# [WI001175] Consumer Behaviour Research Methods

# Legends

Definitions

## Fundamentals

### Variable types

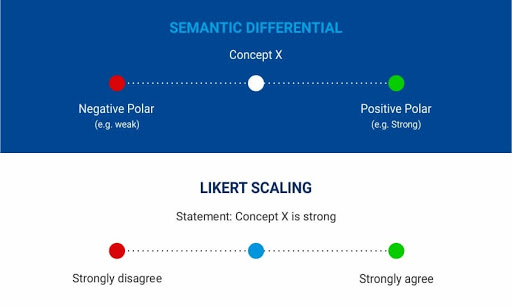
|  |  |  |  |
| --- | --- | --- | --- |
| Variable types for numerical data | | | |
| Qualitative question | Quantitative question | | |
| Qualitative research - obtaining insights into the situation about problems where you don’t have a lot of knowledge.  (Exploratory research) | Categorical var   * [Contingency tables](#_Contingency_table_1) | | Metric var |
|  | **Nominal**: has 2 or more categories (nationality, eye color, brand name) | **Ordinal** (education level, ranks, categories which can be ranked) | **Interval** (income – values, are equally spaced) |
|  | Frequency distribution, mode | Median, (Intermediate) rage | Mean, standard deviation |

* Structuring variables depending on the data type

### Scales

|  |  |  |  |
| --- | --- | --- | --- |
| Important Scales | | | |
| Likert scale | Semantic differential | Continuous rating scale | Multi-item scales |
| indicate the extent to which they agree or disagree with each of series of statements about the stimulus object (attitude, beliefs, behaviour) | endpoints are associated with bipolar labels that have semantic meaning | placing a mark at the appropriate position on a line that runs from one extreme of the criterion. no restriction to select from marks, identified by the researcher | used to measure constructs, which are hard to observe directly. For example, love to product can be measured with 3 items by constructing a scale “warm”, “full of love”, “full of passion”.  You need stat techniques like [factor analysis](#_L4_Factor_analysis). |

*Examples:*

Diagram

Description automatically generated

**Additional notes on scales**

1. # of scale categories (5 sufficient for the most (Brace), 5-7 points commonly used, up to 9-10 for scaling the task)
2. Balance the scales: equal positive and negative, Unbalanced scales may be appropriate when the distribution of responses is likely to be skewed.
3. Even or odd points on a scale
4. Forced versus nonforced scales
5. Verbal description for each category?!
6. Strength of anchors (very/somewhat/…)

### Questionnaire’s basics

* + You need to design a questionnaire which can motivate your respondents
  + Use conditional probabilities
  + Third variable problem. For example, people with higher income could drink more red wine and be healthier

### Improve questionnaires

* Make a pretesting on a small sample (15-30) from the same population
  + Multiple waves: first - personal interview, second – phone / mail
  + Commonly used procedures in pretesting: protocol analysis, debriefing
* Determine the order of questions
  + Make questions sequential
  + Use filter questions to ask before the questions about the topics
    - Definition. **Filter questions** - an initial question in a questionnaire that screens potential respondents to ensure they meet the requirements of the sample.
  + Sometimes screening questions are useful
    - Definition. **Screening questionnaires** to identify illegible, incomplete, inconsistent, or ambiguous responses.
  + Add “thank you” in the end
* Include consent form
* Always think about constraints
* Always think if the question is necessary
* Combine and divide questions if needed
* Don’t ask questions which responded can not remember
* Overcoming unwillingness to answer
  + Place sensitive topics at the end of the questionnaire
  + Preface the question with a statement that the behaviour of interest is common.
  + Ask the question using the third-person technique: phrase the question as if it referred to other people.
  + Hide the question in a group of other questions which respondents are willing to answer. The entire list of questions can then be asked quickly.
  + Provide response categories rather than asking for specific figures
* Choose the question wording
  + Use ordinary, unambiguous words
* Avoid leading questions, implicit alternatives, implicit assumptions, generalizations, estimates

## Marketing research

### Marketing research types

|  |  |
| --- | --- |
| Marketing research | |
| Problem-Identification Research | Problem-Solving Research |
| * Market Potential Research * Market Share Research * Image Research * Market Characteristics Research * Sales Analysis Research * Forecasting Research * Business Trends Research | * Segmentation Research * Product Research * Pricing Research * Promotion Research * Distribution Research |

### MR examples

|  |  |  |  |
| --- | --- | --- | --- |
| Problem-solving examples | | | |
| Segmentation research | Product research | Promotional research | Pricing research |
| * Determine the basis of segmentation * Establish market potential and responsiveness for various segments * Select target markets * Create lifestyle profiles: demography, media, and product image characteristics | * Test concept * Determine optimal product design * Package tests * Product modification * Brand positioning and repositioning * Test marketing * Control score tests | * Optimal promotional budget * Sales promotion relationship * Optimal promotional mix * Copy decisions * Media decisions * Creative advertising testing * Evaluation of advertising effectiveness * Claim substantiation | * Pricing policies * Importance of price in brand selection * Product line pricing * Price elasticity of demand * Initiating and responding to price changes |

Definition. **Marketing research:** is the systematic and objective identification, collection, analysis, dissemination, and use of information for the purpose of improving decision making related to the identification and solution of problems and opportunities in marketing. We want to know what consumers do, think, and feel.

### Limitations of MR

* Figure out what consumers want before they want it

### Marketing research process (Malhotra)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Marketing research process | | | | |
| Problem definition | Research design | Fieldwork and data collection | Analysis | Report and presentation |
| **Management** decision problem: The problem the decision makers face   * E.g. Should a new product be introduced? Should we advertise more?   **Marketing** research problem: The problem that researchers face.   * What information is needed? How it can be obtained?  1. [Defining a problem](#_How_to_define) 2. [Context of the problem](#_Context_of_the) | Research design - specifies details of implementing the developed approach. It defines:   * Necessary information * Type of design * Measurement * Questionnaire * Sample * Data analysis  1. [Types of research design](#_Types_of_research) | 1. [Data collection: primary and secondary data](#_Data_collection:_Primary) 2. [Cross-section, time series, panel data](#_Cross-section,_time_series,) |  |  |

### Defining the problem

* + Discuss with decision markets
  + Interview with experts
  + [Secondary data analysis](#_Data_collection:_Primary)
  + [Qualitative](#_Variable_types) data analysis on a small sample

### Context of the problem

* + Past information - Market, sales, demographics
  + Resource constraints - Cost and timing
  + Objectives - Decision maker‘s objective
  + Buyer behavior - Demographics, consumption habit, price sensitivity
  + Legal constraints - Law and government regulation
  + Economic environment - Income and purchasing power
  + Technological constrains - Marketing and production technology

## Theory

* “There is nothing so practical as a good theory.” “Experience alone does not create knowledge.” (Kurt Lewin)

### Good theory:

* + - Describes
    - Explains
    - Predicts
  + Theory allows for (some) prediction: given the circumstances, if I do (A), then (B) should (most likely) happen
  + In doing so, having theory allows for:
    - Abstraction: problem as specific representation of general class of problems
    - **Generalization**: reducing complexity to enable decision making
      * Having a good theory allows for better decision-making
    - Intervention: change things (but: objective vs. subjective views!)
    - (Causal) Learning: theory worked or did not work (à find out why!)
    - Power: predicting eclipses in the Middle Ages, leadership, motivation…
    - Survival: from animal tamers to astronauts
  + Components of a theory (Bacharach, 1989; AMR)

### What makes theory good theory?

* **Falsifiability** 
  + Accuracy: which theory is better at explaining/predicting?
    - Also: generalizability, i.e., “predicts more”
  + Parsimony: if they explain/predict equally well, which theory needs to make fewer assumptions?
    - Similar: fewer “auxiliary hypotheses”
    - Occam’s razor
* **Evidence of a causal relationship:** 
  + 1. Association: show connection between A and B, free of bias (statistical significance)
  + 2. Temporal precedence: A takes place before B (logic, theory, research design)
  + 3. Non-spuriousness: alternative explanations eliminated (using control variables)

Diagram

Description automatically generated

### What is theory?

* + Causal: it explains why and/or how
  + Aggregate of propositions
    - A proposition is one causal statement linking two constructs
    - A construct is usually a not directly observable - For example, an illness is not directly observable, the symptoms are - Key constructs: intelligence, motivation, performance, … - Vs. objects that “exist” like rocks, scissors, or pencils Testable through hypotheses
    - A hypothesis is one testable statement derived from the theory linking two measurable variables - For example: if intelligence is said to impact exam performance, we can test that by measuring whether IQ impacts grades - Raises first big issue: operationalization Based on assumptions (you have to start somewhere…)
    - But: how can you compare two theories that rest on different assumptions?
    - Think: ethics, human rights, religion, …

### The wheel of science (Wallance, 1971)

Diagram

Description automatically generated

### Theory generation via the scientific method

* The process relies on **empiricism**: things need to be observable
  + Only what is empirically observed can become the basis of theory
  + Difference in “what is observable:” science vs. philosophy (next slide)
* Observable data need to be collected **objectively**
  + However, objectivity is impossible!
  + Replace by: **replicable/reproducible** and testable intersubjectively
  + Means that research methods have to be fully disclosed and described
* **Control**: data needs to be collected without bias
  + Correct selection of people and method
  + Correct application of selected method

### Limits of the hypothetico-deductive method

* **Deduction** – process of reasoning from general proposition/theory to specific implications
  + A hypothesis can then be tested to draw inferences on the larger theory
* **Induction** – abstracting a proposition from data − Empirical generalization
* Theory usually moves forward by **combining both approaches**
  + Philosophical issue: induction cannot be proven to be correct
  + Just hope that the sun will rise tomorrow, again…
* Remember for “choice of research design!”

### The value chain of empirical research

|  |  |
| --- | --- |
| State of current research | * [Definition of the research field](file:///G:\My%20Drive\Masha_Studies\TUM-Master%20in%20Management%20and%20Technology\03_SoSe_2022\Sustainable%20Entrepreneurship%20-%20Theoretical%20Foundations\Lectures%20summary_Strategic%20Entrepreneurship.docx#_heading=h.gjdgxs) * Definition of the research question |
| Research design | * Selection of method of data collection * Operationalization − (Developing a measure) * Inspecting the criteria of goodness |
| Data collection | * Sampling * Pretesting/pilot testing * Data collection |
| Data analysis | * Data preparation (coding, data entry, data “cleaning,”…) * Descriptive statistics * Inferential statistics |
| Publication | * Interpretation of results * Writing a research paper * Submitting the paper to a scientific journal * Revision… revision… revision… * Publication |

### Layout of academic paper / introduction

Graphical user interface, text, application

Description automatically generated

## Research as conversation

### Identifying research field

* + Who is talking?
  + Where are they?
  + What are they talking about?
  + What have they spoken about in the past? What not?
  + What would they be interested in?
  + What do you have to say?
  + How will saying what you want to say change the conversation (can be both interpretation of past and present or future direction)?

Answer these questions to contribute that is (more on this later) − Relevant: it relates to an ongoing academic conversation − Novel: it says something that has not yet been said − Interesting: the thing that has not been said is actually interesting

### Contribution should be:

* New
* Interesting
* Relevant

### Finding interesting questions

* Start from practice: try to explain a phenomenon that boggles you
* Similar: “Grand challenges”
* Start from the literature (note that the following items may be related, co-appear)
  + Conflicting results: previous studies on same topic do not converge (very smart)
  + Boundary conditions: under what conditions does a theory hold? (smart)
  + Increase specificity: theory has not been applied here (boring?)
  + Study new phenomena: has not been explored before (really new?)
  + Note: the latter can also be purely descriptive (i.e., documentation without explanation) depending on the audience and the novelty of the topic
  + Suggestions for future research (boring?)

Text

Description automatically generated

### Designing actual research questions

|  |  |
| --- | --- |
| Do/Is… -> Yes/no | Often extremely boring, but exceptions possible:   * Conflict in the literature * Novel effects * Isolated variables (experiment?) |
| What…/how… | * Usually studies a process * Usually means qualitative work |
| Why… | * Causal reasons: will often need additional specification! |
| To what extent… | * Explains how differences in inputs drive different outcomes * Usually means quantitative work |

### Good research questions come with precision

Either the question itself, or a derivative of it, must be precise

* What are you studying?
  + Processes, decisions, behavior, …
  + Inputs, outputs, transformation, …
* What are you not studying?
  + Key assumptions
* Who are you studying?
  + Industry/country/economy/culture >> firm >> individuals >> brain
* How are you studying?

… The question is usually formulated at the level of constructs

* Discover relationships and/or explain the why: qualitative work
* Corroborate relationships, measure effects: quantitative work

### From research question to the theory section

Start with research question

* Choose which conversation(s) you participate in
* Quantitative & experimental work
  + From this foundation, build an argument through logic
    - Work toward a not-yet-tested causal relationship between two constructs –
    - Explain: why should these to concepts be linked?
    - Note: huge differences in the “degree of storytelling” between disciplines
* State the expected, novel relationship as formal hypothesis
  + - Quantitative management papers usually contain about 3-5 hypotheses
* Qualitative work
  + Summarize and explain the state of the art
  + Make clear how there is still a need to look deep into the process of what is going on

### Basic paper structure

* Introduction
* Theory / Literature review
* Methodology / Research design / Data and methods
* Findings / Results
* Discussion / Conclusions
* References
* Appendices

## What is good research

Publication process

Diagram

Description automatically generated

### Some facts around the publication process

At good journals (note the crucial term here: “good.” Think hard about how you would define it!)

* Rejection rate in 1st round >> 80%
* Rejection rate in 2nd round ~ 50%

Good and bad arguments against the publication process

* “Reviewers never agree”
  + Wrong: editor specifically selects to address different aspects of the paper
* “Bad papers still published”
  + Of course: minimizing, but not eliminating error
* “Politicizing of the process”, “Friends with editors”
  + True, but: (1) word gets around, (2) editors select friends, too
* “Wasted efforts”, “wasted resources”
  + Wrong, as per “progress of science:” (1) results cannot be predicted a priori, (2) quality and novelty thresholds must be met, (3) non-finding is a finding
* “I should have been published”
  + More often than not, it is not the reviewers and editors who are at fault…

### Research ethics

* Reliability
  + Simply: If I measure this again (even with different method, questionnaire, …), I will get the same results
  + Usually means I need at least “two of something” (studies, methods, coders, items, …)
* Validity
  + Simply: I am doing what I say I am doing
  + Free of bias (recall Week 1: “control”) − Many different forms of validity (next slide)
* Reliability and validity need to be shown
  + Research methods: 2nd half of our course

### Some more validity considerations

* “Internal validity is the basic minimum without which any experiment [or other empirical study] is uninterpretable: Did in fact the experimental treatments make a difference in this specific instance?” – similar: is any causal conclusion warranted?
* “External validity asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?”
* Face validity: does the item measure the construct? Does this make sense?
* Construct validity: item, test, etc. measures what it claims to be measuring
  + Convergent validity: correlates with similar scales/constructs?
  + Discriminant validity: is distinct from other scales/constructs?
  + Content validity: is the scale/construct really fully captured?

## Variable definitions

Definition. **Dependent variables** are the variables that measure the effect of the independent variables on the test units. These variables may include sales, profits, and market shares.

Definition. **Extraneous variables** are all variables other than the independent variables that affect the response of the test units. These variables can confound the DV measures in a way that weakens or invalidates the results of the experiment.

### Types of research design (3)

|  |  |  |
| --- | --- | --- |
| Types of research design | | |
| Exploratory | Descriptive | Explanatory (causal) |
| exploring new phenomena, ideas, insights, might involve literature research or conducting qualitative interviews or focus group interviews.  [Non-probability sampling](#_Sampling_approaches:) could be good. | Providing accurate description of a given phenomena | Explanation of the nature of certain relationships |
| Key var are not defined | Key var are defined | Key var and key relationships are defined |
| Flexible, versatile | Planned and structured design | Manipulation of one or more independent variables |
| * Expert surveys * Pilot surveys * Secondary data * Qualitative analysis * Quantitative analysis | * Secondary data * Quantitative analysis * Surveys * Panels * Observation and etc | * Experiments |
| Can [use non-probability sampling](#_Sampling_approaches) to generate ideas, insights, or hypotheses (e.g., low cost, convenient, appropriate) |  |  |

### Research design

* Longitudinal
  + Prospective: individuals are followed over time
  + Retrospective: individuals are sampled, and information is collected about their past
* Dynamic
* Deductive

### Three archetypes of methodological field

|  |  |  |  |
| --- | --- | --- | --- |
| Three Archetypes of Methodological Fit in Field Research | | | |
| State of Prior Theory and Research | Nascent | Intermediate | Mature |
| Research question | Open-ended inquiry about a phenomenon of interest | Proposed relationships between new and established constructs | Focused questions and/or hypotheses relating existing constructs |
| Type of data collected | Qualitative, initially  open-ended data that need to be interpreted for meaning | Hybrid (both qualitative and quantitative) | Quantitative data, focused measures where extent or amount is meaningful |
| Illustrative methods for collecting data | Interviews; observations; obtaining documents or other material from field sites relevant to the phenomena of interest | Interviews; observations; surveys; obtaining material from field sites relevant to the phenomena of interest | Surveys; interviews or observations designed to be systematically coded and quantified; obtaining data from field sites that measure the extent or amount of salient constructs |
| Constructs and measures | Typically, new constructs, few formal measures | Typically, one or more new constructs and/or new measures | Typically relying heavily on existing constructs and measures |
| Goal of data analyses | Pattern identification | Preliminary or  exploratory testing of new propositions and/or new constructs | Formal [hypothesis testing](#_Hypothesis_testing_steps) |
| Data analysis methods | Thematic content  analysis coding for evidence of constructs | Content analysis, exploratory statistics, and preliminary tests | Statistical inference, standard statistical analyses |
| Theoretical contribution | A suggestive theory, often an invitation for further work on the issue or set of issues opened up by the study | A provisional theory, often one that integrates previously separate bodies of work | A supported theory that may add specificity, new mechanisms, or new boundaries to existing theories |

### Data collection: Primary and Secondary Data

|  |  |
| --- | --- |
| Primary data | Secondary data |
| Primary data can be [biased](#_Biases) | It can go out of control of the researcher |

### Cross-section, time series, panel data

|  |  |  |
| --- | --- | --- |
| Cross-section | Time series | Panel data |
| Many observation units at one point in time | One observation unit over time | Several observation units over time |

### Sampling

* **Target population**. Most of the time, the objective of a research project is to obtain information about the characteristics or parameters of a population. Since you cannot examine the entire population, you need a sample.
  + Definition. A **sample** is a “smaller” collection of units from a population used to determine truths about that population. Sometimes the sample does not partially represent the target population.
  + Definition. **Sampling frame** - a representation of the elements of the target population. It consists of a list or set of directions for identifying the target population.

Diagram

Description automatically generated

* Definition. **Census**. Complete enumeration of all the elements of the population. If you want to know exactly how many people live in Germany, then you have to count each citizen living in Germany

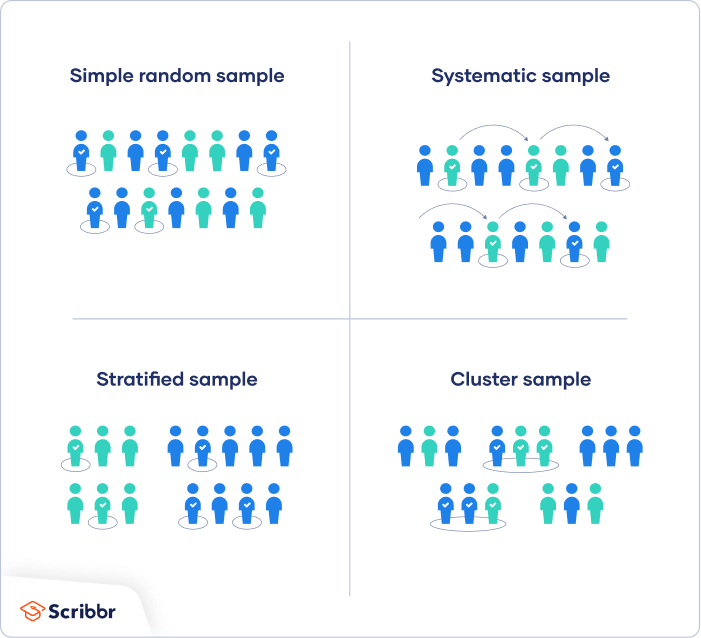
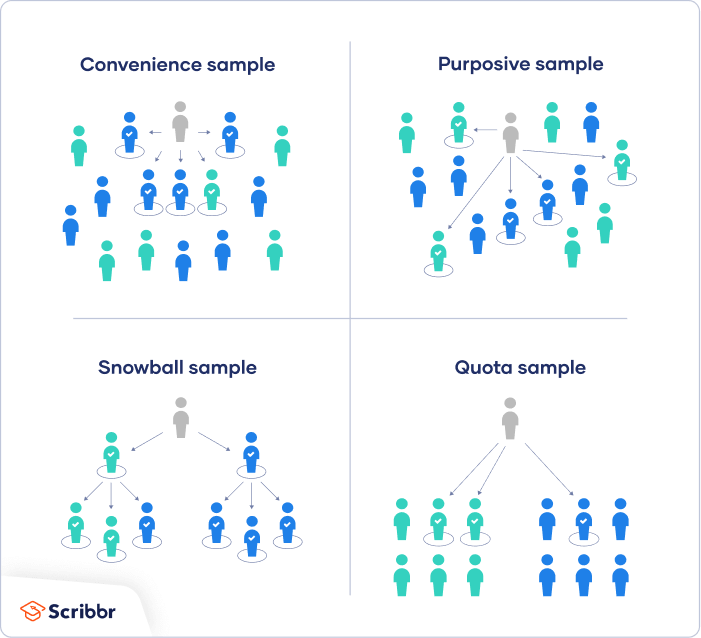
### Sampling approaches

P = probability

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sampling approaches** | | | | | | |
| **Non-probability** sampling | | | **Probability** sampling | | | |
| Population elements are selected on the basis of their availability or based on the researcher's own judgement of their representativeness → not random, there is a subjectivity. It is not recommended for descriptive or causal research but can be used for exploratory research to generate ideas, insights, or hypothesis. | | | Sampling units are selected by chance and randomness is built into the sampling design. | | | |
| **Convenience** sampling | **Judgmental** sampling | **Snowball** sampling | **Simple random** sampling | **Stratified** sampling | **Cluster** sampling | **Systematic** sampling |
| Respondents are selected because they happen to be in the right place at the right time | A form of convenience sampling in which population elements are purposely selected based on researcher’s judgement | Initial group is selected (at random) and subsequent respondents are selected based on the referrals or information provided by the initial respondents | Each element has a known and equal **p** of selection, and is drawn at random independently of all other elements   * time consuming, costly data collection * can be difficult to construct a list that will permit a random sample to be drawn * samples drawn will represent the population well on avg, a given random sampling may grossly misrepresent the target population | * population is partitioned into subpopulations, or strata (e.g. age, income, nationality). * after, elements are selected from each stratum by a random procedure (usually simple random sampling) | * population is divided into mutually exclusive and collectively exhaustive clusters (e.g. geographic areas such as counties, or blocks). * after, a random sample of clusters is selected based on a **p**-sampling. * for each selected cluster, either all the elements are included in the sample, or a sample of elements is drawn at random. | * first element is selected at random * every i-th element is picked in the succession * the sample interval “I” is determined by the population size and sample size. * If the order of elements is related to characteristics of interest, systematic sampling improves representativeness, otherwise it will yield results similar to simple random sampling |

### Stratified sampling vs cluster sampling

|  |  |
| --- | --- |
| Stratified | Cluster |
| * all the subpopulations (strata) are selected for further sampling (e.g., select 2 members from each group) | * only a sample of subpopulations (clusters) is chosen |
| * elements within a stratum should be as homogeneous as possible, but the elements in different stratum should be as heterogeneous as possible | * elements within a cluster should be as heterogeneous as possible, but clusters should be as homogeneous as possible |
| When to use:   * heterogenous, several different groups (strata) |  |

### Sampling errors

|  |  |
| --- | --- |
| Definition. **Total error** is the variation between the true mean value in the population of the variable of interest and the observed mean value obtained in the marketing research project. | |
| **Random** Sampling Error | **Nonsampling** Error |
| Random sampling error occurs because our sample is an imperfect representation of the population, we are interested in | Can be attributed to sources other than sampling. They result from a variety of reasons, including errors in problem definition, approach, scales, questionnaire design, interviewing methods, and data preparation and analysis. |
| Solution =   * Standard error computation * Confidence interval |  |

## Statistical analysis

### Types of statistical data analysis

|  |  |
| --- | --- |
| Types of statistical data analysis | |
| [Descriptive](#_Descriptive_statistics) | [Inferential](#_Inferential_statistics) |
| Method to quantitatively summarize and describe the main features of a data set | Methods to estimate the features of a population based on the analysis of a sample   * Chi-Square |

## Descriptive statistics

### Statistical techniques

|  |  |
| --- | --- |
| Statistical techniques | |
| **Univariate** statistics analysis: specific variable in data set on its own | **Bivariate** statistics analysis: two variables analysis |
| * [Summary statistics](#_Summary_statistics)   + [Measures of centricity](#_Measures_of_centricity)   + [Measures of dispersion](#_Measures_of_dispersion)   + [Shape of distribution](#_Shape_of_distribution) | * Categorical var + categorical var => [Contingency tables](#_Contingency_table_1) * Quantitative var + quantitative var => [Conditional probability](#_Conditional_probability) * [Associative Analysis](#_Associative_Analysis) * Chi-Squire test: show dependency between two nominal variables |

## Summary statistics

### Measures of centricity

|  |  |
| --- | --- |
| Measures of centricity | * + - * Arithmetic mean (avg) * When is the arithmetic mean problematic as a measure of central tendency?   ...can be influenced by the extreme values of the dataset (outliers), since it is based on all the information of the dataset     * + - * Median (central value) * The median is not influenced by outliers of a dataset.   Rule 1: If the dataset has an odd number of observations, the median is the central value  Rule 2: If the dataset has an even number of observations, the median is the average of the two central observations       * + - * Mode (most frequent value)         + The mode is not influenced by outliers       * Quartiles (measure of non-central tendency) * Definition. **Quartiles** are descriptive measures that divide the ordered sequence of data in four parts. * + RULE 1: If Q is an integer number, the quartile is the value of the correspondent observation   + RULE 2: If Q is exactly between two integer numbers (e.g 3.5), the quartile is the average of the values of the two correspondent observations   + RULE 3: If Q is not an integer and not even exactly between two integer numbers (e.g., 3.2 or 3.8), you approximate it to the nearest integer and then follow rule 1     - Definition **Interquartile mean** is the mean between the first and the third quartile of the dataset.      * + - * The interquartile mean is not influenced by the outliers.       * Interquartile mean, median, mode, not influenced by outliers, are robust measures |

### Measures of dispersion (quantity of dispersion of the data)

|  |  |
| --- | --- |
| Measures of dispersion (quantity of dispersion of the data) | * + - * Range   Text  Description automatically generated with low confidence   * A limit of the range - does not consider how the data distributed between the biggest and the smallest value. * This - in case of outliers it is not a proper measure of variability   + Interquartile range   Chart  Description automatically generated with medium confidence   * + - This variability measure summarizes the dispersion of the 50% of data that occupy the central positions and is therefore not influenced by outliers       * Variance   Text, letter  Description automatically generated   * + - * Standard deviation * The standard deviation is the square root of the variance.   Graphical user interface  Description automatically generated with low confidence   * The variance and the sd measure: how the biggest observations fluctuate above the mean and how the smallest observations distribute below the mean. * The sd helps us to define if and how much the data is concentrated or scattered around their mean. * The more the data is dispersed (concentrated) -> the bigger (smaller) is the range, the interquartile range, variance, and standard deviation.   + - * Coefficient of variance * to compare results from two different surveys or tests that have different measures or values |

### Shape of distribution

|  |  |
| --- | --- |
| Shape of distribution | * + - * Symmetric and asymmetric distribution     - Mean = Median → symmetric (b)     - Mean < Median → negatively or left-skewed (a)     - Mean > Median → positively or right-skewed (c)   Mean - BIOLOGY FOR LIFE   * + - * Definition. **Boxplot** is a possible graphical representation of some summary statistics is the boxplot. It describes in a graphical and solid way the distribution of a function.   Chart, box and whisker chart  Description automatically generated   * Skewness (y): it indicates the degree of symmetry or asymmetry of a distribution   + - * + y=0 → symmetric distribution         + y<0 → left-skewness (mean < median)         + y>0 → right-skewness (mean > median) * Kurtosis index (ß): Kurtosis is the degree of peakiness of a distribution   + - * + ß=0 → normal distribution         + ß<0 → flat distribution         + ß>0 → peak distribution     Relation between the skewness and boxplot  Shape, polygon  Description automatically generated |

### Plots categorized by purpose

Graphical user interface, application

Description automatically generated

### Contingency table

|  |  |
| --- | --- |
| Contingency tables are double-entry tables. Help to understand distribution of categorical var.  Contingency tables, specific type of frequency distribution table, where two variables are shown simultaneously. Contingency tables (also called crosstabs or two-way tables) are used in statistics to summarize the relationship between several categorical variables.  The values in the table are the absolute joint frequencies, and their sum is equal to the total of observed cases.  Relative joint frequencies = absolute joint frequency / total  JASP -> frequencies, NO SCALE VARS | Diagram  Description automatically generated |

### Conditional probability

Definition. **Subordinate distributions**: the frequency to observe the phenomenon x given the phenomenon y and vice versa.

Text

Description automatically generated

Definition. **Statistical independence**: Two events are statistically independent if the occurrence of one does not affect the probability of the other.

* If stat independent => the relative joint frequency is equal to the product of the correspondent marginal distributions.



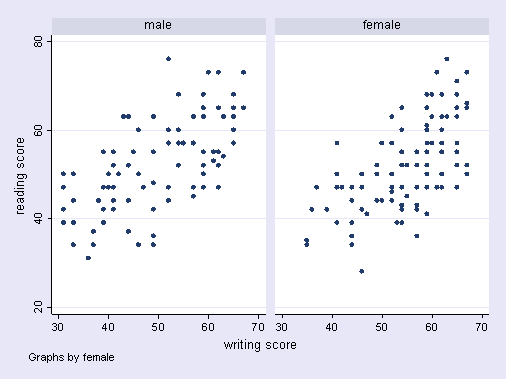
* Testing the association between rows and columns for the contingency table:
  + Definition. **Chi Square Test** = **X2** test for the statistical independence
    - The Chi-Square can be computed by comparing the frequency that we would have in the case of perfect statistical independence between the two studied variables and the observed frequency.



* + - Types of Chi-Square test:
      * Goodness of fit: H0, H1 (X is independent from Y, X is dependent on Y)
      * Test of homogeneity

### Associative Analysis

* Two metric continuous (interval) variable: two-way scatter plot



* Covariance – Cov (x,y) - Determining the **direction** of a linear relationship between two variables
  + - Cov (x,y) **>** 0 -> If x and y tend to increase or decrease together, the coefficient is positive
    - Cov (x,y) **<** 0 -> If one variable tends to increase as the other decreases or vice versa, the coefficient is negative
    - Cov (x,y) **=** 0 -> discordance between x and y

A picture containing text, watch, clock

Description automatically generated

* + - It is an absolute index: it indicates the existence and direction of the relation between two variables, but **not the strength of their relation**
* Linear correlation coefficient - R (x,y) – Determining the **strength** of a linear relationships
  + It is a relative index that solves the problem of the covariance, and it can have values between -1 and 1:
    - – r(x,y) = **1** -> y is a linear function of x, and the observation are on a straight line with positive slope
    - – r(x,y) = **-1** -> y is a linear function of x, and the observation are on a straight line with negative slope
    - – r(x,y) = **0** -> the is no linear relation between the variables x and y

Text

Description automatically generated



* Pearson correlation coefficient
  + - Both variables metric (interval-continuous/ratio)
    - Linear relationship
* Spearman rank correlation coefficient
  + - One or both variables are ordinal
    - Non-monotonic relationships

For both correlation coefficients: -1 ≤ r ≤ 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pearson = +1, Spearman = +1 | Pearson = +0.851, Spearman = +1 (This is a monotonically increasing relationship, thus Spearman is exactly 1) | Pearson = −0.093, Spearman = −0.093 | Pearson = −1, Spearman = −1 | Pearson = −0.799, Spearman = −1 (This is a monotonically decreasing relationship, thus Spearman is exactly 1) |

JASP: Pearson correlation coefficient

Graphical user interface

Description automatically generated

JASP: Spearman’s correlation coefficient

Graphical user interface, application

Description automatically generated

## Inferential statistics

### Hypothesis testing steps

|  |  |
| --- | --- |
| 1. Formulate the hypothesis H0 and H1 | * The **H0** variable is **≥ or ≤** * The **H1** variable **< or >**   + The alternative hypothesis is the opposite of the null hypothesis |
| 1. Choose the a-level (related to confidence interval) | * Probability of falsely rejecting H0 (Type I error), typically 0.05 or 0.01 |
| 1. Find test to test H0 | * One-tailed / two-tailed – depending on the hypothesis |
| 1. Calculate test statistics, find the p-value, assuming H0 is true | * Confidence interval   + The larger the sample size -> the smaller is the confidence interval   + Lower confidence level = smaller interval   + More variation = larger interval |
| 1. Use the decision rule p < 0.05 to determine if you reject H0 | * If **p ≤ 0.05 (a-level)**, thenH0 is rejected as the parameter is significantlydifferent from a specific value (one sample) or across groups (two or more samples). * If **0.05 < p ≤ 0.10**, then H0 is rejected but marginally, the parameter is marginally significantly different from a specific value or across groups. * If **p > 0.10**, then H0 is **not** rejected, the parameter is not statistically different from a specific value or across groups. |

Table

Description automatically generated

**Type I error** - You reject the null hypothesis when the null hypothesis is actually true.

**Type II error** - You fail to reject the null hypothesis when the alternative hypothesis is true.

**Significant != substantive**

* Significant depends on magnitude, sample size

### Choosing statistical test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measurement level, dependent variable | One sample | Independent samples | | Related samples | |
| K = 1 | K=2 | K>2 | K=2 | K>2 |
| **Interval / ratio** |  |  |  |  |  |
|  | t-test  (mean) | t-test  (means) | ANOVA (means) | t-test (paired) | ANOVA (repeated measurement) |
| SPSS: Compare means | t-test  (one sample) | t-test  (independent sample) | ANOVA  (one-way) | t-test (paired samples) | General linear model > repeated measures |
| **Ordinal** |  |  |  |  |  |
|  | Kormogorov-Smirnov | Mann-Whitney u-test | Kruskal-Wallis test | Wilcoxon test  Sign test | Friedman test |
| SPSS: nonparametric tests | 1-Sample K-S\*  (by hand) | 2 Independent samples | K Independent samples | 2 Related samples | K Related samples |
| **Nominal** |  |  |  |  |  |
| Multiple choice  SPSS | Chi-square  (one sample)  Nonparametric tests, Chi-square or by hand | Chi-square (cross tabs)  Descriptive statistics,  Cross tabs | Chi-square (cross tabs)  Descriptive statistics,  Cross tabs | --- | --- |
| Dichotomy (two things that are or are represented as being opposed or entirely different) | z-test  (proportion)  by hand | z-test (proportion)  by hand | Chi-square (cross tabs)  Descriptive statistics  Cross tabs |  |  |

### T-test

|  |  |  |
| --- | --- | --- |
| Single sample | Paired samples | Independent samples |
| tests whether a sample mean is significantly different from a pre-existing value | tests the relationship between 2 linked samples, e.g., means obtained in 2 conditions by a single group of participants | tests the relationship between 2 independent populations |

|  |  |  |
| --- | --- | --- |
| How many samples does the hypothesis involve | | |
| One sample | Independent samples | Related samples |
| Comparing the parameter in a given group  Ex.: if mean liking = 4 | Comparing one parameter across two or more separate groups  Ex.: if mean beer liking is different between men and women | Compare the responses of the same individual amongst each other  Ex.: if women like branded beer as much as they like unbranded beer |
| * Compare 1 var (column) to a specific value. E.g. Does the mean = 40? | * Compare 1 var (column) between different samples (group of respondents / lines) | * Compare 2 or more vars (columns) within respondents (lines) |

### Test examples

1. What is the dependent variable - DV?
2. What is its [measurement level](#_Variable_types)?
3. What and how many samples are there in the hypothesis?

|  |  |
| --- | --- |
| Problem | Test |
| The average household size in the population is the Netherlands is 2.8. The mean in the sample is 2.92. Is the sample representative of the population? | H0: mean in the population where the same sample came from = 2.28   1. DV: household size 2. DV measure: interval 3. Sample: one sample  * T-test (mean) |
| Compare proportion of a dichotomous var (yes/no) to a specific value in one sample (e.g. the choice of the brand in the sample equal to 50%) | H0: is the proportion of men in the sample equal to 50?   1. DV: proportion of men 2. DV measure: nominal (dichotomous) 3. Sample: one-sample  * Z-test (one sample) |
| Do women like branded beer as much as men? | H0: for each branded beer, women’s mean liking = men’s mean liking   1. DV = liking 2. DV measure = interval 3. Sample = **two** independent samples  * T-test (means) independent sample test, ANOVA (means), one-way ANOVA |
| Do consumers in different regions (big cities, south, west, and east) like each beer equally? | H0: liking of beer is equal across regions   1. DV = liking 2. DV measurement = interval 3. Sample = **more than two** **independent** samples  * ANOVA (means), One-way ANOVA |
| Are evaluations of beer A different when it is branded vs unbranded? | H0: liking of branded beer = liking of unbranded beer   1. DV = liking 2. DV measurement = interval 3. Sample = **two** **related** samples  * T-test (paired), paired sample t-test; ANOVA (repeated means) General linear model     ANOVA repeated means |
| Is the gender of the respondents equally distributed? | H0: proportion of males (females) is equal across regions   1. DV = proportion of males (females) 2. DV measure = nominal (dichotomous: male or female) 3. Sample = more than two independent samples  * Crosstabs (contingency tables), chi-square (cross tabs), descriptive stat cross tabs |
| H0: The mean liking of unbranded beer A is **=** to 6.5 | p-value of 0.007 < 0.05  We reject H0. We assume that H1 is true |
| H0: women's mean liking of branded beer A is **=** to men's liking of branded beer A  H0: women like branded beer A as much as men. | Test: Independent samples t-test  p-value of 0.631 > 0.05  The result is not significant. We do not reject H0. |
| H0: People like branded beer B equally across regions: | Test: ANOVA |
| The population in the Netherlands is 2.28. The mean in the sample is 2.92. Is the sample representative of the population?  H0: mean in the population where the sample came from = 2.28. | Test: One sample t-test    p-values < .05 -> H0 is rejected |

### P-value

* The larger the test statistic (absolute value), the smaller the p-value
* The smaller the p=value, the less likely we make type 1 error
* In SPSS p-value is denoted as “Sig.”

### Standardizing data

If you have standardized the data, it is easy to determine the probabilities of observing such a value or a smaller/greater one using a standard normal distribution table.

* Definition. The **normal distribution** is bell shaped and symmetrical. Its mean, median, and mode (measures of central tendency) are identical.

Chart, histogram

Description automatically generated

* Definition. The **z score** for a point is the number of standard errors a point is away from the mean or proportion.
  + z = (X – μ)/σ; x –observation, μ – mean, σ – standard deviation
  + Graphical user interface, text, application

    Description automatically generated with medium confidence

Table

Description automatically generated

## Factor analysis

Requirements to do factor analysis:

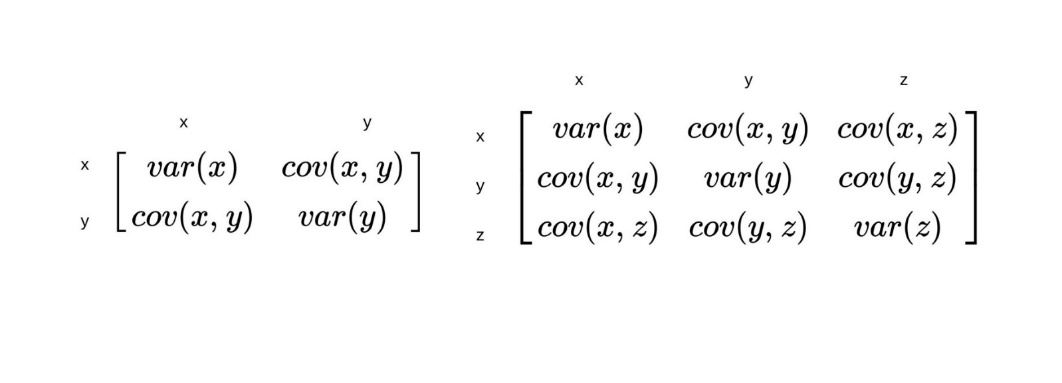
* Sample size: The reliability of factor analysis is also dependent on sample size
  + In general, over 300 cases is probably adequate (see Field, 2005).
* JASP will nearly always find a factor solution to a set of variables. However, the solution is unlikely to have any real meaning if the variables analysed are not “sensible”.

### Inter-correlation between variables

* Test question measure the same dimension, then we expect correlation
* Exclude variables, which do not correlate with any other (or very few), before running factor analysis
* Too high correlation – extreme multicollinearity
* Perfectly correlated variables – singularity

### Variance-covariance matrix

Definition. A **variance-covariance matrix** is a square matrix that contains the variances and covariances associated with several variables.



Correlation matrix

Graphical user interface, application, table, Excel

Description automatically generated

Definition. **Factor analysis** reduces the original number of attributes to a smaller number of factors, each containing a set of attributes that “hang together”. It allows us to identify dimensions that explain correlations between variables that are correlated to each other. FA can be don’t with different techniques:

* Principal component analysis (PCA);
  + To identify a reduced number of linear combinations of the original variables that explain most of the variance of the same variables.
  + Every linear combination is a function of all the original variables, but it is correlated in particular to some of them

Text

Description automatically generated with medium confidence

* It not regression, this is a different minimalization problem
  + PCA allows to compute factor **weights** to extract the max possible variance from the original variables and to continue factoring until there is no more meaningful variance left.
  + The higher the correlation of the original variables – the higher is the power of the factor analysis, the less info you loose by performing it.
* Canonical factor analysis
* Common factor analysis

### Types of factor analysis

* Explanatory factor analysis (EFA): to identify relationships among variables
* Confirmatory factor analysis (CFA): to test hypothesis that variables are associated with specific factors

### Factor analysis process

|  |  |
| --- | --- |
| 1. Formulate the problem |  |
| Number of factors to consider   * + Percentage of total variance explained by each component (70-80%) | Table  Description automatically generated |
| * + Eigenvalue > 1 (Keiser’s criterion)     - Hair et. Al (1998) recommends the following guidelines for practical significance: ±0.3 Minimal ±0.4 More Important ±0.5 Practically Significant     - Costello and Osborne (2005) - .50     - Guadagnoliand Velicer (1988) - .60 or greater. | Table  Description automatically generated  With these 3 factors you can explain 68% of the variance |
| * + Scree test   Definition. The scree plot is a graph that reports the percentage of variance explained by each factor.  Recommendable to stop to the number of components that corresponds on the graph with a significant reduction of the slope of the line. | Here it is 3-4 |
| * + A-priori determination   Say in the JASP how many factors you wanna look at   * + When the number of factors is determined by theory   + Interpretation |  |
| 1. Construct the correlation matrix 2. Determine the method of factor analysis 3. Determine the number of factors. | |
| 1. Rotate the factors & interpret the factors | |
| * + Rotated component matrix and factor loadings – always used to interpret the factors | Rotation methods:   * + Orthogonal   + Oblique   Orthogonal Rotation: **VARIMAX method** (different ones can be used)  Minimizes the # of items with high **loadings** (factors should be similar) on each factor -> improves interpretation of factors  Maximizes the squared factor loadings for each factor differentiates original items according to extracted factors  Factor loadings can be positive or negative (depends on item scaling) -> consider absolute values of factor loadings for interpretation.  Factor loadings = correlation between factor and the item.    Highlight the highest factor loadings per row, then look at highlighted items for each factor  Sometimes an item loads equally high on 2 factors   * Decide on content to which factor it belongs * Include it in the subsequent item analysis of both factors to see where it fits best statistically (one at a time!) * This may mean that there are correlated factors use oblique rotation (note: not part of this course, do not use in the assignment) |
| 1. Calculate the factor scores / Select the surrogate variables 2. Determine the model fit | |

* The factor analysis extract factors also called “latent variables” that concentrate the information of a high number of variables. The aim is - to identify an underlying structure of a set of observed variables. It studies the interrelations between variables.
* Once found the optimal solution, it is possible to consider the factors as new “macro-variables” to use instead of the original variables in further analyses, can also be an input for cluster analysis.

Main applications:

* Marketing
* Risk management

JASP PC1 (principal components), PC2 – there is no difference

Table

Description automatically generated

## Experiments

### Causality

Causality -> Solution to causality: Instrumental variables and advanced statistics, Experiments

* RQ: Does reducing package size (X) affect sales (Y)?
* When can we infer that X (large menus) causes Y (more eating)?
  + X and Y must be related (relationships)
  + First X, then Y, at the limit (time order)
  + Elimination of other possible causal factors (keep constant or controlled)

### Main elements of experiments

|  |  |
| --- | --- |
| Experimenter | * + If feasible, blind to hypotheses |
| Participants | * + Respondents, subjects |
| Independent variables | * + Manipulated variables (e.g., different package sizes for the dog food example) |
| Experimental conditions / cells / treatments |  |
| Experimental stimuli | * + everything held constant (i.e., controlled for) except the factors (i.e.   + Independent variables) |
| Treatment groups |  |
| Control group: | * + No treatment |
| Dependent variables | * + Variables of interest   E.g., the effect of consumer identity on sustainable behaviour assignment |
| Control variables | * + Non-manipulated variables |
| Random assignment to conditions | * + To control for extraneous variables |

### Types of validity (goals of experiment)

* Definition **Internal validity:** to draw valid conclusions about the effects of independent variables:
  + Did X really cause the effect on Y?
  + Threads to internal validity: [selection bias](#_Biases)
    - **Robustness**: can it be replicated (generalizability)
    - **Relevance**: does it matter– potential usefulness of results
* Definition “**External validity** asks the question of generalizability: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?”
  + Can your effect be replicated?
  + Threads to external validity: interactive testing effect (when you show more adds - people will be more knowledgeably of it); reactive testing and demand effect (demand artifacts - people might know about your hypothesis) (workers were more productive because they knew that they were observed).
* Definition. **Face validity**: does the item measure the construct? Does this make sense?
* Definition. **Construct validity**: item, test, etc. measures what it claims to be measuring

### Types of experiments

|  |  |
| --- | --- |
| Laboratory experiments (controlled) | * High internal validity, external validity may suffer * Popular in academics (usually students as participants) * Artificial environment * Case: Effect of Consumer Identity on Sustainable Behavior |
| Field experiments (‘less’ controlled) | * High external validity, internal validity may suffer * Natural environment: test market(s) * Online marketing: A/B Testing |

### Threats to internal validity

* History: events happening at the same time as treatment
* Maturating: changes in participants
* Testing: prior observation -> later observation, consistency or reactance
* Instrumentation: change in measuring instrument during experiment
* Mortality: loss of participants
* Selection bias: improper assignment of respondents to conditions (groups differ on DV before treatment)
* Statistical regression: regression to the mean over time

### Threats to external validity

* Interactive testing effect: prior measurement affects the response to independent variable
  + When people are asked to evaluate a brand, they will pay more attention to it afterwards (for example ad)
  + Measuring intention to buy a product, increases actual buying probability (cars - 30% increase, financial services)
* Reactive testing or Demand effects: abnormal behaviour of participants because they know they are in an experiment and interpret the purpose of the experiment
  + Example: Hawthorne-effect -> People change their behavior because they are aware that they are observed

Controlling threats

* + Randomisation: random assignment of participants to conditions
  + Matching: manual assignment of participants to conditions based on relevant characteristics
  + Statistical control: adjusting for extraneous effects in analysis (e.g., Multivariate Regression, ANCOVA; control for income, dog breed etc.)
  + Experimental design: design such that extraneous variables are controlled for (e.g., economic games, incentive to play the “same way”)

### Administering treatments

|  |  |
| --- | --- |
| Posttest: observation after treatment | Pretest-posttest : observation before and after treatment |
| * Watch, then rate their attitudes twrds brand * With control group: add a group that does not watch an ad | * Attitudes before and after exposure to the ad * With control group: add a group that does not watch an ad |

### Experimental design

* Full factorial designs:
  + Two or more factors (treatment, independent variables)
  + Control condition (no treatment)
  + Types

|  |  |  |
| --- | --- | --- |
| Between-subjects designs | Within-subjects designs | Mixed design |
| * Treatments on different groups of respondents * The most common * Infeasible if there are too many conditions (factors / factor levels)   Between subjects design | Treatments on same group of  Respondents  Between subjects design | Some between-subjects treatments, some  within-subjects treatments |
| * N-way ANOVA | * Repeated measures   + Paired-sample t-test   + ANOVA for repeated measures |  |

Diagram, application

Description automatically generated with medium confidence

* Random assignment to conditions, same # participants in all conditions
* Check successful manipulation:
  + Pre-test: before experiment, helps choose best stimuli
  + Manipulation checks: during experiment, after treatment, check effectiveness of selected stimuli (e.g., ask subjects about valence of the ads)

### Ethics (debriefing and deception)

* Informed consent before the experiment
* Withdrawal at any point for promised compensation
* Learning experience
* Debriefing
* Participants’ awareness of hypotheses
* Cover story or deception? (Not in Economics!)

### Analysing experimental data

* Basic idea of ANOVA: Decomposes total variation in dependent variable to variation due to factor(s) and variation due to random error.
* F-statistic compares the mean variation due to each factor (main effect, interaction) to the mean variation due to error

### Interaction effect

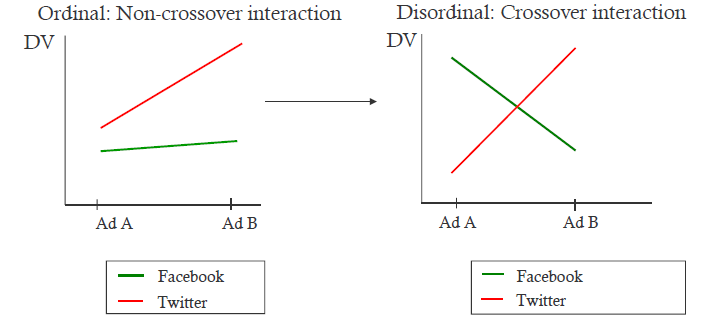
An interaction effect occurs when the effect of an independent variable on a dependent variable is different for different categories or levels of another independent variable.

Diagram

Description automatically generated Chart, line chart

Description automatically generated

* + Definitions. **Ordinal interaction** - an interaction where the rank order of the effects attributable to one factor does not change across the levels of the second factor.
  + Definitions. **Disordinal interaction** - the change in the rank order of the effects of one factor across the levels of another.
    - Crossover
    - Non-crossover



Examples

* + Mood matching
  + Session 2 dataset

# Exercises

## E1 Questionnaire preparation

* Giving user too much information can introduce the [selection bias](#_Biases)
* Gift cards can be a good idea to motivate people
* Example of the survey introduction
  + Comments on the introduction:
    - Saying how long the questionnaire is fine, it’s better to be honest and write the exact time for the questionnaire. Omitting the time - trying to get rid of the problem. Could also be that 25 minutes is too long for a bachelor thesis. If you do not pay people, it’s hard to find people who will answer this. It always depends on the case.
    - Writing that it is your bachelor thesis can be super annoying and might bias the respondents.
  + You can also write that we are conducting a research project for TUM Chair of Marketing, but maybe not writing that it is your bachelor’s thesis.
  + Time also should be added here.
* Cluster analysis questions - to identify clusters of consumers. Understand what if the best channel to reach your consumer.

## E2 Introduction to SPSS & Descriptive statistics

|  |  |  |  |
| --- | --- | --- | --- |
| SPSS | STATA | JASP | Excel |
| * Doing any statistical analysis * No codes are required | * Can write code * Harder to learn to use then SPSS * Expensive | * Free * Reads CSV files * Can use R | * Also can do statistics |
| * If multiple answers are possible, I should create a column in Excel for each possible response category. |  |  | * Options -> add ons - > analysis toolpack -> exit, then data -> data analysis * Do not forget to check the data types (in case if you get an error) |

### JASP

1. Make csv
2. Synchronize changes excel with JASP (menu -> sync data)
3. Check labels – click on column **headers** in JASP
4. Check variable types - clock on column header **icons** in JASP
5. Descriptive statistics – add variables, see missing values
6. Create new variables in csv file

## E3 Inferential statistics

* Exercise: do women like branded beer same as men?
  + What is H1 and H0 (we wanna accept H0)?
    - H0 - H0 should somehow specify that something is equal. Women like branded beer as much as men.
    - H1 - Alternative, which we want to reject. Women don't like branded beer as much as men.
  + Which test should you use?
    - Use independent samples t-test (If we have two independent groups - we should use t-test)

## E4 Factor analysis

## E5 Experiments

## E6 Exercise Exam (online)

# Research report format

A picture containing timeline

Description automatically generated

# Biases

|  |  |  |
| --- | --- | --- |
| Name of the bias | Happens in… | Associated problems… |
| **Selection bias / sampling bias** | Non-probability sampling | Has problems of causal interpretation. If you see some set of data which is not randomly selected.  In non-probability sampling makes it impossible to make inferences about a population, e.g., ruins the representativeness of a sample |
| **Retrospective bias** | longitudinal research | Recall bias is a [systematic error](https://en.wikipedia.org/wiki/Systematic_error) caused by differences in the accuracy or completeness of the [recollections](https://en.wikipedia.org/wiki/Recollection) retrieved |

# Possible questions

## All

1. What types of the **marketing research** are there? -> Marketing research types
2. What are the different approaches/**research design** to marketing research? -> Types of research design (3)
3. Questions about **sampling** procedure -> Sampling approaches
4. Questions about different **variable types**? -> Variable types
5. Which diagram will you use to display the frequency / mean / the relationship between the two variables?
   1. [Bivariate analysis](#_Statistical_techniques) involves the analysis of two variables, e.g., covariance can be used to determine the relations between two variables
   2. Two-way scatter plot for 2 variables -> Associative Analysis
   3. For frequency data -> histogram
   4. Example plots categorized by purpose
6. Which are the main mistakes in the questionnaire? How can the questionnaire be improved? -> Improve questionnaires
7. Is the data left or right skewed? -> Shape of distribution
8. What is crossover, what is non cross over? -> Interaction effect
9. What is interactions effect? When does it happen? -> Interaction effect
10. Validity of experiments -> Types of validity (goals of experiment)
11. Calculate z-values

Table

Description automatically generated

1. **Interpreting p-values from the table**
2. Tests: Chi-square tests, one-sample t-test, z-proportion test, independent sample t-test, 2x2 ANOVA
3. **Interpreting factor analysis**

## Other

1. Deutsche Bahn market research example. What was the goal, problems, lessons learned.
2. Interracial marriage raise example.
3. COVID example.
4. Thicker people live longer than thinner people / Thicker people tend to die earlier than thinner example.
5. Screening tests for diseases example.
6. Drinking more red wine is good for your health example.
7. Thesis result writing example.
8. Projects study description example.
9. What are existing methods in empirical consumer behaviour research (book)?
10. What important scales exist?
11. What are dos and donts of the questionnaires design?
12. Harris Interactive Boeing example.

## Trial exam questions

Please perform a factor analysis with the energy drink data set (Principial Component Analysis, Varimax Rotation) with variables of question 5: If you consume energy drinks, which IMPORTANCE do the following attributes have in your choice? (1=less and 9 =a lot):

q5\_Benefits, q5\_Brand, q5\_Packaging, q5\_Availability, q5\_Price, q5\_Variety, q5\_Taste,q5\_Healthiness, q5\_Sparkling, q5\_Freshness, q5\_Color, q5\_Calories, q5\_Digest

**Question**: For these questions, do we meet the requirements for performing a factor analysis in terms of the Kaiser-Meyer-Olkin measure of sampling adequacy

Please find the appropriate test and type in the value below:

**Question**: Please perform the appropriate test to examine whether the correlation matrix is an identity matrix.

Which of statements below is correct?

* The p-value of the appropriate test is < .001. Therefore we conclude that the correlation matrix is an identity matrix
* The p-value of the appropriate test is > .001. Therefore we reject the hypothesis that thecorrelation matrix is an identity matrix
* The p-value of the appropriate test is < .001. Therefore we reject the hypothesis that thecorrelation matrix is an identity matrix

**Question:** Based on (i) the eigenvalue criterion of Kaiser, how many factors would you extract?

* 3
* 4
* 5

**Question:** On which factor does the variable "brand" load most highly on (use the rotated component matrix)?

* 1
* 2
* 3
* 4

**Question:** On which factor does the variable "brand" load most highly on (use the rotated component matrix)?

Please type in the value of the factor loading below:

**Question:** Please briefly explain why for some variables it does not make sense to compute the median. How is this related to the level of measurement?

**Question:** Does satisfaction with favourite energy drink 1 (q10\_statis\_drink1) depend on whether participants are male or female (gender = factor 1) and on whether people consume energy drinks in the disco (q2\_disco = factor 2)?

Attention: Filter the gender variable and only include females and males

Please type in the p-value of the interaction between gender and disco:

**Question**: Is the average importance of the attribute “Brand” (q5\_Brand) equal to 4.5?State H0 (use the text box below):

**Question**: Are women as much satisfied with their favorite energy drink (q10\_satis\_drink1) as men?

Please perform the appropriate statistical test and type in the p-value below:

**Question**: Is the average importance of the attribute “Brand” (q5\_Brand) equal to 4.5?Which ststement is true?

* The p-value is > .001
* The p-value is < .001
* Both statements are wrong

**Question**: Please analyse the relationship between the answers to the filter question ("Filter") and "Gender" using a Chi-square test

Please type in the "Chi-Square value" you obtain from this test: