por qué usar un one-sample t-test?

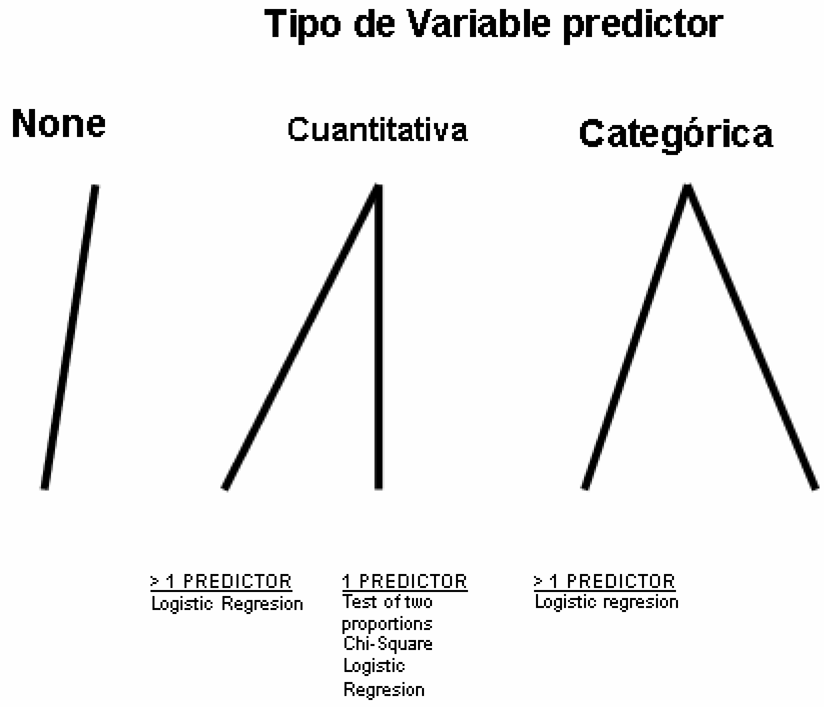
Un one-sample t-test te puede ayudar a responder preguntas tales como:

- ¿Esta el proceso en el objetivo?
- ¿El producto de tu proveedor cumple con tu criterio?

Por ejemplo,

- Es el ancho de la media de las navajas mayor o menor que el objetivo?
- Es la resistencia de la media de los tornillos de tu proveedor menor de lo requerido?

Guías al Escoger las Herramientas de Estadística
Tipo de Variable de Respuesta



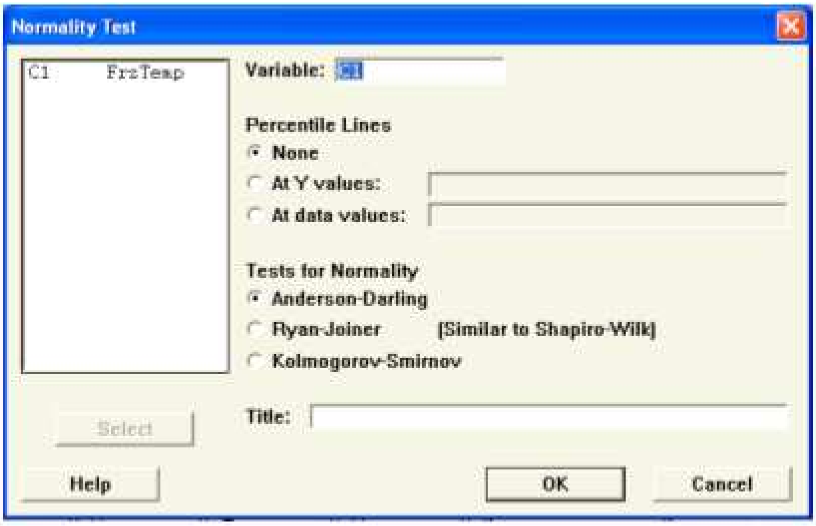
<u>1 PREDICTOR</u>
Logistic Regresion
Test of One Proportion
Chi-Square
One-Sample t-Test
Correlación (dos respuestas)
<u>> 1 PREDICTOR</u>
Múltiple Regresión
AN OVA Factorial Desings

<u>1 PREDICTOR</u>
Twp-Sample t-Test
One-Way >NOVA
<u>> 1 PREDICTOR</u>
Múltiple regresión
Respuesta Surface
<u>1 PREDICTOR</u>
Regresión Simple

Probando la suposición de una normalidad
La prueba de Estadística apropiada para los datos de la temperatura congelante es un one-sample t-test. Esta prueba asume que la población esta normalmente distribuida.
Usa una prueba de normalidad para determinar si la suposición de la normalidad es valida para esos datos.

Prueba de normalidad

1. Abre el proyecto CHEESE.MPJ
2. Elige **Stat> Basic statistics> Normality Test**
3. Completa el recuadro como se indica a continuación:

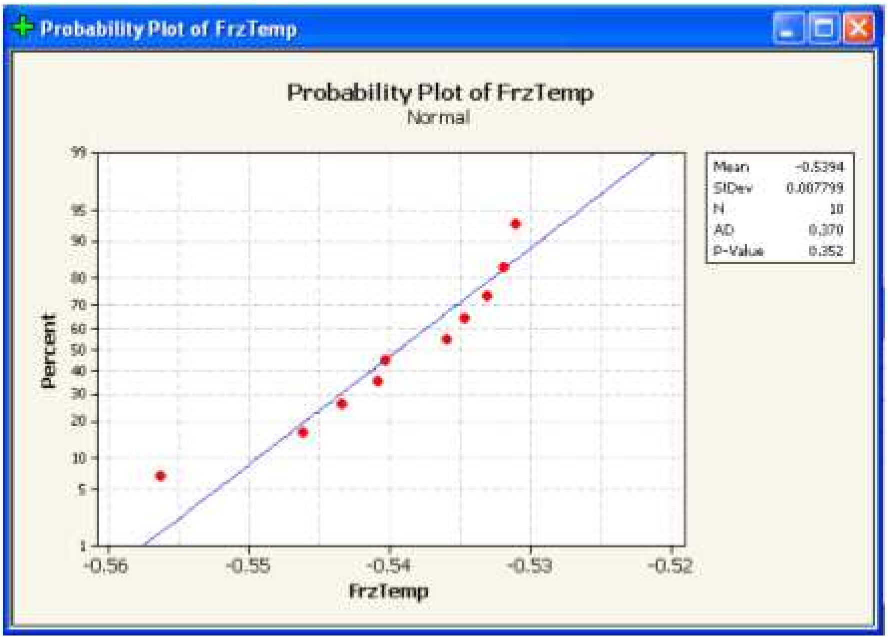


4. Haz clic en **OK**

Interpretando tus resultados
Usa el normal probability plot para verificar que tus datos no se desvíen significativamente de una distribución normal.

- Si los datos vienen de una distribución normal, los puntos muy apenas seguirán la línea de referencia.
- Si los datos no vienen de una distribución normal, los puntos no seguirán la línea.

Anderson-Darling normality test
Un p-value de Anderson-Darlin Test (0.0352) accesa a la probabilidad que los datos son de una población con distribución normal. Usando en α de 0.05, no hay suficiente evidencia para sugerir que los datos no son de una población normal.
Conclusión
Basado en el argumento y en la prueba es razonable asumir que tus datos no se desvían substancialmente de una distribución normal. Puedes proceder con el t-test.



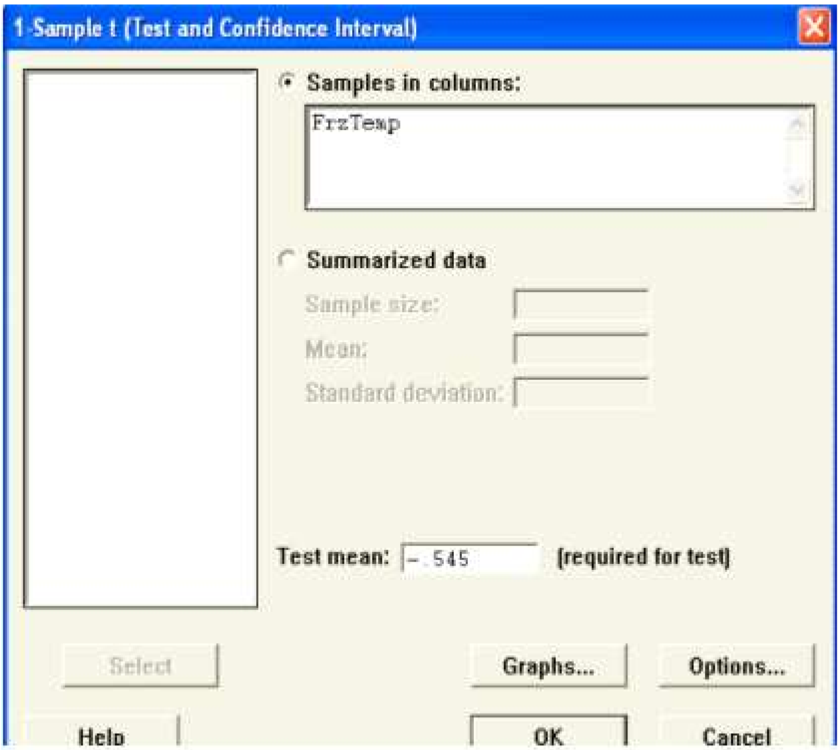
Conduciendo el 1-sample t-test
Conducir un 1-sample t-test para determinar si la temperatura congelante de la leche del proveedor es mayor a –0.545º C.
No hay razón para sospechar que el proveedor quitara el agua de la leche. Así, no necesitas probar si la temperatura congelante es menor que –0.545º C. En esta situación, puedes usar una prueba 1-tailed (en la cual H1 es direccional):

- H0 :µ = -0.545
- H1 :µ > -0.545 (En una prueba 2-tailed, H1 No es direccional: µ es diferente a –0.545)

La ventaja de la prueba 1-tailed es que te da mas Power para detectar la diferencia especificada. Sin embargo, una prueba 1-tailed no puede detectar una diferencia en la dirección contraria que especifica en H1. De esta manera si hay diferencias en ambas direcciones son de interés, deberás usar una 2 tailed test.

1-Sample t

1. Escoge **Stat> Basic Statistics> Sample t.**
2. Completa el recuadro como se indica a continuación:



3. Haz click **Options.**
4. De **Alternative**, Escoge **greater than**.
5. Click **OK** en cada recuadro.

Interpretando tus resultados
Usa un nivel α de 0.05 para la prueba.
T
El t-statistic (2.28) es calculado de esta manera:
T = (muestra de la media – prueba de la media) / SE media
Donde SE media es el error estándar de la media (una medida de variabilidad). Como el valor de t se incremente, el p-value se hace mas pequeño.
P
El p-value es 0.024. Porque este valor es menor que α(0.05), puedes rechazar la hipótesis nula. El resultado sugiere que el agua o cualquier otro liquido halla sido añadido a la leche.
Power
Cuando sea apropiado, una prueba 1-tailed es mas poderosa que una prueba 2-tailed. Por ejemplo, una prueba 2-tailed (H1 : µ es diferente a –0.545) regresa a p-value de 0.048, que es mayor que 0.024.

One-Sample T: FrzTemp
Test of mu = -0.545 vs > -0.545

					95% Lower		
Variable	N	Mean	StDev	SE Mean	Bound	T	P
FrzTemp	10	-0.539368	0.007799	0.0024	-0.5438	2.28	0.024

Consideraciones finales
Conclusiones prácticas
El 1-tailed, 1-sample t-test sugiere que la temperatura congelante de la leche del proveedor es mayor a la que debe ser, indicando que se le pudo haber añadido agua. Esta es una acusación muy seria para el proveedor. Podría ser mejor evaluar que tan cierto es antes de tomar una decisión.
Con un nivel α de 0.05, las probabilidades de haber concluido que se le ha añadido agua cuando no es así son de 5%. Para estar seguro que no rechaces H0 incorrectamente, deberás Escoger valores menores para α, tales como 0.01 o hasta 0.001. Con un α de 0.01, no concluirás que se le allá añadido agua a la leche (p = 0.024).

Consideraciones estadísticas
Cuando uses una 1-sample t-test:
- Tu muestra debe de ser al azar.
- Los datos de muestra deben de ser continuos .
- Los datos de muestra deben de distribución normal.

Debe de ser notado que los procedimientos del t-test son muy justas a las violaciones de las suposiciones de normalidad, dadas esas observaciones son recolectadas al azar y los datos son continuos y racionalmente simétricos. *(ver Box, Hunter & Hunter (1978). Statistics for Experimenters, John Wiley & Sons, Inc.).*

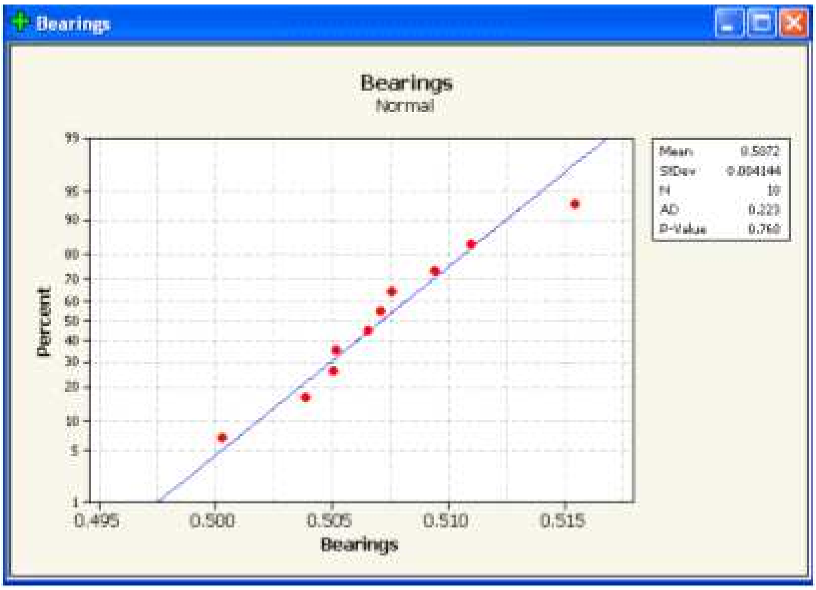
Una prueba 1-tailed es mas poderosa que una prueba 2-tailed. A menos que la diferencia no este en la dirección esperada, Por ejemplo una prueba 1-tailed con una hipótesis alternativa, H1 : µ > -0.545 nunca será capaz de detectar la diferencia si alguien disminuye la temperatura congelante de la leche.

Ejercicio 1.1 Diámetro de los Valeros de Bola
Ejercicio
Tu compañía produce Valeros de bola y necesitas verificar que el tamaño del Balero que esté en las especificaciones. La especificación del diámetro para los Valeros es de 0.5cm.
Usa un nivel α de 0.05 para todas las pruebas.
Recolección de datos
10 Valeros son escogidos al azar y medidos.
Instrucciones

1. Prueba la muestra de normalidad usando **Stat> Basic Statics> Normality Test.**
2. Usa **Stat> Basic Statistics> 1-sample t** para determinar si el proceso esta en el objetivo. Conduce una prueba 2-tailed (H1 : µ es diferente a 0.5) y crea un boxplot de los datos.
3. Usando la desviación estándar de la muestra como un estimado de σ, ¿cuál es el Power de la prueba para detectar una diferencia de 0.005cm.?
4. ¿Cuál es el tamaño mínimo para la muestra requerida para detectar la misma diferencia con un Power de 0.80?

Data set
BEARINGS.MPJ

Nombre	Data type	Variable type
Bearings	Numeric	Response



Two-Sample t-Test
Ejemplo 2 Resistencia plástica
Ejercicio
Tu compañía hace estuches de plástico para calculadoras. Necesitas comparar muestras de plásticos de 2 proveedores en cuanto a su resistencia. El proveedor A dice tenar el plástico mas fuerte, pero cuesta mas que del proveedor B.
Recolección de datos
Pellets seleccionadas al azar de un grupo de plástico son prensadas en agua hasta ser barquillas del mismo grueso. La resistencia para romperlos(en psi, libra por pulgada cuadrada) es tomada para cada barquilla.
Herramientas
Stat> Basic Statistics> Normality Test
Stat> Basic Statistics> 2 variances
Stat> Basic Statistics> 2-sample t

Set de Datos
PLASTIC.MPJ

Nombre	Data type	Variable type
SupplrA	Numeric	Response
SupplrB	Numeric	Response

Two-sample t-test
¿Qué es un two-sample t-test?
Una two-sample t-test te ayuda a determinar si 2 poblaciones de la media son iguales.

La prueba usa las desviaciones estándar de la muestra para estimar σ para cada población. Si la diferencia entre la muestra de la media es grande relativamente para la variabilidad estimada entre las poblaciones, entonces la media de la población son improbables a ser iguales.
Un two-sample t-test también te puede ayudar a evaluar si la media de 2 poblaciones es diferente por una cantidad especifica.

¿Cuándo usar una prueba two-sample t-test?

Usa una prueba two-sample t-test cuando tengas datos continuos de 2 muestras al azar independiente. Las muestras son independientes si las observaciones de un one.sample no están relacionadas a las observaciones de la otra muestra. Por ejemplo, 2 medidas son tomadas por un mismo operador no son independientes.
La prueba también asume que tus datos vienen de una población normalmente distribuida. Sin embargo es muy justo hacia las violaciones de esta suposición proveídas las observaciones son recolectadas al azar y los datos son continuos y razonablemente simétricos. *(ver Box, Hunter & Hunter (1978). Statistics for Experimenters, John Wiley & Sons, Inc.).*

¿Por que usar una prueba two-sample t-test?

Un two-sample t-test te puede ayudar a contestar preguntas tales como:

- ¿Son los productos de dos proveedores comparables?
- ¿Es la formula de un producto mejor que el otro?

Por ejemplo,

- ¿Es similar la viscosidad del aceite de dos proveedores?
- ¿Es la formula de una tinta más brillante que otra?

Probando las suposiciones de la normalidad

La prueba de estadística mas apropiada para los datos del proveedor es la two-sample t-test. Esta prueba asume que los datos son de poblaciones distribuidas normalmente.
Usa la prueba de normalidad para determinar si la suposición de la normalidad es valida para estos datos.

Prueba de normalidad

1. Abre el proyecto PLASTIC.MPJ.
2. Escoge **Stat> Basic statistics> Normality Test.**
3. En Variable, enter ´SupplrA´.
4. Click **OK.**
5. Escoge **Stat > Basic Statistics > Normality Test**, or press ctrl. + E.
6. En **Variable,** enter ´SupplrB.
7. Click **OK**

Interpretando tus resultados

Usa la normal probability plot para verificar que tus datos no se desvíen significativamente de una distribución normal.

- Si los datos vienen de una distribución normal, los puntos muy apenas seguirán la línea de referencia.
- Si los datos no vienen de una distribución normal, los puntos no seguirán la línea.

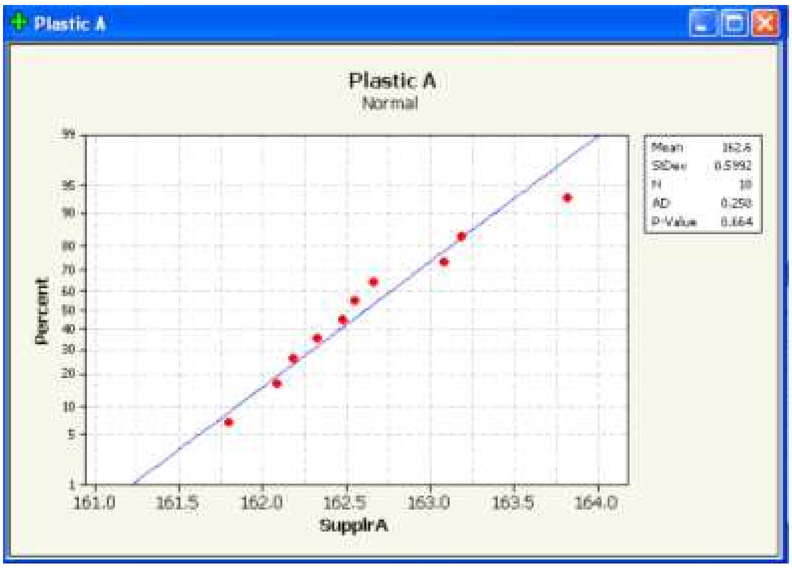
El plot para el SupllrA indica que la distribución de la muestra es razonablemente normal; todos los puntos están cerca de la línea.
El plot para el SupplrB sin embargo aparentemente muestra desviación de la normalidad.

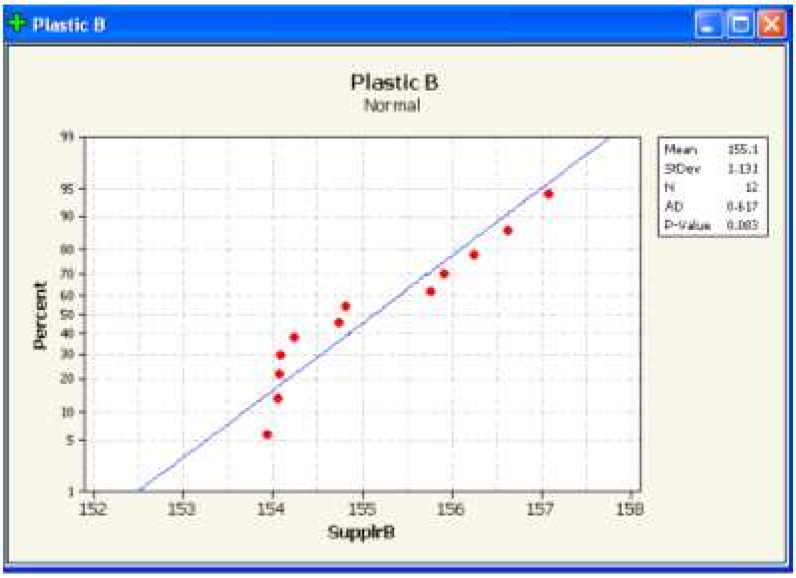
Anderson-Darling Normality test

La desviación de la normalidad observada de SupplrB no es significante en un nivel α de 0.05. Ambas p-values(0.664 para SupplrA, y 0.083 para SupplrB) son mayor que 0.05.

Conclusión

Basado en el plot y las pruebas, es racional asumir que tus datos no se desvían substancialmente de una distribución normal. La suposición de una normalidad es relativamente satisfecha, así que puedes proceder con el t-test.





Comparando las variaciones

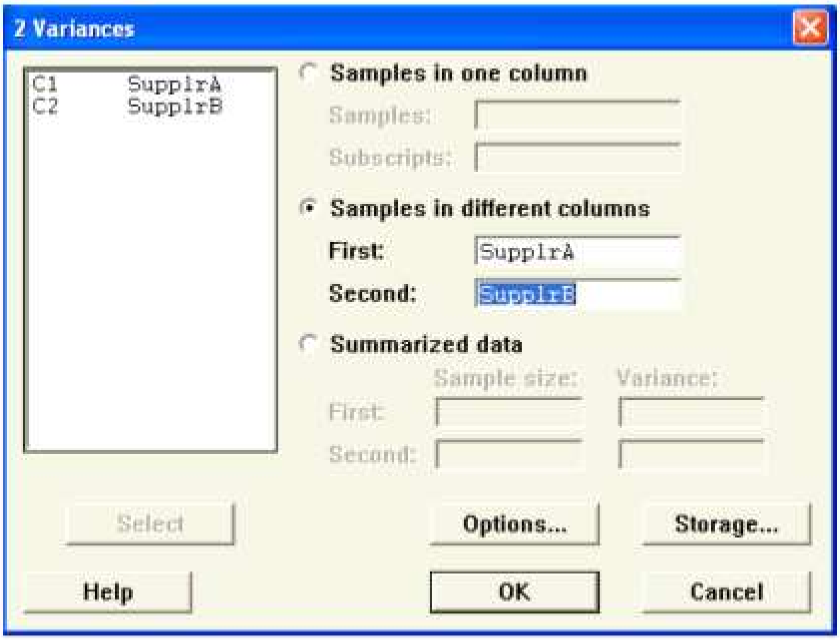
Antes de conducir el t-test, debes evaluar las variaciones de las 2 distribuciones para ver si difieren. Hay dos razones para esto:

- Es importante saber sí el producto de un proveedor varia mas que el del otro
- Los cálculos para el two-sample t-test depende si las variaciones de las muestras son iguales o diferentes.

Para asegurar que encontraste una diferencia entre 2 variaciones si es que existe una. Debes usar un nivel α de 0.10 para esta prueba en lugar de la normal de 0.05. Esto incrementara el power de la prueba.

2 Variances

1. Escoge **Stat> Basic statistics> 2 Variances.**
2. Completa el recuadro como se indica a continuación:



3. Click **Options.**
4. En **Confidence level**, enter 90.
5. Click **OK** en cada recuadro.

Interpretando tus resultados

Intervalos de confianza

Los intervalos de confianza son útiles para comparar σ de las 2 poblaciones. Sin embargo, tu decisión acerca de sí las 2 variaciones son iguales será basadas en una apropiada prueba de variación.

Pruebas de variación

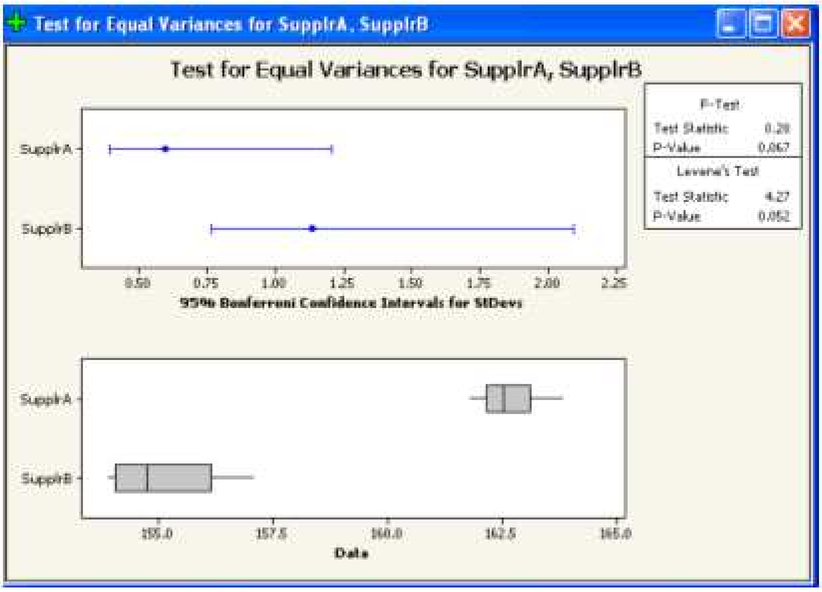
Los resultados incluyen 2 pruebas de variación separadas. El uso de la prueba depende de tus datos.

- Si tus datos son continuos y de distribución normal, usa el F-test.
- Si tus datos son continuos pero no necesariamente de distribución normal, usa el Levene`s test.

Los datos dados son racionalmente normales, así que puedes usar el F-test. Sin embargo, porque el p-value de la prueba de normalidad del proveedor B fue muy baja (0.083), hay que revisar los resultados de la prueba de Levene`s también.

Conclusión

Los p-values para ambos F-test (0.067) y la Levene`s (0.052) son menos que α (0.10), así que puedes rechazar la hipótesis nula que las variaciones son iguales. Los resultados sugieren que las variaciones del plástico del proveedor A son más pequeños que las del proveedor B.



F-Test (normal distribution)
Test statistic = 0.28, p-value = 0.067
Levene's Test(any continuous distribution)
Test statistic = 4.27, p-value = 0.052
Test for Equal Variances for SupplrA, SupplrB

Interpretando tus resultados
Los mismos intervalos de confianza y las pruebas estadísticas incluidas en la ventana de resultados de la grafica también son proveídas en la ventana de sesión.

Test for Equal Variances: SupplrA, SupplrB
95% Bonferroni confidence intervals for standard deviations

	N	Lower	StDev	Upper
SupplrA	10	0.391949	0.59920	1.20658
SupplrB	12	0.764926	1.13118	2.09100

F-Test (normal distribution)
Test statistic = 0.28, p-value = 0.067

Levene's Test(any continuous distribution)
Test statistic = 4.27, p-value = 0.052

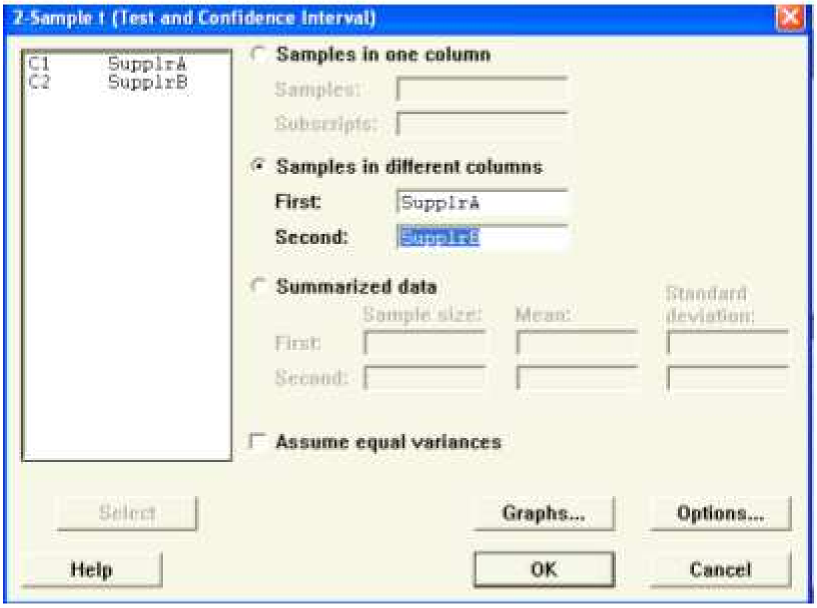
Conduciendo el Two-Sample t-test
Por que los datos son razonablemente normales, tu puedes usar 2 Sample t -to test ya sea la resistencia del plástico de los dos diferentes proveedores.
La prueba de Hipótesis es:

- H0 : µ A – µ B = 0
- H1 : µ A – µ B ≠ 0

Elabora dotplots y boxplots para ayudar a visualizar los datos.

Asumir discrepancias desiguales
Si asumes que las varianzas de las dos poblaciones son iguales, tu t-test será más confiable. Sin embargo, si asumes que la varianza es igual cuando no lo son, los resultados de tu t-test serán falsos. Así, si hay alguna duda, es mejor no asumir que son iguales.
Porque la variance test indica que la población de la varianza es diferente, no asuma que las varianzas son iguales.

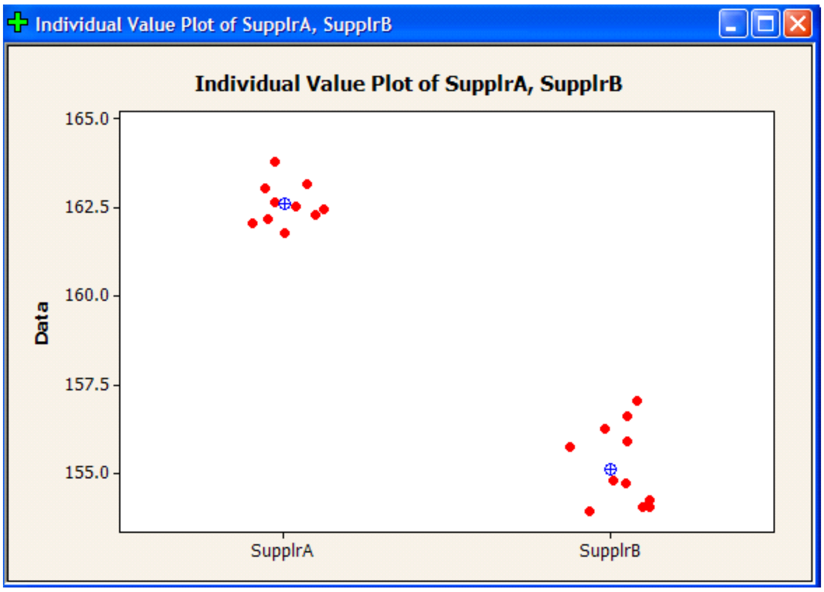
2-Sample t
1. Escoge **Stat > Basic Statistics > 2-Sample t.**
2. Complementa el recuadro como se indica a continuación:

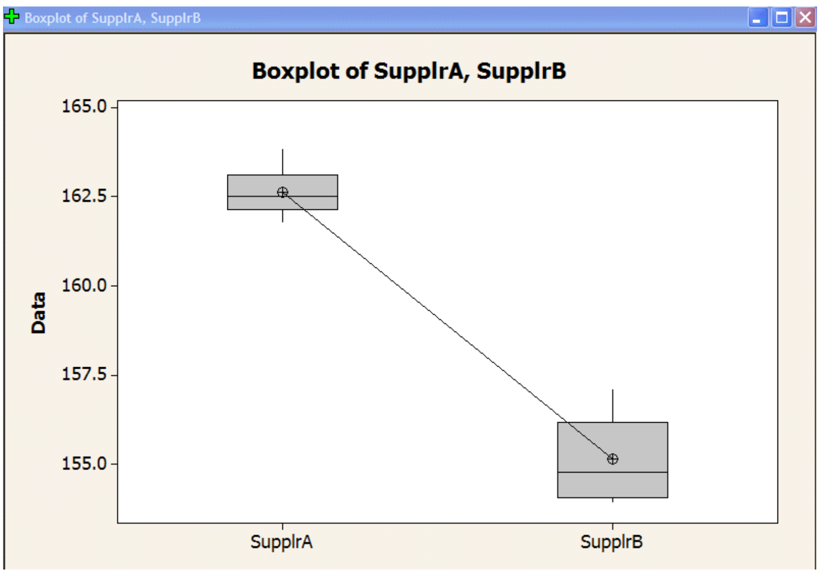


3. Click **Graphs.**
4. Revisa **Dotplots of data** y **Boxplots of data.**
5. Click **OK** en cada recuadro.

Interpretando los Resultados
Las gráficas ilustran dos puntos :

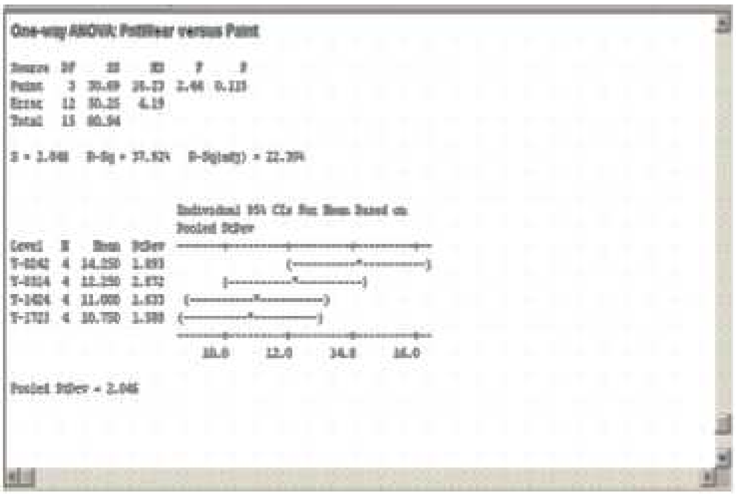
- El plástico del proveedor A se muestra más resistente que el del proveedor B.
- Hay mas variabilidad en la resistencia del Plástico del Proveedor B que del Proveedor A.





Interpretando tus Resultados
Two-Sample T-Test and CI: SupplrA, SupplrB

Two-Sample T-Test and CISupplrA, SupplrB



Individual Value Plot of SupplrA, SupplrB
Boxplot of SupplrA, SupplrB
El promedio del punto de quiebre del plástico (media) y dos medidas de la variabilidad—la desviación estándar (StDev) y el error estándar de la media (SE Mean)—se presentan en cada Proveedor.
Intervalos de Confianza
La diferencia entre la muestra de la media (7.484) se utiliza para estimar la diferencia entre la población de la media (mu SupplrA—mu SupplrB). El intervalo de confianza por la diferencia se basa en esta estimación y la variabilidad de las muestras.
Puede ser 95% confiable que la diferencia entre la población de la media es entre 6.687 y 8.281 psi.

T-value y p-value
El T-value para la prueba es 19.82, lo cual se asocia con un p-value menor que 0.0005 (lo cual se redondea a 0.000)
Así, puedes rechazar la Hipótesis nula en 0.05 α-level, donde concluye que las resistencias son diferentes.

Consideraciones finales
Conclusiones prácticas
El proveedor de plástico A es significativamente resistente y menos variable que el proveedor B. Sin embargo, observamos que el Proveedor A también nos cobra mas por el producto. Ahora tienes que decidir si la diferencia entre los Proveedores es significativa. Se cuenta con el 95% de confianza de la verdadera diferencia entre el proveedor 6.687 y 8.281 psi. Tu decides pagar o no un precio alto por una pequeña diferencia en la resistencia.

Consideraciones estadísticas
Cuando utilizas two-sample t-test :
- Las muestras deben ser al azar.
- Las muestras deben ser independientes.
- Las muestra deben ser continuas.
- Las muestras deber ser de distribución normal.

Debe acentuarse que el procedimiento para la t-test es lo suficientemente veraz a las violaciones de la Asunción de la normalidad, proveídas estas observaciones los datos son recolectados al azar, son continuos, unimodal y razonablemente sistemáticos. *(ver Box, Hunter & Hunter (1978). Statistics for Experimenters, John Wiley & Sons, Inc.).*

Interpretando tus resultados
Dos-ejemplos T para Supp1 rA vs Supp1 Rb

	N	Media	StDev	SE Media
Supp1Ra	10	162.614	0.599	0.19
Suppl1rB	12	155.13	1.13	0.33

Diferencia = mu Supp1rA – mu Supp1rB
Estimación por diferencia: 7.484
95% CI para diferencia: (6.687, 8.281)
T- Testo de diferencia = 0 (vs no =): T- Valor = 19.82 P-Valor = 0.000 DF = 17

La resistencia a ruptura media (medio), y dos medidas de desviación estándar de la variabilidad-(StDev) y del error de estándar del medio (el SE Mean)- se presenta para cada surtidor.

Los intervalos de confianza.

La diferencia entre los medios de la muestra (7.484) se utilizan para estimar la diferencia entre los medios de la población (mu SupplrA-mu SupplrB). El intervalo de la confianza para la diferencia se basa en esta estimación la variabilidad dentro de las muestras.
Usted puede tener una confianza del 95% que la diferencia entre los medios de la población está entre 6.687 y 8.281 PSI.

T-valor y el p-valor

El t-valor para la prueba es 19.82, que se asocia a un p-valor de menos de 0.0005 (que redondeado a 0.000).
Así, usted puede rechazar la hipótesis nula en el 0.05 ά - nivel, y concluye que las fuerzas son diferentes.

Consideraciones Finales.

Conclusiones prácticas.

El plástico de A`s del surtidor es perceptiblemente más fuerte y menos variable que el surtidor B`s. sin embargo recuerda que el surtidor A también carga más para su producto. Ahora usted debe decidir si la diferencia entre los surtidores es de significación práctica.
Usted es el 95% confiable que la diferencia verdadera entre los surtidores es entre 6.687 y 8.281 pis. Usted decide que no está dispuesto a pagar el precio alto más elevado para la pequeña fuerza de diferencia.

Consideraciones Estadísticas.

Al usar una t-prueba de la dos-muestra:

- La muestra debe ser al azar.
- Las muestras deben ser independientes.
- Los datos de la muestra deben ser continuos.
- Los datos independientes de la muestra deben ser distribuidas normalmente

Debe ser observado que el procedimiento de la t-prueba es bastante robusto a las violaciones de la asunción de la normalidad, la condición de que las observaciones se recogen aleatoriamente y los datos son continuos, unimodal, y razonablemente simétricas (véase a la caja, al cazador, y a Cazador (1978). Estadística para Experimentos, John Wiley & Sons, Inc.).

Prueba- t Pareada

3 Ejemplos del Estacionamiento de los Carros.

Problema

Un grupo de consumidor desea determinar si hay una diferencia en la manipulación de capacidad entre dos coches populares. Para medir la capacidad de dirección de los coches, el tiempo lleva conductores el parque paralelo que cada uno de los coches se registra.

Recolección de datos

Veinte conductores parquean ambos coches (en orden al azar), y el tiempo del estacionamiento registrado (en segundos).

Herramientas

Stat > Estadísticas Básicas > Paired

Set de Datos

CARCLT.MPJ

Nombre	Tipo de Dato	Tipo de Variable
Carro - A	Numérico	Respuesta
Carro - B	Numérico	Respuesta

Prueba-t Pareada

¿Qué es una prueba t pareada?

En una prueba t pareada tu puedes determinar si la media de la diferencia entre las observaciones pareadas es significativa Estadísticamente, es equivalente a realizar una Prueba t de una-muestra de una diferencia. Una t-prueba pareada se puede también utilizar para evaluar si la diferencia es igual al valor específico.
Las observaciones pareadas se relacionan de una cierta manera. Los ejemplos incluyen:

- Pesos registrados para los individuos antes y después un programa de ejercicio.
- Muestras tomadas de la misma parte con dos diferentes dispositivos de medida.

¿Cuándo utilizar una prueba t pareado?

Use una Prueba t pareada cuando tengas una muestra escogida al azar de observaciones pareadas. Los datos deben ser continuos.

¿Porqué usar una prueba t pareada?

Las pruebas t pareadas t puede ayudar a responder preguntas tales como:

- ¿Un nuevo tratamiento causa la diferencia en el producto?
- ¿Dos instrumentos de medida hacen lo mismo?

Para el ejemplo:

- ¿Tratando la madera de construcción con ciertos productos químicos aumenta su vida útil?
- ¡Pueden dos calibradores medir idénticas partes de la misma manera?

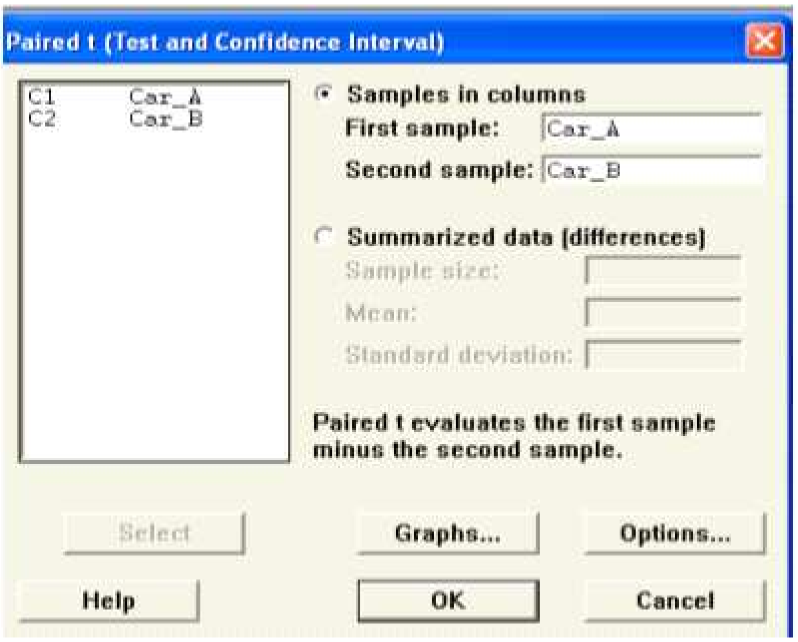
Conduciendo una prueba t de pareada

Tu estas intentando determinar si un coche se puede estacionar más rápidamente que otro. Porque se emparejan los datos (cada individuo estaciono ambos coches), tu utilizaras una prueba t pareado para probar las hipótesis siguientes:

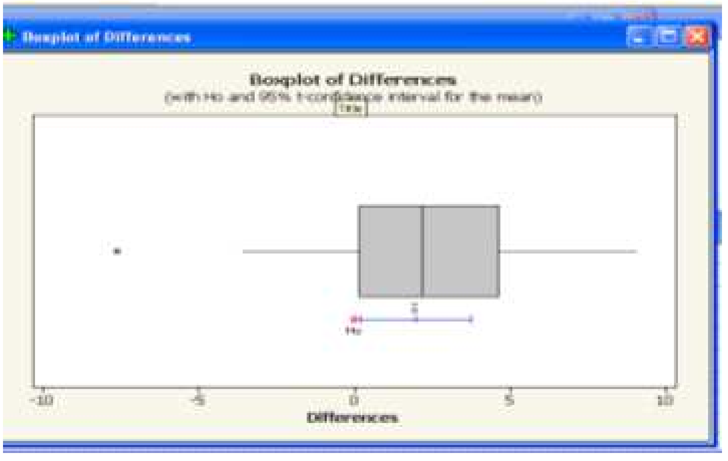
- Ho: La diferencia de la media entre las observaciones pareadas en la población es cero.
- H1: La diferencia de la media entre las observaciones pareadas en la población no es cero.

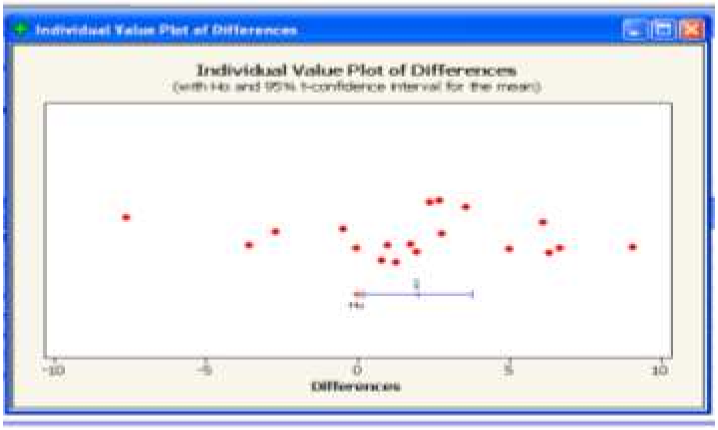
Cree los dotplots y los boxplots para ayudar a visualizar los datos. Utilice el nivel de la confianza del defecto del 95% para la prueba.

t Pareadas
1.- Abre el Project **CARCLT.MPJ.**
2.- Elija **Stat** > Estadísticas básicas > Pareo t.
3.- Completa el recuadro como se indica a continuación:



4.- Click **Graficas.**
5.- Elija **Doplot de diferencias** y **Boxplo**t **de diferencias.**
6.- Click **OK** en cada recuadro.
Interpretando tus resultados
El boxplot y el dotplot ilustran las diferencias entre las observaciones pareadas.
La diferencia de la media (aproximadamente 2) es representa por el X. H_0 representa la diferencia de la población que estas probando (cero.
El intervalo de confianza
MINITAB también dibuja el intervalo de confianza para la diferencia de la media de la población. Así que la hipótesis nula es verdad, tu esperarais que Ho estuviera dentro de este intervalo.
Porque el intervalo de la confianza no esta incluido en Ho, tu puedes rechazar la hipótesis nula y concluir que al coche A le toma mas tiempo estacionar que al coche B.





Interpretando tus resultados
Las medias de los tiempos para estacionarse son 34.87 segundos para el coche A y 32.90 segundos para el coche B. La diferencia es 1.967 segundos.
Los puntos finales para el intervalo de confianza del 95% para la diferencia de la media son de 0.171 y 3.764.
T-valor y p-valor
La prueba da un valor de t de 2.29, se asocia con un p-valor de 0.034. Así, tu puedes rechazar la hipótesis nula en el nivel 0.05 ά y concluir que el tiempo requerido para estacionar el coche A es mayor que el tiempo requerido para estacionar el coche B.

Prueba T para Carros A – Carros –B

One-way ANOVA: PntWear versus Paint

```
Source  DF     SS     MS     F      P
Paint    3  30.69  10.23  2.44  0.115
Error   12  50.25   4.19
Total   15  80.94

S = 2.046   R-Sq = 37.92%   R-Sq(adj) = 22.39%

                             Individual 95% CIs For Mean Based on
                             Pooled StDev
Level   N    Mean  StDev  -------+---------+---------+---------+--
Y-0242  4  14.250  1.893                 (----------*----------)
Y-0314  4  12.250  2.872          (----------*----------)
Y-1424  4  11.000  1.633    (----------*----------)
Y-1723  4  10.750  1.500  (----------*----------)
                          -------+---------+---------+---------+--
                              10.0      12.0      14.0      16.0

Pooled StDev = 2.046
```

Consideraciones finales
Conclusiones prácticas
En promedio, a los conductores les tomo 1.967 segundos mas estacionar el coche A que el coche B. Esta diferencia aunque pequeña es estadísticamente significativa.
¿Es una diferencia de 2-segundos de importancia practica?. Esto lo decides tu.
Los tiempos levemente más largos para estacionarse se asocian a la frustración creciente del conductor, los 2 segundos pueden ser importantes. También, esta diferencia puede ser de mayor importancia a los conductores que seguido se estacionan paralelo.
Consideraciones Estadísticas
Cuando usar una prueba t pareada:

- Las observaciones deben ser pareadas.
- Los datos deben ser continuos.
- Las diferencias deben ser distribuidas normalmente.

Debe ser observado que el procedimiento de la prueba t es bastante robusto para las violaciones de las suposiciones de la normalidad, a condición de que los pares de observaciones se recojan aleatoriamente y los datos sean continuos, unimodal, y razonablemente simétricos (véase a la caja, al cazador, y a Cazador (1978). Estadística para Experimentos, John Wiley & Sons, Inc.).

Utilizando observaciones pareadas eliminas la variabilidad causada por individuos**.** Por ejemplo, al conductor 1 le tomo 18.9 segundos para estacionar el coche y 18.2 segundos para estacionar el coche B. En contraste, al conductor 18 le tomó 43.8 y 41.1 segundos para estacionar los mismos coches. Obviamente, hay mucha variabilidad entre los conductores. Pero analizando las diferencias para cada conductor, tu eliminas esta variabilidad de los cálculos, aumentando el power de tu prueba.

Ejercicio 3.1 Comparaciones de Calibradores
Ejercicio
Tu estás considerando la compra de dos diversos gage para medir válvulas: Calibradores por EasyGage y Too1It. Tu deseas comparar las dos marcas de fábrica del calibrador para determinarse si ofrecen las mismas medidas de promedio.
Utilice un ά-nivel de 0.05 para todas las pruebas.
Recolección de datos

Doce operadores cada uno midieron la misma válvula con los dos diversos calibradores. (El orden en la cual utilizaron el calibrador fue seleccionado aleatoriamente.)

Instrucciones

1.- Use una prueba t pareada para determinar si las medidas de cada calibrador son diferentes.

2.- Con la desviación de estándar de la diferencia de la muestra como estimación de ά , calcule la energía de la prueba al detectar una media de la diferencia de 0.005 cm.. (Indirecta: Conducir una t-test paired es lo mismo que conducir una t-prueba de la una-muestra es la diferencia entre las observaciones pareadas.

Por lo tanto, tu puedes utilizar **Stat > Power and sample size > 1- Sample t** para evaluar el power de la prueba t pareada.

3.- ¿cuál es la energía de la prueba de detectar una diferencia de la media de 0.001 centímetro?

Set de datos
CALIPERS.MPJ

Nombre	Tipo de dato	Tipo de variable
Operator	Numérico	Respuesta
Easy gage	Numérico	Respuesta
Toollt	Numérico	Respuesta
Diff	Numérico	Respuesta

Prueba de una Proporción

Ejemplo 4 Televisiones Reparadas por Tarifa

Ejercicio

Tu quieres determinar si la proporción de tu sistema de televisión de 35- pulgada necesitara ser reparado en el plazo de 4 años de la compra, es diferente que el índice de la industria 6.8% (0.068)

Recolección de datos

Aproximadamente 100,000 encuestas fueron enviadas a los clientes que compraron una televisión 35-plagadas. De los 2,856 clientes que regresaron las encuestas, 236 indicaron que su televisión había requerido la reparación en el plazo de 4 años de la compra.

Herramientas

Stat > Estadísticas Básicas > 1 Proporción

Set de datos

Ninguno

Prueba de una proporción

¿Qué es una prueba de proporción?

Una prueba de una proporción te ayuda determina si una proporción de la población es diferente de un valor específico (proporción de la prueba.)

¿Cuándo utilizar una prueba de una proporción?

Usa una prueba de proporción para evaluar la proporción de los datos de una sola muestra.

¿Porqué usar una prueba de una proporción?

Una prueba de una proporción te puede ayudar a contestar preguntas tales como:

- ¿Es una población diferente de 0.5?
- b) ¿Es una proporción mayor o menor que el criterio?

Por el ejemplo,

- ¿En un programa de inteligencia artificial es posible contestar **Sí / No** preguntas con mayor exactitud del 50%?.
- ¿Está el porcentaje de averías de los sujetadores plásticos debajo del máximo aceptable?.

Conduciendo una prueba de una proporción

Tu estás evaluando los resultados de un examen enviado a los clientes que compraron una de sus televisiones.

La proporción de los que respondieron con televisión que la necesito reparación dentro de los 4 años es 236 / 2856 = 0.0826. El promedio de la industrial es 0.068.

Utiliza una prueba de una proporción para determinar si esta diferencia es significativa.

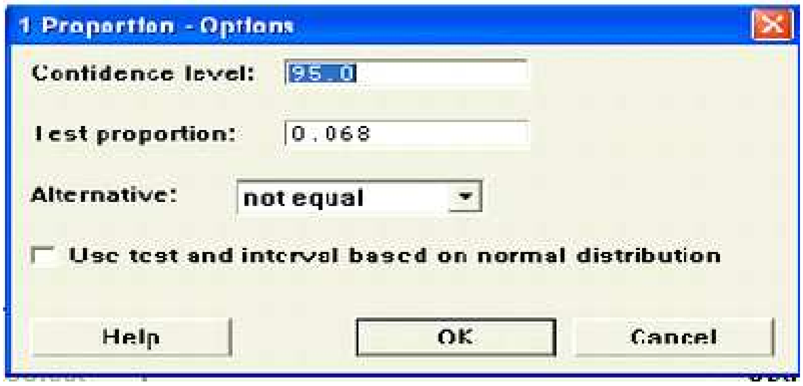
Las hipótesis para la prueba es:

- Ho: la proporción de la población para sus clientes es igual a 0.068.
- H1: la proporción de la población para los clientes no es igual a 0.068.

Utilice un nivel de la confianza del 95%.

1 Proporción:

1.- Elija **Stat** > La Estadística Básica > 1 Proporción.
2.- Seleccione **Summarized data.**
3.- En el Número de ensayos, tipo 2856.
4.- En el **Número de éxitos**, tipo 236.
5.- Click **Opciones.**
6.- Complete el recuadro como se indica a continuación:



7.- Click **OK** en cada recuadro.

Interpretando tus resultados
Utilice ά de 0.05 para la prueba.
Los resultados sugieren que el índice de la reparación para su televisión (muestra p = 0.083) sea más alta que el índice a nivel industrial de 0.068

- El intervalo de confianza del 95% (0.0727992 A 0.093339) no incluye 0.068.
- El p-valor (0.003) es menos que ά (0.005.)

Tu debes rechazar la hipótesis nula, ya que el índice de tu reparación igual que el índice a nivel industrial.
Mas / Para cálculos de intervalos de confianza, vea ayuda de Minitab.

Test y CI para una Proporción

Sample	X	N	Sample p	95% CI	Exact P-Value
1	236	2856	0.082633	(0.072792, 0.093339)	0.003

Consideraciones finales
Conclusiones prácticas
La evidencia sugiere que la proporción de que tu televisión requiera reparación dentro de los de 4 años de la compra es mayor que la proporción del índice a nivel industrial de 0.068.
Por supuesto, la mayoría de los clientes que recibieron el examen no lo devolvió. Es siempre posible que los clientes que han tenido un problema en su televisión son los más probables en devolver el examen. Si ésta es la causa, la proporción real puede ser mucho menos de 0.082633.
Consideraciones estadísticas
Cuantas más observaciones tu tengas, más power tendrá tu prueba de una proporción.
También puedes aumentar el power aumentando **ά.** Sin embargo esto también aumenta la posibilidad de que ocurra el error tipo 1.

REGRESIÓN

Objetivos:

- Mida el grado de la asociación linear entre dos variables usando los gráficos y la estadística.
- Modelo para la relación entre variables de respuestas continuas y unas o más variables de predicción.
- Determine la fuerza de la relación entre variable de respuesta continua y unas o más variables de predicción.

Contenidos

Ejemplos y ejercicios	Propósito	Página
Correlación		69-77
Ejemplo 1 comparaciones al medir los sistemas	Medida del grado de la asociación linear entre dos variables usando correlación	
Regresión simple		77-91
Ejemplo 2 Impurezas en pintura	Evaluación de la relación linear entre dos variables usando Fitted line Plot.	
Ejercicio 2.1 Erosión protectores	Evalúe la relación linear entre dos variables usando la. Fitted line Plot.	
Regresión polinomial		92-108
Ejercicio 3 Caudal de la corriente	Evalúe la relación cuadrática entre dos variables Fitted line Plot.	
Ejercicio 3.1 Extractor del diesel	Evalúe la relación cuadrática entre dos variables Fitted line Plot.	
Regresión Múltiple		109-124
Ejemplo 4 Reducción de golpes del motor	Evalúe la relación linear entre variables múltiples usando la regresión	
Mejor regresión de los subconjuntos		125-133
Ejemplo 5 Reducción de golpes del motor	Seleccione un sistema de variables para incluir en una regresión múltiple mejor subconjuntos	

Correlación
Ejemplo 1 de Sistemas de Medias
Ejercicio
Tu has desarrollado un sistema de medida en línea que usted cree medirán el pH como exactamente el sistema actual en su laboratorio. El sistema en línea proporcionaría una regeneración más rápida y la capacidad de ajustar los sistemas en tiempo real. Tu quieres saber si los dos sistemas producen lecturas similares del pH.
Recolección de datos
La colección de datos ambos sistemas se utiliza para medir el pH de 20 Jornadas aleatoriamente seleccionadas del producto de limpieza.
Herramientas
Graph > Plot
Stat > Basic Statistics > Correlation

Set de Datos
LABSTEST.MPJ

Nombre	Tipo De datos	Tipo de Variable
Laboratorio	Numérica	Respuesta
En línea	Numérica	Respuesta

Correlación
¿Qué es correlación?
La muestra del coeficiente de correlación r, mide el grado de la asociación linear entre dos variables (el grado en la cual una variable cambia con otra).
Una correlación positiva indica que ambas variables tienden a incrementarse juntas. Una correlación negativa indica que una variable se incrementa, y la otra decrece.
¿Cuándo utilizar la correlación?
Utiliza la correlación cuando tengas datos para que dos variables continuas y desees determinen si hay una relación linear entre ellas. La correlación no dirá si estas variables están relacionadas de una manera no lineal.
Algunos estadísticas creen que la correlación no debe ser utilizado si una variable y es dependiente de la respuesta de la otra.
¿Porqué usar la correlación?
La correlación te puede ayudar a contestar preguntas tales como:
- ¿Están dos variables relacionadas en una manera linear?.
- ¿Cuál es fuerza de la relación?.

Por ejemplo,
- ¿Hay una relación entre la temperatura y la viscosidad del aceite de cocina?.
- ¿Es fuerte la relación entre la exposición ultravioleta y la fuerza reducida en el material de nylon de la tienda?.

Dibujando los datos

Creando un diagrama te ayudará a visualizar la relación entre las medidas tomadas por los dos sistemas que estás utilizando para medir el pH.

Graficando las variables

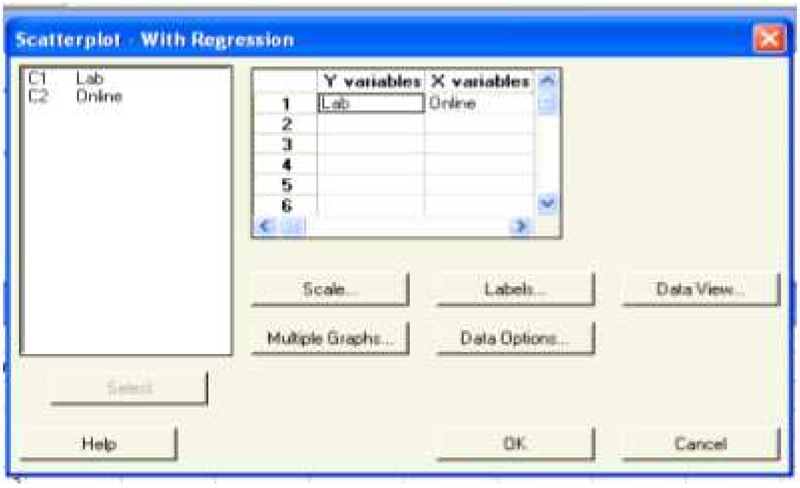
Grafica el laboratorio y el Online de la variable en X y Y respectivamente.

Plot

1.- Abre el proyecto LABTEST.MPJ.

2. Escoge **Graph > Plot**

3.- Completa el recuadro como se indica a continuación:



4.- Click **OK**

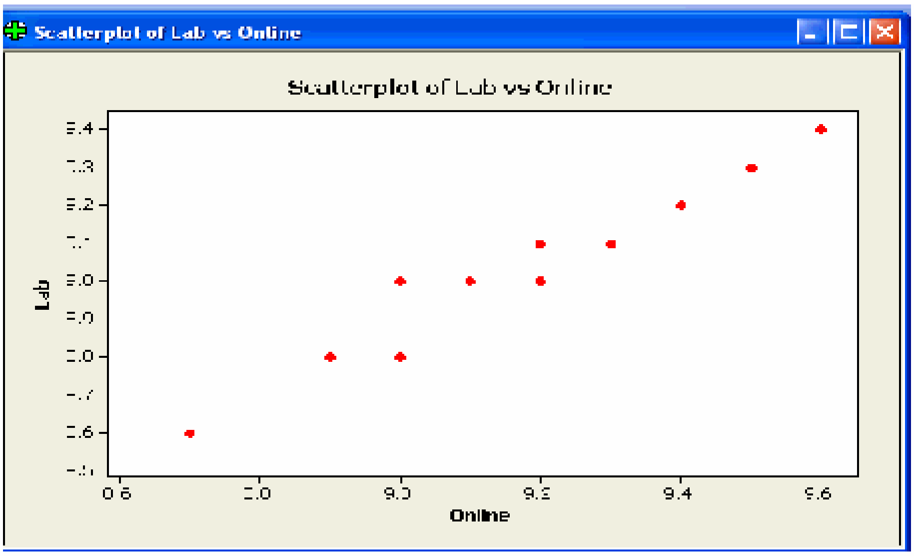
Interpretando tus resultados

El diagrama Online contra medidas del laboratorio indica:

- Hay una relación fuerte entre los dos sistemas que miden. Cuando los valores para el laboratorio cambian, también lo hacen los valores para Online.
- Los datos siguen una línea bastante recta que sugiere que la relación es linear.
- Los altos valores del sistema en línea se asocian a altos valores del sistema del laboratorio, indicando que la relación es positiva.

¿Que se hace después?

Porque la relación es linear, usted puede calcular la correlación para cuantificar la fuerza de la asociación.



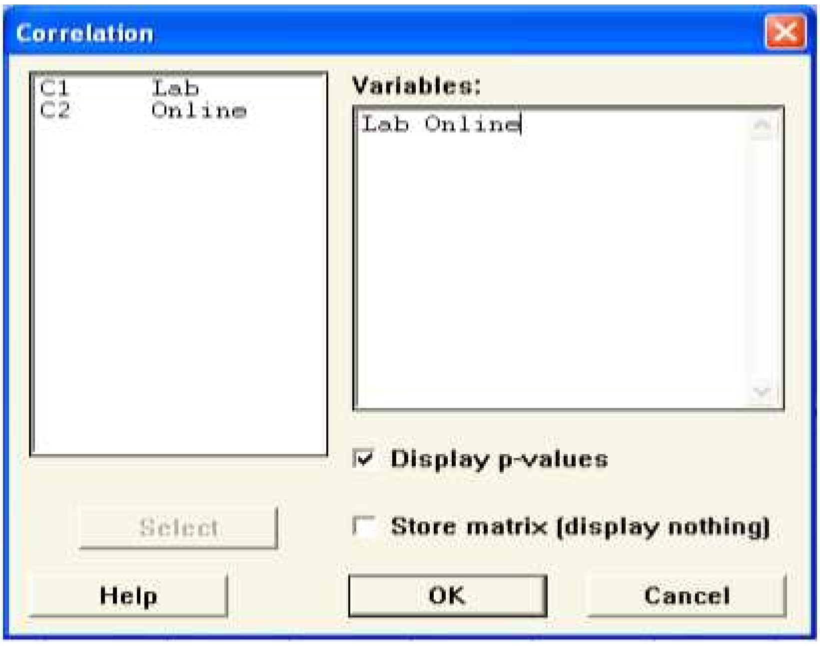
Calculando la correlación

Tu deseas calcular el coeficiente de correlación de Pearson para determinar la fuerza de la asociación linear entre las medidas en Online y del laboratorio.

Correlación

1.- Escoge **Stat > Basic Statistics > Correlacion.**

2.- Enter en Variable Laboratorio.



3.- Click **OK.**

Interpretando tus resultados
Correlación: Laboratorio En línea.
Prueba de la Correlación Laboratorio En línea = 0.959
P – Valor = 0.000

Use una ά 0.05 para el texto.
Pearson correlación
El coeficiente de la correlación de la muestra (r) es calculado por la fórmula:

$$R = \frac{\sum (X_i - X)(\ddot{Y}_i - \ddot{Y})}{\sqrt{\sum (X_i - X)^2 \sum (\ddot{Y}_i - \ddot{Y})^2}}$$

El valor de r estará siempre entre -1 y 1:
- 1 indica una correlación positiva perfecta.
- 0 indica ninguna correlación.
- -1 indica una correlación negativo perfecto.

P-valor
La prueba del p-valor las hipótesis siguientes:
Ho: El coeficiente de correlación (p o rho) para la relación entre las poblaciones es igual a cero.
H1: p no es igual a cero.

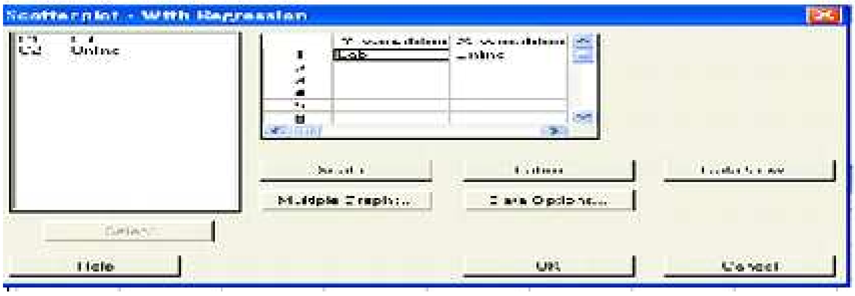
Conclusión:
El coeficiente de correlación (0.959) indica que ahí una fuerte asociación lineal positiva entre la media del Laboratorio y la del Online. Además el p-valor (0.000) es menos que & (0.05), entonces tus puede rechazar la hipótesis nula, ya que no existe ninguna asociación lineal.
¿Que se hace después?
Antes de sustituir el sistema de laboratorio con el sistema en línea, tu necesitas evaluar dos aspectos adicionales entre la relación de los dos. Incluso si la correlación era perfecto (r=1), todavía podrían haber diferencias importantes entre los sistemas:
- Las medidas de un sistema podrían ser coherentemente más altas que las medidas del otro.

El coeficiente de correlación no dirige estas cuestiones de tendencia y sensibilidad.

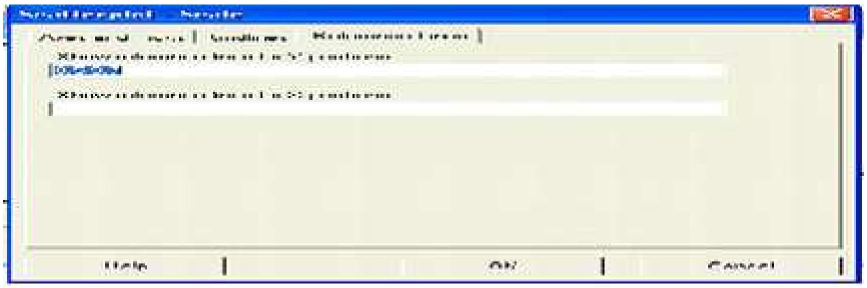
Anotación de la gráfica
Utilice el diagrama del argumento para ayudarte a evaluar si la medida de los dos sistemas es similar, y si los sistemas son igualmente sensibles:
- Traza con los datos los mismos valores del mínimo y del máximo para ambas X.
- Y Agregue una línea para indicar donde **X**=**Y.** (usted podría agregar la línea después de que el gráfico sea creado usando las herramientas para graficas de MINITAB`s. Sin embargo, la caja del subdiálogo de la anotación proporciona una manera más exacta de agregar líneas al gráfico).

Plot
1.-Elige **Graph > Plot**.
2.- Del c**apítulo**, elige el **minuto y el máximo.**
3.-Elige **el mismo mínimo y máximo para las X de X y de Y.**
4.- Click **OK**

5.- Para **anotación**, elija la **línea.**



6. – Completa el recuadro como se indica a continuación:



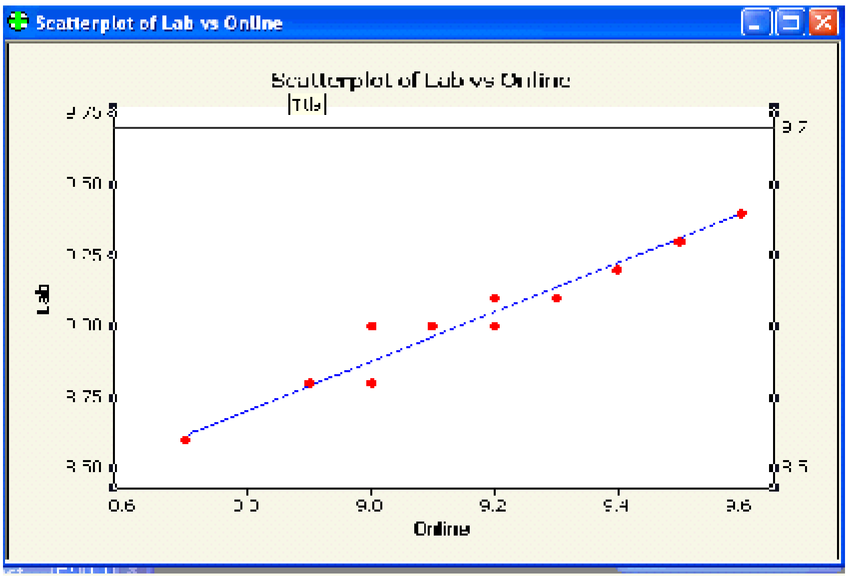
7.- Clic **OK** en cada recuadro.

Interpretando tus resultados

Para cada punto referente en la línea, **X** es igual a **Y**. Si ambos sistemas van encima con las mismas medidas para cada muestra, entonces todos los puntos de referencias caerán en esta línea.

Comparando los datos en la línea de referencia revela lo siguiente:

- Todos, menos un punto está debajo de la línea, indicando que el sistema en línea produce medidas constantemente más altas que el sistema del laboratorio.
- La línea que los datos siguen tiene básicamente la misma cuesta que la línea de referencia. Esto indica que los valores de los rangos indican que los dos sistemas son similares.



Consideraciones finales

Conclusiones prácticas

Hay una fuerte correlación positiva del (0.959) entre las medidas tomadas con el laboratorio y con los sistemas en línea.

Sin embargo el sistema en línea rinde medidas constantemente más altas que lo hace el sistema del laboratorio. Esto puede indicar la necesidad de recalibración.

Los resultados de los límites en el experimento indican que es menos costoso y más fácil utilizar el sistema en línea y puede ser un reemplazo conveniente para el sistema de medida del laboratorio.

Consideraciones estadísticas

La Correlación cuantifica el grado de la asociación linear entre dos variables.

Una correlación fuerte no implica una relación de causa y efecto. Para el ejemplo, una correlación fuerte entre dos variables puede ser debido a la influencia de una tercera variable, no bajo consideración.

Un coeficiente del correlación cerca de cero no significa necesariamente ninguna asociación, sólo que esa asociación no es linear. Tu debes trazar siempre sus datos de modo que puedas identificar relaciones lineares cuando estás se presenten.

Algunos estadísticos discuten que la correlación que sea utilizada si una variable es una respuesta dependiente de la otra.

La correlación asume que los valores de ambas variables están libres de variar. Tu no puedes utilizar la correlación si fijas los valores de una variables una para estudiar cambios en otra.

Regresión Simple
Ejercicio
Tu sospechas que la revoltura tiene un impacto en el nivel de impurezas en tu producto de pintura.
Recolección de datos
Las impurezas fueron medidas para lotes de pintura revueltas en rangos de movimientos a partir de 20 a 42 RPM (revoluciones por minuto.)
Herramientas
Stat > Regression > Fitted Line Plot.
Stat > Regression > Fitted Line Plots.

Set de Datos
PAINT.MPJ

Nombre	Tipo de Datos	Tipo de Variable
Stirrate	Numérica	Predictor
Impureza	Numérica	Respuesta

REGRESIÓN SIMPLE
¿Qué es la regresión simple?
La regresión simple examina la relación entre una variable de respuesta continua (Y) y una variable de predicción (X). La ecuación general para un modelo de regresión simple es:

$Y = Y = \beta_o + \beta_1 X + \acute{\epsilon}$

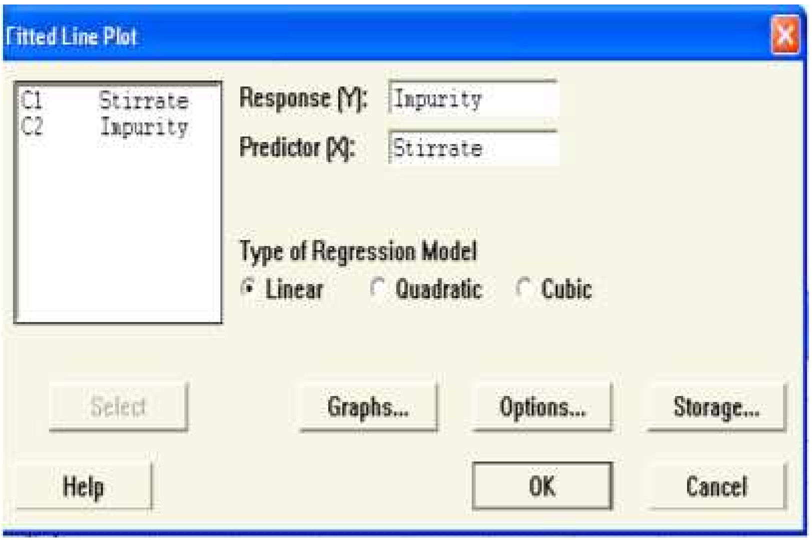
Donde Y es la respuesta, X es la predicción, β_o la es el interceptor (el valor de Y cuando X iguala el cero), β_1 la es la cuesta, y $\acute{\epsilon}$ es el error aleatorio.
¿Cuándo usar la regresión simple?
Usa la regresión simple cuando tu tengas Y continua y solo una X. Las siguientes condiciones deben ser encontradas:
- X puede ser ordinal, o continúa.
- En la teoría, X debería ser fijada. En la práctica, sin embargo, a menudo le permiten para variar.
- Cualquier variación arbitraria en la medida de X es asumida para ser insignificante comparada con el rango en cual X es medido.

Los valores de Y obtenidos en su muestra se diferenciarán de estas predicciones por el modelo de regresión (a no ser que todos los puntos resulten caer sobre la línea perfectamente recta.). Llaman residual a estas diferencias.
Antes de la aceptación de los resultados de un análisis de regresión, tu debes verificar que las suposiciones siguientes sobre los residuales son válidas para tus datos:
- Ellos son independientes (y así arbitrarios).
- Ellos están distribuidos normalmente.
- Ellos tienen constantes variaciones a través de todos los valores de X.

¿Por qué usar la regresión simple?
La regresión simple te puede ayudar a contestar preguntas tales como:
- ¿Cómo importante es X en la predicción Y?
- ¿Qué valor puedes tu esperar para Y cuándo X es 20?
- ¿Cuánto es que cambio de Y si X en una unidad?

Por ejemplo,
- ¿Cómo el proceso de la temperatura de tratamiento se relaciona con la dureza de su acero?
- ¿Que fuerza tendrá su acero si usted lo trata a una temperatura particular?
- ¿Cuánto más difícil tratar será su acero si aumentas la temperatura en 100? º?

Ajustando el modelo lineal
Tu quieres determinar el efecto de tarifa de movimiento sobre la cantidad de impurezas en la pintura. Utiliza **Fitted Line Plot** para calcular y graficar la ecuación de la regresión
Fitted Line Plot
1.- Abre el Project PAINT.MPJ.
2.- Escoge **Stat > Regression > Fitted Plot.**
3.- Completa el recuadro como se indica a continuación:



4.- ClicK **OK**

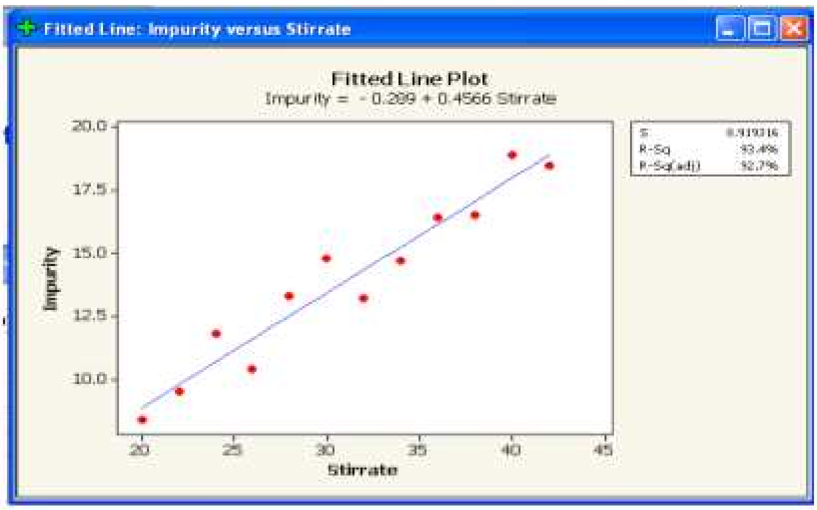
Interpretando tus resultados
Regresión la ecuación
La ecuación de Regresión relaciona la predicción (stirrate) con la respuesta (la impureza):

Impureza =-0.289277 + 0.456643 stirrate

La inclinación de la línea de regresión, 0.456643, indica cuanto un cambio en la impureza es asociado con cada cambio de una unidad de stirrate.
S
S es una estimación del promedio de variabilidad media sobre la línea de regresión. La S es la raíz cuadrada positiva de MSE. La mejor ecuación predice la respuesta, mas bajo S será.

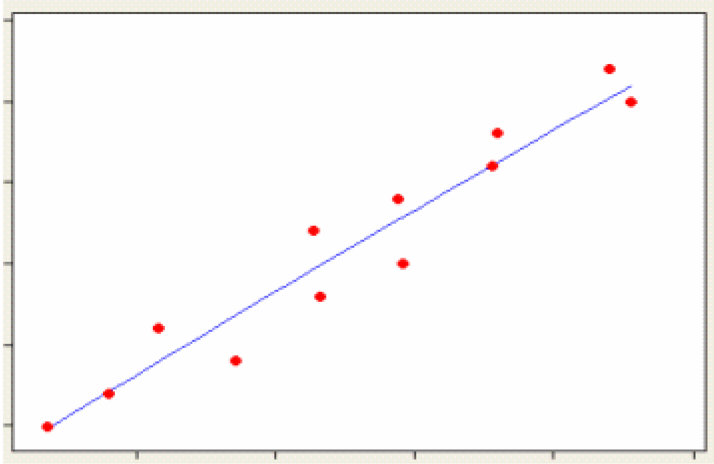
R^2 (R-Sq)

La **R^2** (Cuadrada de r) es la proporción de la variabilidad en la respuesta que es explicada por la ecuación. Así, el 93.4 % de la variación en la impureza puede ser explicado por su relación lineal con el Stirrate.
Valores aceptables para **R^2** varían dependiendo del estudio. Por ejemplo, los Ingenieros que estudian reacciones químicas pueden requerir una **R^2** del 90 % o más. Sin embargo, alguien estudiando el comportamiento humano (que es más variable) puede estar satisfecho con valores de **R^2** inferiores.
Ajustada (cuadrado de r (adj))
R^2 ajustada es sensible al número de términos (condiciones) en el modelo y es importante comparar los modelos con diferentes números de términos (Ver 3-58).



La menor línea de regresión cuadrada
Los coeficientes para la ecuación de regresión son escogidos para reducir al mínimo la suma de las diferencias cuadriculadas entre los valores de respuesta observados en la muestra y aquellos predichos por la ecuación.
En otras palabras, las distancias verticales entre los puntos y la línea son reducidas al mínimo, ilustrado a la derecha. El resultado es llamado: La menor parte de la línea de regresión cuadrada.
Esté atento a esta líneas de fuera usando los procedimientos de la regresión. Algún líneas de fuera (llamados altos puntos) tienen un efecto grande sobre el cálculo de la menor parte de línea de regresión de cuadrado. En tales casos, la línea más puede puede

representar al resto de datos muy bien.
Note: este gráfico ha sido corregido para la ilustración.



Interpretando tus resultados
Use el análisis de varianza (ANOVA) resultados para evaluar si su modelo de regresión simple es útil. El ANOVA compara su modelo a un modelo restringido que no usa Stirrate (**X**) para predecir la impureza (**Y**):

- Modelo de Regresión: Y = $\beta_o + \beta_1 X + \acute{\varepsilon}$
- Modelo Restringido: Y = Y = $\beta_o + \acute{\varepsilon}$

El modelo restringido declara que los cambios de Y están previstos únicamente al error arbitrario ($\acute{\varepsilon}$). Es equivalente a un modelo de regresión simple con una cuesta (β_1) de cero. Así, las hipótesis para el ANOVA son,

- H_o: β_1 es igual para cero.
- H_1: β_1 no es igual a cero.

Interprete el p-valor (P) así:

- Si el p-valor es menos que o igual a ά , deseche H_o. El modelo de regresión explica considerablemente más variabilidad en la respuesta que hace el modelo restringido. La β_1 no iguala el cero.
- Si el p-valor es a mayor, usted no puede rechazar la H_o. β_1 no es considerablemente diferente del cero.

Conclusión
Usando un ά 0.05, tu puedes rechazar el modelo simple restringido y afirmar que Stirrate realmente tiene un efecto significativo lineal sobre la Impureza.

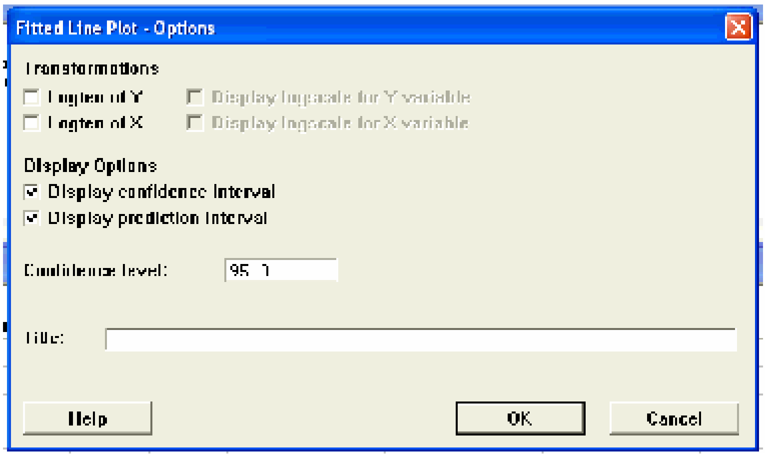
Análisis de Regresión: Impureza contra Stirrate

```
———————— 28/04/2006 18:13:04 ————————————————

Welcome to Minitab, press F1 for help.
Retrieving project from file: 'A:\PAINT.MPJ'

Results for: Paint

Regression Analysis: Stirrate versus Impurity

The regression equation is
logten(Stirrate) = 0.4679 + 0.8965 logten(Impurity)


S = 0.0287770   R-Sq = 93.1%   R-Sq(adj) = 92.5%


Analysis of Variance

Source       DF        SS        MS       F      P
Regression    1  0.112430  0.112430  135.77  0.000
Error        10  0.008281  0.000828
Total        11  0.120711


Fitted Line: Stirrate versus Impurity
```

Adicionando confianza y Predicción
Confidencialidad y bandas de predicción.
Tu también quieres saber si confiar que la media y los puntos individuales en la variable Y, Impurezas caen dentro de ciertos límites de variabilidad.
Residuales y Fits
El Residual es la diferencias entre los valores ajustados de su modelo y los valores observados. Son las estimaciones de punto de la respuesta estimados para cada nivel de la variable independiente. Tu debes almacenar estos valores para usar más tarde.

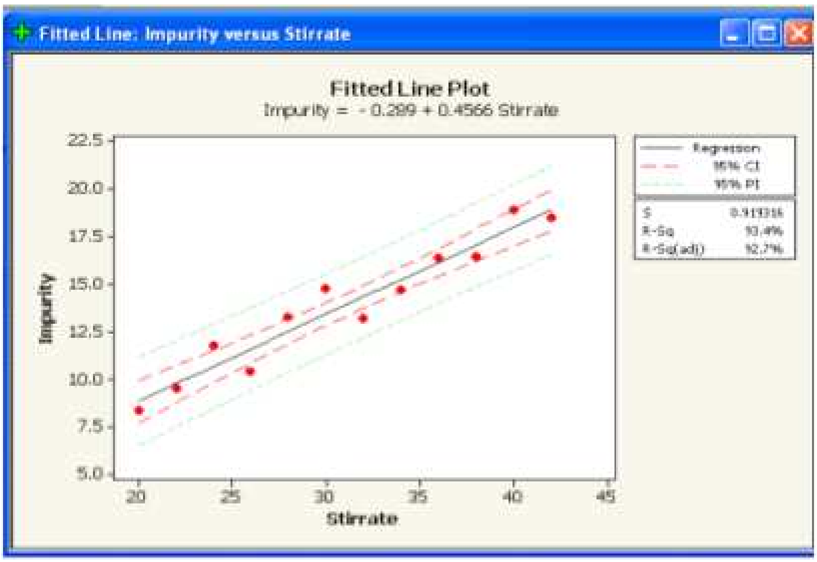
Use un nivel de confianza de falta del 95 %.

Fitted Line Plot
1.- Escoja **Stat** > **La Regresión** > **Fitted Line Plot** presione Ctrl + E para volver al dialogo **Fitted Line Plot** al cuadro.
2.- . **Click** Opciones.
3.- Completa el recuadro de diálogo como se indica a continuación:



4.- Click **OK.**
5. - Click **Storage.**
6.- Elige **Residuals y Fits.**
7- Click **OK** en cada recuadro.

Interpretando tus resultados
Intervalo de confianza
El 95% de confianza define un excelente rango de valores en la población de la media de Y. Para cualquier valor dado de **X,** tu pues confiar que la población de la media de Y esta entre las líneas indicadas.
Intervalo de predicción
El intervalo de predicción del 95 % define una gama probable de valores de **Y** para observaciones individuales. Para cualquier valor dado de **X**, tu puedes confiar con un 95 % que el valor correspondiente de **Y** para una observación estará entre las líneas indicadas.
Regresión Plot

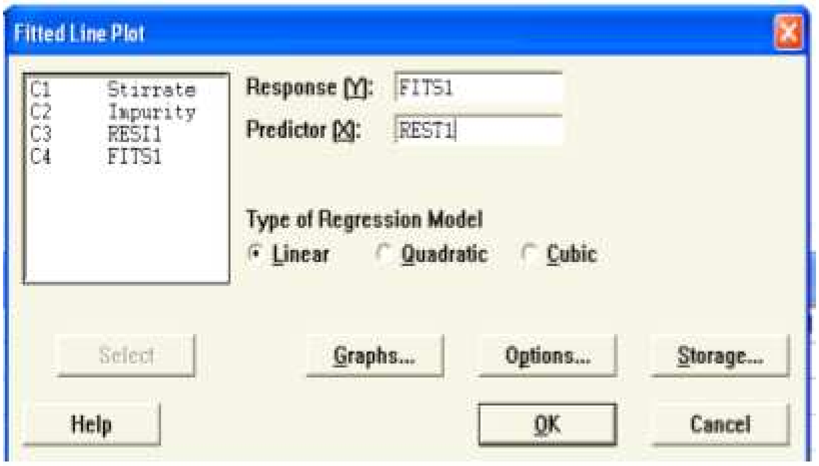


Creando una gráfica del Residual
El residual para cada observación es la diferencia entre el valor observado de la respuesta y el valor predictivo por el modelo (el valor ajustado). Por ejemplo, si el valor de respuesta observado es 12 y el modelo predice 10, el residual es 2.
Suposiciones
Para confirmar que tu análisis de regresión es válido, tu debes verificar todas las suposiciones sobre los residuales. Usa las graficas de la residual para comprobar que los residuales:

- Sean aleatorios (independientes el uno del otro)
- Estén normalmente distribuidos.
- Tienen la misma discrepancia a través de todos los valores de **X**

Nota : Si tienes más que una columna de residuales y Fits sobre su hoja de trabajo, se cuidadoso al seleccionar las columnas correctas cuando crees las graficas de la residuales.

Residual Plot
1. Escoja **Stat** > **Regresión** > **Residuales Plots0**
2. Completar el recuadro como se indica a continuación:

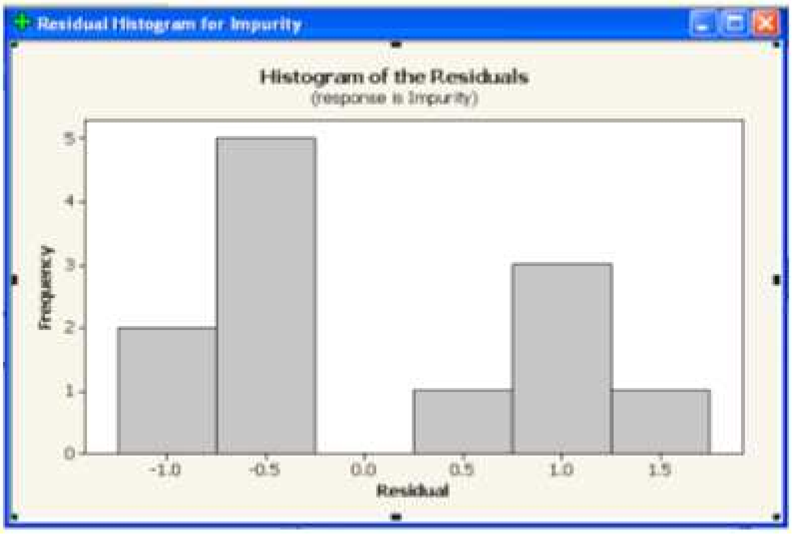


3. –Click **OK**

Interpretando tus residuales
La gráfica de probabilidad de Normalidad
Usa la grafica de la probabilidad normal de la residual para verificar que el residual no este desviada sustancialmente de la distribución normal.

- Si los residuales viene de una distribución normal, los puntos aproximadamente seguirá una línea directa.
- Si los residuales no vienen de una distribución normal, los puntos no seguirá una línea directa.

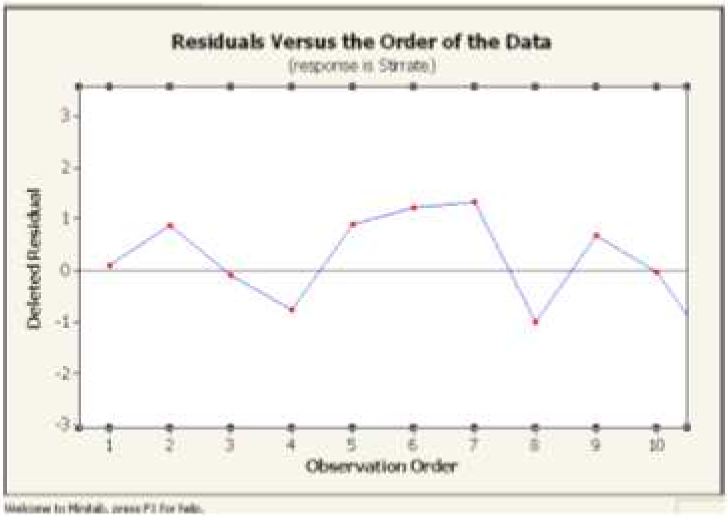
Basado en este grafico, es razonable asumir que los residuales para sus datos no se desvía considerablemente de una distribución normal. Una prueba de normalidad para estos datos (no mostrado dio un p-valor de 0.252.)
Histograma
Tu puedes usar el histograma de las residuales para evaluar la normalidad. Sin embargo, la grafica de probabilidad normal es generalmente más fácil para hacer de interpretar, sobre todo para pequeñas muestras.



Interpretación sus resultados
Gráfica
La gráfica presenta los residuales en el orden de la recolección de los datos (proporcionando los datos que fueron entrados en el misma orden en la cual ellos fueron recogidos). Use la grafica para verificar que el residual es independiente.

- Si hay un efecto debido a la orden de recolección de datos, el residual no estará disperso aleatoriamente sobre el cero. Tu debes ser capaz de detectar este patrón en la gráfica.
- Si no hay ningún efecto debido al orden de recolección de datos, los residuales estará disperso aleatoriamente sobre el cero

No aparece haber en ningún momento efecto de orden en el set de datos presentes.



Interpretando tus resultados

Residuales Versus Fits

Use el grafico de los residuales versus ajustes para verificar que:

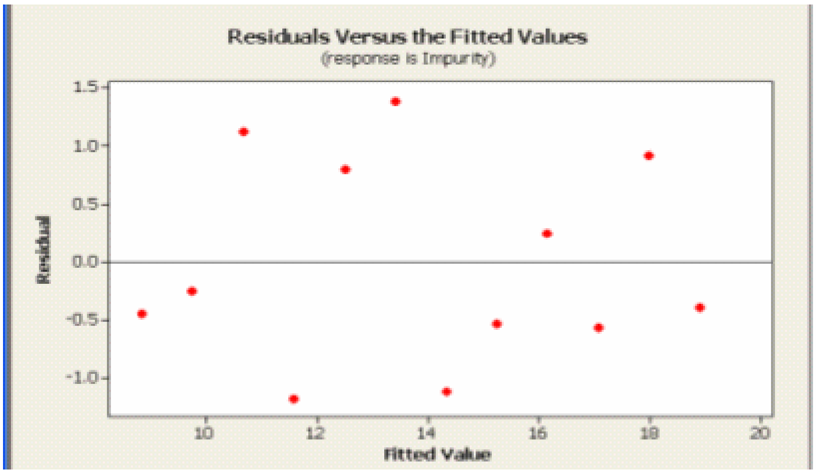
- El modelo no omite ningún termino cuadrático.
- La variación es constante es constante a través de todos los valores ajustados.
- No hay valores fuera de línea en tus datos.

Si tu puedes ver cualquier tipo de patrón en esta grafica, una de estas suposiciones ha sido violada.

La tabla debajo resume el patrón típico que puedes ver.

El patrón	**Indica**
Curvilíneo	Un término cuadrático puede fallar en tu modelo
La extensión desigual de las residuales a través de los diferentes valores ajustados.	La variación de las residuales no es constante.
Un punto está situado muy lejos del cero.	Esta fuera de línea.

La grafica de los datos no parece revelar ningún patrón.



Consideraciones finales

Conclusiones prácticas

El análisis simple de la regresión linear reveló que el aumento de los ritmos de revoltura está asociados a los niveles crecientes de impurezas en su pintura

La pendiente de la ecuación de la regresión indica que cuando tu aumentas el ritmo del revolvimiento en 1 rpm, el nivel de impurezas aumenta en 0.456643.

Tu puedes usar la ecuación para determinar qué las impurezas serán diferentes para las mezclas de pintura. Sin embargo, la ecuación es solamente válida para la gama de datos que usted ha hecho un muestreo (revolviendo entre 20 y 42).

Consideraciones estadísticas

Tu no puedes utilizar el análisis de la regresión para afirmar que los cambios en la predicción fueron causados por la respuesta, a menos que los valores del predictor fueran fijos en los niveles predeterminados en un experimento controlado. Si los valores de los apalancamiento) tienen un efecto grande en el cálculo de la línea menor de la regresión de los cuadrados. En tales casos, la línea puede no representar el resto de los datos muy bien.

Ejercicio 2.1
Protectores de Erosión
Ejercicio
Tu estas intentando predecir cómo los protectores de acero de la erosión para las turbinas de vapor resisten la pérdida de la abrasión.
La resistencia directamente que mide a la abrasión es difícil, costosa, y destructiva. Por lo tanto, tu esperas poder predecir la resistencia a la abrasión usando la dureza de acero, que es más conveniente y menos costosa medir.
Recolección de datos:
La pérdida y la dureza de la abrasión de la recolección de datos fueron medidas para 24 protectores aleatoriamente seleccionados de la erosión.
Instrucciones
1. Utilice la Fitted Line Plot para ajustar el modelo simple de la regresión linear con la abrasión como la respuesta y la dureza como el predictor: Incluya la confianza y la predicción en sus resultados, y asegúrate de almacenar las residuales y los ajustes.
2. Utiliza los diagramas residuales de validar las suposiciones necesarias.
Set de Datos
EROSION.MPJ

Nombre	Tipo de datos	Tipo de Variable
Stirrate	Numérico	Predictor
Impureza	Numérico	Respuesta

REGRESIÓN POLINOMIAL
Ejemplo 3 del caudal de la corriente.
Ejercicio
Tu estás conduciendo un estudio de los impactos para el medio ambiente y quieres utilizar la profundidad de una corriente para estimar el caudal.
Recolección de datos:
La profundidad y el flujo fueron registrados para una sola corriente en un periodo de 6 meses.
Herramientas
Graph > plot.
Stat > Regression > Fitted Line Plot.
Stat > Regresión > Argumentos Residuales.
Set de Datos:
FLOW.MPJ

Nombre	Tipo de datos	Tipo de Variable
Stirrate	Numérico	Respuesta
Impureza	Numérico	Predictor

Regresión Polinomial
¿Que es Regresión polinomial?
Como la regresión lineal, La regresión polinomial examina la relación entre una variable continua de la respuesta (**Y** y una variable del predictor (**X**). Es diferente de la regresión simple, sin embargo, un modelo polinomial puede incluir los términos para los exponentes de **X**:

Ecuación	Tipo de Modelo
$Y = \beta_o + \beta_1 X + \acute{\varepsilon}$	Linear
$Y = \beta_o + \beta_1 X + \beta_2 X^2 + \acute{\varepsilon}$	Polinomio cuadrático
$Y = \beta_o + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \acute{\varepsilon}$	Polinomio cúbico

Donde **Y** es la respuesta, **X** es el predictor, β_o es el coeficiente para el término linear, β_1 es el coeficiente para el término ajustado, β_2 es el coeficiente para el término cuadriculado, β_3 es el coeficiente para el término cubicado, y $\acute{\varepsilon}$ es error al azar.

¿Cuándo utilizar la Regresión Polinomial?
Usa la Regresión Polinomial cuando tengas tiene una **Y** continua y un solo **X**, y evidencia o teoría sugiriendo no-linealidad.

- **X** Puede ser ordinal o continuo o en la teoría.

- En teoría **X** debe ser fijo. En la práctica, sin embargo, se permite a menudo variar.
- Cualquier variación aleatoria en la medida de **X** se asume como insignificante comparado en el rango en donde es medida X.

Después de aceptar los resultados en el análisis de la regresión simple, tu debes verificar que las siguientes suposiciones acerca del residual sean validas en tus datos.

- Deben ser independientes (y así al azar).
- Deben ser distribuidos normalmente.
- Deben tener variación constante a través de todos los valores de **X**.

¿Porqué usar la Regresión Polinomial?

La Regresión Polinomial te puede ayudar a responder preguntas tales como:

- ¿Incrementando **X** incrementa **Y** para algunos valores del rango y disminuye para otras?
- ¿Qué valor puedes tu esperar para **Y** cuando **X** es 20?

Por ejemplo,

- ¿Agregando más cobre a su aleación siempre es mas fuerte o la fuerza disminuye en concentraciones más altas?
- Como puedes esperar que tu aleación sea de 0.01% de cobre.

Dibujando los datos

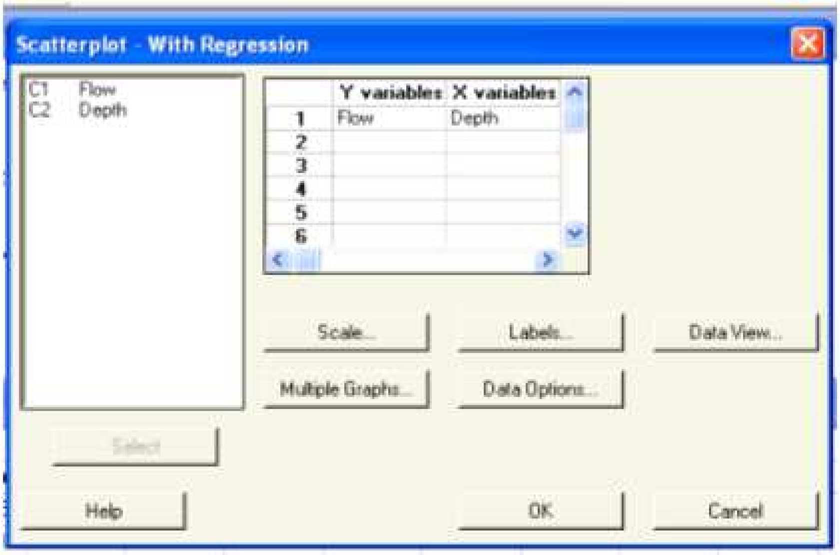
Para visualizar la relación entre la profundidad de la corriente y el caudal, utilice el diagrama para crear un scatterplot con la respuesta (flujo) en el y-axis y el predictor (profundidad) en el x-axis.

Plot

1.- Abre el Project FLOW.MPJ.

2.- Elija El **Graph > Plot**

3.- Completa el recuadro como se indica a continuación:

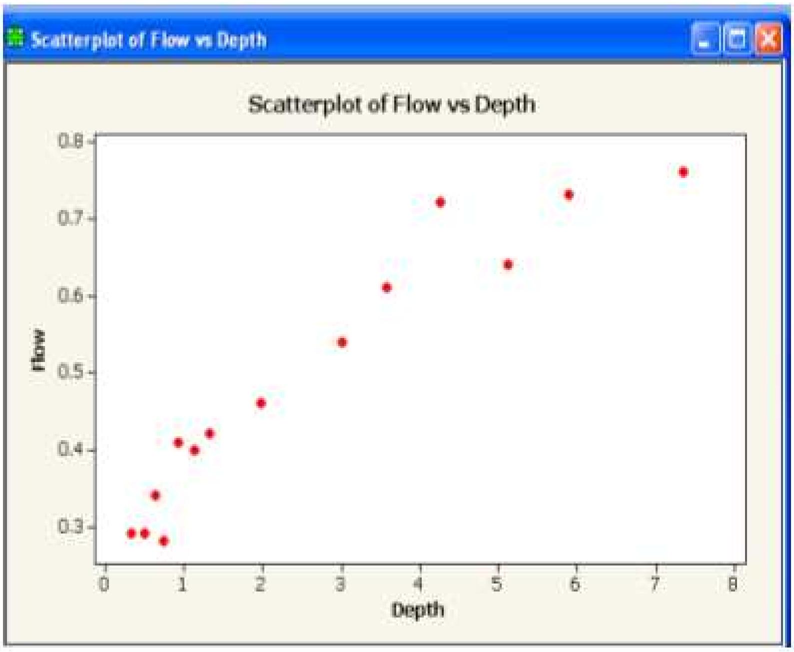


4.- Click **OK**

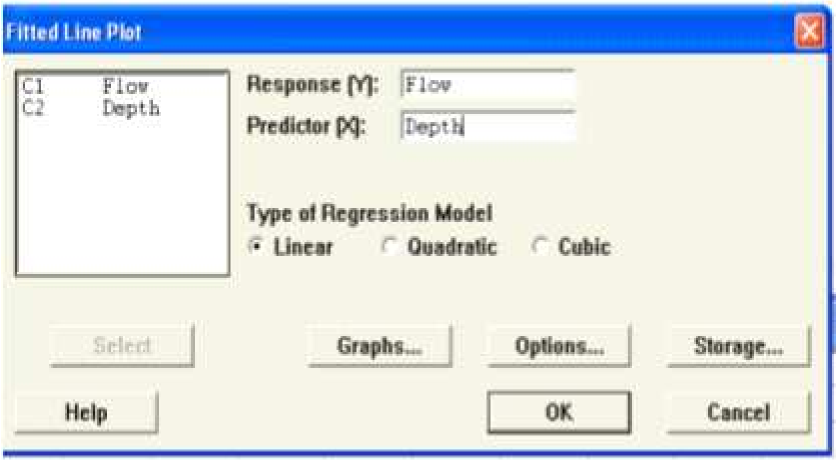
Interpretando tus resultados

La gráfica revela una relación potencialmente no lineal entre la profundidad y el flujo.

Por ejemplo, Note que un aumento 1.5- pies en profundidad a partir de la 0.5 a 2 pies parece aumentar dos veces el flujo tanto como un aumento en profundidad a partir del 6.0 a 7.5 pies.



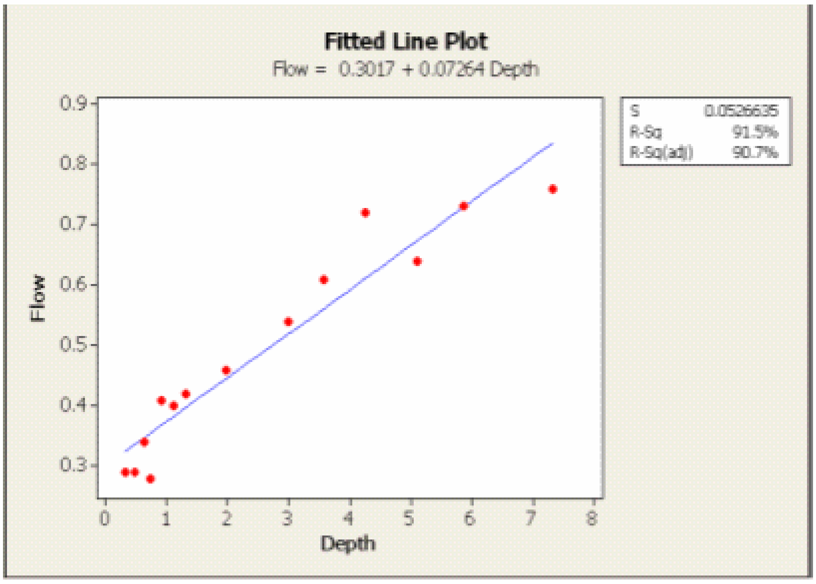
Ajustando el modelo linear
Usa la grafica del modelo linear, para evaluar que tan bien   esta el modelo de regresión linear en los datos.
Fitted Line Plot
1.- Elige **Stat >** Regression **> Fitted Line Plot.**
2.- Completa el recuadro como se indica a continuación:



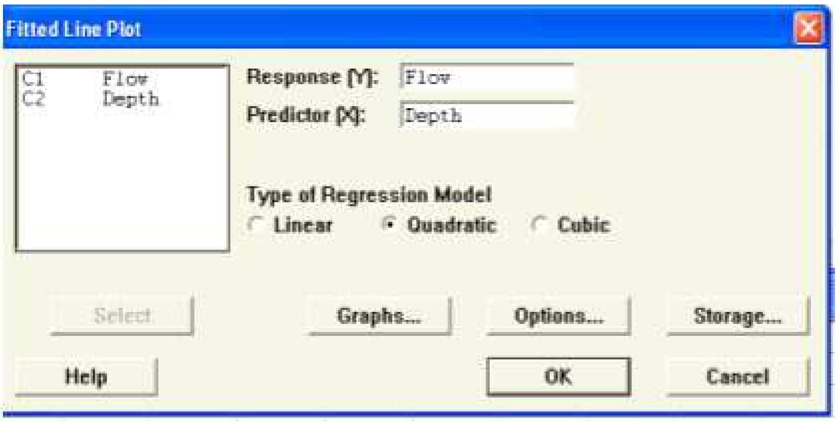
3.- Click **OK**

Interpretando tus resultados
La ecuación linear que mejor describe los datos es:
Flujo = 0.301672 + 0.0726395 Profundidad
R^2 (R-r-Sq)
El R^2 para el modelo linear indica que 91.5% de la variabilidad en flujo es explicado por profundidad de la corriente.
¿Que sigue?
Mientras que un porcentaje de la variabilidad es explicado por el modelo linear, parece una línea levemente curvada cabría incluso mejor. Tu debes evaluar cómo en modelo cuadrático caben estos datos.

Regresión Plot



Ajustando el Modelo Cuadrático
Usa el Fitted Line Plot para cuadrar tu modelo de regresión cuadrático. Almacene los ajustes y las residuales para una reexaminación más futura.
Fitted Line Plot
1.- Elige el **Stat > Regresión > Fitted Line Plot** o presiona Ctrl+E para volver a **Fitted Line Plot** del recuadro.
2.- Completa el recuadro como se indica a continuación:



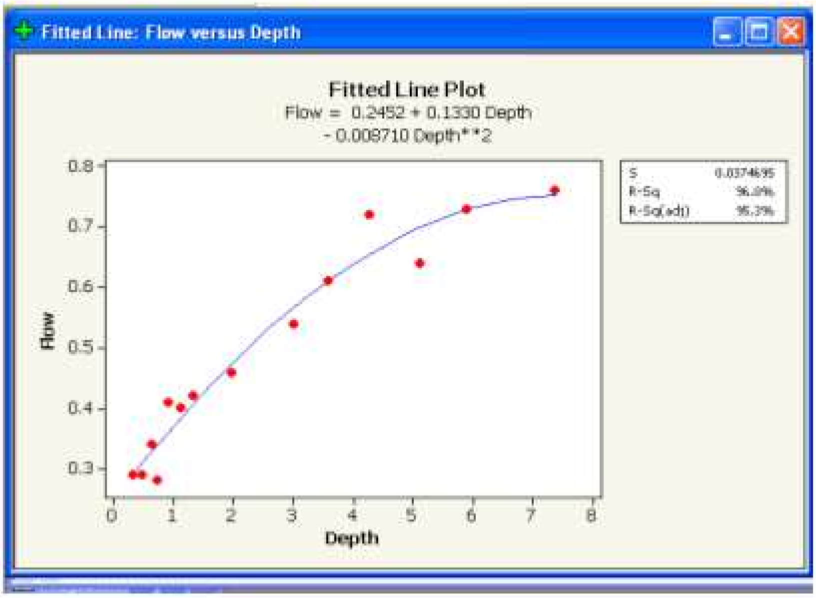
3.- Clic **Storage.**
4.- Compruebe las **Residuals** y los **Fits**.
5.- Click **OK** en cada recuadro.

Interpretando tus resultados
Ecuación de Regresión
La Regresión cuadrática que mejor describe los datos es:
Flujo = 0.245230 + 0.133027 Profundidad
- 0.0087100 Depth**2

R^2 (R-r-Sq) y R^2- adjuntos (R - Sq(adj))

R^2 indica que el modelo cuadrático considera 96.0% de la variabilidad en el caudal. Éste es algo más que el R^2 de 91.5% obtenidos con el modelo linear (véase 3-32).
La estadística ajustada de R^2 estadística ajustada es ajusta según el número de términos en el modelo, y debe ser utilizada al comparar modelos con diversos números de predictores.
El R^2 ajustado para el modelo cuadrático (95.3%) es mayor que el R^2 ajustado para el modelo linear (90.7%), indicando que el término adicional mejora la predicción.

Regresión Plot



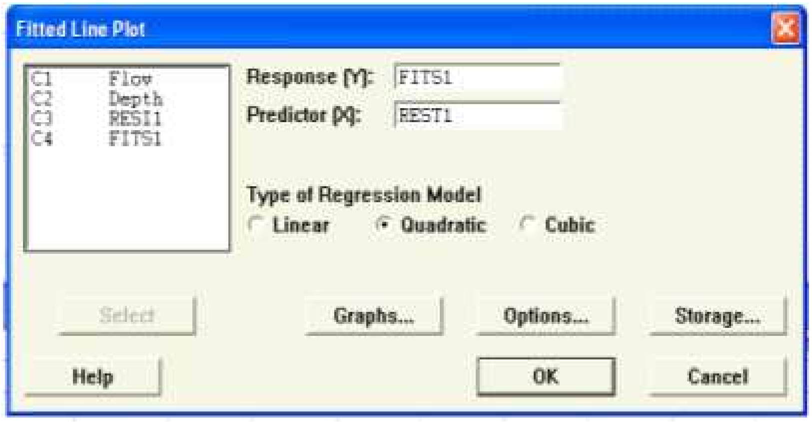
Interpretando tus resultados
Utilizan una ά 0.05 para todas las pruebas.
Análisis de Varianza
El p-valor para el modelo en su totalidad (0.000) es significativo, indicando que el modelo es útil.
El p-valor para el término linear (0.000) es también significativo, indicando que explica una cantidad significativa de variabilidad. Pasado, el p-valor para el término cuadrático (0.004) es significativo, indicando eso que agrega este término al modelo linear mejora la predicción perceptiblemente.

Análisis Polinómial De la Regresión:

Polynomial Regression Analysis: Flow versus Depth

```
The regression equation is
Flow = 0.2452 + 0.1330 Depth - 0.008710 Depth**2


S = 0.0374695   R-Sq = 96.0%   R-Sq(adj) = 95.3%


Analysis of Variance

Source      DF        SS        MS       F      P
Regression   2  0.374192  0.187096  133.26  0.000
Error       11  0.015444  0.001404
Total       13  0.389636


Sequential Analysis of Variance

Source     DF        SS       F      P
Linear      1  0.356354  128.49  0.000
Quadratic   1  0.017838   12.71  0.004
```

Graficando los residuales
Utilizan las residuales y los ajustes correr graficas de diagnóstico en el modelo cuadrático. Tu estás utilizando diagramas residuales para verificar que las suposiciones sobre el término del error en el modelo de la regresión han sido encontradas.
Residual Plots
1.- Elige **Stat > Regresión > Residual Plots.**
2.- Completa el recuadro como se indica a continuación:



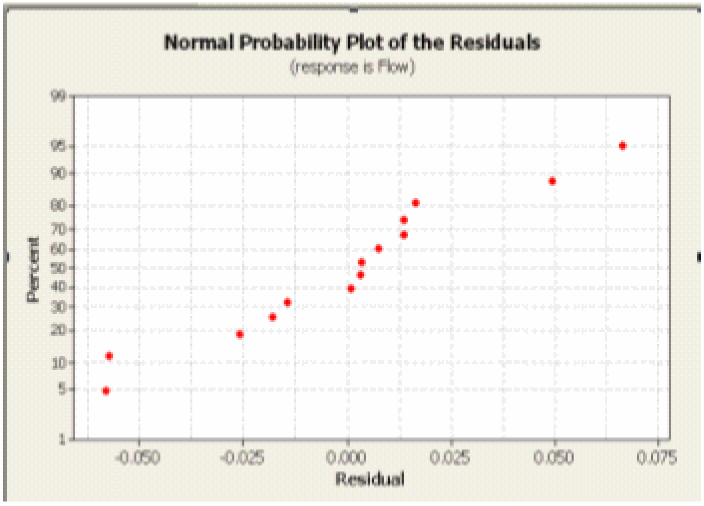
3.- Click **OK**

Interpretando tus resultados
Diagramas de Probabilidad Normal
Usa los diagramas de probabilidad normal de los residuos para verificar que tus residuos no se desvían substancialmente de una distribución normal.

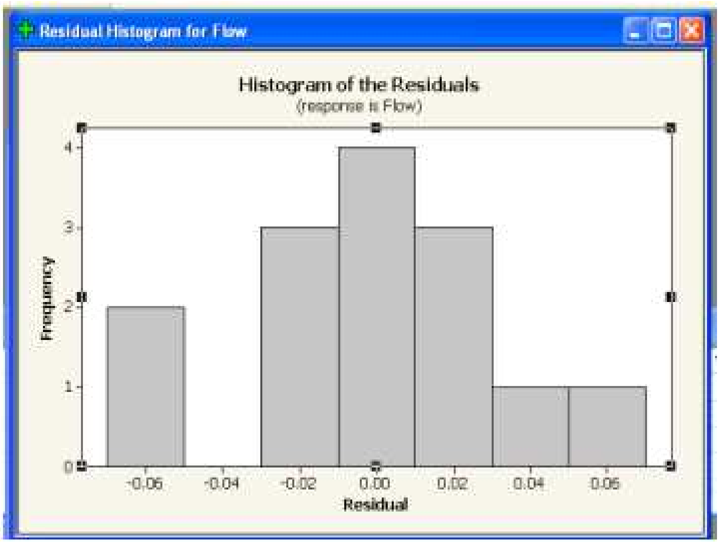
- Si los residuos vienen de una distribución normal, los puntos seguirán una línea recta aproximadamente
- Si los residuos no vienen de una distribución normal, los puntos no seguirán una línea recta

Basado en este diagrama, es razonable asumir que los residuos para tus datos no se desvían substancialmente una distribución normal. (Una prueba de normalidad para estos datos (no mostrado) permitió un p-valor de 0.340.)

Histograma
Puedes también usar el histograma de los residuos para evaluar la normalidad. Sin embargo, la probabilidad normal es generalmente más fácil de interpretar sobre todo para las muestras pequeñas.



Interpretando tus resultados
Grafica 1



En la Grafica 1 se presenta los residuales en el orden de la recolección de datos (los datos que entraron en el mismo orden en los que fueron recolectados) usa esta grafica para verificar que los residuos son independientes.

- Si hay un efecto debido al orden de colección de datos los residuales en ceros no serán esparcidos al azar. Podrás detectar una tendencia en el plot.

- Si no hay efecto debido al orden de colección de datos, los residuales en ceros se esparcirán al azar

Los datos no aparecerán en ningún tiempo o los efectos del orden de los datos presentes.

Interpretando tus resultados

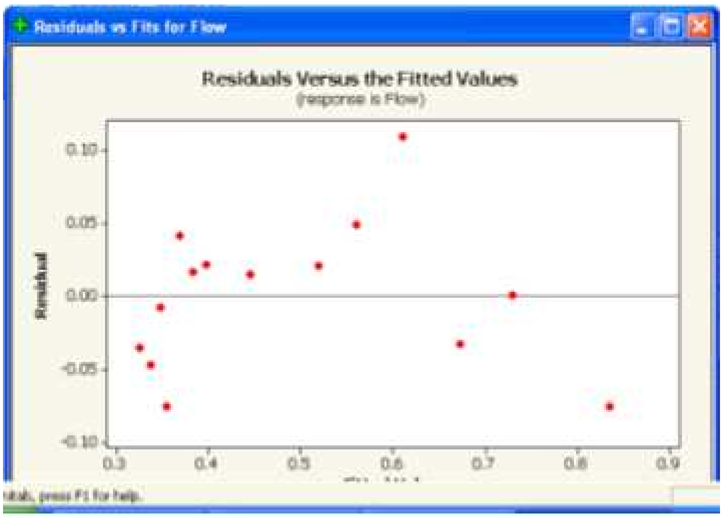
Los residuales vs Fits

Use el plot de los residuales vs los Fits para verificarlo:

- El modelo no está perdiendo ninguna condición cuadrática
- La variación es constante por todo los valores de los Fits.
- No hay datos fuera de línea.

Si ves cualquier tipo de modelo en el plot uno de estas asunciones ha sido violada. Tu puedes ver en el siguiente cuadro debajo el resumen de los modelos típicos.

Este modelo	Indica…
Curvilíneo	Un término cuadrático puede estar perdiendo su modelo.
La extensión desigual de las residuales a través de los diferentes valores ajustados.	La variación de los residuales no es constante
Un punto está situado muy lejos del cero.	Fuera de línea



Agregando confianza y predicción a las Cintas

Creando una nueva fitted line plot del modelo agregando confianza y predicción a las cintas. Mostrando las cintas y los intervalos te da una mejor idea de la variabilidad y estabilidad del modelo cuadrático.

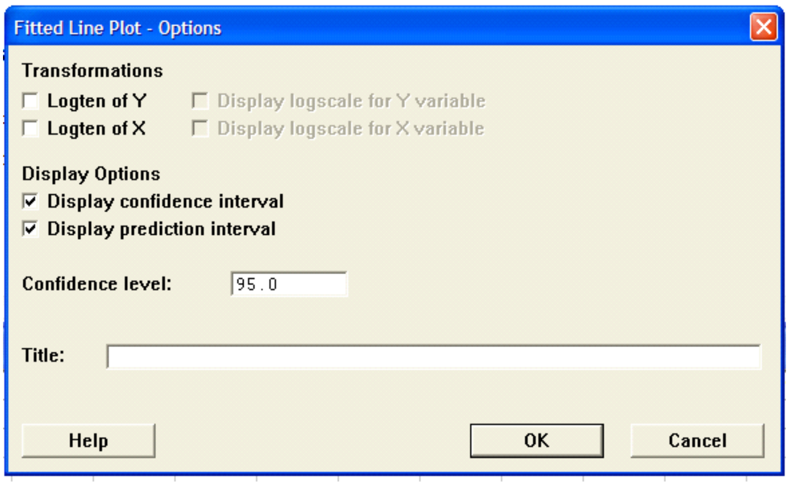
Fitted Line Plot

1.- Escoge Stat **> Regression > Fitted Line Plot.**

2. -Bajo **Type of Regresión Model**, Escoge **Quadratic.**

3.-Pulse el botón las O**pcions.**

4.-Completa el recuadro como se indica a continuación:

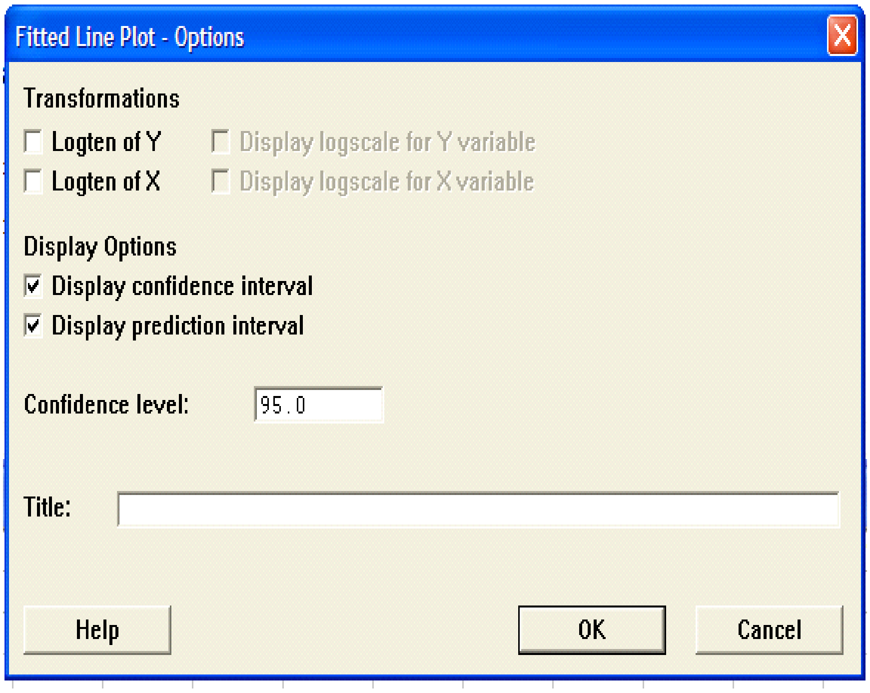


5.- pulse el botón OK en cada cuadro de diálogo

Interpretando tus resultados

El intervalo de confianza

de las líneas indicadas.



Consideraciones Finales
Conclusiones prácticas
El análisis indica que la relación entre la profundidad de la corriente y proporción del flujo es más bien cuadrática que lineal. Cuando la corriente es baja, pequeños incrementos se muestran en los resultados de la profundidad y grandes incrementos en el flujo. Sin embargo, cuando la corriente llega a ser mas profundo, los mismos incrementos en la profundidad causan menos cambios en el flujo.
Consideraciones Estadísticas
Tu no puedes usar la análisis de la regresión para afirmar que los cambios en las predicciones cambian las causas en la respuesta, a menos que el valor predictivo fuere arreglado en la predeterminación de niveles en un experimento controlado. Si los valores predictivos se permiten variar al azar, otros factores pueden influenciar en ambos los predictivos y la respuesta.
No debes aplicar los resultados de la regresión para responder a los valores que están fuera del rango de la muestra.

Ejercicio 3.1 Descarga Diesel
Estas investigando los efectos de humedad en las emisiones de la descarga de camiones diesel
Recolección de datos
Los datos son de la Hare C.T. (1997). “Light Duty Diesel Emisión Correction Factors for Ambient Conditions” el informe final a la Agencia de protección del ambiente bajo contrato No. 68-02-1777, Instituto de la investigación sudoeste, San Antonio, TX.
Instrucciones

1.- La información de la grafica visualiza la relación entre las variables.
2.- Usa Fitted Line Plot para adaptar el modelo apropiado de regresión.
3.- Asegúrate de verificar las asunciones necesarias con las graficas los residuales.

Set de Datos
EL DIESEL. MPJ

Nombre	Tipo de dato	Tipo de variable
Nox	Numérica	Respuesta
Humidity	Numérica	Predicción

REGRESIÓN MÚLTIPLE
Ejemplo 4 Reduciendo el golpe del Motor
Problema
Trataras de identificar las llaves predoctoras del golpe del motor.
Las siguientes variables están bajo las siguientes consideraciones:

- La elección del momento adecuado de la chispa
- La proporción de aire-combustible (AFR)
- La temperatura de la succión
- La temperatura de la descarga

Recolección de los datos
Los datos son recolectados al azar de 13 motores seleccionados, todos trabajan con gasolina con un octanaje tasa de 87.
Herramientas

Graph > Matrix plot
Stat > Basic Statistics > Correlation
Stat > Regression > Regression

Set de Datos
KNOCK.MPJ

Nombre	Tipo de dato	Tipo de variable
Spark	Numérica	Predicción
AFR	Numérica	Predicción
Intake	Numérica	Predicción
Exhaust	Numérica	Predicción
Knock	Numérica	Respuesta

Regresión Múltiple
¿Cuál es la regresión múltiple?
La regresión múltiple examina la relación entre una respuesta continua variable (Y) y más de un predictor (X) de variables. La ecuación general para un modelo de la regresión múltiple es:

$Y .= \beta 0 + \beta 1\ X1 + \beta 2\ X2 + \beta 3\ X3 + \ldots. + \varepsilon$

Donde Y es que la respuesta, $\beta 0$ es el intercepte cada Xi es un predictor variable con una cuesta de βi, y ε es el error aleatorio.
Cuándo usar la regresión múltiple
Use la regresión múltiple cuando tienes un Y continuo y más de una X.

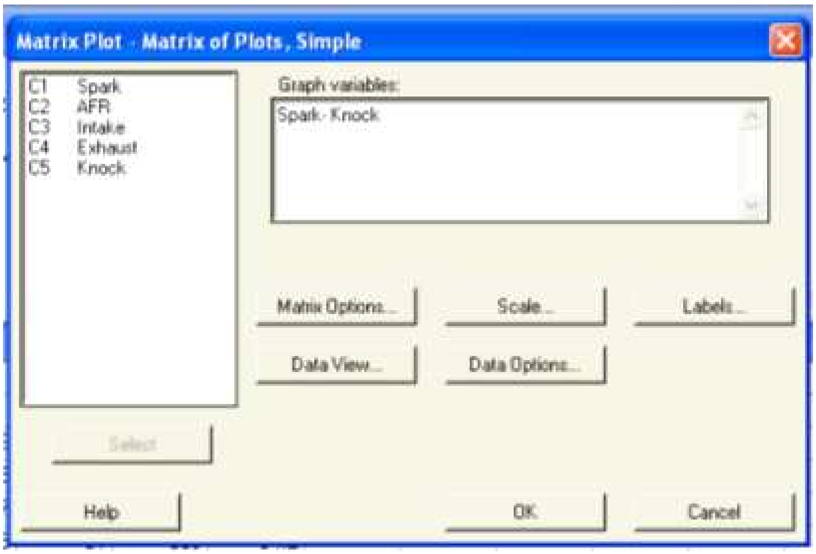
- X puede ser ordinal, o continua.
- En teoría, X debería arreglarse. En la practica, sin embargo, con frecuencia permite la varianza.
- Cualquier variación aleatoria en la medida de X se asume que es una comparación insignificante en el rango en el cual X es medido.

Antes de aceptar los resultados de análisis de la regresión, debes verificar las siguientes asunciones sobre los residuales que son válidos para la información:

- Ellos deben ser independientes (y así aleatorios).
- Ellos deben ser de distribución normal.
- Ellos deben tener una variación constante por todo los valores de X .

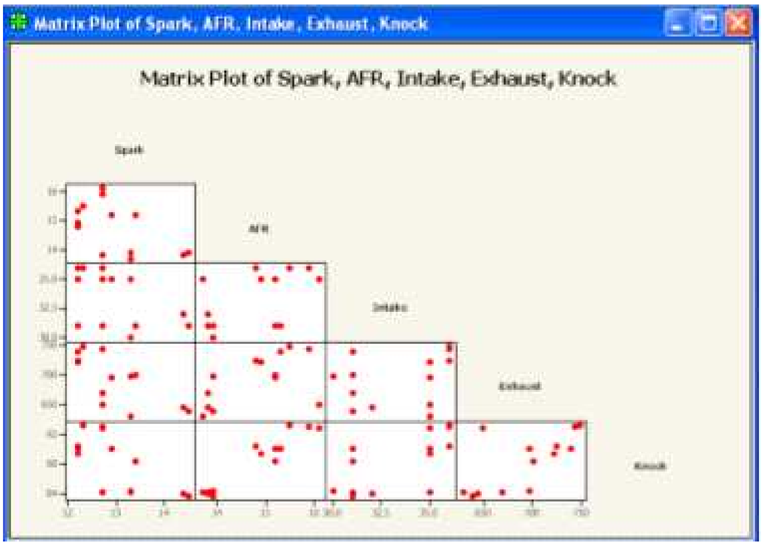
Por qué usa la regresión múltiple
La regresión múltiple puede ayudar a contestar las siguientes preguntas:
¿Qué tan importantes son tus variables X en predicción con tus valores Y?
¿Qué valor esperas de Y cuando X1 es 20 y X2 es 3?
¿Cuánto cambiarán Y si aumentas X3 por una unidad?
Por ejemplo,
¿Cómo procesas la temperatura y porosidad relacionada a la dureza del acero?
¿Qué tan duro esperas que tu acero esta si tu proceso se encuentra a cierta temperatura por cierto tiempo?
¿Qué tan resistente es la dureza del acero si incrementas la temperatura a 100 º?
Creando una Matriz Plot
Usaras primero una matriz plot y coeficientes de correlación primero para ver si las relaciones existen entre la contestación inconstante y las variables de la predicción.
Variables del gráfico
Es más fácil mirar la relación entre la respuesta y la predicción cuando entras en la respuesta de la ultima variable en las variables del gráfico.

Matriz Plot
1.- Abre el proyecto KNOCK.MPJ
2. - Escoge **Graph > Matriz Plot.**
3.- Complete el recuadro como se indica a continuación:

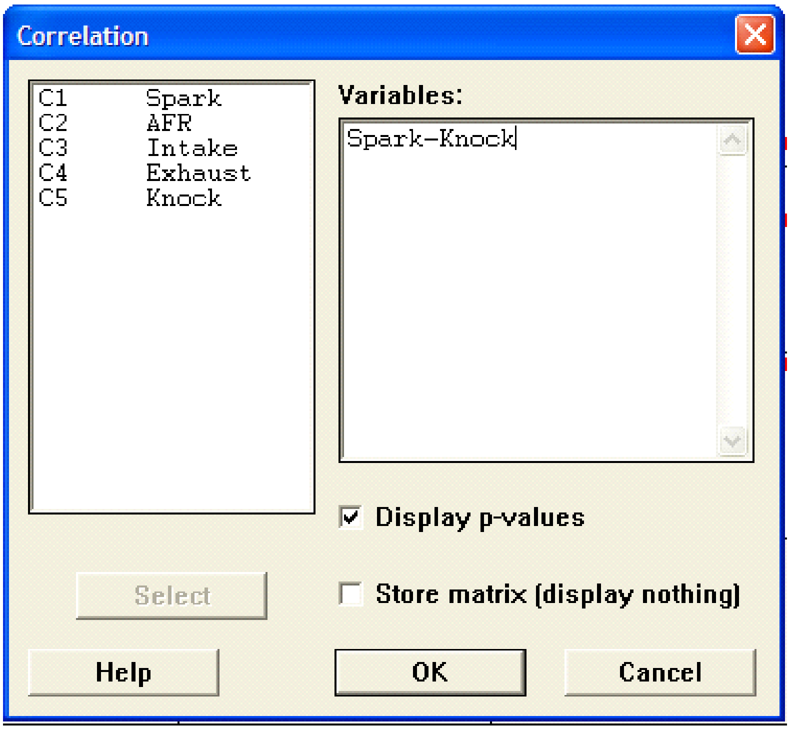


4. - Click Options**.**
5. - Bajo **Matriz Display** Escoge **Lower Left.**
6. - Click **OK** cada recuadro.

Interpretando tus resultados
Los resultados incluyen los gráficos para cada combinación de variables.
Fíjate para evaluar la relación entre el golpe y las predicciones.
Parece ser una correlación negativa entre el golpe y chispa. Allí también parece ser correlaciones positivas entre el golpe y cada uno de las predicciones restante
¿Qué sigue?
Usa la correlación para evaluar las fuerzas de relación lineal.



Cálculo de las Correlaciones múltiples
Cree una matriz de correlación para evaluar las asociaciones entre el golpe y las predicciones.
Correlación
1.- Escoge **Stat>Basic Statistics>Correlation**
2.- Completa el recuadro como se indica a continuación:



3.- Click **OK**

Interpretando tus resultados
La salida incluye el coeficiente de correlación y el p-valor para cada par de variables. (Use un 0.05 para todas las comparaciones.) Una sugerencia en la matriz plot, hay una correlación negativa significante entre el golpe y chispa (r = -0.699, p=0.008). Hay también, correlaciones positivas significantes entre el golpe y cada uno de las predicciones restante:

- AFR(r = 0.961,P = 0.000)
- Intake (r =0.673.P = 0.012)
- Exhaust (r = 0.682, P = 0.010)

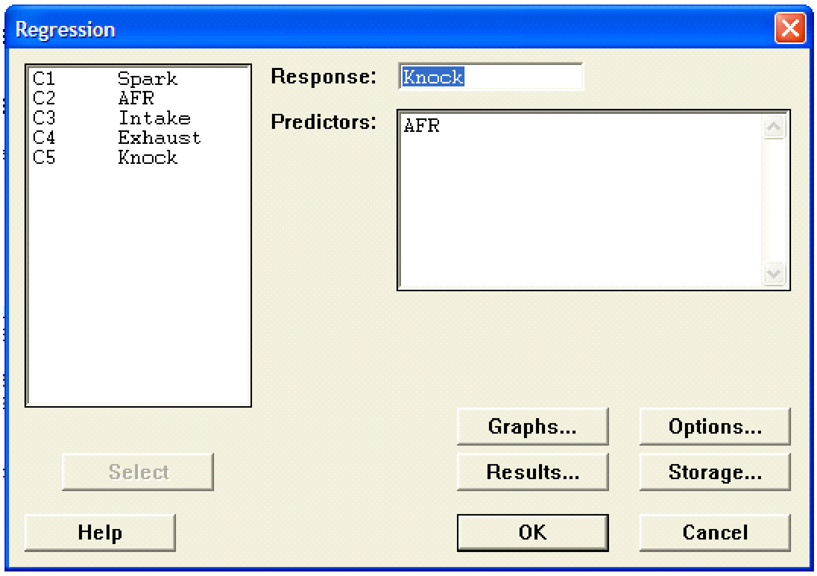
Que sigue.
Porque AFR tiene la relación lineal más fuerte con la regresión de uso de golpe para ajustarse a un modelo de la regresión lineal simple con el golpe como la contestación y AFR como las predicciones.

Correlations: Spark, AFR, Intake, Exhaust, Knock

	Spark	AFR	Intake	Exhaust
AFR	-0.580 0.038			
Intake	-0.500 0.082	0.521 0.068		
Exhaust	-0.723 0.005	0.587 0.035	0.291 0.335	
Knock	-0.699 0.008	0.961 0.000	0.673 0.012	0.682 0.010

Cell Contents: Pearson correlation
P-Value

Encajando a un modelo de la regresión simple
Usa la regresión para realizar un análisis de la regresión lineal simple para el golpe y AFR. Podrías también usar Fitted Line Plot antes de realizar un análisis.
Regresión
1.- Escoja **Stat > Regresión > Regression**
2.- Completa el recuadro como se indica a continuación:



3.- Click **OK**

Interpretando tus resultados
Ecuación de la Regresión
La ecuación relacionada con la respuesta y la predicción es:

Knock = 25.5+4.25 AFR

Esto indica que el golpe aumenta 4.25 veces por el aumento de la unidad en AFR
Tabla de coeficientes
Las hipótesis para cada coeficiente es:

- Ho: el coeficiente es igual a cero
- H1: el coeficiente no es igual a cero

El valor-p para la constante (β0 , la intercepción) y el coeficiente de AFR (β1, la cuesta) ambos son menores de 0.05. Así nosotros podemos rechazar Ho para cada uno a los 0.05 α-level y concluimos que estos coeficientes no son cero. En este modelo, AFR es una predicción significativamente estadística del golpe.

El análisis de la regresión: el golpe contra AFR

Regression Analysis: Knock versus AFR

```
The regression equation is
Knock = 25.5 + 4.25 AFR


Predictor     Coef   SE Coef       T      P
Constant    25.506     5.483    4.65  0.001
AFR         4.2543    0.3707   11.48  0.000


S = 1.09886   R-Sq = 92.3%   R-Sq(adj) = 91.6%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  159.05  159.05  131.72  0.000
Residual Error  11   13.28    1.21
Total           12  172.33
```

Interpretando tus resultados
R²(R-Sq)

El R² indica los 92.3% de la variabilidad del golpe predicho por este modelo.
El Análisis de la varianza
Llamada que las hipótesis para un modelo de la regresión lineal simple son:

- Ho: β1 es igual a cero
- H1: β1 no es igual a cero

¿Qué sigue?
El modelo de la regresión simple con AFR es útil para la predicción del golpe. Sin embargo, es posible que el Power de la

predicción adicional puede ser ganada incluyendo otras predicciones en el modelo de regresión.

El análisis de la regresión: el golpe contra AFR

Regression Analysis: Knock versus AFR

```
The regression equation is
Knock = 25.5 + 4.25 AFR


Predictor     Coef  SE Coef      T      P
Constant    25.506    5.483   4.65  0.001
AFR         4.2543   0.3707  11.48  0.000


S = 1.09886   R-Sq = 92.3%   R-Sq(adj) = 91.6%


Analysis of Variance

Source          DF      SS      MS       F      P
Regression       1  159.05  159.05  131.72  0.000
Residual Error  11   13.28    1.21
Total           12  172.33
```

Examinando la asociación residual

Los residuos contra las variables

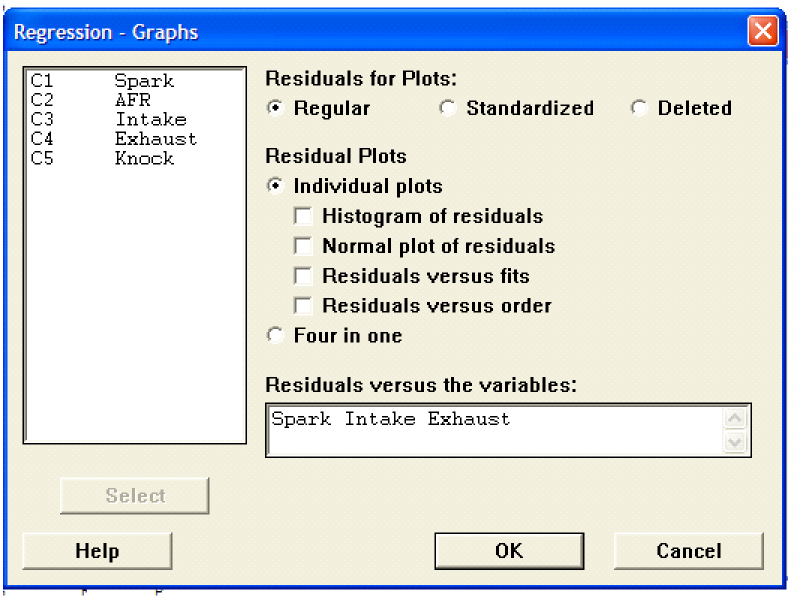
Una técnica por determinar si otras variables pueden ser importantes en predecir la respuesta es la grafica de los residuales contra cada predicción potencial.

Regresión

1.- Escoja el **Stat > Regresión > Regression** o presione ctrl+E para regresar al cuadro de diálogo.

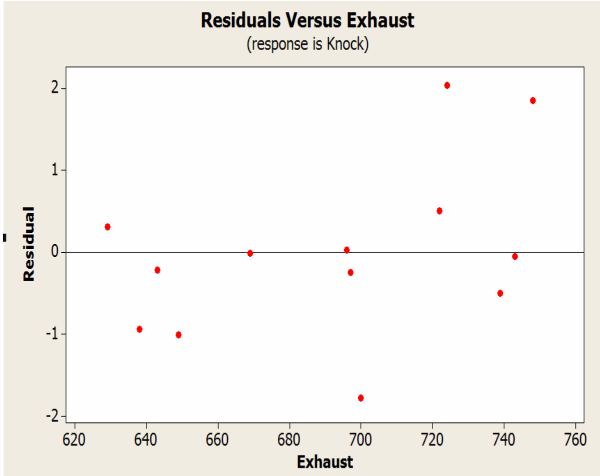
2.- Click **Graphics.**

3.- Completa el recuadro como se indica a continuación:

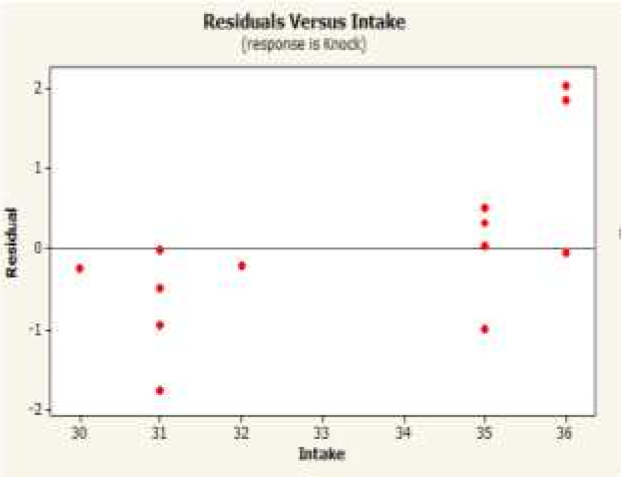


4.- Pulse el botón OK en cada recuadro.

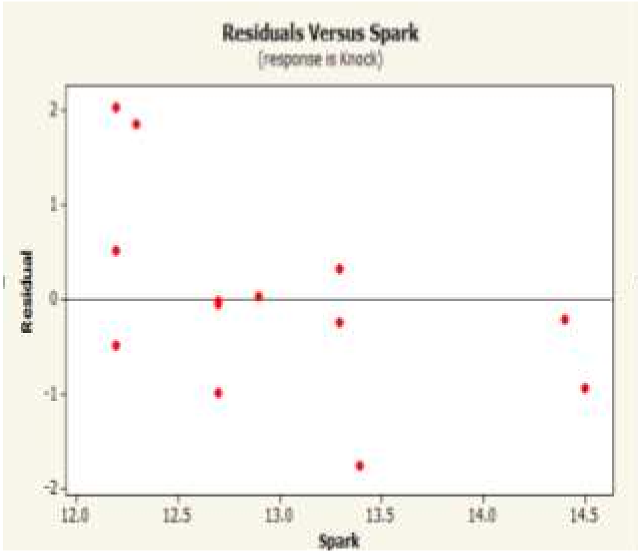
Interpretando tus resultados



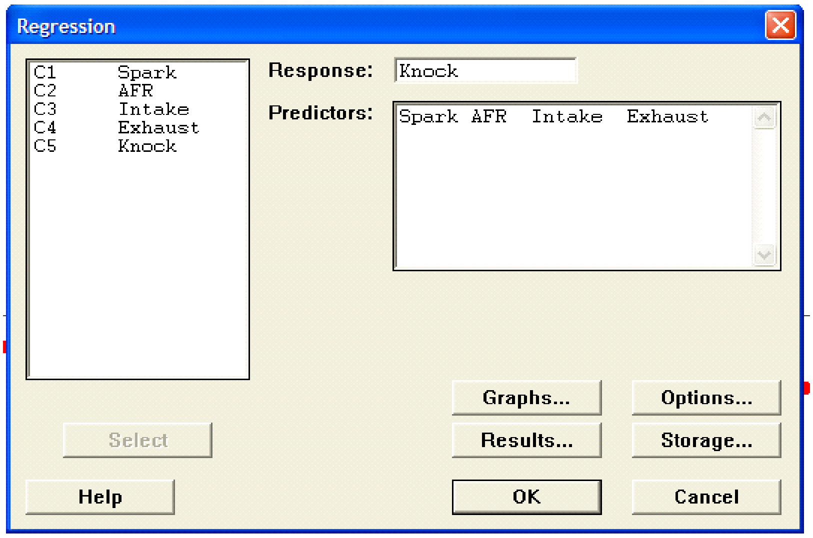
Cuando el plotted en contra de la descarga, los residuales no parecen completamente aleatorios. Los residuales aparecen más grandes para los valores de descarga más grandes. Esto indica que la descarga puede ser útil respondiendo a la variabilidad adicional en el golpe.
La entrada y la chispa también parecen ser relacionadas con el golpear y pueden responder a la variabilidad adicional.
Es posible para dos o más variables explicar la misma variabilidad en la respuesta. En este caso, el modelo final puede que no incluya todas las variables.



Encajando a un modelo de la regresión múltiple
Usa la regresión para analizar el modelo de la regresión múltiple con todos las cuatro predicciones.



Regresión
1.- Escoge **Stat > Regresión > Regresión** o presione Ctrl.+E para regresar a la **Regresión** en el recuadro.
2.- Presione F3 para borrar el cuadro de diálogo
3.- Completa el recuadro como se indica a continuación:



4.-Click **OK**.

Interpretando tus resultados
Use un α de 0.05 para todos los análisis.
Ecuación de regresión
La ecuación que relaciona la respuesta y la predicción es:
Knock = 23.8 - 0.296 +3.19 AFR +0.359 entrada + 0.0134 descarga

Tabla de coeficientes
Tenga el cuidado al interpretar los coeficientes de la regresión múltiple.
El p-valor para cada variable sólo indica si es significante en el modelo presente.
Por ejemplo, la chispa no es una predicción significante en el modelo presente (p = 0.363). Sin embargo, si quitas la descarga del análisis, la chispa se hace significativa, está altamente correlacionados (r = - 0.723,p = 0.005, vea página 3.49) y así explica la misma variación en el golpe.

El análisis de la regresión: el golpe contra la chispa, succión, la descarga,

```
Constant     23.815     8.137   2.93  0.019
Spark       -0.2965    0.3072  -0.97  0.363
AFR          3.1918    0.2398  13.31  0.000
Intake      0.35870   0.07848   4.57  0.002
Exhaust    0.013376  0.005421   2.47  0.039


S = 0.510560   R-Sq = 98.8%   R-Sq(adj) = 98.2%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       4  170.245  42.561  163.28  0.000
Residual Error   8    2.085   0.261
Total           12  172.331


Source   DF  Seq SS
Spark     1  84.250
AFR       1  80.029
Intake    1   4.380
Exhaust   1   1.587
```

Interpretando tus resultados
Cuidado con multicolinealidad
Cuando las predicciones son sumamente correlacionadas, la estimación del coeficiente de regresión puede ser inestable (Significa que varían ampliamente de un ejemplo al siguiente). Esta condición es llamada multicolinealidad, y eso hace que la evaluación sea importante en términos individuales en la dificultad del modelo.
Puedes usar la correlación para tratar de identificar las fuentes potenciales de la multicolinealidad. Si hay multicolinealidad extrema en un modelo, MINITAB mostrará un mensaje en la ventana de la sesión y quita una o más variables para reducir el problema.
Nunca quites mas de una predicción en ningún momento
Una buena forma de Escoger las predicciones de un modelo de regresión múltiple es tratar a todas las combinaciones potenciales usando el modelo de procedimientos de comparación como el mejor subconjuntos o una regresión gradual.

El análisis de la regresión: el golpe contra la chispa, succión, la descarga

```
Constant     23.815     8.137   2.93  0.019
Spark       -0.2965    0.3072  -0.97  0.363
AFR          3.1918    0.2398  13.31  0.000
Intake      0.35870   0.07848   4.57  0.002
Exhaust    0.013376  0.005421   2.47  0.039


S = 0.510560   R-Sq = 98.8%   R-Sq(adj) = 98.2%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       4  170.245  42.561  163.28  0.000
Residual Error   8    2.085   0.261
Total           12  172.331


Source   DF  Seq SS
Spark     1  84.250
AFR       1  80.029
Intake    1   4.380
Exhaust   1   1.587
```
Ir

Interpretando tus resultados

R^2(R-Sq) y R^2 ajustó (R-Sq(adj))

El nuevo modelo explica 98.8% de la variabilidad en la respuesta, lo cual es una mejora sobre el R^2 logrando usar solamente AFR para predecir el golpe

Sin embargo R^2 nunca disminuirá cuando aumente la predicción al modelo, aun cuando eso no resulte un buen modelo. La estadística de R^2 ajustada (R-Sq(adj) = 98.2%) es ajustado para el número de condiciones en el modelo, y debe usarse cuando son comparados los modelos con diferente números las predicciones.

El R^2 ajustado para el modelo con sólo AFR como el predictor tenía 91.6% años. así, el modelo actual con un R^2 esta ajustado 98.2% se mejora.

Análisis de Variación

Las hipótesis para un modelo de la regresión múltiple es:

Ho: todo $\beta 1$ (a excepción de $\beta 0$) son iguales a cero

H1: al menos uno βi (no incluye $\beta 0$) no es igual a cero

Porque p (0.000) es menos que α (0.05), puedes rechazar Ho.

El modelo de la regresión, con la chispa, AFR, Succión, y descarga como las predicciones, es significativamente mejor que la restricción del modelo el cual incluye no predicciones.

El análisis de la regresión: el golpe contra la chispa, succión, la descarga,

```
Constant     23.815     8.137   2.93  0.019
Spark       -0.2965    0.3072  -0.97  0.363
AFR          3.1918    0.2398  13.31  0.000
Intake      0.35870   0.07848   4.57  0.002
Exhaust    0.013376  0.005421   2.47  0.039


S = 0.510560   R-Sq = 98.8%   R-Sq(adj) = 98.2%


Analysis of Variance

Source          DF       SS      MS       F      P
Regression       4  170.245  42.561  163.28  0.000
Residual Error   8    2.085   0.261
Total           12  172.331


Source   DF  Seq SS
Spark     1  84.250
AFR       1  80.029
Intake    1   4.380
Exhaust   1   1.587
```

Consideraciones Finales

Conclusiones prácticas

La ecuación de la regresión para la Chispa usando ejemplar, AFR, Succión y Descarga para predecir el Golpe es:

El golpe = 23.8-0.296 Chispa + 3.19 AFR + 0.359 Succión +0.0134 Descarga.

Este modelo responde de 98.8% de la variabilidad en el Golpe.

Hay problemas del multicolieanidad con el modelo. Sin embargo, la chispa sumamente correlacionado con la Descarga.

En el próximo ejemplo, usarás los mejores Subconjuntos para procesar a todos los posibles modelos con estas cuatro predicciones y Escoger el mejor.

Consideraciones estadísticas

No puedes usar el análisis de la regresión para afirmar que los cambias en las predicciones causan cambios en la respuesta, a menos que los valores de las predicciones cambien niveles predeterminados en un experimento controlado. Si los valores de las predicciones variar al azar, otros factores pueden influir en las predicciones y la respuesta.

No deberías aplicar los resultados de regresión y los valores de respuesta que son salidas de tu rango de los ejemplos.

La precisión de medida es importante. La falta de precisión te lleva a la inexactitud estimada de los coeficientes.

Ten cuidado de no pasar por alto los factores potencialmente importantes al diseñar un estudio de regresión.

Tenga cuidado con multicolinealidad.

Cuando las variables de la predicción están sumamente correlacionadas:

- Los coeficientes estimados de la regresión pueden ser inestables (Ellos pueden variar ampliamente de una muestra a la siguiente muestra)
- Puede ser difícil evaluar la importancia de las condiciones individuales del modelo.

Nunca quite más de una predicción en ningún momento.

Una buena forma de Escoger las predicciones de un modelo de regresión múltiple es tratar a todas las combinaciones potenciales usando el modelo de procedimientos de comparación como el mejor sub conjuntos o una regresión gradual.

Mejores Subconjuntos de la Regresión

El ejemplo 5 Reduciendo el Golpe del Motor

Problema

Estás intentando identificar las variables importantes que efectúan el Golpe del motor. Las siguientes variables están bajo las consideraciones:

- La elección del momento adecuado de la chispa
- La proporción de aire-combustible (AFR)
- La temperatura de la succión
- La temperatura de la descarga

Recolección de datos

Los datos son recolectados al azar de 13 motores seleccionados

Herramientas

Stat > Regressions>Best subsets.

Stat>Regressions>Regressions.

Set de Datos

KNOCK.MPJ

Nombre	Tipo de dato	Tipo de variable
Spark	Numérico	Predictor
AFR	Numérico	Predictor
Intake	Numérico	Predictor
Exhaust	Numérico	Predictor
Knock	Numérico	Respuesta

REGRESIONES DE LOS MEJORES SUBCONJUNTOS

¿Cuál es el mejor subconjunto de regresión?

La regresión de los mejores subconjuntos evalúa todas las posibles combinaciones de las predicciones para ayudarle a determinar qué combinación hace al mejor modelo de las regresiones. MINITAB usa un criterio de R2 máximo para Escoger al mejor modelo . Otro criterio puede proporcionar a un modelo diferente.

¿Cuándo usar los mejores conjuntos de regresión?

Use la regresión de los mejores subconjuntos cuando usted tiene mucho potencial de predicciones y así varios modelos de regresión para Escoger.

¿Por qué usar el mejor subconjunto de regresión?

Los mejores subconjuntos pueden disipar las siguientes preguntas:

- ¿ Qué combinación de tus factores es él más eficaz para predecir tu respuesta?
- ¿Cuál es el mejor modelo de regresión posible usando de 5 a 20 predicciones?

Por ejemplo,

- ¿Está un modelo usando 10 variables para predecir la suavidad del helado mas que uno que usa sólo temperatura y velocidad en la mezcla?

Escogiendo un modelo apropiado

Use los mejores Subconjuntos para ayudarle a Escoger a un modelo de las regresiones múltiples para el Golpe y evita los problemas siguientes:

- Los modelos incómodos e ineficaces son el resultado de muchas predicciones.
- Coeficientes inestables que resultan de redundante y predicciones correlacionadas.
- Habilidad inadecuada de predicciones que resulta pocas predicciones.

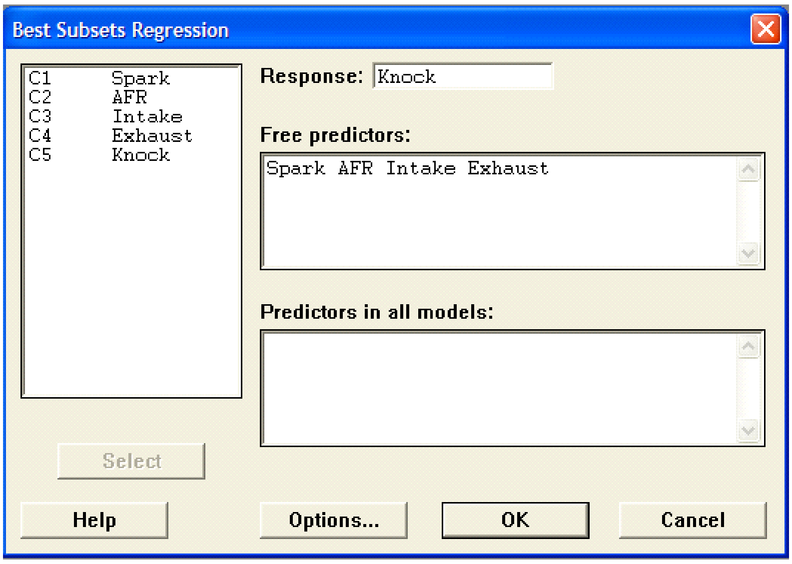
Predicciones libres

Entre todas las cuatro variables en las predicciones Libre. MINITAB probará todas las posibles combinaciones de estas variables y el reporte estadístico para los mejores modelos. (Variables de entradas de las Predicciones en todos los modelos serán incluidas en cada modelo.)

Best Subsets

1.- Seleccione **Stat>Regressions>Best**.

2.-Completa el recuadro como se indica a continuación:



3.- Click **OK**

Interpretando tus resultados

Los Xs al derecho de la tabla indica qué predicciones son incluidas en cada modelo.

Variables

La columna de Vars indica el número de predicciones en el modelo.

R2 (R-Sq) y R2 ajustó (R-Sq(adj))
Al comparar a modelos:
Si el número de predicciones es el mismo, busque al modelo con el R2 más alto.
Si el número de predicciones es diferente, busque al modelo con el R2 más alto.
Cp
Busque a modelos dónde Cp es pequeño y acerca el número de parámetros en el modelo. Por ejemplo, para modelo con 3 predicciones y el interceptor, busque a un modelo con un Cp cerca de 4 La fórmula Para Cp es:

Cp = (SSEp/MSEm)-(n-2p)

Donde SSEp son las sumas de error de los cuadrados para el modelo con los parámetros de p (incluso el interceptor), MSEm el error de la media cuadrada para el modelo con toda las predicciones de m, y n es el número de observaciones.
Los mejores Subconjuntos de regresión: el Golpe contra la Chispa, AFR, la Succión, la Descarga,
La contestación es el Golpe

Best Subsets Regression: Knock versus Spark, AFR, Intake, Exhaust

Response is Knock

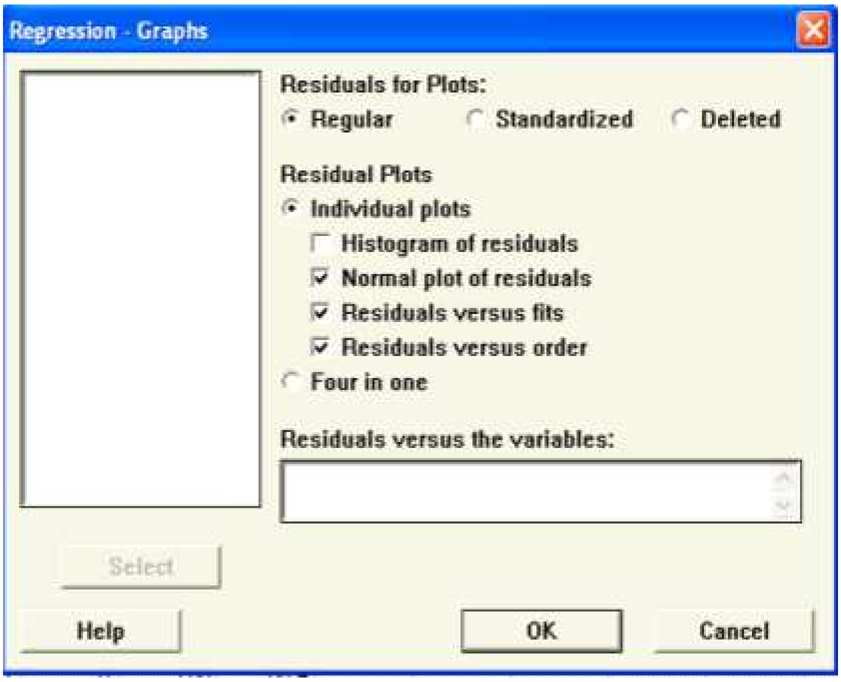
Vars	R-Sq	R-Sq(adj)	Mallows C-p	S	Spark	AFR	Intake	Exhaust
1	92.3	91.6	42.0	1.0989		X		
1	48.9	44.2	328.9	2.8297	X			
2	96.4	95.7	16.9	0.78871		X	X	
2	95.3	94.4	23.9	0.89735	X	X		
3	98.6	98.2	3.9	0.50862		X	X	X
3	97.9	97.2	9.1	0.63879	X	X	X	
4	98.8	98.2	5.0	0.51056	X	X	X	X

Interpretación tus resultados
Variabilidad
S es una estimación de la media variabilidad sobre la línea de las regresiones. Matemáticamente, S es la raíz cuadrada positiva del MSE. En general, tu quieres que S sea tan pequeño como posible.
Conclusión
Basado en éstos criterios, el modelo con AFR, la Succión, y la Descarga es el mejor. El modelo
Conteniendo todos las cuatro predicciones es comparable, pero S para este modelo es ligeramente más grande y allí no parece ser cualquier ganancia en R2 ajustado para usar el modelo. Es generalmente sabio Escoger al modelo más simple a menos que un modelo más complicado sea claramente mejor.
Los mejores Subconjuntos de regresión: el Golpe contra la Chispa, AFR, la Succión, la Descarga
La contestación es el Golpe

Best Subsets Regression: Knock versus Spark, AFR, Intake, Exhaust

```
Response is Knock

                                                E
                                              I x
                                            S n h
                                            p t a
                                            a A a u
                           Mallows          r F k s
Vars  R-Sq  R-Sq(adj)       C-p        S    k R e t
   1  92.3       91.6      42.0   1.0989      X
   1  48.9       44.2     328.9   2.8297    X
   2  96.4       95.7      16.9  0.78871      X X
   2  95.3       94.4      23.9  0.89735    X X
   3  98.6       98.2       3.9  0.50862      X X X
   3  97.9       97.2       9.1  0.63879    X X X
   4  98.8       98.2       5.0  0.51056    X X X X
```

Evaluando el último Modelo
Usa **la Regresión** para evaluar al último modelo. Calcule la ecuación de regresión y confirme que todas las asunciones sobre los residuales sean conocidas.
Regression
1.- Escoge **Stat > Regresión > Regresión**
2.- En **Response**, enter knock.
3.- En **Predictors**, enter AFR intake exhaust
4.- Click **Graphs**
5.- Complete el recuadro como se indica a continuación:



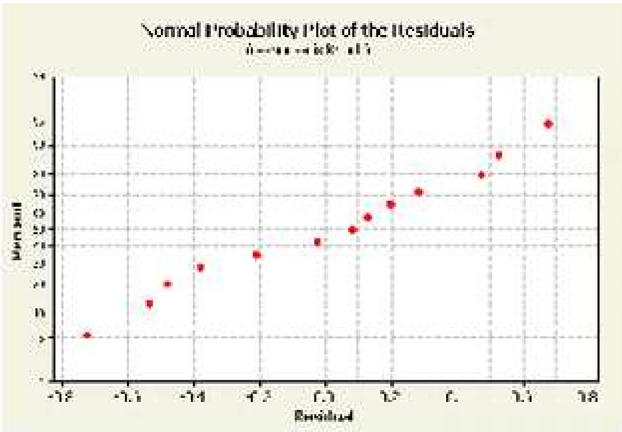
6.- click **OK** en cada recuadro.

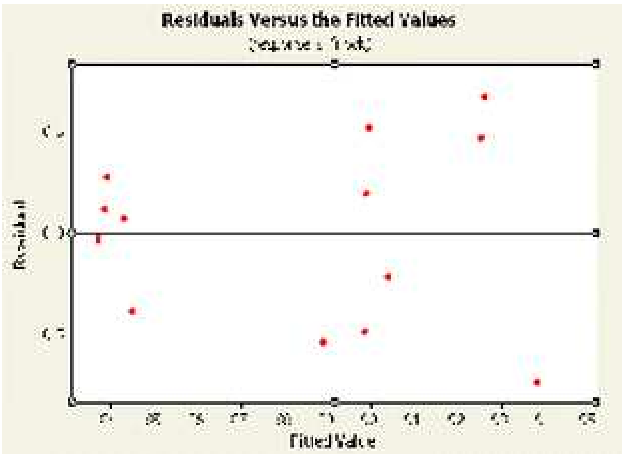
Interpretando tus resultados
Use una α de 0.05 para todos los análisis.
La ecuación de regresión
La ecuación de regresión es:
El golpe = 16.5 +3.21 AFR +O.386 Succión +0.0166 Descarga
Tabla del coeficiente
El valor de p más bajos ($p < 0.05$) en la tabla del coeficiente indica que todas las condiciones en el modelo son significantes.
Análisis de variación
Porque p (0.000) es menor que α (0.05) puedes rechazar H0. El modelo de la regresión que incluye AFR; las Succiones y la

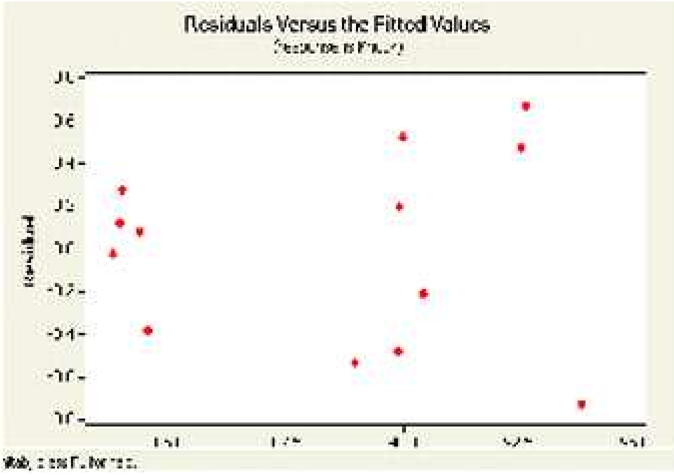
Descarga son significativamente buenas que el modelo restringido que no incluye ninguna predicción.

```
ıark        -0.2965     0.3072   -0.97   0.363
'R           3.1918     0.2398   13.31   0.000
ıtake        0.35870    0.07848   4.57   0.002
:haust      0.013376   0.005421   2.47   0.039

 = 0.510560   R-Sq = 98.8%   R-Sq(adj) = 98.2%

ıalysis of Variance

ıurce            DF        SS       MS        F       P
:gression         4   170.245   42.561   163.28   0.000
:sidual Error     8     2.085    0.261
ıtal             12   172.331


ıurce    DF   Seq SS
ıark      1   84.250
'R        1   80.029
ıtake     1    4.380
:haust    1    1.587
```

Interpretando tus resultados







Las gráficas residuales verifican que se han reunido todas las asunciones acerca de los residuales. Los residuales:

- No parta substancialmente de la normalidad.
- Aparece la distribución aleatoria a cero.
- Aparece tener la variación constante por los todos valores de ajustes.
- No exhiba un tiempo - el efecto del orden.

Consideraciones finales
Conclusiones prácticas
El mejor modelo para predecir el golpe es:
Knock= 16.5 +3.21 AFR+0.386 Intake + 0.0166 Exhaust

Consideraciones estadísticas
Antes de usar el mejor subconjunto de regresión para evaluar los modelos de regresión que son diferentes, asegúrate de que tus predicciones y respuestas son validas para todo el modelo potencial sean modelos validos de regresión.
Todos reglas y las guías también pertenecen a los modelos de la regresión múltiple también aplican cuando Escoge un modelo que usa este procedimiento.

EL ANÁLISIS DE VARIACIÓN

Objetivos:

- **Compare grupos de variables usando una prueba de varianza.**
- **Compare las medias de las muestras recolectadas en diferentes niveles de un solo factor utilizando One-Way ANOVA.**
- **Compare las medias de las muestras recolectadas en diferentes niveles de un solo factor usando el análisis de la media.**
- **Compare las medias de las muestras recolectadas en diferentes niveles de uno o más factores utilizando Balance ANOVA.**
- **Compare las medias de las muestras recolectadas en diferentes niveles en mas de un factor utilizando el Modelo Lineal General.**

Contenidos

Ejemplos y ejercicios	Propósito	Pagina
ANOVA sentido único		**136-149**
Ejemplo 1 El Precalentamiento del CRT Time	Evalúe la diferencia entre los medios del grupo para un solo factor que usa un ANOVA One-Way.	
El análisis de la media		**150-154**
Ejemplo 2 El Precalentamiento del CRT Time Revisited	Evalúe la diferencia entre los grupos de medias usando Análisis de la Media.	
ANOVA equilibrado		**155-167**
Ejemplo 3 El Uso de la pintura	Emplee el blocking de variables para reducir la variación en un análisis usando Balanced ANOVA.	
El Modelo Lineal General		**168-185**
Ejemplo 4 La Distancia de frenado	Evalúe la diferencia entre las medias del grupo para factores múltiples que usan el General el Modelo Lineal.	
Ejercicio 4.1 Prueba de Vino	Evalúe la diferencia entre las medias del grupo para factores múltiples que usan el General el Modelo Lineal.	
Ejercicio 4.2 El Volumen de fosfato	Evalúe la diferencia entre las medias del grupo para factores múltiples que usan el General el Modelo Lineal.	

One Way ANOVA

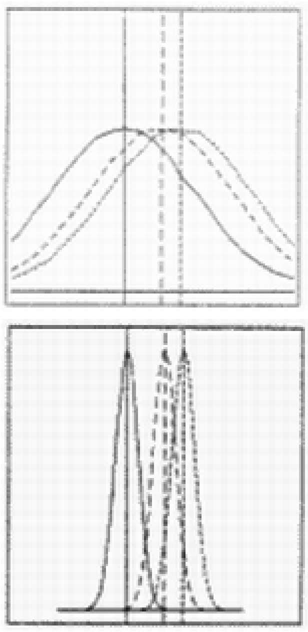
Stat > ANOVA > Test for Equal Variances.
Stat > ANOVA > One-Way.

Set de datos
CRT.MPJ

Nombre	Tipo de dato	Tipo de Variable	Niveles
Tubetype	Numérico	Factor	1,2,3
Time_sec	Numérico	Respuesta	

ANOVA One Way
¿Qué es ANOVA One-Way?
One-Way ANOVA (Análisis de varianza) procedimiento es una generalización independiente de las pruebas t. Al contrario de la prueba t. Sin embargo, One- Way ANOVA puede usarse para analizar las medias de más de dos grupos (muestras) de una vez.

La lógica básica detrás de ANOVA es:



- La variación dentro del grupo sólo es debida al error aleatorio.
- Por consiguiente, si la cantidad de variación de los grupos es similar dentro de los grupos (lo alto dela grafica), es probable que la media del grupo sólo difiera también debido al error aleatorio.
- Sin embargo, si la variación dentro del grupo es relativamente grande dentro del grupo de variación (gráfico) es probable que las diferencias entre las medias del grupo sean causadas por las diferencias por las marcadas de los niveles de factor.

¿Cuándo usar One-Way ANOVA?
Use One-Way ANOV (también llamado el solo factor ANOVA) cuando tengas respuestas continuas de datos de dos o más niveles fijos de un solo factor.
Antes de aceptar los resultados en el ANOVA, debes verificar las siguientes suposiciones acerca del residual y validar los resultados.

- La residual debe ser independiente (y ser la azar).
- La residual no tiene una desviación sustancial de la distribución normal.
- La residual debe tener constantes variaciones a través de los niveles de factor.

¿Por qué usar One Way ANOVA?
One-Way ANOVA te puede ayudar las preguntas de la respuesta como:

- ¿Hay diferencias entre los productos de tus proveedores?
- ¿Hay diferencias entre los tratamientos de los grupos?

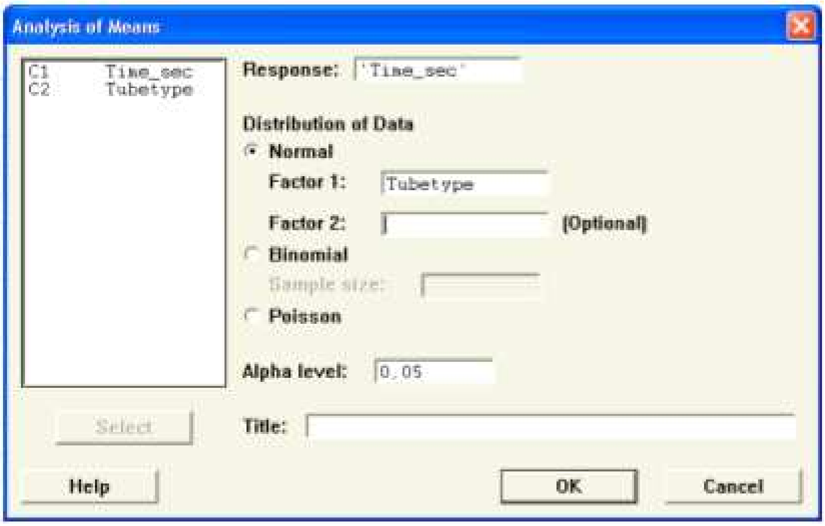
Por ejemplo

- ¿ La dureza de las muestras de plásticos de tus cuatro o proveedores son diferentes?
- ¿La combustión es más eficaz cuándo usas el aditivo de combustible A, B o ningún aditivo de combustible?
- ¿Las fuerzas de las muestras plásticas son de sus cuatro proveedores diferente? ¿La combustión es más eficaz cuándo usted usa el aditivo de combustible UN, combustible B aditivo, o ningún aditivo de combustible?

Validando la Variación Iguales
Usa la prueba para las Variaciones iguales para validar las suposiciones que las variaciones de todos los grupos comparados sean iguales.
Las respuestas de datos de cada grupo deben estar en la misma columna, con el nivel de factor indicado en otra columna.
Use el nivel de confianza de 95% por default.

Test for Equal Variances
1.- Abre el proyecto **CRT.MPJ.**
2.- Escoge **Stat >ANOVA >Test for Equal Variances.**

3.- Completo el cuadro de diálogo como indica a continuación>

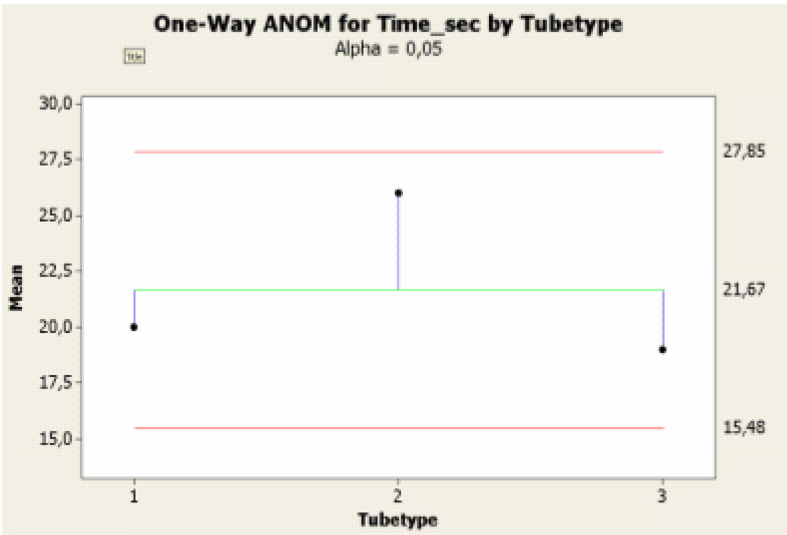


4.- Pulse el botón **OK**

Interpretando sus resultados
Intervalos de Confianza
Los intervalos de confianza son útiles para comparar la σ para las diferentes poblaciones. Sin embargo, tu decisión acerca si las variaciones son iguales deben ser basado en una prueba de varianza apropiada.
Del gráfico, aparece una sigma para el tubo tipo 2, más largo que para los otros grupos.
Pruebas de las variaciones
Los resultados incluyen dos pruebas de la variación separadas.
Qué prueba uses depende de tus datos:

- Si tus datos son continuos y normalmente distribuidos, use la Prueba de Bartlett"s. (Sí solamente se comparan dos grupos, un F-prueba reportara instantáneamente una prueba Bartlett"s).
- Si sus datos son continuos, pero no necesariamente normalmente distribuidos, use la Prueba de Levene"s.

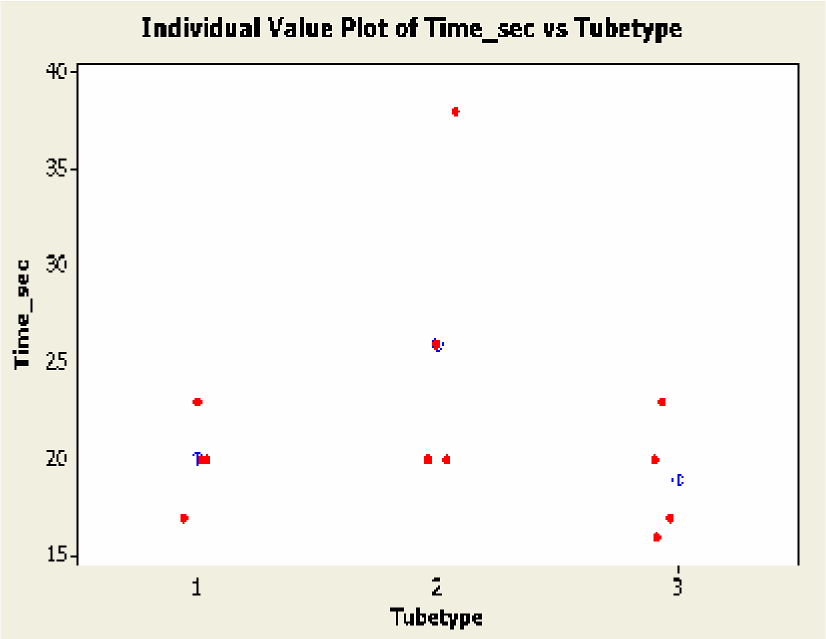
Conclusión
Los p-valores para ambas pruebas (p=0.100 para Bartlett"s Test; p=0.248 para Levene"s Test) es mayor que 0.05. Pero no hay bastante evidencia así (que con un nivel de 0.05 σ) se concluye que las variaciones no son iguales.

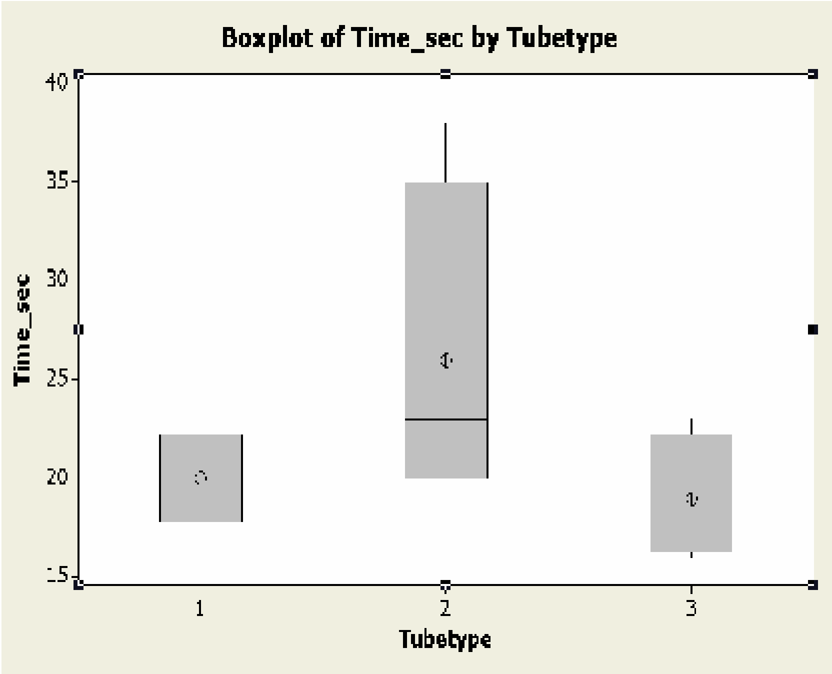


Ejecutando One-Way ANOVA
Usa One-Way ANOVA para comparar la media del tiempo de calentamiento para diferentes tipos de tubos de rayo catódico, y crea los gráficos para visualizar los datos.
One- Way
1.- Escoge el **Stat>ANOVA>One-Way.**
2.- Complete el recuadro como se indica a continuación:

3.- Click **Graphs.**
4.- Selecciona **Doplots of data and Boxplots of data.**
5.- Click **OK** en cada cuadro de diálogo.

Interpretando sus resultados
Boxplots
La grafica boxplot muestra que el rango de valores en el Grupo 2 es más grande, que el de los otros grupos.
Dotplots
La grafica dotplot revela que el grupo 2 contiene una sola observación con un extraordinario valor alto. Con sólo 4 observaciones en cada grupo, tal línea de fuera tiene largos efectos en la media y una desviación de la muestra.
Tales fuera de línea como esta puede ser el resultado de una variación aleatoria, o ellos pueden indicar que algún empalme pasó en tu proceso. Tu debes investigar las líneas de fuera para determinar que causo que eso fuera posible.
Para el presente análisis, asume que todas las observaciones son válidas.





Interpretando sus resultados
El análisis de varianza
La primera fila en la tabla del análisis de varianza contiene todas las estadísticas asociadas con el factor: tybetype. La siguiente fila contiene todas las estadísticas asociadas con el error aleatorio (error).
Los grados de libertad
Los grados de libertad (DF) se refieren al número de valores usados para calcular la suma de los cuadrados (SS) para cada fuente.
La suma de cuadrados
La suma de cuadrados (SS) es la medida de la cantidad de variabilidad que cada fuente contribuye a los datos. Note que el importe global de variabilidad en los datos (SS suman, 378.7) es igual al SS para el tubetype (114.7) más el SS para el Error (264.0).
Media cuadrada
(MS) para cada fuente es igual al SS dividió por el DF.

- El MS para el factor es una estimación del promedio de la media junto con el grupo de variabilidad.
- El MS para el error es una estimación del promedio dentro del grupo.

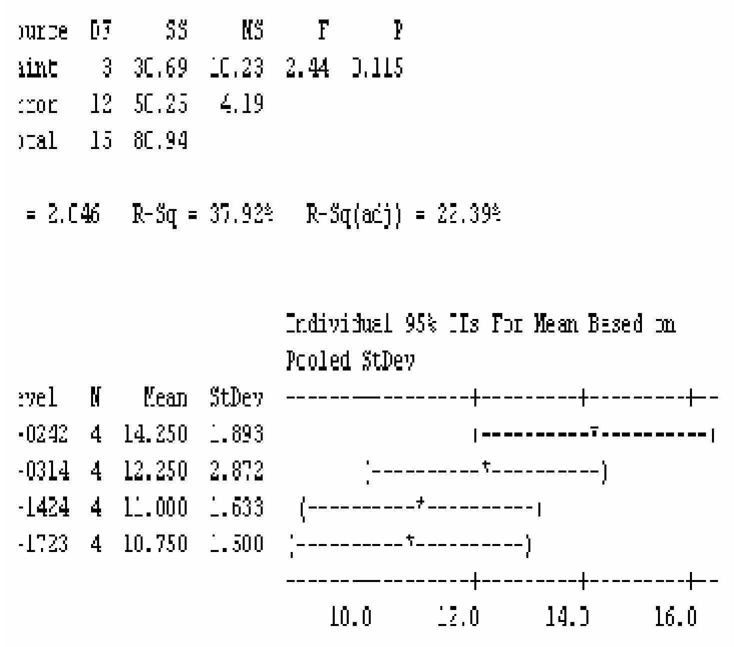
ANOVA sentido único: el time_sec contra el tubetype

```
urce  DF     SS     MS     F      P
aint   3  30.69  10.23  2.44  0.115
rror  12  50.25   4.19
tal   15  80.94

= 2.046   R-Sq = 37.92%   R-Sq(adj) = 22.39%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
evel  N    Mean  StDev  -------+---------+---------+---------+--
-0242 4  14.250  1.893                 (----------*----------)
-0314 4  12.250  2.872        (----------*----------)
-1424 4  11.000  1.633   (----------*----------)
-1723 4  10.750  1.500  (----------*----------)
                        -------+---------+---------+---------+--
                            10.0      12.0      14.0      16.0
```

Interpretando sus resultados
F-estadística

F es el radio de la variabilidad contribuida por el factor de la variabilidad contribuida por el error. Es calculado como el MS para el factor (el tubetype) dividió por el MS para el error.
Cuando las diferencias entre el nivel de factor de la media es similar a las diferencias entre las observaciones de cada nivel. F será cerrado a 1.
Si la variabilidad entre el nivel de factor de la media es mas larga que la variabilidad entre las observaciones dentro del factor, F será más grande que 1.

El P-valor
P-valor es la probabilidad que F sería tan grande como es (o más grande) si su factor no tiene los efectos. Cuando F es grande, sugiere que el nivel de factor de la media es más diferente que los esperados para la ocasión. Así que p-valor es pequeño.
Use el p - el valor de probar las hipótesis lo siguiente:
Ho(hipótesis nula) todos los factores del nivel de la son iguales.
H1(la hipótesis alternativa) todos los factores del nivel de la son diferentes.
Conclusión:
Porque P es mayor que (0.05), tu no puedes rechazar Ho. No hay suficiente evidencia para sugerir que los niveles de las medias son diferentes.



```
ource  DF     SS     MS     F      P
aint    3  30.69  10.23  2.44  0.115
rror   12  50.25   4.19
otal   15  80.94

 = 2.046   R-Sq = 37.92%   R-Sq(adj) = 22.39%

                         Individual 95% CIs For Mean Based on
                         Pooled StDev
evel   N    Mean  StDev  --------------+---------+---------+-
-0242  4  14.250  1.893                 (---------*----------)
-0314  4  12.250  2.872          (---------*---------)
-1424  4  11.000  1.633   (---------*---------)
-1723  4  10.750  1.500  (---------*---------)
                         --------------+---------+---------+-
                             10.0      12.0      14.0      16.0
```

Interpretando sus resultados
95% CIs individuales Para la Media
Para cada nivelado de tu factor MINITAB despliega el intervalo de confianza., Así como lo siguiente las estadísticas:

- N--------- Número de observaciones.
- Mean--- Media de las observaciones.
- StDev--- Desviación estándar de las observaciones

Los intervalos de confianza
Los intervalos de confianza representan rangos de valores probables para la media de cada nivel. Tu puedes estar seguro en un 95% que µ (de la población de la media) para cada nivel esta dentro del rango indicado.
Calculando los intervalos, MINITAB combina las desviaciones estándar de cada nivel con la estimación agrupada de σ (desviación estándar de la población) también llamada desviación estándar agrupada (Pooled StDev).
Note que hay mucho traslapo entre los intervalos para los tres los tipos de tubos diferentes. Ésta es una buena indicación de que las medias no son significativamente diferentes uno del otro
Sin embargo, la prueba de la comparación es necesaria antes de que cualquier conclusión pueda figurar.

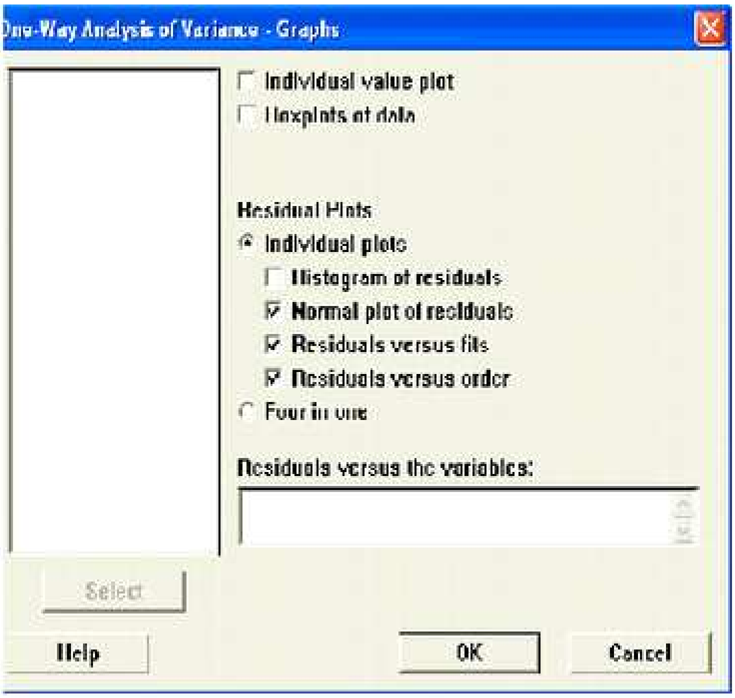
ANOVA sentido único: el time_sec contra el tubetype

```
)urce  DF     SS     MS     F      P
aint    3  30.69  10.23  2.44  0.115
:ror   12  50.25   4.19
)tal   15  80.94

 = 2.046   R-Sq = 37.92%   R-Sq(adj) = 22.39%


                          Individual 95% CIs For Mean Based on
                          Pooled StDev
:vel   N    Mean  StDev  -------+---------+---------+---------+--
-0242  4  14.250  1.893                 (----------*----------)
-0314  4  12.250  2.872       (----------*----------)
-1424  4  11.000  1.633   (----------*----------)
-1723  4  10.750  1.500  (----------*----------)
                         -------+---------+---------+---------+--
                             10.0      12.0      14.0      16.0
```

Validando las Suposiciones de la Residual
Antes de que tu puedas confiar en los resultados de un One-Way ANOVA, tu debes revisar que todas las suposiciones acerca de la residual han sido encontradas.
Usa One Way para crear unan grafica de residuales.

One-Way

1.-Escoja **Stat >ANOVA > One-Way** o presiona **Ctrs + E** para regresar al recuadro de **One Way.**
2.-Click **Graphs.**
3.- Complete el recuadro como se indica a continuación



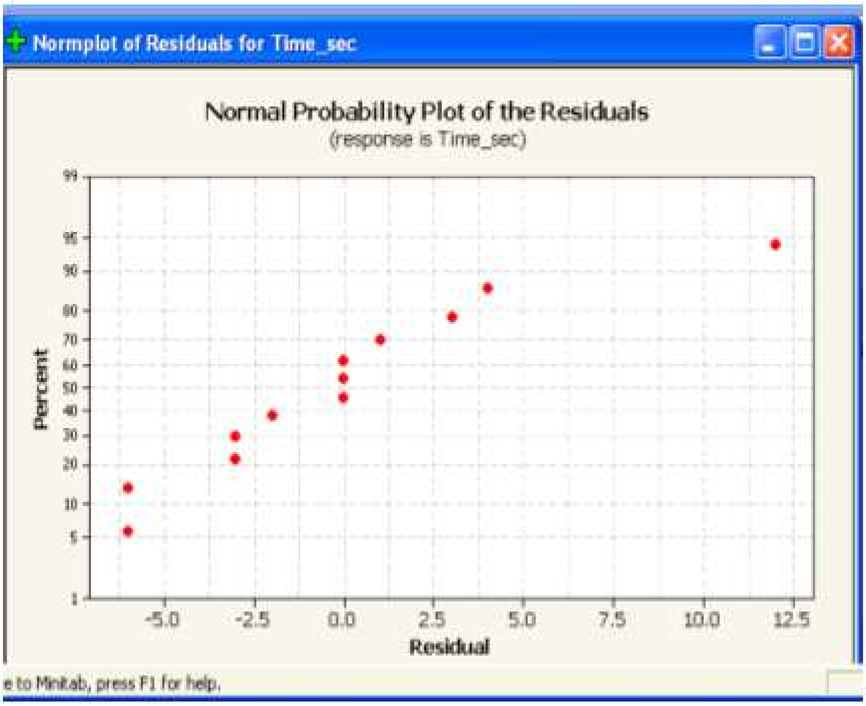
3 **-** Click **OK** en cada cuadro de diálogo.

Interpretando sus resultados
La grafica de probabilidad normal
Usa la grafica de probabilidad normal de la residual para verificar que tu residual no este desviado sustancialmente de la distribución normal.

- Si la residual viene de la distribución normal, los puntos seguirán una línea recta.
- Si la residual no viene de la distribución normal, los puntos no seguirán una línea recta.

Basado en esta grafica, es razonable asumir que la residual de los datos de CRT no están desviados sustancialmente de la

distribución normal.
Como notaste previamente, hay una línea de fuera en el conjunto de datos. Tu debes investigar la línea de fuera para determinar que fue lo que la hizo posible.
Alternativas
Tu también puedes usar un histolograma de la residual para evaluar la normalidad. Sin embargo la grafica de probabilidad normal es generalmente fácil de interpretar, especialmente para muestras pequeñas.

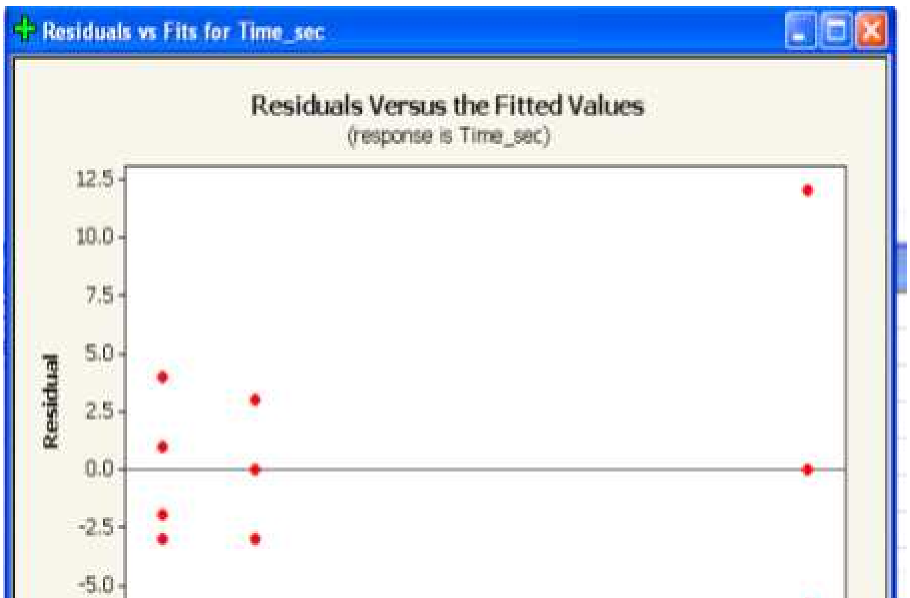


Interpretando sus resultados
Residuales contra Fits
Use la grafica de la residual versus las fits para verificar que las siguientes suposiciones han sido encontradas:
- Variaciones constantes a través de la combinación de todos los factores.
- No están fuera de línea los datos.

Si tú ves cualquier tipo de patrón en la grafica, una de estas suposiciones encontradas han sido violada. La tabla abajo resume los tipos de patrones que tú puedes ver

Los patrones:	**Indica…**
La extensión desigual de las residuales a través de los diferentes valores ajustados.	La variación de tu residual no es constante.
Un punto está situado muy lejos del cero.	Fuera de línea.

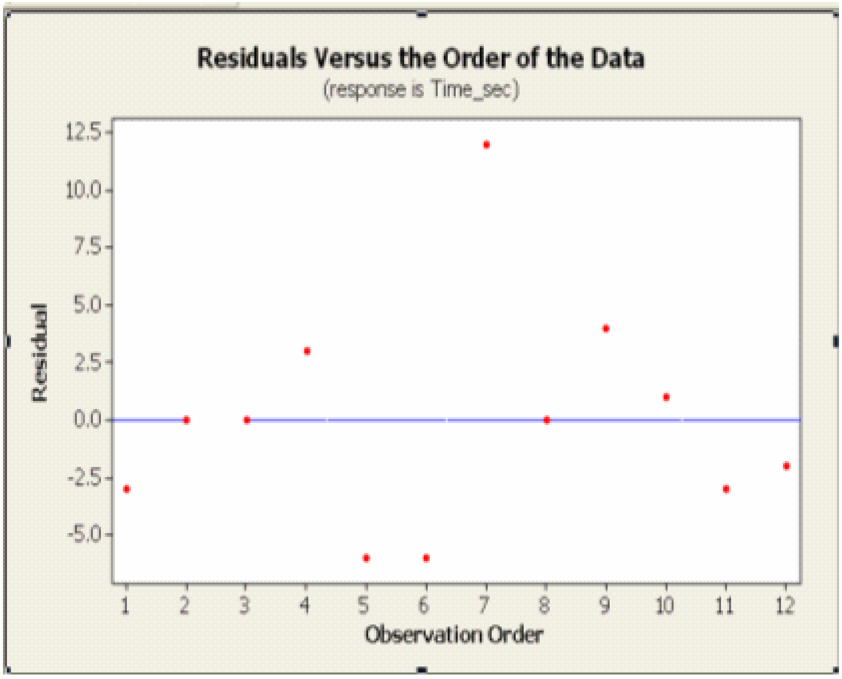
Hay un residual extraordinariamente alto da que la apariencia de una variación no constante. Tu debes poder determinar que causo esta línea de fuera. Tal vez es apropiado volver a analizar los datos sin esta línea. Sin embargo tu solamente deberías remover la observación para estabilizarla, sin puede establecer que no era representativo de la población.



Interpretando sus resultados
La residual versus el orden
Utiliza la grafica de la residual contra el orden para verificar que la residual es independiente.

- Si hay un efecto debido al orden de la recolección de los datos, los residuos no se esparcirán aleatoriamente cerca del cero. Tu debes ser capaz de detectar este patrón en la grafica.
- Si hay un efecto debido al orden de la recolección de los datos, la residual esta aleatoriamente cerca del cero.

La grafica revela la misma fuera de línea identificada en la grafica de residual contra el Fitted values plot.
Ignorando la línea de fuera por un momento, hay dos valores muy bajos que ocurren uno después del otro.
Tal vez en alguna causa especial causo que la recolección de los tubos y las pruebas del tiempo de calentamiento fueran mas rápidamente que los otros tubos.
Tal vez valga la pena investigar.
Puede haber también evidencia de un aumento sistemático en el precalentamiento para los primeros cuatro tubos probados.



Consideraciones finales
Conclusiones prácticas
El análisis los tubos de rayo catódico no revelaron ninguna diferencia en el tiempo de calentamiento. Sin embargo existen problemas potenciales con el estudio:

- Fuera de línea--- Un valor aparece fuera de línea y debe ser investigado. Puede haber también evidencia de un incremento sistemático en el tiempo de calentamiento de los primeros cuatro tubos probados.
- Los Patrones--- dos tubos consecutivos tienen más corto el tiempo de calentamiento que lo normal.
- Bajo Power--- basado en una estimación de un 5.416, el power de la prueba para descubrir una diferencia de 7 segundos (a los 0.05 nivel) es solamente de 0.2642 . Esto es está menos de un 27% de oportunidad para descubrir una diferencia. De hecho en el orden tu tienes un power de 0.80, y deberías tener una diferencia de .80 segundos/

Basado en estos resultados, quizás el curso mejor de acción sería asegurar el proceso bajo control y recolectar muestras grandes y realizar la prueba nuevamente. Con una sigma reducida de 3.0 y una recolección de 6 muestras de tubos de cada lote, tu puedes detectar una diferencia de 7 segundo con un power de 0.9133

Consideraciones estadísticas
Comparando el nivel del factor múltiple con un solo ANOVA es preferible a hacer una comparación de dos niveles del tiempo con dos muestras separadas Esto es porque dirigiendo los aumentos de las pruebas extras incrementa tu posibilidad de error tipo 1 (rechazando Ho, cuando Ho es verdadera.
Las suposiciones de la independiente son criticas. Si las observaciones son afectadas sistemáticamente por otros factores que el que usted este estudiando los resultados de este ANOVA son sin sentido.
La suposición de la normalidad no es generalmente crucial especialmente en muestras grandes.

ANÁLISIS DE LA MEDIA
Ejemplo 2 CRT Revisión de tiempo de calentamiento
Problema
En el ejemplo 1 (página 4-3) realizaste una prueba de tres lotes de tubo de rayo catódico para determinar si los periodos de calentamiento son consistentes.
Colección de datos:
Una muestra al azar de los cuatro tubos es tomada de cada grupo y probadas para determinar el tiempo de calentamiento (Los datos de la muestra son tomados del ejemplo 1, Pagina 4-3).
Herramientas:

Stat > ANOVA >Analysis of Means

Set de Datos
CRT.MPJ

Nombre	Tipo de dato	Tipo de Variable	Niveles
Tubetype	Numérico	Factor	1,2,3
Time_sec	Numérico	Respuesta	

Análisis de la media
¿Qué es un análisis de la media?

No es nada parecido que el ANOVA, el cual es usado para determinar si el nivel de la media difiere de algún otro, el análisis de la media (ANOM) es usado para determinar si el nivel de la media es diferente de la gran media.

La gran media es la media de todas las observaciones sin tener en cuenta el nivel. Por ejemplo: si tú tienes cuatro observaciones por cada 3 niveles en el factor, la gran media es la suma de las 12 observaciones divididas entre 12.
El resultado de un análisis de la media es usualmente similar al obtenido con el ANOVA. Sin embargo:

- El análisis de la media es generalmente más sensible que el ANOVA, cuando un nivel de la media es difiere del resto.
- ANOVA es generalmente más sensible que el análisis de la media cuando los niveles de los grupos de la media son diferentes cada uno de los otros.

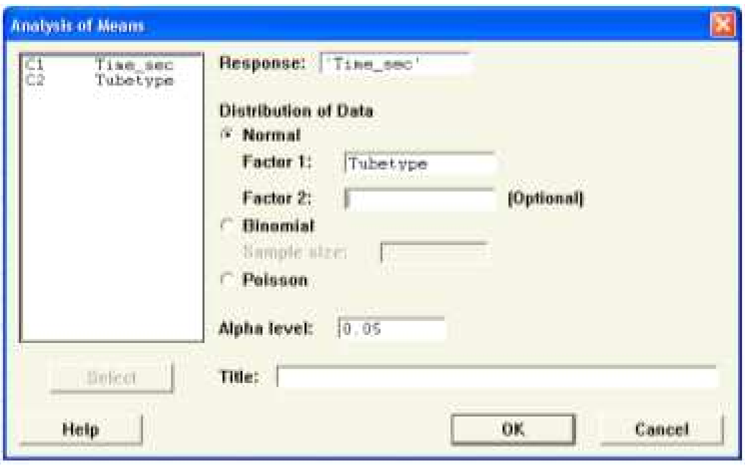
<u>¿Cuándo usar el análisis de la media?</u>
Usa el análisis de la media cuando tengas datos de uno o dos factores. Los datos deben de proceder de la distribución normal, Binomial o Distribución de Poisson.
<u>¿Porqué usar el análisis de la media?</u>
El análisis de la media te ayuda a responder preguntas tales como:

- ¿Es un tratamiento mejor que el promedio?

Por ejemplo:

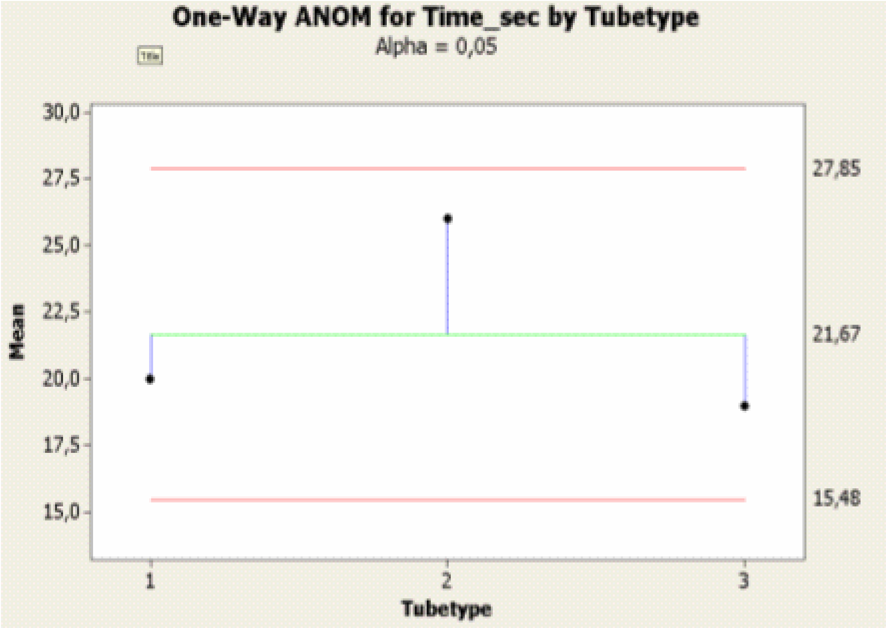
- Entre tablas tratadas con diferentes acabados, un acabado confiere mejor que el promedio de las otras características utilizadas.

Desarrollo del análisis de la media
Use el análisis de la media para evaluar los datos. Use un ∞ de 0.05 para la prueba
Análisis de la media:
1. - Elija: **Stat > ANOVA >Analysis of Means.**
2.- Complete el recuadro como se indica a continuación



3.- Presione **OK**

Interpretando tus resultados
El análisis de la media comparado con el nivel de la media (puntos negros) con la gran media (línea central verde. La gran media es la media de las 12 observaciones.
Si un nivel de la media es mayor o menor que el valor critico representado por la línea decisiva (en rojo) esta es significativamente diferente de la gran media.
Conclusiones:
Ninguno de los niveles individuales de la media de los datos de la CRT son significativamente diferentes de la gran media.



Consideraciones finales
Consideraciones prácticas:
Diferencias no significativas fueron detectadas en medio del nivel del factor de la media y de la gran media.
Sin embargo recordemos que cuando los mismos datos fueron analizados únicamente con el procedimiento de ANOVA se observaron algunos problemas con el control del proceso. La mayor desventaja del procedimiento de Análisis de la media es que no suministra graficas de residual para ayudarte ubicar problemas en los datos.

Consideraciones estadísticas
Los resultados del análisis de la media son usualmente similares a los obtenidos con ANOVA, Sin embargo:

- El análisis de la media es generalmente más sensible que el ANOVA, cuando un nivel de la media es diferente del resto.
- ANOVA es generalmente más sensible que el análisis de la media cuando los niveles de los grupos de la media son diferentes cada uno de los otros.

Si bien usando ANOVA con dos factores, el diseño debe tener balance (esto es, debe tener los mismos números de observaciones por cada combinación de niveles de factor.
Si bien usando ANOVA con datos binomiales, np y n(1-p) deben los dos tener por lo menos cinco, porque ANOVA usa una aproximación normal de la binomial (n es el tamaño de muestra que deben tener todas las muestras, y P es generalmente proporcional.
Para datos de Poisson, la media debe ser por lo menos de 5.

Balance ANOVA
Ejemplo 3 Desgaste de pintura
Problema
Estas estudiando las características que deben llevar los cuatros diferentes tipos de pintura amarilla para carreteras.
Recolección de datos:
Pruebas de restos de cada pintura que fueron aplicadas en las carreteras de Filadelfia, Pittsburg, Harrisburg y Scranton, Pensilvania.
Después de un tiempo conveniente de exposición al mal tiempo y al tráfico, el desgaste de la pintura fue medido en cada una de las cuatro localizaciones. Un alto puntaje de la media arrojo que la pintura estaba erosionada.

Nombre	Tipo Dato	Tipo Variable	Niveles
Locacion	Texto	Factor	Philadelphia
			Pittsburg
			Harrisburg
			Scranton
Pintura	Texto	Factor	Y-0242
			Y-0314
			Y-1723
			Y-1424
PntWear	Numérico	Response	

Herramientas:
Stat > Anova-One-way.
Graph > Chart.
Stat-ANOVA > Balanced ANOVA.

Set de Datos
PNTWEAR.MPJ

Balanced ANOVA
¿Qué es el balance ANOVA?
Balance ANOVA es similar al ANOVA, excepto que las respuestas pueden ser clasificadas en dos o más factores al mismo tiempo. Por ejemplo: La tabla de abajo es el diseño de un estudio para evaluar si la temperatura del motor funcionando es un factor de ambos: Peso del aceite y las revoluciones por minuto del motor.

	RPM		
Aceite	1,000	2,000	3,000
5w30	RPM = 1,000 Aceite = 5w30	RPM = 2,000 Aceite = 5w30	RPM = 3,000 Aceite = 5w30
10w30	RPM = 1,000 Aceite =10w30	RPM = 2,000 Aceite =10w30	RPM = 3,000 Aceite = 10w30

La información de cada celda de la tabla representa la única combinación de RPM y aceite. Las observaciones son clasificadas en las dos variables.
¿Cuándo usar Balance ANOVA?
Usa Balance ANOVA cuando tengas respuestas continuas de datos fijos en uno o más factores. Los datos deben estar balanceados, Esto es, “deben tener los mismos números de observaciones en cada celda del diseño.
Antes de aceptar los resultados en el ANOVA, debes verificar las siguientes suposiciones acerca del residual y validar los resultados.

- La residual debe ser independiente (y ser la azar).
- La residual no tiene una desviación sustancial de la distribución normal.
- La residual debe tener constantes variaciones a través de los niveles de factor.

¿Por qué que usar Balance ANOVA?
Balance ANOVA te puede ayudar a responder preguntas tales como:

- ¿Hay diferencias en tus productos a causa de varios factores identificados?
- ¿Son ciertas combinaciones de los niveles de factor ideales?

Por ejemplo,

- ¿El nivel de temperatura del motor funcionando cambia por el factor de RPM o Peso del aceite?.
- ¿Existen ciertos cambios de maquinas en tus planta que son mas productivos que otros, o es una maquina mas productiva en ciertos cambios, pero no en otras?

Ejecutando One Way ANOVA
Para una comparación, primero ejecuta one-way ANOVA, para analizar el desgaste de la pintura como un factor del tipo de pintura solamente.

One-Way ANOVA
1.- Abre el proyecto PNTWEAR.MPJ.
2.- Escoge **Stat > ANOVA-ONE WAY .**
3.- En **Respuesta** enter PntWear.
4.- En **Factor** enter Paint.
5.- Click **OK.**

Interpretando tus resultados
Aunque la mejor pintura es la de Y-0242 (media = 14.25) y la peor de Y-1725m (media = 10.75) estas diferencias fueron no significativas “en el nivel 0.05 ∞ (p =0.115).

One-way ANOVA: PntWear versus Paint

```
Source  DF     SS     MS     F      P
Paint    3  30,69  10,23  2,44  0,115
Error   12  50,25   4,19
Total   15  80,94

S = 2,046   R-Sq = 37,92%   R-Sq(adj) = 22,39%

                               Individual 95% CIs For Mean Based on
                               Pooled StDev
Level   N    Mean  StDev  -------+---------+---------+---------+--
Y-0242  4  14,250  1,893                  (----------*----------)
Y-0314  4  12,250  2,872         (----------*----------)
Y-1424  4  11,000  1,633   (----------*----------)
Y-1723  4  10,750  1,500  (----------*----------)
                          -------+---------+---------+---------+--
                              10,0      12,0      14,0      16,0

Pooled StDev = 2,046
```

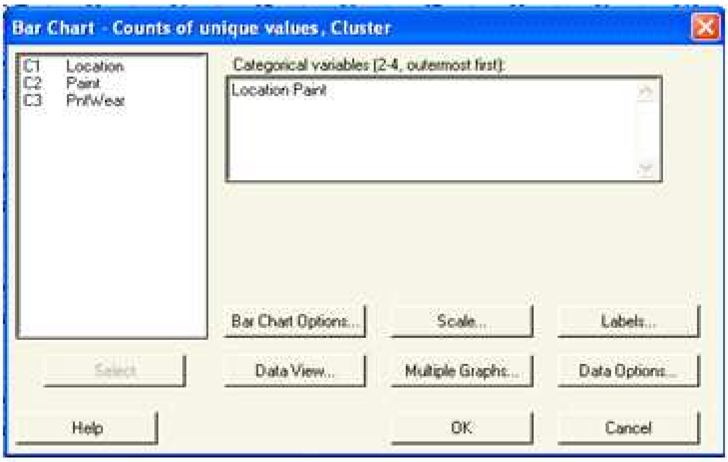
Graficando los datos

Información importante se pierde cuando no incluyes el factor de localización en tu análisis. Para ilustrar esto, usa el chart para crear unas barras agrupadas ilustrando el desgaste de la pintura en función de ambos tipos: de pintura y locacion.

Chart

1.- Escoge **Graph > Chart.**

2.- Complete el recuadro como se indica a continuación.



* Asegúrate de cambiar para cada uno de **Graph** a **Group.**

3.- Click: **Options**.

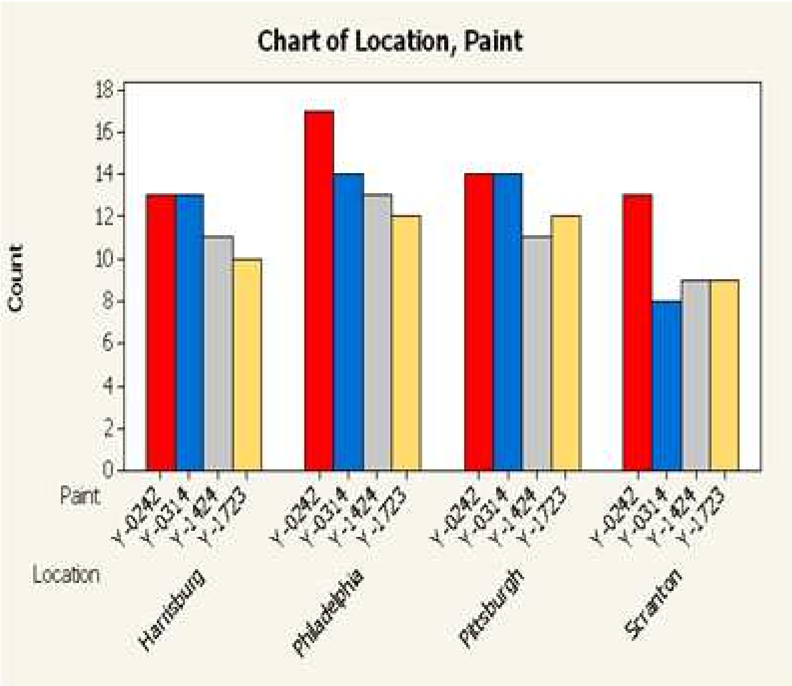
4.- Revisa el **Group** y enter Paint.

5.- Click **OK** en cada recuadro.

Interpretando tus resultados

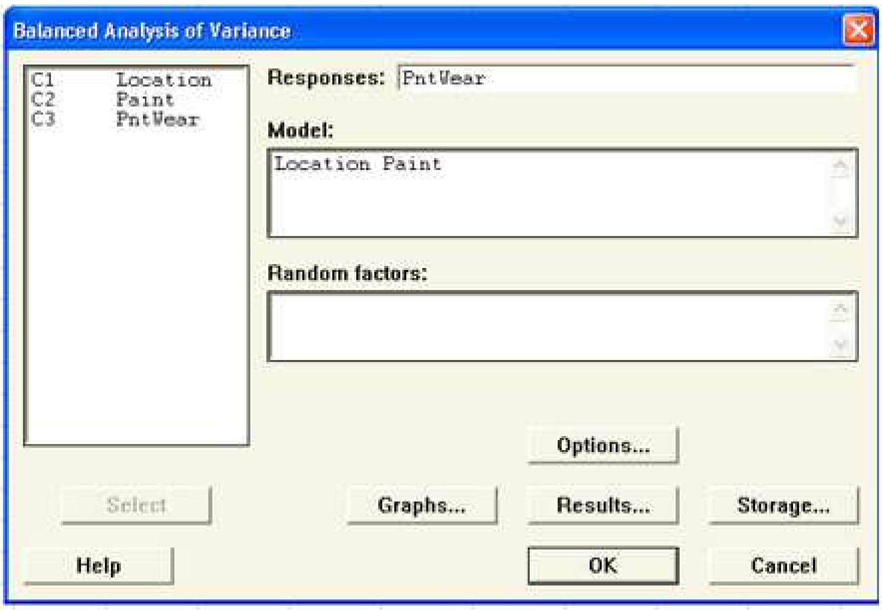
La grafica muestra una gran cantidad de variabilidad en los datos asociados con Localización. En general, el menor desgaste de pintura se presenta en Philadelphia y la peor en Scranton.

Cuando tu no incluyes “Localización” en tu análisis, esta variabilidad es atribuida al error (dentro del grupo con variación). Esto disminuye tu F ratio.



Usando una segunda variable del bloque de variación de Locación.
Incluyendo locación en tu análisis quieres prevenir la variabilidad asociada con este factor, donde se les atribuye el error. Esto incrementara tu habilidad para detectar diferencias en desgaste de cada pintura.
Usa Balance ANOVA para analizar si el desgaste de pintura es un factor de ambos Pintura y Localización. Esto es llamado: Diseño aleatorio del bloque, porque el factor de interés es la pintura, y Localización es incluido solamente para reducir el error de variabilidad, Tal factor es llamado "Variable de bloque".
La respuesta debe estar en una columna con descripciones y columnas adicionales indicando los niveles de cada factor por cada observación.

Balance ANOVA
1.- Escoge **Stat > ANOVA > Balance ANOVA.**
2.- Complete el recuadro como se indica a continuación



3.- Click **OK.**
Interpretando tus resultados
En este modelo el efecto de ambos: Pintura y Localización son significantes en un nivel 0.05 ∞ (p = 0.003 y 0.007 respectivamente).
Note que el MS del error (también llamado MSE) en este modelo es solamente 1.285, comparado con el 4.190 del modelo que no incluyo localización como factor, porque el MSE es el denominador de todos los F-ratios, reduciendo el MSE, incrementa el F-Values.

ANOVA: PntWear versus Location; Paint

Factor	Type	Levels	Values
Location	fixed	4	Harrisburg; Philadelphia; Pittsburgh; Scrant
Paint	fixed	4	Y-0242; Y-0314; Y-1424; Y-1723

Analysis of Variance for PntWear

Source	DF	SS	MS	F	P
Location	3	38,688	12,896	10,04	0,003
Paint	3	30,688	10,229	7,96	0,007
Error	9	11,563	1,285		
Total	15	80,938			

S = 1,13346 R-Sq = 85,71% R-Sq(adj) = 76,19%

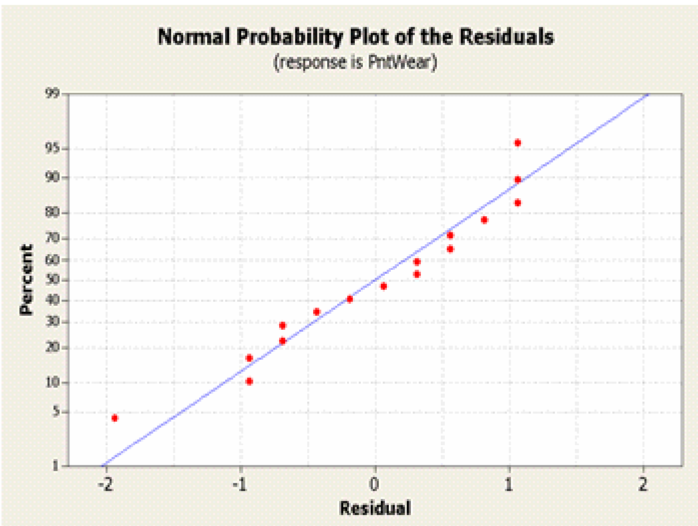
Analizando el residual
Después de que confíes en los resultados de ANOVA, debes revisar, para estar seguro de todas las suposiciones acerca del residual que hayas encontrado.
Usa Balance ANOVA para crear una tabla de Residual.

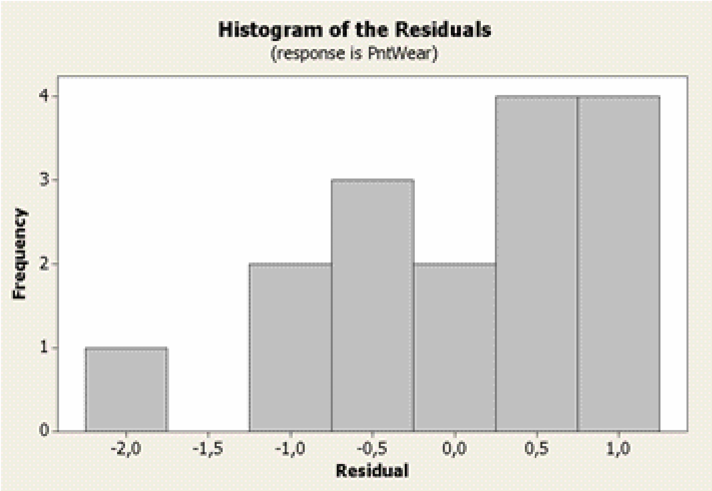
Balanced ANOVA
1.- Escoge **Stat > ANOVA > Balanced ANOVA** y presiona Ctrl + E para regresar **a Balance Analysis of Variance** de los recuadros.
2.- Click **Graphs.**
3.- Marca las cuatro opciones bajo **Residual Plots.**
4.- Click **OK** en cada recuadro.

Interpretando tus resultados
Usa un cuadro de distribución normal y un histolograma para verificar que tu residual no este desviado considerablemente de la distribución normal.

- Si la residual viene desde la distribución normal, los puntos tienen una línea recta en la grafica normal y en el histolograma no tiene una forma de campana recta.
- Si el residual no viene desde una distribución normal, los puntos no continúan en línea recta en la grafica Normal y en el histolograma no tiene forma de campana.

Basado en los recuadros, es razonable asumir que la residual no esta desviada sustanciablemente de la distribución normal. (Normalmente pruebas ejecutadas del residual (no presentadas en este ejemplo) produce que p = valor menos de 0.423.)





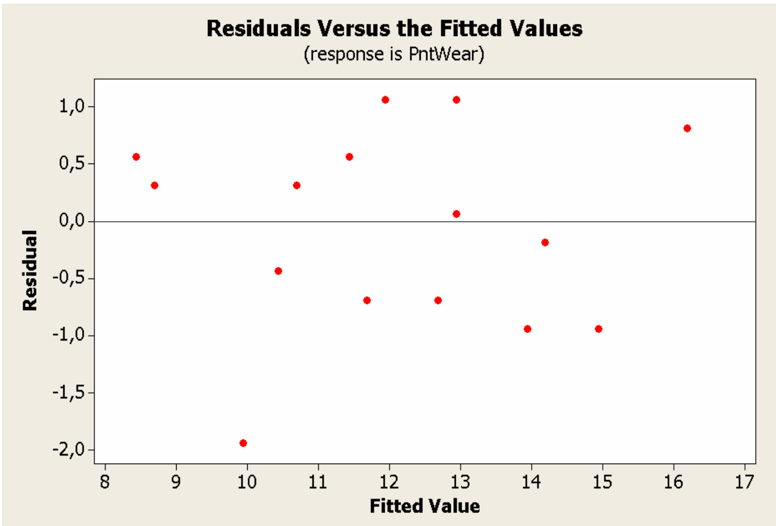
Interpretando tus resultados

Usa las graficas de residual contra el ajuste para verificar que las siguientes suposiciones han sido encontradas:

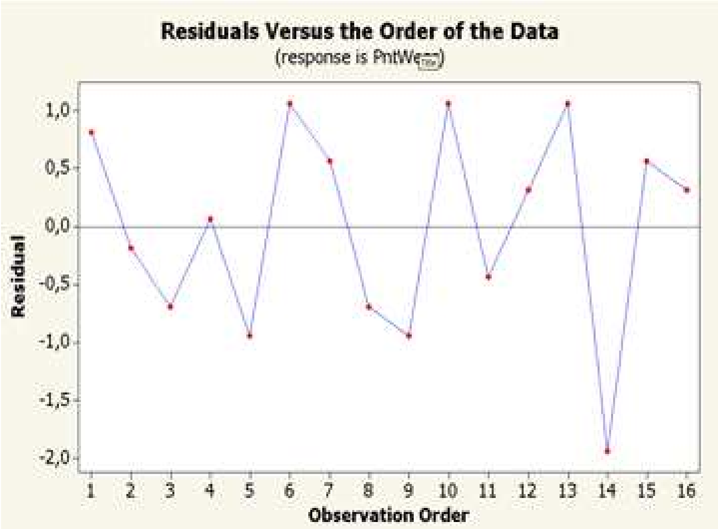
- Variaciones constantes a través de la combinación de todos los factores.
- No están fuera de línea los datos.

Si tú ves cualquier tipo de patrón en la grafica, una de estas suposiciones encontradas has sido violada. La tabla abajo resume los tipos de patrones que tú puedes ver:

Los Patrones	Indican………
La extensión desigual de las residuales a través de los diferentes valores ajustados.	La variación de tu residual no es constante.
Un punto está situado muy lejos del cero.	Fuera de línea.



Para los datos del desgaste de pintura, la varianza de la residual aparece bastante constante.



Interpretando tus resultados
Usa la grafica de la residual contra el orden de los datos para verificar que las residuales son independientes.

- Si hay algún efecto debido a la recolección de los datos, la residual no será aleatoriamente dispersa cerca del cero. Tú serás capaz de detectar este patrón en la grafica.
- Si no hay algún efecto debido al orden de la recolección de datos las residuales serán aleatoriamente dispersas cerca del cero.

No hay ningún patrón que aparezca en las residuales, por lo tanto las suposiciones de independencia son validas.

Consideraciones finales
Conclusiones prácticas:
Inicialmente un simple análisis de los tipos de pintura no revelo diferencias significativas. Sin embargo cuando Locación fue incluida en el análisis de un bloque de variable, efectos significativos de la pintura fueron revelados.
Esto ilustra una de las ventajas del modelo de Multi-Factor ANOVA:

- Tú puedes algunas veces disminuir la cantidad de un error inexplicable de variación en la respuesta, si incluyes factores adicionales en el modelo.
- Tú puedes algunas veces aprovechar recursos investigando dos o más factores en el mismo tiempo.
- Con un modelo apropiado, tú puedes evaluar interacciones entre dos factores. Esta interacción ocurre cuando los efectos de un factor cambian basándose en el nivel de algún otro factor.

Consideraciones estadísticas:
Para usar Balance ANOVA tus datos deben estar balanceados.
Es importante validar las suposiciones de la residual después dibujando las conclusiones finales de los resultados de cualquier ANOVA.

Modelo Lineal General
Ejemplo 4 Distancia de frenado.
Problema:
Tú quieres conocer si la distancia que le toma a un carro para frenar en un pavimento mojado es afectada por los siguientes factores:

- El Modelo de la llanta (Llanta).
- La banda de rodadura (Tread).
- Si el seguro de freno es adecuado (ABS).

Colección de datos
El mismo carro es utilizado para la recolección de los datos. La distancia requerida para frenar a una velocidad de 40 millas por hora en un pavimento mojado fue medida con cada combinación de llantas, banda de rodadura y el ABS. Corridas experimentales fueron realizadas en orden aleatorio.

Herramientas:
Stat-Table > Cross Tabulation.

Stat-ANOVA > Balanced ANOVA.

Stat-ANOVA > General Linear Model.

Set de datos
BRAKEDIS.MPJ

Nombre	Tipo de datos	Tipo de variable	Niveles
Llanta	Texto	Factor	MX,GT,LS

Banda de rodadura	Numérico	Factor	10, 1.5
ABS	Texto	Factor	Capaz No capaz
Distancia	Numérico	Respuesta	

Modelo lineal General

¿Qué es el Modelo lineal general (GLM)?

GLM es similar al Balance ANOVA excepto que:

- Puede ser usado en pruebas de diseño no balanceadas.
- Puede ser usado también para comparar los niveles individuales de las medias.

¿Cuándo usar el Modelo lineal general?

Usa el GLM cuando tengas respuesta continúas de datos de niveles estables de uno o más factores. El diseño no debe ser balanceado.

Después de aceptar los resultados del ANOVA, tú debes verificar que las siguientes suposiciones acerca de la residual sean validas en tus datos:

- La residual debe ser independiente (y de esta manera al azar).
- La residual no debe estar desviada sustanciablemente de la distribución normal.

La residual debe tener constantes variaciones a través de todos los niveles de factor.

¿Por qué usar el Modelo lineal general?

GLM te ayuda a responder preguntas tales como:

- ¿Hay diferencia en tu producto debido a diferentes factores identificados?
- ¿Son ciertas combinaciones de niveles de factor ideales?

Por ejemplo,

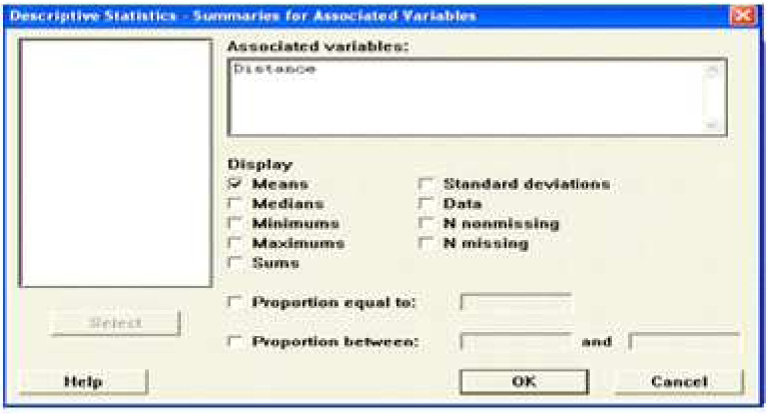
- ¿El color de tu plástico cambia en función de la temperatura, humedad o presión?
- ¿Es el color del plástico generalmente mejor cuando la presión es alta, o depende este de la presión?

Observando los datos en la tabla

Usa CROSS tabulación para crear una tabla con los datos para ver posibles diferencias entre doce combinaciones tratadas.

Cross tabulación

1.- Abre el proyecto BRAKEDIS.MPJ.
2.- Escoge **Stat > Table-Statistics descriptive.**
3.- En **Classification variables** enter Tire Tread ABS.
4.- En **Asociated variables** enter Distancia.
5.- Complete el recuadro como se indica a continuación.



6.- Click **OK** en cada recuadro.

Interpretando tus resultados

```
Rows: Tire    Columns: Tread

           1.5    10.0     All

GT       68.50   67.00   67.75
             2       2       4

LS       82.50   78.00   80.25
             2       2       4

MX       76.00   76.50   76.25
             2       2       4

All      75.67   73.83   74.75
             6       6      12

Cell Contents:   Distance   :   Mean
                                Count
```

requiere para frenar con ABS (74.750) la diferencia es llamada ***Principales efectos* del ABS.**

Efectos de interacción

Note que:

- Cuando el ABS fue no utilizado, el 1.5 mm de la banda de rodadura frena mas rápidamente (media = 90.16) que el 10.mm de banda de rodadura (media =91.167).
- Cuando el ABS fue utilizado, el 10.0 mm de banda de rodadura (media = 73.833) frena mas rápidamente que el 1.5 mm de la banda de rodadura"(media = 75.667).

Los efectos son llamados *interacción* del ABS by tread, porque los efectos de la banda de rodadura dependen del nivel del ABS. (Sin embargo las diferencias no son grandes. Una prueba apropiada revelara que la interacción no es significante.

```
Tabulated statistics: Tire. Tread. ABS

Results for ABS = disabled

Rows: Tire    Columns: Tread

          1.5     10.0     All

GT       83.00    90.00    86.50
             2        2        4

LS       96.50    90.50    93.50
             2        2        4

MX       91.00    93.00    92.00
             2        2        4

All      90.17    91.17    90.67
             6        6       12

Cell Contents:    Distance   :   Mean
                                 Count
```

Analizando tu modelo completo

Usa el balance ANOVA para analizar el modelo completo. El modelo completo contiene todos los posibles efectos e interacciones.

Principales efectos	Interacción en dos direcciones	Interacción en tres direcciones
Llanta	Llanta*Tread	Llanta*Tread*ABS
Tread	Llanta*ABS	
ABS	Tread*ABS	

Anotaciones

Para indicar los términos de interacción, simplemente junta los nombre con un asterisco. De esta manera el modelo completo de frenado a distancia de los datos contiene los siguientes términos:

Como un atajo el modelo completo puede ser usado también cargando los datos en barra vertical.

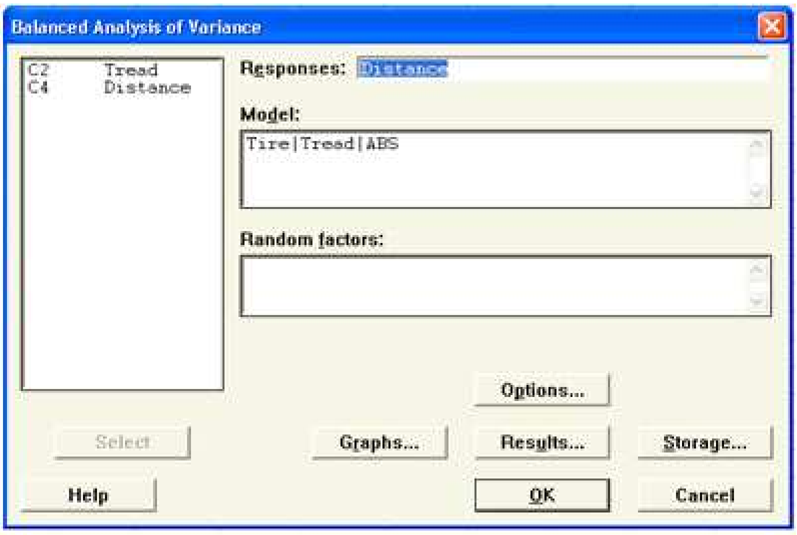
Tire Tread I ABS

Las barras verticales le dicen al minitab que considere los efectos principales e interacciones contenidas en medio de los términos indicados.

Balance ANOVA

1.- Escoge **Stat > ANOVA > Balanced ANOVA**.

2.-Complete el recuadro como se indica a continuación



3.- Click **OK**.

Interpretando tus resultados

Usa el Valor de P, para la prueba de significancía de cada término. En este modelo los siguientes efectos son significantes en un

0.05 ∞.

- Llanta (p = 0.003).
- ABS (p = 0.000).

Porque hubo solamente dos niveles del ABS (capaz y no capaz), tú sabes que la significancía en los términos refleja diferencias significativas entre los dos niveles.

Prueba de comparación

Porque hubo tres niveles en las llantas, tu quieres realizar una comparación estadística en orden para determinar cuales niveles son diferentes entre cada uno.

El procedimiento Balanced ANOVA no permite comparaciones individuales entre niveles de la media, por eso vas a usar el Modelo lineal general para dirigir las pruebas después de eliminar las no significancías y revisar el residual.

ANOVA: Distance versus Tire; Tread; ABS

Factor	Type	Levels	Values
Tire	fixed	3	GT; LS; MX
Tread	fixed	2	1,5; 10,0
ABS	fixed	2	disabled; enabled

Analysis of Variance for Distance

Source	DF	SS	MS	F	
Tire	2	404,33	202,17	9,68	0,00
Tread	1	1,04	1,04	0,05	0,82
ABS	1	1520,04	1520,04	72,82	0,00
Tire*Tread	2	72,33	36,17	1,73	0,21
Tire*ABS	2	30,33	15,17	0,73	0,50
Tread*ABS	1	12,04	12,04	0,58	0,46
Tire*Tread*ABS	2	26,33	13,17	0,63	0,54
Error	12	250,50	20,88		
Total	23	2316,96			

Ajustando el modelo reducido

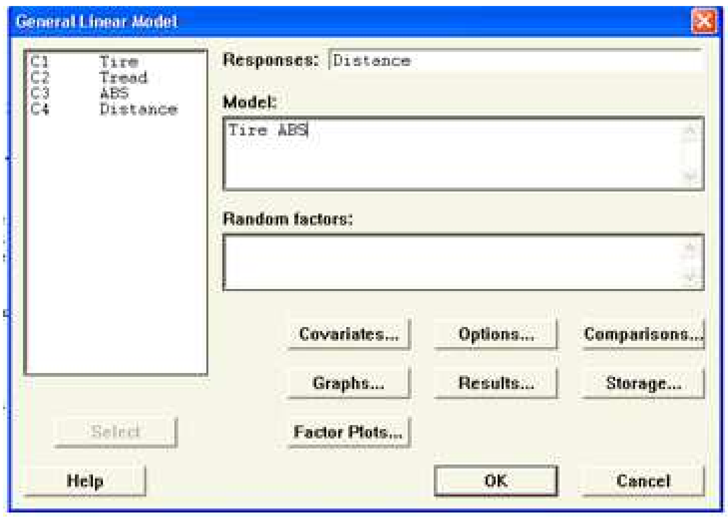
El paso siguiente para ajustar el modelo reducido es removiendo los términos no significantes. Usa el modelo lineal general para ajustar el modelo con solo la llanta y el ABS. (El modelo proporciona opciones para analizar las diferencias entre niveles individuales de la media).

Tú debes también crear graficas de residual, para poder validar las suposiciones de la prueba.

Modelo Lineal General

1.- Escoge Stat > ANOVA > General Linear Model.

2.- Complete el recuadro como se indica a continuación.



3.- Click Graphs.

4.- Revisa las opciones bajo **Residual Plots.**

5.- Click **OK** en cada recuadro.

Interpretando tus resultados

Como es esperado, ambos: Llanta y ABS, tengan una significancía de 0.005 ∞ del nivel en el modelo ajustado.

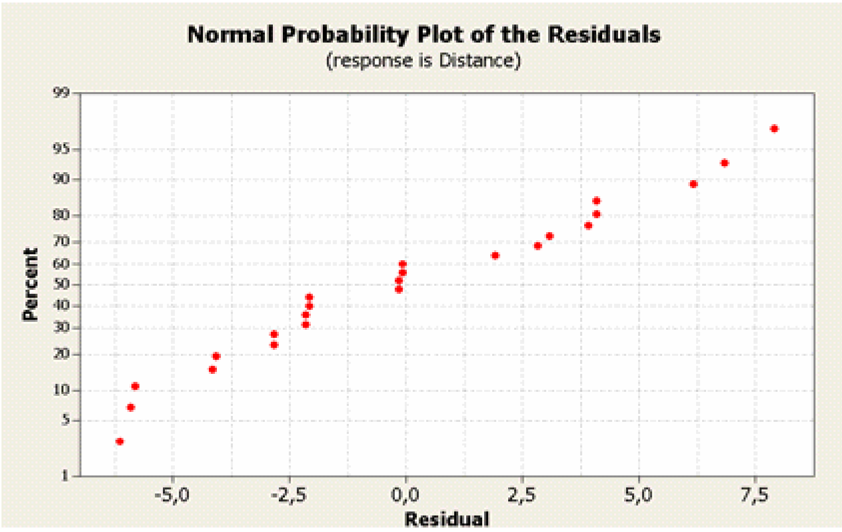
General Linear Model: Distance versus Tire; ABS

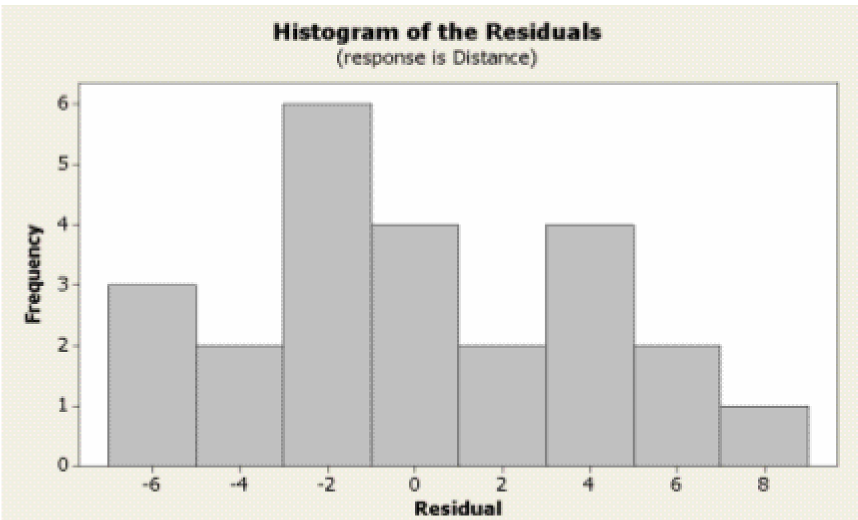
Factor	Type	Levels	Values
Tire	fixed	3	GT; LS; MX
ABS	fixed	2	disabled; enabled

Analysis of Variance for Distance, using Adjusted SS f

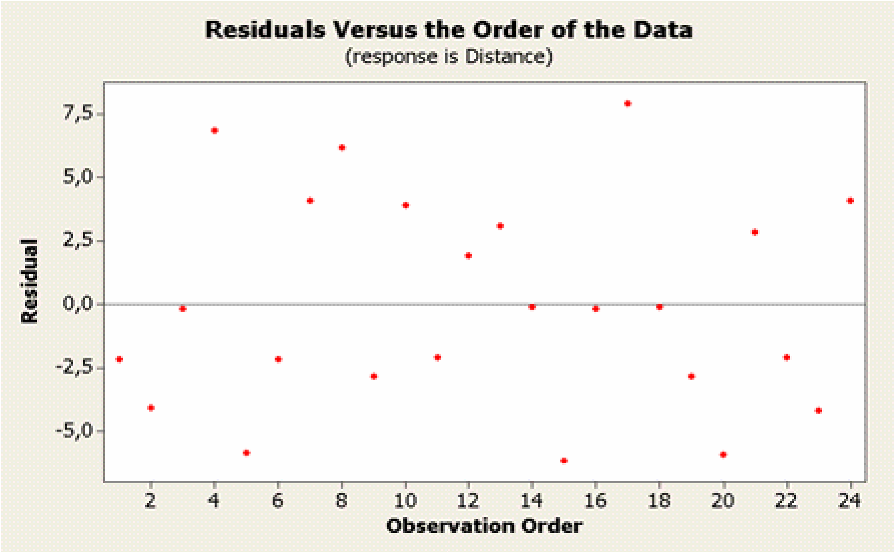
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Tire	2	404,33	404,33	202,17	10,30	0,001
ABS	1	1520,04	1520,04	1520,04	77,44	0,000
Error	20	392,58	392,58	19,63		
Total	23	2316,96				

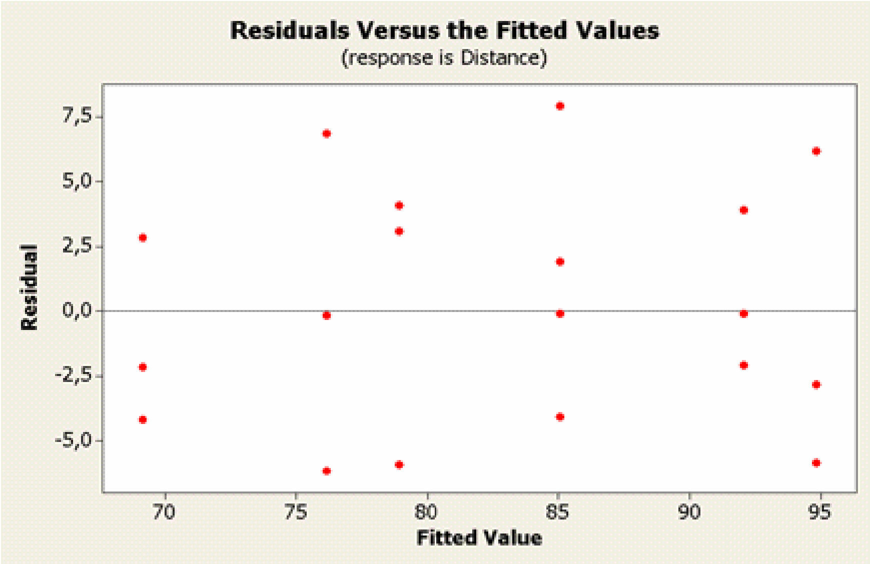
Interpretando tus resultados
Basado en la grafica de probabilidad normal y en el histolograma es razonable asumir que el residual no esta desviado sustanciablemente de la distribución normal.





Interpretando tus resultados
La grafica de la residual contra el ajuste, confirma que la variación de la residual es bastante constante.
La grafica de la residual contra el orden de los datos no presenta ningún patrón, de esa manera las suposiciones de independencia parecen ser validas.
No parece haber ninguno fuera de línea.





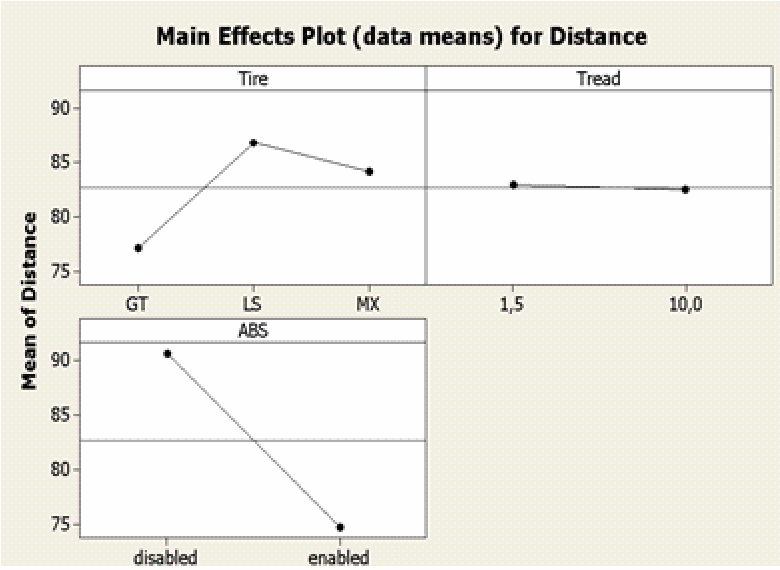
Graficando los principales efectos e interacciones
Comparación gráfica de la media
Ahora que tú te has establecido en un modelo, este es usado para visualizar los resultados del análisis de la media utilizando las graficas y los principales efectos e interacciones.
Aunque tú has elegido no incluir todos los términos del modelo final, es de mucha ayuda incluir todos los factores en las graficas para visualizar términos significantes y no significantes.

Principales efectos e interacciones graficas
1.- Elige **Stat > ANOVA > Main Effects Plot.**
2.- En **Response** presiona **Distance.**
3.- En **Factors** presiona Tire Tread ABS.
4.- Presiona **OK.**
5.- Elige **Stat > ANOVA > Interactions Plot.**
6.- En **Response** presiona Distance.
7.- En **Factors** presiona Tire Tread ABS.
8.- Presiona **OK.**

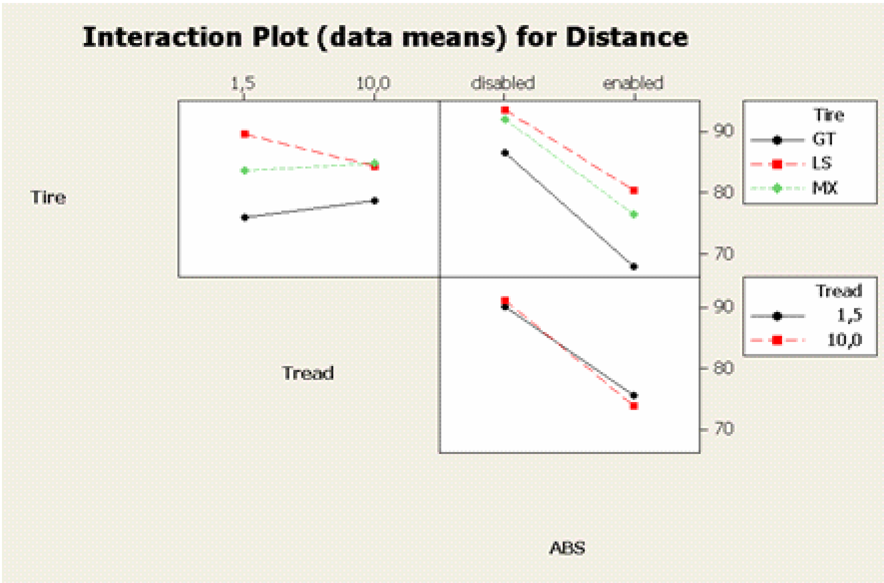
Interpretando tus resultados
La gráfica de efectos principales de llanta y ABS tienen líneas inclinadas, indicando que estos efectos son importantes.
En esta gráfica esta claro que:

- La Llanta con la distancia mas corta de frenado fue la GT.
- Las distancias de frenado mas cortas fueron llevadas a cabo con el sistema ABS.

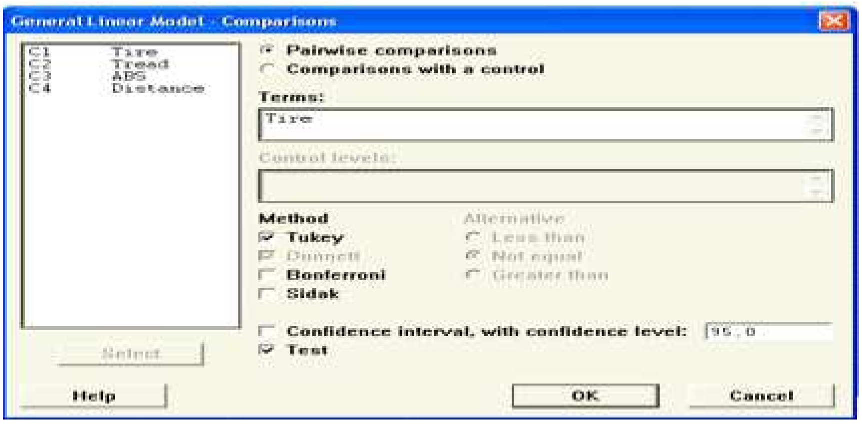
La grafica de la banda de rodadura muestra una pequeña inclinación sugiriendo que los efectos no son significativos. Las bandas de rodadura producen casi idéntica frenada a distancia.



Interpretando tus resultados
Las interacciones de la grafica muestran los dos sentidos de la interacción. Los nombre de los factores a la izquierda de cada grafica son representados en el eje Y, y todos los factores llamados por debajo de cada grafica son representados en el eje X. Las líneas de cada grafica en su mayor parte paralelas, sugieren que no hay interacción en medio de ningún término. Pero es quizás alguna evidencia de una interacción entre las llantas y la banda de rodadura pero el ANOVA índico que esto no fue significante.



Conduciendo la comparación del desgaste de pintura
Utilice el modelo lineal para probar las diferencias entre los niveles de factores significantes.



General Linear Model
1.- Elija **Star > ANOVA > General Linear Model.**
2.- Presione F3 para regresar a las opciones generales
3.- En Response enter Distancia.

4.- En **Model** enter Tire ABS.
5.- Presiona **Comparison.**
6.- Completa el recuadro como se indica a continuación:
7.- Presiona **OK** en cada recuadro

Interpretando tus resultados
Use un ∞ de 0.05 para todas las pruebas.
La primera tabla compara las llantas GT con las LS y las MX. Los resultados revelan que el promedio de distancia de frenado obtenidos con las llantas GT es significativamente mas bajo que los obtenidos con cualquier de las otras llantas LS (p = 0.0008) y MX (p = 0.0131).
La segunda tabla compara las llantas LS y MX, donde no hubo diferencias significativas (p = 0.4436)

Tukey Simultaneous Tests
Response Variable Distance
All Pairwise Comparisons among Levels of Tire
Tire = GT subtracted from:

Tire	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
LS	9,750	2,215	4,401	0,0008
MX	7,000	2,215	3,160	0,0131

Tire = LS subtracted from:

Tire	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
MX	-2,750	2,215	-1,241	0,4436

Consideraciones finales
Conclusiones prácticas
En los términos de frenado y pavimento mojado, la mejor llanta es la GT. También es mejor tener el sistema ABS disponible, y no parece muy importante que la banda de rodadura sea de 10. mm o 1.5 mm de profundidad.
Consideraciones estadísticas
Las ventajas del procedimiento incluyen lo siguiente:
- Puede ser usado con un modelo desbalanceado.
- Puede ser usado para evaluar las diferencias entre niveles individuales de medias.

Es importante validar las suposiciones del residual, antes de dibujar cualquier conclusión final de los resultados del ANOVA.
Este análisis incluye factores estables significando que los niveles incluidos fueron de interés directo y no significo que fueran generalizados a otros niveles. El procedimiento del modelo lineal general puede también ser usado con factores aleatorios, los cuales son factores para que los niveles sean seleccionados aleatoriamente y sean proyectados para representar una mayor población de posibles niveles.
Un ejemplo excelente es un estudio de un gage R&R.
Todos los factores en este análisis fueron cruzados significando por ejemplo, que cada nivel de llanta puede ser probado con cada nivel de banda de rodadura. Los factores son considerados jerarquizados si todos lo niveles de un factor suceden completamente dentro de un nivel de otro.

Ejercicio 4.1 Prueba de vinos
Problema:
Tú tratas de determinar si existen diferencias significativas en la calidad de tres vinos: Matador, Conquistador y Saeta.
Colección de datos
10 jueces de vinos probaron tres vinos cada uno y cada uno califico su calidad. El orden de las pruebas fue aleatorio y cada juez pruebo los vinos en diferente orden.
Instrucciones:
1.- Use el modelo lineal General para analizar los puntajes en función del vino.
- Incluye los jueces en el bloque de variables para reducir la variabilidad.
- Realiza el ajuste (Fits) y el residual.
- Incluye el pairwise comparación en el factor del vino en tus resultados.
-

2.- Genera los efectos principales en la gráfica de vinos.
3.- usa **Stat > Regression > Residual Plots** para validar las suposiciones del modelo.

Set de Datos
RIOJA.MPJ

Nombre	Tipo de dato	Tipo de variable	Niveles
Juez	Texto	Factor	Antonio
			Ballardo
			Carmen
			Esmeralda
			Fernanda
			Gerardo
			Hernan
			Irma
			Josefina
Vino	Texto	Factor	Matador
			Conquistador
			Saeta
Puntaje	Numérico	Respuesta	

Ejercicio 4.2 Contenido de fosfato
Problema
Tú quieres evaluar cuanto tiempo toma el uso del gravímetro contra el método spectometrico para medir el contenido fosfato que contienen dos tipos de material.
Colección de datos
Seis ingenieros tomaron muestras del contenido de fosfato de cada material usando cada método.
Los datos son de J. Neter, W. Wasserman y M.H. Kutner (1985) Aplicación lineal del modelo estadístico, segunda edición Irwin , In. Pagina 936.
Instrucciones
1.- Use el modelo lineal general para analizar el tiempo como una función del material y del método.

- Incluye a los ingenieros en el bloque de variables para reducir la variabilidad.
- Realiza el ajuste (Fits) y el residual.
- Incluye el pairwise comparación y la interacción entre el método * material en tus resultados.

2.- Genera una grafica de interacción del material * método.
3.- Usa **Stat > Regression > Residual Plot** para validar el modelo de suposiciones.

Set de datos
PHOSPHAY.MPJ

Nombre	Tipo de dato	Tipo de variable	Niveles
Tiempo	Numérico	Response	
Ingeniero	Texto	Factor	Jones
			Williams
			Adams
			Dixon
			Erickson
			Mynes
Material	Texto	Factor	A,B
Método	Texto	Factor	Gravimeteric
			Spectro

Trabajo elaborado por
Lic. Cecilia C. Díaz García
ccdiagar@hotmail.com
Lic. Angélica Esquivel
Ing. Maria Valle
Maestría en Administración y Liderazgo
Materia Estadística para administradores
Alumnas de la Universidad del Noreste de Coahuila