Data Analysis in Empirical Research - Overview

Prof. Dr. Hariet Köstner
WS 2017/2018

**Hochschule
Augsburg** University of
Applied Sciences

Business Research

Primary research

Secondary Research

Observation

Internal Data

Survey

External Data

Survey methods

Hochschule Augsburg University of Applied Sciences

Proven approaches

- Mail surveys

- Personal/Face-to-face surveys

- Telephone surveys (CATI)

- Online/Mobile surveys (CAWI)

New approaches

- App-based surveys

- Mixed-mode Surveys

- (N)ethnography

- Big Data Analytics

- Online-Communities

- Social Media Listening/ Monitoring

1. Data Collection Methods

2. Scales

3. Data Analysis

From a question to statistics

Satisfaction with the new car?

Levels of scale measurement

Hochschule
Augsburg University of
Applied Sciences

		Scale level	Permissible relations	Permissible operations	Permissible statistics
Possibilites of Analysis	Non-metric Data	Nominal	$A = A \neq B$	Counting	
		Ordinal	$A<B>A$	Ordering	
	Metric Data	Intervall	$A>B>C$ and $A - B = B - C$	Common arithmetic operations	
		Ratio	Proportion $(x_1/x_2)>(x_3/x_4)$	All arithmetic operations	

Every operation/statistic appropriate to a certain scale level is suitable for a SUBORDINATE scale.
Caution: Loss of information!

Rating Scales

☐	☐	☐	☐	☐
Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**

Most popular answer type in business research

Alternative: Semantic Differential

Rating Scales

☐	☐	☐	☐	☐
Strongly Agree	**Agree**	**Neutral**	**Disagree**	**Strongly Disagree**

Most popular answer type in business research

- Advantages
    - Easy to understand
    - Flexible (numeric, verbal, grafical, combinations)
    - (Nearly) metric data
    - Universal usage
- Disadvantages
    - No real metric data (actual only ordinal)
    - Answer patterns: tendency towards the middle/extremes/uniform

Rating Scales - Examples

Hochschule
Augsburg University of
Applied Sciences

Können Sie sich grundsätzlich vorstellen, Geld in einen Investmentfonds von | anzulegen?

- ○ ganz bestimmt
- ○ wahrscheinlich
- ○ eventuell
- ○ wahrscheinlich nicht
- ○ bestimmt nicht

weiter

Wie beurteilen Sie die folgenden Websites insgesamt?

	Ausgezeichnet	Sehr gut	Gut	Weniger gut	Schlecht
ortal.de	○	○	○	○	○
citymile.de	◉	◉	◉	◉	◉
berlin.de	○	○	○	○	○
muenchen.de	◉	◉	◉	◉	◉
hamburg.de	○	○	○	○	○

Weiter

Contents

1. Data Collection Methods

2. Scales

3. Data Analysis

Raw data:

Excel spreadsheet showing raw data:

	A	B	C	D	E	F	G
1	Responent #	Question 1	Question 2	Question 3	Age	Gender	
2		To what extend do you like the page?	Which layout would you recommend?	Which was your first impression of this page?			
3	1	2	1	nice	18	0	
4	2	3	1	chaotic	21	1	
5	3	1	2	didn't know, where to search	50	1	
6	4	2	2	confusing, don't like the colours	21	1	
7	5	2	3	askdfjaö	21	0	
8	6	1	1	a website	20	0	
9	7	3	3	good structure, company name dominant	51	0	
10	8	1	1	ok	55	1	
11	9	2	3		25	0	
12	10	5	3	fine	47	0	
13	11	2	1	looks pretty	24	1	
14	12	4	2	common, nothing special	19	0	
15							
16							

For all closed questions make a note somewhere which number is representing which answer (e.g. Value labels in SPSS)

For open questions create a **codeplan** (summarize the answers in meaningfull categories)

Raw data:

Mappe1 - Excel

DATEI START EINFÜGEN SEITENLAYOUT FORMELN DATEN ÜBERPRÜFEN ANSICHT Foxit PDF

A16

	A	B	C	D	E	F	G
1	Respondent #	Question 1	Question 2	Question 3	Age	Gender	
2		To what extend do you like the page?	Which layout would you recommend?	Which was your first impression of this page?			
3		1=like very much - 5 don't like at all	4 None of them			0 = male 1 = female	
4	1	2	1	nice	18	0	
5	2	3	1	chaotic	21	1	
6	3	1	2	didn't know, where to search	50	1	
7	4	2	2	confusing, don't like the colours	21	1	
8	5	2	3	askdfjaö	21	0	
9	6	1	1	a website	20	0	
10	7	3	3	good structure, company name dominant	51	0	
11	8	1	1	ok	55	1	
12	9	2	3		25	0	
13	10	5	3	fine	47	0	
14	11	2	1	looks pretty	24	1	
15	12	4	2	common, nothing special	19	0	
16							
17							

General categories of statistical Analysis

Descriptive Analysis

- -

Univariate

Bivariate

Inferential statistics

Multivariate

Descriptive Analysis

Hochschule
Augsburg University of
Applied Sciences

Descriptive
Analysis

Univariate

Bivariate

Multi-
variate

Useful Analysis

Appropriate Scale Level

Frequency Table
Simple Means (Mode, Median)
Cross-tabulations
Histogram

Nominal
Ordinal

Statistics of the Mean
Statistics of Deviation

Metric

Back to our example:

Scale level:

Frequency table:

Zeilenbeschriftungen	Anzahl
1	3
2	5
3	2
4	1
5	1
Gesamtergebnis	12

Mode:

Median:

(Arithmetic) Mean:

Back to our example:

Crosstabulations

Per Excel…

	Spaltenbeschriftungen ⊤		
Anzahl			
Zeilenbeschriftungen ⊤	**0**	**1**	**Gesamtergebnis**
1	1	2	3
2	3	2	5
3	1	1	2
4	1		1
5	1		1
Gesamtergebnis	**7**	**5**	**12**

Per SPSS…

Kreuztabelle Question_1*Gender

|			Gender 0	Gender 1	Gesamtsumme
Question_1	1,00	Anzahl	1	2	3
		% in Gender	14,3%	40,0%	25,0%
		% des Gesamtergebnisses	8,3%	16,7%	25,0%
	2,00	Anzahl	3	2	5
		% in Gender	42,9%	40,0%	41,7%
		% des Gesamtergebnisses	25,0%	16,7%	41,7%
	3,00	Anzahl	1	1	2
		% in Gender	14,3%	20,0%	16,7%
		% des Gesamtergebnisses	8,3%	8,3%	16,7%
	4,00	Anzahl	1	0	1
		% in Gender	14,3%	0,0%	8,3%
		% des Gesamtergebnisses	8,3%	0,0%	8,3%
	5,00	Anzahl	1	0	1
		% in Gender	14,3%	0,0%	8,3%
		% des Gesamtergebnisses	8,3%	0,0%	8,3%
Gesamtsumme		Anzahl	7	5	12
		% in Gender	100,0%	100,0%	100,0%
		% des Gesamtergebnisses	58,3%	41,7%	100,0%

Basics on statistical tests

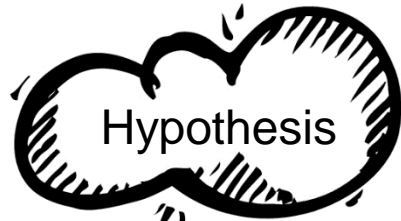*„Is there a difference between two groups?"*

„Does a correlation exist between … and …?"

„Does … influence …?"

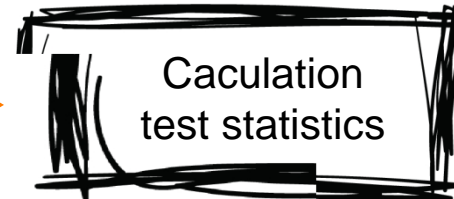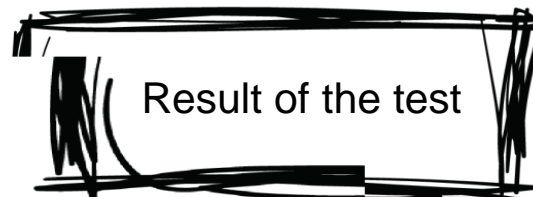# The way of testing a statistical hypothesis
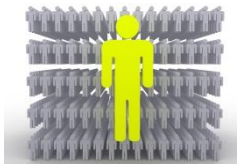
Sample

Hypothesis

Does the relation in the sample hold in the basic population as well?

Alternative hypothesis

Caculation test statistics

How likely is it, to find a relation in the sample, if the alternative hypothesis is true?

This is the hypotheses, the statistical test is going test!

Result of the test

Refusal of the **alternative hypothesis**, if the calculated probability is „relative" low ⇨ Data are indicating, that the assumed relation is true in a more general context.

Simple t-Test in Excel: Are males and females different with regard to the favor of the page?

	Respondent #	Question 1	Question 2	Question 3	Age	Gender	
1		To what extend do you like the page?	Which layout would you recommend?	Which was your first impression of this page?			
2							
		1=like very much - 5 don't like at all	4 None of them			0 = male 1 = female	
3							
4	1	2	1	nice	18	0	
5	5	2	3	askdfjaö	21	0	
6	6	1	1	a website	20	0	
7	7	3	3	good structure, company	51	0	
8	9	2	3		25	0	
9	10	5	3	fine	47	0	
10	12	4	2	common, nothing special	19	0	
11	2	3	1	chaotic	21	1	
12	3	1	2	didn't know, where to sea	50	1	
13	4	2	2	confusing, don't like the c	21	1	
14	8	1	1	ok	55	1	
15	11	2	1	looks pretty	24	1	
16							

Sort the data with reference to the group variable

$\bar{x}_{male}$ = 2,71 $\bar{x}_{female}$ = 1,8   ➡️ Significant difference?

➡️ Resulting p-Value: 0,22

Funktionsargumente

T.TEST

Matrix1	B4:B10	=	{2;2;1;3;2;5;4}
Matrix2	B11:B15	=	{3;1;2;1;2}
Seiten	2	=	2
Typ	2	=	2

= 0,219809855
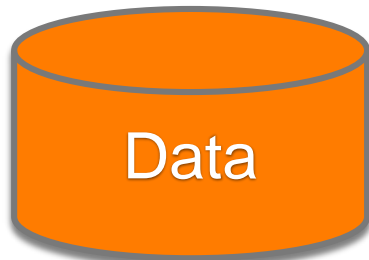
Gibt die Teststatistik eines Studentschen t-Tests zurück.

Typ bestimmt die Form des durchzuführenden t-Tests.
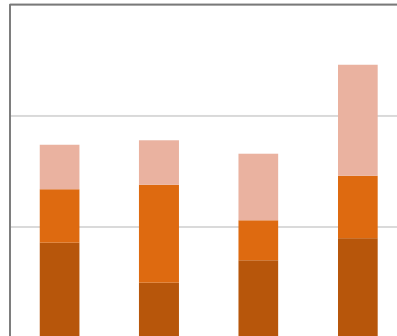
Formelergebnis = 0,219809855

Hilfe für diese Funktion   OK   Abbrechen

⬇️

No statistical significance attested, i.e. no difference between male and female according to the favor of the page.

Visualizing results

Data → Which statement is intended?

↓

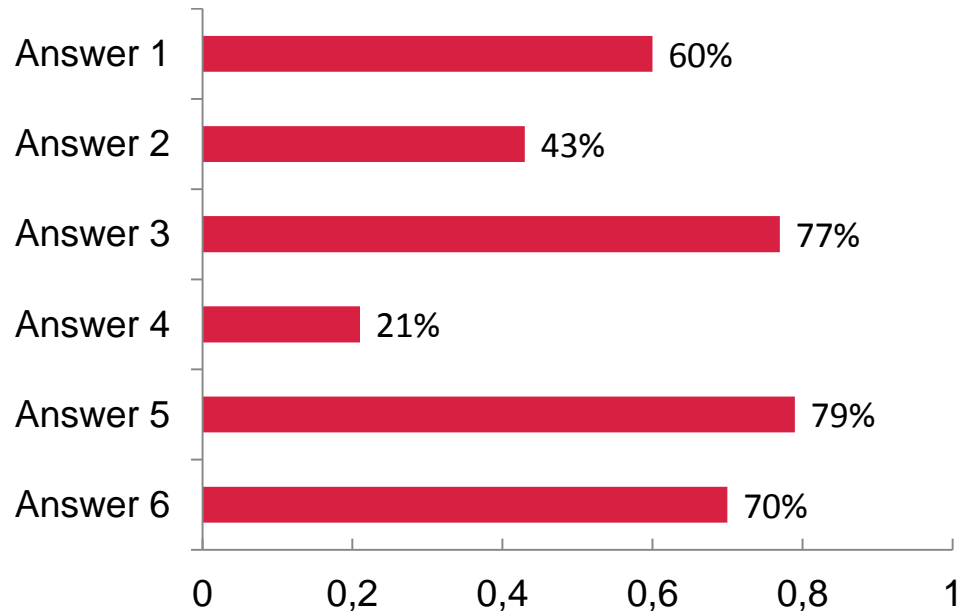Which type of diagramm is appropriate?

↓

Easy and unique interpretation

Hochschule Augsburg University of Applied Sciences

Types of diagramms

- Pie-chart

- Bar-chart

- Stacking diagramm

- Line chart

- …

Visualizing descriptive results

Hochschule
Augsburg University of
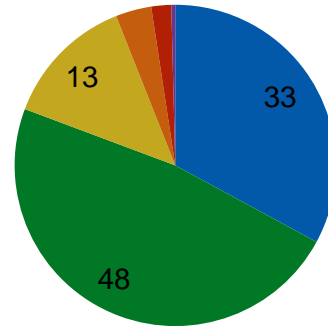Applied Sciences

„Text of the question"

Basis: All respondents, n = 250; depiction in %

Bar charts: suitable for questions with many answers, group comparisons; frequencies, percentages, means can be shown
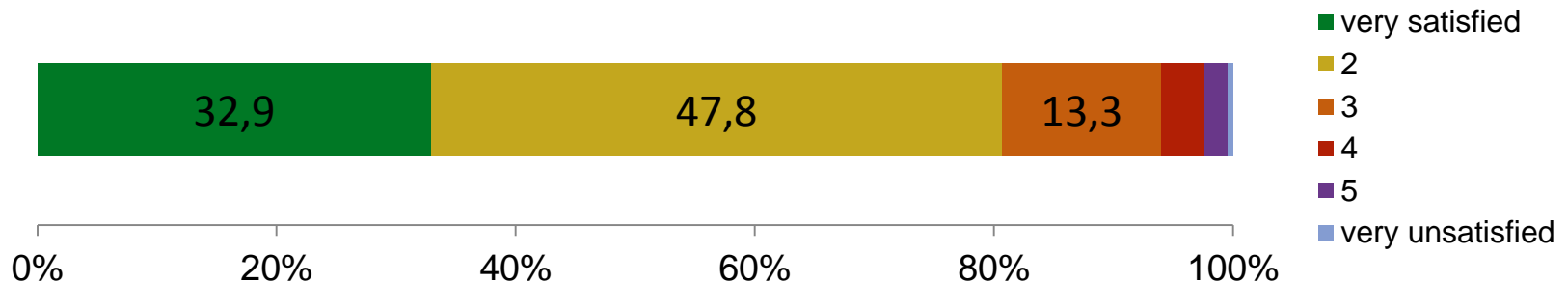
Visualizing descriptive results

„Text of the question"
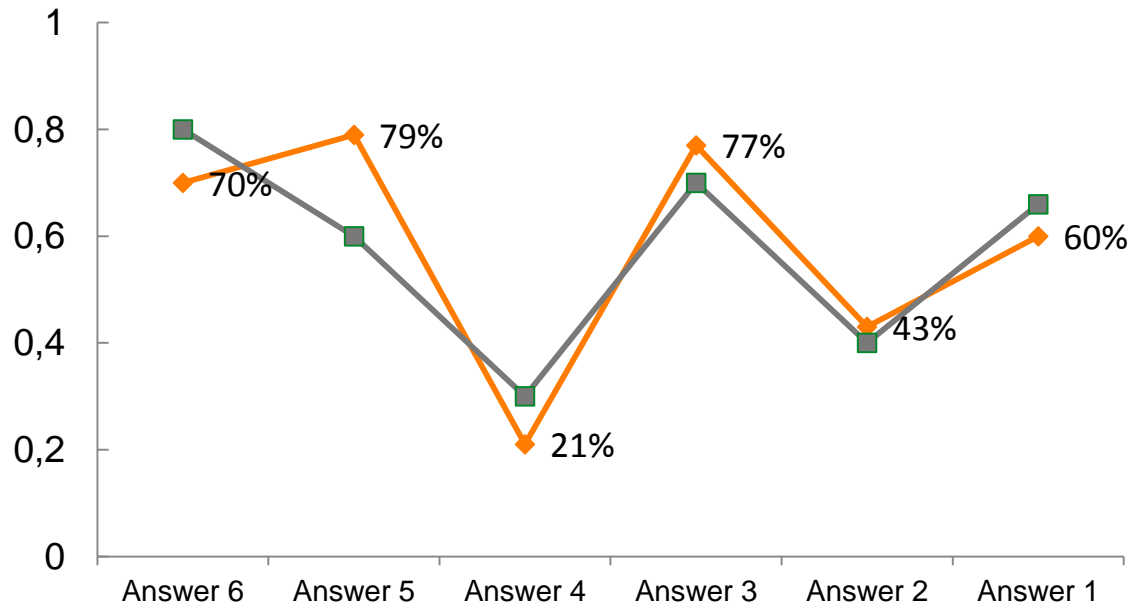

Basis: All respondents, n = 250; depiction in %

Pie charts: suitable for answers, which sum up to 100 and not to many categories.

■	very satisfied
■	2
■	3
■	4
■	5
■	very unsatisfied

Stacking diagramms: see pie chart; as bar chart for comparison of several answers/questions as well.

Visualizing descriptive results

„Text of the question"

Basis: All respondents, n = 250; figures in %

Line charts: suitable for questions with many answers, group comparisons; frequencies, percentages, means can be shown.
Compact graphical representation of a lot of information.

Prof. Dr. Hariet Köstner

Empirische Marktforschung / Marketing Research

Fakultät für Wirtschaft / Faculty of Business

Hochschule Augsburg / University of Applied Sciences

An der Hochschule 1

86161 Augsburg

Tel.: +49 821 5586 2950

Fax: +49 821 5586 2902

E-Mail: hariet.koestner@hs-augsburg.de

www.hs-augsburg.de