Design of Experiments - Exam Summary

Research Design

Observational vs. Experimental

Researchers simply observe and measure variables without actively intervening. In experiments, variables are purposefully manipulated to determine a cause-and-effect relationship.

Randomization

The effect and aim of randomisation is to eliminate selection bias and confounding factors, and to ensure comparability between groups at the start of the study (baseline data).

Confounder

Confounding occurs when a factor (confounder) that has not been investigated is associated with both the independent and dependent variables, causing a spurious correlation between them. (Example: Firefighters -> Damage, Confounder = Size of Fire)

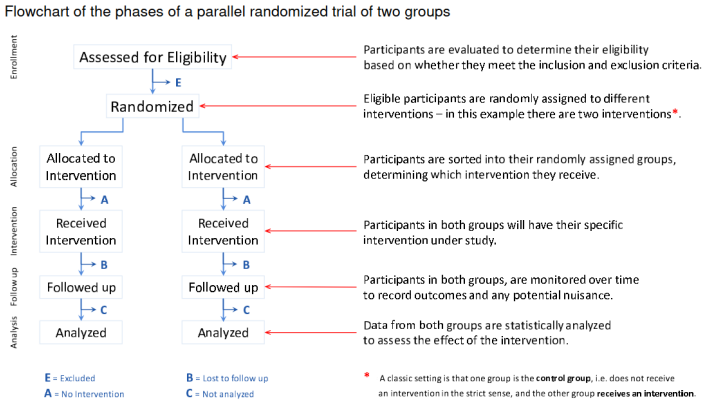

Figure 1: Randomized Controlled Tree

Blinding

Blinding is a suitable technique for avoiding distortions. It eliminates conscious and unconscious influences on the treatment result.

Types of blinding

- Open: No blinding
- Single-blind: Participants don't know their group assignment (e.g., whether they're receiving the real treatment or a placebo).
- Double-blind: Neither the participants nor the researchers administering the treatment know the group assignments.
- Triple-blind: Participants, researchers, and the data analysts are all unaware of the group assignments.

Decision Tree for Research Design

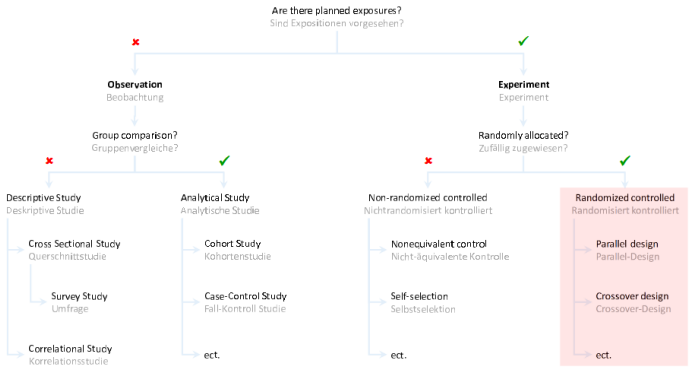

Figure 2: Decision Tree for Research Design

Principles

Introduction to scientific theory

Scientific Theory

It is a branch of philosophy that deals with the theory of scientific knowledge and scientific methods, as well as with research.

Methodology

Provides the instructional framework on how to proceed in order to gain scientific knowledge. Systematized procedures and approaches for obtaining knowledge.

Selection of positions held in scientific theory

- Classical rationalism: Reason precedes experience and there are so-called "innate" concepts of reason.
- Inductive empiricism: Findings are derived inductively based on observations and experiences.
- Logical positivism: The use of logic makes it possible to separate science from metaphysics.
- Critical rationalism: Findings are derived deductively based on observations.
- (Social) constructivism: Individuals construct their reality by relating their thinking and actions.

  Note: Inductive empiricism and critical rationalism belong to **empirical research**.

What is empirical research?

Empirical research examines the environment by means of observation and experiment. There are many research methods for conducting observations and experiments (Interview, Case study, Survey, experiments, etc.)

Landscape of empirical research

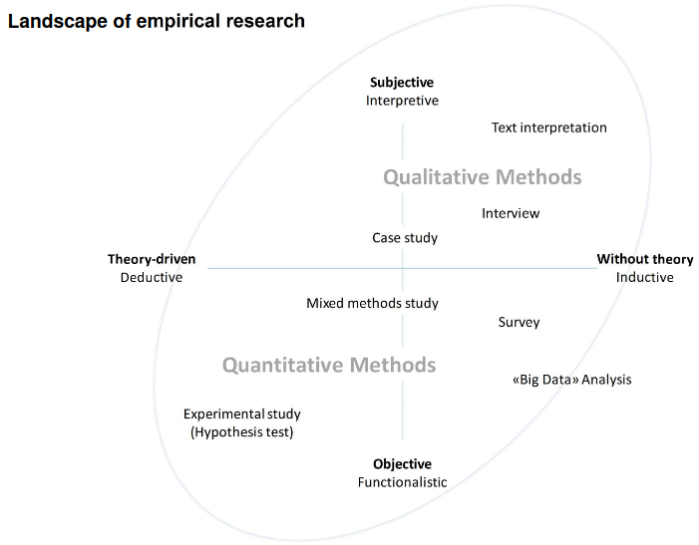

Figure 3: Landscape of empirical research

Quantitative Methodes

The approach assumes that a theory can never be finally verified, it can only be falsified.

Descriptive statistics

Describes the data to be analyzed. Is limited to a sample as a subset of the population. Does not allow for conclusions to be drawn about the population.

Inferential statistics

For drawing conclusions about the population based on information obtained from a sample. Use statistical hypothesis tests, especially, as the main component.

Hypothesis Testing

Hypothesis testing is a statistical method used to determine if there is enough evidence in a sample of data to support a specific hypothesis about a population.

- Alternative hypothesis (H_A): Research hypothesis to be tested that postulates the presence of a certain effect (e.g. a difference) in the population.
- Null hypothesis (H_0): Postulates the opposite, namely the absence of an effect.

Research process

Phases

1. Formulation of the research problem & study design
2. Planning and preparation of the study
3. Data collection
4. Data Analysis
5. Reporting

Measuring Instrument

A process that uses a given set of circumstances to define and specify subsequent research steps with a view to better understanding these circumstances.

Sampling procedure

Process of selecting a representative subset of cases from a larger population to draw empirical conclusions about the whole.

Study types

- Descriptive study: Descriptive character. Suitable for forming hypotheses (Surveys).
- Analytical study: Identification and quantification of effects / verification of relationships. Not fully suitable for hypothesis testing (Cohot).
- Randomized controlled: Suitable for hypothesis testing (RCT).

Introduction to Design of Experiments (DoE)

Cause and Effect

A trial / experiment is carried out to discover a cause-and-effect relationship in a process.

Terms

- Input: Trial objects, test objects, test persons, ect.
- Process: Process in which controllable and non-controllable factors influence the input.
- Output (aka. **Dependet variable: DV**): Input changed by the process, result of the test/experiment.
- Controllable factors (aka **Independet variables: IV**): Influencing factors whose strength can be adjusted within defined limits.
- Non-controllable factors (aka. **Nuisance variables**): Influencing factors whose strength cannot be determined but that can be measured / cannot be determined and that cannot be measured.

Causality in observational and experimental study designs

Observational studies cannot directly prove causality, but only show correlations or associations. Since the assignment is not random, there is always a risk that the results are distorted by unknown confounding factors. Experimental studies (e.g. RCTs) can prove causality because they control for confounding factors through randomisation, thereby isolating the effect of the cause. **They are the gold standard**.

Variance

The variance describes the mean square deviation of the individual measured values from the empirical mean.

- Primary variance: Impact of (experimental) factors in an experiment on the change / variation of the output to be examined.
- Secondary variance: Variation of the output to be examined, caused by nuisance variables. Not in the focus of the study.
- Error variance: Variation caused by measurement errors and random processes.

  Note: Secondary and error variances are grouped to the residual variance.

Variation of Variance

The variance of the dependent variable (DV) (primary variance) should be attributed to the systematic variation of the independent variable (IV). The secondary variance should be controlled and the error variance minimized.

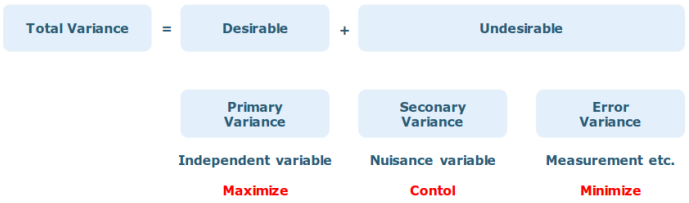

Figure 4: Summary Variance

Maximizing the primary variance

- Relationship is linear: Selecting of extreme values in the IV.
- Relationship were curvilinear: Selecting optimal increments of IV.
- Relationship were unknown: Selecting as many increments of IV in the smallest steps as possible.

Control of the secondary variance

- Keeping constant: Keeping the experimental setup constant.
- Repetition: Several measurements are repeated on the same trial objects.
- Randomization: Trial objects are assigned randomly to Treatment and Control groups to eliminate systematic bias
- Blocking: Trial objects are grouped into homogeneous blocks based on one or more influential variables to reduce variability.
- Covariate adjustment: Nuisance variables are included as covariates in the statistical model to account for their effects.

Minimizing the error variance

- Reliable measurement setup: Standardization of the experimental conditions
- Sample size: Larger sample sizes reduce the impact of individual measurement errors
- Suitable analytical methods: Use of robust estimators to account for heterogeneous error variance

Properties of measurement instruments

- Objectivity: Objectivity of an instrument is given when the results are independent of personnel and calculation methods.
- Reliability: Reliability is the degree to which an instrument produces the same result each time under comparable conditions.
- Validity: Validity is the extent to which an instrument measures what was intended.

Properties of DoE

Design of Experiments Types

Trial and error

Combination of parameters have no structure and are mixed randomly. No idea what factors influence how.

One-factor-at-a-time

Vary the first factor and then measure fuel consumption. Keep the setting with the lowest consumption and then vary the next factor. Easy to implement, but interaction between factors are not recognized. Research question is answered neither systematically nor exhaustively.

Full factorial design

Two levels (+/-) are defined per factor. All possible combinations of factor levels are varied. All main effects and all interactions can be determined. Can be used as a screening experiment to identify potentially important variables. The effort involved increases rapidly as the number of factors increases. Each additional factor doubles the number of combinations.

Profile Plot

Impact of the factors on the dependent variable x. Based model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \cdots + \beta_n x_n$$

Factorial design with interactions

Interactions can occur in experiments with two or more independent variables. An interaction of two factors means that the two factors interact in a complex way. If there is an interaction, the effect of one factor depends on the levels of the other factor. Interaction terms are written as multiplication.

- Two way interaction: $x_1 \times x_2$
- Three way interaction: $x_1 \times x_2 \times x_3$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2$$

Full factorial designs

Generalization of two-level full factorial design with k factors and n levels. All possible factor combinations are varied.

$$\text{combinations} = n^k$$

Fractional factorial designs

Only a (balanced) part of the possible combinations of factors are varied.

$$\text{combinations} = n^{k-1}$$

Design

- Procedure: The factor levels are determined before the experiment
- Factor combinations: Only a part of the possible combinations of factors are selected
- Restrictions: In fractional factorial designs, interactions can only be partially measured because not all possible combinations of factor levels are tested.
- Advantages: The effort involved is significantly lower compared to full factorial designs
- Statistical analysis: As in the case of full factorial design, but without interactions

Quality criteria of experiments

Internal Validity

Exists when changes in dependent variables (DV) are attributed to independent variables (IV). Increases with decreasing impact of nuisance variables.

Population Validity

Degree to which the results of a study can be generalized from the sample to the whole population.

Situation Validity

Degree to which the findings of a study can be applied to different situations.

External Validity

Exists when experimental results from a sample can be generalized to the entire population. Increases with increasing naturalness.

Construct Validity

Effectiveness of the measurement methods in precisely capturing the intended construct

Relationship between Internal vs. External Validity

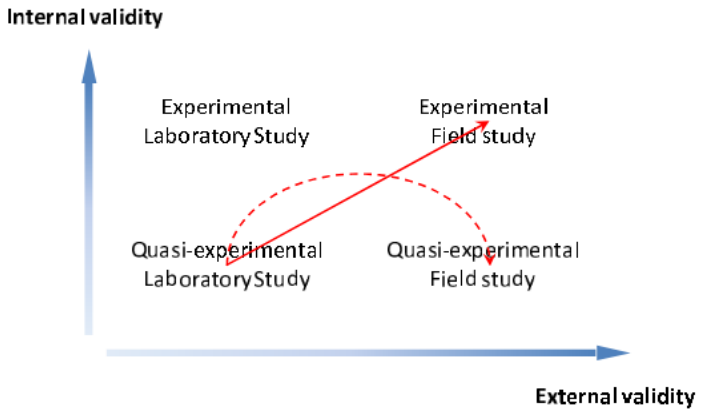

Figure 5: Relationship between Internal vs. External Validity

The lowest general level of validity is at the bottom left for the quasi-experimental laboratory study, and the highest is at the top right for the experimental field study. A well-controlled lab experiment may maximize internal validity by eliminating confounding variables.

Sampling

Population

Set of all (potentially explorable) elements that have a common characteristic or a common combination of characteristics. The population is often very large or not fully accessible. The population can be defined but not identified.

Characteristics of a sample

A sample is a subset of all observation units and should reflect the relevant aspects of the population as accurately as possible. Three elements contribute to creating or describing representativeness:

- The sample is drawn randomly.
- Estimation procedure for generalizing from the sample to the population is reported.
- Accuracy is reported, which is influenced by the sample size, among other things.

Point estimate of the mean

An "estimator" is a function that calculates a value.

- $\bar{x}$: Mean of the sample
- μ_0: True mean in the population (generally unknown)

The mean value x of a sample is an unbiased, efficient and consistent estimator of the true mean value in the population: $\mu_0 = E(\bar{x})$

Sampling methods

- Probabilistic (random) sampling procedures: Selecting elements based on a random mechanism
- Non-probabilistic (non-random, purposive, arbitrary) sampling procedures: Selection of the elements is not based on a random mechanism, but is made by certain decisions (purposive or arbitrary selection of elements)

Sampling Map

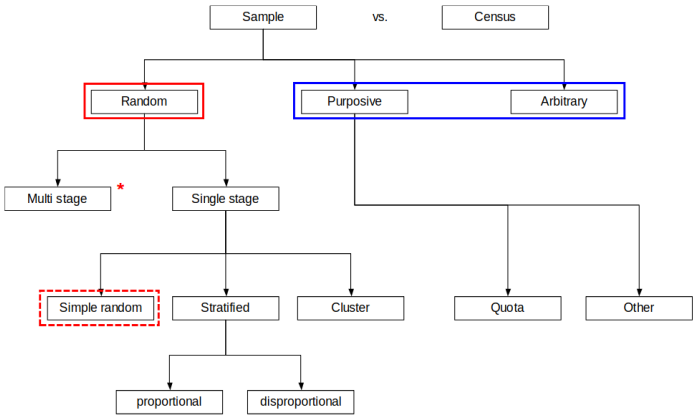

Figure 6: Sampling procedure Overview

Simple random sampling (SRS)

Random selection of n elements from the N elements of the population. Each element has the same probability of being included in the sample.

```
data <- rnorm(100) # 100 random numbers
View(data)

data_sample <- sample(data, 10, replace=FALSE) # Take 10 random sample from data
View(data_sample)
```

Stratified random sampling

Structure of the population regarding certain characteristics is known in advance. The population is stratified in accordance with these characteristics.

- Proportional: Selection set in each Disproportional: Selection rate per stratum is the same (self-weighting)
- Disproportional: Selection rate per stratum is the same (self-weighting) is different (samples will be weighted)

Cluster sampling

The elements are selected at a higher level. Variability within clusters is small (cluster elements are very similar) while variability between clusters is large (clusters differ greatly).

Arbitrary sampling

Arbitrary sampling occurs when a sample is selected based on the researcher's discretion, convenience or ease of availability, without applying a specific, structured or random procedure.

Targeted sampling

For populations that are difficult to reach and whose members are not closely networked. Preferred locations or places of residence of the members are identified and then systematically recruited on site.

Respondent Driven Sampling

Multiple waves of peer-to-peer recruitment with statistical adjustments are conducted to approximate a random sample. Recruited individuals are only allowed to recruit a limited number of other individuals and are only rewarded for each person actually recruited.

Representativeness

The degree of representativeness is not measurable. The sample should be representative with regard to key characteristics of the study.

Sampling errors

Non-sampling error

Difference in the mean value between the defined ideal population and the real population that cannot be attributed to deficiencies in the random selection of the sample.

- Coverage error: Part of the population cannot be identified.
- Systematic non-response: Lack of information on certain individual elements.

Sampling Error

Difference between the estimated mean value from a randomly drawn sample and the real mean value of the population.

- Selection error: Not all elements of the population have the same selection probability.
- Use of an unsuitable estimator.

Variability of Sample Means

In randomly drawn equal samples, the sample means vary depending on:

- Attribute: The more heterogeneously an attribute is distributed in the population, the greater the variability of the sample means among many samples.
- Sample size: The smaller the sample size, the greater the variability of sample means among many samples.

Standard Error

The standard error is a measure of the variability of the sample means among many samples. It quantifies the spread of sample means from repeated random samples of the same size around the population mean μ_0

$$\hat{\sigma}_{\bar{X}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

The larger the sample size, the smaller the standard error, and therefore, the larger the sample size, the more precise the sample mean is as an estimator of the population mean.

Effect size & Power analysis

Statistical significance and importance of an effect

The chance of having a significant result in a hypothesis test is:

- larger if sample size n increases.
- smaller if standard deviation s increases.

An effect in the population can be specified by an effect measurement. This measurement results in what is referred to as effect size.

> Note: The name effect size (ES) comes from Cohen (1992)

Effect size

Effect size is a statistical measure that quantifies the magnitude (or strength) of a phenomenon. A tiny p-value in a very large study might show a difference that is statistically significant, but practically meaningless. Effect size expresses zhe difference in terms of standard deviations.

- A Cohen's d of 1.0 means the means of the two groups are separated by exactly one standard deviation.
- A Cohen's d of 0.0 means there is no difference between the means.

Table 1
ES Indexes and Their Values for Small, Medium, and Large Effects

Test	ES index	Effect size		
		Small	Medium	Large
1. m_A vs. m_B for independent means	$d = \frac{m_A - m_B}{\sigma}$	.20	.50	.80
2. Significance of product–moment r	r	.10	.30	.50
3. r_A vs. r_B for independent rs	$q = z_A - z_B$ where z = Fisher's z	.10	.30	.50
4. $P = .5$ and the sign test	$g = P - .50$	.05	.15	.25
5. P_A vs. P_B for independent proportions	$h = \phi_A - \phi_B$ where ϕ = arcsine transformation	.20	.50	.80
6. Chi-square for goodness of fit and contingency	$w = \sqrt{\sum_{i=1}^{k} \frac{(P_{1i} - P_{0i})^2}{P_{0i}}}$	.10	.30	.50
7. One-way analysis of variance	$f = \frac{\sigma_m}{\sigma}$	.10	.25	.40
8. Multiple and multiple partial correlation	$f^2 = \frac{R^2}{1 - R^2}$	.02	.15	.35

Figure 7: Effect Size Table

Power analysis

Power analysis is the crucial step in research design that connects your expected effect size with your desired power (80%) and significance (5%) to determine the minimum **sample size** you need to run a valid study.

Paradigms

The Fourth Paradigm

1. Empirical Science: Description of natural phenomena.
2. Theoretical Science: Modelling and generalization.
3. Computational Science: Simulation of complex phenomena.
4. Data-Intensive Science (eScience): Synthesis of information technology and science.

Quantitative empirical research

1. Formulation of teh research of the study
2. Planning and preparation of the study
3. Data collection
4. Data Analysis
5. Reporting

Note: Deductive approach → Conclusion from the general to the specific.

Limitations of quantitative empirical research

- Meaning of significant hypotheses vs. meaning of effect size.
- p-hacking (looking for data subsets and configurations until the p-value is less than 5%).
- Assumptions about the distribution of variables are violated.
- Assumption of homogeneity of variance is violated.

Data-driven research

- Visual Analytics: Exploratory Data Analysis & Descriptive Statistics.
- Machine Learning and Predictive Modelling: Regression / Classification / Decision Trees.
- Advanced Analytics for Unstructured Data.

Note: Inductive approach → Conclusion from the specific to the general

Limitations of data-driven research

- Big Data Hubris: Correlation is understood as causality.
- Sparse data: Although the data basis is "big," it contains little information.
- Data analysis: A whole range of methodical errors.

Abductive approach and combination of approaches

Induction: Search and generation of theories that fit the research context. Induction shows that something actually is operative. Deduction: Verification or falsification of existing theories. Deduction proves that something must be. Abduction: Search and generation of new, also speculative theories. Abduction merely suggests that something may be.

> Note: Ideally, all three methods are used cyclically.

ANOVA

The Analysis of Variance (ANOVA) is a statistical test used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups.

- One-Way ANOVA: Used when comparing means across groups based on one single categorical independent variable (factor).
- Two-Way ANOVA: Used when comparing means based on two or more independent variables (factors), allowing for the testing of interaction effects between the factors.

Key Steps in Analysis of Variance

1. Design of experiment
2. Calculating differences and sum of squares
3. Verification of the model
4. Considering other aspects (post hoc tests)
5. Testing of assumptions (homogeneity of variance, etc.)
6. Interpretation of the model and reporting (profile plots)

> Note: Step 2 is done by the ANOVA model

One-way ANOVA in R

```r
oneway.test(salary ~ factor(experience), var.equal=TRUE, data=data)
```

```r
fit <- aov(salary ~ factor(experience), data=data)
summary(fit, intercept=TRUE)
```

Two-Way ANOVA in R

```r
# Only main effects, no interaction
fit <- avo(salary ~ factor(experience) + factor(position), data=data)
summary(fit, intercept=TRUE)

# With Interaction
fit <- avo(salary ~ factor(experience) * factor(position), data=data)
summary(fit, intercept=TRUE)
```

Main Effects

The direct effect of an independent variable on the dependent variable is called main effect. Profile plots are used as visualization.

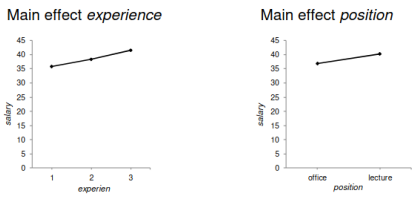

Figure 8: Profile Plots of Main Effect

> Note: If the profile plot shows a (nearly) horizontal line, the main effect in question is probably not significant.

Interaction Effect

An interaction between experience and position means there is dependency between the two variables. The independent variables have a complex influence on the dependent variable. The factors do not just function additively but act together in a different manner.

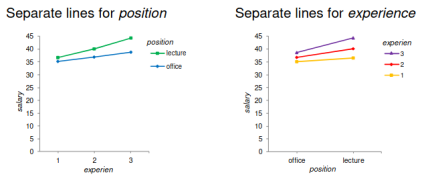

Figure 9: Profile Plot of Interaction Effect

Note: If there is no interaction, the lines are parallel. If there is an interaction, the lines are not parallel.

Prerequisites of ANOVA

- Sampling: Randomly assigning participants to the treatment or control groups.
- Distribution of residuals: Residuals are normally distributed.
- Homogeneity of variances: Residuals have constant variance
- Balanced design: Same sample size in all groups.

Note: ANOVA is relatively robust against violations of prerequisites.

Post Hoc Tests

There are different methods to compare groups in pairs. All methods are similar, however, in that they solve the problem of multiple testing.

```
pairwise.t.test(x, y, p.adj="bonf")
```

Effect size

Partial eta squared η_p^2 relates the variance explained by one factor to the variance not explained by other factors in the model.

```
library(effectsize)
eta_squared(fit)$Eta2 # Use your ANOVA fit
```

Overview over Statistical Hypothesis Tests

Choosing the type of analysis depending on level of measurement

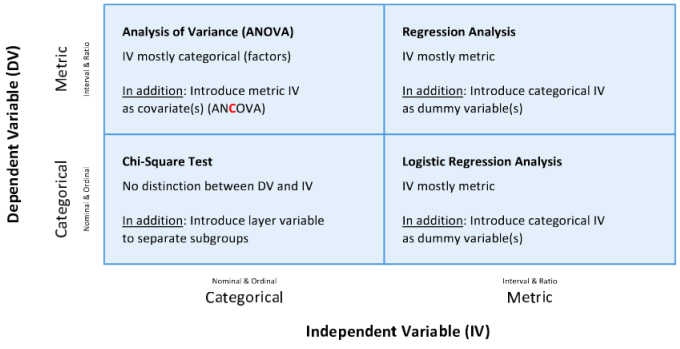

Figure 10: Overview over Statistical Hypothesis Tests

A/B Testing

What is A/B testing?

A/B testing is an experiment. The research questions are applied to two (A/B) or more randomized groups. The statistical analysis is done by t-test, ANOVA and more advanced methods.

Carrying out A/B testing

The aim is to determine which version achieves a better result, in terms of click-through rate etc. Two versions A and B are tested in parallel in a live environment. The generated data becomes the basis for decisions.

Bandit algorithm

Several variants (treatments A, B, C, …) are run in parallel. The variant with the highest "success" (according to metrics) get more data traffic. Other variants are refined and tested with proportionally less traffic. Traffic allocation is continuously adjusted based on the observed success. As the test progresses, more and more information about the performance of the variants becomes available, so that the most successful variant can be identified dynamically.

> Note: A key advantage of bandit algorithms is that they can achieve higher overall profit while still collecting data on the other variants.

Measurement / Metrics / KPI

- Measurement: Method for obtaining one or more measured values that can be assigned to a quantity
- Metrics: Calculation from measured values
- KPI: Quantifiable metric that shows how effectively the most important company goals are achieved

Error sources and pitfalls

A/B testing cannot be used in all research questions. Example: For a complete website redesign (further elements have to be included). Population is unknown, sampling procedure is not suitable, sampling bias, ect.

Dark Pattern / Deceptive Pattern

Dark / deceptive patterns are patterns of persuasion and influence. They may increase short-term gains for the provider but can harm brand image, user experience, and customer satisfaction. A/B testers must ensure that the variants they test are ethically acceptable and aim to create genuine added value for the user. The problem with testing dark patterns is that A/B tests typically measure short-term metrics (such as clicks or conversions). They are not good at capturing long-term damage.

Factorial designs

Full factorial designs

A full factorial design tests every possible combination of all factor levels.

Single factorial (One)

One independent variable IV (factor) acts on the dependent variable DV. Two groups are compared:

- 1 IV: one IV with two levels (dichotomous)
- 1 DV: metric scaled
- Examples: A/B test, RCT with treatment and control or two treatments
- Analysis: One-way ANOVA

One independent variable IV (factor) acts on the dependent variable DV. Several groups are compared:

- 1 IV: one IV with several levels
- 1 DV: metric scaled
- Examples: RCT with combinations of more than two treatments and control
- Analysis: One-way ANOVA

Multifactorial (Several)

Several factors act on the dependent variable DV. Several groups are compared:

- X IV → several IV with two or more levels each
- 1 DV → metric scaled
- Examples: Dwell time with two IV
- Analysis: Multi-factorial ANOVA (two-way ANOVA, three-way ANOVA, …)

Fractional factorial designs

A fractional factorial design tests only a carefully selected subset (a fraction) of the possible combinations.

The more factors and the more characteristics, the larger the number of groups. Provided that there are no interactions, the experiment can also be conducted successfully with a reduced number of groups. By using Latin squares or related designs, the number of groups required can be significantly reduced compared to full factorial designs.

In principle multifactorial (Several)

Several factors act on the dependent variable DV. Several groups are compared: - X IV → several IV with two or more levels each - 1 DV → metric scaled - BUT Not all possible combinations are considered

Latin square

A Latin squares design makes it possible to study the main effects of factors without having to observe all combinations of treatment levels. All combinations lead to a total of 27 (full factorial).

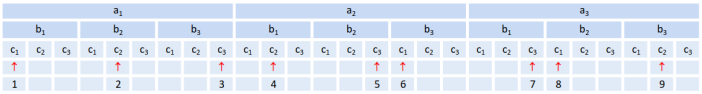

Figure 11: Latin Square

Here we only use 9 combinations (fractional factorial).

> Note: A Latin squares design can only be used if it follows from theory or empirical evidence that the joint effect of the factors does not produce interactions.

Large data quantities

What are large data quantities?

- Samples are taken as part of a study. Primary goal: To answer research questions
- Administrative data are collected for various reasons. Primary goal: To serve documentary and administrative purposes
- Grey area Data from full surveys (census), from social media and from "representative" surveys lie somewhat between data from a sample and administrative data.

This terminology has become established in many fields of research: - Made data → Data is generated by researchers ("made"). - Found data → Data are obtained administratively and technically ("found")

What is bias?

Deviation between mean μ_0 in the population and sample mean $\bar{x}$. Three elements determine the bias:

1. Data quality measure
2. Data quantity measure
3. Problem difficulty measure

How can bias be quantified?

$$\text{Bias} = \bar{x} - \mu_0 = \rho_{R,G} \times \sqrt{\frac{1-f}{f}} \times \sigma_G$$

- $\rho_{R,G}$: Data Quality
- $\sqrt{\frac{1-f}{f}}$: Data Quantity
- σ_G: Problem Difficulty

Data defect correlation $\rho_{R,G}$

In $\rho_{R,G}$ the R is a function that shows how data is obtained from the population. In simple random sampling, the R function generates a randomly generated sequence of elements drawn from the population. Because of the random process, the selection of an element is independent of G.

Statistical paradises and paradoxes in relation to administrative data sets

Measure for the size of the bias → Mean-squared error (MSE). The MSE measures the deviation (bias) of the estimator $\bar{x}$ from the mean μ_0 in the population. The bias goes to 0 only, if the size n of the administrative dataset goes against N ($n \to N$). The absolute size n of the administrative dataset is meaningless without specifying N.

Spurious Correlation

A spurious correlation is a relationship between two variables that appears to be statistically significant and related, but is not actually caused by one another. The relationship is deceptive or coincidental, and often results from either a hidden factor or simply random chance. The primary risk of spurious correlation is that people might mistakenly conclude that a causal relationship exists based solely on the correlation, leading to incorrect policies or decisions.

Experiments in social media

Communication in SoMe

- One – e.g. X: Specific information is sent from one source to «everyone»
- Two – e.g. SMS / WhatsApp: Exchange of information between two individuals via a certain medium.
- Many – e.g. Facebook / WhatsApp: Information is exchanged in a group via social media actively or passively

Mode

The mode of interaction is also changed during a session. Users can speak, text, email, video chat, and post items on social media channels or blogs. The communication / interaction is not always carried out in the same mode. For example, a telephone call can be answered with a text message.

Population bias / Selection bias

Bias essentially means that the population being examined does not correspond to the defined population. If the defined population refers to the entire population and their subgroups, social media is by definition subject to bias. This is mainly due Sampling frame and Sampling procedure. Certain manifestations of these elements can be observed mainly (in some cases exclusively) in social media.

Sources of bias in social media

- Activity bias I: Active (only) at the time of the study / data collection / storage.
- Activity bias II: A few users are very active on social media, while most users use social media only passively.
- Activity with bots: Programs that behave like users or react to specific triggers.

Standardization of bias in social media

- Medium & platform: Restrictions on access to data / (unknown) ways and methods on how data is stored
- Survey / sampling & representativeness: Not taking into account the circumstances, e.g. in the case of a storm on X
- Ways of managing data and sources: Reproducibility limited by restrictive access to data / by deletion of items

Research with social media

Studies using data from social media have great potential for investigating research questions in social science / psychological research questions that are new or place special demands on the study design.

- Rapid availability of data / information and continuous updating
- Simple and low-cost extraction processes compared to classical surveys

  Note: Data from social media is most likely «found data».

Methods of data collection in SoMe

- Using the functionalities of social media: Use individual snowball sampling
- Paid access to survey: Place ads in social media, Use survey tools or Include specialized platforms (Amazon Mechanical Turk)

Exercises

5.3 - Data Sampling

Given is an address list address.csv Use R to draw a simple random sample with n = 50 elements from this data set. Insert also the R-code, the R-output and if necessary, R-plots in your answer.

```
library(readr)
data <- read_csv("address.csv")
data_sample <- sample(data, 50, replace=FALSE) # Take 50 random sample from data
View(data_sample)
```

6.1: Effect Size

$$\hat{d} = \frac{|\bar{x}_{\text{new}} - \bar{x}_{\text{current}}|}{\sigma_0} = \frac{1.0}{1.5} = 0.67$$

6.2 - Sample Size in R

```
library(pwr)
alpha <- 0.05
beta <- 0.2
power <- 1 - beta
mean <- 1.0
sigma <- 1.5
d <- mean/sigma
pwr.t.test(d = d, power = power, sig.level = alpha)
# n = 36.3 -> 72.6
```

6.3 - Sample Size ANOVA

```
library(pwr)
alpha <- 0.05
beta <- 0.2
power <- 1 - beta
d <- 0.5 # From table
pwr.anova.test(f = 0.25, k = 2, power = power, sig.level = alpha)
# n = 63.8 -> 127.6
```

8.2 - Effect Analysis with ANOVA

```
# install.packages("readxl")
library(readxl)
data <- read_excel(path="/home/nils/dev/mscids-notes/hs25/doe/data/clickrate.xlsx")
oneway.test(DV ~ factor(IV1), var.equal=TRUE, data=data)
```

```
    One-way analysis of means

data:  DV and factor(IV1)
F = 77.928, num df = 2, denom df = 2301, p-value < 2.2e-16
```

```
fit <- aov(DV ~ factor(IV1), data=data)
summary(fit, intercept=TRUE)
```

```
             Df  Sum Sq Mean Sq  F value Pr(>F)
(Intercept)    1 3509768 3509768 41265.38 <2e-16 ***
factor(IV1)    2   13256    6628    77.93 <2e-16 ***
Residuals   2301  195708      85
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

9.2 - Latin Square

```
install.packages("agricolae")
library(agricolae)
my_design_lsd <- design.lsd(trt = c("c1","c2","c3"))
my_design_lsd$sketch
```

he column names of the Latin square ([,1] [,2] [,3]) correspond to the levels of factor A. The row names of the Latin square ([1,] [2,] [3,]) correspond to the levels of factor B. The content in the cells of the Latin square (“c1”, “c2”, “c3”) correspond to the levels of factor C We can reduce the combination of experiments from 27 (Full Factorial) to 9 (Fractional Factorial).

```
library(readxl)
dwelltime_Latin <- read_excel("/home/nils/dev/mscids-notes/hs25/doe/data/dwelltime_latin.xlsx")
View(dwelltime_Latin)
fit <- aov(DV ~ factor(A) + factor(B) + factor(C), data = dwelltime_Latin)
summary(fit)
```

```
               Df Sum Sq Mean Sq F value Pr(>F)
factor(A)       2  76057   38028 465.761 <2e-16 ***
factor(B)       2  27159   13580 166.320 <2e-16 ***
factor(C)       2    552     276   3.383  0.034 *
Residuals   10338 844075      82
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There is a main effect on A, B and C. Even with p = .034 the factor C is relatively close to the significance limit of p = .050.