

## ESTADÍSTICA Y LOS PROYECTOS CON DATOS

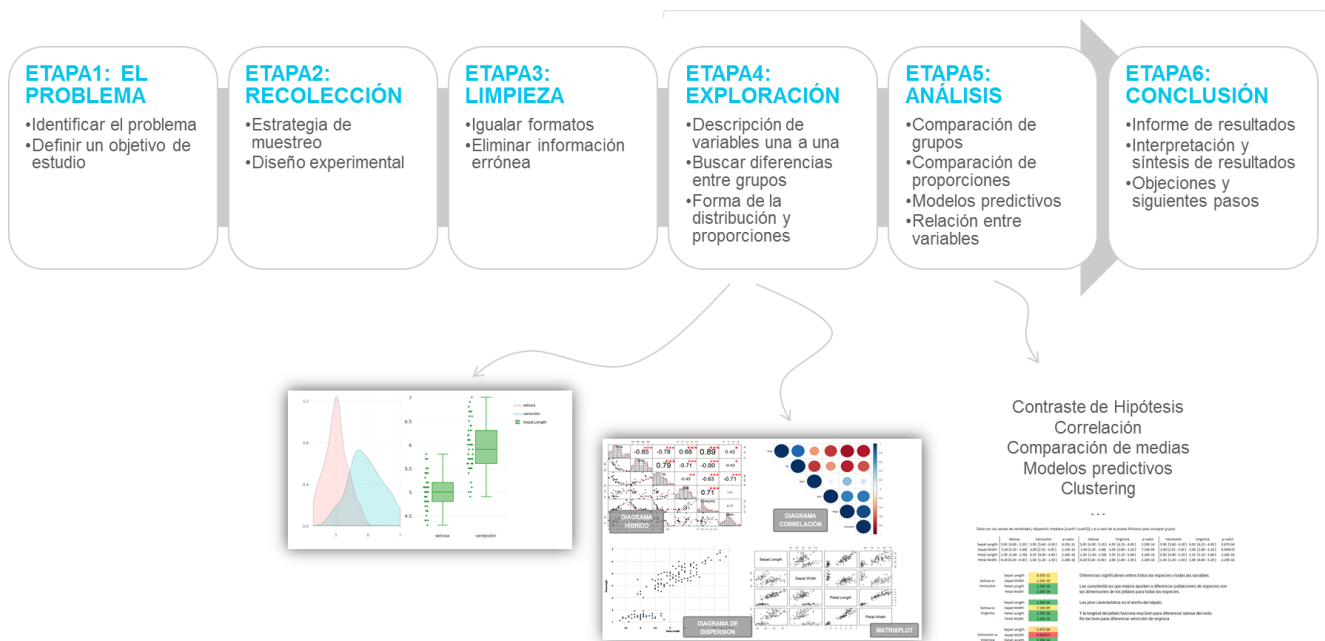
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

El papel de la estadística

La estadística modela los datos para poderlos interpretar.

Etapas de un proyecto con datos

LA ETAPA DE ANÁLISIS DE DATOS (ETAPA4 + ETAPA5 + ETAPA 6) ESTÁ DIRECTAMENTE RELACIONADA CON UN SÓLO OBJETIVO ALINEADO CON EL PROBLEMA DEFINIDO EN LA ETAPA 1



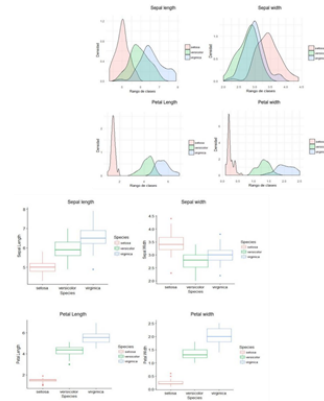
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Las dos estadísticas: descriptiva e inferencial**Estadística Descriptiva**

Medimos y observamos  
datos mediante variables

Fisher's Iris Data

| Largo de sépalo | Ancho de sépalo | Largo de pétalo | Ancho de pétalo | Especies             |
|-----------------|-----------------|-----------------|-----------------|----------------------|
| 5.0             | 2.0             | 3.5             | 1.0             | <i>I. versicolor</i> |
| 6.0             | 2.2             | 4.0             | 1.0             | <i>I. versicolor</i> |
| 6.0             | 2.2             | 5.0             | 1.5             | <i>I. virginica</i>  |
| 6.2             | 2.2             | 4.5             | 1.5             | <i>I. versicolor</i> |
| 4.5             | 2.3             | 1.3             | 0.3             | <i>I. setosa</i>     |
| 5.0             | 2.3             | 3.3             | 1.0             | <i>I. versicolor</i> |
| 5.5             | 2.3             | 4.0             | 1.3             | <i>I. versicolor</i> |
| 6.3             | 2.3             | 4.4             | 1.3             | <i>I. versicolor</i> |
| 4.9             | 2.4             | 3.3             | 1.0             | <i>I. versicolor</i> |
| 5.5             | 2.4             | 3.7             | 1.0             | <i>I. versicolor</i> |
| 5.5             | 2.4             | 3.8             | 1.1             | <i>I. versicolor</i> |
| 4.9             | 2.5             | 4.5             | 1.7             | <i>I. virginica</i>  |
| 5.1             | 2.5             | 3.0             | 1.1             | <i>I. versicolor</i> |
| 5.5             | 2.5             | 4.0             | 1.3             | <i>I. versicolor</i> |
| 5.6             | 2.5             | 3.9             | 1.1             | <i>I. versicolor</i> |
| 5.7             | 2.5             | 5.0             | 2.0             | <i>I. virginica</i>  |
| 6.3             | 2.5             | 4.9             | 1.5             | <i>I. versicolor</i> |
| 6.3             | 2.5             | 5.0             | 1.9             | <i>I. virginica</i>  |
| 6.7             | 2.5             | 5.8             | 1.8             | <i>I. virginica</i>  |
| 5.5             | 2.6             | 4.4             | 1.2             | <i>I. versicolor</i> |
| 5.7             | 2.6             | 3.5             | 1.0             | <i>I. versicolor</i> |
| 5.8             | 2.6             | 4.0             | 1.2             | <i>I. versicolor</i> |
| 6.1             | 2.6             | 5.6             | 1.4             | <i>I. virginica</i>  |



Interpretar la tabla de  
datos con gráficos y  
Tablas

**Estadística Inferencial**

```
call:
glmFormula = l1feap ~ log(gdpPerCap) + continent, data = data2007)
Deviance residuals:
      min       25       50       75      100      125      150      175      200
-19.4927  -2.3248  -0.9432   2.5498  14.8818
Coefficients:
(Intercept)      Estimate Std. Error t value Pr(>|t|)
log(gdpPerCap)  4.0376     4.0376    1.000 1.000e+00 ***
log(gdpPerCap)  4.0308     0.5274    7.662 6.14e-15 ***
continentAmericas 1.0542     1.0542    1.000 1.000e+00 ***
continentAsia    10.1144    1.4761    6.852 2.35e-10 ***
continentEurope  12.1682     1.8936    6.428 2.16e-08 ***
continentOceania 12.9293     4.5211    2.860 0.00491 **
---
Signif. codes:  0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1
Diagnostics:
(1) Residuals: Normal Q-Q plot, for gaussian family taken to be 35.15473)
on 141 degrees of freedom
on 138 degrees of freedom
Wald tests:
(1) Wald tests: 2
```



Interpretar resultados  
y extraer  
conclusiones

- La estadística DESCRIPTIVA traduce dibujo la tabla de datos
- La estadística INFERENCIAL modelo los datos para obtener conclusiones generales de la población

## TABLA DE DATOS Y VARIABLES

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Observaciones o individuos:

Distintas mediciones de la realidad (FILAS)

## Variables

Características que estamos midiendo (COLUMNAS)

## Estructura de una tabla de datos

| Fisher's Iris Data |                 |                 |                 |                  |
|--------------------|-----------------|-----------------|-----------------|------------------|
| Largo de sépalo    | Ancho de sépalo | Largo de pétalo | Ancho de pétalo | Especies         |
| 5.1                | 3.5             | 1.4             | 0.2             | <i>I. setosa</i> |
| 4.9                | 3.0             | 1.4             | 0.2             | <i>I. setosa</i> |
| 4.7                | 3.2             | 1.3             | 0.2             | <i>I. setosa</i> |
| 4.6                | 3.1             | 1.5             | 0.2             | <i>I. setosa</i> |
| 5.0                | 3.6             | 1.4             | 0.2             | <i>I. setosa</i> |
| 5.4                | 3.9             | 1.7             | 0.4             | <i>I. setosa</i> |
| 4.6                | 3.4             | 1.4             | 0.3             | <i>I. setosa</i> |
| 5.0                | 3.4             | 1.5             | 0.2             | <i>I. setosa</i> |
| 4.4                | 2.9             | 1.4             | 0.2             | <i>I. setosa</i> |
| 4.9                | 3.1             | 1.5             | 0.1             | <i>I. setosa</i> |
| 5.4                | 3.7             | 1.5             | 0.2             | <i>I. setosa</i> |
| 4.8                | 3.4             | 1.6             | 0.2             | <i>I. setosa</i> |
| 4.8                | 3.0             | 1.4             | 0.1             | <i>I. setosa</i> |
| 4.3                | 3.0             | 1.1             | 0.1             | <i>I. setosa</i> |

- Variable Numérica "Largo de pétalo"
- Individuo número 4 (fila 4) Se miden las características de cada variable
- Header. Nombres de las variables
- Valores de las variables de todas las observaciones

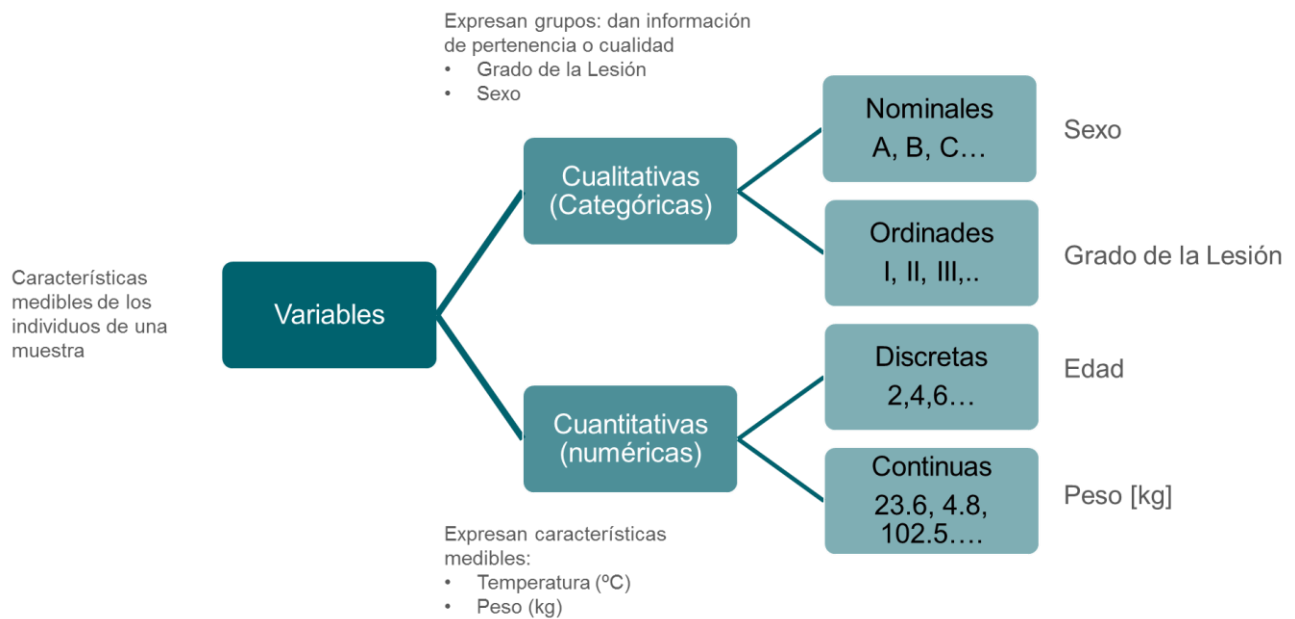
**FILAS**  
individuos u  
observaciones

**COLUMNAS**  
variables

- Variables: columnas
- Observaciones: filas
- Header (nombre de las variables): primera fila

## Tipos de variables

- o Cuantitativa discreta
- o Cuantitativa continua
- o Cualitativa nominal
- o Cualitativa ordinal



## ESTADÍSTICA DESCRIPTIVA PARA VARIABLES CUANTITATIVAS

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Tabla de frecuencia e histograma (continuas)

Para describir una variable cuantitativa continua podemos utilizar

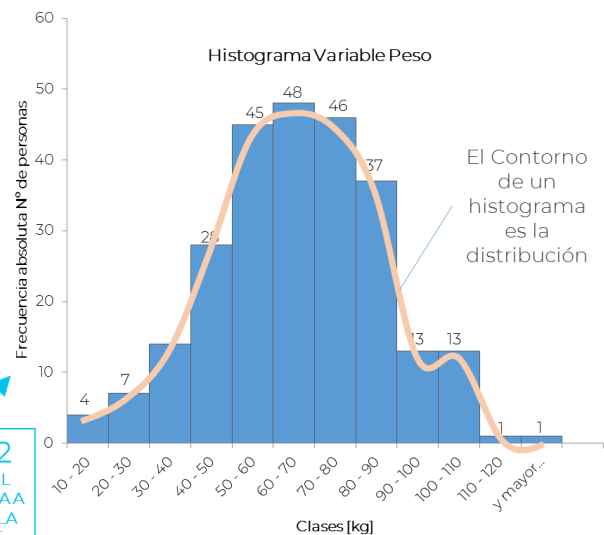
Tabla de Datos

| Sexo   | Peso [kg] |
|--------|-----------|
| HOMBRE | 72.72     |
| HOMBRE | 69.12     |
| HOMBRE | 65.26     |
| MUJER  | 49.35     |
| MUJER  | 61.84     |
| HOMBRE | 64.82     |
| MUJER  | 72.07     |
| HOMBRE | 71.48     |
| HOMBRE | 82.92     |
| MUJER  | 51.07     |
| HOMBRE | 90.47     |
| HOMBRE | 82.24     |
| MUJER  | 47.70     |
| MUJER  | 64.32     |
| MUJER  | 38.07     |
| MUJER  | 59.48     |
| ...    | ...       |

Tabla de Frecuencias

| Clases [kg] | Frecuencia |
|-------------|------------|
| 10 - 20     | 4          |
| 20 - 30     | 7          |
| 30 - 40     | 14         |
| 40 - 50     | 28         |
| 50 - 60     | 45         |
| 60 - 70     | 48         |
| 70 - 80     | 46         |
| 80 - 90     | 37         |
| 90 - 100    | 13         |
| 100 - 110   | 13         |
| 110 - 120   | 1          |
| y mayor...  | 1          |

Histograma



**PASO1**  
CONSTRUIR  
UNA TABLA DE  
FRECUENCIAS  
y UN  
HISTOGRAMA  
CON CLASES  
DE 10 EN 10 kg

**PASO2**  
DIBUJAR EL  
HISTOGRAMA A  
PARTIR DE LA  
TABLA DE  
FRECUENCIAS

- Las clases / rangos o intervalos son las fronteras de la tabla de frecuencias ¡La dimensión del intervalo depende de ti!
- Las frecuencias son el número de individuos por rango de clase
- El histograma es la herramienta gráfica para pintar una tabla de frecuencias
- El contorno de un histograma es la distribución

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Frecuencias absolutas, relativas y acumuladas

| Es un intervalo de edad | Contar las personas de cada rango de edad | La proporción de la frecuencia absoluta al total de personas de |                 | Sumas una fila con la anterior de la frecuencia absoluta v vas |                        | Sumas una fila con la anterior de la frecuencia relativa v vas acumulando |                 |
|-------------------------|---|---|-----------------|--|------------------------|---|-----------------|
| CLASE                   | $n_i$<br>Frecuencia Absoluta              | $f_i$<br>Frecuencia Relativa                                    |                 | $N_i$<br>Fr. Abs. Acumulada                                    |                        | $F_i$<br>Fr. Rel. Acumulada   |                 |
| AÑOS                    | HOMBRES MUJERES                           | HOMBRES MUJERES   | HOMBRES MUJERES | HOMBRES MUJERES  | HOMBRES MUJERES        | HOMBRES MUJERES   | HOMBRES MUJERES |
| 0-5                     | 197054 185453                             | 0.0534 0.0486   | 0.0534 0.0486   | 197054 185453  | 0.053382406 0.04861632 |   |                 |
| 5-10                    | 216624 203631                             | 0.0587 0.0534   | 0.0587 0.0534   | 413678 389084  | 0.112066373 0.101998   |   |                 |
| 10-15                   | 195043 184911                             | 0.0528 0.0485   | 0.0528 0.0485   | 608721 573995  | 0.164903995 0.15047224 |   |                 |
| 15-20                   | 178372 165651                             | 0.0483 0.0434   | 0.0483 0.0434   | 787093 739646  | 0.213225402 0.19389749 |   |                 |
| 20-25                   | 185122 177674                             | 0.0501 0.0466   | 0.0501 0.0466   | 972215 917320  | 0.263375401 0.24047455 |   |                 |
| 25-30                   | 210456 213789                             | 0.0570 0.0560   | 0.0570 0.0560   | 1182671 1131109  | 0.320388442 0.29651913 |   |                 |
| 30-35                   | 271812 269694                             | 0.0736 0.0707   | 0.0736 0.0707   | 1454483 1400803  | 0.394022971 0.36721915 |   |                 |
| 35-40                   | 344746 324997                             | 0.0934 0.0852   | 0.0934 0.0852   | 1799229 1725800  | 0.487415499 0.4524168  |   |                 |
| 40-45                   | 333839 312522                             | 0.0904 0.0819   | 0.0904 0.0819   | 2133068 2038322  | 0.577853293 0.53434415 |   |                 |
| 45-50                   | 297218 285965                             | 0.0805 0.0750   | 0.0805 0.0750   | 2430286 2324287  | 0.65837037 0.60930959  |   |                 |
| 50-55                   | 260913 263357                             | 0.0707 0.0690   | 0.0707 0.0690   | 2691199 2587644  | 0.729052335 0.67834838 |   |                 |
| 55-60                   | 226590 238244                             | 0.0614 0.0625   | 0.0614 0.0625   | 2917789 2825888  | 0.790436115 0.74080381 |   |                 |
| 60-65                   | 193585 209227                             | 0.0524 0.0548   | 0.0524 0.0548   | 3111374 3035115  | 0.842878761 0.79565247 |   |                 |
| 65-70                   | 179496 201102                             | 0.0486 0.0527   | 0.0486 0.0527   | 3290870 3236217  | 0.891504663 0.84837116 |   |                 |
| 70-75                   | 141809 165957                             | 0.0384 0.0435   | 0.0384 0.0435   | 3432679 3402174  | 0.929921064 0.89187663 |   |                 |
| 75-80                   | 102973 134753                             | 0.0279 0.0353   | 0.0279 0.0353   | 3535652 3536927  | 0.9578167 0.927202     |   |                 |
| 80-85                   | 88489 134824                              | 0.0240 0.0353   | 0.0240 0.0353   | 3624141 3671751  | 0.981788584 0.96254598 |   |                 |
| 85-90                   | 47485 90284                               | 0.0129 0.0237   | 0.0129 0.0237   | 3671626 3762035  | 0.994652386 0.98621384 |   |                 |
| 90-95                   | 16971 42444                               | 0.0046 0.0111   | 0.0046 0.0111   | 3688597 3804479  | 0.999249871 0.9973405  |   |                 |
| 95-100                  | 2769 10145                                | 0.0008 0.0027   | 0.0008 0.0027   | 3691366 3814624  | 1 1                    |   |                 |
| TOTAL                   | 3691366 3814624                           |   |                 |  |                        |   |                 |
|                         | HOMBRES MUJERES                           |   |                 |  |                        |   |                 |
|                         | $N$                                       |   |                 |  |                        |   |                 |
|                         | El total de hombres y mujeres             |   |                 |  |                        |   |                 |

Frecuencias Absolutas ( $n_i$ ): el número de individuos de cada clase

Frecuencias Relativas ( $f_i$ ): la proporción de la frecuencia absoluta al total de individuos

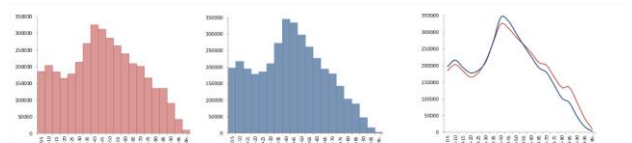
$N$ : total de individuos

Frecuencias Acumuladas: sumar las frecuencias acumulando por filas (clases)

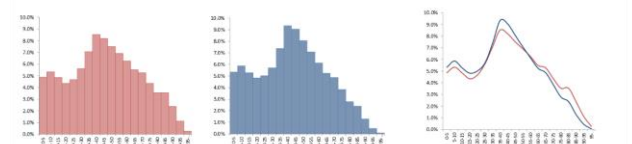
| Es un intervalo de edad | Contar las personas de cada rango de edad | La proporción de la frecuencia absoluta al total de personas de |                 | Sumas una fila con la anterior de la frecuencia absoluta v vas |                 | Sumas una fila con la anterior de la frecuencia relativa v vas acumulando |                 |
|-------------------------|---|---|-----------------|--|-----------------|---|-----------------|
| CLASE                   | $n_i$<br>Frecuencia Absoluta              | $p_i$<br>Porcentaje   |                 | $N_i$<br>Fr. Abs. Acumulada                                    |                 | $P_i$<br>% Acumulado  |                 |
| AÑOS                    | HOMBRES MUJERES                           | HOMBRES MUJERES   | HOMBRES MUJERES | HOMBRES MUJERES  | HOMBRES MUJERES | HOMBRES MUJERES   | HOMBRES MUJERES |
| 0-5                     | 197054 185453                             | 5.3% 4.9%   | 5.3% 4.9%       | 197054 185453  | 5.3% 4.9%       |   |                 |
| 5-10                    | 216624 203631                             | 5.9% 5.3%   | 5.9% 5.3%       | 413678 389084  | 11.2% 10.2%     |   |                 |
| 10-15                   | 195043 184911                             | 5.3% 4.8%   | 5.3% 4.8%       | 608721 573995  | 16.5% 15.0%     |   |                 |
| 15-20                   | 178372 165651                             | 4.8% 4.3%   | 4.8% 4.3%       | 787093 739646  | 21.3% 19.4%     |   |                 |
| 20-25                   | 185122 177674                             | 5.0% 4.7%   | 5.0% 4.7%       | 972215 917320  | 26.3% 24.0%     |   |                 |
| 25-30                   | 210456 213789                             | 5.7% 5.6%   | 5.7% 5.6%       | 1182671 1131109  | 32.0% 29.7%     |   |                 |
| 30-35                   | 271812 269694                             | 7.4% 7.1%   | 7.4% 7.1%       | 1454483 1400803  | 39.4% 36.7%     |   |                 |
| 35-40                   | 344746 324997                             | 9.3% 8.5%   | 9.3% 8.5%       | 1799229 1725800  | 48.7% 45.2%     |   |                 |
| 40-45                   | 333839 312522                             | 9.0% 8.2%   | 9.0% 8.2%       | 2133068 2038322  | 57.8% 53.4%     |   |                 |
| 45-50                   | 297218 285965                             | 8.1% 7.5%   | 8.1% 7.5%       | 2430286 2324287  | 65.8% 60.9%     |   |                 |
| 50-55                   | 260913 263357                             | 7.1% 6.9%   | 7.1% 6.9%       | 2691199 2587644  | 72.9% 67.8%     |   |                 |
| 55-60                   | 226590 238244                             | 6.1% 6.2%   | 6.1% 6.2%       | 2917789 2825888  | 79.0% 74.1%     |   |                 |
| 60-65                   | 193585 209227                             | 5.2% 5.5%   | 5.2% 5.5%       | 3111374 3035115  | 84.3% 79.6%     |   |                 |
| 65-70                   | 179496 201102                             | 4.9% 5.3%   | 4.9% 5.3%       | 3290870 3236217  | 89.2% 84.8%     |   |                 |
| 70-75                   | 141809 165957                             | 3.8% 4.4%   | 3.8% 4.4%       | 3432679 3402174  | 93.0% 89.2%     |   |                 |
| 75-80                   | 102973 134753                             | 2.8% 3.5%   | 2.8% 3.5%       | 3535652 3536927  | 95.8% 92.7%     |   |                 |
| 80-85                   | 88489 134824                              | 2.4% 3.5%   | 2.4% 3.5%       | 3624141 3671751  | 98.2% 96.3%     |   |                 |
| 85-90                   | 47485 90284                               | 1.3% 2.4%   | 1.3% 2.4%       | 3671626 3762035  | 99.5% 98.6%     |   |                 |
| 90-95                   | 16971 42444                               | 0.5% 1.1%   | 0.5% 1.1%       | 3688597 3804479  | 99.9% 99.7%     |   |                 |
| 95-100                  | 2769 10145                                | 0.1% 0.3%   | 0.1% 0.3%       | 3691366 3814624  | 100.0% 100.0%   |   |                 |
| TOTAL                   | 3691366 3814624                           |   |                 |  |                 |   |                 |
|                         | HOMBRES MUJERES                           |   |                 |  |                 |   |                 |
|                         | $N$                                       |   |                 |  |                 |   |                 |
|                         | El total de hombres y mujeres             |   |                 |  |                 |   |                 |

HISTOGRAMAS DE FRECUENCIA ABSOLUTA

DIAGRAMA DE FRECUENCIAS



HISTOGRAMAS DE PORCENTAJES



Porcentajes ( $P_i$ ): la frecuencia multiplicada por 100

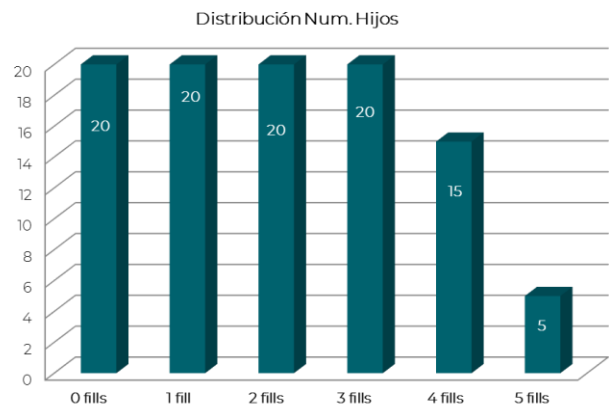
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Tabla de frecuencias y diagrama de barras (discretas)

Tabla de Frecuencias

| Num. Hijos | $n_i$ | $f_i$ | $p_i$ | $N_i$ | $F_i$ | $P_i$ |
|------------|-------|-------|-------|-------|-------|-------|
| 0 Hijos    | 4     | 0,2   | 20    | 4     | 0,2   | 20    |
| 1 Hijo     | 4     | 0,2   | 20    | 8     | 0,4   | 40    |
| 2 Hijos    | 4     | 0,2   | 20    | 12    | 0,6   | 60    |
| 3 Hijos    | 4     | 0,2   | 20    | 16    | 0,8   | 80    |
| 4 Hijos    | 3     | 0,15  | 15    | 19    | 0,95  | 95    |
| 5 Hijos    | 1     | 0,05  | 5     | 20    | 1     | 100   |
| Total      | 20    | 1     | 100   |       |       |       |

Diagrama de Barras



- El diagrama de barras se utiliza para pintar las frecuencias de variables discretas con poca resolución o variables cualitativas

Centralidad y dispersión

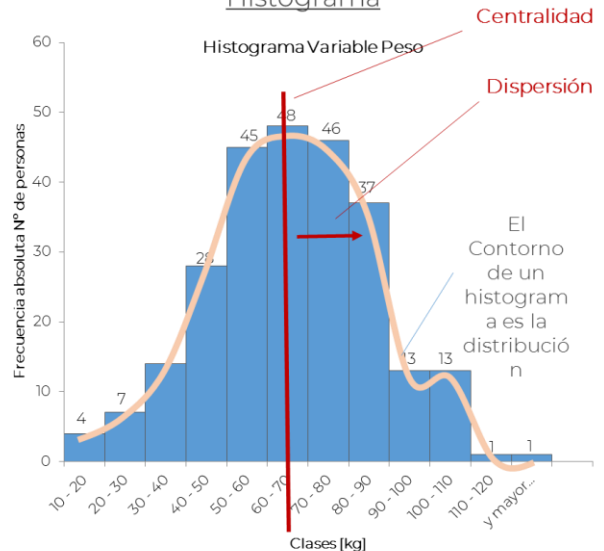
Tabla de Datos

| Sexo   | Peso [kg] |
|--------|-----------|
| HOMBRE | 72.72     |
| HOMBRE | 69.12     |
| HOMBRE | 65.26     |
| MUJER  | 49.35     |
| MUJER  | 61.84     |
| HOMBRE | 64.82     |
| MUJER  | 72.07     |
| HOMBRE | 71.48     |
| HOMBRE | 82.92     |
| MUJER  | 51.07     |
| HOMBRE | 90.47     |
| HOMBRE | 82.24     |
| MUJER  | 47.70     |
| MUJER  | 64.32     |
| MUJER  | 38.07     |
| MUJER  | 59.48     |
| ...    | ...       |

Tabla de Frecuencias

| Clases [kg] | Frecuencia |
|-------------|------------|
| 10 - 20     | 4          |
| 20 - 30     | 7          |
| 30 - 40     | 14         |
| 40 - 50     | 28         |
| 50 - 60     | 45         |
| 60 - 70     | 48         |
| 70 - 80     | 46         |
| 80 - 90     | 37         |
| 90 - 100    | 13         |
| 100 - 110   | 13         |
| 110 - 120   | 1          |
| y mayor...  | 1          |

Histograma



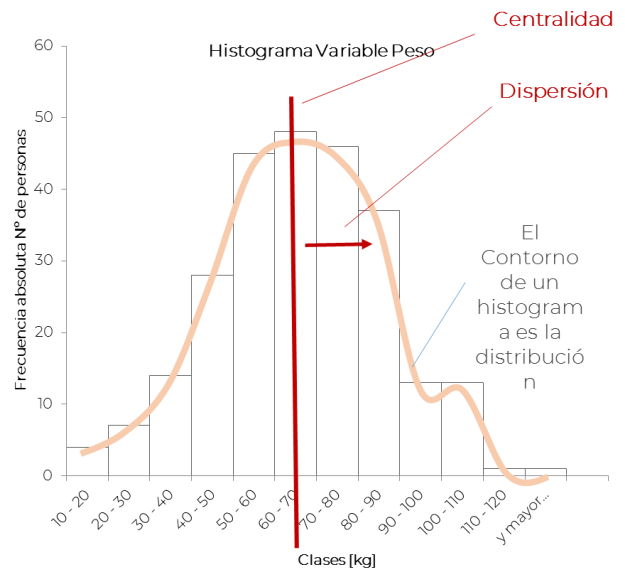
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

- La Centralidad es el valor dónde de se concentra los individuos de la muestra (valor central)

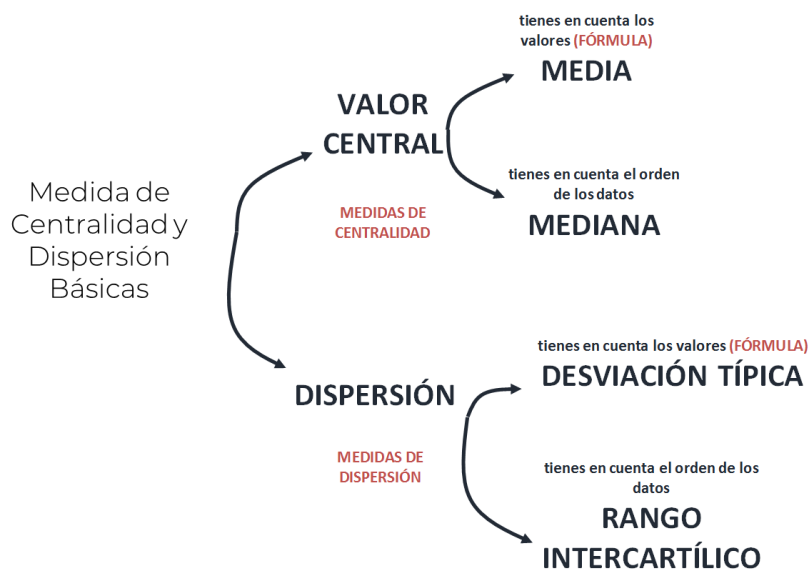
*Ejemplo: media. Más o menos entre 60 y 70 kg*

- La Dispersión es el valor que indica cuánto de lejos están los individuos del centro de la distribución

*Ejemplo: desviación estándar*



## Medidas de centralidad y dispersión



- Las medidas de centralidad y dispersión que tienen en cuenta los valores (fórmulas matemáticas)

- Media o promedio (centralidad)
- Desviación Estándar (dispersión)

- Las medidas de centralidad y dispersión que tienen en cuenta el orden (la posición)

- Mediana (centralidad)
- Rango Intercuartílico (dispersión)

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Media o promedio

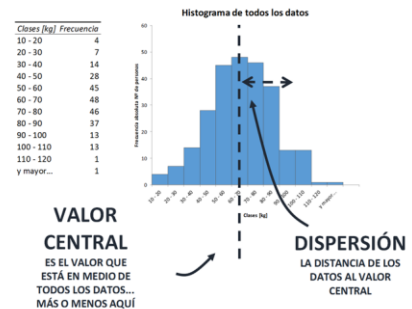
| Sexo   | Peso [kg] | x               |
|--------|-----------|-----------------|
| HOMBRE | 72.72     | x <sub>1</sub>  |
| HOMBRE | 69.12     | x <sub>2</sub>  |
| HOMBRE | 65.26     | x <sub>3</sub>  |
| MUJER  | 49.35     | x <sub>4</sub>  |
| MUJER  | 61.84     | x <sub>5</sub>  |
| HOMBRE | 64.82     | x <sub>6</sub>  |
| MUJER  | 72.07     | x <sub>7</sub>  |
| HOMBRE | 71.48     | x <sub>8</sub>  |
| HOMBRE | 82.92     | x <sub>9</sub>  |
| MUJER  | 51.07     | x <sub>10</sub> |
| HOMBRE | 90.47     | x <sub>11</sub> |
| HOMBRE | 82.24     | x <sub>12</sub> |
| MUJER  | 47.70     | x <sub>13</sub> |
| MUJER  | 64.32     | x <sub>14</sub> |
| ...    |           | x <sub>i</sub>  |

ES LA SUMA DE TODOS  
LOS VALORES DIVIDIDOS  
POR LA CANTIDAD DE  
DATOS QUE TENGAS

$$\frac{x_1 + x_2 + x_3 + \dots + x_N}{N}$$

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

SE PUEDE EXPRESAR  
CON UN SUMATORIO



Media [kg] 66.21

## Desviación Estándar y Varianza

| Sexo   | Peso [kg] | x               | Peso [kg <sup>2</sup> ] | (x-media) <sup>2</sup>                 |
|--------|-----------|-----------------|-------------------------|--|
| HOMBRE | 72.72     | x <sub>1</sub>  | 42.43                   | (x <sub>1</sub> - Media) <sup>2</sup>  |
| HOMBRE | 69.12     | x <sub>2</sub>  | 8.49                    | (x <sub>2</sub> - Media) <sup>2</sup>  |
| HOMBRE | 65.26     | x <sub>3</sub>  | 0.90                    | (x <sub>3</sub> - Media) <sup>2</sup>  |
| MUJER  | 49.35     | x <sub>4</sub>  | 284.14                  | (x <sub>4</sub> - Media) <sup>2</sup>  |
| MUJER  | 61.84     | x <sub>5</sub>  | 19.07                   | (x <sub>5</sub> - Media) <sup>2</sup>  |
| HOMBRE | 64.82     | x <sub>6</sub>  | 1.92                    | (x <sub>6</sub> - Media) <sup>2</sup>  |
| MUJER  | 72.07     | x <sub>7</sub>  | 34.38                   | (x <sub>7</sub> - Media) <sup>2</sup>  |
| HOMBRE | 71.48     | x <sub>8</sub>  | 27.81                   | (x <sub>8</sub> - Media) <sup>2</sup>  |
| HOMBRE | 82.92     | x <sub>9</sub>  | 279.35                  | (x <sub>9</sub> - Media) <sup>2</sup>  |
| MUJER  | 51.07     | x <sub>10</sub> | 229.11                  | (x <sub>10</sub> - Media) <sup>2</sup> |
| HOMBRE | 90.47     | x <sub>11</sub> | 588.72                  | (x <sub>11</sub> - Media) <sup>2</sup> |
| HOMBRE | 82.24     | x <sub>12</sub> | 257.08                  | (x <sub>12</sub> - Media) <sup>2</sup> |
| MUJER  | 47.70     | x <sub>13</sub> | 342.49                  | (x <sub>13</sub> - Media) <sup>2</sup> |
| MUJER  | 64.32     | x <sub>14</sub> | 3.56                    | (x <sub>14</sub> - Media) <sup>2</sup> |
| ...    |           | x <sub>i</sub>  |                         | (x <sub>i</sub> - Media) <sup>2</sup>  |

ES LA MEDIA SUMAS LAS  
DISTANCIAS AL  
CUADRADO  
PONDERADAS AL  
NÚMERO DE DATOS QUE  
TENGAS

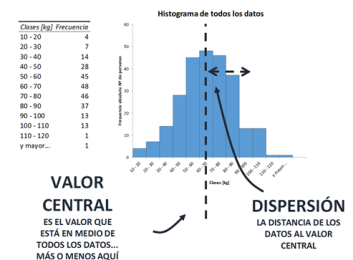
$$\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

$$S_x^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

**VARIANZA**  
SE PUEDE EXPRESAR  
CON UN SUMATORIO

$$S_x = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

**DESVIACIÓN  
TÍPICA O  
ESTÁNDAR**  
PARA TENER UNIDADES  
REALES MEJOR HACER  
LAS RAIZ CUADRADA

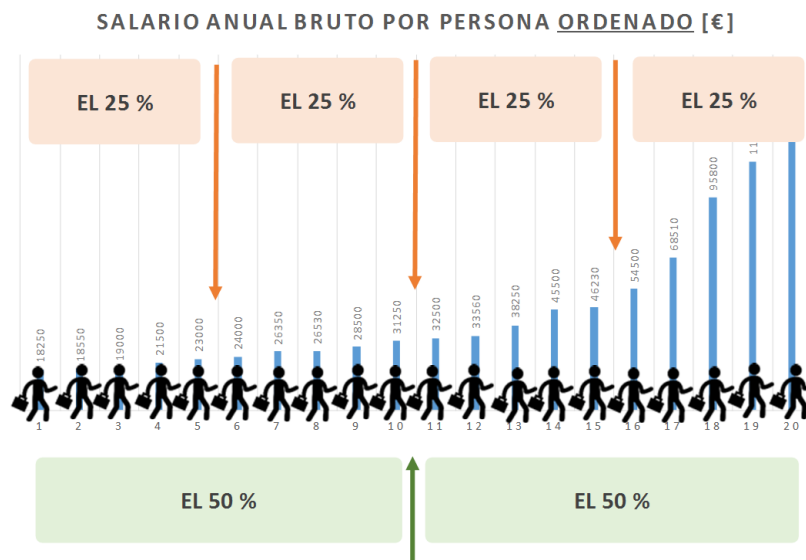


Media [kg] 66.21

398.27 kg<sup>2</sup>

20.00 kg

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Medidas de posición (cuartiles, percentiles)

Los cuartiles son las 3 fronteras cuando se divide la variable ordenada en 4 grupos igual en número

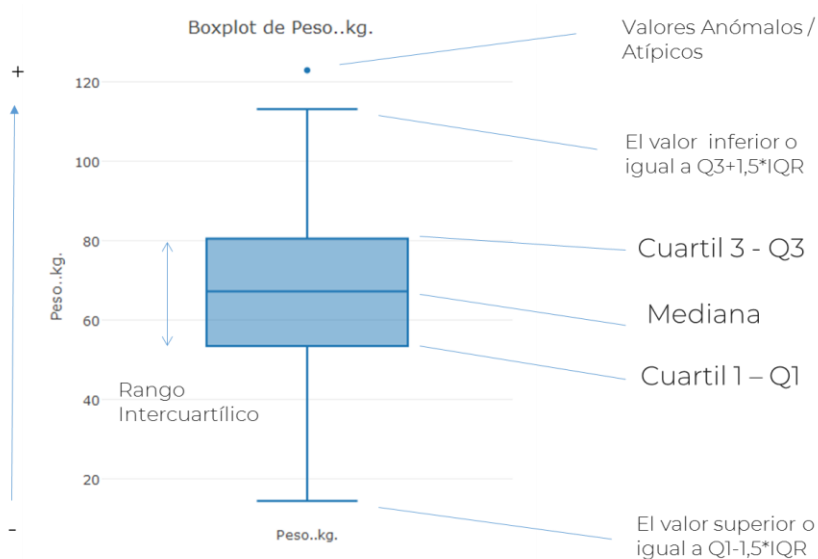
EL 25 % DE 20 = 5  
-> 4 GRUPOS DE 5 PERSONAS

La mediana es la frontera que divide en dos grupos de igual número

EL 50 % DE 20 = 10  
-> 2 GRUPOS DE 10 PERSONAS

Ordenamos los datos de menor a mayor

- Partimos los datos en el 50% → mediana
- Partimos los datos en partes del 25% → cuartiles
- Partimos los datos en partes del 10% → deciles
- Partimos los datos en partes del 1% → percentiles

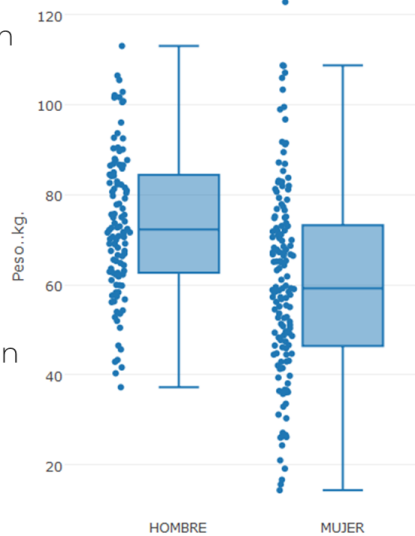
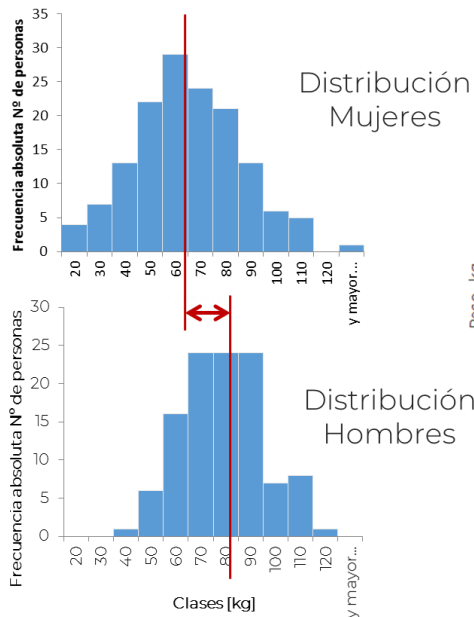
Boxplot

- El boxplot o diagrama de cajas es un gráfico para describir la distribución en función de los rangos
  - Mediana
  - Cuartiles 1 y 3
  - Valores Máximos y atípicos (extremos)

- Se utiliza para encontrar valores extremos de la muestra

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Describiendo una variable cuantitativa



- Centralidad y dispersión

## Mujeres

- Media y Desv. Estándar:  $60.63 \pm 21.14$
- Mediana y IQR:  $59.20 \pm 26.57$

## Hombres

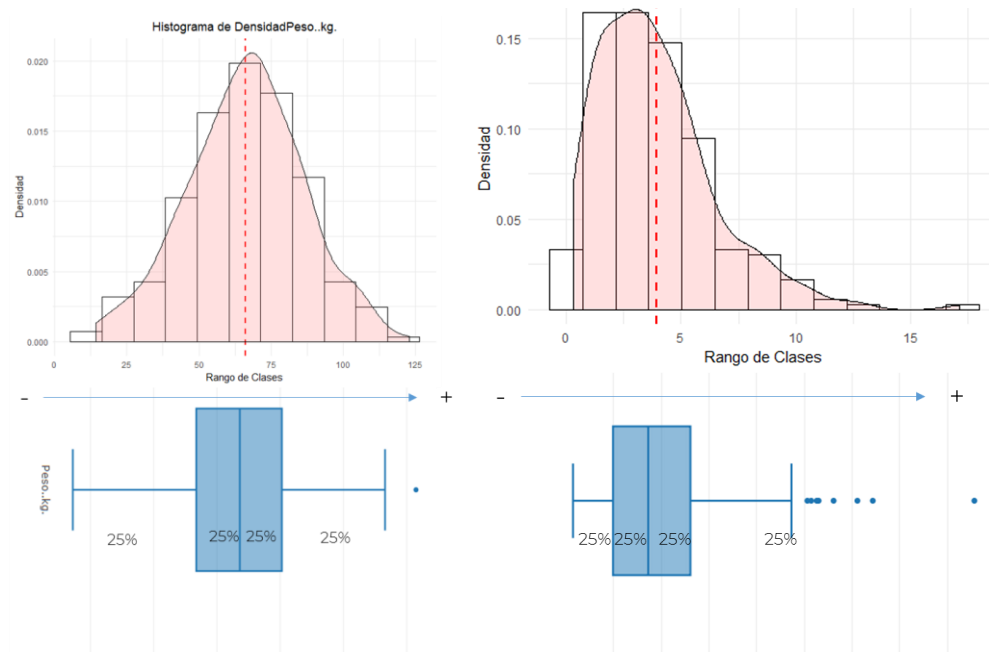
- Media y Desv. Estándar:  $73.42 \pm 15.79$
- Mediana y IQR:  $72.27 \pm 21.69$

|               | Mujeres | Hombres |
|---------------|---------|---------|
| Media         | 60.63   | 73.42   |
| Mediana       | 59.20   | 72.27   |
| Desv Estándar | 21.14   | 15.79   |
| IQR           | 26.57   | 21.69   |

- Histograma y boxplot → distribución
- Desv. Estándar y media → centralidad y dispersión

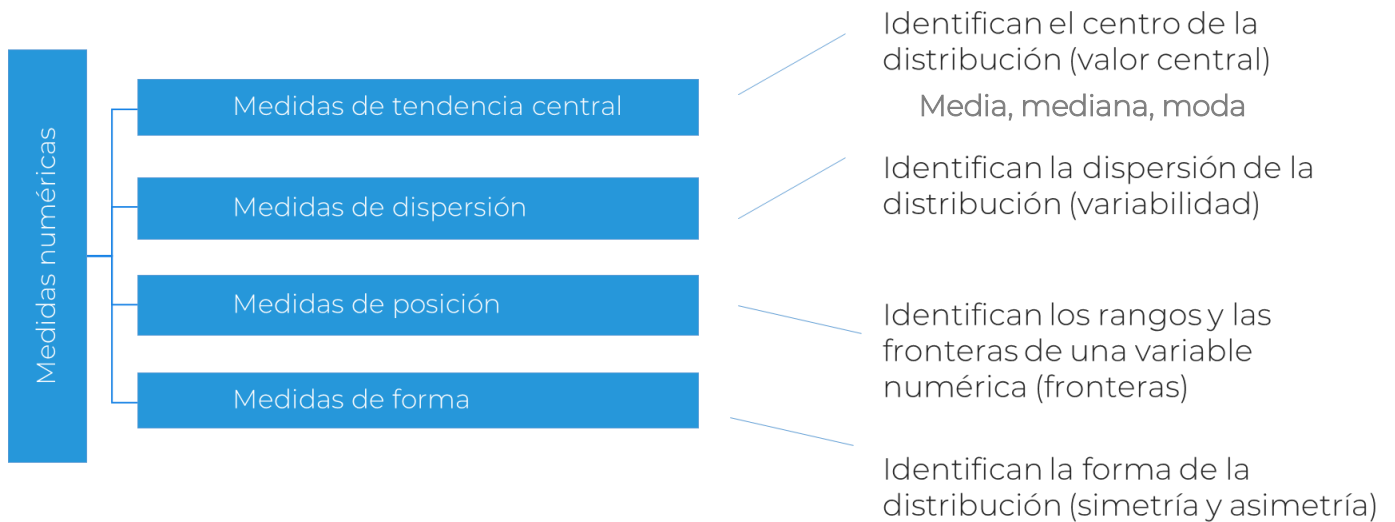
○

- Mediana y IQR → centralidad y dispersión



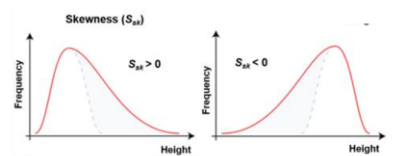
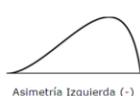
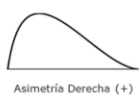
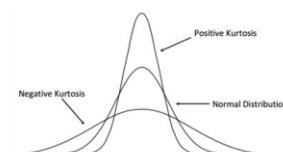
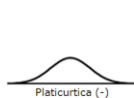
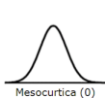
- La simetría de la distribución se puede observar con el histograma o el boxplot

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Características numéricas

Rango, IQR, Desv Estándar (varianza), Error estándar

- Centralidad
  - Media: promedio
  - Mediana: el valor frontera el 50%
  - Moda: el que más se repite
- Dispersión
  - Desviación estándar
  - Rango intercuartílico (IQR)
  - Varianza
  - Error Estándar
  - Coeficiente de variación (CV)
- Posición
  - Cuartiles
  - Percentiles
- Forma
  - Kurtosis
  - Skewness (asimetría)

**Asimetría:** El Coeficiente de Asimetría de Pearson.**Apuntamiento o Curtosis:** Se mide con el coeficiente de curtosis.

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Correlación (relación entre mediciones)

La correlación es la asociación entre mediciones. (generalmente dos variables cuantitativas)

- El coeficiente de correlación mide el grado de asociación
  - Entre -1 y 1
  - El coeficiente 0 significa que no hay asociación

1.0 y -0.8 → la correlación es muy alta y es inversa

-0.8 y -0.6 → la correlación es alta y es inversa

-0.6 y -0.4 → la correlación es normal y es inversa

-0.4 y -0.2 → la correlación es baja y es inversa

-0.2 y 0.2 → la correlación es muy baja o nula. No hay correlación

0.2 y 0.4 → la correlación es baja y es directa

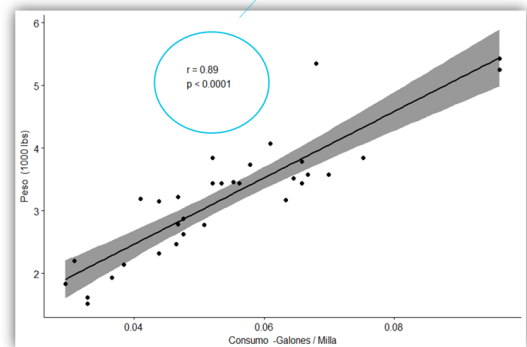
0.4 y 0.6 → la correlación es normal y es directa

0.6 y 0.8 → la correlación es alta y es directa

0.8 y 1.0 → la correlación es muy alta y es directa

| consumo    | peso  |
|------------|-------|
| 0.04761905 | 2.620 |
| 0.04761905 | 2.875 |
| 0.04385965 | 3.320 |
| 0.04672897 | 3.215 |
| 0.05347594 | 3.440 |
| 0.05524862 | 3.460 |
| 0.06993007 | 3.570 |
| 0.04098361 | 3.190 |
| 0.04385965 | 3.150 |
| 0.05208333 | 3.440 |
| 0.05617978 | 3.440 |
| 0.06097561 | 4.070 |
| 0.05780347 | 3.730 |
| 0.06578947 | 3.780 |
| 0.09615385 | 5.250 |
| 0.09615385 | 5.424 |
| 0.06802721 | 5.345 |
| 0.03086420 | 2.200 |
| 0.03280474 | 1.615 |

Coeficiente de Correlación = 0.89



- El Coeficiente de Pearson entre dos variables

Matriz de Correlación

|                    | X-Partidos Ganados: Y-Goles a favor |        |
|--------------------|-------------------------------------|--------|
| X-Partidos Ganados | 1                                   | 0.8429 |
| Y-Goles a favor    | 0.8429                              | 1      |

$$\begin{vmatrix} 1 & r_{xy} \\ r_{yx} & 1 \end{vmatrix}$$

LA CORRELACIÓN INDICA EL GRADO DE RELACIÓN LINEAL

$$r_{xy} = \frac{S_{xy}}{S_x S_y}$$

No tiene dimensiones

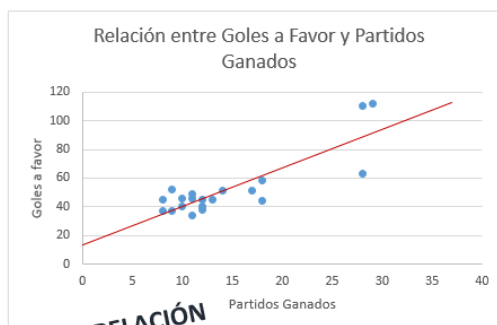
Siempre estará entre -1 y 1

Negativo >> relación lineal decreciente

Positivo >> relación lineal creciente

El valor 1 o -1 >> súper relacionadas linealmente

El valor 0 >> no están nada relacionadas linealmente



LA CORRELACIÓN ES LA RELACIÓN DE DEPENDENCIA LINEAL

ES UNA RATIO ENTRE LA COVARIANZA Y LAS DESVIACIONES TÍPICAS

$$r_{xy} = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^N (x_i - \bar{x})^2 \sum_i^N (y_i - \bar{y})^2}}$$

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## La Matriz de correlación

Se puede calcular la correlación cruzada entre más de dos variables.

Matriz de Correlación

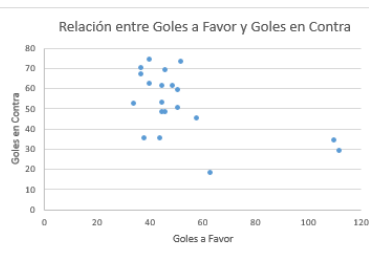
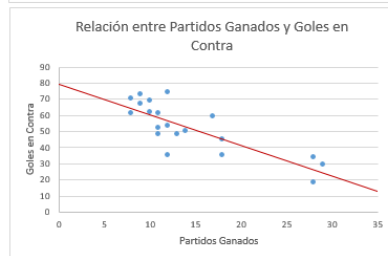
|                    | X-Partidos Ganados | Y-Goles a Favor | Z-Goles en Contra |
|--------------------|--------------------|-----------------|-------------------|
| X-Partidos Ganados | 1.000              | 0.843           | -0.808            |
| Y-Goles a Favor    | 0.843              | 1.000           | -0.545            |
| Z-Goles en Contra  | -0.808             | -0.545          | 1.000             |

|          | 1 | $r_{xy}$ | $r_{xz}$ |
|----------|---|----------|----------|
| $r_{yx}$ |   | 1        | $r_{yz}$ |
| $r_{zx}$ |   | $r_{zy}$ | 1        |

Correlación Goles en Contra vs Partidos Ganados

HeatMap Correlación

|                    | X-Partidos Ganados | Y-Goles a Favor | Z-Goles en Contra |
|--------------------|--------------------|-----------------|-------------------|
| X-Partidos Ganados | 1                  | 0.842853495     | -0.808188528      |
| Y-Goles a Favor    | 0.842853495        | 1               | -0.545116187      |
| Z-Goles en Contra  | -0.808188528       | -0.545116187    | 1                 |



## Correlación

- No tiene dimensiones
- Siempre estará entre -1 y 1
- Negativo >> relación lineal decreciente
- Positivo >> relación lineal creciente
- El valor 1 o -1 >> super relacionadas linealmente
- El valor 0 >> no están nada relacionadas linealmente

|          | 1 | $r_{xy}$ | $r_{xz}$ |
|----------|---|----------|----------|
| $r_{yx}$ |   | 1        | $r_{yz}$ |
| $r_{zx}$ |   | $r_{zy}$ | 1        |

Correlación Goles en Contra vs Partidos Ganados

## Correlación

- No tiene dimensiones
- Siempre estará entre -1 y 1
- Negativo >> relación lineal decreciente
- Positivo >> relación lineal creciente
- El valor 1 o -1 >> super relacionadas linealmente
- El valor 0 >> no están nada relacionadas linealmente

## Varianza

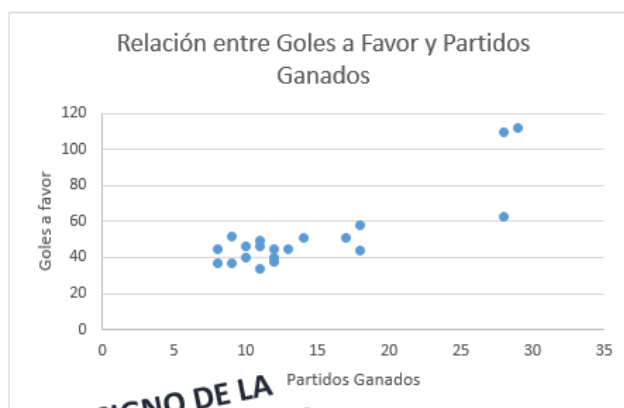
Matriz de covarianza

|                    | X-Partidos Ganados | Y-Goles a favor |
|--------------------|--------------------|-----------------|
| X-Partidos Ganados | 42.44              | 114.34          |
| Y-Goles a favor    | 114.34             | 433.6275        |

Varianza Partidos Ganados

|          | $S_x^2$ | $S_{xy}$ |
|----------|---------|----------|
| $S_{yx}$ |         | $S_y^2$  |

Varianza Goles a Favor



## Covarianza

## Partidos Ganados vs Goles a Favor

Covarianza Positiva >> Relación directa (cuando una crece la otra también)

Covarianza Negativa >> Relación Inversa (cuando una crece la otra decrece)

El valor de la Covarianza depende de las unidades  
Sólo fíjate en el signo

EL SIGNO DE LA COVARIANZA INDICA SI LA RELACIÓN ES DIRECTA O INVERSA

EL VALOR NO ES IMPORTANTE SINO EL SIGNO

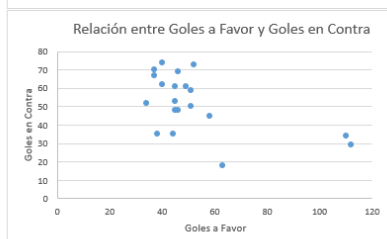
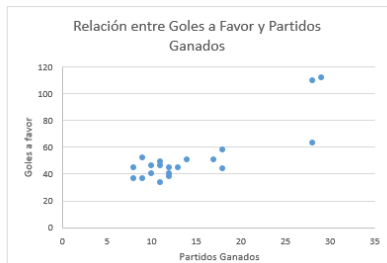
$$s_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Matriz de covarianza

Matriz de Covarianza

|                    | X-Partidos Ganados | Y-Goles a Favor | Z-Goles en Contra |
|--------------------|--------------------|-----------------|-------------------|
| X-Partidos Ganados | 42.44              | 114.34          | -81.11            |
| Y-Goles a Favor    | 114.34             | 433.6275        | -174.8725         |
| Z-Goles en Contra  | -81.11             | -174.8725       | 237.3275          |



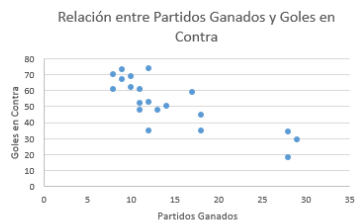
Varianza Partidos Ganados Varianza Goles a Favor

$$\begin{bmatrix} S_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{bmatrix}$$

Covarianza Goles en Contra vs Partidos Ganados Varianza Goles a Favor

## Covarianza

Partidos Ganados vs Goles a Favor

Covarianza Positiva >> Relación directa  
(cuando una crece la otra también)Covarianza Negativa >> Relación Inversa  
(cuando una crece la otra decrece)El valor de la Covarianza depende de las unidades  
Sólo fíjate en el signo

Matriz de Covarianza

|                    | X-Partidos Ganados | Y-Goles a Favor | Z-Goles en Contra |
|--------------------|--------------------|-----------------|-------------------|
| X-Partidos Ganados | 42.44              | 114.34          | -81.11            |
| Y-Goles a Favor    | 114.34             | 433.6275        | -174.8725         |
| Z-Goles en Contra  | -81.11             | -174.8725       | 237.3275          |

Varianza Partidos Ganados Varianza Goles a Favor

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 \end{bmatrix}$$

Covarianza Goles en Contra vs Partidos Ganados Varianza Goles a Favor

## Covarianza

Partidos Ganados vs Goles a Favor

Covarianza Positiva >> Relación directa  
(cuando una crece la otra también)Covarianza Negativa >> Relación Inversa  
(cuando una crece la otra decrece)El valor de la Covarianza depende de las unidades  
Sólo fíjate en el signo

## ESTADÍSTICAS DESCRIPTIVA PARA VARIABLES CUALITATIVAS

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Distribución de una variable cualitativa (barras y sectores)

- Diagrama de Barras y sectores para pintar una distribución en forma de proporciones de una variable cualitativa

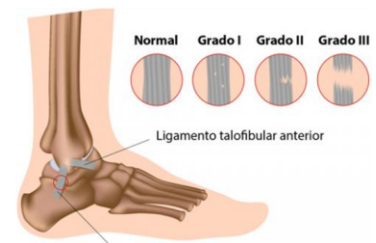
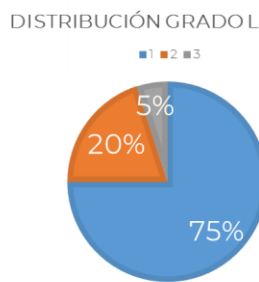
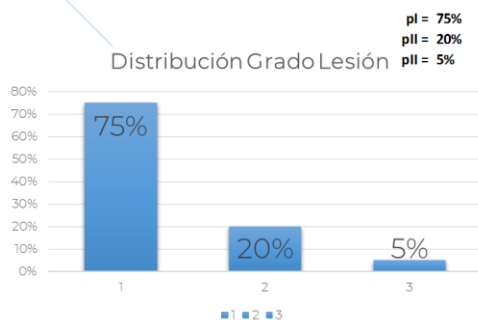
- Variable: Grado Lesión (Ordinal)

Diagrama de Barras  
Distribución De la  
Variable Categórica

| Grado Lesión | f  | fr   | P%   |
|--------------|----|------|------|
| I            | 45 | 0.75 | 75%  |
| II           | 12 | 0.2  | 20%  |
| III          | 3  | 0.05 | 5%   |
|              | 60 | 1    | 100% |

Porcentaje o proporción  
 $P\% = 100 \cdot fr$

Diagrama de Sectores  
Distribución De la  
Variable Categórica

Tablas de contingencias (relativas, absolutas, condicionadas)

- Tabla de contingencias se utiliza para dos variables cualitativas a la vez
- En filas y columnas
- Contamos las frecuencias por parejas de grupos

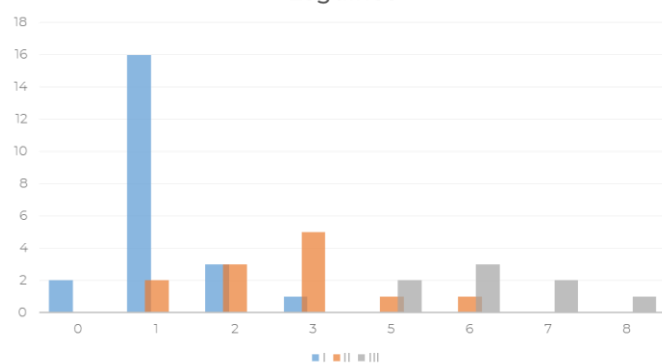
Frecuencias Absolutas:

Dos variables cualitativas:

- Grado Esguince (Grado I, II y III)
- Semanas de Baja (0,1,2,3...8)

| Grado Esguince | Número semanas de Baja |    |   |   |   |   |   |   | Total general |
|----------------|------------------------|----|---|---|---|---|---|---|---------------|
|                | 0                      | 1  | 2 | 3 | 5 | 6 | 7 | 8 |               |
| I              | 2                      | 16 | 3 | 1 |   |   |   |   | 22            |
| II             |                        | 2  | 3 | 5 | 1 | 1 |   |   | 12            |
| III            |                        |    |   |   | 2 | 3 | 2 | 1 | 8             |
| Total general  | 2                      | 18 | 6 | 6 | 3 | 4 | 2 | 1 | 42            |

Diagrama de Barras Por Grado de  
Esguince



Las tablas de contingencias se pueden representar en:

- Frecuencias Absolutas o Relativas
- Frecuencias Condicionadas a los Grupos

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

- **Frecuencia absoluta:** contamos el número absoluto de observaciones por parejas de grupos
- **Frecuencia relativa** es la frecuencia absoluta por el total de observaciones de la muestra
- **Frecuencia condicionada** se puede calcular
  - Por filas o columnas
  - Es la proporción condicionada al total del grupo elegido (o filas o columna)

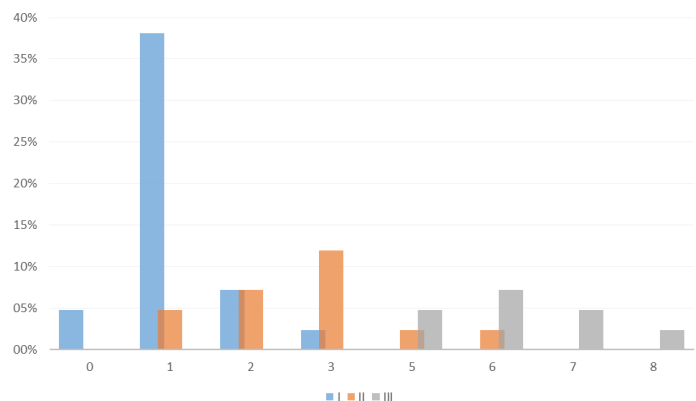
Frecuencias relativas:

Dos variables cualitativas:

- Grado Esguince (Grado I, II y III)
- Semanas de Baja (0,1,2,3...8)

| Grado Esguince | Número semanas de Baja |       |       |       |      |      |      |      | Total general |
|----------------|------------------------|-------|-------|-------|------|------|------|------|---------------|
|                | 0                      | 1     | 2     | 3     | 5    | 6    | 7    | 8    |               |
| I              | 4.8%                   | 38.1% | 7.1%  | 2.4%  | 0.0% | 0.0% | 0.0% | 0.0% | 52.4%         |
| II             | 0.0%                   | 4.8%  | 7.1%  | 11.9% | 2.4% | 2.4% | 0.0% | 0.0% | 28.6%         |
| III            | 0.0%                   | 0.0%  | 0.0%  | 0.0%  | 4.8% | 7.1% | 4.8% | 2.4% | 19.0%         |
| Total general  | 4.8%                   | 42.9% | 14.3% | 14.3% | 7.1% | 9.5% | 4.8% | 2.4% | 100.0%        |

Diagrama de Barras Por Grado de Esguince



Las tablas de contingencias se pueden representar en:

- Frecuencias Absolutas o Relativas
- Frecuencias Condicionadas a los Grupos

$$Frec. Relativa = \frac{Frec. Absoluta}{N}$$

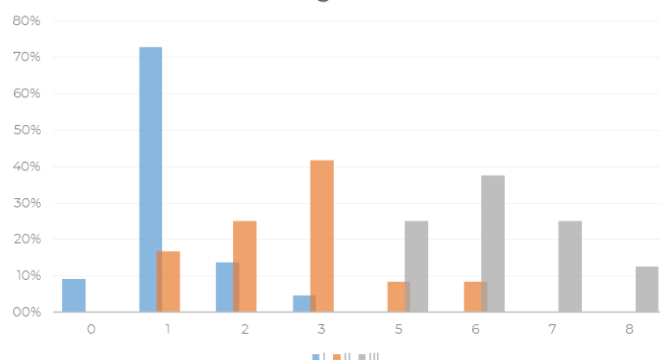
Frecuencia condicionada por filas:

Dos variables cualitativas:

- Grado Esguince (Grado I, II y III)
- Semanas de Baja (0,1,2,3...8)

| Grado Esguince | Número semanas de Baja |       |       |       |       |       |       |       | Total general |
|----------------|------------------------|-------|-------|-------|-------|-------|-------|-------|---------------|
|                | 0                      | 1     | 2     | 3     | 5     | 6     | 7     | 8     |               |
| I              | 9.1%                   | 72.7% | 13.6% | 4.5%  | 0.0%  | 0.0%  | 0.0%  | 0.0%  | 100%          |
| II             | 0.0%                   | 16.7% | 25.0% | 41.7% | 8.3%  | 8.3%  | 0.0%  | 0.0%  | 100%          |
| III            | 0.0%                   | 0.0%  | 0.0%  | 0.0%  | 25.0% | 37.5% | 25.0% | 12.5% | 100%          |

Diagrama de Barras Por Grado de Esguince



Las tablas de contingencias se pueden representar en:

- Frecuencias Absolutas o Relativas
- Frecuencias Condicionadas a los Grupos

$$Frec. Condicionada Grupo I = \frac{Frec. Absoluta}{N Grupo I}$$

## ESTADÍSTICA INFERENCIAL

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

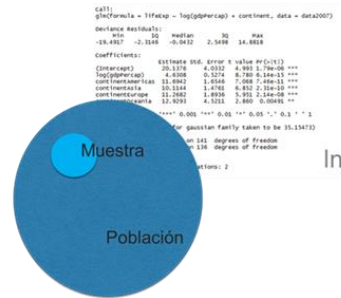
En la estadística inferencial es vital entender:

### Muestra y población

Población



Muestra



Interpretar resultados  
y extraer  
conclusiones

- Población: una muestra infinitivamente grande
- Muestra: un pequeño conjunto de la población (nuestros datos)

### Qué son las distribuciones de variables

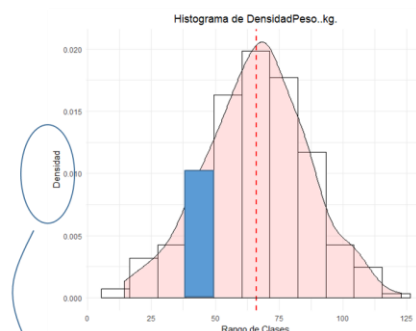
Las distribuciones es la forma de cómo se reparten las variables que hemos medido.

- Cuantitativas continuas – distribución de densidad
- Cualitativas o discretas – diagrama de barras de proporciones

### Variables cuantitativas continuas – histograma de densidad

AreaBarra<sub>i</sub> = Intervalo de Clase \* Altura Densidad

- El histograma de densidad es un histograma que la suma del área de las barras igual a 1
- Es decir, que la suma de las áreas de todas las barras es igual a 1
- La altura del histograma no depende del rango de clases
- Porque el eje vertical está en densidad de frecuencia
- Aunque tengas más clases la altura es la misma.



La densidad o **densidad de frecuencia** es la unidad del eje vertical  
La altura (en densidad de frecuencia) por la amplitud de clase es el área de una barra

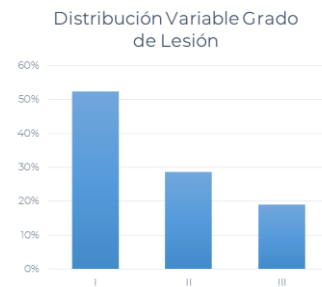
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

### Variables discretas o cualitativas – diagrama de barras de proporciones

- La distribución de una variable cualitativa se puede expresar con un diagrama de barras
- La distribución indica como se reparte los grupos en porcentajes

Ejemplos: grado de lesión

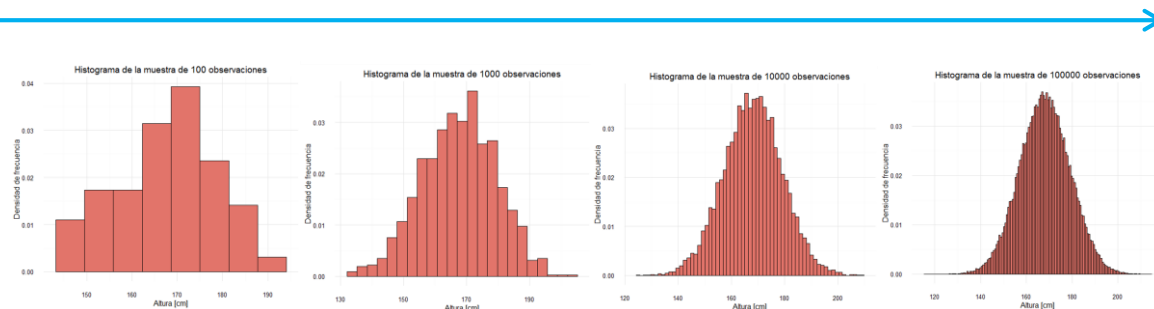
- Alrededor del 50% las lesiones de tobillo son de grado I
- 30% de grado II y 20% de grado III



### Qué es y para qué sirve la distribución de densidad de probabilidad y las probabilidades

La distribución de densidad de probabilidad es un histograma de densidad de una muestra infinitamente grande:

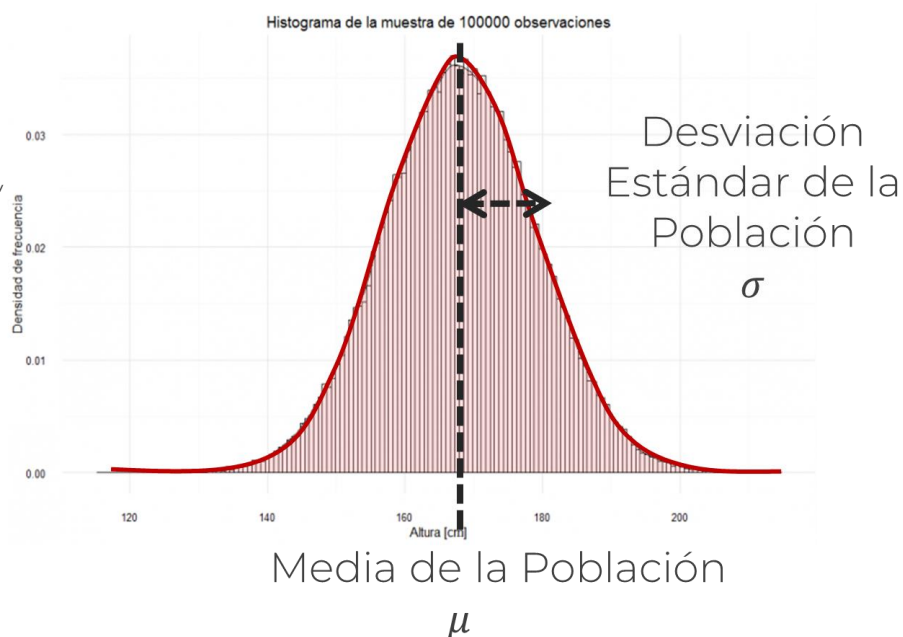
Histograma de densidad de menor a mayor muestra



El contorno del histograma de densidad con una muestra muy muy grande es la distribución de densidad de probabilidad

El área total de esta función es 1

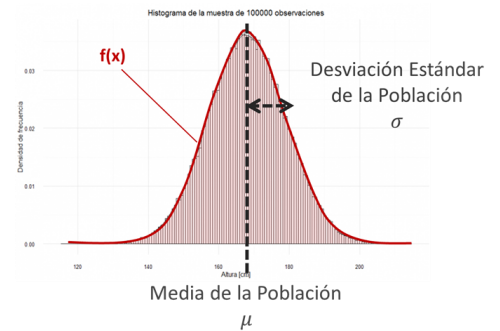
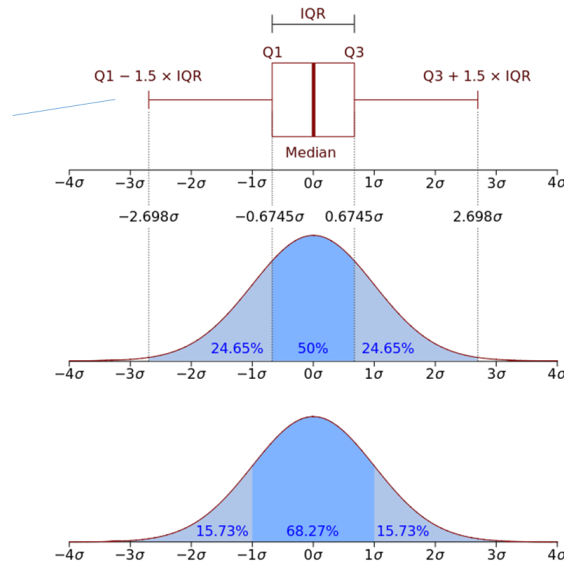
El área es sinónimo de probabilidad



## El poder de la distribución normal

- Media y desviación estándar son sus parámetros

Siempre que te encuentres con un boxplot más o menos simétrico tu variable cuantitativa estará cerca de la distribución normal

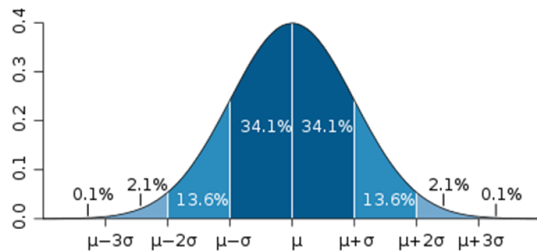


$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Densidad de Frecuencia

Variables Numérica (Altura en cm)

- Las reglas de la distribución normal



Si la media  $\mu = 66,5$  kg y la desviación estándar  $\sigma = 18$  kg  
Entonces:

- El 95% de los pacientes se encuentran entre 31,22 y 101,78 kg  
(cálculo:  $66,5 \pm 1,96 \cdot 18 = [31,22, 101,78] \text{ kg}$ )
- El 99% de los pacientes se encuentran entre 24,56 y 108,44 kg  
(cálculo:  $66,5 \pm 2,33 \cdot 18 = [24,56, 108,44] \text{ kg}$ )

Siempre que la población se comporte como una distribución normal se cumple:

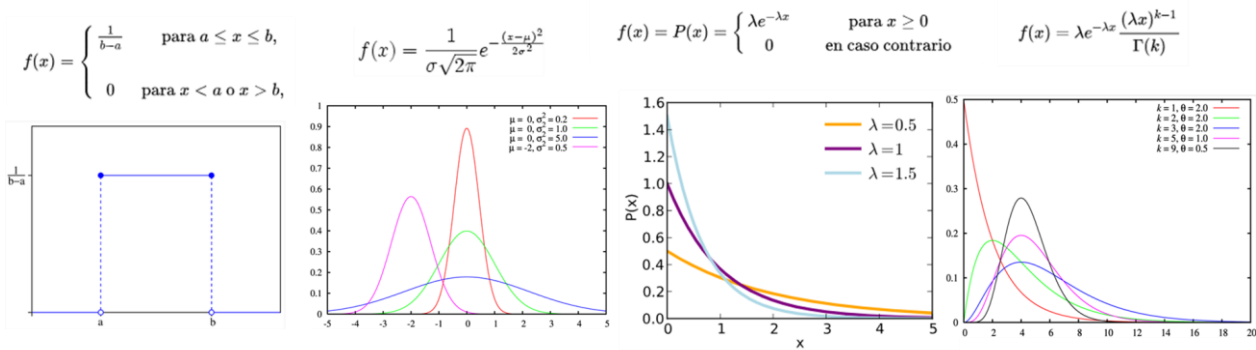
- El 95% de los datos están centrados entre la media  $\pm 1,96$  veces la desviación estándar
- El 99% de los datos están centrados entre la media  $\pm 2,33$  veces la desviación estándar

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

Tipos de distribuciones

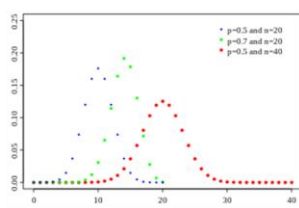
- Continuas → Densidad de probabilidad (variables continuas)

- Uniforme
- Normal
- Exponencial
- Gamma

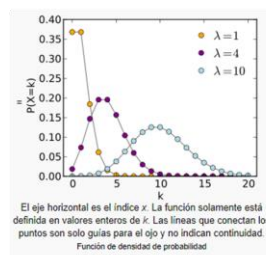


- Discretas → Proporciones (variables cualitativas o de conteo)

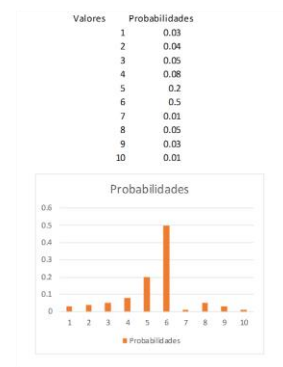
- Binomial
- Poisson
- Personalizada



P: probabilidad de acierto  
N: número de intentos



landa: número de éxitos

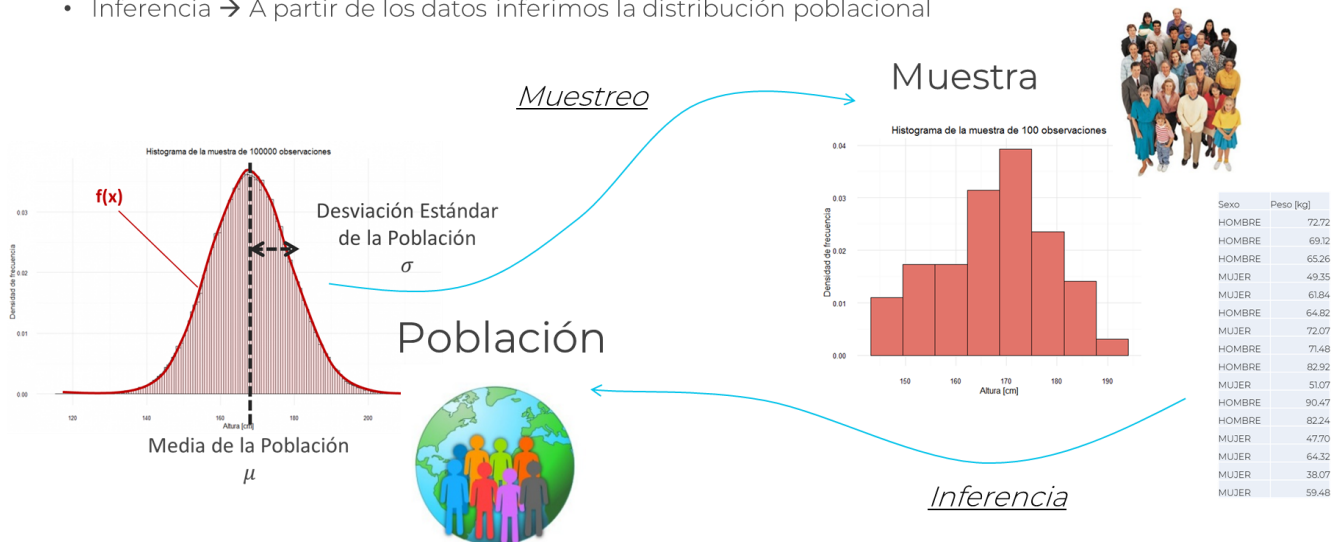


Existen muchas más...

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Muestreo e inferencia

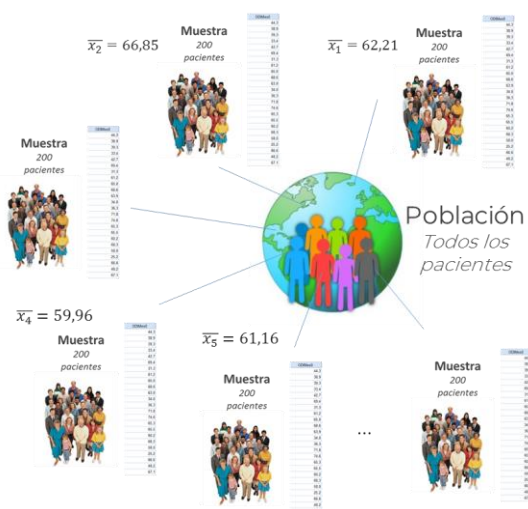
- Muestreo → A partir de la población obtenemos las muestras (tabla de datos)
- Inferencia → A partir de los datos inferimos la distribución poblacional



## Intervalos de confianza

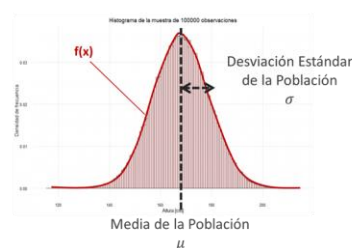
La distribución de medias siempre sigue una distribución normal con:

- La media de las medias = media de la muestra
- La desviación estándar de las medias = error estándar



¿Dónde estará la media de la población?

iii Se ha demostrado que la distribución de las medias de las muestras posibles de una población sigue una distribución Normal!!!



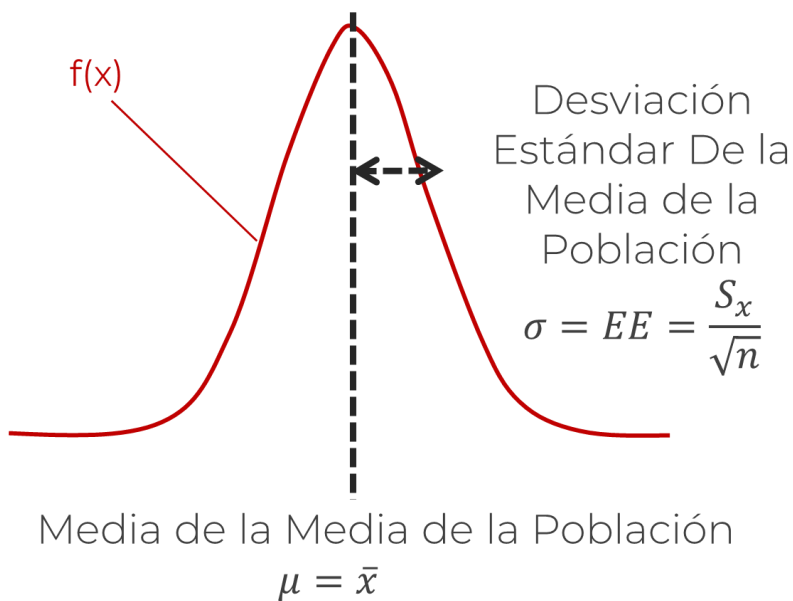
La distribución de la media Muestral

Tiene forma de distribución normal con:

- Media = Media de la Muestra
- Desviación Estándar = Error Estándar de la Media

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

El cálculo del Intervalo de confianza para la media:



¿Dónde estará la media de la población?

La media de la población se encontrará entre los intervalos de confianza

Podemos definir un intervalo de Confianza al 95%

$$\bar{X} \pm 1.96 \cdot EE$$

Podemos definir un intervalo de Confianza al 99%

$$\bar{X} \pm 2.33 \cdot EE$$

## Qué es una hipótesis

Es una pregunta que admite un verdadero o falsa como respuesta. Solo dos opciones

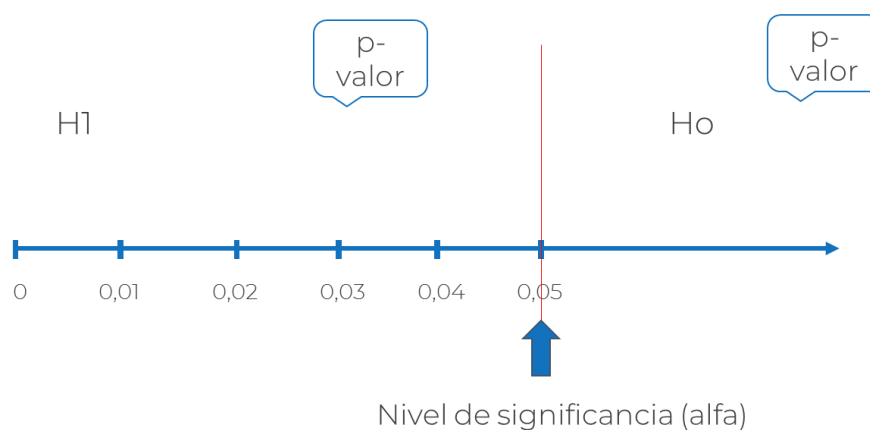
- Hipótesis alternativa (H1)
  - La pregunta que quieres afirmar. Normalmente el caso raro o de diferencias
- Hipótesis nula (H0)
  - La contraria que la H1

## P-valor

La probabilidad de error de afirmar la H1 y en realidad sea falsa

Si el p-valor < 5% → aceptas la H1

Regla de decisiones para la prueba de hipótesis



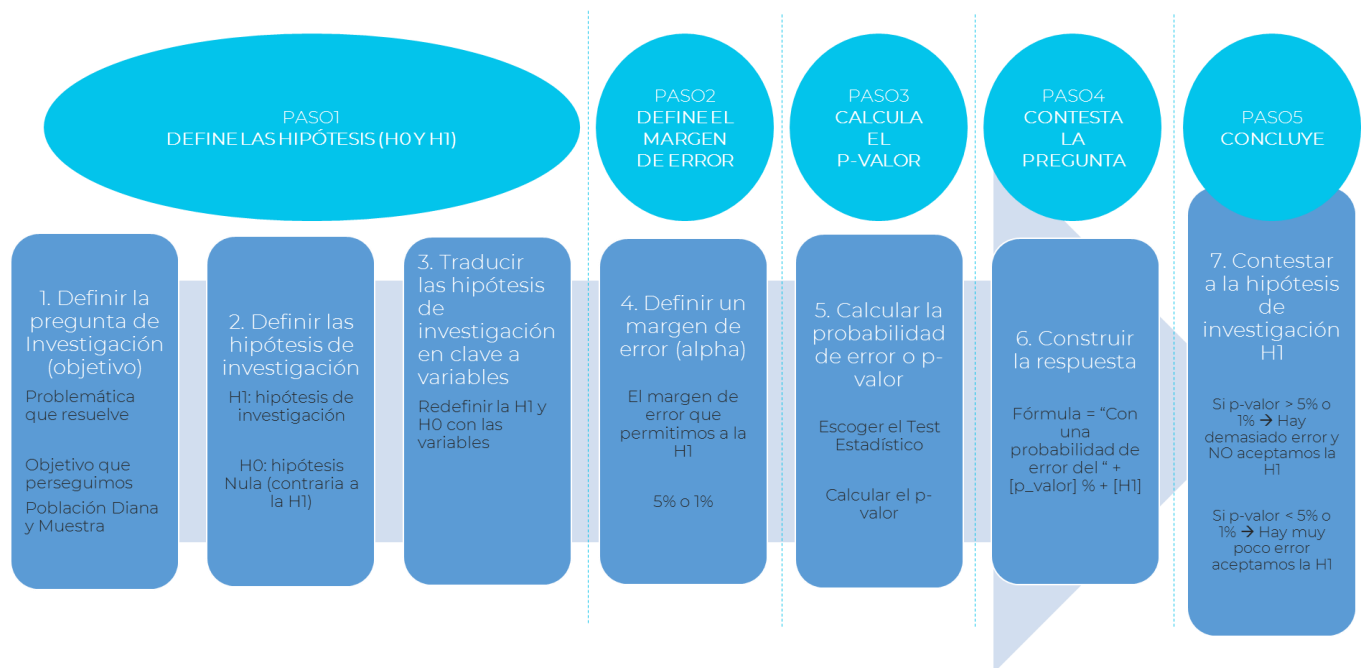
## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Test estadístico

El cálculo del software estadística para calcular el p-valor.

La clave está en saber seleccionar el más adecuado para tu caso.

## El contraste de hipótesis y su metodología



La metodología del contraste es la siguiente:

1. Definir las hipótesis (H1 y H0)
2. Definir el umbral de significación
3. Seleccionar el test estadístico
4. Calcular el p-valor
5. Contestar a la hipótesis
6. Aportar más información con el resultado de contraste de hipótesis y la descripción

## La esencia de los modelos predictivos y sus dos súper poderes

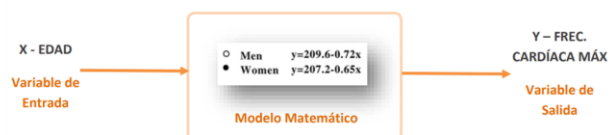
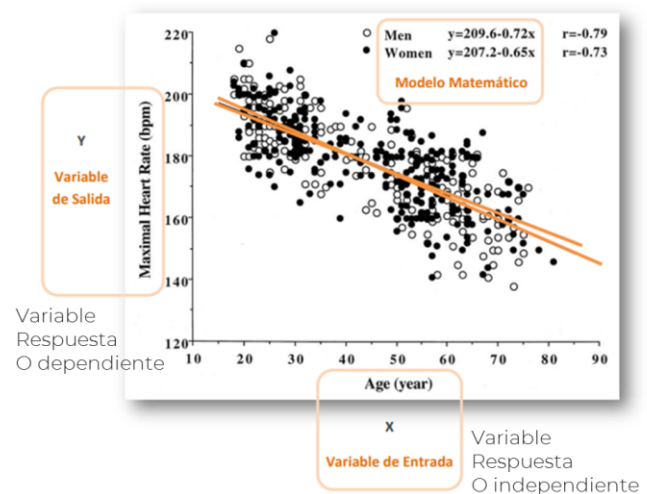
### Las partes de un modelo:

- Función matemática
- Variable/s de entrada
- Variable de salida
- Coeficientes o parámetros

- Un modelo es una función matemática que a partir de unos datos de entrada (variables de entrada) obtienes los datos de la variables de salida (o respuesta)

- Un modelo tiene tres partes bien diferenciadas:

| Variable Salida – Y  | Frecuencia Cardíaca Máxima                       |
|----------------------|--|
| Variable Entrada – X | Edad   |
| Función matemática   | ○ Men $y=209.6-0.72x$<br>● Women $y=207.2-0.65x$ |



### Dos utilidades o dos superpoderes:

- Estimación o predicción
- Relacional (causa-efecto)

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## ¿Para qué sirve?

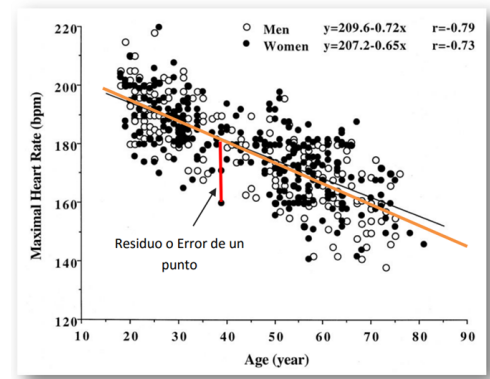
## Utilidad n°1: La Estimación

- Imagina que para uno de los pacientes de 32 años la frecuencia cardíaca máxima es de 145 pulsaciones por minuto
- Para el modelo que hemos calculado la frecuencia cardíaca máxima es la siguiente:

$$\text{Frec. Cardíaca Máxima} = -0,72 \cdot 32 + 209,6 = 186,56$$

- El error o residuo del modelo es el valor observado menos el valor estimado por el modelo

$$\text{observado} - \text{esperado} = y_{\text{obs}} - \hat{y} = 145 - 186,56 = -41,56$$



$$\text{Frec. Cardíaca Máxima} = -0,72 \cdot \text{Edad} + 209,6$$

Variable de Salida

Parámetros del modelo

Variable de Entrada

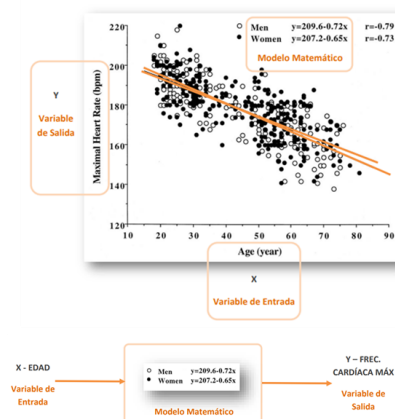
## ¿Para qué sirve?

## Utilidad n°2: Relacionar

- Podemos utilizar el modelo para explicar lo que está pasando con la frecuencia cardíaca máxima y la edad
- Un modelo puede explicar relaciones causa-efecto
- El modelo lineal del ejemplo es la línea naranja del gráfico
- Es una recta con pendiente negativa ya que el valor del coeficiente  $a$  es de  $-0,72$
- Eso significa que cuanto más edad tienen los pacientes menor es la frecuencia cardíaca máxima
- A mayor edad  $\rightarrow$  corazón más débil  $\rightarrow$  frecuencia cardíaca máxima menor
- Además en el estudio se han calculado dos modelos, uno para los datos de hombre y otro para mujeres. También podemos comparar los modelos y ver la influencia de la edad con la frecuencia cardíaca máxima:

$$\text{Frec. Cardíaca Máxima} = -0,72 \cdot \text{Edad} + 209,6 \rightarrow \text{Para Hombres}$$

$$\text{Frec. Cardíaca Máxima} = -0,65 \cdot \text{Edad} + 207,2 \rightarrow \text{Para Mujeres}$$



$$\text{Frec. Cardíaca Máxima} = -0,72 \cdot \text{Edad} + 209,6$$

Variable de Salida

Parámetros del modelo

Variable de Entrada

## EL RESUMEN DEFINITIVO DE LOS CONCEPTOS CLAVE EN ESTADÍSTICA

## Los dos grandes retos de los modelos

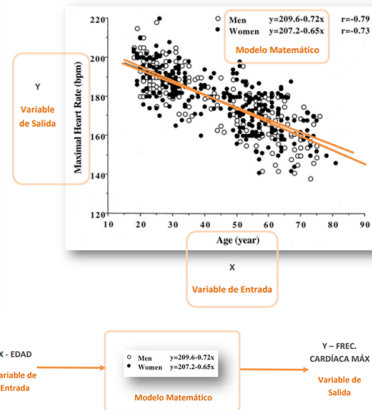
## RETO 1

¿QUÉ MODELO ES EL QUE MEJOR EXPLICA LOS DATOS?

¿Qué función matemática es la más adecuada?

- Tipo de función

Criterio para decidir si el modelo es válido



## RETO 2

¿QUÉ VARIABLES SON LAS MÁS REPRESENTATIVAS?

¿Qué variables introduzco en el modelo?

Criterio para comparar los modelos

Probar varios modelos y escoger el mejor