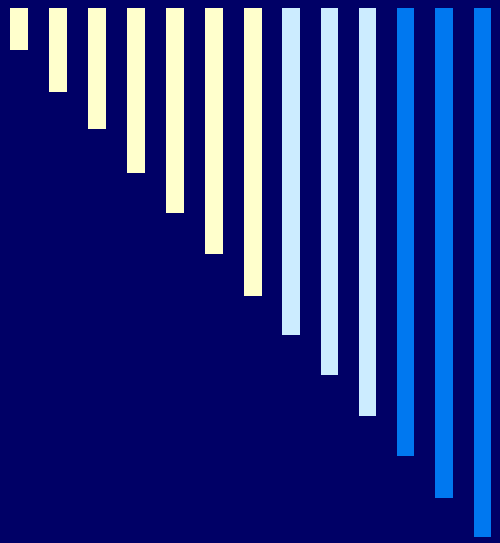


ESE

09 Data Analysis and Interpretation

Data Analysis and Interpretation

Parte 1

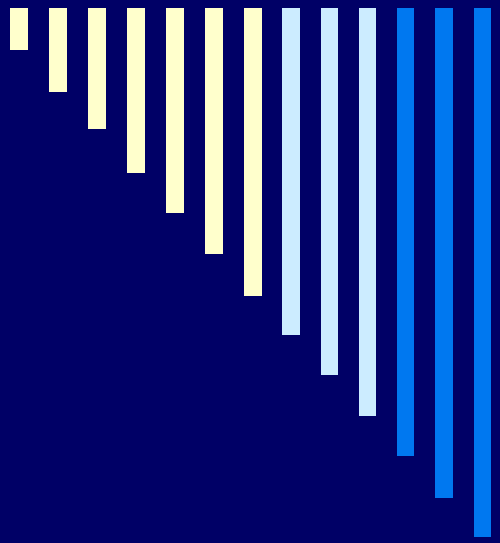
- 09.1a Descriptive statistics
  - Graphical Visualizations
  - Relevant statistics
- 09.1b Data set reduction
  - Subjects not adequate
  - Outliers

Parte 2.1

- 09.3 Statistical basis of statistical tests

Parte 2.2

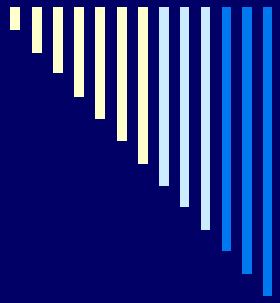
- 09.4 Hypothesis testing
  - Hypothesis test
  - Tools for Hypothesis Tests
  - Statistical errors
  - Statistical power



09.1 Data Analysis

09.1a Descriptive Statistics

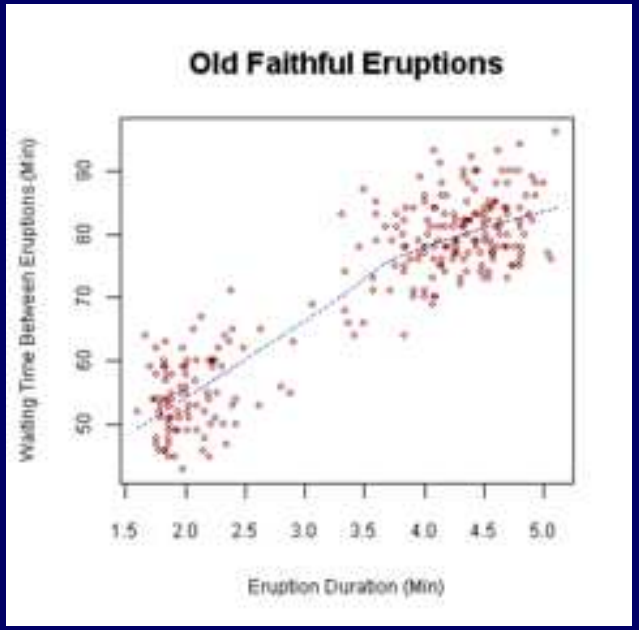
- Graphical Visualizations
- Relevant statistics



Graphical Visualization

- Histogram (No., Frequency)
- Cumulative histogram
- Pie chart
- Scatter plot
- Box plot

Scatter Plot



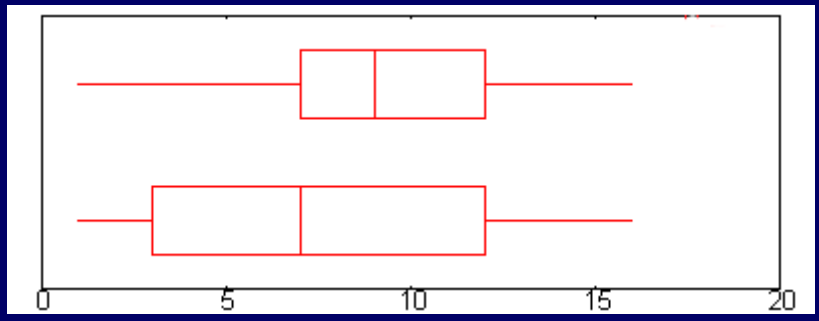
Scatter plot

It is a type of Cartesian diagram.

It illustrates (only) the degree of **correlation** (not causation) between two variables.

Best-fit can be used to study the correlation between the variables.

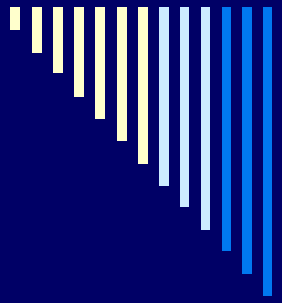
Box Plot



Box plot

Box-plot, also told **Box and Whiskers plot**.

It is a graphical representation used to describe the distribution of a sample by using simple indexes of dispersion and position.

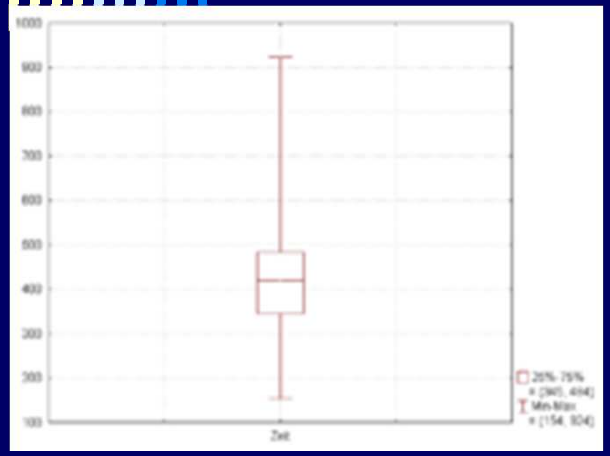


Box Plot

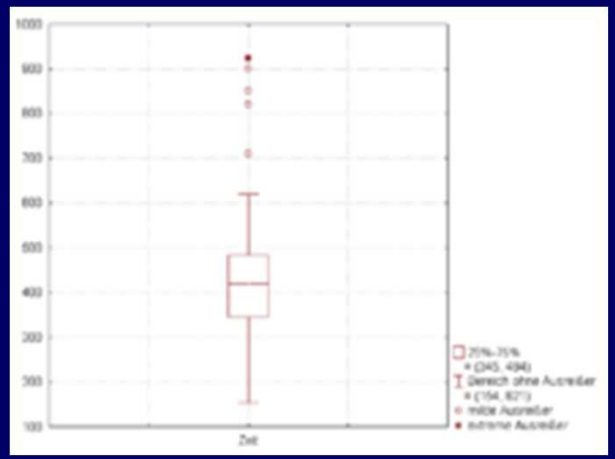
Box and Whiskers plots are uniform in their use of the box:

- the bottom and top of the box are always the first quartile and third quartile, and the band inside the box is always the second quartile (the median).
- But the ends of the whiskers can represent several possible alternative values.

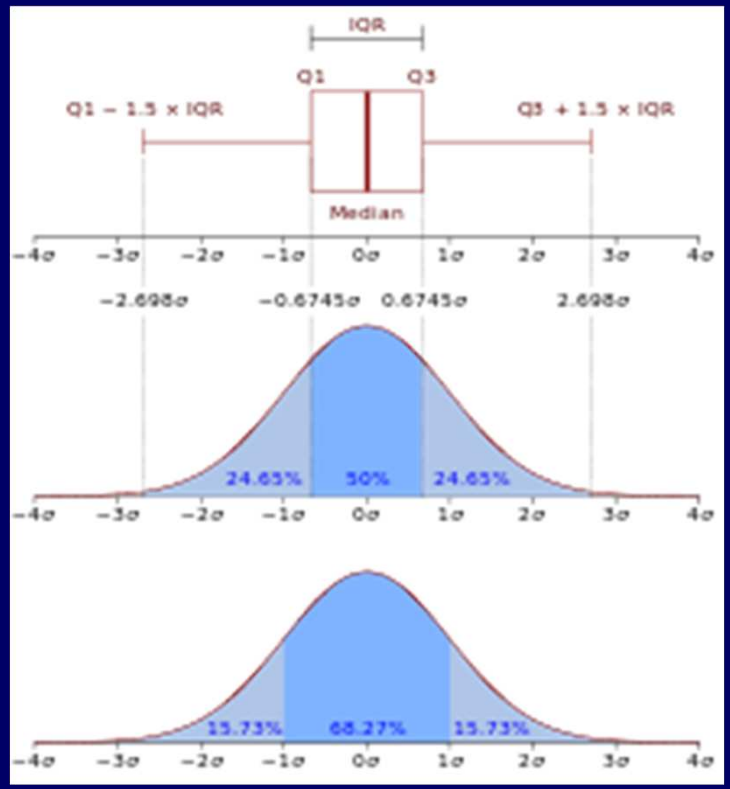
Some Alternative Values at the End of the Whiskers



Whiskers from minimum to maximum



Same Box-plot than above
Whiskers with lowest/highest datum 1.5 IQR



Whiskers with lowest/highest datum 1.5 IQR
Box-plot and a probability density function (pdf) of a Normal $N(0,1\sigma^2)$ Population

Alternative Values for the End of the Whiskers

(See figures in previous slide)

- The minimum and maximum of all of the data
- The lowest datum still within 1.5 Inter Quartile Range, IQR, of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile
- One standard deviation above and below the mean of the data
- The 9th percentile* and the 91st percentile
- The 2nd percentile and the 98th percentile.

* The value below which a given percentage of observations in a group of observations fall.

09.1b Dataset Reduction

- Subjects not adequate
  - Any statistically redundant/spare subject
  - Who did not respected the experiment rules.
- Outliers
  - Any data not included between the whiskers should be plotted as an outlier with a dot, small circle, or star, but occasionally this is not done.