

CAPÍTULO 1

SÍNTESIS

La tesis presentada es la propuesta de un método para verificar el comportamiento de las características especiales de un Producto - Proceso. Esta metodología puede ser aplicada en el desarrollo de nuevos procesos o adecuar por el uso de técnicas estadísticas procesos establecidos, pero que no cuentan con adecuadas herramientas para su monitoreo.

La base de esta investigación es el planteamiento del problema, uso de teorías y modelos estadísticos, formulación de hipótesis, propuesta de método, recopilación y análisis de datos, interpretación de información en base a datos y conclusión.

Como marco teórico se presenta información de el Análisis del Sistema de Medición, sus concepto, así como la interpretación de sus resultados, Amplia sección de conceptos y métodos estadísticos paramétricos y no paramétricos, conceptos de gráficos de control, adicional a esto se incluye una metodología para análisis y solución de problemas.

En anteriores trabajos realizados, no se encuentra el análisis, ni investigación como el que presenta esta tesis, por lo que se consideró necesario incluir el extenso marco teórico presentado; incluyendo figuras, cuadros de referencia, etc.

En la sección de apéndices se encontrarán diversos formatos utilizados en este método, así como tablas de referencia para ser utilizadas en el análisis de datos.

Se incluye la implementación del método en un área piloto con su desglose de actividades paso a paso, resultados y conclusiones. Presentando así un panorama completo del alcance que puede tener este método.

CAPÍTULO 2

INTRODUCCIÓN

2.1. Planteamiento del problema:

En la actualidad la industria automotriz tiene objetivos muy estrictos en cuanto a control de procesos, exige un riguroso control sobre cada característica especial en los componentes de un vehículo, y requiere de cambios rápidos en especificaciones de producto-proceso cuando lo necesite. Aunado a esto solicita planes de reducción de costos, que sólo se puede llegar a cumplir con una alta efectividad en los procesos de producción.

La indefinición de cómo controlar las características especiales en cada producto, de una empresa de la industria automotriz dedicada a la fabricación de componentes de aluminio, resulta en problemas en la interpretación de los resultados obtenidos, dificultando la verificación de su comportamiento.

Como consecuencia de esta indefinición, la falta de control sobre las características especiales del producto es un impedimento para lograr cumplir con las expectativas de la industria.

2.2. Objetivo de la tesis:

Desarrollar una metodología clara y específica para monitorear mediante técnicas estadísticas, las características especiales en un proceso - producto y poder contar con el sistema de medición adecuado para garantizar la confiabilidad en la medición de la variable, de una manera simple, a fin de utilizarla como guía en el desarrollo de procesos.

2.3. Definición de la hipótesis:

La falta de control de las características especiales (en sus respectivos procesos) está altamente relacionado a una inadecuada selección de las técnicas estadísticas a utilizar, además de el sistema de medición.

2.4. Limitaciones del estudio:

Los resultados obtenidos con el Sistema de Control Integral de Procesos estarán basados solamente para el marco teórico de la distribución Normal, cualquier otra, estará fuera de su alcance. Esta metodología propuesta de selección de técnicas estadísticas para control de características especiales está basada en la problemática en una empresa de la industria automotriz dedicada a la fabricación de componentes de aluminio y será aplicada solo en la misma. Por razones de confidencialidad no se mencionan nombres de procesos, cotas, productos, etc. Sólo se muestra en forma ilustrativa el método.

2.5. Justificación del trabajo:

La industria automotriz requiere de un adecuado control sobre las características especiales de calidad en sus productos para asegurar la conformidad del producto terminado.

Además busca la mejora continua en los índices de capacidad de proceso, teniendo como objetivo alcanzar la variación total de sus productos / componentes en niveles de Clase Mundial.

2.6. Metodología a emplear:

Planteamiento del problema.

Uso de teorías y modelos estadísticos.

Formular Hipótesis.

Metodología "Sistema de Control Integral de Procesos".

Recopilar y analizar datos.

Extraer e interpretar información en base a datos.

Conclusión.

2.7. Revisión Bibliográfica:

Los libros Análisis Estadístico y Probabilidad y Estadística para Ingenieros, se tomaron como base para el tratamiento y distribuciones de datos, así mismo fundamentos de pruebas de hipótesis y estadística paramétrica y no paramétrica, base de la metodología propuesta.

El libro Control Estadístico de Calidad, se usa como guía para el análisis e interpretación de gráficas de control.

Se utilizaron apuntes de cursos tomados: Seis Sigma y Metodología de los Sistemas suaves, como complemento para la metodología propuesta en el aspecto de análisis del Sistema de Medición y Análisis y Solución de Problemas.

Tesis elaboradas anteriormente: Implantación de Sistema de Calidad para Manufactura y Aseguramiento de la Calidad a través del Control Estadístico del Proceso, son básicamente referencias bibliográficas de otras propuestas. La tesis Desarrollo de un Sistema de Control Integral de Procesos, propone un método para el correcto uso y funcionamiento de los gráficos de control en la industria.

Para el glosario de términos se usó en Diccionario General de la Lengua Española como referencia.

CAPÍTULO 3

ANÁLISIS DEL SISTEMA DE MEDICIÓN

3.1. Definiciones

Exactitud- ¿El promedio de las mediciones muestra una desviación respecto del valor verdadero?

- El valor verdadero:
 - Valor correcto teórico / estándares.
- Sesgo
 - Distancia entre el valor promedio de todas las mediciones y el valor verdadero.
 - La cantidad, herramienta que está consistentemente fuera del objetivo.
 - Error sistemático o desviación.
- Estabilidad
 - La variación total en las mediciones obtenidas durante un período de tiempo prolongado.
- Linealidad
 - Diferencia en los valores de la tendencia, a través del rango de operación esperado, del calibrador.

Precisión:

- Variación total en el sistema de medición.
- Medida de variación natural en mediciones repetidas.
- Términos: Error al azar, difusión, error en la prueba y en la repetición de la prueba.
- Repetibilidad y reproducibilidad. (R&R del calibrador).

Posibles fuentes de la variación del proceso (Figura 1):

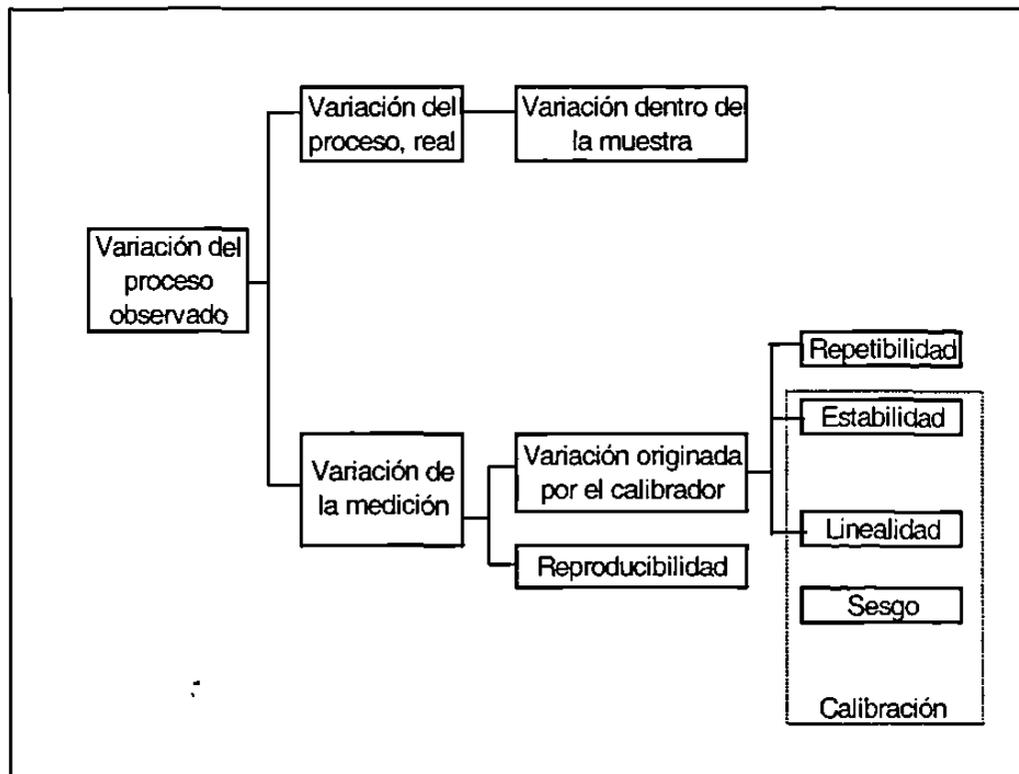


Figura 1. Fuentes de variación del proceso

Definición de la repetibilidad:

Es la variación de las mediciones obtenidas con un instrumento de medición, cuando es utilizado varias veces por un operador, al mismo tiempo que mide las características idénticas en una misma parte. (Figura 2)

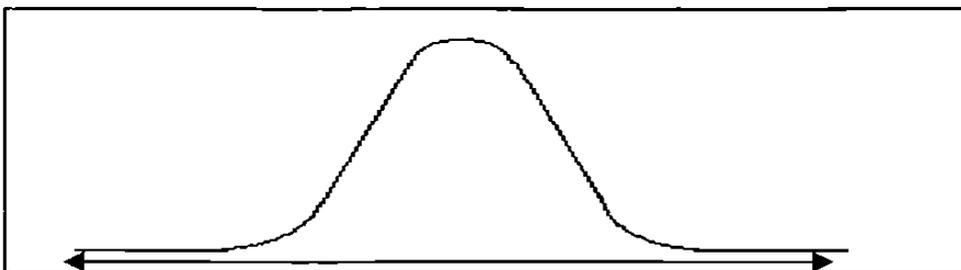


Figura 2 Repetibilidad

Definición de reproducibilidad:

Es la variación, en el promedio, de las mediciones hechas por diferentes operadores que utilizan un mismo instrumento de medición cuando miden características idénticas en una misma parte. (Figura 3)

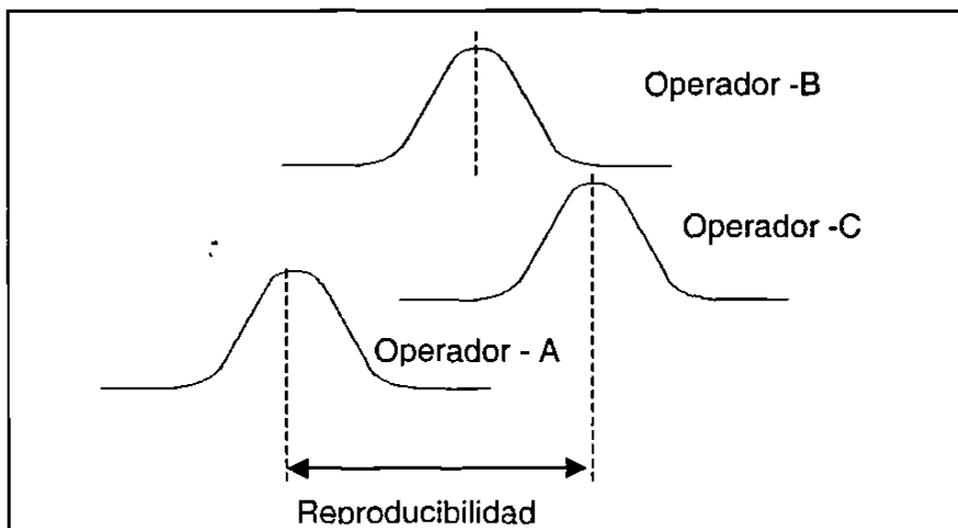


Figura 3 Reproducibilidad

Definición de Sesgo:

Es la diferencia entre el promedio observado de las mediciones y el valor verdadero. (Figura 4)

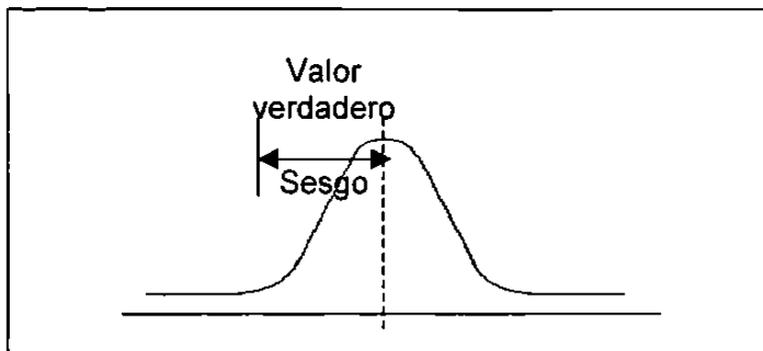


Figura 4 Sesgo

Definición de la estabilidad (o desviación):

Es la variación total de las mediciones obtenidas con un sistema de medición, hechas sobre el mismo patrón o sobre las mismas partes, cuando se mide una sola de sus características, durante un período de tiempo prolongado. (Figura 5)

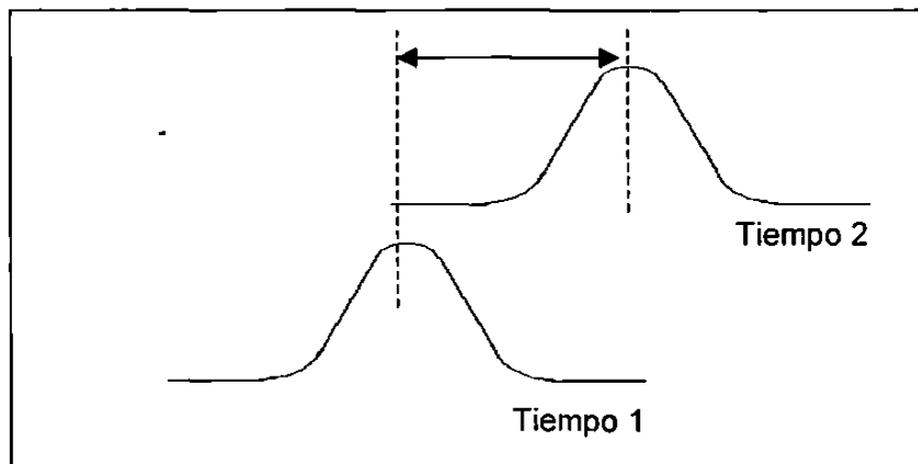


Figura 5 Estabilidad

Definición de Linealidad:

Es la diferencia en los valores del sesgo, a través del rango de operación esperado del calibrador. (Figura 6)

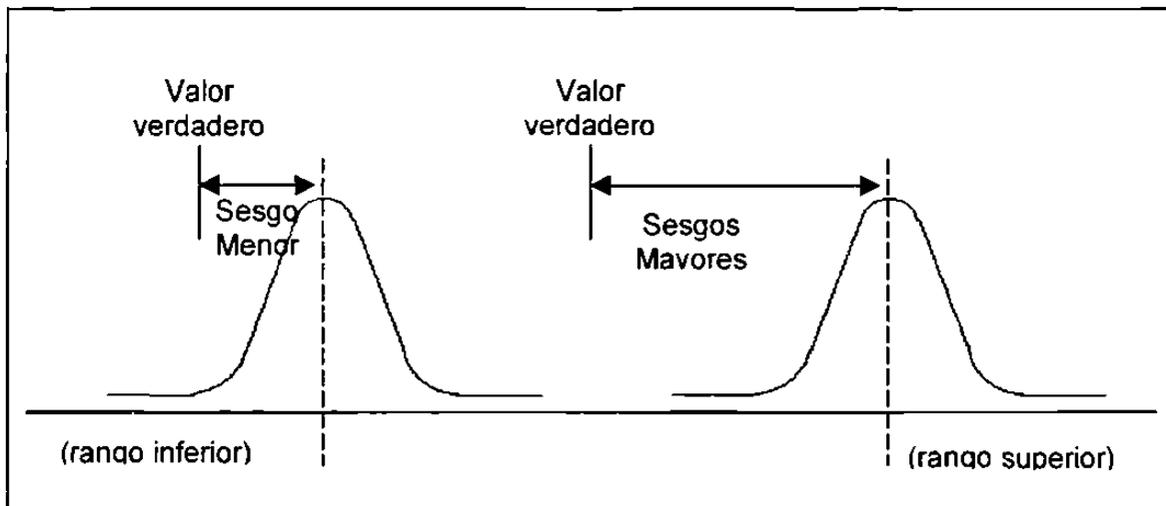


Figura 6 Rango de Operación del Calibrador

Estabilidad del calibrador:

Cómo calcularla:

- Para calibradores que normalmente se utilizan sin ajuste, durante periodos de tiempo relativamente largos.
 - Realizar un segundo estudio R & R del calibrador justo antes de que venza el tiempo de calibración.
 - La estabilidad del calibrador es la diferencia entre los promedios sobresalientes de las mediciones resultantes de los dos estudios.

Posibles causas de poca estabilidad:

- El calibrador no se ajusta tan frecuentemente como se requiere.
- Si es un calibrador de aire, puede necesitar un filtro o un regulador.
- Si es un calibrador electrónico, puede necesitar calentamiento previo.

$$\text{Calibración R \& R} = RPT^2 + REPR^2 \quad (1)$$

- Precisión en relación a la variación total

$$\%R \& R = \frac{R \& R}{\text{Var.Total}} * 100 \quad (2)$$

- Señala qué porcentaje de la variación total debe absorberse como error de medición.

< 10 % Aceptable.

10 - 30% Puede ser aceptable, dependiendo qué tan crítico es el grado de la medición y del número de las distintas categorías.

> 30% Inaceptable.

El valor del R & R es un porcentaje de la variación total del proceso

(Figura 7):

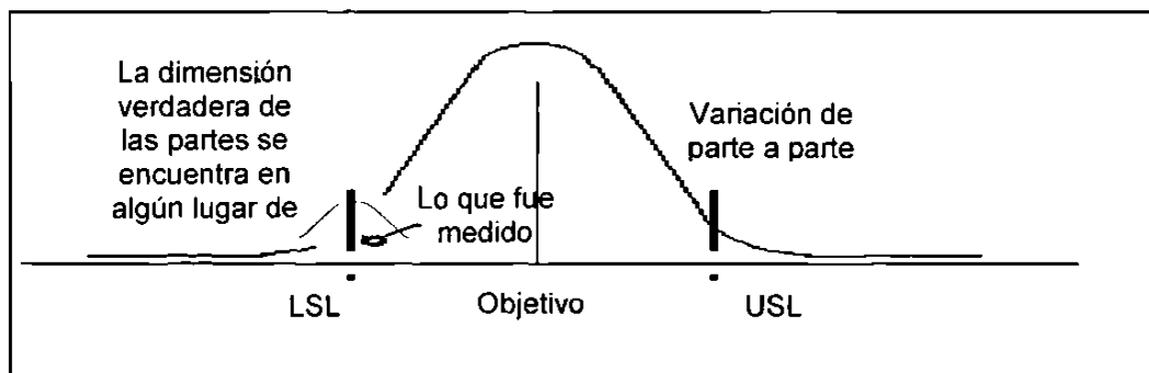


Figura 7 Valor de R & R

Mientras mayor sea el % del R & R, mayor será el área de incertidumbre para conocer la dimensión verdadera de las partes.

Error # 1.- Puede estar aceptando partes que están fuera de especificación.

Error # 2.- Puede estar rechazando partes que están dentro de especificación.

3.2. Llevando a cabo el estudio R & R

Generalmente intervienen de dos a tres operadores.

Generalmente son 10 unidades

Cada unidad es medida por cada operador, 2 ó 3 veces.

Las partes deben seleccionarse al azar, a través del rango total del proceso. Es importante que dichas partes sean representativas del proceso total (80 % de variación).

10 partes no son un tamaño de muestra significativo para una opinión sólida sobre el gage salvo que se cumpla el procedimiento sugerido a continuación.

Procedimiento para realizar un gage R & R

1. Ajustar el calibrador, ó asegurarse de que éste haya sido calibrado.
2. Marcar cada pieza con un número de identificación que no pueda ver la persona que realiza la medición.

3. Hacer que el primer operador mida todas las muestras una sola vez, siguiendo un orden al azar.
4. Hacer que el segundo operador mida todas las muestras una sola vez, siguiendo un orden al azar.
5. Continuar hasta que todos los operadores hayan medido las muestras una sola vez.
6. Repetir los pasos 3 - 4 hasta completar el número requerido de ensayos.
7. Utilizar el formato anexo para determinar las estadísticas del estudio R & R
 - Repetibilidad.
 - Reproducibilidad.
 - % R & R
 - Desviaciones estándar de cada uno de los conceptos mencionados.
 - Análisis del % de tolerancia.
8. Analizar los resultados y determinar los pasos a seguir, si los hay.

Encontrar:

Ancho de tolerancia	X-bar máxima
Numero de ensayos (m)	X-bar mínima
Numero de partes (N)	Diferencia X-bar
Número de operadores	

Alfa $\alpha \Rightarrow$ Para 2 ensayos = 4.56 Beta $\beta \Rightarrow$ Para 2 operadores = 3.65
 Para 3 ensayos = 3.05 Para 3 operadores = 2.7

K3= 1.62

Repetibilidad: La variación del dispositivo de medición (VD) se calcula sobre cada grupo de mediciones tomadas por un operador, en una sola parte.

$$DV = R * \alpha \quad (3)$$

Reproducibilidad: La variación en el promedio de las mediciones (AV) se calcula sobre el rango de los promedios de todas las mediciones, para cada operador, menos el error del calibrador.

$$AV = (Xdif * Beta)^2 - (DV^2 / (m * n)) \quad (4)$$

El componente de varianza para repetibilidad y reproducibilidad (R&R) se calcula combinando la varianza de cada componente.

$$R \& R = DV^2 + AV^2 \quad (5)$$

El componente de varianza para las partes (PV), se calcula sobre el rango de los promedios de todas las mediciones, para cada parte

$$PV = R \text{ part} * K3 \quad (6)$$

La variación total (TV) se calcula combinando la varianza de repetibilidad y reproducibilidad y la variación de la parte.

$$TV = R \& R^2 + PV^2 \quad (7)$$

Basado en la tolerancia:

$$\% DV = 100 * DV / \text{Ancho de tolerancia} \quad (8)$$

$$\% AV = 100 * AV / \text{Ancho de tolerancia} \quad (9)$$

$$\% R \& R = 100 * R \& R / \text{Ancho de tolerancia} \quad (10)$$

Basado en la variación Total:

$$\% DV = 100 * DV / \text{Variación total} \quad (11)$$

$$\% AV = 100 * AV / \text{Variación total} \quad (12)$$

$$\% R \& R = 100 * R \& R / \text{Variación total} \quad (13)$$

$$\% PV = 100 * PV / \text{Variación total} \quad (14)$$

Problemas potenciales de los sistemas de medición:

- Los operadores utilizan diferente metodología para medir las partes.
- Las partes tienen una variación significativa en su interior.
- Resolución insuficiente del calibrador para medir las partes.
- Instrumental inadecuado para medir las partes en forma consistente.
- Problema de calibrador original: el calibrador no es capaz de medir las partes en forma precisa.

Trabajo alrededor del error del gage:

Si se quiere disminuir los errores de el calibrador, tomar ventaja de la raíz cuadrada del error estándar de la muestra:

Básicamente se trata de una técnica de promedio

$$\text{Muestra: } 1/\sqrt{n} \quad n = \text{tamaño de la muestra} \quad (15)$$

Este método se utiliza como una aproximación a corto plazo, para llevar a cabo un estudio , pero se debe ajustar el gage.

Utilizar formato proporcionado en el apéndice 1 (Formato para estudio R&R).

CAPÍTULO 4

MEDIDAS DESCRIPTIVAS PARA DATOS UNIVARIABLES

4.1. Introducción

A menudo, se necesitan medidas descriptivas en forma de números que pueden concentrar mejor la atención en varias propiedades de un conjunto de datos que se investiga.

En general, en el contexto del muestreo, cualquiera función $g(x_1, x_2, \dots, x_n)$ de n observaciones de una muestra se llama una estadística. Funciones paralelas para datos de población se llaman parámetros. Estadísticas y parámetros correspondientes son calculados de datos de muestra y población, respectivamente, con las mismas ecuaciones; por tanto, se toman las siguientes convenciones:

1. Los tamaños de población y muestra se representan por N y n , respectivamente.
2. Los parámetros se representan por letras griegas o mayúsculas, tales como μ y X_5 para media y mediana de población, respectivamente; las estadísticas se representan por letras

minúsculas, tales como \bar{x} y $x_{.5}$ para media y mediana de muestra, respectivamente.

3. Las variables que corresponden a una población o a una muestra se representan por mayúsculas, tales como X, Y, Z; los valores de las variables, por las letras minúsculas correspondientes, tales como x_i , y_i y z_i .

En realidad, raras veces se observan o miden poblaciones enteras, y así, en este punto se definirán todas las medidas sumarias de datos univariados en términos de muestras.

De la definición anterior de una estadística, hay muchas funciones de una muestra que se desea tener. Sin embargo, en términos de análisis estadístico y de aplicación, solo se debe interesar por las cuatro propiedades básicas siguientes que para fines prácticos que a menudo son suficientes para caracterizar las distribuciones de frecuencias de datos univariados:

1. La localización del centro de la distribución, o la medida de tendencia central.
2. El grado de variación de valores individuales alrededor del punto central, o la tendencia de valores individuales a desviarse de la medida de tendencia central.
3. El grado de asimetría: es decir, la falta de simetría de ambos lados del pico (el punto con la mas alta densidad de frecuencias) de una distribución.
4. El grado de variación, o la velocidad con que sube y baja la distribución de izquierda a derecha.

Estas propiedades son significativas especialmente para distribuciones unimodales, pero también se aplican a otros tipos de distribuciones. Para

facilitar la interpretación, se relacionará el estudio principalmente con distribuciones unimodales.

4.2. Medidas de tendencia central

Las medidas de tendencia central se llaman promedios. Un promedio es un valor "típico" en el sentido de que se emplea a veces para representar todos los valores individuales de una serie o de una variable.

Hay muchos tipos de promedios, cada uno de los cuales posee propiedades particulares y cada uno es típico en alguna forma única. Los más frecuentemente encontrados son la mediana, la moda, y la media aritmética. Otros dos promedios, las medias geométrica y armónica, son útiles en algunas situaciones especiales. Sin embargo, se concentrará este estudio en las tres primeras medidas. Iniciando con estas medidas para datos no aplicados.

4.2.1. La mediana

La mediana, representada por x_5 es, como lo indica su nombre, el valor medio de una serie cuando los valores se disponen según su magnitud. Para la serie \$2, \$4, \$5, \$7 y \$8, la mediana es el tercer valor, \$5. Si hay seis valores en una serie, por ejemplo 3, 4, 6, 7, 8 y 10 libras, cualquier valor entre 6 y 7 libras dividiría la serie en dos partes iguales; por tanto, cualquiera de tales valores podría ser la mediana. En la práctica, para un número par de datos, generalmente suponemos que la mediana se encontrará entre los dos valores centrales. Por tanto, en el ejemplo, la mediana sería 6.5 libras. La mediana puede que tenga valores idénticos, con el suyo a ambos lados de él. Por ejemplo, en la serie 1, 2, 3, 4, 5, 5, 5, 6, 7, 8, 9, la mediana es 5.

Debido a estas características, puede definirse ahora formalmente la mediana como aquel valor que divide una serie de tal forma que por lo menos 50 por 100 de los valores son iguales a él o menores que él, y por lo menos 50 por 100 de los valores son iguales o mayores que él. Sin duda, se requiere un conjunto ordenado para localizar el valor de la mediana partiendo de datos no agrupados. Por desgracia, la construcción de un conjunto es a menudo tediosa y lenta.

Como se definió, la mediana no es influida por valores de las colas de una distribución. Por ejemplo, si los datos de muestra 1, 2, 3, 4, 5, cambian por -100, -10, 3, 15, 150, la mediana permanecería en 3. Así, es una medida muy conveniente de localización central para distribuciones asimétricas, tales como al distribución del ingreso, con tal de que no nos importe si la familia más rica, por ejemplo, gana \$500 000 o \$1 000 000 al año.

La mediana también tiene la interesante propiedad de que la suma de las desviaciones absolutas de las observaciones con relación a la mediana es menor que al suma de las desviaciones absolutas con relación a cualquier otro punto de la distribución. En símbolos:

$$\sum_i |x_i - x_{.5}| = \text{un mínimo}$$

y similarmente para datos de población. La mediana es escogida a menudo como una medida de tendencia central debido a esta propiedad.

4.2.2. La moda

La moda, designada por x_m , es aquel valor de una serie que aparece más frecuentemente que cualquier otro. Este valor puede ser descubierto inmediatamente cuando se ordenan los datos. Por ejemplo, en la serie, 1, 2, 3,

4, 4, 5, 6, y 7, la moda es 4. Por consiguiente, se considera la moda como típica en el sentido de que es el valor más "probable" de una serie. La moda para una distribución unimodal (para datos no agrupados) siempre coincide con un valor real de la serie.

Aunque la moda es un concepto sencillo y útil, su aplicación presenta muchos aspectos problemáticos. Primero, una distribución puede revelar que dos o más valores se repiten un número igual de veces, y en tal situación no hay forma lógica de determinar qué valor debe ser escogido como la moda. Hablando en sentido riguroso, cualquier valor se llama una moda si aparece más a menudo que cualquiera de los valores adyacentes. Sin embargo, mientras las frecuencias de los valores modales no sean iguales, se decide escoger el valor con frecuencia más alta como la moda para la serie.

Segundo, puede que no se encuentre ningún valor que aparezca más de una vez.

Tercero, la moda es un valor muy inestable. Puede cambiar radicalmente con el método de redondeo de los datos.

La moda podría ser un valor extremo, como en el caso de una distribución triangular (una distribución en la que la densidad de frecuencias disminuye, o aumenta, continuamente y a un ritmo constante de izquierda a derecha), y entonces difícilmente podría ser considerada como una medida de tendencia central.

Las tres primeras dificultades pueden ser superadas totalmente o en parte cuando los datos se disponen en una distribución por frecuencias. En tal forma, cualquier tendencia de los datos a agruparse alrededor de cierto valor tiende a ser mostrada más claramente, y el problema creado por la primera situación puede ser resuelto. Además, la naturaleza errática de la moda

obtenida de un conjunto debido a diferentes métodos de redondeo puede ser estabilizada en gran parte determinando la moda por la distribución por frecuencias de los mismos datos. Cuando son agrupados en clases, aquellos valores que difieren solo a la derecha de los puntos decimales aún pueden corresponder a la misma clase, cualquiera que sea el método de redondeo empleado. Finalmente, puede aparecer un valor modal en datos agrupados, aunque no haya repetición de ningún valor del conjunto.

4.2.3. La media aritmética

Por su facilidad de cálculo, largo uso y propiedades matemáticas convenientes, es el promedio mejor conocido y de uso más común. A veces, se conoce sencillamente como "la media" o "el promedio", pero deben usarse, siempre adjetivos apropiados cuando el contexto incluye varios tipos de medias.

La media aritmética, representada por \bar{x} , es la suma de los valores individuales de una suma de una muestra dividido por el número de observaciones de la muestra:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i \quad (16)$$

La media tiene algunas interesantes propiedades matemáticas. Primero, es un valor típico porque es el centro de gravedad un punto de equilibrio. También es típica porque su valor puede substituir al valor de cada dato de la serie sin cambiar el total. Esta propiedad puede verse fácilmente por la fórmula para la media. Puesto que $\bar{x} = \sum x_i/n$, se puede derivar, multiplicando ambos miembros de la ecuación por n , el resultado $n\bar{x} = \sum x_i$.

Otra propiedad de la media es que la suma algebraica de las desviaciones con relación a la media (es decir, tomando debida cuenta de los

signos algebraicos) es cero. Simbólicamente $\sum_i (x_i - \bar{x}) = 0$. Esta propiedad se relaciona estrechamente con la anterior, como puede demostrarse:

$$\begin{aligned} (x_1 - \bar{x}) + (x_2 - \bar{x}) + \dots + (x_n - \bar{x}) &= x_1 + x_2 + \dots + x_n - \underbrace{\bar{x} - \bar{x} - \dots - \bar{x}}_{n \text{ veces}} \\ &= \sum_i x_i - n\bar{x} = \sum_i x_i - \sum_i x_i = 0 \end{aligned}$$

La tercera propiedad matemática de la media es que la suma de las desviaciones elevada al cuadrado de los datos de la media es menor que la suma de las desviaciones elevada al cuadrado de cualquier otro punto. Simbólicamente,

$$\sum_i (x_i - \bar{x})^2 = \text{Mínimo}$$

Para concluir el estudio de los promedios, se observa que se verifican las siguientes relaciones entre la media, la mediana y la moda:

1. Para una distribución simétrica y unimodal,

$$\bar{x} = x_5 = x_m$$

2. Para una distribución unimodal positivamente asimétrica, una distribución con una larga cola a la izquierda

$$\bar{x} < x_5 < x_m$$

3. Para una distribución unimodal positivamente asimétrica, una distribución con una larga cola a la derecha,

$$\bar{x} > x_5 > x_m$$

4. Para una distribución unimodal y moderadamente asimétrica,

$$(\bar{x} - x_m) \approx 3 (\bar{x} - x_5)$$

Estas propiedades es ilustran en la figura 8

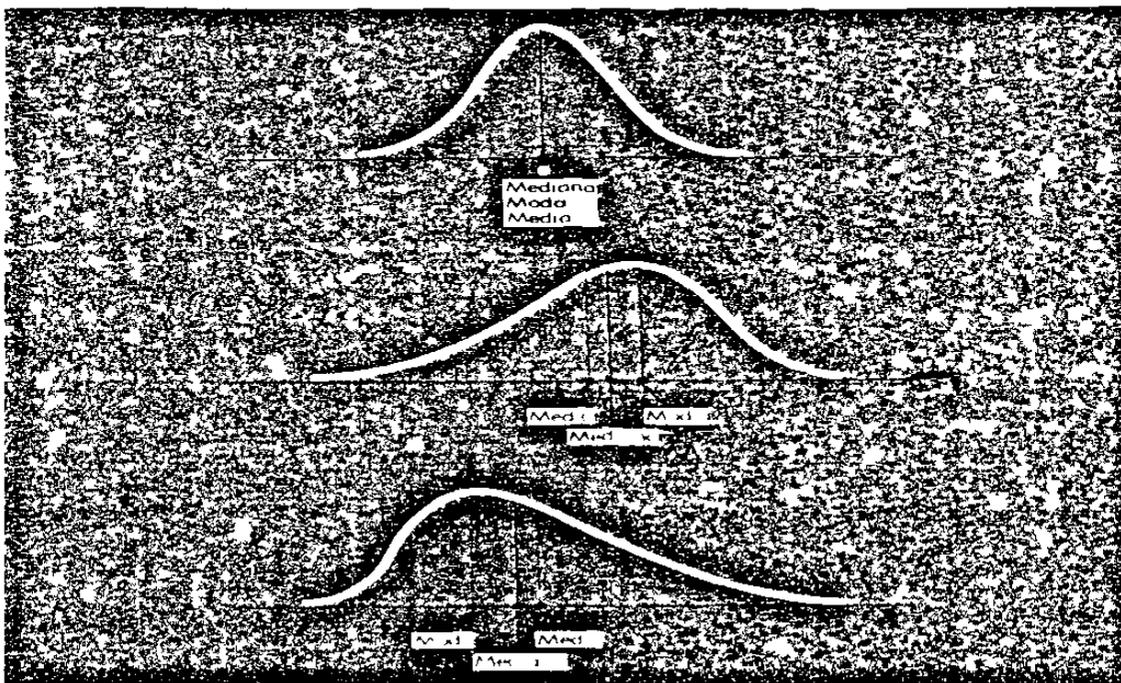


Figura 8 Polígonos de frecuencias hipotéticas que muestran la posición teórica de la moda, la mediana y la media: (a) distribución simétrica, (b) distribución asimétrica a la izquierda, (c) distribución asimétrica a la derecha.

4.3. Medidas de dispersión

La importancia de los promedios a menudo se exagera, un "promedio" sin salvedades puede carecer virtualmente de significado. Un factor que aumenta la confusión es que con algunas distribuciones todos los promedios importantes están estrechamente reunidos, mientras que con otras están muy separados.

4.3.1. La amplitud

La medida más sencilla de dispersión es la amplitud - la diferencia entre los valores máximo y mínimo de la muestra -. El sentido común usa "amplitud" como el máximo y el mínimo de la muestra. El uso estadístico requiere la

diferencia. Para una distribución de frecuencias, la amplitud puede considerarse como la diferencia entre los límites máximo y mínimo de clase, o como la diferencia entre la última y la primera calificaciones de la clase en una tabla de frecuencias.

Juzgando por el método de obtener la amplitud, se observa que no solo es la medida más simple, sino también la más bruta de dispersión. Por consiguiente, la amplitud tiene ciertos defectos. Mas importante, puede ser indebidamente influida por un valor no usual de la muestra. Igualmente, la amplitud no es en modo alguno una medida de dispersión de los datos intermedios con relación al valor típico.

La figura 9 ilustra esto. Finalmente, la amplitud es muy sensible al tamaño de la muestra, la amplitud tiende a cambiar, aunque no proporcionalmente, en la misma dirección en que varía el tamaño de la muestra. Cuando aumenta el número, es posible que algún dato pueda tener mayor valor que el máximo y algún otro dato un valor menor que los valores mínimos de la muestra anterior. Por esta razón, no se puede interpretar la amplitud apropiadamente de la muestra anterior. Por esta razón, se puede interpretar la amplitud apropiadamente sin conocer el número de observaciones.

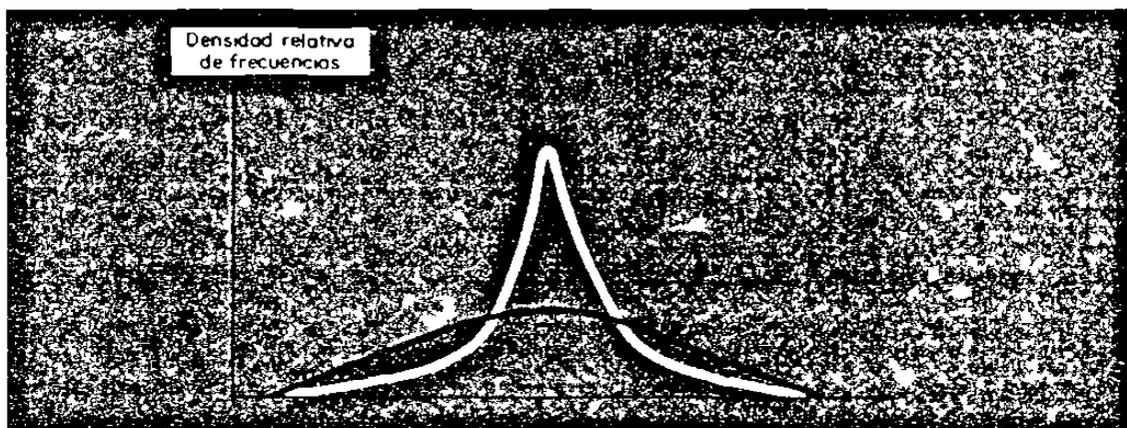


Figura 9 Ilustración de dos distribuciones con la misma amplitud, pero diferente variabilidad

4.3.2. Desviación Media

La búsqueda de una medida de variabilidad que tome en cuenta todos los valores observados y que caracterizaría la dispersión de los valores individuales partiendo de su tendencia central conduce naturalmente a la idea de calcular una medida tal como

$$\sum_i (x_i - \bar{x})$$

pero esta medida sería siempre igual a cero, porque

$$\sum_i (x_i - \bar{x}) = 0$$

y por tanto, difícilmente puede ser considerada como una medida de algo.

Una forma obvia de superar la dificultad es hallar una medida de las desviaciones ignorando la dirección y el signo algebraico correspondiente. Al hacerlo así, obtendríamos lo que se llama la desviación absoluta media, o simplemente la desviación media, de la muestra. Se designará esta medida por dm ; tenemos

$$dm = \frac{1}{n} \sum_i |x_i - \bar{x}| \quad (17)$$

El valor de la desviación media se calcula a veces por la mediana, pero se usará la media. Para una distribución aproximadamente normal, $\bar{x} \pm dm$ es la amplitud que incluirá aproximadamente 58 por 100 de los valores de una distribución. Así, si la desviación media tiene un valor relativamente pequeño, se espera que la distribución sea compacta, porque entonces más de la mitad de los valores de la variable están concentrados dentro de un pequeño espacio alrededor de la media.

La desviación media es útil para tratar situaciones en las que no se requiere un análisis minucioso. Se ha introducido aquí como el escalón lógico para un tratamiento de la variancia, una medida superior de dispersión.

4.3.3. Variancia y desviación estándar

Para superar la propiedad $\sum (x_i - \bar{x}) = 0$, se puede usar cuadrados de las desviaciones a partir de la media, en vez de tomar sus valores absolutos. Las desviaciones al cuadrado no solo previenen obtener cero en la suma; también son mucho más adaptables al análisis matemático que los valores absolutos.

El promedio de las desviaciones al cuadrado de la media se llama variancia de muestra, designada por s^2 . Simbólicamente,

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (18)$$

La variancia de muestra se define a veces también como

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (19)$$

donde $n - 1$ se llama "n - 1 grados de libertad" (Ver CAPÍTULO 6).

La variancia, desde un punto de vista práctico, es un valor muy problemático, porque las unidades asignadas a ella son cuadrados, tales como dólares², libras², etc. Para convertir esta medida de variabilidad en unidades originales, se puede tomar la raíz cuadrada de s^2 ó \hat{s}^2 , obteniendo s ó \hat{s} - la desviación estándar de la muestra.

También ha de hacerse constar que (18) y (19) son ecuaciones de definición de s^2 y \hat{s}^2 . Para facilitar los cálculos se ampliara la suma de desviaciones al cuadrado como sigue:

$$\begin{aligned}\sum_i (x_i - \bar{x})^2 &= \sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_i x_i^2 - 2\bar{x} \sum_i x_i + n\bar{x}^2 \\ &= \sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2\end{aligned}$$

Substituyendo este resultado en (18) se obtiene:

$$\begin{aligned}s^2 &= \frac{1}{n} \sum_i (x_i - \bar{x})^2 \\ &= \frac{1}{n} \left[\sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2 \right] \\ &= \frac{\sum_i x_i^2}{n} - \left(\frac{\sum_i x_i}{n} \right)^2\end{aligned} \tag{20}$$

y substituyéndolo en (19) da

$$\begin{aligned}\hat{s}^2 &= \frac{1}{n-1} \left[\sum_i x_i^2 - \frac{1}{n} (\sum_i x_i)^2 \right] \\ &= \frac{n \sum_i x_i^2 - (\sum_i x_i)^2}{n(n-1)}\end{aligned} \tag{21}$$

Con estas dos últimas ecuaciones se puede calcular la variancia de la muestra simplemente acumulando la suma de los cuadrados de los valores originales y la suma de los valores mismos. Ya no se necesita calcular la media, tomar las desviaciones de la media y elevar al cuadrado dichas desviaciones.

4.3.4. Coeficiente de variación

A menudo, hay interés por comparar las variabilidades entre dos o más conjuntos de datos. Puede hacerse esto fácilmente con sus respectivas variancias o desviaciones estándares cuando las variables se dan en las mismas unidades y cuando sus medias son aproximadamente iguales. Cuando faltan estas condiciones, puede que se desee usar alguna medida relativa de dispersión. Una medida relativa de variabilidad frecuentemente usada se llama coeficiente de variación, designado cv , que es simplemente la razón de la desviación estándar a la media:

$$cv = \frac{s}{\bar{x}} \quad (22)$$

4.3.5. Medidas de asimetría

Dos distribuciones también pueden diferir entre sí en asimetría o agudeza. Las medidas de asimetría y agudeza ganan importancia porque las consideraciones teóricas sobre diferencia estadística y toma de decisiones se basan a menudo en el supuesto de poblaciones normalmente distribuidas. Por tanto, las medidas de asimetría y agudeza son útiles para evitar que, por error, se haga este supuesto.

Existen varias medidas de asimetría, pero solo se mostrará una, que ofrece sencillez de concepto y de cálculo. Esta medida, la mediana pearsoniana de asimetría, se basa en las relaciones entre la media, la mediana y la moda.

Como se mencionó, para una distribución unimodal simétrica, estas tres medidas de asimetría son de idéntico de valor, pero para una distribución asimétrica la media se aleja de la moda hacia la asimetría con la mediana entre ellas. (Ver figura 8). En consecuencia, la distancia entre la media y la moda podría usarse para medir la asimetría. Precisamente,

Asimetría = Media - Moda

Cuanto mayor es esta distancia, negativo o positivo, tanto más asimétrica es la distribución.

Pero tal medida tiene dos efectos en su aplicación. Primero, puesto que es una medida absoluta, el resultado se expresa en unidades originales de la distribución, y así cambia al cambiar la unidad de medición. Segundo, la misma cantidad absoluta de asimetría tiene un diferente significado para distintas series con distintos grados de variabilidad. Para eliminar ambos defectos, se mostrará una medida relativa de asimetría. Esto se logra por el coeficiente pearsoniano de asimetría, designado por sk_p , y expresado simbólicamente como

$$sk_p = \frac{\bar{x} - x_m}{\hat{s}} \quad (23)$$

Pero la aplicación de (23) supone otra dificultad, que surge porque el valor modal de muchas distribuciones solo es una aproximación, pero el lugar de la mediana se encuentra más satisfactoriamente. Como se observó en la sección de medidas de tendencia central, en las distribuciones moderadamente asimétricas, se verifica la relación

$$x_m = \bar{x} - 3(\bar{x} - x_5)$$

De esto se nota que

$$\bar{x} - x_m = \bar{x} - [\bar{x} - 3(\bar{x} - x_5)] = 3(\bar{x} - x_5)$$

Con este resultado, la ecuación (23) puede escribirse ahora

$$sk_p = \frac{3(\bar{x} - x_5)}{\hat{s}} \quad (24)$$

Por esta medida, sk_p sería cero para una distribución simétrica; negativa para una distribución asimétrica a la izquierda; positiva para una distribución asimétrica a la derecha. Más precisamente, esta medida varía dentro de los límites de ± 3 ; sin embargo en realidad, solo en raras ocasiones el valor de sk_p supera los límites de ± 1 .

En conclusión, pueden mencionarse algunos puntos interesantes sobre la asimetría. Las curvas en forma de J y J invertida son ejemplos de asimetría extrema a la izquierda y a la derecha, respectivamente. Es muy común encontrar distribuciones asimétricas positivas en datos de economía y comercio, particularmente en series de producción y precios, que solo pueden ser cero, pero pueden ser infinitamente grandes. Se cree que la asimetría positiva es producida por fuerzas multiplicadoras. Las distribuciones asimétricas negativas son muy raras, y a menudo es difícil ofrecer una explicación racional de su existencia.

4.3.6. Curtosis: una medida de agudeza

En esta sección se presenta una medida de agudeza, el coeficiente de curtosis, representado por k . Esta medida es algebraicamente tratable y geoméricamente interpretable. Se define como la amplitud semiintercuartil y la amplitud 90 - 10 percentil:

$$k = \frac{1}{2} \frac{(x_{75} - x_{25})}{x_{90} - x_{10}} \quad (25)$$

Por medio del coeficiente de curtosis, se pueden clasificar diferentes grados de agudeza en las tres categorías de leptocúrtica, platicúrtica y mesocúrtica (ver figura 10). Una distribución leptocúrtica (curva a) tiene concentradas la mayoría de sus mediciones en el centro. Por ello, la diferencia entre las dos distancias $(x_{75} - x_{25})$ y $(x_{90} - x_{10})$ tiende a ser muy pequeña. Precisamente, dado un grado específico de dispersión, cuanto más agudo es el

pico tanto menor es la diferencia entre estas dos distancias. Puesto que $1/2 (x_{75} - x_{25}) < (x_{90} - x_{10})$ se verifica siempre, para una curva muy aguda, k se aproxima a 0.5 como límite cuando $x_{75} - x_{25} = x_{90} - x_{10}$. En contraste, cuanto más platicútrica es la distribución (curva b), tanto más la amplitud 90 - 10 percentil tiende a superar la amplitud intercuartil. Así cuando al amplitud de una variable se aproxima al infinito y para una curva completamente plana, k se aproxima a cero. En vista de estas consideraciones, parece razonable tomar valores cercanos a ambos lados de 0.25 para representar mesocurtosis (curva c). La elección es reforzada por el hecho de que para cualquier "variable normalmente distribuida", $k = 0.2630$.

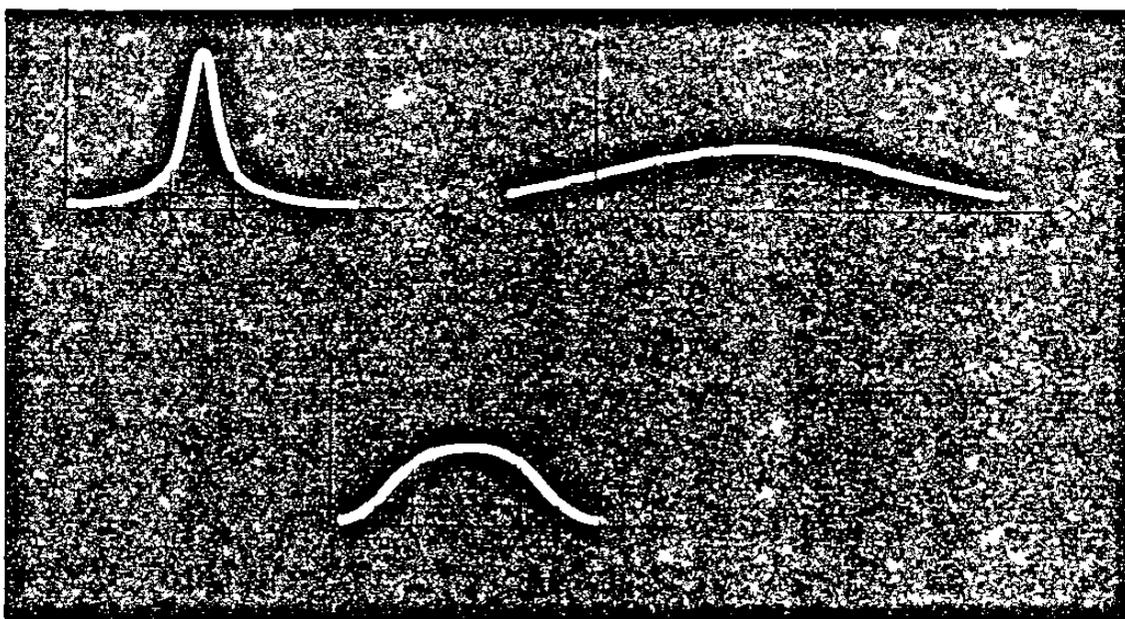


Figura 10. Los tres tipos de kurtosis: a) leptocúrtica, b) platicúrtica, c) mesocúrtica

CAPÍTULO 5

MODELOS DE PROBABILIDADES

5.1. Modelo normal general

El modelo de probabilidad más frecuentemente usado en el análisis económico y comercial es la distribución normal, que puede expresarse en las formas general y estandarizada. Se dice que una variable aleatoria X tiene una distribución normal general si es continua, si existen las constantes $\mu, -\infty < \mu < \infty$ y $\sigma, \sigma > 0$, y si la función densidad es dada por la siguiente expresión:

$$n(x; \mu, \sigma) = n(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty \quad (26)$$

donde μ y σ son, respectivamente, la media y la desviación estándar de la variable normal (los dos parámetros de la distribución normal), $e = 2.718$ y $\pi = 3.142$. Para comprender esto analizar la figura 11. (las relaciones numéricas mostradas en el gráfico aquí han sido redondeadas, no son exactas. Si se desea mayor precisión, ver apéndice 2 [Valores de la función distribución normal estándar]).

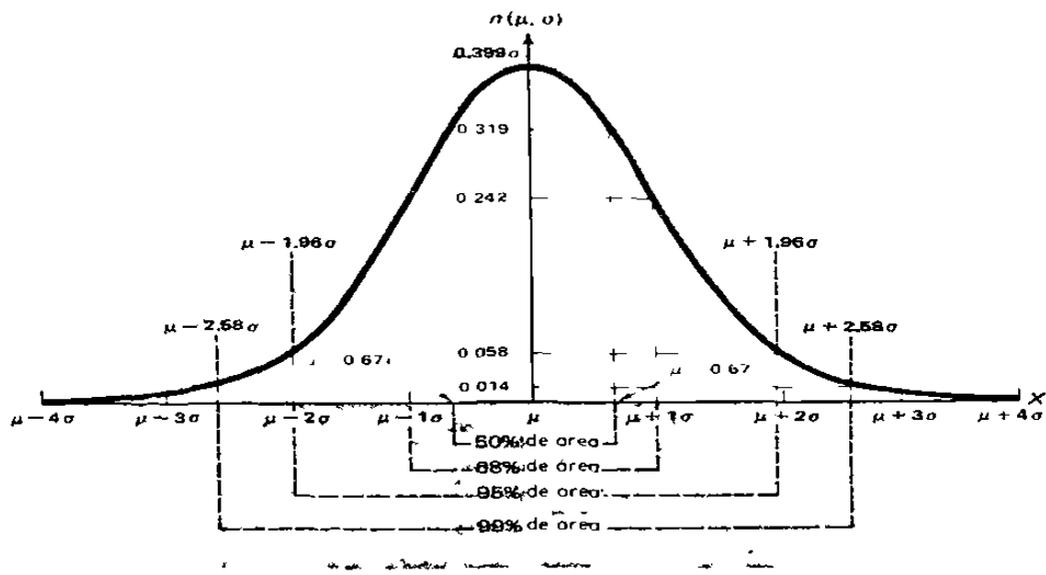


Figura 11. Gráfico de la función general de densidad normal

Primero, una distribución normal tiene como parámetro μ y σ en el sentido de que el área bajo la curva de densidad (probabilidad normal) es definida completamente por los valores de μ y σ . Como tal, la parte práctica de la densidad normal es el exponente $-(x - \mu)^2 / 2\sigma^2$, porque contiene un valor particular de la variable normal, X , y los parámetros μ y σ de la distribución. Cuanto mayor es la desviación de un valor particular de X con relación a μ , tanto menor (es decir, más negativo) es el numerador de este exponente. La desviación es elevada al cuadrado; por tanto, dos valores diferentes de X que muestran la misma desviación absoluta de μ tienen la misma densidad de probabilidad. Esto refleja el hecho de que una distribución normal es simétrica alrededor de μ : $n(\mu - x; \mu, \sigma) = n(\mu + x; \mu, \sigma)$. Por ejemplo,

$$\begin{aligned} P(\mu - 0.67\sigma < X < \mu) &= P(\mu < X < \mu + 0.67\sigma) \\ &= 0.25 \end{aligned}$$

y

$$P(\mu - 0.67\sigma < X < \mu + 0.67\sigma) = 0.50$$

$$P(\mu - \sigma < X < \mu) = P(\mu < X < \mu + \sigma) \\ = 0.34$$

y

$$P(\mu - \sigma < X < \mu + \sigma) = 0.68 \\ P(\mu - 1.96\sigma < X < \mu) = P(\mu < X < \mu + 1.96\sigma) \\ = 0.475$$

y

$$P(\mu - 1.96\sigma < X < \mu + 1.96\sigma) = 0.95$$

Y así sucesivamente.

Segundo, el hecho de que el exponente $-(x - \mu)^2 / 2\sigma^2$ es negativo indica que cuanto mayor es la desviación de X con relación a μ , tanto menor es la densidad de probabilidad de X . Es decir, ambas colas de distribución normal experimentan densidad decreciente, porque cuanto más se aleja X de μ , tanto menor es la altura de la curva de función densidad. A este respecto se observa que, cuando el valor de X es idéntico a μ , el exponente es cero, y así la densidad es $1/\sigma\sqrt{\pi}$, el valor más grande de la densidad normal. Así, la distribución normal es unimodal, con el valor modal en $X = \mu$.

Tercero, la distribución normal tiene una amplitud infinita, por lo que su curva de densidad nunca toca el eje X . Como consecuencia, cualquier intervalo de números tendrá una probabilidad positiva. Sin embargo, la probabilidad de un intervalo muy alejado de μ es despreciablemente pequeña, como se puede observar, el hecho de que la curva de densidad se aleja rápidamente, siendo encerrada más de 99 por 100 de su área por $\mu \pm 3\sigma$. Esta propiedad nos permite usar la distribución normal para aproximar otras distribuciones para las cuales la verdadera amplitud es infinita.

Cuarto, un cambio en el valor de μ desplaza toda distribución normal a la derecha o a la izquierda, mientras que un cambio en el valor de σ altera su forma sin moverla a la derecha o a la izquierda. Estos hechos indican que la distribución normal es realmente una familia de distribuciones. (Ver figura 12). Por supuesto, es cierto para muchas otras distribuciones.

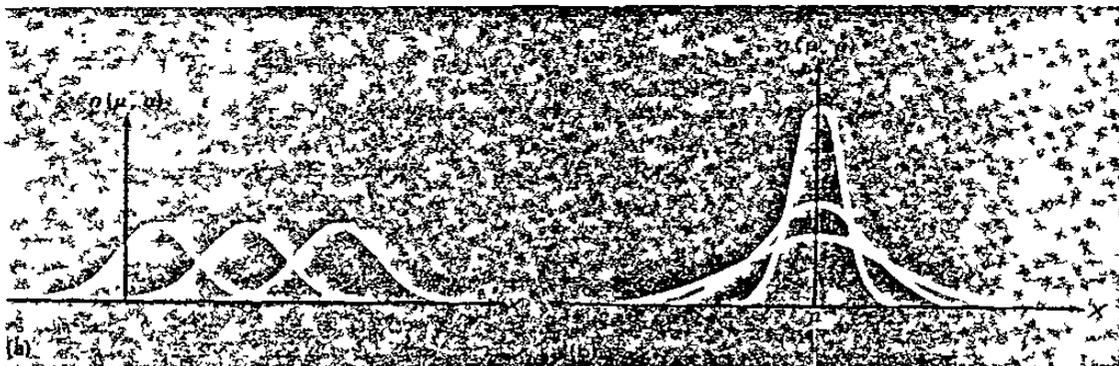


Figura 12 Distribuciones normales: a) con la misma desviación estándar y con diferentes medias; b) con la misma media y con diferentes desviaciones estándar

Quinto, una transformación lineal de una variable normal resulta en una nueva variable normal. Es decir, si X es una variable normal, entonces $Y = a + bX$ es también una variable normal.

Finalmente, si X_1, X_2, \dots, X_n son variables normales independientes, su suma, S , es también una variable normal. Además debido a la independencia, la propiedad de aditividad se verifica para la expectativa y la variancia en este caso. Es decir, la expectativa de S es la suma de las expectativas de las n variables normales. Igualmente, la variancia de S es la suma de las variancias de las n variables normales.

La distribución normal fue descubierta por De Moivre como la forma limitante del modelo binomial, en 1733. También fue conocida por Laplace no mucho después de 1774, pero por un error histórico ha sido acreditada a Gauss, quién primero hizo referencia a ella en 1809. Durante los siglos XVIII y

XIX se hicieron varios esfuerzos para establecer el modelo normal como la ley básica que rige todas las variables aleatorias continuas: de ahí el nombre "normal". Estos esfuerzos fracasaron. Pero el modelo normal se ha convertido en el más importante modelo de probabilidad en análisis estadístico, por varias razones.

Primero, muchas variables aleatorias continuas, tales como estaturas de varones adultos, inteligencia de niños de edad escolar, diámetros de neumáticos de automóvil de cierta marca, resistencias a la tracción de alambres de acero producidos por cierto proceso, etc. son distribuidas en forma aproximadamente normal debido a lo que se conoce, o que se presume es cierto, de las mediciones mismas. A este respecto, es particularmente interesante observar que "errores" de mediciones repetidas de una dimensión dada se supone por hipótesis que siguen la ley de probabilidades normal. Cualquier medición u observación se supone que representa una magnitud verdadera más un error. Cada error tiene una magnitud propia, resultante de un vasto conjunto de factores que operan en este momento. Cada factor solo ejerce un pequeñísimo efecto en el tamaño y la dirección del error; además, los errores de medición actúan independientemente con igual fuerza para empujar la medición observada hacia arriba o abajo, y, por consiguiente, se anulan a largo plazo. Así, se consideran los errores de medición como reflejos de variaciones casuales que son normalmente distribuidas con expectativa cero. Estas se conocen a menudo como "errores al azar" en laboratorios, a diferencia de "errores sistemáticos", que son muy diferentes.

Segundo, la distribución normal sirve como una buena aproximación para muchas distribuciones discretas.

Tercero, en estadística teórica, muchos problemas pueden ser resueltos fácilmente en el supuesto de una población normal. En trabajo aplicado, se encuentra a menudo que métodos elaborados según la ley de probabilidades

normal dan resultados satisfactorios, aunque no se cumple totalmente el supuesto de una población normal.

Finalmente, las distribuciones de muchas estadísticas de muestra calculadas partiendo de grandes muestras se aproximan a la distribución normal como un límite.

5.2. El modelo normal estándar

A menudo es más fácil trabajar con distribuciones de probabilidades estandarizadas. Esto es especialmente cierto cuando se tiene una función densidad, tal como la de la variable normal, que comprende "integrales" que no pueden ser reducidas a funciones elementales. Para rapidez y eficiencia en el trabajo práctico, el concepto de distribución normal estándar es realmente indispensable.

Se dice que una distribución acumulativa es de la forma estándar si media es cero, y su variancia, y por consiguiente, la desviación estándar, es la unidad. De la función densidad de la distribución normal general, se puede observar que la densidad de probabilidades de la variable normal estándar, Z , debe ser:

$$: \quad n(z;0,1) = n(0,1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (27)$$

La función de distribución acumulativa (FDA) correspondiente a la densidad estándar, $n(0,1)$, como lo hace cualquier otra FDA, da la probabilidad de que la variable normal estándar asuma un valor igual o menor que z , o es simplemente $F(z) = P (Z \leq z)$. Los gráficos de la Función Densidad de Distribución y la FDA de la distribución normal estándar se representa en la figura 13.

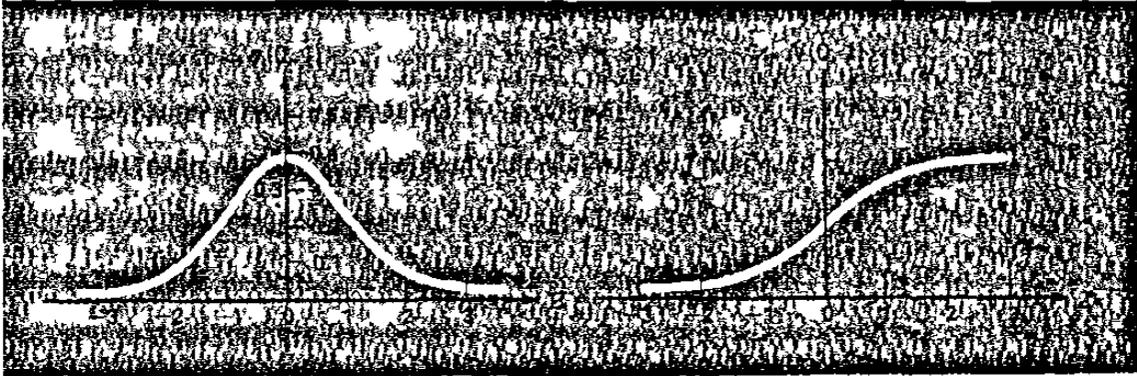


Figura 13 Densidad Normal estándar y funciones de distribución acumulativa

Ahora que si X es $n(\mu, \sigma)$, entonces $Z = (X - \mu)/\sigma$ es $n(0, 1)$. Como se mencionó, esta transformación de X en Z produce el efecto de reducir X a unidades en términos de desviaciones estándares alejadas de la media. En otras palabras, dado un valor de X , el correspondiente valor de Z nos dice cuán alejada está X de su media μ , y en qué dirección, en términos de su desviación estándar, σ . Por ejemplo, $Z = 1.5$ significa que el valor particular de X está 1.5σ arriba (a la derecha) de μ . Análogamente, $Z = 2$ significa que el valor particular de X está 2σ debajo (a la izquierda) de μ . Esta propiedad de la variable normal estándar nos permite evaluar probabilidades normales para cualquier $n(\mu, \sigma)$ de una sola tabla de probabilidades para $n(0, 1)$. Precisamente, cualquier normal general FDA, designada por $N(\mu, \sigma)$, puede ser convertida en una normal estándar FDA, designada por $N(0, 1)$, por la transformación Z como sigue:

$$\begin{aligned}
 N(x; \mu, \sigma) &= P(X \leq x) \\
 &= P[(\mu + \sigma Z) \leq x] \\
 &= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \\
 &= N\left(\frac{x - \mu}{\sigma}; 0, 1\right)
 \end{aligned} \tag{28}$$

Así, dados $N(\mu, \sigma)$ y dos números reales cualesquiera a y b , con $a < b$, tendríamos

$$\begin{aligned} P(a \leq X \leq b) &= N(b) - N(a) \\ &= N\left(\frac{b-\mu}{\sigma}\right) - N\left(\frac{a-\mu}{\sigma}\right) \end{aligned} \quad (29)$$

Se observa que el primer miembro (29) puede tener cualquiera de varias formas exactamente iguales:

$$\begin{aligned} P(a < X < b) &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a \leq X \leq b) \end{aligned}$$

Estas igualdades se verifican porque X tiene una función continua de densidad de probabilidad. Por esta razón, la probabilidad de que $X =$ una constante es cero, aunque puede que la densidad de probabilidad no sea cero. Por consiguiente, $P(X = a) = 0$ y $P(X = b) = 0$, por lo que la probabilidad de que X se encuentre en el intervalo de a y b es igual, tanto si a es incluida como si es excluida, y si b es incluida o excluida. Por supuesto, esta situación es muy diferente de aquella en la que una variable aleatoria es discreta.

Un cuadro de $N(0,1)$ se da como el apéndice 2.

CAPÍTULO 6

DISTRIBUCIONES CHI CUADRADO, F Y t DE STUDENT

6.1. Teoría del muestreo exacto

Hasta ahora se ha tratado de hacer inferencias de muestras grandes. El concepto "muestra grande" es relativo. Se relaciona con la distribución por muestreo de la estadística que se considera. En general, la teoría del muestreo grande se refiere a una clase de situaciones en las que la distribución por probabilidades de una estadística de muestra es normal o aproximadamente normal, porque la población que se investiga es normal o porque el tamaño de la muestra es suficientemente grande para que el teorema de límite central sea operativo.

Pero, a veces, puede que se tenga que hacer inferencias cuando los supuestos subyacentes de la teoría del muestreo grande no pueden cumplirse. Pueden surgir situaciones en las que la desviación estándar de la población es desconocida porque se tiene un nuevo problema o a una teoría, o el tamaño de la muestra es pequeño debido a limitaciones físicas como las relacionadas con

la investigación médica o a limitaciones prácticas como los altos costos de las observaciones de muestra.

Aunque con muestras pequeñas no se tiene tanta información como se desea, no se carece completamente de recursos. El curso apropiado es extraer conclusiones o tomar decisiones de la muestra tomando en cuenta la naturaleza escasa de las pruebas. El estudio de inferencias estadísticas con muestras pequeñas se llama teoría de muestreo pequeño, o teoría del muestreo exacto. Como se mencionó, la principal diferencia entre teoría del muestreo grande y pequeño es entre distribuciones por muestreo: para muestras grandes, las distribuciones por muestreo son normales; para muestras pequeñas, la distribución por muestreo difiere de un caso a otro. Hay tres distribuciones de probabilidades que a menudo son asumidas por una estadística con n pequeña: las distribuciones chi cuadrado, F y t de Student. Los tres modelos se relacionan con el modelo de probabilidad normal y se definen por el "numero de grados de libertad".

6.2. Grados de libertad

La noción de "grados de libertad" se mencionó primero en relación con la definición de \hat{s}^2 en el punto 4.3.2. hasta ahora no se ha explicado el significado de este concepto.

El concepto de grados de libertad es un concepto matemático. Es el nombre dado al número de observaciones linealmente independientes que ocurren en una suma de cuadrados. Una rigurosa discusión de esta definición usa ideas moderadamente avanzadas de álgebra lineal.

En vez de dedicar espacio a estos temas, simplemente se abordará la definición heurísticamente.

Se trata de estimar una medida de población por una muestra de tamaño 1. El valor de la muestra constituiría la estimación, y cualquier inferencia extraída debe basarse necesariamente en este único valor. Ahora, si se trata de calcular la variancia de la muestra con la misma muestra, no se puede hacer. Esto equivale a decir que una observación no proporciona información acerca de la variancia. Para conocer algo sobre la variancia, se necesitan por lo menos una muestra con $n \geq 2$. Ahora, si $n = 2$, uno u otro de los dos valores no dice nada acerca de la variancia: solo un segundo valor proporciona alguna información. Así la variancia se basa en solo un dato de los dos datos de la muestra. En este caso, solo hay $2 - 1 = 1$ grados de libertad. Aquí, en el cálculo de la variancia, el primer valor no proporciona información, y por tanto, se pierde un grado de libertad. Un grado de libertad es perdido de una muestra de tamaño n , por lo que el número de grados de libertad es $n - 1$.

Alternativamente, se observa que cuando se calcula la variancia de la muestra en términos de desviaciones medidas partiendo de la media de la muestra, \bar{x} , se obtiene una suma de n desviaciones al cuadrado $\sum (x_i - \bar{x})^2$, que debe obedecer la relación secundaria o restricción $\sum x_i = n\bar{x}$. Esto implica que, si tenemos la base \bar{x} y cualesquiera desviaciones $n - 1$ de ella, la desviación n -ésima es determinada automáticamente. Es linealmente dependiente de otras y, por tanto, no es "libre" para variar. Si las desviaciones hubieran tomado una μ conocida, la n -ésima desviación seguiría siendo linealmente independiente de las otras $n - 1$ observaciones. En consecuencia, una muestra de n desviaciones al cuadrado tomada alrededor de \bar{x} proporciona menos información útil que el mismo número de desviaciones medida a partir de μ .

Cuando se dispone de dos o más variancias de muestra calculadas de muestras al azar independientes (tomadas de la misma población), a menudo se desea utilizar toda la información para estimar la variancia de la población

calculando una variancia ponderada, o fusionada. Considerando, por ejemplo, el caso de dos muestras independientes. Un estimador insesgado de la variancia de la población puede obtenerse como

$$\hat{s}_w^2 = \frac{(n_1 - 1)\hat{s}_1^2 + (n_2 - 1)\hat{s}_2^2}{n_1 + n_2 - 2}$$

Para el cuál el número de grados de libertad es $n_1 + n_2 - 2$. esto se debe a que \hat{s}_1^2 tiene n_1 comparaciones lineales, pero una medida dada, y por tanto, $n_1 - 1$ comparaciones linealmente independientes; análogamente para \hat{s}_2^2 . Por consiguiente, el número de comparaciones lineales independientes para \hat{s}_2^2 definida antes es $(n_1 - 1) + (n_2 - 1)$.

Para generalizar: el número de grados de libertad representado por δ , puede considerarse como el número de elementos que pueden escogerse libremente; o como el número de variables que pueden variar libremente; ó como el número de variables independientes. Aquí, la independencia es funcional, no estadística. Dado el tamaño de la muestra, el número de grados de libertad es $\delta = n - k$, dónde k es el número de restricciones para los cálculos de una estadística $\hat{\theta}$ que abarca sumas de cuadrados, y las restricciones pueden ser, por ejemplo, el número de estimadores requeridos para calcular la $\hat{\theta}$ en cuestión.

6.3. Distribuciones Chi cuadrado

Si $X_1, X_2, \dots, X_\delta$ son variables normales estándares independientes, la suma de los cuadrados de estas variables se dice que es una variable chi cuadrado, X^2 , con δ grados de libertad. Es decir,

$$X_\delta^2 = X_1^2 + X_2^2 + \dots + X_\delta^2 \quad (30)$$

Hay un número infinito de distribuciones chi cuadrado, una correspondiente a cada entero positivo δ . Una distribución chi cuadrado correspondiente al número δ se designará por χ^2_δ y se define por la siguiente densidad:

$$f(X^2) = \frac{(X^2)^{(\delta/2-1)} e^{-(X^2/2)}}{(2^{\delta/2}) \Gamma\left(\frac{\delta}{2}\right)} \quad 0 \leq X^2 \leq \infty \quad (31)$$

En esta expresión, la cantidad $\Gamma\left(\frac{\delta}{2}\right)$, que se lee "función gamma de $\delta/2$ " se aplica a un valor que es independiente de δ y es $[(\delta/2 - 1)!]$. Esta función de densidad hace que la variable normal parezca simple. Por fortuna, solo se requiere una comprensión completa de ella en la estadística teórica.

Ahora se enunciarán las propiedades más importantes de las distribuciones chi cuadrado para su uso posterior. Algunas de estas propiedades pueden ser deducidas de un detenido estudio de la figura 14, que da las curvas de densidad de distribuciones chi cuadrado para unos pocos valores escogidos de δ .

- 1.- Si X tiene una distribución normal estándar, X^2 tiene una distribución χ^2 con $\delta = 1$ - es decir, χ^2_1 .
- 2.- Si X_1 es $\chi^2_{\delta_1}$ y X_2 es $\chi^2_{\delta_2}$, y si X_1 y X_2 son independientes, entonces $X_1 + X_2$ es $\chi^2_{\delta_1 + \delta_2}$. A esto se le llama comúnmente propiedad aditiva de las distribuciones chi cuadrado.
- 3.- Si X es una variable normal estándar y X_i , $i=1,2,\dots,n$, son n observaciones que constituyen una muestra al azar, siendo cada x_i un valor de X , entonces $\sum_i X_i^2$ se distribuye como χ^2_n .
- 4.- Si X es una variable normal y si x_i son n observaciones que constituyen una muestra al azar, siendo cada x_i un valor de X , entonces $\sum_i [(x_i - \mu) / \sigma]^2$ se distribuye como χ^2_n .

Se observa que las propiedades 3 y 4 son expresiones equivalentes. En la propiedad 3 una muestra al azar es escogida de $n(0, 1)$; en la propiedad 4 se toma una muestra al azar de $n(\mu, \sigma)$, y la operación de $(x_i - \mu) / \sigma$ transforma la propiedad 4 en la propiedad 3.

5.- Una variable chi cuadrado varía en valor de 0 a infinito, porque es la suma de cuadrados valuados.

6.- Una distribución chi cuadrado se define completamente por el número de grados de libertad. Si X es X_{δ}^2 , su media y su variancia, respectivamente, son

$$E(X_{\delta}^2) = \mu = \delta \quad (32)$$

$$V(X_{\delta}^2) = \sigma^2 = 2\delta \quad (33)$$

7.- Las distribuciones chi cuadrado son positivamente asimétricas. Sin embargo, cuando aumenta δ , se aproxima a la distribución normal $n(\delta, \sqrt{2\delta})$. En la práctica, las probabilidades chi cuadrado, cuando $d > 30$, pueden calcularse empleando aproximadamente normales en la forma usual.

En el apéndice 3 se muestran algunos puntos fractiles escogidos para distribuciones chi cuadrado para $1 \leq \delta \leq 100$. Se observa que esta tabla está en forma acumulativa y da probabilidades de cola superior para distribuciones chi cuadrado.

$$P(X > 7.26) = P(7.26 < X_{15}^2 < \infty) = 0.95$$

$$P(X > 5.23) = P(5.23 < X_{15}^2 < \infty) = 0.99$$

$$P(X > 25.00) = P(25.00 < X_{15}^2 < \infty) = 0.05$$

Y así sucesivamente.

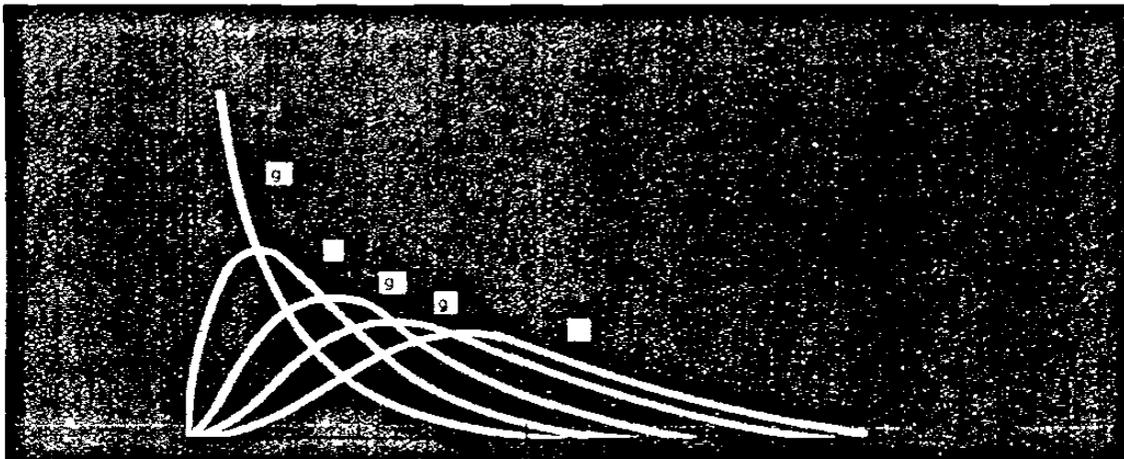


Figura 14 Curvas de distribución Chi cuadrado para grados de libertad escogidos

6.4. Distribuciones F

Si $X_1, X_2, \dots, X_{\delta_1}$ y $Y_1, Y_2, \dots, Y_{\delta_2}$ son variables normales estándares independientes, la estadística

$$F = \frac{(X_1^2 + X_2^2 + \dots + X_{\delta_1}^2)/\delta_1}{(Y_1^2 + Y_2^2 + \dots + Y_{\delta_2}^2)/\delta_2} \quad (34)$$

se dice que tiene una distribución F con (δ_1, δ_2) grados de libertad. El numerador y el denominador de (34) son independientes. Resulta que una variable F es una razón de dos variables chi cuadrado con δ_1 como el número de gl (grados de libertad) para el numerador y con δ_2 como el número de gl para el denominador. Así, una distribución F puede designarse como F_{δ_1, δ_2} y se define por densidad

$$f(F) = \left[\frac{\Gamma\left(\frac{\delta_1 + \delta_2}{2}\right)}{\Gamma\left(\frac{\delta_1}{2}\right)\Gamma\left(\frac{\delta_2}{2}\right)} \right] \left(\frac{\delta_1}{\delta_2} \right)^{\delta_1} (F^{\delta_1/2})^{-1} \left(1 + \frac{\delta_1 F}{\delta_2} \right)^{-(\delta_1 + \delta_2)/2} \quad 0 \leq F \leq \infty \quad (35)$$

La única cosa importante que se debe advertir acerca de esta formidable expresión es que una distribución F tiene dos parámetros, δ_1 y δ_2 . Curvas de densidad para unas pocas distribuciones F escogidas aparecen en la figura 15.

Algunas propiedades importantes de las distribuciones F son las siguientes:

1. Siendo una razón de dos cantidades al cuadrado, una variable F varía de valor de 0 a ∞ .
2. Hay una distribución F por cada par de enteros positivos δ_1 y δ_2 .

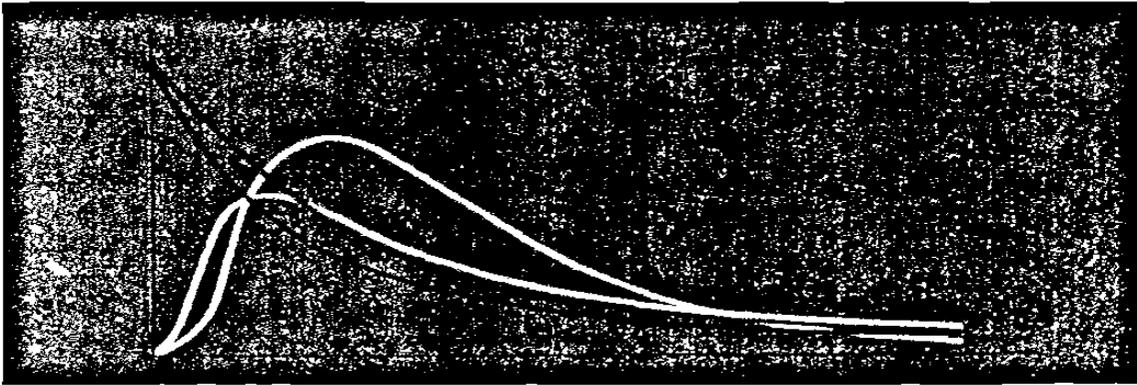


Figura 15. Curvas de distribución F

3. La media y la variancia de F_{δ_1} y δ_2 , respectivamente son

$$E(F) = \frac{\delta_2}{\delta_2 - 2} \quad \text{Para } \delta_2 > 2; \quad (36)$$

$$V(F) = \frac{2\delta_2^2(\delta_1 + \delta_2 - 2)}{\delta_1(\delta_2 - 2)^2(\delta_2 - 4)} \quad \text{Para } \delta_2 > 4; \quad (37)$$

Estas ecuaciones implican que una variable F no tiene media cuando $\delta_2 \leq 2$ y que no tiene variancia cuando $\delta_2 \leq 4$.

4. Como la distribución chi cuadrado, una distribución F es positivamente asimétrica; pero su asimetría se reduce con los aumentos de δ_1 y δ_2 .
5. Si X es F_{δ_1, δ_2} , entonces $Y = 1/X$ es básicamente F_{δ_2, δ_1} . Esta es la propiedad recíproca de F distribuciones, y puede expresarse también exactamente como

$$F_{(1-\alpha), \delta_1, \delta_2} = \frac{1}{F_{\alpha, \delta_2, \delta_1}} \quad (38)$$

donde α y $1 - \alpha$ designan puntos porcentuales de cola inferior de una distribución F, análogos a z_α o $z_{1-\alpha}$ para distribución normal.

Puntos porcentuales para la cola derecha de varias distribuciones F en niveles porcentuales 10, 5 y 1 se dan en el apéndice 3. Cuando X es, por ejemplo, $F_{10,7}$, entonces

$$P(X > 3.64) = P(F_{10,7} > 3.64) = 0.05$$

Decir que $P(F_{10,7} > 3.64) = 0.05$ equivale a decir que

$$\begin{aligned} P(0 \leq X \leq 3.64) &= P(0 \leq F_{10,7} \leq 3.64) \\ &= 1 - 0.05 = 0.95, \end{aligned}$$

y $F_{0.95;10,7} = 3.64$.

En la figura 16 se ilustra esta relación.

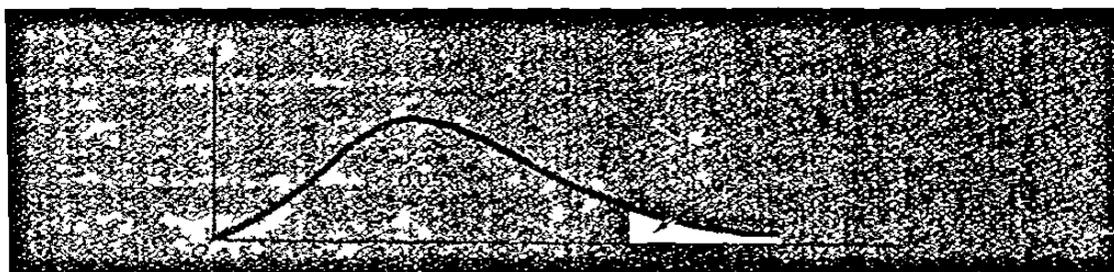


Figura 16. Probabilidad de cola superior de $F_{10,7}$

Análogamente,

$$P(F_{10,7} > 6.62) = 0.01$$

$$P(0 \leq F_{10,7} \leq 6.62) = 0.99$$

Se puede usar también el apéndice 4 para hallar los puntos porcentuales de cola izquierda de distribuciones F siguiendo la propiedad recíproca definida por (38). Por ejemplo, X es $F_{15,10}$; se puede hacer que Y sea $F_{10,15}$. De el apéndice 4 se tiene que

$$P(Y > 2.54) = P(F_{10,15} > 2.54) = 0.05$$

Entonces, por la propiedad recíproca,

$$P\left(\frac{1}{Y} \leq \frac{1}{2.54}\right) = 0.05$$

Como $1/Y$ tiene la misma distribución que X, se tiene entonces:

$$\begin{aligned} P\left(X \leq \frac{1}{2.54}\right) &= P(F_{15,10} \leq 0.394) \\ &= 0.05 \end{aligned}$$

Se puede obtener el mismo resultado usando (38).

$$\begin{aligned} F_{0.05, 15, 10} &= \frac{1}{F_{0.95, 10, 15}} \\ &= \frac{1}{2.54} = 0.394 \end{aligned}$$

Que dice que $P(F_{15,10} \leq 0.394) = 0.05$. Ver figura 17.

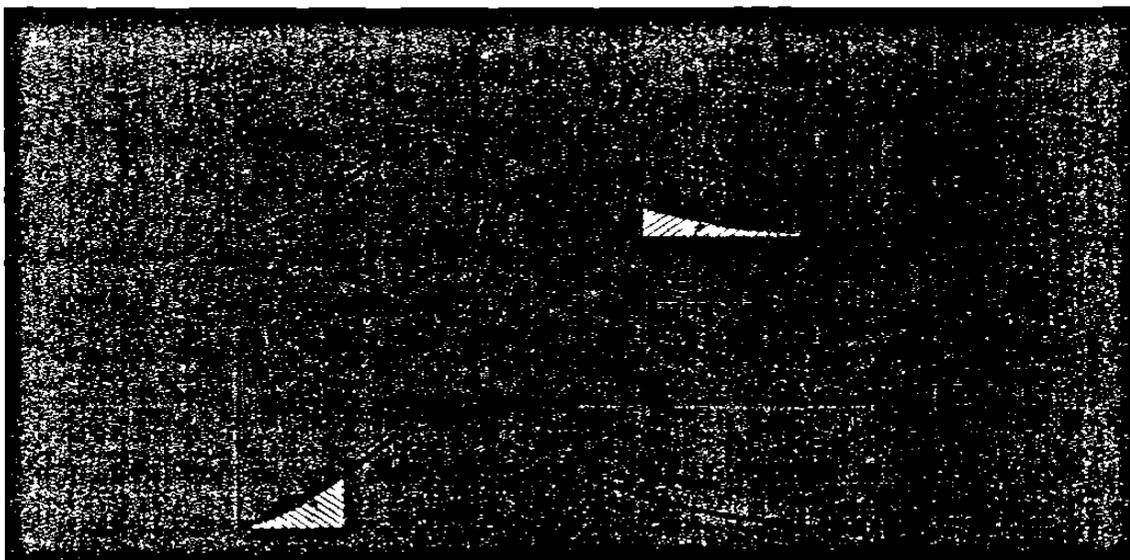


Figura 17. Relación recíproca entre $F_{15,10}$ y $F_{10,15}$

6.5. Distribuciones t Student

Si $X_0, X_1, X_2, \dots, X_\delta$ son $\delta + 1$ variables normales estándares independientes, la estadística

$$t_\delta = \frac{X_0}{\sqrt{\frac{1}{\delta} (X_1^2 + X_2^2 + \dots + X_\delta^2)}} \quad (39)$$

se dice que tiene una distribución t de Student, o simplemente t, con δ grados de libertad. La variable t es una razón de la variable normal estándar a la raíz cuadrada de una variable chi cuadrado dividido por su número de grados de libertad. Es decir (39) es equivalente a $t_\delta = \frac{z}{\sqrt{X^2/\delta}}$. Nuevamente, el numerador y el denominador de (39) son independientes.

Hay una distribución t correspondiente a cada entero positivo. La función densidad para t_δ se da como

147447

$$f(t) = \left(\frac{1}{\delta n} \right) \left[\frac{\Gamma\left(\frac{\delta+1}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)} \right] \left(1 + \frac{t^2}{\delta} \right)^{-(\delta+1)/2} \quad -\infty \leq t \leq \infty \quad (40)$$

Dos curvas de densidad de densidad para distribuciones t, una con $\delta = 2$ y otra con $\delta = 20$, son representadas gráficamente contra la distribución normal estándar en la figura 18 .

Propiedades importantes de las distribuciones t.

1. Como una variable normal, una variable t varía de valor de $-\infty$ a ∞ .

2. Una distribución t es simétrica con

$$E(t) = 0, \quad \text{para } \delta > 1 \quad (41)$$

Y

$$V(t) = \frac{\delta}{\delta - 2} \quad \text{para } \delta > 2 \quad (42)$$

Así una distribución t no posee media cuando $\delta = 1$ y no posee variancia cuando $\delta \leq 2$.

3. Una distribución t es similar a la distribución normal estándar porque ambas varían en valor de $-\infty$ a ∞ , ambas son simétricas y ambas tienen media cero; sin embargo, una distribución t tiene una mayor dispersión que la distribución normal estándar. Esta propiedad puede verse fácilmente de la desviación estándar de t_δ , que es $\sqrt{\delta / (\delta - 2)}$. Esta cantidad es siempre mayor que 1, pero se acerca cada vez más a 1 a medida que aumenta δ . En la práctica, se puede tratar como $n(0, 1)$ cuando $d > 30$.

El apéndice 5, da valores críticos en los extremos de ambas colas de distribuciones t para algunos puntos porcentuales escogidos. Por ejemplo, si X es t_2 , entonces

$$P(-4.303 \leq X \leq 4.303) = P(-4.303 \leq t_2 \leq 4.303) \\ = 0.95$$

De esto se tiene que

$$P(t_2 < -4.303) = 0.025$$

Y

$$P(t_2 > 4.303) = 0.025$$

Ver figura 18 para $\delta = 2$

Análogamente, si X es t_{20} se tiene

$$P(-2.845 \leq t_{20} \leq 2.845) = 0.99,$$

$$P(t_{20} < -2.845) = 0.005,$$

$$P(t_{20} > 2.845) = 0.005,$$

$$P(t_{20} > 2.528) = 0.01,$$

$$P(t_{20} < -1.725) = 0.05,$$

Y así sucesivamente. En realidad, los encabezados columnares de el apéndice 5 son niveles de significación para pruebas bilaterales. Las anotaciones en el cuerpo son los valores críticos de t_δ para tales pruebas. Si se desea tener una prueba de una sola cola con, por ejemplo, $\alpha = 0.05$, el valor crítico se halla en la columna con $2\alpha = 0.10$ para el número apropiado de gl. Por ejemplo, para $P(t_{20} < -1.725) = 0.05$, el número -1.725 se halla en la columna encabezada por 0.10 correspondiente a $\delta = 20$.

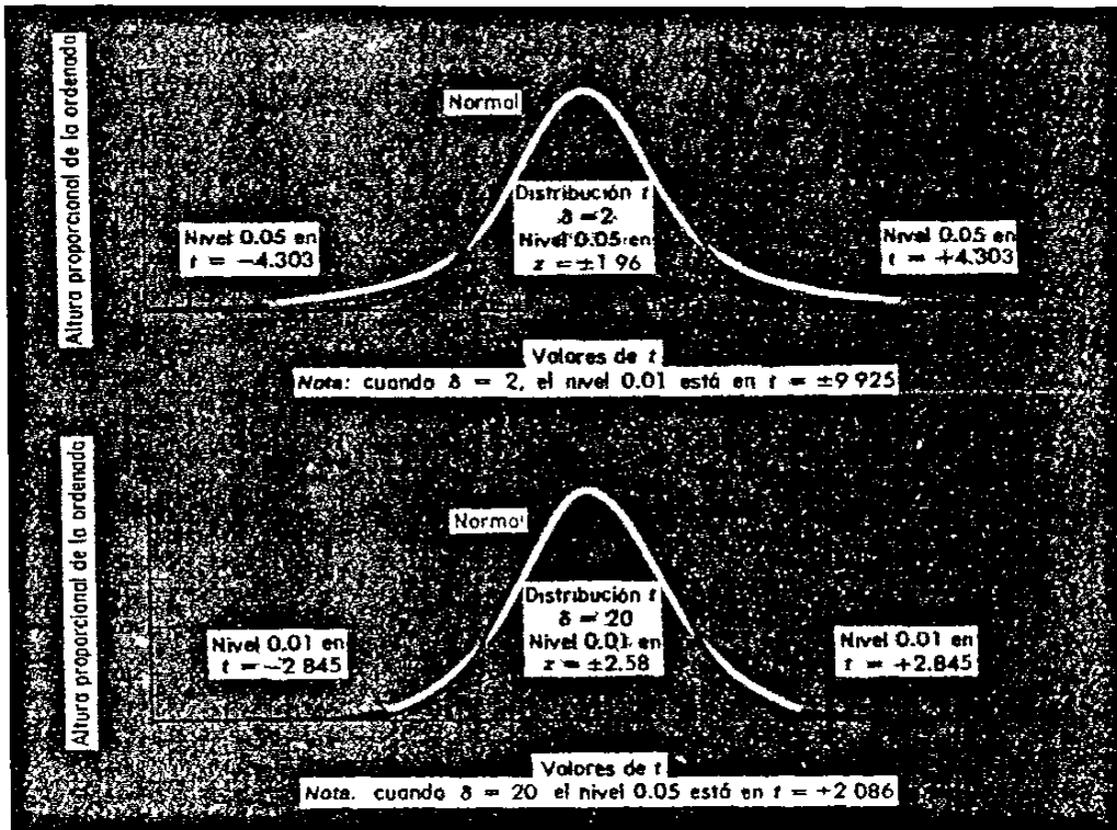


Figura 18 Comparación de distribuciones t como la distribución normal estándar para $d=2$ y $d=20$

CAPÍTULO 7

TEORÍA CLÁSICA DE LA PRUEBA

7.1. Hipótesis estadísticas

Una hipótesis estadística, en términos formales, es un supuesto acerca de la distribución de una variable aleatoria. Se puede especificar una hipótesis dando el tipo de distribución y el valor o los valores del parámetro o los parámetros que la definen. Ejemplos de este tipo de hipótesis serían:

1. X está normalmente distribuida con $\mu = 100$ y $\sigma = 10$.
2. Y es una variable binomial con $\pi = 0.25$.

Pero en la práctica, la distribución de la población es a menudo implícitamente supuesta, especificándose una hipótesis con el valor o los valores del parámetro. Ejemplos de este tipo de hipótesis son

3. La tasa salarial media de los obreros de la industria del automóvil es \$185 por semana; es decir, $\mu = \$185$ por semana.
4. La fracción de unidades defectuosas de producción por cierto proceso es igual o menor de 5 por 100; es decir, $\pi \leq 0.05$.

Una hipótesis estadística puede considerarse como un conjunto de hipótesis "elementales". A este respecto, una hipótesis estadística puede ser simple o compuesta. Una hipótesis simple es una especificación completa de una distribución de probabilidades. La hipótesis (1) anterior es tal hipótesis, porque con la información que contiene se puede escribir la función de densidad específica de X . Una hipótesis compuesta es cualquier hipótesis estadística que no es una hipótesis simple. La hipótesis (2) anterior es una ilustración de esto, porque la distribución de Y no es completamente definida sin la especificación de n .

Puede definirse también una hipótesis simple como aquella que solo contiene un estado, o elemento, del conjunto de parámetros. En contraste, una hipótesis compuesta es aquella que contiene dos o más estados, o elementos, del conjunto de parámetros. Así, la hipótesis (3) es simple, mientras que la hipótesis (4) es compuesta.

En general, se considera que una hipótesis compuesta está constituida por el conjunto de todas las hipótesis simples compatibles con ella. Por tanto, la hipótesis $\pi \leq 0.05$, por ejemplo, ha de ser interpretada como "compuesta" de todas las hipótesis simples de la forma $\pi = \pi_0$, donde π_0 es cualquier número de 0 a 0.05. Cuando se dice que una hipótesis compuesta es verdadera se quiere decir que alguna de las hipótesis simples que la constituyen es verdadera.

Una forma cómoda de especificar lo que se requiere de un procedimiento de prueba es concentrar la atención en dos conjuntos posibles de valores del parámetro, o dos hipótesis estadísticas, a tal par de conjuntos se le llama, respectivamente, hipótesis nula, designada por H_0 , e hipótesis alternativa, designada por H_1 . Aquí H_0 es el conjunto de hipótesis simples que, si cualquiera de ellas es cierta, haría a_0 (aceptando H_0 y actuando de acuerdo con ello) un acto mejor, H_1 es el conjunto de hipótesis simples que, si cualquiera de ellas es cierta, haría a_1 (rechazando H_0 o aceptando H_1 y actuando de acuerdo con ello)

un acto mejor. Así, la elección entre a_0 y a_1 se asocia con la aceptación o rechazo de H_0 .

Las designaciones entre las hipótesis nula y alternativa son arbitrarias. Pero, típicamente, la hipótesis nula es especificada en una forma exacta, tal como "no hay diferencia en ingreso medio entre médicos y profesores", "el proceso de producción está bajo control" o "no ha habido aumento real del ingreso per capita en la India desde que llegó a la independencia". En otras palabras, una hipótesis nula "nulifica" el efecto de un "tratamiento" y corresponde a la ausencia de efectos de la variable que se investiga. La hipótesis alternativa generalmente es formulada con menos precisión: a menudo se especifica como una variación de valores que prevalecería si la variable que se estudia ejerció algún efecto. Así, la hipótesis nula se especifica con frecuencia en una forma opuesta a la que se supone cierta, y la alternativa es expresada como la opuesta a la hipótesis nula.

En la terminología de prueba, hablamos de probar la hipótesis nula contra una alternativa en el supuesto tentativo de que la hipótesis nula es cierta. Pero se debe comprender que realmente se está tomando una decisión entre dos acciones, o entre H_0 y H_1 . Hay tres tipos principales de pruebas, cada uno de los cuales es identificado por la forma en que se formulan H_0 y H_1 . Primero, hay una prueba de doble cola dos lados, para la cual las hipótesis son de la siguiente forma:

$$H_0: \theta = \theta_0; \quad H_1: \theta \neq \theta_0$$

Este tipo de prueba es apropiado si se tiene presente que el valor de un parámetro puede ser demasiado pequeño o demasiado grande para algún fin específico. Se tiene, por ejemplo, que es producida una pieza de máquina y que ha de ser acoplada a otras piezas para montar una máquina completa. Una de las dimensiones críticas, por ejemplo la longitud, de esta pieza es especificada

como de una medida de 2.5 pulgadas. Si la pieza es demasiado corta o demasiado larga (de acuerdo con límites de tolerancia especificados), puede que no encaje y sea inútil. Así el productor debe probar el siguiente par de hipótesis de cuando en cuando con datos de muestra para ver si el proceso de fabricación está bajo control o no:

$$\begin{array}{ll} H_0: \mu = 2.5, & a_0: \text{el proceso está bajo control;} \\ H_1: \mu \neq 2.5, & a_1: \text{el proceso está fuera de control} \\ & \text{y debe emprenderse una acción} \\ & \text{correctiva.} \end{array}$$

Segundo, hay la prueba de cola inferior, o cola izquierda, para la cual pueden formularse hipótesis en una de las dos formas:

1. $H_0: \theta \geq \theta_0, H_1: \theta < \theta_0$
2. $H_0: \theta = \theta_0, H_1: \theta = \theta_0 - a$, para $a > 0$.

Este tipo de prueba se emplea cuando el valor de un parámetro no es bastante pequeño para algún objeto específico. Se tiene, por ejemplo, que ha quedado bien establecido por los científicos médicos que si el contenido medio de nicotina de los cigarrillos es de 25 miligramos o más, es probable que el fumador contraiga cáncer. Pero si el contenido de nicotina medio es menor de 25 miligramos, el usuario se encuentra relativamente seguro. Se supone que se tiene un fumador habitual y está dispuesto a arriesgar su vida; ciertamente, debe tratar de encontrar una marca de cigarrillos cuyo contenido medio de nicotina sea menor de 25 miligramos. En este caso, debe formular sus hipótesis como sigue:

$$\begin{array}{ll} H_0: \mu \geq 25, & a_0: \text{no compre la marca;} \\ H_1: \mu < 25, & a_1: \text{compre la marca.} \end{array}$$

El tercer tipo de prueba se llama prueba de cola superior, o cola derecha. Aquí, las hipótesis suelen expresarse en dos formas generales como sigue:

1. $H_0: \theta \leq \theta_0, H_1: \theta > \theta_0$
2. $H_0: \theta = \theta_0, H_1: \theta = \theta_0 + a$, para $a > 0$.

Se usa este tipo de pruebas cuando se tiene duda de que el valor de un parámetro sea bastante grande para algunas metas predeterminadas. Para construir un puente, por ejemplo, una firma de construcción necesita cables de acero que tengan una resistencia media a la rotura de más de 10 000 libras. Para decidir si debe adquirirse una marca determinada de cables de acero, serían apropiadas las siguientes hipótesis:

$$\begin{array}{ll} H_0: \mu \leq 10,000, & a_0: \text{no compre la marca;} \\ H_1: \mu > 10,000, & a_1: \text{compre la marca.} \end{array}$$

Deben observarse varias precauciones acerca de las designaciones de H_0 y H_1 . Primero, las pruebas del tipo segundo y tercero, por razones obvias, se llaman pruebas de una sola cola o de un lado. Segundo, aunque H_0 y H_1 son siempre mutuamente exclusivas, no es necesario que aparezcan colectivamente exhaustivas. Es decir, $\{\theta \in H_0\}$ y $\{\theta \in H_1\}$ son conjuntos desunidos, pero $\{\theta \in H_0\} \cup \{\theta \in H_1\}$ no agotan necesariamente todos los valores posibles del estado (o parámetro). (Aunque puede suponerse que H_0 y H_1 son en realidad colectivamente exhaustivos aun si no lo parecen.) Tercero, la práctica clásica de formular hipótesis estadísticas simples, tal como $\mu = 50$ o $\pi = 0.25$, etc. Para variables aleatorias continuas puede ser de dudosa corrección.

7.2. Errores y riesgos de la prueba

La práctica de probar la hipótesis nula contra una alternativa sobre la base de información obtenida de muestra conducirá a dos tipos posibles de errores, debido a fluctuaciones al azar en el muestreo. Por una parte, la hipótesis nula es en realidad cierta, pero como los datos de la muestra son incompatibles con ella, es rechazada. El error de rechazar una H_0 cierta se conoce como error de la primera clase, o error de tipo I. Evidentemente, cuando cometemos uno u otro tipo de errores somos inducidos a escoger el acto inapropiado. (Ver cuadro 2).

Las probabilidades de cometer errores del tipo I y del tipo II pueden considerarse como los riesgos de decisiones incorrectas. Precisamente, la probabilidad máxima de cometer un error del tipo I se llama nivel de significación y se representa por α . Y

$$\alpha = \max P (I) = \max P (a_1 | H_0) = \max P (H_1 | H_0).$$

La probabilidad máxima de cometer un error del tipo II no tiene nombre especial, pero se designa universalmente por β . Y

$$\beta = \max P (II) = \max P (a_1 | H_1) = \max P (H_0 | H_1).$$

Muchas veces, α y β no indican probabilidades máximas, sino meras probabilidades. El contexto indicará qué significado pretende darse a α y β .

Se observa que los riesgos α y β son probabilidades condicionales y que están estrechamente relacionados. En particular, cuando el tamaño de la

Acción emprendida sobre la base de los datos de la muestra	Estado de naturaleza	
	$\theta \in H_0$	$\theta \in H_1$
A_0 : acepte H_0	Acto correcto	Error de tipo II
A_1 : rechace H_0	Error de tipo I	Acto correcto

Tabla 1 Errores y decisiones de tipo I y tipo II

muestra es fijo, α y β varían inversamente al variar el "valor crítico"; cuando se aleja de un peligro, aumenta el otro peligro. Debido a esta relación, un estadístico clásico a menudo puede dirigir una prueba especificando α sola sin la evaluación explícita de β . Pero, idealmente, deben especificarse los riesgos de cometer ambos tipos de errores que se desea asumir en una situación dada de prueba. Se hace esto permitiendo variar el tamaño de la muestra y escogiendo el tamaño de la muestra que puede satisfacer ambas especificaciones. A continuación, se introducirá primero el procedimiento de prueba de especificar α sola en el supuesto de un tamaño de muestra fijo, luego la evaluación explícita de β , finalmente la determinación de un tamaño de muestra dados valores específicos de α y β .

Cuando solo ha de especificarse riesgo α , generalmente se fija en cierto valor convencional tal como 0.05 ó 0.01. Estos dos números son usados tan frecuentemente que cuando H_0 es rechazada en $\alpha=0.05$, se puede decir que el resultado es "significativo", y cuando H_0 es rechazada en $\alpha=0.01$, y que el resultado es "altamente significativo". Estas dos descripciones son usadas sin definiciones explícitas, pero el término "significativo" tiene un significado preciso al emplearlo aquí. La noción de "nivel de significación" indica meramente que una prueba estadística evalúa el "significado" del resultado de la muestra comparado con el supuesto contenido en la hipótesis nula. Además, cuando una hipótesis nula es rechazada en, por ejemplo $\alpha = 0.05$, significa que podría

esperarse que los resultados de la muestra causaran rechazo de las verdaderas hipótesis nulas 5 por 100 de las veces a largo plazo. Similares interpretaciones pueden darse a otros valores de α .

En cuanto a si el riesgo α debe ser 0.05 ó 0.01, o algún otro valor, los estadísticos clásicos no dan respuestas definitivas porque sostienen que esto es principalmente una decisión de la gerencia y no la prerrogativa del estadístico. Sin embargo, cuando se le pide consejo a un estadístico, se han seguido dos reglas aproximadas.

1. Si se tiene un alto grado de creencia en la corrección de H_0 , debe asignarse una pequeña probabilidad de rechazarla cuando es realmente cierta; es decir, debe fijarse en 0.01 o un valor más pequeño. Si se cree muy firmemente que H_0 es cierta, debe ser aceptada, a menos que haya fuerte evidencia de lo contrario. Por supuesto, esto indica que si se tiene un alto grado de creencia en la corrección de H_1 , debe fijarse α en 0.05 o aún en valores mucho más altos. Pero este procedimiento está sujeto a severa crítica por los partidarios de la teoría clásica de probabilidades u objetivistas. Afirman que el investigador puede cargar la interpretación de los resultados de su muestra de tal forma que una hipótesis nula virtualmente nunca puede ser rechazada, cualquiera que sea la evidencia de la muestra, simplemente fijando α bastante cerca de cero, cuando el investigador desee aceptar su hipótesis nula. Y, sin duda, el investigador puede hacerse exactamente eso. Los objetivistas dicen con gran vigor que la prueba de la muestra debe hablar por sí misma y no debe ser sometida a ser enterada o desplazada por conceptos preconcebidos sobre la población de la que fue extraída la muestra. Se considera que hay mérito en esta posición y, en lo sucesivo, solo se usará el procedimiento mencionado a continuación.

2. Cuando un rechazo erróneo de H_0 sería costoso o grave, el riesgo α debe ser pequeño; cuando una aceptación errónea de H_0 sería costosa o grave, el riesgo β debe ser pequeño (o debe fijarse α en un nivel alto tal como 0.25 y aún 0.75).

Por desgracia los criterios expuestos arriba son muy elusivos en el sentido de que, dentro del contexto de decisión clásico, no se ha establecido ninguna norma objetiva para medir las consecuencias de decisiones erróneas.

7.3. Estadísticas de prueba y reglas sobre decisiones

Puesto que la elección entre H_0 y H_1 de hacerse basándose en pruebas de muestra, es necesario escoger una función de las n observaciones de muestra como estadística de prueba. En general, la estadística de prueba debe ser una cuya distribución por muestreo sea conocida en el supuesto de que la hipótesis nula es cierta. La estadística de prueba generalmente resulta el estimador convencional del parámetro previsto en H_0 . Por ejemplo, dado que H_0 es de la forma, por ejemplo $\mu = 100$, y que la desviación estándar de la población es, por ejemplo 10, la media de la muestra debe ser escogida como la estadística de prueba. Esta elección es apropiada porque si $\mu = 100$ es cierta, la distribución de \bar{x} será normal (porque la población es normal o porque el teorema de límite central es operativo), con $E(\bar{x}) = \mu = 100$ y $\sigma_{\bar{x}} = 10 / \sqrt{n}$.

El hecho de que la distribución por muestreo de la estadística de prueba es definida completamente en el supuesto de que H_0 es verdadera proporciona un criterio para evaluar la validez de H_0 dado el resultado de una muestra particular. Por ejemplo, si la media de la muestra difiere en una cantidad grande de la media de la población especificada en H_0 , y si H_0 todavía es considerada como cierta, se deduce que la media de la muestra debe ser uno de los valores

extremos de la distribución por muestreo cuya expectativa es μ_0 . Pero una diferencia tan grande no puede explicarse razonablemente por variaciones casuales en el muestreo al azar solamente; es muy probable que la muestra fue extraída de una población diferente. La evidencia de la muestra permite dudar mucho de la corrección de H_0 y, por consiguiente, rechazarla. Pero si la media de la muestra difiere de μ_0 en una cantidad bastante pequeña, se dice que la diferencia puede ser fácilmente el resultado de variaciones casuales, la media de la muestra no es incompatible con H_0 , y H_0 no puede ser rechazada con tal evidencia de la muestra.

Para generalizar: Una estadística de prueba $\hat{\theta}$ es una variable aleatoria cuya distribución es conocida, en el supuesto (tentativo) de que la hipótesis nula $\theta = \theta_0$ es cierta. Reglas de decisión sobre la aceptación o rechazo de H_0 pueden ser establecidas entonces respecto a la amplitud de $\hat{\theta}$ y un resultado particular de la muestra. Se hace esto clasificando la amplitud de $\hat{\theta}$ en dos subconjuntos desunidos, R , la región de rechazo, o la región crítica, que contiene los resultados menos favorables a H_0 , y A , la región de aceptación, que contiene los resultados más favorables a H_0 , tales que si $\hat{\theta} \in R$, rechazamos H_0 , y si $\hat{\theta} \in A$, aceptamos H_0 . El valor de $\hat{\theta}$ que separa R de A se llama valor crítico de la estadística de prueba, representado por $\hat{\theta}_c$. (A veces, hay más de un valor crítico.). El valor de $\hat{\theta}_c$ en cualquier prueba con tamaño fijo muestra y riesgo α depende de la forma de la hipótesis alternativa, del tamaño del riesgo α o del nivel de significación, y de la distribución de la estadística de prueba. En toda esta parte se supondrá que la distribución de $\hat{\theta}$ es normal y que $\hat{\theta}$ es insesgada. Ahora se demostrará cómo ha de ser hallado el valor crítico de la estadística de prueba para cada uno de los tres tipos de pruebas mencionadas anteriormente.

Primero, se considera la prueba de la cola superior. En este caso, se probará $H_0: \theta < \theta_0$ contra $H_1: \theta > \theta_0$. Naturalmente, la validez de H_0 ha de ser apoyada por pequeños valores de $\hat{\theta}$ solamente; es decir, la región de rechazo

se encuentra en la cola derecha de la distribución de θ , como se ilustra por la figura 19. Así, dado el nivel de significación, se determina el valor de $\hat{\theta}_c$ de modo que $P(\hat{\theta} > \hat{\theta}_c | \theta_0) = \alpha$. Esta condición es satisfecha con

$$\hat{\theta}_c = \theta_0 + z_{1-\alpha} \sigma_{\hat{\theta}} \quad (43)$$

donde z es una variable normal estándar y $z_{1-\alpha}$ es el valor de la variable normal estándar que separa R (región de rechazo) de A (región de aceptación). Por ejemplo, si $\alpha = 0.05$, entonces $z_{1-\alpha} = 1.65$; si $\alpha = 0.01$, entonces $z_{1-\alpha} = 2.33$. El término $z_{1-\alpha}$ es el valor de z_0 en la ecuación $P(Z \leq z_0) = 1 - \alpha$. Así, cuando se ocupa de una prueba de cola derecha, la regla de decisión es: rechazar H_0 si, y solo si $\hat{\theta}_c = \theta_c$; en otro caso, aceptar H_0 . Puesto que, en (43), z es una variable normal estándar, esta regla de decisión también puede enunciarse en una forma más conveniente "estandarizando" la estadística de la prueba, rechace H_0 si, y solo si,

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_{1-\alpha} \quad (44)$$

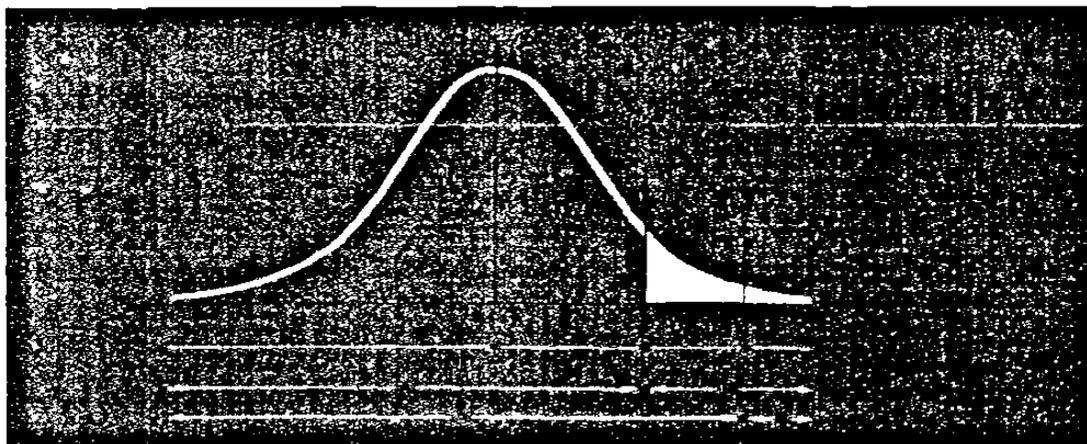


Figura 19 Regiones de rechazo para probar $\theta \leq \theta_0$ contra $\theta > \theta_0$ en $\alpha=0.05, 0.01$

Luego , en el caso de una prueba de cola izquierda, se probará $H_0: \theta \geq \theta_0$ contra $H_1: \theta < \theta_0$. Ahora solo valores grandes de θ apoyarían la validez de H_0 , y la región de rechazo se encontraría en la cola izquierda de la distribución de θ^{\wedge} , como se representa en la figura 20. Por tanto, el valor crítico debe ser escogido de modo que $P(\theta^{\wedge} < \theta^{\wedge}_c | \theta_0) = \alpha$. Esto queda satisfecho si

$$\theta^{\wedge}_c = \theta_0 + z_{\alpha} \sigma_{\theta^{\wedge}} \quad (45)$$

donde z_{α} es el valor crítico -el valor que separa R de A. Nuevamente, si $\alpha = 0.05$, entonces $z_{\alpha} = -1.65$; si $\alpha = 0.01$, entonces $z_{\alpha} = -2.33$. Por tanto, para una prueba de cola izquierda la regla de decisión es: rechace H_0 si, y solo si, $\theta < \theta_c$; en otro caso, acepte H_0 . Alternativamente , este criterio puede enunciarse en términos de la variable normal estándar sola como: rechace H_0 si, y solo si,

$$\frac{\theta^{\wedge} - \theta_0}{\sigma_{\theta^{\wedge}}} > z_{\alpha} \quad (46)$$

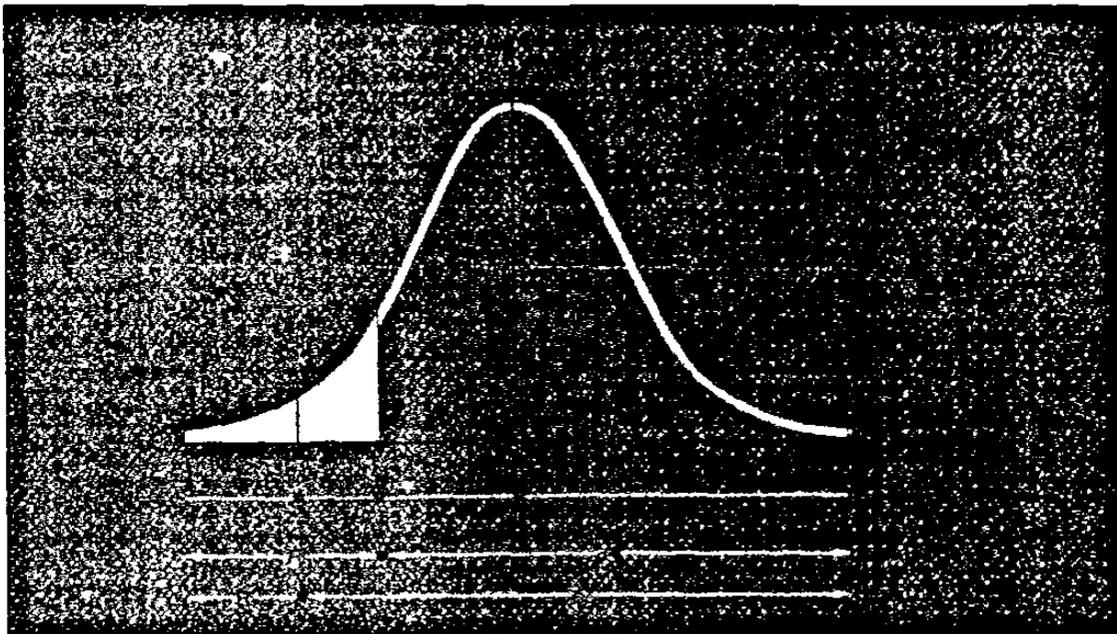


Figura 20 Regiones de rechazo para probar $\theta \geq \theta_0$ contra $\theta < \theta_0$ en $\alpha=0.05, 0.01$

Finalmente, para una prueba de dos lados, la evaluación será entre $H_0: \theta = \theta_0$ y $H_1: \theta \neq \theta_0$. En este caso, valores grandes y pequeños de la estadística de prueba tienden a apoyar la validez de H_1 . La convención es dividir la región crítica en dos partes iguales que son las dos colas de la función densidad de $\hat{\theta}$ representada en la figura 21. Ahora dos valores críticos, $\hat{\theta}_{c1}$ y $\hat{\theta}_{c2}$, en vez de uno solo, por determinar. Es fácil ver de la exposición anterior que, dado un riesgo α ,

$$\hat{\theta}_{c1} = \theta_0 + z_{\alpha/2} \sigma_{\hat{\theta}} \quad (47)$$

y

$$\hat{\theta}_{c2} = \theta_0 + z_{1-\alpha/2} \sigma_{\hat{\theta}} \quad (48)$$

para estas expresiones, si $\alpha = 0.05 = z_{\alpha/2} = -1.96$ y $z_{1-\alpha/2} = +1.96$; si $\alpha = 0.01$, $z_{\alpha/2} = -2.58$ y $z_{1-\alpha/2} = +2.58$. con estas consideraciones, se formula la tercera regla de decisión como: rechace H_0 si, y solo si, $\hat{\theta} < \hat{\theta}_{c1}$ o $\hat{\theta} > \hat{\theta}_{c2}$; en otro caso, aceptar H_0 . Equivalentemente, se puede afirmar: rechazar H_0 si, y solo si,

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_{\alpha/2} \quad (49)$$

o

$$\frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} > z_{1-\alpha/2} \quad (50)$$

Para recapitular lo anterior, puede indicarse que el procedimiento de prueba clásico, para tamaño fijo de muestra, implica los siguientes pasos:

1. Formular hipótesis de acuerdo con el problema que se tiene ante sí.
2. Escoger un nivel de significación, o riesgo α , con referencia al juicio propio sobre los "costos" relativos de las consecuencias de cometer los errores de tipo I y tipo II.

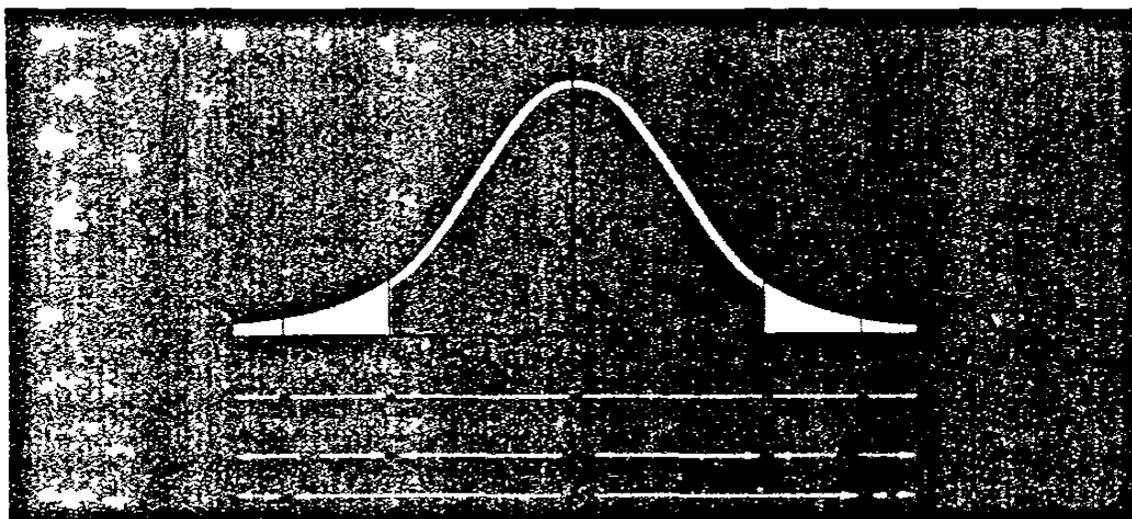


Figura 21 Regiones de rechazo para probar $\theta = \theta_0$ contra $\theta \neq \theta_0$ en $\alpha=0.05, 0.01$
 Se puede observar que solo se han dado valores de z para los diversos tipos de pruebas con $\alpha = 0.05, 0.01$. los valores de z con otros niveles de significación pueden hallarse fácilmente en el apéndice 1.

3. Escoger la estadística de prueba cuya distribución por muestreo es conocida en el supuesto de que H_0 sea cierta; es decir, la estadística de prueba tiene la función de probabilidad condicional de la forma $f(\hat{\theta} | \theta_c)$.
4. Establecer la regla de decisión que depende de la forma de la hipótesis alternativa y el nivel de significación. Esto supone la determinación del valor o los valores críticos de z .
5. Calcular los valores de la estadística de prueba y el error estándar de la estadística de prueba partiendo de una muestra al azar, de modo que la estadística de prueba estandarizada pueda ser comparada con el valor o los valores críticos de z .
6. Decidir rechazar H_0 (o tomar a_1) si $\hat{\theta}$ o z es un miembro de R , o aceptar H_0 (o tomar a_0) si $\hat{\theta}$ o z es un miembro de A .

En relación con la conclusión o decisión inferencial final, debe tenerse presente que las pruebas estadísticas están sujetas a errores y, por consiguiente, no pueden probar nada acerca de poblaciones. Si la regla que define la prueba dirige a rechazar H_0 , en un caso dado, no significa necesariamente que H_0 sea falsa.

Ni tampoco ha de creerse que H_0 sea falsa.

Es requerido a actuar como si fuera falsa, asumiendo con ello el riesgo de una decisión errónea.

CAPÍTULO 8

ESTADÍSTICA NO PARAMÉTRICA

8.1. Pruebas de aleatoriedad

Dos preguntas frecuentemente encontradas en aplicaciones estadísticas son:

- 1) ¿Es la muestra una muestra al azar?,
- 2) 2) ¿Es este proceso un proceso al azar?.

La primera pregunta puede ser contestada solamente examinando el método de elección - no los resultados de la muestra mismos -. Si se asignan números de serie a miembros del conjunto y se usan dígitos al azar para números de serie escogidos, la muestra es una muestra al azar, con tal, por supuesto, de que no se cometan errores durante este procedimiento. La segunda pregunta surge cuando se desea conocer si una muestra que se sabe no fue escogida al azar puede ser tratada como si fuera una muestra al azar. Un ejemplo de tales datos sería el número de piezas defectuosas producidas cada día por una máquina recién instalada. Suponiendo que el número total de piezas producidas cada día es una constante, podemos trabajar entonces con el número de piezas defectuosas o la proporción de piezas defectuosas, para

control de calidad. Hipotéticamente, a juicio del capataz y de los ingenieros supervisores, la máquina ha sido apropiadamente instalada y ajustada, y debe producir en la forma que ha de producir el resto de su vida útil. Se puede entonces reunir datos de unas pocas semanas sobre el número de piezas defectuosas y ver si estos datos se comportan como una serie al azar. La pregunta es si este proceso es al azar. No se puede esperar contestar plenamente esta pregunta, porque no hay límite a los aspectos de aleatoriedad que deben ser probados antes de que los datos hayan sido "completamente" probados en cuanto a aleatoriedad. Pero se pueden probar algunos aspectos de la aleatoriedad. La hipótesis nula es que los datos son extracciones independientes al azar de una población estable.

Este capítulo muestra dos pruebas no paramétricas sencillas de aleatoriedad basadas en el concepto de "series":

- 1) Series por encima y por debajo de la mediana, y
- 2) Series arriba y abajo.

8.1.1. Series por encima y por debajo de la mediana.

Se tiene una muestra de, por ejemplo 26 observaciones reunidas en el siguiente orden:

97,89,25,81,11,83,16,96,44,32,98,19,68,
33,25,54,74,82,17,49,33,22,62,20,92,80

La mediana es $(54+49)/20 = 51.5$. Se busca ahora si cada observación está arriba o abajo del valor de esta mediana. Si está arriba de la mediana, se marca como a, y si está debajo de la mediana se marca como b. Para la muestra anterior, tenemos la siguiente sucesión de a y b:

aa-b-a-b-a-b-a-bb-a-b-a-bb-aaa-bbbb-a-b-aa

Cada sucesión de a ó b, no interrumpida por la otra letra, es una serie, y tenemos entonces 17 series para nuestros datos.

Ahora, el número de series arriba y debajo de la mediana es una variable aleatoria. Por la hipótesis nula, R es distribuida aproximada y normalmente para $n \geq 25$ con:

$$E(R) = \frac{n+2}{2} \quad (51)$$

y

$$V(R) = \frac{n(n-2)}{4(n-1)} \quad (52)$$

La hipótesis alternativa es a menudo de la variedad de cola izquierda. Esto es especialmente apropiado si se considera agrupación de observaciones semejantes, es decir, demasiado pocas series, porque tal situación con frecuencia indica un cambio de la media de la población. Ocasionalmente puede que se desee efectuar una prueba de cola derecha para buscar oscilaciones sistemáticas arriba y debajo de la mediana de una observación o otra. Cuando no tenemos ideas específicas, excepto que H_1 debe ser lo opuesto de H_0 , podemos proceder entonces a hacer una prueba de dos lados. En todo caso, la estadística de prueba es:

$$z = \frac{R - \frac{n+2}{2}}{\sqrt{\frac{n(n-2)}{4(n-1)}}} \quad (53)$$

Se tiene una prueba de dos lados para nuestros datos ilustrativos en $\alpha=0.01$; se tiene:

$$E(R) = \frac{26+2}{2} = 14;$$

$$V(R) = \frac{26(26-2)}{4(26-1)} = 6.24;$$

Puesto que el valor observado de z se encuentra entre -2.58 y $+2.58$, H_0 no puede ser rechazada en el nivel de significación de 1 por 100. La máquina está produciendo piezas defectuosas al azar, independientemente, y a un ritmo estable, por la información que podemos obtener de esta prueba.

Cuando hay valores de muestra ligados al valor de la mediana (iguales a ella) pueden ser despreciados, de modo que el número de series y el tamaño de la muestra se reduzcan, o puede asignárseles a o ó b por un proceso al azar tal como el lanzamiento de una moneda. Nada se dijo acerca de la forma de la distribución de la población.

8.1.2. Series arriba y abajo

Una prueba para series arriba y abajo es similar a la prueba anterior. Dada una muestra de n observaciones registradas en el orden obtenido, colocando un signo (+) o un signo (-) entre cada par de observaciones sucesivas, según si la última observación es mayor o menor que la observación anterior. Ahora, cada serie de + o -, no interrumpida por el otro signo, se cuenta como una serie. Por ejemplo, usando los mismos datos que la muestra tenemos:

- - + - + - + - - + - + - - + + + - + - - + - + - ,

un total de 19 series. Estos signos son determinados entre dos observaciones sucesivas, y hay $(n-1)$ de tales signos para una muestra de tamaño n . Nuevamente pueden ser rotos por un proceso aleatorio.

Por la hipótesis nula, R es distribuida aproximada y normalmente para $n \geq 20$ con

$$E(R) = (1/3)(2n-1) \quad (54)$$

y

$$V(R) = (1/90)(16n-29) \quad (55)$$

La estadística de prueba es

$$(56)$$

Suponiendo que se va a probar contra una alternativa de cola izquierda, con los datos ilustrativos, en $\alpha=0.01$, entonces

$$R = 19$$

$$E(R) = 1/3 [2(26)-1] = 17$$

$$V(R) = 1/90 [16(26)-29] = 4.3$$

Se tiene que $0.96 > -2.33$, y la aleatoriedad es aceptada en $\alpha=0.01$

8.2. La razón Von Neumann: Una prueba de independencia

Las pruebas de independencia son iguales que las pruebas de aleatoriedad. El procedimiento llamado la razón de Von Neumann para probar aleatoriedad (independencia) en datos de series de tiempo.

:

Los datos comerciales y económicos ordenados en secuencias de tiempo se conocen como muestras no al azar. A menudo, una observación sobre precio, producción, ingreso nacional, etc., para un mes o año es estadísticamente dependiente, en cierto grado, del valor de dicha variable en el período de tiempo anterior. Esta relación de dependencia o "desplazamientos" entre términos sucesivos de una serie cronológica se llama correlación serial. A veces, se emplean también los términos autocorrelación y correlación de

desplazamiento para esta relación de dependencia. Cuando se hace una distinción, a menudo se considera la autocorrelación como relación de desplazamiento de la población, correlación serial significa relación de desplazamiento en una muestra, y correlación de desplazamiento significa la relación de desplazamiento entre dos series cronológicas diferentes. Sin embargo, éstos términos se usan mucho indistintamente y esta es la práctica que se adoptará.

Una cuestión clave en las pruebas de una correlación serial es si $P(X_{i+1}=a \mid X_i=b) = P(X_{i+1}=a)$. Si las probabilidades son iguales para todos los valores de i , no hay correlación serial, en lo que se refiere a esta prueba. La cuestión clave es si el conocimiento del valor de la i -ésima observación ayuda a predecir el valor de la $(i+1)$ -ésima observación. Si ayuda, la correlación serial está claramente presente. Si no ayuda, la correlación serial está ausente en lo que se refiere a esta prueba. Por ejemplo, al checar repetidamente en dado perfecto, el conocimiento de que la 64-ésima echada fué cinco no nos ayuda nada a predecir el resultado de la 65-ésima echada. $P(X_{65}=a \mid X_{64}=5) = P(X_{65}=a) = 1/6$ para cada valor de a . La correlación serial está ausente en esta prueba.

La versión general de la cuestión clave no restringe el desplazamiento a solo una observación. Es decir, la versión general pregunta si $P(X_{i+j}=a \mid X_i=b) = P(X_{i+j}=a)$ para todas las i y j , $j \geq 1$. Si es así, la correlación serial está ausente. Si, para cualesquiera valores de i y j , las probabilidades son desiguales, la correlación serial está presente.

Si la correlación serial está presente, puede ser "positiva" o "negativa". Hablando intuitivamente, correlación serial positiva significa que una observación alta es probable que vaya seguida de otra observación alta y una observación baja es probable que vaya seguida de otra observación baja. Correlación serial negativa significa que una observación extraordinariamente alta es probable que vaya seguida de una observación extraordinariamente

baja, y una observación extraordinariamente baja es probable que vaya seguida de una observación extraordinariamente baja. Podría considerarse que tales conjuntos de datos son raros, pero en cualquier caso existen. Puede haber una explicación obvia para tales datos, que a menudo supone un proceso correctivo que no es ajustado apropiadamente.

La prueba de independencia por la razón Von Neuman se basa en la diferencia entre medias sucesivas elevadas al cuadrado definida como sigue:

$$K = \frac{\delta^2}{s_y^2} \quad (57)$$

Donde

$$\delta^2 = \frac{1}{n-1} \sum_i (y_{i+1} - y_i)^2$$

y

$$s_y^2 = \frac{1}{n} \sum_i (y_i - \bar{y})^2$$

Como queda implícito por la definición de δ^2 , la razón K está estrechamente relacionada con la variancia de las diferencias entre observaciones sucesivas. Cuando estas diferencias son pequeñas, resultará una K pequeña y se indica una correlación serial positiva. Cuando estas diferencias son grandes, resultará una K grande y se revela una correlación serial negativa. Así, valores muy grandes y muy pequeños de K nos conducirían a rechazar la aleatoriedad o la independencia.

Para efectuar la prueba de independencia por la razón de Von Neuman, puede usarse la tabla presentada en el apéndice 6. Esta tabla da los puntos 5 por 100 y 1 por 100 de la distribución de K. Da dos valores críticos, k y k', para cada nivel de significación y cada tamaño de muestra.

8.3. La prueba chi cuadrado sobre frecuencias

Chi cuadrado es una estadística asombrosamente útil. Además de su uso para inferencia con variancias, también puede ser usada para probar cualquier hipótesis relativa a datos categóricos, de los cuales la forma binomial solo es un caso especial. Aquí se muestran datos categóricos constituidos por frecuencias absolutas que ocurren en las diversas clases o compartimentos de una tabla de clasificaciones simples o múltiples. Debido a las diferencias en clasificación y en razones para probar, las pruebas chi cuadrado pueden ser clasificadas como: pruebas de bondad de ajuste, la independencia de tablas de clasificación cruzada (o contingencia), y homogeneidad.

Generalmente, con una prueba chi cuadrado se formula primero la hipótesis nula con que son determinadas las frecuencias esperadas, o teóricas. Después, con los datos de la muestra se establecen las frecuencias observadas. A continuación, se comparan estos dos conjuntos de frecuencias tomando diferencias entre ellos. Finalmente, sobre la base de estas diferencias, se especifica un criterio de decisión para juzgar si las frecuencias observadas, en promedio, divergen significativamente de las diferencias esperadas. Se busca determinar, como un criterio apropiado, si estas diferencias se deben a variaciones casuales en el muestreo al azar. Si es así, la hipótesis nula es aceptada; en otro caso, es rechazada.

Para este grupo de problemas puede demostrarse que la suma de la razón de las diferencias al cuadrado entre frecuencias observadas, representada por o_i , y las frecuencias esperadas, representado por e_i , a las frecuencias esperadas es distribuida aproximadamente como una variable chi cuadrado. Es decir,

$$X_{\delta}^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad (58)$$

es distribuida aproximadamente como chi cuadrado, con grados de libertad iguales al número de clases o categorías comparadas, menos el número de restricciones impuestas a la comparación. Más precisamente, por la ecuación anterior, $\delta=k-1$ si se puede calcular frecuencias esperadas sin tener que estimar parámetros de población o estadística de muestra. Aquí, k es el número de clases; así, si conociendo $K-1$ de las diferencias esperadas, la frecuencia restante es determinada singularmente porque hay que satisfacer la condición de que, $\sum o_i = \sum e_i = n$ donde n es el tamaño de la muestra o frecuencia total. Sin embargo, tenemos $\delta=k-1-m$ si se pueden calcular las frecuencias esperadas solo después de estimar m parámetros de población por estadísticas de muestra.

Pueden hacerse aquí tres observaciones en relación con el uso de la fórmula (58) para varios tipos de pruebas chi cuadrado. Primero, cuando $\delta=1$, especialmente cuando la frecuencia total es muy pequeña, por ejemplo, menor de 50, es aconsejable introducir un factor de corrección de continuidad de $1/2$ al calcular el valor de chi cuadrado. Es decir, cuando $\delta=1$ se modifica la ecuación (58) como sigue:

$$X_{\delta}^2 = \sum_{i=1}^k \frac{(|o_i - e_i| - 1/2)^2}{e_i} \quad , \quad |o_i - e_i| \geq 1/2 \quad (59)$$

Cuando $|o_i - e_i| < 1/2$, tomamos $|o_i - e_i| - 1/2$ como igual a cero. Teniendo presente esto (59) se aplica a todas las o_i y e_i . Segundo, como regla empírica debe haber por lo menos cinco datos en cada clase de frecuencia teórica. La finalidad en este caso es asegurar que la aproximación chi cuadrado es buena, es decir, que realmente se aplica la tabla chi cuadrado. En una situación en la que una o más clases tienen frecuencias teóricas menores de 5, se combinan categorías antes de calcular las diferencias, $o_i - e_i$. Al hacerlo así, es importante recordar que el número de grados de libertad es determinado con el número de clases después de la reagrupación. Por ejemplo, si el número inicial es diez

clases y tres de estas tienen pequeñas frecuencias teóricas, se pueden reunir estas tres clases en una y obtenemos $k=8$ clases que comparar. Tenemos, o_i y e_i deben ser frecuencias absolutas, no frecuencias relativas o cualquier clase de densidad.

Se considera ahora la aplicación de chi cuadrado a hipótesis de prueba acerca de frecuencias.

8.3.1. Prueba de la bondad del ajuste

En muchas situaciones problemáticas las condiciones prevalecientes hacen que parezca razonable que la población siga un modelo de probabilidad específico; sin embargo, es benéfico efectuar un procedimiento de prueba estadística para substanciar la suficiencia de tal modelo. Suponiendo la normalidad en los datos. Se duda del supuesto de normalidad y se busca un procedimiento apropiado para aprobarlo o desaprobarlo. El procedimiento estadístico empleado para este objeto se llama prueba de bondad del ajuste.

Comienza, como lo hace el procedimiento general de prueba, formulando la hipótesis nula de que una población dada tiene una probabilidad específica, o función densidad. Entonces, se toma una muestra al azar de la población y se observa, lo que proporciona frecuencias observadas (una distribución empírica) en términos de $\hat{\chi}$ cuadrado. Entonces, la distribución teórica, como la especificada en la hipótesis nula, es ajustada a la distribución empírica. Los valores de probabilidad teóricos para las clases, cuando se multiplican por el tamaño de la muestra, se convierten en frecuencias esperadas. Después de esto, la prueba chi cuadrado de bondad del ajuste sigue exactamente el mismo procedimiento que el de frecuencias, excepto que las frecuencias teóricas son calculadas ahora generalmente por estadística de muestra como estimadores de los correspondientes parámetros de población.

8.3.2. Pruebas de independencia: Pruebas con tablas de contingencias

Las pruebas de frecuencias y bondad del ajuste se ocupan de poblaciones multinomiales. En ambos casos, las poblaciones y las muestras son clasificadas de acuerdo con un solo atributo. Ahora se mostrará que la misma técnica de pruebas chi cuadrado puede ser aplicada a datos distintos de las distribuciones multinomiales.

Cuando la población y la muestra son clasificadas según dos o más atributos, se pueden usar pruebas de independencia para determinar si los atributos son estadísticamente independientes. Por ejemplo, una muestra al azar de n tiendas detallistas puede ser clasificada cruzadamente por magnitud de capitalización y por tipo de sociedad. La proporción de cada una de las clases en la población es desconocida. El interés sería establecer si existe alguna relación de dependencia entre la capitalización de una tienda y su tipo de propiedad. Claramente, en un caso como este, se desea probar la hipótesis de que la magnitud del capital es independiente del tipo de propiedad contra la hipótesis de que se relaciona n o son dependientes.

Las pruebas de independencia se llaman también pruebas con tablas de contingencias. Hasta ahora, solo se han mostrado las tablas de clasificación en un sentido, porque, en cada caso, las frecuencias observadas han ocupado una sola fila o una sola columna. También porque las frecuencias observadas son distribuidas en k clases (columnas o filas), y las tablas de clasificación en un sentido se llaman tablas $1 \times k$ (léase 1 por k) o $k \times 1$. Extendiendo estas ideas, se pueden obtener tablas en las que las frecuencias observadas ocupan r filas y c columnas. Tales tablas se llaman a menudo tablas de contingencias. Correspondiendo a cada frecuencia observada en una tabla $r \times c$, hay una frecuencia esperada calculada por la hipótesis nula especificada. Las

frecuencias, observadas o esperadas, que ocupan las casillas de una tabla de contingencias se llaman frecuencias de casillas. La frecuencia total de cada fila o cada columna se llama frecuencia marginal.

Para evaluar diferencias entre frecuencias observadas y esperadas contenidas en tablas de contingencias, se emplea la misma estadística que para las pruebas expuestas en el punto anterior.

$$X_{\delta}^2 = \sum_{i=1}^{rc} \frac{(o_i - e_i)^2}{e_i}$$

Se suman todas las casillas rc de una tabla de contingencias; y, en general,

$$\delta = (r-1)(c-1)$$

1. Para probar la independencia con una tabla de contingencias 2×2 , entonces $\delta = (2-1)(2-1) = 1$. Como la nota anterior, el factor de corrección de continuidad de $1/2$ debe de ser usado para calcular el valor chi cuadrado. Para muestras grandes, los valores de chi cuadrado corregidos y no corregidos pueden ser prácticamente iguales, por lo que el factor de corrección de continuidad puede ser ignorado.

8.3.3. Fórmulas eficientes para calcular chi cuadrado

Existen fórmulas eficientes para calcular valores de chi cuadrado de tablas de contingencias 2×2 que solo abarcan frecuencias observadas.

Representando las frecuencias de casillas por A, B, C y D, las frecuencias marginales por m_1 , m_2 , m_3 y m_4 , y el número total de observaciones por n , mediante una tabla de contingencias 2×2 como la siguiente:

Datos observados

| | | |
|-------|-------|-------|
| A | B | m_3 |
| C | D | m_4 |
| m_1 | m_2 | n |

Puede demostrarse que:

$$X_1^2 = \frac{n(AD-BC)^2}{m_1 m_2 m_3 m_4} \quad (60)$$

Cuando la muestra es pequeña, es conveniente calcular X_1^2 con un factor de corrección de continuidad. En este caso, la fórmula eficiente es

$$X_1^2 = \frac{n(|AD - BC| - n/2)^2}{m_1 m_2 m_3 m_4} \quad (61)$$

En general, se observa que

$$\begin{aligned} X_\delta^2 &= \sum_i \frac{(o_i - e_i)^2}{e_i} = \sum_i \frac{(o_i - 2o_i e_i + e_i^2)}{e_i} = \sum_i \frac{o_i^2}{e_i} - 2 \sum_i \frac{o_i e_i}{e_i} + \sum_i \frac{e_i^2}{e_i} \\ &= \left(\sum_i \frac{o_i^2}{e_i} \right) - n \end{aligned} \quad (62)$$

Esta fórmula es válida para cualquier tabla de contingencias, excepto cuando se usa una corrección de continuidad para $\delta=1$.

8.3.4. Pruebas de Homogeneidad

Las pruebas de homogeneidad se usan para determinar si dos o más muestras independientes al azar son extraídas de la misma población o de diferentes poblaciones.

La prueba chi cuadrado de homogeneidad es una extensión de la prueba chi cuadrado de independencia. En ambos casos, se trata con datos clasificados cruzadamente. Como se observa inmediatamente también, la misma estadística de prueba usada para pruebas de independencia es usada para pruebas de homogeneidad. Pero estos dos tipos de prueba son diferentes en algunos aspectos.

Primero, se asocian con diferentes clases de problemas. Las pruebas de independencia se aplican al problema de si un atributo es independiente de otro. Mientras que las pruebas de homogeneidad se aplican cuando se desea saber si diferentes muestras proceden de la misma población.

Segundo, las primeras suponen una sola muestra tomada de una población; pero las segundas suponen dos o más muestras independientes, una de cada una de las poblaciones en cuestión. Este segundo hecho también implica que, en el caso de independencia, todas las frecuencias marginales son cantidades al azar, mientras que, en el caso de homogeneidad, los totales de filas son tamaños de muestra que son números escogidos.

8.3.5. La prueba del signo para pares emparejados

Con mucha frecuencia se tienen dos conjuntos más de datos que son muestras al azar dependientes, emparejados en términos de uno o más factores. Se han encontrado antes este tipo de datos.

Primero aparecieron en la prueba t para dos muestras dependientes. Cuando solo hay dos tratamientos y las observaciones emparejadas, puede emplearse un sencillo procedimiento no paramétrico, llamado prueba de los

signos, para probar si los dos tratamientos son igualmente eficaces. Esta prueba se expone a continuación.

Sean A y B dos tratamientos con n pares emparejados de observaciones (x_i, y_i) . Se observa el signo de $d_i = x_i - y_i$. Si se tienen datos que no son medibles, sino solo comparables, d_i se considera como positiva si A es mejor que B y negativa si A es peor que B. Cuando existe un vínculo (es decir, $d_i = 0$), el par de observaciones vinculadas es desechado y n se reduce de conformidad con ello.

Por la hipótesis nula de que los dos tratamientos son igualmente eficaces que cualquier diferencia en un par de observaciones se debe a efectos casuales, la probabilidad de un signo + para cualquier par particular es $1/2$, y también lo es la probabilidad de un signo -. Así, si se hace que S sea el número de veces que aparece el signo menos frecuente en la serie de d_i , entonces S tiene la distribución binomial con $p = 1/2$.

Es importante observar que, para que la regla de los signos sea aplicable, no es necesario suponer que las diferencias están idénticamente distribuidas como para la prueba t . Una ventaja de la prueba de los signos es que se permite compilar cada par de observaciones en diferentes condiciones, teniendo así diferentes distribuciones. Es suficiente que H_0 sea aplicada separadamente a cada par de observaciones, y que los pares sean estadísticamente independientes (como en el muestreo binomial).

El valor crítico para una alternativa de dos lados en $\alpha = 0.05$ puede hallarse bien por la expresión

$$K = \frac{(n-1)}{2} - (0.98)^n \quad (63)$$

H_0 es rechazada si $S \leq K$ para la prueba de los signos.

Se observa que la prueba de los signos es aplicable solo cuando n no es demasiado pequeña. La ecuación (63) se deriva de una prueba chi cuadrado de bondad del ajuste de una distribución binomial. La regla empírica de que $e_i \geq 5$ en tal prueba chi cuadrado implica que $n \geq 10$ es la regla empírica análoga para la validez de (63). Para n más pequeña, puede usarse la distribución binomial exacta, pero la potencia de la prueba de los signos debe ser considerada. Para $n=5$, la probabilidad de que aparezca cero veces el signo menos frecuente y que el signo más frecuente aparezca cinco veces, calculada como antes usando la distribución binomial exacta, es $(0.5)^5 = 0.03125$, y la probabilidad análoga para $n=6$ es $(0.5)^6 = 0.015625$. así, para $n=5$, la hipótesis nula será aceptada siempre en un nivel de significación de 0.05, independientemente de los que muestren los datos, porque $0.03125 > 0.05/2$, y n debe ser por lo menos seis para un nivel de significación de 0.05, para tener alguna posibilidad de rechazar la hipótesis nula.

8.4. Prueba Wald-Wolfowitz para dos muestras independientes

La prueba Wald-Wolfowitz ha sido ideada para probar la hipótesis nula de identidad entre distribuciones de dos poblaciones con dos muestras independientes sobre la base de series. La prueba comienza clasificando las observaciones de las dos muestras, con n_1 y n_2 , en un solo conjunto. Por comodidad, si se identifica una observación de la primera muestra del conjunto como X y se identifica una observación de la primera muestra del conjunto como X y se identifica la de la segunda muestra como Y , se obtendrá entonces series o grupos de X y Y . Por ejemplo, considerando estas dos muestras,

Primera muestra: 36, 20, 37, 40, 77, 61, 90, 74, 30 y 52;

Segunda muestra: 11, 16, 23, 39, 43, 41, 55, 22, 66 y 27.

Ahora, combinando estas observaciones y disponiéndolas en un conjunto, junto con sus designaciones de X y Y, tenemos doce series en función de X y Y:

Y Y - X- Y Y Y - X X X -Y- X- Y Y- X - Y- X - Y - X X X
 11 16 20 22 23 27 30 36 37 39 40 41 43 52 55 61 66 74 77 90

Obteniendo, en primer lugar, en cualquiera ordenación de observaciones procedentes de dos muestras debe haber por lo menos dos series y, como máximo $n_1 + n_2$ series en el conjunto. Después, si las dos muestras son extraídas independientemente de dos poblaciones con distribuciones idénticas, las observaciones de las muestras deben ser mezcladas muy bien cuando se ordenen según magnitudes. En otras palabras, hay que tener un gran número de series. Pero si las muestras son escogidas de poblaciones con diferentes distribuciones, entonces se obtendrá un pequeño número de series. Así, solo un pequeño número de series puede conducir a rechazar la hipótesis nula. Estos argumentos proporcionan la razón de las pruebas basadas en cálculos de probabilidades bastante sencillos.

Sea R el número de series entre las $n_1 + n_2$ observaciones; la distribución de R por muestreo se da en el apéndice 7 para $n_1 \leq 10$ y $n_2 \leq 1$, y también para $n_1 = n_2 = 11$ a 100. Por ejemplo, para la ilustración, $n_1 = n_2 = 10$ y $R = 12$ corresponde al 75.8-ésimo percentil de la distribución de R . Este valor se encuentra entre 0.01 y 1.00; por consiguiente, se acepta la hipótesis nula de que las dos poblaciones progenitoras están idénticamente distribuidas en $\alpha=0.01$. Aproximadamente, r está normalmente distribuida cuando $n_1 \geq 10$ y $n_2 \geq 10$, con

$$E(R) = \frac{2n_1n_2}{n_1+n_2} + 1 \quad (64)$$

$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} \quad (65)$$

Así, cuando las muestras son grandes, la prueba Wald-Wolfowitz puede efectuarse por la estadística de prueba

$$z = \frac{R - E(R)}{\sigma^2} \quad (66)$$

en la forma usual.

La prueba Wald-Wolfowitz también puede aplicarse a una sola muestra, y cualquier criterio de producir una ordenación de X y Y puede ser empleado.

8.5. Prueba Wilcoxon para dos muestras emparejadas

El problema de tratar los pares emparejados fue abordado en la prueba de los signos, que toma en consideración simplemente el signo de una diferencia entre cada par de valores. La prueba de los signos ignora el hecho de que un par de puntuaciones no solo tiene una diferencia, sino también una calificación entre tales diferencias. La prueba Wilcoxon toma en cuenta ambas características, signos y calificaciones de diferencias; por tanto, se la llama a veces prueba de los signos y calificaciones.

La prueba de los signos y las calificaciones es quizá la más poderosa de las pruebas no paramétricas de la identidad entre dos distribuciones de poblaciones usando pares emparejados. Cuando las observaciones se dan en n pares, primero se observan sus diferencias, d_i , y luego se clasifican estas diferencias según sus magnitudes absolutas. A continuación, se asignan los signos de estas calificaciones a sus correspondientes diferencias. Finalmente, se halla la estadística de la prueba, T. Para esto, se calculan dos sumas de calificaciones, una para todas las calificaciones con signos + y la otra para todas las calificaciones con signo -. Ambas sumas se consideran positivas. La menor de estas dos sumas de calificaciones es T.

La hipótesis nula es que las dos FDA son idénticas; es decir, que $F(x)=G(x)$ para todas las x . Cuando la hipótesis nula es cierta, cada uno de los 2^n conjuntos posibles de calificaciones con signos (obtenidos asignando arbitrariamente signos + y - a los pares 1 a n) tiene igual probabilidad de ocurrir. Sobre esta base, puede hallarse la distribución exacta de T .

Para probar $F(x) = G(x)$ contra $F(x) \neq G(x)$, puede obtenerse el valor crítico aproximado para la menor calificación total por la siguiente expresión:

$$T_c = z_{\alpha/2} \sqrt{\frac{n(n+1)(2n+1)}{24} + \frac{n(n+1)}{4}} - 1/2 \quad (67)$$

Esta aproximación se acerca a $n > 16$. Para $n < 16$, debe usarse el cuadro 4.

La prueba de los signos y las calificaciones es de dos colas, aunque de la definición de T solo trabaja con la cola inferior de la distribución de T por muestreo. Así, el cuadro especial solo muestra el valor crítico inferior de T , y así hace (67). En (67), $z_{\alpha/2}$ es siempre negativa. Por ejemplo, para $\alpha=0.05$, $z_{0.05/2} = z_{0.025} = -1.96$.

Cuando $n \geq 16$, la distribución T por muestreo es aproximadamente normal con:

$$E(T) = \frac{n(n-1)}{4} \quad (68)$$

Y

$$V(T) = \frac{n(n-1)(2n+1)}{24} \quad (69)$$

| N= numero de pares | Rechace H0 si es menor o igual a | |
|--------------------|----------------------------------|----------------|
| | $\alpha= 0.05$ | $\alpha= 0.01$ |
| 7 | 2 | 0 |
| 8 | 2 | 0 |
| 9 | 6 | 2 |
| 10 | 8 | 3 |
| 11 | 11 | 5 |
| 12 | 14 | 7 |
| 13 | 17 | 10 |
| 14 | 21 | 13 |
| 15 | 25 | 16 |

Tabla 2 Valores críticos de T cuando $n < 16$ en la prueba Wilcoxon de los signos y las calificaciones

Por tanto, para muestras grandes:

$$z = \frac{T - E(T)}{\sigma_T} \quad (70)$$

como estadística de prueba en la prueba de los signos y calificaciones. Una mejor estadística de prueba aquí incluye un factor de corrección de continuidad, y es,

$$z = \frac{T + 1/2 - E(T)}{\sigma_T} \quad (71)$$

Cuando hay vínculos del tipo $d_i=0$ en esta prueba, son ignorados y n es ajustada hacia abajo correspondientemente. Cuando hay vínculos del tipo, por ejemplo, $d_i=-5$ y $d_j=+5$, y cuando las calificaciones para ellos son, por ejemplo, 3 y 4, se asignará una calificación de $(3+4)/2 = 3.5$ a cada uno. Para -5 , la calificación con signos es -3.5 , y para 5 , $+3.5$.

Hay formas de modificar esta prueba para hipótesis alternativas de una sola cola, pero no se presentarán aquí.

8.6. Prueba H Kruskal - Wallis

W.H. Kruskal y W. A. Wallis han ideado una prueba como la prueba Wilcoxon para dos muestras dependientes que cubren c muestras independientes. Este procedimiento, llamado prueba H, ha sido ideada para probar la hipótesis nula de que las c muestras independientes han sido extraídas de poblaciones de distribuciones idénticas, es decir FDA idénticas.

La prueba H comienza calificando todas las observaciones; luego se halla la suma de las calificaciones de cada muestra, T_i , y después el total de todas las sumas de calificaciones T , donde $T = \sum T_i$. Puede verse entonces que

$$T = \frac{n(n+1)}{2} \quad (72)$$

donde $n = n_1 + n_2 + \dots + n_c$. Se sabe que si $n_i \geq 3$, la estadística

$$H = \left[\frac{12}{n(n+1)} \sum_{i=1}^c \frac{T_i^2}{n_i} \right] - 3(n+1) \quad (73)$$

está distribuida aproximadamente como chi cuadrado con $\delta = c - 1$.

Cuando aparecen vínculos en la calificación, se asignará un promedio de las calificaciones contiguas a cada una de las observaciones vinculadas. Por ejemplo, si dos observaciones están vinculadas y si las calificaciones para ellas deben ser 5 y 6, se asignarán a ambas observaciones una calificación de $(5+6 / 2) = 5.5$. Nuevamente, si hay un vínculo de tres observaciones, siendo las calificaciones, por ejemplo, de 10, 11 y 12, la calificación será de 11 a cada una de las tres observaciones. Y así sucesivamente.

Un valor ajustado de H, designado H', debe ser calculado cuando hay vínculos. Entonces:

$$H' = \frac{H}{C} \quad (74)$$

Donde C, el factor de corrección, se define como

$$C = 1 - \frac{\sum_{j=1}^k (t_j^3 + t_j)}{n^3 - n} \quad (75)$$

En la definición de C, k es el número de conjuntos vinculados y t_j es el número de observaciones vinculadas con el conjunto j.

La prueba H se compara razonablemente bien con la prueba F en un análisis de variancia. Pero, cuando las muestras son grandes, es muy pesado calificar las observaciones, a menos que se disponga de una computadora.

8.7. Pruebas de Kolmogorov - Smirnov y Anderson Darling

Estas pruebas son para diferencias entre distribuciones acumulativas. La prueba de una muestra se refiere al acuerdo entre una distribución acumulativa observada, o empírica, de valores muestrales y una función de distribución continua especificada; por lo tanto, es una prueba de bondad del ajuste. La prueba de dos muestras se refiere al acuerdo entre dos distribuciones acumulativas observadas; prueba la hipótesis de si las dos muestras independientes proceden de distribuciones continuas idénticas, y es sensible a diferencias de población respecto de ubicación, dispersión o sesgadura.

La prueba de una muestra de Kolmogorov -Smirnov es generalmente mas eficiente que la prueba chi cuadrado para bondad del ajuste de muestras pequeñas, y puede servir para muestras muy pequeñas en las que no se aplica la prueba chi cuadrado. Debe recordarse, sin embargo, que a diferencia de la prueba de Kolmogorov - Smirnov, la prueba chi cuadrado sí se puede usar en relación con distribuciones discretas.

La prueba de una muestra se basa en la diferencia máxima absoluta D entre los valores de la distribución acumulativa empírica de una muestra aleatoria de tamaño n y una distribución acumulativa teórica especificada. Para determinar si esta diferencia es mayor de lo que razonablemente podría esperarse para un nivel dado de significación, consultamos el valor crítico de D en el apéndice 8.

A pesar de su atractivo intuitivo, las pruebas Kolmogorov - Smirnov carecen de eficacia. Las diferencias en las colas pueden detectarse más fácilmente si la diferencia entre la distribución acumulativa empírica, F_n , y F se divide entre $F(x)(1-F(x))$. En particular, la prueba Anderson - Darling se basa en grandes valores de la estadística

$$A^2 = \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)(1-F(x))} f(x) dx \quad (76)$$

A primera vista, el cálculo numérico de esta estadística parece difícil pero es posible demostrar que

$$A^2 = - \left[\sum_{i=1}^n (2i-1)(\ln(u_i) + \ln(1-u_{n+1-i})) \right] / n \quad (77)$$

donde $u_i = f(x(i))$ es el valor de la distribución acumulativa teórica en la i -ésima observación más grande $x(i)$.

La hipótesis nula es rechazada para grandes valores de la estadística A^2 . Como pauta general, el punto de 5% de muestras grandes es de 2.492 y el punto de 1% es de 3.857. Se ha sugerido que estos valores críticos son sumamente exactos incluso para muestras tan pequeñas como 10.

CAPÍTULO 9

CONTROL ESTADÍSTICO DEL PROCESO

9.1. Como funciona la gráfica de control.

9.1.1. Empleo de las gráficas de control para juzgar si existe o no un sistema constante de causas especiales.

Las gráficas \bar{X} , R y s, ofrecen una base para un criterio de juicio de gran importancia práctica, que se puede expresar en diferentes formas tales como:

¿Se extrajeron todas las muestras de la misma población?

¿Hay algún universo del cual parezcan provenir estas muestras?

¿Indican estas cifras un patrón estable de variación?

¿Es esta variación el resultado de un sistema de causa constante?

O simplemente ¿Indican estas mediciones que hay control estadístico?.

Cualquier regla que se pudiera establecer para dar una respuesta definitiva de "sí" o "no", puede dar en ocasiones, una respuesta errónea. La decisión de trazar la línea que separe una respuesta afirmativa de una negativa, debe estar basada en la acción que se tomará en cada caso.

En el control de calidad en la manufactura, la respuesta "no, no es un sistema de causa constante", provoca la búsqueda de una causa de variación asignable y un intento para eliminarla, si es posible. La respuesta "sí, es un sistema de causas constantes", hace que no se haga nada al proceso solo, sin ningún esfuerzo por hallar las causas de la variación. La regla para establecer los límites de control que determinan el "sí" o el "no", en cualquier caso repercute en un equilibrio económico entre los costos debidos a dos clases de error:

1.- El error de buscar un problema cuando no lo hay (siempre que la respuesta "no es incorrecta) y,

2.- El error de dejar el proceso solo para no tener que buscar el problema cuando en realidad está allí presente (siempre que la respuesta "sí" sea correcta).

Cualquier regla para establecer límites de control en la manufactura debería ser práctica y basarse en el punto de equilibrio o balanceo económico. La experiencia indica que en la mayoría de los casos, los límites de 3-sigmas, en realidad, dan un equilibrio económico satisfactorio entre estos dos tipos de errores (Aunque, en principio, la elección de límites es cuestión de minimizar la suma de ciertos costos, hay grandes dificultades prácticas para efectuar buenas estimaciones de los costos unitarios pertinentes).

9.1.2. Patrones de variación estable e inestable

Para ejemplificar esto en la figura 22 se presenta la distribución de frecuencias de 260 pesos drenados del obtenido de latas de tomates obtenidos en un periodo de 11 días. A primera vista, parecería que esta distribución presenta una imagen satisfactoria de un patrón de variación que se produjo durante los 11 días.

Tal enunciado no es verdadero en su totalidad. Puede ser que el patrón de variación haya tenido diversos cambios durante el período. La primera pregunta que se debe responder es si hubo o no un patrón estable. Si hay pruebas de falta de estabilidad, puede responder si hubo o no un patrón estable. Si hay pruebas de falta de estabilidad, puede ser cierto que los 260 pesos representan muestras de cierto número de universos diferentes que estuvieron presentes en diferentes momentos; en este caso, la distribución de frecuencia resultante representaría un promedio ponderado de los diversos universos que había cuando se tomaron las muestras.

Siempre que va a emplear una distribución de frecuencias como base para estimar las capacidades de un procedimiento de manufactura, hacer un cambio propuesto en un procedimiento o modificar las especificaciones, la gráfica de control se vuelve un instrumento importante para establecer si hay estabilidad. Una vez establecida ésta, se puede sentir una confianza relativa un que está midiendo un universo y no un promedio ponderado de muchos.

Al anotar los datos para establecer distribución de frecuencias, sí se debe conservar el orden en que se obtuvieron.

| | |
|------|---|
| 25.5 | |
| 25.0 | |
| 24.5 | I |
| 24.0 | |
| 23.5 | |
| 23.0 | |
| 22.5 | I |
| 22.0 | |
| 21.5 | I |
| 21.0 | I |
| 20.5 | |
| 20.0 | |
| 19.5 | |
| 19.0 | |
| 18.5 | |
| 18.0 | |
| 17.5 | |
| 17.0 | |
| 16.5 | I |
| 16.0 | |
| 15.5 | |

Figura 22. Hoja de verificación para determinar la distribución de frecuencia de datos del apéndice 9

9.1.3. Empleo de la gráfica de control para interpretar una distribución de frecuencia.

En la figura 23, se presentan las gráficas de control de \bar{X} y R de los datos básicos de los pesos drenados contenidos de las latas de tomate del apéndice 9. Ninguno de los puntos de \bar{X} está fuera de los límites de control; uno de los 52 puntos de R está fuera de los límites. La aplicación de las reglas secundarias basadas en la teoría de las corridas, que se describe más adelante, indican ciertos ligeros desplazamientos durante el período de 11 días. Aunque el empleo de estos datos para estimar μ y σ pudiera estar sujeto a deliberaciones, no por necesidad quiere decir que esta gran desviación desde un patrón constante de variación sea insatisfactoria desde un punto de vista práctico. Muchos procesos de producción ni siquiera se aproximan a este comportamiento.

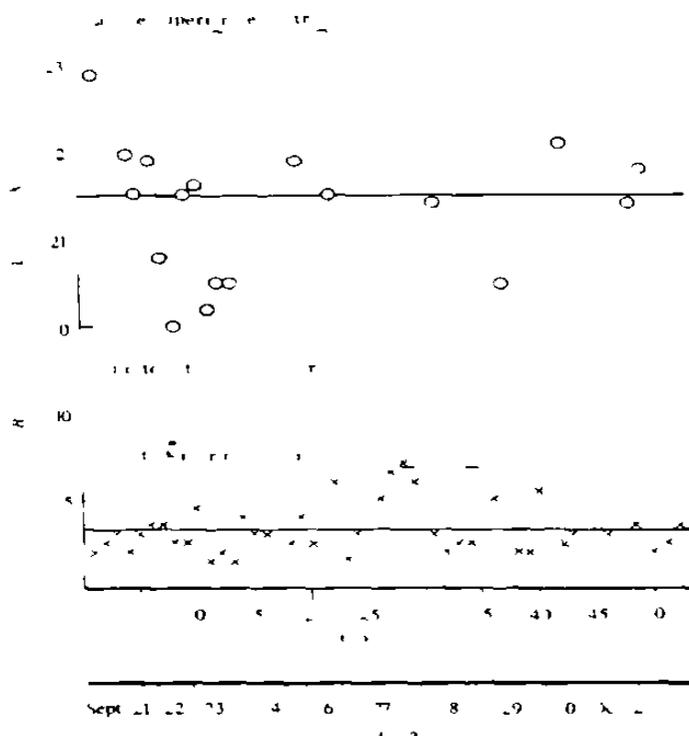


Figura 23 Gráficas de control para \bar{X} y R para los pesos drenados del contenido de latas de tomate, datos del apéndice 9.

9.1.4. Cálculo de los límites de 3 - sigmas en las gráficas de control \bar{X}

El empleo de los apéndices 10, 11 y 12 para el cálculo de los límites de control se puede ilustrar con la utilización de los subgrupos de la tabla del apéndice 13.

Después de calcular los promedios y los recorridos de los subgrupos, el siguiente paso en el cálculo de los límites es encontrar $\bar{\bar{X}}$ y \bar{R} . Para los subgrupos, son:

$$\hat{\mu} = \bar{\bar{X}} = \frac{\sum \bar{X}_i}{20} = \frac{577.75}{20} = 28.9$$

$$\bar{R} = \frac{\sum R_i}{20} = \frac{368}{20} = 18.4$$

Si se utiliza el apéndice 10, el siguiente paso es calcular σ . Para ello, es necesario encontrar el factor d_2 del tamaño del subgrupo. En este caso $n=4$ y se encuentra que $d_2 = 2.059$.

$$\text{Estimación de } \sigma, \hat{\sigma} = \frac{\bar{R}}{d_2} = \frac{18.4}{2.059} = 8.94$$

Después se puede calcular $3\sigma_{\bar{X}}$ a partir de la relación $\sigma_{\bar{X}} = \hat{\sigma} / \sqrt{n}$

$$3\sigma_{\bar{X}} = \frac{3\hat{\sigma}}{\sqrt{n}} = \frac{3(8.94)}{\sqrt{4}} = 13.4$$

$$\text{Límite superior de control } \bar{\bar{X}} + 3\sigma_{\bar{X}} = 28.9 + 13.4 = 43.3$$

$$\text{Límite inferior de control } \bar{x} = \bar{X} - 3\sigma_{\bar{x}} = 28.9 - 13.4 = 15.5$$

Los dos pasos en el cálculo de $3\sigma_{\bar{x}}$ se podrían consolidar como

$$3\sigma_{\bar{x}} = \frac{3R}{d_2 \sqrt{n}} = \frac{3}{2.059 \sqrt{4}} \bar{R} = 0.73 \bar{R} = 0.73 (18.4) = 13.4$$

Para abreviar el cálculo de los límites de control a partir de \bar{R} se ha calculado este factor $3/d_2\sqrt{n}$ multiplicador de \bar{R} en el cálculo anterior, para cada valor de n desde 2 hasta 20 y aparece en el apéndice 11. Este factor se designa A_2 . Entonces, las formulas para los límites de control de 3-sigmas en las gráficas para \bar{X} quedan en la siguiente forma:

$$UCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = \bar{X} + A_2\bar{R}$$

$$LCL_{\bar{x}} = \text{Límite inferior de control } \bar{x} = \bar{X} - A_2\bar{R}$$

Si se van a calcular los límites de control a partir de \bar{s} en vez de \bar{R} , los cálculos de los subgrupos del apéndice 13 son:

$$\bar{X} = 28.9$$

$$\bar{s} = \frac{\sum S_i}{20} = \frac{165.2}{20} = 8.26$$

Con el empleo del factor c_4 del apéndice 10 para calcular σ .

$$\text{Estimación de } \sigma, \hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{8.26}{0.9213} = 8.97 \quad 3\sigma_{\bar{x}} = \frac{3\hat{\sigma}}{\sqrt{n}} = \frac{3(8.97)}{\sqrt{4}} = 13.5$$

$$UCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = 28.69 + 13.5 = 43.4$$

$$LCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = 28.9 - 13.5 = 15.4$$

Al igual que en los cálculos a partir de \bar{R} los dos pasos en el cálculo de $3\sigma_{\bar{x}}$ se puede consolidar como

$$3\sigma_{\bar{x}} = \frac{3}{c_4} \bar{s} = 1.63 (8.26) = 13.5$$

Para abreviar estos cálculos de los límites de control a partir de \bar{s} , se ha calculado el factor $3/c_4\sqrt{n}$ multiplicador de \bar{s} para cada valor de n desde 2 hasta 25 y, luego, de 5 en 5 hasta 100 y se incluye en el apéndice 12. Este factor se designa A_3 . Las fórmulas para los límites de control 3-sigmas con el empleo de este factor son:

$$UCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = \bar{\bar{X}} + A_3\bar{s}$$

$$LCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = \bar{\bar{X}} - A_3\bar{s}$$

Para las situaciones en que se desea calcular los límites de control en forma directa a partir de los valores conocidos o estándar de σ y μ , se ha calculado el factor $3/\sqrt{n}$ y se incluye en el apéndice 14. Este factor se designa A . Las fórmulas para los límites de control de 3-sigmas con el empleo de este factor son:

$$UCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = \mu + A\sigma \text{ o } \bar{X}_0 + A\sigma_0$$

$$LCL_{\bar{x}} = \text{Límite superior de control } \bar{x} = \mu - A\sigma \text{ o } \bar{X}_0 - A\sigma_0$$

Las diversas ecuaciones para las líneas centrales y los límites de 3-sigmas en las gráficas de control de \bar{X} , R y s se han reunido para facilitar la consulta, en el cuadro 5. Los factores (a, A₁, etc.) que se mencionaron se dan en la sección de apéndices.

Se aprecia que la dispersión de los límites en las gráficas para \bar{X} y en las R o s depende de la dispersión del proceso. Los límites en todas las gráficas se pueden calcular en forma directa a partir de σ conocida o supuesta o al estimar σ a partir de \bar{R} o de \bar{s} . En casi todos los casos, en la industria, los límites se calculan a partir de \bar{R} .

| Método | Gráfica \bar{X} | Gráfica R | Gráfica s |
|--|---|--|--|
| μ y σ conocidas o supuestas (X_0, σ_0) | $CL = \bar{X}_0 = \mu$
$UCL_X = \mu + A\sigma$
$LCL_X = \mu - A\sigma$
Apéndice 13 | $CL = R_0 = d_2\sigma$
$UCL_R = D_2\sigma$
$LCL_R = D_1\sigma$
Apéndices 9 y 13 | $CL = s_0 = c_4\sigma$
$UCL_s = B_6\sigma$
$LCL_s = B_5\sigma$
Apéndices 9 y 13 |
| μ y σ estimadas a partir de \bar{X} y R | $CL = \bar{X}$
$UCL_X = \bar{X} + A_2\bar{R}$
$LCL_X = \bar{X} - A_2\bar{R}$
Apéndice 10 | $CL = \bar{R}$
$UCL_R = D_4\bar{R}$
$LCL_R = D_3\bar{R}$
Apéndice 10 | |
| μ y σ estimadas a partir de \bar{X} y \bar{s} | $CL = \bar{X}$
$UCL_X = \bar{X} + A_3\bar{s}$
$LCL_X = \bar{X} - A_3\bar{s}$
Apéndice 11 | | $CL = \bar{s}$
$UCL_s = B_4\bar{s}$
$LCL_s = B_3\bar{s}$
Apéndice 11 |

Tabla 3. Ecuaciones para calcular los límites 3-sigma en las gráficas de control para variables.

9.1.5. Gráficas de control para recorrido y desviación estándar de la muestra

Una formula general para los limites de control en la gráfica \bar{X} es $\mu + 3\sigma_{\bar{x}}$. En forma similar, las formulas generales para las gráficas de control de las mediciones de la dispersión de subgrupos son:

1. Para la gráfica de R : $\bar{R} \pm 3\sigma_R$
2. Para la gráfica de s : $\bar{s} \pm 3\sigma_s$

Pero, cuando se aplican estas tres formulas al cálculo de los límites inferiores de control (LIC), resultará que esos límites son menores de 0 cuando n es de 6 o menos en la gráfica R y cuando n es menor de 5 para la gráfica s. Debido a que R y s no pueden ser menores que 0, en estos casos no se utiliza el límite inferior de control.

Las fórmulas para el cálculo de los limites de control para recorrido y desviación estándar de la muestra se muestran en la tabla 3.

9.1.6. Clasificación de las formas que puede ocurrir la falta de control :

Debido a que la falta de control corresponde a un universo de la muestra distinto, se puede hacer una clasificación de los diferentes tipos de falta de control, es una clasificación de las formas en que pueden diferir entre sí dos universos de muestra.

1. Pueden diferir sólo en el promedio.
3. Pueden diferir sólo en la dispersión.
4. Pueden diferir en el promedio y en la dispersión.

Los cambios o desplazamientos en el promedio del universo influyen en las gráficas de control de \bar{X} y R en una forma; los cambios en la dispersión del universo influyen en ellas en otras formas.

Los cambios en el universo pueden ser continuos durante cierto tiempo, como si se extrajeran muchos subgrupos de un universo y, luego, muchos otros más de otro universo. O bien, los cambios pueden ser frecuentes e irregulares como si hubiera una gran cantidad de universos y las extracciones de cada universo fueran a continuar en periodos de diferente duración. O bien, los cambios, pueden ser graduales y sistemáticos.

Esto se representa en la figura 24

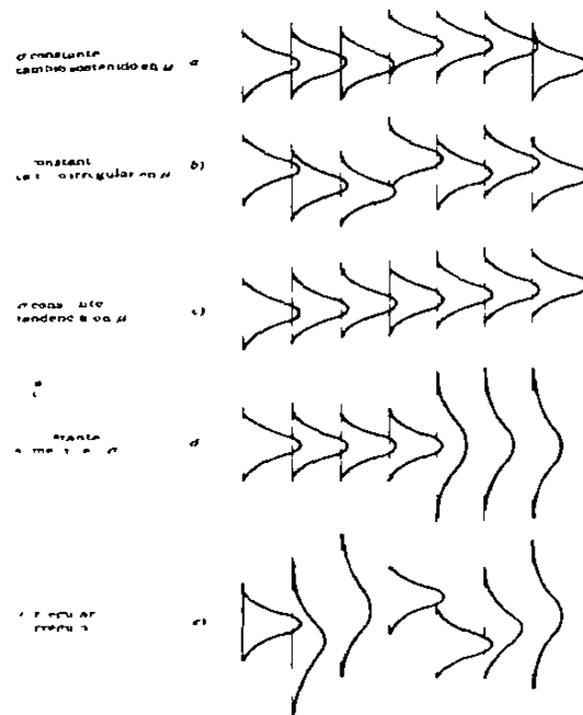


Figura 24. Los sistemas de causa ocasionales

9.1.7. Pruebas de falta de control basadas en corridas de puntos por encima o por debajo de la línea central en la gráfica de control.

Para poder detectar cambios en un parámetro de un universo en las aplicaciones comunes de las gráficas de control en la manufactura, el plan más práctico es utilizar unas cuantas reglas sencillas que dependen sólo de las corridas extremas. Por ejemplo:

Se considera que hay bases para sospechar que ha cambiado el parámetro del universo cuando:

- De 7 puntos consecutivos en la gráfica de control, estén todos en el mismo lado de la línea central.
- De 11 punto consecutivos, cuando menos 10 caen en el mismo lado de la línea central.
- De 14 puntos consecutivos, cuando menos 12 caen en el mismo lado de la línea central.
- De 17 puntos consecutivos en la gráfica de control, cuando menos 14 estén en el mismo lado de la línea central.
- De 20 puntos consecutivos en la gráfica de control, cuando menos 16 estén en el mismo lado de la línea central.

Las secuencias mencionadas en cada una de estas reglas, ocurrirán en forma causal, sin que cambie el universo y con más frecuencia que un punto fuera de los límites de control 3-sigmas. Por esta razón, constituyen una base menos confiable para buscar los problemas que la aparición de un punto fuera de los límites de control.

Además, si se emplean todas estas reglas diferentes para juzgar si se ha cambiado el parámetro del universo, las posibilidades de una falsa indicación del cambio son mayores que si sólo se emplea una de las reglas.

Las secuencias de los puntos en un lado de la línea central pueden ser útiles en relación con una posible acción en la gráfica de control, aun cuando no se emplean para buscar el problema en el proceso. Son de particular utilidad en el caso en que el límite inferior de control es cero. Por ejemplo, una reducción en el valor de σ , requiere un cambio en la ubicación de los límites superior e inferior de control en la gráfica de \bar{X} y un cambio en la línea central y límites en la gráfica de R. Pero, en todas las gráficas de R en donde el tamaño del subgrupo no es mayor de seis, es imposible que se muestre el cambio en σ con un punto que quede fuera del límite inferior de control de cero. Pero, se puede obtener una indicación de la disminución en la dispersión del proceso si se aplica la teoría de las corridas mencionadas anteriormente.

9.1.8. Interpretación de los patrones de variación en las gráficas \bar{X} y R

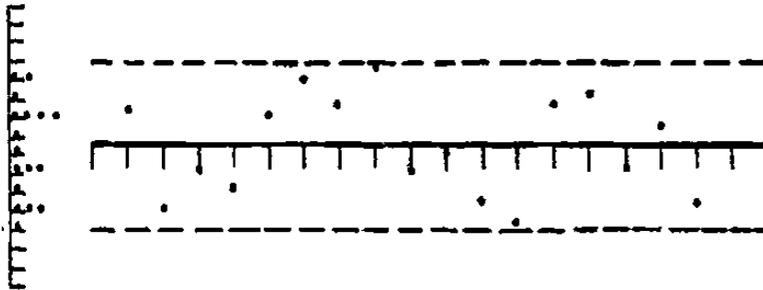
Un aspecto importante del empleo de las gráficas de control es la interpretación de los muchos patrones de datos que puedan tener. Además de la regla principal de buscar el problema, también se han comentado reglas secundarias o complementaria, basadas en la teoría de las corridas.

La gráfica de control indica cuándo buscar el problema, pero no puede indicar por sí sola dónde se debe buscar ni qué causa se va a encontrar.

En la figura 25 se ilustran algunos de los patrones que se ven con frecuencia en las gráficas de control y se expresan algunas causas posibles de estos patrones.

Algunos cambios que alteran la gráfica de \bar{X} (variabilidad estabilizada)

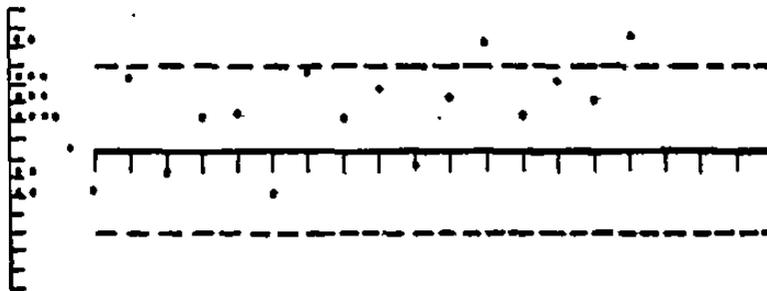
Algunas causas que alteran la gráfica R



(a) Ciclos recurrentes

1. Cambios en la temperatura u otros recurrentes en el entorno físico
2. Fatiga de los operarios
3. Diferencias en los aparatos para medición o pruebas que se utilizan al efecto
4. Rotación normal de máquinas o de operarios
5. Combinación de subconjuntos u otros procesos

1. Mantenimiento preventivo programado
2. Fatiga de los operarios
3. Herramientas gastadas



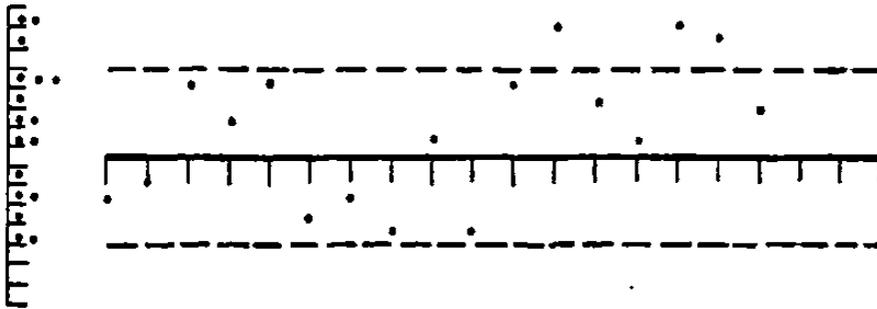
(b) Tendencias

1. Deterioro gradual del equipo que influye en todos los artículos
2. Fatiga de los operarios
3. Acumulación de productos de desecho
4. Deterioro de las condiciones ambientales

1. Mejora o detrimento de la pericia del operario
2. Fatiga de los operarios
3. Cambio en las proporciones de subconjuntos que alimentan una línea de ensamble
4. Cambio gradual en la homogeneidad de la calidad del material recibido.

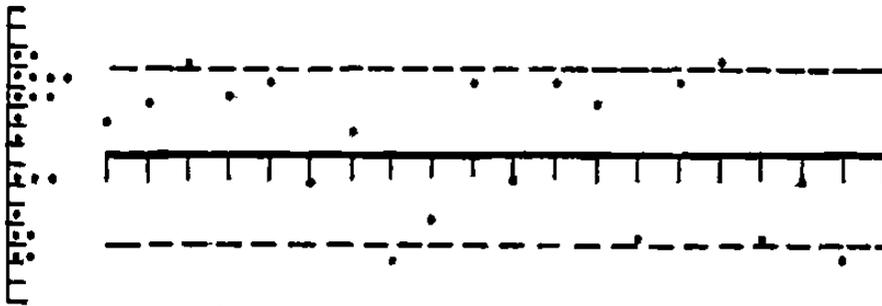
Algunos cambios que alteran la gráfica de \bar{X} (variabilidad estabilizada)

Algunas causas que alteran la gráfica R



(c) Cambios bruscos en el grado del proceso

- | | |
|---|--|
| <ol style="list-style-type: none"> 1. Cambio en las proporciones de material o subconjuntos que vienen de diversas fuentes 2. Operario o máquina nuevos 3. Modificación del método o proceso de producción 4. Cambio en el aparato o método para inspección | <ol style="list-style-type: none"> 1. Cambio en el material 2. Cambio en el método 3. Cambio del operario |
|---|--|

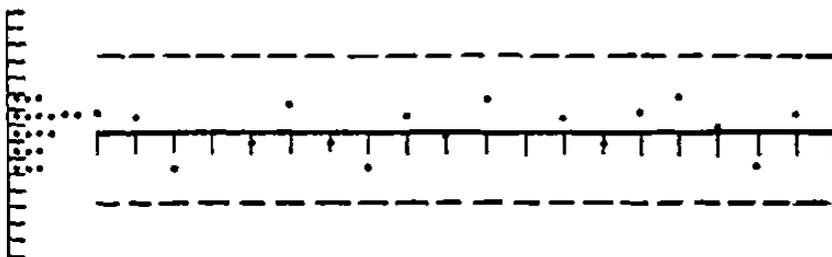


(d) Alta proporción de puntos cerca o fuera de los límites

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Exceso de control 2. Grandes diferencias sistemáticas en la calidad del material 3. Grandes diferencias sistemáticas en el método o equipo para prueba 4. Llevar control de dos o más procesos en la misma gráfica | <ol style="list-style-type: none"> 1. Mezcla de materiales de calidades notoriamente diferentes 2. Empleo de una sola gráfica de R para diversos operarios 3. Datos de procesos en diferentes condiciones, trazados en la misma gráfica |
|--|---|

Algunos cambios que alteran la gráfica de \bar{X} (variabilidad estabilizada)

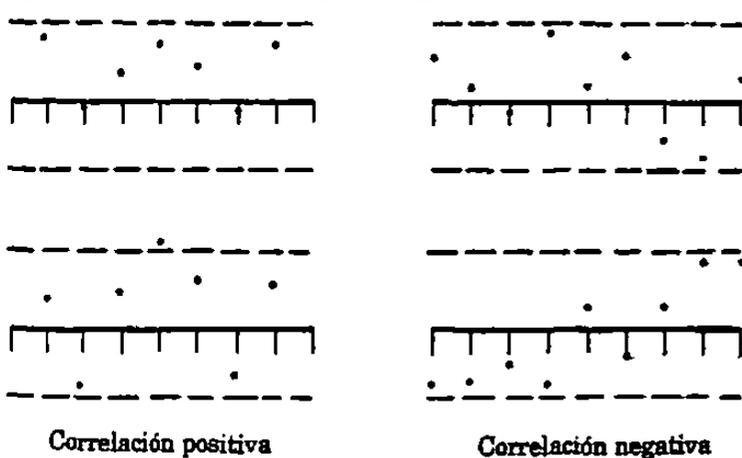
Algunas causas que alteran la gráfica R



(e) *Estratificación o falta de variabilidad*

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Cálculo incorrecto de los límites de control 2. Subgrupos incorrectos; la gráfica R recibe más variabilidad que la gráfica \bar{X} | <ol style="list-style-type: none"> 1. Recopilación, en cada muestra, de cierto número de mediciones provenientes de universos muy variados |
|--|---|

Correlación entre las gráficas



(f) *Correlación entre gráficas \bar{X} y R del mismo proceso*

1. Sesgamiento del universo subyacente
2. Puntos generados desde la misma muestra

(g) *Correlación entre gráficas \bar{X} o R diferentes*

1. Puntos generados desde la misma muestra
2. Relaciones desconocidas entre causa y efecto

Figura 25 Algunas interpretaciones de los patrones de las gráficas de \bar{X} y R

9.2. Formación racional de subgrupos. Análisis de capacidad del proceso

9.2.1. La información dada por la gráfica de control depende del criterio empleado para la selección de los subgrupos.

Un posible aspecto de la gráfica de control es que se suministra una prueba estadística para determinar si la variación entre un subgrupo y otro no es congruente con la variación promedio dentro de los subgrupos. Si se desea determinar un grupo de mediciones tiene o no homogeneidad estadística (es decir, si parecen provenir de un sistema constante de causas ocasionales), los subgrupos se deberían seleccionar en una forma que parezca dar la máxima probabilidad de que las mediciones en cada subgrupo sean iguales así como los subgrupos difieran entre sí.

Para demostrar esto, se puede tomar un conjunto de mediciones que ya se hayan clasificado en subgrupos de acuerdo con su orden de producción y que muestren una clara falta de control con base en las pruebas de las gráficas \bar{X} y R. Se escribe cada medición en un pedazo de papel separado, se colocan en un a caja o tazón, se revuelven y extraen uno por uno sin volverlos al tazón. Se anotan los valores escritos en los pedazos de papel en el orden en que se extrajeron. Con estos valores registrados, se trazan las gráficas \bar{X} y R. Si se ha revuelto bien los pedazos de papel, estas nuevas gráficas mostrarán que hay control. Al revolver los papeles se han sustituido las causas asignables originales por las causas ocasionales como base para la diferencia entre los subgrupos.

La base para la formación de subgrupos requiere un estudio cuidadoso con miras a obtener la cantidad máxima de información útil de cualquier gráfica

de control. Como se mencionó, la base racional para la formación de subgrupos, es el orden de producción.

9.2.2. Dos métodos en que se emplea el orden de producción como base para la formación de subgrupos

Cuando se emplea el orden de producción como base para la formación de subgrupos, hay dos enfoques fundamentales muy diferentes:

1. El primer subgrupo consta de todo el producto producido lo más cerca posible que se pueda a cierta hora; el siguiente subgrupo consta de todos los productos producidos lo más cerca posible de cierto tiempo después que el primero y así sucesivamente. Por ejemplo, si el tamaño del subgrupo es cinco, el inspector que efectúa las mediciones cada hora, podría medir los últimos cinco artículos producidos justo antes de cada una de sus visitas programadas a la máquina. Esto es posible cuando se trata de piezas hechas con la máquina, por ejemplo, si se las coloca en una bandeja en el orden de producción. De lo contrario, se podría lograr el mismo resultado si el inspector espera a que salgan cinco piezas de la máquina y las mide conforme salen.
2. Un subgrupo consta de un producto que se pretende sea representativo de toda la producción durante un tiempo dado; el siguiente subgrupo consta del producto que se pretende sea el representativo de la producción de más o menos la misma cantidad del producto más tarde y así sucesivamente. Cuando se acumula el producto en el punto de producción, el inspector puede tomar una muestra al azar de todo el producto manufacturado desde su última visita. Si no resulta práctico, podrían ser cinco visitas (Si $n=5$), a intervalos de producción o de tiempo iguales y se efectúa una

medición en cada visita; estas cinco mediciones constituyen un subgrupo.

En el primer método se sigue la regla para la selección de subgrupos racionales, que es de permitir una mínima posibilidad de variación dentro de un subgrupo y una máxima posibilidad de variación entre un subgrupo y otro. Se puede esperar que produzca la mejor estimación posible de un valor de σ que represente las capacidades ideales de un proceso, obtenibles si se pueden eliminar las causas asignables o atribuibles de variación, las causas especiales, entre un subgrupo y otro. Además, permite una medición más sensible de los cambios en el promedio del proceso; hace que la gráfica de control sea una mejor guía de los ajustes de la máquina o de las acciones destinadas a mantener un promedio dado del proceso. Por ello, el primer método es el más adecuado para el análisis de un proceso y para control de procesos.

Pero, si la formación de subgrupos ha sido de acuerdo con el primer método y ocurre un cambio en el promedio del proceso después de tomar un subgrupo y de corregir aquél antes del siguiente subgrupo, este cambio no se reflejará en la gráfica de control. Por esta razón, a veces se prefiere el segundo método cuando una de las finalidades de la gráfica de control es influir en las decisiones para la aceptación del producto. En tales casos, una pregunta debe hacerse antes de seleccionar uno de los dos métodos, si es posible en realidad que ocurran o no los cambios compensadores en el promedio del proceso entre subgrupos. En tal caso y si no hay otra forma de detectar un producto malo que se podría haber producido entre un subgrupo y otro (por ejemplo, una pieza que no se puede ajustar en un conjunto), es deseable el segundo método de formación de subgrupos a pesar de las otras ventajas del primer método. En los trabajos de mecanización (maquinado) esos cambios compensadores entre los subgrupos, es más fácil que ocurran en máquinas controladas a mano que en las automáticas.

Cuando se utiliza el segundo método de formación de subgrupos, la interpretación de los puntos fuera de control en la gráfica R es un tanto diferente a la del primer método. Con el segundo método, un cambio en el promedio de un proceso durante el período abarcado por un subgrupo, puede ocasionar puntos fuera de control en la gráfica R aunque no haya habido ningún cambio real en la dispersión del proceso.

La selección del tamaño de subgrupo se debe guiar, en parte, por el deseo de permitir una mínima posibilidad de variación dentro de un subgrupo. En casi todos los casos se obtendrá información más útil con, por ejemplo, cinco subgrupos de cinco que con un subgrupo de 25. En los subgrupos grandes, como el de 25, es probable que haya demasiada oportunidad para un cambio en el proceso dentro de un subgrupo. No obstante, si los cambios en el proceso no ocurren dentro de los subgrupos, entonces los subgrupos de tamaño grande tienen la ventaja, mencionada anteriormente, de disminuir el riesgo de los llamados errores de Tipo II en la interpretación de las gráficas de control. Siempre que se emplean tamaños relativos grandes de subgrupos para esta finalidad, es muy deseable emplear el primer método para formación de subgrupos.

En muchos casos en los cuales el primer método de formación de subgrupos es realmente el mejor, puede ser necesario emplear el segundo método por las razones prácticas relacionadas con la toma de las mediciones.

9.2.3. Objetivos de un análisis de la capacidad de un proceso

El problema estadístico básico en el control de calidad de un proceso es el de establecer un estado de control sobre éste, es decir, eliminar las causas especiales de variación y, luego, mantener ese estado de control en el curso del tiempo. Esto se ilustra en la figura 26-a. De no menor importancia es el

problema de ajustar el proceso al punto en el cual la casi totalidad del producto cumple con las especificaciones. El segundo 26-b tiene como objetivo el análisis de capacidad. Es decir, una vez que se ha establecido un estado de control, entonces se desvía la atención hacia la pregunta ¿Cumple el producto con las especificaciones y si no es así, se puede ajustar el proceso a un grado en que sea posible lograrlo?

Las acciones que producen un cambio o ajuste en un proceso, que están dirigidas a eliminar causas comunes, a menudo son el resultado de algún estudio de la capacidad. La comparación de los límites de tolerancia natural con los límites de especificación y del intervalo de las especificaciones pueden dar lugar a cualquiera de las siguientes posibilidades de acción:

1. Ninguna acción: Si los límites de tolerancia natural quedan bien dentro de los límites de especificaciones, por lo general, no se requiere ninguna acción. A menudo, en estos casos, se pueden hacer menos riguroso el control.
2. Acción para ajustar el centramiento: Cuando el intervalo de tolerancia natural es más o menos el mismo que el de las especificaciones, quizá un ajuste sencillo en el centramiento del proceso será todo lo que se necesite para que la casi totalidad del producto quede dentro de las especificaciones.
3. Acción para producir la variabilidad: Suele ser la acción más compleja. En los casos en que dos o más corrientes del producto se fusionan en una línea antes de la inspección, la acción puede incluir el trabajo poco complicado de poner las diversas corrientes, por separado, bajo control a un cierto valor de \bar{X}_0 . En otros casos, se puede requerir un complejo análisis de las fuentes de variación, que

dará por resultado cambios en los métodos, herramental, materiales o equipo.

4. **Acciones para cambiar las especificaciones:** Es una decisión para diseñadores y proyectistas, pero el personal de control de calidad no la debe pasar por alto. El hecho de que las especificaciones estén por escrito, no significa por necesidad que sean intocables. Por otra parte, el personal de manufactura y control de calidad no pueden simplemente hacer caso omiso de ellas, sin correr el riesgo de ocasionar graves problemas.

5. **Resignación a las pérdidas.** Cuando todo lo demás fracasa, la administración se debe resignar a tener un elevado porcentaje de pérdidas. La atención a este aspecto se concentra en los costos de los derechos y de volver a trabajar ("retrabajar"), el grado económico de control que se pueda ejercer en el proceso de producción y los costos relacionados con la selección y examen de todo el producto, a sabiendas de que todavía hay la posibilidad de aceptar artículos no conformes. En este caso, es de máxima importancia tener buenos cálculos de la forma y tamaño de la distribución del producto a fin de ayudar al centramiento correcto del proceso.

9.2.4. Etapas en el análisis de la capacidad del proceso.

En los siguientes pasos se exponen las diferentes etapas o pasos que se pueden efectuar para realizar un estudio de capacidad. Desafortunadamente, no hay dos estudios idénticos; por ello, se pueden eliminar algunas de esas etapas o pasos, o modificar el orden de ellas para adaptarlas a las necesidades de una situación particular. Se necesita un buen conocimiento de las técnicas

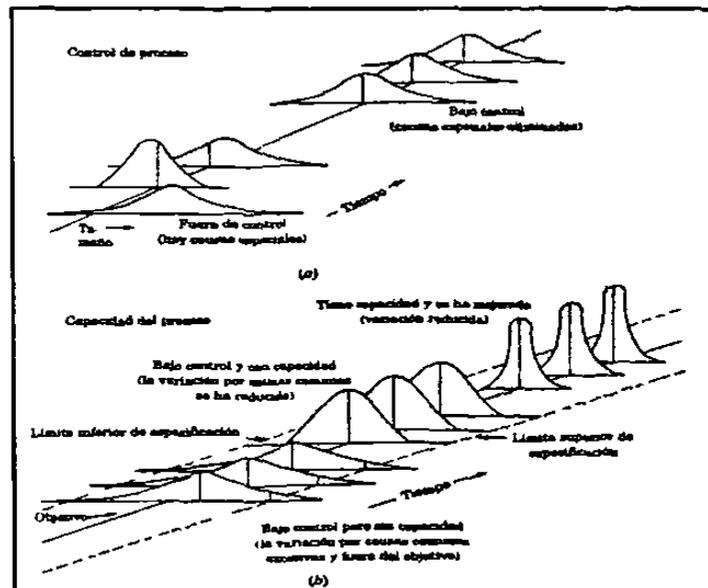


Figura 25 Ilustración del efecto de la eliminación : a) causas especiales, b) causas comunes de variación.

que se emplearán y del funcionamiento del proceso de manufactura. Cada estudio se debe planear con anticipación. Se deben tomar decisiones con respecto a los procedimientos para muestreo, los horarios, y el acopio de datos y estos se deben registrar en tal forma que no haya confusión de cómo, dónde o cuándo se tomaron.

1. Establecer control sobre el proceso. Las pruebas estadísticas descritas se inician con la suposición de que todas las muestras se extrajeron de un solo universo. Hasta que las gráficas de control indiquen que esa suposición es cierta, las estimaciones o cálculos de los parámetros de la distribución en que se basa el proceso, carecen de significado. Se aplican todos los tropiezos relacionados con la formación racional de subgrupos y, en especial, a los estudios de capacidad.

2. Analizar los datos del proceso. Se efectúan cálculos del promedio y dispersión del proceso y se pueden trazar histogramas de frecuencia para tener una idea aproximada de la forma de distribución del producto. Este paso se puede efectuar aunque haya o no indicación de control. Aunque no haya indicación de control, un histograma de frecuencia da una idea de la extensión o dispersión del producto. Cuando se combinan con la información de las gráficas de control, se apreciarán los aspectos principales para reducir la dispersión.

3. Analizar las fuentes de variación. El estudio de las fuentes componentes de la variación y su magnitud puede fluctuar entre pruebas de relativa sencillez hasta diseños experimentales sumamente complicados que requieren un largo tiempo. Este es el punto en que es más importante conocer las características del proceso de manufactura con el procedimiento general ilustrado en la figura 26, se deberá determinar los factores y procedimientos que es más probable se reflejen en la distribución del proceso. Después se planea el acopio de datos, de forma que primero, se aislen estas posibles fuentes y se evalúen. Dado que los resultados de cualquier estudio de capacidad pueden aconsejar mayor investigación en otros aspectos, es importante reconocer con cuidado las horas y la fuente de todas las mediciones. Esto permitirá volver atrás y combinar otra vez los datos con base en la penetración lograda con ese estudio.

9.2.5. Índices de capacidad del proceso

Hasta ahora se ha recalcado la importancia de comparar 6σ de un proceso bajo control estadístico y la de las llamadas tolerancias naturales del proceso, con el intervalo superior (U) - inferior (L) de la especificación. Cuanto

menor sea 6σ en relación de superior - inferior mayor será la capacidad del proceso, siempre y cuando esté bien centrado en la dimensión nominal (objetivo). Cuando se trabaja de acuerdo con el criterio de W. Edwards Deming de un mejoramiento de la calidad que nunca termina, un objetivo de la manufactura sería buscar métodos para reducir en forma continua 6σ , aunque quede bien dentro de los límites de superior - inferior.

Los japoneses hicieron más formal este tipo de comparación y la incluyeron en una serie de índices de Capacidad del Proceso que cada vez que se emplean más en todo el mundo. El primero de ellos, llamado C_p tiene una relación más estrecha con el potencial de capacidad del proceso que se ha comentado hasta ahora. Se puede contar con:

$$C_p = \frac{U - L}{6\hat{\sigma}} \quad (99)$$

Con cualquier valor mayor de 1, el proceso tiene posibilidad de cumplir con las especificaciones si se mantiene bajo control con un \bar{X}_0 de $(U - L)/2$. Se ha sugerido un objetivo mínimo de 1.33 como valor de C_p .

La relación entre μ , estimación del centramiento del proceso y el valor objetivo \bar{X}_0 se designa k y se encuentra con:

$$k = \frac{|\bar{X}_0 - \hat{\mu}|}{\min(\bar{X}_0 - L, U - \bar{X}_0)} \quad (100)$$

en donde las líneas verticales (|) antes y después de $\bar{X}_0 - \hat{\mu}$ significan el "valor absoluto", es decir, no tomar en cuenta el signo del valor y $\min(a,b)$ significa seleccionar el valor mínimo de a o b . El índice k siempre será mayor o igual que cero y tiene un valor objetivo de cero.

Donde las tolerancias de la especificación están simétricas en torno del valor nominal, la fórmula para k se reduce a:

$$k = \frac{|\bar{X}_0 - \hat{\mu}|}{(U - L) / 2} \quad (101)$$

Esta es la forma en que se ven los índices con mayor frecuencia, pero no es válida para las tolerancias unilaterales de especificaciones o cuando hay un solo límite de especificación. También se debería tomar en cuenta que cuando se utilizan tolerancias unilaterales para las especificaciones, el valor objetivo de \bar{X}_0 no será igual a la dimensión nominal, salvo que no haya dispersión posible más allá de esa dispersión.

Hay los tres índices adicionales, que se relacionan más con el comportamiento actual del proceso que con potencial de capacidad y son:

$$CpL = \frac{\hat{\mu} - L}{3\hat{\sigma}} \quad (102)$$

$$CpU = \frac{U - \hat{\mu}}{3\hat{\sigma}} \quad (103)$$

$$Cpk = \min (CpL, CpU) \quad (104)$$

Al igual que con el factor k en cada uno de estos índices se tienen en cuenta la estimación más reciente del centramiento del proceso. CpL y CpU relacionan la diferencia entre la estimación actual de la media del proceso y los límites de especificación, L y U a la mitad de la dispersión natural del proceso. El valor objetivo mínimo de ellos es de uno. Cpk suministra el mínimo de estos dos valores. En las empresas que has adoptado los índices de capacidad del proceso en sus fábricas, suelen especificar su empleo en los contratos de compra - venta con sus proveedores, en especial Cpk.

Se debe tener en cuenta que en cada índice se pretende implicar que la distribución de unidades de producto de un proceso controlado se puede representar con una distribución normal de referencia, es decir, la casi totalidad del producto quedará entre la media μ y $\pm 3\sigma$. (Un examen al apéndice 1 indicará que toda la curva normal, excepto 0.27% queda entre $\pm 3\sigma$). Se necesita un histograma de un número considerable de mediciones de un proceso controlado para determinar si la casi totalidad del producto queda o no dentro de los límites de $\pm 3\sigma$. Estos límites simétricos no encajarán en procesos que puedan producir distribuciones sesgadas. Además, su empleo y la interpretación del valor objetivo se deben modificar en situaciones en donde se especifica un solo límite de especificación, sea U o L

En la etapa de validación o arranque de equipo se debe usar el cálculo de Pp (Similar a Cp), considerando la desviación estándar de todas las mediciones incluyendo puntos dentro y fuera de tolerancia. Y cuando demuestra estabilidad, iniciar el análisis en Ppk (Similar a Cpk) considerando la desviación estándar de todas las mediciones.