

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Denkfallen und Manipulationen in der Statistik

verfasst von | submitted by

Daniel Malinowski BEd

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 199 510 520 02

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Lehramt Sek (AB) Unterrichtsfach
Geographie und wirtschaftliche Bildung
Unterrichtsfach Mathematik

Betreut von | Supervisor:

Dipl.-Ing. Dr.techn. Volkmar Putz

Abstrakt

In dieser Masterarbeit werden verschiedenste Fehleinschätzungen und Manipulationen in der Statistik thematisiert. Aufbauend auf grundlegenden Formeln der Statistik werden einfache Beispiele aus dem Alltag, aber auch komplexere Paradoxa näher beleuchtet.

Wie können alle Daten darauf hinweisen, dass ein Medikament wirksam ist, während sich bei genauerer Betrachtung derselben Daten herausstellt, dass das Medikament bei keinem Geschlecht wirksam ist? In dieser Arbeit wird die Frage beforscht, wie Erhebungen und Grafiken manipuliert und bestehende Daten missinterpretiert werden können. Die Auswahl und der Ausschluss bestimmter Fakten können zu verschiedensten gegenseitigen Argumenten führen, während die Mathematik stichhaltig bleibt. Solche und viele weitere Fragen und Aspekte werden im Rahmen dieser Arbeit untersucht und es wird aufgezeigt, dass in beinahe jedem Aspekt der Begriffe und Formeln zur Statistik eine Gefahr der Manipulation und/oder Denkfalle liegt, sogar (oder erst recht) auf einem Unterstufen- bis hin zum Maturalevel.

Das Ziel ist es somit, allen Lesenden möglichst einfach und verständlich die verschiedenen Themenbereiche der Statistik und deren Wichtigkeit derartig näher zu bringen, dass diese nicht nur verstanden, sondern auch selbstständig im Alltag mit einem kritischen Blick angewandt werden können.

Abstract

This master's thesis deals with a wide range of misinterpretations and manipulations in statistics. Building on basic statistical formulas, simple examples from everyday life as well as more complex paradoxes are examined in more detail.

How can all data indicate that a medication is effective while a closer look at the same data reveals that the medication is not effective for any gender? This paper explores the question of how surveys and graphs can be manipulated and existing data can be misinterpreted. The selection and exclusion of certain facts can lead to various mutual arguments, while the mathematics remains valid. These and many other questions and aspects are examined in this work and it is shown that there is a danger of manipulation and/or mental traps in almost every aspect of the terms and formulas of statistics, even (or even more so) at a lower school level up to GCSE level.

The aim is thus to make the various topics of statistics and their importance as simple and understandable as possible for all readers so that they can not only understand them but also apply them independently in everyday life with a critical eye.

Inhaltsverzeichnis

1	Einleitung	1
2	Geschichte der Statistik	2
3	Grundbegriffe der Statistik	3
3.1	Typisierung von Daten	5
3.2	Skalierung von Daten	6
4	Auswahl und Ausschluss bestimmter Daten	10
4.1	Woher wir unsere Informationen beziehen	10
4.2	Was alle denken, muss stimmen	11
4.3	Das Beste/Schlechteste wovon?	12
4.4	Welche Merkmalsträger:innen zählen?	12
4.5	Welche Merkmalsausprägungen zählen?	12
4.6	Mögliches Fehlwissen nutzen	12
4.7	Zahlen ohne Relation	13
4.8	Zahlen ohne Zeitangabe	14
4.9	Exakte Zahlen als Bluff	14
4.10	Ungenau ermittelte Zahlen	14
4.11	Große Zahlen klein portioniert und umgekehrt	15
4.12	'Statistische' Aussagen ohne Bedingung	15
4.13	Das 'statistisch kleinere Übel'	16
4.14	'Statistische' Aussagen mit falschen Merkmalsträger:innen	17
4.15	Aus Groß wird Klein und umgekehrt	17

4.16	'Statistische' Aussagen unter Ausschluss verschiedener Faktoren	18
4.17	Zusammenfassung	20
4.18	Exkurs 1: Rundungen	21
4.19	Exkurs 2: 'Postfaktisch'	21
5	Häufigkeiten und Änderungen	22
5.1	Absolute und relative Häufigkeit bzw. Änderung	22
5.2	Kleines n	26
6	Prozente	28
6.1	'Um das Doppelte' vs. 'Auf das Doppelte'	28
6.2	Mehrfache relative Änderung	28
6.3	Grundwert wovon?	30
6.4	Falscher Grundwert	30
6.5	(Ver-)Tauschen von Grundwert und Anteil	32
6.6	Prozentpunkte	34
6.7	Geometrisches Mittel	36
6.8	Zusammenfassende fiktive Diskussion	38
7	Lagemaße	41
7.1	Arithmetisches Mittel	41
7.2	Arithmetisches Mittel von Skalen	43
7.3	Harmonisches Mittel	44
7.4	Median	46
7.5	Modus	47
7.6	Quantil/Perzentil	48

8	Streumaße	51
8.1	Spannweite	51
8.2	Varianz vs. durchschnittliche Abweichung	52
9	Korrelation vs. Kausalität	54
9.1	Verwechseln von Ursache und Wirkung	56
9.2	Regression	57
9.3	Zufällige zeitliche Korrelation	59
9.4	Subjektive Kausalität	60
9.5	Selbsterfüllende Korrelation	61
9.6	Ergänzung zur Kriminalität von Personen mit Migrationshintergrund und ausländischer Herkunft	61
10	Statische Prognosen und Extrapolation	63
10.1	Mathematisches Modell vs. Realität	63
10.2	Gleichbleibende Grenzen von Kategorien	64
10.3	Später weniger, jetzt mehr	64
11	Darstellungen von Statistiken	66
12	Bekannte Paradoxa in der Statistik	75
12.1	Will-Rogers-Phänomen	75
12.2	Aufteilungsparadoxon	78
12.3	Alabama-Paradoxon	80
12.4	Braess'sches Paradoxon	82
12.5	Populationsdynamik	88

12.6 Benford's Law	91
12.7 Pareto-Prinzip	93
12.8 Simpson-Paradoxon	94
12.9 Exkurs: Gender Pay Day	100
13 (Eigene) Datenerhebungen	103
13.1 Eigene Erinnerungen als Quelle	103
13.2 Versuchsplanung	106
13.3 Erhebungsarten	107
13.4 Qualität und gleiche Bedingungen	109
13.5 Mehrfach gezählte Merkmalsträger:innen	110
13.6 Verwechseln der Merkmalsträger:innen	111
13.7 Manipulieren der Merkmalsträger:innen	112
13.8 Wenn 'geheime' Absichten einer Datenerhebung erkannt werden	116
13.9 Floor und ceiling effect	117
13.10 Konfidenzintervalle	117
13.11 Hypothesentest	119
14 Fazit	125

1 Einleitung

Eine vertraute Situation: der eine Freund oder die eine Freundin schickt über Social Media eine Grafik, auf der ein Sachverhalt äußerst überzeugend und mit einer scheinbar unbestreitbaren Eindeutigkeit 'bewiesen' wird. Je komplizierter, umso überzeugender: wichtig ist die Aussage in der Titelzeile in Kombination mit einer ansprechenden Grafik. Dass dabei jedoch ein langer Prozess von der Hypothesenformulierung über Datenerhebung und Datenauswertung bis hin zur Interpretation sowie grafischen Darstellung des untersuchten Sachverhalts dahintersteht, wird oft übersehen. Doch können entlang dieses 'Weges' einige Fehler unterlaufen, diese müssen nicht einmal beabsichtigt sein - in den schlimmsten Fällen wird dieser Prozess sogar stark abgekürzt und/oder gar unprofessionell durchgeführt.

Das Ziel dieser Masterarbeit ist es daher, ein Bewusstsein für die alltäglichen Gefahrenquellen der Manipulationen sowie Denkfallen zu schaffen und vor allem dabei auch Nicht-Fachkundigen ein Repertoire an simplen Werkzeugen in der Flut an Statistiken mitzugeben. Beim Lesen dieser Masterarbeit werden Sie also in eine Welt geführt, in der vieles (auf den ersten Blick) 'Offensichtliche' sich als Manipulation oder gar als Irrtum entpuppt. Nicht zuletzt wird auch auf bekannte Paradoxa eingegangen, die sich mit Trugschlüssen in der Statistik beschäftigen.

Anspruch dieser Masterarbeit ist es explizit nicht, sämtliche Statistiken als falsch anzuprangern. Wie der Titel jedoch schon verrät, liegt der Fokus dieser Arbeit auf Denkfallen und Manipulationen. Die Beispiele, die dabei herangezogen werden, sind breitgefächert und deren Auswahl und Fehler-Erklärungen explizit nicht als politische Stellungnahmen des Autors zu verstehen.

Diese Arbeit bietet die Möglichkeit, in weitgehend einfachen und umgangssprachlichen Erklärungen einen Blick hinter die Kulissen der Statistik zu werfen und somit eine geschulte Brille zu gewinnen, die sich von bloßen Titelzeilen und Grafiken nicht vereinnahmen lässt, ohne vorher einen genaueren Blick darauf geworfen zu haben. Im Umkehrschluss wird auch ausdrücklich davon abgeraten, diese Manipulationsweisen zum eigenen Zweck anzuwenden, um andere Personen zum eigenen Vorteil zu beeinflussen.

Doch wie werden wir denn im Alltag durch Statistiken beeinflusst oder was missinterpretieren wir denn überhaupt? Um dies genauer zu betrachten, benötigen wir ein bestimmtes Grundwissen, auf welches sich in dieser Masterarbeit immer wieder bezogen wird.

2 Geschichte der Statistik

Zunächst noch zur Geschichte der Statistik, diese hatte schon frühe Beginne. Statistik kam bereits vor Christi Geburt vor: schätzungsweise 3800 Jahre vor Christus wurden Tonscherben im alten Babylon gefunden, welche auf Volkszählungen hinweisen. Solche amtlichen Statistiken gibt es bis heute. Könige und Kaiser ließen zählen oder auch schätzen, wie viele Soldaten es im Reich gab oder wie viel Einnahmen bei der Steuererhebung zu erwarten waren. Im Mittelalter gab es dann kaum systematische Erhebungen - in Klöstern und Kirchen wurde vereinzelt Register geführt, unter anderem bei Eheschließungen (vgl. [39]).

In der Neuzeit taten sich in der Statistik neue Türen auf. Im 16. Jahrhundert wurde beispielsweise empirisch festgestellt, dass ein idealer Würfel (also ein Würfel mit sechs gleichgroßen Seitenflächen) essenziell für ein faires Würfeln sei (vgl. [34], S. 10).

Im 17. Jahrhundert wurde die Statistik ein Universitätsfach und entwickelte sich im Rahmen der Wissenschaft immer weiter: einerseits die Mathematisierung (Tabellenstatistik, politische Arithmetik) und 'Moralstatistik' (Merkwürdigkeiten im Sinne der Sittenbilder, Eindrucksschilderung etc.), welche bis zum 19. Jahrhundert immer stärker an inhaltlicher Bestimmung verlor und schlussendlich als Hilfswissenschaft für Soziologie, Psychologie, Politikwissenschaften und dergleichen zu einem Zweig der angewandten Mathematik wurde. Die 'alte' Statistik, die anfänglich hauptsächlich als politisches Instrument diente, entwickelte sich im Laufe der Neuzeit immer weiter und breitete sich auch auf die Wahrscheinlichkeitsrechnung aus, als unter anderem Laplace die Kenntnisse der Statistik systematisch anwendet (die Laplace-Formel wird von allen Maturant:innen einmal verwendet worden sein) (vgl. [56], S. 82).

Den Entwicklungen der vergangenen Jahrhunderte haben wir es zu verdanken, heute das Thema Statistik relativ einfach und intuitiv in den Grundzügen verstehen zu können - stellen Sie sich vor, wie aufwändig es gewesen sein musste, die Welt davon zu überzeugen, dass ein Würfel mit unterschiedlich großen Flächen (selbst bei kleinen Abweichungen) statistische Unterschiede hervorbringt oder wie statistische Abweichungen zu erklären sind. Heute dürfen wir auf all dem aufbauen und müssen uns gleichzeitig in Acht nehmen, nicht zu leichtgläubig in das Thema zu stolpern. Um Schritt für Schritt unser Basiswissen in der Statistik aufzubauen, gehen wir im nächsten Kapitel die Grundbegriffe durch.

3 Grundbegriffe der Statistik

In diesem Kapitel werden die grundlegendsten Begriffe in der Statistik und deren Ziel betrachtet, angelehnt an [8] (S. 5 ff).

Ziel der Statistik

Ziel der Statistik ist es, bestimmte Merkmale der Realität der Welt zu erfassen und durch Zahlen zu beschreiben. Diese Zahlen sollen oft zu sogenannten Kennzahlen verdichtet werden, was mit einem gewissen Informationsverlust einhergeht - es bringt aber den Vorteil, solche Kennzahlen miteinander vergleichen zu können.

Merkmalsträger:innen

Merkmalsträger:innen sind bestimmte Objekte, die bezüglich bestimmter Eigenschaften gewisse Gemeinsamkeiten aufweisen, weshalb sie zusammen untersucht werden. Dabei werden die restlichen Eigenschaften der Wirklichkeit, die nicht näher untersucht werden wollen, abgegrenzt.

Diese Merkmalsträger:innen können sowohl sachlich als auch zeitlich und räumlich betrachtet werden:

- Sachlich:
Die Aspekte der Merkmalsträger:innen, die betrachtet werden.
- Zeitlich:
Der Beobachtungszeitraum wird angegeben.
- Räumlich:
Oft werden die Merkmalsträger:innen, die sich innerhalb eines bestimmten Gebiets/Areals befinden, untersucht.

Ein Beispiel für eine Abgrenzung von Merkmalsträger:innen: es soll im Jahr 2024 laufend das monatliche Netto-Einkommen der Bewohner:innen eines Wiener Studierendenwohnheims erfragt werden.

- Sachlich: Bewohner:innen
- Zeitlich: Kalenderjahr 2024
- Räumlich: Studierendenwohnheim

Merkmalsausprägung

Die Ausprägung eines Merkmals ist der jeweilige Wert x_i eines Merkmals X bei einem oder einer bestimmten Merkmalsträger:in i .

Auch hierzu ein Beispiel: alle Bewohner:innen aus dem Beispiel oberhalb sind die Merkmalsträger:innen und werden in der Abfrage mit Nummern (z.B. Max Mustermann mit der Nummer $i = 24$) gekennzeichnet. Das bei allen Merkmalsträger:innen untersuchte Merkmal X ist die Höhe des Einkommens.

Daraus ergibt sich die Merkmalsausprägung x_{24} (monatliches Netto-Einkommen von Max Mustermann), die angibt, welchen Wert das Merkmal X (monatliches Netto-Einkommen) bei dem Merkmalsträger $i = 24$ (Max Mustermann) annimmt.

Urliste

Dies ist die Sammlung/Liste sämtlicher Merkmalsausprägungen x_i aller Merkmalsträger:innen.

Es kann durchaus vorkommen, dass mehrere Merkmalsträger:innen den selben Wert (Merkmalsausprägung) aufweisen. Die Merkmalsträger:innen mit jeweils gleichen Ausprägungen kann man auch zusammenfassen, wodurch sich Häufigkeitsverteilungen ergeben, beispielsweise, wenn mehrere Personen jeweils dasselbe monatliche Einkommen haben.

Grundgesamtheit

Alle Merkmalsträger:innen werden untersucht, die nach der zeitlichen, räumlichen und sachlichen Abgrenzung noch in Frage kommen, ergeben die Grundgesamtheit. Oftmals stehen jedoch nicht die zeitlichen, strukturellen und finanziellen Mittel zur Verfügung, um alle Merkmalsträger:innen der Grundgesamtheit zu untersuchen (beispielsweise alle Einwohner:innen Niederösterreichs).

Stichprobe

Wie eben festgestellt, können oft nicht alle Merkmalsträger:innen untersucht werden (z.B. bei politischen Umfragen zu Wahlen in Österreich). Daher muss man mit einem Teil der Merkmalsträger:innen vorlieb nehmen. In diesem Fall spricht man von einer Stichprobe.

Oft unterscheiden sich Berechnungen je nachdem, ob es sich um eine Stichprobe oder Grundgesamtheit handelt.

Ein weiteres, weniger offensichtliches Beispiel für den Unterschied zwischen Stichprobe und Grundgesamtheit:

Alle Schüler:innen der 2B des Schuljahres 2023/2024 werden auf mathematische Kenntnisse untersucht. Wenn lediglich die mathematischen Fähigkeiten der 2B untersucht und darüber Aussagen getätigt werden sollen, so handelt es sich bei der 2B um eine Grundgesamtheit (vorausgesetzt, alle Schüler:innen der Klasse sind anwesend). Wenn jedoch aufgrund der 2B Aussagen über alle 2. Klassen der Schule getroffen werden sollen, so ergeben die Schüler:innen der 2B eine Stichprobe - wobei solche Schlussfolgerungen auf die Grundgesamtheit gefährlich sein können - zum Beispiel bei unterschiedlichen Mathematik-Lehrkräften und somit unterschiedlichen Prioritäten und Unterrichtsstilen in den 2. Klassen - abgesehen von den unterschiedlichen Schüler:innen.

Daher sollte man sich bemühen, bei einer Stichprobe immer eine möglichst heterogene Auswahl an Merkmalsträger:innen untersuchen zu können. Dazu später mehr, wenn es um das Aufzeichnen der Daten geht - zunächst beschäftigen wir uns mehr mit bereits vorhandenen Daten.

3.1 Typisierung von Daten

Daten sind im Wesentlichen 'beobachtete Ausprägungen eines Merkmals'. Dabei gibt es die Typisierung von diskreten und stetigen Merkmalen (vgl. [52], S. 5 f).

Diskrete Merkmale

Ein Merkmal ist diskret, wenn es abzählbar unendlich viele oder endlich viele Ausprägungen geben kann. Umgangssprachlich gesagt bedeutet 'abzählbar unendlich viele oder endlich viele' einfach nur, dass die Anzahl an Ausprägungen eindeutig zählbar ist, also dass die Ausprägung exakt bestimmbar ist - beispielsweise die Anzahl der Reservierungen in einem Restaurant, die Anzahl der Studiensemester von Absolvent:innen oder die Nationalität respektive der Familienstand.

Diese Merkmale sind eindeutig bestimmbar und es gibt nichts 'dazwischen', also gibt es 20 oder 21 Reservierungen, aber nicht 20,7214... Reservierungen - somit handelt es sich um ein diskretes Merkmal.

Stetige Merkmale

Ein Merkmal ist stetig, falls jeder beliebige Wert in einem Intervall 'getroffen' werden respektive 'vorkommen' kann, zum Beispiel jeder beliebige Wert in den reellen Zahlen. Die Ausprägungen sind dabei nicht abzählbar, sie werden durch Messvorgänge näherungsweise ermittelt.

Stellen Sie sich eine Zahlengerade vor: auf dieser Zahlengerade befinden sich Messwerte zum Tagesmittel der Lufttemperatur. Dabei gibt es die Werte 11°C und 12°C. Theoretisch können unendlich viele Werte zwischen 11°C und 12°C auftreten: 11,92°C, 11,921°C, 11,9211111°C - also können zwischen zwei Werten noch immer unendlich viele Werte getroffen werden - wobei dies meistens durch die Qualität der Messgeräte limitiert wird, welche lediglich Näherungswerte angeben. Weitere Beispiele für stetige Merkmale sind die Entfernung zwischen Arbeitsplatz und Wohnort oder das Alter (nicht schlicht in Jahren gedacht, sondern bis in Millisekunden und noch genauer).

3.2 Skalierung von Daten

Für unsere Daten gibt es auch verschiedene Skalierungen (vgl. [8], S. 8 ff).

Nominalskaliert

Wenn Merkmale nicht als messbare Zahlenwerte vorliegen respektive nicht in eine sinnvolle auf-/absteigende Reihenfolge gebracht werden können, sind sie nominalskaliert. Dabei kann man mit den Daten meist nicht viel mehr machen, als die Häufigkeit bestimmter Ausprägungen zu zählen - ein arithmetisches Mittel zu bilden ist hierbei wenig sinnvoll - beispielsweise bei der Haarfarbe oder dem Wohnort.

Ordinalskaliert/rangskaliert

Wenn die Merkmale nach der Größe ihrer Ausprägungen sortiert werden (aber zwischen verschiedenen Werten das Rechnen mit den Grundrechenarten wenig sinnvoll ist), spricht man von ordinalskalierten Daten. Dabei ist es wichtig, zu betonen, dass in dieser Kategorie die Abstände zwischen den Werten kein klares Verhältnis widerspiegeln - Schulnoten können von 1 bis 5 (respektive von 'Sehr gut' bis 'Nicht genügend') sortiert werden, aber ein Zweier ist nicht doppelt so gut wie ein Vierer. Er ist lediglich zwei Notengrade besser, wobei auch die Differenz zwischen Zweier und Vierer nicht dieselbe ist wie zwischen Einser und Dreier. Ein geeigneter Mittelwert wäre der Median (dazu später mehr).

Metrisch

Bei messbaren Merkmalsausprägungen, bei denen auch die Grundrechenarten sinnvoll angewandt werden können, geht es um metrische Skalierung. Auch diese muss jedoch unterschieden werden:

Verhältnisskaliert:

Wenn Differenzen und Verhältnisse zwischen Messwerten sinnvoll ermittelt werden

können und ein Nullpunkt existiert, liegen verhältnisskalierte Daten vor. Diese können beispielsweise Längen von Bauteilen oder die Lebensdauer eines Produktes sein: 24 cm sind doppelt so lang wie 12 cm und 0 cm ist theoretisch der Nullpunkt. Ein Produkt, welches zehn Jahre hält, hat eine halb so lange Nutzungsdauer als ein anderes Gerät mit 20 Jahren und 0 Jahre (oder 0 Sekunden) sind der Nullpunkt.

Es gelten also folgende Aussagen:

- Verhältnisse zwischen Werten/Ausprägungen können gebildet werden ('dreimal so groß wie', ...)
- Abstände zwischen Werten/Ausprägungen können gebildet werden ('um acht Einheiten kleiner' ...)
- Es gibt einen natürlichen Nullpunkt.

Intervallskaliert:

Tatsächlich gibt es metrische Daten, die nicht verhältnisskaliert sind. Da es keinen natürlichen Nullpunkt gibt, kann auch kein sinnvolles Verhältnis gebildet werden - während die Abstände weiterhin Sinn ergeben.

Das praktischste Beispiel hierfür ist der Vergleich zwischen den Temperatureinheiten 'Grad Celsius' und 'Kelvin' - eine Einheit steht für die verhältnisskalierte Skala und die andere für die intervallskalierte Skala. Der Unterschied zwischen den beiden Einheiten besteht in der Existenz des natürlichen Nullpunkts (oder der Nicht-Existenz): während 30°C keinesfalls doppelt so heiß wie 15°C sind, sind 30K doppelt so warm wie 15 K. Denn 0K ist der Nullpunkt der Temperatur. Auf der anderen Seite ist die Grenze der 0°C von Menschen gewählt, nämlich aufgrund des Gefrieren des Wassers. Weiterhin ergeben bei Grad Celsius auch Differenzen zwischen Temperaturen Sinn (beispielsweise 'globale Erwärmung um 2°C') - doch 'um zehn Prozent wärmer' wäre eine eher untreffende Aussage.

Zusammengefasst: alle verhältnisskalierten Daten sind automatisch auch intervallskaliert. Intervallskalierte Daten müssen einfach keinen natürlichen Nullpunkt besitzen und somit muss kein multiplikatives Verhältnis zwischen den verschiedenen vorhandenen Werten/Ausprägungen bestehen.

Der einfacheren Übersicht halber lassen sich die verschiedenen Skalen/Merkmale in einer Tabelle zusammenfassen:

Art	Skala	Merkmal	Wert/Ausprägung
Qualitativ	Nominal	Haarfarbe	braun, schwarz, ...
		Wahrheitswert	wahr, falsch
		Spielausgang Toto	1, 2, 0
Quantitativ	Ordinal	Schulnoten	1, 2, 3, 4, 5
		Hausnummern	..., 28, 29, 30, ...
		Game-Level	Level 1, ..., Finale
Quantitativ	Metrisch (Verhältnis)	Alkoholgehalt	12%, 13,2%, ...
		Entfernung	1,5 km, 3 km, ...
		Temperatur in Kelvin	0K, 3K ...
Quantitativ	Metrisch (Intervall)	Uhrzeit	2:30, 5:00, ...
		Temperatur in °C	-2°C, 3°C, ...

Reduktion von Skalenniveaus

Es ist möglich, Skalenniveaus zu reduzieren. Bereits damit kann man bewusst manipulieren, indem man beispielsweise Gruppen mit unterschiedlich großen Altersklassen bildet. Hierbei kann man das auf Minuten genaue Alter auf die Altersklasse '31 bis 40 Jahre' reduzieren. Dabei gehen Informationen verloren - aus einer Verhältnisskalierung wurde eine Ordinalskala.

Ein anderes Beispiel dafür wäre eine Schularbeit: genaue Punktezahlen werden zu Noten reduziert. Oder bei einem Formel 1-Rennen: Geschwindigkeiten werden auf Platzierungen reduziert, anhand derer man nicht mehr auf die Geschwindigkeit zurückschließen kann (wenn man nur die Platzierung weiß).

Exkurs zur Dezibel-Skala

Angaben zur Lautstärke sind mit Vorsicht miteinander zu vergleichen, denn doppelt so viele Dezibel, auch als dB(A) angeführt, bedeuten nicht eine doppelte Lautstärke. Während eine Kettensäge sich bei etwa 110-120 dB(A) ansiedelt, so kommt eine gewöhnliche Konversation auf ungefähr 50-60 dB(A) (vgl. [65]). Vermutlich käme niemand auf die Idee, zu behaupten, dass eine Kettensäge nur doppelt so laut sei wie ein Gespräch.

Solch offensichtliche Beispiele sind leicht zu durchschauen. Aber wenn Sie Ihr Kind beispielsweise im Technischen Museum einen Schrei-Test machen lassen und es auf 100 dB(A) käme, läge die Versuchung nahe, Folgendes zu behaupten:

Fehlvorstellung:

Das Kind mit einer Lautstärke von 100 dB(A) ist fast so laut wie eine Kettensäge mit

110 dB(A).

Erklärung des Fehlers:

Die empfundene Lautstärke ist logarithmisch skaliert, daher sind 10 dB(A) nicht als lediglich ein Zehntel Unterschied zu betrachten.

Korrekte Betrachtung des Sachverhalts:

Tatsächlich wäre die Kettensäge noch immer doppelt bis viermal so laut, da eine Erhöhung der Lautstärke um 10 dB(A) stets einer Verdoppelung der Lautstärke entspricht.

Satz 3.1. *Für eine Schallpegeländerung von ΔL (in dB(A)) ist $x = 2^{\frac{\Delta L}{10}}$ der Änderungsfaktor der (empfundenen) Lautstärke (vgl. [62]).*

Wenn die Kettensäge also auf 10 dB(A) respektive 20 dB(A) mehr käme, so wäre die empfundene Lautstärke $x = 2^{\frac{10}{10}} = 2^1 = 2$ -mal respektive $x = 2^{\frac{20}{10}} = 2^2 = 4$ -mal so laut.

4 Auswahl und Ausschluss bestimmter Daten

Ab nun beschäftigen wir uns genauer mit den Manipulationen und Denkfallen. In diesem Abschnitt widmen wir uns zunächst den einfachsten und intuitiver lösbaren Beispielen, die sich damit auseinandersetzen, welche Daten ausgewählt oder bewusst nicht außen vor gelassen werden - und wie damit manipuliert werden kann.

Behalten Sie während der gesamten Arbeit folgende Kernfragen im Gedächtnis:

- Welchen (politischen) Hintergrund weist die publizierende respektive die Statistik teilende Person/Institution auf?
- Welche Interessen pflegt die Person/Institution?
- Wer wird adressiert, wer ist die Zielgruppe?
- Was ist Zweck der statistischen Aussage und weshalb wird sie vorgelegt? Also was möchten die Autor:innen bewirken und aus welchem Grund?

Und zu guter Letzt noch ein Anwendungshinweis: ein gesundes Hinterfragen ist durchaus erwünscht - lassen Sie sich jedoch von all den Manipulationen nicht gänzlich dem Misstrauen gegenüber der Statistik hingeben, denn auch das Gegenteil kann in eine unangenehme Richtung überschlagen. So wird beispielsweise das Misstrauen gegenüber dem Staat für sämtliche Verschwörungstheorien gerne ausgenutzt (vgl. [5], S. 11). Ein gesundes Maß an Vorsicht ist wiederum äußerst hilfreich im Nebel der Statistiken.

Wes Brot ich ess', des Lied ich sing'.

(Redewendung, Ursprung nicht bekannt, aus [5], S. 90)

Um Statistiken und Daten mathematisch zu interpretieren, stellt sich natürlich zunächst die Frage, woher Sie Ihre Daten beziehen respektive woher die Daten bezogen wurden, die Ihnen gesendet wurden.

4.1 Woher wir unsere Informationen beziehen

Bestimmte Zeitschriften vertreten bestimmte Meinungen und politische Einstellungen. Obwohl Statistiken objektiv sein sollen, geschieht das Filtern und Auswählen der Daten

auch bereits, bevor etwas veröffentlicht und gedruckt wird. Welche Informationen bei den Leser:innen landen, ist somit auch teilweise schon von anderen Personen vorherbestimmt - je nachdem, in welcher Zielgruppe Sie sich befinden.

Die zwei Schlagzeilen 'Kellner verlangt 1€ für Besteck!' und 'Kalte Progression - Auswirkungen auf private Haushalte' können intuitiv zwei Zeitschriften-Redaktionen zugeordnet werden.

Genauso verhält sich das mit anderen Quellen: welche Freund:innen senden Ihnen welche Art von Statistiken und Fakten? Welche Positionen vertreten diese? Welche politischen Bewegungen erreichen Sie? Denn jedes 'Lager' hat immer Fakten parat, um andere zu überzeugen.

Wenn also der Spruch 'man ist, was man isst' übertragen werden darf, so dürfte sich bei genauerer Überlegung abzeichnen, dass man auch im Spektrum der Meinungen das ist, was man an Statistiken unüberlegt isst. In der Statistik wird also die Mathematik eng mit der Subjektivität verwoben, denn in der Statistik kommt selten das eine ohne das andere zum Vorschein, irgendwo liegen subjektive Entscheidungen, aufgrund derer die Mathematik und ihre Statistik zu eigenen Belieben genutzt wird.

4.2 Was alle denken, muss stimmen

Eine lange Zeit über war der Spinat das Gemüse schlechthin: unsere Eltern und Großeltern sagten uns, Spinat mache uns stark und sei besonders gesund, um uns groß und stark werden zu lassen - ein legitimer Grund, uns den Spinat Löffel für Löffel zwanghaft in den Mund zu schieben. Und auch Popeye mutierte zu einem Muskelpaket, nachdem er Spinat zu sich nahm. Interessanterweise beruht dieser Mythos auf einer fehlerhaften Berechnung (vgl. [6], S. 241):

Ein Schweizer Wissenschaftler setzte 1890 ein falsches Komma, als es um Angaben zu Lebensmitteln ging. 100 Gramm Spinat enthalten 3,5 Milligramm Eisen, nicht 35. Fehler können jemandem unterlaufen, aber scheinbar keine:r der Kollegschaft machte sich die Mühe, diesen Wert zu überprüfen und so startete die Reise eines Mythos, der sich noch lange am Leben erhalten sollte (und vielleicht auch noch vielen Menschen im Gedächtnis blieb). Alle übernahmen diesen Fakt und gaben ihn bedenkenlos weiter. Nun ist Spinat sicherlich gesund, aber kein Wundermittel in jenem Ausmaß, in dem es für eine lange Zeit angenommen war.

4.3 Das Beste/Schlechteste wovon?

Angelehnt an einen realen Fall (vgl. [5], S. 16) ist hier ein simples, aber selten hinterfragtes Beispiel:

Eine Zeitschrift schildert die in diesem Werk fiktive Situation: 'Die Stadt XYZ hat im Schnitt die ärmsten Menschen Österreichs' klingt für die Bürger:innen von XYZ frustrierend, doch hat sie leider niemand darüber aufgeklärt, dass nur von 20 Städten die Datensätze vorhanden waren und XYZ an letzter Stelle lag - doch viele andere Städte könnten auch schlechter abschneiden, wenn es nur die Daten gäbe. Somit stellt sich oftmals die Frage: das Beste/Schlechteste wovon?

4.4 Welche Merkmalsträger:innen zählen?

Dieselbe Zeitschrift weist nun darauf hin, dass die Menschen in XYZ die faulsten seien. Auf welcher Grundlage wird diese Aussage getätigt? Meistens wohl die durchschnittlichen Arbeitsstunden - doch weshalb ist man produktiver und fleißiger bei mehr Arbeitsstunden? Effizienz und Output können wichtiger sein, jedoch suchen sich die Autor:innen aus, welche statistischen Merkmalsausprägungen relevanter sind und welche nicht, beziehungsweise welche auch für eine Rangordnung relevanter sind. 'Lebenswerteste Stadt' - was bedeutet das, woran wird das bemessen und welche Faktoren spielen eine wie große Rolle? Dennoch wird ein solches Faktum gerne weitererzählt, ohne dabei wirklich etwas zu hinterfragen.

4.5 Welche Merkmalsausprägungen zählen?

Häufig gibt es Vergleiche wie 'CO₂-Emissionen nach Staaten', die aber massiv manipulieren. Denn natürlich wird ein Land wie Österreich mehr Emissionen verzeichnen als Liechtenstein. Wichtig ist der Pro-Kopf-Vergleich, der mehr über den Verbrauch aussagt als die gesamte Emission.

4.6 Mögliches Fehlwissen nutzen

Definitionen, welche nicht ganz bei der Bevölkerung angekommen sind, können starke Auswirkungen auf die subjektive Wahrnehmung von Statistiken haben. Bei einer Ar-

beitslosenquote von zehn Prozent ist man schnell dazu verleitet, sich vorzustellen, dass jeder:r Zehnte im Land arbeitslos sei. Dabei besagt die Definition der Arbeitslosenquote, dass nicht alle Österreicher:innen zur Berechnung herangezogen werden, sondern lediglich die Erwerbspersonen, also Erwerbstätige und Arbeitslose (vgl. [66]), was den Anteil der Arbeitslosen an der gesamten Bevölkerung kleiner erscheinen lässt als noch kurz zuvor. Darauf im Rahmen der Präsentation der Statistik und deren Kennzahlen hinzuweisen, klärt viele Missverständnisse auf, jedoch mag dies nicht immer im Interesse der Autor:innen liegen.

4.7 Zahlen ohne Relation

Wahrscheinlich haben Sie bereits gehört, dass wir 98,5 Prozent unseres Erbguts mit Schimpansen teilen - dieser Fakt allein ist beachtlich, 'wir sind zu 98,5 Prozent Affe' ist man verleitet zu sagen. Wussten Sie aber auch, dass wir 50 Prozent unseres Erbguts mit einer Banane teilen (vgl. [23])? Niemand würde nun aber glauben, dass wir zu 50 Prozent Banane sind, weil die Ähnlichkeit gar nicht bemerkbar ist. Dabei soll keine Ähnlichkeit zu Schimpansen abgestritten werden - aber zwei Argumente aufgezeigt werden:

- Die Gefahr der Missinterpretation von '98,5 Prozent unseres Erbguts mit etwas teilen' als 'zu 98,5 Prozent aus etwas bestehen'.
- Und der schlagende Punkt: der Fakt der 98,5 Prozent ist weitaus beeindruckender, solange wir noch nichts von den 50 Prozent der Banane gehört haben, denn das schmälert die empfundene Ähnlichkeit.

Wenn Sie also jemand von einer Meinung überzeugen möchte, gelingt dies leichter, indem man beeindruckende Zahlen präsentiert, ohne diese in Relation zu setzen - dann wirken sie oftmals beeindruckender.

(Interessanterweise teilen Sie übrigens 99,5 Prozent Ihres Erbguts mit allen anderen Menschen auf diesem Planeten - die 0,5 Prozent sorgen bereits für eine unglaubliche Vielfalt.)

4.8 Zahlen ohne Zeitangabe

Die Regierung hat ein 2-Milliarden-Euro-Paket geschnürt? Was zunächst nach sehr viel klingt, kann in Relation gebracht werden, wenn man nur eine Frage stellt: 2 Milliarden in welchem Zeitraum? Und wenn sich herausstellt, dass es sich um einen 10 Jahres-Plan handelt, werden pro Jahr nicht mehr 2 Milliarden Euro vom Budget verfügbar gemacht, sondern nur noch 200 Millionen. Ein weiterhin beträchtlicher Betrag, aber nur ein Bruchteil der 2 Milliarden.

4.9 Exakte Zahlen als Bluff

Mit der Aussage 'wir können mit unserem Budget nur 39 210 Asylanträge stemmen' kann mehr Überzeugung vermittelt werden als mit einer runderen Zahl wie 40 000. Auch wenn es der Mathematik gleich ist, ob den Menschen eine Zahl schön erscheint: in der subjektiven Wahrnehmung ist die Zahl 39 210 genauer und somit auch vertrauenswürdiger (vgl. [5], S. 241 f), auch wenn beides auf Schätzungen beruht oder eine Zahl gar willkürlich modifiziert wurde, um mehr Vertrauen zu wecken.

4.10 Ungenau ermittelte Zahlen

Wenn in einem Staat Schwarzarbeit 14,65 Prozent der gesamten wirtschaftlichen Leistung einnehme, so klingt dies erstmals nach viel. Dabei stellt sich jedoch die Frage, wie exakt die Schwarzarbeit erhoben werden kann, da es sich hierbei um Dunkelziffern handelt, die nie in ihrer Gänze erfasst werden können (vgl. [6], S. 178 f).

Ein spannender Fall ergab sich um 1900 in einer chinesischen Provinz (vgl. [6], S. 178, Sekundärzitat): zweimal kurz hintereinander wurde die Bevölkerungszahl erhoben. Bei der ersten kam man auf 28 Millionen Einwohner:innen, fünf Jahre später waren es sage und schreibe 105 Millionen. Huff begründet dies mit den Hintergründen der Zählung: die erste erfolgte zu Steuer- und Militärszwecken, die zweite zugunsten der Verteilung von Hilfsgütern.

4.11 Große Zahlen klein portioniert und umgekehrt

1 100 Euro für eine Jahreskarte für die öffentlichen Verkehrsmittel? Eine solche Summe klingt zuerst nach einer horrenden Zahl - 'drei Euro pro Tag' hingegen nicht. Mit solchen Werbetricks kann man günstige Preise suggerieren. Ein bekanntes Beispiel sind 12-Monats-Abonnements, welche als Einmalzahlung viel kosten, aber auf den Monat umgerechnet sind diese Beträge harmloser und günstiger als ein 1-Monats-Abo.

Umgekehrt kann auch eine kleine Zahl groß gemacht werden: wussten Sie, dass wir 3 Jahre unseres Lebens auf dem WC verbringen? Umgerechnet sitzen wir tatsächlich über 20 Minuten pro Tag auf der Toilette (vgl. [16]). Die täglichen 20 Minuten klingen nach weitaus weniger als drei Jahre des Lebens. Wenn man nun dieses Spiel fortführt, so addieren wir bei schätzungsweise 80 Jahren Lebenserwartung nun 27 Jahre Schlaf (acht Stunden täglich), 12 Jahre Arbeit und Ausbildung, über 33 Jahre mit dem Smartphone/Fernseher/Spielekonsole oder anderen Bildschirmen, zwei Jahre im Auto und ein Jahr Haushalt, drei Jahre am WC. So könnten wir fortfahren und dabei sind wir aber bereits schon bei 76 Jahren angelangt, sozusagen längst in der Pension. Ihnen bleiben nicht einmal 6 Jahre, um die Natur und die Welt zu erleben.

Bemerkung: die über 33 Jahre vor dem Bildschirm sind keine eigene Schätzung, sondern tatsächliche Einstufungen von Forschungen, die unser Verhalten untersuchten und auf sage und schreibe täglich zehn Stunden vor dem Bildschirm kommen (vgl. [31]). Ebenso muss erwähnt werden, dass in der Schätzung eine kleine Manipulation vorliegt: die 25 Jahre könnten sich mit den 12 Jahren Ausbildung/Arbeit überlappen (beispielsweise, wenn Sie regelmäßig in der Arbeit einen Computer verwenden).

4.12 'Statistische' Aussagen ohne Bedingung

Wussten Sie, dass es statistisch gesehen umweltfreundlicher ist, das Pausenbrot in Alufolie einzupacken als in eine Plastikdose? Diese Aussage kann sowohl falsch als auch richtig sein - dies ist nämlich abhängig von der Prämisse.

Bei einmaliger Benutzung mag es umweltschonender sein, ein gebrauchtes Stück Alufolie zu verwenden, als extra einen Kunststoffbehälter zu kaufen. Bei regelmäßiger Benutzung relativiert sich diese Aussage natürlich. Doch der Spruch ist eindrucksvoller als die Fußnote 'solange sie für jedes Pausenbrot einen neuen Behälter kaufen'. So offensichtlich dieses Beispiel auch sein mag, so hinterlistig können solche Fälle in diversen

Alltagssituationen sein:

'95 Prozent unserer Mitglieder sind sehr zufrieden mit unserer Dienstleistung.'

Was bedeutet Mitglied? Solange Sie nur den Spruch gesehen haben, ist es nicht klar, was das genau bedeutet. Falls die Mitgliedschaft jedoch voraussetzt, bereits zwei Jahre lang das monatliche Premium-Abo erworben zu haben, so ist das weniger beeindruckend, da immerhin eigentlich nur zufriedene Kund:innen zwei Jahre lang bei derselben Dienstleistung bleiben.

Es ist daher wichtig, bei statistischen Aussagen daran zu denken, ob diese 'Fakten' auf bestimmten Voraussetzungen beruhen.

4.13 Das 'statistisch kleinere Übel'

Ein weiteres Beispiel mit anderem Aspekt ergibt sich häufig bei Werbungen: 'Sie sparen hier Geld und CO₂-Emissionen gegenüber den anderen Anbieter:innen.'

Aber trotzdem entstehen Emissionen - wir sparen nicht, sondern emitieren lediglich weniger. Wenn Sie sich ein Auto mit weniger Sprit-Verbrauch zulegen, weil es 'umweltschonender' sei, dann gilt es, zwei Aspekte besonders genau zu bedenken:

- Sie kaufen sich noch immer ein Auto, welches weiterhin mit vielen Emissionen einhergeht, während Bahn oder Rad Fahren noch sparsamer wären.
- Falls Sie sich langfristig Ausgaben sparen, werden Sie vermutlich Ihr Geld anderweitig 'investieren' - andere Produkte, die ebenfalls mit Emissionen einhergehen.

Kurz gesagt: den Verbrauch zu senken, ist ein sicherlich sinnvolles Ziel, doch nicht immer muss man sich auf die Schulter klopfen, wenn man anstelle von 'extrem bedenklich hohe' nur noch 'bedenklich hohe' Emissionen verschuldet, unabhängig vom Auto in allen Bereichen des Alltagslebens - das kleinere Übel ist nach wie vor noch immer ein Übel.

Am Ende bleibt es Ihre Entscheidung - denken Sie aber nächstes Mal daran, wenn Sie eine Werbung hören, die damit werbt, dass Sie sich damit Geld sparen, obwohl Sie kurz davor stünden, etwas zu kaufen, das Sie sonst womöglich erst gar nicht gekauft hätten.

4.14 'Statistische' Aussagen mit falschen Merkmalsträger:innen

Oft werden in Statistiken Merkmalsträger:innen berücksichtigt, welche nicht zur untersuchten Gruppe gehören - beispielsweise bei der sogenannten Ausländer:innenkriminalität (vgl. [6], S. 10).

Fehlvorstellung:

In Österreich ansässige Ausländer:innen sind krimineller als Österreicher:innen.

Erklärung des Fehlers:

Bei solchen Aussagen werden in Österreich alle Straftaten durch Ausländer:innen (darunter auch Tourist:innen) 'in einen Topf geworfen'. Was ausländische Tourist:innen in Österreich zu Werke bringen, sollte keine Berücksichtigung finden, wird aber dazugekommen. Österreichische Einwohner:innen können auch im Ausland Straftaten begehen, dies wird aber auch nicht berücksichtigt.

Korrekte Betrachtung des Sachverhalts:

Um Aussagen über die Grundgesamtheit 'in Österreich Ansässige' zu treffen, dürfen auch nur Merkmalsträger:innen mit dem Merkmal 'in Österreich ansässig' herangezogen werden.

4.15 Aus Groß wird Klein und umgekehrt

Zwischen Einheiten zu wechseln, kann bei uns unterschiedliche Vorstellungen hervorrufen. Ein Volumen in Kubikmetern oder Kubikkilometern anzugeben, kann erhebliche Unterschiede in unserem Gehirn abbilden.

So nahm sich eine Partner:innenbörse vor, eine Hoffnung verbreitende Statistik (vgl. [14]) zu veröffentlichen, es verliebe sich auf besagter Börse alle 11 Minuten ein Single. Das mag beeindruckend klingen, doch bei genauerer Betrachtung ist dem nicht so.

Bei 4 500 000 Mitgliedern dauert es eine Weile, bis Sie an der Reihe wären: bei 11 Minuten pro Mitglied wären Neukund:innen somit im Schnitt nach $4\,500\,000 \cdot 11 = 49\,500\,000$ Minuten an der Reihe. 49 500 000 Minuten bzw. 825 000 Stunden, 34 375 Tage oder auch...94,11 Jahre (Schaltjahre berücksichtigt).

Aus den 11 Minuten wurden nun 94,11 Jahre, die wir auf das Verlieben warten - und dann haben erst einmal wir uns verliebt, aber niemand sich in uns.

Zuvor wurde in dieser Arbeit über Prämissen (Voraussetzungen) einer Statistik gesprochen. Tatsächlich gab es auch bei dieser Statistik eine in der Werbung nicht erwähnte Voraussetzung - nämlich, dass man dafür Premium-Mitglied sein müsse.

Beim Kündigen der Premium-Mitgliedschaft wird nach einem Grund gefragt und 2013 gaben 48 000 Personen an, fündig geworden zu sein. Auch hier stellt sich die Frage, ob sich 'fündig' auf genau diese Börse beziehen muss oder diese Option auch angeklickt wird, wenn man woanders fündig wurde.

Auf jeden Fall bedeuten 48 000 Personen pro Jahr dann wiederum 10,95 Minuten pro Personen.

Es bedarf also $\frac{1}{48\,000}$ eines Jahres, damit sich eine Person verliebt, so ergeben sich

$$\frac{1 \text{ Jahr}}{48\,000 \text{ Personen}} = \frac{36525 \text{ Tage}}{48\,000 \text{ Personen}} = \frac{8\,766 \text{ Stunden}}{48\,000 \text{ Personen}} = \frac{525\,960 \text{ Minuten}}{48\,000 \text{ Personen}} = 10,9575 \text{ Min./Person}$$

Sozusagen beruht die Statistik darauf, dass einfach 48 000 Personen innerhalb eines Jahres angaben, fündig geworden zu sein. Spannenderweise kündigten 78 000 Singles ihre Premium-Mitgliedschaft ohne Dating-Erfolg.

Das sind $\frac{78\,000}{48\,000+78\,000} = \frac{78\,000}{126\,000} = 0,619 = 61,9\%$ aller Premium-Kündigungen in dem Jahr.

4.16 'Statistische' Aussagen unter Ausschluss verschiedener Faktoren

Eine weitere motivierende Dating-Statistik liegt für männliche Konsumenten eines Anti-Haarausfall-Mittels vor: 'Männer mit Haaren bekommen doppelt so viele Dates' (vgl. [2]).

Das klingt im ersten Moment devastierend für Männer mit Glatze, doch auch hier liegen einige Denkfallen vor, denn es wurden mehrere Faktoren nicht berücksichtigt.

Die Lektion aus dieser Situation ist mitunter eine der wichtigsten dieser Arbeit: Bei jeder Statistik muss erwogen werden, ob andere Faktoren nicht berücksichtigt wurden, die jedoch sehr wohl einen Einfluss auf die Statistik haben könnten.

Den Angaben zufolge wurden 400 Frauen in zwei Testgruppen befragt: einer Gruppe wurde ein Foto eines Mannes mit vollem Haar vorgelegt, der anderen Gruppe wiederum



Abbildung 1: Den Versuchsgruppen vorgelegte Fotos [2]

ein Foto, bei dem die Haare professionell wegretuschiert wurden und stattdessen nun eine (Halb-)Glatze zu sehen war.

Daraufhin mussten die je 200 Frauen angeben, ob sie den Mann auf dem Foto daten würden. 65,5 Prozent der ersten, aber nur 30,5 Prozent der zweiten Gruppe hätten den Mann gedatet.

Es ist ein wenig verwunderlich, dass die Schlagzeile der Werbung daraufhin nicht lautete: 'Männer mit Haaren bekommen **mehr** als doppelt so viele Dates', da dies noch beeindruckender gewesen wäre.

Dafür wurde hingegen empfohlen, dass das Mittel auch bei vollem Haar genommen werden solle, man wolle ja nicht aus dem 'Beuteschema' der Frauen fallen, wenn sich ohne Prävention Haarausfall anbahne. Doch was ist denn nun falsch an der Statistik? Die Datenanalyse der bei der Befragung erhobenen Daten scheint immerhin korrekt abgelaufen zu sein.

Hier wurden Annahmen getroffen und Faktoren ausgelassen, die sich auf die Aussagekraft erheblich auswirken.

- Faktor 1 - im echten Leben kennenlernen:

Dating-Chancen hängen nicht gleich von einem Foto ab. In vielen 'Live'-Situationen kann man im Cafe, Club oder woanders auch mit der eigenen Persönlichkeit überzeugen, um Sympathien zu wecken. Und selbst bei Online-Dating-Plattformen besteht die Möglichkeit, durch ein interessantes Profil hervorzustechen, wohingegen ein 'unsympathischer' Mann mit vollem Haar an Chancen einbüßen könnte.

- Faktor 2 - Männer und der gefallene Begriff 'Beuteschema':
Die Versuchsgruppe bestand aus Frauen im Alter von 18 bis 30 Jahren. Allgemein bekannt sei, dass Männer mit zunehmenden Alter eher von Haarausfall betroffen sind. Aber möchte ein beispielsweise 50-jähriger automatisch eine 18 bis 30 Jahre alte Frau daten? Vielleicht wollen diese Männer auch gleichaltrige Partnerinnen suchen, die höchst wahrscheinlich auch andere Prioritäten setzen und nicht so viel Wert auf volles Haar setzen.
- Faktor 3 - Männer außerhalb des 'Beuteschemas':
Sei nun ein 50-jähriger Mann an das Dating mit Damen im Alter von 18 bis 30 Jahren interessiert. Kann man davon ausgehen, dass die getroffene Aussage zu dem Foto auch auf seine Altersgruppe zutrifft?
- Faktor 4 - die Größe und andere Merkmale:
Diese Statistik vermittelt das Gefühl, dass andere optische Attribute keine nennenswerte Rolle spielen: die Körpergröße, das Lächeln und viele mehr.

Dies waren zunächst vier offensichtliche Faktoren - bei näherer Betrachtung kann man diese noch um weitere Aspekte ergänzen. Es mag in der Dating-Welt helfen, volles Haar zu besitzen - aber dass das 'Ja' oder 'Nein' zu einem Date ausschließlich daran geknüpft ist, scheint bei genauerer Betrachtung doch ein wenig weit hergeholt.

Auf selbiger Website wird behauptet, dass 50,3 Prozent der Single-Frauen bei einem Date auf die Haare und Frisur achten - doch achten diese vermutlich (und hoffentlich) auch auf viel anderes: das Auftreten, die Maniküre, die Kleidung, den Humor, das Lächeln, die Art zu reden und vieles mehr. Auch hier gilt: gibt es nicht auch andere Faktoren, die die Dating-Chancen bedingen?

4.17 Zusammenfassung

Die bewusste Auswahl, welche Informationen wir preisgeben oder betonen wollen, kann starke Auswirkungen auf unsere Argumentationen haben. Umgekehrt ist daher das Bewusstsein wichtig, darüber nachzudenken, ob bei einer uns präsentierten Statistik etwaige Fakten oder beeinflussende Faktoren ausgelassen respektive übersehen worden sein konnten.

Dabei sind in einer Diskussion bezüglich einer vorliegenden Datensammlung vor allem zwei Aspekte essenziell: welche der vorliegenden Attribute wir oder die andere Partei in

einer Situation als relevant (bzw. am relevantesten) bewerten und ob deren Bedeutung auch von beiden Seiten korrekt interpretiert werden.

Aufbauend auf dem bisher Aufgezeigten werden nun Schritt für Schritt mathematisch fortgeschrittenere Beispiele näher beleuchtet - dabei gilt es, diese 'Leitfragen' zu beachten: welche Fakten sind relevant und wie bzw. in welcher Relation sind sie zu interpretieren?

4.18 Exkurs 1: Rundungen

In einem fiktiven Staat mit neun Millionen Einwohner:innen treffe eine Aussage auf 43 Prozent zu. Nun lautet die Schlagzeile: 'Auf jede:n Zweite:n trifft ... zu.'

Das klingt fast nach dem selben Anteil, doch die sieben Prozent sind immerhin 630 000 Menschen.

Auch kleine Rundungsfehler beim Auf- und Abrunden können als absolute Zahl sinngemäß eine gesamte Stadt ausmachen.

4.19 Exkurs 2: 'Postfaktisch'

Leider gibt es auch politische Argumentationsmuster jenseits der Fakten (respektive unter Ausschluss sämtlicher Statistiken), wo solche jedoch notwendig wären - solche Aussagen nennt man auch 'postfaktisch' (vgl. [5], S. 7). Dabei geht es primär um Emotionen (beispielsweise das Wecken von Ängsten) anstelle rationaler Argumentation, eventuelle statistische Widersprüche finden hier kein Gehör.

5 Häufigkeiten und Änderungen

Ab diesem Abschnitt wird mehr gerechnet, mehr Formeln kommen zum Einsatz und die Denkfallen sowie Manipulationen werden systematisch und durchgehend mithilfe der Überschriften 'Fehlvorstellung', Erklärung des Fehlers' und 'Korrekte Betrachtung des Sachverhalts' behandelt.

5.1 Absolute und relative Häufigkeit bzw. Änderung

Satz 5.1. *[Absolute Häufigkeit] Sei n die Anzahl an Merkmalsträger:innen zu einem bestimmten Merkmal X . Dabei treten k verschiedene Werte $x_1, x_2, \dots, x_k$ auf (manche Werten können schließlich mehrfach auftreten).*

So ist H_i die absolute Häufigkeit von x_i mit $H_1 + H_2 + \dots + H_k = n$ (vgl. [10]).

Umgangssprachlich erklärt: die absolute Häufigkeit eines Wertes gibt an, wie oft dieser Wert in der Urliste vorkommt.

Absolute Häufigkeiten können sehr beeindruckend auf uns wirken. Die folgenden Daten beziehen sich auf die Polizeiliche Kriminalstatistik 2022 (vgl. [72]):

Wussten Sie, dass im Jahre 2021 satte 410 957 Straftaten in Österreich vermeldet wurden? Demnach klingt Österreich nach einem kriminellen Staat - tatsächlich war dies jedoch der niedrigste Wert seit 2013. Weiteres vermeldet das Bundeskriminalamt im darauf folgenden Jahr 488 949 Straftaten.

Fehlvorstellung:

Österreich wurde krimineller.

Erklärung des Fehlers:

Es reicht nicht, nur die Änderung der absoluten Häufigkeit (absolute Änderung) zu betrachten. Dabei gibt es mehrere sich verändernde Parameter: zum Beispiel sollte man bedenken, dass es im Jahr 2021 noch Lockdowns gab, welche sicherlich zur Reduzierung der Kriminalität beitrugen.

Ebenfalls muss generell bedacht werden, dass Einwohner:innenzahlen meistens steigen (vgl. [67]).

Fehlvorstellung:

Wenn das Jahr vor der Pandemie (also 2019) mit 2022 verglichen wird, so könnte schnell schlussgefolgert werden, dass die Zahl der Straftaten von 488 912 auf 488 949 stieg und Österreich krimineller wurde.

Erklärung des Fehlers:

Wir sehen bereits, dass die absolute Änderung eher gering ist. Und trotz der Steigerung wäre es nicht ratsam, von 'krimineller' zu sprechen, da die Einwohner:innenzahl sich vom 01.01.2019 bis 01.01.2022 von 8 858 775 auf 8 978 929 stieg. Die sogenannte relative Änderung wäre hilfreicher, um den Sachverhalt zu betrachten:

Korrekte Betrachtung des Sachverhalts:

Satz 5.2. *[Relative Änderung] Die relative Änderung ist die absolute Änderung relativ zum Grundwert (vgl. [43]).*

Die Formel *relative Änderung* = $\frac{\text{neuer Wert} - \text{alter Wert}}{\text{alter Wert}}$ beschreibt die relative Änderung auf umgangssprachliche Weise für die meisten Beispiele in der Alltagsrealität, bei denen der 'alte Wert' der Grundwert und der 'neue Wert' der Anteil ist.

Da aber prinzipiell auch der alte Wert ausgehend vom neuen Wert betrachtet werden kann, ist hier eine umgangssprachliche, allgemeingültige Formel:

$$\text{relative Änderung} = \frac{\text{Vergleichswert} - \text{Ausgangswert}}{\text{Ausgangswert}},$$

wobei der Ausgangswert der Grundwert ist, von dem aus wir andere Werte mit diesem vergleichen wollen. Der Vergleichswert ist dann jener Wert, den wir mit dem Grundwert vergleichen möchten.

Nun betrachten wir die relative Änderung der Straftaten in der Zeitspanne von 2019 bis 2022:

$$\frac{48849 - 488912}{488949} = \frac{37}{488912} = 0,000076 = 0,0076\%$$

Die Anzahl der Straftaten stieg demnach um 0,0076 Prozent.

Nun zur relativen Änderung der Einwohner:innenzahl in der Zeitspanne von 2019 bis 2022:

$$\frac{8\,978\,929 - 8\,858\,775}{8\,858\,775} = \frac{120\,154}{8\,858\,775} = 0,01356 = 1,356\%$$

Die Einwohner:innenzahl stieg um 1,356 Prozent. So gesehen kann behauptet werden, dass Österreich nicht krimineller wurde, da die Einwohner:innenzahl prozentuell gesehen stärker stieg als die Anzahl der Straftaten.

Doch auch hier stellt sich die Frage, weshalb ausgerechnet 2019 und 2022 in dieser Arbeit miteinander verglichen wurden - beim Vergleich mit anderen Jahren sähe das womöglich anders aus. In einem fiktiven Zeitungsartikel sind Sie jedoch meist der Wahl der Fakten des oder der Autor:in ausgeliefert.

Satz 5.3. *[Relative Häufigkeit] Sei n die Anzahl an Merkmalsträger:innen zu einem bestimmten Merkmal X und H_i die absolute Häufigkeit von x_i .*

So ist $h_i = \frac{H_i}{n}$ die relative Häufigkeit von H_i mit $h_1 + h_2 + \dots + h_k = 1$ (vgl. [10]).

Umgangssprachlich erklärt: die relative Häufigkeit gibt an, mit welchem Anteil ein bestimmter Wert in der Urliste vorkommt. Wenn man diesen Anteil auf Prozent umrechnen möchte, muss der entsprechende Wert mit 100 multipliziert werden. Die Summe der relativen Häufigkeiten zu den Merkmalen muss 1 (bzw. 100 Prozent) ergeben, da ja auch 100 Prozent der Merkmalsträger:innen betrachtet werden.

Mit diesem Wissen fahren wir in Kürze fort, zuvor aber noch eine kleine Manipulation durch Wortwahl. Was klingt eindrucksvoller: '32,5 Prozent aller Menschen leiden unter...' oder 'Jeder und jede Dritte von uns leidet unter...'? Letztere Aussage mag stärker an die Vorstellungskraft appellieren: drei Menschen nebeneinander - eine Person leidet unter etwas Bestimmten. 32,5 Prozent hingegen wirken nicht so greifbar, auch wenn es beinahe dasselbe ist.

Nun zum Vergleich der absoluten und relativen Häufigkeit:

2021 gab es in Österreich insgesamt 525 Verkehrsunfälle aufgrund der Missachtung einer roten Ampel (vgl. [53]), hingegen 32 774 Verkehrsunfälle insgesamt (vgl. [68]).

Fehlvorstellung:

Es ist sicherer, bei Rot über die Straße zu gehen als bei Grün.

Erklärung des Fehlers:

Es passieren 'nur' so wenige Unfälle bei roter Ampel, da sich so viele Personen an die Gesetze halten (bzw. auch Unfälle unabhängig von Ampel-Signalen geschehen).

Korrekte Betrachtung des Sachverhalts:

Man müsste ermitteln, wie viele Unfälle bei Grün passieren bzw. wie oft niemand bei Rot/Grün zu Schaden kommt. Doch auch ohne Nachweis dürfte klar sein, dass es weiterhin ratsam ist, sich an die Ampeln zu halten.

So einfach dieses Beispiel nun vorgekommen sein mag, so unklar war vielen dasselbe Prinzip in der Impfdebatte (vgl. [42], S. 8 f):

2021 entstand eine umfangreiche Debatte bezüglich der Wirksamkeit von Impfungen. Häufig machten Aussagen wie 'mehr Geimpfte als nicht Geimpfte sind im Krankenhaus' die Runde - warum diese absoluten Häufigkeiten irreführend sind, zeigt sich relativ einfach anhand einer fiktiven Impfung gegen einen fiktiven Virus.

Es gäbe also eine fiktive Stadt mit 200 000 Einwohner:innen, davon seien 180 000 geimpft und 20 000 seien es nicht. Dem aktuellen Stand zufolge befinden sich 15 000 Personen im Krankenhaus, davon wiederum 10 000 Geimpfte bzw. 5 000 nicht Geimpfte. Weiteres wird angenommen, dass die Personen im Krankenhaus einzig aufgrund des Virus dort sind.

Nun wird die nachstehende Tabelle publiziert:

Impf-Wirksamkeit?	Geimpft	Nicht geimpft	Summe
Im Krankenhaus	15 000	10 000	25 000
Nicht im Krankenhaus	165 000	10 000	175 000
Summe	180 000	20 000	200 000

Nun veröffentlichen zwei verschiedene Zeitschriften einen Artikel mit den Schlagzeilen '50% mehr Geimpfte als Nicht-Geimpfte im Krankenhaus' respektive 'Impf-Wirksamkeit bestätigt - Hälfte aller Nicht-Geimpften im Krankenhaus'. Tatsächlich stimmen beide Schlagzeilen und werden auch die gewünschten Wirkungen erzielen.

Fehlvorstellung:

10 000 der 15 000 Personen im Krankenhaus sind geimpft, daher schadet die Impfung.

Erklärung des Fehlers:

Es geht nicht nur um die bloßen Anzahlen der Geimpften und Nicht-Geimpften im Krankenhaus, sondern auch darum, wie viele der Geimpften und Nicht-Geimpften prozentuell betroffen sind. Darum müssen die relativen Häufigkeiten anstelle der absoluten Häufigkeiten verglichen werden.

Korrekte Berechnung des Sachverhalts:

$\frac{10\,000}{180\,000} = 0,05556 = 5,556\%$ der Geimpften sind im Krankenhaus.

$\frac{10\,000}{20\,000} = 0,5 = 50\%$ der Nicht-Geimpften sind im Krankenhaus.

Das bedeutet, dass unter den Nicht-Geimpften ein größerer Anteil im Krankenhaus liegt als unter den Geimpften, also die Impfung wirksam ist.

Bemerkung: genau genommen gibt es auch hier weitere Aspekte zu beachten. Ein Beispiel: davon ausgehend, dass die Impfrate mit zunehmenden Alter tendenziell steigt, ist es ebenfalls irreführend, Jung und Alt in einen 'Topf zu werfen'. Falls tendenziell Pensionist:innen eher geimpft sind als Jugendliche, sind diese gleichzeitig aber auch aufgrund des Alters und der Gebrechlichkeit bei Virus-Infektionen stärker gefährdet als Junge und somit auch noch mehr Geimpfte im Krankenhaus. Somit ist es auch fahrlässig, aufgrund der Gesamtstatistik Empfehlungen gleichermaßen für alle Altersgruppen abzugeben - empfehlenswert wäre eine Aufschlüsselung nach Altersgruppen, die die relativen Häufigkeiten hinsichtlich der Altersklassen oder weiteren Eigenschaften wie Vorerkrankungen miteinander zu vergleichen, doch so detailliert sind die meisten Statistiken nicht. Mehr zu diesem Aspekt finden Sie im Kapitel Simpson-Paradoxon.

Das Thema der absoluten Häufigkeit im Vergleich zur relativen Häufigkeit kommt häufiger vor als viele denken würden und spannenderweise brachte dies auch einer der bekanntesten Stand Up-Comedians der USA auf den Punkt:

Dieser Typ hat versucht, mich zum Tauchen zu überreden. Ich sagte: 'Ich gehe nicht. Ich will nicht von einem Hai gefressen werden.' Er sagte: '90 Prozent der Haiangriffe passieren im flachen Wasser.' Ich dachte: 'Schlaumeier! Da sind ja die Leute!' Verstehst du? Man nennt es Strand!
(Bull Burr, [12], aus dem Englischen übersetzt)

Hiermit machte Bill Burr mehr oder minder bewusst auf der Bühne klar, worum es beim Vergleich der absoluten und relativen Häufigkeit geht.

5.2 Kleines n

Bei kleinen Anzahlen können kleine Abweichungen zu erheblichen relativen Änderungen führen. Wenn in einer fiktiven Großstadt 20 Personen von einer fiktiven, seltenen Krankheit betroffen sind und im darauffolgenden sind es 27 Personen, so beträgt die relative Änderung:

$$\frac{27-20}{20} = 0,35 = 35\%$$

Diese sieben Personen können die Schlagzeile 'Krankheit verbreitet sich dramatisch - 35 Prozent mehr Betroffene' zur Folge haben. Dabei könnte die Zahl auch durch einen Zufall oder gar durch Bevölkerungswachstum gestiegen sein.

6 Prozente

Aus Daten und Statistiken Prozentangaben zu bestimmten Sachverhalten und Änderungen zu generieren, wirkt relativ simpel durchzuführen. Tatsächlich lauern auch hierbei viele Denkfallen und Manipulationstechniken.

6.1 'Um das Doppelte' vs. 'Auf das Doppelte'

Möglicherweise kennen Sie Aussagen wie 'Wahnsinn, der Preis ist um das Doppelte gestiegen!' Dabei ist oftmals zu hinterfragen, ob die Person tatsächlich die richtige Formulierung verwendete.

Fehlvorstellung:

Wenn sich ein Preis verdoppelt, so ist dieser um das Doppelte gestiegen.

Erklärung des Fehlers:

Die Quelle dieses Fehlers dürfte in der umgangssprachlichen Verwendung der deutschen Sprache liegen. Wenn sich ein Preis verdoppelt, so ist er um 100 Prozent gestiegen (also beispielsweise von 250€ auf 500€) - er ist also **auf** das Doppelte gestiegen. Würde der Preis **um** das Doppelte steigen, so müssten wir zum ursprünglichen Preis das Doppelte dieses Preises addieren und hätten somit eine Verdreifachung des Preises vorliegen.

Korrekte Betrachtung des Sachverhalts:

Auch hier ist die Formel zur Berechnung der relativen Änderung praktisch:

$$\frac{\text{neuer Wert} - \text{alter Wert}}{\text{alter Wert}} = \frac{500 - 250}{250} = 1 = 100\%$$

Dies bedeutet eine Erhöhung des Preises um 100 Prozent, der Preis ist also auf das Doppelte gestiegen (oder um 100 Prozent), aber nicht **um** das Doppelte, denn das bedeutete eine relative Änderung von 200 Prozent.

6.2 Mehrfache relative Änderung

Wir betrachten den Wert eines fiktiven Materials, es sei pro Kubikcentimeter 100€ wert. Über die Jahre verändert sich der Marktwert wie folgt:

Jahr	Änderung des Wertes zum Vorjahr
2061	+1 %
2062	-7 %
2063	+10 %
2064	+12 %
2065	+3 %

Nun stellt sich die Frage, um wie viel Prozent sich der Preis von 2060 bis 2065 verändert hat.

Fehlvorstellung:

$1\% - 7\% + 10\% + 12\% + 3\% = 19\%$, daher stieg der Preis um exakt 19 Prozent.

Erklärung des Fehlers:

Wenn man nun wissen möchte, wie sich der Preis im Vergleich von 2060 zu 2065 verändert hat, reicht es nicht, die einzelnen relativen Änderungen zu addieren, weil sich der Grundwert jedes Jahr ändert. Die 100€/cm³ stiegen auf 101€/cm³ und sanken dann um sieben Prozent. Aber diese sieben Prozent beziehen sich auf die 101€/cm³ und nicht die anfänglichen 100€/cm³, daher darf man nicht einfach die Prozente über die Jahre addieren.

Korrekte Berechnung des Sachverhalts:

Wir multiplizieren die Änderungsfaktoren (zum Vergleich: die relativen Änderungen ergeben sich, wenn von den Änderungsfaktoren jeweils die Zahl 1 subtrahiert wird, da es ja nur um die Änderung ginge). Es ergeben sich also die Änderungsfaktoren 1,01 (Steigerung um ein Prozent), 0,93 (Reduzierung um sieben Prozent) usw.

So ergibt sich

$$1,01 \cdot 0,93 \cdot 1,1 \cdot 1,12 \cdot 1,03 = 1,1919 = 119,19\%$$

Der Preis erhöhte sich also über die fünf Jahre um 19,19 Prozent. Das ist zwar nicht viel mehr als die 'falschen' 19 Prozent, aber kann doch situationsspezifisch einen bedeutenden Unterschied machen.

6.3 Grundwert wovon?

Wenn die Schlagzeile einer Zeitung lautet: 'Lehrer:innen verdienen 30 Prozent mehr', stellt sich die Frage: 30 Prozent mehr als wer oder 30 Prozent mehr als wann? Zunächst wirkt es nach einer drastischen Gehaltserhöhung, aber das kann sich auf Vieles beziehen: 30 Prozent mehr als vor 20 Jahren? 30 Prozent mehr als der Mindestlohn? 30 Prozent mehr als in Polen?

Zudem stellt sich auch die Frage, ob die 30 Prozent auch berücksichtigen, ob der Arbeitsaufwand derselbe blieb. Wenn ein neuer Dienstvertrag mehr Unterrichtseinheiten vorsieht, so ist dies ohnehin in Relation zu betrachten.

So meinte auch der ehemalige US-Präsident George W. Bush, die USA würden den Ausstoß der klimaschädlichen Treibhausgase um mindestens 50 Prozent absenken (vgl. [6], S. 90 f). Leider vergaß er, das Basisjahr anzugeben. Sollen die jährlichen Emissionen halb so hoch sein wie zum Zeitpunkt der Aussage oder halb so hoch wie im schlimmsten Jahr oder gar ein ganz anderes Jahr?

6.4 Falscher Grundwert

In diesem Abschnitt geht es um etwas Triviales wie den Grundwert, wobei die Annahme eines falschen Grundwertes zur Verbreitung eindeutiger Fehlinformation führen kann.

Das Harding-Zentrum für Risikokompetenz veröffentlicht die Unstatistik des Monats und erklärt einen unfassbaren Fehler (vgl. [32]):

Stern, aber auch andere Medien, berichteten davon, dass der Gardasee kaum zur Hälfte gefüllt sei respektive der Pegel um etwa 60 Prozent gesunken sei. Dabei beruht die gesamte Prozentrechnung auf einem falschen Grundwert.

Fehlvorstellung:

Wenn der Pegel im März 2022 ganze 99 cm betrug, im darauffolgenden Jahr jedoch nur mehr 46 cm, so ist der See nicht einmal mehr zur Hälfte gefüllt.

Erklärung des Fehlers:

Wir setzen an der falschen Skala an: der Pegelnullpunkt orientiert sich an einem Pegel in der italienischen Gemeinde Peschiera und nicht am tiefsten Punkt des Sees. Im Abschnitt über Skalierung von Daten wurde bereits der Vergleich zwischen Intervallskala und Verhältnisskala hergestellt. Zur Erinnerung: nur bei einer Verhältnisskala kann

man von 'doppelt so viel' oder '20 Prozent weniger' sprechen, da es einen natürlichen Nullpunkt gibt. In der italienischen Gemeinde wurde einfach an einer von Menschen festgelegten Stelle ein Pegel zur Feststellung des Wasserstandes angebracht, von dem aus der Wasserstand 99 cm betrug. Dabei weist der See eine durchschnittliche Tiefe von 135 Metern auf - in Relation dazu fallen die 53 cm nicht so sehr auf und man kann eindeutig nicht von 'halb leer' sprechen.

Korrekte Betrachtung des Sachverhalts:

Aufgrund des Pegelnullpunkts darf man nicht mit Prozenten rechnen (das wäre so, als wären 20°C doppelt so warm wie 10°C). Auch ausgehend vom tiefsten Punkt des Sees darf keine Prozentrechnung angewandt werden, da der See kein mit Wasser gefüllter Quader, sondern unförmig ist. Wenn, dann kann ausgehend von einer möglichst genauen Durchschnittstiefe näherungsweise eine prozentuelle Veränderung des Volumens mithilfe des Pegelstandes angegeben werden - nur wird eine Reduzierung des Wasservolumens um ein Prozent nicht für dieselbe Anzahl an Klicks sorgen.

Schlussendlich revidierte Stern die Aussage, der See sei nur zu 38 Prozent gefüllt (vgl. [69]). Ein weiterer Gedanke: bei welchem Pegel soll denn ein See zur Gänze gefüllt sein - beim Pegelnullpunkt oder einen Tropfen vor dem Übergehen? Oder lässt sich hier einfach kein Grundwert finden, es sei denn, er sei von Menschenhand festgelegt? Und dies führt uns wieder zum Pegelnullpunkt, der aber nun einmal nicht für 100 Prozent steht.

Zum Glück gibt es keine auf den Gardasee konzentrierte, epische Klimakrise - dennoch gab der Rückgang von 53 cm zu denken und man befürchtete bei Stern eine Verschlimmerung der stärksten Dürre seit 70 Jahren. Ironischerweise ist auch dies wiederum zu entkräften, da der See seit 1949 künstlich reguliert wird (vgl. [38]) und keinerlei Sorge für Tourist:innen bestehe, den Badeurlaubs-Spaß begrenzen zu müssen.

Natürlich kann der Klimawandel auch einen Teil zum Wasserpegel beitragen und das Ausmaß bleibt weiterhin im Auge zu behalten - auf wissenschaftlichere Art und Weise. Doch wie viele der Leser:innen wurden schlussendlich auf den Fehler des Grundwerts hingewiesen und wie oft wurde die negative Schlagzeile weitererzählt?

Bemerkung: das Beispiel mit der Temperatur wurde nun bereits zweimal erwähnt. Falls Sie sich fragen: so etwas wurde tatsächlich berichtet (vgl. [75]). Im Jahre 2016 behauptete wetter.at, dass der Jänner mit 3,5°C statt 0,8°C einer Temperaturzunahme um 337 Prozent entspricht:

$$\frac{35-08}{08} = 3,375 = 337,5\%$$

Nicht bedacht wurde, dass die 0,8°C nicht den Grundwert darstellen sollen, wenn der Nullpunkt willkürlich, also eben einfach von Menschen, gesetzt wurde. Beim Rechnen in Grad Fahrenheit oder Kelvin kämen andere Werte zur prozentuellen Steigung zum Vorschein (wobei letztere Einheit, wie bereits erklärt, noch eher zulässig wäre).

6.5 (Ver-)Tauschen von Grundwert und Anteil

Beim Vergleichen zweier Kennzahlen im Verlauf der Jahre sind wir es gewohnt, den neuen Wert als Anteil anzunehmen und ihn mit dem ursprünglichen Ausgangswert als Grundwert zu vergleichen. Ob die Bedeutung von Grundwert und Anteil in diesem Szenario eindeutig ist, sei dahingestellt. Denn schließlich kann man auch umgekehrt den alten Wert ausgehend vom neuen Wert betrachten - also Grundwert und Anteil vertauschen. Dabei können sich aber andere statistische Aussagen ergeben.

Um dies anschaulicher zu gestalten, hier ein Beispiel:

In einer fiktiven Stadt gibt es zunächst 3 000 Arbeitslose, im Jahr darauf sind es 2 000. Nun stehen sich zwei Politiker:innen gegenüber und eine:r der beiden meint: 'Letztes Jahr hatten wir 50 Prozent mehr Arbeitslose. Noch einmal: unter meiner Führung konnte die Zahl der Arbeitslosen um 50 Prozent gesenkt werden'.

Fehlvorstellung:

Wenn wir letztes Jahr 50 Prozent mehr Arbeitslose hatten, dann ist die Zahl der Arbeitslosen um 50 Prozent gesunken.

Erklärung des Fehlers:

Wenn wir ausgehend vom aktuellen Wert zurückblicken, stimmen die 50 Prozent. Wenn wir aber zunächst ausgehend vom letzten Jahr nun Aussagen über das Hier und Jetzt treffen wollen, so müssen wir beachten, dass der Grundwert nun der Wert des letzten Jahres ist und nicht umgekehrt.

Korrekte Betrachtung des Sachverhalts:

Auch hier betrachten wir die relative Änderung. Erste Aussage: 'Letztes Jahr hatten wir 50 Prozent mehr Arbeitslose.'

$$\frac{\text{Vergleichswert} - \text{Ausgangswert}}{\text{Ausgangswert}} = \frac{3000 - 2000}{2000} = 0,5 = 50\%$$

Die Aussage stimmt, vergangenes Jahr gab es über 50 Prozent mehr Arbeitslose als nun.

Zweite Aussage: 'Unter meiner Führung konnte die Zahl der Arbeitslosen um 50 Prozent gesenkt werden.' Wir müssen nun also die Zahlen des Grundwerts (Ausgangswert) und Anteils (Vergleichswert) tauschen, die Formel wiederum bleibt aber gleich:

$$\frac{\text{Vergleichswert} - \text{Ausgangswert}}{\text{Ausgangswert}} = \frac{2\,000 - 3\,000}{3\,000} = -0,3333 = -33,33\%$$

Die Aussage muss korrigiert werden: die Arbeitslosenzahl sank um rund 33 Prozent, nicht 50 Prozent.

Nun kann es jedoch gut sein, dass ein und dieselbe Tatsache in einem politischen Diskurs anders ausgedrückt wird. Die Führungspartei wird sich im Normalfall auf 'Als wir anfangen, die Ärmel hochzukrempeln, gab es noch 50 Prozent mehr Arbeitslose' stützen, während die Opposition eher auf die (dennoch weiterhin beeindruckend wirkenden) '33 Prozent weniger' referenzieren sollten - anstelle der 'vorher 50 Prozent mehr'.

Bermerkung: eine kleine Erinnerung - dass es nun 33 Prozent weniger Arbeitslose gibt, bedeutet nicht immer automatisch, dass auch die Arbeitslosenquote gesunken ist! Es könnten immerhin die Einwohner:innenzahl oder die Zahl der Erwerbspersonen (prozentuell) noch stärker gesunken sein, beispielsweise aufgrund von Abwanderung.

Sie sehen: einzeln ausgewählte Fakten können unfassbar beeindruckend sein - und im Kontext dann wieder nichtig.

Ein weiteres Beispiel für das Verwechseln von Grundwert und Anteil sind die Mehrwertsteuern: wenn auf einen Nettopreis n 20 Prozent MwSt. aufgeschlagen werden, bedeutet das nicht, dass 20 Prozent des Bruttopreises b die Mehrwertsteuer m sind. Mithilfe der folgenden zwei Gleichungen können die MwSt. durch den Bruttopreis ausgedrückt werden:

$$n + m = b$$

$$m = 0,2 \cdot n$$

Die zweite Gleichung wird umgeformt:

$$5 \cdot m = n$$

Nun wird in der ersten Gleichung n durch $5 \cdot m$ ersetzt und diese Gleichung umgeformt:

$$5 \cdot m + m = b$$

$$6 \cdot m = b$$

$$m = \frac{b}{6} = 0,1667 \cdot b$$

Die MwSt. machen also nur 16,67 Prozent des Bruttopreises aus. Von diesem Bruttopreis 20 Prozent abzuziehen, um den Nettopreis zu erhalten, ist daher ein Irrtum.

6.6 Prozentpunkte

Wir haben uns bereits mit der relativen Änderung von absoluten Häufigkeiten beschäftigt - doch wenn wir die relative Änderung von Prozentangaben (Prozentsätzen) betrachten möchten, gilt es, dies nicht mit Prozentpunkten zu verwechseln (und umgekehrt).

Satz 6.1. *[Prozentpunkte (vgl. [8], S. 72 f)] Wenn Kennzahlen in Prozent angegeben und miteinander verglichen werden (oftmals der Vergleich zum Vorjahr), so ist die Änderung zum Vorjahr in Prozentpunkten durch*

$$\text{Kennzahl}_{\text{Jahr}}\% - \text{Kennzahl}_{\text{Vorjahr}}\%$$

gegeben. Man sagt auch: 'Die Kennzahl stieg um ... Prozentpunkte.'

Möchte man jedoch die Änderung in Prozent wissen (und nicht in Prozentpunkten), so muss die Formel zur relativen Änderung herangezogen werden.

Nehmen wir nun wieder ein Beispiel zur Arbeitslosenquote heran. In einer anderen fiktiven Stadt betrug die Arbeitslosenquote im vergangenen Jahr vier Prozent, dieses Jahr jedoch fünf Prozent.

Fehlvorstellung:

Wenn die Arbeitslosenquote im vergangenen Jahr vier Prozent betrug und nun fünf Prozent, so ist die Arbeitslosenquote um ein Prozent gestiegen.

Erklärung des Fehlers:

Bei einer solchen Aussage wurde offensichtlich der Begriff 'Prozentpunkte' nicht berücksichtigt, aber unterbewusst eben dieser gemeint.

Korrekte Betrachtung des Sachverhalts:

Einerseits reicht es, zu bemerken, dass die Arbeitslosenquote um einen **Prozentpunkt** gestiegen sei.

Man kann aber auch sehr wohl die relative Änderung dieser beiden prozentuellen Kennzahlen ermitteln:

$$\frac{5-4}{4} = 0,25 = 25\%$$

Zurecht kann nun behauptet werden, dass die Arbeitslosenquote um 25 Prozent gestiegen ist, was natürlich weitaus dramatischer klingt.

Wenn sich nun zwei Politiker:innen wieder im Gespräch miteinander befinden, so wird die führende Partei von dem einen Prozentpunkt sprechen, während die Opposition sich auf die 25 Prozent stützen würde.

Ironischerweise könnte eine solche Aussage möglicherweise zu einer negativen Rückwirkung führen, da vielen der Unterschied zwischen Prozentpunkten und relativer Änderung nicht geläufig ist und dann die 25 Prozent an den Haaren herbeigeführt wirken könnten.

Wie im vorigen Kapitel darf auch hier nicht geschlossen werden, dass es nun 25 Prozent mehr Arbeitslose gibt - die Arbeitslosenzahl kann sehr wohl gesunken sein, obwohl die Arbeitslosenquote stieg.

Dazu ein kleiner Reminder aus dem bisher Gelernten:

Fehlvorstellung:

Wenn die Arbeitslosenquote steigt, gibt es auch mehr Arbeitslose.

Erklärung des Fehlers:

Nur weil eine relative Häufigkeit steigt, bedeutet dies nicht, dass auch die absolute Häufigkeit steigt.

Korrekte Betrachtung des Sachverhalts:

Angenommen, es gab zuerst 1 000 Arbeitslose und eine Arbeitslosenquote von vier Prozent. Nun gibt es 800 Arbeitslose und eine Arbeitslosenquote von fünf Prozent.

Durch triviale Prozentrechnung erhalten wir im vorigen Jahr einen Grundwert von $\frac{1000}{4} \cdot 100 = 25\,000$ Erwerbspersonen. Aufgrund verschiedener Umstände wanderten nun enorm viele Fachkräfte ab und so gab es im Folgejahr nur noch 16 000 Erwerbspersonen, aber auch nur mehr 800 Arbeitslose. Nun erhalten wir eine Arbeitslosenquote von:

$$\frac{800}{16\,000} = 0,05 = 5\%.$$

Somit sank die Arbeitslosenzahl, jedoch stieg die Arbeitslosenquote. Und auch hier können Sie sich vermutlich bereits vorstellen, auf welche Formulierung sich die führende Partei respektive die Opposition wohl stützen würde.

6.7 Geometrisches Mittel

Wenn die durchschnittliche prozentuelle Änderung ermittelt werden soll, so passieren auch hierbei Fehler.

Wir betrachten wieder einmal den Wert des fiktiven Materials, das pro Kilogramm 100€ wert ist. Über die Jahre veränderte sich der Marktwert wie folgt:

Jahr	Änderung zum Vorjahr
2061	+1 %
2062	-7 %
2063	+10 %
2064	+12 %
2065	+3 %

Fehlvorstellung:

$\frac{1-7+10+12+3}{5} = \frac{17}{5} = 3,4\%$, daher ändert sich der Wert im Schnitt um 3,4% pro Jahr.

Erklärung des Fehlers:

Man darf hier nicht einfach den Durchschnittswert der fünf Prozentzahlen ermitteln (mehr zum arithmetischen Mittel und dessen Anwendungen im nächsten Kapitel). Da es sich um prozentuelle Änderungen handelt, benötigen wir eine spezielle Formel.

Korrekte Berechnung des Sachverhalts:

Satz 6.2. *[Geometrisches Mittel 1 (vgl. [8], S. 36)]*

Seien $x_1, x_2, \dots, x_n$ die Änderungsfaktoren (respektive die relativen Änderungen dann dementsprechend $x_1 - 1, x_2 - 1, \dots, x_n - 1$), so ist

$$x_{geom} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

das geometrische Mittel, welches den durchschnittlichen Änderungsfaktor vom ersten Zeitabschnitt (Zeitpunkt $t=0$) bis zum n -ten Zeitabschnitt angibt.

So berechnen wir den Wachstumsfaktor über $n = 5$ Jahre:

$$\bar{x}_{geom} = \sqrt[5]{1,01 \cdot 0,93 \cdot 1,1 \cdot 1,12 \cdot 1,03} = 1,035739$$

Und daraus lässt sich wiederum die durchschnittliche relative Änderung $1,035739 - 1 = 0,035739 = 3,5739\%$ ermitteln.

Nochmals zum Vergleich die Umkehrung, also das Berechnen des Änderungsfaktors vom Jahr 2060 zu 2065, wenn nur der durchschnittliche Änderungsfaktor pro Jahr gegeben ist:

Fehlvorstellung:

Der jährliche Änderungsfaktor muss nur mit n multipliziert werden.

Erklärung des Fehlers:

Unterhalb der obigen Tabelle wurden die einzelnen Faktoren beim Berechnen des Änderungsfaktors miteinander multipliziert. Wenn wir nun fünfmal den durchschnittlichen Änderungsfaktor annehmen, so haben wir fünfmal denselben Faktor ($a \cdot a \cdot a \cdot a \cdot a = a^5$ und keine Addition ($a + a + a + a + a = 5 \cdot a$)).

Korrekte Betrachtung des Sachverhalts:

$1,035739^5 = 1,1919$, daher kommen wir auch über den durchschnittlichen Änderungsfaktor zum Schluss, dass der Wert über die fünf Jahre um 19,19 Prozent stieg.

Dieselben Denkfallen finden wir vor, wenn anstelle der prozentuellen Änderungen nur der Anfangswert und der Endwert nach n Zeiteinheiten (meist Jahre) gegeben sind - nur dass dann der Änderungsfaktor von der Rechnung $\frac{\text{Anteil}}{\text{Grundwert}}$ her rührt. Um dann den durchschnittlichen Änderungsfaktor zu ermitteln, gibt es auch eine passende Formel:

Satz 6.3. [Geometrisches Mittel 2 (vgl. [8], S. 36)]

Seien ein Endwert und ein Anfangswert gegeben, so ist

$$\bar{x}_{geom} = \sqrt[n]{\frac{\text{Endwert}}{\text{Anfangswert}}}$$

das geometrische Mittel, welches den durchschnittlichen Änderungsfaktor über n Zeitperioden (meistens in Jahren) angibt.

Bei genauerer Betrachtung wird klar, dass dies praktisch nichts anderes ist, als sich den Änderungsfaktor zu berechnen und dann die n-te Wurzel daraus zu ziehen wie bereits vorhin bei der ersten Variante zum geometrischen Mittel.

6.8 Zusammenfassende fiktive Diskussion

Mit allem erlangten Wissen zum Rechnen mit Prozenten und Prozentpunkten widmen wir uns nun einer fiktiven Diskussion zweier Politiker:innen. Dabei ginge es nun um folgende Statistik:

	2040	2041	2042
Einwohner:innenzahl	1 000 000	1 000 000	1 300 000
Arbeitslosenzahl	90 000	90 435	104 000
Arbeitslosenquote	5%	7%	8%

Die Diskussion werde nun im nationalen Fernsehen vor der Wahl live übertragen. Stellen Sie sich dabei vor, dass alle folgenden Aussagen in Sekundenschnelle getätigt werden:

Führende Partei: 'Wir als führende Partei konnten die Steigung der Arbeitslosenquote von 2041 auf 2042 halbieren!'

Opposition: 'Ihr redet das nur schön, ihr seid daran schuld, dass die Arbeitslosenquote von 2041 auf 2042 um 14 Prozent gestiegen ist!'

Führende Partei: 'Und ihr redet es schlecht - die ist ja nur um einen Prozentpunkt gestiegen!'

Opposition: 'Im Schnitt stieg die Quote in den vergangenen beiden Jahren um 1,5 Prozent und wir haben 15 Prozent mehr Arbeitslose als letztes Jahr!'

Führende Partei: 'Ja, aber die Einwohner:innenzahl ist ja auch gestiegen! Während die um ganze 30 Prozent stieg, ist die Zahl der Arbeitslosen nur um 15 Prozent gestiegen.'

Opposition: 'Na wenn jetzt jede zweite Person ohne Arbeit daherkommt, dann sind wir beim nächsten Problem, dass Sie verursacht haben!'

Bei einem solchen Live-Gespräch kämen hier unfassbar viele Informationen auf. Könnten Sie in Live-Geschwindigkeit die korrekten von den inkorrekten Aussagen unterscheiden?

In diesem Abschnitt geht es hauptsächlich darum, aufzuzeigen, dass viele Falschaussagen oder Missinterpretationen in einer hohen Geschwindigkeit verbreitet werden können, sodass die Richtigstellung im Nachhinein kaum noch zu bewerkstelligen ist - durch Meinungsverbreitung über Social Media können weitere Schäden durch solche fehlerhaften Aussagen kaum behoben werden.

Bemerkung: ist Ihnen aufgefallen, dass in der Statistik und der Diskussion die Einwohner:innenzahl verwendet wurde, obwohl diese nicht direkt im Zusammenhang mit der Arbeitslosenquote steht? Dafür wird die Zahl der Erwerbstätigen benötigt, die in dieser Statistik leider nicht gegeben ist. Denn mit der Einwohner:innenzahl kann man hier nichts beweisen, höchstens vermuten.

Sehen wir uns nun die einzelnen mathematischen Behauptungen hinter den Aussagen an:

- 'Wir konnten die Steigung der Arbeitslosenquote von 2041 auf 2042 halbieren':
Korrekt, da die Steigung der Arbeitslosenquote von 2040 auf 2041 zunächst $7\% - 5\% = 2$ Prozentpunkte betrug (also von fünf auf sieben Prozent), von 2041 auf 2042 schließlich $8\% - 7\% = 1$ Prozentpunkt. Somit konnte die Steigung von 2 Prozentpunkten auf einen Prozentpunkt halbiert werden.
- 'Die Arbeitslosenquote ist von 2041 auf 2042 um 14 Prozent gestiegen':
Korrekt, da die relative Änderung $\frac{8\% - 7\%}{7\%} = 0,14 = 14\%$ betrug.
- 'Die Arbeitslosenquote ist von 2041 auf 2042 um einen Prozentpunkt gestiegen':
Korrekt, da $8\% - 7\% = 1$ Prozentpunkt.
- 'Im Schnitt stieg die Quote in den vergangenen beiden Jahren um 1,5 Prozent':
Inkorrekt, da $\sqrt{\frac{108}{105}} = 1,014185105674 = 101,4185\%$. Somit stieg die Arbeitslosenquote über die zwei Jahre im Schnitt um 1,4185 Prozent. Man darf also nicht $\frac{8-5}{2} = 1,5$ Prozentpunkte berechnen.
- 'Wir haben 15 Prozent mehr Arbeitslose als letztes Jahr':
Korrekt, da $\frac{104\,000 - 90\,435}{90\,435} = 15\%$.
- 'Die Einwohner:innenzahl stieg um 30 Prozent, während die Zahl der Arbeitslosen nur um 15 Prozent stieg'

Zwar gilt: $\frac{1\,300\,000-1\,000\,000}{1\,000\,000} = 30\%$ und die 15 Prozent wurden bereits nachgewiesen, allerdings lässt sich im Gegensatz zur (nicht vorhandenen) Zahl der Erwerbstätigen keine sichere Schlussfolgerung ziehen, da sich die Einwohner:innenzahl ganz anders als die Zahl der Erwerbstätigen verändern kann.

- 'Jede zweite zugewanderte Person ist arbeitslos':
Inkorrekt. Nur weil die Einwohner:innenzahl um 30 Prozent stieg und die Arbeitslosenzahl um 15 Prozent, heißt das längst nicht, dass jede:r zweite Zuwanderer oder Zuwanderin arbeitslos ist, da einerseits die Anzahl der Zugewanderten viel größer sein könnte als jene der Arbeitslosen - andererseits gilt auch wie im vorigen Punkt: wir können keine sichere Schlussfolgerung ziehen.

Das geometrische Mittel wurde vor dem nächsten Kapitel 'Lagemaße' angeschnitten, da es bereits auf den vorherigen Seiten essenziell war. Nun zu weiteren Lageparametern.

7 Lagemaße

Obgleich der Begriff 'Lageparameter' nicht sehr geläufig ist, sind einige davon doch allseits bekannt - und dennoch lauern hier weitere Gefahren der Manipulation und Denkfallen.

7.1 Arithmetisches Mittel

Umgangssprachlich '(Durch-)Schnitt' genannt, ist das arithmetische Mittel ein wichtiger Bestandteil der Statistik.

Satz 7.1. *[Arithmetisches Mittell 2 (vgl. [52], S. 23 f)]*

Bei n vorliegenden Einzelwerten ist

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

das arithmetische Mittel.

So einfach diese Formel wirken mag, so einfach kann sie falsch angewandt werden.

Seien nun in einer Stadt zwei Fabriken. In einer Fabrik herrscht ein Stundenlohn von 8€ pro Stunde, in der anderen sind es 15€ pro Stunde.

Fehlvorstellung:

Der durchschnittliche Stundenlohn der Fabriks-Arbeiter:innen in dieser Stadt liegt bei 11,5€ pro Stunde.

Erklärung des Fehlers:

Es wurde das arithmetische Mittel anhand der beiden Werte berechnet:

$$\bar{x} = \frac{8+15}{2} = 11,5\text{€}$$

Doch diese Berechnung beruht auf der Annahme, dass in beiden Fabriken gleich viele Personen arbeiten würden. Stellen Sie sich vor, dass in einer Fabrik 1 000 Personen arbeiten und in der anderen lediglich 200.

Korrekte Betrachtung des Sachverhalts:

Satz 7.2. *[Gewichtetes arithmetisches Mittel (vgl. [52], S. 24f)]*

Wenn wir k verschiedene Merkmalsausprägungen vorliegen haben (also k verschiedene, vorkommende Werte - von denen viele Werte auch öfters vorkommen können), so ist das gewichtete arithmetische Mittel wie folgt zu berechnen:

$$\bar{x} = \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n} = \frac{1}{n} \cdot \sum_{i=1}^k x_i \cdot n_i$$

Wobei x_i der i -te Wert ist und n_i die Anzahl, wie oft der jeweilige Wert x_i vorkommt.

Nun zur korrekten Berechnung:

$$\bar{x} = \frac{8 \cdot 1000 + 15 \cdot 200}{1200} = 9,17\text{€}$$

Aus dem einstig fälschlicherweise angenommenen Stundenlohn von 11,5€ wurden nun 9,17€.

Der Unterschied zwischen dem 'normalen' und dem gewichteten arithmetischen Mittel ist, dass wir beim 'normalen' arithmetischen Mittel beispielsweise siebenmal vorkommende Werte auch siebenmal in der Formel aufschreiben, wohingegen wir beim gewichteten arithmetischen Mittel bereits die Anzahl berücksichtigen.

Das gewichtete arithmetische Mittel können wir auch mittels relativer Häufigkeiten berechnen:

$$\bar{x} = x_1 \cdot h_1 + x_2 \cdot h_2 + \dots + x_k \cdot h_k$$

Wobei h_i die relative Häufigkeit des i -ten Wertes ist. Die Formel ergibt sich durch Umformen der vorherigen Formel zum gewichteten arithmetischen Mittel:

$$\begin{aligned} \bar{x} &= \frac{x_1 \cdot n_1 + x_2 \cdot n_2 + \dots + x_k \cdot n_k}{n} = \frac{x_1 \cdot n_1}{n} + \frac{x_2 \cdot n_2}{n} + \dots + \frac{x_k \cdot n_k}{n} = x_1 \cdot \frac{n_1}{n} + x_2 \cdot \frac{n_2}{n} + \dots + x_k \cdot \frac{n_k}{n} = \\ &= x_1 \cdot h_1 + x_2 \cdot h_2 + \dots + x_k \cdot h_k \end{aligned}$$

7.2 Arithmetisches Mittel von Skalen

Unter anderem im Schulunterricht wird oft ein Notenschnitt angegeben, um die Leistung der Klasse zu reflektieren - doch wie aussagekräftig ist dieser?

Hierzu ein Beispiel, ausgehend vom folgenden Notenschlüssel:

Punkte	Note
22-24	Sehr gut
19,5-21,5	Gut
16-19	Befriedigend
12,5-15,5	Genügend
0-12	Nicht genügend

Nun gab es bei einer Schularbeit sechs Nicht genügend, acht Genügend, fünf Befriedigend, vier Gut und zwei Sehr gut.

Fehlvorstellung:

Der Notenschnitt ist aussagekräftig und wird daher einfach mit dem arithmetischen Mittel berechnet:

$$\bar{x} = \frac{5 \cdot 6 + 4 \cdot 8 + 3 \cdot 5 + 2 \cdot 4 + 1 \cdot 2}{6 + 8 + 5 + 4 + 2} = 3,48$$

Erklärung des Fehlers:

Es wird so getan, als wäre jedes 'Befriedigend' gleichgut und jedes 'Nicht genügend' gleichschlecht, wobei zwischen den Punkteergebnissen aller 'Nicht genügend' eine große Spannweite liegen kann.

Korrekte Betrachtung des Sachverhalts:

Empfehlenswerter wäre das Berechnen des Punkteschnitts anstelle des Notenschnitts.

Dazu ein Beispiel: sei bei gleichbleibendem Notenspiegel immer die höchste Punktezahl innerhalb eines Notengrades erzielt worden, so ist der Punkteschnitt

$$\bar{x} = \frac{5 \cdot 12 + 4 \cdot 15 + 3 \cdot 19 + 2 \cdot 21 + 1 \cdot 24}{6 + 8 + 5 + 4 + 2} = 17,12 \text{ Punkte}$$

In einem anderen Fall sei bei gleichbleibendem Notenspiegel immer die niedrigste Punktezahl innerhalb eines Notengrades erzielt worden, so beträgt der Punkteschnitt

$$\bar{x} = \frac{5 \cdot 0 + 4 \cdot 12,5 + 3 \cdot 16 + 2 \cdot 19,5 + 1 \cdot 22}{6 + 8 + 5 + 4 + 2} = 12,08 \text{ Punkte}$$

Sie sehen, dass wir hier einen Notensprung von einem 'Befriedigend' auf ein 'Nicht genügend' vorliegen haben. Natürlich ist das Eintreten beider Extremfälle unwahrscheinlich, jedoch zeigt es auf, wie stark die Schwankung bei - nach außen hin - gleichen Notenspiegeln aussehen kann. Dabei sei noch hinzugefügt, dass die Notenschlüssel bis zu einem gewissen Grad beinahe willkürlich gezogen werden können und somit den Notenspiegel ohnehin im Vorhinein schon beeinflussen.

Mitunter fiel Ihnen bereits auf, dass auch eine Notenskala speziell ist - der Abstand zwischen der Note '1' und der Note '3' ist (punktemäßig) nicht derselbe wie zwischen der Note '3' und der Note '5' - anders als die Differenz zwischen Geldbeträgen (die Differenz zwischen 10€ und 15€ respektive 21€ und 26€ ist dieselbe). Man sagt auch nicht 'mein Notengrad hat sich um 25 Prozent verschlechtert', während dies bei Umsätzen durchaus erwähnt werden kann.

Es mag weiterhin Geschmackssache sein, ob der Notenschnitt oder der Punkteschnitt betrachtet wird - seien Sie sich lediglich der Tatsache bewusst, dass der Notenschnitt wichtige Informationen auf der Strecke liegen lässt.

7.3 Harmonisches Mittel

Eher unbekannt, aber im Alltag durchaus relevant ist das harmonische Mittel.

Wir fahren mit dem Auto von Ortschaft A nach Ortschaft B - die Distanz zwischen den beiden Orten beträgt 150 km. Am Hinweg fuhren wir mit einer durchschnittlichen Geschwindigkeit von 100 km/h, am Rückweg waren es 80 km/h.

Fehlvorstellung:

Die Durchschnittsgeschwindigkeit betrug 90 km/h.

Erklärung des Fehlers:

Da der Rückweg länger als der Hinweg dauerte (mit 80 km/h anstelle von 100 km/h), fuhren wir auch über einen längeren Zeitraum mit einer Geschwindigkeit von 80 km/h, als wir es mit 100 km/h taten.

Korrekte Betrachtung des Sachverhalts:

Satz 7.3. *[Harmonisches Mittel 1 (vgl. [52], S. 34f)]*

Das harmonische Mittel hilft, Durchschnittswerte verschiedener Verhältniszahlen $\frac{n_i}{x_i}$ (beispielsweise bei km/h oder €/Stück) zu berechnen:

$$\bar{x}_{\text{harm}} = \frac{n_1 + n_2 + \dots + n_k}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \dots + \frac{n_k}{x_k}}$$

Nun zur Berechnung der Durchschnittsgeschwindigkeit. Ein Blick auf den Nenner zeigt, dass dort die jeweiligen Zeitdauern und in Folge dessen auch die Gesamtzeit berechnet werden. Im Zähler wird die Gesamtstrecke ermittelt - und so ergibt sich die durchschnittliche Geschwindigkeit:

$$\bar{x}_{\text{harm}} = \frac{150+150}{\frac{150}{100} + \frac{150}{80}} = 88,89 \text{ km/h}$$

Ein weiteres Beispiel dazu:

Nehmen wir an, dass uns bei einem Erntedankfest an zwei aufeinanderfolgenden Jahren jeweils ein Budget von 20€ zur Verfügung steht, welches wir auch zur Gänze ausgeben. Davon kaufen wir uns immer nur Softdrinks. Im ersten Jahr kostet ein solcher 4€, im Jahr darauf 5€.

Fehlvorstellung:

Der Durchschnittspreis der erworbenen Getränke lag bei 4,50€.

Erklärung des Fehlers:

Wie zuvor ist auch hier zu beachten, dass wir im ersten Jahr mehr Getränke um das selbe Geld bestellen konnten als im darauffolgenden Jahr.

Korrekte Berechnung des Sachverhalts:

$$\bar{x}_{\text{harm}} = \frac{20+20}{\frac{20}{4} + \frac{20}{5}} = 4,44\text{€}$$

Sie sehen, dass diese Berechnung sehr wohl auch für stetige Werte durchführbar ist, nicht nur diskrete Werte.

Bemerkung: alternativ kann man sich die Durchschnittswerte über die relativen Häufigkeiten errechnen, wenn man beispielsweise weiß, wie viel Prozent der Autofahrt mit 100 km/h zurückgelegt wurden und welcher Anteil mit 80 km/h (oder wie viel Prozent der konsumierten Getränke zu einem Preis von 4€ respektive 5€ erworben wurden.)

7.4 Median

Der Median, auch Zentralwert genannt, ist ein weiterer Mittelwert, welcher regelmäßig zur Anwendung kommt.

Satz 7.4. *[Median (vgl. [8], S. 29)]*

Zunächst müssen alle Werte der Datenreihe aufsteigend sortiert werden, anschließend können die Indizes den Werten nacheinander zugewiesen werden, sodass nun der Größe nach die Werte $x_1, x_2, \dots, x_n$ aufgelistet sind. Dann ist

$$x_{ZW} = \begin{cases} x_{(n+1)/2} & n \text{ ungerade} \\ \frac{x_{n/2} + x_{n/2+1}}{2} & n \text{ gerade} \end{cases}$$

der Median der Datenreihe.

Umgangssprachlich erklärt ist der Median also jener Wert, der in der Mitte steht, wenn alle Werte aufsteigend sortiert sind. Bei einer geraden Anzahl an Werten muss das arithmetische Mittel der beiden Werte in der Mitte berechnet werden, um den Median zu erhalten.

Oft ist das arithmetische Mittel empfindlich gegenüber sogenannten Ausreißern - beispielsweise, wenn auf das Einkommen weniger Reicher und vieler Armer geblickt wird.

So seien in der fiktiven Fabrik, die bereits beim gewichteten arithmetischen Mittel mit dem Stundenlohn vorkam, 1 000 Mitarbeiter:innen angestellt. Die Fabrik präsentiert nun stolz in einer Anzeige, dass sie den Durchschnittslohn auf 20€ erhöhen, um mitarbeiter:innenfreundlicher zu sein.

Fehlvorstellung:

Die meisten Personen werden in der Fabrik 20€ verdienen.

Erklärung des Fehlers:

Das Problem ist, dass nur wenige Mitarbeiter:innen viel verdienen müssen, um den Schnitt heben zu können.

Korrekte Betrachtung des Sachverhalts:

Wir benötigen nähere Infos zu den Stundenlöhnen. Nun erhalten 740 'einfache' Arbeiter:innen einen Stundenlohn von 10€, 250 Nachtschicht-Arbeiter:innen einen Stundenlohn von 16€ und 10 Personen in Führungspositionen einen Stundenlohn von 860€. Nun reihen wir die 1 000 Werte aufsteigend der Reihe nach und an 500. bzw. 501. Stelle steht natürlich eine Person mit einem Stundenlohn von 10€, wodurch auch der Median 10€ beträgt.

Also verdient der oder die 'Durchschnittsmitarbeiter:in' lediglich 10€ pro Stunde, während zehn Top-Manager:innen 860€ pro Stunde erbeuten. Die angegebenen 12€ pro Stunde sind demnach leider irreführend.

Der Median repräsentiert die Durchschnittsperson respektive den oder die Durchschnitts-Merkmalsträger:in besser als das arithmetische Mittel.

7.5 Modus

Der Modus wird auch Modalwert genannt und ist jene Merkmalsausprägung, welche am häufigsten vorkommt (vgl. [52], S. 26). Beim Modus können auch nominale Daten verwendet werden, beispielsweise Eigenschaften als Worte.

Auch bei der Angabe des häufigsten Wertes kann manipuliert werden oder ein umgangssprachlicher Fehler begangen werden. So erklärt ein fiktives Unternehmen anhand des Modus der Antworten einer Umfrage stolz: 'Die meisten unserer Mitarbeiter:innen sind zufrieden bei uns!'

Fehlvorstellung:

Die meisten Mitarbeiter:innen sind zufrieden, da 'Zufrieden' die häufigste Antwort war.

Erklärung des Fehlers:

'Zufrieden' ist der häufigste Wert, also der Modus. Das bedeutet jedoch nicht, dass die absolute Mehrheit der Werte (also über 50 Prozent) 'Zufrieden' sind. Was 'die meisten' bedeutet, ist hier also in Frage zu stellen.

Korrekte Betrachtung des Sachverhalts:

Die Zufriedenheits-Umfrage ergab folgendes Resultat:

Antwort	Anzahl
Sehr zufrieden	4
Zufrieden	30
Wenig zufrieden	29
Nicht zufrieden	27
Summe	90

Tatsächlich sind 30 der 90 Werte, also $\frac{30}{90} = 0,3333 = 33,33\%$ der Antworten 'Zufrieden'. 33 Prozent sind nicht 'die meisten', wenn man nach absoluter Mehrheit geht. Dies dürfte man erst ab mehr als 50 Prozent behaupten. Es gibt lediglich keinen Wert, der häufiger vorkam, dies spricht aber nicht automatisch für die Mehrheit. Denn so ergibt sich hingegen eine Mehrheit von 62,22 Prozent, welche die Umfrage mit einer negativen Antwortmöglichkeit abgeschlossen hatte.

Dass man auch bei solchen Umfragen bereits durch die Fragestellung und Antwortmöglichkeiten manipulieren kann, ist keinerlei Mysterium. Mehr dazu finden Sie später in der Arbeit, wenn wir Umfragen und eigene Datenerhebungen in den Fokus legen.

7.6 Quantil/Perzentil

Zwei Schüler:innen möchten es beim Känguru-Wettbewerb österreichweit in die jeweiligen Top 100 der zugewiesenen Kategorie schaffen - eine:r in der fünften, der oder die andere in der 13. Schulstufe. Angenommen, beide schaffen es in die Top 100: sind beide Leistungen gleichermaßen beeindruckend?

2023 gab es 22 016 Teilnehmende in der fünften Schulstufe, hingegen 133 Teilnehmende der 13. Schulstufe (vgl. [41]) - ein aussagekräftigerer Vergleich der Leistung wäre das Perzentil.

Das Perzentil bzw. der Prozentrang ist der Anteil der Werte einer Urliste, der kleiner oder gleich einem bestimmten Wert ist (vgl. [74]).

So gibt ein 80%-Perzentil an, dass in einer Datenreihe (mindestes) 80 Prozent aller vorliegenden Werte unterhalb oder auf dieser Grenze liegen.

In unserem Beispiel bedeutet das, dass aufgrund der Teilnehmendenzahl die Top 100 unterschiedlich beeindruckend sind und wir uns daher die Perzentile ansehen.

Der Einfachheit halber seien beide Teilnehmenden auf dem 100. Platz der jeweiligen Schulstufe.

In der fünften Schulstufe liegen $22\,016 - 100 = 21\,916$ Teilnehmende unterhalb der Top 100 (vorausgesetzt, dass an der 100. Stelle keine Punktegleichheit herrscht). Wie groß der prozentuelle Anteil der Teilnehmenden außerhalb der Top 100 ist, kann mittels der relativen Häufigkeit ermittelt werden:

$\frac{21\,916}{22\,016} = 0,9954578 = 99,54578\%$ haben also weniger oder gleich viele Punkte wie der 100. Platz. Man kann demnach im 99%-Perzentil nicht in die Top 100 vordringen.

Währenddessen sieht das Perzentil des oder der Schüler:in in der 13. Schulstufe anders aus: $133 - 100 = 33$ haben schlechter abgeschlossen und somit ergibt sich, dass $\frac{33}{133} = 0,24812 = 24,812\%$ weniger oder gleich viele Punkte erzielt haben. Während der oder die eine es unter die (gerundet) besten 75 Prozent geschafft hat, hat der oder die andere Schüler:in sogar die Top 0,5 Prozent geknackt. Was ist beeindruckender?

Haben Sie gerade bemerkt, wie die 75 Prozent formuliert waren? Man hätte genauso gut sagen können, dass 75 Prozent der Teilnehmenden mehr Punkte erreichten - bzw. genau genommen sogar über 75 Prozent. 'Top 75 Prozent' klingt zumindest nicht ganz so schlimm, wie 'über 75 Prozent waren besser' oder gar 'unter den schlechtesten 25 Prozent'.

Aus der Perspektive des oder der Schüler:in der 13. Schulstufe jedoch empfehle es sich, weiterhin nur von dem Erzielen der Top 100 und nicht der gesamten Teilnehmendenzahl oder dem Prozentrang zu sprechen.

Umgekehrt ginge das genauso, falls man etwas schlechtreden möchte: anstatt bei einem 80%-Perzentil zu sagen, dass nur 20 Prozent höhere Werte erzielten, kann man auch sagen: 'diese Person gehört zu den schlechtesten 80 Prozent', ohne zu erwähnen, dass man sich dabei ja schon auf der 80%-Marke befindet - es klingt einfach verheerender, wenn man in einem Topf mit den schlechtesten 80 Prozent ist. 'Unter den schlechtesten 80%' klingt negativ und so kann man - ohne zu lügen - mit einer simplen Formulierung die Meinungen vieler Menschen beeinflussen.

Tatsächlich liegen manchmal nur bestimmte Perzentil-Angaben vor, woher man nur in grobe Bereiche trennen kann. Wenn man beispielsweise erfährt, in der Kategorie der besten zehn bis 30 Prozent zu liegen, so ist das ein breiteres Spektrum. In einem anderen Fall kann 'unter den besten 30 Prozent' auch bedeuten, dass es nur drei Kategorien gibt und man einfach in der Kategorie 'besten 30 Prozent' liegt. Vielleicht läge man

sogar beim 99%-Perzentil, wenn man auf die genaueren Daten zugreifen könnte.

Somit liegt auch hier ein dehnbarer und der Manipulation dienlicher Aspekt vor: welche Perzentile nehme ich heran, welche Klassenbreiten wende ich dabei an (in 1%-Schritten, 10%-Schritten, 25%-Schritten oder gar unterschiedliche Schritte/Klassenbreiten)?

Anmerkung: zu beachten ist, dass höhere Perzentile nicht automatisch bessere Werte sind. Beim systolischen Blutdruck oder Anzahl der Parkstrafen wird ein 95%-Perzentil nicht erstrebenswert sein.

8 Streumaße

In diesem Abschnitt werden die Manipulationsgefahren und Denkfallen mancher Streumaße beleuchtet.

8.1 Spannweite

Satz 8.1. *[Spannweite (vgl. [8], S. 39)]*

In einer Urliste ist die Differenz zwischen dem größten Wert x_{max} und dem kleinsten Wert x_{min}

$$Span = x_{max} - x_{min}$$

die sogenannte Spannweite (oder auch Spannbreite).

Sogar mit etwas Simplen wie der Spannweite kann man schnell unterbewusst in die Irre geführt werden. Wenn beispielsweise ein:e Verkäufer:in behauptet, dass die Preise der Schuhe eine Spannweite von 20€ bis 250€ haben, so könnten gleich zwei Missverständnisse vorliegen: zuerst könnte damit gemeint sein, dass die Spannweite der Spannweite groß sei, dabei beträgt sie $250 - 20 = 230$ € und nicht 20€ bis 250€ - man könnte eher sagen: 'Dank unserer großen Spannweite der Preise finden alle ein passendes Paar Schuhe im Wert von 20 bis 250€'. Nun aber zur größeren Fehlvorstellung.

Fehlvorstellung:

Die Werte der Merkmalsträger sind zwischen x_{max} und x_{min} in etwa gleichverteilt, also findet man sicherlich einige Modelle in jeder Preisklasse.

Erklärung des Fehlers:

Natürlich ist den meisten bewusst, dass dies nicht der Fall sein muss, jedoch suggerieren uns dies x_{max} und x_{min} unterbewusst. Tatsächlich können x_{max} und x_{min} bloß starke Ausreißer sein.

Korrekte Betrachtung des Sachverhalts:

In solchen Fällen wäre es praktischer, eine Urliste zur Verfügung zu haben (oder hier eine Preisliste). Es kann sich herausstellen, dass das Paar um 20€ lediglich ein aussortiertes Modell ist und alle anderen Kollektionen bei 110€ starten, dennoch darf sich das Geschäft mit einer großen Spannweite rühmen.

8.2 Varianz vs. durchschnittliche Abweichung

Die Varianz kann auch mittlere quadratische Abweichung genannt werden, da (zumindest für die Grundgesamtheit) das arithmetische Mittel von allen quadrierten Abweichungen der Werte ermittelt wird:

Satz 8.2. *[Varianz für die Grundgesamtheit]*

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n} \quad (\text{vgl. [8], S. 41})$$

Satz 8.3. *[Varianz für eine Stichprobe]*

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} \quad (\text{vgl. [8], S. 41})$$

Eine plausible Erklärung dafür, weshalb die Summe der quadratischen Abweichungen bei der Varianz für eine Stichprobe durch $n - 1$ gemittelt wird, liefert [22] (S. 65):

Da $\sum(x_i - \bar{x}) = 0$ gilt, ist z. B. die letzte Abweichung $x_n - \bar{x}$ bereits durch die ersten $n - 1$ bestimmt. Somit variieren nur $n - 1$ Abweichungen frei und man mittelt deshalb, indem man durch die Anzahl $n - 1$ der sogenannten Freiheitsgrade dividiert. Diese Plausibilitätserklärung wird im Rahmen der induktiven Statistik formalisiert.

Satz 8.4. *[Standardabweichung für die Grundgesamtheit]*

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} \quad (\text{vgl. [8], S. 43})$$

Satz 8.5. *[Standardabweichung für eine Stichprobe]*

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1}} \quad (\text{vgl. [8], S. 43})$$

Fehlvorstellung:

Wenn die Varianz σ^2 die mittlere quadratische Abweichung ist, dann ist die Standardabweichung die durchschnittliche Abweichung.

Erklärung des Fehlers:

Dieser Fehler ist bei enorm vielen Plattformen vorzufinden und wird von vielen Mathematiker:innen fälschlicherweise verbreitet. Dabei ist die korrekte Berechnung der durchschnittlichen absoluten Abweichung ganz einfach: wir summieren die Beträge der

Abweichungen und dividieren durch n , also ermitteln wir das arithmetische Mittel der Abweichungen.

Korrekte Berechnung des Sachverhalts:

Bei genauerer Betrachtung der Formeln fällt bei s ohnehin sofort auf, dass im Nenner $n - 1$ steht und bei einer Berechnung von einem arithmetischem Mittel von n Werten müsste dort n stehen. Doch für σ^2 bedarf es einer genaueren Erklärung.

Satz 8.6. [*Durchschnittliche absolute Abweichung*]

$$DAA = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \dots + |x_n - \bar{x}|}{n} \quad (\text{vgl. [52] S. 32})$$

bei n Werten.

Tatsächlich ist die durchschnittliche Abweichung weitaus weniger geläufig und seltener verwendet als die Standardabweichung, welche in verschiedenen Bereichen der Statistik und Wahrscheinlichkeitsrechnung vorzufinden (und wichtig) ist (vgl. [52] S. 32). Anschaulich interpretieren kann man die Standardabweichung beispielsweise bei den normalverteilten Daten (die Gauß'sche Glockenkurve ist nicht nur in allen Schulen bekannt), da dort beispielsweise etwa 68 Prozent der Merkmalsausprägungen im Intervall $[\bar{x} - \sigma; \bar{x} + \sigma]$ liegen (vgl. [8], S. 43).

Vielen stellt sich die Frage, weshalb die Standardabweichung nicht die durchschnittliche absolute Abweichung angibt. Die Standardabweichung gibt zwar Auskunft über die Abweichungen - je größer die Standardabweichung, umso stärker schwanken im Wesentlichen die Werte um das arithmetische Mittel. Doch eine Standardabweichung von 3,1 bedeutet nicht, dass die einzelnen Werte im Schnitt um 3,1 vom arithmetischen Mittel abweichen.

Wir rechnen das anhand eines einfachen Beispiels mit drei Werten als Grundgesamtheit: $x_1 = 5, x_2 = 6, x_3 = 7$

Wir berechnen die Standardabweichung:

$$s = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2}{n}} = \sqrt{\frac{(5-6)^2 + (6-6)^2 + (7-6)^2}{3}} = \sqrt{\frac{1+0+1}{3}} = 0,8165$$

Im Vergleich dazu die durchschnittliche absolute Abweichung:

$$DAA = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}|}{n} = \frac{|5-6| + |6-6| + |7-6|}{3} = \frac{1+0+1}{3} = \frac{2}{3} = 0,667$$

9 Korrelation vs. Kausalität

Ein äußerst spannendes Thema liegt in der Untersuchung von Korrelation und Kausalität, dafür ist es unerlässlich, den Unterschied zwischen beiden zu wissen. In diesem Kapitel wird der lineare Zusammenhang näher betrachtet, wobei die Manipulation auch für andere Untersuchungen (Chi-Quadrat-Test, ...) hinsichtlich der Korrelation vorkommen kann.

Wenn nun also für wachsende Werte eines Merkmals X gilt, dass auch die Werte eines Merkmals Y wachsen (Korrelation), so kann es nahe liegen, dass es einen Zusammenhang zwischen den beiden Größen gibt (vgl. [22], S. 126). Am besten lässt sich das durch eine Punktwolke visualisieren und beschreiben, bei der sich die Punktwolke im Streudiagramm 'in eine Richtung zu bewegen' scheint.

Satz 9.1. *Der Bravais-Pearson-Korrelationskoeffizient ist ein Maß für die Stärke eines Zusammenhangs zweier Merkmale X und Y mit den Werten x_i und y_i und wird mit folgender Formel berechnet:*

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

mit $-1 \leq r \leq 1$ ([22], S. 126)

Dabei wird diese Formel bei intervallskalierten Daten angewandt. Ein positiver Korrelationskoeffizient bedeutet: 'je höher der x-Wert, umso höher der y-Wert', ein negativer Korrelationskoeffizient bedeutet: 'je höher der x-Wert umso kleiner der y-Wert'. Wichtig ist es dabei, zu beachten, dass $r = 0,06$ zwar ein positiver Korrelationskoeffizient ist, aber ein Zusammenhang nicht stark ausgeprägt ist (vgl. [48]).

Ein Beispiel dazu wäre die Vermutung unter drei Freund:innen: 'Je älter, umso höheres monatliches Nettoeinkommen'. So werden die Daten zusammengetragen:

Name	Alter (in Jahren)	Einkommen (in €)
Madeleine	25	1 500
Patrick	27	2 000
Marcus	35	3 940
Arithmetisches Mittel	$\bar{x} = 29$	$\bar{y} = 2 480$

Nun wird der Korrelationskoeffizient berechnet, wobei x das Alter und y das Einkommen ist:

$$\begin{aligned}
 r &= \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = \\
 &= \frac{(25-29) \cdot (1\,500-2\,480) + (27-29) \cdot (2\,000-2\,480) + (35-29) \cdot (3\,940-2\,480)}{\sqrt{((25-29)^2 + (27-29)^2 + (35-29)^2) \cdot ((1\,500-2\,480)^2 + (2\,000-2\,480)^2 + (3\,940-2\,480)^2)}} = \\
 &= \frac{4 \cdot (-980) + (27-29) \cdot (2\,000-2\,480) + (35-29) \cdot (3\,940-2\,480)}{\sqrt{((25-29)^2 + (27-29)^2 + (35-29)^2) \cdot ((1\,500-2\,480)^2 + (2\,000-2\,480)^2 + (3\,940-2\,480)^2)}} = \\
 &= \frac{13\,640}{13\,640,17595} = 0,9999871
 \end{aligned}$$

Es liegt ein starker Zusammenhang zwischen dem Alter und dem Einkommen vor.

In einer anderen Situation liegt nun ein anderer Korrelationskoeffizient ($r = 0,06$) vor.

Fehlvorstellung:

$r = 0,06$, also wurde nachgewiesen, dass mit steigendem x auch y steigt.

Erklärung des Fehlers:

Ein häufiger Fehler liegt bereits darin, einen Korrelationskoeffizienten überzubewerten (beispielsweise $r = 0,06$ als einen signifikanten Zusammenhang zu bezeichnen).

Korrekte Betrachtung des Sachverhalts:

Es handelt sich um keinen signifikanten Zusammenhang, denn es lässt sich sagen:

- $|r| < 0,3$: kein signifikanter Zusammenhang
- $0,3 \leq |r| < 0,5$: schwacher Zusammenhang
- $0,5 \leq |r| < 0,8$: mittelstarker Zusammenhang
- $|r| > 0,8$: starker Zusammenhang

(vgl. [48]).

Bemerkung: wie man zwei Größen auf Korrelation überprüft, hängt von der Skalierung der Daten ab. In diesem Thema ging es um zwei intervallskalierte Größen (auch die

gängigere oder in der Schule sowie Universität am ehesten betrachtete angewandte Situation), wie es häufig vorkommt und auch einfach in Microsoft Excel zu bedienen ist.

Es gibt wie gesagt auch weitere Tests wie den Chi-Quadrat-Test und andere, je nach Skalierung der Daten (ordinal/metrisch, nominal/metrisch, ...) - doch dies ist ein äußerst umfangreiches Fachgebiet und dem Thema 'Denkfallen und Manipulationen in der Statistik' nicht weiter dienlich - wichtig war die Art der Denkfalle/Manipulation.

9.1 Verwechseln von Ursache und Wirkung

Eine Korrelation bedeutet nicht automatisch eine Kausalität.

Wenn in Staaten mit höheren Geburtenraten auch mehr Störche leben, so mag es eine Korrelation geben, aber wir gehen nicht davon aus, dass die Störche die Babys bringen - die Kausalität ist nicht automatisch erwiesen. Ein Zusammenhang zeigt auch nicht automatisch, welches Merkmal das andere beeinflusst, es besteht lediglich ein Zusammenhang. Wenn beispielsweise Menschen mit höherem monatlichen Einkommen auch mehr Geld im gastronomischen Bereich ausgeben und so ein Zusammenhang besteht, lässt sich noch nicht sagen, ob sie aufgrund ihres Einkommens mehr ausgeben oder ob sie aufgrund ihrer Neigung zur Gastronomie mehr Geld verdienen möchten, um sich dies leisten zu können.

Ebenso kann es auch Störfaktoren geben, die einen Zusammenhang der beiden Merkmale vorgaukeln, während der nicht erkannte Störfaktor im 'Verborgenen' die beiden Größen beeinflusst. Mehr dazu finden Sie im Kapitel Simpson-Paradoxon. Dies ist wichtig zu begreifen - manchmal gibt es Zusammenhänge, die eine Kausalität vermuten lassen, die aber einen Störfaktor übersehen - eine der wichtigsten Thematiken dieser Arbeit.

Wir erinnern uns: wenn zwei Datensätze miteinander korrelieren, so muss man sich dennoch hüten, daraus sofort eine Kausalität zu ziehen - aber auch, Ursache und Wirkung nicht zu verwechseln.

Ein Beispiel dafür zeigt sich 1809 bei einer Seuche in Düsseldorf (vgl. [6], S. 51 ff). Bei der Untersuchung der Patient:innen fiel auf, dass Personen mit vielen Läusen signifikant weniger Fieber haben als Personen mit wenigen Läusen. Die Zeitung empfahl den Kontakt zu Läusen, das Schlafen in verlausten Betten und Ähnliches (ohne daran zu denken, dass vermutlich einfach eine mit den Läusen zusammenhängende Krankheit überwunden wurde und dadurch das Fieber sank statt durch Läuse).

Fehlvorstellung:

Wenn zwei Variablen miteinander korrelieren, so können beide auf jeden Fall Ursache und Wirkung sowie umgekehrt sein.

Erklärung des Fehlers:

Wenn wir bereits wissen, dass Kausalität nicht automatisch durch eine Korrelation bewiesen wird, dann wissen wir auch, dass wir keine der beiden Variablen einfach so als Ursache oder Wirkung erklären dürfen.

Korrekte Betrachtung des Sachverhalts:

Die Kausalität muss genauer untersucht werden.

Einige weitere fiktive, aber durchaus mögliche Szenarios zu diesen voreiligen Schlüssen:

- 'Je mehr Nachhilfe-Stunden jemand nimmt, umso mehr Prüfungen werden bestanden' versus 'je mehr Wiederholungs-Prüfungen zu bestehen sind, umso mehr Nachhilfe-Stunden werden genommen'
Würden Sie präventiv mehr Nachhilfe nehmen oder einfach dann, wenn Sie sie aufgrund Ihrer Noten benötigen?
- 'Je mehr Medikamente genommen werden, umso mehr Heilungen finden statt' versus 'je mehr Krankheiten es gibt, umso mehr Medikamente muss man nehmen'
Würden Sie präventiv mehr Medikamente nehmen, wenn Sie diese aufgrund Ihrer Krankheit benötigen?
- 'Je länger ein Tumor bestrahlt wird, umso höher die Sterberate' versus 'die Sterberate ist bei länger zu bestrahlenden Tumoren höher (weil größerer Tumor)'
Würden Sie auf die Behandlung verzichten oder bei schwerer Krankheit zu längerer Behandlung greifen?
- ...

9.2 Regression

Wenn bei einer linearen Korrelation die unabhängige und abhängige Variable bekannt sind, so lohnt es sich bei einer signifikanten Korrelation, eine Regressionsgerade zu ermitteln, welche für beliebige x-Werte einen näherungsweisen y-Wert ausgibt (vgl. [51]):

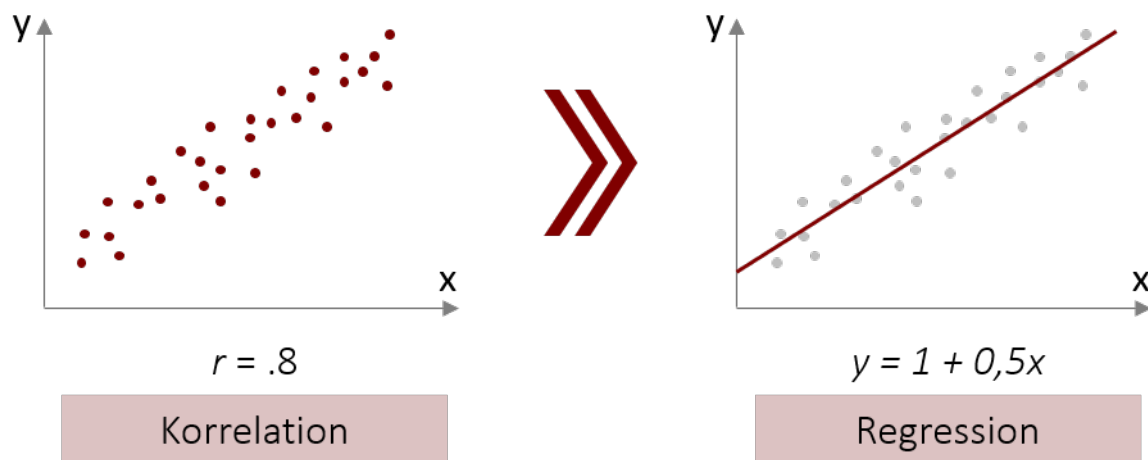


Abbildung 2: Lineare Regressionsgerade [51]

Die Ausgleichsgerade soll den Zusammenhang möglichst gut beschreiben (es wird die Gerade gesucht, bei der die quadrierten vertikalen Abstände minimal sind). Wenn Sie also eine solche sehen, denken Sie daran, dass dies nur eine Annäherung an die Realität ist. Die lineare Funktionsgleichung der Gerade hat eine Variable und oft kann auch klar ersichtlich sein, welche Größe dabei nicht die unabhängige sein kann. Die Anzahl der verkauften Kugeln Eis in Abhängigkeit der Tagestemperatur ergeben mehr Sinn als die Tagestemperatur in Abhängigkeit von der Anzahl an verkauften Kugeln Eis (als könnte man damit das Wetter beeinflussen). Weiterhin sei Vorsicht geboten, ob x auch tatsächlich y beeinflusst oder doch ein Störfaktor eine Rolle spielt.

Fehlvorstellung:

Eine Regressionsgerade beschreibt einen signifikanten Zusammenhang.

Erklärung des Fehlers:

Wenn jemand die Regressionsgerade in die Punktwolke einträgt, heißt dies nicht, dass auch eine hohe Korrelation vorliegt. Regressionsgeraden können auch für $r = 0,06$ ermittelt werden, sind dann jedoch obsolet (dies entspräche eine Punktwolke, die dann keinen Zusammenhang erkennen lässt).

Korrekte Betrachtung des Sachverhalts:

Der Korrelationskoeffizient muss $\geq 0,3$ sein, damit eine Regressionsgerade sinnvoll ist.

Die schnellste Methode zur Ermittlung (vor allem bei umfangreichen Datensätzen) ist das Verwenden eines Programmes wie Microsoft Excel, GeoGebra oder anderer Statistikprogramme.

Bemerkung: dieser Vorgang funktioniert auch bei nicht-linearen Zusammenhängen, beispielsweise bei exponentiellen, indem man die y-Achse logarithmisch skaliert und die Rechenregeln für den Logarithmus (vgl. [10], S. 4) anwendet.

$$y = c \cdot a^x$$

$$\ln(y) = \ln(c \cdot a^x)$$

$$\ln(y) = \ln(c) + \ln(a^x)$$

$$\ln(y) = \ln(c) + x \cdot \ln(a)$$

$$\ln(y) = \ln(a) \cdot x + \ln(c)$$

So wurde aus $y = kx + d$ bei linearen Zusammenhängen schließlich $\ln(y) = \ln(a) \cdot x + \ln(c)$ bei exponentiellen - wobei y nun $\ln(y)$ entspricht, k wiederum $\ln(a)$ und d letztendlich $\ln(c)$ entspricht.

9.3 Zufällige zeitliche Korrelation

Wenn ein:e Minister:in behauptet, dass sich seit seinem oder ihrem Amstantritt eine Kennzahl respektive ein Zustand erheblich verbessert hat, so kann dies auch ein Zufall sein. Es mag eine Korrelation vorliegen, aber noch nicht automatisch eine Kausalität.

So ähnlich kann es sich mit einem fiktiven Fußballtrainer verhalten, der eine neue Mannschaft trainiert, welche zuvor häufig verloren hatte und sich nun des alten Trainers entledigte. Dass die Mannschaft nun die ersten beiden Matches gewinnt, kann Zufall sein oder auch tatsächlich am neuen Trainer liegen. Sympathiepunkte wird er auf jeden Fall gesammelt haben, vielleicht auch nur dank eines Zufalls.

Weniger zufällig ist die Statistik zu Lebenserwartungen, wenn nach einem fiktiven Krieg die Lebenserwartung unter der neuen Führung steigt. Dies ist kein Zufall, liegt aber auch nicht an der neuen Führung - Grund dafür ist offensichtlich unter anderem die Tatsache, dass keine Staatsangehörigen mehr im Krieg sterben müssen und somit die Lebenserwartung wieder steigt.

9.4 Subjektive Kausalität

Von einer Korrelation den Schluss auf eine Kausalität zu ziehen, ist bereits gefährlich. Doch dies noch weiter zu führen, in dem man weitere Kausalitäten damit verbindet, ist noch fahrlässiger - hierzu ein Beispiel (vgl. [5], S. 66). Ein Unternehmen möchte den Mitarbeiter:innen nicht noch mehr Lohn zahlen, mit folgender Begründung:

Fehlvorstellung:

'Höhere Löhne sind niedrigere Gewinne, die sind gleichbedeutend mit weniger Investitionen und das resultiert in weniger Arbeitsplätzen.'

Erklärung des Fehlers:

Hier wurden viele unzulässige Schlüsse gezogen, die auf keinerlei Statistik oder Mathematik bauen. Selbst wenn eine negative Korrelation zwischen den Löhnen und den Gewinnen vorläge, so darf nicht sofort eine Kausalität beansprucht werden. Weiteres wurde keinerlei Korrelation bezüglich Investitionen oder der Anzahl an Arbeitsplätzen untersucht.

Korrekte Betrachtung des Sachverhalts:

Zunächst müsste untersucht werden, ob eine Korrelation zwischen den Löhnen und der Anzahl an Arbeitsplätzen besteht. Erst falls eine solche zutrifft, kann auch auf eine Kausalität untersucht werden.

Ein reales Beispiel aus eigener Erfahrung: im Call-Center einer Bank wurde die Vorgabe auferlegt, die Anrufe dürften im Schnitt nicht mehr 180 Sekunden, sondern 120 Sekunden dauern. Der Grund: 'Geringere Anrufdauer = geringere Kund:innen-Wartezeit = mehr Kund:innenzufriedenheit sowie weniger Ausgaben für Mitarbeiter:innen.'

Dies mag im ersten Moment schlüssig klingen, doch kann man auch argumentieren: 'Geringere Anrufdauer = weniger Kund:innen-Service aufgrund von Zeitmangels = mehr unbehobene oder mangelhaft behobene Probleme = Frustration der Kund:innen = schlechtere Reputation des Unternehmens = Verlust von Kund:innen = weniger Umsatz.'

Tatsächlich sank unter anderem aufgrund dieser Maßnahme die Mitarbeiter:innen-Zufriedenheit und so etliche Kunden und Kundinnen wurden wütend zurückgelassen, da ihren Beschwerden und Anliegen nicht ausreichend Aufmerksamkeit geschenkt werden konnte. Ob diese 60 Sekunden schlussendlich wirtschaftlich sinnvoll waren, bleibt offen und kaum messbar.

9.5 Selbsterfüllende Korrelation

Signifikante Zusammenhänge können sich auch wie ein Kreislauf selbst erfüllen. Beispielsweise wieder anhand der Kriminalität in einem fiktiven Staat.

Fehlvorstellung:

Ob Migrant:innen signifikant krimineller sind als die Staatsbürger:innen (auch in relativer Häufigkeit der jeweiligen Gruppen), lässt sich anhand der vorhandenen Daten simpel berechnen und beurteilen.

Erklärung des Fehlers:

Ein:e Ladenhüter:in kann dazu verleitet sein, die Taschen 'ausländisch Aussehender' eher zu untersuchen als jene der 'inländisch Aussehenden'. Das kann natürlich auch zu vermehrt auftretenden Fällen von Diebstählen der Ausländer:innen führen, da viele einheimische Kriminelle einfach nicht untersucht werden und somit erst gar nicht in die Statistik eingehen können.

Korrekte Betrachtung des Sachverhalts:

Auch wenn es mathematisch 'bewiesen' wurde, dass eine Signifikanz vorliegt, so stellt sich die Frage, ob in der Urliste auch die Merkmalsträger:innen aller Gruppen repräsentativ vertreten sind - sprich, entsprechend ihrer relativen Häufigkeit in der Grundgesamtheit.

Wenn 90 Prozent der Untersuchten dann Ausländer:innen oder Personen mit Migrationshintergrund sind, ändert sich auch das Ergebnis. Doch wie genau sich die Stichproben auf die Ergebnisse auswirken können, wird im Kapitel zu (eigenen) Datenerhebungen genauer besprochen.

9.6 Ergänzung zur Kriminalität von Personen mit Migrationshintergrund und ausländischer Herkunft

Merken Sie sich: bei einem Anteil von beispielsweise zehn Prozent ausländischer Herkunft an der Bevölkerung bei gleichzeitigen 15 Prozent Tatverdächtigen ausländischer Herkunft, müssen diese nicht gleich schlechtere oder kriminellere Menschen sein (vgl. [6], S. 63 ff). Viele Faktoren spielen eine Rolle, wenn es bei diesem Thema zu einer Korrelation kommt:

- **Alter:**

Kriminalität ist bei jungen Menschen stärker ausgeprägt als bei älteren. Nun ist

der Anteil junger Menschen unter den Migrant:innen signifikant höher als unter Einheimischen.

- **Geschlecht:**

Die meisten Kriminellen sind männlich, doch auch der Anteil an Männern unter Migrant:innen ist höher als unter Einheimischen.

- **Soziale Schicht:**

Von Migrant:innen begangene Straftaten sind tendenziell die sichtbaren, da es hierbei oft um Diebstahl, Körperverletzung und Raub geht. Die kriminellen Taten von Personen ohne Migrationshintergrund können öfters auch anderer Natur sein: Steuerhinterziehung, Umweltvergehen und weitere. Darauf geht die örtliche Polizei aber auch seltener ein.

- **Rassismus:**

Das zuvor beschriebene Beispiel zur Untersuchung durch Ladenhüter:innen erklärt diesen Faktor.

- **Großstädte:**

Kriminalität konzentriert sich auf Großstädte. Es bleibt hier zu fragen, ob ein kleiner Anteil an Migrant:innen Schuld an der Großstadt-Kriminalität sei oder ob sich in einer Großstadt generell mehr abspielt und nun eben auch mehr Migrant:innen in einer Großstadt sind.

Diese Argumentation der Autoren Bosbach und Korff konnte einen ortsansässigen Polizisten davon überzeugen, die Argumentationslinie seines Vortrages zu ändern beziehungsweise zu revidieren. Wie groß die Einflüsse dieser Faktoren sind, lässt sich schwer sagen, aber Sie sehen, dass die Mathematik nicht immer eindeutige Interpretationen zulässt.

Das ist nicht der Fehler der Mathematik, diese ist in sich einwandfrei, doch mit welchen Werten man die Mathematik füttert und mit welchem Motiv man sie betrachtet, kann doch erhebliche Auswirkungen mit sich bringen.

10 Statische Prognosen und Extrapolation

10.1 Mathematisches Modell vs. Realität

Jedes mathematische Modell kann die Realität nur so weit beschreiben und die zu beschreibenden Sachzusammenhänge sind meist derartig komplex, dass die mathematischen Modelle durch vereinfachte Annahmen aufgestellt werden - denn es können kaum alle Faktoren berücksichtigt werden, die derzeitig und künftig die Umstände beeinträchtigen könnten.

Ein Beispiel dazu wäre die Entwicklung der Weltbevölkerung:

Fehlvorstellung:

Die weitverbreitete Meinung lautet, dass die Zahl der Menschen auf dieser Welt immer weiter exponentiell steigen würde. Dabei beruht dies lediglich darauf, das angenommen wird, der Wachstumsfaktor bliebe stets gleich. Und so gab es allgemein auch Befürchtungen, dass bald die Nahrung ausgehen könnte, wenn die Erde 'gleichviel Nahrung hervorbrächte', während die Zahl der Menschen auf der Welt exponentiell stiege.

Erklärung des Fehlers:

Dabei wird häufig vergessen, dass die Menschheit immer wieder mit neuen Innovationen die Umstände und Bedrohungen ändert oder abwendet. Das bedeutet nicht automatisch, dass alle Innovationen zur Vergrößerung der Nahrungsmittelressourcen auch gesund sind - es heißt aber auch nicht, dass allen Menschen die Nahrung ausgehen wird.

Dennoch werden gerne statische Prognosen abgegeben, laut derer alles gleich bleibt: die Zahl der Menschen steigt immer exponentiell, Nahrungsmittelressourcen bleiben hingegen konstant (oder werden weniger).

Korrekte Berechnung des Sachverhalts:

Tatsächlich werden die Prognosen zum exponentiellen Wachstum immer stärker abgelehnt und der Verlauf durch ein logistisches Wachstum angenommen - es wird also eine Sättigungsgrenze angenommen und die Weltbevölkerung könnte auch innerhalb dieses Jahrhunderts sogar auch ein Maximum erreicht haben, wodurch auch dann ein logistisches Modell nur begrenzt die Realität und deren Zukunft beschreiben kann (vgl. [61]).

10.2 Gleichbleibende Grenzen von Kategorien

Ein einfaches Beispiel aus Sicht als Student:in: wer Studienbeihilfe bezieht, darf nur bis zu einem gewissen Betrag Einkommen beziehen, der Überschuss muss wieder zurückgezahlt werden (vgl. [11]).

Fehlvorstellung:

Man 'überwindet' die Inflation, da auch das Gehalt dementsprechend steigt.

Erklärung des Fehlers:

Wenn man nun aufgrund der Inflation die Ausgaben steigen und somit auch (im Normalfall) das Einkommen steigt, bringt es den Studierenden wenig, wenn alles, das man mehr als 15 000€ im Jahr verdient, wieder zurückzahlen muss, da die Zuverdienstgrenze sich nicht verändert hat. Dies ist ein echtes, selbst erlebtes Beispiel - das Prinzip ist auch allseits bekannt unter 'kalter Progression'.

Korrekte Betrachtung des Sachverhalts:

Mehr als die 15 000€ plus die Studienbeihilfe kann man de facto nicht verdienen - egal, wie hoch die Gehaltserhöhung ist.

10.3 Später weniger, jetzt mehr

Angaben zu Geldeinheiten in der Zukunft klingen oftmals beeindruckender, wenn nicht bedacht wird, dass aufgrund der Inflation natürlich die Ausgaben steigen.

Fehlvorstellung:

'Ich kann in zehn Jahren 20 000€ sparen. Was ich mir mit 20 000€ (jetzt) alles kaufen könnte...'

Erklärung des Fehlers:

Mit 20 000€ können Sie sich zu diesem Zeitpunkt mehr kaufen als in 20 Jahren, da die Kaufkraft dieses Betrages sinkt (vgl. [21]).

Korrekte Betrachtung des Sachverhalts:

Sei die Inflationsrate konstant mit 3% p.a. angenommen. Nun gibt es ein fiktives Gut, welches nach 20 Jahren 20 000€ kostet, den derzeitigen Preis x wissen wir nicht. Was wir hingegen wissen ist, dass dieser Preis nach 20 Jahren mit je drei Prozent Inflation 20 000€ ausmacht.

$$x \cdot 1,03^{20} = 20\,000\text{€}$$

$$x = \frac{20\,000}{1,03^{20}} = 11\,073,52\text{€}$$

Mit den 20 000€, die Sie in 20 Jahren erhalten, könnten Sie sich leider nur etwas in der Rangordnung kaufen, das derzeitig rund 11 000€ kostet.

Wir überschätzen den Wert des Geldes in 10, 20, 30 Jahren - denn es klingt nach mehr, als es sich dann tatsächlich anfühlen wird.

Bemerkung: natürlich gibt es auch Produkte, deren Preis sinken könnte, weil sich die Technologie entwickelt. Ein High-End-Computer (Stand jetzt) wird in zehn Jahren trotz Inflation womöglich weniger wertvoll sein, da die neueste Technologie unter Umständen andere und besser entwickelte Komponenten bevorzugen wird - andere Güter hingegen könnten auch 'Retro'-Gegenstände werden und deren Preis noch stärker steigen.

11 Darstellungen von Statistiken

Da die Manipulationen bei der Darstellung sehr augenscheinlich sind, werden die Erklärungen in diesem Kapitel eher kurz ausfallen. Lassen Sie uns von einem fiktiven Wert in einer fiktiven Zukunft und Geldeinheit ausgehen (beispielsweise dem Wert einer limitierten Edition eines Goldbarrens oder etwas ganz anderem), der sich relativ wenig verändert. Nun möchte aber jemand Panik und ein negatives Bild verbreiten, dass der Wert bedrohlich sinke - wie ließe sich das anstellen?

Bemerkung: manche dieser Maßnahmen können zwar in manchen Situationen durchaus sinnvoll und angebracht sein, doch mit Maß und Ziel - und einer Begründung. (Beinahe) all diese Schritte können mittels weniger Mausklicks in einer eigenen Microsoft Excel-Datei vollzogen werden.

Hier unser Wert für die Jahre 2023 bis 2029:

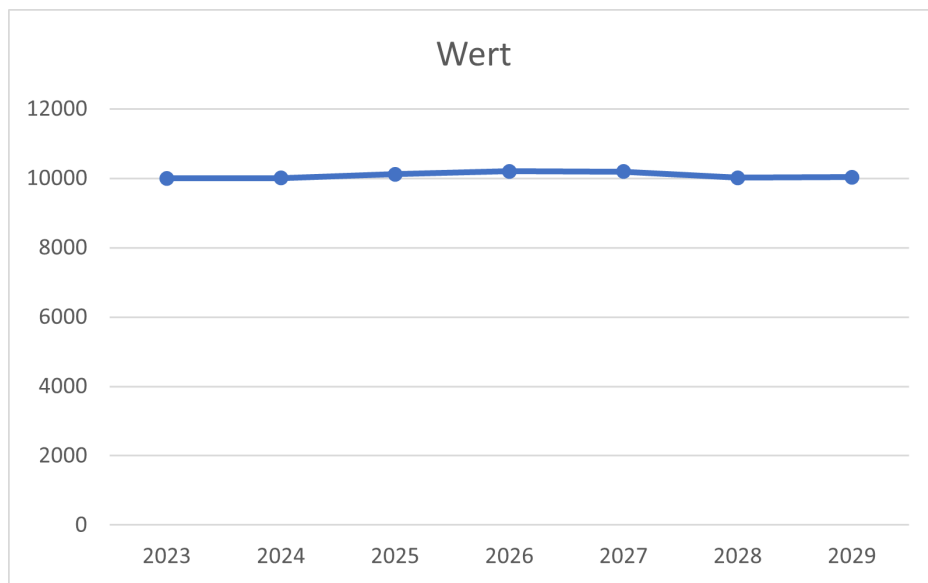


Abbildung 3: 'Neutrales' Diagramm

Der erste Schritt wäre nun, das Minimum der y-Achse zu ändern - beispielsweise auf 9950. So 'zoomen wir hinein' und erhalten bereits eine weitaus dramatischere Darstellung des Verlaufs mit einem eindeutigen Hoch und einem rasanten Abstieg:

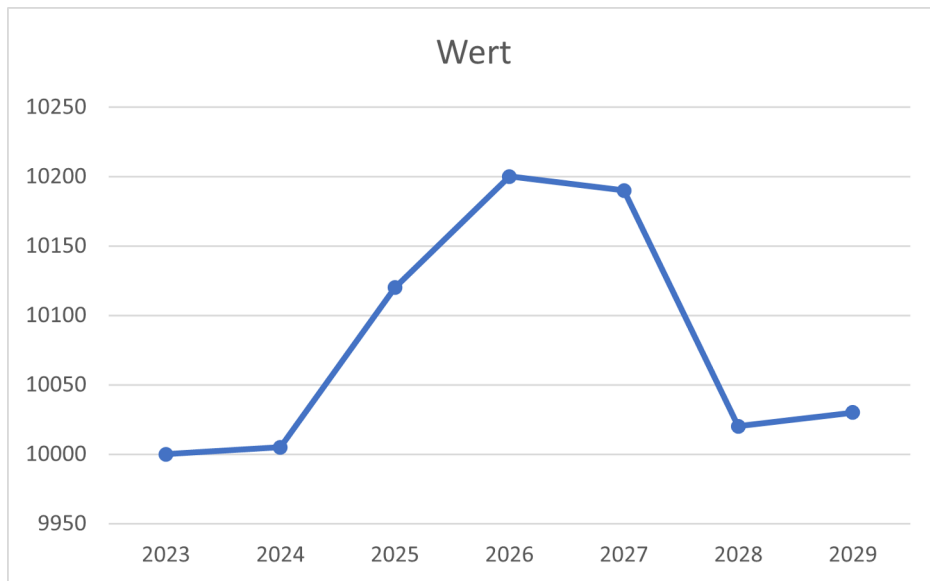


Abbildung 4: Diagramm mit 'abgeschnittener' y-Achse

Doch nicht genug: wer behauptet, wir müssen den Wertanstieg ab dem Jahr 2023 anzeigen? Schließlich wollen wir nur von der Wertminderung sprechen - also ignorieren wir die Daten vor 2026:

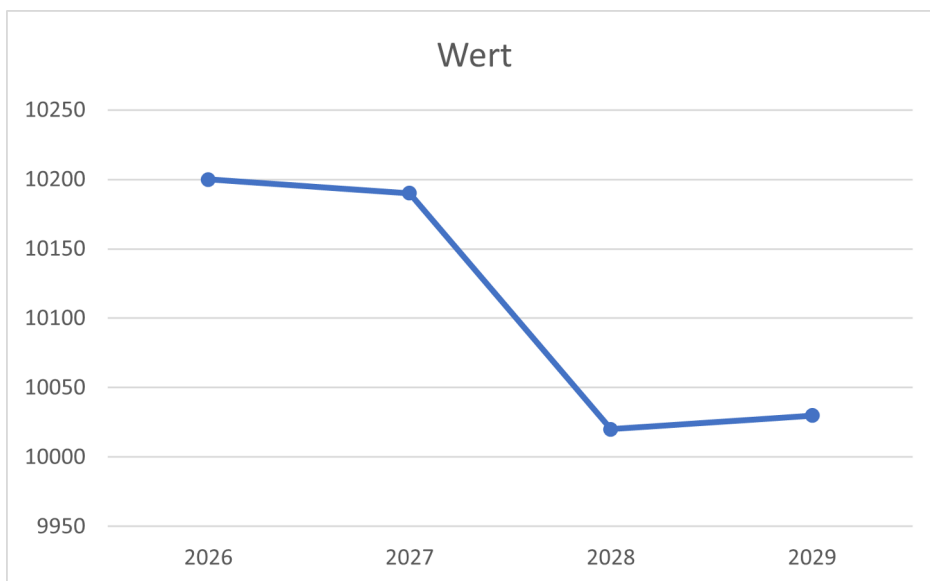


Abbildung 5: Diagramm mit eingeschränktem Zeitraum

Ein 'Problem' besteht noch, immerhin gab es von 2028 auf 2029 eine kleine Wertsteigerung, daher löschen wir das Jahr 2028 aus:

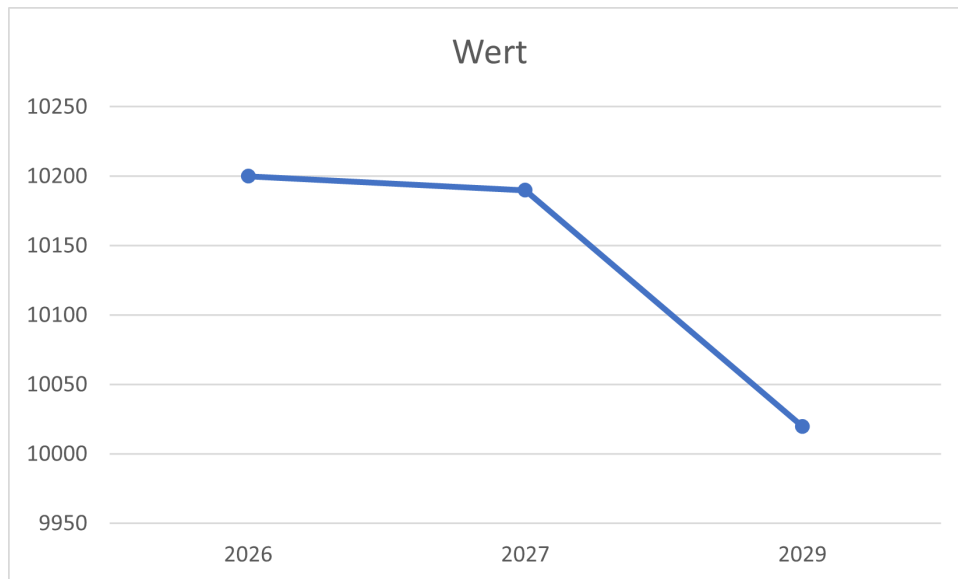


Abbildung 6: Diagramm mit fehlenden Jahreswerten

Nicht nur der Wert zum Jahre 2028 fehlt hier, sondern wurde auch das Jahr 2029 näher an das Jahr 2027 gerückt. Vielen Leser:innen dürfte so etwas auffallen, dennoch bleibt das Bild der fallenden Werte im Kopf, immerhin wissen die Betrachter:innen nicht, welcher Wert genau aus dem Jahr 2028 vorlag. Und hier noch eine weitere Veränderung:

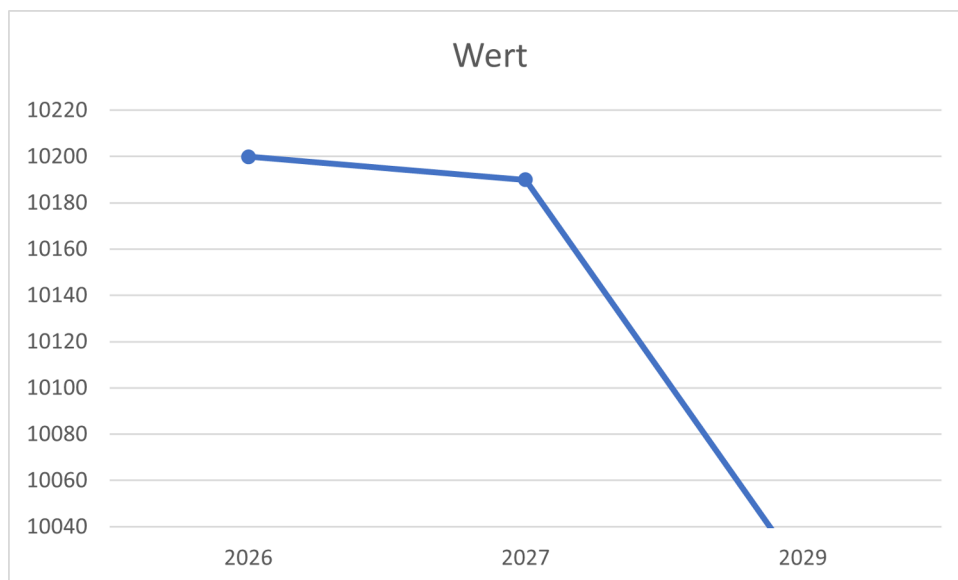


Abbildung 7: Werte entgleiten bis nach außerhalb des 'Rahmens'

Dramatischer wirkt eine Grafik, wenn die Werte außerhalb des Diagrammbereichs 'entgleiten'. Natürlich ist das Minimum der y-Achse selbst gewählt, aber dennoch wird

der Eindruck vermittelt, dass der Wert immer zwischen 10 040 und 10 220 schwankte und nun zu entgleiten droht. Blicken Sie noch einmal auf die erste Grafik dieses Beispiels und sehen Sie, wie in wenigen Schritten aus 'alles geordnet' schließlich eine dramatische Darstellung entstehen kann. Und solche Beispiele sind nicht an den Haaren herbeigezogen - vor allem das Abschneiden der y-Achse und das Beschränken des Zeitraums sind häufig angewandte Tricks, die bei vielen Arten von Diagrammen angewandt werden können.

Wir sehen uns ein weiteres fiktives Beispiel an, bei dem ein bedrohlich steigender Wert beschwichtigt werden soll (sei es bezüglich des Klimas oder etwas anderem). Hier sind die Rohdaten in einem 'neutralem Diagramm':

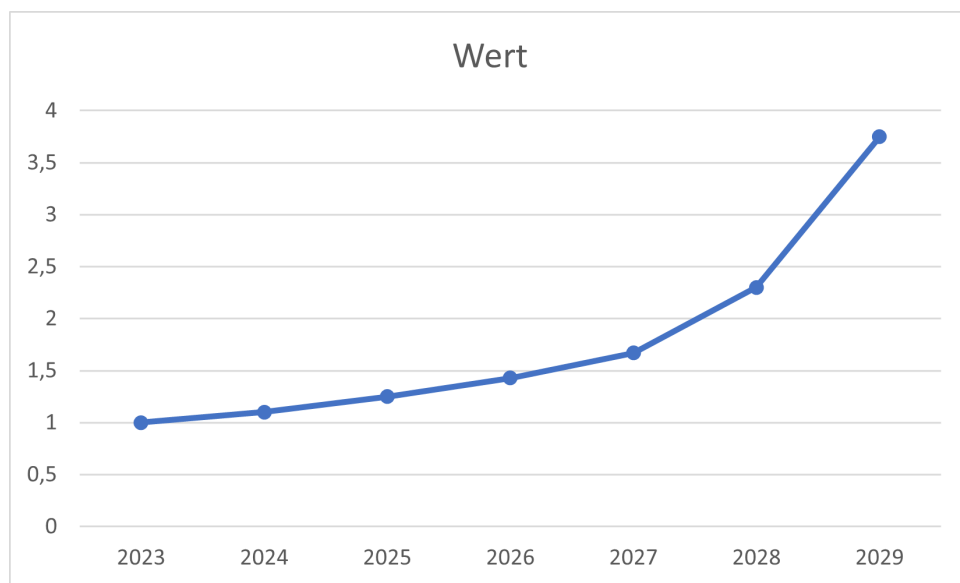


Abbildung 8: 'Neutrales Diagramm' 2

Ein klassischer und einfacher Trick besteht im Strecken respektive Stauchen der x- und y-Achse - mit wenig Aufwand ist der erste Schritt getan. Beachten Sie: manche der gezeigten grafischen Manipulationen können bei beiden Situationen (dramatisieren/entdramatisieren) angewandt werden:

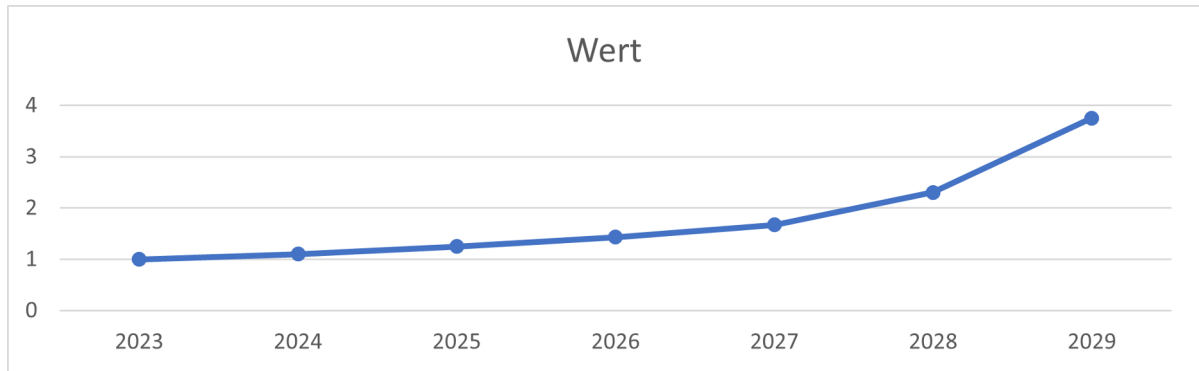


Abbildung 9: Diagramm mit gestreckter/gestauchter x-/y-Achse

Ein weniger bekannter Schritt ist das Anwenden einer logarithmischen Skalierung - dafür ist tatsächlich nur eine Option in Excel zu setzen und wir erhalten:

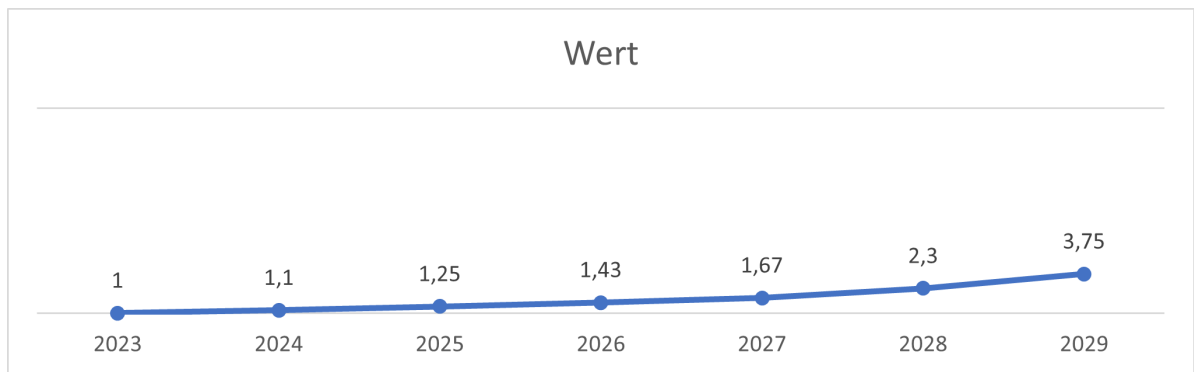


Abbildung 10: Diagramm mit logarithmischer Achsenskalierung

Praktischerweise entfernt man die visuelle Komponente der Skalierung der y-Achse und lässt lediglich die einzelnen Werte der jeweiligen Jahre anzeigen. Jede:r merkt die Steigung, aber nicht automatisch auch die logarithmische Skalierung.

Wir gehen nun zurück zum ersten Diagramm dieses Beispiels (Abbildung 8) und von dort aus betrachten wir zwei weitere Manipulationsweisen, die bei genauerer Betrachtung zwar stark auffallen, doch beim Überfliegen der Nachrichten durchaus eine Wirkung hinterlassen können. Zunächst invertieren wir die y-Achse: oben ist unten und unten ist oben. Falls Sie glauben, dass so etwas nicht vorkäme, irren Sie: genau das passierte, als eine Statistik zu 'Gun deaths in Florida' erschien, die y-Achse wurde invertiert, um ein positiveres Bild anstelle der steigenden Zahlen zu vermitteln (vgl. [19], Original-Quelle von Reuters nicht mehr abrufbar). Man bemerke die rote Farbe, die die Manipulation kräftig unterstützt:

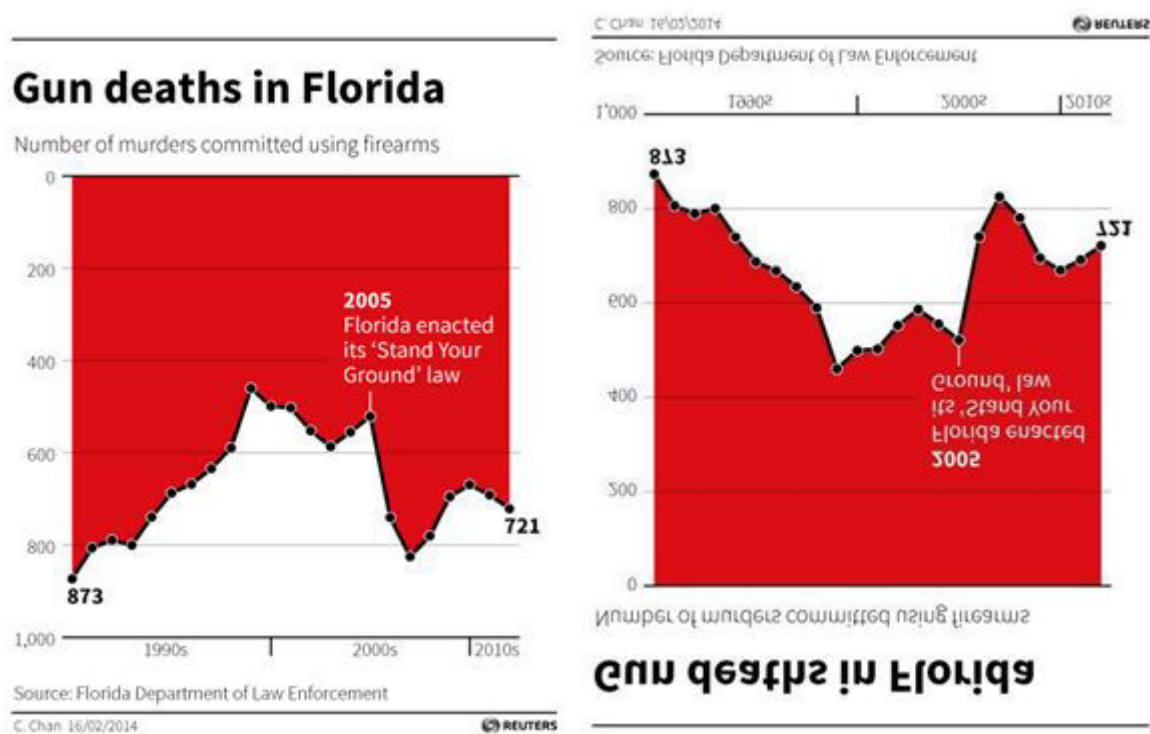


Abbildung 11: Diagramm zu 'Gun deaths in Florida' original und mit gewohnter y-Achse [73]

Bei unserem Beispiel können wir in einem Excel-Diagramm durch einen Klick auf etwas Ähnliches kommen:

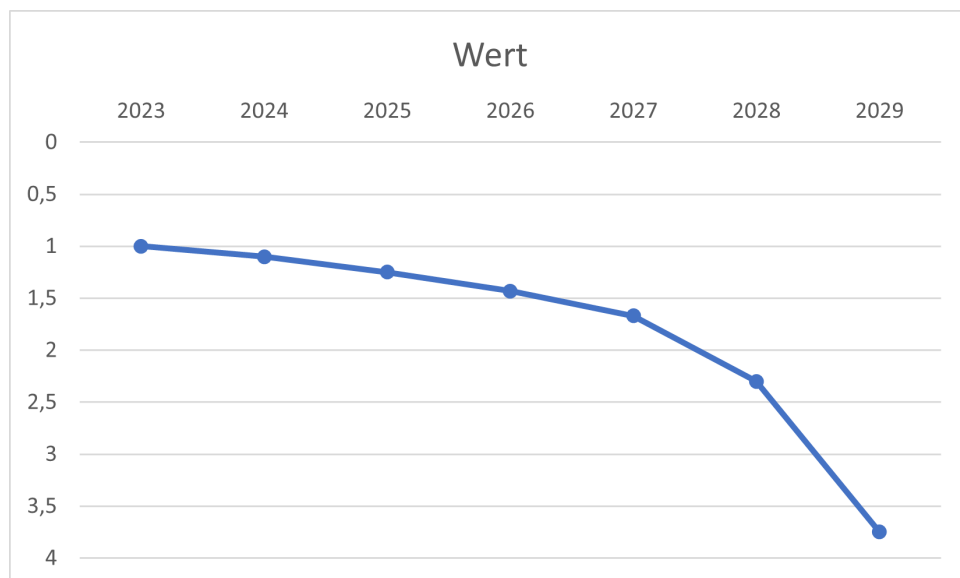


Abbildung 12: Diagramm mit invertierter y-Achse

Ein letzter Trick zu diesem Beispiel ergibt sich, wenn wir anstelle der y-Achse die x-Achse invertieren und von spät zu früh gehen anstatt umgekehrt. Dafür muss man lediglich die Reihenfolge der Wertepaare in der Tabelle umändern.

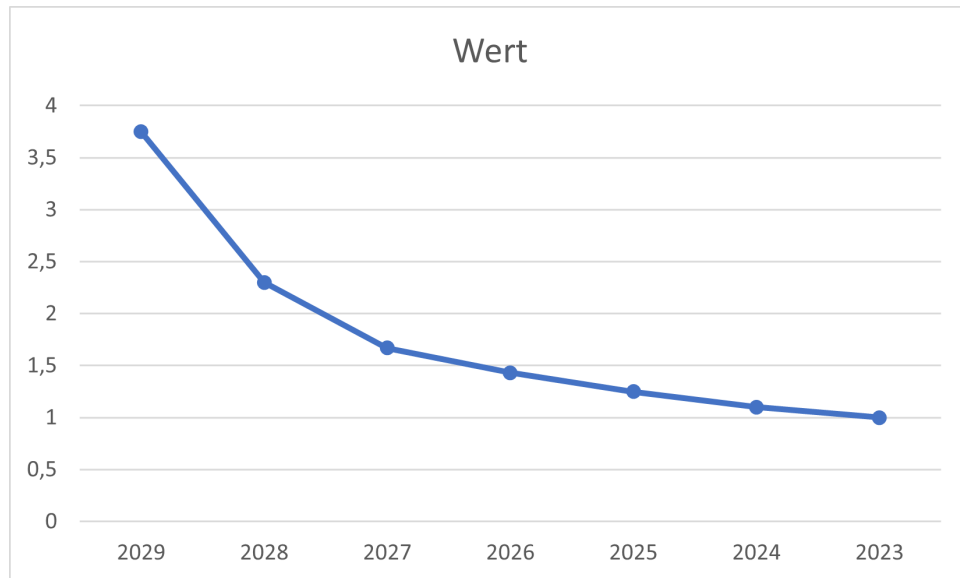


Abbildung 13: Diagramm mit invertierter x-Achse

Zwei andere Beispiele betrachten wir noch: Anteile sollen dargestellt werden, dabei wird bei einem Kreisdiagramm ein 3D-Effekt gewählt:

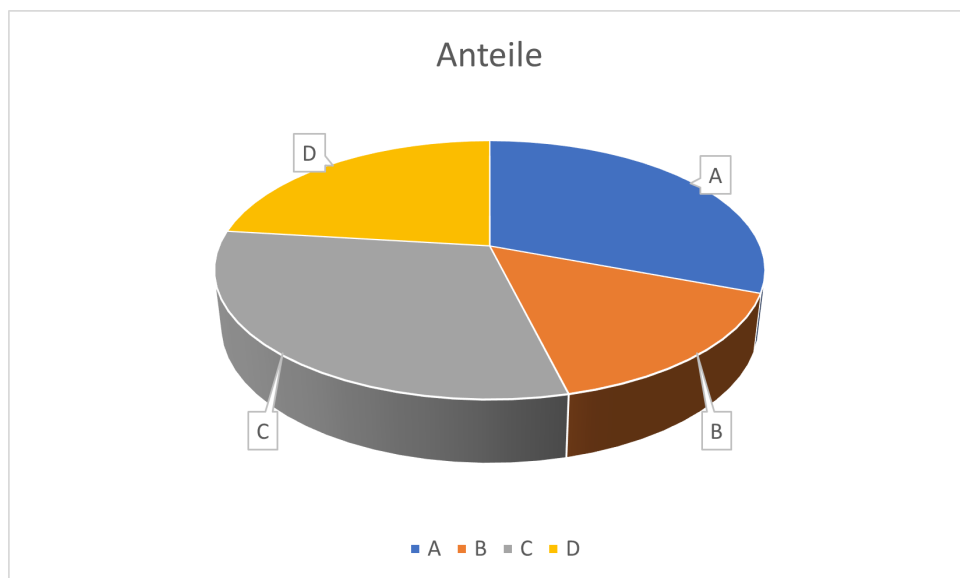


Abbildung 14: Kreisdiagramm mit 3D-Effekt

Dies ist trügerisch, da der graue Anteil größer wirkt als der blaue, obwohl er tatsächlich

kleiner ist.

Und 'last but not least', hier eine fiktive Statistik zu Beschäftigungsverhältnissen:

Beschäftigungsverhältnis	relative Häufigkeit
Arbeiter	25 %
Angestellter	40 %
Beamte	15 %
Selbstständige	20 %

Abbildung 15: Beschäftigungsverhältnis [3]

Die Größen werden durch Piktogramm wiedergespiegelt - prinzipiell eine faszinierende Idee, da dies innovativ wirkt. Doch eine Gefahr verbirgt sich dahinter:

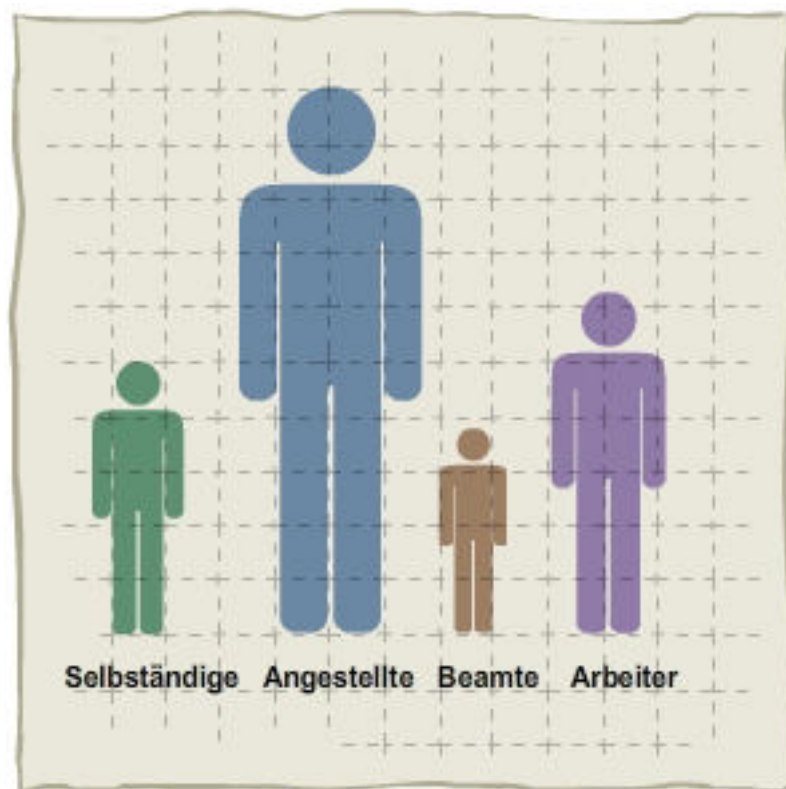


Abbildung 16: Diagramm mit Piktogrammen [3]

Es gibt doppelt so viele Angestellte wie Selbstständige und dafür wurde die Person in Blau auch doppelt so hoch dargestellt wie jene in Grün. Doch betrachtet man die Fläche,

so nimmt die blaue Person nun eindeutig mehr als doppelt so viel Platz ein. Es verhält sich ähnlich wie beim Würfel: eine doppelt so lange Seitenkante bedeutet nicht ein doppeltes Volumen. Solche Probleme lassen sich meist dann finden, wenn beispielsweise Geldeinheiten durch Geldsäcke dargestellt werden oder Müllabfälle durch Müllcontainer. Nachfolgend sehen Sie anhand der schwarzen Piktogramme den Unterschied zwischen der falschen Skalierung (links) und einer korrekten doppelt so großen Fläche (in der Mitte bzw. rechts):

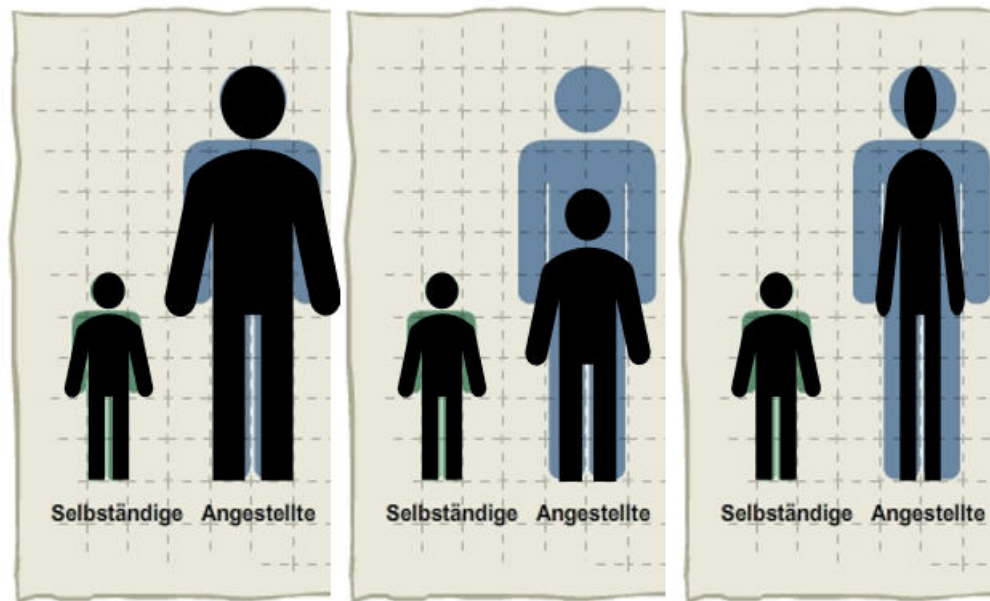


Abbildung 17: Verzerrte Darstellung (links) vs. korrekte Verdoppelung der Fläche (in der Mitte und rechts)

In der mittleren Darstellung wurden sowohl die Breite als auch die Höhe mit dem Faktor $\sqrt{2}$ multipliziert werden, um eine Verdoppelung der Fläche zu bewirken:

$$(b \cdot \sqrt{2}) \cdot (h \cdot \sqrt{2}) = 2 \cdot b \cdot h$$

Die rechte Darstellung ist auch legitim, hier wurde lediglich die Höhe verdoppelt.

Mit Grafiken zu manipulieren, ist einfach - obwohl es am wenigsten mit dem Forschungsprozess und den mathematischen Berechnungen zu tun hat.

12 Bekannte Paradoxa in der Statistik

12.1 Will-Rogers-Phänomen

Das erste Paradoxon, welches untersucht wird, ist das Will-Rogers-Phänomen (vgl. [6], S. 146 ff) Der US-Amerikaner hinterließ einen Fußabdruck in der Statistik - dabei war er kein Mathematiker, sondern Schauspieler, welcher mit einem Scherz über Oakies (Personen aus Oklahoma) mehr oder minder unbewusst auf eine Manipulationstechnik hinwies:

«Wenn die Oakies Oklahoma verlassen und nach Kalifornien ziehen, steigt der IQ in beiden Staaten.»

(Will Rogers, [57], aus dem Englischen übersetzt)

Was hat es damit (mathematisch gesehen) auf sich?

Ein fiktives Unternehmen habe zwei Zweigstellen: Nord und Süd. Die jeweiligen Umsätze der sieben Mitarbeiter:innen im Jahr 2045 sehen wie folgt aus:

Umsätze 2045	Nord	Süd
	2 000	6 000
	3 000	9 000
	4 000	10 000
		20 000
Durchschnitt:	3 000	11 250

In der Hoffnung, die Umsatzzahlen stärker zu steigern, wird die derzeitige Managerin durch eine neue ersetzt. Nach einem Jahr liefert diese die gewünschten Ergebnisse:

Umsätze 2046	Nord	Süd
	2 000	6 000
	3 000	10 000
	4 000	20 000
	9 000	
Durchschnitt:	4 500	12 000

In beiden Zweigstellen stieg der durchschnittliche Umsatz: von 3 000 auf 4 500 respektive von 11 250 auf 12 000. Daraufhin verlängert der Vorsitz den Vertrag um fünf weitere Jahre.

Kurz danach kommt die Überraschung: an den gesamten Umsatzzahlen hat sich nichts geändert! Wie kann es dann sein, dass die durchschnittlichen Umsätze in allen Zweigstellen gesteigert werden konnten?

Fehlvorstellung:

Wenn die Durchschnittswerte der Umsätze steigen, dann muss der Umsatz auch insgesamt gestiegen sein.

Erklärung des Fehlers:

Die Durchschnittswerte dürfen nicht einfach so hingenommen werden, sondern die Merkmalsträger:innen der Gruppen müssen näher betrachtet werden, denn diese haben die Gruppen getauscht.

Korrekte Berechnung des Sachverhalts:

Die neuen Umsatzzahlen werden genauer betrachtet:

Umsätze 2046	Nord	Süd
	2 000	6 000
	3 000	10 000
	4 000	20 000
	9 000	
Durchschnitt:	4 500	12 000

Die durchschnittlichen Umsatzzahlen beider Zweigstellen stiegen, jedoch liegt dies nur daran, dass ein:e Mitarbeiter:in von Süd nach Nord versetzt wurde. Alle Verkaufszahlen aller Mitarbeiter:innen blieben gleich, doch die arithmetischen Mittel stiegen - begründbar mit einer Voraussetzung:

Wann immer Sie in einer 'Gruppe' einen Wert finden, der kleiner als das arithmetische Mittel dieser 'Gruppe' ist, aber gleichzeitig größer als das arithmetische Mittel einer anderen 'Gruppe', dann können Sie beide Durchschnittswerte erhöhen, indem Sie diesen einen Wert in die andere 'Gruppe' versetzen. So hat die neue Managerin einfach den oder die Mitarbeiter:in mit einem Umsatz von 9 000 Geldeinheiten (kleiner als der Schnitt

von 11 250) in die Zweigstelle mit dem Durchschnitt von 3 000 versetzt ($9\,000 > 3\,000$), wodurch dort der Schnitt automatisch steigt. In der ursprünglichen Zweigstelle steigt der durchschnittliche Umsatz ebenso, da die 9 000 nicht mehr den Durchschnitt 'belasten'.

Besser wäre es, sich das arithmetische Mittel der Umsätze aller Mitarbeiter:innen in Obhut der Managerin zu berechnen, anstatt die beiden arithmetischen Mittel der Gruppen zu betrachten. Man könnte sogar gegen die neue Managerin argumentieren, indem man bedenkt, dass trotz Inflation die Umsatzzahlen gleichblieben und somit in gewisser Weise ein negativer Trend vorherrscht - vielleicht wäre die ursprüngliche, gekündigte Managerin doch die bessere Wahl gewesen.

Zurück zu Will Rogers: wenn er meint, dass der IQ in Kalifornien und Oklahoma stiege, wenn Oakies nach Kalifornien zögen, so meint er, dass selbst Oakies mit einem niedrigen IQ klüger als die durchschnittliche Person in Kalifornien sei.

Dieses Phänomen kann in allen möglichen Bereichen der Wissenschaft und des Alltags vorgefunden werden - ein paar der vielen Beispiele dafür:

- **Sport:**

Ein:e der schlechteren Spieler:innen der A-Mannschaft wird in die B-Mannschaft versetzt. Daraufhin steigen die Durchschnittsleistungen beider Mannschaften.

- **Politik:**

Eine geschickte Verschiebung eines Dorfes von einem Wahlkreis in einen benachbarten Wahlkreis kann bewirken, dass der Wähler:innenanteil in beiden Kreisen steigt und die führende Partei somit eventuell ein zusätzliches Mandat für sich gewinnen kann.

- **Medizin:**

Wenn in der Diagnostik Fortschritte erzielt werden, da die Technologie sich weiterentwickelt, so kann sich zum Beispiel die Kategorisierung von Tumoren ändern: bisher als 'klein' eingestufte Tumore sind nun von 'mittlerer Größe, da noch kleinere Tumore entdeckt werden können, die nun 'das neue Klein' sind. Die frühzeitige Erkennung besonders kleiner Tumore kann zu besseren Heilungschancen dieser führen. Gleichzeitig wird trotz gleichbleibender Heilungschancen der anderen Tumore bald festgestellt werden, dass auch mittelgroße Tumoren bessere Heilungschancen aufzeigen - einfach lediglich aufgrund der Tatsache, dass ehemals kleine Tumoren nun als mittelgroß zählen und mittelgroße Tumoren als groß. Es haben sich nur die Grenzen der Kategorien verändert und auf einmal haben alle

Tumoren trotz gleichbleibender Behandlung eine erhöhte Heilungschance. Und wenn diese Nachricht allen Patient:innen mehr Hoffnung gibt, so kann der Placebo-Effekt bekanntlich zu noch weiter erhöhten Heilungschancen führen (Stichwort 'selbsterfüllende Prophezeiung').

- **Ernährung:**

Bei einer fiktiven Diät gibt es verschiedene Gewichtsklassen und nur die Durchschnittswerte der Kategorien (aufgrund von Datenschutzbestimmungen) können von den Teilnehmenden eingesehen werden. Als das Experiment startet, nehmen alle Teilnehmer:innen erfolgreich ab und die Durchschnittswerte aller Gewichtsklassen sinken. Doch als die leichteste Person der obersten Gewichtsklasse nun in die daruntergelegene Klasse fällt, steigen beide Durchschnittswerte, obwohl alle abnehmen.

- ...

Solche Vorfälle müssen also nicht böswilliger Natur sein, sondern können auch auf natürliche oder unbeabsichtigte Weise auftreten.

12.2 Aufteilungsparadoxon

Ein Paradoxon, welches als Geburtsstunde der Wahrscheinlichkeitstheorie gilt, ergibt sich in einer Situation, bei der nicht 'zu statistisch' entschieden werden darf (vgl. [40], S. 6):

Zwei Spieler:innen spielen um 1 000€ ein Spiel im Modus 'Best of 11', also auf sechs Gewinnsätze. O.B.d.A. handle es sich um ein gerechtes Spiel (also mit einer jeweiligen Siegeschance von 50 Prozent). Beim Stand von 5:3 muss jedoch abgebrochen werden. Wie wird der Gewinn aufgeteilt?

Fehlvorstellung:

Statistisch gesehen hat der oder die Spieler:in A fünf von acht gespielten Sätze gewonnen (also 62,5 Prozent), somit werden die 1 000€ wie folgt aufgeteilt: A erhält 62,5 Prozent (625€) und B wiederum 37,5 Prozent (375€).

Erklärung des Fehlers:

Wichtiger als die Anzahl der gewonnenen Sätze ist die Tatsache, dass der oder die Spieler:in A viel kürzer vor dem endgültigen Sieg als der oder die Spieler:in B stand.

Anstatt die relative Häufigkeit der gewonnenen Sätze zu betrachten, muss hier die Wahrscheinlichkeit ins Spiel gebracht werden: mit welcher Wahrscheinlichkeit würde Spieler:in A oder B laut aktuellem Stand das gesamte Spiel gewinnen?

Korrekte Betrachtung des Sachverhalts:

Die einzige Möglichkeit für Spieler:in B, das Match noch zu drehen, besteht darin, die nächsten drei Sätze hintereinander zu gewinnen, das ist genau ein möglicher Weg.

Sehen wir uns die Wege an, die Spieler:in A zum Sieg führen:

- A gewinnt den 9. Satz
- B gewinnt den 9. Satz, A gewinnt den 10. Satz
- B gewinnt den 9. Satz, B gewinnt den 10. Satz, A gewinnt den 11. Satz

Achtung, eine weitere Falle besteht in folgender Annahme: in drei von vier Fällen gewinnt A, daher muss der Gewinn wie folgt aufgeteilt werden: A erhält 750€ und B lediglich 250€.

Dies stimmt nicht, da nicht alle vier Möglichkeiten gleich wahrscheinlich sind:

- A gewinnt den 9. Satz: 50 Prozent Wahrscheinlichkeit
- B gewinnt den 9. Satz, A gewinnt den 10. Satz: 25 Prozent ($0,5 \cdot 0,5 = 0,25$) Wahrscheinlichkeit
- B gewinnt den 9. Satz, B gewinnt den 10. Satz, A gewinnt den 11. Satz: 12,5 Prozent ($0,5 \cdot 0,5 \cdot 0,5 = 0,125$) Wahrscheinlichkeit

So ergibt sich in Summe eine Siegeswahrscheinlichkeit von 87,5 Prozent und A müsste somit 875€ erhalten. B stehen lediglich 125€ zu, denn die Gewinnchance (drei Sätze hintereinander zu gewinnen) beträgt $0,5 \cdot 0,5 \cdot 0,5 = 0,125 = 12,5$ Prozent.

Bemerkung: der mathematische Hintergrund dieser Berechnungen beruht auf der Pfadregel (vgl. [35], S. 17). Ausgehend davon, dass alle Satzgewinne unabhängig voneinander sind (die Wahrscheinlichkeit, einen Satz zu gewinnen, bleibt ja konstant bei 50 Prozent), kommt man schließlich auf die berechneten Werte.

Anhand dieses Beispiels lässt sich gut erkennen, dass auch Aussagen zur Gewinnverteilung aufgrund der Statistik mit Vorsicht zu genießen sind und eine korrekte Berechnung

tatsächlich schon in den Bereich der mit der Statistik verwandten Wahrscheinlichkeitsrechnung fällt, auf die wir aber nicht näher eingehen möchten beziehungsweise in der Bachelorarbeit 'Paradoxa in der Wahrscheinlichkeit' ([42]) mehr zur Wahrscheinlichkeitstheorie und anderen Paradoxa nachgelesen werden kann.

12.3 Alabama-Paradoxon

Gewisse Sitzzuteilungsverfahren für den Landtag in Deutschland beruhten auf der Grundlage von dem Juristen Thomas Hare und dem Mathematiker Horst Niemeyer beziehungsweise halten sie sich noch in Landtagen mancher Deutschen Bundesländer. Deren Verfahren soll der Benachteiligung kleinerer Parteien vorbeugen, da die Methode des größten Rests angewandt wird - die Sitzplätze werden nach ganzzahligen, zustehenden Sitzplatzanteilen vergeben und die verbleibenden Plätze nach der Größe der Reste der Anteile vergeben (vgl. [15]).

Vereinfacht: wenn eine Partei also laut Berechnung 0,45 Sitzplätze erhalten sollte und eine der anderen 2,23, so kann der Rest ($45 > 23$) unter Umständen dazu verhelfen, der kleineren Partei doch noch einen Platz im Landtag zu verschaffen, während die andere Partei bei zwei Plätzen bleibt.

Mit diesem Vorwissen blicken wir auf eine besondere Situation, bei der trotz Erhöhung der gesamten Sitzplatzanzahl die Anzahl der Sitzplätze einer Partei sinkt.

Fehlvorstellung:

Wenn mehr Sitzplätze im Landtag existieren, wird eine Partei gleich viele oder mehr Plätze erhalten.

Erklärung des Fehlers:

Durch den oder die zusätzlichen Sitz(e) müssen die Anteile und Reste neu berechnet werden.

Korrekte Betrachtung des Sachverhalts:

Wir nehmen ein fiktives Beispiel heran, dass an einem Fall in Alabama angelehnt ist (vgl. [76]) - so soll das Repräsentantenhaus eines Staates, bestehend aus drei Bundesländern, mit einer Hausgröße von zehn respektive 11 Sitzplätzen zusammengestellt werden:

Bundesland	Bevölkerungsanteil	Gerechter Sitzplatzanteil	Sitzplätze
Südliches Gebiet	0,42857	4,2857	4
Zentrales Gebiet	0,42857	4,2857	4
Nördliches Gebiet	0,14286	1,4286	2

Bei zehn Sitzplätzen erhalten die Gebiete zunächst vier, vier und einen Sitzplatz - das sind die ganzzahligen Anteile. Wenn nun die Reste der Größe nach gereiht werden, so erhält das nördliche Gebiet den verbleibenden Platz und hat somit zwei.

Anders sieht es bei 11 Sitzplätzen aus:

Bundesland	Bevölkerungsanteil	Gerechter Sitzplatzanteil	Sitzplätze
Südliches Gebiet	0,42857	4,71427	5
Zentrales Gebiet	0,42857	4,71427	5
Nördliches Gebiet	0,14286	1,57146	1

Auf einmal steht dem nördlichen Gebiet nur noch ein Sitzplatz zu, denn beim Rest überholten nun die anderen Gebiete das nördliche Gebiet - von dem zusätzlichen Platz steht nämlich den anderen beiden Gebieten anteilmäßig mehr zu und somit kann der Rest stärker steigen als bei dem kleinsten Gebiet.

Bemerkungen zu ähnlichen Sonderfällen mit Reihung der Reste:

- So kann es sich theoretisch auch im US-Repräsentantenhaus zutragen (vgl. [15]), daher auch der Name 'Alabama-Paradoxon'.
- So ähnlich kann es sich auch zutragen, wenn eine Partei wegfällt und die Anteile neu berechnet werden müssen (vgl. [24]), das nennt sich 'Parteizuwachsparadoxon'.
- In Extremfällen kann eine Partei (theoretisch) durch Stimmengewinn sogar einen Sitz verlieren, wenn sich die Reste der Anteile ungünstig verändern - dabei spricht man vom 'Populations-Paradoxon' (vgl. [25]).

All diese Beispiele hängen eng miteinander zusammen und zeigen die Obskuritäten mancher Sonderfälle auf.

12.4 Braess'sches Paradoxon

Man würde meinen, dass der Ausbau eines Verkehrsnetzes eine Begünstigung der Fahrzeiten mit sich bringt. Leider liegt ein Phänomen vor, bei dem es sich tatsächlich zutragen kann, dass durch eine scheinbar sinnvolle Ergänzung von Verkehrswegen mehr Staus entstehen - dazu das beim 1973 vom deutschen Mathematiker Dietrich Braess entdeckten Braess'schen Paradoxon (vgl. [4]).

Im vorliegenden Szenario wird ein an [70] angelehntes Verkehrsnetz verwendet, dass die Fahrtzeit (in Minuten) von A nach B für eine bestimmte Anzahl von Verkehrsteilnehmer:innen beschreibt:

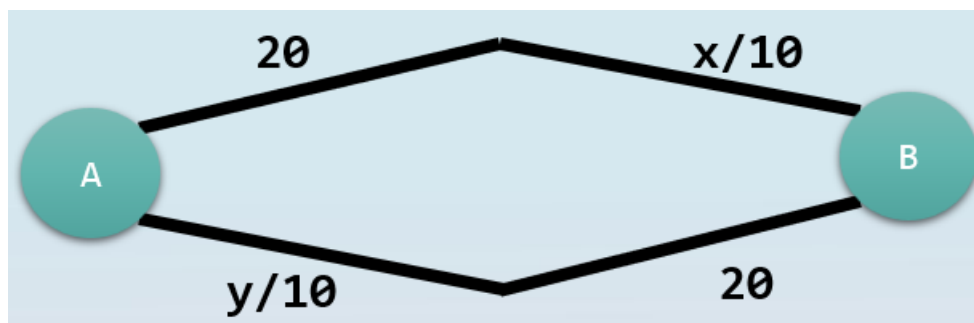


Abbildung 18: Vereinfachtes Verkehrsnetz

Dabei gibt es Abschnitte, deren Fahrtdauer konstant bleibt und solche, deren Fahrtdauer abhängig vom Verkehrsaufkommen sind. Nachdem beide Routen 'gleichwertig' sind, werden sich die Autos gleichmäßig auf beide Optionen verteilen. In diesem Beispiel gehen wir o.B.d.A. von 200 Fahrzeugen aus - davon nehmen 100 die obere Route und die anderen 100 die untere Route.

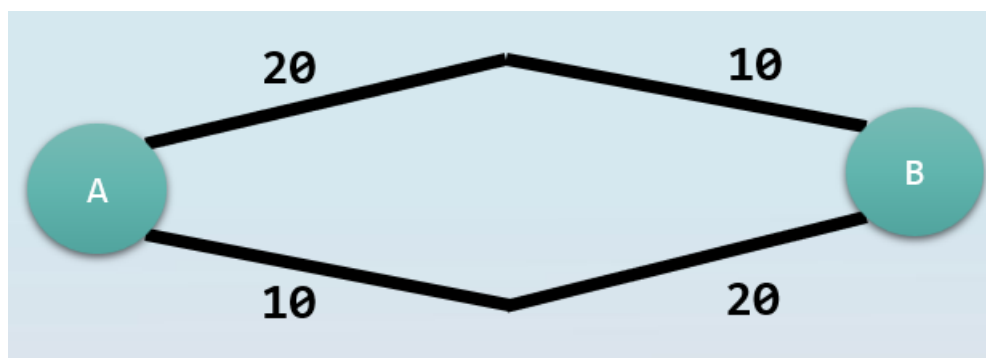


Abbildung 19: Vereinfachtes Verkehrsnetz mit 200 Fahrzeugen

Eine simple Rechnung ergibt, dass beim oberen Weg die Fahrtdauer $20 + \frac{100}{10} = 20 + 10 = 30$ Minuten beträgt und beim unteren Weg ebenfalls $\frac{100}{10} + 20 = 10 + 20 = 30$ Minuten.

Nun stellt sich die Frage, wieso nicht die beiden schnelleren Abschnitte miteinander verknüpft werden, um die Fahrtzeiten zu reduzieren. So wird eine Zubringerstrecke eingerichtet:

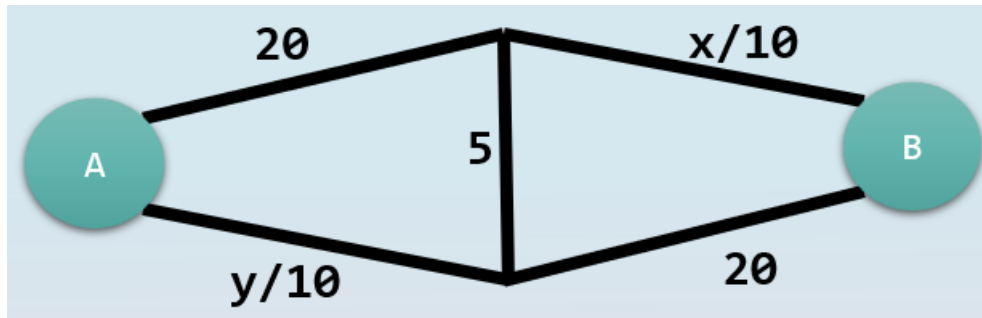


Abbildung 20: Neues Verkehrsnetz mit Zubringer

'Auf dem Papier' kann nun die schnellste Route von A nach B so aussehen, dass man zuerst den unteren Weg wählt und dann über den Zubringer-Weg nach oben kommt, von wo aus man nach B fährt.

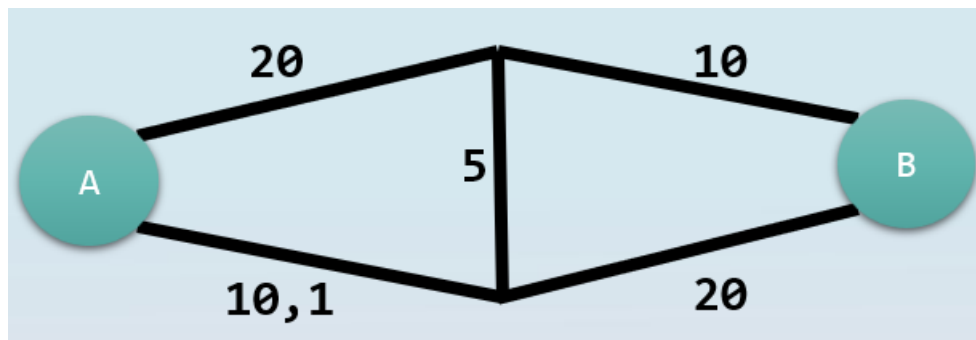


Abbildung 21: Neues Verkehrsnetz bringt scheinbaren Nutzen

Und so trägt es sich zu, dass der oder die erste Fahrer:in auf die Idee kommt, den unteren anstelle des oberen Weges zu wählen und dann den eben beschriebenen Weg zu wählen, wodurch sich eine Fahrtzeit von $\frac{101}{10} + 5 + \frac{100}{10} = 10,1 + 5 + 10 = 25,1$ Minuten beziehungsweise 25 Minuten und 6 Sekunden ergibt. Eine Ersparnis von beinahe 5 Minuten!

Danach kommen aber auch die anderen auf die Idee (und dank Navigations-Programmen wie Google Maps dürfte dies ziemlich schnell vorangehen), ihre Route auf die schnellere Option zu ändern.

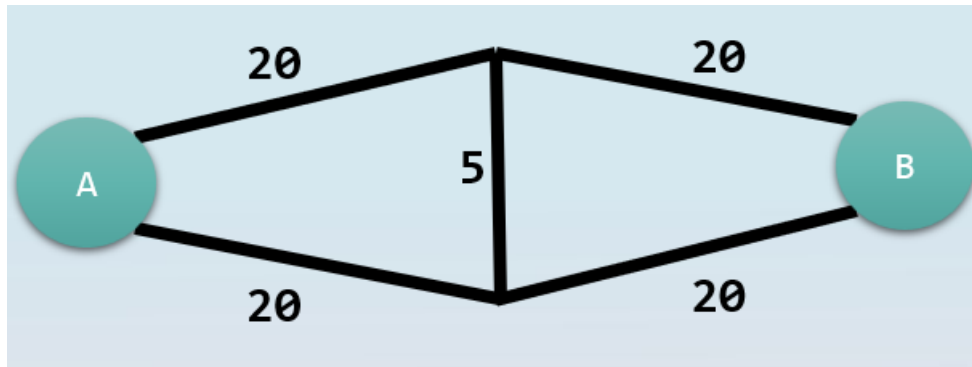


Abbildung 22: Neue Route suboptimal

Es ergibt sich nun, dass immer mehr den eben erklärten Weg anstreben, was aber zu einem negativen Effekt umschlägt, die Fahrtdauer beträgt immerhin $\frac{200}{10} + 5 + \frac{200}{10} = 45$ Minuten, wenn alle diese Option wählen - sie sind somit langsamer als sie ursprünglich waren.

Doch das ist noch nicht alles: anschließend werden einige Verkehrsteilnehmer:innen wieder zu ihrer gewohnten Route zurückkehren, da diese momentan nur $20 + 20 = 40$ Minuten beträgt. Schlussendlich würde sich ein neues Gleichgewicht einstellen, bei dem die alte Route (immer oben fahren) und die neue Route (unterer Weg, Zubringer, dann oberer Weg) den selben Zeitaufwand darstellen.

Daraus ergibt sich eine Gleichung für das neue Gleichgewicht, bei dem n Fahrzeuge die 'unten, Zubringer, oben'-Route nehmen und sich der Rest $(200 - n)$ auf die obere bzw. die untere Route gleichmäßig verteilt:

$$20 + \frac{n + \frac{200-n}{2}}{10} = \frac{n + \frac{200-n}{2}}{10} + 5 + \frac{n + \frac{200-n}{2}}{10}$$

$$20 = 5 + \frac{n + \frac{200-n}{2}}{10}$$

$$15 = \frac{n + \frac{200-n}{2}}{10}$$

$$150 = n + \frac{200-n}{2}$$

$$300 = 2 \cdot n + 200 - n$$

$$300 = n + 200$$

$$n = 100$$

Bei genau 100 Fahrzeugen entlang der 'neuen Route' und je 50 Fahrzeugen entlang der oberen und der unteren Route sind unter den neuen Bedingungen alle am 'schnellsten', nämlich beträgt die Fahrtzeit über die alten Routen $20 + \frac{150}{10} = 20 + 15 = 35$ Minuten und jene über die neue Route ebenfalls $\frac{150}{10} + 5 + \frac{150}{10} = 15 + 5 + 15 = 35$ Minuten (offensichtlich lohnt sich nicht die Route 'oben, Zubringer, unten', welche 45 Minuten dauert und daher vernachlässigt werden kann).

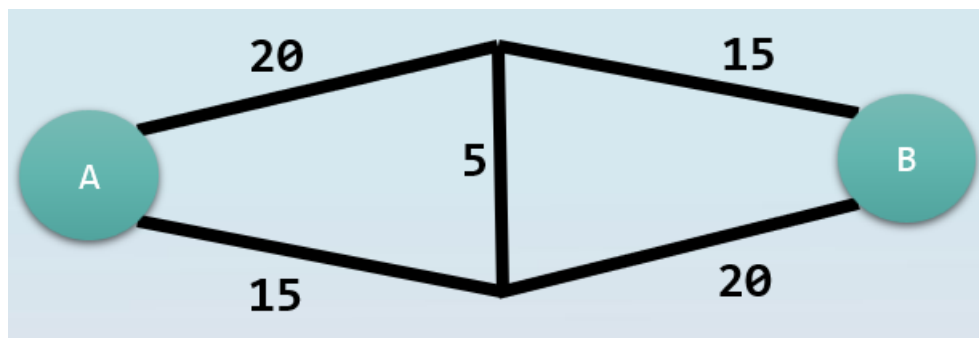


Abbildung 23: Neues Verkehrsnetz schlussendlich für alle suboptimal

Ein neues Gleichgewicht klingt optimal, alle fahren gleich lang. Doch wie lange betrug die Fahrt vorher? 30 Minuten, also fünf Minuten weniger. Wir wollten unser Verkehrssystem optimieren und gleichzeitig brauchen nun alle länger. Leider funktioniert die neue Route nur für den Gebrauch Einzelner, aber da wir alle davon profitieren wollen, verlieren am Schluss alle.

Diese 'Patt'-Situation lässt sich nicht freiwillig beheben, denn niemand würde freiwillig von der neuen Route nicht mehr Gebrauch machen und stattdessen zur Gänze den unteren Weg wählen. Dies würde immerhin bedeuten, dass man bei einer Fahrtdauer von $15 + 20 = 35$ Minuten bliebe, während durch unser Handeln die anderen Autos von unserer Abwesenheit auf den anderen Abschnitten profitieren und wieder schneller werden (wenn auch nur ein wenig) - und Google Maps würde dann natürlich auch wieder eben diesen Weg automatisch als schnellsten ausgeben, bis dort wieder mehr Autos fahren. Ein Kreislauf, dem man nicht mehr entrinnen kann.

Die einzige realistische Lösung dieser Situation ist das Sperren der Zubringerstraße, also ein Eingriff von außen. So lässt sich auch sagen, dass nicht nur der Bau neuer Straßen zu mehr Staus führen kann, sondern im 'Positiven' auch das Sperren von Straßen zu weniger Staus führen kann.

Dieses Beispiel war ein relativ simples, doch solche Fälle treten auch in echten Verkehrsnetzen und auch in anderen Bereichen des Lebens auf. Hierzu noch einige Bemerkungen:

Kritisch wird es laut Martin Treiber (Institut für Wirtschaft und Verkehr der TU Dresden) bei der Verkehrsplanung im Allgemeinen häufig, wenn neue und gut ausgebaute Abschnitte an andere Abschnitte mit niedriger Kapazität angebunden werden, beispielsweise beim Neubau von Brücken, weshalb die Kosten für die Anbindung an das Verkehrsnetz mitunter höher sein können als die für den Bau der Brücke selbst (vgl. [4]). Glücklicherweise gibt es mittlerweile Verkehrsplanungs-Softwares, die solche Risiken deutlich minimieren können.

Solche Phänomene kennt man auch bereits aus New York (42nd Street), Boston (Main Street) und London (U-Bahn-Stationen Borough und Farringdon) (vgl. [71]).

Auch ein amerikanischer Psychologe, Amnon Rapoport, ließ seine Proband:innen anhand eines virtuellen Straßennetzes am Computer Fahrstrecken auswählen. Und obwohl die Teilnehmer:innen mit echtem Geld für schnelle Verbindungen belohnt werden konnten, waren sie alle nach dem Bau einer neuen Straße langsamer als noch zuvor. Auch nachdem sie darüber informiert worden waren, was sie mit welchen Verbindungen verdient hätten, gelang es ihnen nicht, ihre Strecken effektiver zu wählen. Spannenderweise galt dies ebenso umgekehrt: nachdem die neue Verbindung wieder geschlossen wurde, konnte der Verkehr wieder flüssiger vonstattengehen (vgl. [4]).

Weitere Fälle und Aspekte des Braess'schen Paradoxons:

- Strom:
So wie bei den Fahrzeugen kann es sich auch bei Stromlieferanten zutragen: zusätzliche Leitungen im Netz können zu einer Überlastung des Systems führen (vgl. [60], S. 72).
- Sport:
Hier findet man die sogenannte 'Ewing-Theorie' (vgl. [63]), welche besagt, dass manche Sportmannschaften bessere Ergebnisse erzielen, wenn ihre Stars verletzungsbedingt fehlen. Patrick Ewing war ein Basketball-Profi, welcher als Starspieler angesehen wurde. Als er aufgrund einer Verletzung ausfiel, schaffte es sein Team zum erst zweiten Mal in 26 Jahren in die NBA Finals. Oder die Los Angeles Lakers, welche in der '71-'72-Saison nach der 'Pension' der Legende Elgin Baylor auf einmal einen Rekord mit einer 33-fachen Siegesserie erzielten und ihren ersten NBA-Titel gewinnen konnten. Aber auch in Sportarten wie Baseball, Eishockey

und anderen wird dieser Effekt bemerkt, welcher auch mit einer Überbewertung der Stars und dem Abschreiben der Teams durch die Medien zusammenhängen kann. Auch stellte der ESPN-Autor die Theorie auf, dass dieses Phänomen im Einzelsport wie Tennis auftreten kann - sein Beispiel im Mai 2001: Pete Sampras war mit Abstand an der Spitze des Herren-Tennis und sein Abtreten aus dem Tennis-Geschehen könne den Jungstars helfen, mehr ins Rampenlicht zu rücken. Das kann zur Folge haben, dass diese unter anderem mehr Erfahrungen sammeln, ihr Können weiter steigern und das Rennen um die Nummer Eins spannender wird.

- Physik:

Auch in der Physik kann das Braess-Paradoxon (dort 'Spring-Paradox' genannt) durch ein Experiment mit einem Gewicht an zwei (durch ein Stück Seil verbundene) Federn und einer Art 'Sicherheitsseil' beschrieben werden (vgl. [47]).

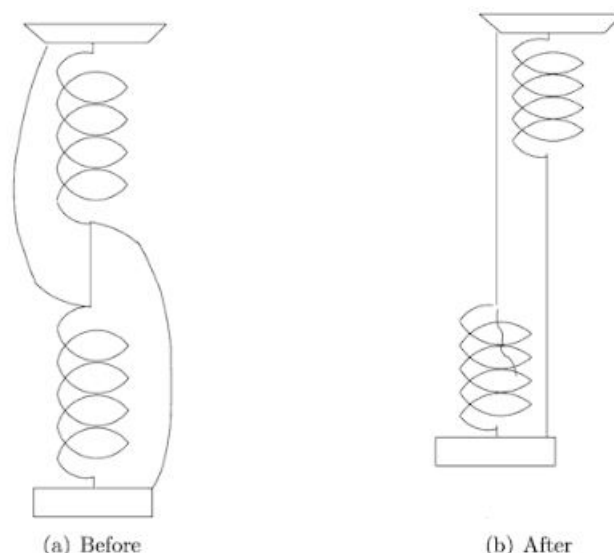


Abbildung 24: Braess-Paradoxon anhand einer Feder (Spring-Paradoxon) [47]

Man würde zunächst vermuten, dass das Gewicht nach unten sinkt, wenn das Stück Seil durchtrennt wird. Tatsächlich wandert das Gewicht nach oben, da vorher auf jeder Feder das volle Gewicht lastete und nach dem Durchtrennen der Verbindung auf jeder Feder jeweils nur noch das halbe Gewicht lastet.

12.5 Populationsdynamik

Statistiken zur Ausbeutung von Tieren sind mit Vorsicht zu betrachten. 'Fische maximal auszubeuten, gefährdet deren Bestand' - das stimmt nicht im Allgemeinen. 'Steigende Zahlen der Ausbeutung, Menschen werden also immer grausamer' - das stimmt ebenso nicht unbedingt. 'Nur noch die Hälfte der Kapazitätsgrenze dieser Spezies vorhanden' - das ist nicht schlecht, sondern kann tatsächlich das optimale Ziel sein.

Bemerkung: es geht lediglich um die Erklärung, bei welchem Bestand mathematisch gesehen langfristig möglichst viel Ertrag erzielt werden kann und das im Optimalfall unter 'möglichst würdigen' Bedingungen, beispielsweise bei eigens in einem Teich gezüchteten Fischen, die möglichst würdevoll behandelt werden.

Fehlvorstellung:

Nun sei die Behauptung: 'Der Fischbestand in meinem Teich liegt an der Hälfte der Kapazitätsgrenze, jetzt sollte ich in den nächsten Jahren so wenig wie möglich Fischen.'

Erklärung des Fehlers:

Die Populationsdynamik verhält sich (ähnlich wie bei den Menschen) bei verschiedensten Tierarten entsprechend eines logistischen Wachstums: zunächst wächst die Population immer schneller und zuletzt flacht diese Kurve wieder ab, bis die Kapazitätsgrenze beinahe erreicht wird (vgl. [33]). Dementsprechend liegt kein lineares Wachstum vor, bei dem die Population jedes Jahr um eine konstante Zahl an Fischen bis in die Unendlichkeit wächst.

Korrekte Betrachtung des Sachverhalts:

Wir nehmen eine fiktive Spezies, deren Populationsgröße N zum Zeitpunkt t durch folgende Funktionsgleichung beschrieben sei:

$$N(t) = \frac{100}{1 + e^{-0.1 \cdot (t-40)}}$$

Dabei kann die Einheit der y-Achse beispielsweise 'Prozent der Kapazitätsgrenze G ' sein (da die Kapazitätsgrenze hier ohnehin bei 100 liegt) oder 10 000 Tiere oder eine andere beliebige Werteinheit und die Einheit der x-Achse kann eine beliebige Zeiteinheit sein - je nach Spezies Jahre, Monate, Wochen oder etwas anderes. Jede Spezies verhält sich anders, aber in etwa nach einem ähnlichen logistischen Wachstums-Modell:

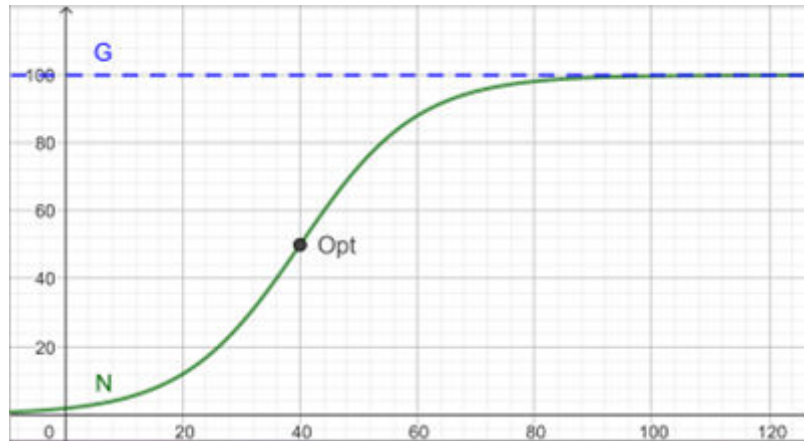


Abbildung 25: Logistisches Wachstum einer Population

Der Grafik ist zu entnehmen, dass die Population am stärksten steigt, wo sich der Wendepunkt 'Opt' befindet. Dies lässt sich unter anderem dadurch erklären, dass optimale Bedingungen für die Tiere vorliegen (ausreichend Nahrung, Raum und Population zur Fortpflanzung) und noch nicht eine Phase erreicht wurde, wo der Raum eng wird, die Nahrung knapper wird und die Population nicht mehr so gut wachsen kann - dies betrifft nicht nur den eigenen Teich, sondern generell das Habitat verschiedener Spezies.

Dass das Wachstum (in diesem Beispiel) nach 40 Jahren am größten ist, ließe sich mithilfe der Differentialrechnung einfach beweisen, wir greifen jedoch auf die uns bisher zur Verfügung stehenden Formeln zurück, indem wir die absolute Änderung betrachten:

Dass die Steigung bei 'Opt' am größten ist, konnten wir grafisch bereits identifizieren und auch die absolute Änderung innerhalb einer Zeiteinheit ist um den Zeitpunkt $t = 40$ herum am höchsten:

$$N(40,5) - N(39,5) = \frac{100}{1+e^{-0,1 \cdot (40,5-40)}} - \frac{100}{1+e^{-0,1 \cdot (39,5-40)}} = 2,5$$

Oder an saisonales Fischen (also an den Beginn einer Zeitperiode) gebunden:

$$N(40) - N(39) = \frac{100}{1+e^{-0,1 \cdot (40-40)}} - \frac{100}{1+e^{-0,1 \cdot (39-40)}} = 2,5$$

respektive

$$N(41) - N(40) = \frac{100}{1+e^{-0,1 \cdot (41-40)}} - \frac{100}{1+e^{-0,1 \cdot (40-40)}} = 2,5$$

Es lässt sich sagen, dass die Population um diesen Zeitpunkt herum innerhalb der nächsten Zeitperiode um 2,5 Werteinheiten steigt. Zum Vergleich: nach 62 Zeiteinheiten hat die Populationsgröße erstmals über 90 Prozent ihrer Kapazitätsgrenze erreicht ($N(62) = 90,02$), dort beträgt der Zuwachs jedoch nicht mehr so viel:

$$N(62) - N(61) = \frac{100}{1+e^{-0,1 \cdot (62-40)}} - \frac{100}{1+e^{-0,1 \cdot (61-40)}} = 0,93$$

Um 0,93 Werteinheiten stieg die Populationsgröße im Intervall [61;62], was eindeutig weniger ist als beim optimalen Zeitpunkt.

So verhält es sich auch, wenn wir einen zu geringen Bestand fast 'leerfischen': $N(15) - N(14) = 0,67$. Wenn wir also eine Spezies zu früh ausbeuten (oder einen zu geringen Bestand), dann ist das Wachstum derartig gering, dass wir einerseits weniger davon profitieren und obenherein auch noch die Spezies respektive den Bestand gefährden.

Wenn wir immer darauf achten, dass die Population sich in einem Bereich befindet, in dem die besten Bedingungen für ein maximales Wachstum herrschen - und nicht einfach nur darauf, dass die Population möglichst ihre maximale Kapazität erreicht hat - so nutzen wir den Bestand am besten.

Das Fazit aus diesem Paradoxon ist, dass die Akteur:innen langfristig denken müssen, dann können sie auch maximal ausbeuten, ohne die Spezies an den Rande des Aussterbens zu drängen.

Wobei sich auch wie beim Braess'schen Paradoxon hier die Frage stellt, ob solche Lösungen ohne Eingriff von außen funktionieren - immerhin würde in einer freien Marktwirtschaft das Abwarten des oder der Einen auf den richtigen Zeitpunkt bedeuten, dass ein:e andere:r womöglich früher zuschlägt und sich noch mehr nimmt. Solch komplexe Situationen sind deutlich schwieriger zu regeln, doch für unseren Teich sei dies genügend beschrieben.

Es lässt sich also sagen, dass steigende Ertragszahlen bei der Ausbeutung nicht automatisch bedeuten müssen, dass eine Spezies hinsichtlich der Population gefährdet ist und ebenso muss ein 'halb leeres' Habitat nichts Negatives sein - ohne die Augen davor zu verschließen, dass es auch vom Aussterben bedrohte Spezies gibt.

12.6 Benford's Law

Ein spannendes und für alle mit Zugriff auf Microsoft Excel selbst experimentierbares Paradoxon ist Benford's Law (vgl. [29]):

In unserem Leben begegnen wir sehr vielen Zahlen. Dabei stellt sich nun die Frage, ob die Ziffern 1 bis 9 gleich häufig als erste Ziffer einer Zahl ('führende Ziffer') auftreten.

Fehlvorstellung:

Es gibt so viele verschiedene Zahlen in verschiedensten Dokumenten und Publikationen, da muss im Großen und Ganzen jede der Ziffern 1 bis 9 mehr oder weniger gleich oft die führende Ziffer sein.

Erklärung des Fehlers:

Tatsächlich tritt die führende Ziffer 1 wesentlich häufiger auf als die anderen führenden Ziffern und es trifft sogar zu, dass jede Ziffer häufiger eine führende Ziffer als ihr Nachfolger ist (1 häufiger als 2 bis hin zu 8 häufiger als 9).

Korrekte Betrachtung des Sachverhalts:

Benford's Law besagt, dass die relative Häufigkeit der führenden Ziffer d wie folgt annähernd beschrieben werden kann:

$$P(d) = \log_{10}(1 + \frac{1}{d})$$

Dadurch ergeben sich folgende Häufigkeiten:

Führende Ziffer	rel. Häufigkeit
1	30,1 %
2	17,61 %
3	12,49 %
4	9,69 %
5	7,92 %
6	6,69 %
7	5,8 %
8	5,12 %
9	4,58 %

Ein wesentlicher Aspekt für diese Verteilung ist das exponentielle Wachstum: Inflation, Populationen (in ihren frühen Phasen), Aktienkurse und so weiter. Denn, um eine Zahl von 100 auf 200 zu ändern (oder 1 000 000 auf 2 000 000, bedarf es einer größeren prozentuellen Änderung als beim Wachstum von 800 auf 900 (oder 8 000 000 auf 9 000 000) (vgl. [45]).

Die benötigte relative Änderung zur höheren führenden Ziffer ($d + 1$) beträgt dann $\frac{d+1-d}{d} = \frac{1}{d}$.

Führende Ziffer	benötigte rel. Änderung zur höheren führenden Ziffer
1	100 %
2	50 %
3	33,33 %
4	25 %
5	20 %
6	16,67 %
7	14,29 %
8	12,5 %
9	11,11 %

Und wenn wir nun die Werte an die vorigen 30,1 Prozent für die führende Ziffer 1 anpassen, erhalten wir durch

$$N(d) = \frac{1}{d} \cdot P(d)$$

folgende Werte (Achtung - keine relativen Häufigkeiten):

Führende Ziffer	Wert nach neuem Maßstab
1	30,1
2	8,8
3	4,16
4	2,42
5	1,58
6	1,12
7	0,83
8	0,64
9	0,51

Nun folgen die Werte demselben Schema, indem sie fallen, doch für die Ziffern 2 bis 9 ergeben sich kleinere Werte. Dies ist nachvollziehbar, da nicht alle Zahlen in unserer Welt auf exponentiellem Wachstum beruhen - doch Sie sehen, dass dies wohl ein mitbestimmter Faktor ist. In unterschiedlichen Datensätzen kann dieses Paradoxon also auch unterschiedlich stark ausgeprägt sein.

Dieses Phänomen kann tatsächlich auch selbst mithilfe der von [29] zur Verfügung gestellten (oder einer eigenen) Excel-Datei auf die Probe gestellt werden - je nach Beispiel variieren die relativen Häufigkeiten, doch nähern sie sich den oben genannten an, je größer die Anzahl an Zahlen ist.

Spannenderweise dürfen aber keine bloßen Zufallszahlen gewählt werden, da dann die Verteilung nach dem Zufallsgenerator auf alle Ziffern in etwa gleich wäre (bei großer Anzahl). Und so gelang es auch, diversem Betrug auf die Schliche zu kommen (vgl. [64]): Identitätsdiebstahl konnte aufgedeckt werden sowie gefälschte Berichte an Investor:innen, bei denen die Zahlen nicht den Test von Benford's Law bestanden.

12.7 Pareto-Prinzip

Der Italiener Vilfredo Pareto untersuchte im 20. Jahrhundert den Bodenbesitz in Italien und erkannte, dass 20 Prozent der Einwohner:innen 80 Prozent des Landes besaßen und konnte schlussfolgern, dass auch die Bank 80 Prozent ihrer Zeit für die 'wertvollsten' 20 Prozent der Kunden und Kundinnen aufbringt - und dieses Phänomen kommt in unserem Leben häufig vor (vgl. [1]). Bekannt im alltäglichen Gebrauch ist die Aussage,

dass bei einem Projekt 80 Prozent der Aufgaben in 20 Prozent der Zeit erledigt werden können.

Doch auch bei anderen Beispielen wie '63 von 100 Programmen installiert' bedeutet dies nicht einen Fortschritt von 63 Prozent (unabhängig vom Pareto-Prinzip) - es können sehr wohl weniger oder mehr als 63 Prozent Fortschritt sein.

Andere Situationen, bei denen das Pareto-Prinzip unter Umständen ebenfalls näherungsweise auftreten kann, sind:

- Unternehmen erzielen 80 Prozent ihres Umsatzes mit 20 Prozent ihrer Kunden und Kundinnen.
- In Lagerhäusern können durchaus 20 Prozent der Produkte 80 Prozent des genutzten Raumes einnehmen.
- 80 Prozent des Verkehrs finden auf 20 Prozent der Straßen statt.
- 80 Prozent der Einwohner:innen eines Staates leben in 20 Prozent der Städte.
- Kinder spielen 80 Prozent ihrer Zeit mit 20 Prozent ihrer Spielzeuge.

12.8 Simpson-Paradoxon

Bei diesem Abschnitt handelt es sich vielleicht um den geheimtückischsten Denkfehler, dem dramaturgischen Höhepunkt dieser Arbeit. Eine der komplexesten, aber wichtigsten und spannendsten Situationen, an denen sich zwei Meinungslager im gemeinsamen Disput die Zähne ausbeißen können, stellt das Simpson-Paradoxon dar (vgl. [6], S. 151 ff).

In diesem Abschnitt werden einander widersprechende Tatsachen mathematisch korrekt sein, auch wenn sie gänzlich gegensätzliche Aussagen sind. Dies geht noch einen Schritt weiter als das Beispiel der Impf-Wirksamkeit mit der relativen vs. absoluten Häufigkeit.

Das Simpson-Paradoxon zeigt sich hin und wieder in Situationen, bei denen wir einen Zusammenhang zwischen zwei Größen 'feststellen', aber einen ausschlaggebenden Faktor übersehen. Wir sehen: 'je mehr x, umso mehr y', in Wahrheit ist es jedoch umgekehrt, wenn Sie den Störfaktor finden: 'je mehr x, umso weniger y'.

Ein anschauliches Beispiel zeigt eine fiktive Situation: wir untersuchen verschiedene Personen hinsichtlich zweier Aspekte: die wöchentlich aufgebrauchte Zeit zur Ausübung von Sport und der Fast Food-Konsum einer jeden Person werden beobachtet und am Schluss in ein Punktediagramm eingetragen.

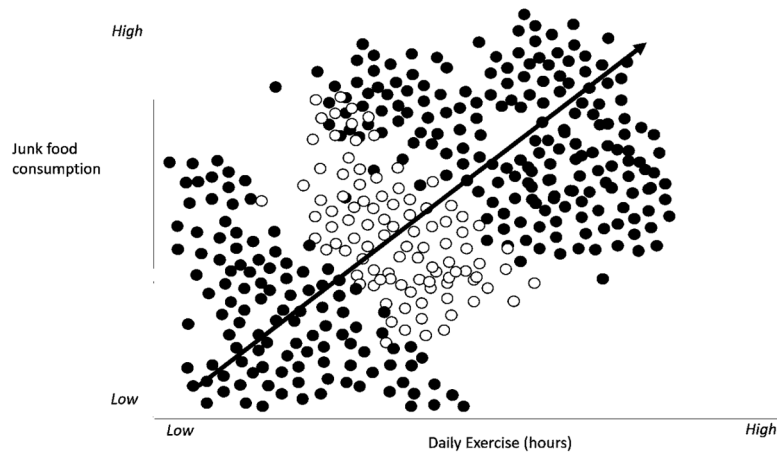


Abbildung 26: Punktwolke vor dem Simpson-Paradoxon [26]

Betrachten Sie die einzelnen Punkte - jeder steht für die Daten eines oder einer Merkmals-träger:in und die Punktwolke zeigt etwas Spannendes: je sportlicher die Menschen sind, umso mehr Junk Food nehmen sie zu sich. Daraufhin lautet womöglich die Schlagzeile:

Fehlvorstellung:

’Statistik zeigt: Sport umsonst! Wir bauen im Training Kalorien ab, um sie nachher wieder mit Junk Food hereinzuholen’.

Erklärung des Fehlers:

Wir haben einen Störfaktor nicht berücksichtigt, der unser ganzes Ergebnis auf den Kopf stellt, wenn wir ihn berücksichtigen.

Korrekte Betrachtung des Sachverhalts:

Betrachten wir die Daten nochmals, aber unterscheiden nach Bevölkerungsgruppen, in diesem Fall nach Altersklassen:

Was Sie vorher vielleicht schon an der Grafik in schwarz-weiß erahnen konnten, zeigt sich nun in Farbe deutlich: für alle einzelnen Gruppen zeigt sich, dass der umgekehrte Fall vorliegt! Je mehr Sport betrieben wird, umso weniger Junk Food wird konsumiert.

Das Gefährliche an dem Simpson-Paradoxon: meist denkt man nicht an einen möglichen Störfaktor und verbreitet vermeintliche Tatsachen. Man geht von einer positiven Korre-

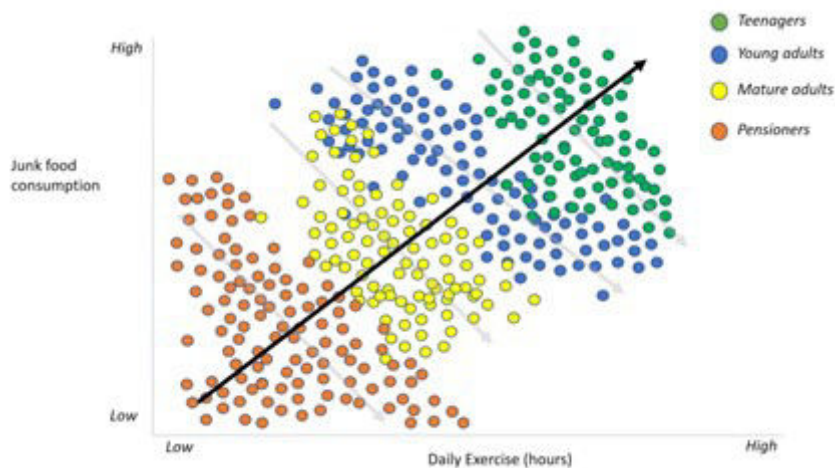


Abbildung 27: Simpson Paradoxon [26]

lation der Gesamtheit aus, dabei herrscht für alle Teilgruppen eine negative Korrelation vor!

Oft treten solch paradoxe Situationen jedoch in Tabellenform auf, wo eine Unterscheidung in der Punktwolke schwer wird. Dazu ein Beispiel:

Einer Universität wird Diskriminierung gegenüber Frauen vorgeworfen, da eindeutig mehr Männer als Frauen aufgenommen wurden (vgl. [6], S. 151 ff). Der oder die trainierte Statistiker:in könnte nun meinen, dass dies daran liegen könnte, dass sich kaum Frauen beworben haben, aber dafür viel mehr Männer. Doch ein Blick auf die Tabelle zeigt: es haben sich auch mehr Frauen als Männer beworben.

Folgende Tabelle liegt vor:

Gesamt	Aufgenommen	Nicht aufgenommen	Aufnahmequote
Männer	109	101	52%
Frauen	90	120	43%

Fehlvorstellung:

Die Frauen wurden benachteiligt - die Aufnahmequote ist bei Männern viel besser als bei Frauen (und auch die Männerquote höher als die Frauenquote). Die Universität muss sich rechtfertigen und untersucht nochmals die Daten, um der Ursache auf den Grund zu gehen.

Erklärung des Fehlers:

Auch hier liegt ein Störfaktor vor, doch welcher?

Korrekte Betrachtung des Sachverhalts:

Nun betrachten wir die einzelnen Studiengänge (in diesem Fall zwei), um zu sehen, was die Ursache ist - und stellen fest: bei beiden Studiengängen haben Frauen eine bessere Aufnahmequote als Männer:

Fach 1	Aufgenommen	Nicht aufgenommen	Aufnahmequote
Männer	100	62	62%
Frauen	38	18	68%

Fach 2	Aufgenommen	Nicht aufgenommen	Aufnahmequote
Männer	9	39	19%
Frauen	42	102	29%

Bei beiden Studienrichtungen zeigt sich, dass Frauen eine bessere Aufnahmequote aufweisen als Männer. Der Störfaktor sind hier die Studiengänge: Fach 1 hat 138 Plätze zu vergeben und Fach 2 nur 51 Plätze - von 210 Frauen haben sich aber 144 für Fach 2 beworben. Durch den Störfaktor, dass bei Fach 1 insgesamt 63 Prozent der Bewerber:innen aufgenommen wurden und bei Fach 2 nur 27 Prozent, scheint es in der ursprünglichen Tabelle so, als wären Frauen strukturell diskriminiert worden.

Ein reales Beispiel (vgl. [30]): um die aggressiven Triebe des Menschen nachzuweisen, die für unsere Fortpflanzung sorgen, zeigte ein Forscher bei der Untersuchung von Yanomami-Dörfern, dass Mörder mehr Kinder haben.

Fehlvorstellung:

Mörder in diesen Dörfern haben dreimal so viele Kinder wie Nicht-Mörder, daher kann geschlossen werden, dass sich in der Natur und Evolution die aggressiveren Gene besser durchsetzen.

Erklärung des Fehlers:

Dabei liegt auch hier ein Störfaktor vor, das Alter der beiden Gruppen.

Korrekte Betrachtung des Sachverhalts:

Das Durchschnittsalter betrug bei den Mördern über 10 Jahre mehr als jenes der

Nicht-Mörder. Und da wird uns auf einmal klar: nicht 'mehr Morde, mehr Kinder', sondern 'je älter, umso mehr Kinder' trifft zu.

Zum Abrunden des Themas und für eine Aussicht auf präventive Maßnahmen vor Veröffentlichung einer solchen Statistik sehen wir uns noch ein Beispiel aus der Medizin an, welches uns vor Augen hält, dass nicht immer alles so klar scheint - hier geht es um die Wirksamkeit von Medikamenten:

Ein Medikament wird an Frauen und Männern getestet und das Ergebnis zeigt:

Gesamt	Wirkt	Wirkt nicht	Wirksamkeit
Medikament	1200	1200	50%
Placebo	960	1440	40%

Fehlvorstellung:

Das Medikament wirkt besser als die Placebos, somit sollten wir das Medikament auf den Markt bringen.

Erklärung des Fehlers:

Es liegt auch hier wieder ein geschlechtsspezifischer Störfaktor vor.

Korrekte Betrachtung des Sachverhalts:

Es wird nach dem biologischen Geschlecht unterteilt eine Tabelle für Frauen und eine für Männer erstellt und separat betrachtet:

Frauen	Wirkt	Wirkt nicht	Wirksamkeit
Medikament	120	480	20%
Placebo	540	1260	30%

Männer	Wirkt	Wirkt nicht	Wirksamkeit
Medikament	1080	420	60%
Placebo	420	180	70%

'Wenn Sie das Geschlecht nicht wissen, verabreichen Sie das Medikament - wenn es sich aber um einen Mann oder eine Frau handelt, dann nicht.' Diese Aussage lässt sich

nicht schließen, daher müssen die Daten näher betrachtet werden - im Zweifelsfall sind die detaillierteren Daten und deren Ergebnisse heranzuziehen (die nach biologischem Geschlecht unterteilten Daten).

Was kann man nun tun, um das Simpson-Paradoxon zu vermeiden? Die Leser:innen und Forscher:innen müssen bewusst darauf achten, ob im Hintergrund Störfaktoren existieren könnten. Für Forscher:innen sei weiteres empfohlen:

- Im Test können gleich viele Frauen getestet werden wie Männer, um so diesen einen Störfaktor zu beseitigen.
- Das Alter kann auch untersucht werden - vielleicht zeigt sich ja bei Frauen und Männern eine gewisse Wirksamkeit in verschiedenen Altersklassen.
- Überlegen Sie sich, ob noch weitere Störfaktoren (Vorerkrankungen, Unverträglichkeiten, Gewicht, ...) vorliegen können und überprüfen Sie diese.
- Treffen Sie unter Umständen lieber Aussagen über homogene Gruppen (beispielsweise nur ein Geschlecht in einer Altersgruppe o.Ä.), um so Störfaktoren zu vermeiden.
- Erhöhen Sie den Stichprobenumfang, falls möglich.
- Achten Sie auf eine korrekte Durchführung des Experiments.

Dieses Beispiel ist kein Plädoyer für die Unglaubwürdigkeit der medizinischen Schule in Österreich, es zeigt lediglich die Gefahren auf - in der Praxis wird so etwas viel gefährlicher, wenn beispielsweise Journalist:innen auf einzelne Tabellen aus Publikationen zugreifen, ohne sich näher mit der Thematik beschäftigt zu haben. Denken Sie an die intensivere Covid-Zeit zurück: die Nachrichten waren voll mit Statistiken zur Impf-Wirksamkeit - und ohne der Impfung jegliche Wirksamkeit abzusprechen, zeugte es von mangelnder Professionalität, dass die Medien stets alle Bürger:innen 'in einen Topf warfen'. Eine Impf-Wirksamkeit ist für einen 25-Jährigen womöglich nicht dieselbe wie für eine Pensionistin (das spräche unter Umständen gegen eine Impfung als junger Mensch) und womöglich nahmen auch tendenziell Menschen mit Vorerkrankung eher die Impfung an als solche ohne (das spräche unter Umständen für eine bessere Impfwirksamkeit als angenommen) - weitere Störfaktoren, die in der Meinungspolitik der Emotionen untergingen. Auch die zugehörigen Daten zu Untergruppen - nach Geschlecht, Alter,

Vorerkrankungen - ließen sich im Rahmen eigener Recherchen auch bei hochrangigen Institutionen lange Zeit nicht finden.

Die letzten drei Punkte der obigen Aufzählung sind Teil des folgenden Kapitels zu Datenerhebungen, wenn es darum geht, wie man eigene Datenerhebungen einwandfrei durchführt und welche Gefahrenquellen dort liegen.

12.9 Exkurs: Gender Pay Day

Weniger Paradoxon, mehr Mythos: '22. Oktober, 71 Tage Gratisarbeit' (vgl. [13]) - die Schlagzeilen variieren, doch die meisten davon suggerieren: Frauen verdienen für dieselbe Arbeit weniger als Männer.

Vielleicht vermuten Sie bereits bereits der Gedanke, dass mehr dahinter stecken müsste, denn sonst würden die meisten Unternehmen fast ausschließlich Frauen einstellen, wenn dies so viel günstiger wäre.

'71 Tage Gratisarbeit' (wobei 71 Kalendertage nicht 71 Arbeitstage sind, wieder eine Manipulation) bezieht sich auf den Equal Pay Day, also jenen Tag, an dem Männer im Kalenderjahr das verdient haben, wofür Frauen bis an das Ende des Jahres arbeiten müssen - dieser Tag fiel 2023 in Österreich auf den 31. Oktober (vgl. [28]). Immerhin sind dies 'nur noch' 62 Kalendertage - doch zum Glück gibt es keine Diskriminierung in dem Ausmaß, in dem sie verbreitet wird. Dies bedeutet nicht, dass Frauen generell keinem Nachteil erlegen sind und es gibt auf jeden Fall allgemein diverse Umstände, die für Frauen verbessert werden müssen - doch sehen wir uns die Sachlage zum Equal Pay genauer an.

Zunächst gibt es den unbereinigten Gender Pay Gap:

Berücksichtigt wird die Differenz vom durchschnittlichen Bruttostundenverdienst aller Arbeitnehmer:innen. Diese wird in Relation zum durchschnittlichen Bruttostundenverdienst aller männlichen Arbeitnehmer gesetzt. Andere lohnbestimmende Faktoren, beispielsweise die Berufserfahrung oder der Bildungsabschluss, werden nicht gesondert in der Berechnung berücksichtigt (vgl. [44]).

In dieser Statistik geht es um Arbeitnehmer:innen, Selbstständige werden also nicht berücksichtigt (das schließt somit beispielsweise Influencer:innen aus). Das schränkt somit einmal die Aussage über Frauen und Männer im Allgemeinen ein.

Im Jahre 2018 war der Bruttostundenverdienst aller Arbeitnehmerinnen um 20,1 Prozent geringer als jener der Arbeitnehmer. Nun gibt es aber im Gegensatz dazu auch den 'Bereinigten Gender Pay Gap', dazu berichtet Frau Mischler vom Statistischen Bundesamt, dass sich Teile davon statistisch erklären lassen.

Bei Betrachtung des erklärten Teils des Gender Pay Gap [...] zeigt sich, dass 6,2 Prozentpunkte auf den höheren Anteil von Frauen in eher niedrig bezahlten Tätigkeiten zurückzuführen sind. „Tätigkeit“ bezeichnet in diesem Zusammenhang den gemeinsamen Erklärungsbeitrag der Faktoren „Beruf“ und „Wirtschaftszweig“, da beide Beiträge inhaltlich kaum zu trennen sind. Zudem lassen sich 4,3 Prozentpunkte auf Unterschiede in den Leistungsgruppen zurückführen, während 2 Prozentpunkte damit zusammenhängen, dass Frauen eher in Teilzeit arbeiten. Weitere 0,6 Prozentpunkte lassen sich darauf zurückführen, dass Frauen auch eher geringfügig beschäftigt sind.
[44]

Die drei größten Faktoren für den Unterschied sind die Faktoren Jobwahl (6,2 Prozentpunkte), Position im Unternehmen (4,3 Prozentpunkte) und durch Teilzeit-/geringfügige Anstellung bedingte geringere Bruttostundenverdienste (2,6 Prozentpunkte).

Und hier könnte man auch ansetzen. Statt zu behaupten, dass Frauen für dieselbe Tätigkeit weniger verdienen, könnten folgende Ziele erstrebenswert sein:

- Soziale Jobs (und Jobs anderer Branchen) besser bezahlen
- Mehr Frauen in Führungspositionen
- Mehr Frauen in Vollzeitbeschäftigung

Diese Themen zu fokussieren, könnte ertragsreicher sein, als weiterhin von 71 unbezahlten Tagen zu reden - das vermittelt nicht nur ein falsches Bild, sondern überschattet Fakten, die zielgerichteten Strategien helfen könnten.

Wenn Faktoren wie die eben erwähnten berücksichtigt werden, so erhält man den bereinigten Gender Pay Gap. Mit diesem stehen wir nun bei 5,9 Prozent im Vergleich zu den ursprünglichen 20,1 Prozent. Dabei steht der bereinigte Gender Pay Gap für den 'unerklärbaren Teil' - also jene Aspekte, welche aufgrund der Datenlage statistisch nicht eindeutig nachgewiesen werden können.

Dabei sind die 5,9 Prozent eine Obergrenze - nicht eindeutig messbare Faktoren umfassen beispielsweise Erwerbsunterbrechungen, individuelles Verhalten in Lohnverhandlungen oder tatsächliche Berufserfahrung.

Das Verhalten in Lohnverhandlungen wurde untersucht und so ergab sich (vgl. [20], S. 407 ff): Männer verhandeln signifikant öfters als Frauen. Es stellt sich die Frage, ob sich Frauen durch die Gesellschaft in eine Rolle gedrängt fühlen, in der sie bescheiden agieren sollen. Doch spannenderweise verhandeln Frauen sogar noch weniger, wenn die Person, mit der sie verhandeln, ebenfalls eine Frau ist.

Erwerbsunterbrechungen und tatsächliche Berufserfahrungen könnten durch die Geburt und Karenz beeinträchtigt sein - dies ist ein weiterer Faktor, an dem man arbeiten könnte: sollten Väter auch öfters in Karenz gehen?

Kurzum: zu behaupten, dass Frauen einfach aufgrund ihres Geschlechts generell 20 Prozent weniger verdienen als Männer, ist nicht zutreffend, denn mithilfe der Faktoren Jobwahl, Position im Unternehmen, Teilzeit-/geringfügige Anstellung, Erwerbsunterbrechungen, individuelles Verhalten in Lohnverhandlungen und tatsächliche Berufserfahrung kann der Unterschied auf einen unbekannten Wert (auf jeden Fall unter 5,9 Prozent) reduziert werden, der direkt dem Geschlecht zugeordnet werden kann - der genaue Wert bleibt weiterhin offen - nicht zu vergessen der Faktor 'Es geht nur um Arbeitnehmer:innen und keine Selbstständigen'.

Es besteht kein Zweifel daran, dass Frauen für dieselbe Arbeit bei dem oder der selben Arbeitgeber:in dasselbe verdienen sollen.

Wichtig ist, dass das Aufgreifen der unbereinigten Gender Pay Gap und 'gleiche Arbeit, gleiches Geld' zu vielen Irritationen führen kann. Empfehlenswert wäre es, alle möglichen Faktoren zu untersuchen und dabei zu beobachten, ob es einen Zusammenhang mit dem Geschlecht gibt. Das bedeutet beispielsweise: wieso verdient man in sozialen Jobs weniger, wieso verdient man in Teilzeit weniger pro Stunde, ...

13 (Eigene) Datenerhebungen

Grundsätzlich verwendet man statistische Verfahren, wenn man bestimmte Forschungsfragen empirisch überprüfen möchte (vgl. [22], S. 18).

Nachdem wir uns reichlich mit vorhandenen Daten (oder Extrapolation dieser) beschäftigt haben, sehen wir uns an, welche Fehler schon im Schritt davor passieren können. Denn bereits bei der Gewinnung der Daten, auf der die daraufhin getroffenen statistischen Aussagen beruhen, können bei eigenen Datenerhebungen verheerende Fehler geschehen, daher wird es in diesem Kapitel tendenziell auch um nicht direkt mathematische Aspekte gehen, welche sich aber schlussendlich auf statistische Ergebnisse auswirken - und es fängt bereits in unserem Kopf an:

13.1 Eigene Erinnerungen als Quelle

Es ist immer wieder spannend, wie sehr unser Kopf imaginäre Statistiken bereitstellt. Wenn wir ein Schloss öffnen wollen, klappt es immer erst beim letzten Schlüssel. Wenn wir etwas suchen, finden wir es immer am letzten Ort (abgesehen davon, dass das daran liegt, dass wir dann ja nicht mehr weitersuchen müssen). Unsere - in diesem Fall - negativen Erfahrungen brennen sich in unser Gehirn ein und so entstehen auch bereits verfälschte Statistiken, die wir an andere weitergeben - nur, dass es dann nicht mehr um Schlüssel geht, sondern um Sonderfälle.

Dabei beruhen viele subjektive, aber starke Empfindungen oft nur auf einzelnen Sonderfällen, die aber überproportional oft weitererzählt (oder von bestimmten Zeitschriften veröffentlicht) werden.

Dieses Phänomen der falschen Einschätzung geht auf Beurteilungsfehler zurück und davon gibt es viele verschiedene - die nicht immer zutreffen müssen, aber doch öfter als erwartet vorkommen können. Wichtig ist es dabei, vorab zu betonen, dass solche Fehler sich massiv auf Statistiken auswirken können, die dann aber neutral und mathematisch einwandfrei wirken - obwohl dem eine Manipulation vorausgegangen ist, welche die Präsentation aktueller Daten oder auch zukünftige Werte und Statistiken beeinflusst. Diese Beurteilungsfehler sind (vgl. [50]):

- Wahrnehmungsverzerrungen: Überbewertung bestimmter Aspekte

– Halo-Effekt:

Ein Attribut überschattet andere - wie ein Heiligenschein. So haben attraktivere Menschen statistisch gesehen mehr Erfolg in der Berufswelt, da ihre hohe Attraktivität zugleich mit hoher Intelligenz assoziiert wird. Auch Menschen mit Brillen werden oftmals als intelligenter eingeschätzt.

– Regency-Effekt:

Bei einem beispielsweise jährlichen Mitarbeiter:innengespräch bezieht sich die Gesprächsleitung auf die statistischen Ergebnisse, die kurz vor dem Gespräch erzielt wurden - nicht auf die Jahresbilanz. Dieser Effekt wird auch 'Nikolaus-Effekt' genannt, da für den Nikolaus und Krampus scheinbar das Benehmen der Kinder in den letzten Tagen und Wochen im November wichtiger sind als die 'Jahresbilanz'. Daher rührt auch eine der subjektiv interessantesten Redewendungen unter Musiker:innen:

'You're only as good as your last album.' ([27])

– Kleber-Effekt:

Frühere Statistiken haben Einfluss auf spätere Ereignisse - Mitarbeiter:innen, die sehr lange nicht befördert wurden, werden oft unterschätzt.

– Lorbeer-Effekt:

Hierbei handelt es sich quasi um den umgekehrten Kleber-Effekt: frühe Lorbeeren für Mitarbeiter:innen führen zur weiterhin positiven Einschätzung - ein:e Verkäuferin, welche:r die oder den bisher größte:n Kunden oder Kundin an Land ziehen konnte, wird auch zwei Jahre später als gute:r Mitarbeiter:in eingestuft, obwohl er oder sie seither in der betriebsinternen Statistik nur noch schlechte Umsatzzahlen erzielt hat.

– Andorra-Effekt:

Die subjektiv spannendste Falle liegt hier - die selbsterfüllende Prophezeiung! Eine Mutter, die aus der Karenz zurückkehrt, wird unter Umständen als weniger leistungsfähig befürchtet, da sie sich um ihr Kind kümmern muss. Dass sie und ihr Partner sich diese Aufgaben teilen und sehr motiviert sind, wird dabei außer Acht gelassen - dabei kann sich eine negative Beurteilung tatsächlich negativ auf ihre Motivation und statistisch auf die folgenden Ergebnisse auswirken.

Man kann diesen Effekt auch positiv nutzen: mehr Lob und Unterstützung können sich positiv auf die folgenden Zahlen auswirken.

Ein selbst erlebtes Beispiel: in einer betriebsinternen E-Mail wird ein Erste-Hilfe-Kurs angeboten mit dem Hinweis: 'Am besten schnell anmelden, die

Plätze sind schnell vergriffen!' Wenig wunderlich sind die Kurs-Plätze dann auch schnell ausgebucht - wie wäre es wohl, wenn da stünde: 'Keine Eile, so schnell sind die Kurse nicht ausgebucht.'

Noch ein Beispiel, dass Sie vermutlich selbst in Pandemie-Zeiten erlebt haben: Eine Zeitung vermeldet: 'Das Klopapier wird uns ausgehen.' Was passierte daraufhin? Wir kauften erst recht mehr Klopapier, da es ja bald ausgehen könnte - und dann ging es aus. So kann es mit Öl, Eiern und so weiter gehen - die Umsatzzahlen steigen noch stärker, weil Knappheit vorhergesagt wird und dadurch resultiert eine tatsächliche Knappheit - die statistische Prognose hat sich durch eine selbst erfüllende Prophezeiung bewahrheitet.

Ebenfalls kann es auch zu einem umgekehrten Effekt kommen: 'Die Schularbeit wird die leichteste des Jahres, ihr braucht nicht viel lernen!' kann dazu führen, dass statistisch gesehen die schlechtesten Ergebnisse zustande kommen, weil niemand mehr wirklich lernt.

– Hierarchie-Effekt:

Mitarbeiter:innen auf höheren Stufen werden als besser/produktiver bewertet als jene in unteren Stufen.

- Maßstabsprobleme: unterschiedliche Ansprüche an unterschiedliche Personen

– Tendenz zur Milde/Strenge:

Unterschiedliche Anspruchsniveaus führen bei denselben vorliegenden Daten zu unterschiedlichen Bewertungen. Ein:e Einser-Schüler:in wird mit einem Dreier daheim weniger Lob erhalten als ein:e Fünfer-Schüler:in mit einem Dreier.

– Tendenz zur Mitte:

Die Beurteilenden tendieren zu keinen Extremen - es gibt weder Gut noch Schlecht, nur die Mitte oder geringe Abweichungen davon.

– Antipathie-/Sympathiefehler:

Sympathische Personen werden besser bewertet als unsympathische Personen - auch wenn sie dieselben Ergebnisse erzielt haben. Bei diesem Phänomen werden sich vermutlich einige Lehrkräfte an der eigenen Nase nehmen müssen - die sympathische Person hatte nur ein Blackout, die unsympathische vielleicht geschummelt. Zwei Schüler:innen stehen zwischen zwei Noten, die Sympathie kann jedoch das Zünglein an der Waage sein, dass nur eine:r der beiden die bessere Jahresnote erhält.

- Bewusste Verfälschungen:

Denken Sie nicht, dass alle Statistiken mit bestem Gewissen verarbeitet und dargestellt werden.

Weiteres kann es selbstverständlich auch dazu kommen, dass wir Statistiken kognitiv falsch auffassen oder uns falsch merken. In manchen Fällen werden Fakten, wie bereits erwähnt, auch absichtlich falsch wahrgenommen und/oder weitergegeben.

13.2 Versuchsplanung

Einer der größten Fehler in der empirischen Forschung ist es, zuerst Daten ohne Forschungsfrage zu erheben und erst anschließend zu schauen, was 'sich finden lässt'. Vergleichsweise schießt man dabei zuerst einen Pfeil und malt danach eine Zielscheibe um den Pfeil herum.

In einem Versuchsplan (vgl. [22], S. 20 ff) muss möglichst präzise festgelegt werden:

- Inhaltliche Fragestellung, welches Ziel wird verfolgt?
- Wie wird das Ziel erreicht?
- Welche statistischen Methoden sind zur Auswertung geeignet?

Wichtig ist dabei die verbindliche Festlegung aller Aspekte. Im Laufe der Forschung darf nichts mehr geändert werden, denn sonst leidet die Qualität der Forschungsergebnisse - dies geschieht dennoch zu häufig.

Wir haben bereits über Störfaktoren gehört - die Forschungsfrage muss möglichst eingegrenzt werden, um diese Faktoren zu vermeiden. Eine Forschungsfrage könnte in Wien andere Ergebnisse als in Salzburg bringen - durch Homogenisierung (gleiche Gruppen wie Alter, Geschlecht, Ort, ...) und Randomisierung (Zufallsauswahl) lässt sich dies besser reduzieren. Dann kann eine Forschungsfrage lauten: 'Wie wirksam ist das Medikament XYZ unter Männern im Alter von 20 bis 30 Jahren mit Vorerkrankungen?'

Sie verkleinern zwar die Gruppe, über die eine Aussage getroffen wird, jedoch ist die Aussage mehr von Störfaktoren befreit, was eine zuverlässigere Aussage ermöglicht als beim Versuch, mittels der Daten etwas über alle Menschen zu sagen - denn meistens gibt es eben Störfaktoren, die man nicht eindeutig sieht. Und man will vermutlich bei

einem Medikament lieber eine Aussage zur Wirksamkeit der eigenen Gruppe (männlich, 25, keine Vorerkrankungen), als in einen Topf mit allen anderen geworfen zu werden, die anders auf das Medikament reagieren, weil sie ein anderes Geschlecht aufweisen oder Vorerkrankungen haben und so weiter. Ein anschauliches Beispiel dazu haben Sie beim Simpson-Paradoxon gesehen.

13.3 Erhebungsarten

Dieses Kapitel orientiert sich an [22] (S. 21 ff).

Es gibt drei Erhebungsarten:

- Primärstatistische Erhebung:
Die Erhebung wird selbst durchgeführt, neue Daten werden gesammelt.
- Sekundärstatistische Erhebung:
Es liegen Originaldaten vor (beispielsweise aus statistischen Jahrbüchern), welche herangezogen werden.
- Tertiärstatistische Erhebung:
Bereits komprimierte oder transformierte Daten werden entnommen (beispielsweise Mittelwerte statt Einzelwerte oder Altersklassen statt genauen Alters).

Dabei ist zu beachten, dass bei sekundärstatistischen und noch mehr bei tertiärstatistischen Erhebungen mit wachsenden Fehlerquellen zu rechnen ist.

Fehlende Daten können bei allen drei Arten ein Problem sein - eine Person verweigert eine Aussage bei einer Unterfrage, eine Krankenhausakte ist unvollständig ausgefüllt und so weiter.

Bei den Datenerhebungen versucht man meist, anhand einer untersuchten Teilgruppe eine Aussage über die gesamte Gruppe (Grundgesamtheit) zu treffen. Dabei hat man meist nicht die Möglichkeit, alle Merkmalsträger:innen zu untersuchen, weshalb Stichproben durchgeführt werden.

Stichprobe:

Wenn nicht die gesamte zu betrachtende Grundgesamtheit erhoben werden kann, greift man auf Stichproben zurück - also auf einen Teil der Grundgesamtheit, der zufällig gezogen wird, um repräsentativ für die Grundgesamtheit zu sein.

Einfache Zufallsstichprobe:

Es werden Teilmengen oder einzelne Merkmalsträger:innen so erhoben, dass alle mit derselben Wahrscheinlichkeit gezogen werden können. Stellen Sie sich vor, dass Sie alle Merkmalsträger:innen mit Nummern versehen und anschließend zufällige Nummern ziehen, wie bei einer Lottoziehung.

Geschichtete Zufallsstichprobe:

Es werden Schichten gebildet, welche sich nicht überlappen und anschließend werden Stichproben erhoben.

Beispielsweise könnten in Österreich solche Schichten die Bundesländer sein und in jedem Bundesland wird eine Zufallsstichprobe durchgeführt.

Klumpensstichprobe:

Es bilden sich auf natürliche Weise Klumpen (beispielsweise Gemeinden). Aus der Grundgesamtheit werden nun einige Klumpen komplett erhoben.

Bewusste Auswahl:

Oft gibt es als Forscher:in keinen Zugang zu einem Mechanismus, der randomisiert Bürger:innen ziehen kann. Sie können sich vermutlich denken, dass viele Forschungen bereits an dieser Hürde scheitern. Wenn ein:e Student:in eine Umfrage startet, so wird gerne einmal der Link zur Umfrage den Verwandten und Freund:innen geschickt, vielleicht auch der Kollegschaft oder einer WhatsApp-Gruppe. Das beeinflusst die Ergebnisse bereits massiv, wenn eine Aussage über alle Österreicher:innen getroffen werden soll. Denn die Wahrscheinlichkeit für eine wildfremde Person - über die ja auch eine Aussage getroffen werden soll - zur Erhebung herangezogen zu werden, ist praktisch Null. Natürlich verfügen Studierende nicht über die nötigen finanziellen, strukturellen und zeitlichen Mittel, weshalb die Forschung zu weniger fundierten Aussagen kommt.

Umfragen im 1. Wiener Bezirk können ganz anders ausgehen als im Waldviertel - die Aussagekraft über die gesamte Bevölkerung ist also weniger beeindruckend als das Befragen von Personen in möglichst vielen Ortschaften. Oder eine Umfrage vor den Türen einer Universität wird ganz andere Merkmalsträger:innen hervorbringen als vor einem Fußballstadion. Das Finanzministerium wird vermutlich bessere Möglichkeiten für eine Zufallswahl haben als ein:e Schüler:in in der fünften Klasse.

Allenfalls kann man die Forschungsfrage auf eine kleinere Grundgesamtheit reduzieren (vor der Erhebung!) - beispielsweise statt über die gesamte Bevölkerung lieber über Studierende in Wien Aussagen zu treffen, wieder einmal auch hier: Störfaktoren be-

seitigen. Dies geschieht aber relativ selten - die meisten Einladungen zu Umfragen in Studierenden-Gruppen zielen auf Aussagen über die breite Masse ab, befragen gleichzeitig aber hauptsächlich Studierende desselben Ortes, sogar derselben Fakultät, sogar desselben Studienganges.

Man kann versuchen, durch Quoten zu einer besseren Aussagekraft zu kommen - wenn beispielsweise 40 Prozent einer Grundgesamtheit weiblich sind, so sollte bei der Erhebung der Anteil an Frauen möglichst ähnlich groß sein. Alter, Geschlecht, Wohnort, Beschäftigungsgrad und alle dieser Aspekte gemäß der Quoten zu erheben, ist natürlich äußerst schwer.

Kurzum: repräsentative Stichproben sind schwer zu erheben - das ist auch verständlich. Denken Sie einfach daran, dass eine Forschung unter Umständen nicht so repräsentativ ist, wie man gerne hätte. Zum Glück sind wichtige Forschungen durch renommierte Institutionen seltener von diesem Problem befallen, als wir selbst bei eigenen Forschungen.

Studiendesigns:

Querschnittsstudien untersuchen bestimmte Merkmale zu einem Zeitpunkt (Wahlumfragen, ...), Zeitreihen betrachten eine:n Merkmalsträger:in im zeitlichen Verlauf (Aktienkurse, ...) und Längsschnittstudien betrachten eine Gruppe von Merkmalsträger:innen zu verschiedenen Zeitpunkten (Vorher-/Nachher-Betrachtung, ...).

13.4 Qualität und gleiche Bedingungen

Wenn beispielsweise Messwerte miteinander verglichen werden sollen, müssen die Messungen qualitativ sein und unter möglichst gleichen Bedingungen geschehen.

So kam es dazu, dass die Auswertung von 30 Jahre lang in Klimastationen Österreichs gesammelten Daten ergab, dass im Hinblick auf die Sonnenstunden Großraming das absolute Schlusslicht war (vgl. [58]).

Fehlvorstellung:

Mit 217 Sonnenscheinstunden von November bis März steht die Ortschaft satte 45 Stunden hinter dem vorletzten Platz Wels. Die interviewte Person konnte sich nicht genau erklären, warum Wels am vorletzten Platz liegt, vermutete daraufhin Nebelanfälligkeit.

Erklärung des Fehlers:

Die Messstation befand sich im Schatten eines Berges (vgl. [55], originaler Zeitungsartikel nicht mehr abrufbar).

Korrekte Betrachtung des Sachverhalts:

Die Datenqualität lässt zu wünschen übrig, da die Bedingungen unterschiedlich sind - und somit ist ein Vergleich nicht zulässig.

13.5 Mehrfach gezählte Merkmalsträger:innen

'Es gab in dieser Woche 70 000 Besucher:innen', regelmäßig sind ähnliche Meldungen zu lesen, doch stellt sich dabei die Frage, ob es tatsächlich 70 000 verschiedene Personen sind oder einfach 70 000 Tickets verkauft wurden.

Fehlvorstellung:

Es gab bei dem Tennis-Turnier innerhalb einer Woche 70 000 Zuschauer:innen.

Erklärung des Fehlers:

Ausgehend davon, dass eine Person an mehreren Tagen anwesend sein kann, muss die Zuschauer:innenzahl nicht der Anzahl an verkauften Tickets entsprechen.

Korrekte Betrachtung des Sachverhalts:

Es seien die genauen Daten der Zuschauer:innen vorhanden (H_n sei die Anzahl der Personen mit Tickets für n verschiedene Tage):

n	H_n
1	12 000
2	6 000
3	4 000
4	3 000
5	2 500
6	1 000
7	500

In Summe wurden

$$1 \cdot 12\,000 + 2 \cdot 6\,000 + 3 \cdot 4\,000 + 4 \cdot 3\,000 + 5 \cdot 2\,500 + 6 \cdot 1\,000 + 7 \cdot 500 = 70\,000$$

Tickets verkauft, jedoch gab es nur

$$12\,000 + 6\,000 + 4\,000 + 3\,000 + 2\,500 + 1\,000 + 500 = 29\,000$$

verschiedene Zuseher:innen bei dem Turnier. Wenn also das nächste Mal bei einem Tennis-Turnier in einer Woche '70 000 Personen vor Ort zusahen', so könnten es theoretisch auch nur 29 000 gewesen sein, wenn viele der Zuschauer:innen an mehreren Tagen anwesend waren. '70 000 verkaufte Tickets' wäre die korrekte Beschreibung.

13.6 Verwechseln der Merkmalsträger:innen

Ein Beispiel, an dem diese Problematik gut gezeigt werden kann, wäre eine Umfrage, anhand der ermittelt respektive eingeschätzt werden soll, wie viele Ein-, Zwei-, Drei- oder Mehrpersonenhaushalte es gibt. Nun geht das Forschungsteam an einem umsatzstarken Samstag in ein Einkaufszentrum. Abgesehen davon, dass es sich hierbei schon um eine vorsortierte Stichprobe handelt, da nicht jede Bevölkerungsgruppe mit gleicher relativer Häufigkeit in das Einkaufszentrum geht, wird nun der Fehler in der Umfrage (wobei H_i die Anzahl der Befragten aus i -Personenhaushalten ist) betrachtet:

i	H_i
1	20
2	40
3	60
4	80

Fehlvorstellung:

Von 200 Befragten sind 60 aus Dreipersonenhaushalten, daher sind $\frac{200}{60} = 0,3 = 30$ Prozent der Haushalte Drei-Personen-Haushalte.

Erklärung des Fehlers:

Aus jedem Dreipersonenhaushalt können drei Personen angeben, in einem solchen zu leben - aus einem Einpersonenhaushalt kann dies nur eine Person angeben.

Korrekte Betrachtung des Sachverhalts:

Die Haushalte müssten kontaktiert werden, nicht die Bewohner:innen derer befragt, denn die zu untersuchenden Merkmalsträger:innen sind die Haushalte und nicht deren Bewohner:innen, die Umfrage ist daher irreführend.

Von der anderen Seite betrachtet: angenommen, es gäbe in einem Dorf 80 Haushalte, wobei H_i Haushalte i -Personenhaushalte sind:

i	H_i
1	20
2	20
3	20
4	20

Nun befragt man alle Bewohner:innen, wie viele Personen in ihrem Haushalt leben, so ergibt sich wieder das anfangs erwähnte, verfälschte Bild: 20 Personen antworten mit 'Einpersonenhaushalt', $2 \cdot 20 = 40$ Personen antworten mit 'Zweipersonenhaushalt', $3 \cdot 20 = 60$ Personen antworten mit 'Dreipersonenhaushalt' und $4 \cdot 20 = 80$ Personen antworten mit 'Vierpersonenhaushalt'. Daraus den Schluss zu ziehen, dass $\frac{60}{200} = 0,3 = 30$ Prozent der Haushalte Dreipersonenhaushalte sind, ist inkorrekt, denn tatsächlich sind es lediglich $\frac{20}{20+20+20+20} = 0,25 = 25$ Prozent.

13.7 Manipulieren der Merkmalsträger:innen

Dies ist wohl einer der wichtigsten Abschnitte dieser Arbeit in Hinsicht auf Beeinflussung der breiten Masse und wie sich falsche Ergebnisse über Jahrzehnte hinweg halten können (ähnlich wie die falsche Information zum Spinat).

Dafür betrachten wir eines der berühmtesten Experimente in der Psychologie, deren Fehler in einem Werk von Rutger Bregman mit dem Thema 'Im Grunde gut' ([7]) aufgezeigt werden, um so das Bild vieler aufzulösen, der Mensch sei schlicht böse. Dabei geht es hier weniger um rechnerische Abläufe, sondern um das Prinzip, dass wir alle möglichen statistischen Aussagen erlangen können, wenn die 'richtige' Manipulation vor oder während der Datenerhebung angewandt wird.

Stanford-Experiment:

Am 15. August 1971 werden neun 'Häftlinge' jeweils zu dritt in eine Zelle gesteckt

und von 'Wärtern' gefangen gehalten, die für Ordnung sorgen sollten. Dann beobachtete man, was geschah: nach wenigen Tagen musste das Experiment aufgrund der brutalen Ausuferungen seitens der Wärter abgebrochen werden - seither ging dieses Experiment und dessen Ausgang um die Welt. Dabei wurde im ganzen Medienrummel übersehen, welche Manipulationen geschahen, die solche Gewaltausmaße ermöglichten - der Versuchsleiter Philip Zimbardo wurde weltberühmt, doch sein Ruhm beruht auf Manipulationen (vgl.[7], S. 167 ff).

Dazu sehen wir uns einen Ausschnitt aus den vielen Transkripten der Audio-Aufzeichnungen an - denken Sie dabei daran, dass die Prämisse dieses Experiments war, dass sich die Wärter ohne Anweisungen der Leitenden zurecht finden sollen, kein Eingriff von außen - offiziell.

Hier ist der Auszug eines Gespräches, welches zwischen einem der Leiter, Philip D. Jaffe (J) und einem Wärter (RG) stattfand:

J: We noticed this morning that you weren't really lending a hand and I was wondering if there was anything wrong?

RG: Well this morning it seemed like there was pretty many guys in there anyhow [...]

J: But generally, you've been kind of in the background, part of this is my fault, 'cos I've gone along you know when you've wanted to sit outside while they were doing Count and that sort of thing. We really want you to get active and involved because the guards have to know that every guard is going to be what we call a tough guard, and so far —

RG: I'm not too tough.

J: Well, you have to kind of try and get it in you.

RG: Well, I don't know about that.

J: See the thing is, what I mean is tough is, you have to be thorough and you have to be in the action and that sort of thing. It's really important for the workings of the experiment [...] the thing depends largely on the guards' behaviour.

RG: Oh, I can see that. I can see what it's doing to the experiment. [...]
[...]

RG: [...] What I would do [...] if this was just entirely up to me, is I wouldn't do anything, I'd just let it cool off.

J: At some point that might be necessary, but our purpose here is not to devise a better prison. [...]

[...]

J: Well, we'd like you to try a little bit more, to get into the act.

[77], zweites Interview, S. 1 ff

Im Protokoll wird der Wächter 'RG' genannt: 'reluctant guard' (widerwilliger Wächter), dies verdeutlicht die Einstellung der Protokollierenden gegenüber des Probanden.

Die zweite Lektion dieses Beispiels ist eine wiederholte: statistische Aussagen müssen vorsichtig behandelt und überprüft werden, bevor man sie anderen weitererzählt. Der dadurch angerichtete Schaden könnte groß sein.

Erst über 30 Jahre später erwähnte Zimbardo in seinem Buch sogar ein Treffen mit den Wärtern noch **vor** dem Start des Experiments, hier ein kleiner Auszug:

Wir werden ihnen auf verschiedene Weise ihre Individualität nehmen. Sie werden Uniformen tragen, und niemand wird sie mit ihrem Namen ansprechen; sie werden Nummern haben und nur nach ihren Nummern gerufen werden. Im Allgemeinen sollte all dies bei ihnen ein Gefühl der Machtlosigkeit hervorrufen.

(aus dem Englischen nach [78], S. 55)

Ironischerweise spricht Zimbardo zunächst von der Forschungsfrage, wie sich Gefangene wohl unter solchen Umständen verhalten (vgl. [78], S. 55) - erst im Nachhinein entwickelte sich das Statement 'wir ließen die Wärter machen, was sie wollten, ohne einzugreifen' - wobei auch das obige Gespräch ein Beweis dagegen ist.

Dieser Abschnitt fokussiert sich zwar mehr auf Psychologie, doch es mag kaum ein besseres Beispiel zu finden sein, wie stark eine Studie verfälscht sein kann, um die Menschen auf dieser Welt zu beeinflussen und nach genau diesem Prinzip lassen sich alle möglichen 'mathematischen' und statistischen Aussagen im Zuge einer Studie verfälschen. Wenn eine Studie behauptet, '70 Prozent aller Menschen würden...', dann denken Sie daran, dass bereits während der Datenerhebung massive Manipulationen aufgetreten

sein können, aber nicht müssen. Ein ähnliches Beispiel dafür ist das Milgram-Experiment, dazu später mehr.

Suggestivfragen:

Wenn Sie das Prinzip solcher Manipulationen verstehen, dann werden Sie auch skeptischer bei statistischen Umfragewerten sein, wenn nicht die genaue Fragestellung vorliegt, denn auch hier kann die Manipulation unserer Merkmalsträger:innen - wie beim Stanford-Experiment - zu verheerend verfälschten statistischen Aussagen führen.

Ein fiktives Beispiel: eine Zeitschrift veröffentlicht Umfrageergebnisse: '90 Prozent finden, unser Steuersystem ist in Ordnung.'

Fehlvorstellung:

Das Volk ist zufrieden mit unserem Steuersystem und möchte nichts ändern.

Erklärung des Fehlers:

Solange wir nicht wissen, welche Frage exakt gestellt wurde, ist das Ergebnis relativ bedeutungslos. Verschiedene Fragestellungen können zu verschiedenen Ergebnissen führen:

- Finden Sie es in Ordnung, dass jemand Steuern zahlen muss?
- Finden Sie die Millionärssteuer in Ordnung?
- Finden Sie die Idee von Steuerklassen in Ordnung?
- Was wäre Ihnen lieber: unser derzeitiges Steuersystem oder dass die reichen zehn Prozent die Steuergesetze allein bestimmen?
- Ist es nicht so, dass die kalte Progression angesichts des Rückflusses einiger Steuergelder sinnvoll ist, da diese unter Umständen wieder redistributiert werden können? (Diese Frage soll die Befragten überrumpeln, einfach zuzustimmen)
- ...

Manche dieser Fragen sind überspitzt zeigen jedoch den suggestiven Charakter, mithilfe dessen man bewusst in eine Richtung der Antwortgebung getrieben wird. Daher sind möglichst neutrale und verständliche Fragen zu stellen, bei denen es ebenfalls mehrere Antwortmöglichkeiten gibt, welche ebenso möglichst nicht manipulativ sind.

Bemerkung: auch bei den Antwortmöglichkeiten ist die Skala wichtig - mit Skalen zu manipulieren, haben wir bereits kennengelernt. Erinnern Sie sich an die möglichen Abweichungen der Punkte allein bei der Notenskala zu einer Schularbeit. Wenn nun anstelle eines genauen Wertes eine von verschiedenen Klassen anzukreuzen ist, so können im Prozess der eigenen Datenerhebung auch bei einer unvoreilhaften Skalenwahl zwei komplett unterschiedliche Ergebnisse zustande kommen.

Korrekte Betrachtung des Sachverhalts:

Es eignet sich die Likert-Skala mit beispielsweise den Antwortmöglichkeiten 'stimme überhaupt nicht zu', 'stimme nicht zu', 'stimme weder zu noch lehne ab', 'stimme zu' und 'stimme voll und ganz zu' (vgl. [54]), wobei nach wie vor die Frage möglichst neutral und leicht verständlich bleiben muss. Achten Sie dabei auf die erwähnte Gefahr der Tendenz zur Mitte (daher gibt es auch eine Variante mit vier Antwortmöglichkeiten ohne 'Mitte', bei der sich zumindest tendenziell für eine der beiden Meinungen entschieden werden muss).

13.8 Wenn 'geheime' Absichten einer Datenerhebung erkannt werden

Milgram-Experiment:

Benannt nach dem Psychologen Stanley Milgram wurde 1961 ein Experiment durchgeführt, bei dem ein 'Schüler' durch die Versuchsperson mit Stromschlägen bestraft werden soll, wenn er bei einem Test eine falsche Antwort gibt - die Stromschläge werden immer stärker bis hin zu 'lebensbedrohlich' (wobei der Schüler nur ein Schauspieler war) (vgl. [7], S. 186 ff).

Zwei Drittel der Versuchspersonen sollen unter der Anleitung des Kittelträgers immer mehr Schocks bis zum tödlichen Schock abgegeben haben. Tatsächlich brachte [49] hervor: aufgrund von Aufzeichnungen durch Milgrams Assistenten stellte sich heraus, dass in 18 von 23 Fällen weniger starke Schocks verabreicht wurden, wenn man an das Experiment nicht voll glaubte, als wie wenn man daran voll glaubte.

Manche Teilnehmer:innen glaubten also nicht an die Echtheit des Experiments. Aus den Aufzeichnungen des Assistenten ging hervor, dass mit 56,07 Prozent nur knapp über die Hälfte der Teilnehmer:innen keinerlei Zweifel an der Echtheit des Experiments hatte. Und unter dieser Gruppe gab es auch die höchste Abbruchrate von 62,3 Prozent. Zugegebenermaßen sind die anderen 37,7 Prozent noch immer viel - doch bedenkt man

die anderen, die zweifelten und sich dadurch anders verhielten, waren es schlussendlich doch 'nur' noch 21,12 Prozent der Teilnehmenden, die keinerlei Zweifel an der Echtheit dieser Situation hatten und dennoch bis zum tödlichen Schock gingen.

Erneut: auch 21,12 Prozent sind zu viel, doch mathematisch gesehen ist es ein bedeutender Unterschied - aus den zunächst beinahe zwei Dritteln wurden zu guter Letzt rund 21 Prozent. Ein perfektes Beispiel dafür, wie der Prozess der Erhebung gravierende Auswirkungen auf das Endergebnis hat - und wie man sich als Forscher:in dennoch durch alternative Interpretationen (anstelle des Eingestehens einer mangelnden Qualität oder des Ausschlusses gewisser Merkmalsträger:innen aufgrund eines quasi gescheiterten Versuchs) selbst einreden kann, doch zu fundierten wissenschaftlichen Ergebnissen zu kommen.

Doppelblind-Studien:

Seien Sie jedoch getröstet - oft wird viel Wert auf Doppelblindstudien gelegt, um die Objektivität zu gewährleisten. Dabei wissen sowohl der/die Versuchsleiter:in während des Versuchs nicht, wer ein Placebo oder ein echtes Medikament erhält (vgl. [46]). So kann dies auch bei Fachliteratur (vgl. [36]) zur Anwendung kommen, wenn bei '(single) blind peer reviewed articles' ein:e Spezialist:in anonym die Validität des Artikels prüft, bevor dieser für eine Fachzeitschrift freigegeben wird.

13.9 Floor und ceiling effect

Beim Erstellen von Tests gilt es, einen guten Schwierigkeitsgrad zu finden. Wenn der Schwierigkeitsgrad eines Tests dermaßen hoch ist, dass sogar Merkmalsträger:innen mit einer starken Ausprägung des untersuchten Merkmals (beispielsweise IQ oder mathematische Fähigkeiten) im Test nur niedrige Werte erzielen, so spricht man vom floor effect respektive vom Bodeneffekt (vgl. [18]). Daraus resultiert auch, dass signifikante Unterschiede zwischen den Merkmalsträger:innen oder Gruppen schwer feststellbar sind.

Umgekehrt gibt es den ceiling effect (respektive Deckeneffekt), welcher auftritt, wenn der Test viel zu einfach ist (vgl. [17]).

13.10 Konfidenzintervalle

Nun liegen nach einer Erhebung die Daten vor. Doch erhoben wurden dabei aus Aufwands-, Zeit-, finanziellen oder anderen Gründen meist keine Grundgesamtheiten,

sondern repräsentative Stichproben. Und aufgrund der erhobenen Stichproben wollen wir mit den n Merkmalsträger:innen dann auch möglichst genau etwas über die 'breite Masse' aussagen, doch geht dies nur begrenzt genau.

Fehlvorstellung:

Wenn bei einer Umfrage von 150 Personen mit einer relativen Häufigkeit von 0,38 angegeben wurde, dass regelmäßig Brillen oder Kontaktlinsen getragen werden, so werden 38 Prozent der Bevölkerung Brillen tragen.

Erklärung des Fehlers:

Auch ausgehend davon, dass die Stichprobe repräsentativ ist, muss beachtet werden, dass die beobachtete relative Häufigkeit nicht automatisch der tatsächlichen relativen Häufigkeit entspricht. Daher geben wir ein Intervall an, das den wahren Anteil mit möglichst großer Wahrscheinlichkeit abdeckt.

Korrekte Betrachtung des Sachverhalts:

Satz 13.1. *Das γ -Konfidenzintervall für einen unbekannten relativen Anteil p an der Grundgesamtheit ist definiert durch*

$$\left[h - z \cdot \sqrt{\frac{h \cdot (1-h)}{n}}; h + z \cdot \sqrt{\frac{h \cdot (1-h)}{n}} \right]$$

wobei h die relative Häufigkeit in einer Stichprobe ist und der z -Wert abhängig vom γ -Vertrauensniveau ist, die gängigsten davon sind:

γ	90%	95%	99%
z	1,645	1,960	2,576

(vgl. [9], S. 11)

Das zufällige (da von der Stichprobe abhängige) Intervall überdeckt den unbekannten Anteil p mit einer Wahrscheinlichkeit von γ (vgl. [37], S. 6).

In unserem Fall geben wir anstelle der 38 Prozent ein Konfidenzintervall an, welches den wahren Anteil mit 95-prozentiger Wahrscheinlichkeit überdeckt:

$$\left[0,38 - 1,960 \cdot \sqrt{\frac{0,38 \cdot (1-0,38)}{150}}; 0,38 + 1,960 \cdot \sqrt{\frac{0,38 \cdot (1-0,38)}{150}}\right]$$

$$[0,374; 0,386]$$

Bemerkung: es ist eher ungeschickt, zu sagen, dass p mit einer Wahrscheinlichkeit von γ im Konfidenzintervall landet/liegt - so wird es jedoch oft erklärt. Irreführend dabei ist nämlich, dass p nirgendwo hinfällt oder variabel ist, sondern eben fest ist und eher das Konfidenzintervall 'zufällig' entstanden ist (vgl. [37], S. 6).

Achtung, Manipulation: offensichtlich trägt n zur Breite des Konfidenzintervalls bei. Bei einem Stichprobenumfang von 1500 anstelle der ursprünglichen 150 verkleinert sich das Intervall auf $[0,379; 0,381]$ - daher empfiehlt sich immer ein möglichst großer Stichprobenumfang.

Doch auch mit der Wahl von γ lässt sich viel beeinflussen - denn die Wahl obliegt jenen, die das Konfidenzintervall ermitteln/angeben. Wenn die Wahrscheinlichkeit groß sein soll, dass das Konfidenzintervall p abdeckt, dann muss das Intervall breiter angegeben werden, als wenn die Sicherheit der Aussage geringer sein darf.

Konfidenzintervalle können ebenfalls ermittelt werden, wenn nicht Anteile, sondern Mittelwerte zum Tragen kommen (vgl. [10], S. 18).

13.11 Hypothesentest

Ein sehr umfangreiches Thema wird hier in einem Kapitel kurz und bündig zusammengefasst - Denkfallen und Manipulationen stechen auch hier hervor. Zunächst zum Denkfehler:

Durch Forschungen und Datenerhebungen sollen mitunter veraltete Hypothesen widerlegt und durch neue ersetzt werden - beispielsweise bei einem bekannten p - dazu ein Beispiel (vgl. [37], S. 30): ein Bauer verwendet spezielle Nahrungsmittel, die seinen Angaben nach die Wahrscheinlichkeit erhöhen sollen, weibliche Kälber zu gebären anstatt männlicher.

Fehlvorstellung:

Und tatsächlich: in einer Versuchsserie kamen bei den Geburten 141 weibliche Kälber (im Gegensatz zu 115 männlichen) zur Welt - der Bauer scheint Recht zu haben.

Erklärung des Fehlers:

Es kann immer wieder Abweichungen geben (umgekehrt gedacht werden wohl auch ohne die spezielle Nahrung nie genau 50 Prozent der Geburten weibliche Kälber hervorbringen - 'mal mehr und 'mal weniger').

Korrekte Betrachtung des Sachverhalts:

Daher sehen wir uns an, wie unwahrscheinlich es ist, dass bei 256 Geburten 55,08 (oder mehr) Prozent der Kälber weiblich sind. Wenn dies unwahrscheinlich genug ist, dann verwerfen wir unsere bisherige Annahme und gehen davon aus, dass der Bauer tatsächlich etwas anders macht, das diesen Unterschied hervorruft.

Ein kurzer Ausflug zu den Begrifflichkeiten der Hypothesen (vgl. [37], S. 21 ff):

H_0 : *Nullhypothese* - die ursprüngliche Annahme, dass 50 Prozent weibliche Kälber geboren werden.

H_1 : *Alternativhypothese* - die neue Annahme, dass bei dem Bauern mehr als 50 Prozent weibliche Kälber geboren werden.

Dabei kann getestet werden ob der beobachtete Anteil: 1) größer als, 2) kleiner als, 3) ungleich p ist.

Die Nullhypothese kann also verworfen werden, wenn eben das Ereignis '141 oder mehr der 256 Geburten waren weiblich' unwahrscheinlich genug ist.

Auch hier muss ein Konfidenzniveau/Sicherheitsniveau angegeben werden - es sei vorab ein 95-prozentiges Konfidenzniveau gewählt worden. Demnach muss also die Wahrscheinlichkeit, dass mindestens 55,08 Prozent der 256 Geburten weiblich waren, unter fünf Prozent liegen. Da sich diese Arbeit kaum mit Wahrscheinlichkeitsrechnung auseinandersetzt, wird die Herleitung und Berechnung über die Standardnormalverteilung nicht berücksichtigt - viel mehr wird ein einfaches und allen zugängliches Werkzeug (GeoGebra) gezeigt, dass die Antwort auf Hypothesentests sehr schnell liefert.

Das Ergebnis zeigt, dass diese Wahrscheinlichkeit ($p = 5,21\%$) noch zu hoch ist, um die Nullhypothese zu verwerfen.

Vielleicht fiel Ihnen bald auf, dass die Hypothese verworfen worden wäre, wenn wir ein 90-prozentiges Konfidenzniveau angenommen hätten. Und genau hier liegt eine Gefahr der Manipulation: Forscher:innen (aber auch Studierende und andere) könnten versucht sein, einfach nachträglich das Konfidenzniveau zu ändern - denn die Schlagzeile 'Neues Tierfutter bringt mehr weibliche Kälber' erregt mehr Aufsehen.

Verteilung	Statistik
Gauß-Test für Anteile ▼	
Nullhypothese $p =$ <u>0.5</u>	
Alternativhypothese ☐ $<$ ☒ $>$ ☐ $\neq$	
Stichprobe	
Erfolge	<u>141</u>
n	<u>256</u>
<u>Ergebnis</u>	
Gauß-Test für Anteile	
Erfolge	141
n	256
z	1.625
p	0.0521

Abbildung 28: Hypothesen-Test für Anteile

Auch in diesem Beispiel zeigt sich: wenn der Bauer die Versuchsserie größer aufzieht (also n deutlich größer wird, während die relative Häufigkeit der weiblichen Geburten gleichbleibt), so wird auch der Fall eindeutiger:

Hier ist die Wahrscheinlichkeit praktisch Null, also kann auch bei einem Konfidenzniveau von 99 Prozent die Nullhypothese verworfen werden.

Und hier liegt die nächste Gefahr der Manipulation vor: durch einfaches Kopieren respektive Verfielfachen der erhobenen Daten (also das Vortäuschen weiterer Merkmals-träger:innen anstelle des Erhebens weiterer Daten) können beispielsweise Signifikanzen plötzlich 'bewiesen' (und Hypothesen verworfen) werden, was zuvor noch nicht möglich war. Wenn 600 von 1 000 Personen etwas zustimmen, wirkt dies zuverlässiger als bei sechs von zehn befragten Personen oder wenn man drei von vier Münzwürfe korrekt vorhersagt, ist es bei weitem nicht so beeindruckend wie 300 von 400 korrekte Vorhersagen, obgleich die relative Häufigkeit jeweils dieselbe bleibt. Denn Glückstreffer und Schwankungen wirken sich bei kleinen Anzahlen schnell aus, weshalb Forschungen

Verteilung	Statistik
Gauß-Test für Anteile ▼	
Nullhypothese $p =$ <u>0.5</u>	
Alternativhypothese ☐ $<$ ☒ $>$ ☐ $\neq$	
Stichprobe	
Erfolge	<u>1410</u>
n	<u>2560</u>
 <u>Ergebnis</u>	
Gauß-Test für Anteile	
Erfolge	1410
n	2560
z	5.1387
p	0

Abbildung 29: Hypothesen-Test für Anteile, mit 10-fachem n

immer eine möglichst große Anzahl an Datensätzen erheben wollen. Wenn nun der Bauer seine Daten kopiert und somit vervielfältigt, so ergibt sich auf einmal ein anderes Ergebnis und wir würden die Alternativhypothese annehmen, was selbstverständlich alles andere als eine wissenschaftliche Herangehensweise darstellt. Dieser Betrug kann durch Leser:innen nicht nachvollzogen werden und die Forschung sollte daher möglichst gut betreut und kontrolliert werden.

Anmerkung: auch beim Testen von Hypothesen kann man nicht nur Anteile, sondern auch Mittelwertunterschied auf Signifikanz überprüfen.

Ein weiteres Beispiel, welches dieses Kapitel zusammen mit den 'Quellen in unserem Kopf' abrundet, beruht im 'Mythos des 7. Games' (vgl. [59]): die Rede ist von Tennis. Dort werden auf Profi-Ebene die Sätze auf sechs gewonnene Games mit zwei Games Abstand gespielt (beispielsweise 6:3 6:4) - Ausnahme ist der Stand von 5:5, doch dies touchiert dieses Thema nicht. Worum geht es?

Regelmäßig hört man Coaches und Kommentator:innen davon sprechen, welche Bedeutung das 7. Game (jedoch nur beim Stand von 3:3) hat - der Satz käme bald zu Ende und der Druck steige, den eigenen Aufschlag durchzubringen oder zu breaken. 'Break' bedeutet, dass der oder die Aufschläger:in es nicht schafft, den eigenen Aufschlag zum 4:3 zu verwerten, sondern der oder die Gegner:in mit 3:4 davonzieht. Und genau im 7. Game soll bei Gleichstand der Knackpunkt sein. Die Denkfalle, die unser Kopf uns stellt, ist einerseits 'was alle denken, muss stimmen', andererseits auch das Ausblenden der Seltenheit beziehungsweise das bewusstere Wahrnehmen eines Breaks bei 3:3 als bei anderen Spielständen.

Tatsächlich besteht hier keine hinreichende Tatsache: jemand untersuchte Sätze in fünfstelliger Anzahl (vgl. [59]). Bei 11 000 Matches der ATP-Tour (männliche Tennis-Profi-Tour) erreichten 11 421 Sätze den Stand von 3:3. 9 080-mal konnte der Aufschläger seinen Aufschlag zum 4:3 durchsetzen.

Gauß-Test für Anteile ▼

Nullhypothese $p =$ 0.802

Alternativhypothese ☐ $<$ ☐ $>$ ☒ $\neq$

Stichprobe

Erfolge 9080

n 11421

Ergebnis

Gauß-Test für Anteile

Erfolge	9080
n	11421
z	-1.8701
p	0.0615

Abbildung 30: Hypothesen-Test '7. Game'

Ausgehend davon, dass die untersuchten Spieler erfahrungsgemäß 80,2 Prozent ihrer Aufschlag-Games durchbringen sollten, dies aber laut Alternativhypothese besser oder schlechter tun und wir ein 95-prozentiges Konfidenzniveau festlegen, stellen wir fest: in

diesem Fall wird die Nullhypothese beibehalten.

Bemerkung: dabei muss man erwähnen, dass die Quelle kein Konfidenzniveau verzeichnete, dieses wurde in dieser Masterarbeit selbst gewählt. Andererseits muss jedoch auch berücksichtigt werden, dass der Andorra-Effekt einen Teil zur Verfälschung der Ergebnisse beitragen könnte: wenn alle davon sprechen, wie wichtig das 7. Game sei, so kann man erst recht unter dem Druck versagen und eben dieses Game verlieren - doch dies lässt sich natürlich schwer testen.

Und zu guter Letzt: auch hier könnte manipuliert werden, indem die Alternativhypothese nicht lautet, dass der Anteil ungleich 0,82 ist - sondern, dass der Anteil kleiner als 0,82 ist, dadurch ändert sich der Test und bei gleichem Konfidenzniveau könnte die Nullhypothese verworfen werden. Wieder einmal zeigt sich, wie wichtig es ist, zuerst alle Entscheidungen für den Test festzulegen, bevor die Forschung startet. Denn sonst gilt, wie bereits erwähnt: man schießt dabei zuerst einen Pfeil und malt danach eine Zielscheibe um den Pfeil herum.

14 Fazit

In dieser Arbeit wurde die Frage beforscht, wie Erhebungen und Grafiken manipuliert und bestehende Daten missinterpretiert werden können. Die verschiedenen Anwendungsfehler und -strategien zeigten, dass die bewusste Auswahl und der Ausschluss bestimmter Fakten zu verschiedensten Argumenten führen kann. Auch unbeabsichtigte Fehler und mangelndes Vorwissen zeigen noch auf, dass jede Statistik genau betrachtet werden sollte und der Ablauf von Forschungen muss genau unter die Lupe genommen werden, um aussagekräftige Ergebnisse zu erlangen.

Zum Abschluss dieser Arbeit erhalten Sie noch eine Sammlung an praktischen Tipps. Die Gefilde der Statistik können sehr komplex sein - wenn Ihnen eine Statistik vorgesetzt wird, kann es also nicht schaden, sich an eine Checkliste zu halten:

- Ruhe bewahren und nichts überstürzen
- Daten aus ähnlichen Situationen besorgen und vergleichen
- Details zu den Daten anfordern
- Prüfen: ist die vorgelegte Zahl überhaupt ermittelbar?
- Prüfen: wie sicher kann eine Prognose sein?
- Überschlagsrechnung durchführen
- Prüfen: was genau wurde untersucht oder belegt?
- Doppelte Vorsicht bei besonders großen Geschäften
- Prüfen: ist die Quelle seriös?
- Grafiken prüfen
- Daten hinter den Grafiken anfordern und prüfen
- Querdenken statt stromlinienförmig
- Trauen Sie sich, selbst nachzuforschen
- Die fünf größten Fehler vermeiden:
 - Ich habe keine Zeit

- Das sagen alle, also muss es stimmen
- Ich kann das ohnehin nicht überprüfen
- Da steht '2,6' - daher wurde der Wert genau gemessen/berechnet
- Wenn das immer so weiter geht, dann wird ja [...] passieren...

entnommen von [6] (S. 270)

Die einzelnen Punkte der Checkliste müssen nicht in dieser Reihenfolge abgearbeitet werden, doch es empfiehlt sich, möglichst viele dieser Vorschläge zu berücksichtigen.

'Last but not least' bleibt zu sagen, dass Sie nun in Berührung mit allen möglichen Denkfallen und vor allem auch Manipulationstechniken gekommen sind. Explizit soll nochmals an dieser Stelle erwähnt werden, dass mit diesem Wissen rücksichtsvoll umgegangen werden muss. Sich selbst vor Manipulationen und Denkfallen in der Statistik zu schützen, war das primäre Ziel dieser Arbeit. Dabei wurden mancherlei reale und fiktive Situationen angeführt, welche die Aspekte näher beleuchten. Mit dieser Arbeit liegt eine umfassende Sammlung vor, auf die Sie schnell zurückgreifen können, wenn Sie nicht mehr alle Denkfallen und Manipulationen im Kopf haben - hier sind sämtliche davon nennenswerten vorhanden, die im Alltag auftreten.

Nicht hinter jeder Statistik verbirgt sich ein Fehler oder eine Manipulation, doch auf jeden Fall sei Vorsicht geboten, wenn Sie einer Statistik über den Weg laufen - wir schließen mit dem wohl bekanntesten Zitat bezüglich der Statistik:

Traue keiner Statistik, die du nicht selbst gefälscht hast.
(Ursprung unbekannt, Winston Churchill zugeschrieben)

Literatur

- [1] Absolventa. *Pareto-Prinzip im Zeitmanagement*. URL: <https://www.absolventa.de/karriereguide/zeitmanagement/pareto-prinzip> (besucht am 16. 11. 2023).
- [2] Alpecin. *Der Alpecin Faktencheck*. URL: <https://www.alpecin.com/de-de/faktencheck>.
- [3] Bettermarks. *Manipulationen*. URL: <https://de.bettermarks.com/mathe/manipulationen/> (besucht am 23. 11. 2023).
- [4] W. Blum. *Ewig lockt die Schnellstraße*. URL: <https://www.sueddeutsche.de/wissen/ewig-lockt-die-schnellstrasse-1.913440> (besucht am 14. 11. 2023).
- [5] G. Bosbach und J. J. Korff. *Die Zahlentricks - Das Märchen von den aussterbenden Deutschen und andere Statistiklügen*. München: Heyne Verlag, 2017.
- [6] G. Bosbach und J. J. Kroff. *Lügen mit Zahlen: Wie wir mit Statistiken manipuliert werden*. ger. 4. Aufl. 2012. Neumarkter Str. 28, München: Wilhelm Heyne Verlag, 2011.
- [7] R. Bregman, U. Faure und G. Busse. *Im Grunde gut: Eine neue Geschichte der Menschheit*. Rowohlt E-Book, 2020. URL: <https://books.google.at/books?id=41DADwAAQBAJ>.
- [8] C. Brell, J. Brell und S. Kirsch. *Statistik von Null auf Hundert: Mit Kochrezepten schnell zum Statistik-Grundwissen*. ger. 2. Aufl. 2017. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016.
- [9] Bundesministerium für Bildung, Wissenschaft und Forschung. *Formelsammlung Mathematik (AHS)*. 2019.
- [10] Bundesministerium für Bildung, Wissenschaft und Forschung. *Formelsammlung*. 2022.
- [11] Bundesministerium für Bildung, Wissenschaft und Forschung. *Zuverdienstgrenzen bei Bezug der Studienbeihilfe*. URL: https://www.oesterreich.gv.at/themen/bildung_und_neue_medien/universitaet/2/Seite.160806.html (besucht am 09. 11. 2023).
- [12] B. Burr. <https://scrapsfromtheloft.com/comedy/bill-burr-people-2012-full-transcript/>. URL: <https://scrapsfromtheloft.com/comedy/bill-burr-people-2012-full-transcript/> (besucht am 13. 09. 2023).
- [13] B. Csoka. *Equal Pay Day 2020: 22. Oktober, 71 Tage Gratisarbeit*. URL: <https://awblog.at/equal-pay-day-2020/> (besucht am 04. 12. 2023).

- [14] Der Westen. *Parship, die Singles und der 11-Minuten-Irrglaube*. URL: <https://www.derwesten.de/panorama/parship-die-singles-und-der-11-minuten-irrglaube-id11443266.html> (besucht am 08.09.2023).
- [15] Die Bundeswahlleiterin. *Hare/Niemeyer*. URL: <https://www.bundeswahlleiterin.de/service/glossar/h/hare-niemeyer.html> (besucht am 10.11.2023).
- [16] Die Presse. *Jeder Mensch verbringt drei Jahre seines Lebens auf dem Klo*. URL: <https://www.diepresse.com/4864125/jeder-mensch-verbringt-drei-jahre-seines-lebens-auf-dem-klo> (besucht am 27.10.2023).
- [17] Dorsch. *ceiling effect*. URL: <https://dorsch.hogrefe.com/stichwort/ceiling-effect> (besucht am 25.11.2023).
- [18] Dorsch. *floor effect*. URL: <https://dorsch.hogrefe.com/stichwort/floor-effect> (besucht am 25.11.2023).
- [19] P. Engel. *This Chart Shows An Alarming Rise In Florida Gun Deaths After 'Stand Your Ground' Was Enacted*. URL: <https://www.businessinsider.com/gun-deaths-in-florida-increased-with-stand-your-ground-2014-2> (besucht am 24.11.2023).
- [20] K. H. Eriksson und A. Sandberg. „Gender Differences in Initiation of Negotiation: Does the Gender of the Negotiation Counterpart Matter?“ In: *Negotiation Journal* 28.4 (2012), S. 407–428. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1571-9979.2012.00349.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1571-9979.2012.00349.x>.
- [21] extraETF. *Inflation berechnen*. URL: <https://extraetf.com/de/finanzrechner/inflations-rechner> (besucht am 09.11.2023).
- [22] L. Fahrmeir u. a. *Statistik: Der Weg zur Datenanalyse*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2016. URL: <https://books.google.at/books?id=rKveDAAAQBAJ>.
- [23] M. Falkner. *Erbgut - der Mensch ist zur Hälfte eine Banane*. URL: <https://www.morgenpost.de/kultur/berlin-kultur/article115517396/Erbgut-der-Mensch-ist-zur-Haelfte-eine-Banane.html> (besucht am 08.09.2023).
- [24] M. Fehndrich. *Parteizuwachsparadoxon*. URL: <https://www.wahlrecht.de/verfahren/paradoxien/partezuwachs.html> (besucht am 10.11.2023).
- [25] M. Fehndrich. *Populations-Paradoxon*. URL: <https://www.wahlrecht.de/verfahren/paradoxien/population.html> (besucht am 10.11.2023).

- [26] N. Fenton und M. Neil. *Simpson's paradox again: fixing an example from Pearl's "Book of Why"(with video)*. URL: <https://probabilityandlaw.blogspot.com/2020/03/simpsons-paradox-again-fixing-example.html> (besucht am 24.11.2023).
- [27] S. Filan. *Shane Filan Quotes*. URL: <https://quotefancy.com/quote/1731610/Shane-Filan-You-re-only-as-good-as-your-last-album>.
- [28] Frauenserviceportal. *Equal Pay Day 2023 in Österreich*. URL: <https://www.bundeskanzleramt.gv.at/frauenserviceportal/aktuell/equal-pay-day-2023-in-oesterreich.html> (besucht am 04.12.2023).
- [29] J. Frost. *Benford's Law Explained with Examples*. URL: <https://statisticsbyjim.com/probability/benfords-law/> (besucht am 15.11.2023).
- [30] D. P. Fry. *A Critique of the Killers-Have-More-Kids Idea*. URL: <http://naturalbornnonkillers.blogspot.com/> (besucht am 23.11.2023).
- [31] Futurezone. *Bildschirmzeit in Österreich: So viel Zeit verbringen wir täglich am Computer, Smartphone Co*. URL: <https://futurezone.at/cm/bildschirmzeit-in-oesterreich-so-viel-zeit-verbringen-wir-taeglich-am-computer-smartphone-co/402496234> (besucht am 26.12.2023).
- [32] G. Gigerenzer und A. Bartel. *Unstatistik des Monats: Der Gardasee ist halb leer*. URL: <https://www.hardingcenter.de/de/unstatistik/unstatistik-des-monats-der-gardasee-ist-halb-leer> (besucht am 17.10.2023).
- [33] T. Grams. *Von der Zinseszinsrechnung zur Populationsdynamik*. URL: <https://www2.hs-fulda.de/~grams/OekoSim0/OeSi3.html> (besucht am 15.11.2023).
- [34] J. Hartung, B. Elpelt und K. Klösener. „Zur Geschichte der Statistik“. ger. In: *Statistik*. Germany: Walter de Gruyter GmbH, 2002.
- [35] F. Hofbauer. *Stochastik. Eine Vorlesung für das Lehramtsstudium*. 2017.
- [36] C. Hornung. *peer review – Von Doppelblindverfahren, Herausgeberentscheidungen und anderen Formen der Qualitätsprüfung*. URL: <https://blog.fid-romanistik.de/2019/02/05/peer-review-von-doppelblindverfahren-herausgeberentscheidungen-und-anderen-formen-der-qualitaetspruefung/> (besucht am 24.11.2023).
- [37] H. Humenberger. *C Beurteilende Statistik*. 2020.

- [38] Kleine Zeitung. *Kann man am Gardasee bei so niedrigem Wasserstand noch Urlaub machen?* URL: https://www.kleinezeitung.at/lebensart/reise/6284337/Gute-Frage_Kann-man-am-Gardasee-bei-so-niedrigem-Wasserstand-noch#:~:text=Der%20Gardasee%20wird%20seit%201949,der%20Bew%C3%A4sserung%20in%20der%20Landwirtschaft. (besucht am 17.10.2023).
- [39] J. Köehler. *Hintergrund und Geschichte der Statistik*. URL: <https://de.linkedin.com/learning/statistik-grundlagen-1-grundbegriffe-und-deskriptive-statistik/hintergrund-und-geschichte-der-statistik#:~:text=Die%20ersten%20Anf%C3%A4nge%20der%20Statistik,die%20es%20bis%20heute%20gibt.> (besucht am 10.10.2023).
- [40] W. König. *Paradoxa in der Wahrscheinlichkeitsrechnung*. 2008.
- [41] Känguru der Mathematik. *Statistiken 2023*. URL: <https://www.kaenguru.at/ergebnisse.html> (besucht am 05.05.2023).
- [42] D. Malinowski. *Paradoxa in der Wahrscheinlichkeitsrechnung*. 2022.
- [43] maths2mind. *Relative Änderung*. URL: <https://www.maths2mind.com/schluesselwoerter/relative-aenderung.>
- [44] F. Mischler. „VERDIENSTUNTERSCHIEDE ZWISCHEN MÄNNERN UND FRAUEN“. In: Statistisches Bundesamt, 2021.
- [45] J. Murtagh. *What Is Benford’s Law? Why This Unexpected Pattern of Numbers Is Everywhere*. URL: <https://www.scientificamerican.com/article/what-is-benfords-law-why-this-unexpected-pattern-of-numbers-is-everywhere/> (besucht am 15.11.2023).
- [46] Öffentliches Gesundheitsportal Österreich. *Doppel-blinde Studie*. URL: <https://www.gesundheit.gv.at/lexikon/D/lexikon-doppel-blinde-studie.html#:~:text=Doppel%2Dblind%20bedeutet%2C%20dass%20weder,m%C3%B6glichst%20objektive%20Auswertung%20der%20Ergebnisse.> (besucht am 24.11.2023).
- [47] A. O’Dwyer. *Two counterintuitive results*. URL: <https://quuxplusone.github.io/blog/2021/08/12/roughgarden/> (besucht am 25.03.2024).
- [48] F. Pahlke. *Korrelationskoeffizienten interpretieren*. URL: <https://www.evidat.com/faqs/correlation-coefficient/> (besucht am 27.11.2023).
- [49] G. Perry u. a. „Credibility and Incredulity in Milgram’s Obedience Experiments: A Reanalysis of an Unpublished Test“. In: *Social Psychology Quarterly* 83.1 (2020), S. 88–106. URL: <https://journals.sagepub.com/doi/10.1177/0190272519861952> (besucht am 10.10.2023).

- [50] Personio. *Beurteilungsfehler: Arten von Fehlern, Folgen und Beispiele*. URL: <https://www.personio.de/hr-lexikon/beurteilungsfehler/#was-sind-beurteilungsfehler-und-warum-machen-wir-sie>.
- [51] P. Planing. *Regression*. URL: <https://statistikgrundlagen.de/ebook/chapter/regression/> (besucht am 27.11.2023).
- [52] J. Puhani. *Statistik*. Springer Fachmedien Wiesbaden, 2020.
- [53] Puls24. *Jährlich rund 525 Unfälle wegen Missachtung roter Ampeln*. URL: <https://www.puls24.at/news/chronik/jaehrlich-rund-525-unfaelle-wegen-missachtung-roter-ampeln/278714> (besucht am 08.09.2023).
- [54] Qualtrics XM. *Likert-Skala: Definition, Beispiel und Vorteile*. URL: <https://www.qualtrics.com/de/erlebnismanagement/marktforschung/likert-skala/> (besucht am 23.11.2023).
- [55] A. Quatember. *Unsinn in den Medien – Vom allzu sorglosen Umgang mit Daten: Datenqualität*. URL: <https://www.jku.at/fileadmin/gruppen/118/groramng.pdf> (besucht am 09.11.2023).
- [56] M. Rassem und J. Stagl. „ZUR GESCHICHTE DER STATISTIK UND STAATSBESCHREIBUNG IN DER NEUZEIT: Grundriß für einen Arbeitskreis“. In: *Zeitschrift für Politik* 24.1 (1977), S. 81–86. URL: <http://www.jstor.org/stable/24224414> (besucht am 10.10.2023).
- [57] W. Rogers. *Will Rogers Quotes*. URL: https://www.brainyquote.com/quotes/will_rogers_128074.
- [58] M. Rohrdorfer. *Großraming und Welssehen zur Winterzeit am wenigsten Sonne*. URL: <https://www.nachrichten.at/oberoesterreich/Grossraming-und-Welssehen-zur-Winterzeit-am-wenigsten-Sonne;art4,27944> (besucht am 09.11.2023).
- [59] J. Sackmann. *How Important is the Seventh Game of the Set?* URL: <https://www.tennisabstract.com/blog/2015/09/24/how-important-is-the-seventh-game-of-the-set/> (besucht am 03.12.2023).
- [60] B. Schäfer und M. Timme. „Stromnetze für die Zukunft“. In: *Spektrum der Wissenschaft* 10 (2023), S. 72–77. URL: www.spektrum.de/artikel/2169867 (besucht am 10/2023).
- [61] P. Schmitt. *Kein unendliches Wachstum*. URL: <https://www.faz.net/aktuell/gesellschaft/menschen/prognose-zur-weltbevoelkerung-kein-unendliches-wachstum-16862257.html> (besucht am 09.11.2023).

- [62] sengpielaudio. *Tontechnik-Rechner - sengpielaudio*. URL: <http://www.sengpielaudio.com/Rechner-pegelaenderung.htm> (besucht am 07.11.2023).
- [63] B. Simmons. *Ewing Theory 101*. URL: <https://www.espn.com/espn/page2/story?page=simmons/010509a> (besucht am 14.11.2023).
- [64] S. Singh. *Alex Through the Looking-Glass review – adventures with a maths demon*. URL: <https://www.theguardian.com/books/2014/may/04/alex-through-the-looking-glass-review-bellos-maths-demon> (besucht am 16.11.2023).
- [65] M. Speck. *DEZIBEL (DB) VERGLEICH – BEI WELCHEM GERÄUSCHPEGEL WIRD ES ZU LAUT?* URL: <https://www.berlinearguard.com/blogs/gehoerschutz-ratgeber/dezibel-vergleich> (besucht am 07.11.2023).
- [66] Statistik Austria. *Arbeitslose, Arbeitssuchende*. URL: <https://www.statistik.at/statistiken/arbeitsmarkt/arbeitslosigkeit/arbeitslose-arbeitssuchende> (besucht am 20.10.2023).
- [67] Statistik Austria. *Bevölkerung zu Jahres-/Quartalsanfang*. URL: <https://www.statistik.at/statistiken/bevoelkerung-und-soziales/bevoelkerung/bevoelkerungsstand/bevoelkerung-zu-jahres--quartalsanfang> (besucht am 08.09.2023).
- [68] Statistik Austria. *Straßenverkehrsunfälle*. URL: <https://www.statistik.at/statistiken/tourismus-und-verkehr/unfaelle/strassenverkehrsunfaelle> (besucht am 08.09.2023).
- [69] Stern. *Beliebtes Urlaubsziel: Pegel des Gardasees bedenklich niedrig*. URL: <https://www.stern.de/panorama/wetter/italien--pegel-des-gardasees-bedenklich-niedrig---alarmierende-duerre--video--33399934.html> (besucht am 17.10.2023).
- [70] P. Talwalkar. *Braess's Paradox*. URL: <https://datafilos.blogspot.com/2019/10/braesss-paradox.html> (besucht am 14.11.2023).
- [71] P. Talwalkar. *How Closing Roads Could Speed Up Traffic - The Braess Paradox*. URL: <https://www.youtube.com/watch?v=8mlH9bnvWVE> (besucht am 14.11.2023).
- [72] R. Tofan. *Polizeiliche Kriminalstatistik 2022*. URL: [https://bmi.gv.at/magazin/2023_05_06/01_Polizeiliche_Kriminalstatistik_2022.aspx#:~:text=2022%20wurden%20%C3%B6sterreichweit%2078.836%20Gewaltdelikte,wurde%20\(2021%3A%2067.441\).](https://bmi.gv.at/magazin/2023_05_06/01_Polizeiliche_Kriminalstatistik_2022.aspx#:~:text=2022%20wurden%20%C3%B6sterreichweit%2078.836%20Gewaltdelikte,wurde%20(2021%3A%2067.441).) (besucht am 08.09.2023).

- [73] L. Wade. *How to Lie with Statistics: Stand Your Ground and Gun Deaths*. URL: <https://thesocietypages.org/socimages/2014/12/28/how-to-lie-with-statistics-stand-your-ground-and-gun-deaths/> (besucht am 24.11.2023).
- [74] Welt der BWL. *Perzentil*. URL: <https://welt-der-bwl.de/Perzentil> (besucht am 12.12.2023).
- [75] wetter.at. *Wärmster Winter aller Zeiten*. URL: <https://www.wetter.at/wetter/oesterreich-wetter/waermster-winter-aller-zeiten/223267664> (besucht am 20.10.2023).
- [76] Wikibrief. *Aufteilungsparadoxon*. URL: https://de.wikibrief.org/wiki/Apportionment_paradox (besucht am 10.11.2023).
- [77] P. G. Zimbardo. *Audio transcript-tape E*. Stanford University, 1971. URL: purl.stanford.edu/tb853qw7615.
- [78] P. G. Zimbardo. *The Lucifer effect : how good people turn evil*. eng. London [u.a.]: Random House, 2007.

Abbildungsverzeichnis

1	Den Versuchsgruppen vorgelegte Fotos [2]	19
2	Lineare Regressionsgerade [51]	58
3	'Neutrales' Diagramm	66
4	Diagramm mit 'abgeschnittener' y-Achse	67
5	Diagramm mit eingeschränktem Zeitraum	67
6	Diagramm mit fehlenden Jahreswerten	68
7	Werte entgleiten bis nach außerhalb des 'Rahmens'	68
8	'Neutrales Diagramm' 2	69
9	Diagramm mit gestreckter/gestauchter x-/y-Achse	70
10	Diagramm mit logarithmischer Achsenskalierung	70
11	Diagramm zu 'Gun deaths in Florida' original und mit gewohnter y-Achse [73]	71
12	Diagramm mit invertierter y-Achse	71
13	Diagramm mit invertierter x-Achse	72
14	Kreisdiagramm mit 3D-Effekt	72
15	Beschäftigungsverhältnis [3]	73
16	Diagramm mit Piktogrammen [3]	73
17	Verzerrte Darstellung (links) vs. korrekte Verdoppelung der Fläche (in der Mitte und rechts)	74
18	Vereinfachtes Verkehrsnetz	82
19	Vereinfachtes Verkehrsnetz mit 200 Fahrzeugen	82
20	Neues Verkehrsnetz mit Zubringer	83
21	Neues Verkehrsnetz bringt scheinbaren Nutzen	83

22	Neue Route suboptimal	84
23	Neues Verkehrsnetz schlussendlich für alle suboptimal	85
24	Braess-Paradoxon anhand einer Feder (Spring-Paradoxon) [47]	87
25	Logistisches Wachstum einer Population	89
26	Punktewolke vor dem Simpson-Paradoxon [26]	95
27	Simpson Paradoxon [26]	96
28	Hypothesen-Test für Anteile	121
29	Hypothesen-Test für Anteile, mit 10-fachem n	122
30	Hypothesen-Test '7. Game'	123