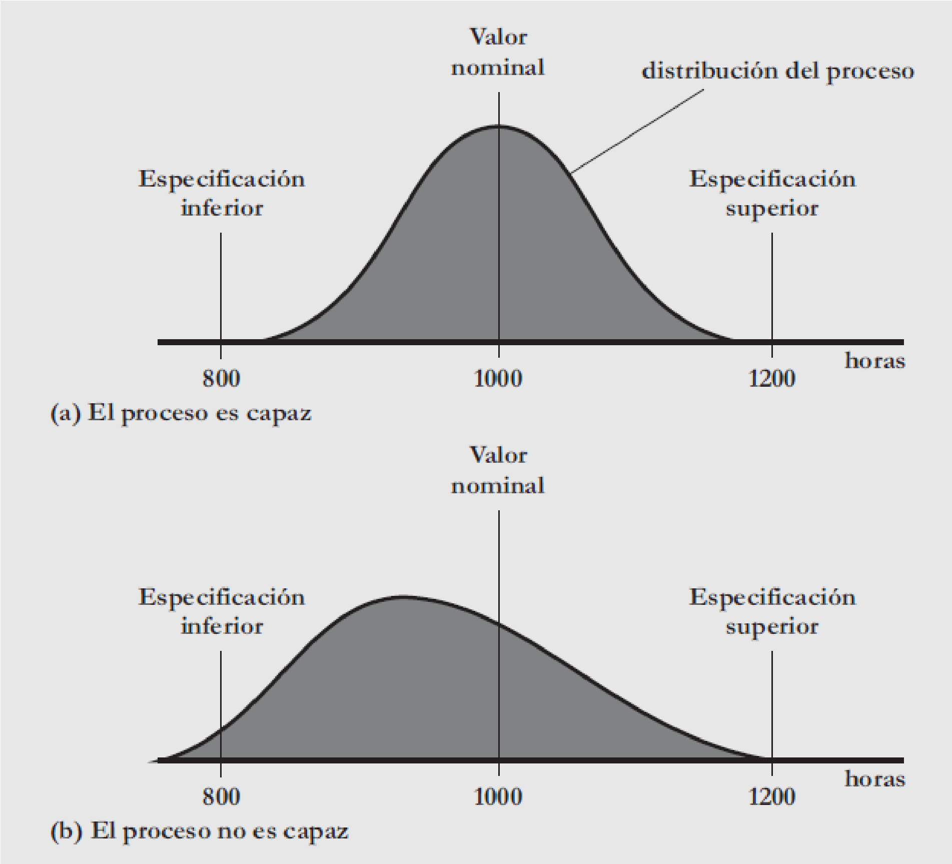
Control Estadístico de Calidad

2022.

Control Estadístico de Calidad

En la Planta Industrial de cualquier empresa todo el tiempo se están ejecutando procesos que podemos “medir” para poder conocer si cumplen con los requisitos del producto.

Diagrama de Pareto



Este grafico se refiere a las especificaciones de diseño referentes a la vida útil de una lampara de bajo consumo con un valor nominal de 1000 hs y una tolerancia de +/-200 hs. Esta tolerancia arroja una especificación no superior a 1200 hs. y una especificación no inferir a 800hs. El proceso de producción debe ser capaz de producirlas dentro de esas especificaciones de diseño.

Figura a) El proceso es capaz. Figura b) el proceso no es capaz, porque produce demasiadas lámparas de corta vida útil.

Control Estadístico de Calidad

En la Planta Industrial de cualquier empresa todo el tiempo se están ejecutando procesos que podemos “medir” para poder conocer si cumplen con los requisitos del producto.

Todo proceso está afectado por un gran número de factores sometidos a variabilidad (oscilaciones de las características del material utilizado, variaciones de temperatura y humedad ambiental, variabilidad introducida por el operario, entre otras) que inciden en el proceso y que inducen una variabilidad de las características del producto fabricado.

Si el proceso está operando de manera que existen pequeñas oscilaciones de todos estos factores, de modo tal que ninguno de ellos tiene un efecto preponderante frente a los demás, entonces en virtud del Teorema del Límite Central (TLC) es esperable que la característica de calidad del producto fabricado se distribuya de acuerdo con una ley normal. Al conjunto de esta multitud de factores se denominan **causas comunes o aleatorias**.

Por el contrario, si circunstancialmente incide un factor con un efecto preponderante, se dice que está presente una **causa especial o asignable**. Por ejemplo, si en un proceso industrial se está utilizando materias primas procedentes de un lote homogéneo y se continúa la fabricación con materias primas procedentes de otro lote, cuyas características son muy diferentes de las anteriores.

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

La herramienta estadística que permite evaluar la variabilidad de un proceso es el Control Estadístico de Procesos. Comprende un conjunto de herramientas tales como:

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

La herramienta estadística que permite evaluar la variabilidad de un proceso es el Control Estadístico de Procesos. Comprende un conjunto de herramientas tales como:

histogramas de frecuencia,

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

La herramienta estadística que permite evaluar la variabilidad de un proceso es el Control Estadístico de Procesos. Comprende un conjunto de herramientas tales como:

histogramas de frecuencia,
gráfico de Pareto,

Control Estadístico de Calidad

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

La herramienta estadística que permite evaluar la variabilidad de un proceso es el Control Estadístico de Procesos. Comprende un conjunto de herramientas tales como:

histogramas de frecuencia,
gráfico de Pareto,
diagrama causa-efecto,

Por definición, se dice que un proceso está bajo control estadístico cuando NO hay presente causas asignables.

Los procesos que operan bajo **causas asignables** se consideran **fuera de control estadístico** y se requiere de un sistema de control que mitigue el impacto de las causas verdaderamente significativas de una manera práctica y económicamente viable.

La herramienta estadística que permite evaluar la variabilidad de un proceso es el Control Estadístico de Procesos. Comprende un conjunto de herramientas tales como:

histogramas de frecuencia,
gráfico de Pareto,
diagrama causa-efecto,

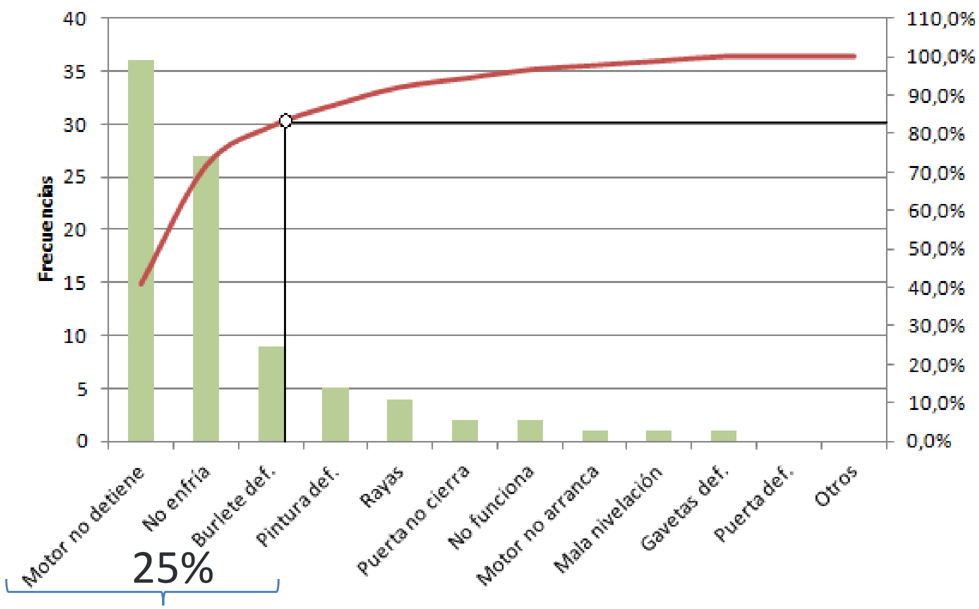
Gráficos de control.

Diagrama de Pareto

Un diagrama de Pareto es una técnica que permite clasificar gráficamente la información de mayor a menor relevancia, con el objetivo de reconocer los problemas más importantes en los que se debería enfocar y por lo tanto solucionar. Es decir permite representa en forma ordenada el grado de importancia que tienen los diferentes factores en un determinado problema, tomando en consideración la frecuencia con que ocurre cada uno de dichos factores.

El diagrama de Pareto es una variación del histograma tradicional, puesto que en el Diagrama de Pareto se ordenan los datos por su frecuencia de mayor a menor.

Esta técnica se basa en el principio de Pareto o regla 80-20, donde el 80 % de las consecuencias provienen del 20 % de las causas.



En éste caso el 81,8% de las consecuencias (defectos del proceso) corresponden al 25% de las causas (los tipos de defectos), es decir que tan solo solucionando las 3 principales inconformidades se solucionarían el 81,8% de unidades defectuosas.

Diagrama de Pareto

Supongamos que un proceso que produce Heladeras desea establecer controles sobre los defectos que aparecen en las unidades que salen como producto terminado en la línea de producción. Para ello se hace imperativo determinar cuáles son los defectos más frecuentes.

En primer lugar se clasifican todos los defectos posibles:

- Motor no detiene
- No enfría
- Burlete def.
- Pintura def.
- Rayas
- No funciona
- Puerta no cierra
- Gavetas def.
- Motor no arranca
- Mala nivelación
- Puerta def.
- Otros

Diagrama de Pareto

Supongamos que un proceso que produce Heladeras desea establecer controles sobre los defectos que aparecen en las unidades que salen como producto terminado en la línea de producción. Para ello se hace imperativo determinar cuáles son los defectos más frecuentes.

En primer lugar se clasifican todos los defectos posibles:

- Motor no detiene
- No enfría
- Burlete def.
- Pintura def.
- Rayas
- No funciona
- Puerta no cierra
- Gavetas def.
- Motor no arranca
- Mala nivelación
- Puerta def.
- Otros

Después de inspeccionar 88 Heladeras defectuosas, se obtuvo la siguiente tabla de frecuencias:

Tipo de defecto	N°
Burlete def.	9
Pintura def.	5
Gavetas def.	1
Mala nivelación	1
Motor no arranca	1
Motor no detiene	36
No enfría	27
No funciona	2
Otros	0
Puerta def.	0
Puerta no cierra	2
Rayas	4
Total	**88**

Diagrama de Pareto

Ordenamos los datos y anexamos una columna de frecuencias y otra de frecuencias acumuladas:

Tipo de defecto	N°	Frecuencia	Frecuencia Acumulada
Motor no detiene	36	40,9%	40,9%
No enfría	27	30,7%	71,6%
Burlete def.	9	10,2%	81,8%
Pintura def.	5	5,7%	87,5%
Rayas	4	4,5%	92,0%
Puerta no cierra	2	2,3%	94,3%
No funciona	2	2,3%	96,6%
Motor no arranca	1	1,1%	97,7%
Mala nivelación	1	1,1%	98,9%
Gavetas def.	1	1,1%	100,0%
Puerta def.	0	0,0%	100,0%
Otros	0	0,0%	100,0%
Total	**88**	**100,0%**	

Diagrama de Pareto

Lo que obtenemos es lo que se conoce como Diagrama de Pareto:

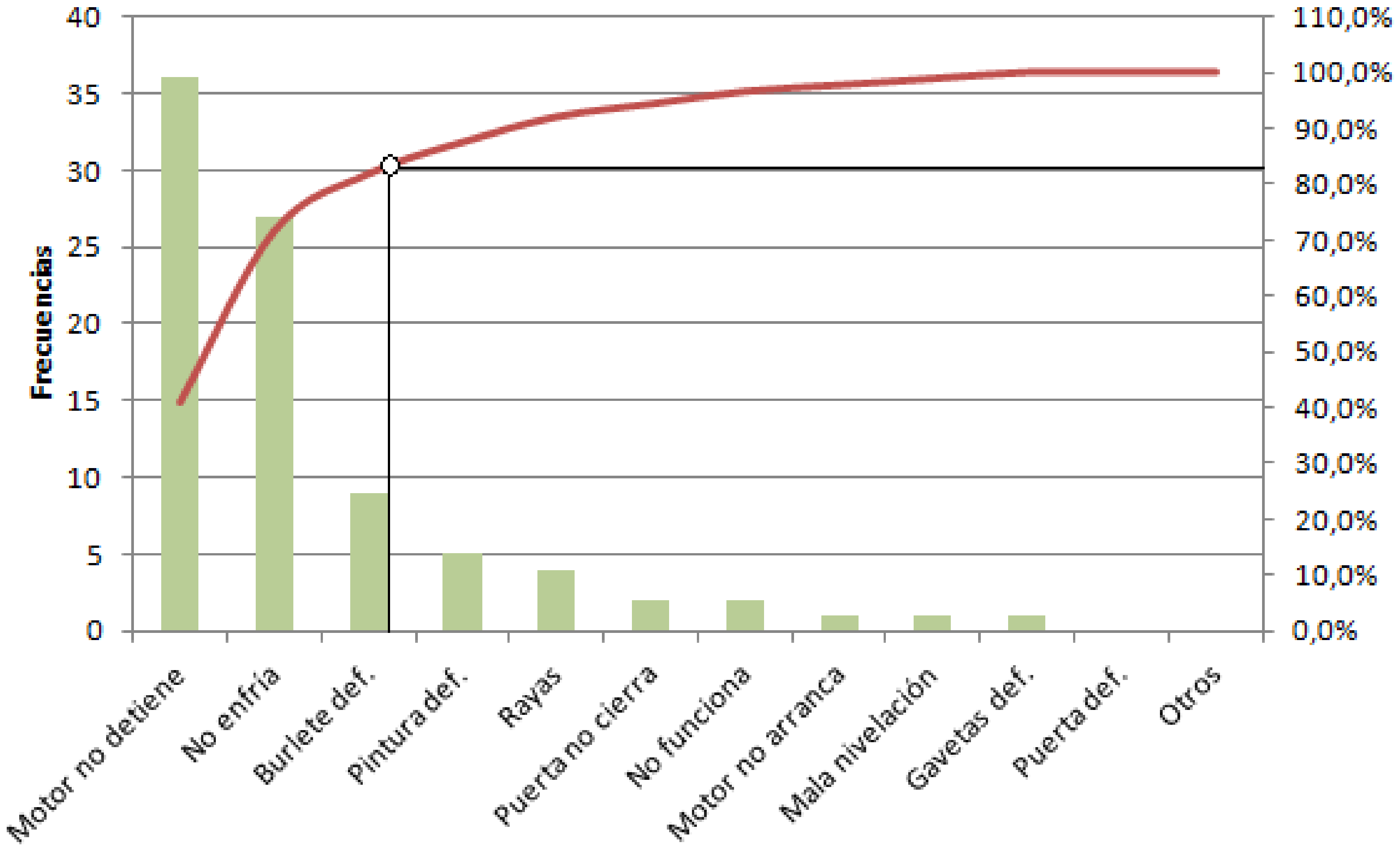
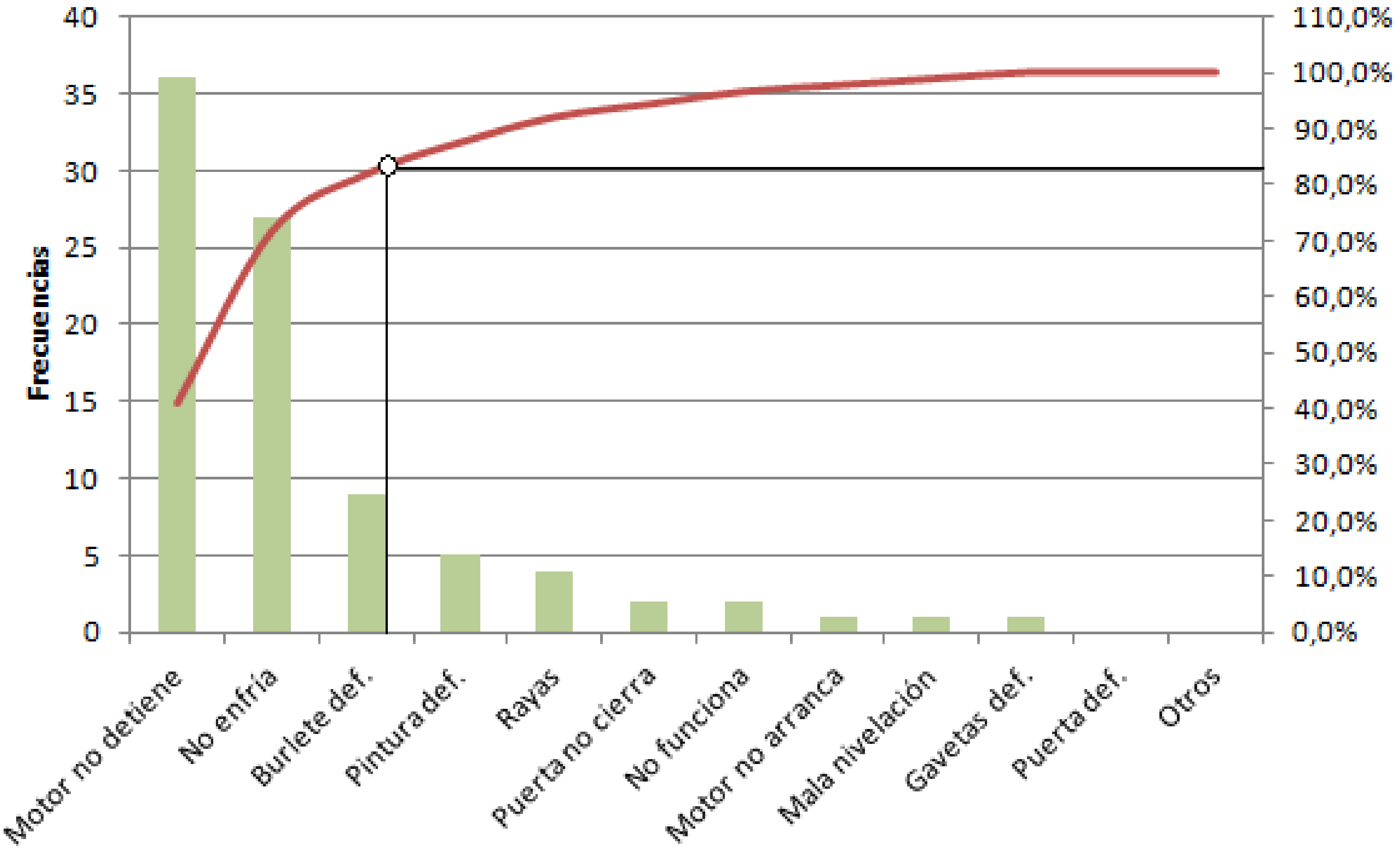


Diagrama de Pareto

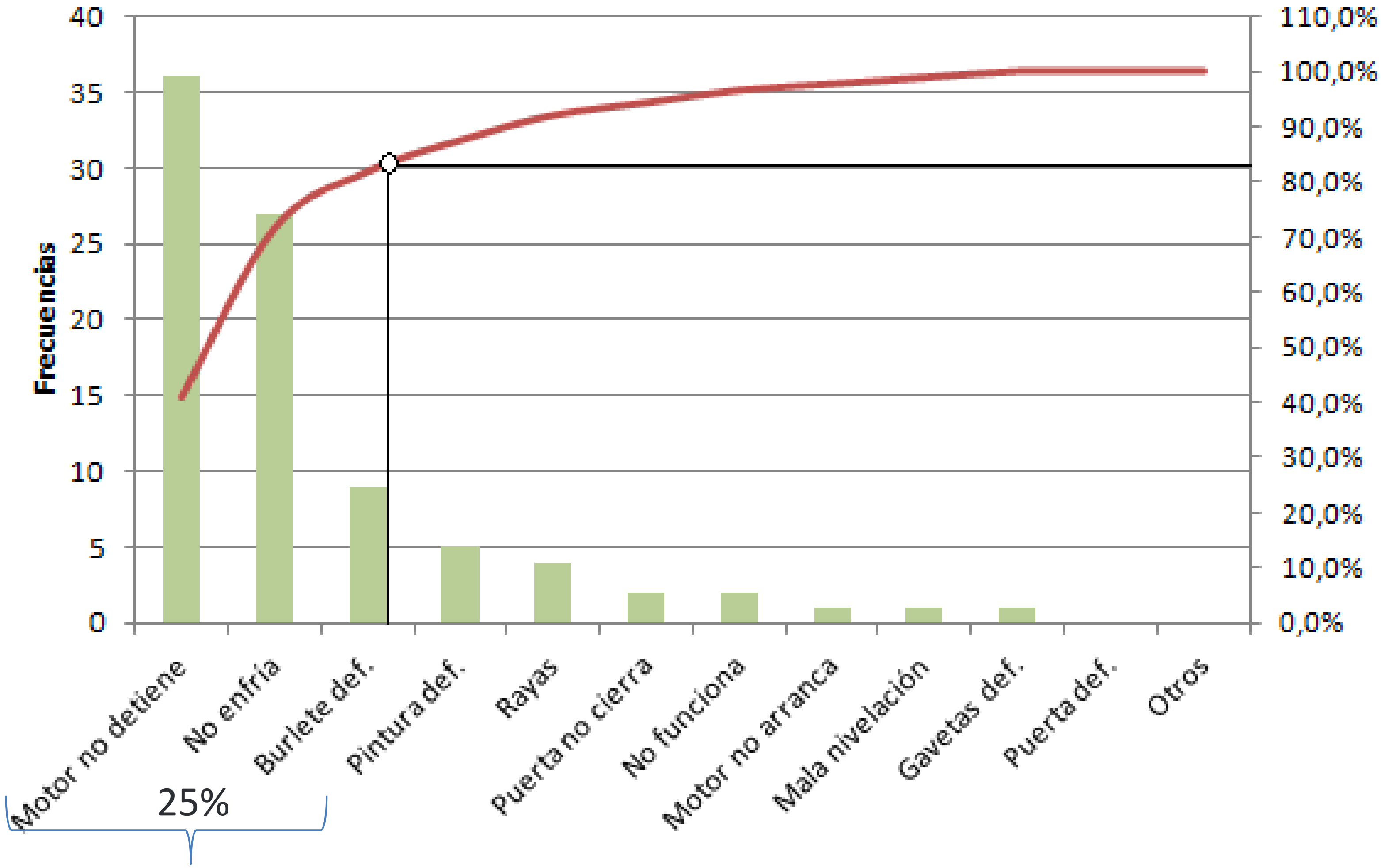
Lo que obtenemos es lo que se conoce como Diagrama de Pareto:



En éste caso el 81,8% de los defectos del proceso corresponden al 25% de los tipos de defectos, es decir que tan solo solucionando las 3 principales inconformidades se solucionarían el 81,8% de unidades defectuosas.

Diagrama de Pareto

Lo que obtenemos es lo que se conoce como Diagrama de Pareto:

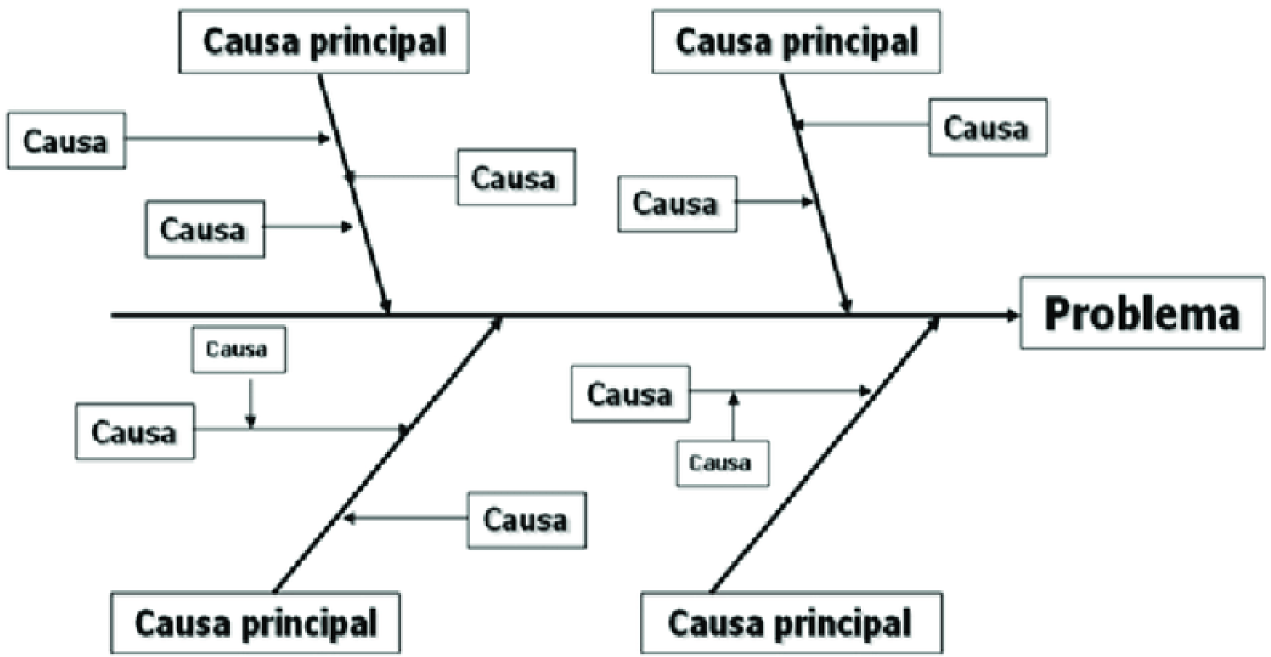


En éste caso el 81,8% de los defectos del proceso corresponden al 25% de los tipos de defectos, es decir que tan solo solucionando las 3 principales inconformidades se solucionarían el 81,8% de unidades defectuosas.

DIAGRAMAS DE CAUSA – EFECTO

La variabilidad de una característica de calidad es una consecuencia de múltiples causas, por ello es importante detallar las posibles causas de la inconsistencia.

Un diagrama de causa-efecto es una herramienta visual que se utiliza para organizar de forma lógica las posibles causas de un problema o efecto específico, mostrándolas gráficamente de forma cada vez más detallada, sugiriendo relaciones causales entre las distintas hipótesis.



DIAGRAMAS DE CAUSA – EFECTO

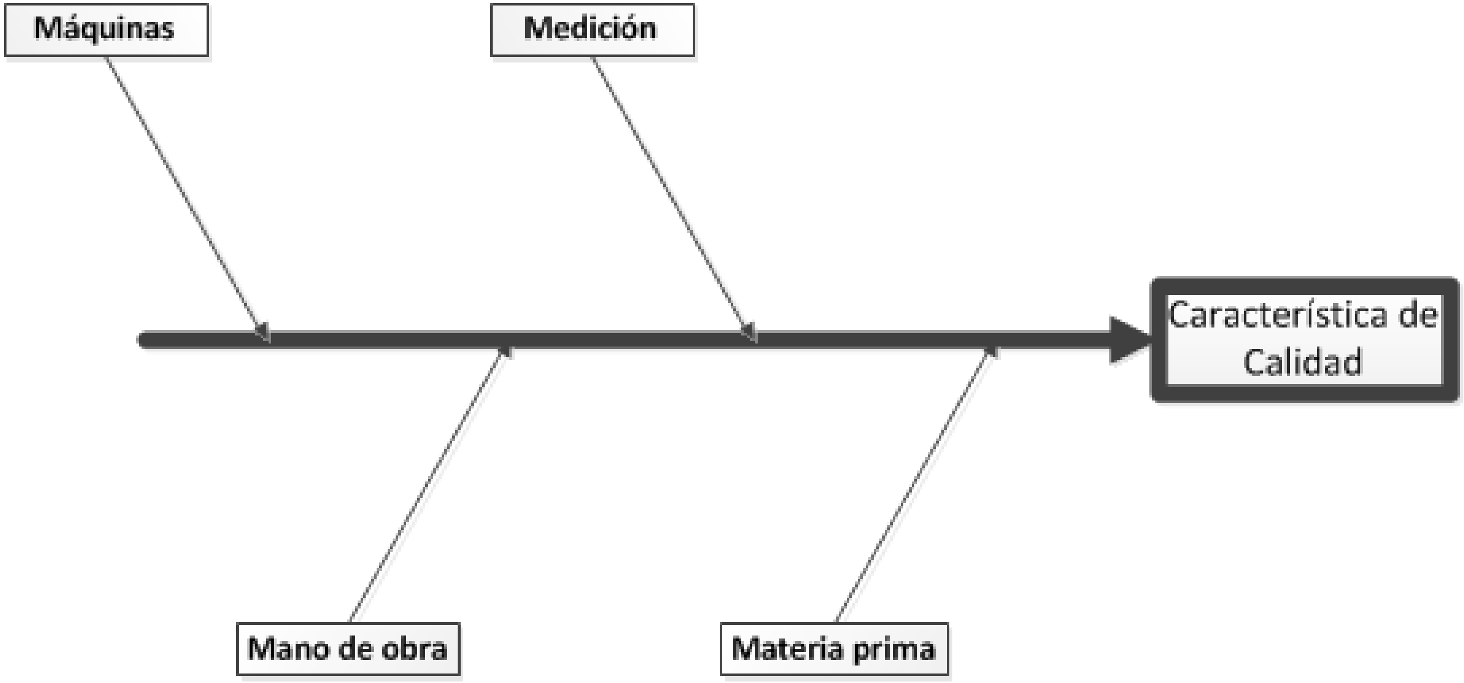
Para hacer un diagrama de causa – efecto se recomienda seguir los siguientes pasos

1. **Elegir la característica de calidad que se va a analizar**: Por ejemplo, en la producción de frascos de mermelada, la característica podría ser el peso del frasco lleno, la densidad del producto, etc. Trazamos una flecha horizontal gruesa en sentido izquierda a derecha, que representa el proceso y a la derecha de ésta escribimos la característica de calidad.

DIAGRAMAS DE CAUSA – EFECTO

Para hacer un diagrama de causa – efecto se recomienda seguir los siguientes pasos

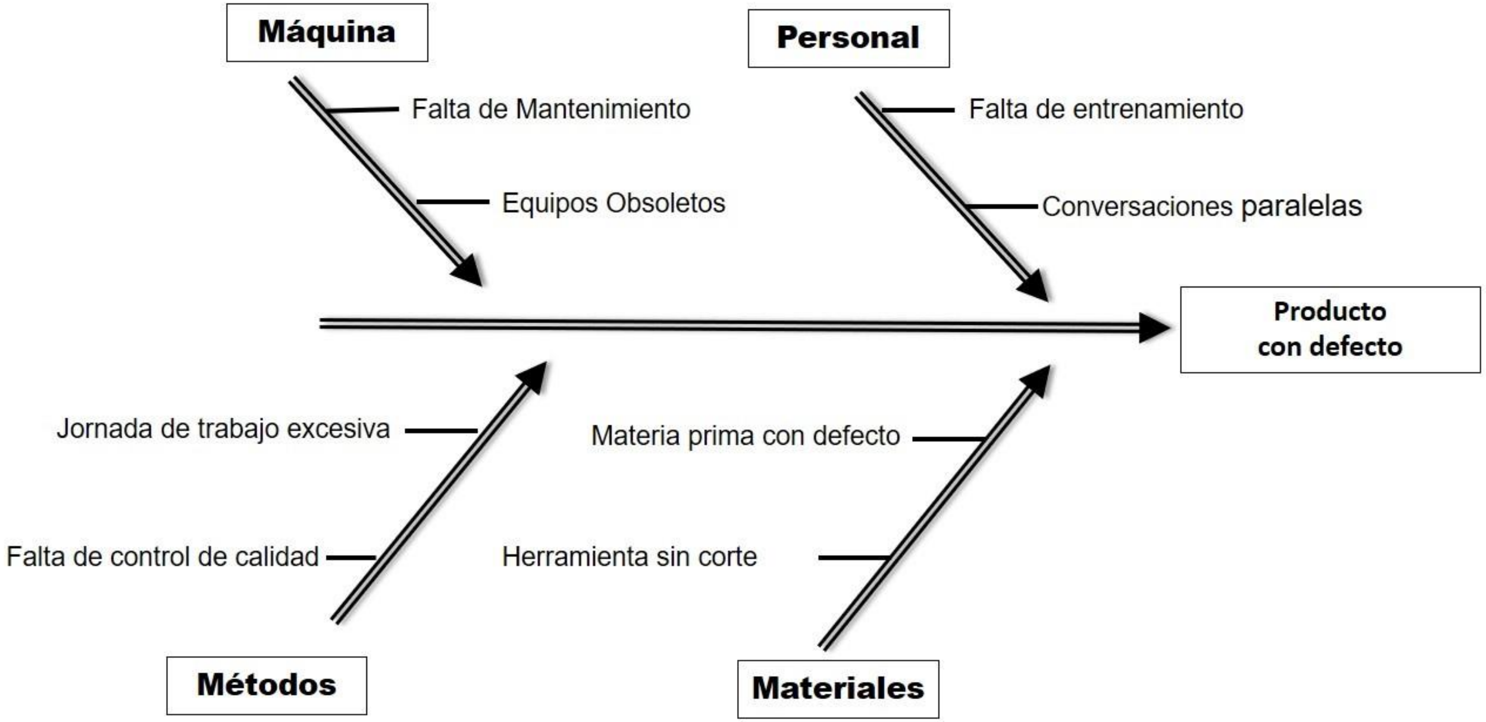
2. Indicamos los factores causales más importantes que puedan generar la fluctuación de la característica de calidad: Trazamos flechas secundarias diagonales en dirección de la flecha principal. Usualmente estos factores causales se ven representados en Materias primas, Máquinas, Mano de obra, Métodos de medición, etc.



DIAGRAMAS DE CAUSA – EFECTO

Para hacer un diagrama de causa – efecto se recomienda seguir los siguientes pasos

3. Anexamos en cada rama factores causales más detallados de la fluctuación de la característica de calidad: Para simplificar ésta labor podemos recurrir a la técnica del interrogatorio. De ésta forma seguimos ampliando el diagrama hasta asegurarnos de que contenga todas las posibles causas de dispersión.



DIAGRAMAS DE CAUSA – EFECTO

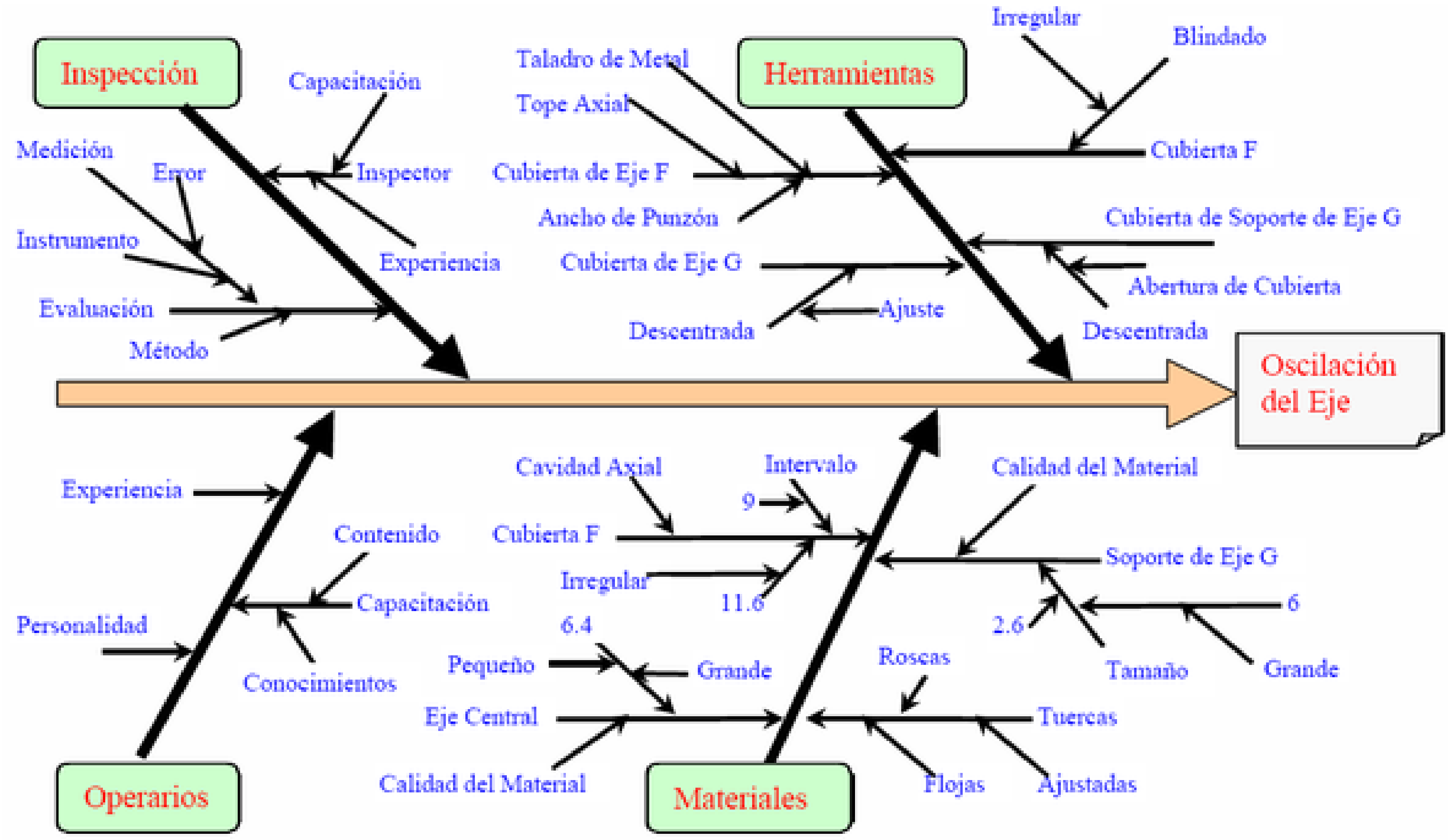
Para hacer un diagrama de causa – efecto se recomienda seguir los siguientes pasos

4. **Verificamos que todos los factores causales de dispersión hayan sido anexados al diagrama**: Una vez establecidas de manera clara las relaciones causa y efecto, el diagrama estará terminado.

DIAGRAMAS DE CAUSA – EFECTO

Para hacer un diagrama de causa – efecto se recomienda seguir los siguientes pasos

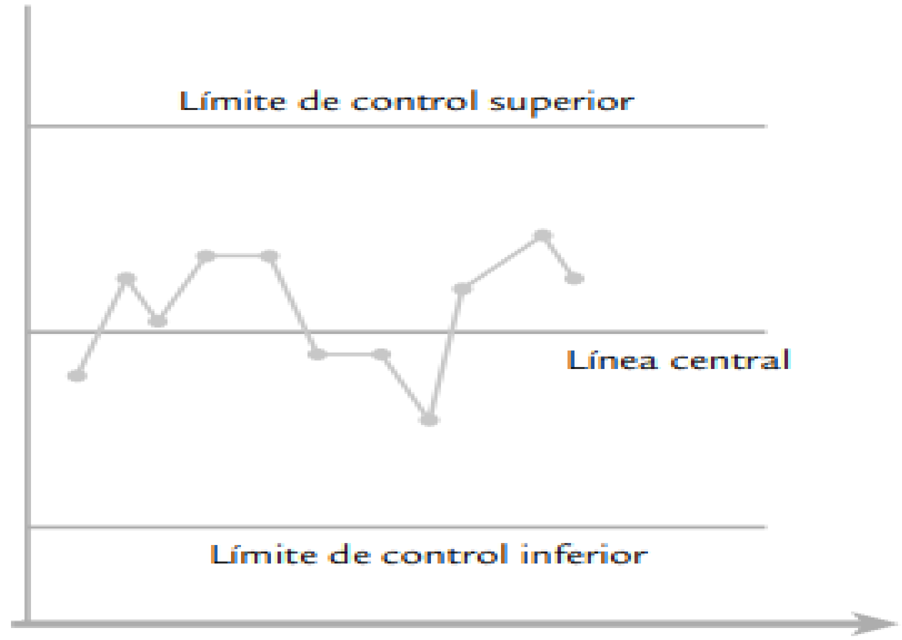
El siguiente gráfico corresponde a un ejemplo de diagrama de causa – efecto. El proceso corresponde a una máquina en la que se observa un defecto de rotación oscilante, la característica de calidad es la oscilación de un eje durante la rotación:



Gráficos de Control

Gráficos de control. La importancia del gráfico de control radica en su capacidad para **detectar causas asignables durante un proceso de fabricación**.

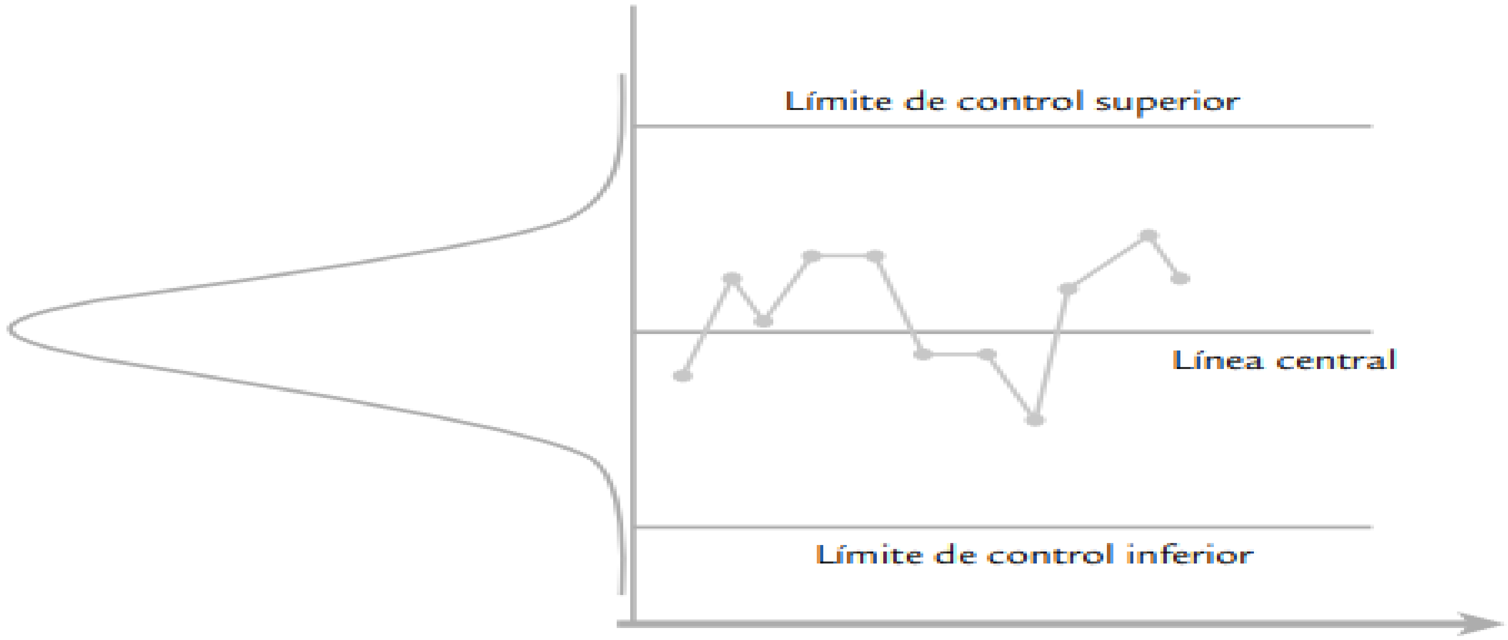
Un gráfico de control consiste de una línea central que representa el valor promedio de la característica de calidad correspondiente al estado bajo control y dos líneas que representan los límites de control inferior y superior.



Gráficos de Control

Gráficos de control. La importancia del gráfico de control radica en su capacidad para **detectar causas asignables durante un proceso de fabricación**.

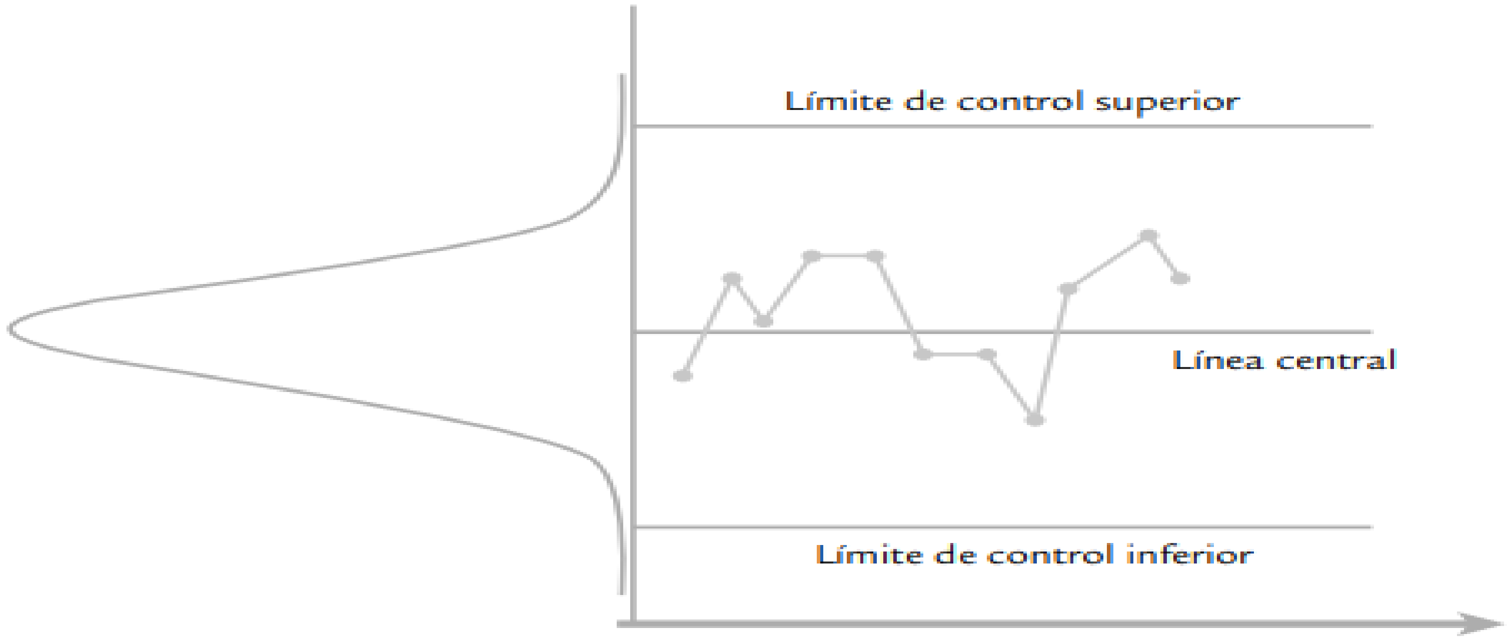
Un gráfico de control consiste de una línea central que representa el valor promedio de la característica de calidad correspondiente al estado bajo control y dos líneas que representan los límites de control inferior y superior.



Gráficos de Control

Gráficos de control. La importancia del gráfico de control radica en su capacidad para **detectar causas asignables durante un proceso de fabricación**.

Un gráfico de control consiste de una línea central que representa el valor promedio de la característica de calidad correspondiente al estado bajo control y dos líneas que representan los límites de control inferior y superior.



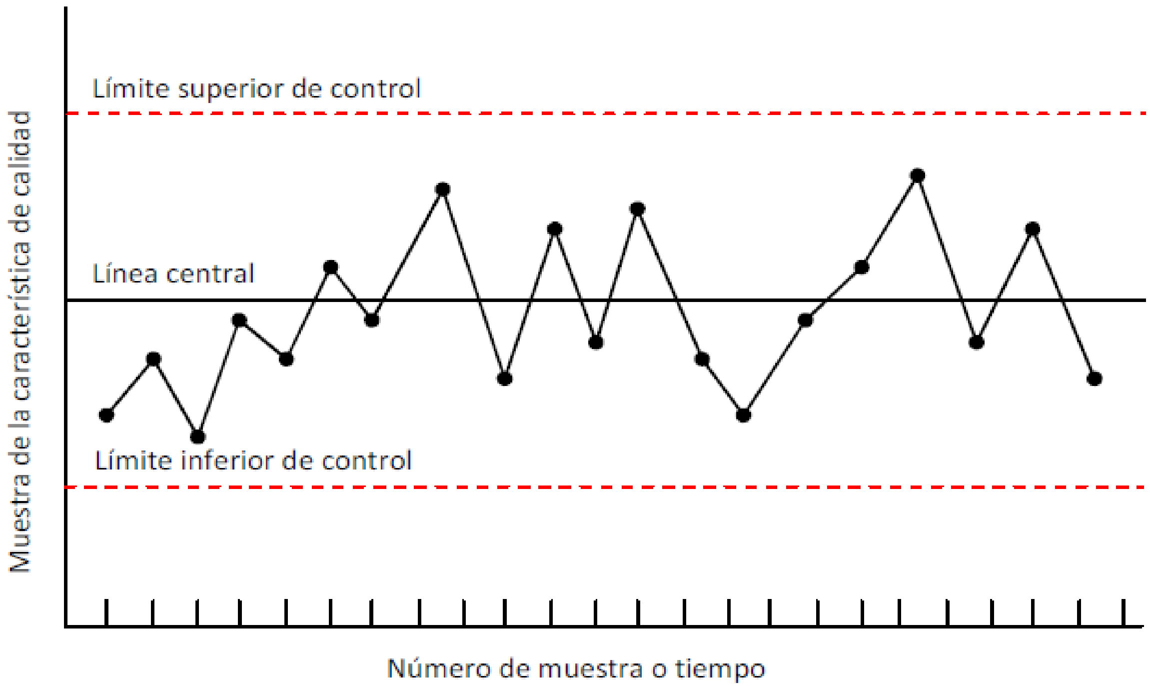
LCS = promedio de proceso **+** 3 desviaciones estándar

LCI = promedio de proceso **-** 3 desviaciones estándar

Gráficos de Control

Gráficos de control. La importancia del gráfico de control radica en su capacidad para **detectar causas asignables durante un proceso de fabricación**.

Un gráfico de control consiste de una línea central que representa el valor promedio de la característica de calidad correspondiente al estado bajo control y dos líneas que representan los límites de control inferior y superior.



GRÁFICOS DE CONTROL

Los gráficos de control pueden clasificarse en dos categorías, según el tipo de variable o característica de calidad que se desee monitorear:

GRÁFICOS DE CONTROL

Los gráficos de control pueden clasificarse en dos categorías, según el tipo de variable o característica de calidad que se desee monitorear:

1) **Gráficos de control para variables.**

GRÁFICOS DE CONTROL

Los gráficos de control pueden clasificarse en dos categorías, según el tipo de variable o característica de calidad que se desee monitorear:

1) **Gráficos de control para variables.**

2) **Gráficos de control por atributos.**

GRÁFICOS DE CONTROL

Los gráficos de control pueden clasificarse en dos categorías, según el tipo de variable o característica de calidad que se desee monitorear:

1) Gráficos de control para variables: Este tipo de gráficos se emplea cuando la característica de calidad puede medirse y expresarse como un número en alguna escala continua de medición, por ejemplo, el diámetro de un objeto, la longitud de una piza, el peso de un producto, etc.

GRÁFICOS DE CONTROL

1) Gráficos de control para variables más usadas en procesos industriales son los siguientes:

a) $\bar{X} - R$ .Se utilizan para controlar y analizar un proceso, empleando valores continuos de calidad del producto.

$\bar{X}$ es la media aritmética de los valores en subgrupos pequeños (una medida del promedio del proceso); R es el rango de los valores dentro de cada subgrupo (una medida de la variación del proceso). Las gráficas $\bar{X} - R$ son las gráficas más comunes, aunque pueden no ser las más apropiadas para todas las situaciones.

GRÁFICOS DE CONTROL

1) Gráficos de control para variables más usadas en procesos industriales son los siguientes:

a) $\bar{X} - R$ .Se utilizan para controlar y analizar un proceso, empleando valores continuos de calidad del producto.

$\bar{X}$ es la media aritmética de los valores en subgrupos pequeños (una medida del promedio del proceso); R es el rango de los valores dentro de cada subgrupo (una medida de la variación del proceso). Las gráficas $\bar{X} - R$ son las gráficas más comunes, aunque pueden no ser las más apropiadas para todas las situaciones.

b) Gáfico x

GRÁFICOS DE CONTROL

2) Gráficos de control para atributos más usadas en procesos industriales son los siguientes:

Basados en la distribución Binomial:

- **Gráfico np**: número de unidades defectuosas en una muestra de n artículos para N muestras producidas.
- **Gráfico p**: proporción de unidades defectuosas en una muestra de n artículos para N muestras producidas.

Basados en la distribución de Poisson:

- **Gráfico c**: Número de defectos por unidad para N muestras de igual tamaño (n).
- **Gráfico u**: Número de defectos por unidad para N muestras de distintos tamaños.

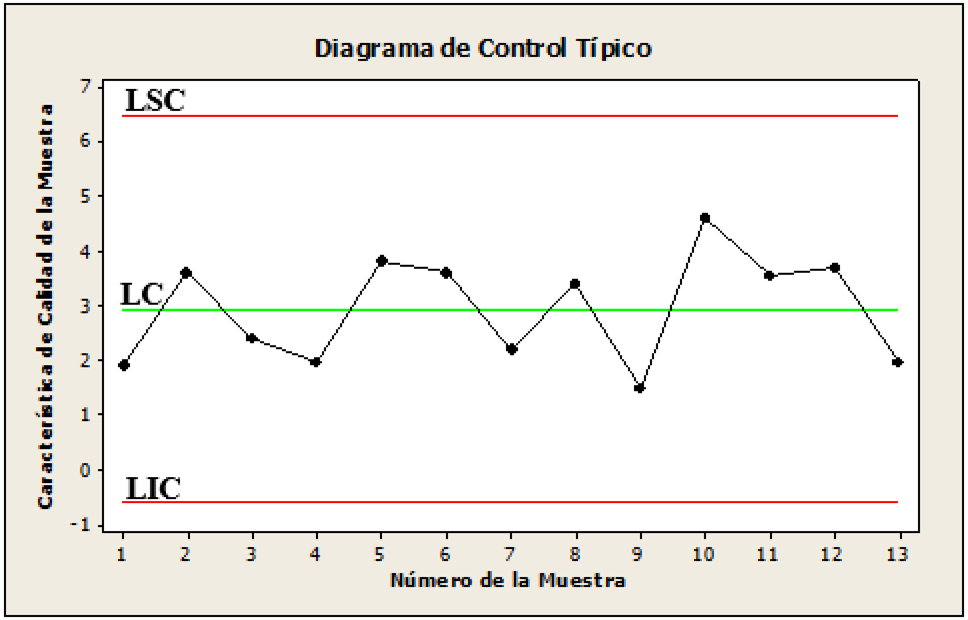
Gráficos de Control

La forma típica de un gráfico de control establece límites de control que se encuentran dentro de $\pm 3\sigma$ o sea dentro de ±3 desviaciones estándar de la medida estadística de interés (puede ser el promedio, la porción, etc.). En general puede establecerse como:

<promedio de proceso> ±<3 desviaciones estándar>

LCS = promedio de proceso **+** 3 desviaciones estándar

LCI = promedio de proceso **-** 3 desviaciones estándar



Gráficos de Control

GRÁFICO	LÍNEA CENTRAL = LC	LÍNEA CONTROL SUP. = LCS	LÍNEA CONTROL INF. = LCI
$\bar{x}$	$\bar{\bar{x}}$	$\bar{\bar{x}} + A_2\bar{R}$	$\bar{\bar{x}} - A_2\bar{R}$
R	$\bar{R}$	$D_4\bar{R}$	$D_3\bar{R}$
x	$\bar{x}$	$\bar{x} + 2{,}66\bar{R}_s$	$\bar{x} - 2{,}66\bar{R}_s$
p	$\bar{p}$	$\bar{p} + 3\sqrt{\bar{p}(1-\bar{p})/\bar{n}}$	$\bar{p} - 3\sqrt{\bar{p}(1-\bar{p})/\bar{n}}$
np	$n\bar{p}$	$n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})}$	$n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})}$
c	$\bar{c}$	$\bar{c} + 3\sqrt{\bar{c}}$	$\bar{c} - 3\sqrt{\bar{c}}$
u	$\bar{u}$	$\bar{u} + 3\sqrt{\bar{u}/\bar{n}}$	$\bar{u} - 3\sqrt{\bar{u}/\bar{n}}$

Gráficos de Control $\bar{X} - R$

	Factor for Control Limits						
	$\bar{X}$ Chart			R Chart		S Chart	
n*	A_1	A_2	d_2	D_3	D_4	c_4	n
2	3.760	1.880	1.128	0	3.267	0.7979	2
3	2.394	1.023	1.693	0	2.575	0.8862	3
4	1.880	.729	2.059	0	2.282	0.9213	4
5	1.596	.577	2.326	0	2.115	0.9400	5
6	1.410	.483	2.534	0	2.004	0.9515	6
7	1.277	.419	2.704	.076	1.924	0.9594	7
8	1.175	.373	2.847	.136	1.864	0.9650	8
9	1.094	.337	2.970	.184	1.816	0.9693	9
10	1.028	.308	3.078	.223	1.777	0.9727	10
11	.973	.285	3.173	.256	1.744	0.9754	11
12	.925	.266	3.258	.284	1.716	0.9776	12
13	.884	.249	3.336	.308	1.692	0.9794	13
14	.848	.235	3.407	.329	1.671	0.9810	14
15	.816	.223	3.472	.348	1.652	0.9823	15
16	.788	.212	3.532	.364	1.636	0.9835	16
17	.762	.203	3.588	.379	1.621	0.9845	17
18	.738	.194	3.640	.392	1.608	0.9854	18
19	.717	.187	3.689	.404	1.596	0.9862	19
20	.697	.180	3.735	.414	1.586	0.9869	20
21	.679	.173	3.778	.425	1.575	0.9876	21
22	.662	.167	3.819	.434	1.566	0.9882	22
23	.647	.162	3.858	.443	1.557	0.9887	23
24	.632	.157	3.895	.452	1.548	0.9892	24
25	.619	.153	3.931	.459	1.541	0.9896	25

*$n > 25$: $A_1 = 3/\sqrt{n}$ where n = number of observations in sample.

n tamaños de cada muestra

GRÁFICOS DE CONTROL

2) Gráficos de control para atributos más usadas en procesos industriales son los siguientes:

Basados en la distribución Binomial:

• **Gráfico np**: número de unidades defectuosas en una muestra de n artículos para N muestras producidas.
• **Gráfico p**: proporción de unidades defectuosas en una muestra de n artículos para N muestras producidas.

Basados en la distribución de Poisson:

• **Gráfico c**: Número de defectos por unidad para N muestras de igual tamaño (n).
• **Gráfico u**: Número de defectos por unidad para N muestras de distintos tamaños.

GRÁFICOS DE CONTROL POR ATRIBUTOS

a) Gráfico ***p*** (fracción o proporción defectuosa o no concordante)

A menudo es deseable **clasificar un producto como defectuoso o no defectuoso** sobre la base de la comparación con un estándar. Por ejemplo, el diámetro de una bola puede verificarse determinando si pasará a través de un medidor compuesto por agujeros circulares cortados en una plantilla. Esto sería mucho más simple que medir el diámetro con un micrómetro. Sin embargo, **los diagramas de control de atributo requieren un tamaño de muestra bastante más grande que en el caso de las mediciones de contrapartes.**

El gráfico se realiza tomando muestras de tamaño n_i (no tienen por qué ser todas de igual tamaño) y contando el número de artículos defectuosos d_i . Nuestro interés está en la evolución de la proporción de artículos defectuoso, es decir,

$$\widehat{p}_i = \frac{d_i}{n_i} = \frac{\text{número de defectuosos en la muestra i-ésima}}{\text{tamaño de la muestra i-ésima}}$$

GRÁFICOS DE CONTROL POR ATRIBUTOS

a) Gráfico ***p*** (fracción o proporción defectuosa o no concordante)

Como en gráficos anteriores, para el cálculo de los límites de control se utilizan las siguientes fórmulas:

$$\bar{p} = \frac{\sum_{i=1}^{N}\left(\frac{d_i}{n_i}\right)}{N} \quad y \quad \bar{n} = \frac{\sum_{i=1}^{N} n_i}{N}$$

$$d_i = \textit{cantidad de productos defectuosos de la muestra i de tamaño } n_i,$$

$$n_i = \textit{número de productos de la muestra i} \quad y \quad N = \textit{número de muestras}$$

$$\boxed{\begin{cases} LSC = \bar{p} + 3\sqrt{\dfrac{\bar{p}(1-\bar{p})}{\bar{n}}} \\ LC = \bar{p} \\ LIC = \bar{p} - 3\sqrt{\dfrac{\bar{p}(1-\bar{p})}{\bar{n}}} \end{cases}}$$

Gráficos de Control np

b) Gráfico ***np*** (tamaño de la maestra constante)

Número de unidades defectuosas en una muestra de n artículos para N muestras producidas. Estas muestras se seleccionan periódicamente en el proceso de producción, por ejemplo: cada hora, cada 15 minutos, cada mañana, cada turno, etc.

Los límites de control se calculan mediante:

$$NP \Rightarrow \begin{cases} LSC = n\overline{p} + 3n\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}} = n\overline{p} + 3\sqrt{n\overline{p}(1-\overline{p})} \\ LC = n\overline{p} \\ LIC = n\overline{p} - 3n\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}} = n\overline{p} - 3\sqrt{n\overline{p}(1-\overline{p})} \end{cases}$$

donde $n\bar{p} = \frac{\sum_{i=1}^{N} n\, p_i}{N}$

Gráficos de Control *c*

c) Gráfico ***c***

En algunas situaciones puede ser necesario **controlar el número de defectos en una unidad de producto**, en vez de la fracción de defectos. En estas situaciones podemos emplear el gráfico de control c.

Supóngase que en la producción de ropa es necesario controlar el número de defectos por metro cuadrado, o que en el ensamblado de un ala de avión el número de remaches faltantes debe controlarse.

Muchas situaciones de defectos por unidad pueden modelarse por medio de la **distribución de Poisson**.

Las muestras incluidas en un gráfico **c** son productos individuales de tamaño constante.

El número de defectos en cada producto representados por c, se cuenta y se registra como el valor de una muestra.

Gráficos de Control *c*

c) Gráfico *c*

En algunas situaciones puede ser necesario controlar el número de defectos en una unidad de producto, en vez de la fracción de defectos. En estas situaciones podemos emplear el gráfico de control c.

Sea c el número de defectos en una unidad, donde c es una variable aleatoria de Poisson con parámetro α. Por lo tanto, si se disponen k unidades y c_i es el número de defectos en la unidad i, la línea central del diagrama de control es:

$$\bar{c} = \frac{1}{k}\sum_{i=1}^{k} c_i \qquad \text{y} \qquad C \Rightarrow \begin{cases} LSC = \bar{c} + 3\sqrt{\bar{c}} \\ LC = \bar{c} \\ LIC = \bar{c} - 3\sqrt{\bar{c}} \end{cases}$$

Gráficos de Control u

d) Gráfico *u*

El gráfico ***u*** se utiliza cuando no es posible tener siempre la misma unidad de medida para contar el número de defectos (o no-conformidades, o clientes, etc...). Entonces, se controla el número medio de defectos por unidad de medida.

Entonces, una carta de control ***u***, cantidad de defectos por unidad en una muestra, es el equivalente a una carta de control ***c***, en la que el tamaño de la muestra puede ser variable o constante.

Gráficos de Control $\boldsymbol{u}$

d) Gráfico *u*

Cuando las muestras no son de igual tamaño los límites de control superior o inferior se calculan por separado para cada tamaño siendo:

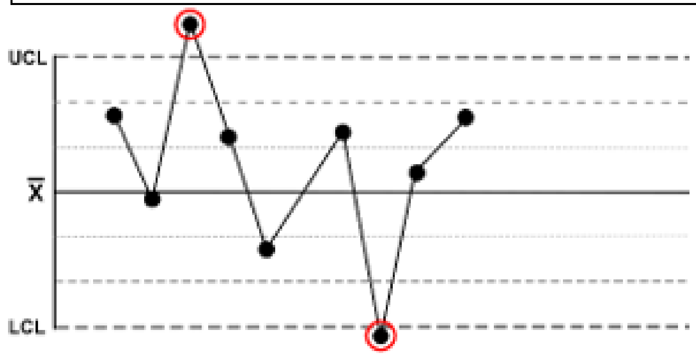
$$U \Rightarrow \begin{cases} LSC = \bar{u} + 3\sqrt{\dfrac{\bar{u}}{n_i}} \\ LC = \bar{u} \\ LIC = \bar{u} + 3\sqrt{\dfrac{\bar{u}}{n_i}} \end{cases} \quad donde \quad \begin{array}{l} \bar{u} = \dfrac{\sum_{i=1}^{K} u_i}{K} \quad siendo \quad K = número\ de\ muestras \\ u_i = número\ de\ defectos\ por\ unidad\ de\ la\ muestra\ "i" \\ n_i = número\ de\ productos\ o\ elementos\ de\ la\ muestra\ "i" \end{array}$$

En la práctica, para simplificar los cálculos, se reemplaza n_i por $\bar{n} = \dfrac{\sum_{i=1}^{K} n_i}{K}$ y de esta forma, permanecen constantes $\forall i$ los valores de LSC y LIC.

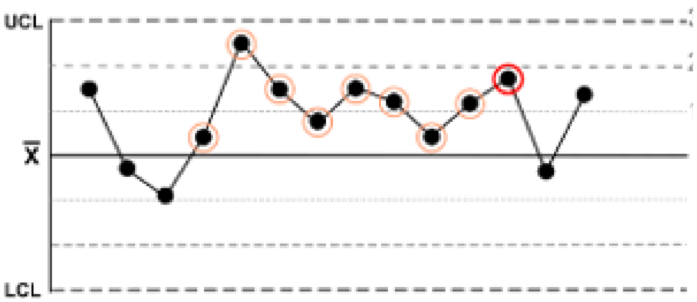
$$U \Rightarrow \begin{cases} LSC = \bar{u} + 3\sqrt{\dfrac{\bar{u}}{n_i}} \approx \bar{u} + 3\sqrt{\dfrac{\bar{u}}{\bar{n}}} \quad \forall i \\ LC = \bar{u} \\ LIC = \bar{u} + 3\sqrt{\dfrac{\bar{u}}{n_i}} \approx \bar{u} - 3\sqrt{\dfrac{\bar{u}}{\bar{n}}} \quad \forall i \end{cases}$$

Interpretación de las gráficas de control

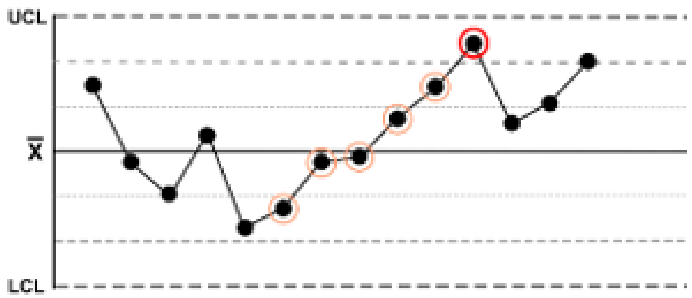
Un punto del gráfico excede los límites de 3σ



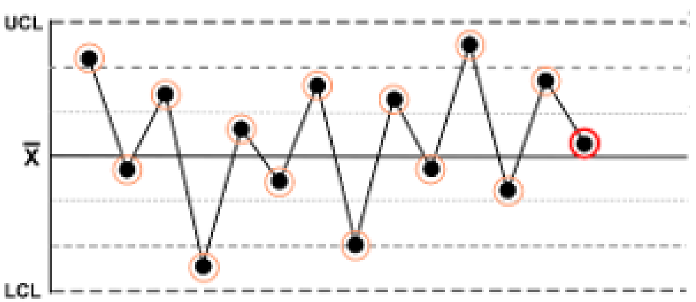
7 (ó más) puntos consecutivos ubicados del mismo lado de la línea central



7 (ó mas) puntos consecutivos en escalera ascendente o descendente

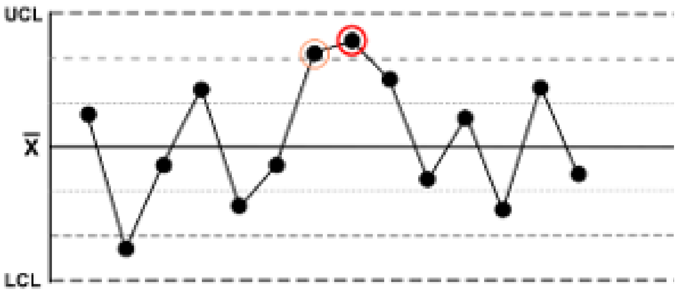


14 (ó más) puntos consecutivos oscilando alternativamente hacia arriba y hacia abajo

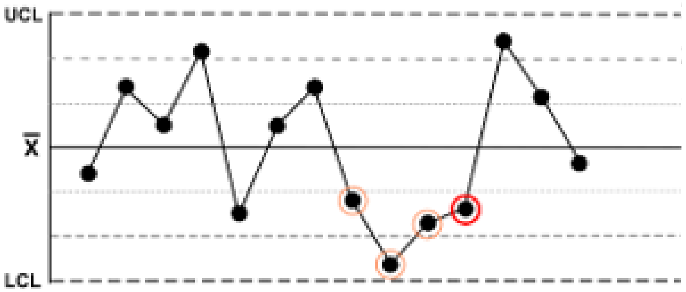


Interpretación de las gráficas de control

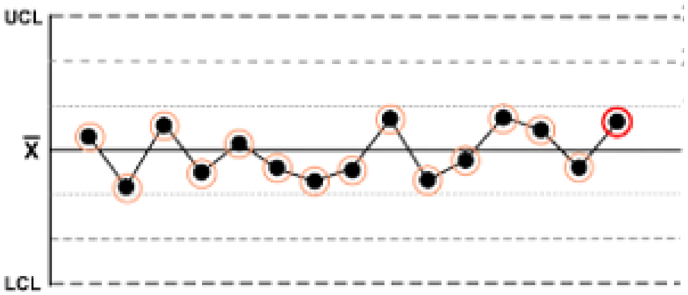
2 puntos (ó 3 puntos) de 3 consecutivos, exceden los límites de 2σ en la misma dirección



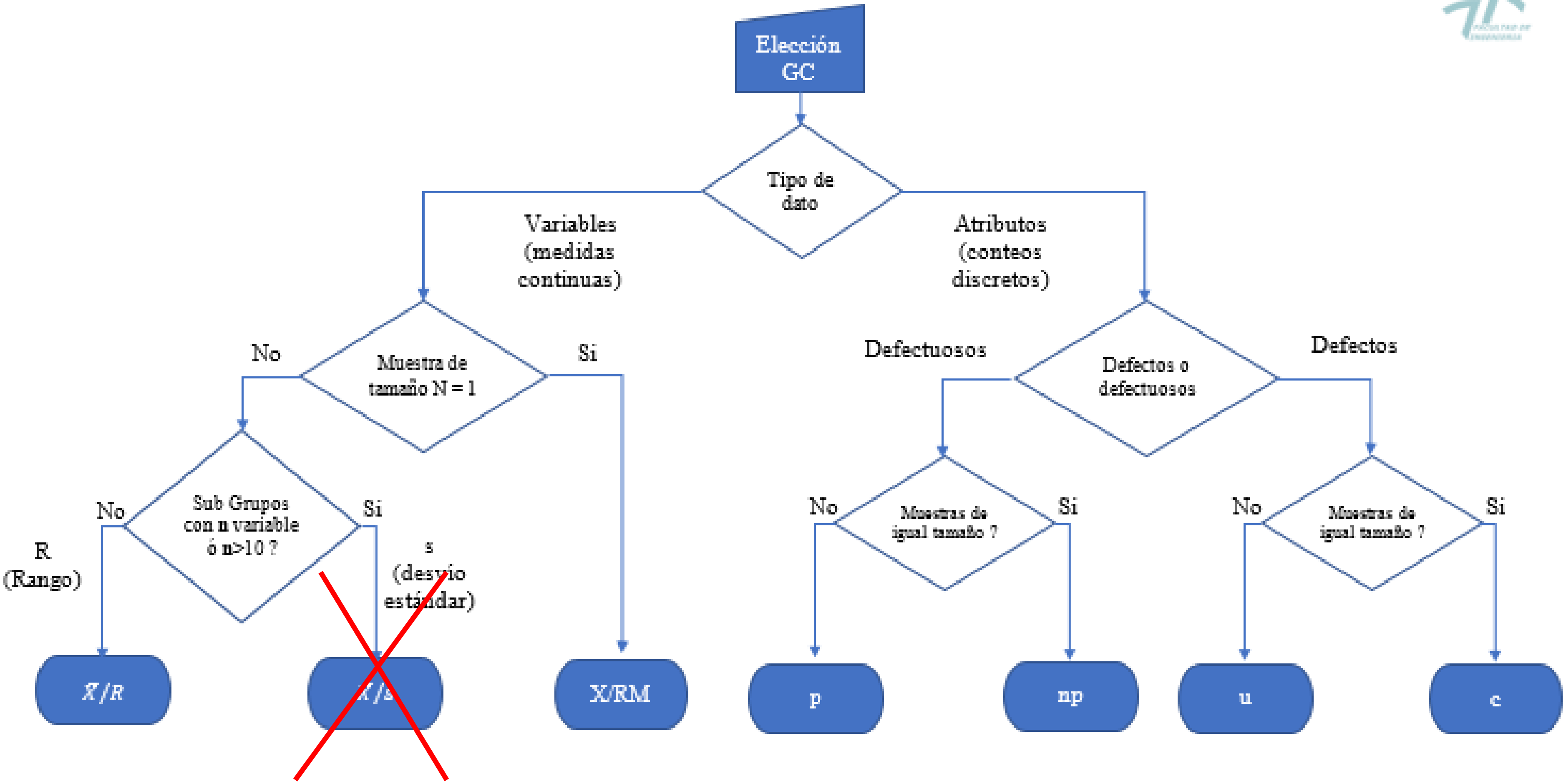
4 puntos (ó 5 puntos) de 5 consecutivos, exceden los límites de 1σ en la misma dirección



15 puntos consecutivos dentro del límite de control de ±1,5σ



Control de Calidad



Nota1: Se suele usar un GC del tipo X/RM cuando la única muestra de tamaño N=1 tiene entre 15 y 25 datos

Nota2: Se suele usar un GC del tipo $\overline{X}/R$ cuando se pueden conseguir valores de N entre 15 y 25, c/u de ellos formados por hasta 10 subgrupos, preferentemente 3, 4, ó 5

Nota3: Para esta Materia, no usaremos GC del tipo $\overline{X}/s$ pues requieren muchos cálculos y se adaptan mejor cuando se dispone de sistemas estadísticos computarizados.

Desigualdad de Tchebycheff.
Ley de los grandes números.

2022.

Desigualdad de Tchebycheff

Desigualdad de Tchebycheff

Si la esperanza y la varianza de una v. a. son finitas, para cualquier número real positivo $k > 1$, la probabilidad de que la v. a. X esté en el intervalo $(\mu - k\sigma, \mu + k\sigma)$ es mayor o igual a $1 - 1/k^2$.

Formalmente:

$$P\left[x \in (\mu - k\sigma, \mu + k\sigma)\right] \geq 1 - \frac{1}{k^2}$$

o por sucesos contrarios:

$$P\left[x \notin (\mu - k\sigma, \mu + k\sigma)\right] < \frac{1}{k^2}$$

Interpretación: la probabilidad de que X asuma un valor que se encuentra entre las k dispersiones de su esperanza es por lo menos $1 - 1/k^2$ para todo $k > 1$.

Desigualdad de Tchebycheff

Observaciones:

1) Esta desigualdad puede ser aplicada a cualquier variable aleatoria.

2) Si se conoce la distribución de probabilidades de la variable aleatoria (discreta o continua), podemos calcular, si existe, su esperanza y varianza. La recíproca no vale, es decir, si se conocen la esperanza y la varianza no es posible reconstruir la distribución de probabilidades, por lo que no podemos calcular la probabilidad exacta de la desigualdad.

En este caso a partir de este resultado podemos dar una cota superior o inferir de dicha probabilidad.

Formalmente:

$$P\left[x \in (\mu - k\sigma, \mu + k\sigma)\right] \geq 1 - \frac{1}{k^2} \quad \equiv \quad P\left[|X - \mu| < k\sigma\right] \geq 1 - \frac{1}{k^2}$$

o por sucesos contrarios:

$$P\left[x \notin (\mu - k\sigma, \mu + k\sigma)\right] < \frac{1}{k^2} \quad \equiv \quad P\left[|X - \mu| \geq k\sigma\right] < \frac{1}{k^2}$$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

X = demanda de paquetes de cigarrillos $E(X) = \mu = 100 \;\; y \;\; V(X) = \sigma = 40$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

X = demanda de paquetes de cigarrillos $E(X) = \mu = 100 \quad y \quad V(X) = \sigma = 40$

$$P[50 \leq X \leq 150] = \; P[|X - 100| \leq 50] \geq 1 - \frac{1}{k^2}$$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

X = *demanda de paquetes de cigarrillos* $\quad E(X) = \mu = 100 \;\; y \;\; V(X) = \sigma = 40$

$$P[50 \leq X \leq 150] = \; P[|X - 100| \leq 50] \geq 1 - \frac{1}{k^2}$$

$$\boxed{P[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

$X = \text{demanda de paquetes de cigarrillos} \qquad E(X) = \mu = 100 \;\; y \;\; V(X) = \sigma = 40$

$$P[50 \leq X \leq 150] = \; P[|X - 100| \leq 50] \geq 1 - \frac{1}{k^2}$$

$$\boxed{P[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

$$Tomando \;\; k\sigma = 50 \;\; se\ obtiene \;\; k = \frac{50}{40} = \frac{5}{4} = 1{,}25$$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

X = *demanda de paquetes de cigarrillos* $\qquad E(X) = \mu = 100 \;\; y \;\; V(X) = \sigma = 40$

$$P[50 \leq X \leq 150] = \; P[|X - 100| \leq 50] \geq 1 - \frac{1}{k^2}$$

$$\boxed{P[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

$$Tomando \;\; k\sigma = 50 \;\; se\ obtiene \;\; k = \frac{50}{40} = \frac{5}{4} = 1{,}25$$

$$Así\ se\ obtine\ que \;\; 1 - \frac{1}{k^2} \; = \; 1 - \frac{1}{\left(\frac{5}{4}\right)^2} = 1 - \left(\frac{4}{5}\right)^2 = \frac{9}{25} = 0{,}36$$

Desigualdad de Tchebycheff

<u>Ejemplo</u>: En un kiosco céntrico, la demanda media de cigarrillos (de una marca A) es de 100 paquetes por día, con una desviación de 40 paquetes.

Hallar una cota inferior para la probabilidad que, diariamente, se demanden entre 50 y 150 paquetes

Solución

$X = demanda\ de\ paquetes\ de\ cigarrillos \qquad E(X) = \mu = 100 \ \ y \ \ V(X) = \sigma = 40$

$$P[50 \leq X \leq 150] = \ P[|X - 100| \leq 50] \geq 1 - \frac{1}{k^2}$$

$$\boxed{P[|X - \mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

$$Tomando \ \ k\sigma = 50 \ \ se\ obtiene \ \ k = \frac{50}{40} = \frac{5}{4} = 1{,}25$$

$$Así\ se\ obtine\ que \ \ 1 - \frac{1}{k^2} \ = \ 1 - \frac{1}{\left(\frac{5}{4}\right)^2} = 1 - \left(\frac{4}{5}\right)^2 = \frac{9}{25} = 0{,}36$$

$$\boxed{P[50 \leq X \leq 150] \geq 0{,}36}$$

Desigualdad de Tchebycheff

Ley de los grandes números.

Sea ε un experimento y A un evento asociado a ε. Consideremos n repeticiones independientes del experimento ε y sea n_A el número de veces que A ocurre en las n repeticiones.

Si $f_A = \dfrac{n_A}{n}$ entonces $f_A \underset{x \to \infty}{\rightarrow} P(A)$; **convergencia en sentido probabilístico**.

Interpretación: En toda sucesión de pruebas de Bernoulli, la frecuencia relativa converge en sentido probabilístico a la probabilidad.

Ley de los grandes números:

Cuando el número de repeticiones de un experimento aleatorio aumenta, la frecuencia relativa asociada al suceso A (f_A) converge en sentido probabilístico a la probabilidad teórica P(*A*)

Supongamos que $P(A) = p$ para todas las repeticiones del experimento ε. Para todos $\delta > 0$, tenemos:

$$\lim_{n \to \infty} \left\{ P\left(\left| f_A - p \right| < \delta \right) \right\} = 1$$

La probabilidad de que la frecuencia relativa del éxito en n pruebas independientes difiera de la probabilidad p en una cantidad menor que δ tiende a 1 cundo la cantidad de pruebas tiende a ∞. De manera equivalente podemos expresar:

$$\lim_{n \to \infty} \left\{ P\left(\left| f_A - p \right| \geq \delta \right) \right\} = 0$$

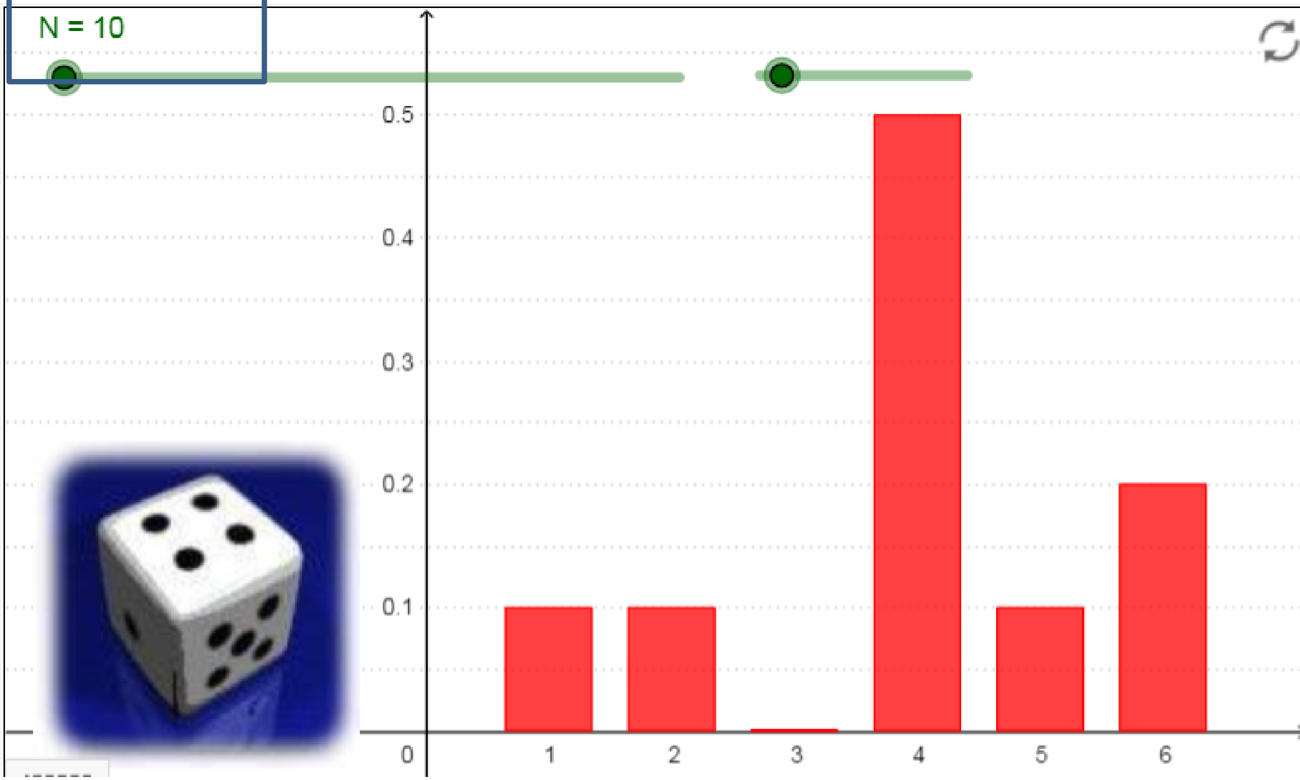
Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

Ley de los grandes números

Frecuencia Relativa.

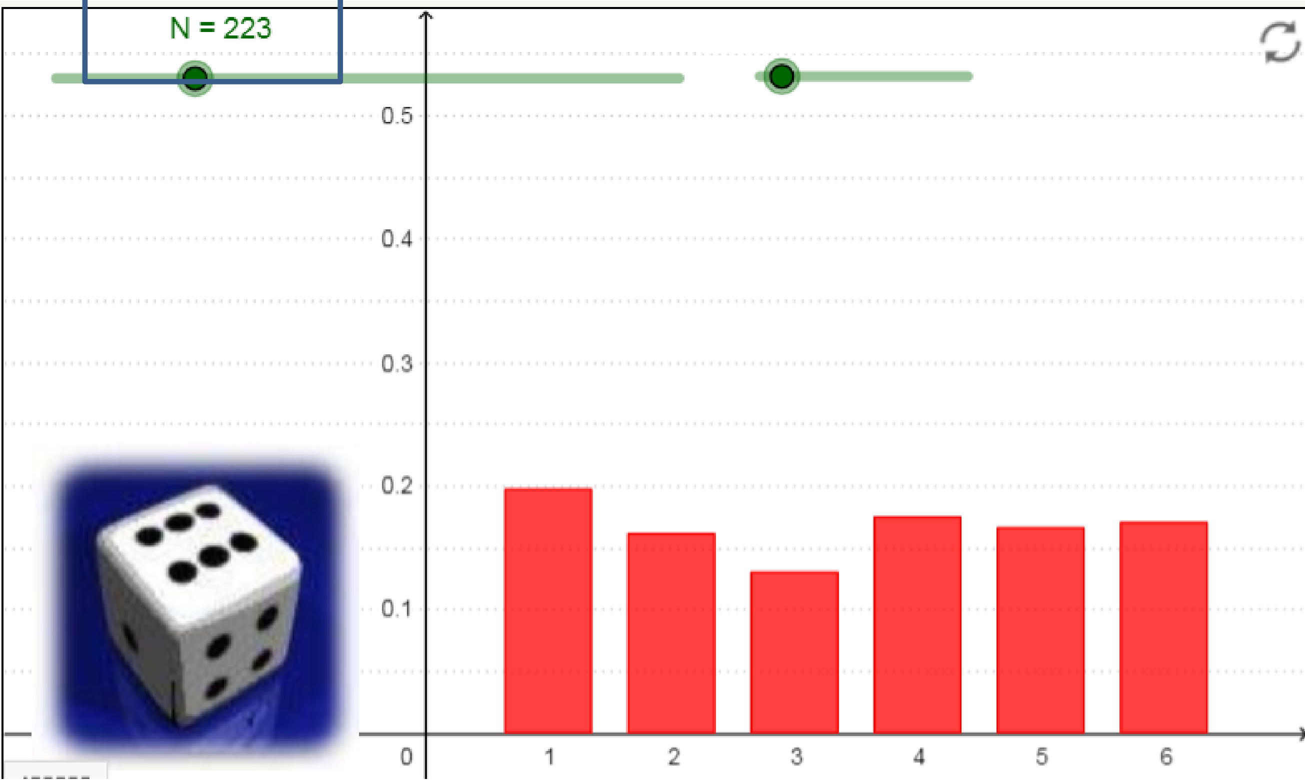


	A	B	C	D
1		F Absoluta	F. Relativas	Prob.T.
2	1	1	0.1	0.1667
3	2	1	0.1	0.1667
4	3	0	0	0.1667
5	4	5	0.5	0.1667
6	5	1	0.1	0.1667
7	6	2	0.2	0.1667
8	total	10		
9				
10				
11				
12				
13				
14				
15				

$$\boxed{0.2 - 0.1667 = 0.0333}$$

$$f_r = \frac{n_A}{n}$$

Frecuencia Relativa.

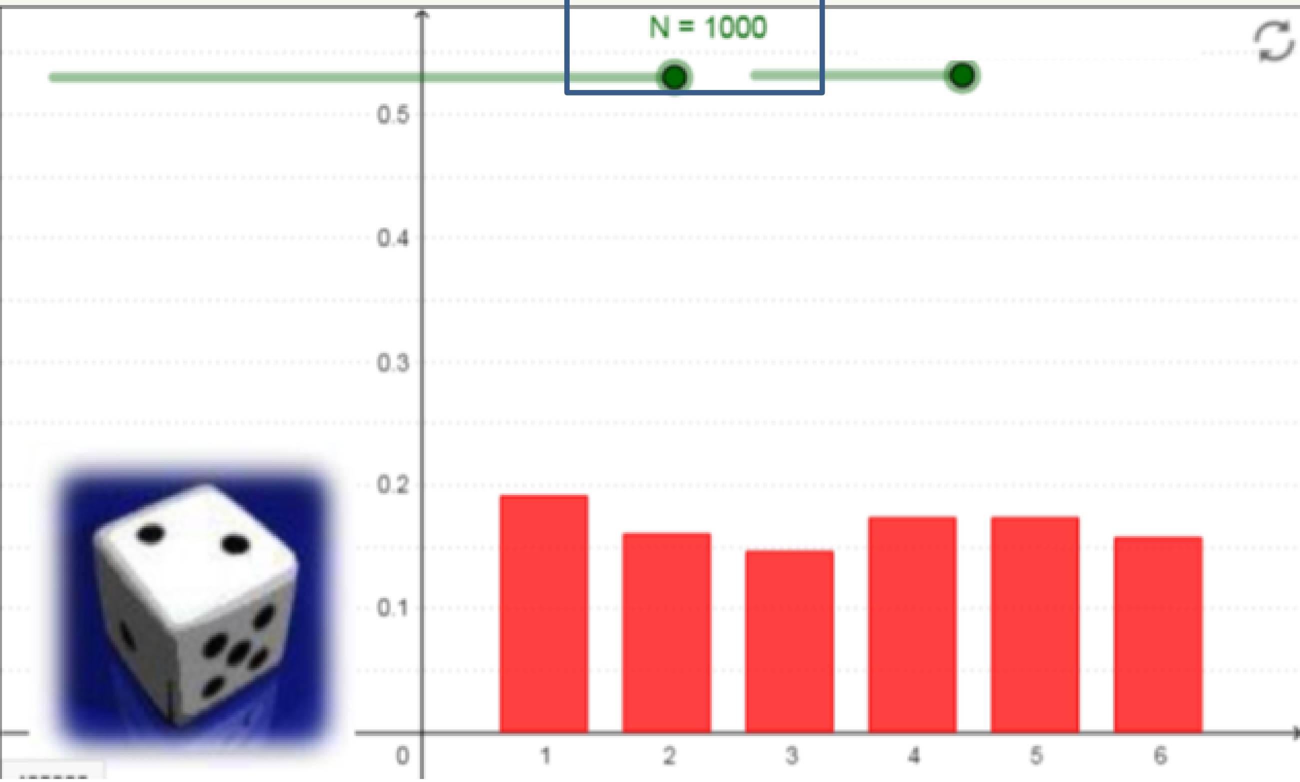


	A	B	C	D
1		F Absoluta	F. Relativas	Prob.T.
2	1	44	0.1973	0.1667
3	2	36	0.1614	0.1667
4	3	29	0.13	0.1667
5	4	39	0.1749	0.1667
6	5	37	0.1659	0.1667
7	6	38	0.1704	0.1667
8	total	223		
9				
10				
11				
12				
13				
14				
15				

$$0.1704 - 0.667 = 0.0037$$

$$f_r = \frac{n_A}{n}$$

<u>Frecuencia Relativa</u>.



	A	B	C	D
1		F Absoluta	F. Relativas	Prob.T.
2	1	191	0.191	0.1667
3	2	160	0.16	0.1667
4	3	146	0.146	0.1667
5	4	173	0.173	0.1667
6	5	173	0.173	0.1667
7	6	157	0.157	0.1667
8	total	1000		
9				
10				
11				
12				
13				
14				
15				

$$f_r = \frac{n_A}{n}$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

X: número de 6 obtenidos en n repeticiones del experimento.

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$E(X) = np$$

$$V(X) = np(1-p)$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n}$$

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2{}_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

$$P\left[\left|f_A - \frac{1}{6}\right| < 0.01\right] \geq 0.95$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2{}_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

$$\boxed{P[|X-\mu| \le k\sigma] \ge 1 - \frac{1}{k^2}}$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2{}_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

$$P\left[\left|f_A - \frac{1}{6}\right| < 0.01\right] \geq 0.95$$

$$\boxed{P[|X-\mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

$$P\left[\left|f_A - \frac{1}{6}\right| < 0.01\right] \geq 0.95$$

$$\boxed{P[|X-\mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

$$1 - \frac{1}{k^2} = 0.95 \Rightarrow k = 4.47$$

$$k\sigma_{f_A} = 0.01 \Rightarrow 4.47 \times \sqrt{\frac{\frac{5}{36}}{n}} = 0.01 \Rightarrow n > 27.778$$

Ley de los grandes números

Ejemplo: ¿Cuántas veces se debe lanzar un dado regular para tener una probabilidad de 0.95, de que la frecuencia relativa del número de 6 obtenidos difiera de la probabilidad teórica en menos de 0.01?

Solución/

$$E(X) = np$$

X: número de 6 obtenidos en n repeticiones del experimento. $X \sim B\left(n, \frac{1}{6}\right)$

$$V(X) = np(1-p)$$

$$f_A = \frac{X}{n} \qquad E(f_A) = \frac{1}{6} \qquad V(f_A) = \sigma^2{}_{f_A} = \frac{\frac{1}{6}\left(1-\frac{1}{6}\right)}{n} = \frac{\frac{5}{36}}{n}$$

$$P\left[\left|f_A - \frac{1}{6}\right| < 0.01\right] \geq 0.95$$

$$\boxed{P[|X-\mu| \leq k\sigma] \geq 1 - \frac{1}{k^2}}$$

$$1 - \frac{1}{k^2} = 0.95 \Rightarrow k = 4.47$$

$$k\sigma_{f_A} = 0.01 \Rightarrow 4.47 \times \sqrt{\frac{\frac{5}{36}}{n}} = 0.01 \Rightarrow n > 27.778$$

Rta: n mayor o igual que 28

Distribuciones Muestrales

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

La inferencia estadística está formada por un conjunto de métodos utilizados para **tomar decisiones u obtener conclusiones sobre una población a partir de la información contenida en una muestra** obtenida en forma aleatoria de la población.

Inferencia Estadística

Normalmente, por distintas razones, es imposible estudiar la totalidad de la población para tomar decisiones u obtener conclusiones. **Por tal razón, para obtener conclusiones sobre la población se recurre a una muestra de la misma.**

La inferencia estadística está formada por un conjunto de métodos utilizados para **tomar decisiones u obtener conclusiones sobre una población a partir de la información contenida en una muestra** obtenida en forma aleatoria de la población.

Cualquier inferencia o conclusión obtenida de la población, necesariamente, estará basada en la información proporcionada por la muestra.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Sabemos que la variable de interés es binomial pero desconocemos la probabilidad de éxito poblacional o el número de pruebas de Bernoulli.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Sabemos que la variable de interés es binomial pero desconocemos la probabilidad de éxito poblacional o el número de pruebas de Bernoulli.

Sabemos que se trata de un proceso de Poisson pero desconocemos la frecuencia media de ocurrencia por unidad.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Sabemos que la variable de interés es binomial pero desconocemos la probabilidad de éxito poblacional o el número de pruebas de Bernoulli.

Sabemos que se trata de un proceso de Poisson pero desconocemos la frecuencia media de ocurrencia por unidad.

Presumimos que la variable es exponencial pero desconocemos el parámetro que define la distribución exponencial poblacional.

Inferencia Estadística

En muchos casos sabemos o presumimos conocer cómo se distribuye una población.

Sabemos por ejemplo que la población es aproximadamente normal; pero desconocemos la media y la varianza poblacionales.

Sabemos que la variable de interés es binomial pero desconocemos la probabilidad de éxito poblacional o el número de pruebas de Bernoulli.

Sabemos que se trata de un proceso de Poisson pero desconocemos la frecuencia media de ocurrencia por unidad.

Presumimos que la variable es exponencial pero desconocemos el parámetro que define la distribución exponencial poblacional.

La estimación de uno o varios parámetros poblacionales desconocidos es posible construyendo funciones de probabilidad de variables aleatorias muestrales, más conocidos como estimadores muestrales.

Población y Muestra

Población o universo: Es el conjunto de individuos o elementos con características comunes que se desean investigar. Normalmente es demasiado grande para poder abarcarla. **El estudio de toda la población se denomina censo**.

Muestra: es un subconjunto de la población al que tenemos acceso y sobre el que realmente hacemos las observaciones (mediciones). El objetivo principal de la toma de una muestra aleatoria es obtener información sobre los parámetros no conocidos de la población.

La **inferencia estadística** consiste en generalizar las conclusiones extraídas de una muestra sobre la población.

Muestreo Aleatorio Simple (M.A.S.)

Muestreo. Es el proceso para extraer una muestra de una población. Es necesario utilizar ***"muestras representativas"*** del total de la población, es decir, muestras en las que exista alguna garantía de que cualquier elemento de la población queda representado.

En el ***Muestreo Aleatorio Simple***, cada elemento de la población tiene la misma probabilidad de ser elegido para formar parte de la muestra. Cada muestra del mismo tamaño tiene la misma probabilidad de ser seleccionada.

- El muestreo aleatorio simple en poblaciones finitas se realiza "***con sustitución***".
- Este procedimiento **garantiza la independencia de las observaciones**.

Muestreo Aleatorio Simple (M.A.S.)

Formalmente:

Muestra aleatoria simple: Si X es una variable aleatoria con distribución de probabilidades o *fdp* $f(x)$, el conjunto de n observaciones tomadas de la variable X, $X_1, X_2, \ldots, X_n$, con resultados numéricos $x_1, x_2, \ldots, x_n$, es una muestra aleatoria de tamaño n si:

a) Las X_i son variables aleatorias independientes.

b) Cada observación X_i tiene la misma distribución de probabilidades.

El objetivo de tomar una muestra es obtener información sobre los parámetros no conocidos de la población.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Inferencia Estadística

Estimación puntual: Es el valor numérico que toma un estimador. Se calcula con los datos de la muestra, del cual se espera que estime un parámetro poblacional.

Parámetro Poblacional	Estimador	Estimación
Media μ	$\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$	$\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$
Varianza σ^2	$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$
Proporción p	$\hat{p} = \dfrac{X}{n} = \dfrac{\text{númeroéxitos}}{\text{númeropruebas}}$	$\hat{p} = \dfrac{x}{n}$
Parámetros poblacionales, estimadores y estimaciones		

Distribuciones de Muestreo.

Las muestras aleatorias obtenidas de una población son, por naturaleza propia, impredecibles. No se debería esperar que dos muestras aleatorias del mismo tamaño y tomadas de la misma población tengan, por ejemplo, la misma media muestral (o que sean completamente parecidas). Por lo tanto es de esperar que cualquier estadístico cambie su valor de una muestra a otra, por ello, se quiere estudiar la distribución de todos los valores posibles de un estadístico.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Inferencia Estadística

<u>Distribuciones de Muestreo</u>.

Cómo los valores de un estadístico, varían de una muestra aleatoria a otra, se lo puede considerar como una **variable aleatoria** con su correspondiente distribución de probabilidades. La distribución de probabilidades de un estadístico se conoce como ***distribuciones de muestreo***.

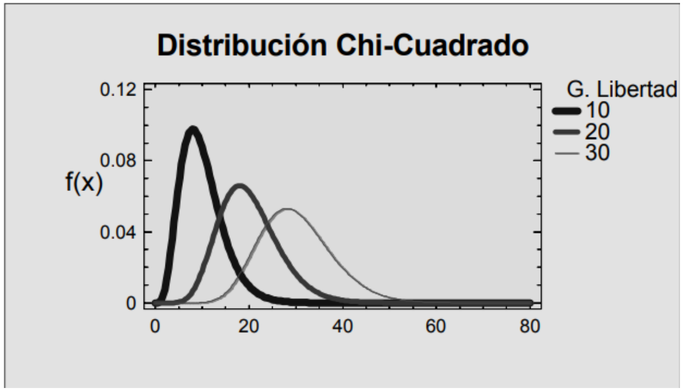
La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Objetivo: ***Conocer la distribución muestral de los siguientes estadísticos bajo ciertas condiciones.***

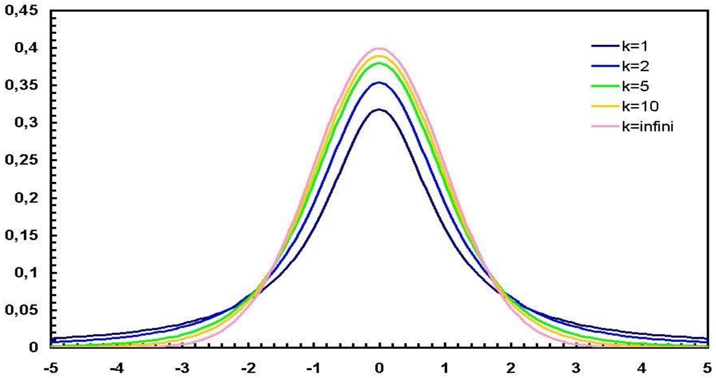
Estimador
$\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$
$\hat{\sigma}^2 = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
$\hat{p} = \dfrac{X}{n} = \dfrac{\text{número éxitos}}{\text{número pruebas}}$

Antes de continuar veamos algunas distribuciones asociadas al muestreo

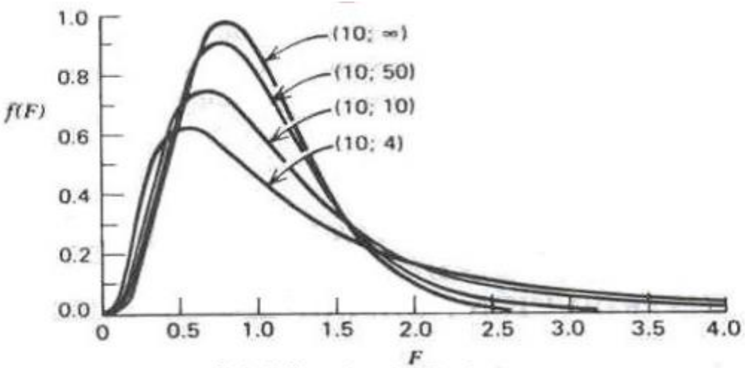
Distribución Chi-Cuadrado



Distribución t-Student



Distribución F de Fisher



Distribución F para varios grados de libertad.

Distribuciones asociadas al muestreo

Distribución Ji-Cuadrado

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias distribuidas normalmente e independientemente con media $\mu = 0$ y varianza $\sigma^2 = 1$. Entonces la variable aleatoria $\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$, tiene una distribución **Ji-Cuadrado** con n grados de libertad, cuya fdp está definida por:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

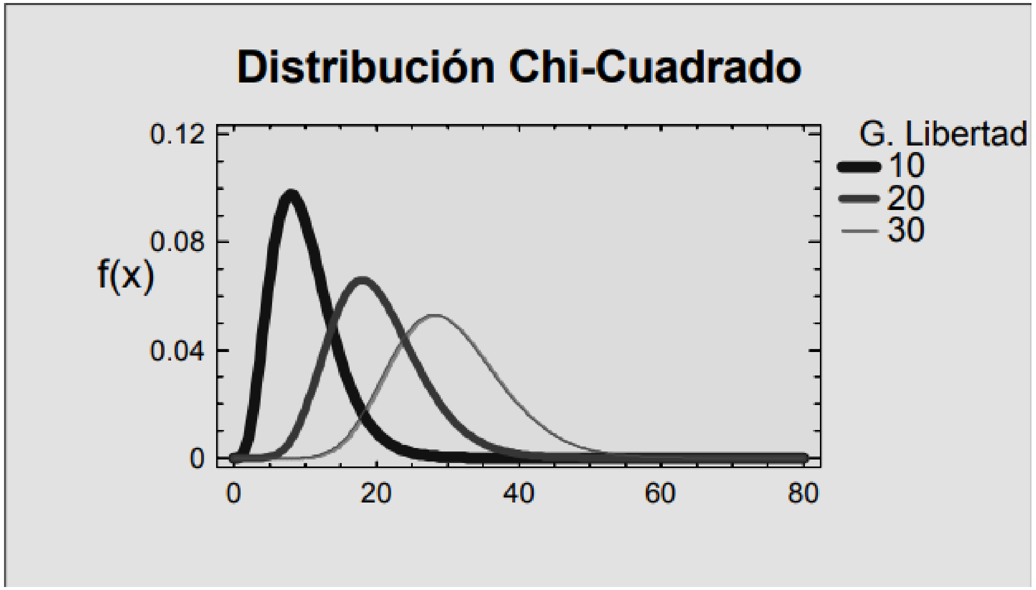
Donde Γ es la función Gamma.

En matemáticas, la función gamma denotada por $\Gamma(z)$, extiende el concepto de factorial a los números complejos. Se define como:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$$

Inferencia Estadística

La siguiente figura muestra la gráfica de la función ji-Cuadrado para distintos grados de libertad



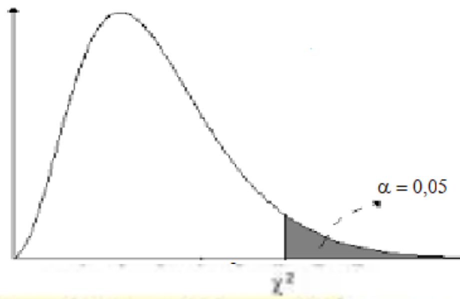
Observación:

- ✓ Si los grados de libertad $n \to \infty$ la forma límite de χ^2 es la distribución normal.
- ✓ La media y la varianza de la distribución ji-cuadrado son $\mu = n$ y $\sigma^2 = 2n$.
- ✓ Calculo de probabilidades a partir de la tabla correspondiente

$$P\left(\chi_n^2 \geq \chi_{(\alpha,n)}^2\right) = \int_{\chi_{(\alpha,n)}^2}^{\infty} f_{\chi_n^2}(x)\,dx = \alpha$$

Ejemplo: uso de tabloa ji-Cuadrado.

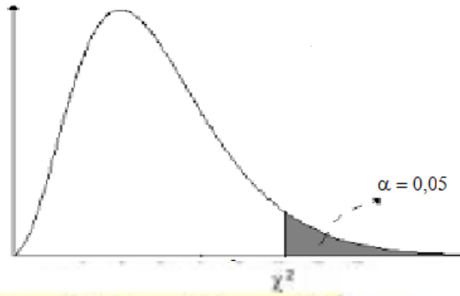
$$P\left(\chi_{10}^{2} > \chi^{2}\right) = 0.05$$



g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

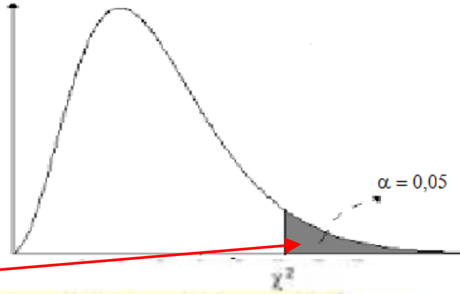
$$P\left(\chi^2_{10} > \chi^2\right) = 0.05$$



g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

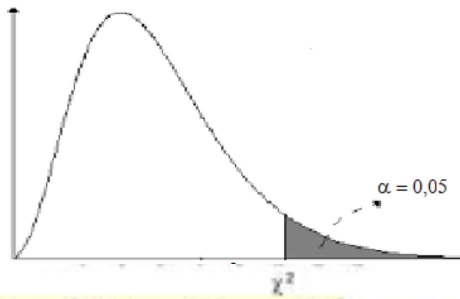
$$P\left(\chi_{10}^{2} > \chi^{2}\right) = 0.05$$



g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
$\boxed{10}$	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Ejemplo: uso de tabloa ji-Cuadrado.

$$P\left(\chi^2_{10} > \chi^2\right) = 0.05$$

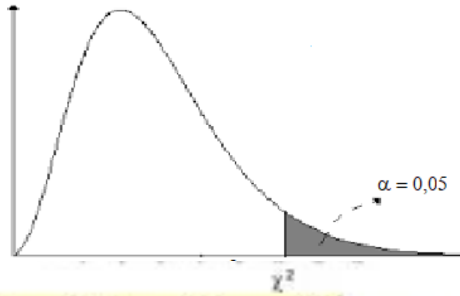


g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Inferencia Estadística

Ejemplo: uso de tabloa ji-Cuadrado.

$$P\left(\chi^2_{10} > \chi^2\right) = 0.05 \;\Rightarrow\; \chi^2 = 18.307$$



g. d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
$\boxed{10}$	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
14	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Distribuciones asociadas al muestreo

Distribución T de Student

Si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

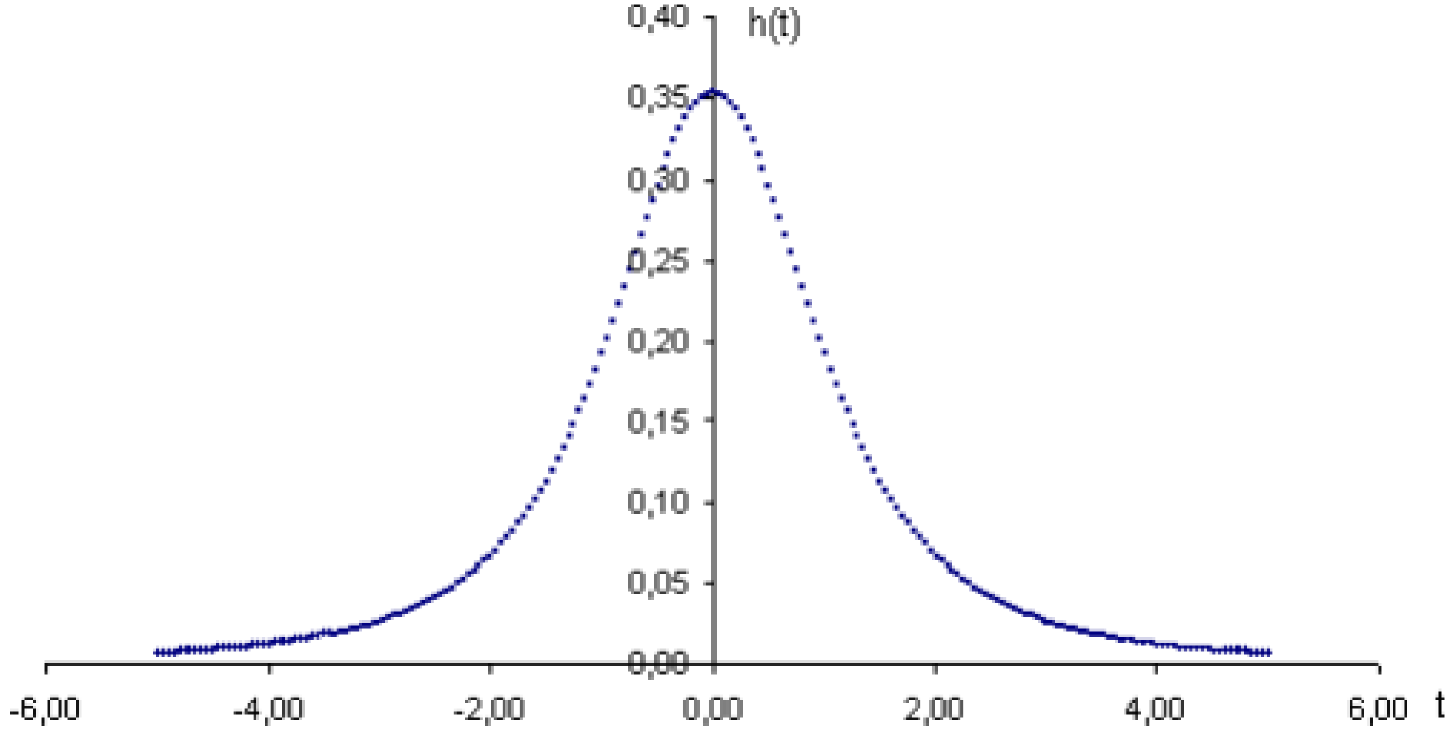
$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

Tiene la siguiente fdp:

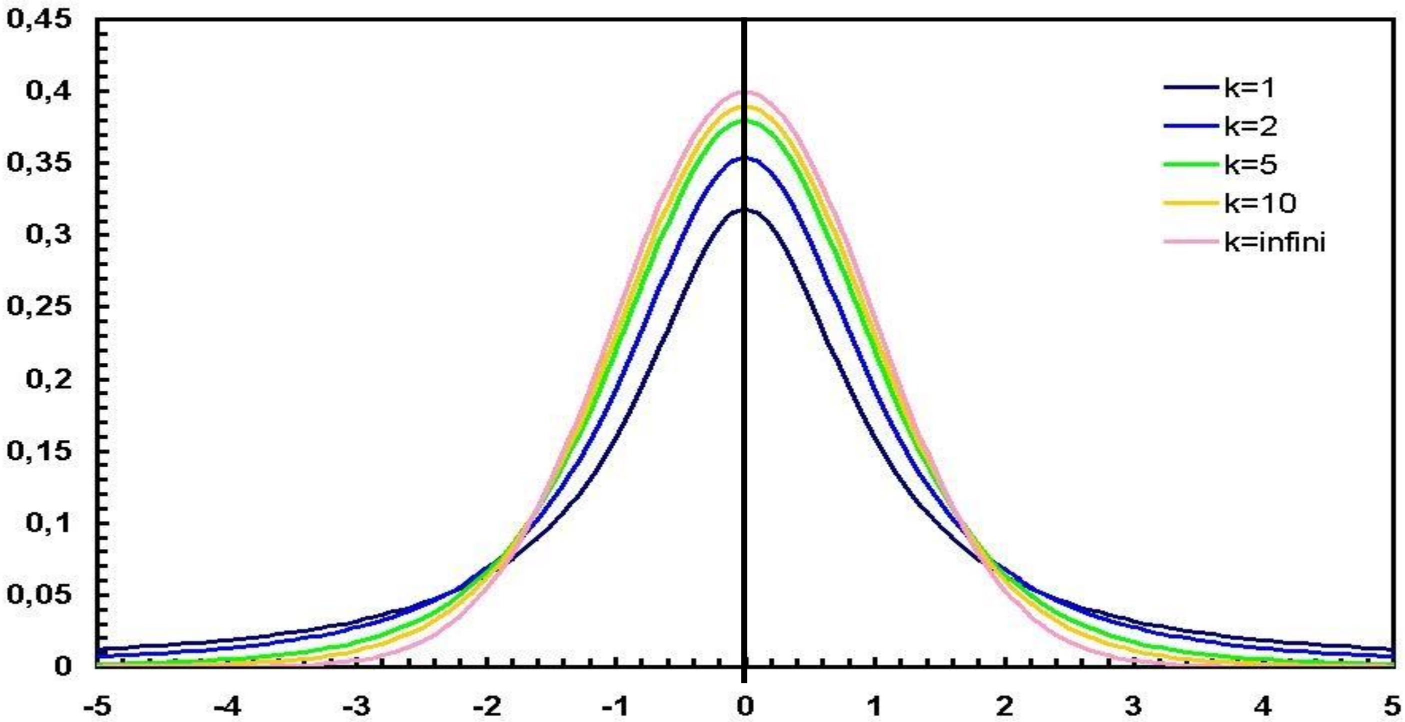
$$f(x) = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(\frac{x^2}{n}+1\right)^{(n+1)/2}} \qquad -\infty < x < \infty$$

La distribución de *T* se llama ahora la **distribución *t* de Student**. El parámetro n representa el número de *grados de libertad*.

La gráfica de esta función de densidad es simétrica, respecto del eje de ordenadas, con independencia del valor de **k**, y de forma algo semejante a la de una distribución normal:



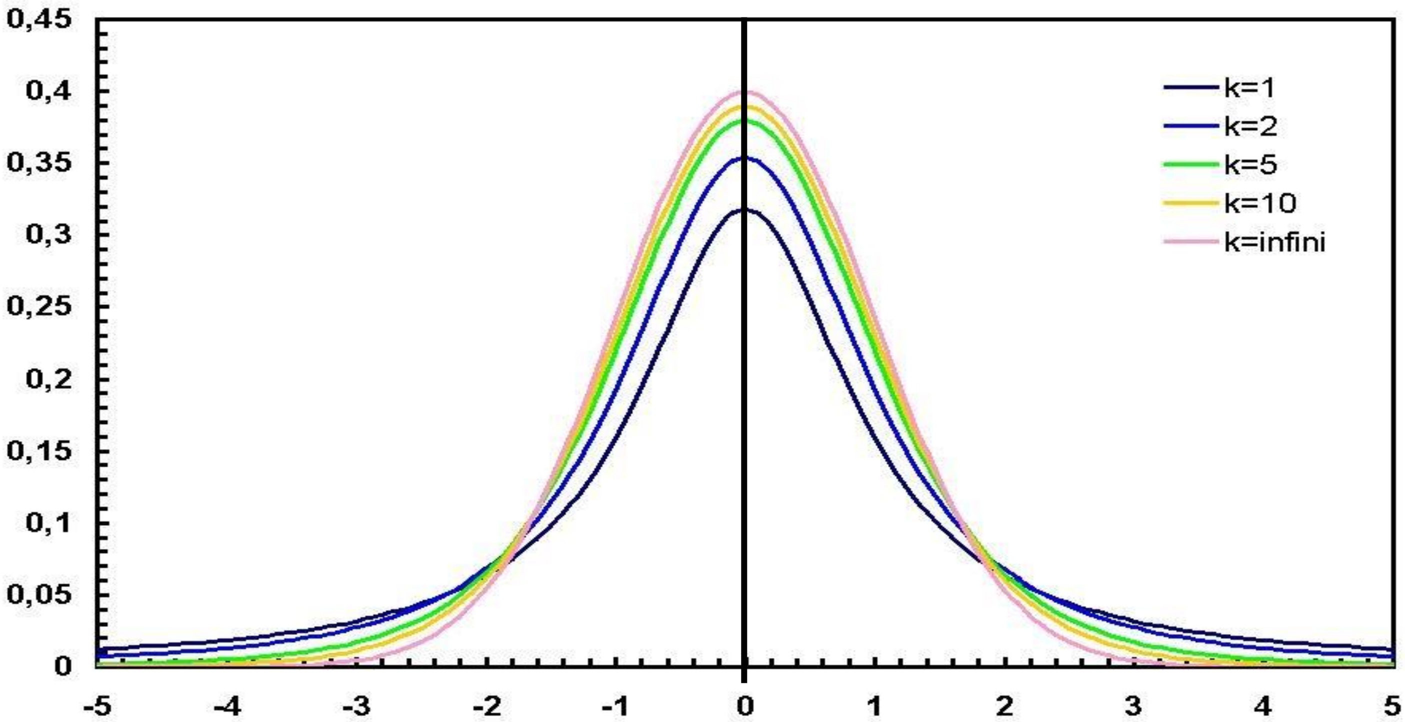
Distribución t de Student con 10 grados de liberta



Propiedades de la distribución t

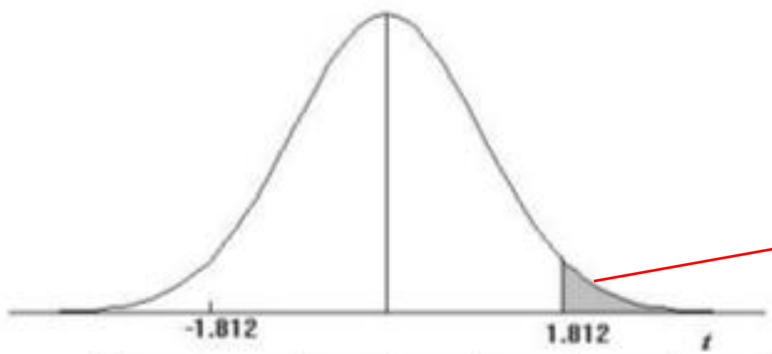
- Cada curva t tiene forma de campana con centro en 0.
- A medida que los grados de libertad (k) aumentan, la dispersión de la curva t correspondiente disminuye.
- A medida que los grados de libertad (k) tienen a infinito, la curva t se aproxima a la curva normal estándar

Inferencia Estadística



Observación: A diferencia de la distribución normal que depende de la media y la varianza, la distribución t solo depende de los grados de libertad, del inglés, *degrees of freedom* (df). En otras palabras, controlando los grados de libertad, controlamos la distribución.

Puntos de porcentaje de la distribución t



Ejemplo

Para $\phi = 10$ grados de libertad:

$P[\,t > 1.812\,] = 0.05$
$P[\,t < -1.812\,] = 0.05$

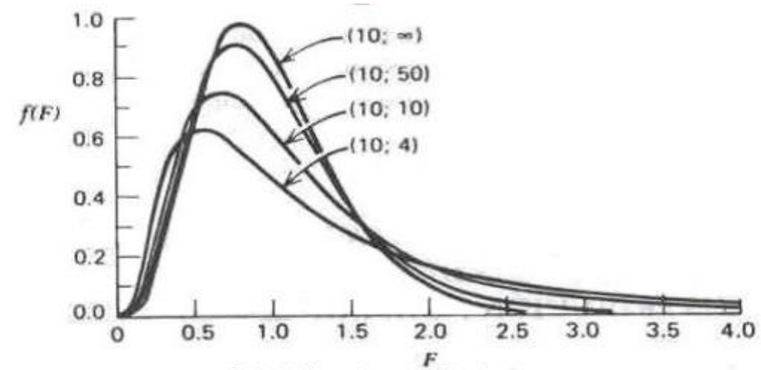
r \ α	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,0005
1	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,656	636,578
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,600
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850

Distribuciones asociadas al muestreo

La distribución F de Fisher

Si χ_1^2 y χ_2^2 son v.a. independientes que tienen distribución ji-cuadrado, con n_1 y n_2 grados de libertad, respectivamente, la variable aleatoria definida por:

$$F = \frac{\dfrac{\chi_1^2}{n_1}}{\dfrac{\chi_2^2}{n_2}}$$



Distribución F para varios grados de libertad.

Tiene una distribución F con n_1 y n_2 grados de libertad.

Distribuciones de Muestreo.

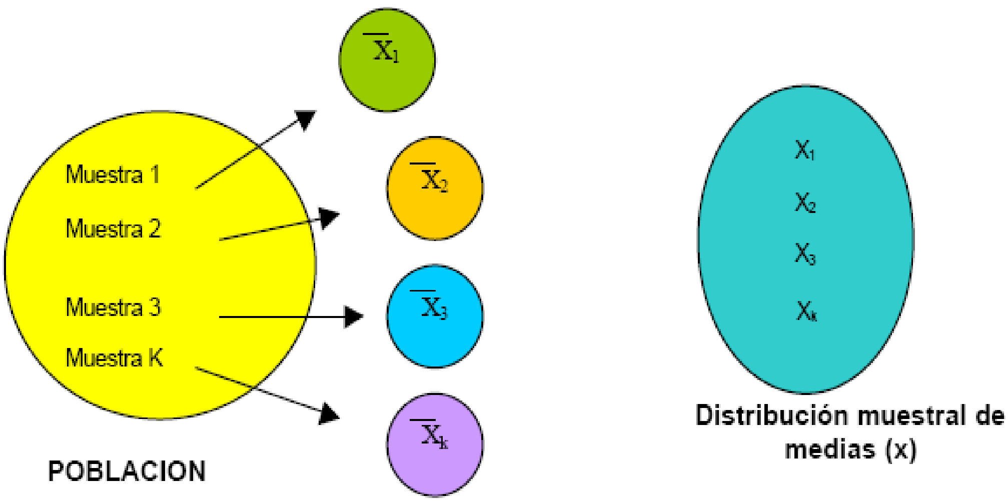
Cómo los valores de un estadístico, varían de una muestra aleatoria a otra, se lo puede considerar como una **variable aleatoria** con su correspondiente distribución de probabilidades. La distribución de probabilidades de un estadístico se conoce como ***distribuciones de muestreo***.

La distribución de probabilidades de un estadístico se denomina ***distribución muestral***.

Objetivo: ***Conocer la distribución muestral de los siguientes estadísticos bajo ciertas condiciones.***

Estimador
$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
$\hat{p} = \frac{X}{n} = \frac{\text{número éxitos}}{\text{número pruebas}}$

<u>Determinación empírica de la distribución muestral de un estadístico:</u> Para hallar empíricamente la distribución muestral de un estadístico es necesario seleccionar todas las muestras de tamaño ***n*** de dicha población y a partir de dicha información construir la distribución de frecuencia relativa de los valores del estadístico, la cual es considerada como su distribución muestral.



Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$\overline{X}$: promedio de 2 observaciones elegidas al azar de entre las 4

Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que los parámetros poblacionales son: la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$\overline{X}$: promedio de 2 observaciones elegidas al azar de entre las 4

$$S_{\overline{X}} = \{2,3,4,5,6,7,8\}$$

Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

Tomamos **muestras de tamaño 2, con reposición**. Cada muestra es de la forma donde: (X_1, X_2)

Consideremos la variable aleatoria: .

$\overline{X}$: promedio de 2 observaciones elegidas al azar de entre las 4

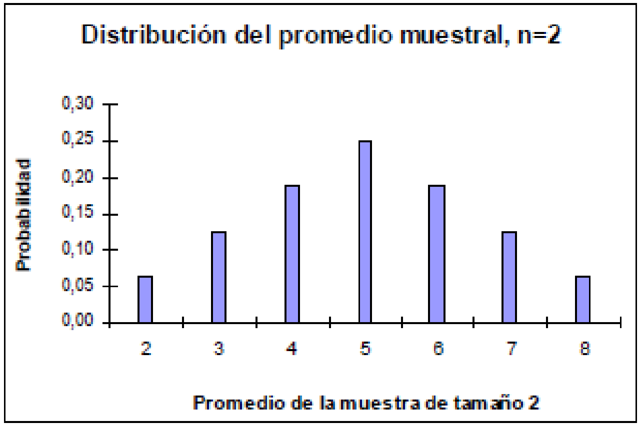
Tabla 2		
Muestra, n=2		
x_1	x_2	Promedio
2	2	2
2	4	3
2	6	4
2	8	5
4	2	3
4	4	4
4	6	5
4	8	6
6	2	4
6	4	5
6	6	6
6	8	7
8	2	5
8	4	6
8	6	7
8	8	8

La distribución de probabilidad es:

$$E\left(\overline{X}\right) = 5$$

$$V\left(\overline{X}\right) = \frac{5}{2}$$

Tabla 3: n=2	
$\overline{X}$	*Probabilidad*
2	0,0625
3	0,1250
4	0,1875
5	0,2500
6	0,1875
7	0,1250
8	0,0625
	1,0000



Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

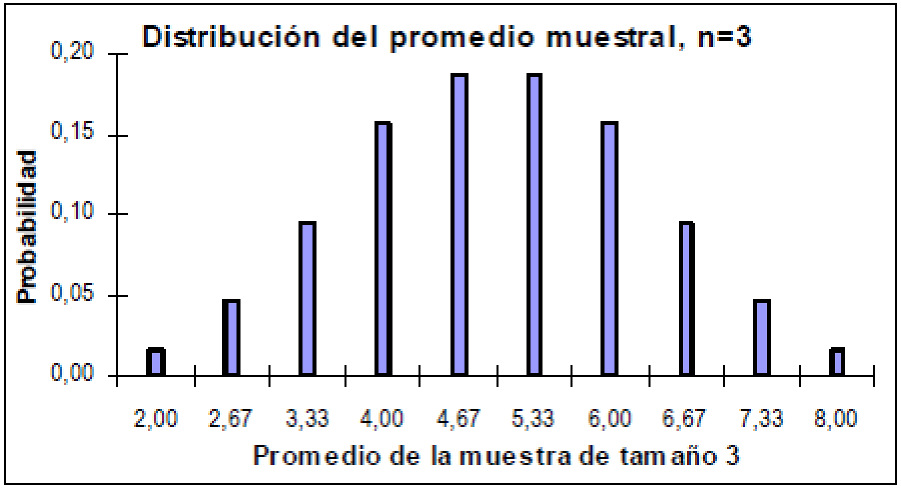
Repitiendo la experiencia, con muestras de tamaño **3**, obtenemos la distribución de los promedios que mostramos a continuación, acompañada de la gráfica:

n = 3

Promedio	*Probabilidad*
2,00	0,015625
2,67	0,046875
3,33	0,093750
4,00	0,156250
4,67	0,187500
5,33	0,187500
6,00	0,156250
6,67	0,093750
7,33	0,046875
8,00	0,015625
	1

$$E\left(\overline{X}\right)=5$$

$$V\left(\overline{X}\right)=\frac{5}{3}$$



Distribuciones de Muestreo

Ejemplo: Supongamos tenemos una población de tamaño cuatro y las mediciones sobre una característica de la población son las siguientes: 2, 4, 6, 8.

Se puede verificar fácilmente que la media poblacional es igual a 5 y la varianza poblacional es igual a 5.

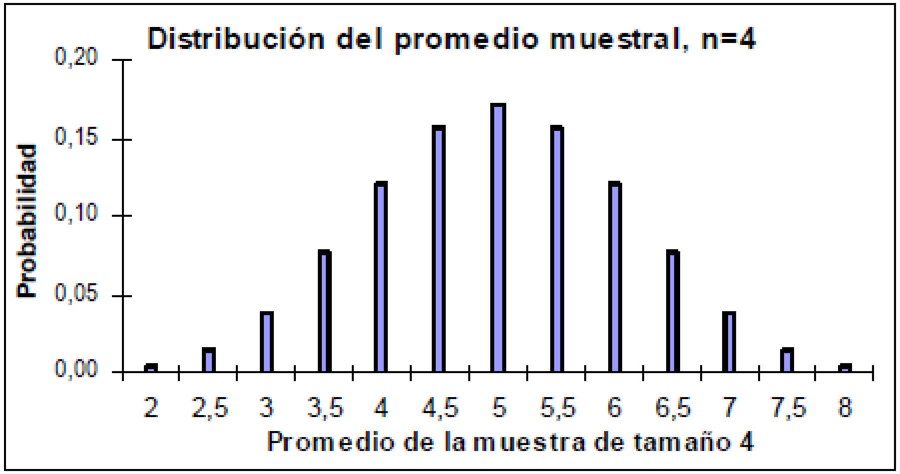
Repitiendo la experiencia, con muestras de tamaño **4**, obtenemos la distribución de los promedios que mostramos a continuación, acompañada de la gráfica:

n = 4

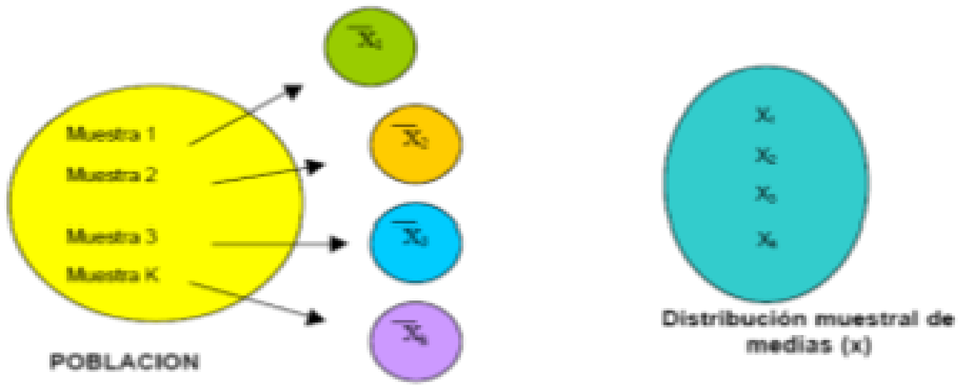
Promedio	*Probabilidad*
2	0,00390625
2,5	0,01562500
3	0,03906250
3,5	0,07812500
4	0,12109375
4,5	0,15625000
5	0,17187500
5,5	0,15625000
6	0,12109375
6,5	0,07812500
7	0,03906250
7,5	0,01562500
8	0,00390625
	1

$$E\left(\overline{X}\right)=5$$

$$V\left(\overline{X}\right)=\frac{5}{4}$$



Distribución de muestreo de la media $\bar{x}$



Si $X_1, X_2, X_3, \ldots, X_n$ son v.a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Sea $\overline{X} = \sum_{i=1}^{n} \frac{x_i}{n}$, *entonces* $\boxed{E(\overline{X}) = \mu \,,\ V(\overline{X}) = \frac{\sigma^2}{n} \text{ y } \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}}$.

$$E(\overline{X}) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}} \cancel{n} \mu = \mu$$

$$V(\overline{X}) = V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^{\cancel{2}}} \cancel{n} \sigma^2 = \frac{\sigma^2}{n}$$

Distribuciones de Muestreo

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza** σ^2 **conocida.**

b) **Distribución muestral de la media para una población normal con varianza** σ^2 **desconocida.**

Distribuciones de Muestreo

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) Distribución muestral de la media para una población normal con varianza σ^2 conocida.

b) Distribución muestral de la media para una población normal con varianza σ^2 desconocida.

Resultado: Al estudiar la distribución normal se consideraron algunas propiedades que posee dicha distribución, una de ellas era referente a la distribución de una combinación lineal de variables aleatorias normales. Así pues, sabemos que si $X_1, X_2, X_3, \ldots, X_n$ son v.a. independientes distribuidas normalmente $X_i \sim N\left(\mu_i, \sigma_i^2\right)$ $\forall i = 1, \ldots, n$ y si $a_i, \ldots, a_n$ son números reales, entonces la variable aleatoria:

$$X = \sum_{i=1}^{n} a_i X_i$$

Tiene una distribución normal $X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$

Este resultado nos será de utilidad para obtener la distribución de la media muestral,

a) **Distribución muestral de la media para una población normal con varianza σ^2 conocida.**

Teorema: Si $X_1, X_2, X_3, \ldots, X_n$ es una muestra aleatoria extraída de una población que se distribuye normalmente $X \sim N(\mu, \sigma^2)$ entonces la media muestral se distribuye normalmente con $E(\overline{X}) = \mu$ y $V(\overline{X}) = \frac{\sigma^2}{n}$, es decir $\overline{X} - N\left(\mu, \frac{\sigma^2}{n}\right)$.

Sea $\boxed{\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} \ ,}$

$$E(\overline{X}) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}}\cancel{n}\mu = \mu$$

$$V(\overline{X}) = V\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^{\cancel{2}}}\cancel{n}\sigma^2 = \frac{\sigma^2}{n}$$

Cómo la población se distribuye normalmente con varianza σ^2 conocida entonces por el resultado anterior para cualquier tamaño de la muestra $\overline{X} - N\left(\mu, \frac{\sigma^2}{n}\right)$.

Ejemplo: Las notas de cierto examen se distribuyen normalmente con media 5.8 puntos y desviación estándar de 2.4 puntos. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7 puntos.

X: notas de los estudiantes de cierto examen. $X \sim N\left(5.8, 2.4^2\right)$

Ejemplo: Las notas de cierto examen se distribuyen normalmente con media 5.8 puntos y desviación estándar de 2.4 puntos. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7 puntos.

X: notas de los estudiantes de cierto examen. $X \sim N\left(5.8, 2.4^2\right)$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow N\left(5.8, \underbrace{\frac{2.4^2}{16}}_{0.36}\right)$$

Ejemplo: Las notas de cierto examen se distribuyen normalmente con media 5.8 puntos y desviación estándar de 2.4 puntos. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7 puntos.

X: notas de los estudiantes de cierto examen. $X \sim N\left(5.8, 2.4^2\right)$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow N\left(5.8, \underbrace{\frac{2.4^2}{16}}_{0.36}\right)$$

$$P\left(5 < \overline{X} < 7\right) =$$

Ejemplo: Las notas de cierto examen se distribuyen normalmente con media 5.8 puntos y desviación estándar de 2.4 puntos. Hallar la probabilidad de que la media de una muestra tomada al azar de 16 estudiantes esté comprendida entre 5 y 7 puntos.

X: notas de los estudiantes de cierto examen. $X \sim N\left(5.8, 2.4^2\right)$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow N\left(5.8, \underbrace{\frac{2.4^2}{16}}_{0.36}\right)$$

$$P\left(5 < \overline{X} < 7\right) = P\left(\frac{5-5.8}{0.6} < \underbrace{\frac{\overline{X}-5.8}{0.6}}_{N(0,1)} < \frac{7-5.8}{0.6}\right) = P\left(\frac{5-5.8}{0.6} < Z < \frac{7-5.8}{0.6}\right) = P\left(-1.33 < Z < 2\right) = \ldots$$

Ley de los grandes números

Teorema de Bernoulli generalizado

Sean $x_1, x_2, x_3, \ldots, x_n$ v. a. **independendientes idénticamente distribuidas** con esperanza y varianza finita, $E(x_i) = \mu$ y $V(x_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Sea $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$, para todo $\delta > 0$, $\lim_{n \to \infty} \left\{ P\left(\left| \bar{x} - \mu \right| < \delta \right) \right\} = 1$ o $\lim_{n \to \infty} \left\{ P\left(\left| \bar{x} - \mu \right| \geq \delta \right) \right\} = 0$

Interpretación: Pensando $x_1, x_2, x_3, \ldots, x_n$ como medidas independientes de una característica numérica X que producen el promedio $\bar{x}$. Entonces el límite en probabilidad, de la media muestral para $n \to \infty$ es igual a la media poblacional de la que se extrajo la muestra.

En la práctica, esto significa que si a una muestra "grande" le calculamos su media, este valor numérico puede tomarse como un valor muy próximo a la media de la población.

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza** σ^2 **conocida.**

b) **Distribución muestral de la media para una población normal con varianza** σ^2 **desconocida.**

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

a) **Distribución muestral de la media para una población normal con varianza σ^2 conocida.**

b) **Distribución muestral de la media para una población normal con varianza σ^2 desconocida.**

Antes de analizar este punto analicemos como se distribuye la varianza muestral.

Inferencia Estadística

Distribución de la varianza muestral.

Definamos la varianza muestral como: $$S^2 = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{n-1}$$

Si la población de la cual se extrae la muestra tiene una distribución normal, la variable

$$\left(\sum_{i=1}^{n} \underbrace{\left(\frac{\left(X_i - \overline{X}\right)}{\sigma} \right)^2}_{N(0,1)} \right) \sim \chi^2_{n-1} \qquad \Longrightarrow \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

• La distribución de la varianza muestral no es simétrica: tiene asimetría positiva.

Si se extraen todas las muestras posibles de una población normal y a cada una de ellas se le calcula su varianza, se obtendrá la distribución muestral de la varianza con esta distribución.

Distribución Ji-Cuadrado

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias distribuidas normalmente e independientemente con media $\mu = 0$ y varianza $\sigma^2 = 1$. Entonces la variable aleatoria $\chi^2 = X_1^2 + X_2^2 + \ldots + X_n^2$, tiene una distribución **Ji-Cuadrado** con n grados de libertad, cuya fdp está definida por:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Donde Γ es la función Gamma.

En matemáticas, la función gamma denotada por $\Gamma(z)$, extiende el concepto de factorial a los números complejos. Se define como:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$$

Ejemplo: Una máquina de llenado opera con una varianza de 0.83 gr^2. Si se toma una muestra de 15 unidades, ¿cuál es la probabilidad de tener una varianza muestral superior a 1,249 gr^2?

$$P\left(S^2 \geq 1{,}249\right) = P\left(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}} \geq \frac{14 \times 1{,}249}{0{,}83} \right) = P\left(\chi^2_{14} \geq 21{,}067\right) = 0{,}10$$

g.d.l	0,001	0,005	0,01	0,02	0,025	0,03	0,04	0,05	$\boxed{0,10}$	0,15	0,20	0,25	0,30	0,35	0,40	g.d.l
1	10,828	7,879	6,635	5,412	5,024	4,709	4,218	3,841	2,706	2,072	1,642	1,323	1,074	0,873	0,708	1
2	13,816	10,597	9,210	7,824	7,378	7,013	6,438	5,991	4,605	3,794	3,219	2,773	2,408	2,100	1,833	2
3	16,266	12,838	11,345	9,837	9,348	8,947	8,311	7,815	6,251	5,317	4,642	4,108	3,665	3,283	2,946	3
4	18,467	14,860	13,277	11,668	11,143	10,712	10,026	9,488	7,779	6,745	5,989	5,385	4,878	4,438	4,045	4
5	20,515	16,750	15,086	13,388	12,833	12,375	11,644	11,070	9,236	8,115	7,289	6,626	6,064	5,573	5,132	5
6	22,458	18,548	16,812	15,033	14,449	13,968	13,198	12,592	10,645	9,446	8,558	7,841	7,231	6,695	6,211	6
7	24,322	20,278	18,475	16,622	16,013	15,509	14,703	14,067	12,017	10,748	9,803	9,037	8,383	7,806	7,283	7
8	26,124	21,955	20,090	18,168	17,535	17,010	16,171	15,507	13,362	12,027	11,030	10,219	9,524	8,909	8,351	8
9	27,877	23,589	21,666	19,679	19,023	18,480	17,608	16,919	14,684	13,288	12,242	11,389	10,656	10,006	9,414	9
10	29,588	25,188	23,209	21,161	20,483	19,922	19,021	18,307	15,987	14,534	13,442	12,549	11,781	11,097	10,473	10
11	31,264	26,757	24,725	22,618	21,920	21,342	20,412	19,675	17,275	15,767	14,631	13,701	12,899	12,184	11,530	11
12	32,909	28,300	26,217	24,054	23,337	22,742	21,785	21,026	18,549	16,989	15,812	14,845	14,011	13,266	12,584	12
13	34,528	29,819	27,688	25,472	24,736	24,125	23,142	22,362	19,812	18,202	16,985	15,984	15,119	14,345	13,636	13
$\boxed{14}$	36,123	31,319	29,141	26,873	26,119	25,493	24,485	23,685	21,064	19,406	18,151	17,117	16,222	15,421	14,685	14
15	37,697	32,801	30,578	28,259	27,488	26,848	25,816	24,996	22,307	20,603	19,311	18,245	17,322	16,494	15,733	15

Distribución muestral de la media $\bar{x}$

En esta sección vamos a determinar la distribución muestral de la media solo en el caso en que la población sea normal, tomando en consideración los casos en que la varianza es conocida y la varianza es desconocida.

b) Distribución muestral de la media para una población normal con varianza σ^2 **desconocida.**

Hasta ahora estábamos admitiendo que se conoce la varianza de la población de la que se extrae la muestra, pero esta no sería la situación general, sino que la mayoría de las veces no conocemos la varianza de la población, entonces cómo se dispone de una muestra aleatoria de tamaño n, podemos, calcular la varianza muestral S^2 para estimar la varianza poblacional σ^2 desconocida.

Cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$

depende del tamaño de la muestra:

a) <u>El tamaño de la muestra es grande $n \geq 30$</u>.

Cuando el tamaño de la muestra es grande, es decir $n \geq 30$, la distribución del estadístico $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue siendo aproximadamente $N(0,1)$

b) <u>El tamaño de la muestra es pequeño $n < 30$.</u>

Si el tamaño de la muestra es pequeño, $n < 30$, los valores de la varianza muestral S^2 varían considerablemente de muestra en muestra, pues S^2 disminuye a medida que n aumenta, y la distribución del estadístico ya no sería una distribución normal. Por lo tanto, cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$

se ajusta a una **distribución *t* de Student**, con $n-1$ grados de libertad.

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

a) El tamaño de la muestra es grande $n \geq 30$.

Cuando el tamaño de la muestra es grande, es decir $n \geq 30$, la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$ sigue siendo aproximadamente $N(0,1)$

b) El tamaño de la muestra es pequeño $n < 30$.

Si el tamaño de la muestra es pequeño, $n < 30$, los valores de la varianza muestral S^2 varían considerablemente de muestra en muestra, pues S^2 disminuye a medida que n aumenta, y la distribución del estadístico ya no sería una distribución normal. Por lo tanto, cuando la varianza poblacional σ^2 es desconocida la distribución del estadístico $\frac{\overline{X}-\mu}{S/\sqrt{n}}$

se ajusta a una **distribución *t* de Student**, con $n-1$ grados de libertad.

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$$

Recordemos que si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

$$t_{n-1} \sim \frac{\overbrace{\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}}^{N(0,1)}}{\sqrt{\left[\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}}\right] / (n-1)}} =$$

Recordemos que si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

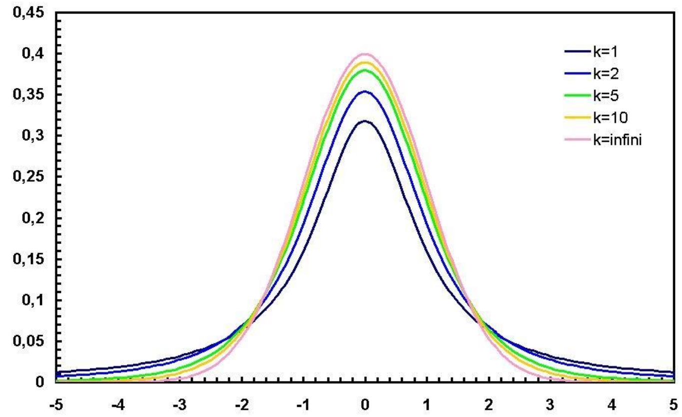
$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$

$$t_{n-1} \sim \frac{\overbrace{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}^{N(0,1)}}{\sqrt{\left[\underbrace{\dfrac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}}\right]\Big/(n-1)}} = \frac{\dfrac{\overline{X}-\mu}{\cancel{\sigma}}\sqrt{n}\,\cancel{\sqrt{n-1}}}{\cancel{\sqrt{n-1}}\;S/\cancel{\sigma}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

Distribución T de Student

Si Z es una variable aleatoria con distribución normal $N(0,1)$, V una variable aleatoria con distribución Ji-Cuadrado con n grados de libertad y además Z y V son independientes, entonces la variable aleatoria:

$$T = \frac{Z}{\sqrt{\frac{V}{n}}} \sim t_n$$



Tiene la siguiente fdp:

$$f(x) = \frac{\Gamma\left(\frac{(n+1)}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(\frac{x^2}{n}+1\right)^{(n+1)/2}} \qquad -\infty < x < \infty$$

La distribución de *T* se llama ahora la **distribución *t* de Student**. El parámetro n representa el número de *grados de libertad*.

DISTRIBUCION DE LA PROPORCIÓN MUESTRAL $\overline{p}$

Consideremos una v.a $X \sim B(n,p)$ donde p es la proporción de "éxito" en la población. Para tamaños grandes de n, n>30, la distribución Binomial se aproxima a una distribución normal.

$$X \cong N\left(np, np(1-p)\right)$$

Definamos el estadístico $\hat{p} = \dfrac{X}{n}$, es el estimador puntual de la proporción poblacional.

La distribución de muestreo de $\hat{p}$ es aproximadamente normal con esperanza p y varianza $\dfrac{p(1-p)}{n}$ con p no cerca de 0 y 1.

Este resultado se obtiene de la aplicación directa del TLC y el Teorema de Bernouli Generalizado.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Parámetros y Estimadores

Parámetro: Es una característica numérica que describe una variable observada en la población. Se calcula sobre la población.

Estadístico o estimador: Es cualquier operación que se hace con la muestra. Por eso es una función de las observaciones contenidas en una muestra.

Si X es una variable aleatoria con distribución de probabilidades o fdp $f(x)$, caracterizada por el parámetro desconocido θ y si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de tamaño n, entonces $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ es un estimador del parámetro θ.

Ejemplos: la media muestral, la proporción muestral y la varianza muestral.

Inferencia Estadística

Estimación puntual: Es el valor numérico que toma un estimador. Se calcula con los datos de la muestra, del cual se espera que estime un parámetro poblacional.

Parámetro Poblacional	Estimador	Estimación
Media μ	$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Varianza σ^2	$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$
Proporción p	$\hat{p} = \frac{X}{n} = \frac{\text{númeroéxitos}}{\text{númeropruebas}}$	$\hat{p} = \frac{x}{n}$
Parámetros poblacionales, estimadores y estimaciones		

Estimación de parámetros

Puede haber varios estimadores puntuales diferentes para un mismo parámetro. Por ejemplo, si deseamos estimar la media poblacional, podríamos considerar:

$$\hat{\mu}_1 = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{\mu}_2 = \frac{X_1 + X_n}{2}$$

Teniendo en cuenta esto:

- **¿Qué características queremos que posea un buen estimador?**
- **¿Cómo decimos que un estimador es mejor que otro?**

Propiedades de los estimadores

Nuestro objetivo ahora será dar algunas propiedades deseables de los estimadores puntuales, con el fin de poder conocer la bondad de los mismos.

Propiedad de insesgadura:

Si tenemos un gran número de muestras de tamaño n y obtenemos el valor del estimador en cada una de ellas, sería deseable que la media de todas estas estimaciones coincidiera con el valor medio de la población . Se dice que un estimador es insesgado si su esperanza matemática coincide con el valor del parámetro a estimar.

Formablemente: Un estimador $\hat{\theta}$ es un **estimador insesgado** del parámetro desconocido θ si $E\left(\hat{\theta}\right)=\theta$. Es decir, la media de su distribución muestral es el parámetro.

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Estimación de parámetros

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Formablemente: Un estimador $\hat{\theta}$ es un **estimador insesgado** del parámetro desconocido θ si $E(\hat{\theta}) = \theta$. Es decir, la media de su distribución muestral es el parámetro.

<u>Ejemplo</u>:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

Ejemplo:

La media muestral $\overline{X}$ y la varianza muestral S^2 son estimadores insesgados de la media poblacional μ y la varianza poblacional σ^2, respectivamente.

Dem/

Si $X_1, X_2, X_3, \ldots, X_n$ son v. a. independientes idénticamente distribuidas (iid) con esperanza y varianza finita, $E(X_i) = \mu$ y $V(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Veamos en primer lugar que $\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n}$ es un estimador insesgado de μ, es decir de la media poblacional.

$$E(\overline{X}) = E\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{\cancel{n}} \cancel{n} \mu = \mu$$

Estimación de parámetros

Veamos ahora que $S^2 = \sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n-1}$ es un estimador insesgado de σ^2, es decir de la varianza poblacional.

$$\boxed{E\left(S^2\right)} = E\left(\sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n-1}\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} \left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)\right) =$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} -2X_i\overline{X} + \sum_{i=1}^{n} \overline{X}^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}\underbrace{\sum_{i=1}^{n} X_i}_{n\overline{X}} + n\overline{X}^2\right) =$$

$$= \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - 2n\overline{X}\,\overline{X} + n\overline{X}^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - 2n\overline{X}^2 + n\overline{X}^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right) =$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} E\left(X_i^2\right) - E\left(n\overline{X}^2\right)\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} E\left(X_i^2\right) - nE\left(\overline{X}^2\right)\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n}\left(\mu^2 + \sigma^2\right) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right) =$$

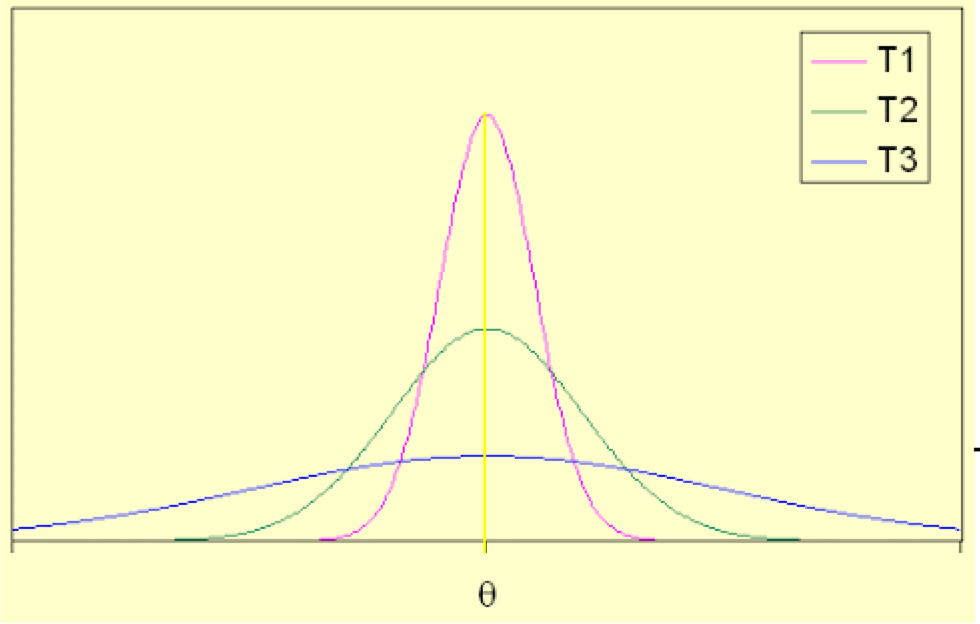
$$= \frac{1}{n-1}\left(\cancel{n\mu^2} + n\sigma^2 - \cancel{n\mu^2} - \cancel{n}\left(\frac{\sigma^2}{\cancel{n}}\right)\right) = \frac{1}{n-1}\left(n\sigma^2 - \sigma^2\right) = \frac{1}{\cancel{n-1}}\sigma^2\,\cancel{(n-1)} = \boxed{\sigma^2}$$

Estimación de parámetros

Propiedad de eficiencia:

Se dice que los estimadores son eficientes cuando generan una distribución muestral con el mínimo error estándar, es decir, entre dos estimadores insesgados de un parámetro dado es más eficiente el de menor varianza.

Formalmente: Un estimador ***insesgado*** $\hat{\theta}_1$ es más eficiente que otro $\hat{\theta}_2$ Si son insesgados de θ y la varianza de $\hat{\theta}_1$ es menor que la varianza de $\hat{\theta}_2$.



$$E(T_1) = \theta$$

$$E(T_2) = \theta$$

$$E(T_3) = \theta$$

T_1 es el de varianza mínima

Estimación de parámetros

Estimador consistente:

Un estimador se dice ***consistente*** cuando su valor tiende hacia el verdadero valor del parámetro a medida que aumenta el tamaño de la muestra. Es decir, la probabilidad de que la estimación sea el verdadero valor del parámetro tiende a 1.

Propiedad de suficiencia:

Se dice de un estimador que es ***suficiente*** cuando es capaz de extraer de los datos toda la información importante sobre el parámetro.

Un estimador es ***suficiente*** si utiliza toda la información de la muestra.

Estimación de parámetros

Estimador consistente:

Un estimador se dice ***consistente*** cuando su valor tiende hacia el verdadero valor del parámetro a medida que aumenta el tamaño de la muestra. Es decir, la probabilidad de que la estimación sea el verdadero valor del parámetro tiende a 1.

Propiedad de suficiencia:

Se dice de un estimador que es ***suficiente*** cuando es capaz de extraer de los datos toda la información importante sobre el parámetro.

Un estimador es ***suficiente*** si utiliza toda la información de la muestra.

$$\hat{\mu}_1 = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \hat{\mu}_2 = \frac{X_1 + X_n}{2}$$