

Lecture

Empirical Model Building and Methods (Empirische Modellbildung und Methoden)

Dr. Andreas Jedlitschka

SS 2017

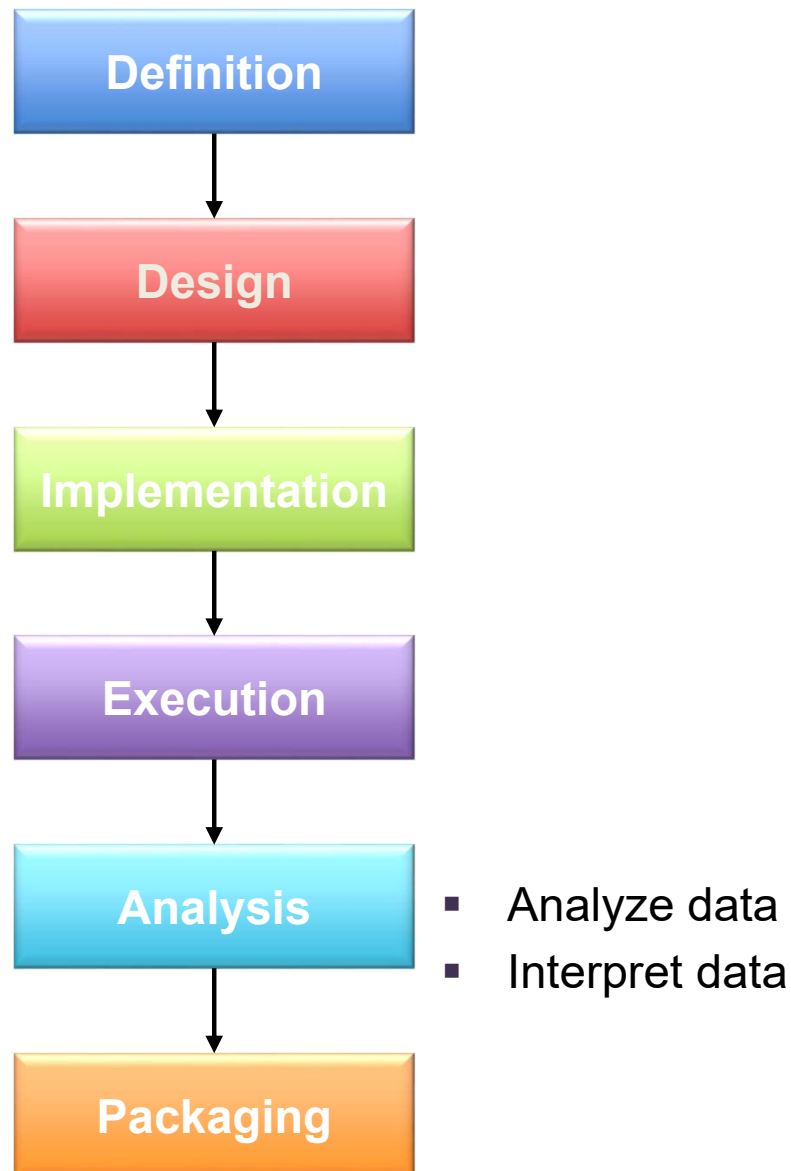
Chapter 3.6 – Analysis

Chapter objectives

At the end of this chapter, you should

- know the steps for analyzing results from empirical studies.
- understand practical issues concerning the analysis.

Empirical Study - Data analysis



3.5.1 Overview

3.5.2 Descriptive Statistics

3.5.3 Data Reduction

3.5.4 Hypothesis Testing

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- 3.6 Data analysis**
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Purpose

- Analyze collected data for
 - Describing sample
 - Testing hypothesis
 - Answering research questions

Steps

- Describe the data (Descriptive Statistics)
 - Are they sensible?
 - Find unusual values
 - Missing values, outliers and inconsistent values
 - Data reduction
- Conduct statistical analysis (Inferential Statistics)
- Interpret results

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

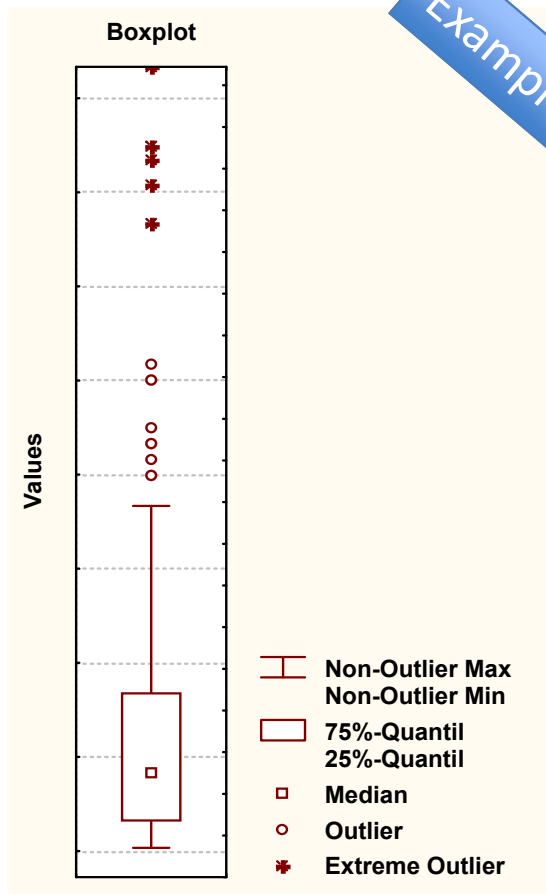
Describing data

- Frequencies
 - Absolute
 - Relative
- Measure of central tendency
 - Mode: “most often”
 - Median: “middlest value”
 - Mean: “arithmetic average”
- Measures of distribution
 - Minimum/Maximum: “smallest/largest value”
 - Interquartile range: “range containing the middle 50% of data points”
 - Variance and standard deviation
- If there is more than one variable
 - Their relationship might be of interest (bivariate analysis)
 - Contingency (cross-) table
 - Correlation: “Pearson’s r ” for continuous variables else “Spearman’s ρ ”
 - ...

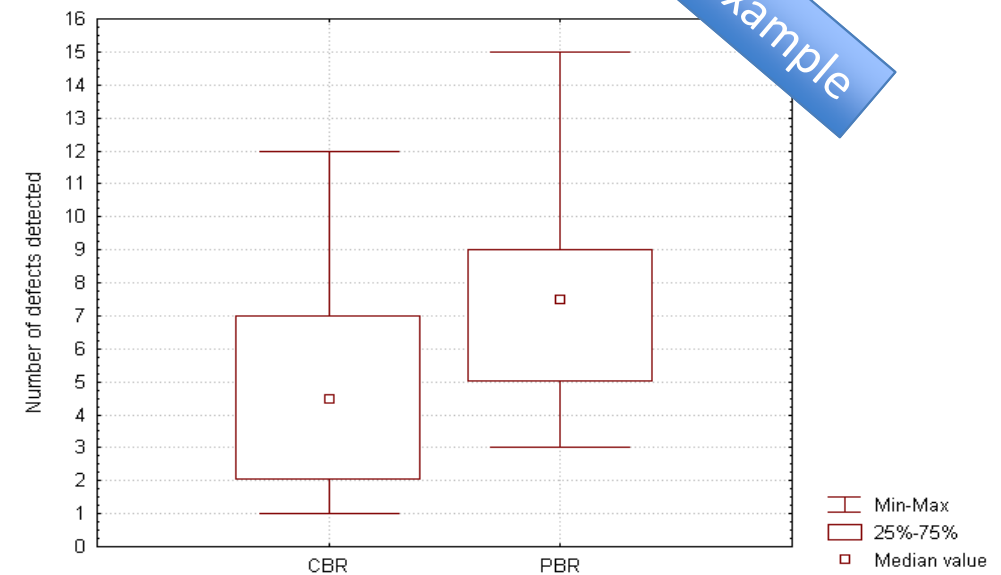
- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Visualization

e.g., Boxplot



Example

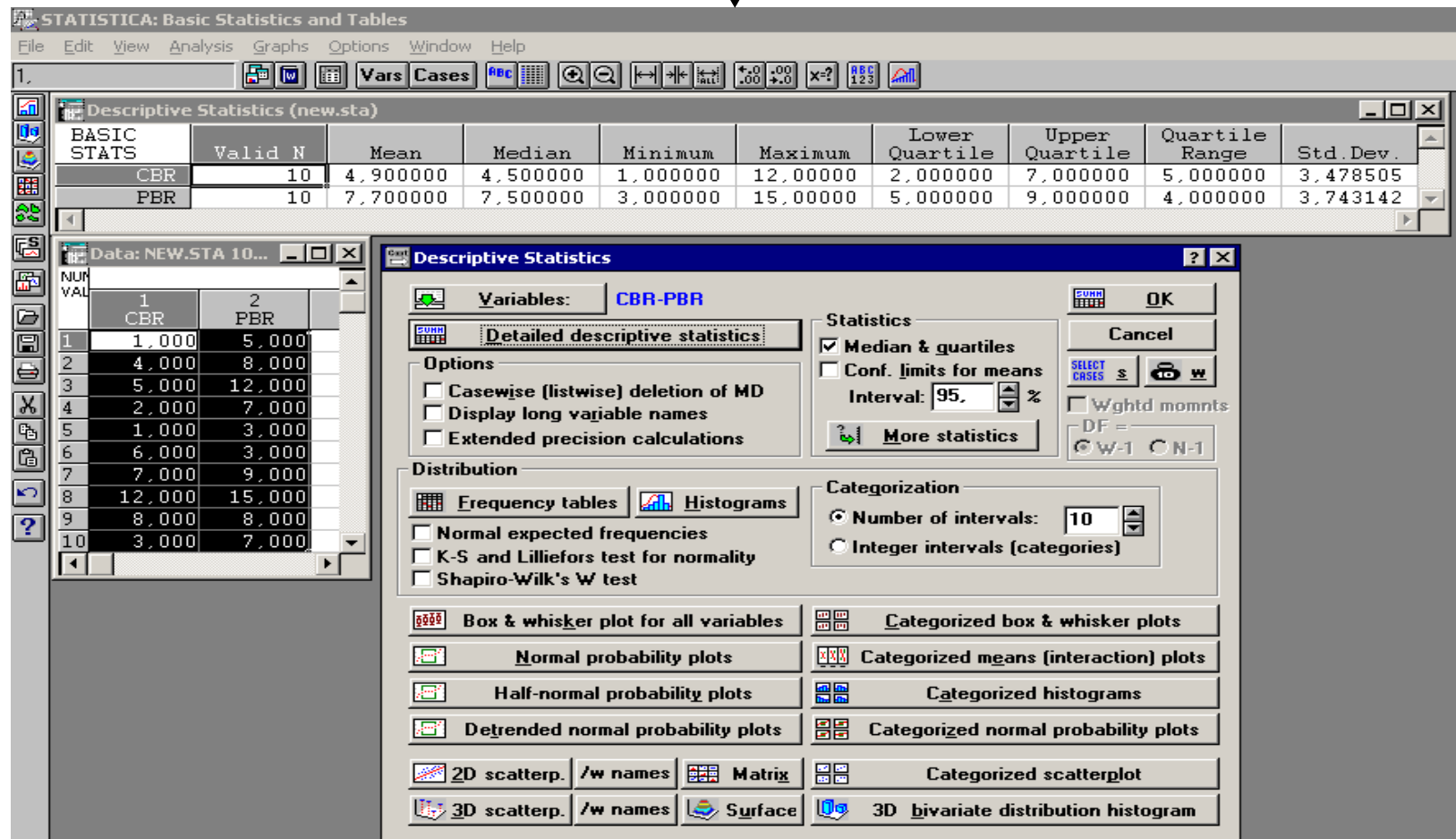


Example

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- 3.6 Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Tool support

Statistica



STATISTICA: Basic Statistics and Tables

File Edit View Analysis Graphs Options Window Help

1. Vars Cases

BASIC STATS	Valid N	Mean	Median	Minimum	Maximum	Lower Quartile	Upper Quartile	Quartile Range	Std.Dev.
CBR	10	4,900000	4,500000	1,000000	12,000000	2,000000	7,000000	5,000000	3,478505
PBR	10	7,700000	7,500000	3,000000	15,000000	5,000000	9,000000	4,000000	3,743142

Data: NEW.STA 10...

NUM VAL	1 CBR	2 PBR
1	1,000	5,000
2	4,000	8,000
3	5,000	12,000
4	2,000	7,000
5	1,000	3,000
6	6,000	3,000
7	7,000	9,000
8	12,000	15,000
9	8,000	8,000
10	3,000	7,000

Descriptive Statistics

Variables: CBR-PBR

Detailed descriptive statistics

Options

- Casewise (listwise) deletion of MD
- Display long variable names
- Extended precision calculations

Statistics

- Median & quartiles
- Conf. limits for means
- Interval: 95. %
- Wghtd momnts
- DF =
- GW-1 CN-1

Distribution

- Frequency tables
- Histograms
- Normal expected frequencies
- K-S and Lilliefors test for normality
- Shapiro-Wilk's W test

Categorization

- Number of intervals: 10
- Integer intervals (categories)

Box & whisker plot for all variables

Normal probability plots

Half-normal probability plots

Detrended normal probability plots

2D scatterp. /w names **Matrix**

3D scatterp. /w names **Surface**

Categorized box & whisker plots

Categorized means (interaction) plots

Categorized histograms

Categorized normal probability plots

Categorized scatterplot

3D bivariate distribution histogram

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Purpose of data validation

- Checking consistency and credibility
- Ensuring the correctness and completeness of the collected data
- Identifying missing variables
- Identifying outliers

# Defects Subject 1	Effort Subject 1 (minutes)	# Defects Subject 2	Effort Subject 2 (minutes)	Total Effort
13333	0	12	120	120
23	60	10	65	125
	20	23	80	80

Exceptionally high/low values

Null values reasonable?

Missing values

Missing records

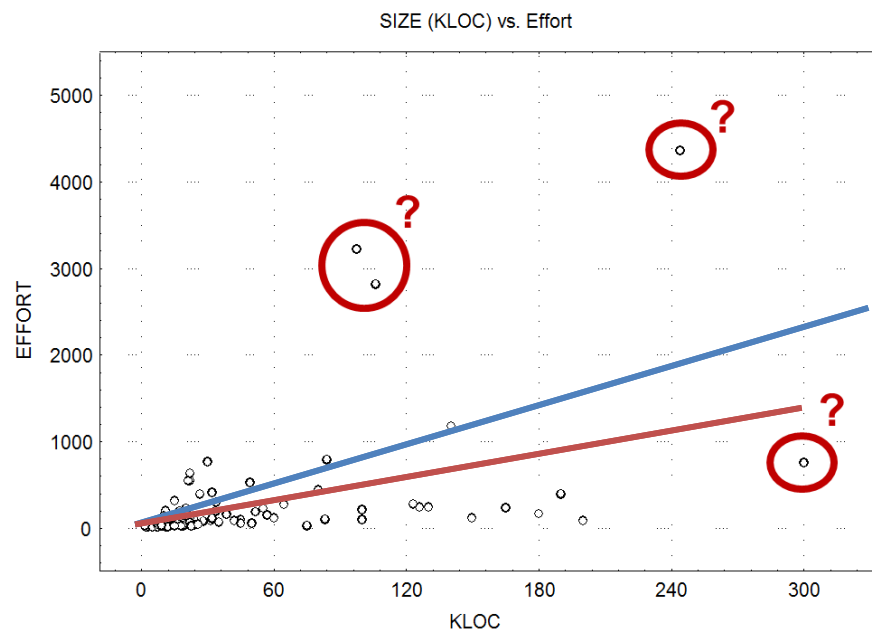
Inconsistent values

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Outliers

- Data points that are much larger or much smaller than one could expect looking at the other points

- Problems
 - They may be hinting at hidden influences
 - Comparison with the rest of data may make no sense
 - Most statistical techniques are not robust against outliers.



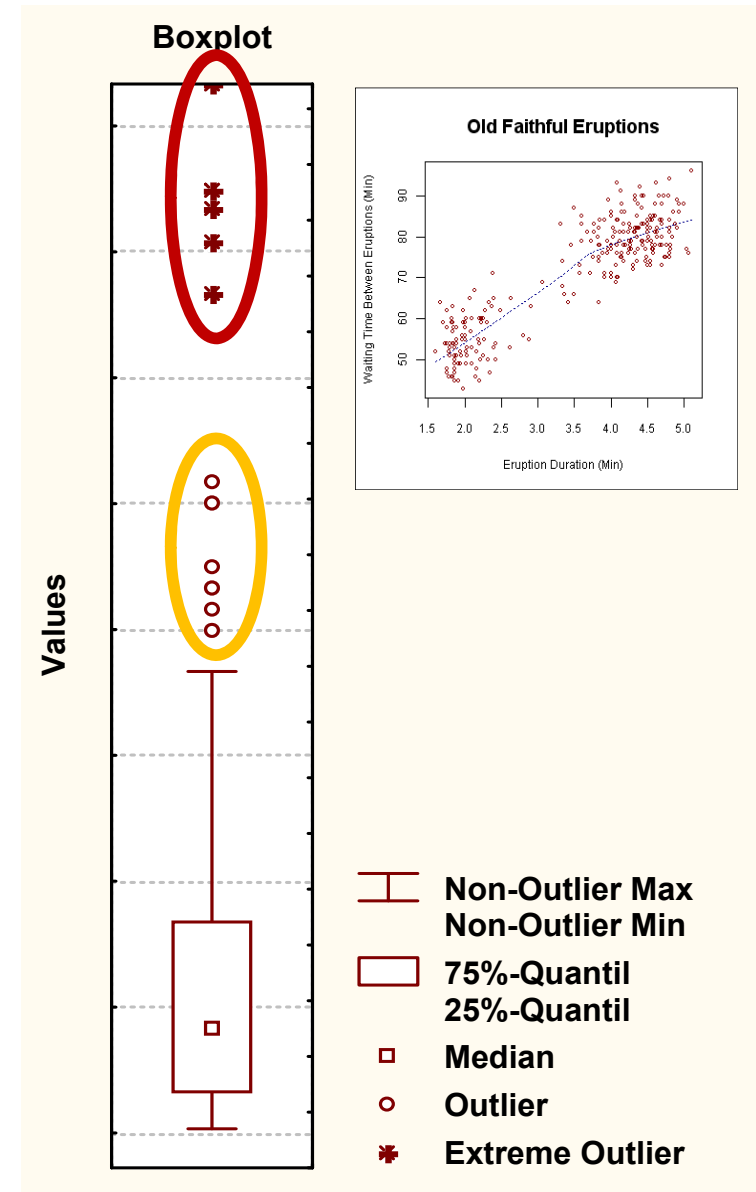
- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Identifying outliers

- Create box-plots
- Analyse if there are systematic explanations for these values
- Perform sensitivity analysis with respect to measures of central tendency and distribution (e.g., using a scatter plot)

Options for dealing with outliers

- Exclude the outlying observations
- Report two analysis (with and without outliers)
- (Transform the variable)
- Gather more observations



- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Definition

- Systematic procedure for determining whether experimental results (of a sample) provide support for the validity of a particular theory/hypothesis in the overall population.
- It implies decision making based on probabilities!

Basic questions

- Is the relationship between two or more variables due to chance?
 - Mathematical Statistics
- If chance can be excluded, what does it mean?
 - Interpretation

Approach (Falsification principle)

- “Opposite-of-what-you-predict” reasoning
- Drawing conclusions by
 - “identifying the probability of rejecting the inverse hypothesis” (e.g., > 0.95) or
 - “of not rejecting the inverse hypothesis although the hypothesis is true” (e.g., < 0.05)

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Logic of significance test: Example (1/2)

- Research hypotheses
 - H2: Individuals applying each PBR perspective (PERSP) respectively perform better than individuals applying Ad-hoc reading (N) with respect to their mean defect detection rate (DDR).
 - $PERSP = \{\text{tester perspective (T), user perspective (U), designer perspective (D)}\}$
 - $DDR = \text{number of defects found by individual} / \text{total number of defects in the document}$

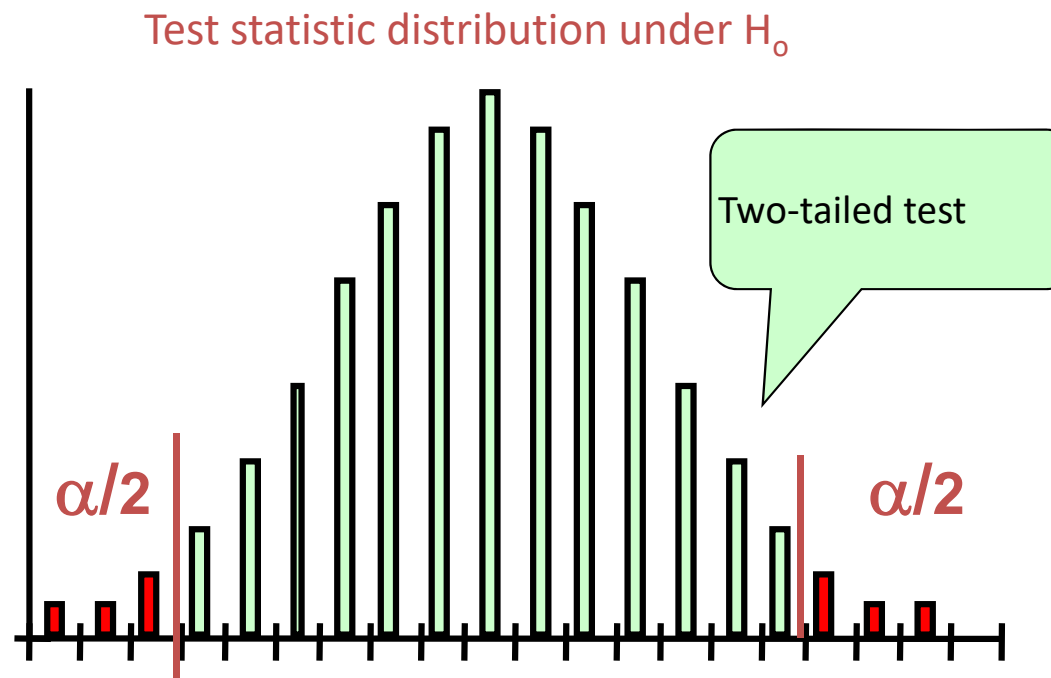
- Statistical hypotheses
 - Null hypothesis
 - Individuals applying ad-hoc reading perform better than individuals applying each PBR perspective
 - $H_{02} = DDR(PBR\ PERSP) \leq DDR(N)$

 - Alternative hypothesis
 - $H_{12} = DDR(PBR\ PERSP) > DDR(N)$

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Logic of significance test: Example (2/2)

- Statistical hypotheses
- alpha (α) is typically 0.05 (better 0.01 or smaller but rarely used)



- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Procedure (1/2)

1. Formulate the alternative and null hypothesis
2. Select statistical test (as defined during design) now considering data distribution
 - Normal distribution → Parametric tests
 - Non-normal or ordinal/nominal distribution → Non-parametric tests
3. Use significance level (α -value) and perform power analysis
 - α conventionally 0.05 or 0.01
 - Power = $1 - \beta$ (β conventionally 0.2)
 - Determine optimal sample size based on α , effect size and power

		In the population ...	
		H_0 is true	H_0 is false
Decision	H_0 is <u>not</u> rejected	Correct outcome True negative	Type II error (β) False negative
	H_0 is rejected	Type I error (α) False positive	Correct outcome True positive

Statistical Test Selection (1/2)

depends on further assumptions

What do you want to do?	Number of variables	Subjects in condition	Parametric Test	Non parametric Test
Differences between conditions	One variable: two treatments	Independent	Independent t-test	Mann-Whitney U test
		Dependent	Paired t-test	Wilcoxon matched pairs test
	One variable: > 2 treatments	Independent	One factor independent ANOVA	Kruskal-Wallis-One way ANOVA
		Dependent	One factor repeated measures ANOVA	Friedman ANOVA
	Two or more treatments	Independent/Dependent	Variants of ANOVA-Analysis	
Compare frequency counts (in categories)				Chi-square
Correlate variables	Two variables		Pearson's r	Spearman's r
	More than two variables		Multiple correlation R	

Statistical tests exist for different purposes, with different preconditions, and different power!

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Statistical Test Selection (2/2)

- Each test makes a number of assumptions
 - Experimental design
 - Distribution of data
 - Outliers
- Be aware that your data set fulfills the assumptions.
- Statistical testing is just the means to an end not an end in itself.
- More difficult than running statistical tests
 - Interpretation of the results.
 - What does the results mean?

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Statistical Tools

- SPSS (<http://www.spss.com/>)
- Statistica (<http://www.statsoftinc.com>)
- R Package (<http://www.r-project.org/>, Open source)
- JMP (<http://www.jmpdiscovery.com/>)
- S-Plus (<http://www.splus.mathsoft.com>)
- Minitab (www.minitab.com/)
- Excel
- ...

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

Interpretation

- If null-hypothesis is rejected
 - There is an effect.
 - Calculate „real“ effect size.
 - Is effect in concordance with theory?
 - Do you need to modify the theory?

- If null-hypothesis is not rejected
 - It is not possible to conclude there is no effect!
 - There is not sufficient evidence to accept there is an effect.
 - Discuss descriptive statistics.
 - Calculate „real“ effect size. Determine sample size required to reject null hypothesis, if effect is in „right“ direction.
 - **Formulate new hypothesis**

		In the population ...	
		H_0 is true	H_0 is false
Decision	H_0 is not rejected	Correct outcome True negative	Type II error (β) False negative
	H_0 is rejected	Type I error (α) False positive	Correct outcome True positive

Discuss and interpret the results in the context of your “theory”!

- 3.1 Introduction
- 3.2 Definition
- 3.3 Design
- 3.4 Implementation
- 3.5 Execution
- Data analysis
 - 3.6.1 Overview
 - 3.6.2 Descriptive Statistics
 - 3.6.3 Data Reduction
 - 3.6.4 Hypothesis Testing
- 3.7 Packaging

- **Wohlin, Runeson, Höst, Ohlsson, Regnell, Wesslén (2012). Experimentation in Software Engineering, Springer**
- **J. Bortz, and N. Döring (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4 Auflage). Berlin: Springer Verlag.**
- **N. Juristo and A. Moreno. (2001). Basics of Software Engineering Experimentation, Kluwer Academic Publishers.**
- **Jedlitschka, A., Ciolkowski, M., Pfahl, D. (2008). Reporting Controlled Experiments in Software Engineering. In: Shull, F., Singer, J., Sjoberg, D.I. (Eds.). Guide to Advanced Empirical Software Engineering. Springer.**