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

Participants

Don’t do

Performance of test users

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 5 1 3 1 5 5 2 4 0 0 4 5 4 4 5 5

<table>
<thead>
<tr>
<th>Days</th>
<th>Frequency</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>13%</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>10%</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>17%</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>17%</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>20%</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>23%</td>
</tr>
</tbody>
</table>
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

Archives of Psychology 140, 55
Mean and Median

Mean
If $x_1, x_2, \ldots, x_n$ are the data in a sample, the mean is the sum divided by $n$.

Median
If $x_1, x_2, \ldots, x_n$ 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

If $E(X)$ is the expected value of the random variable $X$ then the variance $\text{Var}(X)$ is defined as: 
$\text{Var}(X) = E(X^2) - E(X)^2$.

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

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

Standard Deviation

The standard deviation $s$ is the square root of the variance: 
$s = \sqrt{\text{Var}(X)}$
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.
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 than 1.5 box length below the lower quartile or more than 1.5 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.

### T-Test Example in Excel

Real data from a user study

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>751</td>
<td>1097</td>
</tr>
<tr>
<td>K2</td>
<td>1007</td>
<td>971,5</td>
</tr>
<tr>
<td>K3</td>
<td>716</td>
<td>1121</td>
</tr>
<tr>
<td>K4</td>
<td>1066,5</td>
<td>1096,5</td>
</tr>
<tr>
<td>K5</td>
<td>871</td>
<td>932</td>
</tr>
<tr>
<td>K6</td>
<td>1256,5</td>
<td>926,5</td>
</tr>
<tr>
<td>K7</td>
<td>957</td>
<td>1111</td>
</tr>
<tr>
<td>K8</td>
<td>1327</td>
<td>1211,5</td>
</tr>
<tr>
<td>K9</td>
<td>1482</td>
<td>1062</td>
</tr>
<tr>
<td>K10</td>
<td>881</td>
<td>976</td>
</tr>
</tbody>
</table>

**Mean** 1031,5 1050,5

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1</td>
<td>826,5</td>
<td>1382</td>
</tr>
<tr>
<td>K2</td>
<td>806</td>
<td>1066</td>
</tr>
<tr>
<td>K3</td>
<td>791</td>
<td>1276,5</td>
</tr>
<tr>
<td>K4</td>
<td>896,5</td>
<td>1352</td>
</tr>
<tr>
<td>K5</td>
<td>696</td>
<td>1191</td>
</tr>
<tr>
<td>K6</td>
<td>1121</td>
<td>1066</td>
</tr>
<tr>
<td>K7</td>
<td>891</td>
<td>1217</td>
</tr>
<tr>
<td>K8</td>
<td>1327</td>
<td>1412</td>
</tr>
<tr>
<td>K9</td>
<td>1277</td>
<td>1266,5</td>
</tr>
<tr>
<td>K10</td>
<td>656</td>
<td>1101</td>
</tr>
</tbody>
</table>

**Mean** 928,8 1233

**T-test** 0,8236863 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%