

## Objetivos

- ✓ Recordar los conocimientos elementales de las herramientas estadísticas que se aplican en el Control de Calidad.
- ✓ En base al conocimiento de los modelos probabilísticos más utilizados y teniendo en cuenta el planteamiento del problema y el estudio descriptivo previo, elegir un modelo de probabilidad, estimar sus parámetros, chequear que el modelo de probabilidad ajustado a los datos es adecuado y que se verifican las hipótesis supuestas en el estudio.

## Epígrafes

- ✓ Estadística Descriptiva:
  - ✓ Tipos de datos. Representación: tablas, gráficos,
  - ✓ Medidas de centralización, de posición, de dispersión, simetría
- ✓ Inferencia Estadística
  - ✓ Distribuciones discretas
  - ✓ Distribuciones continuas:
    - ✓ Distribución normal
    - ✓ Teoría Límite Central

## Estadística aplicada al Control de Calidad

La **Estadística** es la Ciencia que estudia los **procedimientos** destinados a la recogida, resumen, análisis e interpretación de un conjunto de datos, así como los conducentes a la obtención de deducciones científicas a partir de ellos.

### Ramas de la Estadística

- ✓ **Estadística Descriptiva:** Se ocupa de los sucesos que ya han ocurrido, es decir de los datos. *Ejemplos básicos de descriptores numéricos son la **media** y la **desviación estándar**. Resúmenes gráficos incluyen varios tipos de tablas, figuras y gráficos.*
- ✓ **Inferencia Estadística:** Se dedica a la generación de los **modelos, inferencias y predicciones** asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para **modelar patrones** en los datos y extraer inferencias acerca de la población bajo estudio, es decir **que generaliza a toda la población la información obtenida a partir de una muestra reducida**.

## Estadística Descriptiva

La interpretación de estos datos se realiza a través de diversas técnicas y cálculos:

- ✓ **Organización de los datos en tablas:** Es el primer paso para poder analizar los datos y realizar la interpretación de éstos. Son la base para el resto del estudio.
- ✓ **Representación gráfica de los datos:** Nos da una información visual sobre la distribución de éstos datos. Los distintos tipos de gráficos dependerán de las variables a estudiar.
- ✓ **Medidas de centralización:** Nos resumen el conjunto de datos en un único valor, localizado aproximadamente en el centro de la distribución.
- ✓ **Medidas de posición:** Complementan las medidas de centralización y nos ofrecen valores relativos a otras posiciones no centrales.
- ✓ **Medidas de dispersión:** Ofrecen información sobre la homogeneidad de los datos y la dispersión de éstos. También informan sobre el grado de representatividad de las medidas de centralización.
- ✓ **Medidas de simetría y curtosis:** Nos dan valores relativos a la forma de la distribución de datos.
- ✓ **Identificación y tratamiento de valores atípicos:** En situaciones en las que aparecen valores atípicos, nos establece criterios para su localización y algunas correcciones en el cálculo de los parámetros estadísticos.

## Tipos de Variables

Las variables estadísticas se clasifican en dos tipos:

- ✓ **Variables cualitativas o atributos:** Los valores que toma ***no son de tipo numérico*** (por ej, *sexo, estado civil, nivel de estudios,...*). Se pueden clasificar a su vez:
  - **Ordenables**: Existe un criterio lógico para realizar una ordenación de los distintos valores: ***p.ej; nivel de estudios.***
  - **No ordenables**: No existe un criterio que permita colocar unos valores antes que otros: ***p.ej; nacionalidad.***
  
- ✓ **Variables cuantitativas:** Los distintos valores que puede tomar ***son de tipo numérico*** (por ej., *número de hijos, salario, edad,...*). Dentro de éstas podemos distinguir dos tipos:
  - **Discretas**: Toman valores aislados, es decir, entre dos valores consecutivos no existen valores intermedios: ***p.ej.; número de hijos.***
  - **Continuas**: Entre dos valores siempre es posible encontrar un valor intermedio: ***p.ej; peso, estatura.***

## Tablas de frecuencia

Realizado el experimento y tomados los datos, **se ordenan** en una **tabla**, denominada de **frecuencias**, que permite ver cómo están distribuidos, su construcción depende del tipo de variable.

### Tipos de Frecuencia:

- ✓ **Frecuencia absoluta:** Número de veces que se presenta el valor de la variable. Habitualmente se representa como  $n_i$ .
- ✓ **Frecuencia relativa:** Cociente entre la frecuencia absoluta y el número total de casos. La representaremos como  $f_i$ . La frecuencia relativa nos da el tanto por uno relativo a ese valor.
- ✓ **Porcentaje:** Resultado de multiplicar por 100 la frecuencia relativa. Representado por  $p_i$ , indica el tanto por ciento de la población que corresponde a ese valor de la variable.
- ✓ **Frecuencia absoluta acumulada:** Suma de las frecuencias absolutas de todos los valores menores o iguales que el correspondiente valor de la variable. Representada como  $N_i$ .
- ✓ **Frecuencia relativa acumulada:** Cociente de la frecuencia absoluta acumulada y el número total de casos. También se puede obtener sumando las frecuencias relativas de todos los valores menores o iguales que el correspondiente de la variable. La representaremos como  $F_i$ .
- ✓ **Porcentaje acumulado:** Resultado de multiplicar por 100 la frecuencia relativa acumulada. Lo representamos como  $P_i$ , y también los podríamos obtener acumulado los porcentajes.

[Ejemplo numérico](#)

## Gráficos Estadísticos

### DIAGRAMA DE BARRAS

#### Ejemplo-barra

- ✓ Se utiliza para variables cualitativas y **cuantitativas discretas**.
- ✓ Habitualmente se representa sobre el eje horizontal las distintas modalidades de la variable, sobre el eje vertical se sitúan las frecuencias (absolutas, relativas, porcentajes o acumuladas).
- ✓ Partiendo del eje horizontal, sobre cada valor de la variable se levanta una barra de altura correspondiente a la frecuencia.
- ✓ Debemos tener en cuenta que la escala en el eje vertical debe comenzar con el valor cero.
- ✓ En ocasiones podemos ver representados los gráficos horizontalmente.

### DIAGRAMA DE SECTORES

#### Ejemplo-sectores

- ✓ Utilizado habitualmente para variables cualitativas (cuantitativas discretas si presentas pocos valores)
- ✓ El gráfico consiste en un círculo sobre el que se van trazando distintos sectores circulares de amplitud proporcional a la frecuencia absoluta.
- ✓ Para hallar la amplitud del sector circular multiplicamos  $360^\circ$  por la frecuencia relativa:  
 $360^\circ \cdot f_i = 360^\circ \cdot n_i / N$ .
- ✓ También podemos encontrarnos este gráfico representado sobre media circunferencia, el cálculo sería análogo al anterior pero cambiando la amplitud de  $360^\circ$  correspondiente al círculo completo por  $180^\circ$

## Gráficos Estadísticos

### HISTOGRAMA

### Ejemplo-Histograma

- ✓ Se utiliza con variables **cuantitativas continuas**, o agrupadas en intervalos, representando en el eje X los intervalos de clase y levantando rectángulos de base la longitud de los distintos intervalos y de altura tal que el área sea proporcional a las frecuencias representadas.
- ✓ El polígono de frecuencias se obtiene uniendo los puntos medios de las bases superiores de los rectángulos.
- ✓ Los histogramas permiten compara datos de una forma rápida (basta mirar la gráfica).

### OTRAS REPRESENTACIONES SON:

- ✓ Pictogramas
- ✓ Pirámides de Población
- ✓ Cartogramas
- ✓ Diagramas de Pareto
- ✓ Diagramas de Causas y Efectos

## Medidas de Tendencia central

### MEDIA

- ✓ Se llama media aritmética (habitualmente la denominamos media) a la suma de todos los valores de la variable dividida por el número total de valores.
- ✓ Para su cálculo, se utiliza la siguiente expresión:

$$\bar{x} = \frac{\sum_{i=1}^n x_i n_i}{N}$$

## Medidas de Dispersión

### VARIANZA

- ✓ La varianza va a ser la media del cuadrado de la distancia de los valores de los datos a la media, así evitamos que se anulen diferencias positivas y negativas.
- ✓ Para el cálculo se suele aplicar la expresión:

$$S^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{N}$$

### DESVIACIÓN TÍPICA

- ✓ Definimos la desviación típica como la raíz cuadrada de la varianza.
- ✓ El hecho de realizar la raíz cuadrada tiene por objetivo que la medida de dispersión esté expresada en la misma unidad que la variable.
- ✓ Para su cálculo, se utiliza la siguiente expresión:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 n_i}{N}}$$



## Inferencia Estadística

Comprende los **métodos y procedimientos** para deducir propiedades (hacer **inferencias**) de una población, a partir de una pequeña parte de la misma (**muestra**).

- ✓ Se definen con precisión **la población**, la **característica a estudiar**, las variables, etcétera.

### Elaboración de un modelo

- ✓ Se establece un modelo teórico de comportamiento de la variable de estudio. Los posibles modelos son **distribuciones de probabilidad**.

### Extracción de la muestra

- ✓ Se usa alguna **técnica de muestreo** o un **diseño experimental** para obtener información de una pequeña parte de la población.

### Tratamiento de los datos

- ✓ Se tabulan los datos y se calculan los valores, como la **media**, la **varianza**, etc.

### Estimación de los parámetros

- ✓ Con determinadas **técnicas** se realiza una predicción sobre cuáles podrían ser los **parámetros de la población**.

### Contraste de hipótesis

- ✓ Los **contrastes de hipótesis** son técnicas que permiten simplificar el modelo matemático.

### Conclusiones

- ✓ Las **conclusiones** obtenidas en este punto pueden servir para **tomar decisiones** o **hacer predicciones**.

## Variable Aleatoria

### Definición

Una variable es aleatoria si su valor está determinado por el azar, es decir, se sabe qué valores puede tomar la variable pero no se tiene certeza de su ocurrencia, solo se sabe que puede ocurrir con cierta probabilidad.

### Tipos de Variable Aleatoria

VARIABLES ALEATORIAS DISCRETAS

Diremos que una variable aleatoria es discreta si su recorrido es finito o infinito numerable, es decir, si sólo puede tomar un conjunto numerable de valores aislados.

Generalmente, este tipo de variables van asociadas a experimentos en los cuales se cuenta el número de veces que se ha presentado un suceso o donde el resultado es una puntuación concreta.

#### Ejemplos:

- ✓ El número de automóviles vendidos en un mes.
- ✓ El número de accidentes ocurridos en una determinada semana en una planta de manufactura, también determinada
- ✓ El número de clientes esperando servicio en la caja de un supermercado.
- ✓ El número de tubos electrónicos de televisión producidos en una hora determinada.

## Distribuciones discretas

Se denomina distribución de variable discreta a aquella cuya función de probabilidad sólo toma valores positivos en un conjunto de valores de  $X$  finito o infinito numerable. A dicha función se le llama función de masa de probabilidad. En este caso la distribución de probabilidad es el sumatorio de la función de masa, por lo que tenemos entonces que:

$$F(x) = P(X \leq x) = \sum_{k=-\infty}^x f(k)$$

Las distribuciones de variable discreta más importantes son:

[Distribución binomial](#)

[Distribución binomial negativa](#)

[Distribución Poisson](#)

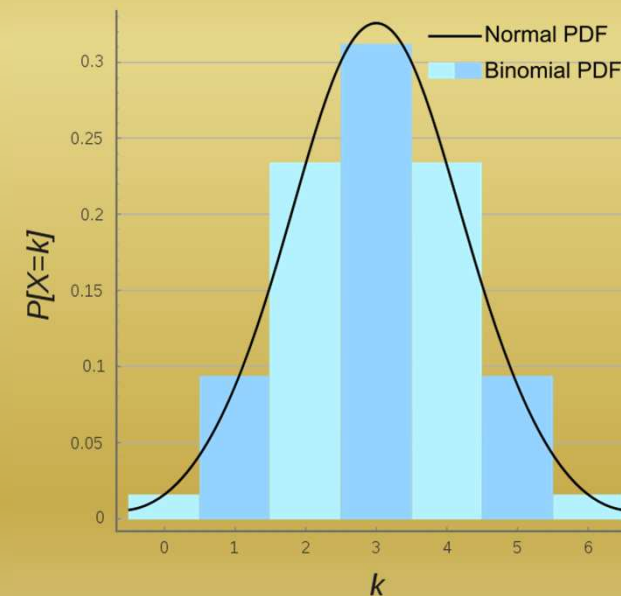
[Distribución geométrica](#)

[Distribución hipergeométrica](#)

[Distribución de Bernoulli](#)

[Distribución Rademacher](#), que toma el valor 1 con probabilidad  $1/2$  y el valor -1 con probabilidad  $1/2$ .

[Distribución uniforme discreta](#), donde todos los elementos de un conjunto finito son equiprobables.



## Tipos de Variable Aleatoria

### Variables aleatorias continuas

Generalmente, se corresponden con variables asociadas a experimentos en los cuales la variable medida puede tomar cualquier valor en un intervalo de la recta.

#### Ejemplos:

- ✓ El tiempo necesario para completar el ensamblaje de un artículo en una planta.
- ✓ La cantidad de petróleo bombeado cada hora en un pozo.
- ✓ La cantidad en miligramos de monóxido de carbono contenido en un metro cúbico de aire.
- ✓ La cantidad de energía eléctrica producida en una planta hidroeléctrica en un día.

## Distribución de Probabilidad

La **distribución de probabilidad** es un **modelo matemático** que relaciona cada valor de la variable aleatoria con la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los eventos rango de valores de la variable aleatoria.

### Distribuciones continuas

En el caso de variables continua la distribución es la integral de la función de densidad

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

Las distribuciones de variable continua más importantes son:

[Distribución ji cuadrado](#)

[Distribución exponencial](#)

[Distribución t de Student](#)

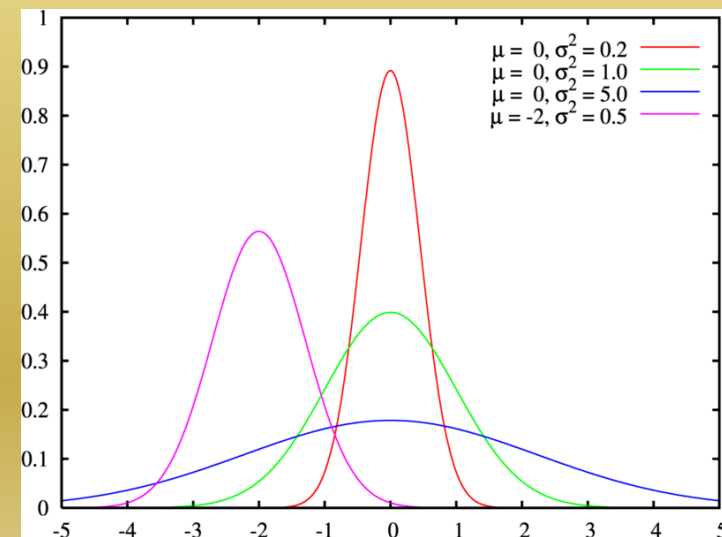
[Distribución normal](#)

[Distribución Gamma](#)

[Distribución Beta](#)

[Distribución F](#)

[Distribución uniforme \(continua\)](#)



## Distribuciones continuas Importantes

### Distribución Normal

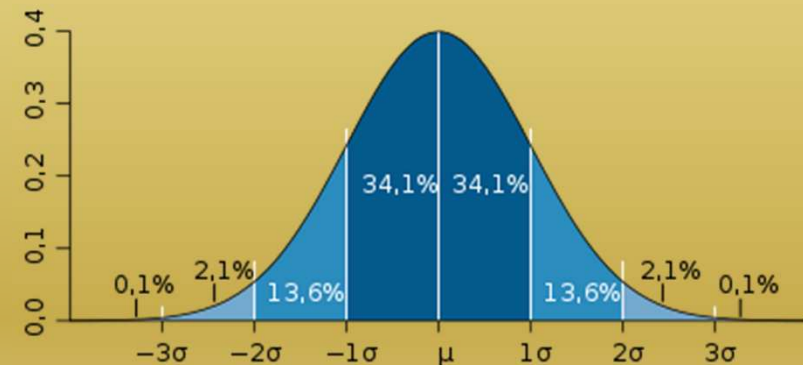
Se dice que una variable aleatoria continua  $X$  sigue una distribución normal de parámetros  $\mu$  y  $\sigma$  y se denota  $X \sim N(\mu, \sigma)$  si su función de densidad está dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

donde  $\mu$ , es la media y  $\sigma$  es la desviación típica ( $\sigma^2$  es la varianza).

#### Propiedades de la distribución normal :

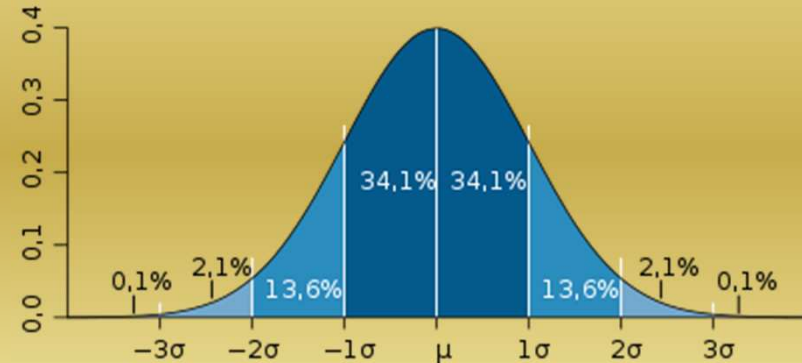
- ✓ Es simétrica respecto de su media,  $\mu$ ;
- ✓ Distribución de probabilidad alrededor de la media en una distribución  $N(\mu, \sigma)$ .
- ✓ La moda y la mediana son ambas iguales a la media,  $\mu$ ;
- ✓ Los puntos de inflexión de la curva se dan para  $x = \mu - \sigma$  y  $x = \mu + \sigma$ .



## Distribuciones continuas Importantes

### Propiedades de la distribución normal :

- ✓ Distribución de probabilidad en un entorno de la media:
  - en el intervalo  $[\mu - \sigma, \mu + \sigma]$  se encuentra comprendida, aproximadamente, el 68,26% de la distribución;
  - en el intervalo  $[\mu - 2\sigma, \mu + 2\sigma]$  se encuentra, aproximadamente, el 95,44% de la distribución;
  - por su parte, en el intervalo  $[\mu - 3\sigma, \mu + 3\sigma]$  se encuentra comprendida, aproximadamente, el 99,74% de la distribución.
  
- ✓ No existe una única distribución normal, sino una familia de distribuciones con una forma común, diferenciadas por los valores de su media y su varianza. De entre todas ellas, la más utilizada es la **distribución normal estándar**, que corresponde a una distribución de media 0 y varianza 1.



$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right); \quad -\infty < z < \infty$$

- ✓ La distribución normal estándar está tabulada (habitualmente en la forma de el valor de la función de distribución  $\Phi$ ). [Tabla](#)

- ✓ Otras **distribuciones normales** pueden obtenerse como transformaciones simples, como se describe más abajo, de la **distribución normal estándar**. De este modo se pueden usar los valores tabulados de la función de distribución normal estándar para encontrar valores de la función de distribución de cualquier otra distribución normal.
- ✓ Por tanto, a partir de cualquier **variable X** que siga una **distribución  $N(\mu, \sigma)$** , se puede obtener otra **característica Z** con una **distribución normal estándar**, sin más que efectuar la transformación:

$$Z = \frac{X - \mu}{\sigma}$$

- ✓ Esto nos permite a partir de las **tablas tabuladas** para la **distribución normal estándar** obtener de modo sencillo la **probabilidad de observar** un dato menor o igual a un cierto valor  $z$ , y que permitirán resolver preguntas de probabilidad acerca del comportamiento de variables de las que se sabe o se asume que siguen una distribución aproximadamente normal.

$$\Pr(X \leq x) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{x - \mu}{\sigma\sqrt{2}}\right)\right).$$

### Ejemplo



## Teorema Central del Límite

- ✓ Este teorema nos dice que si una muestra es lo bastante grande (generalmente cuando el tamaño de la muestra ( $n$ ) supera los 30), sea cual sea la distribución de la media de la muestra, seguirá aproximadamente una distribución normal.

### Enunciado

Sea  $X_1, X_2, \dots, X_n$  un conjunto de variables aleatorias, independientes e idénticamente distribuidas con media  $\mu$  y varianza  $\sigma^2$  distinta de cero. Sea

$$S_n = X_1 + \dots + X_n$$

Entonces

$$\lim_{n \rightarrow \infty} \Pr \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right) = \Phi(z)$$

- ✓ Por tanto, dada cualquier variable aleatoria, si extraemos muestras de tamaño  $n$  ( $n > 30$ ) y calculamos los promedios de las muestras, dichos promedios seguirán una distribución normal. Además, la media será la misma que la de la variable de interés, y la desviación estándar de la media de la muestra será aproximadamente el error estándar.

### Ejemplo