# Reproducible Research Using RMarkdown

Basic Statistics with R for Biomedical Research

UEB – VHIR

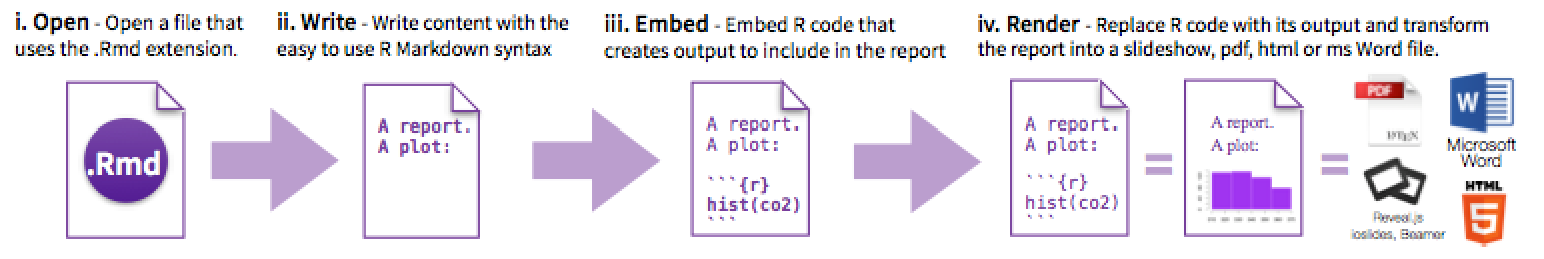
**Miriam Mota-Foix**

Miriam.mota@vhir.org

# Why Rmarkdown?

- Reproducbility  (With one click, literally)
- Organization
- Share-ability (easily share or publish knitted files to HTML/RPubs)
- Annotate-ability (now I’m just making these up… Markdown makes it easy to annotate your code)
- Insert chunks of Python, Bash, SQL code

# Workflow



**i. Open** - Open a file that uses the .Rmd extension.

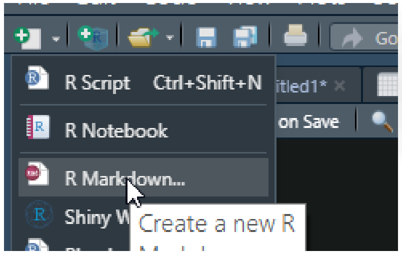
**ii. Write** - Write content with the easy to use R Markdown syntax

**iii. Embed** - Embed R code that creates output to include in the report

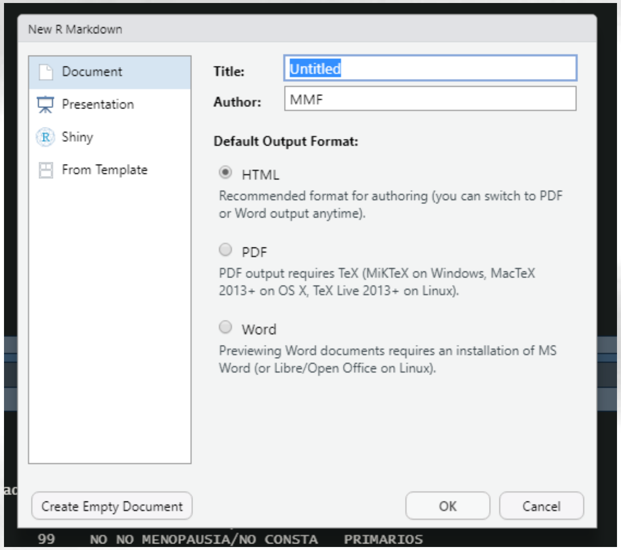
**iv. Render** - Replace R code with its output and transform the report into a slideshow, pdf, html or ms Word file.

.Rmd
A report.
A plot:
```{r}
hist(co2)
```
=
A report.
A plot:
=
PDF
LaTeX
Microsoft Word
Reveal.js
ioslides, Beamer
HTML

## Open an Rmd File



## Choose output



New R Markdown

- Document
- Presentation
- Shiny
- From Template

Title: Untitled

Author: MMF

Default Output Format:

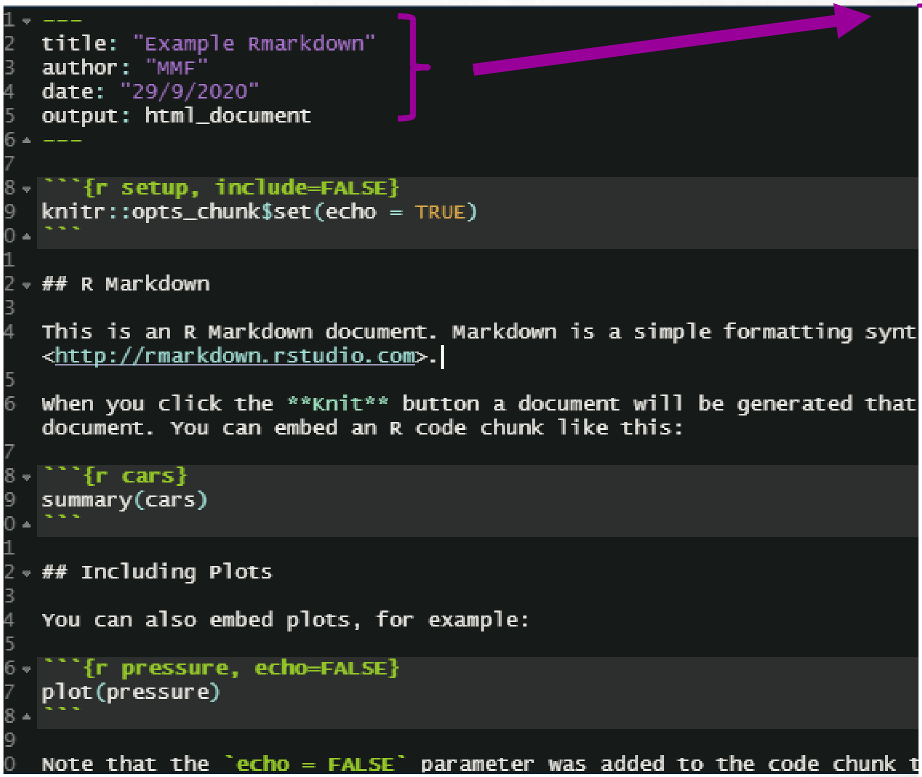
● HTML
Recommended format for authoring (you can switch to PDF or Word output anytime).

○ PDF
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

○ Word
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

Create Empty Document OK Cancel

# Rmarkdown Example



~~~
---
title: "Example Rmarkdown"
author: "MMF"
date: "29/9/2020"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
~~~

## R Markdown

This is an R Markdown document. Markdown is a simple formatting synt
<http://rmarkdown.rstudio.com>.

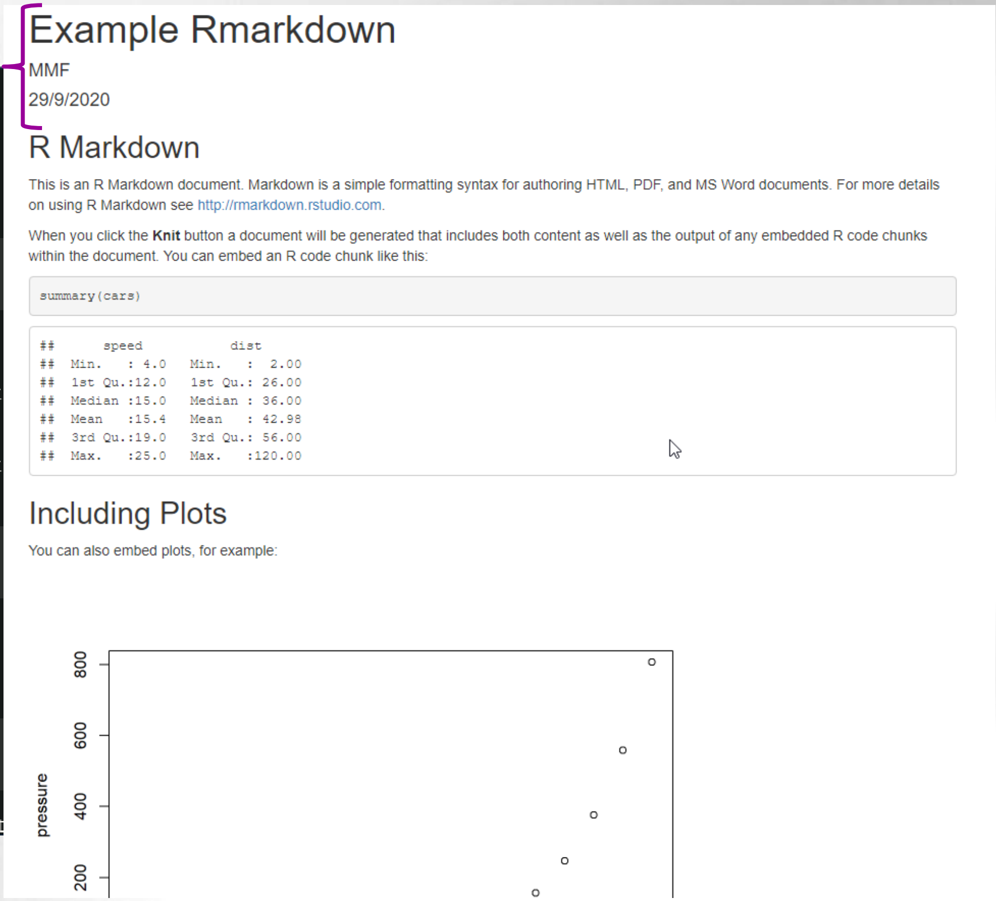
When you click the **Knit** button a document will be generated that
document. You can embed an R code chunk like this:

```{r cars}
summary(cars)
```

## Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk t
```



# Example Rmarkdown

MMF

29/9/2020

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

## Including Plots

You can also embed plots, for example:
```

# Descriptive Statistics: Summaries and Graphs

Basic Statistics with R for Biomedical Research

UEB – VHIR

Miriam Mota-Foix and Santiago Pérez-Hoyos

Miriam.mota@vhir.org santi.perezhoyos@vhir.org

# Index

1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
    1. Measures of location/central tendency
    2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
    1. Contingency Tables
    2. Graphs
5. Examples & exercises

# Index

1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
    1. Measures of location/central tendency
    2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
    1. Contingency Tables
    2. Graphs
5. Examples & exercises

# GENERAL CONCEPTS

- **Population**: It represents the largest group of individuals who want to study and generally usually inaccessible.
- **Sample:** Subset of the population in which measurements are done. This sample should be representative of the original population (any individual has equal opportunity to be selected).
- **Variable**: Feature measurable and observable that represents a concept of study
- **Measure:** Procedure for assign quantitative or qualitative values to the characteristics of objects, people or events. If these procedures are not well measured the validity of the results is not guaranteed.

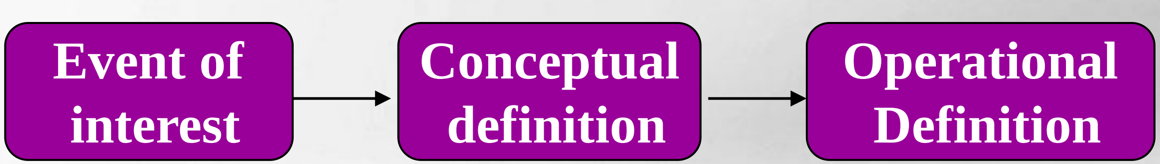
# STEPS IN A STATISTICAL STUDY ANALSYIS

1. **Make hypothesis about a population**
2. **Decide which data collect (Experimental design)**
   - Which individuals will be part of the study (samples)
   - Which data must be collected for each individual (variables)
3. **Collect Data**
4. **Describe (summarize) collected data**
   - Summary measures and graphs
   - Point estimations and confidence intervals
5. **Establish relations between two variables**
   - Set up Statistical Hypothesis test
   - Check application conditions
   - Calculate intensity relationship measures
6. **Multivariable analysis . Modelling**
   - Consider effects of several variables on an outcome
   - Regression models
   - More complex models

# Index

1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
   1. Measures of location/central tendency
   2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
   1. Contingency Tables
   2. Graphs
5. Examples & exercicis

# TYPES OF VARIABLES


Event of interest → Conceptual definition → Operational Definition

## QUALITATIVE

### NOMINAL

Measure qualities of an individual

**Examples: Sex, Treatment, Disease**

### ORDINAL

Measure qualities but they are ordered

**Examples: Educational level, Stage, Severity**

## QUANTITATIVE

### DISCRETE

Take only a finite possible values

**Examples: Nº of admissions, Nº of programmed visits**

### CONTINOUS

Can take an infinite number of values. Between two measures always can be another

**Example: Stay time, Age, Cholesterol level**

# Variable classification in a Study

- **Response, dependent or outcome variable**
  - One that answer the research question
- **Explain, independent or exposure variables**
  - They are those that are related to the causes of the events we want to study
- **Confounding or effect modifier variables**
  - Are those that can affect the relation between exposure and outcome variables
- **Universal variables**
  - Are those that can be exposures or confounders that always have to be considered. For example: sex, age, residence location, ethnic, etc.

# Descriptive analysis

- Data have to be **organized** to be useful (frequency or contingency tables)
- **Graph** data before calculating summary measures
- This actions can help to:
  - Select the **best summary** measure
  - **Transform** variables
  - Detect **outliers**

# Index

1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
    1. Measures of location/central tendency
    2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
    1. Contingency Tables
    2. Graphs
5. Examples & exercises

# Quantitative Variables

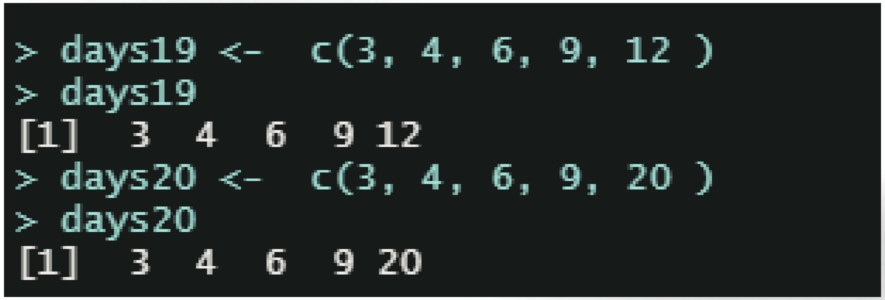
- We have a new variable (i.e. a biomarker and we want to summarize information)
  - Around which values is the variable ?
  - Values vary greatly between different individuals
  - Data are grouped or not

# Summary Measures

- Location
  - **Mean**
  - **Median**
  - Mode
- Dispersion
  - Range (Maximum-Minimum)
  - **Variance**
  - **Standard Deviation**
  - Variation Coefficient
  - Percentile
  - **Interquartile range (IQR)** or Interquartile interval
- Shape
  - Asymmetry
  - Kurtosis

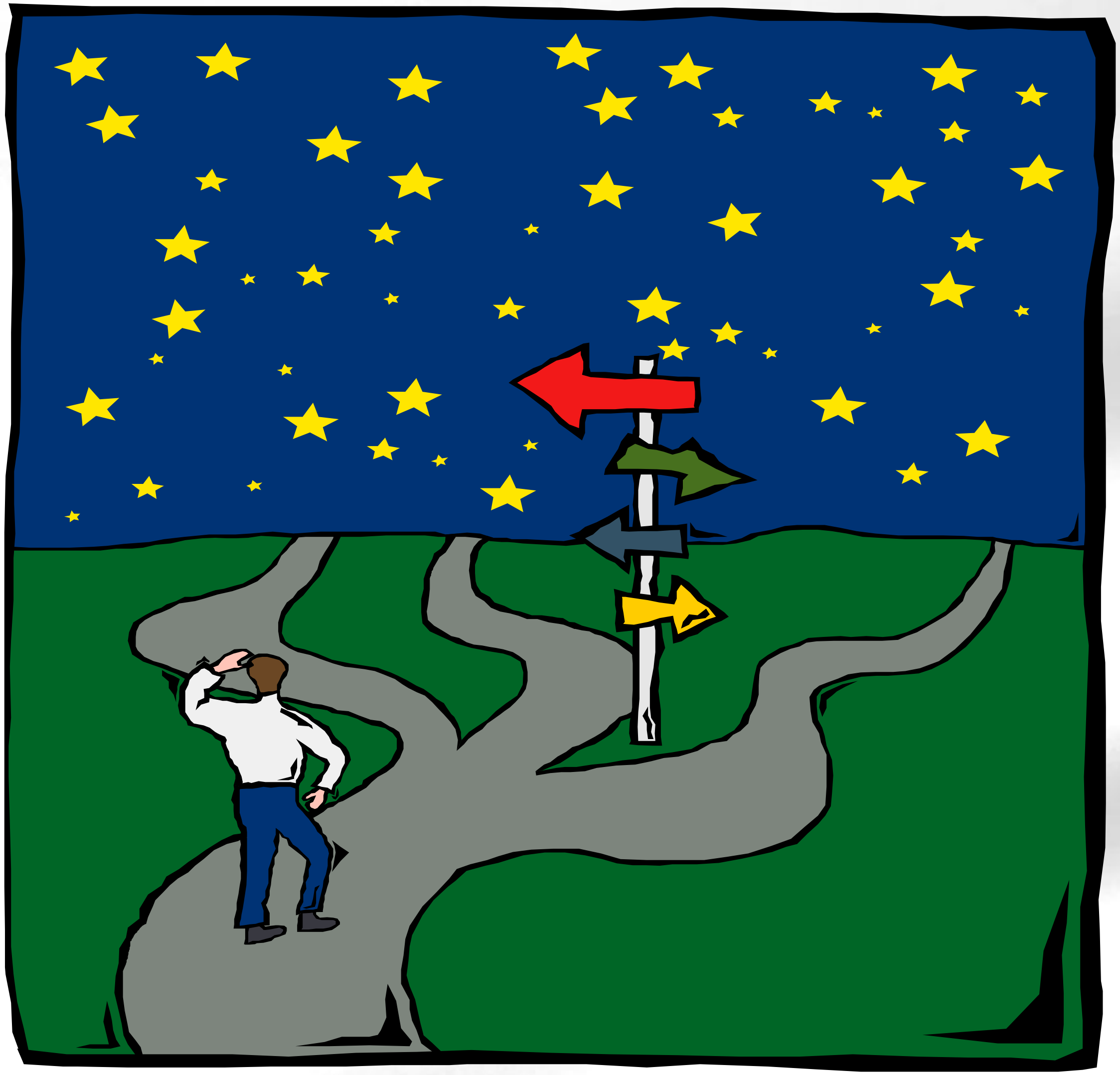
# Data

```
days19 <-  c(3, 4, 6, 9, 12 )

days19

days20 <-  c(3, 4, 6, 9, 20 )

days20
```



```
> days19 <-  c(3, 4, 6, 9, 12 )
> days19
[1]  3  4  6  9 12
> days20 <-  c(3, 4, 6, 9, 20 )
> days20
[1]  3  4  6  9 20
```

# Location measures



Mean

Median

Mode

# Mean

- Useful to locate data .
- Is the **sum** of observed values **over sample size**
- Can be altered by extreme values

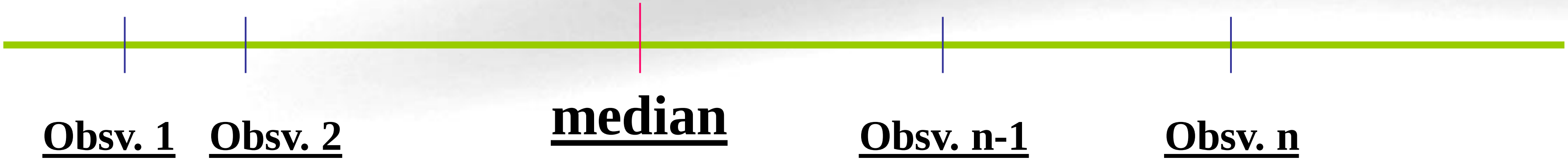
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

**Example Stay days**

| | |
|---|---|
| **3, 4, 6, 9, 12** | **Mean=6,8** |
| **3, 4, 6, 9, 20** | **Mean=8,4** |

# Median

- Is the point that divied in **two parts** the observations
- Observations are ordered from lowest to highest and median is the **central point**
- It is not altered by extreme observations



Obsv. 1 Obsv. 2 **median** Obsv. n-1 Obsv. n

**Example Stay days**

**3, 4, 6, 9, 12** **Median=6**

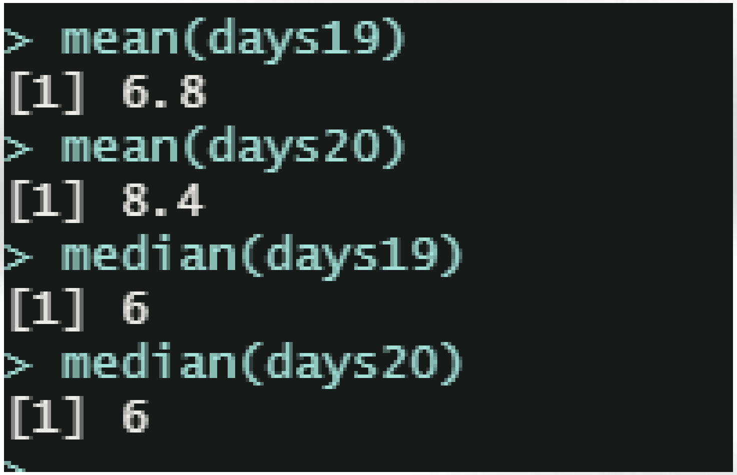
**3, 4, 6, 9, 20** **Median=6**

# Example: Location measures

CODE:

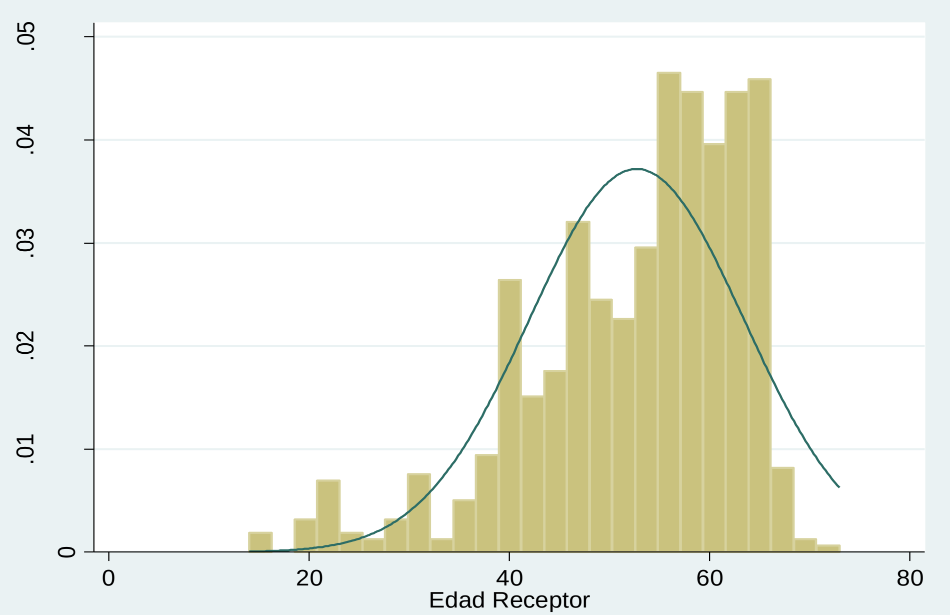
```
mean(days19)

mean(days20)

median(days19)

median(days20)
```

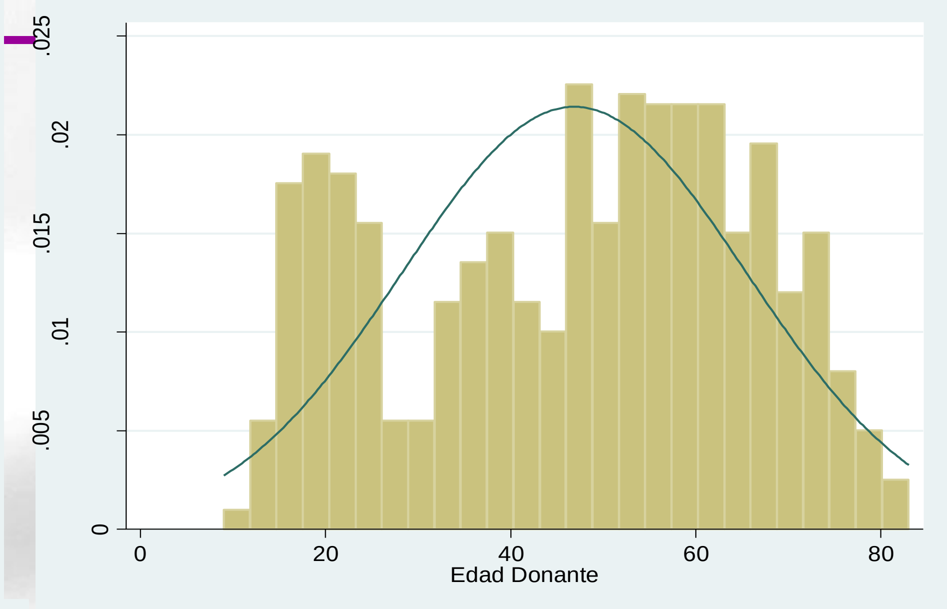
RESULT:

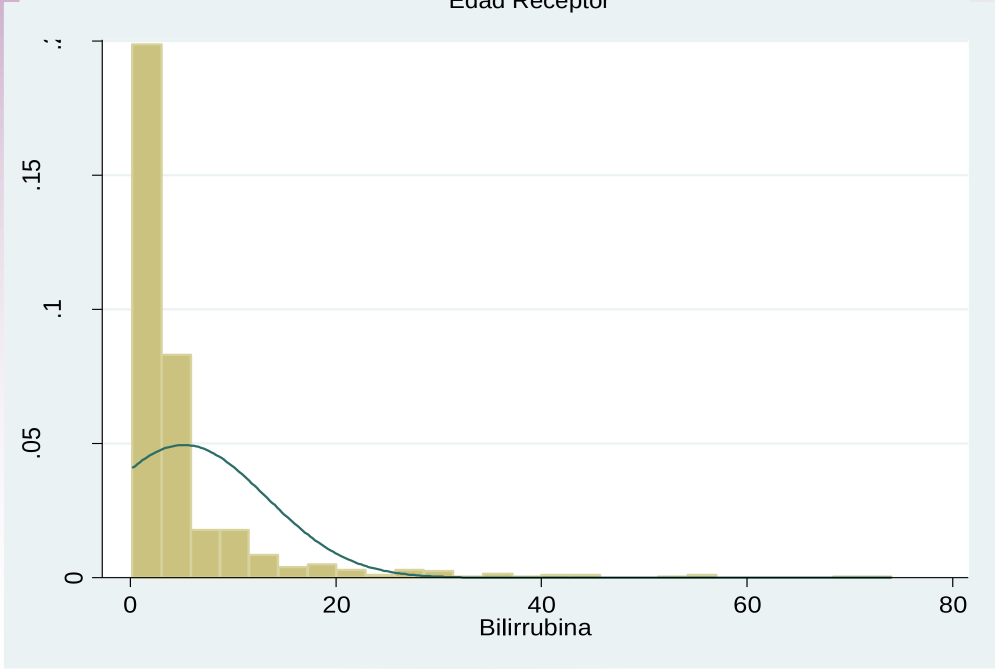


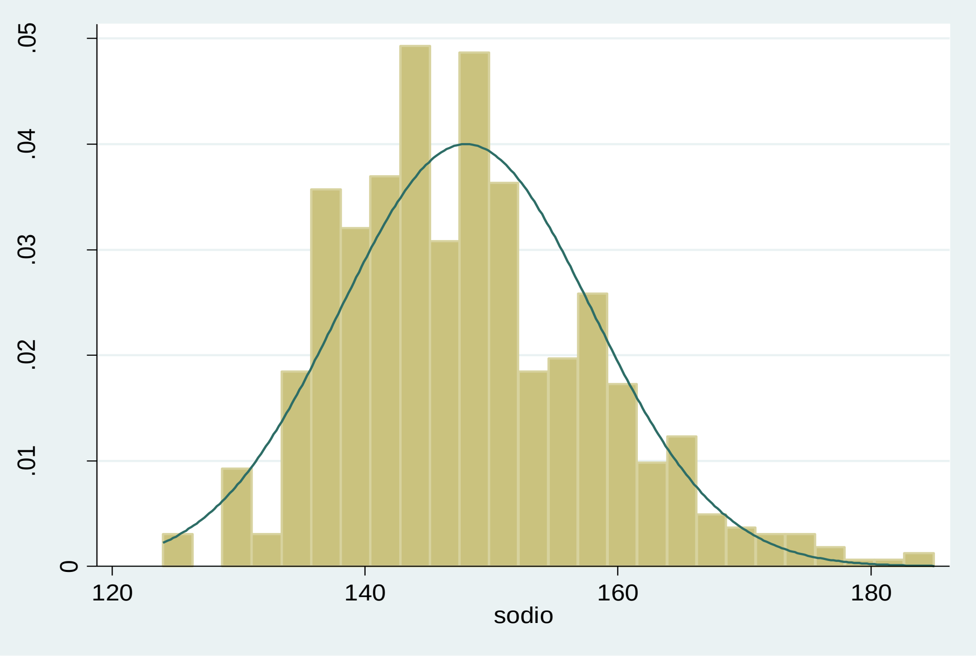
```
> mean(days19)
[1] 6.8
> mean(days20)
[1] 8.4
> median(days19)
[1] 6
> median(days20)
[1] 6
```

# Transplant study: Mean or Median



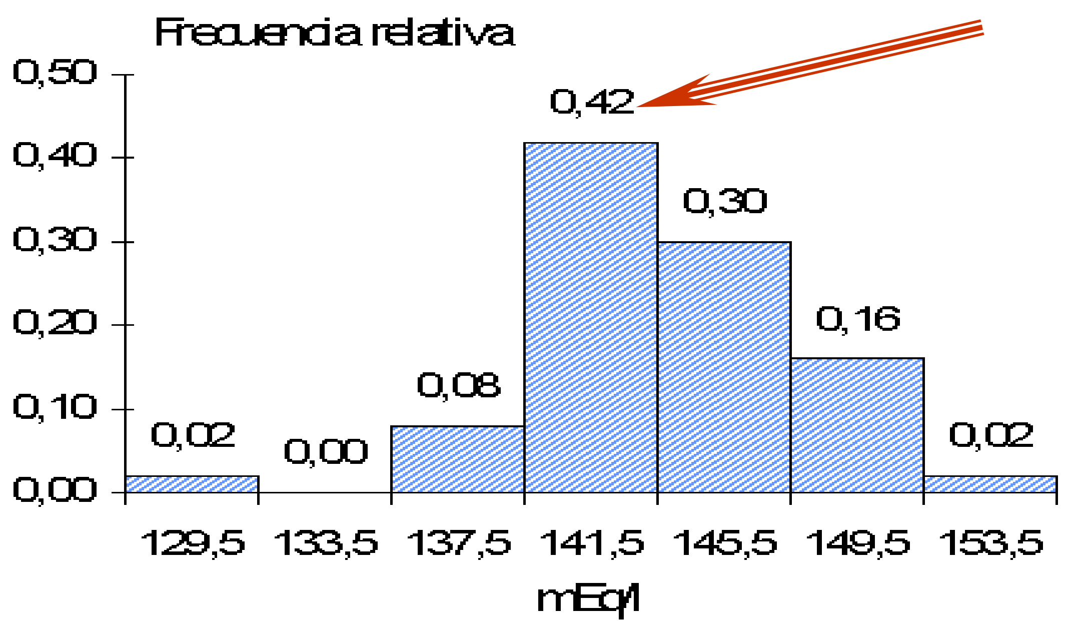






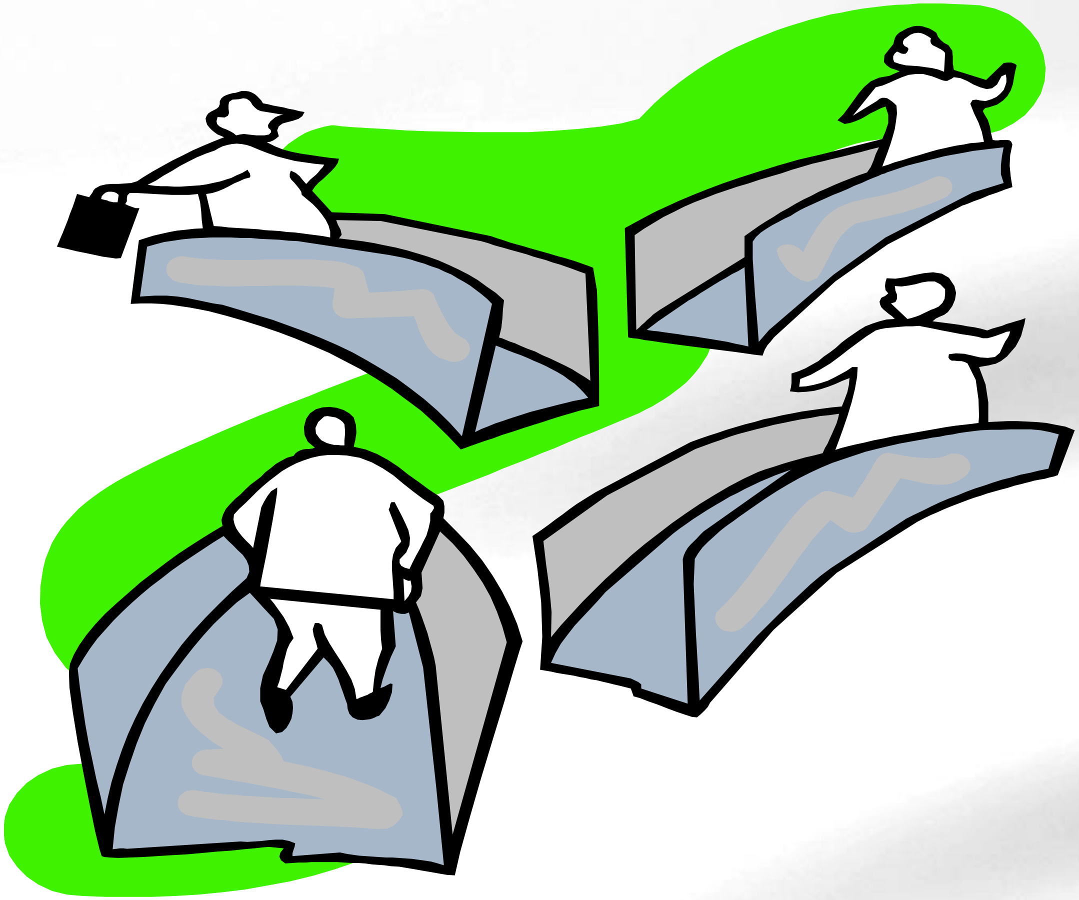
# Mode

- The most frequent value
- May be not unique
- In a quantitative variable is the maximum values of an histogram



Determinaciones de sodio

# Dispersion or variability measures



Range ( Maximum-Minimum)

Variance

Standard Deviation

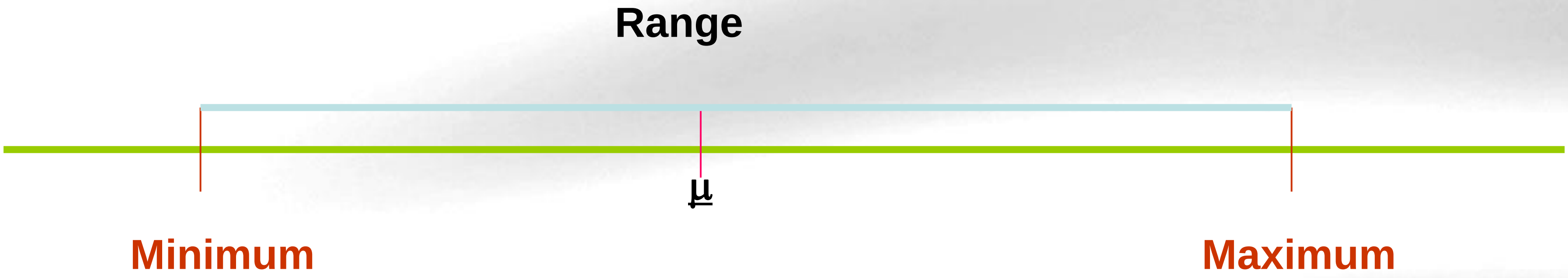
Variation Coefficient

Percentile

Interquartile Range (IQR)

# Range

- Simplest measure of dispersion
- Is the difference between maximum and minimum value of the observations



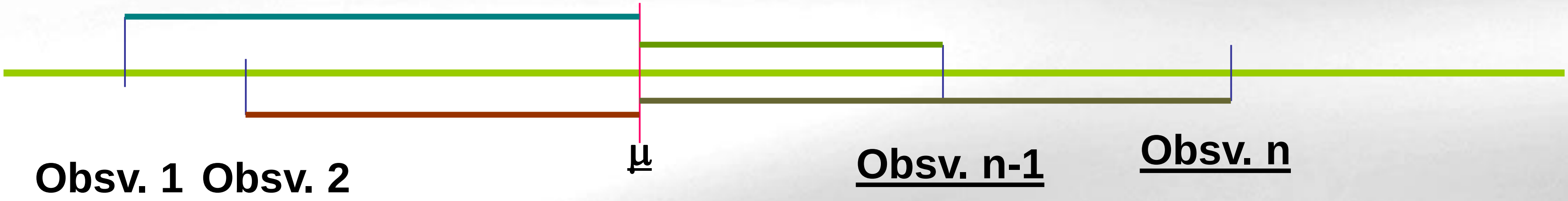
**Example Stay days**

**3, 4, 6, 9, 12** **Range=12-3=9**

**3, 4, 6, 9, 20** **Range=20-3=17**

# Variance

- Mean difference of observations from mean in squared scale



| | d | d-mean | (d-mean)^2 | d | d-mean | (d-mean)^2 |
|---|---|---|---|---|---|---|
| | 3 | -3,8 | 14,44 | 3 | -5,4 | 29,16 |
| | 4 | -2,8 | 7,84 | 4 | -4,4 | 19,36 |
| | 6 | -0,8 | 0,64 | 6 | -2,4 | 5,76 |
| | 9 | 2,2 | 4,84 | 9 | 0,6 | 0,36 |
| | 12 | 5,2 | 27,04 | 20 | 11,6 | 134,56 |
| Sum | | 0 | 54,8 | | 0 | 189,2 |
| Sum/5 | 6,8 | 0 | 10,96 | 8,4 | 0 | 37,84 |

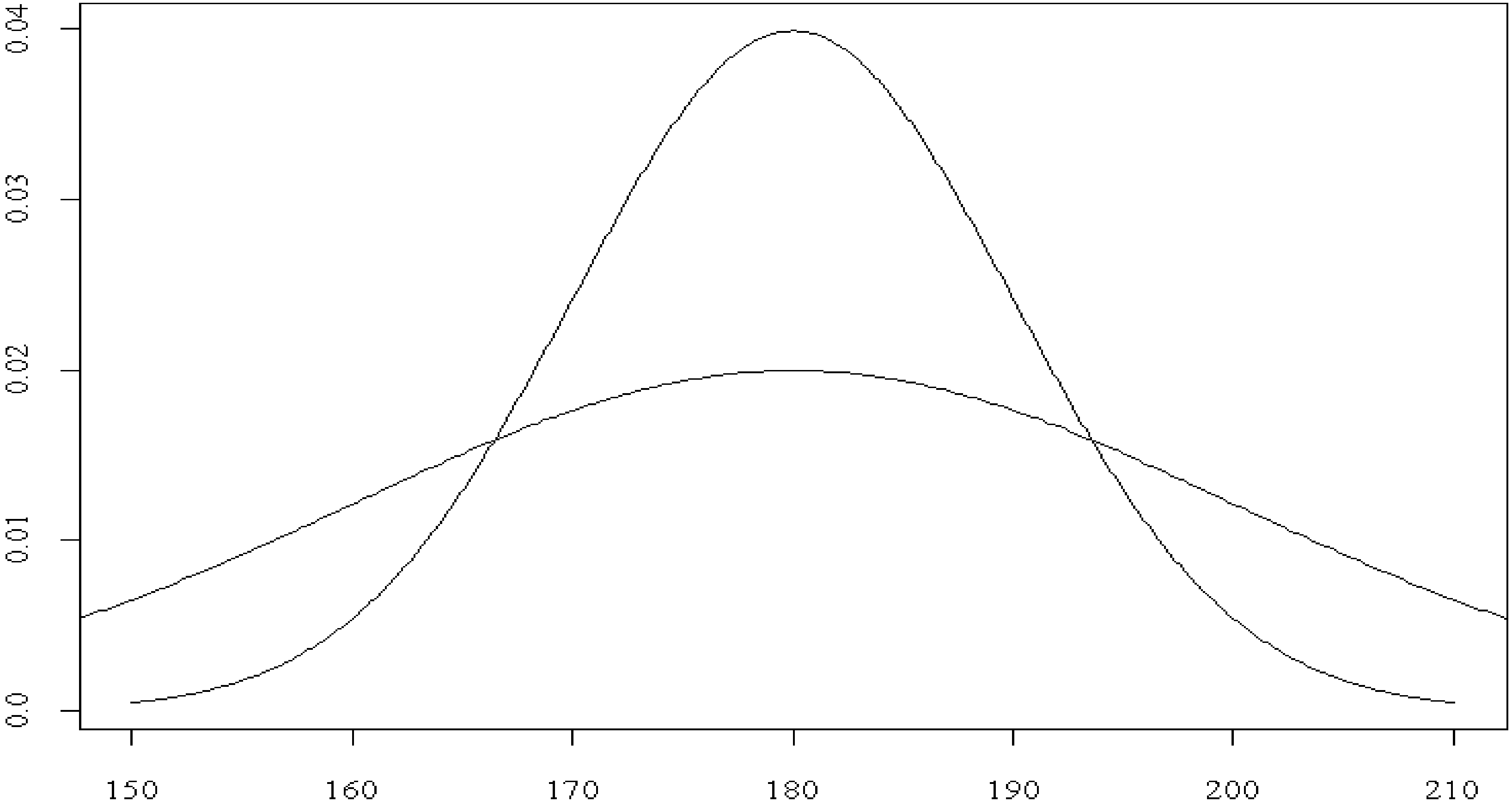
# Standard Deviation

- Squared root of the variance
- It is measured in the same units than the variable

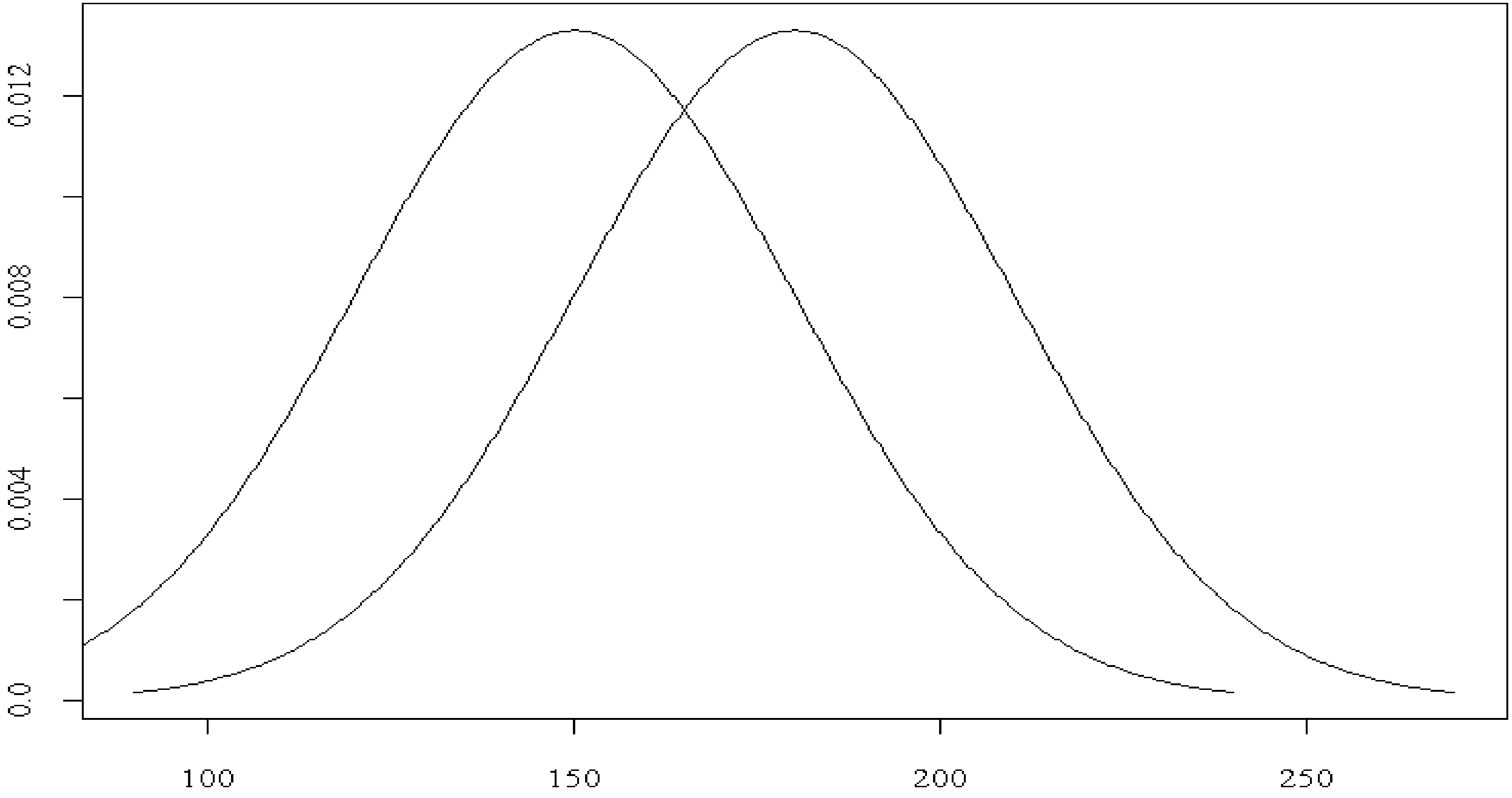
**Example Stay days**

| | |
|---|---|
| **3, 4, 6, 9, 12** | **Variance =10.96** |
| | **Std. Dev.=3.31** |
| **3, 4, 6, 9, 20** | **Variance= 37.84** |
| | **Std. Dev.= 6.15** |

# Same mean , different variances



# Same variances, different means



# Coefficient of Variation (CV)

- It is the ratio between standard deviation and mean
- Allows to compare the variability of variables measured in different scales

**Example Stay Days**

**3, 4, 6, 9, 12**

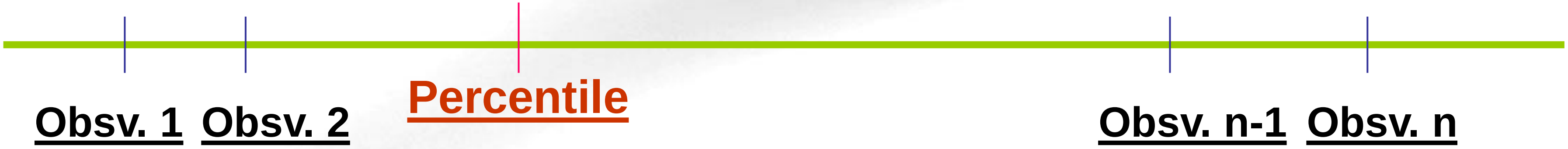
**Std. Dev.= 3.31**
**Mean = 6.8**
**Variation Coef.= 0.49**

**3, 4, 6, 9, 20**

**Std. Dev. = 6.15**
**Mean = 8.4**
**Variation Coef.= 0.73**

# Percentiles

- Observations are ranked from minimum to maximum and the point that leaves below p% of observations is selected
- There are some special percentiles
  - Deciles are percentiles 10, 20, 30, 40, 50, 60, 70, 80, 90
  - Quartiles are  percentiles 25, 50, 75
  - Quintiles are percentiles 20,40, 60,80
  - They are not affected by extreme observations
  - Interquartile range is difference between 25 and 75 percentile



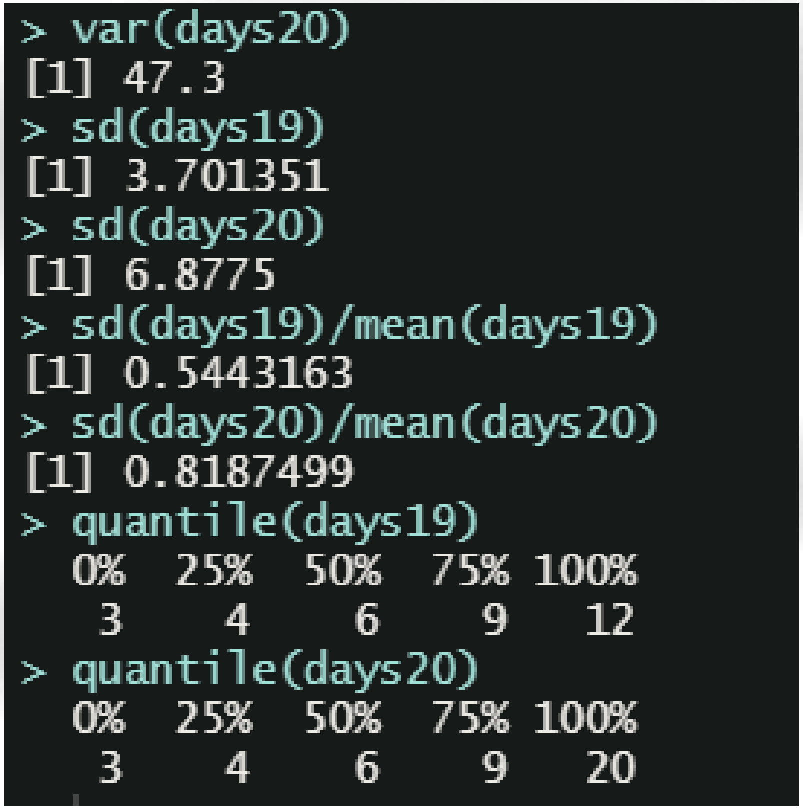
Obsv. 1 Obsv. 2 Percentile Obsv. n-1 Obsv. n

# Example: Variability measures

CODE:

```
var(days19)

var(days20)

sd(days19)

sd(days20)

sd(days19)/mean(days19)

sd(days20)/mean(days20)

quantile(days19)

quantile(days20)
```

RESULT:



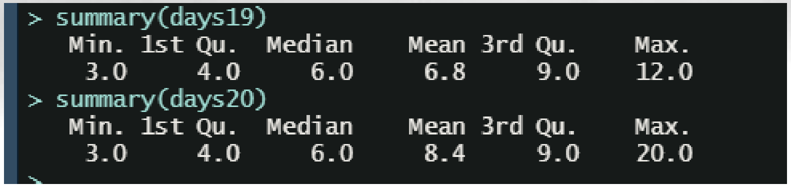
```
> var(days20)
[1] 47.3
> sd(days19)
[1] 3.701351
> sd(days20)
[1] 6.8775
> sd(days19)/mean(days19)
[1] 0.5443163
> sd(days20)/mean(days20)
[1] 0.8187499
> quantile(days19)
  0%  25%  50%  75% 100% 
   3    4    6    9   12 
> quantile(days20)
  0%  25%  50%  75% 100% 
   3    4    6    9   20 
```

# Summary measures in R

CODE:

```
summary(days19)

summary(days20)
```

RESULT:



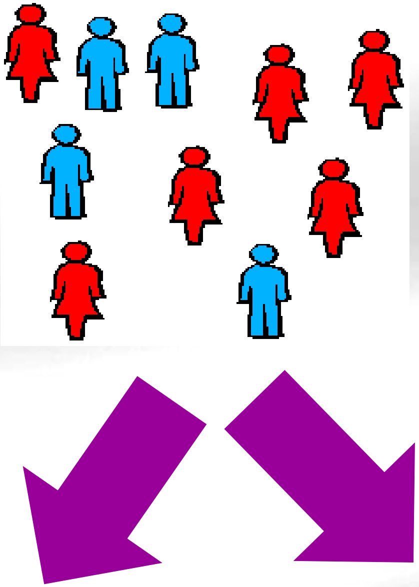
```
> summary(days19)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0     4.0     6.0     6.8     9.0    12.0
> summary(days20)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    3.0     4.0     6.0     8.4     9.0    20.0
>
```

# Syllabus

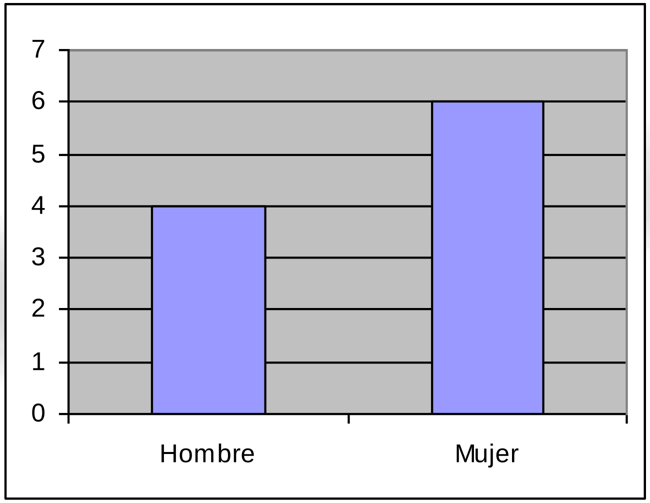
1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
    1. Measures of location/central tendency
    2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
    1. Contingency Tables
    2. Graphs
5. Examples & exercises

# Summary of variables

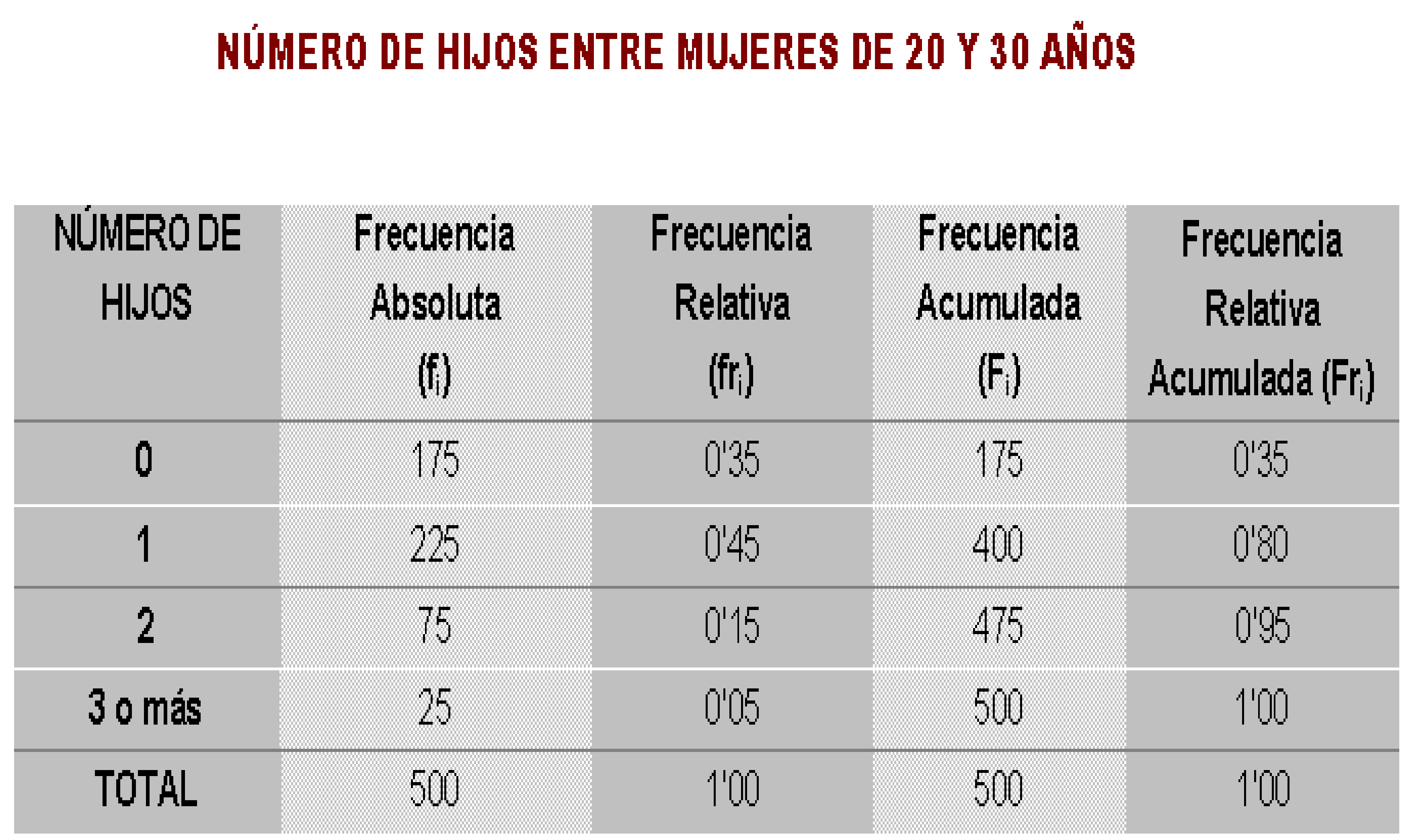
Frequency tables and graphs are two equivalent ways to present information. Both expose in an ordered way the collected data.a.



| Género | Frec. |
|---|---|
| Hombre | 4 |
| Mujer | 6 |



# Frequency table



NÚMERO DE HIJOS ENTRE MUJERES DE 20 Y 30 AÑOS

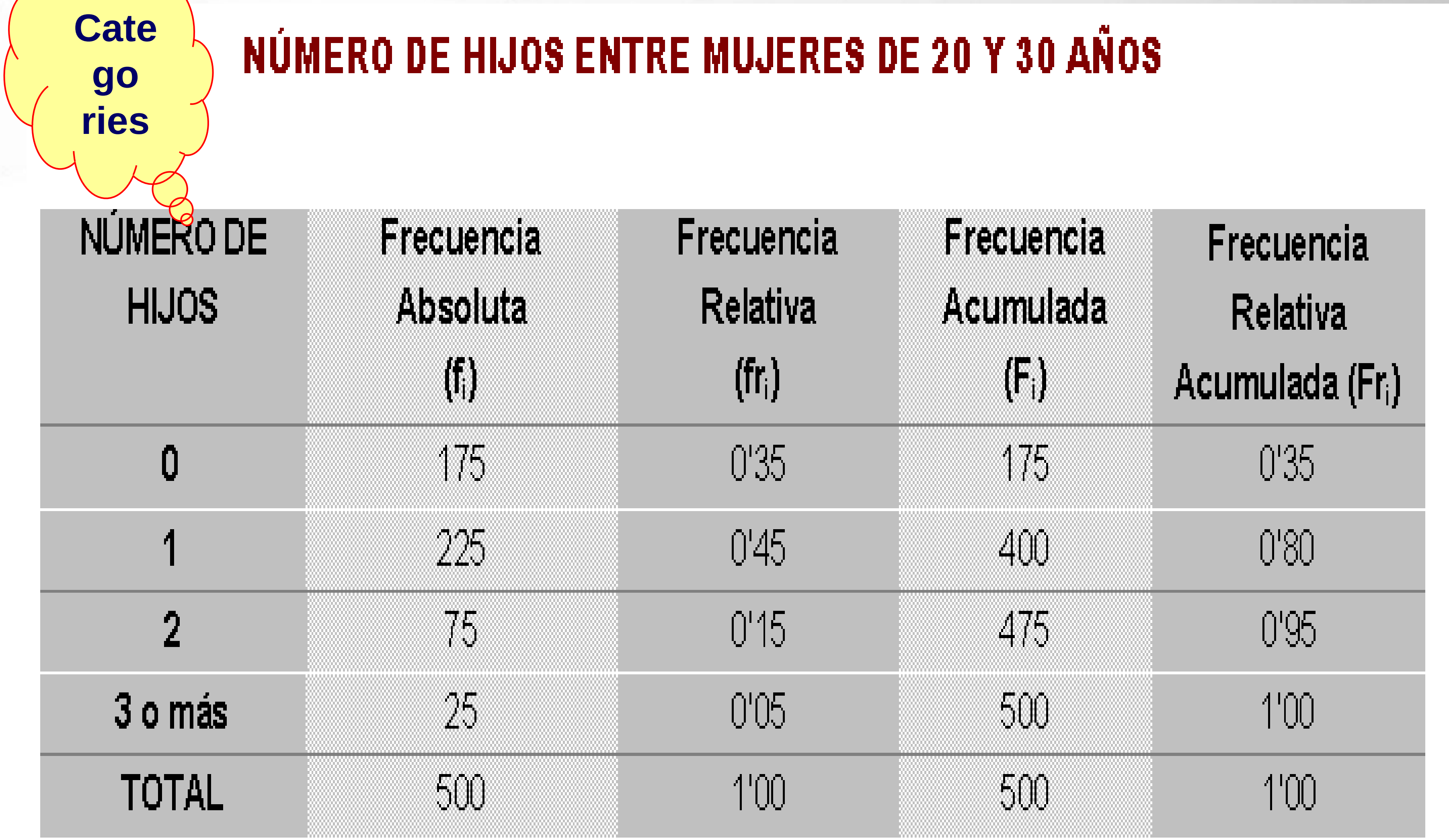
| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'95 |
| 3 o más | 25 | 0'05 | 500 | 1'00 |
| TOTAL | 500 | 1'00 | 500 | 1'00 |

# Frequency table

UNITAT D'ESTADÍSTICA I BIOINFORMÀTICA

Vall d'Hebron Institut de Recerca

Cate
go
ries



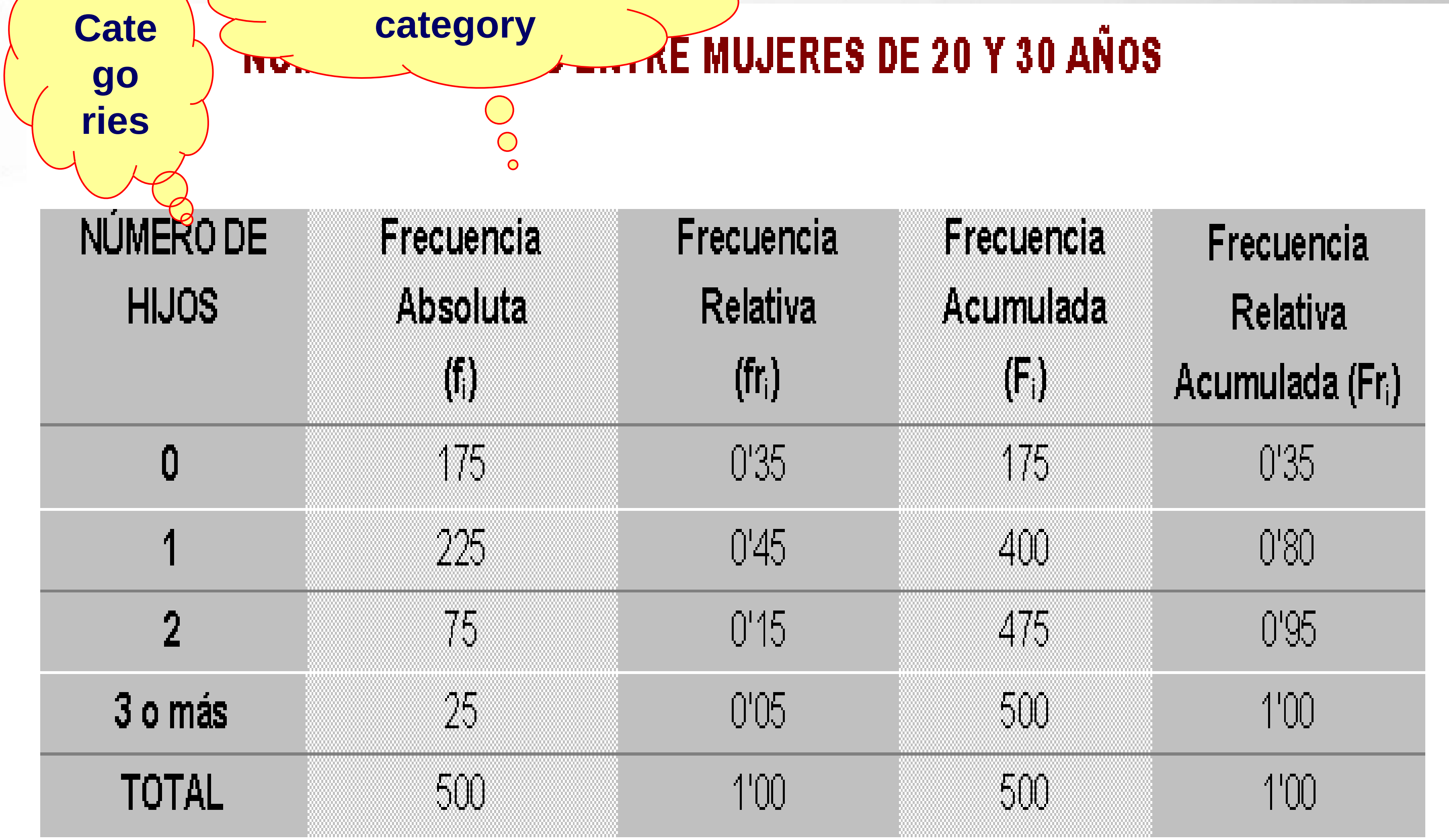
NÚMERO DE HIJOS ENTRE MUJERES DE 20 Y 30 AÑOS

| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'95 |
| 3 o más | 25 | 0'05 | 500 | 1'00 |
| TOTAL | 500 | 1'00 | 500 | 1'00 |

# Taula de Freqüencia

Categories

Nº of subjects by category



NU… ENTRE MUJERES DE 20 Y 30 AÑOS

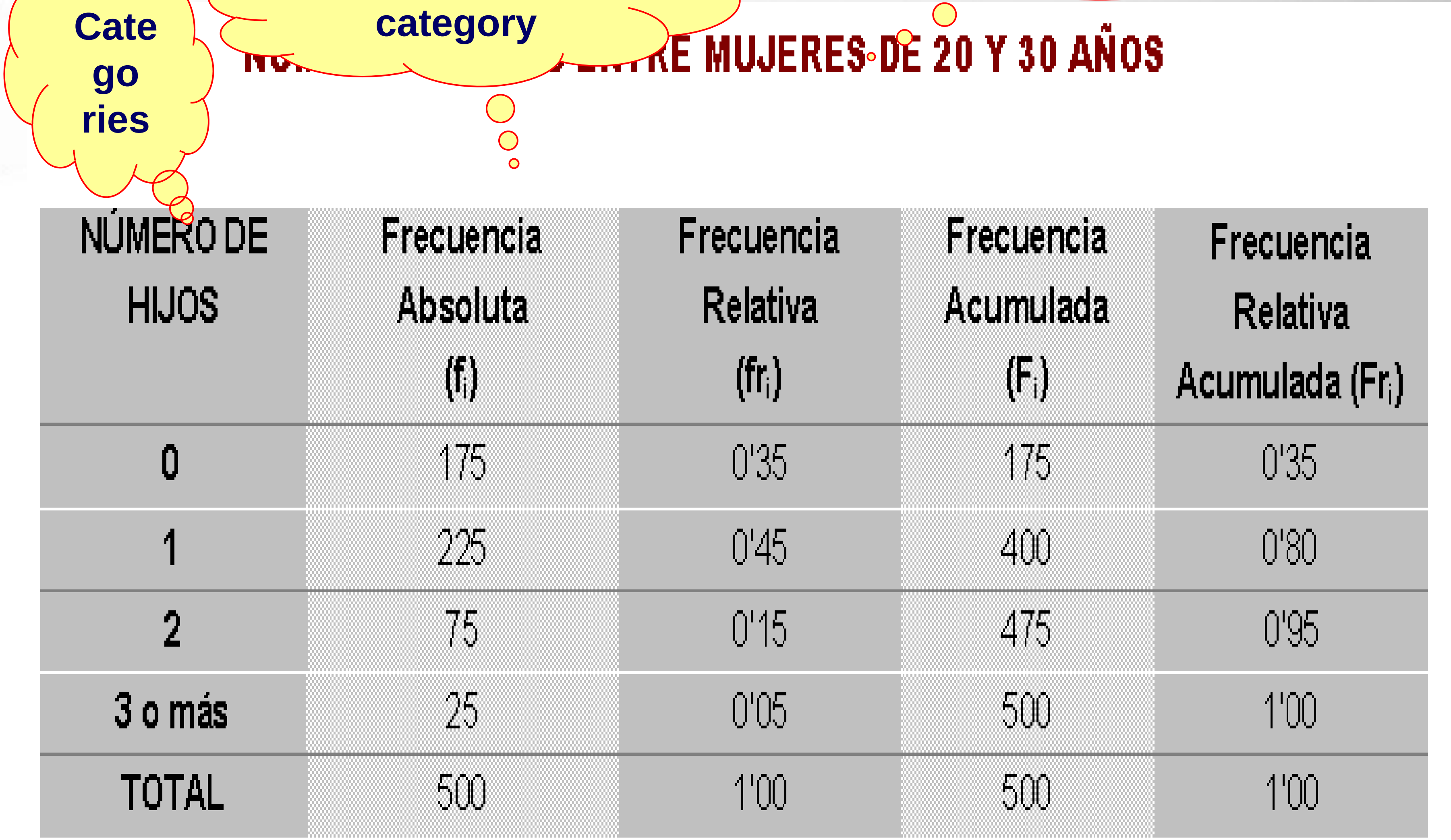
| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'95 |
| 3 o más | 25 | 0'05 | 500 | 1'00 |
| TOTAL | 500 | 1'00 | 500 | 1'00 |

# Frequency table

Cate go ries

Nº of subjects by category

Percentage of subjects
Freq /Total



NÚ... ...RE MUJERES DE 20 Y 30 AÑOS

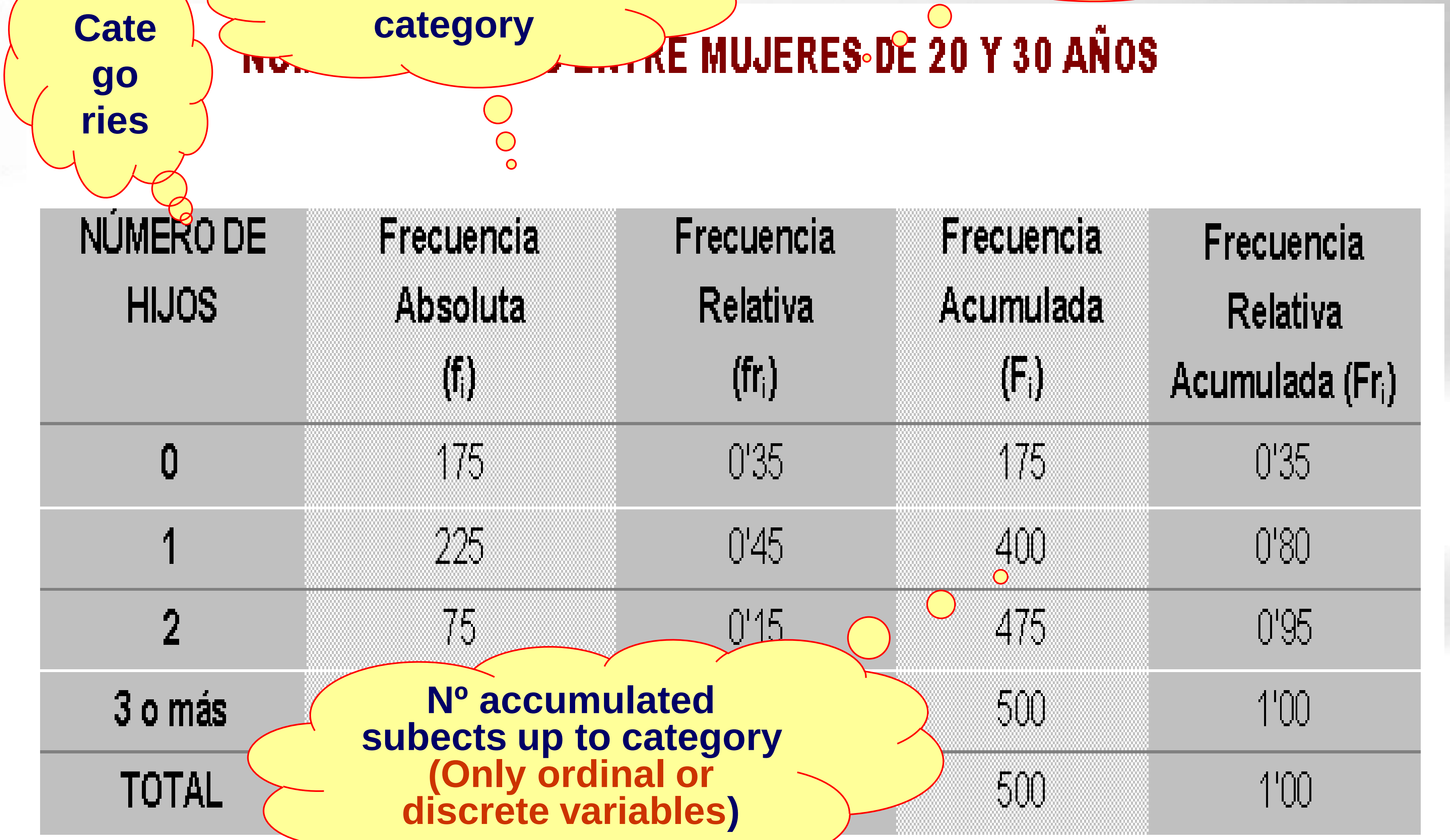
| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'95 |
| 3 o más | 25 | 0'05 | 500 | 1'00 |
| TOTAL | 500 | 1'00 | 500 | 1'00 |

# Frequency table

Cate
go
ries

Nº of subjects by category

Percentage of subjects
Freq /Total



NU ENTRE MUJERES DE 20 Y 30 AÑOS

| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'95 |
| 3 o más | | | 500 | 1'00 |
| TOTAL | | | 500 | 1'00 |

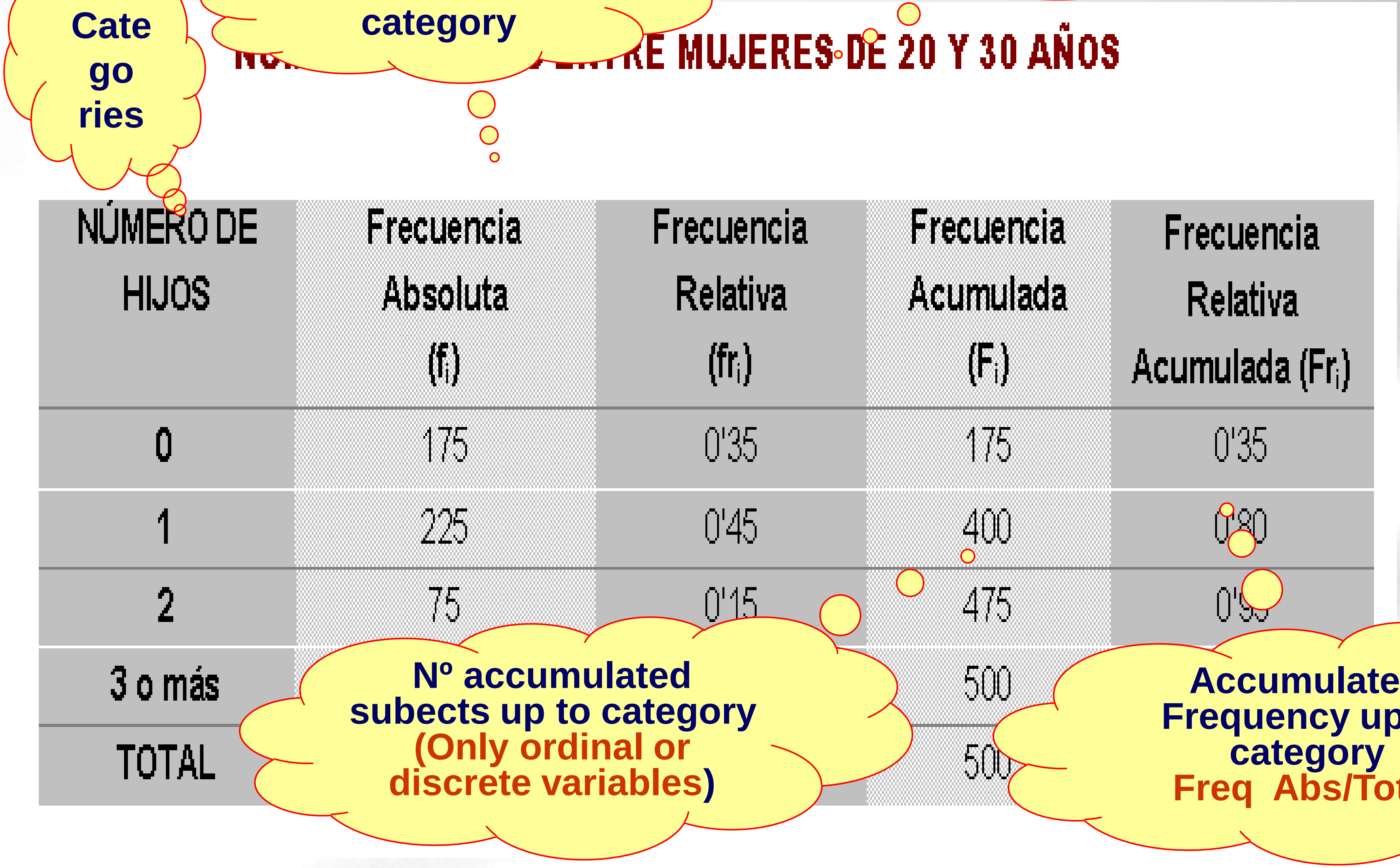
Nº accumulated subects up to category
(Only ordinal or discrete variables)

# Frequency table

Percentage of subjects
Freq /Total

Cate
go
ries

Nº of subjects by
category



NU[illegible]RE MUJERES DE 20 Y 30 AÑOS

| NÚMERO DE HIJOS | Frecuencia Absoluta ($f_i$) | Frecuencia Relativa ($fr_i$) | Frecuencia Acumulada ($F_i$) | Frecuencia Relativa Acumulada ($Fr_i$) |
|---|---|---|---|---|
| 0 | 175 | 0'35 | 175 | 0'35 |
| 1 | 225 | 0'45 | 400 | 0'80 |
| 2 | 75 | 0'15 | 475 | 0'9[illegible] |
| 3 o más | [illegible] | [illegible] | 500 | [illegible] |
| TOTAL | [illegible] | [illegible] | 500 | [illegible] |

Nº accumulated
subects up to category
(Only ordinal or
discrete variables)

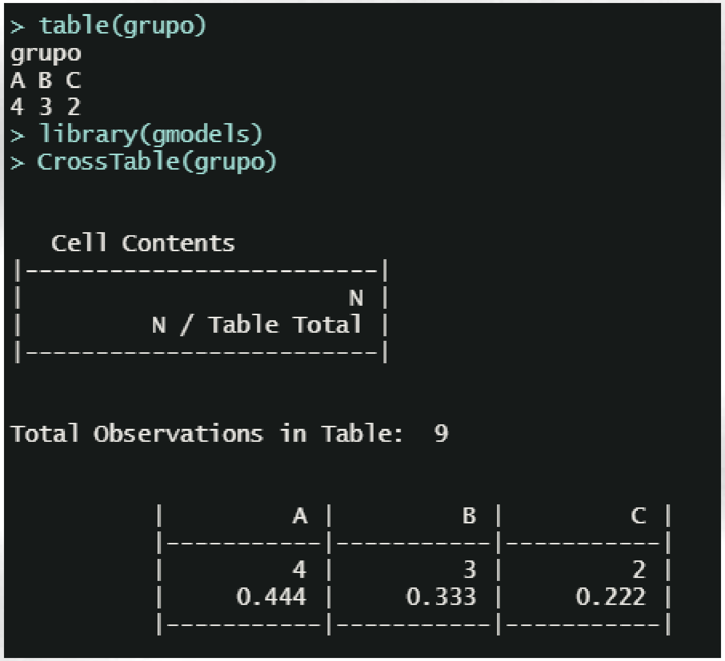
Accumulated
Frequency up to
category
Freq Abs/Total

# Example: Frequency table

CODE:

```
grupo <- factor(c("A",
"A","B","B","A","C","C",
"A", "B"))
table(grupo)
library(gmodels)
CrossTable(grupo)
```

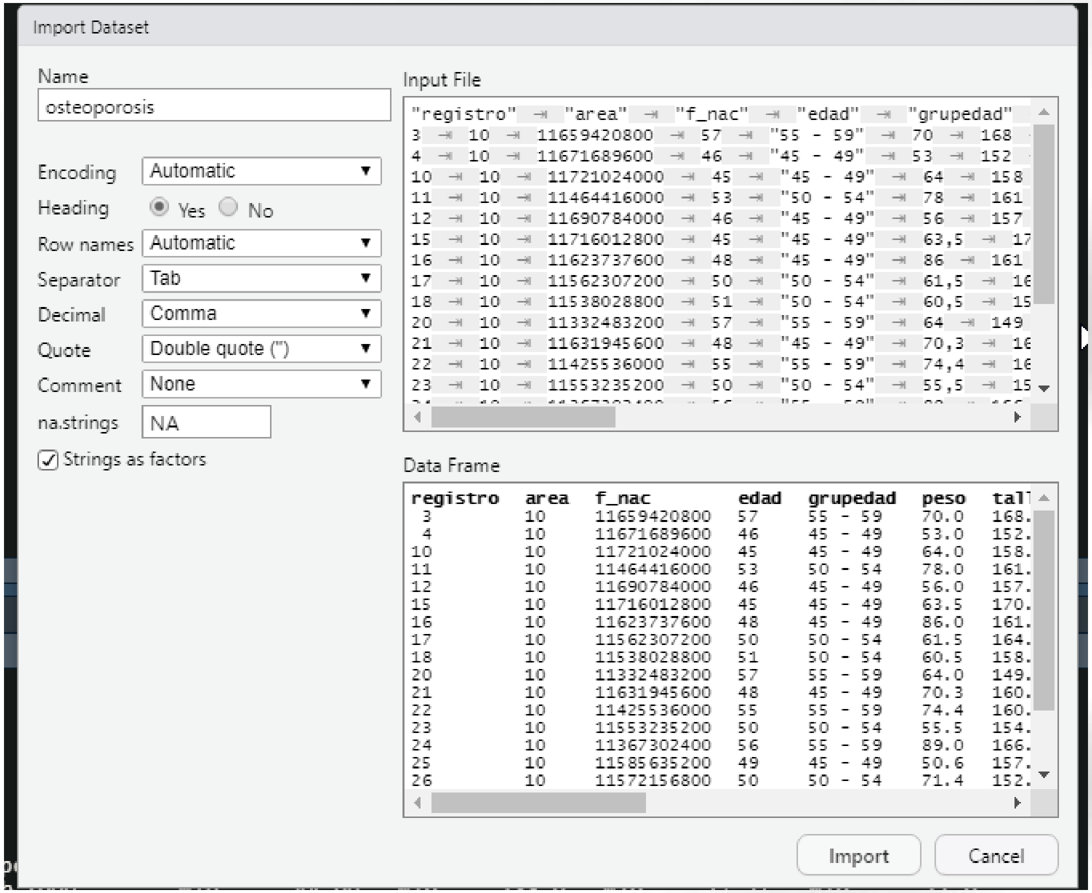
RESULT:



```
> table(grupo)
grupo
A B C
4 3 2
> library(gmodels)
> CrossTable(grupo)


   Cell Contents
|-------------------------|
|                       N |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  9


          |         A |         B |         C |
          |-----------|-----------|-----------|
          |         4 |         3 |         2 |
          |     0.444 |     0.333 |     0.222 |
          |-----------|-----------|-----------|
```

# Dataset osteoporosis

```
osteo <- read.delim2("/osteoporosis.csv")
```



Import Dataset

Name: osteoporosis

Encoding: Automatic

Heading: ● Yes ○ No

Row names: Automatic

Separator: Tab

Decimal: Comma

Quote: Double quote (")

Comment: None

na.strings: NA

☑ Strings as factors

Input File

```
"registro" ⇥ "area" ⇥ "f_nac" ⇥ "edad" ⇥ "grupedad"
3  ⇥ 10 ⇥ 11659420800 ⇥ 57 ⇥ "55 - 59" ⇥ 70   ⇥ 168
4  ⇥ 10 ⇥ 11671689600 ⇥ 46 ⇥ "45 - 49" ⇥ 53   ⇥ 152
10 ⇥ 10 ⇥ 11721024000 ⇥ 45 ⇥ "45 - 49" ⇥ 64   ⇥ 158
11 ⇥ 10 ⇥ 11464416000 ⇥ 53 ⇥ "50 - 54" ⇥ 78   ⇥ 161
12 ⇥ 10 ⇥ 11690784000 ⇥ 46 ⇥ "45 - 49" ⇥ 56   ⇥ 157
15 ⇥ 10 ⇥ 11716012800 ⇥ 45 ⇥ "45 - 49" ⇥ 63,5 ⇥ 17
16 ⇥ 10 ⇥ 11623737600 ⇥ 48 ⇥ "45 - 49" ⇥ 86   ⇥ 161
17 ⇥ 10 ⇥ 11562307200 ⇥ 50 ⇥ "50 - 54" ⇥ 61,5 ⇥ 16
18 ⇥ 10 ⇥ 11538028800 ⇥ 51 ⇥ "50 - 54" ⇥ 60,5 ⇥ 15
20 ⇥ 10 ⇥ 11332483200 ⇥ 57 ⇥ "55 - 59" ⇥ 64   ⇥ 149
21 ⇥ 10 ⇥ 11631945600 ⇥ 48 ⇥ "45 - 49" ⇥ 70,3 ⇥ 16
22 ⇥ 10 ⇥ 11425536000 ⇥ 55 ⇥ "55 - 59" ⇥ 74,4 ⇥ 16
23 ⇥ 10 ⇥ 11553235200 ⇥ 50 ⇥ "50 - 54" ⇥ 55,5 ⇥ 15
```

Data Frame

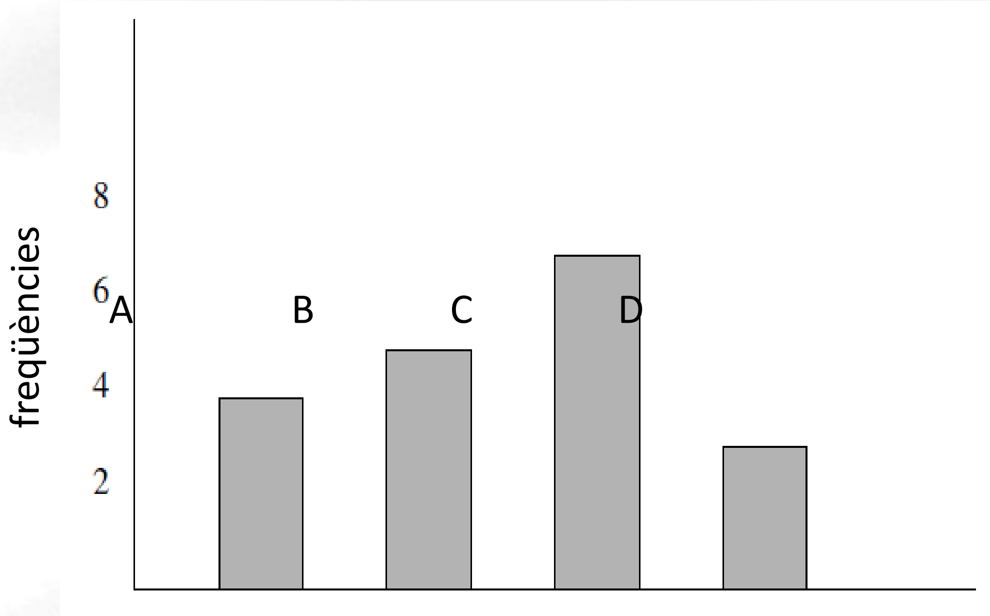
```
registro  area  f_nac        edad  grupedad  peso  tall
 3        10    11659420800  57    55 - 59   70.0  168.
 4        10    11671689600  46    45 - 49   53.0  152.
10        10    11721024000  45    45 - 49   64.0  158.
11        10    11464416000  53    50 - 54   78.0  161.
12        10    11690784000  46    45 - 49   56.0  157.
15        10    11716012800  45    45 - 49   63.5  170.
16        10    11623737600  48    45 - 49   86.0  161.
17        10    11562307200  50    50 - 54   61.5  164.
18        10    11538028800  51    50 - 54   60.5  158.
20        10    11332483200  57    55 - 59   64.0  149.
21        10    11631945600  48    45 - 49   70.3  160.
22        10    11425536000  55    55 - 59   74.4  160.
23        10    11553235200  50    50 - 54   55.5  154.
24        10    11367302400  56    55 - 59   89.0  166.
25        10    11585635200  49    45 - 49   50.6  157.
26        10    11572156800  50    50 - 54   71.4  152.
```

```
#All dataset
summary(osteo)

# quantitative variable
mean(osteo$edad)
sd(osteo$edad)
median(osteo$edad)
IQR(osteo$edad)

# qualitative variable
library(gmodels)
CrossTable(osteo$grupedad)
```

# Bar Graph

- Categorys are representened in X axis and frequencies in Y axis

- For comparing two population better use relative frequencies

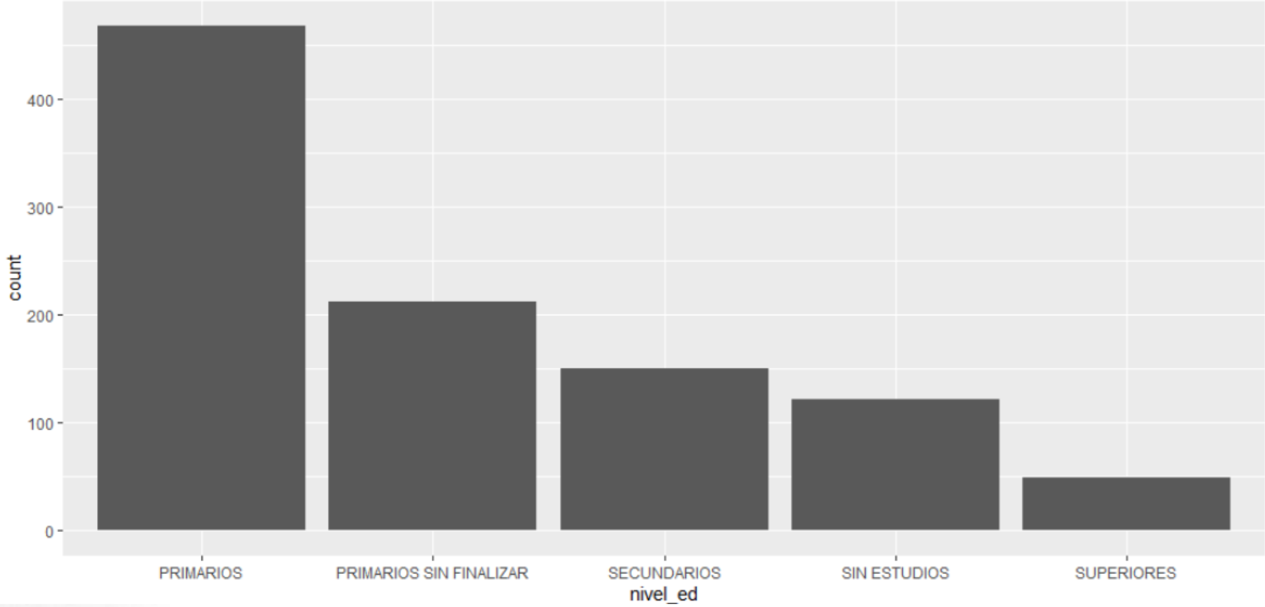


# Example. Bar Graph

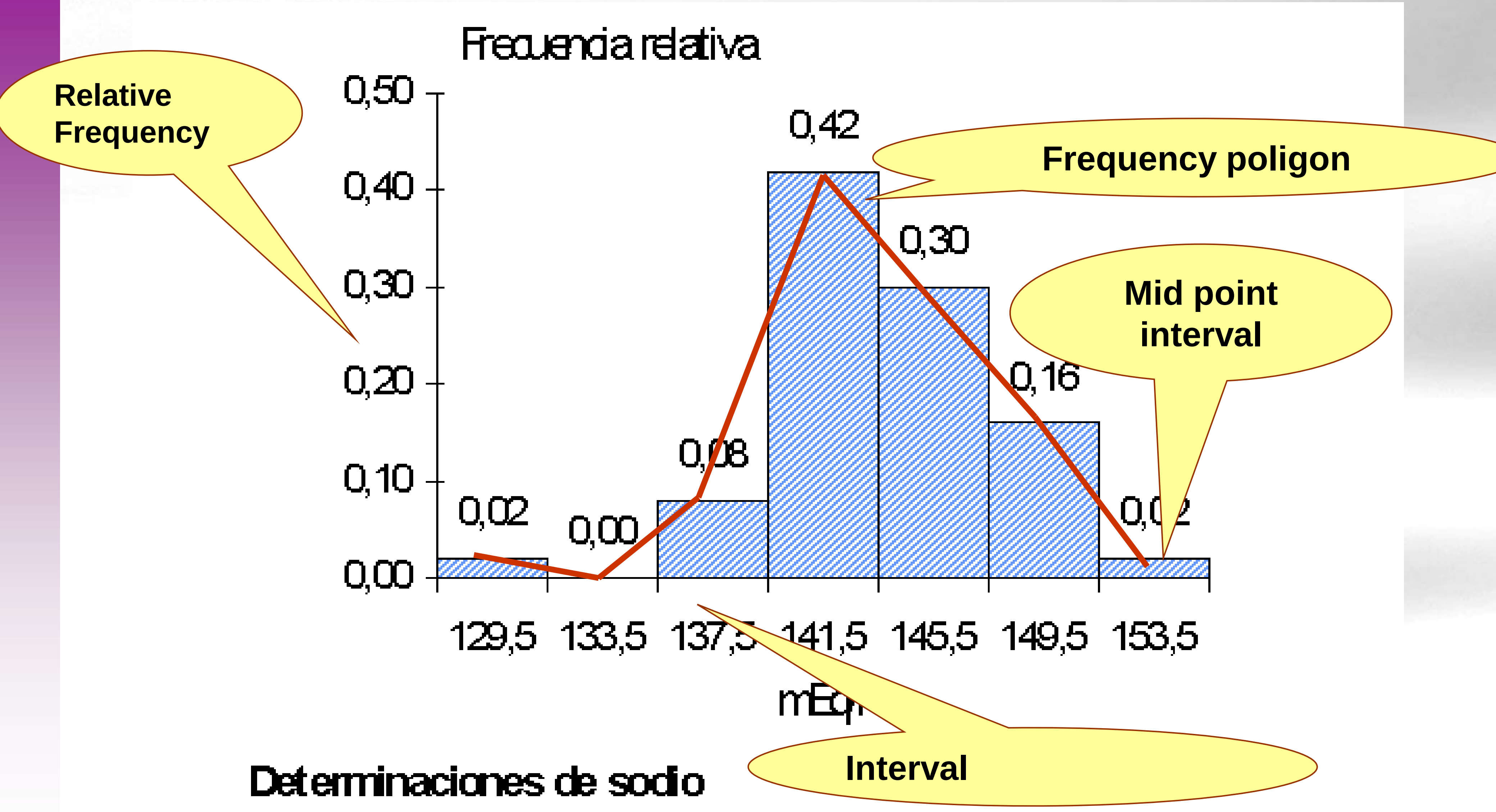
CODE:

```
require(ggplot2)

ggplot(data = osteo) +

  geom_bar(mapping = aes(x = nivel_ed))
```

RESULT:



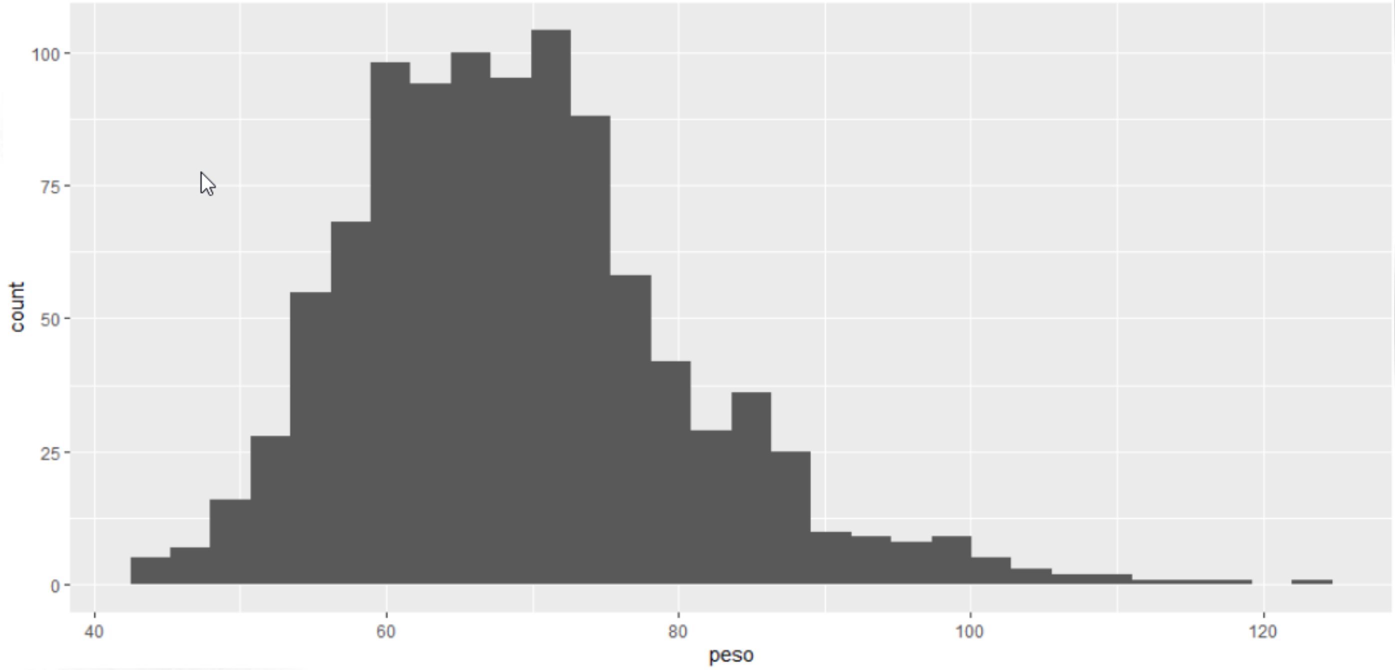
# Histogram


Frecuencia relativa
0,50
0,40
0,30
0,20
0,10
0,00
0,02
0,00
0,08
0,42
0,30
0,16
0,02
129,5
133,5
137,5
141,5
145,5
149,5
153,5
mEq/l
Determinaciones de sodio
Relative Frequency
Frequency poligon
Mid point interval
Interval

CODE:

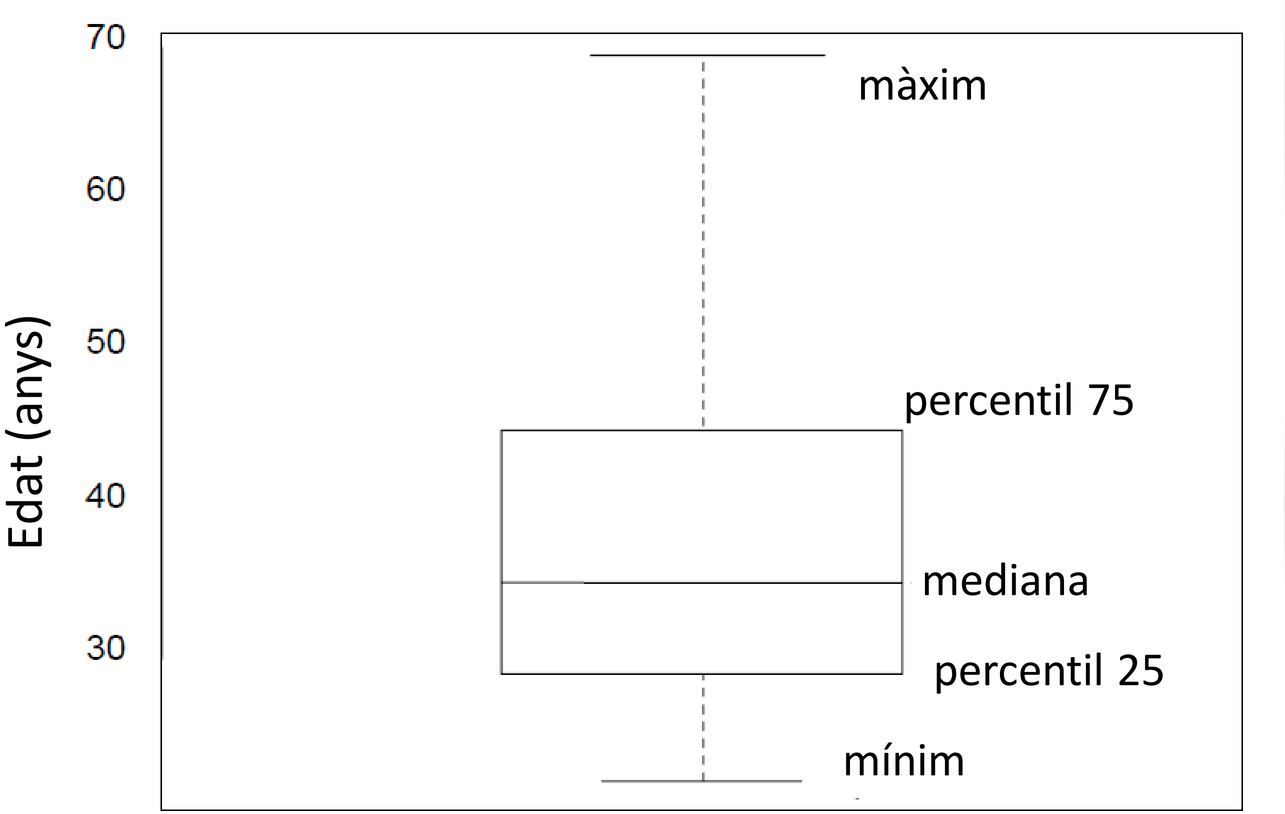
```
ggplot(data = osteo) +
  geom_histogram(mapping = aes(x = peso))
```

RESULT:



# Boxplot

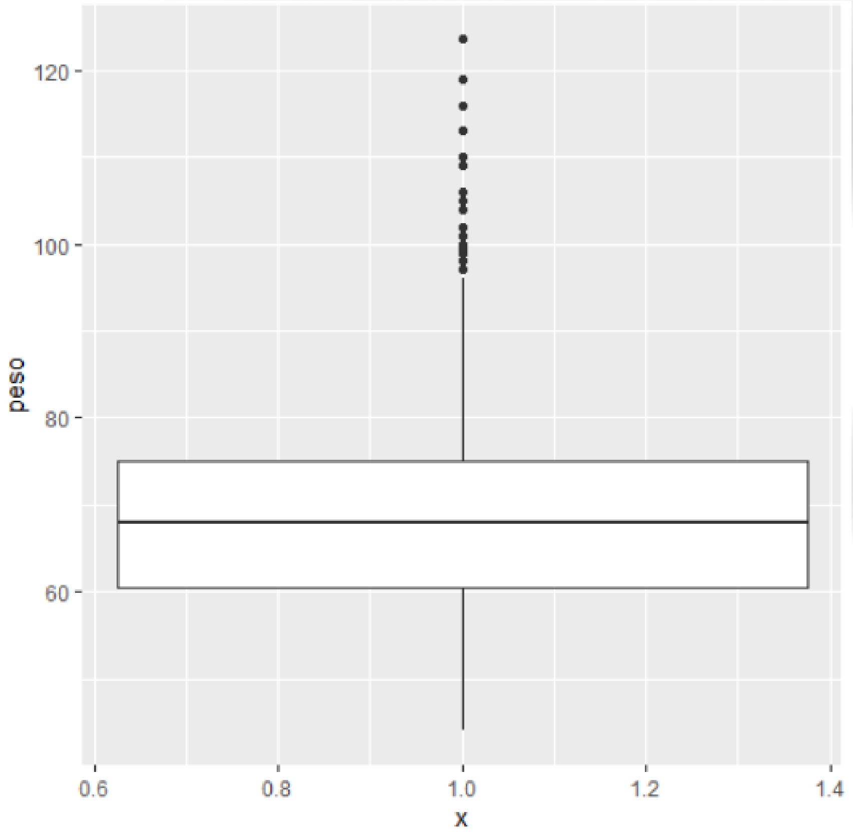
It is graphically represented the “five numbers": box are 25th and 75th percentiles, the middle line is the median (50th percentile) and the ends are the minimum and maximum values.



CODE:

```
ggplot(osteo,aes(y=peso,x=1))+
        geom_boxplot()
```

RESULT:



Vall d'Hebron
Institut de Recerca

UNITAT
D'ESTADÍSTICA I
BIOINFORMÀTICA

# Index

1. INTRODUCTION. ANALYSIS STRATEGY
2. VARIABLES CLASSIFICATION
3. SUMMARY MEASURES
    1. Measures of location/central tendency
    2. Measures of variability/dispersion
4. SUMMARY OF VARIABLES
    1. Contingency Tables
    2. Graphs
5. Examples & exercises

# Exercise

- Import diabetes.sav

BMI (Body mass index) of 149 patients →

$\mathrm{MEAN}_{\mathrm{BMI}}$ =

$\mathrm{SD}_{\mathrm{BMI}}$ =

SPB(Diastolic Pression Blood) of 149 patients →

$\mathrm{MEAN}_{\mathrm{SPB}}$ =

$\mathrm{SD}_{\mathrm{SPB}}$ =

Which variable have more variation? BMI or SPB

Calculate CV

$\mathrm{CV}_{\mathrm{BMI}}$ = %

$\mathrm{CV}_{\mathrm{SPB}}$ = %

- Create a boxplot with BMI

- Resume “MORT” variable. Create a frequency table and a barplot.