



**AICADEP**  
ACADEMIA INTERNACIONAL DE CAPACITACIÓN  
Y DESARROLLO PROFESIONAL

# **CURSO ONLINE: DE DOCENTE A AUTOR CIENTÍFICO: PUBLICA ARTÍCULOS ACADÉMICOS DE ALTO IMPACTO**

## **Módulo 3**

**Estadística elemental para la investigación**



**Scopus®**



**WEB OF SCIENCE**

# Contenido

## Unidad III: Estadística elemental para la investigación.

- Introducción a la estadística
- Estadística descriptiva
- Análisis de datos
- Estadística inferencial simple
- Redacción de resultados y gráficos



- **Objetivo**

Analizar y aplicar los fundamentos de la estadística descriptiva e inferencial en el proceso de investigación, mediante la organización, análisis y representación de datos, con el fin de redactar e interpretar resultados de manera clara, coherente y basada en evidencia empírica.

- **Resultado de Aprendizaje**

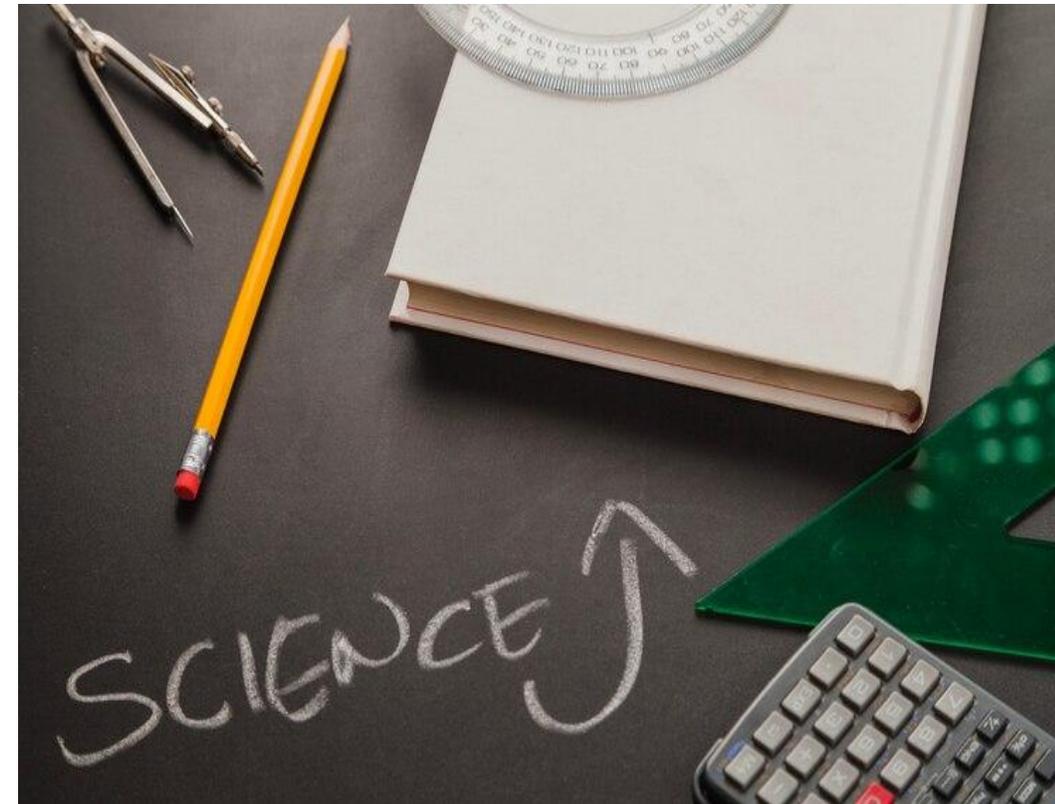
Al finalizar la unidad, el estudiante será capaz de utilizar herramientas estadísticas básicas para organizar y analizar datos cuantitativos, interpretar resultados e inferencias simples, y presentarlos de manera efectiva a través de gráficos y redacción técnica adecuada en contextos de investigación.



## Unidad III: Introducción

La estadística es una herramienta fundamental en la investigación científica, ya que permite organizar, describir, analizar e interpretar datos de manera sistemática y objetiva. Su aplicación no solo facilita la comprensión de fenómenos observables, sino que también posibilita la toma de decisiones basadas en evidencia. Esta unidad tiene como propósito familiarizar al estudiante con los conceptos y métodos básicos de la estadística, abordando desde su carácter descriptivo hasta sus aplicaciones inferenciales más simples.

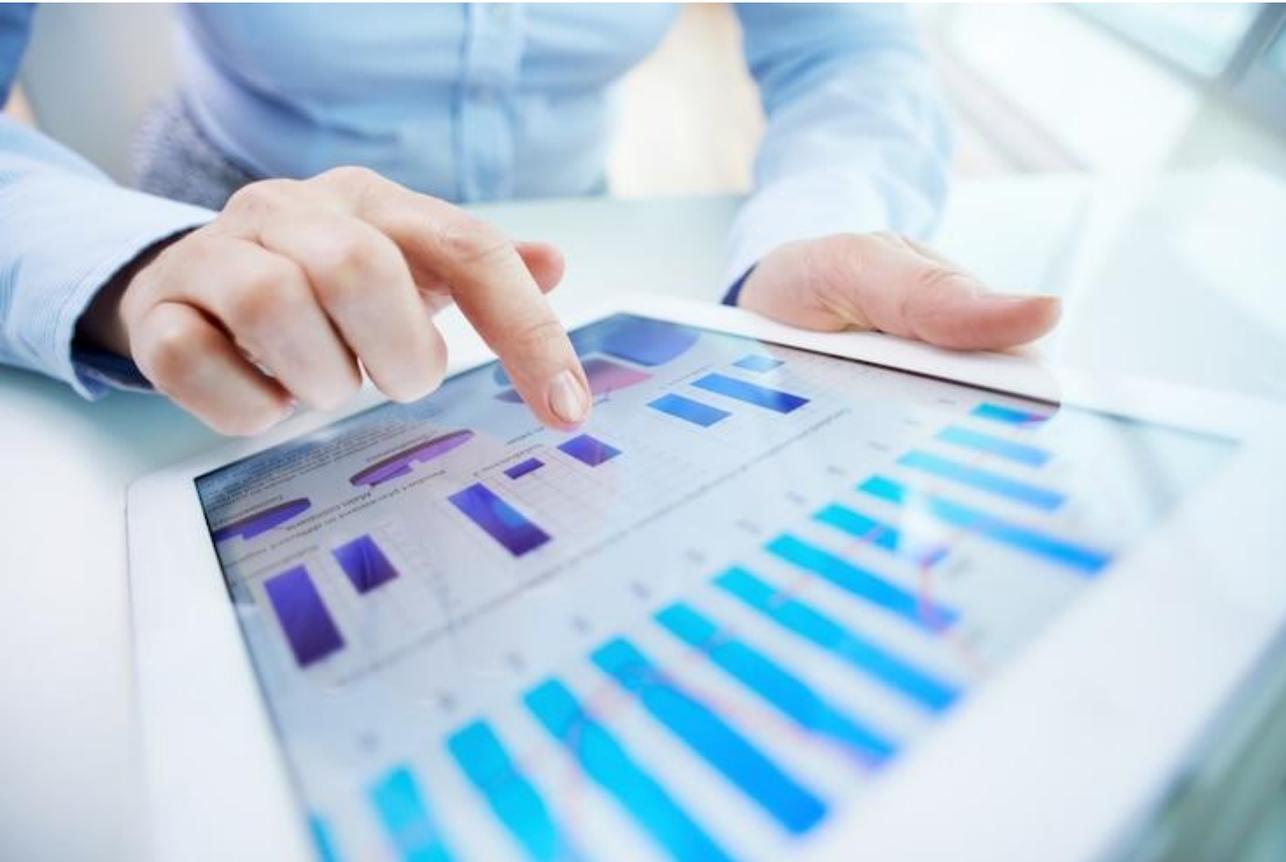
En el ámbito académico, el manejo adecuado de la estadística favorece la validez de los hallazgos, fortalece la redacción científica y promueve una mejor comunicación de los resultados mediante gráficos y representaciones claras (Hernández et al., 2014). La comprensión de estos fundamentos es esencial para el desarrollo de investigaciones rigurosas y confiables.



## 1. Introducción a la estadística

La estadística es la ciencia que se encarga de recolectar, organizar, analizar e interpretar datos con el propósito de tomar decisiones fundamentadas. Existen dos grandes ramas: la estadística descriptiva, que resume y organiza datos, y la estadística inferencial, que permite generalizar conclusiones de una muestra hacia una población (Triola, 2018). “La estadística es más que una herramienta matemática; es un lenguaje que permite comunicar hallazgos de manera estructurada y objetiva” (Martínez, 2020, p. 12).





El terreno de la estadística puede dividirse a grandes rasgos en dos campos de acción:

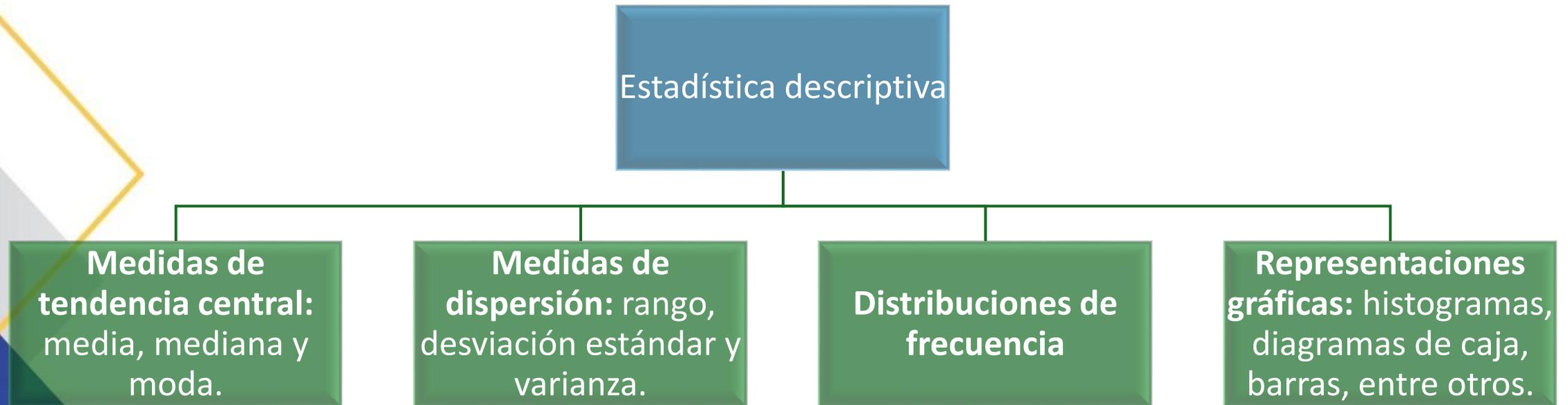
- ✓ Estadística descriptiva y
- ✓ Estadística inferencial.

La *estadística descriptiva* Incluye la obtención, presentación y descripción de los datos muestrales.

El término *estadística inferencial* se refiere a la técnica de interpretación de los valores resultantes de las técnicas descriptivas y la toma de decisiones, así como a la obtención de conclusiones relativas a la población.

## 2. Estadística descriptiva

La estadística descriptiva incluye técnicas que permiten representar datos de manera sencilla. Estas herramientas permiten una primera aproximación al comportamiento de los datos recolectados y sientan las bases para análisis más profundos (Navarro, 2019).

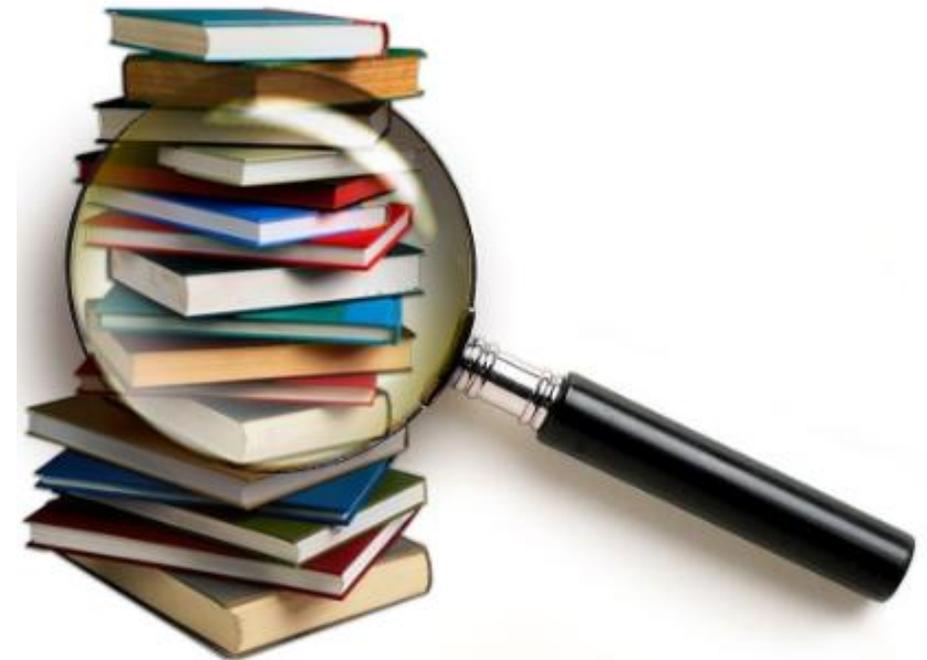


# Estadística descriptiva e Inferencial. Uso de software

## Estadística descriptiva

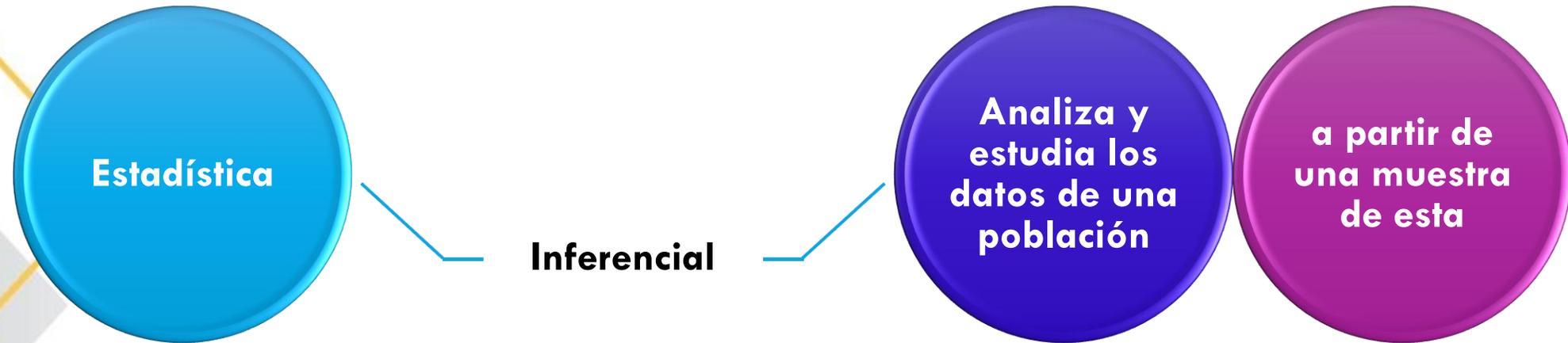
Se refiere al análisis, el resumen y la presentación de los resultados relacionados con un conjunto de datos derivados de una muestra o de toda la población

La estadística descriptiva comprende tres categorías principales: distribución de frecuencias, medidas de tendencia central y medidas de variabilidad.



Rustom (2012)

## Estadística descriptiva e Inferencial. Uso de software



Efectúa **estimaciones e hipótesis** basándose en **probabilidades** y argumentando sus resultados a partir de las **muestras de una población**.

Céspedes, (2015)

## Estadística descriptiva e Inferencial. Uso de software



Céspedes, (2015)

# Estadística inferencial Estimaciones con intervalos de confianza y pruebas de hipótesis

## Estimación de parámetros:

Hace referencia al estudio de una muestra de una población a partir de los diferentes parámetros que permite caracterizar los valores hallados en las poblaciones y las muestras



Cespedes, (2015)

## Esquema de las etapas de un estudio estadístico



Tipos de datos y escalas de medición.

## Variables, dimensiones, indicadores



Mendoza, & Garza, (2009)

Variable: corresponde a la característica de la Unidad de Análisis

## TIPOS DE VARIABLES

### Variables Cuantitativas

#### CONTINUO

Intervalo

Toma cualquier valor dentro de un intervalo  
Ejemplos: *Peso; Estatura; Temperatura, etc.*

#### DISCRETA

Toma valores enteros  
Ejemplos: *Número de Hijos, Número de empleados de una empresa, Número de asignaturas aprobadas en un semestre, etc.*

### Variables Cualitativas

#### NOMINAL

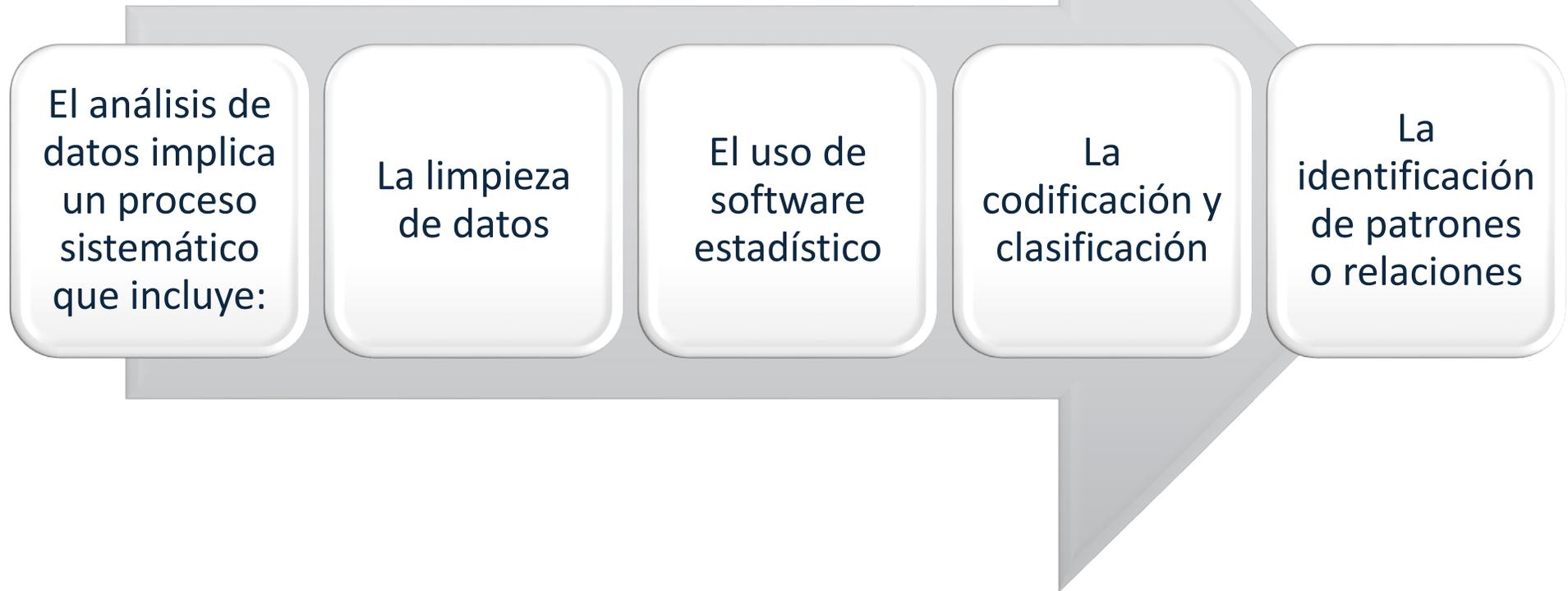
Característica o cualidad cuyas categorías no tienen un orden preestablecido.  
Ejemplos: *Sexo, Deporte Favorito, etc.*

#### ORDINAL

Característica o cualidad cuyas categorías tienen un orden preestablecido.  
Ejemplos: *Calificación (S, N, A); Grado de Interés por un tema, etc.*

Unidad de Medida: Gramos o Kilos para la variable Peso; Grados C o F para Temperatura

### 3. Análisis de datos



Este proceso es crucial en la investigación cuantitativa, pues transforma los datos en información útil para responder preguntas o contrastar hipótesis (Field, 2018).

## Organización y presentación de datos (tablas, gráficos).

**DATOS EN BRUTO** Los *datos en bruto* son los datos **recolectados** que aún **no** se han **organizado**. Por ejemplo, las estaturas de 100 estudiantes tomados de la lista alfabética de una universidad.



**ORDENACIONES** *Ordenación* se le llama a los datos numéricos en bruto dispuestos en orden creciente o decreciente de magnitud.



**Rango.** A la diferencia entre el número mayor y el número menor se le conoce como el *rango* de los datos. Por ejemplo, si la estatura mayor en los 100 estudiantes es 74 pulgadas (in) y la menor es 60 in, el rango es  $74 - 60 = 14$  pulgadas (in).



**DISTRIBUCIONES DE FRECUENCIA** Al organizar una gran cantidad de datos en bruto, suele resultar útil distribuirlos en *clases* o *categorías* y determinar la **cantidad de datos que pertenece a cada clase**; esta cantidad se conoce como la **frecuencia de clase**. A la disposición tabular de los datos en clases con sus respectivas frecuencias de clase se le conoce como **distribución de frecuencias** o **tabla de frecuencias**.

**INTERVALOS DE CLASE Y LÍMITES DE CLASE** Al símbolo que representa una clase, como 60-62 en la tabla, se le conoce como *intervalo de clase*.

**LIMITES DE CLASES.** A los números de los extremos, 60 y 62, se les conoce como *límites de clase*; el número menor (60) es el *límite inferior de clase*, y el número mayor (62) es el *límite superior de clase*. Los términos *clase* e *intervalo de clase* se suelen usar indistintamente, aunque el intervalo de clase en realidad es un símbolo para la clase.

**TAMAÑO O AMPLITUD DE UN INTERVALO DE CLASE** El tamaño, o la amplitud, de un intervalo de clase es la diferencia entre sus fronteras superior e inferior y se le conoce también como *amplitud de clase*, *tamaño de clase* o *longitud de clase*. Si en una distribución de frecuencia todos los intervalos de clase tienen la misma amplitud, esta amplitud común se denota  $c$ . En este caso,  $c$  es igual a la diferencia entre dos límites inferiores de clases sucesivas o entre dos límites superiores de clases sucesivas. Por ejemplo, en los datos de la tabla, el intervalo de clase es  $c = 62.5 - 59.5 = 65.5 - 62.5 = 3$

**LA MARCA DE CLASE.** La *marca de clase* es el punto medio del intervalo de clase y se obtiene sumando los límites de clase inferior y superior y dividiendo entre 2. Así, la marca de clase del intervalo 60-62 es  $(60 + 62)/2 = 61$ . A la marca de clase también se le conoce como *punto medio de clase*.

Estatura (in)	Cantidad de estudiantes
60-62	5
63-65	18
66-68	42
69-71	27
72-74	8
Total	100

## REGLAS GENERALES PARA FORMAR UNA DISTRIBUCIÓN DE FRECUENCIAS

- 1. En el conjunto de los datos en bruto, se determina el número mayor y el número menor y se halla, así, el **rango** (la diferencia entre los números mayor y menor).
- 2. Se divide el rango en una cantidad adecuada de **intervalos de clase** de una misma amplitud. Si esto no es posible, se usan intervalos de clase de diferentes amplitudes o intervalos de clase abiertos. La cantidad de intervalos suele ser de 5 a 20, dependiendo de los datos. Los intervalos de clase también suelen elegirse de manera que las marcas de clase (o puntos medios de clase) coincidan con datos observados. Esto tiende a disminuir el llamado *error de agrupamiento* en los análisis matemáticos subsiguientes. En cambio, las fronteras de clase no deben coincidir con datos observados.
- 3. Se determina la cantidad de observaciones que caen dentro de cada intervalo de clase; es decir, se encuentran las **frecuencias de clase**. La mejor manera de hacer esto es utilizando una *hoja de conteo*.

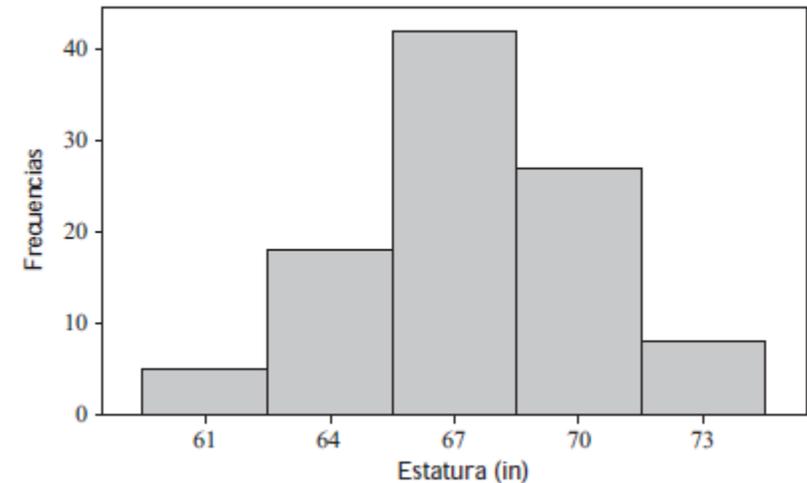
## HISTOGRAMAS Y POLÍGONOS DE FRECUENCIAS

### HISTOGRAMAS Y POLÍGONOS DE FRECUENCIAS

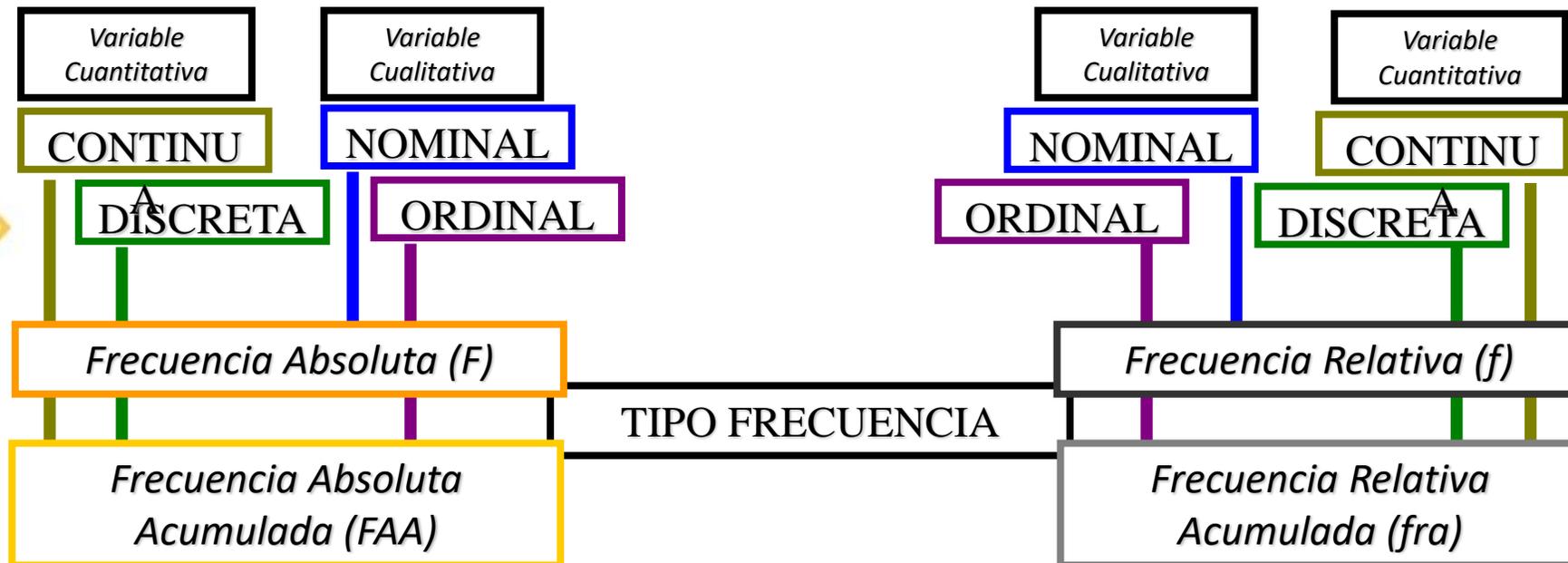
Los histogramas y los polígonos de frecuencias son dos representaciones gráficas de las distribuciones de frecuencias.

1. Un *histograma* o *histograma de frecuencias* consiste en un conjunto de rectángulos que tienen: *a*) sus bases sobre un eje horizontal (el eje  $X$ ), con sus centros coincidiendo con las marcas de clase de longitudes iguales a la amplitud del intervalo de clase, y *b*) áreas proporcionales a las frecuencias de clase.

2. Un *polígono de frecuencias* es una gráfica de línea que presenta las frecuencias de clase graficadas contra las marcas de clase. Se puede obtener conectando los puntos medios de las partes superiores de los rectángulos de un histograma.



**Frecuencia:** desde un conjunto de unidades, corresponde al Número o Porcentaje de veces que se presenta una característica.



## DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

La *frecuencia relativa* de una clase es la frecuencia de la clase dividida entre la suma de las frecuencias de todas las clases y generalmente se expresa como porcentaje

Si en una tabla las frecuencias se sustituyen por frecuencias relativas, la tabla que se obtiene es una ***distribución de frecuencias relativas, distribución porcentual o tabla de frecuencias relativas.***

Las representaciones gráficas de las distribuciones de frecuencias relativas se obtienen a partir de los histogramas o polígonos de frecuencias, cambiando únicamente, en la escala vertical, las frecuencias por las frecuencias relativas y conservando la gráfica exactamente igual.

## DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS Y OJIVAS

### DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS Y OJIVAS

Estatura (in)	Cantidad de estudiantes
Menos de 59.5	0
Menos de 62.5	5
Menos de 65.5	23
Menos de 68.5	65
Menos de 71.5	92
Menos de 74.5	100

A la suma de todas las frecuencias menores que la frontera superior de un intervalo de clase dado se le llama **frecuencia acumulada** hasta ese intervalo de clase inclusive. Por ejemplo, en la tabla 2.1, la frecuencia acumulada hasta el intervalo de clase 66-68 inclusive es  $5 + 18 + 42 = 65$ , lo que significa que 65 estudiantes tienen una estatura menor a 68.5 in.

- A una tabla en la que se presentan las frecuencias acumuladas se le llama **distribución de frecuencias acumuladas**, *tabla de frecuencias acumuladas* o simplemente *distribución acumulada*, y se presenta en la tabla.

## DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS RELATIVAS Y OJIVAS PORCENTUALES

•La **frecuencia acumulada relativa** o **frecuencia acumulada porcentual** es la frecuencia acumulada dividida entre la suma de todas las frecuencias (frecuencia total). Por ejemplo, la frecuencia acumulada relativa de las estaturas menores que 68.5 in es  $65/100 = 0.65$  o 65%, lo que significa que 65% de los estudiantes tienen estaturas menores a 68.5 in. Si en una tabla se emplean las frecuencias acumuladas relativas en lugar de las frecuencias acumuladas, se obtiene una **distribución de frecuencias acumuladas relativas** (o **distribución acumulada porcentual**) y una **gráfica de frecuencias acumuladas relativas** (u **ojiva porcentual**), respectivamente.

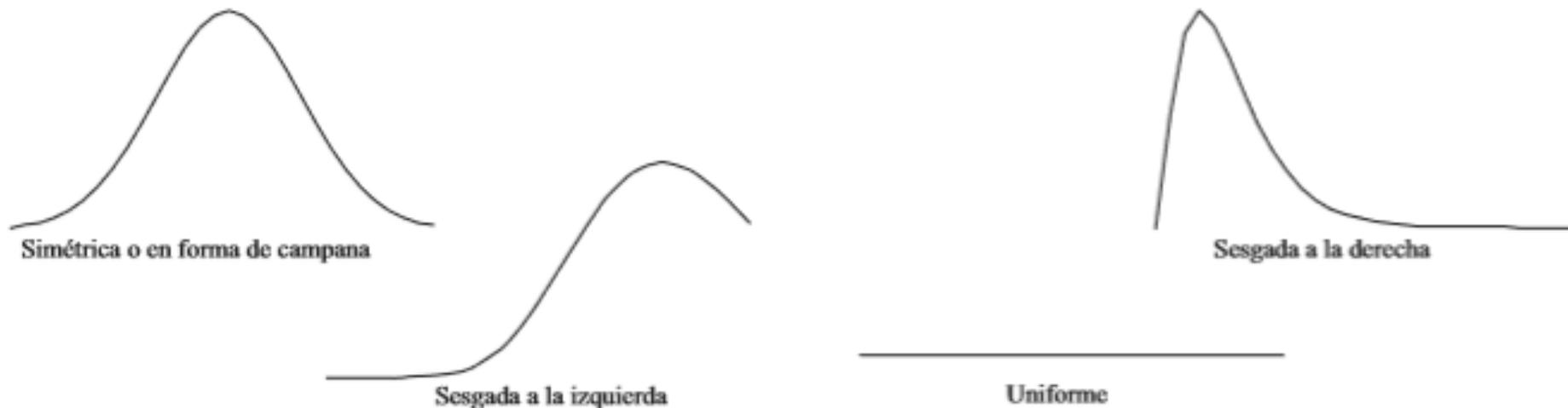


## CURVAS DE FRECUENCIAS Y OJIVAS SUAVIZADAS

Cuando se tienen poblaciones grandes puede esperarse que los polígonos de frecuencias, o los polígonos de frecuencias relativas, correspondientes a estas poblaciones estén formados por una gran cantidad de pequeños segmentos de recta de manera que sus formas se aproximen a las de unas curvas, a las cuales se les llama curvas de frecuencias o curvas de frecuencias relativas, respectivamente.

Es razonable esperar que estas curvas teóricas puedan ser aproximadas suavizando los polígonos de frecuencias o los polígonos de frecuencias relativas de la muestra; esta aproximación mejorará a medida que aumenta el tamaño de la muestra. Ésta es la razón por la que a las curvas de frecuencias se les suele llamar polígonos de frecuencias suavizados.

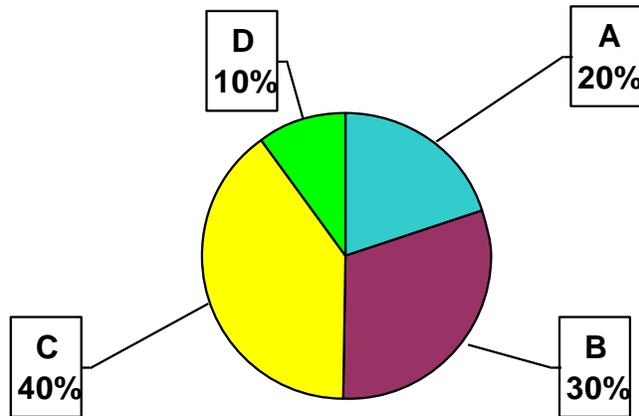
Tipos de curvas de frecuencias



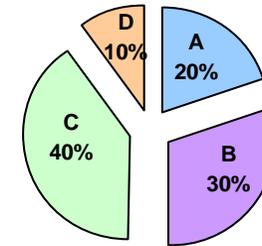
## TIPOS DE GRÁFICOS

### Gráfico de Sectores Circulares (de *Torta*)

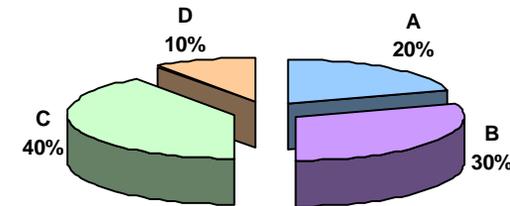
Distribución de las unidades de análisis de acuerdo a variable 1



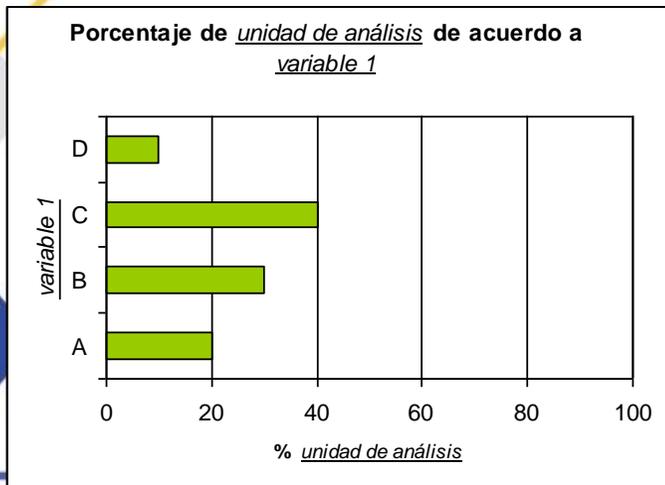
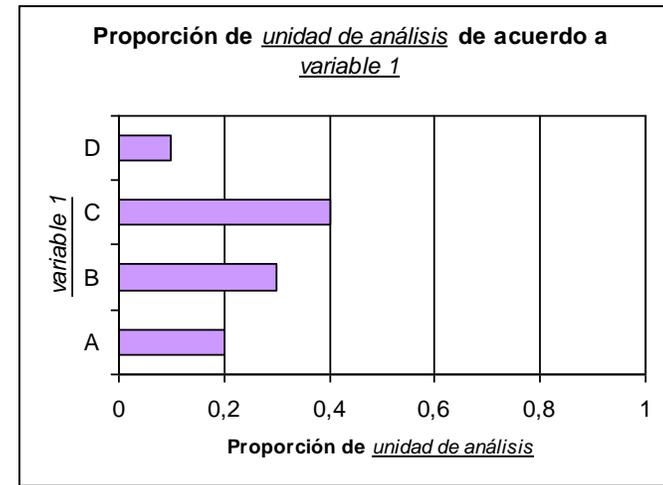
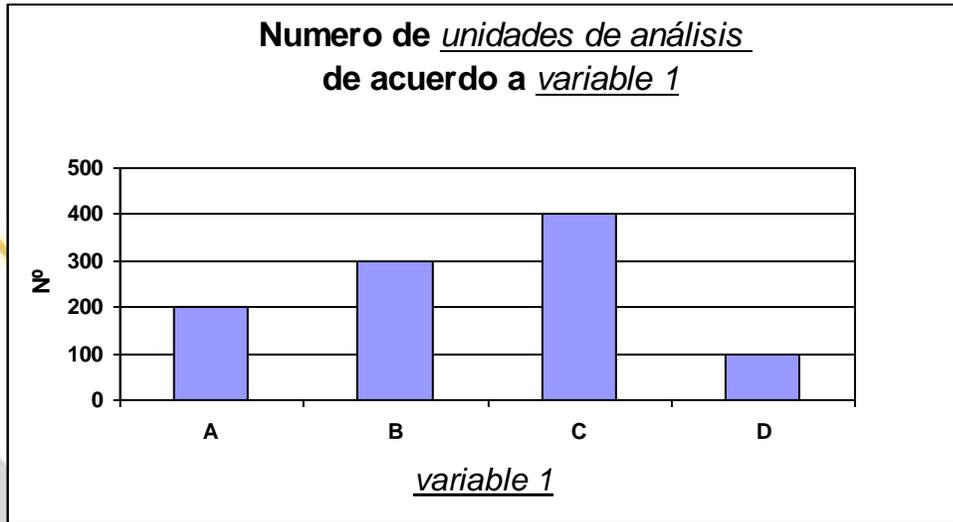
Distribución de las unidades de análisis de acuerdo a variable 1



Distribución de las unidades de análisis de acuerdo a variable 1

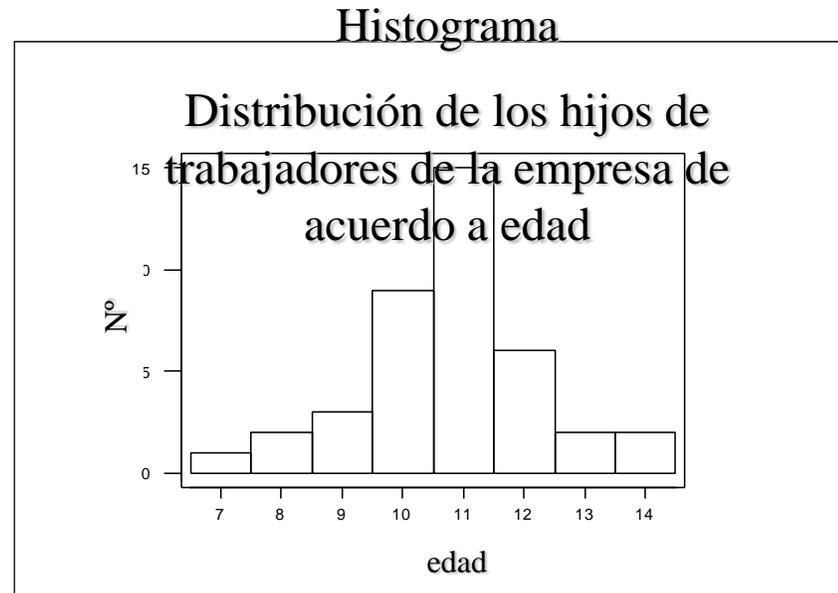


## Gráfico de Barras



- Este tipo de gráfico se utiliza generalmente para *representar la frecuencia* de las categorías de una variable cualitativa.
- Cuando una variable es cuantitativa se puede utilizar este tipo de gráfico sólo si la variable se ha transformada en categorías.
- Hay distintas versiones de estos gráficos (por ejemplo en Excel), y en algunos casos son muy útiles para describir el comportamiento de una variable en distintos grupos.

## Histograma



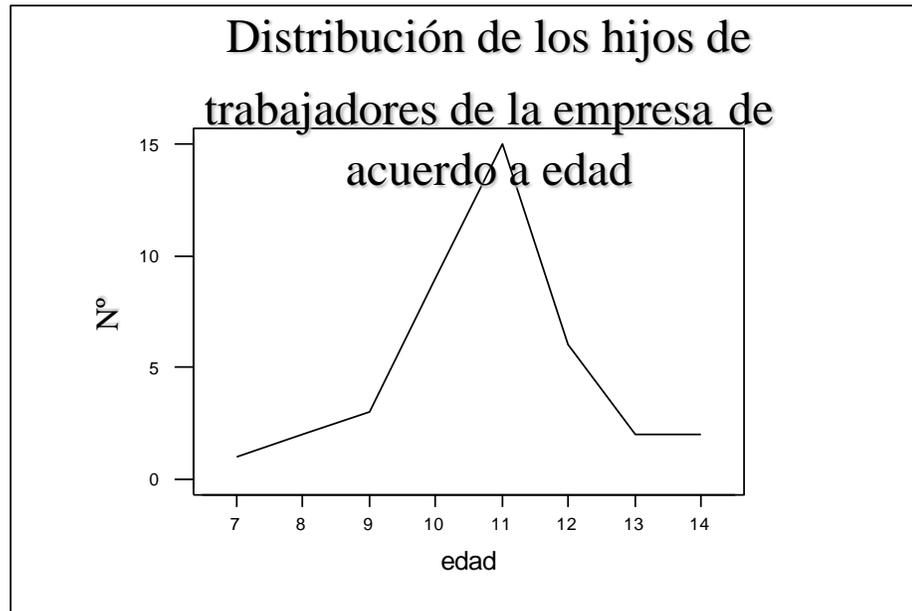
### *Ejemplo*

En el gráfico se puede observar el número de hijos , de menor edad (7-8 años), las de mayor edad (13-14 años); y además que la mayoría de... hijos de los trabajadores están entre los 10 y 12 años.

### Histograma

- Permite la representación de la frecuencia de una variable Cuantitativa.
- El eje x se refiere a la variable.
- El eje y se refiere a la frecuencia (Nº , %).
- Cada barra representa la frecuencia de la variable en la población en estudio (o la muestra).
- El histograma se puede construir desde los datos de la tabla de frecuencia de la variable en estudio.

## Polígono de Frecuencia



-Esta representación se basa en el Histograma.

-Sólo es útil para variables cuantitativas.

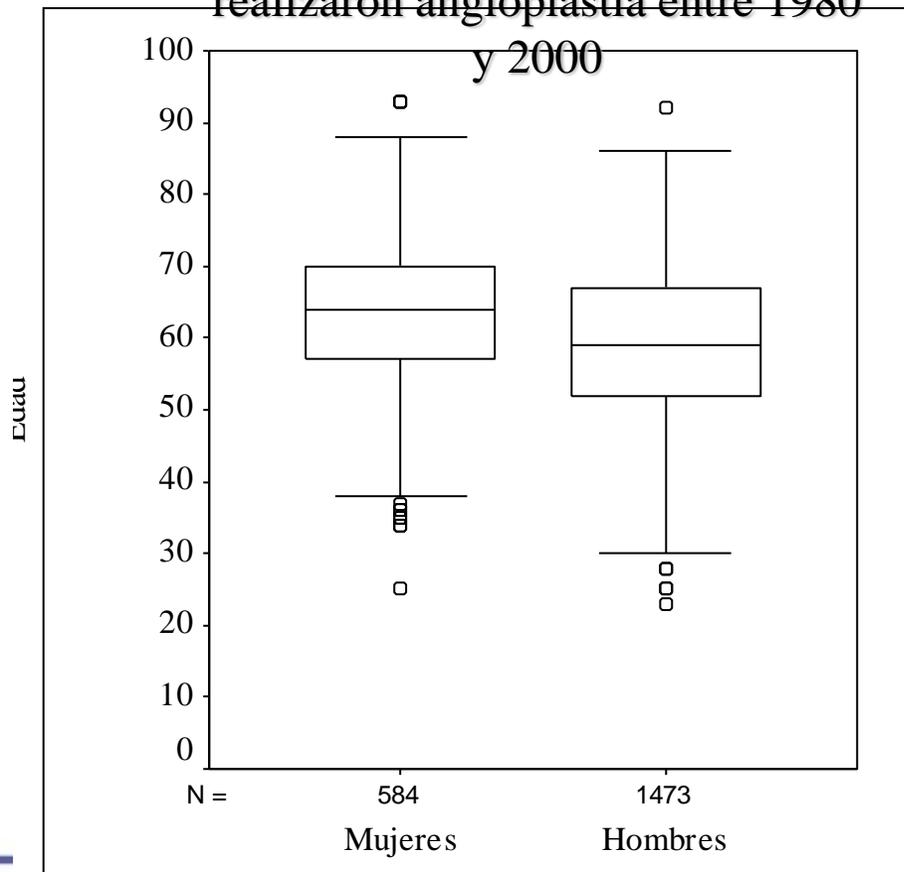
-El eje x se refiere a la variable.

- El eje y se refiere a la frecuencia (Nº, %).

-Los puntos que permiten la unión de las líneas representa el *centro de clase* (o *marca de clase*).

## Diagrama de Caja

Edad de las personas que se  
realizaron angioplastía entre 1980  
y 2000



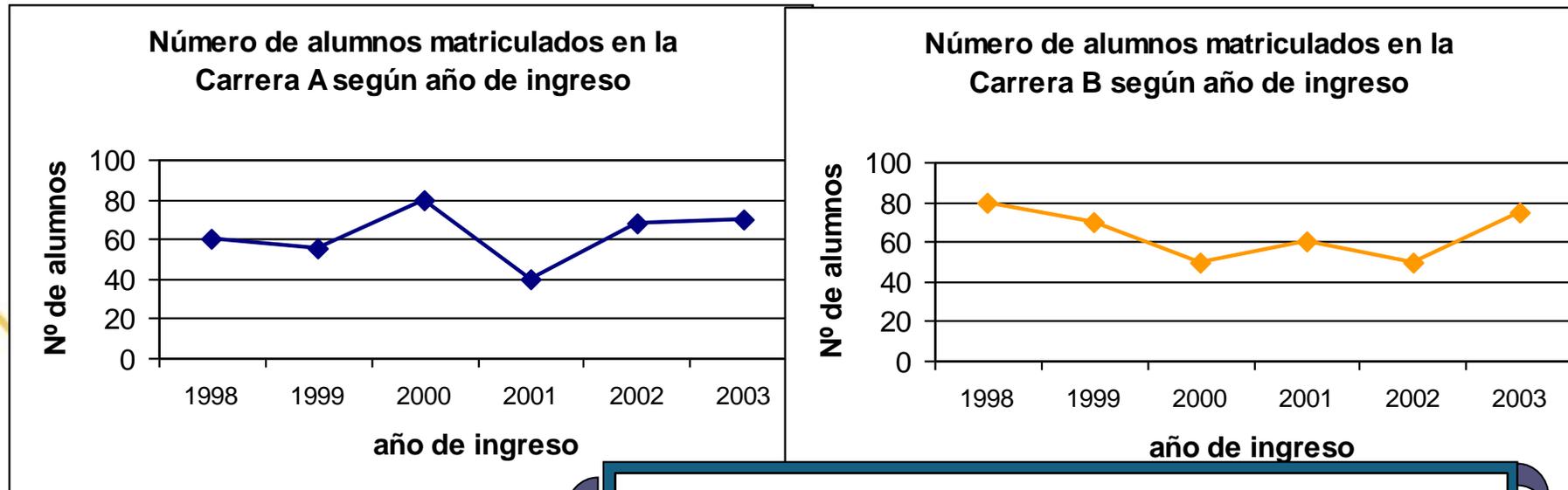
- Permite identificar gráficamente la mediana, los cuartiles 1 y 3 (percentiles 25 y 75), mínimo y máximo de una variable.

- Sólo es útil para variables cuantitativas.

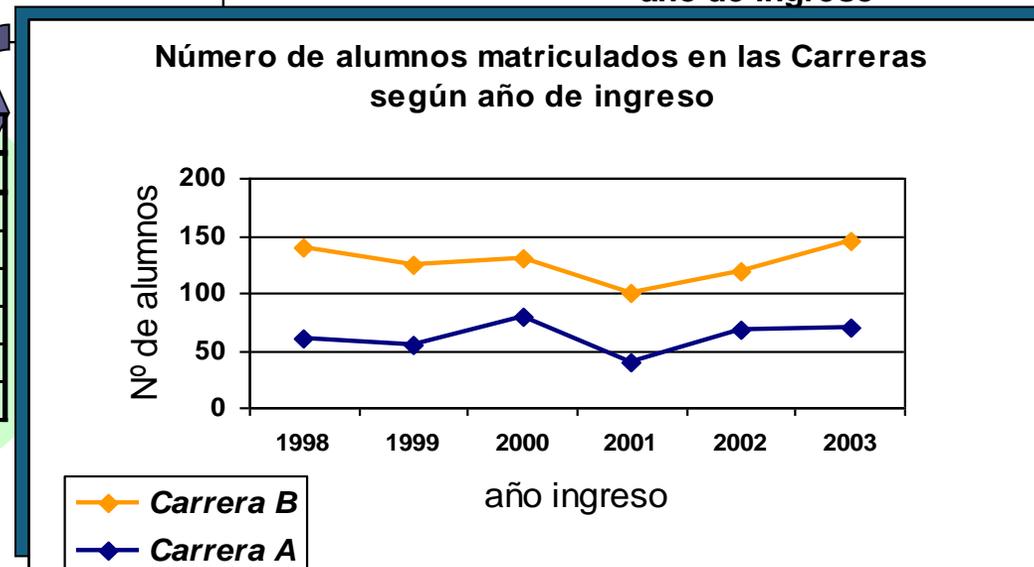
- El eje x permite identificar la población en estudio.

- El eje y representa los valores de la variable en estudio.

### Otros



año de ingreso	Nº de alumnos	
	Carrera A	Carrera B
1998	60	80
1999	55	70
2000	80	50
2001	40	60
2002	68	50
2003	70	75



## 4. Estadística inferencial simple

La estadística inferencial permite estimar o predecir características de una población a partir de una muestra. Incluye conceptos como:

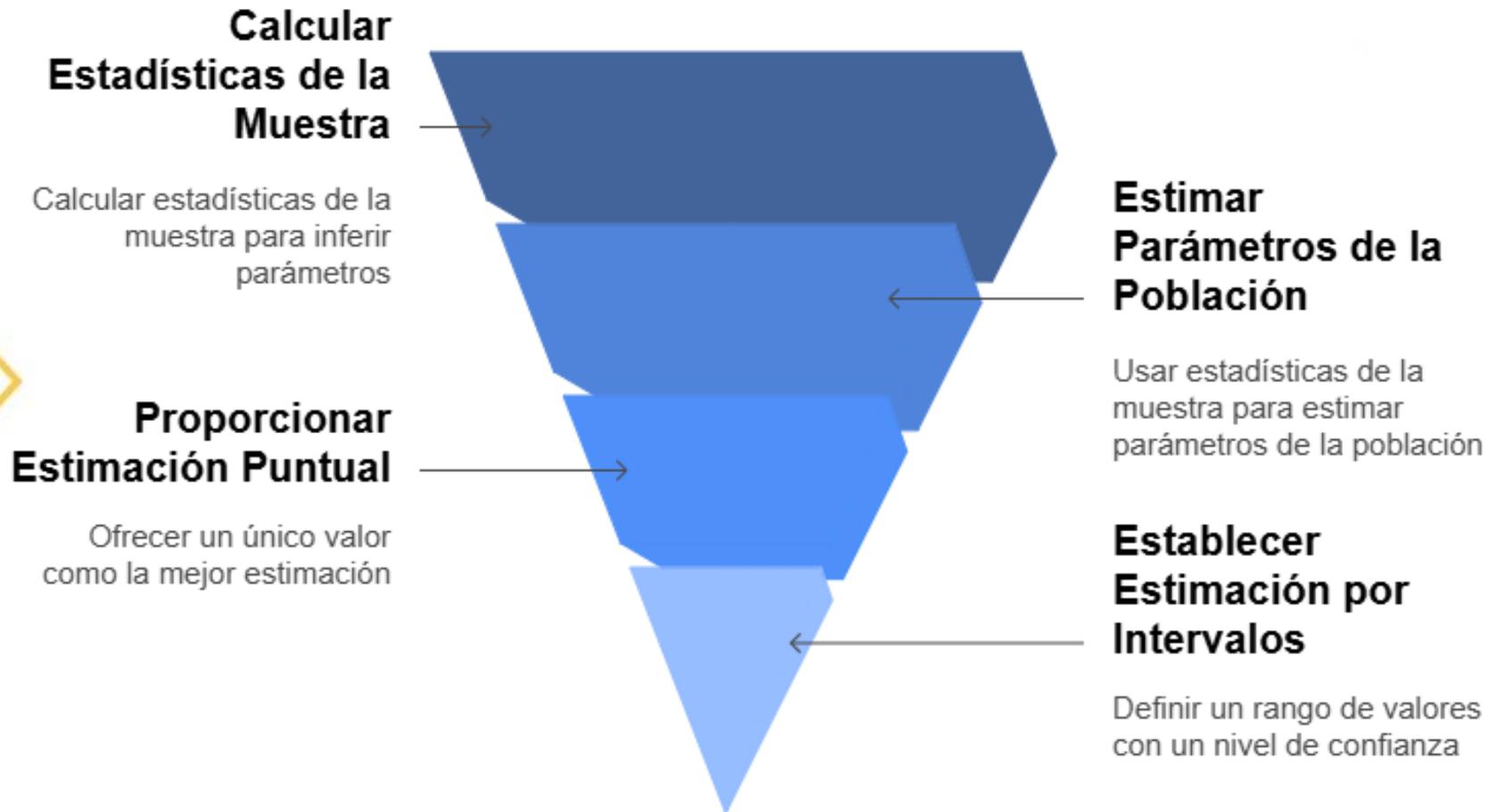
- **Estimación de parámetros**
- **Pruebas de hipótesis**
- **Intervalos de confianza**
- **Errores tipo I y tipo II**

Se trata de un enfoque probabilístico que introduce un grado de incertidumbre controlado en la interpretación de los resultados (Pagano & Gauvreau, 2011).



# Conceptos de estimación puntual y por intervalos

## Proceso de Estimación Estadística



# Tipos de Estimación

$$\bar{x} \pm Z_c \frac{\sigma}{\sqrt{N}} \text{ Población infinita}$$

$$\bar{x} \pm Z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}} \text{ Población finita}$$

## Estimación por Intervalos

Ofrece un rango de valores dentro del cual se espera que se encuentre el parámetro con una determinada probabilidad.

Nivel de confianza, usualmente 95%

## Estimación Puntual

Proporciona un único valor numérico como estimación. Por ejemplo, la media muestral se usa como estimación.

EP: se utiliza la información de la muestra  $(x_1, x_2, \dots, x_n)$ , a través de un **estimador**.

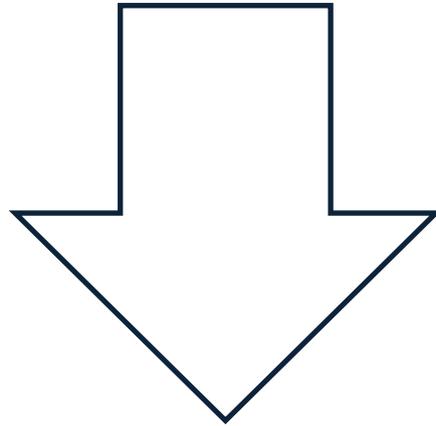
$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

$$\hat{p} = \frac{x_1 + \dots + x_n}{n}$$

$$S^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

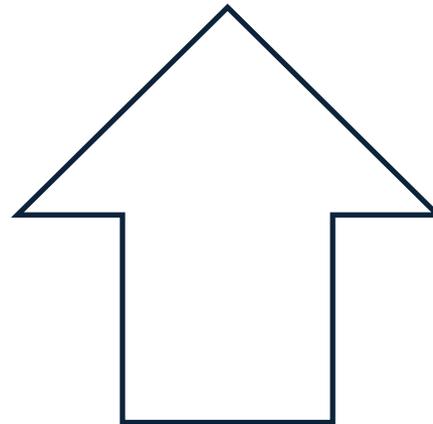
- Media muestral ( $\bar{x}$ ) para estimar  $\mu$ .
- Proporción muestral ( $\hat{p}$ ) para estimar  $p$ .
- Varianza muestral  $S^2(\sigma^2)$
- $\sigma$ : Desviación estándar poblacional.
- $Z_c$  coeficiente nivel de confianza
- $n$  representa el tamaño de la muestra

## Pruebas de hipótesis paramétricas



### Pruebas de hipótesis paramétricas

Las pruebas de hipótesis paramétricas son procedimientos estadísticos que permiten tomar decisiones sobre parámetros poblacionales (como medias o proporciones) cuando se conocen o se asumen ciertas condiciones sobre la distribución de los datos (por ejemplo, normalidad).



### ¿Deberíamos usar pruebas de hipótesis paramétricas?

#### Usar Pruebas Paramétricas

Adecuado cuando se cumplen los supuestos sobre la distribución de datos, como la normalidad.

#### Usar Pruebas No Paramétricas

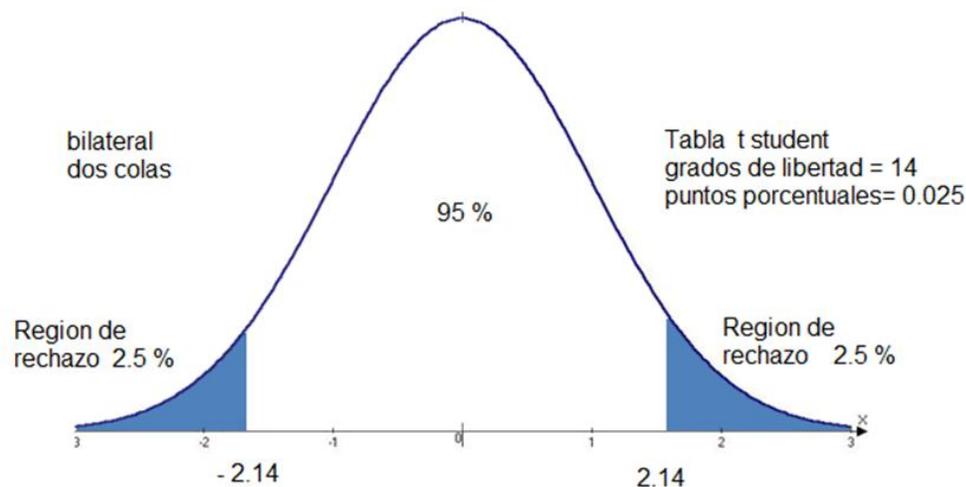
Necesario cuando los supuestos sobre la distribución de datos no se cumplen.



### Estadística inferencial. Estimaciones con intervalos de confianza y pruebas de hipótesis

#### Pasos para la prueba de hipótesis

- Plantear las hipótesis ( $H_0$  y  $H_1$ )
- Establecer un nivel de significancia: valor de alfa, nivel de probabilidad de equivocarse, lo fija el investigador



# Estadística inferencial. Estimaciones con intervalos de confianza y pruebas de hipótesis

## Pasos para la prueba de hipótesis

### 3. Seleccionar el estadístico de prueba a aplicar

Puede ser, la Prueba Z de la curva normal, la prueba T de student, la prueba de correlación o la de regresión



### 4. Establecer una regla de decisión

Esta señala que se debe ser rechazada  $H_0$  si el valor de la estadística de prueba que se calcula a partir de la muestra es uno de los valores de la región de rechazo, y que no se debe rechazar si el valor calculado es uno de los valores de la región de no rechazo.

## Estadística inferencial. Estimaciones con intervalos de confianza y pruebas de hipótesis

### Pasos para la prueba de hipótesis

#### 5. Tomar una decisión o conclusión

Decidir si  $H_0$  se rechaza, entonces se concluye que  $H_1$  es verdadera

Si  $H_0$  no se rechaza, se concluye que  $H_0$  puede ser verdadera.

Es importante aclarar que cuando la  $H_0$  no es rechazada, tampoco se puede decir que se acepta. Se debe decir que «*no se rechaza*»



Cespedes, (2015)



## Análisis Multivariado de Datos

El **Análisis Multivariado de Datos (AMD)** es un conjunto de **técnicas estadísticas** que permiten examinar simultáneamente múltiples variables con el objetivo de **identificar patrones, relaciones o estructuras subyacentes en los datos**. A diferencia del análisis univariado o bivariado, el análisis multivariado considera el efecto conjunto de varias variables, lo cual es fundamental cuando se trabaja con datos complejos (Hair et al, 2019).

## Prueba *t* de Student

$$T = \frac{\bar{x} - \mu_0}{\hat{S}_{n-1}/\sqrt{n}} \in t_{n-1}.$$

Diseñada para **comparar medias** bajo el supuesto de normalidad.

- Es una prueba para evaluar hipótesis en torno a una media
- Se utiliza cuando se tiene una muestra menor de 30 mediciones
- Se quiere saber si existe diferencia significativa entre la media de la muestra y la media poblacional

Tipos:

- *t para una muestra*: comparar la media muestral con un valor conocido.
- *t para muestras independientes*: comparar medias entre dos grupos diferentes.
- *t para muestras relacionadas (pareadas)*: comparar medidas antes y después en un mismo grupo.
- ( $n < 30$ )

Hernández et al (2014)

## Supuestos de ANOVA

### **ANOVA (Análisis de la Varianza):**

- Permite comparar simultáneamente las medias de tres o más grupos.
- Evalúa si al menos una de las medias difiere significativamente de las demás.
- Se basa en la relación entre la varianza entre grupos y la varianza dentro de los grupos.



#### **Objetivo**

Comparar medias de tres o más grupos.



#### **Normalidad**

Los datos deben seguir una distribución normal.



#### **Homocedasticidad**

Las varianzas entre grupos deben ser iguales.



#### **Independencia**

Las observaciones deben ser independientes entre sí.

# Regresión Lineal

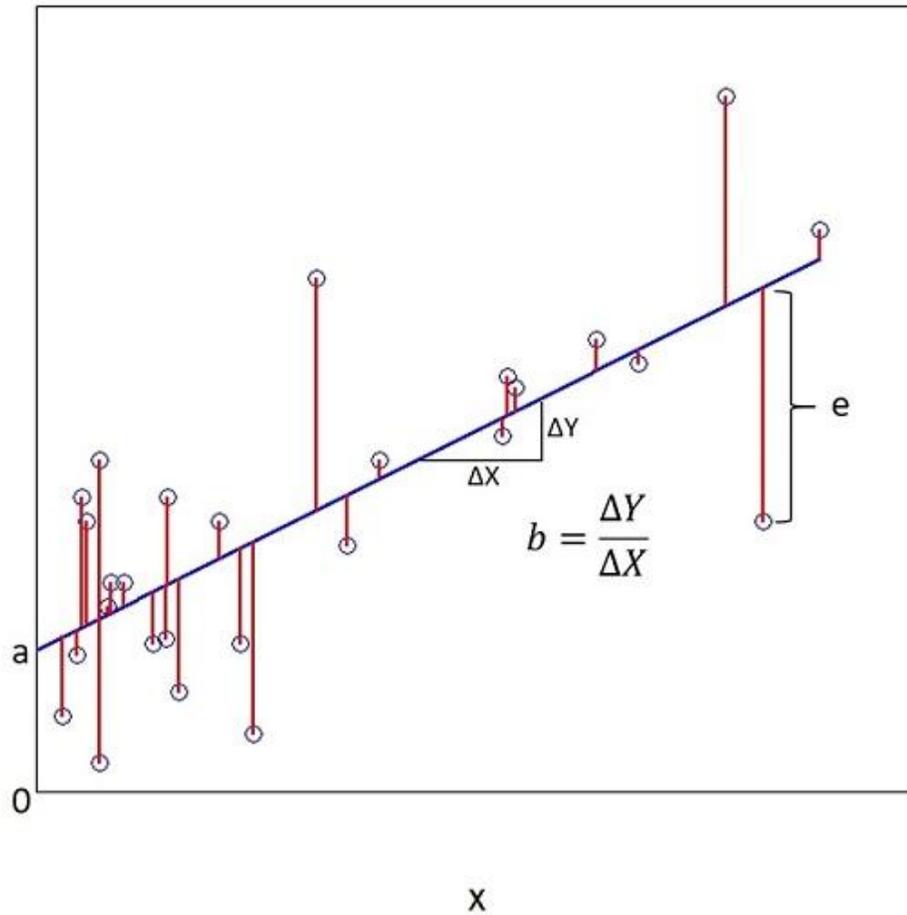
$$y = a + bx + e$$

## Correlación entre dos variables

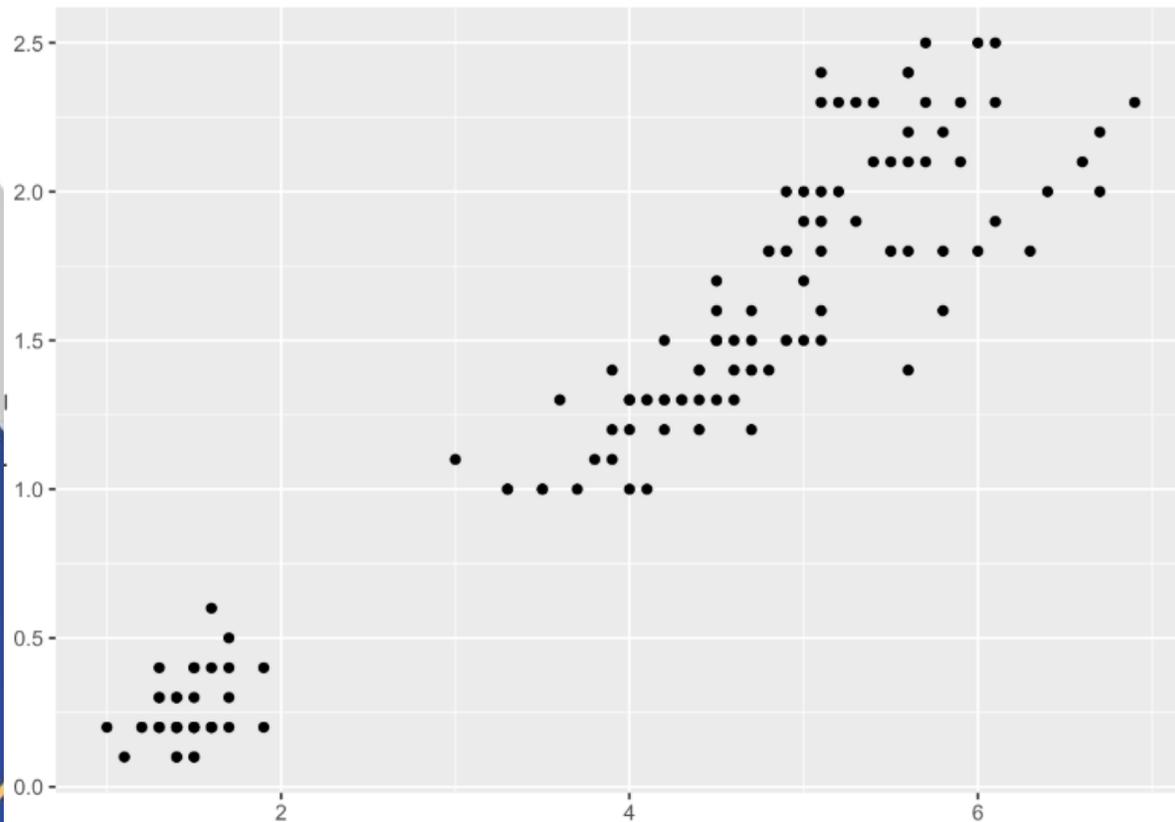
El establecimiento de una *correlación entre dos variables* es importante, pero esto se considera un primer paso para predecir una variable a partir de la otra. (U otras, en el caso de la regresión múltiple.)

Claro está, si sabemos que la variable **X** está muy relacionada con **Y**, ello quiere decir que podemos predecir **Y** a partir de **X**. Estamos ya en el terreno de la predicción. (Evidentemente si, **X** no está relacionada con **Y**, **X** no sirve como *predictor* de **Y**.)

Nota: Emplearemos los términos “regresión” y “predicción” como casi sinónimos. (La razón del uso del término “regresión” es antigua, y se ha mantenido como tal.)



## Interpretación



---

Si la covarianza es:

---

*Positiva*, significa que ambas variables tienden a variar de la misma forma, hay una asociación positiva. Proporcionalidad directa

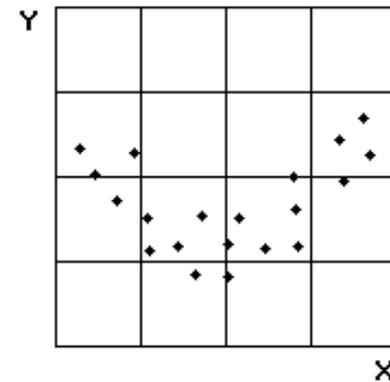
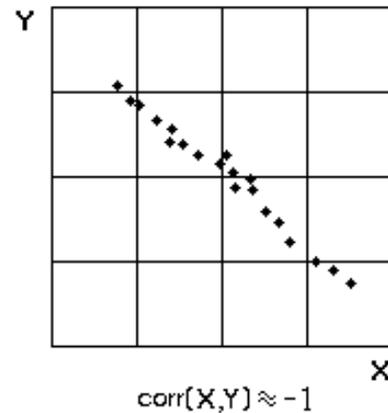
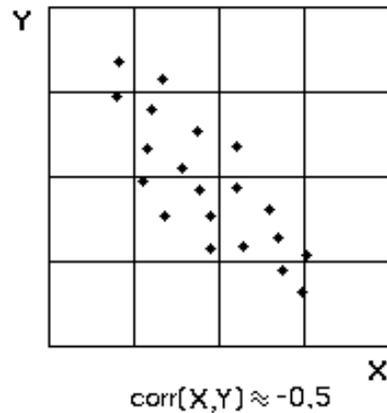
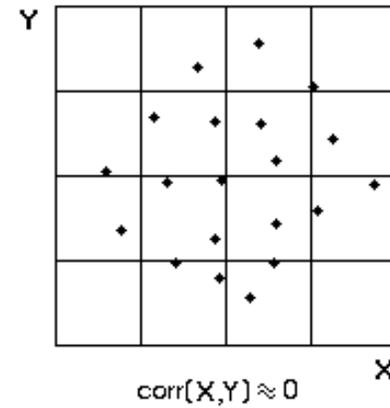
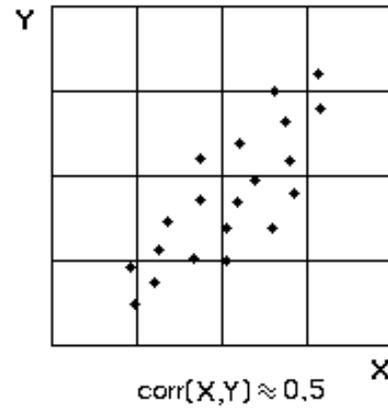
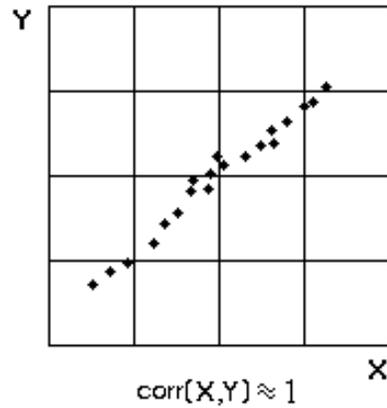
---

*Negativa*, significa que si una aumenta, la otra tiende a disminuir, y vice versa. Proporcionalidad inversa

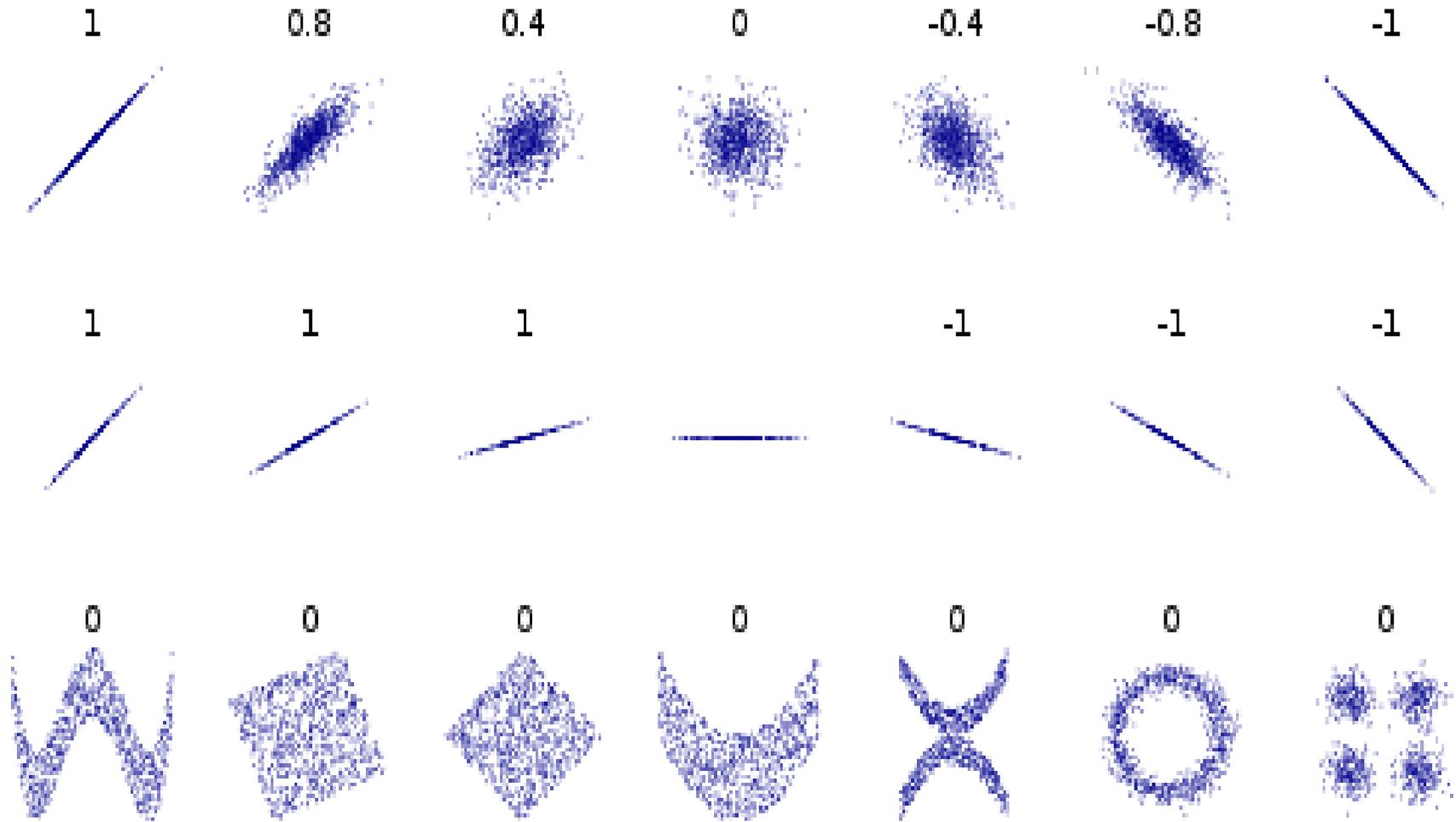
---

*Tiende a cero*, indica que no hay asociación entre las variables.

La interpretación del coeficiente de correlación puede ilustrarse mediante los siguientes gráficos.

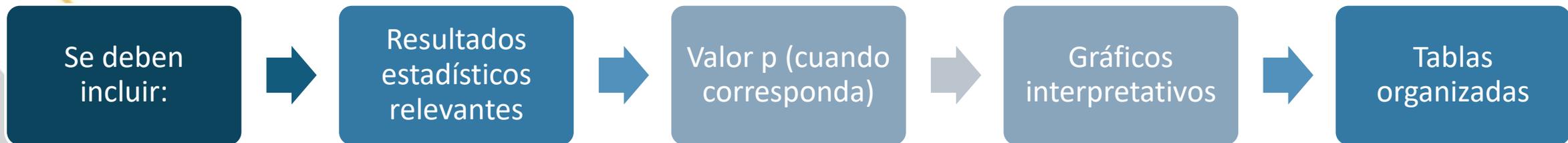


En este caso no hay una asociación lineal, por lo tanto no tiene sentido la correlación



## 5. Redacción de resultados y gráficos

- Una vez analizados los datos, la redacción de resultados debe ser clara, precisa y basada en evidencias.



El uso de gráficos no solo apoya la visualización, sino que contribuye a comunicar los hallazgos de forma más comprensible para audiencias técnicas y no técnicas (APA, 2020). “La presentación visual de los datos es una parte clave del proceso de análisis, ya que permite detectar tendencias, comparar grupos y comunicar hallazgos de forma eficaz” (Field, 2018, p. 215).

# Actividad práctica en clase



# CONTACTOS



**WhatsApp**  
0982901923



**Correo electrónico**  
cursos@aicadep.com



**Sitio web**  
www.aicadep.com



**Redes sociales**  
@aicadep



**AICADEP**

ACADEMIA INTERNACIONAL DE CAPACITACIÓN  
Y DESARROLLO PROFESIONAL

# GRACIAS



[WWW.AICADEP.COM](http://WWW.AICADEP.COM)



[CURSOS@AICADEP.COM](mailto:CURSOS@AICADEP.COM)



0982901923