# INFERENCIA ESTADÍSTICA

La **Estadística Inferencial** se ocupa de inferir o deducir las características de la población a partir de las características de las muestras.

Distinguiremos :

- **Parámetros poblacionales**. Son los índices centrales y de dispersión que definen a una población.
- **Estadísticos muéstrales**. Son los índices centrales y de dispersión que definen a una muestra.

Población : Media $\mu$ , Desviación típica: $\sigma$

Muestra: Media $\bar{x}$ Desviación típica: s

1.-Los sueldos, en euros, de los empleados de una fábrica se distribuyen
N (1 200, 400). Se elige al azar una muestra de 25 de ellos. ¿Cuál es la probabilidad de que su sueldo sea superior a 1500 €?

2. Sabemos que la desviación típica de los pesos de los pollos adultos es 300 g.
Queremos estimar el peso medio de los pollos adultos de una granja con un error menor que 100 g, y para ello, tomamos una muestra de 50 individuos.
¿Con qué nivel de confianza podremos realizar la estimación?

**MUESTREO:**

**Simple**: Es el más sencillo y sirve de base para todos los demás. Se parte de un listado de los elementos de la población y se seleccionan aleatoriamente n de ellos que constituyen la muestra. La elección se puede hacer asignándoles un número a cada elemento de la población y utilizar una urna o una tabla de números aleatorios.

EJEMPLO ;
Una ganadería tiene 3 000 vacas. Se quiere extraer una muestra de 120. Explica cómo se obtiene la muestra:
Mediante muestreo aleatorio simple.

— Se numeran las vacas del 1 al 3000.
— Se sortean 120 números de entre los 3000.
— La muestra estará formada por las 120 vacas a las que correspondan los números obtenidos.

**Sistemático**: Es una variante del simple. Conocidos N (tamaño de la población) y n (tamaño de la muestra), se divide $\frac{N}{n}$ y la parte entera del cociente k, nos indica que hemos de seleccionar los elementos de k en k, eligiendo al azar previamente el primero de ellos entre los k primeros elementos. La ventaja es que solo hay que determinar al azar un elemento.

Con el ejemplo anterior:

Coeficiente de elevación: h = 3000/120= 25
— Se sortea un número del 1 al 25. Supongamos que sale el 9.
— Las vacas seleccionadas para la muestra serían las que correspondieran a los números 9, 34, 59, 84, 109, …, 2984.

**Estratificado**: Se divide la población en subgrupos o estratos homogéneos en los cuales se toman muestras aleatorias simples. La ventaja es que todas las partes en que la población se divide estarán representadas adecuadamente.

Si $N_1, \ldots, N_k$ es el nº de elementos en cada estrato ($N_1 + \ldots + N_k = N$) se elige el tamaño de la muestra $n_i$ ($n_1 + \ldots + n_k = n$) de forma que

$$\frac{n_1}{N_1} = \ldots = \frac{n_k}{N_k} = \frac{n}{N}$$

Este método recibe el nombre de muestreo estratificado proporcional. Dentro de cada estrato se puede aplicar el muestreo simple o sistemático para escoger los $n_i$ elementos de la muestra.

Una ganadería tiene 2 000 vacas. Son de distintas razas: 853 de A, 512 de B, 321 de C, 204 de D y 110 de E.

Queremos extraer una muestra de 120:

a) ¿Cuántas hay que elegir de cada raza para que el muestreo sea estratificado con reparto proporcional?

b) ¿Cómo ha de ser la elección dentro de cada estrato?

a) Llamamos $n_1$ al número de vacas que debemos elegir de la raza A, $n_2$ al de raza B, $n_3$ al de C, $n_4$ al de D y $n_5$ al de E.

Ha de cumplirse que:

$$\frac{120}{2000} = \frac{n_1}{853} = \frac{n_2}{512} = \frac{n_3}{321} = \frac{n_4}{204} = \frac{n_5}{110}$$

Así, obtenemos:

$$n_1 = 51{,}18 \quad n_2 = 30{,}72 \quad n_3 = 19{,}26 \quad n_4 = 12{,}24 \quad n_5 = 6{,}6$$

La parte entera de estos números suma:

51 + 30 + 19 + 12 + 6 = 118. Faltan 2 para llegar a 120.

Por tanto, debemos elegir:

51 vacas de raza A, 31 vacas de B, 19 de C, 12 de D y 7 de E.

USO DE CALCULADORA PARA OBTENER NÚMEROS ALEATORIOS.

Tecla Ran# (secundaria)

Por ejemplo:

Ran# [0.226] × 95 + 1 = [22.47] → 22

Ran# [0.048] × 95 + 1 = [5.56] → 5

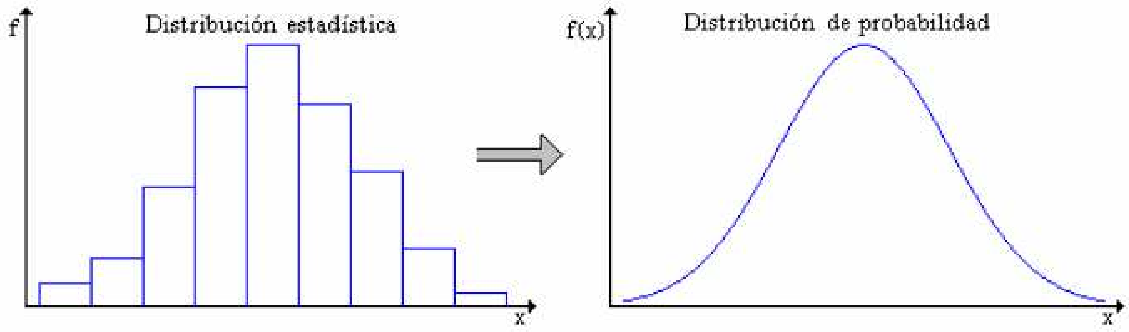
Ran# [0.277] × 95 + 1 = [27.315] → 27

Ran# [0.842] × 95 + 1 = [80.99] → 80

**DISTRUBUCIÓN NORMAL**

**Distribuciones de probabilidad**

Las distribuciones de probabilidad de variable continua son distribuciones teórica e idealizadas de las distribuciones estadísticas de variable continua.

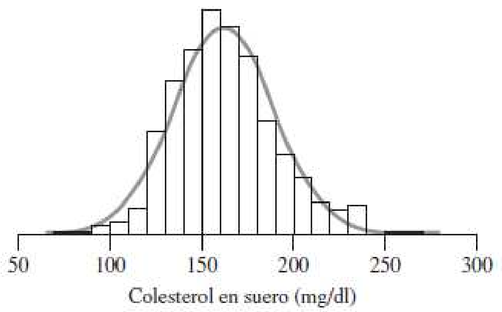


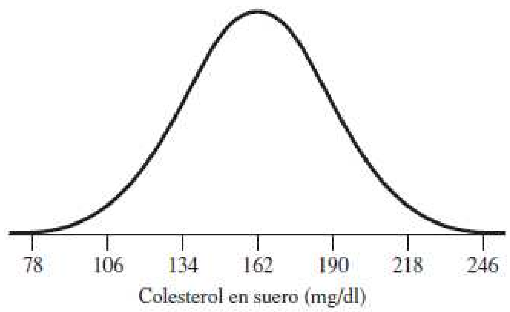

Las distribuciones de probabilidad de variable continua se definen mediante una función($y = f(x)$) denominada función de probabilidad o **función de densidad**.

Las funciones de densidad de variable aleatoria deben de cumplir una serie de propiedades:

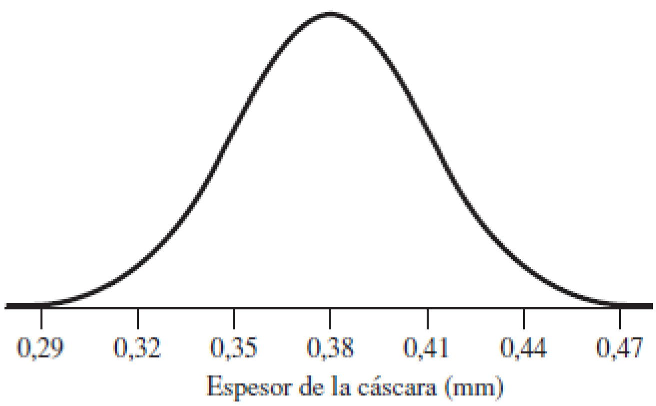
- Deben ser positivas($f(x) > 0$) para cualquier valor de la variable aleatoria en su campo de definición.
- Debe ser una función normalizada, es decir, el área encerrada por la curva en su campo de existencia y el eje de abcisas debe ser la unidad

Ejemplo 1: La concentración de colesterol en la sangre se puede aproximar bastante bien mediante una curva normal de media 162 mg/dl y desviación típica 28 mg/dl.




Ejemplo 2: Espesor de cáscaras de huevo se distibuye según una normal de media 0,38 mm y desviación típica 0,03 mm.

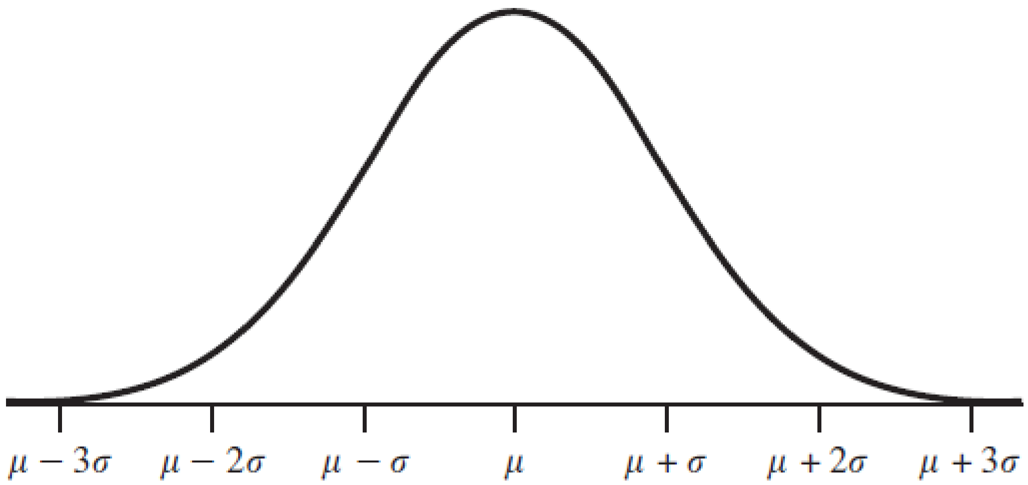


Cada curva normal concreta se caracteriza por su media y su desviación típica. Si la variable *X* sigue una distribución normal de media μ y desviación típica σ entonces es habitual escribir que **X es N (μ , σ)**
Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son: la estatura; el efecto de un fármaco, l consumo de cierto producto por un mismo grupo de individuos, el coefiente intelectual, etc.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado a la media de la distribución .
Esta curva se conoce como **campana de Gauss.**

La expresión de su función de densidad :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

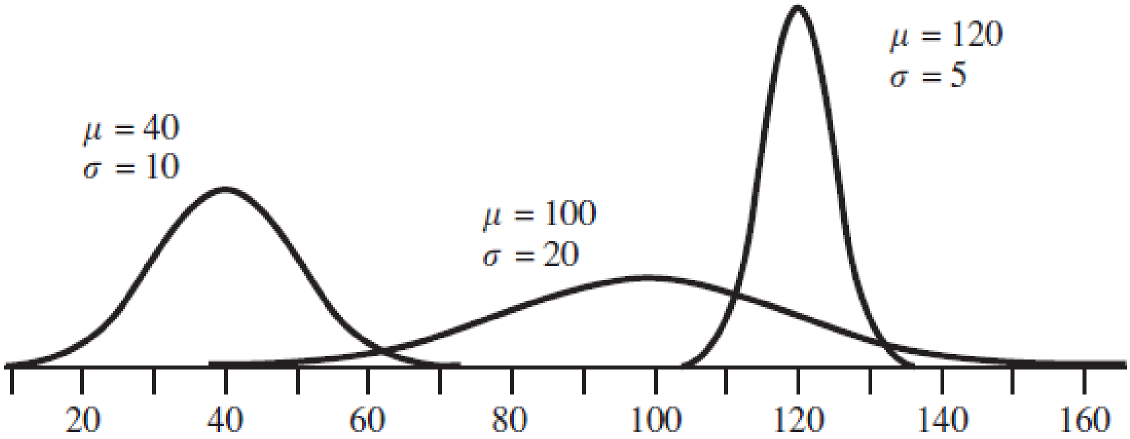



## Una propiedad de la distribución normal

Si $X \sim N(\mu, \sigma)$, entonces

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &\approx 0{,}683 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &\approx 0{,}955 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &\approx 0{,}997 \end{aligned}$$

**DISTRIBUCIONES NORMALES :**



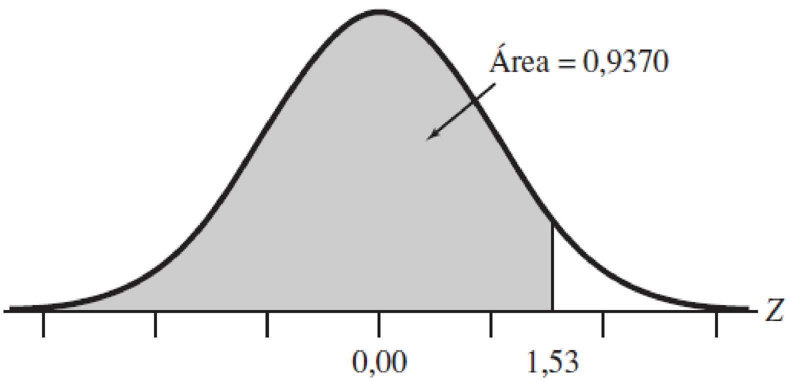

**TIPIFICAR**

Transformar la variable **X** que sigue una distribución **N(μ, σ)** en otra variable **Z** que siga una distribución **N(0, 1)**, se aplica el cambio de variable:

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

**DISTRIBUCIÓN NORMAL ESTÁNDAR : N (0, 1)**

La variable Z se denomina normal tipificada o normal estándar y su distribución sigue una curva normal de media cero y desviación típica uno.
La función de densidad es simétrica respecto x=0




Por ejemplo, para $z_{1,53}$, el área tabulada es 0,9370.