Statistics made easy

Basic principles
Assume we are interested in getting some information about the average height of people in this room.

One basic information would be provided by the **mean**: adding individual heights and divide it by the number of people.

The mean gives us some idea about the average height, but how much does it really say?
Mean

- Which information can be derived by the mean of the class?
- Can we generalize this information?
- What sort of generalization can we really made?
  - This group is not truly a random sample.
  - It is predominant male but there is a percentage of women.
  - It is multi-cultural.
How statistics can go wrong

- Statistics are often used in “pop science”.
- The idea is that we select a representative sample of a population, perform some tests and then convert the results into numbers and perform statistics, i.e. calculating the mean.
- How can we be sure that the sample does represent the behavior of the population?
- How do we interpret the mean?
Second order Statistics

- Even if the sample is representative of the population, is the mean (first order statistics) enough to provide information?
- Think of i.e. certain countries where there is a social polarization, i.e. the rich are very rich, the poor are very poor and there is no middle class.
- This is why we introduce the meaning of variance.
Variance and Standard Deviation

- **Variance:**
  \[
  \frac{\sum_{i}^{N} (x_i - \mu)^2}{N}
  \]
  When variance is computed over a sample: denominator N-1

- **Standard Deviation:** the square root of variance
The so-called “normal” distribution
Correlation

- Second order statistics
- Assume that we have two observed variables. Correlation reveals the degree of linear relationship between them.
- The correlation coefficient may take on any value between plus and minus one.
Matlab functions

- Mean: `mean(x)`, x: a vector or matrix
- Standard deviation: `std(x)`
- Correlation coefficients: `corrcoef(x,y)`, where x and y two column vectors representing the observations
Matlab examples (1)

- \( X=[0:10] \); \text{mean}(X)
  
  \[
  \text{ans} = 5
  \]

- \text{std}(X)
  
  \[
  \text{ans} = 2.8872
  \]

- \( Y=5*X; \text{corrcoef}(x,y) \)
  
  \[
  \text{ans} = \\
  1.0000 \quad 1.0000 \\
  1.0000 \quad 1.0000
  \]
Matlab examples (2)

- Y = 10*x; corrcoef(X,Y)
  ans =
  1.0000  1.0000
  1.0000  1.0000

- Y = 10*(x.^5); corrcoef(X,Y)
  ans =
  1.0000  0.8207
  0.8207  1.0000
Matlab examples (3)

- \[ Y = 10 \times (X^{10}); \text{ corrcoef}(X,Y) \]
  
  \[
  \begin{array}{cc}
  1.0000 & -0.0000 \\
  -0.0000 & 1.0000
  \end{array}
  \]

- \[ Y = -10 \times X; \text{ corrcoef}(X,Y) \]
  
  \[
  \begin{array}{cc}
  1.0000 & -1.0000 \\
  -1.0000 & 1.0000
  \end{array}
  \]