

# 1 Wegweiser

## 1.1 Ausgangspunkt

Statistik Klausuren sind ein Ärgernis – das finden in überraschender Einmütigkeit viele Studierende und Dozenten von Statistikkursen für Hörer verschiedenster Fächer. Allerdings mit unterschiedlicher Begründung: Die Studierenden monieren, dass sie leere Formalismen pauken müssen, deren Sinn ihnen verschlossen bleibt und die sie nach bestandener Klausur baldmöglichst vergessen. Die Dozenten bedauern, dass die Studierenden auf die Klausuren fixiert und an dem „eigentlichen“ Stoff gar nicht interessiert sind. Diese doppelt beklagenswerte Situation enthält freilich auch eine Chance und trägt gewissermaßen schon einen Schlüssel zu ihrer Auflösung in sich. Wenn nämlich die Klausuraufgaben gerade den „eigentlichen“ Stoff einfordern, also das widerspiegeln, was vermittelt werden soll, ist die Motivation der Studierenden von alleine in die richtige Richtung gelenkt. Wenn zudem dieser „eigentliche“ Stoff auch einsehbar praxisrelevant ist, werden die Studierenden ihn auch nicht als leeren Formalismus empfinden.

Erfolgreiche Lehre für Anwender muss demnach bei den **Klausuraufgaben** ansetzen. Das ist der Ausgangspunkt dieses Buches, das auf 16 Klausuren basiert, die der Autor mit eben dieser Ausrichtung in den vergangenen Jahren gestellt hat. Die Klausuraufgaben sind meist als kleine Fallstudien konzipiert und manchmal sogar direkt aus der Tagespresse entlehnt. Ihre Relevanz ist ziemlich leicht erkennbar. In der Tat benötigt man statistische Grundkenntnisse in nahezu **allen Lebensbereichen**. Man betrachte dazu einmal die Titelseite einer beliebigen (renommierten) Tageszeitung und ermittle – schon wieder eine statistische Information! – den Prozentsatz, den statistische (also im weitesten Sinne quantitative) Information dort einnimmt. Schon um diese Information sinnvoll nutzen und bewerten zu können, benötigt man statistische Kenntnisse – man benötigt sie allerdings im **aktiven Wissensschatz**, nicht als halbvergessene tote Formeln von „damals“.

Natürlich klingt das fast zu schön, um wahr zu sein. Und in der Tat wird man schon wegen der Prüfungssituation einige Kompromisse machen müssen. So lassen sich ganz einfach nicht alle für die statistische Praxis relevanten Fertigkeiten in Aufgaben fassen. Auch müssen Klausuraufgaben in vertretbarer Zeit korrigierbar sein und sollten – schon aus Gründen der rechtlichen Überprüfbarkeit – scharf umrissene Lösungen zulassen. Man wird das Ideal also sicher nicht ganz erreichen. Aber das ist kein Grund, sich nicht wenigstens in die richtige Richtung zu bewegen.

## 1.2 „Statistik-Fahrschule“: 16 Doppelstunden und etwas Theorie

Das Konzept dieses Buches lässt sich gut mit der Analogie zur Fahrschule beschreiben. Dort gibt es im Wesentlichen drei Komponenten: die Fahrstunden, die theoretische Ausbildung und den Fahrlehrer. Gehen wir sie einmal der Reihe nach durch.

### 1.2.1 Die Fahrstunden

Wohl für jeden Fahrschüler sind diese das **Herzstück** seiner Ausbildung. Man stelle sich einmal vor, die Fahrausbildung bestünde nur in theoretischer Unterweisung oder nur im praktischen Drill gewisser Standardtechniken (etwa Einparken oder Tachometer ablesen)! (Die Analogie zu gewissen Statistikkursen und vor allem Statistikklausuren kann jeder Leser selbst bilden.) Das Besondere an den Fahrstunden ist gerade, dass sie den Fahrschüler dem realen Verkehr mit seinen ständig neuen Situationen aussetzen und ihn zwingen, viele verschiedene Standardtechniken und Standardwissen zu koordinieren und situationsgemäß einzusetzen. Nicht zuletzt deswegen ist man in den ersten Stunden komplett überfordert. Später hat man den Durchblick erworben und das Zusammenspiel gelernt. Dann geht es wie von selbst, und man kann nebenher Radio hören.

Die folgenden **Übungsklausuren** sind ähnlich gehalten. Die Teilnehmer werden darin, soweit möglich, den Bedingungen der Praxis ausgesetzt. Es geht nicht darum, nur schematisch etwas auszurechnen. Gute Statistik besteht ja (wie gutes Autofahren) gerade darin, sich in unübersichtlichen Situationen zurechtzufinden und die Aufmerksamkeit auf das jeweils Wichtige zu lenken. Wie die erste Fahrstunde mag der Leser auch die erste Klausur als unübersichtlich und als Überforderung empfinden, weil sie nicht schematisch ist. Aber genau auf die Fähigkeit, damit zurecht zu kommen, kommt es an. Weil eben auch die Realität nicht schematisch ist.

Aus diesem Grunde sind die Klausuraufgaben auch als **ganze Klausuren** belassen und nicht etwa nach Themengebieten sortiert. Denn wenn man weiß, welches Themengebiet gerade behandelt wird, ahnt man meist schon die Lösung. Diese scheinbare Unübersichtlichkeit ist also ganz bewusst gewählt. Der Leser – besser: Löser – muss selbst herausfinden, welche Technik jeweils angebracht ist – ganz wie im richtigen Leben.

### 1.2.2 Die theoretische Ausbildung

Diese ist idealerweise bereits durch einen Statistikkurs erfolgt oder erfolgt parallel. Der Autor würde dazu das hervorragende Lehrbuch [6] von Freed-

man/Pisani/Purves empfehlen, aber auch für jeden anderen Kurs ist das vorliegende Buch eine nützliche Begleitlektüre. Wo Grundkenntnisse fehlen oder aufgefrischt werden sollten, kann dies durch einen Blick in das [Auffrischkapitel](#) geschehen. Dort sind eigentlich alle benötigten Kenntnisse in knapper Form zusammengestellt, so dass das Buch auch ganz eigenständig benutzt werden kann. Das Auffrischkapitel dient zudem dem Zweck, die [Sprachregelungen](#) zu vereinheitlichen und einige besondere Begriffe dieses Kurses (wie z. B. „zweitschgenförmiges Streudiagramm“) einzuführen. Eine weitere kleine Eigenheit des Buches ist es, dass als [Dezimaltrennzeichen](#) statt eines Kommas durchgängig (außer in wörtlichen Zitaten) ein Punkt verwendet wird. Der Grund dafür ist, dass in der Statistik oftmals Programmpakete benutzt werden, die nur in international üblicher Form vorliegen. Wer Lehrbücher in deutscher Sprache bevorzugt, kann zu vielen bewährten Titeln – wie beispielsweise [1], [5], [8] oder [13] – greifen. Ein sehr ausführliches und lesenswertes neueres Buch ist [7]. Auch ältere Darstellungen – wie [10], [11] oder [17] – lohnen wegen ihrer vielen guten Gedanken und Beispiele nach wie vor einen Blick. Ganz modern und ganz gezielt kann man sich zu Einzelfragen im Internet kundig machen, wobei der Autor insbesondere [Wikipedia](#) hervorheben möchte.

### 1.2.3 Der Fahrlehrer

Der Fahrschüler sitzt nicht allein im Auto. Neben ihm sitzt der Fahrlehrer und passt auf. Natürlich haben in diesem Buch die [Lösungsvorschläge](#) diese Funktion. Der Lernende kann hieraus selbst ersehen, inwieweit er wichtige Aspekte der Aufgabenstellung erfasst hat. Die Lösungen sind absichtlich sehr ausführlich gehalten. Nach Erfahrung des Autors gibt es nämlich viele Defizite bei der Darstellung von Sachverhalten, die man „im Prinzip“ verstanden hat. Zwischen einer numerisch richtigen Rechnung oder einem Schlagwort und einer akzeptablen Präsentation von Ergebnissen können Welten liegen. Aber Letzteres ist eine für die Anwendung der Statistik im Berufsleben unabdingbare Kompetenz. Deswegen werden Multiple-Choice-Aufgaben auch weitgehend vermieden. Wer nur Multiple-Choice-Aufgaben anklickt – dies ist oft bei Lernprogrammen der Fall – entwickelt nicht die Fähigkeit, Sachverhalte eigenständig und korrekt zu formulieren. (Auch die Fragen- und Antwortenkataloge etwa für Abiturprüfungen sind nach Meinung des Autors in dieser Hinsicht eher kontraproduktiv, weil dabei meist nur ein Rechenergebnis, aber keine ausführliche verbale Herleitung angegeben wird.) Idealerweise sollte der Leser also die Lösung selbst ausformulieren und erst nachher mit dem Lösungsvorschlag vergleichen. Dies ist übrigens auch der Grund, weshalb die Lösungen nicht unmittelbar unterhalb der jeweiligen Aufgabe angegeben sind. Es ist sonst noch schwerer, der Versuchung zu widerstehen, schon einmal in die Lösung zu schauen.

Kurz und knapp: Die hier gesammelten Klausuren legen Wert auf **korrektes Denken** und **korrektes Darstellen**. Wenig Wert wird auf kompliziertes Rechnen gelegt, denn das kann der Computer besser. Deswegen sind alle Klausuren auch „im Kopf“ (d. h. ohne Taschenrechner) zu bearbeiten. Lediglich die im Anhang A abgedruckte Tabelle der Normalverteilung wird benötigt. Das Buch kann insofern überall – insbesondere in Pausen oder bei Bahnfahrten – schnell hervorgeholt und benutzt werden.

Zusätzlich geben viele Fahrlehrer mehr oder weniger kluge Weisheiten an den Fahrschüler weiter (z. B. über „Herren mit Hut“ am Steuer oder ähnliches). Auch dafür haben wir ein Analogon: die **Kommentare** zu einigen Lösungen. Hier finden sich Anmerkungen, die die jeweilige Aufgabenlösung in einen weiteren Kontext einordnen, auf zusätzliche wichtige Aspekte aufmerksam machen oder die der Autor einfach irgendwie loswerden wollte. Sie gehören nicht wirklich zur Aufgabenlösung, können aber für manche Leser ganz besonders wertvoll sein. Formal sind sie durch farbige Unterlegung gekennzeichnet, damit man sie entweder leicht finden oder leicht überspringen kann. Hier ist ein Beispiel:

Kommentar: Es ist gut, bei Klausuren möglichst viele und „in Flensburg“ möglichst wenige Punkte zu erzielen.

Ein mitteilungsfreudiger Fahrlehrer kommentiert nicht nur während der Fahrstunden, sondern auch während des theoretischen Unterrichts. Analog gibt es Kommentare auch im Auffrischungskapitel.

#### 1.2.4 Fazit

Der Ansatz dieses Buches unterscheidet sich möglicherweise von anderen Kursen, die Sie besuchen. Es geht hier nicht so sehr um **Informationsvermittlung**, sondern um **Kompetenzerwerb**. Sie sollen also nicht in erster Linie über etwas „orientiert“ oder „informiert“ werden. Der Anspruch ist vielmehr, sich eine neue Fertigkeit so anzueignen, dass man sie in verschiedensten Situationen einsetzen kann. Wie beim Schreiben und Lesen oder beim Autofahren ist das zeitaufwendig und erfordert Übung. Aber es lohnt sich. Kompetenzen verlernt man – im Unterschied zu Informationen – auch nicht so schnell wieder.

Die Analogie zum Schreiben und Lesen oder zum Autofahren ist noch in einer anderen Hinsicht passend: Statistisches Denken ist ebenfalls eine **Schlüsselqualifikation**, die in den verschiedensten Gebieten und Kontexten benötigt wird.

## 1.3 Tipps zum Umgang mit diesem Buch

Arbeiten Sie mit diesem Buch, um statistisches Denken für Ihre sonstige Arbeit zu erlernen oder als Vorbereitung, um eine Pflichtklausur zu bestehen? Glücklicherweise stellt sich diese Frage gar nicht, weil die Empfehlungen in beiden Fällen gleich lauten. Denn unser Ausgangspunkt ist ja gerade, dass es keinen Unterschied zwischen den Klausuren und der eigentlich benötigten Statistik und folglich auch keine Diskrepanz zwischen den beiden oben genannten Zielen geben sollte – außer natürlich dem formalen Prüfungscharakter einer Klausur. Die Empfehlungen in Abschnitt 1.3.1 sind also für alle Leser gültig. In Abschnitt 1.3.2 gibt es dann noch einige spezielle Tipps für die Stresssituation „Klausur“.

### 1.3.1 Tipps zur Klausurvorbereitung

1. Blättern Sie als erstes das [Auffrischungskapitel](#) durch. Das meiste wird Ihnen aus Ihrem Statistikkurs bekannt sein oder unmittelbar einleuchten. Sollten dennoch Lücken verbleiben, können Sie diese gegebenenfalls durch die empfohlene Literatur füllen.
2. Bearbeiten Sie dann die [Übungsklausuren](#). Diese sind in keiner speziellen Reihenfolge angeordnet; daher spielt es überhaupt keine Rolle, mit welcher Sie beginnen. Ähnlich wie bei Fahrstunden kommt es nur darauf an, sich den fallstudienartigen Aufgaben auszusetzen und daraus zu lernen.
3. Am meisten profitieren Sie, wenn Sie die Lösungen selbst ausarbeiten und erst danach mit den [Lösungsvorschlägen](#) vergleichen. Sie benötigen nur Bleistift und Papier sowie die Normalverteilungstabelle im Anhang A. Alle Rechnungen lassen sich leicht ohne Taschenrechner ausführen. Es ist sinnvoll, die Lösungen recht ausführlich auszuarbeiten und nachher selbstkritisch zu prüfen, welche Einzelschritte man erkannt oder vielleicht übersehen hat.
4. Geben Sie nicht zu schnell auf, wenn Ihnen die Klausuren zu Beginn schwer fallen. Denken Sie an Ihre erste Fahrstunde! Nach einigen Klausuren werden Sie von alleine mehr Übersicht auch in komplizierteren Fragen entwickeln. Das ist genau die Kompetenz, die wir entwickeln wollen. Denken Sie auch daran, dass [Kompetenzerwerb](#) mehr Zeit und Übung benötigt als reine [Informationsvermittlung](#).
5. Sollten Sie [gezielt](#) ein Themengebiet durch Aufgaben wiederholen wollen oder umgekehrt bei einer Aufgabe ganz festhängen, können Sie die [Kreuzreferenztablette](#) aus Anhang B zu Rate ziehen. Davon sollten Sie aber nur sehr sparsam Gebrauch machen. Deswegen wurde diese Tabelle

auch bewusst auf die Aufgaben der Klausuren mit den Abschnittsnummern 3.1, 3.3, 3.5, 3.7, 3.9 und 3.11 beschränkt.

### 1.3.2 Tipps zur Klausurbearbeitung

Statistik Klausuren sind Prüfungs- und Stresssituationen. Deswegen sind hier noch einige Tipps zusammengefasst, die der Autor beim Betreuen und Auswerten von vielen Klausuren immer wieder bestätigt gefunden hat.

1. Lesen Sie die Aufgaben **genau** durch. Oftmals ist es viel einfacher als man glaubt.
2. Bearbeiten Sie diejenigen Aufgaben, zu denen Sie eine Lösung **direkt** sehen, als erste. Sie schaffen sich so ein Polster und können gelassen und mit einem Erfolgserlebnis an die schwierigeren Aufgaben gehen. Zudem sind Sie für die weiteren Aufgaben durch eine positive Erfahrung „beflügelt“. Wenn Sie mit (für Sie) schwierigen Aufgaben beginnen, haben Sie nachher eventuell nicht genügend Zeit und geraten in Panik. Außerdem können Sie die negative Erfahrung mit der schwierigen Aufgabe vielleicht nicht schnell genug überwinden.
3. Verbeißen Sie sich in einer realen Klausur nicht in eine Aufgabe, mit der Sie **nicht zurechtkommen**. Jede Klausur enthält einige Redundanz. Es wird also nicht erwartet, dass man alle Aufgaben bearbeitet. In jeder Klausur können 120 Punkte erreicht werden. Die Punktezahl einer Aufgabe gibt in etwa die Bearbeitungszeit in Minuten an, falls man mit dem Stoff gut vertraut ist. Viel mehr Zeit sollten Sie (zumindest nach einer Gewöhnungsphase) nicht darauf verwenden.
4. Wenn Sie zu Nervosität bei Klausuren neigen, ist es eine gute Idee, die Übungsklausuren als „echte Probedurchläufe“ zu inszenieren. Schirmen Sie sich dazu für genau 120 Minuten von Störungen ab, und stellen Sie sich vor, die Übung sei schon der Ernstfall. Besser noch: Machen Sie dies in einer Arbeitsgruppe, in der dann später Ihre Klausur von jemand anderem korrigiert wird. Es ist einfach etwas anderes, ob man eine Lösung nur für sich notiert oder für eine zweite Person aufschreibt. Zudem hilft die **inszenierte Klausursituation** tatsächlich, eine gewisse „Routine“ und damit mehr Gelassenheit auch für die reale Klausur zu erreichen. Das Motto ist hier: Es ist besser, sich vorher selbst ein wenig unter Druck zu setzen, als nachher wirklich unter Druck zu kommen.

## 2 Auffrischungen aus der Theorie

### 2.1 Grundlagen

#### 2.1.1 Ein einführendes Beispiel statistischen Denkens

Wie gewinnt man gesicherte Erkenntnisse? Keine einfache, aber doch schon eine sehr alte Frage. Betrachten wir dazu ein Beispiel aus der Bibel [2] (Daniel 1, 8–17), das der Statistiker Mosteller (in [9], S. 881) einmal analysiert hat:

Daniel war entschlossen, sich nicht mit den Speisen und dem Wein der königlichen Tafel unrein zu machen, und er bat daher den Oberkämmerer darum, sich nicht unrein machen zu müssen. Gott ließ ihn beim Oberkämmerer Wohlwollen und Nachsicht finden. Der Oberkämmerer aber sagte zu Daniel: Ich fürchte mich vor meinem Herrn, dem König, der euch die Speisen und Getränke zugewiesen hat; er könnte finden, dass ihr schlechter aussieht als die anderen jungen Leute eures Alters; dann wäre durch eure Schuld mein Kopf beim König verwirkt. Da sagte Daniel zu dem Mann [...]: Versuch es doch einmal zehn Tage lang mit deinen Knechten! Lass uns nur pflanzliche Nahrung und Wasser zu trinken geben. Dann vergleiche unser Aussehen mit dem der jungen Leute, die von den Speisen des Königs essen.[...] Der Aufseher nahm ihren Vorschlag an und machte mit ihnen eine zehntägige Probe. Am Ende der zehn Tage sahen sie besser und wohlgenährter aus als all die jungen Leute, die von den Speisen des Königs aßen. Da ließ der Aufseher ihre Speisen und auch den Wein, den sie trinken sollten, beiseite und gab ihnen Pflanzenkost.

Der Text ist sehr aufschlussreich. Erkenntnis wird hier durch einen **Versuch**<sup>1</sup>, genauer durch einen **Vergleich**<sup>2</sup>, gewonnen.

Sollte man daraufhin zum Vegetarier werden? Für Daniel und seine Freunde ist das sicher der richtige Weg, aber gilt dies auch allgemein? Da gibt es Zweifel, denn das Ergebnis ist nicht zwingend: Daniel lehnte die nichtvegetarischen Speisen aus religiösen Gründen ab. Er wünschte sich eine vegetarische Ernährung, und es könnte durchaus sein, dass einem diejenigen Speisen gut bekommen, die man gerne essen möchte. Oder seine stärkere Religiosität könnte zur besseren Gesundheit geführt haben. Der Vergleich müsste für eine verallgemeinerungsfähige Aussage also anders gestaltet werden.

---

<sup>1</sup>Versuch macht klug.

<sup>2</sup>Vergleich macht reich.

Aus heutiger Sicht würde man zunächst eine Gruppe von Probanden suchen, die an einem entsprechenden Versuch teilnehmen würden. Diese Gruppe würde man dann in eine **Kontrollgruppe** und eine **Behandlungsgruppe** unterteilen. Die Kontrollgruppe würde wie bisher gepflegt, die Behandlungsgruppe mit vegetarischer Kost. Die Aufteilung darf nicht den Probanden selbst überlassen bleiben, damit sich deren Vorlieben nicht auswirken können. Sie muss also vom Versuchsleiter vorgenommen werden. Dies ist charakteristisch für ein sogenanntes **kontrolliertes Experiment**. Am besten geschieht die Aufteilung **randomisiert**, d. h. durch einen Zufallsmechanismus (z. B. Münzwurf), um jegliche Verzerrung soweit als möglich auszuschließen. Man wird weiterhin ausschließen wollen, dass Vorurteile der Beurteiler das Resultat verfälschen, und daher eine **Verblindung** vornehmen. Die Beurteiler wissen dann also nicht, ob ein Teilnehmer, dessen Zustand sie beurteilen, aus der Kontroll- oder der Behandlungsgruppe stammt. Wünschenswert wäre, dass auch die Versuchsteilnehmer selbst das nicht wissen. Bei Ernährungsversuchen ist das nicht durchführbar, bei Medikamententests kann es aber oft durch Gabe eines Scheinmedikamentes (**Placebo**) erreicht werden. Gesicherte Erkenntnisse gewinnt man also durch einen Vergleich unter möglichst (bis auf die zu untersuchende Behandlung) identischen Bedingungen.

Nicht immer lässt sich ein kontrolliertes Experiment durchführen. Soll etwa die Gefährlichkeit von Asbest untersucht werden, ist es wenig wahrscheinlich, dass sich Teilnehmer finden, die sich der Behandlungsgruppe zuweisen lassen. Man muss sich dann mit **Beobachtungsstudien** begnügen, bei denen die Zuordnung zur Behandlungs- oder Kontrollgruppe nicht in der Hand des Versuchsleiters liegt. Dadurch besteht aber die Gefahr, dass die beiden Gruppen sich auch in anderer Hinsicht als nur der zu untersuchenden Behandlung unterscheiden – nämlich zum Beispiel in den Kriterien, die die Aufteilung bewirkt haben. In Beobachtungsstudien können sich also weitere Einflüsse (sogenannte **vermengende Faktoren**) unauflösbar mit dem Behandlungseinfluss vermischen. Aus diesem Grunde kann man aus Beobachtungsstudien zwar nützliche Hinweise erhalten, aber keine beweiskräftigen Schlüsse ziehen. Auch das obige Bibelbeispiel ist eine Beobachtungsstudie. In diesem Fall hatte sich die Behandlungsgruppe durch Eigenauswahl gebildet, da Daniel und seine Freunde um vegetarische Ernährung gebeten hatten. Man kann aber z. B. nicht sagen, ob sich das bessere Aussehen wirklich aus der vegetarischen Ernährung oder einer stärkeren Beachtung religiöser Vorschriften ergab, die ja auch andere Lebensbereiche betreffen dürfte.

Um Entwicklungen im Zeitablauf zu betrachten, eignen sich besonders **Längsschnittstudien**, bei denen eine Gruppe im Zeitablauf beobachtet wird. Hat man etwa eine Gruppe von Menschen mit Geburtsjahr 1950 in den Zeitpunkten 1970 und 2010 untersucht, wird man feststellen, dass deren Haare im Zeitablauf vom Alter zwanzig bis zum Alter sechzig grauer geworden sind.



Hat man stattdessen im Rahmen einer **Querschnittstudie** im Jahre 2010 eine Gruppe von Sechzigjährigen und eine Gruppe von Zwanzigjährigen untersucht, wird man auch feststellen, dass die Haare der Sechzigjährigen grauer sind. Man kann aber nicht sagen, ob Menschen im Laufe ihres Lebens grauere Haare bekommen oder ob vielleicht Menschen im Jahre 1950 mit grauere Haaren zur Welt kamen als im Jahre 1990.

Kommentar: „Denken heißt Vergleichen“ ist ein bekannter Aphorismus von Walther Rathenau ([12], S. 32). Dieser Spruch überrascht zunächst. Denn natürlich muss man beim Vergleichen denken, aber dass beides nahezu identisch ist, erschließt sich nicht sofort. Wenn der Leser aber zum Beispiel darüber nachdenkt, ob er „reich“ ist, dann werden die meisten das mit Blick auf die Weltbevölkerung bejahen, mit Blick auf Fußballstars aber verneinen. Man sieht: Außer in Formalwissenschaften (Logik, Mathematik) ist relevantes Denken ohne Vergleichen in der Tat kaum möglich. Für uns hat das eine aufmunternde Konsequenz: Wenn relevantes Denken und Vergleichen identisch sind, dann ist die „Wissenschaft vom Vergleichen“ (also die Statistik) eben auch die „Wissenschaft vom relevanten Denken“ – ein erhebendes Gefühl!

### 2.1.2 Grundstruktur statistischer Überlegungen

Etwas abstrakter ist die Grundstruktur aller statistischen Überlegungen in der Abbildung 2.1 beschrieben:

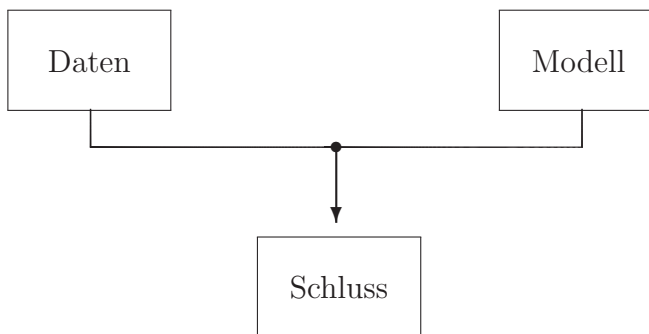


Abbildung 2.1: Grundmodell der Statistik

Im Wesentlichen ist es stets so, dass aus **Daten** mit Hilfe von **Modellen** ein **Schluss** gezogen wird. Man muss sich dabei jederzeit bewusst sein, in welchem Ausmaß Daten und Modell letztlich zur Schlussfolgerung beitragen. Insbesondere die Rolle des Modells wird häufig unterschätzt und muss daher

jeweils genau diskutiert werden. Stellt man etwa durch Marktforschung fest, dass in einem Land niemand Taschenuhren besitzt, kann die Schlussfolgerung sein, dass dies kein Markt für Taschenuhren ist, da niemand welche kauft. Sie kann aber auch ganz im Gegenteil lauten, dass dies ein großartiger Markt für Taschenuhren ist, weil noch niemand eine hat. Trotz gleicher Daten gelangt man also zu völlig verschiedenen Schlüssen. Das Modell ist gewissermaßen der Scheinwerfer, mit dem die Daten beleuchtet werden. Je nach der Farbe des Scheinwerferlichtes erscheinen auch die Daten in der entsprechenden Farbe. Deswegen ist die Wahl des richtigen Modells von ausschlaggebender Bedeutung.

Kommentar: Ein bemerkenswerter Fall unterschiedlicher Modellannahmen ergab sich während des Zweiten Weltkrieges bei Studien zur besseren Panzerung amerikanischer Flugzeuge gegen Flakbeschuss. Der Statistiker A. Wald (vgl. [16]) ließ die Militärs zunächst feststellen, wo sich bei beschossenen Maschinen die Einschussstellen befanden. Anschließend schlug er zur allgemeinen Überraschung vor, diejenigen Stellen stärker zu panzern, an denen man keine Einschüsse festgestellt hatte. Der Grund für die Verwunderung der Militärs waren unterschiedliche Modellannahmen. Die Militärs gingen davon aus, dass die Flugzeuge eine Stichprobe aus den beschossenen Flugzeugen waren und dass daher Flugzeuge an den festgestellten Einschussstellen besonders häufig getroffen wurden. Wald ging dagegen davon aus, dass die Treffer in etwa gleichmäßig über das Flugzeug verteilt waren. Flugzeuge, die an wirklich gefährlichen Stellen getroffen wurden, kehrten aber nicht zurück, so dass an diesen Stellen auch keine Einschüsse festgestellt wurden. Die untersuchten Flugzeuge waren also seiner (realistischeren) Modellvorstellung nach eine Stichprobe von Flugzeugen, die nur an relativ unproblematischen Stellen getroffen worden waren. Je nach Modell markieren also die Daten (Einschussstellen) entweder besonders gefährdete oder relativ ungefährdete Flugzeugteile.

Mit den Daten und ihrer Aufbereitung beschäftigt sich die [deskriptive Statistik](#). Die Ausarbeitung und Untersuchung von Modellen ist die Domäne der [Wahrscheinlichkeitstheorie](#). In der [analytischen Statistik](#) werden dann beide Elemente zusammengeführt.

Daten begegnen dem Statistiker zumeist in numerisch kodierter Form. Es ist aber wesentlich, sich über das Messskalenniveau im Klaren zu sein, auf dem sie aufgenommen wurden. Auf einer [Nominalskala](#) werden einfach nur Eigenschaften nebeneinander gestellt wie z. B. „männlich“ und „weiblich“. Durch eine [Ordinalskala](#) ist hingegen eine Rangordnung vorgegeben, z. B. eine Klassifikation als „schwacher“, „mittlerer“ oder „starker“ Raucher. Die Abstände zwischen den Kategorien müssen dabei nicht unbedingt interpretierbar sein. Sind sie dies, liegt eine [Intervallskala](#) vor. Den Unterschied kann man gut an