



VNIVERSITAT DE VALÈNCIA

ANEXO: CONCEPTOS FUNDAMENTALES DE ESTADÍSTICA

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

- ▶ Universo: totalidad de individuos que pueden ser representados por una determinada característica a estudiar
- ▶ Población: Conjunto sobre el que se realizará el estudio estadístico (contenido, lugar y tiempo determinado), es una parte del universo.
- ▶ Individuo o Unidad Estadística: Cada uno de los elementos que componen la población.
- ▶ Muestra: Conjunto representativo de la población pero más pequeño que esta.
- ▶ Valor: Resultados que podemos obtener.

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

- ▶ Variable: Característica observable que varía entre los diferentes individuos
- ▶ Tipos de variables:
 - ▶ Cualitativas, categóricas o de atributos
 - ▶ Nominal
 - ▶ Ordinal
 - ▶ Cuantitativas
 - ▶ Discreta
 - ▶ Continua

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

- ▶ Dato: Valor de la variable para un individuo concreto
- ▶ Parámetro: Cantidad calculada sobre una población
- ▶ Estadístico: Cantidad calculada sobre una muestra. Puede ser representativo del parámetro
- ▶ Censo: Listado de características de la población

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

- ▶ La esperanza matemática:

$$\mu = E(x) = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n x_i^* \cdot f_i}{n}$$

- ▶ Propiedades:

- ▶ Si X es siempre positiva, entonces siempre lo es E(X).
- ▶ La esperanza matemática de una constante es igual a esa misma constante
- ▶ Si X está delimitada por dos números reales, a y b, tal que: $a < X < b$, entonces también lo está su media: $a < E(X) < b$
- ▶ Linealidad: Si $Y = a + bX$, $E(Y) = a + b \cdot E(X)$

- ▶ Otras medias:

- ▶ Ponderada
- ▶ Geométrica
- ▶ Armónica
- ▶ Cuadrática

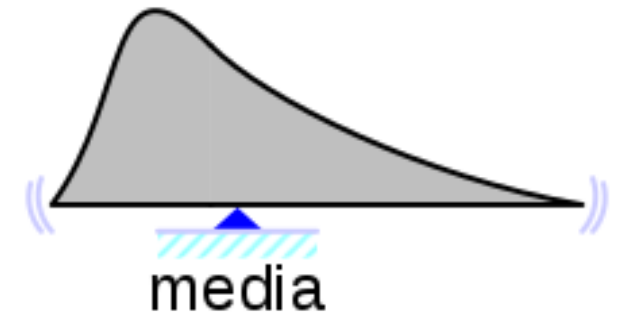
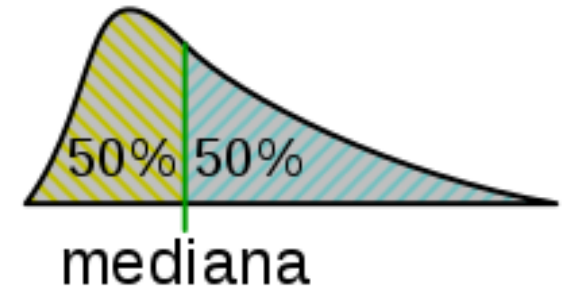
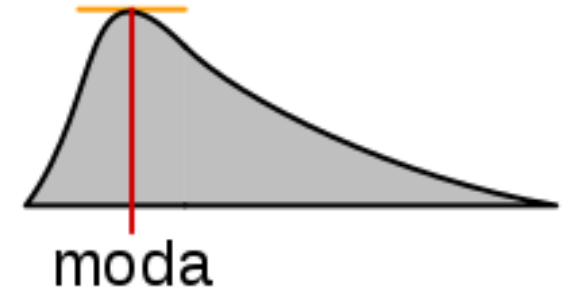
ESTADÍSTICOS Y CONCEPTOS BÁSICOS

► Mediana:

- Representa el valor de la variable de posición central en un conjunto de datos ordenados.
- Uno de los estimadores más importantes
- Si estamos con valores ordinales, es más representativa que la media

► Moda y multi-moda

- Moda: valor más repetido
- Puede haber más de una



ESTADÍSTICOS Y CONCEPTOS BÁSICOS

► La varianza y la desviación típica

$$\sigma^2 = V(x) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$$

$$\sigma = \sqrt{V(x)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2}$$

Propiedades:

$$V(X) \geq 0$$

$$V(aX+b) = a^2V(X)$$

$$V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X,Y)$$

$$V(X-Y) = V(X) + V(Y) - 2\text{Cov}(X,Y)$$

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

► Cuantiles: mediana, cuartiles, deciles y percentiles

- Cuantil: El cuantil de orden p de una distribución (con $0 < p < 1$) es el valor de la variable x_p que marca un corte de modo que una proporción p de valores de la población es menor o igual que x_p
- Mediana: valor que se encuentra en la posición central. Divide la distribución de datos en 2 partes iguales
- Cuartil: valores dividen la distribución de datos en 4 partes iguales
- Decil: valores dividen la distribución de datos en 10 partes iguales
- Percentil: valores dividen la distribución de datos en 100 partes iguales
- Rango intercuartílico: $Q_3 - Q_1$

► Correlación $x \sim y$

- Mide el grado de dependencia entre dos variables
- Generalmente se mide la dependencia lineal

ESTADÍSTICOS Y CONCEPTOS BÁSICOS

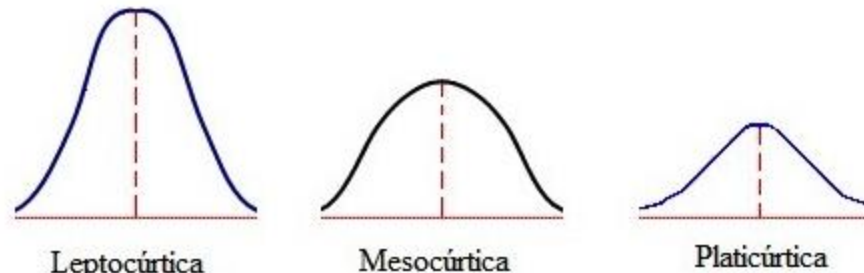
► Asimetría y curtosis:

► Asimetría:

- A la izquierda: $\text{media} < \text{mediana} < \text{moda}$
- A la derecha: $\text{media} > \text{mediana} > \text{moda}$
- Simétrica: $\text{media} = \text{mediana} = \text{moda}$

► Curtosis

- Habría que calcular los momentos centrales
- Formulas

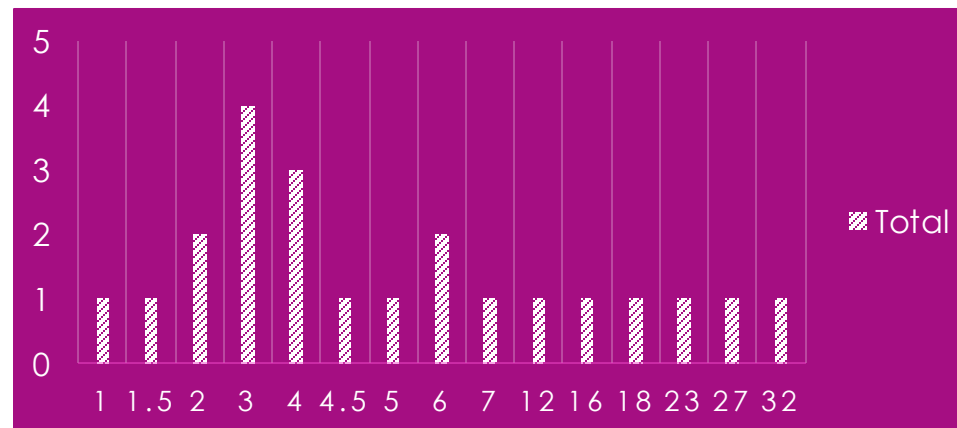


PREPARACIÓN DE LOS DATOS

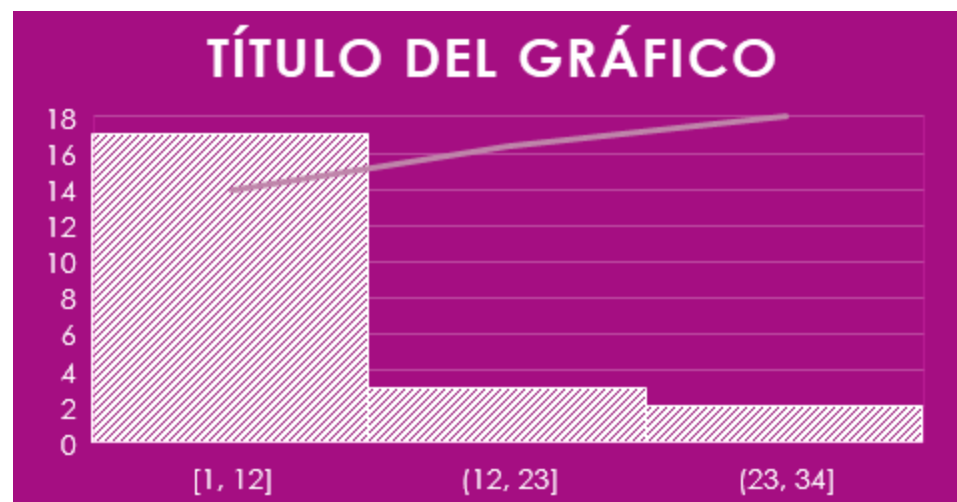
- ▶ Estadísticos básicos: media, mediana, desviación típica, cuantiles...
- ▶ Missings: registros nulos, muy frecuentes en estudios de mercado, BBDD de clientes...
- ▶ Outliers: valores que se salen de lo “normal”
- ▶ Escalado:
 - ▶ Min-Max: $(x - \min) / (\max - \min)$
 - ▶ Z-score: $(x - \mu) / \sigma$

GRÁFICOS MÁS IMPORTANTES

► Frecuencias

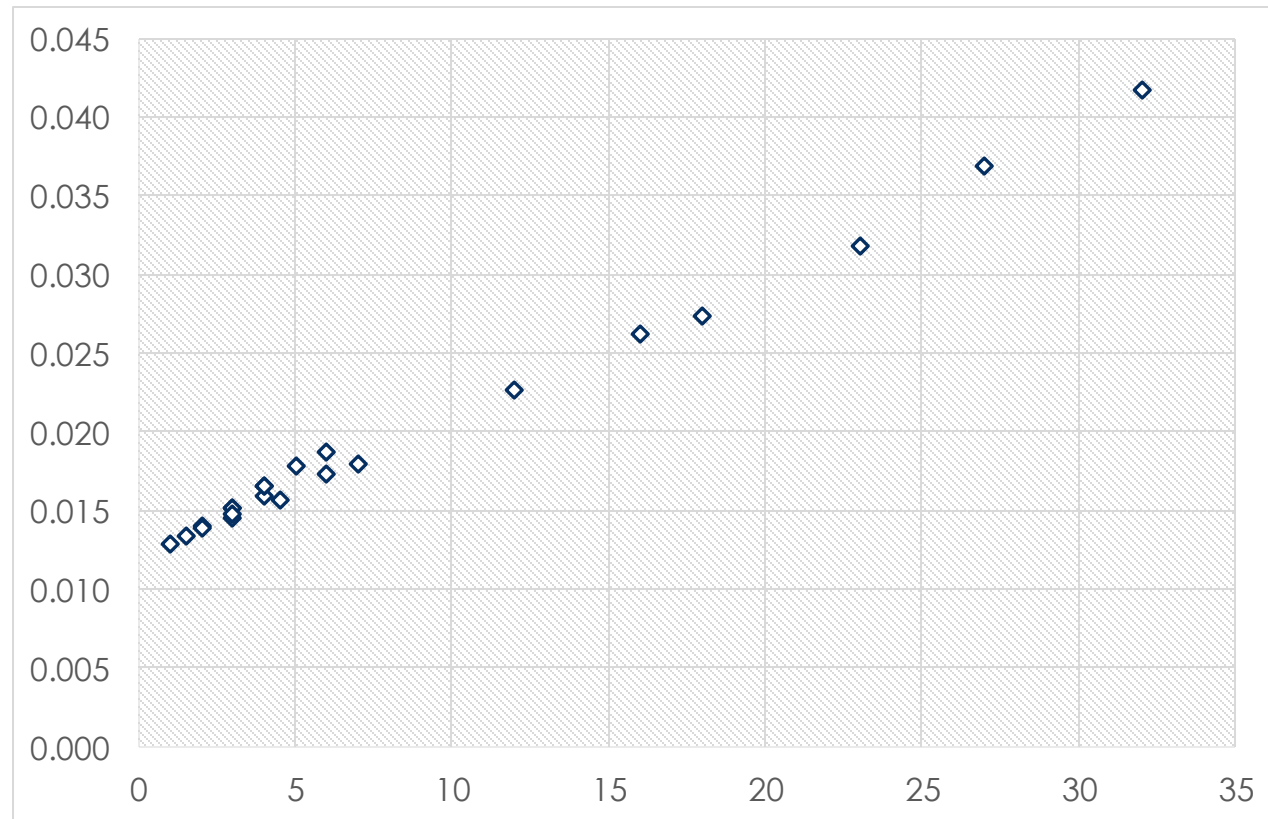


► Histograma



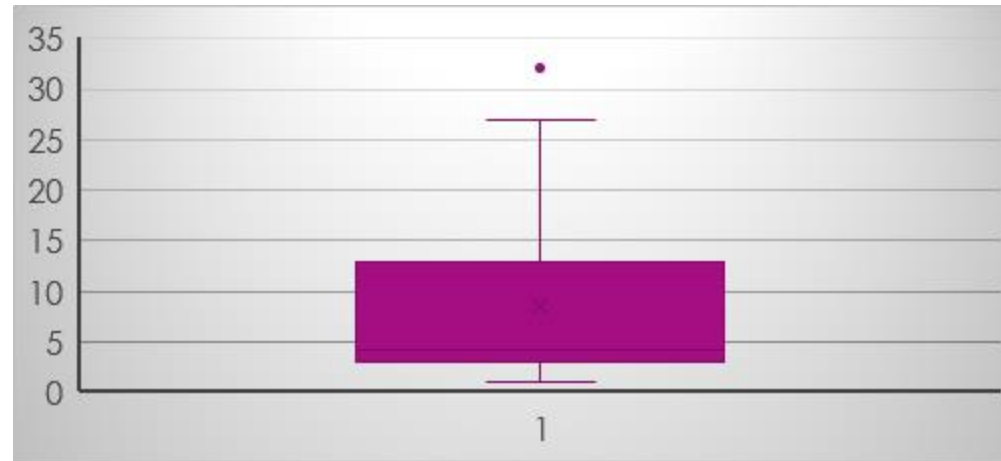
GRÁFICOS MÁS IMPORTANTES

► Dispersión

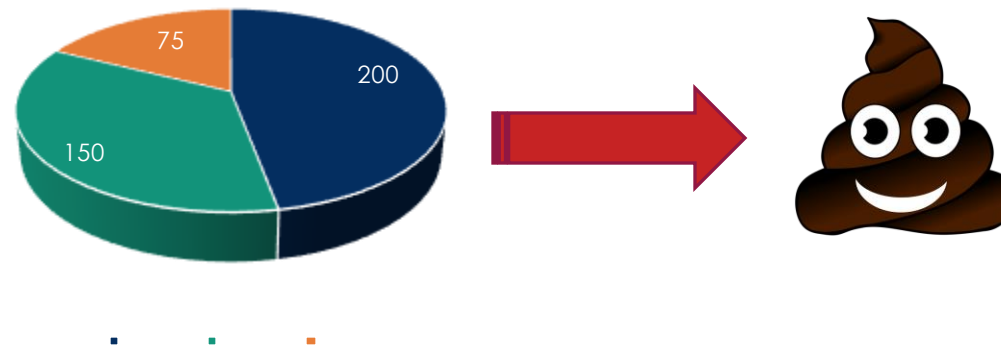


GRÁFICOS MÁS IMPORTANTES

►Boxplot

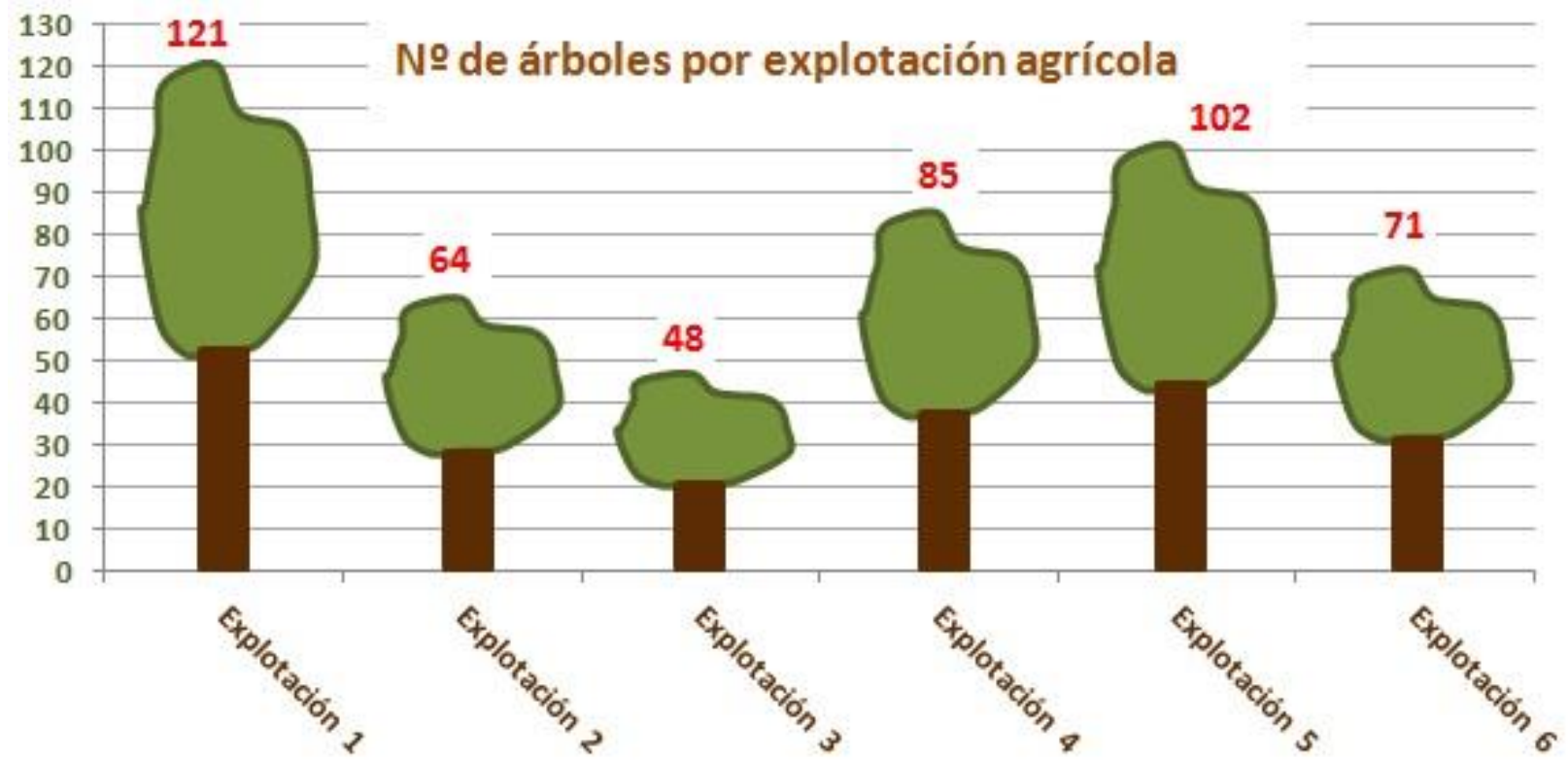


►Sector

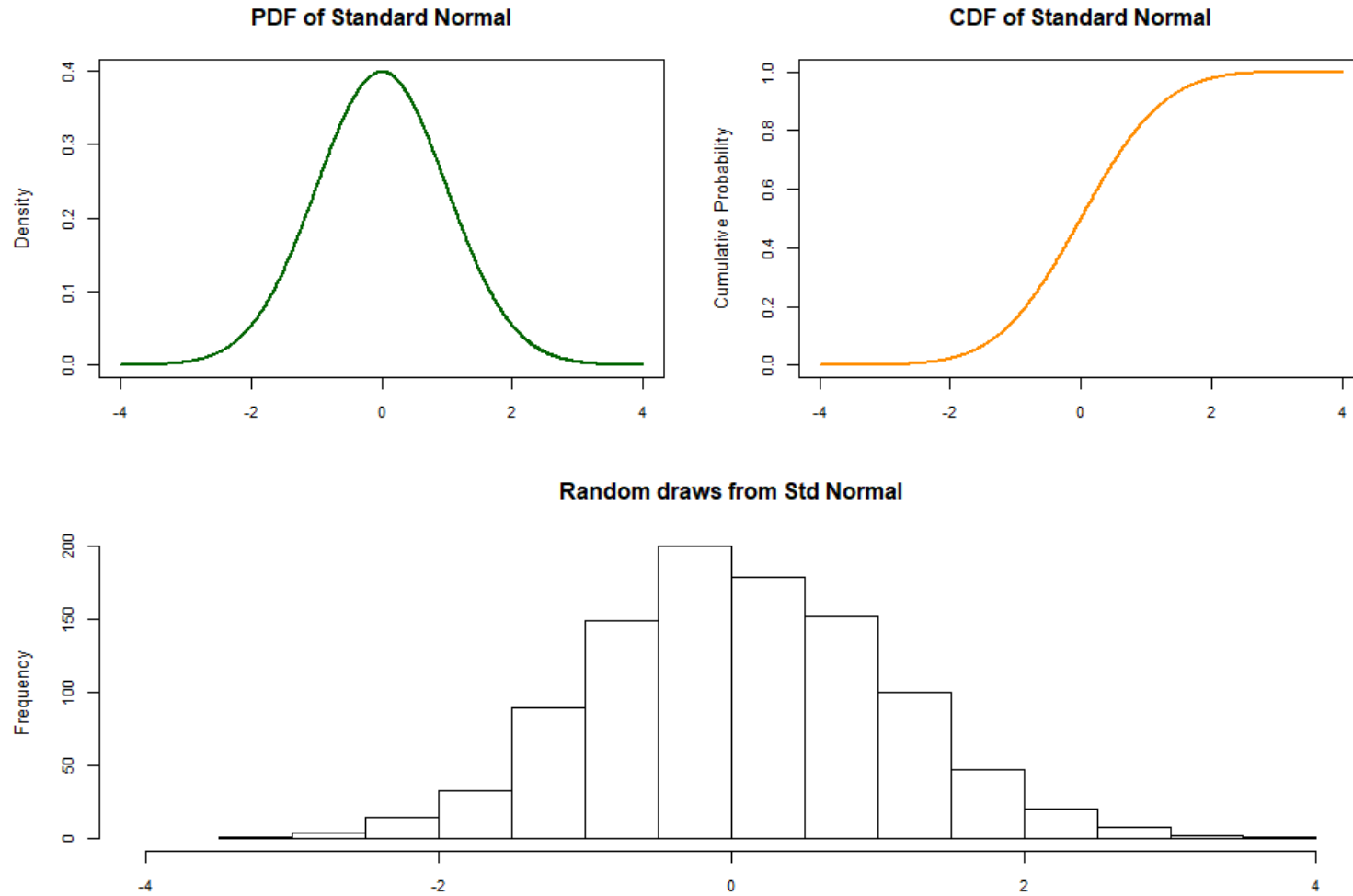


GRÁFICOS MÁS IMPORTANTES

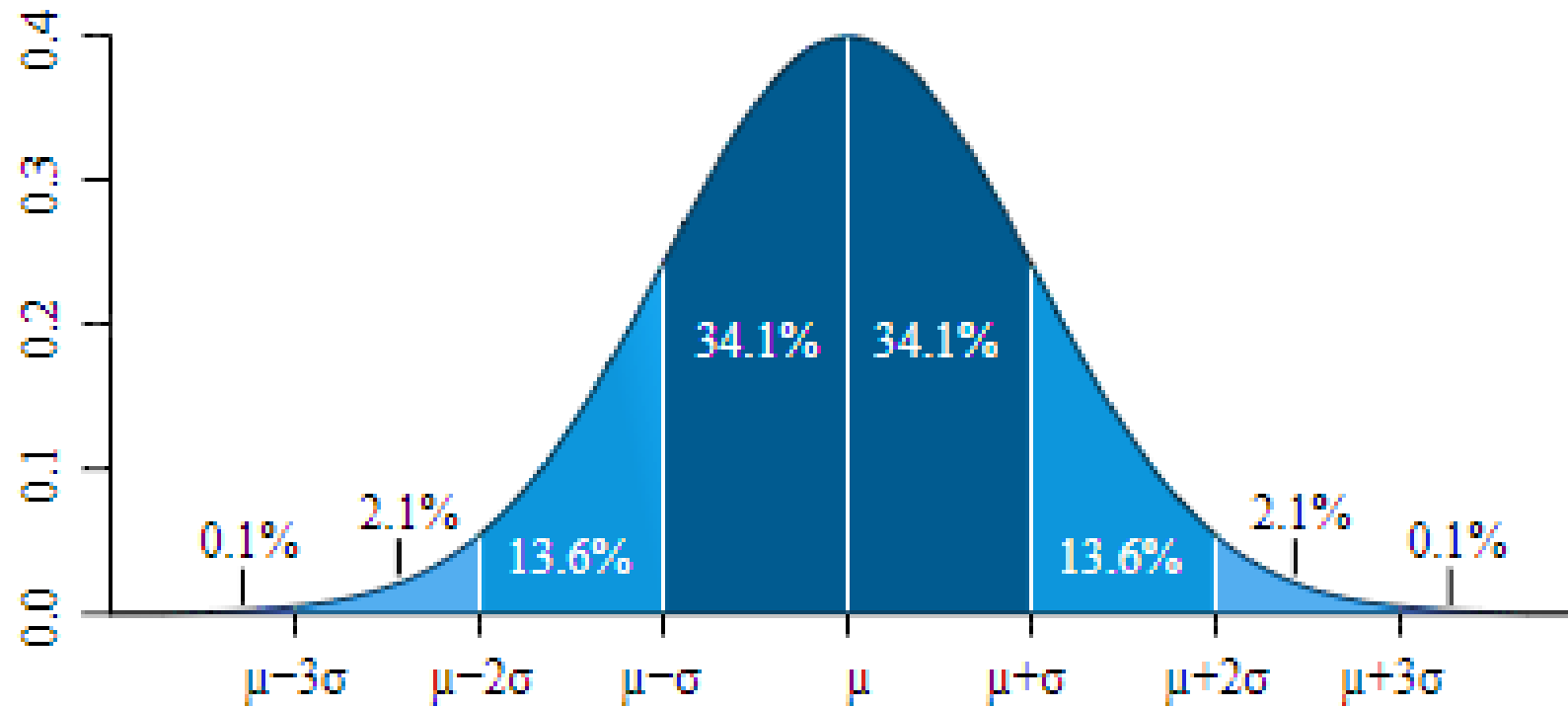
► Pictogramas



NORMALIDAD Y SIGNIFICATIVIDAD



NORMALIDAD Y SIGNIFICANCIA

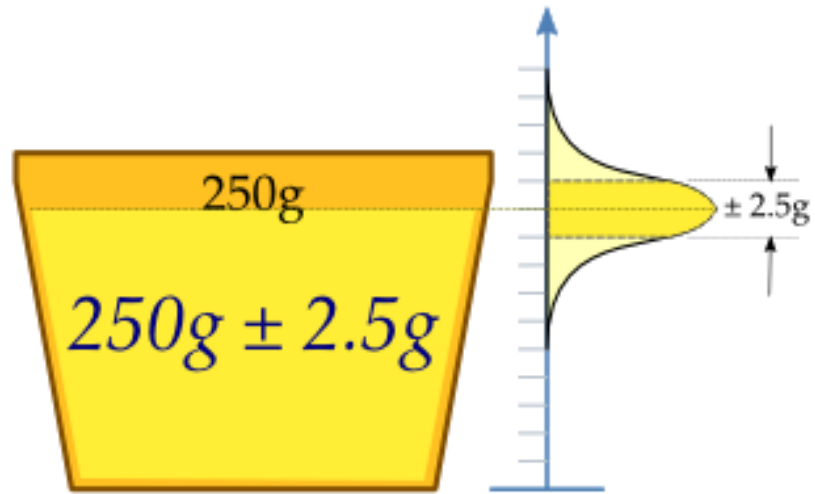


H_0 : Hipótesis nula $\rightarrow (1-\alpha)$: nivel de confianza

H_1 : Hipótesis Alternativa $\rightarrow \alpha$: nivel de significatividad

INTERVALO DE CONFIANZA

$$\hat{X} = \bar{X} \pm k \cdot \sigma$$



IC = 1- α	α	k
0,68	0,32	0,994
0,85	0,15	1,440
0,9	0,1	1,645
0,95	0,05	1,960
0,99	0,01	2,576
0,999	0,001	3,291

¿Cuál sería el IC al 95 % de una media de 5 y una desviación típica de 1,5?

$$\hat{X} = 5 \pm 1,96 \cdot 1,5 = 5 \pm 2,94 = [2,06; 7,94]$$

EL CONCEPTO DE ESTIMADOR

- ▶ Un estimador es una función utilizada para conocer un parámetro de una población determinada



- ▶ Cada parámetro puede tener varios estimadores, escogeremos el que cumpla mejor las siguientes propiedades:
 - ▶ Sesgo: diferencia entre el valor real y el esperado
 - ▶ Eficiencia: diferencia entre varianza de estimadores
 - ▶ Consistencia: a medida que el la muestra crece el estimador tiende al valor real del parámetro
 - ▶ Robustez: si violando los supuestos de partida en los que se basa la estimación, el resultado no cambia significativamente
 - ▶ Suficiencia: el estimador resume de manera relevante toda la información de la muestra
 - ▶ Invarianza: el estimador de la función del parámetro coincide con la función del estimador del parámetro