

# Estadística

## Introducción

# ¿Qué es la estadística?

Métodos y procedimientos destinados a recoger, clasificar, resumir, analizar los datos y realizar inferencias con el fin de ayudar a la toma de decisiones y en su caso formular predicciones.

Podemos responder preguntas cómo:

- ¿Cómo fijar el salario del CEO, en función de las ventas o en función del ROE?
- ¿Cómo se ve afectado los años de formación en el salario de los empleados?
- ¿Qué efecto tienen los gastos de las campañas electorales sobre los resultados de las votaciones?
- ¿Qué efecto tienen los gastos de escolarización sobre el rendimiento de los alumnos?
- ¿Cuál es la máxima pérdida que puede tener una empresa en base a sus datos históricos? ((Valor en Riesgo)
- ¿Cómo se relacionan diferentes hechos o sucesos?

# Clasificación

## ESTADÍSTICA DESCRIPTIVA



Describe, analiza y representa un grupo de datos utilizando métodos numéricos y gráficos que resumen y presentan la información contenida en ellos.

## ESTADÍSTICA INFERENCIAL



A partir del cálculo de probabilidades y datos muestrales, efectúa estimaciones, decisiones, predicciones u otras generalizaciones sobre un conjunto mayor de datos.



**ANÁLISIS DE DATOS, INFERENCIA Y TOMA DE DECISIONES**

# Conceptos básicos



- **Individuos o elementos:** Personas u objetos que contienen cierta información que se desea estudiar.
- **Población:** Conjunto de individuos o elementos que cumplen ciertas propiedades comunes.
- **Muestra:** Subconjunto representativo de una población.
- **Muestreo:** Métodos para la recolección de la muestra.
- **Datos:** Conjunto de valores de una variable para cada uno de los elementos de la muestra.

# Conceptos básicos

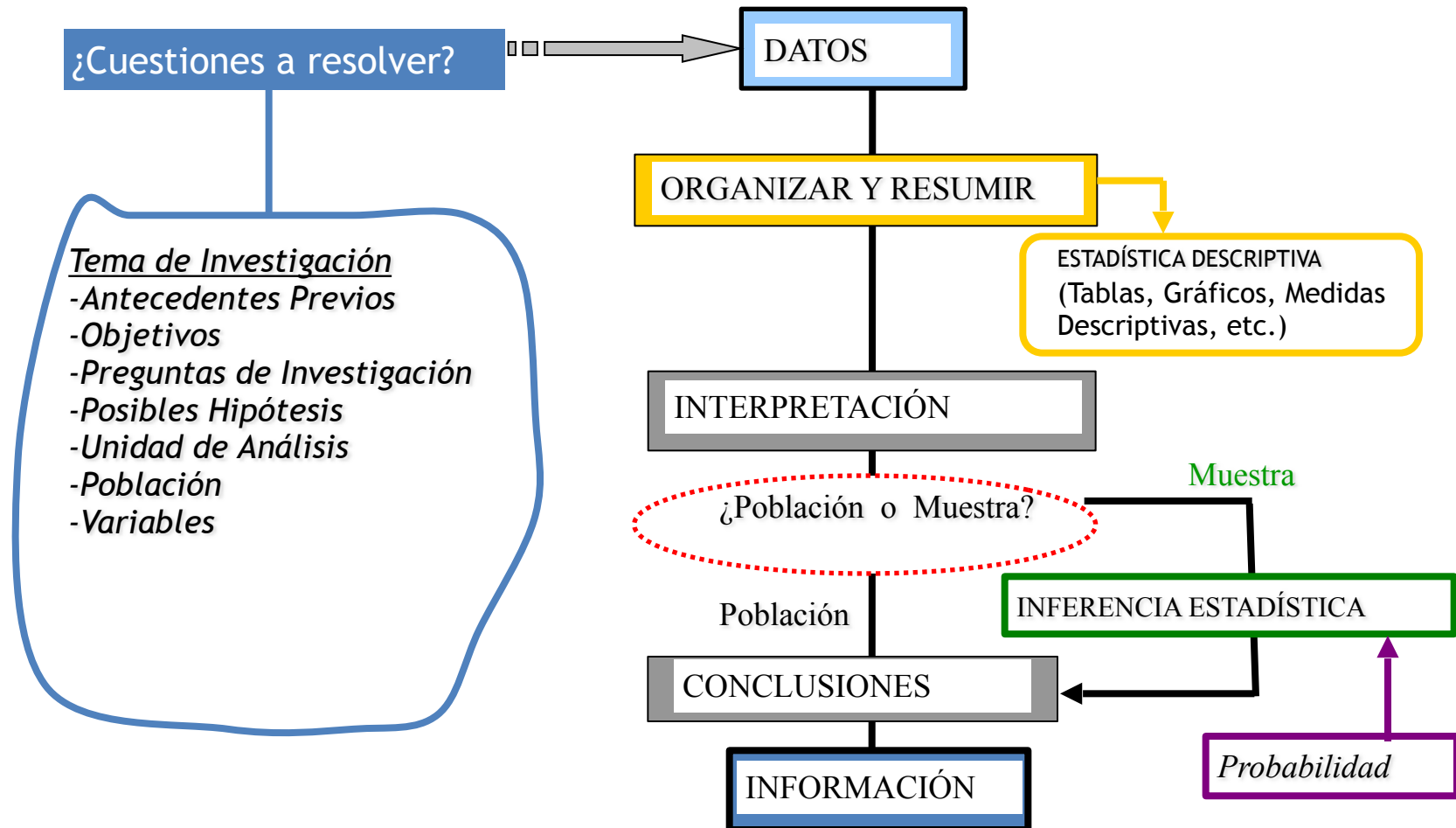


## Etapas de un estudio estadístico

Un análisis estadístico se lleva a cabo siguiendo las etapas habituales en el llamado método científico cuyas etapas son:

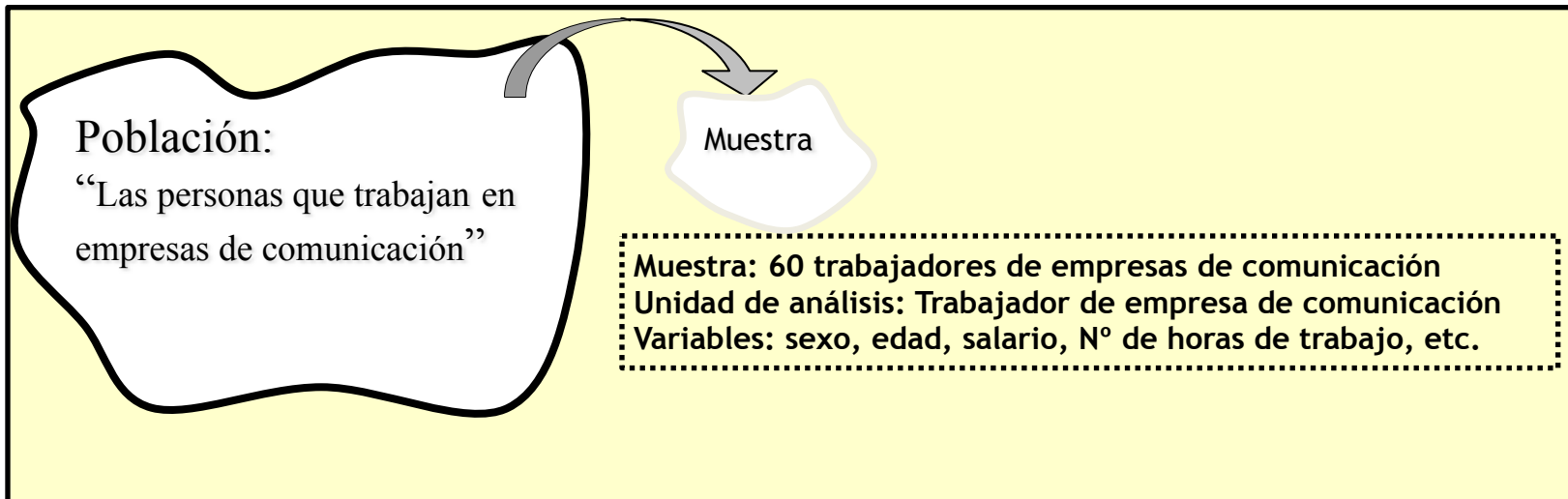
- 1) **Planteamiento del problema:** consiste en definir el objetivo de la investigación y precisar el universo o población.
- 2) **Recogida de la información:** consiste en recolectar los datos necesarios relacionados al problema de investigación.
- 3) **Análisis descriptivo:** consiste en resumir los datos disponibles para extraer la información relevante en el estudio.
- 4) **Inferencia estadística:** consiste en suponer un modelo para toda la población partiendo de los datos analizados para obtener conclusiones generales.
- 5) **Diagnóstico:** consiste en verificar la validez de los supuestos del modelo que nos han permitido interpretar los datos y llegar a conclusiones sobre la población

# Programa de la asignatura



# Conceptos básicos

- **VARIABLE:** es lo que se va a medir y representa una característica de la UNIDAD DE ANÁLISIS.
- **¿QUIÉNES VAN A SER MEDIDOS?:** Los sujetos u objetos o Unidades de Análisis de una Población o una Muestra
- **POBLACIÓN :** Es el total de unidades de análisis que son tema de estudio.
- **MUESTRA:** Es un conjunto de unidades de análisis provenientes de una población.



# Conceptos básicos

Variable: corresponde a la característica de la Unidad de Análisis

## TIPOS DE VARIABLES

### Variables Cuantitativas

#### CONTINUA

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, Grado de Interés por un tema, etc.*



# Conceptos básicos

## MEDIDAS DE TENDENCIA CENTRAL

- Media Aritmética (Promedio)
- Mediana
- Moda

### Datos Cuantitativos

$x$
$x_1$
$x_2$
?
$x_n$

### Media Aritmética o Promedio

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

### Datos Cuantitativos ordenados de menor a mayor

### Mediana

$$M_E = x_{(k)} \quad \text{Si } n \text{ es impar}$$

$$M_E = \frac{x_{(k)} + x_{(k+1)}}{2} \quad \text{Si } n \text{ es par}$$

$$x_{(k)} = \text{dato del centro}$$

Datos  
Cualitativos y Cuantitativos

### Moda

$M_o$  = "el dato que más se repite"

# Conceptos básicos

## Percentiles, Deciles o Cuartiles

- Percentil (ejemplo: 25, 50, 75)
- Decil (ejemplo: 4, 5, 8)
- Cuartil (ejemplo: 1, 2, 3)

Percentil, Decil o Cuartil: corresponde al valor que toma la variable (cuantitativa), cuando los  $n$  datos están ordenados de **Menor** a **Mayor**

El Percentil va de 1 a 100

El percentil 25 (25/100): es el valor de la variable que reúne al menos el 25% de los datos

Ejemplo: Si  $N=80$ , el 25% de 80 es 20; por lo tanto, se busca el dato que este en la posición 20.

Si  $N=85$ , el 25% de 85 es 21,25; por lo tanto se busca el dato que este en la posición 22.

El Decil va de 1 a 10

El Decil 4 (4/10): es el valor de la variable que reúne al menos el 40% de los datos

Ejemplo: Si  $N=80$ , el 40% de 80 es 32; por lo tanto, se busca el dato que este en la posición 32.

Si  $N=85$ , el 40% de 85 es 34; por lo tanto se busca el dato que este en la posición 34.

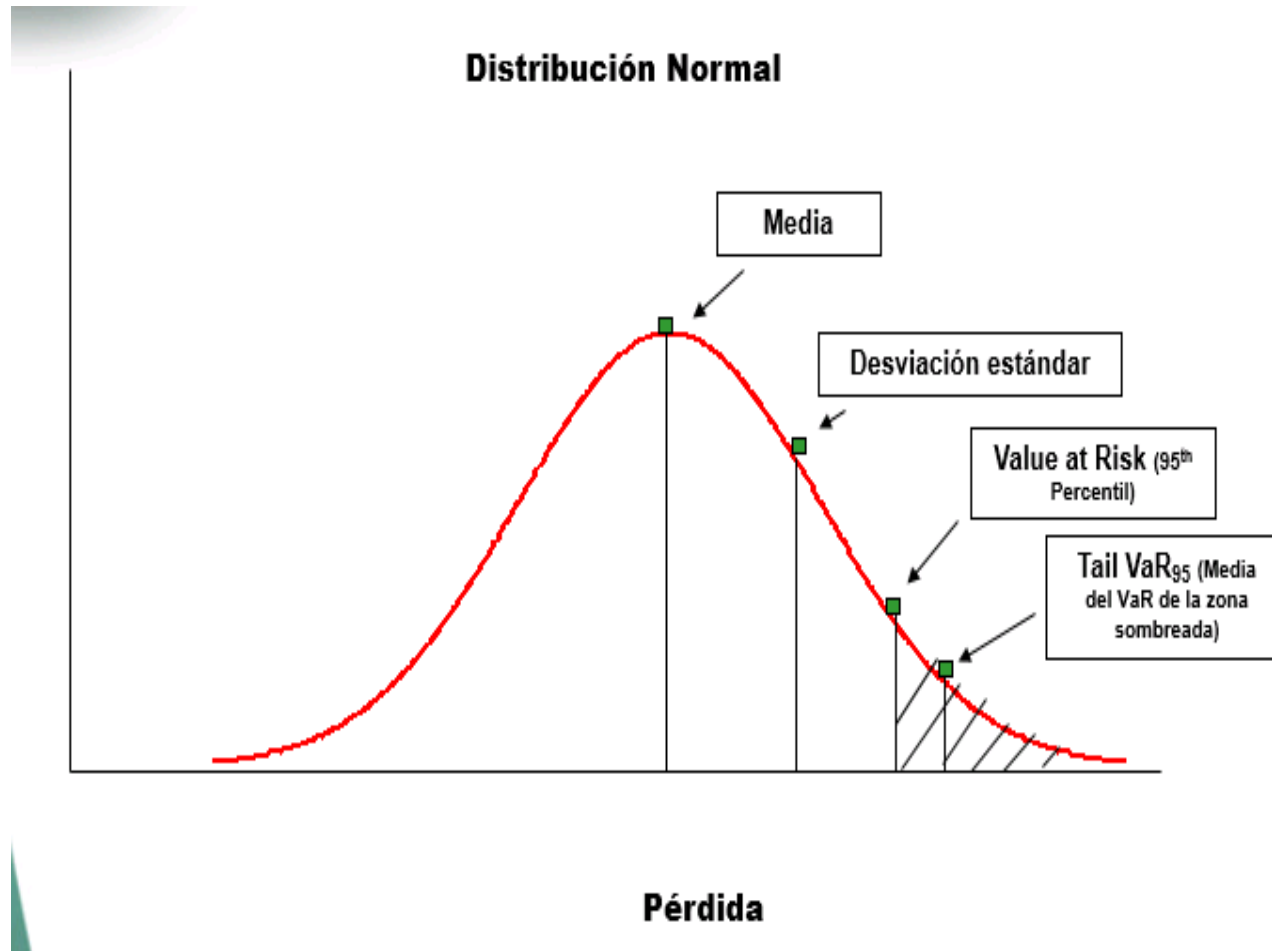
El Cuartil va de 1 a 4

El Cuartil 3 (3/4): es el valor de la variable que reúne al menos el 75% de los datos

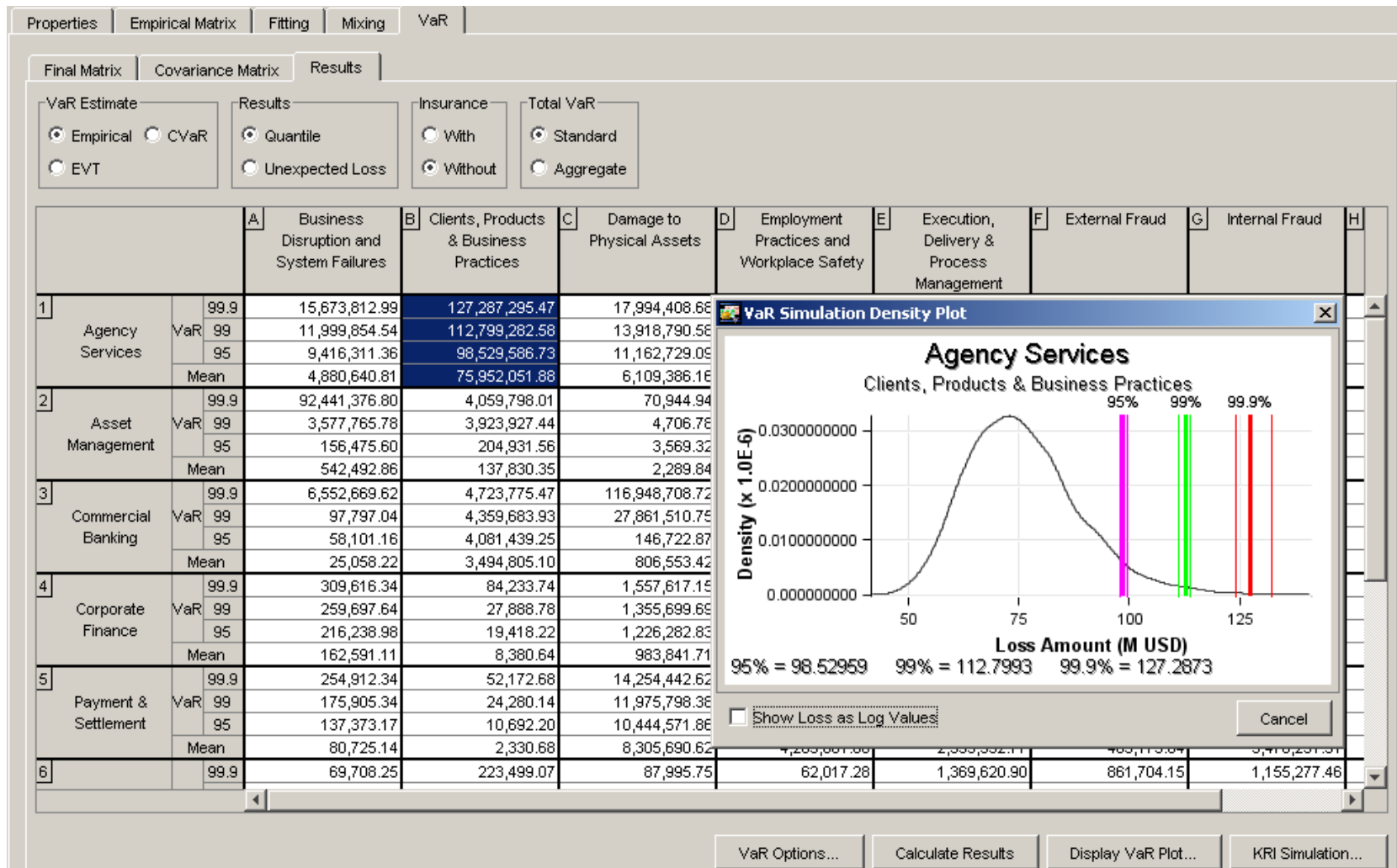
Ejemplo: Si  $N=80$ , el 75% de 80 es 60; por lo tanto, se busca el dato que este en la posición 60.

Si  $N=85$ , el 75% de 85 es 63,75; por lo tanto se busca el dato que este en la posición 64.

# Valor en Riesgo- Aplicación



# Cuantificación del riesgo

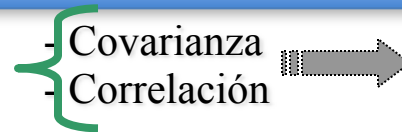


# Ejemplo: Indicadores de Riesgo



# Conceptos básicos

## MEDIDAS DE ASOCIACIÓN LINEAL



Datos  
Cuantitativos

$x$
$x_1$
$x_2$
?
$x_n$

Recordemos que: *Hasta ahora hemos estudiado las medidas tendencia central (Media, Mediana, Moda) y dispersión (Varianza y Desviación Estándar) para una Variable Cuantitativa (x).*

**Covarianza:** *Es una medida de Variabilidad Conjunta entre dos variables ( $x_1, x_2$ ) o bien ( $x, y$ )*

$x$	$y$
$x_{(1)}$	$y_{(1)}$
$x_{(2)}$	$y_{(2)}$
?	?
$x_{(n)}$	$y_{(n)}$

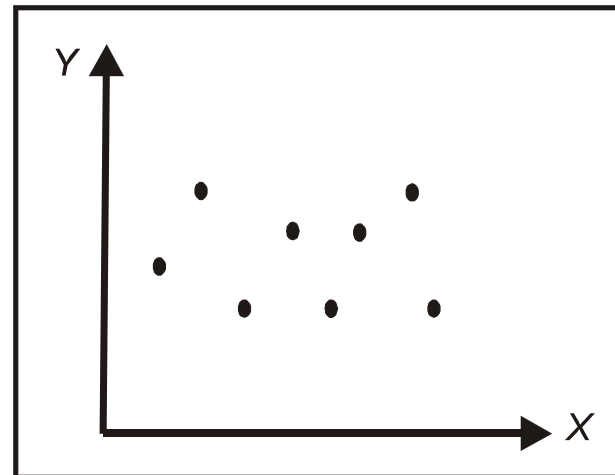
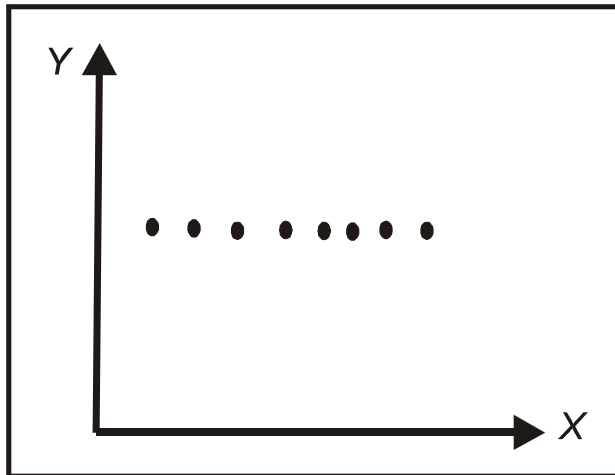
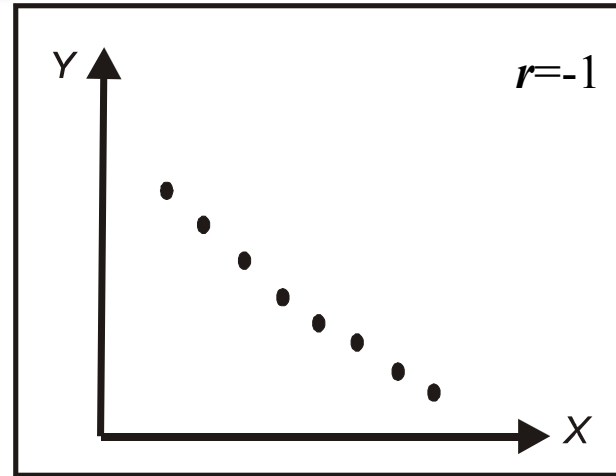
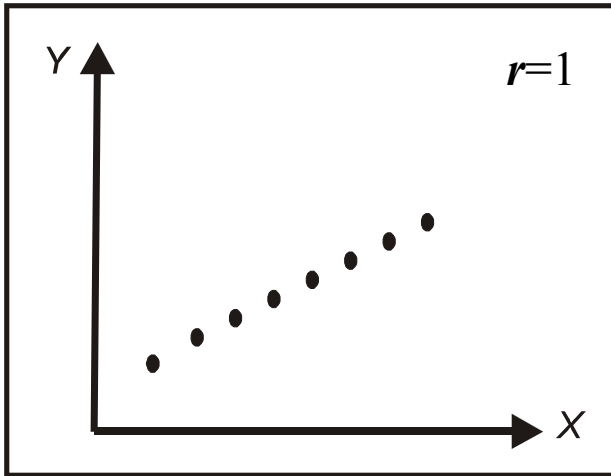
$$cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Si Cov(x,y) es positiva: la asociación entre x e y es directamente proporcional, es decir que cuando x aumenta y también aumenta; y viceversa.

Si Cov(x,y) es negativa: la asociación entre x e y es inversamente proporcional, es decir que cuando x aumenta y disminuye; y viceversa.

Si Cov(x,y) es cero: no existe asociación entre x e y.

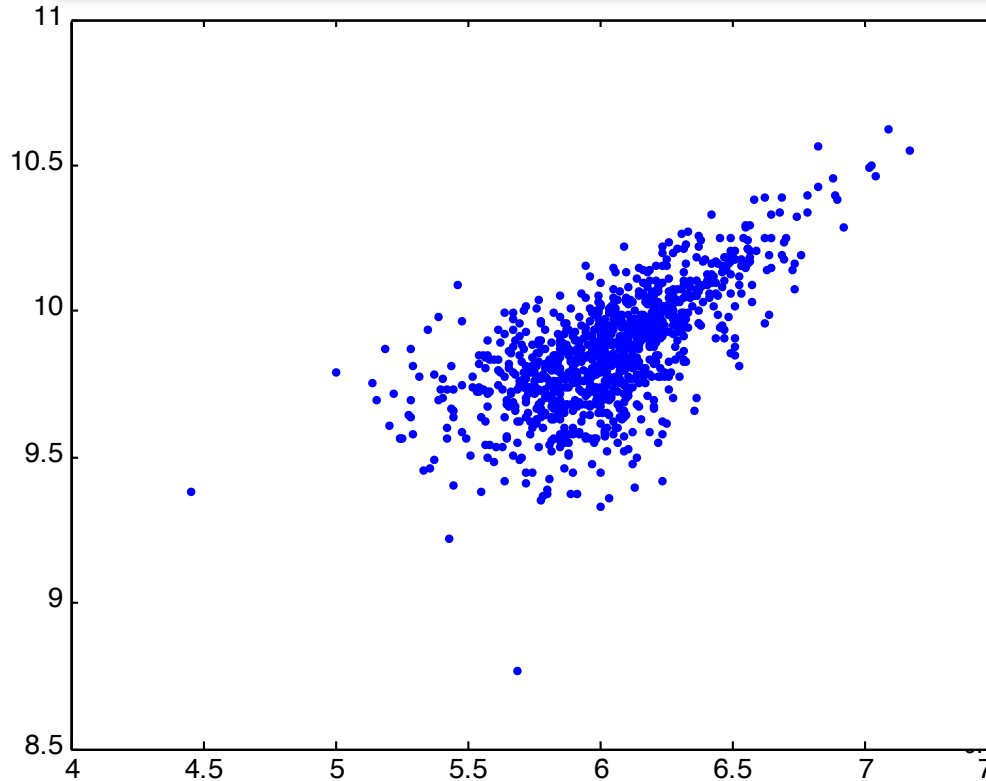
# Conceptos básicos



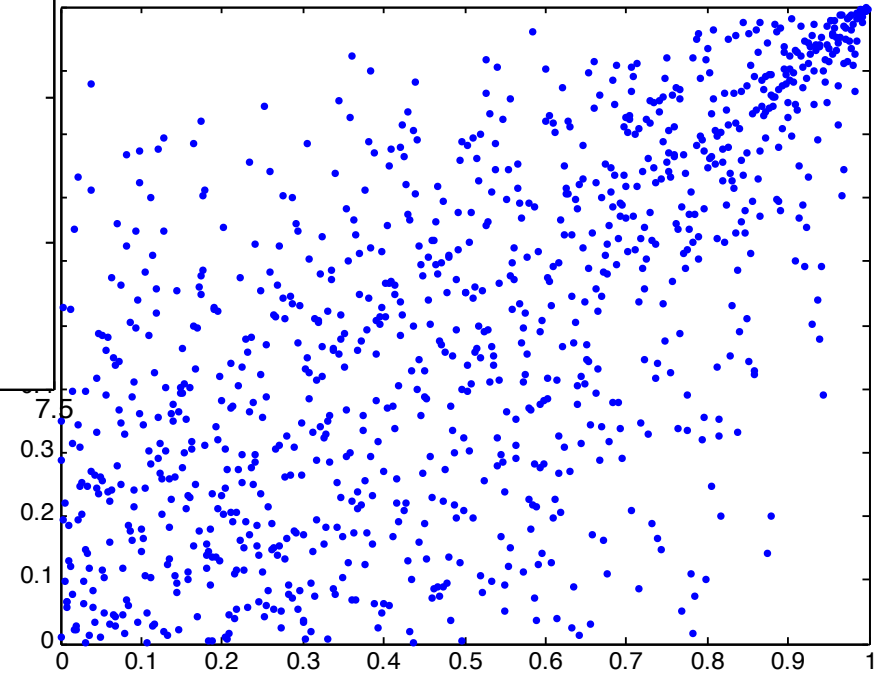
EJEMPLO : Representación gráfica de las variables x e y

# Conceptos básicos

x vs. y



u vs v





## Análisis de datos

Para la presentación de este análisis van a ser utilizados los que aparecen en el archivo: supervivencia2

```
> supervivencia2
```

	V1	V2	V3	V4
1	25.5	1.74	5.30	10.8
2	31.2	6.32	5.42	9.4
3	25.9	6.22	8.41	7.2
4	38.4	10.52	4.63	8.5
5	18.4	1.19	11.60	9.4
6	26.7	1.22	5.85	9.9
7	26.4	4.10	6.62	8.0
8	25.9	6.32	8.72	9.1
9	32.0	4.08	4.42	8.7
10	25.2	4.15	7.60	9.2
11	39.7	10.15	4.83	9.4
12	35.7	1.72	3.12	7.6
13	26.5	1.70	5.30	8.2

# Análisis de datos

```

> y<-supervivencia[[1]]
> y
[1] 25.5 31.2 25.9 38.4 18.4 26.7 26.4 25.9 32.0 25.2 39.7 35.7 26.5
> x1<-supervivencia[[2]]
> x1
[1] 1.74 6.32 6.22 10.52 1.19 1.22 4.10 6.32 4.08 4.15 10.15 1.72 1.70
> x2<-supervivencia[[3]]
> x2
[1] 5.30 5.42 8.41 4.63 11.60 5.85 6.62 8.72 4.42 7.60 4.83 3.12 5.30
> x3<-supervivencia[[4]]
> x3
[1] 10.8 9.4 7.2 8.5 9.4 9.9 8.0 9.1 8.7 9.2 9.4 7.6 8.2

```

```

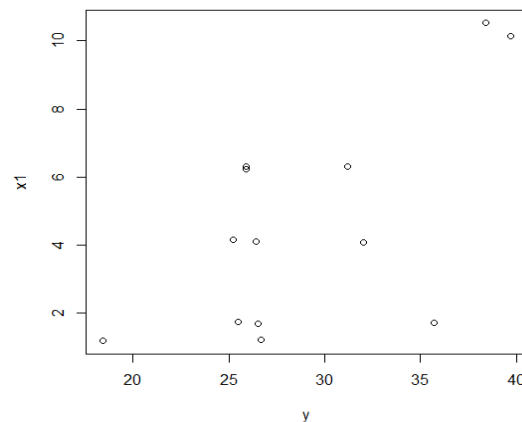
> plot(x1,y)

```

```

> cor(y,x1)
[1] 0.6538538
> cor(y,x2)
[1] -0.785805
> cor(y,x3)
[1] -0.186275
> cor(x1,x2)
[1] -0.153499
> cor(x1,x3)
[1] -0.1452203
> cor(x2,x3)
[1] 0.0748444

```



# Conceptos básicos: Regresión lineal

## *EJEMPLO: Aplicación del Modelo de Regresión Lineal Simple*

Se cuenta con las mediciones sobre la edad y la talla de 14 niños, y estamos interesados en determinar si existe algún tipo de relación entre la talla del niño y su edad.

### **Modelo Estimado**

$$\hat{y} = a + bx$$

$$b = 2,44 \quad a = 53,64$$

$$\hat{y} = 53,64 + 2,44x$$

### Interpretación de los resultados

- Existe asociación o dependencia entre la Talla del niño y la edad ( $r=0,88$ ); a medida que la edad aumenta la talla aumenta.
- Desde los resultados del modelo de regresión lineal simple, se tiene que la talla media de un niño es de 53,64 cm. Cuando la edad del niño (meses) aumenta en una unidad la talla se incrementa en 2,44 cm.

# Regresión lineal

•**Ejemplo:** Cursos de formación profesional y la productividad de los trabajadores.

- Cuestión: **Un empresario quiere examinar los efectos de los cursos de formación sobre la productividad de los trabajadores.** Se tiene en cuenta que el salario de los trabajadores se establece según su productividad. La ecuación sería la siguiente:

$$wage = f(educ, exper, training)$$

**Educación:** Años de educación, formación académica

**Exper:** Años de experiencia en el trabajo

**Training:** Semanas de los cursos de formación

# Introducción teórica-básica de conceptos

## Regresión Lineal

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 training + u,$$

**Educación:** Años de educación, formación académica

**Exper:** Años de experiencia en el trabajo

**Training:** Semanas de los cursos de formación

$B_0$ = Intercepto. Parámetro del modelo econométrico

$B_1, B_2, B_3$ = Parámetro del modelo econométrico

$u$ = Término error, todos aquellos factores que influyen en el salario/productividad que no han sido variables incluidas en la ecuación. Por ejemplo: habilidad innata, calidad de la educación o el entorno familiar.

