Statistics for User Studies

A Practical Approach

This lecture is not enough!

We strongly recommend to teach yourself

There is plenty of materials in the internet. Many universities have public content on the topic. Wikipedia is a good source too.

Literature in German language

- Christel Weiß, Basiswissen Medizinische Statistik, 3.te Auflage, Springer-Verlag
- Lothar Sachs, Jürgen Hedderich, Angewandte Statistik, 12.te Auflage, Springer-Verlag

Dealing with the Raw Data

- All data have an accuracy – think about it
  - do a statement on accuracy of all measures
- All data are noisy
  - if there is noise, more data are required
  - always do a statement of the range of the data or give the standard deviation

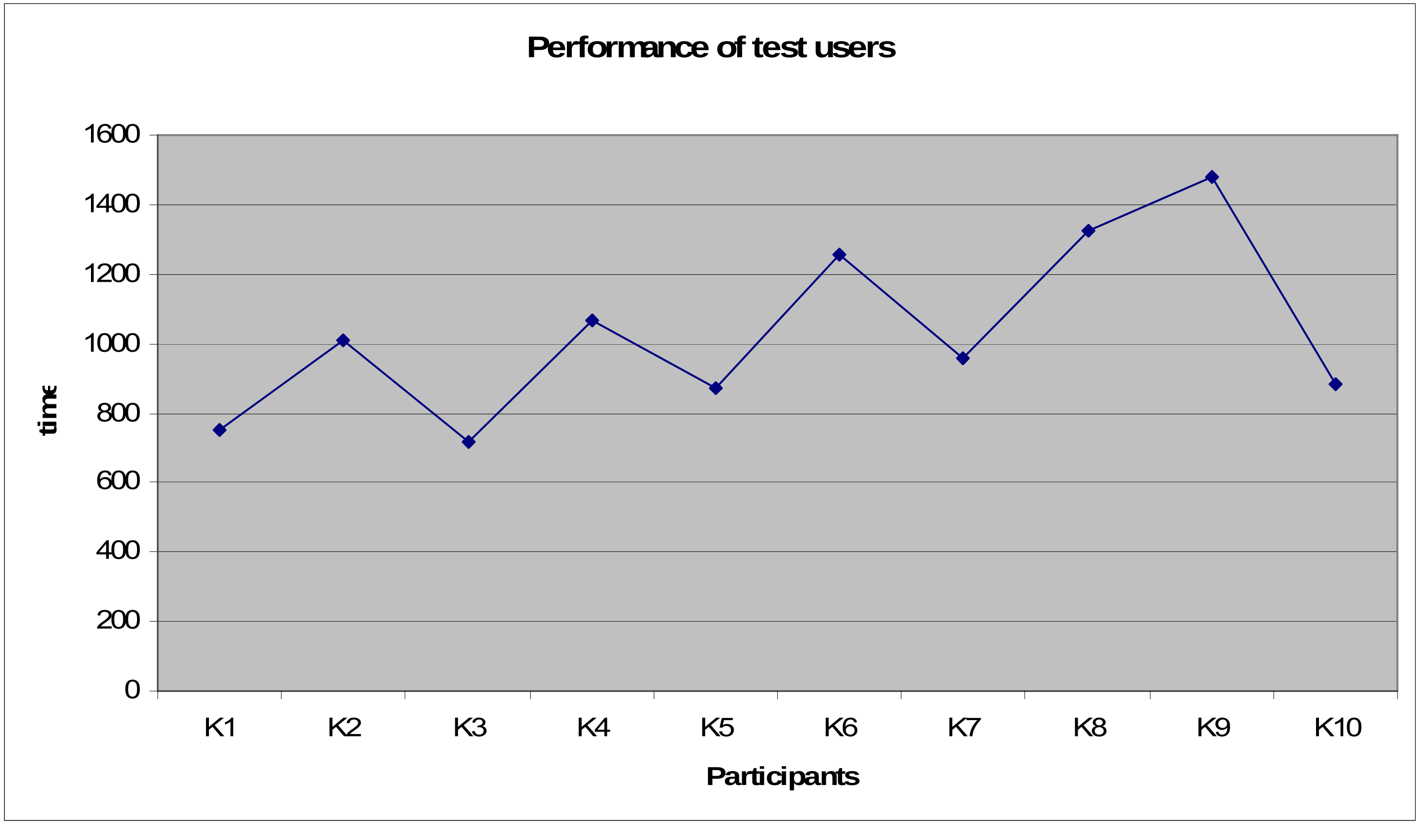
Types of Data (1)

- Categorical / Nominal Data (alternatives in non-overlapping subsets, A=B, A!=B)
  - Gender: male/female
  - Color of hair: blonde/brown/black/grey/white
- Ordinal Data (ranking/ordering A>B, A<B, A=B)
  - Marks in school 1, 2, 3, 4, 5, 6
  - Type of education school, high school, university
- Interval Scale (zero point is arbitrary, A-B)
  - Tide, Celsius scale
- Ratio Scale (fixed zero point A / B)
  - weight

Types of Variables (2)

- Discrete Data
  - distinct and separate
  - can be counted
- Continuous Data
  - any value within a finite or infinite interval
  - always have a order

Don‘t do



Performance of test users
1600
1400
1200
1000
800
600
400
200
0
time
K1
K2
K3
K4
K5
K6
K7
K8
K9
K10
Participants

Frequency Table

Data can be summarized in form of a frequency table

- well suited for discrete data
- continuous data have to be divided in groups

Example: days to answer my email

Data: 5 2 2 3 4 4 3 2 0 3 0 3 2 1 51 3 1 5 5 2 4 0 0 4 5 4 4 5 5

Days	*Frequency*	*Frequency (%)*
0	4	13%
1	3	10%
2	5	17%
3	5	17%
4	6	20%
5	7	23%

Likert Scale

Examples:

PowerPoint presentations are the best way to teach. State your opinion.

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

This year I buy a new computer.

No	Uncertain	Yes
○	○	○

Likert, R. (1932). "A Technique for the Measurement of Attitudes" Archives of Psychology 140, 55

Mean and Median

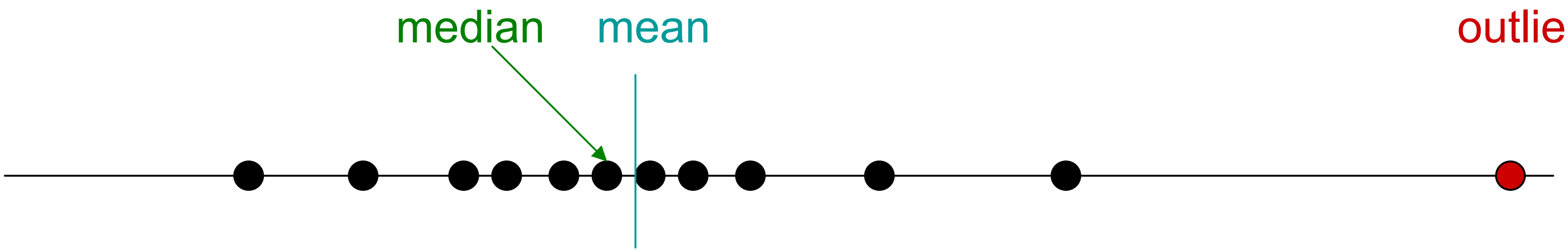
Mean

If x_1, x_2, … x_n are the data in a sample, the mean is the sum divided by n.

Median

If x_1, x_2, … x_n are the **ordered** data in a sample, the median is $x_{(n+1)/2}$ if n is odd, and $(x_{n/2} + x_{n/2+1}) / 2$ if n is even. It is the value halfway through the ordered data set.

The median is less sensitive on outliers



Variance and Standard Deviation

Variance

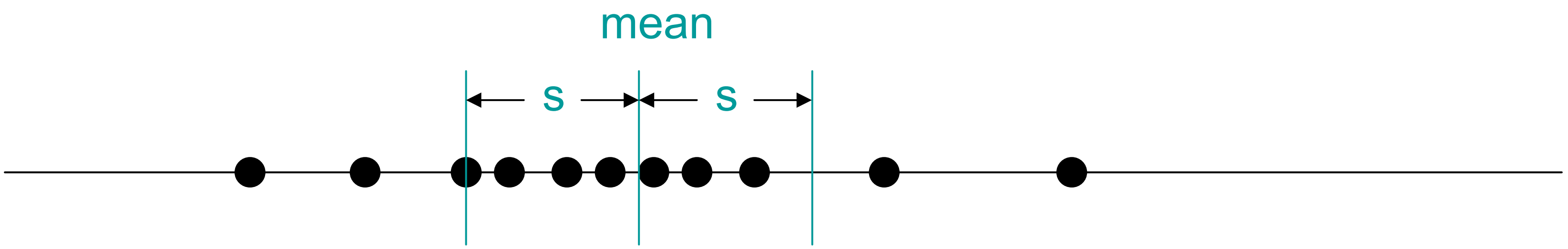
If $E(X)$ is the expected value of the random variable X then the variance $Var(X)$ is defined as: $Var(X) = E(X^2) – E(X)^2$.

If $x_1, x_2, \ldots x_n$ are the data in a sample with mean m, then the sample variance s^2 is: $s^2 = (\sum(x_i – m)^2) / (n – 1)$

The larger the variance, the more scattered the observations on average.

Standard Deviation

The standard deviation s is the square root of the variance: $s = \sqrt{Var(X)}$



Quantile, Quartile, Percentile

Quantile

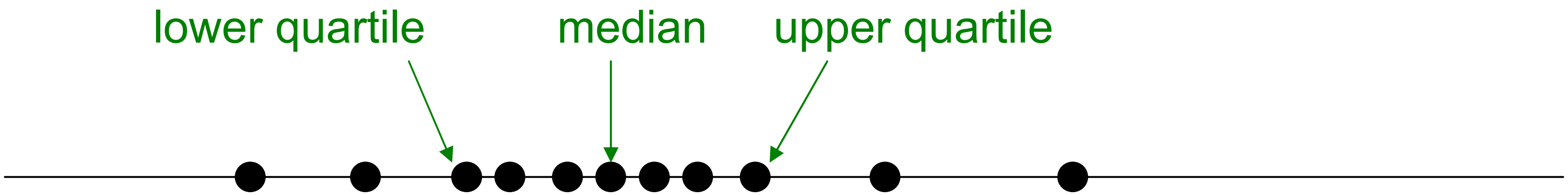
Quantiles are a set of 'cut points' that divide a sample of data into groups containing (as far as possible) equal numbers of observations.

Quartile

Quartiles are values that divide a sample of data into four groups containing (as far as possible) equal numbers of observations

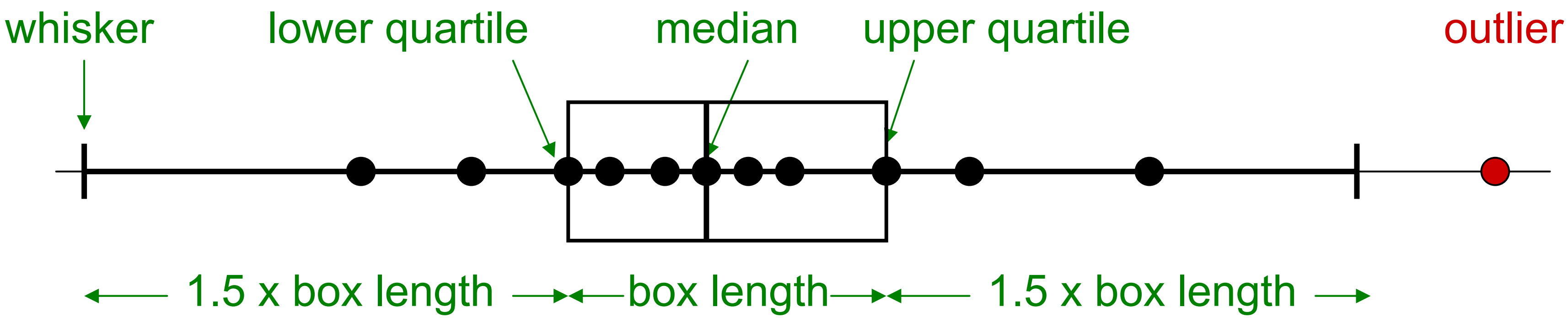
Percentile

Quartiles are values that divide a sample of data into hundred groups containing (as far as possible) equal numbers of observations



Boxplot

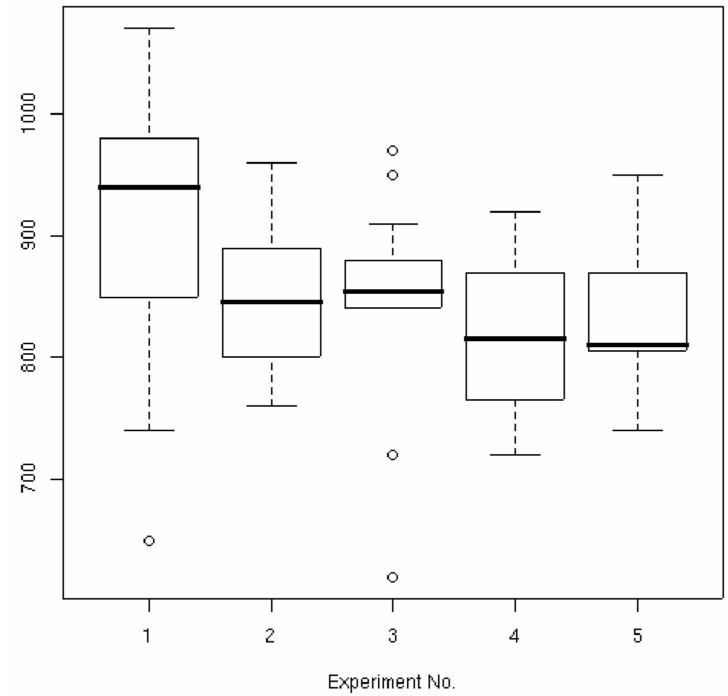
The **boxplot** is also known as **box-and-whisker diagram** or **candlestick chart**.



Outliers

Try to avoid outliers

- Improve your test equipment
- Eliminate sources of disturbances
- Repeat parts of your experiment in case of disturbance



Outliers are values that are more than1.5 box length below the lower quartile or more than1.5 box length above the upper quartile.

Some Excel Functions

MEDIAN(Matrix)

- Matrix Data row

QUARTILE(Matrix; Quartil)

- Matrix Data row
- Quartil 0 = min, 1=lower quartile, 2 = median, 3 = upper quartile, 4 = max.

QUANTIL(Matrix; Alpha)

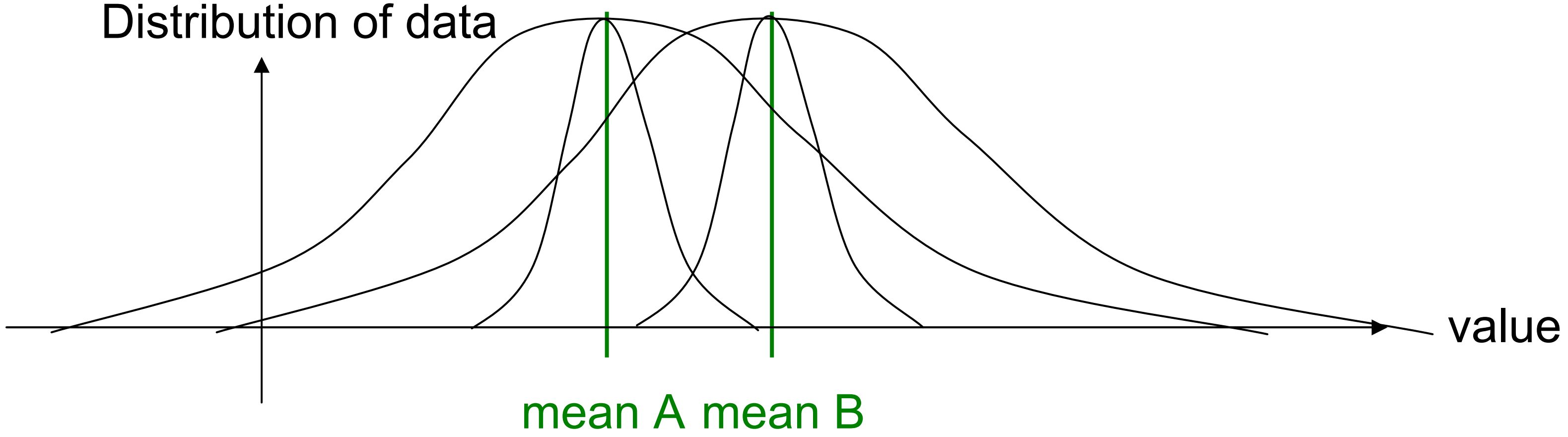
- Matrix Data row
- Alpha value form 0 to 1.

Comparing Values

A typical situation is to compare the means of two data sets.

The means are never exactly the same. But is the difference significant?

The answer depends on the difference and the variances of the data sets.



Don‘t do

With version A the test users needed 25 seconds in average to complete the task, but with version B it took only 21 seconds. Thus, our user study showed that version B is the better way to solve the task.

Is the difference significant?

Student‘s T-Test

(Mostly from wikipedia.org)

The t statistic was introduced by William Sealy Gosset for cheaply monitoring the quality of beer brews. "Student" was his pen name. Gosset was a statistician for the Guinness brewery in Dublin.

The t-test is a test of the null hypothesis that the means of two normally distributed populations are equal. The t-test gives the probability that both populations have the same mean.

Student [William Sealy Gosset] (March 1908). "The probable error of a mean". Biometrika 6 (1): 1–25.

T-Test Example in Excel

Real data from a user study

	A	B
K1	751	1097
K2	1007	971,5
K3	716	1121
K4	1066,5	1096,5
K5	871	932
K6	1256,5	926,5
K7	957	1111
K8	1327	1211,5
K9	1482	1062
K10	881	976
Mean	**1031,5**	**1050,5**

T-test	**0,8236863**

	A	B
K1	826,5	1382
K2	806	1066
K3	791	1276,5
K4	896,5	1352
K5	696	1191
K6	1121	1066
K7	891	1217
K8	1327	1412
K9	1277	1266,5
K10	656	1101
Mean	**928,8**	**1233**

T-test	**0,0020363**

Excel functions used:

=MITTELWERT(C4:C13)
=TTEST(C4:C13;D4:D13;2;1)

(function names are localized)

TTEST(…) Parameters:

- Data row 1
- Data row 2
- Ends (1 or 2)
- Type (1=paired, 2=same variance, 3=different variance)

Significance

In statistics, a result is called significant if it is unlikely to have occurred by chance.

In the case of hypothesis testing the significance level is the probability that the null hypothesis will be rejected in error when it is true.

The t-test gives the probability that both populations have the same mean. A result of 0.05 from a t-test is a 5% chance for the same mean.

Popular levels of significance are 5%, 1% and 0.1%