

Quantitative Methods

© Prof. Dr. Stephan Huber

March 11, 2024

Table of contents

1 Preface	1
1.1 About the notes	1
1.2 About the author	1
1.3 Contact	2
1.4 About this course	2
1.5 Personal note	4
2 Doing research	5
2.1 What is research	5
2.2 Everybody can do research	5
2.3 It's difficult to do good research	8
2.4 Asking questions like a good researcher	10
2.5 Features of good research	11
2.5.1 Reliability and validity	11
2.5.2 Generalizability	13
2.5.3 Replicability, reproducibility, transparency, and other criteria	13
2.6 The role of resources, data and ethics	14
Glossary	17
2.7 Solutions to exercises	17
3 Identification	19
3.1 Data acquisition	19
3.1.1 Interviews	19
3.1.2 Surveys	20
3.1.3 Case studies	20
3.1.4 Experiments	21
3.1.5 Observational data	23
3.2 Correlation does not imply causation	24
3.3 The fundamental problem of causal inference	25
3.4 It's difficult to overcome the fundamental problem	29
3.4.1 Ignorability	29
3.4.2 Unconfoundedness	29
3.5 Statistical control requires causal justification	30
3.6 Solutions to exercises	30
References	34
4 Appendix	35

List of Figures

1.1 Prof. Dr. Stephan Huber	1
1.2 Richard P. Feynman's badge photo from Los Alamos National Laboratory	3
2.1 Zora Neale Hurston, 1891-1960	5
2.2 Causal Inference: The Mixtape	6
2.3 Children as little researcher	6
2.4 John Maynard Keynes (1883-1946)	8
2.5 Konrad Adenauer (1876-1967)	8
2.6 Sir Alexander Fleming (1881-1955)	9
2.7 The Effect: An Introduction to Research Design and Causality	11
3.1 Three methods of data acquisition	19
3.5 Daniel Kahneman and his best selling book	21
3.8 Randomized Controlled Trials (RCTs)	22
3.9 Observational data	23
3.10 Correlation does not imply causation	24
3.11 (#fig:cunbook) The book cover of “Causal Inference: The Mixtape?” by Scott Cunningham (2021)	25
3.12 (#fig:berkley) Graph taken from Bickel et al. (1975, p. 403)	26
3.13 (#fig:nealrct) Average treatment effect (ATE)	28

1 Preface

1.1 About the notes

- These notes aims to support my lecture at the HS Fresenius but are incomplete and no substitute for taking actively part in class. - A pdf version of these notes is available here
- I appreciate you reading it, and I appreciate any comments.
- This is work in progress so please check for updates regularly.
- Do not distribute without permission.

1.2 About the author

Figure 1.1: Prof. Dr. Stephan Huber¹



I am a Professor of International Economics and Data Science at HS Fresenius, holding a Diploma in Economics from the University of Regensburg and a Doctoral Degree (summa cum laude) from the University of Trier. I completed postgraduate studies at the Interdisciplinary Graduate Center of Excellence at the Institute for Labor Law and Industrial Relations in the European Union (IAAEU) in Trier. Prior to my current position, I worked as a research assistant to Prof. Dr. Dr. h.c. Joachim Möller at the University of Regensburg, a post-doc at the Leibniz Institute for East and Southeast European Studies (IOS) in Regensburg, and a freelancer at Charles University in Prague.

Throughout my career, I have also worked as a lecturer at various institutions, including the TU Munich, the University of Regensburg, Saarland University, and the Universities of Applied Sciences in Frankfurt and Augsburg. Additionally, I have had the opportunity to teach abroad for the University of Cordoba in Spain, the University of Perugia in Italy, and the Petra Christian University in Surabaya, Indonesia. My published work can be found in international journals such as the Canadian Journal of Economics and the Stata Journal. For more information on my work, please visit my private homepage at hubchev.github.io.

¹Source: Picture is taken from <https://sites.google.com/view/stephanhuber>

1.3 Contact

Hochschule Fresenius für Wirtschaft & Medien GmbH
Im MediaPark 4c
50670 Cologne

Office: 4e OG-3
Telefon: +49 221 973199-523
Mail: stephan.huber@hs-fresenius.de
Private homepage: www.hubchev.github.io
GitHub: <https://github.com/hubchev>

1.4 About this course

Workload of M-IBS 8 Quantitative & Qualitative Methods for Business

125 h = 56 h (in-class) + 21 h (guided private study hours) - 48 h (private self-study).

Workload of M-IBS 8.1 Quantitative Methods

62.5 h = 28 h (in-class) + 10.5 h (guided private study hours) - 24 h (private self-study).

Assessment

Students complete this module with a written exam of 120 minutes where 50% of the points stem from *M-IBS 8.1 Quantitative Methods* and 50% from *M-IBS 8.2 Qualitative Methods*. A passing grade in this module is achieved when the overall grade is greater than or equal to 4.0.

Learning outcomes:

After successful completion of the module, students are able to:

- assess and discuss coherent research paradigms, based on quantitative, qualitative, and mixed-methods research approaches,
- explain a broad set of quantitative and qualitative methods to collect, gather, illustrate, analyze, and interpret data,
- distinguish and discuss empirical strategies to identify causal mechanisms, causes, and effects.

How to prepare for the exam:

I am convinced that reading the lecture notes, preparing for class, taking actively part in class, and trying to solve the exercises without going straight to the solutions is the best method for students to

- maximize leisure time and minimize the time needed to prepare for the exam, respectively,
- getting long-term benefits out of the course,

- improve grades, and
- have more fun during lecture hours.

Literature:

Cunningham (2021), Huntington-Klein (2022), Illowsky & Dean (2018), Békés & Kézdi (2021), Paldam (2021)

Content:

- Research design
 - Research design
 - How to measure socio-economical reality
 - How to identify causes of effects
 - How to identify effects of causes
 - The selection problem and ways to solve it (matching, natural experiments, laboratory experiments)
- Statistical toolbox
 - Types of data (cross-section, panel, time-series, georeferenced)
 - Types of variables (continuous, count, ordinal, categorical, qualitative)
 - Data sampling methods
 - Descriptive methods (data visualization, statistical moments, correlation)
 - Methods of statistical inference (distribution, statistical tests)
 - Mathematical and statistical software packages (R, Stata, SPSS, Excel, WolframAlpha, etc.)
- Methods
 - Data mining (graphical visualizations, cluster analysis, factor analysis)
 - Regression analysis (matching, instrument variables, difference in difference, fixed effects, regression discontinuity)
 - Other methods (time series analysis, spatial analysis, simulations, qualitative comparative analysis, etc.)

About how to learn (and prepare for the exam)

Figure 1.2: Richard P. Feynman's badge photo from Los Alamos National Laboratory²



²Source: Picture is taken from <https://repository.aip.org/islandora/object/nbla%3A299600>

Richard P. Feynman (see Figure 1.2):

“I don’t know what’s the matter with people: they don’t learn by understanding; they learn by some other way by rote, or something. Their knowledge is so fragile!”

Stephan Huber:

“I agree with Feynman: The key to learning is understanding. However, I believe that there is no understanding without practice, that is, solving problems and exercises by yourself with a pencil and a blank sheet of paper without knowing the solution in advance.”

- Study the lecture notes, i.e., try to understand the exercises and solve them yourself.
- Study the exercises, i.e., try to understand the logical rules and solve the problems yourself.
- Test yourself with past exams that you will find on ILIAS. The structure of the exam is more or less the same every semester.
- If you have the opportunity to form a group of students to study and prepare for the exam, make use of it. It is great to help each other, and it is very motivating to see that everyone has problems sometimes.
- If you have difficulties with some exercises and the solutions shown do not solve your problem, ask a classmate or contact me. I will do my best to help.

1.5 Personal note

Dear students,

If the title of this course “Quantitative & Qualitative Methods for Business” seems uninteresting to you, I assure you that it is actually quite exciting because it focuses on how we can use information to understand how the world and business works and how to interpret facts. The course will enhance your data literacy, help you think critically, and improve your personal decision-making skills.

One way we can do this is by understanding the differences between quantitative and qualitative data and how they can be used to inform our choices.

Quantitative data is information that can be measured, such as numbers and statistics, while qualitative data is information that cannot be measured and is often expressed in words or other non-numerical forms.

Both forms of information are crucial for making good decisions. Without sufficient information, it can be difficult to evaluate the options and potential outcomes of a decision, leading to poor or uninformed choices. In general, the more information a decision-maker has and the faster and better the information can be used, the better they will be to make a sound decision.

The methods we discuss in this course will help you systematically gather information and make sense of it.

Enjoy the course!

2 Doing research

2.1 What is research

Research often involves exploring unknown territory and seeking out new information through methods such as attending conferences, conducting interviews and experiments, and reading related research. This process can lead to the discovery of valuable techniques or insights that address important issues in society or science. Zora Neale Hurston (2010) (see Figure 2.1) paraphrased it beautifully:

“Research is formalized curiosity. It is poking and prying with a purpose.” (Hurston, 2010)

Figure 2.1: Zora Neale Hurston, 1891-1960¹



Effective research is based on the principles of honesty, transparency and much more. A pithy yet profound quote from Scott Cunningham’s book (see Figure 2.2) sums up this idea :

“True scientists do not collect evidence in order to prove what they want to be true or what others want to believe. That is a form of deception and manipulation called propaganda, and propaganda is not science. Rather, scientific methodologies are devices for forming a particular kind of belief. Scientific methodologies allow us to accept unexpected, and sometimes undesirable, answers.” (Cunningham, 2021, p. 10)

2.2 Everybody can do research

Before I go into how empirical research can and should be conducted, I would like to assert that each of us is a researcher in some sense and that you don’t need a degree or a higher

¹Source: Photography is taken from Library of Congress: Prints & Photographs Division, Carl van Vechten Collection, Reproduction Number LC-USZ62-54231, see: <https://www.loc.gov/pictures/item/2004663047/>

²Source: Cunningham (2021)

³Source: Image by macrovector on Freepik, see: https://www.freepik.com/free-vector/kindergarten-set-isolated-icons-with-toys-characters-kids-practicing-with-teacher-playing-games-vector-illustration_26760074.htm

Figure 2.2: Causal Inference: The Mixtape²

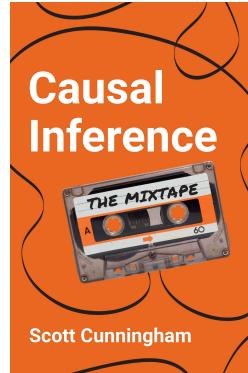


Figure 2.3: Children as little researcher³



2 Doing research

education to be a (good) researcher. Each of my four children (ages 2, 5, 6, and 8 (at the time of writing this)), for example, explores the world and learns something new every day. Even though none of my children is yet able to verify the novelty of their acquired knowledge and write it down in scientific form, I will claim that mine, like practically all children, are already little scientists. Why? Well, they explore unknown territory and search for information to discover new techniques that will make their lives pleasant, see Figure 2.3. Of course, they don't attend conferences or read journals to do this. They have never heard terms such as ontology, epistemology, axiology, or quantitative and qualitative methods. They are using methods that they have mastered for their age. They interview me, my wife and all other people around and they conduct experiments. For example, all my children liked to throw plates, cutlery, cups and alike from the table when they were about one year old. At first the throwing was just an accident, but they quickly found out that each throw was followed by a sound when the object touched the stone floor. My first son, in particular, took great delight in making these sounds. He threw everything within reach to the ground and giggled with joy at the clink he made when the object hit the ground. Perhaps he was also enjoying the attention he was getting from us parents through these actions. In any case, the behavior annoyed us. Wiping food scraps off the floor is not a nice thing to do. Unfortunately, at that time my son did not accept any argument to refrain from throwing. Neither a stern look nor a definite "no" helped to stop this behavior. Too great was the joy at the relationship he had figured out, which was, "I throw something off the table and it always clangs beautifully loud." So I started to do some research to figure out what I could do to stop him. The short answer I found can be summed up pretty well as "nothing". There is practically no good method to change the behavior without possibly negatively influencing his early childhood development. The reason is he did some research and we should not suppress that. Besides nature and material research he did social research: He found out that things fall to the ground (gravity), that things break and make different sounds (material research), and that other people notice him when he throws things (social research).

Once, when we were eating at a friend's house, my son (once again) threw everything off the table one after the other in unobserved moments. This time, however, it made no noise. The carpet under the table muffled everything. My son was irritated and at some point became really angry. Why? Well, his surely believed reality and his law "I throw something from the table and then it always clangs beautifully loud." was falsified. Soon he understood that his law only had to be adapted a little. It was then: "I throw something from the table and it clangs then beautifully loudly if a stone floor is under me." He repeated his experiments for a few more weeks, to check its validity. In the meantime he does other experiments trying to contribute to his own knowledge.

In general, the purpose of research is to find new knowledge or discover new ways to use existing knowledge in a creative way so as to generate new concepts, methodologies, inventions and understandings that -now or later- may be of some value for the human mankind. In simple terms, we aim to find something out. We aim to find a new law, a new relationship, a new insight. Or, we aim to challenge and revise existing insights on how the world works. You don't need a degree to do that. All you need is interest, open-mindedness, and a willingness to revise your ideas about how the world works. The latter is perhaps the most important skill you need to be a good researcher. Otherwise, one is a narrow-minded, and bigoted person who is too proud to follow up an insight with a change of mind.

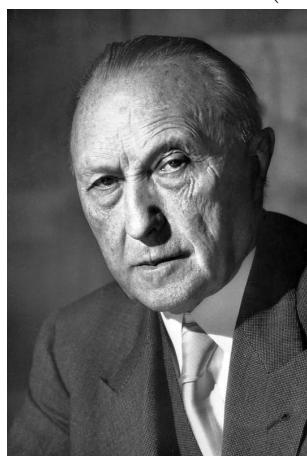
I myself have a quick and happy tendency to change my views because it is a statement of a fresh understanding. Here are two more quotes from Mr. Keynes (see Figure 2.4) and Mr. Adenauer (see Figure 2.5), two historically slightly more significant people than me that are along the same lines and should convince you that changing your mind is not a sign of weakness, but of strength. Especially in science, the willingness to change one's mind is essential.

Figure 2.4: John Maynard Keynes (1883-1946)⁴



“When the facts change, I change my mind. What do you do, sir?”¹

Figure 2.5: Konrad Adenauer (1876-1967)⁵



“What do I care about the rubbish I said yesterday? No one can stop me from getting smarter every day.” (“Was interessiert mich mein Geschwätz von gestern? ... es kann mich doch niemand daran hindern, jeden Tag klüger zu werden.”)²

2.3 It's difficult to do good research

Simply trying something and seeing what happens, like my children do, is a research method that relies on luck and chance. Before I go into more grown-up ways of doing research, I want to emphasize that the role of chance and serendipity in research is often downplayed and not acknowledged. The most well-known example of such research is the discovery of penicillin by Alexander Fleming (see Figure 2.6). In 1928, Fleming was studying the properties of staphylococcus bacteria when he noticed that a mold called Penicillium notatum had contaminated one of his bacterial cultures. He noticed that the mold seemed to be inhibiting the growth of the bacteria, and he began to investigate this further. Eventually, he was able to isolate and purify the active ingredient in the mold, which he named penicillin, and he discovered that it had powerful antibiotic properties. This discovery revolutionized the field of medicine and has saved countless lives.

⁴Source: Photography is public domain and stems from https://de.wikipedia.org/wiki/John_Maynard_Keynes#/media/Datei:Ke

⁵Source: This photography from 1952 is public domain and stems from the Bundesarchiv, B 145 Bild-F078072-0004, Katherine Young, CC BY-SA 3.0 DE.

⁶Source: Photography is public domain and stems from https://en.wikipedia.org/wiki/File:Synthetic_Production_of_Penicillin_T

Figure 2.6: Sir Alexander Fleming (1881-1955)⁶



Doing something on purpose and observing how things respond to the action can be considered a research strategy. Acting like a child or just waiting for something to happen by chance can also be considered a research strategy, and of course this can contribute greatly to knowledge. However, it are a naïve and poorly targeted strategies to conduct research. There are more grown-up research methods that are targeting more precisely the gaps in our knowledge and speed up innovation in the field where progress is desperately needed.

The processes of research and observation of phenomena should aim to maximize the probability of discovering new and intriguing findings. They should also ensure a high degree of confidence in the validity of our findings and reduce the likelihood that they will be disproved shortly afterwards. Transparency, scientific collaboration and open competition are crucial for efficient progress in science.

Take, for example, the scenario of a fatal disease. A naïve approach to finding a cure might be to try different things and observe who falls ill and who dies or is cured, hoping to stumble upon a cure through serendipitous observation. However, this method is unlikely to be effective or practical. A more promising strategy would be to systematically study the disease and openly communicate research plans before they are implemented. This avoids unnecessary efforts and costs and accelerates the achievement of results.

For example, a laboratory should first seek to isolate the causative virus or bacterium in order to be able to grow and study it outside the danger to humans. Once this is done, we need a precise plan on how we can use all the available knowledge to cure the disease, protect people from infection, or help them survive the disease. In short, we need a strategic way to conduct research, i.e., a research strategy or design.

A *research strategy* is a general plan for conducting a study and a *research design* is a detailed plan for conducting the study. These words are frequently used interchangeably. A research strategy depends on many things including the question, the resources available, the current state of knowledge, the ambitions, whether quantitative or qualitative data are used, and what is considered to be the criteria of good research.

Before discussing some research strategies that can provide reasonable answers to certain types of questions, we should clarify how to ask a research question and what qualifies a research question.

2.4 Asking questions like a good researcher

Unfortunately, there is no one research strategy that is appropriate for all questions and, what is worse, there is still controversy about what constitutes good research and how to properly ask a research question. In particular, this controversy takes place between researchers who use quantitative data and statistical methods and researchers who use qualitative data and methods.

Quantitative researchers are interested in both the causes of effects and the effects of causes. Experimental setups can allow to validate causes of effects and to measure the effects of causes. With observational data, however, it is often difficult to investigate the causes of effects. Thus, often quantitative research is more interested to quantify the effects of causes. Qualitative researchers also try to determine the causes of effects. However, their data analysis does rely less on statistical inference. A qualitative data set not necessarily requires (large) random samples or structured data (all the data that you can structure in a spreadsheet) in general, but allows to analyze selective and unstructured data (that is data in form of audio, video, text, images and alike). Qualitative research methods allow to classify these data into patterns or to interpret them in a meaningful way in order to arrive at results. Qualitative researchers are more concerned with the *why and how* of decision making and examine people's behavior, beliefs, perceptions of events, experiences, attitudes, interactions, and more in great depth.

In empirical research, inductive and deductive are two different approaches to reasoning. *Inductive reasoning* is a process of collecting data from various sources, such as interviews, surveys or observations, and then use this data to identify patterns, themes, or relationships that can form the basis of a new hypothesis or theory. The goal of these exploratory studies, is to generate new ideas or insights about a topic, rather than testing a specific hypothesis. *Deductive reasoning* is a process in which the researcher starts with a general theory or hypothesis with the goal to test a specific hypothesis or theory. In most cases, a combination of both inductive and deductive reasoning may be used to formulate the research question and to design the empirical identification strategy.

In what follows, however, we focus on the criteria for *good research* that are more commonly used in evaluating the quality of quantitative research.

Exercise 2.1. The Effect ch.1+2

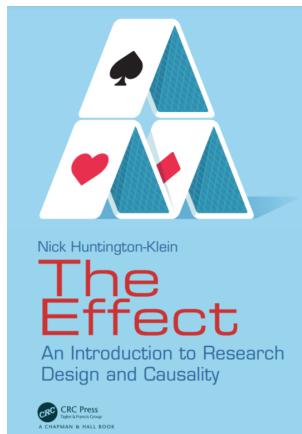
Read chapter 1 and 2 of Huntington-Klein (2022) and answer the questions below. The book (see Figure 2.7) is freely available at <https://theeffectbook.net> and here is the link to chapter 1: <https://theeffectbook.net/ch-TheDesignofResearch.html>

1. What is the main focus of the book the author is writing about?
 - a) Philosophy of science
 - b) Qualitative research methods
 - c) Empirical research and quantitative methods to identify and measure causal effects
 - d) Statistics

2. What is the main challenge faced by quantitative empirical research, according to the author?
 - a) Difficulty in obtaining accurate measurements
 - b) Difficulty in interpreting measurements
 - c) Difficulty in obtaining data that allows to answers the research question

⁷Source: Huntington-Klein (2022)

Figure 2.7: The Effect: An Introduction to Research Design and Causality⁷



- d) Difficulty in designing a research that gets a lot of attention
3. What is the author's main point about research questions?
- a) They should be well-defined, answerable, and understandable
 - b) They should be simple and easy to answer
 - c) They should be related to the world of traffic
 - d) They should be related to the field of quantum mechanics

Please find solution here: Solution 2.1.

2.5 Features of good research

In order to make you a competent researcher who does not have to wait for a lucky chance but has a clear strategy, let's discuss the criteria of a good research. Before I do that, however, I must make a disclaimer: there is a lack of consensus on what constitutes high-quality research in social sciences. In my experience, the practical benefits of such a tedious discussion are quite small. All I like to put forward is that I believe that all social science disciplines such as sociology, anthropology, psychology, economics, business administration, and education using quantitative methods agree that good research should be replicable, reproducible, transparent, reliable, and valid.

2.5.1 Reliability and validity

A research design is a plan to examine information in a systematic and controlled way so that the results of the research are *valid and reliable*.

Validity refers to the accuracy and truthfulness of research findings. In other words, if a study is valid, it should measure what it is intended to measure and produce results that are representative of the population being studied. Validity is important because it helps to ensure that the conclusions drawn from a study are supported by the data and are not based on flawed or biased methods.

Reliability refers to the consistency and stability of research findings. In other words, if a study is reliable, it should produce similar results if it is repeated using the same methods and

conditions. Reliability is important because it helps to ensure that the results of a study are not simply due to chance or random error.

Both reliability and validity are important considerations in research, and researchers strive to maximize both in their studies. However, it is important to note that it is often difficult to achieve both at the same time, and trade-offs may need to be made between the two.

 Note 1

A good research design should aim to minimize bias and maximize the reliability and validity of the research. It should also be appropriate for the research question being asked and the resources available to the researcher.

High reliability and low validity

An example of a study that has high reliability but low validity is a study that measures the weight of a group of people using a digital scale. If the scale is consistently accurate and produces the same weight measurements each time it is used, then the study has high reliability. However, if the scale is not calibrated correctly and produces inaccurate weight measurements, then the study has low validity.

Another example of a research design that has high reliability but low validity is a study that uses a highly reliable measurement tool, such as a standardized test, to measure a concept that is not directly related to the research question being asked. For example, a study that uses a standardized math test to measure students' critical thinking skills may have high reliability because the test is consistently accurate and produces similar scores each time it is administered. However, the study may have low validity because the math test is not an appropriate tool for measuring critical thinking skills. As a result, the results of the study may not be representative of the students' true critical thinking abilities.

High validity and low reliability

An example of a study that has high validity but low reliability is a study that asks people to self-report their eating habits. While the study may produce accurate and representative results about people's eating habits, the self-reported data may vary from person to person and may not be consistent over time. As a result, the study has high validity but low reliability.

Another example of a study that has high validity but low reliability is a study that uses a highly valid measurement tool, such as a survey, to measure a concept that is directly related to the research question being asked. However, the study may have low reliability because the survey is not administered consistently or the responses are not accurately recorded. For example, a study that uses a survey to measure students' attitudes towards school may have high validity because the survey is relevant to the research question and accurately measures the students' attitudes. However, if the survey is not administered consistently or the responses are not accurately recorded, the study may have low reliability. As a result, the results of the study may not be representative of the students' true attitudes towards school.

Trade-offs between reliability and validity

In research design, trade-offs may need to be made between reliability and validity. For example, a study that uses a highly reliable measurement tool may not be valid if the tool is not appropriate for the research question being asked. Similarly, a study that uses a highly valid measurement tool may not be reliable if the tool is prone to producing inconsistent results. As a result, researchers must carefully consider both reliability and validity when designing a study and make trade-offs as necessary to maximize the overall quality of the research.

2.5.2 Generalizability

Coming back to my little son who threw everything within reach to the ground and giggled with joy at the clink he made when the object hit the ground. He identified a cause-and-effect relationship through an experiment in a controlled environment. His law “I throw something off the table and it always clangs” worked in our home. To our regret, it was replicable and he really tried hard to falsify it. Moreover, his study was reasonable valid as his study design, conduct, and analysis could answer his research questions without bias (at least ignoring the other noises that his sibling and parents make coincidentally during his experiment). Scientist call this *internal validity*. However, he also found out that when he leaves our home, things are sometimes a bit different, for example, if there is a carpet under the table. Thus, his insights from our home findings can't be generalized to other contexts, at least not without further specifications. Scientist call this *external validity*.

Note 2

Internal validity examines whether the study design, conduct, and analysis answer the research questions without bias. External validity examines whether the study findings can be generalized to other contexts.

2.5.3 Replicability, reproducibility, transparency, and other criteria

It must be possible to repeat the research conducted for several reasons. For example, if you can repeat a study with slightly changed parameters, you are able to improve its external validity and show that the conclusions drawn are reliable. To be able to repeat a study, everything that is important for drawing a conclusion from the research has to be mentioned. This is what we call *transparency*. Moreover, everything in the study must have been done in such a way that we can check the results for truth. In the best case, it is possible to *reproduce* the results in the same way they were obtained in the study. Sometimes this is not possible because, for example, we can never really ask the same people again in a survey, and even if we found the same people, they would have gotten older and not be the same people as before. In such a case, it should at least be possible to *replicate* the research. This means that we can basically do the same thing in a setting that differs only in those things that we cannot avoid to be different. For example, by interviewing a group of people who match the people in the study to replicate them on all the important characteristics like age.

In an empirical quantitative research study, for example, the data and the code written to process the data and analyze it should be accessible to everyone.

In a qualitative study, all sources of information should be stated, and the circumstances leading to a conclusion should be fully explained. For example, all transcripts of interviews conducted should be made available. The researcher should provide rich and detailed descriptions of

the data and the context in which it was collected. Research should be provided with rich, nuanced, and multi-layered accounts of social phenomena by describing and interpreting the meanings, beliefs, and practices of the people being studied. That is known as *thick description*. Researchers typically employ a variety of methods such as participant observation, in-depth interviews, and document analysis, and they often use multiple sources of data to triangulate their findings. The goal is to provide a holistic and broad understanding of the phenomenon being studied, rather than a narrow view from the researcher's perspective.

There are some other criteria of good research that are worth mentioning:

Credibility

The research should be trustworthy and believable, and the researcher should provide detailed descriptions of the methods used to ensure transparency.

Reflective Practice

The researcher should engage in reflexive practice throughout the research process, which means to be critically aware of oneself, one's own assumptions, and one's own role in the research process.

Triangulation

The researcher should use multiple methods, sources, and perspectives to increase the credibility of the findings (also see *thick description* above).

Transferability

The conclusions drawn from looking at mostly unstructured data in qualitative research can hardly be generalized in a strict sense, as they depend crucially on the context of the object of study. For example, generalizability is essentially impossible in a qualitative case study, since everything depends on the specific situation of an individual, a company, or a group of people considered in the specific setting. This means that in a case study or interview, we may be looking at only a few or even a single observation that cannot be considered *representative* of the larger population, as generalizability does. Transferability, on the other hand, gives the reader the ability to transfer the findings into other contexts. The ability to transfer contextual findings to other cases is a goal of qualitative research, and the author of a study should attempt to offer the information in a way that allows the reader to transfer the findings to the setting or situation with which he or she is familiar.

2.6 The role of resources, data and ethics

There are several types of research designs, including experimental designs, quasi-experimental designs, and observational designs. Each of these designs took advantage of various empirical methods and statistical procedures. We will discuss some of them later on. The choice of research design, of course, should depend on the research question being asked, the resources available, and the type of data that is being collected. The research design should also take into account any ethical considerations that may be relevant to the research. The research

2 Doing research

design should be chosen so that it is well suited to answer the research question. For example, if one is interested in the question “Why do some people get sick with a certain disease and others do not?” then an observational study design to determine possible *causes of effect* may be appropriate. These identified potential causes should then be verified followed by an experimental study. Relatively, a statistical analysis should be used which would allow the *effects of causes* to be evaluated. The aim should be to identify necessary and sufficient circumstances to develop a disease. Also circumstances should be described that favor a disease.

If the question is a “how” question, for example, “How do parents feel when their child throws everything off the table?” then interviews might be an appropriate study design. If available resources such as time, funding, and staff are limited, you might also consider conducting an (online) survey in which parents are asked standardized questions about their feelings. In any way, the chosen research design must be feasible given the resources available.

In answering a question, a researcher should know, state, and discuss all the assumptions and unexamined beliefs that led him to his conclusion. However, since resources for conducting and explaining research are limited, special attention should be paid to what are called *critical assumptions*. These are assumptions that must be true in reality, otherwise the research is meaningless. Therefore, researchers should make great efforts to identify and validate these assumptions.

The type of data that is being collected is another important factor to consider when choosing a research design. Different types of data, such as quantitative data, qualitative data, or a combination of both, may require different methods of collection and analysis. For example, quantitative data, such as numerical data, can be collected through methods such as surveys and analyzed using statistical techniques, whereas qualitative data, such as interview transcripts, may require more interpretive methods of analysis.

Finally, the researcher should also take into account any ethical considerations that may be relevant to the research. For example, if the study involves human subjects, the researcher must ensure that the study is conducted in accordance with ethical principles such as informed consent and confidentiality. Additionally, the researcher should ensure that the potential benefits of the study outweigh any potential risks to the subjects.

Exercise 2.2. Features of research

1. Which of the following best defines reliability in research?
 - a) The extent to which a measurement tool produces consistent results
 - b) The extent to which a study’s results accurately reflect the concept being measured
 - c) The extent to which a study’s results can be generalized to other populations
 - d) The extent to which a study’s results are statistically significant
2. Which of the following best defines validity in research?
 - a) The extent to which a measurement tool produces consistent results
 - b) The extent to which a study’s results accurately reflect the concept being measured
 - c) The extent to which a study’s results can be generalized to other populations
 - d) The extent to which a study’s results are statistically significant
3. Which of the following is an example of a study with high reliability but low validity?
 - a) A study that uses a highly reliable measurement tool to measure a concept that is directly related to the research question being asked
 - b) A study that uses a highly valid measurement tool to measure a concept that is not directly related to the research question being asked

2 Doing research

- c) A study that uses a highly reliable measurement tool to measure a concept that is not directly related to the research question being asked
 - d) A study that uses a highly valid measurement tool to measure a concept that is directly related to the research question being asked
4. Which of the following is an example of a study with high validity but low reliability?
- a) A study that uses a highly reliable measurement tool to measure a concept that is directly related to the research question being asked
 - b) A study that uses a highly valid measurement tool to measure a concept that is not directly related to the research question being asked
 - c) A study that uses a highly reliable measurement tool to measure a concept that is not directly related to the research question being asked
 - d) A study that uses a highly valid measurement tool to measure a concept that is directly related to the research question being asked
5. What does internal validity examine in a study?
- a) The ability to replicate the study
 - b) The generalizability of the study's findings
 - c) Whether the study design, conduct, and analysis answer the research questions without bias
 - d) All of the above
6. What does external validity examine in a study?
- a) The ability to replicate the study
 - b) The generalizability of the study's findings
 - c) Whether the study design, conduct, and analysis answer the research questions without bias
 - d) None of the above
7. What is transparency in research?
- a) The ability to replicate a study
 - b) The generalizability of the study's findings
 - c) The availability and accessibility of the data and materials used in a study for others to review
 - d) The ethical considerations of the research
8. What are the different types of research design discussed in the text?
- a) Experimental designs, quasi-experimental designs, and observational designs
 - b) Experimental designs and descriptive designs
 - c) Quasi-experimental designs and observational designs
 - d) None of the above
9. Why is replicability important in a study?
- a) To be able to repeat a study with slightly changed parameters and thus improve the external validity
 - b) To be able to check the results of the study for truth.
 - c) To be able to reproduce the results in the same way they were obtained in the study
 - d) All of the above

Please find solution here: Solution 2.2.

Glossary

- **Generalizability:** The extent to which the results of a study can be applied to other populations or contexts.
- **Internal validity:** The degree to which a study's results can be attributed to the specific variables or factors being studied, and not to other extraneous factors.
- **External validity:** The degree to which a study's results can be generalized to other populations or contexts outside of the specific sample or setting of the study.
- **Quantitative data:** Data that can be measured and quantified.
- **Qualitative data:** Data that cannot be easily measured or quantified.
- **Quantitative research:** A research approach that uses statistical methods and experiments to determine the causes of effects, to quantify the effects of causes, or to describe data.
- **Qualitative research:** A research approach that uses unstructured data and methods to examine, for example, people's behavior, beliefs, and experiences in depth, rather than quantifying results.
- **Reflective Practice:** A form of self-evaluation used to analyze one's own thoughts and actions.
- **Reliability:** The consistency of a study's results to produce similar results when repeated.
- **Research design:** A detailed plan for conducting a study, frequently used interchangeably with research strategy.
- **Research method:** A procedure used to conduct a study or investigation to gain knowledge or understanding about a particular topic.
- **Research question:** A question or problem that a study aims to answer or solve.
- **Research strategy:** A general plan for conducting a study, frequently used interchangeably with research design.
- **Replicability:** The ability of a study to be repeated with new data.
- **Reproducibility:** The ability of a study to be repeated and produce the same results, often used interchangeably with replicability.
- **Serendipity:** The role of luck and unexpected events in research.
- **Thick Description:** A detailed narrative used to explain a situation and its context.
- **Credibility:** A quality criterion in qualitative research, which refers to confidence in the truth value of the data and interpretations of them.
- **Transparency:** The degree to which a study's methods and data are easily accessible and understandable to others, allowing for the study to be independently evaluated and replicated.
- **Triangulation:** A method used in qualitative research to verify the accuracy of data by combining multiple sources of information.
- **Validity:** The degree to which a study measures what it is intended to measure, and the extent to which the results of the study can be considered accurate and meaningful.
- **Structured data:** Data that can be easily organized and analyzed in a structured format, such as a spreadsheet.
- **Unstructured data:** Data that cannot be easily organized and analyzed in a structured format, such as text, images, and audio.

2.7 Solutions to exercises

Solution 2.1. Solution to exercise Exercise 2.1.

1. c), 2. c), 3. a)

Glossary

Solution 2.2. Solution to exercise Exercise 2.2

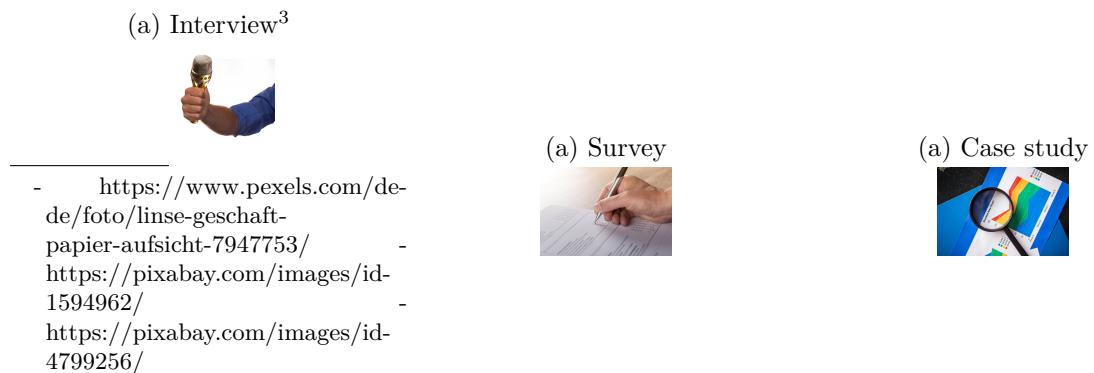
- 1. a), 2. b), 3. c), 4. d), 5. c), 6. b), 7. c), 8. a), 9. d)

3 Identification

In empirical research, *identification* refers to the process of establishing a clear and logical relationship between a cause and an effect. This involves demonstrating that the cause is responsible for the observed effect, and that there are no other factors that could potentially explain the effect. The goal of identification is to provide strong evidence that a particular factor is indeed the cause of a particular outcome, rather than simply coincidentally happen. In order to identify a cause-and-effect relationship, researchers can use experimental or non-experimental, i.e., observational data, or both. The next section discusses how to generate or collect this data.

3.1 Data acquisition

Figure 3.1: Three methods of data acquisition¹



As Figure 3.1 shows, there are several ways to get data which allows you to (hopefully) identify a cause-and-effect relationship:

3.1.1 Interviews

An interview is normally a one-on-one verbal conversation. Interviews are conducted to learn about the participants' experiences, perceptions, opinions, or motivations. The relationship between the interviewer and interviewee must be taken into account and other circumstances (place, time, face to face, email, etc.) should be taken into account. There are three types of interviews structured, semi-structured, and unstructured. *Structured interviews* use a set list of questions and hence are like a verbal surveys. In *unstructured interviews* the interviewer doesn't use predetermined questions but only a list of topics to address. Semi-structured interviews are the middle ground. Semi-structured interviews require the interviewer to have a list of questions and topics pre-prepared, which can be asked in different ways with different interviewee/s. Semi-structured interviews increase the flexibility and the responsiveness of the interview while

3 Identification

keeping the interview on track, increasing the reliability and credibility of the data. Semi-structured interviews are one of the most common interview techniques.

Structured interviews use a predetermined list of questions that must be asked in a specific order, improving the validity and trustworthiness of the data but lowering respondent response. Structured interviews resemble verbal questionnaires. In unstructured interviews, the interviewer has a planned list of subjects to cover but no predetermined interview questions. In exchange for less reliable data, this makes the interview more adaptable. Long-term field observation studies may employ unstructured interviews. The middle ground are interviews that are semi-structured. In semi-structured interviews, the interviewer must prepare a list of questions and themes that can be brought up in various ways with various interviewees.

Interviews allow you to address a cause-and-effect relationship fairly directly, and it can be a good idea to interview experts and ask some *why* and *how* questions to gather initial knowledge about a particular topic before further elaborating your research strategy. For example, I interviewed kindergarten teachers with many years of experience working with children, as well as other parents, to get information on how to solve the problem of my children throwing plates around the dining room. However, findings based on interviews are not very valid or reliable because the personal perceptions of both the interviewer and the interviewee can have an impact on the conclusions drawn. For example, I received very different tips and explanations because of the personal experiences of the people I interviewed. Unfortunately, I could not really ask my son why he was misbehaving. His vocabulary was too limited at the time, and even if he could speak, he would probably refuse to tell me the truth.

3.1.2 Surveys

In contrast to an interview a survey can be sent out to many different people. Surveys can be used to identify a cause-and-effect relationship by asking questions about both the cause and the effect and examining the responses. For example, if a researcher wanted to determine whether there is a relationship between a person's level of education and their income, they could conduct a survey asking participants about their education level and their income. If the data shows that participants with higher levels of education tend to have higher incomes, it suggests that education may be a cause of higher income. However, it is important to note that surveys can only establish a correlation between variables, but it is difficult to claim that correlations found through the survey imply a causal relationship. To establish a causal relationship, a researcher would need to use other methods, such as an experiment, to control for other potential factors that might influence the relationship that the respondent does not see.

3.1.3 Case studies

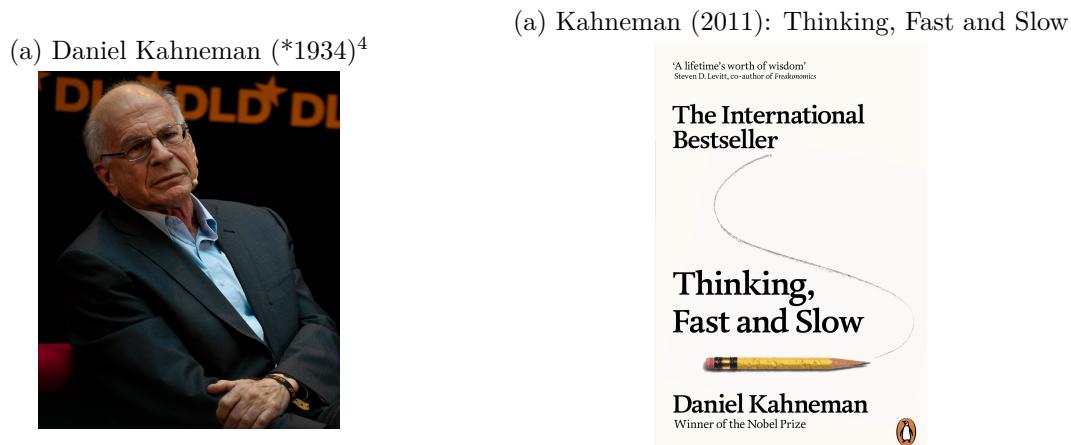
Case studies involve in-depth examination of a single case or a small number of cases in order to understand a particular phenomenon. Case studies can be conducted using both quantitative and qualitative methods, depending on the research question and the data being analyzed. While it is reasonable to find causal effects in the particular case, it is problematic to generalize the causal relationship. To establish a general causal relationship, a researcher would need to use other methods, such as an experiment, to control for other potential factors that might influence the relationship that the respondent does not see.

3.1.4 Experiments

One way to clearly identify a cause-and-effect relationship is through experiments, which involve manipulating the cause (the independent variable) and measuring the effect (the dependent variable) under controlled conditions (we will later on define precisely what is meant here). Experiments can be conducted using both quantitative and qualitative methods. Here are some examples:

- A medical study in which a new drug is tested on a group of patients, while a control group receives a placebo.
- An educational study in which a group of students is taught a new method of learning, while a control group is taught using the traditional method.
- An agricultural study in which a group of crops is treated with a new fertilization method, while a control group is not treated.
- A study to determine the effect of a new training program on employee productivity might involve randomly assigning employees to either a control group that does not receive the training, or an experimental group that does receive the training. By comparing the productivity of the two groups, the researchers can determine if the new training program had a causal effect on employee productivity.
- A study to determine the effect of a new advertising campaign on sales might involve randomly assigning different groups of customers to be exposed to different versions of the campaign. By comparing the sales of the different groups, the researchers can determine if the advertising campaign had a causal effect on sales.
- In *experimental economics*, experimental methods are used to study economic questions. In a lab-like environment data are collected to investigate the size of certain effects, to test the validity of economic theories, to illuminate market mechanisms, or to examine the decision making of people. Economic experiments usually motivates and rewards subjects with money. The overall goal is to mimic real-world incentives and investigate things that cannot be captured or identified in the field.
- In *behavioral economics*, laboratory experiments are also used to study decisions of individuals or institutions and to test economic theory. However, it is done with a focus on cognitive, psychological, emotional, cultural, and social factors.

Figure 3.5: Daniel Kahneman and his best selling book



In 2002 the Nobel Prize of Economics was awarded to Vernon L. Smith, I quote The Royal Swedish Academy of Sciences (2002), “for having established *laboratory experiments* as a tool

3 Identification

in empirical economic analysis, especially in the study of alternative market mechanisms” and Daniel Kahneman “for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty”.

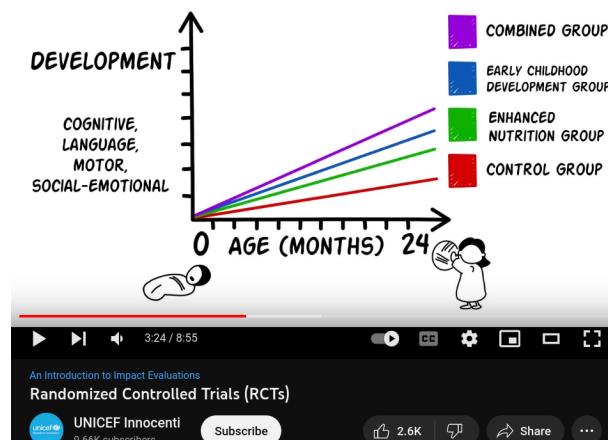
The strength of evidence from a controlled experiment is generally considered to be strong. However, the external validity, i.e., the generalizability, should be considered as well. External validity is sometimes low because effects that you can identify and measure in a lab are sometimes only of minor importance in the field.

There are different types of experiments:

Randomized controlled trials (RCTs) are a specific type of an experiment that involve randomly assigning participants to different treatment groups and comparing the outcomes of those groups. RCTs are often considered the gold standard of experimental research because they provide a high degree of control over extraneous variables and are less prone to bias.

For a better explanation and some great insights into what an RCT actually is, please watch the video produced by UNICEFInnocenti and published on the YouTube channel of UNICEF’s dedicated research center <https://youtu.be/Wy7qpJeoze>

Figure 3.8: Randomized Controlled Trials (RCTs)



Quasi-experiments involve the manipulation of an independent variable, but do not involve random assignment of participants to treatment groups. Quasi-experiments are less controlled than RCTs, but can still provide valuable insights into cause-and-effect relationships.

Natural experiments involve the observation of naturally occurring events or situations that provide an opportunity to study cause-and-effect relationships. Natural experiments are often used when it is not possible or ethical to manipulate variables experimentally.

In a **laboratory experiment**, researchers manipulate an independent variable and measure the effect on a dependent variable in a controlled laboratory setting. This allows for greater control over extraneous variables, but the results may not generalize to real-world situations.

In a **field experiment**, researchers manipulate an independent variable and measure the effect on a dependent variable in a natural setting, rather than in a laboratory. This allows researchers to study real-world phenomena, but it can be more difficult to control for extraneous variables.

Figure 3.9: Observational data²



3.1.5 Observational data

Observational data are data that had been observed before the research question was asked or being collected independently from the study. To understand how observational data can be used to constitute a causal relationship is a bit tricky because there is only one world and only one reality at a time. In other words, we usually miss a counterfactual which we can use for a comparison. Take, for example, the past COVID-19 pandemic, where you chose to be vaccinated or not. Regardless of what you chose, we will never find out what would have happened to you if you had chosen differently. Maybe you would have died, maybe you would have gotten more or less sick, or maybe you wouldn't have gotten sick at all. We don't know, and it's impossible to find out because it's impossible to observe the counterfactual outcomes. This makes it difficult to establish causality from observational data. However, ingenious minds have found reasonable procedures and methods to extract some level of knowledge from observational data that allows us to infer causal relationships from observational data where we cannot directly observe the counterfactual outcome. We will come back to these methods later on.

In the upcoming sections, however, we will discuss experimental research designs including *randomized controlled trials* (RCTs) which are considered to be the “gold standard for measuring the effect of an action” (Taddy, 2019, p. 128). RCTs can be used, for example, to study the effectiveness of drugs by observing people randomly assigned to three groups, one taking the pill (or treatment), a second receiving a placebo, and a third taking nothing. If the first group responds in any way differently than the other groups, the drug has an effect. Before explaining an RCT in more detail, we need to be clear about the fundamental problem of causal inference. This will be discussed in the following.

Exercise 3.1. Methods used in economic research

Read Paldam (2021) which is freely available, see <https://doi.org/10.1515/econ-2021-0003> and answer the following questions:

- a) List the eight types of research methods described in the paper and provide the description found in the paper.
- b) Read the following statements and discuss whether they are true or not, and if the latter, correct them:
 - i) The annual production of research papers in economics in the year 2017 has reached about 100 papers in top journals, and about 1,400 papers in the group of good journals. The production has grown with 3.3% per year, and thus it has doubled the last twenty years.
 - ii) The upward trend in publication must be due to the large increase in the importance of publications for the careers of researchers, which has greatly increased the production of papers. There has also been a large increase in the number of researches,

²The picture is free to use under the Pixabay license, see: <https://pixabay.com/images/id-5029286/>

3 Identification

- but as citations are increasingly skewed toward the top journals it has not increased demand for papers correspondingly.
- iii) Four trends are significant: The fall in theoretical papers and the rise in classical papers. There is also a rise in the share of statistical method and event studies. It is surprising that there is no trend in the number of experimental studies.
 - iv) Book reviews have dropped to less than 1/3. Perhaps, it also indicates that economists read fewer books than they used to. Journals have increasingly come to use smaller fonts and larger pages, allowing more words per page. The journals from North-Holland Elsevier have managed to cram almost two old pages into one new one. This makes it easier to publish papers, while they become harder to read.
 - v) About 50% of papers in the sample considered in ? belong to the economic theory class, about 6% are experimental studies, and about 43% are empirical studies based on data inference.
 - vi) The papers in economic theory have increased from 33.6% to 59.5% – this is the largest change for any of the eight subgroups. It is highly significant in the trend test.
- c) Explain what is meant with *theory fatigue* and discuss the reasons that lead to that fatigue.
d) According to Paldam (2021): What factors contribute to the immediate relevance of research papers for policymakers?

Please find solution to the exercise in the appendix (Solution 4.1).

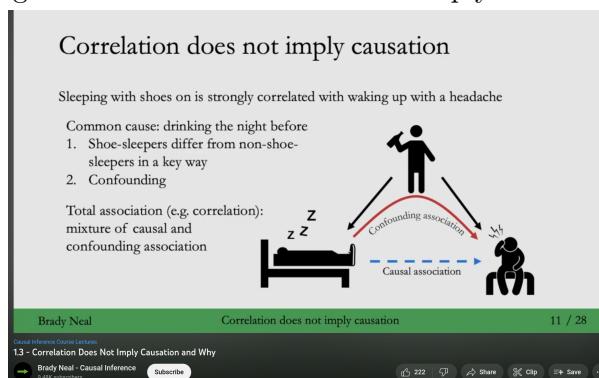
3.2 Correlation does not imply causation

Correlation refers to a statistical relationship between two variables, where one variable tends to increase or decrease as the other variable also increases or decreases. However, just because two variables are correlated does not necessarily mean that one variable causes the other. This is known as the *correlation does not imply causation* principle.

For example, it may be observed that the number of storks in a particular area is correlated with the birth rate of babies in that area. However, this does not mean that the presence of storks causes an increase in the birth rate. It is possible that both the number of storks and the number of babies born are influenced by other factors, such as the overall population density or economic conditions in the area.

Therefore, it is important to carefully consider all possible explanations (confounders) for a correlation and to use empirical evidence to determine the true cause-and-effect relationship between variables.

Figure 3.10: Correlation does not imply causation³



Exercise 3.2. Watch the video of Brady Neal's lecture *Correlation Does Not Imply Causation and Why*. Alternatively, you can read chapter 1.3 of his lecture notes (Neal, 2020) which you find here.

3.3 The fundamental problem of causal inference

Figure 3.11: (#fig:cunbook) The book cover of “Causal Inference: The Mixtape?” by Scott Cunningham (2021)



Cunningham (2021, ch. 1.3): *“It is my firm belief, which I will emphasize over and over in this book, that without prior knowledge, estimated causal effects are rarely, if ever, believable. Prior knowledge is required in order to justify any claim of a causal finding. And economic theory also highlights why causal inference is necessarily a thorny task.”*

As Cunningham (2021) explains in his book, it is very hard to claim causality. In the following section, I will paraphrase briefly two aspects why it is so difficult to claim to have found a causal effect. One reason for that is, that it is rather difficult to find or generate the right data and to use them properly so that the result is not biased. First, I will discuss Simpson’s Paradox as an example how easy it is to interpret the data falsely. It will provide an idea on how difficult it is to analyze observational data meaningful and that we need to have a theory when looking on data. Above that, we should try to challenge the assumptions on which the theory is build on. After that I will briefly discuss the fundamental problem of causal inference as a problem of missing counterfactual data.

::: {.cell}

Causal inference ch.1

Before going on with the content in these notes, please read chapter 1 (Introduction) of @Cu

1. What are some common misconceptions about causality that the author addresses in chapter

1. What is the role of randomization in causal inference, as described in the book?

Please find solution to the exercise [in the appendix.] (#sol:causalinf1)

:::

³Picture is taken from https://youtu.be/DFPm_a-uJM

3 Identification

Simpsons Paradox

![(\#fig:waitingroom) Discrimination^[The photography is public domain and stems from the]

Discrimination is bad. Whenever we see it, we should try to find ways to overcome it.

Figure 3.12: (#fig:berkley) Graph taken from Bickel et al. (1975, p. 403)

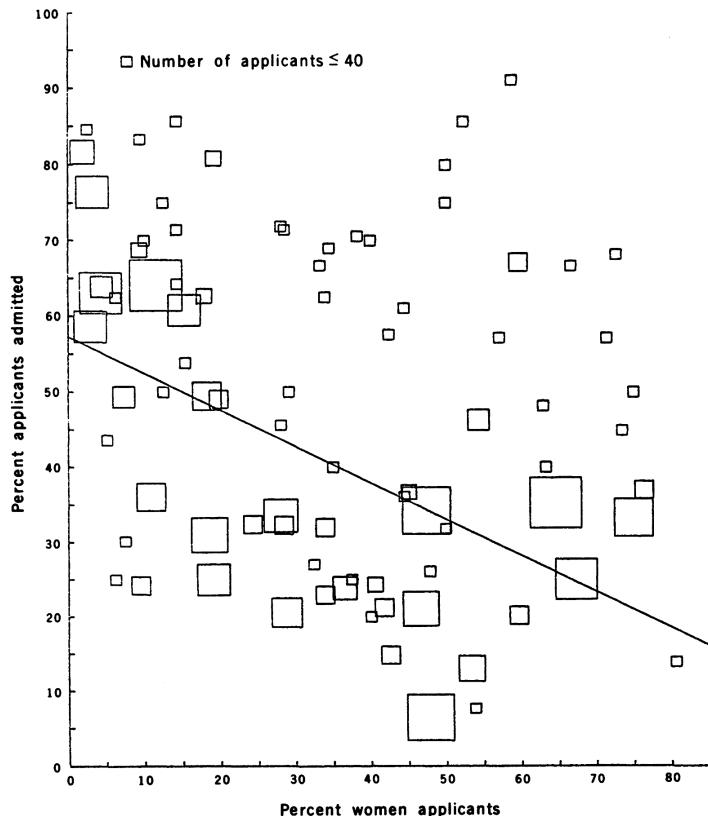


Fig. 1. Proportion of applicants that are women plotted against proportion of applicants admitted, in 85 departments. Size of box indicates relative number of applicants to the department.

Here you can read the summary of their infamous study:

"Examination of aggregate data on graduate admissions to the University of California, Berkeley, for fall 1973 shows a clear but misleading pattern of bias against female applicants. Examination of the disaggregated data reveals few decision-making units that show statistically significant departures from expected frequencies of female admissions, and about as many units appear to favor women as to favor men. If the data are properly pooled, taking into account the autonomy of departmental decision making, thus correcting for the tendency of women to apply to graduate departments that are more difficult for applicants of either sex to enter, there is a small but statistically significant bias in favor of women. The graduate departments that are easier to enter tend to be those that require more mathematics in the undergraduate preparatory curriculum. The bias in the aggregated data stems not from any pattern of discrimination on the part of admissions committees, which seem quite fair on the whole, but apparently from prior screening at earlier levels of the educational system. Women are shunted by their socialization and education

3 Identification

toward fields of graduate study that are generally more crowded, less productive of completed degrees, and less well funded, and that frequently offer poorer professional employment prospects.”

Graduate admissions

Read the first three pages of @Bickel1975Sex, i.e., pages 398–400, and answer the following questions:

- a) Describe the two assumptions that must be true in order to prove that UC Berkeley discriminated against women.
- b) Table 1, shows that 277 fewer women and 277 more men were admitted than we would have expected under non-discriminatory admissions.
- c) Explain the analogy with fish that illustrates the danger of pooling data.

Please find solution to the exercise [in the appendix.] (#sol:graduateadmission)

Simpson's Paradox

1. What is Simpson's Paradox?
 - a) A phenomenon in which the direction of a relationship between two variables changes when combined across groups.
 - b) A phenomenon in which the strength of a relationship between two variables changes when combined across groups.
 - c) The phenomenon where correlation appears to be present in different groups of data, but disappears when the groups are combined.
1. What is a potential cause of Simpson's Paradox?
 - a) Differences in the variance of the two variables
 - b) Differences in the correlation of the two variables
 - c) Confounding variables
 - d) Differences in the sample size of the two variables

Please find solution to the exercise [here] (#sol:simpsonpara)

Rubin causal model

Keele (2015, p. 314): “*An identification analysis identifies the assumptions needed for statistical estimates to be given a causal interpretation.*”

If we are interested in the causal effect of a certain treatment on an outcome, we need to make causal assumptions.

For example, if we want to measure the effect of a vaccine we never can have a person who is vaccinated and not vaccinated at the same time. Formally, we have either $Y_i(1)$ or $Y_i(0)$, where Y_i denotes the effect/output of individual i in case of being vaccinated (1) and not vaccinated (0).

Thus, the so-called individual treatment effect (ITE) does not exist for person i :

$$ITE_i = Y_i(1) - Y_i(0)$$

The Rubin Causal Model, also known as the potential outcomes framework, is a statistical framework for causal inference.

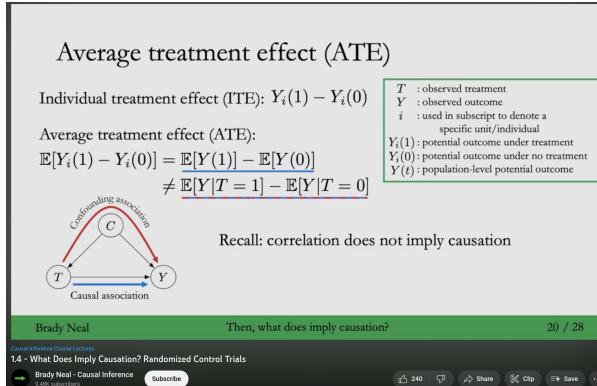
3 Identification

Table @ref(tab:tab21) is taken from Neal (2020) and shows some example data to illustrate that the fundamental problem of causal inference is actually a missing data problem. The Model goes back to Donald B. Rubin (*1943) a statistician and is now a widely used method for causal inference. The basic premise of the Rubin Causal Model is that for each individual in a study, there are two potential outcomes: the outcome that would occur if the individual were exposed to a certain treatment or intervention (the “treatment group”), and the outcome that would occur if the individual were not exposed to that treatment (the “control group”). The key idea is that these potential outcomes can be used to infer causality by comparing the outcomes between the treatment and control groups even if we do not have a full set of data.

Table 3.1: (#tab:tab21) Example data to illustrate that the fundamental problem of causal inference

i	T	Y	Y(1)	Y(0)	Y(1)-Y(0)
1	0	0	?	0	?
2	1	1	1	?	?
3	1	0	0	?	?
4	0	0	?	0	?
5	0	1	?	1	?
6	1	1	1	?	?

Figure 3.13: (#fig:nealrct) Average treatment effect (ATE)⁴



Watch the video of Brady Neal's lecture [[What Does Imply Causation? Randomized Control Trials](#)]

Under certain assumptions, the Rubin Causal Model allows for the estimation of the Average Treatment Effect (ATE), which is the difference in the expected outcomes between the treatment and control groups, given by the formula:

$$ATE \triangleq E[Y(1) - Y(0)]$$

There are several ways to estimate the ATE under the Rubin Causal Model some of which will be part of this course. Applied correctly, it can provide valuable insights into causality and inform decision making. However, the Rubin Causal Model has its limitations and assumptions that need to be met in order for the inferences to be valid.

⁴Picture is taken from <https://youtu.be/gGaWU8XEoGk>

To get the average treatment effect (ATE) we can take the average of the individual treatment effects (ITE):

$$ATE \triangleq \mathbb{E}[Y(1) - Y(0)] = \underbrace{\mathbb{E}[Y_i(1) - Y_i(0)]}_{ITE} (\#eq : feq) \quad (3.1)$$

3.4 Its difficult to overcome the fundamental problem

In the following we will discuss conditions that need to hold in order to empirically draw causal conclusions from the ATE without bias. This is important because equation @ref(eq:feq) does very often not hold when using observational data.

3.4.1 Ignorability

Referring to table @ref(tab:tab21), Brady Neal (2020) wrote:

“what makes it valid to calculate the ATE by taking the average of the $Y(0)$ column, ignoring the question marks, and subtracting that from the average of the $Y(1)$ column, ignoring the question marks?” This ignoring of the question marks (missing data) is known as ignorability. Assuming ignorability is like ignoring how people ended up selecting the treatment they selected and just assuming they were randomly assigned their treatment” (Neal, 2020, p. 9)

Randomized controlled trials (RCTs) are characterized by randomly assigning individuals to different treatment groups and comparing the outcomes of those groups. Thus, they are built on the assumption of *ignorability*

$$(Y(1), Y(0)) \perp T$$

which allows to write the ATE as follows:

$$\mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] = \mathbb{E}[Y(1) | T = 1] - \mathbb{E}[Y(0) | T = 0] \quad (3.2)$$

$$= \mathbb{E}[Y | T = 1] - \mathbb{E}[Y | T = 0].(\#eq : feq2) \quad (3.3)$$

Another perspective on this assumption is the concept of exchangeability. Exchangeability refers to the idea that the treatment groups can be interchanged such that if they were switched, the new treatment group would have the same outcomes as the old treatment group, and the new control group would have the same outcomes as the old control group.

3.4.2 Unconfoundedness

While randomized controlled trials (RCTs) assume the concept of ignorability, most observational data present challenges in drawing causal conclusions due to the presence of confounding factors that affect both (1) the likelihood of individuals being part of the treatment group and (2) the observed outcome. For instance, regional factors can affect both the number of storks and the number of babies born in a region. These factors are typically referred to as confounders, which we discussed in section @ref(cornotcaus) as having the potential to create the illusion of a causal impact where none exists. However, empirical methods are available to control for these confounders and prevent the violation of the ignorability assumption.

Treatment effects

Read sections 2.1 and 2.3 of @Neal2020Introduction.

1. What is the individual treatment effect (ITE)?
1. What is the average treatment effect (ATE)?
1. How is the ATE calculated?
1. Can the ATE be used to determine the effect of a treatment on an individual level?
1. What are some potential sources of bias when estimating the ATE?

Please find solution to the exercise [in the appendix.] (#sol:treatmenteffects)

3.5 Statistical control requires causal justification

Read Wysocki et al. (2022) which is freely available here.

Scientific research revolves around challenging our own views and findings. A good researcher does not merely present their results; instead, they engage in discussions about potential limitations and pitfalls to draw valid conclusions. Engaging in polemics goes against the essence of good research. We should not conceal potential weaknesses in our scientific strategy or empirical approach; rather, we should emphasize their existence. Even if this disappoints individuals seeking easy answers, it is crucial to acknowledge these limitations. The Catalogue of Bias is an excellent resource that provides insight into various potential pitfalls and challenges encountered during research, which may sometimes be difficult to completely rule out.

3.6 Solutions to excercises

Solution 3.1. Solution to exercise Exercise 3.1

a) List the eight types of research methods described in the paper and provide the description found in the paper

1. **Economic theory:** Papers are where the main content is the development of a theoretical model. The ideal theory paper presents a (simple) new model that recasts the way we look at something important.
2. **Statistical technique, incl. forecasting** Papers reporting new estimators and tests are published in a handful of specialized journals in econometrics and mathematical statistics. Some papers compare estimators on actual data sets. If the demonstration of a methodological improvement is the main feature of the paper, it belongs to this subgroup, but if the economic interpretation is the main point of the paper, it belongs to the classical empirical studies or newer techniques group.
3. **Surveys, incl. meta-studies** When the literature in a certain field becomes substantial, it normally presents a motley picture with an amazing variation, especially when different schools exist in the field. They are of two types, where the second type is still rare:

1. **Assessed surveys** where the author reads the papers and assesses what the most reliable results are. Such assessments require judgment that is often quite difficult to distinguish from priors, even for the author of the survey.
2. **Meta-studies** which are quantitative surveys of estimates of parameters claimed to be the same. These types of studies have two levels: The basic level collects and codes the estimates and studies their distribution. This is a rather objective exercise where results seem to replicate rather well. The second level analyzes the variation between the results. This is less objective.
4. **Experiments in laboratories** Most of these experiments take place in a laboratory, where the subjects communicate with a computer, giving a controlled, but artificial, environment. A number of subjects are told a (more or less abstract) story and paid to react in either of a number of possible ways. A great deal of ingenuity has gone into the construction of such experiments and in the methods used to analyze the results. Lab experiments do allow studies of behavior that are hard to analyze in any other way, and they frequently show sides of human behavior that are difficult to rationalize by economic theory. However, everything is artificial – even the payment while participants usually receive real money for participation and their performance. In some cases, the stories told are so elaborate and abstract that framing must be a substantial risk. In addition, experiments cost money, which limits the number of subjects. It is also worth pointing to the difference between expressive and real behavior. It is typically much cheaper for the subject to ‘express’ nice behavior in a lab than to be nice in the real world.
5. **Event studies (field experiments and natural experiments)** Event studies are studies of real world experiments. They are of two types:
 1. **Field experiments** analyze cases where some people get a certain treatment and others do not. The ‘gold standard’ for such experiments is double blind random sampling, where everything (but the result!) is announced in advance. Experiments with humans require permission from the relevant authorities, and the experiment takes time too. In the process, things may happen that compromise the strict rules of the standard. Controlled experiments are expensive, as they require a team of researchers.
 2. **Natural experiments** take advantage of a discontinuity in the environment, i.e., the period before and after an (unpredicted) change of a law, an earth-quake, etc. Methods have been developed to find the effect of the discontinuity. Often, such studies look like classical empirical studies with many controls that may that may or may not belong. Thus, the problems discussed under the classic empirical studies also apply here.
6. **Descriptive, deductions from data** In a descriptive study, researcher use an existing sample and hence, they have no control over the data generating process as it is usually the case with experiments. Descriptive studies are deductive. The researcher describes the data aiming at finding structures that tell a story, which can be interpreted. The findings may call for a formal test. If one clean test follows from the description, the paper can still be classified as a descriptive study. If more elaborate regression analysis is used, however, it can also be classified as a classical empirical study. Descriptive studies often contain a great deal of theory. Some descriptive studies present a new data set developed by the author to analyze a debated issue. In these cases, it is often possible to make a clean test, so to the extent that biases sneak in, they are hidden in the details of the assessments made when the data are compiled.

7. **Classical empirical studies** Typically have three steps: It starts by a theory, which is developed into an operational model. Then it presents the data set, and finally it runs regressions. The significance levels of the t-ratios on the coefficient estimated assume that the regression is the first meeting of the estimation model and the data. In practice, we all know that this is rarely the case. The classical method is often just a presentation technique. The great virtue of the method is that it can be applied to real problems outside academia. The relevance comes with a price: The method is quite flexible as many choices have to be made, and they often give different results. Preferences and interests, may affect these choices.
8. **Newer techniques** Partly as a reaction to the problems of classical empirical methods, the last 3–4 decades have seen a whole set of newer empirical techniques. They include different types of vector autoregression (VAR)⁵, Bayesian techniques, causality and co-integration tests, Kalman filters, hazard functions, etc. The main reason for the lack of success for the new empirics is that it is quite bulky to report a careful set of co-integration tests or VARs, for example, and they often show results that are far from useful in the sense that they are unclear and difficult to interpret.

b) Read the following statements and discuss whether they are true or not, and if the latter, correct them:

Statements i) and vi) are false, all others are correct.

- i) *The numbers are wrong:* The annual production of research papers in economics in the year 2017 has now reached about **1,000** papers in top journals, and about **14,000** papers in the group of good journals. The production has grown with 3.3% per year, and thus it has doubled the last twenty years.
- ii) *Statement is correct:* The upward trend in publication must be due to the large increase in the importance of publications for the careers of researchers, which has greatly increased the production of papers. There has also been a large increase in the number of researches, but as citations are increasingly skewed toward the top journals it has not increased demand for papers correspondingly.
- iii) *Statement is correct:* Four trends are significant: The fall in theoretical papers and the rise in classical papers. There is also a rise in the share of statistical method and event studies. It is surprising that there is no trend in the number of experimental studies.
- iv) *Statement is correct:* Book reviews have dropped to less than 1/3. Perhaps, it also indicates that economists read fewer books than they used to. Journals have increasingly come to use smaller fonts and larger pages, allowing more words per page. The journals from North-Holland Elsevier have managed to cram almost two old pages into one new one. This makes it easier to publish papers, while they become harder to read.
- v) *Statement is correct:* About 50% of papers in the sample considered in ? belong to the economic theory class, about 6% are experimental studies, and about 43% are empirical studies based on data inference.
- vi) *Economic theory is not on the rise:* The papers in economic theory have **dropped from 59.5% to 33.6%** – this is the largest change for any of the eight subgroups. It is highly significant in the trend test.

3 Identification

- vii) “Theory fatigue” is a term used to describe the decreasing attractiveness of theoretical research among journals, researchers and political decision-makers. This trend goes hand in hand with the increasing importance of empirical research. Policy makers are finding it increasingly difficult to engage with variations of existing theoretical models, and researchers often struggle to systematically summarize the findings of theoretical work, making it difficult to draw definitive conclusions on specific topics. In addition, theoretical work can be unconvincing to a wider audience that must rely on the reasonableness of complex and sometimes unrealistic assumptions. The credibility of theoretical research often depends on how realistic the initial assumptions are and how plausible the conclusions are. If neither aspect is grounded in reality, there is a danger that the research becomes an abstract exercise that provides new insights into the real world, but which are difficult to communicate to the layperson.
- viii) A research paper that policymakers find appealing typically offers estimates of a crucial effect that decision-makers outside of academia are keen to understand. Papers that target policymakers should put an emphasis on distilling the core findings into a short executive summary tailored for decision-makers, facilitating their understanding and application of the research insights.

References

- Békés, G., & Kézdi, G. (2021). *Data analysis for business, economics, and policy*. Cambridge University Press.
- Bickel, P. J., Hammel, E. A., & O'Connell, J. W. (1975). Sex bias in graduate admissions: Data from berkeley: Measuring bias is harder than is usually assumed, and the evidence is sometimes contrary to expectation. *Science*, 187(4175), 398–404. <https://doi.org/10.1126/science.187.4175.398>
- Cunningham, S. (2021). *Causal inference: The mixtape*. Accessed January 30, 2023; Yale University Press. <https://mixtape.scunning.com/>
- Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. Accessed January 30, 2023; CRC Press. <https://theeffectbook.net>
- Hurston, Z. N. (2010). *Dust tracks on a road*. HarperCollins.
- Illowsky, B., & Dean, S. (2018). *Introductory statistics*. Openstax.
- Kahneman, D. (2011). *Thinking, fast and slow*. Penguin.
- Keele, L. (2015). The statistics of causal inference: A view from political methodology. *Political Analysis*, 23(3), 313–335.
- Neal, B. (2020). *Introduction to causal inference from a machine learning perspective: Course lecture notes*. Accessed January 30, 2023. https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf
- Paldam, M. (2021). Methods used in economic research: An empirical study of trends and levels. *Economics*, 15(1), 28–42.
- Taddy, M. (2019). *Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions* (1st ed.). McGraw Hill Education.
- The Royal Swedish Academy of Sciences. (2002). Press Release. <https://www.nobelprize.org/prizes/economic-sciences/2002/press-release/>
- Weymar, P. (1955). *Konrad Adenauer: Die autorisierte Biographie*. Kindler.
- Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, 5(2). <https://doi.org/10.1177/25152459221095823>

4 Appendix

Solution 4.1. Solution to exercise Exercise 3.1

a) List the eight types of research methods described in the paper and provide the description found in the paper

1. **Economic theory:** Papers are where the main content is the development of a theoretical model. The ideal theory paper presents a (simple) new model that recasts the way we look at something important.
2. **Statistical technique, incl. forecasting** Papers reporting new estimators and tests are published in a handful of specialized journals in econometrics and mathematical statistics. Some papers compare estimators on actual data sets. If the demonstration of a methodological improvement is the main feature of the paper, it belongs to this subgroup, but if the economic interpretation is the main point of the paper, it belongs to the classical empirical studies or newer techniques group.
3. **Surveys, incl. meta-studies** When the literature in a certain field becomes substantial, it normally presents a motley picture with an amazing variation, especially when different schools exist in the field. They are of two types, where the second type is still rare:
 1. **Assessed surveys** where the author reads the papers and assesses what the most reliable results are. Such assessments require judgment that is often quite difficult to distinguish from priors, even for the author of the survey.
 2. **Meta-studies** which are quantitative surveys of estimates of parameters claimed to be the same. These types of studies have two levels: The basic level collects and codes the estimates and studies their distribution. This is a rather objective exercise where results seem to replicate rather well. The second level analyzes the variation between the results. This is less objective.
4. **Experiments in laboratories** Most of these experiments take place in a laboratory, where the subjects communicate with a computer, giving a controlled, but artificial, environment. A number of subjects are told a (more or less abstract) story and paid to react in either of a number of possible ways. A great deal of ingenuity has gone into the construction of such experiments and in the methods used to analyze the results. Lab experiments do allow studies of behavior that are hard to analyze in any other way, and they frequently show sides of human behavior that are difficult to rationalize by economic theory. However, everything is artificial – even the payment while participants usually receive real money for participation and their performance. In some cases, the stories told are so elaborate and abstract that framing must be a substantial risk. In addition, experiments cost money, which limits the number of subjects. It is also worth pointing to the difference between expressive and real behavior. It is typically much cheaper for the subject to ‘express’ nice behavior in a lab than to be nice in the real world.
5. **Event studies (field experiments and natural experiments)** Event studies are studies of real world experiments. They are of two types:

4 Appendix

1. **Field experiments** analyze cases where some people get a certain treatment and others do not. The ‘gold standard’ for such experiments is double blind random sampling, where everything (but the result!) is announced in advance. Experiments with humans require permission from the relevant authorities, and the experiment takes time too. In the process, things may happen that compromise the strict rules of the standard. Controlled experiments are expensive, as they require a team of researchers.
2. **Natural experiments** take advantage of a discontinuity in the environment, i.e., the period before and after an (unpredicted) change of a law, an earth-quake, etc. Methods have been developed to find the effect of the discontinuity. Often, such studies look like classical empirical studies with many controls that may or may not belong. Thus, the problems discussed under the classic empirical studies also apply here.
6. **Descriptive, deductions from data** In a descriptive study, researcher use an existing sample and hence, they have no control over the data generating process as it is usually the case with experiments. Descriptive studies are deductive. The researcher describes the data aiming at finding structures that tell a story, which can be interpreted. The findings may call for a formal test. If one clean test follows from the description, the paper can still be classified as a descriptive study. If more elaborate regression analysis is used, however, it can also be classified as a classical empirical study. Descriptive studies often contain a great deal of theory. Some descriptive studies present a new data set developed by the author to analyze a debated issue. In these cases, it is often possible to make a clean test, so to the extent that biases sneak in, they are hidden in the details of the assessments made when the data are compiled.
7. **Classical empirical studies** Typically have three steps: It starts by a theory, which is developed into an operational model. Then it presents the data set, and finally it runs regressions. The significance levels of the t-ratios on the coefficient estimated assume that the regression is the first meeting of the estimation model and the data. In practice, we all know that this is rarely the case. The classical method is often just a presentation technique. The great virtue of the method is that it can be applied to real problems outside academia. The relevance comes with a price: The method is quite flexible as many choices have to be made, and they often give different results. Preferences and interests, may affect these choices.
8. **Newer techniques** Partly as a reaction to the problems of classical empirical methods, the last 3–4 decades have seen a whole set of newer empirical techniques. They include different types of vector autoregression (VAR)⁶, Bayesian techniques, causality and co-integration tests, Kalman filters, hazard functions, etc. The main reason for the lack of success for the new empirics is that it is quite bulky to report a careful set of co-integration tests or VARs, for example, and they often show results that are far from useful in the sense that they are unclear and difficult to interpret.

b) Read the following statements and discuss whether they are true or not, and if the latter, correct them:

Statements i) and vi) are false, all others are correct.

- i) *The numbers are wrong:* The annual production of research papers in economics in the year 2017 has now reached about **1,000** papers in top journals, and about **14,000** papers

4 Appendix

in the group of good journals. The production has grown with 3.3% per year, and thus it has doubled the last twenty years.

- ii) *Statement is correct:* The upward trend in publication must be due to the large increase in the importance of publications for the careers of researchers, which has greatly increased the production of papers. There has also been a large increase in the number of researches, but as citations are increasingly skewed toward the top journals it has not increased demand for papers correspondingly.
- iii) *Statement is correct:* Four trends are significant: The fall in theoretical papers and the rise in classical papers. There is also a rise in the share of statistical method and event studies. It is surprising that there is no trend in the number of experimental studies.
- iv) *Statement is correct:* Book reviews have dropped to less than 1/3. Perhaps, it also indicates that economists read fewer books than they used to. Journals have increasingly come to use smaller fonts and larger pages, allowing more words per page. The journals from North-Holland Elsevier have managed to cram almost two old pages into one new one. This makes it easier to publish papers, while they become harder to read.
- v) *Statement is correct:* About 50% of papers in the sample considered in ? belong to the economic theory class, about 6% are experimental studies, and about 43% are empirical studies based on data inference.
- vi) *Economic theory is not on the rise:* The papers in economic theory have **dropped from 59.5% to 33.6%** – this is the largest change for any of the eight subgroups. It is highly significant in the trend test.
- vii) “Theory fatigue” is a term used to describe the decreasing attractiveness of theoretical research among journals, researchers and political decision-makers. This trend goes hand in hand with the increasing importance of empirical research. Policy makers are finding it increasingly difficult to engage with variations of existing theoretical models, and researchers often struggle to systematically summarize the findings of theoretical work, making it difficult to draw definitive conclusions on specific topics. In addition, theoretical work can be unconvincing to a wider audience that must rely on the reasonableness of complex and sometimes unrealistic assumptions. The credibility of theoretical research often depends on how realistic the initial assumptions are and how plausible the conclusions are. If neither aspect is grounded in reality, there is a danger that the research becomes an abstract exercise that provides new insights into the real world, but which are difficult to communicate to the layperson.
- viii) A research paper that policymakers find appealing typically offers estimates of a crucial effect that decision-makers outside of academia are keen to understand. Papers that target policymakers should put an emphasis on distilling the core findings into a short executive summary tailored for decision-makers, facilitating their understanding and application of the research insights.

Solution 4.2. Solution to exercise ?@exr-causalinf1

1. Some common misconceptions about causality that the author addresses in chapter 1 include confusion between correlation and causality, and the belief that observational studies cannot (hardly) establish causality without prior knowledge. He says that human beings “engaging in optimal behavior are the main reason correlations almost never reveal causal relationships, because rarely are human beings acting randomly” which is crucial for identifying causal effects.

4 Appendix

2. The role of randomization in causal inference, as described in the book, is that it helps to control for confounding variables and allows for the estimation of causal effects.

Solution 4.3. Solution to exercise ?@exr-graduateadmission

- a) Assumption 1 is that in any given discipline male and female applicants do not differ in respect of their intelligence, skill, qualifications, promise, or other attribute deemed legitimately pertinent to their acceptance as students. It is precisely this assumption that makes the study of “sex bias” meaningful, for if we did not hold it any differences in acceptance of applicants by sex could be attributed to differences in their qualifications, promise as scholars, and so on. (...) Assumption 2 is that the sex ratios of applicants to the various fields of graduate study are not importantly associated with any other factors in admission. (Bickel et al., 1975, p. 398)
- b) Expectations were taken based on the overall acceptance rate of about 0.416 and multiplied by the total observed numbers of applicants admitted and rejected. For example: $(3738 + 4704) \cdot 0.41666 \approx 3460$ and $(3738 + 4704) \cdot (1 - 0.41666) \approx 4981$. Taking the difference of these two measures gives the number to be explained.
- c) The analogy is explained on page 400:

“Picture a fishnet with two different mesh sizes. A school of fish, all of identical size (assumption 1), swim toward the net and seek to pass. The female fish all try to get through the small mesh, while the male fish all try to get through the large mesh. On the other side of the net all the fish are male. Assumption 2 said that the sex of the fish had no relation to the size of the mesh they tried to get through. It is false.”

The UC Berkley case is just one of many examples to illustrate that uniformity of group assignment of individuals is a necessary condition to ensure that pooling of data does not lead to misleading conclusions when using statistics. The phenomenon of obtaining different results depending on whether one considers the data pooled or unpooled is often referred to as the *Simpson Paradox*.

Solution 4.4. Solution to exercise (exr:simpsonpara?)

1. a), 2. c) and d)

Solution 4.5. Solution to exercise ?@exr-treatmenteffects

1. The individual treatment effect (ITE) is a measure of the effect of a treatment or intervention on an individual level. It represents the difference in the outcome for an individual who receives the treatment versus the outcome for that same individual if they had not received the treatment.
2. The average treatment effect (ATE) is a measure of the difference in the expected outcomes between a treatment group and a control group. It represents the overall effect of a treatment on the population as a whole.
3. The ATE is calculated by taking the difference between the average outcome for the treatment group and the average outcome for the control group.
4. No, the ATE is a population-level measure and cannot be used to determine the effect of a treatment on an individual level. To determine the effect of a treatment on an individual level, you would need to use techniques such as propensity score matching or instrumental variables.

4 Appendix

5. Some potential sources of bias when estimating the ATE include selection bias, measurement bias, and unobserved confounding variables. To mitigate these biases, researchers may use randomization or other advanced statistical techniques such as propensity score matching or instrumental variables to control for these potential sources of bias.

Solution 4.6. Solution to exercise ?@exr-randomsampling

Random sampling is a sampling procedure by which each member of a population has an equal chance of being included in the sample. Random sampling ensures a representative sample. There are several types of random sampling. In simple random sampling, not only each item in the population but each sample has an equal probability of being picked. In systematic sampling, items are selected from the population at uniform intervals of time, order, or space (as in picking every one-hundredth name from a telephone directory). Systematic sampling can be biased easily, such as, for example, when the amount of household garbage is measured on Mondays (which includes the weekend garbage). In stratified and cluster sampling, the population is divided into strata (such as age groups) and clusters (such as blocks of a city) and then a proportionate number of elements is picked at random from each stratum and cluster. Stratified sampling is used when the variations within each stratum are small in relation to the variations between strata. Cluster sampling is used when the opposite is the case. In what follows, we assume simple random sampling. Sampling can be from a finite population (as in picking cards from a deck without replacement) or from an infinite population (as in picking parts produced by a continuous process or cards from a deck with replacement). The larger the sample gets, the closer we get to the population and hence, we reduce the bias of having a non-randomly selected sample.

Solution 4.7. Solution to exercise (exr_summarystatistics?)

Metric	1	2
Mode	21	3
Median	29.5	3
P20	21	2
Range	71	5
IQR	25	1.5
Arithmetic mean	37	3.5
σ^2	485.33	1.5591
σ	22.030	1.248621
COV	0.5954	2.4027

Solution 4.8. Solution to exercise ?@exr-guessstat

varlabel	a	b	c	d	e
Variance	2 (4.6666)	4 (466.66)	5 (800.33)	1 (4.6666)	3 (466.66)
COV	1 (.02160)	3 (.21602)	5 (.56580)	4 (.54006)	2 (.02160)
Mean	3 (100)	4 (100)	2 (50)	1 (4)	5 (1000)
Median	3 (100)	4 (100)	2 (50)	1 (4)	5 (1000)
Range	1 (6)	4 (60)	5 (98)	2 (6)	3 (60)
SD	1 (2.1602)	4 (21.602)	5 (28.290)	2 (2.1602)	3 (21.602)

4 Appendix

Solution 4.9. Solution to exercise ?@exr-destatexcel

Mittelwert	56
Standardfehler	11,4697670227235
Modus	0
Median	55
Erstes Quartil	42,5
Drittes Quartil	85
Varianz	1315,555555555556
Standardabweichung	36,2705880232945
Kurtosis	-0,731538352420953
Schräge	-0,417051115341008
Bereich	100
Minimum	0
Maximum	100
Summe	560
Anzahl	10

Solution 4.10. Solution to exercise ?@exr-DiD

Answers: 1. b), 2. a), 3. c)