UNIVERSITÄT PASSAU
Fakultät für Informatik und Mathematik

Analysis

Homework Assignment

- Falsifiability
- Validity

Why Hypotheses?

- Provide guidance
- Avoid *Fishing for Results*
- Combines theory and empirical research
  - Derived from theory
  - Evaluated with empirical research
- Falsifiability:
  - When a hypothesis is wrong, it must be possible to disprove it

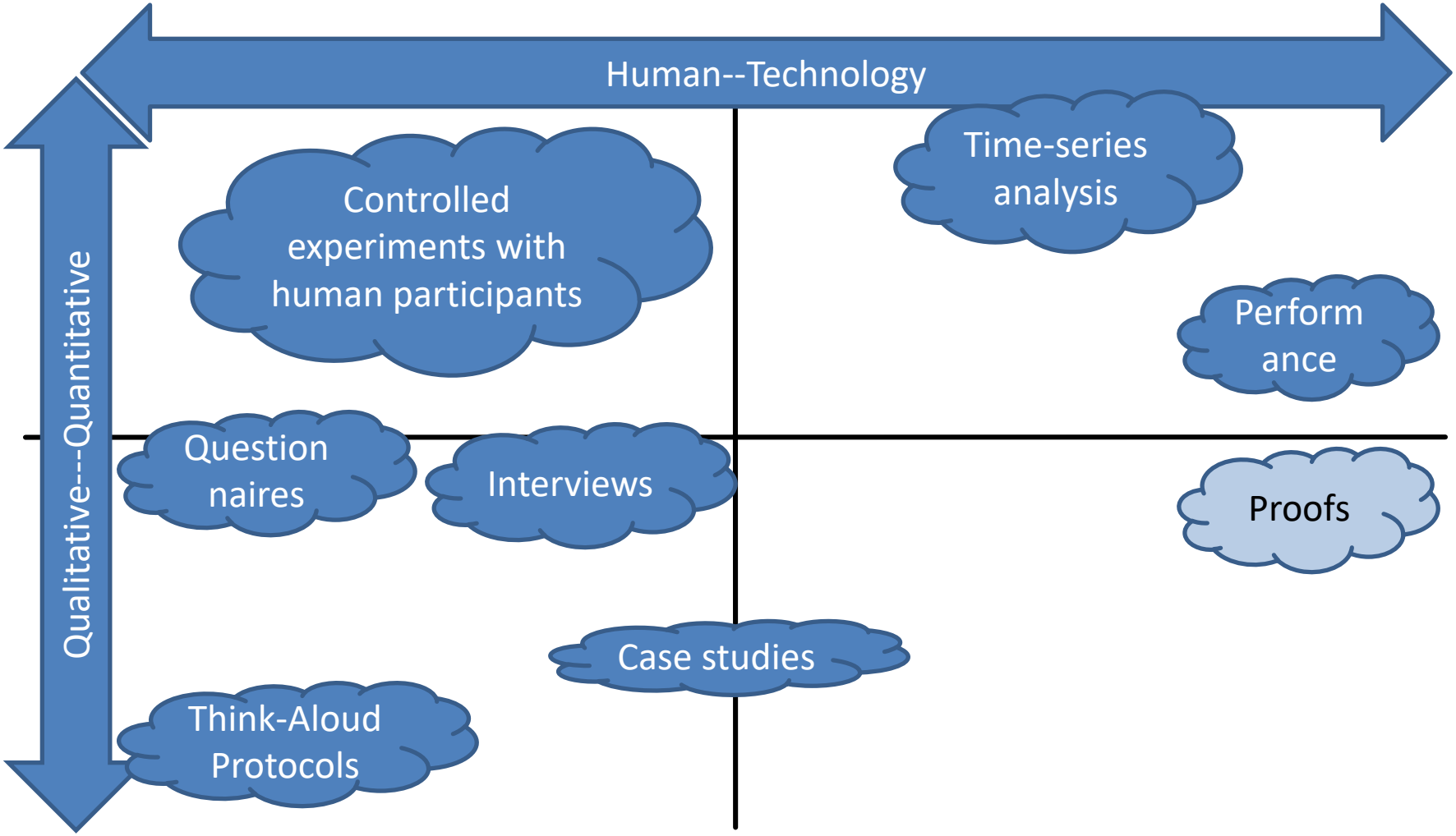
Further Kinds of Validity

- Construct validty
  - Describes how well the construct is being measured
- Statistical Conclusion Validity
  - Describes how suitable the statistical methods are

Learning Goals

- Understand the principle of standard statistical tests
- Understand the principle of correlations

Overview


Human--Technology
Qualitative---Quantitative
Controlled experiments with human participants
Time-series analysis
Perform ance
Question naires
Interviews
Proofs
Case studies
Think-Aloud Protocols

Analysis

- Descriptive statistics
- Vizualization
- Significance tests

Arithmetic Mean

- Repeat measurement
- Compute mean:

$$\bar{x}_{arithm} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- R:
  - mean(rt)

Suggested Homework

- Familiarize yourself with R (or another data-analysis tool, such as SPSS or PSPP)

Median

- Value that is in the middle
- Robust against outliers
- R:
  - median(rt)

- Even number of measurements:
  - Arithmetic mean of the two middle values
  - Use one of the two middle values

Median or Arithmetic Mean?

- Median, if:
  - Ordinal Data*
  - Few measurement values
  - Non-normal distribution
  - Outliers

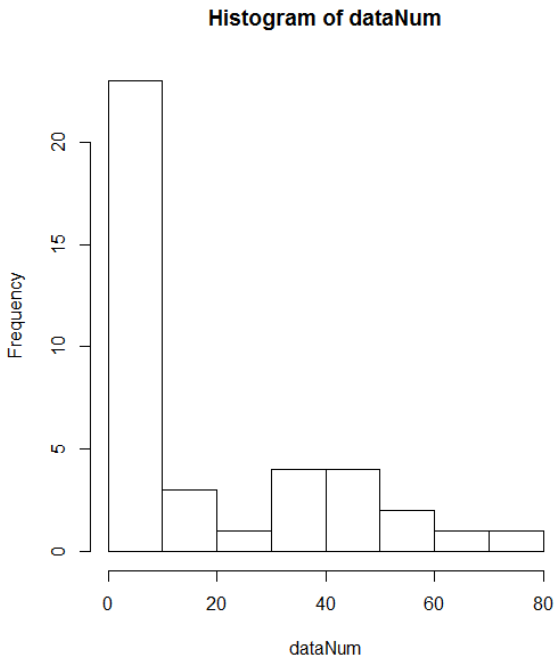
- *Scale types
  - Nominal (z.B. Gender)
  - Ordinal (z.B. Ranking)
  - Metric (z.B. Temperature, response time)

Look at the Data

- Get an overview
- Estimate distribution and outliers

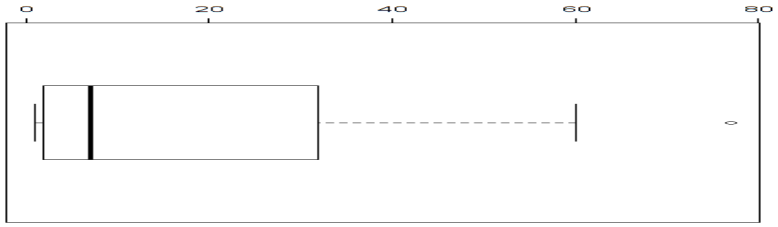
Histogram

- Frequency of values in defined buckets
  - R
    - rtNum <-as.numeric(unlist(rt))
    - hist(rt)



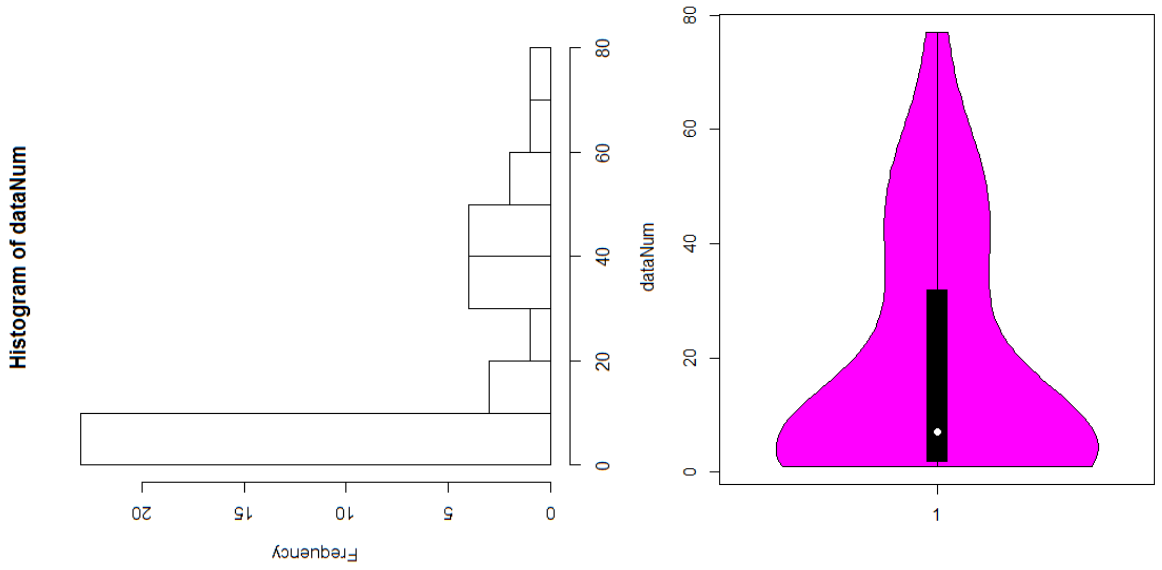
Boxplots

- Boxplot shows
  - Median as thick line
  - Quartiles as Box (50% of all values are in the box)
  - Whiskers (25% of lowest and 25% of highest values)
  - (Outliers as dots)
- Gives a hint about the distribution
- R: boxplot(rt)



Violin-Plot

- Like boxplots, only that it additionally shows the distribution
- R:
  - install.packages("vioplot")
  - library(vioplot)
  - vioplot(rtNum)



Measurement Model

- $y = \tau + \varepsilon$

- y: Observed value
- τ: True value
- ε: Error

- Population: greek letters
- Sample: german letters

Error Model

- True mean: 10

- 1 random error, influence of +/- 1
- Measurement: 9 (50%) and 11 (50%)

- 2 random errors, each +/- 1
- Measurement: 8 (25%), 10 (50%), 12 (25%)

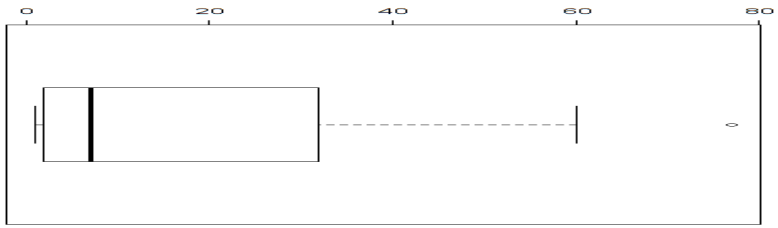
- 3 random errors, each +/- 1
- Measurement: 7 (12.5%), 9 (37.5%), 11 (37.5%), 13 (12.5%)

- N random errors, each +/- 1
- Normal distribution

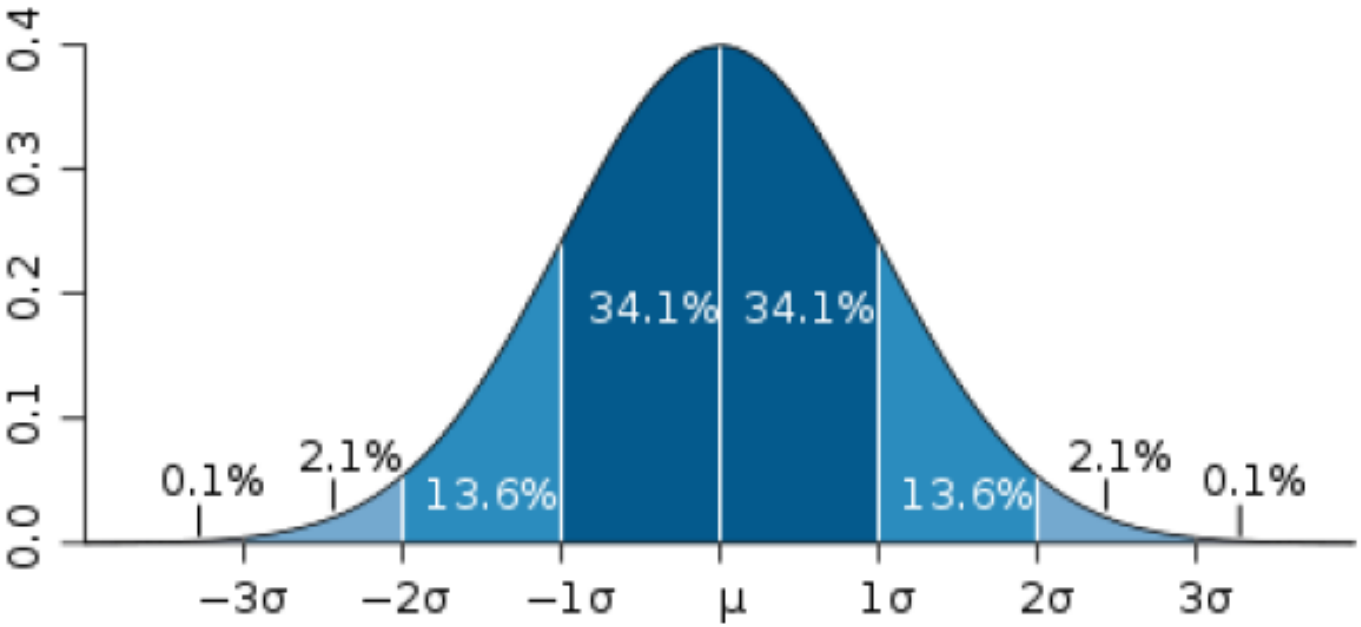
Normal Distribution

Dispersion

- Mean: 45,55
- Boxplot



Standard Deviation



Standard Deviation

- R:
  - sd(rtNum)
  - 21,55
- Mean: 45,55

- 24 –> 45,55 (34 % of measurement values)
- 45,55 –> 67,1 (34% of measurement values)

Use cases for Standard Deviation

- Define outlier
- Define giftedness
- Announce the discovery of the Higgs-Boson

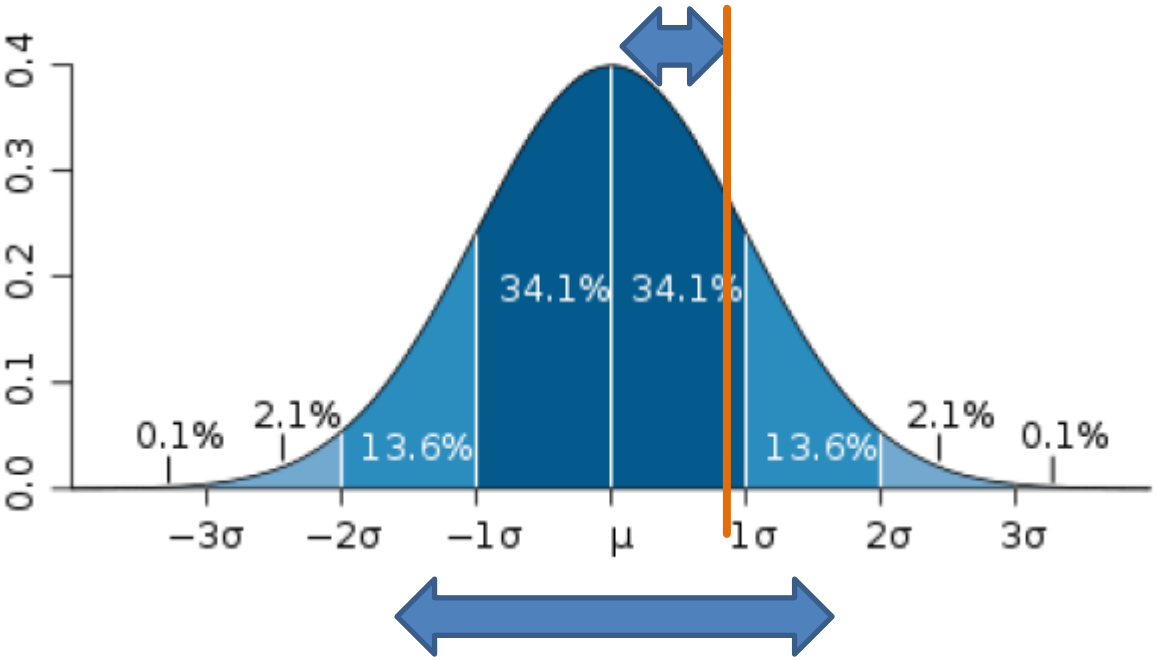
Variance

- Is the squared standard deviation

Accuracy vs. Precision

Accuracy:
Deviation of observed mean from true mean

Important when measuring response time



Precision:
Dispersion around mean

Cause of measurement errors is unclear

Random vs. Systematic Errors

- Systematic errors: Errors of the epxeriment/measurement methods
  - CPU speed: measurement during different tempertatures
  - State not resetted for second measurement
  - Low variance, or constant variance for all measurements
  - Need to be excluded during design, which requires practice and experience

  → Affect accuracy
- Random errors:
  - Cannot be controlled
  - Requires statistical methods

  → Affect precision

Significance Tests

- To evaluate whether an observed result appeared rather randomly or not

T-Test

- Designed by Student (William Sealy Gosset)
- Comparision of two measurements

Null hypothesis (H_0)	**Alternative hypothesis (H_1)**
Statistical hypotheses	
Measurements do not differ, i.e., they come from the same population	Data of both measurements are from different populations
Formal: $H_0 : \overline{x}_1 = \overline{x}_2$	Formal: $H_1 : \overline{x}_1 \neq \overline{x}_2$

T-Test: Result

- Determines probability of observed result, under the assumption that the null hypothesis is valid -> conditional probability
- If probability is smaller than:
  - 0.001 very very significant
  - 0.01 very significant
  - 0.05 typical significance level
  - 0.10 for exploraty/initial studies

  null hypothesis must be wrong
- Significance level must be defined in advance!

T-Test: Conclusion

- What does significant result mean?
- Is null hypothesis incorrect? -> No
- Is alternative hypotheses correct? -> No

- There is no evidence that the null hypothesis is valid (thus, I can only make statements about the null hypothesis)
- Writing a report:
  - Reject/could not reject null hypothesis
  - Never: Confirmation of null or alternative hypothesis

T-Test by Hand (1)

- Computation of test value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$$

Datensatz (rt.csv):
t = 1.522

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \bullet \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

T-Test by Hand (2)

- Degrees of freedom, df
  - for t-Test: $n_1 + n_2 - 2$ (in this example: 11)
- Table with t distribution (e.g., wikipedia)

$$t_{\alpha/2,df=11} = 2{,}201$$

- Comparison with calculated value (t_{emp} = 1.522)
  - is $t_{emp} > t_{\alpha/2,df=11}$?
  - no, so not significant

One-tailed vs. Two-tailed

- Two-tailed:
  - No assumption about direction of effect (e.g., which of two UIs is more usable
  - Compute half of significance level
- One-tailed:
  - Assumption that one UI is more usable
  - No need to cut significance level in half

$$t_{\alpha, df=11} = 1,796$$

T-Test: R

- t.test(rt1, rt2)
- Output:

```
Welch Two Sample t-test

data:  dataPC1 and dataPC2
t = 1.5222, df = 10.566, p-value = 0.1573
alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
 -5.095727 27.583584

sample estimates:
mean of x mean of y
 50.74243  39.49850
```

- p value: conditional probability of having observed result under the assumption that nul hypothesis is valid
- If p value is smaller than defined significance level, result is significant and null hypothesis can be rejected

Types of errors

Valid is:

Decisic

error

T-Test: Variants

- T test for independent samples:
  - Creation of samples must not be dependent on each other
  - E.g., random assignment of participants to one or the other sample
- T test for dependent samples:
  - Creation of samples depends on each other
  - E.g., in a within-subjects design, or when spouses are distributed to different samples

T-Test: Prerequisits

- Metric scale type
- Normally distributed data (e.g., Shapiro-Wilk)
- Or: sample size >= 30

Mann-Whitney-U

- Non-parametric test
- Ordinal data (or non-normal distributed metric data)
- Computation of test:

$$U = n_1 \bullet n_2 + \frac{n_1(n_1+1)}{2} - T_1$$

$$T = \sum_{i=1}^{n} r_i$$ – r_i : Ranks in the sample

χ^2-Test

- Compares frequencies
- Can answer two questions:
  - Do observed frequencies deviate from expected frequencies?
  - Do observed frequencies deviate from each other?

χ^2-Test by Hand

- Do observed values deviate from expected vaules?
- Throwing a dice at a bar (20 times)

	3	Not 3
Expected	3,33	16,66
Observed	0	20

$$\chi^2 = \sum_{i=1}^{n} \frac{(f_{oi} - f_{ei})^2}{f_{ei}} \qquad \frac{(0-3{,}33)^2}{3{,}33} + \frac{(20-16{,}66)^2}{16{,}66} = 3{,}99$$

χ^2-Test by Hand

- Compare calculated value with value from table: $\chi^2_{df=1,\alpha=.05} = 3.84$
- 3.84 < 3.99; significant

- One-tailed or two-tailed?

χ^2-Test by Hand

	G 1		G 2		G 3		
Task 1	6	4,5	18	16,5	16	19	40
Task 2	3	4,5	15	16,5	22	19	40
Sum	9		33		38		80

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} \frac{(f_{oij} - f_{eij})^2}{f_{eij}}$$

- Calculate expected frequencies (Sum of row*Sum of column/Overall sum)
- 2.22
- Degrees of freedom: (Number of rows - 1)*(Number of columns - 1)

$$\chi^2_{df=2,\alpha=.05} = 5.99$$

χ^2-Test with R

- Define a matrix:
- freqs <- matrix(c(6,3,18,15,16,22),nrow=2)
- chisq.test(freqs)

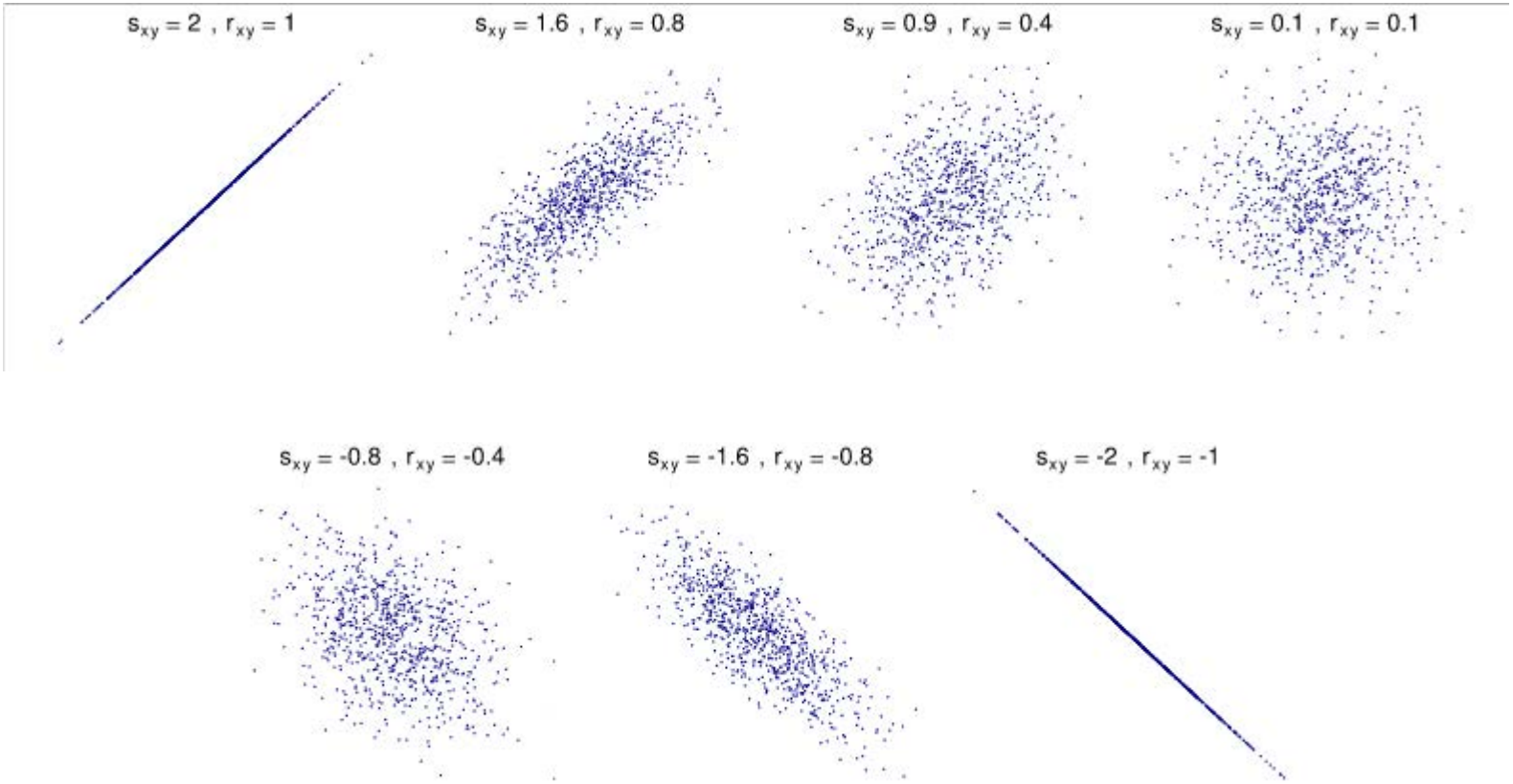
χ^2-Test - Prerequisits

- Comparison of frequencies
- Expected frequencies > 5 (Fisher‘s exact test, otherwise)
- Nominal scale type

Correlation

- Value for relationship in data
- No causality
- Values range from: -1 <= r <=+1
- $|r|$: 0.0-0.1: no relationship
- $|r|$: 0.1-0.3: weak relationship
- $|r|$: 0.3-0.5: median relationship
- $|r|$: >0.5: strong relationship

Visualisierung



Significance Tests for Correlation

- Depending on correlation there are different tests
- Null hypotheses:
  - H0: r = 0
- Significance means that correlation is most likely different from 0s

Be Careful with Small, but Significant Correlations!

- Back to the t test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)}}$$

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{\sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2}{(n_1 - 1) + (n_2 - 1)}} \bullet \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Pearson‘s r

- Metric-Metric

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \bullet (y_i - \bar{y})}{n \bullet s_x \bullet s_y}$$

Spearman - Correlation

- Rank korrelation
- Ordinal-ordinal, ordinal-metric

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n \bullet (n^2 - 1)}$$

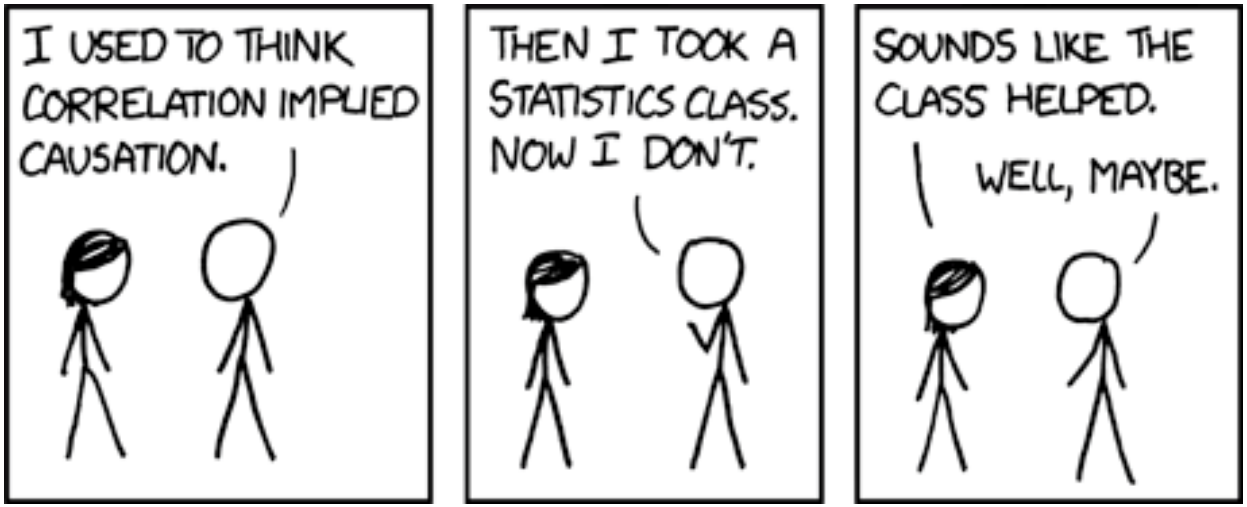
- d: difference in ranks between two observations

Contingency Coefficient

- Nominal-nominal

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$$

Correlation != Causality



Types of errors

Valid is:

Decisi

s

Multiple Testing - Example (1)

- One factor with 4 levels, pairwise comparison between all levels
- Altogether: $\binom{4}{2} = 6$

- Probability of correctly accepting null hypothesis: 0.95
- Probability of correctly accepting **two** null hypotheses: 0.95*0.95
- Probability of correctly accepting **six** null hypotheses: $0.95^6 = 0.74$

Multiple Testing – Example (2)

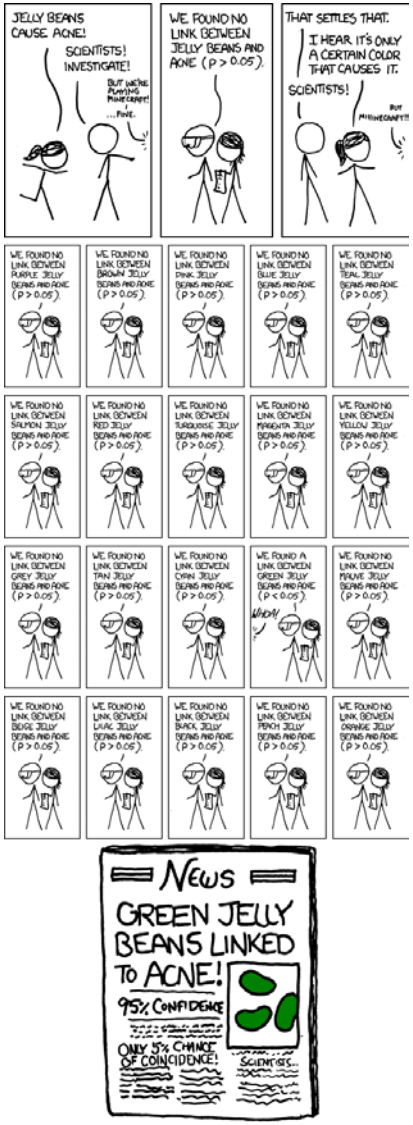
- Probability that for six tests, at least one is significant:
- $1 - 0.95^6 = 0.26$

Multiple Testing

- When conducting multiple significance tests, the significance level needs to be adapted
- Bonferoni correction:
  - t: Number of tests
  - $\alpha' = \alpha / t$
  - $\alpha / 6 = 0.0083$

- How many independent variables? How many levels?
- α-Fehler (Type-1 error), that green jelly beans cause acne: $0.95^{25} = 64\%$
- Adjusted significance level:

  $0.05/20 = 0.0025$



Analysis of Variances

- ANOVA
- Analysis to evaluate how much the variance in dependent variable can be explained by variance in independent variable
- Variance is divided to treatment and error variance

- H_0: Mean of all groups is the same
- H_1: At least to means differ

ANOVA: Steps

1. Total sum of squares
2. Treatment sum of squares
3. Error sum of squares
4. Degrees of freedom and variances
5. F value
6. Pairwise comparison

Step 1: Total Sum of Squares

- Squared difference of all measurement values from total mean
- $\overline{M}$: 4
- $SS_{tot} = 100$

A_1		A_2		A_3		A_4	
2	4	3	1	6	4	5	1
1	9	4	0	8	16	5	1
3	1	3	1	7	9	5	1
3	1	5	1	6	4	3	1
1	9	0	16	8	16	2	4

Step 2: Treatment Sum of Squares

- Part that can be attributed to 4 levels of independent variable
- Assumption: Only independent variable causes variance in result
- Squared difference of all measurement value from total mean (4)
- $SS_{treat} = 70$

A_1	A_2	A_3	A_4
2	3	6	5
1	4	8	5
3	3	7	5
3	5	6	3
1	0	8	2
2	3	7	4

A_1	A_2	A_3	A_4
2 **4**	3 **1**	7 **9**	4 **0**
2 **4**	3 **1**	7 **9**	4 **0**
2 **4**	3 **1**	7 **9**	4 **0**
2 **4**	3 **1**	7 **9**	4 **0**
2 **4**	3 **1**	7 **9**	4 **0**

Schritt 3: Error Sum of Squares

- Assumption: Differences in result per group are only caused by confounding factors
- Squared differences of individual measurement values from group mean
- $SS_{error} = 30$

A_1		A_2		A_3		A_4	
2	**0**	3	**0**	6	**1**	5	**1**
1	**1**	4	**1**	8	**1**	5	**1**
3	**1**	3	**0**	7	**0**	5	**1**
3	**1**	5	**4**	6	**1**	3	**1**
1	**1**	0	**9**	8	**1**	2	**4**
2		3		7		4	

Relationship of Sum of Squares

- $SS_{tot} = 100$
- $SS_{treat} = 70$
- $SS_{error} = 30$

- $SS_{tot} = SS_{treat} + SS_{error}$

Step 4a: Degrees of Freedom

- df_{tot}: Levels * Number of particpants per group – 1 (=19)
- df_{treat} : Number of levels – 1 (=3)
- df_{error}: Number of levels * (Number of particpants per group – 1) (=16)

- $df_{tot} = df_{treat} + df_{error}$

Step 4b: Variances

$$\hat{\sigma}^2_{treat} = \frac{SS_{treat}}{df_{treat}} = \frac{70}{3} = 23.33$$

$$\hat{\sigma}^2_{error} = \frac{SS_{error}}{df_{error}} = \frac{30}{16} = 1.88$$

$$\hat{\sigma}^2_{tot} = \frac{SS_{tot}}{df_{tot}} = \frac{100}{19} = 5.26$$

Step 5: F value

- H_0: Mean in all groups is the same
- H_0: $\hat{\sigma}^2_{treat} = \hat{\sigma}^2_{error}$

$$F = \frac{\hat{\sigma}^2_{treat}}{\hat{\sigma}^2_{error}} = \frac{23.33}{1.88} = 12.41$$

$$F_{dfnumerator=3, dfDenominator=16, \alpha=.05} = 3.24$$

- Signiciant difference, i.e.:
  At least two values differ

Step 6: Pairwise Comparisons

	A_1	A_2	A_3	A_4
Mean	2	3	7	4

$$\sum_i c_i = 0$$

$$D = 1 \cdot \overline{A_1} + 1 \cdot \overline{A_2} + 1 \cdot \overline{A_4} - 3 \cdot \overline{A_3} = 2 + 3 + 4 - 21 = -12$$

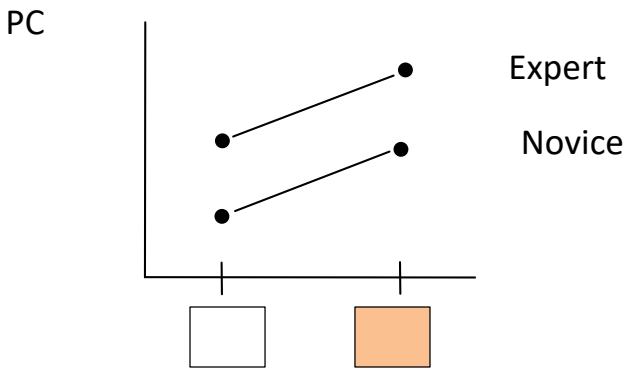
$$F = \frac{n \cdot D^2}{\sum_{i=1}^{p} c_i^{\,2} \cdot \hat{\sigma}^2_{error}} = \frac{5 \cdot 12^2}{(1+1+1+9) \cdot 1.88} = 31.91$$

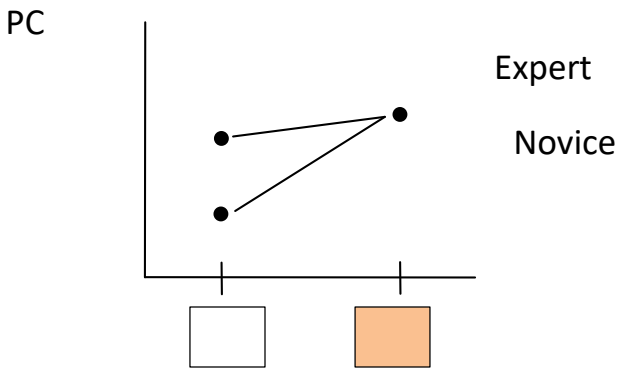
$$F_{dfNumerator=1, dfDenominator=16, \alpha=.05} = 4.49$$

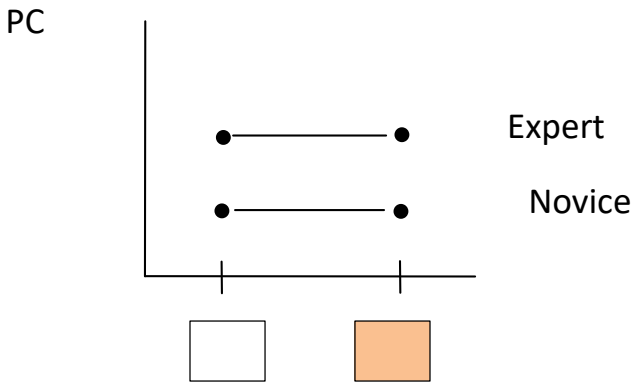
Two factorial ANOVA

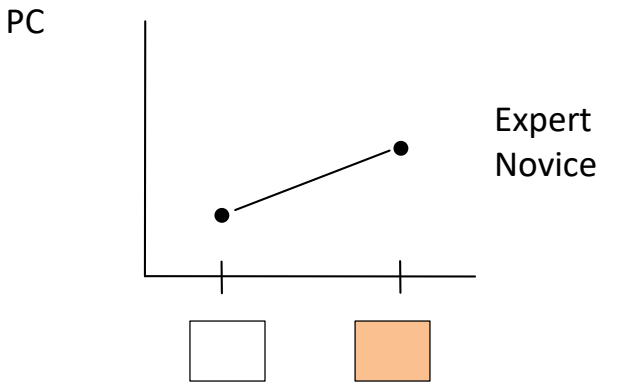
- Test whether influence of two factors is significant
- Difference to one factorial ANOVA:
  - Main effects
  - Interaction effects

Main- and Interaction Effects









Two factorial ANOVA

- Null hypotheses:
  - H_{0A}: Same mean for levels of factor A
  - H_{0B}: Same mean for levels of factor B
  - H_{0AxB}: No interaction between A und B

Step 1: Total Sum of Squares

	A_1		A_2		A_3	
B_1	22	26.01	16	0.81	13	15.21
	25	65.61	16	0.81	12	24.01
	22	26.01	16	0.81	12	24.01
	21	16.81	15	3.61	13	15.21
	22	26.01	15	3.61	12	24.01
B_2	18	1.21	19	4.41	16	0.81
	19	4.41	20	9.61	14	8.41
	17	0.01	17	0.01	16	0.81
	21	16.81	16	0.81	13	15.21
	19	4.41	16	0.81	14	8.41

- Squared difference of all values from total mean
- $\overline{M}$: 16,9
- $SS_{tot} = 348.7$

Step 2: Sum of Squares per Cell

	A_1	A_2	A_3
B_1	22.4	15.6	12.4
B_2	18.8	17.6	14.6

B_1A_1	$5*(22.4-16.9)^2$
B_1A_2	$5*(18.8-16.9)^2$
B_1A_3	$5*(15.6-16.9)^2$
B_2A_1	$5*(17.6-16.9)^2$
B_2A_2	$5*(12.4-16.9)^2$
B_2A_3	$5*(14.6-16.9)^2$

- Squared difference of group mean from total mean
- $SS_{cells} = 307.9$

Schritt 3: Error Sum of Squares

- Assumption: Differences in result per group are only caused by confounding factors
- Squared differences of individual measurement values from group mean
- SS_{error} = 40,80

	A_1		A_2		A_3	
B_1	22	0.16	16	0.16	13	0.36
	25	6.76	16	0.16	12	0.16
	22	0.16	16	0.16	12	0.16
	21	1.96	15	0.36	13	0.36
	22	0.16	15	0.36	12	0.16
Mean	22.4		15.6		12.4	
B_2	18	0.64	19	1.96	16	1.96
	19	0.04	20	5.76	14	0.36
	17	3.24	17	0.36	16	1.96
	21	4.84	16	2.56	13	2.56
	19	0.04	16	2.56	14	0.36
Mean	18.8		17.6		14.6	

Relationship Sum of Squares

- $SS_{tot} = 348.7$
- $SS_{cells} = 307.9$
- $SS_{error} = 40.8$

- $SS_{tot} = SS_{cells} + SS_{error}$

Schritt 4: Main Effects Sum of Squares

- Factor A: Squared difference of group mean from total mean
- Factor B is ignored
- $\overline{M}$: 16.9
- Weighted with number of participants per group (5) * number of levels of B (2)

- QS_A: 253.4

	A_1		A_2		A_3	
	22		16		13	
	25		16		12	
	22		16		12	
	21		15		13	
	22		15		12	
	18		19		16	
	19		20		14	
	17		17		16	
	21		16		13	
	19		16		14	
Mean	20.6	**13.69**	16.6	**0.09**	13.5	**11.56**

Schritt 4: Main Effects Sum of Squares

- Factor B (the same as for factor A)
  - Squared difference of group mean from total mean
  - Factor A is ignored
  - Weighted with number of participants per group (5) * number of levels of A (3)
  - SS_B: 0.30

Relationship of Main Effect Square Sums

- $SS_{cells} = 307.9$
- $SS_A = 253.4$
- $SS_B = 0.30$

- $SS_{cells} = SS_A + SS_B + SS_{AxB}$

Observed means

	A_1	A_2	A_3
B_1	22.4	15.6	12.4
B_2	18.8	17.6	14.6

Expected means

	A_1	A_2	A_3	
B_1	20.5	16.5	13.4	16.8
B_2	20.7	16.7	13.6	17.0
	20.6	16.6	13.5	16.9

Schritt 5: Quadratsumme Interaktionseffekt

	A_1		A_2		A_3		Mean B_j
B_1	20.5	3.61	16.5	0.81	13.4	1	16.8
B_2	20.7	3.61	16.7	0.81	13.6	1	17.0
Mean A_i	20.6		16.6		13.5		

	A_1	A_2	A_3
B_1	22.4	15.6	12.4
B_2	18.8	17.6	14.6

- Squared difference of expected and observed group means
- Weighted with number of participants per group (5)
- $QS_{AxB} = 54.2$

Relationship Sum of Squares

- $SS_{tot} = SS_{cells} + SS_{error}$
- $SS_{tot} = SS_A + SS_B + SS_{AxB} + SS_{error}$
- $348.7 = 253.4 + 0.3 + 54.2 + 40.8$

Step 6a: Degrees of Freedom

- df_{tot}: Levels (A) * Levels (B) * Number of participants per group – 1 (=29)
- df_A: Levels (A) – 1 (=2)
- df_B: Levels (B) – 1 (=1)
- df_{AxB}: (Levels (A) – 1)*(Levels (B) – 1) (=2)
- df_{error}: Levels (A) * Levels (B) * (Number of participants per group – 1) (=24)

- $df_{tot} = df_A + df_B + df_{AxB} + df_{error}$

Step 6b: Variances

$$\hat{\sigma}^2_{tot} = \frac{SS_{tot}}{df_{tot}} = \frac{348.7}{29} = 12.2$$

$$\hat{\sigma}^2_{error} = \frac{SS_{error}}{df_{error}} = \frac{40.8}{24} = 1.7$$

$$\hat{\sigma}^2_{A} = \frac{SS_{A}}{df_{A}} = \frac{253.4}{2} = 126.7$$

$$\hat{\sigma}^2_{B} = \frac{SS_{B}}{df_{B}} = \frac{0.3}{1} = 0.3$$

$$\hat{\sigma}^2_{AxB} = \frac{SS_{AxB}}{df_{AxB}} = \frac{54.2}{2} = 27.1$$

Step 7: Significance Tests

$$F_A = \frac{126.7}{1.7} = 74.53 \qquad F_{dfNumerator=2,dfDenominator=24,\alpha=.05} = 3.4$$

$$F_B = \frac{0.3}{1.7} = 0.18 \qquad F_{dfNumerator=1,dfDenominator=24,\alpha=.05} = 4.26$$

$$F_{AxB} = \frac{27.1}{1.7} = 15.94 \qquad F_{dfNumerator=2,dfDenominator=24,\alpha=.05} = 3.4$$

What does that mean?

	A_1	A_2	A_3
B_1	22	16	13
	25	16	12
	22	16	12
	21	15	13
	22	15	12
B_2	18	19	16
	19	20	14
	17	17	16
	21	16	13
	19	16	14

- Pairwise comparisons, e.g., with a t test

Effect Sizes

- ANOVA: $\eta_A^2 = \frac{SS_A}{SS_{tot}} = \frac{253.4}{348.7} = 0.73$
- Metric data, normally distributed: Cohen‘s d
- Ordinal data (or metric non-normal data): Cliff‘s delta
- Depending on concrete measure, different thresholds count for weak, medium, or strong effects
- Overview: https://www.psychometrica.de/effect_size.html

Writing a Report

- Literature:
  - *Reporting Experiments in Software Engineering*. Andreas Jedlitschka, Marcus Ciolkowski, and Dietmar Pfahl. In Shull, F., Singer, J., and Sjøberg, D.I. (Edtrs.): Advanced Topics in Empirical Software Engineering, Springer, 2007.

Introduction

- Explain context
- Motivate reader to read your paper

Background

- Special knowledge on area, e.g., type systems, program comprehension, usability heuristics
- Everything that the common computer scientist does not know
- (You can omit that)

Objective

- What is the goal of the experiment
- Where does the goal come from?
- What are the research questions/hypotheses

Variables

- Independent and dependent variable, including operationalization
- Important counfounding factors, including control techniques

Material

- Source code
- Tools
- Questionnaires
- Justification for selection

Tasks

- What were participants supposed to do?
- Why should they do this? How does it help to answer the hypotheses?

Participants

- Where do they come from?
- What characteristical values? -> Everything that might be a confounding factor, e.g., age, programming experience, gender
- Why are the participants suitable to answer the research questions/hypotheses?

Design

- What experimental design did you choose?
- Why?

Conduct

- Procedure of experiment
- Was there training/warm up?
- Why?
- Were there deviations? Why?

Analysis

- Descriptive statistics
  - Describing
  - Mean, standard deviation, box plots
- Significance tests
- Only analysis of data, no interpretation!

One-Way ANOVA with R

- http://rtutorialseries.blogspot.de/
- One-Way Omnibus ANOVA:

```
> anova(lm(Values ~ Group, dataOneWay))
Analysis of Variance Table

Response: Values
          Df Sum Sq Mean Sq F value    Pr(>F)
Group      1     60  60.000  64.444 5.503e-11 ***
Residuals 58     54   0.931
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R: Two-Way ANOVA

- Two-Way Omnibus ANOVA:
  - anova(lm(Values ~ Group * Gender, dataTwoWay))

```
> anova(lm(Values ~ Group * Gender, dataTwoWay))
Analysis of Variance Table

Response: Values
             Df Sum Sq Mean Sq F value  Pr(>F)
Group         1 60.000  60.000 67.9245 3.1e-11 ***
Gender        1  0.267   0.267  0.3019 0.58489
Group:Gender  1  4.267   4.267  4.8302 0.03212 *
Residuals    56 49.467   0.883
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R: t-Test

```
> shapiro.test(rt)

 Shapiro-Wilk normality test

data:  rt
W = 0.9472, p-value = 0.5559
```

```
> input <- read.csv("G:/work/lehre/EMCS/rt.csv", sep=";", dec=",")
> rt <- input[,'time']
> rt1 <- subset(input,group==1)[,'time']
> rt2 <- subset(input,group==2)[,'time']
> t.test(rt1, rt2)

 Welch Two Sample t-test

data:  rt1 and rt2
t = 1.5222, df = 10.566, p-value = 0.1573
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -5.095727 27.583584
sample estimates:
mean of x mean of y
 50.74243  39.49850
```

R: Mann-Whitney-U Test

```
> wilcox.test(rt1,rt2,paired=FALSE)

 Wilcoxon rank sum test

data:  rt1 and rt2
W = 31, p-value = 0.1807
alternative hypothesis: true location shift is not equal to 0
```

R: Chi^2

- http://ww2.coastal.edu/kingw/statistics/R-tutorials/independ.html

```
> row1 = c(91,90,51)                    # or col1 = c(91,150,109)
> row2 = c(150,200,155)                 # and col2 = c(90,200,198)
> row3 = c(109,198,172)                 # and col3 = c(51,155,172)
> data.table = rbind(row1,row2,row3)    # and data.table = cbind(col1,col2,col3)
> data.table
     [,1] [,2] [,3]
row1   91   90   51
row2  150  200  155
row3  109  198  172
> chisq.test(data.table)

        Pearson's Chi-squared test

data:  data.table
X-squared = 25.086, df = 4, p-value = 4.835e-05
```

R: Correlation

```
> cor.test(rt,rtTask2, method="pearson")

 Pearson's product-moment correlation

data:  rt and rtTask2
t = 4.6652, df = 11, p-value = 0.0006878
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4792664 0.9426838
sample estimates:
      cor
0.8150282
```

R: Correlation

```
> cor.test(rt,rtTask2, method="spearman")

 Spearman's rank correlation rho

data:  rt and rtTask2
S = 102.9219, p-value = 0.005786
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7172475

Warning message:
In cor.test.default(rt, rtTask2, method = "spearman") :
  Kann exakten p-Wert bei Bindungen nicht berechnen
>
```

	A	B	C	D	E
1	probCode	group	time	time2	
2	ATM	1	42,744	52	
3	BQV	1	60,1	42	
4	cno	1	30,139	40	
5	ikx	1	77,047	90	
6	KQR	1	58,231	50	
7	LOF	1	48,54	48	
8	OLCAA	1	38,396	45	
9	BTM	2	48,438	42	
10	mdp	2	48,245	55	
11	RPR	2	26,052	30	
12	TZX	2	50,436	55	
13	VND	2	30,077	40	
14	vtd	2	33,743	35	
15					

Interpretation

- What do the numbers mean?
- What do results mean for research questions/hypotheses
- Which further questions emerge?
- Come back to introduction and context

Related Work

- What have others done in this area?
- Differences/similarities to own work
- Can often help to interpret own results
- (You can omit this)

Threats to Validity

- What threatens the validity of the experiment?
- Distinction between internal and external validity
- Often also construct and stastical conclusion validity

Report

- Style guide:
  - https://www.acm.org/publications/proceedings-template
- Max. 10 pages (recommendation: at least 6)
- Submit one or two weeks before exam
- Background and related work is not necessary

Suggested Homework

- Revisit the three research papers of the first session
- Evaluate the paper