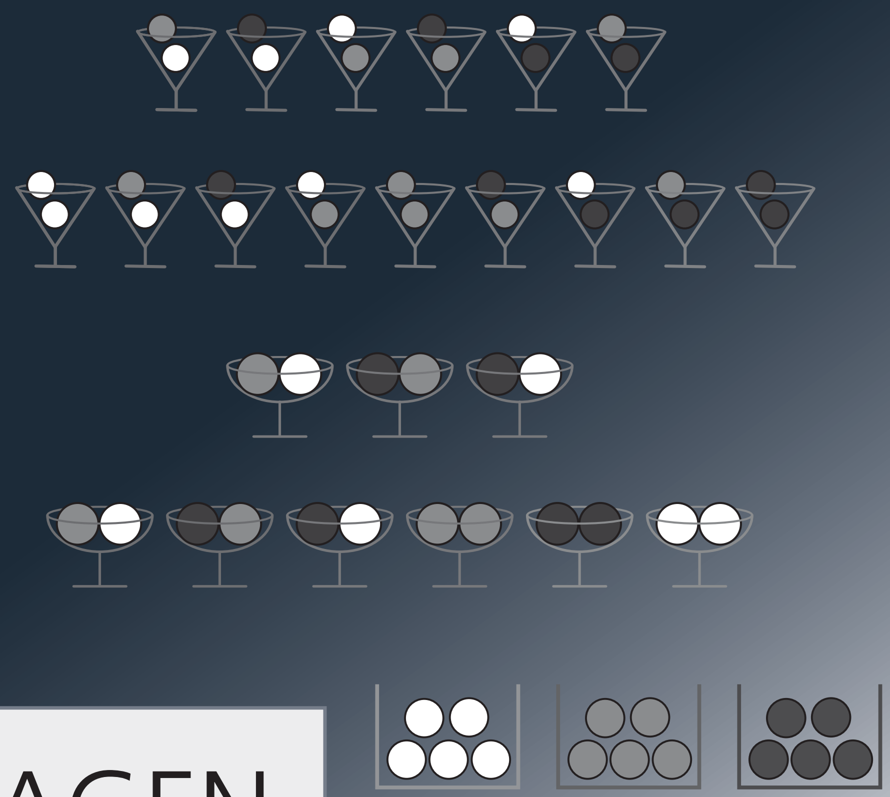
Clauß
Finze
Partzsch



GRUNDLAGEN DER STATISTIK

Für Soziologen, Pädagogen, Psychologen und Mediziner

Edition Harri Deutsch

EUROPA LEHRMITTEL

EUROPA LEHRMITTEL

Edition
Harri
Deutsch

Grundlagen der Statistik

Für Soziologen, Pädagogen, Psychologen und Mediziner

von

G. Clauß
F.-R. Finze
L. Partzsch

7. Auflage

VERLAG EUROPA-LEHRMITTEL · Nourney, Vollmer GmbH & Co. KG
Düsselberger Straße 23 · 42781 Haan-Gruiten

Europa-Nr.: 55422

Die Autoren

Prof. Dr. phil. habil. Günter Clauß †
Dr. rer. nat. Falk-Rüdiger Finze, Technische Universität Dresden
Dr. rer. nat. Lothar Partzsch, Technische Universität Dresden

7. Auflage 2017

Druck 5 4 3 2 1

ISBN 978-3-8085-5543-9

ISBN 978-3-8085-5876-8 (E-Book)

http://www.europa-lehrmittel.de
Satz: Satzherstellung Dr. Naake, 09618 Brand-Erbisdorf
Umschlaggestaltung: braunwerbeagentur, 42477 Radevormwald
Druck: Medienhaus Plump GmbH, 53619 Rheinbreitbach

Vorwort

Die Statistik kann bei der Auswertung empirischer Untersuchungen in der Psychologie, Medizin, Pädagogik, Soziologie und in den angrenzenden Wissenschaften ein hilfreiches methodisches Instrumentarium sein.

Die Autoren vertreten die Meinung, dass auch und gerade in Zeiten moderner Rechentechnik und entsprechender Softwareangebote das Grundverständnis für die Interpretation der Ergebnisse dieser technischen Hilfsmittel voraussetzt, dass man als Nutzer die elementaren Verfahrensschritte einmal kennengelernt und selbst nachvollzogen hat. Nur auf diesem Weg erarbeitet sich der Anwender auch die Möglichkeit, Computerergebnisse angemessen zu interpretieren und kritisch zu hinterfragen.

Um diesen nächsten Schritt, die Nutzung moderner Computersoftware gehen zu können, befindet sich in diesem Buch bei der Berechnung jedes statistischen Parameters und bei jedem vorgestellten inferenzstatistischen Verfahren im Sinne einer Brücke auch ein Verweis darauf, wie man die Lösung der entsprechenden Fragestellung mithilfe von Software-Angeboten realisieren kann.

Das vorliegende Lehrbuch wendet sich in erster Linie an Leser, die die genannten Disziplinen studieren oder auf diesen Gebieten arbeiten, und verfolgt das Ziel, dem Leser in möglichst verständlicher Form die entsprechenden Verfahren vorzustellen und ihn zu deren sachkundiger Anwendung zu befähigen.

Es eignet sich auch zum Selbststudium, da die statistischen Verfahren mit vollständig durchgerechneten Zahlenbeispielen behandelt werden und es auf allgemeinen Schulkenntnissen im Fach Mathematik aufbaut.

Dresden, September 2017 — Falk-Rüdiger Finze, Lothar Partzsch

Geleitwort

Wer sich zum Studium der Psychologie oder der Soziologie entschließt, interessiert sich vor allem für das Erleben und Handeln der Menschen, er will sich und andere beobachten, verborgene Motive ergründen, Ursachen für Konflikte entdecken, seelische Leiden mindern, durch Beratung helfen usw. Mit Mathematik hat er vielfach nichts im Sinne. Sie erscheint ihm oft trocken, lebensfremd und irrelevant für seinen Beruf. Die Vorlesungen zur mathematischen Psychologie und Statistik stoßen daher zunächst überwiegend auf wenig Gegenliebe. Das erlebte ich in vielen Jahren, in denen ich – gemeinsam mit meinem Kollegen Heinz Ebner – Studierende der Psychologie und Pädagogik in die Statistik einführte. Um die emotionalen Barrieren abzubauen, bemühten wir uns, die Inhalte möglichst ansprechend darzustellen, narrensicher zu erklären und ihre Nützlichkeit an praktischen Beispielen eindringlich zu zeigen. Das ist uns offenbar gelungen; denn die Studierenden besuchten unsere Vorlesungen und Übungen gern und erwarben zumeist eine positive Einstellung zur empirischen Methodik, für die statistische Verfahren unverzichtbar sind.

Die positive Resonanz der Lehrveranstaltungen veranlasste uns, ein Lehrbuch zu schreiben. 1967 erschienen die „Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen“ erstmals im Berliner Verlag Volk und Wissen. Eine erweiterte Fassung fand Zugang zu deutschen Universitäten, wurde auch in fremde Sprachen übersetzt und erwarb in einem Vierteljahrhundert das Label „Clauß-Ebner“.

Die Langlebigkeit des Titels dürfte vor allem auf den hohen Grad an Verständlichkeit zurückzuführen sein. Wir setzten beim Leser nur das mathematische Schulwissen voraus, behandelten die Kennwerte und Prozeduren in enger Bindung an praktisch relevante Sachverhalte und gaben stellenweise rezeptartige Handlungsanweisungen, die bei der Anwendung statistischer Verfahren zu beachten sind und dann zwangsläufig zum richtigen Ergebnis führen. Eine solche didaktische Vereinfachung mag manchem Mathematiker klippschulenhaft erscheinen; der Mehrzahl der Leser kam die Redundanz und straffe Lenkung sehr entgegen.

Als das Buch zwei Jahrzehnte lang im Handel war, hielten wir es für angebracht, eine gründliche Neubearbeitung vorzunehmen. Dazu kam es jedoch nicht. Heinz Ebner verstarb plötzlich, und ich glaubte, die mittlerweile erschienene Fachliteratur könne unser Buch ablösen. Jedoch die Nachfrage blieb bestehen. Ein unveränderter Nachdruck kam nicht zustande. Um die Marktlücke rasch zu schließen, gewann der Verlag Harri Deutsch zwei erfahrene Dresdner Autoren, Falk-Rüdiger Finze und Lothar Partzsch. Sie sollten in der Tradition des Clauß-Ebner ein Statistiklehrbuch verfassen, das zum Gebrauch an Universitäten sowie zum Selbststudium geeignet ist, die bewährten Verfahren in verständlicher Form behandelt und durch neue Methoden – vor allem multivariate und parameterfreie – ergänzt.

Das vorliegende Buch ist das Produkt. Ich identifiziere mich uneingeschränkt mit seinem Inhalt und der Art der Darstellung. Es ehrt mich, dass mein Name den Autorennamen beigefügt wird, um dadurch eine gewisse Kontinuität der Lehrbuchentwicklung zu signalisieren. Freilich ist das Werk eine durchaus eigenständige Leistung der Verfasser. Sie nötigen den Leser zum selbstständigen Mitdenken – mehr, als wir ihm das abverlangten, – geben ihm aber klare Anleitung und die Möglichkeit, anhand von Beispielen zu prüfen, ob er den Text richtig verstanden hat und anwenden kann. Während wir in den 60er-Jahren allenfalls die Nutzung einer Tischrechenmaschine empfehlen konnten, stehen dem Leser heute elektronische Taschenrechner und Computer zur Verfügung. Spezielle Statistikrechner enthalten Programme für Signifikanztests und Varianzanalysen. Ihr Einsatz vereinfacht und beschleunigt die Arbeit außerordentlich. Es wäre töricht, von der Nutzung moderner Rechentechnik abzuraten. Sie kann sehr hilfreich sein und befreit von der Mühsal stumpfsinniger Routine. Aber sie ist und bleibt dienendes Hilfsmittel im Prozess wissenschaftlichen Problemlösens. Der Nutzer muss mit Einsicht und Sachverstand entscheiden, ob bei einem gegebenen Datensatz die Anwendung eines bestimmten Prüfverfahrens statthaft ist und welche inhaltliche bedeutsame Frage auf diese Weise beantwortet werden kann. Ein solches eindringendes Verständnis setzt voraus, dass man sich einige mathematische Grundbegriffe und Kernaussagen der Wahrscheinlichkeitstheorie aneignet. Sie werden im 3. und 6. Kapitel behandelt, soweit sie für statistische Methodik unentbehrlich sind. Der Leser, der beim ersten Zugriff diese Kapitel überspringt, tut gut daran, sich ihnen später aufmerksam zuzuwenden. Andernfalls läuft er Gefahr, in Praktizismus abzusinken und ernste Fehler zu begehen.

Ich halte das Buch von Finze und Partzsch für eine gut gelungene Einführung, die den heutigen Ansprüchen genügt und dem gegenwärtigen Entwicklungsstand der psychologischen Statistik entspricht. Möge das Buch dankbare Leser finden und dazu beitragen, dass die Studierenden humanwissenschaftlicher Disziplinen ihre eventuell vorhandene Aversion überwinden und die Mathematik in angemessener Weise dafür nützen, methodisch kontrolliert Hypothesen zu prüfen, Aussagen über gesetzmäßige Zusammenhänge zu sichern und Ungewissheit in Wissen zu verwandeln.

Leipzig, Dezember 1993 — Günter Clauß

Inhaltsverzeichnis

1 Einleitung

Für wissenschaftliche Untersuchungen in der Psychologie, Pädagogik, Soziologie, Medizin und weiteren artverwandten Disziplinen haben statistische Methoden in zunehmendem Maße an Bedeutung gewonnen. Das Anliegen des vorliegenden Lehrbuches besteht darin, einerseits den Leser mit den Methoden der deskriptiven Statistik und darauf aufbauend mit einem angemessenen Fundus an statistischen Verfahren vertraut zu machen und ihm andererseits die wahrscheinlichkeitstheoretischen Grundlagen, die zum Verständnis der Statistik erforderlich sind, zu vermitteln. Entsprechend dieser Zielstellung ist das Buch aufgebaut.

Die Statistik findet man in den Kapiteln 2, 4 und 5. Wir beginnen in Kapitel 2 mit einer umfassenden Darstellung der Methoden der deskriptiven Statistik, die in der Verwendung von tabellarischen Übersichten, grafischen Darstellungen und geeigneten Kennziffern bestehen. Das Kapitel 4 ist das eigentliche Kernstück dieses Buches und befasst sich mit den in der Psychologie, Pädagogik und den Sozialwissenschaften gebräuchlichsten statistischen Testverfahren. Die Struktur dieses Kapitels ist nach dem Charakter der vorhandenen Daten aufgebaut. Die einzelnen statistischen Verfahren werden dadurch vorgestellt, dass weniger die theoretischen Details ausgeführt werden, sondern vielmehr anhand repräsentativer Anwendungssituationen die konkrete Testdurchführung „rezeptähnlich“ vorgestellt und an einem konkreten Rechenbeispiel nachvollzogen wird. Um beim Lösen einer statistischen Problemstellung die Suche nach einem geeigneten Testverfahren zu erleichtern, befinden sich am Ende jedes größeren Abschnittes tabellarische Übersichten. In der heutigen Zeit gibt es eine Vielzahl von Computerprogrammen zur Bearbeitung statistischer Problemstellungen. Die Autoren vertreten den Standpunkt, dass das Verständnis der Statistik wesentlich gefördert wird, wenn der Lernende an ausgewählten, typischen Beispielen die erforderlichen Rechenschritte wenigstens einmal „zu Fuß“ ausgeführt hat.

In den Kapiteln 3 und 6 werden, anknüpfend an das Schulwissen, die notwendigen Grundlagen behandelt. In Kapitel 3 verfolgen wir das Ziel, die wichtigsten Begriffe der Wahrscheinlichkeitsrechnung möglichst einfach und klar zu beschreiben und darauf aufbauend ein angemessenes Verständnis für dieses Gebiet zu entwickeln. Im einzelnen handelt es sich hier im wesentlichen um das Grundmodell der Wahrscheinlichkeitsrechnung und den Begriff der Zufallsgröße. Kapitel 6 enthält in zusammenfassender Form Grundkenntnisse aus der Mengenlehre, über reelle Funktionen sowie aus der Kombinatorik, und gibt eine elementare Einführung in die Matrizenrechnung.

Die multivariate Statistik betrachtet in zunehmendem Maße derart komplexe Fragestellungen, dass für den Anwender die Nutzung des Computers erforderlich wird. Aus diesem Grund wird im Kapitel 5 anhand einfacher Beispiele versucht, eine erste Einführung in die Korrelations- und Regressionsanalyse, die Faktorenanalyse, die Clusteranalyse und die Varianzanalyse zu vermitteln. Damit soll zum Verständnis von Lösungen, die man durch die Benutzung von Computerprogrammen erhält, beigetragen werden. Das abschließende Kapitel 7 enthält eine Zusammenstellung der notwendigen Tabellen.

1.1 Grundanliegen der Statistik

Bei vielen wissenschaftlichen Untersuchungen, insbesondere unter anderem in der Psychologie, Pädagogik, Soziologie und Medizin, hat man es zum Teil mit großen Anzahlen von Daten zu tun, aus denen man nicht mit Sicherheit auf Aussagen über vorhandene Beziehungen und Phänomene schließen kann. Es handelt sich aber häufig um wiederholbare Erfahrungen, so dass man hoffen kann, immer noch Allgemein gültiges, Gesetzmäßiges in derartigen „Massenerscheinungen“ herauszufinden.

Während man z. B. bei einem „gezinkten“ Würfel durch eine einzige Messung sein Gewicht mit ausreichender Genauigkeit zweifelsfrei ermitteln kann, lässt sich aus 10 Wurfergebnissen nicht eindeutig beurteilen, welche Wahrscheinlichkeiten den einzelnen Augenzahlen zugeordnet werden sollen. Es gibt aus prinzipiellen oder pragmatischen Gründen zu viele unkontrollierbare Einflüsse, die eine zuverlässige Aussage über die interessierende Eigenschaft nicht gestatten. Wir sprechen dann davon, dass der „Zufall“ seine Hand im Spiel hat, die Messergebnisse streuen mehr oder weniger unvorhersehbar. Es entstehen gewisse Unschärfeeffekte und Grauzonen. Trotzdem ist man bestrebt, immer noch Allgemein gültiges zu erkennen, eventuelle Gesetzmäßigkeiten zu entdecken, einen Kern in den vielen Grauzonen streuender Werte zu finden. Man möchte gerne „hinter die Kulissen“ schauen, dem Zufall „auf die Finger sehen“, um mehr Sicherheit in der Unsicherheit zu erzielen. Die Statistik liefert hierzu ein geeignetes methodisches Instrumentarium. Sie gibt **Hilfen zur Entscheidung** bei der Auswertung empirischer Daten.

Wir unterscheiden zwischen der beschreibenden (oder auch deskriptiven) und schließenden Statistik. Das Anliegen der deskriptiven Statistik besteht darin, interessierende Daten von großen Anzahlen von Objekten, Personen usw., die man in diesem Zusammenhang auch Grundgesamtheit nennt, anschaulich, übersichtlich und verständlich darzustellen. Dies erfolgt in Listen und Tabellen, in Grafiken oder mithilfe von typischen Maßzahlen wie z. B. Mittelwerten und Streuungen. Dabei ist man bestrebt, auf möglichst umfassenden Erhebungen aufzubauen, wie es beispielsweise bei der Erfassung von Einwohnerzahlen in statistischen Jahrbüchern der Fall ist. Mithilfe der Methoden der deskriptiven Statistik soll eine Datenvoranalyse gefördert werden. Aufgrund deren unmittelbarer Verständlichkeit stellen wir die deskriptive Statistik an den Anfang (vgl. Kapitel 2).

Bei der schließenden Statistik wird im Unterschied dazu auf der Grundlage von Informationen aus einer Teilmenge (auch Stichprobe genannt) der Grundgesamtheit auf Aussagen über die Grundgesamtheit geschlossen. Es werden also Aussagen über den empirischen Beobachtungsbereich hinaus getroffen. Genauer heißt dies, man stellt ein theoretisches Modell auf und vergleicht es mit den empirischen Informationen. Mit diesem theoretischen Modell versucht man, das Allgemein gültige des Zufallsgeschehens dadurch zu repräsentieren, dass man bestimmte Wahrscheinlichkeitsverteilungen des Auftretens der Beobachtungswerte postuliert. Dieser Vergleich erfolgt auf zwei Wegen. Zum einen gibt es die Möglichkeit, die postulierte Wahrscheinlichkeitsverteilung oder auch nur interessierende Parameter näherungsweise zu bestimmen. Entsprechende Verfahren werden in der Schätzstatistik entwickelt. Zum anderen stellt man Hypothesen über die Modellverteilung auf und überprüft diese auf der Grundlage der empirischen Daten der Stichprobe. Dieses große Teilgebiet nennt man Inferenzstatistik (oder auch Teststatistik). So könnte man z. B. die Behauptung prüfen, ob die Leistung nach einem Trainingskurs besser ist als vorher. Es sei aber explizit darauf verwiesen, dass die Richtigkeit von Aussagen wegen der unvollständigen Informationen – es wurde ja nur eine Teilmenge untersucht – nicht mit absoluter Sicherheit, sondern nur mit einer bestimmten Wahrscheinlichkeit garantiert werden kann. Überdies sollte man beim Durchführen von statistischen Tests die zu überprüfende Hypothese nicht erst in Abhängigkeit vom vorliegenden Datenmaterial formulieren. Ausgehend von den typischen Erfordernissen der Psychologie, Soziologie und Pädagogik werden wir uns in diesem Lehrbuch vorwiegend mit der Inferenzstatistik (vgl. Kapitel 4) beschäftigen. Deren Ergebnis besteht in der Beurteilung empirischer Daten.

Eine weitere Differenzierung der schließenden Statistik ist durch den Grad der Komplexität gegeben. Wir unterscheiden in diesem Zusammenhang zwischen uni- und multivariater Statistik. Zu letzterer zählt man auch Analysen komplexerer Systeme, die dann im Sinne von so genannten Datenvoranalysen der explorativen Statistik durchgeführt werden. Kapitel 5 enthält einige Ausführungen über dieses große Gebiet der Statistik.

1.2 Die Relativität statistischer Aussagen

Statistische Aussagen vermitteln, und das wird leider viel zu oft vergessen, immer nur Erkenntnisse über zufallsabhängige Massenerscheinungen und müssen gerade deshalb im Einzelfall nicht zwingend zutreffen. Die Aussage gilt nur für den Bereich **insgesamt**, über den sie gemacht wird – sie ist eine Globalaussage. Stellen wir uns die Situation vor, dass in einem großen Unternehmen der Industrie von **allen** Mitarbeitern die tatsächliche wöchentliche Arbeitszeit bestimmt wurde, und es ergaben sich im Mittel 40 Stunden. Dann gilt die statistische Aussage: „Der Mittelwert der wöchentlichen Arbeitszeit liegt in diesem Unternehmen bei 40 Stunden." nur dann, wenn man vom gesamten Unternehmen spricht. Es ist leicht einsichtig, dass sie bei den einzelnen Arbeitern, Angestellten, Managern usw. verschieden sein kann. Hier variiert sie vielleicht zwischen 38 und 60 Stunden. Es wäre also falsch, aus der statistischen Aussage auf den Einzelfall zu schließen.

Deshalb unterscheiden wir zwischen statistischen und **kasuistischen** Aussagen, d. h., Aussagen über die Grundgesamtheit und Aussagen über den **Einzelfall**. Eine kasuistische Aussage könnte dann z. B. lauten: „Der Betriebsklempner hat eine wöchentliche Arbeitszeit von 42 Stunden." Sie kann demnach nur eine Information darüber liefern, dass ein bestimmtes Element (Person) einer bestimmten Grundgesamtheit (Unternehmensmitarbeiter) ein bestimmtes Merkmal (wöchentliche Arbeitszeit) in einer bestimmten Ausprägung (42 Stunden) aufweist. Es besteht die Möglichkeit, aus einer Vielzahl kasuistischer Aussagen eine statistische Aussage abzuleiten.

Hervorhebenswert ist an dieser Stelle, dass der Unterschied zwischen einer kasuistischen und einer statistischen Aussage immer nur **relativ** ist. Erheben wir z. B. die wöchentliche Arbeitszeit jedes Mitarbeiters und bilden davon pro Abteilung Mittelwerte, dann sind die Aussagen zur Abteilungsarbeitszeit statistische Aussagen. Erheben wir nun aber nur die Arbeitszeit pro Abteilung und bilden daraus die mittlere Unternehmensarbeitszeit, dann sind die Aussagen zur Abteilungsarbeitszeit kasuistische Aussagen. Ein und dieselbe Aussage kann demnach einmal eine kasuistische und einmal eine statistische Aussage sein. Die Entscheidung darüber, welche Art einer Aussage vorliegt, hängt also immer vom einzelnen Betrachter und von den Randbedingungen der Untersuchung und Interpretation ab.

1.3 Zur Anwendung der Statistik in der Psychologie

Die Psychologie ist eine Wissenschaft, die sich mit dem Menschen, seinen Fähigkeiten und Fertigkeiten, seinen Motiven und Zielen, seinen Emotionen und Gefühlen, aber auch mit seinen psychischen Störungen und Beeinträchtigungen auseinandersetzt. Entsprechend breit sind die Methoden wissenschaftlicher Analyse und Synthese gefächert. Die Statistik ist **eine** Möglichkeit, Ergebnisse gegen zufällige Einflüsse und Schwankungen zu sichern, systematische Unterschiede aufzudecken und im Rahmen ihrer Grundannahmen zu verifizieren oder zu falsifizieren. Um die Statistik aber sowohl im Einzelfall als auch insgesamt sinnvoll einsetzen zu können, müssen die Daten, die statistisch weiter verarbeitet werden sollen, einige Bedingungen erfüllen, auf die wir nachfolgend näher eingehen wollen.

1.3.1 Forderungen an empirische Daten

Die Statistik, d. h., die in ihr genutzten mathematischen Modelle, benötigt im weitesten Sinne quantifizierbare Daten. Es besteht die Notwendigkeit, dass sich die empirischen Daten und Beobachtungen numerisch, d. h. durch Zahlen, kennzeichnen lassen. Häufig treffen wir aber, speziell im Einzelfall (z. B. der Persönlichkeitspsychologie), auf ausschließlich qualitative Angaben. In solchen Situationen, wie etwa beispielsweise bei einem qualitativen Merkmal wie „soziale Herkunft“, benötigen wir wenigstens als quantifizierbare Daten die beobachteten Häufigkeiten der einzelnen „Stufen“ des Merkmals.

Eine weitere, notwendige Voraussetzung besteht in einer exakten Definition dessen, was gemessen werden soll. In den so genannten „exakten Wissenschaften“ wie Physik oder Chemie ist das sicher leichter als in den Human- und Sozialwissenschaften. Betrachten wir als Beispiel das Merkmal „soziale Herkunft“. Hier sind verschiedene Kategoriensysteme wie etwa Arbeiter, Angestellter, Bauer, Angehöriger der Intelligenz, Unternehmer usw. vorstellbar. Will man aber beispielsweise wissenschaftliche Aussagen aus dem Vergleich verschiedener Untersuchungen ableiten, dann hat das nur einen Sinn, wenn die Inhalte der untersuchten Kategorien vergleichbar, also genau definiert sind. Welche Berufe zählen wir nun zum Beispiel zur Kategorie „Arbeiter“? Der Anwender der Statistik sollte also auch Definitionen hinterfragen.

Schließlich sollen die empirischen Daten durch geeignete Messverfahren erhoben worden sein. Diese Forderung ist nicht neu und wohl in allen Wissenschaften gleich. Dennoch können wir nicht a priori davon ausgehen, dass sie immer Berücksichtigung findet. Misst man beispielsweise bei einem Weitsprungwettbewerb die erreichten Weiten mithilfe eines Maßbandes aus Gummi, dann würde man offensichtlich an der Seriosität der Ergebnisse zweifeln. Aus diesem Grunde sollten wir, die Anwender, stets darauf achten, dass wenigstens nachfolgende 3 Forderungen an die verwendeten Messverfahren erfüllt sind:

1. Das Kriterium der Objektivität:
   Damit ist gemeint, dass das jeweilige Messverfahren unabhängig vom Anwender sein soll. Auch wenn ein anderer Versuchsleiter oder Befrager die Analyse durchführt, sollte dies ohne Einfluss bleiben. Der interessierte Leser findet in der psychologischen Fachliteratur dazu viele Hinweise unter dem Schlagwort „Versuchsleiterfehler“.
2. Das Kriterium Reliabilität (auch Zuverlässigkeit):
   Hierbei geht es darum, dass das Messverfahren reproduzierbar, d. h. wiederholbar sein muss. Jetzt wird der eine oder andere Leser einwenden, dass z. B. gerade so genannte projektive Tests in der Psychologie diesem Anspruch nicht uneingeschränkt genügen. Das stimmt – aber der qualitative Einzelfall kann ohnehin nicht statistisch ausgewertet werden.
3. Das Kriterium Validität (auch Gültigkeit):
   Die eingesetzten Messverfahren müssen wirklich das messen, was sie messen sollen und vorgeben zu messen. Es ist eben beispielsweise unsinnig, mithilfe der Lösung eines Kreuzworträtsels die Intelligenz eines Probanden messen zu wollen. Auch wenn dieses Beispiel sicher etwas drastisch ist, so verdeutlicht es doch das mitunter zu unkritische Umgehen mit dem Problem der Validität.

Bemerkung: *Die Anwendung statistischer Verfahren setzt voraus, dass diese Forderungen erfüllt sind – wobei unterstellt wird, dass es ein wissenschaftstheoretisch weitestgehend ungeklärtes Validitäts-Reliabilitäts-Dilemma gibt.*

1.3.2 Vorteile und Grenzen beim Einsatz der Statistik

Wie bereits im Abschnitt 1.2 festgestellt wurde, stellt die Statistik ein methodisches Instrumentarium zur wissenschaftlichen Analyse empirischer Daten bereit. Wir wollen in diesem Abschnitt zusammenfassend sowohl einige wesentliche Vorteile für ihren Einsatz benennen als auch auf Einschränkungen kritisch aufmerksam machen.

1. Vorteile:
   - Die deskriptive Statistik bietet uns die Möglichkeit der Präzision, d. h. der genauen Beschreibung der Beobachtungen und deren Zusammenfassung. Damit können wir unüberschaubare Mengen detaillierter Daten anschaulich darstellen und weiterverarbeiten.

- Durch den Einsatz der schließenden Statistik können wir die auf der Grundlage von Stichproben gewonnenen Aussagen auf die gesamte zugehörige Grundgesamtheit verallgemeinern.
- Wir sind aufgrund der verwendeten mathematischen Modelle in der Lage, selbst Aussagen zur Genauigkeit und zur Sicherheit der festgestellten Schlussfolgerungen zu treffen.
- Es besteht die Möglichkeit, Aussagen sowohl theoretisch als auch empirisch zu überprüfen.
- Schließlich können wir durch den Gebrauch mathematischer Methoden auf der Grundlage der empirisch gewonnenen Daten weitere Modellrechnungen durchführen, d. h., wir können ohne ökonomische, soziale oder andere Konsequenzen weitere Fallbeispiele exemplarisch durchspielen.

2. Kritisch zu beachten sind:
   - Wie immer man die Statistik auch einsetzt, sie liefert niemals Aussagen zur **inhaltlichen Bedeutsamkeit** der durchgeführten Untersuchung. Die Verantwortung dafür liegt ausschließlich beim Anwender.
   - Die Statistik liefert für den Untersuchungsansatz und die -durchführung keine Kriterien darüber, **welche Beobachtungsgrößen** zu verwenden sind. Also selbst wenn das Kreuzworträtsel zur „Messung der Intelligenz" missbraucht wird, kann die Statistik darauf keinen Einfluss nehmen.
   - Auch die Frage, welches Messverfahren im Einzelfall zu verwenden ist, kann mit den Mitteln der Statistik nicht beantwortet werden.
   - Die Statistik kann eine inhaltliche Interpretation nicht leisten. Beispielsweise werden zwar Zusammenhänge und Abhängigkeiten mathematisch ermittelt, aber die Einschätzung ihrer Bedeutung, etwa die Unterscheidung zwischen Ursache und Wirkung, obliegt dem Benutzer.
   - Jede Anwendung von statistischen Verfahren ist abhängig von bestimmten Voraussetzungen. Für deren Beachtung ist der Anwender verantwortlich, d. h., wenn man bei vorliegendem empirischen Datenmaterial den falschen Parameter berechnet oder den falschen Test einsetzt, dann kann sich die Statistik nicht dagegen „wehren".

2 Deskriptive Statistik

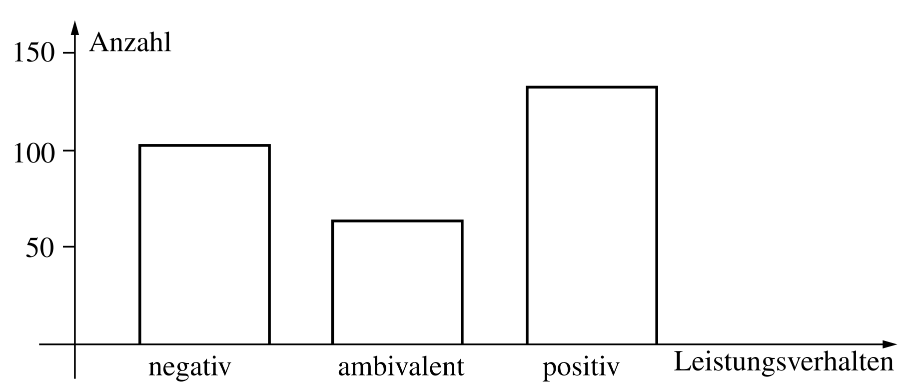
Bei der Anwendung statistischer Methoden im Rahmen psychologischer, pädagogischer, soziologischer oder medizinischer Untersuchungen fallen häufig große Datenmengen an. Die beschreibende Statistik verfolgt das Ziel, diese in geeigneter Weise darzustellen, um damit Übersichtlichkeit zu erreichen, Vergleiche zu ermöglichen und das Erkennen eventueller „Gesetzmäßigkeiten“ zu erleichtern. Die mit ihrer Hilfe gewonnenen empirischen Erkenntnisse können dann durch geeignete Prozeduren der schließenden Statistik, insbesondere der Prüfstatistik, theoretisch gesichert werden. Mittel der beschreibenden Statistik sind Tabellen, Grafiken und die Angabe von Kennwerten der durch die Datenmenge erzeugten Verteilungen.

Eine erste Einteilung der Methoden der beschreibenden Statistik berücksichtigt die **Art der Daten**, die sich auf der Grundlage der verwendeten Skalen ergibt. Wir unterscheiden zwischen Nominalskalen (z. B. Berufe), Ordinalskalen (z. B. Stärken von Stürmen), Intervallskalen (z. B. Celsius-Skala der Temperaturmessung) und Absolutskalen (z. B. Kelvin-Skala der Temperaturmessung). Wir werden uns im Abschnitt 2.1 mit dieser Klassifikation nach Datentyp ausführlich beschäftigen.

Ein zweiter Gesichtspunkt betrifft die **Anzahl der gleichzeitig untersuchten** Variablen. Im einfachsten Fall wird nur eine Variable untersucht, z. B. das Leistungsverhalten von Jugendlichen in der Berufsausbildung. Dies führt zu so genannten **monovariablen** Verteilungen, die wir im zweiten Teil dieses Kapitels behandeln wollen. Man könnte bei der Befragung von 300 Jugendlichen beispielsweise nachfolgende Ergebnisse erhalten haben:

Leistungsverhalten	Anzahl n_i
positiv	133
ambivalent	64
negativ	103
Σ	300

Eine grafische Darstellung, das so genannte Histogramm, sieht dann wie folgt aus:



Außer der bildlichen Darstellung hat man die Möglichkeit, charakteristische Kennziffern wie Mittelwerte und Streuwerte zur Beschreibung und zum Vergleich monovariabler Verteilungen zu benutzen.

Im Unterschied dazu führt die gleichzeitige Betrachtung von 2 zu untersuchenden Größen zu **bivariablen** Verteilungen. Würde z. B. der Zusammenhang zwischen Leistungsverhalten und Erfüllung des Berufswunsches den Schwerpunkt der Untersuchung bilden, so hätte man diese beiden Variablen gemeinsam zu berücksichtigen. Im obigen Zahlenbeispiel ist dann das Verhalten der 2. Variable zu ergänzen, was zu nachfolgender Tabelle führen könnte:

	Berufswunsch erfüllt	teilweise erfüllt	nicht erfüllt	$\sum$
positiv	70	43	20	133
ambivalent	19	24	21	64
negativ	12	30	61	103
$\sum$	101	97	102	300

Eine entsprechende grafische Darstellung der Häufigkeiten erfolgt dann in einem dreidimensionalen Koordinatensystem. Wir wollen uns im letzten Teil dieses Kapitels mit bivariablen Verteilungen beschäftigen und dabei auch Kennziffern des Zusammenhanges, so genannte Korrelationskoeffizienten, diskutieren.

2.1 Arten der Daten

2.1.1 Das Messen

Der Begriff des Messens ist uns aus dem Alltagsverständnis und aus der bisherigen Ausbildung zweifelsfrei verständlich. **MESSEN ist die Zuordnung von Zahlen zu Beobachtungen durch den Vergleich mit einer Maßeinheit.** Dazu fallen uns sofort entsprechende Situationen ein, beispielsweise die Temperaturmessung in der Physik. Eine solche physikalische Messung unterliegt bestimmten Eigenschaften, nämlich:

1. Es gibt eine exakte Definition. Eine bestimmte Temperatur führt zu einer wohldefinierten Ausdehnung z. B. von Quecksilber und wird dadurch gegenständlich ablesbar. Es liegt also eine definierte Messvorschrift zugrunde, die in der Zuordnung einzelner Zahlen zur Länge z. B. der Quecksilbersäule besteht.
2. Wir verfügen über ein objektives Messgerät, das Thermometer.
3. Uns steht eine festgelegte Einheit als Maßeinheit, z. B. Grad Celsius, zur Verfügung.
4. Für die Temperatur gibt es einen „absoluten Nullpunkt" bei 0 K oder $-273{,}15$ °C.

Sind die oben genannten Voraussetzungen erfüllt, dann sprechen wir vom **Messen im engeren Sinne**. Im Ergebnis eines solchen Messvorganges erhalten wir Zahlen relativ zu einer Maßeinheit. Wir sprechen dann von metrischen Daten und unterscheiden dabei noch zwischen Skalen mit Nullpunkt (so genannte Absolutskalen) und Skalen ohne einen solchen (so genannte Intervall- bzw. Differenzskalen). So entspricht z. B. die Angabe des Geburtsjahres einer Versuchsperson einer Intervallskala, das Alter jedoch einer Verhältnisskala.

In der Psychologie, Soziologie, Pädagogik usw. gibt es nun aber auch für viele Untersuchungsgegenstände keine so klar definierten Bedingungen. Aus diesem Grund ist eine Verallgemeinerung und Erweiterung des Messbegriffes erforderlich, weil eine unabdingbare Voraussetzung für den Einsatz statistischer Methoden **quantifizierbare** Merkmale sind. Die einfachste Vorgehensweise der Quantifizierung besteht im Zählen. Wir zählen beispielsweise ab, wie häufig eine bestimmte, uns interessierende Beobachtung auftritt. Das ergibt sich immer dort, wo wir Ereignisse, Beobachtungen oder Objekte im engeren Sinn nicht messen, sondern nur klassifizieren können, d. h., wir ordnen bestimmte Beobachtungen oder Ereignisse einzelnen, wohldefinierten Klassen zu. Es entstehen dann Häufigkeiten als quantitative Daten. In diesen Situationen verfügen wir also nicht über eine Maßeinheit und erhalten Messergebnisse im Sinne von Zählergebnissen durch eine exakte Beschreibung der beobachteten Klassen. Beispiele sind die Zuordnung des Geschlechts oder der Augenfarbe bei Versuchspersonen. Noch mehr Informationen erhalten wir, wenn zwischen den Klassen in sinnvoller Weise ein Ordnen möglich ist. Wir können dann auch noch je zwei Untersuchungsobjekte vergleichen. Eine solche Situation liegt z. B. vor, wenn wir Untersuchungen zur Sympathiestruktur in einer Gruppe durchführen.

Wegen dieser Sachlage ist es notwendig, den uns bislang bekannten Messbegriff zu erweitern. Nach **Stevens** (1951) definieren wir: Messen entspricht der Zuordnung von Zahlen zu Beobachtungen/Objekten nach bestimmten Regeln. Wir sprechen vom **Messen im weiteren Sinn**. Ein so verstandener Messbegriff schließt einerseits das Messen im engeren Sinn ein, umfasst aber andererseits auch das Ordnen und das Klassifizieren. Wenn wir nachfolgend vom Messen sprechen, dann meinen wir Messen im weiteren Sinne.

Allerdings soll an dieser Stelle keine Fortsetzung des großen Theorienstreites darüber erfolgen, ob psychologisch, soziologisch, pädagogisch usw. relevante Sachverhalte durch Zahlen abbildbar sind oder ob sich in diesen Disziplinen überhaupt gesicherte wissenschaftliche Aussagen nur durch Zahlen belegen lassen. Zahlen können ein geeignetes Hilfsmittel sein. Die alleinige quantitative Analyse führt im Allgemeinen nicht zwingend zur Wahrheit. Es gibt in der gesellschaftlichen und wissenschaftlichen Praxis Beispiele, in denen sowohl „rein quantitative“ als auch „rein qualitative“ Vorgehensweisen falsche Schlussfolgerungen ergaben. Dieser Gefahr müssen wir uns stets bewusst sein.

Die Statistik ist eine Möglichkeit, einen Beitrag zur gesicherten Wissensgewinnung zu leisten. Entscheidende Voraussetzung ist dann allerdings, dass eine **gegenstandsangemessene Quantifizierung** stattfindet. Ob eine Untersuchung und deren Ergebnis relevant ist, das hängt nicht davon ab, ob Statistik zum Einsatz kam oder nicht, sondern ob eine Fragestellung bearbeitet wurde, die bisher bekanntes Wissen erweitert, aus diesem Grund Einfluss auf die Theorie nimmt und uns schließlich zur besseren Bewältigung unserer täglichen Aufgaben befähigt.

2.1.2 Klassifikation der Skalen

Im vorangegangenen Abschnitt hatten wir das Messen im engeren und im weiteren Sinne kennengelernt. Die für die Psychologie, Soziologie, Pädagogik usw. notwendige Erweiterung des Messbegriffs hat nun Konsequenzen auf die weitere Verarbeitung des empirischen Materiales. In Abhängigkeit davon, ob die untersuchten Merkmale nämlich im engeren Sinne gemessen, geordnet oder nur klassifiziert wurden, verwenden wir unterschiedliche Skalen. Eine Skala ist ein Ausdruck für die, wie auch immer, systematische Einteilung von Beobachtungen aus unserer Umwelt. Wir alle kennen die Skalen an Fieberthermometern, auf Bandmaßen, Uhren usw., d. h. auf Messinstrumenten. Häufig treffen wir aber auch auf Skalen in anderen Formen, bei denen der Charakter der Skala nicht so deutlich in den Vordergrund tritt. Denken wir z. B. an Tabellen für die Beurteilung von Windstärken (etwa schwach /mittel/stark) oder an Möglichkeiten zur Einordnung nach dem Geschlecht (etwa männlich/weiblich). Wir kennen vier große, voneinander unterscheidbare Skalen, die nachfolgend näher beschrieben werden sollen. Dabei stellt der Prozess der Zu- und Einordnung der Daten zu einer/in eine Skala die Skalierung dar. Es wird unterschieden zwischen der Absolut- (oder auch Verhältnis-) Skala, der Intervallskala, der Ordinalskala und der Nominalskala. Diese Skalen sind die Grundlage für die Quantifizierung der Daten.

2.1.2.1 Nominalskalen

Mithilfe einer Nominalskala werden meist qualitative Merkmale (z. B. Farben) dargestellt. Klassifizieren heißt hier, dass eine Beobachtung, ein Ereignis oder ein Objekt einer bestimmten Klasse zugeordnet wird, und wir zählen dann, wie viele der zu untersuchenden Objekte in jeder Klasse enthalten sind. Nominalskalen verwenden wir also, wenn eine Klassifikation noch möglich ist. **Klassifizieren ist das Zuordnen eines Objektes zu einer von mehreren genau definierten Klassen.** Beispielsweise erfolgt eine derartige Zuordnung zu den Geschlechtskategorien „männlich“ oder „weiblich“ bzw. zu den Farbkategorien „rot“, „grün“, „gelb“ usw. Die Kategorien untereinander können wir nicht mehr begründet ordnen, d. h., wir können zum Beispiel nicht sagen, welche Kategorie an erster, zweiter oder auch letzter Stelle angeordnet werden müsste. Deshalb ist die Reihenfolge der Aufzählung ohne jede Bedeutung, d. h. ohne Wertung. Ob im obigen Beispiel „rot“, „grün“, „gelb“ oder eben „gelb“, „rot“, „grün“ aufgezählt wird, das spielt keine Rolle. Bezogen auf die Skala enthält die Reihenfolge der qualitativen Ausprägungen keine Information für uns.

Entscheidend für den Gebrauch einer Nominalskala ist die Möglichkeit der genauen inhaltlichen Beschreibung oder Kennzeichnung derjenigen Merkmale, die zur Klassenbildung führen. Nur wenn wir die Klassen – auch Kategorien genannt – hinsichtlich mindestens nachfolgender Eigenschaften charakterisieren können, führt die Quantifizierung (also das Abzählen) zu einer Nominalskala:
1. Was ist das Gemeinsame aller Elemente einer Klasse?
2. Worin unterscheiden sich die Klassen voneinander?

Die Klassen selbst können entweder durch Ziffern oder durch Begriffe bezeichnet werden.

Wir können in zwei oder in mehr als zwei Klassen unterscheiden. Für den ersten Fall trennen wir zwischen „Element der Klasse A“ und „Nicht-Element der Klasse A“. Eine derartige Zweiteilung nennt man auch **dichotom**, z. B. „Raucher“ und „Nicht-Raucher“. Die Aufteilung kann aber auch in mehr als zwei Klassen erfolgen. Denken wir beispielsweise an verschieden Kraftfahrzeuge, so fällt die Einteilung in Klassen wie „Lastkraftwagen“, „Personenkraftwagen“, „Sportwagen“, „Jeeps“ usw. relativ leicht. Ohne große Probleme sind wir in der Lage, die „Klassenkennzeichen“ und die Unterschiede zu anderen Klassen zu definieren.

Wir sollten uns bei der Entscheidung über die Anzahl der Kategorien von inhaltlichen Gesichtspunkten leiten lassen. Letztlich ist es die untersuchte Fragestellung, die uns Informationen darüber liefert, unter welchen Gesichtspunkten die zu zählenden Objekte oder Gegenstände zusammenzufassen sind. Zu beachten ist dabei nur, dass sich die Klassen nicht überschneiden dürfen und dass alle auftretenden Beobachtungen eindeutig einer Klasse zugeordnet werden können. Damit entsteht aber auch die Notwendigkeit, die Weite oder Enge des gruppenbildenden Kriteriums bei der Interpretation zu berücksichtigen.

Im Bereich der Nominalskala besteht die Form der Quantifizierung im einfachen Auszählen der Häufigkeit, d. h., wie viele Objekte fallen in jede der genannten Klassen. Beim Zählen bedienen wir uns der so genannten Kardinalzahlen. Dem Zahlencharakter von Nominalskalen entsprechen Nominalzahlen. Sie berücksichtigen nur die Verschiedenheit. Nun brauchen wir nur noch abzuzählen, wie viele Objekte oder Ereignisse in den jeweiligen Klassen enthalten sind. Diese Anzahlen sind dann die Häufigkeiten, die wir zur statistischen Weiterverarbeitung verwenden. Nominalskalen enthalten nur die Information, ob jeweils zwei Objekte gleich oder nicht gleich sind (z. B. bei der Aufteilung in „männlich“ und „weiblich“), also eine Aussage zur **Verschiedenheit**. Die Invarianz (Unveränderlichkeit) besteht hier darin, dass die Aussage der Verschiedenheit bei einer Permutation der Klassen (d. h. bei einer entsprechenden Umbenennung der Klassen) unverändert bleiben muss.

2.1.2.2 Ordinalskalen

Wir wollen uns den Begriff der Ordinalskala an einem Beispiel verdeutlichen und erinnern uns dazu noch einmal an die im vorhergehenden Abschnitt getroffenen Aussagen zu den Nominalskalen. Wenn wir z. B. eine Reihe verschiedener Erdbeersorten zu vergleichen haben, dann liegt damit zunächst eine Klassifikation mithilfe der Nominalskala „Erdbeersorte“ vor. Betrachten wir nun zusätzlich neben dem Namen der jeweiligen Erdbeersorte noch deren Geschmack, dann verfügen wir über eine weitere Information. Und obwohl es weder ein geeichtes Messinstrument noch eine definierte Maßeinheit für den Geschmack von Erdbeeren gibt, sind wir in der Lage, mehrere Erdbeersorten nach ihren Geschmackseigenschaften in eine Rangreihe (in eine Reihe der „Bevorzugung“) zu bringen. Ähnliche

Situationen liegen beispielsweise bei Abstufungen von Verbrennungsgraden bei Brandverletzungen vor. Im Unterschied zur Nominalskala ist also eine Ordinalskala „informativer“.

Ordinalskalen sind dadurch ausgezeichnet, dass sie ein sinnvolles Ordnen der Beobachtungen ermöglichen. Die Zuordnungsvorschrift besteht imVergleichen, d. h. einer Klassifikation des Unterschiedes in z. B. „kleiner“, „gleich“ und „größer“. Ein Beispiel für ein solches Ordnungsprinzip sind Urteile beim Vergleich zweier visueller Reize hinsichtlich der Stärke. Jeder Sehende ist in der Lage, die Helligkeit einer Lampe mit einer Leistung von 25 Watt mit der Helligkeit einer Lampe von 1000 Watt zu vergleichen und zu beurteilen, welches Licht heller ist. In der Folge solcher Aussagen entstehen Rangreihen über Ausprägungsgrade bestimmter Merkmale an einer Anzahl von Objekten.

Wichtig für Ordinalskalen ist die Tatsache, dass sie keine definierte Maßeinheit erfordern. Was inhaltlich mit einer Ordinalskala erfasst werden kann, bezieht sich immer auf den gleichen Sachverhalt (z. B. die Sympathie eines bestimmten Politikers, die Helligkeit von Lampen oder den Geschmack von Erdbeeren), der in unterschiedlicher Ausprägung auftreten oder vorliegen kann. Solche Ausprägungen können die Stärke, die Intensität, die Größe usw. sein. Es ist nicht festgelegt, wie groß die Unterschiede zwischen den verschiedenen Merkmalsausprägungen sind. Ob z. B. auf einer Ordinalskala der Form „sehr klein“ – „klein“ – „mittel“ – „groß“ – „sehr groß“ die Abstände zwischen „sehr klein“ und „klein“ bzw. zwischen „mittel“ und „groß“ die gleiche absolute Differenz haben, ist nicht definiert. Mit anderen Worten heißt das, dass die Distanzen der einzelnen Skalenpunkte zueinander bei einer Ordinalskala überhaupt nicht festgelegt sind. Darin besteht der Hauptmangel dieser Skala. Es sollten aber auch hier durchaus klare Definitionen darüber vorliegen, unter welchen Bedingungen wir von einer „sehr kleinen“, einer „kleinen“, einer „mittleren“ usw. Ausprägung sprechen.

Dem Zahlencharakter von Ordinalskalen entsprechen **Rangplätze** (Ordinalzahlen). Die Folge der einzelnen Rangplätze (z. B. Rangplatz 1, Rangplatz 2, Rangplatz 3 usw.) ergibt dann die so genannte **Rangreihe**. Sie liefert uns eine Aussage zur Reihenfolge und zur Verschiedenheit. Die in ihnen enthaltene Information bezieht sich auf die **Verschiedenheit** und auf die **Art der Verschiedenheit** von je 2 Objekten. Damit ist gemeint, dass wir nicht nur feststellen können, dass sich zwei Objekte voneinander unterscheiden (wie bei der Nominalskala), sondern wir können auch noch eine Aussage darüber treffen, welches der beiden Objekte – je nach Definition – vor dem anderen Objekt zu platzieren ist. Bei so genannten ordnungserhaltenden Skalenänderungen (wie sie sich durch Anwenden monoton wachsender Transformationen ergeben, vgl. auch Abschnitt 6.2) bleiben diese Informationen, d. h. die Rangreihe, unverändert erhalten.

2.1.2.3 Intervallskalen

Wir hatten eingangs dieses Abschnittes festgestellt, dass wir aus inhaltlichen Gründen den klassischen Messbegriff der Physik für Belange der Psychologie, der Soziologie, der Pädagogik usw. erweitern müssen. Nun sind wir in unserer Betrachtung an einen Punkt gelangt, wo wir uns wieder an die klassische Definition erinnern sollten, d. h., es geht nachfolgend um das Messen im engeren Sinn.

Intervallskalen liegen vor, wenn wir Merkmale messen. Die Zuordnungsvorschrift ist dementsprechend der Vergleich mit einer Maßeinheit, die aber keinen absoluten Nullpunkt besitzt. Ein Beispiel dafür ist die Temperaturmessung auf der Celsius-Skala . Bei Intervallskalen sind also die **Differenzen** (Intervalle) zwischen zwei beliebigen, aufeinander folgenden Werten der Skala immer gleich groß. Deshalb können wir genauere Aussagen als bei Ordinalskalen machen. Wenn uns also nicht nur interessiert, ob ein Unterschied besteht und welcher Art ein Unterschied zwischen zwei Objekten ist, sondern auch, wie groß die Differenz ist, dann benutzen wir Intervallskalen.

Allerdings sei in diesem Zusammenhang anhand eines Beispiels auf eine Gefahr hingewiesen, die bei der Skalierung mithilfe von Intervallskalen auftreten kann: Leistungen in der Schule werden mit den Noten Eins bis Sechs bewertet. Diese Skala wird oft als Intervallskala interpretiert, d. h., es werden