FACHBEREICH INFORMATIK

Lecture
Empirical Model Building and Methods
(Empirische Modellbildung und Methoden)

Prof. Dr. Dr. h.c. Dieter Rombach
Dr. Andreas Jedlitschka

SS 2015

Chapter 3.6 – Analysis

Fraunhofer IESE

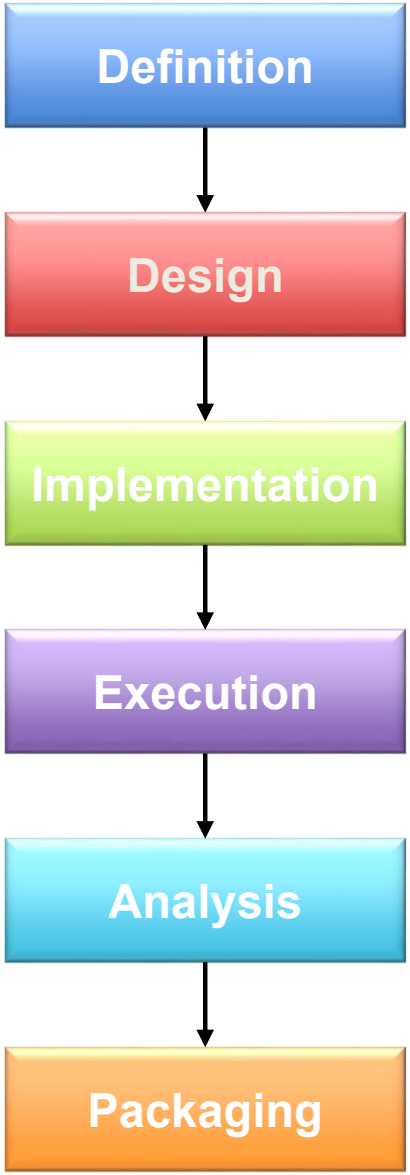
AG·SE

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



- Analyze data
- Interpret data

Empirical Study - Data analysis

3.5.1 Overview

3.5.2 Descriptive Statistics

3.5.3 Data Reduction

3.5.4 Hypothesis Testing

Overview

Purpose

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

Steps

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

Descriptive Statistics

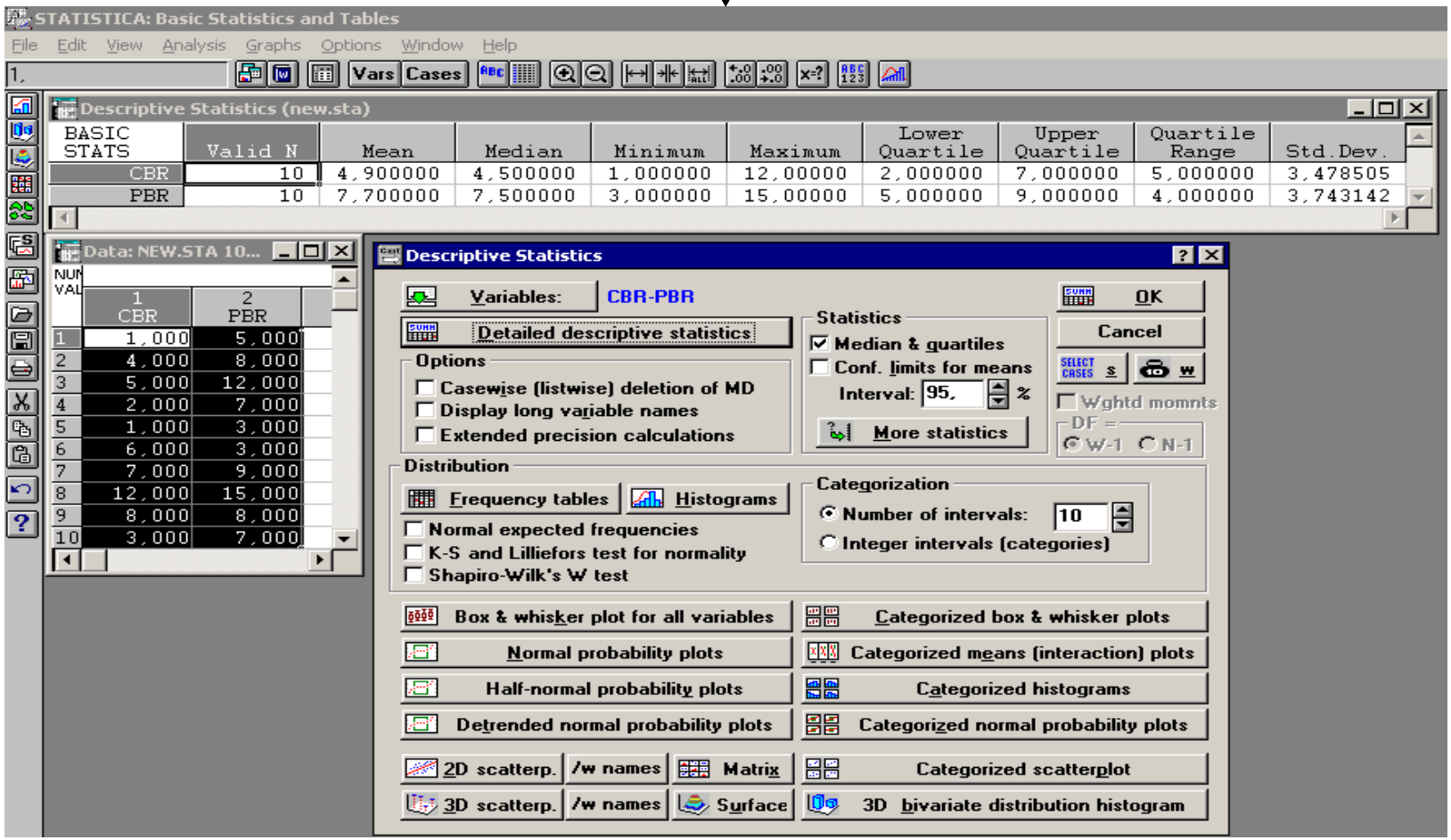
Exploring and describing data

- 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”
  - Standard deviation

Descriptive Statistics

Tool support

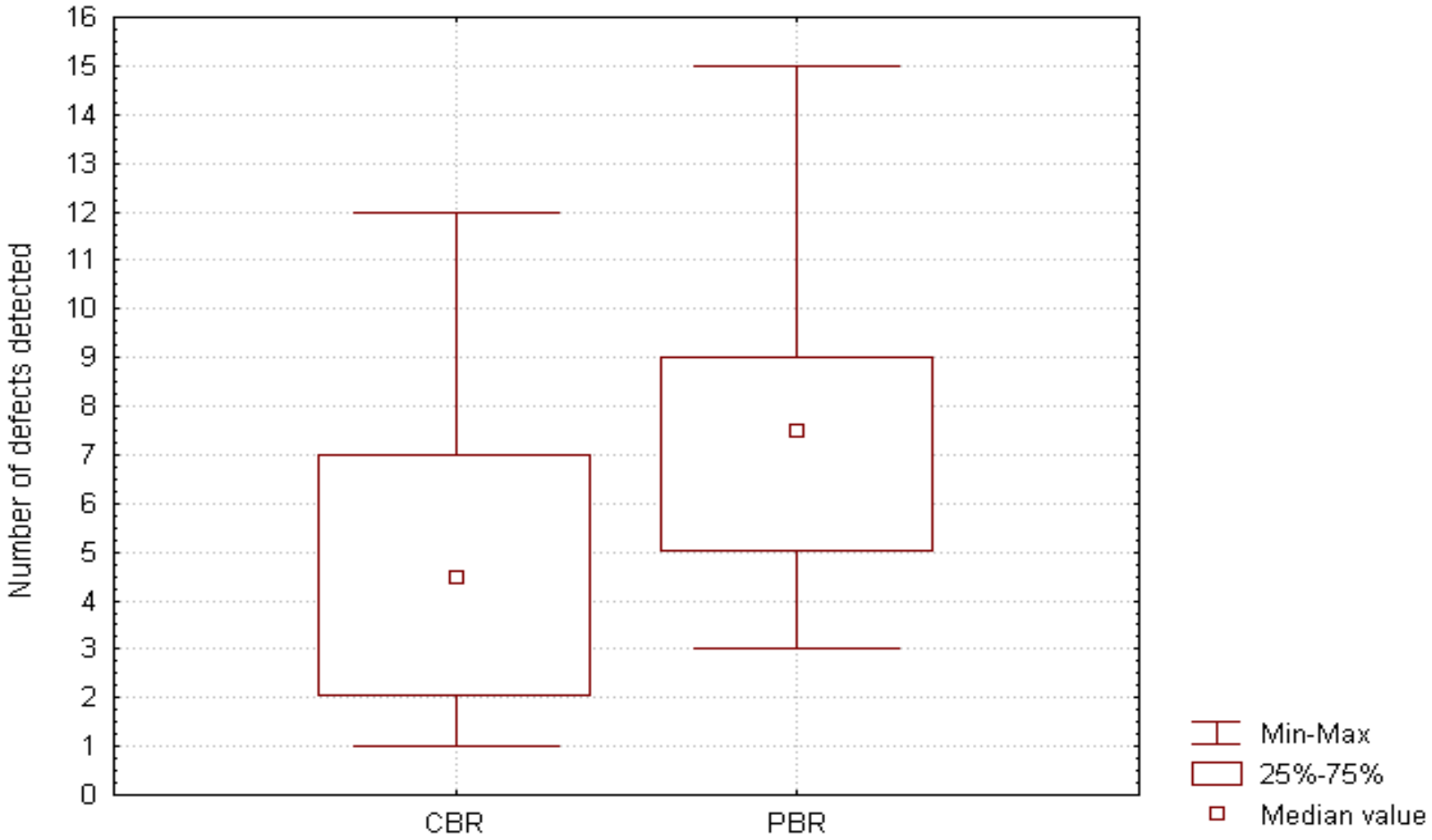
Statistica



Descriptive Statistics

Visualization

e.g., Boxplot



Data Validation

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

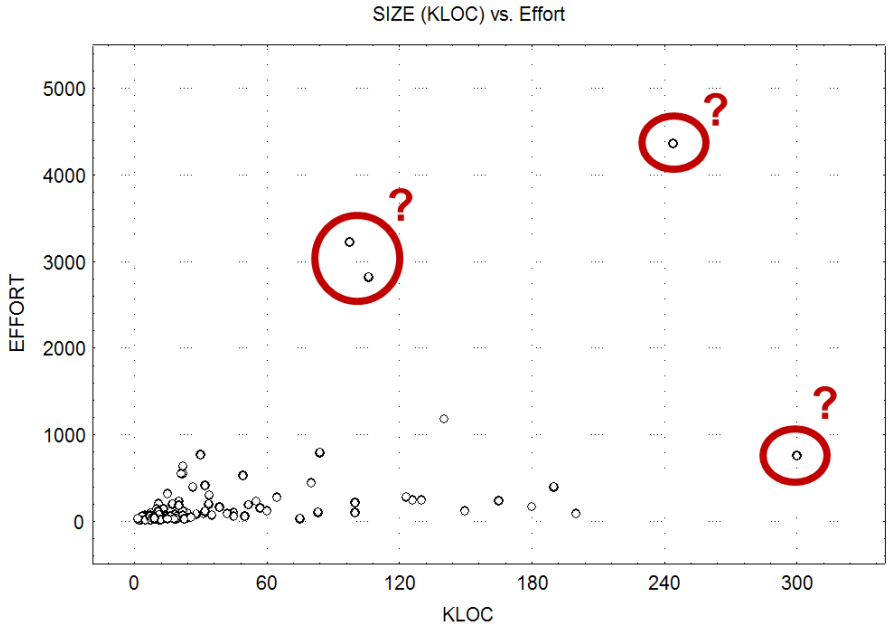
Inconsistent values

Data Validation

Outliers

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

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



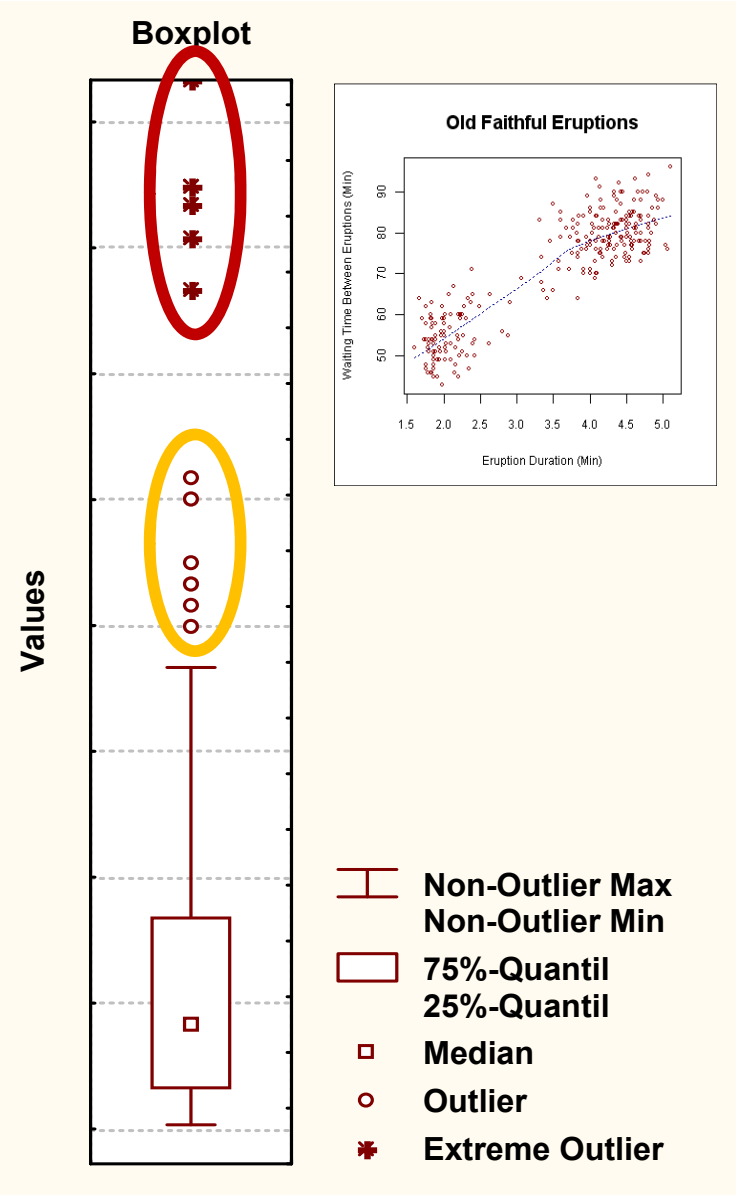
Data Validation

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

Options for dealing with outliers

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



Hypothesis testing

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

Hypothesis testing

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)

Hypothesis testing

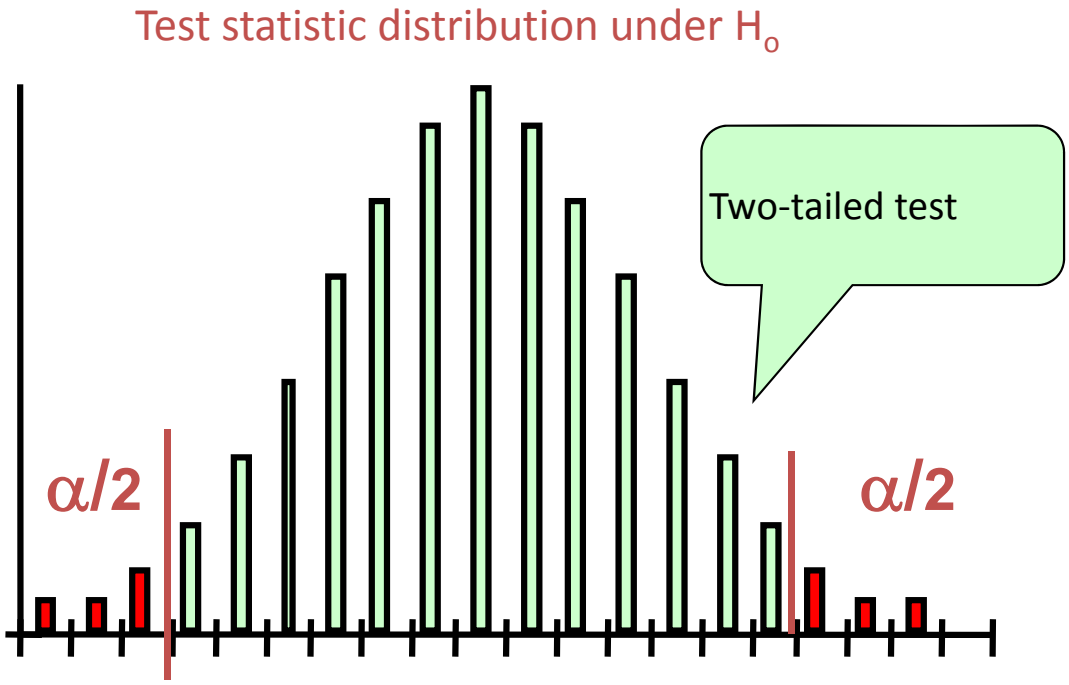
Logic of significance test: Example (1/2)

- 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={tester perspective (T), user perspective (U), designer perspective (D)}
    - DDR = number of defects found by individual / 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) <= DDR(N)
  - Alternative hypothesis
    - H_{12} = DDR(PBR PERSP) > DDR(N)

Hypothesis testing

Logic of significance test: Example (2/2)

- Statistical hypotheses



Hypothesis testing

Procedure (1/2)

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

		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

How to find the right test?

Purpose

- (a) compare treatments or (b) correlate variables
  - If compare treatments
    - How many independent variables did you manipulate
    - How many do you wish to compare at any one time
    - Related or unrelated sample

Type of Data

- (a) nominal, (b) ordinal or (c) at least interval
  - If interval
    - Are the scores normally distributed

Hypothesis testing

Procedure (1/2)

1. Determine test statistic of your sample
   - e.g., t-value for student t-test
2. Compare test statistic of your sample with comparison distribution

Tool supported

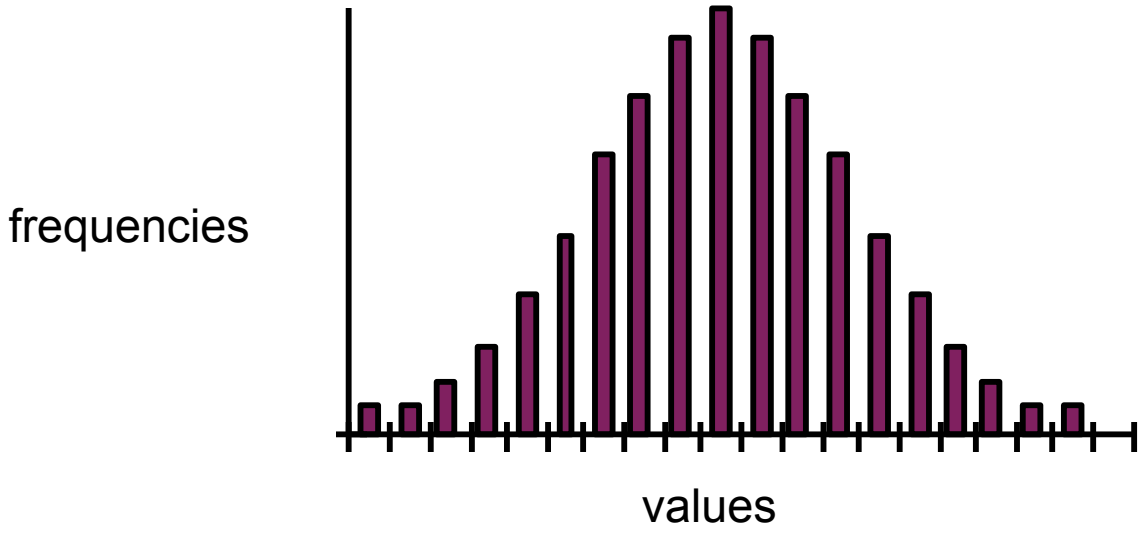
WS95/96	N	df. Effect	MS effect	df. Error	MS error	F	P-value
D vs. N	9	1	355.5	16	63.26	5.620	.301
U vs. N	8	1	6.250	14	66.36	.094	.736
T vs. N	8	1	115.5	14	99.81	1.158	.300

WS96/97	N	df. Effect	MS effect	df. Error	MS error	F	P-value
D vs. N	8	1	246.7	14	55.27	4.464	.053
U vs. N	9	1	547.0	16	71.49	7.651	.014
T vs. N	9	1	101.0	16	70.07	1.442	.247

Hypothesis testing

Assumption: Normal Distribution

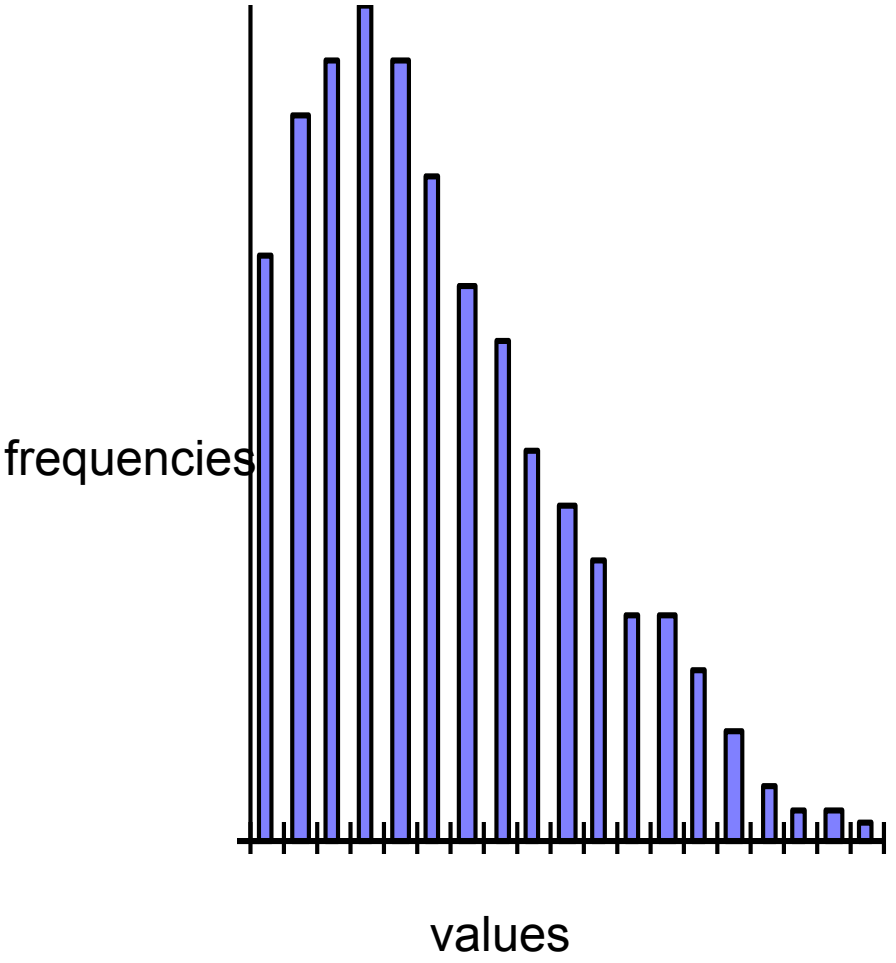
- Symmetrical, no beginning or ending point, 68.26% of the area within one standard deviation from the mean.
- Rare in software engineering
- Many variables have properties that are close enough, but frequently some variables are skewed to the right.



Hypothesis testing

Other distributions:

- e.g., Skewed distribution
  - Typical distribution in SE
  - Examples
    - Productivity
    - Change density



Hypothesis testing

What do we do when the populations sampled are not normally distributed?

Central Limit Theorem (simplified):

- Irrespective of the distribution of the parent population - given that its mean m and a variance s^2, and so long as the sample size n is large, the distribution of sample means is approximately normal with mean m and variance s^2 /n.

Beware of small sample sizes!

- Consider nonparametric tests.

Hypothesis testing

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

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	Variation 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	

Hypothesis testing

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?

Hypothesis testing

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
- ...

Hypothesis testing

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 hypothese

		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!

References

- Bortz, J. and Döring, N. (2006). Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler (4 Auflage). Berlin: Springer Verlag.
- Juristo, N., and Moreno, A. (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.
- Wohlin, Runeson, Höst, Ohlsson, Regnell, Wesslén (2000). Experimentation in Software Engineering: An Introduction, Kluwer Academic Publishers.