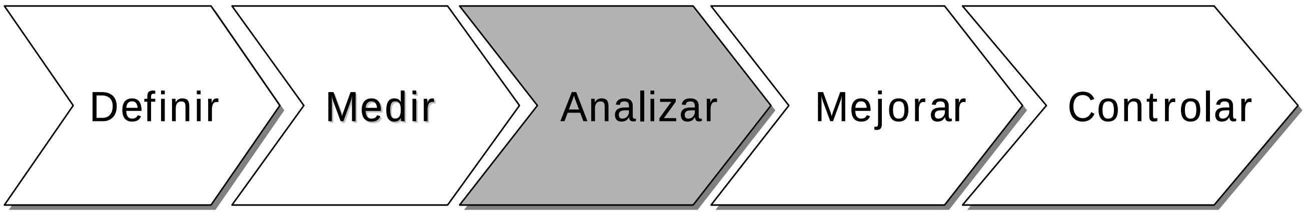
CAPÍTULO 4 FASE DE ANÁLISIS



4.1 Introducción

En esta fase se efectúa el análisis de los datos obtenidos en la etapa de Medición, con el propósito de conocer las relaciones causales o causas raíz del problema. La información de este análisis nos proporcionará evidencias de las fuentes de variación y desempeño insatisfactorio, el cual es de gran utilidad para la mejora del proceso.

Los objetivos de esta fase son:

- Determinar el nivel de desempeño del proceso actual.
- Identificar cuales son las fuentes de variación. Por ejemplo mediante el análisis Multi-Vari podemos determinar las fuentes que presentan mayor variación, a través de la descomposición de los componentes de variabilidad del proceso. las cuáles pueden ser, por ejemplo: de lote a lote, dentro del lote, de turno a turno, entre turnos, dentro del turno, de máquina a máquina, dentro de la máquina, de operador a operador, dentro del operador, entre operadores, etc.
- Una vez identificadas las causas potenciales por medio de una lluvia de ideas y un diagrama de causa efecto, se realiza un proceso de validación estadística de las mismas apoyándose en Análisis de regresión, Pruebas de Hipótesis y Análisis de varianza.

Las herramientas incluidas en está fase se muestran en la tabla 4.1:

Herramienta	¿Para qué es utilizada?
Capacidad del proceso.	Determinar cual es el desempeño del proceso para cumplir con los límites de especificaciones.
Mediciones para seis sigma	Mide la capacidad del proceso en términos cuantificables, revisándolo para su mejora a través del tiempo.
AMEF *(FMEA)*	Identificar las maneras en las cuales un proceso puede fallar para alcanzar los requerimientos críticos del cliente. Estimar el riesgo de las causas específicas en relación con estas fallas.
Intervalos de confianza y Pruebas de hipótesis	Herramienta utilizada para ser inferencias de la población a partir de una muestra.
Tablas de contingencia	Comparación de valores esperados vs observados.
Análisis Multivari	El objetivo general de las cartas Multi-Vari es, descubrir los componentes de variación en el proceso y cuantificar las diferentes fuentes de variabilidad.
Análisis de Varianza (ANOVA)	Metodología para analizar la variación entre muestras y al interior de las mismas con varianzas. Nos sirve para comparar dos o más medias poblacionales.
Análisis de Regresión	Sirve para predecir el valor de una

	variable a partir de una o más variables. Es usada para conocer las relaciones que existen entre las variables dependientes e independientes.

Tabla 4.1 Herramientas de análisis

4.2 Etapas de la fase de análisis

La fase de análisis consta de las siguientes etapas:

4.2.1 Determinar la capacidad del proceso

La capacidad del proceso mide la habilidad del proceso para cumplir con los requerimientos. Compara la variación del proceso contra la variación permitida por el cliente.

Utilizando la herramienta "mediciones para Seis Sigma" se calcula:

DPU: es la relación de la cantidad de defectos, entre el número de unidades producidas.

DPO: es similar al DPU excepto porque considera el numero total de oportunidades que existen para que un defecto ocurra.

DPMO (PPM) = es el producto de DPO X 1,000,000.

Sigma a corto plazo ($_{ST}$): es el nivel de desviaciones estándar (sigma) en el cual se encuentra nuestro proceso.

Sigma a Largo plazo ($_{LT}$): Es el sigma a corto plazo – 1.5, por el desplazamiento que tiene la media a lo largo del tiempo, si no se toma antes una medida preventiva.

4.2.2 Definir el objetivo de desempeño

En esta etapa se define la meta que perseguimos, es decir el *nivel de sigma* esperado. Una opción es realizar un Benchmarking, este es un mecanismo para identificar quien tiene el mejor desempeño, ya sea dentro o fuera de la organización y comparamos nuestros valores contra ese parámetro de referencia y determinar la "brecha" existente.

4.2.3 Identificar las fuentes de variación

Cuando un proceso esta fuera de las especificaciones, hay evidencia de que existe variación. Para comprobarlo usamos alguna de las herramientas de análisis, según sea el caso, por ejemplo, el análisis Multi-Vari es una herramienta estadística que nos permite determinar cuales fuentes presentan mayor variación, a través de la descomposición de los componentes de variabilidad del proceso. Una vez determinadas las causas de variación, nos enfocaremos en los *"pocos vitales X"* que están afectando la variable de respuesta "Y". Esto se puede priorizar usando el *"diagrama de Pareto".*

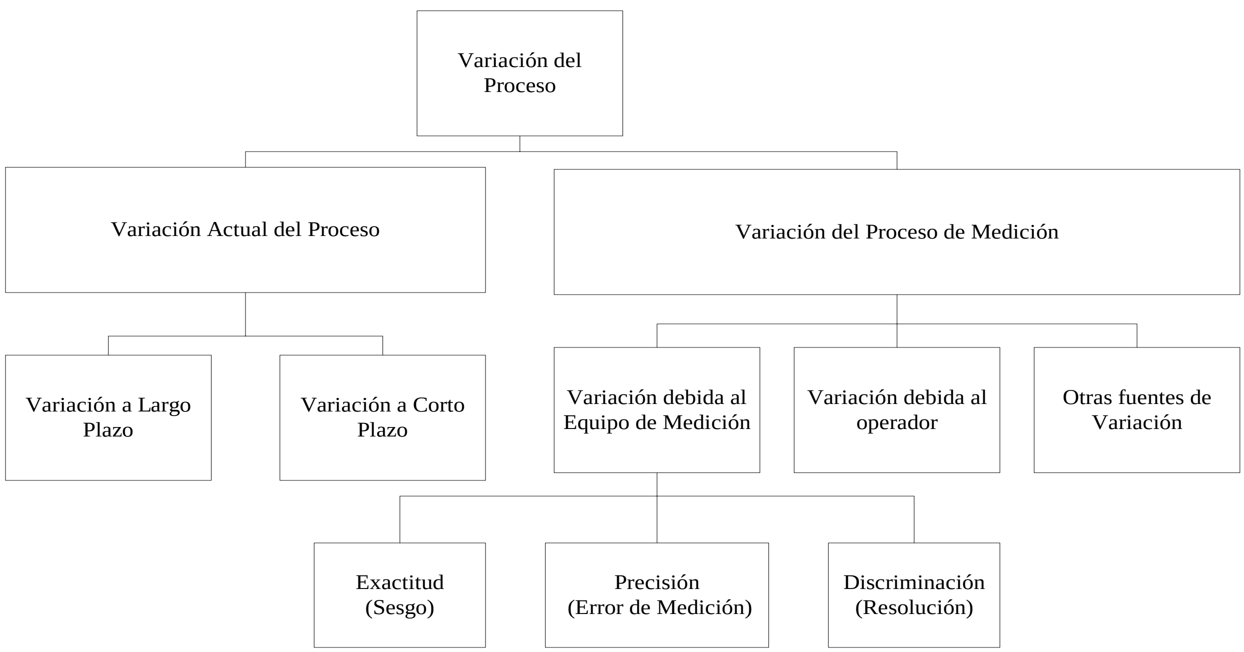


Figura 4-1 Posibles fuentes de variación del proceso.

Para una validación estadística de causas se utilizan diversas herramientas estadísticas como las que se explican a continuación.

4.3 Regresión lineal

La Regresión lineal es la predicción del valor de una variable a partir de una o más variables. La variable predecida es la variable dependiente (***Y***) y la variable independiente (***x***) variable de predicción.

En muchos problemas hay dos o más variables relacionadas, y es necesario explorar la naturaleza de esta relación. El análisis de regresión puede emplearse para construir un modelo que exprese el rendimiento como una función de la temperatura. Este modelo puede usarse para predecir el rendimiento a un nivel determinado de temperatura.

Supuestos para el modelo de regresión lineal

1. Cada valor de *x*, la variable aleatoria ε se distribuye normalmente.
2. Para cada valor de *x*, la media o valor esperado de ε es 0; esto es, $E(\varepsilon) = \mu_\varepsilon = 0$.
3. Para cada valor de *x,* la varianza de ε es la constante σ^2 (llamada varianza del error).
4. Los valores del término de error ε son independientes.
5. Para un valor fijo de *x*, la distribución muestral de ***Y*** es normal, porque sus valores dependen de los de ε.
6. Para un valor fijo *x*, es posible predecir el valor de ***Y***.
7. Para un valor fijo *x*, es posible estimar el valor promedio de ***Y.***

Ejemplo 1:

La revista *Motor Trend*[27] presenta a menudo, datos del rendimiento de automóviles, comparando el tamaño del motor en pulgadas cúbicas de desplazamiento (pcd) y las millas por galón (mpg) estimadas para ocho modelos de automóviles subcompactos modelo 1984.

[27] Motor Trend es marca registrada

coches compactos	tamaño del motor (pcd) *x*	millas/galón (mpg), y
Chevrolet Cavalier	121	30
Datsun Nissan Stanza	120	31
Dodge Omni	97	34
Ford Escort	98	27
Mazda 626	122	29
Plymouth Horizon	97	34
Renault Alliance/Encore	85	38
Toyota Corolla	122	32

Tabla 4.2 Relación entre el rendimiento y el desplazamiento cúbico de un motor

Graficando los datos en un "diagrama de dispersión" podemos observar la colección de los ocho pares de datos *(x, y)* como muestra de una población de pares, donde las medidas pulgadas cúbicas de desplazamiento (pcd) "*x*" pueden tomar cualquier valor en el rango de valores que se extiende de 85 a 122. Para cada pcd posible hay muchos millajes asociados con ella. Por ejemplo para un tamaño del motor de 98 hay un gran número de millajes asociados, uno por cada coche cuyo tamaño sea 98 pcd. Asumamos que existe una relación lineal para la población de pares de datos de pcd y mpg. (Se entiende por relación lineal cuando la variable *y* tiene una tendencia a crecer o decrecer, cuando la variable x aumenta y el valor r tiende a ± 1).

Cálculo por medio de Minitab

Las instrucciones para Minitab son:

> Stat > Regression > Regression

Seleccionar Response = Columna de las Y's

Seleccionar Predictors = Columnas de las Xs

OK

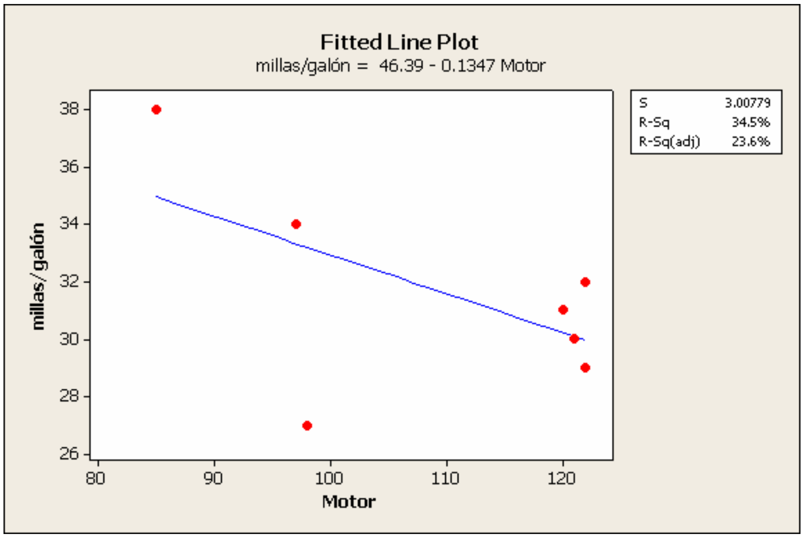


Figura 4.2 Regresión Lineal

Usamos el **modelo probabilístico** para explicar el comportamiento de los millajes para las ocho medidas de tamaño de motor, este se llama **modelo de regresión lineal**[28], y expresa la relación lineal entre tamaño de motor (x) y millas por galón (y).

Modelo de regresión lineal

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (4.1)$$

Donde:

y = variable dependiente

β_0 = ordenada al origen

β_1 = pendiente

x = variable independiente

ε = Error aleatorio

La expresión $\beta_0 + \beta_1 x$ se denomina **componente determinística** del modelo de regresión lineal. La muestra de pares de datos se usará para estimar los parámetros $\beta_0 y \beta_1$ de la componente determinística.

[28] Portus, Lincoyán. Curso Práctico de Estadística. México, p 25, México, Editorial Mc Graw Hill. 1988

4.4 Cartas Multi Vari

Técnica de pre-experimentación diseñada para aislar y cuantificar los mayores componentes de variabilidad en procesos de producción. El análisis Multi-Vari permite determinar las fuentes que presentan mayor variación, a través de la descomposición de los componentes de variabilidad del proceso, fueron desarrolladas por Dorian Shainin[29], encaminadas a la mejora del proceso.

El objetivo general de las cartas Multi-Vari es, descubrir los componentes de variación en el proceso, por ejemplo: de lote a lote, dentro del lote, de turno a turno, entre turnos, dentro del turno, de máquina a máquina, dentro de la máquina, etc.

Las cartas Multi-Vari identifican tres principales familias de variación que pueden influenciar en la variabilidad del proceso, éstas son: ***variación posicional, cíclica y temporal*.**

1. Variación posicional. Se refiere a variaciones dentro de una misma pieza, variaciones producidas de un troquel a otro, dentro de un lote de piezas, de una máquina a otra, de un operador a otro, o de una plata a otra.

2. Variación cíclica. Es la variación entre unidades de un mismo proceso, o variación entre grupos de unidades (lotes).
3. Variación temporal. Variación de diferencia de tiempo (por ejemplo: de hora a hora, de día a día, de semana a semana, etc.), variación de una corrida de producción a otra, o variación de turno a turno.

[29] Dorian Shainin (1914 – 2000). Precursor de herramientas como Lot-plot, Multi-Vari, Pre-control, B vs C, (prueba no paramétrica de significancia estadística basada en la técnica de John Tukey), paired comparisons, y otras.

Una vez identificados las fuentes de variación, el análisis Multi-Vari esta diseñado para identificar la variable independiente de mayor influencia dentro de las familias de variación descritas anteriormente.

Pasos para elaborar una carta Multi-Vari

1. Selección del proceso. Cualquier proceso que se presente fuera de las especificaciones permitidas es candidato a ser seleccionado para posteriormente ser analizado por medio de cartas Multi-Vari.

2. Plan de muestreo y toma de mediciones. El objetivo es tomar mediciones que otorguen información con respecto a las principales familias de variación que pueden influenciar en la variabilidad del proceso. Es decir, para el caso de variación posicional, tomar mediciones en una máquina y en otra, éstas mediciones deben ser tomadas de manera que resulten en un muestreo aleatorio. El muestreo aleatorio puede ser llevado a cabo con tablas de números aleatorios, dichas tablas contienen una serie de columnas y renglones de números, por ejemplo, el último dígito de un número nos puede dar el intervalo de minutos que se debe esperar para tomar la siguiente medición.

3. Toma de mediciones para variación simultánea. Es recomendable también realizar una toma de mediciones que cubra más de una familia de variabilidad, por ejemplo es posible incluir la variación temporal o la variación cíclica a una carta Multi-Vari que este enfocada en la variación posicional.

4. Determinar los limites de variación. Obtener el máximo y el mínimo del total de las mediciones tomadas anteriormente. Así como el máximo y el mínimo para cada familia de variabilidad seleccionada.

5. Graficar los rangos. Las cartas Multi-Vari están formadas por cuatro ejes dispuestos de manera rectangular.

    a. **Ejes.** Los ejes verticales (izq. y der.) grafican la escala de la medidas tomadas, use como guía los límites de variación obtenidos en el paso 4. El eje horizontal inferior es usado para identificar las unidades muestreadas, la posición, etc. El eje horizontal superior es usado para agrupar por lotes, para agrupar por turno, por máquina, etc.

    b. **Cajas.** La representación es similar a una gráfica de “Box and Whiskers”, con la diferencia que el eje horizontal superior define los límites horizontales de las cajas. Los límites verticales de las cajas están definidos por la medición máxima y el mínima de la familia de variabilidad seleccionada.

    c. **Mediciones.** Las mediciones se ordenan clasificando por la familia de variabilidad representada en el eje horizontal superior. Posteriormente, son agrupadas por la familia de variabilidad del eje horizontal inferior. Finalmente son representadas en la gráfica y unidas por una línea horizontal.

    d. Medias. La media de cada clasificación de la familia de variabilidad se representada en el eje horizontal superior. La media de cada caja se grafica en el centro de cada caja. Las medias son unidas con una línea sólida ver figura

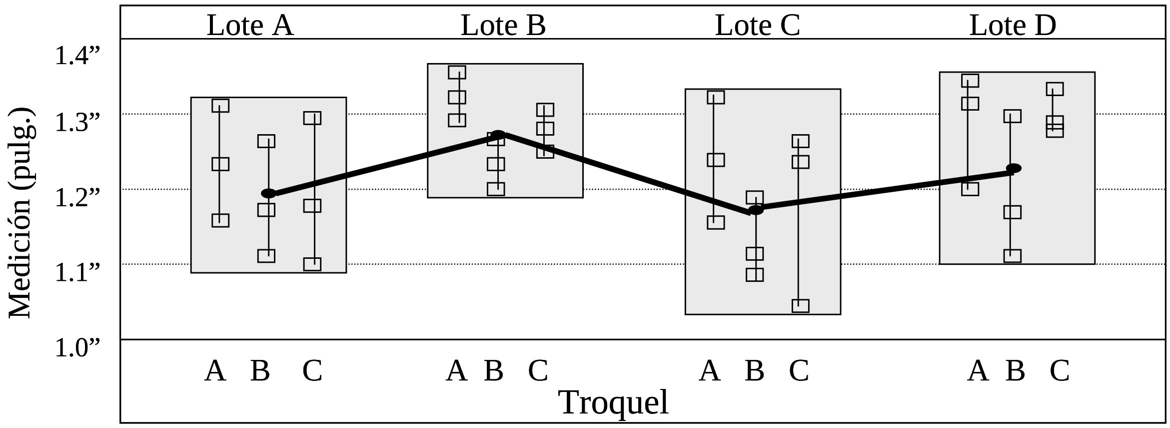


Figura 4.3 Gráfica Multi Vari

Ejemplo de cartas Multi-Vari

Se detectó que el proceso de producción de rondanas de plástico se encuentra fuera de especificación. El diámetro exterior es de 1.50” + 0.01”. Se realiza un proceso de pre-experimentación a través de cartas Multi-Vari para identificar la fuente de variabilidad. Se planifica un muestreo aleatorio de cinco diferentes lotes de material a ser procesadas en dos diferentes inyectoras de plástico. Se toman cinco muestras para cada una y para cada lote. Sumando 50 observaciones.

La figura 4.4, muestra la carta Multi-Vari obtenida. El software estadístico usado para obtener esta carta es Minitab.

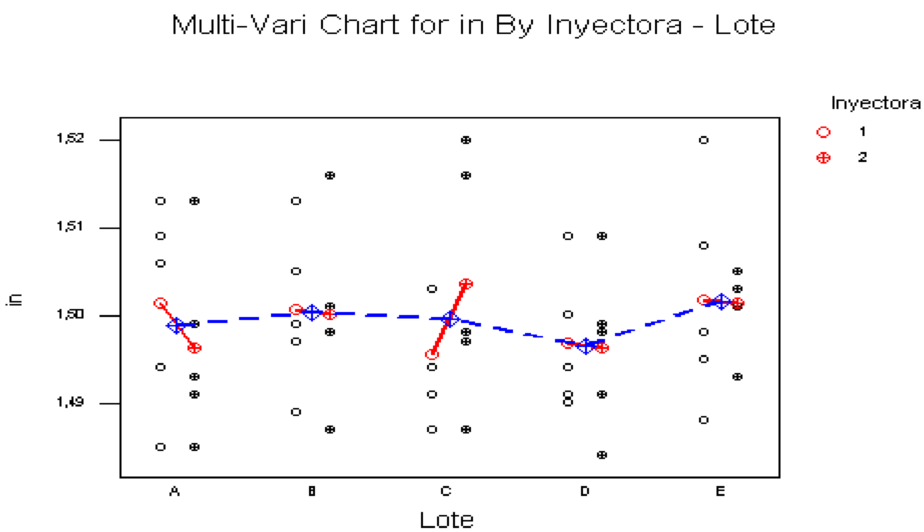


Figura 4.4 Gráfica Multi Vari en Minitab

Una carta Multi-Vari adicional correspondería a los ejes horizontales invertidos, esto es, ordenados por lote y luego por inyectora.

4.5 Pruebas de hipótesis para una población[30]

Se trata de probar una afirmación sobre parámetros de la población (media μ; varianza σ^2 o proporción π) con base en datos de estadísticos de una muestra ($\overline{X}$, s^2 o p respectivamente); probar las afirmaciones en los parámetros se usan los estadísticos:

En una población:

La media poblacional	μ **= 12**;	estadístico Zc
La varianza poblacional	σ^2 **= 12**;	estadístico χ_c^2
La proporción poblacional	π **= 0.3**	estadístico Zc

En dos poblaciones:

Las medias poblacionales son iguales $\mu_1 = \mu_2$ o $\mu_1 - \mu_2 = 0$; estadístico Zc o Tc

Las varianzas poblacionales son iguales $\sigma_1^2 = \sigma_2^2$ o $\sigma_1^2 - \sigma_2^2 = 0$; estadístico Fc

Las proporciones poblacionales son iguales $\pi_1 = \pi_2$ o $\pi_1 - \pi_2 = 0$; estadístico Zc

- Hipótesis nula H_o, complemento de la Hipótesis alterna H_a:
    - Es la hipótesis o afirmación a ser probada
    - Puede ser por ejemplo μ **=,** $\leq$ o $\geq$ a 5
    - Sólo puede ser rechazada o no rechazada
- Hipótesis alterna H_a, complemento de la hipótesis nula:
    - Es la hipótesis que se acepta como verdadera cuando se rechaza H_o, es su complemento

[30] Lind, Douglas, Mason, Robert. Estadística para la Administración y Economía. México. Editorial Mac Graw Hill. 2004

- Si el signo de la hipótesis alterna es ≠ entonces se trata de una prueba de dos colas; si es **>** de cola derecha y si es **<** de cola izquierda.
- Puede ser por ejemplo $\mu \neq$ **5** para prueba de dos colas
- μ **< 5** para prueba de cola izquierda
- μ **> 5** para prueba de cola derecha

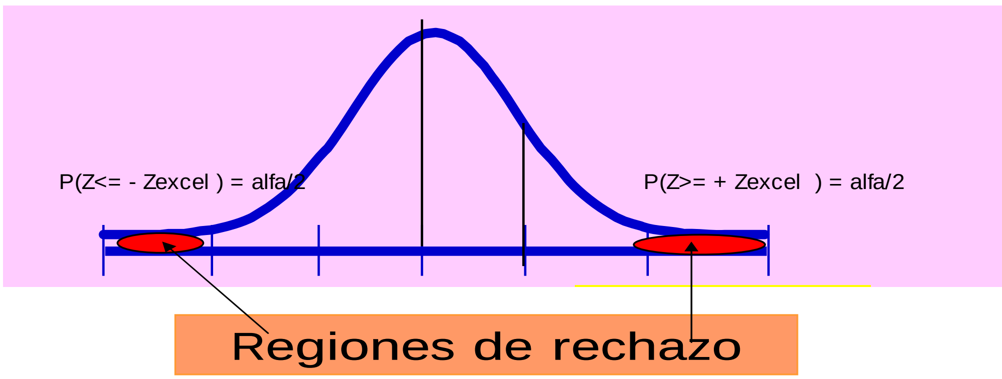


Figura 4.5 Representación de la Prueba de Hipótesis

- Se plantea la H_a si en el problema se muestra la afirmación de ser menor o mayor a un valor establecido histórico.

- Se plantea la H_o si en el problema se muestra la afirmación igual (es o históricamente ha sido); mayor o igual (cuando menos) o menor o igual (a lo más) a un valor establecido histórico.

No importa cual se plantee primero, siempre la conclusión se hace contra la H_o (se rechaza o no se rechaza)

- El intervalo de confianza es el intervalo donde se estima que se encuentre el parámetro de la población (media μ; varianza σ^2 o proporción π) para un cierto nivel de confianza o de significancia.

Estadístico de prueba

- Para probar la hipótesis nula se calcula un estadístico de prueba con la información de la muestra el cual se compara a un valor crítico apropiado. De esta forma se toma una decisión sobre rechazar o no rechazar la Ho

- **Error tipo I** (alfa = nivel de significancia, es común = 0.05). **Alfa (α)= 1- Nivel de confianza**
  - o Se comete al rechazar la H_o cuando en realidad es verdadera. También se denomina riesgo del productor.
- **Error tipo II beta (β)**
  - o Se comete cuando no se rechaza la hipótesis nula siendo en realidad falsa. Es el riesgo del consumidor.

Pruebas de Hipótesis de dos colas:

Si la H_o: μ = que un valor poblacional, entonces el riesgo alfa se reparte en ambos extremos de la distribución. Por ejemplo si Ha: $\mu \neq 10$ se tiene:

H_o: a = b

H_a: $a \neq b$

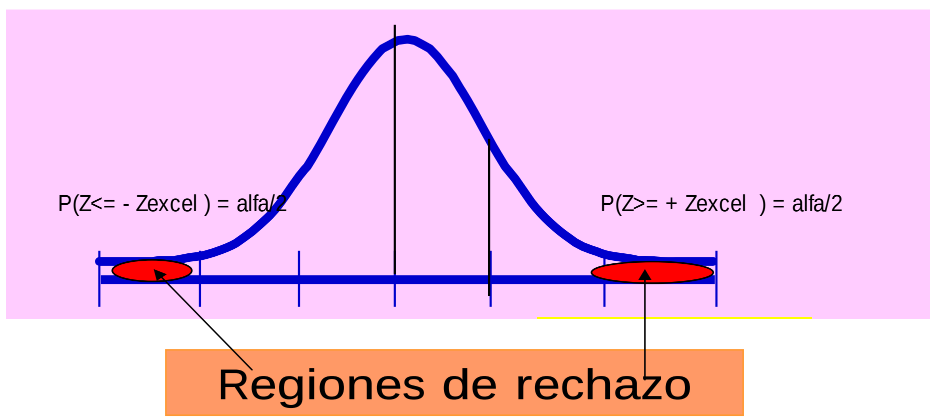


Figura 4.6 Prueba de Hipótesis de dos colas

Pruebas de Hipótesis de cola derecha:

Si la Ho: $\mu \leq$, que un valor poblacional, entonces el riesgo alfa es en el extremo derecho de la distribución. Por ejemplo si Ho $\mu \leq 10$ y Ha: $\mu > 10$ se tiene una prueba de cola derecha:

H_o: $a \leq b$

H_a: $a > b$

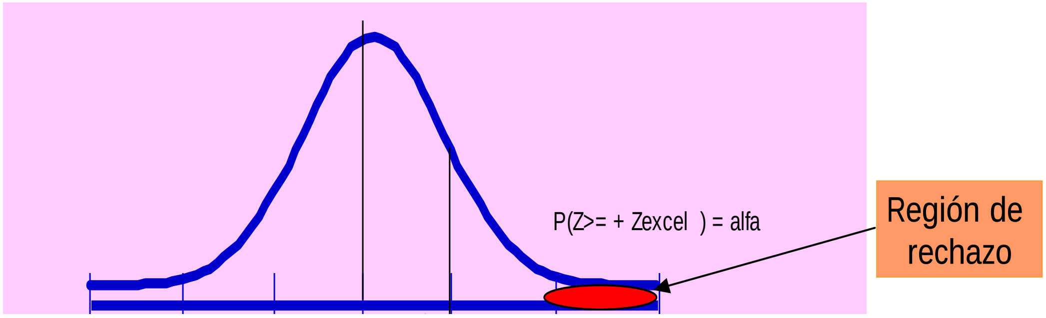


Figura 4.7 Prueba de Hipótesis de cola derecha

Pruebas de Hipótesis cola izquierda:

Si la Ho: $\mu \geq$ que un valor poblacional, entonces el riesgo alfa se coloca en el extremo izquierdo de la distribución. Por ejemplo si Ho $\mu \geq 10$ y Ha: $\mu < 10$ se tiene una prueba de cola izquierda:

H_o: a $\geq$ b

H_a: a < b

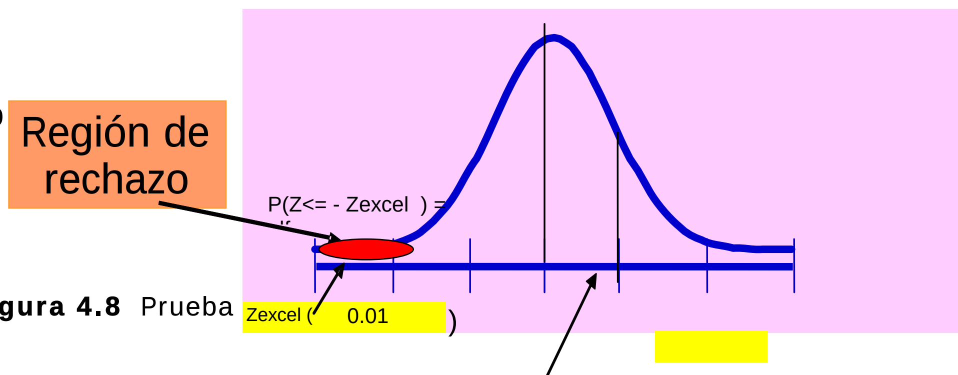


Figura 4.8 Prueba

Pasos para probar una Prueba de Hipótesis

a) Probar la hipótesis de igualdad de una media μ para n > 30

1) Establecer las hipótesis e identificar el nivel de significancia alfa o 1-Nivel de confianza (NC)

H_o: $\mu = \mu_o$

H_a: $\mu \neq \mu_o$

2) Calcular el estadístico de prueba Zc o Tc con formula

$$Zc = \frac{\overline{X} - \mu_{HIPOTESIS}}{\sigma / \sqrt{n}} \tag{4.2}$$

3) Determine el estadístico de tablas Z_t o T_c de Excel para una alfa

4) Establecer la región de rechazo con Z_t y ver si cae ahí Z_c

Las regiones de rechazo prueba de 2 colas: - $Z_{\alpha/2}$ y + $Z_{\alpha/2}$

5) Determinar el Intervalo de confianza para la media y ver si incluye a la media de la hipótesis, si no rechazar H_o

$IC\ para\ \mu = \overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ o si no se conoce Sigma $IC\ para\ \mu = \overline{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$

6) Determinar el valor P correspondiente a Zc y comparar contra Alfa/2, si es menor rechazar Ho

P value en Excel =DISTR.NORM.ESTAND(Zc)

Prueba de hipótesis en Minitab:

>Stat >Basic statistics > 1-Sample Z

Utilizar Sample in columns o Summarized data

Sample size 49 Mean 11.5 Standar deviation 1.1 Test Mean 12

Graphs – Seleccionar / Individual value plot

Confidence level 90% Alternative not equal

OK

Resultados:

One-Sample Z

Test of mu = 12 vs not = 12

The assumed standard deviation = 1.1

N	Mean	SE Mean	90% CI	Z	P
49	11.5000	0.1571	(11.2415, 11.7585)	-3.18	0.001

Criterios de rechazo de H_o:

Si Z_c cae en la zona de rechazo

El valor de la Hipótesis no se encuentra en el Intervalo de confianza

El valor P es menor que el valor de alfa (prueba de una cola) o de alfa/2 (dos colas).

a) Estadístico t_c para muestras pequeñas (n<30) y la σ es desconocida:

$$t_C = \frac{\overline{X} - \mu_{HIPOTESIS}}{s/\sqrt{n}} \qquad (4.3)$$

Estadístico de tablas T_α o $T_{\alpha/2}$ en Excel =DISTR.T.INV(2*alfa o alfa, grados de libertad n-1)

Intervalo de confianza para estimar μ con muestras pequeñas (n < 30; grados de libertad (gl.) = n –1):

$$IC\ para\ \mu = \overline{X} \pm t_{\alpha/2,gl}\frac{S}{\sqrt{n}} \qquad (4.4)$$

El valor p de probabilidad correspondiente al estadístico T_c se determina como sigue:

P value en Excel =DISTR.T(T_c, grados de libertad, 1 o 2 colas)

Prueba de hipótesis en Minitab:

>Stat >Basic statistics > 1-Sample t

Utilizar Sample in columns o Summarized data

Sample size 10 Mean 11277 Standar deviation 3772

Test Mean 12000

Graphs – Seleccionar º! Individual value plot

Confidence level 95% Alternative not equal

OK

Resultados:

One-Sample T

Test of **μ** = 12000 vs not = 12000

N	Mean	St Dev	SE Mean	95% CI	T	P
10	11277.0	3772.0	1192.8	(8578.7, 13975.3)	-0.61	0.559

Criterios de rechazo de Ho:

Si Tc cae en la zona de rechazo

El valor de la Hipótesis no se encuentra en el Intervalo de confianza

El valor P es menor que el valor de alfa (prueba de una cola) o de alfa/2 (dos colas).

c) Estadístico Zc para proporciones y muestras grandes (n >= 30):

$$\sigma_p = \sqrt{\frac{\pi_{HIPOTESIS}(1-\pi_{HIPOTESIS})}{n}}$$

$$Zc = \frac{p-\pi_{HIPOTESIS}}{\sigma_p} \qquad (4.5)$$

Estadístico de tablas Zalfa o Zalfa/2 en Excel

=DISTR.NORM.ESTAND.INV(alfa o alfa/2)

Intervalo de confianza para estimar π proporción poblacional:

$$S_p = \sqrt{\frac{p(1-p)}{n}}$$

$$IC\ para\ \pi = p \pm Z_{\alpha/2} S_p \qquad (4.6)$$

El valor p de probabilidad correspondiente al estadístico Zc se determina como sigue:

P value en Excel =DISTR.NORM.ESTAND(Zc)

Prueba de hipótesis en Minitab

>Stat >Basic statistics > 1- Proportion

Summarized data

Number of trial 500 Number of events 225

Graphs – Seleccionar º! Individual value plot

Confidence level 98% Test proportion 0.40

Alternative Less than

º! Use test and interval based on normal distribution

OK

Resultados:

Test and CI for One Proportion

Test of p = 0.4 vs p < 0.4

Sample	X	N	Sample p	Bound	Z-Value	P-Value
1	225	500	0.450000	0.495693	2.28	0.989

Criterios de rechazo de H_o:

Si Z_c cae en la zona de rechazo

El valor de la Hipótesis no se encuentra en el Intervalo de confianza

El valor P es menor que el valor de alfa (prueba de una cola) o de alfa/2 (dos colas).

d) Estadístico tc para muestras pequeñas (n < 30) y la σ es desconocida:

$$t_C = \frac{\overline{X} - \mu_{HIPOTESIS}}{s/\sqrt{n}} \tag{4.7}$$

Estadístico de tablas Talfa o Talfa/2 en Excel =DISTR.T.INV(2*alfa o alfa, grados de libertad n-1)

Intervalo de confianza para estimar μ con muestras pequeñas (n < 30; grados de libertad (gl.) = n –1):

$$IC \;\; para \;\; \mu = \overline{X} \pm t_{\alpha/2, gl} \frac{S}{\sqrt{n}} \tag{4.8}$$

El valor p de probabilidad del estadístico Tc se determina como sigue;

P value en Excel =DISTR.T(Tc, grados de libertad, 1 o 2 colas)

Prueba de hipótesis en Minitab:

>Stat >Basic statistics > 1-Sample t

- Utilizar Sample in columns o Summarized data

  Sample size 10 Mean 11277 Standar deviation 3772

  Test Mean 12000

-Graphs – Seleccionar º! Individual value plot

-Confidence level 95% Alternative not equal → OK

Resultados:

One-Sample T

Test of mu = 12000 vs not = 12000

N	Mean	StDev	SE Mean	95% CI	T	P
10	11277.0	3772.0	1192.8	(8578.7, 13975.3)	-0.61	0.559

Criterios de rechazo de Ho:

Si T_c cae en la zona de rechazo

El valor de la Hipótesis no se encuentra en el Intervalo de confianza

El valor P es menor que el valor de alfa (prueba de una cola) o de alfa/2 (dos colas).

e) Estadístico Z_c para proporciones y muestras grandes (n >= 30):

$$\sigma_p = \sqrt{\frac{\pi_{HIPOTESIS}(1-\pi_{HIPOTESIS})}{n}}$$

$$Zc = \frac{p-\pi_{HIPOTESIS}}{\sigma_p} \tag{4.9}$$

Estadístico de tablas Zalfa o Zalfa/2 en Excel

=DISTR.NORM.ESTAND.INV(alfa o alfa/2)

Intervalo de confianza para estimar π proporción poblacional:

$$S_p = \sqrt{\frac{p(1-p)}{n}}$$

$$IC\ para\, \pi = p \pm Z_{\alpha/2} S_p \tag{4.10}$$

El valor p de probabilidad correspondiente al estadístico Zc se determina como sigue:

P value en Excel =DISTR.NORM.ESTAND(Zc)

Prueba de hipótesis en Minitab

>Stat >Basic statistics > 1- Proportion

Summarized data

Number of trial 500 Number of events 225

Graphs – Seleccionar º! Individual value plot

Confidence level 98% Test proportion 0.40

Alternative Less than

º! Use test and interval based on normal distribution → OK

Resultados:

Test and CI for One Proportion

Test of p = 0.4 vs p < 0.4

Sample	X	N	Sample p	Bound	Z-Value	P-Value
1	225	500	0.450000	0.495693	2.28	0.989

Criterios de rechazo de H_o:

Si Zc cae en la zona de rechazo

El valor de la Hipótesis no se encuentra en el Intervalo de confianza

El valor P es menor que el valor de alfa (prueba de una cola) o de alfa/2 (dos colas).

f) Tamaño de muestra para estimar μ en función del error $(\bar{X}-\mu)$:

$$n = \frac{Z^2_{\alpha/2}}{(\bar{X}-\mu)^2} \tag{4.11}$$

g) Tamaño de muestra para estimar π en función del error $(p-\pi)$, en el peor caso π = 0.5:

$$n = \frac{Z^2_{\alpha/2}(\pi)(1-\pi)}{(p-\pi)^2} \tag{4.12}$$

4.6 Pruebas de hipótesis para dos poblaciones

Pruebas de igualdad de dos varianzas

Presentaremos ahora pruebas para comparar dos varianzas. Supóngase que son dos las poblaciones de interés, por ejemplo X_1 y X_2, donde $\mu_{1,}\sigma_1^2, \mu_2, \sigma_2^2$, se desconocen. Deseamos probar hipótesis relativas a la igualdad de las dos varianzas, $H_0 : \sigma_1^2 = \sigma_2^2$. Considérese que se disponen dos muestras aleatorias de tamaño n_1 de la población 1 y de tamaño n_2 de la población 2, y sean $S_1^2 y S_2^2$ las varianzas de muestra. Para probar la alternativa de dos lados

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Utilizamos el hecho de que la estadística

$$F_0 = \frac{S_1^2}{S_2^2} \tag{4.13}$$

Se distribuye como F, con n_1-1 y n_2 –1 grados de libertad.

Rechazaríamos H_0 si:

$$F_0 > F_{\alpha/2, n_1-1, n_2-1}$$

o si

$$F_0 < F_{1-\alpha/2, n_1-1, n_2-1}$$

Donde $F_{\alpha/2, n_1-1, n_2-1}$ y $F_{1-\alpha/2, n_1-1, n_2-1}$ son los puntos porcentuales $\alpha/2$ superior e inferior de la distribución F con n_1-1 y n_2-2 grados de libertad. La tabla F proporciona sólo los puntos de la cola superior de F, por lo que para determinar $F_{1-\alpha/2, n_1-1, n_2-1}$ debemos emplear

$$F_{1-\alpha/2, n_1-1, n_2-1} = \frac{1}{F_{\alpha/2, n_1-1, n_2-1}} \tag{4.14}$$

La misma estadística de prueba puede utilizarse para probar hipótesis alternativas de un lado. La hipótesis alternativa de un lado es:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 > \sigma_2^2$$

Si $F_0 > F_{\alpha, n_1 - 1, n_2 - 1}$, rechazaríamos $H_0 : \sigma_1^2 = \sigma_2^2$.

Ejemplo: Los siguientes son tiempos de quemado (en minutos) de señales luminosas de dos tipos diferentes.

Tipo 1	Tipo 2
63	64
81	72
57	83
66	59
82	65
82	56
68	63
59	74
75	82
73	82

Tabla 4.3 Igualdad de Varianzas

Pruebe la hipótesis de dos varianzas sean iguales. Use $\alpha = 0.05$

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

$\bar{X}_1 = 70.6$

$\bar{X}_2 = 70$

$S_1^2 = 88.71$

$S_2^2 = 100.44$

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{88.71}{100.44} = 0.877$$

$F_{\alpha/2, n_1 - 1, n_2 - 1}$ = $F_{.025,9,9}$ = 4.03

$F_{1-\alpha/2, n_1 - 1, n_2 - 1}$ =0.248

0.877 no es mayor que 4.03, por lo que no se rechaza la hipótesis nula

$$H_0 : \sigma_1^2 = \sigma_2^2 .$$

Con alfa = 0.05

Prueba de dos poblaciones cuando las varianzas son iguales

Sean X_1 y X_2 dos poblaciones normales independientes con medias desconocidas $\mu_1 \, y \, \mu_2$, y varianzas conocidas pero iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Deseamos probar:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Sean X_1, X_2, S_1^2, S_2^2, las medias y las varianzas de las muestras, respectivamente. Puesto que tanto $S_1^2 como \, S_2^2$ estiman la varianza común σ^2, podemos combinarlas para producir una sola estimación, mediante la siguiente fórmula:

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \qquad (4.15)$$

Para probar $H_0 : \mu_1 = \mu_2$ calcúlese la estadística de prueba

$$t_0 = \frac{\overline{X}_1 - \overline{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (4.16)$$

Si $t_0 > t_{\alpha/2, n_1 + n_2 - 2}$ o si $t_0 < -t_{\alpha/2, n_1 + n_2 - 2}$, rechazamos $H_0 : \mu_1 = \mu_2$

Las alternativas de un lado se tratan de modo similar. Para probar:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Calcúlese la estadística de prueba t_0 y rechácese $H_0 : \mu_1 = \mu_2$ si:

$$t_0 > t_{\alpha, n_1 + n_2 - 2}$$

Para la otra alternativa de un lado,

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

Calcúlese la estadística de prueba y rechácese $H_0 : \mu_1 = \mu_2$ si:

$$t_0 < -t_{a,n_1+n_2-2}$$

Ejemplo: Se está investigando la resistencia de dos alambres, con la siguiente información de muestra.

Alambre	Resistencia (ohms)					
1	0.140	0.141	0.139	0.140	0.138	0.144
2	0.135	0.138	0.140	0.139	-	-

Tabla 4.4 Resistencia de dos alambres

Suponiendo que las dos varianzas son iguales, ¿qué conclusiones puede extraerse respecto a la resistencia media de los alambres?

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Calculando la media y la desviación estándar de la muestra:

$\bar{x}_1 = .140$

$\bar{x}_2 = .138$

$S_1 = .0021$

$S_2 = .0022$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} = 0.0021$$

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 1.72$$

Se busca en la tabla de distribución t el valor $t_{\alpha/2,n_1+n_2,-2}$ = $t_{.025,8}$ =2.306 o se determina con Excel.

Utilizando el criterio de rechazo $t_0 > t_{\alpha/2,n_1+n_2-2}$, 1.72 no es mayor que 2.306, por lo tanto no rechazamos H_0.

4.7 **Análisis de varianza de una vía (ANOVA)**

El análisis de la varianza de un criterio (ANOVA) es una metodología para analizar la variación entre muestras y la variación al interior de las mismas mediante la determinación de varianzas. Es llamado de un criterio porque analiza un variable independiente o Factor por ejemplo: Velocidad. Como tal, es un método estadístico útil para comparar dos o más medias poblacionales. El ANOVA de una vía nos permite poner a prueba hipótesis tales como:

$$H_0 = \mu_1 = \mu_2 = \mu_3 = ... = \mu_k$$

$$H_1 : \textit{Al menos dos medias poblacionales son diferentes.}$$

Los supuestos en que se basa la prueba *t* de dos muestras que utiliza muestras independientes son:

1. Ambas poblaciones son normales.
2. Las varianzas poblacionales son iguales, esto es, $\sigma_1^2 = \sigma_2^2$.

Los supuestos para el ANOVA de una vía son:

1. Todas las poblaciones *k* son normales.

2. $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = ... = \sigma_k^2 \left(= \sigma^2\right)$

El Procedimiento es el siguiente[31]:

1. Determinar si las muestras provienen de poblaciones normales.
2. Proponer las hipótesis.
3. Encontrar las medias poblacionales y las varianzas.
4. Encontrar la estimación de la varianza al interior de las muestras s_w^2 y sus grados de libertad asociados $gl_{w.}$
5. Calcular la gran media para la muestra de las medias muéstrales.

[31] C. Weimer Richard. Estadística, p 126. México. Editorial CECSA. 2000

6. Determinar la estimación de la varianza entre muestras s_b^2 y sus grados de libertad asociados.
7. Hallar el valor del estadístico de la prueba F.
8. Calcular el valor crítico para F basado en gl_b y gl_w.
9. Decidir si se rechaza H_0.

Por ejemplo:

Cuatro catalizadores que pueden afectar la concentración de un componente en una mezcla líquida de tres componentes están siendo investigados. Se obtienen las siguientes concentraciones:

Catalizador

A	B	C	D
58.2	56.3	50.1	52.9
57.2	54.5	54.2	49.9
58.4	57	55.4	50
55.8	55.3		51.7
54.9			

Tabla 4.5 Catalizadores

Utilizando el paquete Minitab se tiene:

Stat > ANOVA > One Way (Unstacked)

Response in separate columns A, B, C

Seleccionar Store Residuals y Store Fits Confidence level 95%

Graphs: Seleccionar Normal plot of residuals

Comparisons: Seleccionar Tukey's Family error rate → OK

Se obtuvieron los resultados siguientes:

Los residuos se distribuyen normalmente, validando el modelo ANOVA.

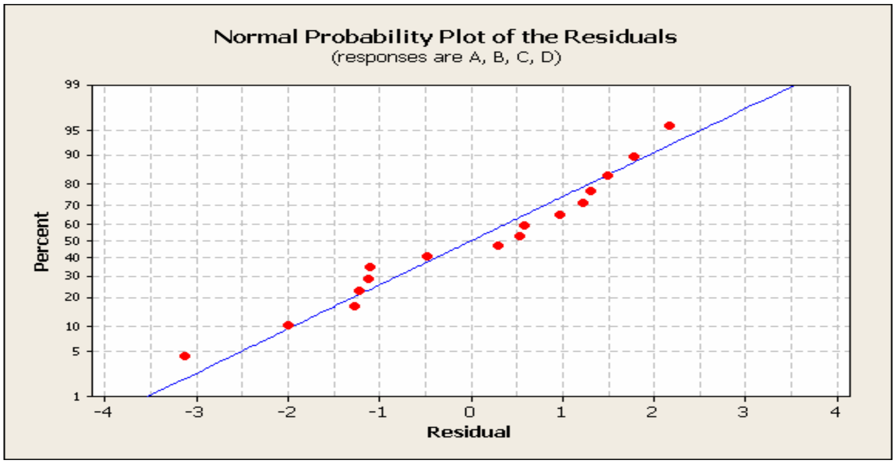


Figura 4.9 Curva de Probabilidad Normal

De la tabla de ANOVA se observa que el valor P es menor a alfa de 0.05, por lo que al menos dos medias son diferentes:

One-way ANOVA: A, B, C, D

```
Source  DF      SS     MS     F      P
Factor   3   85.68  28.56  9.92  0.001
Error   12   34.56   2.88
Total   15  120.24

S = 1.697   R-Sq = 71.26%   R-Sq(adj) = 64.07%
                      Individual 95% CIs For Mean Based on
                        Pooled StDev
Level  N    Mean  StDev  ---+---------+---------+---------+------
A      5  56.900  1.520                        (------*-----)
B      4  55.775  1.100                   (------*------)
C      3  53.233  2.779       (--------*-------)
D      4  51.125  1.443  (-------*------)
                         ---+---------+---------+---------+------
                         50.0      52.5      55.0      57.5
Pooled StDev = 1.697
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 98.83%

A subtracted from:

    Lower  Center   Upper  --------+---------+---------+---------+-
B  -4.506  -1.125   2.256          (------*------)
C  -7.347  -3.667   0.014     (-------*------)
D  -9.156  -5.775  -2.394  (-----*------)
                           --------+---------+---------+---------+-
                                -5.0       0.0       5.0      10.0
```

De aquí se puede observar que el catalizador A es diferente del catalizador D (el cero no se encuentra en el intervalo de confianza)

```
B subtracted from:

    Lower  Center   Upper  --------+---------+---------+---------+-
C  -6.391  -2.542   1.308       (-------*-------)
D  -8.214  -4.650  -1.086    (------*------)
                           --------+---------+---------+---------+-
                                -5.0       0.0       5.0      10.0
```

De aqui se puede observar que el catalizador B es diferente del catalizador D (el cero no se encuentra en el intervalo de confianza)

```
C subtracted from:

    Lower  Center  Upper  --------+---------+---------+---------+-
D  -5.958  -2.108  1.741        (-------*------)
                          --------+---------+---------+---------+-
                               -5.0       0.0       5.0      10.0
```

Por lo anterior el catalizador D es diferente de los demás, para mejores resultados se puede utilizar el catalizador A o B.

Conclusiones

En este capítulo se mostraron las etapas de la Fase de Análisis, adicionalmente se trato la regresión lineal, también se revisaron las cartas Multi Vari, así como, las pruebas de hipótesis de una y dos poblaciones y por último, el Análisis de Varianza (ANOVA)

A continuación se analizará la fase de mejora de la metodología DMAIC y en particular el Diseño de Experimentos.