CONTROL ESTADÍSTICO DE PROCESOS

Dr. Josué Álvarez Borrego

1 Introducción

El control estadístico de procesos tuvo su inicio en los años 20's con el Dr. Shewhart, pero no fue hasta la Segunda Guerra Mundial cuando comenzó a aplicarse prácticamente en la industria.

Su desarrollo comenzó en Estados Unidos con la producción de artículos militares de bajo costo, buena calidad y en gran volumen. Sin embargo, no fue precisamente en Estados Unidos en donde tuvo su auge.

Posterior a la Segunda Guerra Mundial, los norteamericanos impusieron a la industria japonesa de telecomunicaciones que aplicara el control estadístico de calidad debido a que las fallas en el servicio telefónico eran excesivas. Japón utilizó las enseñanzas aprendidas en la industria de telecomunicaciones y las aplicó a las demás industrias. En un principio se realizaron diversos seminarios sobre control estadístico de procesos enfocándolos principalmente a la alta dirección, esto es, a directores y presidentes. El principal conferencista fue el Dr. Edwards Deming, el cual se considera que introdujo el control de calidad en Japón. Posteriormente, se creó una conciencia de participación total, es decir, se vieron involucrados todos los niveles dentro de las empresas y no únicamente la alta dirección.

En México, el tema de "control total de calidad" y otras de sus acepciones se ha puesto de moda en los últimos años. Lo anterior no se debe a la casualidad sino a la necesidad de subsistencia en mercados abiertos y competitivos.

Anteriormente, los consumidores, ya sea de productos o servicios, mostraban interés únicamente en el precio. Actualmente, se evalúa precio y calidad, debiendo existir un equilibrio entre estas dos variables.

Lo cierto es que la calidad ha pasado de ser un lujo a ser una necesidad. Hasta hace unos años, la documentación sobre el tema de "calidad" se basaba en industrias de manufactura. Sin embargo, la cultura de calidad, y en particular el control estadístico de procesos, se pueden aplicar a cualquier industria o sector económico.

Control estadístico del proceso

El método general es prescriptivo y descriptivo, no es analítico. Al controlar estadísticamente los procesos no se trata de moldear la distribución de datos reunidos en un proceso dado. Lo que se trata es de controlar el proceso con ayuda de reglas de decisión que localicen discrepancias apreciables entre los datos observados y las normas del proceso que se controla.

Se dice que un proceso esta bajo control estadístico cuando sólo se producen variaciones debidas a causas comunes. En otras palabras el objetivo y razón de ser control Estadístico de Procesos es ayudar a identificar las causas especiales que producen variaciones en el proceso y suministrar información para tomar decisiones.

Aplicaciones del control estadístico del proceso

Se aplica a todo: a las cosas, a las personas y a los actos. Determina y analiza rápidamente las causas que pueden originar desviaciones para que no vuelvan a presentarse en el futuro.

Existen cuatro factores que deben ser considerados al aplicar el proceso de control.

Cantidad, Tiempo, Costo y Calidad.

Su aplicación incide directamente en la racionalización de la administración y consecuentemente, en el logro de la productividad de todos los recursos de la empresa.

Este control estadístico de puede aplicar en todos los tipos de empresas donde se tiene un conjunto de operaciones materiales ejecutadas para la obtención, transformación o transporte de uno o varios productos.

Posibles ventajas de control estadístico de procesos.

Localiza los sectores responsables de la administración, desde el momento en que se establecen medidas correctivas.

Proporciona información acerca de la situación de la ejecución de los planes, sirviendo como fundamento al reiniciarse el proceso de la planeación.

Reduce costos y ahorra tiempo al evitar errores.

Determinar las causas asignables a este comportamiento y atacarlas y de esta manera mantener el proceso en control. Son herramientas de dirección que permiten:

1. Identificar en la muestra inicial del proceso las observaciones atípicas, a fin de excluirlas una vez detectadas las causas asignables y no tomarlas en consideración para estimar los parámetros del proceso.
2. Detectar a tiempo anormalidades en el proceso, tanto por corrimientos de la media, como incrementos en la desviación por encima de sus límites naturales, para impedir la producción de piezas fuera de especificación.

Economía en la realización de la investigación y la rapidez en la obtención de resultados.

El aumento creciente de calidad de los productos.

La desviación puede ser identificable y posible de eliminar.

Adopción de decisiones a corto y largo plazo. Las decisiones a corto plazo se dan cuando se inicia una labor de investigación como resultado de un síntoma de anomalía indicado por el control estadístico. Las decisiones a largo plazo son consecuencia de una decisión de incluir o excluir ciertos datos en el estándar y los limites de control futuro.

Proporciona evidencias para investigar la causa de malos resultados.

La mayor ventaja es el de detectar un deterioro no deseado del proceso. El deterioro puede obedecer a múltiples causas. Es fácil detectar y ajustar el tipo de deterioro que pasa a uno de otro nivel.

La desviación que es una variación aleatoria es calculada y esperada a que ocurra k veces de cada mil.

Posibles perjuicios del control estadístico de proceso.

Si no se utilizan o interpretan adecuadamente los gráficos de control o se toman datos erróneos se puede tener una gran ineficiencia en el control estadístico de proceso.
Si se toma un muestreo del proceso existe un cierto porcentaje de error y de confiabilidad en todos esos elementos de muestra.

Conceptos y fundamentos estadísticos.

Para el entendimiento del *Control Estadístico de Procesos* no es necesario ser un experto en estadística, pero es preciso recordar al menos los puntos que se describen a continuación.

a) Distribución Normal o Campana de Gauss.

La distribución normal es desde luego la función de densidad de probabilidad "estrella" en estadística. Depende de dos parámetros μ y s, que son la media y la desviación típica respectivamente. Tiene una forma acampanada (de ahí su nombre) y es simétrica respecto a μ. Llevando múltiplos de s a ambos lados de μ, nos encontramos con que el 68% de la población está contenido en un entorno ±1s alrededor de μ, el 95% de la población está contenido en un entorno ±2s alrededor de μ y que el 99,73% está comprendido en ±3s alrededor de μ.

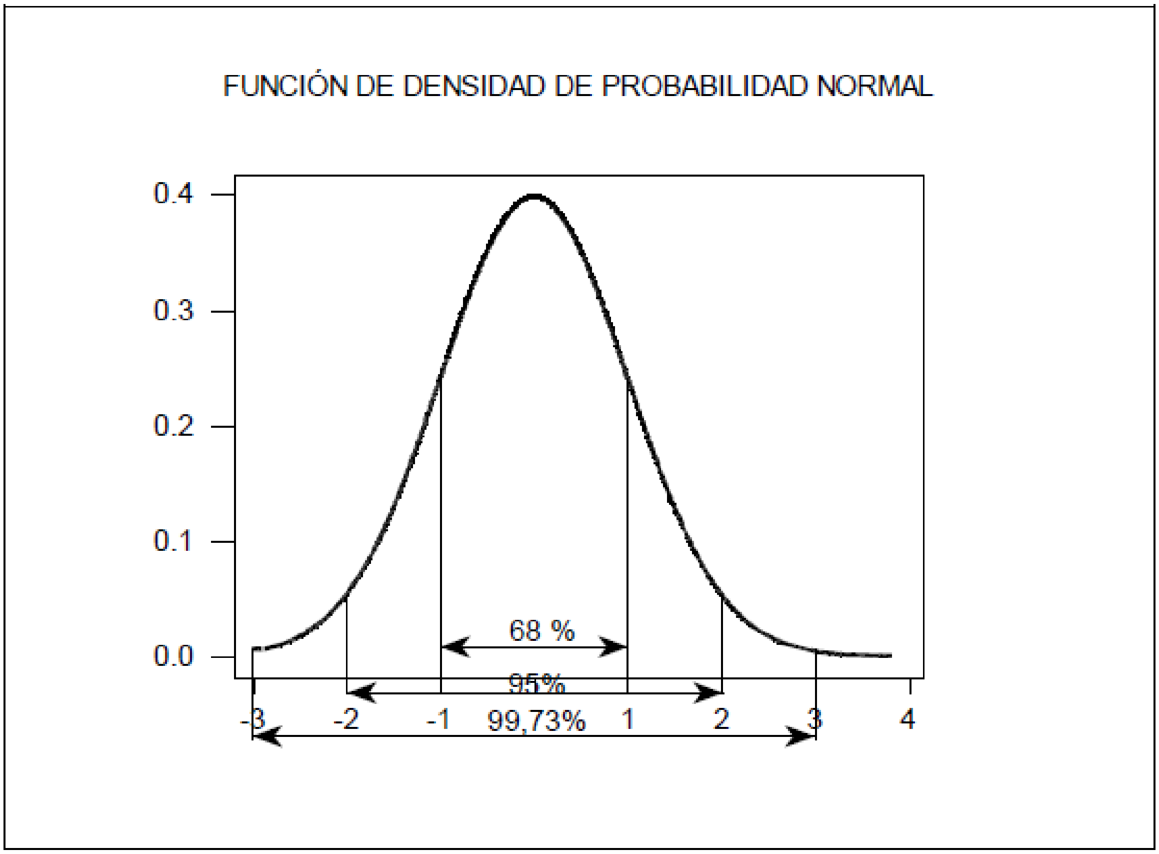


Figura1. Función de densidad de Probabilidad Normal.

2 Describiendo y resumiendo datos.

Dado un grupo de datos organizados en una distribución de frecuencias (o bien una serie de observaciones sin ordenar), pretendemos describirlos mediante dos o tres cantidades sintéticas.

En este sentido pueden examinarse varias características, siendo las más comunes:

a) La tendencia central de los datos;
b) La dispersión o variación con respecto a este centro;
c) Los datos que ocupan ciertas posiciones.
d) La simetría de los datos.
e) La forma en la que los datos se agrupan.

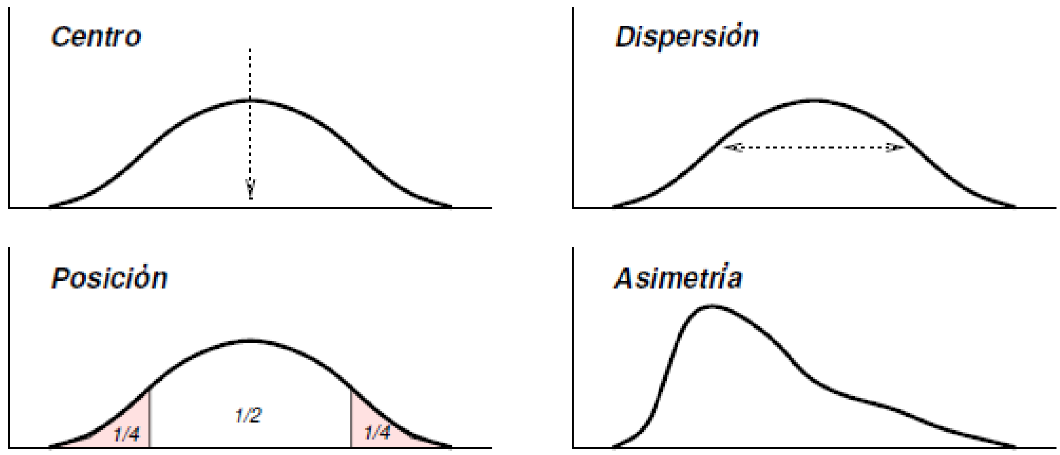


Figura 2. Medidas representativas de un conjunto de datos estadísticos.

Estadísticos de tendencia central

Las tres medidas más usuales de tendencia central son:

a) la media,
b) la mediana,
c) la moda.

En ciertas ocasiones estos tres estadísticos suelen coincidir, aunque generalmente no es así. Cada uno de ellos presenta ventajas e inconvenientes que precisaremos más adelante. En primer lugar vamos a definir los conceptos anteriores.

La media

La media aritmética de una variable estadística es la suma de todos sus posibles valores, ponderada por las frecuencias de los mismos. Es decir, si la tabla de valores de una variable X es

X	n_i	f_i
x_1	n_1	f_1
$\dots$	$\dots$	$\dots$
x_k	n_k	f_k

la media es el valor que podemos escribir de las siguientes formas equivalentes:

$$\begin{aligned}\overline{x} &= x_1\, f_1 + \ldots + x_k\, f_k \\ &= \frac{1}{n}\,(x_1\, n_1 + \ldots x_k\, n_k) \\ &= \frac{1}{n}\sum_{i=1}^{k} x_i\, n_i\end{aligned}$$

Si los datos no están ordenados en una tabla, entonces

$$\boxed{\overline{x} = \frac{x_1 + \ldots + x_n}{n}}$$

Algunos inconvenientes de la media

La media presenta inconvenientes en algunas situaciones:

Uno de ellos es que es muy sensible a los valores extremos de la variable: ya que todas las observaciones intervienen en el cálculo de la media, la aparición de una observación extrema, hará que la media se desplace en esa dirección. En consecuencia,

- No es recomendable usar la media como medida central en las distribuciones muy asimétricas;

- Si consideramos una variable discreta, por ejemplo, el número de hijos en las familias mexicanas el valor de la media puede no pertenecer al conjunto de valores de la variable; Por ejemplo x = 1, 2 hijos.

Otras medias: Medias generalizadas

En función del tipo de problema varias generalizaciones de la media pueden ser consideradas. He aquí algunas de ellas aplicadas a unas observaciones $x_1, ..., x_n$:

La media geométrica $\overline{x}_g$, es la media de los logaritmos de los valores de la variable:

$$\log \overline{x}_g = \frac{\log x_1 + \ldots + \log x_n}{n}$$

Luego

$$\overline{x}_g = \sqrt[n]{x_1 \, x_2 \, \ldots \, x_n}$$

Si los datos están agrupados en una tabla, entonces se tiene:

$$\overline{x}_g = \sqrt[n]{x_1^{n_1} \, x_2^{n_2} \, \ldots \, x_k^{n_k}}$$

La media armónica $\overline{x}_a$, se define como el recíproco de la media aritmética de los recíprocos, es decir,

$$\frac{1}{\overline{x}_a} = \frac{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}{n}$$

Por tanto

$$\overline{x}_a = \frac{n}{\frac{1}{x_1} + \ldots + \frac{1}{x_n}}$$

La mediana

Consideramos una variable discreta **X** cuyas observaciones en una tabla estadística han sido ordenadas de menor a mayor. Llamaremos **mediana**, M_{ed} al primer valor de la variable que deja por debajo de sí al 50% de las observaciones.

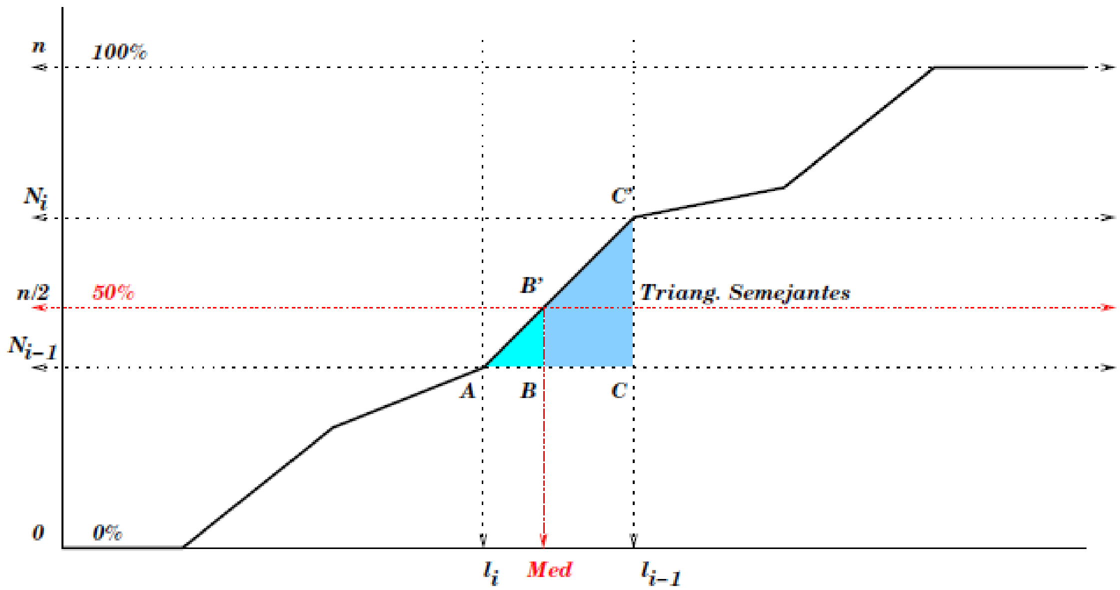


Figura 3. Cálculo geométrico de la mediana.

En el caso de variables continuas, las clases vienen dadas por intervalos, y aquí la fórmula de la mediana se complica un poco más (pero no demasiado): Sea $(l_{i-1}, l_i]$ el intervalo donde hemos encontrado que por debajo están el 50% de las observaciones. Entonces se obtiene la mediana a partir de las frecuencias absolutas acumuladas, mediante interpolación lineal (teorema de Thales) como sigue (figura 3):

$$\frac{CC'}{AC} = \frac{BB'}{AB} \quad \Longrightarrow \quad \frac{n_i}{a_i} = \frac{\frac{n}{2} - N_{i-1}}{M_{ed} - l_{i-1}}$$

$$\Longrightarrow \quad \boxed{M_{ed} = l_{i-1} + \frac{\frac{n}{2} - N_{i-1}}{n_i} \cdot a_i}$$

Esto equivale a decir que la mediana divide al histograma en dos partes de áreas iguales a 1/2 .

Propiedades de la mediana

Entre las propiedades de la mediana, vamos a destacar las siguientes:

- Como medida descriptiva, tiene la ventaja de no estar afectada por las observaciones extremas, ya que no depende de los valores que toma la variable, sino del orden de las mismas. Por ello es adecuado su uso en distribuciones asimétricas.

- Es de cálculo rápido y de interpretación sencilla.

- A diferencia de la media, la mediana de una variable discreta es siempre un valor de la variable que estudiamos (ej. La mediana de una variable número de hijos toma siempre valores enteros).

Un ejemplo de cálculo de mediana

Sea X una variable discreta que ha presentado sobre una muestra las modalidades

$$X \leadsto 2, 5, 7, 9, 12 \Longrightarrow \overline{x} = 7, \qquad M_{ed} = 7$$

Si cambiamos la última observación por otra anormalmente grande, esto no afecta a la mediana, pero si a la media:

$$X \rightsquigarrow 2, 5, 7, 9, 125 \Longrightarrow \overline{x} = 29,6; \qquad M_{ed} = 7$$

En este caso la media no es un posible valor de la variable (discreta), y se ha visto muy afectada por la observación extrema. Este no ha sido el caso para la mediana.

Un ejemplo de cálculo de media y mediana

Obtener la media aritmética y la mediana en la distribución adjunta. Determinar gráficamente cuál de los dos promedios es más significativo.

$l_{i-1} - l_i$	n_i
$0 - 10$	60
$10 - 20$	80
$20 - 30$	30
$30 - 100$	20
$100 - 500$	10

Solución:

$l_{i-1} - l_i$	n_i	a_i	x_i	$x_i n_i$	N_i	$n_i{}'$
$0 - 10$	60	10	5	300	60	60
$10 - 20$	80	10	15	1.200	140	80
$20 - 30$	30	10	25	750	170	30
$30 - 100$	20	70	65	1.300	190	2,9
$100 - 500$	10	400	300	3.000	200	0,25
	$n = 200$			$\sum x_i n_i = 6{,}550$		

La media aritmética es:

$$\overline{x} = \frac{1}{n} \sum x_i = \frac{6{,}550}{200} = 32,75$$

La primera frecuencia absoluta acumulada que supera el valor n/2 = 100 es Ni = 140. Por ello el intervalo mediano es [10; 20). Así:

$$M_{ed} = l_{i-1} + \frac{n/2 - N_{i-1}}{n_i} \cdot a_i = 10 + \frac{100 - 60}{80} \times 10 = 15$$

Para ver la representatividad de ambos promedios, realizamos el histograma de la figura 4, y observamos que dada la forma de la distribución, la mediana es más representativa que la media.

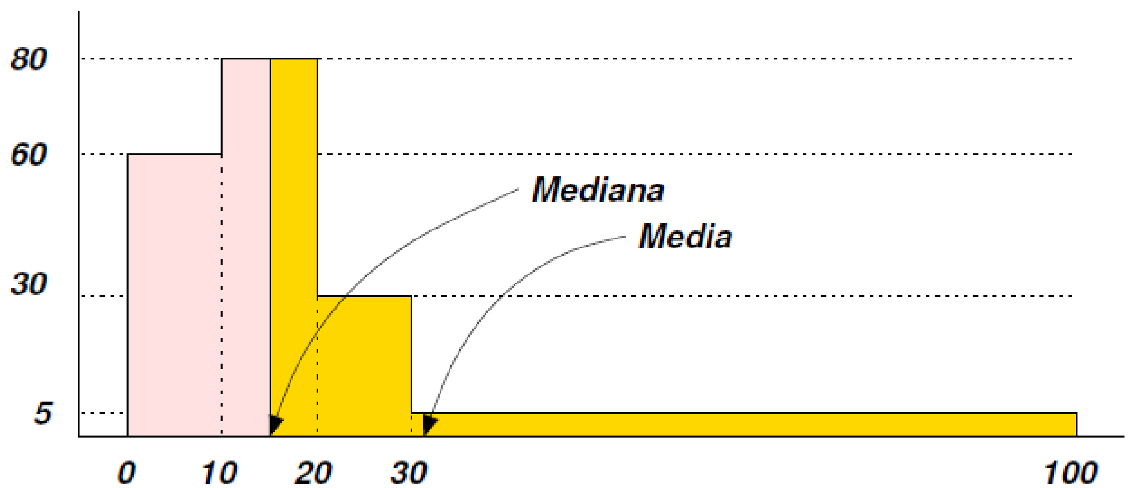


Figura 4. Para esta distribución de frecuencias es más representativo usar como estadístico de tendencia central la mediana que la media.

La moda

Llamaremos moda a cualquier máximo relativo de la distribución de frecuencias, es decir, cualquier valor de la variable que posea una frecuencia mayor que su anterior y su posterior.

Observación

De la moda destacamos las siguientes propiedades:

Es muy fácil de calcular.

Puede no ser única.

	Medidas de tendencia central	
	Datos sin agrupar	Datos agrupados
	(ordenados) $x_1,\ x_2,\ \ldots,\ x_N$	$\begin{array}{cccc} \text{Interv.} & x_i & n_i & N_i \\ \hline l_0—l_1 & x_1 & n_1 & N_1 \\ l_1—l_2 & x_2 & n_2 & N_2 \\ \ldots & \ldots & \ldots & \ldots \\ l_{k-1}—l_k & x_k & n_k & N_k \end{array}$
Media	$\overline{x} = \dfrac{x_1 + \cdots + x_n}{N}$	$\overline{x} = \dfrac{n_1\, x_1 + \cdots + n_k\, x_k}{N}$
Mediana	Primera observación que deja debajo de sí estrictamente a las $[N/2]$ observaciones menores: $x_{[N/2]+1}$	$M_{ed} = l_{i-1} + \dfrac{\frac{N}{2} - N_{i-1}}{n_i} \cdot a_i$
Moda	$M_{oda} = x_i$ de mayor frecuencia	$M_{oda} == l_{i-1} + \frac{n'_i - n'_{i-1}}{(n'_i - n'_{i-1}) + (n'_i - n'_{i+1})}\, a_i$

Cuadro 1: Resumen de las medidas de posiciones centrales.

Relación entre media, mediana y moda

En el caso de distribuciones unimodales, la mediana está con frecuencia comprendida entre la media y la moda (incluso más cerca de la media).

En distribuciones que presentan cierta inclinación, es más aconsejable el uso de la mediana. Sin embargo en estudios relacionados con propósitos estadísticos y de inferencia suele ser más apta la media.

Medidas de variabilidad o dispersión

Los estadísticos de tendencia central o posición nos indican donde se sitúa un grupo de puntuaciones. Los de variabilidad o dispersión nos indican si esas puntuaciones o valores están próximas entre sí o si por el contrario están muy dispersas.

Rango

Una medida razonable de la variabilidad podría ser la amplitud o rango, que se obtiene restando el valor más bajo de un conjunto de observaciones del valor más alto.

Propiedades del rango

- Es fácil de calcular y sus unidades son las mismas que las de la variable.
- No utiliza todas las observaciones (sólo dos de ellas);
- Se puede ver muy afectada por alguna observación extrema;
- El rango aumenta con el número de observaciones, o bien se queda igual. En cualquier caso nunca disminuye.

Varianza

La varianza, S^2, se define como la media de las diferencias cuadráticas de n puntuaciones con respecto a su media aritmética, es decir

$$\boxed{S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Esta medida es siempre una cantidad positiva, con propiedades interesante para la realización de inferencia estadística. Como sus unidades son las del cuadrado de la variable, es más sencillo usar su raíz cuadrada, que es la que vemos en la siguiente sección.

Desviación típica o estándar

La varianza no tiene la misma magnitud que las observaciones (ej. si las observaciones se miden en metros, la varianza lo hace en metros cuadrados.

Si queremos que la medida de dispersión sea de la misma dimensionalidad que las observaciones bastarán con tomar su raíz cuadrada. Por ello se define la **desviación típica**, S, como

$$\mathcal{S} = \sqrt{\mathcal{S}^2}$$

Ejemplo de cálculo de medidas de dispersión

Calcular el rango, varianza y desviación típica de las siguientes cantidades medidas en metros:

3, 3, 4, 4, 5

<u>Solución</u>: El rango de esas observaciones es la diferencia entre la mayor y menor de ellas, es decir, 5−3 = 2. Para calcular las restantes medidas de dispersión es necesario calcular previamente el valor con respecto al cual vamos a medir las diferencias. Éste es la media:

$$\overline{x} = (3+3+4+4+5)/5 = 3,8 \text{ metros}$$

La variancia es

$$\mathcal{S}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}^2 = \frac{1}{5}\left(3^2+3^2+4^2+4^2+5^2\right) - 3,8^2 = 0,56 \text{ metros}^2$$

Siendo la desviación típica su raíz cuadrada

$$\mathcal{S} = \sqrt{\mathcal{S}^2} = \sqrt{0,56} = 0,748 \text{ metros}$$

Propiedades de la varianza y desviación típica

- Ambas son sensibles a la variación de cada una de las puntuaciones, es decir, si una puntuación cambia, cambia con ella la varianza. La razón es que si miramos su definición, la varianza es función de cada una de las puntuaciones.
- La desviación típica tiene la propiedad de que en el intervalo

$$(\overline{x} - 2\,S, \overline{x} + 2\,S) \overset{\text{def}}{\sim} \overline{x} \pm 2\,S$$

se encuentra, al menos, el 75% de las observaciones Incluso si tenemos muchos datos y estos provienen de una distribución normal (se definirá este concepto más adelante), podremos llegar al 95 %.

- No es recomendable el uso de ellas, cuando tampoco lo sea el de la media como medida de tendencia central.

Coeficiente de variación

Hemos visto que las medidas de centralización y dispersión nos dan información sobre una muestra. Nos podemos preguntar si tiene sentido usar estas magnitudes para comparar dos poblaciones. Por ejemplo, si nos piden comparar la dispersión de los pesos de las poblaciones de elefantes de dos circos diferentes, S nos dará información útil. ¿Pero qué ocurre si lo que comparamos es la altura de unos elefantes con respecto a su peso? Tanto la media como la desviación típica, $\overline{x}$ y S, se expresan en las mismas unidades que la variable. Por ejemplo, en la variable altura podemos usar como unidad de longitud el metro y en la variable peso, el kilogramo. Comparar una desviación (con respecto a la media) medida en metros con otra en kilogramos no tiene ningún sentido.

El problema no deriva sólo de que una de las medidas sea de longitud y la otra sea de masa. El mismo problema se plantea si medimos cierta cantidad, por ejemplo la masa, de dos poblaciones, pero con distintas unidades. Este es el caso en que comparamos el peso en toneladas de una población de 100 elefantes con el correspondiente en miligramos de una población de 50 hormigas.

El problema no se resuelve tomando las mismas escalas para ambas poblaciones. Por ejemplo, se nos puede ocurrir medir a las hormigas con las mismas unidades que los elefantes (toneladas). Si la ingeniería genética no nos sorprende con alguna barbaridad, lo lógico es que la dispersión de la variable peso de las hormigas sea prácticamente nula (¡Aunque haya algunas que sean 1.000 veces mayores que otras!)

En los dos primeros casos mencionados anteriormente, el problema viene de la dimensionalidad de las variables, y en el tercero de la diferencia enorme entre las medias de ambas poblaciones. El coeficiente de variación es lo que nos permite evitar estos problemas, pues elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre medias y desviación típica. Se define del siguiente modo:

$$\boxed{CV = \frac{S_X}{\overline{x}}}$$

Propiedades del coeficiente de variación

- Sólo se debe calcular para variables con todos los valores positivos. Todo índice de variabilidad es esencialmente no negativo. Las observaciones pueden ser positivas o nulas, pero su variabilidad debe ser siempre positiva. De ahí que sólo debemos trabajar con variables positivas, para la que tenemos con seguridad que $\overline{x} > 0$.
- No es invariante ante cambios de origen. Es decir, si a los resultados de una medida le sumamos una cantidad positiva, $b > 0$, para tener $Y = X + b$, entonces $CVY < CVX$.
- Es invariante a cambios de escala. Así por ejemplo el coeficiente de variación de una variable medida en metros es una cantidad adimensional que no cambia si la medición se realiza en centímetros.

Población versus muestra.

La teoría del muestreo tiene por objetivo, el estudio de las relaciones existentes entre la distribución de un carácter en dicha población y las distribuciones de dicho carácter en todas sus muestras.

Las ventajas de estudiar una población a partir de sus muestras son principalmente:

Coste reducido: Si los datos que buscamos los podemos obtener a partir de una pequeña parte del total de la población, los gastos de recogida y tratamiento de los datos serán menores. Por ejemplo, cuando se realizan encuestas previas a un referéndum, es más barato preguntar a 4,000 personas su intención de voto, que a 30,000,000;

Mayor rapidez: Estamos acostumbrados a ver cómo con los resultados del escrutinio de las primeras mesas electorales, se obtiene una aproximación bastante buena del resultado final de unas elecciones, muchas horas antes de que el recuento final de votos haya finalizado;

Más posibilidades: Para hacer cierto tipo de estudios, por ejemplo el de duración de cierto tipo de bombillas, no es posible en la práctica destruirlas todas para conocer su vida media, ya que no quedaría nada que vender. Es mejor destruir sólo una pequeña parte de ellas y sacar conclusiones sobre las demás.

De este modo se ve que al hacer estadística inferencial debemos enfrentarnos con dos problemas:

- Elección de la muestra (muestreo), que es a lo que nos dedicaremos en esta sección.
- Extrapolación de las conclusiones obtenidas sobre la muestra, al resto de la población (inferencia).

El tipo de muestreo más importante es el muestreo aleatorio, en el que todos los elementos de la población tienen la misma probabilidad de ser extraídos; Aunque dependiendo del problema y con el objetivo de reducir los costes o aumentar la precisión, otros tipos de muestreo pueden ser considerados como veremos más adelante: muestreo sistemático, estratificado y por conglomerados.

Muestreo aleatorio

Consideremos una población finita, de la que deseamos extraer una muestra. Cuando el proceso de extracción es tal que garantiza a cada uno de los elementos de la población la misma oportunidad de ser incluidos en dicha muestra, denominamos al proceso de selección muestreo aleatorio.

El muestreo aleatorio se puede plantear bajo dos puntos de vista:

- Sin reposición de los elementos;
- Con reposición.

Muestreo aleatorio sin reposición

Consideremos una población E formada por N elementos. Si observamos un elemento particular, $e \in E$, en un muestreo aleatorio sin reposición se da la siguiente circunstancia:

- La probabilidad de que e sea elegido en primer lugar es $1/N$;
- Si no ha sido elegido en primer lugar (lo que ocurre con una probabilidad de $(N-1)/N$), la probabilidad de que sea elegido en el segundo intento es de $1/(N-1)$.
- en el $(i + 1)$–ésimo intento, la población consta de $N - i$ elementos, con lo cual si e no ha sido seleccionado previamente, la probabilidad de que lo sea en este momento es de $1/(N-i)$.

Muestreo aleatorio con reposición

Sobre una población E de tamaño N podemos realizar extracciones de n elementos, pero de modo que cada vez el elemento extraído es repuesto al total de la población. De esta forma un elemento puede ser extraído varias veces.

El muestreo aleatorio con reposición es también denominado muestreo aleatorio simple, y se caracteriza porque cada elemento de la población tiene la misma probabilidad de ser elegido, y las observaciones se realizan con reemplazamiento. De este modo, cada observación es realizada sobre la misma población (que no disminuye con las extracciones sucesivas).

Muestreo aleatorio estratificado

Un muestreo aleatorio estratificado es aquel en el que se divide la población de N individuos, en k subpoblaciones o estratos, atendiendo a criterios que puedan ser importantes en el estudio, de tamaños respectivos $N_1, \ldots, N_k$,

$$N = N_1 + N_2 + \cdots + N_k$$

y realizando en cada una de estas subpoblaciones muestreos aleatorios simples de tamaño n_i $i = 1, \ldots, k$.

A continuación nos planteamos el problema de cuantos elementos de muestra se han de elegir de cada uno de los estratos. Para ello tenemos fundamentalmente dos técnicas: la asignación proporcional y la asignación óptima.

Asignación proporcional

Sea n el número de individuos de la población total que forman parte de alguna muestra:

$$n = n_1 + n_2 + \cdots + n_k$$

Cuando la asignación es **proporcional** el tamaño de la muestra de cada estrato es proporcional al tamaño del estrato correspondiente con respecto a la población total:

$$n_i = n \cdot \frac{N_i}{N}$$

Asignación óptima

Cuando se realiza un muestreo estratificado, los tamaños muestrales en cada uno de los estratos, n_i, los elige quien hace el muestreo, y para ello puede basarse en alguno de los siguientes criterios:

- Elegir los n_i de tal modo que se minimice la varianza del estimador, para un coste especificado, o bien,
- habiendo fijado la varianza que podemos admitir para el estimador, minimizar el coste en la obtención de las muestras.

Así en un estrato dado, se tiende a tomar una muestra más grande cuando:

- El estrato es más grande;
- El estrato posee mayor variabilidad interna (varianza);
- El muestreo es más barato en ese estrato.

Muestreo sistemático

Cuando los elementos de la población están ordenados en fichas o en una lista, una manera de muestrear consiste en

- Sea k =[N/n];
- Elegir aleatoriamente un número m, entre 1 y k;
- Tomar como muestra los elementos de la lista:

$$\left\{e_m,\ e_{m+k},\ e_{m+2k},\ \ldots,e_{m+(n-1)k}\right\}$$

Esto es lo que se denomina muestreo sistemático. Cuando el criterio de ordenación de los elementos en la lista es tal que los elementos más parecidos tienden a estar más cercanos, el muestreo sistemático suele ser más preciso que el aleatorio simple, ya que recorre la población de un modo más uniforme. Por otro lado, es a menudo más fácil no cometer errores con un muestreo sistemático que con este último.

El método tal como se ha definido anteriormente es sesgado si N/n no es entero, ya que los últimos elementos de la lista nunca pueden ser escogidos. Un modo de evitar este problema consiste en considerar la lista como si fuese circular (el elemento N + 1 coincide con el primero) y:

- Sea k el entero más cercano a N/n;
- Se selecciona un número al azar m, entre 1 y N;
- Se toma como muestra los elementos de la lista que consisten en ir saltando de k elementos en k, a partir de m, teniendo en cuenta que la lista es circular.

Se puede comprobar que con este método todos los elementos de la lista tienen la misma probabilidad de selección.

Muestreo por conglomerados

Si intentamos hacer un estudio sobre los habitantes de una ciudad, el muestreo aleatorio simple puede resultar muy costoso, ya que estudiar una muestra de tamaño n implica enviar a los encuestadores a n puntos distintos de la misma, de modo que en cada uno de ellos sólo se realiza una entrevista.

En esta situación es más económico realizar el denominado muestreo por conglomerados, que consiste en elegir aleatoriamente ciertos barrios dentro de la ciudad, para después elegir calles y edificios. Una vez elegido el edificio, se entrevista a todos los vecinos.

Distribución normal o gaussiana

La distribución gaussiana, recibe también el nombre de distribución normal, ya que una gran mayoría de las v.a continuas de la naturaleza siguen esta distribución. Se dice que una v.a. X sigue una distribución normal de parámetros μ y σ^2, lo que representamos del modo $X \to N\left(\mu, \sigma^2\right)$ si su función de densidad es:

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad \forall\, x \in I\!R}$$

Observación

Estos dos parámetros μ y σ^2 coinciden además con la media (esperanza) y la varianza respectivamente de la distribución como se demostrará más adelante:

$$\begin{aligned} \mathbf{E}[X] &= \mu \\ \mathbf{Var}[X] &= \sigma^2 \end{aligned}$$

La forma de la función de densidad es la llamada campana de Gauss.

Para el lector es un ejercicio interesante comprobar que ésta alcanza un único máximo (moda) en μ, que es simétrica con respecto al mismo, y por tanto P[X $\leq$ μ] = P[X $\geq$ μ] = 1/2, con lo cual en μ coinciden la media, la mediana y la moda, y por último, calcular sus puntos de inflexión.

El soporte de la distribución es todo *IR*, de modo que la mayor parte de la masa de probabilidad (área comprendida entre la curva y el eje de abscisas) se encuentra concentrado alrededor de la media, y las ramas de la curva se extienden asintóticamente a los ejes, de modo que cualquier valor "muy alejado" de la media es posible (aunque poco probable).

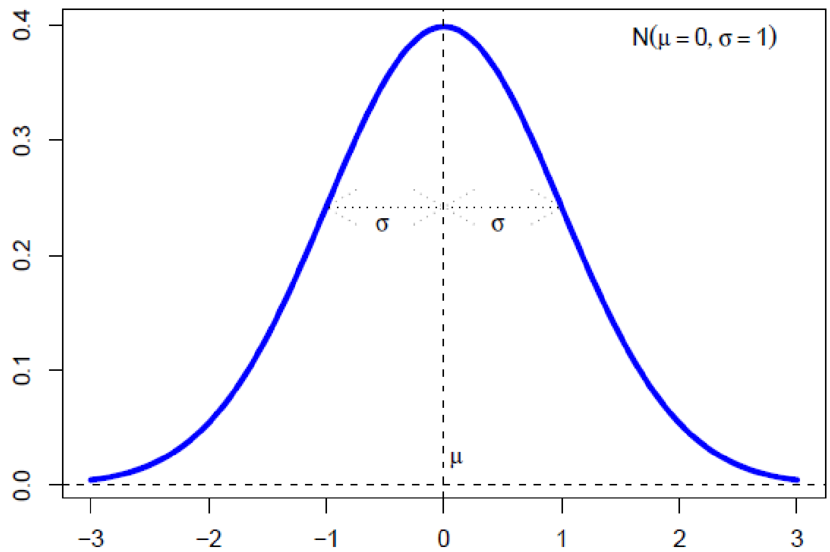


Figura 5. Campana de Gauss o función de densidad de una v. a. de distribución normal. El parámetro μ indica el centro y σ la dispersión. La distancia del centro a los puntos de inflexión es precisamente σ.

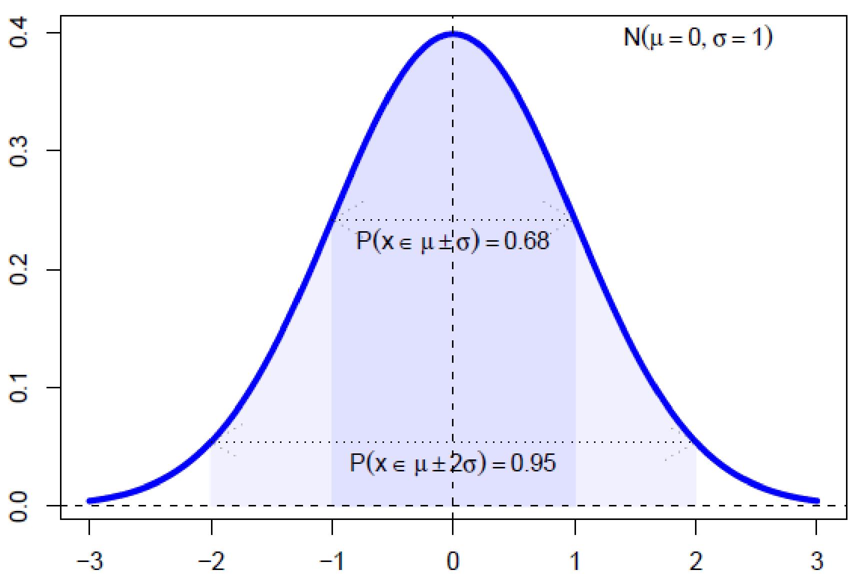


Figura 6. A una distancia que no supera en una desviación de la media tenemos una probabilidad de 68%. A dos desviaciones tenemos el 95%.

La forma de la campana de Gauss depende de los parámetros μ y σ:

- μ indica la posición de la campana (parámetro de centralización);

- σ^2 (o equivalentemente, σ) será el parámetro de dispersión. Cuanto menor sea, mayor cantidad de masa de probabilidad habrá concentrada alrededor de la media (grafo de f muy apuntado cerca de μ) y cuanto mayor sea "más aplastado" será.

Aproximación a la normal de la ley binomial

Se demuestra que una v.a. discreta con distribución binomial, X→B(n, p) se puede aproximar mediante una distribución normal si n es suficientemente grande y p no está ni muy próximo a 0 ni a 1. Como el valor esperado y la varianza de X son respectivamente n p y n p q, la aproximación consiste en decir que $X \to N(np, npq)$. El convenio que se suele utilizar para poder realizar esta aproximación es:

$$X \rightsquigarrow \mathbf{B}(n,p) \quad \text{donde} \quad \begin{cases} n > 30 \\ n\,p > 4 \\ n\,q > 4 \end{cases} \quad \Longrightarrow \quad X \overset{\approx}{\rightsquigarrow} \mathbf{N}(n\,p, n\,p\,q)$$

aunque en realidad esta no da resultados muy precisos a menos que realmente n sea un valor muy grande o p ≈ q≈ 1/2. Como ilustración obsérvense las figuras 7 y 8.

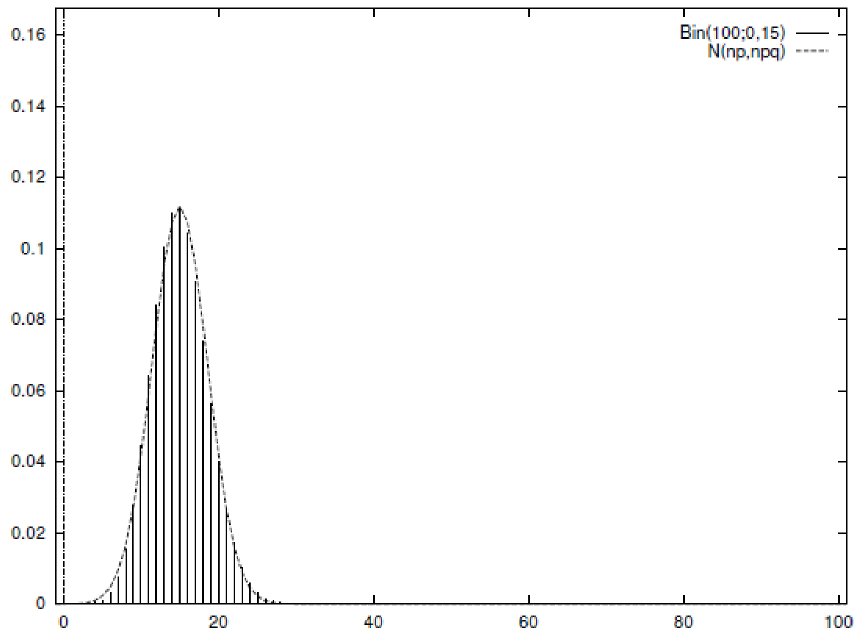


Figura 7. Comparación entre la función de densidad de una v.a. continua con distribución N(n p, n p q) y el diagrama de barras de una v.a. discreta de distribución B(n, p) para casos en que la aproximación normal de la binomial es válida. Es peor esta aproximación cuando p está próximo a los bordes del intervalo [0, 1].

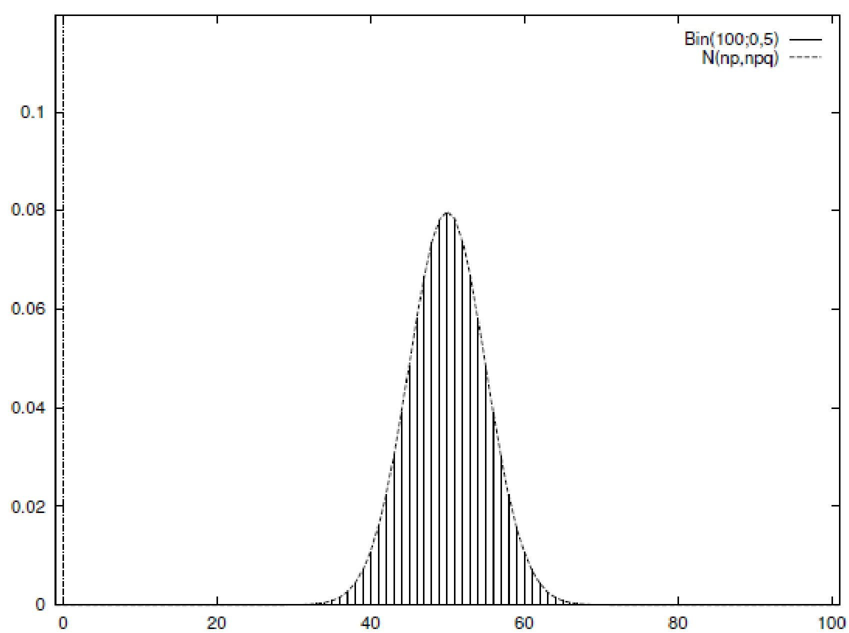


Figura 8. La misma comparación que en la figura anterior, pero realizada con parámetros con los que damos la aproximación normal de la binomial es mejor.

3 Control estadístico de procesos

Este capítulo esta destinado a presentar técnicas estadísticas utilizadas para realizar un seguimiento en tiempo real, de la calidad de un proceso productivo.

Estas técnicas son conocidas como Control Estadístico de Procesos, o control de proceso en línea. El control de calidad en línea o control estadístico de procesos produce mayores beneficios:

• Al vigilar continuamente el funcionamiento del proceso reduce la fabricación de productos defectuosos.

• El control de producto terminado tiene poca capacidad de generar información útil para ajustar y mejorar el proceso. Mediante el control estadístico de proceso podemos saber en qué momento se empezó a desajustar el proceso, por lo que se puede analizar las circunstancias que llevaron al desajuste para que no vuelvan a ocurrir.

La herramienta más utilizada dentro del control estadístico de procesos son los gráficos de control.

Los gráficos de control los podemos clasificar:

1. Gráficos de control por variables.
2. Gráficos de control por atributos.
3. Gráficos de control por el número de defectos.

3.1 Causas de variabilidad de un proceso

Uno de los axiomas fundamentales de fabricación es que nunca dos objetos pueden fabricarse exactamente iguales, producimos con cierta variabilidad.

Cada proceso está sujeto a la influencia de muchos factores (muchos de los cuales no podemos controlar) y por lo tanto, hay variabilidad en el resultado final del proceso.

Ejemplo 1:

Medición de la longitud de un cilindro de un modelo concreto, llamémosle C, producido por una máquina determinada, digamos M. Cada vez que midamos un cilindro del mismo modelo y procedente de esa máquina estaremos repitiendo este experimento. En el experimento definido en este ejemplo son dos los factores cuya influencia mantenemos constante: modelo de cilindro y máquina. Hay, por contra, infinidad de factores que en dicho experimento no controlamos. Por ejemplo, la presión con que dicha máquina corta el cilindro no será siempre exactamente igual. La estructura microscópica de la aleación metálica no será la misma. Las vibraciones que se produzcan durante el proceso serán también irrepetibles. La persona que se encargue de la manipulación de dicho proceso puede cambiar. El desgaste de las piezas de la máquina que intervienen en el corte será un factor que tendrá una influencia en la longitud final. En este ejemplo, aunque los cilindros sean aparentemente idénticos las repeticiones del experimento proporcionarán, con mucha probabilidad, longitudes diferentes. Las diferencias observadas en cada repetición son debidas al resto de los factores que no se controlan.

Ejemplo 2:

Resultado de un juego de azar, como el resultado del lanzamiento de un dado, lanzamiento de una moneda, giro de una ruleta o la extracción de una carta. Cada vez que repitamos el experimento puede salir un resultado diferente, aunque el factor que se controla se mantenga constante: mismo dado, moneda, mesa de juego, baraja e incluso misma persona que participa en la acción. Los factores que no se controlan son, en realidad, muy pocos: el movimiento de la mano de la persona que ejecuta el experimento; pero el movimiento es tan complejo que es prácticamente imposible repetirlo ni hacerlo de manera premeditada para conseguir un resultado prefijado. Un esquema útil para resumir los diferentes factores que afectan a un proceso es el de las 6´M.

$$\left\{\begin{array}{c} \text{Materia prima} \\ \text{Metodos de producción} \\ \text{Mano de obra} \\ \text{Máquinas} \\ \text{Medio Ambiente} \\ \text{Medición} \end{array}\right. \Rightarrow \text{Producto Final}$$

En la práctica siempre habrá factores que no podremos controlar, por lo que el producto final tendrá una calidad variable. Por tanto, cuando se habla de controlar la calidad nos estamos refiriendo a controlar la variabilidad de esa calidad. Aunque no se puede dar una relación exhaustiva de las causas que provocan la variabilidad, es interesante hacer una clasificación entre causas asignables a factores concretos (y controlables) y causas no asignables. No obstante, la frontera entre ambas no estará, en general, perfectamente definida.

1. Causas no asignable o naturales: causas cuyos efectos individuales son pequeños y difíciles de eliminar, producen una variabilidad estable y predecible, por ejemplo:

- Proceso: la imprecisión de las herramientas, la vibración de las máquinas, fluctuaciones hidráulicas y eléctricas,...
- Materiales: cambios de materia prima, espesor, resistencia,...
- Ambiente: diferencias en las condiciones atmosféricas, limpieza, iluminación,...
- Personas: diferencia en el estado físico, experiencia, motivación,...
- ...

2. Causas asignables: causas que actúan en determinados momentos produciendo gran variabilidad, sus efectos son predecibles y definidos. Sus efectos son eliminados cuando se elimina la causa, por ejemplo:

- desajuste de la máquina
- defecto de la materia prima
- operario no cualificado
- ...

El control estadístico de procesos consiste en observar periódicamente el proceso y descubrir mediante gráficos cuando en el proceso están actuando causas asignables, con objeto de descubrirlas y eliminarlas del proceso.

Cuando en el proceso sólo actúan causas no asignables se dice que el proceso está bajo control.(Cuando en el proceso actúan causas asignable se dice que el proceso está fuera de control).

Gráficos de Control por variables

Suponemos que la variable de calidad del proceso es una variable continua.

Supongamos que cuando el proceso está bajo control la variable de calidad, X, se distribuye normalmente

$$X \; ; \; N(\mu, \sigma)$$

Intervalo de tolerancia e índice de capacidad

El intervalo de tolerancia es un conjunto de valores de X que se consideran admisibles o aceptables.

$$\text{intervalo de tolerancia} = (LT_1, LT_2)$$

Suele ser definido por el cliente o en ocasiones por el diseñador del producto (suele representase por un valor objetivo θ y un error aceptable L; θ ± L.)

Ejemplo:

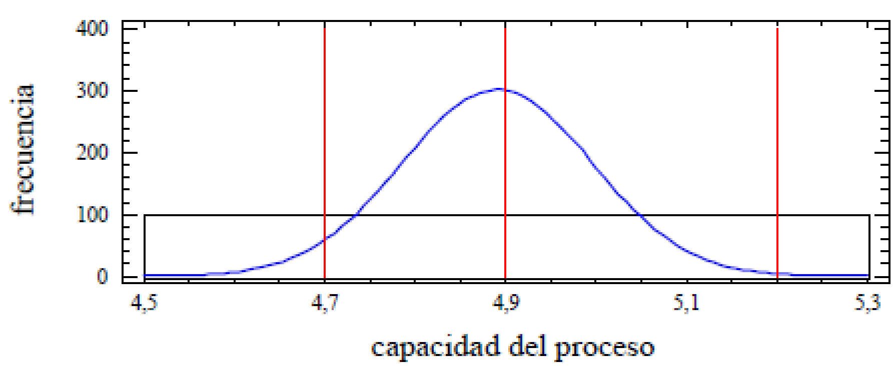
Supongamos que producimos piezas metálicas y una variable de calidad es el diámetro de las piezas. Cuando el proceso está bajo control el diámetro de las piezas se distribuye

$$X \; ; \; N(4.9, 0.1)$$

El cliente establece un intervalo de tolerancia entre 4.7 y 5.2

¿Qué proporción de defectuosos producimos cuando el proceso está bajo control?

P(defectuosos)=1-P(4.7≤ X ≤ 5.2) = 1 − P(−2 ≤ Z ≤ 3) = 0.024



La proporción de defectuosos en estado de control indica la adecuación para la fabricación de productos con especificaciones o tolerancias definidas.

Definición índice de capacidad

Cuando la media de la producción está centrada en el centro del intervalo de tolerancia ($\mu = \theta$), el índice de capacidad indica la adecuación de la fabricación de productos.

$$C_p = \frac{LT_2 - LT_1}{6\sigma}$$

observación:

• Si Cp < 1 Se producen más del 3 por mil de artículos defectuosos.

• Si Cp > 1 Se producen menos del 3 por mil de artículos defectuosos.

Ver anexo (Indices de capacidad)

Presencia de causas asignables

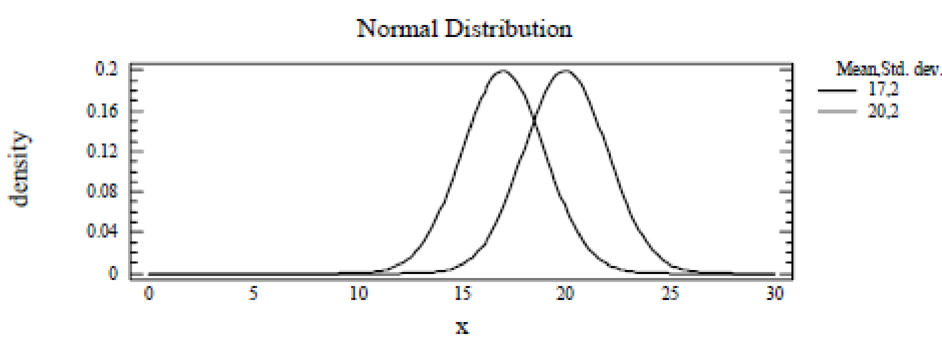
La presencia de una causa asignable produce un cambio en la media o en la varianza del proceso, o en ambas, lo que supone un aumento de la proporción de artículos defectuosos.

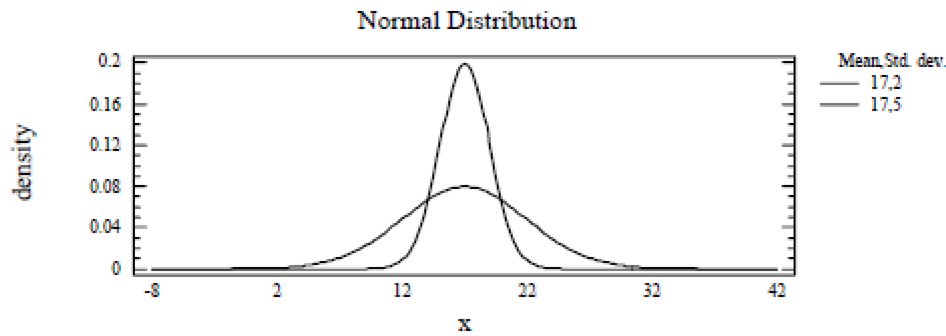
Ejemplo:

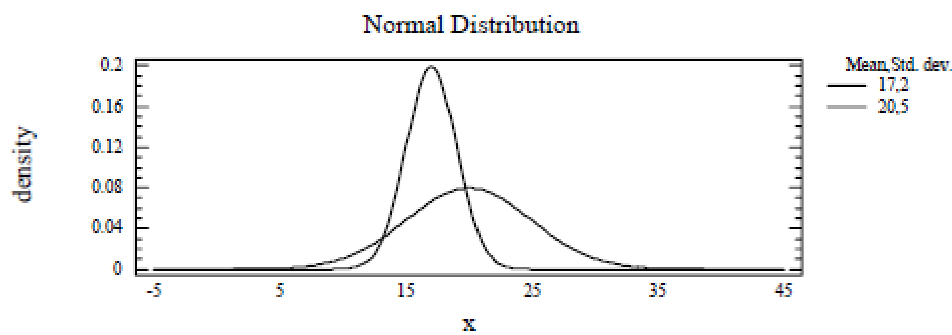
Supongamos que la variable de calidad cuando el proceso está bajo control se distribuye ;X ; N(17, 2)

El cliente especifica un intervalo de tolerancia (10,24)

En los gráficos siguientes mostramos la proporción de defectuosos cuando en el proceso varia la media, o la varianza, o ambas.







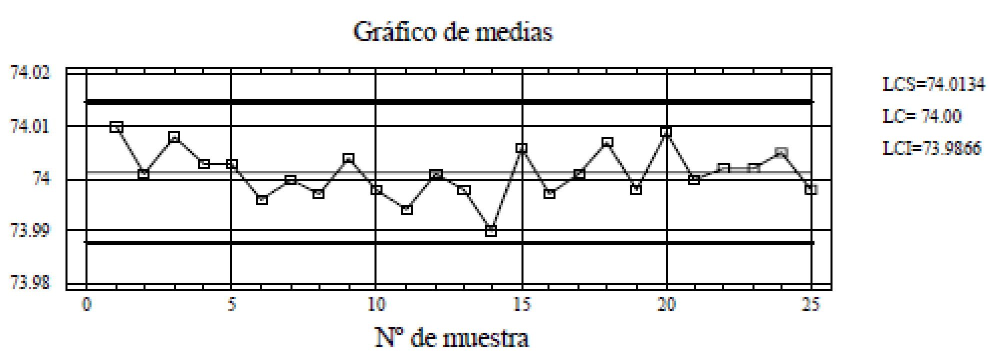
Los gráficos de control por variables permiten detectar la presencia de causas asignables en el proceso.

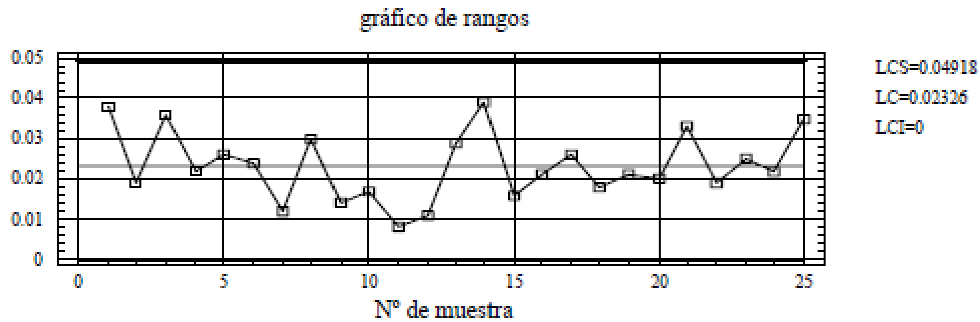
Gráficos de Medias y Rangos (o desviaciones)

El objetivo de los gráficos de control por variables es controlar el proceso para que no se vaya de control, es decir, controlar el proceso para detectar la presencia de causas asignables. Para ello se toman muestras del proceso periódicamente y se representa, en dos diferentes gráficos:

• la media muestral .

• el rango muestral (o la desviación típica muestral).





Si la media muestral o el rango muestral (o la desviación) se sale de los gráficos (cuando un punto se sale de los límites establecidos por las líneas gruesas) entonces tenemos una evidencia estadística de que la media o la variabilidad del proceso han cambiado, es decir, el proceso está fuera de control.

En los **gráficos de medias** detectamos cambios en la media o variabilidad del proceso.

En los **gráficos de rangos** (o de desviaciones) detectamos cambios en la variabilidad del proceso.

Gráficos de $\overline{X}$ y R (ó S) (μ, σ conocidas)

Gráfico de medias.

Supongamos que cuando el proceso está bajo control

$$X \;;\; N(\mu, \sigma)$$

Entonces

$$\overline{X} \;;\; N(\mu, \frac{\sigma}{\sqrt{n}})$$

Así

$$p(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}}) = 0.997$$

Si observamos una media y no está dentro de ese intervalo puede deberse a:

• El proceso esté bajo control y hallamos extraído una muestra cuya media esté fuera de ese intervalo.

• El proceso esté fuera de control.

La primera razón es poco probable, sólo el 3 por mil de las muestras tienen media muestral fuera de ese intervalo, luego suponemos que el proceso está fuera de control, es decir, ha cambiado la media o la varianza, o ambas.

Si para una muestra su media no pertenece al intervalo ($\mu - 3\frac{\sigma}{\sqrt{n}}$, $\mu + 3\frac{\sigma}{\sqrt{n}}$) tenemos evidencia estadística de que el proceso está fuera de control.

$$\text{Gráfico de control de medias}$$
$$LCS \text{ (limite de control superior)} = \mu + 3\frac{\sigma}{\sqrt{n}}$$
$$LC \text{ (linea central)} = \mu$$
$$LCI \text{ (limite de control inferior)} = \mu - 3\frac{\sigma}{\sqrt{n}}$$

Gráfico de Rangos

Supongamos que

$$X \;;\; N(\mu, \sigma)$$

R = rango muestral = $X_{\max} - X_{\min}$ es un estimador sesgado de la desviación típica

$$\begin{aligned} E(R) &= d_2\sigma \\ Var(R) &= d_3^2\sigma^2 \end{aligned}$$

donde d_2 y d_3 son constantes que dependen del tamaño de la muestra y su valor está tabulado (Anexo 2).

Y su distribución es aproximadamente normal, luego si el rango muestral está fuera del intervalo

$$(d_2\sigma - 3d_3\sigma, d_2\sigma + 3d_3\sigma) = (D_1\sigma, D_2\sigma)$$

hay evidencia estadística de que la variancia del proceso ha variado.

$$\text{Gráfico de control de rangos}$$
$$LCS = D_2\sigma$$
$$LC = d_2\sigma$$
$$LCI = D_1\sigma$$

Gráfico de desviaciones típicas corregidas $\left(\hat{S} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \right)$

Supongamos que

$$X \;;\; N(\mu,\sigma)$$

$\widehat{S} =$ desviación típica corregida muestral es un estimador sesgado de la desviación típica

$$\begin{aligned} E(\widehat{S}) &= c_4\sigma \\ Var(\widehat{S}) &= \sigma^2(1-c_4^2) \end{aligned}$$

donde c_4 es una constante que depende del tamaño de la muestra y su valor está tabulado (Anexo 2).

Se puede mostrar que si la desviación típica muestral está fuera del intervalo

$$(c_4\sigma - 3\sqrt{1-c_4^2}, c_4\sigma + 3\sqrt{1-c_4^2}) = (B_5\sigma, B_6\sigma)$$

hay evidencia estadística de que la variancia del proceso ha variado.

$$\begin{gathered} \text{Gráfico de control de desviaciones} \\ LCS = B_6\sigma \\ LC = c_4\sigma \\ LCI = B_5\sigma \end{gathered}$$

Ejemplo:

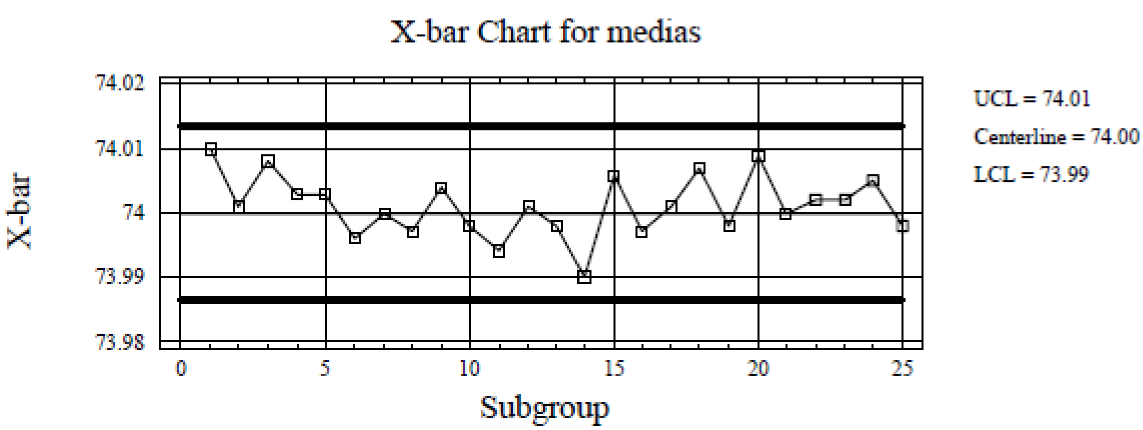
Se fabrican anillos de pistón para motor de automóvil mediante un proceso de forjado. Se desea controlar el diámetro de éstos por medio de gráficos de medias y rangos. Se toman muestras de tamaño 5 cada media hora, los datos aparecen en la siguiente tabla. Cuando el proceso está bajo control el diámetro de los anillos se distribuye

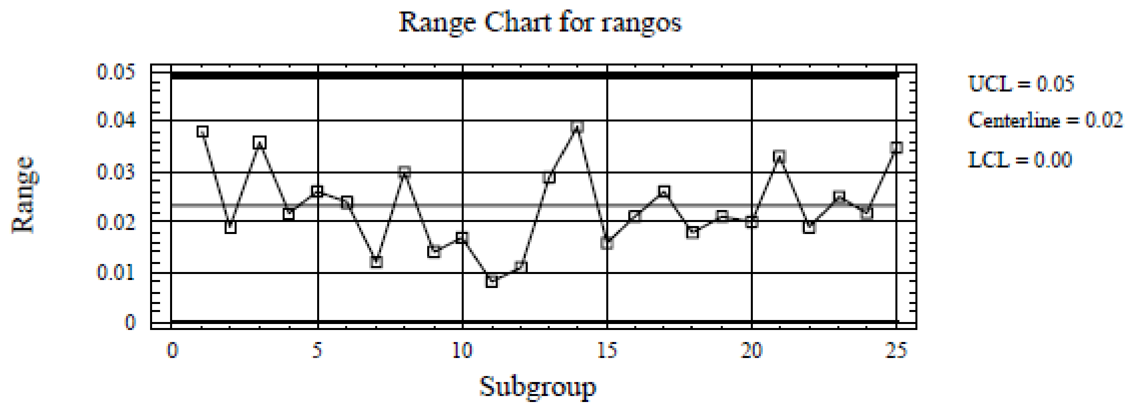
$$X \;;\; N(74, 0.01)$$

Nº de muestra	Media	Rango	Nº de muestra	Media	Rango
1	74.01	0.038	14	73.99	0.039
2	74.001	0.019	15	74.006	0.016
3	74.008	0.036	16	73.997	0.021
4	74.003	0.022	17	74.001	0.026
5	74.003	0.026	18	74.007	0.018
6	73.996	0.024	19	73.998	0.021
7	74	0.012	20	74.009	0.02
8	73.997	0.03	21	74	0.033
9	74.004	0.014	22	74.002	0.019
10	73.998	0.017	23	74.002	0.025
11	73.994	0.008	24	74.005	0.022
12	74.001	0.011	25	73.998	0.035
13	73.998	0.029			

Los gráficos de control de medias y rangos

grafico de medias	gráfico de rango
$LCS = 74 + 3 * \frac{0.01}{\sqrt{5}} = 74.0134$	$LCS = D_2\sigma = 4.918 * 0.01$
$LC = 74$	$LC = d_2\sigma = 2.326 * 0.01$
$LCI = 74 - 3 * \frac{0.01}{\sqrt{5}} = 73.9866$	$LCI = D_1\sigma = 0 * 0.01$





Como podemos observar ningún punto se sale de los gráficos por lo que suponemos que el proceso se encuentra bajo control.

Sensibilidad del gráfico $\overline{X}$

Hemos definido, anteriormente, el gráfico de control para las medias (supuesto conocido la media y la variancia del proceso en estado de control)

$$\begin{aligned} LCS &= \mu + 3\frac{\sigma}{\sqrt{n}} \\ LC &= \mu \\ LCI &= \mu - 3\frac{\sigma}{\sqrt{n}} \end{aligned}$$

Esta regla se denomina de las 3 sigmas y es común en todos los gráficos de control estándares.

Con el gráfico de medias así definido podemos cometer dos errores, representados en la siguiente tabla

	decisión	
	bajo control	fuera de control
bajo control	-	Error I
fuera de control	Error II	-

Error I: decimos que los parámetros del proceso han cambiado cuando en realidad no se han modificado, falsa alarma.

Error II: decimos que el proceso no ha cambiado cuando en realidad si se ha modificado.

Al establecer los límites 3 sigma podemos calcular la probabilidad de falsa alarma, es decir la probabilidad del error I:

$$\begin{aligned} p(error I) &= p(\text{decir que esta fuera de control / esta bajo control}) = \\ &= 1 - p(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \overline{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}} \ / \ X \ ; \ N(\mu, \sigma) = \\ &= 0.003 \end{aligned}$$

Si se quiere reducir la probabilidad de falsa alarma se puede aumentar el intervalo más allá de 3 sigma, dando un valor K>3, los límites se establecen en

$$\mu \pm K\frac{\sigma}{\sqrt{n}}$$

Lógicamente esto presenta un inconveniente, que es: si el proceso cambia y sale fuera de su estado de control, la probabilidad de que los puntos caigan dentro de los limites, y por tanto, equivocarse aceptando que el proceso está bajo control, también es mayor.

Tamaño muestral

Habitualmente el tamaño muestral está entre 4 y 10, siendo muy habitual en la industria tomar muestras de tamaño 4 o 5. Cuanto mayor sea el tamaño muestral más probabilidad de detectar cambios pequeños en el proceso.

Frecuencia de muestreo: consiste en determinar cada cuanto tiempo se toman las muestras. Depende fundamentalmente del coste de inspección, del coste de producir artículos defectuosos, de la probabilidad de cambio en el estado de control del proceso y, también, del índice de capacidad del proceso.

Gráficos de X y R (ó S) (μ, σ desconocidas)

Cuando el proceso lleva funcionando un periodo largo (años) es aceptable suponer conocidos los valores de μ, σ , en otras ocasiones esto no es así y los valores de los parámetros del proceso en situación de control deben estimarse con los datos disponibles. Vamos a suponer que se dispone de K muestras, cada una de tamaño n, que se han tomado todas ellas con el proceso bajo control y durante un periodo suficientemente largo en el cual han intervenido todas las fuentes de variabilidad natural.

N muestra	observaciones	media	rango	desviación
1	$x_{11}, \ldots, x_{1n}$	$\overline{x}_1$	R_1	$\widehat{S}_1$
2	$x_{21}, \ldots, x_{2n}$	$\overline{x}_2$	R_2	$\widehat{S}_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$x_{k1}, \ldots, x_{kn}$	$\overline{x}_k$	R_k	$\widehat{S}_k$

1 Estimación

$$\widehat{\mu} = \overline{\overline{x}} = \frac{\overline{x}_1 + \ldots + \overline{x}_k}{k}$$

$$\widehat{\sigma} = \frac{\overline{R}}{d_2} = \frac{R_1 + R_2 + \ldots + R_k}{k * d_2} \text{ (utilizando los rangos)}$$

$$\widehat{\sigma} = \frac{\widehat{S_T}}{c_4} = \frac{\widehat{S}_1 + \widehat{S}_2 + \ldots + \widehat{S}_k}{k * c_4} \text{ (utilizando los desviaciones corregidas)}$$

2 Gráficos de control, utilizando las estimaciones

Sustituyendo las estimaciones en las fórmulas de las líneas del gráfico de control para el caso de μ y σ conocidos obtenemos:

- Si utilizo rangos

gráfico de medias	gráfico de rangos
$LCS = \overline{\overline{x}} + 3\frac{\overline{R}}{d_2\sqrt{n}}$ $LC = \overline{\overline{x}}$ $LCI = \overline{\overline{x}} - 3\frac{\overline{R}}{d_2\sqrt{n}}$	$LCS = \overline{R} + 3d_3\frac{\overline{R}}{d_2} = \overline{R}(1 + 3\frac{d_3}{d_2}) = D_4\overline{R}$ $LC = \overline{R}$ $LCI = \overline{R} - 3d_3\frac{\overline{R}}{d_2} = \overline{R}(1 - 3\frac{d_3}{d_2}) = D_3\overline{R}$

- Si utilizo desviaciones

gráfico de medias	gráfico de desviaciones
$LCS = \overline{\overline{x}} + 3\frac{\widehat{S_T}}{c_4\sqrt{n}}$ $LC = \overline{\overline{x}}$ $LCI = \overline{\overline{x}} - 3\frac{\widehat{S_T}}{c_4\sqrt{n}}$	$LCS = \widehat{S_T} + 3 * \frac{\widehat{S_T}}{c_4}\sqrt{1 - c_4^2} = (1 + 3\frac{\sqrt{1-c_4^2}}{c_4})\widehat{S_T} = B_4\widehat{S_T}$ $LC = \widehat{S_T}$ $LCI = \widehat{S_T} - 3 * \frac{\widehat{S_T}}{c_4}\sqrt{1 - c_4^2} = (1 - 3\frac{\sqrt{1-c_4^2}}{c_4})\widehat{S_T} = B_3\widehat{S_T}$

ESTUDIO INICIAL (determinación de la capacidad del proceso)

En el apartado anterior hemos supuesto que las muestras han sido tomadas con el proceso en estado de control. A continuación mostramos un procedimiento iterativo que se utiliza para garantizar (al menos de forma aproximada) que las muestras empleadas para hacer la estimación de parámetros corresponden al proceso en estado de control.

Pasos:

1. Se toman k muestras (k entre 20 o 30) de tamaño n (n entre 4 o 5).
2. Se estiman los parámetros

$$\widehat{\mu} = \overline{\overline{x}} = \frac{\overline{x}_1 + \ldots + \overline{x}_k}{k}$$

$$\widehat{\sigma} = \frac{\overline{R}}{d_2} = \frac{R_1 + R_2 + \ldots + R_k}{k * d_2} \quad \text{(utilizando los rangos)}$$

$$\widehat{\sigma} = \frac{\widehat{S_T}}{c_4} = \frac{\widehat{S}_1 + \widehat{S}_2 + \ldots + \widehat{S}_k}{k * c_4} \quad \text{(utilizando los desviaciones corregidas)}$$

3. Se construyen los gráficos de control.
4. Representar los rango (o desviaciones corregidas) observados en el gráficos de rangos (o desviaciones corregidas) construido en el paso 3.

Si algún punto se encuentra fuera de los limites de control, eliminar la muestra del conjunto. En este caso volver al paso 2 para recalcular las estimaciones de los parámetros con el conjunto modificado de muestras. Si todos los puntos están dentro del gráfico de control ir al paso siguiente

5. Representar las medias en el gráfico de medias construido en el paso 3.

Si algún punto se encuentra fuera de los limites de control, eliminar la muestra del conjunto. En este caso volver al paso 2 para recalcular las estimaciones de los parámetros con el conjunto modificado de muestras. Si todos los puntos están dentro del gráfico de control ir al paso siguiente.

6. Con las observaciones restantes, estimar los parámetros y contrastar la normalidad de las observaciones.

Ejemplo:

Supongamos que fabricamos anillos de pistón. Queremos determinar la capacidad de nuestro proceso. Se toman 40 muestras de tamaño 5 obteniendo los datos presentados en la siguiente tabla.

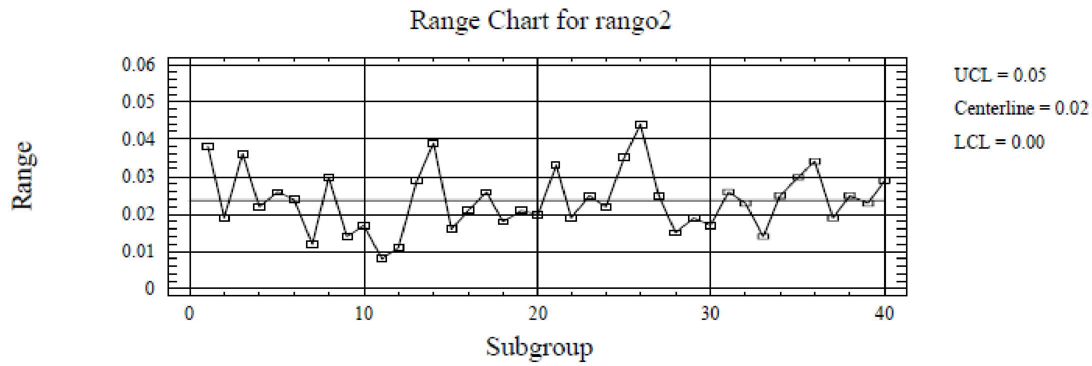
Nº muestra	Media	Rango	Nº muestra	Media	Rango
1	74.01	0.038	21	74	0.033
2	74.001	0.019	22	74.002	0.019
3	74.008	0.036	23	74.002	0.025
4	74.003	0.022	24	74.005	0.022
5	74.003	0.026	25	73.998	0.035
6	73.996	0.024	26	74.009	0.044
7	74	0.012	27	74.002	0.025
8	73.997	0.03	28	73.992	0.015
9	74.004	0.014	29	74.004	0.019
10	73.998	0.017	30	73.997	0.017
11	73.994	0.008	31	74.007	0.026
12	74.001	0.011	32	74.006	0.023
13	73.998	0.029	33	73.998	0.014
14	73.99	0.039	34	74.011	0.025
15	74.006	0.016	35	74.013	0.03
16	73.997	0.021	36	74.004	0.034
17	74.001	0.026	37	74.017	0.019
18	74.007	0.018	38	74.02	0.025
19	73.998	0.021	39	74.023	0.023
20	74.009	0.02	40	74.015	0.029

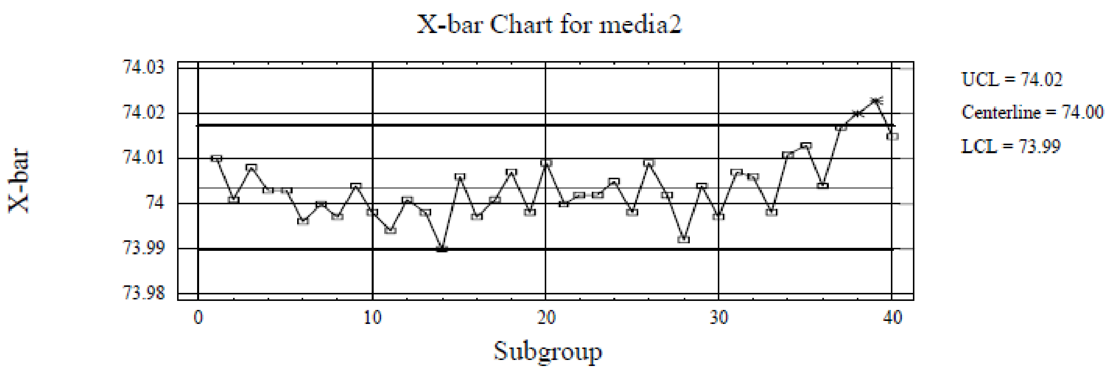
Pasos para determinar la capacidad del proceso:

1 Estimación de parámetros

$$\begin{aligned} \widehat{\mu} &= \overline{\overline{x}} = 74.0037 \\ \widehat{\sigma} &= \frac{\overline{R}}{d_2} = 0.0101999 \end{aligned}$$

2 Gráficos de control



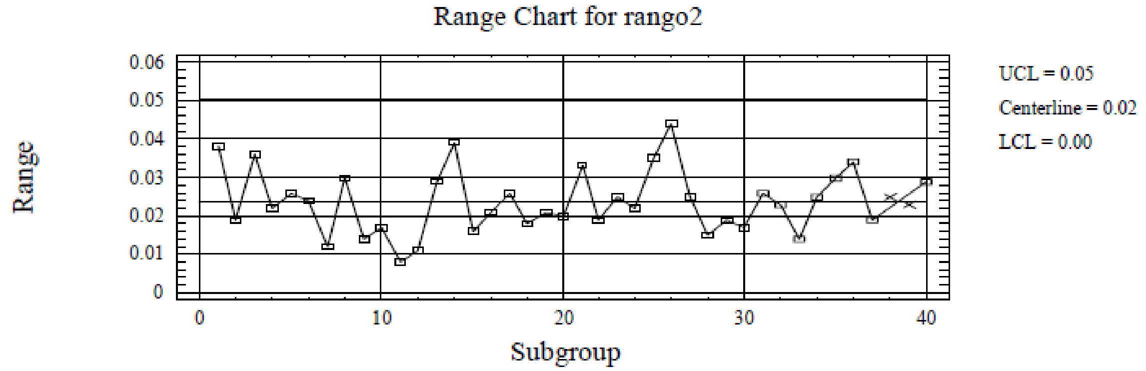


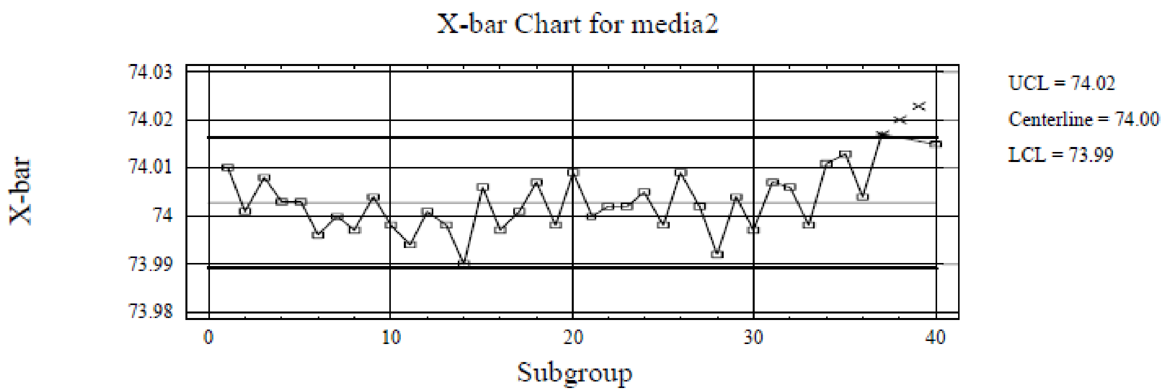
Observamos que hay dos puntos fuera del gráfico de medias, correspondientes a las muestras 38 y 39. Eliminamos estas muestras y volvemos a estimar los parámetros

$$\widehat{\mu} = \overline{\overline{x}} = 74.0027$$

$$\widehat{\sigma} = \frac{\overline{R}}{d_2} = 0.0101937$$

volvemos a representar los gráficos utilizando estas estimaciones

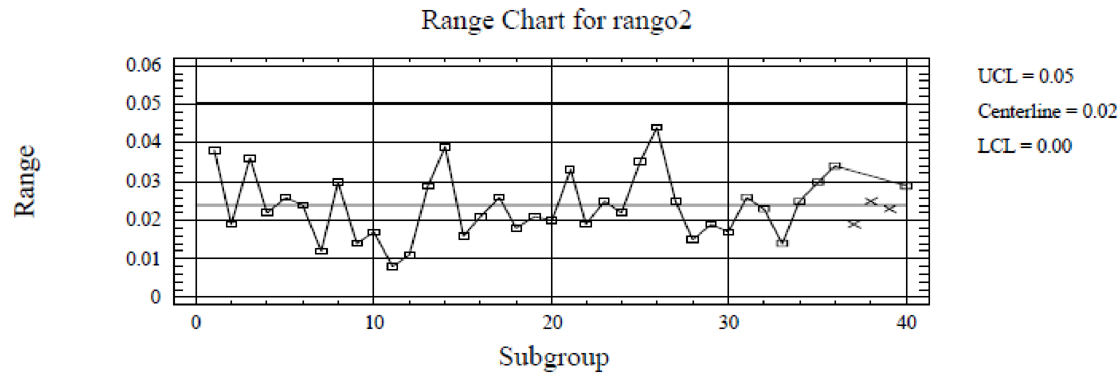


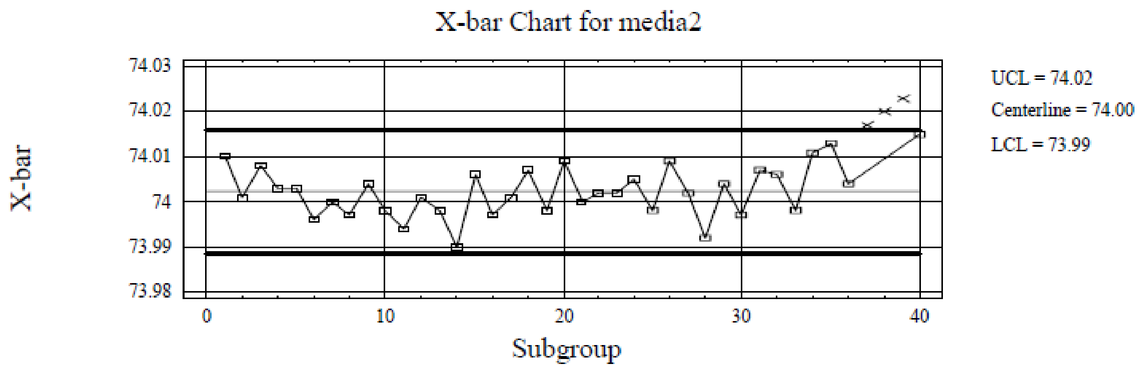


Observando que hay un punto fuera del gráfico de medias, correspondiente a la muestra 37, por lo tanto, eliminamos esta muestra y volvemos a estimar los parámetros

$$\hat{\mu} = \overline{\overline{x}} = 74.0023$$
$$\hat{\sigma} = \frac{\overline{R}}{d_2} = 0.0102484$$

Volvemos a representar los gráficos de control utilizando estas nuevas estimaciones, obteniendo:

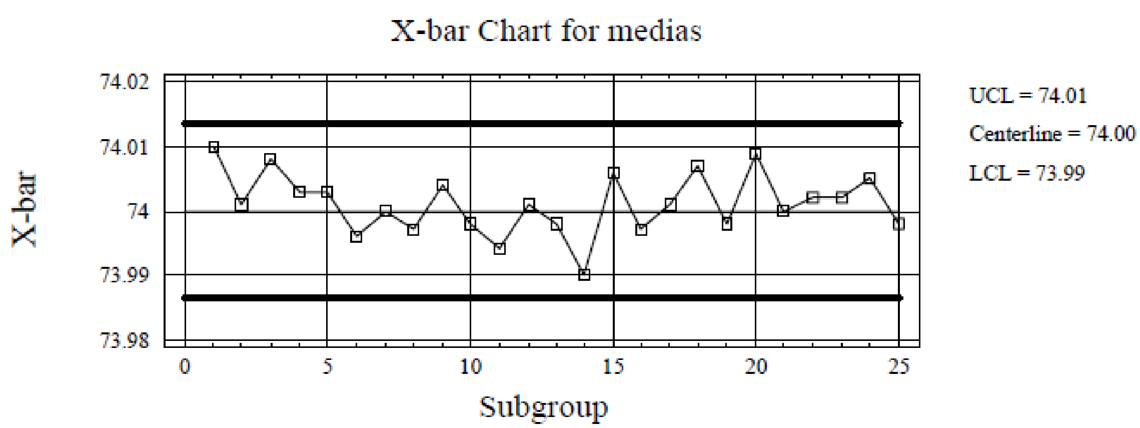




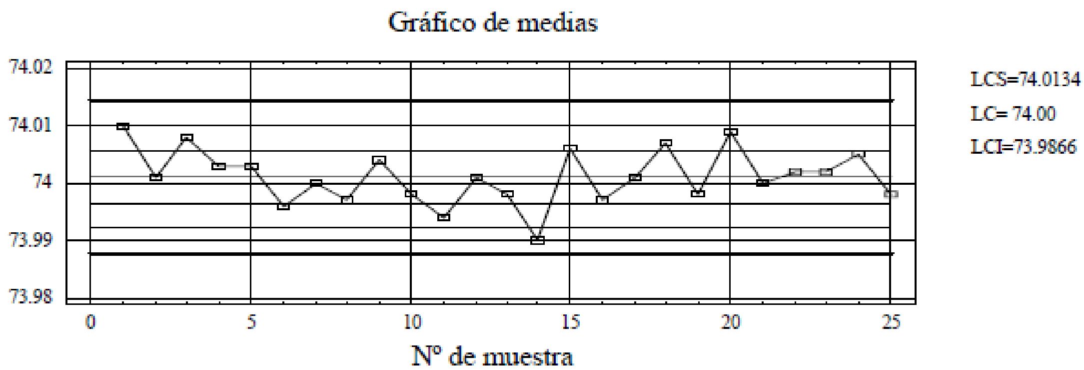
Ahora no hay algún punto fuera del gráfico entonces las estimaciones de los parámetros del proceso en situación de control son las obtenidas en último lugar.

Interpretación de los gráficos de control.

Lo que se debe esperar en un gráfico de control es una distribución de puntos aleatorios alrededor de la línea central, patrón no aleatorio o natural, por ejemplo:



El objetivo principal del gráfico de control es determinar cuando un proceso está fuera de control de manera que es necesaria la acción de ajuste. La regla principal es tomar como indicación de falta de control cuando un punto se sale fuera de los limites. A esta regla se le suele añadir alguna que aumentan la sensibilidad del gráfico si en la identificación de señales de la evolución del proceso no sigue un patrón natural. Para definir estas reglas entre los limites de control se añaden líneas de aviso equiespaciadas, a distancia de una desviación típica.



Regla 1: el proceso está fuera de control si un punto se sale de los límites de control.

Regla 2: (2 de 3): un proceso se considera fuera de control si de tres puntos consecutivos, dos están más allá de la línea de aviso de 2σ.

Regla 3: (4 de 5): un proceso se considera fuera de control si de cinco puntos consecutivos, cuatro están a más allá de la línea de aviso de σ.

Regla 4: (racha) el proceso está fuera de control si ocho o más puntos consecutivos están en una mitad del grafico.

Regla 5: (tendencia): Ocho puntos consecutivos o más en ascenso (o descenso)

GRÁFICOS DE MEDIAS MOVILES

El gráfico de control es relativamente insensible a pequeños cambios en la media del proceso. Se ha sugerido diversas modificaciones y criterios suplementarios con el fin de mejorar su capacidad para detectar cambios pequeños.

Supongamos que se obtienen muestras de tamaño n, y sean $(\bar{x}_1, \ldots\ldots, \bar{x}_t, \ldots)$ sus medias muestréales correspondientes. La media móvil de alcance w en el momento t se define como

$$M_t = \frac{\overline{x}_t + \overline{x}_{t-1} + ... + \overline{x}_{t-w+1}}{w}$$

(es decir, en el momento t se suprime la media muestral más antigua y se añade la más reciente. Como utilizamos más observaciones es más fácil detectar pequeños cambios en el sistema)

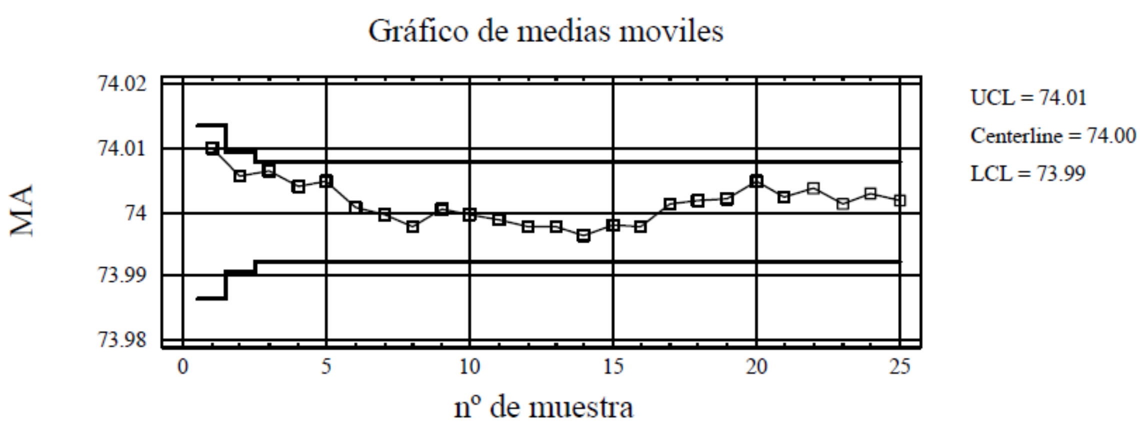
Como

$$\begin{aligned} E(M_t) &= \mu \\ Var(M_t) &= \frac{1}{w}\frac{\sigma^2}{n} \end{aligned}$$

el gráfico de control de medias móviles tiene límites de control

$$\begin{aligned} LCS &= \mu + 3\frac{\sigma}{\sqrt{wn}} \\ LC &= \mu \\ LCI &= \mu - 3\frac{\sigma}{\sqrt{wn}} \end{aligned}$$

representando el límite de control para el primer ejemplo del diámetro de los pistones, utilizando amplitud 3.



Gráficos de control por atributos

Muchas características de calidad no se pueden representar en forma conveniente por números. En tales ocasiones, cada artículo o producto es inspeccionado y suele clasificarse como conforme o disconforme con las especificaciones para tal característica de calidad (defectuoso o no defectuoso).

Si el proceso de producción está en control estadístico. Sea X una variable aleatoria que indica si un artículo producido es o no defectuoso.

$$X = \begin{cases} 1 \text{ si el artículo es defectuoso} \\ 0 \text{ si el artículo no es defectuoso} \end{cases}$$

Si el proceso de producción está bajo control estadístico (no actúan causas asignables) entonces

X: Bernoulli (p)

(p es la proporción de artículos defectuosos cuando el proceso está bajo control).

Defino, la capacidad del proceso como 1-p.

Si en el proceso de producción intervienen n causas asignables la proporción de elementos defectuosos aumenta. El objetivo de los gráficos de control por atributos es detectar rápidamente la presencia de causas asignables en el proceso.

Se toman muestras en periodos regulares de tiempo y las voy representando en el gráfico de control por atributos, este gráfico puede ser de dos formas:

1. Gráfico de número de defectos en la muestra de tamaño n (si represento el número de defectos).

2. Gráfico de proporción de defectos en la muestra (si represento la proporción de defectos en la muestra).

Gráfico de número de defectuosos (p conocida)

Sea r el número de defectuosos en la muestra r ; b(n, p) cuando el proceso está bajo control $\Leftrightarrow$ r; N(np, $\sqrt{np(1-p)}$ cuando el proceso está bajo control y n es grande.

Luego p(np − 3$\sqrt{np(1-p)}$ ≤ r ≤ np + 3$\sqrt{np(1-p)}$) = 0.997 si el proceso está bajo control, así

gráfico de control por nº ºde defectuosos

$$LCS = np + 3\sqrt{np(1-p)}$$
$$LC = np$$
$$LCI = np - 3\sqrt{np(1-p)}$$

Ejemplo

Se envasa jugo de naranja en botes de cartón. Se considera que el bote es disconforme si al llenarlo gotea. Se sabe que la proporción de botes defectuosos es 0.23.

Para controlar el proceso se establece un control por atributos, tomando muestras de tamaño 50 cada hora.

Las muestras observadas en las últimas 30 horas presentan los siguientes resultados.

nº muestra	nº defectuosos	nº muestra	nº defectuosos
1	12	16	8
2	15	17	10
3	8	18	5
4	10	19	13
5	4	20	11
6	7	21	20
7	16	22	18
8	9	23	24
9	14	24	15
10	10	25	9
11	5	26	12
12	6	27	7
13	17	28	13
14	12	29	9
15	22	30	6

El gráfico de control por número de defectos

$$\begin{aligned} LCS &= np + 3\sqrt{np(1-p)} = 50 * 0.23 + 3\sqrt{50 * 0.23 * 0.77} = 20.43 \\ LC &= np = 50 * 0.23 = 11.5 \\ LCI &= np - 3\sqrt{np(1-p)} = 50 * 0.23 - 3\sqrt{50 * 0.23 * 0.77} = 2.57 \end{aligned}$$

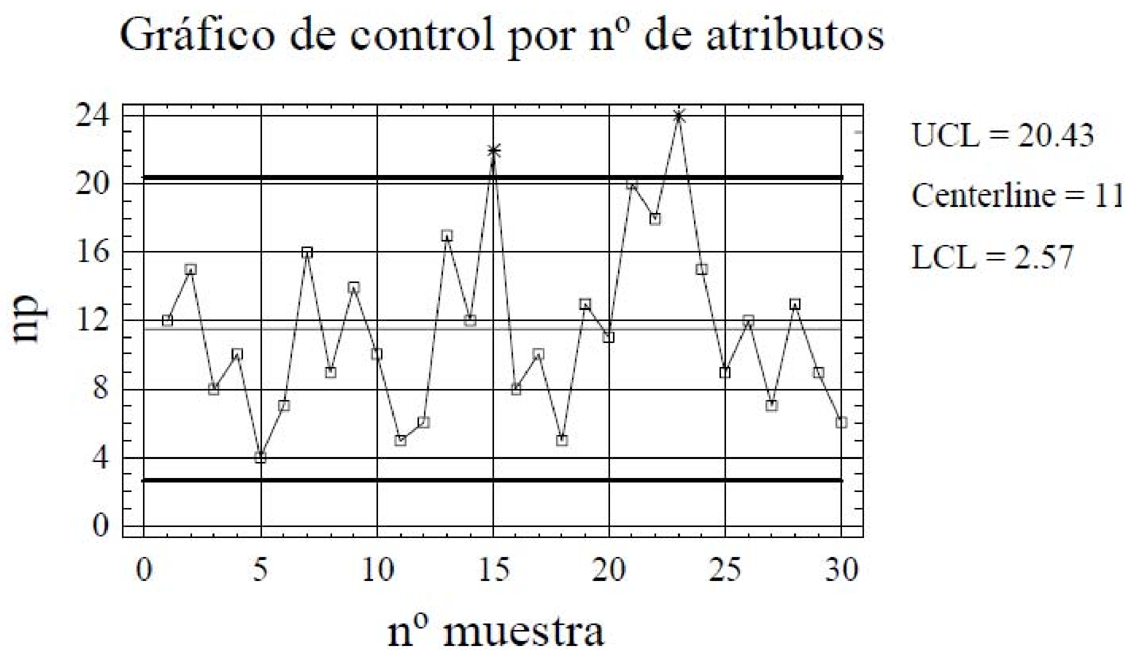


Gráfico de número de defectuosos (p desconocida)

De forma análoga a procedimientos estudiados anteriormente. Se toman k muestras de tamaño n (suponemos que son tomadas cuando el proceso está bajo control).

1 Estimamos p utilizando las k muestras

$$\widehat{p} = \frac{r_1 + r_2 + \ldots + r_k}{k * n}$$

2 Sustituyendo las estimaciones en las fórmulas de las líneas del gráfico de control para el caso de p conocida obtenemos:

gráfico para nº ºde defectuosos
$LCS = n\widehat{p} + 3\sqrt{n\widehat{p}(1-\widehat{p})}$
$LC = n\widehat{p}$
$LCI = n\widehat{p} - 3\sqrt{n\widehat{p}(1-\widehat{p})}$

Estudio inicial (capacidad del proceso)

En el apartado anterior hemos supuesto que las muestras han sido tomadas con el proceso en estado de control. A continuación mostramos un procedimiento iterativo que se utiliza para garantizar (al menos de forma aproximada) que las muestras empleadas para hacer la estimación de parámetros corresponden al proceso en estado de control.

Pasos:

1. Se toman k muestras (k al menos 25) de tamaño n (n mayor que 50).
2. Se estiman los parámetros

$$\widehat{p} = \frac{r_1 + r_2 + \dots + r_k}{k * n}$$

3. Se construyen el gráfico de control.
4. Representar en el gráfico el número de defectuosos en cada muestra. Si algún punto se encuentra fuera de los límites de control, eliminar la muestra del conjunto. En este caso volver al paso 2 para recalcular las estimaciones de los parámetros con el conjunto modificado de muestras. Si todos los puntos están dentro del gráfico de control ir al paso siguiente.
5. Al encontrarse todas las muestras dentro del gráfico suponemos que todas han sido tomadas cuando el proceso estaba bajo control. De esta forma estimamos la capacidad del proceso como $1 - \widehat{p}$.

Ejemplo. Determinación de la capacidad del proceso.

1 Iteración

1. Se toman 30 muestras de tamaño 50.

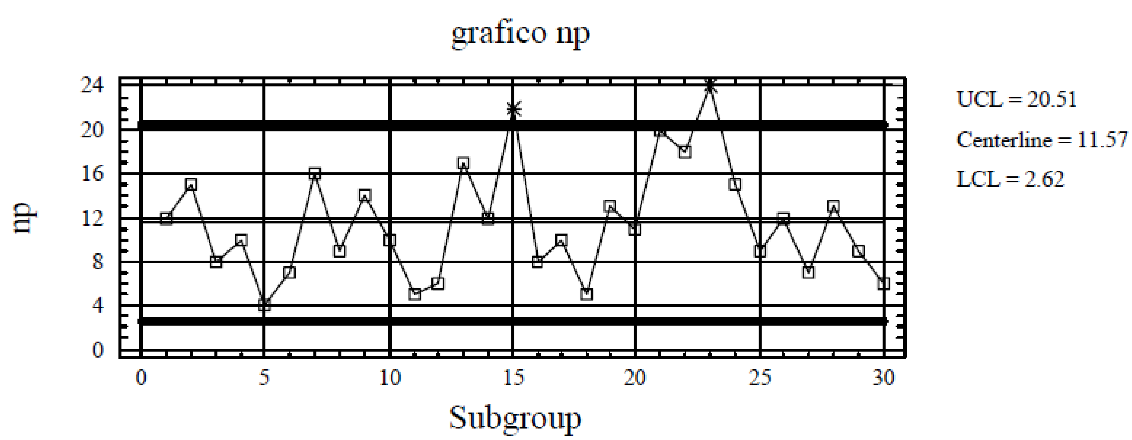
nº muestra	nº de defec	nº muestra	nº de defec	nº muestra	nº de defec.
1	12	11	5	21	20
2	15	12	6	22	18
3	8	13	17	23	24
4	10	14	12	24	15
5	4	15	22	25	9
6	7	16	8	26	12
7	16	17	10	27	7
8	9	18	5	28	13
9	14	19	13	29	9
10	10	20	11	30	6

2. Estimación

$$\widehat{p} = \frac{\sum r_i}{n * k} = \frac{347}{50 * 30} = 0.2313$$

3. Gráfico de control para el número de defectos

$$\begin{aligned} LCS &= n\widehat{p} + 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.2313 + 3\sqrt{50 * 0.2313 * 0.7687} = 20.51 \\ LC &= n\widehat{p} = 11.57 \\ LCI &= n\widehat{p} - 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.2313 - 3\sqrt{50 * 0.2313 * 0.7687} = 2.62 \end{aligned}$$



4. Dos muestras se salen del gráfico 15 y 23. Se eliminan.

2 Iteración

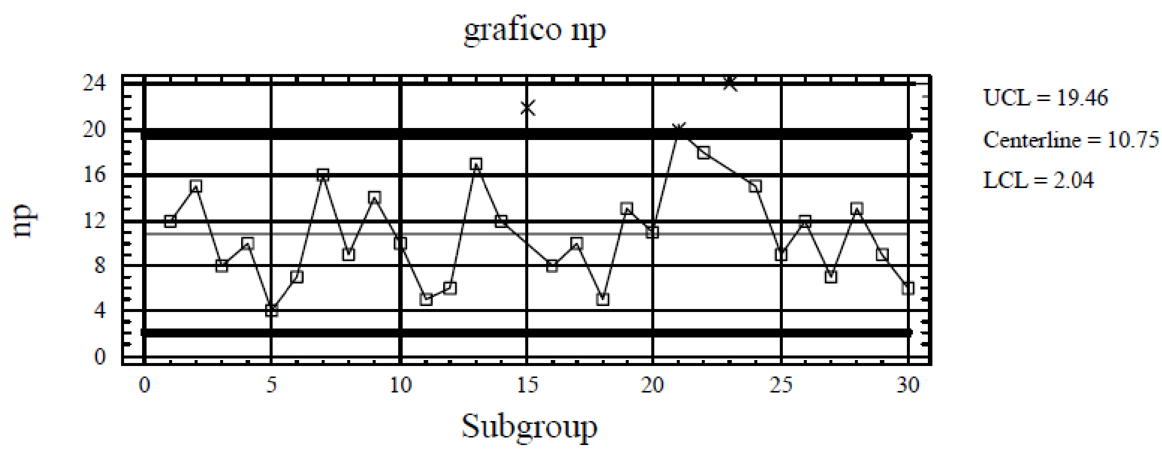
1. Se vuelve a estimar p

$$\widehat{p} = \frac{301}{50 * 28} = 0.215$$

2. Nuevo gráfico

$$\begin{aligned} LCS &= n\widehat{p} + 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.215 + 3\sqrt{50 * 0.215 * 0.785} = 19.46 \\ LC &= n\widehat{p} = 10.5 \\ LCI &= n\widehat{p} - 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.215 - 3\sqrt{50 * 0.215 * 0.785} = 2.04 \end{aligned}$$

3. Gráfico



4. Se sale la muestra 21. La eliminamos.

3 Iteración

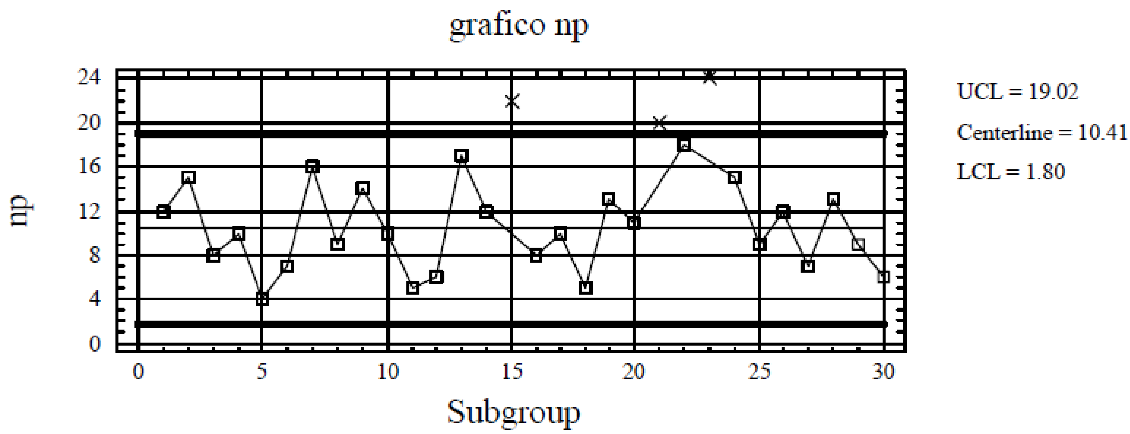
1. Nueva estimación

$$\widehat{p} = \frac{281}{50 * 27} = 0.208$$

2. Gráfico

$$\begin{aligned} LCS &= n\widehat{p} + 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.208 + 3\sqrt{50 * 0.208 * 0.792} = 19.02 \\ LC &= n\widehat{p} = 10.41 \\ LCI &= n\widehat{p} - 3\sqrt{n\widehat{p}(1-\widehat{p})} = 50 * 0.208 - 3\sqrt{50 * 0.208 * 0.792} = 1.20 \end{aligned}$$

3. Gráfico



4. Todas las muestras están dentro del gráfico por lo que suponemos que todas han sido tomadas cuando el proceso estaba bajo control.

Estimación de la capacidad del proceso: 1- $\hat{p}$ =1-0.208

4 Aplicaciones estadísticas en administración de calidad y productividad.

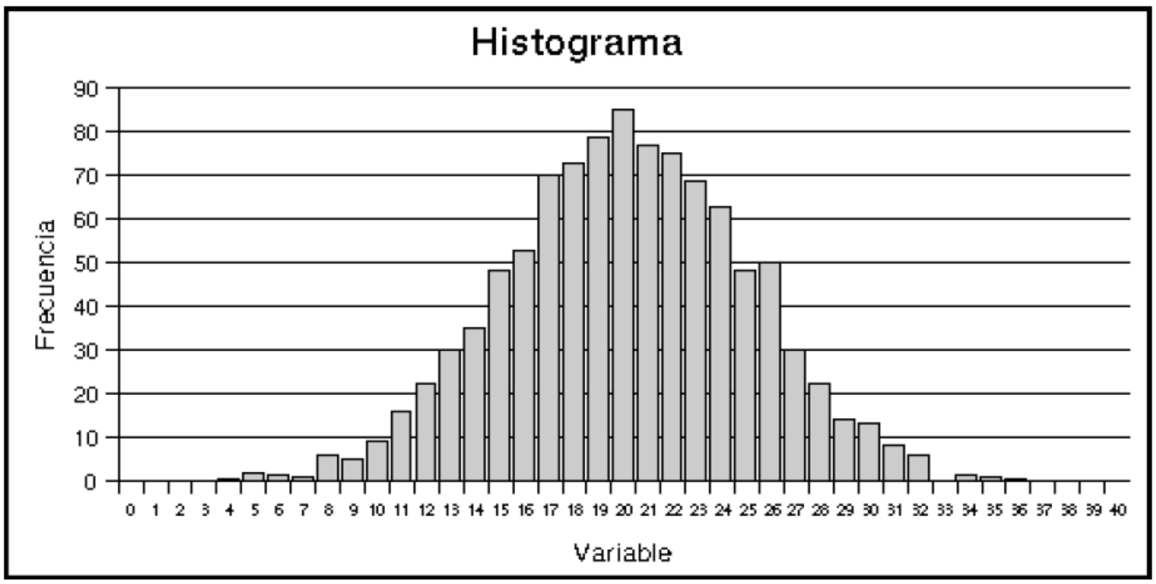
Antes de aplicar cualquier técnica estadística, es necesario establecer algunas hipótesis bajo las cuales se va a desarrollar el análisis. En primer lugar, vamos a suponer que la característica de calidad (Variable aleatoria) es continua y de distribución normal. En segundo lugar, consideraremos que el proceso está bajo control estadístico, es decir que la variabilidad se debe solamente a un sistema constante de causas aleatorias (No intervienen causas asignables).

Al realizar una sucesión de mediciones de la característica de calidad sobre muestras del producto fabricado, encontramos que los valores fluctúan alrededor de un valor central. Esto es lo que llamamos la fluctuación natural y esperable del proceso. Esta variación de la característica de calidad medida se debe a un conjunto muy grande de causas que afectan el proceso, cuyo efecto individual es pequeño y que actúan en forma aleatoria (Sistema constante de causas aleatorias). La fluctuación natural del proceso es inherente al mismo y no puede eliminarse, sólo puede reducirse realizando modificaciones al proceso mismo, lo cual significa, como ya hemos dicho, trabajar con otro proceso. La fluctuación natural de un proceso puede cuantificarse a través de la desviación estándar del mismo, con la cual podemos calcular Límites de Tolerancia Natural del proceso. Se debe insistir en que estos límites no pueden fijarse voluntariamente, dependen del proceso y de las variables no controlables del mismo. Generalmente se toma un rango para la fluctuación natural de 6 sigmas.

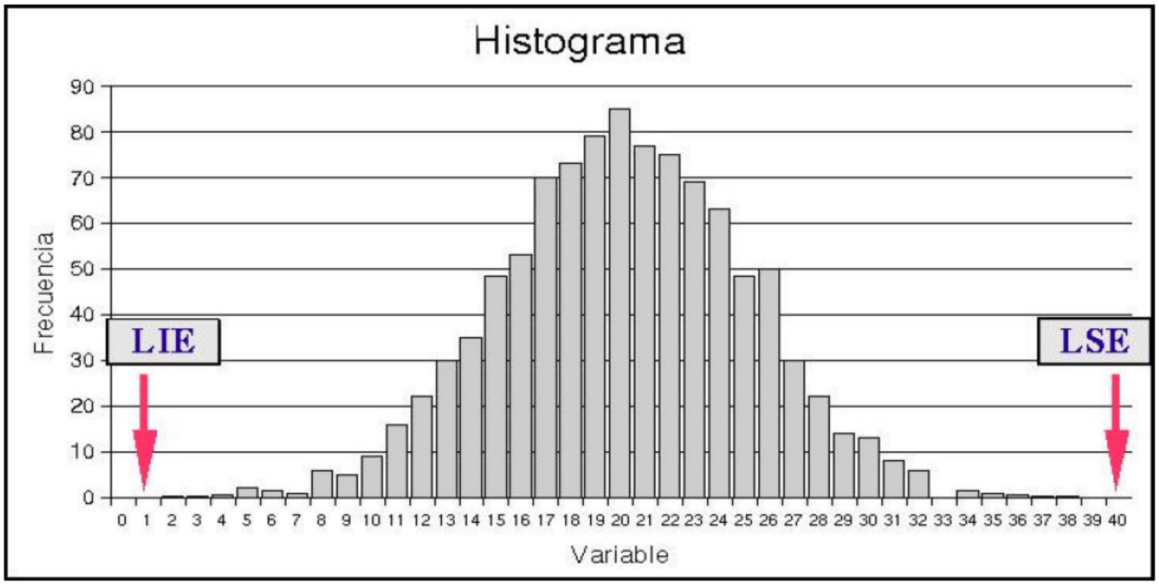
Los Límites de Especificación de un producto son fijados voluntariamente por el cliente, por el fabricante o por alguna norma. Estos límites constituyen un requisito a cumplir por el producto y no deben confundirse en ningún caso con los Límites de Control o con los Límites de Tolerancia Natural del proceso.

La Capacidad de un proceso es la aptitud para generar un producto que cumpla con determinadas especificaciones. En el mejor de los casos, es conveniente que los Límites de Tolerancia Natural del proceso se encuentren dentro de los Límites de Especificación del producto. De esta manera nos aseguramos que toda la producción cumplirá con las especificaciones.

Para analizar la capacidad del proceso se puede utilizar un histograma de frecuencias. Si se dispusiera de todos los datos del universo para la característica de calidad medida y se hiciera un histograma este permitiría tener una idea exacta de la fluctuación natural del proceso. Como esto es imposible, es necesario tomar un cierto número de mediciones (Mínimo 100-200) y efectuar con ellas un histograma de frecuencias.



Este es el histograma de una muestra y por lo tanto es sólo una estimación del verdadero histograma del universo. Si representamos en las abscisas los Límites de Especificación del producto, podemos ver gráficamente si el proceso tiene aptitud (Capacidad) para fabricar dicho producto.



Para cuantificar la Capacidad de Proceso se utilizan coeficientes que permiten comparar el rango de especificaciones con la fluctuación natural del proceso.

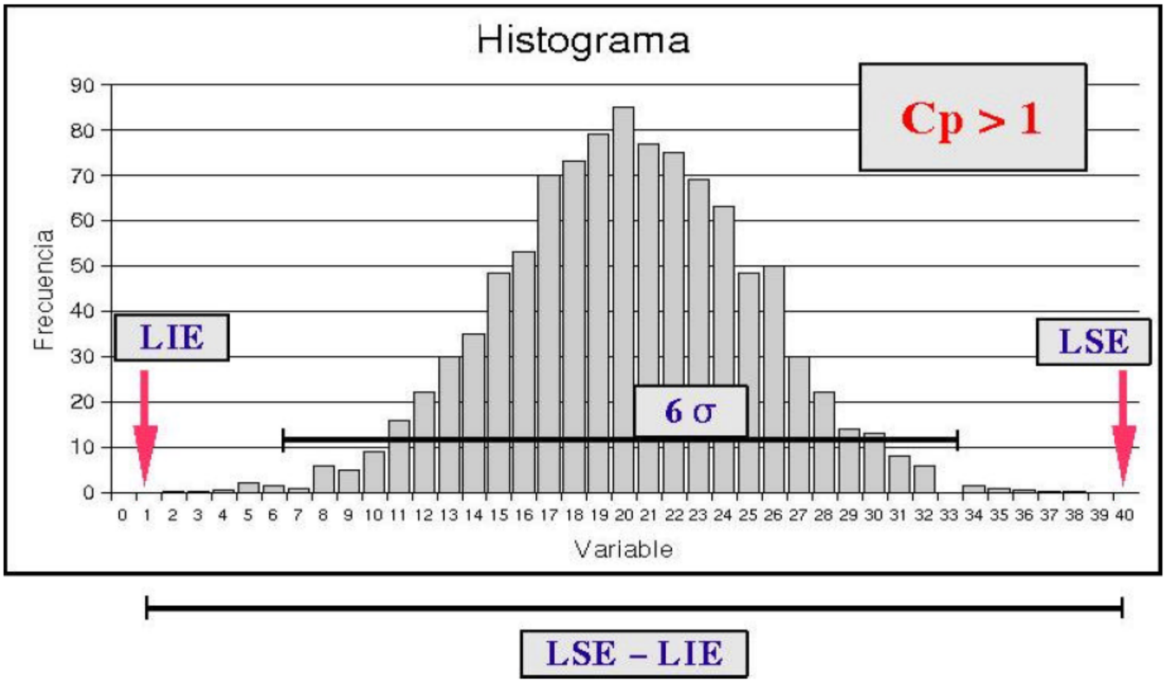
Uno de ellos es Cp.

$$C_p = \frac{LT_2 - LT_1}{6\sigma}$$

Otros autores lo escriben como

$$Cp = \frac{(LSE - LIE)}{6 * \sigma}$$

Si el proceso tiene capacidad para fabricar el producto, entonces Cp > 1. En general se exige Cp > 1.30 para mayor seguridad.

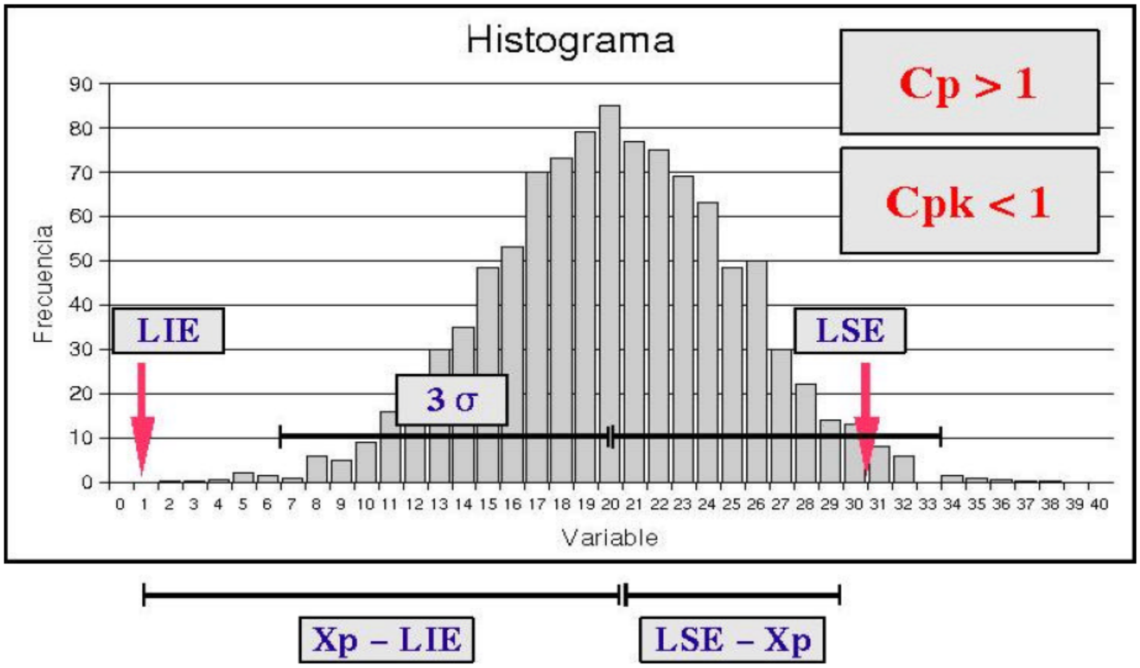


Este coeficiente tiene el inconveniente de que para poder aplicarlo, el centro de gravedad del rango de especificaciones debe coincidir con la tendencia central de las mediciones del proceso. Cuando esto no ocurre se emplea el Cpk:

$$C_{pk} = \frac{\Delta}{3*\sigma}$$

Donde

$$\Delta = \text{Mínimo entre} \left[LSE - \overline{X} \right] \text{ y } \left[\overline{X} - LIE \right]$$



En el gráfico podemos observar que una buena parte del producto está por encima del Límite Superior de Especificación (LSE). Aún así resulta Cp > 1, indicando erróneamente que el proceso tiene capacidad suficiente. En este caso se debe usar el segundo coeficiente que muestra claramente que el proceso no tiene capacidad suficiente (Cpk < 1), tal como se puede observar en el gráfico.

El uso de un histograma para analizar la capacidad de un proceso tiene la ventaja de que se puede apreciar la forma de la distribución, con lo cual se puede confirmar o rechazar la hipótesis de que la misma es normal. Pero el problema es que no se puede detectar la presencia de patrones no aleatorios, con lo cual no es posible confirmar o rechazar la hipótesis de que el proceso está bajo control estadístico. Si el proceso no está bajo control estadístico los resultados del análisis de la capacidad de proceso no serán válidos y pueden llevar a conclusiones equivocadas.

Otra manera de analizar la capacidad de un proceso es por medio de los gráficos de control.

El Diagrama de Pareto

El **diagrama de Pareto** consiste en un **método gráfico para determinar cuáles son los problemas más importantes de una determinada situación** y por consiguiente, **las prioridades de intervención**.

Permite identificar los factores o problemas más importantes en función de la premisa de que pocas causas producen la mayor parte de los problemas y muchas causas carecen de importancia relativa.

Para la construcción del diagrama de Pareto se procede según las fases que son las siguientes:

• Decidir cómo clasificar los datos
• Elegir el período de observación
• Obtener los datos y ordenarlos
• Preparar los ejes cartesianos del diagrama
• Diseñar el diagrama

Para mayor claridad se examina un caso como ejemplo. Se supone que en un departamento de montaje en una industria se producen determinadas fallas. Se aplicará el diagrama de Pareto con las siguientes fases.

• **Fase 1: Decidir cómo clasificar los datos**: Se pueden clasificar por tipo de problema, por cadena de montaje, por turno de trabajo, por fase de trabajo, etc. Se establece por tipo de problema.

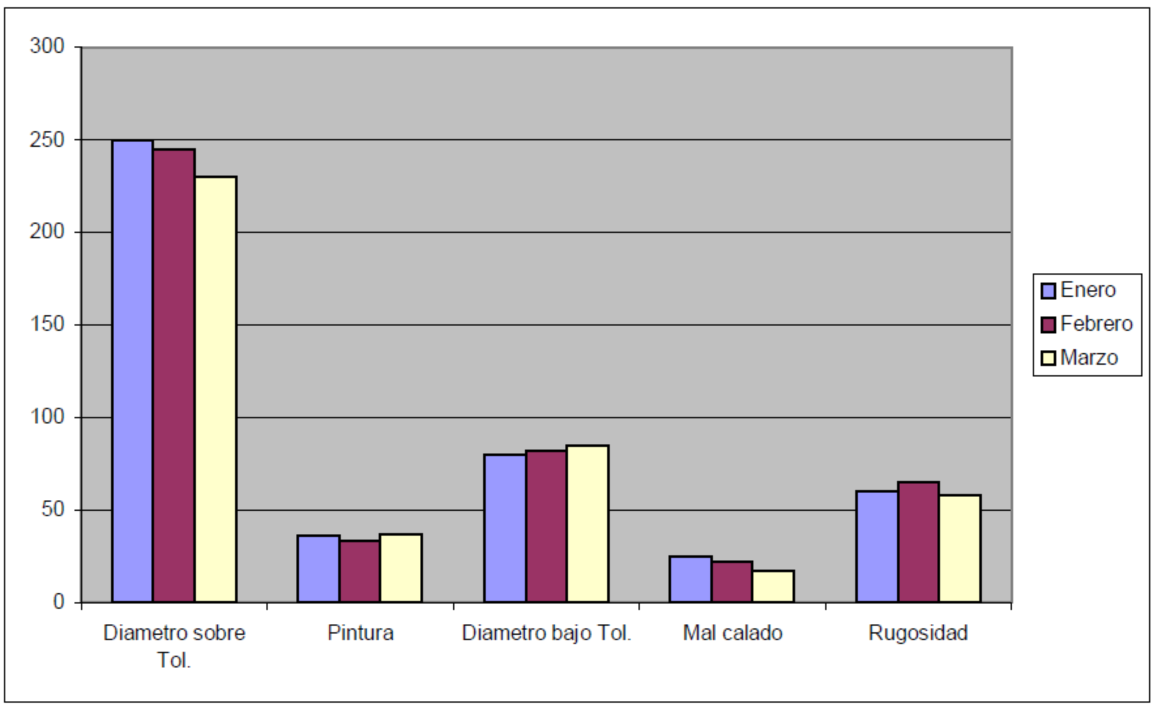
• **Fase 2: Elegir el período de observación**: En el caso del ejemplo dependerá de la cantidad de productos fabricados. Si la cantidad diaria es elevada, será suficiente un período breve, por el contrario, cuando la producción es reducida será necesario un período más prolongado. Se decide realizar el relevamiento por 3 meses.

• **Fase 3: Obtener los datos y ordenarlos:** En ésta fase se tendrá que preparar una hoja para recoger los datos según las pautas establecidas en las fases precedentes: tipo de problema y un período de 3 meses.

Durante el período de relevamiento de datos se completará el formulario con un trazo por cada defecto encontrado y luego se determinan los totales, según se muestra en el formulario.

Tabla

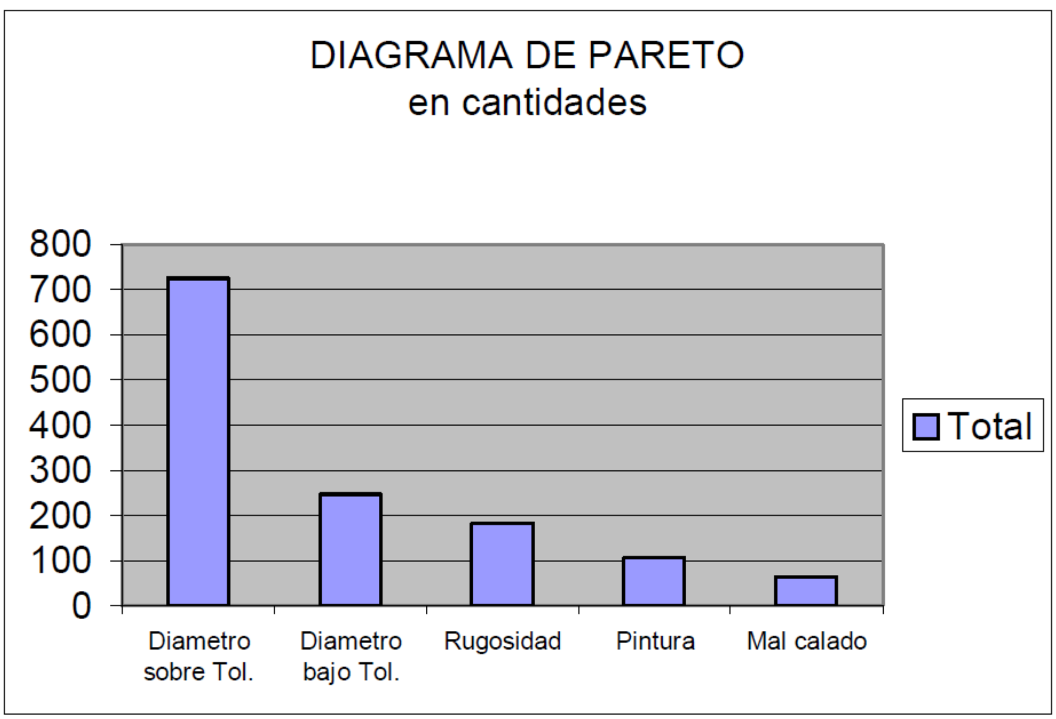
Num.	Defecto	Meses			Total
		Enero	**Febrero**	**Marzo**	
1	Diámetro sobre Tol.	250	245	230	725
2	Pintura	36	33	37	106
3	Diámetro bajo Tol.	80	82	85	247
4	Mal calado	25	22	17	64
5	Rugosidad	60	65	58	183
	TOTAL				1325



A continuación se procede a ordenar los conceptos por orden de importancia en una tabla como se muestra en la figura, el defecto más numeroso se dispone en primer lugar, en segundo lugar el defecto que le sigue por orden de frecuencia, y así sucesivamente, etc. En la última columna se indica la cantidad total de problemas.

TABLA		
Num.	**Defecto**	**Total**
1	Diámetro sobre Tol.	725
3	Diámetro bajo Tol.	247
5	Rugosidad	183
2	Pintura	106
4	Mal calado	64
	TOTAL	1325

* **Fase 4: Preparar los ejes cartesianos para el diagrama:** En el eje **X** se dispondrán los tipos de defectos y en el **Y** las cantidades de defectos. Se pueden graficar los problemas según las cantidades o en función de los porcentajes con respecto al total de problemas.



Los defectos se ordenan en forma similar a la tabla, en orden de mayor a menor frecuencia.
Para definir la escala del eje Y, se tiene que considerar que el valor más grande corresponde al primer defecto, según lo determinado, y será la base para la escala de valores absolutos o porcentual. El eje X se divide proporcionalmente según la cantidad de grupos de problemas a graficar, en el ejemplo son 5 grupos.

Fase 5: Diseñar el diagrama: Se procede a representar en escala, con bastones los valores absolutos y/o porcentuales que se han determinado en la tabla.
Otra forma de visualizar los problemas es proceder a ordenar también los conceptos por orden de importancia en una tabla y en la ultima columna se indica la cantidad en porcentaje.

Tabla

Num.	Defecto	%
1	Diámetro sobre Tol.	54,72
3	Diámetro bajo Tol.	18,64
5	Rugosidad	13,81
2	Pintura	8,00
4	Mal calado	4,83
	TOTAL	100

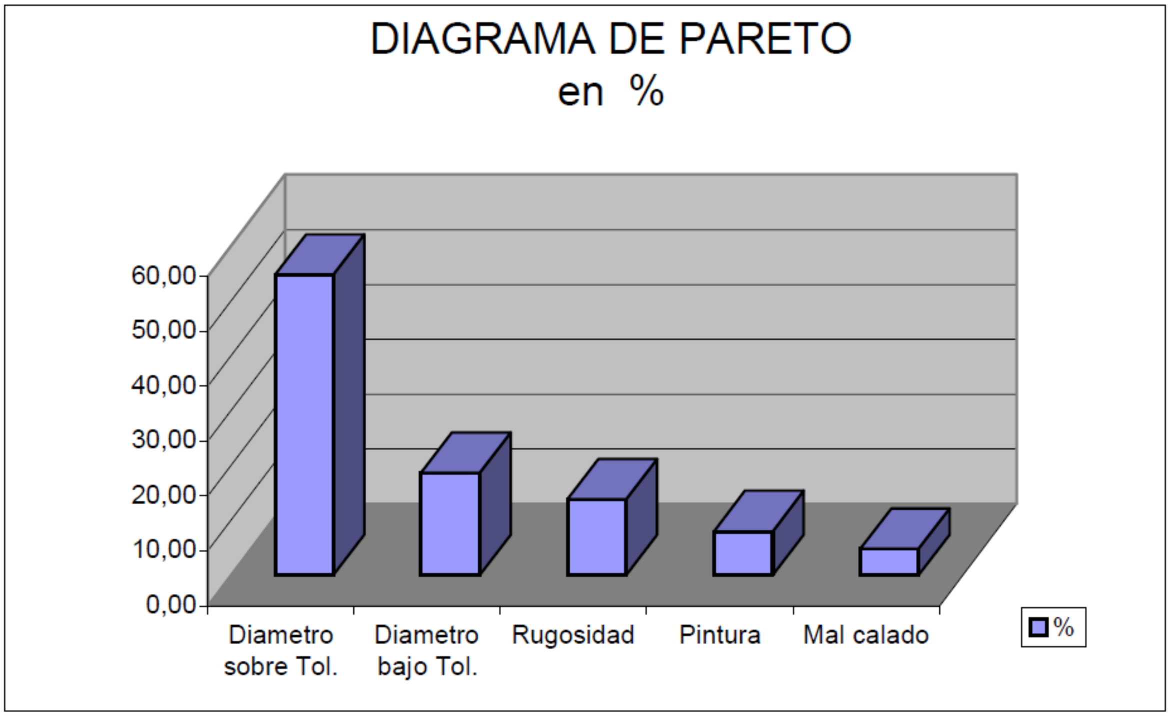


DIAGRAMA DE PARETO
en %
60,00
50,00
40,00
30,00
20,00
10,00
0,00
Diametro sobre Tol.
Diametro bajo Tol.
Rugosidad
Pintura
Mal calado
%

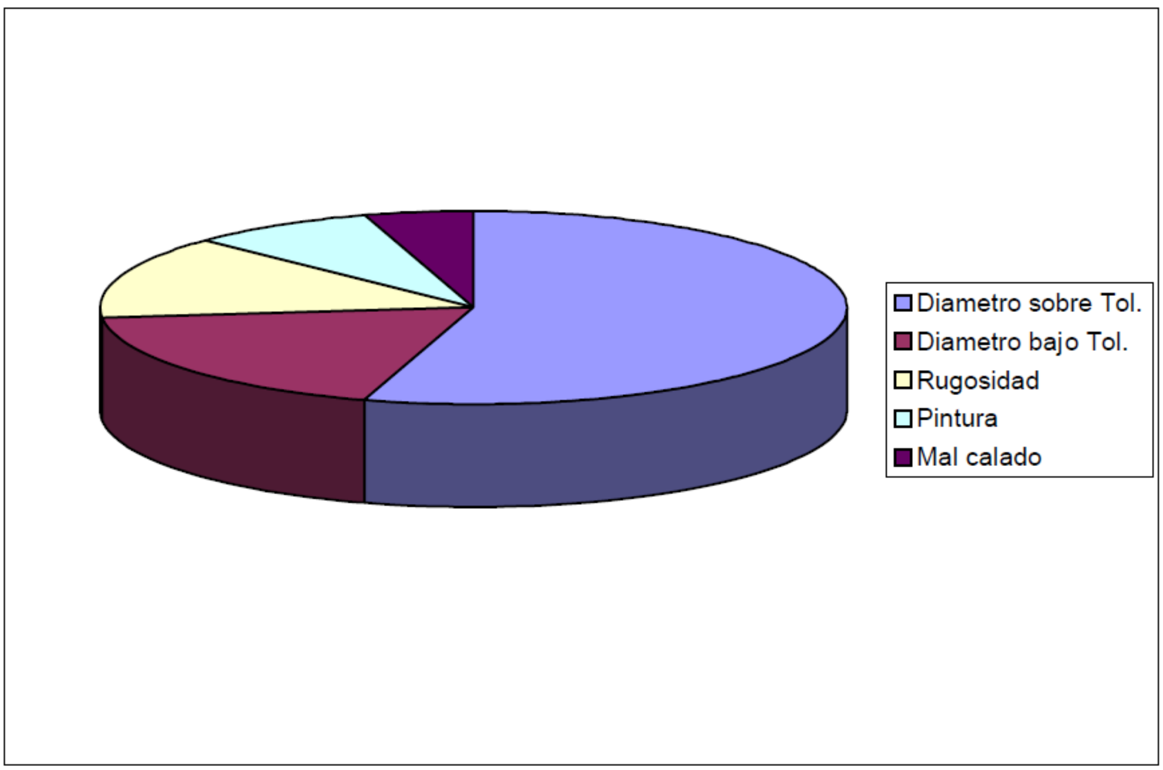
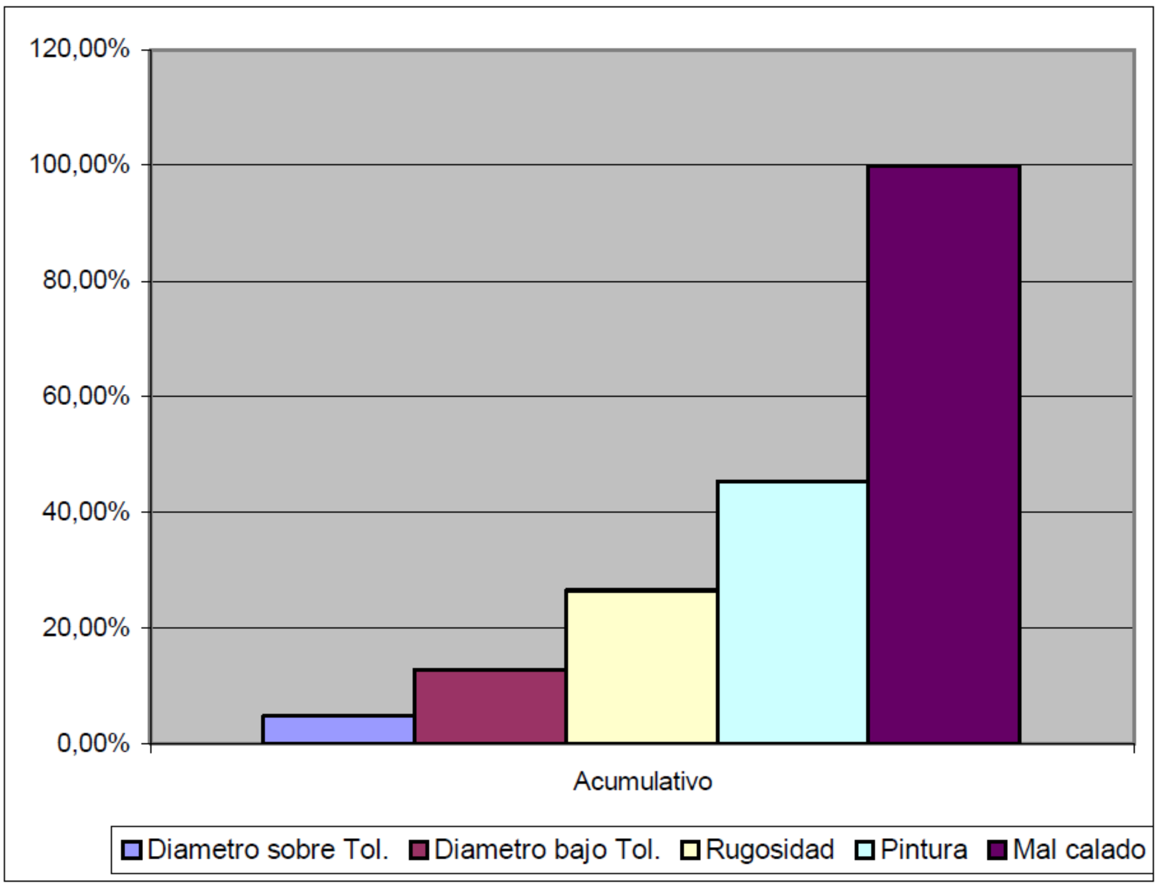
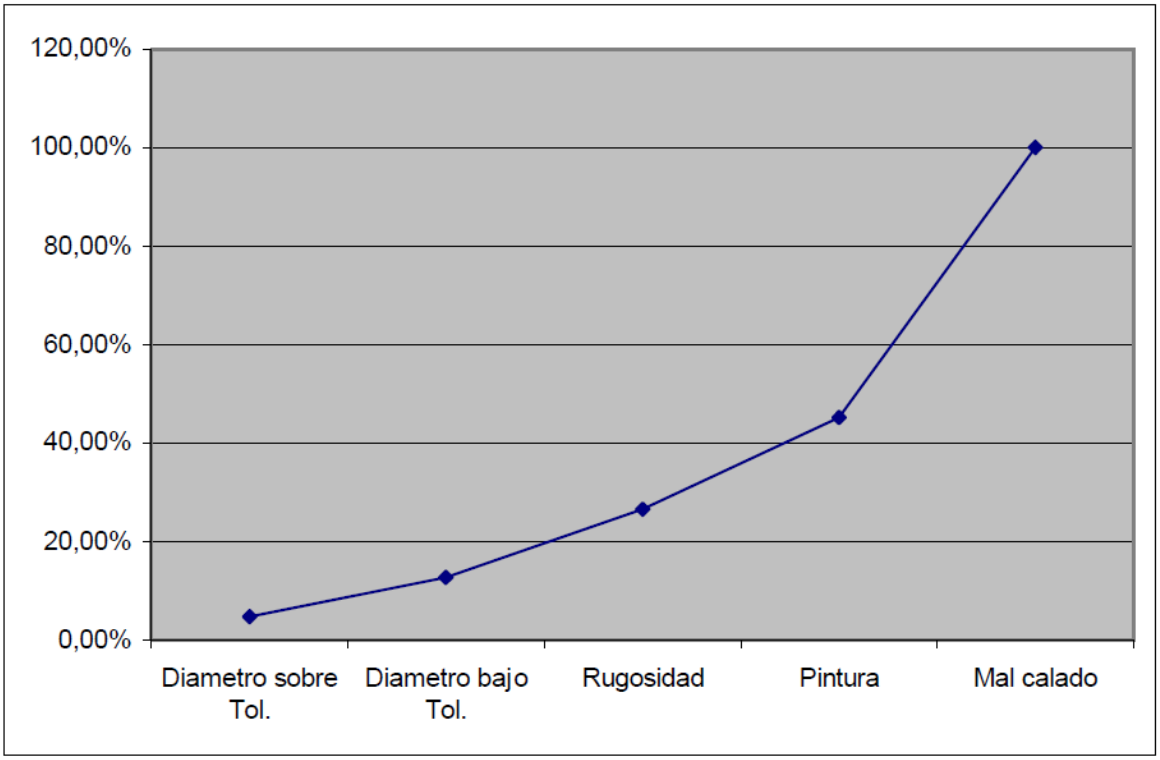

Diametro sobre Tol.
Diametro bajo Tol.
Rugosidad
Pintura
Mal calado

TABLA		
Num.	Defecto	Acumulativo
1	Diámetro sobre Tol.	4,83%
3	Diámetro bajo Tol.	12,83%
5	Rugosidad	26,64%
2	Pintura	45,28%
4	Mal calado	100,00%



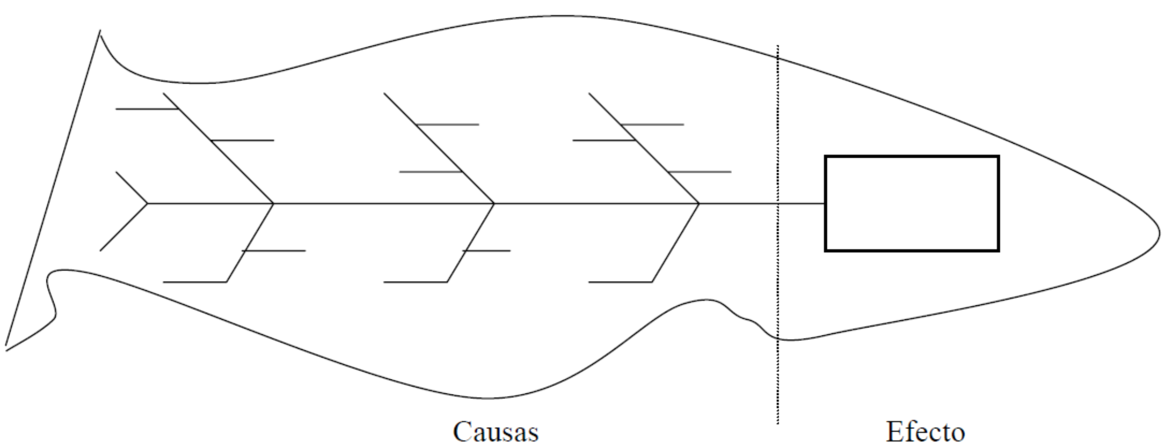


El Diagrama Causa-Efecto

Este diagrama **se utiliza para representar la relación entre algún efecto y todas las causas posibles que lo pueden originar**.

Todo tipo de problema, como el funcionamiento de un motor o una lámpara que no enciende, puede ser sometido a éste tipo de análisis.

Generalmente, se lo presenta con la forma del espinazo de un pez, de donde toma el nombre alternativo de Diagrama de espina de pescado. También se lo llama de Diagrama de Ishikawa que es quién lo impulsó.



Los diagramas de causa efecto se construyen para ilustrar con claridad cuáles son las posibles **causas** que producen el problema. Un eje central se dirige al **efecto**. Sobre el eje se disponen las posibles **causas**. El análisis causa-efecto, es el proceso mediante el que se parte de una definición precisa del efecto que se desea estudiar.

Posteriormente, se disponen todas las causas que pueden provocar el efecto.

A las causas conviene agruparlas por tipos, al modo de ejemplo las originadas por motivos eléctricos, otras por elementos mecánicos, hidráulicos, etc. Cada grupo se dispone en un subeje.

La construcción de este diagrama presenta un esquema gráfico que permite efectuar un análisis de las causas que influyen sobre el efecto objeto de estudio.

El análisis causa-efecto puede dividirse en tres etapas:

• Definición del efecto que se desea estudiar.
• Construcción del diagrama causa-efecto.
• Análisis causa-efecto del diagrama construido.

La definición del efecto que se desea estudiar representa la base de un eficaz análisis.

Efectivamente, siempre es necesario efectuar una precisa definición del efecto objeto de estudio. Cuanto más definido se encuentre éste, tanto más directo y eficaz podrá ser el análisis de las causas. Así si el motor del automóvil no arranca, ¿cuáles pueden ser las causas de la falta de arranque? Evidentemente, las causas posibles pueden ser múltiples. **Si se definiera el efecto como el motor no arranca cuando esta muy frío y el vehículo se encuentra a la intemperie, en este caso el análisis será más preciso y estamos eliminando una serie de causas que no corresponden a la situación del vehículo.**

Invirtiendo el razonamiento se puede decir que cuando más indefinido se exprese el efecto que se desea estudiar, tanto más amplio e indeterminado será el diagrama causa-efecto y por lo tanto, más vago y de mayor complejidad el análisis y resolución del problema. Cuando se tiene bien definido el efecto que se desea estudiar, se puede proceder a las dos fases sucesivas si se tiene la prudencia de separar la fase segunda -construcción del diagrama- de la fase tercera -análisis y valoración de las diversas causas-.

De este modo es posible garantizar que la definición de las posibles causas sea innovadora y creativa, mientras que el análisis crítico de las causas debe ser lo más realista posible. En realidad cuanto más ideas y sugerencias contenga el diagrama causa-efecto, tanto más eficaz será para la determinación de la causa o las causas, ya que el problema puede ser originado por más de una.

Construcción del Diagrama Causa-Efecto

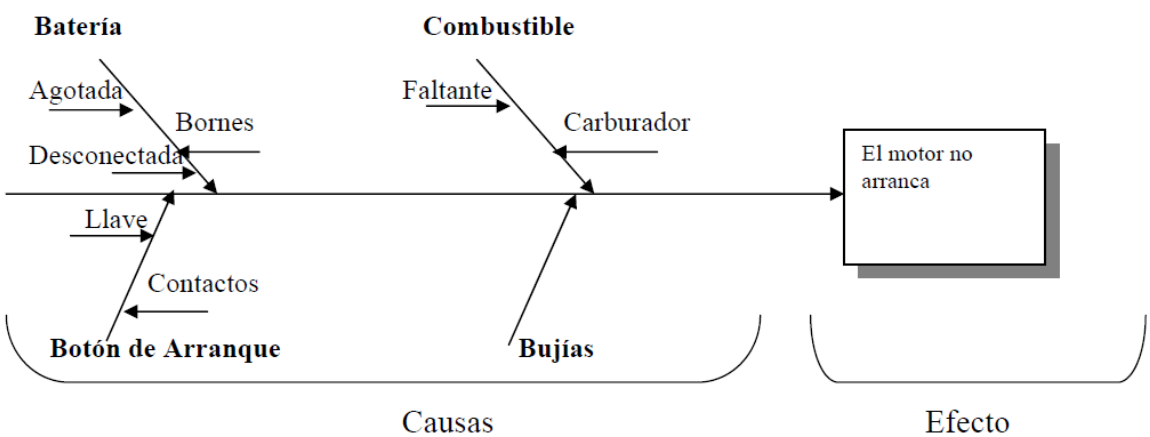
La construcción del diagrama causa-efecto se inicia escribiendo el efecto que se desea estudiar en el lado derecho de una hoja de papel. A ello debe seguir la búsqueda de todas las posibles causas que sobre él influyen. Para esa búsqueda se pueden seguir **tres métodos**, que se diferencian por la forma en que se realizan.

Son los siguientes:

• Método de la **Clasificación de las Causas**.

• Método por **Fases del Proceso**.

• Método por **Enumeración de las Causas**.



Método Brainstorming

Este método **Brainstormig** que traducido a nuestro idioma significa "**Tormenta de Ideas**" consiste básicamente en que todos los participantes expongan sus ideas, que las mismas sean anotadas, luego comentadas, para finalmente llegar a conclusiones.

Para llevar a cabo ésta actividad es conveniente establecer un orden de prioridades, y seguir los siguientes pasos:

- Nombrar a un **moderador del grupo**, quien debe asegurar que todos comprendan el problema. Será el encargado de observar que se anoten las ideas que se propongan, en un lugar visible, preferentemente construyendo el diagrama.
- Antes de iniciar la propuesta de ideas dar 5 a 6 minutos en **silencio** pensando en **el problema** en forma **individual**.
- Por turnos, **cada miembro enuncia una idea**. No se permiten comentarios ni críticas. En ésta etapa sólo pueden intervenir el encargado de anotar las ideas y a quien le corresponde el turno.
- Cuando alguno de los participantes no tenga idea para sugerir el moderador esperará poco tiempo y pasará al turno de quien continua. Cuando las ideas hayan comenzado a agotarse -aproximadamente a los 30 minutos-, el grupo analiza y discute las ideas anunciadas. Las ideas duplicadas o relacionadas se agrupan. Se pueden descartar las ideas que no tienen fundamento serio, siempre sin realizar críticas.
- De todas las ideas **se analizan cuáles pueden ser las más probables**. Se puede aplicar el diagrama de Pareto y sobre las causas que concentran la atención realizar un relevamiento de datos.

Cuando el problema es complejo se puede construir el diagrama de las 5 M a lo largo de varias sesiones.

En algunos casos la causa puede estar en más de alguna categoría, según la decisión del grupo se la dispone por mayoría en las distintas categorías o en la que se considere más indicada.

La revisión directa del diagrama puede impulsar al grupo a decidir una profundización de la investigación en un área determinada.

Método de las 5 M

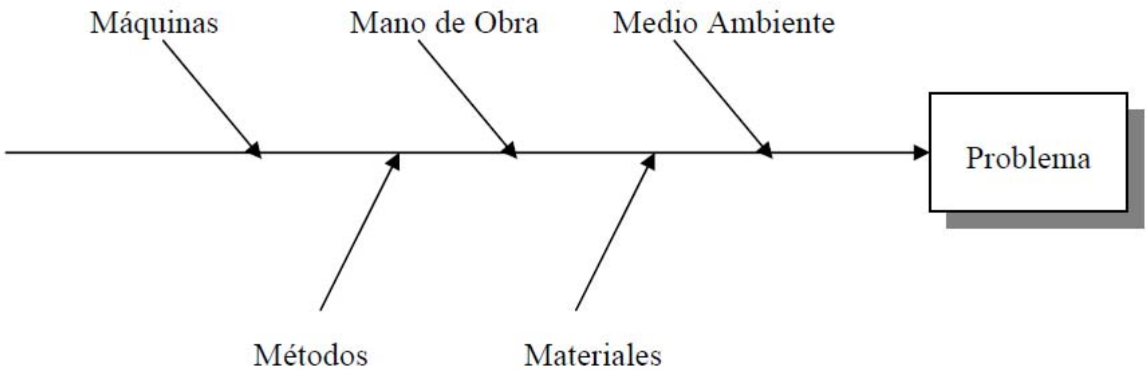
Conforme al presente método se procede a analizar el problema y a definir las posibles causas, generalmente este proceso se realiza con el grupo de trabajo encargado de la resolución del problema.
Para la aplicación de este método se sigue un orden para considerar las causas de los problemas, partiendo de la premisa que estas, están agrupadas según cinco criterios y por ello se denomina de las **5 M**.

Las **M** corresponden a:
- Máquinas
- Mano de Obra
- Métodos
- Materiales
- Medio Ambiente

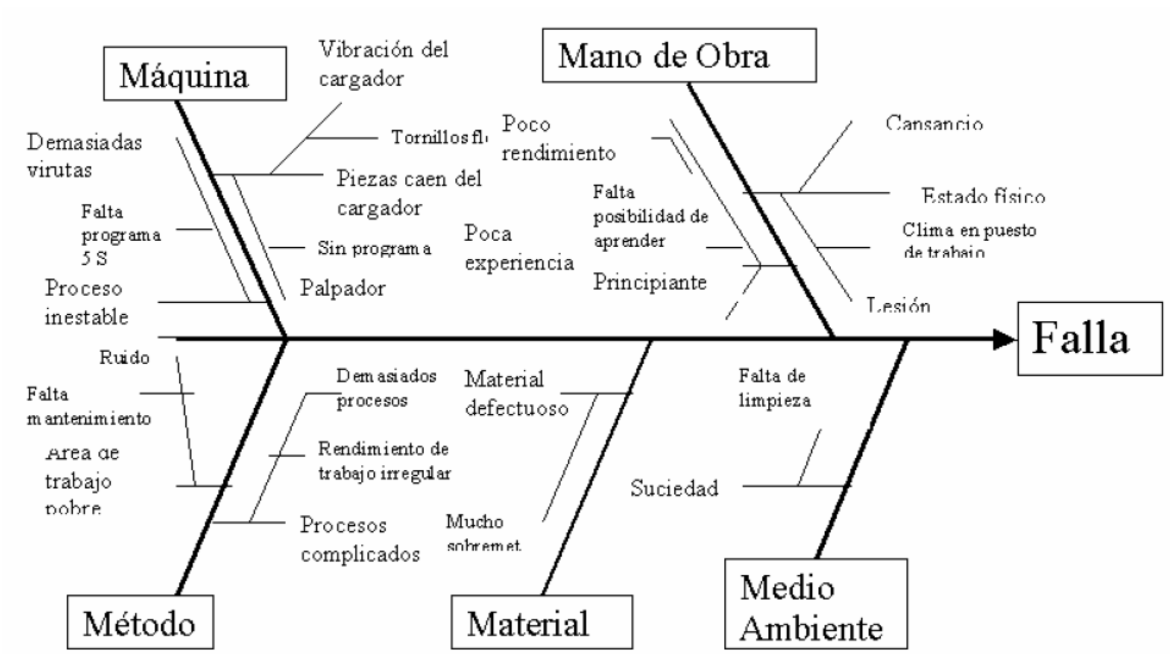
Las 5 M suelen ser generalmente un punto de referencia que abarca casi todas las principales causas de un problema, por lo que constituyen los brazos principales del diagrama causa-efecto.

Estructura Básica de las 5 M

A continuación se puede proceder una “Lluvia o Tormenta de Ideas” – Brainstorming que consiste en generar tantas ideas como sea posible dejando que el pensamiento creativo de cada persona del grupo las exponga libremente.
Las subdivisiones en base a las 5 M, además de organizar las ideas estimulan la creatividad. En ésta fase quienes intervienen deben liberarse de preconceptos, en caso contrario se puede condicionar la búsqueda a las soluciones que ya se han propuesto o probado y que no han aportado la solución. Las causas sugeridas se incluyen situándolas en el brazo correspondiente. En el ejemplo se ilustra con algunas de las posibles causas en forma genérica.



Ejemplo del diagrama de las 5 M.



Procedimientos Básicos para Analizar los Problemas

Antes de investigar un problema, es fundamental asegurarse de que se lo comprende perfectamente. Esto supone definir los síntomas del problema y comprender el proceso que lo provoca., así se evita desperdiciar esfuerzos innecesariamente.

Cuando se comprende y define un problema se ha avanzado bastante en su resolución.

Las herramientas que más se utilizan para ayudar a definir un problema son las listas de comprobación y los diagramas de flujo.

Lista de comprobación.

¿Qué?	¿Cuál es el problema? ¿Qué se ha observado?
¿Quién?	¿Quién interviene en el problema? ¿Quién está antes o después del problema en el flujo de trabajo?
¿Dónde?	¿Dónde se manifiesta? ¿Dónde se origina?
¿Cuándo?	¿En qué ocasión aparece? ¿En qué momentos y por cuánto tiempo?
¿Cómo?	¿Cómo se manifiesta? ¿Con cuánta frecuencia ocurre? ¿Cuál es la importancia del problema? ¿Cuál es la importancia en tiempo perdido? ¿Cuál es la importancia en costos? ¿Cuál es la importancia en cuanto a la frecuencia?
¿Por qué?	¿Por qué ocurre el problema? **Pregunta clave que se debe responder.**

Este tipo de consideraciones centra la atención sobre el problema, y contribuye a dar cohesión al grupo de trabajo.

Apendice A: CONTROL POR VARIABLES

$$X \ ; \ N(\mu, \sigma)$$

1. μ, σ CONOCIDAS

1.1 GRÁFICO DE MEDIAS

$$\begin{aligned} LCS &= \mu + 3\frac{\sigma}{\sqrt{n}} \\ LC &= \mu \\ LCI &= \mu - 3\frac{\sigma}{\sqrt{n}} \end{aligned}$$

1.2 GRÁFICO DE DESVIACIONES(corregidas) O RANGOS

1.2.1 DESVIACIONES CORREGIDAS

$$\begin{aligned} LCS &= B_6\sigma \\ LC &= c_4\sigma \\ LCI &= B_5\sigma \end{aligned}$$

1.2.2 RANGOS

$$\begin{aligned} LCS &= d_2\sigma + 3d_3\sigma = D_2\sigma \\ LC &= d_2\sigma \\ LCI &= d_2\sigma - 3d_3\sigma = D_1\sigma \end{aligned}$$

2. μ, σ DESCONOCIDAS

2.1 GRÁFICO DE MEDIAS (utilizando desviaciones corregidas)

$$
\begin{aligned}
LCS &= \overline{\overline{x}} + 3\frac{\overline{S}}{c_4\sqrt{n}} \\
LC &= \overline{\overline{x}} \\
LCI &= \overline{\overline{x}} - 3\frac{\overline{S}}{c_4\sqrt{n}}
\end{aligned}
$$

2.2 GRÁFICO DE MEDIAS (utilizando rangos)

$$
\begin{aligned}
LCS &= \overline{\overline{x}} + 3\frac{\overline{R}}{d_2\sqrt{n}} \\
LC &= \overline{\overline{x}} \\
LCI &= \overline{\overline{x}} - 3\frac{\overline{R}}{d_2\sqrt{n}}
\end{aligned}
$$

2.3 GRÁFICO DE DESVIACIONES O RANGOS

2.3.1 DESVIACIONES

$$
\begin{aligned}
LCS &= \overline{S} + 3 * \frac{\overline{S}}{c_4}\sqrt{1 - c_4^2} = (1 + 3\frac{\sqrt{1 - c_4^2}}{c_4})\overline{S} = B_4\overline{S} \\
LC &= \overline{S} \\
LCI &= \overline{S} - 3 * \frac{\overline{S}}{c_4} = (1 - 3\frac{\sqrt{1 - c_4^2}}{c_4})\overline{S} = B_3\overline{S}
\end{aligned}
$$

2.3.2 RANGOS

$$
\begin{aligned}
LCS &= \overline{R} + 3\frac{d_3}{d_2}\overline{R} = D_4\overline{R} \\
LC &= \overline{R} \\
LCI &= \overline{R} - 3\frac{d_3}{d_2}\overline{R} = D_3\overline{R}
\end{aligned}
$$

Apéndice B

Tablas para gráficos de control

Observaciones	Gráficos de medias con dispersión basada en			Gráficos para Rangos				Gráficos para deviaciones típicas (corregidas)			
	$\bar{s}$:	$\hat{s}_T$:	$\bar{R}$:								
en la muestra,n	c_2	c_4	d_2	D_1	D_2	D_3	D_4	B_3	B_4	B_5	B_6
2	0.5642	0.7979	1.128	0	3.686	0	3.267	0	3.267	0	2.606
3	0.7236	0.8862	1.693	0	4.358	0	2.575	0	2.568	0	2.276
4	0.7979	0.9213	2.059	0	4.698	0	2.282	0	2.266	0	2.088
5	0.8407	0.9400	2.326	0	4.918	0	2.115	0	2.089	0	1.964
6	0.8686	0.9515	2.534	0	5.078	0	2.004	0.030	1.970	0.029	1.874
7	0.8882	0.9594	2.704	0.204	5.204	0.076	1.924	0.118	1.882	0.113	1.806
8	0.9027	0.9650	2.847	0.388	5.306	0.136	1.864	0.185	1.815	0.179	1.751
9	0.9139	0.9693	2.970	0.547	5.393	0.184	1.816	0.239	1.761	0.232	1.707
10	0.9227	0.9727	3.078	0.687	5.469	0.223	1.777	0.284	1.716	0.276	1.669
11	0.9300	0.9754	3.173	0.811	5.535	0.256	1.744	0.321	1.679	0.313	1.637
12	0.9359	0.9776	3.258	0.922	5.594	0.283	1.717	0.354	1.646	0.346	1.610
13	0.9410	0.9794	3.336	1.025	5.647	0.307	1.693	0.382	1.618	0.374	1.585
14	0.9453	0.9810	3.407	1.118	5.696	0.328	1.672	0.406	1.594	0.399	1.563
15	0.9490	0.9823	3.472	1.203	5.741	0.347	1.653	0.428	1.572	0.421	1.544
16	0.9523	0.9835	3.532	1.282	5.782	0.363	1.637	0.448	1.552	0.440	1.526
17	0.9551	0.9845	3.588	1.356	5.820	0.378	1.622	0.466	1.534	0.458	1.511
18	0.9576	0.9854	3.640	1.424	5.856	0.391	1.608	0.482	1.518	0.475	1.496
19	0.9599	0.9862	3.689	1.487	5.891	0.403	1.597	0.497	1.503	0.490	1.483
20	0.9619	0.9869	3.735	1.549	5.921	0.415	1.585	0.510	1.490	0.504	1.470
21	0.9638	0.9876	3.778	1.605	5.951	0.425	1.575	0.523	1.477	0.516	1.459
22	0.9655	0.9882	3.819	1.659	5.979	0.434	1.566	0.534	1.466	0.528	1.448
23	0.9670	0.9887	3.858	1.710	6.006	0.443	1.557	0.545	1.455	0.539	1.438
24	0.9684	0.9892	3.895	1.759	6.031	0.451	1.548	0.555	1.445	0.549	1.429
25	0.9696	0.9896	3.931	1.806	6.056	0.459	1.541	0.565	1.435	0.559	1.420

ANEXO : INDICES DE CAPACIDAD

Los indices 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 desrecomiendan 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.

1. Índice de Capacidad, C_p

$$C_{\mathrm{p}} = \frac{LTS - LTI}{6\sigma}$$

- Si $C_{\mathrm{p}} < 1 \Rightarrow$ Se producen más del 3 por mil de artículos defectuosos.
- Si $C_{\mathrm{p}} > 1 \Rightarrow$ Se producen menos del 3 por mil de artículos defectuosos.

Cuando el $C_{\mathrm{p}} < 1$ se dice que el proceso no es capaz de cumplir las especificaciones. Actualmente, sin embargo, un índice de capacidad cercano a uno se considera insuficiente. Es frecuente utilizar el valor $C_{\mathrm{p}} = 4/3 \approx 1,33$ como límite inferior de la calidad que debe tenerse en la práctica. Esto implica que

$$\frac{LTS - LTI}{6\sigma} = \frac{4}{3} \Rightarrow LTS - LTI = 8\sigma$$

por tanto, serían defectuosos aquellos artículos que estén a más de 4σ de la medía; esto es aproximadamente , 64 piezas por millon (bajo normalidad). Por esta razón se dice que:

- Si $C_{\mathrm{p}} < 1 \Rightarrow$ el proceso no es capaz.
- Si $C_{\mathrm{p}} > 1.33 \Rightarrow$el proceso es capaz
- Si $1 \leq C_{\mathrm{p}} \leq 1.33 \Rightarrow$el proceso es capaz pero requiere un seguimiento muy estricto

2. Índices de capacida unilaterales C_{pL} y C_{pU}

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 C_{pL} cuando existe una especificación mínima, pero no máxima, mientras que C_{pU} se usa cuando se debe cumplir con una especificación máxima.

$$C_{\mathrm{pL}} = \frac{\mu_{\mathrm{N}} - LTI}{3\sigma}$$

$$C_{\mathsf{pU}} = \frac{LTI - \mu_{\mathsf{N}}}{3\sigma}$$

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 media muestral $\overline{x}$.

3. Índice de capacidad recíproco C_r

Este índice es el invero de C_{p},es decir, $C_{\mathsf{r}} = 1/C_{\mathsf{p}}$. Su interpretación es la siguiente:

- Si $C_{\mathsf{r}} > 1$, el proceso no es capaz
- Si $C_{\mathsf{r}} < 0.75$, el proceso es capaz
- Si $0.75 \leq C_{\mathsf{r}} \leq 1$, el proceso es capaz pero precisa de un control estricto

4. Coeficiente K

Este coeficiente es una medida de descentrlización del proceso. Se define como

$$K = \frac{\overline{x} - \mu_{\mathsf{N}}}{\frac{1}{2}(LTS - LTI)}$$

- Si $K > 0$ el proceso tiene un sesgo hacia valores superiores al nominal (sesgo positivo)
- Si $K < 0$ el proceso tiene un sesgo hacia valores inferiores al nominal (sesgo negativo)

5. Coeficiente n-sigma

Este coeficiente consiste en la expresión de los límites de tolerancia en función de la desviación típica del proceso de forma $LTS - LTI = 2n\sigma$. Esta definición equivale a

$$\begin{aligned} LTS &= \mu + n\sigma \\ LTI &= \mu - n\sigma \end{aligned}$$

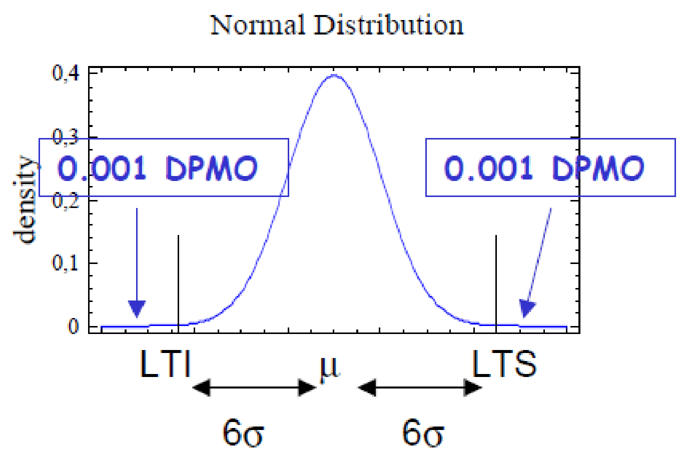
Por tanto, un índice de capacidad $C_{\mathsf{p}} = 1$ equivale a $3 - sigma$ y $C_{\mathsf{p}} = 2$ equivale a $6 - sigma$.

Es frecuente interpretar el nivel de calidad n-sigma en terminos de número de artículos defectuosos por millon. Para ello hay dos posibles alternativas:

1. Consiste en emplear estrictamente su significado (suponiendo normalidad). Es decir, interpretarlo como el nº de defectuosos que se producen por millón de oportunidades.
2. Otra interpretación popularizada por la empresa motorola, consiste en interpretarlo como el nº de DPMO cuando se produce un desajuste ne la media de $\pm 1,5\sigma$

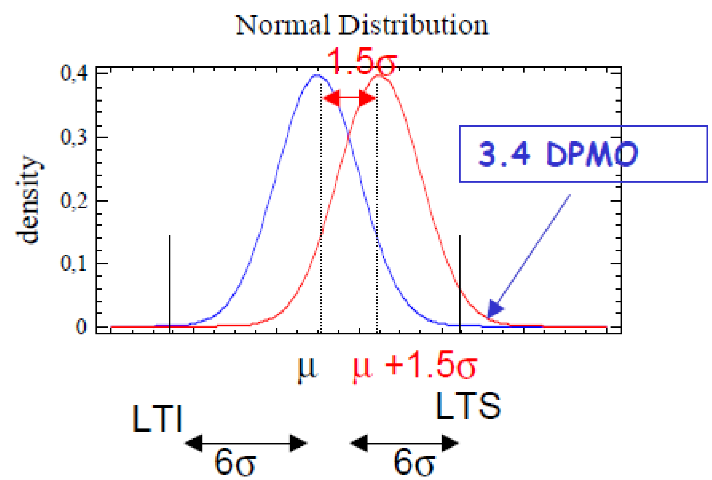
1ª Interpretación

Defectos por millón de oportunidades



2ª Interpretación

Defectos por millón de oportunidades, cuando se produce un cambio en la media de 1,5 σ



Nivel σ	DPMO (1º interpretación)	DPMO (2º Interpretación)
2	45500	308537
3	2700	66807
4	63	6210
5	0.57	233
6	0.002	3.4