FACULTAD DE INGENIERIA U.N.A.M.
DIVISION DE EDUCACION CONTINUA

DIPLOMADO DE CALIDAD EN LA CONTRUCION

MODULO I: METODOS ESTADISTICOS PARA CONTROL DE CALIDAD

TEMA: INTRODUCCION AL CURSO

EXPOSITOR: DR. OCTAVIO RASCON CHAVEZ

DIPLOMADO DE CALIDAD EN LA CONSTRUCCIÓN

MÓDULO I: MÉTODOS ESTADÍSTICOS PARA CONTROL DE CALIDAD

1.1 INTRODUCCIÓN AL CURSO

Prof: Dr. Octavio A. Rascón Chávez

1. *PRESENTACIÓN*

La necesaria modernización del país requiere inversiones cada vez mayores en infraestructuras, pero no pueden ir más allá de lo que las condiciones económicas puedan permitir. Ello exige **aumentar la eficacia y la eficiencia** de los recursos disponibles.

La consecución de estos requerimientos, frente a unos usuarios cada vez más exigentes, con un adecuado respeto a los medios físico y ambiental, y con una adecuada credibilidad, obliga a **mejorar las formas de control de calidad de las obras**.

La **calidad**, desde el punto de vista del ciudadano, es esencialmente **calidad de uso**. Es decir el conjunto de propiedades y características que dan a una obra la capacidad de cubrir de modo satisfactorio, tanto las necesidades implícitas como las explícitas involucradas en la utilización del bien.

Para las administraciones públicas y las empresas existe un factor crítico de la calidad en el proceso de concepción y construcción de sus obras, y es que las mismas no cumplan o no se adapten a la demanda y necesidades de los usuarios.

Toda obra es el resultado de un **proceso** de planeación, diseño, proyecto, construcción y conservación posterior, que requiere por sí mismo una **calidad global**, misma que no se puede obtener a un costo adecuado si en el proceso se trasladan responsabilidades propias de cada etapa a la siguiente, porque esto da lugar a modificaciones, actuaciones complementarias y elevación de las necesidades de conservación. Todo ello no sólo produce una mala imagen de la capacidad técnica de los diferentes agentes (Administración, Ingenierías Consultoras y Empresas), sino unos costos adicionales de la "no calidad", muchas veces difícilmente explicables e incontrolables.

En una obra la calidad requerida se establece en el marco contractual de unas **especificaciones**.

La obtención de la calidad se basa en la **convicción** de que la misma es rentable y de que no cuesta cara. La aparente carestía, apreciada por algunos, se fundamenta en no analizar adecuadamente el costo de los problemas ulteriores.

La calidad se consigue a través de una **adecuada realización del proceso** que debe estar bien planificado, programado y ejecutado, de modo que se imposibiliten los errores, en su caso se prevengan y se evite lo más posible el tener que corregirlos a partir de los resultados de los controles.

De las ideas anteriores, se derivan los dos grandes sistemas de control, complementarios y no contradictorios, los **controles de producción** y los **controles de aceptación**, necesarios los primeros para alcanzar las calidades requeridas y los segundos para evitar que se sacrifique la calidad a los costos.

La calidad es el resultado de la aplicación de un **sistema de gestión de la calidad**, que afecta a **todas las etapas del proceso** de inversión (planificación, programación, proyecto, licitación y obra), por lo que los comportamientos de todos los que intervienen en el proceso son esenciales. Algunos tanto, incluso, como los involucrados en el proceso de construcción.

La calidad se fundamenta en la competencia profesional de todos los que participan, por lo que la atención continua a lo largo del proceso es esencial.

Entre los beneficios que la calidad produce, se pueden señalar que el constructor tendrá ahorros porque:

- Si aplican sistemáticamente las medidas necesarias para obtener su calidad, ahorrará tiempo, materiales y mano de obra.

- Si da una buena calidad, se ahorra discusiones con la Administración o el cliente.

- Si consigue calidad, mejorará su capacidad para que sean aceptables sus ofertas en contratos posteriores.

La Administración tendrá ventajas porque:

- Los usuarios se sentirán más satisfechos.
- La mejora de los procesos permitirá unos precios más ajustados en posteriores ofertas.
- Una mayor fiabilidad del constructor permitirá a la Administración aligerar el contenido de los controles exteriores.

Un sistema de aseguramiento de la calidad requiere:

- Escribir lo que se va a hacer.
- Hacer lo que se ha escrito.
- Escribir lo que se ha hecho.
- Archivar lo escrito.

Exige también tener **conciencia** de que la calidad es necesaria, conciencia que se debe **poseer y transmitir, desde los más altos niveles** de las organizaciones implicadas a todos los participantes.

2. *FILOSOFIA*

Las dos **premisas** o **pilares** en que se fundamenta la filosofía y esencia del **Aseguramiento Total de la Calidad** son:

1) Cualquier operación o actividad de trabajo debe verse como un PROCESO.

2) La persona más importante relacionada con un proceso es el CLIENTE.

De esta forma el Aseguramiento Total de la Calidad puede definirse como:

El estilo de trabajo basado en una metodología operativa, y totalmente comprometido con el continuo mejoramiento en la calidad de los productos y servicios para <u>maximizar</u> <u>la</u> <u>satisfacción</u> de los clientes.

3. *ELEMENTOS Y CARACTERISTICAS.*

Los elementos operativos fundamentales del Aseguramiento Total de la Calidad son:

a) ENFOQUE en el continuo mejoramiento de los procesos.

- Cualquier actividad es un proceso.
- Empleo de datos y métodos científicos de análisis.
- Su meta es alcanzar la perfección.

b) REQUIERE de participación universal.

- Todas las personas pueden y deben practicarlo indepedientemente de su posición y funciones.
- Debe aplicarse en todas partes en una organización.
- Necesita, y a la vez propicia, un trabajo en equipo efectivo.

c) PRODUCE la satisfacción de los clientes.

- Excediendo sus necesidades y expectativas.
- Eliminando las preocupaciones de clientes externos e internos.

Las CARACTERISTICAS principales del Aseguramiento Total de la Calidad que complementan su definición, pueden sintetizarse de la siguiente manera:

1) Representa una alternativa para generar nuevas ideas y utilizar enfoques diferentes, que rompan con la peligrosa costumbre de hacer las cosas siempre de la misma forma.

2) Ofrece una metodología estructurada para identificar y resolver problemas en lugar de vivir "apagando fuegos".

3) Para resultar convincente y exitoso en una organización, requiere antes que nada el compromiso y evidencia de ser aprendido y utilizado **por los directivos de más alto nivel**, y continuar su expansión hacia abajo hasta llegar al último de los empleados.

4) Utiliza conceptos y técnicas de **control estadístico** como soporte a la toma de decisiones encaminada al mejoramiento de los procesos.

5) Es una solución permanente que en forma paulatina se convierte en un estilo de vida.

Los BENEFICIOS internos y externos que ofrece la aplicación del Aseguramiento Total de la Calidad en un sistema productivo, se ilustran en la siguiente figura:

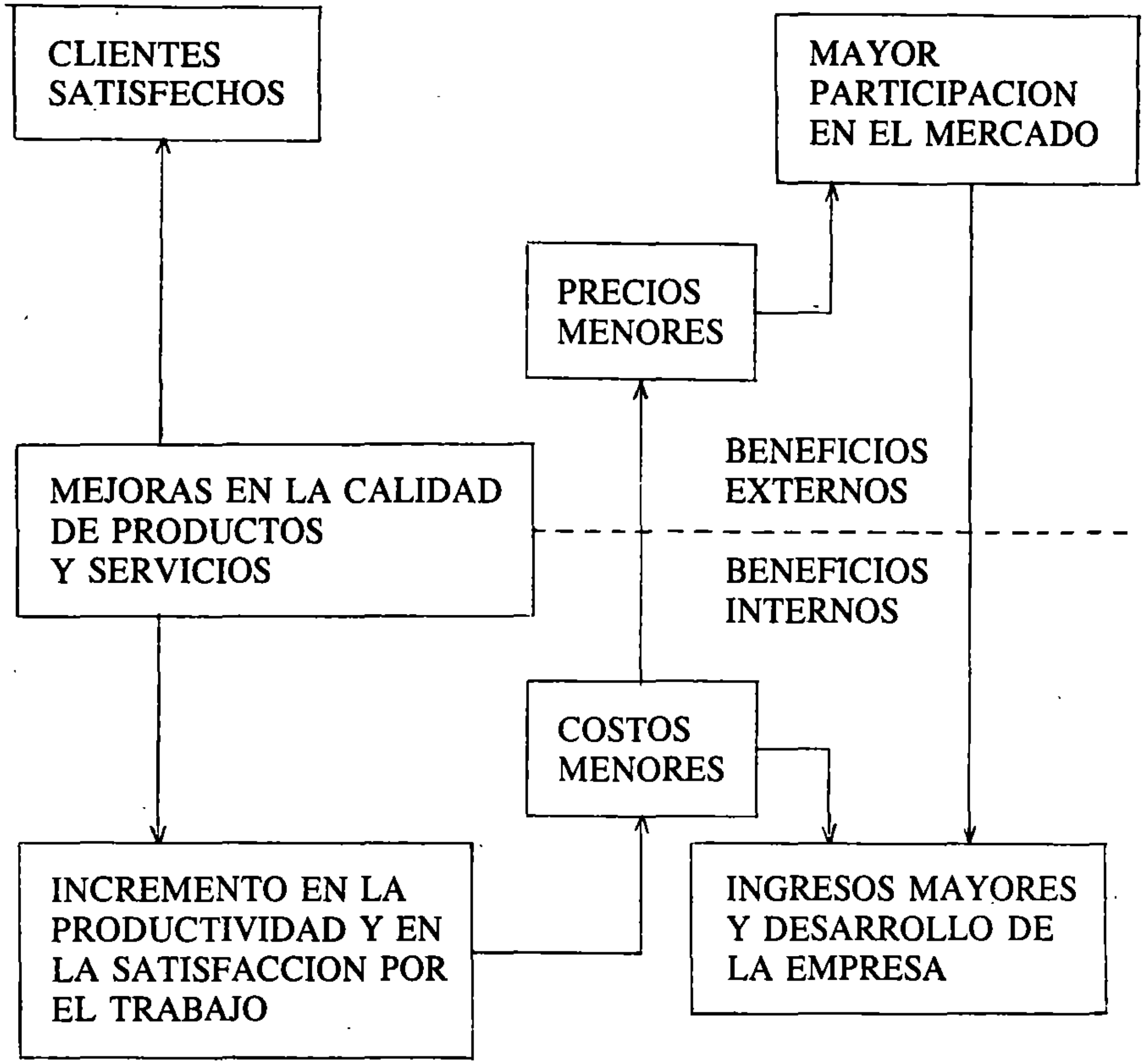


FIGURA 1. BENEFICIOS INTERNOS Y EXTERNOS DEL ASEGURAMIENTO TOTAL DE LA CALIDAD.

4. *METODOLOGIA PARA EL MEJORAMIENTO CONTINUO DE LOS PROCESOS*

El Aseguramiento Total de la Calidad asume cualquier función operativa o actividad, y es un proceso con un propósito determinado cuya **misión primordial es satisfacer los requerimientos de sus clientes**. En este contexto, el papel que juega el Aseguramiento Total de la Calidad, consiste en "asegurar" el **continuo mejoramiento** en la calidad de los procesos y sus resultados, para garantizar el cumplimiento de la misión.

El mejoramiento continuo en los procesos, solamente se puede lograr con un **mecanismo de monitoreo y retroalimentación**, también continuo y permanente.

Esquemáticamente, el sistema puede conceptualizarse de la forma que se indica en la figura 2:

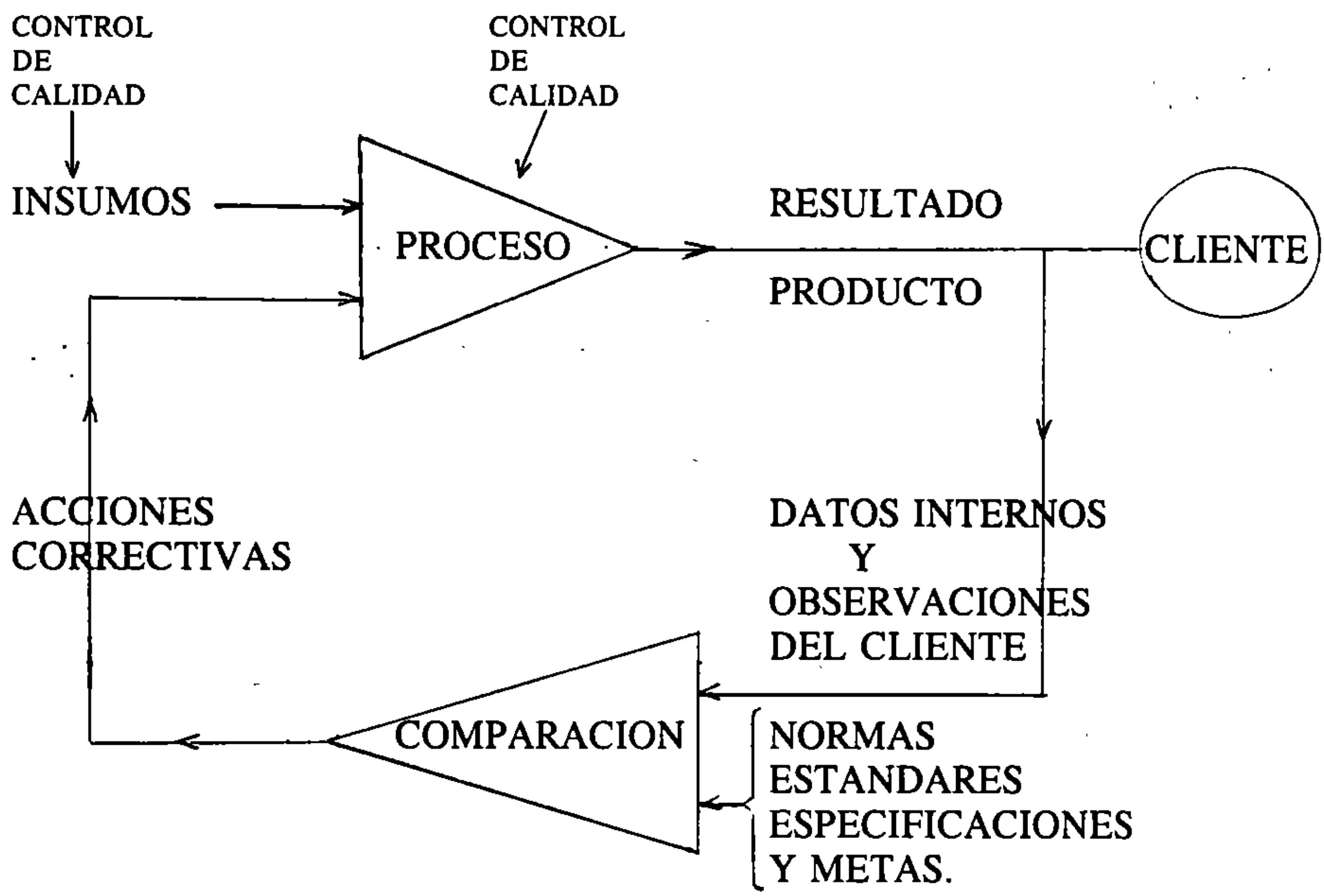


FIGURA 2. SISTEMA DE MONITOREO Y RETROALIMENTACION EN EL ASEGURAMIENTO TOTAL DE LA CALIDAD.

5. *INTRODUCCIÓN A LOS MÉTODOS ESTADÍSTICOS*

* LA CALIDAD DE UN PRODUCTO O MATERIAL SE CARACTERIZA POR EL COMPORTAMIENTO DE UNA O MÁS CARACTERÍSTICAS (ATRIBUTOS O VARIABLES) DEL MISMO

* LOS ATRIBUTOS Y LAS VARIABLES SE REPRESENTAN MEDIANTE ALGUNA **MEDIDA**

* PARA **MEDIR** SE REQUIEREN MÉTODOS, TÉCNICAS Y APARATOS ADECUADOS, CON EL FIN DE OBTENER DATOS FIDEDIGNOS

* LAS MEDICIONES DEBEN PODER REPETIRSE PARA OBTENER EL MISMO TIPO DE INFORMACIÓN EN CADA UNIDAD DEL PRODUCTO O MATERIAL, Y OBTENER ASÍ **COLECCIONES DE DATOS O MUESTRAS**

* LOS **MÉTODOS ESTADÍSTICOS** DE CONTROL DE CALIDAD DAN LOS PROCEDIMIENTOS PARA OBTENER LAS MUESTRAS, PROCESARLAS Y PRESENTAR E INTERPRETAR LOS RESULTADOS, CON EL PROPOSITO DE LOCALIZAR LAS CAUSAS QUE PROVOCAN LA ELABORACIÓN DE PRODUCTOS DEFECTUOSOS

* EN GENERAL, HAY DOS TIPOS DE PROBLEMAS QUE SE ESTUDIAN CON LOS MÉTODOS ESTADÍSTICOS:

    1.- LOS QUE SE PRESENTAN **DURANTE** LA PRODUCCIÓN.
    ESTOS SE ANALIZAN MEDIANTE LAS **CARTAS DE CONTROL.**

    2.- LOS CORRESPONDIENTES A MATERIAS PRIMAS Y A PRODUCTOS YA ELABORADOS.
    ESTOS SE ANALIZAN MEDIANTE **INSPECCIÓN POR MUESTREO.**

* PARA ESTUDIAR AMBOS TIPOS DE PROBLEMA, SE HACEN MEDICIONES DE LA O LAS CARACTERÍSTICAS QUE SE DESEAN CONTROLAR, LAS CUALES SE REALIZAN DE ACUERDO CON UN PROGRAMA PREDEFINIDO.

* LOS **DATOS** QUE SE OBTIENEN AL REALIZAR LAS MEDICIONES, **VARÍAN** DE UNA A OTRA VERIFICACIÓN.

  LAS VARIACIONES PUEDEN DEBERSE AL INSTRUMENTO DE MEDICIÓN, AL PROCEDIMIENTO DE PRUEBA, A DIFERENCIAS " RAZONABLES " EN LAS PROPIEDADES DE LAS MATERIAS PRIMAS, A DESGASTES GRADUALES DE LA MAQUINARIA, ETC.

* LOS MÉTODOS ESTADÍSTICOS CONSTITUYEN LAS "HERRAMIENTAS" PARA IDENTIFICAR OPORTUNAMENTE LAS **VARIACIONES** EN LOS DATOS, QUE REFLEJAN QUE EL PROCESO DE PRODUCCIÓN ESTÁ SALIÉNDOSE DE CONTROL, O SI LOS PRODUCTOS O MATERIALES INSPECCIONADOS CUMPLEN CON LOS REQUISITOS Y ESPECIFICACIONES ESTABLECIDAS.

* SE DICE QUE SE REALIZA UNA INSPECCIÓN POR **VARIABLES** CUANDO LAS MEDICIONES DE LAS CARACTERÍSTICAS BAJO CONTROL SE REGISTRAN EN TÉRMINOS **CUANTITATIVOS O NUMÉRICOS**.
  TAL ES EL CASO DE MEDICIÓN DE RESISTENCIAS, DIMENSIONES, VOLÚMENES, DEFORMACIONES, PROPIEDADES MECÁNICAS, ETC.

* SE DICE QUE SE REALIZA UNA INSPECCIÓN POR **ATRIBUTOS** CUANDO LA INFORMACIÓN QUE SE OBTIENE O REGISTRA SE SEÑALA EN FORMA **CUALITATIVA**.
  POR EJEMPLO, SI LOS DATOS SE REPORTAN COMO **BUENO** O **DEFECTUOSO**; O SI **PASA** O **NO PASA**, ETC.

5.1 *HERRAMIENTAS PARA EL CONTROL DE CALIDAD ESTADÍSTICO*

1. **CARTAS DE CONTROL**, PARA **CARACTERÍSTICAS DE CALIDAD MESURABLES**. SE TRABAJA CON LOS PROMEDIOS ARITMÉTICOS, LAS DESVIACIONES ESTÁNDAR Y LOS RANGOS DE LAS MUESTRAS QUE SE OBTIENEN PARA MONITOREAR LA CALIDAD DEL PROCESO. (GRÁFICA DE $\bar{X}$, R y σ).

2. **CARTAS DE CONTROL** PARA LA **FRACCIÓN O PORCENTAJE DE ELEMENTOS DEFECTUOSOS**. (GRÁFICA P).

3. **CARTAS DE CONTROL** PARA EL **NÚMERO DE DEFECTOS** POR UNIDAD. (GRÁFICAS C).

4. **MUESTREO DE ACEPTACIÓN**, PARA EVALUAR ESTADÍSTICAMENTE LA CALIDAD DE LAS MATERIAS PRIMAS Y LOS PRODUCTOS TERMINADOS O EN ALGUNA ETAPA DEL PROCESO DE PRODUCCIÓN.

5.2 *CARTAS DE CONTROL*

LA CALIDAD DE UN PRODUCTO MANUFACTURADO ESTÁ SIEMPRE SUJETA A UNA **CIERTA VARIACIÓN**, COMO RESULTADO DEL **AZAR**.

SIEMPRE EXISTE **UN PATRÓN** DE CAUSAS CASUALES ESTABLE, QUE ES INHERENTE A CUALQUIER ESQUEMA DE PRODUCCIÓN Y DE INSPECCIÓN.

LA VARIACIÓN DENTRO DE ESTE PATRÓN ESTABLE ES INEVITABLE Y MEDIBLE EN TÉRMINOS ESTADÍSTICOS. TAMBIÉN EXISTEN CAUSAS **ADICIONALES** DE VARIACIÓN **QUE SON EXTERNAS A ESTE PATRÓN**, PERO QUE PUEDEN SER DETECTADAS Y CORREGIDAS.

LAS **CARTAS DE CONTROL** SON HERRAMIENTAS ESTADÍSTICAS QUE PERMITEN DETERMINAR CUÁNDO SE PRESENTAN LAS VARIACIONES EXTERNAS. POR TANTO, HACEN POSIBLE EL **DIAGNÓSTICO** Y **CORRECCIÓN** DE MUCHOS PROBLEMAS DE PRODUCCIÓN, POR LO QUE DAN PAUTAS PARA REALIZAR ACCIONES EN EL PROCESO DE PRODUCCIÓN, QUE LLEVEN A MEJORAS CONSIDERABLES EN LA CALIDAD DEL PRODUCTO Y A LA REDUCCIÓN DE DESPERDICIOS Y REPROCESADO.

AL DETERMINAR ESTADÍSTICAMENTE CUÁLES SON LAS BANDAS DE VARIACIÓN DE LA CALIDAD, OCASIONADA POR ELEMENTOS INEVITABLES, ASOCIADAS A **CIERTAS PROBABILIDADES** DE QUE LOS RESULTADOS DE LAS MEDICIONES SE MANTENGAN DENTRO DE ELLAS, LA CARTA DE CONTROL INDICA CUÁNDO EL PROCESO ESTÁ **BAJO CONTROL** Y, DE ESTA FORMA, EVITA AJUSTES INNECESARIOS AL MISMO, O CUANDO SE HA SALIDO DE CONTROL Y ES NECESARIO TOMAR ACCIONES CORRECTIVAS.

5.3 *CARTAS DE CONTROL: BENEFICIOS*

1. **LA VIABILIDAD BÁSICA DE LA CARACTERÍSTICA DE CALIDAD.**

   CUANDO SE HAN ESPECIFICADO TANTO UN **VALOR SUPERIOR** COMO UNO **INFERIOR** TOLERABLES PARA UNA CARACTERÍSTICA DE LA CALIDAD, UN PROBLEMA TÉCNICO IMPORTANTE QUE SE PRESENTA, CONSISTE EN DETERMINAR SI LA VARIABILIDAD BÁSICA (CASUAL) DEL PROCESO DE PRODUCCIÓN, ES TAN GRANDE QUE SEA MUY DIFÍCIL FABRICAR "TODO" EL PRODUCTO DENTRO DE LOS LÍMITES ESPECIFICADOS.

2. **EL NIVEL GENERAL DE LAS CARACTERÍSTICAS DE CALIDAD.**

   AUN CUANDO LA VARIABILIDAD BÁSICA DE UN PROCESO, SEA TAL QUE LA GAMA NATURAL DE VARIACIÓN SEA MÁS ESTRECHA QUE LA GAMA DE TOLERANCIA ESPECIFICADA, Y QUE EL PROCESO SE ENCUENTRE BAJO CONTROL, EL PRODUCTO **PUEDE SER NO SATISFACTORIO** PARA UNA CLIENTELA EN PARTICULAR, DADO QUE EL NIVEL DE CALIDAD ESPECIFICADO ES DEMASIADO BAJO.

3. **LA CONSISTENCIA DEL RENDIMIENTO.**

   LA VARIABILIDAD DE LA CARACTERÍSTICA DE CALIDAD PUEDE SEGUIR UN PATRÓN **CASUAL** O PUEDE COMPORTARSE ERRÁTICAMENTE DEBIDO A LAS **CAUSAS ASIGNABLES**. AL DETECTARSE CUÁL SITUACIÓN PREVALECE, SE DA LA PAUTA PARA DECIDIR SOBRE DEJAR AL PROCESO COMO ESTÁ O TOMAR ACCIONES PARA CORREGIR LOS MOTIVOS DE LAS DIFICULTADES.

5.4 *MUESTREO DE ACEPTACION*

LA **INSPECCIÓN DE ACEPTACIÓN** ES UNA PARTE NECESARIA DE LA MANUFACTURA, Y PUEDE SER APLICADA A LOS MATERIALES QUE SE RECIBEN, A LOS PRODUCTOS PARCIALMENTE ACABADOS EN DIFERENTES ETAPAS INTERMEDIAS DEL PROCESO DE MANUFACTURA, Y AL PRODUCTO FINAL. LA INSPECCIÓN DE ACEPTACIÓN PUEDE LLEVARSE A CABO EXTERIORMENTE POR EL COMPRADOR.

MUCHA DE ESTA INSPECCIÓN DE ACEPTACIÓN SE LLEVA A CABO MEDIANTE **MUESTREO**, YA QUE A MENUDO LA INSPECCIÓN DEL TOTAL RESULTA IMPRACTICABLE O CLARAMENTE ANTIECONÓMICA.

DEBE RECONOCERSE QUE MIENTRAS UNA PARTE DEL PRODUCTO SEA DEFECTUOSA, ES POSIBLE QUE ALGUNOS ELEMENTOS SEAN PASADOS POR ALTO, CUALQUIERA QUE SEA EL ESQUEMA DEL MUESTREO DE ACEPTACIÓN.

EL MUESTREO DE ACEPTACIÓN PERMITE VALUAR EL RIESGO ASUMIDO CON PROCEDIMIENTOS DE MUESTREO ALTERNOS, Y TOMAR UNA DECISIÓN ACERCA DEL **GRADO DE PROTECCIÓN** NECESARIO EN CUALQUIER CASO, EL CUAL SE DETERMINA MEDIANTE PROCEDIMIENTOS ESTADISTICOS.

ES POSIBLE, ENTONCES, SELECCIONAR UN MODELO DE MUESTREO DE ACEPTACIÓN, QUE PROPORCIONE UN **GRADO DESEADO DE PROTECCIÓN**, CON LA DEBIDA CONSIDERACIÓN A LOS DIFERENTES COSTOS INVOLUCRADOS.

5.5 *BENEFICIOS DEL CONTROL ESTADÍSTICO DE CALIDAD*

- AHORROS EN TIEMPO, EQUIPO, MATERIALES Y DINERO

- REDUCCIÓN EN LOS COSTOS Y TIEMPOS DE INSPECCIÓN Y PRUEBA

- MEJORÍA EN EL USO DE LOS RECURSOS

- MEJORÍA EN LA PRODUCTIVIDAD

- MEJORÍA EN LAS RELACIONES HUMANAS INTERNAS Y CON LOS PROVEEDORES Y CLIENTES

- MEJOR PRESTIGIO Y COMPETITIVIDAD

FACULTAD DE INGENIERIA U.N.A.M.
DIVISION DE EDUCACION CONTINUA

DIPLOMADO DE CALIDAD EN LA CONTRUCION

MODULO I: METODOS ESTADISTICOS PARA CONTROL DE CALIDAD

TEMA: CARTAS DE CONTROL

EXPOSITOR: M. EN I. AUGUSTO VILLARREAL A.

Palacio de Minería Calle de Tacuba 5 Primer piso Deleg. Cuauhtemoc 06000 Mexico, D.F. APDO Postal M-[illegible]
Telefonos 512-8955 512-5121 521-7335 521-1987 Fax 510-0573 512-5121 521-4020 AL [illegible]

CARTAS DE CONTROL

Por: M en I Augusto Villarreal A.

INTRODUCCION

Aunque existe la tendencia generalizada a pensar que el Control de Calidad es de desarrollo reciente, realmente no existe nada nuevo en la idea básica de elaborar un producto caracterizado por un alto grado de uniformidad.

Durante siglos, hábiles artesanos han procurado elaborar productos que se distingan por su superior calidad, y una vez que han logrado obtener un cierto estándar de calidad óptimo, eliminar dentro de lo posible la variación entre productos que nominalmente deben resultar iguales.

La idea de que la Estadística puede resultar un instrumento muy útil para asegurar un estándar adecuado de calidad para los productos manufacturados, se remonta no más allá del advenimiento de la producción masiva, y el uso extendido de los métodos estadísticos para resolver problemas de control de calidad es aún más reciente.

Muchos problemas que aparecen durante la elaboración de un producto son susceptibles de ser resueltos empleando tratamientos estadísticos, por lo que al hablar de control estadístico de calidad, nos estaremos refiriendo esencialmente a las dos técnicas especiales que se discutirán en esta parte del curso: uso de las Cartas de Control y muestreo de aceptación.

Conviene mencionar que la palabra calidad, al ser empleada de aquí en adelante, se referirá a alguna propiedad medible o contable de algún producto, tal como el diámetro de un balín de acero, la resistencia de una viga de concreto, el número de defectos en una pieza de tela, la eficacia de cierta droga, etc.

IDEAS SOBRE CARTAS DE CONTROL

A muchos individuos les puede sorprender el hecho de que dos artículos aparentemente idénticos, elaborados bajo condiciones cuidadosamente controladas, de las mismas materias primas, y por una misma máquina con diferencia de pocos segundos, puedan, sin embargo, diferir en muchos aspectos.

En efecto, cualquier proceso de manufactura, aun siendo muy bueno, se encuentra caracterizado por una cierta cantidad de variación que es de naturaleza aleatoria, y que no puede ser eliminada en forma completa.

Cuando la variabilidad presente en un proceso de producción se limita a variación aleatoria se dice que el proceso se encuentra en un estado de control estadístico.

Tal estado se puede alcanzar cuando se eliminan aquellos problemas causados por otro tipo de variación, llamada variación sistemática, que es de naturaleza más bien determinística, y que se puede achacar, por ejemplo, a operadores mal entrenados, materia prima de baja calidad, máquinas en mal estado, etc.

Ya que los procesos de manufactura se encuentran rara vez libres de estos problemas, conviene contar con algún método sistemático para detectar desviaciones serias de un estado de control estadístico cuando ocurren, o inclusive antes de que ocurran, tales desviaciones.

Ese método sistemático de detección se puede tener mediante el empleo de las llamadas Cartas de Control.

TIPOS DE CARTAS DE CONTROL

En lo que sigue distinguiremos entre las cartas de control para mediciones o variables ($\bar{X}$, R, σ) y las cartas de control para atributos (p, c), dependiendo de que las observaciones que estemos analizando sean mediciones o datos contados o calculados, respectivamente.

Un ejemplo del primer caso sería la longitud de las varillas de acero de una muestra. Como ejemplo del segundo caso tendríamos el número de focos defectuosos en una muestra de tamaño dado.

CONFIGURACION DE LAS CARTAS DE CONTROL

En cualquiera de los casos mencionados, una carta de control consiste de una Línea Central, correspondiente a la calidad promedio a la que el proceso debe funcionar, y dos líneas que corresponden al Límite Superior de Control (LSC) y al Límite Inferior de Control (LIC), respectivamente, tal como se muestra en la Fig 1.

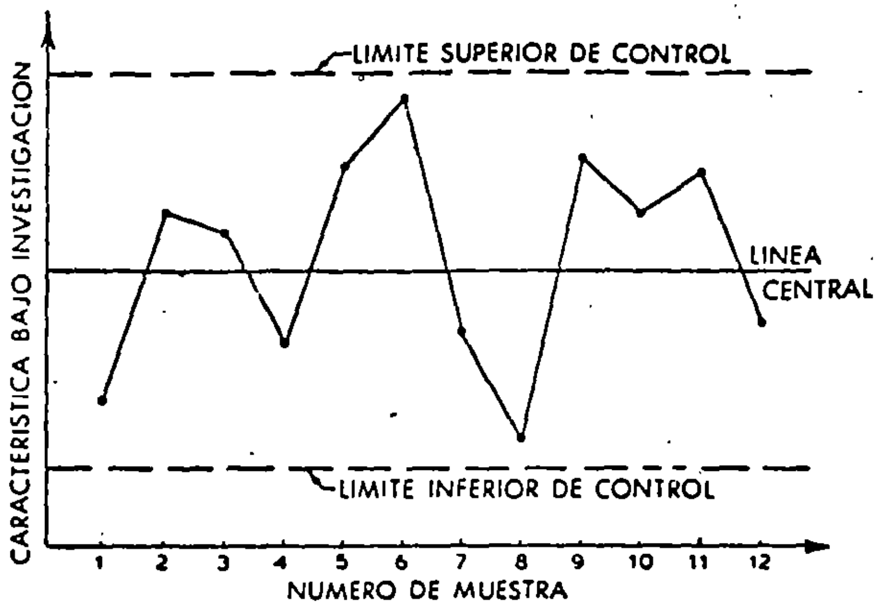


Fig 1. Aspecto general de una carta de control

Estos límites se escogen en forma tal que los valores que se encuentren dentro de ellos se puedan atribuir al azar, en tanto que los valores que caigan fuera de ellos se puedan considerar como indicaciones de falta de control.

No obstante la idea anterior, conviene mencionar que en la Fig 2 que se presenta a continuación se pueden considerar otras posibles situaciones de "falta de control" que ameritan investigarse:

1. Cuando dos de tres puntos sucesivos caen en la zona A.
2. Cuando cuatro de cinco puntos sucesivos caen en la zona B o más allá.
3. Cuando ocho puntos sucesivos caen en la zona C o más allá.

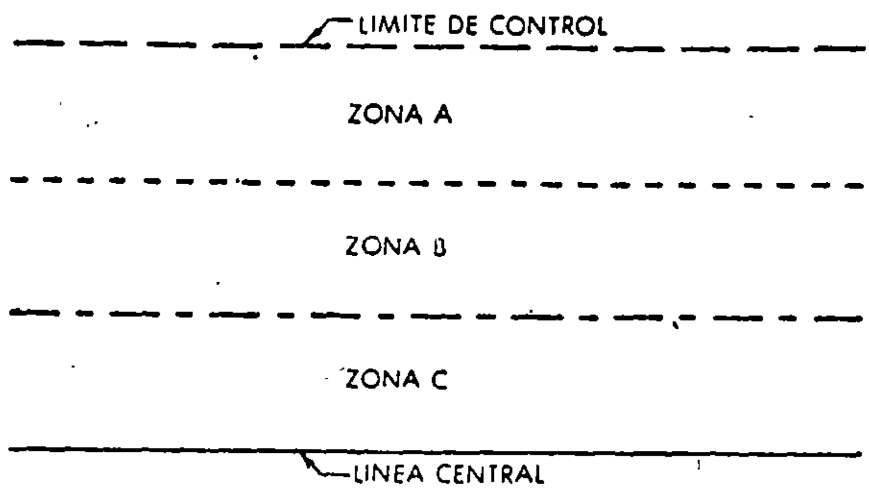


Fig 2 Diagrama que define las zonas A, B y C usadas en el análisis de Cartas de Control.

Debe hacerse notar que cada una de las zonas A, B y C constituye la tercera parte del área entre la línea central y un límite de control, y que las pruebas mencionadas se aplican a ambas mitades de la carta de control, pero se aplican separadamente para cada mitad, y nunca a las dos mitades en combinación.

EXPLICACION DEL EMPLEO DE LAS CARTAS DE CONTROL

Si se grafican en una carta los resultados obtenidos a partir de muestras tomadas periódicamente a intervalos frecuentes, es posible verificar por medio de ella si el proceso se encuentra bajo control, o si se encuentra presente en el proceso la variación sistemática del tipo descrito anteriormente.

Cuando un punto graficado cae fuera de los límites de control, es necesario encontrar el problema que causó tal evento dentro del proceso. Pero aun si los puntos caen dentro de los límites mencionados, alguna tendencia, o cierto patrón de los mismos, puede indicar que se debe llevar a cabo alguna acción para prevenir y así evitar algún problema serio.

La habilidad para "leer" las cartas de control y para determinar a partir de ellas cuál acción correctiva debe llevarse a cabo, - se obtiene a partir de la experiencia y del juicio altamente desarrollado. Un practicante del control estadístico de la calidad debe no sólo comprender los fundamentos estadísticos de la - materia, sino también encontrarse identificado plenamente con los procesos que desea controlar.

CARTAS DE CONTROL PARA MEDICIONES (VARIABLES)

Cuando se requiere establecer control estadístico de la calidad de algún producto en términos de mediciones o variables, es costumbre ejercer tal control sobre la calidad media del proceso, - al igual que sobre su variabilidad.

La primera meta se logra al graficar los promedios de muestras - extraídas periódicamente en la llamada carta de control para los promedios, o simplemente carta $\bar{X}$. La variabilidad se puede controlar de igual forma si se grafican los rangos o las desviaciones estándar de las muestras en las llamadas cartas R o cartas σ, respectivamente, dependiendo de cuál estadística se emplee para estimar la desviación estándar de la población.

Si se conocen la media μ y la desviación estándar σ de la población (proceso) y es razonable suponer las mediciones obtenidas - como muestras extraídas de una población normal, se puede asegurar que con probabilidad $1 - \alpha$ el promedio aritmético de una - - muestra aleatoria de tamaño n se encontrará entre

$$\mu - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{y} \quad \mu + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

ó

$$\mu - z_{\alpha/2}\, \sigma_{\bar{X}} \quad \text{y} \quad \mu + z_{\alpha/2}\, \sigma_{\bar{X}}$$

puesto que $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ para el caso de la distribución muestral del promedio aritmético, cuando se muestrea de una población infinita. La suposición de que la extracción de muestras aleatorias se hace de una población infinita es válida en el caso presente, puesto que, por ejemplo, la producción de cierto producto en una fábrica tiende a infinito conforme pasa el tiempo.

Los dos límites anteriores ($\mu \pm z_{\alpha/2}\sigma_{\bar{X}}$) proporcionan entonces límites inferiores y superiores de control y, bajo las suposiciones - anteriores, permiten al practicante del control de calidad determinar si se debe o no llevar a cabo algún ajuste en el proceso, - al graficar los promedios aritméticos obtenidos de muestras de tamaño n en una carta como la que se muestra en la Fig 1.

Conviene establecer en este momento que al emplear una carta de control para los promedios, lo que se hace realmente es probar hipótesis nulas de que a un cierto nivel de confianza $1-\alpha$ el valor de la media de la distribución muestral de los promedios sea igual al valor de la calidad nominal del proceso, o al de la calidad media calculada para el mismo, μ_o. Para estas pruebas secuenciales de hipótesis, s emplean como estadísticas de prueba los valores de los promedios - aritméticos obtenidos de muestras aleatorias extraídas de la población (o proceso). Es decir, se realizan pruebas de hipótesis para las cuales

$$H_0 \quad : \quad \mu = \mu_o$$

$$H_1 \quad : \quad \mu \neq \mu_o$$

(Prueba de dos colas; cada prueba se - realiza con el valor $\bar{X}_i$ de la muestra i)

en donde μ es la media de la distribución muestral del promedio aritmético, μ_o la calidad nominal o calidad media calculada del proceso, y $\bar{X}_i$ (i=1,2,3,...) el valor del promedio aritmético obtenido de la iésima muestra aleatoria. La forma secuencial de estas pruebas de hipótesis se muestra en la Fig 3 que se presenta a continuación.

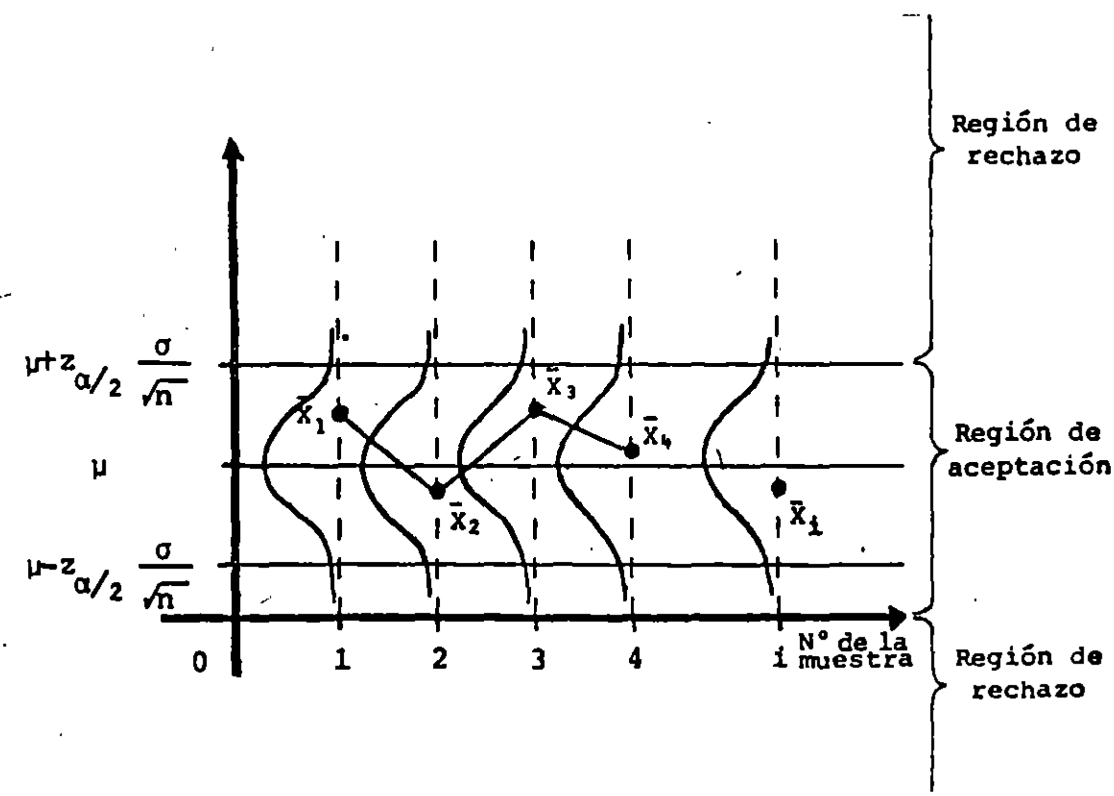


Fig 3 Pruebas de hipótesis que se realizan al emplear una carta de control para los promedios

Si se consideran problemas prácticos, los valores de μ y σ del proceso se desconocen, y es entonces conveniente estimar sus valores a partir de muestras tomadas mientras el proceso se encuentre "bajo control", tal como se explica más adelante. En la práctica es entonces difícil llegar a establecer límites de control del tipo $\mu \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ al desconocerse μ y σ, independientemente de que en muchos casos es demasiado arriesgado considerar a las mediciones como muestras aleatorias extraídas de una población normal.

En lugar de lo anterior, en el control de calidad industrial se emplean comúnmente los límites de control de "tres desviaciones estándar" o de "tres sigmas", que se obtienen al sustituir a $z_{\alpha/2}$ por un 3 al calcular los límites de control.

Conforme a lo anterior, con los límites de control

$$\mu \pm 3\sigma_{\bar{X}} \qquad \text{ó} \qquad \mu \pm 3 \frac{\sigma}{\sqrt{n}}$$

se puede confiar en que en el 99.73% de los casos el proceso no será declarado "fuera de control", cuando de hecho se encuentra "bajo control".

En otras palabras, estos límites de control permiten considerar que la probabilidad máxima de rechazar la hipótesis

$$H_0 \ : \ \theta = \theta_o$$

cuando debería de ser aceptada (probabilidad de cometer un error de tipo I) es de 0.27%, siendo θ_o un valor de calidad fijo del proceso, y θ el del parámetro correspondiente de la distribución muestral de la estadística bajo consideración.

ELABORACION DE LA CARTA DE CONTROL PARA LOS PROMEDIOS ($\bar{X}$)

a. Caso en que se conocen la media μ y la desviación estándar σ de la población.

Línea central ———————— μ

Límites de control ———————— $\mu \pm 3\sigma_{\bar{x}}$ ó $\mu \pm 3\frac{\sigma}{\sqrt{n}}$

ó $\mu \pm A\sigma$, siendo $A = \frac{3}{\sqrt{n}}$.

en donde los valores de A se obtienen de la tabla I, en función de n, el tamaño de la muestra.

<u>Ejemplo</u>: Sea el proceso de elaboración de varillas de acero para las cuales se sabe que el diámetro medio es de 2.5 cm, con una desviación estándar de 0.01 cm. Se desea efectuar control del diámetro de las mismas, para lo cual se extraen periódicamente muestras de cinco varillas. Se pide establecer la línea central y los límites de control para una carta $\bar{X}$.

<u>Solución</u>. Siendo $\mu = 2.5$ cm, $\sigma = 0.01$ y $n = 5$, se tiene que:

Línea central = μ=2.5

Límites de control:

$$2.5 \pm 3\frac{\sigma}{\sqrt{n}} = 2.5 \pm \frac{3(0.01)}{\sqrt{5}} = 2.5 \pm 0.0134 \Longrightarrow 2.5134, \quad 2.4866$$

o, de la tabla I

$$2.5 \pm A\sigma = 2.5 \pm 1.342(0.01) = 2.5 \pm 0.01342 \Longrightarrow 2.51342, \quad 2.48658$$

b. Caso en que se desconocen μ y σ.

Para este caso, que es el más común, es necesario estimar, como se dijo anteriormente, tales parámetros con base en muestras preliminares. Para el caso, normalmente se acostumbra emplear un mínimo de 20 a 25 muestras de 4 ó 5 elementos, obtenidas consecutivamente cuando el proceso está "bajo control".

Sin embargo, como veremos más adelante, se pueden emplear procedimientos estadísticos más formales para determinar el número de muestras (y de elementos en las mismas) más adecuado para las cartas $\bar{X}$.

Entonces, si se utilizan κ muestras preliminares, cada una de tamaño n, se puede estimar con adecuada precisión el valor de μ mediante

$$\bar{\bar{X}} = \frac{1}{\kappa}\sum_{i=1}^{\kappa} \bar{X}_i$$

siendo $\bar{\bar{X}}$ un estimador insesgado y consistente de μ, donde $\bar{X}_i$ denota al promedio aritmético de la iésima muestra, y $\bar{\bar{X}}$ es el promedio de los promedios de las muestras.

El valor de σ de la población puede ser estimado a partir de las desviaciones estándar o de los rangos de las muestras. Si el tamaño de las mismas es pequeño, usualmente el rango proporciona un estimador eficiente de σ, además de que el proceso de cálculo del mismo es bastante más simple que el de la desviación estándar para las muestras.

Sin embargo, es conveniente, cuando se requiere bastante precisión en el cálculo de los límites de control, estimar a σ mediante las desviaciones estándar de las muestras. Tal es el caso, por ejemplo, de muestras de productos que son caros, y que deben destruirse al momento de tomar las mediciones.

b.1 Estimando a σ mediante los rangos de las muestras

Hay que obtener primero el valor $\bar{R}$, que es el rango promedio de los rangos de las κ muestras, es decir,

$$\bar{R} = \frac{1}{\kappa} \sum_{i=1}^{k} R_i$$

Puesto que la estadística $\bar{R}$ siempre estima por encima de su valor real a la desviación estándar de la población, se obtiene un estimador sesgado. Debido a ello, es indispensable afectar el valor de $\bar{R}$ en forma tal de obtener un estimador insesgado de σ, para lo cual se hace

$$\text{Estimador insesgado de } \sigma = \frac{\bar{R}}{d_2}$$

El factor d_2 en la expresión anterior se obtiene experimentalmente al identificar el valor de la media en las distribuciones muestrales del cociente R/σ para distintos valores de n, considerando una población en la cual el valor de σ es conocido. Por ejemplo, para muestras de tamaño cinco (n=5), se ha obtenido experimentalmente el valor d_2=2.326, tal como se muestra en la Fig 4.

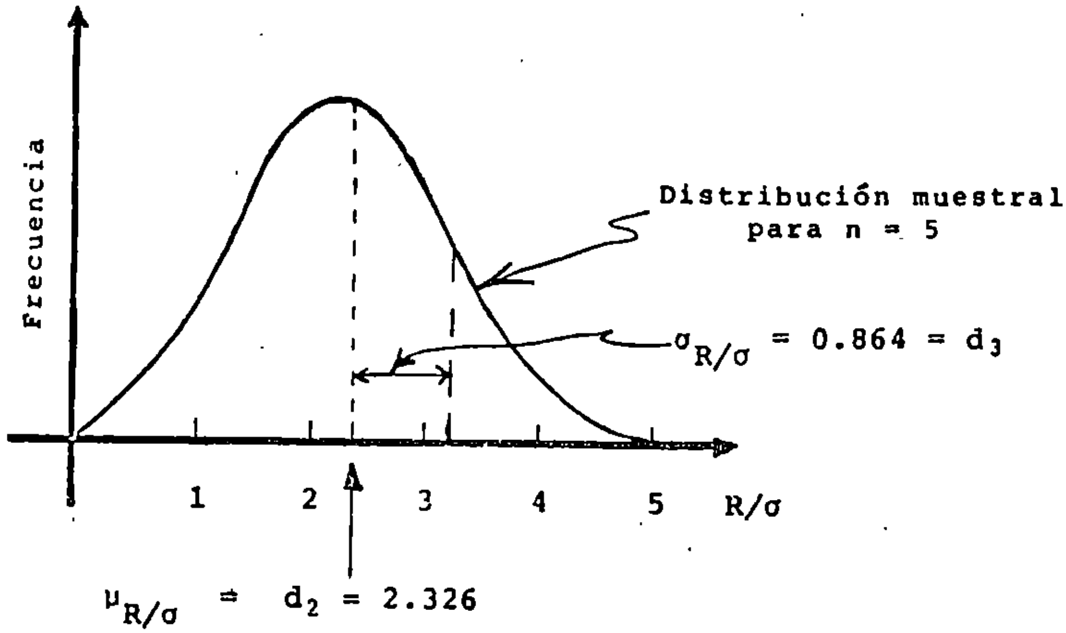


Fig 4. Distribución muestral de R/σ para $n=5$, suponiendo σ conocida.

En la tabla I se presentan los valores del factor d_2 para distintos tamaños de muestra, observándose que conforme se incrementa el valor de n aumenta el de ese factor, lo cual permite concluir que el rango estima mejor a la desviación estándar cuando las muestras son pequeñas.

De acuerdo con lo anterior, se pueden emplear las siguientes expresiones en la elaboración de la carta de control para los promedios:

Línea Central —— $\bar{X}$

Límites de Control —— $\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}}$ ó $\bar{\bar{X}} \pm \frac{3\bar{R}}{d_2\sqrt{n}}$

Para abreviar el cálculo de los límites de control a partir de los rangos de las muestras, se ofrece en la tabla I el factor

$$A_2 = \frac{3}{d_2\sqrt{n}}$$

cuyo empleo permite establecer los límites de control como

$$\bar{\bar{X}} \pm A_2 \bar{R}$$

b.2 Estimando a σ mediante las desviaciones estándar de las muestras

Se debe obtener primero el valor de $\bar{\sigma}$, que es el promedio de las desviaciones estándar de las muestras, es decir

$$\bar{\sigma} = \frac{1}{K} \sum_{i=1}^{k} S_i$$

En donde S_i denota la desviación estándar de la iésima muestra. No siendo tampoco $\bar{\sigma}$ un estimador insesgado de la desviación estándar de la población, ya que siempre la estima por abajo de su valor real, hay que afectar dicho valor por un cierto factor para hacerlo insesgado, es decir

$$\text{Estimador insesgado de } \sigma = \frac{\bar{\sigma}}{c_2}$$

Los valores de c_2 se reportan en la tabla I en función del tamaño de la muestra, y se obtienen mediante un procedimiento similar al explicado para el factor d_2.

Con base en lo anterior, los parámetros de la carta de control para los promedios son los siguientes:

Línea Central —— $\bar{\bar{X}}$

Límites de Control —— $\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}}$ ó $\bar{\bar{X}} \pm \frac{3\bar{\sigma}}{c_2\sqrt{n}}$

De nuevo, para abreviar el cálculo de los límites de control para la carta $\bar{X}$, obtenidos ahora a partir de las desviaciones estándar de las muestras, se puede emplear el factor dado en la tabla I

$$A_1 = \frac{3}{c_2\sqrt{n}}$$

con el cual los límites de control quedan como

$$\bar{\bar{X}} \pm A_1 \bar{\sigma}$$

NUMERO MINIMO DE MUESTRAS REQUERIDO PARA LA ELABORACION DE CARTAS $\bar{X}$

En este momento conviene establecer el número mínimo de muestras preliminares, m, así como el tamaño de las mismas, n, que es necesario considerar para estimar adecuadamente los parámetros de una carta de control para los promedios.

El asegurar que un mínimo de 20 o 25 muestras con 4 o 5 elementos cada una son necesarias para obtener los valores de $\bar{\bar{X}}$, $\bar{R}$ o $\bar{\sigma}$, frecuentemente choca con el argumento de que por razones de costo, tiempo, etc., se debe emplear un número menor de ellas. Por ello, se han preparado tablas como las II y III que se presentan al final, que permiten obtener una solución cuantitativa para este problema.

Cuando se emplea el rango $\bar{R}$ como estimador de σ para la elaboración de una carta $\bar{X}$, y como se verá más adelante, para una carta R, la tabla II permite determinar el número mínimo, m, de muestras de tamaño n que se deben emplear para tener poco más de un 98% de nivel de confianza de que los promedios aritméticos obtenidos de las muestras se encuentren dentro de los límites de control que se calculen para la carta $\bar{X}$, suponiendo únicamente la presencia de variación aleatoria.

De la misma manera, se establecen en la tabla III los valores óptimos de m y n, cuando se emplean las desviaciones estándar de las muestras para obtener el estimador $\bar{\sigma}$ de la desviación estándar de la población.

<u>Ejemplo:</u> Sea una fábrica que produce varillas de acero, en la cual se desea ejercer control sobre el peso de las mismas. Para ello, se seleccionan veinte muestras aleatorias de cinco varillas cada una, obteniéndose los valores que se reportan en la tabla siguiente:

Número de la muestra	Valores individuales del peso, Kg					Promedio Aritmético $\bar{X}$	Rango R	Desviación estándar S_x
	X_1	X_2	X_3	X_4	X_5			
1	11.1	9.4	11.2	10.4	10.1	10.44	1.8	0.6651
2	9.6	10.8	10.1	10.8	11.0	10.46	1.4	0.5276
3	9.7	10.0	10.0	9.8	10.4	9.98	0.7	0.2400
4	10.1	8.4	10.2	9.4	11.0	9.82	2.6	0.8727
5	12.4	10.0	10.7	10.1	11.3	10.90	2.4	0.8832
6	10.1	10.2	10.2	11.2	10.1	10.36	1.1	0.4224
7	11.0	11.5	11.8	11.0	11.3	11.32	0.8	0.3059
8	11.2	10.0	10.9	11.2	11.0	10.86	1.2	0.4454
9	10.6	10.4	10.5	10.5	10.9	10.58	0.5	0.1720
10	8.3	10.2	9.8	9.5	9.8	9.52	1.9	0.6493
11	10.6	9.9	10.7	10.2	11.4	10.56	1.5	0.5083
12	10.8	10.2	10.5	8.4	9.9	9.96	2.4	0.8357
13	10.7	10.7	10.8	8.6	11.4	10.44	2.8	0.9562
14	11.3	11.4	10.4	10.6	11.1	10.96	1.0	0.3029
15	11.4	11.2	11.4	10.1	11.6	11.14	1.5	0.5352
16	10.1	10.1	9.7	9.8	10.5	10.04	0.8	0.2800
17	10.7	12.8	11.2	11.2	11.3	11.44	2.1	0.7116
18	11.9	11.9	11.6	12.4	11.4	11.84	1.0	0.3382
19	10.8	12.1	11.8	9.4	11.6	11.14	2.7	0.9708
20	12.4	11.1	10.8	11.0	11.9	11.44	1.6	0.6086
SUMA						213.20	31.80	11.321

<u>Solución</u>: Puesto que se desconoce la media del proceso, esta se puede estimar en forma insesgada mediante

$$\bar{\bar{X}} = \frac{1}{20} \sum_{i=1}^{20} \bar{X}_i$$

Los valores de los promedios aritméticos $\bar{X}_i$ (i=1,2,...20) de las muestras se reportan en la tabla anterior, por lo cual la línea central es

$$\bar{\bar{X}} = \frac{1}{20} (213.20) = 10.66$$

Se obtendrán ahora los límites inferior y superior de control estimando primero a σ mediante los rangos de las muestras, y después mediante las desviaciones estándar correspondientes.

a. Estimando a σ mediante los rangos de las muestras

El valor de $\bar{R}$ es

$$\bar{R} = \frac{1}{20} \sum_{i=1}^{20} R_i$$

Los valores R_i para i=1,2,...,20 se encuentran en la tabla inicial, por lo que

$$\bar{R} = \frac{1}{20} (31.80) \doteq 1.59$$

Los límites de control para la carta de los promedios son

$$\bar{\bar{X}} \pm A_2 \bar{R}$$

Y, de la tabla I, para n=5, se obtiene $A_2 = 0.577$, quedando

$$10.66 \pm \underbrace{0.577\ (1.59)}_{0.92}$$

O sea

Línea Central —— 10.66

Límites de Control —— $10.66 \pm 0.92 \Rightarrow 11.58,\ 9.74$

b. Estimando a σ mediante las desviaciones estándar de las muestras

El valor de $\bar{\sigma}$ es

$$\bar{\sigma} = \frac{1}{20}\ (11.3211) = 0.57$$

Los límites de control son ahora

$$\bar{\bar{X}} \pm A_1\ \bar{\sigma}$$

De la tabla I, para n=5, se obtiene

$A_1 = 1.596$, quedando

$$10.66 \pm \underbrace{1.596(0.57)}_{0.91}$$

O sea

Línea Central —— 10.66

Límites de Control —— $10.66 \pm 0.91 \Rightarrow 11.57,\ 9.75$

En la Fig 5 que se presenta a continuación se muestra la carta de control obtenida empleando ambos procedimientos.

cionadas permite ejercer control estadístico sobre la variabilidad de un proceso, usualmente se prefiere la carta para los rangos, R, ya que su elaboración es más sencilla que la de σ, que corresponde a las desviaciones estándar. Por otra parte, la carta R conduce a resultados altamente confiables, a la vez que muestra con claridad ciertas tendencias de los valores de las muestras que deben investigarse.

IMPORTANCIA DEL CONTROL DE LA VARIABILIDAD DE UN PROCESO

La importancia del control sobre la variabilidad de un proceso mediante el empleo de las cartas para los rangos o las desviaciones estándar, se hace evidente al considerar que un cambio brusco en aquella característica es de consecuencias más serias que un cambio similar en la "calidad media". Si el proceso experimenta un cambio en ésta última, normalmente se puede regresar al punto de partida efectuando ajustes simples en los dispositivos de producción (por ejemplo, recalibración de herramientas de corte, dosificadoras, etc).

Sin embargo, si el proceso sufre un cambio brusco en su variabilidad, para regresar al punto de partida son necesarios ajustes más costosos y tardados, tales como reparaciones mayores en los dispositivos de producción, o inclusive la compra de un nuevo dispositivo de procesamiento.

Los cambios efectivos en la variabilidad de un proceso afectan necesariamente el desempeño de una carta $\bar{X}$, ya que, como se recordará, los límites de control para la carta de los promedios se establecen

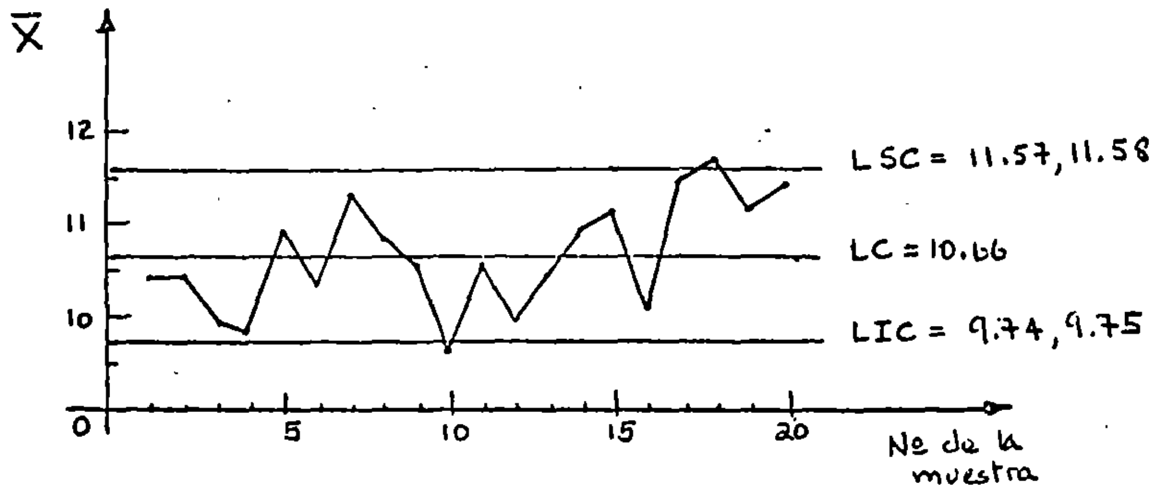


Fig 5 Carta de control $\bar{X}$ obtenida para el ejemplo de las varillas de acero

CARTAS PARA CONTROLAR LA VARIABLILIDAD DE UN PROCESO

Al controlar estadísticamente un proceso puede no ser suficiente - fijar la atención en su "calidad media", sino también en la variabilidad del mismo. Aun cuando es razonable suponer que un incremento en las fluctuaciones de los valores de los promedios aritméticos graficados en una carta $\bar{X}$ se relaciona con un incremento en la variabilidad del proceso, es posible determinar con mayor objetividad y precisión los cambios que experimenta ésta mediante el empleo de las llamadas cartas R y σ, que se elaboran a partir de los rangos y las desviaciones estándar de las muestras, respectivamente.

Conviene mencionar que aun cuando cualquiera de las dos cartas mena partir de los valores $\bar{R}$ o $\bar{\sigma}$, que se suponen, después de ser afectados por los factores de corrección correspondientes, como buenos estimadores de la desviación estándar del proceso. Si los valores del rango y la desviación estándar de las muestras aumentan, se hace evidente que la carta $\bar{X}$ no operará correctamente.

En contraste con lo anterior, los cambios significativos que se verifican en la carta $\bar{X}$ no necesariamente provocan efectos similares en las cartas R y σ, ya que en la elaboración de ellas no intervienen los promedios aritméticos de las muestras, tal como se verá a continuación.

Por lo anteriormente expuesto, es conveniente ejercer, cuando así sea posible, control simultáneo sobre la "calidad media" y la variabilidad de un proceso.

ELABORACION DE LAS CARTA DE CONTROL PARA LOS RANGOS (CARTA R)

Al igual que para la carta $\bar{X}$, se pueden considerar dos casos distintos en la elaboración de la carta para los rangos: cuando se conoce la desviación estándar σ del proceso y cuando ésto no sucede. En cualquiera de los casos anteriores, se debe observar siempre que el procedimiento de obtención de la línea central y de los límites de control para la carta R, se basa en la distribución muestral de los rangos de muestras aleatorias de tamaño n, extraídas de una población normal.

a. Caso en el que se conoce la desviación estándar σ de la - - Población

De acuerdo con lo anterior, es fácil comprender que los pará metros de la carta de control para los rangos son

Línea Central—— μ_R

Límites de Control —— $\mu_R \pm 3\sigma_R$

Sin embargo, normalmente no conocen los valores de la media y la desviación estándar de la distribución muestral de los rangos. En esta situación, la lógica indica que para estimar el valor de μ_R se debe emplear el de $\bar{R}$, el promedio de los rangos de muestras preliminares. Sin embargo, si se recuerda que

$$\sigma = \frac{\bar{R}}{d_2}$$

entonces

$$\bar{R} = d_2\ \sigma$$

Y, puesto que se conoce el valor de σ, se puede escribir

Línea Central —— $\bar{R}$ o $d_2\sigma$

quedando finalmente

Línea Central —— $d_2\sigma$

en donde los valores de d_2 se presentan en la tabla I.

Por lo que respecta a σ_R, si se observa nuevamente la Fig 4 se puede ver que la desviación estándar de la distribución muestral de la estadística R/σ, para el caso de muestras de tamaño 5 es, en forma experimental

$$\sigma_{R/\sigma} = d_3 = 0.864$$

Lo anterior permite considerar que si σ es conocida (y por tanto constante) es válido escribir

$$\sigma_{R/\sigma} = \frac{\sigma_R}{\sigma} = d_3$$

o sea

$$\sigma_R = \sigma_{R/\sigma}\ \sigma = d_3\sigma = 0.864\ \sigma$$

En el caso en que n sea diferente de cinco, los valores del factor d_3 se pueden obtener de la tabla I.

Empleando el valor de σ_R así obtenido, los límites de control son, en general, los siguientes

$$d_2\sigma \pm 3d_3\sigma$$

o sea

$$d_2\sigma - 3d_3\sigma \Rightarrow (d_2 - 3d_3)\sigma \Rightarrow D_1\ \sigma$$

$$d_2\sigma \quad 3d_3\sigma \Rightarrow (d_2 + 3d_3)\sigma \Rightarrow D_2\ \sigma$$

en donde

$$D_1 = d_2 - 3d_3 \quad y \quad D_2 = d_2 + 3d_3$$

Los valores de D_1 y D_2 se reportan también en la tabla I en función de n, el tamaño de la muestra.

Conforme a lo anterior, los parámetros de la carta de control para los rangos, cuando σ es conocida, son

Línea Central —— $d_2\ \sigma$

Límite Inferior de Control —— $D_1\ \sigma$

Límite Superior de Control —— $D_2\ \sigma$

b. Caso en el que se desconoce la desviación estándar σ de la población

En este caso en necesario estimar a μ_R de la distribución muestral de los rangos mediante $\bar{R}$, empleando un número adecuado de muestras preliminares, normalmente el mismo que se emplea para la elaboración de una carta $\bar{X}$. Al respecto, conviene recordar que la carta R (o la σ) generalmente se construye <u>después</u> de la carta $\bar{X}$, y que, por lo tanto, se emplean para su elaboración las mismas muestras aleatorias. De acuerdo con ésto, la línea central resulta ser

$$\text{Línea Central} \;—\; \bar{R}$$

En este caso se requieren límites de control del tipo

$$\bar{R} \pm 3\sigma_R$$

Puesto que ahora se desconocen σ_R y σ, se pueden hacer, para el límite inferior de control

$$\bar{R} - 3\sigma_R = \bar{R} - \frac{3\,\bar{R}\,\sigma_R}{\bar{R}} = \left(1 - 3\,\frac{\sigma_R}{\bar{R}}\right)\bar{R}$$

$$= \left(1 - 3\,\frac{\frac{\sigma_R}{\sigma}}{\frac{\bar{R}}{\sigma}}\right)\bar{R} = \left(1 - 3\,\frac{d_3}{d_2}\right)\bar{R}$$

$$= \left(\frac{d_2 - 3d_3}{d_2}\right)\bar{R} = \left(\frac{D_1}{d_2}\right)\bar{R}$$

Para el límite superior de control se obtiene

$$\bar{R} + 3\sigma_R = \bar{R}\left(\frac{D_2}{d_2}\right)$$

En la tabla I se presentan los valores de

$$D_3 = \frac{D_1}{d_2} \qquad y \qquad D_4 = \frac{D_2}{d_2}$$

en función de n .

Finalmente, los parámetros de la carta R cuando se desconoce el valor de σ de la población son los siguientes:

Línea Central —— $\bar{R}$

Límite Inferior de Control —— $D_3\bar{R}$

Límite Superior de Control —— $D_4\bar{R}$

ELABORACION DE LA CARTA DE CONTROL PARA LAS DESVIACIONES ESTANDAR (CARTA σ)

En la elaboración de la carta para las desviaciones estándar también se deben considerar los dos casos posibles: cuando se conoce la desviación estándar de la población y cuando ésto no es así. De igual manera, el procedimiento para obtener los paráme‿tros de la carta se fundamenta en la distribución muestral de las desviaciones estándar de muestras aleatrias de tamaño n, extraídas de una población normal.

a. Caso en el que se conoce la desviación estándar σ de la población

Con base en la distribución muestral de las desviaciones estándar de las muestras, se pueden establecer los parámetros de la carta σ, a saber

$$\text{Línea Central} \longrightarrow \mu_{S_X}$$

$$\text{Límites de Control} \longrightarrow \mu_{S_X} \pm 3\sigma_{S_X}$$

Al desconocerse, como ocurre normalmente, los valores de μ_{S_X} y σ_{S_X} de la distribución muestral, se debe estimar primero μ_{S_X} a partir de $\bar{\sigma}$, el promedio de las desviaciones estándar de las muestras preliminares. Sin embargo, no es necesario realizar en este caso ese cálculo si se recuerda que

$$\sigma = \frac{\bar{\sigma}}{c_2}$$

o sea

$$\bar{\sigma} = c_2\ \sigma$$

Y, en virtud de que el valor de σ es conocido, se llega a

$$\text{Línea Central} \longrightarrow \bar{\sigma} \quad \text{o} \quad c_2\sigma$$

quedando finalmente

$$\text{Línea Central} \longrightarrow c_2\sigma$$

en donde los valores de c_2 se pueden obtener de la tabla I.

Bajo la suposición de que la población de la cual se extraen las muestras aleatorias se encuentra distribuida en forma normal (o aproximadamente normal), se puede demostrar que la desviación estándar de la distribución muestral de las desviaciones estándar es

$$\sigma_{S_X} = \frac{\sigma}{\sqrt{2n}}$$

en donde n denota al tamaño de las muestras. Empleando el va lor de σ_{S_X} anterior, los límites de control se pueden establecer como

$$\mu_{S_X} \pm 3\sigma_{S_X} = c_2\sigma \pm 3 \frac{\sigma}{\sqrt{2n}}$$

o sea

$$c_2\sigma - 3 \frac{\sigma}{\sqrt{2n}} = \left(c_2 - \frac{3}{\sqrt{2n}} \right) \sigma = B_1\sigma$$

$$c_2\sigma + 3 \frac{\sigma}{\sqrt{2n}} = \left(c_2 + \frac{3}{\sqrt{2n}} \right) \sigma = B_2\sigma$$

en donde

$$B_1 = c_2 - \frac{3}{\sqrt{2n}}$$

$$B_2 = c_2 + \frac{3}{\sqrt{2n}}$$

Los valores de B_1 y B_2 se proporcionan en la tabla I, en función del valor de n. Entonces, los parámetros de la carta σ son, finalmente

Línea Central —— $c_2\sigma$

Límite Inferior de Control —— $B_1\sigma$

Límite Superior de Control —— $B_2\sigma$

b. Caso en el que se desconoce la desviación estándar σ de la población

En este caso es necesario estimar a μ_{S_X} mediante $\bar{\sigma}$, empleando un número suficiente de muestras aleatorias preliminares.

De acuerdo con lo anterior, la línea central de la carta σ es

Línea Central —— $\bar{\sigma}$

Los límites de control serán entonces del tipo

$$\bar{\sigma} \pm 3\sigma_{s_x}$$

Puesto que ahora se desconoce el valor de σ, pero se sabe que

$$\sigma = \frac{\bar{\sigma}}{c_2}$$

el límite inferior de control resulta ser

$$\bar{\sigma} - 3\sigma_{s_x} = \bar{\sigma} - 3\,\frac{\sigma}{\sqrt{2n}} = \bar{\sigma} - 3\,\frac{\bar{\sigma}}{c_2\sqrt{2n}}$$

$$= \left(1 - \frac{3}{c_2\sqrt{2n}}\right)\bar{\sigma}$$

Para el límite superior de control se obtiene

$$\bar{\sigma} + 3\sigma_{s_x} = \left(1 + \frac{3}{c_2\sqrt{2n}}\right)\bar{\sigma}$$

En la tabla I se presentan los valores de

$$B_3 = 1 - \frac{3}{c_2\sqrt{2n}} \quad \text{y} \quad B_4 = 1 + \frac{3}{c_2\sqrt{2n}}$$

en función del valor de n.

Finalmente, los parámetros de la carta σ, cuando no se conoce la desviación estándar de la población, quedan como

Línea Central —— $\bar{\sigma}$

Límite Inferior de Control —— $B_3\bar{\sigma}$

Límite Superior de Control —— $B_4\bar{\sigma}$

<u>Ejemplo</u>: Sea el proceso de elaboración de varillas de acero mencionado en la página 10 de estos apuntes. En él se informa que el diámetro medio de las varillas es igual a 2.5 cm , con desviación estándar de 0.01 cm. En este caso se pide establecer los parámetros de las cartas de control R y σ, considerando que se extraen periódicamente muestras de cinco varillas.

<u>Solución</u>:

a. Carta R

Puesto que se conoce el valor de la desviación estándar de la población, y en virtud de que n=5, se obtiene, empleando la tabla I

$$LC \text{ —— } d_2\sigma = 2.326(0.01) = 0.02326$$

$$LIC \text{ —— } D_1\sigma = 0(0.01) = 0.0000$$

$$LSC \text{ —— } D_2\sigma = 4.918(0.01) = 0.04918$$

b. Carta σ

En este caso, puesto que σ=0.01 y n=5, se obtiene, con el uso de la tabla I

$$LC \text{ —— } c_2\sigma = 0.8407(0.01) = 0.008407$$

$$LIC \text{ —— } B_1\sigma = 0(0.01) = 0.00000$$

$$LSC \text{ —— } B_2\sigma = 1.756(0.01) = 0.01756$$

<u>Ejemplo</u>: Con el fin de investigar la variabilidad en el proceso de producción de varillas de acero mencionado en la página - 16, se desea elaborar las cartas de control R y σ correspondientes, considerando la información contenida en la tabla de la misma página.

<u>Solución</u>:

En este caso se desconoce la desviación estándar de la población, por lo cual es indispensable emplear los valores de $\bar{R}$ y $\bar{\sigma}$, considerando que el tamaño de la muestra es 5.

a. Carta R

El valor de $\bar{R}$, obtenido durante el proceso de elaboración de la carta $\bar{X}$ correspondiente, es $\bar{R} = 1.59$. - Considerando este valor, y empleando la tabla I, los parámetros de la carta de control R resultan

$$LC \text{ —— } \bar{R} = 1.590$$

$$LIC \text{ —— } D_3\bar{R} = 0(1.59) = 0.000$$

$$LSC \text{ —— } D_4\bar{R} = 2.115(1.59) = 3.363$$

En la Fig 6 se presenta la carta R para este problema.

b. Carta σ

Considerando que al calcular para este problema los parámetros de la carta $\bar{X}$ se obtuvo $\bar{\sigma} = 0.57$, la carta σ queda definida con

$$LC \text{ —— } \bar{\sigma} = 0.57$$

$$LIC \text{ —— } B_3\bar{\sigma} = 0(0.57) = 0.00$$

$$LSC \text{ —— } B_4\bar{\sigma} = 2.089(0.57) = 1.19$$

En la Fig 7 se muestra la carta de control σ correspon diente.

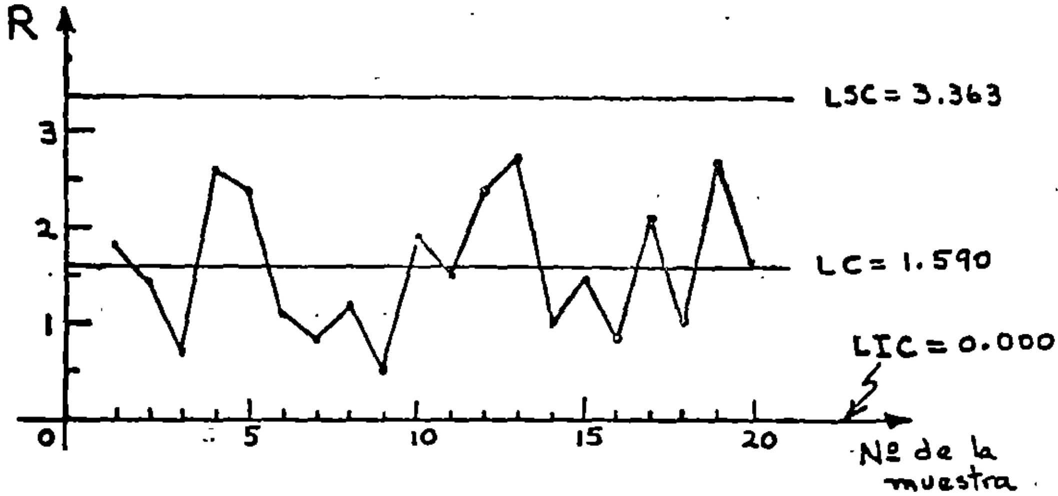


Fig 6 Carta de control R obtenida para el ejemplo de las varillas de acero

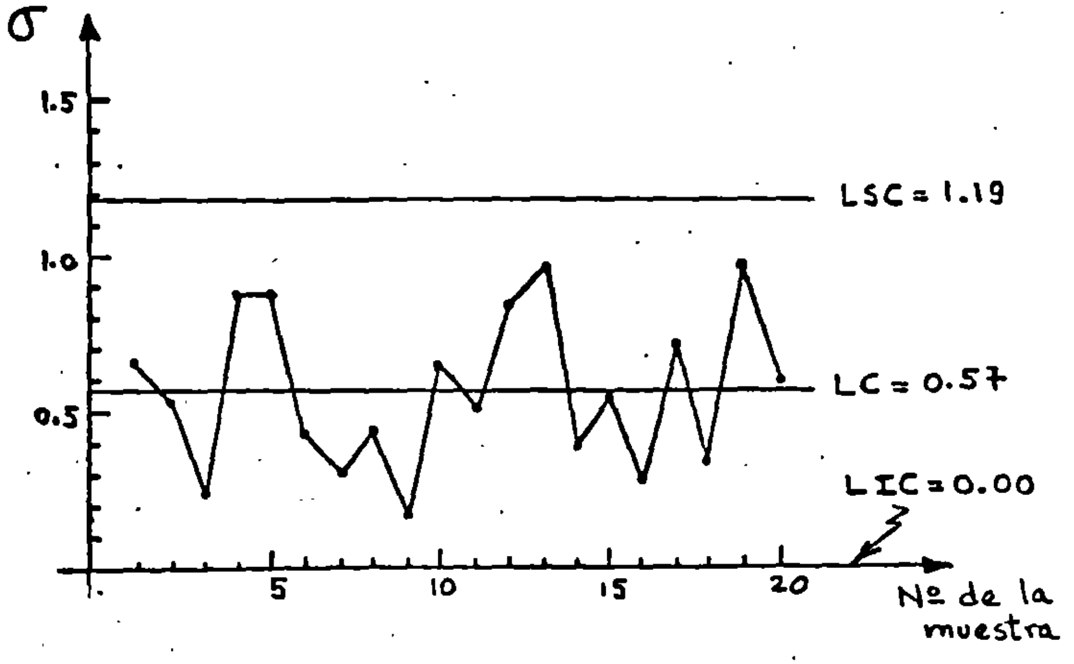


Fig 7 Carta de control σ obtenida para el ejemplo de las varillas de acero

CARTAS DE CONTROL PARA MEDICIONES (ELEMENTOS INDIVIDUALES)

Se han establecido las cartas $\bar{X}$, R y σ considerando que existe la posibilidad de conocer la media μ y/o la desviación estándar σ de la población (proceso), o bien, cuando estos parámetros se desconocen, que es posible obtener un número adecuado de muestras aleatorias de ella, cuyos tamaños sean cuando menos igual a dos, con el fin de estimar con buena precisión los valores de dichos parámetros.

Sin embargo, en muchas ocasiones no se conocen los parámetros del proceso, y únicamente es posible contar con muestras de tamaño uno, es decir, muestras con un solo elemento. Cuando ésto sucede, la técnica para calcular los límites de control en las cartas para mediciones se fundamenta en el empleo de los llamados rangos móviles, que se explican a continuación.

Si, por ejemplo, se cuenta con el conjunto de datos X_i $(i=1,2,\ldots,n)$ registrados en orden, se definen los rangos móviles de orden dos como

$$\left|X_i - X_{i+1}\right| \qquad ; \qquad 1 \leq i \leq n-1$$

es decir

$$\left|X_1 - X_2\right| , \left|X_2 - X_3\right| , \ldots , \left|X_{n-1} - X_n\right|$$

Si se trata de rangos móviles de orden tres, éstos se definen como

$$\left|X_i - X_{i+2}\right| \qquad ; \qquad 1 \leq i \leq n-2$$

es decir

$$\left|X_1 - X_3\right| , \left|X_2 - X_4\right| , \ldots , \left|X_{n-2} - X_n\right|$$

La obtención de los rangos móviles de orden superior al tres se hace siguiendo las ideas anteriores.

En forma numérica, si se tienen los datos registrados en orden 4, 6, 4, 3 y 7, los rangos móviles de orden dos son

$$|4 - 6| = 2 \ , \quad |6 - 4| = 2 \ , \quad |4 - 3| = 1 \ , \quad |3 - 7| = 4$$

y los de orden tres son

$$|4 - 4| = 0 \ , \quad |6 - 3| = 3 \ , \quad |4 - 7| = 3$$

El empleo de los rangos móviles para la obtención de los límites de control es importante en este caso, debido a que, si se trata de rangos móviles de orden dos, se puede considerar que el valor de cualquiera de ellos debe obtenerse a partir de los valores de dos elementos individuales registrados en orden. Dicho de otra manera, un rango móvil de orden dos debe provenir de una muestra "ficticia" de tamaño dos. En la misma forma, un rango móvil de orden tres tiene que obtenerse a partir de tres elementos individuales, lo cual permite "crear" muestras de tamaño tres.

De acuerdo con lo anterior, es factible establecer los límites de control para las cartas de control, en el caso de elementos individuales, empleando los factores de la tabla I, que se encuentran tabulados a partir de muestras de tamaño dos.

a. Elaboración de la carta X (elementos individuales)

En este caso, la línea central está dada por

$$\bar{X} = \frac{1}{K} \sum_{i=1}^{K} X_i$$

en donde X_i $(i=1,2,\ldots,K)$ denota a los valores de los datos -

individuales.

Los límites de control requeridos son

$$\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{n}}$$

Puesto que el tamaño real de la muestra es uno, la expresión anterior se puede escribir

$$\bar{\bar{X}} \pm 3\frac{\sigma}{\sqrt{1}} = \bar{\bar{X}} \pm 3\sigma$$

Debido a que el valor de σ se desconoce, pero es posible obtener el de $\bar{R}$ (promedio de los rangos móviles), la última expresión puede transformarse algebraicamente de la siguiente manera:

$$\bar{\bar{X}} \pm 3\sigma = \bar{\bar{X}} \pm \frac{3\sigma\ \bar{R}}{\bar{R}} = \bar{\bar{X}} \pm \frac{3\bar{R}}{\frac{\bar{R}}{\sigma}}$$

$$\bar{\bar{X}} \pm \frac{3\ \bar{R}}{d_2} = \bar{\bar{X}} \pm E_2\bar{R}$$

en donde

$$E_2 = \frac{3}{d_2}$$

Los valores de E_2 se pueden obtener de la tabla I en función de n, que representa ahora el tamaño "ficticio" de la muestra, o el orden de los rangos móviles.

De acuerdo con lo anterior, los parámetros de la carta de control X para elementos individuales son

Línea Central —— $\bar{\bar{X}}$

Límite Inferior de Control —— $\bar{\bar{X}} - E_2\bar{R}$

Límite Superior de Control —— $\bar{\bar{X}} + E_2\bar{R}$

b. Elaboración de la carta R* (rangos móviles)

En este caso, la línea central está dada por el valor del promedio de los rangos móviles, es decir

$$\bar{R} = \frac{1}{K} \sum_{i=1}^{K} R_i$$

En donde R_i $(i=1,2,\ldots,K)$ denota a los valores de los rangos móviles, obtenidos a partir de los datos individuales registrados en orden.

Los límites de control se obtienen considerando que se desconoce el valor de la desviación estándar de la población, en la forma ya explicada para la carta R.

De acuerdo con lo anterior, los parámetros de la carta de control R* para los rangos móviles son

Línea Central —— $\bar{R}$

Límite Inferior de Control —— $D_3\bar{R}$

Límite Superior de Control —— $D_4\bar{R}$

en donde los valores de D_3 y D_4 se obtienen de la tabla I en función de n, el tamaño "ficticio" de la muestra, u orden de los rangos móviles.

<u>Ejemplo</u>: Considérese un proceso de destilación y mezclado de alcohol, para el cual se desea ejercer control sobre el porcentaje de metanol existente. Se extraen 26 lotes sucesivos de alcohol, y se obtiene el porcentaje de metanol correspondiente para cada uno de ellos. Los valores

se presentan en la tabla siguiente, y se pide construir cartas X y R* considerando rangos móviles de orden dos.

Lote	Porcentaje de metanol, X	Rango móvil, R	Lote	Porcentaje de metanol, X	Rango móvil, R
1	4.6	---	14	5.5	0.1
2	4.7	0.1	15	5.2	0.3
3	4.3	0.4	16	4.6	0.6
4	4.7	0.4	17	5.5	0.9
5	4.7	0	18	5.6	0.1
6	4.6	0.1	19	5.2	0.4
7	4.8	0.2	20	4.9	0.3
8	4.8	0	21	4.9	0
9	5.2	0.4	22	5.3	0.4
10	5.0	0.2	23	5.0	0.3
11	5.2	0.2	24	4.3	0.7
12	5.0	0.2	25	4.5	0.2
13	5.6	0.6	26	4.4	0.1
			SUMA	128.1	7.2

Solución: El valor del promedio de los rangos móviles de orden dos es

$$\bar{R} = \frac{1}{25} \sum_{i=1}^{25} R_i = \frac{1}{25} (7.2) = 0.288$$

a. Carta X

La línea central de esta carta es $\bar{X}$, cuyo valor es

$$\bar{X} = \frac{1}{26} \sum_{i=1}^{26} X_i = \frac{1}{26} (128.1) = 4.927$$

De la tabla I se obtiene $E_2 = 2.66$ para n=2, - siendo los límites de control

$$\bar{X} \pm E_2\bar{R} = 4.927 \pm 2.66(0.288)$$

$$= 4.927 \pm 0.7661$$

Finalmente, los parámetros de la carta X quedan como

LC —— 4.927

LIC—— 4.927 - 0.7661 = 4.161

LSC—— 4.927 + 0.7661 = 5.693

En la Fig 8 se presenta la gráfica correspondiente.

b. Carta R*

La línea central para esta carta es $\bar{R} = 0.288$, y los límites de control se obtienen empleando la tabla I considerando que n=2. De ahi que

LC —— 0.288

LIC—— $D_3\bar{R} = 0(0.288) = 0.000$

LSC—— $D_4\bar{R} = 3.267(0.288) = 0.941$

La Fig 9 muestra la carta R* para este problema.

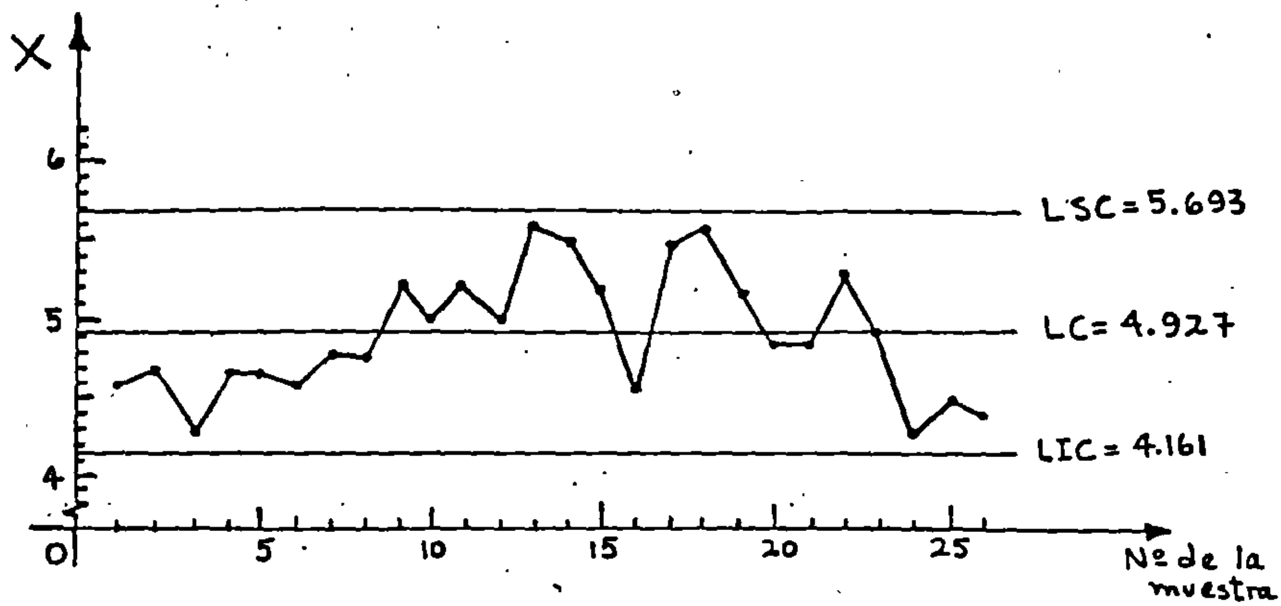


Fig 8   Carta de control X obtenida para el ejemplo de los lotes de alcohol

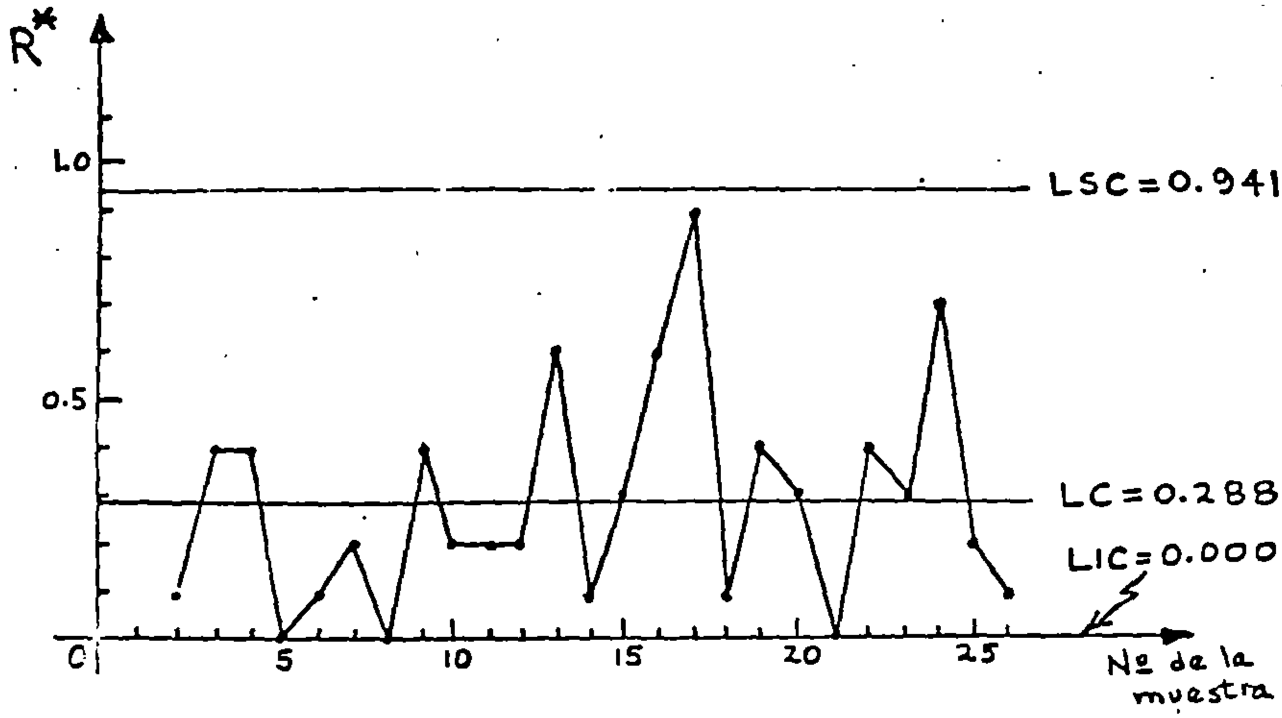


Fig 9   Carta de ccntrol R* obtenida para el ejemplo de los lotes de alcohol

CARTAS DE CONTROL PARA ATRIBUTOS

El término atributo, tal como se emplea en el control de calidad, indica la propiedad que tiene un producto de ser bueno o malo, es decir, permite reconocer si la característica de calidad del mismo se encuentra dentro de ciertos requerimientos específicos o no.

Aunque generalmente se puede obtener información más completa de las mediciones hechas a productos terminados, a menudo consume menos tiempo y dinero el comparar la calidad de un producto en contra de ciertas especificaciones mínimas, sobre la base, por ejemplo, de considerar que sirve o no, o que es bueno o malo.

Por ejemplo, al ejercer control sobre el diámetro de un balín de acero, es más simple y rápido el determinar si éste pasa por un agujero hecho en una placa de acero templado con el diámetro adecuado, que realizar la medición del diámetro con un micrómetro.

Se establecerán ahora los dos tipos fundamentales de cartas de contr que se utilizan en conexión con el muestreo por atributos: la carta para la proporción de elementos defectuosos, o carta p, y la carta para el número de defectos, o carta c.

Considérese por ejemplo una muestra de 50 fusibles en la cual se encontró, después de probar todos ellos, que contiene dos elementos defectuosos. En este caso, la proporción de fusibles defectuosos en la muestra es de 2/50 = 0.04.

Por otra parte, debe observarse que si se prueba una sola unidad producida, esta puede tener varios defectos pero, sin embargo, puede o no ser una unidad defectuosa. Tal es el caso, por ejemplo, de rollos (unidades) de tela de determinada longitud, que pueden tener cierto número de imperfecciones pero no necesariamente ser considerados como defectuosos. No obtante, en muchas aplicaciones prácticas una unidad producida se considera defectuosa si tiene cuando menos un defecto.

La distribución de la proporción y del número de elementos defectuosos en un proceso es obviamente binomial, en tanto que la del número de defectos es de Poisson. Sin embargo, para la elaboración de la carta p se aprovecha la propiedad que tiene la distribución muestral de las proporciones de ser aproximada mediante una distribución normal cuando el tamaño de la muestra es grande, y la proporción de elementos defectuosos no se acerca a cero o a uno.

ELABORACION DE LAS CARTAS DE CONTROL p Y np PARA LA PROPORCION DE DEFECTUOSOS Y EL NUMERO DE DEFECTUOSOS

Los límites de control que se requieren en este caso son

$$\mu_p \pm 3\sigma_p$$

en donde μ_p es la media de la distribución muestral de las proporciones, y σ_p la desviación estándar correspondiente. Como μ_p de esta distribución es igual al parámetro P de la población, la estadística p de la muestra estima en forma insesgada a este último.

Si no se conoce el valor de P de la población, lo cual en la práctica es frecuente, se debe disponer de K muestras de tamaño n constante para obtener el valor del estimador insesgado

$$\bar{p} = \frac{1}{K} \sum_{i=1}^{K} p_i$$

en donde p_i $(i=1,2,\ldots,K)$ denota el valor de la proporción en la muestra i. Empleando el valor así obtenido, la línea central es

$$\text{Línea Central} \text{ --- } \bar{p}$$

En textos de estadística se demuestra que la desviación estándar de la distribución muestral de las proporciones es

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

por lo cual los límites de control son

$$\bar{p} \pm 3\sigma_p = \bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Finalmente, los parámetros de la carta de control p quedan como

$$\text{Línea Central} \text{ --- } \bar{p}$$

$$\text{Límite Inferior de Control} \text{ --- } \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\text{Límite Superior de Control} \text{ --- } \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

A partir de los parámetros anteriores se pueden derivar los de la llamada carta np, o sea, para el número de defectuosos. Para ello, es necesario multiplicar dichos parámetros por n para así obtener, en el caso de los límites de control

$$n\left(\bar{p} \pm 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right) = n\bar{p} \pm 3n\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$= n\bar{p} \pm 3\sqrt{\frac{n^2\bar{p}\ (1-\bar{p})}{n}}$$

$$= n\bar{p} \pm \sqrt{3\ n\ \bar{p}\ (1-\bar{p})}$$

y los parámetros resultan ahora

Línea Central —— $n\bar{p}$

Límite Inferior de Control —— $n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$

Límite Superior de Control —— $n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$

Ejemplo: Para un proceso de elaboración de fusibles se desea ejercer control sobre la proporción de elementos defectuosos, así como sobre el número de ellos. Para ello, se seleccionan 40 muestras aleatorias de 50 fusibles cada una, y se obtienen los valores reportados en la tabla siguiente.

Se desea construir las cartas p y np correspondientes.

Número de la muestra	Número de fusibles defectuosos	Proporción de defectuosos, p	Número de la muestra	Número de fusibles defectuosos	Proporción de defectuosos, p
1	2	0.04	21	1	0.02
2	1	0.02	22	1	0.02
3	2	0.04	23	4	0.08
4	0	0.00	24	2	0.04
5	2	0.04	25	2	0.04
6	3	0.06	26	4	0.08
7	4	0.08	27	1	0.02
8	2	0.04	28	3	0.06
9	0	0.00	29	3	0.06
10	3	0.06	30	2	0.04
11	0	0.00	31	3	0.06
12	1	0.02	32	6	0.12
13	2	0.04	33	2	0.04
14	2	0.04	34	3	0.06
15	3	0.06	35	2	0.04
16	5	0.10	36	3	0.06
17	1	0.02	37	1	0.02
18	2	0.04	38	0	0.00
19	3	0.06	30	2	0.04
20	1	0.02	40	0	0.00
			S U M A		1.68

Solución: El valor de $\overline{p}$ es

$$\overline{p} = \frac{1}{40} \sum_{i=1}^{40} p_i = \frac{1}{40} (1.68) = 0.042$$

a. Carta p

Los límites de control son, para n=50

$$0.042 \pm 3\sqrt{\frac{(0.042)\ (1-0.042)}{50}} = 0.042 \pm 0.0851$$

por lo cual

LC ——— 0.0420

LIC——— 0.042 - 0.0851 = -0.0431 $\Rightarrow$ 0.0000

LSC——— 0.042 + 0.0851 = 0.1271

En este caso, y como se verá a continuación para la carta np, la expresión para el cálculo del límite inferior de control conduce a un valor negativo del mismo. Puesto que no tiene sentido físico hablar de una proporción menor de cero o de un número de defectuosos negativo, en forma arbitraria se asigna a ese límite el valor cero.

En la Fig 10 se presenta la carta de control p correspondiente.

b. Carta np

Puesto que $n\overline{p}$ = 50(0.042)=2.1, los límites de control son ahora

$$2.1 \pm 3\ \sqrt{50(0.042)\ (1-0.042)} = 2.1 \pm 4.255$$

o sea

LC —— 2.1

LIC —— 2.1-4.255 = -2.155 ⇛ 0.000

LSC —— 2.1+4.255 = 6.355

En la Fig 10 se presenta la carta np para este problema.

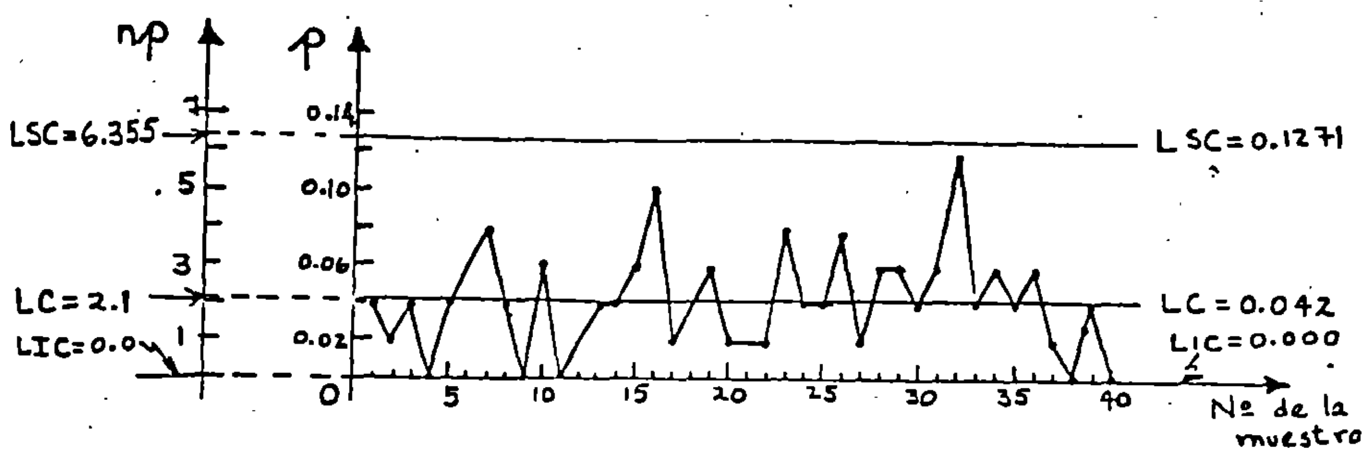


Fig 10 Cartas de control p y np obtenidas para el ejemplo de los fusibles

ELABORACION DE LA CARTA DE CONTROL c PARA EL NUMERO DE DEFECTOS

Existen ocasiones en las que es necesario controlar el número de defectos por unidad en un proceso. Por ejemplo, en la producción de alfombras es importante controlar el número de defectos por metro cuadrado; en la elaboración de papel se requiere controlar el número de defectos por rollo, etc. En estos casos, la variable aleatoria c asociada al número de defectos por unidad tiene una distribución de Poisson.

De lo anterior se desprende que la línea central de la carta de con trol para el número de defectos es el parámetro λ de la distribución de Poisson correspondiente, cuyo valor usualmente se desconoce. En tal situación, se acostumbra estimar en forma insesgada el valor de λ a partir de un mínimo de 20 valores de c, observados previamente en igual número de unidades producidas. De acuerdo con ésto, el valor de

$$\bar{c} = \frac{1}{K} \sum_{i=1}^{K} c_i$$

en donde c_i $(i=1,2,\ldots,K)$ representa el número de defectos observados en la unidad i, se puede emplear como estimador de λ.

Los límites de control requeridos ahora son del tipo

$$\bar{c} \pm 3\sigma_c$$

Puesto que en este caso se observa el número de defectos por unidad, se puede suponer que el tamaño de la muestra es unitario. Por tal motivo, se puede considerar que la desviación estándar de la distribución muestral del número de defectos c es igual a la desviación estándar de la distribución de Poisson y, puesto que $\bar{c}$ estima el valor de λ

$$\sigma_c = \sqrt{\lambda} = \sqrt{\bar{c}}$$

De acuerdo con lo anterior, los parámetros de la carta de control c son

Línea Central —— $\bar{c}$

Límite Inferior de Control —— $\bar{c} - 3\sqrt{\bar{c}}$

Límite Superior de Control —— $\bar{c} + 3\sqrt{\bar{c}}$

<u>Ejemplo</u>: Considérese el proceso de soldadura de dos placas de acero en una fábrica. Diariamente se alcanzan a soldar 8 juntas, y en cada una de ellas se observa el número de defectos existente. Con la información correspondiente a tres días de labor que se presenta en la tabla siguiente, se desea elaborar una carta de control para el número de defectos por junta soldada

Número de la junta soldada	Fecha	Número de defectos
1	Julio 18	2
2		4
3		7
4		3
5		1
6		4
7		8
8		9
9	Julio 19	5
10		3
11		7
12		11
13		6
14		4
15		9
16		9
17	Julio 20	6
18		4
19		3
20		9
21		7
22		4
23		7
24		12
SUMA................		144

Solución: Empleando los valores reportados en la tabla anterior, el valor de $\bar{c}$ resulta

$$\bar{c} = \frac{1}{24} \sum_{i=1}^{24} c_i = \frac{1}{24} (144) = 6$$

Siendo $\bar{c} = 6$, los límites de control quedan como

$$6 \pm 3\sqrt{6} = 6 \pm 7.35$$

Finalmente, los parámetros de la carta c son

$$\begin{aligned} &LC \text{ — } 6 \\ &LIC \text{ — } 6 - 7.35 = -1.35 \Rightarrow 0.00 \\ &LSC \text{ — } 6 + 7.35 = 13.35 \end{aligned}$$

Puesto que el número de defectos no puede ser negativo, se fija el valor del límite inferior de control igual a cero.

En la Fig 11 se presenta la carta de control c que corresponde al ejemplo.

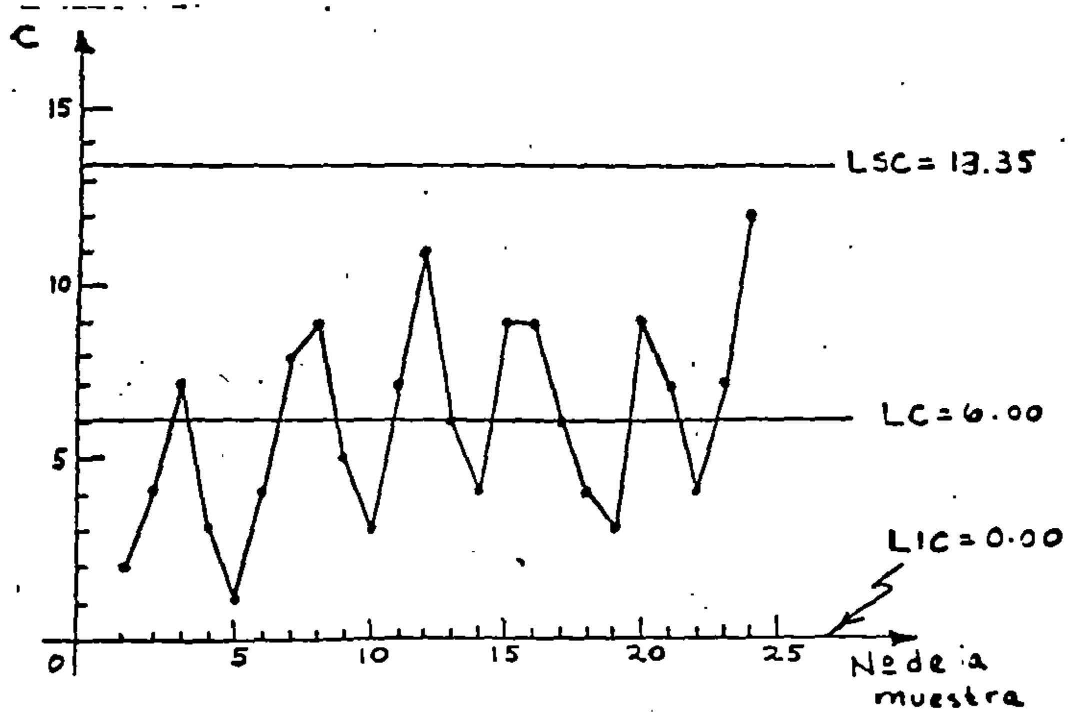


Fig 11 Carta de control c obtenida para el ejemplo de las juntas soldadas

B I B L I O G R A F I A

1. Hansen, B., "Quality Control: Theory and Applications", Prentice Hall, Inc. (1964)

2. Grant, E.L., "Statistical Quality Control", Mc Graw-Hill Book Co. (1971)

3. Ostle, B. "Estadística aplicada", Limusa-Wiley (1973)

4. Miller, I. y Freund, J., "Probability and Statistics - for Engineers", Prentice Hall, Inc. (1965)

FACULTAD DE INGENIERIA U.N.A.M.
DIVISION DE EDUCACION CONTINUA

DIPLOMADO DE CALIDAD EN LA CONTRUCION

MODULO I: METODOS ESTADISTICOS PARA CONTROL DE CALIDAD

TEMA: INFERENCIA ESTADISTICA

EXPOSITOR: M. EN I. AUGUSTO VILLARREAL A.

INFERENCIA ESTADISTICA

Por: M en I Augusto Villarreal Aranda*

1. Introducción

La parte de la estadística que proporciona las reglas para inferir ciertas características de una población a partir de muestras extraídas de ella, junto con indicaciones probabilísticas de la veracidad de tales inferencias, se llama *inferencia estadística*.

En la inferencia estadística se estudian las relaciones existentes entre una población, las muestras obtenidas de ella, y las técnicas para estimar parámetros, tales como la media y la varianza, o bien para determinar si las diferencias entre dos muestras son debidas al azar, etc.

2. Distribuciones muestrales

Si se consideran todas las muestras posibles de tamaño

* *Secretario Académico*, División de Estudios Superiores, Facultad de Ingeniería, UNAM y *Profesor investigador*, Instituto de Ingeniería, UNAM

n que pueden extraerse de una población, y para cada una se calcula el valor del promedio aritmético, este seguramente variará de una muestra a otra, ya que depende de los valores de los datos que se hayan obtenido en cada muestra. Por lo tanto, el promedio aritmético es en sí una variable aleatoria, como también lo son, por la misma razón, el rango y la variancia de la muestra.

A todo elemento que es función de los valores de los datos que se tienen en una muestra se le denomina *estadística;* toda estadística es, entonces, una variable aleatoria cuya distribución de probabilidades se conoce como *distribución muestral*. Si, por ejemplo, la estadística considerada es la variancia de la muestra, su densidad de probabilidades se llama *distribución muestral de la variancia.*

En forma similar se pueden obtener las distribuciones muestrales de la desviación estándar, del rango, etc., cada una de las cuales tendrá sus propios parámetros, lo que permite hablar de la media y la desviación estándar de la variancia, etc.

3. Muestreo con y sin remplazo

Cuando se efectúa un muestreo en una población de tal manera que cada elemento de la misma se pueda escoger más de una vez, se dice que el muestreo es *con remplazo;* en caso contrario, el muestreo es *sin remplazo*. Si de una urna se quiere extraer una muestra de bolas de colores, se puede proceder de dos maneras: se saca al azar una bola, se anota su color y se regresa a la urna antes de obtener otra, y así sucesivamente; en este caso el muestreo es *con remplazo*. La segunda forma consiste en extraer al azar todas las bolas que constituyen la muestra sin regresarlas a la urna, siendo entonces un muestreo *sin remplazo*.

4. Distribucion muestral del promedio aritmético

Supóngase que se extraen sin remplazo todas las muestras posibles de tamaño n de una población finita de tamaño $N_p > n$. Si la media y la desviación estándar de la distribución muestral del promedio aritmético se denotan con $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$, y la media y la desviación estándar de la población con μ y σ, respectivamente, entonces es posible demostrar que se cumplen las siguientes ecuaciones

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Además, si la población es infinita (o el muestreo es con remplazo), los resultados anteriores se reducen a

$$\mu_{\bar{X}} = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

puesto que

$$\lim_{N_p \to \infty} \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}} = \frac{\sigma}{\sqrt{n}}$$

Para valores grandes de $n(n \geqslant 30)$ se demuestra, empleando el teorema del límite central, que la distribución muestral del promedio aritmético es aproximadamente una distribución normal con media $\mu_{\bar{X}}$ y desviación estándar $\sigma_{\bar{X}}$, independientemente de cuál sea la densidad de probabilidades de X, la variable aleatoria asociada a la población. Si esta variable tiene distribución normal, la distribución muestral del promedio aritmético también es normal, aun para valores pequeños de n $(n < 30)$.

Ejemplo 4.1

Supóngase que se tiene una población finita formada por los datos 1,2,3,4,5. Se desea conocer la media y la desviación estándar de la distribución muestral del promedio aritmético, considerando las muestras de tamaño 3 obtenidas sin remplazo.

Primer procedimiento.

Siendo la población finita y el muestreo sin remplazo, es posible obtener la distribución muestral correspondiente para calcular después sus parámetros, considerando que el número total de muestras distintas de tamaño 3 que pueden obtenerse a partir de una población de 5 elementos es

$$\frac{5!}{3!(5-3)!} = 10$$

Dichas muestras son las siguientes, junto con sus promedios aritméticos correspondientes:

	$\bar{X}_i$		$\bar{X}_i$
1, 2, 3	6/3	3, 4, 5	12/3
1, 2, 4	7/3	3, 4, 1	8/3
1, 2, 5	8/3	4, 5, 1	10/3
2, 3, 4	9/3	4, 5, 2	11/3
2, 3, 5	10/3	5, 1, 3	9/3

Para calcular la media y la desviación estándar, se emplea la siguiente tabla

$\bar{X}_i$	6/3	7/3	8/3	8/3	9/3	9/3	10/3	10/3	11/3	12/3
$\bar{X}_i^2$	36/9	49/9	64/9	64/9	81/9	81/9	100/9	100/9	121/9	144/9

$$\sum_{i=1}^{10} \bar{X}_i = 90/3 \qquad \sum_{i=1}^{10} \bar{X}_i^2 = 840/9$$

$$\mu_{\bar{X}} = \bar{\bar{X}} = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i = \frac{1}{10} \quad \frac{90}{3} = 3$$

$$\sigma_{\bar{X}}^2 = \frac{1}{10} \sum_{i=1}^{10} \bar{X}_i^2 - \bar{\bar{X}}^2 = \frac{1}{10} \quad \frac{840}{9} - (3)^2 =$$

$$= 9.333 - 9.000 = 0.333 \Rightarrow \sigma_{\bar{X}} = \sqrt{0.333} = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Segundo procedimiento.

Por tratarse de una población finita, se verifica que

$$\mu_{\bar{X}} = \mu \qquad \text{y} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en donde $N_p = 5$, $n = 3$ y $\mu = 3$.

El valor de σ^2 de la población es

$$\sigma^2 = \frac{1+4+9+16+25}{5} - (3)^2 = \frac{55}{5} - 9 = 11-9 = 2$$

Por lo tanto, $\sigma = \sqrt{2} = 1.4145$ y

$$\sigma_{\bar{X}} = \frac{1.4145}{\sqrt{3}} \sqrt{\frac{5-3}{5-1}} = (0.8164)(0.7071) = 0.577$$

Es decir, $\mu_{\bar{X}} = 3$ y $\sigma_{\bar{X}} = 0.577$

Comparando los resultados, se puede observar que ambos procedimientos conducen a la obtención de los mismos valores de $\mu_{\bar{X}}$ y $\sigma_{\bar{X}}$ para la distribución muestral del promedio aritmético.

Ejemplo 4.2

En una bodega se tienen cinco mil varillas de acero; el valor medio del peso, X, de cada varilla es de 5.02 kg, y la desviación estándar 0.3 kg. Hallar la probabilidad de que una muestra de cien varillas, escogida al azar, tenga un peso total

a. entre 496 y 500 kg

b. de más de 510 kg.

Para la distribución muestral del promedio, se tiene que $\mu_{\bar{X}} = \mu = 5.02$ kg y, por tratarse de una población finita,

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \frac{0.30}{\sqrt{100}} \sqrt{\frac{5000 - 100}{5000 - 1}} = 0.027$$

a. El peso total de la muestra estará entre 496 y 500 kg si el peso promedio de las cien varillas se encuentra entre 4.96 y 5.00 kg. Puesto que la muestra es mayor de 30 elementos se puede considerar como aproximadamente normal a la distribución muestral, y los valores estándar correspondientes a $\bar{X} = 4.96$ y a $\bar{X} = 5.00$ se obtienen mediante la transformación

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

es decir,

$$Z_1 = \frac{4.96 - 5.02}{0.027} = -2.22$$

$$Z_2 = \frac{5.00 - 5.02}{0.027} = -0.74$$

En la fig 4.1 se puede apreciar que

$$P[496 \leq X \leq 500] = P[-2.22 \leq Z \leq -0.74] =$$

$$= P[-2.22 \leq Z \leq 0] - P[-0.74 \leq Z \leq 0]$$

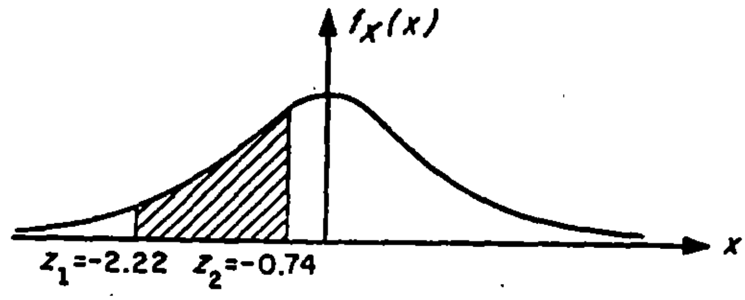


Fig 4.1 Distribución normal correspondiente al ejemplo

Recurriendo a la tabla de áreas bajo la curva normal estándar entre 0 y Z queda finalmente

$$P[496 \leq X \leq 500] = 0.4868 - 0.2704 = 0.2164$$

b. El peso total de la muestra excederá de 510 kg si el peso promedio de las cien varillas pasa de 5.10 kg.

Estandarizando dicho valor, queda

$$Z_3 = \frac{5.10 - 5.02}{0.027} = 2.96$$

Calculando el área bajo la curva normal a la derecha de este valor (fig 4.2), se tiene que

$$P[X \geq 510] = P[Z \geq 2.96] = P[Z > 0] - P[0 \leq Z \leq 2.96] =$$

$$= 0.5 - 0.4985 = 0.0015$$

6. Teoría estadística de la estimación

En la práctica profesional a menudo resulta necesario inferir información acerca de una población mediante el uso de muestras extraídas de ella; una parte básica de dicha inferencia consiste en *estimar* los valores de los parámetros de la población (media, variancia, etc.) a partir de las estadísticas correspondientes de la muestra, como se explica a continuación.

7. Estimadores puntuales. Clasificación

Si un estimador de un parámetro de la población consiste en un solo valor de una estadística, se le conoce como *estimador puntual* del parámetro.

Cuando la media de la distribución muestral de una estadística es igual al parámetro que se está estimando de la población, entonces la estadística se conoce como *estimador insesgado* del parámetro; si no sucede así, entonces se denomina *estimador sesgado*. Ambos estimadores son puntuales, y sus valores correspondientes se llaman estimaciones insesgadas o sesgadas, respectivamente. Dicho de otra manera, si S es una estadística cuya distribución muestral tiene media μ_S, y el parámetro correspondiente de la población es θ, se dice que S es un estimador insesgado de θ si

$$\mu_S = \theta$$

Por otra parte, si la estadística S_n de la muestra tiende a ser igual al parámetro θ de la población a medida que se hace más grande el tamaño de la muestra, entonces la estadística recibe el nombre de *estimador consistente* del parámetro.

Empleando símbolos, si

$$\lim_{n \to \infty} S_n = \theta$$

resulta que la estadística S_n es un estimador consistente. Por ejemplo, el promedio aritmético es un estimador insesgado y consistente de la media, y la variancia de la muestra es un estimador sesgado y consistente de la variancia de la población.

Si las distribuciones muestrales de varias estadísticas tienen el mismo valor de la media, se dice que la estadística que cuenta con la menor variancia es un *estimador eficiente* de dicha media, en tanto que las estadísticas restantes se conocen como *estimadores ineficientes* del parámetro.

Por ejemplo, las distribuciones muestrales del promedio aritmético y de la mediana cuentan con medias que son, en ambos casos, iguales a la media de la población. Sin embargo, la variancia de la distribución muestral del promedio aritmético es menor que la de la distribución de la mediana, por lo que el promedio aritmético obtenido de una muestra aleatoria proporciona un estimador eficiente de la media de la población, en tanto que la mediana obtenida de la muestra proporciona un estimador ineficiente de dicho parámetro.

8. Estimación de intervalos de confianza para los parámetros de una población

La estimación de un parámetro de una población mediante un par de números entre los cuales se encuentra, con cierta probabilidad, el valor de dicho parámetro, se llama estimación del intervalo del mismo.

Sea S una estadística obtenida de una muestra de tamaño n para estimar el valor del parámetro θ, y sea σ_S la desviación estándar (conocida o estimada) de su distribución muestral. La probabilidad, $1 - \alpha$, de que el valor de θ se localice en el intervalo de $S - z_c \sigma_S$ a $S + z_c \cdot \sigma_S$, donde z_c es una constante, se escribe en la forma

$$P[S - z_c \sigma_S \leqslant \theta \leqslant S + z_c \sigma_S] = 1 - \alpha$$

Si se fija el valor de $1 - \alpha$, se puede obtener el valor de z_c necesario para que se satisfaga la ecuación anterior, con lo cual queda definido el *intervalo de confianza* del parámetro θ, $(S - z_c \sigma_S, S + z_c \sigma_S)$, correspondiente al nivel de *confianza* $1 - \alpha$.

La constante z_c que fija el intervalo de confianza se conoce como *valor crítico*. Si la distribución de S es normal, el valor de z_c correspondiente a uno de α se obtiene de la tabla de áreas bajo la curva normal o de la tabla 8.1 siguiente.

TABLA 8.1 VALORES DE z_c PARA DISTINTOS NIVELES DE CONFIANZA

Nivel de confianza, en porcentaje	z_c
99.73	3.00
99.00	2.58
98.00	2.33
96.00	2.05
95.45	2.00
95.00	1.96
90.00	1.64
80.00	1.28
68.27	1.00
50.00	0.674

Ejemplo 8.1

Sea el promedio aritmético $\bar{X}$ una estadística con distribución normal. Las probabilidades o niveles de confianza de que $\mu_{\bar{X}}$ (o μ de la población) se encuentre localizada entre los límites $\bar{X} \pm \sigma_{\bar{X}}$, $\bar{X} \pm 2\ \sigma_{\bar{X}}$ y $\bar{X} \pm 3\ \sigma_{\bar{X}}$ son 68.26, 95.44 y 99.73%, respectivamente, obteniéndose dichos valores de la tabla de áreas bajo la curva normal. Lo anterior significa que el intervalo $\bar{X} \pm 3\ \sigma_{\bar{X}}$ contendrá a $\mu_{\bar{X}}$ en el 99.73 por ciento de las muestras de tamaño *n*, por lo que los intervalos de confianza de 68.26, 95.44 y 99.73 por ciento para estimar a μ son $(\bar{X} - \sigma_{\bar{X}}, \bar{X} + \sigma_{\bar{X}})$ $(\bar{X} - 2\ \sigma_{\bar{X}}, \bar{X} + 2\ \sigma_{\bar{X}})$ y $(\bar{X} - 3\ \sigma_{\bar{X}}, \bar{X} + 3\ \sigma_{\bar{X}})$, lo cual se aprecia en la *fig* 8.1 siguiente.

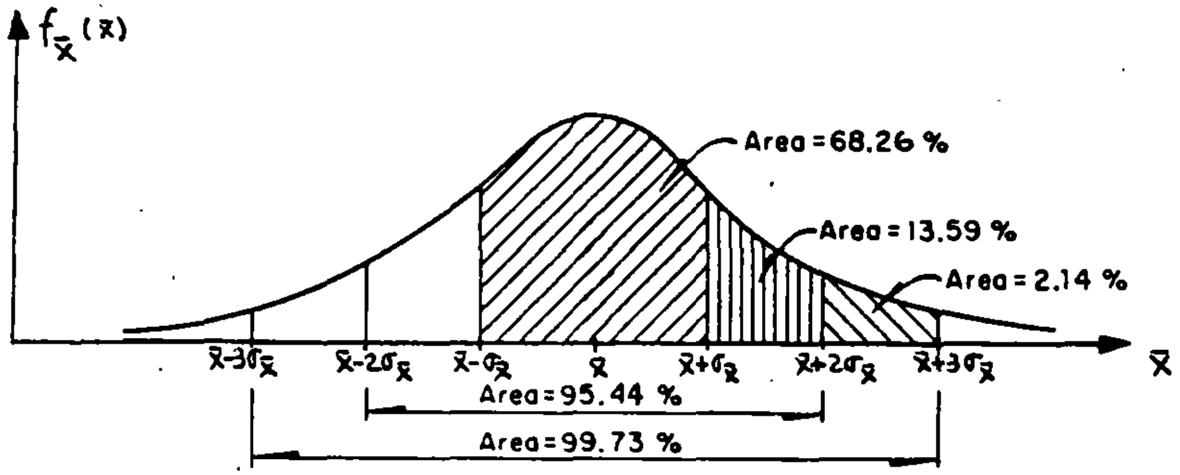


Fig 8.1

9. Estimación de intervalos de confianza para la media

Los límites de confianza para la media de una población con variable aleatoria X asociada están dados por

$$\bar{X} \pm z_c \, \sigma_{\bar{X}}$$

en donde z_c depende del nivel de confianza deseado. Si $\bar{X}$ tiene distribución normal, z_c puede obtenerse en forma directa de la tabla 8.1. Por ejemplo, los límites de confianza de 95 y 99 por ciento para estimar la media, μ, de la población son $\bar{X} \pm 1.96\sigma_{\bar{X}}$ y $\bar{X} \pm 2.58\ \sigma_{\bar{X}}$, respectivamente. Al obtener estos límites hay que usar el valor calculado de $\bar{X}$ para la muestra correspondiente.

Entonces, los límites de confianza para la media de la población quedan dados por

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{n}}$$

en caso de que el muestreo se haga a partir de una población infinita o de que se efectúe con remplazo a partir de una población finita, o por

$$\bar{X} = z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

si el muestreo es sin remplazo a partir de una población finita de tamaño N_p.

Ejemplo 9.1

Las mediciones de los diámetros de una muestra aleatoria de 100 tubos de albañal mostraron una media de 32 cm y una desviación estándar de 2 cm. Obténganse los límites de confianza de

a. 95 por ciento

b. 97 por ciento

para el diámetro medio de todos los tubos.

a. De la tabla 8.1, los límites de confianza del 95 por ciento son

$$\bar{X} \pm 1.96\sigma/\sqrt{n} = 32 \pm 1.96(2/\sqrt{100}) = 32 \pm 0.392 \text{ cm}$$

o sea 31.608 y 32.392, en donde se ha empleado el valor de S_X para estimar el de σ de la población, puesto que la muestra es suficientemente grande (mayor de 30 elementos). Esto significa que con una probabilidad de 95 por ciento, el valor de μ_X se encuentra entre 31.608 y 32.392 cm.

b. Si $Z = z_c$ es tal que el área bajo la curva normal a la derecha de z_c es el 1.5 por ciento del área total, entonces el área entre 0 y z_c es $0.5 - 0.015 = 0.485$, por lo que de la tabla de áreas bajo la curva normal se obtiene $z_c = 2.17$. Por lo tanto, los límites de confianza del 97 por ciento son:

$$\bar{X} \pm 2.17\sigma/\sqrt{n} = 32 \pm 2.17(2/\sqrt{100}) = 32 \pm 0.434 \text{ cm}$$

y el intervalo de confianza respectivo es (31.566 cm, 32.434 cm).

Ejemplo 9.2

Una muestra aleatoria de 50 calificaciones de cierto examen de admisión tiene un promedio aritmético de 72 puntos, con desviación estándar igual a 10. Si el examen se aplicó a 1018 personas, obtener

a. El intervalo de confianza del 95% para la media del total de calificaciones.

b. El tamaño de muestra necesario para que el error en la estimación de la media no exceda de 2 puntos, considerando el mismo nivel de confianza.

c. El nivel de confianza para el cual la media de la población sea 72 ± 1 puntos.

a. Si se estima a σ de la población con S_X de la muestra y se considera que la población es finita, los límites de confianza son, puesto que $\bar{X} = 72$, $Z_c = 1.96$, $S_X = 10$, $N_p = 1018$ y $n = 50$,

$$72 \pm 1.96 \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm 1.96\ (1.4142)\ (0.9755)$$

$$72 \pm 2.704$$

y el intervalo de confianza respectivo es

$$(69.296,\ \ 74.704)$$

b. Puesto que el error en la estimación de la media es, para población finita,

$$\text{Error en la estimación} = Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

en este caso se tendría

$$Z_c \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} < 2$$

o sea, para un nivel de confianza de 95%,

$$1.96 \frac{10}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

$$\frac{19.6}{\sqrt{n}} \sqrt{\frac{1018 - n}{1018 - 1}} < 2$$

Elevando al cuadrado la desigualdad, queda

$$\frac{384.16}{n} \quad \frac{1018 - n}{1017} < 4$$

o sea

$$87.85 < n$$

Por lo cual, se requieren al menos 88 elementos en la muestra para que el error en la estimación no exceda de 2 puntos, para $1 - \alpha = 0.95$.

c. Los límites de confianza son, en este caso

$$72 \pm Z_c \frac{10}{\sqrt{50}} \sqrt{\frac{1018 - 50}{1018 - 1}}$$

$$72 \pm Z_c \ (1.4142)(0.9755)$$

o sea

$$72 \pm 1.3795 \ Z_c$$

Puesto que se desea que el valor de la media sea 72 ± 1 puntos, se verifica que

$$1 = 1.3795 \ Z_c$$

Es decir

$$Z_c = \frac{1}{1.3795} = 0.725$$

El área bajo la curva normal estándar entre 0 y $Z_c = 0.725$ es, por interpolación lineal, igual a 0.2657. Por lo tanto, el nivel de confianza es igual al doble del área anterior, es decir, $2(0.2657) = 0.5314$ (o 53.14%), tal como se muestra en la fig 9.1.

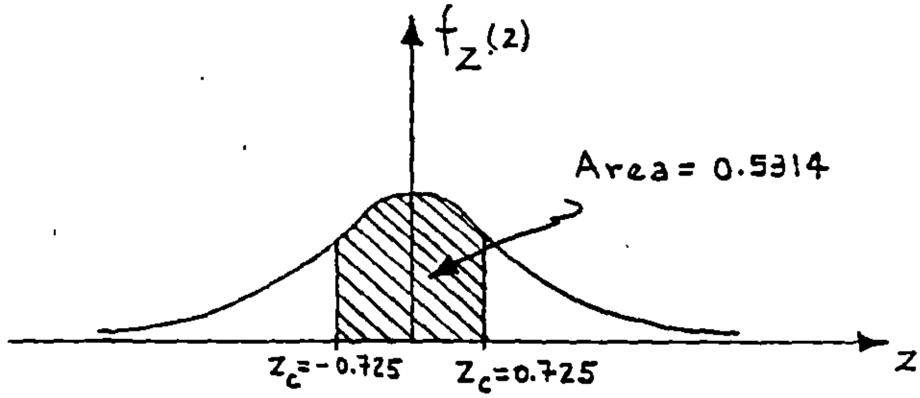


Fig 9.1

12. Errores de los tipos I y II. Nivel de significancia

En muchas ocasiones se presenta el caso de que se rechaza una hipótesis nula cuando en realidad debería ser aceptada; cuando esto sucede se dice que se ha cometido un *error de tipo I*. En otras ocasiones se acepta una hipótesis nula siendo en realidad falsa; en este caso se dice que se ha cometido un *error de tipo II*.

Al probar una hipótesis nula, a la máxima probabilidad con la que se está dispuesto a cometer un error del tipo I se le llama *nivel de significancia*, α, de la prueba, el cual dentro de la práctica se acostumbra establecer de 5 por ciento (0.05) o 10 por ciento (0.1). El complemento del nivel de significancia, $1 - \alpha$, se conoce como *nivel de confianza*.

Si, por ejemplo, al realizar una prueba de hipótesis se escoge un nivel de significancia de 10 por ciento, significa que existen 10 posibilidades en 100 de que se rechace ésta cuando debería ser aceptada; es decir, que se rechaza a un nivel de significancia del 10 por ciento, y que la probabilidad de que la decisión haya sido errónea es de 0.1.

13. Comportamiento de los errores tipos I y II

Supóngase que se trata de probar la hipótesis nula de que la media, μ_S, de la distribución muestral de la estadística S es μ_1, en contra de la hipótesis alternativa que establece que $\mu_S = \mu_2$, donde $\mu_2 > \mu_1$, es decir

$$H_0 : \mu_S = \mu_1$$

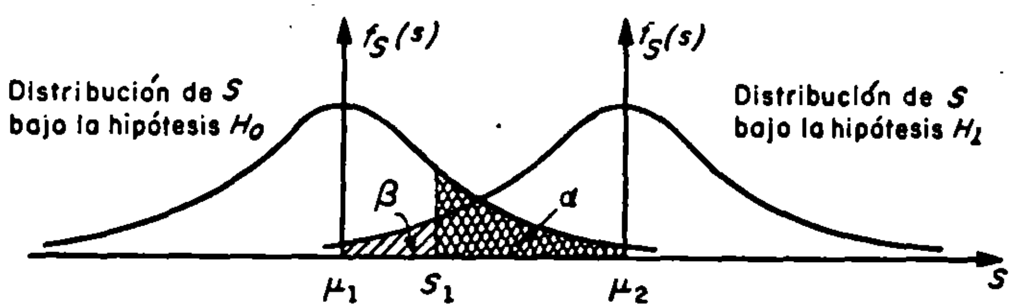
$$H_1 : \mu_S = \mu_2$$

En la fig 13.1 se muestra en forma gráfica la relación entre los errores tipos I y II en el caso en el que la regla de decisión para aceptar o rechazar H_0 es la siguiente:

> *Si el valor de la estadística S obtenido de una muestra excede de cierto valor crítico S_1, rechácese H_0; en caso contrario, acéptese.*

Es evidente que si H_0 es verdadera, entonces α (área con rayado doble) es la probabilidad de que $S > S_1$, o sea la de rechazar a H_0 siendo verdadera (error tipo I). Por otro lado, si H_1 es verdadera, entonces β (área con rayado sencillo) es la probabilidad de que $S < S_1$, o sea la de aceptar H_0 siendo falsa (error tipo II).

Obsérvese que si se aumenta el valor de S_1 se reduce la probabilidad α, pero se incrementa la β; lo contrario sucede si se disminuye el valor de S_1.



$P[S>S_1] = \alpha$ (error tipo I)

$P[S<S_1] = \beta$ (error tipo II)

Fig. 13.1 Probabilidades de los errores tipos I y II en pruebas de hipótesis.

En realidad, la única forma posible en la cual se pueden minimizar simultáneamente los errores de tipos I y II es aumentando el tamaño de la muestra, para hacer más "picudas" las distribuciones muestrales de la estadística bajo las hipótesis H_0 y H_1.

Al observar la fig 13.2 siguiente, es posible concluir que el tamaño de los errores I y II es menor para un tamaño de muestra igual a 100 que para un tamaño igual a 50, considerando la misma regla de decisión anterior.

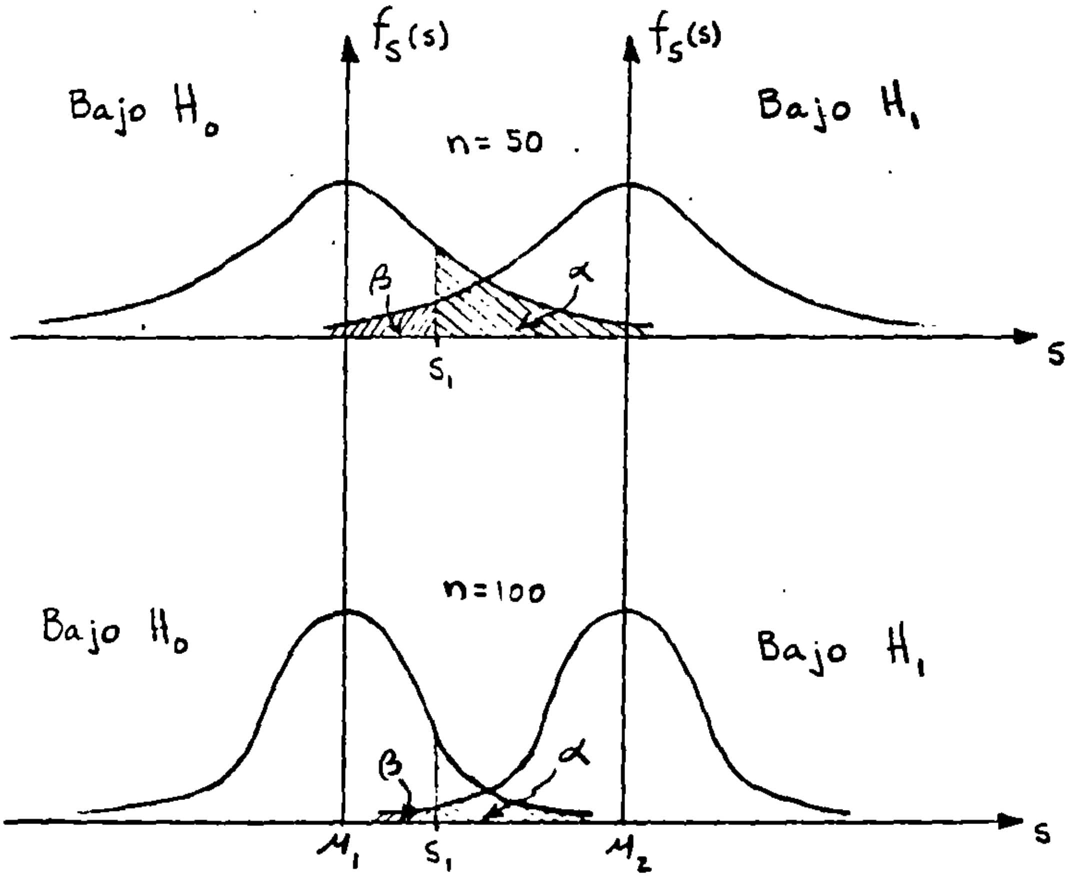


Fig 13.2

Sin embargo, esta técnica de reducción simultánea de ambos tipos de errores no siempre puede ponerse en práctica, debido a razones de costo, tiempo, etc.

14. Regiones críticas, de rechazo o de significancia. Regiones de aceptación.

Cuando una hipótesis nula no se acepta se dice que se *rechaza a un nivel de significancia del α por ciento*, o que el valor estandarizado de la estadística involucrada es *significativo a un nivel de significancia* α.

Al conjunto de los valores de la estadística en el que se rechaza la hipótesis nula se le denomina *región crítica, de rechazo, o de significancia.* Por el contrario, al conjunto de los valores de la estadística en que se acepta la hipótesis, se le llama *región de aceptación.*

Considérese que la distribución muestral de la estadística S es normal con desviación estándar σ_S, que la variable Z resulta de estandarizar a S, que la hipótesis nula, H_0, es que la media de S vale μ_S, y que la hipótesis alternativa H_1 es que dicha media es diferente de μ_S, es decir, que

$$Z = \frac{S - \mu_S}{\sigma_S}$$

H_0: media de la distribución muestral de $S = \mu_S$

H_1: media de la distribución muestral de $S \neq \mu_S$

Si se adopta la regla de decisión de aceptar la hipótesis H_0, si el valor de Z cae dentro del intervalo central que encierra al 99 por ciento del área de la distribución de probabilidades, entonces H_0 se aceptará en el caso en que

$$-2.58 \leq Z \leq 2.58$$

empleando la tabla de áreas bajo la curva normal estándar. Pero si el valor estandarizado de la estadística se encuentra fuera de dicho intervalo, se concluye que el evento puede ocurrir con probabilidad de 0.01 si la hipótesis H_0 es verdadera (área rayada total de la fig 14.1). En tal caso, el valor Z de la variable estándar difiere *significativamente* del que se podría esperar de acuerdo con la hipótesis nula, lo cual inclina a rechazarla a un nivel de confianza del 99 por ciento.

De lo anterior de deduce que el área total rayada de la fig 14.1 es el nivel de significancia α de la prueba, y representa la probabilidad de cometer un error del tipo I. Por ello, la región de aceptación de H_0 es $-2.58 \leq Z \leq 2.58$, y la de rechazo es $Z > 2.58$ y $Z < -2.58$.

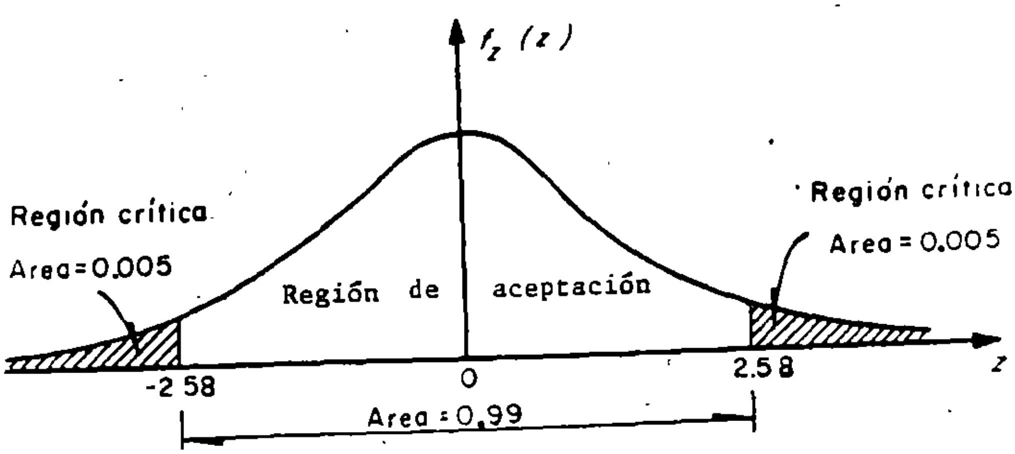


Fig 14.1 Región de significancia

En la tabla 14.1 se presentan los valores de la variable estandarizada, Z, que limitan las regiones de aceptación y de rechazo para el caso en el que la estadística involucrada en la prueba tenga distribución muestral normal. Cuando en alguna prueba de hipótesis se consideren niveles de significancia diferentes a los que aparecen en la tabla mencionada, resulta necesario emplear la de áreas bajo la curva normal estándar.

TABLA 14.1 VALORES CRITICOS DE z

Nivel de significancia, α	Valores de z para pruebas de una cola	Valores de z para pruebas de dos colas
0.1	−1.281 o 1.281	−1.645 y 1.645
0.05	−1.645 o 1.645	−1.960 y 1.960
0.01	−2.326 o 2.326	−2.575 y 2.575
0.005	−2.575 o 2.575	−2.810 y 2.810

15. Pruebas de una y de dos colas

En la prueba de hipótesis del ejemplo anterior, la región de rechazo de la hipótesis nula quedó en ambos extremos (colas) de la distribución muestral de la estadística involucrada en la prueba; a las pruebas de este tipo se les denomina *pruebas de dos colas*. Cuando la región de rechazo se encuentra solamente en un extremo de la distribución muestral en cuestión, se les llama *pruebas de una cola*.

Las pruebas de dos colas se presentan cuando en la hipótesis alternativa aparece el signo $\neq$ (diferente de), como en el siguiente caso

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S \neq \mu_1$$

en donde μ_S es la media de la estadística S, y μ_1 es un valor fijo.

En los casos

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S < \mu_1$$

y

$$H_0 : \mu_S = \mu_1$$

$$H_1 : \mu_S > \mu_1$$

las pruebas resultan de una cola.

16. Pruebas de hipótesis para la media

Para el caso de una población infinita (o finita en que se muestree con remplazo), cuya desviación estándar σ se conoce o se puede estimar adecuadamente, si se tiene que la estadística S obtenida de la muestra es el promedio aritmético, entonces la media de su distribución muestral es $\mu_S = \mu_{\bar{X}} = \mu$, y su desviación estándar es $\sigma_S = \sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ son, respectivamente, la media y la desviación estándar de la variable aleatoria X asociada a la población, y n es el tamaño de la muestra. En tal caso, si $\bar{X}$ tiene distribución normal, la variable estandarizada correspondiente será

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Para el caso de muestreo sin remplazo de población finita, se tiene que $\sigma_S = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$, en donde N_p es el tamaño de la población, por lo que la variable estandarizada será

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}}$$

En los dos casos anteriores, el valor de Z correspondiente al de $\bar{X}$ de la muestra es el que se debe comparar con el valor crítico correspondiente al nivel de significancia fijado, para así aceptar o no la hipótesis nula (prueba de una cola). Si se trata de una prueba de dos colas, el valor de Z se debe comparar con los dos valores críticos que corresponden al valor de α seleccionado. En cualquiera de los casos anteriores, el valor o valores críticos se pueden obtener de la tabla 14.1, para valores comunes de α.

Ejemplo 16.1.

Se sabe que el promedio de calificaciones de una muestra aleatoria de tamaño 100 de los estudiantes de tercer año de ingeniería civil es de 7.6, con una desviación estándar de 0.2. Si μ denota la media de la población de esas calificaciones, X, y si se supone que $\bar{X}$ tiene distribución normal, probar la hipótesis

$\mu = 7.65$ en contra de la hipótesis alternativa $\mu \neq 7.65$, usando un nivel de significancia de

a. 0.05

b. 0.01

Para la solución se deben considerar las hipótesis

$$H_0 : \mu = 7.65$$

$$H_1 : \mu \neq 7.65$$

Puesto que $\mu \neq 7.65$ incluye valores menores y mayores de 7.65, se trata de una prueba de dos colas.

La estadística bajo consideración es el promedio aritmético, $\bar{X}$, de la muestra, que se supone extraída de una población infinita. La distribución muestral de $\bar{X}$ tiene media $\mu_{\bar{X}} = \mu$, y desviación estándar $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, en donde μ y σ denotan, respectivamente, la media y la desviación estándar de la población de calificaciones.

Bajo la hipótesis H_0 (considerándola verdadera), se tiene que

$$\mu_{\bar{X}} = 7.65 = \mu$$

y utilizando la desviación estándar de la muestra como una estimación de σ, lo cual se supone razonable por tratarse de una muestra grande,

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.2/\sqrt{100} = 0.2/10 = 0.02$$

a. Para la prueba de dos colas a un nivel de significancia de 0.05 se establece la siguiente regla de decisión

Aceptar H_0 si el valor Z correspondiente al valor del promedio de la muestra se encuentra dentro del intervalo de -1.96 a 1.96 (tabla 14.1). En caso contrario, rechazar H_0.

Puesto que

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.6 - 7.65}{0.02} = -2.5$$

se encuentra fuera del rango de -1.96 a 1.96, se rechaza la hipótesis H_0 a un nivel de significancia de 0.05.

b. Si el nivel de significancia es 0.01, el intervalo de -1.96 a 1.96 de la regla de decisión del inciso a se remplaza por el de -2.58 a 2.58 tabla (14.1). Entonces, puesto que el valor muestral $Z = -2.5$ se encuentra dentro de este intervalo, se acepta la hipótesis H_0 a un nivel de significancia de 0.01.

Ejemplo 16.2

La resistencia media a la ruptura de cables de acero fabricados por la empresa X es de 905 kg. Una empresa consultora sugiere a X que cambie su proceso de manufactura, con lo cual incrementará la resistencia de sus cables. Se prueba el nuevo proceso, y se extrae una muestra aleatoria de 50 cables, obteniéndose para ellos una resistencia promedio de 926 kg, con desviación estándar igual a 42 kg. ¿Se puede considerar que el nuevo proceso realmente incrementa la resistencia, con un nivel de confianza de 99%?

En este caso, se debe plantear una prueba de hipótesis de una cola, para la cual

$$H_0 : \mu = 905 \text{ kg}$$

$$H_1 : \mu > 905 \text{ kg}$$

Puesto que el tamaño de la muestra es suficientemente grande, se puede aproximar la distribución muestral de la resistencia promedio mediante una normal, y estimar el valor de σ de la población mediante S_X de la muestra.

Considerando a la población infinita, y suponiendo como verdadera a H_0, se tiene que

$$\mu_{\overline{X}} = \mu = 905 \text{ kg}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{42}{\sqrt{50}} = 5.94$$

Para la prueba de una cola a un nivel de significancia de $\alpha = 1 - (1 - \alpha) = 1 - 0.99 = 0.01$, la regla de decisión es

Aceptar H_0 si el valor estandarizado de $\overline{X}$ de la muestra es menor o igual a $Z_c = 2.326$ (tabla 14.1); en caso contrario, rechazar H_0.

En virtud de que

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{926 - 905}{5.94} = 3.535$$

es mayor de 2.326, se rechaza H_0 a un nivel de significancia de 1%, concluyéndose que en realidad el nuevo proceso sí incrementa la resistencia de los cables.

3.4 Muestras pequeñas

Como ya se indicó, para muestras grandes ($n > 30$) las distribuciones muestrales de muchas estadísticas son aproximadamente normales, siendo tanto mejor la aproximación cuanto mayor es el tamaño de n. Sin embargo, cuando se trata de muestras en las que $n < 30$, llamadas *muestras pequeñas*, la aproximación no es suficientemente buena, por lo que resulta necesario introducir una teoría apropiada para su estudio.

Al estudio de las distribuciones muestrales de las estadísticas para muestras pequeñas se le llama *teoría estadística de las muestras pequeñas*. Existen al respecto tres distribuciones importantes: *Ji cuadrada*, *F* y *t de Student*.

3.4.1 Distribución *Ji* cuadrada (χ^2)

Hasta ahora solo se ha tratado la distribución muestral de la media. En esta sección se verá lo concerniente a la distribución muestral de la variancia, S_X^2, para muestras aleatorias extraídas de poblaciones normales. Puesto que S_X no puede ser negativa, es de esperarse que su distribución muestral no sea una curva normal, ya que esta tiene ordenadas mayores de cero en el lado de las abscisas negativas. De hecho, la estadística S_X^2 se puede estudiar si se consideran muestras aleatorias de tamaño n extraídas de una población normal con desviación estándar σ_X y si para cada muestra se calcula el valor de la estadística.

$$\chi^2 = \frac{n\, S_X^2}{\sigma^2} \qquad (3.14)$$

donde S_X^2 es la variancia de la muestra.

El número de grados de libertad, ν, de una estadística se define como

$$\nu = n - k$$

siendo n el tamaño de la muestra y k el número de parámetros de la población que deben estimarse a partir de ella.

La distribución muestral de la estadística χ^2 está dada por la ecuación

$$f(\chi^2) = U\, \chi^{\nu-2}\, e^{-\frac{1}{2}\chi^2}$$

en la que U es una constante que hace que el área total bajo la curva resulte igual a uno, y $\nu = n - 1$ es el número de grados de libertad. Esta distribución se llama *Ji cuadrada*, misma que se presenta en la fig 21 para distintos valores de ν.

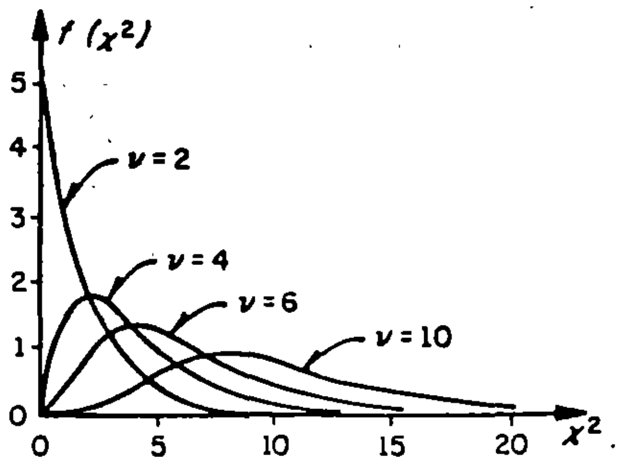


Fig 21. Distribución Ji cuadrada para distintos valores de ν

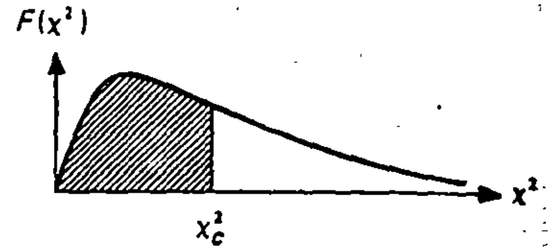


TABLA 8. VALORES CRITICOS χ^2_c

ν	$\chi^2_{.995}$	$\chi^2_{.99}$	$\chi^2_{.975}$	$\chi^2_{.95}$	$\chi^2_{.90}$	$\chi^2_{.75}$	$\chi^2_{.50}$	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.01}$	$\chi^2_{.005}$
1	7.88	6.63	5.02	3.84	2.71	1.32	.455	.102	.016	.0039	.0010	.0002	.0000
2	10.6	9.21	7.38	5.99	4.61	2.77	1.39	.575	.211	.103	.0506	.0201	.0100
3	12.8	11.3	9.35	7.81	6.25	4.11	2.37	1.21	.584	.352	.216	.115	.072
4	14.9	13.3	11.1	9.49	7.76	5.39	3.36	1.92	1.06	.711	.483	.297	.207
5	16.7	15.2	12.8	11.15	9.2	6.63	4.35	2.67	1.61	1.15	.831	.554	.413
6	18.5	16.8	14.4	12.6	10.6	7.84	5.35	3.45	2.20	1.64	1.24	.872	.676
7	20.3	18.5	16.0	14.1	12.0	9.04	6.35	4.25	2.83	2.18	1.69	1.24	.989
8	22.0	20.1	17.5	15.5	13.4	10.2	7.34	5.07	3.49	2.73	2.18	1.65	1.34
9	23.6	21.7	19.0	16.9	14.7	11.4	8.34	5.90	4.17	3.33	2.70	2.09	1.73
10	25.2	23.2	20.5	18.3	16.0	12.5	9.34	6.74	4.87	3.94	3.25	2.56	2.16
11	26.8	24.7	21.9	19.7	17.3	13.7	10.35	7.57	5.58	4.57	3.82	3.05	2.60
12	28.3	26.2	23.2	21.0	18.5	14.8	11.3	8.44	6.30	5.23	4.40	3.57	3.07
13	29.8	27.7	24.7	22.4	19.8	16.0	12.3	9.30	7.04	5.89	5.01	4.11	3.57
14	31.3	29.1	26.1	23.7	21.1	17.2	13.3	10.2	7.79	6.57	5.63	4.66	4.07
15	32.7	30.6	27.5	25.1	22.3	18.2	14.3	11.0	8.55	7.26	6.25	5.22	4.60
16	34.3	32.0	28.8	26.3	23.5	19.4	15.3	11.9	9.31	7.96	6.91	5.81	5.14
17	35.7	33.4	30.2	27.6	24.8	20.5	16.3	12.8	10.1	8.67	7.56	6.41	5.70
18	37.2	34.8	31.5	28.9	26.0	21.6	17.3	13.7	10.9	9.39	8.23	7.01	6.26
19	38.6	36.2	32.9	30.1	27.2	22.7	18.3	14.6	11.73	10.1	8.91	7.63	6.84
20	40.0	37.6	34.2	31.4	28.45	23.8	19.3	15.5	12.4	10.9	9.59	8.26	7.43
21	41.4	38.8	35.6	32.7	29.6	24.9	20.3	16.3	13.2	11.6	10.3	8.90	8.02
22	42.8	40.3	36.8	33.9	30.8	26.0	21.3	17.2	14.0	12.3	11.0	9.54	8.64
23	44.2	41.6	38.1	35.2	32.0	27.1	22.3	18.1	14.8	13.1	11.7	10.2	9.26
24	45.6	43.0	39.4	36.4	33.2	28.2	23.3	19.0	15.7	13.8	12.4	10.9	9.89
25	46.9	44.3	40.6	37.7	34.4	29.3	24.3	19.9	16.5	14.5	13.15	11.5	10.5
26	48.3	45.6	41.9	38.9	35.6	30.4	25.3	20.8	17.3	15.4	13.8	12.2	11.2
27	49.6	47.0	43.2	40.1	36.7	31.5	26.3	21.7	18.1	16.2	14.6	12.9	11.8
28	51.0	48.3	44.5	41.3	37.9	32.6	27.3	22.7	18.9	16.9	15.3	13.6	12.5
29	52.3	49.6	45.7	42.5	39.1	33.7	28.3	23.6	19.8	17.7	16.0	14.3	13.1
30	53.7	50.9	47.0	43.8	40.3	34.8	29.3	24.5	20.6	18.5	16.8	15.0	13.8
40	66.8	63.7	59.3	55.8	51.8	45.7	39.3	33.7	29.1	26.5	24.4	22.2	20.7
50	79.5	76.2	71.4	67.5	63.2	56.3	49.3	43.0	37.7	34.8	32.4	29.7	28.0
60	92.0	88.4	83.3	79.1	74.4	67.0	59.3	52.3	46.5	43.2	40.5	37.5	35.5
70	104.2	100.4	95.0	90.5	85.5	77.6	69.3	61.7	55.3	51.7	48.8	45.4	43.3
80	116.3	112.3	106.6	101.9	96.6	88.1	79.3	71.1	64.3	60.4	57.2	53.5	51.2
90	128.3	124.1	118.1	113.1	107.6	98.6	89.3	80.6	73.3	69.1	65.6	61.8	59.2
100	140.2	135.8	129.6	124.3	118.5	109.1	99.3	90.12	82.4	77.9	74.2	70.1	67.3

No obstante que la distribución *Ji* cuadrada solo se ha presentado en el estudio de las muestras pequeñas, cabe aclarar que es válida para aquellas mayores de 30 si la variable aleatoria involucrada tiene distribución normal.

3.4.1.1 Intervalo de confianza para la variancia

Tal como se hizo para la distribución normal, se pueden establecer intervalos de confianza para la variancia de la población en términos de la variancia de una muestra extraída de ella, a un nivel de confianza dado $1 - \alpha$, si se hace uso de los valores críticos χ^2_c de la tabla 8. Por lo tanto, un intervalo de confianza para la estadística χ^2, estaría dado por

$$\chi^2_{c'} < \frac{n\, S_X^2}{\sigma^2} < \chi^2_c$$

donde $\chi^2_{c'}$ y χ^2_c son los valores críticos para los cuales el $(1 - \alpha)/2$ por ciento del área se encuentra en los extremos izquierdo y derecho de la distribución, respectivamente.

Con base en lo anterior, se concluye que

$$\frac{n\, S_X^2}{\chi^2_c} < \sigma^2 < \frac{n\, S_X^2}{\chi^2_{c'}}$$

es un intervalo de confianza para estimar a σ^2 a un nivel de confianza $1 - \alpha$.

3.4.1.2 Prueba de hipótesis para la variancia

La prueba de hipótesis para la variancia de una población normal se efectúa calculando el valor de la estadística χ^2 y estableciendo las hipótesis H_0 y H_1 apropiadas, es decir, se adoptan reglas de decisión similares a las usadas para la estadística Z.

Ejemplo

La variancia del tiempo de elaboración de cierto producto es igual a 40 min; sin embargo, su proceso de manufactura se modifica y se toma una muestra de veinte tiempos, para la cual la variancia resulta ser igual a 62 min. ¿Es significativo el aumento del tiempo de elaboración a un nivel de significancia de

a) 0.05

b) 0.01?

Se debe decidir de entre las hipótesis

$$H_0 : \sigma^2 = 40 \text{ min}$$

$$H_1 : \sigma^2 > 40 \text{ min}$$

Suponiendo que la hipótesis nula es correcta, el valor de la estadística χ^2 para la muestra considerada es

$$\chi^2 = \frac{n S_X^2}{\sigma^2} = \frac{(20)\,(62)}{40} = 31$$

a) Como se trata de una prueba de una cola, la hipótesis H_0 se rechazaría si el valor de la estadística χ^2 fuera mayor que el de χ^2 para un nivel de significancia igual a 0.05, el cual, para $\nu = 20 - 1 = 19$ grados de libertad resulta ser 30.1 (tabla 8). Como $31 > 30.1$, H_0 se rechaza a un nivel de significancia de 0.05.

b) En este caso, el valor de χ^2 para un nivel de significancia de 0.01 y 19 grados de libertad es igual a 36.2. Puesto que $31 < 36.2$, se acepta H_0 a un nivel de significancia de 0.01.

3.4.3 Distribución t de Student

Si se consideran muestras de tamaño n extraídas de una población normal con media μ y variancia desconocida, para cada muestra se puede calcular la estadística T definida mediante la fórmula

$$T = \frac{\overline{X} - \mu}{S_X} \sqrt{n - 1} \tag{3.16}$$

donde $\overline{X}$ es el promedio y S_X la desviación estándar de la muestra.

La distribución muestral de T (fig 23) está dada por la ecuación

$$f(t) = \frac{U}{\left(1 + \frac{t^2}{\nu}\right)(\nu + 1)/2}$$

ojo: va de exponente

en la que U es una constante que hace que el área bajo la curva sea igual a uno, y $\nu = n - 1$ es el número de grados de libertad.

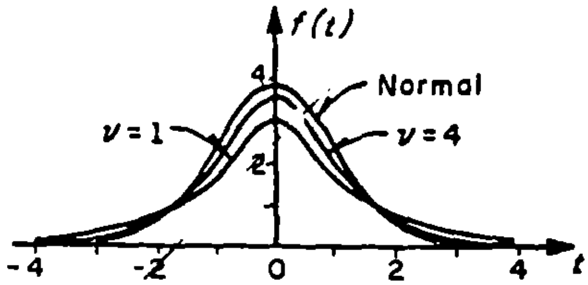


Fig 23. Distribución t de Student para distintos valores de ν

En la fig 23 se aprecia que conforme ν (o n, el tamaño de la muestra) aumenta, la distribución de $f(t)$ se aproxima a la distribución normal.

3.4.3.1 Límites e intervalos de confianza

De manera similar a como se hizo con la distribución normal, es posible estimar los límites de confianza de la media, μ, de una población mediante los *valores críticos*, t_c, de la distribución t, que dependen del tamaño de la muestra y del nivel de confianza deseado, encontrándose dichos valores en la tabla 10.

Así pues,

$$- t_c < \frac{\bar{X} - \mu}{S_X} \sqrt{n - 1} < t_c$$

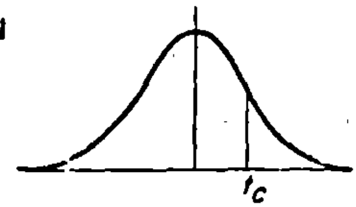
representa un intervalo de confianza para t, a partir del cual se puede estimar que μ se encuentra dentro del intervalo

$$\bar{X} - t_c \frac{\sigma_X}{\sqrt{n-1}} < \mu < \bar{X} + t_c \frac{\sigma_X}{\sqrt{n-1}}$$

En términos generales, los límites de confianza para la media de la población se representan como

$$\bar{X} \pm t_c \frac{\sigma_X}{\sqrt{n-1}}$$

TABLA 10. VALORES t_c PARA LA DISTRIBUCION t DE STUDENT



ν	$t_{.995}$	$t_{.99}$	$t_{.975}$	$t_{.95}$	$t_{.90}$	$t_{.80}$	$t_{.75}$	$t_{.70}$	$t_{.60}$	$t_{.55}$
1	63.66	31.82	12.71	6.31	3.07	1.376	1.000	.727	.325	.158
2	9.92	6.96	4.30	2.92	1.89	1.061	.816	.617	.289	.142
3	5.84	4.54	3.18	2.35	1.64	.978	.765	.584	.275	.138
4	4.60	3.75	2.78	2.13	1.53	.941	.741	.569	.271	.134
5	4.04	3.36	2.58	2.02	1.48	.920	.727	.560	.267	.132
6	3.71	3.14	2.45	1.94	1.44	.906	.718	.553	.265	.131
7	3.50	3.00	2.36	1.91	1.43	.896	.711	.549	.263	.130
8	3.36	2.90	2.31	1.86	1.40	.889	.706	.546	.262	.130
9	3.25	2.82	2.26	1.83	1.38	.883	.703	.543	.261	.129
10	3.17	2.76	2.23	1.81	1.37	.879	.700	.542	.260	.129
11	3.11	2.72	2.20	1.80	1.36	.876	.697	.540	.260	.129
12	3.06	2.68	2.18	1.78	1.36	.873	.695	.539	.259	.128
13	3.01	2.65	2.16	1.77	1.36	.871	.694	.538	.259	.128
14	2.98	2.62	2.14	1.76	1.34	.868	.693	.537	.258	.128
15	2.95	2.61	2.13	1.75	1.34	.866	.691	.536	.258	.128
16	2.92	2.58	2.12	1.75	1.34	.865	.690	.535	.258	.128
17	2.90	2.57	2.11	1.74	1.33	.863	.689	.534	.257	.128
18	2.88	2.55	2.10	1.73	1.33	.862	.688	.534	.257	[illegible]
19	2.87	2.54	2.09	1.73	1.33	.861	.688	.533	.257	.127
20	2.84	2.53	2.09	1.72	1.32	.860	.687	.533	.257	.127
21	2.83	2.52	2.08	1.72	1.32	.859	.686	.532	.256	.127
22	2.82	2.51	2.07	1.72	1.32	.858	.686	.532	.256	.127
23	2.81	2.50	2.07	1.71	1.32	.858	.685	.532	.256	.127
24	2.80	2.49	2.06	1.71	1.32	.857	.685	.531	.256	.127
25	2.79	2.48	2.06	1.71	1.32	.856	.684	.531	.256	[illegible]
26	2.78	2.48	2.05	1.71	1.32	.856	.684	.531	.256	[illegible]
27	2.77	2.47	2.05	1.71	1.31	.855	.683	.531	.256	[illegible]
28	2.76	2.47	2.05	1.70	1.31	.855	.683	.530	.256	.127
29	2.76	2.46	2.04	1.70	1.31	.854	.683	.530	.256	.127
30	2.75	2.46	2.04	1.70	1.30	.853	.683	.530	.256	.127
40	2.70	2.43	2.02	1.68	1.30	.851	.681	.529	.255	[illegible]
60	2.66	2.39	2.00	1.67	1.30	.848	.679	.528	[illegible]	[illegible]
120	2.62	2.36	1.98	1.66	1.29	.845	.677	.526	.254	[illegible]
∞	2.58	2.33	1.96	1.645	1.28	.842	.674	.524	.253	[illegible]

3.4.3.2 Pruebas de hipótesis

La prueba de hipótesis para la media de una población se puede efectuar con muestras pequeñas en forma análoga a la de muestras de tamaño mayor de 30 si en lugar de utilizar a la estadística Z se emplea la T. Entonces, si se consideran dos muestras aleatorias cuyos tamaños, desviaciones estándar y promedios son n_X, S_X, $\bar{X}$ y n_Y, S_Y, $\bar{Y}$, respectivamente, extraídas de poblaciones normales de igual variancia ($\sigma_X^2 = \sigma_Y^2$), se puede probar la hipótesis, H_0, de que las muestras provienen de una misma población, es decir, de que también sus medias son iguales, utilizando la estadística T definida por

$$T = \frac{\bar{X} - \bar{Y}}{s\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}}} \tag{3.17}$$

donde

$$s = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} \tag{3.18}$$

cuya distribución es la t de Student, con $\nu = n_X + n_Y - 2$ grados de libertad.

Ejemplo

Conforme al plan de desarrollo agrícola de una región, se probó un nuevo fertilizante para maíz. Para ello se escogieron 24 ha de terreno, aplicándose dicho producto a la mitad de ellas. El promedio de producción de maíz en la zona que se usó fertilizante fue de 5.3 ton, con una desviación estándar de 0.40 ton, en tanto que en la otra zona el promedio fue de 5.0 ton, con desviación estándar de 0.36 ton.

De acuerdo con los resultados, ¿se puede concluir que existe un aumento significativo en la producción de maíz al usar fertilizante, si se utiliza un nivel de significancia de

a) 0.01

b) 0.05?

Solución

Para probar la hipótesis de igualdad de medias es indispensable saber primero si las muestras provienen de dos poblaciones normales de igual variancia. En ese caso, si σ_X^2 y σ_Y^2 denotan a las variancias de la producción de maíz en la zona tratada y en la no tratada, respectivamente, se debe probar la hipótesis nula H_0: $\sigma_X^2 = \sigma_Y^2$ en contra de la hipótesis alternativa H_1: $\sigma_X^2 > \sigma_Y^2$ a los dos niveles de significancia establecidos.

El valor de la estadística F es, de la ec 3.15,

$$F = \frac{S_X^2}{S_Y^2} = \frac{(0.40)^2}{(0.36)^2} = 1.27$$

y el valor crítico de F (11, 11), obtenido de la tabla 9 mediante interpolación lineal, resulta 4.47. Por lo tanto, como $1.27 < 4.47$, se acepta la hipótesis nula a un nivel de significancia de 0.01.

El valor crítico de F (11, 11) a un nivel de significancia de 0.05 (ref 9) es 2.82, de ahí que como $1.27 < 2.82$, también se acepta la hipótesis H_0.

Con base en lo anterior, se debe decidir entre las hipótesis

H_0: $\mu_X = \mu_Y$ (la diferencia en los promedios se debe al azar)

H_1: $\mu_X > \mu_Y$ (el fertilizante mejora la producción)

Bajo la hipótesis H_0, se tiene que

$$s = \sqrt{\frac{n_X S_X^2 + n_Y S_Y^2}{n_X + n_Y - 2}} = \sqrt{\frac{12(0.40)^2 + 12(0.36)^2}{12 + 12 - 2}} = 0.397$$

por lo cual

$$t = \frac{5.3 - 5.0}{0.397\sqrt{\dfrac{1}{12} + \dfrac{1}{12}}} = 1.85$$

a) Puesto que se trata de una prueba de una cola a un nivel de significancia de 0.01, se rechaza la hipótesis H_0 si t es mayor que el valor crítico, t_c, correspondiente a dicho nivel, el cual para $\nu = n_X + n_Y - 2 = 12 + 12 - 2 = 22$ grados de libertad, se obtiene de la tabla 8 como $t_c = 2.51$. Como $t < t_c$, la hipótesis H_0 no se puede rechazar a un nivel de significancia de 0.01.

b) Si el nivel de significancia de la prueba es de 0.05, se rechaza H_0 si t es mayor que el valor t_c respectivo que para 22 grados de libertad es $t_c = 1.72$, por lo que de acuerdo con lo anterior, H_0 se rechaza a un nivel de significancia de 0.05.

4. TAMAÑO DE LA MUESTRA

Por: M en I Augusto Villarreal Aranda

INTRODUCCION

Dentro de un plan de muestreo, cuando ya se ha establecido la característica (o características) a estimar, así como el nivel de confianza y el grado de precisión requeridos, se debe decidir cuál debe ser el tamaño de la muestra o número de elementos a seleccionar por el procedimiento de muestreo que vaya a emplearse, en forma tal que los resultados que se obtengan no sean en exceso costosos o imprecisos.

Una vez que se ha fijado el error máximo admisible, que representa la precisión mínima que se exige tengan los resultados, así como el nivel de confianza $P_K = 1 - \alpha$, se requiere conocer además, en la forma más precisa posible, la variabilidad de la población, ya que cuanto más dispersos estén los valores de la variable asociada a ella más arriesgado será el utilizar una muestra de tamaño pequeño.

A continuación se expondrá el procedimiento para seleccionar el tamaño de muestra más adecuado en el caso del muestreo aleatorio simple o irrestrictamente aleatorio (sin remplazo). Más adelante se estudiarán los métodos para calcular el tamaño de la muestra para otros procedimientos de muestreo.

4.1 Tamaño de una muestra aleatoria simple (Medias)

En este caso se trata de estimar la media μ de una población con variable aleatoria asociada X mediante el empleo del promedio aritmético $\bar{X}$, obtenido de una muestra aleatoria de tamaño n con un error máximo admisible absoluto e y un nivel de confianza P_K. Es natural que a la probabilidad P_K le corresponderá un cierto valor de desviación K, obtenido a partir de la desigualdad de Chebyshev, o bien considerando a K como el número de desviacio- nes estándar para una distribución normal o para una t de Student. El procedimiento para obtener el tamaño de la muestra se fundamen ta en el hecho de que

$$P (\bar{X} - K\sigma_{\bar{X}} \leq \mu \leq \bar{X} + K\sigma_{\bar{X}}) = P_K = 1 - \alpha$$

sea que con probabilidad o nivel de confianza P_K se puede ase- ar que el valor de μ de una población se encuentra dentro del

$(1-\alpha)$ % de los intervalos formados a partir de muestras de tamaño n, de la forma siguiente

$$(\bar{X} - K\sigma_{\bar{X}} \; , \; \bar{X} + K\sigma_{\bar{X}})$$

Lo anterior implica que los límites de confianza del P_K % para estimar a μ son

$$\bar{X} \pm K\sigma_{\bar{X}}$$

es decir, que el error en la estimación del valor de μ es, en valor absoluto,

$$|\text{error en la estimación de } \mu| = K\sigma_{\bar{X}} \qquad (4.1)$$

Por lo tanto, es posible escribir

$$|\text{error máximo admisible}| = |\text{error en la estimación de } \mu| = e$$

4.1.1 Muestreo de una población finita

De la inferencia estadística, el valor de $\sigma_{\bar{X}}$, <u>la desviación estándar de la distribución muestral de $\bar{X}$ (o error estándar de $\bar{X}$)</u> cuando la población es finita es

$$\sigma_{\bar{X}} = \sqrt{\frac{N_p - n}{N_p - 1} \; \frac{\sigma_X^2}{n}}$$

pudiéndose escribir entonces

$$e = K\sigma_{\bar{X}} = K \sqrt{\frac{N_p - n}{N_p - 1} \; \frac{\sigma_X^2}{n}}$$

siendo K la desviación correspondiente al nivel de confianza P_k, N_p el tamaño de la población, σ_x^2 la variancia de esta última y n el tamaño de la muestra.

Puesto que se desea conocer el tamaño de la muestra, éste se puede obtener despejando de la ecuación anterior el valor de n. Para ello, se requiere elevar al cuadrado ambos miembros, es decir

$$e^2 = K^2 \frac{N_p - n}{N_p - 1} \frac{\sigma_x^2}{n}$$

$$e^2 = \frac{K^2 \sigma_x^2 N_p - K^2 \sigma_x^2 n}{(N_p - 1)\, n}$$

despejando a n:

$$ne^2 (N_p - 1) = K^2\sigma_x^2 N_p - K^2\sigma_x^2 n$$

$$ne^2 N_p - ne^2 = K^2\sigma_x^2 N_p - K^2\sigma_x^2 n$$

$$ne^2 N_p - ne^2 + K^2\sigma_x^2 n = k^2\sigma_x^2 N_p$$

$$n(e^2 N_p - e^2 + K^2\sigma_x^2) = k^2\sigma_x^2 N_p$$

$$\therefore n = \frac{K^2\sigma_x^2 N_p}{e^2 N_p - e^2 + K^2\sigma_x^2} \qquad (4.2)$$

La fórmula anterior permite obtener el tamaño de la muestra considerando conocidos K, e, N_p y σ_X^2. Puesto que el valor de σ_X^2 de la población usualmente se desconoce, se debe estimar previamente en forma adecuada considerando la información disponible de poblaciones semejantes a la que deberá muestrearse, o tomando una muestra preliminar suficientemente grande de dicha población.

Puesto que el tamaño de la muestra debe corresponder a un número entero positivo, se deberá asignar a n el valor entero más próximo por exceso al obtenido mediante la fórmula 4.2.

4.1.2 Muestreo de una población infinita

Cuando el muestreo se realiza a partir de una población infinita, el valor de $\sigma_{\bar{X}}$, la desviación estándar de la distribución muestral de $\bar{X}$, es

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$

en donde σ_X es la desviación estándar de la población y n el tamaño de la muestra.

considerando la ecuación 4.1, se puede escribir en este caso

$$|\text{error en la estimación de } \mu| = e = K\sigma_{\bar{X}} = K\,\frac{\sigma_X}{\sqrt{n}}$$

Para obtener el valor de n, se elevan al cuadrado ambos miembros de la expresión anterior, es decir,

$$e^2 = \frac{K^2\ \sigma_X^2}{n}$$

Por lo cual

$$n = \frac{K^2 \sigma_X^2}{e^2}$$

Para resaltar el hecho de que en este caso el tamaño de la muestra se obtiene a partir de una población infinita, en lugar de emplear n se puede emplear n_∞, es decir

$$n_\infty = \frac{K^2 \sigma_X^2}{e^2} \qquad (4.3)$$

Al igual que en el caso de una población finita, el tamaño de la muestra dado por la ec 4.3 debe corresponder a un número natural, por lo cual se debe aproximar por exceso al valor entero más cercano.

4.1.3 Comparación entre n y n_∞

Si se divide entre $N_p e^2$ el numerador y el denominador del miembro izquierdo de la ecuación 4.2, se obtiene

$$n = \frac{\frac{K^2\sigma_X^2 N_p}{N_p e^2}}{\frac{e^2 N_p - e^2 + K^2\sigma_X^2}{N_p e^2}} = \frac{\frac{K^2\sigma_X^2}{e^2}}{1 - \frac{1}{N_p} + \frac{K^2\sigma_X^2}{N_p e^2}}$$

$$n = \frac{\frac{K^2\sigma_X^2}{e^2}}{1 + \frac{1}{N_p}\left(\frac{K^2\sigma_X^2}{e^2} - 1\right)}$$

y, considerando el valor de n_∞ dado por la ec 4.3, se obtiene finalmente

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}\ (n_\infty - 1)} \qquad (4.4)$$

Como se puede apreciar de la ec 4.4, el valor de n es menor que el de n_∞, a menos que $N_p = \infty$.

4.1.4 Empleo adecuado de n y n_∞

Para una población finita, se definirá la <u>fracción de muestreo</u> como

$$\text{fracción de muestreo} = \text{fm} = \frac{n_\infty}{N_p}$$

siendo n_∞ el tamaño de la muestra calculada con la ec 4.3, y N_p el tamaño de la población.

Al obtener el tamaño de la muestra cuando se trata de una población finita, usualmente se acostumbra emplear la fórmula 4.3, que proporciona dicho tamaño para población infinita, y considerar como bueno dicho valor siempre que se cumpla la condición

$$\text{fm} \leqslant 0.05$$

Lo anterior quiere decir que en la práctica se calcula el valor de n_∞, y si n_∞ / N_p cumple con la condición mencionada, entonces se considera que n_∞ es una aproximación satisfactoria de n. Si la condición no se cumple, entonces se emplea la ec 4.4 para obtener el valor de n.

Es claro que tomando como tamaño de la muestra a n_∞ siempre se estará del lado más prudente, en el sentido de que se toma una muestra igual o mayor que la necesaria. Sin embargo, la eficiencia del diseño exige que el gasto y el tiempo de muestreo no sean superiores a los que haya que efectuar.

Ejemplo 4.1

Sea una población normal finita con variancia aproximadamente igual a 500. Se desea obtener una muestra aleatoria para estimar mediante $\bar{X}$ a la media poblacional μ_X, con error en la estimación no mayor de 10 y nivel de confianza igual a 90%. Obténgase el valor de n considerando que el tamaño de la población es igual a

a. 1000

b. 100

Solución

a. Puesto que $\sigma_X^2 = 500$, $e = 10$ y $1 - \alpha = 0.90$, tratándose de una población normal se tiene que

$$K = Z_{0.45} = 1.645$$

por lo cual

$$n_\infty = \frac{K^2\sigma_X^2}{e^2} = \frac{(1.645)^2\ (500)}{10^2}$$

$$= (2,706)\ (5) = 13.53$$

$$\therefore\ n_\infty = 14$$

En virtud de que en este caso

$$fm = \frac{n_\infty}{N_p} = \frac{14}{1000} = 0.014 < 0.05$$

se considera que n = 14.

b. En este caso

$$fm = \frac{14}{100} = 0.14 > 0.05$$

por lo cual se emplea la ec 4.4 para obtener el valor de n, es decir,

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}\ (n_\infty - 1)} = \frac{14}{1 + \frac{1}{100}\ (14 - 1)}$$

$$= \frac{14}{1 + \frac{13}{100}} = \frac{14}{1.13} = 12.389$$

$$\therefore\ n = 13$$

Ejemplo 4.2

Cierta universidad cuenta con 4726 estudiantes, y se desea conocer el rendimiento académico medio de todos ellos, en términos de una escala de calificación que va de cero a cien puntos. En estudios semejantes en otras universidades, se obtuvo que la desviación estándar de las calificaciones es aproximadamente igual a 7 puntos. Si el error en la estimación de la media de calificaciones no debe ser mayor de un punto en valor absoluto, y el nivel de confianza es igual a 99%, ¿cuál debe ser el tamaño de la muestra para realiaar la estimación?

Solución

En este caso, aproximando la distribución muestral de $\bar{X}$ mediante la distribución normal, se debe considerar que

$$P_K = 1 - \alpha = 0.99 \quad \therefore \quad K = Z_{0.495} = 2.58$$

$$\sigma_X^2 = (7)^2 = 49 \quad ; \quad e = 1 \text{ punto}$$

Por lo tanto,

$$n_\infty = \frac{Z_C^2\ \sigma_X^2}{e^2} = \frac{(2.58)^2\ (49)}{(1)^2}$$

$$= \frac{(6.656)\ (49)}{1} = 326.144$$

O sea $n_\infty = 327$

Puesto que

$$fm = \frac{n_\infty}{N_p} = \frac{327}{4726} = 0.0692 > 0.05$$

se procede a calcular n, es decir,

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}(n_\infty - 1)} = \frac{327}{1 + \frac{1}{4726}(327 - 1)}$$

$$= \frac{327}{1 + \frac{326}{4726}} = \frac{327}{1.069} = 305.89$$

$$\therefore n = 306$$

Ejemplo 4.3

Una muestra aleatoria de 14 observaciones de la altura alcanzada por cierto tipo de planta arrojó los siguientes datos:

Nº de elemento	Altura, X, en pulgadas
1	52.3
2	48.1
3	55.7
4	56.8
5	50.1
6	49.2
7	47.7
8	50.8
9	57.9
10	52.5
11	54.7
12	49.6
13	53.9
14	56.0

Obténgase el tamaño de muestra necesario para asegurar, con una probabilidad igual a 0.95, que el error en la estimación de la media de alturas de esta variedad de planta no sea mayor del 2.86%.

Solución

Se deben obtener primero los valores de $\bar{X}$ y S_X^2 de la muestra, con los cuales se estimarán los de μ_X y σ_X^2 de la población. Para ello, se dispone la información en la forma siguiente :

X_i	X_i^2
52.3	2735.3
48.1	2313.6
55.7	3102.5
56.8	3226.2
50.1	2510.0
49.2	2420.6
47.7	2275.3
50.8	2580.6
57.9	3352.4
52.5	2756.2
54.7	2992.1
49.6	2460.2
53.9	2905.2
56.0	3136.0
Σ 735.3	38766.2

Por lo tanto,

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{n} X_i = \frac{1}{14}(735.3) = 52.52 \text{ pulgadas}$$

$$S_X^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{14}(38766.2) - (52.52)^2$$

$$= 2769.01 - 2758.35 = 10{,}66 \text{ pulgadas}$$

Puesto que el error en la estimación de la media no debe ser mayor del 2.86%, y el estimador de μ_X es $\bar{X} = 52.52$, se tiene que

$$e = 52.52\ (0.0286) \cong 1.5 \text{ pulgadas}$$

Por otra parte, se desconoce el valor real de σ_X^2 de la población, además de que S_X^2, su estimador, se ha obtenido de una muestra menor de 30 elementos. Por lo tanto, la distribución teórica a la cual se debe aproximar la muestral debe ser la t de Student, siendo en este caso $K = t_c$. Sin embargo, puesto que en este caso se estima σ_X^2 mediante S_X^2 de la muestra, se debe tener presente que el error en la estimación de μ_X es

$$e = K\ \sigma_{\bar{X}} = t_c\ \sigma_{\bar{X}} = t_c\ \frac{S_X}{\sqrt{n-1}}$$

O sea, elevando al cuadrado

$$e^2 = t_c^2\ \frac{S_X^2}{n-1}$$

y, despejando a n,

$$n - 1 = \frac{t_c^2\ S_X^2}{e^2}$$

$$n = \frac{t_c^2\ S_X^2}{e^2} + 1$$

Por ser muestreo de población infinita, se puede escribir finalmente

$$n_\infty = \frac{t_c^2 \, S_x^2}{e^2} + 1 \qquad (4.5)$$

Ya que el valor de t_c depende del número de grados de libertad de la muestra υ, y este último depende del tamaño de la muestra (ya que $\upsilon = n - 1$), la fórmula anterior para obtener el valor de n_∞ contiene dos incógnitas. Por ello, se sigue el siguiente proceso iterativo para obtener el valor de n_∞:

1. Se hace $t_{0.025} = Z_{0.475}$, es decir

$$t_{0.025} = 1.96$$

Con dicho valor de t_c se obtiene

$$n_\infty = \frac{(1.96)^2 \; (10.66)}{(1.5)^2} + 1 = 18.2 + 1 = 19.3 \Rightarrow 20$$

De la tabla de la distribución t, se obtiene $t_{0.025} = 2.09$, para $\upsilon = 20-1 = 19$ grados de libertad.

2. Se toma ahora $t_{0.025} = 2.09$, y se obtiene

$$n_\infty = \frac{(2.09)^2 \; (10.66)}{(1.5)^2} + 1 = 20.7 + 1 = 21.7 \Rightarrow 22$$

De la tabla de la distribución t, se obtiene $t_{0.025} = 2.08$, para $\upsilon = 22-1 = 21$ grados de libertad.

3. Se toma ahora $t_{0.025} = 2.08$, y se obtiene

$$n_\infty = \frac{(2.08)^2\ (10.66)}{(1.5)^2} + 1 = 20.5 + 1 = 21.5 \Rightarrow 22$$

En este paso se obtiene un valor de n_∞ igual al del paso anterior, por lo que se puede considerar que el tamaño de muestra adecuado es igual a 22 plantas.

En este caso la población es infinita, por lo cual no s requiere hacer la corrección para población finita con la ec 4.4. Sin embargo, debe aclararse que es posible emplear la ec 4.5 para obtener n_∞ primero y, si la pob ción de la que se muestrea es finita, usar desp la 4.4 para obtener el valor de n corregido.

4.2 Tamaño de una muestra aleatoria simple (Totales)

Una característica o parámetro poblacional de gran interés es e total, que corresponde a la suma de todos los valores y_i que c tituyen la población, es decir,

$$Y = \sum_{i=1}^{N_p} y_i$$

en donde Y denota al total, y N_p es el número de elementos de misma.

Si se multiplica y divide por N_p el 2° miembro de la e ción rior, se obtiene

$$Y = \frac{N_p}{N_p} \sum_{i=1}^{N_p} y_i = N_p \, \mu_Y$$

Es decir, el total de una población es igual al tamaño de la misma multiplicado por la media correspondiente.

Cono estimador puntual del total de la población se puede tomar el de la estadística

$$\hat{Y} = N_p \, \bar{Y}$$

en donde $\bar{Y}$ es el promedio aritmético de la muestra, y $\hat{Y}$ un estimador insesgado en virtud de que

$$E\{\hat{Y}\} = E\{N_p \bar{Y}\} = N_p E\{\bar{Y}\} = N_p \, \mu_Y = Y$$

Por otra parte, la variancia de la distribución muestral de $\hat{Y}$ es

$$\sigma^2_{\hat{Y}} = \sigma^2_{N_p \bar{Y}} = \mathrm{Var}\,\{N_p \bar{Y}\} = N_p^2 \,\mathrm{Var}\,\{\bar{Y}\} = N_p^2 \, \sigma^2_{\bar{Y}}$$

y la desviación estándar es

$$\sigma_{\hat{Y}} = \sigma_{N_p \bar{Y}} = N_p \sigma_{\bar{Y}} = N_p \frac{\sigma_{Y.}}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

De igual manera a como se hizo para las medias, el valor del tamaño de muestra para estimar el total con un nivel de confianza y un error absoluto dados, se obtiene en la forma siguiente

$$e = K\,\sigma_{\hat{Y}} = K\,N_p\,\frac{\sigma_Y}{\sqrt{n}}\sqrt{\frac{N_p - n}{N_p - 1}}$$

Elevando al cuadrado y realizando operaciones algebraicas,

$$e^2 = K^2\,N_p^2\,\frac{\sigma_Y^2}{n}\,\frac{N_p - n}{N_p - 1}$$

$$e^2 = \frac{K^2\,N_p^3\,\sigma_Y^2 - K^2\,N_p^2\,\sigma_Y^2\,n}{n(N_p - 1)}$$

$$n\left(1 + \frac{K^2\,N_p^2\,\sigma_Y^2}{e^2(N_p - 1)}\right) = \frac{K^2\,N_p^3\,\sigma_Y^2}{e^2(N_p - 1)}$$

O sea

$$n = \frac{K^2\,N_p^3\,\sigma_Y^2}{e^2(N_p - 1) + K^2\,N_p^2\,\sigma_Y^2}$$

Dividiendo el numerador y denominador de la expresión anterior entre $N_p e^2$, se obtiene

$$n = \frac{\dfrac{K^2 N_p^3\,\sigma_Y^2}{N_p e^2}}{\dfrac{e^2 N_p - e^2 + K^2 N_p^{\,2}\,\sigma_Y^2}{N_p e^2}}$$

$$= \frac{N_p^2\,\dfrac{K^2\sigma_Y^2}{e^2}}{1 - \dfrac{1}{N_p} + \dfrac{N_p^2}{N_p}\,\dfrac{K^2\,\sigma_Y^2}{e^2}}$$

Considerando la ec 4.3, queda finalmente

$$n = \frac{N_p^2\,n_\infty}{1 + \frac{1}{N_p}\,(N_p^2\,n_\infty - 1)} \qquad (4.6)$$

Ejemplo 4.4

Con el fin de hacer una solicitud al Gobierno, se recogieron firmas de habitantes de una ciudad en 676 hojas. Cada hoja tenía espacio suficiente para 42 firmas, pero en varias hojas se recolectó un número menor de ellas. Para obtener una estimación del total de firmas, se contó el número de firmas por hoja en una muestra aleatoria de 50 hojas, obteniéndose los datos que aparecen en la tabla siguiente:

Número de firmas, y_i	Número de hojas, f_i
42	23
41	4
36	1
32	1
29	1
27	2
23	1
19	1
16	2
15	2
14	1
11	1
10	1
9	1
7	1
6	3
5	2
4	1
3	1

Obtener el tamaño de muestra necesario para estimar el valor del total de firmas con un error absoluto igual al 5%, considerando un nivel de confianza igual a 95%.

Solución : Por conveniencia para realizar los cálculos, se dispone la información en la forma siguiente:

y_i	f_i	y_i^2	$f_i y_i$	$f_i y_i^2$
42	23	1764	966	40572
41	4	1681	164	6724
36	1	1296	36	1296
32	1	1024	32	1024
29	1	841	29	841
27	2	729	54	1458
23	1	529	23	529
19	1	361	19	361
16	2	256	32	512
15	2	225	30	450
14	1	196	14	196
11	1	121	11	121
10	1	100	10	100
9	1	81	9	81
7	1	49	7	49
6	3	36	18	108
5	2	25	10	50
4	1	16	4	16
3	1	9	3	9
Σ	50		1471	54497

$$\bar{Y} = \frac{1}{50} \sum_{i=1}^{19} f_i \, y_i = \frac{1471}{50} = 29.42$$

$$S_Y^2 = \frac{1}{50} \sum_{i=1}^{19} f_i \, y_i^2 - (\bar{Y})^2 = \frac{54497}{50} - (29.42)^2 =$$

$$= 1089.94 - 865.44 = 224.5$$

Entonces

$$\hat{Y} = N_p \, \bar{Y} = 676 \times 29.42 = 19888 \text{ firmas}$$

y, puesto que el error absoluto debe ser igual al 5%, se tendría

$$e = (0.05)\ (19888) = 995$$

Por otra parte, el tamaño inicial de muestra igual a 50 permite suponer que la estimación de σ_Y^2 de la población es suficientemente buena con S_Y^2 , y que la distribución muestral de totales puede aproximarse mediante la normal. Por lo tanto,

$$K = Z_{0.475} = 1.96$$

$$N_p = 676$$

$$\sigma_Y^2 \doteq S_Y^2 = 224.5$$

$$N_p^2 n_\infty = N_p^2 \frac{K^2 \sigma_Y^2}{e^2} = \frac{(676)^2\ (1.96)^2\ (224.5)}{(995)^2} = 397.9$$

$$n = \frac{N_p^2 n_\infty}{1 + \frac{1}{N_p}(N_p^2 n_\infty - 1)} = \frac{397.9}{1 + \frac{1}{676}(397.9 - 1)}$$

$$= \frac{397.9}{1 + 0.58} = \frac{397.9}{1.58} = 251.83$$

$$\therefore \quad n = 252 \text{ hojas}$$

4.3 Tamaño de una muestra aleatoria simple (Proporciones)

4.3.1 Antecedentes

Supóngase una población binomial de tamaño N_p tal que cada uno de sus elementos únicamente puede estar en una de dos clases: A o B (buenos o malos, negros o blancos, grandes o chicos, etc). La proporción de elementos de la población que están en la clase A es

$$P = \frac{A}{N_p}$$

y la proporción de elemtos que están en B es

$$Q = \frac{B}{N_p}$$

por lo cual

$$P + Q = \frac{A}{N_p} + \frac{B}{N_p} = 1 \quad ; \quad (A + B = N_p)$$

Si a todos los elementos X_i de la población que están en A se les asigna el valor 1 y a los de B el 0, se obtiene

$$P = \frac{A}{N_p} = \frac{\sum_{i=1}^{N_p} X_i}{N_p} = \mu_X$$

Es decir, la proporción puede considerarse un caso particular de la media cuando los elementos de la población son unos y ceros.

La variancia es

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} (X_i - P)^2$$

o sea

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} X_i^2 - P^2$$

Sin embargo, como X_i sólo puede ser igual a uno o cero, se tiene que $X_i = X_i^2$, por lo cual

$$\sigma_X^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} X_i - P^2 = P - P^2 = P(1 - P) = PQ$$

En virtud de lo anterior, si se muestrea sin remplazo y con tamaño n de una población binomial finita, para estimar la proporción de elementos con cierta característica, se obtienen, considerando que la proporción se puede calcular como una media, los siguientes parámetros de la distribución muestral de proporciones

$$\mu_p = P$$

$$\sigma_p = \frac{\sigma_X}{\sqrt{n}} \sqrt{\frac{N_p - n}{N_p - 1}} = \sqrt{\frac{PQ}{n}} \cdot \sqrt{\frac{N_p - n}{N_p - 1}}$$

Si la población es infinita, se obtiene

$$\mu_p = P$$

$$\sigma_p = \frac{\sigma_x}{\sqrt{n}} = \sqrt{\frac{PQ}{n}}$$

estimándose P en ambos casos con el valor de p de la muestra, si se desconoce P de la población.

En la práctica se considera que la distribución muestral de proporciones es aproximadamente igual a la normal para tamaños de muestra mayores o iguales a 30 elementos.

4.3.2 Obtención del tamaño de la muestra

Aprovechando el hecho de que la proporción se puede calcular como una media simple, las ecs 4.3 y 4.4 se pueden emplear en este caso para obtener el tamaño de la muestra haciendo $\sigma_x^2 = PQ$. Entonces,

$$n_\infty = \frac{K^2\ PQ}{e^2} \tag{4.7}$$

para muestreo de población infinita, y

$$n = \frac{n_\infty}{1 + \frac{1}{N_p}(n_\infty - 1)}$$

para muestreo de población finita con tamaño N_p.

Usualmente se calcula primero el valor de n_∞, y si la fracción de muestreo es mayor de 0.05, se calcula a continuación el valor de n.

Ejemplo 4.5

En una colonia con 4000 casas se desea estimar el porcentaje de inquilinos que son a la vez propietarios de su casa, con un <u>error estándar</u> en la estimación no mayor del 1%. Se supone, de estudios semejantes, que el porcentaje real de inquilinos-propietarios se acerca al 10%. ¿Cuántas casas se deben muestrear para que se satisfaga la condición establecida?

Solución

El error estándar en la estimación de P de la población es

$$\sigma_p = \sqrt{\frac{PQ}{n}} \sqrt{\frac{N_p - n}{N_p - 1}}$$

y no debe ser mayor en este caso del 1%. Por lo tanto, siendo $N_p = 4000$, $P = 0.1$ y $Q = 1 - P = 0.9$, se obtiene

$$0.01 = \sqrt{\frac{(0.1)\ (0.9)}{n}} \sqrt{\frac{4000 - n}{4000 - 1}}$$

Elevando al cuadrado y realizando operaciones algebraicas

$$0.0001 = \frac{0.09}{n} \frac{4000 - n}{3999}$$

$$0.0001 = \frac{360 - 0.09\ n}{3999\ n}$$

$$0.3999\ n = 360 - 0.09\ n$$

$$n(0.3999 + 0.09) = 360$$

$$n = \frac{360}{0.4899} = 734.84$$

$$\therefore\ n = 735 \text{ casas}$$

Ejemplo 4.6

En un estudio antropológico para estimar el porcentaje de habitantes de una isla con sangre del grupo O, se obtuvo una muestra aleatoria de 50 isleños, en la cual 22 de ellos pertenecen al grupo sanguíneo mencionado. Si en la isla habitan 3208 gentes, ¿cuál debe ser el tamaño de muestra mínimo para estimar con un error absoluto del 5% el valor real de P, suponiendo que el nivel de confianza es del 95%?

Solución

En este caso la proporción de la muestra es

$$p = \frac{22}{50} = 0.44$$

o sea

$$q = 1 - p = 1 - 0.44 = 0.56$$

Considerando que la muestra inicial es suficientemente grande, se aproxima mediante la distribución normal, obteniéndose

$$K = Z_{0.475} = 1.96$$

por lo cual

$$n_{\infty}' = \frac{K^2\ PQ}{e^2} = \frac{K^2\ pq}{e^2} = \frac{(1.96)^2\ (0.44)\ (0.50)}{(0.05\ )^2}$$

$$= \frac{0.84515}{0.0025} = 338.06$$

$$\therefore\ n_{\infty} = 339$$

Como

$$fm = \frac{n_{\infty}}{N_p} = \frac{339}{3208} = 0.106 > 0.05$$

se corrige el valor anterior, obteniéndose finalmente

$$n = \frac{n_{\infty}}{1 + \frac{1}{N_p}\ (n_{\infty} - 1)} = \frac{339}{1 + \frac{1}{3208}\ (339-1)}$$

$$= \frac{339}{1.105} = 306.787$$

$\therefore$ **n = 307 habitantes**

FACULTAD DE INGENIERIA U.N.A.M.
DIVISION DE EDUCACION CONTINUA

DIPLOMADO DE CALIDAD EN LA CONTRUCION

MODULO I: METODOS ESTADISTICOS PARA CONTROL DE CALIDAD

TEMA: MUESTREO DE INSPECCION

EXPOSITOR: M. EN I. AUGUSTO VILLARREAL A

Palacio de Minería Calle de Tacuba 5 Primer piso [illegible] 06000 México, D.F. APDO [illegible]
Teléfonos: 512-8955 512-5121 521-73[illegible] [illegible] 510-0573 512-5121 521-4020 [illegible]

MUESTREO DE INSPECCION

Por: M en I Augusto Villarreal Aranda*

1. Introducción

El muestreo de inspección (o de aceptación) se define como el conjunto de todas las acciones que realiza el receptor de producto "terminado" para asegurar la calidad de éste, después de recibirlo del productor.

Este tipo de muestreo puede ser aplicado por un consumidor a los productos que recibe de un vendedor, por un departamento de inspección de producto terminado a los productos recibidos de los departamentos de producción, etc, es decir, se aplica en aquellas ocasiones en que un número grande de unidades producidas se presenta para inspección en forma de lotes, y en donde la forma

* *Secretario Académico*, División de Estudios Superiores, Facultad de Ingeniería, UNAM y *Profesor investigador*, Instituto de Ingeniería, UNAM

lógica de realizar esa tarea es mediante el empleo de la técnica que usa atributos (sirve, no sirve, o pasa, no pasa), con el fin de evitar la tan costosa y tardada inspección al 100%.

Generalmente, con la inspección de lote por lote del producto terminado, existe el acuerdo entre productor y receptor en que

a. los lotes aceptados por el plan de muestreo que se emplee serán aceptados por el receptor como buenos a excepción de aquellas unidades detectadas como defectuosas en todos los lotes durante el proceso de muestreo, las cuales serán reemplazadas por unidades buenas por el productor.

b. los lotes rechazados por el plan de muestreo le serán devueltos al productor para su rectificación.

Sin embargo, existen algunas variantes sobre el acuerdo mencionado. Por ejemplo, algunos receptores de producto terminado emplean la opción de inspeccionar al 100% los lotes rechazados para eliminar los elementos defectuosos, y trasladar el costo de esa operación al productor. Lo anterior se realiza con frecuencia cuando el receptor tiene urgencia de emplear las unidades que recibe del productor. En última instancia el objetivo que se persigue es responsabilizar al productor por la deficiente calidad de un producto terminado.

Para determinar la calidad de un lote, es factible seleccionar una, dos o múltiples muestras aleatorias del mismo, lo cual conduce a considerar planes de muestreo simples, dobles, o múltiples para aceptarlo o rechazarlo. La explicación de cómo y cuando se emplean estos tipos de muestreo se discutirá en esta parte del curso.

2. Plan de muestreo simple

Cómo se dijo anteriormente, el muestreo de aceptación se aplica a las producciones en masa cuando un *productor* abastece de lotes de artículos a un *receptor*. En situaciones como ésta, se debe decidir individualmente sobre la aceptación o rechazo de cada lote.

En este caso particular, la decisión que se toma se basa en el resultado que se obtiene al inspeccionar una muestra de tamaño "n" que se toma de un lote de "N" artículos, de la cual se determina el número de defectuosos, "X", esto es, de artículos que no cumplen las especificaciones nominales (tamaño, color, resistencia, etc.)

Si el número "X" de artículos defectuosos en la muestra es menor o igual que un número especificado "c" *menor* que "n", se acepta el lote; si el número de defectuosos es mayor que "c", se rechaza. A "c" se le llama el número tolerable de artículos defectuosos o *número de aceptación*. Por lo tanto, las alternativas son

$X \leq c$ se acepta el lote

$X > c$ se rechaza el lote

Resulta evidente que el productor y el receptor deben quedar de acuerdo en cierto *plan de muestreo*, es decir, en cierto tamaño n de muestra y cierto número de aceptación c. Puesto que en este caso el acuerdo se basa en la extracción de una muestra aleatoria única del lote de N artículos, el plan de muestreo a emplearse se denomina *plan de muestreo simple*.

2.1 Probabilidad de aceptación de un lote

Supóngase que si $X \leq c$ se acepta un lote, es decir, ocurre el evento A = {el número de artículos defectuosos en la muestra extraída del lote es menor o igual que el número de aceptación}. En este caso, la probabilidad de dicho evento no depende únicamente del tamaño n de la muestra y del número de aceptación c, sino también del número total de artículos defectuosos que se encuentran en el lote,"M". Si se supone además que el muestreo se realiza sin remplazo, la probabilidad de dicho evento es hipergeométrica, es decir

$$P(A) = P\{X \leq c\} \sum_{X=0}^{c} \frac{C_X^M \; C_{n-X}^{N-M}}{C_n^N} \qquad (2.1)$$

Si no hay artículos defectuosos en el lote, entonces $M = 0$, y el único valor posible que puede asumir X es también 0, por lo cual

$$P(A) = P\{X \leq c\} = \frac{C_o^o \; C_n^N}{C_n^N} = 1$$

Es decir, la probabilidad de aceptar un lote en el cual no hay elementos defectuosos es igual a la unidad.

Si todos los artículos en un lote son defectuosos, entonces $M = N$, y el valor de X debe ser igual a n, por lo que

$$P(A) = P\{X \leq c\} = P(\emptyset) = 0$$

en virtud de que la condición inicial es que $c < n$. Lo anterior indica que la probabilidad de aceptar un lote en el cual todos los artículos son defectuosos es nula.

Conviene hacer notar también que si se mantienen fijos el tamaño de la muestra y el número de aceptación al incrementarse el valor de M, el número de artículos defectuosos en un lote, decrece la probabilidad $P(A)$ de aceptación de este último.

Ejemplo 2.1

Considérese un plan de muestreo simple para el cual $N = 10$, $c = 0$ y $n = 5$. Obténganse los valores de $P(A)$ cuando

a. $M = 1$

b. $M = 3$

Solución

a. En este caso, la probabilidad de aceptación es

$$P(A) = P\{X = 0\} = \frac{C_0^1 \; C_{5-0}^{10-1}}{C_5^{10}} =$$

$$= \frac{\dfrac{1!}{0!(1-0)!} \; \dfrac{9!}{5!(9-5)}}{\dfrac{10!}{5!(10-5)!}} = \frac{\dfrac{9\times8\times7\times6}{4\times3\times2\times1}}{\dfrac{10\times9\times8\times7\times6}{5\times4\times3\times2\times1}} = 0.5$$

b. Para este caso, se obtiene

$$P(A) = P\{X \leq 0\} = P\{X = 0\} = \frac{C_0^3 \; C_{5-0}^{10-3}}{C_5^{10}} =$$

$$= \frac{\dfrac{3!}{0!(3-0)!} \; \dfrac{7!}{5!(7-5)!}}{\dfrac{10!}{5!(10-5)!}} = \frac{\dfrac{7\times6}{2\times1}}{\dfrac{10\times9\times8\times7\times6}{5\times4\times3\times2\times1}} = 0.0833$$

Lo anterior indica que un plan de muestreo simple para el cual se mantenga fijo el tamaño de la muestra, aun cuando se incremente el número de elementos defectuosos en los lotes, o el número total de elementos en estos últimos, proporciona buena protección en contra de la aceptación errónea de lotes malos.

2.2 Curva característica de operación

Dentro de un plan de muestreo simple, al considerar un número fijo de aceptación, c, y cuando se obtiene una muestra aleatoria de n artículos de un lote para saber si éste se acepta o no, es evidente que se desconoce el número total de artículos defectuosos, M, dentro del mismo. Para que este número se pudiera conocer en forma precisa, se requeriría haber realizado una inspección al 100% en el lote, pero entonces no tendría caso el considerar un plan de muestreo simple.

Por lo anterior, para realizar el cálculo de la probabilidad de aceptación de un lote determinado cuando se desconoce el valor de M, se debe introducir una modificación dentro de la fórmula 2.1. Para ello, considérese que si se divide el número de elementos defectuosos entre el total de elementos para un lote determinado, se obtiene la *fracción de defectuosos*

$$p = \frac{M}{N} \tag{2.2}$$

en el lote. Si p se multiplica por 100, se obtiene el *porcentaje de elementos defectuosos* en dicho lote.

Puesto que M puede tomar dentro de un lote de tamaño N cualquiera de los N + 1 valores 0,1,2,3,...,N-1,N, p puede asumir entonces los N + 1 valores, 1/N, 2/N, 3/N,..., $^{N-1}$/N, 1. Por lo tanto, la probabilidad de aceptación P (A) únicamente se puede definir para los valores mencionados de p.

Si en la ec 2.2 se despeja el valor de M, se obtiene

$$M + Np$$

en forma tal que la ec 2.1 se puede escribir como

$$P\ (A;\ p) = P\ \{X \leq c\} = \sum_{X=0}^{c} \frac{C_X^{Np}\ \ C_{n-X}^{N-Np}}{C_n^N} \tag{2.3}$$

siendo las probabilidades así obtenidas hipergeométricas.

Si se mantienen fijos los valores de n y c, se pueden graficar las probabilidades de aceptación de un lote en función de los valores de la fracción de elementos defectuosos en el mismo, es decir, de los valores de p. Dicha gráfica contendrá N + 1 puntos, a través de los cuales se puede dibujar la llamada *curva característica de operación* (o curva CO) de un plan de muestreo simple.

Ejemplo 2.2

La fábrica Z elabora cartuchos de dinamita, y los empaca en cajas de 20 unidades. El comprador W acepta cada caja únicamente si al extraer una muestra de dos cartuchos encuentra que ambos son buenos. Elaborar la curva característica de operación correspondiente.

Solución

En este caso, se tiene que N = 20, n = 2 y c = 0. Por lo tanto, las probabilidades de aceptación son, empleando la ec 2.3

$$P\ (A;p) = P\ \{X \leq 0\} = \frac{C_0^{20p}\ \ C_{2-0}^{20-20p}}{C_2^{20}}$$

$$= \frac{\dfrac{20p!}{0!\,(20p-0)!}\;\dfrac{(20-20p)!}{2!\,(20-20p-2)!}}{\dfrac{20!}{2!\,(20-2)!}} =$$

$$= \frac{\dfrac{20p!}{0!\,20p!}\;\dfrac{(20-20p)!}{2\times 1\times (18-20p)!}}{\dfrac{20!}{2\times 1\times 18!}} = \frac{18!\,(20-20p)!}{20!\,(18-20p)!} =$$

$$= \frac{(20\;-\;20p)\;(19\;-\;20p)}{380}$$

Si se le asignan a p los 21 valores 0, 1/20, 2/20, 3/20, ...,19/20,1, se obtienen los correspondientes de P (A; p). Por ejemplo, para p = 10/20 = 0.5, la probabilidad de aceptación es

$$P\ (A;\ 0.5) = \frac{[20\;-\;20(10/20)]\;[19\;-\;20\;(10/20)]}{380}$$

$$= \frac{(20\;-\;10)\;(19\;-\;10)}{380} = \frac{(10)\,(9)}{380} = \frac{90}{380} = 0.237$$

Siguiendo el procedimiento anterior, se obtienen los puntos siguientes:

p	P (A; p)
0/20 = 0.00	1.000
1/20 = 0.05	0.900
2/20 = 0.10	0.305
3/20 = 0.15	0.716
4/20 = 0.20	0.632
5/20 = 0.25	0.553
6/20 = 0.30	0.479
7/20 = 0.35	0.411
8/20 = 0.40	0.347
9/20 = 0.45	0.289
10/20 = 0.50	0.237
11/20 = 0.55	0.189
12/20 = 0.60	0.147
13/20 = 0.65	0.111
14/20 = 0.70	0.079
15/20 = 0.75	0.053
16/20 = 0.80	0.032
17/20 = 0.85	0.016
18/20 = 0.90	0.005
19/20 = 0.95	0.000
20/20 = 1.00	1.000

La curva característica de operación correspondientes es la que se hace pasar por los puntos anteriores, y se presenta en la Fig 2.1.

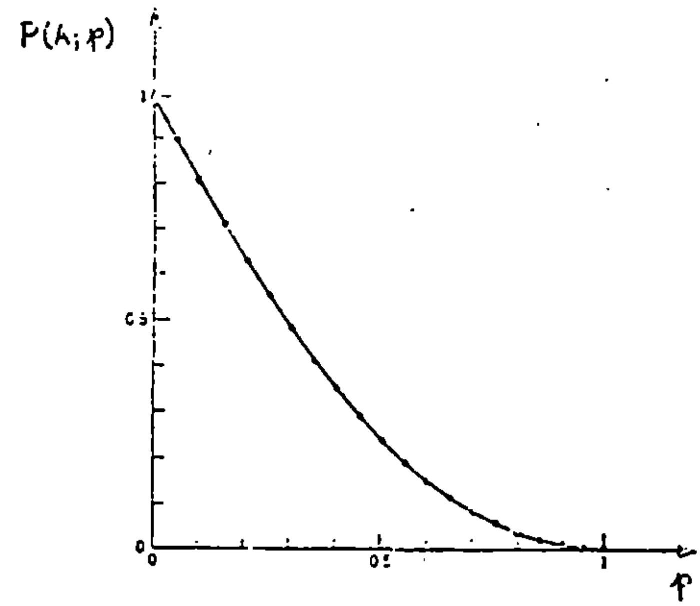


Fig 2.1 Curva CO para un plan de muestreo simple, con N = 20, n = 2 y c = 0.

En la Fig 2.1 se puede observar que a medida que se hace más grande la fraccion de defectuosos en el lote (o el número de artículos defectuosos), la probabilidad de aceptación del mismo se va haciendo cada vez menor. Los casos extremos se dan en p = 0, en que la aceptación del lote es un evento seguro, y en p = 1, cuando es imposible aceptarlo.

2.3 Empleo de la aproximación binomial para construir la curva CO

En la mayor parte de los casos prácticos, el porcentaje de artículos defectuosos en un lote será pequeño (menor del 10%), en tanto que el tamaño del mismo será muy grande (1000 elementos, 10000 elementos, etc), y el de la muestra usualmente será varias veces menor, de tal manera que es posible aproximar las probabilidades dadas por la distribución hipergeométrica (ecs 2.1 y 2.3) empleando la distribución binomial. En particular, la aproximación es buena cuando $N \leq 10\ n$. En estos casos, se puede escribir

$$P\ (A;\ p) = P\ \{X \leq c\} \doteq \sum_{X=0}^{c} C_X^n\ p^X\ (1-p)^{n-X} \qquad (2.4)$$

Se debe observar que siempre se define a p como en la ec 2.2, y que serán mejor aproximadas por la ecuación anterior aquellas probabilidades de aceptación para las cuales el valor de p sea pequeño.

Ejemplo 2.3

En el caso del ejemplo 2.2 anterior, aproxímense las probabilidades de aceptación hipergeométricas para los distintos valores de p mediante la distribución binomial.

Solución

En este caso sí es posible realizar la aproximación pedida, ya que se verifica la condición $N \geq 10n$, porque siendo $N = 20$ y $n = 2$, se tiene que $20 \geq 10(2)$. Por ejemplo, para $p = 0.2$, la aproximación binomial dada por la ec 2.4 conduce al valor

$$P\ (A\ ;\ 0.2) = P\ \{X \leq 0\} = C_0^2\ \ (0.2)^0\ \ (1-0.2)^{2-0}$$

$$= \frac{2!}{0!\,(2-0)!}\ \ (0.8)^2 = 0.640$$

en contra del valor exacto 0.632 obtenido mediante la ec 2.3.

Procediendo en forma similar se calculan los restantes valores de P (A; p), los cuales se presentan de 0.1 en 0.1 en la tabla siguiente, junto con los anteriormente obtenidos en el ejemplo 2.2 para fines de comparación.

p	Hipergeométrica P (A; p)	Binomial P (A; p)
0.00	1.000	1.000
0.10	0.805	0.810
0.20	0.632	0.640
0.30	0.479	0.490
0.40	0.347	0.360
0.50	0.237	0.250
0.60	0.147	0.160
0.70	0.075	0.090
0.80	0.032	0.040
0.90	0.005	0.010
1.00	0.000	0.000

En la tabla se puede observar que las probabilidades de aceptación se aproximan bastante más a las exactas cuando el valor de p se encuentra en la vecindad de p = 0.10.

2.4 Empleo de la aproximación de Poisson para construir la curva CO

Como ya se vio, la distribución hipergeométrica se puede aproximar adecuadamente mediante la binomial cuando $N \geq 10$ y $p \leq 0.1$. A su vez, la distribución binomial puede aproximarse suficientemente bien mediante la de Poisson cuando se cumple lo anterior y np es menor de 15, lo cual evita en ocasiones la gran cantidad de labor numérica que se requiere para calcular las probabilidades de aceptación mediante las distribuciones hipergeométrica y binomial.

Entonces, si se hace $\lambda = np$ para la distribución de Poisson, se puede escribir

$$P\ (A;\ p) = P\ \{X \leq c\} \doteq e^{-np} \sum_{X=0}^{c} \frac{(np)^{X}}{X!}$$

La aproximación anterior es muy útil cuando los lotes son grandes, ya que como se puede apreciar, la ec 2.4 no requiere del manejo de dicho dato para el cálculo de las probabilidades de aceptación que se emplean para construir la curva CO.

Ejemplo 2.4

Obténganse los valores de P (A; p) para p = 0, 0.1, 0.2, 0.3, 0.5 y 1.0 en el caso del plan de muestreo simple del ejemplo 2.2, aproximando mediante la distribución de Poisson.

Solución

Se sabe que n = 2 y c = 0, por lo que

$$np = 2(0) = 0 \; ; \qquad P\ (A;\ 0) = \frac{e^{-0}\ \ 0^{0}}{0!} = 1$$

$$np = 2(0.1) = 0.2; \quad P\ (A;\ 0.1) = \frac{e^{-0.2}\ \ 0.2^{0}}{0!} = 0.818$$

$$np = 2(0.2) = 0.4; \quad P\ (A;\ 0.2) = \frac{e^{-0.4}\ \ 0.4^{0}}{0!} = 0.670$$

$$np = 2(0.3) = 0.6; \quad P\ (A;\ 0.3) = \frac{e^{-0.6}\ \ 0.6^{0}}{0!} = 0.549$$

$$np = 2(0.5) = 1.0; \quad P\ (A;\ 0.5) = \frac{e^{-1.0}\ \ 1.0^{0}}{0!} = 0.367$$

$$np = 2(1.0) = 2.0; \quad P\ (A;\ 1.0) = \frac{e^{-2.0}\ \ 2.0^{0}}{0!} = 0.135$$

En la siguiente tabla se comparan los valores hipergeométricos exactos con los obtenidos mediante las aproximaciones binomial y de Poisson.

p	P (A;p) Hipergeométrica	P (A; p) Binomial	P (A;p) Poisson
0	1.000	1.000	1.000
0.1	0.805	0.810	0.818
0.2	0.632	0.640	0.670
0.3	0.479	0.490	0.549
0.5	0.237	0.250	0.367
1.0	0.000	0.000	0.135

Como se puede observar en la tabla anterior, las probabilidades de aceptación calculadas con la fórmula de Poisson difieren bastante de las exactas y de las binomiales cuando p no se encuentra cercano al valor 0.1. Sin embargo, hay que considerar que en el problema anterior los tamaños del lote y la muestra son bastante pequeños, por lo que la aproximación de Poisson no puede ser muy buena.

De hecho, la forma práctica para construir las curvas CO se fundamenta en el método aproximado de Poisson, considerando que los lotes que entrega el productor son muy grandes, y haciendo uso de la tabla 2.1 que se presenta adelante, en la cual se proporcionan, en función del número de aceptación c y del valor $\lambda = np$, las probabilidades de aceptación

$$P(A;\, p) = P\{X \leq c\} = e^{-np} \sum_{x=0}^{c} \frac{(np)^x}{x!}$$

multiplicadas por mil.

TABLA 2.1

TERMINOS ACUMULATIVOS DE LA APROXIMACION DE POISSON A BINOMIAL

np \ c	0	1	2	3	4	5	6	7	8	c / np
0.02	980	1000	....	....	....	....	....	....	....	0.02
0.04	961	999	1000	....	....	....	....	....	....	0.04
0.06	942	998	1000	....	....	....	....	....	....	0.06
0.08	923	997	1000	....	....	....	....	....	....	0.08
0.10	905	995	1000	....	....	....	....	....	....	0.10
0.15	861	990	999	1000	....	....	....	....	....	0.15
0.20	819	982	999	1000	....	....	....	....	....	0.20
0.25	779	974	998	1000	....	....	....	....	....	0.25
0.30	741	963	996	1000	....	....	....	....	....	0.30
0.35	705	951	994	1000	....	....	....	....	....	0.35
0.40	670	938	992	999	1000	....	....	....	....	0.40
0.45	638	925	989	999	1000	....	....	....	....	0.45
0.50	607	910	986	998	1000	....	....	....	....	0.50
0.55	577	894	982	998	1000	....	....	....	....	0.55
0.60	549	878	977	997	1000	....	....	....	....	0.60
0.65	522	861	972	996	999	1000	....	....	....	0.65
0.70	497	844	966	994	999	1000	....	....	....	0.70
0.75	472	827	959	993	999	1000	....	....	....	0.75
0.80	449	809	953	991	999	1000	....	....	....	0.80
0.90	407	772	937	987	998	1000	....	....	....	0.90
1.00	368	736	920	981	996	999	1000	....	....	1.00
1.10	333	699	900	974	995	999	1000	....	....	1.10
1.20	301	663	879	966	992	998	1000	....	....	1.20
1.30	273	627	857	957	989	998	1000	....	....	1.30
1.40	247	592	833	946	986	997	999	1000	....	1.40
1.50	223	558	809	934	981	996	999	1000	....	1.50
1.60	202	525	783	921	976	994	999	1000	....	1.60
1.70	183	493	757	907	970	992	998	1000	....	1.70
1.80	165	463	731	891	964	990	997	999	1000	1.80
1.90	150	434	704	875	956	987	997	999	1000	1.90
2.00	135	406	677	857	947	983	995	999	1000	2.00
2.10	122	380	650	839	938	980	994	999	1000	2.10
2.20	110	354	622	819	927	974	993	998	1000	2.20
2.30	100	331	596	799	916	970	991	997	999	2.30
2.40	091	308	570	779	904	964	988	997	999	2.40
2.50	082	287	544	758	891	958	986	996	999	2.50
2.60	074	267	518	736	877	951	983	995	999	2.60
2.70	067	249	494	714	863	943	979	993	998	2.70
2.80	061	231	469	692	848	935	976	992	998	2.80
2.90	055	215	446	670	832	926	971	990	997	2.90
3.00	050	199	423	647	815	916	966	988	996	3.00
3.10	045	185	401	625	798	906	961	986	995	3.10
3.20	041	171	380	603	781	895	955	983	994	3.20
3.30	037	159	359	580	763	883	949	980	993	3.30
3.40	033	147	340	558	744	871	942	977	992	3.40
3.50	030	136	321	537	725	858	935	973	990	3.50
3.60	027	126	303	515	706	844	927	969	988	3.60
3.70	025	116	285	494	687	830	918	965	986	3.70
3.80	022	107	269	473	668	816	909	960	984	3.80
3.90	020	099	253	453	648	801	899	955	981	3.90
4.00	018	092	238	433	629	785	889	949	979	4.00
4.10	017	085	224	414	609	769	879	943	976	4.10
4.20	015	078	210	395	590	753	867	936	972	4.20
4.30	014	072	197	377	570	737	856	929	968	4.30
4.40	012	066	185	359	551	720	844	921	964	4.40
4.50	011	061	174	342	532	703	831	913	960	4.50
4.60	010	056	163	326	513	686	818	905	955	4.60
4.70	009	052	152	310	495	668	805	896	950	4.70
4.80	008	048	143	294	476	651	791	887	944	4.80
4.90	007	044	133	279	458	634	777	877	938	4.90

np \ c	0	1	2	3	4	5	6	7	8	c / np
5.00	007	040	125	265	440	616	762	867	932	5.00
5.10	006	037	116	251	423	598	747	856	925	5.10
5.20	006	034	109	238	406	581	732	845	918	5.20
5.30	005	031	102	225	390	563	717	833	911	5.30
5.40	004	029	095	213	373	546	702	822	903	5.40
5.50	004	027	088	202	358	529	686	809	894	5.50
5.60	004	024	082	191	342	512	670	797	886	5.60
5.70	003	022	077	180	327	495	654	784	877	5.70
5.80	003	021	072	170	313	478	638	771	867	5.80
5.90	003	019	067	160	299	462	622	758	857	5.90
6.00	002	017	062	151	285	446	606	744	847	6.00
6.10	002	016	058	143	272	430	590	730	837	6.10
6.20	002	015	054	134	259	414	574	716	826	6.20
6.30	002	013	050	126	247	399	558	702	815	6.30
6.40	002	012	046	119	235	384	542	687	803	6.40
6.50	002	011	043	112	224	369	527	673	792	6.50
6.60	001	010	040	105	213	355	511	658	780	6.60
6.70	001	009	037	099	202	341	495	643	767	6.70
6.80	001	009	034	093	192	327	480	628	755	6.80
6.90	001	008	032	087	182	314	465	614	742	6.90
7.00	001	007	030	082	173	301	450	599	729	7.00
7.20	001	006	025	072	156	276	420	569	703	7.20
7.40	001	005	022	063	140	253	392	539	676	7.40
7.60	000	004	019	055	125	231	365	510	648	7.60
7.80	...	004	016	048	112	210	338	481	620	7.80
8.00	...	003	014	042	100	191	313	453	593	8.00
8.20	...	003	012	037	089	174	290	425	565	8.20
8.40	...	002	010	032	079	157	267	399	537	8.40
8.60	...	002	009	028	070	142	246	373	509	8.60
8.80	...	001	007	024	062	128	226	348	482	8.80
9.00	...	001	006	021	055	116	207	324	456	9.00
9.20	...	001	005	018	049	104	189	301	430	9.20
9.40	...	001	005	016	043	093	173	279	404	9.40
9.60	...	001	004	014	038	084	157	258	380	9.60
9.80	...	001	003	012	033	075	143	239	356	9.80
10.00	...	001	003	010	029	067	130	220	333	10.00
10.20	...	000	002	009	026	060	118	203	311	10.20
10.40	...	...	002	008	023	053	107	186	290	10.40
10.60	...	...	002	007	020	045	097	171	269	10.60
10.80	...	...	001	006	018	042	087	157	250	10.80
11.00	...	...	001	005	015	037	079	143	232	11.00
11.20	...	...	001	004	013	033	071	131	215	11.20
11.40	...	...	001	004	012	029	064	119	198	11.40
11.60	...	...	001	003	010	026	057	108	183	11.60
11.80	...	...	001	003	009	023	051	098	169	11.80
12.00	...	...	001	002	008	020	046	090	155	12.00
12.20	...	...	000	002	007	018	041	081	143	12.20
12.40	...	...	...	002	006	016	037	073	131	12.40
12.60	...	...	...	001	005	014	033	066	120	12.60
12.80	...	...	...	001	004	012	029	060	109	12.80
13.00	...	...	...	001	004	011	026	054	100	13.00
13.20	...	...	...	001	003	009	023	049	091	13.20
13.40	...	...	...	001	003	008	020	044	083	13.40
13.60	...	...	...	001	002	007	018	039	075	13.60
13.80	...	...	...	001	002	006	016	035	068	13.80
14.00	...	...	...	000	002	006	014	032	062	14.00
14.20	...	...	...	...	002	005	013	028	056	14.20
14.40	...	...	...	...	001	004	011	025	051	14.40
14.60	...	...	...	...	001	004	010	023	046	14.60
14.80	...	...	...	...	001	003	009	020	042	14.80
15.00	...	...	...	...	001	003	008	018	037	15.00

A continuación se presenta un ejemplo práctico de construcción de una curva CO mediante el método descrito, haciendo uso de la tabla 2.1.

Ejemplo 2.5

Supóngase que un receptor de producto terminado adopta el plan de muestreo simple siguiente:

a. Recibe lotes de ciertos artículos con 1000 unidades c/u.

b. Extrae de cada lote una muestra aleatoria de 20 artículos.

c. Si la muestra extraída contiene dos o más artículos defectuosos, rechaza el lote. De no ser así, lo acepta.

Constrúyase la curva CO correspondiente.

Solución

Puesto que el tamaño de los lotes es grande, se pueden aproximar adecuadamente las probabilidades de aceptación mediante la distribución de Poisson. Para ello, se considera en la práctica que con los valores

$$P\ (A;\ p) = 0.98,\ 0.95,\ 0.70,\ 0.50,\ 0.20,\ 0.10,\ 0.05,\ 0.02$$

se puede definir suficientemente bien la curva CO.

Para construir la curva del plan de muestreó simple indicado, considérese que $c = 1$ y $n = 20$. En la columna para la cual $c = 1$ en la tabla 2.1, se puede ver que el valor más cercano a 980 (0.98 de probabilidad) es 982. Para dicho valor, el correspondiente de np es 0.2, siendo por lo tanto $p = \frac{np}{n} = \frac{0.2}{20} = 0.01$.

El valor más cercano a 950 (0.95 de probabilidad) es en la tabla el 951. Para este valor, $np = 0.35$ y $p = \frac{0.35}{20} = 0.0175$.

Siguiendo el procedimiento anterior, se llega a

P (A;p)	np	p
1.000	0.00	0.000
0.982	0.20	0.010
0.951	0.35	0.0175
0.699	1.10	0.055
0.493	1.70	0.085
0.199	3.00	0.150
0.099	3.90	0.195
0.052	4.70	0.235
0.021	5.80	0.290
0.000	20.00	1.000

En la Fig 2.2 siguiente se presenta la curva característica de operación correspondiente al problema.

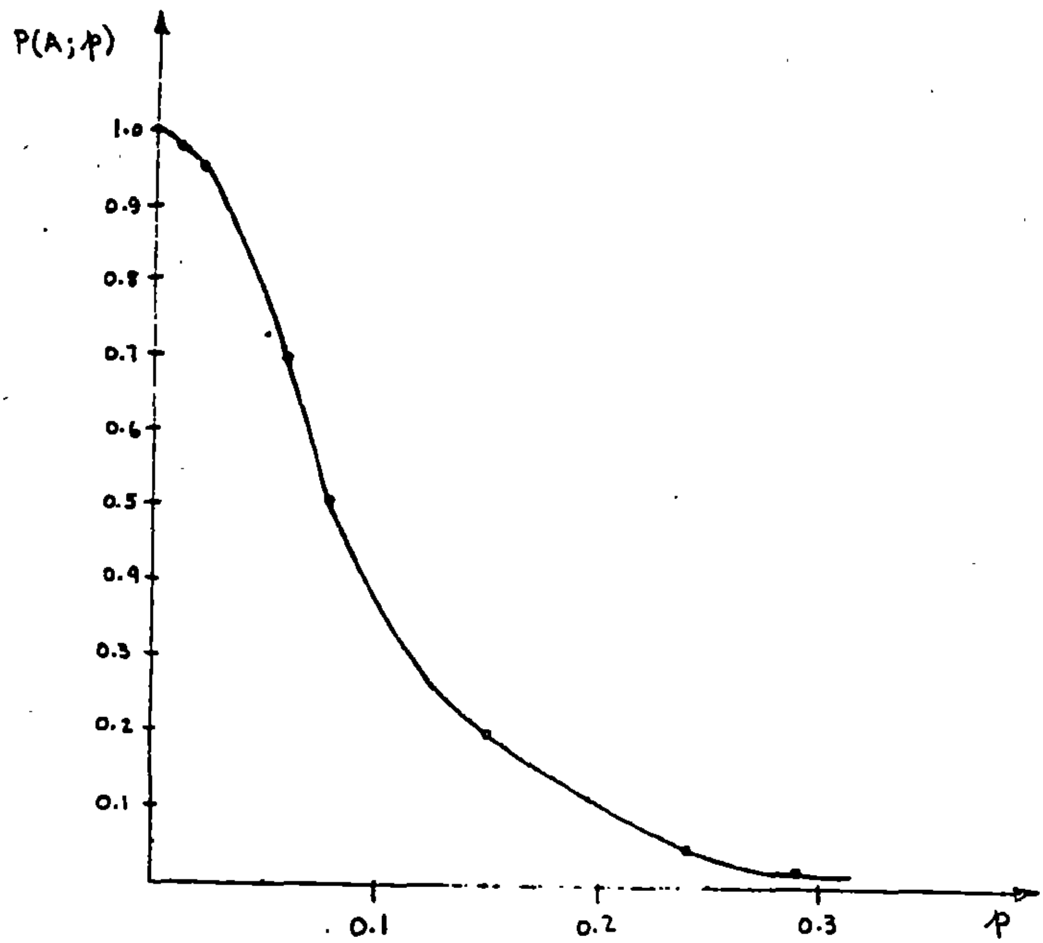


Fig 2.2 Curva característica de operación para plan de muestreo simple con lote grande, c = 1 y n = 20.

2.5 Riesgos en el muestreo de aceptación

Al realizarse los muestreos de aceptación, el productor y el receptor de lotes de artículos tienen intereses distintos al definir un plan de muestreo. El productor puede pedir que la probabilidad, α, de rechazar un lote "bueno" o "aceptable" sea pequeña. Por su parte, el receptor puede exigir que la probabilidad de aceptar un lote "malo" o "no aceptable" sea una cantidad pequeña β.

Para cumplir con ambos compromisos, supóngase que productor y receptor deciden que un lote para el cual p es menor o igual que ciero número p_o es un *lote aceptable*, en tanto que un lote para el que p es mayor o igual que cierto número p_1 $(p_1 > p_o)$ es un lote *no aceptable*, es decir

Si $p \leq p_o$ lote aceptable

Si $p \geq p_1$ lote no aceptable

De acuerdo con lo anterior, α es la probabilidad de rechazar un lote con $p \leq p_o$ y se llama *riesgo del productor*, correspondiendo al error de tipo I que se comete al probar una hipótesis estadística. Por otra parte, β es la probabilidad de aceptar un lote con $p \geq p_1$, se llama *riesgo del receptor*, y corresponde al error de tipo II que se comete al realizar una prueba de hipótesis.

A p_o se le acostumbra llamar *nivel de calidad aceptable* (NCA), y a p_1 *nivel de calidad rechazable* (NCR), o *porcentaje de defectuosos tolerable en un lote* (PDTL). A un lote con $p_o < p < p_1$ se le llama *lote indiferente*.

En la práctica es usual que el acuerdo entre productor y receptor establezca lo siguiente

$$\alpha = \text{Riesgo del productor} \approx 1 - P(A;\ p)_{0.95} = 0.05$$

$$\beta = \text{Riesgo del receptor} \approx P\ (A;\ p)_{0.10} = 0.10$$

Ejemplo 2.6

Para un plan de muestreo simple en el que n = 300 y c = 5, obténganse los valores de p_o y p_1.

Solución

Empleando la tabla 2._. y considerando los valores P (A; p) que definen adecuadamente a la curva CO, se obtiene:

P (A;p)	np	p
1.000	0.00	0.0000
0.980	2.10	0.0070
0.951	2.60	0.0087
0.703	4.50	0.0150
0.495	5.70	0.0190
0.210	7.80	0.0260
0.104	9.20	0.0307
0.048	10.60	0.0353
0.020	12.00	0.0400
0.000	300.00	1.0000

De acuerdo con la tabla, se tiene que

$$\alpha = 1 - P(A;\ p)_{0.951} = 0.0499 \ ;\ p_o = 0.0087$$

$$\beta = P(A;\ p)_{0.104} = 0.104 \ ;\ p_1 = 0.0307$$

En la Fig 2.3 que se presenta a continuación, se muestra la curva CO del plan simple en cuestión, así como los valores del NCA y del NCR.

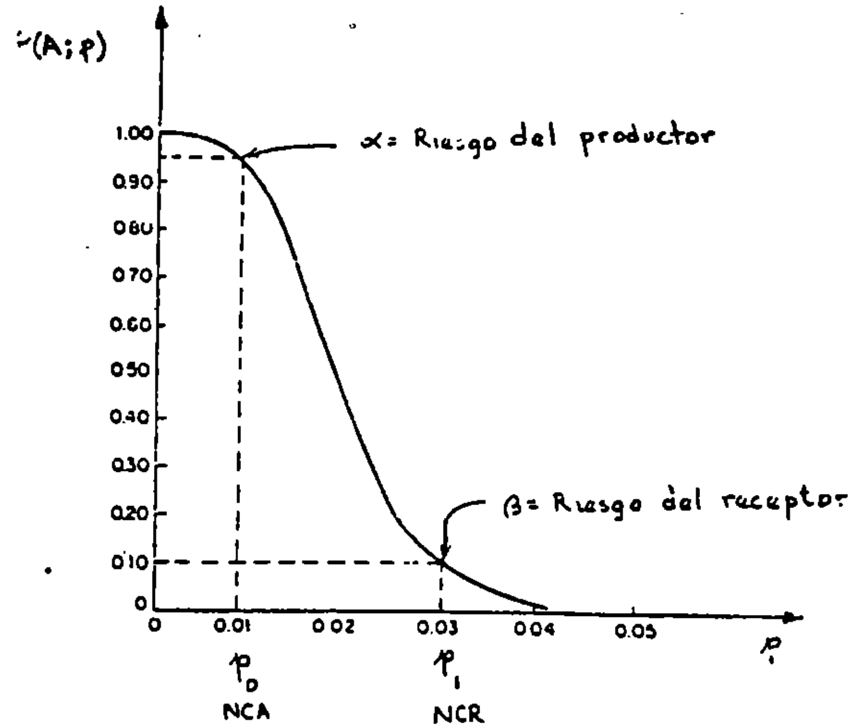


Fig 2.3 Curva CO para plan de muestreo simple con $n = 300$ y $c = 5$.

2.6 Cálculo de n y c a partir de p_0, p_1, α y β.

Al observar la Fig 2.3 se puede concluir que los puntos $(p_0, 1-\alpha)$ y (p_1, β) se localizan en la curva CO. Tomando ello en cuenta, existe un método iterativo aproximado para determinar los valores de n y c, considerando conocidos los de p_0, p_1, α y β, de manera que la curva CO pase muy cerca de los puntos mencionados. Dicho procedimiento se expondrá en el ejemplo que sigue, haciendo uso de la tabla 2.1.

Ejemplo 2.7

Para cierto plan de muestreo simple, se fijan los riesgos siguientes:

a. Productor: Aquellos lotes que contengan un 1% de artículos defectuosos se rechazarán en el 5% de los casos.

b. Receptor: Los lotes que contengan un 6% de artículos defectuosos se aceptarán en el 10% del total de casos.

¿Cuáles son los valores del tamaño de la muestra y del número de aceptación que se deben emplear para dicho plan?

Solución

De acuerdo con los datos del problema, se desprende que

$$\alpha = 0.05 \quad ; \quad p_o = 0.01$$

$$\beta = 0.10 \quad ; \quad p_1 = 0.06$$

a. Se considera c = 0, con lo cual, de la tabla 2.1,

np_o (para $\alpha = 0.05$ o $P\ (A;\ 0.01) = 0.95) \doteq 0.05$

np_1 (para $\beta = 0.10$) = 2.30

Entonces

$$n_\alpha = \frac{np_o}{p_o} = \frac{0.05}{0.01} = 5$$

$$n_\beta = \frac{np_1}{p_1} = \frac{2.30}{0.06} \approx 38$$

Obviamente, se debe verificar que $n_\alpha = n_\beta$; no siendo este el caso, se hace ahora c = 1.

b. Se considera c = 1, obteniéndose ahora de la tabla 2.1 lo siguiente

$$np_o \text{ (para } \alpha = 0.05) \doteq 0.35$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 3.90$$

Por lo tanto

$$n_\alpha = \frac{0.35}{0.01} = 35$$

$$n_\beta = \frac{3.90}{0.06} = 65$$

Tampoco se verifica que $n_\alpha = n_\beta$; por lo tanto, se hace c = 2.

c. Se considera c = 2, y

$$np_0 \text{ (para } \alpha = 0.05) \doteq 0.82$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 5.32$$

Ahora, se tiene que

$$n_\alpha = \frac{0.82}{0.01} = 82$$

$$n_\beta = \frac{5.30}{0.06} = 88$$

Ahora n_α y n_β se parecen bastante, pero aún no son iguales. Por lo tanto, se hace c = 3 para saber si la diferencia se hace más pequeña.

d. Se considera c = 3, y se obtiene

$$np_0 \text{ (para } \alpha = 0.05) \doteq 1.37$$

$$np_1 \text{ (para } \beta = 0.10) \doteq 6.68$$

Luego

$$n_\alpha = \frac{1.37}{0.01} = 137$$

$$n_\beta = \frac{6.68}{0.06} = 112$$

Se observa que ahora la diferencia se hace más grande. Por lo que el valor real de n se debe encontrar entre 82 y 88 elementos para $c = 2$. Con el fin de ajustar adecuadamente el valor de n, se puede hacer

$$n = \frac{n_\alpha + n_\beta}{2} = \frac{82 + 88}{2} = 85$$

Por lo tanto, el plan de muestreo simple es el siguiente

$$\alpha = 0.05 \quad ; \quad \beta = 0.10$$

$$p_o = 0.01 \quad ; \quad p_1 = 0.06$$

$$n = 85 \quad ; \quad c = 2$$

cuya curva CO se muestra en la Fig 2.4.

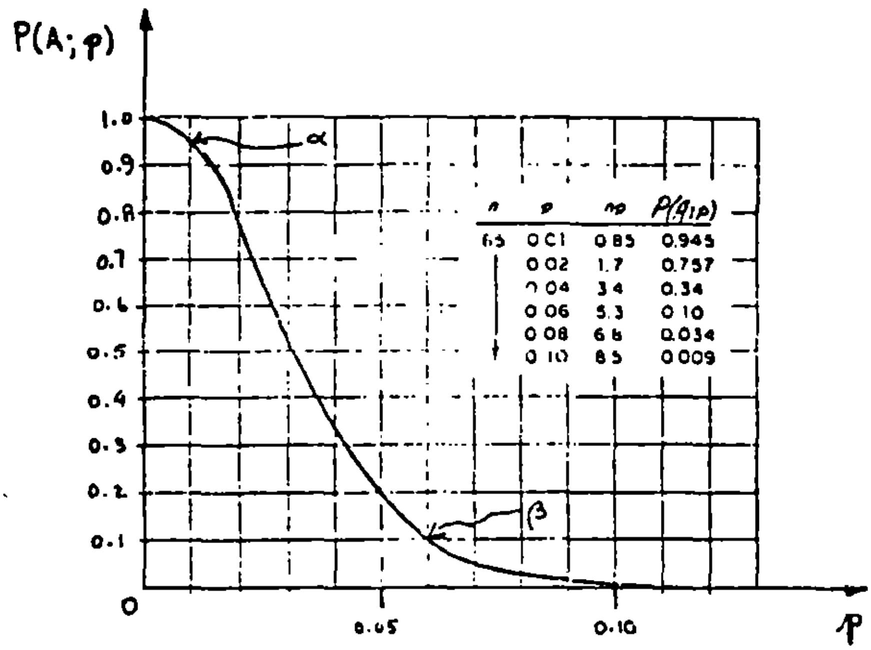


Fig 2.4 Curva CO ajustada para α, β, p_o y p_1 conocidos.

2.7 Comentarios sobre la curva CO

Al comparar las curvas CO de las Figs 2.3 y 2.4, se puede observar que, no obstante el número más grande de artículos defectuosos que permite en la muestra el plan de muestreo asociado a la curva CO de la Fig 2.3, se trata de un mejor plan de aceptación de lotes, en el sentido de que proporciona riesgos más favorables al receptor.

En efecto, ambos planes consideran $\alpha = 0.05$, $\beta = 0.10$ y $p_o \approx 0.01$, pero el plan de la Fig 2.4 aceptará lotes con 6% de defectuosos ($p_1 \approx 0.06$) en el 10% del total de casos, en tanto que el de la Fig 2.3 aceptará lotes con 3% de defectuosos ($p_1 \approx 0.03$)

en el mismo número de casos.

En muchas ocasiones no se comprende con claridad el porqué de un número de aceptación mayor de cero en los planes de muestreo. Si se observa la Fig 2.5, se puede apreciar que las curvas CO (a), (b) y (c) corresponden a planes de muestreo que evitan los artículos defectuosos en la muestra ($c = 0$), pero que tienen riesgos de productor y receptor distintos. Los planes de las curvas CO (d) y (e) consideran 4 y 7 defectuosos en la muestra, respectivamente.

Se observa que las curvas CO con $c = 0$ se caracterizan por patrones cóncavos, en tanto que aquellas con $c \neq 0$ semejan curvas S invertidas.

Los planes de muestreo con $c = 0$ usualmente penalizan más al productor. Asimismo, aquellos planes en que c es mayor de cero proporcionan riesgos más favorables al productor o al receptor, y en muchos casos a ambos.

Se puede afirmar que el riesgo para el receptor se hace más pequeño conforme se incrementa el tamaño de la muestra, en tanto que el riesgo para el productor decrece conforme se permiten uno o más artículos defectuosos en la misma. Esto se puede aclarar si se observan los riesgos en las curvas (c) y (d) de la Fig 2.5.

Las curvas (d) y (e) consideran esencialmente el mismo riesgo para el productor (NCA $\approx$ 0.01 en $\alpha = 0.05$), pero la (e) considera un tamaño de muestra mayor, por lo que el receptor corre un riesgo menor. La curva (f) corresponde a la curva ideal CO, ya que ese plan de muestreo acepta todos los lotes con uno por ciento o menos de artículos defectuosos y rechaza todos los lotes que contengan más del 1% de defectuosos. Dicha curva obviamente no se puede obtener con las técnicas usuales de muestreo de aceptación.

Lo anterior indica que un plan de muestreo simple será más efectivo en tanto su curva CO correspondiente se asemeje más a la curva ideal de operación.

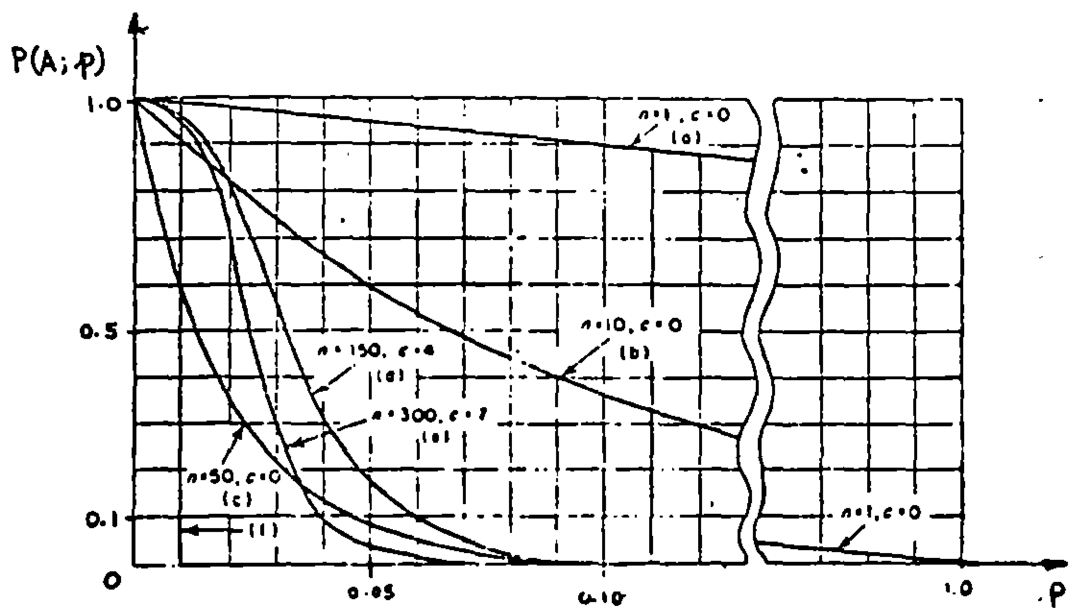


Fig 2.5 Distintos planes de muestreo con $c = 0$ y $c \neq 0$.

3. Plan de muestreo doble

Un plan de muestreo simple requiere que se tome una decisión sobre la aceptación o rechazo de un lote tomando como base la evidencia de una muestra extraída del mismo.

Sin embargo, un *plan de muestreo doble* implica la posibilidad de posponer la decisión sobre la aceptación o rechazo de un lote hasta que *una segunda* muestra haya sido extraída. Dicho lote podrá ser aceptado inmediatamente si la primera muestra es muy buena, o rechazado enseguida si la primera muestra es bastante mala. Si la primera muestra no es ni muy buena ni muy mala, la decisión se basa en la evidencia de la primera y segunda muestras combinadas.

En general, los planes de muestreo doble conducen a menos inspección total que los planes sencillos, y también proporcionan la ventaja sicológica que conlleva la idea de dar una segunda oportunidad a los lotes dudosos.

3.1 Símbolos en el muestreo doble

Los siguientes son los símbolos empleados en conexión con el muestreo doble:

N = tamaño del lote

n_1 = tamaño de la primera muestra

c_1 = número de aceptación para la primera muestra

n_2 = tamaño de la segunda muestra

$n_1 + n_2$ = tamaño de la muestra combinada

c_2 = número de aceptación para la muestra combinada

3.2 Interpretación del plan de muestreo doble

Considérese un plan de muestreo doble para el cual se fijan los valores de N, n_1, c_1, n_2 y c_2 ($c_2 > c_1$). La interpretación del proceso que se realiza con dicho plan es la siguiente:

a. Se inspecciona una primera muestra de tamaño n_1 extraída del lote de tamaño N.

b. Se acepta el lote si la muestra anterior contiene c_1 o menos artículos defectuosos.

c. Se rechaza el lote si el número de defectuosos en la muestra excede el valor c_2

d. Si la primera muestra contiene $c_1 + 1$, $c_1 + 2$, ... o c_2 artículos defectuosos, se extrae e inspecciona una segunda con n_2 elementos.

e. Se acepta el lote sobre la base de la muestra combinada con $n_1 + n_2$ elementos si dicha muestra contiene c_2 artículos defectuosos o menos.

f. Se rechaza el lote si la muestra combinada contiene más de c_2 defectuosos.

3.2 Curva CO de un plan de muestreo doble

De acuerdo con lo que se ha explicado, existen cuatro posiblilidades de que se acepte o se rechace un lote sometido para muestreo doble. Dichas posibilidades son

a. Aceptación después de la primera muestra.

b. Rechazo después de la primera muestra.

c. Aceptación después de la segunda muestra.

d. Rechazo después de la segunda muestra

Tomando como base lo anterior, se explicará a través del ejemplo siguiente la forma como se construye la curva CO para el plan de muestreo doble.

Ejemplo 3.1

Considérese el plan de muestreo doble para el cual tamaño del lote es muy grande, $n_1 = 50$, $c_1 = 1$, $n_2 = 100$ y c_2 ...

Constrúyase la curva CO correspondiente.

Solución

Para determinar los puntos de la curva CO, es necesario calcular las probabilidades de que si se toma una segunda muestra el lote sea aceptado, para distintos valores de p. Para ilustrar lo anterior considérese inicialmente el valor $p = 0.02$.

Entonces, un lote puede ser aceptado según el plan anterior en cualquiera de las formas siguientes:

a. un defectuoso o menos en la primera muestra

b. dos defectuosos en la primera muestra, seguido de cero o un defectuoso en la segunda muestra

c. tres defectuosos en la primera muestra, seguidos de cero defectuosos en la segunda muestra.

La probabilidad de aceptar un lote es entonces igual a la suma de las probabilidades de estos diferentes modos por los cuales puede ser aceptado.

Inicialmente, se deben calcular las probabilidades de tener uno o menos, dos o menos y tres o menos defectuosos en la primera muestra. Lo anterior equivale a considerar un plan de muestreo simple para el cual $n_1 = 50$ y $c = 1, 2, 3$. A continua-

... se deben calcular las probabilidades de tener exactamente dos y tres defectuosos en la primera muestra.

Entonces, con $n_1 p = 50(0.02) = 1.00$, se obtiene, empleando la tabla 2.1 y siendo X el número de elementos defectuosos

$P\{X \leq 1\}_1 = 0.736$ ... $c = 1$, $n_1 p = 1.00$

$P\{X \leq 2\}_1 = 0.920$ ... $c = 2$, $n_1 p = 1.00$

$P\{X \leq 3\}_1 = 0.981$ ... $c = 3$, $n_1 p = 1.00$

$$P\{X = 2\}_1 = P\{X \leq 2\}_1 - P\{X \leq 1\}_1 = 0.920 - 0.736 = 0.184$$

$$P\{X = 3\}_1 = P\{X \leq 3\}_1 - P\{X \leq 2\}_1 = 0.981 - 0.920 = 0.061$$

El subíndice fuera de la llave indica que la probabilidad del evento se calcula con base en la primera muestra.

Ahora bien, si en la primera muestra hay dos defectuosos, los cálculos relacionados con la segunda muestra deberán basarse en $n_2 p = 100(0.02) = 2$. El tomar la segunda muestra e inspeccionarla equivale, para efectos de los cálculos, a considerar un nuevo plan de muestreo simple para el resto del lote con número de aceptación igual a 1, ya que este elemento, sumado a los dos defectuosos considerados, permite la aceptación del lote.

Por lo tanto,

$$P\ \{X \leq 1\}_2 = 0.406 \qquad c = 1, \quad n_2p = 2$$

Si en la primera muestra hay tres defectuosos, los cálculos para la segunda muestra se deben basar en $n_2p = 100(0.02)$ y un número de aceptación igual a cero, es decir

$$P\ \{X \leq 0\}_2 = 0.135 \qquad c = 0\ , \ n_2p = 2$$

La probabilidad de aceptación es, empleando el concepto de independencia de eventos, la suma de las probabilidades siguientes:

P {un defectuoso o menos en la primera muestra} $= P\ \{X \leq 1\}_1 = 0.736$

+ P {dos defectuosos en la primera muestra, seguidos de cero o un defectuoso en la segunda} $= P\ \{X = 2\}_1\ P\{X \leq 1\}_2 = (0.184)(0.406) =$

$= 0.075$

+ P {tres defectuosos en la primera muestra, seguidos de cero defectuosos en la segunda} $= P\ \{X = 3\}_1\ P\{X \leq 0\}_2 = (0.061)(0.135) =$

$= 0.008$

Entonces,

$$P\ (A;\ 0.02) = 0.736 + 0.075 + 0.008 = 0.819$$

es decir, el punto (0.02, 0.819) se encuentra sobre la curva CO del plan de muestreo doble.

En la forma descrita anteriormente, se pueden calcular también los puntos restantes para definir la curva CO, quedando finalmente

P (A; p)	p
0.98	0.012
0.95	0.015
0.82	0.020
0.70	0.027
0.50	0.037
0.20	0.065
0.10	0.080
0.05	0.100
0.02	0.136

La gráfica de la curva CO correspondiente al plan de muestreo doble propuesto se presenta en la Fig 3.1

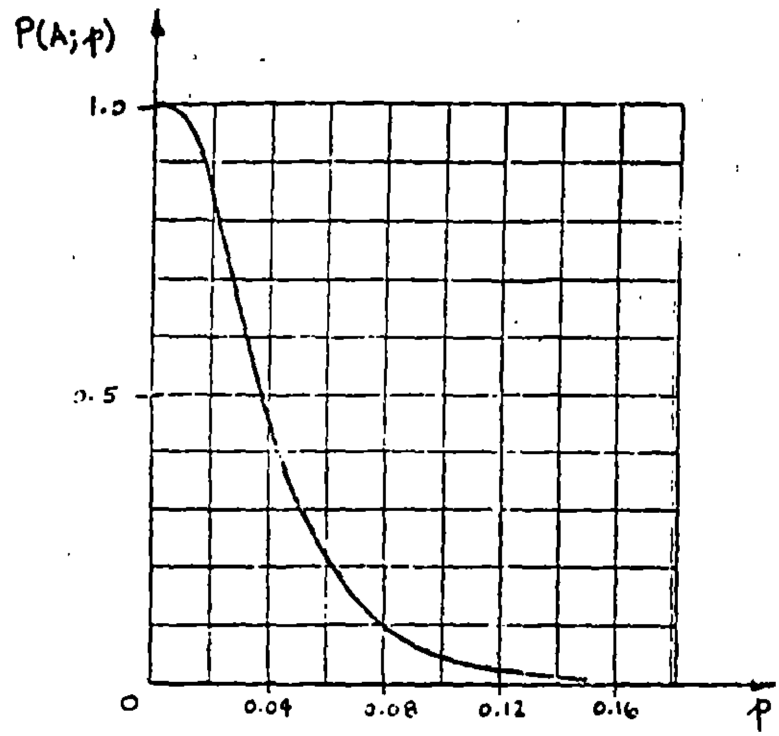


Fig 3.1 Curva CO para plan de muestreo doble con $n_1 = 50$, $c_1 = 1$, $n_2 = 100$, $c_2 = 3$.

4. Plan de muestreo múltiple

De la misma manera que los planes de muestreo doble pueden diferir la decisión sobre la aceptación o rechazo de un lote hasta que haya sido tomada una segunda muestra, otros planes pueden permitir la extracción de cierto número de muestras antes de que una decisión sea tomada.

Los planes de *muestreo múltiple* son usados cuando se permite la extracción de tres o más muestras de un tamaño preestablecido, y cuando la decisión sobre la aceptación o rechazo de un lote se debe tomar después de la séptima muestra extraída, considerando que no es permitida la aceptación de ese lote con la evidencia obtenida de la primera muestra.

4.1 Interpretación de un plan de muestreo múltiple

Considérese el siguiente plan de muestreo múltiple:

Número de la muestra	Tamaño de la muestra individual	Tamaño de la muestra combinada	Número de aceptación, c	Número de rechazo, r
1	20	20	-	2
2	20	40	0	3
3	20	60	1	3
4	20	80	2	4
5	20	100	2	4
6	20	120	2	4
7	20	140	[illegible]	[illegible]

La forma de interpretar el plan anterior es la siguiente:

a. Se extrae e inspecciona una muestra de 20 elementos. Si dos o más son defectuosos, se rechaza el lote; si hay uno o cero defectuosos, se extrae e inspecciona una segunda muestra de 20 elementos. (La aceptación del lote no se permite con la primera muestra.)

b. Si en la muestra combinada (20 + 20 = 40), no hay ningún defectuoso, se acepta el lote; si 3 o más artículos son defectuosos se rechaza. De encontrarse uno o dos defectuosos, se toma una tercera muestra de 20 elementos.

c. Si en la muestra combinada (40 + 20 = 60) hay un defectuoso, se acepta el lote; si 3 o más artículos son defectuosos, se rechaza. De encontrarse dos defectuosos, se toma una cuarta muestra de 20 elementos.

d. Si en la muestra combinada (60 + 20 = 80) hay dos defectuosos, se acepta el lote; si 4 o más son defectuosos, se rechaza. De encontrarse tres defectuosos, se toma una quinta muestra de 20 elementos.

•

•

•

g. Si en la muestra combinada (120 + 20 = 140) hay tres defectuosos, se acepta el lote. Si hay cuatro defectuosos o más, se rechaza.

4.2 Curva CO de un plan de muestreo múltiple

La curva característica de operación de un plan de muestreo múltiple se puede obtener siguiendo un procedimiento semejante al empleado en el caso del muestreo doble, haciendo uso de probabilidades condicionales y suponiendo la descomposición del plan múltiple en varios planes sencillos. Desde luego, el cálculo de las probabilidades de aceptación es bastante más complejo, pero el razonamiento es básicamente el mismo.

A continuación, se describirá mediante un ejemplo el procedimiento para la construcción de la curva CO.

Ejemplo 4.1

Considérese el plan de muestreo múltiple descrito anteriormente, y constrúyase la curva CO correspondiente, suponiendo un lote de tamaño grande.

Solución

Los siguientes cálculos corresponden a un solo punto de la curva, para el cual $p = 0.02$. Cada una de las muestras contiene 20 artículos, por lo que para cada una de llas se tendrá $np = 20(0.02) = 0.4$. Entrando con este valor a la tabla 3.1, y considerando que X denota el número de artículos defectuosos, se obtienen, también para cada muestra las probabilidades incondicionales siguientes:

$$P_o = P\{X = 0\} = P\{X \leq 0\} = 0.670$$

$$P_1 = P\{X = 1\} = P\{X \leq 1\} - P\{X \leq 0\} = 0.938 - 0.670 = 0.268$$

$$P_2 = P\{X = 2\} = P\{X \leq 2\} - P\{X \leq 1\} = 0.992 - 0.938 = 0.054$$

Tomando en cuenta que A = aceptación, R = rechazo y CM = continúa muestreo, se hace enseguida el análisis muestra por muestra para obtener la probabilidad P (A; 0.02).

a. Muestra 1 (M1)

número de aceptación = c = no hay
número de rechazo = r = 2

0 def M1 → P_0 = 0.670 → CM (0 def)

1 def M1 → P_1 = 0.268 → CM (1 def)

2 def M1 → → R (2 def)

Probabilidad de aceptación = 0.000

b. Muestra 2 (M2)

c = 0
r = 3

0 def M1, 0 def M2 → P_{00} = (0.670)(0.670) = 0.449 → A (0 def)

0 def M1, 1 def M2 → P_{01} = (0.670)(0.268) = 0.1795 → CM (1 def)

0 def M1, 2 def M2 → P_{02} = (0.670)(0.054) = 0.0362 → CM (2 def)

0 def M1, 3 def M2 → → R (3 def)

1 def M1, 0 def M2 → P_{10} = (0.268)(0.670) = 0.1795 → CM (1 def)

1 def M1, 1 def M2 → P_{11} = (0.268)(0.268) = 0.0718 → CM (2 def)

1 def M1, 2 def M2 → → R (3 def)

Probabilidad de aceptación = 0.449

Nuevos valores:

P_1 = P {un defectuoso en M2} = 0.1795 + 0.1795 = 0.359

P_2 = P {dos defectuosos en M2} = 0.0362 + 0.0718 = 0.108

c. Muestra 3 (M3)

c = 1

r = 3

1 def M2, 0 def M3 $\Rightarrow$ P_{10} = (0.359)(0.670) = 0.2405 > A ([illegible] def)

1 def M2, 1 def M3 $\Rightarrow$ P_{11} = (0.359)(0.268) = 0.0962 -> CM ([illegible] def)

1 def M2, 2 def M3 $\Rightarrow$ -> R (3 def)

2 def M2, 0 def M3 $\Rightarrow$ P_{20} = (0.108)(0.670) = 0.0723 $\Rightarrow$ CM (2 def)

2 def M2, 1 def M3 $\Rightarrow$ -> R ([illegible] def)

Probabilidad de aceptación = 0.2405

Nuevo valor:

P_2 = P {dos defectuosos en M3} = 0.0962 + 0.0723 = 0.1685

d. Muestra 4 (M4)

c = [illegible]

r = 4

2 def M3, 0 def M4 $\Rightarrow$ P_{20} = (0.1685)(0.670) = 0.11[illegible] > [illegible] def)

2 def M3, 1 def M4 $\Rightarrow$ P_{21} = (0.1685)(0.[illegible]8) = 0.0452 > [illegible]

2 def M3, 2 def M4 $\Rightarrow$ -> R (4 def)

Probabilidad de aceptación = 0.1129

Nuevo Valor:

P_3 = P {3 defectuosos en M4} = 0.0451

e. Muestra 5 (M5)

c = 2
r = 4

3 def M4, 0 def M5 ⇒ P_{30} = (0.0451)(0.670) = 0.0302 ⇒ CM (3 def)

3 def M4, 1 def M5 ⇒ ⇒ R (4 def)

Probabilidad de aceptación = 0.000

Nuevo valor:

P_3 = P {3 defectuosos en M5} = 0.0302

f. Muestra 6 (M6)

c = 2
r = 4

3 def M5, 0 def M6 ⇒ P_{30} = (0.0302)(0.670) = 0.0202 ⇒ CM (3 def)

3 def M5, 1 def M6 ⇒ ⇒ R (4 def)

Probabilidad de aceptacion = 0.000

Nuevo valor

P_3 = P (tres defectuosos en M6) = 0.0202

g. Muestra 7 (M7)

c =

r = 4

3 def M6, 0 def M7 → P_{30} = (0.0202)(0.670) = 0.0135 → ... (3 def)

3 def M6, 1 def M7 → → R (4 def)

Probabilidad de aceptación = 0.0135

De acuerdo con lo anterior, la probabilidad de aceptación de un lote, sujeto al plan de muestreo múltiple propuesto con p = 0.02, es

P (A; 0.02) = 0.449 + 0.2405 + 0.1129 + 0.0135 = 0.8159

Siguiendo el método descrito, se pueden calcular los valores de las probabilidades de aceptación para distintos valores de p, con los cuales se definen los puntos necesarios para construir la curva característica de operación correspondiente, que se presenta en la Fig 4.1.

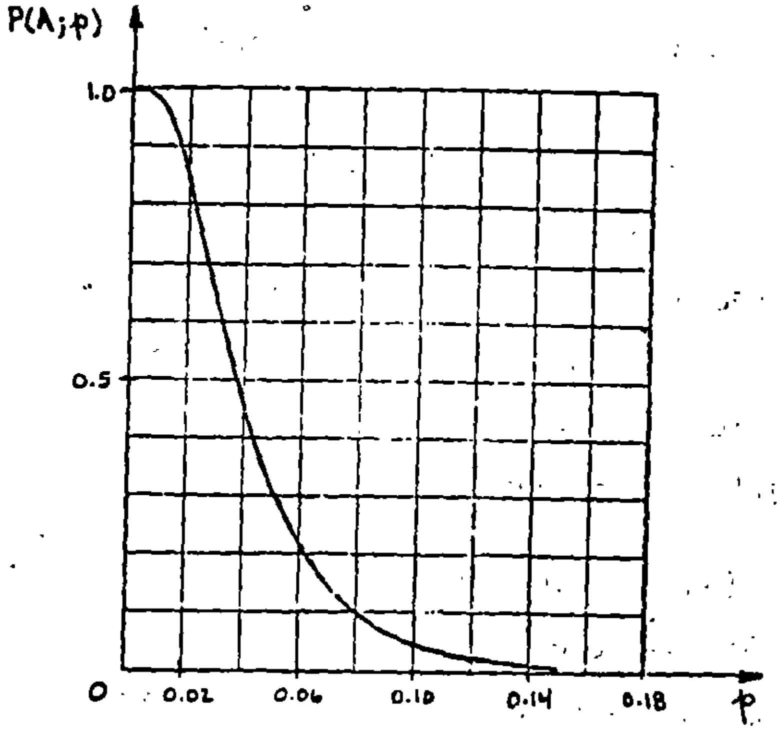


Fig 4.1 Curva CO para un plan de muestreo múltiple

5. Ventajas y desventajas de los planes de muestreo simples, dobles y múltiples

En general, los tres esquemas de muestreo de aceptación que se han presentado se pueden ajustar para proporcionar a lotes con valores de p determinados prácticamente la misma probabilidad de ser aceptados; es decir, si se desea, se puede lograr que las curvas características de operación para los planes simples, dobles y múltiples sean muy parecidas.

No obstante lo anterior, puede suceder que un plan de muestreo de aceptación que ha dado buen resultado para un productor o producto, resulte no tan efectivo para otros. La efectividad de los distintos planes de muestreo expuestos se puede juzgar si se analizan las ventajas y desventajas de cada uno de ellos, en términos de cuatro factores importantes: El número medio de artículos inspeccionados, el costo de administración del plan, la aceptación por parte del producto. y la información sobre calidad de los lotes obtenida a largo plazo. En la tabla 5.1 se compara la efectividad de los tres planes estudiados.

Los factores mencionados en la tabla 5.1 deben ser considerados al seleccionar un plan de muestreo. Por ejemplo, en aquellos casos en que el costo de inspección de cada artículo es elevado, la reducción en el número de artículos inspeccionados puede justificar el empleo del muestreo múltiple no obstante su gran complejidad y elevado costo de administración.

Por otro lado, el muestreo simple puede ser el adecuado si el costo de entrenamiento de personal es muy apreciable. Finalmente, si el problema es de acuerdo entre receptor y productor del plan a emplear, posiblemente la solución sea el muestreo doble, ya que es sicológicamente bien aceptado por ambas partes.

TABLA 5.1

COMPARACION ENTRE LOS PLANES DE
MUESTREO SIMPLE, DOBLE Y MULTIPLE

Factor	Plan simple (PS)	Plan doble (PD)	Plan múltiple (PM)
Número medio de artículos inspeccionados	El más grande de todos	De 5 a 40% menos que en PS	Aproximadamente 25% menos que en PD
Costos de administración (entrenamiento, registros, personal, etc.)	El más bajo de todos	Mayor que el de PS	El más alto de todos
Aceptación por parte del productor	Regular	Adecuada	Poca
Información a largo plazo sobre calidad de los lotes	La mayor	Menos que en PS	La menor

Ejemplo 3.1 (con $p = 0.02$)

a. Muestra 1 (M1)

$c = 1$

$r = 4$

$np = 50(0.02) = 1.0$; $P_0 = 0.368$; $P_1 = 0.368$; $P_2 = 0.184$; $P_3 = 0.061$

0 def M1 $\Rightarrow P_0 = 0.368$ $\Rightarrow$ A (0 def)

1 def M1 $\Rightarrow P_1 = 0.368$ $\Rightarrow$ A (1 def)

2 def M1 $\Rightarrow P_2 = 0.184$ $\Rightarrow$ CM (2 def)

3 def M1 $\Rightarrow P_3 = 0.061$ $\Rightarrow$ CM (3 def)

4 def M1 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.736

b. Muestra 2 (M2)

$c = 3$

$r = 4$

$np = 100(0.02) = 2$; $P_0 = 0.135$; $P_1 = 0.271$; $P_2 = 0.271$; $P_3 =$ [illegible]

2 def M1, 0 def M2 $\Rightarrow P_{20} = (0.184)(0.135) = 0.0248 \Rightarrow$ A (2 def)

2 def M1, 1 def M2 $\Rightarrow P_{21} = (0.184)(0.271) = 0.0498 \Rightarrow$ A (3 def)

2 def M1, 2 def M2 $\Rightarrow$ $\Rightarrow$ R (4 def)

3 def M1, 0 def M2 $\Rightarrow P_{30} = (0.061)(0.135) = 0.0082 \Rightarrow$ A (3 def)

3 def M1, 1 def M2 $\Rightarrow$ $\Rightarrow$ R (4 def)

Probabilidad de aceptación = 0.0628

$\therefore$ $P(A; 0.02) = 0.736 + 0.0828 = 0.8188 \simeq 0.819$

APÉNDICE 3 – TABLAS

TABLA K. LETRAS CLAVE PARA EL TAMAÑO DE MUESTRAS —MIL-STD-105 (ESTÁNDAR ABC)

Tamaño del lote o partida	Niveles de inspección generales		
	I	II	III
2-8	A	A	B
9-15	A	B	C
16-25	B	C	D
26-50	C	D	E
51-90	C	E	F
91-150	D	F	G
151-280	E	G	H
281-500	F	H	J
501-1 200	G	J	K
1 201-3 200	H	K	L
3 201-10 000	J	L	M
10 001-35 000	K	M	N
35 001-150 000	L	N	P
150 001-500 000	M	P	Q
500 001 y más	N	Q	R

PAQUETES ESTADISTICOS PARA LA FAMILIA IBM PC Y COMPATIBLES

Traducido de la primera edición en inglés de
STATISTICAL PACKAGES FOR THE IBM PC FAMILY

ISBN: 0-07-056320-9

ISBN: 84-7615-190-X
Depósito legal. M. 39.475/1987

Compuesto en Fernández Ciudad, S. L.
Impreso en LAVEL, Industria gráfica

De esta edición se han impreso 3 500 ejemplares en diciembre de 1987

PRINTED IN SPAIN - IMPRESO EN ESPAÑA

Contenido