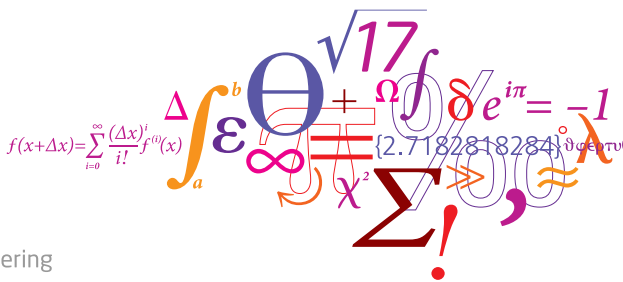
Lecture 2 - More Python, Remembering statistics



DTU Management Engineering
Department of Management Engineering

Outline

- Numpy
- Descriptive statistics

Numpy

- One of your best friends. Does a lot of algebra leg-work for you!
- Important to know:
  - Vector and matrix manipulation
  - Random number generators
  - Mathematical functions
  - Deep VS shallow copy (careful!)

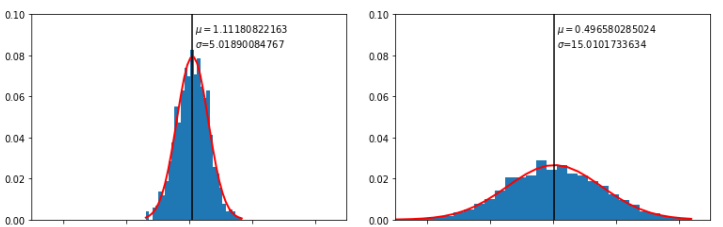
Descriptive statistics

- Mean
- Standard deviation
- Mode
- Quantiles
- Skewness
- Correlation

The mean

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

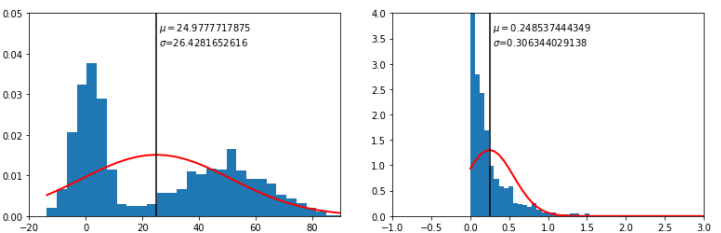
$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{(\bar{x} - x_i)^2}{N}}$$



The mean

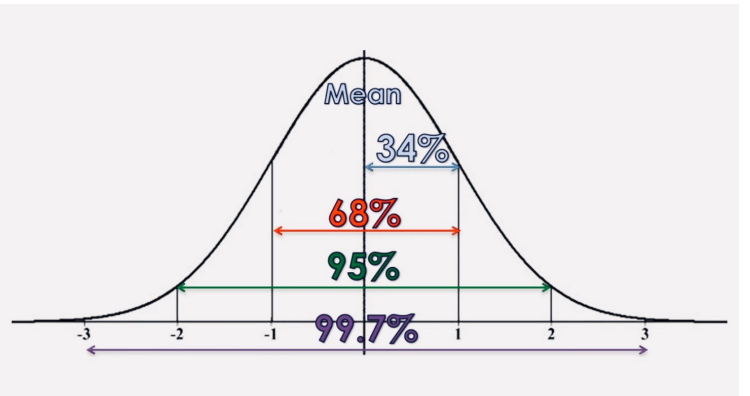
$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N}$$

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

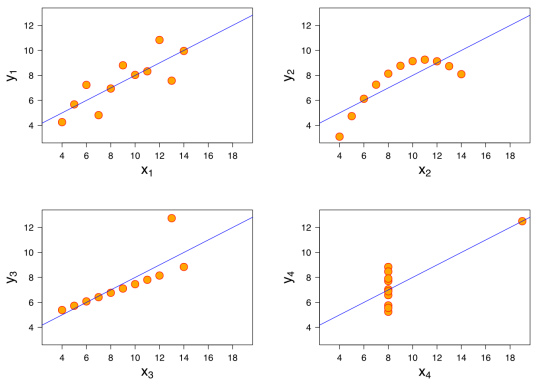


Variance and standard deviation

$$variance = \sigma^2$$

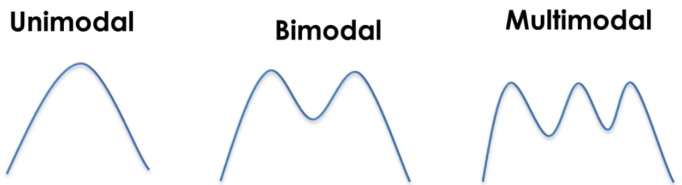


The Ancombes quartet

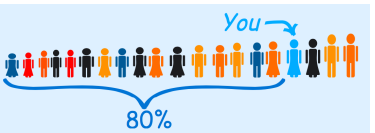


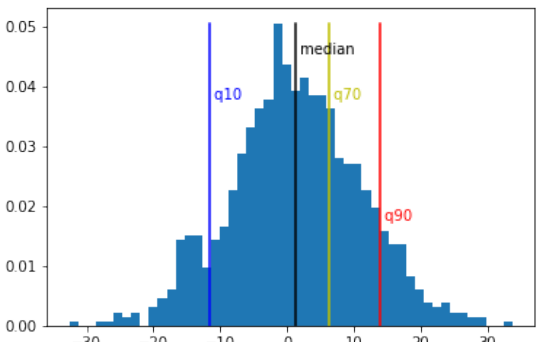
Modality

- Mode - Point with high probability

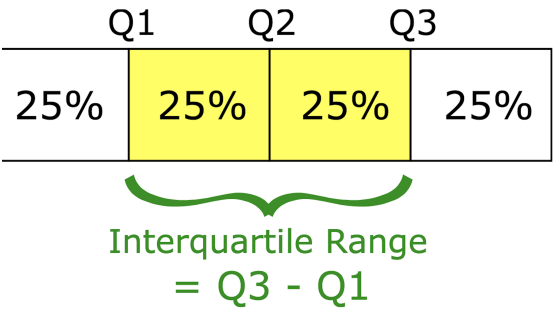


The quantiles



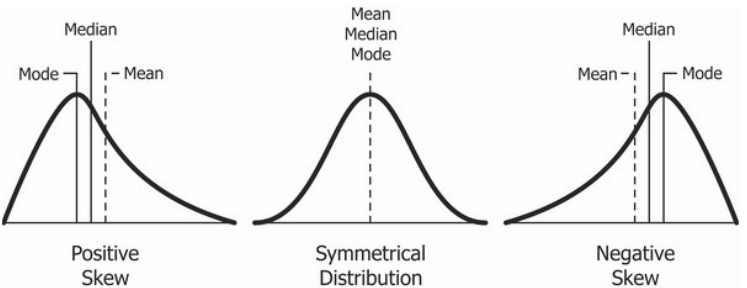


Interquartile range



- Where the majority of the data lies

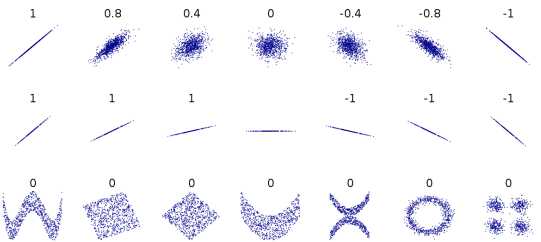
Skewness


Median
Mode
Mean
Positive Skew
Mean
Median
Mode
Symmetrical Distribution
Median
Mean
Mode
Negative Skew

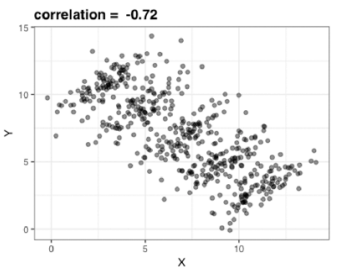
Correlation

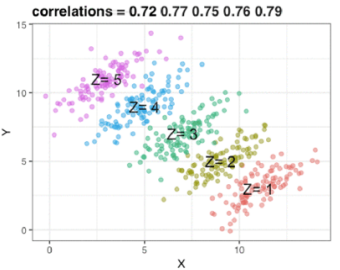
$$cov_{x,y} = \sum_{i=1}^{N} \frac{(\bar{x} - x_i)(\bar{y} - y_i)}{N}$$

$$\rho_{x,y} = \frac{cov_{x,y}}{\sigma_x \sigma_y}$$



Correlation - Simpson's paradox





More to read...

- “Numpy Tutorial Part 1 – Introduction to Arrays",
  https://www.machinelearningplus.com/python/numpy-tutorial-part1-array-python-examples/
- “Numpy Tips and Tricks part 1",
  http://arogozhnikov.github.io/2015/09/29/NumpyTipsAndTricks1.html
- “Intro to Descriptive Statistics" by Niklas Donges
  https://towardsdatascience.com/intro-to-descriptive-statistics-252e9c464ac9
- “Python Pandas - descriptive statistics",
  https://www.tutorialspoint.com/python_pandas/python_pandas_descriptive_statistics.htm
- “Python for Data Science - descriptive statistics",
  https://pythonfordatascience.org/descriptive-statistics-python/