UNIVERSITAT JAUME·I

MANUAL DE CONTROL ESTADÍSTICO DE CALIDAD: TEORÍA Y APLICACIONES

Pablo Juan Verdoy, Jorge Mateu Mahiques,
Santiago Sagasta Pellicer, Raúl Sirvent Prades



COL·LECCIÓ
«TREBALLS D'INFORMÀTICA I TECNOLOGIA»
Núm. 21

Col·lecció
«Treballs d'Informàtica i Tecnologia»
Núm. 21

MANUAL DE CONTROL ESTADÍSTICO DE CALIDAD: TEORÍA Y APLICACIONES

**Pablo Juan Verdoy, Jorge Mateu Mahiques,
Santiago Sagasta Pellicer, Raúl Sirvent Prades**

UNIVERSITAT
JAUME·I

BIBLIOTECA DE LA UNIVERSITAT JAUME I. Dades catalogràfiques

MANUAL de control estadístico de calidad : teoría y aplicaciones / Pablo Juan Verdoy ... [et al.]. — Castelló de la Plana : Publicacions de la Universitat Jaume I, D.L. 2006
p. : gràf. ; cm. — (Treballs d'Informàtica i tecnologia ; 21)
Bibliografia.
ISBN: 978-84-15443-85-8
1. Control de qualitat — Mètodes estadístics. I. Juan Verdoy, Pablo, coaut. II. Universitat Jaume I. Publicacions, ed. III. Sèrie.
658.562:519.2



Cap part d'aquesta publicació, incloent-hi el disseny de la coberta, no pot ser reproduïda, emmagatzemada, ni transmesa de cap manera, ni per cap mitjà (elèctric, químic, mecànic, òptic, de gravació o bé de fotocòpia) sense autorització prèvia de la marca editorial

© Del text: Els autors, 2006

© De la present edició: Publicacions de la Universitat Jaume I, 2006

Edita: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions
Campus del Riu Sec. Edifici Rectorat i Serveis Centrals. 12071 Castelló de la Plana
Fax 964 72 88 32
www.tenda.uji.es e-mail: publicacions@uji.es

ISBN: 978-84-15443-85-8

DOI: http://dx.doi.org/10.6035/INFiTEC.2006.21

ÍNDICE

1. Introducción 11
1.1. Conceptos de calidad 11
1.2. Apartados a los que se aplica la calidad 12
1.3. Evolución del control de calidad 13
1.4. Situación actual de la calidad 16
1.4.1. Gestión de la calidad 16
1.4.2. Calidad total 17
1.5. Costes que genera la calidad, costes que genera la no calidad 17
1.6. La normalización y la certificación en el marco de la calidad 18
1.7. Infraestructura para la calidad y la seguridad industrial 20
1.8. Incertidumbre de medida y calibración de equipos 20
1.9. Modelo europeo de gestión de la calidad 23
1.10. Equipos de mejora de la calidad 24

2.Técnicas estadísticas para el control de calidad 29
2.1. Descripción de una muestra 29
2.1.1. Introducción 29
2.1.2. Distribución de frecuencias 30
2.1.3. Métodos gráficos 32
2.1.4. Medidas descriptivas 34
2.1.5. Transformaciones lineales 48
2.2. Descripción conjunta de varias variables 50
2.2.1. Distribuciones de frecuencias bivariantes 51
2.2.2. Representación gráfica 52
2.2.3. Medidas de dependencia lineal 52
2.2.4 Recta de regresión 53
2.3. Espacio muestral y sucesos 60
2.3.1. Espacio muestral. Álgebra de sucesos 60
2.3.2. Frecuencias 63
2.3.3. Probabilidad 63
2.4. Variables aleatorias 71
2.4.1. Variables aleatorias. Función de distribución. Función de densidad .71
2.4.2. Modelos de probabilidad 77

2.5. Introducción a la indiferencia estadística 85
2.5.1. Introducción 85
2.5.2. Distribuciones en el muestreo 88
2.6. Estimación 94
2.6.1. Estimación puntual 94
2.6.2. Propiedades de los estimadores 95
2.6.3. Métodos de cálculo de los estimadores 97
2.6.4. Estimación por intervalos de confianza 99
2.6.5. Precisión y tamaño de la muestra 108

3. Control estadístico de la calidad. Diagramas de control 111
3.1. Gráficos de control por variables 111
3.1.1. ¿Qué es un gráfico de control? 111
3.1.2. Gráficos de control para variables 113
3.1.3. Gráficos $\bar{X}$-R 114
3.1.4. Gráficos $\bar{X}$-S 117
3.1.5. Gráfico X-σ 119
3.1.6. Gráfico de observaciones individuales y gráfico de rangos móviles 122
3.1.7. Otros tipos de gráficos de control por variables 124
3.1.8. Factores para la construcción de gáficos de control 130
3.1.9. Interpretación de los gráficos de control 131
3.1.10. Ejemplos 135
3.2. Gráficos de control por atributos 143
3.2.1. Diagrama P: Diagrama de control para el porcentaje o fracción de unidades defectuosas 144
3.2.2. Diagrama de control del número de unidades defectuosas. Diagrama «np» 152
3.2.3. Diagrama de control del número de defectos o diagrama «C» 158
3.2.4. Diagrama de control del número de defectos por unidad o diagrama «U» 164

4. Estudio de capacidad 177
4.1. Introducción 177
4.2. Índices de capacidad 179
4.3. Capacidad y potencial de proceso y máquina 182
4.4. Otros tipos de índices de capacidad 184
4.5. Ejecución de ensayos de capacidad y su cálculo 184
4.5.1.Ejemplo 185

5. Técnicas básicas para el análisis y mejora continua 191
5.1. Ideas generales de Mejora Continua 191
5.2. Herramientas de resolución de problemas y mejora continua 191
5.2.1. *Brainstorming* (Tormenta de ideas) 191
5.2.2. Diagrama causa-efecto 194
5.2.3. Histograma 197
5.2.4. Gráficos de control 201
5.2.5. Gráficos desarrollo 202
5.2.6. Diagramas de dispersión 202
5.2.7. Diagrama de Pareto 204
5.3. Diagrama de Pareto 204
5.3.1. Introducción 204
5.3.2. Concepto 205
5.3.3. El análisis de Pareto 205
5.3.4. Construcción paso a paso del Diagrama de Pareto 207
5.4. Otras herramientas 213
5.4.1. Diagrama de flujo 213
5.4.2. Diagramas de Gantt 214
5.4.3. Hoja de verificación e inspección 217
5.4.4. Los 5 por que 218
5.5. Herramientas avanzadas de gestión de calidad 219
5.5.1.Diagrama de afinidades 219
5.5.2. Diagrama de interrrelaciones 222
5.5.3. Diagrama de árbol 224
5.5.4. Diagrama matricial 228
5.5.5. Diagrama de decisiones de acción (PDPC) 229
5.5.6. Diagrama de flechas 230
5.5.7.Análisis factorial de datos 231
5.5.8. 6-Sigma 235

6. Normas de aceptación i rechazo ISO 10545 de baldosas cerámicas, (parte 1). Aplicación práctica de Poisson 241
6.1. Las normas de inspección y recepción por atributos. UNE EN 66-020-73 241
6.2. Definiciones y conceptos generales 241
6.3. Ejemplo de aplicación. Tablas de consulta 247
6.4. Consulta de curvas características 251
6.5. Protección del límite de calidad y NCA para series de lotes 252
6.6. Las normas de calidad de las baldosas cerámicas 254
6.7. Tipos de normas 254

6.8. Comparativa entre la normativa antigua y la actual 256
6.8.1. Normas (clasificación, características y marcado) 256
6.9. Norma UNE 67-087-85. Definiciones, clasificación, características y marcado 260
6.10. Normas de muestreo y condiciones de aceptación 261
6.11. Normas de ensayo de las características 267

7. Ejemplos y aplicación 269
7.1. Seguimiento de los ejercicios 269
7.2. Aplicaciones de control estadístico de la calidad 271
7.2.1. Descriptiva 272
7.2.2. Control estadístico de calidad 275
7.3. Fichero de datos para los ejercicios 291

8. Solución de los ejercicios 295

9. Ejercicios propuestos 331

10. Bibliografía 333

11. Anexo tablas 335
11.1. Tabla distribución normal tipificada 335
11.2. Tabla Poisson 336
11.3. Tabla binomial 339

PRÓLOGO

LA FINALIDAD de las sociedades modernas ha sido la búsqueda de la perfección, de la calidad de vida y del bienestar. La evolución durante el siglo XX de estas sociedades, ha hecho que se introdujera en la sociedad el término de calidad.

A principios del siglo XX, la introducción de tecnologías a las industrias, implicaba un aumento de la producción, pero también un aumento de los productos defectuosos. La calidad, en esta época, consistía en que no se produjeran piezas defectuosas.

En los últimos años la calidad ha adquirido una gran importancia a todos los niveles de la sociedad. Actualmente se define la calidad como satisfacción total de todas la partes del tejido de la sociedad. Las empresas cada vez más se han sumado a las políticas de calidad, ya no solo por el prestigio que otorga la concesión de una marca de calidad, sino por los grandes beneficios económicos que generan las inversiones en materia de calidad.

Los consumidores, cada vez más concienciados, exigen que sus productos cumplan una serie de especificaciones, y tengan por tanto unas marcas de calidad asignadas, que les garantice la satisfacción personal.

La necesidad de la matemática estadística en los procesos de control de calidad, se materializa en la mayoría de los casos, aunque especialmente cuando se trata de productos de alto valor añadido o de producciones muy elevadas donde es imposible controlar todos los números. El uso de los lotes, como muestras de la población, ha facilitado el trabajo de control en las empresas, consiguiéndose resultados mas que beneficiosos.

El desarrollo de todas estas técnicas estadísticas aplicadas al proceso productivo, ha motivado una necesidad de preparación en los técnicos encargados de calidad en las empresas. Generalmente estos trabajadores están fuera del entorno de la matemática o tienen poca preparación estadística. La claridad de los apuntes y la cantidad de ejemplos que se presentan, pueden ayudar a esta preparación.

Por tanto, el trabajo que el lector tiene en su mano, tiene como objetivo ser una herramienta de matemática estadística de usos múltiples. En primer lugar puede servir en asignaturas de matemáticas como ejemplo práctico de las teorías estadísticas,

en segundo lugar puede servir como libro de consulta para las empresas que apuestan por estas políticas de calidad, por la gran cantidad de herramientas estadísticas que se barajan y por último, puede servir como libro de texto de las asignaturas de calidad impartidas en los módulos de formación profesional, especialmente en los ciclos formativos de grado superior, donde el técnico cualificado puede tener responsabilidades en el apartado de calidad. Además puede servir de manual teórico y práctico para algunas asignaturas impartidas en diplomaturas y licenciaturas universitarias.

La organización del libro que aquí se presenta es la siguiente:

Un capítulo 1 de introducción a los conceptos de calidad, su importancia y su aplicación. El capítulo 2 aporta una introducción a la estadística necesaria para poder comprender todo el libro, desde su visión más técnica. El capítulo 3 presenta el desarrollo de los gráficos de control propiamente dicho, sin incluir en ellos los índices de capacidad y otras partes que serán desarrolladas posteriormente. El capítulo 4, como hemos dicho, es específico para los estudios de capacidad tanto de máquinas como de procesos. El capítulo 5 analiza un desarrollo de las técnicas básicas para el desarrollo del control de calidad. El capítulo 6 desarrolla, de forma teórica, los distintos conceptos necesarios para la calidad y muestreo en el campo de la cerámica.

El capítulo 7 es claramente una novedad en el campo de las aplicaciones de control de calidad estadístico, ya que se introduce un ejemplo y es desarrollado según unos pasos establecidos y que nos sirve como ejemplo de aplicación de todo lo explicado anteriormente. Finalmente aparece una serie de ejercicios propuestos como complemento a los ya resueltos.

Nuestro agradecimiento a la Unitat de Suport Educatiu, que fue, quien con el apoyo para realizar los distintos proyectos de mejora educativa, ha hecho posible que este libro tenga forma y pueda ser desarrollado. También nuestro agradecimiento al servicio de publicaciones de la UJI y a los revisores, cuyos comentarios han mejorado sensiblemente la versión inicial del libro.

LOS AUTORES

Castellón de la Plana, mayo de 2006

1 INTRODUCCIÓN

1.1. CONCEPTOS DE CALIDAD

TODOS HEMOS oído este concepto y en el entorno industrial es una de las palabras más manejadas en los últimos tiempos. Aunque la percepción de la calidad tiene una parte subjetiva y otra objetiva, se han propuesto varias definiciones más o menos acertadas.

Así, según la ASQC (American Society For Quality Control) «la calidad es el conjunto de propiedades y características de un producto o servicio que le confieren su aptitud para satisfacer unas necesidades expresadas o implícitas.»

Uno de los países que más ha trabajado el concepto de calidad, ha sido Japón, y el mejor exponente de la eficacia de la misma es el avanzado progreso económico logrado y la alta competitividad de sus productos en occidente. Dos personas de este país tienen editados varios libros sobre el tema, como son Genichi Taguchi y Kaorou Kurosawa. Las definiciones que ambos hicieron de calidad se relacionan con:

- La **pérdida** que la no calidad ocasiona al **usuario** del producto
- La **pérdida** que la no calidad ocasiona a la propia **empresa** productora
- La posibilidad de **diseñar, producir y servir** un producto lo más **útil y económico** posible.
- La **satisfacción** que ocasiona al usuario.

Una definición muy aceptada es:

CALIDAD ⇒ CONCORDANCIA ENTRE UTILIDADES Y EXPECTATIVAS

El significado de esta definición es que hay que fabricar un producto que satisfaga todo aquello que un usuario espera de él, lográndolo además con el menor coste económico.

Esta definición estaría aplicada sobre todo a la calidad de producto o servicio, pero no es el único tipo de calidad, ya que tanto o más importante es la gestión integral y to-

tal de la calidad, involucrando a todo el personal y todas las secciones de un sistema de producción. La nueva calidad se basa sobre todo en la prevención de la no calidad.

1.2. APARTADOS A LOS QUE SE APLICA LA CALIDAD

Para lograr la calidad tiene que haber un flujo de información entre lo que el cliente espera y lo que se le proporciona. Si esta comunicación no es correcta se incurrirá en falta de calidad. Los errores pueden ser debidos a:

- Diferencia entre lo que el cliente espera y lo que el productor cree que el cliente espera. Esta diferencia intenta evitarse mediante la **calidad de concepto.**
- **La calidad de concepto** vendrá determinada por los medios que el fabricante emplee en captar lo que el cliente espera.
- Diferencia entre lo que se quiere diseñar y lo que realmente diseñamos. La calidad aplicada a esta fase se denomina **calidad de diseño.**
- Se alcanzará la **calidad en el diseño** cuando el diseño del producto satisfaga las expectativas que el cliente tiene en él.
- Diferencias entre lo que se diseña y lo que se produce realmente. Se controlan mediante la **calidad de conformidad.**
- Este tipo de calidad se ocupa en verificar que lo que se produce esté de acuerdo con lo diseñado. Se logrará con un buen flujo de información entre diseño y fabricación.
- Diferencia entre lo que se produce y lo que se comunica a los clientes. Es la **calidad de servicio.**
- Es la respuesta del productor frente a un caso de fallo del producto, demanda de información sobre el producto y su uso, y la atención posventa.

Además de estos tipos de calidades, existen otros tipos que tienen su hueco en un plan de calidad. Algunas de ellas son:

- **Calidad de disponibilidad**
- **Calidad implícita**
- **Calidad especificada o explícita**
- **Calidad que sorprende**

Existen distintos trabajos de autores que nos pueden servir de ayuda para comprender los distintos conceptos de calidad, como Duncan (1974), Juran y otros (1990) y Deming (1992), junto con los distintos trabajos de Crosby (1979) e Ishikawa (1985). Respecto a trabajos más actuales que resumen los distintos conceptos tenemos el trabajo de De Domingo-Acinas, J. y Arranz-Molinero, A. (1997).

1.3.EVOLUCIÓN DEL CONTROL DE CALIDAD

A lo largo del tiempo, el concepto y la forma de llevar a cabo la calidad han ido evolucionando. Las fases, ordenadas de más antiguas a más modernas han sido:

TABLA 1: Fases de gestión de la calidad

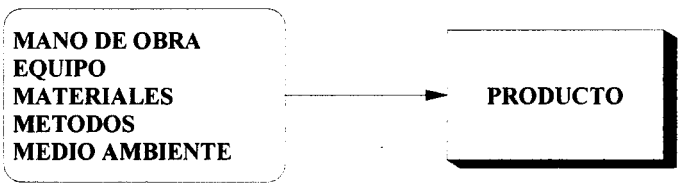
Indicadores de calidad	FASES DE LA GESTIÓN DE LA CALIDAD		
	1ª FASE Control de calidad	2ª FASE Aseguramiento de la calidad	3ª FASE Gestión de la calidad total
Finalidad principal	Conseguir la conformidad con las especificaciones	Lograr satisfacer necesidades de los clientes	Satisfacer los intereses de todas las personas relacionadas con la organización y la sociedad en general
Visión de la calidad	Problema que se ha de resolver	Problema que se ha de resolver actuando sobre él	Posibilidad de disponer de mayor capacidad competitiva
Métodos	Inspecciones y mediaciones de productos acabados	Sistemas de calidad. Control de procesos	Fijar objetivos claros y movilizar la organización
Responsables de la calidad	El departamento de control de calidad	Todos los departamentos	La totalidad de los componentes de la organización, sobre todo la dirección
Énfasis de la calidad	En el producto y en la producción	Todas la áreas	En el cliente externo, interno y la sociedad
Orientación	La calidad ha de comprobarse	La calidad ha de conseguirse	La calidad debe gestionarse
Época	Hacia 1920	A partir de 1950	A partir de 1970

La evolución del concepto de calidad a o largo del tiempo, ha supuesto también diferentes maneras de llevar el control de procesos y de productos.

Índice

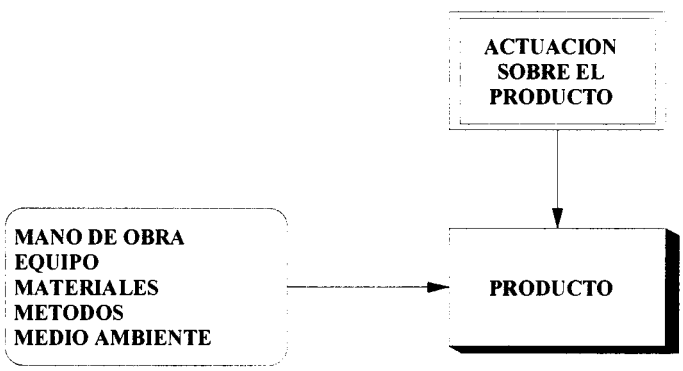
SIN CONTROL

En este sistema no hay ningún tipo de control ni sobre el proceso ni sobre el producto. Evidentemente salen a la venta igual productos buenos que malos y hay muchas reclamaciones



SISTEMAS DE DETECCIÓN

En este sistema se actúa viendo las no conformidades en el producto cuando ya está fabricado. Lo único que logramos es impedir que salga a la calle un producto no conforme a las especificaciones. Evidentemente hay que hacer algo diferente a lo inicialmente planificado con los no conformes (tirarlo, venderlo de saldo, reciclarlo, etc.)



SISTEMAS DE PREVENCIÓN

En este sistema, no sólo se actúa sobre el producto acabado, sino sobre el proceso, pero ambas actuaciones son independientes.

De esta manera, si el proceso se va perfeccionando salen menos unidades malas.

Por otra parte, el sistema de detección de defectos previene que salgan al mercado unidades no conformes (ver Ilustración 1).

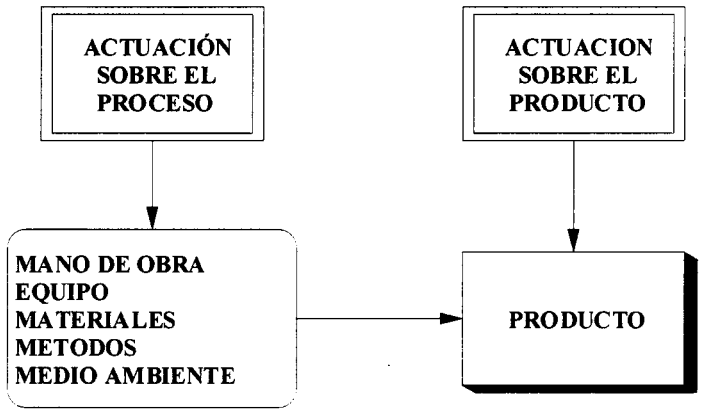


ILUSTRACIÓN 1: Sistemas de prevención

SISTEMA DE CONTROL

El paso siguiente es relacionar las dos actuaciones sobre el producto y sobre el proceso, estableciendo sistemas de información que relacionen los defectos con el proceso y permita actuar eficazmente para prevenir los problemas a partir de la información y la experimentación (ver Ilustración 2).

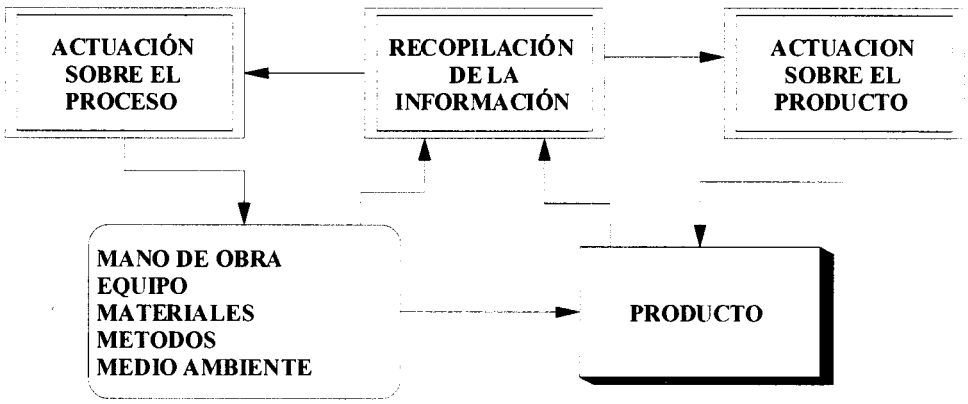


ILUSTRACIÓN 2: Sistemas de control

A lo largo del desarrollo industrial ha habido muchas aportaciones a la calidad. Entre estas aportaciones destacan las siguientes:

- Walter Shewart. Desarrolló el control estadístico de procesos *SPC,* cuyo objetivo es dominar el proceso.

- Edwards Deming. Impulsó el uso del control estadístico de procesos y realizó numerosas aportaciones que se encuentran en pleno vigor, que luego veremos.
- Joseph M. Juran. escribió su trilogía *Planificación, Control y Mejora de la Calidad.* Su filosofía era que la calidad se consigue por la mentalización y no por la inspección
- Kaoru Ishikawa. Fue el creador de los círculos de calidad para mejorar la calidad de los procesos. Inventó el diagrama causa-efecto que luego veremos.
- Masaaki Imai. Describe el método de mejora continua (Kaizen).
- Seuchi Nakajima. Escribió el tratado *Mantenimiento productivo total TPM.*
- Shigeo Shing. Desarrolló un sistema de maximizar el uso de equipos por medio del Cambio Rápido de Utillaje (SMED, cambio de útiles al minuto).
- Genichi Taguchi. Investigó una metodología para reducir costes y mejorar la calidad (Método de Taguchi DDE de diseño estadístico de experimentos).
- Philip B. Crosby. Elaboró un programa para mejorar la calidad que incluye las cuaro calidades: Definición, sistema de calidad, cero defectos y medición de la calidad.
- Taiichi Ohno. Desarrolló las técnicas de gestión de productos Justo a Tiempo (Just in Time). Elaboró una lista con diferentes tipos de despilfarro (7).
- Kiyoshi Suzaki. Impulsó la recogida de información mediante la Gestión visual in situ, partiendo del hecho de que las personas de la organización poseen capacidad de mejorar la calidad aportando sus soluciones.

1.4. SITUACION ACTUAL DE LA CALIDAD

Actualmente conviven dos conceptos de calidad diferentes:

- Sistemas de **gestión de la calidad**
- Gestión de **calidad total**

1.4.1. GESTIÓN DE LA CALIDAD

Definición: Es el conjunto de **actividades** de la función general de la dirección que determina

- la política de la calidad,
- los objetivos,
- las responsabilidades,

y se implanta por medios tales como la planificación de la calidad, el control y el aseguramiento de la calidad y la mejora de la calidad dentro del marco de un sistema de gestión.

Definición: Aseguramiento de la calidad. Es el conjunto **de acciones** planificadas y sistemáticas que son necesarias para proporcionar la confianza adecuada de que un producto o servicio va a satisfacer los requisitos dados sobre la calidad. El procedimiento para conseguir la calidad podemos verlo en la Ilustración 3.

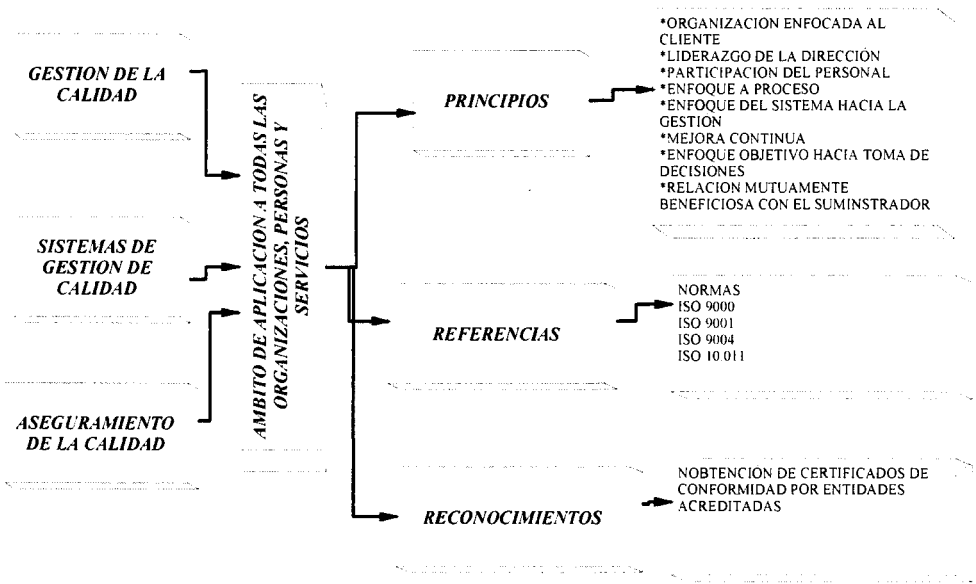


ILUSTRACIÓN 3: Procedimientos para conseguir la calidad

1.4.2. CALIDAD TOTAL

Definición: Es una estrategia de gestión que tiene como objetivo lograr de una manera equilibrada la satisfacción de los clientes, empleados, accionistas y sociedad en general. Representa la calidad de la gestión de empresa contemplada en su totalidad.

1.5. COSTES QUE GENERA LA CALIDAD. COSTES QUE GENERA LA NO CALIDAD

Este apartado pretende demostrar que la aplicación de los sistemas de calidad generan costes. Por otra parte, la no calidad en la empresa genera también costes, entre ellos el despilfarro que hemos visto en uno de los apartados anteriores. Como premisa fundamental, se deberá lograr un aumento de beneficio claro y medible cuando se pongan en marcha los planes de la calidad.

Los costes que genera la calidad pueden ser:

- Costes de conformidad (Inversiones necesarias para la aplicación de planes)
- Costes de prevención
- Costes de detección

La no calidad por su parte:

- Costes de no calidad internos
- Costes de no calidad externos (De cara al cliente).

Como en cualquier proceso humano, hay que determinar un equilibrio entre los costes necesarios para asegurar la calidad, definiendo la calidad óptima. Esta calidad es aquella que obtiene un producto con las especificaciones cumplidas que garantizan su aceptación, con un coste de aseguramiento de la calidad moderado. Tan malo es producir con baja calidad como ser excesivamente riguroso con la misma y tener unos costos que no se traducen en una mayor aceptación del producto.

La calidad mínima se da cuando se cumplen las características implícitas del producto. A partir de ahí, incrementando costes se puede producir con más calidad, pero a hay que sopesar la mayor aceptación del producto en función de los costes de la elevación de la calidad. Esto significa que la calidad óptima es aquella que genera el máximo beneficio.

1.6. LA NORMALIZACION Y LA CERTIFICACIÓN EN EL MARCO DE LA CALIDAD

Se define la normalización como:

> La actividad que tiene por objetivo instaurar un proceso a través del cual se **unifican criterios** respecto a determinadas materias y se posibilita la utilización de un lenguaje común en un determinado campo de actuación.

Por otra parte se define la certificación como:

> El procedimiento mediante el cual un determinado organismo acredita mediante un documento que un servicio, producto o proceso resulta conforme a una exigencias concretas.

En los apartados anteriores, hemos visto las normas que sirven para normalizar la calidad de una actividad. Hay diferentes entidades certificadoras y marcas de acreditación. Una de las más importantes es AENOR que dispone de varias marcas.

Entre las que destacan las siguientes:

Certifica la conformidad del sistema de gestión de la calidad de una empresa, respecto a los requisitos contenidos en las normas UNE-EN ISO 9000.

En un futuro no lejano, será imprescindible para conseguir la certificación de los productos de cada empresa.

Para conseguir el certificado de registro de empresa, es necesario realizar la evaluación y el control de las disposiciones de aseguramiento de la calidad de la propia empresa, lo cual no supone una certificación de sus productos, procesos o servicios. En la página siguiente se representa esquemáticamente el proceso de obtención de este certificado.



Marca de conformidad, que señala la conformidad del sistema de gestión medioambiental de una determinada empresa con la Norma ISO 14000, elaborada para establecer un sistema normalizado de gestión, que tenga la posibilidad de ser integrado en las actuaciones generales de la empresa, al margen del tipo de organización y tamaño de la misma.

Los productos que hayan obtenido esta marca llevan colocadas sobre los mismos unas estampillas, marcas, etiquetas u otros procedimientos como el que se muestra.



A través de la marca medio ambiente, se prueba, que un producto es conforme con las normas UNE de criterios ecológicos, y que ha sido obtenido por medio de un proceso de fabricación, cuyo efecto medioambiental se encuentra perfectamente controlado.

Los productos que hayan obtenido esta marca llevan colocadas sobre los mismos unas estampillas, marcas, etiquetas u otros procedimientos como el que se muestra.

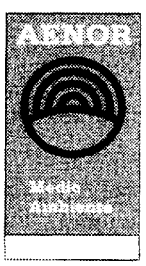


ILUSTRACIÓN 4: Sellos de calidad

En la siguiente ilustración, se representa el esquema para lograr el certificado de registro de empresa de AENOR. Este certificado es diferente de la obtención de la certificación de los productos o servicios, que requiere otra dinámica complementaria.

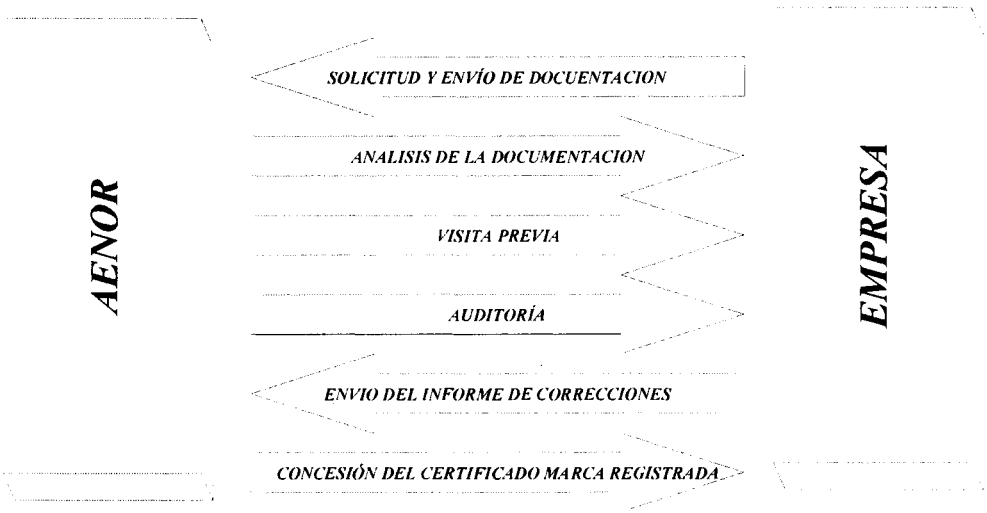


ILUSTRACIÓN 5: Relación AENOR - Empresa

1.7. INFRAESTRUCTURA PARA LA CALIDAD Y LA SEGURIDAD INDUSTRIAL

Con el objetivo de garantizar la calidad de los productos y servicios, es necesario que los siguientes organismos se encuentren coordinados y acreditados para realizar la verificación con unos criterios comunes y normalizados. Esto trae consigo una estructura compleja.

- Todas las entidades de certificación
- Laboratorios de ensayo
- Entidades auditoras y de inspección
- Organismos de control
- Verificadores medio ambientales

Existen muchas entidades de normalización, certificación y acreditación, tanto a nivel del estado como de la comunidad europea.

Merece la pena destacar AENOR, IQNET (red europea) EQS, IONET (Red internacional) BUREAU VERITAS, ENAC, CEN, ISO, ETC.

1.8. INCERTIDUMBRE DE MEDIDA Y CALIBRACIÓN DE EQUIPOS

Asegurar la calidad implica que hay que comprobarla. Generalmente se tienen que tomar medidas con aparatos adecuados para comprobar que se cumplen los re-

quisitos (por ejemplo, una longitud con pie de rey). Ya que vamos a tomar decisiones sobre si un producto o proceso cumple los requisitos de calidad en base a estas medidas, es necesario asegurarse que se reaizan correctamente y sin error.

Por esta razón, en todo sistema de calidad debe de hacerse referencia a la identidad y conservación de patrones de referencia y de mantenimiento de los equipos para que sus medidas aseguren la exactitud.

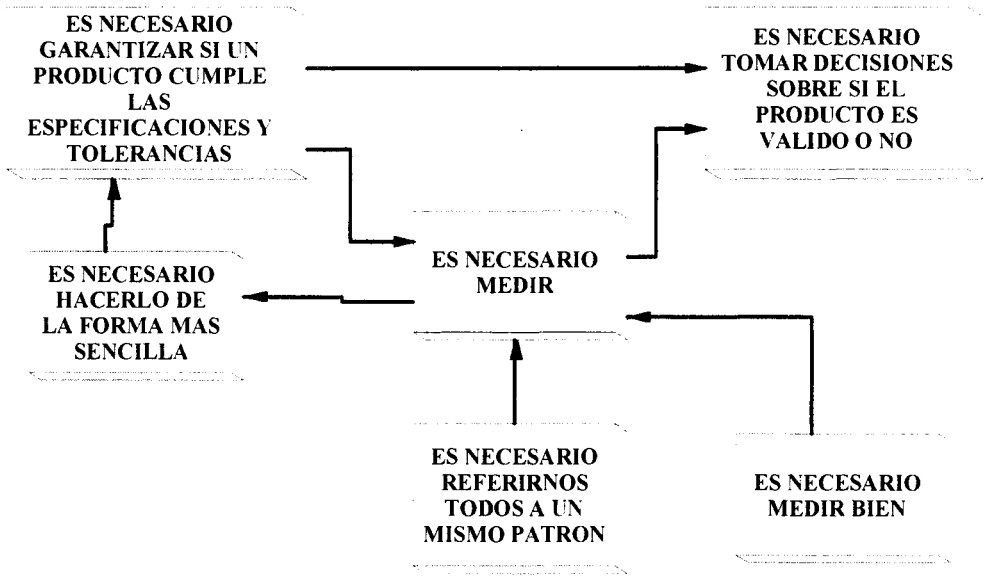


ILUSTRACIÓN 6: Necesidad de medir

En esencia significa que cuando hagamos una medida, es necesario asegurarse que nuestro equipo está contrastado con un patrón correcto, y del que se conoce toda su historia de calibraciones respecto de los patrones internacionales de medida.

Las normas UNE definen la calibración como el conjunto de operaciones que permiten establecer en condiciones especificadas, la relación existente entre los valores indicados por un instrumento de medida o sistema de medida y los valores correspondientes de una magnitud obtenidos a partir de un patrón de referencia.

El resultado de una calibración sirve para:

- Estimar los errores de indicación de un instrumento o sistema de medida o la asignación de valores a las marcas de las escalas
- Plasmar en un documento (certificado de calibración) el registro de la propia operación de calibración
- Se obtiene un factor de corrección o una curva de calibración que permite obtener resultados válidos con nuestro instrumento o equipo.

En todo el mundo, las unidades de medida deben de tener idéntica definición (por ejemplo, el metro), y además las definiciones de estas unidades de medida se han de materializar en patrones internacionales cuyo valor no debe de diferir en más de un valor definido (por ejemplo, el patrón de metro es una barra de platino iridio con dos marcas hechas en ella, y cuyo valor es de 1.00000 m con un error especificado (por ejemplo de +- 0.000001) y que además se debe de conservar en unas condiciones de temperatura por la dilatación.

Los patrones de medida tienen varios niveles, entre los que destacan:

- Patrón internacional (como el metro antes definido)
- Patrón nacional
- Patrón de referencia
- Patrón de trabajo
- Material de referencia. Este material está medido y se especifica el valor exacto. Con nuestros instrumentos lo podemos medir y de esta manera saber el error que cometemos
- Instrumentos de medida

El proceso mediante el cual se atraviesa el camino entre la definición de la unidad y los patrones o instrumentos que usamos en el día a día se denomina diseminación de la unidad.

La cadena ininterrumpida de comparaciones entre patrones incluyendo las incertidumbres de cada paso, se denomina cadena de trazabilidad.

En este apartado he hecho referencia a la incertidumbre de una medida. Las normas UNE-EN definen este concepto como la estimación que determina el intervalo de valores en el que se ubica, con una probabilidad elevada, el valor verdadero de la magnitud medida.

Cuando tomamos una medida de una magnitud (por ejemplo la longitud del lado de una baldosa), usamos un instrumento que tiene una cierta incertidumbre (por ejemplo, un pie de rey cuya incertidumbre es ± 0.0016mm). Además de la incertidumbre, tenemos que tener en cuenta la tolerancia de la medida. Supongamos que se trata de medir piezas de 10.5mm de calibre, con una tolerancia de ±0.01mm (ver Ilustración 7).

Esto implica que la pieza debe de medir entre 10.51 y 10.49mm. Pero como el aparato de medida nos puede cometer un error de 0.0016, tendremos que descontarlo de los extremos que nos hemos impuesto para la medida de la baldosa:

Tolerancia superior = 10.49+0.0016 =10.4916 mm

Tolerancia inferior = 10.51-0.0016 = 10.5084 mm

Por lo tanto la medida de la pieza con ese instrumento será válida si el valor obtenido está comprendido entre esos dos valores.

$$T_{as} = T_s - I = 10.51 - 0.0016 = 10.5084$$

$$T_{ai} = T_i + I = 10.49 + 0.0016 = 10.4916$$

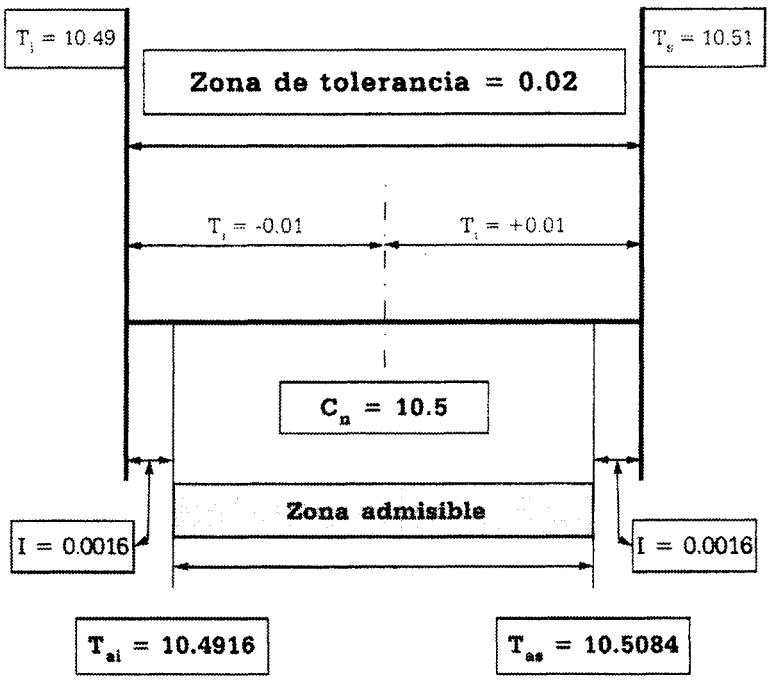


ILUSTRACIÓN 7: Gráfico de tolerancias

Es evidente que los patrones de calibración, materiales de referencia y equipos de medida se deben de mantener en perfecto estado para garantizar la precisión original. (por ejemplo, un pie de rey con los dos palpadores con golpes y muescas, no medirá ni tendrá la precisión de cuando estaba nuevo).

La calibración no es una operación única, sino que se debe de definir la periodicidad con que se debe de llevar a cabo, y el mantenimiento de los certificados de calibración correspondientes en un registro.

1.9. MODELO EUROPEO DE GESTION DE LA CALIDAD

Es un modelo que se basa en la autoevaluación de la organización. Para ello tiene marcados unos criterios para su realización y la asignación de unas puntuaciones a cada sector. De esta manera se puede saber cuales son los puntos fuertes y las debilidades.

Este modelo tiene nueve criterios divididos en 5 criterios y 4 resultados. Para la autoevaluación de la organización se designan equipos heterogéneos de trabajo, que pasan un cuestionario estándar y asignan puntuaciones.

Una vez conocida mediante la autoevaluación cuáles son las áreas débiles, se confecciona un plan de actuación determinado en el tiempo. Pasado este tiempo se efectúa una nueva evaluación, valorando si se ha mejorado. El proceso sigue indefinidamente, encontrándonos de nuevo en un bucle de MEJORA CONTINUA.

El modelo de autoevaluación se puede resumir en el siguiente esquema (Ilustración 8):

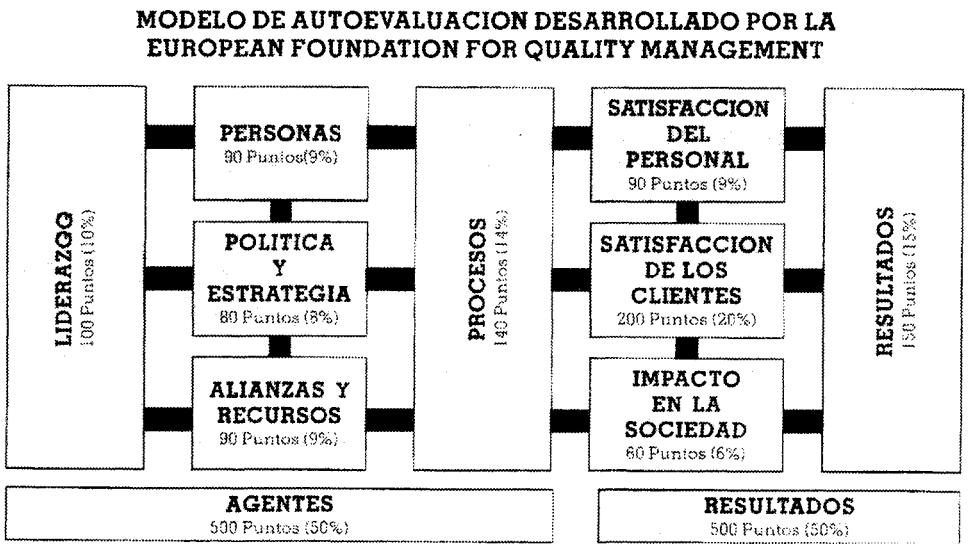


ILUSTRACIÓN 8: Modelo de autoevaluación

Este sistema está menos implantado que las normas de la serie ISO, por lo que no vamos a profundizar más en él. Simplemente decir que en cada apartado de los agentes y de los resultados, existe un cuestionario muy concreto para asignarle una puntuación. Entre todas las puntuaciones, y según el porcentaje de participación, se obtiene la calificación global de la organización. Aquellos factores que peores puntuaciones obtienen, son los que tienen que mejorarse y en consecuencia se elaboran planes concretos de actuación.

1.10. EQUIPOS DE MEJORA DE LA CALIDAD

Una constante de todos los sistemas de calidad es la mejora continua. Esta mejora no se logra de una manera mágica, sino que se tiene que implicar una serie de personas en la consecución de los objetivos marcados.

La norma ISO define la mejora de la calidad como *las acciones que se toman en una organización para aumentar la eficacia y el rendimiento de las actividades y*

procesos con el fin de aportar ventajas añadidas tanto a la organización como a los clientes.

Una mejora se considera consolidada cuando es posible mantener en condiciones normales la actividad, proceso o sistema mejorado en la situación controlada durante un largo período de tiempo. Esto significa que la mejora debe de ser permanente.

> La mejora no puede ser fruto de la improvisación, y se deberán establecer procedimientos en los que habrá consultas, análisis de todo tipo y propuestas de actuación. Así mismo se deberán emplear herramientas adecuadas que ayuden a la consecución de este fin.

Las normas ISO definen un proceso, de manera que se puede ver en un esquema los elementos de la organización donde se podrá actuar (Ilustración 9):

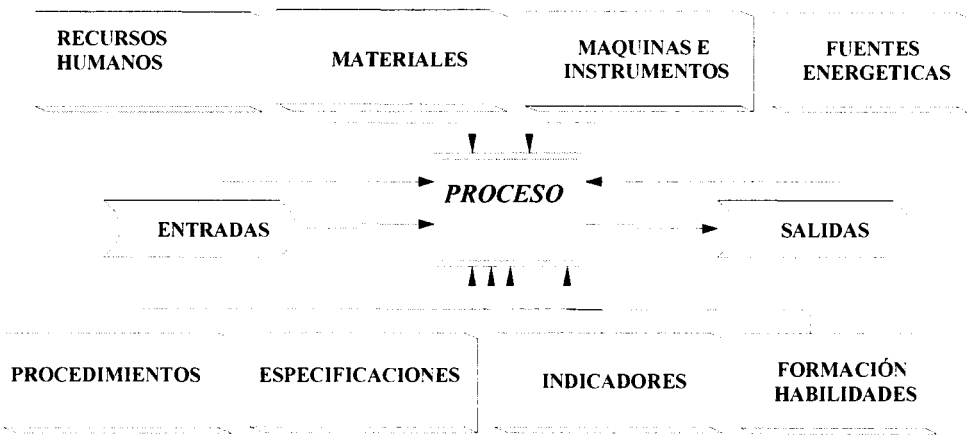


ILUSTRACIÓN 9: Proceso de definición de norma ISO

La misión de la mejora es analizar los procesos y determinar qué actividades del mismo generan costes sin añadir beneficio, y qué modificaciones se pueden hacer para aumentar la competitividad del mismo.

Para lograr estos objetivos se deben de

- Establecer planes de mejora
- Crear equipos que lleven a cabo las mejoras
- Emplear herramientas y técnicas de análisis aplicables a la mejora y resolución de problemas y procesos

Los objetivos de un equipo de mejora son:

- Analizar y mejorar los procesos
- Proponer soluciones ante desviaciones o problemas que se plantean
- Levar a cabo mejoras que aumenten la satisfacción de los clientes
- Favorecer el desarrollo personal, el compromiso con la organización y la plena satisfacción de los componentes del equipo

La secuencia para la creación y el funcionamiento de un equipo de mejora puede verse en las dos siguientes Ilustraciones:

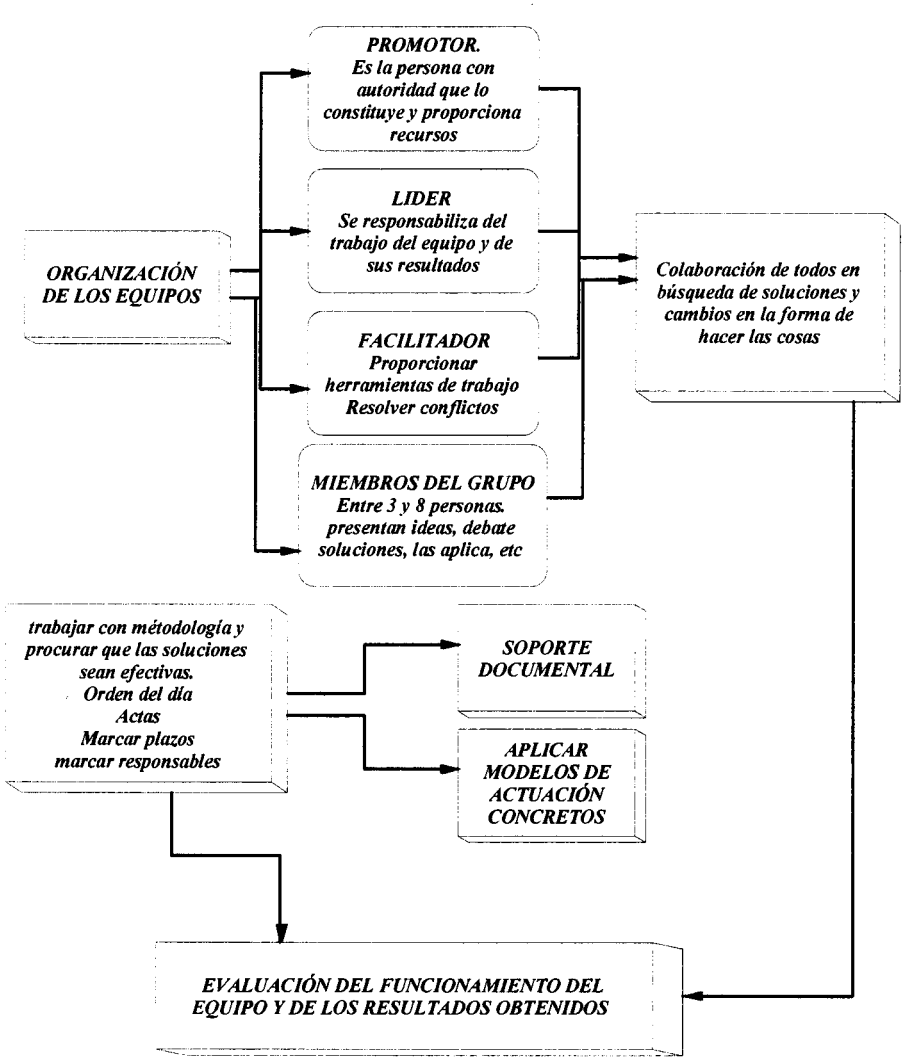


ILUSTRACIÓN 10: Creación y funcionamiento de equipo de mejora (1)

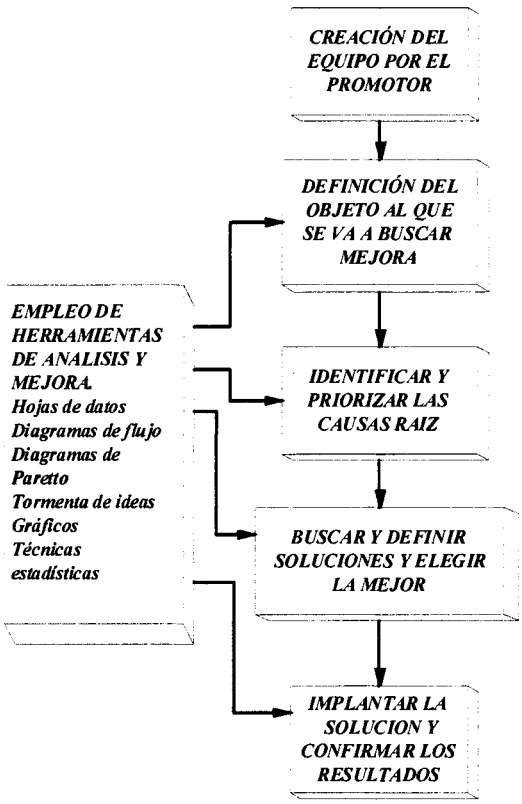


ILUSTRACIÓN 11: Creación y funcionamiento de equipo de mejora (2)

2 TÉCNICAS ESTADÍSTICAS PARA EL CONTROL DE CALIDAD

2.1. DESCRIPCIÓN DE UNA MUESTRA

2.1.1. INTRODUCCIÓN

LA ESTADÍSTICA es la ciencia que se encarga de la recopilación, representación y el uso de datos sobre una o varias características de interés para, a partir de ellos, tomar decisiones o extraer conclusiones generales.

Los pasos aconsejados para determinar un modelo estadístico son:

Paso 0: Planteamiento del problema en términos precisos: ámbito de aplicación (población) y característica(s) a estudio (variable(s)).

Paso 1: Recogida de datos de la población de interés (muestreo).

Paso 2: Organización, Presentación y Resumen de los datos (o de la muestra). (*Estadística descriptiva*).

Paso 3: Confección de modelos matemáticos. (*Teoría de la probabilidad*).

Paso 4: Obtener conclusiones generales o verificar hipótesis. (*Inferencia estadística*).

Estadística descriptiva: es la parte de la estadística que se encarga de organizar, resumir y dar una primera descripción (sin conclusiones generales) de los datos.

Algunos conceptos de interés también son:

Población: Es el conjunto de individuos o entes sujetos. Algunas poblaciones son finitas y pueden conocerse; otras pueden ser infinitas y abstractas.

Variable: Característica que estamos analizando. Las variables se suelen denotar por letras mayúsculas: X, Y,...

Tipos de variables:

Cualitativas o Categóricas: aquellas que no son medibles numéricamente. Expresan cualidades o categorías.

Cuantitativas: aquellas que son medibles, es decir sus observaciones tienen carácter numérico. Estas se dividen a su vez en:

Discretas: toman valores en un conjunto finito numerable.

Continuas: toman valores en un conjunto no numerable (los números reales o un intervalo).

Nota: La distinción entre variables discretas y continuas es más teórica que práctica, puesto que la limitación de los aparatos de medida hace que todas las variables se comporten como discretas cuando se pretende observarlas. De momento haremos más flexible el concepto de variable continua considerando continua a aquella variable que toma un gran número de valores diferentes.

Muestra: es un subconjunto finito de elementos seleccionados de la población.

Tamaño muestral: número de observaciones en la muestra. Habitualmente se denotará por n.

Dato: cada valor observado de la variable. Si representamos por X a la variable, representaremos por x_i cada dato diferente observado en la muestra, el subíndice i indica el lugar que ocupa si los ordenamos de menor a mayor.

Denotaremos por k al número de valores distintos.

2.1.2. DISTRIBUCIONES DE FRECUENCIAS

Consideramos las siguientes definiciones:

Frecuencia absoluta (n_i): es el número de veces que se repite un determinado valor (x_i) de la variable.

Propiedad: la suma de todas las frecuencias absolutas es igual al tamaño muestral.

Este tipo de frecuencias no son comparables con las obtenidas en otras muestras de distinto tamaño.

Frecuencia relativa (f_i): es igual a la frecuencia absoluta dividida por el número total de datos, es decir por el tamaño muestral $f_i = n_i/n$.

Propiedad: la suma de todas las frecuencias relativas es igual a la unidad.

Frecuencia absoluta acumulada (N_i): Informa sobre el número de datos que hay igual o inferiores a uno determinado. Se calcula de la siguiente forma:

$$N_i = \sum_{j=1}^{i} n_j = N_{i-1} + n_i$$

Propiedad: La última frecuencia acumulada absoluta es el tamaño muestral.

Frecuencia relativa acumulada (F_i): Es el resultado de dividir cada frecuencia absoluta acumulada por el número total de datos:

$$F_i = \frac{N_i}{n} = \sum_{j=1}^{i} f_j$$

Propiedad: La última frecuencia relativa acumulada es la unidad.

TABLA DE FRECUENCIAS

Llamamos así a una tabla conteniendo el conjunto de diferentes valores que ha tomado una variable (los datos sin repetir) ordenados de menor a mayor con sus correspondientes frecuencias.

DISTRIBUCIONES DE FRECUENCIAS AGRUPADAS

En aquellos en que los valores distintos que toma la variable son muchos, es decir k es grande, la tabla obtenida es muy poco manejable y por tanto poco clarificadora. Esto nos va a ocurrir frecuentemente en el caso en que la variable a estudiar sea continua. La solución es agrupar los diferentes valores de la variable en intervalos o intervalos de clase. Teniendo en cuenta que lo que ganamos en manejabilidad lo perdemos en información, con lo que los resultados serán aproximados.

Agrupar en intervalos de clase consiste en agrupar los datos en un número relativamente pequeño de intervalos que cumplan:

a) No se superpongan entre sí, de forma que no exista ambigüedad con respecto a la clase a que pertenece una observación particular.

b) Cubran todo el rango de valores que tenemos en la muestra.

Llamaremos:

A las fronteras del intervalo, *límites inferior* y *superior* de la clase y los denotaremos por L_{i-1}, L_i.

Marca de clase (c_i) al punto medio del intervalo, es decir, al promedio aritmético entre el límite inferior y superior:

$$c_i = \frac{L_i + L_{i-1}}{2}$$

Es el valor que tomamos como representativo.

Amplitud (a_i) a la diferencia entre el extremo superior e inferior: $a_i = L_i - L_{i-1}$.

Al número de observaciones de una clase se le llama frecuencia de clase (n_i). Si dividimos esta frecuencia por el número total de observaciones, obtenemos la frecuencia relativa de clase (f_i), y del mismo modo que lo hacíamos para datos sin agrupar definiríamos N_i, y F_i.

COMO CONSTRUIR UNA DISTRIBUCIÓN DE FRECUENCIAS AGRUPADA EN INTERVALOS

1. Empezamos determinando el recorrido de la variable o rango de valores que tenemos en la muestra. Se define como la diferencia entre el mayor y el menor valor de la variable, $Re = x_{max} - x_{min}$

2. Número de clases: depende del tamaño de la muestra. Para muestras de tamaño moderado, n<50, se suele elegir un número de clases igual a $\sqrt{n}$, o bien se usa la formula de Sturtges, (se toma el resultado de calcular el logaritmo de n , dividir por el logaritmo de 2 y sumar 1: $m = \frac{\log(n)}{\log(2)} + 1$); en general el número de clases no debe sobrepasar de 15 o 20, en casos de muestras muy grandes.

3. Determinamos la amplitud de los intervalos. Es más cómodo que la amplitud de todas las clases sea la misma (siempre que sea posible),

$$a_i = \frac{Re}{n^o \, de \, intervalos}$$

Tomaremos como regla, a no ser que se indique lo contrario, coger el intervalo cerrado por la izquierda y abierto por la derecha.

2.1.3. MÉTODOS GRÁFICOS

La forma de la distribución de frecuencias se percibe más rápidamente y quizás se retiene durante más tiempo en la memoria si la representamos gráficamente.

A) FRECUENCIAS NO ACUMULADAS

DIAGRAMA DE BARRAS (Ilustración 12): Es la representación gráfica usual para variables cuantitativas sin agrupar o para variables cualitativas. En el eje de or-

denadas representamos los diferentes valores de la variable (x_i). Sobre cada valor levantamos una barra de altura igual a la frecuencia (absoluta o relativa).

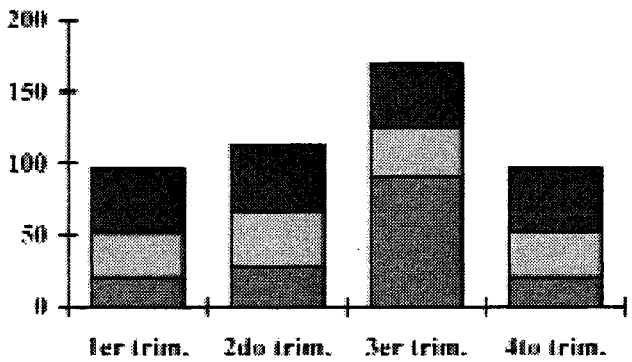


ILUSTRACIÓN 12: Diagrama de barras

DIAGRAMA DE SECTORES O DE PASTEL: Es el más usual en variables cualitativas. Se representan mediante círculos. A cada valor de la variable se le asocia el sector circular proporcional a su frecuencia.

Para hallar el ángulo usamos la siguiente proporción: al tener una circunferencia 360°, el cociente entre la frecuencia absoluta (o relativa) total y la frecuencia absoluta (o relativa) que queramos representar será igual al cociente entre los 360° de la circunferencia y el ángulo a determinar, así:

$$\frac{n}{n_i}=\frac{360^\circ}{\alpha} \qquad \frac{1}{f_i}=\frac{360^\circ}{\alpha}$$

donde α es el ángulo a determinar.

PICTOGRAMA: Se usa también para variables cualitativas, expresan con dibujos alusivos al tema de estudio las frecuencias de las modalidades de la variable. Estos gráficos se hacen representando en diferentes escalas el mismo dibujo. La escala de los dibujos tiene que ser tal que el área de cada uno de ellos sea proporcional a la frecuencia de la modalidad que representa.

HISTOGRAMA (Ilustración 13): Es la representación gráfica equivalente al diagrama de barras para datos agrupados, en el eje de ordenadas representarnos las clases y levantarnos sobre cada clase rectángulos unidos entre sí de altura igual a la frecuencia de la clase (absolutas o relativas).

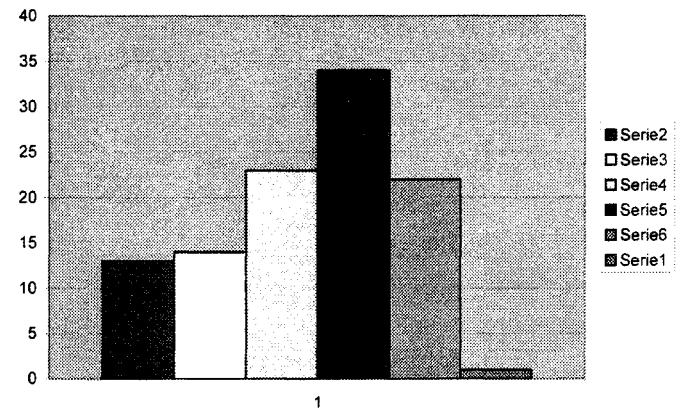


ILUSTRACIÓN 13: Histograma

El histograma o diagrama de barras proporcionan mucha información respecto a la estructura de los datos (y si la muestra es representativa de la población): el valor central de la distribución, su dispersión y la forma de la distribución. Cuando nos encontramos en distribuciones donde los intervalos no tienen la misma amplitud, las barras del histograma tienen que tener un área proporcional a la frecuencia que queramos representar.

B) FRECUENCIAS ACUMULADAS

POLÍGONO DE FRECUENCIAS: Es la representación habitual para datos cuantitativos agrupados de las frecuencias acumuladas (absolutas o relativas), mediante puntos se representan las frecuencias en el eje de ordenadas y la marca de clase en el de abscisas. Después se unen estos puntos por trozos de rectas.

2.1.4 MEDIDAS DESCRIPTIVAS

Para datos cualitativos, la distribución de frecuencias proporciona un resumen conciso y completo de la muestra, pero para variables cuantitativas puede complementarse este resumen utilizando medidas descriptivas numéricas extraídas de los datos.

Las medidas descriptivas son valores numéricos calculados a partir de la muestra y nos resumen la información contenida en ella.

2.I.4.1. MEDIDAS DE POSICIÓN

Nos dan el valor que ocupa una determinada «posición» respecto al resto de la muestra.

A) MEDIDAS DE TENDENCIA CENTRAL

Nos proporcionan el centro de la distribución de frecuencias, es un valor que se puede tomar como representativo de todos los datos. Hay diferentes caminos para definir el «centro» de las observaciones en un conjunto de datos. Por orden de importancia, son:

a) MEDIA ARITMÉTICA ($\bar{x}$) (o simplemente media): Es el promedio aritmético de las observaciones, es decir, el cociente entre la suma de todos los datos y el número de ellos, es decir:

$$\bar{x}=\frac{\sum_i x_i n_i}{n}$$

Si los datos están agrupados utilizamos las marcas de clase, es decir c_i en vez de x_i. Es la medida de centralización más importante.

PROPIEDADES

1. La suma de las diferencias de los valores de la variable con respecto a la media es cero.

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

2. La suma de las desviaciones al cuadrado de los valores de la variable respecto a una constante k cualquiera, se hace mínima cuando esa constante es la media. Es decir:

$\sum_i (x_i-\bar{x})^2 n_i \leq \sum_i (x_i-k)^2 n_i$, para cualquier constante k.

b) MEDIANA (Me): es el valor que separa por la mitad las observaciones ordenadas de menor a mayor, de tal forma que el 50% de estas son menores que la me-

diana y el otro 50% son mayores. Si el número de datos es impar la mediana será el valor central, si es par tomaremos como mediana la media aritmética de los dos valores centrales.

Distinguiremos entre distribuciones no agrupadas y distribuciones agrupadas.

DISTRIBUCIONES NO AGRUPADAS

Los pasos para el cálculo de Me en el caso de distribuciones no agrupadas son los siguientes:

1) Calculamos n/2

2) Buscamos en la tabla $N_{i-1}<n/2 \leq N_i$ (es decir aquel valor cuya frecuencia acumulada más se acerca a n/2 por arriba)

Si $n/2<N_i$ la mediana es aquel valor de la variable cuya frecuencia acumulada es N_i es decir:

$$Me = x_i \text{ tal que } n/2 < N_i$$

Si $n/2=N_i$ la mediana será la media aritmética de aquellos valores cuya frecuencia acumulada es N_i y N_{i+1} respectivamente, es decir:

$$Me = (x_i+x_{i+1})/2, \text{ tal que } N_i=n/2$$

DISTRIBUCIONES AGRUPADAS

Los pasos para el cálculo de Me en el caso de distribuciones agrupadas son los siguientes:

1) Se calcula n/2

2) Buscamos en la tabla el intervalo, $[L_{i-1}, L_i)$, que cumple $N_{i-1}<n/2<N_i$ (a este intervalo lo llamamos intervalo mediano)

3) A continuación para encontrar la mediana, aplicaremos la siguiente fórmula:

$$Me=L_{i-1}+\frac{\left(\frac{n}{2}-N_{i-1}\right)a_i}{n_i}$$

El razonamiento es el siguiente: la frecuencia acumulada hasta el intervalo anterior al mediano es N_{i-1}; para llegar a la mitad de los datos, es decir, n/2 necesitamos tomar $n/2 - N_{i-1}$ del intervalo mediano, el cual tiene n_i datos repartidos en una amplitud a_i; como a cada dato le corresponde una longitud a_i/n_i, a los $n/2 - N_{i-1}$ datos les corresponderá:

$$\frac{\left(\frac{n}{2}-N_{i-1}\right)a_i}{n_i}$$

PROPIEDAD: La mediana hace mínima la suma de todas las desviaciones absolutas de los valores de la variable respecto a una constante k cualquiera. Es decir,

$$\sum_i |x_i - Me| n_i \leq \sum_i |x_i - k| n_i$$

para cualquier constante k.

c) MODA (Mo): es el valor de la variable que más veces se repite, es decir, aquella cuya frecuencia absoluta es mayor. No tiene porque ser única.

Distinguiremos entre distribuciones no agrupadas y agrupadas.

DISTRIBUCIONES NO AGRUPADAS

Simplemente observamos en la columna de las frecuencias absolutas y aquel o aquellos valores (no tiene porque ser única) de la variable a los que corresponde la mayor frecuencia será la moda. Cuando encontramos dos modas decimos que es una distribución bimodal, (n) tres, trimodal, etc.

DISTRIBUCIONES AGRUPADAS

Es importante distinguir aquí también entre intervalos de igual amplitud, o distribuciones de frecuencias donde los intervalos no tengan la misma amplitud.

Intervalos de igual amplitud

Observando las frecuencias absolutas, determinamos el intervalo con mayor frecuencia $[L_{i-1}, L_i)$, a este intervalo le llamaremos intervalo modal.

A continuación para encontrar la moda aplicamos la siguiente fórmula:

$$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} a_i$$

El razonamiento es el siguiente: Consideramos los intervalos anterior y posterior al modal, con frecuencias n_i y n_{i-1}. Si estas frecuencias son iguales, la moda sería el centro del intervalo modal, en caso contrario, la moda estaría más cerca de aquel intervalo contiguo cuya frecuencia es mayor, es decir, las distancias de la moda a los intervalos contiguos son inversamente proporcionales a las frecuencias de dichos intervalos.

Como consecuencia $Mo = L_{i-1+m}$ con:

$$\frac{m}{a_i - m} = \frac{n_{i+1}}{n_{i-1}}$$

Despejando m y sustituyendo obtenemos la fórmula anterior.

Intervalos de distinta amplitud

Tendremos que hallar en primer lugar la densidad de frecuencia de cada intervalo que se define como: $d_i = n_i / a_i$.

El intervalo modal $[L_{i-1}, L_i)$ será ahora el intervalo con mayor densidad de frecuencia y para hallar la moda de nuevo aplicamos la fórmula anterior pero sustituyendo las frecuencias por las densidades de frecuencia:

$$Mo = L_{i-1} + \frac{d_{i+1}}{d_{i-1} + d_{i+1}} a_i$$

COMPARACIÓN ENTRE MEDIA, MODA Y MEDIANA

Estas tres medidas de tendencia central son las más importantes y las más usuales. ¿Cuando utilizamos una u otra?

La media es la mejor por que utiliza toda la información, es decir, tiene en consideración todos los valores de la distribución, tiene también como ventaja que es única. Como desventaja más importante está el hecho de que es muy sensible a la presentación de datos anómalos o atípicos que hacen que la media se desplace ha-

cia ellos y como consecuencia no es recomendable usar la media en estos casos. Otra desventaja es que puede no coincidir con uno de los valores de la variable.

La mediana utiliza menos información que la media puesto que no depende de los valores de la variable sino del orden que ocupa. Por este motivo tiene la ventaja de no estar afectada por observaciones extremas. La mediana la utilizaremos cuando la media falle. Otra ventaja frente a la media es que es un valor de la variable.

La moda es la que menos información maneja y por tanto la peor. Tiene la ventaja de que puede calcularse incluso para datos cualitativos. Otra desventaja es que no es única.

Si la distribución es simétrica y campaniforme coinciden. En el caso de distribuciones campaniformes, la mediana se encuentra entre la media y la moda (algo más cerca de la media). La siguiente relación nos permite calcular una de estas medidas de centralización en función de las otras:

$$Mo \approx 3Me - 2\bar{x}$$

Las siguientes medidas de centralización tienen un significado estadístico menos intuitivo y se utilizan en situaciones más específicas:

d) MEDIA GEOMÉTRICA (G): Se define como la raíz n-ésima del producto de los n datos. Así:

$$G=\sqrt[n]{\prod_i x_i^{n_i}}$$

PROPIEDAD: El logaritmo de la media geométrica es igual a la media aritmética del logaritmo de los valores de la variable.

La media geométrica se suele utilizar para promediar porcentajes, tasas y números índices.

e) MEDIA ARMÓNICA (H): Se define como el recíproco de la media aritmética de los recíprocos de los datos:

$$H=\frac{n}{\sum_i \frac{1}{x_i} n_i}$$

Se suele utilizar para promediar velocidades, rendimientos y en general magnitudes expresadas en términos relativos.

Si los datos están agrupados, para calcular las medidas anteriores utilizamos las marcas de clase, es decir x_i indicará el punto medio del intervalo.

La relación existente entre la media, la media geométrica, y la media armónica viene dada por:

$$H \leq G \leq \bar{x}$$

B) MEDIDAS DE TENDENCIA NO CENTRAL: CUANTILES

Los cuantiles son valores de la distribución que la dividen en partes iguales, es decir, en intervalos, que comprenden el mismo número de valores. Los más usados son los cuartiles, los deciles y los percentiles.

a) PERCENTILES (P_i): Son 99 valores que dividen en cien partes iguales el conjunto de datos ordenados.

El percentil de orden p (P_p) es el menor valor superior al p% de los datos (ordenados de menor a mayor los datos, deja el p% de datos por delante). La forma más cómoda de calcularlos es a partir de las frecuencias acumuladas.

DISTRIBUCIONES NO AGRUPADAS: El percentil p es aquel valor cuya frecuencia acumulada más se acerca por arriba al p% de n,es decir:

$$P_p=X_i \text{ tal que } N_{i-1} < pn/100 \leq N_i$$

DISTRIBUCIONES AGRUPADAS: Usamos la misma idea que cuando calculábamos la mediana, buscamos en primer lugar el intervalo $[L_{i-1},L_i)$ cuya frecuencia acumulada sea $N_{i-1} < pn/100 \leq N_i$, a continuación para hallar el percentil aplicamos la siguiente fórmula:

$$P_p=L_{i-1}+\frac{\left(\frac{pn}{100}-N_{i-1}\right)a_i}{n_i}$$

b) CUARTILES (C_i): Son los tres valores que dividen al conjunto de datos ordenados en cuatro partes iguales, son un caso particular de los percentiles:

$C_1=P_{25}$
$C_2=P_{50}$
$C_3=P_{75}$.

c) DECILES (D_i): Son los nueve valores que dividen al conjunto de datos ordenados en diez partes iguales, son también un caso particular de los percentiles.

$D_1=P_{10}$
$D_2=P_{20}$
.....
$D_9=P_{90}$

La Mediana también es un caso particular de percentil: $Me = P_{50}$.

2.1.4.2. MEDIDAS DE DISPERSIÓN

Las medidas de tendencia central tenían como objetivo el sintetizar los datos en un valor representativo, las medidas de dispersión nos dirán hasta que punto estas medidas de tendencia central son representativas como síntesis de la información. Las medidas de dispersión cuantifican la separación, la dispersión, la variabilidad de los valores de la distribución respecto al valor central.

Distinguiremos entre medidas de dispersión absolutas, que no son comparables entre diferentes muestras y las relativas que nos permitirán comparar varias muestras.

A) MEDIDAS DE DISPERSIÓN ABSOLUTAS

Por orden de importancia tenemos:

a) VARIANZA (s^2): promedio del cuadrado de las distancias entre cada observación y la media aritmética del conjunto de observaciones

$$s^2=\frac{\sum_i (x_i-\bar{x})^2 n_i}{n}$$

Si los datos están agrupados utilizamos las marcas de clase, es decir c_i en vez de x_i.

En el caso extremo en que todas las observaciones fueran iguales, la media coincidiría con ese valor común y la varianza sería cero. En general, cuanto más dispersas sean las observaciones, mayores serán las diferencias dentro de los cuadrados y por tanto mayor será el valor de s^2.

PROPIEDADES:

1. La varianza nunca puede ser negativa, $s^2 > 0$.
2. Otra forma más sencilla de calcular la varianza es:

$$s^2 = \frac{\sum_i x_i^2 n_i}{n} - \bar{x}^2$$

Demostración:

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n} = \frac{\sum_i (x_i^2 - 2\bar{x}x_i + \bar{x}^2) n_i}{n} =$$

$$\frac{\sum_i x_i^2 n_i}{n} - 2\bar{x}\frac{\sum_i x_i n_i}{n} + \frac{\bar{x}^2 \sum_i n_i}{n} = \frac{\sum_i x_i^2 n_i}{n} - 2\bar{x}\bar{x} + \frac{\bar{x}^2 n}{n} = \frac{\sum_i x_i^2 n_i}{n} - \bar{x}^2$$

Otras medidas de dispersión directamente relacionadas con la varianza son las dos siguientes.

b) DESVIACIÓN TÍPICA (s): la varianza vendría dada por las mismas unidades que la variable pero al cuadrado, para evitar este problema podemos usar como medida de dispersión la desviación típica que se define como la raíz cuadrada positiva de la varianza $s = \sqrt{s^2}$

PROPIEDAD: Se observa a partir de la definición que $s \geq 0$

c) CUASI-VARIANZA (s^{*2}): se define de forma muy parecida a la varianza pero dividiendo por n-1:

$$s^{*2} = \frac{\sum_i (x_i - \bar{x})^2 n_i}{n-1} = \frac{n}{n-1} s^2$$

d) DESVIACIÓN MEDIA RESPECTO A LA MEDIA ($D_{\bar{x}}$): se define como el promedio de las desviaciones en valor absoluto respecto a la media aritmética:

$$D_{\bar{x}} = \frac{\sum_i |x_i - \bar{x}| \, n_i}{n}$$

Si toma valores grandes significa que los valores de la variable se distribuirán en valores alejados de la media.

e) DESVIACIÓN MEDIA RESPECTO A LA MEDIANA (D_{Me}): Se define como el promedio de las desviaciones en valor absoluto respecto a la mediana:

$$D_{Me}=\frac{\sum_i |x_i - Me| n_i}{n}$$

Si D_{Me} es grande los valores están dispersos respecto de la mediana.

f) RECORRIDO O RANGO MUESTRAL (Re). Es la diferencia entre el valor de las observaciones mayor y el menor, $R_e = x_{max} - x_{min}$

g) RECORRIDO INTERCUARTíLICO (RQ). Es la diferencia entre el primer y el tercer cuartil, $RQ = C_3 - C_1$

B) MEDIDAS DE DISPERSIÓN RELATIVAS

a) COEFICIENTE DE VARIACIÓN DE PEARSON: Cuando se quiere comparar el grado de dispersión de dos distribuciones que no vienen dadas en las mismas unidades o que las medias no son iguales se utiliza el coeficiente de variación de Pearson que se define como el cociente entre la desviación típica y el valor absoluto de la media aritmética

$$CV=\frac{s}{|\bar{x}|}$$

Al hacer el cociente eliminamos las unidades.

CV representa el número de veces que la desviación típica contiene a la media aritmética y por lo tanto cuanto mayor es CV mayor es la dispersión y menor la representatividad de la media.

2.1.4.3. OTRAS MEDIDAS DESCRIPTIVAS: TIPIFICACIÓN

Supongamos que hacemos la siguiente transformación a los datos:

$$z_i=\frac{x_i-\bar{x}}{s}$$

es decir, a cada valor de la variable le restamos la media y lo dividirnos por la desviación típica.

Se trata de una transformación lineal $z_i = a + bx_i$ con $b = \frac{1}{s}$ y $a = \frac{-\bar{x}}{s}$

Usando las propiedades de la media y de la desviación típica, es fácil demostrar que la nueva distribución de frecuencias tiene media aritmética cero y desviación típica 1. Diremos entonces que la muestra o la distribución de frecuencias está tipificada y a la transformación anterior se le llama tipificación.

2.1.4.4. MEDIDAS DE FORMA

Comparan la forma que tiene la representación gráfica, bien sea el histograma o el diagrama de barras de la distribución, con la distribución normal.

A) ASIMETRÍA

Nos miden la simetría de la distribución. Supongamos que hemos representado gráficamente una distribución de frecuencias: tracemos una perpendicular al eje de las x por $\bar{x}$. Diremos que la distribución es simétrica si existe a ambos lados el mismo número de valores, equidistantes dos a dos y cada par de puntos equidistantes con la misma frecuencia.

COEFICIENTE DE ASIMETRÍA DE FISHER:

$$g_1 = \frac{\sum_i (x_i - \bar{x})^3 n_i}{n s^3} = \frac{m_3}{s^3}$$

Si la distribución es simétrica en el denominador tendremos el mismo número de desviaciones positivas como negativas y por tanto $g_1 = 0$.

Si $g_1 > 0$ la distribución es asimétrica positiva o asimétrica a derechas.

Si $g_1 < 0$ la distribución es asimétrica negativa o asimétrica a izquierdas.

COEFICIENTE DE ASIMETRÍA DE PEARSON: Es mucho más fácil de calcular que el anterior pero sólo es aplicable a aquellas distribuciones que tienen una sola moda y cuya distribución tiene forma de campana. Se define:

$$A_s = \frac{\bar{x} - M_o}{s}$$

Si la distribución es simétrica $\bar{x} = Me$ y por tanto $A_s = 0$. Si $A_s > 0$ la distribución es asimétrica positiva. Si $A_s < 0$ la distribución es asimétrica negativa.

B) APUNTAMIENTO O CURTOSIS

Miden la mayor o menor cantidad de datos que se agrupan en torno a la moda. Solo tienen sentido en distribuciones campaniformes, es decir, unimodales simétricas o ligeramente asimétricas.

Si para valores próximos a la moda las frecuencias son más altas que en la distribución normal, la gráfica será muy apuntada en esa zona, y se dice que es de tipo *leptocúrtico*. Cuando son más bajas que en la distribución normal se dice que es de tipo *platicúrtico*. Cuando la distribución de frecuencias es igual de apuntada que la normal se dice que es *mesocúrtica*.

COEFICIENTE DE APUNTAMIENTO DE FISHER. Se define como:

$$g_2 = \frac{\sum_i (x_i - \bar{x})^4 n_i}{n s^4} - 3 = \frac{m_4}{s^4} - 3$$

- si $g_2 > 0$ leptocúrtica
- si $g_2 < 0$ platicúrtica
- si $g_2 = 0$ mesocúrtica o normal.

2.1.4.5. MEDIDAS DE CONCENTRACIÓN

Las medidas de concentración tratan de poner de manifiesto el mayor o menor grado de igualdad en el reparto total de los valores de la variable. Son por tanto, indicadores del grado de equidistribución de la variable. Estas medidas tienen especial aplicación a variables económicas (rentas, salarios, etc.).

Supongamos que tengamos n sujetos cuyos valores de la variable (rentas, salarios, etc.) son:

$$x_1 \leq x_2 \leq x_3 < \ldots < x_n$$

Nos interesa estudiar hasta que punto la suma total de valores (rentas, salarios, etc.) está equitativamente repartida.

Las dos situaciones extremas serían:

1. Concentración máxima: de los n sujetos, sólo uno percibe el total y los demás nada:

$$x_1 = x_2 = x_3 = ... = x_{n-1} = 0 \text{ y } x_n \neq 0$$

2. Concentración mínima o equidistribución: todos tienen el mismo valor

$$x_1 = x_2 = x_3 = ... = x_{n-1} = x_n$$

Hay que tener en cuenta que desde el punto de vista estadístico los términos dispersión y concentración no son opuestos, recordemos que el primero hacía referencia a la variabilidad de los datos con respecto al promedio, mientras que el segundo, como acabamos de definir, a la no equidad en el reparto de la suma total de la variable.

a) ÍNDICE DE CONCENTRACIÓN DE GINI (I_G): El índice de concentración de Gini se construye a partir de las siguientes cantidades:

1. Los productos $x_i n_i$ que nos indicarán el total percibido (renta total, ganancia total, etc.) por los n_i sujetos con valor (renta,...) x_i. A este producto le llamaremos riqueza del grupo.
2. Las riquezas acumuladas de la variable (u_i) , se calculan de la siguiente forma:

$$u_1 = x_1 n_1$$
$$u_2 = x_1 n_1 + x_2 n_2$$
$$u_3 = x_1 n_1 + x_2 n_2 + x_3 n_3$$
$$.....$$
$$u_k = x_1 n_1 + x_2 n_2 + ... + u_k n_k$$

3. Las riquezas acumuladas (u_i) las expresamos en tanto por ciento del total u_k.

$$q_i = \frac{u_i}{u_k} x100$$

4. Las frecuencias relativas acumuladas, expresadas en tanto por ciento:

$$p_i = \frac{N_i}{n} x100 = F_i x100$$

A partir de todo esto se define el índice de concentración de Gini mediante la fórmula:

$$I_G = \frac{\sum_{i=1}^{k-1} (p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

Podemos observar que:

a) Si $q_i = 0$, para i = 1,2,...,k-1, y $q_k \neq 0$ entonces:

$$I_G = \frac{\sum_{i=1}^{k-1} p_i}{\sum_{i=1}^{k-1} p_i} = 1$$

y la concentración es máxima.

b) Si para cada i tenemos que $p_i = q_i$, $I_G = 0$ y el reparto es equitativo, ya que cada porcentaje de individuos posee el mismo porcentaje de riqueza.

c) CURVA DE LORENZ: Una forma de estudiar gráficamente la concentración es mediante la curva de Lorenz que se construye representado en el eje de abcisas el porcentaje de frecuencias acumuladas (p_i) y en el eje de ordenadas los porcentajes acumulados del total de la variable (q_i). Al unir estos puntos obtenemos la curva de Lorenz.

Como para $p_i = 0$, la gráfica pasa por el punto (0,0), y para $p_i = 100\%$ se tiene que $q_i = 100\%$, la gráfica pasa por los puntos O = (0,0) y P(100,100). Por otra parte, al ser $p_i \leq q_i$, por estar ordenados los datos en sucesión creciente, la gráfica está siempre situada por debajo de la diagonal del cuadrado o coincidente con ella. En el caso de existir reparto equitativo, es decir concentración mínima, la curva coincide con la diagonal (OB), pues en ese caso $p_i = q_i$. Si la concentración es máxima la curva de Lorenz estaría formada por los lados OA y OB.

Se demuestra que aproximadamente:

$$I_G = \frac{\text{Área entre la curva y la diagonal OB}}{\text{Área del triángulo OAB}}$$

COMPARACIÓN ENTRE LAS DOS MEDIDAS

Si bien el índice de Gini tiene la ventaja de resumir la información en una sola cifra y por tanto comparar más fácilmente que la curva de Lorenz, esta ventaja tiene como contrapartida el que dos distribuciones con aspecto muy diferente pueden tener el mismo índice de Gini.

2.1.5. TRANSFORMACIONES LINEALES

En este apartado veremos como quedan afectadas algunas de las medidas de una variable cuando le sumamos o multiplicamos alguna cantidad. Es decir, calculamos una transformación lineal de la variable original, sobre la que obtenemos queremos saber cuanto vale su media, mediana, varianza y desviación típica.

2.1.5.1. EN LA MEDIA

1. Si a todos los valores de una variable les sumamos una constante k, la media aritmética queda aumentada en esa constante, (la media aritmética queda afectada por los cambios de origen).

Es decir, si $y_i = k + x_i$ entonces: $\bar{y}=k+\bar{x}$

Demostración:

$$\bar{y}=\frac{\sum_i y_i n_i}{n}=\frac{\sum_i (k+x_i)n_i}{n}=\frac{k\sum_i n_i+\sum_i x_i n_i}{n}=\frac{kn}{n}+\frac{\sum_i x_i n_i}{n}=k+\bar{x}$$

2. Si todos los valores de una variable los multiplicamos por una constante k, su media aritmética queda multiplicada por la misma constante (la media aritmética queda afectada por los cambios de escala).

Es decir, si $y_i = k\, x_i$ entonces $\bar{y}=k\bar{x}$

3. Como consecuencia de las anteriores, si consideramos la transformación lineal $y_i = a + bx_i$ siendo a y b dos constantes cualesquiera, la media aritmética de la nueva variable quedaría:

$$\bar{y}=a+b\bar{x}$$

2.1.5.2. EN LA MEDIANA

1. Si a todos los valores de una variable les sumamos una constante k, la mediana queda aumentada en esa constante. Es decir, la mediana queda afectada por los cambios de origen.

Es decir, si $y_i = k + x_i$, entonces: $Me_y = k + Me_x$

2. Si todos los valores de una variable los multiplicamos por una constante k, su mediana queda multiplicada por la misma constante. Es decir, la mediana queda afectada por los cambios de escala.

Es decir, si $y_i = k\, x_i$, entonces: $Me_y = kMe_x$

3. Como consecuencia de las anteriores, si consideramos la transformación lineal $y_i = a + bx_i$ siendo a y b dos constantes cualesquiera, la nueva mediana quedaría:

$$Me_y = a + bMe_x$$

2.1.5.3. EN LA VARIANZA

1. Si a todos los valores de una variable les sumamos una constante k, la varianza no varía. Es decir:

Si $y_i = k + x_i$, entonces: $s_y^2 = s_x^2$

2. Si todos los valores de una variable los multiplicamos por una constante k, su varianza queda multiplicada por el cuadrado de la constante.

Si $y_i = kx_i$, entonces: $s_y^2 = k_2\, s_x^2$

3. Como consecuencia de las anteriores, si consideramos la transformación lineal $y_i = a + bx_i$ siendo a y b dos constantes cualesquiera, la nueva varianza quedaría

$$s_y^2 = b_2\, s_x^2$$

2.1.5.4. EN LA DESVIACIÓN TÍPICA

1. Si $y_i = k + x_i$ entonces: $s_y = s_x$
2. Si $y_i = k\, x_i$ entonces: $s_y = |k|s_x$.
3. Si $y_i = a + bx_i$ entonces: $s_y = |b|s_x$.

2.2. DESCRIPCIÓN CONJUNTA DE VARIAS VARIABLES

En muchos procesos de la vida ordinaria se hace necesario estudiar simultáneamente dos características, dos variables. Su estudio conjunto permite determinar las relaciones entre ellas. Supondremos inicialmente que estamos observando dos variables aunque el tratamiento que se presenta se generaliza sin dificultad a cualquier número de variables.

2.2.1. DISTRIBUCIONES DE FRECUENCIAS BIVARIANTES

a) DISTRIBUCIÓN CONJUNTA

Cuando queremos describir conjuntamente dos variables, el primer paso al igual que en el caso univariante será la representación de los datos en una tabla de frecuencias.

FRECUENCIA ABSOLUTA CONJUNTA (n_{ij}): número de veces que se presenta en la muestra el valor x_i, de la variable X con el valor de y_j de la variable Y.

PROPIEDAD: La suma de las frecuencias absolutas es igual a n.

FRECUENCIA RELATIVA CONJUNTA (f_{ij}) $f_{ij} = n_{ij} / n$.

PROPIEDAD: La suma de las frecuencias relativas es igual a la unidad.

TABLA DE FRECUENCIAS CONJUNTAS: Llamamos así a una tabla de doble entrada donde se representa en las columnas los diferentes valores observados para la variable X (x_i) ordenados de menor a mayor y en las filas los diferentes valores observados para la variable Y (y_j), y en el interior sus correspondientes frecuencias conjuntas, (n_{ij}, o f_{ij} o ambas).

En el caso en que las variables sean cualitativas la tabla de distribución conjunta también recibe el nombre de tabla de contingencia.

b) DISTRIBUCIONES MARGINALES

Para distinguir las frecuencias de cada variable al estudiarlas aisladamente llamaremos frecuencias marginales a las de cada variable por separado. De esta forma tendríamos dos distribuciones unidimensionales a partir de las conjuntas.

FRECUENCIA ABSOLUTA MARGINAL

Para la X sería el número de veces que se repite el valor x_i sin tener en cuenta los valores de Y, la representamos por n_i.

Para la Y sería el número de veces que se repite el valor y_j sin tener en cuenta los valores de X, la representamos por n_j.

FRECUENCIAS RELATIVAS MARGINALES: A partir de las anteriores, y del mismo modo se construirían estas frecuencias. f_i y f_j.

Sobre la tabla de doble entrada que contiene las frecuencias conjuntas, situamos:

en la última columna de la tabla conjunta, las frecuencias marginales de X es decir, n_i, añadiendo tantas columnas como otros tipos de frecuencias marginales se deseen añadir.

en la última fila de la tabla conjunta, las frecuencias marginales de Y, es decir, n_j, añadiendo tantas columnas como otros tipos de frecuencias marginales se deseen añadir.

c) DISTRIBUCIONES CONDICIONADAS

A partir de la distribución de frecuencias conjuntas podemos definir otro tipo de distribuciones unidimensionales, tanto para X como para Y. Estas distribuciones se obtendrán al fijar el valor de la otra variable y reciben el nombre de distribuciones condicionadas.

FRECUENCIA ABSOLUTA CONDICIONADA para X = x_i dado que Y = y_j es el número de veces que se repite el valor x_i teniendo en cuenta solo aquellos valores en que Y = y_j ;así es: $n_{i(j)} = n_{ij}$ $\forall i = 1,...,k$.

FRECUENCIA ABSOLUTA CONDICIONADA para Y = y_j dado que X = x_i es el número de veces que se repite el valor y_j teniendo en cuenta sólo aquellos valores en que X = x_i; así es: $n_{(i)j} = n_{ij}$ $\forall j = 1,...,h$

En las distribuciones condicionadas no se suelen utilizar las frecuencias absolutas, puesto que, como sabemos, éstas dependen del número de datos y el número de datos será diferente para cada distribución, pues dependerá de la frecuencia del valor que fijamos de la otra variable. Son mucho más útiles las frecuencias selectivas que se definen:

FRECUENCIA RELATIVA CONDICIONADA para X dado que $Y = y_j$ es $f_{i(j)} = n_{ij}/n_{\cdot j}$

FRECUENCIA RELATIVA CONDICIONADA para Y dado que $X = x_i$ es $f_{(i)j} = n_{ij}/n_{i\cdot}$

Si la tabla resulta muy grande deberemos agrupar una o las dos variables en intervalos de clase.

INDEPENDENCIA ESTADÍSTICA

DEFINICIÓN 1: Dos variables X e Y se dicen que son independientes estadísticamente cuando la frecuencia relativa conjunta es igual al producto de las frecuencias relativas marginales, es decir: $f_{ij} = f_{i\cdot} f_{\cdot j}$ ó $\frac{n_{ij}}{n} = \frac{n_{i\cdot}}{n} \cdot \frac{n_{\cdot j}}{n}$

DEFINICIÓN 2: Dos variables X e Y se dicen que son independientes estadísticamente cuando todas las frecuencias relativas condicionadas son iguales a sus correspondientes frecuencias marginales, es decir: $f_{i(j)} = f_{i\cdot}$ para todo j, y $f_{(i)j} = f_{\cdot j}$ para todo i.

2.2.2. REPRESENTACIÓN GRÁFICA

Al igual que en el caso univariante, la forma de la distribución conjunta se aprecia a primera vista y se retiene más fácilmente en la memoria con una adecuada representación gráfica.

DIAGRAMA DE DISPERSIÓN O NUBE DE PUNTOS: Se obtiene representando cada par observado (x_i, y_j) como un punto en el plano cartesiano. Se utiliza con los datos sin agrupar y sobre todo para variables continuas. Si los datos están agrupados tomaríamos las marcas de clase. Es el más útil porque nos permite determinar visualmente la relación entre las dos variables.

2.2.3. MEDIDAS DE DEPENDENCIA LINEAL

COVARIANZA-CORRELACIÓN

Al analizar dos variables cuantitativas de forma conjunta, el objetivo que se pretende es, por lo general, determinar si existe o no algún tipo de variación conjunta o covarianza entre ellas: si una variable aumenta, la otra también, o lo contrario.

La cantidad se denomina COVARIANZA (S_{xy}) y tiene la siguiente expresión:

$$S_{xy} = \frac{\sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y}) n_{ij}}{n}$$

y ayuda a cuantificar la covariación entre dos variables del siguiente modo:

Cuando $S_{xy} > 0$, hay una tendencia directa, en el sentido que a mayores observaciones de X correspondan mayores observaciones de Y.

Cuando $S_{xy} < 0$, la tendencia resulta inversa; es decir, a mayor valor de X solemos encontrar menores valores de Y.

Este valor depende de la magnitud de las variables, por tanto de sus unidades. Para poder eliminar las unidades y tener una medida adimensional utilizamos el COEFICIENTE DE CORRELACIÓN (r_{xy})

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

siendo también invariante frente a transformaciones lineales (cambio de origen y escala) de las variable. Citamos las siguientes propiedades:

- Es un coeficiente adimensional. Los valores de r_{xy}:

$$-1 \leq r_{xy} \leq 1$$

- Si hay relación lineal positiva $r_{xy} > 0$ y próximo a 1.
- Si hay relación lineal negativa $r_{xy} < 0$ y próximo a -1.
- Si no hay relación lineal r_{xy} se aproxima a 0.
- Si X e Y son independientes $S_{xy} = 0$ y por tanto $r_{xy} = 0$.

2.2.4. RECTA DE REGRESIÓN

A) INTRODUCCIÓN AL MÉTODO DE MÍNIMOS CUADRADOS

El diagrama de dispersión o nube de puntos nos permitía visualizar la relación entre dos variables X e Y. Al representar el diagrama de dispersión podemos encontrar las siguientes situaciones.

a) Distribuciones estadísticas para las que la nube de puntos se dispone de tal forma que existe una función matemática cuyos puntos son una parte de su representación gráfica.

b) Sin coincidir exactamente sus puntos con las de una gráfica de una función matemática, se aproximan a ella con mayor o menor intensidad.

c) La nube de puntos presenta un aspecto tal que no existe concentración de puntos hacia ninguna función matemática, distribuyéndose de una forma uniforme en una región del plano.

En el primer caso se dice que existe una dependencia funcional o exacta entre las variables X e Y, es decir existe una función matemática tal que $Y = f(X)$. En el segundo caso se dice que existe una dependencia estadística o aproximada entre las dos variables, $Y \approx f(X)$. Y en el último caso diríamos que las variables son independientes.

Es el segundo caso del que se ocupa la teoría de la regresión.

Las técnicas de regresión tienen por objeto modelizar, es decir, encontrar una función que aproxime lo máximo posible la relación de dependencia estadística entre variables y predecir los valores de una de ellas: Y (Variable dependiente o explicada) a partir de los de la otra (o las otras): X (variables(s) independiente(s) ó explicativa(s)).

Llamaremos regresión de Y sobre X a la función que explica la variable Y (dependiente) para cada valor de la X (independiente).

$$Y \approx f(X)$$

Llamaremos regresión de X sobre Y a la función que explica la variable X (dependiente) para cada valor de la Y (independiente).

$$X \approx f(Y)$$

La regresión es lineal cuando el modelo función de regresión seleccionado es una recta. En cualquier otro caso se dice regresión no lineal. La regresión será simple cuando sólo tengamos una variable independiente. Cuando tengamos dos o más variables independientes, la regresión será múltiple.

El procedimiento a seguir es el siguiente:

a) Elegir un tipo de función o curva que creamos que mejor relaciona las dos variables; esto lo podemos hacer observando la nube de puntos.

b) Obtener la ecuación de la curva, de entre las infinitas de dicho tipo que hay en el plano, que mejor se adapte al conjunto de puntos. El objetivo de obtener esa ecuación será predecir el valor de la variable Y dado un valor x_0 de la variable X.

c) Obtener una medida del grado de esta asociación o correlación. Esto me dará la fiabilidad de las predicciones que se determinen con esta ecuación.

Notar que los dos primeros apartados se engloban dentro de lo que se conoce como la teoría de la regresión mientras que el tercero es lo que se conoce como teoría de la correlación.

El segundo apartado se conoce como el problema del ajuste y se pueden utilizar diferentes métodos matemáticos como por ejemplo:

- Método de los mínimos cuadrados.
- Método de los polinomios ortogonales.
- Método de los momentos.
- Método de la curva logística.

Sólo vamos a desarrollar el método de los mínimos cuadrados. Trabajaremos con la muestra original, sin ordenar ni agrupar en una tabla de frecuencias.

2.2.4.1. MÉTODO DE MÍNIMOS CUADRADOS

Dados los puntos (x_1,y_1), (x_2,y_2),...,(x_n,y_n), supongamos que hemos elegido una función $y = f(x|a_1,...,a_r)$ que queremos ajustar a ese conjunto de puntos y en la que intervienen r parámetros $(a_1,...,a_r)$. Consideramos la nube de puntos correspondiente:

Para cada valor de X, (x_i) tenemos dos valores de Y:

- El valor observado en la muestra (o en la nube de puntos) y_i.
- Otro que denominamos teórico, y_i^*, que se obtendría al sustituir $x = x_i$ en la función.

Como se puede observar, para cada x_i tenemos una diferencia entre los dos valores de Y, que llamaremos residuo: $e_i = y_i - y_i^*$.

El método de los mínimos cuadrados consiste en determinar los parámetros $(a_1,...,a_r)$ de tal forma que los residuos sean mínimos. Es decir, buscaremos minimizar la expresión:

$$\Psi = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - y_i^*\right)^2 = \sum_{i=1}^{n} \left(y_i - f\left(x_i | a_1,...,a_r\right)\right)^2$$

es decir, minimizamos la suma de las distancias verticales de los puntos a la curva.

La condición necesaria para obtener el mínimo es que las primeras derivadas parciales respecto a cada uno de los parámetros se anulen, es decir,

$$\left.\begin{aligned} &\frac{\partial \Psi}{\partial a_1}=0 \\ &\frac{\partial \Psi}{\partial a_2}=0 \\ &\cdot \\ &\cdot \\ &\cdot \\ &\frac{\partial \Psi}{\partial a_r}=0 \end{aligned}\right\}$$

resolviendo este sistema, denominado sistema de ecuaciones normales, quedan determinados $(a_1,...,a_r)$, así como la correspondiente función.

2.2.4.2. MODELO DE REGRESIÓN LINEAL SIMPLE

RECTA DE REGRESIÓN DE Y SOBRE X: En el modelo de regresión lineal simple la función elegida para aproximar la relación entre las variables es una recta, es decir $y = a + bx$, donde a, b son los parámetros.

Vamos a deducir su ecuación usando el método de los mínimos cuadrados. Dado un valor de X, x_i, tenemos los dos valores de Y, el observado, y_i, y el teórico, $y_i^* = a + bx_i$. Así pues, hemos de minimizar:

$$\Psi=\sum_{i=1}^{n}(y_i-(a+bx_i))^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2$$

que derivando respecto a ***a*** y a ***b*** e igualando a cero:

$$\left.\begin{aligned} &\frac{\partial \Psi}{\partial a}=-2\sum_i(y_i-a-bx_i)=0 \\ &\frac{\partial \Psi}{\partial b}=-2\sum_i(y_i-a-bx_i)x_i=0 \end{aligned}\right\}$$

nos dará un sistema de dos ecuaciones y dos incógnitas (a, b). A este sistema se le llama de *ecuaciones normales.*

Resolviendo el sistema:

$$\left.\begin{aligned}\sum_i y_i &= \sum_i a + b\sum_i x_i \\ \sum_i x_i y_i &= a\sum_i x_i + b\sum_i x_i^2\end{aligned}\right\}$$

$$-na = -\sum_i y_i + b\sum_i x_i \longrightarrow \underline{\underline{a = \bar{y} - b\bar{x}}}$$

$$\sum_i x_i y_i = (\bar{y} - b\bar{x})\sum_i x_i + b\sum_i x_i^2$$

$$\sum_i x_i y_i = \frac{\sum_i y_i}{n}\sum_i x_i - b\bar{x}n\bar{x} + b\sum_i x_i^2$$

$$\sum_i x_i y_i - \bar{y}n\bar{x} = b\left(\sum_i x_i^2 - n\bar{x}^2\right)$$

$$S_{xy} = bS_x^2 \longrightarrow \underline{\underline{b = \frac{S_{xy}}{S_x^2}}}$$

y obtenemos que la recta de regresión de Y sobre X es y = a + bx con los valores a y b anteriormente calculados, o bien la siguiente expresión:

$$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

A la pendiente b de la recta de regresión de Y sobre X se le denomina coeficiente de regresión de Y sobre X.

RECTA DE REGRESIÓN DE X SOBRE Y: Aplicando el mismo razonamiento llegaríamos a la expresión de la recta de regresión de X sobre Y, x = a' + b'y con

$$b' = \frac{S_{xy}}{S_y^2} \; y \; a' = \bar{x} - b'\bar{y}$$

o bien

$$x - \bar{x} = \frac{S_{xy}}{S_y^2}(y - \bar{y})$$

Igualmente a la pendiente b' de la recta de regresión de X sobre Y se le denomina coeficiente de regresión de X sobre Y.

Hay que tener en cuenta que la recta de regresión de X sobre Y no se obtiene despejando X de la recta de regresión de Y sobre X.

PROPIEDADES:

Tanto el signo de b como el de b' será el signo de la covarianza (pues las varianzas son siempre positivas). Una covarianza positiva nos dará dos coeficientes de regresión positivos y sus correspondientes rectas de regresión crecientes. Si la covarianza es negativa, las dos rectas de regresión serán decrecientes al ser negativas sus pendientes. En caso de que la covarianza valga cero, las rectas de regresión serán paralelas a los ejes coordenados y perpendiculares entre sí.

2.2.4.3. MEDIDAS DE BONDAD DE AJUSTE: CORRELACIÓN

a) VARIANZA RESIDUAL: Para cada valor x_i de X, obteníamos una diferencia (el residuo) entre el valor observado de Y en la nube de puntos y el correspondiente valor teórico obtenido en la función. Si todos los puntos de la nube están en la función, la dependencia será funcional; el grado de dependencia será el máximo posible. Cuanto más se alejen los puntos observados de la función (mayores sean los residuos) iremos perdiendo intensidad en la dependencia.

Se define la VARIANZA RESIDUAL como la media de todos los residuos elevados al cuadrado:

$$S_e^2 = \frac{\sum_i (y_i - y_i^*)^2}{n} = \frac{\sum_i (y_i - a - bx_i)^2}{n}$$

Si la varianza residual es grande los residuos serán grandes y la dependencia será pequeña, *el ajuste será malo.*

Si la varianza residual es pequeña (cerca de cero), la dependencia será grande, *el ajuste será bueno.*

Es fácil demostrar que la media de los residuos en la regresión lineal de Y sobre X es cero, es decir, $\bar{e} = 0$. Por tanto la varianza residual recibe este nombre por ser la varianza de los residuos.

b) VARIANZA DEBIDA A LA REGRESIÓN: Nos sirve para ver en qué medida mejora la descripción de una variable a través de la otra. Llamaremos *VARIANZA DEBIDA A LA REGRESIÓN* a la varianza de los valores teóricos, es decir de los y_i^*. Se demuestra que $\bar{y}^* = \bar{y}$, y así pues la varianza debida a la regresión será:

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

se demuestra que $S_y^2 = S_e^2 + S_{y^*}^2$

Es decir, la varianza total de la variable Y es la suma de dos varianzas: la varianza de Y*, que representaría la parte de la dispersión o variabilidad de la variable Y explicada por la regresión, o sea, por la relación lineal con la variable X y la varianza residual que representaría la parte de la variabilidad no explicada por la regresión.

Así pues, cuando aumenta la varianza debida a la regresión, disminuye la varianza residual y el ajuste es bueno y al contrario.

c) COEFICIENTE DE DETERMINACIÓN: El problema de la varianza residual es que viene afectada por las unidades de medida y esto imposibilita la comparación de la dependencia entre grupos de variable. Teniendo en cuenta la relación entre los diferentes tipos de varianzas, podemos obtener una medida relativa (es decir, que no dependa de las unidades y esté entre cero y uno) de la bondad de ajuste dividiendo la varianza debida a la regresión entre la varianza total de Y.

Se define el *COEFICIENTE DE DETERMINACIÓN* como:

$$R^2 = \frac{S_{y^*}^2}{S_y^2} \quad o\ bien \quad R^2 = 1 - \frac{S_e^2}{S_y^2}$$

representa el porcentaje de la variabilidad de Y explicada por la recta de regresión, es decir por su relación con la variable X. Los valores de este coeficiente:

$$0 \leq R^2 \leq 1$$

Si $R^2 = 1$ todos los residuos valen cero y el ajuste es perfecto; si $R^2 = 0$ el ajuste es inadecuado.

PROPIEDADES:

El coeficiente de determinación de la recta de regresión de Y sobre X es el mismo que el de la recta de regresión de X sobre Y, cumpliéndose que: $R^2 = bb'$. Es decir, el coeficiente de determinación es una medida del grado de relación lineal entre las variables. Se observa rápidamente, ya que si por definición $y_i^* = a + bx_i$, aplicando las propiedades de la varianza $S_{y^*}^2 = b^2 S_x^2$, y así pues:

$$R^2=\frac{b^2S_x^2}{S_y^2}=b\frac{\frac{S_{xy}}{S_x^2}S_x^2}{S_y^2}=b\frac{S_{xy}}{S_y^2}=bb'$$

El coeficiente de determinación es el cuadrado del coeficiente de correlación lineal, es decir: $R^2 = r_{xy}^2$, lo que es obvio ya que:

$$R^2=bb'=\frac{S_{xy}}{S_x^2}\frac{S_{xy}}{S_y^2}=\left(\frac{S_{xy}}{S_xS_y}\right)^2=r_{xy}^2$$

2.2.4.4. PREDICCIÓN

El objetivo último de la regresión es la predicción de una variable para un valor determinado de la otra. La predicción de Y para $X = x_0$ será simplemente el valor obtenido en la recta de regresión de Y sobre X al sustituir el valor de x por x_0.

Es claro que la fiabilidad de esta predicción será tanto mayor cuando mayor sea la correlación entre las variables (es decir mayor sea R^2 o r_{xy}).

2.3. ESPACIO MUESTRAL Y SUCESOS

2.3.1. ESPACIO MUESTRAL. ÁLGEBRA DE SUCESOS

FENÓMENOS ALEATORIOS

Sus características más notables son:

a) Es posible repetir el experimento indefinidamente, sin cambiar las condiciones iniciales.

b) Una pequeña modificación en las condiciones iniciales altera el resultado final.

No se puede predecir un resultado particular, pero se puede dar el conjunto de todos los resultados posibles.

Si el experimento se repite un gran número de veces aparece un modelo de regularidad.

Definición

Se dice que un experimento es aleatorio, estocástico o estadístico, si, pudiéndose repetir indefinidamente en análogas condiciones, es imposible predecir el resultado, aún conociendo las condiciones iniciales.

En un experimento aleatorio no conocemos el resultado hasta que se ha realizado la prueba.

Ejemplos

- Sacar una carta de la baraja
- Lanzar un dado
- Lanzar una moneda
- Sacar una bola de un bombo de la lotería

NO SON EXPERIMENTOS ALEATORIOS

- El resultado de una reacción química
- La velocidad de llegada de un cuerpo a tierra al dejarlo caer desde una torre

Nota

Llamaremos *prueba* a cada realización de un experimento.

Definición

El conjunto de todos los resultados posibles a que puede dar lugar un experimento aleatorio se llama *espacio muestral.* Suele representarse por E ó Ω ; y diremos que es *finito* si el número de resultados posibles es finito.

Definición

Dado un experimento aleatorio cuyo espacio muestral es E, se llama *suceso* a cada uno de los subconjuntos de E.

Distinguimos los siguientes tipos de sucesos:

- Suceso simple o elemental: sólo consta de un elemento
- Suceso compuesto: consta de dos o más elementos
- Suceso imposible: es el que nunca puede realizarse (viene determinado por el conjunto vacío, $\varnothing$)
- Suceso seguro: es el que siempre se cumple (viene determinado por el conjunto total, E)
- Sucesos disjuntos o mutuamente excluyentes: aquellos sucesos A y B que no pueden realizarse a la vez, $A \cap B = \varnothing$

Definición

Llamaremos *Álgebra de sucesos* al conjunto de las partes del espacio muestral, $\wp(\Omega)$, o sea, al conjunto de todos los sucesos.

Nota

card $\wp(\Omega) = 2^{\text{card}(\Omega)}$

Así, el número de subconjuntos que tenemos al lanzar un dado es 2^6; y el número de subconjuntos que tenemos al tirar una moneda es 2^2, veamos este último:

$\{ \varnothing \} \longrightarrow$ 1 subconjunto
$\{ C \}, \{ X \} \longrightarrow$ 2 subconjuntos
$\{ C, X \} = \Omega \longrightarrow$ 1 subconjunto

Total, 4 subconjuntos $4 = 2^2$

Nota

Teniendo en cuenta que los sucesos son subconjuntos de E (de Ω), podemos aplicarles la teoría general de conjuntos. Nos interesarán las uniones, intersecciones, diferencias y complementarios entre conjuntos.

Propiedades de la teoría de conjuntos

- Conmutativa:
$A \cup B = B \cup A \quad A \cap B = B \cap A$

- Asociativa:
$A \cup (B \cup C) = (A \cup B) \cup C \quad A \cap (B \cap C) = (A \cap B) \cap C$

- Leyes de Morgan:
$\overline{A} \cup \overline{B} = \overline{A \cap B} \quad \overline{A} \cap \overline{B} = \overline{A \cup B}$

- Distributivas:
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Además:
$A \cap \overline{A} = \varnothing \quad A \cup \overline{A} = \Omega \quad A - B = A \cap \overline{B}$

2.3.2. FRECUENCIAS

Se llama *frecuencia absoluta* de un suceso aleatorio al número de veces que ocurre dicho suceso al realizar un experimento. Se denota por F .

Se llama *frecuencia relativa* de un suceso aleatorio al cociente entre la frecuencia absoluta y el número de veces que se ha realizado el experimento. Se denota por f .

Acotaciones de las frecuencias:

Consideremos un resultado elemental del experimento aleatorio, y observemos en n realizaciones la frecuencia con la que se ha presentado este suceso, que llamaremos r.

Evidentemente: $0 \leq F_n (x = r) \leq n$

Si dividimos entre n: $0 \leq [F_n (x = r) / n] \leq 1$

Por lo tanto, $0 \leq f_n (x = r) \leq 1$

2.3.3. PROBABILIDAD

2.3.3.1. CONCEPTO DE PROBABILIDAD

El concepto de probabilidad se aplica a los elementos de una población homogénea.

Supongamos una población finita con N elementos, k de los cuales tienen la característica A. Llamaremos «probabilidad de la característica A en la población» a la frecuencia relativa k / N. Se escribe:

$$P (A) = k / N$$

AXIOMÁTICA DE LA PROBABILIDAD

AXIOMA 1: La probabilidad del suceso seguro vale 1, $P (\Omega) = 1$.

AXIOMA 2: La probabilidad de cualquier otro suceso S es no negativa, $P (S) \geq 0$.

AXIOMA 3: La probabilidad de la unión de dos sucesos, A y B, *mutuamente excluyentes,* es la suma de sus probabilidades.

Si $A \cap B = \varnothing$, entonces:

$$P(A \cup B) = P(A) + P(B)$$

Generalizando este último axioma:

La probabilidad de la unión de un conjunto infinito numerable de sucesos mutuamente excluyentes es igual a la suma de sus probabilidades.

$$P (\cup A_i) = \Sigma P (A_i) = P (A_1) + P (A_2) + ...$$

PROPIEDADES DE LA PROBABILIDAD

De estos axiomas podemos deducir una serie de propiedades:

Propiedad 1

Si $A_1, A_2, \ldots\ldots, A_n$ son sucesos disjuntos dos a dos con $n > 2$ (o sea, $A_i \cap A_j = \varnothing$ con $i \neq j$) , entonces :

$$P (A_1 \cup A_2 \cup \ldots\ldots \cup A_n) = P (A_1) + P (A_2) + \ldots\ldots + P (A_n)$$

Demostración

Es inmediata por el Axioma 3, ya que el número de sucesos que hemos tomado es n (un nº finito), y ya teníamos que se cumple para dos sucesos y para una cantidad infinita numerable. Por tanto se cumple para una cantidad finita.

Propiedad 2

$P (\overline{A}) = 1 - P (A)$, siendo A un suceso cualquiera.
($\overline{A}$ es el complementario de A).

Demostración

$A \cup \overline{A} = \Omega \longrightarrow P (A \cup \overline{A}) = P (\Omega) = 1$
Como $A \cap \overline{A} = \varnothing$ —AXIOMA 3→ $P (A \cup \overline{A}) = P (A) + P (\overline{A})$
De ambas consecuencias, $P(A) + P(\overline{A}) = 1 \longrightarrow P(\overline{A}) = 1 - P (A)$

Propiedad 3

$P (\varnothing) = 0$

Demostración

$\varnothing = \overline{\Omega} \longrightarrow P (\varnothing) = P (\overline{\Omega})$
Por la Propiedad 2, $P (\overline{\Omega}) = 1 - P (\Omega) = 1 - 1 = 0$
Por lo tanto, $P (\varnothing) = 0$.

Propiedad 4

$P (S) \leq 1$, siendo S un suceso cualquiera.

Demostración

Por reducción al absurdo, supongamos que $P(S) > 1$.
Como por la Propiedad 2 se tiene que $P(S) + P(\overline{S}) = 1$, deberá ser
$P(S) < 0$, pero esto no puede ser, ya que por el *AXIOMA 2*, la probabilidad de cualquier suceso siempre es ≥ 0.

Nota

Del *AXIOMA 2* y de la Propiedad 4 deducimos:
$0 \leq P(S) \leq 1$, siendo S un suceso cualquiera.

Propiedad 5

Dados dos sucesos A y B tales que $A \subset B \longrightarrow P(A) \leq P(B)$.

Demostración

Luego $B = A \cup (B \cap \overline{A})$
Además, $A \cap (B \cap \overline{A}) = \varnothing \longrightarrow$ A y $(B \cap \overline{A})$ son disjuntos
Por lo tanto, por el *AXIOMA 3* : $P(B) = P(A) + P(B \cap \overline{A})$.
Como, por el *AXIOMA 1*, $P(B \cap \overline{A}) \geq 0 \longrightarrow P(B) \geq P(A)$.

Propiedad 6

$$\forall A, B \subset \Omega, P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Demostración

Intuitivamente, al hacer A + B, tomamos dos veces $A \cap B$, luego para calcular lo que queremos hemos de restar una vez $A \cap B$.

Definición

Llamaremos *espacios muestrales finitos* a los espacios muestrales que provengan de experimentos para los cuales sólo existe un número finito de resultados posibles, así $\Omega = \{ w_1, w_2, \ldots, w_n \}$.

En un experimento aleatorio con un espacio muestral finito, una distribución de probabilidad se especifica asignando una probabilidad p_i a cada resultado $w_i \in \Omega$, $p_i = P(\{ w_i \})$. Debe cumplirse:

a) $p_i \geq 0$

b) $P(\Omega) = 1 \longrightarrow \sum p_i = 1$

En estas condiciones, si $A = \{ w_{i1}, w_{i2}, \ldots, w_{ir} \}$, se tiene $P(A) = \sum p_{ij}$

Definición

Llamaremos *espacios muestrales simples* a los espacios muestrales finitos en los que todos los resultados son equiprobables (tienen la misma probabilidad).

Si $\Omega = \{ w_1, w_2, \ldots, w_n \}$, entonces $P(\{w_i\}) = 1 / n$, para todo $i = 1, \ldots, n$

En estos espacios muestrales simples, dado un suceso $A = \{w_1, w_2, \ldots, w_k \}$ con $k < n$ se tiene:

$$P(A) = \frac{casos\ favorables}{casos\ posibles} = \frac{k}{n}$$

Esto está estrechamente relacionado con la *Fórmula de Laplace:*

$$P(A) = \frac{número\ de\ elementos\ de\ S}{número\ de\ elementos\ de\ \Omega} = \frac{casos\ favorables}{casos\ posibles}$$

siendo S un suceso cualquiera.

2.3.3.2. CONTEO DE ELEMENTOS

A veces, contar el número de elementos puede ser difícil. Para ello utilizaremos lo que se conoce con el nombre de *combinatoria.*

Llamaremos «n factorial» (o «factorial de n»), designándolo por n! , al producto de los n primeros números naturales. Es decir, $n! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots\ldots \cdot (n-1) \cdot n$

Nota: Se define $0! = 1$

Se deducen las siguientes relaciones:

$n! \cdot (n + 1) = (n + 1)!$

$$\frac{n!}{r!} = (r+1) \cdot (r+2) \cdot (r+3) \cdot \ldots\ldots \cdot (n-1) \cdot n$$

Suponiendo todos los elementos distintos, tenemos:

a) *Variaciones:* Dados n elementos, llamaremos variaciones de orden k a todos los conjuntos distintos que podamos formar con esos n elementos, tomados de k en k y teniendo en cuenta el orden. El número de tales variaciones es:

$$V_{n,k} = \frac{n!}{(n-k)!}$$

b) *Permutaciones*: Dados k elementos, llamaremos permutaciones de orden k a todos los conjuntos distintos que podamos formar con esos k elementos, tomados de k en k. El número de tales permutaciones es:

$$P_k = k!$$

c) *Combinaciones:* Dados n elementos, llamaremos combinaciones de orden k a todos los conjuntos distintos que podamos formar con esos n elementos, tomados de k en k y sin tener en cuenta el orden. Su número es:

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

Se cumple la siguiente propiedad: $V_{n,k} = C_{n,k} \cdot P_k$

d) *Variaciones con repetición:* A partir de n elementos distintos formamos variaciones de orden k tales que 2, 3, ..., los k con la posibilidad de repetir elementos. El número de tales variaciones, que designaremos por $VR_{n,k}$, es:

$$VR_{n,k} = n^k$$

e) *Permutaciones con repetición:* Sean k elementos, de los que k_1 son iguales entre sí, k_2 son iguales entre sí, ..., k_r son iguales entre sí, con $k_1 + k_2 + ... + k_r = k$. El número de tales permutaciones es igual a:

$$PR^{k,k,..k} = \frac{k!}{k_1! \cdot k_2! \ldots \cdot k_r!}$$

f) *Combinaciones con repetición:* A partir de n elementos distintos, formamos combinaciones de orden k con la posibilidad de repetir elementos. El número de tales combinaciones es:

$$CR_{n,k} = \frac{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$

2.3.3.3. PROBABILIDAD CONDICIONADA

Dados dos sucesos A y B $\in \wp(\Omega)$, se llama probabilidad de A condicionada a B y se escribe P (A / B) a la probabilidad que existe de que ocurra el suceso A considerando que antes ha ocurrido el suceso B.

Veamos cómo calcular P (A / B):

Si suponemos que ha ocurrido B, tendremos un nuevo espacio muestral, Ω_B = $= \Omega \cap B = B$, y así:

$$P(A/B) = \frac{n^o \text{ de casos favorables en } A \cap B}{n^o \text{ de casos posibles en } B}$$

$$= \frac{\dfrac{n^o \text{ de casos favorables en } A \cap B}{\text{número de casos en } \Omega}}{\dfrac{n^o \text{ de casos posibles en } \Omega}{\text{número de casos en } \Omega}} = \frac{P(A \cap B)}{P(B)}$$

Por lo tanto:

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

2.3.3.4. PROBABILIDAD COMPUESTA (TEOREMA DEL PRODUCTO)

Sea un espacio muestral Ω, dados dos sucesos A y B $\in \wp(\Omega)$ tal que $P(A) > 0$ y $P(B) > 0$, se cumple:

$P(A \cap B) = P(A/B) \cdot P(B)$
$P(B \cap A) = P(B/A) \cdot P(A)$

Esto es así porque por la definición de la probabilidad condicional,

$$P(A/B) = \frac{P(B \cap A)}{P(B)} \longrightarrow P(B \cap A) = P(A/B) \cdot P(B)$$

Análogamente,

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \longrightarrow P(B \cap A) = P(B/A) \cdot P(A)$$

Si en vez de 2 sucesos tenemos n sucesos:

Sean $A_1, A_2, A_3, A_4, \ldots\ldots, A_n \in \wp(\Omega)$:

$$P[\cap A_i] = P(A_1)\cdot P(A_2/A_1)\cdot P(A_3/A_1\cap A_2)\cdot P(A_4/A_1\cap A_2\cap A_3)\cdot\ldots\cdot P(A_n/\cap A_i)$$

2.3.3.5. PROBABILIDAD TOTAL

Dado un espacio muestral Ω, y siendo $\{A_i\} \in \wp(\Omega) / \cup A_i = \Omega$ y $A_i \cap A_j = \varnothing\ \forall\ i \neq j$, y siendo B un suceso del que se conoce $P(B/A_i)$, $\forall$ i, se tiene que:

$$P(B) = \sum P(B/A_i)\cdot P(A_i)$$

Demostración

$$B = (B\cap A_1)\cup(B\cap A_2)\cup(B\cap A_3)\cup\ldots\cup(B\cap A_i)\cup\ldots\cup(B\cap A_n)$$

Como son todos disjuntos:

$$P(B) = P(B\cap A_1) + P(B\cap A_2) + P(B\cap A_3) + \ldots\ldots + P(B\cap A_n)$$

Y aplicando el Teorema del Producto:

$$P(B) = P(B/A_1)\cdot P(A_1) + P(B/A_2)\cdot P(A_2) + \ldots. + P(B/A_n)\cdot P(A_n) =$$
$$= \sum P(B/A_i)\cdot P(A_i) \longrightarrow P(B) = \sum P(B/A_i)\cdot P(A_i)$$

2.3.3.6 TEOREMA DE BAYES

Sea Ω un espacio muestral.
Sean $\{A_i\} \in \wp(\Omega) / \cup A_i = \Omega$, $A_i \cap A_j = \varnothing\ \forall\ i \neq$, conociéndose $P(A_i)\ \forall\ i$, $P(A_i) > 0$
Sea B un suceso tal que $P(B) > 0$ y del que se conocen $P(B/A_i)\ \forall i$.
Entonces:

$$P(A_1/B) = \frac{P(B/A_1)\cdot P(A_1)}{P(B/A_1)\cdot P(A_1) + P(B/A_2)\cdot P(A_2) + \ldots\ldots + P(B/A_n)\cdot P(A_n)}$$

Es decir:

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{\sum P(B/A_k) \cdot P(A_k)}$$

Demostración

$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$, por la probabilidad condicionada.

Si en el numerador aplicamos el Teorema del Producto, y en el denominador la Probabilidad Total, queda:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B/A_i) \cdot P(A_i)}{\sum P(B/A_k) \cdot P(A_k)}$$

2.3.3.7. INDEPENDENCIA DEL SUCESO

Teorema de Caracterización:

Dos sucesos A y B son independientes sii $P(A \cap B) = P(A) \cdot P(B)$
Veámoslo:
($\Rightarrow$)
$P(A \cap B) = P(A/B) \cdot P(B)$
Si son independientes, se tiene que $P(A/B) = P(B)$
Uniendo ambas cosas, $P(A \cap B) = P(A) \cdot P(B)$

($\Leftarrow$)
Ahora se tiene que $P(A \cap B) = P(A) \cdot P(B)$
Como $P(A \cap B) = P(A/B) \cdot P(B)$, sustituyendo :
$P(A) \cdot P(B) = P(A/B) \cdot P(B)$
Por lo tanto, $P(A) = P(A/B)$, y así los sucesos A y B son independientes.

Consecuencia:

$P(A/B) = P(A) \longrightarrow P(B/A) = P(B)$.

Propiedades de la independencia estocástica:

1. Si A y B son independientes $\Rightarrow$ $\overline{A}$ y B también lo son.
2. Si A y B son independientes $\Rightarrow$ A y $\overline{B}$ también lo son.

3. Si A y B son independientes $\Rightarrow$ $\bar{A}$ y $\bar{B}$ también lo son.

4. Si existe implicación entre A y B $\Rightarrow$ No existe independencia (salvo que $A = \Omega$ ó $B = \Omega$).

5. Si dos sucesos son incompatibles $\Rightarrow$ No existe independencia (salvo que $P(A) = 0$ ó $P(B) = 0$).

Nota:

Diremos que tres sucesos A_1, A_2 y A_3 son independientes si, y sólo si, verifican las relaciones:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1) \cdot P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2) \cdot P(A_3) \text{ y } P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3)$$

Esta definición se puede generalizar a n sucesos.

2.4. VARIABLES ALEATORIAS

2.4.1. VARIABLES ALEATORIAS. FUNCIÓN DE DISTRIBUCIÓN. FUNCIÓN DE DENSIDAD

VARIABLE ALEATORIA

Lo que se pretende con la variable aleatoria es sustituir el espacio de resultados por uno numérico, para facilitar la comprensión. Se llama variable aleatoria a aquella cuyo valor está determinado por el valor del experimento.

TIPOS DE VARIABLES ALEATORIAS

Se llama *v.a. discreta* a aquella que sólo puede tomar valores en nº finito de números reales distintos.

Se llama *v.a. continua* a aquella que toman valores en un intervalo, o una unión de intervalos sobre la recta de los números reales. Por tanto, puede tomar ∞ valores.

Al conjunto de valores que toma la v.a. X se le denomina rango de X , rg (X).

DISTRIBUCIÓN DE PROBABILIDAD

Se dice que tenemos una distribución de probabilidad de la v.a. X, cuando asociamos probabilidades a la v.a. que procede de un espacio de resultados probabilizado.

FUNCIÓN DE DISTRIBUCIÓN

Definimos la *función de distribución,* F, de una v.a. X, como una función definida para cada nº real x , de la forma :

$$F(x) = P(X \leq x) \qquad -\infty < x < \infty$$

Es decir, una función que nos indica la probabilidad de que la v.a. X tome valores menores o iguales al x dado.

PROPIEDADES

1) $0 \leq F(x) \leq 1$, por ser una probabilidad
2) La función de distribución es *no decreciente* a medida que crece x, o sea:

$$\text{Si } x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$$

Veámoslo:

$$\text{Si } x_1 < x_2 \Rightarrow P(X \leq x_1) \leq P(X \leq x_2) \Rightarrow F(x_1) \leq F(x_2)$$

En la función de distribución se cumple:

$$\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to +\infty} F(x) = 1$$

Veámoslo:

$$\text{Si } x \to -\infty \Rightarrow P(X-\infty) = P(\varnothing) = 0$$
$$\text{Si } x \to +\infty \Rightarrow P(X+\infty) = P(\Omega) = 1$$

4) $P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

Demostración

Descomponemos el intervalo $(-\infty, x_2\,]$ en $(\infty, x_1\,] \cup (x_1, x_2]$, disjuntos:

$F(x_2) = P(X \leq x_2) = P(X \in (-\infty, x_2\,]) = P(X \in (-\infty, x_1\,]) +$
$+ P(X \in (x_1, x_2\,]) = F(x_1) + P(x_1 < X \leq x_2\,)$

Despejando:

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1)$$

5) La función de distribución es siempre continua por la derecha, o sea:

$$F(x) = F(x^+) = F(\lim\,(x + \varepsilon))\ \forall\ x$$

Es inmediata, ya que $F(\lim\,(x + \varepsilon)) - F(x) = F(x) - F(x) = 0 \Rightarrow$
$\Rightarrow F(x) = F(\lim(x + \varepsilon)) = F(x^+)$

Nota: Por la izquierda no tiene por qué ser continua.

FUNCIÓN DE DISTRIBUCIÓN DE UNA V.A. DISCRETA

Si la v.a. es discreta, la función de distribución es escalonada, y los saltos del escalonamiento tendrán un tamaño igual a la probabilidad del punto en que estemos, manteniéndose constante hasta el siguiente punto de la variable.

FUNCIÓN DE DISTRIBUCIÓN DE UNA V.A. CONTINUA

Si la v.a. es continua, la función de distribución F(x) es continua por la derecha y por la izquierda.

CÁLCULO DE PROBABILIDADES A PARTIR DE LA FUNCIÓN DE DISTRIBUCIÓN

1) $\forall$ x, se cumple: $P(X > x) = 1 - F(x)$

Es inmediato, ya que: $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$

2) $\forall$ x se cumple: $P(X < x) = F(x^-)$

La diferencia entre $P(X < x)$ y $P(X \leq x)$ es que en esta última hemos añadido la probabilidad de x, luego $P(X < x) = F(x^-)$, que es la función de distribución hasta el valor de X infinitésimamente anterior al x dado.

Si la v.a. es continua, no hay saltos, luego: $P(X < x) = P(X \leq x) = F(x)$

3) Dados los valores de X, x_1, x_2 tales que $x_1 < x_2 \Rightarrow P(x_1 < X < x_2) =$
$= F(x_2^-) - F(x_1)$

Demostración

$P(x_1 < X < x_2) = P(X < x_2) - P(X \leq x_1)$
Como $P(X < x_2) = F(x_2^-)$ y $P(X \leq x_1) = F(x_1)$, se tiene que:
$P(x_1 < X < x_2) = F(x_2^-) - F(x_1)$

4) $\forall$ x se cumple: $P(X = x) = F(x^+) - F(x^-)$

Demostración

$P(X = x) = P(X \leq x) - P(X< x) \Rightarrow P(X = x) = F(x^+) - F(x^-)$
Si la v.a. es continua:
$P(X = x) = F(x) - F(x) = 0$

FUNCIÓN DE PROBABILIDAD (O DE CUANTÍA)

Vamos a definir una función que nos especificará la probabilidad de los diversos valores de la variable, cuando es *discreta.*

Definición

Dada una v.a. X de tipo discreto, se llama función de probabilidad P(X) a una función real, tal que cuando la v.a. X toma un determinado valor x_i, es igual a la probabilidad de que ocurra el suceso que viene asociado a dicho valor x_i de la variable.

Es decir:

$P(x_i) = P($ suceso al que representa x_i, según la v.a. X $)$

PROPIEDADES

1) $0 \leq P(x) \leq 1 \ \forall$ x
2) $\Sigma P(x) = 1 \ \forall$ x

3) $P(x_1 \leq X \leq x_2) = \sum_{x1}^{x2} P(x)$

4) $F(x_1) = \sum_{x<x_1} P(x)$

FUNCIÓN DE DENSIDAD

Cuando el experimento da lugar a una v.a. X continua, la variable da lugar a una función de distribución F(x) de tipo continuo, no existiendo probabilidad para un valor concreto de la variable.

La probabilidad de que la variable esté comprendida en el intervalo [A, B] es F(B) - F(A).

Si dividimos la probabilidad de que un valor de X esté en un intervalo, por la longitud del intervalo, obtenemos la *densidad media de probabilidad* en dicho intervalo, que será:

$$\frac{F(B) - F(A)}{B - A}$$

Definición

Se define la *función de densidad* en un punto x, como el límite de la densidad media de probabilidad de un intervalo, cuando la longitud del intervalo tiende a cero. Se representa por f(x) y será:

$$f(x)=\lim \frac{F(x+\Delta x)-F(x)}{\Delta x}=F'(x)$$

Vemos pues que la función de densidad es la derivada de la función de distribución, f(x) = F'(x)

PROPIEDADES

1) La función de densidad es siempre positiva

Demostración

Como la variable es continua ⇒ la función de distribución es siempre creciente ⇒ su derivada (función de densidad o de cuantía) es positiva.

2) $f(+\infty) = f(-\infty) = 0$

Veámoslo:

$$f(+\infty)=\lim\frac{F(+\infty+\Delta x)-F(+\infty)}{\Delta x}=\lim\frac{(1-1)}{\Delta x}=0$$

$$f(-\infty)=\lim\frac{F(-\infty+\Delta x)-F(-\infty)}{\Delta x}=\lim\frac{(0-0)}{\Delta x}=0$$

3) La función de distribución es una primitiva de la función de densidad, ya que $f(x) = F'(x)$

Por tanto:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

Demostración

$$\int_{-\infty}^{x} f(t)dt = [F(t)]_{-\infty}^{x} = F(x) - F(-\infty) = F(x) - 0 = F(x)$$

4) Dada una v.a. X con función de densidad f(x), se cumple:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Veámoslo:

$$\int_{-\infty}^{+\infty} f(x)dx = F(+) - F(-\infty) = 1 - 0 = 1$$

La superficie entre la curva y el eje OX tiene valor 1.

La probabilidad de que x pertenezca a cierto intervalo está en la superficie generada por la curva y el eje OX, y las verticales de los extremos del intervalo.

ESPERANZA MATEMÁTICA

También se le llama valor esperado de la distribución. Es el valor central sobre el que se concentra la distribución de probabilidad. Es semejante a la media de una distribución de frecuencias.

Dada una v.a. X, definimos su **esperanza** como:

$$E(x) = \begin{cases} \sum x_i P(x_i) & \forall i \quad si\ la\ v.a.\ es\ discreta \\ \int xf(x)dx & si\ la\ v.a.\ es\ continua \end{cases}$$

Sirve para saber las posibles ganancias o pérdidas en los juegos de azar. Se dice que un juego es justo cuando la esperanza es cero.

A partir de la esperanza de la v.a. X, se puede obtener su varianza:

$$Var(X) = E(X^2) - [E(X)]^2$$

2.4.2. MODELOS DE PROBABILIDAD

INTRODUCCIÓN

Si un conjunto dado de variables aleatorias (distribuciones) tienen sus funciones de cuantía o de densidad con la misma estructura funcional matemática, diremos que pertenecen a la misma familia de distribuciones o al mismo *modelo de probabilidad.*

La estructura matemática depende de 1 ó más parámetros, y se les llama *parámetros de la distribución.*

Para estudiar los tipos de modelos de distribución utilizaremos el llamado *proceso experimental,* que es el conjunto de características que rigen la realización de un fenómeno aleatorio. Un proceso quedará definido por una serie de características o hipótesis. A partir de estas características, podremos estudiar y determinar la estructura matemática de una distribución.

2.4.2.1. DISTRIBUCIONES DISCRETAS DE PROBABILIDAD

A) DISTRIBUCIÓN DE BERNOULLI

Experimento:

Realizar una prueba con únicamente dos resultados posibles, A (éxito) y $\bar{A}$ (fracaso).

Llamaremos $p = P(A)$ y $q = 1 - p = P(\bar{A})$

Variable:

$$X = n^{\circ}\ de\ éxitos\ rg\ (X) = \{0,1\}$$

Se denota X → Be(p)

La función de cuantía es de la forma:

$$f(x) = \begin{cases} p^x q^{1-x} & x = 0,1 \\ 0 & en \text{ otro caso} \end{cases} \qquad E(X) = p \qquad Var(X) = p(1-p)$$

Ejemplos:
Tirar una moneda (cara o cruz)
Aprobar o suspender un examen
Recibir o no una llamada

B) DISTRIBUCIÓN BINOMIAL

Experimento:

n repeticiones independientes de un experimento Bernoulli; es decir, experimentos con dos resultados posibles (éxito y fracaso) y probabilidades p = P(éxito) y q = 1 – p = P(fracaso). Dichas probabilidades se mantienen constantes a lo largo de las n repeticiones.

Si hay extracción, debe realizarse con reemplazamiento.

Variable:

X = nº de éxitos en las n pruebas rg(X) = {0, 1, 2, 3,....n}

Se denota X → Bi(n, p)

La función de cuantía es de la forma:

$$f(x) = \begin{cases} C_{n,x} p^x q^{n-x} & x = 0,1,2,.....,n \\ 0 & en \text{ otro caso} \end{cases} \qquad E(X) = np \qquad Var(X) = npq$$

TEOREMA DE ADICIÓN

Se dice que una distribución verifica el Teorema de Adición para alguno de sus parámetros, o que es *reproductiva,* cuando dadas dos o más variables aleatorias independientes que sigan todas ellas una distribución de ese tipo con parámetros iguales o distintos, la v.a. suma de todas ellas sigue también una distribución de ese tipo con parámetros la suma de los parámetros de cada una de las variables originales.

En particular, la distribución binomial verifica el Teorema de Adición para el parámetro n, debiendo ser el parámetro p de las variables originales el mismo.

$X_1 \rightarrow Bi(n_1,p)$, $X_2 \rightarrow Bi(n_2,p)$,...., $X_m \rightarrow Bi(n_m,p)$ independientes entre ellas $\Rightarrow$ $\sum X_i \rightarrow Bi(n_1 + n_2 + ... + n_m, p)$

C) DISTRIBUCIÓN HIPERGEOMÉTRICA

Experimento:

n repeticiones no independientes de un experimento con dos resultados posibles (éxito y fracaso) y probabilidades p = P(éxito) y q = P(fracaso) , que no se mantienen constantes a lo largo del experimento . Equivale a un modelo de urnas sin reemplazamiento (sería como una Bernoulli con probabilidades no constantes)

Variable:

X = n° de éxitos en las n pruebas rg (X) = {0, 1, 2,...}
Se denota $X \rightarrow HG(N, n, p)$, siendo:

N = n° de elementos totales
n = n° de pruebas
p = P(éxito)

La función de cuantía es de la forma:

$$f(x) = \begin{cases} \dfrac{C_{k,x} C_{N-k,n-x}}{C_{N,n}} & x = 0,1,2,....., n \\ 0 & en \text{ otro caso} \end{cases}$$

k : n° aciertos población

$$E(X) = \frac{n \cdot k}{N} \qquad Var(X) = npq \frac{N-n}{N-1}$$

D) DISTRIBUCIÓN DE POISSON

Experimento:

Observar un cierto fenómeno físico de naturaleza aleatoria durante un cierto periodo de tiempo o una región del espacio, con las siguientes características:

- el n° de ocurrencias en dos intervalos disjuntos es independiente
- la probabilidad de que haya una ocurrencia en un intervalo pequeño es proporcional a la longitud del intervalo
- la probabilidad de que haya dos o más ocurrencias en un intervalo pequeño debe ser de menor orden que la probabilidad de que haya sólo una ocurrencia

Variable:

X = nº de ocurrencias en un intervalo de tiempo en un intervalo de tiempo o espacio. rg(X) = {0, 1, 2, 3, ...}

Se denota $X \rightarrow Po(\lambda)$

El parámetro λ que distingue una distribución de otra suele ser una especie de «intensidad», y se corresponde con la media de la distribución.

La función de cuantía es de la forma:

$$f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0,1,2,.....,n \\ 0 & en \text{ otro caso} \end{cases} \qquad E(X) = Var(X) = \lambda$$

TEOREMA DE ADICIÓN

La distribución de Poisson verifica el Teorema de Adición para el parámetro λ.

$X_1 \rightarrow Po(\lambda_1)$, $X_2 \rightarrow Po(\lambda_2)$, ..., $X_m \rightarrow Po(\lambda_m)$,
independientes entre ellas $\Rightarrow \sum X_i \rightarrow Po(\lambda_1 + \lambda_2 + ... + \lambda_m)$

APROXIMACIÓN DE LA POISSON POR LA BINOMIAL

Si X tiene una distribución binomial con n grande y p pequeño ($n \rightarrow \infty$, $p \rightarrow 0$), X se puede aproximar por una distribución de Poisson de parámetro $\lambda = np$. Para que sea buena la aproximación, $np < 5$. Utilizaremos la aproximación cuando el valor n no esté en tablas y el valor de p sea pequeño.

2.4.2.2. DISTRIBUCIONES CONTINUAS DE PROBABILIDAD

A) DISTRIBUCIÓN UNIFORME

Experimento:

Cualquier experimento cuyo resultado sea un nº en el intervalo [a, b] de manera que cualquiera de los infinitos puntos del intervalo tenga la misma probabilidad de ser elegido. Es decir, la probabilidad está uniformemente repartida por todo el intervalo.

Variable:

X = valor obtenido, rg(X) = [a, b]
Se denota X → U(a, b)
La función de densidad es de la forma:

$$f(x) = \frac{1}{b-a} \qquad si \;\; a \leq x \leq b$$

La función de distribución toma la forma:

$$F(x) = \frac{1}{b-a} dx$$

B) DISTRIBUCIÓN EXPONENCIAL

Experimento:

Proceso experimental con las mismas características que estudiábamos al definir el modelo de Poisson, es decir, evaluación de un fenómeno a lo largo de un tiempo determinado.

Variable:

X = «tiempo de espera hasta que se produce el primer fenómeno»,o «tiempo de espera entre dos fenómenos consecutivos»

rg(X) = [0, +∞[Se denota X → Ex(β)

Las funciones de densidad, distribución, y valor esperado y varianza son:

$$f(x) = \beta \cdot e^{-\beta x} \; x \geq 0$$

$$F(x) = P(X \leq x) = \beta \cdot e^{-\beta x} \, dx = 1 - e^{-\beta x}$$

$$E(X) = \frac{1}{\beta} \qquad Var(X) = \frac{1}{\beta^2}$$

B.1) FUNCIÓN DE SUPERVIVENCIA

Si el significado de la v.a. es «tiempo que transcurre hasta que se produce el primer fallo», entonces $F(x) = 1 - e^{-\beta x}$ y calculará la probabilidad de que el fallo ocurra antes o en el instante x.

Si consideramos la función $S(x) = 1 - F(x) = P(X > x) = e^{-\beta x}$ (*función de supervivencia*) ésta calculará la probabilidad de que el fallo transcurra después del tiempo x, por tanto, será la probabilidad de que el elemento considerado «sobreviva» al tiempo x.

PROPIEDAD

La distribución exponencial no tiene memoria

$$P(X>s+t / X>s) = \frac{P(X > s+t)}{P(X > s)} = P(X > t)$$

C) DISTRIBUCIÓN NORMAL. TEOREMA CENTRAL DEL LÍMITE

La distribución normal es la más importante de todas las distribuciones de probabilidad. Hay tres razones principales:

a) por sus propiedades matemáticas. En la parte de Inferencia, podremos conocer y trabajar con la distribución de los estimadores si vienen de una población normal.
b) por su aplicación. Un gran número de fenómenos reales se pueden modelizar con esta distribución (como la distribución de alturas y pesos de una población homogénea de personas)
c) por el Teorema Central del Límite. La distribución normal sirve para aproximar la suma y la media de cualquier otro tipo de distribuciones.

Definición

Una v.a. X tiene una distribución normal con parámetros μ y σ^2, $(-\infty < \mu < +\infty, \sigma^2 > 0)$, si X tiene una distribución continua cuya función de densidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in R$$

Se denota $X \rightarrow N(\mu, \sigma^2)$, y se tiene:

$E(X) = \mu$ $Var(X) = \sigma^2$

Esta función de densidad es simétrica respecto a la media μ y la media coincide con la moda y mediana de la distribución. Tiene forma de campana; la densidad decrece a ambos lados de μ más o menos rápido según el valor de σ^2.

Teorema (de las transformaciones lineales)

Sea $X \rightarrow N(\mu, \sigma^2)$, y sea $Y = a X + b$, siendo a y b constantes. Entonces $Y \rightarrow N(a\mu+b, a^2\sigma^2)$

Corolario

Sea $X \rightarrow N(\mu, \sigma^2)$,entonces la nueva v.a. $Z = \frac{X - \mu}{\sigma} \rightarrow N(0,1)$

DISTRIBUCIÓN NORMAL TIPIFICADA

La distribución normal con media 0 y varianza 1 se denomina distribución normal tipificada, $Z \rightarrow N(0,1)$. La función de distribución de esta v.a. se denota por Φ, con $\Phi(z) = P(Z \leq z)$, y se utiliza para el cálculo de probabilidades con tablas, pues las probabilidades acumuladas para la v.a. Z están tabuladas.

Por ser la distribución simétrica se cumple:

$P(Z \leq z) = P(Z > -z) = 1 - P(Z \leq -z)$
Así, $\Phi(-z) = 1 - \Phi(z)$

Ejemplo

Sea $Z \rightarrow N(0,1)$. Calcular : $P(Z \leq 2)$, $P(Z > 1.5)$, $P(1 \leq Z \leq 2.1)$, $P(Z \leq -0.65)$

$P(Z \leq 2) = \Phi(2) = 0.9773$
$P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - \Phi(1.5) = 1 - 0.9332 = 0.0668$
$P(1 \leq Z \leq 2.1) = P(Z \leq 2.1) - P(Z \leq 1) = \Phi(2.1) - \Phi(1) = 0.9821 - 0.8413 = 0.1408$
$P(Z \leq -0.65) = \Phi(-0.65) = 1 - \Phi(0.65) = 1 - 0.7422 = 0.2578$

Ejemplo

Sea $X \rightarrow N(10,25)$. Calcular: $P(X > 15)$, $P(X \leq 12)$, y $P(10 < X \leq 16)$

Primero hay que tipificar la v. a. X. Es decir, convertirla en una v.a. de media 0 y varianza 1.

$$P(X > 15) = 1 - P(X \leq 15) = 1 - P\left(\frac{X - 10}{\sqrt{25}} \leq \frac{15 - 10}{\sqrt{25}}\right)$$

Ahora $Z = \frac{X - 10}{5} \rightarrow N(0,1)$, y por tanto, la probabilidad anterior se puede escribir

$1 - P(Z \leq 1) = 1 - \Phi(1) = 1 - 0.8413 = = 0.1587$

Análogamente se obtiene:

$P(X \leq 12) = 0.6554$

$P(10 < X \leq 16) = 0.3849$

TEOREMA DE ADICIÓN

Sean $X_1, X_2, ..., X_n$ n v.a. independientes, tales que cada una de ellas se distribuye $X_i \rightarrow N(\mu_i, \sigma_i^2)$. Entonces:

$X_1 + X_2 + ... + X_n \rightarrow N(\mu_1 + \mu_2 + ... + \mu_n, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)$

Corolario

Sean $X_1, X_2, ..., X_n$ n v.a.'s independientes, tales que cada una de ellas se distribuye $X_i \rightarrow N(\mu_i, \sigma_i^2)$. Consideramos $a_1, a_2, ..., a_n$, b constantes con al menos algún $a_i \neq 0$.

Entonces la v.a. $Y = a_1 X_1 + a_2 X_2 + ... + a_n X_n + b$ se distribuye $N(\mu_Y, \sigma_Y^2)$, donde los parámetros son:

$\mu_Y = a_1\mu_1 + a_2\mu_2 + ... + a_n\mu_n + b$, $\sigma_Y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + ... + a_n^2 \sigma_n^2$

TEOREMA CENTRAL DEL LÍMITE

Sean $X_1, X_2, ..., X_n$ n v.a. independientes y con la misma distribución de media μ y varianza σ^2, entonces para x fijo se tiene:

$$\lim_{n\to\infty} P(\overline{X} \leq x) = \Phi\left(\frac{x-\mu}{\sigma/\sqrt{n}}\right)$$

Interpretación

Si se selecciona una muestra aleatoria grande de cualquier distribución con media μ y varianza σ^2, independientemente de si la distribución es discreta o continua, entonces la distribución de la v.a. será aproximadamente una normal tipificada.

Esto equivale a que:

a) $\bar{X}$ tiene aproximadamente una distribución $N(\mu, \frac{\sigma^2}{n})$

b) $X_1 + X_2 + ... + X_n$ se distribuye aproximadamente como $N(n\,\mu, n\,\sigma^2)$

Por consiguiente, podemos aproximar las distribuciones discretas Binomial y Poisson a una normal de la siguiente manera:

Si $X \rightarrow Bi(n, p)$ con $n \rightarrow \infty$, p alrededor de 0.5, entonces,

$$\frac{X - np}{\sqrt{np(1-p)}} \rightarrow N(0,1)$$

Si $Y \rightarrow Po(\lambda)$ con $\lambda \rightarrow \infty$ entonces, $\frac{Y - \lambda}{\lambda} \rightarrow N(0,1)$

En ambos casos se puede aplicar para valores de n y λ que no estén en las tablas correspondientes.

2.5. INTRODUCCIÓN A LA INFERENCIA ESTADÍSTICA

2.5.1. INTRODUCCIÓN

Inferir es, en general, establecer un nuevo conocimiento a partir de uno ya dado. En nuestro contexto, nos interesa, basándonos en la información contenida en una muestra, **inferir** información sobre una población.

La **Inferencia Estadística** es la parte de la Estadística que incluye los métodos utilizados para tomar decisiones o para obtener conclusiones sobre una característica desconocida de la población a partir de la información contenida en una o más muestras representativas de esa población. La herramienta teórica que utiliza es la teoría de la probabilidad.

La Inferencia Estadística estudia principalmente dos tipos de problemas:

a) La **Estimación**: consiste en determinar una característica desconocida de la población.

Puede ser:

puntual: determinar el valor concreto.

por intervalos: determinar un intervalo en el que esté contenida con cierto grado de probabilidad.

b) El **Contraste de hipótesis:** determinar si es aceptable, a partir de los datos muestrales, que la característica estudiada tome un valor predeterminado o pertenezca a un intervalo concreto.

Conceptos básicos

a) **Población**: es el conjunto de todos los individuos sujetos a estudio.

b) **Muestra**: es el subconjunto finito de elementos selecionados de la población.

Para que las inferencias sean válidas, las muestras deben ser representativas de la población.

c) **Muestreo:** procedimiento de obtención de una muestra. Podemos describir los sigientes tipos:

Muestreo Opinático: la selección de los elementos muestrales se realiza según el criterio del investigador. la muestra no es representativa de la población.

Muestreo Aleatorio: se seleciona de forma que cada elemento de la población tiene una probabilidad positiva de ser elegido.

Muestreo Aleatorio Simple: Cada elemento de la población tiene la misma probabilidad de ser elegido y esta probabilidad se mantiene constante a lo largo del proceso. La técnica del muestreo puede asimilarse a un modelo de extracción con reemplazamiento. Un mismo dato puede ser muestreado más de una vez. Los datos muestrales serán estocásticamente independientes .

Muestreo Aleatorio Irrestricto: Cada elemento de la población tiene en un principio la misma probabilidad de ser elegido, pero posteriormente, la probabilidad de obtener un valor en cada selección viene influida por los resultados anteriores. Se corresponde con un modelo de extracción sin reemplazamiento.

Muestreo Estratificado: Se divide a la población en estratos, niveles o grupos según criterios prefijados y la muestra se toma asignando una proporción de miembros a cada estrato y escogiendo los elementos dentro de cada estrato por muestreo aleatorio simple (m.a.s.)

El muestreo aleatorio debe utilizarse cuando los elementos de la población son homogéneos respecto a la característica a estudiar. Cuando dispongamos de información sobre la población conviene tenerla en cuenta al seleccionar la muestra. Un ejemplo son las encuestas de opinión, donde los elementos (personas) son hetereogéneos en razón a su sexo, edad, profesión, etc. Interesa en estos casos que la muestra tenga composición análoga a la población y esto se consigue con un muestreo estratificado.

Muestreo por conglomerados: cuando los elementos de la población se encuentran «de manera natural» agrupados en conglomerados, cuyo número se conoce, y podamos suponer que cada uno de estos conglomerados es una muestra representativa de la población respecto de la variable que se estudia. El muestreo con-

siste en seleccionar uno de estos conglomerados al azar y, dentro de ellos, analizar todos sus elementos o una muestra aleatoria simple.

Aquí vamos a suponer que la muestra ha sido obtenida por muestreo aleatorio simple (m.a.s.) puesto que la teoría bajo este tipo de muestreo es la más sencilla. Sin embargo, en la práctica se suele realizar el muestreo irrestricto. De todas formas, si la población tiene un número grande de elementos, la probabilidad de que un elemento salga repetido es muy pequeña y los dos tipos de muestreo serán equivalentes.

Si el muestreo es aleatorio, seleccionar un elemento de la población es realizar un experimento aleatorio y cada observación de la muestra es el valor observado de una variable aleatoria. La distribución de probabilidad de cada una de estas variables aleatorias viene determinada por la distribución de los elementos de la población. Así podemos definir:

Muestra aleatoria simple de tamaño n: es una variable aleatoria n-dimensional $X = (x_1, x_2,...,x_n)$ donde cada representa el valor observado en la i-ésima extracción y podrá tomar cualquier valor de la población. Por tanto, una muestra concreta realizada, será un valor particular, una realización de la muestra genérica.

En la medida en el que en el m.a. cada elemento de la población tiene una probabilidad de ser elegido, cada dato muestral genérico será una variable aleatoria que tendrá asociada una función de probabilidad (de cuantía o de densidad) según una determinada distribución que llamaremos **distribución de la población.** Si trabajamos con un m.a.s. cada observación es estocásticamente independiente y entonces la **función de probabilidad o de densidad conjunta** de la muestra será:

$$f(X) = f(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i)$$

Ya hemos mencionado en la introducción que uno de los propósitos es estimar cantidades desconocidas de la población (tales como la media poblacional, la varianza, etc.) llamadas parámetros poblacionales, o más brevemente **parámetros**, a partir del conocimiento de las correspondientes cantidades muestrales (tales como la media muestral, la varianza muestral, etc.) llamadas estadísticos muestrales, o más brevemente **estadísticos.**

Así pues **estadístico** es cualquier función de los valores muestrales que depende exclusivamente de éstos. Como los valores muestrales son variables aleatorias, también lo son los estadísticos.

Todas las medidas descriptivas son ejemplos de estadísticos como la media muestral $\bar{x}$, la varianza muestral s^2, ó $x_2 - x_1$; pero no es estadístico $\frac{n\sigma^2}{s^2}$ porque σ^2 es la varianza poblacional, y por tanto, no depende exclusivamente de los valores muestrales.

Consideremos todas las posibles muestras que pueden extraerse de una población dada. Para cada muestra se puede calcular un estadístico, tal como la media, la desviación típica, etc. que variará de una muestra a otra. De esta forma se obtiene una distribución del estadístico que se conoce como **distribución muestral o distribuciones en el muestreo.**

2.5.2. DISTRIBUCIONES EN EL MUESTREO

2.5.2.1. MODELOS DE DISTRIBUCIÓN DE PROBABILIDAD EN EL MUESTREO

A) DISTRIBUCIÓN CHI-CUADRADO DE PEARSON

La distribución χ^2 de Pearson con n grados de libertad se define como la distribución que sigue la variable definida como suma de los cuadrados de n variables normales tipificadas independientes.

Sean $Z_1,Z_2,...,Z_n$ un conjunto de variables aleatorias independientes, $Z_i \rightarrow N(0,1)$ entonces

$$X=\sum_{i=1}^{n} Z_i^2 \rightarrow \chi_n^2$$

Características

$$E[X] = n \qquad Var[X] = 2n$$

Para el cálculo de probabilidades, en las tablas podremos encontrar el valor de x tal que $P(X \leq x) = p$

Propiedad (Teorema de adición)

Dadas $\{X_i\}_{i=1}^{k} \rightarrow \chi_{n_i}^2$ independientes $\Rightarrow \sum_{i=1}^{k} X_i \rightarrow \chi_{n_1+n_2+...+n_k}^2$

Relación con la normal

Si $X \rightarrow \chi_n^2$, entonces $Y=\sqrt{2X} \approx N(\sqrt{2n-1},1)$ si n es grande.

B) DISTRIBUCIÓN t DE STUDENT

Dadas las variables aleatorias independientes X e Y tales que $X \rightarrow N(0,1)$ e $Y \rightarrow \chi^2_n$, la nueva variable aleatoria definida como $T = \frac{X}{\sqrt{\frac{Y}{n}}}$ sigue una distribución t de Student con n grados de libertad, que denotamos por t_n

Características

$$E[T] = 0 \text{ para } n > 2; \text{ si } n = 1 \text{ no existe media}$$

$$Var[T] = \frac{n}{n-2} \text{ para } n > 3$$

Para el cálculo de probabilidades, en las tablas podremos encontrar el valor de x tal que $P(T \leq x) = p$

Es una distribución simétrica y por tanto se cumple que f(x) = f(-x). Con lo cual se deduce que $P(T \leq x) = 1 - P(T \geq -x)$

La distribución t de Student converge a una normal cuando los grados de libertad tienden a infinito (sirve para n > 30).

C) DISTRIBUCIÓN F DE SNEDECOR

Sus principales usos son los del centrado de la igualdad de varianzas de dos poblaciones normales y, fundamentalmente, el análisis de la varianza y el diseño de experimentos, técnicas que permiten detectar la existencia o inexistencia de diferencias significativas entre muestras diferentes.

Dadas dos variables aleatorias independientes tales que sus distribuciones son $X_1 \rightarrow \chi^2_n$ y $X_2 \rightarrow \chi^2_m$, definimos la nueva variable:

$$F = \frac{\frac{X_1}{n}}{\frac{X_2}{m}}$$

que sigue una distribución F de Snedecor con n grados de libertad en el numerador y m grados de libertad en el denominador, que denotamos por $F_{n,m}$.

Para el cálculo de probabilidades, en las tablas podremos encontrar el valor de x tal que $P(F_{n,m} \leq x) = p$ para p = 0.9 y 0.99.

Se cumple que $P(F_{n,m} \leq x) = P(F_{n,m} \geq \frac{1}{x}) = 1 - P(F_{n,m} \leq \frac{1}{x})$

Un resultado particular interesante es que si $X \rightarrow t_n \Rightarrow X^2 \rightarrow F_{1,n}$

E) TEOREMA CENTRAL DEL LÍMITE (Recordatorio)

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con la misma distribución y con media μ y varianza σ^2. Entonces

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow[(aprox)]{n \rightarrow \infty} N(\mu, \frac{\sigma^2}{n})$$

Un enunciado más riguroso sería:

$$\lim_{n \rightarrow \infty} P(\overline{X} \leq x) = \Phi\left(\frac{x - \mu}{\sigma / \sqrt{n}} \right)$$

2.5.2.2. DISTRIBUCIONES MUESTRALES DE LOS PRINCIPALES ESTADÍSTICOS

EN UNA POBLACIÓN CUALQUIERA

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de variables aleatorias idénticamente distribuidas con $E[X_i] = \mu$ y $Var[X_i] = \sigma^2$. Si no conocemos la distribución de la población, no podemos, en general, calcular la distribución de los estadísticos, pero sí se podrá, en cualquier caso, determinar la esperanza y varianza de los principales estadísticos en función de los parámetros de la distribución de la población. En todos los casos denotaremos la media muestral por $\overline{x} = \overline{x}_n$.

A) Media muestral

$E[\overline{X}_n] = \mu$ (tanto para muestreo simple como irrestricto)

$Var\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$ (para m.a.s.) $Var\left[\overline{X}_n\right] = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$ (para irrestricto)

Siempre que n sea muy grande (n > 30) podemos aplicar el TCL:

$$\overline{X}_n \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

B) Varianza muestral

$$s^2 = \frac{\sum (X_i - \overline{X}_n)^2}{n} = \frac{\sum X_i^2}{n} - \overline{X}_n^2$$

Los siguientes resultados serán para m.a.s:

$$E\left[s^2\right] = \frac{n-1}{n}\sigma^2$$

Llamando $\mu_n = E[X^n]$ tenemos

$$Var\left[s^2\right] = \frac{\mu_4 - \mu_2}{n} - \frac{2\left(\mu_4 - 2\mu_2^2\right)}{n^2} - \frac{\mu_4 - 3\mu_2^2}{n^3}$$

EN UNA POBLACIÓN NORMAL

A partir de ahora supondremos que la población tiene una distribución normal, ya que los resultados que vamos a obtener por el Teorema Central del Límite, así lo requieren. Si no lo tuviera serían una aproximación cuando la muestra que tomemos sea de tamaño grande.

En estas circunstancias, ya no sólo podremos averiguar los valores de la esperanza y varianza, sino que podremos hablar de sus modelos de distribución.

A) UNA MUESTRA

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de variables aleatorias tales que $X_i \rightarrow N(\mu, \sigma^2)\ \forall i$

a) Media muestral

$$\overline{X}_n \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

Podemos ver que $E[\overline{X}_n] = \mu$ $\qquad Var\left[\overline{X}_n\right] = \frac{\sigma^2}{n}$

ya que si $X_i \to N(\mu, \sigma^2)$, por el teorema de la adición del tema anterior se cumple que $X_1 + X_2 + ...+ X_n \to N(n\mu, n\sigma^2)$, y por las transformaciones lineales

$$\frac{X_1 + X_2 + ... + X_n}{n} \to N\left(\mu, \frac{\sigma^2}{n}\right)$$

b) Varianza muestral

$$s^2 = \frac{\sum(X_i \ \overline{X}_n)^2}{n} = \frac{\sum X_i^2}{n} \ \overline{X}_n^2$$

No conocemos exactamente la distribución de s^2, solamente:

$$\frac{ns^2}{\sigma^2} \to \chi^2_{n-1}$$

pero podemos calcular

$$E\left[s^2\right] = \frac{n-1}{n}\sigma^2 \qquad Var\left[s^2\right] = \frac{2(n-1)\sigma^4}{n^2}$$

Debemos recordar que una distribución chi-cuadrado se definía como una suma de normales tipificadas elevadas al cuadrado.

c) Media muestral con varianza desconocida

Si no conocemos el valor de σ^2, la distribución que hemos deducido anteriormente para el estadístico media muestral no nos sirve, podemos utilizar entonces:

$$\frac{\overline{X_n} - \mu}{s}\sqrt{n-1} \to t_{n-1}$$

Recordemos que la distribución t de Student se definía como el cociente entre una normal tipificada y la raiz cuadrada de una chi-cuadrado dividida por sus grados de libertad.

d) Proporción muestral

Tenemos una población Bernoulli y elegimos una muestra donde $X_i \rightarrow Be(p)$.

Consideremos la variable aleatoria X = «nº de éxitos en la muestra», que sigue una distribución Bi(n,p). Definimos P = «proporción de éxitos en n extracciones» entonces P = X/n

$$E[P]=p \qquad Var[P]=\frac{pq}{n}$$

B) DOS MUESTRAS INDEPENDIENTES

Supongamos ahora que queremos comparar dos variables o dos poblaciones. Para ello necesitamos tomar dos muestras y vamos a suponer que estas muestras se toman independientemente.

Seguimos suponiendo que la distribución de las dos poblaciones es Normal.

Sean $X_1,X_2,...,X_n$ una muestra aleatoria de variables aleatorias independientes tales que $X_i \rightarrow N(\mu_x, \sigma_x^2)\ \forall i = 1,2,...,n$ e $Y_1,Y_2,...,Y_m$ una muestra aleatoria de variables aleatorias tales que $Y_j \rightarrow N(\mu_y, \sigma_y^2)\ \forall j = 1,2,...,m$.

a) Diferencia de medias muestrales con varianzas conocidas

$$\overline{X_n}-\overline{Y_m} \rightarrow N\left(\mu_x-\mu_y, \frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}\right)$$

Un caso particular sería la diferencia de proporciones

$$P_1-P_2 \rightarrow N\left(p_1-p_2, \frac{p_1q_1}{n}+\frac{p_2q_2}{m}\right)$$

b) Diferencia de medias muestrales con varianzas desconocidas pero iguales

$$\frac{\left(\overline{X_n}-\overline{Y_m}\right)-\left(\mu_x-\mu_y\right)}{\sqrt{\frac{ns_x^2+ms_y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}} \rightarrow t_{n+m-2}$$

c) Cociente de varianzas muestrales

No conocemos la distribución de $\frac{s_x^2}{s_y^2}$ exactamente, sólo:

$$\frac{\frac{n \cdot s_x^2}{\sigma_x^2 (n-1)}}{\frac{m \cdot s_y^2}{\sigma_y^2 (m-1)}} \rightarrow F_{(n-1),(m-1)}$$

C) DOS MUESTRAS RELACIONADAS

Supongamos ahora que queremos comparar dos variables pero que las dos muestras que se toman no son independientes (los datos están apareados).

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de variables aleatorias tales que $X_i \rightarrow N(\mu_x, \sigma_x^2)\ \forall i = 1,2,\ldots,n$ e $Y_1, Y_2, \ldots, Y_n$ una muestra aleatoria de variables aleatorias tales que $Y_j \rightarrow N(\mu_y, \sigma_y^2)\ \forall j = 1,2,\ldots,n$. Las muestras están relacionadas. Notar que en este caso n = m.

a) Diferencia de medias muestrales

Sea $D_i = X_i - Y_i$ y sea su s_d desviación típica muestral. Podemos utilizar el siguiente resultado:

$$\sqrt{n-1}\,\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{s_d} \rightarrow t_{n-1}$$

2.6. ESTIMACIÓN

2.6.1. ESTIMACIÓN PUNTUAL

Supongamos que estamos estudiando una población cuya distribución es conocida excepto en un parámetro (μ, σ^2, λ,...) al que llamaremos θ. A la distribución de la población la denotaremos por f(x).

Diremos que nos encontramos ante un **problema de estimación** cuando, dada una población con una distribución f(x, θ) donde θ es un parámetro desconocido, aventuremos o infiramos en base a los datos muestrales $X_1, X_2, \ldots, X_n$ el valor de θ. Si al inferir el parámetro damos un único valor estaremos ante un problema de *estimación puntual.*

Estimador puntual: $\hat{\theta} = f(X_1, X_2, ..., X_n)$ será una función de la muestra aleatoria (un estadístico) que utilizaremos para estimar el valor del parámetro.

Estimación: $\hat{\theta}$ valor obtenido del estimador al sustituir por los valores de una muestra completa.

Cuando no haya lugar para la confusión designaremos al estimador simplemente por $\hat{\theta}$.

Un estimador es, por tanto, un estadístico y, por ello, es una v.a. con una determinada distribución de probabilidad llamada distribución muestral.

Dado un parámetro, podríamos utilizar distintos estimadores puntuales para estimarlo. Por ejemplo, para estimar la varianza de la población podemos utilizar la varianza muestral o la cuasi-varianza muestral. ¿Cuál es mejor? Veamos a continuación como comprobar si un estadístico es un buen estimador de un parámetro. Para ello le exigiremos una serie de propiedades. Como el estadístico es una variable aleatoria, las propiedades se las tenemos que exigir a su distribución de probabilidad.

2.6.2. PROPIEDADES DE LOS ESTIMADORES

Un estadístico se considera un buen estimador de un parámetro si cumple: ser insesgado, ser consistente y ser eficiente.

A) INSESGADEZ

Un estimador $\hat{\theta}$ es insesgado si $E[\hat{\theta}] = \theta$. Es decir, su distribución está centrada en el parámetro a estimar.

Ejemplos: Sea una m.a.s. $X_1, X_2, ..., X_n$ tal que $E[X_i] = \mu$ y $Var[X_i] = \sigma^2$

a) Consideremos como estimador de la media poblacional a la media muestral. Es decir $\hat{\mu} = \overline{X}_n$. Sabemos que $E[\overline{X}_n] = \mu$. Por tanto, la media muestral es estimador insesgado de la media poblacional.

b) Supongamos como estimador de la varianza poblacional a la varianza muestral, $\hat{\sigma}^2 = s_n^2$. Sabemos que $E[s_n^2] = \frac{n-1}{n}\sigma^2$. Por tanto, la varianza muestral no es un estimador insesgado de la varianza poblacional.

c) Consideremos ahora como estimador de la varianza poblacional a la cuasi-varianza muestral, $\hat{\sigma}^2 = s_n^{*2}$. Sabemos que $E[s_n^{*2}] = \sigma^2$. Por tanto, la cuasi-varianza muestral es un estimador insesgado de la varianza poblacional.

B) CONSISTENCIA

Diremos $\hat{\theta}$ que es un estimador consistente de θ si cumple:

$$\lim_{n\to\infty} E[\hat{\theta}]=\theta \text{ y } \lim_{n\to\infty} Var[\hat{\theta}]=0 \Leftrightarrow \lim_{n\to\infty} P[\hat{\theta}=\theta]=1$$

Esto significa que si tomáramos la mayor muestra posible, el estimador coincidiría con el valor del parámetro.

Ejemplo: Veamos que los estimadores de los cuales hemos hablado en el apartado anterior son consistentes.

a) Si consideramos $\hat{\mu}=\overline{X}_n$. Se cumple que $E[\overline{X}_n]=\mu$ y $Var[\overline{X}_n]=\frac{\sigma^2}{n}$. Pero tomando límites, $\lim_{n\to\infty}\frac{\sigma^2}{n}=0$. Por tanto, $\overline{X}_n$ es estimador consistente.

b) Si consideramos $\hat{\sigma}^2=s_n^2$. Se cumple que:

$$E[s_n^2]=\frac{n-1}{n}\sigma^2 \text{ y } Var[s_n^2]=\frac{2(n-1)}{n^2}\sigma^4$$

Tomando límites,

$$\lim_{n\to\infty}\frac{n-1}{n}\sigma^2=\sigma^2 \text{ y } \lim_{n\to\infty}\frac{2(n-1)}{n^2}\sigma^4=0$$

Por tanto s_n^2 es un estimador consistente de σ^2.

c) Si consideremos $\sigma^2=s_n^{*2}$. Se cumple que

$$E[s_n^{*2}]=\sigma^2 \text{ y } Var[s_n^{*2}]=\frac{2}{n-1}\sigma^4$$

Tomando límites,

$$\lim_{n\to\infty}\sigma^2=\sigma^2 \text{ y } \lim_{n\to\infty}\frac{2}{n-1}\sigma^4=0$$

Por tanto, s_n^{*2} es un estimador consistente de la varianza poblacional.

C) EFICIENCIA

Dados dos estimadores de θ, $\hat{\theta}_1$ y $\hat{\theta}_2$, decimos que $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$, si $Var\,[\hat{\theta}_1] < Var\,[\hat{\theta}_2]$. Nos interesa el que tenga menos dispersión. Para comparar la eficiencia se construye el cociente $\frac{Var\,[\hat{\theta}_1]}{Var\,[\hat{\theta}_2]}$. Si es mayor que 1, entonces $\hat{\theta}_2$ es más

más eficiente; si es igual a 1, entonces ambos estimadores son igual de eficientes; si es menor que 1, entonces $\hat{\theta}_1$ es más eficiente.

Ejemplo: Consideremos como estimadores de σ^2 a $s_n^2 = s_n^{*2}$ y . Calculamos el cociente de las varianzas:

$$\frac{Var[s_n^2]}{Var[s_n^{*2}]} = \frac{\frac{2(n-1)}{n^2}\sigma^4}{\frac{2}{(n-1)}\sigma^4} = \frac{(n-1)^2}{n^2} < 1$$ Por tanto, s_n^2 es más eficiente.

2.6.3. MÉTODOS DE CÁLCULO DE LOS ESTIMADORES PUNTUALES

De los diferentes métodos de cálculo de los estimadores, nos centramos en el de máximo verosímil

ESTIMACIÓN MÁXIMO-VEROSÍMIL

Sea X una variable aleatoria con distribución $f(x;\theta)$ donde θ es el parámetro desconocido. Sean $X_1, X_2, \ldots X_n$ n variables aleatorias independientes con la misma distribución que X; es decir, sea $(X_1, X_2, \ldots X_n)$ una m.a.s. Bajo estas condiciones la distribución conjunta de las variables aleatorias $X_1, X_2, \ldots X_n$ será igual al producto de las marginales

$$f(x_1, x_2, \ldots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \ldots \cdot f(x_n; \theta)$$

Si consideramos $X_1, X_2, \ldots X_n$ fijos y estudiamos esta función como función de θ recibe el nombre de **función de verosimilitud** y se denota por $V(\theta)$.

Sean $\theta_1 = u_1(X_1, X_2, \ldots X_n)$, $\theta_2 = u_2(X_1, X_2, \ldots X_n)$, etc. diversos estimadores de θ. De todos ellos pretendemos elegir el que haga máxima la función de verosimilitud. Es decir, un estimador $\hat{\theta}$ será estimador máximo-verosímil (EMV) de θ si maximiza $V(\theta)$.

Debido a que la función de verosimilitud es no negativa, continua y creciente, alcanzará su máximo en los mismos puntos que su logaritmo y por ello, y por razones de cálculo se suele maximizar ln $V(\theta)$ cuando esta depende de exponenciales. Así pues, deberemos resolver la siguiente ecuación:

$$\frac{d(\ln V(\theta))}{d\theta} = 0$$

En el caso de dos o más parámetros desconocidos, el procedimiento es el mismo. Por ejemplo, si tuviéramos $V(\theta_1,\theta_2,\theta_3)$ los tres estimadores máximo verosímiles serán los que maximizan la función $V(\theta_1,\theta_2,\theta_3)$ o su logaritmo. Se obtendrían al resolver las ecuaciones siguientes:

$$\frac{d(\ln V(\theta_1,\theta_2,\theta_3))}{d\theta_1}=0$$

$$\frac{d(\ln V(\theta_1,\theta_2,\theta_3))}{d\theta_2}=0$$

$$\frac{d(\ln V(\theta_1,\theta_2,\theta_3))}{d\theta_3}=0$$

Propiedades de los EMV

a) Son consistentes.

b) Son asintóticamente eficientes. Es decir, tienen la varianza mínima cuando el tamaño muestral tiende a infinito.

c) Si $\hat{\theta}$ es estimador suficiente de θ, el EMV de θ es función de $\hat{\theta}$.

d) Son asintóticamente normales. Es decir, su distribución tiende a la distribución normal cuando tiende a infinito el tamaño de la muestra.

e) Si $\hat{\theta}$ es EMV de θ, entonces $g(\hat{\theta})$ es EMV de $g(\theta)$, siendo g una aplicación biyectiva.

Ejemplo: Obtener el EMV del parámetro de una v.a. X que sigue una distribución de Bernoulli, $X \rightarrow Be(p)$. Su función de cuantía es $f(x; p) = p^x (1 - p)^{1-x}$

Si elegimos una muestra de tamaño n, la función de verosimilitud correspondiente será:

$$V(p)=f(x_1;p)\cdot f(x_2;p)\cdot\ldots\cdot f(x_n;p)=$$
$$p^{x_1}(1-p)^{1-x_1}\,p^{x_2}(1-p)^{1-x_2}\ldots p^{x_n}(1-p)^{1-x_n}=$$
$$=p^{\sum x_i}(1-p)^{n-\sum x_i}$$

Tomando logaritmos tenemos

$$\ln V(p)=\sum x_i \ln p+(n-\sum x_i)\ln(1-p)$$

Para obtener el EMV de p debemos resolver la ecuación $\frac{d(\ln V(p))}{dp}=0$

En este caso:

$$\frac{d(\ln V(p))}{dp}=\sum x_i\frac{1}{p}+(n-\sum x_i)\frac{(-1)}{(1-p)}=0$$

Haciendo operaciones e igualando denominadores obtenemos:

$$(1-p)\sum x_i-p(n-\sum x_i)=$$
$$\sum x_i-p\sum x_i-np+p\sum x_i=\sum x_i-np=0$$

Despejando el valor de p obtenemos el estimador $\hat{p}=\frac{\sum x_i}{n}$

Por tanto EMV(p) = $\frac{\sum x_i}{n}$. Es decir, la proporción de éxitos de la muestra.

2.6.4. ESTIMACIÓN POR INTERVALOS DE CONFIANZA

En la estimación puntual atribuimos al parámetro el valor correspondiente del estimador obtenido en la muestra aleatoria de tamaño n. Es claro, que dicho valor dificilmente coincidirá con el verdadero valor del parámetro aunque el tamaño de la muestra sea muy grande.

La **estimación por intervalos** consiste en atribuir al parámetro desconocido un rango de posibles valores (en base a los datos muestrales) que tengan una alta probabilidad de incluir entre ellos al valor del parámetro desconocido. Para ello será imprescindible conocer la distribución muestral del estadístico utilizado.

El intervalo estimado que debe contener al parámetro se llama **intervalo confidencial o de confianza**. Denominamos **límites confidenciales** a los extremos de dicho intervalo. Llamaremos **nivel de confianza** a la probabilidad de que un intervalo contenga al parámetro desconocido y se suele denotar por $1-\alpha$. Se llama **nivel de riesgo o de significación** al valor de α.

Es decir, $P(\theta \in [a,b]) = 1-\alpha$. Esto indica que el $(1-\alpha)\%$ de intervalos construidos contendrán al parámetro desconocido.

Denominaremos **error muestral máximo** a la diferencia entre el valor de la estimación muestral y el valor del parámetro; es decir,

$$E=|\hat{\theta}-\theta|=\frac{b-a}{2}.$$

Ejemplo: Sea θ el parámetro desconocido y $\hat{\theta}$ el estimador que consideramos el cual sigue una distribución $N(\theta,\sigma^2)$. Supongamos un error muestral máximo $E=|\hat{\theta}-\theta|=2\sigma$.

Si calculamos la probabilidad de tener ese error o uno menor, obtendremos:

$$P([\hat{\theta}-\theta]\leq 2\sigma)=P(-2\sigma\leq\hat{\theta}-\theta\leq 2\sigma)=P\left(-2\leq\frac{\hat{\theta}-\theta}{\sigma}\leq 2\right)=\phi(2)-\phi(-2)=$$

$$=2\phi(2)-1=0.9544$$

ya que el estimador seguía una distribución normal.

Esta probabilidad podemos escribirla también de la siguiente forma:

$$P(-2\sigma\leq\hat{\theta}-\theta\leq 2\sigma)=P(-2\sigma+\hat{\theta}\leq\theta\leq 2\sigma+\hat{\theta})=$$
$$P(\theta\in[\hat{\theta}-2\sigma,\hat{\theta}+2\sigma])=0.9544$$

Por tanto, el intervalo de confianza $[\hat{\theta}-2\sigma,\hat{\theta}+2\sigma]$ tiene un nivel de confianza $1-\alpha=0.9544$ o un nivel de significación de $\alpha=0.0456$. Esto equivale a decir que tenemos la confianza 0.9544 de que, extraída una muestra y calculado el valor de $\hat{\theta}$, éste no se aleja del parámetro más de dos desviaciones típicas o un riesgo de 0.0456 de que se aleja más de esa cantidad.

Normalmente lo que se hace es fijar de antemano el nivel de confianza y se busca el intervalo correspondiente a ese nivel de confianza utilizando la distribución muestral del estadístico.

2.6.4.1. CASOS PATICULARES

A) INTERVALO DE CONFIANZA PARA LA MEDIA

Si desconocemos la distribución de la población, podemos hallar un intervalo de confianza para la media, basándonos en un resultado que conocemos como *Desigualdad de Tchebychev.*

Sea X una v.a. cualquiera con media μ y varianza σ^2. Se cumple que:

$$P\left(\mu-\frac{k}{\sigma}\leq X\leq\mu+\frac{k}{\sigma}\right)\geq 1-\frac{1}{k^2}$$

Usando el anterior resultado, aplicándolo a la variable aleatoria X y tomando $\alpha=\frac{1}{k^2}$, obtendríamos que el intervalo para un nivel 1 – α sería:

$$\left[\overline{X}_n-\sigma/\sqrt{n\alpha},\overline{X}_n+\sigma/\sqrt{n\alpha}\right]$$

Para analizar los resultados que presentamos a continuación, supongamos una población que se distribuye según una normal de media μ y varianza poblacional σ^2. También servirán cuando la población no es normal pero el tamaño muestral es grande.

a) Si σ^2 es conocida

Ya sabemos que $\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\rightarrow N(0,1)$. Sea $z_{1-\frac{\alpha}{2}}$ el percentil de la distribución normal; es decir, $\phi(z)=1-\frac{\alpha}{2}$.

Entonces:

$$P\left(-z_{1-\frac{\alpha}{2}}\leq\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}\leq z_{1-\frac{\alpha}{2}}\right)=1-\alpha$$

Haciendo operaciones $P\left(\overline{X}_n-z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\leq\mu\leq\overline{X}_n+z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\right)=1-\alpha$

Por tanto, el intervalo de confianza para μ será:

$$\left[\overline{X}_n-z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n},\overline{X}_n+z_{1-\frac{\alpha}{2}}\sigma/\sqrt{n}\right]$$

b) Si σ^2 es desconocida

En este caso tenemos que $\frac{\overline{X}_n-\mu}{s_n}\sqrt{n-1}\rightarrow t_{n-1}$

Por el mismo razonamiento anterior, si llamamos $t_{n-1\,1-\frac{\alpha}{2}}$ al percentil de la distribución t de Student tal que $P(t_{n-1}\leq x)=1-\frac{\alpha}{2}$, el intervalo de confianza al nivel de significación α (o equivalentemente, al nivel de confianza 1- α) será:

$$\left[\overline{X}_n - t_{n-1\,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n-1}}, \overline{X}_n + t_{n-1\,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n-1}}\right]$$

Ejemplo: Extraemos una m.a.s. de 61 azulejos. Responden a una prueba de medida de longitud, en la que alcanzan una media de 80 cm y una varianza de 100. ¿Entre qué límites se hallará la verdadera longitud media de los azulejos, a un nivel de confianza del 99%?

$$1-\alpha=0'99 \Rightarrow \alpha=0.01 \Rightarrow 1-\frac{\alpha}{2}=0.995$$

La varianza poblacional es desconocida y la población no es normal, pero el tamaño muestral es mayor que 30, por tanto, el intervalo correspondiente será:

$$\left[\overline{X}_n - t_{n-1\,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n-1}}, \overline{X}_n + t_{n-1\,1-\frac{\alpha}{2}}\frac{s_n}{\sqrt{n-1}}\right]$$

Buscamos en las tablas la distribución t de Student $t_{60_{0,995}}=2.66$.

Sabemos que $\overline{X}_n=80$ y $s_n = 10$. Sustituyendo en el intervalo de confianza tenemos:

$$\left[80-2.66\frac{10}{\sqrt{60}}, 80+2.66\frac{10}{\sqrt{60}}\right]$$

por tanto, $\mu \in$ [76.57,83.43], con un nivel de confianza del 99%.

B) INTERVALO DE CONFIANZA PARA LA PROPORCIÓN

Si en una población Bernoulli de parámetro p definimos la v.a. X = nº éxitos en la muestra, X sigue una distribución binomial de parámetros (n,p). Si la muestra es grande, tenemos que la proporción muestral P = X/n se distribuye aproximadamente como una normal $N\left(p, \frac{pq}{n}\right)$ y podremos usar el teorema central del límite.

En una población Bernoulli, $\mu = p$, $\sigma^2 = p(1-p)$, y si denotamos por P a la proporción en la muestra $\overline{X}_n = P$. Así pues podemos aplicar el intervalo de confianza pa-

ra la media con varianza conocida visto anteriormente, sustituyendo lo anterior y aproximando p(1-p) por P(1-P), un intervalo de confianza aproximado para p a nivel $1-\alpha$ sería:

$$\left[P-z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}},P+z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}\right]$$

Ejemplo: Uno de los líderes de un colectivo laboral desea plantear una cuestión a todos los miembros del grupo. Si más de la mitad respondieran NO entonces preferiría no plantearla para no minar su prestigio. Para salir de dudas, elige aleatoriamente a 100 trabajadores a los que hace la pregunta y sólo 30 responden NO. ¿Entre qué límites se hallará la verdadera proporción al nivel del 95%?

Como el tamaño muestral es grande, podemos aplicar el teorema central del límite. Tenemos $1-\alpha=0.95\Rightarrow 1-\frac{\alpha}{2}=0.975\Rightarrow z_{1-\frac{\alpha}{2}}=1.96$

Sustituyendo los valores en el intervalo correspondiente:

$$\left[0.3-1.96\sqrt{\frac{0.3\cdot 0.7}{100}},0.3+1.96\sqrt{\frac{0.3\cdot 0.7}{100}}\right]=[0.2102,0.3898]$$

Por tanto, la verdadera proporción está en el intervalo [0.2102,0.3898] con un nivel de confianza del 95%.

C) INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE MEDIAS

Suponemos dos poblaciones independientes

$$X\rightarrow N(\mu_1,\sigma_1^2), \qquad Y\rightarrow N(\mu_2,\sigma_2^2)$$

Tomamos muestras de tamaño n_1 y n_2, respectivamente.

a) Si σ_1^2 y σ_2^2 son conocidas, como $\overline{X}_{n_1}-\overline{Y}_{n_2}\rightarrow N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$, el intervalo de confianza será:

$$\mu_1-\mu_2\in\left[\overline{X}_{n_1}-\overline{Y}_{n_2}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\overline{X}_{n_1}-\overline{Y}_{n_2}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]$$

b) Si σ_1^2 y σ_2^2 son desconocidas pero iguales,

como $\dfrac{\overline{X}_{n_1}-\overline{Y}_{n_2}-(\mu_1-\mu_2)}{\sqrt{\dfrac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}} \to t_{n_1+n_2-2}$, el intervalo de confianza será:

$$\left[\overline{X}_{n_1}-\overline{Y}_{n_2}-t_{n_1+n_2-2_{1-\frac{\alpha}{2}}}\sqrt{\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},\overline{X}_{n_1}-\overline{Y}_{n_2}+t_{n_1+n_2-2_{1-\frac{\alpha}{2}}}\sqrt{\frac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right]$$

Ejemplo: Dos empresas tienen dos métodos distintos para elegir a sus trabajadores. Los dos desean comprobar el tiempo promedio que toma la inscripción de los trabajadores. En cada universidad se tomaron los tiempos de inscripción de 31 trabajadores tomados al azar. Las medias y desviaciones típicas muestrales fueron: $\overline{x} = 20.3$, $s_x = 2.5$, $\overline{y} = 23$, $s_y = 3$. Si se supone que el muestreo se llevó a cabo en dos poblaciones normales e independientes, obtener los intervalos de confianza al nivel de riesgo 0.05 para la diferencia entre las medias del tiempo de inscripción para las dos universidades,

a) suponiendo que las varianzas poblacionales son $\sigma_x^2=9$, $\sigma_y^2=10$;

b) suponiendo que las varianzas poblacionales son desconocidas pero iguales.

Para el apartado *a*) $\alpha=0.05\Rightarrow 1-\alpha=0.95\Rightarrow 1-\dfrac{\alpha}{2}=0.975\Rightarrow z_{1-\frac{\alpha}{2}}=1.96$

Sustituyendo los valores en el intervalo obtenemos:

$$\mu_1-\mu_2\in\left[20.3-23-1.96\sqrt{\frac{9}{31}+\frac{10}{31}},20.3-23+1.96\sqrt{\frac{9}{31}+\frac{10}{31}}\right]=$$

$$=[-2.7-1.53,-2.7+1.53]=[-4.23,-1.17]$$

Para el apartado *b*), buscamos en la tabla de la t de Student $t_{31+31-2_{0.975}}=2$.

Sustituyendo los valores en el intervalo obtenemos:

$$\left[20.3-23-2\sqrt{\frac{31\cdot 2.3^2+31\cdot 3^2}{31+31-2}\left(\frac{1}{31}+\frac{1}{31}\right)},20.3-23+2\sqrt{\frac{31\cdot 2.3^2+31\cdot 3^2}{31+31-2}\left(\frac{1}{31}+\frac{1}{31}\right)}\right]=$$

$$=[-2'7-1.4,-2.7+1.4]=[-4.1,-1.3]$$

D) INTERVALO DE CONFIANZA PARA LA DIFERENCIA DE PROPORCIONES

Sean $X \to Be(p_1)$ e $Y \to Be(p_2)$ dos poblaciones independientes con p_1 y p_2 desconocidos. Extraemos muestras de tamaño n_1 y n_2, respectivamente. Como $P_1 - P_2 \to N\left(p_1 - p_2, \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}\right)$ y desconocemos los valores de p_1 y p_2, aproximaremos las proporciones poblacionales por las proporciones muestrales correspondientes. Por tanto, el intervalo de confianza será:

$$p_1 - p_2 \in \left[P_1 - P_2 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1 \cdot Q_1}{n_1} + \frac{P_2 \cdot Q_2}{n_2}}, P_1 - P_2 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{P_1 \cdot Q_1}{n_1} + \frac{P_2 \cdot Q_2}{n_2}}\right]$$

Caso particular: Si tenemos $p_1 = p_2 = p$, entonces $E[P_1 - P_2] = 0$ y $Var[P_1 - P_2] = pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$. Lo que haremos es sustituir p por $\frac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2}$

Ejemplo: En dos grandes empresas se lleva a cabo un estudio sobre la proporción de mujeres entre sus empleados diplomados y licenciados. De cada empresa se toma una m.a.s. de 40 empleados entre los diplomados y licenciados, obteniéndose que en la empresa A, había 16 mujeres y en la empresa B, 22 mujeres. Obtener el intervalo de confianza para la diferencia de proporciones poblacionales al nivel de confianza 0.96 ¿Podemos pensar que la proporción es la misma?

$$1-\alpha = 0.96 \Rightarrow 1 - \frac{\alpha}{2} = 0.98 \Rightarrow z_{1-\frac{\alpha}{2}} = 2.05$$

$$P_1 = \frac{16}{40} = 0.4 \qquad P_2 = \frac{22}{40} = 0.55$$

Sustituyendo en el intervalo:

$$\left[0.4 - 0.55 - 2.05\sqrt{\frac{0.4 \cdot 0.6}{40} + \frac{0.55 \cdot 0.45}{40}}, 0.4 - 0.55 + 2.05\sqrt{\frac{0.4 \cdot 0.6}{40} + \frac{0.55 \cdot 0.45}{40}}\right]$$

$$= [-0.15 - 0.2265, -0.15 + 0.2265] = [-0.3765, 0.0765]$$

El intervalo contiene al cero, pero el extremo inferior se aleja bastante de cero.

E) INTERVALO DE CONFIANZA PARA LA VARIANZA

Si tenemos una población $X \to N(\mu, \sigma^2)$ con σ^2 desconocida, entonces

$$\frac{(n-1)s_{n-1}^2}{\sigma^2} \rightarrow \chi_{n-1}^2$$

El intervalo de confianza para la varianza poblacional al nivel de confianza lo podemos obtener como sigue:

$$P\left(\chi_{n-1\,\frac{\alpha}{2}}^2 \leq \frac{(n-1)s_{n-1}^2}{\sigma^2} \leq \chi_{n-1\,1-\frac{\alpha}{2}}^2 \right) = 1-\alpha$$

Despejando σ^2 tenemos:

$$P\left(\frac{(n-1)s_{n-1}^2}{\chi_{n-1\,1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{(n-1)s_{n-1}^2}{\chi_{n-1\,\frac{\alpha}{2}}^2} \right) = 1-\alpha$$

Es decir,

$$\sigma^2 \in \left[\frac{(n-1)s_{n-1}^2}{\chi_{n-1\,1-\frac{\alpha}{2}}^2}, \frac{(n-1)s_{n-1}^2}{\chi_{n-1\,\frac{\alpha}{2}}^2} \right]$$

Ejemplo: De acuerdo con las tablas de altura, los varones tienen una altura superior a las mujeres en la población española. Según las últimas tablas en el servicio militar, los varones entre 18 y 20 años presentan una varianza de 0.0529 de las mujeres no tenemos información, por ello tomamos una muestra de 101 mujeres entre 18 y 20 años y obtenemos $s_{n-1} = 0.18$ ¿Entre qué valores se encontrará la verdadera varianza a un nivel de 0.95 de confianza?

$$1-\alpha = 0.95 \Rightarrow 1-\frac{\alpha}{2} = 0.975 \Rightarrow \chi_{100_{0.025}}^2 = 74.22$$

Sustituyendo en el intervalo tendremos:

$$\left[\frac{100 \cdot 0.18^2}{129.56}, \frac{100 \cdot 0.18^2}{74.22} \right] = [0.025, 0.0436]$$

F) INTERVALO DE CONFIANZA PARA EL COCIENTE DE VARIANZAS

La distribución muestral del cociente de varianzas muestrales, cuando teníamos dos poblaciones normales e independientes era:

$$\frac{s_{n-1}^2 / \sigma_1^2}{s_{m-1}^2 / \sigma_2^2} \rightarrow F_{n-1,m-1}$$

A partir de aquí deducimos el intervalo de confianza para el cociente de varianzas poblacionales al nivel de $1-\alpha$ y obtenemos

$$\frac{\sigma_1^2}{\sigma_2^2} \in \left[\frac{s_{n-1}^2}{s_{m-1}^2} \frac{1}{F_{n-1,m-1}}, \frac{s_{n-1}^2}{s_{m-1}^2} \frac{1}{F_{n-1,m-1}} \right]$$

Ejemplo: Con los datos del ejemplo de las dos empresas, calcular el intervalo de confianza para el cociente de varianzas al nivel de confianza 0.95. ¿Podríamos aceptar la suposición de que las varianzas poblacionales son iguales?

$$1-\alpha=0.95 \Rightarrow 1-\frac{\alpha}{2}=0.975 \Rightarrow F_{30,30_{0.975}}=2.07 \text{ y}$$

$$\frac{\alpha}{2}=0.025 \Rightarrow F_{30,30_{0.025}}=\frac{1}{F_{30,30_{0.975}}}=\frac{1}{2.07}$$

$$s_{n-1}^2=\frac{n}{n-1}s_n^2=\frac{31}{30}2.3^2=5.47 \quad s_{m-1}^2=\frac{m}{m-1}s_m^2=\frac{31}{30}3^2=$$

Sustituyendo en el intervalo obtenemos

$$\left[\frac{5.47}{9.3} \frac{1}{2.07}, \frac{5.47}{9.3} \frac{1}{1/2.07} \right]=[0.284,1.218]$$

El intervalo contiene al 1 y los extremos están bastante próximos al 1. Hay mayor diferencia por el extremo inferior, lo que indica que la varianza de la población X es menor que la de la población Y.

2.6.5. PRECISIÓN Y TAMAÑO DE LA MUESTRA

En general, cuanto más estrecho es un intervalo de confianza mayor precisión tendrá nuestra estimación (será menor el error muestral máximo). Ahora bien, la amplitud de un intervalo depende de dos factores: el nivel de confianza que decidimos utilizar y el tamaño del error típico (es la desviación típica) del estadístico utilizado como estimador.

Si disminuimos el nivel de confianza, diminuye la amplitud del intervalo, pero aumenta el riesgo. Debemos intentar reducir la amplitud del intervalo manteniendo constante el nivel de confianza; para ello hay que reducir el error típico del estimador.

En el caso de la media, el error típico es $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. Por lo tanto, variando el tamaño muestral variaremos el error típico. Al aumentar n, disminuye $\sigma_{\bar{x}}$. Por tanto, manipulando el tamaño de la muestra podemos obtener los intervalos de la precisión que deseemos.

Para la media:

$$E = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{E} \right)^2 \text{ ó}$$

$$E = t_{n-1\,1-\frac{\alpha}{2}} \frac{s_{n-1}}{\sqrt{n}} \Rightarrow n = \left(t_{n-1\,1-\frac{\alpha}{2}} \frac{s_{n-1}}{E} \right)^2$$

Para la proporción:

$$E = z_{1-\frac{\alpha}{2}} \sqrt{\frac{PQ}{n}} \Rightarrow n = \frac{PQz^2_{1-\frac{\alpha}{2}}}{E^2}$$

Ejemplo: Queremos estimar la media de una población normal con varianza poblacional igual a 4. ¿Qué tamaño muestral debemos tomar para que E = 0.02 al nivel de confianza 0.95?

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow 1-\frac{\alpha}{2} = 0.975 \Rightarrow z_{0.975} = 1.96$$

Como conocemos la varianza poblacional, el tamaño muestral será:

$$n = \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{E} \right)^2 = \left(1.96 \frac{2}{0.02} \right)^2 = (196)^2 = 38416$$

¿y si queremos un error E=1 al mismo nivel de confianza? En este caso

$$n = \left(z_{1-\frac{\alpha}{2}} \frac{\sigma}{E} \right)^2 = \left(1'96 \frac{2}{1} \right)^2 = (3'92)^2 = 15'37.$$

Redondeamos y tenemos que n = 16.

3 CONTROL ESTADÍSTICO DE LA CALIDAD. DIAGRAMAS DE CONTROL

3.1. GRÁFICOS DE CONTROL POR VARIABLES

3.1.1. ¿QUÉ ES UN GRÁFICO DE CONTROL?

Un gráfico de control es un gráfico en el que se representa el comportamiento de un proceso anotando sus datos ordenados en el tiempo.

El objetivo principal de los gráficos de control es detectar lo antes posible cambios en el proceso que puedan dar lugar a la producción de unidades defectuosas, y ello se consigue minimizando el tiempo que transcurre desde que se produce un desajuste hasta que se detecta.

Asimismo, el Control estadístico de procesos (CEP) hay que verlo también como una herramienta de mejora continua de la calidad de los productos, puesto que hoy en día se mide la calidad de un producto como un valor que es inversamente proporcional a la variabilidad que presentan sus características de calidad en el cliente. Puesto que reducir la variabilidad debido a las causas comunes resulta más complejo, en general el CEP ayuda a la detección de causas asignables para tomar acciones en función de su naturaleza.

Respecto a los distintos temas del Control de calidad, podemos hacer referencia a los siguientes trabajos publicados: Montgomery (1985), Peña (1986a, b), Prat-Bartes y otros (1997), Pérez-López (2001, 2003) y como trabajo previo, Juan-Verdoy y otros (2003).

Por lo tanto, el objetivo primordial de los gráficos es la **detección rápida** de la presencia de causas asignables en un proceso y para ello son importantes los siguientes puntos:

- El riesgo que se está dispuesto a admitir cada vez que se decida que una causa asignable ha entrado en el proceso.
- El **cambio mínimo** es el valor del parámetro que se desea detectar.
- El tiempo medio esperado entre **desajustes**.

Se entiende que un gráfico de control da «falsas alarmas» cuando las observaciones de un proceso en estado de control llevadas al gráfico son interpretadas erróneamente como señales de aparición de causas asignables. Para que esto no ocurra con frecuencia, se toman los límites tales que la probabilidad de falsa alarma sea del orden de un 3%, es decir, se ha de estar muy seguro para aceptar que el proceso ha cambiado. Para ello, en el caso de la distribución normal los límites de control se han de situar a una distancia de la línea central de tres desviaciones típicas del estadístico que se situa en el gráfico.

El tiempo medio en que se producen los desajustes en el proceso, determinará la frecuencia de muestreo: cuanto más estable en un proceso menos inversión (tiempo y dinero) debe dedicarse a controlarlo. Puesto que se ha de minimizar el tiempo de detección de un cambio en el proceso, la frecuencia de muestreo ha de ser superior a la del tiempo esperado entre cambios.

Por otra parte, además del tiempo medio entre desajustes, es importante estimar los costes de producir fuera de control en este tiempo, de tomar datos del proceso y de ajustarlo y, en función de ellos, realizar una política de control lo más óptima posible.

En el ejemplo de la Ilustración 14, la opción T(1) es más cara desde el punto de vista del coste de control que la opción T(2), ya que la frecuencia de muestreo es mayor, pero tal vez los beneficios obtenidos al detectar antes un cambio en el proceso sean mayores en T(1).

Cambio en media del proceso

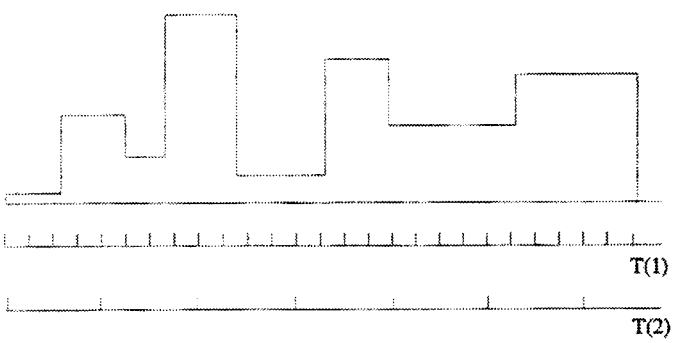


ILUSTRACIÓN 14: Frecuencia de muestreo T(1) y T(2) para control de la medida

Por último, de poco o nada sirve lo hasta aquí expuesto, si la información que emana de los gráficos de control tarda en ser «escuchada» por aquellas personas que directamente pueden actuar sobre el proceso, o no se le presta la debida atención. Es decir, poner en marcha un sistema de control estadístico de procesos **no consiste en**:

- Rellenar plantillas de recogida de datos por parte de los operarios, para ser llevados a final de mes a un gráfico por parte de los técnicos y posteriormente ser archivados
- Realizar un gráfico con los límites obtenidos la primera vez que se implementó el CEP, los cuales se observan que poco tienen que ver con la situación actual del proceso.

El CEP es una técnica que ha de ser utilizada cerca del proceso y, por lo tanto, es recomendable que sean los operarios los que la utilicen. Para ello, los gráficos de control han de ser **sencillos** de construir e interpretar.

En cuanto a la actualización de los límites de control en los gráficos, hemos de decir que los límites no se han de cambiar mientras no se tenga evidencia de que el sistema de causas comunes haya cambiado o de que cambia el procedimiento de toma de datos.

3.1.2. GRÁFICOS DE CONTROL PARA VARIABLES

Se denominan así los gráficos de control para características continuas del producto o del proceso tales como: contenido en cm^3 de un líquido, peso de un saco de pienso, viscosidad de una resina, intensidad de una tinta, temperatura de un horno, etc., las cuales, cuando el proceso está en estado de control, se distribuyen en general según la ley normal, es decir, si no se distribuyen las observaciones individuales, si lo hacen las medias para tamaño suficientemente grande (teorema central del límite).

Un proceso del cual se está controlando una característica continua puede abandonar ese estado de control por verse afectada su media, su variabilidad o ambas a la vez. Por consiguiente, hay que construir gráficos para controlar ambos parámetros por separado, no sólo la media.

El control de la media del proceso se realiza, siempre que se pueda, a través de las medias de muestras de tamaño n.

Si de las mismas muestras se calcula la desviación típica muestral s, se puede construir un gráfico de control para σ utilizando la distribución teórica de s:

$$s^2 \approx \frac{\sigma^2}{(n-1)} X_{n-1}^2$$

Así, tomando los percentiles adecuados de la chi-Cuadrado se tendría un gráfico para s^2 (notar que este gráfico no es simétrico). En la práctica, el cálculo de la varianza muestral resulta difícil para personal poco familiarizado con la estadística y

se recurre a calcular el rango de la variación muestral R, que se define como la diferencia entre las observaciones extremas. Se puede demostrar matemáticamente que existe una relación entre s^2 y el rango muestral para muestras de tamaño n y tal relación es:

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

donde d_2 se puede encontrar en la tabla de índices (apartado 3.1.8).

A continuación se presentan los gráficos más utilizados para características continuas.

3.1.3 GRÁFICOS $\bar{X}$-R

El CEP utilizando gráficos $\bar{X}$-R se lleva a cabo tomando muestras de n individuos (entre dos y seis), calculando la media y el recorrido muestral y llevando estos valores a los gráficos correspondientes.

Las muestras han de ser obtenidas de tal forma que contengan individuos **homogéneos**, es decir, producidos bajo las mismas condiciones, así, los estadísticos que se obtengan de ellos, la media y el recorrido, serán buenos estimadores de los parámetros del proceso. Debe procurarse que, durante el tiempo que el proceso fabrica los individuos que forman parte de una muestra, sólo hayan actuado causas comunes de variabilidad.

Los límites en los gráficos se colocan en

$$\textit{Gráfico } \bar{X} \begin{cases} \textit{Límite superior: } \mu + 3\dfrac{\sigma}{\sqrt{n}} & o & \bar{\bar{x}} + A_2\bar{R} \\ \textit{Límite central: } \mu & o & \bar{\bar{x}} \\ \textit{Límite inferior: } \mu - 3\dfrac{\sigma}{\sqrt{n}} & o & \bar{\bar{x}} - A_2\bar{R} \end{cases}$$

$$\textit{Gráfico } R \begin{cases} \textit{Límite superior: } D_4\bar{R} \\ \textit{Límite central: } \bar{R} \\ \textit{Límite inferior: } D_3\bar{R} \end{cases}$$

donde los valores de A_2, D_3 y D_4 se pueden encontrar en la tabla de índices (apartado 3.1.8).

En aquellos casos en que no se tenga una estimación de los parámetros del proceso en el momento de implementar los gráficos de control, o se desee recalcularlos, se han de seguir los siguientes pasos:

Tomar k (al menos 20) muestras de tamaño n (entre dos y seis) de forma consecutiva y a intervalos de tiempo iguales, calculando la media y el rango de cada muestra:

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \ldots + x_{in}}{n} \quad R_i = máx(x_{ij}) \; mín(x_{ij}) \; j = 1,2,3\ldots,n$$

Calcular la media de las k medias muestrales y la media de los k rangos:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k} \bar{x}_i}{k} \quad \bar{R} = \frac{\sum_{i=1}^{k} R_i}{k}$$

Calcular los límites de control del gráfico mediante las expresiones anteriores.

Llevar los valores de las medias y los rangos de las muestras obtenidas a los gráficos y comprobar que no hay ningún tipo de comportamiento anómalo en ninguno de ellos. En tal caso, pasar al apartado siguiente.

Si existe evidencia de que durante la construcción del gráfico el proceso no ha estado bajo control, se han de buscar las causas asignables y actuar sobre ellas. Sólo en este caso se reconstruirá el gráfico eliminando las anomalías y comenzando en el paso segundo. En aquellos casos en que hayan variado notablemente las características del proceso, debe comenzarse desde el principio.

Mantener los límites de control calculados en el apartado tercero y establecer un plan de control para el futuro con el objetivo de realizar un seguimiento del proceso.

Para ello, dependiendo de las características del proceso (coste de inspección, producción diaria, coste de producir fuera de especificaciones) se toman muestras de tamaño n en intervalos de tiempo determinado y se lleva la media muestral, $\bar{x}_i$, y el recorrido, R_i, a los gráficos correspondientes.

Una llamada de atención en uno de los gráficos, equivale a que una causa asignable ha entrado en el proceso. En este caso, se ha de buscar la causa asignable y deben tomarse las acciones adecuadas para llevar al proceso a su estado de control.

Cuando la construcción de los gráficos se hace manualmente, existen plantillas que han sido adoptadas con pequeñas variantes en las empresas. En ellas, además del espacio reservado para los gráficos, existen casillas donde se debe anotar toda la información que pueda ayudar a una posterior interpretación del gráfico.

Vamos a ver un caso particular del uso del gráfico $\bar{X}$. En él, los límites se han calculado a partir de las primeras 20 observaciones que se encuentran en la siguiente tabla, de la cual calculamos los límites:

Límite superior $= 200 + 0.729 * 1.48 = 201.1$
Límite central $= 200$
Límite inferior $= 200 - 0.729 * 1.48 = 198.9$

Realizando el gráfico del contenido medio de 4 botellas con $\mu = 200$ y $\sigma = 0.7$ (Observaciones 1-20), con $\mu = 201$ y $\sigma = 0.7$ (Observaciones 21-40) y con $\mu = 200$ y $\sigma = 0.7$ (ultimas 20 observaciones), se detecta cambios en el proceso debidos a cambios en la media y también a un incremento de la variabilidad.

TABLA 2: Contenido en cm^3 de 20 muestras de tamaño 4 de botellas de agua

	X_1	X_2	X_3	X_4	$\overline{X}$	R_1
1	202.016	200.218	200.588	199.93	200.688	2.085
2	200.39	200.18	198.385	199.12	199.52	2.005
3	200.08	199.42	200.045	199.985	199.884	0.665
4	198.584	201.011	200.26	200.097	199.988	2.427
5	199.413	199.453	200.012	200.72	199.9	1.306
6	199.701	198.761	200.001	200.118	199.645	1.357
7	199.38	200.491	200.361	200.057	200.072	1.111
8	201.435	200.279	199.727	200.513	200.488	1.707
9	199.44	199.155	199.966	200.129	199.672	0.974
10	200.857	201.021	199.526	200.654	200.515	1.496
11	200.445	199.933	200.03	199.044	199.863	1.401
12	200.109	200.9	200.116	201.751	200.719	1.642
13	199.985	200.006	200.659	200.6	200.312	0.673
14	200.58	199.934	199.789	199.699	200.001	0.881
15	199.796	199.759	199.88	200.34	199.944	0.581
16	199.277	200.722	198.398	200.41	199.702	2.324
17	200.612	198.605	199.194	199.998	199.602	2.007
18	199.899	201.027	199.998	200.806	200.432	1.129
19	200.31	199.998	200.571	199.25	200.032	1.322
20	198.692	198.65	200.686	201.191	199.805	2.541
					$\overline{\overline{X}}=200$	$\overline{R}=1.48$

Para el gráfico de la variabilidad, se llevan los rangos de la tabla a un gráfico, donde los límites de control se han calculado a partir del rango medio y los valores de $D_3 = 0$ y $D_4 = 2.282$ correspondientes a una muestra de tamaño 4.

Gráfico R para la variabilidad del proceso de embotellado. Con $\mu = 200$ y $\sigma = 0.7$ (observación 1-20), con $\mu = 201$ y $\sigma = 0.7$ (obs. 21-40) y con $\mu = 200$ y $\sigma = 1$ (ultimas 20 observaciones)

Los límites así calculados son los siguientes:

Límite superior = 2.282 *1.48 = 3.38
Límite central = 1.48
Límite inferior = 0 *1.48 = 0

Observando el gráfico que se obtendría para R, el aumento de la media del proceso a partir de la observación 20 provoca un ligero incremento en los rangos, aunque este aumento es más manifiesto cuando aumenta la variabilidad del proceso a partir de la observación 40, llegando incluso a salir los rangos fuera de los límites.

3.1.4. GRÁFICOS $\bar{X}$-S

En este gráfico de control se va a representar la media $\bar{X}_i$ de cada submuestra midiendo la variabilidad a través de las desviaciones típicas muestrales, por lo que en primer lugar hay que calcular la media y desviación típica de cada muestra (submuestra):

$$\bar{X}_i = \frac{\sum_{j=1}^{n} X_{ij}}{n} \qquad s = \sqrt{(1/n) * \sum_{j=1}^{n} (X_{ij} - \bar{X}_i)^2}$$

para i = 1,2,...,k.

Supuesto un estado de control, es sabido que la esperanza matemática de $\bar{X}_i$ es μ y su desviación típica $\sigma/\sqrt{n}$:

$$E[\bar{X}_i] = \mu \qquad \sigma[\bar{X}_i] = \sigma/\sqrt{n}$$

por ello el gráfico de control para las medias muestrales debería tener una línea central LC = μ y unos límites de control iguales a $\mu \pm 3\sigma/\sqrt{n}$; ahora bien, como la media μ y la desviación típica σ son desconocidas se estiman con los datos. La media por:

$$\mu = \bar{\bar{X}} = (1/n) * \sum_{i=1}^{k} \bar{X}_i = (1/N) * \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}$$

que es un estimador centrado; en cambio las desviaciones típicas muestrales S_i no son estimadores centrados de la desviación típica teórica σ, sino que tienen un sesgo que depende del tamaño n de la muestra. Puede probarse que $E[S_i] = c_2\sigma$, donde c_2 es un coeficiente que depende sólo de n y cuyos valores se encuentran tabulados; la expresión analítica exacta es la siguiente:

$$E[S_i]=c_2\sigma=\sqrt{2/n}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\sigma$$

donde $\Gamma(z)$ es la función Gamma de Euler. Como para cada muestra i = 1,2,...,k resulta que S_i/c_2 es un estimador centrado de σ, al promediarlos obtenemos el siguiente estimador centrado:

$$\sigma = (1/c_2)^* \overline{S} = (1/kc_2)^* \sum_{i=1}^{k} S_i$$

quedando por tanto las siguientes líneas de control para las medias $\overline{X}_i$:

$$\text{Límite Superior de control} = LCS = \overline{\overline{X}} + A_i\,\overline{S}$$
$$\text{Límite Central} = LC = \overline{\overline{X}}$$
$$\text{Límite Inferior de control} = LC = \overline{\overline{X}} - A_i\,\overline{S}$$

donde A_i es un coeficiente que depende del tamaño n de la muestra y que se halla tabulado; su expresión analítica es:

$$A = 3/(c_2\sqrt{n})$$

Gráfico S de Desviaciones Típicas

En este gráfico de control se va a representar la desviación típica S de cada muestra y va a ser útil para controlar la variabilidad del proceso. Como hemos visto antes se tiene que $E[S_i] = c_2\sigma$, podemos estimar de modo centrado por $\overline{S}$. Por otra parte, puede probarse que la desviación típica de S_i vale:

$$\sigma(S) = \sigma\sqrt{1-(1/n)-c_2^2}$$

lo que nos conduciría a los siguientes límites teóricos de control:

$$c_2\,\sigma \pm 3\sigma\sqrt{1-(1/n)-c_2^2}\ (c \pm 3\sqrt{1-(1/n)-c_2^2})\sigma$$

y estimando σ por $\overline{S}\,c_2$ obtenemos finalmente las siguientes líneas de control para las desviaciones típicas:

$$\begin{aligned}&\text{Límite Inferior de Control} = \text{LIC} = B_3\ \overline{S}\\ &\text{Límite Central} = \text{LC} = \overline{S}\\ &\text{Límite Superior de Control} = \text{LSC} = B_4\ \overline{S}\end{aligned}$$

donde B_3 y B_4 están tabuladas, siendo sus expresiones analíticas:

$$B_3 = 1 - (3/c_2)^*\sqrt{1-(1/n)-c_2^2}\ \text{o cero si es negativo}$$

$$B_4 = 1 + (3/c_2)^*\sqrt{1-(1/n)-c_2^2}$$

3.1.5. Gráfico *X*-σ

En este apartado vamos a suponer que el gráfico de control se realiza sobre una característica medible X cuyo valor medio es μ y cuya desviación típica es σ. Si las observaciones de X son independientes y normales entonces el 95.44 % de ellas se encontrarán dentro de los límites $\mu \pm 2\sigma$, y el 99.73 % dentro de los límites $\mu \pm 3\sigma$.

Esto nos permite trazar la línea central, los límites de control y los de alerta y sobre este gráfico, a medida que transcurra el tiempo, ir representando los valores de la magnitud X para ir observando su evolución. Es frecuente que los valores de μ y σ sean desconocidos y se estimen con datos anteriores; si $X_1, X_2,\ldots,X_n$ es una muestra aleatoria simple de observaciones de X, entonces la media μ se estima por la media de la muestra:

$$\mu = \overline{X} = (1/n)^* = \sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

y la varianza σ^2 con la cuasivarianza de la muestra:

$$\sigma^2 = S_c^2 = (1/(n\text{-}1))^* \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Así como la cuasivarianza muestral S_c^2 es un estimador ligeramente sesgado de la desviación típica teórica σ, pero el tamaño de la muestra es de 20 o más el sesgo puede considerarse despreciable (por ejemplo, para n = 20 el sesgo es de un 1.3 % el valor de S_c, y tiende a disminuir a medida que aumenta el tamaño de la muestra). Si designamos con S la desviación típica muestral:

$$S = \sqrt{(1/n)^* \sum_{i=1}^{n} (X_{ij} - \overline{X})^2} = \sqrt{\frac{n-1}{n}} S_c$$

entonces podemos construir un estimador centrado de σ dividiendo S por cierto coeficiente c_2 cuyo valor sólo depende del tamaño de la muestra:

$$\sigma = S/c_2$$

Los valores de c_2 pueden tabularse para distintos valores de tamaño n de la muestra, siendo su expresión analítica la siguiente:

$$c_2 = \sqrt{2/n} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}$$

En la práctica lo que se hace es observar 20 o más datos retrospectivos de X, que se encuentren en una situación estacionaria, y a partir de esas medidas se estima la media y desviación típica de $\overline{X}$ y S/c_2 respectivamente, con cuyos valores se dibujan las líneas de control y central:

Límite Superior de control = LCS = $\overline{\overline{X}} + 3S/c_2$

Límite Central = LC = $\overline{\overline{X}}$

Límite Inferior de control = LC = $\overline{\overline{X}} - 3S/c_2$

Con objeto de garantizar el carácter estacionario de los datos observados $X_1, X_2, \ldots, X_n$ se les aplica el gráfico de control que se acaba de calcular, y se desechan aquéllos que queden fuera de los límites de control (por considerarlos anómalos), que forman una racha (pues durante ese período es presumible que la media teórica μ haya tenido una fluctuación) o que presenten una tendencia positiva o negativa (ya que en esos casos la media teórica μ está variando en el tiempo y no permanece fija). Con la muestra resultante se vuelve a recalcular $\overline{X}$ y S, así como el nuevo gráfico de control. Una vez que la muestra retrospectiva de trabajo se encuentre dentro de los lí-

mites de control calculados, que no presente rachas y que tampoco muestra tendencias, se utilizan los límites calculados con ella para representar y controlar observaciones futuras.

Diagrama de Control de la Varianza Muestral (S^2)

Lo más frecuente en control de calidad es usar el diagrama de R, o el de S, para controlar la dispersión del proceso, con preferencia por el de S para tamaños muestrales moderadamente grandes.

Pero a veces es recomendable utilizar un diagrama basado directamente en la varianza muestral S^2, los parámetros para el diagrama de control de S^2 se derivan de que, dada una muestra de tamaño n, el estadístico $(n-1)\ S^2/\sigma^2$ tiene una distribución chi-cuadrado con n-1 grados de libertad, en cuyo caso se pueden determinar a_α y b_α tales que:

$P(a_\alpha \leq (n-1)\ S^2/\sigma^2 \leq b_\alpha) = 1 - \alpha$ para un nivel α dado, donde:

$a_\alpha = F^{-1}(\alpha/2)$ y $b_\alpha = F^{-1}(1 - \alpha/2)$

siendo F la función de distribución de una variable Chi-cuadrado con n-1 grados de libertad.

Supongamos que tenemos una característica de calidad X, que sigue una distribución Normal, con media μ y desviación estándar σ, desconocidas.

Se supone que tenemos *k* muestras $(x_{11},.....,x_{1n}),....(x_{k1},.....,x_{kn})$, cada una con n observaciones de la característica de calidad X.

Sean $\overline{x_1}, \overline{x_2},....,\overline{x_k}$ las k medias de las muestras y $S_1, S_2,...,S_k$ las k desviaciones típicas de las muestras.

Para cada una de las muestras se pueden determinar a_α y b_α tales que:

$P(a_\alpha \leq (n-1)\ S_i^2/\sigma^2 \leq b_\alpha) = 1 - \alpha$, con i = 1,.....,k.

Tenemos entonces:

$$P\left(\frac{a_\alpha \cdot \sigma^2}{n-1} \leq S_i^2 \leq \frac{b_\alpha \cdot \sigma^2}{n-1}\right) = 1-\alpha$$

Tomando $\overline{S}^2$ como una varianza muestral media de las S_i^2, es decir:

$$\overline{S}^2 = \frac{1}{k}\sum_{i=1}^{k} S_i^2$$

podemos escribir:

$$P\left(\frac{a_\alpha \cdot \sigma^2}{n-1} \leq \bar{S}^2 \leq \frac{b_\alpha \cdot \sigma^2}{n-1}\right)$$

Además, como es un estimador insesgado de σ^2 para i = 1,...,k, tenemos que $\bar{S}^2$ también es un estimador insesgado de σ^2, con lo que ya tenemos los límites probabilísticos al nivel a de la forma siguiente:

Límite Superior de control probabilistico = $LSCP = \frac{b_\alpha \cdot \bar{S}^2}{n-1}$

Línea Central = $\bar{S}^2$

Límite Inferior de control probabilistico = $LICP = \frac{a_\alpha \cdot \bar{S}^2}{n-1}$

El gráfico resultante quedaría de la siguiente forma (Ilustración 15):

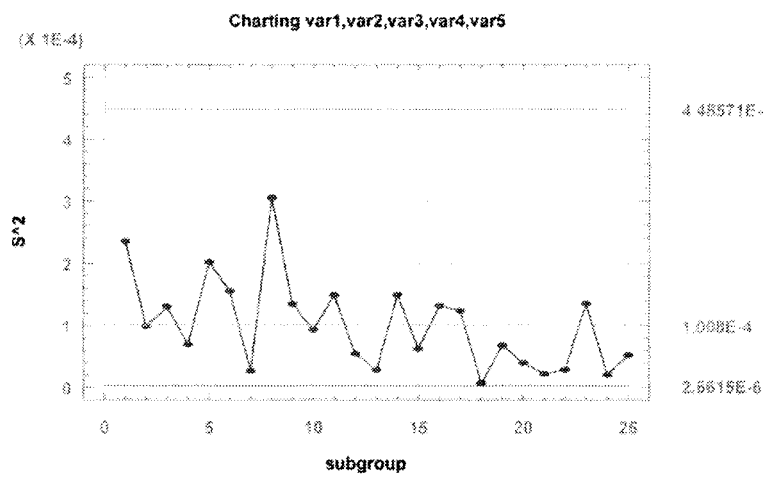


ILUSTRACIÓN 15: Gráfico de control de varianzas

3.1.6. GRÁFICO DE OBSERVACIONES INDIVIDUALES Y GRÁFICO DE RANGOS MÓVILES

Estos gráficos son similares a los gráficos $\bar{X}$-R con la diferencia de que los primeros se utilizan en aquellos casos en que se obtiene una única observación en cada instante. Algunas situaciones en las que esto ocurre son:

- Sólo puede obtenerse una observación por lote o partida de material
- En procesos continuos o de batch en los cuales no tiene sentido hablar de «individuos»
- Se requiere realizar una comparación directa con las especificaciones.

Para la implementación de los gráficos de observaciones individuales y de rangos móviles se han de seguir los pasos presentados para los gráficos $\bar{X}$-R, teniendo en cuenta que, al ser n = 1, se tendrán que realizar algunas modificaciones.

Así, con las k observaciones obtenidas, la media del proceso se etima según,

$$\bar{x}=\mu=\frac{\sum_{i=1}^{k} x_i}{k}$$

x_i observación i-ésima

k total de observaciones

El rango medio se obtiene promediando los rangos móviles obtenidos al hacer muestras de tamaño w de la siguiente manera: Para obtener R_1 se toman las primeras w observaciones $(x_1, x_2,\ldots,x_n)$ y se calcula el rango. R_2 se obtiene a partir de $(x_2, x_3, \ldots,x_w, x_{w+1})$, y así sucesivamente. Así, se obtiene la media de rangos,

$$\bar{R}=\frac{\sum_{i=1}^{k-w+1} R_i}{k-w-1}$$

donde:

k: Total observaciones.

w: Nº de observaciones utilizadas en el cálculo del rango móvil.

R_i: Rango del grupo móvil $(x_i, \ldots,x_{i+w-1})$.

$\bar{R}$: Media de rangos móviles.

De esta manera el control estadístico de la media del proceso se realiza llevando las observaciones individuales a un gráfico que tiene de límites,

Límite superior $\bar{x}+3\frac{\bar{R}}{d_2}$

Límite central $\bar{x}$

Límite inferior $\bar{x} - 3\dfrac{\bar{R}}{d_2}$

Para el control estadístico de la variabilidad se utilizan los rangos móviles calculados anteriormente. En ambos gráficos para el cálculo de los valores d_2, D_3 y D_4 se considera n = w.

Hay que notar que w ha de ser seleccionado de tal manera que los elementos en un mismo grupo sean lo más homogéneos posibles. Un valor de w muy utilizado es w = 2, y en tal caso d_2 es aproximadamente 9/8.

El gráfico de observaciones individuales es menos sensible que el gráfico $\bar{X}$. Además, si la distribución de los datos no sigue una ley normal, hay que tener mucho cuidado en la interpretación del gráfico.

Téngase en cuenta por otra parte, que el hecho de que los valores R_i no sean totalmente independientes también dificulta la interpretación del gráfico de la variabilidad.

3.1.7. OTROS TIPOS DE GRÁFICOS DE CONTROL POR VARIABLES

Los tipos de gráficos que a continuación describimos también llamados Gráficos de Control con Memoria, se caracterizan por ir acumulando la información de muestras anteriores, de manera que un cambio pequeño en los datos es detectado rápidamente, por el contrario cambios grandes son detectados más lentamente.

3.1.7.1. DIAGRAMA DE CONTROL DE SUMA ACUMULATIVA (CUSUM)

El diagrama de control de suma acumulativa o diagrama CUSUM (Cumulative Sum), incorpora directamente al representar el estadístico relativo a cada punto (muestra o subgrupo), toda la información muestral de la sucesión de puntos anteriores, ya que se representan para cada punto las sumas acumuladas para el subgrupo actual y todos los anteriores de las desviaciones del estadístico en que se basa el gráfico de control.

Supongamos que tenemos muestras de tamaño $n \geq 1$, y que $\bar{x}_i$ es el promedio de la i-ésima muestra. Entonces, si μ_o es el valor objetivo de la media del proceso, el gráfico de control de suma acumulativa, que utiliza como estadístico para cada subgrupo la media muestral del mismo, se formará representando las cantidades:

$$S_m = \sum_{i=1}^{m} \left(\bar{x}_i - \mu_o\right) \qquad \text{para } m = 1,2,\ldots,k$$

El grafo del diagrama de control lo constituyen los puntos (m, S_m) para m = 1,2,...,*k*.

El S_m para cada una de las muestras de entre las *k* posibles viene dado en función del número *m* de la muestra o subgrupo e incluye a todas las S_m anteriores ya que:

$$S_m = \sum_{i=1}^{m} (\bar{x}_i - \mu_o) = \bar{x}_m - \mu_o + S_{m-1} = \bar{x}_m - \mu_o + \bar{x}_{m-1} - \mu_o + S_{m-2} = ...$$

S_m recibe el nombre de suma acumulativa hasta la m-ésima muestra o subgrupo inclusive.

En este tipo de gráficos, si el proceso permanece bajo control para el valor objetivo de la media (μ_o) la suma acumulativa S_m tendrá que variar aleatoriamente alrededor de cero.

Por el contrario si la media cambia aumentando hasta cierto valor $\mu_1 > \mu_o$, entonces se observará un desplazamiento en esa dirección (positiva) de las sumas acumuladas. Recíprocamente, si la media cambia disminuyendo hasta cierto valor, entonces se observará un desplazamiento en esa dirección (negativa).

En consecuencia si se desarrolla una tendencia hacia arriba o hacia abajo en los puntos ubicados en el gráfico, se tendrá que considerar esto una evidencia de cambio de la media del proceso y se deberá buscar alguna causa atribuible.

Los gráficos obtenidos serán de la forma que se muestra en Ilustración 16:

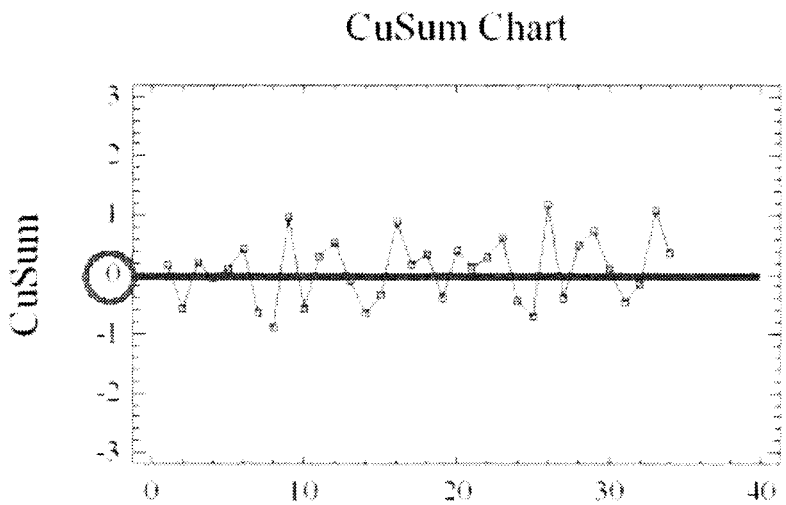


ILUSTRACIÓN 16: Gráfico de CuSum

Mascarilla V (V-Mask)

Como en todo gráfico de control es necesario fijar un procedimiento de decisión para determinar si el proceso está fuera de control o no. Este procedimiento de decisión consiste en aplicar la Mascarilla V al diagrama de suma acumulativa.

Si para cada punto *m* todas las sumas acumuladas anteriores S_1, S_2,..., S_m caen entre los dos brazos de la Mascarilla V, el proceso se encuentra bajo control. Sin embargo, si cualquier S_i $\forall$ i = 1,...,*m* se sitúa fuera de los brazos de la mascarilla, el proceso se considera fuera de control. Un punto por encima del brazo superior de la mascarilla indicaría un cambio hacia abajo en la media, y viceversa si el punto se sitúa por debajo del brazo inferior de la mascarilla.

De esta forma la Mascarilla V constituye un marco de referencia visual para decidir sobre el control, similar a los límites de control de un diagrama de Shewhart convencional.

Para construir la Mascarilla V se considera el diagrama de control con el número de muestra *m* en el eje de abscisas y los valores de las sumas acumuladas S_m para cada muestra en el eje de ordenadas.

Se sitúa el punto O en el último valor de S_m, es decir en S_i, de modo que las coordenadas de O son (i, S_i) y se traza la línea OA paralela al eje horizontal tal y como se indica en la Ilustración siguiente:

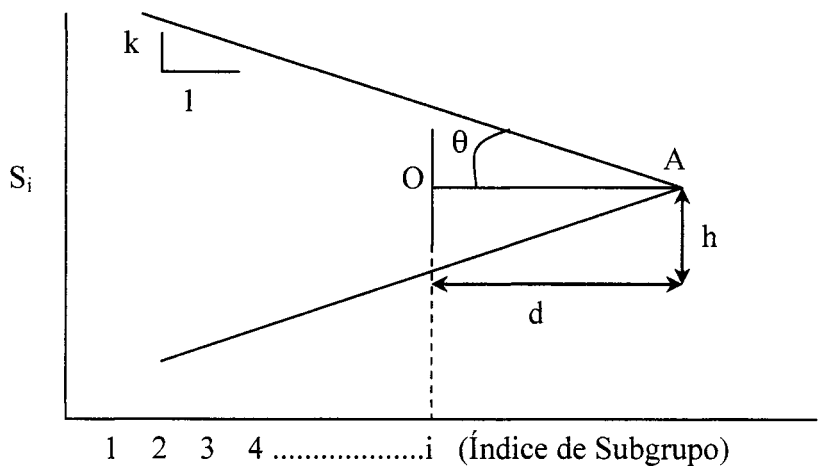


ILUSTRACIÓN 17: Construcción Mascarilla V

La línea OA se traza de forma que su longitud sea la distancia guía **d**. Una vez fijado el segmento OA se utiliza el ángulo **θ** para definir los brazos de la mascarilla.

Por tanto d y θ determinan la mascarilla y se calculan con las siguientes expresiones:

$$d = \left(\frac{2}{\delta^2}\right) \cdot \ln\left(\frac{1-\beta}{\alpha}\right) \quad y \quad \theta = \arctan\left(\frac{\Delta}{2 \cdot r}\right)$$

donde:

- Δ es el cambio en la media del proceso que se quiere detectar.
- α es la probabilidad de concluir incorrectamente que se produjo un cambio (falsa alarma).
- β es la probabilidad de no detectar un cambio en la media.
- δ toma el valor $\delta = \frac{\Delta}{\sigma_{\bar{x}}}$ donde $\sigma_{\bar{x}}$ es la desviación estándar de $\bar{x}$ y k que es un factor de escala que relaciona la unidad de la escala vertical con la unidad de escala horizontal.

Normalmente se recomienda que k esté entre $\sigma_{\bar{x}}$ y $2 \cdot \sigma_{\bar{x}}$, con un valor preferente de $k = 2 \cdot \sigma_{\bar{x}} \cdot \delta$ que represente entonces la magnitud del cambio en la media del proceso que se quiere detectar en unidades de desviación estándar de $\sigma_{\bar{x}}$.

Si β es pequeño ($\beta \rightarrow 0$), entonces el valor de d se reduce y $d \rightarrow -2\ln(\alpha/\delta^2)$.

Como $h = d \cdot \tan(\theta) = d \cdot k$, se puede determinar la mascarilla en función de k y h.

Un ejemplo de gráfico de control con Mascarilla V sería el que se muestra en la siguiente Ilustración:

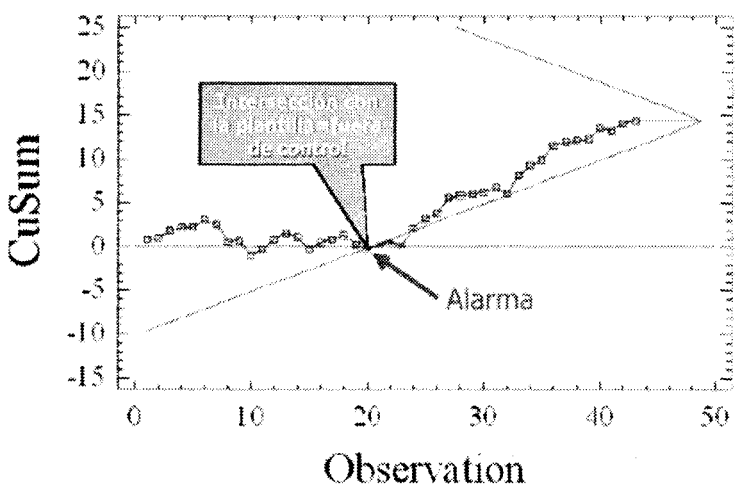


ILUSTRACIÓN 18: Gráfico de control con Mascarilla

3.1.7.2. Diagrama de control de Media Móvil

El gráfico de medias móviles, o gráfico MA (*Moving Average*), es un gráfico para controlar la media del proceso y se emplea en general en aquellos casos en los que, obteniéndose observaciones individuales del proceso, se desea analizar el mismo con la sensibilidad que permite un gráfico de medias. Para ello, las medias móviles se obtienen de igual manera que se han obtenido los rangos móviles en el apartado anterior.

Estos gráficos «suavizan» el comportamiento observado en el de observaciones individuales y muestran mejor la tendencia del proceso. Por otra parte, y tal como ocurría en el gráfico de rangos móviles, las observaciones no son independientes, lo cual dificulta la interpretación.

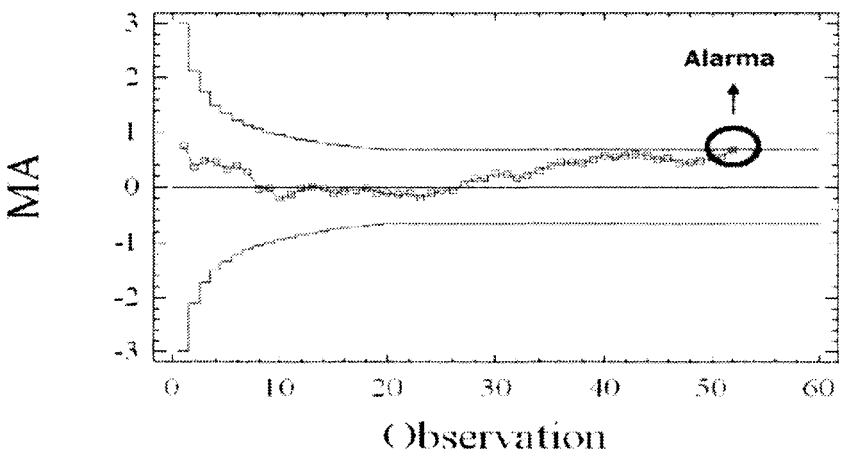


ILUSTRACIÓN 19: Medias Móviles

3.1.7.3. Diagrama de Media Móvil Geométrica (EWMA)

El diagrama de control de media móvil geométrica, también denominado Diagrama de Media Móvil Ponderada Exponencialmente, EWMA (Exponentially Weighted Moving Average), puede considerarse un procedimiento de control de la media ponderada en el que cada media muestral $\bar{x}_i$, $\bar{x}_{i-1}$,..., $\bar{x}_{i-w+1}$ es ponderada por 1/w en el instante i, y las medias muestrales para periodos $t \leq i - w$ son ponderados por cero.

Un gráfico de control basado en la media móvil geométrica de orden r, se define como:

$$Z_t = r\ \bar{x}_t + (1\ \ r)\ Z_{t-1}$$

siendo $i > 1$ y donde r es una constante tal que $0 < r \leq 1$, cuyo valor inicial (necesario para la primera muestra en $i = 1$) es $Z_o = \overline{\overline{X}}$.

Z_i es una media ponderada de todas las medias muestrales anteriores, ya que si sustituimos Z_{i-1} en la parte derecha de Z_i por su valor iterativo, tendremos:

$$Z_i = r \cdot \bar{x}_i + (1-r) \cdot Z_{i-1} = r \cdot \bar{x}_i + (1-r) \cdot \left[r \cdot \bar{x}_{i-1} + (1-r) \cdot Z_{i-2} \right] =$$
$$r \cdot \bar{x}_i + r \cdot (1-r) \cdot \bar{x}_{i-1} + (1-r)^2 \cdot Z_{i-2}$$

Continuando con la sustitución recursiva de Z_{i-j}, $j = 2,3,....,i$, se obtiene:

$$Z_i = r \sum_{j=0}^{i-1} (1-r)^j \cdot \bar{x}_{i-j} + (1-r)^i \cdot Z_o$$

Los pesos $r(1 - r)^j$ decrecen geométricamente con el orden de la media muestral, además la suma de los pesos será igual a 1.

Si las $\bar{x}_i$ son variables aleatorias independientes con varianza σ^2/n entonces la varianza de Z_i es:

$$V(Z_i) = V\left(r \sum_{j=0}^{i-1} (1-r)^j \cdot \bar{x}_{i-j} + (1-r)^i \cdot Z_o \right) = \frac{\sigma^2}{n} \cdot \left(\frac{r}{2-r} \right) \cdot \left[1 - (1-r)^{2i} \right]$$

Se observa que cuando i aumenta, $V(Z_i)$ aumenta hasta alcanzar un valor límite, que suele denominarse valor asintótico y que vale:

$$V(Z_i) = \frac{\sigma^2}{n} \cdot \left(\frac{r}{2-r} \right)$$

Para la media de Zi tenemos que:

$$E(Z_i) = E\left(r \sum_{j=0}^{i-1} (1-r)^j \cdot \bar{x}_{i-j} + (1-r)^i \cdot Z_o \right) = \mu \cdot \left[1 - (1-r)^i \right] + (1-r)^i \cdot \mu = \mu$$

Para construir el diagrama de control de la media móvil hay que estimar μ y σ, para lo que se toman k muestras consecutivas de $n_i = n$ unidades se calcula en cada una la media muestral $\bar{x}_i$, y se considera el estadístico $\overline{\overline{X}}$ como la media de las $\bar{x}_i$, que será el estimador de μ.

Para calcular los límites de control de Shewhart sustituimos μ y σ por sus estimadores $\overline{\overline{X}}$ y $\hat{\sigma}$, obteniendo los límites de control para el caso de muestras del mismo tamaño n, los límites asintóticos son:

$$\text{Límite Superior de control prob.} = \text{LSCP} = \overline{\overline{X}} + k_\alpha \cdot \hat{\sigma} \cdot \sqrt{\frac{r}{n \cdot (2-r)}}$$

$$\text{Límite Central} = \overline{\overline{X}}$$

$$\text{Límite Inferior de control prob.} = \text{LICP} = \overline{\overline{X}} - k_\alpha \cdot \hat{\sigma} \cdot \sqrt{\frac{r}{n \cdot (2-r)}}$$

Para realizar el gráfico de control, en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas situamos las medias móviles geométricas Z_i para cada muestra. Mientras la media móvil para cada muestra quede entre los límites de conrol y la sucesión de puntos definida por cada media no exhiba un patrón sistemático, podemos considerar el proceso estudiado bajo control.

Un ejemplo de diagrama de Media Móvil Geométrica seria el mostrado en la Ilustración 20.

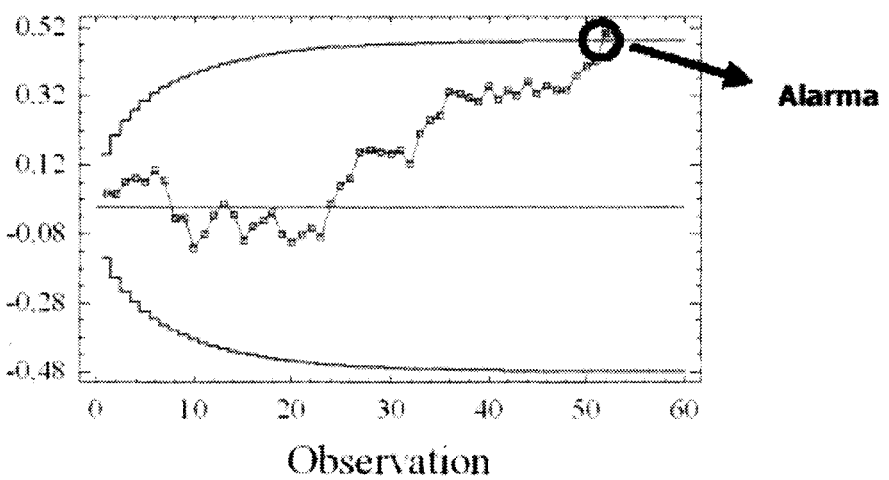


ILUSTRACIÓN 20: Medias Móviles Geométricas

3.1.8. FACTORES PARA LA CONSTRUCCIÓN DE GRAFICO DE CONTROL

A continuación, podemos ver la tabla de los distintos valores de los factores para la construcción de gráficos de control.

TABLA 3: Factores para gráficos de control

N	A_2	D_2	D_3	D_4
2	1.880	1.118	0	3.267
3	1.023	1.693	0	2.575
4	0.729	2.059	0	2.282
5	0.577	2.326	0	2.115
6	0.483	2.534	0	2.004
7	0.419	2.704	0.076	1.924
8	0.373	2.847	0.136	1.864
9	0.337	2.970	0.184	1.816
10	0.308	3.078	0.223	1.777
11	0.285	3.173	0.256	1.744
12	0.266	3.258	0.284	1.716
13	0.249	3.336	0.308	1.692
14	0.235	3.407	0.329	1.671
15	0.223	3.472	0.348	1.652
16	0.212	3.532	0.364	1.636
17	0.203	3.588	0.379	1.621
18	0.194	3.640	0.392	1.608
19	0.187	3.689	0.404	1.596
20	0.180	3.735	0.414	1.586
21	0.173	3.778	0.425	1.575
22	0.167	3.819	0.434	1.566
23	0.162	3.858	0.443	1.557
24	0.157	3.895	0.452	1.548
25	0.153	3.931	0.459	1.541

3.1.9. INTERPRETACIÓN DE LOS GRÁFICOS DE CONTROL

El objetivo de la utilización de los gráficos de control para el seguimiento de un proceso es primordialmente el de detectar cualquier evidencia de que la media y la variabilidad del proceso no se han mantenido constantes a lo largo del tiempo. Es decir, se pretende detectar la aparición de causas asignables a la variabilidad.

Con tal objetivo en el gráfico se han representado dos cotas o límites de variabilidad las cuales evidencian la presencia de tales causas si son sobrepasadas. Este patrón de inestabilidad fue el que se mantuvo durante los primeros años de la implantación de los gráficos Shewart.

Los gráficos así construidos tenían varios inconvenientes:

- Permanecen impasibles ante aquellas causas asignables que afectaban al proceso sin llegar a provocar individuos fuera de límites.
- Detectaban algunas anomalías demasiado tarde.
- No tenían en cuenta la información histórica del proceso.

Para contrarrestar los puntos anteriormente citados, se incorporaron nuevos patrones de inestabilidad. Estos últimos tienen la particularidad de ser tan poco probables de ser presenciados en un proceso bajo control como el hecho de obtener una observación fuera de límites, además, tienen en cuenta el comportamiento histórico del proceso a corto plazo.

Para la detección de tales patrones, se han de dividir las dos áreas alrededor del límite central en tres zonas: A, B y C.

Nótese además, que cuanto más se muestrea más posibilidades existen de obtener falsas alarmas y tomar por lo tanto acciones que, en lugar de disminuir la variabilidad del proceso, la aumentan. Por lo tanto, no se ha de actuar si no se está seguro de la presencia de causas asignables y se conoce su identidad.

Generalmente, cuesta identificar una causa asignable entre todas las posibles. En la práctica lo que se hace es estar atento al proceso cuando éste comienza a mostrar anomalías y tratar de analizar la disposición de los datos entre los cuales van apareciendo estos patrones de inestabilidad.

La manera en que aparecen las anomalías puede ayudar a identificar las causas asignables pero no existe una regla general ya que:

- Cada proceso tiene unas características particulares que hacen que un tipo de inestabilidades sean más frecuentes que otras;
- Dado un patrón de inestabilidad, las causas que pueden provocarlo son variadas y dependen totalmente del proceso con el que se está trabajando.

Por consiguiente, para la interpretación de los gráficos de control es primordial conocer el proceso.

En general, las causas que afectan a la media del proceso son aquellas que cuando intervienen afectan a todo el producto de forma parecida. Por el contrario, las causas que afectan a la variabilidad afectan sólo a una parte del producto.

Por supuesto que ciertas causas pueden afectar a la media y a la variabilidad del proceso a la vez. Por ello, si se trabaja con los gráficos $\bar{X}$-R es recomendable analizar el comportamiento de la media y el recorrido por separado. Primero el comportamiento del gráfico R, que es más sensible a cambios en el proceso, después el gráfico $\bar{X}$ y finalmente los dos a la vez. (No tiene sentido interpretar el gráfico $\bar{X}$ si el gráfico R no está bajo control.)

Por último hemos de insistir en el hecho de que el objetivo en la interpretación de un gráfico de control es analizar si el proceso se comporta de una forma estable a lo largo del tiempo. Este análisis no pretende comparar si el proceso se comporta de una forma estable a lo largo del tiempo. Este análisis no pretende comparar las características del proceso con sus especificaciones o tolerancias. Cuando tal comparación sea necesaria se realizará un estudio de capacidad, el cual será fiable sólo cuando el proceso esté bajo control.

Causas por las que fallan los gráficos de control de procesos por variables (CEP)

Podemos encontrarnos casos, en los que estos gráficos, NO describen adecuadamente lo que pasa o pasará en el proceso, con lo cual se pierde interés y se desacreditan estas técnicas estadísticas tan útiles y poderosas. Puestos a estudiar cuales son las causas por las que sucede esto último, descubrimos 3 grupos principales:

- Instrumentos o Sistemas de Medición inadecuados.
- Muestra, tamaño o frecuencia de extracción inadecuadas.
- El cálculo de los límites naturales de control o de los límites especificados de control, suponiendo que todos los Procesos son Gaussianos o similares.

Instrumentos y Sistemas de Medición inadecuados

A1 Calibrado inadecuado de los instrumentos de medición

Las mediciones hechas con instrumentos que no han sido calibrados adecuadamente no tienen ningún significado Físico, y tampoco significan nada los gráficos CEP, hechos con ellas.

A2 Uso de instrumentos de medición con una apreciación menor a la necesaria

Los gráficos CEP, obtenidos sobre la base de mediciones que surgen de instrumentos con una apreciación del orden de las variaciones del Proceso que se pretenden detectar, no son adecuados para detectar síntomas de «descontrol», salvo cuando estos ya son muy grandes.

Como regla general, se requieren instrumentos con una apreciación un orden de magnitud superior a la variación que se quiere detectar.

A3 Sistemas de medición inadecuados

El uso de instrumentos de medición calibrados y de una apreciación adecuada, es una condición necesaria pero no suficiente para obtener valores cuyo control por CEP tenga algún sentido Físico.

Los instrumentos de medición son normalmente operados por seres humanos, y estos también son fuente de error. La combinación de instrumento/operador, es lo que comúnmente se conoce como «Sistema de medición».

Para definir si un sistema de medición es adecuado o no para controlar un Proceso determinado, debemos asegurar unas condiciones mínimas de Repetibilidad y Reproductibilidad de los datos que estemos tomando, de manera que los errores introducidos por el conjunto Instrumento/Operador influyan mínimamente.

Muestra, tamaño o frecuencia de extracción inadecuadas

B1 Muestras inadecuadas

Es bastante común, sobre todo en Procesos de tipo continuo, que hay variaciones muy importantes en los valores de las mediciones, dependiendo del lugar donde se hacen dichas mediciones.

En un film de polietileno, en una chapa, etc., se suelen tomar la o las muestras para hacer las mediciones, en el lugar más cómodo, es necesario hacer un estudio de homogeneidad y luego si este así lo requiere «corregir» los valores.

B2 Tamaño de la muestra inadecuado

Un número insuficiente de medidas en cada muestra, puede hacer que la variación en la toma de medidas no sea constante, y por tanto introducir variaciones en los gráficos de control, es decir un escaso número de medidas en cada muestra, puede introducir un error al no describir correctamente la realidad del proceso.

B3 Frecuencia de extracción inadecuada

Cada tipo de proceso tiene una frecuencia de extracción determinada, esta viene dada en función del número de medidas por muestra y el promedio de piezas fabricadas entre dos desajustes consecutivos (cada qué intervalo de piezas el proceso se desajusta).

A no ser que el proceso sea constante el intervalo de extracción de muestras no tiene porque ser constante en el tiempo, (cada 1/2 hora p.ej.), sino que este intervalo de extracción dependerá del número total de piezas producidas.

El cálculo de los límites naturales de control o de los límites especificados de control, suponiendo que todos los Procesos son Gaussianos o similares

En estos casos, si se calculan los límites de control, como si los Procesos fueran Gaussianos, encontraremos algunas veces síntomas de «descontrol», cuando el Proceso en realidad no ha cambiado, con lo que perderemos tiempo y dinero buscando causas asignables inexistentes.

En algunos otros casos, no detectaremos síntomas de «descontrol» alguno, pese a que el Proceso claramente ha cambiado. Por tanto, si insistimos en calcular los límites de control suponiendo que estos Procesos son Gaussianos, los resultados obtenidos en la realidad NO coincidirán con la teoría.

Procesos industriales que son claramente NO GAUSSIANOS, y en ellos aplicar dicha teoría lleva a resultados que no coinciden con la realidad.

3.1.10. EJEMPLOS

3.1.10.1. Ejemplo 1 de gráficos de control y R, por variables

No sabemos si el horno en el que se cuecen las salchichas está bien regulado, y queremos conocer si su funcionamiento es uniforme, dentro de unos límites aceptables.

Para ello hemos anotado las temperaturas de la producción diaria de salchichas. Hay 10 cocciones al día y esta labor la hemos desarrollado durante 15 días alternos, tomando una temperatura de cada 2 procesos.

TABLA 4: Temperaturas de producción

Número de muestra	Submuestras				
1	78.5	76.7	79.3	78.9	76.7
2	79.2	79.1	80.3	81.4	80.2
3	80.3	80.8	80.2	79.7	81.2
4	79.8	80.6	80.1	81.4	81.1
5	81.1	80.0	79.5	80.0	79.3
6	79.6	79.7	81.1	80.0	81.5
7	81.3	81.4	80.7	80.2	81.0
8	79.3	81.3	79.0	81.5	80.8
9	78.9	80.9	79.3	79.7	80.1
10	80.7	81.2	79.6	79.8	82.0
11	80.3	80.2	79.9	80.5	81.3
12	80.5	79.1	79.7	80.3	80.1
13	79.9	80.5	81.1	80.9	80.1
14	79.3	82.0	81.5	81.3	80.3
15	80.4	81.7	81.3	79.0	81.8

1° Hallamos las medias y los recorridos de cada muestra.

2° Hallamos la media de las medias y el recorrido medio de las muestras (como podemos ver en la tabla siguiente).

TABLA 5: Cálculo de medias

Número de muestra	Submuestras					Media	Recorrido
1	78.5	76.7	79.3	78.9	76.7	78.2	2.6
2	79.2	79.1	80.3	81.4	80.2	80.04	2.3
3	80.3	80.8	80.2	79.7	81.2	80.44	1.5
4	79.8	80.6	80.1	81.4	81.1	80.6	1.6
5	81.1	80	79.5	80	79.3	79.98	1.8
6	79.6	79.7	81.1	80	81.5	80.38	1.9
7	81.3	81.4	80.7	80.2	81	80.92	1.2
8	79.3	81.3	79	81.5	80.8	80.38	2.5
9	78.9	80.9	79.3	79.7	80.1	79.78	2
10	80.7	81.2	79.6	79.8	82	80.66	2.4
11	80.3	80.2	79.9	80.5	81.3	80.44	1.4
12	80.5	79.1	79.7	80.3	80.1	79.54	1.4
13	79.9	80.5	81.1	80.9	80.1	80.5	1.2
14	79.3	82	81.5	81.3	80.3	80.88	2.7
15	80.4	81.7	81.3	79	81.8	80.84	2.8
$\bar{\bar{X}}$						**80.253**	
$\bar{R}$							**1.9533**

3° Calculamos los límites de control para el gráfico de las medias:

$LCS_X = \bar{\bar{X}} + A_2 \, x \, \bar{R} = 80.2533 + (0.577 \, x \, 1.9533) = 81.3803$
$LC_X = \bar{\bar{X}} = 80.2533$
$LCI_X = \bar{\bar{X}} - A_2 \, x \, \bar{R} = 80.2533 - (0.577 \, x \, 1.9533) = 79.1262$

4° Calculamos los límites de control para el gráfico de recorridos:

$LCS_R = D_4 * \bar{R} = 2.114 * 1.9533 = 4.1292$
$LC_R = \bar{R} = 1.9533$
$LCI_R = D_3 * \bar{R} = 0 \, x * 1.9533 = 0$

5° Representamos los gráficos con los valores de las medias y los recorridos de cada muestra, teniendo en cuenta los límites de control calculados para cada gráfico (Ilustración 21).

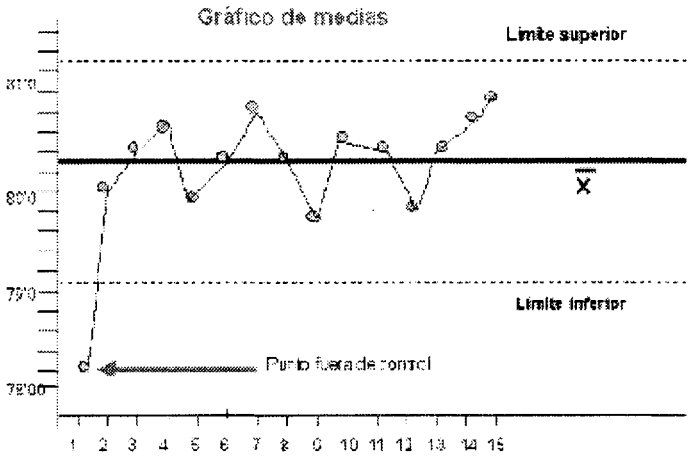
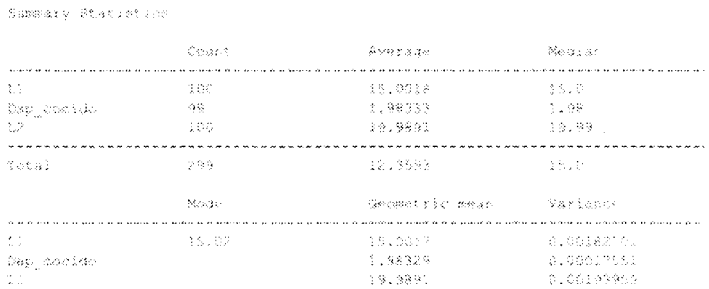


ILUSTRACIÓN 21: Gráfico de medias

6° Vemos en el primer gráfico que hay un punto fuera de control lo que deberá corregirse siempre. Este dato fuera de control puede tener su causa en que se haya modificado el proceso al principio, se haya modificado la forma de recoger datos o el punto marcado sea incorrecto. A continuación el gráfico de los recorridos (Ilustración 22).



	Count	Average	Median
X	100	15.0016	15.0
Desv_estándar	99	1.98033	1.98
LP	100	19.9691	19.99
Total	299	12.3563	15.0

	Mode	Geometric mean	Variance
X	15.07	15.0017	0.0018[illegible]
Desv_estándar		1.98329	0.0001[illegible]
LP		19.9891	0.0010[illegible]

ILUSTRACIÓN 22: Gráfico de recorridos y rangos

7° La interpretación del gráfico de recorridos es más sencilla puesto que no se ha rebasado en ningún momento los límites. Si hubiese un solo punto fuera de control entonces habría que analizar el proceso entero.

Como interpretación de resultados entresacaríamos que, al principio ha existido un valor fuera de control, que pueda ser debido a que se haya modificado la medición o que el proceso de producción fue modificado, pero que en la actualidad está bajo control. Eso no quita para que se proceda a la investigación de los datos fuera de control y el proceso en sí mismo.

3.1.10.2. Ejemplo 2 de Gráficos de Control $\bar{X}$ y R, por variables (sin valores especificados)

En la siguiente tabla se muestran los pesos de los sobres de un determinado alimento. Cada media hora se realizan 4 mediciones por muestra, sumando un total de 20 muestras. Los límites de tolerancia son 0,5360 (LST) y 0,4580 (LIT)

Con esto se pretende evaluar el comportamiento del proceso y hacer un control del mismo respecto a su localización y dispersión, con el objeto que el proceso cumpla con las especificaciones preestablecidas.

TABLA 6: Muestras de peso (gramos)

SUBGRUPO	PESO (g)				PROMEDIO	INTERVALO
Nº	X_1	X_2	X_3	X_4	$\bar{X}$	R
1	0,5053	0,4821	0,5103	0,5090	0,5017	0,0269
2	0,5102	0,5028	0,4958	0,5069	0,5039	0,0144
3	0,5221	0,5142	0,5116	0,5121	0,5150	0,0105
4	0,5074	0,5023	0,4892	0,4954	0,4986	0,0182
5	0,4816	0,5112	0,5223	0,5041	0,5048	0,0407
6	0,4862	0,5028	0,5122	0,4972	0,4996	0,0260
7	0,5111	0,5122	0,5332	0,4951	0,5129	0,0381
8	0,5328	0,5021	0,5125	0,5100	0,5144	0,0307
9	0,4912	0,5145	0,5069	0,4910	0,5009	0,0235
10	0,4652	0,4856	0,4895	0,4555	0,4740	0,0340
11	0,5160	0,4847	0,5095	0,5124	0,5056	0,0313
12	0,5010	0,4795	0,5023	0,5136	0,4991	0,0341
13	0,4864	0,5015	0,5046	0,5045	0,4992	0,0182
14	0,5023	0,5125	0,5012	0,5111	0,5068	0,0113
15	0,5005	0,5055	0,5091	0,5044	0,5049	0,0086
16	0,4952	0,4978	0,4975	0,5124	0,5007	0,0172
17	0,5046	0,4860	0,4965	0,4851	0,4930	0,0195
18	0,5029	0,4850	0,4998	0,4650	0,4882	0,0379
19	0,4721	0,4585	0,4686	0,4925	0,4729	0,0340
20	0,4652	0,4596	0,4681	0,4852	0,4695	0,0256

Primero debemos calcular las medias tanto de la media de cada muestra ($\bar{\bar{X}}$) como la de su amplitud o recorrido ($\bar{R}$):

$(\bar{\bar{X}}) = 0.4970$

$(\bar{R}) = 0.0224$

Para construir los Gráficos de Control por variables, se tiene que tener en cuenta que al determinar si un proceso está bajo «control estadístico», siempre se debe analizar primero la gráfica R.

Como los límites de control en el gráfico $\bar{X}$ dependen de la amplitud promedio, podrían haber causas especiales en la gráfica R que produzcan comportamientos anómalos en el gráfico $\bar{X}$, aún cuando el centrado del proceso esté bajo control.

Para el gráfico R, se tiene que:

Límite de Control Superior (LCS_R)

$LCS_R = D_4 * \bar{R}$

$LCS_R = 0.0511$

donde el valor de D_4 se consigue en una tabla estadística (para este caso es 2,282 con un tamaño de grupo n = 4).

Límite Central (LC_R)

$LC_R = \bar{R}$

$LC_R = 0.224$

Límite de Control Inferior (LCI_R)

$LCI_R = D_3 * \bar{R}$

$LCI_R = 0$

porque para todo proceso en que se considera un n < 7, el LCI no se indica en la gráfica.

El gráfico R de recorridos es el siguiente (Ilustración 23):

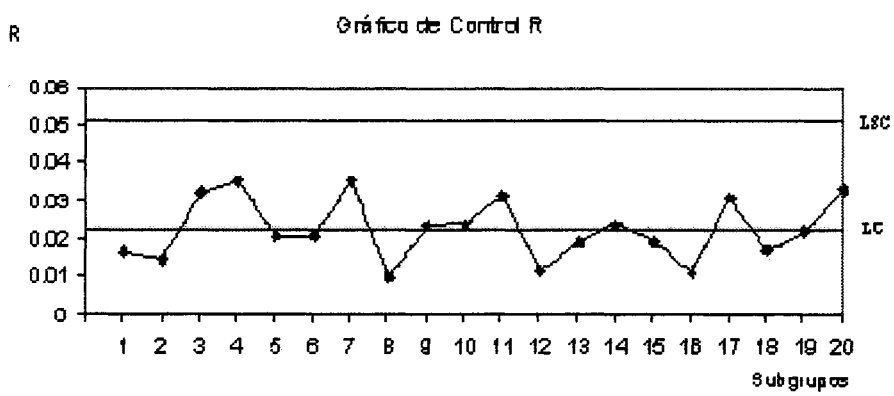


ILUSTRACIÓN 23: Gráfico de recorridos o rangos

Como se puede apreciar, el gráfico R no presenta variaciones fuera del límite superior, por lo tanto la dispersión de los datos es aceptable para calcular el gráfico.

Para el gráfico $\bar{\bar{X}}$, se tiene:

Límite de Control Superior (LCS_X).

$LCS_X = \bar{\bar{X}} + A_2 * \bar{R}$

$LCS_X = 0.5133$

donde el valor de A_2 se consigue en una tabla estadística (para este caso el valor es 0,729 con un tamaño n = 4).

Límite Central (LC_X).

$LC_X = \bar{\bar{X}}$

$LC_X = 0.4970$

Límite de Control Inferior (LCI_X)

$LCI_X = \bar{\bar{X}} - A_2 *$

$LCI_X = 0.4807$

El gráfico $\bar{\bar{X}}$ es el siguiente (Ilustración 24):

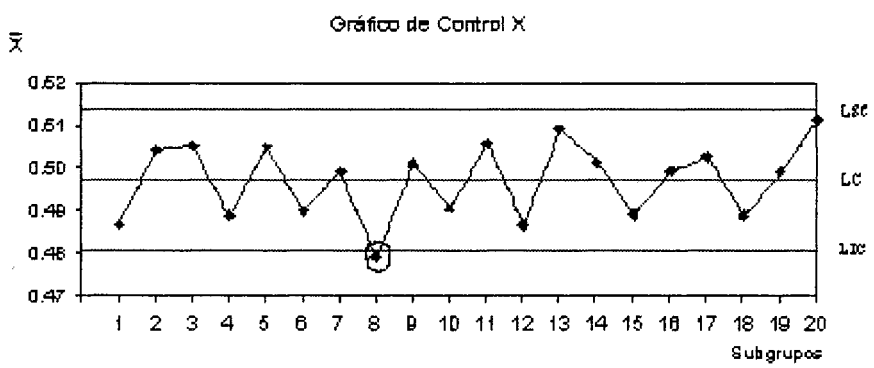


ILUSTRACIÓN 24: Gráfico de la media

Como se puede apreciar un punto queda fuera del rango calculado, por lo tanto el proceso se encuentra fuera de control estadístico.

En este caso, habría que investigar y eliminar la causa asignable, que podría haberse debido al uso de algún material defectuoso o una mala lectura del instrumento. Este dato debe eliminarse de la gráfica y recalcular todo de nuevo, pero sin considerar el subgrupo 8.

Nota. Esto no siempre es así, si los puntos fuera de control son de tal magnitud, entonces no queda más remedio que una vez encontrada y eliminadas las causas en la práctica, habría que repetir el proceso, recogiendo nuevos datos.

Después de la corrección, los resultados son:

Gráfico R corregido

$LCS_R = 0.0527$
$LC_R = 0.0231$
$LCI_R = 0$

Gráfico $\bar{X}$ corregido

$LCS_X = 0.5147$
$LC_X = 0.4979$
$LCI_X = 0.4811$

Los gráficos son los siguientes (Ilustración 25):

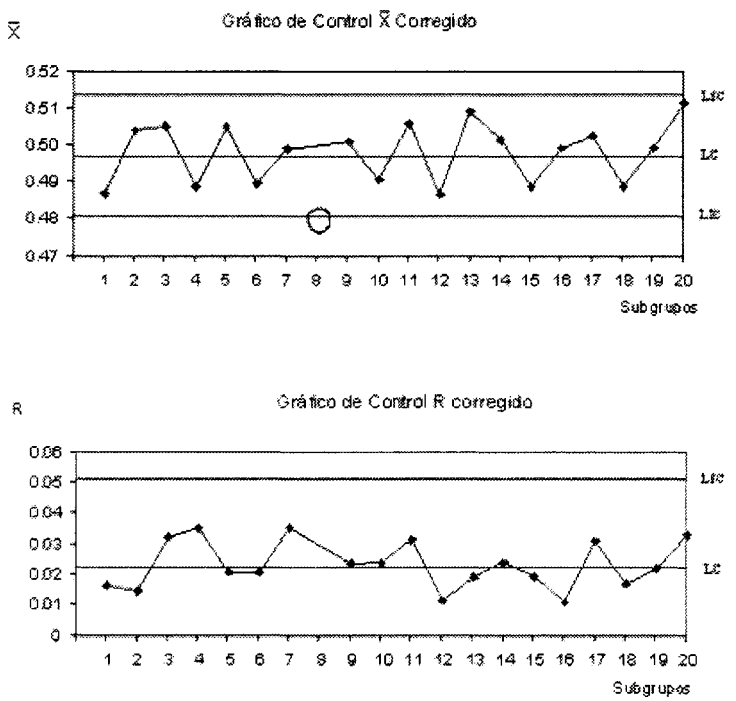


ILUSTRACIÓN 25: Gráficos de $\bar{X}$-R corregidos

Como se puede apreciar en ambos gráficos, ahora el proceso se encuentra en «control estadístico» (bajo control).

3.1.10.3. Ejemplo de gráficos de control $\bar{X}$-S

Se han tomado 26 muestras de tamaño 6 de cada una y se ha calculado la media y la desviación típica de cada una de ellas, dando los resultados que se muestran en la siguiente tabla:

TABLA 7: Cálculo de medias

Muestra	Medias	Desviaciones típicas	Muestra	Medias	Desviaciones típicas
1	53	9	14	51	4
2	51	3	15	48	1
3	48	2	16	49	4
4	48	1	17	52	5
5	48	6	18	50	6
6	51	7	19	51	6
7	49	3	20	51	8
8	51	8	21	49	6
9	48	2	22	50	7
10	49	5	23	49	9
11	47	4	24	48	4
12	47	5	25	47	3
13	47	2	26	49	5

En este ejemplo la media general es:

$$\bar{\bar{X}} = \frac{53 + 51 + \ldots + 49}{26} = 49.5769$$

y la desviación típica media vale:

$$\bar{S} = \frac{9 + 3 + \ldots + 5}{26} = 4.8077$$

en consecuencia, los límites de control para la media son:

$$\bar{X} = \bar{\bar{X}} \pm A_1\bar{S} = 49{,}5769 \pm 1.409978 \times 4.8077 = 49.5769 \pm 6.7788$$

esto es, LIC = 42.80, LC = 56.36, LSC = 56.36, y para la desviación típica obtenemos las siguientes líneas de control:

Límite Inferior de Control = LIC = $B_3\bar{S} = 0.030363 \times 4.8077 = 0.15$
Límite Central = LC = $\bar{S} = 4.81$
Límite Superior de Control = LSC = $B_4\bar{S} = 1.96964 \times 4.8077 = 9.47$

En el gráfico de control para las medias tendremos la muestra nº 13 esta fuera de control, por lo que vamos a prescindir de la misma y a repetir los cálculos con las 25 muestras restantes. Esto nos conduce a una media general $\bar{\bar{X}} = 49.28$ y a una desviación típica media $\bar{S} = 4.92$, a partir de las cuales determinamos las siguientes líneas de control:

Para las medias: LIC = 42.94; LC = 49.28; LSC = 56.22
Para las desviaciones típicas: LIC = 0.15; LC = 4.92; LSC = 9.69

Lo que permite construir gráficos de control, en los que vemos que tanto respecto de las medias como las desviaciones típicas de las 25 muestras se encuentran dentro de los límites de control y no presentan tendencias ni rachas significativas, por lo que el proceso se halla en estado de control.

3.2. GRÁFICOS DE CONTROL POR ATRIBUTOS

Los diagramas de control por atributos constituyen la herramienta esencial utilizada para controlar características de calidad cualitativas, esto es, características no cuantificables numéricamente. Ejemplos clásicos de estas características no medibles son la fracción o porcentaje de unidades defectuosas (p) en la producción, el número de unidades defectuosas (np) en la producción, el número de defectos por unidad (u) producida y el número de defectos de todas las unidades producidas.

Los diagramas de control por atributos son apropiados en casos en los que es necesario reducir el rechazo del proceso. Típicamente se aplican en situaciones en las que el proceso es una operación de montaje complicada y la calidad del producto se mide en términos de la ocurrencia de disconformidades, del funcionamiento exitoso o fallido del producto, etc. Algunos ejemplos prácticos son los procesos de fabricación de computadoras, de equipos de automatización de oficinas, de automóviles y los subsistemas importantes de estos productos. También son aplicables los diagramas de control por atributos cuando se necesita un control del proceso, pero no se pueden obtener prácticamente datos de mediciones.

En muchas aplicaciones será necesario elegir entre un diagrama de control por variables y uno de control por atributos. Si la característica de calidad a estudiar no ofrece dificultades, la selección estará bien definida, pero en otros casos, la selección no será tan evidente, y el analista tendrá que tomar en cuenta varios factores para poder elegir entre diagramas de control por atributos o por variables. Los diagramas de con-

trol por atributos tienen la ventaja de que hacen posible considerar varias características de calidad al mismo tiempo y clasificar los productos como disconformes si no satisfacen las especificaciones de cualquiera de las características. Por otra parte, si se manejan las diversas características de calidad como variables, entonces habrá que medir cada una de ellas y utilizar separadamente un diagrama de $\bar{X}$ y R para cada una, o bien alguna técnica de control multivariante en la que se consideren a la vez todas las características. Hay una evidente sencillez asociada el diagrama de atributos en este caso. Además, con la inspección por atributos pueden evitarse mediciones costosas en recursos y tiempo.

Cuando las características de calidad no se pueden representar en forma conveniente por números, cada artículo o producto estudiado o inspeccionado suele clasificarse como conforme o disconforme con las especificaciones para tal característica de calidad. Las características de calidad de este tipo se llaman *atributos*. Los términos *no defectuoso* o *defectuoso* se utilizan con frecuencia para identificar estas dos clasificaciones de un producto, pero la terminología de *conforme* y *disconforme* se ha extendido mucho. De esta forma hay bibliografía que habla del porcentaje de unidades defectuosas para referirse a p, mientras existe otro tipo de bibliografía que habla del porcentaje de unidades disconformes(o no conformes) para referirse al mismo concepto.

Seguidamente hablaremos de los diagramas de control de atributos más comúnmente utilizados:

- *Diagrama p*: fracción o porcentaje de unidades defectuosas también llamada fracción o porcentaje no conforme o disconforme en la producción.
- *Diagrama np*: el número de unidades defectuosas o número de unidades no conformes o disconformes en la producción.
- *Diagrama u*: el número de defectos por unidad o número de disconformidades por unidad producida.
- *Diagrama c*: número defectos o número de disconformidades de todas las unidades producidas.

3.2.1. DIAGRAMA P: DIAGRAMA DE CONTROL PARA EL PORCENTAJE O FRACCIÓN DE UNIDADES DEFECTUOSAS

El porcentaje o fracción de unidades defectuosas se define como el cociente del número de artículos defectuosos en una población entre el número total de artículos de la citada población. Los artículos pueden tener varías características de calidad que se examinan simultáneamente. Si el artículo no esta conforme con el estándar de una o más de tales características, se clasificará como defectuoso. En general, expresamos la fracción de unidades defectuosas como un número decimal, aunque se usa en ocasiones el llamado porcentaje de unidades defectuosas (que es exactamen-

te el producto de 100% y la fracción de unidades defectuosas). Se utiliza frecuentemente el porcentaje de unidades defectuosas porque tiene una apreciación más intuitiva. No obstante, el diagrama de control para ambos conceptos es similar.

Los principios estadísticos que sirven de base al diagrama de control de la fracción o proporción disconforme se basan en la distribución binomial. Supóngase que el proceso de producción funciona de manera estable, de tal forma que la probabilidad de que cualquier artículo no esté conforme con las especificaciones es p, y que los artículos producidos sucesivamente son independientes. Entonces cada artículo producido es una variable aleatoria de Bernoulli, con parámetro p. Si se selecciona una muestra aleatoria de n artículos del producto, y si D representa el número de artículos no conformes en la muestra, entonces D tiene distribución binomial con parámetro n y p, es decir:

$$P(D = x) = \binom{n}{x} p^x q^{n-x} \qquad x=0,1,...n$$

Ya sabemos que la media y la varianza de la variable aleatoria D son np y np(1-p), respectivamente. Por otra parte, también sabemos que a partir de la variable aleatoria D se puede definir la variable aleatoria $\hat{p} = D / n$ que se denomina *fracción disconforme muestral,* y que lógicamente será el cociente del número de artículos disconformes D en la muestra, entre el tamaño muestral. La distribución de $\hat{p}$ puede calcularse a partir de la binomial, donde su media es p y su varianza es p(1-p)/n.

A) Diagrama «p» con patrón dado

Supongamos que se conoce la verdadera fracción disconforme p en el proceso de fabricación, o que la administración le asigna un valor estándar. Para construir este diagrama de control se toman muestras subsecuentes de n unidades (subgrupos), se calcula en cada una la fracción muestral disconforme $\hat{p}$, y se grafica el estadístico $\hat{p}$ en el diagrama para cada muestra (en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas se colocan las fracciones muestrales disconformes $\hat{p}$).

Para calcular los límites de control de Shewart sabemos que si T es un estadístico muestral que mide la característica de calidad de interés, y la media de T es μ_T, y su desviación estándar, σ_T, entonces la línea central, y los límites superior e inferior del gráfico de control k-sigmas de Shewart son:

$$LSC = \mu_T + k\sigma_T$$

$$\text{Línea central} = \mu_T$$

$$\text{LIC} = \mu_T - k\sigma_T$$

donde k es la distancia entre los límites de control y la línea central, expresada en desviaciones estandars. Como nuestro estadístico es $\hat{p}$ con media p y varianza p(1-p)/n, podemos definir los límites de control 3-sigmas de Shewart para cada muestra como:

$$LSC = p + 3\sqrt{\frac{p(1-p)}{n}}$$
$$Línea\ \ central = p$$
$$LIC = p - 3\sqrt{\frac{p(1-p)}{n}}$$

Mientras la fracción muestral disconforme $\hat{p}$ quede entre los límites de control para cada muestra y la sucesión de puntos definida por $\hat{p}$ para cada muestra no exhiba un patrón sistemático, se concluye que el proceso esta bajo control, al nivel de fracción de defectuosos p. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que la fracción de disconformes del proceso cambio hacia un nuevo nivel y que el proceso está fuera de control.

Para calcular los límites probabilísticos de control al nivel α utilizamos la distribución del estadístico $\hat{p}$, con lo que sabemos que para cada muestra:

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \rightarrow N(0,1)$$

Por lo tanto, mediante la distribución normal N(0,1), para un α dado, podremos calcular k_α tal que:

$$P\left(-k_\alpha \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq k_\alpha\right) = 1 - \alpha$$

Pero de esta expresión se deduce que:

$$F_{N(0,1)}(k_\alpha) - F_{N(0,1)}(-k_\alpha) = 1 - \alpha \Rightarrow F_{N(0,1)}(k_\alpha) - (1 - F_{N(0,1)}(k_\alpha)) = 1 - \alpha \Rightarrow$$
$$2F_{N(0,1)}(k_\alpha) - 1 = 1 - \alpha \Rightarrow F_{N(0,1)}(k_\alpha) = 1 - \alpha/2$$

Con lo que k_α valdrá F, siendo F^{-1} (q – α/2) la función de distribución de la normal N(0,1).

Como para cada muestra se cumple que:

$$\left(-k_\alpha \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq k_\alpha \right) \Leftrightarrow p - k_\alpha \sqrt{\frac{p(1-p)}{n}} \leq \hat{p} \leq p + k_\alpha \sqrt{\frac{p(1-p)}{n}}$$

ya tenemos los límites probabilísticos al nivel para cada muestra de la forma siguiente:

$$LSC = p + k_\alpha \sqrt{\frac{p\,(1-p)}{n}}$$

$$Línea\ central = p$$

$$LIC = p - k_\alpha \sqrt{\frac{p\,(1-p)}{n}}$$

Sabemos que se acostumbra a reemplazar k_α por 3, que aproximadamente equivale a $\alpha = 0.002$, de manera que se usan límites de tres sigmas y se obtienen como caso particular los diagramas de Shewart.

Hasta aquí hemos supuesto que los tamaños de todas las muestras (subgrupos) son iguales a n, pero puede ocurrir que esto no sea cierto. Si el tamaño de cada muestra es n_i para i = 1... k, una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán líneas rectas. Otra opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i = 1...k(n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

B) Diagrama «p» sin patrón dado

Supongamos ahora que no se conoce la verdaderas fracción disconforme p en el proceso de la fabricación, situación más realista que la anterior. Para construir el diagrama de control de la fracción de defectuosos hay que estimar p, para lo que se to-

man k muestras subsecuentes de n unidades (subgrupos), se calcula en cada una la fracción muestral disconforme, y se considera el estadístico $\overline{p}$ de p definido como

$$\overline{p} = \sum_{i=1}^{k} \hat{p}_i$$

Para calcular los límites de control de Shewart sustituimos p por su estimador, obteniendo los límites de control 3-sigmas para cada muestra como:

$$LSC = \overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

$$Línea\ \ central = \overline{p}$$

$$LIC = \overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

Mientras la fracción muestral disconforme para cada muestra (subgrupo) quede entre los límites de control y la sucesión de puntos definida por las fracciones muestrales disconformes para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que la fracción de disconformes del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Hasta aquí hemos supuesto que los tamaños de todas las muestras (subgrupos) son iguales a n, pero puede ocurrir que esto no sea cierto. Si el tamaño de cada muestra es n_i para i = 1...k, una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán líneas rectas. Otra opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i = 1...k(n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

No obstante, cuando los tamaños de los subgrupos o muestra n_i son muy distintos, suelen tomarse para cada muestra los siguientes límites de control 3-sigmas:

$$LSC = Min\left(\overline{p} + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}}, 1\right)$$

$$Línea\ \ central = \overline{p}$$

$$LIC = Max\left(\overline{p} - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n_i}}, 0\right)$$

También es usual tomar como la media de los de todas las muestras, pero ponderada por los tamaños muestrales n_i, es decir:

$$\overline{p} = \frac{\sum_{i=1}^{k} n_i \hat{p}_i}{\sum_{i=1}^{k} n_i}$$

La notación utilizada en el caso más general es la siguiente:

p = fracción o proporción esperada de unidades defectuosas producidas por el proceso
$\hat{p}_i$ = fracción de unidades defectuosas en el iésimo subgrupo (muestra)
D_i = número de unidades defectuosas en el iésimo subgrupo
n_i = número de unidades en el iésimo subgrupo
$\overline{p}$ = fracción media de unidades defectuosas en el proceso
k = número de subgrupos

Para realizar el gráfico de control *p*, en el eje de abscisas de colocan los números de muestra o subgrupo y en el eje de ordenadas se colocan las fracciones de unidades defectuosas $\hat{p}_i$ de cada muestra.

Para calcular los ***límites probabilísticos de control*** podemos utilizar la función de distribución de la variable binomial D_i o también una variable beta incompleta de parámetros α y β, definida como:

$$I_T(\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^T t^{\alpha-1}(1-t)^{\beta-1} dt$$

Para hallar el límite inferior de control utilizamos la relación entre la distribución binomial (D_i es binomial) y la beta de la siguiente forma:

$$P(\hat{p}_i < LICP) = 1 - P(\hat{p}_i \geq LICP) = 1 - P(D_i / n_i \geq LICP) = 1 - P(D_i \geq n_i LICP) =$$
$$1 - I_{\overline{p}}(n_i LICP, n_i + 1 - n_i LICP) = I_{1-\overline{p}}(n_i + 1 - n_i LICP, n_i LICP)$$

Entonces podemos calcular el límite probabilistico inferior de control LICP al nivel α mediante la expresión:

$$I_{1-\overline{p}}(n_i + 1 - n_i LICP, n_i LICP) = \alpha / 2$$

Alternativamente, para hallar el límite superior de control utilizamos la relación entre la distribución binomial y la beta de la siguiente forma:

$$P(\hat{p}_i > LSCP) = P(D_i / n_i > LSCP) = P(D_i > n_i LSCP) = I_{\bar{p}}(n_i LSCP, n_i + 1 - n_i LSCP)$$

Entonces podemos calcular el límite probabilístico superior de control LSCP al nivel α mediante la expresión:

$$I_{\bar{p}}(n_i LSCP, n_i + 1 - n_i LSCP) = \alpha / 2$$

El diagrama de control de la fracción de defectuosos tiene tres parámetros que hay que especificar: el tamaño muestral, la frecuencia de muestreo, y la amplitud de los límites de control. Es útil tener algunas directrices generales para seleccionar estos parámetros.

Es común basar un diagrama de control de la fracción de defectuosos en una inspección de 100% de todo el rendimiento del proceso en un periodo conveniente, como un turno o un día. En este caso, el tamaño muestral y la frecuencia de muestreo están relacionados. En general, se seleccionaría una frecuencia de muestreo apropiada para la tasa de producción, y eso fija el tamaño muestral. El uso de subgrupos racionales también puede ser importante en la determinación de la frecuencia de muestreo. Por ejemplo, si existen tres turnos y se sospecha que los turnos difieren respecto a su nivel general de calidad, entonces habrá que usar la salida de cada turno como un subgrupo en vez de juntar las salidas de los tres turnos para obtener una fracción defectuosas diaria.

Si hay que seleccionar una muestra del rendimiento del proceso, tendremos que escoger el tamaño muestral n. Se han sugerido varias reglas para la selección de n. Si p es muy pequeño, habrá que elegir n lo suficientemente grande para tener una alta probabilidad de encontrar por lo menos una unidad disconforme en la muestra. Es decir, hay que evitar que los límites de control puedan ser tales que la presencia de una sola unidad no conforme indique una condición fuera de control. Como para cualquier $p > 0$ existe una probabilidad positiva de producir algunos artículos defectuosos, no es razonable concluir, en muchos casos, que el proceso está fuera de control con base en la observación de un solo artículo defectuoso.

Para evitar estos problemas se podría escoger el tamaño muestral n_i de manera que la probabilidad de hallar por lo menos un artículo no conforme por muestra sea al menos igual a γ. Si D_i es el número de unidades defectuosas en la muestra i, se determinará el valor de n de modo que

$P\{ D_i \langle 1 \}\langle \gamma$. El valor de n_i se hallará teniendo en cuenta la función de distribución de la variable binomial D_i o mediante la aproximación de Poisson a la binomial.

Algunos autores, como por ejemplo Duncan, han sugerido que el tamaño n tendría que ser suficientemente grande para tener una probabilidad aproximada del 50% de detectar un cambio de alguna cantidad especificada en el proceso. Al suponer que se puede aplicar la aproximación normal a la binomial, habría que elegir n de manera que el límite superior de control coincida exactamente con la fracción no conforme en el estado fuera de control. Si es la magnitud del cambio en el proceso, entonces n tendrá que satisfacer:

$$\delta = k_\alpha \sqrt{\frac{p(1-p)}{n}} \Rightarrow n = \left(\frac{k_\alpha}{\delta}\right)^2 p(1-p)$$

Si es pequeño el valor bajo control de la fracción disconforme, otro criterio útil es escoger n suficientemente grande para que el diagrama de control tenga un límite inferior de control positivo. Esto asegura un mecanismo para obligamos a investigar una o más muestras que contienen un número anormalmente pequeño de artículos disconformes.

Normalmente se usan límites de control de tres sigmas en el diagrama de control de la fracción de defectuosos. Si se utilizan límites de control más estrechos hacen al diagrama de control más sensible a pequeños cambios en p, pero al costo de más frecuentes *falsas alarmas* de proceso en situación de fuera de control. Ocasionalmente, se ha probado el uso de límites más estrechos para tratar de forzar mejoras en la calidad del proceso. Sin embargo, hay que tener cuidado, porque demasiadas falsas alarmas destruirán la confianza del personal operativo en el programa de diagramas de control.

Debe advertirse que el diagrama de control de la fracción disconforme no es un modelo universal para todos los datos respecto a la fracción no conforme. Se basa en el probabilístico binomial en el que se cumple que la probabilidad de ocurrencia de un artículo con disconformidad es constante, y unidades sucesivas en la producción son independientes. En procesos en que las unidades no conformes se agrupan, o en los que la probabilidad de que una unidad sea disconforme depende de la conformidad (o no conformidad) de las unidades anteriores, el diagrama de control de la fracción disconforme es muchas veces de poca utilidad.

Por otra parte, hay que tener cuidado con la interpretación de los puntos del diagrama de control que se hallan por debajo del límite inferior de control. Tales puntos no representan a menudo una mejora real en la calidad del proceso. Frecuentemente son el resultado de errores en el método de inspección o recogida de datos debido a que los inspectores o los operarios han recibido un adiestramiento inadecuado o carece de experiencia, o a que el equipo de prueba e inspección está mal calibrado. No todos los cambios a la baja de la fracción disconforme o fracción de defectuosos p significan una mejora en la calidad.

Curva característica de operación para el diagrama «p»

Existe una relación estrecha entre los diagramas de control y la prueba de hipótesis. Esencialmente, un diagrama de control puede considerarse como una prueba de la hipótesis de que el proceso está bajo control estadístico. Un punto que se ubica entre los límites de control es equivalente a no poder rechazar la hipótesis de que el proceso está bajo control estadístico. Un punto que se ubica fuera de los límites equivale al rechazo de la hipótesis del control estadístico. Como en toda prueba de hipótesis, puede considerarse la probabilidad del error tipo I del diagrama de control (concluir que el proceso está fuera de control cuando en realidad no lo está) y la probabilidad del error tipo II de dicho diagrama (concluir que el proceso está bajo control cuando en realidad no es así). La curva característica de operación de un diagrama de control representa su probabilidad de error tipo II e indica la aptitud del diagrama de control para detectar cambios de diferente magnitud en el proceso.

La curva característica de operación del diagrama de control de la fracción no conforme o fracción de defectuosos es una representación gráfica de la probabilidad de aceptar incorrectamente la hipótesis de un control estadístico, es decir, un error tipo II o β, contra la fracción disconforme del proceso. La curva proporciona una medida de la sensibilidad del diagrama de control; esto es, su capacidad para detectar un cambio en la fracción disconforme del proceso desde su valor nominal $\bar{p}$ hacia algún otro valor p. La probabilidad del error tipo II para el diagrama de control de la fracción no conforme puede calcularse a partir de:

Como D_i es una variable aleatoria binomial con parámetros n_i, y p, β puede obtenerse a partir de la función de distribución de la binomial.

No obstante, es más general y preciso el uso de la relación entre la distribución binomial y la beta, que permite calcular β de la forma siguiente:

$$\beta = P\{\hat{p}_i \leq LSC|p\} - P\{\hat{p}_i < LIC|p\} = P\{D_i \leq nLSC|p\} - P\{D_i < n\,LIC|p\} =$$
$$P\{D_i < nLSC|p\} + P\{D_i = nLSC|p\} - P\{D_i < n\,LIC|p\} =$$
$$I_{1-p}(n+1-nLSC, nLSC) + P\{D_i = nLSC|p\} - I_{1-p}(n+1-nLIC, nLIC) =$$
$$I_p(nLSC, n+1-nLSC) + P\{D_i = nLSC|p\} - I_p(nLIC, n+1-nLIC)$$

En la curva característica de operación suele situarse β en el eje de ordenadas y la fracción de defectuosos p en el eje de abscisas.

3.2.2. DIAGRAMA DE CONTROL DEL NUMERO DE UNIDAD DEFECTUOSAS. DIAGRAMA «np»

Es posible basar un diagrama de control en el número de unidades defectuosas (o número disconforme) en vez de en la fracción de unidades defectuosas o fracción no conforme. Este diagrama de control suele denominarse ***diagrama np.***

A continuación veremos cómo se puede utilizar el estadístico fracción disconforme para desarrollar un diagrama de control de la fracción de defectuosos.

A) Diagrama «np» con patrón dado

Supongamos que se conoce la verdadera fracción disconforme p en el proceso de fabricación, o que la administración le asigna un valor estándar. Para construir el diagrama de control ***np*** se toman muestras subsecuentes de ***n*** unidades (subgrupos), se calcula en cada una la fracción muestral disconforme $\hat{p}$, y se grafica el estadístico $n\hat{p}$ en el diagrama para cada muestra (en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas se colocan los números de unidades defectuosas $n\hat{p}$).

Para calcular los límites de control de Shewart sabemos que si T es un estadístico muestral que mide la característica de calidad de interés, y la media de T es μ_T, y su desviación estándar, es σ_T, entonces la línea central, y los límites superior e inferior del gráfico de control k-sigmas de Shewart son:

$$\begin{aligned} &\text{LSC} = \mu_T + k\sigma_T \\ &\text{Línea central} = \mu_T \\ &\text{LIC} = \mu_T - k\sigma_T \end{aligned}$$

donde k es la distancia entre los límites de control y la línea central, expresada en desviaciones estándares. Como nuestro estadístico es $n\hat{p}$ que es una variable binomial con media np y varianza np(1-p), podemos definir los límites de control 3-sigmas de Shewart para cada muestra como:

$$\begin{aligned} &\text{LSC} = np + 3\sqrt{np(1-p)} \\ &\text{Línea central} = np \\ &\text{LIC} = np - 3\sqrt{np(1-p)} \end{aligned}$$

Mientras el número de unidades defectuosas $n\hat{p}$ quede entre los límites de control para cada muestra y la sucesión de puntos definida por $n\hat{p}$ para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control, al nivel de fracción de defectuosos p. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de unidades defectuosas del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Para calcular los límites probabilísticos de control al nivel α utilizamos la distribución binomial del estadístico $n\hat{p}$, con lo que sabemos que para cada muestra:

$$\frac{n\hat{p}-np}{\sqrt{np(1-p)}} \to N(0,1)$$

Por lo tanto, mediante la distribución normal N(0, 1), para un α dado, podremos calcular k_α tal que:

$$P\left(-k_\alpha \le \frac{n\hat{p}-np}{\sqrt{np(1-p)}} \le k_\alpha\right) = 1-\alpha$$

Pero de esta expresión se deduce que

$$F_{N(0,1)}(k_\alpha) - F_{N(0,1)}(-k_\alpha) = 1-\alpha \Rightarrow F_{N(0,1)}(k_\alpha) - (1 - F_{N(0,1)}(k_\alpha)) = 1-\alpha \Rightarrow$$
$$2F_{N(0,1)}(k_\alpha) - 1 = 1-\alpha \Rightarrow F_{N(0,1)}(k_\alpha) = 1-\alpha/2$$

con lo que k_α valdrá $F^{-1}(1-\alpha/2)$, siendo F la función de distribución de la normal N(0, 1).

Como para cada muestra se cumple que:

$$-k_\alpha \le \frac{n\hat{p}-np}{\sqrt{np(1-p)}} \le k_\alpha \Leftrightarrow np - k_\alpha\sqrt{np(1-p)} \le \hat{p} \le np + k_\alpha\sqrt{np(1-p)}$$

ya tenemos los límites probabilísticos al nivel α para cada muestra de la forma siguiente:

$$LSCP = np + k_\alpha\sqrt{np(1-p)}$$
$$Linea \quad central = np$$
$$LICP = np - k_\alpha\sqrt{np(1-p)}$$

Sabemos que se acostumbra a reemplazar k_α por 3, que aproximadamente equivale a $\alpha = 0.002$, de manera que se usan límites de tres sigmas y se obtienen como caso particular los diagramas de Shewart.

Hasta aquí hemos supuesto que los tamaños de todas las muestras (subgrupos) son iguales a n, pero puede ocurrir que esto no sea cierto. Si el tamaño de cada muestra es n_i para i = 1,...,k, una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las lineas de control

no serán líneas rectas. Otra opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i = 1,...,k(n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

B) Diagrama «np» sin patrón dado

Supongamos ahora que no se conoce la verdadera fracción disconforme p en el proceso de fabricación. Para construir el diagrama de control del número de unidades defectuosas hay que estimar p, para lo que se toman k muestras subsecuentes de n unidades (subgrupos), se calcula en cada una la fracción muestral disconforme $\hat{p}_i$, y se considera el estadístico $\bar{p}$ definido como la media de los $\hat{p}_i$. Este estadístico

$$\bar{p} = \sum_{i=1}^{k} \hat{p}_i$$

será el estimador p.

Para calcular los límites de control de Shewart sustituimos p por su estimador , obteniendo los límites de control 3-sigmas para cada muestra como:

$$LSCP = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$$
$$Linea \quad central = n\bar{p}$$
$$LICP = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$$

Mientras el número de unidades defectuosas para cada muestra (subgrupo) quede entre los límites de control y la sucesión de puntos definida por los números de unidades defectuosas para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de unidades defectuosas del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Hasta aquí hemos supuesto que los tamaños de todas las muestras (subgrupos) son iguales a n, pero puede ocurrir que esto no sea cierto. Si el tamaño de cada muestra es n_i para i = 1,..,k, una vez que tenemos límites de control para cada muestra, po-

demos optar por graficarlos tal y como se obfienen, con lo que las líneas de control no serán líneas rectas. Otra opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i=1,...,k(n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

No obstante, cuando los tamaños de los subgrupos o muestras n_i son muy distintos, suelen tomarse para cada muestra los siguientes límites de control 3-sigmas:

$$\text{LSCP} = n_i\,\bar{\mu} + k_\alpha \sqrt{n_i\,\bar{\mu}}$$
$$\text{Línea central} = n_i\,\bar{\mu}$$
$$\text{LICP} = Max\left(n_i\,\bar{\mu} - k_\alpha \sqrt{n_i\,\bar{\mu}}, 0\right)$$

También es usual tomar $\bar{p}$ como la media de los $\hat{p}_i$ de todas las muestras, pero ponderada por los tamaños muestrales n_i, es decir:

$$\bar{p} = \frac{\sum_{i=1}^{k} n_i \hat{p}_i}{\sum_{i}^{k} n_i}$$

La notación utilizada en el caso más general es la siguiente:

p = fracción o proporción esperada de unidades defectuosas producidas por el proceso
$\hat{p}_i$ = fracción de unidades defectuosas en el iésimo subgrupo (muestra)
D_i = número de unidades defectuosas en el iésimo subgrupo
n_i = número de unidades en el iésimo subgrupo
$\bar{p}$ = fracción media de unidades defectuosas en el proceso
k = número de subgrupos

Para realizar el gráfico de control *np*, en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas se colocan los números de unidades defectuosas $D_i = n_i \hat{p}_i$.

Para calcular los *límites probabilísticos de control* podemos utilizar la función de distribución de la variable binomial D_i o también una variable beta incompleta de parámetros α y β, definida como:

$$I_T(\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\int_0^T t^{\alpha-1}(1-t)^{\beta-1}dt \quad 0<T<1, \quad \alpha>0, \quad \beta>0$$

Para hallar el límite inferior de control utilizamos la relación entre la distribución binomial (D_i es binomial) y la beta de la siguiente forma:

$$P(n_i\hat{p}_i<LICP)=P(D_i<LICP)=1-P(D_i\geq LICP)=$$
$$1-I_{\overline{p}}(LICP,\ n_i+1-LICP)=I_{1-\overline{p}}(n_i+1-LICP,\ LICP)$$

Entonces podemos calcular el límite probabilístico inferior de control LICP al nivel α mediante la expresión:

$$I_{1-\overline{p}}(n_i+1-LICP,\ LICP)=\alpha/2$$

Alternativamente, para hallar el límite superior de control utilizamos la relación entre la distribución binomial y la beta de la siguiente forma:

$$P(n_i\hat{p}_i>LSCP)=(D_i>LSCP)=I_{\overline{p}}(LSCP,\ n_i+1-LSCP)$$

Entonces podemos calcular el límite probabilístico superior de control LSCP al nivel α mediante la expresión:

$$I_{\overline{p}}(LSCP,\ n_i+1-LSCP)=\alpha/2$$

Curva característica de operación para el diagrama «np»

La curva característica de operación del diagrama de control del número de unidades defectuosas es una representación gráfica de la probabilidad de aceptar incorrectamente la hipótesis de un control estadístico (es decir, un error tipo II o β) contra el número de unidades defectuosas del proceso. La curva proporciona una medida de la sensibilidad del diagrama de control; esto es, su capacidad para detectar un cambio en el número de unidades defectuosas del proceso desde su valor nominal $n\overline{p}$ hacia algún otro valor *np*. La probabilidad del error tipo II para el diagrama de control del número de unidades defectuosas puede calcularse a partir de:

$$\beta=P\{D_i\leq LSC|p\}-P\{D_i<LIC|p\}$$

Como D_i es una variable aleatoria binomial con parámetros n_i y p, β puede obtenerse a partir de la función de distribución de la binomial.

No obstante, es más general y preciso el uso de la relación entre la distribución y la binomial y la beta, que permite calcular β de la forma siguiente:

$$\beta = P\{D_i \leq LSC|p\} - P\{D_i < LIC|p\} =$$
$$P\{D_i < LSC|p\} + P\{D_i = LSC|p\} - P\{D_i < LIC|p\} =$$
$$I_{1-p}(n+1-LSC, LSC) + P\{D_i = LSC|p\} - I_{1-p}(n+1-LIC, LIC) =$$
$$I_p(LSC, n+1-LSC) + P\{D_i = LSC|p\} - I_p(LIC, n+1-LIC)$$

En la curva característica de operación suele situarse β en el eje de ordenadas y el número de unidades defectuosas *np* en el eje de abscisas.

3.2.3. DIAGRAMA DE CONTROL DEL NÚMERO DE DEFECTOS O DIAGRAMA «C»

Es posible basar un diagrama de control en el número de defectos (o no conformidades) en vez de en el número de unidades defectuosas. Si se considera el número de defectos por unidad de inspección producida en el proceso se obtienen los ***diagramas u*** o diagramas de control del número de defectos por unidad. Si se considera el número total de defectos en la producción se obtienen los ***diagramas c*** o diagramas del número de defectos.

A continuación nos centraremos en los diagramas *c* para desarrollar un diagrama de control del número de defectos.

A) Diagrama «c» con patrón dado

Supongamos que se conoce el número esperado u de defectos por unidad de inspección producida en el proceso de fabricación, o que la administración le asigna un valor estándar. Para construir el diagrama de control *c* se toman muestras subsecuentes de n_i unidades (subgrupos), se calcula en cada una el número de defectos por unidad u_i, y se grafica el estadístico $n_i\, u_i$ en el diagrama para cada muestra (en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas se colocan los números de defectos $c_i = n_i\, u_i$ en las muestras). Se considera que c_i es el número total de defectos en el i-ésimo subgrupo, con lo que $u_i = c_i/n_i$ (i=1,2,..., k) siendo *k* el número de subgrupos.

En la mayoría de los casos, la unidad de inspección será una sola unidad del producto, aunque no necesariamente siempre es así. La unidad de inspección es sencillamente una entidad apropiada para registrar los defectos. Podría ser un grupo de

5 artículos, etcétera. El modelo probabilistico adecuado para la ocurrencia del número de defectos en una población grande es la distribución de Poisson. Entonces, podemos suponer que para cada subgrupo el número de defectos c_i sigue una distribución de Poisson de parámetro $n_i\, u_i$ donde u es el número esperado de defectos por unidad de inspección producida en el proceso de fabricación.

Para calcular los límites de control de Shewart sabemos que si T es un estadístico muestral que mide la característica de calidad de interés, y la media de T es μ_T, y su desviación estándar, σ_T es , entonces la línea central y los límites superior e inferior del gráfico de control k-sigmas de Shewart son:

$$\text{LSC} = \mu_T + k\sigma_T$$
$$\text{Línea central} = \mu_T$$
$$\text{LIC} = \mu_T - k\sigma_T$$

donde k es la distancia entre los límites de control y la línea central, expresada en desviaciones estándares. Consideramos nuestro estadístico T como $c_i = n_i\, u_i$ que sigue una distribución de Poisson de parámetro $n_i\, u$ cuya media es el propio parámetro $n_i\, u$ y cuya varianza también es el propio parámetro $n_i\, u$. Ya podemos definir los límites de control 3-sigmas de Shewart para cada muestra como:

$$\text{LSC} = n_i u + 3\sqrt{n_i u}$$
$$\text{Línea central} = n_i u$$
$$\text{LIC} = n_i u - 3\sqrt{n_i u}$$

Mientras el número de defectos $c_i = n_i\, u_i$ quede entre los límites de control para cada muestra y la sucesión de puntos definida por $c_i = n_i\, u_i$ para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control, al nivel medio de número de defectos por unidad u. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de unidades defectuosas del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Para calcular los límites probabilísticos de control al nivel α utilizamos la distribución del estadístico $c_i = n_i\, u_i$ que sigue una distribución de Poisson de parámetro $n_i\, u$, con lo que sabemos que para cada muestra:

$$\frac{c_i - n_i \mu}{\sqrt{n_i \mu}} \rightarrow N(0,1)$$

Por lo tanto, mediante la distribución normal N(0, 1), para un α dado, podremos calcular k_α tal que:

$$P\left(-k_\alpha \leq \frac{c_i - n_i\mu}{\sqrt{n_i\mu}} \leq k_\alpha\right) = 1-\alpha$$

Pero de esta expresión se deduce que

$$F_{N(0,1)}(k_\alpha) - F_{N(0,1)}(-k_\alpha) = 1-\alpha \Rightarrow F_{N(0,1)}(k_\alpha) - (1 - F_{N(0,1)}(k_\alpha)) = 1-\alpha \Rightarrow$$
$$2F_{N(0,1)}(k_\alpha) - 1 = 1-\alpha \Rightarrow F_{N(0,1)}(k_\alpha) = 1-\alpha/2$$

con lo que k_α valdrá $F^{-1}(1 - \alpha/2)$, siendo F la función de distribución de la normal N(0, 1).

Como para cada muestra se cumple que:

$$P\left(-k_\alpha \leq \frac{c_i - n_i\mu}{\sqrt{n_i\mu}} \leq k_\alpha\right) = 1-\alpha$$

ya tenemos los límites probabilísticos al nivel α para cada muestra de la forma siguiente:

$$\text{LSCP} = n_i\mu + k_\alpha\sqrt{n_i\mu}$$

$$\text{Línea central} = n_i\mu$$

$$\text{LICP} = n_i\mu - k_\alpha\sqrt{n_i\mu}$$

Sabemos que se acostumbra a reemplazar k_α por 3, que aproximadamente equivale a $\alpha = 0.002$, de manera que se usan límites de tres sigmas y se obtienen como caso particular los diagramas de Shewart.

El tamaño de cada muestra es n_i para i = 1,..., k , y una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán líneas rectas. Una opción útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i = 1,...,k($n = \Sigma ni/k$). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos

límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

B) Diagrama «c» sin patrón dado

Supongamos ahora que no se conoce el número esperado u de defectos por unidad de inspección producida en el proceso de fabricación. Para construir el diagrama de control del número de defectos hay que estimar u, para lo que se toman k muestras subsecuentes de n_i unidades (subgrupos), se calcula en cada una el número de defectos por unidad u_i y se considera el estadísdico $\bar{u}$ definido como la media de los u_i. Este estadístico

$$\bar{u} = \sum_{i=1}^{k} u_i$$

será el estimador de u.

Para calcular los límites de control de Shewart sustituimos u por su estimador $\bar{u}$, obteniendo los límites de control 3-sigmas para cada muestra como:

$$\text{LSCP} = n_i \bar{\mu} + k_\alpha \sqrt{n_i \bar{\mu}}$$
$$\text{Línea central} = n_i \bar{\mu}$$
$$\text{LICP} = n_i \bar{\mu} - k_\alpha \sqrt{n_i \bar{\mu}}$$

Mientras el número de defectos para cada muestra (subgrupo) quede entre los límites de control y la sucesión de puntos definida por los números de defectos para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de defectos del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Ya que el tamaño de cada muestra es n_i para i = 1,...,k , una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán lineas rectas. Una opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño todas las muestras la media de los n_i para i = 1 ,...,k (n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos

límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema reduce.

No obstante, cuando los tamaños de los subgrupos o muestras n_i son muy distintos, suelen tomarse para cada muestra los siguientes límites de control k-sigmas:

$$\text{LSCP} = n_i\,\bar{\mu} + k_\alpha \sqrt{n_i\,\bar{\mu}}$$
$$\text{Línea central} = n_i\,\bar{\mu}$$
$$\text{LICP} = Max\left(n_i\,\bar{\mu} - k_\alpha \sqrt{n_i\,\bar{\mu}}\,, 0\right)$$

También es usual tomar $\bar{\mu}$ como la media de los u_i de todas las muestras, pero ponderada por los tamaños muestrales n_i, es decir:

$$\bar{u} = \frac{\sum_{i=1}^{k} n_i u_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} c_i}{\sum_{i=1}^{k} n_i}$$

La notación utilizada en el caso más general es la siguiente:

u = número esperado de defectos por unidad de inspección producida en el proceso
u_i = número de defectos por unidad de inspección en el iésimo subgrupo (muestra)
c_i = número de defectos en el iésimo subgrupo
n_i = número de unidades en el iésimo subgrupo
$\bar{\mu}$ = media del número de defectos por unidad de inspección en el proceso
k = número de subgrupos

Para realizar el gráfico de control c, en el eje de abscisas se colocan los números de la muestra y en el eje de ordenadas se colocan los números de defectos por muestra $c_i = n_i\,u_i$.

Para calcular los ***límites probabilísticos de control*** podemos utilizar la función de distribución de la variable de Poisson c_i o también la variable gamma que la aproxima de la siguiente forma.

$$\text{P}(c_i < \text{LICP}) = \text{P}\left(\chi^2_{2(LICP+1)} > 2\,n_i\,\bar{\mu}\right)$$

Entonces podemos calcular el límite probabilístico inferior de control LICP al nivel α mediante la expresión:

$$P(\chi^2_{2(LICP+1)} > 2\, n_i\, \bar{\mu}) = \alpha/2$$

Alternativamente, para hallar el límite superior de control utilizamos la relación entre la distribución binomial y la beta de la siguiente forma:

$$P(c_i > LSCP) = 1 - P(c_i \leq LSCP) = 1 - P(\chi^2_{2(LSCP+1)} \geq 2\, n_i\, \bar{\mu})$$

Entonces podemos calcular el límite probabilístico superior de control LSCP al nivel α mediante la expresión:

$$1 - P(\chi^2_{2(LSCP+1)} \geq 2\, n_i\, \bar{\mu}) = \alpha/2$$

Curva característica de operación para el diagrama «c»

La curva característica de operación del diagrama de control del número de defectos es una representación gráfica de la probabilidad de aceptar incorrectamente la hipótesis de un control estadístico (es decir, un error tipo II o β) contra el número de defectos del proceso. La curva proporciona una medida de la sensibilidad del diagrama de control; esto es, su capacidad para detectar un cambio en el número de defectos del proceso desde su valor nominal $n_i\, \bar{\mu}$ hacia algún otro valor $n_i\, u$. La probabilidad del error tipo II para el diagrama de control del número de defectos puede calcularse a partir de:

$$\beta = P\{c_i \leq LSC | n_i u\} - P\{c_i < LIC | n_i u\}$$

Como c_i es una variable aleatoria de Poisson con parámetro $n_i\, u$, β puede obtenerse a partir de la función de distribución de la variable de Poisson.

No obstante, es más general y preciso el uso de la relación entre la distribución de Poisson y la gamma, que permite calcular β de la forma siguiente:

$$\begin{aligned}
&\beta = P\{c_i \leq LSC | n_i u\} - P\{c_i < LIC | n_i u\} = \\
&P\{c_i < LSC | n_i u\} + P\{c_i = LSC | n_i u\} - P\{c_i < LIC | n_i u\} = \\
&P(\chi^2_{2(LSCP+1)} > 2\, n_i u) + P\{c_i = LSC | n_i u\} - P(\chi^2_{2(LICP+1)} > 2\, n_i u)
\end{aligned}$$

En la curva característica de operación suele situarse β en el eje de ordenadas y el número de defectos c_i en el eje de abscisas.

Índice

3.2.4. DIAGRAMA DE CONTROL DEL NÚMERO DE DEFECTOS POR UNIDAD O DIAGRAMA «U»

Es posible basar un diagrama de control en el número de defectos por unidad de inspección en vez de en el número total de defectos. Si se considera el número de defectos por unidad de inspección producida en el proceso se obtienen los ***diagramas u*** o diagramas de control del número de defectos por unidad.

A) Diagrama «u» con patrón dado

Supongamos que se conoce el número esperado u de defectos por unidad de inspección producida en el proceso de fabricación, o bien que la administración le asigna un valor estándar. Para construir el diagrama de control del número de defectos por unidad se toman muestras subsecuentes de n_i unidades (subgrupos), se calcula en cada una el número de defectos por unidad u_i, y se grafica u_i, en el diagrama para cada muestra (en el eje de abscisas se colocan los números de muestra y en el eje de ordenadas se colocan los números de defectos por unidad de inspección u_i en las muestras). Se considera que c_i es el número total de defectos en el i-ésimo subgrupo, con lo que $u_i = c_i / n_i$ (i=1,2,...,k) siendo k el número de subgrupos.

En la mayoría de los casos, la unidad de inspección será una sola unidad del producto, aunque no necesariamente siempre es así. La unidad de inspección es sencillamente una entidad apropiada para registrar los defectos. Podría ser un grupo de 5 artículos, etcétera. El modelo probabilístico adecuado para la ocurrencia del número de defectos en una población grande es la distribución de Poisson. Entonces, podemos suponer que para cada subgrupo el número de defectos c_i sigue una distribución Poisson de parámetro $n_i u$, donde u es el número esperado de defectos por unidad de inspección producida en el proceso de fabricación.

Para calcular los límites de control de Shewart sabemos que si T es un estadístico muestral que mide la característica de calidad de interés, y la media de T es μ_T, y su desviación estándar es σ_T, entonces la línea central y los límites superior e inferior del gráfico de control k-sigmas de Shewart son:

$$\text{LSC} = \mu_T + k\sigma_T$$
$$\text{Línea central} = \mu_T$$
$$\text{LIC} = \mu_T - k\sigma_T$$

donde k es la distancia entre los límites de control y la línea central, expresada en desviaciones estándares. Consideramos nuestro estadístico T como $u_i = c_i / u_i$ que tie-

ne como media μ y como varianza μ/ n_i ya que c_i sigue una distribución de Poisson de parámetro $n_i u$ cuya media es el propio parámetro $n_i u$ y cuya varianza también es el propio parámetro $n_i u$. Ya podemos definir los límites de control 3-sigmas de Shewart para cada muestra como:

$$\text{LSC} = u + 3\sqrt{u / n_i}$$
$$\text{Línea central} = u$$
$$\text{LIC} = u - 3\sqrt{u / n_i}$$

Mientras el número de defectos por unidad u_i quede entre los límites de control para cada muestra y la sucesión de puntos definida por u_i para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control, al nivel medio de número de defectos por unidad u. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de defectos por unidad del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Para calcular los límites probabilísticos de control al nivel α consideramos que u_i tiene como media μ y como varianza μ/ n_i con lo que sabemos que para cada muestra:

$$\frac{u_i - \mu}{\sqrt{\mu / n_i}} \rightarrow N(0,1)$$

Por lo tanto, mediante la distribución normal N(0,1), para un α dado, podremos calcular k_α tal que:

$$P\left(-k_\alpha \leq \frac{u_i - \mu}{\sqrt{\mu / n_i}} \leq k_\alpha\right) = 1 - \alpha$$

Pero de esta expresión se deduce que

$$F_{N(0,1)}(k_\alpha) - F_{N(0,1)}(-k_\alpha) = 1 - \alpha \Rightarrow F_{N(0,1)}(k_\alpha) - (1 - F_{N(0,1)}(k_\alpha)) = 1 - \alpha \Rightarrow$$
$$2F_{N(0,1)}(k_\alpha) - 1 = 1 - \alpha \Rightarrow F_{N(0,1)}(k_\alpha) = 1 - \alpha / 2$$

con lo que k_α valdrá $F^{-1}(1 - \alpha / 2)$, siendo F la función de distribución de la normal N(0, 1).

Como para cada muestra se cumple que:

$$-k_\alpha \le \frac{u_i - \mu}{\sqrt{\mu / n_i}} \le k_\alpha \Leftrightarrow \mu - k_\alpha \sqrt{\mu / n_i} \le u_i \le \mu + k_\alpha \sqrt{\mu / n_i}$$

ya tenemos los límites probabilísticos al nivel α para cada muestra de la forma siguiente:

$$\text{LSCP} = \mu + k_\alpha \sqrt{\mu / n_i}$$
$$\text{Línea central} = \mu$$
$$\text{LICP} = \mu - k_\alpha \sqrt{\mu / n_i}$$

Sabemos que se acostumbra a reemplazar por 3, que aproximadamente equivale a $\alpha = 0.002$, de manera que se usan límites de tres sigmas y se obtienen como caso particular los diagramas de Shewart.

El tamaño de cada muestra es n_i para i = 1,..., k , y una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán líneas rectas. Una opción útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño de todas las muestras la media de los n_i para i = 1,...,k($n=\Sigma ni/k$). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema se reduce.

B) Diagrama «u» sin patrón dado

Supongamos ahora que no se conoce el número esperado u de defectos por unidad de inspección producida en el proceso de fabricación. Para construir el diagrama de control del número de defectos hay que estimar u, para lo que se toman k muestras subsecuentes de n_i unidades (subgrupos), se calcula en cada una el número de defectos por unidad u_i y se considera el estadísfico definido como la media de los u_i. Este estadístico

$$\bar{u} = \sum_{i=1}^{k} u_i$$

será el estimador de u.

Para calcular los límites de control de Shewart sustituimos u por su estimador $\bar{u}$, obteniendo los límites de control 3-sigmas para cada muestra como:

$$\text{LSCP} = \bar{\mu} + k_\alpha \sqrt{\bar{\mu} / n_i}$$
$$\text{Línea central} = \bar{\mu}$$
$$\text{LICP} = \bar{\mu} - k_\alpha \sqrt{\bar{\mu} / n_i}$$

Mientras el número de defectos por unidad para cada muestra (subgrupo) quede entre los límites de control y la sucesión de puntos definida por los números de defectos por unidad para cada muestra no exhiba un patrón sistemático, se concluye que el proceso está bajo control. Si quedan puntos fuera de los límites de control, o si se observa un patrón no aleatorio entre los puntos, habrá que concluir que el número de defectos del proceso cambió hacia un nuevo nivel y que el proceso está fuera de control.

Ya que el tamaño de cada muestra es n_i para i = 1,...,k , una vez que tenemos límites de control para cada muestra, podemos optar por graficarlos tal y como se obtienen, con lo que las líneas de control no serán líneas rectas. Una opción, útil cuando los tamaños n_i no difieren mucho, es tomar como valor común n del tamaño todas las muestras la media de los n_i para i = 1 ,...,k (n = Σni/k). También se puede optar por tomar un n común e igual al mayor valor de los n_i, con lo que obtendríamos unos límites de control bastante sensibles, ya que la amplitud de la franja que indica proceso en estado de control es inversamente proporcional al tamaño de la muestra. Pero en la mayoría de los casos n suele ser igual para todas las muestras, con lo que el problema reduce.

No obstante, cuando los tamaños de los subgrupos o muestras n_i son muy distintos, suelen tomarse para cada muestra los siguientes límites de control k-sigmas:

$$LSCP = \bar{\mu} + k_\alpha \sqrt{\bar{\mu} / n_i}$$
$$Linea \quad Central = \bar{\mu}$$
$$LICP = Max(\bar{\mu} - k_\alpha \sqrt{\bar{\mu} / n_i})$$

También es usual tomar $\bar{u}$ como la media de los u_i de todas las muestras, pero ponderada por los tamaños muestrales n_i, es decir:

$$\bar{u} = \frac{\sum_{i=1}^{k} n_i u_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} c_i}{\sum_{i=1}^{k} n_i}$$

La notación utilizada en el caso más general es la siguiente:

u = número esperado de defectos por unidad de inspección producida en el proceso
u_i = número de defectos por unidad de inspección en el iésimo subgrupo (muestra)
c_i = número de defectos en el iésimo subgrupo
n_i = número de unidades en el iésimo subgrupo
$\bar{u}$ = media del número de defectos por unidad de inspección en el proceso
k = número de subgrupos

Para realizar el gráfico de control u, en el eje de abscisas se colocan los números de la muestra y en el eje de ordenadas se colocan los números de defectos por unidad para cada muestra u_i.

Para calcular los ***límites probabilísticos de control*** podemos utilizar la función de distribución de la variable de Poisson c_i o también la variable gamma que la aproxima de la siguiente forma:

$$P(u_i < LICP) = P(c_i < n_i LICP) = P(\chi^2_{2n_i(LICP+1)} > 2\, n_i\, \bar{\mu})$$

Entonces podemos calcular el límite probabilístico inferior de control LICP al nivel α mediante la expresión:

$$P(\chi^2_{2n_i(LICP+1)} > 2\, n_i\, \bar{\mu}) = \alpha/2$$

Alternativamente, para hallar el límite superior de control utilizamos la relación entre la distribución binomial y la beta de la siguiente forma:

$$P(u_i > LSCP) = 1 - P(u_i \leq LSCP) = 1 - P(c_i \leq n_i LSCP) = 1 - P(\chi^2_{2n_i(LSCP+1)} \geq 2\, n_i\, \bar{\mu})$$

Entonces podemos calcular el límite probabilístico superior de control LSCP al nivel α mediante la expresión:

$$1 - P(\chi^2_{2n_i(LSCP+1)} \geq 2\, n_i\, \bar{\mu}) = \alpha/2$$

Curva característica de operación para el diagrama «u»

La curva característica de operación del diagrama de control del número de defectos es una representación gráfica de la probabilidad de aceptar incorrectamente la

hipótesis de un control estadístico (es decir, un error tipo II o β) contra el número de defectos por unidad del proceso. La curva proporciona una medida de la sensibilidad del diagrama de control; esto es, su capacidad para detectar un cambio en el número de defectos del proceso desde su valor nominal $\bar{u}$ hacia algún otro valor u. La probabilidad del error tipo II para el diagrama de control del número de defectos por unidad puede calcularse a partir de:

$$\beta = P\{u_i \leq LSCIu\} - P\{u_i < LICIu\} = P\{c_i \leq n_i\, LSCIu\} - P\{c_i < n_i\, LICIu\}$$

Como c_i es una variable aleatoria de Poisson con parámetro $n_i\, u$, β puede obtenerse a partir de la función de distribución de la variable de Poisson.

No obstante, es más general y preciso el uso de la relación entre la distribución de Poisson y la gamma, que permite calcular β de la forma siguiente:

$$\beta = P\{c_i \leq n_i\, LSCIu\} - P\{c_i < n_i\, LICIu\} =$$
$$P\{c_i < n_i\, LSCI\, u\} + P\{c_i = n_i\, LSCI\, u\} - P\{c_i < n_i\, LICI\, u\} =$$
$$P(\chi^2_{2n_i(LSCP+1)} > 2\, n_i\, u) + P\{c_i = n_i\, LSCI\, u\} - P(\chi^2_{2n_i(LICP+1)} > 2\, n_i\, u)$$

En la curva característica de operación suele situarse β en el eje de ordenadas y el número de defectos c_i en el eje de abscisas.

A continuación podemos ver dos ejemplos de aplicación de los gráficos de control por atributos.

EJEMPLO 1

ENUNCIADO: Este ejemplo se aplica a un informe cuatrimestral de la inspección al 100% realizada diariamente, de una característica de calidad crítica de una pieza integrante de un dispositivo eléctrico. Principalmente, se intenta dar una exposición de la forma en que se calculan los límites de control con una producción variable, así como del establecimiento y revisión de los valores estándar de la fracción rechazada.

Cuando, después de un cambio en el diseño, se comenzó la producción de ésta pieza a principios de junio, se calculó la fracción rechazada diaria y se representó en un gráfico. Al cabo del mes, se calculó la fracción rechazada media, $p\,\bar{p}$. Asimismo, se calcularon los límites de control de prueba para cada punto. Se estableció un valor estándar de la fracción rechazada p'_o, para su aplicación a la producción futura. Durante el mes de julio se volvieron a calcular y representar diariamente los límites de control, basándose en el número de piezas n inspeccionadas durante el día. Para el mes de agosto, se estableció un solo conjunto de límites de control, en base a la producción diaria media estimada. A finales de este mes, se obtuvo un valor de p'_o, revisado para aplicarlo en septiembre, y el gráfico de control se mantuvo a lo largo de todo ese mes con este nuevo valor.

CÁLCULO DE LOS LÍMITES DE CONTROL DE PRUEBA. La tabla muestra, para cada día del mes de junio, el total de piezas inspeccionadas y el número de ellas rechazadas como fuera de especificación. La fracción rechazada diaria está dada por el cociente entre el número de piezas rechazadas y el total de ellas inspeccionadas ese día. Por ejemplo, para el día 6 de junio, $p = 31/3350 = 0.0093$. El tanto por ciento defectuoso es $100p = 0.93\%$.

Al final del mes, se calcula la fracción rechazada media, $p\,\bar{p}$. Debería hacerse hincapié en que la forma correcta de obtener *p media*, consiste en dividir el número total de rechazadas en el periodo entre el número total de piezas inspeccionadas durante el mismo. Siempre que el tamaño de subgrupo no sea constante (en este caso, el número de piezas inspeccionadas cada día), no es correcto promediar los valores de p .

La desviación estándar se calculó en base a este valor medio observado de p , 0.0145. Téngase en cuenta que el valor $3\sqrt{(p\,\bar{p}\cdot(1\text{-}p\,\bar{p}))}$ sólo se calculó una vez y se aplica a la obtención de todos los límites de control. Así, pues, los límites del mes de junio son, $\pm\, 0.3586/\sqrt{3.350}$, o $\pm\, 0.0062$.

En esta Ilustración podemos ver la representación de los valores diarios p y de los límites de control.

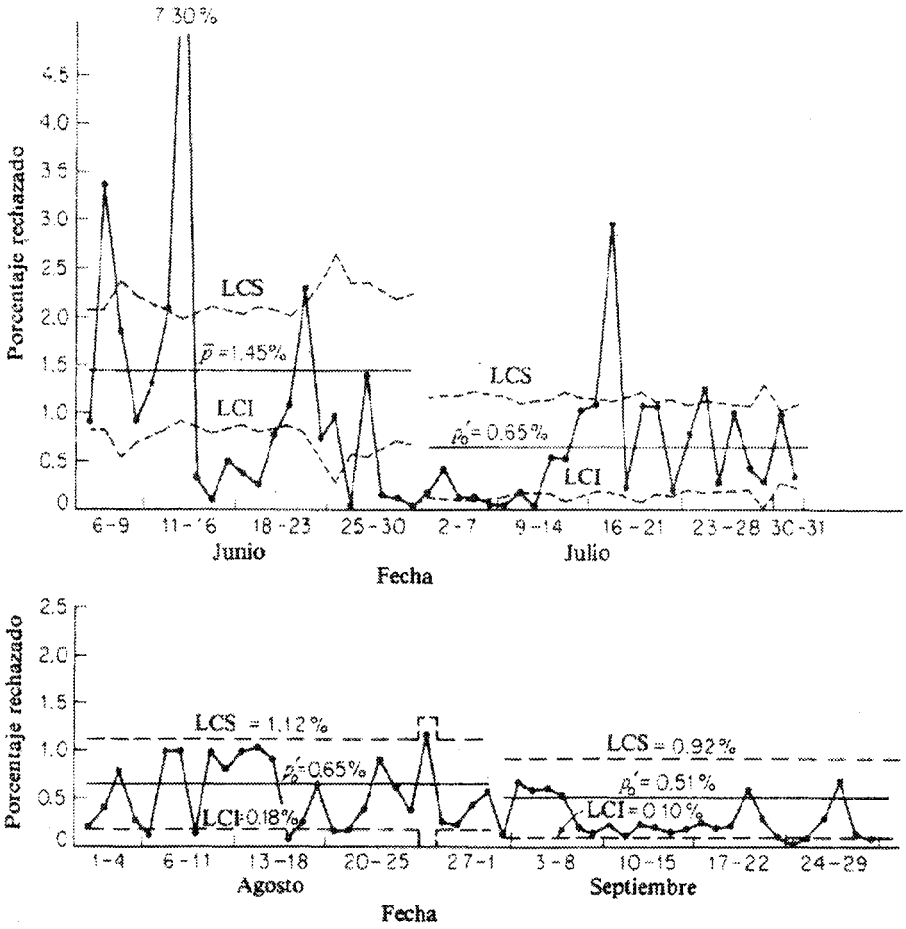


ILUSTRACIÓN 26: Gráficos por atributos

En la tabla 8 podemos ver los resultados de los distintos cálculos para la obtención de los gráficos.

Determinación del valor estándar, p'°. Si todos los puntos cayeran dentro de los límites de control de prueba, el valor estándar podría suponerse igual a la media.

En este ejemplo muchos puntos salieron fuera de dichos límites. En tales casos, la decisión del valor de p'° que se emplea exige conocer que fracción rechazada media puede mantenerse en el futuro, siempre que puedan ser eliminadas las causas atribuibles ocasionales a la mala calidad. Una forma de obtener este valor consiste en calcular un $p\,\bar{p}$ revisado, sin considerar los días en los que se encontraba por encima de límite de control superior. Después de estas consideraciones, y de la experiencia previa con piezas de diseño ligeramente diferente, se decidió suponer un p'° = 0.0065.

TABLA 8: Cálculos de los gráficos de control

Fecha	Número inspeccionado n	Número de artículos rechazados	Fracción rechazada p	$3\sigma = \frac{3\sqrt{\bar{p}(1-\bar{p})}}{\sqrt{n}}$	LCS $\bar{p} + 3\sigma$	LCI $\bar{p} - 3\sigma$
Junio 6	3 350	31	0.0093	0.0062	0.0207	0.0083
7	3 354	113	0.0337	0.0062	0.0207	0.0083
8	1 509	28	0.0186	0.0092	0.0237	0.0053
9	2 190	20	0.0091	0.0077	0.0222	0.0068
11	2 678	35	0.0131	0.0069	0.0214	0.0076
12	3 252	68	0.0209	0.0063	0.0208	0.0082
13	4 641	339	0.0730	0.0053	0.0198	0.0092
14	3 782	12	0.0032	0.0058	0.0203	0.0087
15	2 993	3	0.0010	0.0066	0.0211	0.0079
16	3 382	17	0.0050	0.0062	0.0207	0.0083
18	3 694	14	0.0038	0.0059	0.0204	0.0086
19	3 052	8	0.0026	0.0065	0.0210	0.0080
20	3 477	27	0.0078	0.0061	0.0206	0.0084
21	4 051	44	0.0109	0.0056	0.0201	0.0089
22	3 042	70	0.0230	0.0065	0.0210	0.0080
23	1 623	12	0.0074	0.0089	0.0234	0.0056
25	915	9	0.0098	0.0119	0.0264	0.0026
26	1 644	1	0.0006	0.0088	0.0233	0.0057
27	1 572	22	0.0140	0.0090	0.0235	0.0055
28	1 961	3	0.0015	0.0081	0.0226	0.0064
29	2 440	3	0.0012	0.0073	0.0218	0.0072
30	2 086	1	0.0005	0.0079	0.0224	0.0066
Totales	60 688	880				

$$\bar{p} = \frac{\text{Número total de rechazados}}{\text{Número total inspeccionado}} = \frac{880}{60\,688} = 0.0145$$

$$3\sqrt{\bar{p}(1-\bar{p})} = 3\sqrt{(0.0145)(0.9855)} = 0.3586$$

CÁLCULO DE LOS LÍMITES DE CONTROL BASADOS EN LA FRACCIÓN RECHAZADA ESTÁNDAR, p'_0. La tabla siguiente indica la cantidad de piezas inspeccionadas y rechazadas, diariamente, durante el mes de julio, y muestra el cálculo de los límites de control basados en la fracción rechazada estándar, el cual es casi idéntico al realizado en la tabla anterior.

La diferencia práctica está en que cuando se emplea $p\,\bar{p}$, no se pueden calcular los límites hasta que se conoce dicho valor; es decir, hasta el final del periodo. Cuando se establece de antemano un valor estándar, p'_0, los límites pueden calcularse cada día,y representarse en el gráfico de control mediante el punto correspondiente al día en cuestión. De esta manera, el gráfico de control proporciona un criterio para la acción inmediata siempre que se produzca un punto fuera de los límites de control.

TABLA 9: Cálculos de límites de control

Fecha	Número inspeccionado n	Núm. de artículos rechazados	Fracción rechazada p	$3\sigma = \frac{3\sqrt{p'_0(1-p'_0)}}{\sqrt{n}}$	LCS $p'_0 + 3\sigma$	LCI $p'_0 - 3\sigma$
Julio 2	2 228	4	0.0018	0.0051	0.0116	0.0014
3	2 087	9	0.0043	0.0053	0.0118	0.0012
5	2 088	3	0.0014	0.0053	0.0118	0.0012
6	1 746	2	0.0011	0.0058	0.0123	0.0007
7	2 076	1	0.0005	0.0053	0.0118	0.0012
9	2 164	1	0.0005	0.0052	0.0117	0.0013
10	2 855	5	0.0018	0.0045	0.0110	0.0020
11	2 560	5	0.0020	0.0048	0.0113	0.0017
12	2 545	14	0.0055	0.0048	0.0113	0.0017
13	1 874	1	0.0005	0.0056	0.0121	0.0009
14	2 329	24	0.0103	0.0050	0.0115	0.0015
16	2 744	30	0.0109	0.0046	0.0111	0.0019
17	2 619	77	0.0294	0.0047	0.0112	0.0018
18	2 211	5	0.0023	0.0051	0.0116	0.0014
19	1 746	19	0.0109	0.0058	0.0123	0.0007
20	2 628	28	0.0107	0.0047	0.0112	0.0018
21	2 366	5	0.0021	0.0050	0.0115	0.0015
23	2 954	23	0.0078	0.0044	0.0109	0.0021
24	2 586	32	0.0124	0.0047	0.0112	0.0018
25	2 790	8	0.0029	0.0046	0.0111	0.0019
26	2 968	30	0.0101	0.0044	0.0109	0.0021
27	3 100	13	0.0042	0.0043	0.0108	0.0022
28	1 359	4	0.0029	0.0065	0.0130	0.0000
30	3 940	39	0.0099	0.0038	0.0103	0.0027
31	3 138	11	0.0035	0.0043	0.0108	0.0022
Totales	61 701	393				

La fracción rechazada estándar p'_0 es 0.0065

$3\sqrt{p'_0(1-p'_0)} = 3\sqrt{(0.0065)(0.9935)} = 0.241$

EJEMPLO 2

ENUNCIADO: En la tabla 10 se indican los errores de alineación observados en la inspección final de un determinado modelo de avión. La Ilustración siguiente es la representación del gráfico de control de estas 50 observaciones. Los errores de alineación observados en cada avión constituyen un subgrupo del gráfico 27.

El número total de errores de alineación encontrados en los 25 primeros subgrupos, era de 200. La media $c\ \bar{c} = 200/25 = 8$. Los límites de control de prueba, calculados a partir de este valor medio, son los siguientes:

$$\text{LCS} = c\ \bar{c} + 3\sqrt{c\ \bar{c}} = 8 + 3\sqrt{8} = 16.5$$

$$\text{LCI} = c\ \textit{media} - 3\sqrt{c\ \bar{c}} = 8 - 3\sqrt{8} = \text{negativo, por tanto, no hay LCI}$$

TABLA 10: Cálculos de límites de control

Número del avión	Núm. de errores de alineación	Número del avión	Núm. de errores de alineación
201	7	226	7
202	6	227	13
203	6	228	4
204	7	229	5
205	4	230	9
206	7	231	3
207	8	232	4
208	12	233	6
209	9	234	7
210	9	235	14
211	8	236	18
212	5	237	11
213	5	238	11
214	9	239	11
215	8	240	8
216	15	241	10
217	6	242	8
218	4	243	7
219	13	244	16
220	7	245	13
221	8	246	12
222	15	247	9
223	6	248	11
224	6	249	11
225	10	250	8
Total.....	200	Total.....	236

Cuando para el límite de control inferior de un gráfico c se obtenga un valor negativo, no se utilizará un límite inferior. El gráfico sólo presentará un límite superior de control.

Como ninguno de los 25 primeros puntos de este gráfico sale fuera de los límites de control de prueba, calculados a partir de los mismos puntos, puede tomarse un número estándar de defectos c'_o, igual a $c\ \bar{c}$, continuando el gráfico de control a lo largo del siguiente periodo, con una línea central de 8 y un límite de control superior de 16.5.

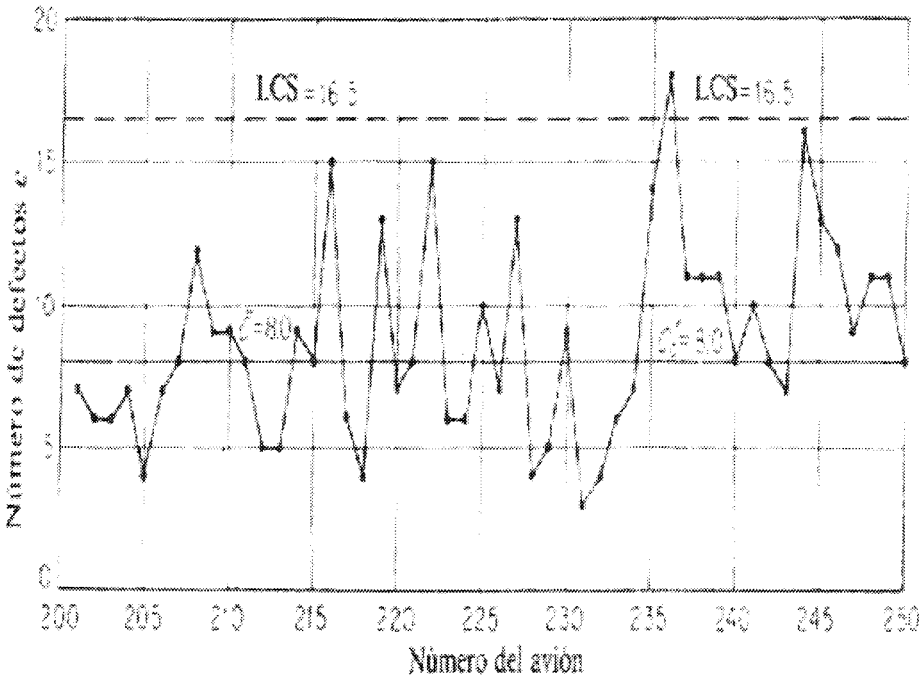


ILUSTRACIÓN 26: Gráficos de control

Del segundo grupo de 25 puntos, uno de ellos (avión número 236) se encuentra por encima del límite de control superior. La media durante este periodo fue 236/25 = 9.44.

(Omitiendo el valor fuera de control, la media es 9.08). De los 16 últimos puntos, correspondientes a los aviones 235 al 250, 12 están por encima del valor estándar c'_o; 3 se encuentran exactamente sobre ese valor, y sólo uno cae por debajo. Parece evidente que se ha producido un empeoramiento ligero, pero definitivo, de la calidad (o un incremento de rigurosidad en la inspección) durante este periodo.

4 ESTUDIO DE CAPACIDAD

4.1 INTRODUCCIÓN

Hoy en día la necesidad del aseguramiento de una determinada calidad, justificando ante el cliente la correcta relación entre capacidad de proceso y capacidad de la máquina en relación con la tolerancia exigida a las piezas, junto con una notable importancia del factor económico hacen del conocimiento y control del proceso una de las armas de diferenciación y ventaja competitiva.

En la fabricación repetitiva, dirigida a obtener lotes o series largas, puede convenir la realización de pruebas estadísticas ocasionales de la calidad de funcionamiento de los equipos de producción. Por su profundidad y duración, estas pruebas se hallan a medio camino entre las de verificación y las de muestreo y control estadístico de procesos. Su evaluación tiene lugar mediante ciertos indicadores y técnicas estadísticas.

La *estabilidad:* es un indicador de la evolución temporal de la distribución, definida por la variación de su posición (deriva) y de la anchura de su dispersión.

Un proceso es estable cuando la posición y la forma de su función de distribución definida por su media ($\overline{x}$) y sus extremos de dispersión (Ds, Di), estén o no dentro del campo de tolerancia, no varían sensiblemente a lo largo del tiempo.

La *capacidad:* cuantifica la relación entre la incertidumbre del proceso y la tolerancia admitida en diseño (Ls, Li). Para evaluar la capacidad se emplean dos parámetros, que reflejan la *estrechez* de la distribución y su *margen* o reserva respecto a los límites de tolerancia.

Un proceso es capaz. o dominado si los limites de su distribucion (D,, D;), se hallan dentro de los limites de tolerancia del producto (Ls, Li).

El término de estabilidad es fácil de explicar, reflejando la distribución en diversos planos temporales. En cambio la capacidad presenta mas matices. Un proceso estable puede ser incapaz, es decir salirse del campo de tolerancia, porque su función de distribución sea ancha (tiene poca estrechez.) o porque este desplazada (queda po-

co margen) hasta uno de los limites de dicho campo. En sentido positivo para que un proceso sea capaz, se deben cumplir simultaneamente dos condiciones:

- debe ser estrecho
- tener margen suficiente hasta los dos limites de tolerancia.

Representación de la estabilidad y capacidad de un proceso

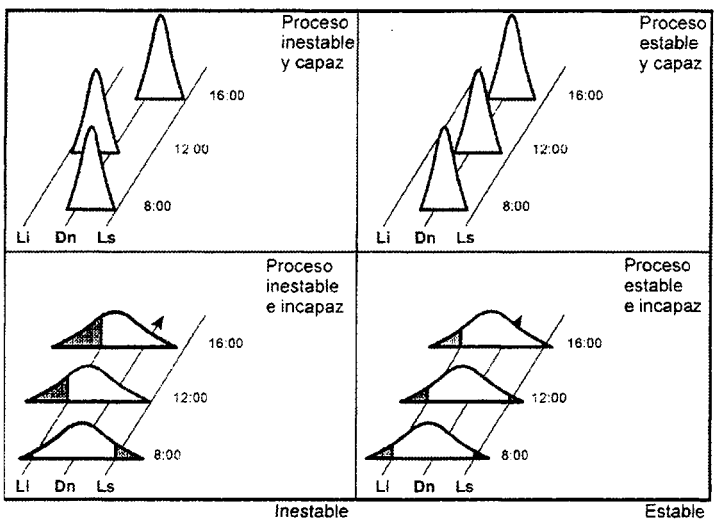


ILUSTRACIÓN 28: Estabilidad y capacidad de un proceso

En la anterior Ilustración, el gráfico del ángulo inferior izquierdo muestra un comportamiento inestable e incapaz El caso del cuadro inferior derecho, es estable pero incapaz, debido a su gran dispersión En el ángulo superior izquierdo, se compara este segundo comportamiento con los límites de tolerancia prefijados por diseño: se observa que es un proceso capaz pero inestable.

Finalmente, en el ángulo superior derecho se presenta un ejemplo de proceso dominado y capaz: se comporta de modo estable y respeta los límites.

¿Cuando se debe hacer un análisis de capacidad?

Debe usarse el análisis de capacidad cuando responda «Si» a las siguientes 3 preguntas:

¿Es estable el sistema? Un sistema inestable está cambiando constantemente, sin embargo la información obtenida de él será inconsistente a través del tiempo. Con una gráfica de control revise primero la estabilidad antes de hacer un estudio de capacidad.

¿Están normalmente distribuídos los valores Individuales? Un análisis de capacidad para distribuciones no-normales se puede realizar con programas de computadora o usando una hoja para análisis de capacidad. Pruebe la normalidad con un Histograma.

¿Desea saber como se comporta el sistema en relación con los límites de especificación? Las gráficas de control (con excepción de la gráfica X-MR) no proporcionan una comparación directa con los límites de especificación, pero tales comparaciones son esenciales para comprender el desempeño completo del sistema, para predecir el desempeño futuro y monitorear las mejoras del sistema

4.2. ÍNDICES DE CAPACIDAD

Los índices de capacidad comparan la capacidad del procesos con las especificaciones técnicas requeridas. La capacidad del proceso (6σ) es una medida de la dispersión natural de la variable que mide la calidad del producto o servicio, pero no dice nada de si dicha calidad se ajusta o no a las especificaciones.

Existen algunos analistas que no recomiendan la utilización de estos índices. Su principal argumento es que resulta un resumen demasiado simplista de la evolución del proceso.

A continuación se muestran los índices más utilizados.

Procesos

Para describir la capacidad de un proceso se emplean dos indices, Cp y Cpk definidos matemáticamente.

- Cp indica la estrechez relativa del proceso respecto a la tolerancia de diseño, igual al séxtuplo de la desviación típica, 6σ; dispuesta simétricamente alrededor del valor medio. Dentro de este campo se encuentran, en un proceso dominado, más del 99 % de los valores. Si la dispersión coincide con la tolerancia (limite superior - limite inferior), resulta Cp = 1. En caso de cotas con tolerancia unilateral como, por ejemplo, en la excentricidad, este parámetro no se puede hallar, por lo que se define un nuevo parametro, Cpk.

C_p = índice de capacidad potencial continua del proceso.

Pretende ver en conjunto, si un proceso esta dentro de los límites, sin tener en cuenta el descentrado de las muestras del valor nominal. Estos límites de capacidad los podemos observar en las siguiente Ilustración.

$$C_p = \frac{LTS - LTI}{6\sigma}$$

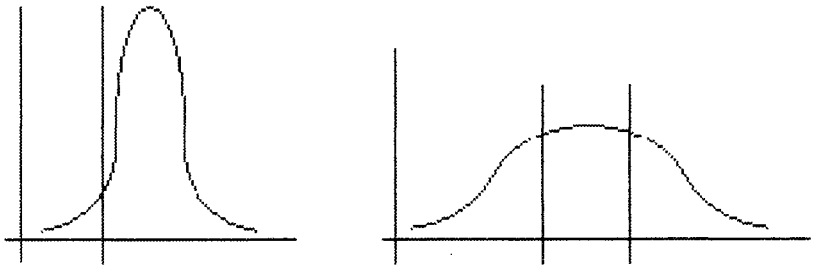


ILUSTRACIÓN 29: Límites de capacidad

Si:

$C_p > 1$ CAPAZ

$C_p = 1$ ESTRICTAMENTE CAPAZ

$C_p < 1$ NO CAPAZ

Es frecuente utilizar el valor Cp = 4/3 ≈ 1, 33 como límite inferior de la calidad que debetenerse en la práctica. Esto implica que:

$$C_p = \frac{LTS - LTI}{6\sigma} = \frac{4}{3} \Rightarrow \text{LTS} - \text{LTI} = 8\sigma$$

Por tanto, serían defectuosos aquellos artículos que estén a más de 4σ de la media; esto es aproximadamente, 64 piezas por millon (bajo normalidad). Por esta razón se dice que:

- Si Cp < 1 → (el proceso no es capaz)
- Si Cp > 1.33 → (el proceso es capaz)
- Si 1 ≤ Cp ≤ 1.33 → (el proceso es capaz pero requiere un seguimiento muy estricto)
- C_{pk} es el *margen relativo de la distribucion*. Es la relación entre la menor distancia del centro de la distribución (posición del proceso) al limite del campo de tolerancia mas próximo, dividida por la mitad de la dispersión del proceso (3σ).

Índice de capacidad real continua del proceso. Este valor es igual a Cp si el proceso esta centrado. Si hay desviación del centro del proceso respecto al centre del campo de tolerancias, Cpk resultara inferior a Cp.

El C_{pk} ha sido un índice para medir la capacidad de un proceso usado por muchos años. Éste índice utiliza solamente un estimador de sigma para medir la variación. Mientras éste estimador sea aceptable, el estimador de sigma puede ser modificado artificialmente para hacerlo más bajo dependiendo del tamaño del subgrupo, intervalo del muestreo, o la planeación del muestreo (su representación en la Ilustración siguiente).

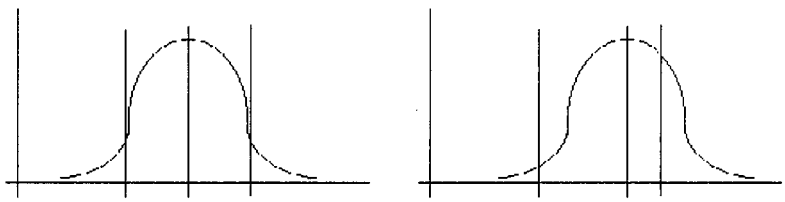


ILUSTRACIÓN 30: Gráfico de rangos de índices

Si Cpk = 1, podemos decir que nuestro proceso es capaz de producir 99.73% dentro de especificaciones, se dice que tenemos una calidad de 3 Sigma.

Si Cpk < 1, nuestro proceso no es capaz de cumplir las especificaciones, se debe reducir la variación.

Si Cpk > 1, podemos decir que nuestro proceso es capaz de producir cuando menos 99.73% de producto dentro de especificaciones.

Si Cpk = 1.33 decimos que tenemos una calidad de 4 Sigma y nuestro proceso es capaz de producir 99.99% de producto dentro de especificaciones.

Si Cpk = 2, decimos que tenemos una calidad de 6 Sigma y nuestro proceso es capaz de producir 99.99999% de producto dentro de especificaciones.

$$C_{pk1} = \frac{T_s - \bar{x}}{3 \cdot \sigma}$$

$$C_{pk2} = \frac{\bar{x} - T_i}{3 \cdot \sigma}$$

El menor valor de los 2, será el más estricto, por lo tanto lo tendremos más en cuenta.

Ejemplo de cálculo de los valores de la capacidad de proceso

El proceso es capaz si ambos índices son superiores a 1. Si C_{pk} es inferior a C_p, el proceso puede volverse capaz desplazándolo hacia el centro. Solo se calcula la capacidad de bajo control estadístico. La exigencia de una capacidad superior a 1 in-

dica una producción que aprovecha toda la tolerancia. Cualquier pequeño desplazamiento del proceso origina piezas defectuosas, por lo que los valores altos revelan la existencia de reservas de capacidad. Así, la primera distribución tiene una desviación típica de 7/3 y su eje esta a una distancia del borde inferior de 5 unidades. Ambos parámetros son iguales (caso centrado, como en el cuarto caso). El valor C_{pk} =0,71 inferior a 1, revela que la distribución desborda los limites, lo que no ocurre en el cuarto caso, con C_{pk} = 5. El segundo y el tercero se parecen (están descentrados y entre limites), pero el segundo es menos disperse y mas descentrado.

Respecto a todo esto, podemos verlo de forma más clara en la Ilustración siguiente, como va variando la forma de la curva.

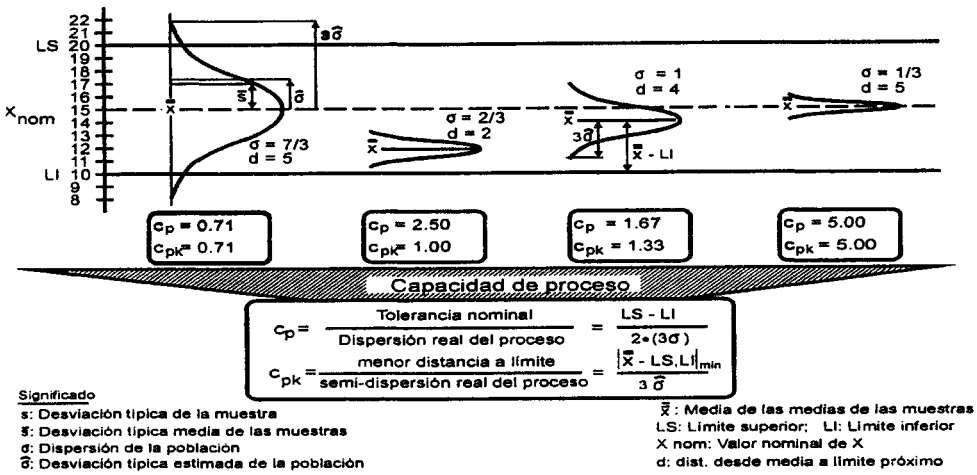


ILUSTRACIÓN 31: Cálculo de valores de capacidad

4.3. CAPACIDAD Y POTENCIAL DE PROCESO Y MÁQUINA

Para realizar un estudio de capacidad de un determinado proceso, se necesita conocer el tamaño de la muestra, el material que se produce, los aparatos de medida. Estos estudios se aplican cuando las máquinas son nuevas y cuando se tienen máquinas con averías.

Los resultados obtenidos del estudio, dependerán del valor de la media y de la variabilidad (desviación típica).Con los valores de la media, recorrido y la desviación típica, podemos realizar ciertas valoraciones del muestreo realizado:

Si desviación típica sale muy alta, una posible propuesta de mejora, sería hacer un nuevo muestreo, eliminando la dispersión de los datos ajustando la máquina.

Los estudios de capacidad, se realizan a un coste bajo y pueden ser útiles para evaluar las situaciones presentes y futuras, debiéndose incluir en la base de datos sobre el desempeño pasado.

Como indicadores de calidad de un proceso, la norma Ford, propone cuatro parámetros, que combinan la *estrechez* y el *margen relativo* de la distribución con las condiciones de *repetibilidad* (pruebas cortas) y *reproducibilidad* (pruebas largas): Ambos tipos se detallan en la tabla siguiente y gozan de amplia difusión en los contratos entre proveedores y compradores. A estos últimos les interesan los valores altos.

Capacidad del proceso (capacidad en pruebas largas)

Es una prueba similar a la reproducibilidad, y sus parámetros se obtienen mediante varias ejecuciones del proceso a lo largo de mas de 20 días de produccion. Al ser más reales, suelen ser los parámetros mas usados. La empresa Ford exige valores de C_p y C_{pk} superiores a 1,33 para pruebas largas.

Potencial de capacidad (capacidad en pruebas cortas, o de maquina)

El potencial de capacidad en pruebas cortas, sobre 20 o mas muestras, también llamado *capacidad de maquina* o ensayo de recepción es un indicador semejante a la repetibilidad.

La empresa Ford exige valores superiores a 1,67 para pruebas cortas, lo que sólo es posible con un ajuste de la maquina, y un dominio de la misma por parte del operador, mas estrictos que los reclamados durante ensayos mas largo. Los resultados se observan en la tabla siguiente.

TABLA 10: Resumen de los valores de capacidad de máquina y de proceso

Concepto	**Pruebas largas, capacidad, proceso reproducibilidad**		**Pruebas cortas, potencial, máquina repetibilidad**		**Fórmulas**	**Significado físico**
Relación campo de toler./distrib.	Símbolo, ecuación	Valor (Ford)	Símbolo, ecuación	Valor (Ford)		
Estrechez relativa de la distribución	c_p (5.19)	>1,33	p_p (5.21)	>1,67	$\frac{T}{6\hat{\sigma}} = \frac{LS-LI}{6\hat{\sigma}}$	Toler. nominal de diseño / Dispersión real de proceso
Margen relativo de la distribución	c_{pk} (5.20)	>1,33	p_{pk} (5.22)	>1,67	$\left\lvert \frac{LS-\bar{\bar{x}}}{3\hat{\sigma}} ; \frac{\bar{\bar{x}}-LI}{3\hat{\sigma}} \right\rvert$	Mín.desv. real de proceso / Semidispers. real de proceso

4.4. OTROS TIPOS DE ÍNDICES DE CAPACIDAD

A) Índices de capacidad unilaterales CpL y CpU

Estos índices se utilizan cuando los límites de tolerancia son unilaterales. Por ejemplo, el tiempo de realización de una oferta, la tensión de rotura de un material, la temperatura máxima que admite un componente electrónico, etc. En estos casos, la tolerancia se mide desde el valor nominal central hasta el extremo inidicado por la especificación. Esta cantidad se compara con 3σ. Se utiliza CpL cuando existe una especificación mínima, pero no máxima, mientras que CpU se usa cuando se debe cumplir con una especificación máxima.

$$C_{pL} = \frac{\mu_N - LTI}{3\sigma}$$

$$C_{pU} = \frac{LTI - \mu_N}{3\sigma}$$

Cuando cualquiera de estos dos índices es menor que 1, el proceso no cumplirá las especificaciones.

Cuando no se conozca el valor nominal del proceso μ_N, se sustituirá por la mediamuestral.

B) Índice de capacidad recíproco Cr

Este índice es el invero de Cp,es decir, Cr = 1/Cp. Su interpretación es la siguiente:

- Si Cr > 1, → el proceso no es capaz
- Si Cr < 0.75, → el proceso es capaz
- Si $0.75 \leq Cr \leq 1$, → el proceso es capaz pero precisa de un control estricto.

4.5. EJECUCIÓN DE ENSAYOS DE CAPACIDAD Y SU CÁLCULO

La siguiente figura detalla el algoritmo y las formulas para el cálculo de la capacidad de un proceso, en pruebas largas, o del potencial del proceso, en pruebas cortas. Un requisito previo consiste en asegurarse de que el proceso es estable, cuya comprobación se realiza con ayuda de los gráficos de control. Si se observan influencias sistemáticas, el ensayo debe detenerse. En la imagen siguiente podemos ver los distintos pasos de un estudio de capacidad.

Primero se comprueba el tipo de distribución de los datos. Las formulas son válidas para distribuciones normales. Si se presentan otras formas de distribución, la dispersión del proceso se calcula de modo adecuado (por ejemplo, mediante los percentiles de la distribución).

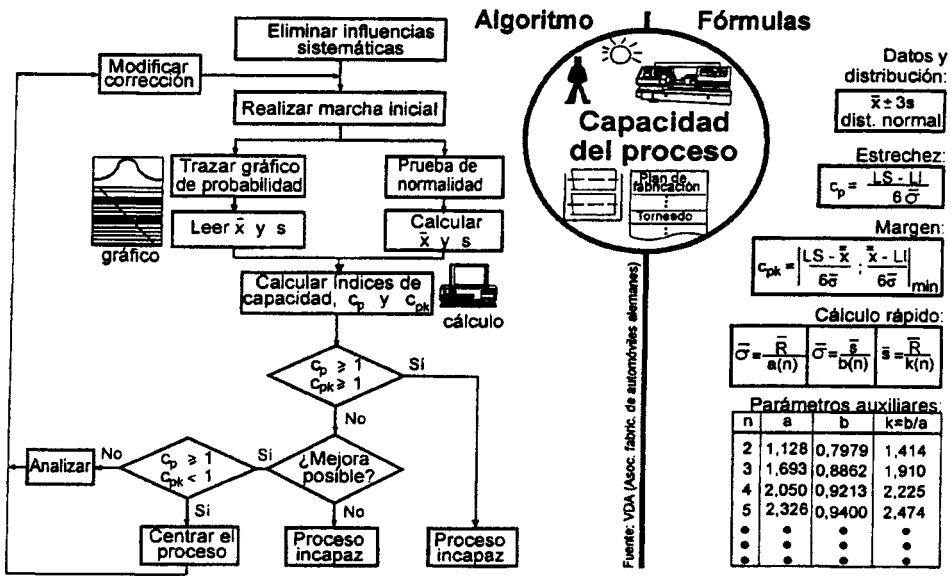


ILUSTRACIÓN 32: Ensayo de capacidad

En la actualidad existen una serie de programas informáticos especializados en cálculos de capacidad que agilizarán y nos ayudarán a determinar y ajustar los equipos y/o procesos frente a las variaciones de demanda del mercado.

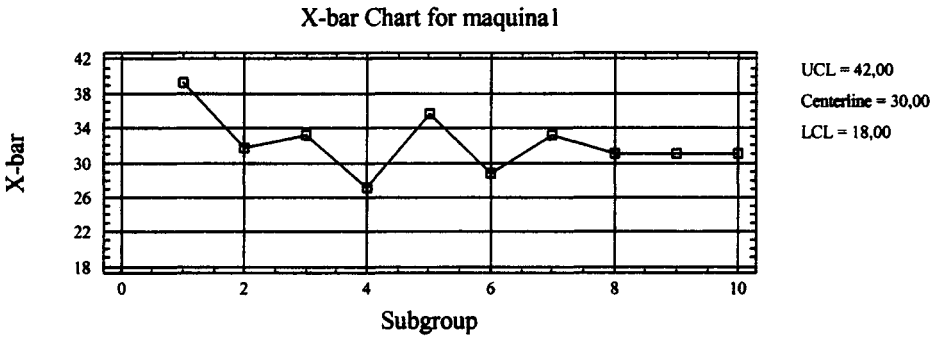
EJEMPLO

Tenemos 3 máquinas cuya finalidad es fabricar tornillos con un valor nominal de 30 mm y con unas tolerancias de ±12 mm. Se toman 10 muestras de cada una de ellas con un tamaño de muestra igual a 5. Se debe analizar la capacidad de cada una de las máquinas.

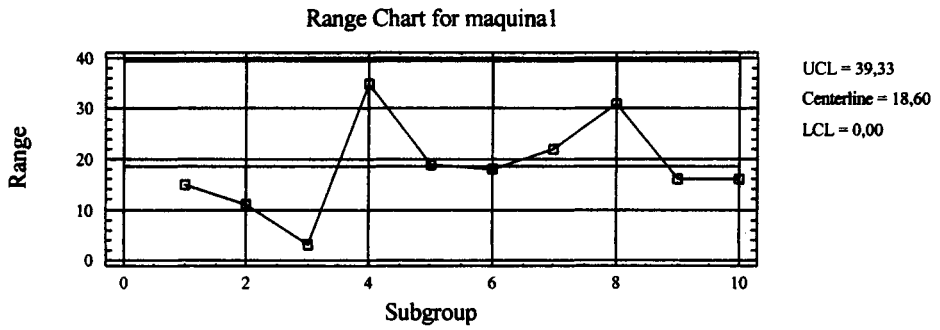
Máquina 1:

Muestra	Datos				
1	47	32	44	35	38
2	33	33	34	24	35
3	34	34	31	34	33
4	12	21	24	47	32
5	35	23	28	40	42
6	19	37	31	27	30
7	23	45	26	37	35
8	33	12	29	43	38
9	25	22	37	33	38
10	25	22	37	33	38

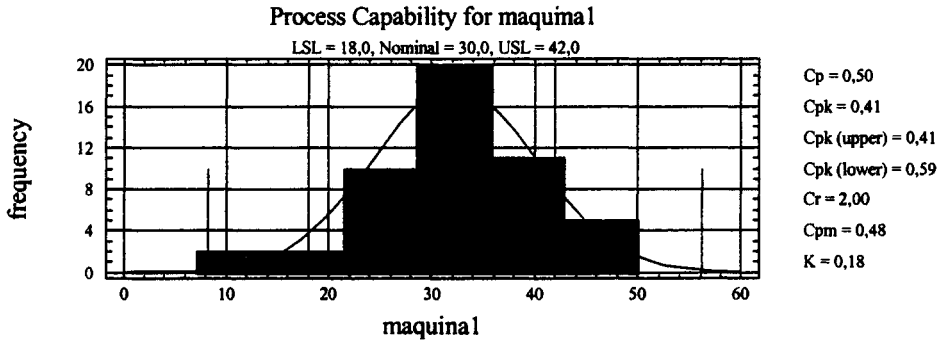
Con estos datos, sacamos la media de cada una de las muestras y las representamos en un gráfico de medias:



Todos los valores los encontramos dentro de los límites admitidos por las tolerancias dadas. También representamos el gráfico de recorrido:



Vemos que en ambos gráficos, todos los puntos se encuentran dentro de los límites, así que podemos decir que el proceso está bajo control. Pero sólo con esto no podemos decidir si la máquina1 es idónea para realizar el trabajo. Debemos comprobar la capacidad del proceso. Para ello, representamos el siguiente gráfico:

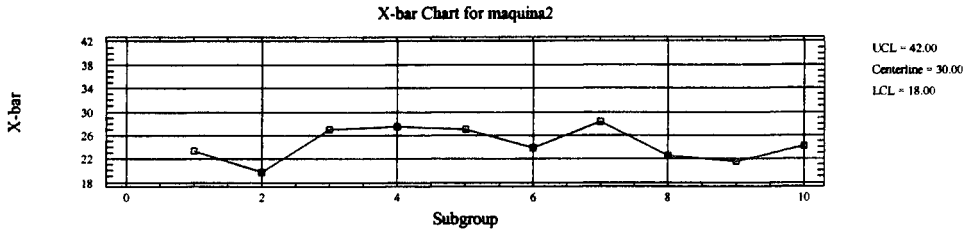


Aquí vemos que pese a que la media de cada muestra estaba dentro de los límites, los valores de gráfica de capacidad se salen fuera de ambos límites, por lo tanto, el proceso no es capaz.

Máquina 2:

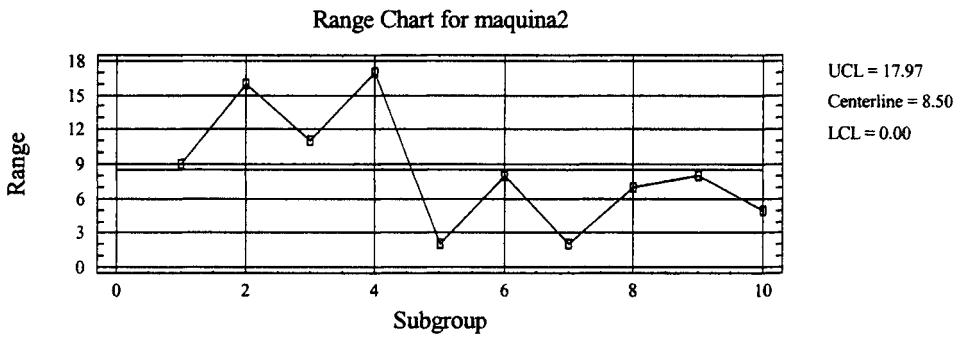
Muestra	Datos				
1	25	26	17	26	23
2	28	26	15	12	18
3	33	31	25	24	22
4	34	24	37	22	20
5	26	27	26	28	28
6	25	25	28	21	20
7	28	29	28	29	27
8	22	26	19	21	25
9	24	22	16	24	22
10	24	21	26	25	25

Al igual que en la máquina anterior, comenzamos representando las medias de los datos de cada muestra:



En esta ocasión, vemos que los datos se sitúan por debajo de la media, aunque están dentro de los límites.

Ahora realizamos el gráfico de recorrido:



En ambos vemos que los puntos están dentro de los límites, pero al igual que en la máquina anterior, no podemos asegurar su capacidad sólo con esto, ya que estas dos gráficas solo nos indican que el proceso está bajo control, por estar todos los valores entre los límites especificados.

Realizamos el gráfico de capacidad:

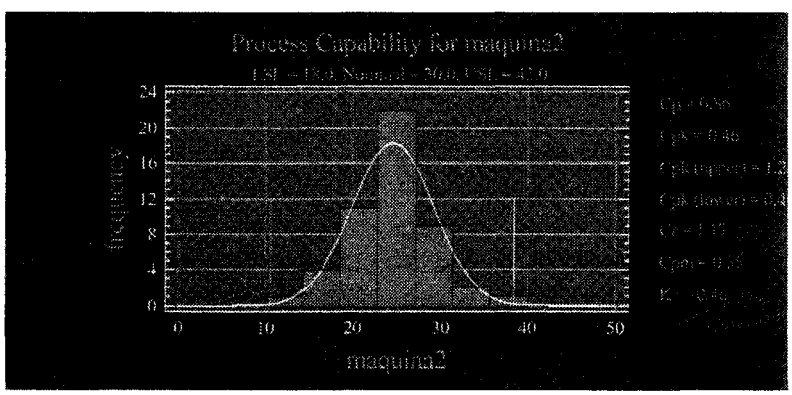


ILUSTRACIÓN 33: Capacidad del proceso

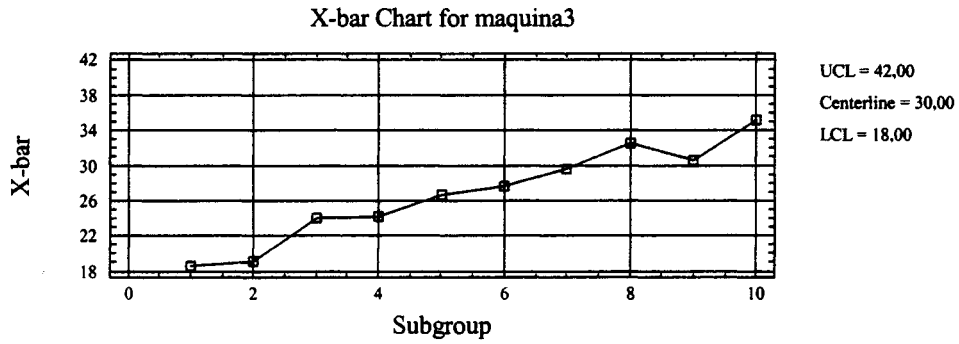
Los valores salen desplazados a la izquierda de la media y límites. El proceso no es capaz por salirse de los límites.

Máquina 3:

Los datos de esta máquina son:

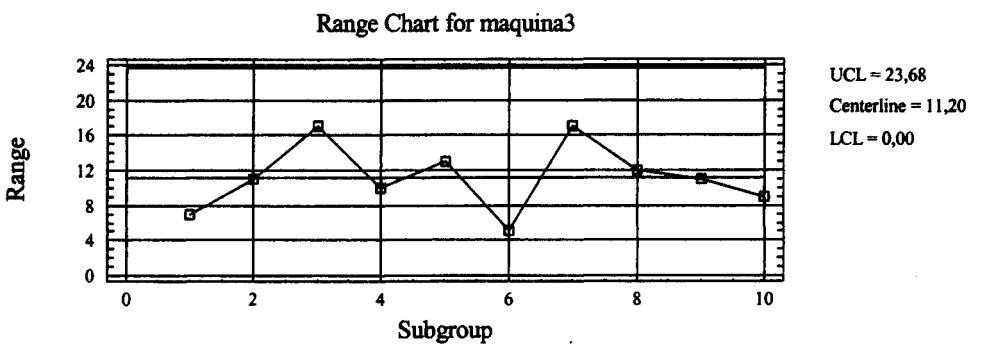
Muestra	Datos				
1	18	15	22	20	18
2	16	13	24	20	22
3	17	25	26	34	18
4	19	25	29	23	25
5	32	24	19	30	28
6	25	26	30	28	29
7	24	23	29	32	40
8	27	30	39	32	35
9	29	37	35	26	26
10	40	38	31	34	33

Hacemos el gráfico de medias:



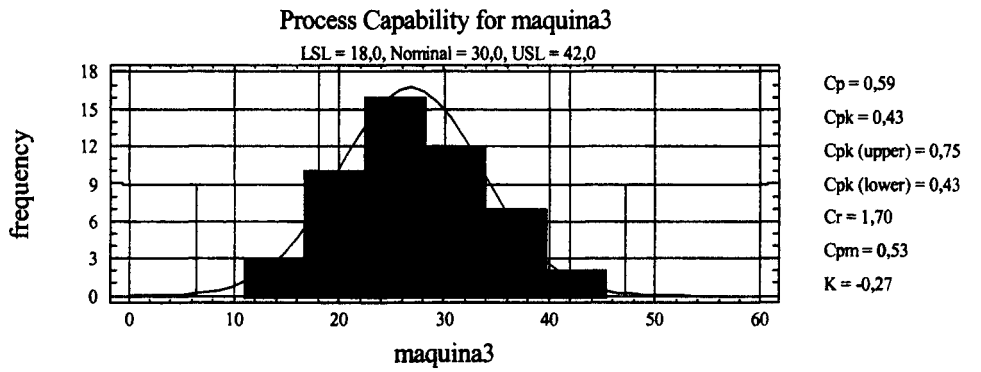
Podemos ver que los valores en las medias siguen una trayectoria ascendente, ha ocurrido un cambio gradual en la producción.

Realizamos el gráfico de recorrido:



Pese a que todos los puntos se encuentran dentro de los límites, la tendencia ascendente de los puntos de las medias, nos da indicaciones de que no tendremos un proceso bajo control.

Debemos realizar el gráfico de capacidad para ver si la máquina es apropiada para nuestro proceso o no:



Como podemos ver en la gráfica, la máquina 3 no es un proceso capaz.

Conclusión:

Si debiéramos elegir una de las tres máquinas propuestas en el problema, escogeríamos la 2, ya que la primera no es capaz por ninguno de los dos lados, la tercera no es capaz por un lado pero posee una tendencia ascendente en las medias no recomendable, y finalmente, la segunda es la que posee unos valores mas ajustados a un mismo punto que aunque no es el de la media, se podría ajustar para que así fuera y pasara a tener un proceso capaz.

5

TÉCNICAS BÁSICAS PARA EL ANÁLISIS Y MEJORA CONTINUA

5.1. IDEAS GENERALES DE MEJORA CONTINUA

Para empezar este capítulo, vamos a conocer la primera idea de mejora la Rueda de Deming, que aparece en la siguiente Ilustración.

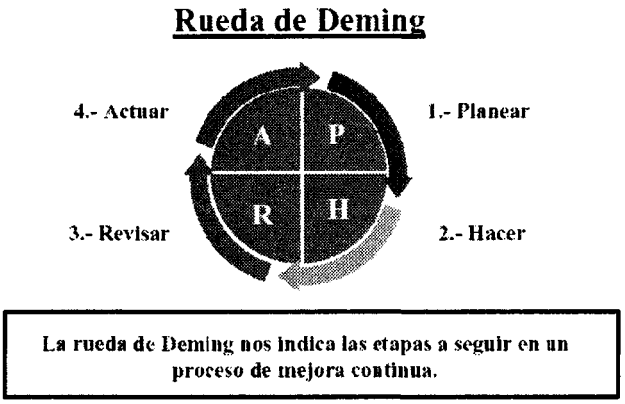


ILUSTRACIÓN 34: Rueda de Deming

Esta es la primera idea que nos introduce en cada una de las siguientes herramientas de la calidad y en las diferentes ideas y conceptos.

5.2. HERRAMIENTAS DE RESOLUCIÓN DE PROBLEMAS Y MEJORA CONTINUA

5.2.1. *BRAINSTORMING* (TORMENTA DE IDEAS)

¿Qué es?

Es una herramienta de trabajo grupal que facilita el surgimiento de nuevas ideas sobre un tema o problema determinado.

La tormenta de ideas (*Brainstorming*), es una técnica de grupo para generar ideas originales en un ambiente relajado.

Esta herramienta fue creada en el año 1941, por Alex Osborne, cuando su búsqueda de ideas creativas resulto en un proceso interactivo de grupo no estructurado que generaba mas y mejores ideas que las que los individuos podían producir trabajando de forma independiente; dando oportunidad de sugerir sobre un determinado asunto y aprovechando la capacidad creativa de los participantes.

¿Cuándo se utiliza?

Se deberá utilizar la lluvia de ideas cuando exista la necesidad de:

- Liberar la creatividad de los equipos
- Generar un numero extensos de ideas
- Involucrar oportunidades para mejorar

Nos permite:

- Plantear y resolver los problemas existentes
- Plantear posibles causas
- Plantear soluciones alternativas
- Desarrollar la creatividad
- Discutir conceptos nuevos
- Superar el conformismo y la monotonía

¿Cómo se utiliza?

- Se define el tema o el problema.
- Se nombra a un conductor del ejercicio
- Antes de comenzar la «tormenta de ideas», explicará las reglas.
- Se emiten ideas libremente sin extraer conclusiones en esta etapa.
- Se listan las ideas
- No se deben repetir
- No se critican
- El ejercicio termina cuando ya no existen nuevas ideas
- Se analizan, evalúan y organizan las mismas, para valorar su utilidad en función del objetivo que pretendía lograr con el empleo de esta técnica.

Modo de uso

La técnica, *Brainstorming*, puede ser empleada a través de 3 diferentes maneras:

No estructurado (flujo libre)

- Escoger a alguien para que sea el facilitador y apunte las ideas
- Escribir en un tablero una frase que represente el problema y el asunto de discusión.
- Escribir cada idea en el menor numero de palabras posible.
- Verificar con la persona que hizo la contribución cuando se este repitiendo la idea.
- No interpretar o cambiar las ideas.
- Establecer un tiempo limite (aproximadamente 25 minutos)
- Fomentar la creatividad
- Construir sobre las ideas de otros.
- Los miembros del grupo de «lluvia de ideas» y el facilitador nunca deben criticar las ideas.
- Revisar la lista para verificar su comprensión.
- Eliminar las duplicaciones, problemas no importantes y aspectos no negociables.
- Llegar a un consenso sobre los problemas que parecen redundantes o no importantes.

Estructurado (en círculo)

Tiene las mismas metas que la tormenta de ideas no estructurada. La diferencia consiste en que cada miembro del equipo presenta sus ideas en un formato ordenado (ej: de izquierda a derecha). No hay problema si un miembro del equipo cede su turno si no tiene una idea en ese instante.

Silenciosa (tormenta de ideas escritas)

Es similar a la lluvia de ideas, los participantes piensan las ideas pero registran en papel sus ideas en silencio. Cada participante pone su hoja en la mesa y la cambia por otra hoja de papel. Cada participante puede entonces agregar otras ideas relacionadas o pensar en nuevas ideas. Este proceso continua por cerca de 30 minutos y permite a los participantes construir sobre las ideas de otros y evitar conflictos o intimidaciones por parte de los miembros dominantes.

Aspectos a tener en cuenta:

Para que una reunión sea útil tiene que estar bien preparada, para ello hay que tener en cuenta los aspectos materiales como los funcionales. Se pueden dividir en 4 pasos:

Definición de los objetivos:

- Fijar los objetivos de la reunión.
- Tipos de objetivos.
- Los que se han de lograr.

Si los objetivos prioritarios son muchos se hacen varias reuniones.

La elección de los participantes:

La eficacia de una reunión dependerá mucho de los participantes. Se tendrá que reunir a las personas compatibles.

Planificar el desarrollo de la reunión:

Vendrá definido en el orden del día que es un desarrollo de las cuestiones que se abordarán.

Organización del material de la reunión. Tenemos que tener en cuenta:

- Los documentos
- Preparación de la sala
- Los aspectos anexos: la botella de agua, papelera, bolígrafos...
- Hacer un recordatorio de la convocatoria
- Asegurarse que asistirán
- Relación con otras herramientas
- Diagrama de afinidad

Es un herramienta que organiza un gran numero de ideas en función de afinidad, es decir, de las relaciones que existen entre ellas.

5.2.2. DIAGRAMA CAUSA - EFECTO

El diagrama causa-efecto es una forma de organizar y representar las diferentes teorías propuestas sobre las causas de un problema. Se conoce también como diagrama de Ishikawa o diagrama de espina de pescado y se utiliza en las fases de Diagnóstico y Solución de la causa.

Cómo interpretar un diagrama de causa-efecto:

El diagrama causa-efecto es un vehículo para ordenar, de forma muy concentrada, todas las causas que supuestamente pueden contribuir a un determinado efecto.

Nos permite, por tanto, lograr un conocimiento común de un problema complejo, sin ser nunca sustitutivo de los datos. Es importante ser conscientes de que los diagramas de causa-efecto presentan y organizan teorías. Sólo cuando estas teorías son contrastadas con datos podemos probar las causas de los fenómenos observables. Errores comunes son construir el diagrama antes de analizar globalmente los síntomas, limitar las teorías propuestas enmascarando involuntariamente la causa raíz, o cometer errores tanto en la relación causal como en el orden de las teorías, suponiendo un gasto de tiempo importante.

A continuación veremos como el valor de una característica de calidad depende de una combinación de variables y factores que condicionan el proceso productivo (entre otros procesos). Vamos a utilizar el ejemplo de fabricación de mayonesa para explicar los Diagramas de Causa-Efecto:

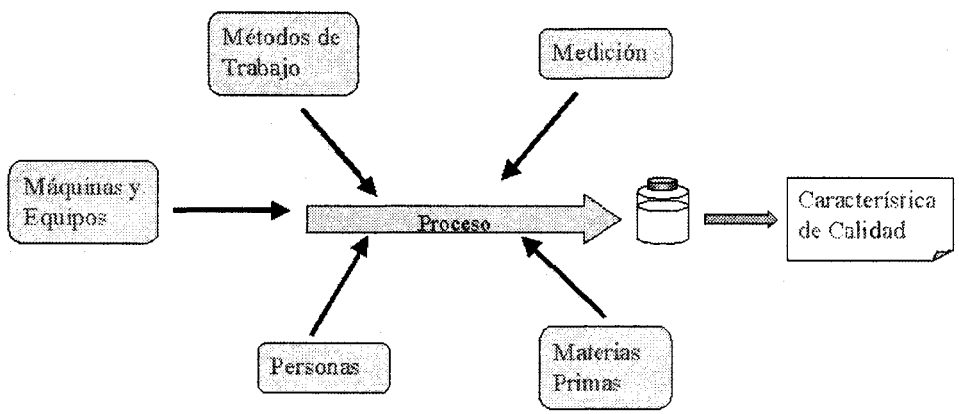


ILUSTRACIÓN 35: Diagrama causa-efecto

La variabilidad de las características de calidad es un efecto observado que tiene múltiples causas. Cuando ocurre algún problema con la calidad del producto, debemos investigar para identificar las causas del mismo. Para hacer un Diagrama de Causa-Efecto seguimos estos pasos:

Decidimos cual va a ser la característica de calidad que vamos a analizar. Por ejemplo, en el caso de la mayonesa podría ser el peso del frasco lleno, la densidad del producto, el porcentaje de aceite, etc.

Trazamos una flecha gruesa que representa el proceso y a la derecha escribimos la característica de calidad:

Indicamos los factores causales más importantes y generales que puedan generar la fluctuación de la característica de calidad, trazando flechas secundarias hacia la principal. Por ejemplo, Materias Primas, Equipos, Operarios, Método de Medición, etc.

Incorporamos en cada rama factores más detallados que se puedan considerar causas de fluctuación. Para hacer esto, podemos formularnos estas preguntas:

- ¿Por qué hay fluctuación o dispersión en los valores de la característica de calidad? Por la fluctuación de las Materias Primas. Se anota Materias Primas como una de las ramas principales.
- ¿Qué Materias Primas producen fluctuación o dispersión en los valores de la característica de calidad? Aceite, Huevos, sal, otros condimentos. Se agrega Aceite como rama menor de la rama principal Materias Primas.
- ¿Por qué hay fluctuación o dispersión en el aceite? Por la fluctuación de la cantidad agregada a la mezcla. Agregamos a Aceite la rama más pequeña Cantidad.
- ¿Por qué hay variación en la cantidad agregada de aceite? Por funcionamiento irregular de la balanza. Se registra la rama Balanza.
- ¿Por qué la balanza funciona en forma irregular? Por que necesita mantenimiento. En la rama Balanza colocamos la rama Mantenimiento.

Así seguimos ampliando el Diagrama de Causa-Efecto hasta que contenga todas las causas posibles de dispersión.

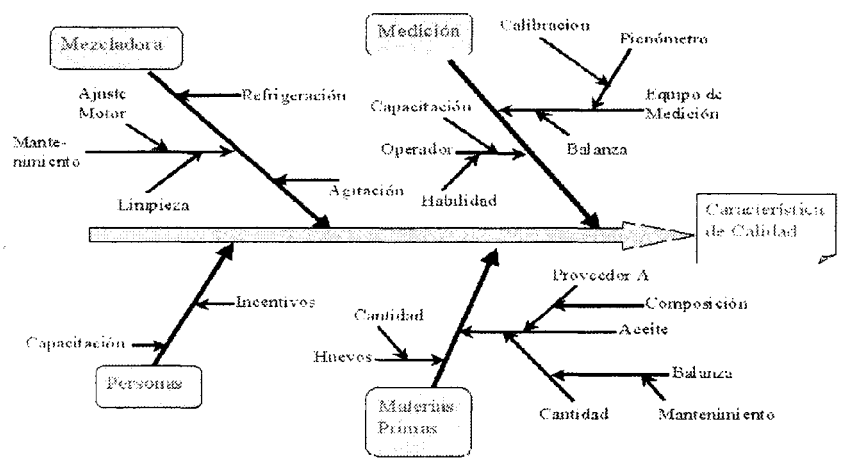


ILUSTRACIÓN 36: Diagrama causa-efecto 2

Finalmente verificamos que todos los factores que puedan causar dispersión hayan sido incorporados al diagrama. Las relaciones Causa-Efecto deben quedar claramente establecidas y en ese caso, el diagrama está terminado.

Un diagrama de Causa-Efecto es de por si educativo, sirve para que la gente conozca en profundidad el proceso con que trabaja, visualizando con claridad las relaciones entre los Efectos y sus Causas. Sirve también para guiar las discusiones, al exponer con claridad los orígenes de un problema de calidad. Y permite encontrar más rápidamente las causas asignables cuando el proceso se aparta de su funcionamiento habitual.

5.2.3. HISTOGRAMA

Un histograma es un resumen gráfico de la variación de un conjunto de datos. La naturaleza gráfica del histograma nos permite ver pautas que son difíciles de observar en una simple tabla numérica. Esta herramienta se utiliza especialmente en la Comprobación de teorías y Pruebas de validez.

Cómo interpretar los histogramas:

Sabemos que los valores varían en todo conjunto de datos. Esta variación sigue cierta pauta. El propósito del análisis de un histograma es, por un lado, identificar y clasificar la pauta de variación, y por otro desarrollar una explicación razonable y relevante de la pauta. La explicación debe basarse en los conocimientos generales y en la observación de las situaciones específicas y debe ser confirmada mediante un análisis adicional. Las pautas habituales de variación más comunes son la distribución en campana, con dos picos, plana, en peine, sesgada, truncada, con un pico aislado, o con un pico en el extremo.

Construcción de un histograma:

Paso 1:

Determinar el rango de los datos: RANGO es igual al dato mayor menos el dato menor; R = > - <

Paso 2:

Obtener en número de clases, existen varios criterios para determinar el número de clases (o barras). Sin embargo ninguno de ellos es exacto. Algunos autores recomiendan de cinco a quince clases, dependiendo de como estén los datos y cuántos sean. Un criterio usado frecuentemente es que el número de clases debe ser aproximadamente ala raíz cuadrada del número de datos, por ejemplo, la raíz cuadrada de 30 (número de artículos) es mayor que cinco, por lo que se seleccionan seis clases.

Paso 3:

Establecer la longitud de clase: es igual al rango entre el número de clases.

Paso 4:

Construir los intervalos de clases: Los intervalos resultan de dividir el rango de los datos en relación al resultado del PASO 2 en intervalos iguales.

Paso 5:

Graficar el histograma: se hace un gráfico de barras, las bases de las barras son los intervalos de clases y altura son la frecuencia de las clases. Si se unen los puntos medios de la base superior de los rectángulos se obtiene el polígono de frecuencias.

Ejemplo: A una fabrica de envases de vidrio, un cliente le está exigiendo que la capacidad de cierto tipo de botella sea de13 ml, con una tolerancia de más menos 1 ml. La fábrica establece un programa de mejora de calidad para que las botellas que se fabriquen cumplan con los requisitos del cliente (Ilustración siguiente).

MUESTREO = 11,12,13,12,13,14,14,15,11,12,13,12,14,15,11,12,16,16,14,13,14,14,13,15,15

1°. RANGO = >-<
16 - 11 = 5

2°. √ 25 = 5

3°. 5 / 5 = 1

4°.

CLASE	INTERVALO	FRECUENCIA	FREC. RELATIVA
1	11 - 12	3	.12
2	12 - 13	5	.25
3	13 - 14	5	.25
4	14 - 15	6	.24
5	15 - 16	6	.24
		25	1.00

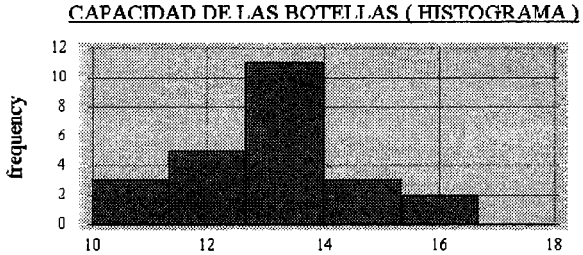


ILUSTRACIÓN 37: Histograma 1

A continuación podemos ver otra forma de representación de histogramas.

Hay histogramas donde se agrupan los datos en clases, y se cuenta cuántas observaciones (frecuencia absoluta) hay en cada una de ellas. En algunas variables (variables cualitativas) las clases están definidas de modo natural, p.e sexo con dos clases: mujer, varón o grupo sanguíneo con cuatro: A, B, AB, O. En las variables cuantitativas, las clases hay que definirlas explícitamente (*intervalos de clase*).

Se representan los intervalos de clase en el eje de abscisas (eje horizontal) y las frecuencias, absolutas o relativas, en el de ordenadas (eje vertical):

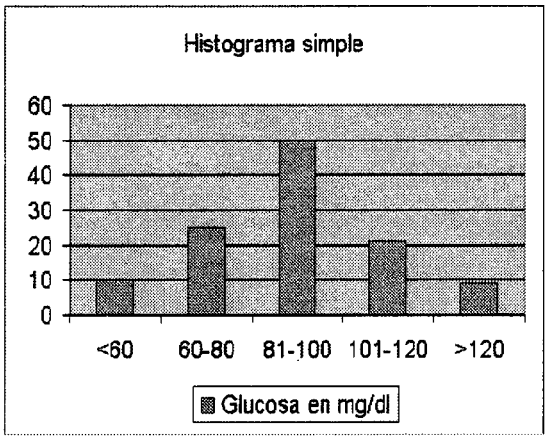


ILUSTRACIÓN 38: Histograma simple

A veces es más útil representar las frecuencias acumuladas:

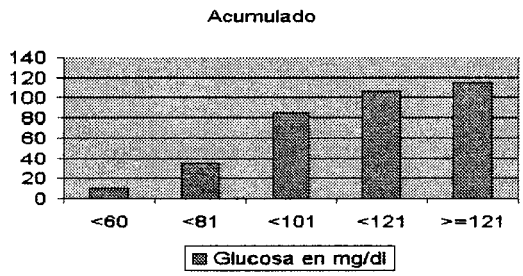


ILUSTRACIÓN 39: Histograma acumulado

Podemos representar en un histograma distintas variables:

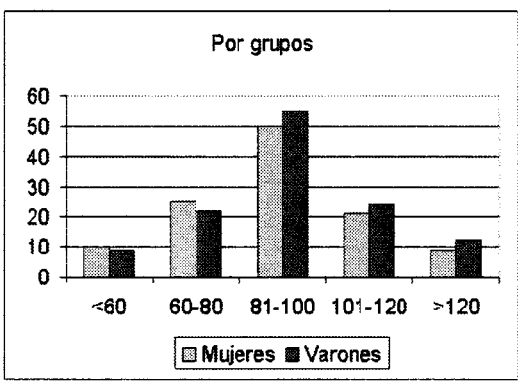


ILUSTRACIÓN 40: Histograma de dos variables

Otra forma muy frecuente, de representar dos histogramas de la misma variable en dos situaciones distintas:

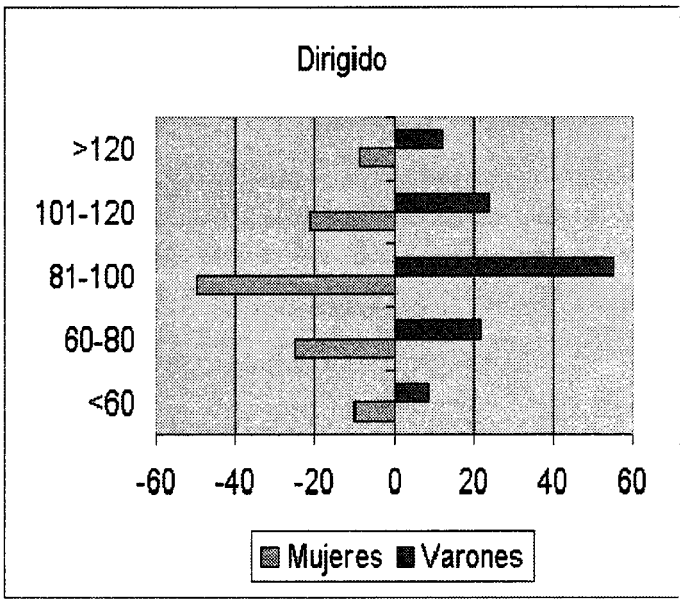


ILUSTRACIÓN 41: Histograma de dos variables comparadas

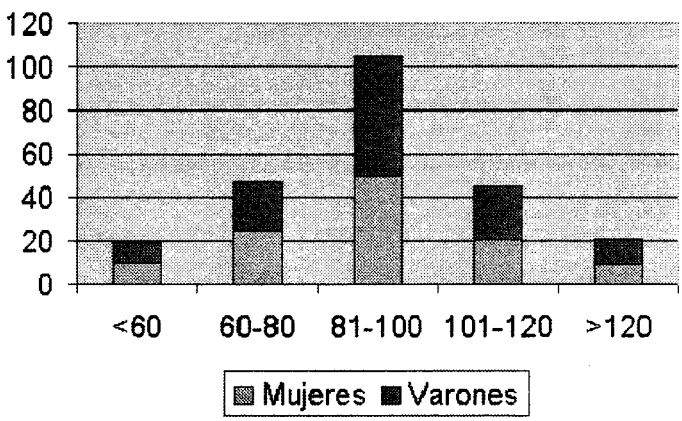


ILUSTRACIÓN 41: Histograma estratificado

En las variables cuantitativas o en las cualitativas ordinales se pueden representar *polígonos* de frecuencia en lugar de histogramas, cuando se representa la frecuencia acumulativa, se denomina *ojiva*.

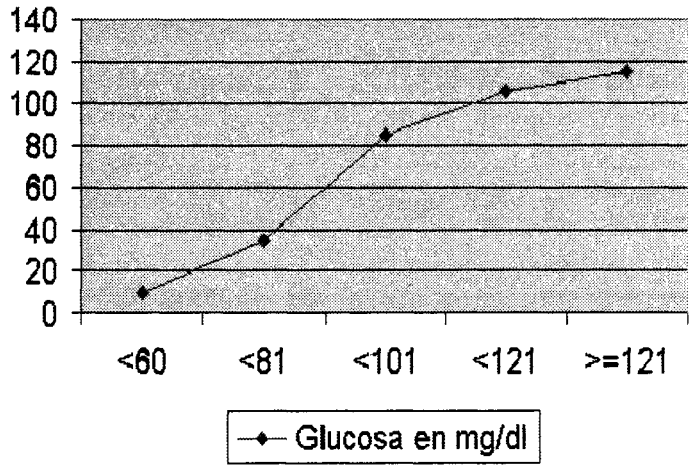


ILUSTRACIÓN 43: Polígono de frecuencias

5.2.4. GRÁFICOS DE CONTROL

Ha sido desarrollado en el capítulo tercero.

Un gráfico de control es una herramienta estadística utilizada para evaluar la estabilidad de un proceso. Permite distinguir entre las causas de variación. Todo proceso tendrá variaciones, pudiendo éstas agruparse en:

Causas aleatorias de variación. Son causas desconocidas y con poca significación, debidas al azar y presentes en todo proceso.

Causas específicas (imputables o asignables). Normalmente no deben estar presentes en el proceso. Provocan variaciones significativas.

Las causas aleatorias son de difícil identificación y eliminación. Las causas específicas sí pueden ser descubiertas y eliminadas, para alcanzar el objetivo de estabilizar el proceso.

Los gráficos de control fueron ideados por Shewhart durante el desarrollo del control estadístico de la calidad. Han tenido una gran difusión siendo ampliamente utilizados en el control de procesos industriales. Sin embargo, con la reformulación del concepto de Calidad y su extensión a las empresas de servicios y a las unidades administrativas y auxiliares, se han convertido en métodos de control aplicables a procesos llevados a cabo en estos ámbitos.

Existen diferentes tipos de gráficos de control:

- De datos por variables. Que a su vez pueden ser de media y rango, mediana y rango, y valores medidos individuales.
- De datos por atributos. Del estilo aceptable / inaceptable, sí / no,...

5.2.5. GRAFICOS DESARROLLO

Son la evolución de los distintos gráficos de control.

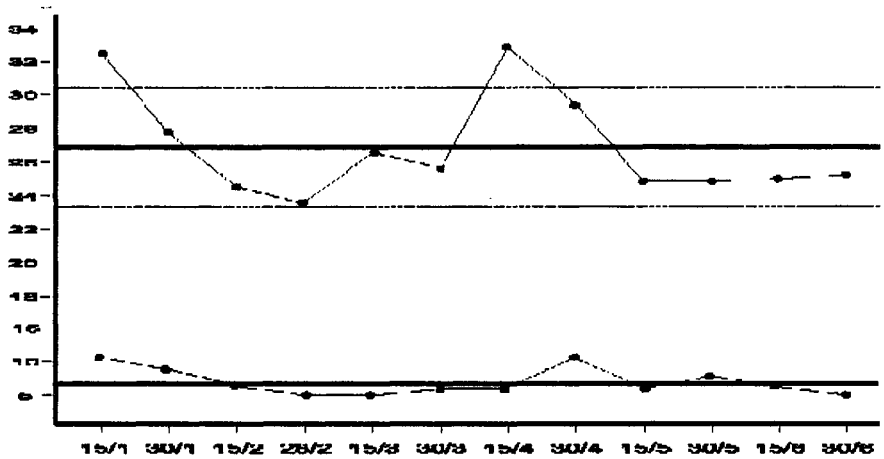


ILUSTRACIÓN 44: Gráficos de desarrollo

5.2.6. DIAGRAMAS DE DISPERSIÓN

Los Diagramas de Dispersión o Gráficos de Correlación permiten estudiar la relación entre dos variables. Dadas dos variables X e Y, se dice que existe una correlación entre ambas si cada vez que aumenta el valor de X aumenta proporcionalmente el valor de Y (Correlación positiva) o si cada vez que aumenta el valor de X disminuye en igual proporción el valor de Y (Correlación negativa).

En un gráfico de correlación representamos cada par X, Y como un punto donde se cortan las coordenadas de X e Y, como podemos ver en la siguiente Ilustración:

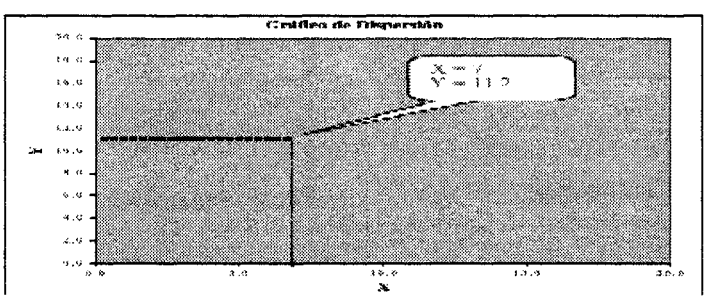


ILUSTRACIÓN 45: Diagrama de dispersión

Veamos un ejemplo.

Supongamos que tenemos un grupo de personas adultas de sexo masculino. Para cada persona se mide la altura en metros (Variable X) y el peso en kilogramos (Variable Y). Es decir, para cada persona tendremos un par de valores X, Y que son la altura y el peso de dicha persona:

TABLA 11: Datos de alturas y pesos

N° Persona	Altura (m)	Peso (Kg.)	N° Persona	Altura (m)	Peso (Kg.)
001	1.91	95.8	026	1.66	74.9
002	1.82	80.5	027	1.96	88.1
003	1.79	78.2	028	1.56	65.3
004	1.63	77.4	029	1.55	64.5
005	1.80	82.6	030	1.71	75.5
006	1.89	87.3	031	1.90	91.3
007	1.57	67.5	032	1.65	66.6
008	1.81	82.5	033	1.78	76.8
009	1.75	82.5	034	1.83	80.2
010	1.63	65.8	035	1.92	97.6
011	1.59	67.3	036	1.67	76.0
012	1.84	88.8	037	1.53	58.0
013	1.82	93.7	038	1.96	95.2
014	1.81	82.9	039	1.66	74.5
015	1.88	98.4	040	1.62	71.8
016	1.68	69.0	041	1.85	91.0
017	1.85	83.4	042	1.53	62.1
018	1.81	89.1	043	1.55	69.8
019	1.89	95.2	044	1.55	64.6
020	1.75	79.1	045	1.97	90.0
021	1.55	62.6	046	1.51	63.8
022	1.71	70.6	047	1.55	62.6
023	1.75	79.4	048	1.67	67.8
024	1.75	78.1	049	1.57	63.3
025	2.03	90.6	050	1.61	65.2

Entonces, para cada persona representamos su altura y su peso con un punto en un gráfico:

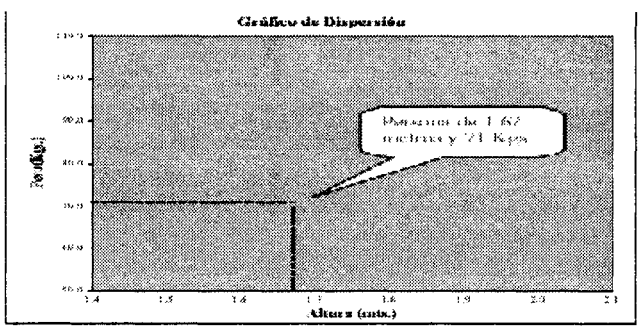


ILUSTRACIÓN 46: Dispersión 2

Una vez que representamos a las 50 personas quedará un gráfico como el siguiente:

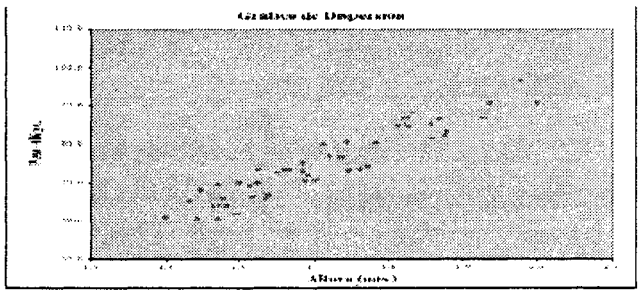


ILUSTRACIÓN 47: Dispersión 3

¿Qué nos muestra este gráfico? En primer lugar podemos observar que las personas de mayor altura tienen mayor peso, es decir parece haber una correlación positiva entre altura y peso. Pero un hombre bajito y gordo puede pesar más que otro alto y flaco. Esto es así porque no hay una correlación totaly absoluta entre las variables altura y peso. Para cada altura hay personas de distinto peso.

Sin embargo podemos afirmar que existe *cierto grado de correlación* entre la altura y el peso de las personas. Cuando se trata de dos variables cualesquiera, puede no haber ninguna correlación o puede existir alguna correlación en mayor o menor grado.

5.2.7. DIAGRAMA DE PARETO

Será desarrollado en el apartado siguiente de forma completa, por su importancia.

5.3. DIAGRAMA DE PARETO

5.3.1. INTRODUCCIÓN

Mediante el Diagrama de Pareto se pueden detectar los problemas que tienen más relevancia mediante la aplicación del principio de Pareto (pocos vitales, muchos triviales) que dice que hay muchos problemas sin importancia frente a solo unos graves. Ya que por lo general, el 80% de los resultados totales se originan en el 20% de los elementos.

La minoría vital aparece a la izquierda de la gráfica y la mayoría útil a la derecha. Hay veces que es necesario combinar elementos de la mayoría útil en una sola cla-

sificación denominada otros, la cual siempre deberá ser colocada en el extremo derecho. La escala vertical es para el costo en unidades monetarias, frecuencia o porcentaje.

El gráfico es muy útil al permitir identificar visualmente en una sola revisión tales minorías de características vitales a las que es importante prestar atención y de esta manera utilizar todos los recursos necesarios para llevar acabo una acción correctiva sin malgastar esfuerzos.

Algunos ejemplos de tales minorías vitales serían:

- La minoría de clientes que representen la mayoría de las ventas.
- La minoría de productos, procesos, o características de la calidad causantes del grueso de desperdicio o de los costos de reelaboración.
- La minoría de rechazos que representa la mayoría de quejas de la clientela.
- La minoría de vendedores que esta vinculada a la mayoría de partes rechazadas.
- La minoría de problemas causantes del grueso del retraso de un proceso.
- La minoría de productos que representan la mayoría de las ganancias obtenidas.
- La minoría de elementos que representan al grueso del costo de un inventarios.

5.3.2. CONCEPTO

El Diagrama de Pareto es una gráfica en donde se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras sencillas después de haber reunido los datos para calificar las causas. De modo que se pueda asignar un orden de prioridades.

5.3.3. EL ANÁLISIS DE PARETO

A principios del siglo XX, Vilfredo Pareto (1848-1923), un economista italiano, realizó un estudio sobre la riqueza y la pobreza. Descubrió que el 20% de las personas controlaba el 80% de la riqueza en Italia. Pareto observó muchas otras distribuciones similares en su estudio. A principios de los años 50, el Dr. Joseph Juran descubrió la evidencia para la regla de «80-20» en una gran variedad de situaciones. En particular, el fenómeno parecía existir sin excepción en problemas relacionados con la calidad.

Una expresión común de la regla 80/20 es que «el ochenta por ciento de nuestro negocio proviene del 20% de nuestros clientes.»

Por lo tanto, el Análisis de Pareto es una técnica que separa los «pocos vitales» de los «muchos triviales». Una Gráfica Pareto es utilizada para separar gráficamen-

te los aspectos significativos de un problema desde los triviales de manera que un equipo sepa dónde dirigir sus esfuerzos para mejorar. Reducir los problemas más significativos (las barras más largas en una Gráfica Pareto) servirá más para una mejora general que reducir los más pequeños. Con frecuencia, un aspecto tendrá el 80% de los problemas.

En el resto de los casos, entre 2 y 3 aspectos serán responsables por el 80% de los problemas.

¿CUÁNDO SE UTILIZA EL DIAGRAMA DE PARETO?

- Al identificar un producto o servicio para el análisis para mejorar la calidad.
- Cuando existe la necesidad de llamar la atención a los problema o causas de una forma sistemática.
- Al identificar oportunidades para mejorar .
- Al analizar las diferentes agrupaciones de datos (ejemplos: por producto, por segmento, del mercado, área geográfica, etc.)
- Al buscar las causas principales de los problemas y establecer la prioridad de las soluciones.
- Al evaluar los resultados de los cambios efectuados a un proceso (antes y después).
- Cuando los datos puedan clasificarse en categorías.
- Cuando el rango de cada categoría es importante.

Pareto es una herramienta de análisis de datos ampliamente utilizada y es por lo tanto útil en la determinación de la causa principal durante un esfuerzo de resolución de problemas. Este permite ver cuáles son los problemas más grandes, permitiéndoles a los grupos establecer prioridades.

En casos típicos, los pocos (pasos, servicios, ítems, problemas, causas) son responsables por la mayor parte el impacto negativo sobre la calidad. Si enfocamos nuestra atención en estos pocos vitales, podemos obtener la mayor ganancia potencial de nuestros esfuerzos por mejorar la calidad.

Un equipo puede utilizar la Gráfica de Pareto para varios propósitos durante un proyecto para lograr mejoras:

- Para analizar las causas
- Para estudiar los resultados
- Para planear una mejora continua

Las Gráficas de Pareto son especialmente valiosas como fotos de «antes y después» para demostrar qué progreso se ha logrado. Como tal, la Gráfica de Pareto es una herramienta sencilla pero poderosa.

CONSEJOS PARA LA CONSTUCCIÓN DEL DIAGRAMA DE PARETO

Un Diagrama de Pareto es un gráfico de barras que enumera las categorías en orden descendente de izquierda a derecha, el cual puede ser utilizado por un equipo para analizar causas, estudiar resultados y planear una mejora continua.

Dentro de las dificultades que se pueden presentar al tratar de interpretar el Diagrama de Pareto es que algunas veces los datos no indican una clara distinción entre las categorías. Esto puede verse en el gráfico cuando todas las barras son más o menos de la misma altura.

Otra dificultad es que se necesita más de la mitad de las categorías para sumar más del 60% del efecto de calidad, por lo que un buen análisis e interpretación depende en su gran mayoría de un buen análisis previo de las causas y posterior recogida de datos.

Debido a que el mismo se ha demostrado como válido en literalmente miles de situaciones, es muy poco probable que se haya encontrado una excepción. Es mucho más probable que simplemente no se haya seleccionado un desglose apropiado de las categorías. Se deberá tratar de estratificar los datos de una manera diferente y repetir el Análisis de Pareto.

Esto nos lleva a la conclusión que para llevar a cabo un proceso de Resolución de Problemas /Toma de Decisiones (RP/TD) es necesario manejar cada una de las herramientas básicas de la calidad, tanto desde el punto de vista teórico como desde su aplicación.

La interpretación de un Diagrama de Pareto se puede definir completando las siguientes oraciones de ejemplo:

«*Existen* (número de categorías) *contribuyentes relacionados con* (efecto). *Pero estos* (número de pocos vitales) *corresponden al* (número) *% del total* (efecto). *Debemos procurar estas* (número) *categorías pocos vitales, ya que representan la mayor ganancia potencial para nuestros esfuerzos.*»

5.3.4. CONSTRUCCIÓN PASO A PASO DEL DIAGRAMA DE PARETO

Paso 1:

En el primer paso se decide la clase de problema que será investigado. Se define el cubrimiento del análisis, si se realiza a una máquina completa, una línea o un sistema de cierto equipo. Se decide que datos serán necesarios y la forma de como clasificarlos. Este punto es fundamental, ya que se pretende preparar la información para facilitar su estratificación posterior.

Paso 2:

Preparar una hoja de recogida de datos. Si la empresa posee un programa informático para la gestión de los datos, se preparará un plan para realizar las búsquedas (sort) y la clasificación de la información que se desea. Es en este punto cuando se puede realizar la estratificación de la información sugerida anteriormente.

Paso 3:

Clasificar en orden de magnitud la información obtenida. Se recomienda indicar con letras (A,B,C,...) los temas que se han ordenado.

Paso 4:

Dibujar dos ejes verticales (izquierdo y derecho) y otro horizontal.

(1) Eje vertical.

- En el eje vertical a la izquierda se marca una escala desde 0 hasta el total acumulado.
- En el eje vertical de la derecha se marca una escala desde 0 hasta l00%

(2) Eje horizontal.

Se divide este eje en un número de intervalos de acuerdo al número de clasificaciones que se pretende realizar. Es allí donde se escribirá el tipo de avería que se ha presentado en el equipo que se estudia.

Paso 5.

Construir el diagrama de barras.

Paso 6:

Marcar con un punto los porcentajes acumulados y unir comenzando desde cero cada uno de estos puntos con líneas rectas obteniendo como resultado la curva acumulada. A esta curva se le conoce como la curva de Lorentz.

Paso 7:

Escribir notas de información del diagrama como título, unidades, nombre de la persona que elaboró el diagrama, período comprendido y número total del datos.

Un diagrama de Pareto es el primer paso para eliminar las averías importantes del equipo. En todo estudio los siguientes aspectos se deben tener en cuenta:

- Toda persona involucrada deberá colaborar activamente
- Concentrarse en la variable que mayor impacto produzca en la mejora.
- Establecer una meta para la mejora

Con la cooperación de todos se podrán obtener excelentes resultados. Uno de los objetivos del Diagrama de Pareto es el de mostrar a todas las personas las áreas prioritarias en que se deben concentrar todas las actividades y el esfuerzo del equipo.

El Diagrama de Pareto presenta claramente la magnitud relativa de los problemas y suministra a los técnicos una base de conocimiento común sobre la cual trabajar (observar la Ilustración siguiente). Una sola mirada vasta para detectar cuales son las barras del diagrama que componen el mayor porcentaje de los problemas. La experiencia demuestra que es más fácil reducir a la mitad una barra alta que reducir una barra de reducida altura a cero.

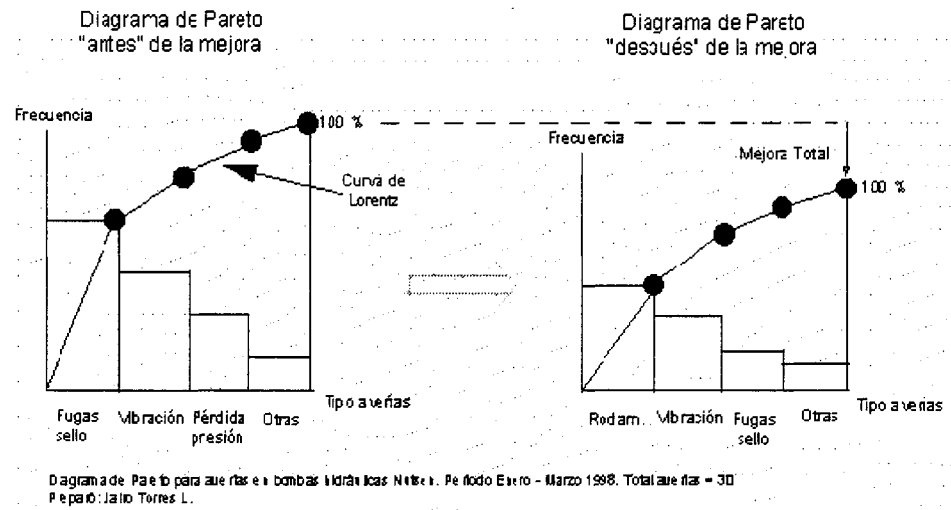


ILUSTRACIÓN 48: Pareto comparativo antes y después de la mejora

EJEMPLOS DE APLICACIÓN

Un fabricante de heladeras desea analizar cuales son los defectos más frecuentes que aparecen en las unidades al salir de la línea de producción.

Para esto, empezó por clasificar todos los defectos posibles en sus diversos tipos (ver tabla):

TABLA 12: Clasificación de los defectos

Tipo de defecto	Detalle del problema
Motor no detiene	No para el motor cuando alcanza temperatura
No enfría	El motor arranca pero la heladera no enfría
Burlete def.	Burlete roto o deforme que no ajusta
Pintura def.	Defectos de pintura en superficies externas
Rayas	Rayas en las superficies externas
No funciona	Al enchufar no arranca el motor
Puerta no cierra	La puerta no cierra correctamente
Gavetas def.	Gavetas interiores con rajaduras
Motor no arranca	El motor no arranca después de ciclo de parada
Mala nivelación	La heladera se balancea y no se puede nivelar
Puerta def.	Puerta de refrigerador no cierra herméticamente
Otros	Otros Defectos no incluídos en los anteriores

Posteriormente, un inspector revisa cada heladera a medida que sale de producción registrando sus defectos de acuerdo con dichos tipos. Después de inspeccionar 88 heladeras, se obtuvo una tabla como esta:

TABLA 13: Frecuencias

Tipo de defecto	Detalle del problema	Frec.
Burlete def.	Burlete roto o deforme que no ajusta	9
Pintura def.	Defectos de pintura en superficies externas	5
Gavetas def.	Gavetas interiores con rajaduras	1
Mala nivelación	La heladera se balancea y no se puede nivelar	1
Motor no arranca	El motor no arranca después de ciclo de parada	1
Motor no detiene	No para el motor cuando alcanza temperatura	36
No enfría	El motor arranca pero la heladera no enfría	27
No funciona	Al enchufar no arranca el motor	2
Otros	Otros defectos no incluidos en los anteriores	0
Puerta def.	Puerta de refrigerador no cierra herméticamente	0
Puerta no cierra	La puerta no cierra correctamente	2
Rayas	Rayas en las superficies externas	4
	Total	88

La última columna muestra el número de heladeras que presentaban cada tipo de defecto, es decir, la frecuencia con que se presenta cada defecto. En lugar de la frecuencia numérica podemos utilizar la frecuencia porcentual, es decir, el porcentaje de heladeras en cada tipo de defecto (siguiente tabla):

TABLA 14: Porcentaje de frecuencias

Tipo de defecto	Detalle del problema	Frec.	Frec. %
Burlete def.	Burlete roto o deforme que no ajusta	9	10.2
Pintura def.	Defectos de pintura en superficies externas	5	5.7
Gavetas def.	Gavetas interiores con rajaduras	1	1.1
Mala divelación	La heladera se balancea y no se puede nivelar	1	1.1
Motor no arranca	El motor no arranca después de ciclo de parada	1	1.1
Motor no detiene	No para el motor cuando alcanza temperatura	36	40.9
No enfría	El motor arranca pero la heladera no enfría	27	30.7
No funciona	Al enchufar no arranca el motor	2	2.3
Otros	Otros defectos no incluidos en los anteriores	0	0
Puerta def.	Puerta de refrigerador no cierra herméticamente	0	0
Puerta no cierra	La puerta no cierra correctamente	2	2.3
Rayas	Rayas en las superficies externas	4	4.5
	Total	**88**	100

Podemos ahora representar los datos en un histograma como el siguiente:

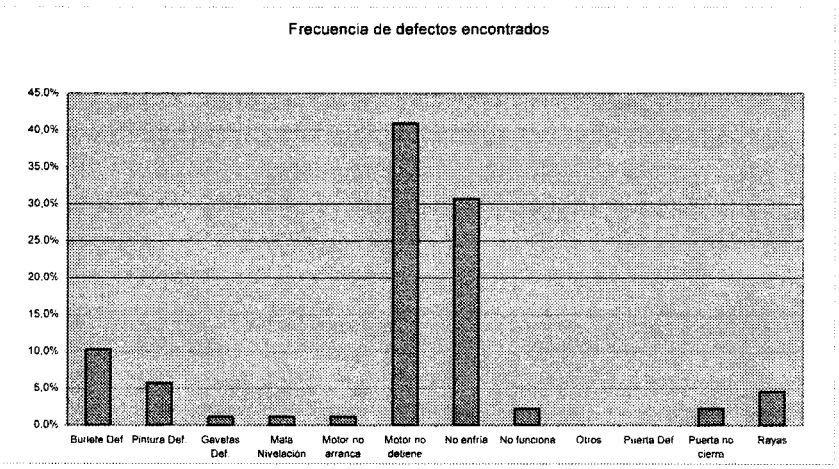


ILUSTRACIÓN 49: Histograma - Pareto

Pero, ¿Cuáles son los defectos que aparecen con mayor frecuencia? Para hacerlo más evidente, antes de graficar podemos ordenar los datos de la tabla en orden decreciente de frecuencia (siguiente tabla):

TABLA 15: Orden de frecuencias

Tipo de defecto	Detalle del problema	Frec.	Frec. %
Motor no detiene	No para el motor cuando alcanza Temperatura	36	40.9
No enfría	El motor arranca pero la heladera no enfría	27	30.7
Burlete def.	Burlete roto o deforme que no ajusta	9	10.2
Pintura def.	Defectos de pintura en superficies externas	5	5.7
Rayas	Rayas en las superficies externas	4	4.5
No funciona	Al enchufar no arranca el motor	2	2.3
Puerta no cierra	La puerta no cierra correctamente	2	2.3
Gavetas def.	Gavetas interiores con rajaduras	1	1.1
Mala nivelación	La heladera se balancea y no se puede nivelar	1	1.1
Motor no arranca	El motor no arranca después de ciclo de parada	1	1.1
Puerta def.	Puerta de refrigerador no cierra herméticamente	0	0.0
Otros	Otros Defectos no incluidos en los anteriores	0	0.0
	Total	88	100

Vemos que la categoría «otros» siempre debe ir al final, sin importar su valor. De esta manera, si hubiese tenido un valor más alto, igual debería haberse ubicado en la última fila.

Ahora resulta evidente cuales son los tipos de defectos más frecuentes. Podemos observar que los 3 primeros tipos de defectos se presentan en el 82 % de las heladeras, aproximadamente. Por el Principio de Pareto, concluimos que: *La mayor parte de los defectos encontrados en el lote pertenece sólo a 3 tipos de defectos, de manera que si se eliminan las causas que los provocan desaparecería la mayor parte de los defectos.*

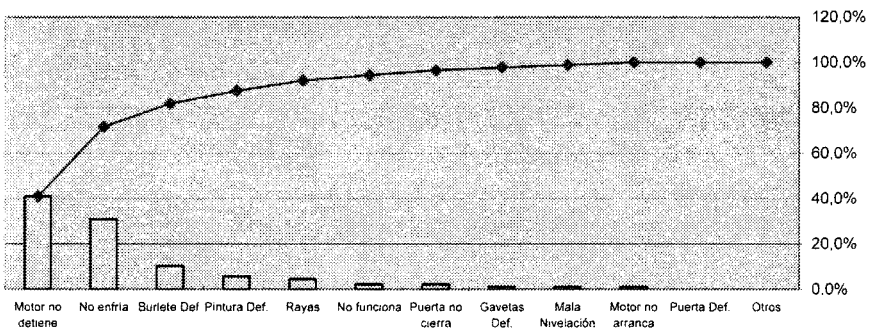


ILUSTRACIÓN 50: Diagrama de Pareto completo

5.4 OTRAS HERRAMIENTAS

5.4.1. Diagrama de flujo

Diagrama de Flujo es una representación gráfica de la secuencia de etapas, operaciones, movimientos, decisiones y otros eventos que ocurren en un proceso. Esta representación se efectúa a través de formas y símbolos gráficos utilizados usualmente.

Los símbolos gráficos para dibujar un diagrama de flujo están más o menos normalizados.

Existen otros símbolos que se pueden utilizar. Lo importante es que su significado se entienda claramente a primera vista. En el ejemplo siguiente, vemos un diagrama de flujo para representar el proceso de fabricación de una resina (Reacción de Polimerización):

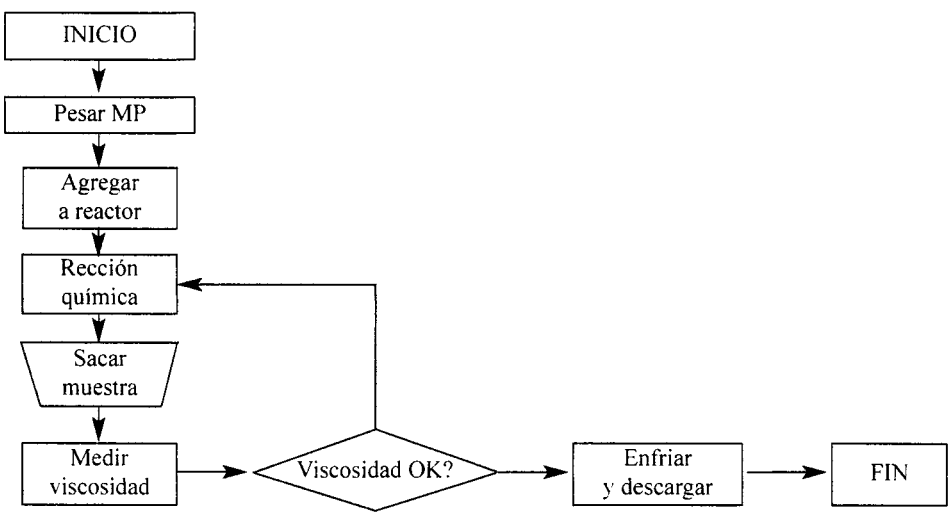


ILUSTRACIÓN 51: Diagrama de flujo

Algunas recomendaciones para construir Diagramas de Flujo son las siguientes:

- Conviene realizar un Diagrama de Flujo que describa el proceso real y no lo que está escrito sobre el mismo (lo que se supone debería ser el proceso).
- Si hay operaciones que no siempre se realizan como está en el diagrama, anotar las excepciones en el diagrama.
- Probar el Diagrama de Flujo tratando de realizar el proceso como está descripto en el mismo, para verificar que todas las operaciones son posibles tal cual figuran en el diagrama.

- Si se piensa en realizar cambios al proceso, entonces se debe hacer un diagrama adicional con los cambios propuestos.

5.4.2. DIAGRAMAS DE GANTT

Los cronogramas de barras o «gráficos de Gantt» fueron concebidos por el ingeniero norteamericano Henry L. Gantt, uno de los precursores de la ingeniería industrial contemporánea de Taylor. Gantt procuro resolver el problema de la programación de actividades, es decir, su distribución conforme a un calendario, de manera tal que se pudiese visualizar el periodo de duración de cada actividad, sus fechas de iniciación y terminación e igualmente el tiempo total requerido para la ejecución de un trabajo. El instrumento que desarrolló permite también que se siga el curso de cada actividad, al proporcionar información del porcentaje ejecutado de cada una de ellas, así como el grado de adelanto o atraso con respecto al plazo previsto.

Este gráfico consiste simplemente en un sistema de coordenadas en que se indica:

- **En el eje Horizontal:** un calendario, o escala de tiempo definido en términos de la unidad más adecuada al trabajo que se va a ejecutar: hora, día, semana, mes, etc.
- **En el eje Vertical:** Las actividades que constituyen el trabajo a ejecutar.
  A cada actividad se hace corresponder una línea horizontal cuya longitud es proporcional a su duración en la cual la medición efectúa con relación a la escala definida en el eje horizontal conforme se ilustra.
- **Símbolos Convencionales:** En la elaboración del gráfico de Gantt se acostumbra utilizar determinados símbolos, aunque pueden diseñarse muchos otros para atender las necesidades específicas del usuario.

Los símbolos básicos son los siguientes:

- Iniciación de una actividad
- Término de una actividad
- Línea fina que conecta las dos «L» invertidas. Indica la duración prevista de la actividad.
- Línea gruesa. Indica la fracción ya realizada de la actividad, en términos de porcentaje. Debe trazarse debajo de la línea fina que representa el plazo previsto.
- Plazo durante el cual no puede realizarse la actividad. Corresponde al tiempo improductivo puede anotarse encima del símbolo utilizando una abreviatura.
- Indica la fecha en que se procedió a la última actualización del gráfico, es decir, en que se hizo la comparación entre las actividades previstas y las efectivamente realizadas.

Utilización

El gráfico de Gantt se presta para la programación de actividades de la más grandes especie, desde la decoración de una casa hasta la construcción de una nave. Desde su creación ha sido un instrumento sumamente adaptable y de uso universal, dad su fácil construcción.

Ventajas y desventajas de los gráficos de Gantt

La ventaja principal del gráfico de Gantt radica en que su trazado requiere un nivel mínimo de planificación, es decir, es necesario que haya un plan que ha de representarse en forma de gráfico. La técnica descrita de este capítulo representa y al mismo tiempo ayuda a la elaboración del plan de trabajo.

Los gráficos de Gantt se revelan muy eficaces en las etapas iniciales de la planificación. Sin embargo, después de iniciada la ejecución de la actividad y cuando comienza a efectuarse modificaciones, el gráfico tiende a volverse confuso. Por eso se utiliza mucho la representación gráfica del plan, en tanto que los ajustes (replanificación) requieren por lo general de la formulación de un nuevo gráfico. Para superar esa deficiencia se crearon dispositivos mecánicos, tales como cuadros magnéticos, fichas, cuerdas, etc., que permite una mayor flexibilidad en las actualizaciones. Aún en términos de planificación, existe todavía una limitación bastante grande en lo que se refiere a la representación de planes de cierta complejidad. El Gráfico de Gantt no ofrece condiciones para el análisis de opciones, ni toma en cuenta factores como el costo. Es fundamentalmente una técnica de pruebas y errores. No permite, tampoco, la visualización de la relación entre las actividades cuando el número de éstas es grande.

En resumen, para la planificación de actividades relativamente simples, el gráfico de Gantt representa un instrumento de bajo costo y extrema simplicidad en su utilización. Para proyectos complejos, sus limitaciones son bastantes serias, y fueron éstas las que llevaron a ensayos que dieron como resultado el desarrollo del CPM, el PERT y otras técnicas conexas. Estas técnicas introdujeron nuevos conceptos que, asociados más tarde a los de los gráficos de Gantt, dieron origen a las denominadas «redes-cronogramas».

Ejemplo

Planificación de tareas para la proyección de una empresa de 3° Fuego

Planificación:

Visita a empresa y entrevista	1 día	11/10/02	11/10/02		Grupo[400%]
Planifiicación del proyecto	1 día	18/10/02	18/10/02	1	Grupo[400%]
Busqueda catálogos maquinaria	3 días	01/11/02	05/11/02	2	Grupo[200%]
Distribución en planta inicial	1 día	06/11/02	06/11/02	3	Grupo[400%]
Busqueda ordenanzas municipales	3 días	21/10/02	23/10/02	2	Grupo[300%]
Selección de polígono y parcela	2 días	24/10/02	25/10/02	5	Grupo[400%]
Adaptación de la distribución en planta a la parcela	1 día	07/11/02	07/11/02	6:4	Grupo[400%]
Catálogos y presupuestos	4 días	28/10/02	31/10/02	6	Grupo
Estudio económico	2 días	15/11/02	18/11/02	8	Grupo[400%]
Entrega anteproyecto	4 días	19/11/02	22/11/02	9	Grupo[400%]
Consulta legislación contra incendios	3 días	Mon 28/10/02	Wed 30/10/02	2	Grupo[200%]
Diseño instalación contra incendios	2 días	25/11/02	26/11/02	«10;11»	Grupo[400%]
Consulta legislación gas	3 días	31/10/02	04/11/02	2	Grupo[200%]
Diseño instalación gas	2 días	13/12/02	16/12/02	«10;13»	Grupo[400%]
Consulta ordenanzas polígono	4 días	28/10/02	31/10/02	6	Grupo
Aislamiento acústico/iluminación	3 días	20/12/02	24/12/02	«15;14;12»	Grupo[400%]
Entrega proyecto	7 días	09/01/03	17/01/03	16	Grupo[400%]

Diagrama de Gantt:

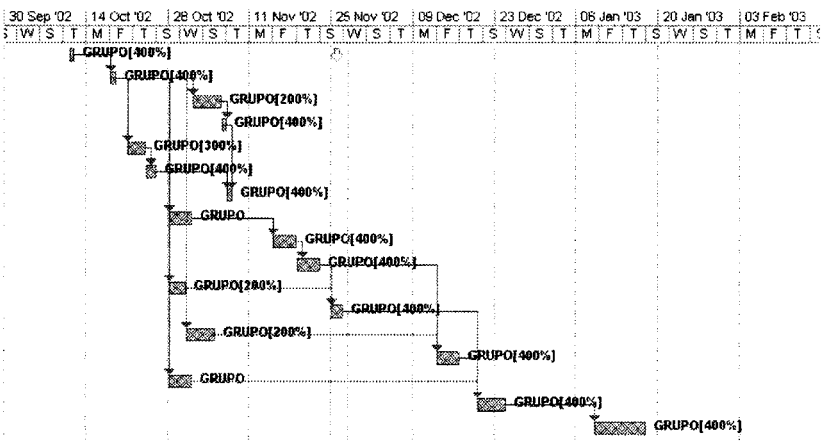


ILUSTRACIÓN 52: Diagrama de Gantt

5.4.3. HOJA DE VERIFICACIÓN E INSPECCIÓN

¿Qué son?

Son formatos sencillos los cuales nos ayudan a recolectar datos.

¿Cómo se elaboran?

Para su elaboración se requiere primeramente que dicho formato conteste a la pregunta:

¿Qué tan frecuente ocurre el evento sujeto a estudio u observación?

Son requisitos indispensables para la elaboración de las hojas de verificación o inspección los siguientes:

- Todas las personas involucradas en la recolección de los datos deben estar de acuerdo sobre el evento que va a ser observado.
- Tomar la decisión del período de tiempo durante el cual serán recolectados del dato (horas, semanas, meses, etc.)
- Diseñar una forma clara y fácil de usar para el registro de los datos.
- Asegurarse de que al obtener los datos este trabajo se haga de una manera consistente y honesta.
- *Asegurarse de que las muestras u observaciones sean tomadas al azar.*

Utilización

Esta herramienta se utiliza cuando se necesita reunir datos basados en la observación con la finalidad de detectar tendencias y comenzar a analizar los problemas subyacentes de éstas; es decir, aquí comienza el proceso de convertir simples opiniones o suposiciones de hechos.

La utilización de las hojas de verificación es el punto lógico de inicio para el análisis de solución de problemas.

Ejemplo de servicio: se realizó un muestreo de los servicios de cajero automático durante un mes en el banco «X» y los resultados de los problemas que se observaron fueron los siguientes (ver tabla):

SERVICIOS DE CAJEROS AUTOMÁTICOS DEL BANCO «X»

TABLA 16: Ejemplo de servicio

PROBLEMAS	FEBRERO (SEMANAS) 1	2	3	4	TOTAL
No dio el dinero solicitado	II	IIII I	III	IIII II	18
No tiene dinero disponible	IIII IIII IIII	IIII IIII IIII IIII	IIII IIII IIII II	IIII IIII IIII IIII II	72
No estaba en red con el banco de la tarjeta usada	III	IIII	II	IIII	14
No estaba en servicio	IIII IIII IIII	IIII IIII IIII IIII IIII	IIII IIII IIII IIII	IIII IIII IIII IIII IIII II	86
No devolvió la tarjetea	I	III	I	IIII	10
					200

Ejemplo de administración / vida diaria: en el lapso de un mes se registraron en el Departamento de Urgencias del hospital «Z» los casos de lesiones que se atendieron, siendo el resultado de estas observaciones el que se muestra en la siguiente hoja de verificación o inspección (ver tabla):

TABLA 17: Hoja de verificación

PARTE DEL CUERPO LESIONADO	TIEMPO DE OBSERVACIÓN: 1 MES	TOTAL
Cuerpo	IIII IIII III	13
Manos	IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII IIII	53
Cabeza	IIII I	6
Dedos	IIII IIII IIII IIII IIII IIII IIII IIII IIII I	61
Pies	IIII IIII	9
		142

5.4.4. LOS 5 POR QUE

Esquema de los 5 por que

Es un método que nos permite identificar la causa raíz y poder así, encontrar soluciones. Esta herramienta es útil cuando el grupo requiere encontrar la causa raíz, cuando se requiere un análisis más profundo y cuando existen muchas causas lo que provoca que el problema se torne confuso.

Cómo se usa?

Ante la presencia de un problema nos preguntamos 5 veces seguidas la pregunta ¿porque?

Ejemplo:

- **1º ¿Por qué?**

¿Por qué se ha detenido la máquina? Porque se ha producido una sobrecarga y ha saltado el fusible.

- **2º ¿Por qué?**

¿Por qué se ha producido la sobrecarga ? El cojinete no estaba suficientemente lubricado y genera un esfuerzo superior al normal.

- **3º ¿Por qué?**

¿Por qué no estaba suficientemente lubricado? La bomba de aceite no bombeaba lo suficiente

- **4º ¿Por qué?**

¿Por qué no bombeaba lo suficiente? Porque el rotor vibraba y hacía perder presión al sistema.

- **5º ¿Por qué?**

¿Por qué vibraba el rotor ? Porque uno de los bujes del eje tiene juego excesivo.

Este método permite hallar causas raíz, evitando intentar la solución de un problema sin un correcto análisis

5.5. HERRAMIENTAS AVANZADAS DE GESTIÓN DE CALIDAD

5.5.1. DIAGRAMA DE AFINIDADES

¿Qué es?

Un diagrama de afinidad es una forma de organizar la información reunida en sesiones de Lluvia de Ideas. Está diseñado para reunir hechos, opiniones e ideas sobre áreas que se encuentran en un estado de desorganización. El diagrama de afinidad ayuda a agrupar aquellos elementos que están relacionados de forma natural. Como resultado, cada grupo se une alrededor de un tema o concepto clave. El uso de un dia-

grama de afinidad es un proceso creativo que produce consenso por medio de la clasificación que hace el equipo en vez de una discusión. El diagrama fue creado por Kawaita Jiro y también es conocido como el método KJ.

¿Cuándo se utiliza?

Se debe utilizar un diagrama de afinidad cuando:

- El problema es complejo o difícil de entender.
- El problema parece estar desorganizado.
- El problema requiere de la participación y soporte de todo el equipo / grupo.
- Se quiere determinar los temas claves de un gran número de ideas y problemas.

¿Cómo se utiliza?

1. Armar el equipo correcto.

El líder del equipo o el facilitador asignado es normalmente responsable por dirigir al equipo a través de todos los pasos para hacer el diagrama de afinidad.

2. Establecer el problema.

El equipo o grupo deberá inicialmente determinar el problema a atender. Es de gran ayuda determinar el problema en la forma de una pregunta.

3. Hacer lluvia de ideas / Reunir datos.

Los datos pueden reunirse en una sección tradicional de Lluvia de Ideas además de los datos reunidos por observación directa, entrevistas y otro material de referencia.

4. Transferir datos a notas Post It.

Los datos reunidos son desglosados en frases independientes con un solo significado evidente y solo una frase registrada en un Post It.

5. Reunir los Post It en grupos similares.

Los Post It deberán colocarse en una pared o rotafilio de tal manera que todos los Post It puedan verse fácilmente. Luego, en silencio, los miembros del equipo agrupan los Post It en grupos similares. Los Post It que sean similares se consideran de «afinidad mutua».

6. Crear una tarjeta de título para cada agrupación.

Los *Post It* deberán leerse y revisarse una vez más con el fin de verificar si han sido agrupados de forma apropiada. Asignar un nombre a cada gr4upo de Post It por medio de una discusión en grupo. Este título deberá transmitir el significado de los



ILUSTRACIÓN 53: Diagrama de afinidad

Post It en muy pocas palabras. Este proceso se repite hasta que todos los grupos tengan un nombre. Cualquier Post It individual que no parezca encajar en ningún grupo puede incluirse en un grupo de «Misceláneos».

7. Dibujar el Diagrama de Afinidad terminado.

Después que los grupos estén ordenados, se deben pegar los Post It en una hoja de rotafolio. Las tarjetas de los títulos se deberán colocar en la parte superior del grupo.

8. Discusión.

El equipo o grupo deberá discutir la relación de los grupos y sus elementos correspondientes con el problema.

A continuación vamos a ver un ejemplo de Diagrama de Afinidad (ilustración 53).

5.5.4. DIAGRAMA DE INTERRELACIONES

¿Qué es?

Un diagrama de interrelaciones presenta las relaciones entre factores / problemas. Toma la idea principal o problema y presenta la conexión entre los ítems relacionados. Al utilizarlo se demuestra que cada ítem puede ser conectado con más de un ítem diferente a la vez. Permite el pensamiento multidireccional.

¿Cuándo se utiliza?

Para comprender y aclarar las interrelaciones entre los diferentes puntos de un problema complejo.

Para identificar puntos clave para mayor investigación.

¿Cómo se utiliza?

1. Reunir el equipo apropiado.
2. Determinar el problema o el asunto clave a solucionar.
3. Utilizar una herramienta de generación de ideas tal como la Lluvia de Ideas para producir ideas.
4. Reunir ideas o tarjetas o notas Post-It y colocarlas en la superficie de trabajo (generalmente dos hojas de rotafolio pegadas hacen una superficie ideal de trabajo) en un patrón circular. Marcar con una letra o número cada tarjeta / Post-It.

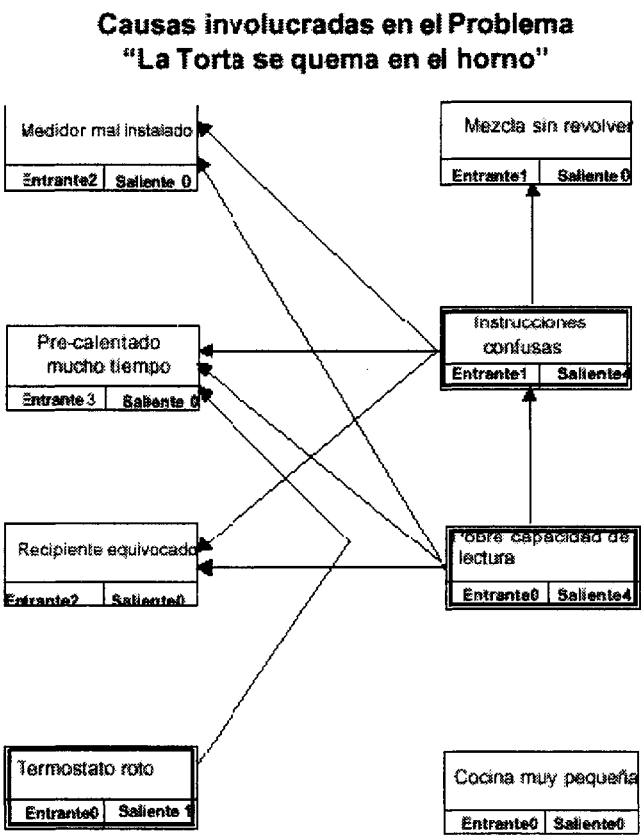


ILUSTRACIÓN 54: Diagrama de interrrelaciones

5. Buscar relaciones entre cada una y todas las ideas. Determinar qué otras tarjetas / Post-Its están influenciadas por esta tarjeta. Dibujar flechas que salgan de la tarjeta / Post-it que influencien otras tarjetas y flechas hacia las tarjetas que estén influenciadas por otras tarjetas.

6. Evitar las flechas de doble vía. Hacer una determinación en cuanto a qué ítem es una mayor influencia.

7. Debajo de cada tarjeta / post-it, totalizar todas las flechas que entran y salen de cada tarjeta. Luego se podrán identificar las causas / impulsos principales (flechas salientes con más frecuencia) y los efectos / resultados claves (flechas entrantes con más frecuencia).

8. Identificar las tarjetas / post-It que son causas o efectos mayores al utilizar casillas dobles o en negrilla. Ver ejemplo.

9. Por consenso, identificar las tarjetas / post-It que sólo tienen pocas flechas hacia adentro o afuera pero todavía pueden ser un ítem o causa clave (ver la caja rota del termostato en el ejemplo). Podemos ver la Ilustración 54 como ejemplo.

5.5.3. DIAGRAMA DE ÁRBOL

Un diagrama de estructura de árbol es la representación de un esquema de la base de datos jerárquica, de ahí el nombre, ya que un árbol esta desarrollado precisamente en orden descendente formando una estructura jerárquica.

Este tipo de diagrama está formado por dos componentes básicos:

- Rectángulos: que representan a los de registros.
- Líneas: que representan a los enlaces o ligas entre los registros.

Un diagrama de árbol tiene el propósito de especificar la estructura global de la base de datos.

Un diagrama de estructura de árbol es similar a un diagrama de estructura de datos en el modelo de red. La principal diferencia es que en el modelo de red los registros se organizan en forma de un grafo arbitrario, mientras que en modelo de estructura de árbol los registros se organizan en forma de un árbol con raíz.

Características de las estructuras de árbol:

- El árbol no puede contener ciclos.
- Las relaciones que existen en la estructura deben ser de tal forma que solo existan relaciones muchos a uno o uno a uno entre un padre y un hijo.

Diagrama de estructura de árbol

En este diagrama podemos observar que las flechas están apuntando de padres a hijos. Un padre (origen de una rama) puede tener una flecha apuntando a un hijo, pero un hijo siempre puede tener una flecha apuntando a su padre.

El esquema de una base de datos se representa como una colección de diagramas de estructura de árbol. Para cada diagrama existe una única instancia de árbol de base de datos. La raíz de este árbol es un nodo ficticio. Los hijos de ese nodo

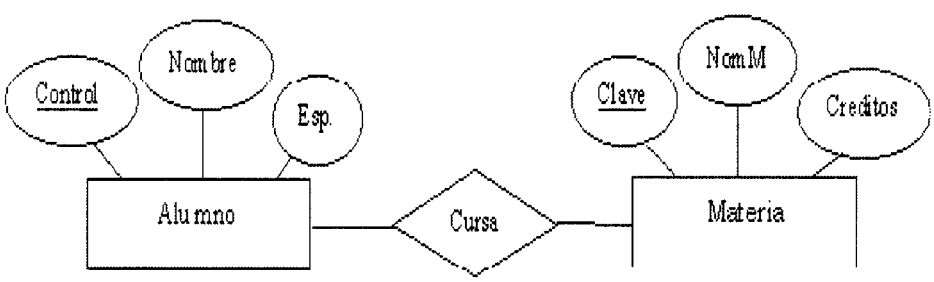


ILUSTRACIÓN 55: Ejemplo alumno-materia

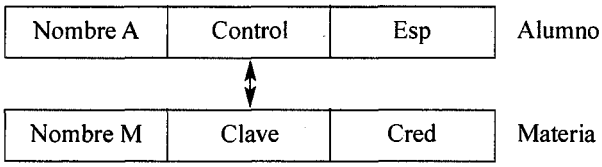
son instancias de los registros de la base de datos. Cada una de las instancias que son hijos pueden tener a su vez, varias instancias de varios registros.

Las representaciones según las cardinalidades son:

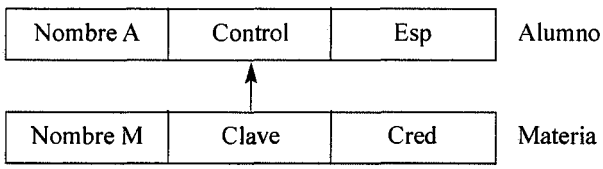
Consideremos la relación alumno-materia sin atributo descriptivo.

La transformación según las cardinalidades seria:

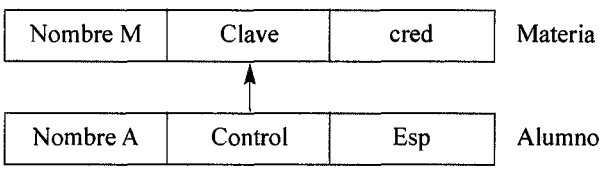
Cuando la relación es uno a uno.



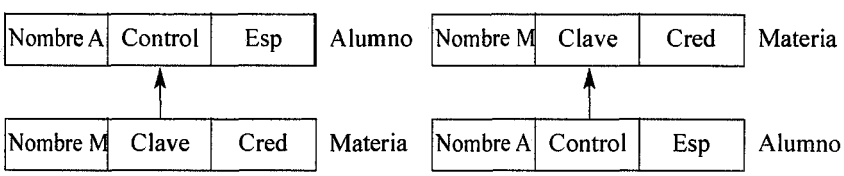
Cuando la relación es uno a muchos.



Cuando la relación es muchos a uno.



Cuando la relación es muchos a muchos.



Cuando la relación tiene atributos descriptivos, la transformación de un diagrama E-R a estructura de árbol se lleva a cabo cubriendo los siguientes pasos:

- Crear un nuevo tipo de registro.
- Crear los enlaces correspondientes.

Consideremos que a la relación Alumno-Materia añadimos el atributo **Cal** a la relación que existe entre ambas, entonces nuestro modelo E-R resulta:

Añadir el diagrama E-R

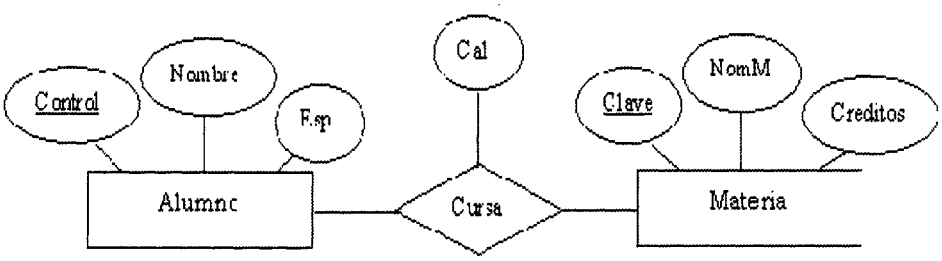
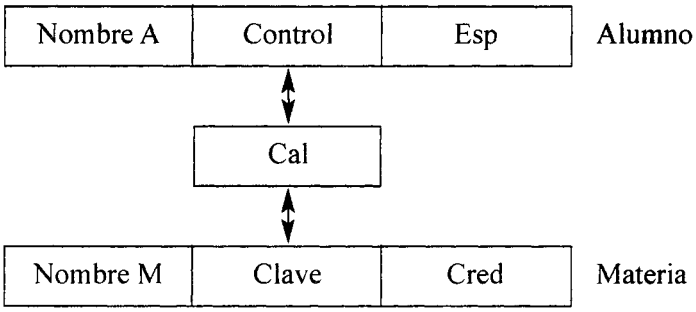


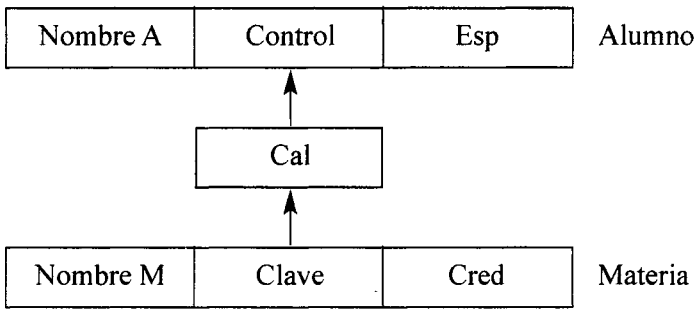
ILUSTRACIÓN 56: Añadiendo Cal

Según las cardinalidades los diagramas de estructura de árbol pueden quedar de la siguiente manera:

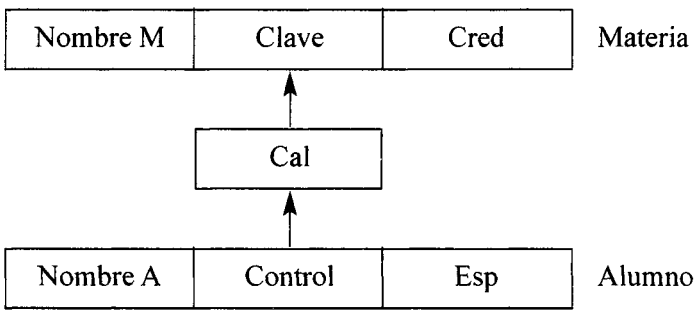
Cuando la relación es uno a uno:



Cuando la relación es uno a muchos:



Cuando la relación es Muchos a uno:



Cuando la relación es Muchos a Muchos:

Si la relación es muchos a muchos entonces la transformación a diagramas de árbol es un poco más compleja debido a que el modelo jerárquico solo se pueden representar las relaciones uno a uno o uno a muchos.

Existen varias formas distintas de transformar este tipo de relaciones a estructura de árbol, sin embargo todas las formas constituyen la repetición de algunos registros.

La decisión de qué método de transformación debe utilizarse depende de muchos factores, entre los que se incluyen:

- El tipo de consultas esperadas en la base de datos.
- El grado al que el esquema global de base de datos que se está modelando se ajusta al diagrama E-R dado.

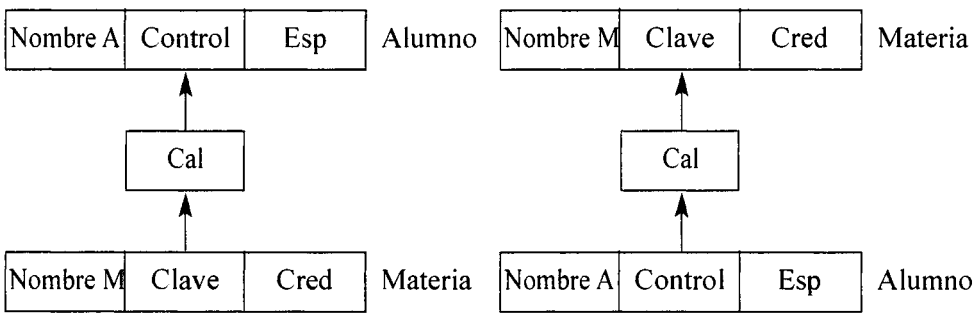
A continuación se describe la forma de transformar un diagrama E-R a estructura de árbol con relaciones muchos a muchos. Suponemos el ejemplo de la relación alumno-materia.

Crear dos diagramas de estructura de árbol distintos T1 y T2, cada uno de los cuales incluye los tipos de registro alumno y materia, en el árbol T1 la raíz es alumno y en T2 la raíz es materia.

Crear los siguientes enlaces:

- Un enlace muchos a uno del registro cuenta al registro Alumno, en T1
- Un enlace muchos a uno del tipo de registro cliente al tipo de registro materia en T2.

Como se muestra en el siguiente diagrama:



5.5.4. DIAGRAMA MATRICIAL

Los diagramas matriciales son herramientas que se emplean para revelar las correlaciones entre ideas, tareas y responsabilidades.

Estos diagramas pueden mostrar la fortaleza estadística y la dirección de influencia de cada relación. Con frecuencia se emplean estos diagramas para mostrar la relación entre las tareas de un diagrama de árbol y otras características o funciones.

Los diagramas matriciales son herramientas de extrema flexibilidad, pueden manejar cualquier tipo de contenido de información y comparar cualquier número de variables. Para lograr esto con éxito, los diagramas matriciales pueden tener cualquiera de las siguientes formas:

Forma de L
Forma de T
Forma de Y
Forma de X
Forma de C

Procedimiento:

1. Reunir a un equipo apropiado
2. Elegir las consideraciones clave.
3. Elegir la forma apropiada de la matriz.
4. Definir los símbolos de relación a emplear y crear una leyenda.
5. Concluir la matriz.

Ejemplo:

El diagrama matricial del ejemplo presenta un programa de capacitación en la calidad. La matriz incluye los nombre de cursos y las listas de asistencia. La gráfica señala quién recibe los talleres detallados y quién requiere sólo una visión global de cada curso.

Por ejemplo, el diagrama muestra que la dirección recibe sólo una visión general de cada curso en tanto que los ingenieros toman un taller detallado de todos ellos. A continuación podemos ver en la Ilustración, un ejemplo.

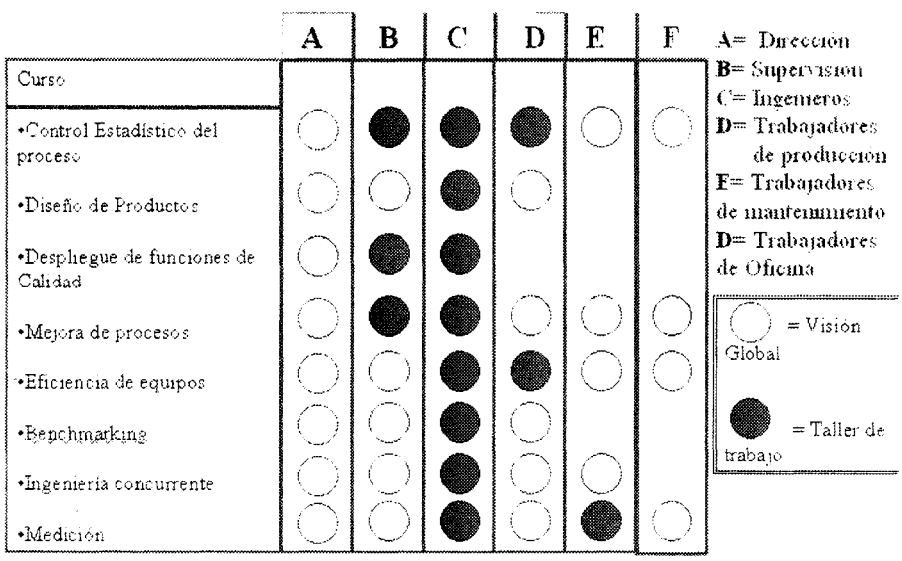


ILUSTRACIÓN 57: Diagrama matricial

5.5.5. DIAGRAMA DE DECISIONES DE ACCIÓN (PDPC)

Es un diagrama que representa el diseño de un plan de actuación, en el que se identifican, dado que se han previsto, el mayor número de acontecimientos aleatorios e

imponderables de todo tipo que pueden aparecer o surgir durante el proceso de implantación, con lo cual es posible:

- Sumar los puntos conseguidos por cada factor o aspecto, para determinar la importancia de cada uno de ellos.
- Predecir posibles situaciones no deseadas.
- Desarrollar y tomar las medidas preventivas necesarias con objeto de evitar que ocurran o poder aplicar, medidas correctoras, en caso de que no haya sido posible su detección.

En el diagrama de decisiones de acción (PDPC) se van proponiendo cuestiones, de tal modo que, en función de las contestaciones recibidas, el propio gráfico señala el camino adecuado a seguir; de este modo es posible evidenciar los riesgos a evitar o las acciones a tomar.

Es un diagrama apropiado fundamentalmente para:

- Implantar planes de actuación de cierta complejidad, facilitando la labor de programación en lo incierto, puesto que se desconocen los resultados.
- Ser capaces de hacer frente a las dificultades.
- Emprender tareas complejas cuya probabilidad de error o fallo es alta, lo cual obliga a detectar fallos con anticipación

5.5.6. DIAGRAMA DE FLECHAS

Un diagrama de flechas es una técnica de administración de redes de uso generalizado para la planeación e implantación de tares complejas , en particular las más comunes que cuentan con subtareas conocidas.

Se muestra el tiempo para implantar la tarea global y vigila los programas de tiempo y costo para cada subtarea. Un diagrama de redes de actividades es una combinación de la Técnica de Revisión y Evaluación de Programas (PERT) y el Método de la Ruta Crítica (CPM). Ofrece una forma de trazar las relaciones entre diversas actividades en cualquier proyecto.

Procedimiento:

1. Reunir el equipo apropiado
2. Identificar todas las tareas que requiere el proyecto
3. Determinar la secuencia de actividades
4. Calcular el tiempo que se requiere para cada actividad

5. Calcular la ruta crítica del proyecto

6. Calcular las fecha más tardía de inicio y más temprana de determinación de cada subtarea. La forma más común para mostrar estas cuatro fechas es el siguiente.

7. Calcular la holgura total

8. Diseñar el diagrama de redes de actividades.

Ejemplo: El diagrama de redes de actividades del ejemplo muestra las actividades que suponen el desarrollo de una clase de capacitación. El diagrama muestra las rutas y actividades, así como las flechas más próximas y más tardías de inicio y terminación. Es necesario señalar que existen actividades paralelas en el diagrama.

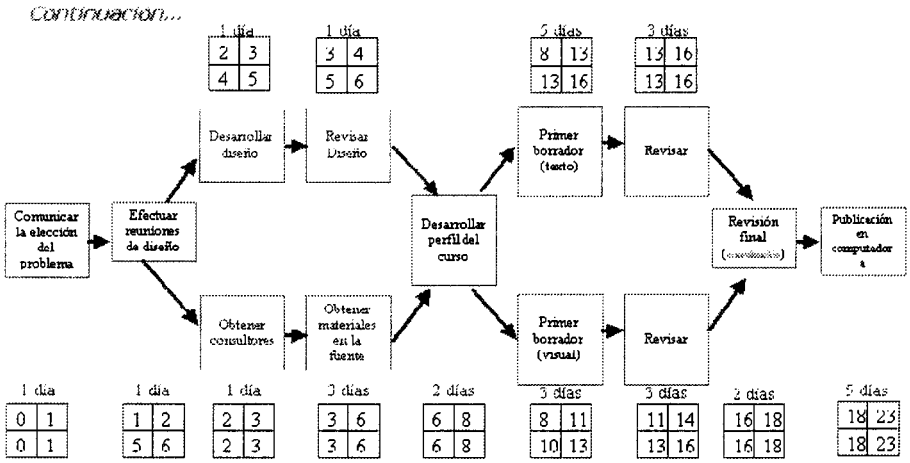


ILUSTRACIÓN 58: Diagrama de flechas

5.5.7. ANÁLISIS FACTORIAL DE DATOS

EL ANÁLISIS FACTORIAL

Es una modalidad del Análisis Multivariado que nos permite reducir una serie de variables a un conjunto menor (factores) que contienen la mayor parte de la información y son suficientes para explicar el modelo.

Puede ser planteada una clasificación del Análisis Multivariado que considera dos grupos básicos: Métodos de Dependencia y Métodos de Interdependencia. Este autor indica que el objetivo de los Métodos de Interdependencia es dar significado a un conjunto de variables o tratar de agrupar las cosas. Por tanto, el Análisis Factorial se ubicaría dentro de éstos últimos.

Queda claro que es un procedimiento mediante el cual se toma un gran número de variables y se investiga para ver si tienen un pequeño número de factores en común que expliquen su intercorrelación.

Hay que señalar que el Análisis Factorial de Componentes Principales analiza tablas de individuos por variables métricas. Es decir, este análisis se aplica a tablas rectangulares de datos, en las que las columnas representan las variables y las filas a los individuos. Esto se puede apreciar en el siguiente tabla:

VARIABLES

SUJETOS	X_1	X_2	X_3	X_4	X_5
1					
2					
3					
4					

Distintos trabajos, precisan que el objetivo del Análisis Factorial es simplificar los datos ya que condensa la información contenida en un gran número de unidades métricas (por ejemplo, escalas de calificación) en un número más pequeño de escalas resumidas, llamadas «factores».

Por ejemplo, una encuesta conformada por 40 items (cada uno de los cuales será considerado una variable de insumo, es decir, de entrada) con seguridad contendrá una serie de redundancias, en el sentido de que varios de los items pueden estar midiendo en parte la misma idea fundamental. Esta idea fundamental es lo que se denomina «factor». Queda definido el «factor» como una variable que no es directamente observable pero que necesita ser inferida de las variables de insumo (entrada). Es decir, el factor es el agrupamiento de aquellas variables de insumo que miden o que son indicadoras del factor.

Es propuesto el siguiente ejemplo, en el cual, las variables talla y peso pueden ser resumidas en el factor «condición física» y las variables ocupación, educación e ingreso pueden ser resumidas en el factor «condición social».

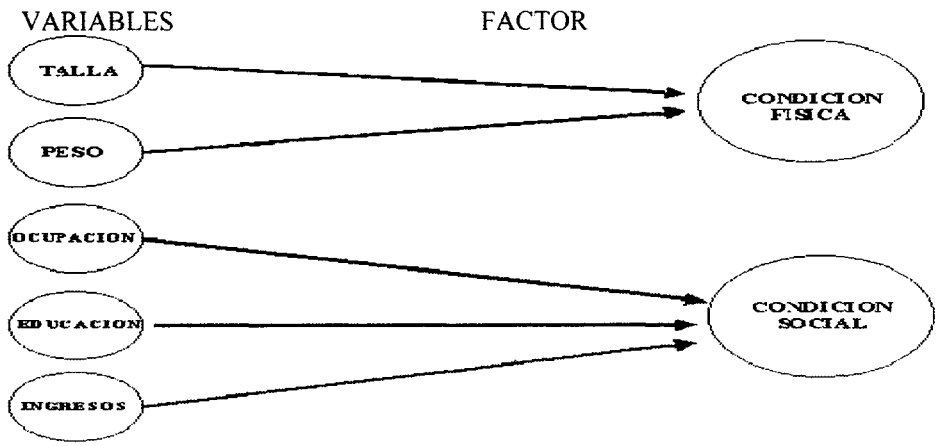


ILUSTRACIÓN 59: Ejemplo variables-factor

UTILIDAD DEL ANÁLISIS FACTORIAL EN MARKETING

El Análisis Factorial puede aplicarse en:

- Segmentación de Mercados: Para identificar las variables subyacentes en las cuales deben agruparse los clientes. Así, por ejemplo, los compradores de automóviles nuevos pueden agruparse sobre la base de la importancia que dan a la economía, la comodidad, el desempeño, el lujo, el servicio postventa, etc.
- Investigación de Productos: Para identificar los atributos de las marcas que influyen en la elección del consumidor. Una marca de pasta dental puede evaluarse por: la protección contra la caries, la blancura que proporciona a los dientes, el sabor, el aliento fresco que otorga, su precio, etc.
- Publicidad: Para comprender los hábitos de consumo del mercado meta. Los consumidores de comida rápida (fast food) pueden espectar determinados programas de TV, preferir determinado tipo de música, gustar determinado tipo de ropa, transporte, etc.
- Estudios sobre Precios: Para identificar las características de los consumidores sensibles al precio. Estos consumidores pueden ser de clase media, preocupados por la economía, caseros, etc.

El Análisis Factorial se utiliza para reducir una gran masa de información a un nivel fácil de manejar. Si se han recolectado 50 atributos de un producto, el Análisis Factorial reducirá los atributos a un número mínimo de factores que sustenten los 50 atributos y que pueden utilizarse posteriormente en vez de los 50 atributos con una pérdida mínima de información. Un nuevo campo de aplicación es el del MARKETING POLÍTICO (Segmentación de Electores, Características Ideales del Candidato, Características de los Electores de cada Candidato, etc.)

PASOS PARA REALIZAR UN ANÁLISIS FACTORIAL

Indicar que es importante que las variables consideradas para el Análisis Factorial se midan con una escala de intervalo y que el tamaño de la muestra sea, cuando menos, el quíntuplo de las variables empleadas. Algunos autores, señalan que es conveniente aplicar una prueba preliminar (Prueba Piloto o Pre–Test) A esta Prueba Piloto se le aplica el Análisis Factorial para «depurarla» de modo que sólo contenga preguntas realmente relevantes, es decir, que midan cosas distintas sobre el objeto de estudio. Con esto se ahorra gastos de tipo y de su aplicación, se abrevia el cuestionario y se incrementa la tasa de respuestas.

Los tres pasos en el proceso de Análisis Factorial son:

a. Elaboración de la Matriz de Correlaciones

Se debe obtener, en primer lugar, una matriz en la que se ubican las correlaciones entre todas las variables consideradas. Es muy conveniente solicitar una serie de pruebas conexas (tests) que nos indicarán si es pertinente, desde el punto de vista estadístico, llevar a cabo el Análisis Factorial con la data y muestras disponibles. La mayoría de programas computarizados que realizan el Análisis Factorial disponen de una serie de tests que permiten verificar si es posible llevar a cabo el mismo. Entre los principales tenemos:

- El Test de Esfericidad de Bartlett: Se utiliza para probar la Hipótesis Nula que afirma que las variables no están correlacionadas en la población. Es decir, comprueba si la matriz de correlaciones es una matriz de identidad. Se puede dar como válidos aquellos resultados que nos presenten un valor elevado del test y cuya fiabilidad sea menor a 0.05. En este caso se rechaza la Hipótesis Nula y se continúa con el Análisis.
- El Índice Kaiser-Meyer-Olkin: Mide la adecuación de la muestra. Indica qué tan apropiado es aplicar el Análisis Factorial. Los valores entre 0.5 y 1 indican que es apropiado aplicarlo.

El Gráfico de Sedimentación representa en el eje de las «x» el número de orden de los factores y en el eje de las «y», los valores propios (eigenvalues). Muestra la forma en que van disminuyendo los valores propios, graficando el hecho de que el primer factor es el que más varianza explica.

La «comunalidad», es la cantidad de varianza que una variable comparte con las demás variables consideradas. Los «eigenvalue» (valores propios) pueden interpretarse como la cantidad de varianza explicada por cada factor. La «carga factorial» es la correlación entre las variables y los factores.

b. Extracción de los Factores Iniciales

Se dispone de muchos métodos para extraer los Factores Iniciales de la matriz de correlación. El más utilizado y el que empleamos en este estudio es el de «Componentes Principales». Este procedimiento busca el factor que explique la mayor cantidad de la varianza en la matriz de correlación. Este recibe el nombre de «factor principal». Esta varianza explicada se resta de la matriz original produciéndose una matriz residual. Luego se extrae un segundo factor de esta matriz residual y así sucesivamente hasta que quede muy poca varianza que pueda explicarse. Los factores así extraídos no se correlacionan entre ellos, por esta razón se dice que estos factores son ortogonales.

c. Rotación de los Factores Iniciales

Con frecuencia es difícil interpretar los factores iniciales, por lo tanto, la extracción inicial se rota con la finalidad de lograr una solución que facilite la interpretación. Hay dos sistemas básicos de rotación de factores: los métodos de rotación ortogonales (mantienen la independencia entre los factores rotados: varimax, quartimax y equamax) y los métodos de rotación no ortogonales (proporcionan nuevos factores rotados que guardan relación entre sí). En el presente estudio aplicaremos los métodos de rotación ortogonales, específicamente el Método de Rotación Varimax. Éste es, actualmente, uno de los métodos más utilizados.

En lo que se refiere a cómo interpretar el «Gráfico de Factores» (Gráfico de Saturaciones, algunos autores señalan las siguientes indicaciones generales:

- Proximidades entre las variables: Dos o más variables muy correlacionadas positivamente se sitúan muy próximas. Dos o más variables muy correlacionadas negativamente están muy alejadas. Si son independientes, su distancia es del tipo medio.
- Cuanto más alejada esté una variable del origen en un plano factorial mayor es la calidad de representación sobre ese plano.

5.5.8. 6-SIGMA

¿Qué es Seis Sigma?

- una medida del rendimiento de los procesos.
- una metodología de mejora del negocio
- una iniciativa de transformación empresarial

«A un nivel 6 sigma, un proceso funciona correctamente el 99.9997% de las veces».

La sigma de un proceso designa su capacidad de operar correctamente, o en terminología Seis Sigma, sin defectos. Un proceso 3 sigma funciona sin errores en el 93% de los casos. Un proceso 6 sigma lo hace en el 99.9997% de los casos, o con solo 3.4 «defectos por millón de oportunidades».

La mayoría de las empresas operan actualmente a niveles entre 3 y 4 sigma. El rendimiento y los niveles de calidad son mucho menores en las áreas de servicios.

Detrás de estas cifras se esconden grandes ineficiencias (las llamadas 'factorías ocultas'), clientes insatisfechos, y una enorme oportunidad de mejorar la productividad y la satisfacción de los clientes. Se estima que los costes asociados a 'problemas' de calidad en empresas que operan entre 3 y 4 sigma, suponen entre el 10 y el 15% de los ingresos.

Seis Sigma utiliza datos y herramientas estadísticas para alcanzar un conocimiento profundo de las necesidades de los clientes y del comportamiento de los procesos; y para diseñar y mejorar procesos, productos y servicios»

HISTORIA DE SEIS SIGMA

Esta filosofia se inicia en los años 80's como una estrategia de negocios y de mejoramiento de la calidad, introducida por Motorola, la cual ha sido ampliamente difundida y adoptada por otras empresas de clase mundial, tales como: G.E., Allied Signal, Sony, Polaroid, Dow Chemical, FeDex, Dupont, NASA, Lockheed, Bombardier, Toshiba, J&J, Ford, ABB, Black & Decker, etc.

Su aplicación requiere del uso intensivo de herramientas y metodologías estadísticas (en su mayoría) para eliminar la variabilidad de los procesos y producir los resultados esperados, con el mínimo posible de defectos, bajos costos y maxima satisfaccion del cliente. Esto contrasta con la forma tradicional de asegurar la calidad, al inspeccionar post-mortem y tratar de corregir los defectos, una vez producidos.

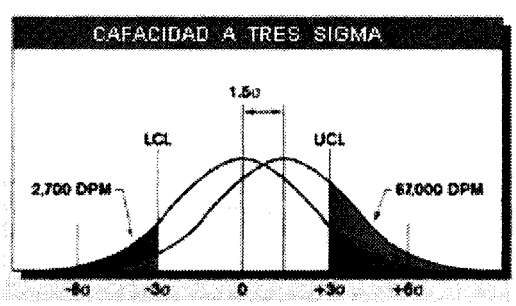


ILUSTRACIÓN 60: Capacidad tres sigma

Un proceso con una curva de capacidad afinada para seis (6) sigma, es capaz de producir con un mínimo de hasta 3,4 defectos por millón de oportunidades (DPMO), lo que equivale a un nivel de calidad del 99.9997 %.

Este nivel de calidad se aproxima al ideal del cero-defectos y puede ser aplicado no solo a procesos industriales de manufactura, sino tambien en procesos transaccionales y comerciales de cualquier tipo, como por ejemplo: en servicios financieros, logísticos, mercantiles, etc.

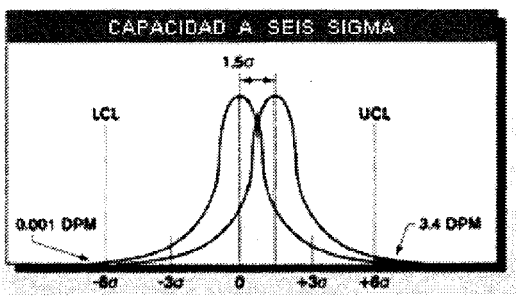


ILUSTRACIÓN 61: Capacidad 6 sigma

Quizá la contribución más importante para el auge y desarrollo actual de Seis Sigma, haya sido el interés y esfuerzo dedicado para su implantación en toda G.E., desde sus divisiones financieras, hasta sus divisiones de equipos médicos y de manufactura. La fuerza impulsora que apuntaló y apoyó esta iniciativa: Jack Welch, CEO de G.E. «Miren, Solamente tengo tres cosas que hacer: tengo que seleccionar a las personas correctas, asignar la cantidad adecuada de doláres y transmitir ideas de una división a otra a la velocidad de la luz. Así que realmente estoy en el negocio de promover y transmitir ideas».

La historia de Seis Sigma se inicia en Motorola cuando un ingeniero (Mikel Harry) comienza a influenciar a la organización para que se estudie la variación en los procesos (enfocado en los conceptos de Deming), como una manera de mejorar los mismos. Estas variaciones son lo que estadísticamente se conoce como desviación estándar (alrededor de la media). Esta iniciativa se convirtió en el punto focal del esfuerzo para mejorar la calidad en Motorola, capturando la atención del entonces CEO de Motorola: Bob Galvin. Con el apoyo de Galvin, se hizo enfasis no sólo en el análisis de la variación sinó también en la mejora continua, estableciendo como meta obtener 3, 4 defectos (por millón de oportunidades) en los procesos; algo casi cercano a la perfección.

Esta iniciativa llegó a oidos de Lawrence Bossidy, quién en 1991 y luego de una exitosa carrera en General Electric, toma las riendas de Allied Signal para transformarla de una empresa con problemas en una máquina exitosa. Durante la implantación de Seis Sigma en los años 90 (con el empuje de Bossidy), Allied Signal multiplicó

sus ventas y sus ganancias de manera dramática. Este ejemplo fué seguido por Texas Instruments, logrando el mismo éxito. Durante el verano de 1995 el CEO de GE, Jack Welch, se entera del éxito de esta nueva estrategia de boca del mismo Lawrence Bossidy, dando lugar a la mayor transformación iniciada en esta enorme organización.

El empuje y respaldo de Jack Welch transformaron a GE en una «organización Seis Sigma», con resultados impactantes en todas sus divisiones. Por ejemplo: GE Medical Systems recientemente introdujo al mercado un nuevo scanner para diagnóstico (con un valor de 1,25 millones de doláres) desarrollado enteramente bajo los principios de Seis Sigma y con un tiempo de scan de sólo 17 segundos (lo normal eran 180 segundos). En otra de las divisiones: GE Plastics, se mejoró dramáticamente uno de los procesos para incrementar la producción en casi 500 mil toneladas, logrando no sólo un beneficio mayor, sinó obteniendo también el contrato para la fabricación de las cubiertas de la nueva computadora iMac de Apple.

LA PERSPECTIVA DE SEIS SIGMA

Seis Sigma es una metodología rigurosa que utiliza herramientas y métodos estadísticos, para Definir los problemas y situaciones a mejorar, Medir para obtener la información y los datos, Analizar la información recolectada, Incorporar y emprender mejoras al o a los procesos y finalmente, Controlar o rediseñar los procesos o productos existentes, con la finalidad de alcanzar etapas óptimas, lo que a su vez genera un ciclo de mejora contínua.

La metodología formal de aplicación de Seis Sigma en general sigue este esquema: DMAIC; sin embargo, algunos practicantes prefieren incorporar otras etapas adicionales, tales como: Reconocer la situación o problema, Estandarizar los nuevos procesos en toda la organización, y finalmente, Integrar los cambios o soluciones a toda la organización.

HERRAMIENTAS UTILIZADAS

Dentro del arsenal de herramientas utilizadas para soportar Seis Sigma, se encuentran casi todas las conocidas en el mundo de la Calidad tradicional, TQM, etc. Se pueden mencionar entre otras:

- CIP, Procesos de Mejora Continua.
- Diseño/Rediseño de Procesos.
- Análisis de Varianza, ANOVA.
- Cuadro de Mando Integral, BSC.
- La Voz del Cliente, VOC.

- Pensamiento Creativo.
- Diseño de Experimentos, DoE.
- Gerencia de los Procesos.
- Control Estadístico de Procesos, SPC.

ELEMENTOS CLAVE

Los elementos clave que soportan la filosofía Seis Sigma y que aseguran una adecuada aplicación de las herramientas, así como el éxito de esta iniciativa como estrategia de negocios, son los siguientes:

- Identificación de los elementos Críticos para la Calidad (CTQ), de los clientes Externos.
- Identificación de los elementos Críticos para la Calidad (CTQ), de los clientes Internos.
- Realización de los análisis de los modos y efectos de las fallas (FMEA).
- Utilización del Diseño de Experimentos (DoE), para la identificación de las variables críticas.
- Hacer Benchmarking permanente y establecer los objetivos a alcanzar, sin ambiguedades.

IMPLEMENTACION DE SEIS SIGMA

Esta filosofía promueve la utilización de herramientas y métodos estadísticos de manera sistemática y organizada, para el logro de mejoras dramáticas y medibles por su impacto financiero. El ingrediente secreto que hace que funcione, reside en la infraestructura que se establece en la organización. Esta infraestructura, es la que motiva y produce una cultura «Seis Sigma» que junto con un «Proceso de Pensamiento» en toda la organización, genera un estilo de «Gerencia Basada en Conocimientos».

El soporte y compromiso por parte de la Alta Gerencia es vital y fundamental, para lo cual se entrenan y definen los Maestros (también conocidos como Champions), quienes son los dueños de los proyectos críticos para la organización. Para desarrollar estos proyectos se escogen y preparan Expertos (conocidos como: Master Black Belt, Black Belt, Green Belt), quienes se convierten en agentes de cambio para impulsar y desarrollar estos proyectos, en conjunto con los equipos de trabajo seleccionados para los mismos.

Diferencias entre Calidad Tradicional y Seis Sigma

CALIDAD TRADICIONAL	SEIS SIGMA
Está centralizada. Su estructura es rígida y de enfoque reactivo.	Está descentralizada en una estructura constituída para la detección y solución de los problemas. Su enfoque es proactivo.
Generalmente no hay una aplicación estructurada de las herramientas de mejora.	Se hace uso estructurado de las herramientas de mejora y de las técnicas estadísticas para la solución de los problemas.
No se tiene soporte en la aplicación de las herramientas de mejora. Generalmente su uso es localizado y aislado.	Se provee toda una estructura de apoyo y capacitación al personal, para el empleo de las herramientasde mejora.
La toma de decisiones se efectúa sobre la base de presentimientos y datos vagos.	La toma de decisiones se basa en datos precisos y objetivos: «Sólo en Dios creo, los demás traigan datos».
Se aplican remedios provisionales o parches. Sólo se corrige en vez de prevenir.	Se vá a la causa raíz para implementar soluciones sólidas y efectivas y así prevenir la recurrencia de los problemas.
No se establecen planes estructurados de formación y capacitación para la aplicación de las técnicas estadísticas requeridas.	Se establecen planes de entrenamiento estructurados para la aplicación de las técnicas estadísticas requeridas.
Se enfoca solamente en la inspección para la detección de los defectos (variables clave de salida del proceso). Post-Mortem.	Se enfoca hacia el control de las variables clave de entrada al proceso, las cuales generan la salida o producto deseado del proceso.

Como resumen:

- Seis Sigma se hace imperativo cuando hay que evaluar y mejorar la capacidad de los procesos.
- Seis Sigma es un medio para reducir la complejidad de diseños de productos y procesos al tiempo que se aumenta su fiabilidad.
- Seis sigma es una puesta en escena para combatir lo que muchas veces se achaca a la «mala suerte»; esta puesta en escena es valida no solo en el taller sino en cualquier lugar de la organización.

6

NORMAS DE ACEPTACIÓN Y RECHAZO ISO 10545 DE BALDOSAS CERÁMICAS, (PARTE 1). APLICACIÓN PRÁCTICA DE POISSON

6.1. LAS NORMAS DE INSPECCION Y RECEPCION POR ATRIBUTOS. UNE EN 66-020-73

En las normas de baldosas, hemos visto la ISO 10545 (parte 1) sobre muestreo y condiciones de aceptación. Ene esta norma se detallaba el sistema de coger muestras de baldosas para determinar la aceptabilidad del lote.

Además de esa norma, existen otras muchas que plantean planes de muestreo con el objetivo de tomar decisiones sobre el lote en función de los resultados obtenidos para una muestra extraída del mismo. Una de estas normas es la UNE 66-020-73, y otras normas de la serie ISO (ISO 2859-1:1989 y la ISO 3951-:1989.)

Vamos a interpretar esta norma UNE como ejemplo, ya que su estructura es muy similar a las ISO y si sabemos aplicar ésta sabremos interpretar otras similares. Esta norma es de inspección por atributos, y puede servir tanto para unidades defectuosas como para defectos por cada 100 unidades.

Esta norma consta de tres partes bien diferenciadas:

- Las definiciones de los conceptos básicos manejados en la norma
- Las tablas de consulta para diferentes tipos de muestreo, aceptación y rechazo de lotes
- Las curvas características de aceptabilidad de lotes según planes de muestreo.

6.2. DEFINICIONES Y CONCEPTOS GENERALES

Es prácticamente imposible conseguir 0 defectos. Por lo tanto son necesarios sistemas de inspección que nos informen de la calidad de los productos, tanto los que recibimos como los que fabricamos. Estos sistemas de inspección afectan a:

- Inspección en recepción de materias primas, suministros de maquinaria o repuestos
- Inspección previa de productos fabricados por otras entidades y que se tienen que procesar en nuestra planta.
- Inspecciones intermedias en el proceso de fabricación
- Inspección anterior a la entrega para asegurarnos de la calidad de nuestros productos
- Inspección por parte del cliente de nuestros productos para decidir la aceptación o no.

La primera pregunta que surge es ¿Inspección al 100% o inspección de muestras sacadas de los lotes?.

La inspección al 100% es evidentemente más segura que la que se hace sobre una muestra extraída. Su finalidad es detectar y separar los artículos defectuosos, y se aplica generalmente a

- Artículos muy costosos
- Artículos cuyo fallo puede provocar graves problemas de funcionamiento, seguridad, etc.
- Si resulta factible y económicamente rentable
- Si el lote es reducido

La inspección por muestreo pretende aceptar o rechazar un lote en función de los resultados obtenidos en la verificación de una muestra extraída del lote. La muestra se extrae siguiendo procesos basados en la estadística y las probabilidades.

Los beneficios y características de este sistema también son claros:

- Menor coste que la inspección al 100%
- Posibilita la inspección cuando el tiempo de inspeccionar es superior al tiempo necesario para fabricar la pieza
- Es adecuado para inspeccionar productos a los que se aplican técnicas de ensayo destructivas
- También se emplea cuando el coste de inspección es mayor que el valor intrínseco de la pieza.

Las desventajas son también evidentes:

- Se emplean técnicas estadísticas, que llevan aparejadas un error, por lo que no se puede saber con toda seguridad la calidad de un lote, ni se puede estar seguro que en un lote aceptado no habremos cometido un error, o bien que pueden resultar rechazados lotes buenos.

1. Inspección. Proceso de medir, examinar, ensayar o comparar de alguna manera la unidad de producto con las prescripciones exigidas al mismo

2. Inspección por atributos. Es aquella en virtud de la cual la unidad de producto se clasifica como defectuosa o no defectuosa, o bien, aquella en la que se computa el número de defectos de dicha unidad de producto, en relación con las especificaciones o normas dadas. Esta inspección es la que se realiza para ver si una baldosa lleva defectos.

3. Inspección por variables, o por valor medio, es cuando se mide una característica a la que se puede asignar un valor numérico, comparándola con el valor que marcan las normas dadas. Si en vez de considerar un valor, se considra la media de varios, se denomina inspección por valor medio. Este procedimiento se emplea para comparar el calibre de una baldosa con el que debería tener para su formato.

4. Unidad de producto. Es aquella que se inspecciona con el fin de determinar su clasificación como defectuosa o no defectuosa, o para computar el número de defectos de la misma. La unidad de producto puede ser un solo objeto, un par, un conjunto, una superficie (1 m^2 por ejemplo) etc.

5. Defecto es cualquier disconformidad de la unidad de producto con las prescripciones exigidas. Una clasificación de defectos es una enumeración de los defectos posibles en la unidad de producto, calsificados de acuerdo con su gravedad.

6. Defecto crítico, Defecto crítico es aquel que según la razón y la experiencia conduce probablemente a condiciones arriesgadas o inseguras para los individuos que utilicen, mantengan o dependan del producto considerado.

7. Defecto principal lo define como aquel defecto que sin ser crítico puede ocasionar un fallo o avería o reducir sustancialmente la idoneidad o utilidad de un producto para la finalidad propuesta.

8. Defecto secundario, es aquel que no reduce sensiblemente la utilidad de la unidad de producto para el fin a que está destinada.

9. Unidad defectuosa crítica, Unidad defectuosa crítica. Es la que tiene uno o más defectos críticos

10. Unidad defectuosa principal. Es aquella que no teniendo ningún defecto crítico tiene uno o más defectos principales. Puede contener defectos secundarios

11. Unidad defectuosa secundaria. Es la que solo tiene uno o más defectos secundarios. La determinación de que un defecto sea crítico o no, puede ser un punto de negociación entre proveedor de baldosas y cliente, que se debe de consensuar para evitar reclamaciones posteriores.

12. Porcentaje de unidades defectuosas y número de defectos por cada 100 unidades. La norma dice que el grado de disconformidad se expresará en en % de unidades defectuosas o

$$\%\ defectuoso = \frac{n^o\ unidades\ defectuosas}{n^o\ unidades\ inspeccionadas}\ x\ 100$$

por el nº de defectos en cada 100 unidades.

$$Defectos\ por\ cada\ 100\ unidades = \frac{n^o\ defectos}{n^o\ unidades\ inspeccionadas}\ x\ 100$$

Así si de 500 unidades inspeccionadas salen 40 unidades malas, el % defectuoso

$$\%\ defectuoso = \frac{40}{500}\ x\ 100 = 8\%$$

Si hubiese un total de 84 defectos en esas 500 baldosas, el nº de defectos por cada 100 unidades será

$$Defectos\ por\ cada\ 100\ unidades = \frac{40}{500}\ x\ 100 = 16.8\%$$

13. Nivel de calidad aceptable. Es el porcentaje máximo de unidades defectuosas o el nº máximo de defectos por cada 100 unidades alcanzable por la media del proceso para poder considerar éste último satisfactorio desde el punto de vista de la inspección por muestreo. Cuando un consumidor establece un valor para el NCA para un cierto defecto o grupo de defectos, le indica al proveedor que aceptará con el plan de muestreo de la norma la mayoría de los lotes que el proveedor presente a inspección, con tal de que el nivel medio del % de unidades defectuosas (o número de defectos por cada 100 unidades) en dichos lotes, no sea superior al valor designado por el NCA. Como se ve el NCA sirve para conjunto de lotes. Para un lote individual se deben de seguir otras pautas.

14. Limitación. La designación de un NCA no imlica que el proveedor pueda suministrar a sabiendas unidades defectuosas

15. Especificación de los NCA. Los hará la administración o se especificará en los términos del contrato entre proveedor y cliente. Se pueden establecer diferentes NCA para ciertos defectos considerados en conjunto o aislados. Los valores de los NCA iguales o inferiores a 10 se podrán expresar en % de unidades defectuosas o % de defectos por cada 100 unidades. Para valores mayores de 10, siempre estaremos expresando % de defectos por cada 100 unidades.

16 NCA preferentes. Esta norma indica unos NCA denominados preferentes.

17. Lote. Este término se aplica en el sentido de lote para inspección, es decir, que es un conjunto de unidades de producto del que se extrae una muestra para ser inspeccionada, con objeto de determinar si el lote es aceptado o rechazado de acuerdo con lo establecido por esta norma.

18. Tamaño del lote es el número de unidades de producto de que consta el lote

19. Formación de lotes. El producto se agrupa en lotes, sublotes identificables o de otra forma que pueda establecerse. Cada lote será homogéneo y constará de un solo tipo de producto de unidades de producto, producidas en las mismas condiciones y en un período de tiempo. Por ejemplo, se pueden formar lotes con las baldosas de porosa tratadas en u turno, siempre que sean del mismo modelo.

20. Presentación de los lotes. Los lotes se formarán de acuerdo a los establecido por la autoridad o los acuerdos proveedor-cliente, y deberá ser identificado claramente. El proveedor dispondrá de un local adecuado para el almacenamiento de cada lote, tener los equipos necesarios para su identificación , y disponer de personal entrenado para la extracción de las muestras

21. Aceptación y rechazo. En este apartado la norma hace reservas sobre los defectos críticos, y además indica que se podrán rechazar unidades de producto defectuosas extraídas en el proceso de inspección.

22. Lotes sometidos de nuevo a inspección. La norma especifica que los lotes rechazados podrán ser presentados a nueva inspección únicamente cuando todas las unidades hayan sido examinadas y ensayadas, eliminando todas las unidades defectuosas o corrigiendo los defectos.

23. Muestra. Se compone de una o más unidades de producto extraídas al azar de un lote. El número de unidades de producto se denomina tamaño de muestra

24. Muestreo representativo. Consiste en determinar el número de unidades de la muestra que se extraerán de cada parte o sublotes que integran el lote, identificados mediante algún criterio racional. Las unidades deben de ser seleccionadas al azar. En la práctica, supongamos un lote de baldosas consistente en 1000 m^2. Este lote está dividido en sublotes (palets por ejemplo). La norma nos dirá cuantas unidades deben de componer cada muestra que se extraiga de cada sublote. Además el muestreo será aleatorio, o sea wue se sacarán las unidades de producto al azar.

25. Muestreo simple, doble o múltiple. La extracción de la muestra podrá ser simple, es decir, una sola muestra, o doble, (dos muestras cuando la primera muestra no es completamente satisfactoria) o múltiple (varias extracciones)

26. Inspecciones normal, rigurosa o reducida. Los sistemas de muestreo y aceptación y rechazo de lotes están divididos en tres clases de inspección, que son normal, reducida y rigurosa. El cambio de unos a otros sistemas de inspección se hará con unos criterios que especifica la norma en función de cómo vayan siendo aceptados o rechazados o lotes. Plan de muestreo. Indica el número de unidades de producto de cada lote que deberán ser inspeccionadas y los criterios para la determinación de la aceptabilidad o rechazo del lote. Estos criterios suelen ser un número de aceptación y un número de rechazo. Estos números suelen ser el número máximo de unidades defectuosas que se pueden encontrar para que el lote sea aceptado (nº aceptación) o el número de unidades defectuosas para el cual el lote es rechazado (nº de rechazo). Si se encuentra un número de unidades defectuosas entre medio de los dos números, se suele coger una segunda muestra, en la que vuelve a haber un número de aceptación y otro de rechazo, pero con las unidades defectuosas acumuladas de las dos extracciones

27. Nivel de inspección. Indica la relación entre el tamaño de lote y el tamaño de muestra. Está claro que cuantas más baldosas se saquen, más segura será la decisión de aceptar o rechazar el lote. La norma da tres niveles de inspección I, II y III. Se suele usar el nivel II. Los tamaños de las muestras se designan por letras código.

28. Obtención de un plan de muestreo. Con el NCA y la letra código se obtienen los planes de muestreo de las tablas correspondientes de la norma. Recordemos que los planes pueden ser simples, dobles o múltiples, debiendo especificarse cual se va a seguir. Las normas de baldosas usan el plan doble.

29. Determinación de la aceptabilidad. Se hace con los números de aceptación y rechazo de las tablas correspondientes.

30. Plan de muestreo simple. Normalmente se compara el número de unidades no conformes con el número de aceptación. Si es menor o igual , se acepta el lote. Si es mayor o igual al número de rechazo, se rechaza el lote

31. Plan de muestreo doble. Si el número de unidades no conformes es menor o igal al número de aceptación se acepta el lote. Si es mayor, pero menor que el número de rechazo se coge una segunda muestra. Si es mayor o igual que el número de rechazo, se rechaza el lote. Si se coge una segunda muestra, se deben de sumar las unidades defectuosas de la primera muestra y de la segunda, y comparar la suma con los números de aceptación y rechazo de la segunda muestra, que los especifica la tabla. El ejemplo que vimos en la ISO 10545 es un plan de muestreo doble.

32. Plan múltiple. Es similar al doble, pero con más extracciones de muestras.

Antes de pasar a ver ejemplos de aplicación de la norma, conviene hacer algunas precisiones sobre esta norma.

PROPIEDADES DE LOS PLANES DE MUESTREO

1. Existirán lotes cuya calidad sea aceptable y sin embargo podrían ser rechazados algunos de ellos. En los planes que se trabajan habitualmente, la probabilidad de rechazar lotes con calidad igual al NCA suele estar alrededor del 5%. (Riesgo del productor)

2. También puede ocurrir que sean aceptados lotes cuya calidad no se a aceptable. El riesgo suele oscilar alrededor del 10%. (Riesgo del cliente)

3. El % de unidades defectuosas de una muestra no tiene porqué coincidir con el % del lote. Lo que si ocurre que la media de % defectuosas de diferentes muestras extraídas del lote tienden al % defectuoso del lote

4. El lote debe de estar compuesto por unidades homogéneas (Baldosas de un mismo turno de producción, de una misma máquina, etc.)

5. Es mejor formar sublotes pequeños e inspeccionar muestras extraídas de ellos, que una única muestra extraída del lote global.

6. Los planes de muestreo inspeccionan más unidades en lotes pequeños que en un solo lote suma de todos.

8. Los defectos críticos implican pactar valores de NCA menores que para otros defectos.

9. Si existen muchas quejas de los clientes, se deberá reducir el valor aplicado de NCA

10. Cuando no se rechazan lotes, es posible aumentar el valor de los NCA.

CALIDAD LÍMITE

La calidad límite se define como el máximo % defectuoso que se puede admitir en un lote individual. Este valor se fija porque hemos dicho que se pueden aceptar lotes si la media de calidad de todos ellos está por debajo del NCA. Pero este valor pone un límite tal que si se rebaja se tiene que rechazar el lote. Normalmente este valor se fija en el 10 % defectuosas.

El nivel de NCA es frecuente colocarlo en el 5% de unidades defectuosas de media, aunque para defectos críticos se puede rebajar.

6.3. EJEMPLO DE APLICACIÓN. TABLAS DE CONSULTA

Supongamos que queremos decidir la aceptación de un lote determinado de una serie de lotes. A este se le aplican los siguientes parámetros:

- Tamaño del lote 3500 baldosas (Si fuesen m^2 la unidad de producto sería 1 m^2)
- Nivel de inspección II
- Inspección normal
- Muestreo doble
- NCA 2,5% de unidades defectuosas

Los pasos a seguir serán:

a) Tamaño de la muestra que se debe extraer. Se consulta la tabla de la norma, que como dice que es de 3201 a 10.000, corresponde una letra L

b) Se busca la tabla de inspección normal y muestreo doble (tabla III-A pag 13 de la norma).

TABLA 18: Niveles de inspección

Tamaño del lote			Niveles de inspección especiales				Niveles de inspección para usos generales		
			S-1	S-2	S-3	S-4	I	II	III
2	a	8	A	A	A	A	A	A	B
9	a	15	A	A	A	A	A	B	C
16	a	25	A	A	B	B	B	C	D
26	a	50	A	B	B	C	C	D	E
51	a	90	B	B	C	C	C	E	F
91	a	150	B	B	C	D	D	F	G
151	a	280	B	C	D	E	E	G	H
281	a	500	B	C	D	E	F	H	J
501	a	1200	C	C	E	F	G	J	K
1201	a	3200	C	D	E	G	H	K	L
3201	a	10000	C	D	F	G	J	L	M
10001	a	35000	C	D	F	H	K	M	N
35001	a	150000	D	E	G	J	L	N	P
150001	a	500000	D	E	G	J	M	P	Q
500001	y	superiores	D	E	H	K	N	Q	R

En dicha tabla nos dice que la muestra 1ª debe de tener un tamaño de muestra de 125 unidades, y que para un NCA del 2,5 % le corresponde un NA de 5 y un NR de 9. Igualmente nos indica que la muestra 2ª en caso de que el número de unidades no conforme esté entre 2 y 5, debe de ser de 125 unidades más, y que los Na y NR de esta muestra más la primera son 12 y 13.

Supongamos que hemos obtenido en la inspección de las 50 unidades extraídas (más o menos el 5%) 2 unidades defectuosas. **LA DECISION SERÍA ACEPTAR EL LOTE.**

Supongamos que hemos obtenido 10 unidades defectuosas. **LA DECISION SERÍA RECHAZAR EL LOTE.**

Supongamos que hemos obtenido 7 unidades defectuosas. **LA DECISION SERIA VOLVER A COGER 125 UNIDADES MÁS DEL LOTE.**

Tabla 19: Niveles de calidad aceptables

NIVELES DE CALIDAD ACEPTABLES (Inspección Normal)

Letra código del tamaño de la muestra	Muestra	Tamaño muestral	Tamaño muestras acumuladas	0.010	0.015	0.025	0.040	0.065	0.10	0.15	0.25	0.40	0.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
				A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R	A R
A				↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	•	↓	↓	•	•	•	•	•	•	•	•	•
B	1ª	2	2	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31
	2ª	3	4																	1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57
C	1ª	3	3	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑
	2ª	3	6																1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57	
D	1ª	5	5	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑	↑
	2ª	5	10															1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57		
E	1ª	8	8	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	17 22	25 31	↑	↑	↑
	2ª	8	16														1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27	37 38	56 57			
F	1ª	13	13	↓	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑
	2ª	13	26													1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27						
G	1ª	20	20	↓	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑
	2ª	20	40												1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27							
H	1ª	32	32	↓	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	32	64											1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27								
J	1ª	50	50	↓	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	50	100										1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27									
K	1ª	80	80	↓	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	80	160									1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27										
L	1ª	125	125	↓	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	125	250								1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27											
M	1ª	200	200	↓	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	200	400							1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27												
N	1ª	315	315	↓	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	315	630						1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27													
P	1ª	500	500	↓		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	500	1000					1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27														
Q	1ª	800	800		↑	↓	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	800	1600				1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27															
R	1ª	1200	1250	↑	↑	0 2	0 3	1 4	2 5	3 7	5 9	7 11	11 16	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
	2ª	1200	2500			1 2	3 4	4 5	6 7	8 9	12 13	18 19	26 27																

↓ = Utilizar el primer plan de muestreo situado debajo de la flecha. Si el tamaño de la muestra es igual o superior al tamaño del lote, efectuar una inspección al 100%.

↑ = Utilizar el primer plan de muestreo situado encima de la flecha.

A = Número de aceptación.

R = Numero de rechazo.

• = Utilizar el Plan de Muestreo Simple (con la alternancia de poder emplear el Plan de Muestreo Doble inmediato disponible).

Índice

Una vez inspeccionadas estas 125 unidades de la muestra segunda, vemos que han aparecido 3 unidades defectuosas más. En total, las sumamos con las 7 defectuosas de la 1ª muestra y obtenemos 10 unidades defectuosas, que las comparamos con el NA y NR de la muestra 2ª de la tabla. Como NA es 12, **LA DECISION SERÍA ACEPTAR EL LOTE.**

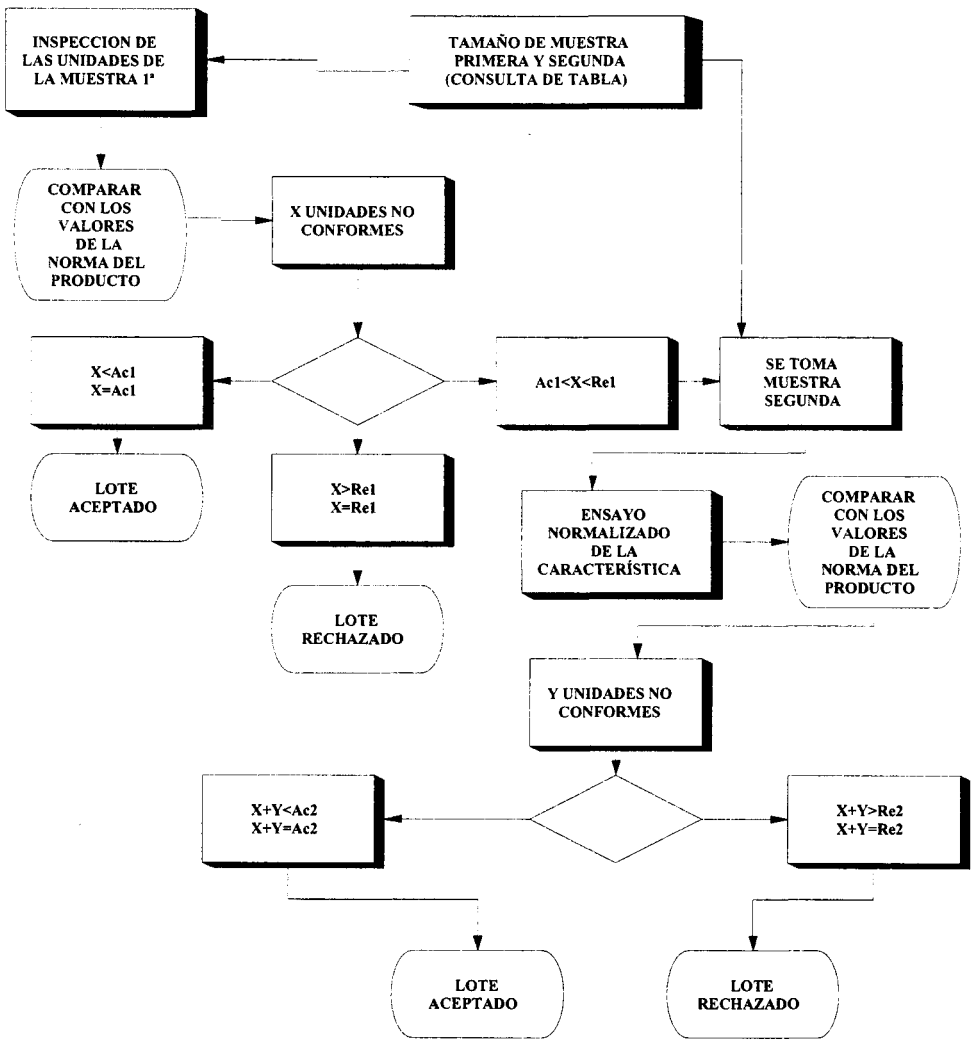


ILUSTRACIÓN 62: Proceso de decisión o rechazo

6.4. CONSULTA DE CURVAS CARACTERÍSTICAS

Una vez tomada la decisión, en función del muestreo llevado a cabo, conviene consultar las curvas características para ver los riesgos de comprador y proveedor.

Para ellos buscamos la curva correspondientes en la norma (Ilustración siguiente).

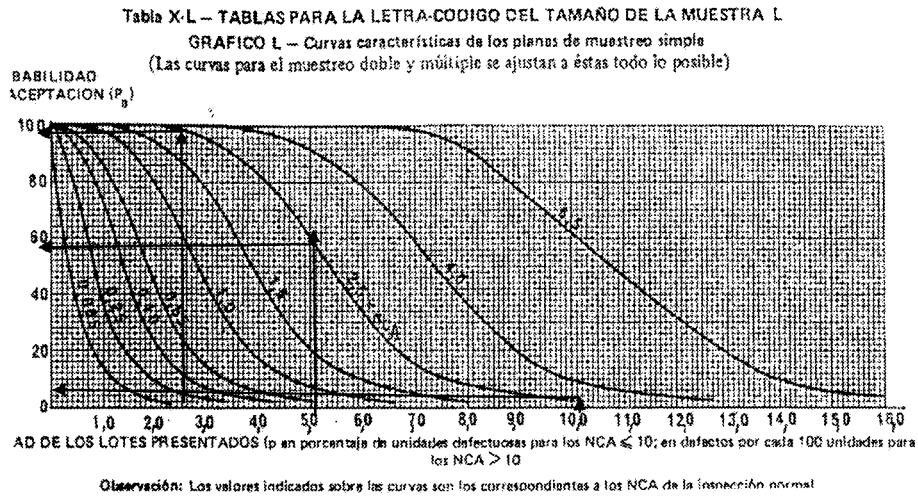


ILUSTRACIÓN 63: Curva característica para la inspección normal simple

Consultando la curva se pueden sacar bastantes datos que interesan a comprador y proveedor.

Probabilidad de aceptación de un lote con una calidad del 5%. Evidentemente, el 5% es una calidad peor que el 2.5 %, y al comprador le interesa ver dicha probabilidad. Para ello levantamos una vertical por el 5% hasta cortar a la curva que nos interesa, consultando con una horizontal a la probabilidad de aceptación. En este caso, esta probabilidad es del 60%. Esto significa que el 60% de los lotes con esta calidad (5%) podrían ser aceptados con el plan de muestreo.

Otra pregunta que se hace el proveedor es que % de lotes con la calidad correcta del 2.5 % le pueden rechazar. Para saberlo, se sube una vertical por el 2.5 hasta cortar a la curva, y luego una horizontal hasta el eje de las Y. Se puede ver que el 95% de los lotes se le aceptarán con este plan de muestreo. El 5% será rechazado, y se llama a este número riesgo de suministrador.

También está el límite de calidad % o riesgo del cliente de aceptar lotes malos, que se fija en el 10%. Esto significa el riesgo de aceptar lotes con una calidad menor del 10%. Trazamos vertical por el 10% y nos da un 7.5 % de probabilidad de aceptar un lote con una calidad inferior al 10%

Los valores de las curvas se pueden consultar numéricamente en la tabla situada debajo de la misma, obteniéndose mayor precisión.

6.5. PROTECCION DEL LIMITE DE CALIDAD Y NCA PARA SERIES DE LOTES

La norma en su apartado 11.6 habla de la protección del límite de calidad. De alguna manera, especifica la calidad para la cual el consumidor tiene un riesgo de aceptación del 5 o del 10% (dependiendo del contrato). Este valor se puede obtener a partir de las curvas características, trazando horizontal por el 5% o por 10 %, o también a partir de tablas calculadas para este fin. La gran utilidad de estas tablas radica en que se pueden usar para aceptar o rechazar lotes individuales, fijando el límite de calidad como criterio para seleccionar el tamaño de muestra y los NCA correspondientes.

TABLA 20: Protección del límite de calidad

Tabla VII-A

PROTECCION DEL LIMITE DE CALIDAD (EN PORCENTAJE DE UNIDADES DEFECTUOSAS) PARA EL CUAL $P_a = 5\%$

(Para muestreo simple en inspección normal)

(Véase el párrafo 11.6)

LC (UNIDADES DEFECTUOSAS) $P_a = 5\%$

Letra código	Tamaño de la muestra	Nivel de calidad aceptable															
		0,010	0,015	0,025	0,040	0,065	0,10	0,15	0,25	0,40	0,65	1,0	1,5	2,5	4,0	6,5	10
A	2															78	
B	3														63		
C	5													45			66
D	8												31			47	60
E	13											21			32	41	50
F	20										14			22	28	34	46
G	32									8,9			14	18	23	30	37
H	50								5,6			9,1	12	15	20	25	32
J	80							3,7			5,8	7,7	9,4	13	16	20	26
K	125						2,4			3,8	5,0	6,2	8,4	11	14	18	24
L	200					1,5			2,4	3,2	3,9	5,3	6,6	8,5	11	15	
M	315				0,95			1,5	2,0	2,5	3,3	4,2	5,4	7,0	9,6		
N	500			0,60			0,95	1,3	1,6	2,1	2,6	3,4	4,4	6,1			
P	800		0,38			0,59	0,79	0,97	1,3	1,6	2,1	2,7	3,8				
Q	1250	0,24			0,38	0,50	0,62	0,84	1,1	1,4	1,8	2,4					
R	2000			0,24	0,32	0,39	0,53	0,66	0,85	1,1	1,5						

Supongamos que queremos para un lote una protección del límite de calidad del 3.8%, para una probabilidad de aceptación del 5% por parte del comprador. Vamos a la tabla y buscamos donde aparece el 3.8 %. Podemos ver que aparece en **K y NCA 0,4** así como en **P y 1.5 de NCA** también en **J y 0.15 de NCA** en **L y 0.65**. Dependiendo de nuestro tamaño de lote, cogeremos un plan de muestreo y lo aplicaremos

para determinar su aceptación o rechazo. Si nuestro lote es por ejemplo de 2000 baldosas, aplicaríamos a 800 baldosas sacadas al azar del mismo un plan de letra P 1,5 de NCA. También podríamos hacer 5 sublotes de 200 baldosas y aplicar un plan de letra L y 0.65 de NCA.

Cuando se trabajan con lotes consecutivos, ya hemos dicho que los NCA corresponden a la calidad media de todos estos lotes juntos. Para establecer la aceptabilidad o no, habría que establecer la calidad real de un lote que ha sido rechazado, para poder hacer la media con otros lotes aceptados.

Para llevar a cabo estas medias, se miran la calidad de los lotes en primera presentación, es decir en muestreo simple.

Supongamos que en una serie de 10 lotes de 1000 baldosas cada uno, se han cogido muestras dadas por la letra J (inspección normal). Supongamos que el número de baldosas extraídas en cada lote ha sido de 80 (página 10 de la norma, tabla II-A). El proveedor y el cliente han pactado un NCA de 2.5%.

Después de la inspección tenemos el siguiente resultado:

TABLA 21: Resultados

Nº de lote	Baldosas no conformes	Decisión	Calidad
Lote 1	2 baldosas malas	Aceptado	1%
Lote 2	5 baldosas malas	Aceptado	2.5
Lote 3	6 baldosas malas	Rechazado	3.2
Lote 4	2 baldosas malas	Aceptado	1
Lote 5	7 baldosas malas	Rechazado	4
Lote 6	0 baldosas malas	Aceptado	0.15
Lote 7	7 baldosas malas	Rechazado	4
Lote 8	4 baldosas malas	Aceptado	2
Lote 9	6 baldosas malas	Rechazado	3.2
Lote 10	4 baldosas malas	Aceptado	2
		Media de calidad	2.3%<2.5

Los valores de calidad aproximada se han obtenido a partir de la tabla considerando que el número baldosas defectuosas se corresponde con el número de aceptación en la columna de la calidad representada.

Según el significado de NCA, esta serie de lotes debería ser aceptada por el comprador, ya que la media de calidad es inferior al 2.5 %. Habría que ver también si se ha pactado algún límite de calidad para cada lote en concreto para la protección del comprador.

En la tabla anterior podemos ver que para este plan de muestreo, el límite de calidad estaría en el 13%, muy lejos del 4% de peor calidad que hemos obtenido.

6.6. LAS NORMAS DE CALIDAD DE LAS BALDOSAS CERÁMICAS

El objetivo de las normas es el de unificar criterios sobre las características y propiedades de un producto y los ensayos que se deben de llevar a cabo para comprobar que las cumple.

Hasta el año 1996 aproximadamente la normativa de las baldosas cerámicas se regia por la serie UNE-EN (aprobadas en su mayor parte en el año 1985), tanto en lo relativo a clasificación, características, etc. como en las normas de ensayo y criterios de aceptación o rechazos de lotes.

Con la aparición de nuevos productos (porcelánico sobre todo), se hizo necesaria una nueva clasificación de las baldosas cerámicas, lo que causó la aparición de un proyecto de norma, la ISO 13006 y las normas de ensayo de la serie UNE-EN ISO 10545, partes 1 a la 17.

A fecha de hoy, las normas de ensayo se encuentra plenamente en vigor, pero la norma de producto y definiciones ISO 13006 que tenia este nombre cuando era proyecto ha pasado a publicarse con el nombre ISO 14411.

Las normas de la serie 67 fueron confeccionadas por el CEN (comité Europeo de Normalización). Este comité constaba de:

- Un Comité director donde existe un representante de cada país miembro
- Una secretaría general
- Grupos de trabajo, estando las secretaría a cargo de los países miembros

El CEN se apoya ampliamente en los trabajos de la Organización Internacional de Normalización ISO, y una de las misiones es la de adaptar las normas internacionales al espacio Europeo, para su posterior aplicación.

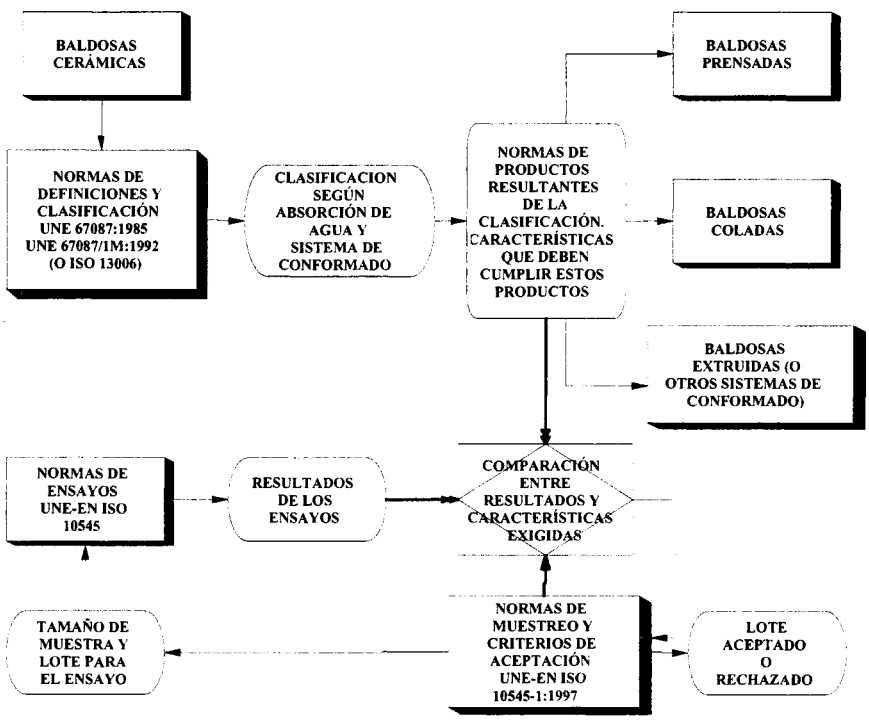
Las normas de la serie ISO implantadas a fecha de hoy, las desarrollo el Comité Técnico ISO/TC 189 «Baldosas Cerámicas» siendo adoptadas como Norma Europea por el Comité Técnico CEN/TC 67 «Baldosas Cerámicas». Estas normas a su vez han sido adoptadas por cada país. En concreto en España AENOR (Asociación Española de Normalización) las tradujo y publicó como normas UNE-EN-ISO, derogando las normas contradictorias.

6.7. TIPOS DE NORMAS

En primer lugar hay que comprender que las normas pertenecen a varias categorías, dependiendo de su contenido. Estas categorías son:

- Normas de **definiciones, criterios de clasificación** y **las características generales** que deben de cumplir según su uso los productos de referencia (En este caso, las baldosas cerámicas)
- **Normas de producto**, en las que se desarrollan las características concretas de cada uno de los productos que resultan de la clasificación dada por la norma anterior. Podríamos decir que es una repetición de la norma anterior, pero personalizada para un producto concreto. En esta norma se informa igualmente de cómo determinar las características exigidas, haciendo referencia a las normas de ensayo correspondientes
- **Normas de ensayo**, donde se nos dice la forma concreta de llevar a cabo las pruebas y procesos para determinar una característica de un producto
- **Normas de muestreo y aceptación** de un lote, en la que se nos dice cómo coger una muestra para la realización de los ensayos, y cómo se toma la decisión de aceptar o rechazar un lote dependiendo de los resultados de los ensayos realizados según la norma anterior.
- Otras normas complementarias, como la de los adhesivos para la colocación de las baldosas.

La estructura pues de las normas de cualquier producto y en concreto de las baldosas se pueden resumir en el cuadro siguiente,



La interpretación del cuadro es como sigue:

- Dada una baldosa, se clasifica según la norma de clasificación, en el grupo de productos que le corresponda.
- Una vez determinado el grupo de productos, se consulta la norma correspondiente a dicho producto. En esa norma se indicará qué características debe de cumplir
- Una vez conocidas las características que debe cumplir, se miran las normas de los ensayos correspondientes. Para seleccionar el tamaño de la muestra a ensayar se consulta la norma de muestreo y condiciones de aceptación.
- Se realizan los ensayos según la norma en las piezas que componen la muestra tomada
- Se miran el número de baldosas que no cumplen las características exigidas en la norma de producto, y se consulta con la norma de aceptación o rechazo. De resultas de la consulta de esta norma, se pueden tomar tres decisiones:
- El lote se acepta
- El lote se rechaza
- Se toma una segunda muestra, para volver a realizar los ensayos. Con los datos de la primera muestra y la segunda, se toma la decisión
- El lote se acepta
- El lote se rechaza

En la normativa de baldosas antiguas, estas normas están publicadas separadamente, mientras que en las normas de la serie ISO solamente hay dos normas fundamentales:

La norma ISO 14411 que abarca las normas de características, definiciones, clasificación y normas de producto de las antiguas normas

La normas UNE-EN-ISO 10545 partes 1 a la 17. Esta norma además de establecer cada uno de los ensayos que se deben de llevar a cabo, incluye la norma de criterios de muestreo y aceptación de lotes

6.8. COMPARATIVA ENTRE LA NORMATIVA ANTIGUA Y LA ACTUAL

6.8.1. NORMAS (CLASIFICACIÓN, CARACTERÍSTICAS Y MARCADO)

La normativa antigua establecía la **norma EN 87 (UNE 67-087** año 1985 y modificada en 1991) como referencia para clasificar las baldosas, además de definirlas. Actualmente se usa de hecho la **norma ISO 14411**

TABLA 22: Normativa antigua y moderna de clasificación, características y marcado de baldosas

Normas de clasificación, características y marcado	Normativa antigua	Normativa actual
Clasifican las baldosas en grupos, dan ciertas definiciones básicas y señalan las características que deben cumplir	En 87 (UNE 67-087) Años 1985, 1991)	ISO 14411

La clasificación de las baldosas se hace por dos criterios:

El sistema de conformado. La norma EN 87 establece tres grupos A (extrusión), B (prensado en seco) y C (colado). La norma ISO 14411 establece los mismos grupos A y B, pero en el C hace referencia a otros sistemas de conformado, y además no desarrolla la norma de estos productos.

La absorción de agua. La absorción de agua se entiende como la porosidad abierta que tiene la baldosa y que hace que sumergida en agua según la norma correspondiente, incremente su peso por su absorción. Las dos normas son diferentes, ya que la norma antigua solo considera la baja absorción de agua por debajo del 3%, sin hacer subgrupos como establece la ISO 14411.

La tablas que resumen estas clasificaciones quedan como sigue:

TABLA 23: Clasificación de las baldosas cerámicas según la norma EN 87 (UNE 67-087)

Absorción agua / Moldeo	Grupo I %A.a ≤ 3%	Grupo II a 3<%A.a≤6	Grupo II b 6<%A.a≤10	Grupo III %A.a > 10 %
A Extruídas	Grupo AI EN 121	Grupo AIIa EN 186	Grupo AIIb EN 187	Grupo AIII EN 188
B Prensadas en seco	Grupo BI EN 176	Grupo BIIa EN 177	Grupo BIIb EN 178	Grupo BIII EN 179
C Coladas	Grupo CI	Grupo CIIa	Grupo CIIb	Grupo CIII

En cada grupo resultante de la clasificación, se designan las normas europeas aplicables.

La nueva norma establece tres grupos:

- Baja absorción de agua. Pone el límite en el 3% y subdivide en dos grupos como se ve en la tabla.
- Media absorción de agua. Pone el límite inferior en 3 y el superior en 10% y también hace dos subgrupos

- Alta absorción de agua. Considera los productos con más del 10% de absorción de agua

TABLA 24: Clasificación de las baldosas cerámicas según la ISO 14411

	Grupo I E ≤ 3%	Grupo IIa 3% ≤ E < 6%	Grupo IIb 6% ≤ E < 10%	Grupo III E > 10%
A. Extrusión	Grupo AI ISO 14411 Anexo A	Grupo AIIa-1 ISO 14411 Anexo B	Grupo AIIb-1 ISO 14411 Anexo D	Grupo AIII ISO 14411 Anexo F
		Grupo AIIa-2 ISO 14411 Anexo C	Grupo AIIb-2 ISO 14411 Anexo E	
B. Prensado en seco	Grupo BIa E ≤ 0,5% ISO 14411 Anexo G	Grupo BIIa ISO 14411 Anexo J	Grupo BIIb ISO 14411 Anexo K	Grupo BIII ISO 14411 Anexo L Solo revestimiento
	Grupo BIb 0,5% < E ≤ 3% ISO 14411 Anexo H			
C. Otros procesos de conformación	Grupo CI No reflejadas en la norma	Grupo CIIa No reflejadas en la norma	Grupo CIIb No reflejadas en la norma	Grupo CIII No reflejadas en la norma

Las normas de producto de cada grupo de clasificación aparecen como anexos de la ISO 14411, de las letras A a la L. Como puede verse, el grupo C trata de «Otros métodos de conformado» incluyendo el colado o cualquier otra tecnología diferente del prensado en seco o extrusionado.

El resto de las normas son muy similares, estableciendo por ejemplo definiciones bastante parecidas:

- Baldosa. Son placas de poco grosor fabricadas en arcilla, sílice, fundentes, colorantes y otros materiales, generalmente utilizadas como revestimientos para suelos, paredes, y fachadas. Se fabrican mediante molturación, tamizado, amasado, humidificación, y se moldean por prensado, extrusionado, colado u otros procedimiento generalmente a temperatura ambiente. Seguidamente son secadas y esmaltadas a altas temperaturas. Las baldosas cerámicas pueden ser esmaltadas GL, no esmaltadas UGL o engobadas y son incombustibles e inalterables por la luz
- Esmalte. Es una cubierta vitrificada y prácticamente impermeable

- Engobe. Es una cubierta a base de arcilla con un acabado mate que puede ser permeable o impermeable
- Monococción y bicocción. Hace referencia al proceso de cocción si se hace una sola vez o dos veces.
- Baldosas prensadas, extruídas y coladas. Hace referencia al proceso de obtención de la pieza, o sea el conformado, y les asigna unas letras A,B,C como veremos en el apartado siguiente.
- La absorción de agua de las baldosas. Hace referencia al poro abierto de la baldosa y a la norma para medirla
- Los conceptos de medida nominal, medida de coordinación, medida real y tolerancias, medidas modulares y no modulares

La norma ISO 14411 establece las medidas de las baldosas. Estas medidas pueden ser:

- Modulares, basadas en la norma ISO 1006/1 en las que las medidas se refieren al modulo M donde M= 100 mm. Se refieren a baldosas con área mayor de 900 mm^2. los ejemplos más importantes de medidas modulares se citan en los anexos de cada tipo de productos.
- No modulares, en las que las medidas se expresan libremente. Son las más comúnmente empleadas en los países miembros del comité de normalización europeo.
- Además se expresan en la norma los siguientes conceptos de medidas:

  - Medida de fabricación, que es la prevista en para la fabricación de una baldosa W
  - Medida actual o real. Es la medida resultante del proceso de determinación de la longitud de una baldosa dada
  - Medida de coordinación C. Es la suma de la medida de fabricación más la junta J

Veamos en la página siguiente unas figuras que ilustran estas definiciones:

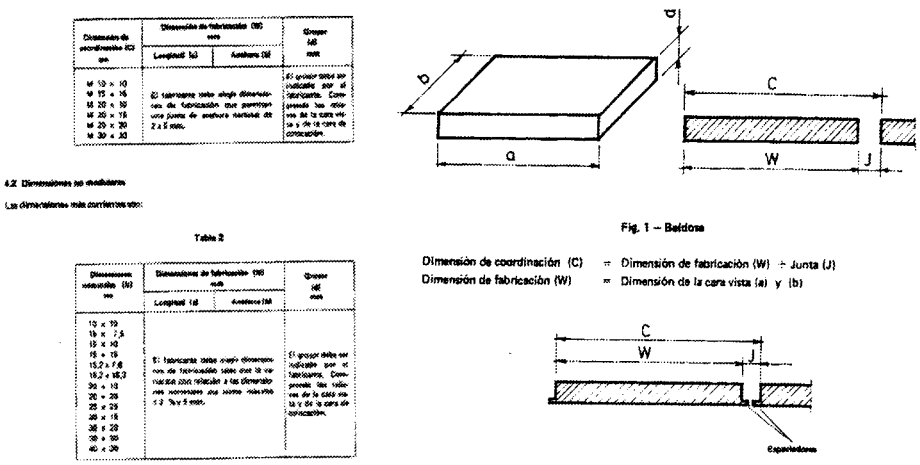


ILUSTRACIÓN 64: Ejemplo definiciones

6.9. NORMA UNE 67-087-85. DEFINICIONES, CLASIFICACIÓN, CARACTERÍSTICAS Y MARCADO

Esta norma define en primer lugar conceptos tan conocidos como:

- Baldosa
- Esmalte
- Engobe
- Monococción y bicocción
- Baldosas prensadas, extruídas y coladas
- La absorción de agua de las baldosas
- Los conceptos de medida nominal, medida de coordinación, medida real y tolerancias, medidas modulares y no modulares

A continuación clasifica las baldosas en varios grupos según el sistema de conformado y la absorción de agua, obteniendo la tabla ya vista.

En el punto 5 indica las características importantes de las baldosas desde el punto de vista dimensional, de las propiedades físicas y de las propiedades químicas, además de decir con qué norma se debe de hacer el ensayo:

Hay que hacer notar que las normas de ensayo a que hace referencia, han cambiado substancialmente, siendo sustituidas por la serie ISO 10545 como ya hemos comentado.

El siguiente punto de la norma hace referencia al marcado de las baldosas y a su identificación, y por último habla de los sistemas de hacer pedidos, y a la norma de muestreo y condiciones de aceptación aplicables a las baldosas.

Esta norma de clasificación de baldosas ha sido sustituida por la norma ISO 13006 que establece la otra tabla de clasificación que también hemos expuesto.

En las páginas siguientes vamos a ver las normas de productos que marca la norma antigua que es la que se aplica en la mayoría de las empresas, y aplicaremos la norma nueva de ensayo ISO 10545 a la que se ha hecho referencia anteriormente.

6.10. NORMAS DE MUESTREO Y CONDICIONES DE ACEPTACIÓN

Estas normas permiten tomar muestras de acuerdo con un procedimiento, y una vez realizados los ensayos nos informan si el lote se ha aceptado o rechazado.

Las normas antiguas de muestreo y condiciones de aceptación son la UNE 67-163-85 y ha sido sustituida con pocas variaciones por la UNE-EN-ISO-10545-1:1997.

Esta última norma es la primera de la serie ISO-10545 que vienen a sustituir los procedimientos de ensayo de la serie UNE antigua. Las partes de que trata son las siguientes:

1. Objeto y campo de aplicación
2. Definiciones
   - 2.1. Pedido: Cantidad de baldosas pedidas de una sola vez
   - 2.2. Partida: Cantidad de baldosas suministradas en un período de dos días
   - 2.3. Sub partida homogénea: Sub partida de baldosas de un mismo fabricante y producidas con propiedades presuntamente uniformes
   - 2.4. Lote de inspección: Cantidad de baldosas sometidas a ensayo, de un mismo fabricante producidas con propiedades presuntamente uniformes
   - 2.5. Muestra: Número de baldosas tomadas de un lote de inspección
   - 2.6. Tamaño de muestra: Número de baldosas sometidas a ensayo para cada propiedad
   - 2.7. Requisito: Característica exigible para una propiedad según la norma de producto correspondiente
   - 2.8. Unidad no conforme: Baldosa que no cumple el requisito para la propiedad en cuestión

La interpretación de estas definiciones es muy importante. Cuando un cliente hace un pedido a la empresa, este conjunto de baldosas es el pedido. Este pedido puede ser de baldosas todas iguales o varios tipos de baldosas. Las baldosas suministradas en el plazo de dos días se denominan partida.

Cuando llega una partida al consumidor, el grupo de baldosas que tienen las mismas propiedades (dimensiones, pavimento o revestimiento, esmalte, etc.) se denominan subpartida. De esta subpartida es de la que se toma el lote de inspección. Este lote de inspección se suele tomar por acuerdo entre fabricante y cliente, y son un con-

junto de baldosas que van a ser de donde se formarán las muestras para los ensayos normalizados.

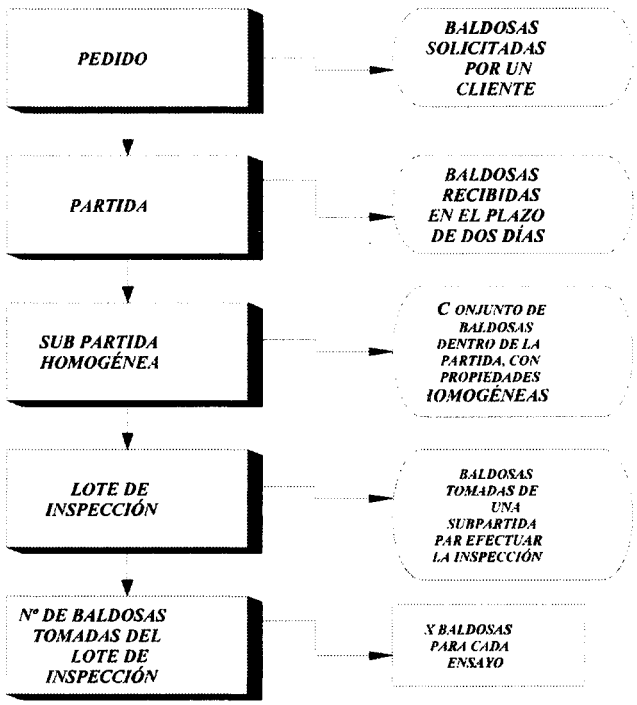


ILUSTRACIÓN 65: Esquema de control

Las muestras se toman por azar del lote de inspección, y de aquí se van cogiendo submuestras de u tamaño dado por la norma de la característica ensayada (Por ejemplo para la absorción de agua se toman 10 baldosas).

Es muy frecuente que el tamaño de muestra se exprese en m^2 y las baldosas que se deben de tomar se calculan en función de la superficie unitaria.

La norma una vez definidos estos conceptos, expone otros puntos como son:

3. Fundamento
4. Formación de lotes de inspección
5. Amplitud de la inspección
6. Muestreo
   6.1.El lugar de la toma de muestras

6.2. Quién estará presente
6.3. Como determinar el nº de baldosas para cada propiedad
7. Inspección
7.1. Cómo se debe de hacer la inspección
7.2. Resultados y la manera de evaluarlos
8. Determinación de la aceptabilidad de los lotes de inspección
8.1. Define la inspección por atributos, con varios de los subapartados que luego trataremos
8.2. Define la inspección por valor medio
9. Indica cómo realizar el informe de la inspección

Uno de los apartados más importantes es el 8, donde se indica cómo aceptar o rechazar los lotes. Este apartado se basa en la tabla expuesta a continuación, donde se pueden ver una serie de columnas.

Como veremos en la norma de inspección por atributos UNE 66-020-73, esta tabla es una particularización de la misma para un nivel de calidad aceptable del 2.5 % en inspección doble. La norma ISO ha especificado pues un nivel de calidad determinado.

- La primera columna indica la propiedad medida
- La segunda indica el tamaño de muestra
- Las columnas 3,4,5,6, indican inspección por atributos, indicando el nº de aceptación, y el nº de rechazo de la muestra inicial, y el nº de aceptación, y el nº de rechazo de la muestra segunda.
- Las columnas 7,8,9,10 indican lo mismo para inspección por valor medio
- La columna 11 indica el método o norma de ensayo para esa propiedad

- El apartado 8.1.1. dice literalmente: Cuando el número de unidades no conformes de la muestra inicial sea igual o menor que el número de aceptación Ac indicado en la columna 3 de la tabla 1, se considera aceptable el lote de inspección del que se haya tomado la muestra (Esto para cada característica)
- El apartado 8.1.2. dice que cuando el nº de unidades no conformes sea igual o mayor que el nº de rechazo Re_1, el lote puede ser rechazado
- El apartado 8.1.3. dice que cuando el nº de unidades no conformes esté entre el nº de aceptación y el de rechazo, se deberá tomar una segunda muestra
- El apartado 8.1.4. dice que se volverán a realizar los ensayos y se totalizarán el nº de unidades no conformes de la muestra inicial y segunda.
- El apartado 8.1.5.vuelve a aplicar el criterio del apartado 8.1.2. al nº total de unidades no conformes, comparándolo con el número de aceptación Ac_2. De esta manera el lote se acepta si el nº de unidades no conformes es igual o menor que el nº de aceptación.

- El apartado 8.1.6. compara el nº de unidades no conformes con el nº de rechazo Re_2, diciendo que si el nº de unidades no conformes es mayor o igual que este número, el lote puede rechazarse.

Tabla 1
Procedimiento de muestreo

1	2		3	4	5	6	7	8	9	10	11
	Tamaño de muestra		Inspección por atributos, si se requiere				Inspección por el valor medio, si se requiere				Método de ensayo
Características			Muestra inicial		Muestra inicial + segunda		Muestra inicial		Muestra inicial + segunda		
	Inicial	Segunda	Número de aceptación Ac_1	Número de rechazo Re_1	Número de aceptación Ac_2	Número de rechazo Re_2	Condición de aceptación	Condición para la toma de segunda muestra	Condición de aceptación	Condición de rechazo	Parte de ISO 10545
Dimensiones[1]	10	10	0	2	1	2	–	–	–	–	2
Calidad superficial[2]	30	30	1	3	3	4	–	–	–	–	2
	40	40	1	4	4	5	–	–	–	–	
	50	50	2	5	5	6	–	–	–	–	
	60	60	2	5	6	7	–	–	–	–	
	70	70	2	6	7	8	–	–	–	–	
	80	80	3	7	8	9	–	–	–	–	
	90	90	4	8	9	10	–	–	–	–	
	100	100	4	9	10	11	–	–	–	–	
	1m²	1m²	4%	9%	5%	> 5%	–	–	–	–	
Absorción de agua[4]	5[5]	5[6]	0	2	1	2	$\bar{x}_1 > L$[5]	$\bar{x}_1 < L$	$\bar{x}_2 > L$	$\bar{x}_2 < L$	3
	10	10	0	2	1	2	$\bar{x}_1 < U$[6]	$\bar{x}_1 > U$	$\bar{x}_2 < U$	$\bar{x}_2 > U$	
Resistencia a la flexión[4]	7[7]	7[7]	0	2	1	2	$\bar{x}_1 > L$	$\bar{x}_1 < L$	$\bar{x}_2 > L$	$\bar{x}_2 < L$	4
	10	10	0	2	1	2					
Carga de rotura[4]	7[8]	7[8]	0	2	1	2	$\bar{x}_1 > L$	$\bar{x}_1 < L$	$\bar{x}_2 > L$	$\bar{x}_2 < L$	4
	10	10	0	2	1	2					
Abrasión profunda UGL	5	5	0	2[9]	1[10]	2[10]	–	–	–	–	6
Coeficiente de dilatación térmica lineal	2	2	0	2[9]	1[9]	2[9]	–	–	–	–	8
Resistencia al cuarteo	5	5	0	2	1	2	–	–	–	–	11
Resistencia química[4]	5	5	0	2	1	2	–	–	–	–	13
Resistencia a las manchas[4]	5	5	0	2	1	2	–	–	–	–	14
Resistencia a la helada[4]	10	–	0	1	–	–	–	–	–	–	12
Resistencia al choque térmico	5	5	0	2	1	2	–	–	–	–	9
Expansión por humedad	5	–	–	Atribución por declaración del fabricante							10

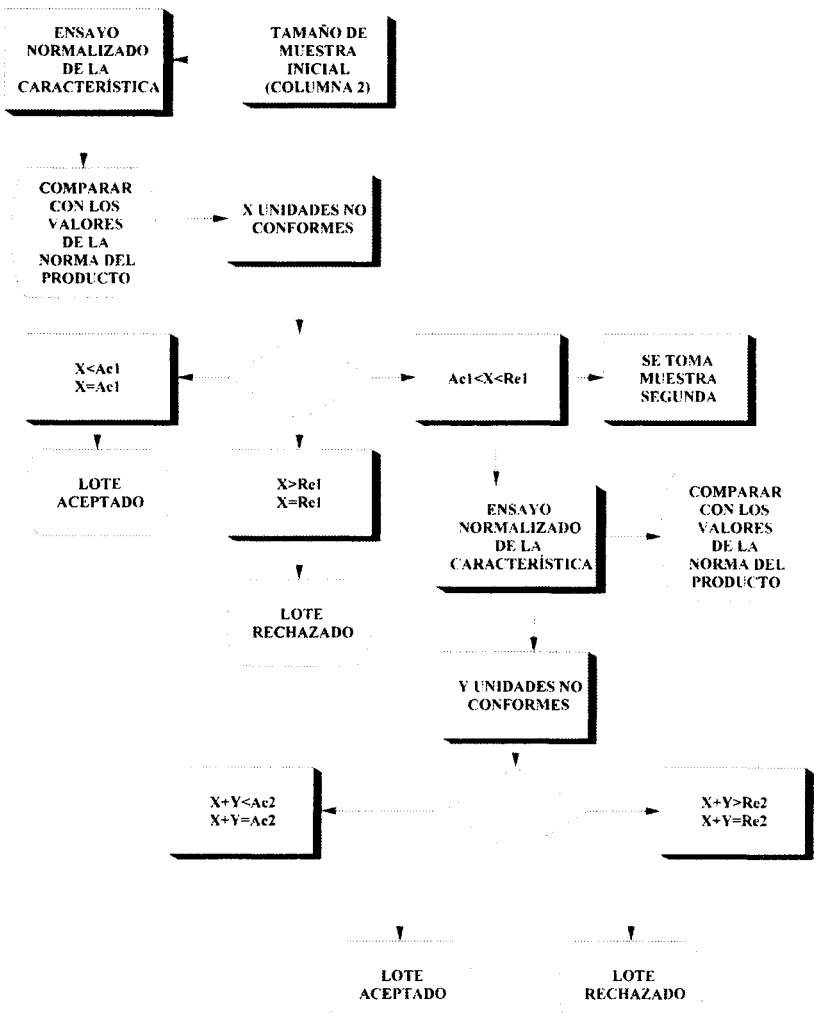
1	2		3	4	5	6	7	8	9	10	11
	Tamaño de muestra		Inspección por atributos, si se requiere				Inspección por el valor medio, si se requiere				Método de ensayo
Características			Muestra inicial		Muestra inicial + segunda		Muestra inicial		Muestra inicial + segunda		
	Inicial	Segunda	Número de aceptación Ac_1	Número de rechazo Re_1	Número de aceptación Ac_2	Número de rechazo Re_2	Condición de aceptación	Condición para la toma de segunda muestra	Condición de aceptación	Condición de rechazo	Parte de ISO 10545
Resistencia a la abrasión GL[11]	11	–	–	Atribución por declaración del fabricante							7
Coeficiente de fricción	–[12]	–	–	Atribución por declaración del fabricante							17
Diferencias de color	5	–	–	Atribución por declaración del fabricante							16
Resistencia al impacto	5	–	–	Atribución por declaración del fabricante							5
Emisión de plomo y cadmio	5	–	–	Atribución por declaración del fabricante							15

1) Solamente para baldosas con una superficie ≥ 4 cm².
2) Someter al ensayo al menos 1 m² o un mínimo de 30 baldosas. Cualquiera que sea el número de baldosas comprendidas en 1 m² la muestra debe ser redondeada a la decena superior. La conformidad con el nivel de calidad aceptable (NCA) del 2,5 % según ISO 2859-1:1989, *Reglas de muestreo para la inspección por atributos. Parte 1: Planes de muestreo para inspecciones lote por lote, tabulados según el nivel de calidad aceptable (NCA)*, o según ISO 3951:1989, *Reglas y tablas de muestreo para la inspección por variables de los porcentajes de unidades defectuosas*, es una alternativa válida al procedimiento de la tabla 1
3) El tamaño de la muestra depende de la dimensión de la baldosa.
4) Solo para baldosas con área superficial ≥ 0,04 m². En el caso de baldosas con una masa < 50 g, debe tomarse un número suficiente como para formar cinco probetas, cada una con un peso comprendido entre 50 g y 100 g.
5) L: Límite inferior de especificación
6) L: Límite superior de especificación.
7) Solo para baldosas con longitud ≥ 48 mm.
8) Número de medidas.
9) Número de probetas.
10) Por cada solución de ensayo
11) No hay doble toma de muestras para estas propiedades.
12) El tamaño de muestra varía con el método de ensayo

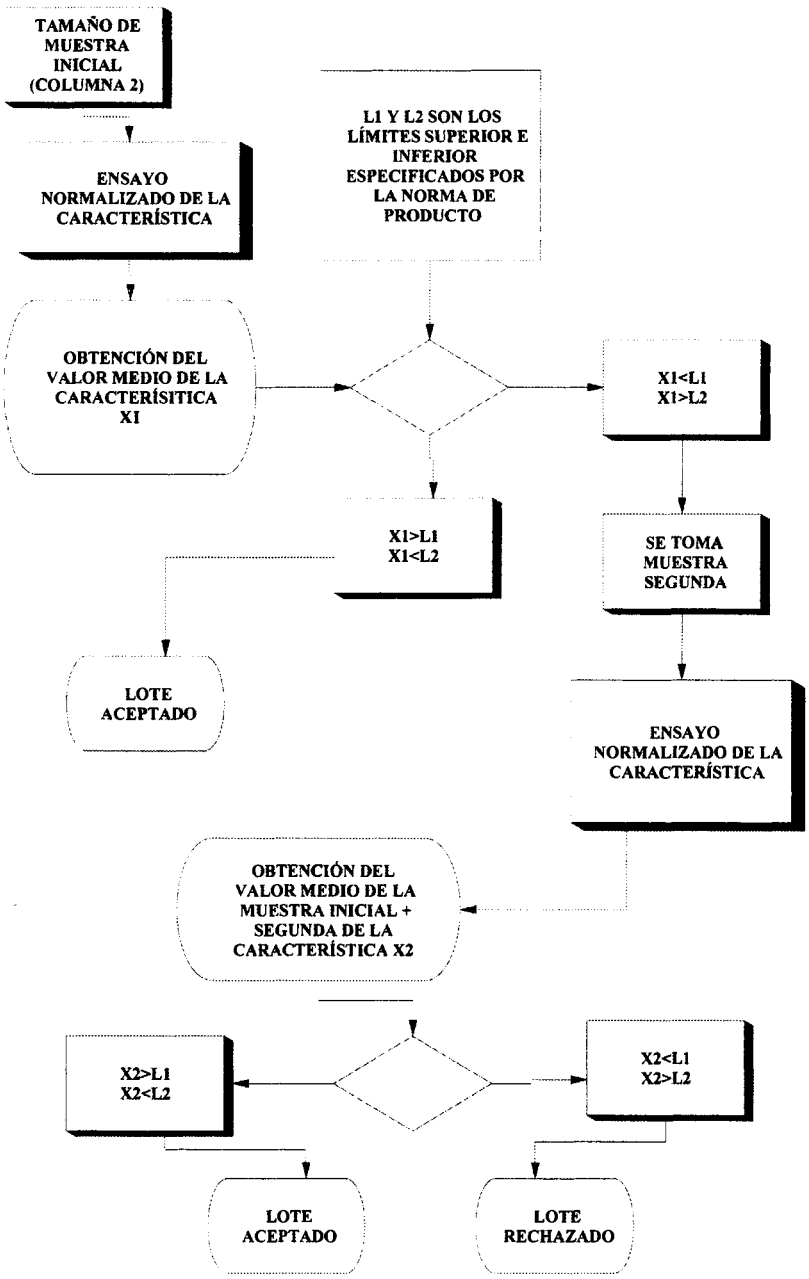
En la inspección por valor medio, ya no se habla de unidades no conformes o conformes, sino que se compara el valor medio de los resultados obtenidos con el valor indicado en las columnas 7, 8 9, 10.

- 8.2.1.Si el valor medio de los resultados de los ensayos (x_1) cumple los requisitos, el lote se considera aceptado
- 8.2.2.Si el valor medio no cumple los requisitos, se deberá tomar una segunda muestra del mismo tamaño que la primera.
- 8.2.3.Si el valor medio combinado de la muestra primera y segunda (x_2) cumple los requisitos, el lote queda aceptado
- 8.2.4.Si el valor medio no cumple los requisitos, el lote puede rechazarse.

En el diagrama siguiente se puede ver el proceso por el cual se puede llegar a aceptar o rechazar un lote aplicando esta norma de aplicación de criterios



Para los ensayos por valor medio, el proceso es prácticamente el mismo, salvo que no existe el concepto de unidades no conformes, sino el valor medio de la muestra que se debe de comparar con el valor que marca la tabla.



Con estos cuadros queda cerrado el sistema de normas de las baldosas cerámicas en cuanto a sus características y aceptación de lotes.

Lo siguiente que vamos a ver son las normas de ensayo, que como ya hemos dicho han sido cambiadas hace un par de años, desde la serie EN antigua a la moderna ISO 10545 que aún tiene algunas normas en preparación.

6.11. NORMAS DE ENSAYO DE LAS CARACTERÍSTICAS

En el siguiente cuadro pueden verse el resumen de los dos grupos de normas:

TABLA 25: Ensayos

UNE 67-098-85. Baldosas cerámicas. Determinación de las características dimensionales y aspecto superficial. UNE 67-099-85. Baldosas cerámicas. Determinación de la absorción de agua. UNE 67-100-85. Baldosas cerámicas. Determinación de la resistencia a la flexión. UNE 67-101-85. Baldosas cerámicas. Determinación de la dureza al rayado de la superficie según Mohs. UNE 67-102-85. Baldosas cerámicas. Determinación de resistencia a la abrasión profunda. Baldosas no esmaltadas UNE 67-154-85. Baldosas cerámicas. Determinación de resistencia a la abrasión. Baldosas esmaltadas. UNE 67-103-85. Baldosas cerámicas. Determinación de la dilatación térmica lineal UNE 67-104-85. Baldosas cerámicas. Determinación de la resistencia al choque térmico UNE 67-105-85. Baldosas cerámicas. Determinación de la resistencia al cuarteo. Baldosas esmaltadas.	ANTECEDENTES El texto de la norma internacional del Comité Técnico ISO/TC 189 Baldosas cerámicas» de la Organización Internacional de Normalización (ISO) ha sido adoptado como Norma Europea por el Comité Técnico CENITC 67 «Baldosas cerámicas» cuya secretaría está desempeñada por UNI. Esta Norma Europea sustituye a la Norma EN 163:1991. La Norma EN ISO 10545, bajo el titulo general Baldosas cerámicas, contiene las siguientes partes: Parte 1: Muestreo y criterios de aceptación. Parte 2 Determinación de las dimensiones y del aspecto superficial Parte 3: Determinación de la absorción de agua, de la porosidad abierta, de la densidad relativa aparente y de la densidad aparente. Parte 4: Determinación de la resistencia a la flexión y de la carga de rotura. Parte 5: Determinación de la resistencia al impacto por medición del coeficiente de restitución

TABLA 25: Ensayos (continuación)

UNE 67-202-85. Baldosas cerámicas. Determinación de la resistencia a la helada. UNE 67-155-85. Baldosas cerámicas. Determinación de la expansión por humedad usando agua hirviendo. Baldosas no esmaltadas UNE 67-106-85. Baldosas cerámicas. Determinación de la resistencia química. Baldosas no esmaltadas UNE 67-122-85. Baldosas cerámicas. Determinación de la resistencia química. Baldosas esmaltadas UNE 67-163-85. Baldosas cerámicas. Muestreo y condiciones de aceptación	Parte 6: Determinación de. de la resistencia a la abrasión profunda de las baldosas no esmaltadas Parte 7: Determinación de la resistencia a la abrasión superficial de las baldosas esmaltadas. Parte 8: Determinación de la dilatación térmica lineal. Parte 9: Determinación de la resistencia al choque térmico. Parte 10: Determinación de la dilatación por humedad Parte 11: Determinación de la resistencia al cuarteo de las baldosas esmaltadas Parte 12: Determinación de la resistencia a la helada. Parte 13: Determinación de la resistencia química. Parte 14: Determinación de la resistencia a las manchas. Parte 15: Determinación de la emisión de plomo y cadmio de las baldosas esmaltadas Parte 16: Determinación de las pequeñas diferencias de color. Parte 17: Determinación del coeficiente de fricción.

A la vista del cuadro puede deducirse que son prácticamente las mismas normas con la inclusión de algunas nuevas en la serie ISO, como las 15, 16 y 17 . De estas tres, la más importante es la del coeficiente de fricción, cada día más exigida en las baldosas destinadas a la exportación.

Aunque el título de las normas son las mismas, hay diferencias que más adelante analizaremos en el tratamiento de cada norma en concreto.

7 EJEMPLOS Y APLICACIÓN

7.1. SEGUIMIENTO DE LOS EJERCICIOS

El esquema que de seguimiento de los ejercicios está relacionado con el orden de obtención de resultados y cuales son las características principales a estudiar, es decir siempre empezar con una análisis descriptivo de los datos, para seguir, según el parámetro de estudio, con gráficos individualizados, gráficos de control de calidad por variables, gráficos de control de calidad por atributos o estudio de capacidad con los índices correspondientes.

En el caso que se quiera primero saber cuales son los defectos que más afectan a al proceso (tanto defectos como efectos en general), deberá realizarse el diagrama de Pareto, antes de elegir el defecto a eliminar.

Un esquema general para el control de calidad de unos parámetros podría ser en general, conocer el caso de estudio de forma que se sepa:

- Definir las variables. Saber de que tipo son (continuas, discretas, etc...)
- Saber si es un proceso de control inicial o estudio a partir de un estándar.
- Saber si se quiere solucionar un problema en concreto, varios a la vez o hay que definir primero el problema.
- Si es referido a un proceso, a una máquina de medida o a al entorno.

Los distintos casos podrían ser (con los posibles pasos a realizar):

1) Máquina de medidas (variable) y sin un estándar previo:

- Tormenta de Ideas (para ver cuales son los posibles errores).
- Diagrama Causa – Efecto (de las posibles causas de error ver cual es el culpable).
- Diagrama de control por variables sin estándar (para ver si está bajo control).
- Para finalizar, estudio de capacidad de la máquina (con estudio de los índices).

2) Máquina de medidas (variable) con un estándar previo (nos servirá para comparar máquinas, la compra de una nueva o solucionar problemas de esta):

- Tormenta de Ideas (para ver cuales son los posibles errores).
- Diagrama Causa – Efecto (de las posibles causas de error ver cual es el culpable).
- Diagrama de control por variables con estándar de patrón (para ver si está bajo control).
- Para finalizar, estudio de capacidad de la máquina (con estudio de los índices).

3) Máquina de medidas (atributos) la cual tiene unos límites de tolerancia, es decir, si cumple o no los requisitos. Un ejemplo sería máquina que corta piezas (o las rectifica como ocurre en el sector cerámico) y son separadas las que están dentro de límites y las que no lo están:

- Tormenta de Ideas (para ver cuales son los posibles errores).
- Diagrama Causa – Efecto (de las posibles causas de error ver cual es el culpable).
- Diagrama de control por atributos (para ver si está bajo control), diferenciando el tipo de gráfico (u, c, p, np) según la necesidad de conocer el porcentaje de error o el número de errores totales o por pieza.
- Para finalizar, estudio de capacidad de la máquina (con estudio de los índices).

4) Se producen una serie de defectos en mi producción y tengo que eliminarlos. Queda claro que yo no puedo eliminar todos los defectos a la vez, por lo que tendré que ir ordenándolos de mayor a menor importancia y según el coste eliminar los de mayor porcentaje de defectos en producción.

- Tormenta de Ideas (para ver cuales son los posibles errores).
- Diagrama Causa – Efecto (de las posibles causas de error ver cual es el culpable).
- Diagrama de Pareto para ordenar los defectos y eliminar según el porcentaje de error que estos produzcan.

Estos sólo son algunos de los posibles casos en los que es necesaria la aplicación del control estadístico de calidad. En muchos casos es solo necesaria la utilización de una de las herramientas de calidad como el Diagrama de Pareto o Diagramas de Control por Varibles o Atributos que nos servirán para ver los defectos que más afectan al proceso o ver si el proceso está bajo control.

En cada uno de ellos es necesario previamente un estudio descriptivo de los datos, incluso cabe la posibilidad de la utilización de los gráficos individualizados (CUSUM, Medias Móviles, Multivariables).

7.2. APLICACIONES DE CONTROL ESTADÍSTICO DE LA CALIDAD

Previamente al inicio del estudio, tenemos que comentar el software utilizado, el Statgraphics. En estos ejemplos, se ha utilizado de forma general la versión 4.1:

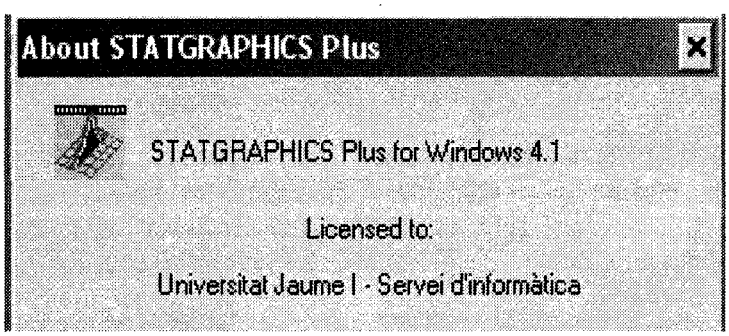


ILUSTRACIÓN 66: Versión Statgraphics

Otra opción es la utilización de la versión en español 5.1: STATGRAPHICS Plus Version 5.1, Abril, 2002

En este capítulo vamos a ver un ejemplo de cada uno de las distintas partes que componen el estudio del Control Estadístico de la Calidad.

Empezaremos por la necesidad del estudio previo descriptivo con el análisis exploratorio de los datos y las medidas de posición, dispersión y forma.

A continuación, posterior a la descripción de los datos a analizar, entraremos en el control estadístico de calidad, es decir sus distintos componentes. Primeramente control del proceso, indicar los factores que afectan a la calidad, y los diagramas Causa-Efecto y el de Pareto. Seguiremos con los gráficos de variables y atributos, para acabar con los gráficos de capacidad y los índices correspondientes.

Los datos que serán utilizados en cada uno de los ejemplos, se encuentran en el cuarto apartado del libro en forma de tabla. Será un fichero general que nos servirá como aplicación de todos los casos.

El caso de estudio que utilizamos en este apartado, es para datos de una empresa cerámica en la cual se están estudiando distintos parámetros de los productos obtenidos. Son piezas cerámicas de revestimiento de formato rectangular de medidas nominales de 20x15 centímetros. El muestreo está realizado en planta durante varios días

de producción, con variación en el número de piezas muestreadas por día y hora (por lo tanto podemos agrupar de la forma que queramos en los gráficos de control).

Las variables de estudio serán:

a) L1: Longitud del lado corto de la pieza cerámica (cm)
b) L2: Longitud del lado largo de la pieza cerámica (cm)
c) Dap_crudo: Densidad aparente de la pieza cruda (g/cc)
d) Tipos de defectos
- Sin defectos
- Descuadre
- Calibre
- Despuntado
- Ahojado
- Planaridad

e) Inspeccionados: número de piezas inspeccionadas
f) Espesor: Espesor de la pieza medida en centímetros
g) Defectos: número de defectos por día. Se inspeccionan 100 piezas día.
h) Problemas producción: Clasificación de los problemas para poder crear un diagrama de causa-efecto.

En el capítulo siguiente a los datos, tenemos ejercicios propuestos como complemento a los ejemplos analizados.

7.2.1. DESCRIPTIVA

7.2.1.1. Tablas de frecuencias

Para crear las tablas de frecuencias de las distintas variables, tenemos que seleccionar la variable y con las opciones tabulares elegir las tablas de frecuencias, como podemos ver en las siguientes ventanas:

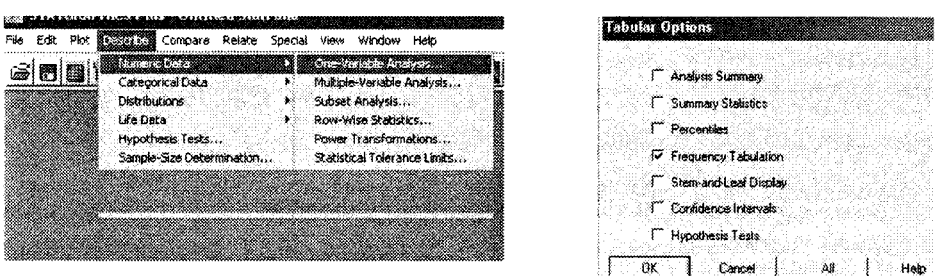


ILUSTRACIÓN 67: creación de tablas de frecuencia

Dentro de las opciones, tenemos la posibilidad de crear las tablas con distinto número de clases, con distintos límites, etc.

Ejercicio 7.1.1

Con los datos de las longitudes de la pieza, lado corto y lado largo,

a) Obtener la tabla de frecuencias en la que se incluya la frecuencia absoluta y relativa de cada una de las medidas.
b) Tabla de frecuencias del lado corto cuando el lado largo es inferior a 20,00 cm.
c) Agrupar en 5 clases, en 4 y en 3 de la variable L1.
d) Agrupar en 5 clases L2, especificando un límite inferior de 19,90 cm y 20,02 cm de superior.

7.2.1.2. Histogramas

Además de crear las tablas de frecuencias, podemos visualizar los datos en forma de histogramas, que recordemos utilizaremos para variables continuas, como en este caso las medidas en centímetros de los lados corto y largo de la pieza cerámica. La forma de obtenerlo es mediante la siguiente opción en Statgraphics:

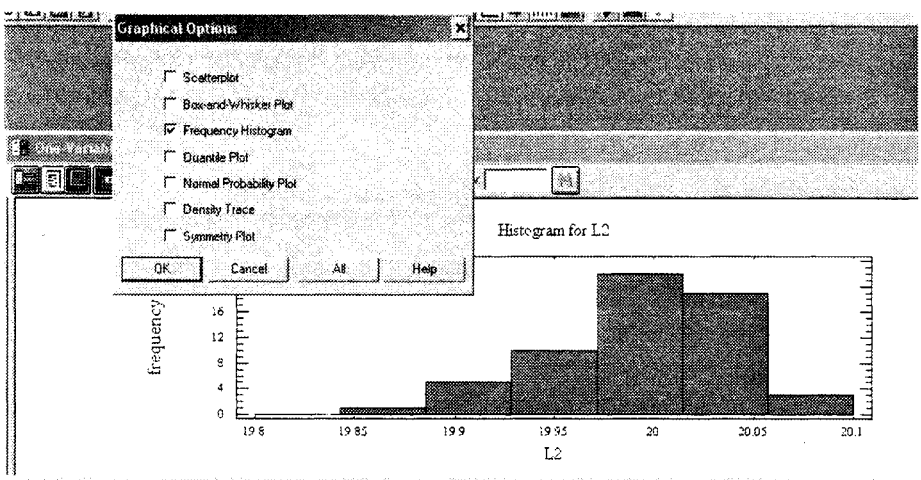


ILUSTRACIÓN 68: Creación de histogramas

Ejercicio 7.1.2

Con los datos de las longitudes de la pieza, lado corto, lado largo, y espesor,

a) Obtener los histogramas de las tres variables por separado. Recordar que es posible obtener histogramas de varias variables a la vez, pero en este caso son demasiado diferentes entre ellas.
b) Analizar los histogramas y ver que posible distribución tienen.
c) Crear los histogramas con distintas clases, 10 y 20 para la variable L2.
d) Crear los histogramas acumulados y relativos en forma de barras y polígono de la variable lado largo. (Recordar que en Statgraphics existen distintas opciones en cada gráfico).

7.2.1.3. Análisis exploratorio de los datos

Ejercicio 7.1.3

Para un simple análisis gráfico de los datos, podemos obtener, para las variables densidad de la pieza en cocido, lado corto y lado largo, (los pasos para obtenerlo, podemos verlo en la ilustración anterior)

a) ¿Cómo están distribuidos los datos de las tres variables, agrupados o con puntos fuera de la distribución? (Crear los gráficos de cajas o *box and whisker plot*)
b) Están distribuidos como distribuciones normales? (Crear el gráfico *Normal Probability Plot* para cada variable)

7.2.1.4. Medidas de posición / dispersión / forma

Ejercicio 7.1.4

Para un análisis de posición, para las variables densidad de la pieza en cocido, lado corto y lado largo, obtener:

a) La media aritmética, la moda, la mediana, los cuartiles, el rango intercuartílico y los percentiles 56, 78 y 98 de las tres variables.
b) Desviación típica, varianza, desviación respecto a la media y el coeficiente de variación de Pearson
c) Coeficientes de asimetría y curtosis, tanto generales como tipificados. Observar la gráfica de simetría.

7.2.1.5. Gráficos de regresión simple

Obtención de los distintos gráficos de regresión simple para la relación entre variables, mediante las ventanas *Relate* > *Simple Regresión*, con las que también es posible obtener la fiabilidad del estudio.

Ejercicio 7.1.5

Obtener:

a) Gráfico de regresión X – Y de las variables lado largo y lado corto.
b) Gráfico de regresión simple entre el lado largo y el espesor de las piezas.
c) Las rectas de regresión y la fiabilidad de ambos gráficos.
d) Distintas posibles relaciones entre las variable y decidir cual es la mejor.

7.2.1.8. Distribuciones multidimensionales

Vamos a relacionar varias variables a la vez, es decir, obtener una a partir de las demás.

Ejercicio 7.1.6

Obtener, mediante Relate > Múltiple Regresión:

a) Regresión múltiple de la longitud del lado largo en función del espesor, densidad aparente y del lado corto.
b) Bondad del ajuste y analizar la recta de regresión múltiple.

7.2.2. CONTROL ESTADÍSTICO DE CALIDAD

7.2.2.1. Indentificar los factores que afectan a la calidad

Los primeros pasos para identificar los factores que afectan a la calidad de una variable, son desarrollados en el inicio del apartado 2 (Esquema de seguimiento de los ejercicios).

Después de saber cuales son las posibles causas podemos pasar la crear los distintos gráficos o diagramas de estudio.

Es importante conocer la idea de la tormenta de ideas (Brainstorming), en la cual, para resolver el problema, primero se aportan todas las ideas de forma ordenada o dando las ideas tal como aparecen.

7.2.2.2. Diagrama Causa – Efecto

También es definido como diagrama de la espina de pescado (en Statgraphics, *Fishbone Diagram*).

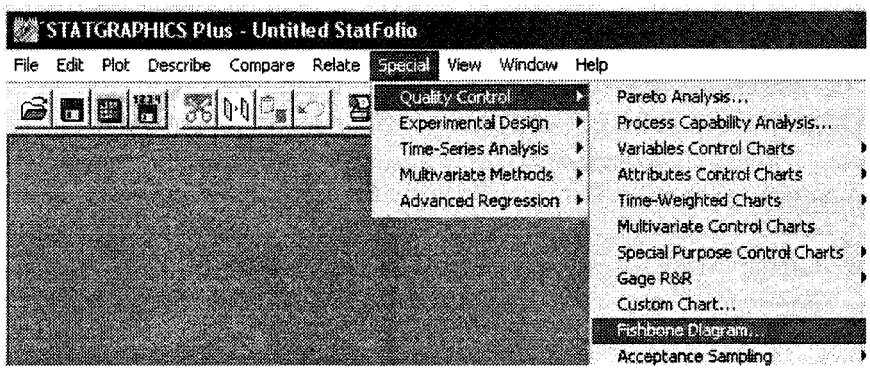


ILUSTRACIÓN 69: Diagrama cusa-efecto

Ejercicio 7.2.1

Obtener:

a) Un posible diagrama de causa – efecto para solucionar el problema de piezas cerámicas defectuosas (piezas rotas).
b) Es posible solucionar el problema sólo con este tipo de diagramas?

7.2.2.3. Pareto

El diagrama de Pareto, también llamado análisis A-B-C, sirve para la clasificación de los elementos o factores que intervienen en un proceso por su orden de importancia para poder tratar cada uno de ellos de una forma distinta según su peso específico.

Los pasos para obtener el gráfico, los podemos ver en la siguiente ilustración:

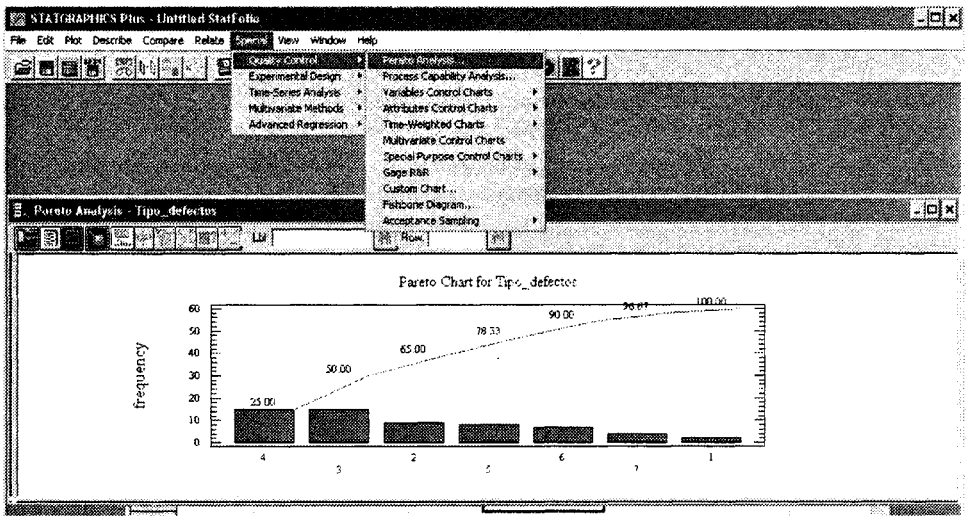


ILUSTRACIÓN 70: Diagrama de Pareto

Ejercicio 7.2.2

Obtener, de la variable Tipos de defectos:

a) Un el diagrama de Pareto para la eliminación de los distintos defectos de las piezas cerámicas, clasificando cada defecto según su peso específico.
b) ¿Qué defecto es el principal causante de los problemas en las piezas?
c) ¿Cómo eliminar aproximadamente el 80 % de los defectos?

7.2.2.4. Gráficos de Control por variables $\overline{x} - R$

La obtención del gráfico de $\overline{x} - R$, se obtendrá a partir de los siguientes pasos:

1) En Statgraphics seguir el siguiente proceso de ventanas:

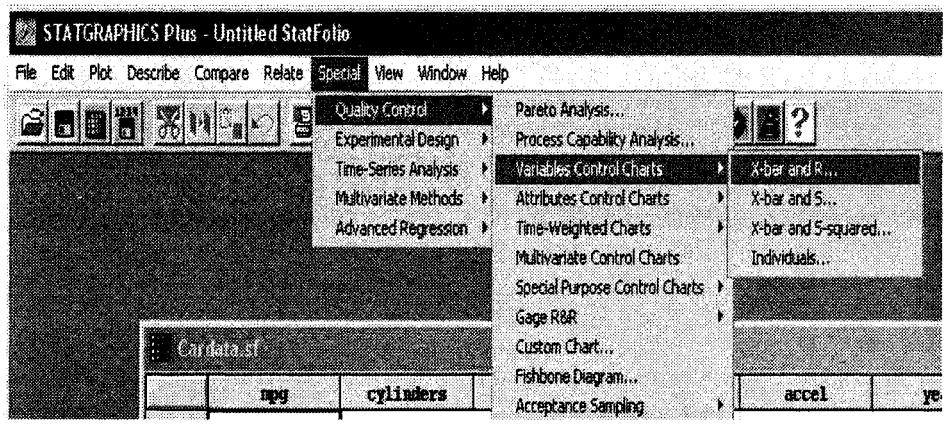


ILUSTRACIÓN 71: Gráfico de medias-recorridos

2) Con la elección de *X-bar and R*, obtendremos la siguiente ventana:

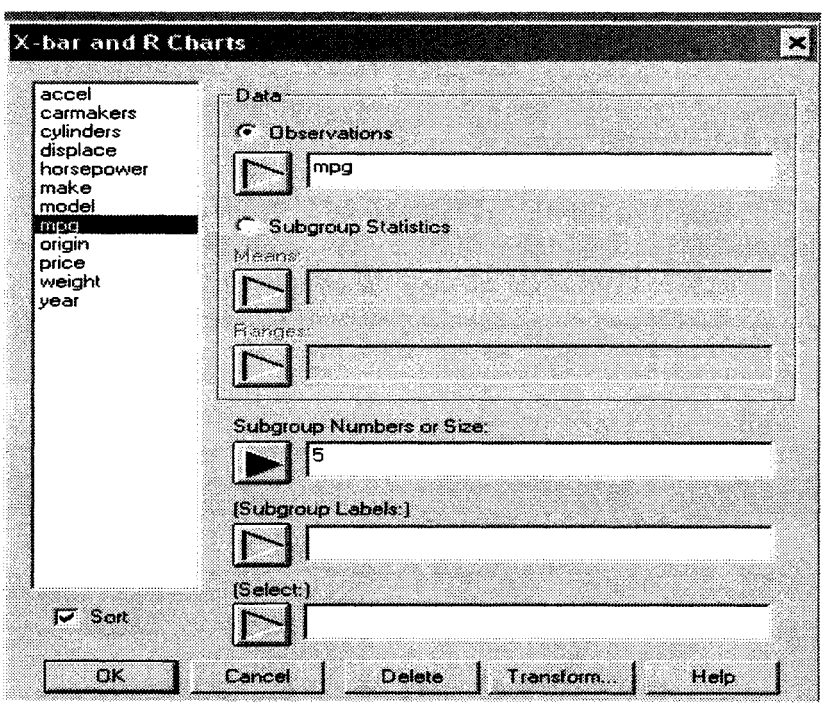


ILUSTRACIÓN 72: Introducción variable en medias- recorridos

Habrá que definir, en data (*observations*) los datos a analizar y en *Subgrup numers of size*, el tamaño de las muestras (por ejemplo las muestras obtenidas cada hora, día, etc.).

En el caso de que tengamos unos datos como estándares, definiremos los datos mediante la introducción del estándar previo y especificar los parámetros:

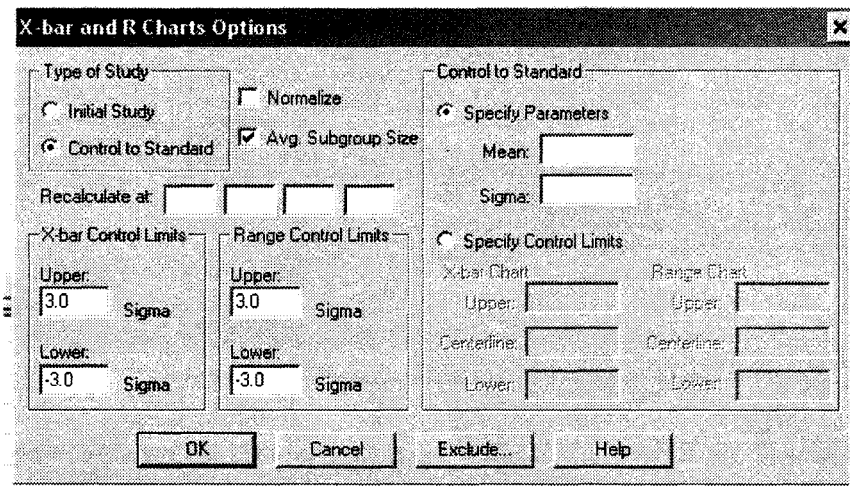


ILUSTRACIÓN 73: Opciones gráfico medias-recorridos

Podemos observar que podemos modificar los límites de control *(x-bar control limits* y *range control limits*), obtener límites de control variables (*avg. Subgroup size*).

También es importante la opción de *Exclude*, para poder eliminar previamente los datos fuera de límites, que hará que estemos con datos fuera de control.

Ejercicio 7.2.3.1

La variable longitud lado largo (L2), puede ser estudiada mediante gráficos de control por variables.

a) Obtener el gráfico de medias - recorridos.
b) Analizar el resultado, eliminando primero los datos fuera de límites, para que esté bajo control y posteriormente ver las tendencias.
c) Si inicialmente tenemos una media de 20,00 cm y desviación típica de 0,05 cm, obtener los gráficos de medias - recorridos.
d) Si existe variación, realizar de nuevo el estudio comentado.
e) Aplicar el estudio de medias-recorridos para la variable de la densidad aparente en crudo.

Ejercicio 7.2.3.1

Otra opción es obtener los índices de capacidad de datos en forma de variable, con el camino de *Special >Quality Control > Variables Control Charts> X -bar and R >* en *tabular options*:

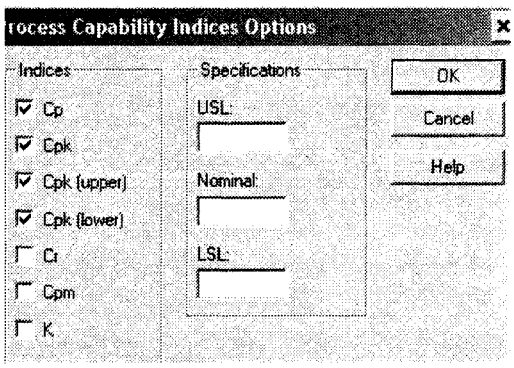


ILUSTRACIÓN 74: Índices de capacidad de procesos

a) Calcular los índices de capacidad Icp e Icpk del lado largo (L2) con las especificaciones 20 ± 0,05 cm.

7.2.2.5. $\bar{x} - s$, $\bar{x} - s^2$

En la siguiente ventana se puede observar la forma de introducir los datos, de igual estilo que con medias-recorridos. Ocurre lo mismo con medias-varianzas.

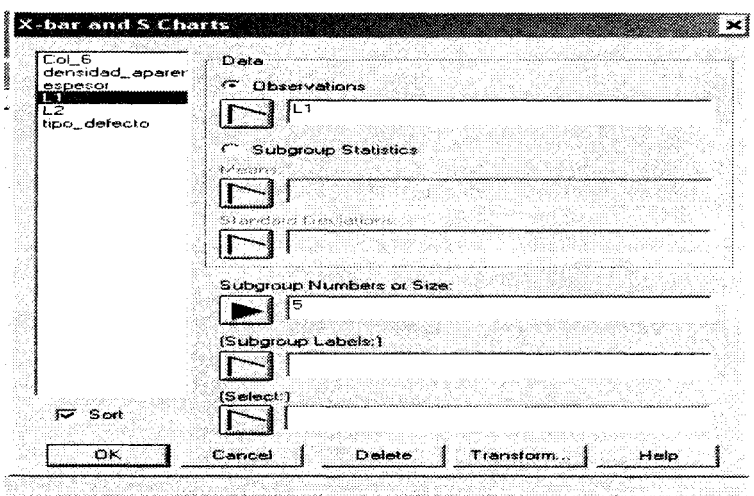


ILUSTRACIÓN 75: Introducción datos medias-desviación típica

Ejercicio 7.2.4.1

De igual forma que tenemos el gráfico de medias-recorridos, con las variables de las longitudes de los lados (L1 y L2) obtener el gráfico de medias-desviaciones típicas de L2. Analizar los resultados.

7.2.2.6. Individuales

Son utilizados para gráficos de control de medias individuales tanto utilizando el gráfico de la media como el gráfico de rangos móviles.

La pantalla siguiente nos muestra el procedimiento para la obtención de gráficos para datos individuales. En el campo de observations se introduce la variable que contiene los datos individuales de la característica de calidad.

En este caso obtendremos el gráfico de las medias (*X chart*) y el de los rangos móviles *(MR (2) chart*). Se puede obtener a partir de la opción de gráficos de la ventana siguiente:

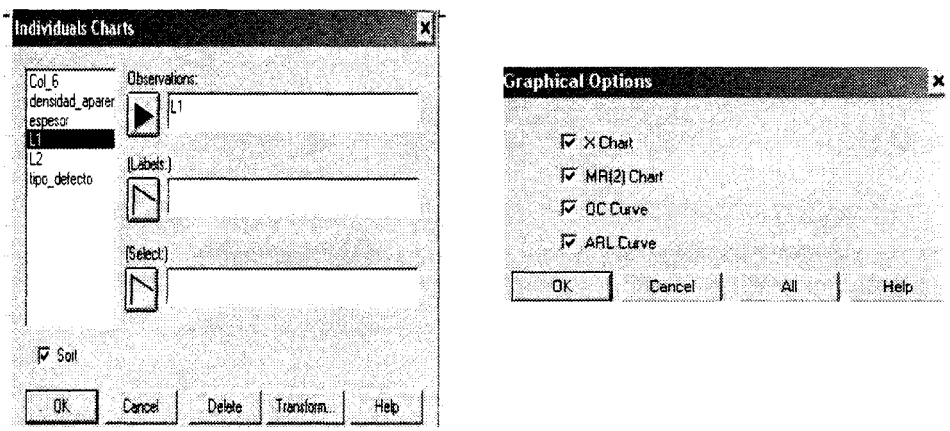


ILUSTRACIÓN 76: Datos en gráficos individuales

Ejercicio 7.2.4.2

La variable densidad aparente en cocido puede ser estudiada mediante gráficos de control por variables individualizada:

a) Obtener el gráfico de datos individualizados.
b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o agrupando los datos para obtener el gráfico de medias-recorridos?

7.2.2.7. CUMSUM

Los gráficos que obtendremos, también llamados V-Máscara, ofrecen la posibilidad de trabajar con gráficos de control de sumas acumulativas a través de la mascarilla V. Sus campos son similares a los descritos para los demás gráficos. La pantalla de entrada es la siguiente:

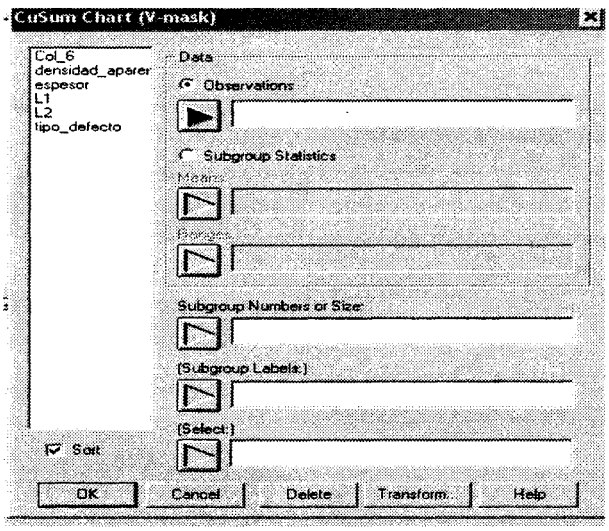


ILUSTRACIÓN 77: Gráfico de CuSum

Las opciones gráficas las podemos ver en la siguiente ventana:

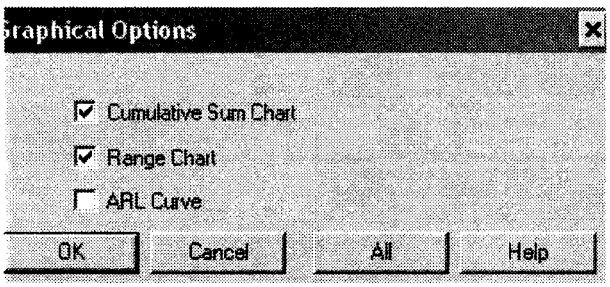


ILUSTRACIÓN 78: Opciones gráficas de CuSum

Ejercicio 7.2.4.3

De la misma forma que hemos estudiado la variable densidad aparente en cocido anteriormente, puede ser estudiada mediante gráficos de CuSum:

a) Obtener los gráficos de la máscara y de los rangos.
b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o agrupando los datos para obtener el gráfico de medias- recorridos?

7.2.2.8. Medias Móviles (MA y EWMA)

Permiten trabajar con gráficos de control de medias móviles y de medias móviles ponderadas exponencialmente. Las opciones son las mismas que las de los gráficos CumSum.

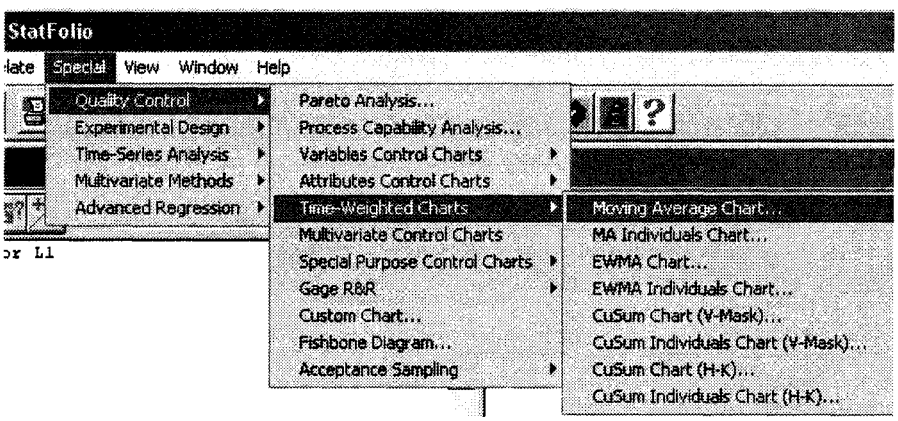


ILUSTRACIÓN 79: Obtención de datos EWMA y MA

Ejercicio 7.2.4.4

Los datos de la variable espesor de las piezas cerámicas, puede ser estudiada con estos gráficos:

a) Obtener los gráficos EM y EWMA del espesor de las piezas.
b) ¿Qué utilidad para la cerámica tienen estos gráficos? ¿Son comparables los resultados con otros estudios anteriores?

7.2.2.9. Gráficos de Medias Móviles y CUSUM Individuales

En la figura del apartado anterior podemos ver los pasos para obtener estos gráficos.

Ejercicio 7.2.4.5

Para comparar con el estudio individualizado anterior, la variable densidad aparente en crudo puede ser estudiada mediante estos gráficos:

a) Obtener el gráfico de medias móviles y CumSum Individuales de la densidad aparente en crudo.
b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o mediante el procedimiento anterior?

7.2.2.10. Gráficos de control multivariante

Se utiliza para gráficos de control multivariante en Statgraphics. La pantalla de entrada es la siguiente:

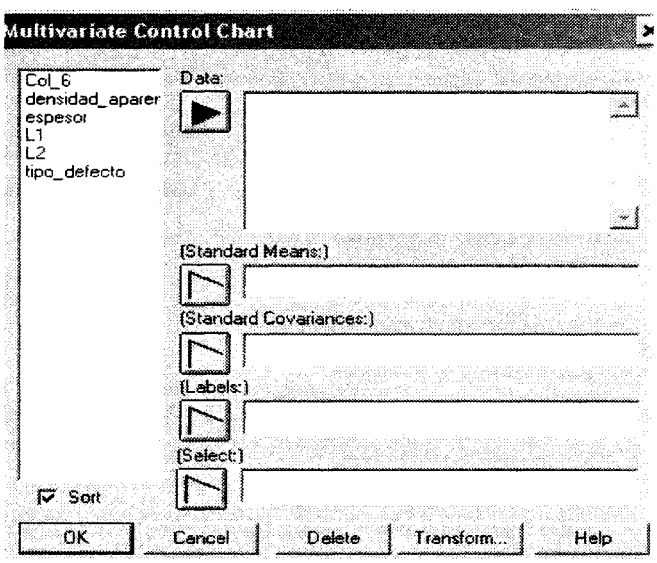


ILUSTRACIÓN 80: Gráfico de control multivariante

En el campo *data* se introducen las variables que contienen los datos de cada característica de calidad. En el campo *Standard Covariances*, se pueden introducir opcionalmente las covarianzas de las variables cuyo gráfico de control simultáneo se realiza.

Ejercicio 7.2.4.6

Estudiamos conjuntamente las longitudes de la pieza L1 y L2,

a) Obtener el gráfico *Multivariate Control Chart* de las dos longitudes.
b) Analizar el resultado y compararlo con los estudios previos de cada variable.

7.2.2.11. Gráficos de control atributos

En este apartado vamos a trabajar con variables en forma de atributos, como son los defectos, número de defectos o porcentajes.

La opción para obtener los gráficos de control por atributos, se permite mediante la siguiente ventana.

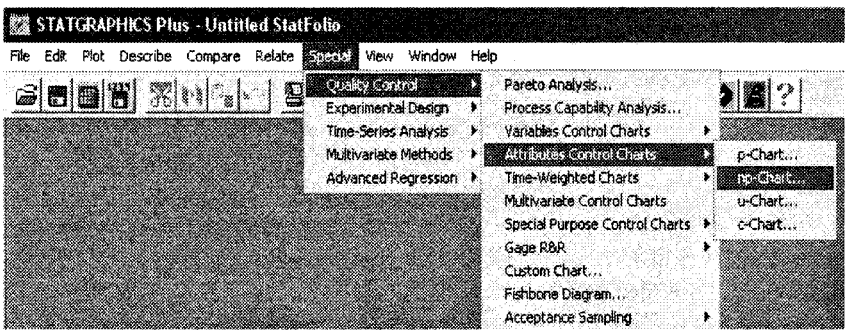


ILUSTRACIÓN 81: Obtención gráfico control por atributos

Se observa que el menú desplegable ofrece la posibilidad de trabajar con gráficos p, np, c y u y dentro de cada uno de ellos tenemos las distintas opciones tabulares y gráficas para las opciones más comunes de estos gráficos.

7.2.2.12. Gráfico p

El procedimiento *p-Chart* nos permite estudiar los datos en formato de números de defectos por unidad. En nuestro caso podemos estudiar los datos de dos formas,

que serían estudiar los defectos que tenemos por pieza o los defectos que tenemos por grupo de muestreo.

La ventana que aparece en este tipo de gráficos es la siguiente:

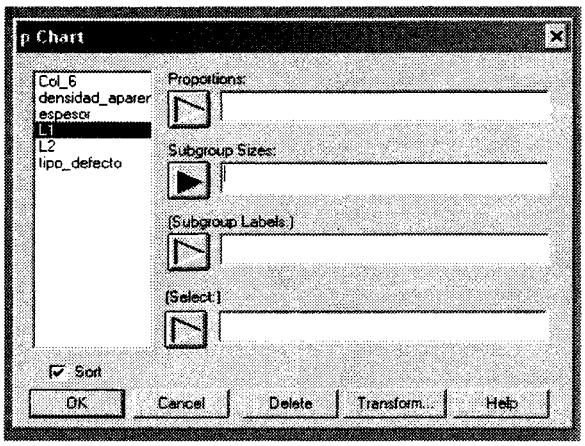


ILUSTRACIÓN 82: Opciones gráfico p

En *Proportions* colocaremos la variable que indique la proporción de unidades defectuosas en cada muestra (número de defectos / número de inspeccionados) y en *Subgroup Sizes* la variable que indica los tamaños de cada muestra.

Ejercicio 7.2.5.1

La variable defectos, nos muestra los defectos que aparecen cada día en el muestreo de las piezas cerámicas, mientras que inspeccionados nos dará las piezas inspeccionadas en cada momento:

a) Obtener el gráfico por atributos «p» para los defectos que se producen cada día (recordar que se muestrean 100 pieza día).
b) Eliminar los datos fuera de límites.

7.2.2.13. Gráficos np

En este caso, las muestras analizadas deben ser constantes a lo largo del tiempo contando número de defectuosos, a diferencia del caso anterior (p-chart) en el que se podía utilizar para diferentes productos o diferentes tamaños muestrales por se proporciones. En el campo *counts*, se introducirán el número de defectos.

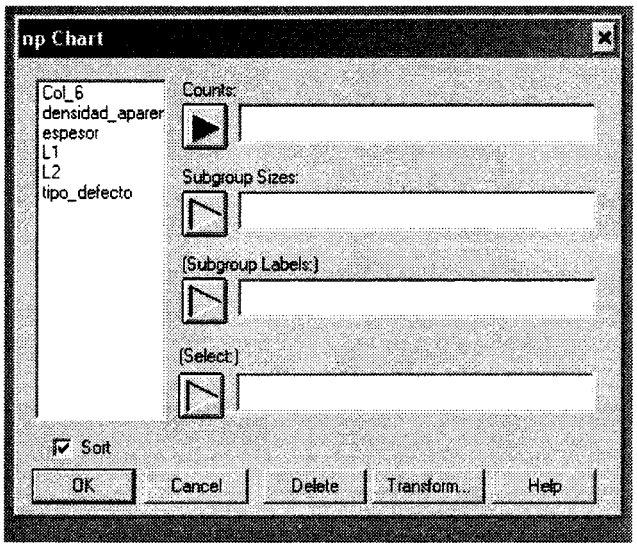


ILUSTRACIÓN 83: Opciones gráfico np

Ejercicio 7.2.5.2

Obtener el gráfico por atributos «np» para los defectos que se producen cada día.

7.2.2.14. Gráficos c

Se utiliza cuando todas las muestras tienen el mismo tamaño o se cuentan los defectos sobre la misma cantidad de material. Se utiliza para analizar el número de defectos.

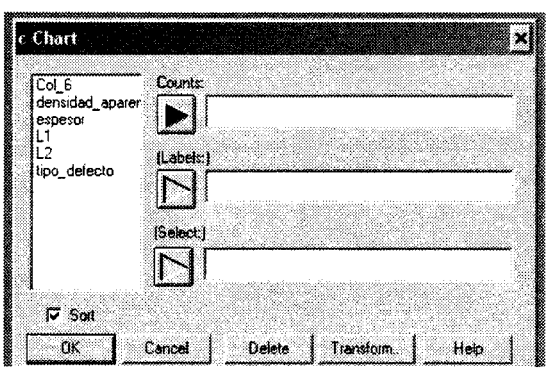


ILUSTRACIÓN 84: Opciones gráficos c

Ejercicio 7.2.5.3

Obtener el gráfico por atributos «c» para los defectos que se producen cada día.

7.2.2.15. Gráficos u

Es de la misma forma que el gráfico c pero prescindiendo de la condición anterior. En Defects per Unit colocaremos la variable «número de defectos por unidad» y en Subgroup Sizes la variable que indica el tamaño de cada muestra.

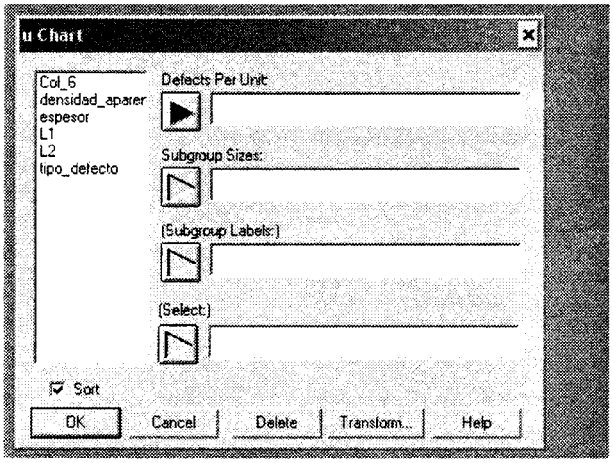


ILUSTRACIÓN 85: Opciones gráficos u

Ejercicio 7.2.5.4

Obtener el gráfico por atributos «u» para los defectos que se producen cada día.

7.2.2.16. Capacidad de los procesos

El procedimiento de Análisis de la Capacidad de un Proceso como una opción de control de calidad, lo podemos observar en las ventanas siguientes:

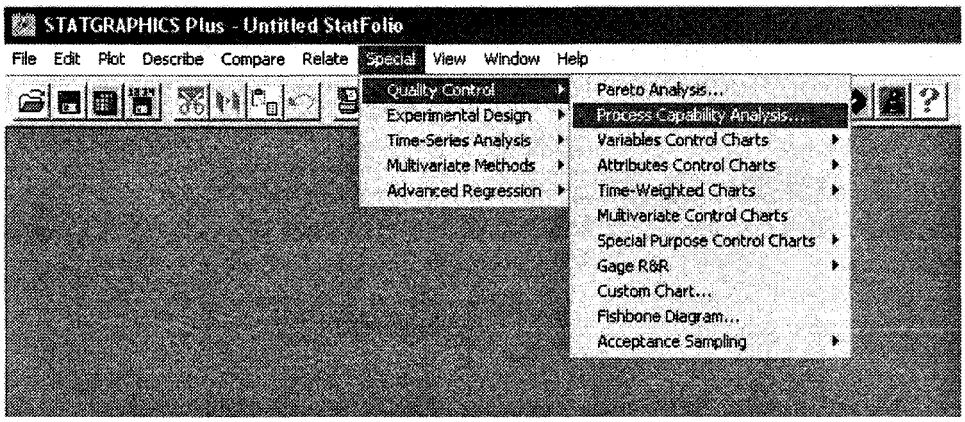


ILUSTRACIÓN 86: Pasos obtención análisis de capacidad

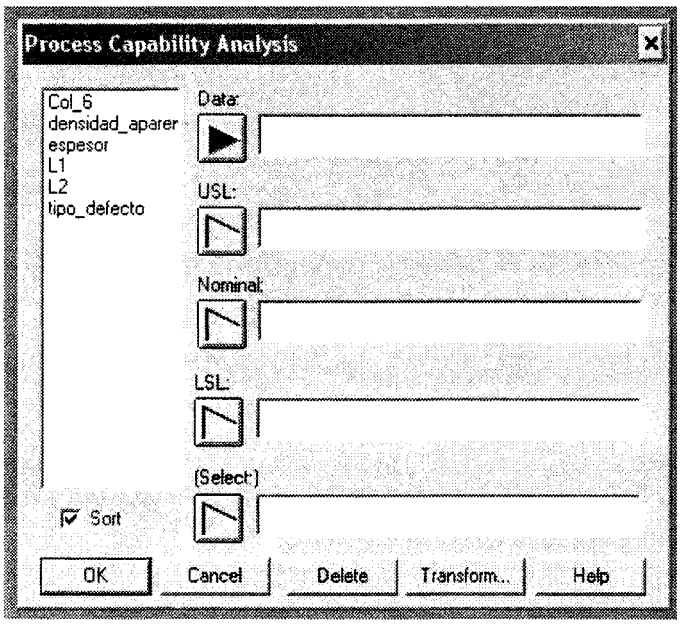


ILUSTRACIÓN 87: Datos del análisis de capacidad

En la ventana de Análisis de Capacidad del Proceso, introduciremos en Data, los datos a analizar y en USL, Nominal y LSL, los valores nominal (el central) y los que tomaremos como límite superior y límite inferior.

Después de elegir estos datos, con la opción de análisis (botón derecho del ratón) para elegir el tipo de distribución y especificar los parámetros:

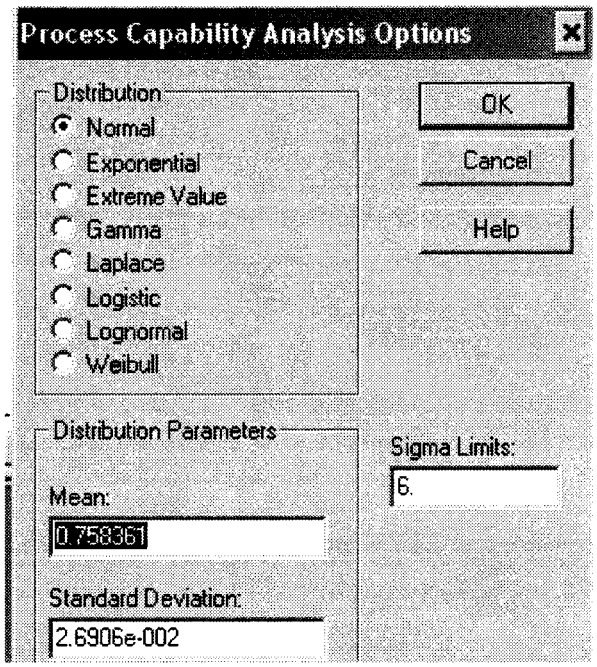


ILUSTRACIÓN 88: Opciones de análisis

Ejercicio 7.2.6.1

Para estudiar las variables, o mejor dicho, el proceso o máquinas utilizadas en el control:

a) Realizar el estudio de capacidad para la variable densidad aparente en crudo (valor nominal = 1.96 g/cc y los límites superior e inferior, ± 0.017).
b) Estudio de capacidad de las dos longitudes (L1 y L2)

7.2.2.17. Índices de capacidad

Dentro del estudio de capacidad de un proceso o máquina, se deben incluir los datos de los distintos índices de capacidad.

Para obtener estos índices, observamos las siguientes ventanas:

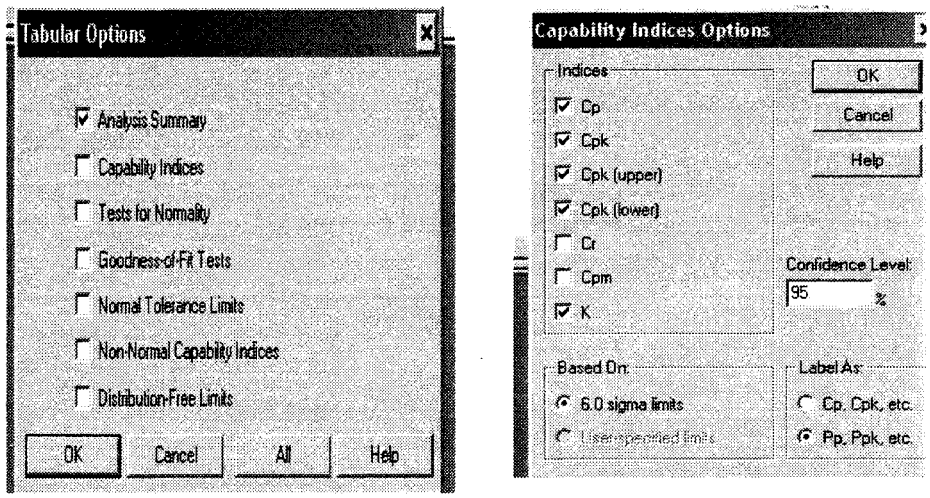


ILUSTRACIÓN 89: Opciones de los índices de capacidad e índices

Ejercicio 7.2.6.2

Obtener los Índices de capacidad del apartado anterior.

7.3. FICHEROS DE DATOS PARA LOS EJERCICIOS

Fichero de datos utilizado para los distintos ejercicios (las variables han sido comentadas en apartados anteriores).

L1	L2	Dap crudo	Espesor	Tipo defecto	Problemas producción	Inspeccionados	Defectuosas
15.00	20.01	1.96	0.80	0	piezas rotas	100	2
15.02	20.03	1.98	0.80	0	equipos	100	0
15.01	20.03	1.98	0.81	0	.prensa	100	4
15.00	20.01	1.97	0.81	0	.secadero	100	3
15.03	20.02	1.96	0.82	0	.línea	100	0
15.03	20.05	1.95	0.83	0	operarios	100	0
15.04	20.07	1.94	0.83	1	.turnos	100	2
15.02	20.04	1.94	0.82	1	.nocturnidad	100	4
15.02	20.03	1.93	0.82	0	.número	100	5
15.03	20.03	1.94	0.83	3	.horarios	100	0
15.02	20.02	1.94	0.83	0	materias primas	100	0

L1	L2	Dap crudo	Espesor	Tipo defecto	Problemas producción	Inspec-cionados	Defectuosas
15.01	20.02	1.94	0.82	0	.pasta c	100	0
15.00	20.00	1.96	0.80	0	.esmaltes	100	5
15.00	20.02	1.96	0.81	0	.grasas	100	4
14.99	20.00	1.96	0.80	2	.serigrafías	100	3
14.99	19.98	1.97	0.79	5	entorno	100	6
15.01	20.03	1.96	0.80	0	.empresa	100	9
15.00	20.03	1.95	0.80	0	.proveedores	100	2
14.98	20.00	1.97	0.79	0	.zona	100	3
14.98	19.96	1.98	0.79	0	.nocturnidad	100	2
14.97	19.94	1.98	0.78	1	.núm de operarios	100	6
14.97	19.95	1.99	0.78	0	.horarios	100	5
14.96	19.94	1.98	0.78	1	materias primas	100	4
14.97	19.96	1.98	0.77	5	.pasta cerámica	100	5
14.97	19.97	1.97	0.77	0	.esmaltes	100	0
15.00	19.99	1.96	0.79	0	.grasas de prensa	100	0
15.02	20.00	1.96	0.79	0	.serigrafías	100	0
15.02	20.03	1.95	0.80	0		100	5
15.04	20.07	1.94	0.82	1		100	4
15.05	20.09	1.92	0.83	2		100	3
15.04	20.09	1.92	0.83	1		100	6
15.04	20.07	1.93	0.83	0		100	4
15.03	20.04	1.93	0.83	0		100	5
15.03	20.04	1.94	0.82	0		100	0
15.00	20.01	1.96	0.82	0		100	0
15.01	20.01	1.97	0.80	0		100	0
14.99	20.00	1.97	0.79	0		100	5
14.98	19.98	1.98	0.79	0		100	4
14.99	19.99	1.98	0.79	0		100	3
15.01	20.00	1.98	0.80	0		100	6
15.00	20.00	1.97	0.80	0		100	2
15.01	20.00	1.96	0.79	0		100	3
15.02	20.03	1.96	0.82	2		100	7
15.00	20.02	1.96	0.82	0		100	5
15.00	20.02	1.96	0.82	0		100	9
15.01	20.03	1.97	0.83	0		100	1
15.00	20.03	1.96	0.83	0		100	2

L1	L2	Dap crudo	Espesor	Tipo defecto	Problemas producción	Inspec-cionados	Defectuosas
14.99	20.01	1.96	0.82	0		100	0
14.97	20.00	1.98	0.80	0		100	0
14.98	19.97	1.98	0.78	4		100	0
14.96	19.95	1.99	0.78	5		100	0
14.96	19.94	1.99	0.77	1		100	0
14.96	19.94	1.99	0.77	5		100	0
14.98	19.97	1.98	0.79	0		100	2
14.99	19.98	1.98	0.78	0		100	3
15.00	20.02	1.97	0.79	0		100	2
15.01	20.02	1.96	0.80	0		100	8
15.01	20.03	1.96	0.82	0		100	6
15.00	20.00	1.96	0.80	0		100	5
15.02	20.01	1.96	0.80	0		100	5
15.02	20.00	1.95	0.81	0		100	9
15.01	20.00	1.95	0.81	0		100	1
15.00	19.98	1.97	0.79	0		100	2
15.00	19.99	1.97	0.80	0		100	0
14.99	19.96	1.98	0.79	4		100	0
14.97	19.96	1.98	0.77	5		100	0
15.00	19.95	1.97	0.77	0		100	0
15.00	19.98	1.97	0.77	0		100	0
15.02	19.99	1.97	0.77	0		100	0
15.01	19.99	1.96	0.77	0		100	2
15.01	20.00	1.96	0.79	0		100	3
15.03	20.04	1.96	0.83	0		100	2
15.04	20.07	1.93	0.83	1		100	1
15.03	20.09	1.93	0.84	0		100	2
15.03	20.08	1.93	0.84	0		100	0
15.02	20.08	1.93	0.83	0		100	0
15.02	20.06	1.93	0.83	3		100	0
15.01	20.05	1.93	0.82	0		100	0
15.01	20.04	1.94	0.83	0		100	0
15.03	20.03	1.96	0.83	2		100	0
15,01	20,02	1,95	0,82	0		100	1
15,00	20,02	1,96	0,83	0		100	2
14,99	20.00	1,96	0,82	0		100	0

L1	L2	Dap crudo	Espesor	Tipo defecto	Problemas producción	Inspec-cionados	Defectuosas
14,99	20,00	1,96	0,81	0		100	0
14,98	20,01	1,95	0,82	0		100	0
14,99	20,02	1,95	0,80	0		100	0
15,00	20,01	1,96	0,81	0		100	0
15,00	19,97	1,97	0,79	5		100	0
15,01	20,00	1,96	0,80	0		100	1
15,02	20,03	1,95	0,82	0		100	0
15,02	20,05	1,95	0,82	3		100	0
15,02	20,04	1,94	0,81	0		100	0
15,03	20,05	1,94	0,80	0		100	7
15,02	20,04	1,96	0,80	0		100	2
15,01	20,00	1,96	0,80	2		100	4
15,02	20,00	1,96	0,79	0		100	3
15,01	19,98	1,96	0,80	0		100	3
15,00	20,00	1,97	0,80	0		100	6
15,00	20,01	1,96	0,79	0		100	3
15,01	20,02	1,96	0,80	0		100	2

En la columna de Problemas de producción, sólo aparecen los datos necesarios para un simple problema. Si es necesario, puede completares con un mayor numero de efectos.

8 SOLUCIÓN DE LOS EJERCICIOS

Solución Ejercicio 7.1.1

a) Tablas de frecuencias para las variables longitud de la pieza, lado largo y lado corto:

Frequency Tabulation for L2

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		19.8		0	0.0000	0	0.0000
1	19.8	19.8375	19.8188	0	0.0000	0	0.0000
2	19.8375	19.875	19.8563	0	0.0000	0	0.0000
3	19.875	19.9125	19.8938	7	0.0700	7	0.0700
4	19.9125	19.95	19.9313	10	0.1000	17	0.1700
5	19.95	19.9875	19.9688	24	0.2400	41	0.4100
6	19.9875	20.025	20.0063	38	0.3800	79	0.7900
7	20.025	20.0625	20.0438	19	0.1900	98	0.9800
8	20.0625	20.1	20.0813	2	0.0200	100	1.0000
above	20.1			0	0.0000	100	1.0000

Mean = 19.9891 Standard deviation = 0.0440407

Frequency Tabulation for L1

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		14.8		0	0.0000	0	0.0000
1	14.8	14.85	14.825	0	0.0000	0	0.0000
2	14.85	14.9	14.875	6	0.0600	6	0.0600
3	14.9	14.95	14.925	4	0.0400	10	0.1000
4	14.95	15.0	14.975	42	0.4200	52	0.5200
5	15.0	15.05	15.025	40	0.4000	92	0.9200
6	15.05	15.1	15.075	8	0.0800	100	1.0000
7	15.1	15.15	15.125	0	0.0000	100	1.0000
8	15.15	15.2	15.175	0	0.0000	100	1.0000
above	15.2			0	0.0000	100	1.0000

Mean = 15.0018 Standard deviation = 0.0427201

b) Tablas de frecuencias del lado corto cuando el lado largo es inferior a 20,00 cm:

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		14.8		0	0.0000	0	0.0000
1	14.8	14.8429	14.8214	0	0.0000	0	0.0000
2	14.8429	14.8857	14.8643	3	0.0566	3	0.0566
3	14.8857	14.9286	14.9071	3	0.0566	6	0.1132
4	14.9286	14.9714	14.95	6	0.1132	12	0.2264
5	14.9714	15.0143	14.9929	18	0.3396	30	0.5660
6	15.0143	15.0571	15.0357	18	0.3396	48	0.9057
7	15.0571	15.1	15.0786	5	0.0943	53	1.0000
above	15.1			0	0.0000	53	1.0000

Mean = 14.9979 Standard deviation = 0.0512294

c) Agrupadas L1 y L2 en 5, 4 y 3 clases:

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		14.8		0	0.0000	0	0.0000
1	14.8	14.88	14.84	3	0.0300	3	0.0300
2	14.88	14.96	14.92	13	0.1300	16	0.1600
3	14.96	15.04	15.0	73	0.7300	89	0.8900
4	15.04	15.12	15.08	11	0.1100	100	1.0000
5	15.12	15.2	15.16	0	0.0000	100	1.0000
above	15.2			0	0.0000	100	1.0000

Mean = 15.0018 Standard deviation = 0.0427201

Frequency Tabulation for L1

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		14.8		0	0.0000	0	0.0000
1	14.8	14.9	14.85	6	0.0600	6	0.0600
2	14.9	15.0	14.95	46	0.4600	52	0.5200
3	15.0	15.1	15.05	48	0.4800	100	1.0000
4	15.1	15.2	15.15	0	0.0000	100	1.0000
above	15.2			0	0.0000	100	1.0000

Mean = 15.0018 Standard deviation = 0.0427201

Frequency Tabulation for L1

Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		14.8		0	0.0000	0	0.0000
1	14.8	14.9333	14.8667	6	0.0600	6	0.0600
2	14.9333	15.0667	15.0	87	0.8700	93	0.9300
3	15.0667	15.2	15.1333	7	0.0700	100	1.0000
above	15.2			0	0.0000	100	1.0000

Mean = 15.0018 Standard deviation = 0.0427201

d)

Frequency Tabulation for L2

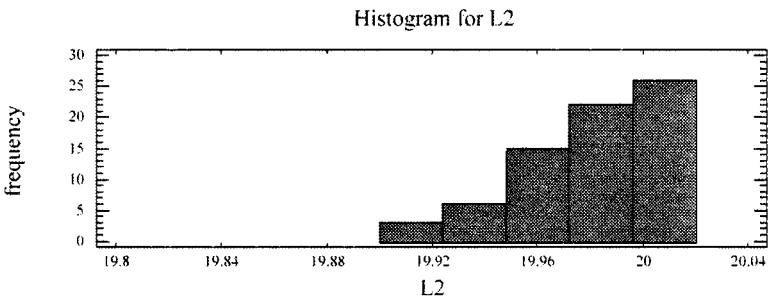
Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		19.9		7	0.0700	7	0.0700
1	19.9	19.924	19.912	3	0.0300	10	0.1000
2	19.924	19.948	19.936	6	0.0600	16	0.1600
3	19.948	19.972	19.96	15	0.1500	31	0.3100
4	19.972	19.996	19.984	22	0.2200	53	0.5300
5	19.996	20.02	20.008	26	0.2600	79	0.7900
above	20.02			21	0.2100	100	1.0000

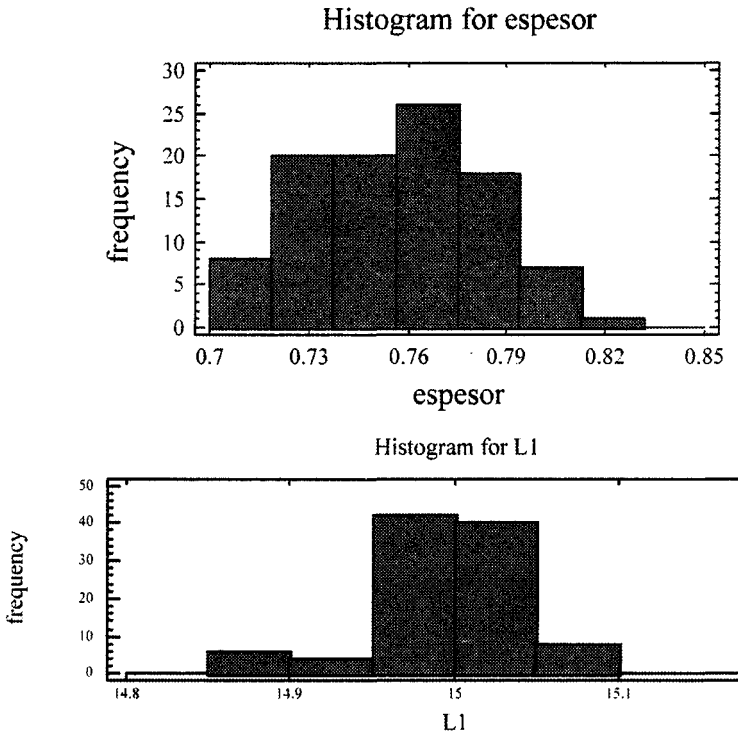
Mean = 19.9891 Standard deviation = 0.0440407

Solución Ejercicio 7.1.2

Con los datos de las longitudes de la pieza, lado corto, lado largo, y espesor:

a) Obtener los histogramas de las tres variables por separado. Recordar que es posible obtener histogramas de varias variables a la vez, pero en este caso son demasiado diferentes entre ellas.

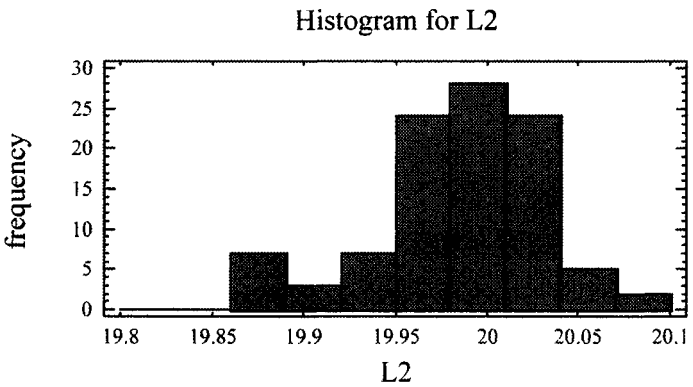




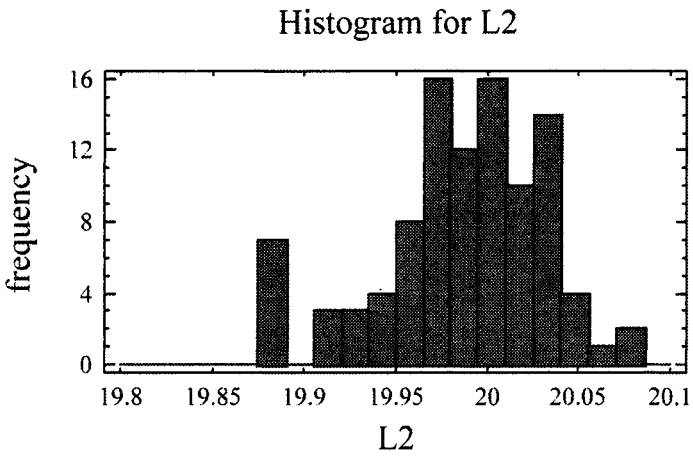
b) Analizar los histogramas y ver que posible distribución tienen:

En todos los casos, sin otra clase de estudio, podemos suponer que son distribuciones normales. Sería necesario otras representaciones (Normal Probability Plot) o distintos tests de normalidad.

c) Crear los histogramas con distintas clases, 10 y 20 para la variable L2:

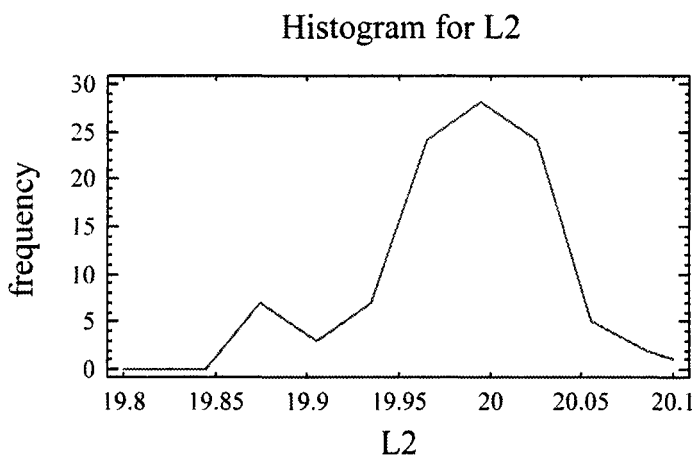


Con 20 clases:

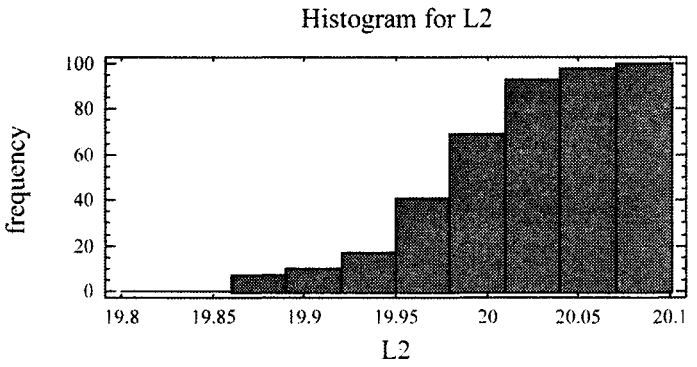


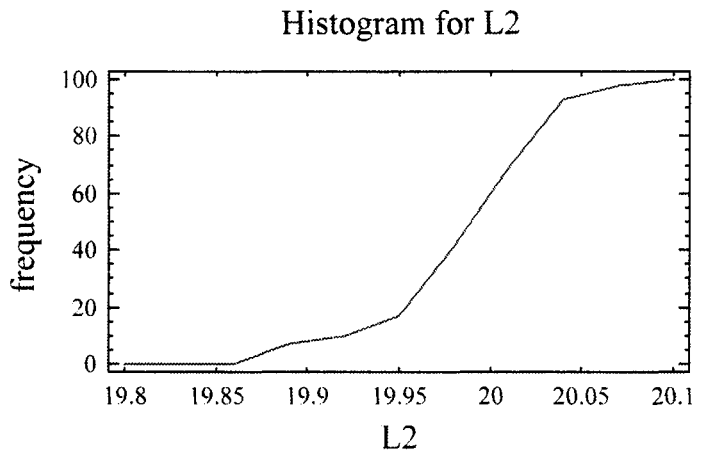
d) Crear los histogramas acumulados y relativos en forma de barras y polígono de la variable lado largo (L2).

Los anteriores son relativos en forma de barras, el siguiente es en forma de polígonos:



Los absolutos de barras y polígonos:

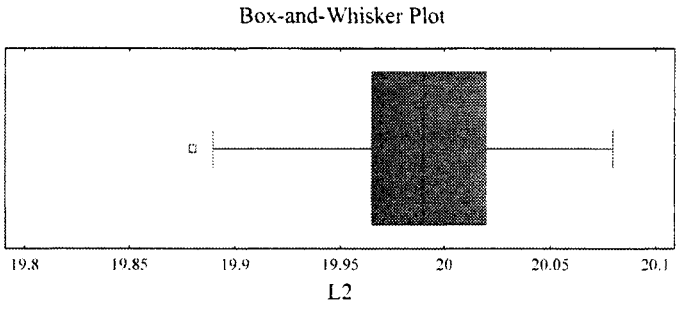


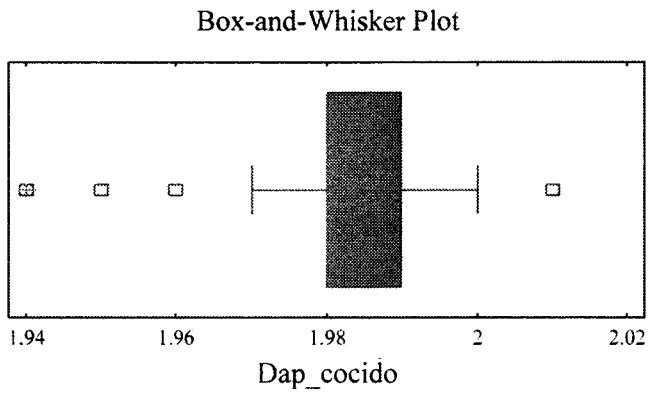


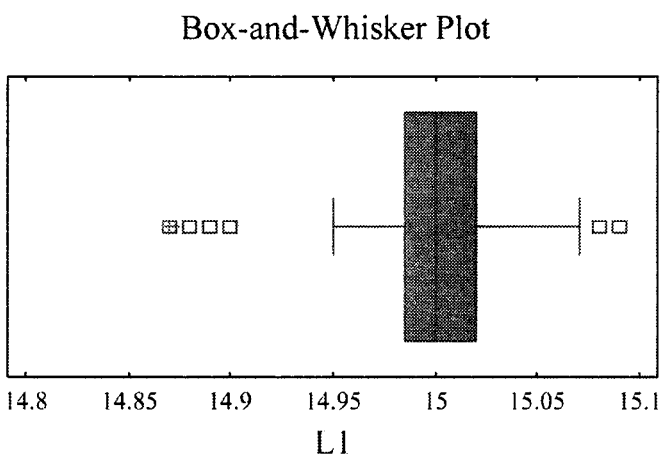
Solución Ejercicio 7.1.3

Variables: densidad en crudo, L1 y L2.

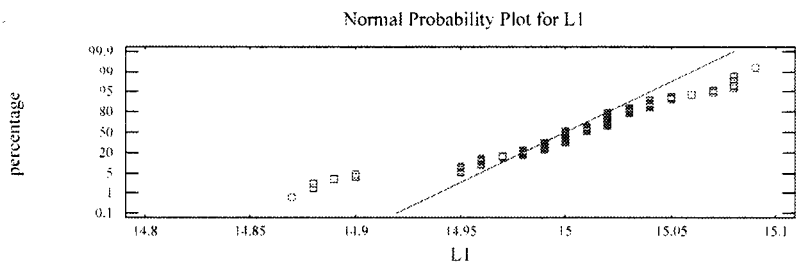
a) Como están distribuidos los datos de las tres variables, agrupados o con puntos fuera de la distribución? (Crear los gráficos de cajas o *box and whisker plot*)

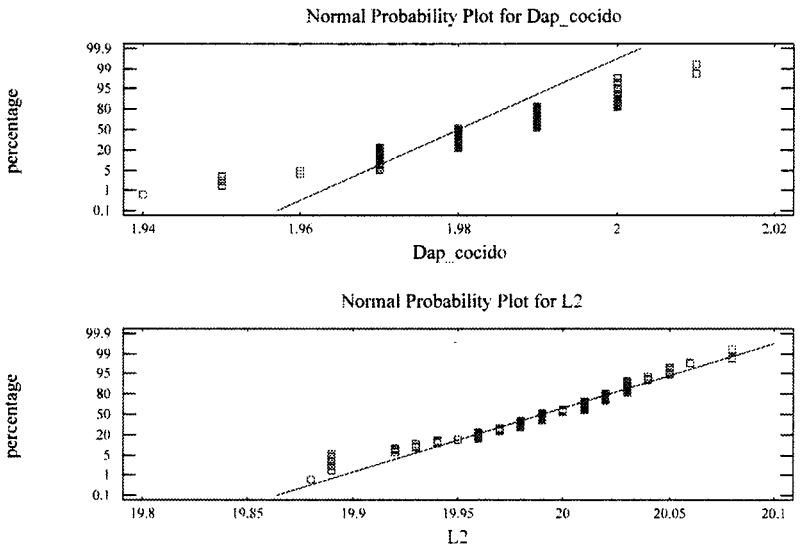






e) Están distribuidos como distribuciones normales? (Crear el gráfico Normal Probability Plot para cada variable)





Solución Ejercicio 7.1.4

Para un análisis de posición, para las variables densidad aparente de la pieza en crudo, lado corto y lado largo, obtener:

a) La media aritmética, la moda, la mediana, los cuartiles, el rango intercuartílico y los percentiles 56, 78 y 98 de las tres variables.

Como son tres variables, podemos obtener los resultados todos juntos en una tabla para los tres apartados:

Summary Statistics

	Count	Average	Median
L1	100	15.0018	15.0
Dap_cocido	99	1.98333	1.98
L2	100	19.9891	19.99
Total	299	12.3593	15.0

	Mode	Geometric mean	Variance
L1	15.02	15.0017	0.00182501
Dap_cocido		1.98329	0.00017551
L2		19.9891	0.00193959
Total		8.45029	57.6458

	Standard deviation	Standard error	Minimum

	Standard deviation	Standard error	Minimum
L1	0.0427201	0.00427201	14.87
Dap_cocido	0.013248	0.00133148	1.94
L2	0.0440407	0.00440407	19.86
Total	7.59249	0.439085	1.94

	Maximum	Range	Lower quartile
L1	15.09	0.22	14.985
Dap_cocido	2.01	0.07	1.98
L2	20.08	0.2	19.965
Total	20.08	18.14	1.99

	Upper quartile	Interquartile range	Skewness
L1	15.02	0.035	-0.823919
Dap_cocido	1.99	0.01	-0.612134
L2	20.02	0.055	-0.613992
Total	19.97	17.98	-0.496519

	Stnd. skewness	Kurtosis	Stnd. kurtosis
L1	-3.36364	1.62522	3.31747
Dap_cocido	-2.4865	0.700052	1.42181
L2	-2.50661	0.133291	0.27208
Total	-3.50506	-1.49381	-5.27259

	Coeff. of variation	Sum
L1	0.284767%	1500.18
Dap_cocido	0.667968%	196.35
L2	0.220324%	1998.91
Total	61.4312%	3695.44

b) Desviación típica, varianza, desviación respecto a la media y el coeficiente de variación de Pearson

c) Coeficientes de asimetría y curtosis, tanto generales como tipificados. Observar la gráfica de simetría.

Los valores de los percentiles 56, 78 y 98:

```
Percentiles for Dap_cocido

56.0% = 1.99
78.0% = 1.99
98.0% = 2.01
```

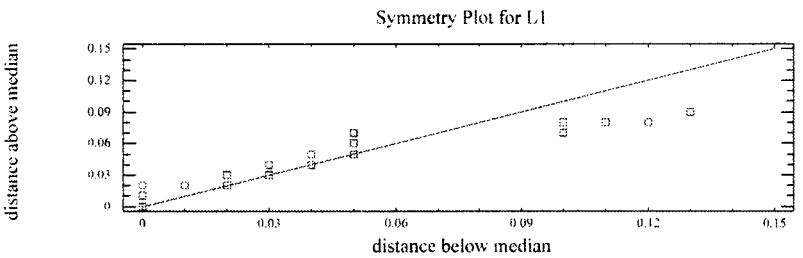
```
Percentiles for L2

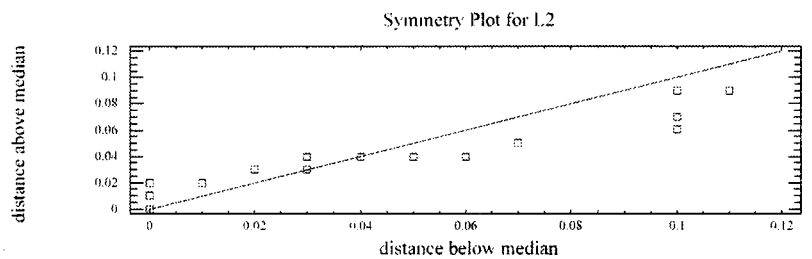
56.0% = 20.0
78.0% = 20.02
98.0% = 20.07

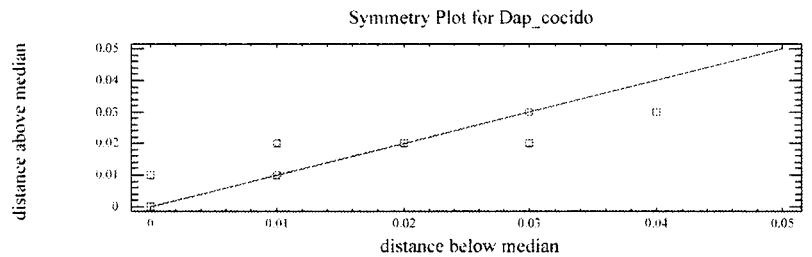

Percentiles for L1

56.0% = 15.01
78.0% = 15.03
98.0% = 15.08
```

Los gráficos de simetría para las tres variables son:



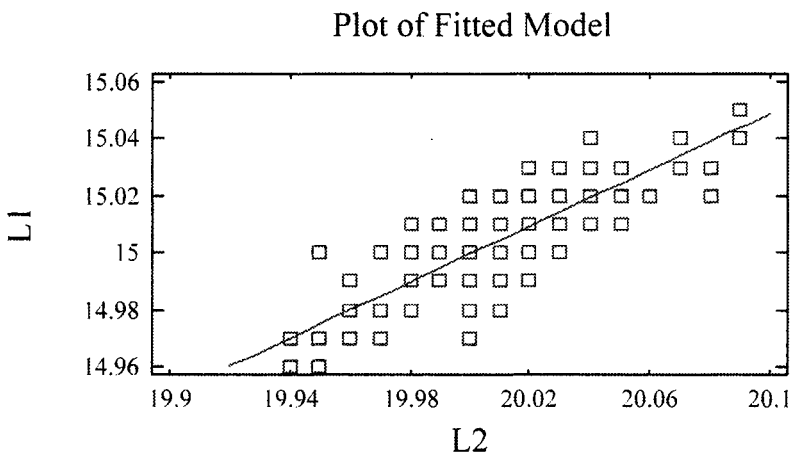




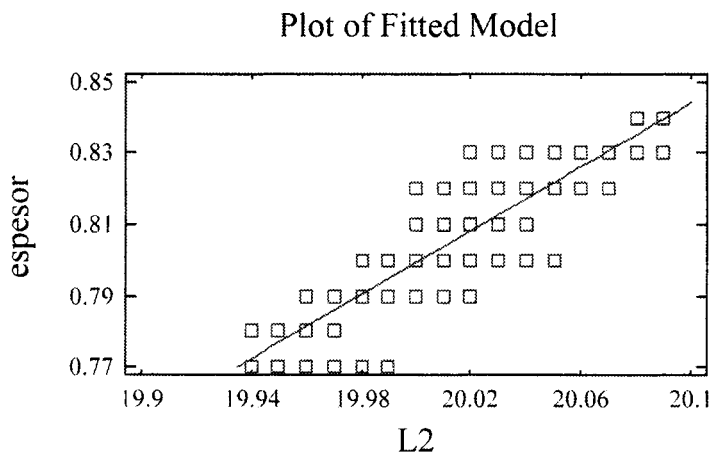
Solución Ejercicio 7.1.5

Obtener,

a) Gráfico de regresión X – Y de las variables lado largo (como X) y lado corto (como Y).



b) Gráfico de regresión simple entre el lado largo y el espesor de las piezas.



c) Las rectas de regresión y la fiabilidad de ambos gráficos.

Índice

Para el apartado *a*):

```
Correlation Coefficient = 0.84197
R-squared = 70.8914 percent
Standard Error of Est. = 0.0113112

L1 = 5.14031 + 0.492961*L2
```

Para el apartado *b*):

```
Correlation Coefficient = 0.833085
R-squared = 69.4031 percent
Standard Error of Est. = 0.0106738

espesor = -8.17948 + 0.448937*L2
```

d) Distintas posibles relaciones entre las variable y decidir cual es la mejor.

Para la relación entre lado largo y lado corto:

Comparison of Alternative Models

Model	Correlation	R-Squared
Reciprocal-X	-0.8421	70.91%
S-curve	-0.8420	70.90%
Logarithmic-X	0.8420	70.90%
Double reciprocal	0.8420	70.90%
Square root-X	0.8420	70.90%
Multiplicative	0.8420	70.89%
Linear	0.8420	70.89%
Square root-Y	0.8420	70.89%
Exponential	0.8419	70.89%
Reciprocal-Y	-0.8419	70.88%
Logistic	<no fit>	
Log probit	<no fit>	

Para la relación entre lado largo y espesor:

Comparison of Alternative Models

Model	Correlation	R-Squared
Reciprocal-X	-0.8332	69.43%
Logarithmic-X	0.8332	69.41%
Log probit	0.8331	69.41%
Square root-X	0.8331	69.41%
Linear	0.8331	69.40%
S-curve	-0.8331	69.40%
Square root-Y	0.8330	69.39%
Multiplicative	0.8330	69.39%
Logistic	0.8330	69.39%
Exponential	0.8329	69.38%
Double reciprocal	0.8328	69.36%
Reciprocal-Y	-0.8326	69.33%

En ambos casos, no mejora el ajuste con el cambio de modelo.

Solución Ejercicio 7.1.6

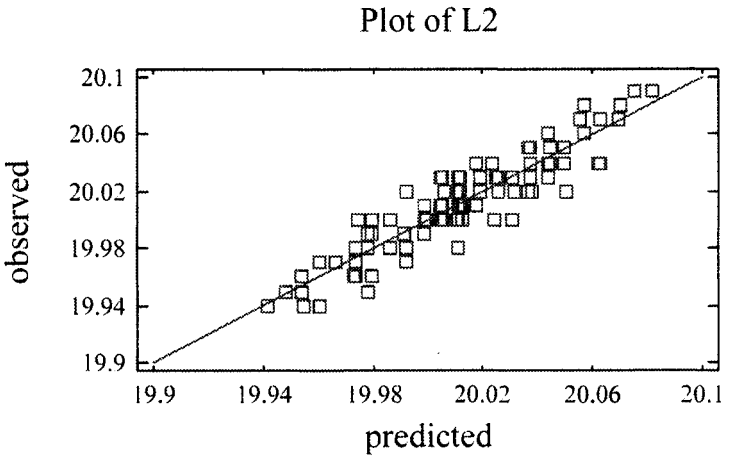
Obtener, mediante *Relate* > *Múltiple Regresión*,

a) Regresión múltiple de la longitud del lado largo en función del espesor, densidad aparente y del lado corto.

La recta de regresión obtenida es la siguiente:

```
L2 = 11.7269 + 0.673457*espesor + 0.600796*L1 - 0.649194*Dap_crudo
```

Uno de los posibles gráficos que podemos obtener es el de los valores predichos frente a los observados:



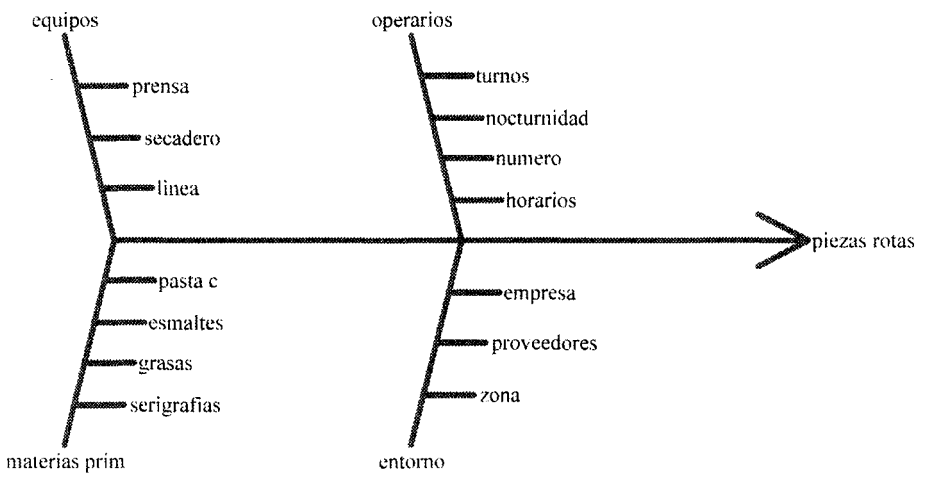
b) Bondad del ajuste y analizar la recta de regresión múltiple.

```
R-squared = 85.7098 percent
R-squared (adjusted for d.f.) = 85.2633 percent
Standard Error of Est. = 0.0136767
Mean absolute error = 0.0106556
Durbin-Watson statistic = 1.23269
```

Solución Ejercicio 7.2.1

Obtener,

a) Un posible diagrama de causa-efecto para solucionar el problema de piezas cerámicas defectuosas.



Índice

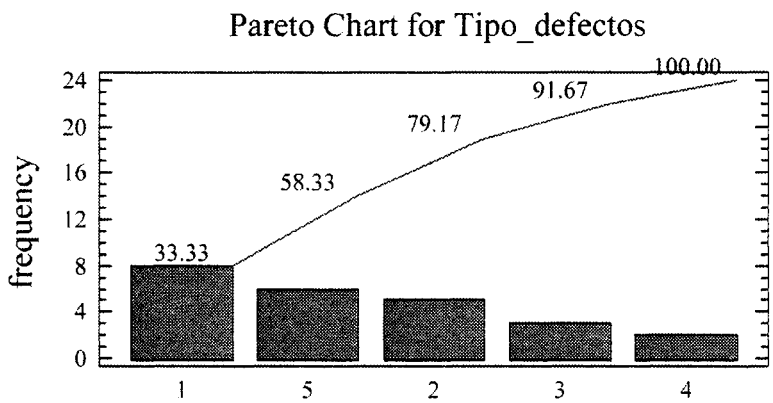
b) Es posible solucionar el problema solo con este tipo de diagramas?
Es necesario otros gráficos y estudios como analizar cada una de las variables.

Solución Ejercicio 7.2.2

Obtener, de la variable Tipos de defectos:

a) Un el diagrama de Pareto para la eliminación de los distintos defectos de las piezas cerámicas, clasificando cada defecto según su peso específico.

Con el fichero de datos que tenemos, para obtener el diagrama de pareto, primero tenemos que crear una columna con los tipos de defectos pero eliminando los ceros, que son las piezas que no tienen defectos.



b) ¿Qué defecto es el principal causante de los problemas en las piezas?

El 1 (Descuadre), seguido del 5 (Planaridad) y del 2 (Calibre).

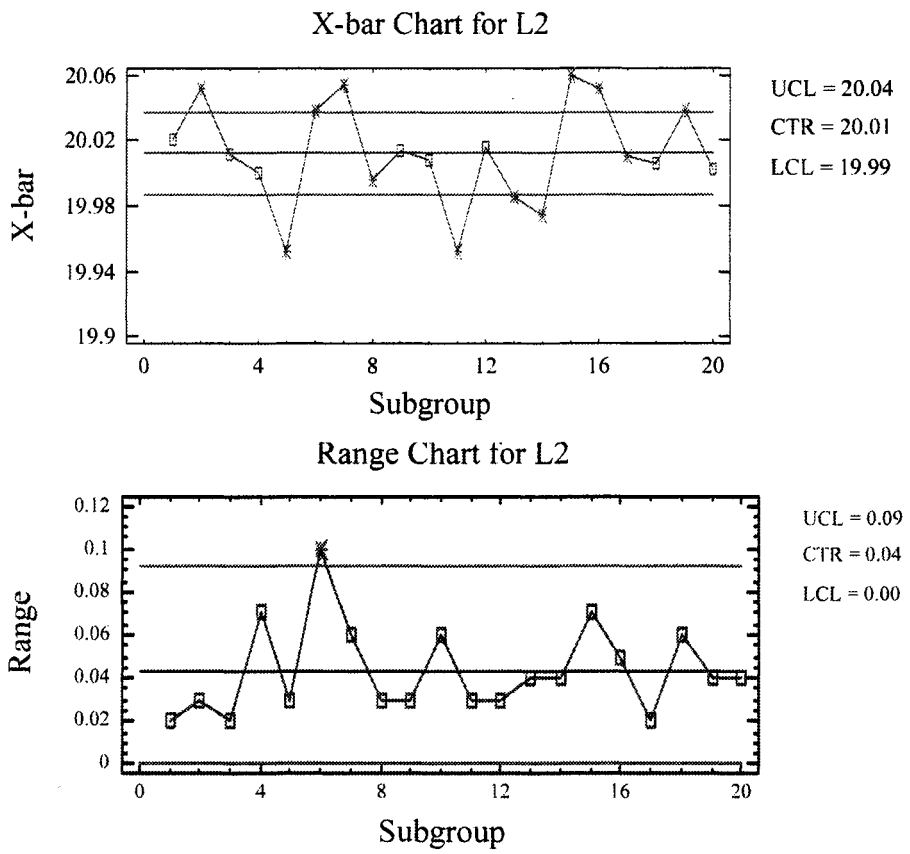
c) ¿Cómo eliminar aproximadamente el 80 % de los defectos?

Eliminando los tres primeros tipos de defectos, desaparece el 80% de los defectos. Esto nos permite gastar dinero, solo en la solución de los defectos principales, sin tener que preocuparnos de los demás.

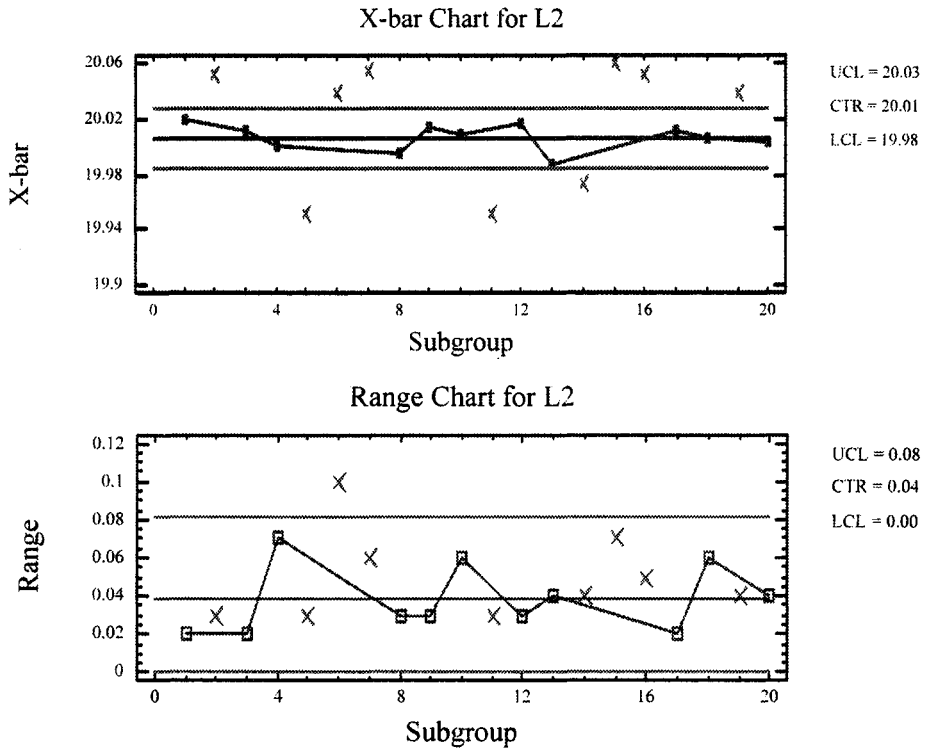
Solución Ejercicio 7.2.3.1

La variable longitud lado largo (L2), puede ser estudiada mediante gráficos de control por variables.

a) Obtener el gráfico de medias-recorridos.



b) Analizar el resultado, eliminando primero los datos fuera de límites, para que esté bajo control y posteriormente ver las tendencias.



Las cruces en rojo son puntos donde teníamos datos y los hemos eliminado por estar fuera de límites.

Hay muchos datos que se encuentran fuera de límites. Esto puede ser debido a la gran variabilidad entre los datos.

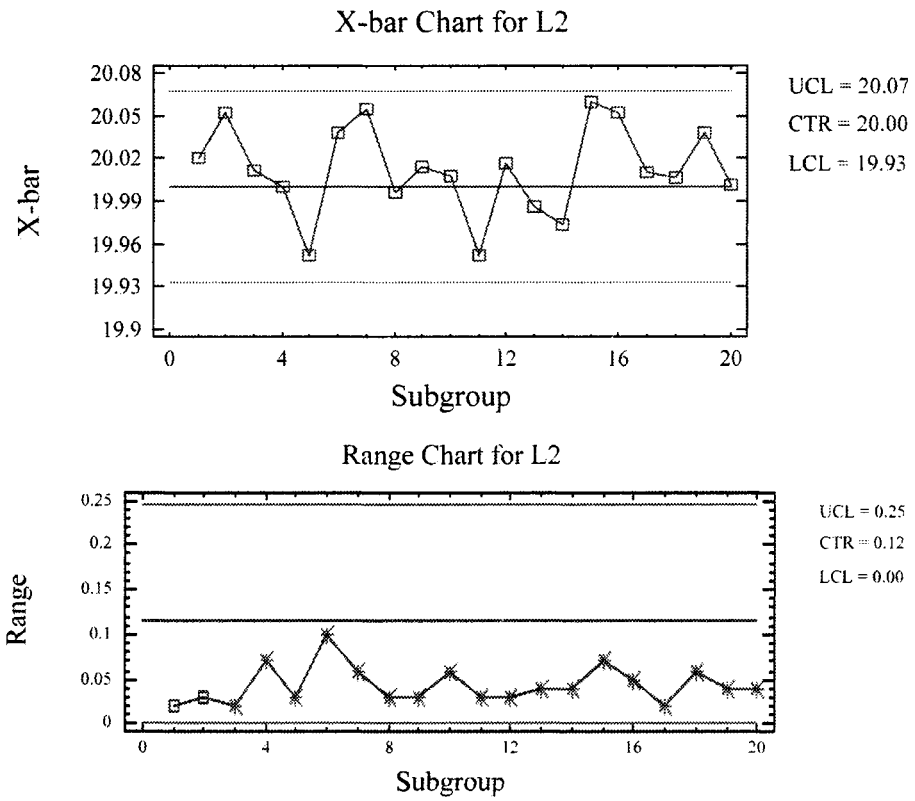
Una posible solución sería un mayor control de producción, para rectificar las tendencias antes de que ocurra el defecto.

Ocurre tanto en el gráfico de medias como en el de recorridos.

Tampoco pueden ser estudiadas series de datos, por la gran cantidad de datos excluidos.

c) Si inicialmente tenemos una media de 20,00 cm y desviación típica de 0,05 cm, obtener los gráficos de medias – recorridos.

Ahora será a partir de un control estándar anterior:



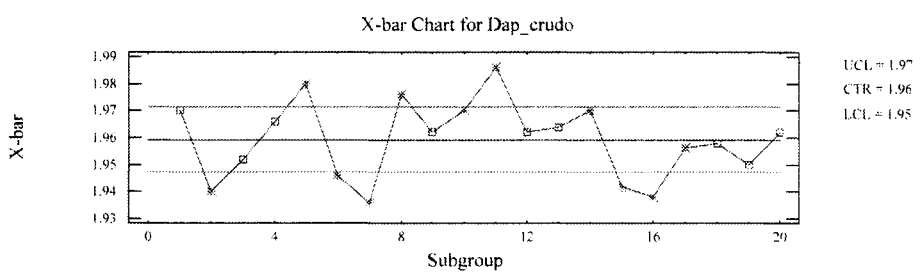
d) Si existe variación, realizar de nuevo el estudio.

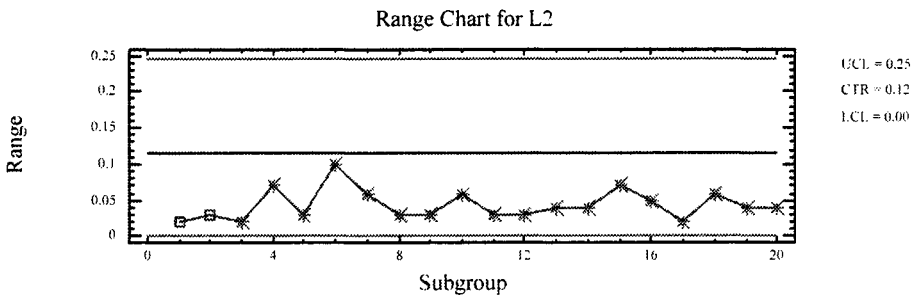
En este caso no tenemos datos fuera de límites.

En el gráfico de los rangos o recorridos, los datos están marcados en rojo ya que los límites son poco exigentes, por lo que en producción se podría ser más estricto, es decir, disminuir el límite superior de los rangos.

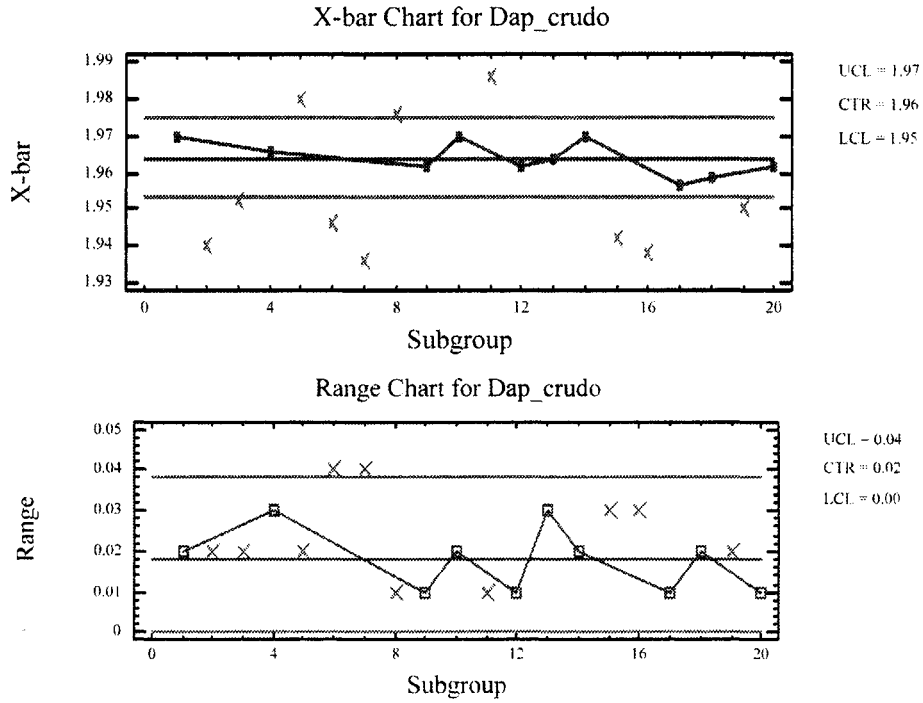
e) Aplicar el estudio de medias – recorridos para la variable de la densidad aparente en crudo.

Sin eliminar datos:





Eliminando los datos fuera de límites:



La explicación es la misma que en el caso anterior.

Solución Ejercicio 7.2.3.2

Otra opción es obtener los índices de capacidad de datos en forma de variable, con el camino de *Special >Quality Control > Variables Control Charts> X-bar and R* > en tabular options:

a) Calcular los índices de capacidad Icp e Icpk del lado largo (L2) con las especificaciones 20 ± 0,05 cm.

```
Capability Indices for L2

Specifications
     USL = 20.05
     Nominal = 20.0
     LSL = 19.95

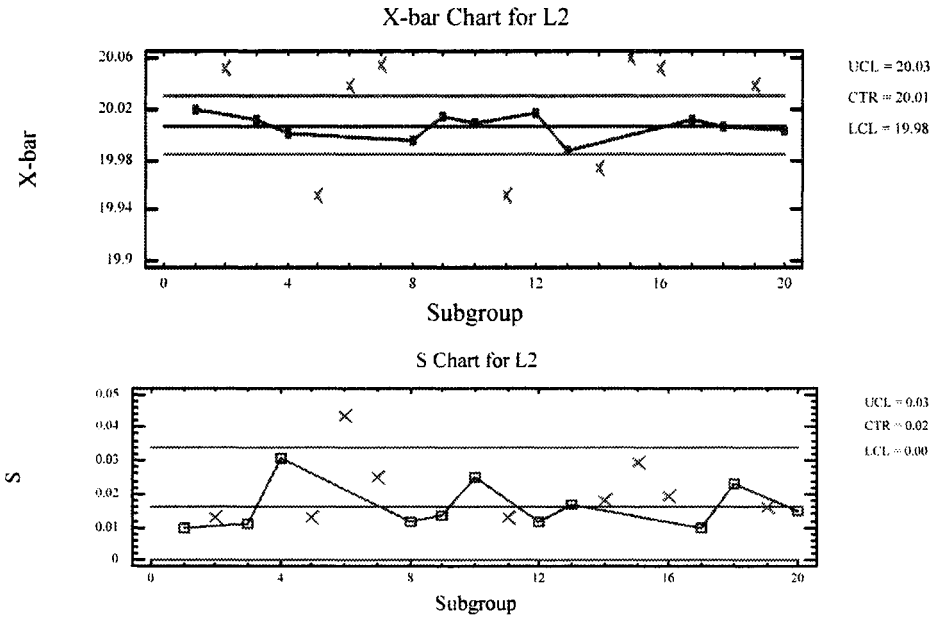
Cp = 0.891188
Cpk = 0.67552
Cpk (upper) = 0.67552
Cpk (lower) = 1.10686

Based on 6.0 sigma limits.


The StatAdvisor
---------------
   Several capability indices have been computed to summarize the
comparison of the data to the specifications.  One common index is Cp,
which equals the distance between the specification limits divided by
6 times the standard deviation.  In this case, Cp equals 0.891188,
which is usually considered to be not good.  Cpk is a one-sided
capability index, which divides the distance from the mean to the
nearer specification limit by 3 times the standard deviation.  In this
case, Cpk equals 0.67552.  The rather large difference between Cp and
Cpk is a sign that the distribution is not centered well between the
specification limits.
```

Solución Ejercicio 7.2.4.1

De igual forma que tenemos el gráfico de medias – recorridos, con las variables de las longitudes de los lados (L1 y L2) obtener el gráfico de medias – desviaciones típicas de L2. Analizar los resultados.

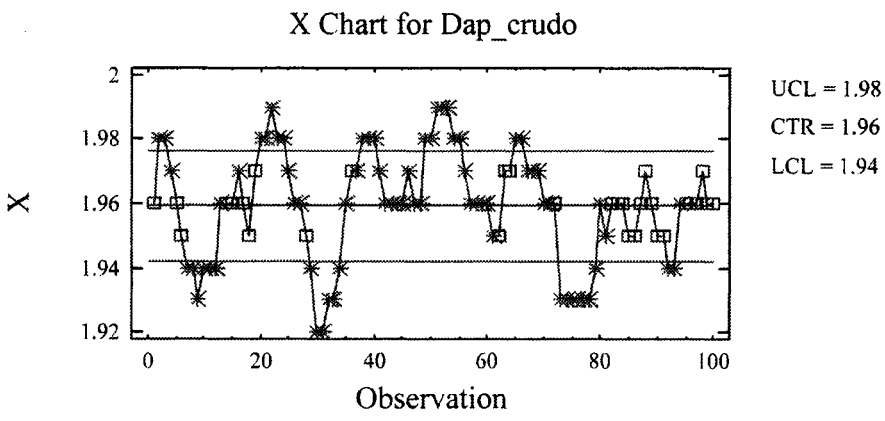


En este caso solo tenemos los gráficos después de eliminar los datos con *Exclude*. Ocurre lo mismo que en el caso anterior.

Solución Ejercicio 7.2.4.2

La variable densidad aparente en crudo puede ser estudiada mediante gráficos de control por variables individualizada .

a) Obtener el gráfico de datos individualizados.



b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o agrupando los datos para obtener el gráfico de medias-recorridos?

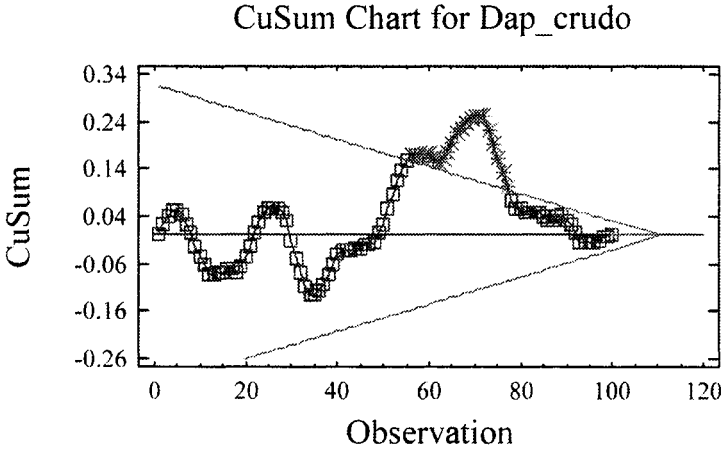
Se puede ver en este caso, que la mayoría de los datos están fuera de límites, existe mucha variación.

Normalmente, no se estudia de forma individualizada, ya que es posible en el los datos analizados tengamos algún problema concreto. Se estudiará mejor agrupando los datos por horas de control, turnos, etc...

Solución Ejercicio 7.2.4.3

De la misma forma que hemos estudiado la variable densidad aparente en crudo anteriormente, puede ser estudiada mediante gráficos de CuSum .

a) Obtener los gráficos de la máscara.




```
CuSum and MR(2) - Initial Study for Dap_crudo

Number of observations = 100
0 observations excluded

CuSum Chart
-----------
Detect shift of 1.0 sigma = 0.00573107
alpha = 0.01
beta = 0.01

V-mask at sample 100
```

```
MR(2) Chart
-----------
UCL: +3.0 sigma = 0.0211305
Centerline      = 0.00646465
LCL: -3.0 sigma = 0.0

1 beyond limits

Estimates
---------
Process mean = 1.9593
Process sigma = 0.00573107
Mean MR(2) = 0.00646465
```

b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o agrupando los datos para obtener el gráfico de medias-recorridos?

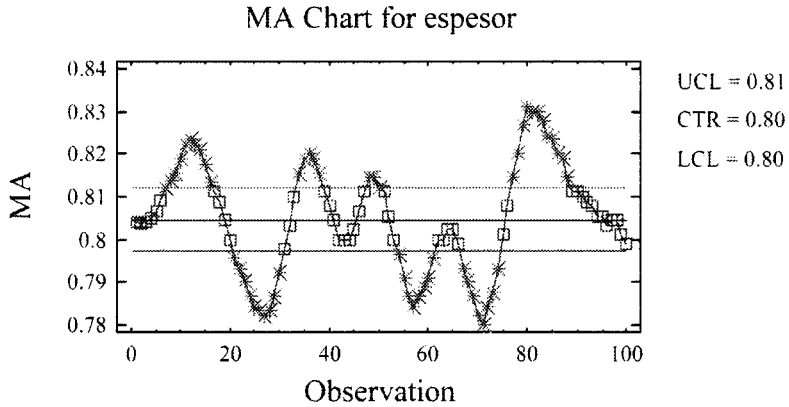
En este gráfico se estudian las sumas acumulativas con la máscara V. Se puede observar en este caso, que tenemos 22 puntos fuera de límites. Esto puede ser debido a un problema en la producción, por ejemplo un aumento de humedad que producirá aumentos puntuales en la densidad aparte. Cuando el operario ha detectado el problema, lo soluciona disminuyendo la presión de prensado.

Una buena opción es analizar ambos tipos de gráficos.

Solución Ejercicio 7.2.4.4

Los datos de la variable espesor de las piezas cerámicas, puede ser estudiada con estos gráficos,

Obtener los gráficos EM y EWMA del espesor de las piezas.



```
Moving Average and MR(2) - Initial Study for espesor

Number of observations = 100
0 observations excluded

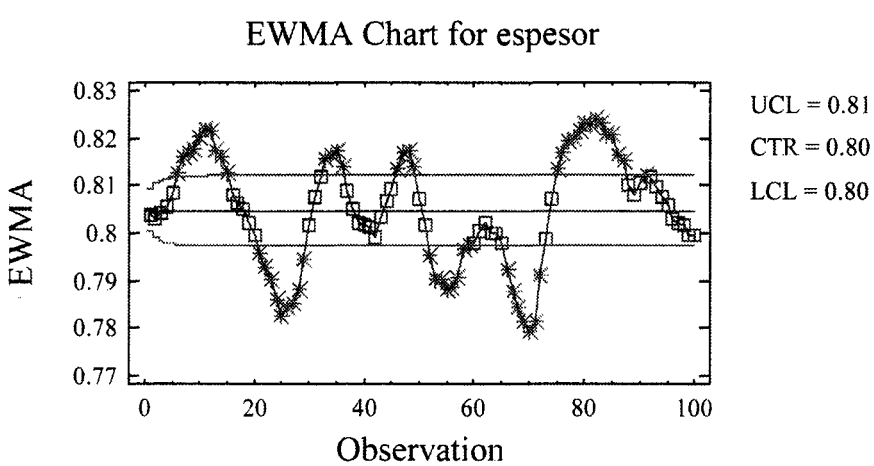
Moving Average Chart - Order = 9
--------------------
UCL: +3.0 sigma = 0.812043
Centerline      = 0.8047
LCL: -3.0 sigma = 0.797357

56 beyond limits

MR(2) Chart
-----------
UCL: +3.0 sigma = 0.0270734
Centerline      = 0.00828283
LCL: -3.0 sigma = 0.0

2 beyond limits

Estimates
---------
Process mean = 0.8047
Process sigma = 0.00734293
Mean MR(2) = 0.00828283
```



```
EWMA and MR(2) - Initial Study for espesor

Number of observations = 100
0 observations excluded

EWMA Chart - Lambda = 0.2
----------
UCL: +3.0 sigma = 0.812043
Centerline      = 0.8047
LCL: -3.0 sigma = 0.797357

56 beyond limits

MR(2) Chart
-----------
UCL: +3.0 sigma = 0.0270734
Centerline      = 0.00828283
LCL: -3.0 sigma = 0.0

2 beyond limits

Estimates
---------
Process mean = 0.8047
Process sigma = 0.00734293
Mean MR(2) = 0.00828283
```

En ambos casos, tanto en el gráfico de Media móviles y el de Medias móviles ponderadas, gran cantidad de datos fuera de control (56 datos). Esto nos indica la gran variabilidad en la variable espesor que solo podría ser detectada después de ver los distintos gráficos de control.

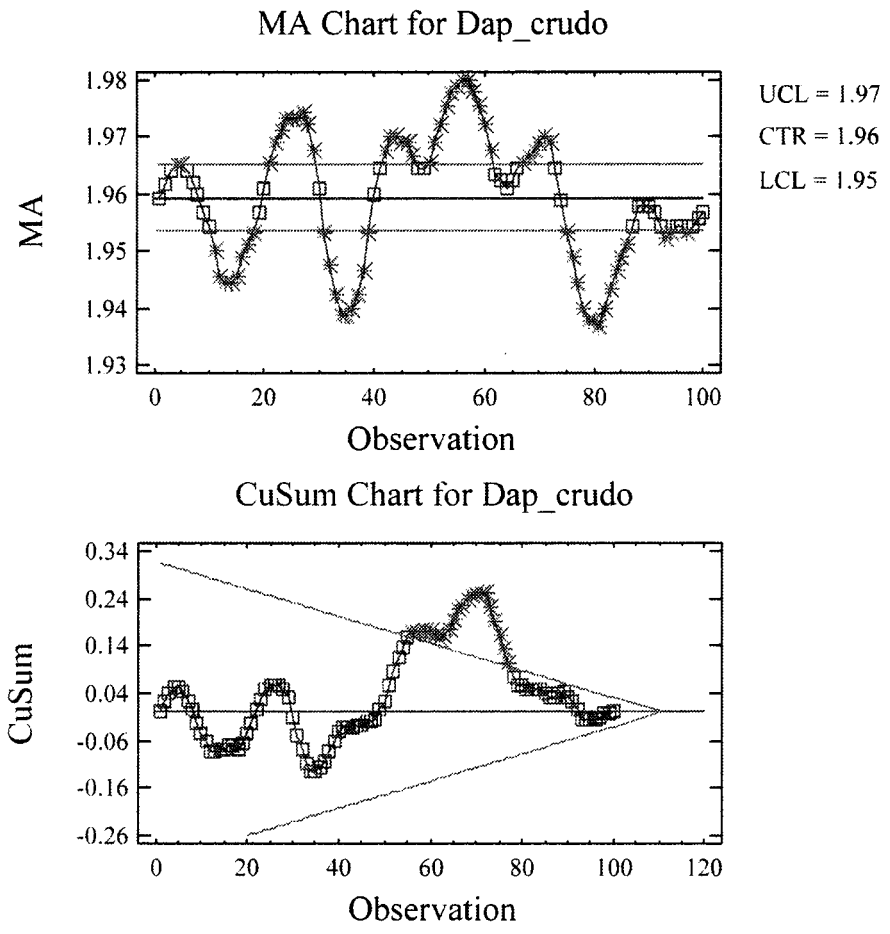
b) ¿Qué utilidad para la cerámica tienen estos gráficos? ¿Son comparables los resultados con otros estudios anteriores?

Pueden ser gráficos que sirvan de complemento a los anteriores. En todo caso, los gráficos más usados son los de medias-recorridos.

Solución Ejercicio 7.2.4.5

Para comparar con el estudio individualizado anterior, la variable densidad aparente en crudo puede ser estudiada mediante estos gráficos:

a) Obtener el gráfico de medias móviles y CuSum Individuales de la densidad aparente en crudo.



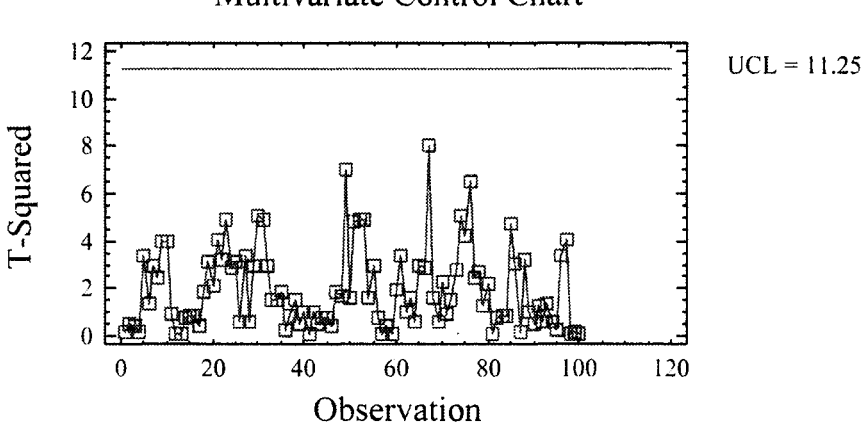
b) Análisis de los resultados. ¿Se estudia mejor esta variable de esta forma o mediante el procedimiento anterior?

Son complementarios

Solución Ejercicio 7.2.4.6

Estudiamos conjuntamente las longitudes de la pieza L1 y L2,

a) Obtener el gráfico Multivariate Control Chart de las dos longitudes.



Analysis Summary

Data variables:
L1
L2

Number of complete cases = 100
0 cases excluded

UCL: 11.2522 for alpha = 0.0027

0 beyond limits

The StatAdvisor

This procedure constructs a single control chart for the 2 data variables. Unlike most control charts which treat variables separately, this procedure takes into account possible correlations between the variables. It is assumed that the variables follow a multivariate normal distribution. The intent of the procedure is to determine whether all cases come from a single multivariate normal distribution.

The procedure constructs a control chart with an upper limit, which you can examine by selecting Multivariate Control Chart from the list of Graphical Options. In this case, the upper limit has been located so as to give a 0.27% false alarm rate. 0 cases were above the control limit. Since the probability of seeing 0 or more points above the control limit just by chance is 1.0 if the data comes from the assumed multivariate normal distribution, we cannot reject the hypothesis that the process is in a state of statistical control at the 95% confidence level.

b) Analizar el resultado y compararlo con los estudios previos de cada variable.

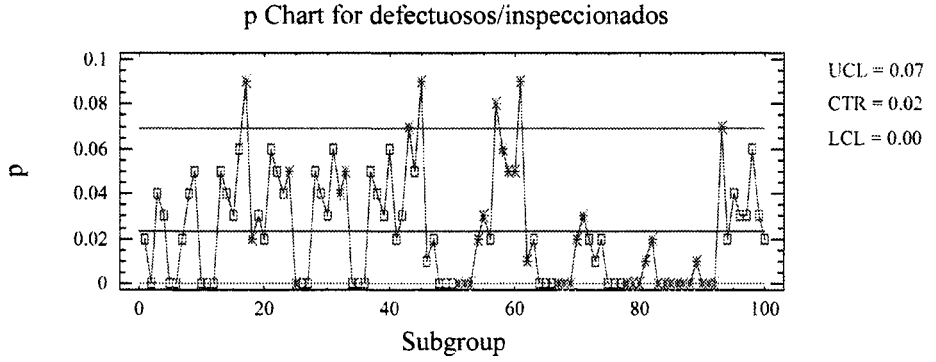
Nos sirve para ver la correlación entre variables, y comparar con el estudio de los gráficos de las variables por separado. En este caso, con los resultados que obtenemos, podemos suponer, al 95%, que el proceso está bajo control.

Solución Ejercicio 7.2.5.1

La variable defectos, nos muestra los defectos que aparecen cada día en el muestreo de las piezas cerámicas, mientras que inspeccionados nos dará las piezas inspeccionadas en cada momento.

a) Obtener el gráfico por atributos «p» para los defectos que se producen cada día (recordar que se muestrean 100 pieza día).

Como son siempre 100 piezas, podemos poner como tamaño del subgrupo 100.




```
p - Initial Study for defectuosos/inspeccionados

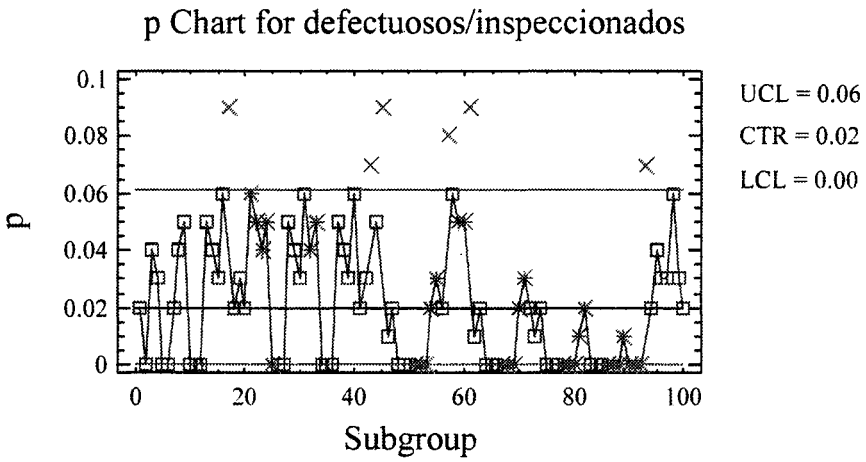
Number of subgroups = 100
Subgroup size = 100.0
0 subgroups excluded

p Chart
-------
UCL: +3.0 sigma = 0.0689455
Centerline      = 0.0235
LCL: -3.0 sigma = 0.0

6 beyond limits

Estimates
---------
Mean p = 0.0235
Sigma = 0.0151485
```

b) Eliminar los datos fuera de límites.
Eliminando los datos fuera de los límites de control:




```
p - Initial Study for defectuosos/inspeccionados

Number of subgroups = 94
Subgroup size = 100.0
6 subgroups excluded

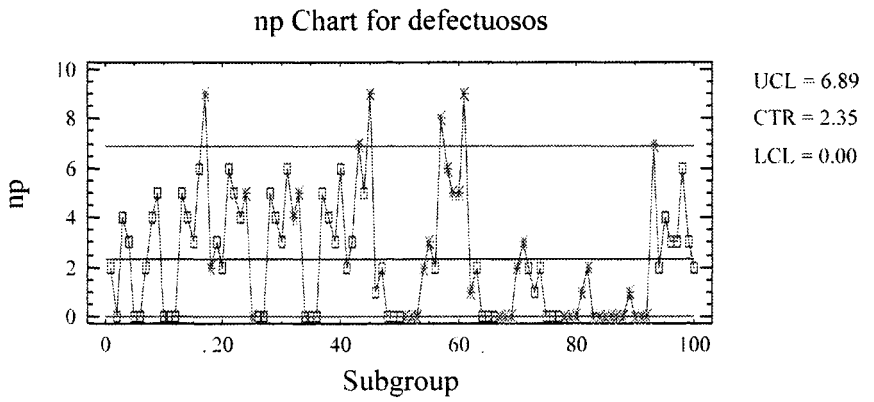
p Chart
-------
UCL: +3.0 sigma = 0.0615678
Centerline      = 0.0197872
LCL: -3.0 sigma = 0.0

0 beyond limits

Estimates
---------
Mean p = 0.0197872
Sigma = 0.0139268
```

Solución Ejercicio 7.2.5.2

Obtener el gráfico por atributos «np» para los defectos que se producen cada día.
Con la variable de defectos, y tomando que se inspeccionan 100 datos por día, obtenemos el siguiente gráfico:



Solución Ejercicio 7.2.5.3

Obtener el estudio por atributos «c» para los defectos que se producen cada día.

El gráfico que se obtiene es similar a los anteriores, solo cambiarán los parámetros:

```
c - Initial Study for defectuosos

Number of observations = 100
0 observations excluded

c Chart
-------
UCL: +3.0 sigma = 6.94891
Centerline      = 2.35
LCL: -3.0 sigma = 0.0

6 beyond limits

Estimates
---------
Mean c = 2.35
Sigma = 1.53297
```

Solución Ejercicio 7.2.5.4

Obtener el gráfico por atributos «u» para los defectos que se producen cada día.

De la misma forma que en los apartados anteriores, los gráficos son iguales. Con «Defects per unt»: defectuosas/100 y tamaño 100:

```
u - Initial Study for defectuosos/100

Number of subgroups = 100
Subgroup size = 100.0
0 subgroups excluded

u Chart
-------
UCL: +3.0 sigma = 0.0694891
Centerline      = 0.0235
LCL: -3.0 sigma = 0.0

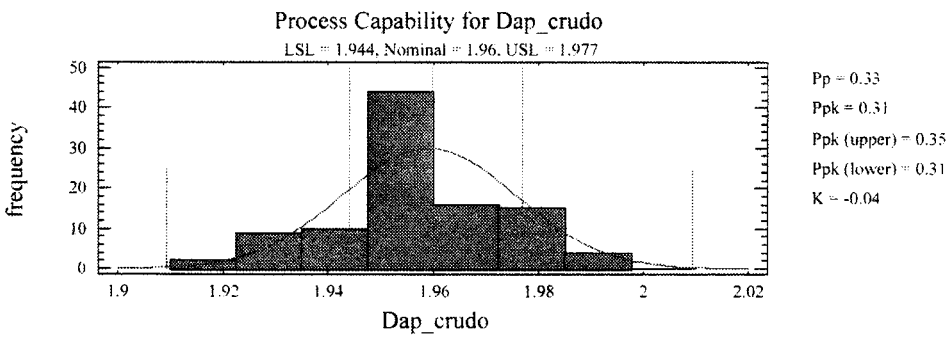
6 beyond limits

Estimates
---------
Mean u = 0.0235
Sigma = 0.0153297
```

Solución Ejercicio 7.2.6.1

Para estudiar las variables, o mejor dicho, el proceso o máquinas utilizadas en el control,

a) Realizar el estudio de capacidad para la variable densidad aparente en crudo valor nominal = 1,96 g/cc y los límites superior e inferior, ± 0,017).

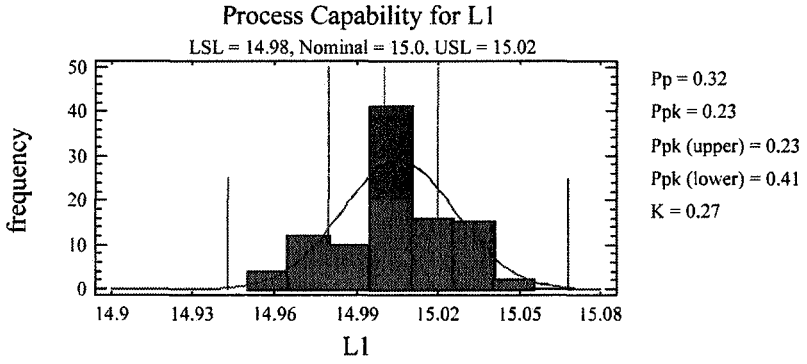


```
Distribution: Normal
     sample size = 100
     mean = 1.9593
     standard deviation = 0.0166518

6.0 Sigma Limits
     +3.0 sigma = 2.00926
     mean = 1.9593
     -3.0 sigma = 1.90934
```

Specifications	Observed Beyond Spec.	Z-Score	Estimated Beyond Spec.	Defects Per Mill
USL = 1.977	19.000000%	1.06	14.390232%	143902
Nominal = 1.96				
LSL = 1.944	21.000000%	-0.92	17.909432%	179094
Total	40.000000%		32.299665%	322996

b) Estudio de capacidad de las dos longitudes (L1 y L2)

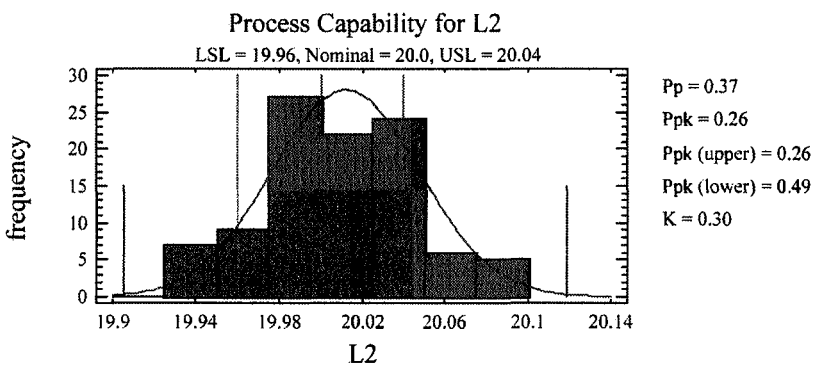


```
Data variable: L1

Distribution: Normal
     sample size = 100
     mean = 15.0055
     standard deviation = 0.0208591

6.0 Sigma Limits
     +3.0 sigma = 15.0681
     mean = 15.0055
     -3.0 sigma = 14.9429
```

Specifications		Observed Beyond Spec.	Z-Score	Estimated Beyond Spec.	Defects Per Million
USL =	15.02	17.000000%	0.70	24.348222%	243482.22
Nominal =	15.0				
LSL =	14.98	10.000000%	-1.22	11.076091%	110760.91
Total		27.000000%		35.424313%	354243.13



```
Data variable: L2

Distribution: Normal
     sample size = 100
     mean = 20.0121
     standard deviation = 0.035627

6.0 Sigma Limits
     +3.0 sigma = 20.119
     mean = 20.0121
     -3.0 sigma = 19.9052
```

Specifications		Observed Beyond Spec.	Z-Score	Estimated Beyond Spec.	Defects Per Million
USL =	20.04	15.000000%	0.78	21.677901%	216779.01
Nominal = 20.0					
LSL =	19.96	7.000000%	-1.46	7.181912%	71819.12
Total		22.000000%		28.859813%	288598.13

Solución Ejercicio 7.2.6.2

Obtener los Índices de capacidad del apartado anterior.

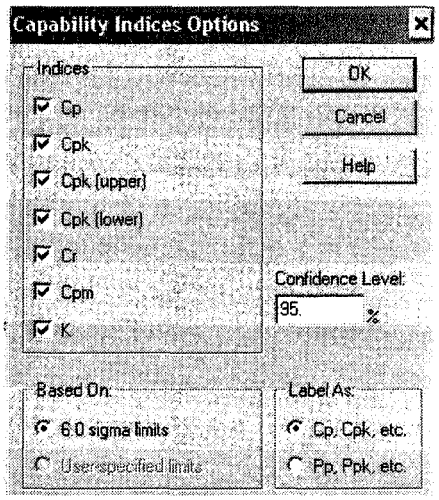
```
Capability Indices for L2

Cp = 0.374248
Cpk = 0.261038
Cpk (upper) = 0.261038
Cpk (lower) = 0.487458
Cr = 2.67202
Cpm = 0.354183
K = 0.3025

Based on 6.0 sigma limits.

95.0% Confidence Intervals
   Cp: (0.322162, 0.426247)
   Cpk: (0.186269, 0.335807)
   Cpm: (0.305387, 0.402891)
```

Con la siguiente ventana, podemos cambiar los índices y pasar de C a P (es decir, nos permite camibar las etiquetas de los índices).



Para L, utilizaremos, los índices en P:

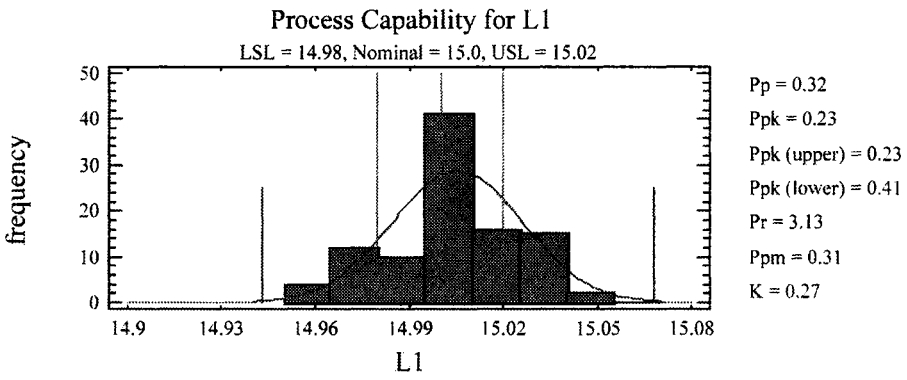
```
Capability Indices for L1

Pp = 0.319605
Ppk = 0.231714
Ppk (upper) = 0.231714
Ppk (lower) = 0.407496
Pr = 3.12886
Ppm = 0.308941
K = 0.275

Based on 6.0 sigma limits.

95.0% Confidence Intervals
   Pp: (0.275124, 0.364012)
   Ppk: (0.158844, 0.304583)
   Ppm: (0.266239, 0.351566)
```

Esto no hará cambiar el gráfico:



9 EJERCICIOS PROPUESTOS

1) Hemos visto distintas relaciones entre las variables, ¿Qué relación existe entre el espesor y la densidad aparente en crudo?

2) Hacer una comparación descriptiva de todas las variables de forma conjunta.

3) Obtener los índices de capacidad de las variables no estudiadas anteriormente.

4) Si se estudiasen otras variables como resistencia mecánica de las piezas o dureza de Mohs de las mismas, como debería afrontarse su estudio? Crear un nuevo fichero de datos con estas dos variables y ver su comportamiento. (Los valores de la resistencia mecánica pueden oscilar entre 15 y 20 kg/cm^2 por ejemplo y la dureza entre 0 y 10).

5) Si la empresa en la que se trabaja quiere implantar un Sistema de Calidad ISO 9000:2000 y quiere hacer un estudio previo de la producción, ¿Qué pasos se recomendarían en el caso de la necesidad de estudiar la estadística de los datos?

6) Obtener el gráfico de medias-recorridos de la variable L1 (longitud lado corto) y ver si la variabilidad es la misma que la observada con L2.

7) ¿Qué tipo de gráfico debería utilizarse en los siguientes casos?

- Viscosidad
- Densidad del esmalte
- Control de piezas de distintas calidades (primera, segunda...)
- Temperatura de salida de piezas del secadero.

8) En la empresa en que trabajas, se ha instalado una máquina que clasifica automáticamente según el tono. Crea un fichero de datos por tonos (distintos colores clasificados desde el 1 al 10) y realizar el estudio de capacidad de la máquina.

10 BIBLIOGRAFÍA

[1] Crosby, P.B. (1979), *Quality is free*, Ed: McGraw-Hill, Nueva York.

[2] De Domingo-Acinas, J. y Arranz-Molinero, A. (1997) *Calidad y Mejora Continua.* Editorial Donostiarra.

[3] Deming, W.E. (1982). *Quality, Productivity and Competitive position, MIT.*

[4] Duncan, A. J. (1974). *Quality control and Industrial Statistics.* Ed: R Irwin.

[5] Ishikawa, K. (1985). *What is total Quality Control?. The Japanese way.* Prentice Hall.

[6] Juan-Verdoy, P., Mateu, J., Sagasta, S. y Sirvent, R. (2003). *Control estadístico de calidad, con aplicaciones en ingeniería técnica mecánica en industria cerámica.* Materiales Docentes 174. Universidad Jaume I de Castellón

[7] Juran, J.M., Gryna, M.F y Bingham, R.S, (1990). *Manual de Control de Calidad.* Dd: Reverté. (2ª edición)

[8] Montgomery, D.C. (1985). *Control Estadístico de la Calidad.* Grupo Editorial Iberoamericana.

[9] Peña, D. (1986a). *Estadística, modelos y métodos.* Alianza Universidad Textos, Madrid.

[10] Peña, D. (1986b). *Control de calidad y diseño de experimentos.* Servicio de Publicaciones ETSII, Madrid.

[11] Pérez-Lopez, C. (2001). *Estadística Práctica con STATGRAPHICS. Ed:* Paerson Educación, S. A.

[12] Pérez-López, C. (2003) *Control Estadístico de la Calidad. Teoría, Práctica y Aplicaciones Informáticas.* Ed. Ra-Ma.

[13] Prat-Bartes, A, Tort-Martorell, X., Grima, P. y Pozueta, L. (1997). *Métodos estadísticos. Control y Mejora de la Calidad.* 1ª edición. Ed: U.P.C.

Páginas Web consultadas:

[1] www.tqm-manager.com.ar/herramientas/controlestadistico.htm

[2] www.tqm-manager.com.ar/ProbemlasCEP.htm

[3] www.uoc.edu/in3/e-math/docs/SPC_3.pdf

[4] www.ispjae.cu/eventos/colaeiq/html/cursos.htm

[5] halweb.uc3m.es/esp/docencia/MetodosMejora/ documentos/Spc.pdf

[6] halweb.uc3m.es/esp/Docencia/MEMC/Cap3_archivos/Cap3pdf.pdf

[7] halweb.uc3m.es/esp/Docencia/Memc/Cap5_archivos/Cap5pdf.pdf

[8] www.ifesaragon.org/obj1999/hera/paginas/conceptos/05hdc/unidad07.html

[9] www.uch.edu.ar/rrhh

[10] www.aiteco.com

[11] www.ceroaverias.com

[12] www.calidad.com

11 ANEXO TABLAS

11.1. DISTRIBUCIÓN NORMAL TIPIFICADA

Area bajo la densidad de la distribución normal estándar a la izquierda de z										
	Z~N(0,1)					P(Z <= z)				
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

11.2. TABLA POISSON

	λ									
x	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
0	0.3329	0.3012	0.2725	0.2466	0.2231	0.2019	0.1827	0.1653	0.1496	0.1353
1	0.6990	0.6626	0.6268	0.5918	0.5578	0.5249	0.4932	0.4628	0.4337	0.4060
2	0.9004	0.8795	0.8571	0.8335	0.8088	0.7834	0.7572	0.7306	0.7037	0.6767
3	0.9743	0.9662	0.9569	0.9463	0.9344	0.9212	0.9068	0.8913	0.8747	0.8571
4	0.9946	0.9923	0.9893	0.9857	0.9814	0.9763	0.9704	0.9636	0.9559	0.9473
5	0.9990	0.9985	0.9978	0.9968	0.9955	0.9940	0.9920	0.9896	0.9868	0.9834
6	0.9999	0.9997	0.9996	0.9994	0.9991	0.9987	0.9981	0.9974	0.9966	0.9955
7	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9996	0.9994	0.9992	0.9989
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9998
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
x	2.10	2.20	2.30	2.40	2.50	2.60	2.70	2.80	2.90	3.00
0	0.1225	0.1108	0.1003	0.0907	0.0821	0.0743	0.0672	0.0608	0.0550	0.0498
1	0.3796	0.3546	0.3309	0.3084	0.2873	0.2674	0.2487	0.2311	0.2146	0.1991
2	0.6496	0.6227	0.5960	0.5697	0.5438	0.5184	0.4936	0.4695	0.4460	0.4232
3	0.8386	0.8194	0.7993	0.7787	0.7576	0.7360	0.7141	0.6919	0.6696	0.6472
4	0.9379	0.9275	0.9162	0.9041	0.8912	0.8774	0.8629	0.8477	0.8318	0.8153
5	0.9796	0.9751	0.9700	0.9643	0.9580	0.9510	0.9433	0.9349	0.9258	0.9161
6	0.9941	0.9925	0.9906	0.9884	0.9858	0.9828	0.9794	0.9756	0.9713	0.9665
7	0.9985	0.9980	0.9974	0.9967	0.9958	0.9947	0.9934	0.9919	0.9901	0.9881
8	0.9997	0.9995	0.9994	0.9991	0.9989	0.9985	0.9981	0.9976	0.9969	0.9962
9	0.9999	0.9999	0.9999	0.9998	0.9997	0.9996	0.9995	0.9993	0.9991	0.9989
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

	λ									
x	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90	4.00
0	0.0450	0.0408	0.0369	0.0334	0.0302	0.0273	0.0247	0.0224	0.0202	0.0183
1	0.1847	0.1712	0.1586	0.1468	0.1359	0.1257	0.1162	0.1074	0.0992	0.0916
2	0.4012	0.3799	0.3594	0.3397	0.3208	0.3027	0.2854	0.2689	0.2531	0.2381
3	0.6248	0.6025	0.5803	0.5584	0.5366	0.5152	0.4942	0.4735	0.4532	0.4335
4	0.7982	0.7806	0.7626	0.7442	0.7254	0.7064	0.6872	0.6678	0.6484	0.6288
5	0.9057	0.8946	0.8829	0.8705	0.8576	0.8441	0.8301	0.8156	0.8006	0.7851
6	0.9612	0.9554	0.9490	0.9421	0.9347	0.9267	0.9182	0.9091	0.8995	0.8893
7	0.9858	0.9832	0.9802	0.9769	0.9733	0.9692	0.9648	0.9599	0.9546	0.9489
8	0.9953	0.9943	0.9931	0.9917	0.9901	0.9883	0.9863	0.9840	0.9815	0.9786
9	0.9986	0.9982	0.9978	0.9973	0.9967	0.9960	0.9952	0.9942	0.9931	0.9919
10	0.9996	0.9995	0.9994	0.9992	0.9990	0.9987	0.9984	0.9981	0.9977	0.9972
11	0.9999	0.9999	0.9998	0.9998	0.9997	0.9996	0.9995	0.9994	0.9993	0.9991
12	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

x	4.10	4.20	4.30	4.40	4.50	4.60	4.70	4.80	4.90	5.00
0	0.0166	0.0150	0.0136	0.0123	0.0111	0.0101	0.0091	0.0082	0.0074	0.0067
1	0.0845	0.0780	0.0719	0.0663	0.0611	0.0563	0.0518	0.0477	0.0439	0.0404
2	0.2238	0.2102	0.1974	0.1851	0.1736	0.1626	0.1523	0.1425	0.1333	0.1247
3	0.4142	0.3954	0.3772	0.3594	0.3423	0.3257	0.3097	0.2942	0.2793	0.2650
4	0.6093	0.5898	0.5704	0.5512	0.5321	0.5132	0.4946	0.4763	0.4582	0.4405
5	0.7693	0.7531	0.7367	0.7199	0.7029	0.6858	0.6684	0.6510	0.6335	0.6160
6	0.8786	0.8675	0.8558	0.8436	0.8311	0.8180	0.8046	0.7908	0.7767	0.7622
7	0.9427	0.9361	0.9290	0.9214	0.9134	0.9049	0.8960	0.8867	0.8769	0.8666
8	0.9755	0.9721	0.9683	0.9642	0.9597	0.9549	0.9497	0.9442	0.9382	0.9319
9	0.9905	0.9889	0.9871	0.9851	0.9829	0.9805	0.9778	0.9749	0.9717	0.9682
10	0.9966	0.9959	0.9952	0.9943	0.9933	0.9922	0.9910	0.9896	0.9880	0.9863
11	0.9989	0.9986	0.9983	0.9980	0.9976	0.9971	0.9966	0.9960	0.9953	0.9945
12	0.9997	0.9996	0.9995	0.9993	0.9992	0.9990	0.9988	0.9986	0.9983	0.9980
13	0.9999	0.9999	0.9998	0.9998	0.9997	0.9997	0.9996	0.9995	0.9994	0.9993
14	1.0000	1.0000	1.0000	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Índice

	λ									
x	6	6.5	7	7.5	8	8.5	9	9.5	10	15
0	0.0025	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000
1	0.0174	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0000
2	0.0620	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0000
3	0.1512	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0002
4	0.2851	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0009
5	0.4457	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0028
6	0.6063	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.0076
7	0.7440	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.0180
8	0.8472	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.0374
9	0.9161	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.0699
10	0.9574	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.1185
11	0.9799	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.1848
12	0.9912	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.2676
13	0.9964	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.3632
14	0.9986	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.4657
15	0.9995	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.5681
16	0.9998	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.6641
17	0.9999	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857	0.7489
18	1.0000	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928	0.8195
19	1.0000	1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965	0.8752
20	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9170
21	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9993	0.9469
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9997	0.9673
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9805
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9888
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9938
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9967
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9983
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
33	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
34	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
35	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

11.3. TABLA BINOMIAL

n	k	p	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,33	0,35	0,40	0,45	0,49	0,50
2	0		0,9801	0,9025	0,8100	0,7225	0,6400	0,5625	0,4900	0,4449	0,4225	0,3600	0,3025	0,2601	0,2500
	1		0,0198	0,0950	0,1300	0,2550	0,3200	0,3750	0,4200	0,4442	0,4550	0,4800	0,4950	0,4998	0,5000
	2		0,0001	0,0025	0,0100	0,0225	0,0400	0,0625	0,0900	0,1109	0,1225	0,1600	0,2025	0,2401	0,2500
3	0		0,9703	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2967	0,2746	0,2160	0,1664	0,1327	0,1250
	1		0,0294	0,1354	0,2430	0,3251	0,3840	0,4219	0,4410	0,4444	0,4436	0,4320	0,4084	0,3823	0,3750
	2		0,0003	0,0071	0,0270	0,0574	0,0960	0,1406	0,1890	0,2219	0,2339	0,2880	0,3341	0,3674	0,3750
	3		0,0000	0,0001	0,0010	0,0034	0,0080	0,0156	0,0270	0,0369	0,0429	0,0640	0,0911	0,1176	0,1250
4	0		0,9606	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1979	0,1785	0,1296	0,0915	0,0677	0,0625
	1		0,0388	0,1715	0,2916	0,3685	0,4096	0,4219	0,4116	0,3953	0,3845	0,3456	0,2995	0,2600	0,2500
	2		0,0006	0,0135	0,0486	0,0975	0,1536	0,2109	0,2646	0,2960	0,3105	0,3456	0,3675	0,3747	0,3750
	3		0,0000	0,0005	0,0036	0,0115	0,0256	0,0469	0,0756	0,0985	0,1115	0,1536	0,2005	0,2400	0,2500
	4		0,0000	0,0000	0,0001	0,0005	0,0016	0,0039	0,0081	0,0123	0,0150	0,0256	0,0410	0,0576	0,0625
5	0		0,9510	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1320	0,1150	0,0778	0,0503	0,0345	0,0313
	1		0,0480	0,2036	0,3281	0,3915	0,4096	0,3955	0,3602	0,3295	0,3124	0,2592	0,2059	0,1657	0,1563
	2		0,0010	0,0214	0,0729	0,1382	0,2048	0,2637	0,3087	0,3291	0,3354	0,3456	0,3369	0,3185	0,3125
	3		0,0000	0,0011	0,0081	0,0244	0,0512	0,0879	0,1323	0,1643	0,1811	0,2304	0,2757	0,3060	0,3125
	4		0,0000	0,0000	0,0005	0,0022	0,0064	0,0146	0,0284	0,0410	0,0488	0,0768	0,1128	0,1470	0,1563
	5		0,0000	0,0000	0,0000	0,0001	0,0003	0,0010	0,0024	0,0041	0,0053	0,0102	0,0185	0,0282	0,0313
6	0		0,9415	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0881	0,0754	0,0467	0,0277	0,0176	0,0156
	1		0,0571	0,2321	0,3543	0,3993	0,3932	0,3560	0,3025	0,2638	0,2437	0,1866	0,1359	0,1014	0,0938
	2		0,0014	0,0305	0,0984	0,1762	0,2458	0,2966	0,3241	0,3292	0,3280	0,3110	0,2780	0,2436	0,2344
	3		0,0000	0,0021	0,0146	0,0415	0,0819	0,1318	0,1852	0,2191	0,2355	0,2765	0,3032	0,3121	0,3125
	4		0,0000	0,0001	0,0012	0,0055	0,0154	0,0330	0,0595	0,0821	0,0951	0,1382	0,1861	0,2249	0,2344
	5		0,0000	0,0000	0,0001	0,0004	0,0015	0,0044	0,0102	0,0164	0,0205	0,0369	0,0609	0,0864	0,0938
	6		0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0007	0,0014	0,0018	0,0041	0,0083	0,0138	0,0156

7	0	0,9321	0,5983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0587	0,0490	0,0280	0,0152	0,0090	0,0078
	1	0,0659	0,2573	0,3720	0,3960	0,3670	0,3115	0,2471	0,2053	0,1848	0,1306	0,0872	0,0604	0,0547
	2	0,0020	0,0406	0,1240	0,2097	0,2753	0,3115	0,3177	0,3074	0,2985	0,2613	0,2140	0,1740	0,1641
	3	0,0000	0,0036	0,0230	0,0617	0,1147	0,1730	0,2269	0,2558	0,2679	0,2903	0,2918	0,2786	0,2734
	4	0,0000	0,0002	0,0026	0,0109	0,0287	0,0577	0,0972	0,1277	0,1442	0,1935	0,2388	0,2676	0,2734
	5	0,0000	0,0000	0,0002	0,0012	0,0043	0,0115	0,0250	0,0383	0,0466	0,0774	0,1172	0,1543	0,1641
	6	0,0000	0,0000	0,0000	0,0001	0,0004	0,0013	0,0036	0,0064	0,0084	0,0172	0,0320	0,0494	0,0547
	7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0005	0,0006	0,0016	0,0037	0,0068	0,0078
8	0	0,9227	0,5634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0392	0,0319	0,0168	0,0084	0,0046	0,0039
	1	0,0746	0,2793	0,3826	0,3847	0,3355	0,2670	0,1977	0,1565	0,1373	0,0896	0,0548	0,0352	0,0313
	2	0,0026	0,0515	0,1488	0,2376	0,2936	0,3115	0,2965	0,2734	0,2587	0,2090	0,1569	0,1183	0,1094
	3	0,0001	0,0054	0,0331	0,0839	0,1468	0,2076	0,2541	0,2730	0,2786	0,2787	0,2568	0,2273	0,2188
	4	0,0000	0,0004	0,0046	0,0185	0,0459	0,0865	0,1361	0,1704	0,1875	0,2322	0,2627	0,2730	0,2734
	5	0,0000	0,0000	0,0004	0,0026	0,0092	0,0231	0,0467	0,0680	0,0808	0,1239	0,1719	0,2098	0,2188
	6	0,0000	0,0000	0,0000	0,0002	0,0011	0,0038	0,0100	0,0170	0,0217	0,0413	0,0703	0,1008	0,1094
	7	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0012	0,0024	0,0033	0,0079	0,0164	0,0277	0,0313
	8	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0002	0,0007	0,0017	0,0033	0,0039
9	0	0,9135	0,5302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0261	0,0207	0,0101	0,0046	0,0023	0,0020
	1	0,0830	0,2985	0,3874	0,3679	0,3020	0,2253	0,1556	0,1174	0,1004	0,0605	0,0339	0,0202	0,0176
	2	0,0034	0,0629	0,1722	0,2597	0,3020	0,3003	0,2668	0,2345	0,2162	0,1612	0,1110	0,0776	0,0703
	3	0,0001	0,0077	0,0446	0,1069	0,1762	0,2336	0,2668	0,2731	0,2716	0,2508	0,2119	0,1739	0,1641
	4	0,0000	0,0006	0,0074	0,0283	0,0661	0,1168	0,1715	0,2045	0,2194	0,2508	0,2600	0,2506	0,2461
	5	0,0000	0,0000	0,0008	0,0050	0,0165	0,0389	0,0735	0,1021	0,1181	0,1672	0,2128	0,2408	0,2461
	6	0,0000	0,0000	0,0001	0,0006	0,0028	0,0087	0,0210	0,0340	0,0424	0,0743	0,1160	0,1542	0,1641
	7	0,0000	0,0000	0,0000	0,0000	0,0003	0,0012	0,0039	0,0073	0,0098	0,0212	0,0407	0,0635	0,0703
	8	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0009	0,0013	0,0035	0,0083	0,0153	0,0176
	9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0003	0,0008	0,0016	0,0020

10	0	0,9044	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0174	0,0135	0,0060	0,0025	0,0012	0,0010
	1	0,0914	0,3151	0,3874	0,3474	0,2684	0,1877	0,1211	0,0870	0,0725	0,0403	0,0207	0,0114	0,0098
	2	0,0042	0,0746	0,1937	0,2759	0,3020	0,2816	0,2335	0,1955	0,1757	0,1209	0,0763	0,0494	0,0439
	3	0,0001	0,0105	0,0574	0,1298	0,2013	0,2503	0,2668	0,2603	0,2522	0,2150	0,1665	0,1267	0,1172
	4	0,0000	0,0010	0,0112	0,0401	0,0881	0,1460	0,2001	0,2274	0,2377	0,2508	0,2384	0,2130	0,2051
	5	0,0000	0,0001	0,0015	0,0085	0,0264	0,0584	0,1029	0,1362	0,1536	0,2007	0,2340	0,2456	0,2461
	6	0,0000	0,0000	0,0001	0,0012	0,0055	0,0162	0,0368	0,0567	0,0689	0,1115	0,1596	0,1966	0,2051
	7	0,0000	0,0000	0,0000	0,0001	0,0008	0,0031	0,0090	0,0162	0,0212	0,0425	0,0746	0,1080	0,1172
	8	0,0000	0,0000	0,0000	0,0000	0,0001	0,0004	0,0014	0,0030	0,0043	0,0106	0,0229	0,0389	0,0439
	9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0005	0,0016	0,0042	0,0083	0,0098
	10	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0003	0,0008	0,0010