

Die Statistik beschäftigt sich mit der Analyse von Daten, in denen gewisse Unsicherheiten vorhanden sind, die wir später im Rahmen der Mathematik des Zufalls modellieren werden. Auf welche Art und Weise die Daten erhoben werden, beeinflusst ihre Qualität und damit auch die Gültigkeit von Analysen dieser Daten. Was bei der Erhebung von Daten zu beachten ist, damit aussagekräftige Ergebnisse erzielt werden können, wird in diesem Kapitel erläutert.

Exemplarisch besprochen wird die Erhebung von Daten im Rahmen von Studien und Umfragen. Zunächst wird dazu auf kontrollierte Studien und Beobachtungsstudien eingegangen, und es werden die dabei auftretenden Probleme besprochen. Anschließend wird die Erhebung von Daten im Rahmen von Umfragen dargestellt.

Auch wenn die meisten Leser dieses Buches niemals selbst Daten erheben werden, ist es doch entscheidend, darüber Bescheid zu wissen: Denn selbst die beste statistische Auswertung eines Datensatzes ist nutzlos, wenn bei der Erzeugung des Datensatzes Fehler gemacht werden, denn dann sind die bei der Analyse des Datensatzes gezogenen Schlüsse ungültig. Insofern sind Kenntnisse über mögliche Fehler bei der Erhebung von Daten auch bei der Interpretation von Ergebnissen statistischer Untersuchungen sehr nützlich.

2.1 Kontrollierte Studien

Kontrollierte Studien werden im Folgenden anhand des Vorgehens bei der Überprüfung der Wirksamkeit eines Medikamentes auf der Basis von Dimethylfumarat (Fumarsäure) zur Behandlung der Multiplen Sklerose eingeführt.¹

Multiple Sklerose ist eine chronisch entzündliche Erkrankung des zentralen Nervensystems. Dabei greifen Zellen des Immunsystems entzündete Nervenzellen an, was zu vielfältigen Beschwerden (u. a. Gehbehinderungen, in extremen Fällen auch Tod durch Ausfall der Atmung) führt. Allein in Deutschland sind ca. 130.000 Menschen an Multipler Sklerose erkrankt. Das Erscheinungsbild bzw. der Verlauf

der Krankheit ist vielfältig, weshalb Multiple Sklerose auch als Krankheit mit den 1000 Gesichtern bezeichnet wird. Meist tritt sie erstmals im Alter zwischen 20 und 30 Jahren auf, hat dann zunächst 10 bis 25 Jahre lang einen schubförmigen Verlauf, bei dem (evtl. im Abstand von mehreren Jahren) Entzündungsschübe auftreten, deren Symptome sich anschließend zumindest teilweise wieder zurückbilden. Nach der schubförmigen Phase kann dann die sogenannte chronisch-progressive Phase folgen, bei der es zu fortschreitenden Behinderungen kommt.

Die Diagnose der Multiplen Sklerose erfolgt durch den Neurologen, der dazu heutzutage primär Bilder des Nervensystems aus dem Kernspintomografen verwendet und darin nach typischen Anzeichen für eine Entzündung schaut. Abgesichert wird diese Diagnose durch eine Untersuchung der Flüssigkeit im Rückenmark des Patienten auf Entzündungszellen.

Die Krankheit ist unheilbar. Seit ca. 1990 wird Multiple Sklerose in der sogenannten Basistherapie mit β -Interferonen behandelt, was das Fortschreiten der Krankheit oft stark verlangsamt. Diese wirken auf das Immunsystem ein und bewirken, dass dieses Nervenzellen nicht oder nur vermindert angreift. Die Einnahme von β -Interferonen ist aber aufwendig, da diese unter die Haut gespritzt werden müssen. Weiter führt die Behandlung zu vielfältigen Nebenwirkungen (z. B. häufigeres Auftreten von grippalen Infekten aufgrund der Einwirkung auf das Immunsystems, oder teilweise auch allergische Reaktionen) und ist extrem teuer (ca. 20.000 Euro pro Jahr pro Patient, wobei die Behandlung im Prinzip lebenslang erfolgen muss). Dies führt dazu, dass entsprechende Medikamente für die Pharmaindustrie extrem lukrativ sind, d. h., es handelt sich um sogenannte Blockbuster mit einem Jahresumsatz von mehr als einer Milliarde Euro.

Längere Zeit hat die Pharmaindustrie versucht, ein Medikament für die Basistherapie der Multiplen Sklerose zu entwickeln, das in Tablettenform eingenommen werden kann, den Verlauf der Multiplen Sklerose mindestens so stark abschwächt wie die bekannten β -Interferone und dabei möglichst wenige Nebenwirkungen hat.

Ein Wirkstoff, der dabei betrachtet wurde, war Dimethylfumarat, das seit 1994 erfolgreich zur Behandlung der sogenannten Schuppenflechte eingesetzt wurde. Schuppenflechte ist eine (nicht ansteckende) chronisch-entzündliche Hautkrankheit. Sie führt zu scharf begrenzten, roten und mit weißen Schuppen bedeckten Hautausschlägen, die teilweise jucken. Ebenso wie bei der Multiplen Sklerose ist auch bei der Schuppenflechte das Immunsystem beteiligt, und daher war es naheliegend, dass dieser Wirkstoff auch bei Multipler Sklerose helfen könnte. Erste Anhaltspunkte dafür hatte man schon recht früh durch die Beobachtung, dass eine Behandlung der Schuppenflechte mit Dimethylfumarat bei gleichzeitig an Multipler Sklerose erkrankten Patienten zu einer Verbesserung der Symptome der Multiplen Sklerose geführt hatte.

Ab dem Jahre 2003 sicherte sich das amerikanische Pharmaunternehmen Biogen das Patent auf die Behandlung von Multipler Sklerose mit Dimethylfumarat und entwickelte ein entsprechendes Medikament.

Zur Zulassung als Medikament musste die Wirksamkeit am Menschen nachgewiesen werden. Dabei ist ein Vorgehen in drei Phasen üblich: In Phase I wird an einer kleinen Gruppe gesunder Menschen getestet, ob es unerwartete

Nebenwirkungen gibt. In Phase II wird die Wirksamkeit des Medikaments an einer kleinen Gruppe Erkrankter überprüft, wobei auch ermittelt wird, was die beste Dosierung ist. Abschließend erfolgt in Phase III ein Test unter realistischen Bedingungen an Hunderten von Menschen.

Da Dimethylfumarat bereits seit Längerem zur Behandlung von Schuppenflechte eingesetzt wurde, waren die Nebenwirkungen am Menschen bereits bekannt. Hier traten überwiegend Magen-Darm-Beschwerden auf sowie in seltenen Fällen eine starke Absenkung der weißen Blutkörperchen. Sofern diese zu stark ausgeprägt war, ist in Einzelfällen eine sogenannte PML (progressive multifokale Leukenzephalopathie) aufgetreten, bei der sich ein im Prinzip harmloses Virus im Gehirn ausbreitet, was zum Tod führen kann. Das Risiko dafür kann man aber durch regelmäßige Blutuntersuchungen stark absenken. Insgesamt war daher die Überprüfung der Verträglichkeit des neuen Medikamentes zur Behandlung der Multiplen Sklerose in Phase I einfach.

Die Überprüfung der Wirksamkeit eines Medikaments in den Phasen II und III erfolgt im Rahmen einer *Studie*. Die Grundidee dabei ist der *Vergleich*: Man vergleicht eine sogenannte *Studiengruppe*, die mit dem Medikament behandelt wurde, mit einer sogenannten *Kontrollgruppe*, die nicht mit dem Medikament behandelt wurde. Man spricht daher von einer *kontrollierten Studie*. Um von Unterschieden zwischen Studien- und Kontrollgruppe (z. B. hinsichtlich des Verlaufs der Erkrankung) auf die Wirksamkeit des Medikaments schließen zu können, muss dabei (abgesehen von der Behandlung mit dem Medikament) die Kontrollgruppe der Studiengruppe möglichst ähnlich sein.

Für die Wahl von Studien- und Kontrollgruppe gibt es verschiedene Möglichkeiten. Bei einer *retrospektiv kontrollierten Studie* werden Eigenschaften der Studiengruppe mit in der Vergangenheit gesammelten Daten verglichen.

Im obigen Beispiel bedeutet dies, dass man als Studiengruppe eine größere Anzahl von Personen auswählt, die an Multipler Sklerose erkrankt sind, und all diese (bzw. nur diejenigen, die mit der Behandlung einverstanden sind) mit dem neuen Medikament behandelt. Dann wartet man einige Zeit ab und bestimmt den durchschnittlichen Verlauf der Krankheit bei den behandelten Patienten (z. B. im Hinblick auf das Auftreten von Schüben oder von neuen Entzündungsherden im Gehirn). Diesen vergleicht man mit dem durchschnittlichen Verlauf der Krankheit von in der Vergangenheit an Multipler Sklerose erkrankten Personen. Da bei diesem Vergleich der durchschnittliche Verlauf der Krankheit zugrunde gelegt wird, können eventuelle Unterschiede in der Größe der beiden Gruppen vernachlässigt werden.

Problematisch an diesem Vorgehen ist, dass sich die Diagnose der Multiplen Sklerose in den letzten Jahrzehnten stark verändert hat. Durch den weit verbreiteten Einsatz des Kernspintomografen wird heute Multiple Sklerose zum Teil auch schon diagnostiziert, bevor sich Symptome überhaupt massiv gezeigt haben, sodass die Schwere der diagnostizierten Fälle im Schnitt deutlich abgenommen hat. Stellt man also fest, dass der durchschnittliche Verlauf der Krankheit bei den mit dem neuen Medikament behandelten Personen geringer ist als bei den in der Vergangenheit traditionell behandelten Personen, so weiß man nicht, ob das an dem neuen Medikament liegt, oder ob der Grund dafür darin liegt, dass die Personen in der

Studiengruppe besonders häufig an einer eher harmlosen Variante der Multiplen Sklerose erkrankt waren. Die Schwere des Verlaufs der Multiplen Sklerose ist daher ein sogenannter *konfundierender Faktor*, d. h., es handelt sich um eine Einflussgröße, deren Einfluss auf den durchschnittlichen Verlauf der Krankheit sich mit dem Einfluss der Art der Behandlung vermengt.

Möchte man diesen konfundierenden Faktor ausschließen, so sollte man statt einer retrospektiv kontrollierten Studie eine *prospektiv kontrollierte Studie* durchführen, bei der Studien- und Kontrollgruppe parallel auf gewisse Merkmale hin untersucht werden. Je nachdem, ob man die Testpersonen dabei *deterministisch* oder *mittels eines Zufallsexperiments* in Studien- und Kontrollgruppe unterteilt, spricht man von *prospektiv kontrollierten Studien ohne* oder *mit Randomisierung*.

Im vorliegenden Beispiel könnte man eine prospektiv kontrollierte Studie ohne Randomisierung so durchführen, dass man zuerst eine größere Anzahl von an Multipler Sklerose erkrankten Personen auswählt und dann alle diejenigen, die der Behandlung zustimmen, mit dem neuen Medikament behandelt. Diese Personen würden die Studiengruppe bilden, der Rest der ausgewählten Personen wäre die Kontrollgruppe. Nach einiger Zeit würde man den durchschnittlichen Verlauf der Krankheit in beiden Gruppen vergleichen.

Bei diesem Vorgehen entscheiden die Erkrankten, ob sie zur Studiengruppe oder zur Kontrollgruppe gehören. Das führt dazu, dass sich die Kontrollgruppe nicht nur durch die Behandlung von der Studiengruppe unterscheidet. Zum Beispiel ist es denkbar, dass besonders viele Erkrankte mit einer massiven Verlaufsform der neuen Behandlung zustimmen. Daher tritt wieder das Problem auf, dass hier der Einfluss der Behandlung *konfundiert* (sich vermengt) mit dem Einfluss der Verlaufsform der Krankheit. Insofern kann man nicht sagen, inwieweit ein möglicher Unterschied bei den durchschnittlichen Verläufen der Krankheit auf die Behandlung zurückzuführen ist (bzw. ein eventuell nicht vorhandener Unterschied nur aufgrund der Unterschiede bei der Schwere der Erkrankung auftritt).

Als möglicher Ausweg bietet sich an, als Kontrollgruppe nur einen Teil der Erkrankten auszuwählen, die der Behandlung mit dem neuen Medikament nicht zustimmen, und diesen Teil so zu bestimmen, dass er z. B. hinsichtlich der Anzahl der in der Vergangenheit aufgetretenen Schübe oder auch des Kernspintomografie-Befundes möglichst ähnlich zur Studiengruppe ist. Dies ist aber sehr fehleranfällig, da man dazu sämtliche Faktoren kennen muss, die Einfluss auf den Verlauf der Krankheit haben. Denkbar ist hier z. B., dass eine fettarme Ernährung Einfluss auf den chronischen Entzündungsprozess bei Multipler Sklerose hat und damit ein solcher Faktor ist.

Das Problem konfundierender Faktoren wird bei einer *prospektiv kontrollierten Studie mit Randomisierung* vermieden. Denn dabei werden nur solche Testpersonen betrachtet, die sowohl für die Studien- als auch für die Kontrollgruppe infrage kommen. Diese werden dann zufällig (z. B. durch Münzwürfe) in Studien- und Kontrollgruppe unterteilt.

Im obigen Beispiel heißt das, dass nur die Erkrankten betrachtet werden, die der Behandlung zustimmen. Diese werden zufällig in Studien- und Kontrollgruppe aufgeteilt. Anschließend werden die Personen in der Studiengruppe mit dem

neuen Medikament behandelt, die in der Kontrollgruppe werden traditionell behandelt, und nach einiger Zeit werden die durchschnittlichen Verläufe der Krankheit verglichen.

Im Rahmen zweier prospektiv kontrollierter Studien mit Randomisierung (DEFINE und CONFIRM) wurde die Wirkung von Dimethylfumarat bei Multipler Sklerose in den Jahren 2010 bis 2012 untersucht. Während die Personen in der Studiengruppe das neue Medikament erhielten, wurde den Personen in der Kontrollgruppe anstelle des Medikaments eine gleich aussehende Kapsel ohne Wirkstoff, ein sogenanntes *Placebo*, verabreicht. Damit sollte verhindert werden, dass es den Personen in der Studiengruppe schon allein deshalb besser ging als denen in der Kontrollgruppe, weil sie eine Tablette eingenommen hatten. Die Besserung von Symptomen durch Einnahme einer Tablette, die nichts mit Wirkstoffen in der Tablette zu tun hat, wird auch als *Placebo-Effekt* bezeichnet. Da es sich um eine *blinde Studie* handelte, wusste keiner der Studienteilnehmer, ob er das Medikament oder ein Placebo bekam. Darüber hinaus wurde auch den behandelnden Ärzten nicht mitgeteilt, ob ein Patient zur Studien- oder zur Kontrollgruppe gehörte, die Studie wurde also als *doppelblinde Studie* durchgeführt. Dies sollte sicherstellen, dass das Wissen des Arztes über die Art der verordneten Tablette (Wirkstoff oder Placebo) keinen Einfluss auf seine Beurteilung der Symptome hatte.

Anfang 2013 waren beide Studien abgeschlossen. Insgesamt wurden 771 Versuchspersonen rekrutiert. Die Auswertung ergab, dass die Einnahme von Fumarsäure die jährliche Schubrate von ca. 0,4 auf ca. 0,2 abgesenkt hat. Auch im Kernspintomografen ist die Anzahl der neu aufgetretenen Entzündungsherde im Gehirn (sogenannte Läsionen) um mehr als 2/3 gesunken.

Aufgrund dieses Ergebnisses ist Fumarsäure im Frühjahr 2014 in Deutschland zur Behandlung von Multipler Sklerose zugelassen worden. Im ersten Jahr durfte das Pharmaunternehmen den Preis dafür noch relativ frei festlegen. Danach wurde dieser aber in Verhandlungen mit einer Kommission der Krankenkassen unter Berücksichtigung des Zusatznutzens des Medikaments (im Vergleich zur bisherigen Therapie) neu festgelegt. Dabei wurde der Preis dann von den anfänglichen 28.300 Euro pro Jahr auf ca. 16.300 Euro pro Jahr abgesenkt.

2.2 Beobachtungsstudien

Beobachtungsstudien zeichnen sich gegenüber kontrollierten Studien dadurch aus, dass die Studienteilnehmer nur beobachtet und gegebenenfalls befragt werden und während der Studie keinerlei Intervention ausgesetzt sind. Dies hat insbesondere zur Konsequenz, dass bei Beobachtungsstudien die Aufteilung in Studien- und Kontrollgruppe immer anhand bestimmter Merkmale der beobachteten Personen oder Objekte durchgeführt wird. Bei diesen Merkmalen kann es sich um bestimmte Eigenschaften, Verhaltensweisen oder andere Charakteristika handeln, durch die sich eine Gruppe von Personen von einer anderen Gruppe unterscheidet. Beispiele für solche Unterteilungen wären etwa die Unterteilung einer Gruppe nach

Geschlecht (Männer und Frauen) oder nach Ernährungsform im Säuglingsalter (gestillt oder nicht gestillt).

Damit ergibt sich – ähnlich wie bei retrospektiv kontrollierten Studien oder prospektiv kontrollierten Studien ohne Randomisierung – die Einteilung in Studien- und Kontrollgruppe aus Unterschieden in Bezug auf bestimmte Merkmale: Die Teilnehmer der Studie bestimmen quasi selbst, ob sie zur Studien- oder Kontrollgruppe gehören. Diese Eigenschaft von Beobachtungsstudien führt wieder zu einer Reihe von Problemen, die sich aus der nicht-zufälligen Einteilung der Studienteilnehmer in Studien- und Kontrollgruppe ergeben.

Eine typische Fragestellung, die mithilfe von Beobachtungsstudien untersucht worden ist, lautet: „Verursacht Rauchen Krankheiten mit Todesfolge?“. Hierzu wählt der Statistiker eine Gruppe von Rauchern und eine Gruppe von Nichtrauchern aus und vergleicht die Todesraten beider Gruppen über einen längeren Zeitraum. Er kann jedoch keinen Einfluss darauf nehmen, wie die Studienteilnehmer in Studien- und Kontrollgruppe aufgeteilt werden, denn er wird kaum Teilnehmer finden, die bereit sind, je nach seiner Anweisung die nächsten 10 Jahre intensiv bzw. gar nicht zu rauchen.

Nun unterscheidet sich aber die Studiengruppe (bestehend aus allen Rauchern) nicht nur hinsichtlich des Rauchens von der Kontrollgruppe (bestehend aus allen Nichtrauchern). Da besonders viele Männer rauchen, sind nämlich unter anderem Männer überproportional häufig in der Studiengruppe vertreten.² Die Todesrate bei Männern ist wegen des häufigeren Auftretens von Herzerkrankungen höher als die von Frauen. Damit ist das Geschlecht ein *kongfundierender Faktor*, d. h. eine Einflussgröße, deren Einfluss auf die Todesrate sich mit dem des Rauchens vermengt. Stellt sich dann heraus, dass die Todesrate in der Studiengruppe deutlich höher ist als in der Kontrollgruppe, so weiß man nicht, ob dies am Rauchen oder an dem kongfundierenden Faktor liegt.

Das Hauptproblem bei Beobachtungsstudien besteht somit darin, dass es schwierig und häufig nahezu unmöglich ist, zu beurteilen, ob die Studiengruppe und die Kontrollgruppe wirklich ähnlich sind.

Wie bei prospektiv kontrollierten Studien ohne Randomisierung kann man auch hier versuchen, das Problem zu lösen, indem man nur Gruppen vergleicht, die bzgl. dieses kongfundierenden Faktors übereinstimmen. Dazu würde man im obigen Beispiel die Todesrate von männlichen Rauchern mit der von männlichen Nichtrauchern und die von weiblichen Rauchern mit der von weiblichen Nichtrauchern vergleichen. Dies löst das Problem aber nicht vollständig, da es weitere kongfundierende Faktoren gibt, wie z. B. Alter (ältere Menschen unterscheiden sich sowohl in ihren Rauchgewohnheiten als auch hinsichtlich des Risikos, an Lungenkrebs zu erkranken, von jüngeren Menschen).² Nötig ist daher die Erkennung aller kongfundierenden Faktoren und die Bildung von vielen Untergruppen.

Dies wird aber oft nicht richtig durchgeführt bzw. kann manchmal gar nicht richtig durchgeführt werden, wie im Folgenden an zwei Beispielen erläutert wird.

In unserem ersten Beispiel geht es um die Frage, ob gestillte Kinder intelligenter sind als Kinder, die nicht gestillt wurden. Seit 1929 kam eine Vielzahl von Beobachtungsstudien zu dem Schluss, dass Stillen die Intelligenz erhöht. Bei diesen Studien

traten natürlich konfundierende Faktoren auf, wie z.B. Ausbildung, Alter und Rauchgewohnheiten der Mutter oder Geburtsgewicht und Vorhandensein von Geschwistern des Kindes. Die meisten der Studien kamen jedoch selbst bei Kontrolle dieser konfundierenden Faktoren zu dem Resultat, dass Stillen des Kindes seine Intelligenz erhöht. Erst im Rahmen einer umfangreichen Auswertung von Daten, die im Rahmen einer Langzeitbeobachtung ab 1979 am Center for Human Resource Research in den USA durchgeführt wurde, hat sich dann herausgestellt, dass dieser positive Effekt auf die Intelligenz des Kindes nicht mehr vorhanden war, sobald man auch die Intelligenz der Mutter als konfundierenden Faktor berücksichtigte.³ Im Nachhinein ist dies intuitiv klar, da die Intelligenz der Mutter vererbt wird und gleichzeitig Mütter mit einer höheren Intelligenz auch häufiger gestillt haben. Jedoch ist es leider meistens schwer, alle relevanten konfundierenden Faktoren zu erkennen.

Unser zweites Beispiel beschäftigt sich mit der Frage, ob sich die regelmäßige Einnahme von Vitamin E positiv auf das Auftreten von Gefäßerkrankungen am Herzen (wie z.B. Verstopfung der Koronararterie, was zu Herzinfarkten führen kann) auswirkt. Ein solcher Zusammenhang wurde z. B. in der Nurses Health Study (Studie zur Untersuchung des Gesundheitszustandes von Krankenschwestern)⁴ festgestellt. In dieser wurden ab dem Jahr 1980 mehr als 87.000 Krankenschwestern in den USA zu ihrer Ernährung befragt und anschließend über 8 Jahre hinweg hinsichtlich ihres Gesundheitszustands beobachtet. Dabei traten in 522 Fällen Gefäßerkrankungen am Herzen auf, wobei in der Gruppe der Krankenschwestern, die viel Vitamin E zu sich nahmen, 34 % weniger von diesen Erkrankungen auftraten als in der Gruppe der Krankenschwestern, die relativ wenig Vitamin E zu sich nahmen. Dieser Effekt trat auch dann noch auf, als mögliche konfundierende Faktoren, wie z. B. Alter, Alkoholkonsum, sportliche Betätigung, Einnahme von Hormonen in den Wechseljahren etc. kontrolliert wurden.

Diese Studie wurde zum Anlass genommen, Diätempfehlungen für Hochrisikopatienten bzgl. Herzerkrankungen aufzustellen. Um die Wirkung dieser Empfehlungen zu überprüfen, wurde in den Jahren 1994 bis 2001 in Großbritannien eine kontrollierte Studie mit Randomisierung durchgeführt. Dabei wurden 20.536 Erwachsene im Alter zwischen 40 und 80 Jahren mit bereits vorgeschädigten Gefäßen am Herzen oder mit Diabetes (was oft zu solchen Erkrankungen führt) ausgewählt und zufällig in Studien- und Kontrollgruppe unterteilt. Anschließend wurde den Personen in der Studiengruppe täglich eine Tablette mit 600 mg Vitamin E, 250 mg Vitamin C und 20 mg Beta-Karotin als Nahrungsergänzung verordnet, während die Kontrollgruppe nur ein Placebo bekam. Diese Behandlung erfolgte über 5 Jahre, anschließend wurden die relativen Häufigkeiten einzelner Krankheits- bzw. Todesfälle bestimmt. Tabelle 2.1 fasst die Resultate zusammen.⁵

Betrachtet man die Zahlen in der Tabelle, so sieht man, dass in der Studiengruppe prozentual sogar etwas mehr Todesfälle im Zusammenhang mit Gefäßerkrankungen und etwas mehr Herzinfarkte auftraten, als in der Kontrollgruppe. Insgesamt sind alle Zahlen in den einzelnen Zeilen bei Studien- und Kontrollgruppe aber recht ähnlich, sodass kein systematischer Unterschied erkennbar ist. Aufgrund dieser Daten kam man zu der Auffassung, dass die zusätzliche Einnahme von

Tab. 2.1 Studie zur Wirkung von Vitaminen bei Herzerkrankungen

	Gesamt	Studiengruppe	Kontrollgruppe
Alle	20.536 (100 %)	10.268 (50 %)	10.268 (50 %)
Todesfälle	2835 (13,8 %)	1446 (14,1 %)	1389 (13,5 %)
Todesfälle in Zusammenhang mit Gefäßerkrankungen	1718 (8,4 %)	878 (8,6 %)	840 (8,2 %)
Herzinfarkt	2110 (10,3 %)	1063 (10,4 %)	1047 (10,2 %)
Schlaganfall	1029 (27,5 %)	511 (5,0 %)	518 (5,0 %)
Erstauftritt schwere Herzerkrankung	4618 (22,5 %)	2306 (22,5 %)	2312 (22,5 %)

Vitaminen bei Hochrisikopatienten für Gefäßerkrankungen des Herzens vermutlich keinen positiven Effekt hat.

Selbstverständlich führen Beobachtungsstudien aber nicht nur zu falschen Vermutungen. Zum Beispiel wurde in einigen Beobachtungsstudien, in denen die Teilnehmer zu ihren Ernährungsgewohnheiten befragt und dann bzgl. ihres Gesundheitszustandes über einen längeren Zeitraum beobachtet wurden, der positive Effekt einer mediterranen Ernährung auf Herz-Kreislauf-Krankheiten festgestellt. Dieser positive Effekt konnte dann auch im Rahmen einer in Indien durchgeführten kontrollierten Studie mit Randomisierung nachgewiesen werden.⁶ Dabei wurden 1000 Hochrisikopatienten für Herz-Kreislauf-Krankheiten zufällig in zwei Gruppen unterteilt. Der einen Gruppe wurde eine spezielle, mediterrane Diät verordnet, während die andere Gruppe die sonst üblichen Diätempfehlungen erhielt. Nach zwei Jahren wurden beide Gruppen hinsichtlich neu aufgetretener Herz-Kreislauf-Krankheitsfälle verglichen. Wie die in Tab. 2.2 aufgeführten Resultate zeigen, traten bei den Teilnehmern mit mediterraner Diät, die die Studiengruppe bildeten, deutlich weniger Herz-Kreislauf-Krankheitsfälle auf, was die positive Wirkung dieser Diät bei Hochrisikopatienten belegt.

Eine Übersicht über die verschiedenen Arten von Studien findet man in Abb. 2.1.

Gemeinsam ist den verschiedenen Arten von Studien, dass zunächst einmal nur das gleichzeitige Auftreten (sogenannte *Assoziation*) zweier Dinge nachgewiesen wird. Was man aber normalerweise gern nachweisen möchte, ist ein kausaler Zusammenhang dieser Dinge. Dies ist allerdings bei Beobachtungsstudien, retrospektiv kontrollierten Studien und prospektiv kontrollierten Studien ohne Randomisierung nicht möglich, da der Grund für das gleichzeitige Auftreten zweier

Tab. 2.2 Studie zur Wirkung von mediterraner Diät bei Herz-Kreislauf-Erkrankungen

	Gesamt	Studiengruppe	Kontrollgruppe
Alle	1000 (100 %)	499 (49,9 %)	501 (50,1 %)
Nicht tödlich verlaufende Myokardinfarkte	64 (6,4 %)	21 (4,2 %)	43 (8,6 %)
Tödlich verlaufende Myokardinfarkte	29 (2,9 %)	12 (2,4 %)	17 (3,4 %)
plötzlicher Herztod	22 (2,2 %)	6 (1,2 %)	16 (3,2 %)

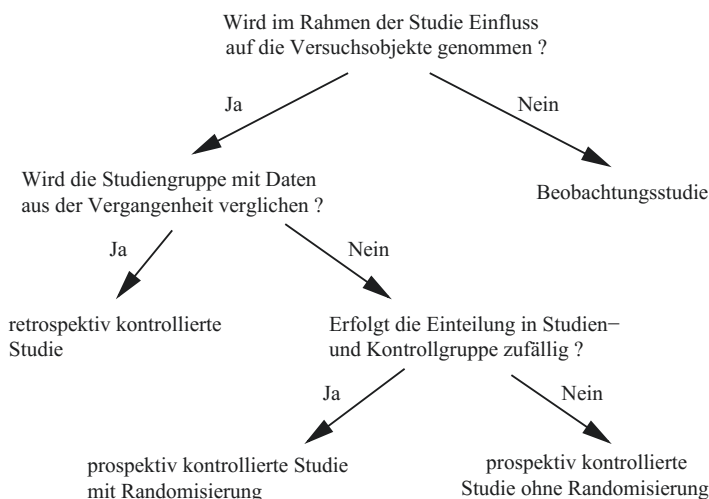


Abb. 2.1 Übersicht über die verschiedenen Arten von Studien

Dinge auch der Einfluss konfundierender Faktoren sein kann. Konfundierende Faktoren beeinflussen sowohl die Aufteilung in Studien- und Kontrollgruppe als auch das beobachtete Resultat. Daher erlauben sowohl Beobachtungsstudien als auch retrospektiv kontrollierten Studien bzw. prospektiv kontrollierte Studien ohne Randomisierung zunächst einmal nur das Aufstellen einer Hypothese über einen kausalen Zusammenhang. Der Nachweis eines solchen kausalen Zusammenhangs ist jedoch nur mithilfe einer prospektiv kontrollierten Studie mit Randomisierung möglich.

2.3 Probleme bei der Durchführung von Studien

Wir haben in den vorigen Abschnitten gesehen, dass die Datenerhebung optimalerweise im Rahmen einer prospektiv kontrollierten Studie mit Randomisierung erfolgen sollte. Denn nur diese Art von Studie erlaubt den zweifelsfreien Nachweis kausaler Zusammenhänge. Allerdings gibt es häufig ethische Gründe, die die Durchführung einer solchen Studie (zumindest am Menschen) verbieten.

Ein Beispiel ist die Fragestellung, ob die von Handys benutzten hochfrequenten elektromagnetischen Felder des Mobilfunks beim Menschen Gehirntumore auslösen. Hierzu wurde im Jahr 2003 in einer indischen Beobachtungsstudie festgestellt, dass diese Art von Strahlen Genschäden an menschlichen Blutzellen verursacht.⁷ Dabei wurden eine Studien- und eine Kontrollgruppe gebildet, bestehend aus je 24 indischen Teilnehmern, die ein Handy regelmäßig benutzten bzw. kein Handy verwendeten (und solche gab es zur Zeit der Entstehung der Studie in

Indien auch noch), und Blutproben der Teilnehmer wurden auf Genschäden, wie z. B. Chromosomenaberrationen und Schwesterchromatidaustausch, hin untersucht. Dabei stellte sich heraus, dass in der Studiengruppe deutlich häufiger Genschäden auftraten als in der Kontrollgruppe. Wie bei jeder Beobachtungsstudie wurde natürlich auch hierbei versucht, potenziell konfundierende Faktoren zu kontrollieren. Dabei wurden die Testpersonen gemäß Alter, Geschlecht, Rauchgewohnheiten und Alkoholkonsum unterteilt, und auch bei Berücksichtigung dieser potenziell konfundierenden Faktoren (z. B. durch Bildung von homogenen Untergruppen) trat das obige Resultat auf. Somit stützte diese Beobachtungsstudie die Hypothese, dass hochfrequente elektromagnetische Felder Chromosomenschäden verursachen.

Der Verdacht, dass hochfrequente Felder gesundheitliche Auswirkungen auf den Menschen haben, war für das Bundesamt für Strahlenschutz der Anlass, eine Reihe von kontrollierten Studien mit Randomisierung (im Rahmen des Deutschen Mobilfunk-Forschungsprogramms) zur Überprüfung dieser Hypothese in Auftrag zu geben.

Im Prinzip könnte man eine solche kontrollierte Studie mit Randomisierung wie folgt durchführen: Man wählt eine Reihe von Testpersonen aus, die sich gegen Bezahlung zur Teilnahme an der Studie bereit erklären, teilt diese zufällig in Studien- und Kontrollgruppe auf und verteilt an beide Gruppen identisch aussehende Mobiltelefone, wobei die Geräte in der Studiengruppe eine deutlich stärkere Magnetstrahlung verwenden als die Geräte in der Kontrollgruppe. Nach längerer Zeit untersucht man dann die Probanden und stellt die Anzahl der auftretenden Genschäden (und bei hinreichend langer Laufzeit der Studie auch die Zahl der auftretenden Krebserkrankungen) in beiden Gruppen fest. Aufgrund des Resultats dieser Studie könnte man nun leicht (bei ausreichender Anzahl der Teilnehmer) auf die kausale Wirkung der Magnetstrahlen zurückschließen. Selbstverständlich verbietet sich dieses Vorgehen aber aus ethischen Gründen.

Stattdessen wurden die kontrollierten Studien mit Randomisierung auf Zellebene durchgeführt. In einer dieser Studien⁸ sollte festgestellt werden, ob hochfrequente elektromagnetische Felder des Mobilfunks menschliche Blutzellen schädigen. Dazu hat man Blutproben von insgesamt 40 Spendern verwendet. Dabei handelte es sich um jeweils 10 männliche und 10 weibliche Spender der Altersgruppen 18–22 Jahre sowie 50–60 Jahre. Pro Proband wurden mehrere identische Zellkulturen gebildet, die jeweils hochfrequenten Feldern in drei verschiedenen Stärken und für zwei unterschiedliche Zeiträume ausgesetzt wurden; zur Kontrolle wurde eine Zellkultur pro Proband keiner Strahlung ausgesetzt.

Die anschließende Untersuchung der Zellkulturen ergab, dass signifikante Veränderungen der Blutzellen nur bei den beiden stärksten hochfrequenten Feldern auftraten; bei dem hochfrequenten Feld mit der geringsten Stärke (dessen Stärke noch über dem geltenden Grenzwert lag) wurden hingegen keine signifikanten Veränderungen gefunden. Das Fazit der Studie war somit, dass kein Handlungsbedarf besteht, was die Senkung der derzeit geltenden Grenzwerte für die Stärke hochfrequenter Felder angeht.

Auch in der Medizin gibt es viele Fragestellungen, die nur schwer mit einer doppelblinden, Placebo-kontrollierten Studie untersucht werden können. Dazu

gehört die Untersuchung von operativen Therapien. Hier würde dem Placebo eine Schein-Operation entsprechen, der sich die Teilnehmer der Kontrollgruppe unterziehen müssten; eine blinde Studie ist aus ethischer Sicht problematisch und eine doppelblinde Studie ist grundsätzlich nicht durchführbar, weil der behandelnde Arzt zwangsläufig erfährt, ob eine Schein-Operation gemacht wurde oder nicht. Wenn eine Studie aber nicht blind durchgeführt wird, kann es Probleme mit der Randomisierung geben, wie das folgende Beispiel zeigt.

In den USA wurde 2005 eine große prospektiv kontrollierte Studie mit Randomisierung durchgeführt, in der eine operative Therapie mit einer konservativen Behandlung bei einem lumbalen Bandscheibenvorfall verglichen wurde.⁹ Es wurden 500 Patienten mit Bandscheibenvorfällen ausgewählt, welche als Symptome radikuläre Ausfälle und Schmerzen über mindestens sechs Wochen hatten. Für die operative Therapie wurden 245 Patienten zufällig ausgewählt, für die konservative Behandlung 256 Patienten. Allerdings entschied sich während der Studiendauer mehr als ein Drittel der Patienten für die jeweils andere Therapieform: 37,5 % der Patienten in der operativen Gruppe wollten lieber konservativ behandelt werden und 41,8 % der Patienten in der Gruppe mit konservativer Behandlung entschieden sich für eine Operation. Der Statistiker konnte somit die Randomisierung nicht bei allen Studienteilnehmern durchsetzen. Dies führt zu erheblichen Problemen bei der Auswertung der Studie: Lässt man die selbst die Gruppen wechselnden Patienten in der Auswertung der Studie einfach weg, so hat sich die Struktur beider Gruppen verändert, sodass man nicht mehr sicher sein kann, dass Studien- und Kontrollgruppe in der Tat gleich aufgebaut sind. Und will man diese Patienten doch noch berücksichtigen, so könnte man z. B. deren Auswertung im Hinblick auf einen potenziellen Nachweis des gewünschten Effektes so abändern, dass die Resultate dieser Patienten gegen den Effekt sprechen, um dann zu sehen, ob in diesem Fall der Effekt nachgewiesen werden kann. In der betrachteten Studie wird aber bei dem großen Prozentsatz von wechselnden Patienten der gewünschte Effekt nicht mehr nachweisbar sein. Das gleiche Problem tritt natürlich auch immer dann auf, wenn Patienten während der Laufzeit einer Studie plötzlich die weitere Teilnahme an der Studie verweigern.

Bei Studien in der Medizin ist auch der Test auf unerwartete Nebenwirkungen eines neuen Medikaments am Menschen kritisch. Zum Beispiel wurde im Jahr 2006 ein neues gentechnisch hergestelltes Medikament an 6 gesunden Probanden ausprobiert. Überraschenderweise traten aber bei allen 6 Probanden sehr schwere Nebenwirkungen auf, die zum Teil zu Amputationen von Gliedmaßen führten.¹⁰ Als Konsequenz wird heute empfohlen, neue Medikamente erst nach und nach einer Gruppe von Probanden zu verabreichen, sodass man den Test beim ersten Anzeichen von starken Nebenwirkungen noch abbrechen kann.

Ein weiteres Problem beim Durchführen einer Studie ist, dass die Art und Weise, wie die Daten beobachtet werden, eine Veränderung der Daten bewirken kann. Zum Beispiel wurde in den 1990er-Jahren in einer Reihe von Studien nachgewiesen, dass die Anzahl der Pinguine am Südpol abnimmt. Dabei wurden über 40.000 Pinguine mit Markierungsbändern gekennzeichnet, und über mehrere Sommer hinweg wurde die Anzahl der noch lebenden markierten Pinguine bestimmt. Der hierbei

beobachtete Rückgang der Population der Pinguine wurde als weiteres Indiz für die Erderwärmung angesehen, die zum Schmelzen des Eises und damit zur Vernichtung des Lebensraumes der Pinguine führt. Schwierig war dabei das Anbringen des Markierungsbandes. Weil Pinguine sich aufgrund ihrer kurzen Beine nicht wie andere Vögel beringen lassen, mussten die Bänder an einem der zur Flosse mutierten Flügel befestigt werden. Im Rahmen einer im Jahr 2007 abgeschlossenen kontrollierten Studie mit Randomisierung, in der je 50 Pinguine mit einem Markierungsband und einem unter die Beinhaut implantierten Chip bzw. nur mit einem Chip versehen und sodann über 4 Brutperioden beobachtet wurden, konnte dann aber nachgewiesen werden, dass gerade durch das Anbringen der Markierungsbänder die Überlebensrate der Pinguine sinkt, was vermutlich auf Behinderung des Schwimmens durch das Markierungsband zurückzuführen war.¹¹

2.4 Umfragen

Bei einer Umfrage betrachtet man eine Menge von Objekten (*Grundgesamtheit*), wobei jedes der Objekte eine Reihe von Eigenschaften besitzt. Feststellen möchte man, wie viele Objekte der Grundgesamtheit eine gewisse vorgegebene Eigenschaft haben.

Ein Beispiel dafür ist die sogenannte Sonntagsfrage, über die regelmäßig in den Medien berichtet wird. Dabei möchte man wissen, wie viele der Wahlberechtigten in der BRD für die aktuelle Bundesregierung stimmen würden, wenn nächsten Sonntag Bundestagswahl wäre.

Tabelle 2.3 beinhaltet die Ergebnisse von Wahlumfragen, die von sechs verschiedenen Meinungsforschungsinstituten ca. eine Woche vor der Bundestagswahl 2013 durchgeführt wurden, sowie das amtliche Endergebnis der Bundestagswahl am 22.09.2013. Wie man sieht, weichen die Umfrageergebnisse zum Teil erheblich vom tatsächlichen Wahlergebnis ab. Daraus kann man allerdings nicht auf Fehler bei den Umfragen schließen, da sich das Wahlverhalten der Deutschen in den Tagen vor der Wahl noch einmal geändert haben könnte. Allerdings sieht man an den

Tab. 2.3 Umfragen zur Bundestagswahl 2013¹²

	SPD	CDU/CSU	FDP	GRÜNE	LINKE	AfD
Allensbach	27,0	39,5	5,5	9,0	9,0	4,5
TNS Emnid	26,0	39,0	6,0	9,0	9,0	4,0
Forsa	26,0	40,0	5,0	10,0	9,0	4,0
Forschungsgruppe Wahlen	27,0	40,0	5,5	9,0	8,5	4,0
Infratest-dimap	28,0	40,0	5,0	10,0	8,0	2,5
INSA	28,0	38,0	6,0	8,0	9,0	5,0
amtliches Endergebnis	25,7	41,5	4,8	8,4	8,6	4,7

Schwankungen der Umfrageergebnisse der verschiedenen Institute, dass zumindest bei einigen davon doch erhebliche Ungenauigkeiten bei der Vorhersage auftraten.

Wie man Umfragen durchführen kann und warum genaue Prognosen häufig schwierig sind, wird im Folgenden behandelt.

Die Bestimmung der Anzahl der Objekte einer Grundgesamtheit mit einer gewissen vorgegebenen Eigenschaft ist zunächst einmal eine rein deterministische Fragestellung, die man im Prinzip durch reines Abzählen entscheiden könnte. Bei vielen Fragestellungen (insbesondere bei der oben erwähnten Sonntagsfrage) ist die Betrachtung aller Objekte der Grundgesamtheit aber nicht möglich bzw. viel zu aufwendig.

Als Ausweg bietet sich an, nur für eine „kleine“ Teilmenge (der Statistiker spricht hier von einer *Stichprobe*) der Grundgesamtheit zu ermitteln, wie viele Objekte darin die interessierende Eigenschaft haben, und dann zu versuchen, mithilfe dieses Resultats die gesuchte Größe näherungsweise zu bestimmen (der Statistiker spricht hier von *schätzen*). Dazu muss man erstens festlegen, wie man die Stichprobe wählt, und zweitens ein Verfahren entwickeln, das mithilfe der Stichprobe die gesuchte Größe schätzt.

Für die oben angesprochene Sonntagsfrage könnte man dazu wie folgt vorgehen: Zuerst wählt man „rein zufällig“ n Personen (z. B. $n = 2000$) aus der Menge aller Wahlberechtigten aus und befragt diese bzgl. ihres Wahlverhaltens. Anschließend schätzt man den prozentualen Anteil der Stimmen für die aktuelle Bundesregierung in der Menge aller Wahlberechtigten durch den entsprechenden prozentualen Anteil in der Stichprobe. Wie wir in den weiteren Kapiteln dieses Buches sehen werden, liefert dies zumindest dann eine gute Schätzung, wenn die Stichprobe wirklich „rein zufällig“ ausgewählt wurde. Damit steht man aber noch vor dem Problem, wie man Letzteres durchführt. Dazu werden im Weiteren die folgenden fünf Vorgehensweisen betrachtet:

Vorgehen 1: Befrage die Studenten einer Statistik-Vorlesung.

Vorgehen 2: Befrage die ersten n Personen, die montagsmorgens ab 10 Uhr einen festen Punkt der Fußgängerzone in Darmstadt passieren.

Vorgehen 3: Erstelle eine Liste aller Wahlberechtigten (mit Adresse). Wähle aus dieser „zufällig“ n Personen aus und befrage diese.

Vorgehen 4: Wähle aus einem Telefonbuch für Deutschland rein zufällig Nummern aus und befrage die ersten n Personen, die man erreicht.

Vorgehen 5: Wähle zufällig Nummern am Telefon und befrage die ersten n Privatpersonen, die sich melden.

Betrachtet man diese Vorgehensweisen bzgl. der praktischen Durchführbarkeit, so stellt sich Vorgehen 3 als sehr aufwendig heraus: Die zu befragenden Personen sind dabei im Allgemeinen nämlich über die gesamte BRD verstreut, zudem werden die Adressen nicht immer aktuell sein. Darüber hinaus gibt es Länder (wie z. B. die USA), wo Listen aller Wahlberechtigten gar nicht erst existieren.

Bei allen anderen Vorgehensweisen tritt eine sogenannte *Verzerrung durch Auswahl* (*sampling bias*) auf. Diese beruht darauf, dass die Stichprobe nicht *repräsentativ* ist, d. h., dass bestimmte Gruppen der Wahlberechtigten, deren Wahlverhalten vom Durchschnitt abweicht, überrepräsentiert sind. Zum Beispiel sind dies bei Vorgehen 1 die Studenten, bei Vorgehen 2 die Einwohner von Darmstadt sowie Personen, die dem Interviewer sympathisch sind, bei Vorgehen 4 Personen mit Eintrag im Telefonbuch und bei Vorgehen 5 Personen, die telefonisch leicht erreichbar sind, sowie Personen, die in einem kleinen Haushalt leben. Bei Vorgehen 5 lässt sich dieses Problem teilweise umgehen, indem man dort bei einzelnen Nummern mehrmals anruft, sofern man nicht sofort jemanden erreicht, und indem man die Person, die man unter dieser Nummer befragt, nach demografischen Aspekten auswählt (wie z. B. „befrage jüngsten Mann, der älter als 18 ist und zu Hause ist“).

Bei allen fünf Vorgehensweisen tritt darüber hinaus noch eine *Verzerrung durch Nicht-Antworten* (*non-response bias*) auf. Diese beruht darauf, dass ein Teil der Befragten die Antwort verweigern wird und dass das Wahlverhalten dieser Personen unter Umständen vom Rest abweicht. Außerdem werden im Allgemeinen nur sehr wenige Personen zugeben, dass sie nicht zur Wahl gehen, und auch deren Wahlverhalten kann vom Rest abweichen.

In Deutschland führt z. B. das Meinungsforschungsinstitut TNS Emnid im Auftrag von n-tv wöchentlich eine telefonische Wahlumfrage durch. Bei dieser werden ca. 1000 Wahlberechtigte befragt. TNS Emnid verwendet dazu eine Liste von 100.000 Telefonnummern, die aus einer zufällig aus Telefonbüchern und CD-ROMs ausgewählten Menge von Telefonnummern durch Modifikation der letzten Ziffer erzeugt wurde. Dabei soll die ebenfalls zufällige erfolgende Veränderung der letzten Ziffer sicherstellen, dass auch nicht in Telefonverzeichnisse eingetragene Haushalte in die Stichprobe gelangen können. Innerhalb der so ausgewählten Haushalte wird die Zielperson durch einen Zufallsschlüssel ermittelt. Dieser soll ausschließen, dass die Personen mit häufiger Anwesenheit eine größere Chance haben, befragt zu werden.¹³ Aus den Angaben der Befragten wird dann durch gewichtete Mittelungen die Wahlprognose erstellt.

Für die Wahl der Gewichte sind eine Reihe von Verfahren üblich: Meist wird zuerst eine sogenannte Transformationsgewichtung verwendet, die sicherstellt, dass Personen aus kleinen Haushalten in der Umfrage nicht überrepräsentiert sind (da die Umfrage ja zunächst Haushalte und nicht Personen zufällig auswählt). Als Nächstes kann man dann auch demografisch motivierte Gewichte verwenden, die die Stichprobe hinsichtlich ihrer sozialstrukturellen Zusammensetzung an die Menge aller Wahlberechtigten (deren sozialstrukturelle Zusammensetzung z. B. im Rahmen des Mikrozensus des Statistischen Bundesamtes approximativ ermittelt wird) anpassen. Schließlich gibt es noch die sogenannte Recall-Gewichtung. Dazu werden die Befragten auch nach ihrem Abstimmungsverhalten bei der letzten Bundestagswahl befragt. Die Recall-Gewichtung unterteilt die Gruppe der Befragten dann entsprechend dem angegebenen Abstimmungsverhalten bei der letzten Bundestagswahl und versucht, diesen Gruppen dann so Gewichte zuzuweisen, dass das (gewichtete) Abstimmungsverhalten bei der letzten Wahl mit dem tatsächlichen Ergebnis

übereinstimmt. Diese Gewichte sollen Antwortverzerrungen der Befragten korrigieren, die dadurch entstehen, dass ein Teil der Befragten die Antwort verweigert oder ihre Antwort an den gesellschaftlichen Konsens anpasst (z. B. hinsichtlich der Wahl von rechtsextremen Parteien). Problematisch an dieser Art der Gewichtung ist allerdings, dass sich viele Menschen nicht mehr genau an ihr Wahlverhalten bei der letzten Bundestagswahl erinnern können (es kommt z. B. vor, dass sie es mit ihrem Wahlverhalten bei der letzten Landtags- oder Gemeinderatswahl verwechseln).¹⁴

Aufgaben

2.1. In der sogenannten PISA-Studie werden in verschiedenen Ländern jeweils Schulen zufällig ausgewählt und Leistungstests für Schüler dieser Schulen in mehreren Fächern durchgeführt. Anschließend werden die Ergebnisse der einzelnen Länder nach Jahrgangsstufen getrennt miteinander verglichen.

- (a) Um welche Art Studie handelt es sich bei der PISA-Studie?
- (b) Inwieweit kann man Unterschiede in den Leistungen der Schüler in den verschiedenen Ländern auf Eigenschaften des Schulunterrichts in den einzelnen Ländern zurückführen? Begründen Sie Ihre Antwort.

2.2. Psychologen der Universität Leipzig beschäftigten sich in einer im Jahr 2008 veröffentlichten Studie mit der Frage, ob bei der Entstehung von Freundschaft eher die Ähnlichkeit der Persönlichkeiten oder der Zufall eine Rolle spielt. Dazu fingen sie einen ganzen Jahrgang neuer Psychologiestudenten vor ihrer ersten Vorlesung ab und teilten ihnen per Losnummer willkürlich Sitzplätze im Hörsaal zu. Ein Jahr später fragten sie die Studenten dieses Jahrgangs, wie gut sie mit ihren Kommilitonen befreundet seien. Dabei stellte sich heraus, dass diejenigen Personen, die zu Beginn des Studiums in der ersten Vorlesung nebeneinander gesessen hatten, im Schnitt besser befreundet waren als der Rest.

- (a) Wenn Sie die obige Studie als Studie im Sinne dieses Buches auffassen, um welche Art Studie handelt es sich dann und welche ist dann die Studiengruppe, welche die Kontrollgruppe?
- (b) Interpretieren Sie das Resultat dieser Studie. Gehen Sie dabei insbesondere auf die Frage ein, inwieweit es diese Studie erlaubt, auf einen Zusammenhang zwischen Freundschaft und Zufall zu schließen.

2.3. Eine Sozialerhebung des Deutschen Studentenwerks hat ergeben, dass in Deutschland 72 Prozent der Kinder aus vermögenden Familien aber nur 8 Prozent der Kinder aus einkommenschwachen Familien einen Studienabschluss erlangen. Kann man daraus schließen, dass ein kausaler Zusammenhang zwischen dem Einkommen der Eltern und dem Erlangen eines Studienabschlusses der Kinder besteht? Begründen Sie *kurz* ihre Antwort.

2.4. Der sogenannte Ig-Nobelpreis (ig ist hier Abkürzung für ignoble – unwürdig, schmachvoll, schändlich) wird vergeben für Forschungsleistungen, die Menschen erst zum Lachen und dann zum Nachdenken verleiten. Im Jahr 2013 wurde er im Bereich Psychologie für eine Studie vergeben, mit der nachgewiesen wurde, dass Menschen, die glauben, betrunken zu sein, auch glauben, attraktiv zu sein.

Wie würden Sie vorgehen, wenn Sie eine Studie entwerfen sollten, mit der Sie obige Aussage nachweisen sollen? Geben Sie dabei insbesondere an, wie Sie Ihre Probanden in Studien- und Kontrollgruppe unterteilen und wie Sie auf diese beiden Gruppen einwirken bzw. was genau Sie beobachten.

2.5. Die Evaluation von Vorlesungen an der TU Darmstadt erfolgt durch eine Umfrage, die gegen Ende des Semesters direkt in einer der Vorlesungsstunden an die Studierenden ausgeteilt wird. Dabei bekommen die Studierenden noch in der Vorlesung Zeit zum Ausfüllen der Umfragebögen, die anschließend eingesammelt und ausgewertet werden.

Beurteilen Sie diese Art der Umfrage hinsichtlich sampling bias und non-response bias.

2.6. Google Consumer Survey bietet an, Umfragen schnell und günstig im Internet durchführen zu lassen. Meist soll der Kunde sich dabei auf eine Frage beschränken, für die nur wenige Antwortmöglichkeiten vorgegeben werden, z. B.:

Wie viel Geld wären Sie bereit, für perfekte Brillengläser zu zahlen? a) Bis zu 50 Euro, b) 50-100 Euro, c) 100-200 Euro, d) Mehr als 200 Euro.

Diese wird dann in eine geeignete Webseite eingebaut, in der der Besucher wertvolle bzw. interessante Informationen finden kann (z. B. ein Nachrichtenportal). Bevor er diese Webseite aber nützen kann, muss er zunächst eine Frage beantworten, und für jede erteilte und an Google weitergeleitete Antwort enthält der Betreiber der Webseite Geld von Google. Der große Vorteil dieser Art von Umfragen ist, dass man sehr schnell Antworten von vielen verschiedenen Personen sammeln kann.

Beurteilen Sie diese Art der Umfrage hinsichtlich des sampling- und des non-response-bias.

Eine Einführung in die Statistik und ihre Anwendungen

Eckle-Köhler, J.; Köhler, M.

2017, X, 333 S. 44 Abb., Softcover

ISBN: 978-3-662-54093-0