

Den Gegenstand Ihrer Untersuchung festlegen

Zwischen diskreten und stetigen Merkmalen sowie zwischen quantitativen und qualitativen Merkmalen unterscheiden,

Einen Kompromiss zwischen Zufallsstichprobe und repräsentativer Stichprobe kennenlernen.

Kapitel 1

Klarmachen zum Datensammeln

Der Begriff »Statistik« ist von dem lateinischen Wort für »stehen« abgeleitet. Die Statistik interpretiert Daten aus der realen Welt und zeigt so, wie es um sie steht (jedenfalls in Hinblick auf die jeweils vorliegende Fragestellung). Daneben existiert die Wahrscheinlichkeitsrechnung als rein abstraktes Denkspiel. In den ersten zwei Teilen dieses Buches stelle ich Ihnen die beiden Gebiete einzeln vor. In Teil III zeige ich Ihnen dann, wie die Wahrscheinlichkeitsrechnung dabei hilft, bei statistischen Daten noch mehr Durchblick zu erzeugen.

Statistik beginnt nicht gleich mit dem Rechnen. Davor kommt erst einmal das Sammeln von Daten. Und davor wiederum kommen erst einmal ein paar grundsätzliche Überlegungen, damit Sie mit den Daten später auch etwas anfangen können. Um solche vorbereitenden Gedanken geht es in diesem Kapitel.

Wer Sie interessiert: Die Beobachtungsmenge

Eine sinnvolle Antwort nimmt ihren Anfang bei einer gut gestellten Frage. Auf die Statistik bezogen heißt dies: Der erste Schritt zu einer aussagefähigen Auswertung von Daten ist die Festlegung, worüber überhaupt eine Einsicht gewonnen werden soll.

Die Menge aller Objekte, über die Sie am Ende eine Aussage machen möchten, heißt die *Beobachtungsmenge*. Dasselbe bedeuten die häufig verwendeten Fachbegriffe *Grundgesamtheit* oder *Population*.

Offt lässt sich die Beobachtungsmenge ganz selbstverständlich festlegen:

- ✓ Mit einem Crashtest möchten Sie etwas über die gesamte Baureihe eines bestimmten Fahrzeugmodells erfahren.
- ✓ Bei der Markteinführung eines Produkts gibt es eine definierte Zielgruppe, deren voraussichtliche Reaktion auf die Neuheit Sie prognostizieren sollen.

Mitunter sind Untersuchungsvorhaben aber auch so komplex, dass Sie sie in einzelne Fragestellungen aufteilen müssen.

Für eine Untersuchung der Verkehrssituation in Ihrer Stadt kommen ganz unterschiedliche Beobachtungsmengen in Frage, beispielsweise

- ✓ die Menschen, die am Verkehr in dieser Stadt teilnehmen
- ✓ die Verkehrsmittel, mit denen der Verkehr in dieser Stadt stattfindet
- ✓ die Verkehrswege, über die der Verkehr in dieser Stadt abläuft.

Je nachdem, wie Sie die Beobachtungsmenge festlegen, ergeben sich dann Fragestellungen, die für diese Beobachtungsmenge sinnvoll beantwortet werden können.

- ✓ Hinsichtlich einer Person können Sie beispielsweise danach fragen, wie lange sie täglich im Verkehr unterwegs ist.
- ✓ Hinsichtlich eines bestimmten Verkehrsmittels können Sie etwa danach fragen, wie bei ihm das Verhältnis zwischen Personen- und Güterverkehr aussieht.
- ✓ Hinsichtlich eines Verkehrsweges wäre schließlich die Frage nach den jährlichen Unterhaltskosten denkbar.

Kaum noch Chance auf Eheglück?

Auf den Internetseiten des statistischen Bundesamtes können Sie nachlesen, dass im Jahr 2016 in Deutschland 410.426 Ehen geschlossen und 162.397 Ehen geschieden wurden. Nahezu unwiderstehlich ist die Versuchung, dies auf die Schlagzeile zu verkürzen »Fast jede zweite Ehe wird geschieden«. Mal abgesehen davon, dass hier geburtenstärkere Jahrgänge der Scheidungspaare mit geburtenschwächeren

der Hochzeitspaare gleichgesetzt werden: Heißt das, dass allenfalls gut die Hälfte der Menschen, die jemals heiraten, auf eine dauerhafte Ehe hoffen darf?

Diese vermeintlich naheliegende Schlussfolgerung lässt außer Acht, dass Geschiedene sich wieder verheiraten können.

Wenn die Ehepaare Alfred und Susanne einerseits und Bernd und Theresa andererseits sich scheiden lassen und anschließend in den Konstellationen Alfred und Theresa sowie Bernd und Susanne erneut heiraten und auf Dauer zusammenbleiben, dann ist offensichtlich von den vier Ehen, von denen hier die Rede ist, die Hälfte geschieden worden. Trotzdem haben alle vier Männer und Frauen im Laufe ihres Lebens eine dauerhafte Ehe geführt.

Es ist also wichtig, dass Sie hier zwischen der Beobachtungsmenge »Ehen« und der Beobachtungsmenge »verheiratete Menschen« sauber trennen – obwohl das eine ohne das andere überhaupt nicht existiert.

Die Elemente der Beobachtungsmenge, werden als *Beobachtungseinheit* bezeichnet – also:

- ✓ ein Auto aus der fraglichen Baureihe des Fahrzeugmodells,
- ✓ eine Person aus der Zielgruppe für das neue Produkt,
- ✓ eine Verkehrsteilnehmerin oder ein Verkehrsteilnehmer in der betrachteten Stadt,
- ✓ ein Verkehrsmittel (Pkw, Bus, Bahn, Fahrrad, Schiff, ...)
- ✓ eine Straße (oder ein Schienenabschnitt oder eine Route über Wasser oder ...)

Zur Durchführung Ihrer Untersuchung nehmen Sie sich Exemplare der Beobachtungseinheit vor und notieren, welche Beobachtungen Sie bei diesen Exemplaren machen.

Was Sie interessiert: Merkmale

Diejenige Eigenschaft der Beobachtungseinheit, auf die Sie Ihr Augenmerk richten, wird als **Beobachtungsmerkmal** bezeichnet, also: welche Farbe eine Testkundin für das neue Produkt vorschlägt, um welche Länge sich die Knautschzone verkürzt hat oder wie viele Menschen im Laufe des Tages in einer bestimmten Buslinie unterwegs waren. Das, was Sie über dieses Merkmal bei Ihrer Beobachtung notieren, wird als *Ausprägung* des Merkmals bezeichnet, also etwa, dass die Farbe Blau

vorgeschlagen wurde, dass sich die Knautschzone um 20,5 cm verkürzt hat oder dass über den gesamten Tag hinweg 1738 Fahrgäste in die Busse dieser bestimmten Linie eingestiegen sind. Ich werde oft aber auch weniger fachmännisch vom Wert des Merkmals sprechen. Was Sie mit den so gewonnenen Notizen anfangen können, hängt davon ab, von welcher Art das Merkmal ist.

Merkmale werden in zweierlei Hinsicht klassifiziert:

1. Diskret vs. stetig: **Diskrete Merkmale** haben Ausprägungen, die sich klar voneinander unterscheiden. Zwischen zwei unterschiedlichen Ausprägungen eines solchen Merkmals gibt es immer einen »Sicherheitsabstand«. Typische Vertreter diskreter Merkmale haben entweder endlich viele mögliche Ausprägungen, oder wenn sie unendlich viele mögliche Ausprägungen haben, dann sind sie so säuberlich voneinander getrennt wie die natürlichen Zahlen. **Stetige Merkmale** hingegen haben Ausprägungen, die stufenlos ineinander übergehen, so wie es beispielsweise bei der Menge aller reellen Zahlen zwischen 0 und 1 der Fall ist.
2. Quantitativ vs. qualitativ: **Quantitative Merkmale** haben Ausprägungen, die Zahlenwerte darstellen. Das allein reicht aber nicht. Es muss darüber hinaus auch sinnvoll sein, mit diesen Zahlenwerten zu rechnen. Als einfachen Test hierfür können Sie sich fragen, ob es sinnvoll ist, die Ausprägungen nach kleineren und größeren Werten zu sortieren. **Qualitative Merkmale** haben entweder Ausprägungen, die keine Zahlen sind, oder es handelt sich um Zahlen, die in keiner sinnvollen kleiner-größer-Beziehung zueinander stehen.

Beispiele:

- ✓ Die Anzahl der leiblichen Kinder einer Person ist ein *diskretes quantitatives* Merkmal: Als Antwort kommen nur die natürlichen Zahlen in Frage, aber keine Zwischenwerte (also ist das Merkmal diskret); außerdem gibt es auf die Frage, ob eine bestimmte Person mehr Kinder hat als eine andere, eine sinnvolle Antwort (also ist das Merkmal quantitativ).
- ✓ Die Ortskennung auf deutschen Autokennzeichen ist ein *diskretes qualitatives* Merkmal: Es gibt nur endlich viele Möglichkeiten (also ist das Merkmal diskret), und die Ausprägungen des Merkmals sind keine Zahlen (also ist das Merkmal qualitativ). Dass Sie Autokennzeichen nach dem Alphabet sortieren können, spricht nicht gegen diese Einstufung, denn trotz dieser Sortierung können Sie mit diesen Buchstaben nun einmal nicht rechnen. Oder könnten Sie sich unter dem arithmetischen Mittel (dieser Begriff wird in Kapitel 3 vorgestellt) der Autokennzeichen in Brandenburg etwas Sinnvolles vorstellen?

- ✓ Die aktuelle Anzahl der Umdrehungen eines Windrades pro Minute ist ein *stetiges quantitatives* Merkmal: Die Anzahl muss nicht ganzzahlig sein, im Gegenteil: Sie steigt und fällt stufenlos mit der Windstärke (also ist das Merkmal stetig), und die Frage danach, welches von zwei Windrädern sich gerade schneller dreht, hat eine sinnvolle Antwort (also ist das Merkmal quantitativ).
- ✓ Die momentanen GPS-Koordinaten eines LKW einer Spedition sind ein *stetiges qualitatives* Merkmal: Die Koordinaten ändern sich stufenlos (also ist das Merkmal stetig), und wenn ein LKW von Berlin nach Rostock fährt und ein anderer von Bremen nach München, dann ist es sinnlos zu fragen, wer von beiden denn jetzt die größeren Koordinaten habe (also ist das Merkmal qualitativ).
- ✓ Eine Verschlüsselung von Angaben durch Zahlen (beispielsweise 0=ledig, 1=verheiratet/Lebenspartnerschaft, 2=geschieden, 3=verwitwet) ist kein quantitatives Merkmal, denn die größer-kleiner-Beziehung zwischen den Zahlen hat keine Bedeutung. Vielmehr können Sie durch eine Vertauschung der Schlüsselwerte jede beliebige Beziehung zwischen den verschlüsselten Begriffen erzeugen.



Es ist sinnvoll, Zeiten, Längen oder Preise als stetige Merkmale anzusehen. Wenn es sich dann ergibt, dass auf Grund der gewählten Maßeinheiten nur ganzzahlige Werte auftreten, sollten Sie diese Angaben als Klasseneinteilung (was das ist, können Sie in Kapitel 2 nachlesen) auffassen. So steht dann beispielsweise die Angabe »zwei Sekunden« nicht für die genaue Zahl 2,000000..., sondern sie repräsentiert irgendeinen Wert zwischen 1,5 und 2,5 Sekunden.

Wen Sie tatsächlich befragen: Stichproben

Wenn Sie alle Elemente der Beobachtungsmenge untersuchen, erhalten Sie natürlich die zuverlässigsten Resultate. Bei einer Wahlabstimmung steht auf diese Weise zumindest die Meinung all derjenigen fest, die ihre Stimme abgegeben haben, und wir alle akzeptieren, dass auf dieser Grundlage für die nächsten Jahre Politik gemacht wird.

In der Regel verbietet sich aber ein solches Vorgehen. Bei einer Marketingstudie beispielsweise wäre es meistens schlicht zu teuer, die gesamte Bevölkerung mit einem Probeexemplar eines neuen Produkts zu versorgen und sie anschließend

nach ihrer Meinung darüber zu befragen. Und auf die Idee, wirklich jedes Auto vor der Auslieferung einem Crashtest zu unterziehen, wird wohl auch niemand kommen.

Immer wenn Sie eine Teilmenge der Beobachtungsmenge untersuchen, aus der Sie dann Rückschlüsse auf die gesamte Beobachtungsmenge ziehen, haben Sie es mit einer *Stichprobe* zu tun. Ob solche Rückschlüsse auf die Gesamtmenge sinnvoll sind, hängt auch davon ab, ob Sie gewisse Vorgehensweisen bei der Auswahl der Teilmenge einhalten.

Im Wesentlichen stehen sich hier zwei Konzepte gegenüber: die **Zufallsstichprobe** und die **repräsentative Stichprobe**.

Die Namen sprechen weitgehend für sich. Eine Zufallsstichprobe wird so aus der Beobachtungsmenge ausgewählt, dass jedes Element dieselbe Chance hat, berücksichtigt zu werden. Bei einer Umfrage vor einer Wahl könnten Sie beispielsweise eine Reihe von Nummern von Personalausweisen auslosen und die entsprechenden Wahlberechtigten befragen. Eine repräsentative Stichprobe hingegen ist so aufgebaut, dass sie alle relevanten Eigenschaften der Beobachtungsmenge widerspiegelt. Bei einer Umfrage vor einer Wahl sollte dies beispielsweise hinsichtlich der Anteile der Geschlechter, der Verteilung auf Altersgruppen, der geografischen Verteilung über das Bundesgebiet und des beruflichen Status der Fall sein (Ihnen fallen sicher ganz schnell noch weitere Gesichtspunkte ein, oder?).



Bei Zufallsauswahlen lassen Sie sich am besten durch einen Zufallszahlengenerator unterstützen. Die menschliche Vorstellung vom Zufall ist hier nicht immer ausreichend. Beispielsweise reagieren viele Menschen auf die Aufforderung, zufällig einen Eintrag im Telefonbuch auszuwählen, so, dass sie das Buch ungefähr in der Mitte aufschlagen und dort auf einen Namen ungefähr in der Mitte der Seite tippen. Das hat mit Zufall nur eingeschränkt etwas zu tun, weil es viele Einträge gibt, die praktisch keine Chance haben, auf diesem Wege ausgewählt zu werden.

Bei Wahlen sind sich die Forschungsinstitute mittlerweile einigermaßen sicher, welche Faktoren Einfluss auf das Wahlverhalten haben. Vielfach ist aber das Aufstellen einer repräsentativen Stichprobe ein zwar edles, doch zugleich auch unerreichbares Ziel, weil gar nicht alle relevanten Kriterien bekannt sind.

Zufallsstichproben, die nur einen kleinen Teil der Beobachtungsmenge umfassen, bergen andererseits das Risiko, dass relevante, aber kleine Gruppen in der Beobachtungsmenge, bei der Auswahl »Pech haben« und in der Stichprobe nicht (oder fast nicht) vorkommen.

Einen guten Kompromiss zwischen diesen beiden Ansätzen stellen **geschichtete Zufallsstichproben** dar.

So geht's

3. Wählen Sie ein diskretes Merkmal aus und notieren Sie, welche Anteile die einzelnen Ausprägungen in der Beobachtungsmenge ausmachen.
4. Jetzt betrachten Sie nacheinander jede Ausprägung für sich und wählen unter denjenigen Elementen der Beobachtungsmenge, die genau diese Ausprägung aufweisen, eine Zufallsstichprobe. Diese Stichprobe soll so groß sein, dass alle von Ihnen gewünschten Auswertungen möglich sind.
5. Werten Sie jede dieser Stichproben aus.
6. Für ein Gesamtergebnis aller Stichproben multiplizieren Sie jedes Teilergebnis mit dem Prozentsatz, den die betreffende Merkmalsausprägung an der gesamten Beobachtungsmenge ausmacht, und addieren dann die Teilergebnisse.

Beispiel:

In etlichen Bundesländern beginnt der Englischunterricht schon in der Grundschule, sodass Abiturientinnen und Abiturienten in diesem Fach auf 10 Unterrichtsjahre zurückblicken können. Im Gegensatz dazu wird die dritte Fremdsprache nur drei Jahre lang unterrichtet. Sie sollen nun untersuchen, welche Sprachkompetenz dieser dreijährige Unterricht im Vergleich zum Lernfortschritt im Langzeitfach Englisch gebracht hat. Dies soll mit einem standardisierten Test gemessen werden, bei dem für jede Fremdsprache bis zu 100 Punkte erzielt werden können.

Bei der Durchführung dieser Untersuchung müssen Sie beachten, dass die Schülerinnen und Schüler unter verschiedenen Drittsprachen wählen dürfen. In dem Bundesland, das Sie mit der Untersuchung beauftragt hat, gibt es folgende Verteilung:

- ✓ 70% der Schülerinnen und Schüler, die überhaupt eine dritte Fremdsprache lernen, wählen dafür Spanisch,
- ✓ 20% wählen Französisch,
- ✓ 8% wählen Russisch und
- ✓ 2% wählen Schwedisch.

Für jede dieser Sprachen wählen Sie per Zufall eine Gruppe von 50 Personen aus (mit dieser Anzahl kann man schon etwas anfangen), die diese Sprache drei Jahre lang gelernt haben. Bei allen Personen testen Sie die Sprachkompetenz und führen in jeder Gruppe die Auswertungen durch, die Sie möchten. Unter anderem werden Sie sicher die durchschnittliche Punktzahl in jeder Gruppe bestimmen.

Die durchschnittliche Punktzahl für die dritte Fremdsprache im Allgemeinen erhalten Sie dann, wenn Sie

- ✓ die Durchschnittspunktzahl für die Sprache Spanisch mit 0,7 (das ist dieselbe Zahl wie 70%, nur anders geschrieben),
- ✓ die Durchschnittspunktzahl für die Sprache Französisch mit 0,2,
- ✓ die Durchschnittspunktzahl für die Sprache Russisch mit 0,08 und schließlich
- ✓ die Durchschnittspunktzahl für die Sprache Schwedisch mit 0,02

multiplizieren und die vier Resultate addieren.

Das steckt dahinter

Die so konstruierte Stichprobe wird als *disproportional geschichtete Stichprobe* bezeichnet, weil die Größe der einzelnen Teil-Stichproben nicht auf die Mengenteile der Merkmalsausprägungen in der gesamten Beobachtungsmenge Rücksicht nimmt.

Auf diese Weise werden zwei Fliegen mit einer Klappe geschlagen:

- ✓ Für jede Merkmalsausprägung können Teilauswertungen vorgenommen werden.
- ✓ Das Gesamtergebnis ist bezüglich des gewählten Merkmals repräsentativ.

Im Falle der dritten Fremdsprache hätte eine repräsentative Auswahl von 200 Personen nur 4 Personen mit Kenntnissen in Schwedisch enthalten. Bei einer Zufallsauswahl hätte es auch sehr gut passieren können, dass überhaupt niemand mit dieser Sprachenwahl dabei ist. Aussagen über die Leistungsfähigkeit des Unterrichts in Schwedisch wären dann in beiden Situationen nicht möglich gewesen.



Bei politischen Umfragen wird die Meinung der Bevölkerung zu einer bestimmten Sachfrage sowohl insgesamt als auch aufgeschlüsselt nach der Anhängerschaft der Parteien dargestellt. Das funktioniert natürlich nur, wenn auch für Parteien, die nur von wenigen Prozent der Wählerschaft unterstützt werden, eine hinreichend große Anzahl von Sympathisantinnen und Sympathisanten in die Befragung einbezogen wird.

Darauf kommt's an

Sie können auch Schichtungen bezüglich mehrerer Merkmale vornehmen. Dann müssen Sie die Teilstichproben für jede Kombination aus Ausprägungen der beiden Merkmale getrennt wählen. Wenn Sie also beispielsweise im Fall der dritten Fremdsprache zusätzlich zwischen den Schulformen Gymnasium und Gesamtschule (ich weiß, dass viele Bundesländer hierfür eigene Bezeichnungen geschaffen haben, aber ich glaube, mit diesem Wort ist über Ländergrenzen hinweg klar, was ich meine) unterscheiden wollen, brauchen Sie insgesamt acht Stichproben für folgende Konstellationen:

- ✓ Spanisch am Gymnasium,
- ✓ Französisch am Gymnasium,
- ✓ Russisch am Gymnasium,
- ✓ Schwedisch am Gymnasium,
- ✓ Spanisch an der Gesamtschule,
- ✓ Französisch an der Gesamtschule,
- ✓ Russisch an der Gesamtschule,
- ✓ Schwedisch an der Gesamtschule.

Für jede dieser acht Kombinationen brauchen Sie dann auch die Prozentanteile an der Gesamtzahl aller derjenigen Schülerinnen und Schüler, die überhaupt eine dritte Fremdsprache gelernt haben.



Bei den PISA-Erhebungen über die Leistungsfähigkeit von deutschen Schulen wird in jedem Bundesland für jede Schulform eine eigene Stichprobe genommen. So kann man auch für kleine Bundesländer und auch für Schulformen, die nur von vergleichsweise wenigen Schülerinnen und Schülern besucht werden, zu Ergebnissen kommen. Für den internationalen Vergleich werden diese Stichproben dann zu einem Ergebnis für ganz Deutschland zusammengesetzt.

Falls Sie nach einem Merkmal absichten wollen, das nicht diskret, sondern stetig ist, müssen Sie zunächst eine Klasseneinteilung durchführen. Dieses Vorgehen wird in Kapitel 2 vorgestellt.

Leider gibt es kein Rezept, mit dem Sie Schwächen bei der Auswahl von Stichproben gänzlich ausschalten können. Die dauernde Frage, wo Sie einen fehlenden Bezug

zur Realität vielleicht doch noch übersehen haben könnten, gehört zu sauberer statistischer Arbeit einfach dazu.



Eine niedrige Rücklaufquote – also eine geringe Zahl von eingehenden Antworten auf eine große Zahl von Einladungen zu einer Umfrage – ist ein starker Hinweis darauf, dass die Ergebnisse einer Erhebung verzerrt oder sogar unbrauchbar sind. Offensichtlich war es für die weitaus meisten angesprochenen Personen herzlich uninteressant, sich an der Befragung zu beteiligen. Bei denen, die trotzdem geantwortet haben, muss ein besonderer Antriebsgrund vorgelegen haben. Wenn aber schon die Motive der Teilnehmer nicht typisch für die Gesamtheit der Befragten sind, dann werden es ihre Angaben zum Gegenstand der Befragung höchstwahrscheinlich ebenfalls nicht sein.

Auch wenn Sie die Auswahl der Stichprobe erst einmal als vertretbar ansehen, lauert eine weitere Fehlerquelle schon hinter der nächsten Ecke – nämlich, bei der Art und Weise, wie Sie der Beobachtungseinheit die gewünschten Auskünfte entlocken. Bei technischen Messungen kann es Fehler im Versuchsaufbau, im Versuchsablauf oder beim Ablesen der Messwerte geben. Bei der Befragung von Menschen können heikle Themen Anlass zu bewusst falschen Antworten sein. Aber auch unbewusste Assoziationen durch bestimmte Formulierungen in der Fragestellung können die Antworten beeinflussen. Die Forderung nach mehr Lehrern in den Schulen wird auf höhere Zustimmung treffen als der Vorschlag, mehr Geld für Lehrergehälter auszugeben – obwohl doch niemand ernsthaft glauben kann, die jetzigen Lehrerinnen und Lehrer würden die neuen Kolleginnen und Kollegen von ihrem eigenen Gehalt mit durchfüttern.

Sechs von sieben Autos vom Marder beschädigt?

Vor mehreren Jahren fragte der ADAC die Leser seiner Mitgliederzeitschrift, ob sie schon einmal einen erheblichen Marderschaden an ihrem Auto gehabt hätten. Fast 86% der Antworten lauteten »ja«. Allerdings hatten von den damals deutlich über zehn Millionen Lesern der Zeitschrift nur 5834 geantwortet – also erheblich weniger als ein Promille. Rückschlüsse vom Ergebnis der Umfrage auf die tatsächliche Schadenshäufigkeit in Deutschland lassen sich auf dieser Grundlage wohl kaum ziehen.

Übungsaufgabe

Welche der folgenden Merkmale einer Wohnung sind diskret, beziehungsweise stetig? Welche sind quantitativ und welche qualitativ?

- a) Das Fabrikat des Kühlschranks,
- b) das Alter des ältesten Mitglieds des Haushalts,
- c) die (Himmels-)Richtung, in die das Schlafzimmerfenster ausgerichtet ist.
- d) die Anzahl der Zimmer.

