Mannheim Master in Management
MKT 520 Market Research

University of Mannheim

Chair of Sales & Services Marketing

Spring Term 2023

Team

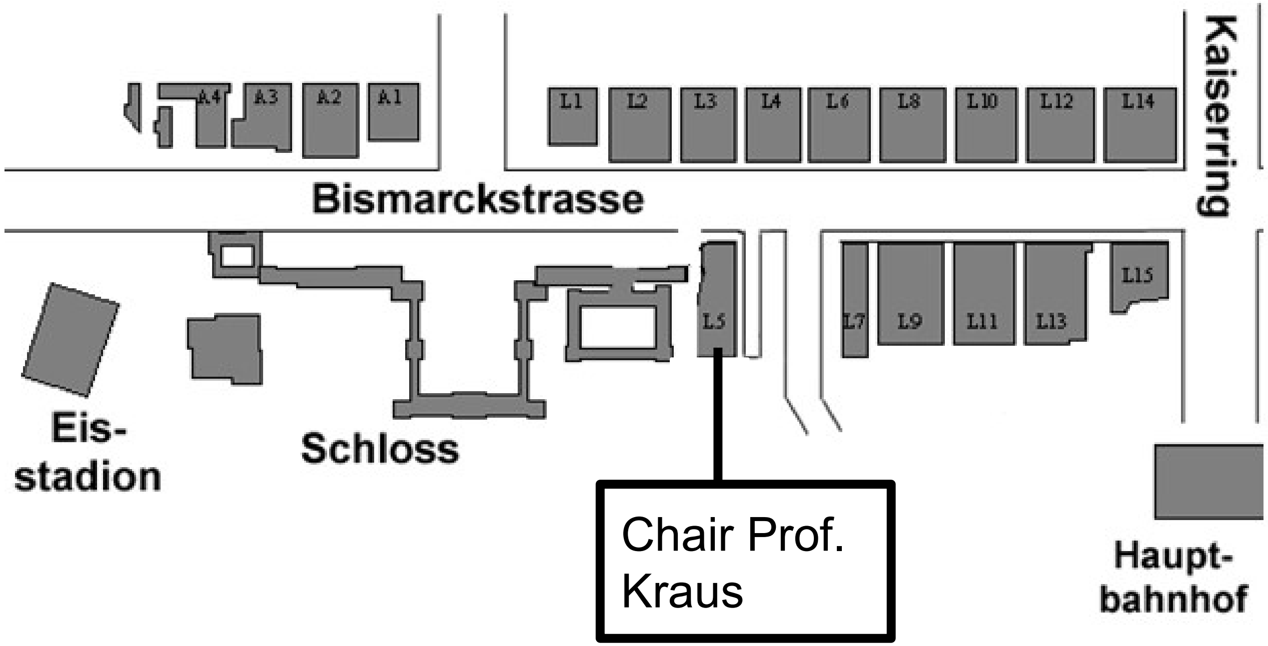
Prof. Dr. Florian Kraus
LECTURES

Room: L5, 2, Room 102-103
Office Hours: by appointment
E-Mail: office-kraus@bwl.uni-mannheim.de

Yasid Soufi, M.Sc.
CONTACT PERSON

Room: L5, 2, Room 007
Office Hours: by appointment
E-Mail: soufi@bwl.uni-mannheim.de

Chair of Sales and Services Marketing



Chair of Sales and Services Marketing

Prof. Dr. Florian Kraus
L5, 2
68161 Mannheim

Office
Ewelina Bartella, Room 104
Mail: office-kraus@bwl.uni-mannheim.de
Office hours: Monday to Friday, 9.30 – 12:00 am and 1.00 – 2.30 pm or by appointment

Follow Us!


Lehrstuhl für Sales & Services Marketing
Universität Mannheim
Fakultät für Betriebswirtschaftslehre
L 5, 2
68161 Mannheim
Tel.: +49 621 181-2662
Fax: +49 621 181-2672
E-Mail: office-kraus(at)bwl.uni-mannheim.de
Foto: Norbert Bach
© 1987–2018 HERE, Deutschland | Terms of use

Internet

- Information about our courses:
  kraus.bwl.uni-mannheim.de
- General information on this course:
  Teaching > Master >
  MKT 520 Market Research

Master Level

Course name	Credits	Details	Term
MKT 520 Market Research	6 ECTS	Course details	Spring
MKT 560 Services Marketing	4 ECTS	Course details	Fall
MKT 612 Business-to-Business Marketing	2 ECTS	Course details	Fall
MKT 613 Negotiation Management	2 ECTS	Course details	Fall
MKT 740 Research Seminar / Seminar Thesis	6 ECTS	Course details	Fall and Spring

Click here to reach the Info-session presentation for MMM students.

E-learning: ILIAS

- Register at Portal² for
  MKT 520 Market Research
  (portal2.uni-mannheim.de)

- You are now automatically registered to the ILIAS group
  MKT 520 Market Research [V] [1. PG] (FSS 2023)
  (ilias.uni-mannheim.de)

- Here you can find
  - Updates on the course
  - Relevant literature for download and
  - Answers to frequently asked questions (FAQs)

Course details

Lectures:	**90 min., Mondays, SN 169 Röchling Hörsaal, 1.45 – 3.15 pm (Room: SN 169)**
Exercise classes:	**90 min., Wednesdays, SN 169 Röchling Hörsaal, 1.45 – 3.15 pm (Room: SN 169)** R-Tutorials (Self-Study): 15.02 –14.03 First Exercise in SN 169: 15.03
Lecture slides	**A reduced version will be available on ILIAS.**

Lecture materials

- Copyright restrictions prohibit us from providing all slides online
- Most slides contain information from books and registered trademarks
- Yet, we offer two options for you…

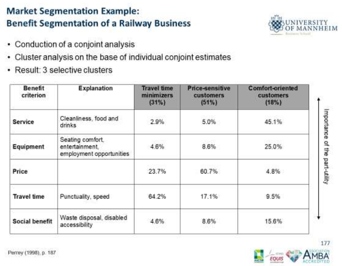
Option 1 BEST CHOICE

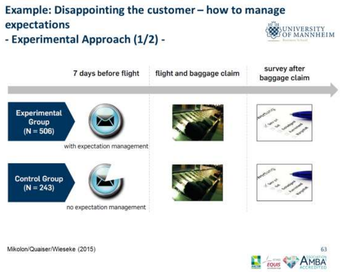
- Entire slide deck (271 slides)
- Printed
- Includes all examples, models and logos

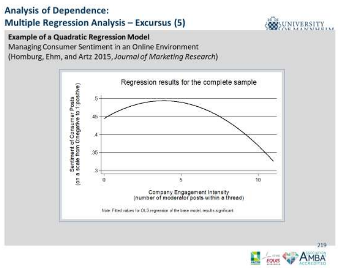
Option 2

- Short version (218 slides)
- Digital

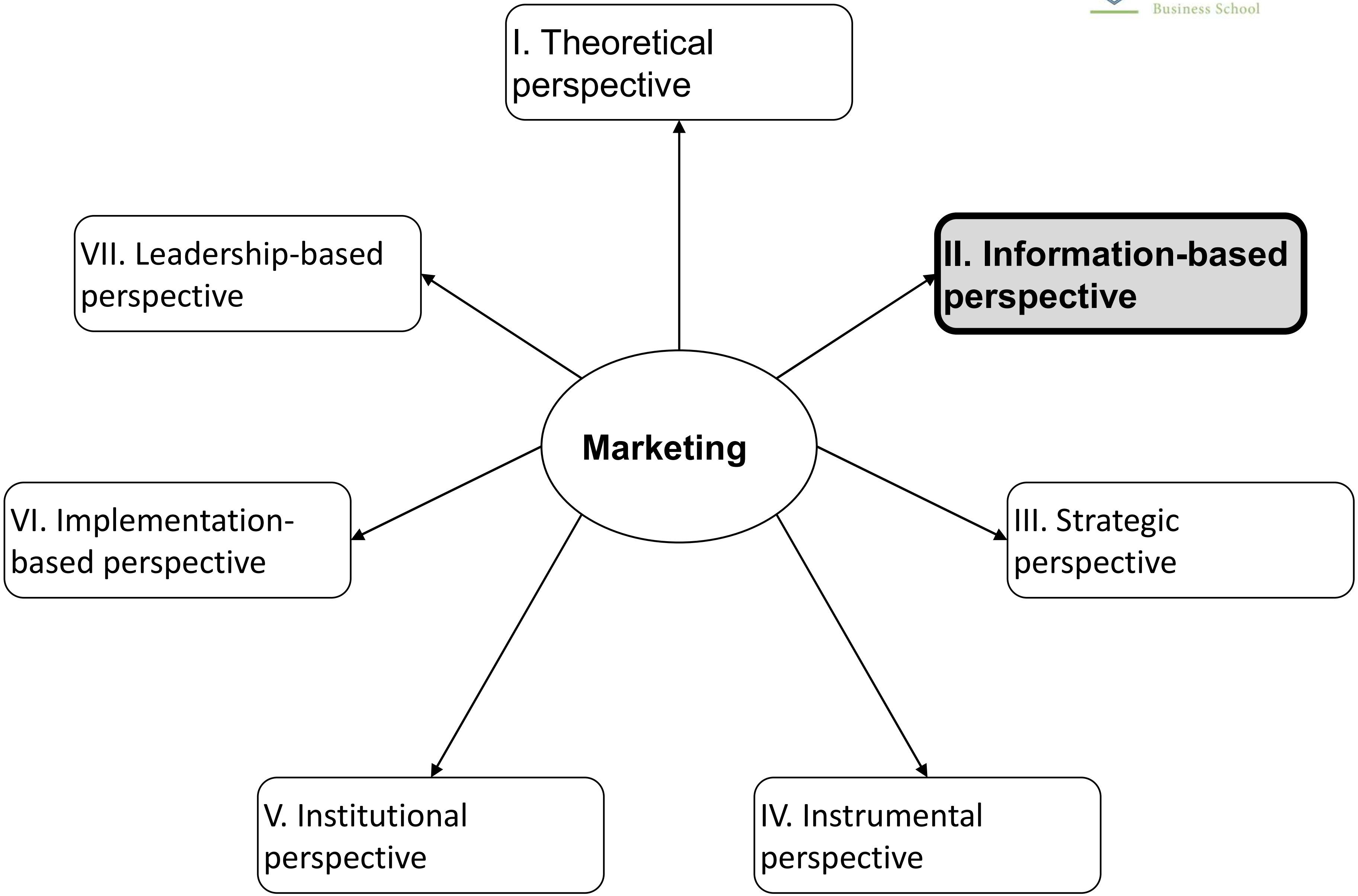
Exemplary slides that will only be provided in Option 1







Classification of the course



Course objectives and requirements

UNIVERSITY OF MANNHEIM
Business School

Course objectives

- Get acquainted with details on the market research process
- Understand and apply major market research methods
- Implement the methods via state-of-the-art statistical programming in R Statistics
- Get prepared for the data collection and analysis during your master thesis
- Get prepared for future employment

→ The final exam addresses these course objectives

Requirements

- Fundamentals in marketing (e.g., Marketing I, Marketing for Business Minors)
- Fundamentals in statistics
- Recommendation: We strongly recommend the successful participation of the course "CC 503 Empirical Methods" before attending this course

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. Sampling
4. Design of the research instrument
5. Data collection
6. Editing and coding of data
7. Analysis and interpretation of data
    1. Exploratory Factor Analysis
    2. Confirmatory Factor Analysis
    3. Cluster Analysis
    4. Multivariate Analysis of Variance
    5. Multiple Regression Analysis
    6. Structural Equation Modeling with latent variables
    7. Conjoint Analysis

The process of market research: A practical example – shampoo product test

- We introduce the practical example of a shampoo product test to illustrate and get you even more familiar with the market research process
- Wherever appropriate, we apply this example for steps within the market research process including data analysis methods
- Whenever you find a shampoo bottle at the right top corner, we illustrate the lecture content with the example of the shampoo product test

Analysis and interpretation of data:
R Statistics

- We offer an introduction and Q&A to the statistical programming language R Statistics (on 15.03.2023)
- Upfront you will self-study R-Tutorials (focus on **first 3 tutorials**)- similar as Marketing Analytics [MKT 511]
  - https://www.datasciencewithr.education/VirtualClassroom/
  - Login: LearningR Passwort: FunW!thR
- Send questions for the R Statistics Q&A until the 10th of March 2023 to soufi@bwl.uni-mannheim.de
- Exemplary R calculations in the lecture are tagged with the R logo
- Throughout the exercises you will learn to implement the methods from the lecture in R
- On ILIAS, we provide you with the slides summarizing the exercise sessions & all the datasets on which the analyses are based

After participating in the lecture and exercises you should have the skillset to conduct Real-World Market Research Projects in R!

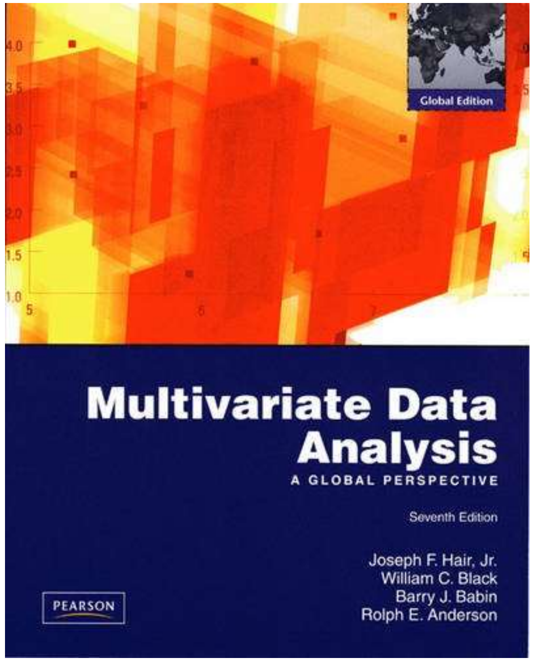
Guest speakers

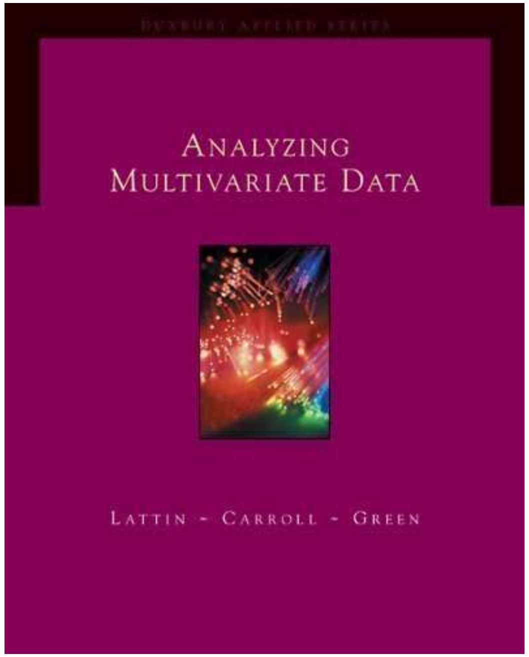
&

Literature (1)

Basic literature

Hair Jr, J.F., Black, W.C., Babin B.J., Anderson, R.E. (2010), Multivariate Data Analysis, 7th edition, Upper Saddle River, NJ.

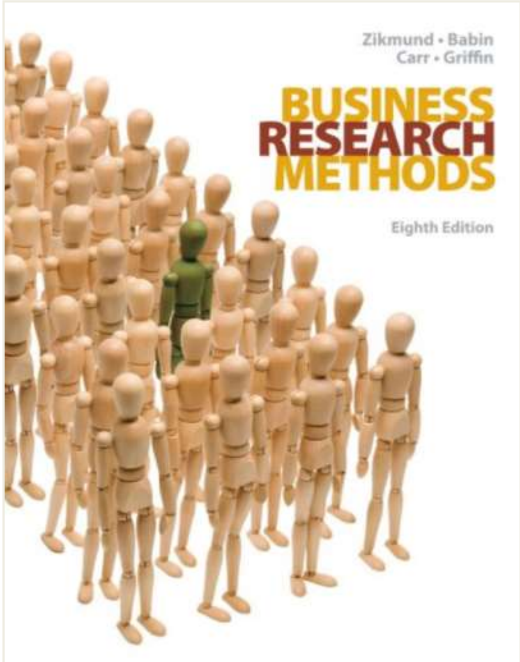




Lattin, J., Carrol, D., Green, P. (2003), Analyzing Multivariate Data, 1st edition, Belmont, CA.

Literature (2)

Basic literature

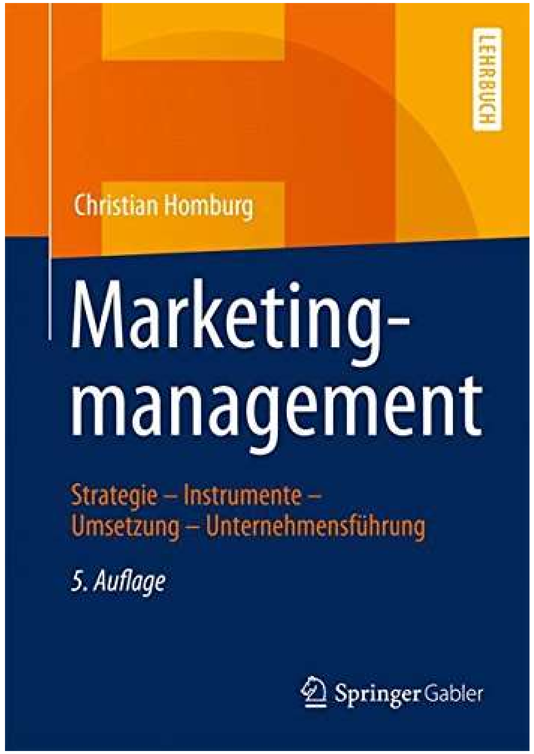


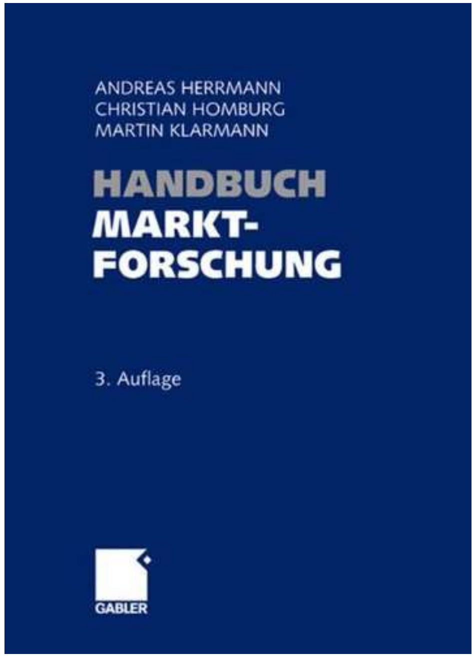
Zikmund, William G., Babin, Barry J., Carr, Jon C., Babin (2009), Business Research Methods, 8th edition, Wadsworth Inc Fulfillment.

Literature (3)

Basic literature in German

Homburg, Ch. (2015), Marketingmanagement: Strategie – Instrumente – Umsetzung – Unternehmensführung, 5th edition, Wiesbaden.





Herrmann, A., Homburg, Ch., Klarmann, M. (2008, Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden.

Literature (4)

- Supplementary literature

•Aaker, D., Kumar, V., Day, G., Leone, R. (2012), Marketing Research, 11th edition, New York

•Churchill, G., Iacobucci, D.(2009), Marketing Research: Methodological Foundations, 10th edition, Fort Worth

•Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. Journal of Marketing Research, 52(5), 629-641.

•Hox, J. J. (2010), Multilevel Analysis – Techniques and Applications, 2nd ed., Routledge, New York

•Iacobucci, D. (2013), Marketing Models: Multivariate Statistics and Marketing Analytics, South Western Cengage Learning

•Malhotra, N. (2010), Marketing Research – An Applied Orientation, 6th edition, Upper Saddle River

•Malhotra, N., Birks, D. (2007), Marketing Research – An Applied Approach, Harlow

• Mikolon, S., Quaiser, B., Wieseke, J. (2015), Don't try harder: using customer inoculation to build resistance against service failures, Journal of the Academy of Marketing Science volume 43, pages 512–527 (2015)

•Raudenbush, S. W., Bryk, A. S. (2002), Hierarchical Linear Models – Applications and Data Analysis Methods, 2nd ed., Sage Publications

•Shadish, W. R., Cook, T. D., Campbell, D. T. (2002), Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Boston

•Witten, I. H., Frank, E., Hall, M. A. (2011), Data Mining, 3rd ed., Morgan Kaufmann

•Wooldrige, J. M. (2012), Introductory Econometrics – A Modern Approach, 5th ed., South Western Educ Pub

Literature (5)

- Supplementary literature in German

•Backhaus, K., Erichson, B., Plinke, W., Weiber, R. (2010), Multivariate Analysemethoden, Eine anwendungsorientierte Einführung, 13th edition, Berlin

•Bamberg, G., Baur, F., Krapp, M. (2012), Statistik, 1th edition, München

•Bauer, H. (1995), Marktabgrenzung, in: Tietz, B., Köhler, R., Zentes, J. (Hrsg.), Handwörterbuch des Marketing, 2nd edition, Stuttgart, 1709-1721

•Berekoven, L., Eckert, W., Ellenrieder, P. (2009), Marktforschung: Methodische Grundlagen und praktische Anwendung, 12th edition, Wiesbaden

•Bleymüller, J., Gehlert, G., Gülicher, H. (2012), Statistik für Wirtschaftswissenschaftler, 16th edition, München

•Bortz, J., Weber, R. (2010), Statistik für Human- und Sozialwissenschaftler, 7th edition, Berlin

•Frenzen, H., Krafft, M. (2008), Logistische Regression und Diskriminanzanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 607-649

•Gensler, S. (2003), Heterogenität in der Präferenzanalyse – Ein Vergleich von hierarchischen Bayes-Modellen und Finite-Mixture-Modellen, Wiesbaden

•Günther, M., Vossebein, U., Wildner, R. (2007), Marktforschung mit Panels: Arten – Erhebung – Analyse – Anwendung, 2nd edition, Wiesbaden

•Herrmann, A., Homburg, Ch., Klarmann, M. (2008), Marktforschung: Ziele, Vorgehensweise und Nutzung, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 3-20

•Homburg, Ch. (2000a), Quantitative Betriebswirtschaftslehre, 3rd edition, Wiesbaden

•Homburg, Ch., Daum, D. (1997), Marktorientiertes Kostenmanagement: Kosteneffizienz und Kundennähe

Literature (6)

- Supplementary literature in German

Homburg, Ch. (2015), Marketingmanagement: Strategie – Instrumente – Umsetzung – Unternehmensführung (Kapitel 6 und 7), 5th edition, Wiesbaden, 241-420

Homburg, Ch., Herrmann, A., Pflesser, Ch., Klarmann, M. (2008), Methoden der Datenanalyse im Überblick, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 151-174

Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 213-240

Homburg, Ch., Klarmann, M., Krohmer, H. (2008), Statistische Grundlagen der Datenanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 213-239.

Homburg, Ch., Klarmann, M., Pflesser, Ch. (2008), Konfirmatorische Faktorenanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 271-304

Homburg, Ch., Pflesser, Ch., Klarmann, M. (2008), Strukturgleichungsmodelle mit latenten Variablen: Kausalanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 547-578

Hüttner, M., Schwarting, U. (2008), Exploratorische Faktorenanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 241-270

Jensen, O. (2008), Clusteranalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 335-372

Karlshaus, J. (2000), Die Nutzung von Kostenrechnungsinformationen im Marketing, Wiesbaden

Klarmann, M. (2008), Methodische Probleme der Erfolgsfaktorenforschung: Bestandsaufnahme und empirische Analysen, Wiesbaden

Literature (7)

- Supplementary literature in German

•Koschate, N. (2008), Experimentelle Marktforschung, in: Herrmann, A., Homburg, Ch., Klarmann, M., Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 107-121

•Meffert, H. (1992), Marketingforschung und Käuferverhalten, 2nd edition, Wiesbaden

•Perrey, J. (1998), Nutzenorientierte Marktsegmentierung, Wiesbaden

•Sarris, V. (1992), Methodologische Grundlagen der Experimentalpsychologie, Band 2, Versuchsplanung und Stadien des psychologischen Experiments, Stuttgart

•Sarris, V., Reiß, S. (2005), Kurzer Leitfaden der Experimentalpsychologie, München

•Skiera, B., Albers, S. (2008), Regressionsanalyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 467-498

•Völckner, F., Sattler, H., Teichert, T. (2008), Wahlbasierte Verfahren der Conjoint-Analyse, in: Herrmann, A., Homburg, Ch., Klarmann, M. (Hrsg.), Handbuch Marktforschung: Methoden – Anwendungen – Praxisbeispiele, 3rd edition, Wiesbaden, 687-711

•Weiber, R., Mühlhaus, D. (2013), Strukturgleichungsmodellierung, 2nd edition, Heidelberg

Please note that taking photos, videos, or any other tape recordings is strictly forbidden in all sessions of the lecture and exercise classes

Course outline

1. Fundamentals of market research

2. Determination of the data collection method

3. Sampling

4. Design of the research instrument

5. Data collection

6. Editing and coding of data

7. Analysis and interpretation of data

Definition and goals of market research

UNIVERSITY OF MANNHEIM
Business School

Definition

"Market research refers to the process of gathering and evaluating information in the market (e.g. from actual and potential customers or competitors) in order to support marketing decision making."

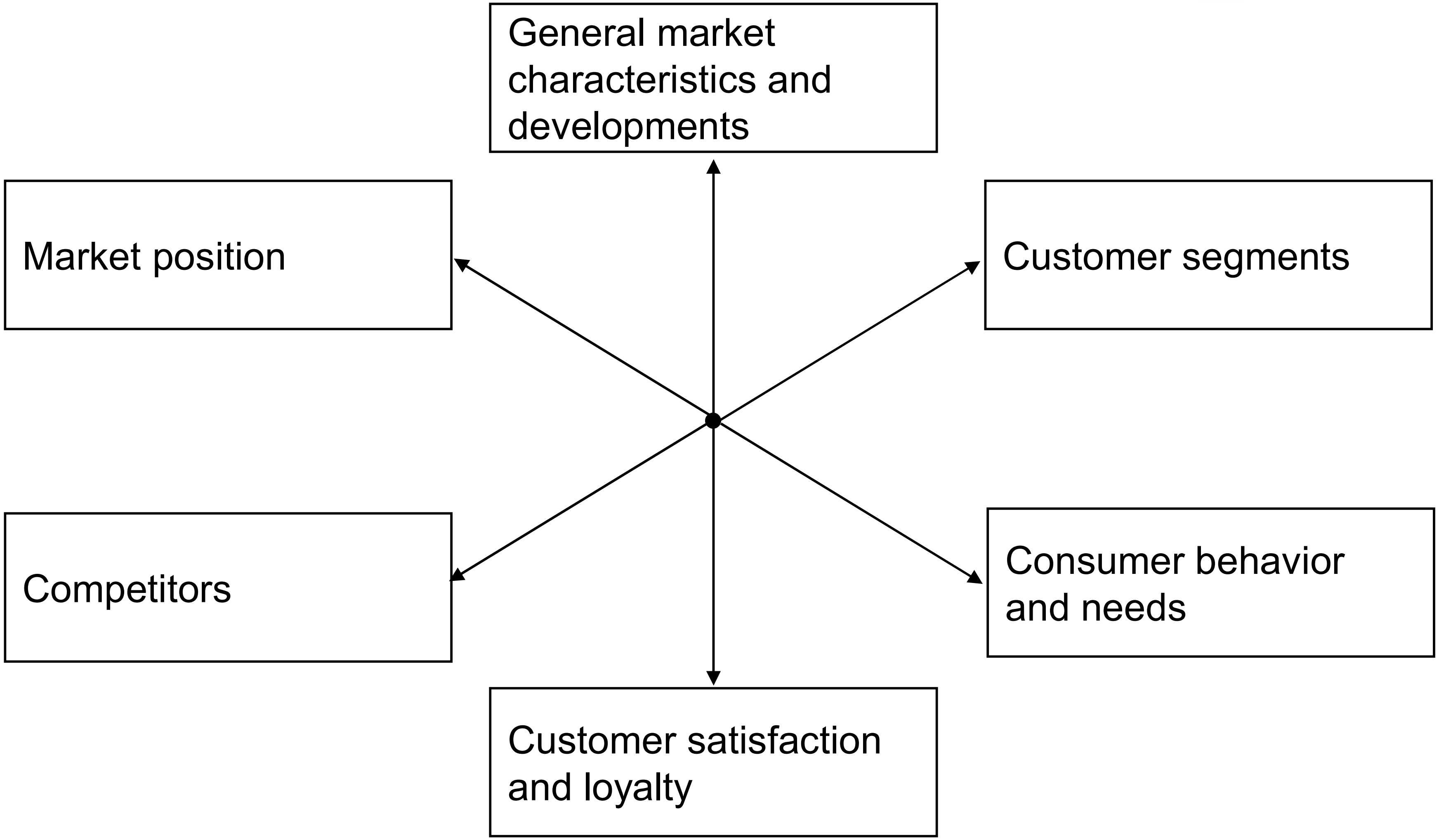
Homburg, Kuester, Krohmer (2013), p. 608

Goals of market research

- Continuous improvement of the information that is crucial for decision making with regard to
  - Actuality
  - Objectivity
  - Precision
  - Relevance
- Satisfying the need for information of the decision maker
- Timely identification of trends, opportunities, and risks in the markets
- Limiting the risk of wrong decisions
- Supporting the decision-making process in the firm

Homburg (2015), p.248

Key areas of investigation in market research



Homburg (2015), p. 250

Quality assessment in market research

Objectivity	• Measurement and results of the analysis are independent from the researcher • Objectivity with regard to – The research process – The analysis – The interpretation of the results
Reliability	• Formal accuracy when measuring the feature characteristics → freedom of random errors • Ability to reproduce the results under constant measuring conditions • Potential causes of measurement error: – Change of (external) conditions – Lack of feature's consistency – Lack of precision of the research instrument
Validity	• Concerns the question whether the research instrument measures what it is supposed to measure → accuracy of the measure with regard to content → freedom of systematic errors • Internal validity: Variation of the dependent variable can only be attributed to the variation of the independent variable (elimination of confounding factors) • External validity: Generalizing of results; representation of the analysis

Homburg (2015), pp. 254

The process of market research

1	**Problem formulation**	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Step 1: Problem formulation
Example shampoo product test

You work as a marketing manager for a major brand of a volumizing shampoo (**Shampoo A**). For half a year, you have been noticing declining sales figures. As there are many volumizing shampoos on the market, you want to compare the performance of your shampoo with nine volumizing shampoos offered by competitors.

The objectives of the study are to

- Derive strategic implications in terms of
  - Market positioning and
  - Product improvement
- Stop the decline of the sales figures as soon as possible, and
- Strengthen the market position of your volumizing shampoo

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	**Determination of the research design**	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Step 2: Determination of the research design

Type of study/research

- Descriptive: Describes characteristics of objects, people, groups, organizations, or environments; tries to "paint a picture" of a given situation
- Exploratory: Conducted to clarify ambiguous situations or discover ideas that may be potential business opportunities
- Explanatory: Allows causal inferences to be made; seeks to identify cause-and-effect relationships

Zikmund et al. (2009), p. 55

Step 2: Determination of the research design
Example shampoo product test

- **Exploratory**
  - We want to explore which characteristics of the volumizing shampoos customers perceive as superordinate factors (e.g., Scent, Volumizing performance, …) (exploratory and confirmatory factor analysis)

- **Explanatory**
  - We want to test whether the respondents' quality assessment of the shampoos vary across customers' hair structure (analysis of variance)
  - We want to calculate in how far the identified factors of the shampoos positively increase customers' buying intention (regression analysis)
  - We want to test the causal relationship CSR reputation → Trust / Customer-company identification → Customer loyalty (structural equation modeling)

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	**Determination of the executing researcher**	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Step 3: Determination of the executing researcher
Example shampoo product test

We decide to work with a market research institute due to its

- Profound knowledge
- Substantial experience
- Lower costs
- Greater objectivity
- Higher capacities in conducting market research

Course outline

1. Fundamentals of market research

2. Determination of the data collection method

3. Sampling

4. Design of the research instrument

5. Data collection

6. Editing and coding of Data

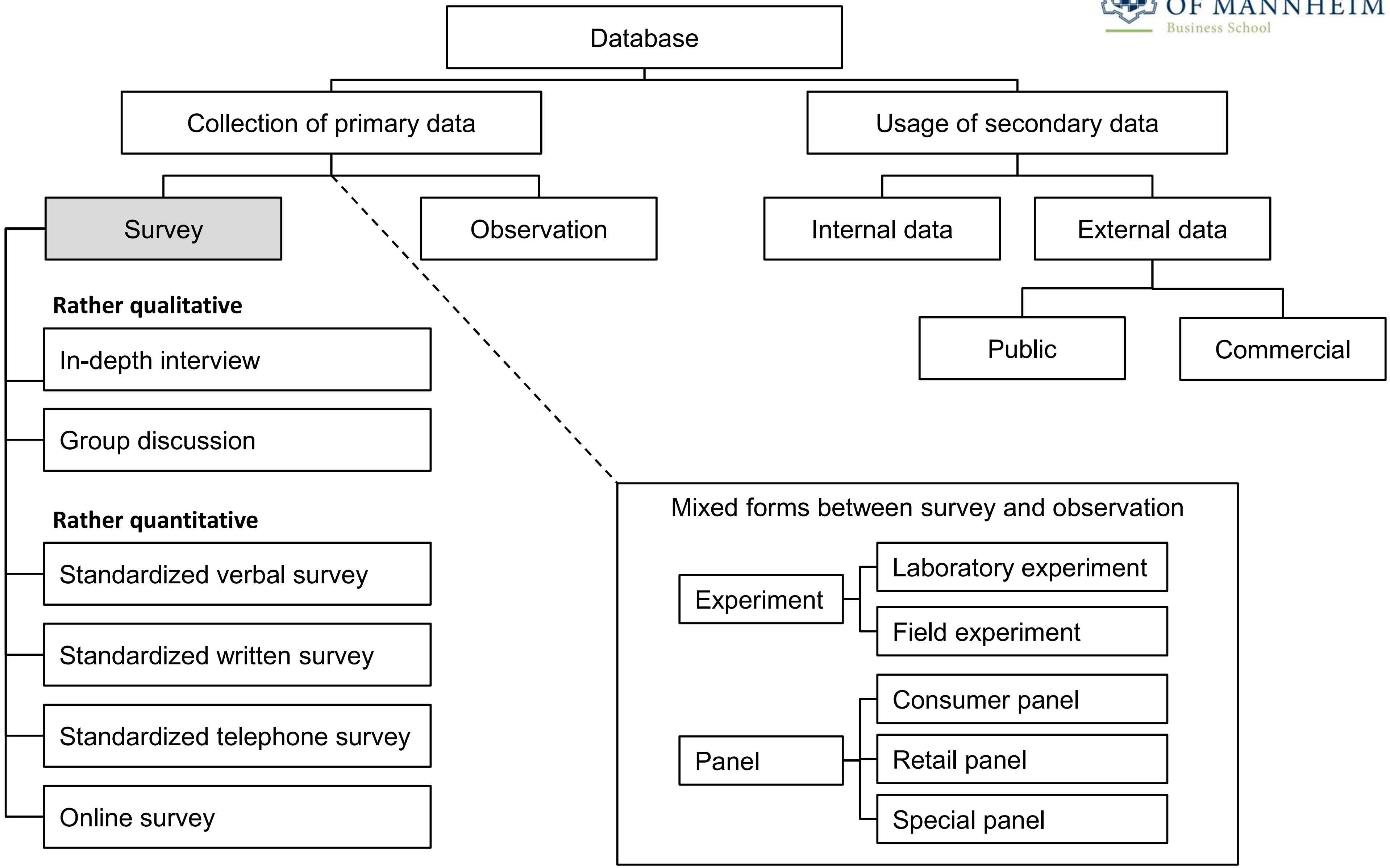
7. Analysis and interpretation of data

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	**Determination of the data collection method**	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Overview of data collection methods



Homburg (2015), p.262

Primary research: Surveys

	Data collection method	Description
Rather qualitative	In-depth interview	• Non-directive: unstructured interview in the form of a personal conversation • Semi-structured: interview guided along a list of subjects (interviewer manual), especially useful for expert surveys • Duration: 1-3 hours
	Group discussion	• Interviewing a moderate-sized group of participants (usually 6-12) guided by a moderator • Loss of inhibitions and mutual animation to respond and state opinions thanks to the presence of other group members • Structured procedure (moderation manual)
Rather quantitative	Standardized verbal interview	• Usually a one-time, representative verbal survey about a pre-defined object of study • Basis: Standardized questionnaire (predetermined questions that allows comparability)
	Standardized written interview	• Paper and pencil questionnaires sent to the participants
	Standardized telephone interview	• Standardized survey, e.g., via CATI (Computer Assisted Telephone Interviewing)
	Online survey	• Email surveys: Invitation of potential participants via email to take part in the survey; reference to the online survey via an individual link or send the questionnaire via email • Internet surveys ("on site"): Implementation of a questionnaire on a website (e.g., of the provider)

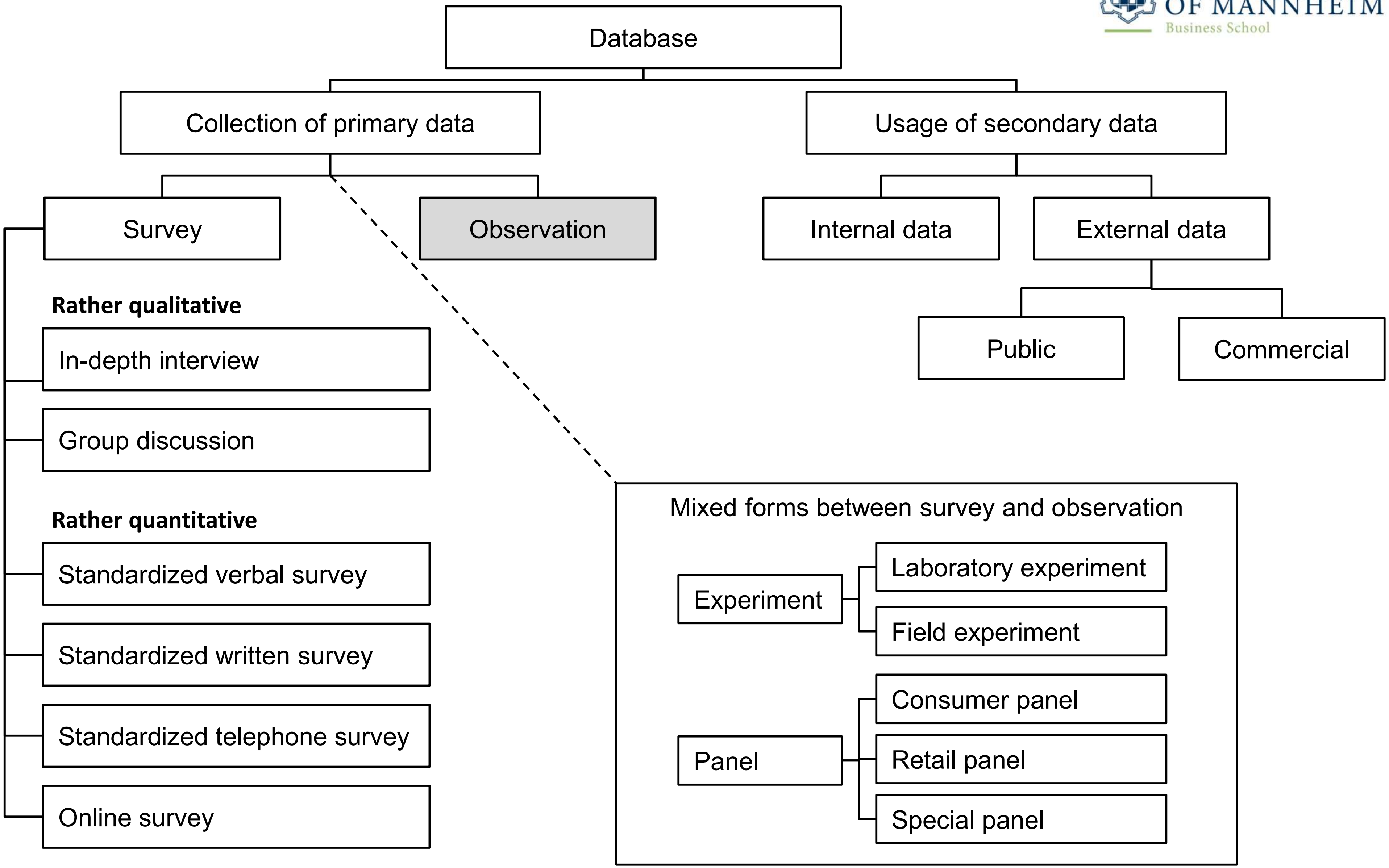
Homburg (2015), pp.263.

Primary research:
Evaluation of quantitative survey methods

Criteria	Standardized verbal survey	Standardized telephone survey	Standardized written survey	Standardized online survey
Suitability for the object of study				
Possibility to explain a complex issue	Very good	Good	Rather low	Rather low
Possibility to use complex scales and filter questions in the questionnaire	Only when computer-aided	Only when computer-aided	Low	Very good
Possibility to demonstrate trial product samples	Very good	Rather low	Rather low	Rather low
Extent of data				
Subjective length of the survey evaluated by the interviewee	Short	Medium	Long	Very long
Possibility to question a large sample	Low	Medium	Very high	High
Response rate	High	Medium	Low	Low
Risk of aborting the interview	Low	Medium	High	Very high
Quality of data				
Possibility of inquiries when comprehension problems occur	Very good	Good	Very low	Very low
Possibility of distortion due to the social interaction with the interviewer	Very high	High	None	None
Possibility to modify the research instruments during the field phase	Very good	Good	Low	Good
Length and costs of the project				
Length of the field phase	Medium	Short	Long	Long
Costs	Very high	High	Low	Low

Modified according to Homburg (2015), p.269

Overview of data collection methods



Homburg (2015), p. 280

Primary research: Observation

Observation: Planned registration of observable issues, behaviors and characteristics of specific persons

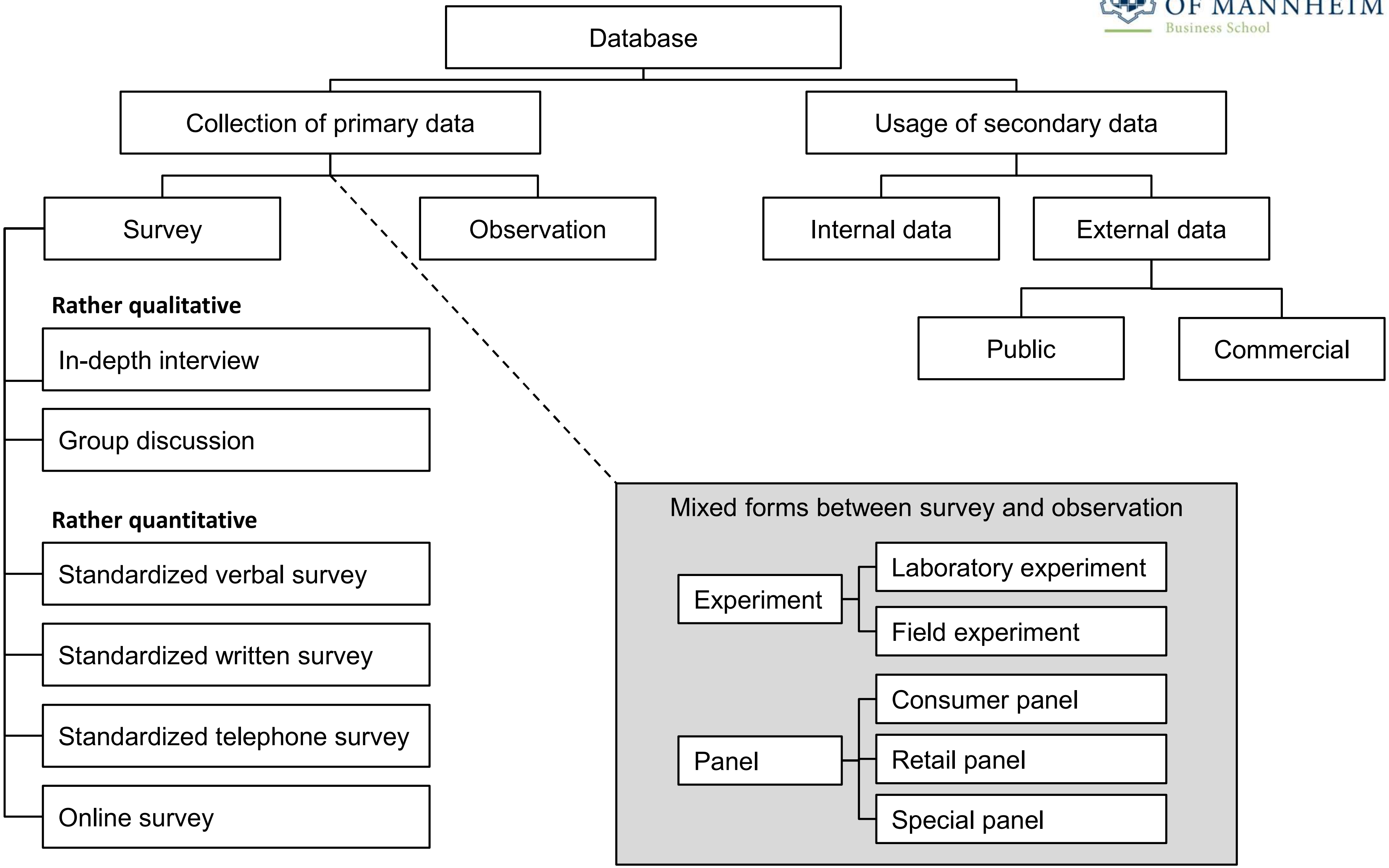
- As opposed to a survey: Behavior of the observed person is directly captured
- However, one cannot capture the motivation behind the action of the observed person
- → Interpretation by the researcher
-

Advantages	**Disadvantages**
• Often the only option	• Observation bias
• Inexpensive	• Hard to reproduce
• No interviewer bias	• Risk of an insufficient observation quality and selective perception

- Examples
  - Hidden camera
  - Personal observer (when visiting a test store or making a test call)
→ Silent shopper (also: mystery shopper) analysis
  - Equipment-based measurements in the context of communication management (e.g., eye-tracking)
  - Observation of click-behavior in the internet (e.g., observation of behavior when making a buying decision/comparison of product alternatives)

Zikmund et al. (2009), pp. 67

Overview of data collection methods



Homburg (2015), p. 280

Primary research: Experiment (1)

Definition

An experiment is characterized as the systematic manipulation of one or more independent variables and the determination of their effect on one or more dependent variables. Hereby, confounding variables can be eliminated or controlled.

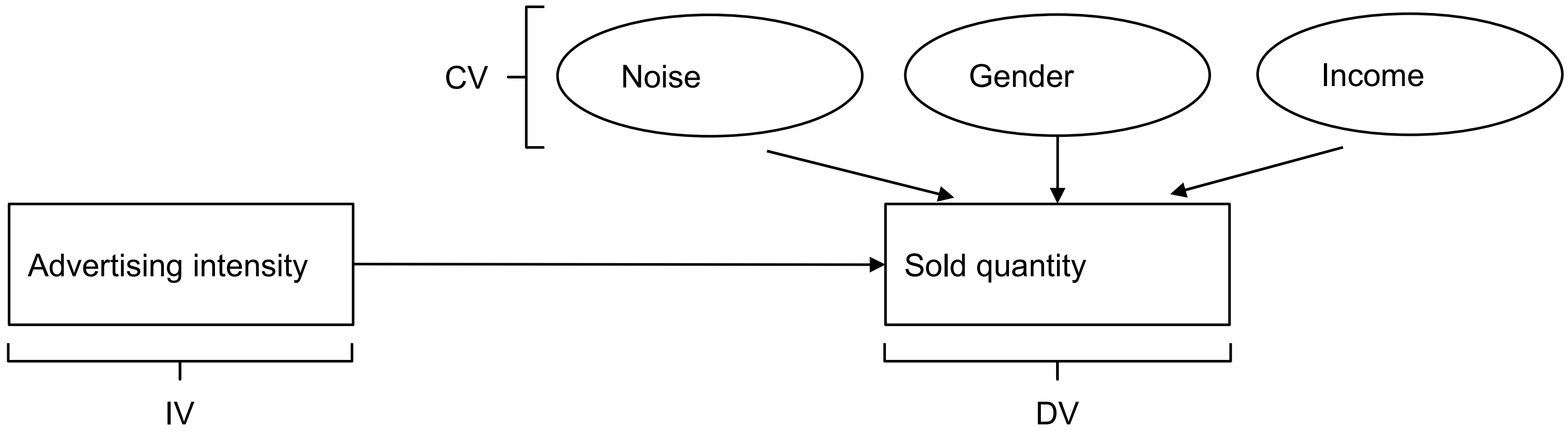
Adapted from Koschate (2008), p. 109

- Mixed form between survey and observation, as usually elements of both methods are apparent
- Central principle: Active variation (manipulation) of one variable and capturing the effect on dependent measures
- Control of confounding variables is very important

Primary research:
Experiment (2) – Types of variables

Types of variables

- Independent variable (IV)
  Variable that is systematically manipulated (e.g., number of sales trainings, price level, advertising intensity)
- Dependent variable (DV)
  Variable for which the effect of the IV should be observed (e.g., number of sold products)
- Confounding variable (CV)
  Confounding variable is any extraneous variable, other than the IV, which may potentially affect the DV and thereby (probably) confound the findings (e.g., awareness of the product brand)



Koschate 2008, p. 109; Zikmund et al. (2009), pp. 59 and pp. 257

Primary research: Experiment (3) – Fundamental design issues

Experimental setting

- Field experiment
- Laboratory experiment

Realistic ⟷ Fictive

Randomization

- True (random) experiment
- Quasi-experiment

Complete ⟷ None

Basis of comparison

- Within Subjects Design
- Between Subjects Design

Same participants ⟷ Different participants

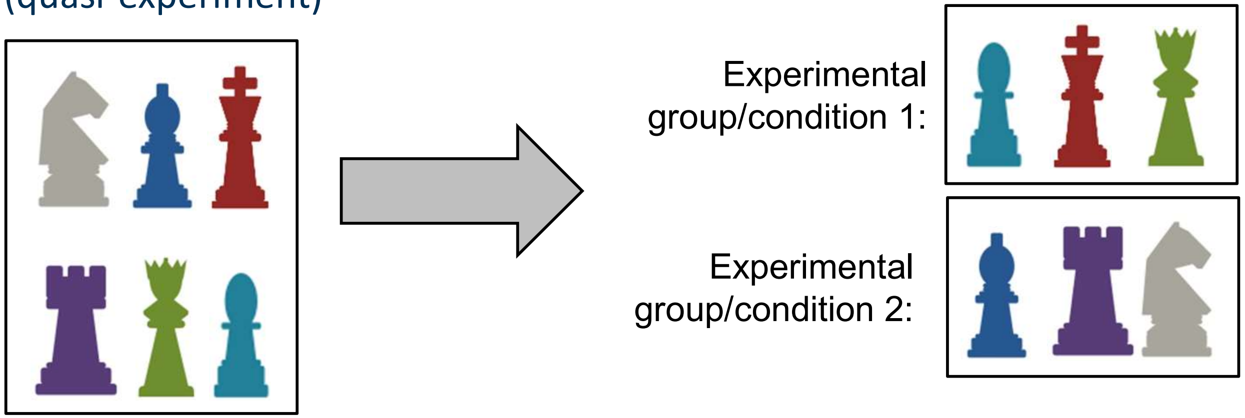
Primary research:
Experiment (4) – Experimental setting

Field experiment	Laboratory experiment
Characteristics: ▪ Takes place in a realistic setting ▪ Subjects are mostly unaware of their participation in an experiment ***Key Advantage:*** ▪ High degree of authenticity increases the possibility of making generalizations about the experimental findings (high external validity) ***Key Disadvantages:*** ▪ Very expensive or impossible to implement ▪ Ethically problematic ▪ Barely replicable (decreases generalizability) ▪ Little control over potential confounding variables (low internal validity) ▪ Manipulation of independent variables often very difficult ***Areas of application:*** ▪ Testing of TV, radio, and print advertising (e.g., after Superbowl) ▪ Regional test market ▪ Micro test markets ▪ Electronic test markets	***Characteristics:*** ▪ Takes place in environment created especially for the experiment ▪ Subjects are aware of their participation in the experiment (although often the actual research question of the study is unknown) ***Key Advantages:*** ▪ Time and costs ▪ High control over potential confounding variables (high internal validity) ▪ Easy to replicate ▪ High level of control over manipulation of the independent variables ***Key Disadvantage:*** ▪ Little authenticity (low external validity) ***Areas of application:*** ▪ Product blind test ▪ Packaging test ▪ Testing of TV and radio advertising ▪ Testing of adverts: e.g. eye-tracking ▪ Test market simulation

Homburg (2015), p. 280

Primary research: Experiment (5) – Randomization

- Participants can be assigned to experimental conditions (e.g., experimental group 1, experimental group 2) either randomized (random experiment) or non-randomized (quasi-experiment)



- **Randomized** can be realized by e.g., a random number table: Assign participants a random number in the order of their appearance. Then allocate participants with even (odd) number to experimental group 1 (2)

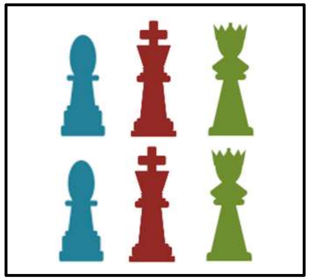
- A non-randomized allocation into groups can e.g., be conducted by considering the first letter of the participants' surnames

- If there is a huge sample size, it is expected that randomization ensures that extraneous variables which could affect the outcome are evenly distributed amongst the groups, in such a manner that there is less systematic differences between the experimental conditions (reduction of confounding variables)

Sarris/Reiß 2005, p. 62, Shadish/Cook/Campbell 2002, pp. 13

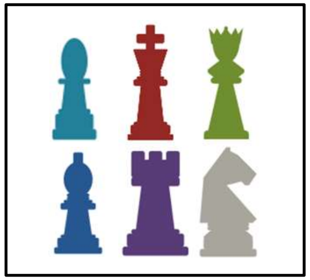
Primary research:
Experiment (6) – Basis of comparison

- Within Subjects Design



	Manipulation of the independent variable	Observation of the dependent variable
Experimental group 1	(X_1): **Advertising message 1**	(O_1): sales
Experimental group 1	(X_2): **Advertising message 2**	(O_2): sales

- Between Subjects Design



	Manipulation of the independent variable	Observation of the dependent variable
Experimental group 1	(X_1): **Advertising message 1**	(O_1): sales
Experimental group 2	(X_2): **Advertising message 2**	(O_2): sales

Primary research: Experiment (7) – Experimental

Definition

Experimental design is defined as a standardized, routinely applicable research plan (scheme or pattern) that underlies and specifies the experimental setup (e.g., number of conditions; kind of experimental treatments; number and kind of dependent variables), the control mechanisms employed (e.g., randomization) and forms the basis of the methodological evaluation of an empirical study with independent and dependent variables.

Sarris 1992, p. 4

Example for an experimental design

"Does the utilization of an advertising message impact the sales of chocolate?"

	Compilation	Manipulation of the independent variable	Observation of the dependent variable
Experimental group	random (R)	(X_A): **advertising message**	(O_1): sales
Control group	random (R)	(X_B): **no advertising message**	(O_2): sales

Primary research:
Experiment (8) – Experimental design

Notation

- X = Manipulation of the independent variable
- O = Observation or measure of the dependent variable
- R = Random allocation of the participants to the experimental groups
-

Pre-experimental designs

- One-Shot Case Study X O
  - No pre-measurement or control group
  - Many possible confounding variables
- One-Group Pretest-Posttest Design O_1 X O_2
  - No control group
  - Many possible influencing factors (e.g. history, maturation, testing)
- Static-Group Comparison
  - Experimental group X O_1
  - Control group O_2
    - No pre-measurement
    - Possibility of a selection bias

Primary research:
Experiment (9) – Experimental design

True experimental design

- Pretest-Posttest Control Group Design

Experimental group	**R**	O_1	**X**	O_2
Control group	**R**	O_3		O_4

→ Desired effect can be calculated by eliminating most effects of possible invalidity

$O_2 - O_1 = TE + H + T + I + R + TM$
$O_4 - O_3 = H + T + I + R + TM$
→ $(O_2 - O_1) - (O_4 - O_3) = TE$

TE = Treatment Effect
H = History
T = Testing
I = Instrumentation
R = Statistical Regression
TM = Test Unit Mortality

- **Example**

Branch group	Observation 1	Treat-ment	Observation 2	Difference
1	12 contracts per week	Yes	19 contracts per week	7
2	12 contracts per week	No	16 contracts per week	4

→ TE = (19 – 12) – (16 – 12) = 3

Homburg (2015), p. 284

Primary research: Panel research (1)

Panel

Definite, consistent circle of people which are in principle surveyed about the same object of study in regular intervals

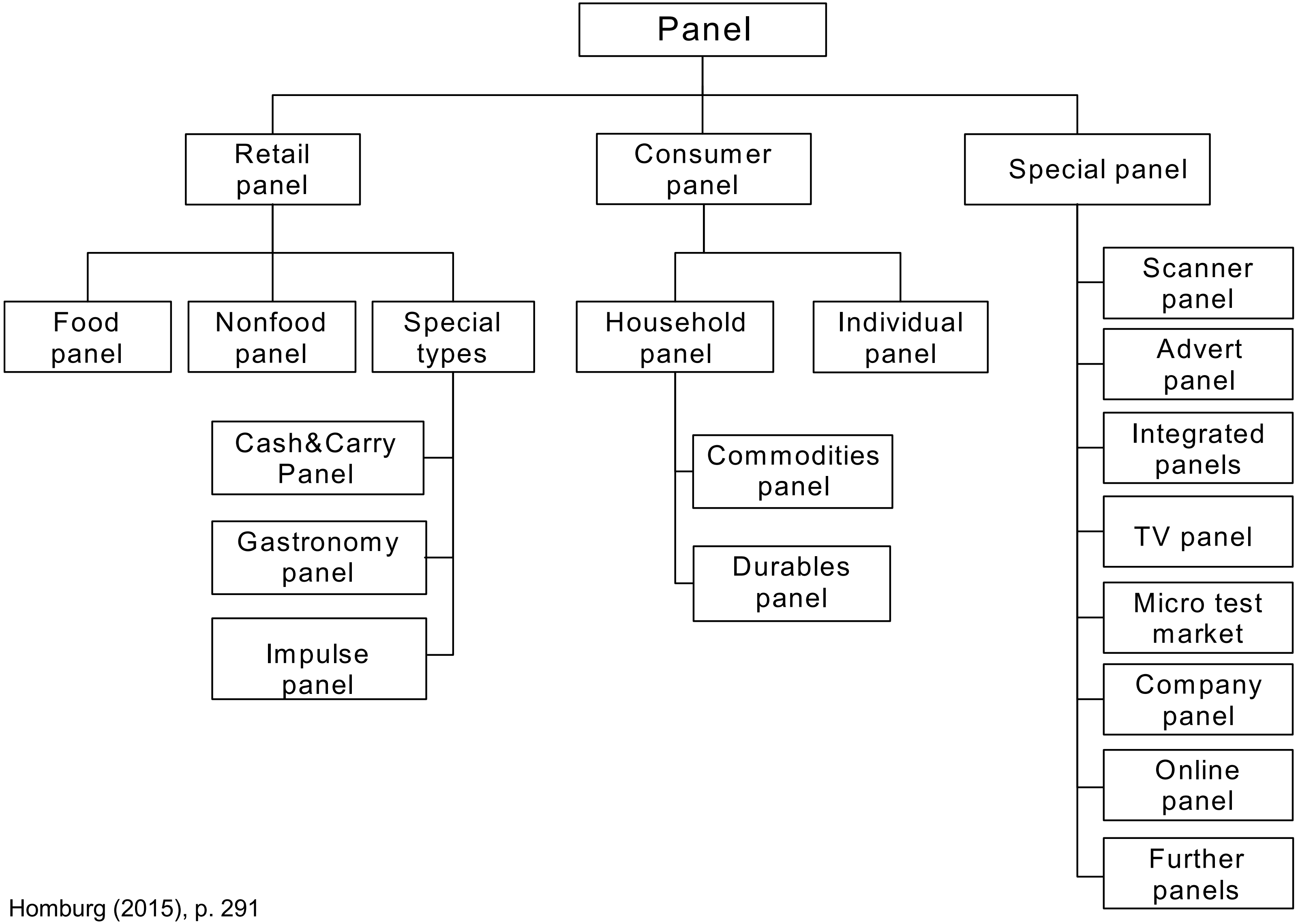
Homburg (2015), p. 290

- Acquisition of longitudinal data (variations during the course of time) and cross-sectional data
- Implementation and maintenance of a panel is time as well as cost intensive
  → market research institutes
- Types of panels (see following slide)

Homburg (2015), p. 291

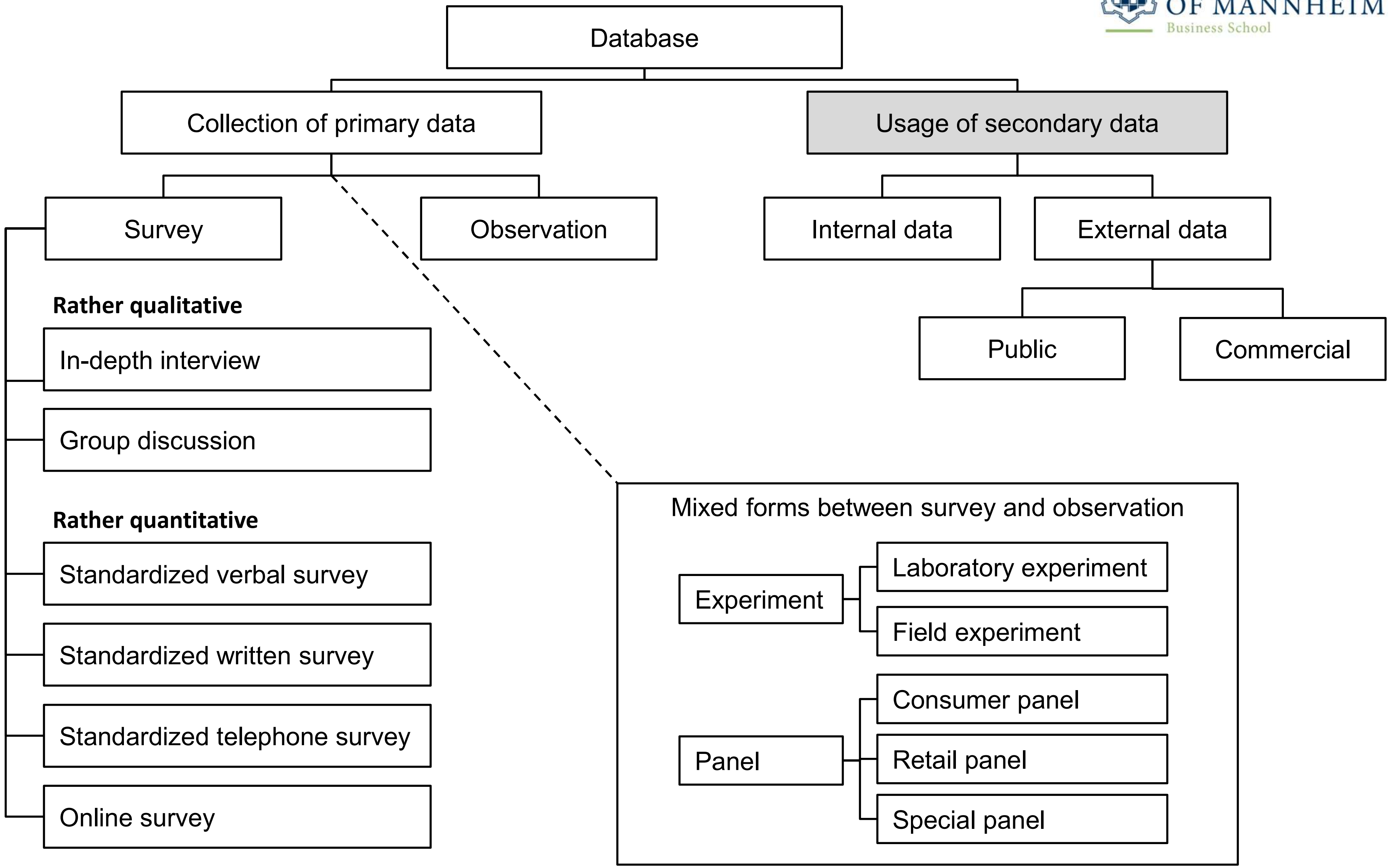
Primary research:
Panel research (2)

- Types of panels



Homburg (2015), p. 291

Overview of data collection methods



Homburg (2015), p.262

Usage of secondary data: Examples of data sources

	Company internal data sources		Company external data sources: non-commercial	Company external data sources: commercial
national	▪ Sales figures ▪ Accounting records (e.g., transfer prices) ▪ Customer and supplier data ▪ Feedback from customers and sales representatives ▪ Data on pricing	**national**	▪ National government ▪ National institute of statistics ▪ Academic institutions ▪ Associations, e.g. Chambers of Commerce and Industry (IHK), VDMA	▪ National market research institutes ▪ Publishing houses (e.g., Gruner + Jahr)
international		**international**	▪ United Nations ▪ Central banks/World Bank ▪ International Monetary Fund (IMF) ▪ World trade organizations ▪ OECD	▪ International market research institutes (e.g., Nielsen) ▪ Internationale agencies ▪ Data bases (e.g., International Market Identifiers, BERI-Index)

Bauer 2009, p. 77ff., Berndt/Fantapié Altobelli/Sandner 2010, p. 56ff; Homburg (2015), p. 295

Overview of online data

Online data

Internet as a source of primary data

- Online focus group
- Online survey
- Online experiment (e.g., A/B testing)
- Crowdsourcing

Internet as a source of secondary data

- Big Data
- Analysis of online communities e.g.,
  - Sentiment Analysis
  - Identification of opinion leaders
- Analysis of other user-generated content (e.g., product reviews)
- Descriptive company statistics (e.g., sales statistics at amazon.com)
- Descriptive industry statistics e.g.
  - U.S. Census Bureau
    European Research Gateway

Online data:
Secondary data – Big Data (1)

Definition

Although it lacks a conclusive definition, big data in general, refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze.

MGI Report (2011)

Characteristics of Big Data

Volume	Velocity	Variety
• Huge amount of data created each day • Drivers for data explosion – Social Media – Multimedia – Growth in transactional databases – Internet of Things	• Real-time information create highest value • Quick response to big data can offer a competitive advantage	• Internal vs. external data sources • Mix of three different data formats – Structured – Unstructured – Semi-Structured

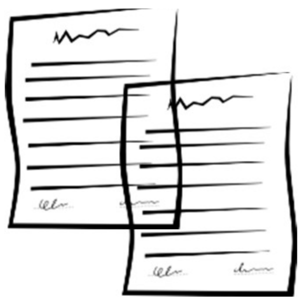
Martin (2012), pp.7; MGI Report (2011), pp.1

Online data:
Secondary data – Sentiment Analysis (1)

Definition

Sentiment analysis is an advanced method for quantifying textual data with the aim of identifying and extracting sentiment (e.g., customer or investor sentiment) in source material. Consumer sentiment measures if a consumer's statement is positive or negative.

Exemplary Posts



- Example 1: “This new drill is just amazing.”
- Example 2: “My drill broke after one day. I hate it!!!!!!!”

?

How do we know which posts express positive and which ones express negative sentiment?

Online data:
Secondary data – Sentiment Analysis (2)

Two-stage procedure

Stage 1: Creation of word lists → **Stage 2: Document analysis**

Step 1: **Pre-processing of data**
Create a list with all words and their number of occurrences in the data set

Step 2: **Generate a balanced training set**
Manually code an equal number of training instances as positive and negative

Step 3: **Edit training data set**
Create a list with all words and their number of occurrences in the training set and eliminate very rare words using a threshold t as the hurdle rate

Step 4: **Classify words according to positive and negative sentiment**
In this step, words that occur more often in positive than in negative posts are treated as positive and vice versa (based on a ratio r)

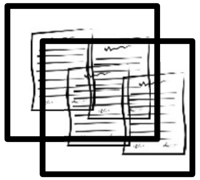
Step 5: **Generate different word lists**
Varying the threshold value t and ratio r, different word lists can be obtained

Witten/Frank/Hall (2011)

Online data:
Secondary data – Sentiment Analysis (4)

UNIVERSITY OF MANNHEIM
Business School

Stage 1: Creation of word lists → **Stage 2: Document analysis**



Input from Stage 1:
Word lists generated in Stage 1 as "learning" input

Step 1:
Machine learning algorithm "learns"

Machine Learning algorithm "learns" inherent text structures and computes the probability of an unseen post being positive or negative for different sub samples of the pre-classified training set.

Step 2:
Evaluation and choice of best word list/ algorithm combination

Step 1 carried out for various feature set size conditions (i.e., different compositions of the optimal word lists)
and usage of different machine learning algorithms

Step 3:
Classification of remaining documents

Obtained values can now be used in further analyses (e.g., regression analyses)

Witten/Frank/Hall (2011)

Step 4: Determination of the data collection method Example shampoo product test

We conduct a standardized written survey since we…

- Need to send out trial product samples by mail

  → costs can be saved by directly attaching the questionnaire to the package

- Need to question a very large sample as we want to test 10 volumizing shampoos

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. **Sampling**
4. Design of the research instrument
5. Data collection
6. Editing and coding of data
7. Analysis and interpretation of data

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	**Sampling/Sample selection**	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Sample – Definition

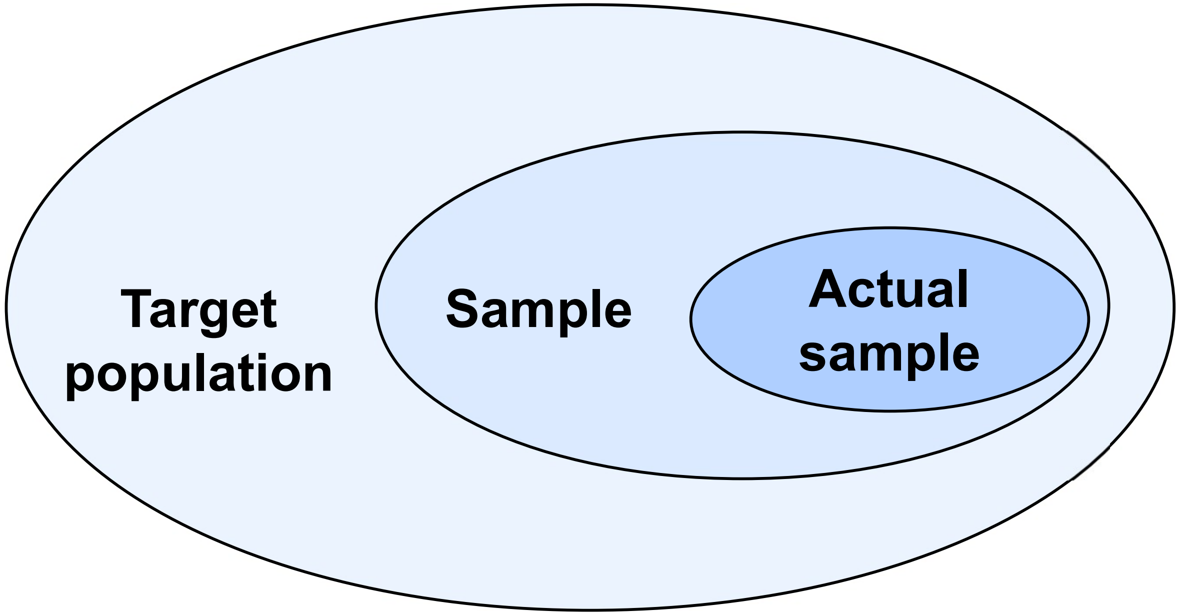
Sample

While the target population refers to any group of entities that share some common set of characteristics and for which the results of market research aim to make an inference, the sample refers to a subset, or some part, of a larger population from which information is gathered during the market research process.

Actual sample

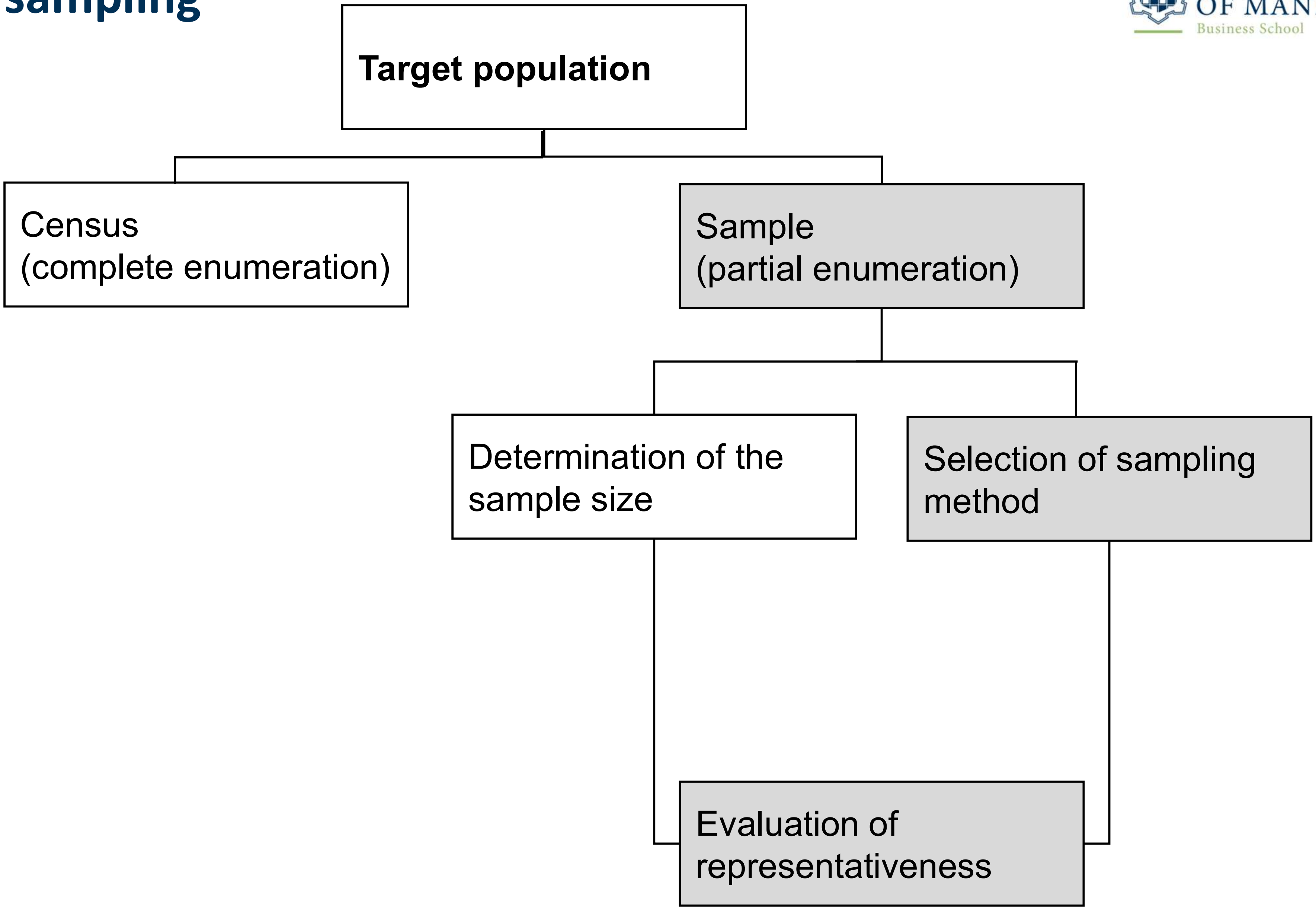
Number of participants from which information is gathered

n(target population) ≥
n(sample) ≥
n(actual sample)

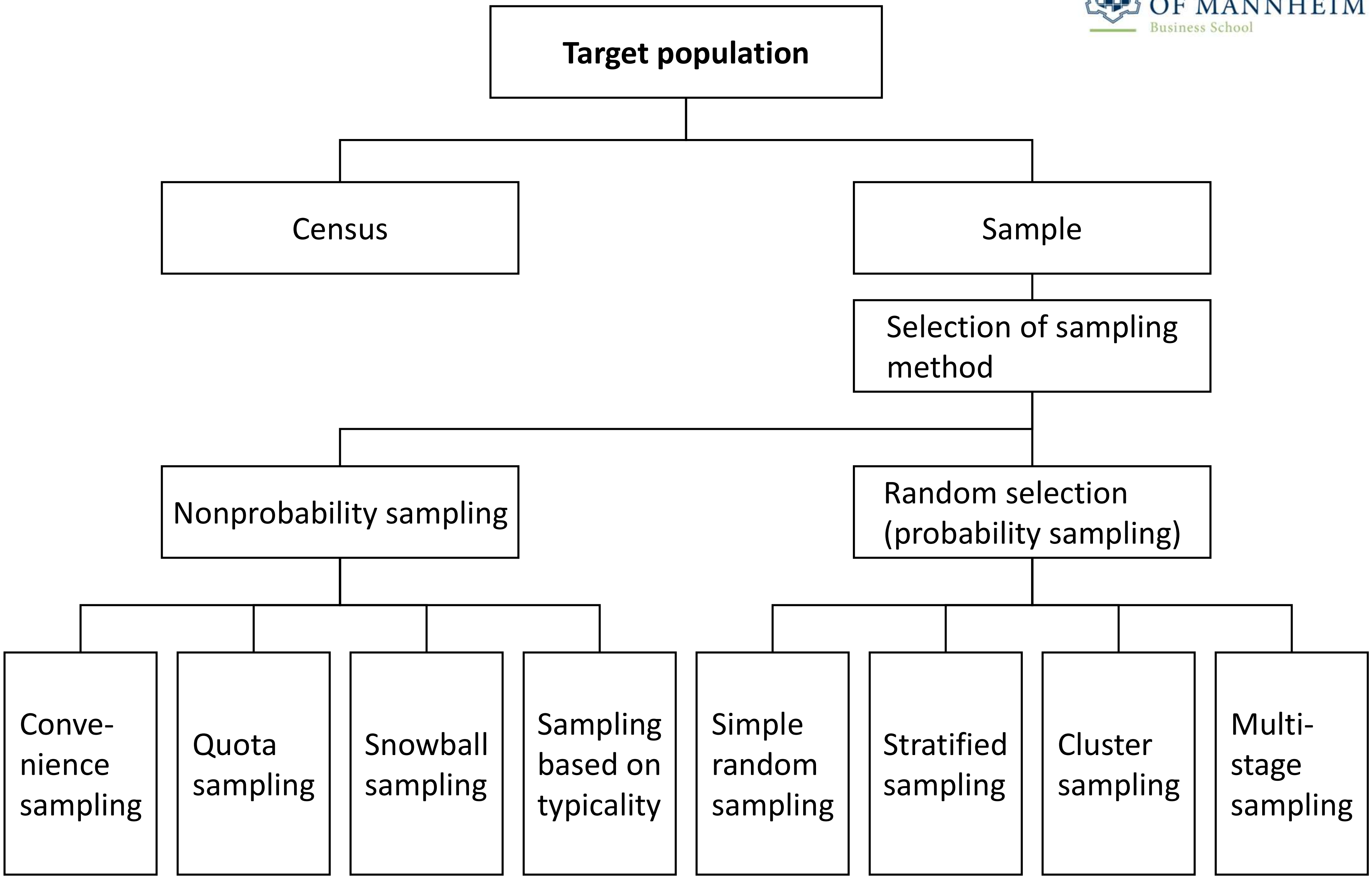


Zikmund et al. (2009), pp. 387

Fundamental decisions with respect to sampling



Sampling procedures



Iacobucci/Churchill 2010, pp. 282

Evaluation of representativeness: Types of error that threaten representativeness

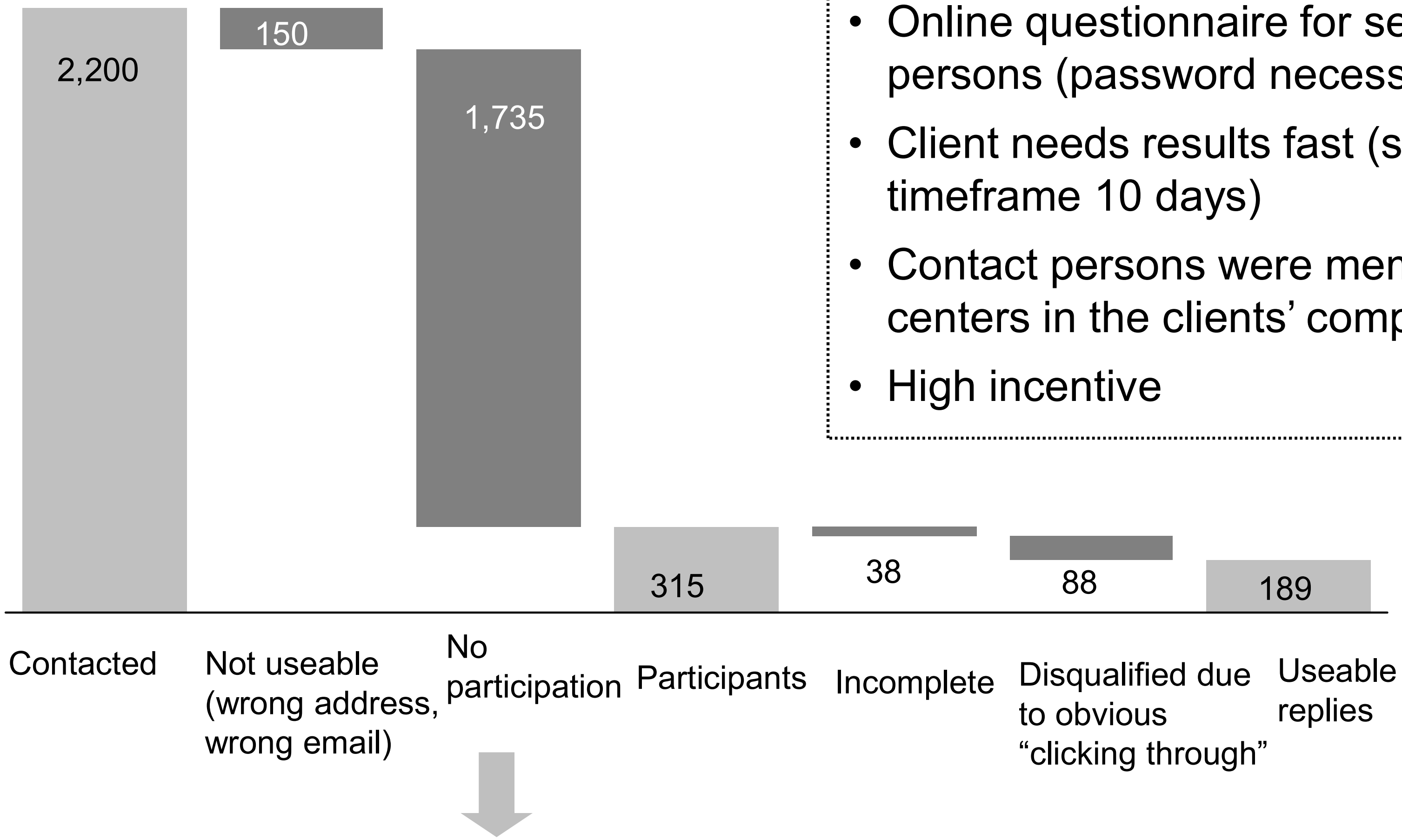
Set of objects	Target population →	sample →	actual sample
Types of errors	Sampling error	Errors due to unbalanced refusal to reply	
Possibilities to check for errors	▪ Examine the structural equivalence of the population and the sample	▪ Examine the structural equivalence of the sample with the effective sample ▪ Early/Late Respondents Test according to Armstrong/ Overton (1977) ▪ Creation of a control group	

Homburg (2015), p. 302

Example:
Survey Non-Response

Survey profile

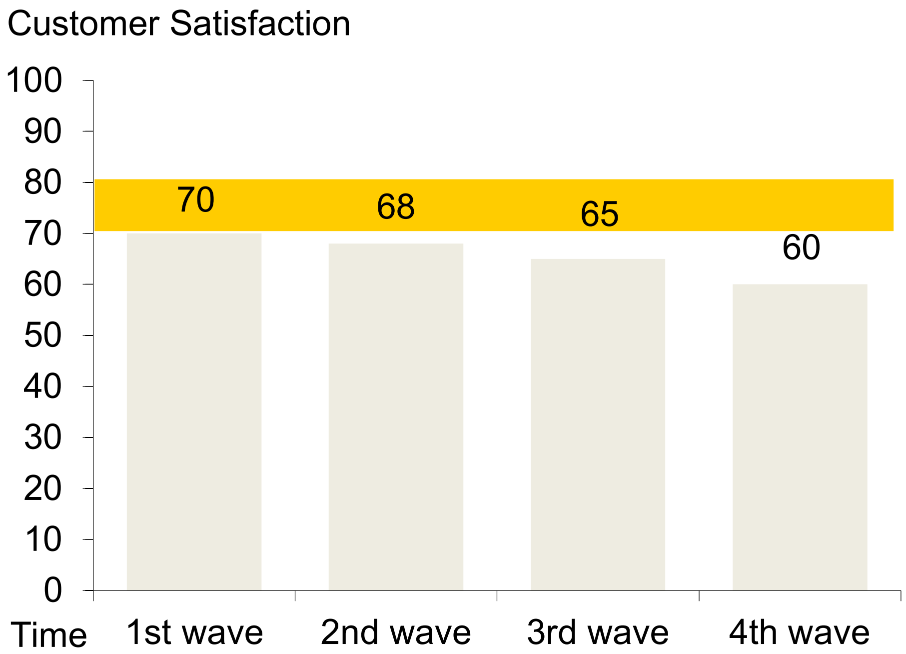
- Online questionnaire for selected contact persons (password necessary)
- Client needs results fast (survey timeframe 10 days)
- Contact persons were members of buying centers in the clients' company
- High incentive



No time for the survey | No contact person due to holidays/carnival | Incentive was not high enough for some contact persons | Missing interest in the topic of the survey

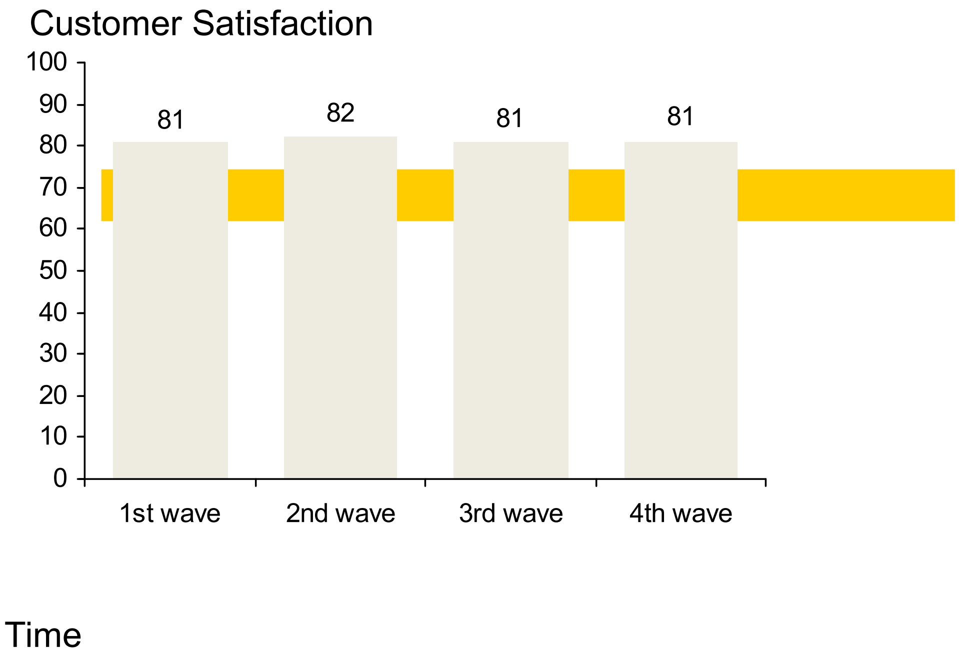
Test for Non-Response-Bias: Early-Late-Respondents-Test

Case a)



Legitimate assumption of a "Non-Response-Bias"

Case b)



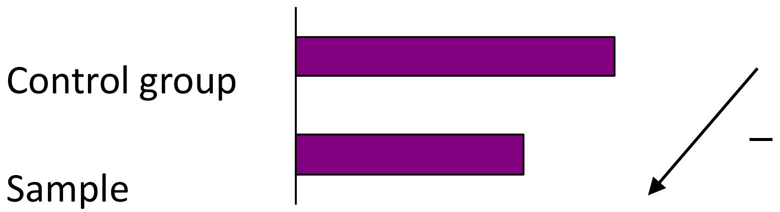
No "Non-Response-Bias"

Test for Non-Response-Bias: Control group with incentive

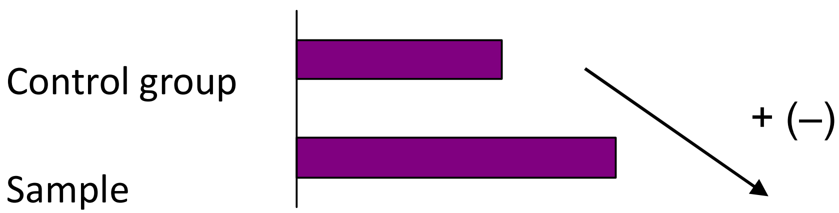
Case a)

Indication
of a non-response-bias

Response rate



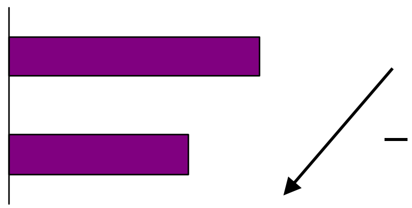
Satisfaction



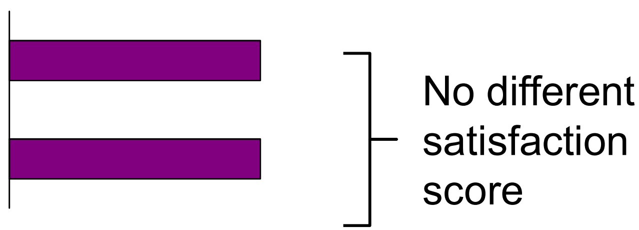
Case b)

No indication
of a non-response-bias

Response rate



Satisfaction



Possible incentives for participation

Type of incentive	Example	Costs	Primary field of application (B2B vs. B2C)	Suitability for low involvement topics
Ex ante incentive	The cover letter already contains a small incentive (e.g., a nice pen)	Low to medium	B2C	High
Payment	10 € for the participation in a mail-intercept-interview	Medium to high	B2C	High
Material incentive	Every participant gets a voucher for a book	Medium to high	B2C	Medium
Lottery	Raffle among all participants to win a flat screen TV	Low	B2C	Medium
Social incentive	For each participant, 10€ are being donated to UNICEF	Medium to high	B2B	Medium
Report of results	Participants receive a report of the results of the survey	Low	B2B	Low
Benchmarking report	Participants receive a report that compares their responses to those of other companies	High	B2B	Low
Workshop	Participants in an expert survey are invited to workshop where results are discussed	Very high	B2B	Low

Step 5: Sampling/Sample selection
Example shampoo product test

Target population

- Women
- Aged 18 – 65 years
- Users of volumizing shampoo

→ Sample (partial enumeration)

→ Determination of the desired sample size approx. 50 for each shampoo

→ Selection of sampling method: Simple random sampling within the market research institute's consumer panel

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. Sampling
4. **Design of the research instrument**
5. Data collection
6. Editing and coding of data
7. Analysis and interpretation of data

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	**Design of the research instrument**	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Operationalization and scale levels

Definitions

Operationalization represents the development of scales for measuring characteristic values of a particular concept/variable.

Scale measurement defines the mathematical characteristics of a scale and thereby the quality of the data to be gathered.

- Distinction of four fundamental scale levels
  - Nominal scale
  - Ordinal scale
  - Interval scale
  - Ratio scale
- Hierarchical order: Higher scales include the features of all lower leveled scales

Zikmund et al. (2009), pp. 293

Scale level and statement quality

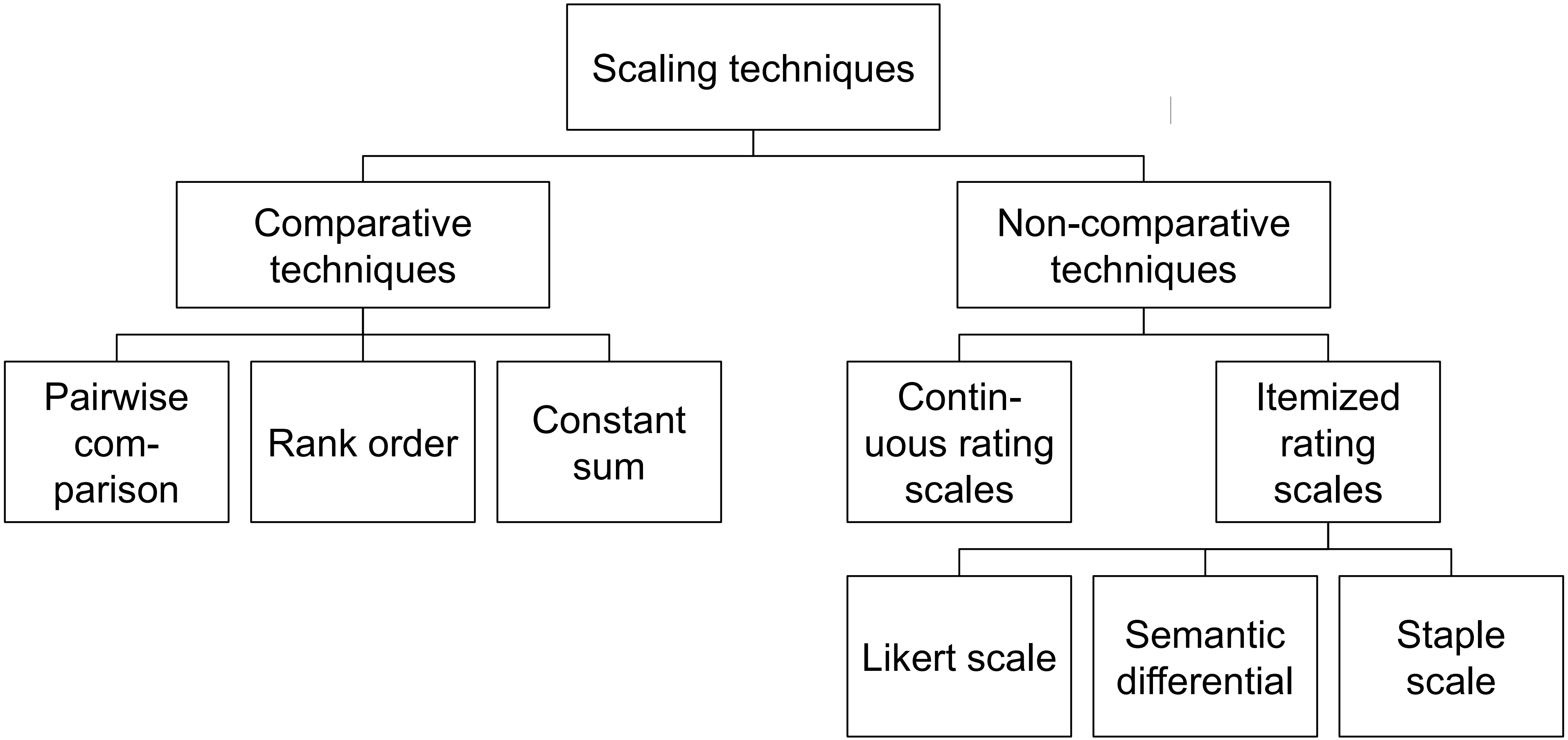
Statement	Scale level			
	Nominal scale	Ordinal scale	Interval scale	Ratio scale
	nonmetric	"quasi-metric"	metric	metric
Equivalence statement: $O_1 = O_2$, $O_1 \neq O_2$	**x**	**x**	**x**	**x**
Ordering statement: $O_1 > O_2$, $O_1 < O_2$		**x**	**x**	**x**
Distance statement: $O_1 - O_2 \geq O_3 - O_4$		**(x)**	**x**	**x**
Ratio statement: $O_1 = b * O_2$				**x**

O = observation

Selected scaling techniques

Scaling techniques determine how scales measure data.

Homburg (2015), p. 311



Homburg (2015), p. 311

Comparative scaling techniques

UNIVERSITY OF MANNHEIM

Pairwise comparisons

- A respondent **compares** 2 of altogether n objects and selects his **preferred alternative**
- Pairwise comparison matrix provides **ordinal data**

Example: Matrix for shampoo brands

	Finesse	Vidal Sas-soon	Head and Shoul-ders	Pert
Finesse	-	0	1	0
Vidal Sassoon	1	-	1	1
Head and Shoulders	0	0	-	0
Pert	1	0	1	-
Number of times preferred	**2**	**0**	**3**	**1**
1: Brand in column was preferred over the brand in row				

Rank order scaling

- A respondent is asked to **rank a number of alternatives** (e.g. according to decreasing utility)

→ Ordinal data

Example: Customer satisfaction criteria

How important are the following criteria for you? (Please rank with 1 = very important and 5 = very unimportant)

Criteria	**Rank** Product
quality	________
Order procedure	________
Technical customer service	________
Sales team/Support	________
Documentation/ Information	________

Constant sum technique

- Respondents allocate a **constant sum of units** (e.g., 100 points) to different attributes of an object (or different alternatives)
- Conclusions about the importance of different attributes are possible

Example: Customer satisfaction criteria

How important are the following four attributes for you when buying a software product? Please allocate a total of 100 points

Performance (number of tools)	_____
Price	_____
Compatibility with existing software	_____
Applicability of existing data	_____
Sum	100

Non-comparative scaling techniques: Itemized rating scales

Likert scale

- Respondents indicate their extent of agreement with a statement
- Assessment on a bipolar rating scale
- On average, an ordinal scale with mostly 5 to 7 scale points

Example:

How satisfied are you with the clarity of your invoice?

How do you rate your invoice?

Very dissatisfied ☐ ☐ ☐ ☐ ☐ Very satisfied

The invoice is fair.

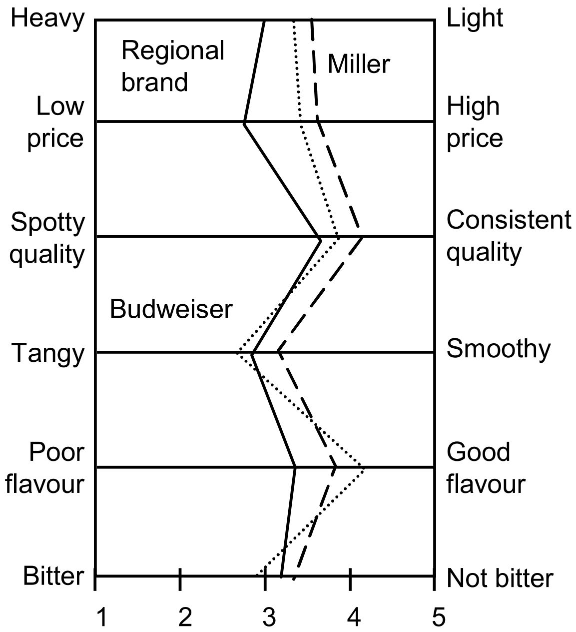
Very expensive ☐ ☐ ☐ ☐ ☐ Very low priced

Strongly disagree ☐ ☐ ☐ ☐ ☐ Strongly agree

Semantic differential

- 5 or 7 point rating scale with end points associated with bipolar labels that have semantic meaning
- Multiple bipolar rating scales with adjective opposite pairs

Example: Profile portray of three beer brands



Stapel scale

- Modification of the semantic differential scale
- Adjectives are tested singularly with a unipolar point scale
- Points in the scale are identified by numbers

Example: Evaluation of the working atmosphere (exemplary)

Stapel scale

Extent of agreement

Open	1	2	3	4	5
Discrete	1	2	3	4	5
Withdrawn	1	2	3	4	5

Comparison: Semantic differential

Extent of agreement

Open	1	2	3	4	5	Withdrawn
Open	1	2	3	4	5	Discrete

Step 6: Design of the research instrument
Example shampoo product test

Measurement scale

- Nominal: e.g., status of the hair (normal, greasy, dry, greasy hairline – dry hair ends)
- Ordinal (quasi-metric): e.g., extent of agreement on various characteristics of the shampoo
- Ratio scales: e.g., age of the respondent

Scaling technique

- 7-point-Likert scales

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. Sampling
4. Design of the research instrument
5. **Data collection**
6. Editing and Coding of data
7. Analysis and interpretation of data

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	**Data collection**	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Step 7: Data Collection
Example shampoo product test

Package with one trial product each and attached questionnaire is sent out to 10 x 55 respondents

↓ *2 weeks of product testing +*
2 weeks for filling out the questionnaire

After shampoo was tested, follow up call or letter

↓ *max. 12 weeks*

Letter of appreciation enclosed
with participation incentives

↓

If applicable, share report with findings and implications within company

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. Sampling
4. Design of the research instrument
5. Data collection
6. **Editing and coding of data**
7. Analysis and interpretation of data

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	**Editing and coding of data**	How can the data be processed using (numerical) codes?
9	Analysis and interpretation of data	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Editing

- **Objective**
  Editing ensures
  - The availability
  - The readability and
  - The absence of errors of the needed data

- **Basis**
  Collected information from questionnaires, written protocols, data from telephone and online surveys

- Check whether it is possible to evaluate questionnaires with respect to
  - Completeness
  - Correct way of replying
  - Contradictory way of responding
  - Inaccuracy due to distortion
  - Wrong respondent

→ **If necessary, collect additional data through surveys and follow-up questions, or sort out questionnaires**

Coding

- Idea and Procedure
  - Technical procedure to categorize raw data
  - Classification of raw data in response categories and transformation in numbers to enable the analysis of the data
  - No automatic classification, but an interpretation through a coding person is required

- Relevant issues when constructing response categories (codes)
  - Mutual exclusion of response categories
  - Completeness of response categories
  - Detailed coding to avoid loss of information and to improve the analytical potential
  - Simple processing of numerical codes
  - Utilization of standard codes for missing values is advisable (e.g., -999)
  - Coding of intervals (size group)
    - Number of intervals is appropriate
    - Size of the categories is appropriate
  - Coding of open questions is difficult (only possible after having reviewed all occurring responses)

Step 8: Editing and coding of data
Example shampoo product test

# of question in the questionnaire	Variable	Code	Characteristic
18	Status of the hair	hair_status	1= normal 2 = greasy 3 = dry 4 = greasy hairline – dry hair ends
19	Structure of the hair	hair_struc	1 = fine/thin 2 = normal/medium 3 = thick/strong
20	Age in Years	Age	1 = 18-39 2 = 40-65
…			

Course outline

1. Fundamentals of market research
2. Determination of the data collection method
3. Sampling
4. Design of the research instrument
5. Data collection
6. Editing and coding of data
7. **Analysis and interpretation of data**

The process of market research

1	Problem formulation	What is the goal of the study? (e.g., analysis of customer satisfaction and loyalty)?
2	Determination of the research design	With which type of study can we attain our goals? (descriptive, exploratory, explanatory)?
3	Determination of the executing researcher	Who should conduct the study (in-house vs. outsourcing to a market research institute)?
4	Determination of the data collection method	What type of data collection method is suitable?
5	Sampling/Sample selection	What methods should be employed to identify the sample? Which scope and/or size should the sample have?
6	Design of the research instrument	How should the questionnaire, the interviewer manual, or the experimental design be devised?
7	Data collection	What should the executing researcher pay attention to while collecting data?
8	Editing and coding of data	How can the data be processed using (numerical) codes?
9	**Analysis and interpretation of data**	Which data analysis methods should be applied?
10	Presentation of the results	How can the data be presented in a meaningful way (complexity and depth)?

Homburg (2015), p. 251

Summary of the market research process for a shampoo product test

1. Problem formulation: We are the manufacturer of Shampoo A and want to compare the performance of our shampoo with those of nine competitive shampoos
2. Determination of research design: explorative and explanatory
3. Determination of the executing researcher: market research institute
4. Determination of the data collection method: standardized written survey
5. Sampling/Sample selection: simple random sampling
6. Design of the research instrument:
   - Scales of measurement: nominal, ordinal and ratio scales
   - Scaling techniques: 7-point-Likert scales
7. Data collection: survey participants answer the questionnaire after having tested one out of 10 different volumizing shampoos
8. Editing and coding of data
9. Analysis and interpretation of data: see following slides

Classifying data analysis methods

- According to the number of variables involved
  - Univariate analyses
  - Bivariate analyses
  - Multivariate analyses

- According to the type of conclusions
  - Descriptive analyses
  - Inductive analyses

Overview of uni- and bivariate descriptive statistics

UNIVERSITY OF MANNHEIM
Business School

Descriptive Statistics

- Univariate
  - Measures of centrality
    - Mean
    - Median
    - Percentiles
    - Mode
  - Measures of spread
    - Variance
    - Standard deviation
    - Coefficient of variation
    - Mean absolute deviation
  - Measures of form
    - Skewness
    - Kurtosis
- Bivariate
  - Measures of centrality
  - Measures of spread
    - Contingency tables
    - Covariance
    - Correlation coefficient
  - Measures of form
    - (Bivariate skewness)
    - (Bivariate kurtosis)

Homburg (2015), pp. 327

Overview of uni- and bivariate inductive statistics

Inductive Statistics (Hypothesis tests)

	univariate			bivariate		
	Tests for central tendency	Tests for spread	Tests for form	Tests for central tendency	Tests for spread	Tests for form
Parametric tests	• One-sample t-test • Test of proportions	• χ2-test for a population variance	-----	• t-test for two population means • Test for mean of differences of paired data or paired t-test	• t-test for independence	-----
Non-parametric tests	• Sign test	-----	• χ^2-test for goodness-of-fit	• Wilcoxon rank sum test or Mann-Whitney U-Test	• χ^2-test for independence	• χ2 -test for homogeneity

Homburg (2015), pp. 327

Overview of multivariate data analysis methods

Analysis of Interdependence

- Metric variables
  - Factor Analysis
    - Exploratory
    - Confirmatory
  - Cluster Analysis (metric model)
- Nonmetric variables
  - Cluster Analysis (nonmetric model)

Analysis of Dependence

- Dependent variables, metric
  - Inde-pendent variables non-metric
    - Multivariate Analysis of Variance
  - Inde-pendent variables metric
    - Multiple Regression Analysis
    - Structural Equation Modeling with latent variables
- Dependent variables, nonmetric
  - Inde-pendent variables metric
  - Inde-pendent variables non-metric

- Conjoint Analysis

Homburg (2015), p. 359; Hair et al. (2010), p. 12

Factor Analysis
Introduction (1)

Analysis of Interdependence

Metric variables

Factor Analysis
- Exploratory
- Confirmatory

Definition

Factor analysis is a prototypical multivariate, interdependence technique that statistically identifies a reduced number of factors from a larger number of measured variables

Zikmund et al. (2009), p. 593

Factor Analysis: Introduction (2)

Definition

A factor is a linear combination of the original variables. Factors also represent the underlying dimensions (constructs) that summarize or account for the original set of observed variables.

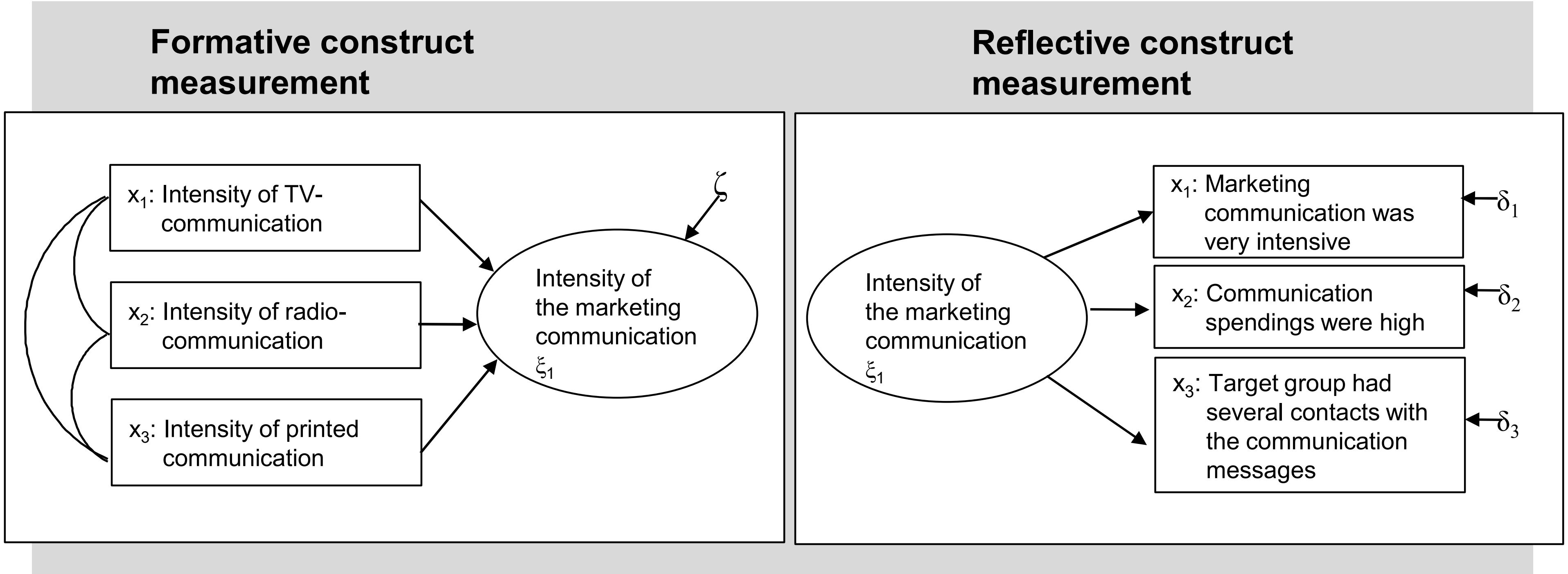
•Hair et al. (2010), p. 92

- Factors are usually latent constructs like attitude or satisfaction, or an index like social class
- A factor is measured through several variables (multi-item measurement) instead of using a single variable (single-item measurement)
- We assume a causal relationship between a factor and its variables
  → Measurement theory
  – Reflective measurement
  – Formative measurement

Factor Analysis: Measurement theory (1)

UNIVERSITY OF MANNHEIM
Business School

Decision on measurement theory



- Formative measurement theory: Measured variables cause the construct
- Reflective measurement theory: Latent constructs cause the measured variables

Klarmann (2008); Hair et al. (2010), pp.701

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Factor Analysis:
Measurement theory (2)

UNIVERSITY OF MANNHEIM
Business School

Distinguishing between reflective and formative constructs

	Indicative of	
	Reflective	**Formative**
Causality of construct	Items are caused by construct	Construct is formed from items
Conceptual relationship among items	All items are related conceptually because they have a common cause	No requirement of conceptual linkage to other items
Domain of items	Representative sample of potential items	Exhaustive inventory of all possible items
Covariance among items	Expected collinearity among items	No expectation of collinearity. High collinearity among formative items can be problematic
Internal consistency	Required	Not required
Forms of construct validity	Internal and external	Only external

Hair et al. (2010), p.753

Exploratory Factor Analysis: Example shampoo product test – Introduction (1)

UNIVERSITY OF MANNHEIM
Business School

Below we have listed various characteristics that may apply to the shampoo you have tested. To which extent do you agree with the following statements?

strongly disagree — strongly agree

The shampoo tested...	1	2	3	4	5	6	7
gives hair long lasting volume							
has a pleasant scent							
has a fresh scent							
keeps hair voluminous till the evening							
has a natural scent							
makes hair visibly voluminous							
has a foam which is nice to apply							
foams up particularly softly							
gives hair a natural shine							
gives hair particularly more volume							
gives hair volume all day long							
makes hair look healthy							
gives hair a silky touch							
makes hair looking well-cared							
has a particular pleasant foam							

Exploratory Factor Analysis: Example shampoo product test – Introduction (2)

Below we have listed various characteristics that may apply to the shampoo you have tested. To which extent do you agree with the following statements?

The shampoo tested...	1 (strongly disagree)	2	3	4	5	6	7 (strongly agree)
gives hair long lasting volume	○	○	○	○	○	○	○
has a pleasant scent	○	○	○	○	○	○	○
has a fresh scent	○	○	○	○	○	○	○
keeps hair voluminous till the evening	○	○	○	○	○	○	○
has a natural scent	○	○	○	○	○	○	○
makes hair visibly voluminous	○	○	○	○	○	○	○
has a foam which is nice to apply	○	○	○	○	○	○	○
foams up particularly softly	○	○	○	○	○	○	○
gives hair a natural shine	○	○	○	○	○	○	○
gives hair particularly more volume	○	○	○	○	○	○	○
gives hair volume all day long	○	○	○	○	○	○	○
makes hair look healthy	○	○	○	○	○	○	○
gives hair a silky touch	○	○	○	○	○	○	○
makes hair looking well-cared	○	○	○	○	○	○	○
has a particular pleasant foam	○	○	○	○	○	○	○

Problem

- We want to interpret the respondents' answers regarding the shampoos' characteristics, e.g., in terms of their importance for the overall satisfaction
- However, the large number of characteristics (variables) in the study makes interpretations difficult and complex
- We want to explore whether interdependencies between the variables exist

Exploratory Factor Analysis

Factor Analysis: Introduction (3)

- **Goal**
  - Examination of underlying relationships for a large number of variables
  - Determination whether the information can be condensed or summarized in a smaller set of factors
- **Application**
  - Identification of broader dimensions (= factors) for segmentation (e.g., market segmentation)
  - Use of identified factors in methods of analysis of dependence e.g., regression analysis
- **Procedure**
  1. Create the data matrix
  2. Calculate of the input data (correlation matrix)
  3. Determine the number of factors to be retained
  4. Rotate and interpret the factors

Hair et al. (2010), pp. 91

Exploratory Factor Analysis: Procedure (1)

Step 1: Create the data matrix

Step 2: Calculate the input data (correlation matrix)

- Calculation of correlations r_{ij} i, j = 1, 2, ..., n
- Correlation matrix constitutes the data set on which analyses are based
  - Assumption that some underlying structure exists
  - Consideration of strength and direction of correlations

Hair et al. (2010), pp. 100

Exploratory Factor Analysis:
Example shampoo product test – Procedure (2)

Step 2: Calculate the correlation matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. gives hair longlasting volume	1	0.199***	0.219***	0.896***	0.257***	0.813***	0.318***	0.194***	0.419***	0.882***	0.877***	0.247***	0.171***	0.223***	0.260***
2. has a pleasant scent		1	0.882***	0.213***	0.857***	0.131**	0.222***	0.239***	0.286***	0.159***	0.142**	0.308***	0.280***	0.270***	0.303***
3. has a fresh scent			1	0.245***	0.858***	0.173***	0.241***	0.256***	0.323***	0.199***	0.190***	0.338***	0.301***	0.298***	0.303***
4. keeps hair voluminous till the evening				1	0.293***	0.783***	0.357***	0.202***	0.431***	0.838***	0.924***	0.266***	0.166***	0.236***	0.275***
5. has a natural scent					1	0.201***	0.291***	0.258***	0.376***	0.242***	0.212***	0.392***	0.356***	0.338***	0.355***
6. makes hair visibly voluminous						1	0.361***	0.205***	0.360***	0.855***	0.778***	0.146**	0.093*	0.157***	0.196***
7. has a foam which is nice to apply							1	0.609***	0.404***	0.385***	0.335***	0.364***	0.300***	0.335***	0.630***
8. foams up particularly softly								1	0.412***	0.259***	0.175***	0.318***	0.287***	0.298***	0.537***
9. gives hair a natural shine									1	0.485***	0.374***	0.645***	0.528***	0.568***	0.417***
10. gives hair particularly more volume										1	0.819***	0.265***	0.195***	0.250***	0.264***
11. gives hair volume all day long											1	0.213***	0.121**	0.194***	0.240***
12. makes hair look healthy												1	0.687***	0.692***	0.441***
13. gives hair a silky touch													1	0.666***	0.394***
14. makes hair looking well-cared														1	0.460***
15. has a particular pleasant foam															1

Computed correlation used pearson-method with listwise-deletion; $^{*}p<0.1$ $^{**}p<0.05$ $^{***}p<0.01$

Exploratory Factor Analysis: Procedure (3)

Step 3a: Determine the number of factors to be retained – Extraction of factors

- **Goal**: Extraction of uncorrelated factors, each of them accounting for as much variance of the original variables as possible

- **Multi-stage process** *(supplementary background information)*

Formulation of a maximization problem: Search for the linear combination *z* of *J* original variables with maximum variance $z = u_1 \cdot x_1 + u_2 \cdot x_2 + ... + u_J \cdot x_J \rightarrow \max Var(z)$
↓
Setting an additional restriction, e.g. $\sum_{j=1}^{J} u_j^2 = 1$, otherwise maximization with infinite large u_j possible
↓
Solving the maximization problem (usually requires the help of statistical software in which numerical methods for solving problems of the matrix algebra are implemented)
↓
Estimation of the residual variation of J original variables which is not determined by the factors

Iteration of the process until total variance of the original variables is explained (maximal K=J factors)

- **Results**
  - Factors z_k and the (unstandardized) factor loadings of the original variables (u_{jk})
  - Factor loadings represent the degree of association (correlation) of each variable with each factor

Exploratory Factor Analysis:
Example shampoo product test – Procedure (4)

Step 3a: Determine the number of factors to be retained – Extraction of factors *(cont'd)*

Total Variance Explained

Factor_n	Eigenvalue	variance_accounted_for	cum_variance_accounted_for
1	6.34687395	0.423124930	0.4231249
2	2.96242939	0.197495293	0.6206202
3	1.79076845	0.119384563	0.7400048
4	1.22870271	0.081913514	0.8219183
5	0.51995407	0.034663604	0.8565819
6	0.38872873	0.025915249	0.8824972
7	0.35092226	0.023394817	0.9058920
8	0.32070349	0.021380233	0.9272722
9	0.29727740	0.019818493	0.9470907
10	0.23714172	0.015809448	0.9629001
11	0.14344932	0.009563288	0.9724634
12	0.14108934	0.009405956	0.9818694
13	0.11627703	0.007751802	0.9896212
14	0.08717093	0.005811395	0.9954326
15	0.06851120	0.004567414	1.0000000

Exploratory Factor Analysis: Procedure (5)

Step 3b: Determine the number of factors to be retained – Decide on the number of factors

Latent Root Criterion

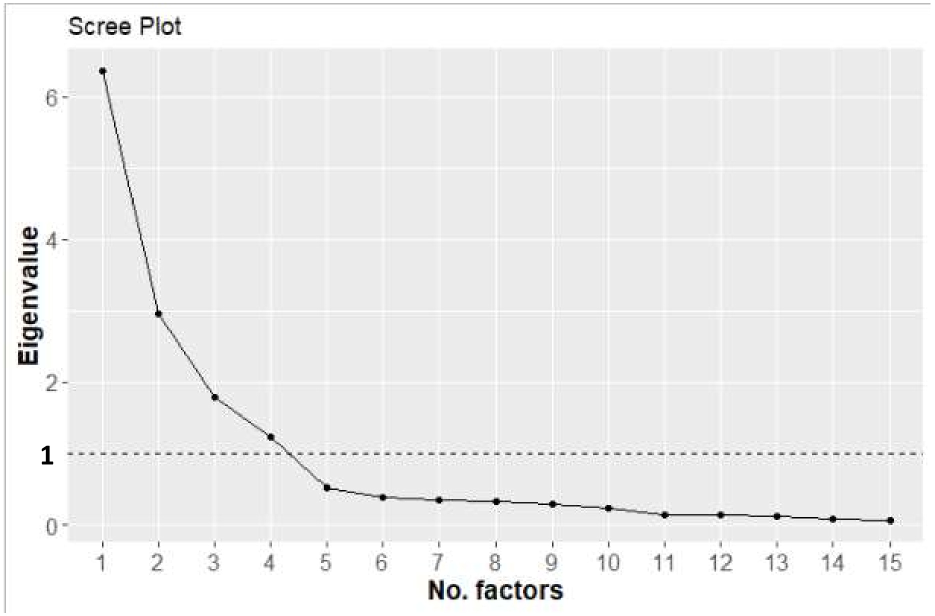
- Factors with eigenvalues (= sum of squared factor loadings) > 1 will be retained
- Any individual factor should account for the variance of at least a single variable (eigenvalue corresponds to the variance of the factor)
- The variance of one single variable is 1 (based on standardization)

→ Factors with eigenvalue ≤ 1 explain less variance than one single variable

Scree Test Criterion

- Sorts eigenvalues in descending order based on their value within coordination system
- Shape of the curve is used to evaluate the cutoff point (elbow) and determines the number of factors that need to be extracted

Hair et al. (2010), pp. 108 and pp. 134



Exploratory Factor Analysis:
Example shampoo product test – Procedure (6)
Step 4: Rotate and interpret the factors

Rotated Component Matrix with Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4	*Communality*
gives hair longlasting volume	**0.93**	0.10	0.13	0.09	*0.9*
has a pleasant scent	0.07	**0.92**	0.14	0.12	*0.89*
has a fresh scent	0.10	**0.90**	0.18	0.13	*0.88*
keeps hair voluminous till the evening	**0.92**	0.13	0.14	0.12	*0.89*
has a natural scent	0.14	**0.86**	0.24	0.15	*0.85*
makes hair visibly voluminous	**0.85**	0.06	0.03	0.16	*0.76*
has a foam which is nice to apply	0.24	0.09	0.19	**0.79**	*0.73*
foams up particularly softly	0.10	0.12	0.20	**0.68**	*0.53*
gives hair a natural shine	0.33	0.16	**0.61**	0.27	*0.58*
gives hair particularly more volume	**0.90**	0.06	0.16	0.17	*0.86*
gives hair volume all day long	**0.91**	0.07	0.09	0.11	*0.85*
makes hair look healthy	0.11	0.17	**0.82**	0.19	*0.75*
gives hair a silky touch	0.03	0.16	**0.77**	0.16	*0.64*
makes hair looking well-cared	0.10	0.14	**0.77**	0.20	*0.66*
has a particular pleasant foam	0.12	0.18	0.34	**0.64**	*0.57*

Interpret the factors

(1) Volumizing Performance
(2) Characteristics of Scent
(3) Shiny/Healthy Appearance of Hair
(4) Foaming Properties

Exploratory Factor Analysis:
Example shampoo product test – Procedure (7)
Step 4: Rotate and interpret the factors

- **Communality of one variable = sum of squared factor loadings across all factors**
  - Amount of variance in a variable that is accounted for by all factors taken together
  - Size of communality = useful index for assessing how much variance in a particular variable is accounted for by the factor solution
  - Example for Variable 1: “Gives hair long lasting volume”
    Communality = .90 → 90 % of variance of variable 1 is explained by the 4-factor solution

- **Percentage of trace of a factor = eigenvalue/number of variables**
  - Percentage of variance of variables explained by a single factor
  - Trace = sum of the eigenvalues of the variable set
  - Example percentage of trace for Factor 1 (Volumizing Performance):
    Percentage of trace = $(0.93^2+0.07^2+0.10^2+\ldots+0.12^2)/15 = 28.76\%$
    → Factor 1 explains 28.76 % of the variance of all 15 variables

- **Total percentage of trace = sum of percentage of trace of a factor**
  - Useful index for the overall solution to determine how well a factor solution accounts for what all the variables together represent
  - Example total percentage of trace for 4-factor solution:
    total percentage of trace = 0.2876+0.1732+0.1738+0.1216 = 0.7562=75.62%
    → Index is high and the variables are highly related to one another

Hair et al. (2010), pp. 135

Exploratory Factor Analysis: Example shampoo product test – Summary

What was the initial problem?

- We conducted a survey in which respondents were asked to evaluate a shampoo on 15 different characteristics (= variables)
- The large number of variables made it difficult to derive any interpretation on these characteristics
- Therefore, we wanted to find underlying relationships between these variables to condense them into factors

What did we do?

- With the help of statistical software (e.g., R Statistics), we …
  - created a data matrix and calculated the correlation coefficients between all variables
  - calculated the total variance explained and determined the number of factors by using the latent root criterion / scree test criterion
  - rotated the factors
- We named the factors by interpreting their variables

What was the outcome?

- Four factors summarize the characteristics on which the shampoos were evaluated in the survey

(1) Volumizing Performance
(2) Characteristics of Scent
(3) Shiny/Healthy Appearance of Hair
(4) Foaming Properties

Confirmatory Factor Analysis:
Example shampoo product test – Introduction (1)

Below we have listed various characteristics that may apply to the shampoo you have tested. To which extent do you agree with the following statements?

The shampoo tested...	1 (strongly disagree)	2	3	4	5	6	7 (strongly agree)
gives hair long lasting volume							
has a pleasant scent							
has a fresh scent							
keeps hair voluminous till the evening							
has a natural scent							
makes hair visibly voluminous							
has a foam which is nice to apply							
foams up particularly softly							
gives hair a natural shine							
gives hair particularly more volume							
gives hair volume all day long							
makes hair look healthy							
gives hair a silky touch							
makes hair looking well-cared							
has a particular pleasant foam							

Problem

- We still have a large number of variables which make interpretation difficult due to the high complexity
- But this time, we make a priori theoretical assignments of the variables to pre-specified factors (constructs)
- We want to **test/confirm** in how far our measurement theory reflects the actual data

Confirmatory Factor Analysis

Confirmatory Factor Analysis: Introduction (2)

- **Goal**:
  Testing the extent to which a researcher's a priori, theoretical assignment of indicators to pre-specified constructs represents the actual data
- **Application**:
  Applied when complex constructs are used that cannot be measured directly (e.g., in consumer behavior)
- **Methodological principle**:
  Covariance structure analysis (models of confirmatory factor analysis correspond to the measurement models of structural equation modeling)
- Before results can be computed, researcher must specify conceptually:
  - Number of factors
  - Assignment of single variables to factors

  → Distinction to exploratory factor analysis

Hair et al. (2010), pp. 91

Confirmatory Factor Analysis: Introduction (3) – Comparison of EFA and CFA

	Exploratory factor analysis	Confirmatory factor analysis
Model	No model formulation	Theoretical model formulation before results can be computed
Goal	Exploration of factors as cause for highly correlated variables → Model for structure exploration	Testing the relationship between indicator variables and hypothetical constructs → Model for structure confirmation
Assignment of indicator variables to factors	Underlying pattern of the data determine the factor structure by statistical criteria	Specified by the researcher in advance
Number of factors	Determined by statistical criteria during the analysis	Specified by the researcher in advance
Rotation of factor loading matrix	Performed to facilitate the interpretation of the factors	Not required (construct structure is specified in advance)
Factor Interpretation	Afterwards with support of the matrix of factor loadings	Determined by the use of constructs in advance (specified by researcher)

Weiber/Mühlhaus (2009); Hair et al. (2010), pp.693

Confirmatory Factor Analysis: Procedure (1) – Overview

UNIVERSITY OF MANNHEIM
Business School

1. Developing the overall measurement model

2. Parameter estimation

3. Assessing measurement model validity

4. If necessary: model respecification

5. Interpretation of results

Homburg (2015), p. 365; Hair et al (2010), pp. 693

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Confirmatory Factor Analysis:
Example shampoo product test – Procedure (2)

Step 1: Developing the overall measurement model

To which extent do you agree with the following statements? The shampoo tested...

strongly disagree … strongly agree

(1) Volume Performance	1	2	3	4	5	6	7
gives hair long lasting volume							
keeps hair voluminous till the evening							
makes hair visibly voluminous							
gives hair particularly more volume							
gives hair volume all day long							

(2) Shiny/Healthy Appearance of Hair							
gives hair a natural shine							
makes hair look healthy							
gives hair a silky touch							
makes hair look well-cared							

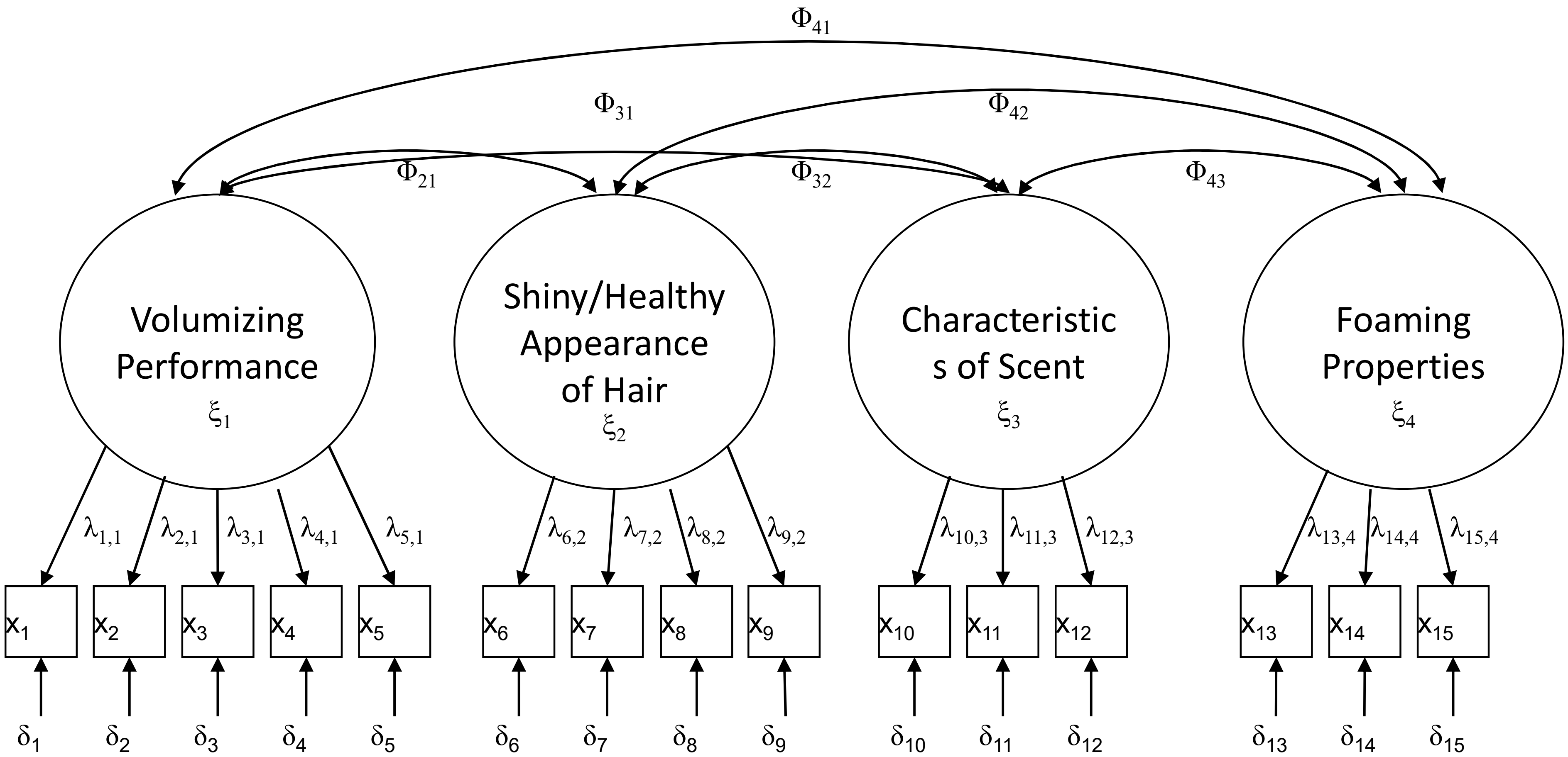
(3) Characteristics of Scent							
has a pleasant scent							
has a fresh scent							
has a natural scent							

(4) Foaming Properties							
has a foam which is nice to apply							
foams up particularly softly							
has a particular pleasant foam							

Confirmatory Factor Analysis:
Example shampoo product test – Procedure (3)

Step 1: Developing the overall measurement model *(cont'd)*

Illustration of the measurement model



Confirmatory Factor Analysis: Example shampoo product test – Procedure (4)

Step 1: Developing the overall measurement model *(cont'd)*

Matrix notation of the measurement model in the example

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \\ x_{15} \end{pmatrix} = \begin{pmatrix} \lambda_{1,1} & 0 & 0 & 0 \\ \lambda_{2,1} & 0 & 0 & 0 \\ \lambda_{3,1} & 0 & 0 & 0 \\ \lambda_{4,1} & 0 & 0 & 0 \\ \lambda_{5,1} & 0 & 0 & 0 \\ 0 & \lambda_{6,2} & 0 & 0 \\ 0 & \lambda_{7,2} & 0 & 0 \\ 0 & \lambda_{8,2} & 0 & 0 \\ 0 & \lambda_{9,2} & 0 & 0 \\ 0 & 0 & \lambda_{10,3} & 0 \\ 0 & 0 & \lambda_{11,3} & 0 \\ 0 & 0 & \lambda_{12,3} & 0 \\ 0 & 0 & 0 & \lambda_{13,4} \\ 0 & 0 & 0 & \lambda_{14,4} \\ 0 & 0 & 0 & \lambda_{15,4} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15} \end{pmatrix}$$

Symbol	Pronoun-ciation	Meaning
x_i		Measured variable i (= used as indicators of latent constructs)
δ_i	Delta	Error term associated with an estimated measured variable
λ_{ij}	Lambda	Factor loading
ξ_j	Ksi	Construct (latent concept)
Φ_{jk}	Phi	Covariance between set of constructs

Homburg/Klarmann/Pflesser (2008); Hair et al. (2010) p.683; Lattin, Carroll, and Green (2003), p. 176

Confirmatory Factor Analysis:
Example shampoo product test – Procedure (5)
Step 2: Parameter estimation

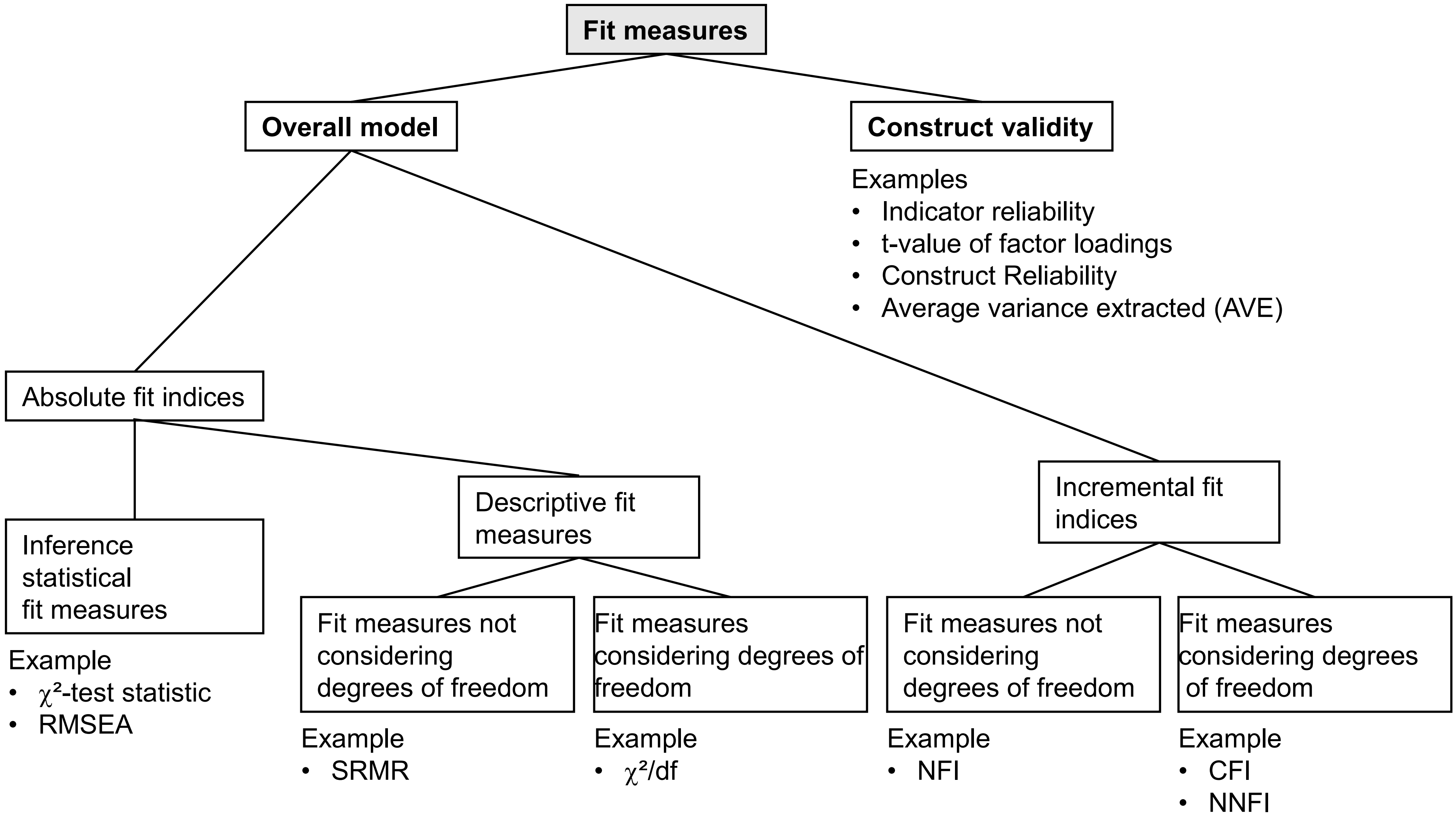
Factor (Construct)	Item No.	Indicator Reliability	Standardized Factor loadings	z-value of factor loadings	Construct reliability	AVE
Factor 1: Volumizing Performance	1 4 6 10 11	0.90 0.90 0.73 0.82 0.88	1 1.14 0.94 0.96 1.10	 45.71 31.13 37.75 42.67	.96	.85
Factor 2: Shiny/Healthy Appearance of Hair	9 12 13 14	0.54 0.75 0.61 0.65	1 1.38 1.05 1.16	 18.33 16.71 17.19	.87	.65
Factor 3: Characteristics of Scent	2 3 5	0.88 0.88 0.84	1 0.98 0.94	 38.88 36.29	.95	.87
Factor 4: Foaming Properties	7 8 15	0.67 0.51 0.61	1 0.76 1.03	 15.19 16.40	.81	.60

Fit measures for the overall model

χ^2	470.353
Degrees of freedom (df)	84
RMSEA	0.097
NNFI	0.928
CFI	0.943
SRMR	0.060
χ^2/df	5.60

Confirmatory Factor Analysis: Procedure (6)

Step 3: Assessing measurement model validity – Overview



Homburg/Klarmann/Pflesser (2008); Hair et al.(2010), pp. 664 and pp. 708

Confirmatory Factor Analysis: Procedure (7)

Step 3: Assessing measurement model validity – Overall model

- **Incremental fit indices**
  – Assessing how well a specified model fits relative to some alternative baseline model. The baseline model is a null model (= all measured variables are unrelated to each other)
- **Absolute fit indices**
  – χ^2- test: Fundamental measure of differences between observed and estimated covariance matrices → Summarizes the fit of a model quite well

$$\chi^2 = (N-1) \cdot \mathrm{F}(\mathrm{S}, \Sigma)$$

N: Overall sample size
F: Discrepancy function F(,)
$\hat{S}$: Observed covariance matrix of measured variables
Σ: Estimated covariance matrix by the model of measured variables

Homburg/Klarmann/Pflesser (2008); Hair et al. (2010) p.665

Confirmatory Factor Analysis: Procedure (8)

Step 3: Assessing measurement model validity – Overall model *(cont'd)*

Absolute fit indices *(cont'd)*

- **RMSEA** (Root Mean Squared Error of Approximation)
  - Represents how well a model fits a population, not just a sample used for estimation

$$\text{RMSEA} = \sqrt{\left(\frac{\chi^2 - \text{df}}{\text{df}(N-1)}\right)}$$

df: Degrees of freedom

- **SRMR** (Standardized Root Mean Square Residual)
  - Measures the deviation of the elements of the observed covariance matrix from the elements of the covariance matrix generated by the model

$$\text{SRMR} = \sqrt{\frac{2\sum_{i=1}^{p}\sum_{j=1}^{i}\left(\frac{s_{ij} - \sigma_{ij}}{s_{ii}s_{jj}}\right)^2}{p(p+1)}}$$

s_{ij}: Elements of the observed covariance matrix S
σ_{ij}: Elements of the estimated covariance matrix Σ by the model
p: Number of indicator variables

Homburg/Klarmann/Pflesser (2008); Hair et al. (2010) pp. 667

Confirmatory Factor Analysis: Procedure (9)

Step 3: Assessing measurement model validity – Construct validity

- **Indicator reliability** (= variance extracted of the item)

  The extent to which a measured variable's variance is explained by a latent factor (and not due to measurement errors)

$$\text{rel}(x_i) = \frac{\lambda_{ij}^2 \phi_{jj}}{\lambda_{ij}^2 \phi_{jj} + \theta_{ii}}$$

- **Construct reliability** (CR) and **Average Variance Extracted** (AVE)

  Extent of reliability about all measured variables which form a construct

$$\mathrm{CR}(\xi_j) = \frac{\left(\sum_{i=1}^{k} \lambda_{ij}\right)^2 \phi_{jj}}{\left(\sum_{i=1}^{k} \lambda_{ij}\right)^2 \phi_{jj} + \left(\sum_{i=1}^{k} \theta_{ii}\right)} \qquad \mathrm{AVE}(\xi_j) = \frac{\sum_{i=1}^{k} \lambda_{ij}^2 \phi_{jj}}{\sum_{i=1}^{k} \lambda_{ij}^2 \phi_{jj} + \sum_{i=1}^{k} \theta_{ii}}$$

Homburg/Klarmann/Pflesser; Hair et al. (2010), pp. 632 and pp.708; Lattin, Carroll, and Green 2003, p. 190

Confirmatory Factor Analysis: Procedure (10)

Step 3: Assessing measurement model validity

Goodness-of-Fit measures	Threshold values
RMSEA	≤ .05 (respectively .10)
NFI	≥ .9
NNFI	≥ .9
CFI	≥ .9
SRMR	≤ .05 (respectively .10)
χ^2/df	≤ 3
Indicator reliability (IR)	≥ .4
Construct reliability (CR)	≥ .6
Average variance extracted (AVE)	≥ .5
Significance test of factor loadings	t ≥ 1.645
χ^2-difference statistic	χ^2- difference ≥ 3.841
Fornell-Larcker criterion	AVE (ξ_i) > squared correlation (ξ_i, ξ_j) for all i ≠ j

Homburg/Klarmann/Pflesser (2008,) p. 288

Confirmatory Factor Analysis:
Example shampoo product test – Procedure (11)
Step 3: Assessing measurement model validity

Factor (Construct)	Item No.	Indicator Reliability	Standardized Factor loadings	z-value of factor loadings	Construct reliability	AVE
Factor 1: Volumizing Performance	1 4 6 10 11	0.90 0.90 0.73 0.82 0.88	1 1.14 0.94 0.96 1.10	 45.71 31.13 37.75 42.67	.96	.85
Factor 2: Shiny/Healthy Appearance of Hair	9 12 13 14	0.54 0.75 0.61 0.65	1 1.38 1.05 1.16	 18.33 16.71 17.19	.87	.65
Factor 3: Characteristics of Scent	2 3 5	0.88 0.88 0.84	1 0.98 0.94	 38.88 36.29	.95	.87
Factor 4: Foaming Properties	7 8 15	0.67 0.51 0.61	1 0.76 1.03	 15.19 16.40	.81	.60
Fit measures for the overall model						
χ^2	470.353					
Degrees of freedom (df)	84					
RMSEA	0.097					
NNFI	0.928					
CFI	0.943					
SRMR	0.060					
χ^2/df	5.60					

Confirmatory Factor Analysis: Procedure (12)

Step 4: Modification of measurement model

- Modification has an exploratory rather than confirmatory character
- Elimination of single indicators from the measurement model
  - Indicator reliability ≥ .4 and path coefficients differ significantly from 0, *otherwise*
  - Successive elimination of the relevant indicator
- Modification of construct structure
  - Construct reliability for each construct ≥ .6 and AVE ≥ .5, *otherwise:*
  - Successive inclusion or elimination of indicators

Weiber/Mühlhaus 2009

Confirmatory Factor Analysis: Example shampoo product test – Summary

What was the initial problem?

- We conducted a survey with 15 different shampoo characteristics (variables) which makes the interpretation difficult and complex
- However, we had an a priori, theoretical assignment of the variables to the pre-specified factors: (1) Volumizing Performance, (2) Shiny/Healthy Appearance of Hair, (3) Characteristics of Scent, (4) Foaming Properties

What did we do?

- We developed the overall measurement model based on our a priori assignment of the variables
- Using a statistical software (e.g., R statistics), we…
  - Estimated the parameters by the help of different fit measures.
  - Assessed the measurements model validity (no respecification was needed)

What was the outcome?

We confirmed the a priori specified model with four factors

(1) Volumizing Performance

(2) Shiny/Healthy Appearance of Hair

(3) Characteristics of Scent

(4) Foaming Properties

Analysis of Interdependence: Cluster Analysis (1)

Analysis of Interdependence

Nonmetric variables

- Cluster Analysis

Cluster Analysis

Goal

- Complexity reduction by dividing a large group of observations into smaller groups (clusters)
- Observations within each cluster should be relatively similar (i.e., possessing largely the same characteristics) and the observations in different clusters should be relatively dissimilar

Application

- Market segmentation: Clustering of buyers according to their buying behavior (e.g., intensity of information search, attitudes)
- Market structure analysis: Clustering of stores based on store-switching behavior of consumers
- Clustering of alternative providers based on their strategic behaviors (e.g., pricing strategy, portfolio range)

Lattin et al. (2003), pp. 264

Analysis of Interdependence: Cluster Analysis (3)

Procedure

1. Selection of cluster variables and creating a data matrix
2. Creating a distance matrix
3. Eliminating outliers
4. Selecting a cluster algorithm
5. Determining the number of clusters
6. Interpretation and labeling of the clusters

Lattin et al. (2003), pp. 264

Analysis of Interdependence: Cluster Analysis (4)

Step 1: Selection of Cluster Variables and Creating a Data Matrix

- Cluster variables
  - No fixed recommended scale level
  - As independent from each other as possible
  - Since different dimensions should be comparably scaled, variables are usually standardized
  - Distinction between active and passive cluster variables
- Display of the characteristics of the cluster variables in a data matrix

Analysis of Interdependence: Cluster Analysis (5)

UNIVERSITY OF MANNHEIM
Business School

Step-by-step Example: Clustering by Product Attributes

You have ten products A,…,J and average product ratings across customers for four product attributes based on a 1 to 7 Likert Scale.

Your goal is to reduce the complexity of the reporting of your product range to upper management by clustering products that perform well for similar attributes.

Product	Attribute 1	Attribute 2	Attribute 3	Attribute 4
A	4.9	5.5	5.7	6.1
B	4.8	5.3	4.5	5.8
C	5.1	5.4	5.4	5.8
D	5.2	5.3	5.3	5.5
E	4.3	5.3	3.9	5.5
F	5.1	5.6	5.8	6.2
G	4.6	5.2	4.7	5.9
H	5.2	5.6	5.6	5.9
I	5.3	5.6	5.5	5.6
J	4.8	5.8	4.2	5.9

Analysis of Interdependence: Cluster Analysis (6)

UNIVERSITY OF MANNHEIM
Business School

Step 2: Creating a Distance Matrix

Determining the similarity or distance with proximity measures

Variables with a nominal scale level – a binary variable structure

	Object 2		**Row sum**
Object 1	**Characteristic exists (1)**	**Characteristic does not exist (0)**	
Characteristic exists (1)	a	c	a + c
Characteristic does not exist (0)	b	d	b + d
Column sum	a + b	c + d	m

General similarity function

$$s_{ij} = \frac{a + \delta \cdot d}{a + \delta \cdot d + \lambda(b + c)}$$

s_{ij} : similarity between the objects i and j
δ, λ : possible weighting factor

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Example:

- A bank would like to determine the degree of similarity between some of its ten most important private customers (customer 1 to 10). Therefore, they collect data regarding five binary coded characteristics (previous complaints, academic education, annual income, duration of business relationship, gender) which they assume to be of relevance and obtain the following information for their customers (with 1 indicating that the characteristic exists):

	Characteristics				
Customer	Any previous complaints yes (1) / no (0)	Academic education yes (1) / no (0)	Annual income > 100.000 Euro yes (1) / no (0)	Duration of business relationship yes (1) / no (0)	gender female (1) / male (0)
1	0	1	0	0	0
2	1	0	1	0	0
3	1	0	1	0	0
4	1	0	0	1	0
5	1	0	1	1	1
6	1	1	1	0	0
7	0	1	1	1	1
8	1	0	1	1	1
9	1	1	1	1	1
10	1	0	1	1	1

Example (2):

Please focus on customer 1, 4, and 8. Particularly, you are interested in which of these customers are most similar to each other. Therefore, you consider the distance matrix for binary variable structure based on the given data for customer pair 1 (customer 1 and customer 4) as well as for customer pair 2 (customer 1 and customer 8) and for customer pair 3 (customer 4 and 8). Which pair is most similar? Please use the general similarity function with δ=0.7 and λ=0.3.

Hint: All calculations and steps of calculations must be evident. Round your results for each calculation to two decimal places. To determine similarity, use the following formula:

$$s_{ij} = \frac{a + \delta d}{a + \delta d + \lambda(b + c)}$$

with generally:

a = characteristic exists for both objects
b = characteristic exists for object b, but not for object a
c = characteristic exists for object a, but not for object b
d = characteristic does not exist for both objects

Example (3):
Solution:

Customer 1,4: $\frac{0+0.7*2}{0+0.7*2+0.3*3} = 0.61$

Customer 1, 8: $\frac{0+0.7*0}{0+0.7*0+0.3*5} = 0$

Customer 4, 8: $\frac{2+0.7*1}{2+0.7*1+0.3*2} = 0.82$

Customer pair 4 and 8 is the most similar.

Analysis of Interdependence: Cluster Analysis (7)

Examples for similarity measures for variables with metric

Properties

Usually, distance measures

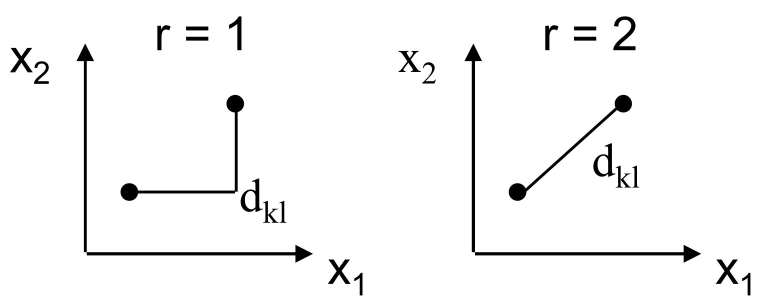
- Minkowski Metric (general form)

$$d_{ij}(r) = \left[\sum_k \left| x_{ik} - x_{jk} \right|^r \right]^{1/r}$$

$d_{i,j}$: Distance between the objects i and j
x_{ik}, x_{jk} : Value of the variables k at object i, j
$r \geq 1$: Minkowski constant

Derived distance measures

- r = 1 → City Block
- r = 2 → Euclidean distance



Lattin et al. (2003), pp. 273

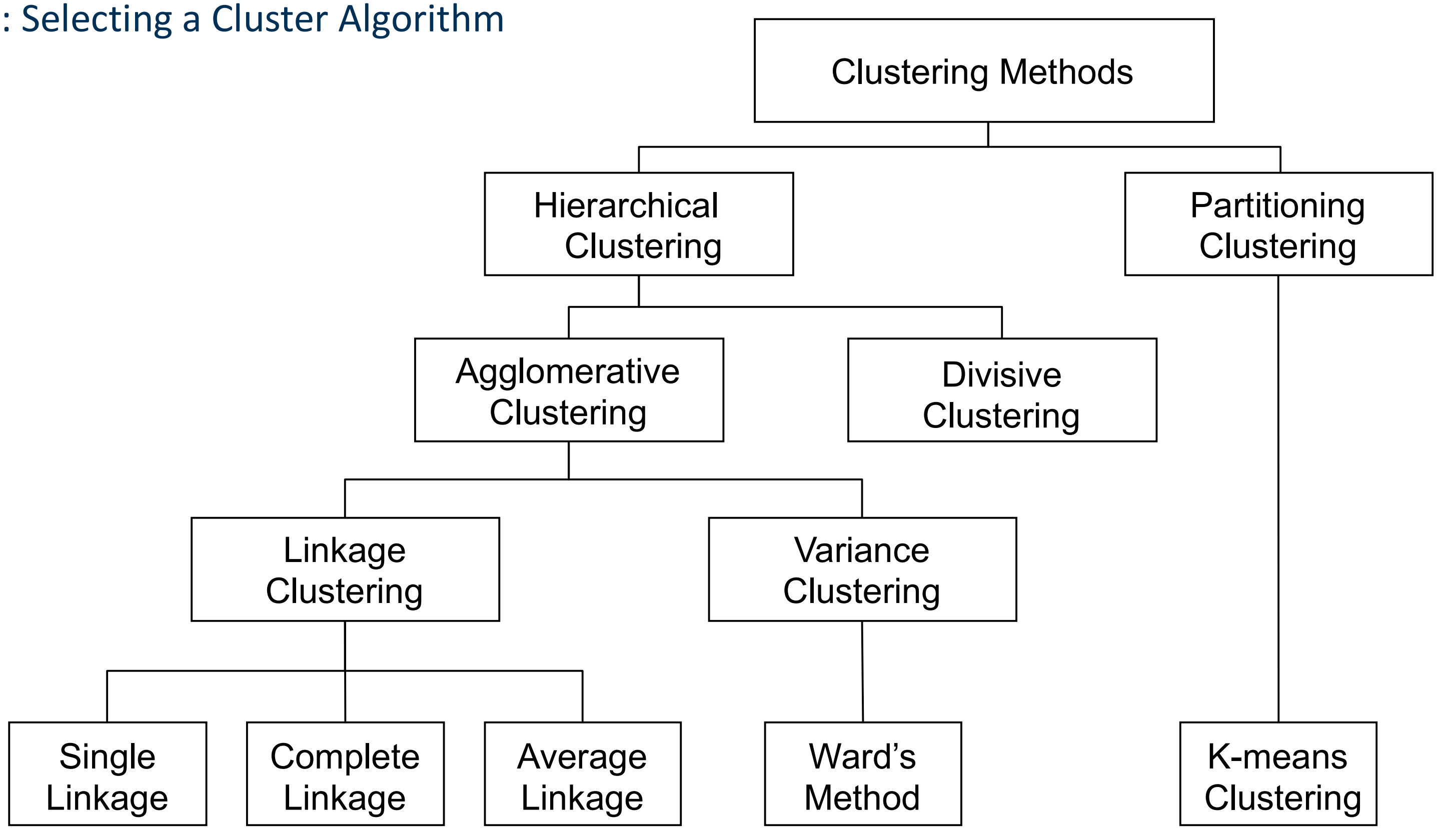
Analysis of Interdependence: Cluster Analysis (8)

Step 3: Eliminating Outliers

Objects with very untypical characteristics for the sample

→ Application of the Single Linkage Clustering (see step 4)

Step 4: Selecting a Cluster Algorithm



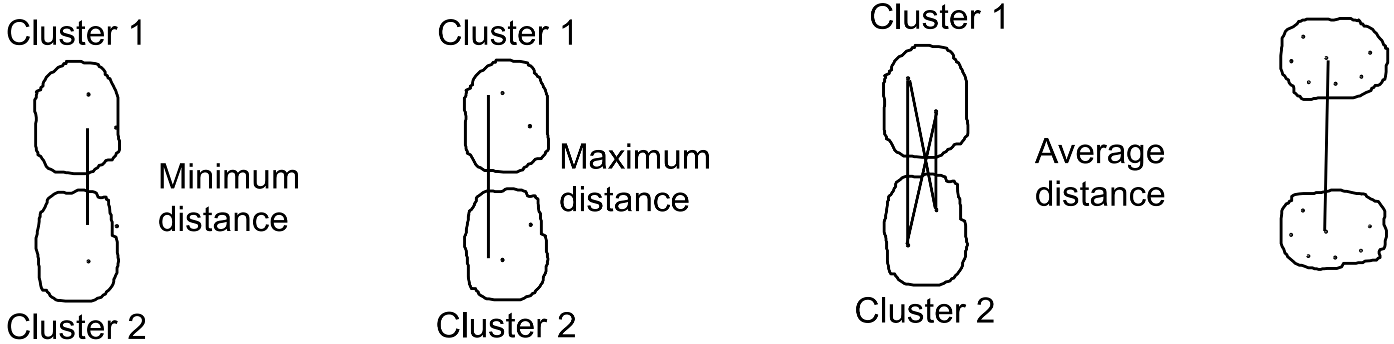
Homburg (2015), p. 375; Jensen (2008), p. 339; Lattin et al. (2003), p. 265

UNIVERSITY OF MANNHEIM
Business School

Hierarchical/Agglomerative Clustering

- Linkage Methods
- Variance Methods
- Centroid Methods

Single Linkage **Complete Linkage** **Average Linkage** **Centroid Methods**

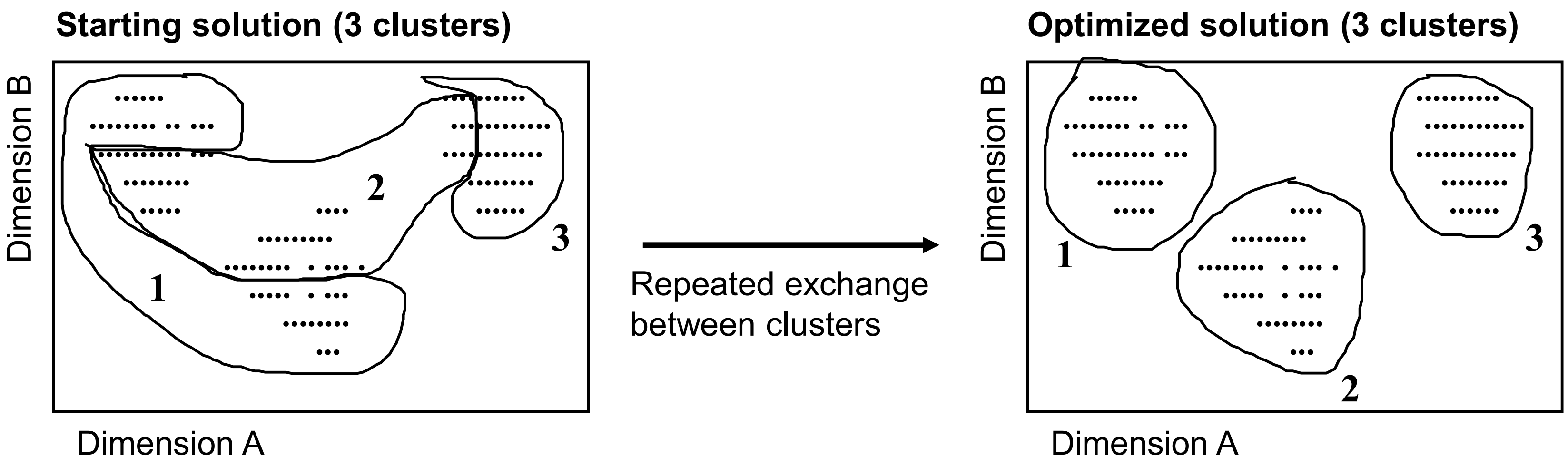


Lattin et al. (2003), p. 288

Analysis of Interdependence: Cluster Analysis (10)

Partitioning Clustering

- K-means algorithm: Select an initial partition of the data into K clusters and calculate the centroid for each cluster; then assign each object to the cluster whose centroid is closest and recalculate the centroids
- Local optimization: The better the starting solution, the better the result



Lattin et al. (2003), p. 288

Step-by-step Example: Clustering by Product Attributes

Cluster analysis is utilized in order to thoroughly examine if groups of products can be built. A **hierarchical cluster analysis** is used in this case **(Minkowski Metric – city block)**, and the **single linkage distance** is applied. Starting from ten groups (= ten products $\sigma_1,\ldots,\sigma_{10}$) $K_1,\ldots,K_{10}$, two groups are combined into one group if their heterogeneity, which is evaluated by

$$v(K_{j_1}, K_{j_2}) = \min\{d_{i_1 i_2}\} \qquad \begin{array}{l} \sigma_{i_1} \varepsilon K_{j_1} \\ \sigma_{i_2} \varepsilon K_{j_2} \end{array}$$

is minimal. In the following step, the same procedure is repeated using the same structure which is reduced by one group. The starting point of the cluster analysis is the distance matrix D=(dij), (i=1,...,10 and j=1,...,10) of the ten products that is determined in accordance with the rule

$$d_{ij} = \sum_{k=1}^{4} p_k \left| a_{ik} - a_{jk} \right|$$

where p_k is the weighting of the attribute k (k=1,...,4), and a_{ik} is the evaluation of the product i with respect to the attribute k.

Analysis of Interdependence:
Cluster Analysis (12)

UNIVERSITY OF MANNHEIM
Business School

Step-by-step Example: Clustering by Product Attributes

Please evaluate and interpret the corresponding cluster analysis.
How many groups can be built?
Which implications can be derived for product A?

Solution

The distance matrix of the ten products A,...,J:

d_{ij}	A	B	C	D	E	F	G	H	I	J
A	0.0	1.8	0.9	1.5	3.2	0.5	1.8	0.7	1.2	2.1
B		0.0	1.3	1.5	1.4	2.3	0.6	1.9	2.0	0.9
C			0.0	0.6	2.7	1.0	1.5	0.6	0.7	2.0
D				0.0	2.3	1.6	1.7	1.0	0.7	2.4
E					0.0	3.7	1.6	3.3	3.0	1.7
F						0.0	2.3	0.6	1.1	2.4
G							0.0	1.9	2.2	1.3
H								0.0	0.5	2.0
I									0.0	2.3
J										0.0

The steps of the hierarchical cluster analysis are displayed below.

Step 1: $\min\{d_{ij}\}_{i \neq j} = d_{AF} = d_{HI} = 0.5$, i.e.

$C_1 = \{\{A,F\}B,C,D,E,G,\{H,I\},J\}$

d_{ij} = Distance between the objects i and j
C_i = Cluster
v_i = Cluster Center

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED

Analysis of Interdependence:
Cluster Analysis (13)

UNIVERSITY OF MANNHEIM
Business School

Step-by-step Example: Clustering by Product Attributes

v^1	{A,F}	B	C	D	E	G	{H, I}	J
{A,F}	0.0	1.8	0.9	1.5	3.2	1.8	0.6	2.1
B		0.0	1.3	1.5	1.4	0.6	1.9	0.9
C			0.0	0.6	2.7	1.5	0.6	2.0
D				0.0	2.3	1.7	0.7	2.4
E					0.0	1.6	3.0	1.7
G						0.0	1.9	1.3
{H,I}							0.0	2.0
J								0.0

Step 2: $\min\{v^1(C_i,C_j)\}_{i\neq j} = v^1[(\{A,F\},C,D,\{H,I\})(B,G)] = 0.6$, i.e.
$C_2 = \{\{A,C,D,F,H,I\},\{B,G\},E,J\}$

v^2	{A,C,D,F,H,I}	{B,G}	E	J
{A,C,D,F,H,I}	0.0	1.3	2.3	2.0
{B,G}		0.0	1.4	0.9
E			0.0	1.7
J				0.0

Step 3: $\min\{v^2(C_i,C_j)\}_{i\neq j} = v^2(\{B,G\},J) = 0.9$ i.e.
$C_3 = \{\{A,C,D,F,H,I\},\{B,G,J\},E\}$

d_{ij} = Distance between the objects i and j
C_i = Cluster
v_i = Cluster Center

Analysis of Interdependence: Cluster Analysis (14)

Step-by-step Example: Clustering by Product Attributes

v^3	{A,C,D,F,H,I}	{B,G,J}	E
{A,C,D,F,H,I}	0.0	1.3	2.3
{B,G,J}		0.0	1.4
E			0.0

Step 4: $\min\{v^3(C_i,C_j)\}_{i \neq j} = v^3(\{A,C,D,F,H,I\},\{B,G,J\}) = 1.3$ i.e.
$C_4 = \{\{A,B,C,D,F,G,H,I,J\},E\}$

v^4	{A,B,C,D,F,G,H,I,J}	E
{A,B,C,D,F,G,H,I,J}	0.0	1.4
E		0.0

Step 5: $\min\{v^4(C_i,C_j)\}_{i \neq j} = v^4(\{A,B,C,D,F,G,H,I,J\},E) = 1.4$ i.e.
$C_5 = \{A,B,C,D,F,G,H,I,J,E\}$

d_{ij} = Distance between the objects i and j
C_i = Cluster
v_i = Cluster Center

Analysis of Interdependence: Cluster Analysis (15)

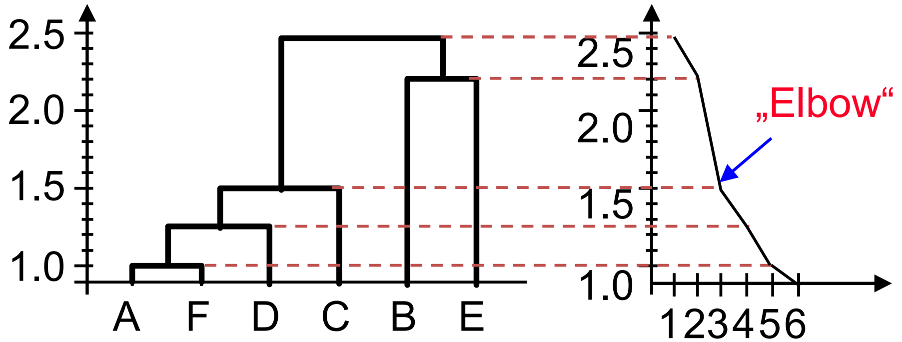
Step 5: Determining the Number of Clusters

Documentation of the cluster formation by means of a dendrogram

Clustering methods are frequently combined

- Elimination of outliers based on the single linkage clustering (high distance to all other objects)
- Cluster assignment based on the Ward's method (minimization of the variance within each cluster)
- Optimization based on the K-means algorithm

Elbow criterion: Number of clusters for which the heterogeneity growth is the strongest



Analysis of Interdependence: Cluster Analysis (16)

Pseudo-F Statistic

- Objective: Reducing the within-group heterogeneity

$$pseudo - F = \frac{SS_B/(K-1)}{SS_W/(n-K)}$$

SS_B: total variance between the clusters
SS_W: total variance within the clusters
K: number of clusters
n: number of objects

Jensen (2008), p. 355

- The higher the pseudo-F statistic, the better the reduction of the within-group heterogeneity
- The pseudo-F statistic does not increase monotonously, but reaches its maximum for the optimal value of K

Step 6: Interpretation and Labeling of the Clusters

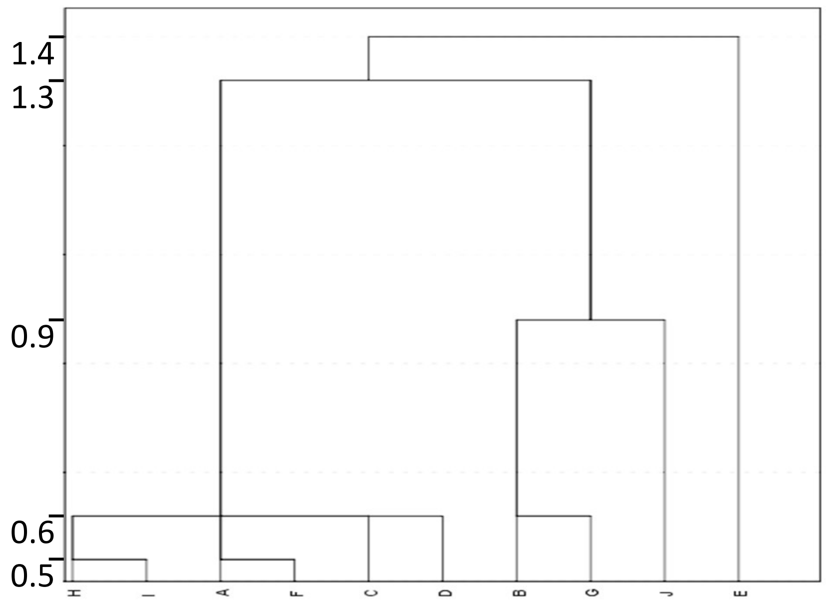
- Starting point: Centroids of the clusters with reference to the characteristics on which the cluster formation is based
- Conducting variance analyses in order to test the significance of mean differences
- Meaningful labeling of the clusters

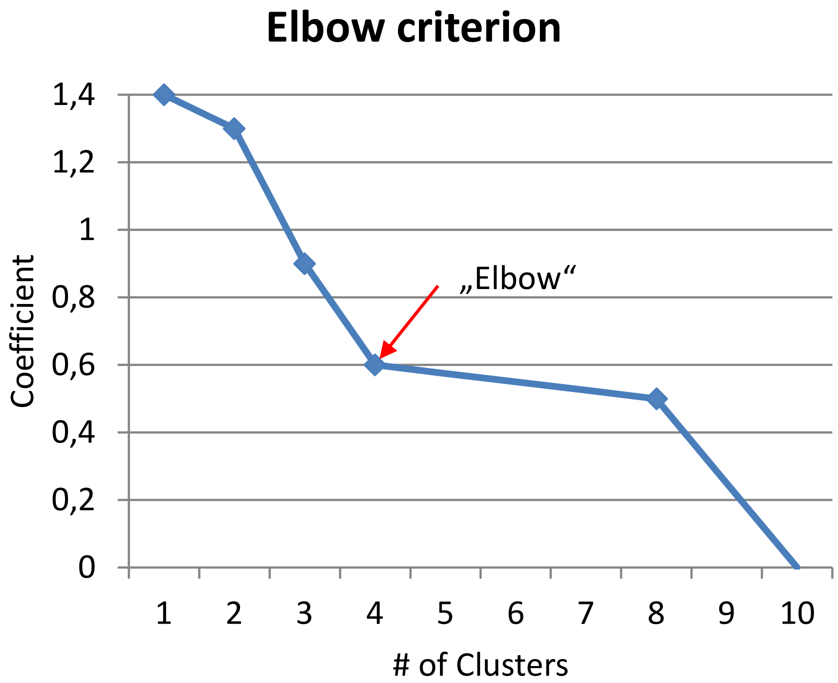
Analysis of Interdependence: Cluster Analysis (17)

Step-by-step Example: Clustering by Product Attributes

A dendrogram and the elbow criterion illustrate the results of the cluster formation

Dendrogram





A formation of four product groups appears reasonable:

1. **Group (A,C,D,F,H,I):** High quality products (highest values in all attributes)
2. **Group (B,G):** Average quality products (average values in all attributes)
3. **Group (E):** Lower quality products (lower values in all attributes)
4. **Group (J):** Quality focus on Attributes 2 & 4

Cluster Analysis:
Clustering by Product Attributes– Summary

UNIVERSITY OF MANNHEIM
Business School

What was the initial problem?

- We have product ratings for ten products and four attributes
- However, due to the high number of products comparisons would be very complex
- Therefore, using the four attributes, we want to group the products based on their similarity in these attributes

What did we do?

- We created a data matrix with the ten products and the average rating they got on the four attributes in the survey
- Using the Minkowski Metric (r = 1), we created the distance matrix
- Under the condition that outliers were already eliminated, we conducted a hierarchical clustering with the single linkage method
- We determined 4 clusters by using the dendrogram/elbow criterion
- By interpreting the mean values of the factors for every cluster, we interpreted and named them

What was the outcome?

A formation of four product groups appears reasonable:

1. **Group (A,C,D,F,H,I):** High quality products (highest values in all attributes)
2. **Group (B,G):** Average quality products (average values in all attributes)
3. **Group (E):** Lower quality products (lower values in all attributes)
4. **Group (J):** Quality focus on Attributes 2 & 4

Excursus:
Market Segmentation

Market segmentation

Division of the heterogeneous market into homogenous sub-markets (segments) with respect to certain characteristics of actual or potential buyers (target groups) as well as
the targeted development of one or more segments by means of segment-specific marketing programs

(Freter (1995), p. 1803; Lattin et al. (2003), pp. 266)

Objectives

- Fulfill various requirements, desires and preferences
- Increase in market share
- Increase in profitability

Requirements for Market Segmentation Criteria

- Relevance for behavior
- Accessibility of segment members
- Discriminatory power
- Measurability of the segmenting criteria
- Stability over time
- Profitability/Efficiency

Market Segmentation Criteria

Demographic Criteria	
Private customers • Gender • Age • Marital Status • Residence	Business customers • Corporate headquarters • Duration of business relationship

Purchasing Behavior Criteria	
Private customers • Selection of shopping location • Product choice • Buying frequency	Business customers • Selection of distribution channel • Buying frequency
• Price sensitivity • Information behavior	

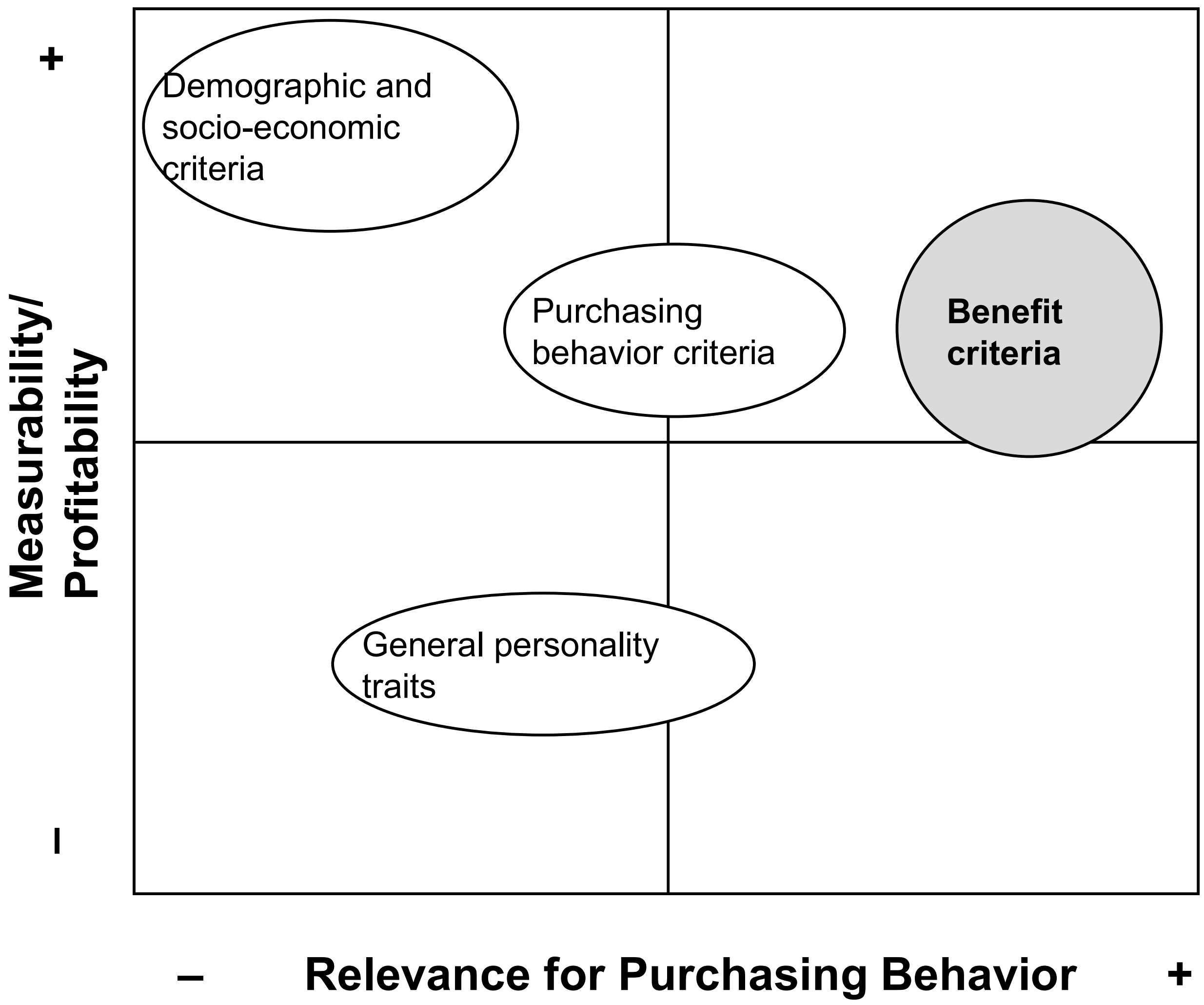
Socio-Economic Criteria	
Private customers • Income • Education • Job	Business customers • Sales volume • Sector

Benefit Criteria	
Private customers • Benefit of price • Benefit of quality • Benefit of image • Benefit of service	Business customers • Benefit of price • Benefit of quality • Benefit of image • Benefit of service

General Personality Traits	
Private customer: • Lifestyle • Attitudes • Interests	

Homburg/Kuester/Kromer (2012)

Market Segmentation: Benefit Segmentation



Benefit segmentation avoids two crucial problems of other segmentation criteria

- Lack of or too complex measurability
- No causal relationship between the segmentation criteria and the purchasing decision or the brand choice

Analysis of Dependence: Analysis of Variance (1)

Analysis of Dependence

Dependent variables, metric

Inde-
pendent
variables
non-
metric

- Multivariate Analysis of Variance

Homburg (2015), p. 359; Hair et al. (2010), p. 12

Analysis of Dependence: Analysis of Variance (2)

Problem

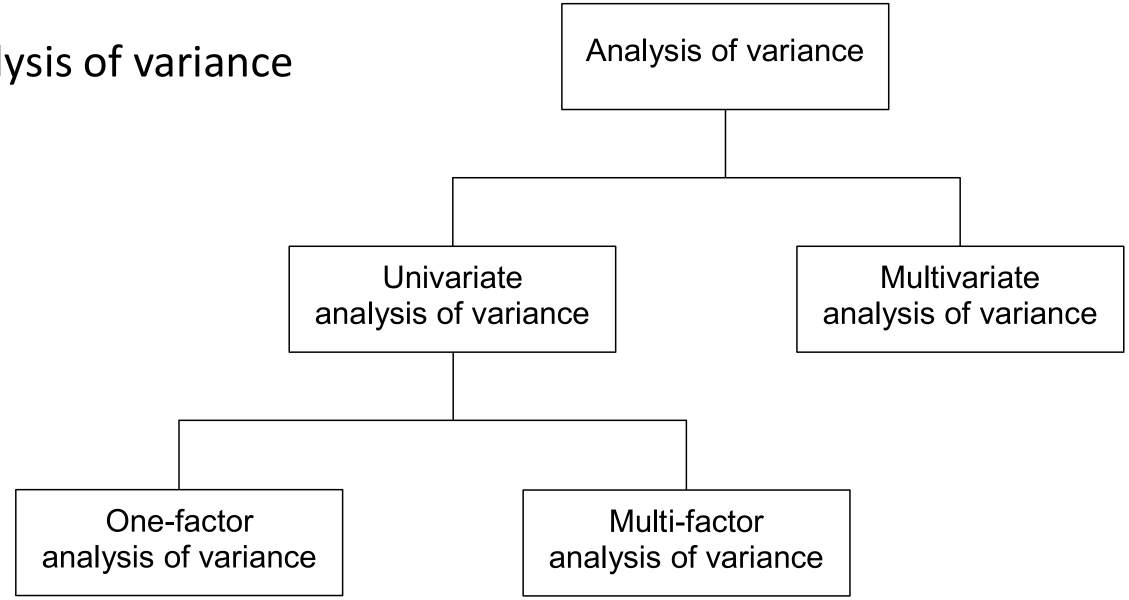
- Within the shampoo product test, we have surveyed respondents with different hair structures
- Now we would like to know for every shampoo A-J if the respondents' quality ratings differ in accordance to their hair structure

Analysis of Variance

Analysis of Dependence: Analysis of Variance (3)

Question: Are there significant differences between groups with respect to the dependent variable?

Forms of the analysis of variance



- One-factor analysis of variance (ANOVA)
- Multi-factor analysis of variance (ANOVA): Several independent variables affect a dependent variable
- Multivariate analysis of variance (MANOVA): Two or more dependent variables and one or more independent variables

Analysis of Dependence: Analysis of Variance (4)

- Application of the one-factor analysis of variance if two sample means are to be compared
- Test for differences in the mean values of a metric scaled variable Y across the groups G (g = 1,…, G)
- **Model:** $Y_{gk} = \mu + \alpha_g + e_{gk}$, where

$Y_{gk} =$ Observation value k in group g

$\mu =$ Mean of the dependent variable Y of the population

$\alpha_g =$ Influence of the affiliation to the group g on the dependent variable Y

$e_{gk} =$ Error term

- Null hypothesis: Group affiliation has no effect on the dependent variable ($\alpha_1 = \alpha_2 = ... = \alpha_G$), i.e. all group means are equal, they come from the same population

Analysis of Dependence: Analysis of Variance (5)

Shampoo Product Test

- F-test to test the null hypothesis that there is no relationship between the group affiliation and the average value of the dependent variables

$$F = \frac{\frac{SS_x}{G-1}}{\frac{SS_e}{K-G}}$$

SS_x: Explained variance = variance between the groups
Ss_e: Unexplained variance = sum of the variances within the groups
G: Number of groups
K: Number of cases

- Example: Quality Rating of Shampoos as a function of the hair structure affiliation of the respondents in the shampoo product test

fine/thin->

```
 Descriptive statistics by group
group: 1
   vars   n mean  sd median trimmed  mad min max range  skew kurtosis  se
x1    1 223 5.31 1.5      6    5.31 1.48   1   7     6 -0.81    -0.13 0.1
--------------------------------------------------------------------------
----------------------------------------------
```

quality assessments…

normal/medium->

```
group: 2
   vars   n mean  sd median trimmed  mad min max range  skew kurtosis  se
x1    1 169 5.44 1.3      6    5.44 1.48   1   7     6 -0.87     0.55 0.1
--------------------------------------------------------------------------
----------------------------------------------
```

thick/strong->

```
group: 3
   vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
x1    1 96 5.23 1.59    5.5    5.23 1.48   1   7     6 -0.78    -0.22 0.16
```

- Different hair status → Might influence quality assessment ?

Analysis of Dependence: Analysis of Variance (6)

Shampoo Product Test: Univariate Analysis of Variance with R

```
                        Df Sum Sq Mean Sq  F value Pr(>F)
(Intercept)              1  13927   13927 6591.841 <2e-16 ***
volshampoo$hair_struc    2      3       2    0.745  0.475
Residuals              485   1025       2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- The explained variance (R^2) is 0.3 %
- The unexplained variance is 99.7%
- Computation of the F value:

$F_{emp} = [SS_x/(G-1)]/[SS_e/(K-G)]$ $\qquad F_{tab;.05;2;485} = 3.01$

$= (SS_x/ SS_e) \cdot [(K-G)/(G-1)]$

$= (0.3 / 99.7) \cdot [(488 -3)/(3-1)]$ $\qquad (n_1 = G - 1 = 3 - 1 = 2)$

$= 0.73$ (rounded) $\qquad (n_2 = K - G = 488 - 3 = 485)$

→ No rejection of H_0. Thus, no proof of significant group differences.

Analysis of Dependence:
Analysis of Variance (7)

Shampoo Product Test: Univariate Analysis of Variance with R

- Check ANOVA for subset of data containing only responses for shampoo A

```
                          Df Sum Sq Mean Sq F value Pr(>F)    
(Intercept)                1 1387.4  1387.4 647.785 <2e-16 ***
only_shampoo_a$hair_struc  2   11.8     5.9   2.761 0.0733 .  
Residuals                 48  102.8     2.1                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

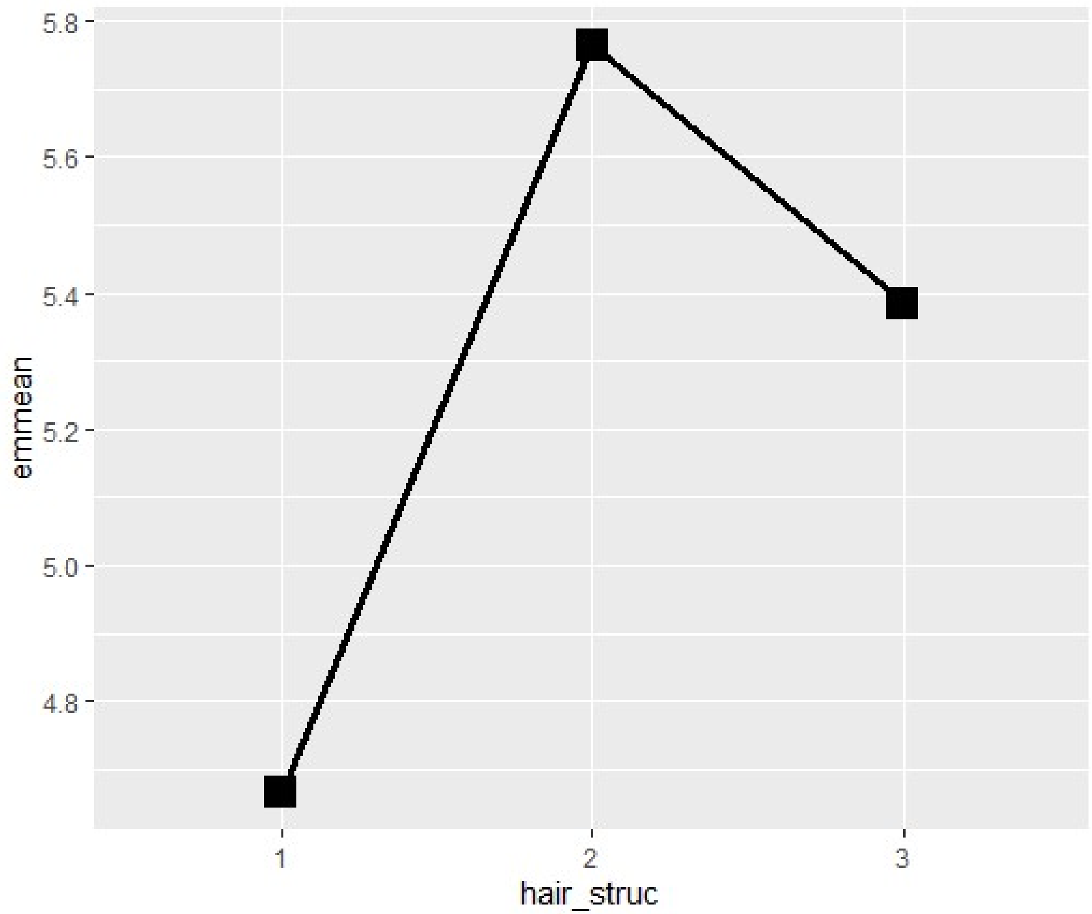
- There seems to be a stronger effect and R-Squared increases to 10.3%.

→ BUT: No rejection of H_0 at 5%-level!

Analysis of Dependence: Analysis of Variance (8)

Shampoo Product Test: Univariate Analysis of Variance with R

Estimated Marginal Means (EMM) of Quality Ratings for Shampoo A



Analysis of Dependence: Analysis of Variance (9)

Procedure of Multi-factor Analysis of Variance (Example with two factors)

1. **Model specification: $Y_{ghk} = \mu + \alpha_g + \beta_h + (\alpha \cdot \beta)_{gh} + \varepsilon_{ghk}$,** where μ is the mean of the dependent variable of the population, α_g and β_h are the main effects of the two independent variables, $(\alpha \cdot \beta)_{gh}$ is the interaction effect, Y_{ghk} is the kth observed value and ε_{ghk} is the error term
2. **Variance decomposition: $SS_Y = SS_e + SS_A + SS_B + SS_{AxB}$,**
   where SS_Y is the total variance, SS_e is the variance within the groups, SS_A is the variance caused by the factor A, SS_B is the variance caused by the factor B and SS_{AxB} is the variance caused by the interaction of the factors
3. **Model assessment**
   - F-test to test the null hypothesis
   - Eta Squared as a measure for the strength of the effect of an independent on the dependent variable: $\eta^2 = SS_{A/}SS_Y$ (here referred to the factor A)
4. **Result interpretation**
   - Interpretation of the results of the F tests
   - Checking of the mean values of the dependent variables in each group with regard to the effect sequence

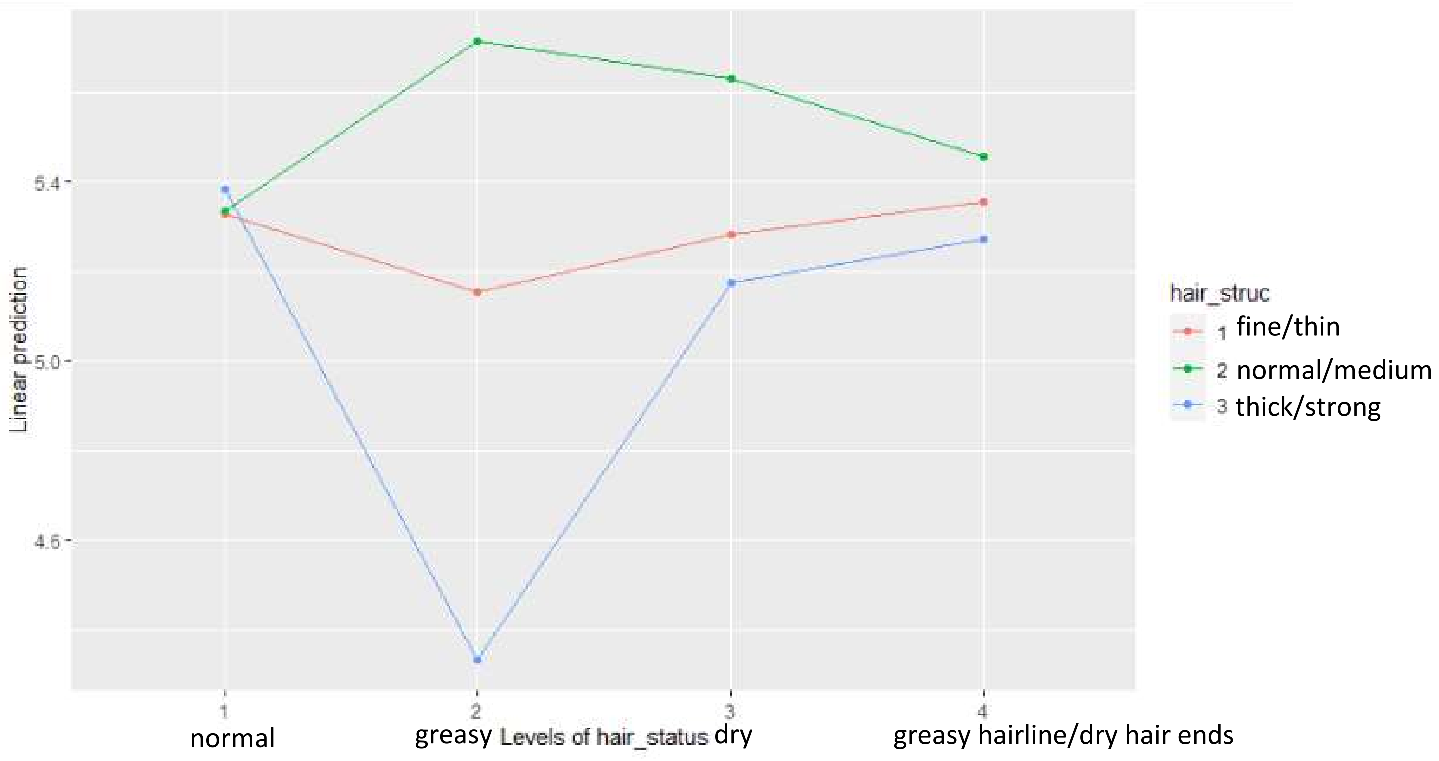
Analysis of Dependence:
Analysis of Variance (10)

Shampoo Product Test: Example for a two-factor analysis of variance

Dependent Variable: Quality Rating

Independent Variable: Structure of the hair, Status of the hair

Estimated Marginal Means (EMMs) of Quality Ratings for all shampoos



Analysis of Dependence:
Analysis of Variance (11)

Shampoo Product Test: Example for a two-factor analysis of variance

Dependent Variable: Quality Rating
Independent Variable: Structure of the hair, Status of the hair

```
                                             Df Sum Sq Mean Sq  F value Pr(>F)    
(Intercept)                                   1  13756   13756 6423.281 <2e-16 ***
volshampoo$hair_struc                         2      3       1    0.586  0.557    
volshampoo$hair_status                        3      1       0    0.216  0.885    
volshampoo$hair_struc:volshampoo$hair_status  6      8       1    0.626  0.710    
Residuals                                   470   1007       2                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- **No significant group differences for Multi-Factor ANOVA!**

Analysis of Variance: Example shampoo product test – Summary

What was the initial problem?

- We wanted to find out if there are significant differences between groups of hair structures with respect to the quality ratings for the shampoos?

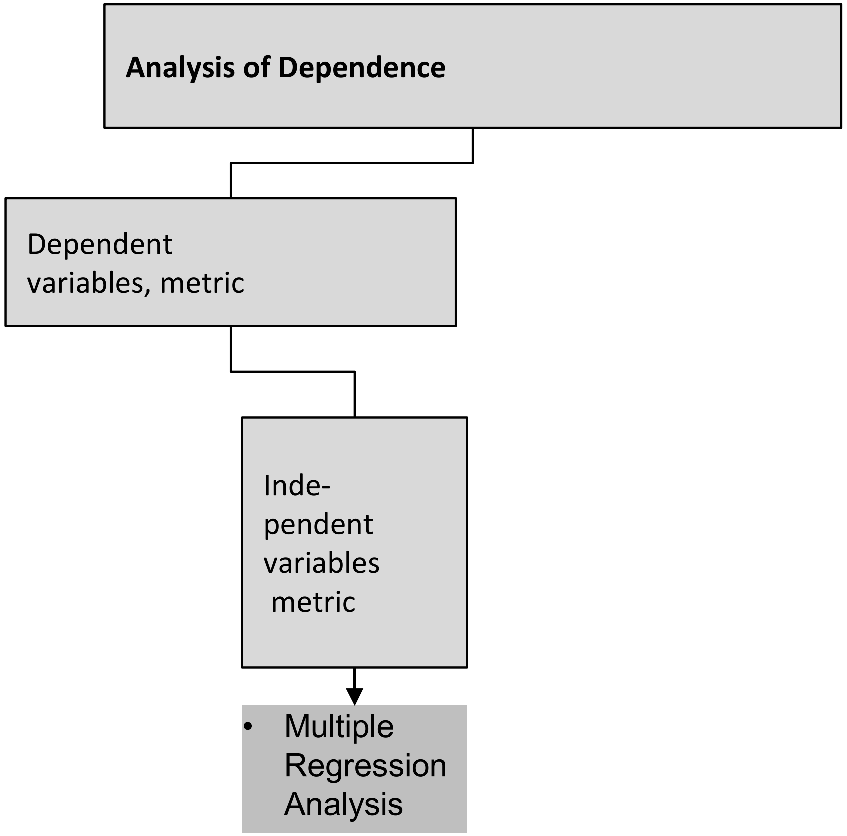
What did we do?

- The quality ratings indicated were not significantly different between the groups of hair structures
- By the help of a statistical software (e.g., R Statistics), we calculated the explained and unexplained variance and
- We were able to calculate the F value for hypothesis testing

What was the outcome?

- We did not reject H_0, i.e., we could not show that there are significant group differences
- The group difference is stronger when analyzing only our shampoo A

Analysis of Dependence: Multiple Regression Analysis



Homburg (2015), p. 359; Hair et al. (2010), p. 12

Analysis of Dependence: Multiple Regression Analysis – Basic Analysis (1)

Problem

- Apart from asking the respondents in the survey to rate the tested shampoo on several characteristics, we also asked them for their buying intention (7-point Likert scale)
- Now, we would like to know in how far the shampoos' characteristics summarized by the four factors have an influence on the respondents' buying intentions
- The values for the four factors have been computed using a confirmatory factor analysis
- Based on this analysis, we want to figure out if Shampoo A already fulfills the most important characteristics for the buying intention to a high extent

Multiple Regression Analysis

UNIVERSITY OF MANNHEIM
Business School

Multiple regression equation: $y_i = b_0 + b_1x_{1i} + b_2x_{2i} + ... + b_Jx_{Ji} + e_i$

- Analyze relationship between a dependent variable and several independent variables
- Applicable to metric and quasi-metric variables
- Parameter estimation with the least squares procedure
- Standardization allows for a direct comparison between regression coefficients

➔ Standardized regression coefficients

$$\beta_j = b_j \frac{\textbf{standard deviation } (x_j)}{\textbf{standard deviation } (y)} \quad \textbf{where } j = 1, 2, ..., J$$

- t-test to determine the significance of individual regression coefficients
  H_0: regression coefficient $b_j = 0$
- Coefficient of determination R^2 to test the significance of the regression model

Hair et al. (2010), p. 155

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (3)

UNIVERSITY OF MANNHEIM
Business School

- Examination of the regression equation using the F-test
- Null hypothesis H_0: $b_1 = b_2 = \dots = b_J = 0$
  → no relationship between independent variables and dependent variable
- Significance of the regression model based on the F-test with significance level α
- F ratio

$$F_{emp} = \frac{\frac{R^2}{J}}{\frac{1 - R^2}{H - J - 1}}$$

J = Number of independent variables
H = Number of observations
Degrees of freedom n1 = J and n2 = H - J - 1

- Critical value: tabled F statistic at a significance level of .05
- Test rule: Rejection of H_0 ←→ $F_{emp} > F$

Assumptions of a regression analysis

- All variables are measured at interval level and free of error
- The expected value of the residuals is zero
- Residual variance is constant across all observations. The residual variance of the independent variables is constant
- Residuals of the independent variables are uncorrelated with each other
- Residuals are uncorrelated with independent variables
- Independent Variables are not perfectly correlated with each other (no perfect collinearity). There is no perfect linear relationship between the independent variables
- Residuals are normally distributed

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (6)

UNIVERSITY OF MANNHEIM
Business School

Assumptions of a regression analysis

- All variables are measured at interval level and free of error
- The expected value of the residuals is zero
- Residual variance is constant across all observations. The residual variance of the independent variables is constant
- Residuals of the independent variables are uncorrelated with each other
- Residuals are uncorrelated with independent variables
- Independent Variables are not perfectly correlated with each other (no perfect collinearity). There is no perfect linear relationship between the independent variables
- Residuals are normally distributed

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (8)

UNIVERSITY OF MANNHEIM
Business School

Assumptions of a regression analysis

- All variables are measured at interval level and free of error
- The expected value of the residuals is zero
- Residual variance is constant across all observations. The residual variance of the independent variables is constant
- Residuals of the independent variables are uncorrelated with each other
- Residuals are uncorrelated with independent variables
- Independent Variables are not perfectly correlated with each other (no perfect collinearity). There is no perfect linear relationship between the independent variables
- Residuals are normally distributed

Assumptions of a regression analysis

- All variables are measured at interval level and free of error
- The expected value of the residuals is zero
- Residual variance is constant across all observations. The residual variance of the independent variables is constant
- Residuals of the independent variables are uncorrelated with each other
- Residuals are uncorrelated with independent variables
- Independent Variables are not perfectly correlated with each other (no perfect collinearity). There is no perfect linear relationship between the independent variables
- Residuals are normally distributed

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (12)

UNIVERSITY OF MANNHEIM
Business School

Shampoo Product Test

Determine the importance of the four factors for the overall buying intention

Concrete question: How strong is the impact of the shampoo characteristics on the overall buying intention?

This question can be answered by using multiple regression analysis

Measurement Independent Variables

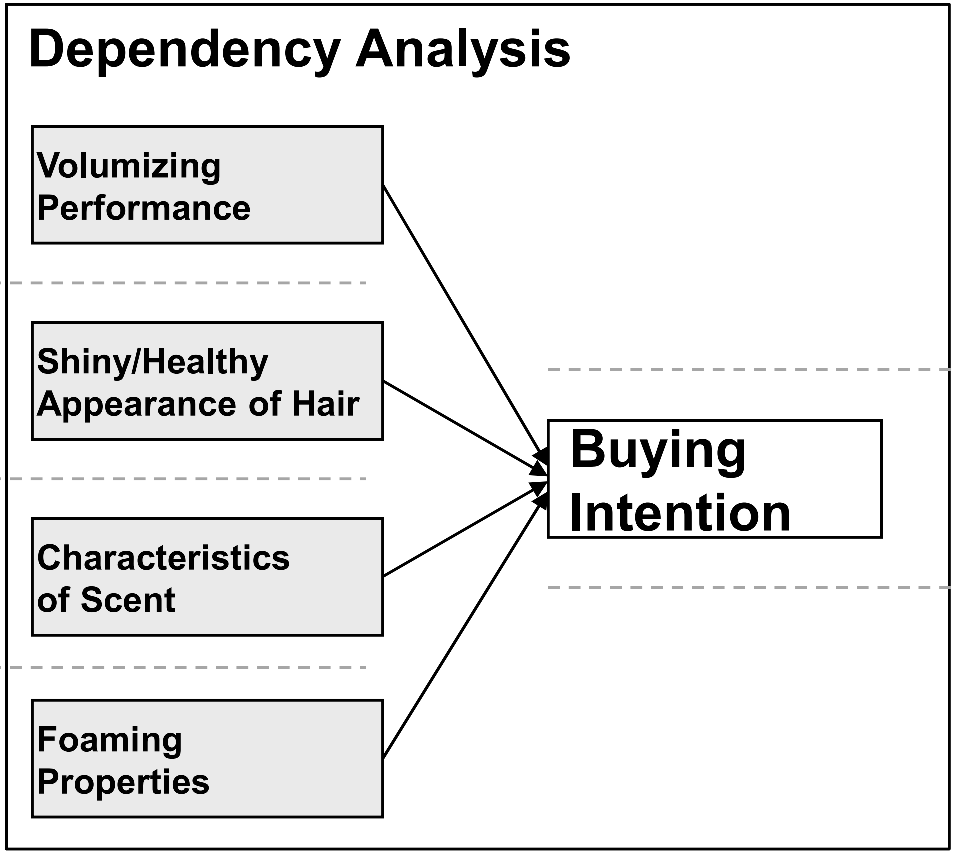
The shampoo tested…

Gives hair long lasting volume
Keeps hair voluminous till the evening
Makes hair visibly voluminous
Gives hair particularly more volume
Gives hair volume all day long

Gives hair a natural shine
Makes hair look healthy
Gives hair a silky touch
Makes hair look well-cared

Has a pleasant scent
Has a fresh scent
Has a natural scent

Has a foam which is nice to apply
Foams up particularly softly
Has a particular pleasant foam



Measurement Dependent Variable

To which extent do you agree with the following statement (1 = strongly disagree; 7 = strongly agree):
I would buy the shampoo tested if it was available in the stores where I normally buy my shampoo.

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (13)
Shampoo Product Test

Model Summary

```
Residuals:
    Min      1Q  Median      3Q     Max 
-3.7045 -0.5382  0.0532  0.7126  3.6143 

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                       4.94845    0.05215  94.892  < 2e-16 ***
Volumizing_Performance            0.52395    0.03496  14.986  < 2e-16 ***
Shiny_Healthy_Appearance_of_Hair  0.82792    0.07207  11.488  < 2e-16 ***
Characteristics_of_Scent          0.10272    0.03415   3.008  0.00277 ** 
Foaming_Properties                0.17366    0.06626   2.621  0.00905 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.148 on 480 degrees of freedom
Multiple R-squared:  0.6681,	Adjusted R-squared:  0.6654 
F-statistic: 241.6 on 4 and 480 DF,  p-value: < 2.2e-16
```

Standardized Coefficients

```
Volumizing_Performance Shiny_Healthy_Appearance_of_Hair 
            0.43991246                       0.41804077 
Characteristics_of_Scent               Foaming_Properties 
              0.09105108                       0.09585675 
```

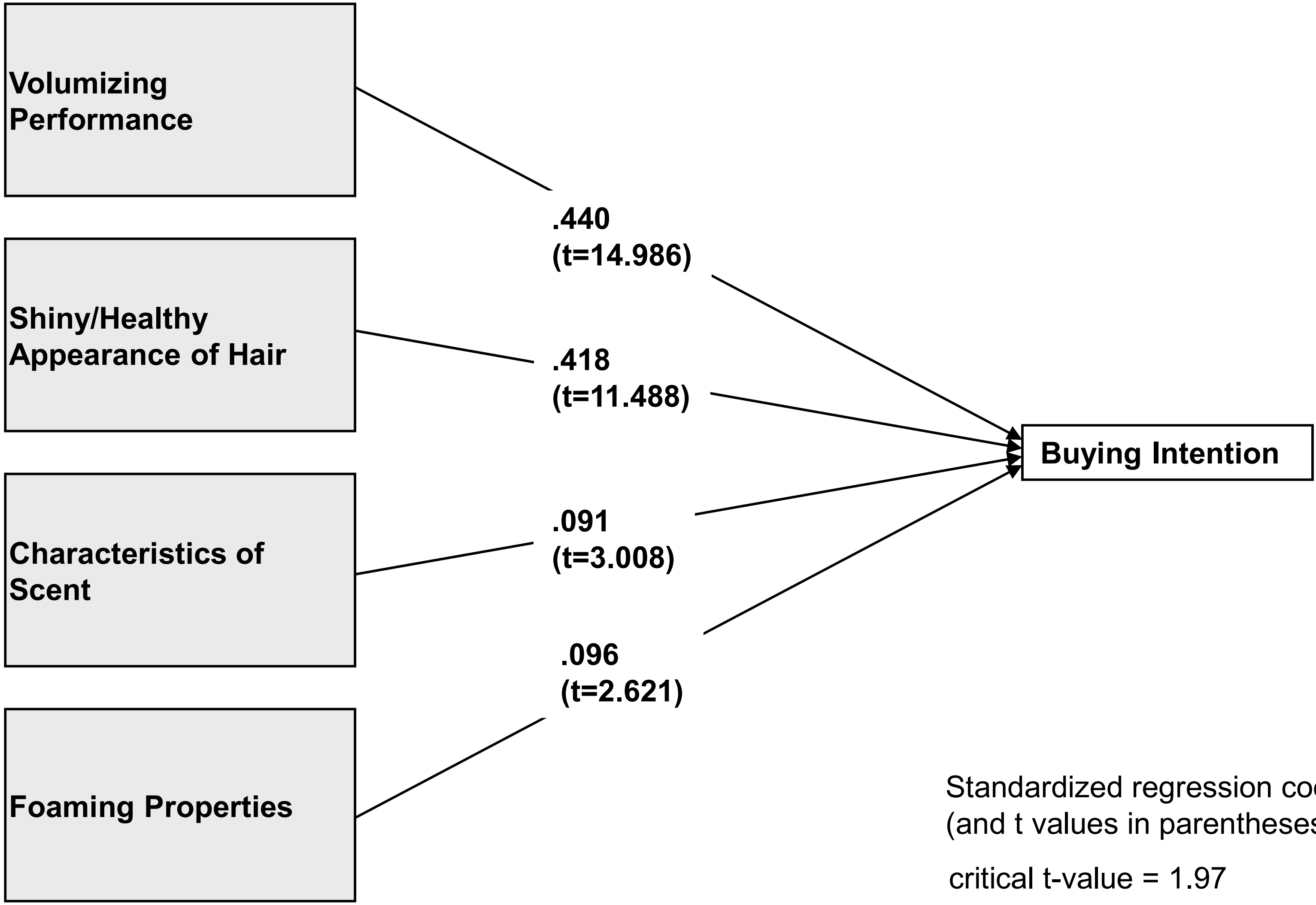
Results

- The four factors (independent variables) explain 66.81% of the variance of the dependent variable → $R^2 = .6681$
- The empirical F statistic is 241.6 with $n_1 = 4$ and $n_2 = 480$ degrees of freedom
- The tabled F statistic at a significance level of 5% yields 2.39
- Null hypothesis is rejected, i.e., there is a significant relationship between the independent and dependent variables
- All factors have significant positive impacts on the buying intention

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (14)

Shampoo Product Test

Determination of the Characteristics' Importance for the Buying Intention



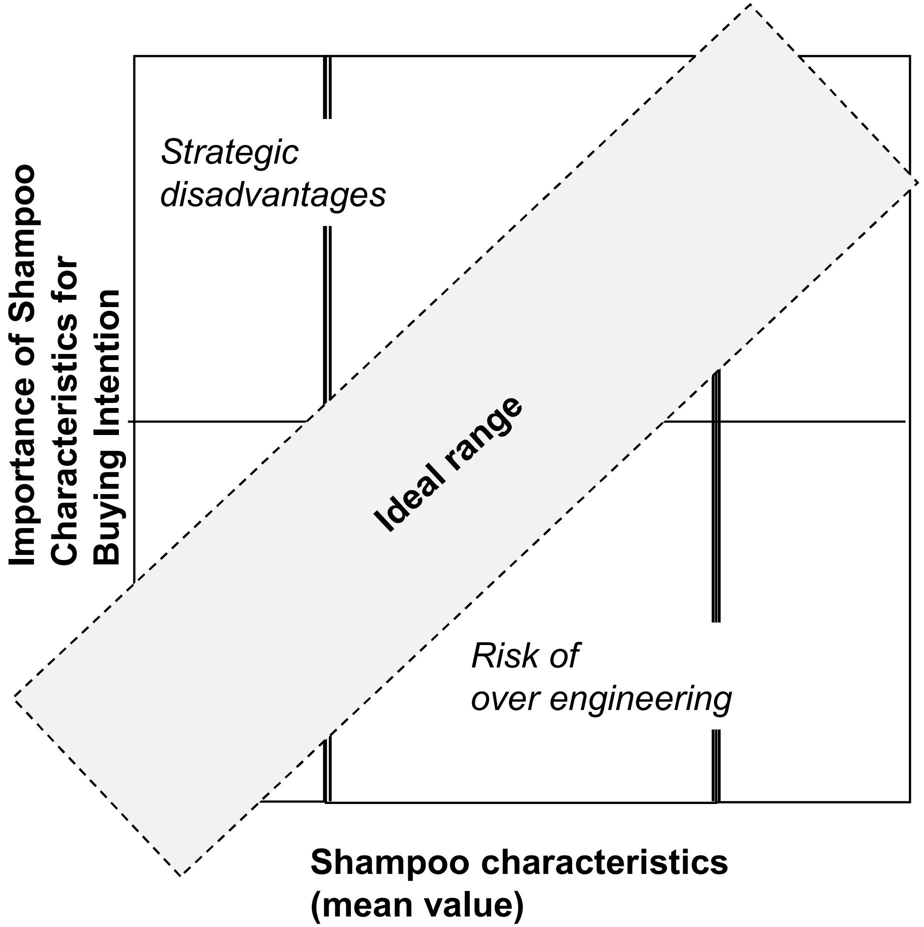
Standardized regression coefficients (and t values in parentheses)

critical t-value = 1.97

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (15)

Shampoo Product Test

Integration of the shampoo's characteristics and their importance for the Buying Intention in one portfolio



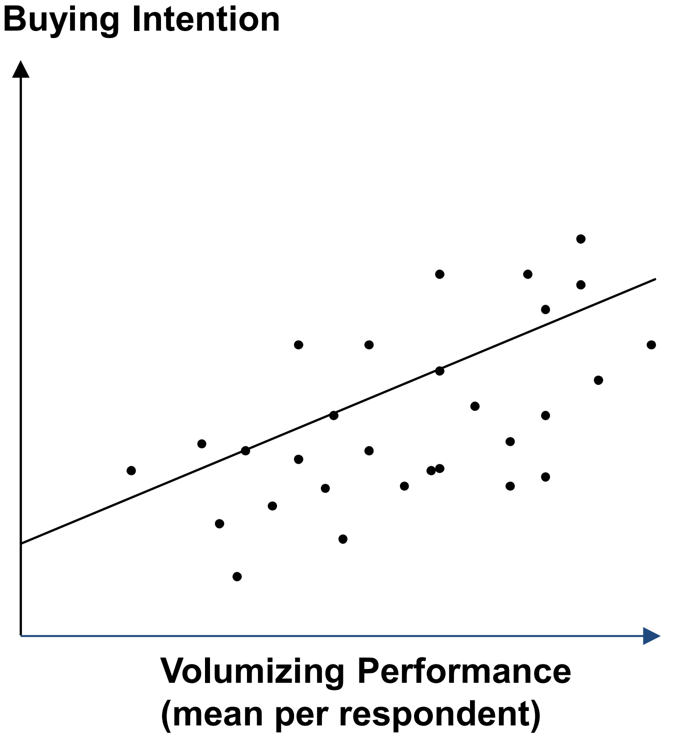
- The graph illustrates the relationship between the shampoo's characteristics and their importance for the respondents' buying intentions
- The position in the profile can be used to derive strategic implications for the shampoo
- In the ideal range, characteristics of a high importance for the buying intention are well-marked and characteristics of a lower importance are less marked

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (16)

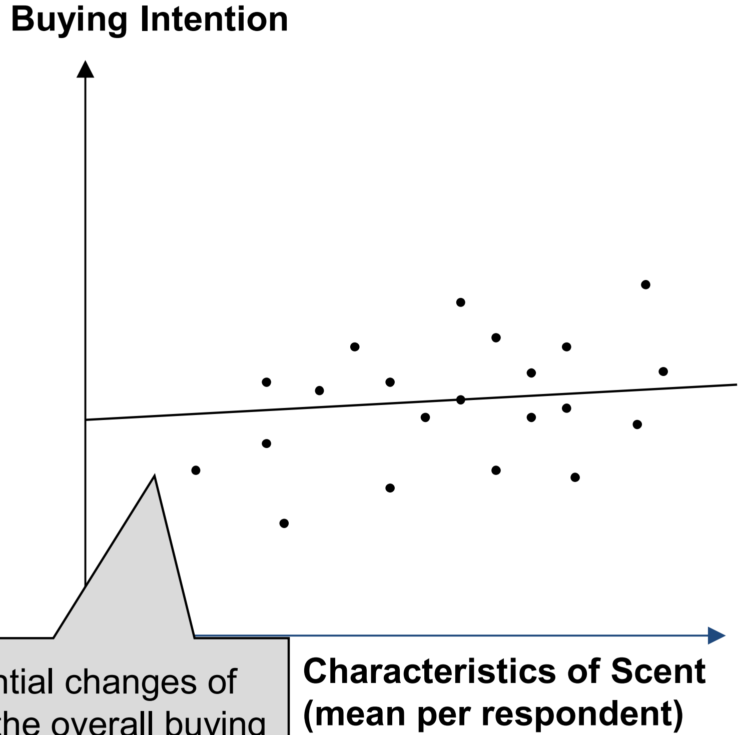
Shampoo Product Test

For shampoo A we should perform well regarding the characteristics 1 and 2. However, for the characteristics 3 and 4 there is a risk of over engineering.

Strong relationship between shampoo characteristics and buying intention → High importance



Weak relationship between shampoo characteristics and buying intention → Low importance

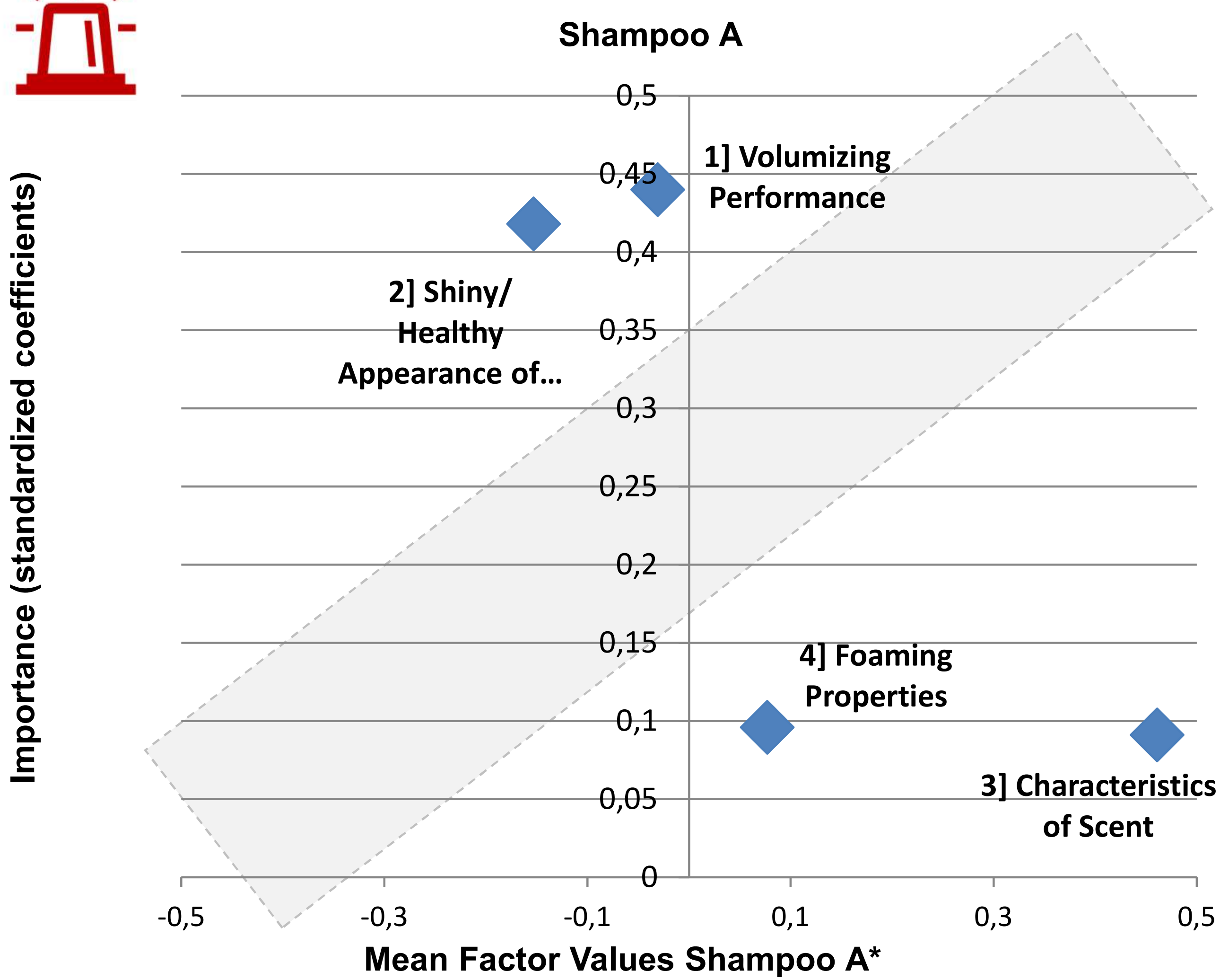


Even for substantial changes of the agreement, the overall buying intention hardly changes

Analysis of Dependence:
Multiple Regression Analysis – Basic Analysis (17)

Shampoo Product Test

We underperform in the important factors and overperform/overengineer in the unimportant factors!



*mean factor values across all shampoos set to zero

Multiple Regression Analysis:
Example shampoo product test – Summary

UNIVERSITY OF MANNHEIM
Business School

What was the initial problem?

- We wanted to find out to which extent the shampoos' characteristics have an influence on the respondents' buying intentions
- Besides, we wanted to know if Shampoo A already fulfills the most important characteristics for the buying intention

What did we do?

- Supposing that all the assumptions for a regression analysis are fulfilled, we conducted a multiple regression analysis with a statistical software (e.g., R Statistics)

What was the outcome?

- The results showed us that all the factors have a significant influence on the buying intention with "Volumizing Performance" being the strongest driver and "Characteristics of Scent" being the lowest driver
- Shampoo A is rated moderately on the most important factors. At the same time, it is overengineered on "Characteristics of Scent" and "Foaming Properties" which are not of a very high importance for the buying intention!

Analysis of Dependence: Multiple Regression Analysis – Excursus (1)

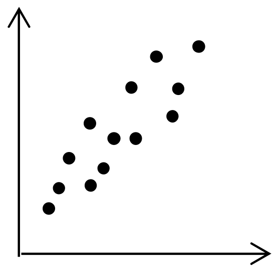
Excursus: Further Models of Regression Analysis

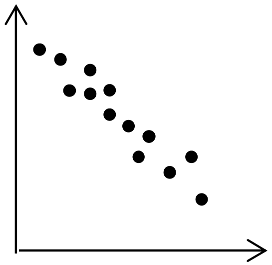
- Non-linear regression analysis
- Regression analysis with interaction terms/moderated regression analysis

Analysis of Dependence:
Multiple Regression Analysis – Excursus (2)

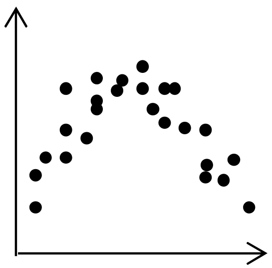
- Non-Linear Regression Analysis

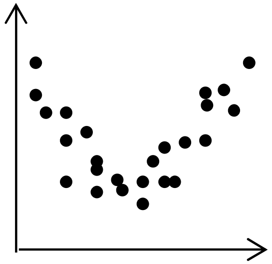
- Linear regression:



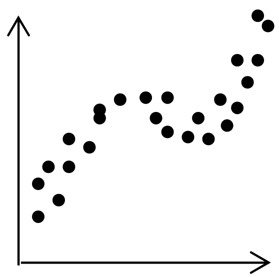


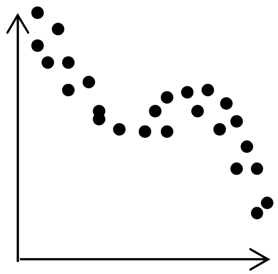
- Quadratic regression:





- Cubic regression:





•Uncovering non-linear relationships by using, for example, scatter plots or polynomial regression (see Klarmann 2008)

Analysis of Dependence: Multiple Regression Analysis – Excursus (3)

Polynomial regression model of order r (one independent variable):

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{1i}^2 + \ldots + b_r x_{1i}^r + e_i$$

- Test for non-linearity
  - H_0: $b_2 = \ldots = b_r = 0$ versus H_1: at least one $b_j \neq 0$, j = 2, …, r
  - F-test
- Identification of polynomial order r
  - Sequence of t tests

    $$H_0^1: \; b_{r_{max}} = 0 \text{ versus } H_1^1: \; b_{r_{max}} \neq 0$$

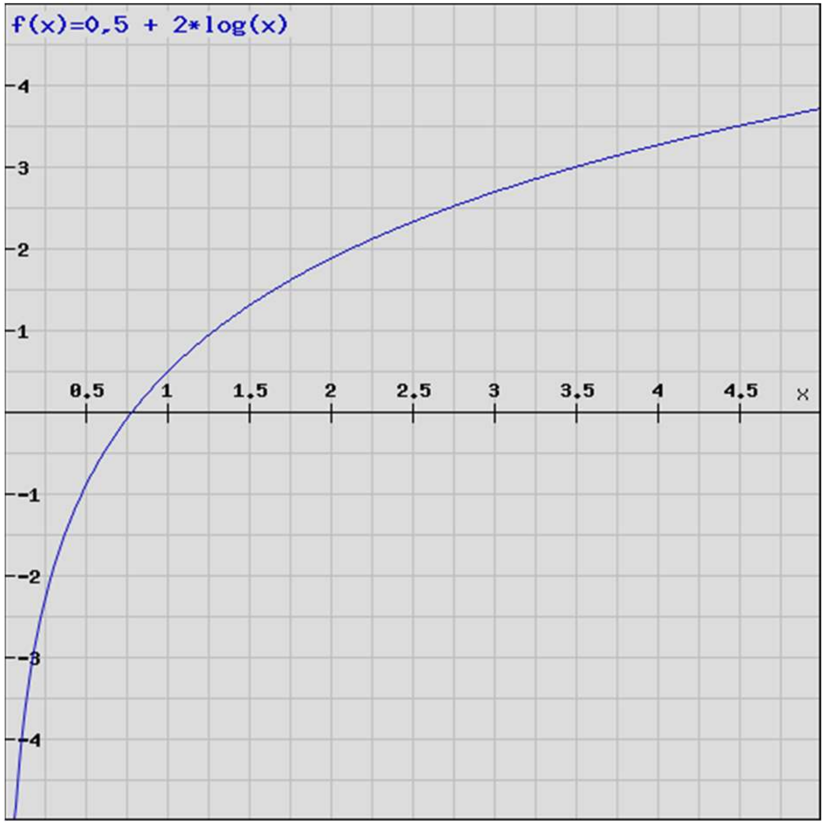
  - In case of a rejection of $H_0^1: r = r_{max}$
  - In case of a non-rejection of H_0^1: Until the first rejection, test for

    $$H_0^2: \; b_{r_{max-1}} = 0 \text{ versus } H_1^2: \; b_{r_{max-1}} \neq 0$$

  - For r = 2: quadratic regression model, for r = 3: cubic regression model

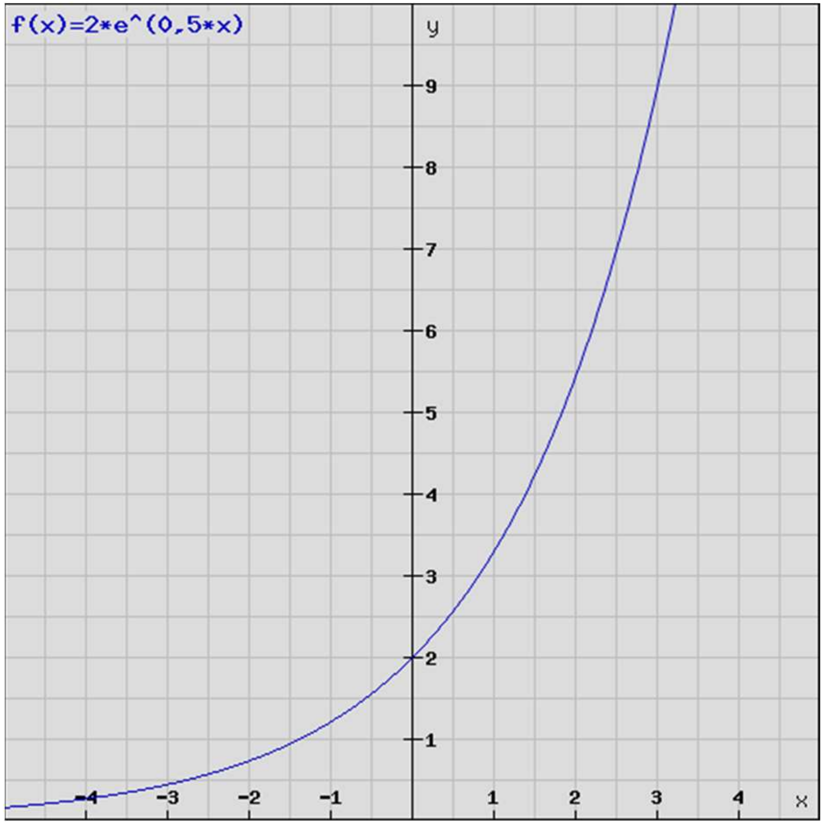
Further Examples of Non-Linear Regression Models

Logarithmic Regression



$$y_i = b_0 + b_1 \cdot \ln(x_{1i})$$

Exponential Regression

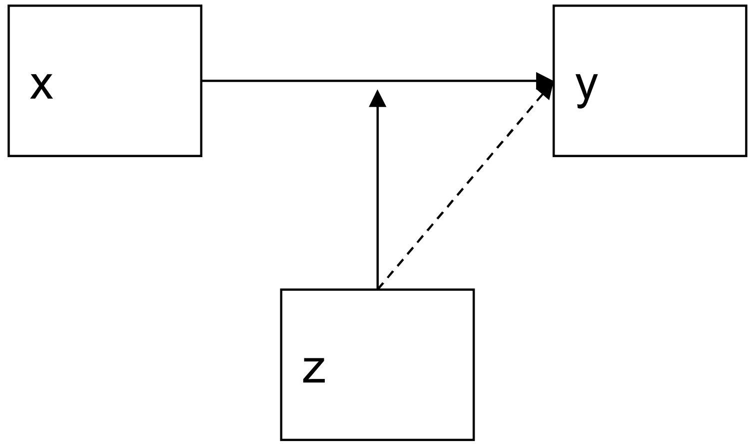


$$y_i = a \cdot e^{b_1 x_{1i}}$$

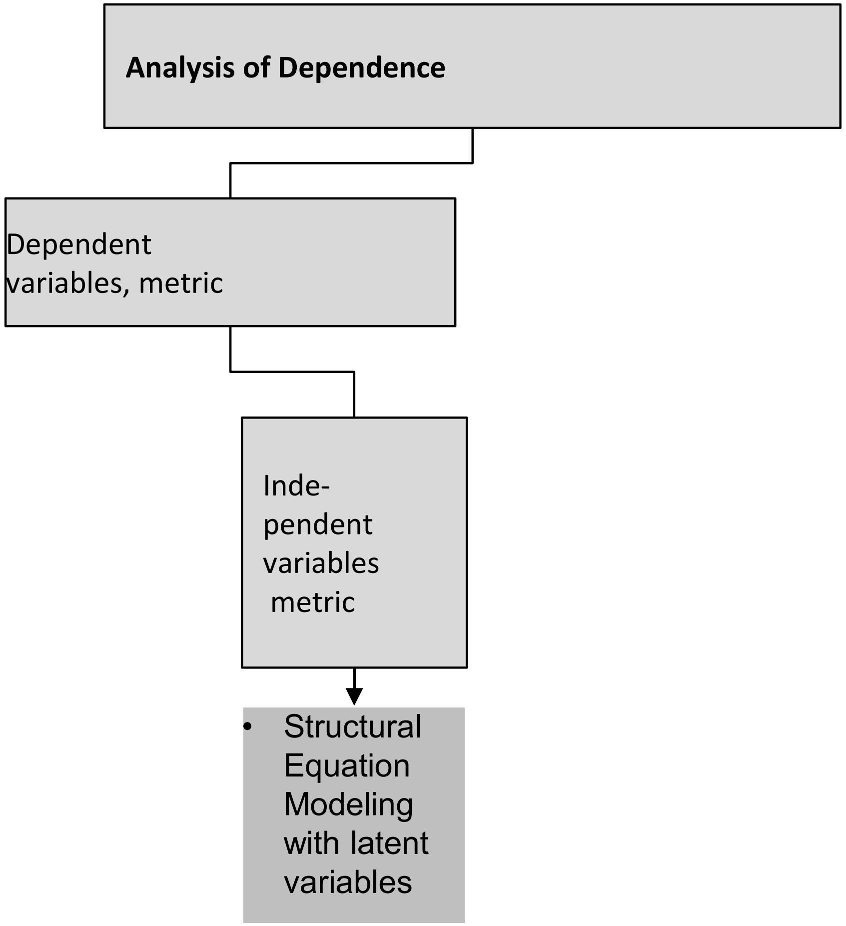
Analysis of Dependence:
Multiple Regression Analysis – Excursus (8)

Moderated Regression Analysis

- The strength and/or direction of the effect of an independent variable x (IV) on a dependent variable y (DV) is influenced by a third variable z (moderator)
  - Formal
    - (1) $y = a + bx + e$
    - (2) $b = c + dz$
    - (2) in (1): $y = a + cx + dzx + e$
    - Estimated model, including the effect of z:
      $y = a + cx + dzx + fz + e$
- Sharma, Durand, and Gur-Arie (1981) distinguish
  - Pure moderators: $f=0$
  - Quasi moderators: $f \neq 0$



Analysis Using Dependence Techniques: Structural Equation Modeling (1)



Homburg (2015), p. 359; Hair et al. (2010), p. 12

Analysis Using Dependence Techniques: Structural Equation Modeling (2)

UNIVERSITY OF MANNHEIM
Business School

Problem

- We conducted another survey about Shampoo A with which we want to find out…
  – In how far the manufacturer's CSR reputation has an influence on the customers' trust and the customer-company identification
  – In how the CSR reputation, customers' trust and customer-company identification affect customer loyalty
- These complex effects cannot be calculated properly with a simple multiple regression analysis
- Besides, we want to estimate the measurement model simultaneously to the structural model

Structural equation modeling

Analysis Using Dependence Techniques: Structural Equation Modeling (3)

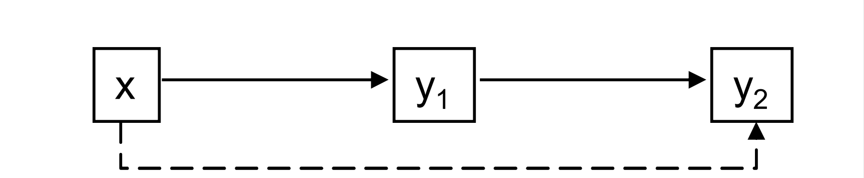
Basic idea: Empirical measurement of variances and covariances of the measured variables (sample)

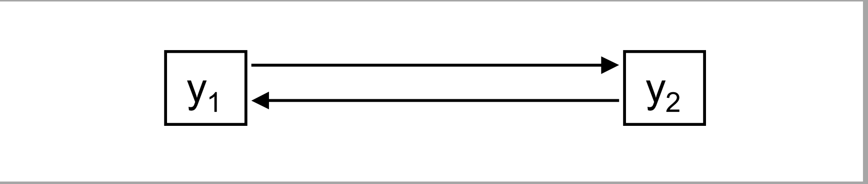
→ Examining the structure of interrelationships among latent factors (parameter estimation)

Method of the 2nd generation→ Overcoming the deficits of the multiple regression analysis

- Assumption that measurement errors can be overlooked→ SEM corrects for the amount of measurement error in the variables
- Problems with “too strong" correlations among the independent variables→ explicit consideration of correlations
- Complex structural models cannot be estimated→ in principle any structural model can be investigate

Examples of complex structural models

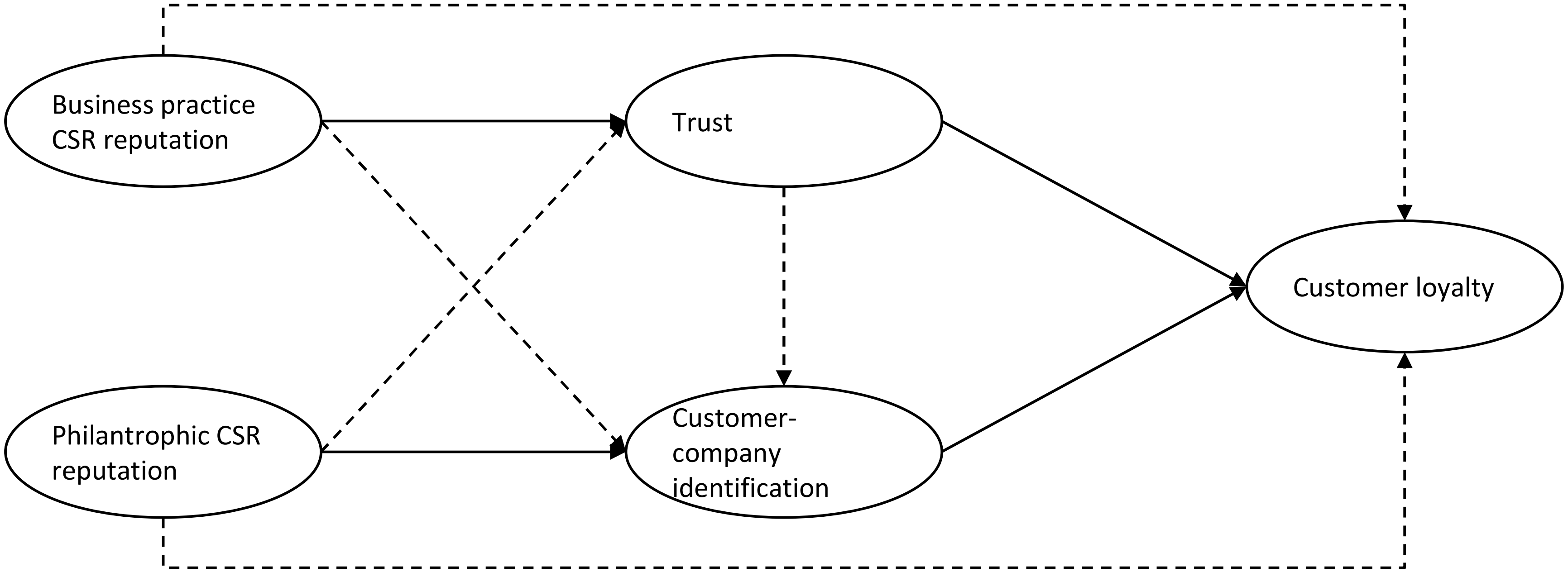




Homburg (2015), p. 395; Hair et al. (2010), p. 634

Analysis Using Dependence Techniques: Structural Equation Modeling (4)

Shampoo Survey: Example of SEM



Information regarding the measured variables **(= indicator variables)** used

- 5 indicators for measuring Business practice CSR reputation
- 4 indicators for measuring Philantrophic CSR reputation
- 4 indicators for measuring Trust
- 5 indicators for measuring Customer-company identification
- 4 indicators for measuring Customer loyalty

Adapted from Homburg, Stierl, and Bornemann 2013, Corporate Social Responsibility in Business-to-Business Markets: How Organizational Customers Account for Supplier Corporate Social Responsibility Engagement, Journal of Marketing 2013

Analysis Using Dependence Techniques: Structural Equation Modeling (5)

Confirmatory Approach

1. Specifying the model
2. Parameter estimation
3. Assessing the model validity
4. If necessary: Model respecification
5. Interpretation of Results

Analysis Using Dependence Techniques: Structural Equation Modeling (6)

Structure of the model

- Differentiation between exogenous (ξ) & endogenous latent variables (η)
- Structural equation model (1) and measuring models (2, 3)

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta \quad (1)$$

$$y = \Lambda_y \eta + \varepsilon \quad (2)$$

$$\mathbf{x} = \Lambda_{\mathbf{x}} \xi + \delta \quad (3)$$

- Table to illustrate the components of the equations:

Symbol	Pronunciation	Meaning
η	Eta	(k x 1)- vector of the endogenous latent variables
y	-	(p x 1)-vector of measured variables to η-variables
Λ_y	Lambda-y	(p x k)-matrix of coefficients λ^y, which specifies the assignment of the η-variables to the y-variables (factor loadings)
ε	Epsilon	(p x 1)-vector of the error term associated with estimated, measured y -variables
ξ	Ksi	(l x 1)-vector the exogenous latent variables
x	-	(q x 1)-Vector of the measured variables to the ξ-variables
Λ_x	Lambda-x	(q x l)- matrix of coefficients λ^x, which specifies the assignment of the ξ-variables to the x-variables (factor loadings)
δ	Delta	(q x 1)- vector of the error term associated with estimated, measured x -variables
B	Beta	(k x k)-matrix of coefficients β, which specifies the causal dependencies between η-variables
Γ	Gamma	(l x l)- matrix of coefficients γ, which specifies the causal dependencies between ξ-variables and η-variables
ζ	Zeta	(k x 1)-vector of error terms ζ in the structural model, each of them belonging to one η-variable

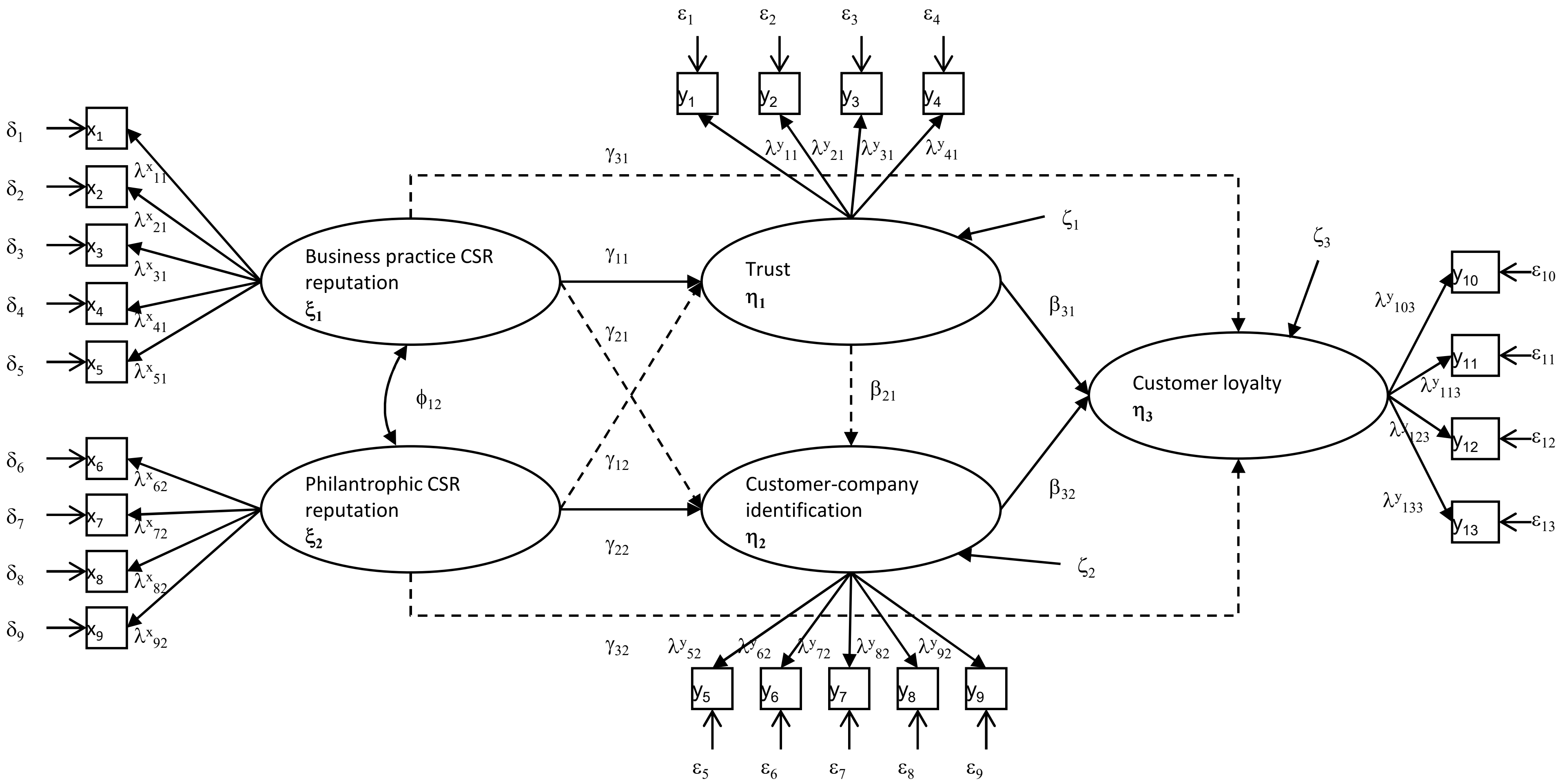
Lattin, Carroll, and Green 2003, pp. 358

Remark:

k = Number of η-Variables
l = Number of ξ-Variables
p = Number of y-Variables
q = Number of x-Variables

Analysis Using Dependence Techniques: Structural Equation Modeling (7)

Shampoo Survey: Complete structural equation model



Analysis Using Dependence Techniques: Structural Equation Modeling (8)

Shampoo Survey: Complete structural equation model

Structural model

$$\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\end{pmatrix}=\begin{pmatrix}0&0&0\\ \beta_{21}&0&0\\ \beta_{31}&\beta_{32}&0\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\end{pmatrix}+\begin{pmatrix}\gamma_{11}&\gamma_{12}\\ \gamma_{21}&\gamma_{22}\\ \gamma_{31}&\gamma_{32}\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}+\begin{pmatrix}\zeta_1\\ \zeta_2\\ \zeta_3\end{pmatrix}$$

Measurement model

$$\begin{pmatrix}x_1\\x_2\\x_3\\x_4\\x_5\\x_6\\x_7\\x_8\\x_9\end{pmatrix}=\begin{pmatrix}\lambda^x_{11}&0\\ \lambda^x_{21}&0\\ \lambda^x_{31}&0\\ \lambda^x_{41}&0\\ \lambda^x_{51}&0\\ 0&\lambda^x_{62}\\ 0&\lambda^x_{72}\\ 0&\lambda^x_{82}\\ 0&\lambda^x_{92}\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}+\begin{pmatrix}\delta_1\\ \delta_2\\ \delta_3\\ \delta_4\\ \delta_5\\ \delta_6\\ \delta_7\\ \delta_8\\ \delta_9\end{pmatrix}$$

$$\begin{pmatrix}y_1\\y_2\\y_3\\y_4\\y_5\\y_6\\y_7\\y_8\\y_9\\y_{10}\\y_{11}\\y_{12}\\y_{13}\end{pmatrix}=\begin{pmatrix}\lambda^y_{11}&0&0\\ \lambda^y_{21}&0&0\\ \lambda^y_{31}&0&0\\ \lambda^y_{41}&0&0\\ 0&\lambda^y_{52}&0\\ 0&\lambda^y_{62}&0\\ 0&\lambda^y_{72}&0\\ 0&\lambda^y_{82}&0\\ 0&\lambda^y_{92}&0\\ 0&0&\lambda^y_{103}\\ 0&0&\lambda^y_{113}\\ 0&0&\lambda^y_{123}\\ 0&0&\lambda^y_{133}\end{pmatrix}\begin{pmatrix}\eta_1\\ \eta_2\\ \eta_3\end{pmatrix}+\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_4\\ \varepsilon_5\\ \varepsilon_6\\ \varepsilon_7\\ \varepsilon_8\\ \varepsilon_9\\ \varepsilon_{10}\\ \varepsilon_{11}\\ \varepsilon_{12}\\ \varepsilon_{13}\end{pmatrix}$$

Analysis Using Dependence Techniques: Structural Equation Modeling (9)

Parameter estimation

- Determining a set of parameters, which reduces the values within the matrix of residuals

  $F(S, \Sigma(\alpha)) \rightarrow$ min.

  S: Observed covariance matrix of indicator variables

  $\Sigma(\alpha)$: Estimated covariance matrix of indicator variables by the model

- Estimation Techniques
  - Maximum Likelihood Estimation (MLE)
  - Method of Unweighted Least Squares (ULS)
  - Method of Weighted Least Squares (WLS)
- Prerequisite: The specified model must be identified, that is
  - The covariance matrix of the indicators provides enough information for an unambiguous estimation of the model parameter
  - The number of parameters to be estimated is not allowed to exceed the number of observed variances and covariances

$$t \leq \frac{q \cdot (q+1)}{2}$$

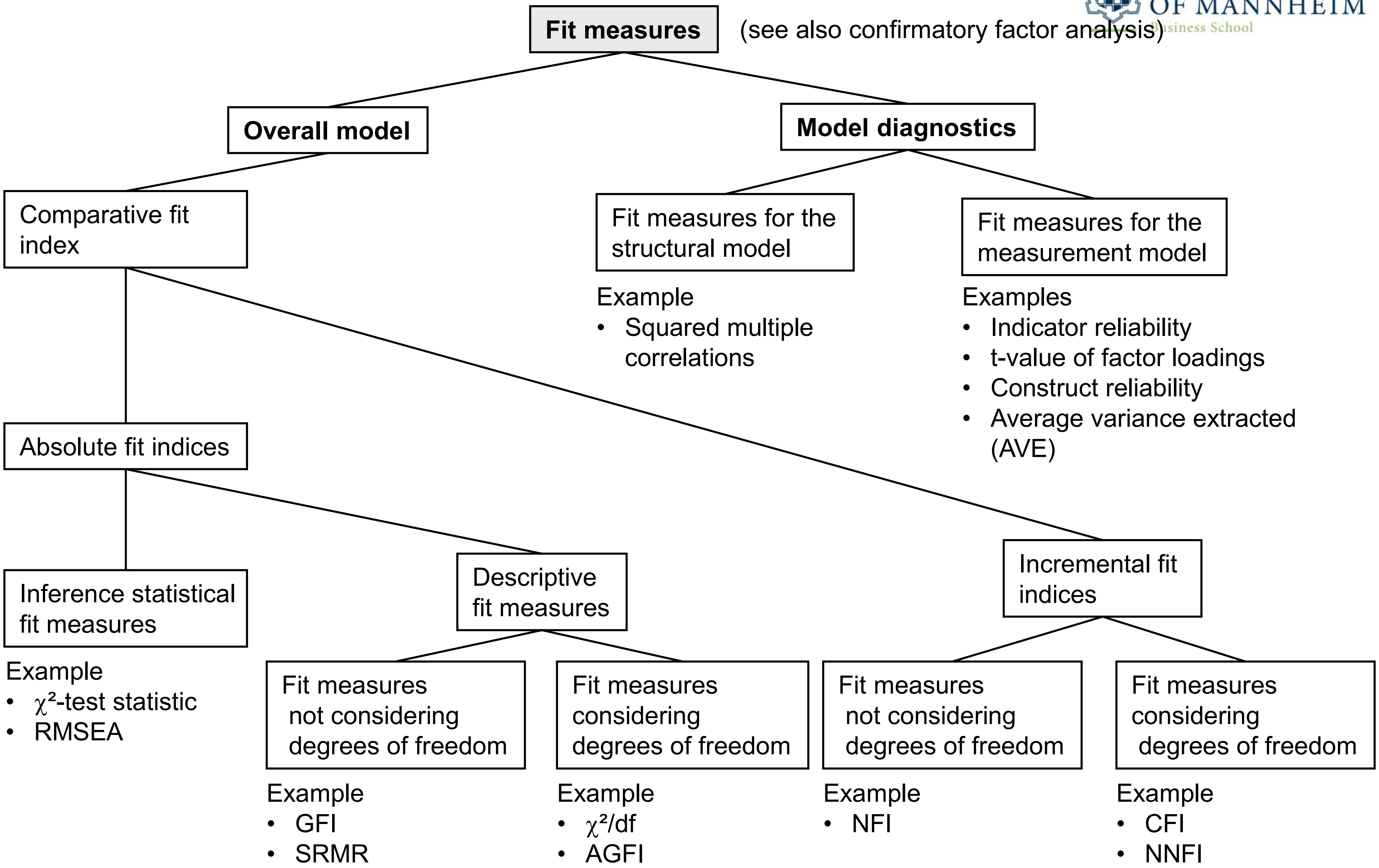
t: Number of parameters to be estimated

q: Number of indicators

→ Indications for not identified models are large standard errors or unclear estimator (e.g. negative error variances).

Hair et al.(2010), pp. 650

Analysis Using Dependence Techniques: Structural Equation Modeling (10)



Homburg/Pflesser/Klarmann (2008)

UNIVERSITY OF MANNHEIM
Business School

Ideal approach for model evaluation

1. **Review of formal aspects**
   Existence of incomprehensible estimator? Model identification given?
2. **Assessing the goodness of fit for the measurement model**
   Construct validity of the measurement model
3. **Assessing the goodness of fit for the overall model**
   Overall fit measures
4. **Assessing the goodness of fit for the structural model**
   Fit measures of the structural model
5. **Comparison of alternative model structures**
   Comparison of fit measures of the model diagnostics with fit measures of the overall model

Homburg/Pflesser/Klarmann (2008); Hair et al. (2010), pp.706

Analysis Using Dependence Techniques: Structural Equation Modeling (12)

Shampoo Product Test: Parameter estimation and Goodness of fit

UNIVERSITY OF MANNHEIM
Business School

Results of the measurement model

		Indicator Reliability	Construct Reliability	AVE
	x_1	.65		
Business practice	x_2	.69		
CSR reputation	x_3	.46	.87	.56
(ξ_1)	x_4	.50		
	x_5	.52		
	x_6	.65		
Philantrophic CSR	x_7	.80		
reputation (ξ_2)	x_8	.86	.94	.79
	x_9	.85		
	y_1	.76		
Trust (η_1)	y_2	.89	.90	.69
	y_3	.66		
	y_4	.45		
	y_5	.72		
Customer-	y_6	.80		
company identification (η_2)	y_7	.70	.92	.69
	y_8	.68		
	y_9	.57		
	y_{10}	.68		
Customer loyalty	y_{11}	.66		
(η_3)	y_{12}	.44	.84	.57
	y_{13}	.51		

Results of the structural model

Model parameter	stand. coefficients
γ_{11}	.316**
γ_{12}	-.020
γ_{21}	.033
γ_{22}	.147*
γ_{31}	.061
γ_{32}	-.088
β_{21}	.560**
β_{31}	.208*
β_{32}	.328**

+ $p < .10$ * $p < .05$ ** $p < .01$

Selected fit measures of the overall model:

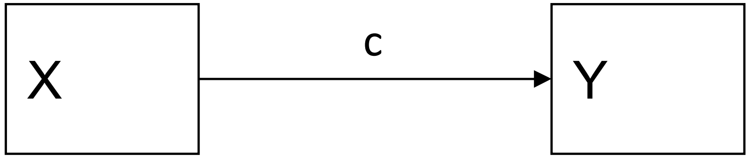
CFI = 0.92

$\chi 2/df = 1.71$

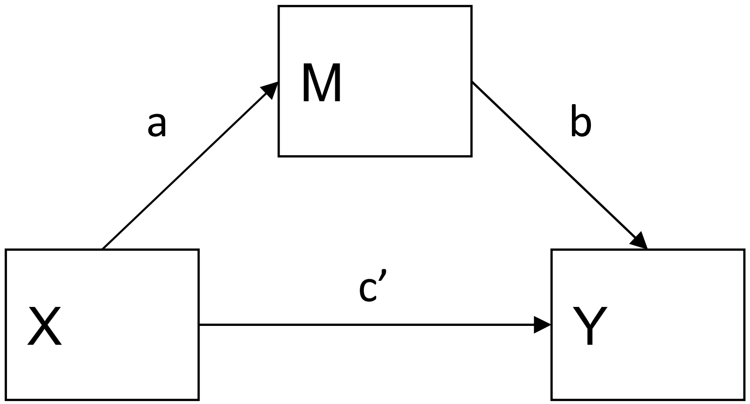
RMSEA = 0.06

Mediated Models

The Unmediated Model



The Simple Mediated Model



We consider three types of variables

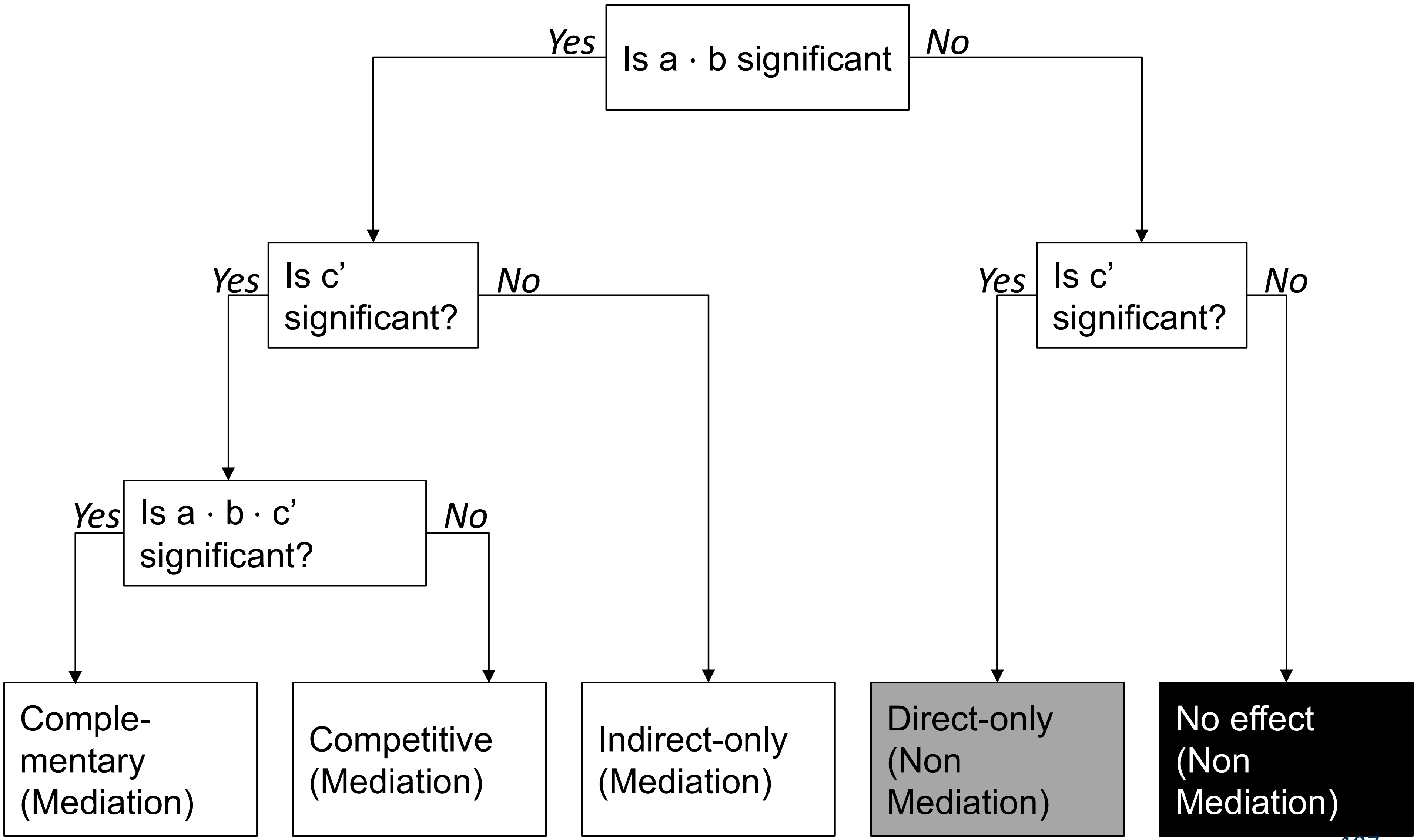
- X = independent variable
- Y = dependent variable
- M = mediating variable

Thus, there are three types of effects on the dependent variable:

- c = total effect
- a * b = indirect effect
- c' = direct effect

Analysis Using Dependence Techniques:
Structural Equation Modeling – Excursus (2)

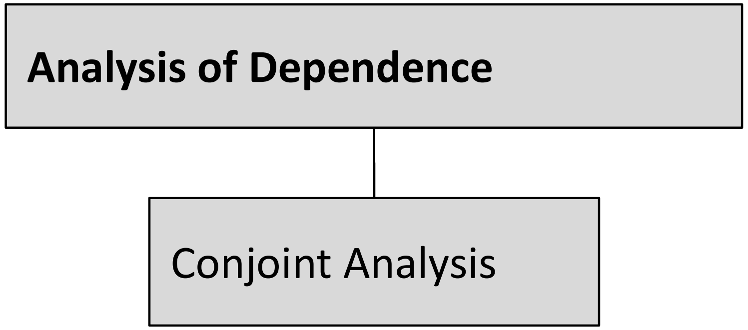
Mediated Models: Establishing Mediation & Classifying Type



Structural Equation Modeling:
Example shampoo product test – Summary

What was the initial problem?	• We wanted to calculate the relationships between several constructs which resulted in a complex structural model that cannot be calculated with a multiple regression analysis
What did we do?	• We specified the structural model as well as the measurement model • We estimated the parameters by the help of R Statistics • We assessed the model validity by the help of several fit measures. Further model specifications were not necessary
What was the outcome?	• The results confirmed our initial model with a high model validity

Analysis of Dependence: Conjoint Analysis (1)



Homburg (2015), p. 359; Hair et al. (2010), p. 12

Analysis of Dependence: Conjoint Analysis (2)

Problem

- We work for the market research department of an automobile manufacturer who wants to launch an new mid-range car
- However, we have endless options to configure this car
- Therefore, we want to find out how a new mid-range car should be constructed in order to fit our potential customers' preferences best

Conjoint Analysis

Analysis of Dependence: Conjoint Analysis (3)

The Conjoint Analysis is a multivariate technique developed specifically to understand how respondents develop preferences for any type of products, services, or ideas. It is based on the simple premise that customers evaluate the value of an object by combining the separate amounts of value provided by each attribute. [...] customers can best provide their estimates of preferences by judging objects formed by combinations of attributes.

(Hair et al. 2010, p. 266)

- *Decompositional Model*
- *Additive Model of Utility:* Total score of Utility = Sum of "part-utilities"
- Managerial Use of Conjoint Analysis
  - *Product Design*, e.g., What kind of product or service is preferred by a customer in comparison to other products/services?
  - *Price Policy*, e.g., How much is a new product‘s attribute utility in terms of costs?
  - *Market Segmentation*, e.g., What product attributes provide an exceptionally high utility in particular market segments?
  - *Simulation Model*, e.g., What is the impact level of a new product on the market share of its rivals?

Analysis of Dependence: Conjoint Analysis (4)

Reference criteria	Traditional Conjoint Analysis	Choice-Based Conjoint Analysis
Consideration of uncertainty in the specified utility function	• Deterministic utility function	• Total utility consists of a deterministic and stochastic component
Information content and realism	• Metric or nonmetric evaluation of stimuli • Relatively high information content	• Selection decision is observed (dichotomous/nominal) • Relatively low information content • More realistic
Interpretation of the estimated part-utilities	• Estimation of individual part-utilities and total utilities • No systematic biases • Researchers makes an assumption concerning the selection decision	• Usually aggregated estimation of part-utilities and total utilities • Systematic bias: utilities could be biased due to the differences between the respondents • Selection decision is integrated into the model

"Yet, in many situations, two or more methodologies are feasible and the researcher has the option of selecting one methodology or combining the methodologies. Only by being knowledgeable about the strengths and weaknesses of each methodology can the researcher make the more appropriate choice." (Hair et al. 2010, p. 312)

According to Gensler 2003, p. 74

Analysis of Dependence: Conjoint Analysis (5)

Procedure

1. Specifying the determinant factors
2. Selecting a Conjoint Analysis methodology
3. Data collection
4. Estimating the preferences and utilities
5. Assessing the overall fit
6. Interpreting the results

Analysis of Dependence: Conjoint Analysis (6)

Example: Mid-Range Car

Step 1: Specifying the determinant factors

Factor: variable a researcher manipulates that represents a specific product attribute

- Factors should be relevant for buying decision (key determinant variables)
- All determinant factors (price, quality,...) must be included
- Factors must be manipulable by firm
- Important competitive products must be representable by factors
- Factors should be independent of each other
- Compensatory relationship between factors
- No exclusion criteria
- Limited number of determinant factors and levels

→ **Factors for mid-range car**: Price, Engine Power, Service, Gas Consumption

Analysis of Dependence: Conjoint Analysis (7)

Example: Mid-Range Car

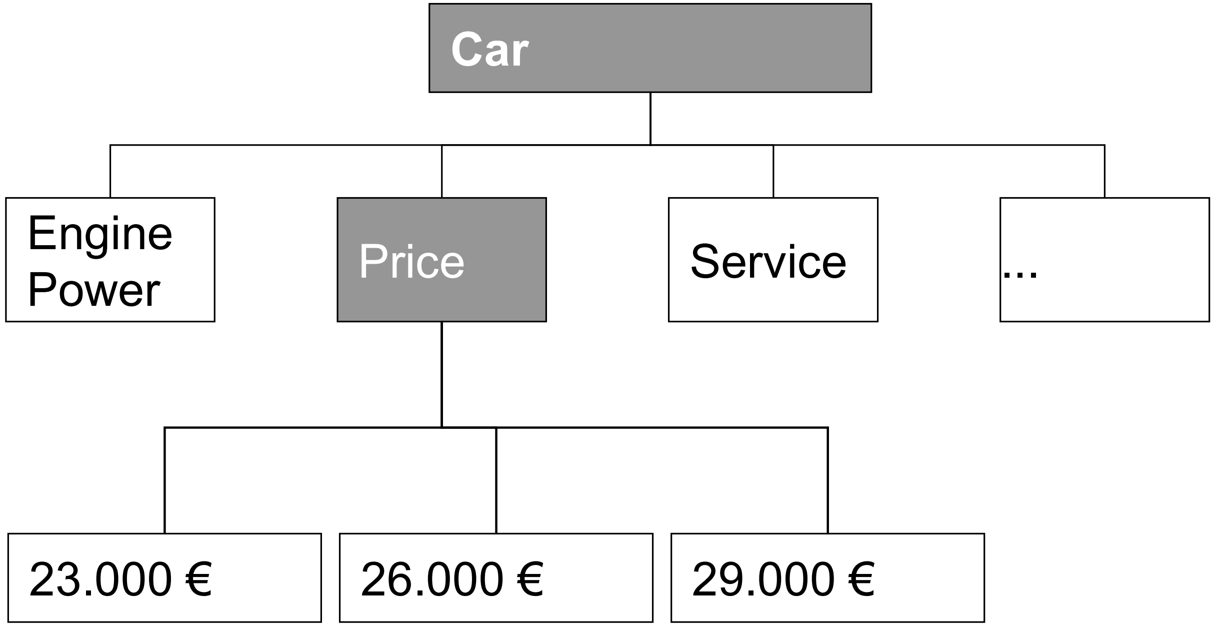
UNIVERSITY OF MANNHEIM
Business School

Basic Idea

Product or Service can be subdivided in certain attributes, e.g. Engine Power, Price, etc.

Each attribute is described by a range of different value levels.

Example



AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS AMBA ACCREDITED

Analysis of Dependence: Conjoint Analysis (8)

Example: Mid-Range Car

•Step 2: Selecting a Conjoint Analysis Methodology

Traditional Conjoint Methodology *(Full-Profile Method):*

•Evaluation of profiles (constructed by each level of every considered factors)

Car A

Engine power: 60 KW
Free services
Price: 23.000 €
Gas consumption: 7 l per 100 km

Car B

Engine power: 70 KW
Fee-based services
Price: 23.000 €
Gas consumption: 9 l per 100 km

Preference – Car A **undecided** **Preference – Car B**

1	2	3	4	5	6	7	8	9

Number of Stimuli

- As the number of factors and levels of the full-profile method increases, the design becomes complex and impractical
- Number of Stimuli: $n_1 \cdot n_2 \cdot ... \cdot n_j$
  (j: Set of factors, n_i: Set of levels of factor i)

Adaptive Conjoint Analysis (ACA)

Goal: Reducing the number of comparisons to accommodate a large number of factors

1. Direct assessment of the importance of each factor and the part-utilities for each factor level
2. Exclusion of profiles which will not be considered by the participant according to his/her factor rating
3. Employment of a computerized process that adapts the choice sets as the choice task proceeds and new preference information are generated

Analysis of Dependence: Conjoint Analysis (10)

Example: Mid-Range Car

Step 3: Data Collection

Example: Data Collection in case of an Adaptive Conjoint Analysis (I)

Assessment of preferences for each attribute

- Measurement of the preference for each single level of an attribute
- Assessing preferences by a 5-point rating scale
- Endpoints: ‘absolutely not desirable‘ – ‘highly desirable‘

0% 100%

This part of the survey introduces you to **fictitious features of a mid-range car**. Please imagine a mid-range car with the corresponding characteristic features.

Please evaluate the following features of a mid-range car on a scale from “absolutely not desirable” to “highly desirable” according to how important it is to you.

	absolutely not desirable				highly desirable
The mid-range car has a low price	○	○	○	○	○
The mid-range car has a free service in the first year after purchase	○	○	○	○	○
The mid-range car has a very high engine power in it’s category	○	○	○	○	○
The mid-range car consumes relatively has a very low fuel consumption in it’s category	○	○	○	○	○

Next question

Analysis of Dependence: Conjoint Analysis (11)

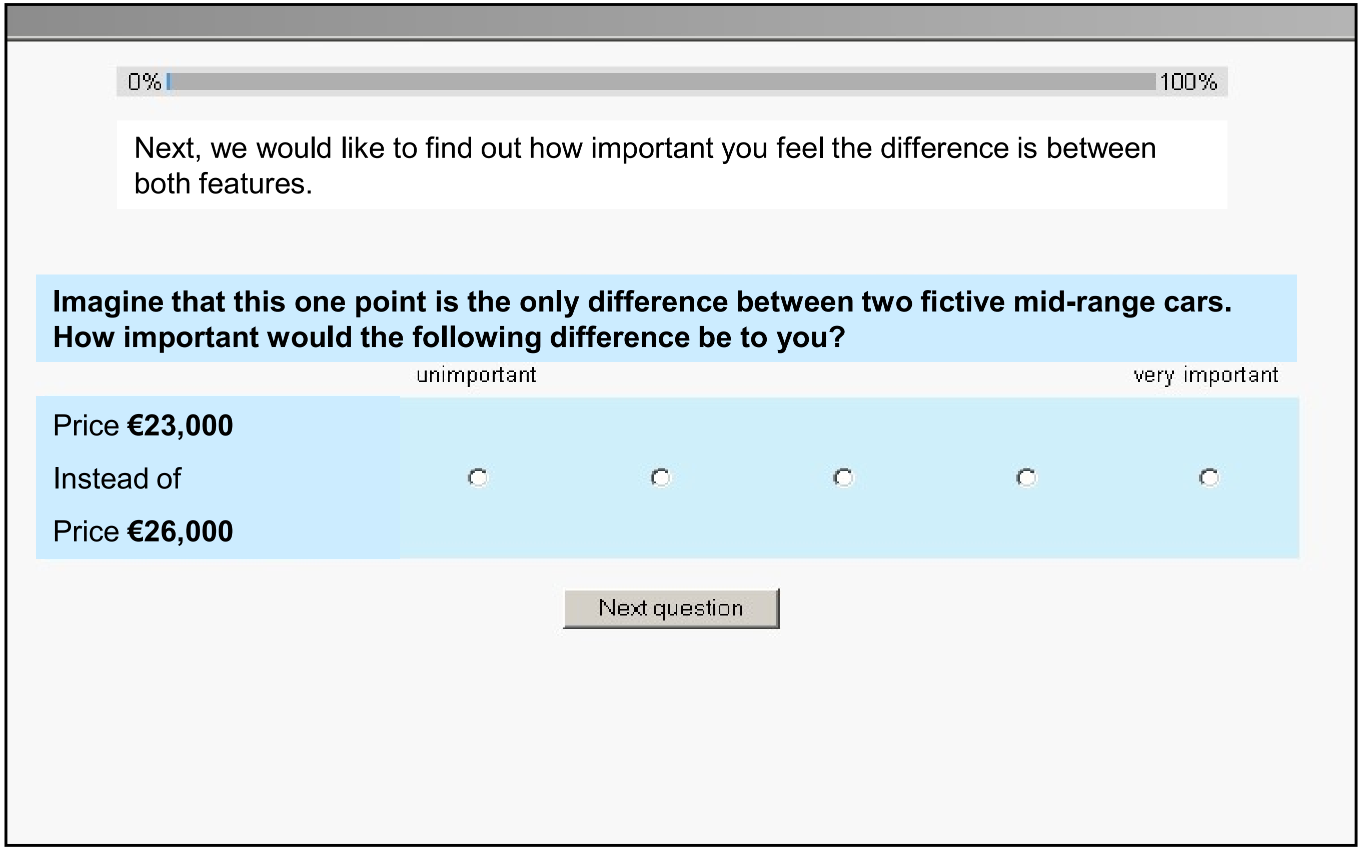
Example: Mid-Range Car

Step 3: Data Collection

Example: Data Collection in case of an Adaptive Conjoint Analysis (II)

Determining the relative importance of each attribute

- Measuring the difference between the most preferred and least preferred level of a certain attribute
- 4- to 9-point rating scale
- Endpoints: ‘unimportant‘ – ‘very important‘



Analysis of Dependence: Conjoint Analysis (12)

Example: Mid-Range Car

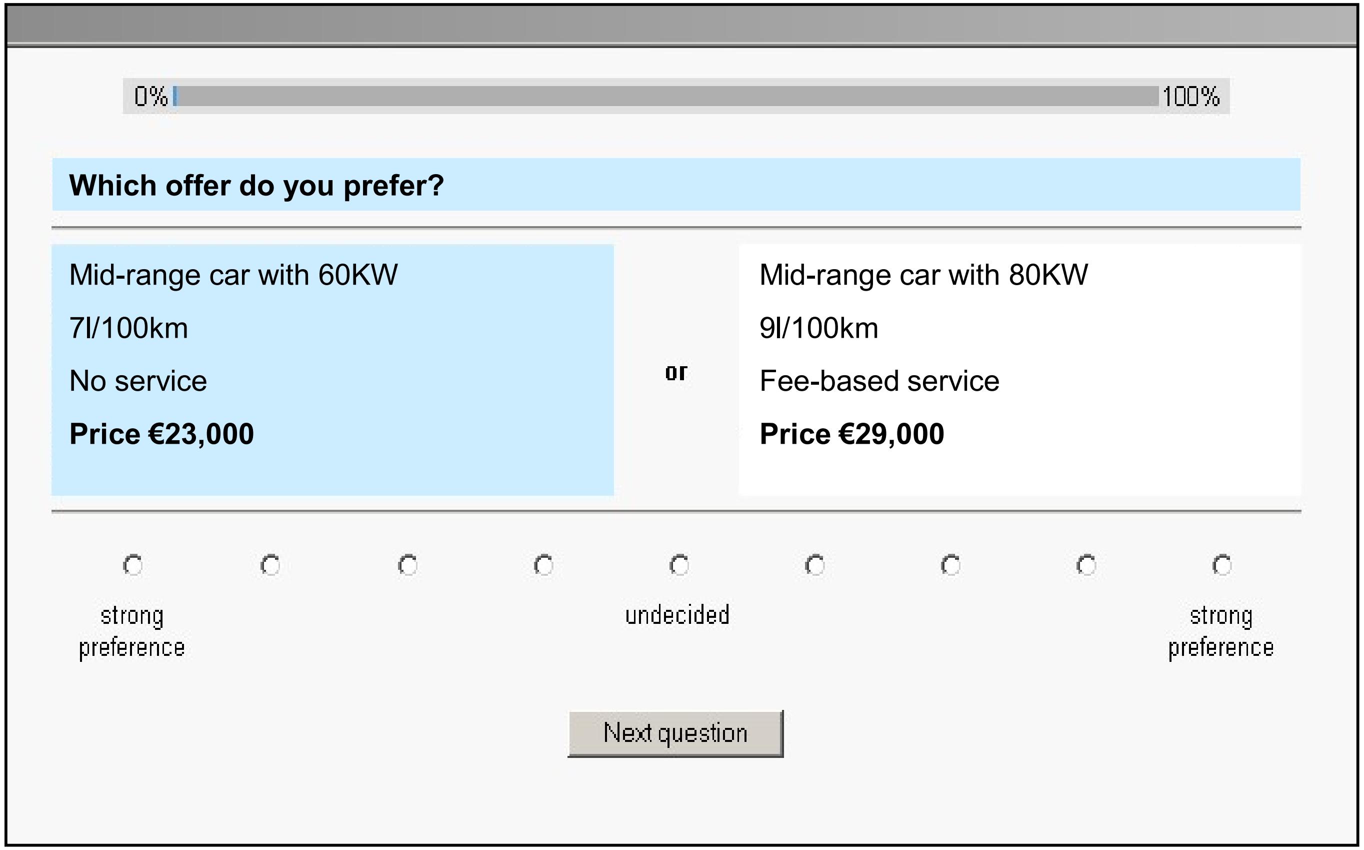
Step 3: Data Collection

Example: Data Collection in case of an Adaptive Conjoint Analysis (III)

Paired Comparisons

- Measuring the preference for the product concepts on the left/right-hand side
- 3 to 5 attributes per product concept
- 7- to 9-point rating scale
- Endpoints: ‘strong preference (left-hand side)‘ – ‘undecided‘ – ‘strong preference (right-hand side)‘

The mid-range car comes with a free service in the first year after purchase.



Analysis of Dependence: Conjoint Analysis (13)

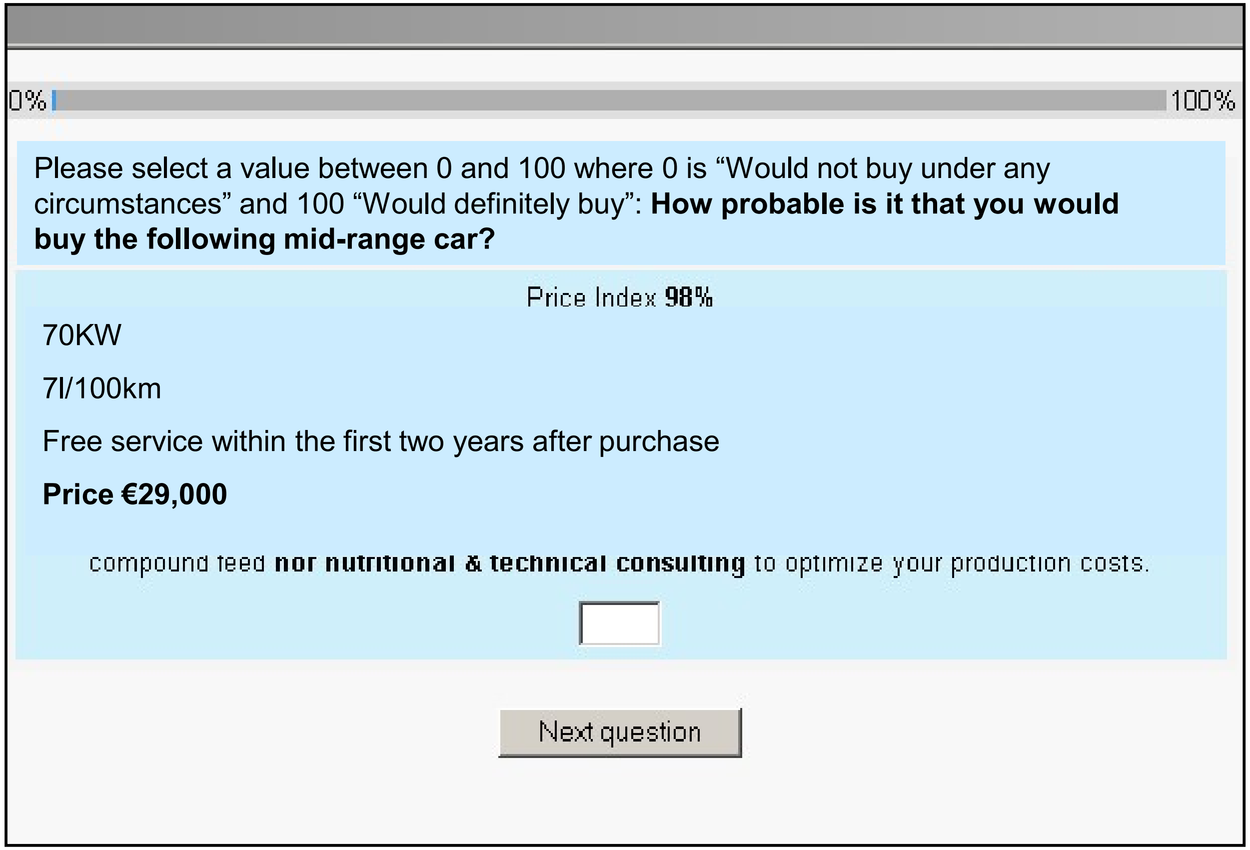
Example: Mid-Range Car

Step 3: Data Collection

Example: Data Collection in case of an Adaptive Conjoint Analysis (IV)

Questions to calibrate the results

- Measuring the disposition to buy for several product concepts
  a. The least preferred concept
  b. The most preferred concept
  c. A concept between both extremes
- Scale from 0% to100%
- Method used to normalize the relative importance and preference



UNIVERSITY OF MANNHEIM
Business School

Step 4: Estimating the preferences and utilites

- Additive Model

$$\mathbf{y_k = \sum_{j=1}^{J} \sum_{m=1}^{M_j} \beta_{jm} \cdot x_{jmk}}$$

y_k : total utility for product **k**
β_{jm} : part-utility of level **m** for factor/attribute **j**

$$x_{jmk} = \begin{cases} 1 & \text{if product } \mathbf{k} \text{ has the attribute } \mathbf{j} \text{ in a certain level m} \\ 0 & \text{otherwise} \end{cases}$$

- Estimation of each part-utility β_{jm} in such a way that the resulting total utility for a product y_k reflects the decision process in an adequate manner
- Objective function: Minimizing the sum of the squares of the deviation between the empirical observed and the estimated total utility value for a product

Analysis of Dependence: Conjoint Analysis (15)

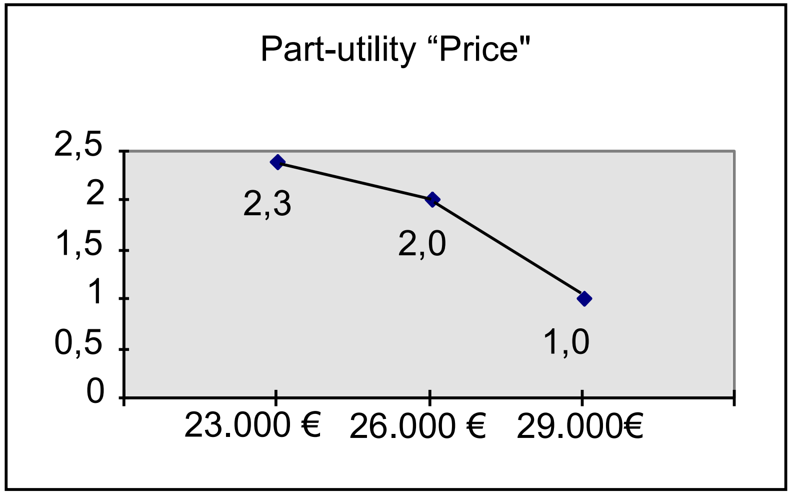
- Different methods of estimation
  - **Aggregated estimation: Part-utility estimation for the whole sample; *requirement: homogenous structure of utility***
  - Segmental estimation: Part-utility estimation for different respondent segments
  - Individual (*Bayesian*) estimation: Part-utility estimation on an individual level (followed by an aggregation to different segments or the whole sample)
- Determining the relative importance of each attribute

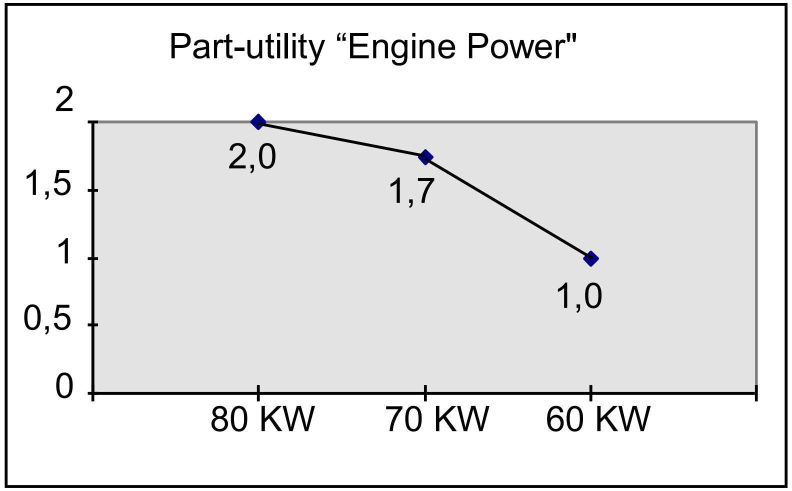
$$w_j = \underset{m}{\mathrm{Max}}\{\beta_{jm}\} - \underset{m}{\mathrm{Min}}\{\beta_{jm}\} \longrightarrow \hat{w}_j = \frac{w_j}{\sum_{j=1}^{J} w_j} \cdot 100\ [\%]$$

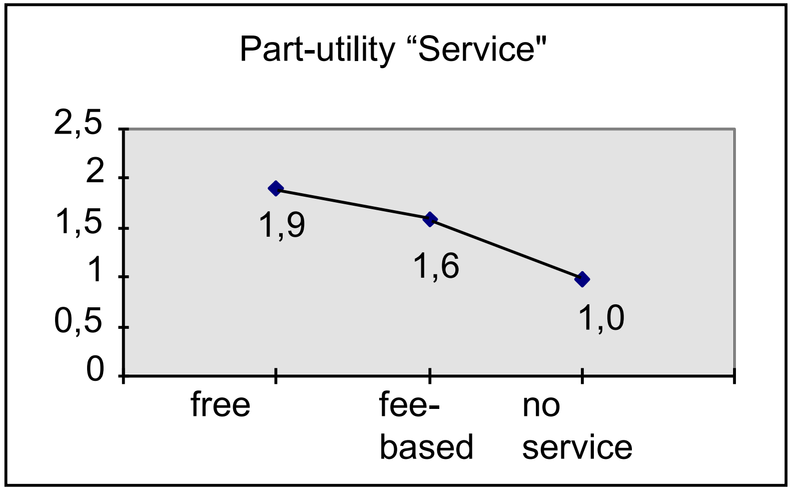
w_j : Importance of attribute **j**
$\hat{w}_j$: Importance of attribute **j** in %

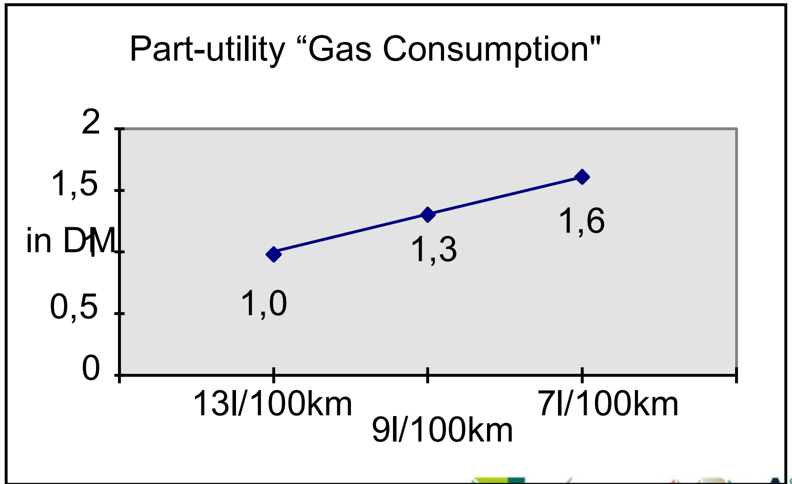
Analysis of Dependence: Conjoint Analysis (16)

- Example: Mid-Range Car
- Part-utility relationship (evaluation on an individual basis):









Analysis of Dependence: Conjoint Analysis (17)

How consistently can the estimated utility values predict the preferences of respondents?

→ Presentation of “Hold Out” concepts (example see next slide)

- Assessment of the disposition to buy regarding realistic product concepts
- “Hold Out” concepts reflect a real product range
- Example: Constant sum scale from 0% - 100%

Analysis of Dependence:
Conjoint Analysis (18)

Example: Mid-Range Car

Step 5: Assessing the overall fit

Example: Assessing the overall fit in case of an Adaptive Conjoint Analysis

0% 100%

Imagine that the market has the following four mid-range cars. Please distribute 100 points according to your preference for the cars among those four cars. You may grant 100 points to one car, divide the points among the four cars, or also assign 0 points to one or more of the cars.
Please enter numbers between 0 and 100 in the appropriate boxes. The fictitious mid-range cars are all the same in all other features not shown here (e.g., product quality).

70KW	60KW	80KW	70KW
7l/100km	7l/100km	9l/100km	9l/100km
Free service within the first two years after purchase	No service	Fee-based service	Fee-based service
Price €29,000	**Price €23,000**	**Price €29,000**	**Price €26,000**

Next question

Analysis of Dependence: Conjoint Analysis (19)

Step 6: Interpreting the results

- Rescaling the individual estimated part-utilities to a common scale
  - Allows easier comparison across and within respondents
  - E.g. conversion of part-utilities into a cash equivalent
- Choice simulator: Simulate the market and derive market share estimates concerning a new product concept
  - Example on the following slides

Conjoint Analysis:
Example mid-range car – Summary

What was the initial problem?	• We wanted to find out how to configure a mid-range car with regard to four factors optimally in order to meet our customers' preferences in the best way possible
What did we do?	• We specified the determinant factors price, gas consumption, power engine and service and conducted a full-profile conjoint analysis • The data were collected with different question designs • We estimated the part-utilities: the factor price is most important • By the help of a constant sum scale we assessed the overall fit of the analysis, i.e. in how far the analysis predicts the customers preferences
What was the outcome?	• The results can predict buying decisions and market shares.