

utb.

Elisabeth Steiner  
Michael Benesch

# Der Fragebogen

Von der Forschungsidee  
zur SPSS-Auswertung

6. Auflage

Mit  
Datensatz  
zum üben



## **Eine Arbeitsgemeinschaft der Verlage**

Böhlau Verlag • Wien • Köln • Weimar  
Verlag Barbara Budrich • Opladen • Toronto  
facultas • Wien  
Wilhelm Fink • Paderborn  
Narr Francke Attempto Verlag / expert verlag • Tübingen  
Haupt Verlag • Bern  
Verlag Julius Klinkhardt • Bad Heilbrunn  
Mohr Siebeck • Tübingen  
Ernst Reinhardt Verlag • München  
Ferdinand Schöningh • Paderborn  
transcript Verlag • Bielefeld  
Eugen Ulmer Verlag • Stuttgart  
UVK Verlag • München  
Vandenhoeck & Ruprecht • Göttingen  
Waxmann • Münster • New York  
wbv Publikation • Bielefeld  
Wochenschau Verlag • Frankfurt am Main

Elisabeth Steiner / Michael Benesch

# Der Fragebogen

Von der Forschungsidee zur SPSS-Auswertung

6., aktualisierte und überarbeitete Auflage

**facultas**

**Elisabeth Steiner**, Prof. (FH), DSA, Mag.a Dr.in, ist Dipl. Sozialarbeiterin und Klinische und Gesundheitspsychologin. Sie ist Studiengangsleiterin des Masterstudiengangs

„Sozialraumorientierte und Klinische Soziale Arbeit“ an der FH Campus Wien.

Ihre Forschungsschwerpunkte liegen in der Sozialarbeitswissenschaft.

Kontakt: [elisabeth.steiner@fh-campuswien.ac.at](mailto:elisabeth.steiner@fh-campuswien.ac.at)

**Michael Benesch**, Dr., ist Wirtschaftspsychologe und Geschäftsführer der M. Benesch Unternehmensberatung. Er ist als Trainer und Berater in der Organisationsentwicklung und sozialwissenschaftlichen Forschung tätig sowie Lehrbeauftragter an mehreren österreichischen Universitäten und Fachhochschulen. Sein Spezialgebiet ist die

Verbindung empirisch-quantitativer mit qualitativen Informationen unter Anwendung der Dialogischen Kommunikation nach David Bohm und Martin Buber. Nähere

Informationen unter [www.benesch.co.at](http://www.benesch.co.at).

Über facultas bei UTB außerdem erschienen:

Michael Benesch: Der Dialog in Beratung und Coaching, 2020 (UTB)

Michael Benesch, Elisabeth Steiner: Klinische Studien lesen und verstehen,

2. Auflage, 2018 (UTB)

Die Lehrbeispiele in diesem Buch wurden mit der SPSS-Version 26 statistisch ausgewertet.

Den **Übungsdatensatz** zum Buch finden Sie unter <https://www.utb-shop.de> beim Zusatzmaterial zu diesem

Buch zum kostenlosen Download.

**Bibliografische Information der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Alle Angaben in diesem Fachbuch erfolgen trotz sorgfältiger Bearbeitung ohne Gewähr,

eine Haftung der Autorin, des Autors oder des Verlages ist ausgeschlossen.

6., aktualisierte und überarbeitete Auflage 2021

Copyright © 2021 Facultas Verlags- und Buchhandels AG

facultas Universitätsverlag, Stolberggasse 26, 1050 Wien, Österreich

Alle Rechte, insbesondere das Recht der Vervielfältigung und der Verbreitung sowie der Übersetzung, sind vorbehalten.

Umschlagfoto: „Excellent Performance © istockphoto/bluestocking

Lektorat: Verena Hauser, Wien

Satz: Facultas Verlags- und Buchhandels AG

Einbandgestaltung: Atelier Reichert, Stuttgart

Druck und Bindung: Friedrich Pustet, Regensburg

Printed in Germany

utb-Nummer 8406

ISBN [978-3-8252-8788-7](#) (Print-Ausgabe)

ISBN 978-3-8385-8788-2 (Online-Leserecht)

ISBN 978-3-8463-8788-7 (E-PUB)

# Vorwort

Vor mittlerweile zwölf Jahren, im Oktober 2008, erschien die erste Auflage unseres Buches, das die Erstellung von Fragebögen und die Herangehensweise an einfachere statistische Auswertungen mit SPSS für Anfänger\*innen zum Inhalt hatte. Dabei wird der Forschungsprozess von der Idee bis zur statistischen Auswertung und Berichterstellung vermittelt, um eine Grundlage für die weitere Beschäftigung mit dem Thema zu schaffen.

Die Idee zu diesem Buch entstand im Zuge unserer langjährigen Lehrtätigkeit in den unterschiedlichsten Bereichen und der dabei gewonnenen Erfahrungen in der Vermittlung statistischer Grundkenntnisse an Einsteiger\*innen. Dabei konnten wir immer wieder eine wesentliche Beobachtung machen, nämlich die, dass eine eher intuitive, auf „Alltagsverständnis“ aufbauende Herangehensweise, welche auf formalistische Zugänge weitestgehend verzichtet, von den Studierenden sehr geschätzt wird und das Interesse am Fach fördert.

Das didaktische Konzept des Buches hat sich auch in der vorliegenden 6. Auflage nicht geändert, wenngleich es zu einigen kleineren Anpassungen im Inhalt gekommen ist, etwa was das Statistikprogramm SPSS (in dieser Auflage auf Basis von SPSS 26) betrifft. Die Rückmeldungen von Studierenden ebenso wie Lehrenden aus österreichischen und deutschen Fachhochschulen und Universitäten, was das Konzept des Buches und die einfachen, eher alltagssprachlichen und intuitiv verständlichen

Erklärungen statistischer Begriffe, den weitestgehenden Verzicht auf Formeln und Ableitungen sowie die gut verständlichen Beispiele betrifft, legen nahe, nichts Wesentliches zu verändern. Der Bedarf an einer 6. Auflage spricht für sich. Dies ist kein Lehrbuch für Profis, sondern soll Studierenden unterschiedlicher Studienrichtungen den Einstieg erleichtern sowie lehrenden Kolleg\*innen, die in der Praxis der Vermittlung von quantitativen Forschungsmethoden stehen, Anregungen für ihren Unterricht bieten. Und last, but not least werden auch Personen aus verschiedenen Anwendungsfeldern, wie Psycholog\*innen, Sozialarbeiter\*innen, Mediziner\*innen, Kollegen\*innen aus unterschiedlichen Gesundheitsberufen, wie Logopädie, Physiotherapie, aus der Krankenpflege u. v. m., deren Ausbildungszeit schon einige Jahre zurückliegt, gerne auf dieses einfach verständliche Buch zurückgreifen, um einige Lücken aus Theorie und Praxis des wissenschaftlichen Arbeitens zu füllen – das zumindest legen die mittlerweile unzähligen Rückmeldungen nahe, die wir in den letzten zwölf Jahren erhalten haben.

Alle im Buch angeführten Beispiele können mithilfe von SPSS selbst nachgerechnet werden – das entsprechende Datenfile finden Sie auf <https://www.utb-shop.de/> beim Zusatzmaterial zu diesem Buch. Die Daten sind fiktiv und beziehen sich auf den im Anhang abgebildeten Übungsfragebogen. Wir empfehlen auch die selbstständige Bearbeitung der jedem Kapitel angehängten Übungsbeispiele – zu Ihrer Kontrolle finden Sie Musterlösungen auf den Seiten 175 bis 186.

Es bleibt uns nun wieder die Hoffnung, dass das vorliegende Buch auch weiterhin Unterstützung bei der Erstellung wissenschaftlicher Arbeiten bzw. Abschlussarbeiten bietet und vielleicht ein wenig Lust auf die Generierung von Daten und deren statistische

Auswertung macht. Zumindest wäre es schon ein wesentlicher Schritt in die richtige Richtung, wenn Ängste bzw. Vorbehalte, die quantitativen Zugängen oft entgegengebracht werden, reduziert bzw. ein wenig ins Positive verändert werden!

Wien, im Dezember 2020

Elisabeth  
Steiner,  
Michael  
Benesch



# Inhaltsverzeichnis

## 1 Elementare Definitionen

- 1.1 Deskriptive Statistik und Inferenzstatistik
  - 1.1.1 Deskriptivstatistik (beschreibende Statistik)
  - 1.1.2 Inferenzstatistik (beurteilende bzw. schließende Statistik)
- 1.2 Stichprobenarten
  - 1.2.1 Einfache Zufallsstichprobe (Random Sample)
  - 1.2.2 Geschichtete Zufallsstichprobe
  - 1.2.3 Klumpenstichprobe (Cluster Sample)
  - 1.2.4 Zufall versus willkürliche Auswahl
  - 1.2.5 Abhängigkeit der Stichproben
- 1.3 Schluss von der Stichprobe auf die Grundgesamtheit
- 1.4 Zusammenfassung des Kapitels
- 1.5 Übungsbeispiele

## 2 Messung in den Sozialwissenschaften

- 2.1 Skalenbzw. Messniveaus
- 2.2 Nominalskala
- 2.3 Ordinalskala
- 2.4 Intervallskala
- 2.5 Verhältnisskala
- 2.6 Zusammenfassung des Kapitels
- 2.7 Übungsbeispiele

## 3 Die Untersuchungsplanung - von der Idee zur empirischen Forschung

- 3.1 Die Themensuche
  - 3.1.1 Das Anlegen einer Ideensammlung
  - 3.1.2 Die Replikation von Untersuchungen
  - 3.1.3 Die Mitarbeit an Forschungsprojekten
  - 3.1.4 Weitere kreative Anregungen
- 3.2 Konkretisierung und Formulierung einer Forschungsfrage
- 3.3 Die Literaturrecherche
- 3.4 Auswahl der Untersuchungsart - Forschungsdesign
- 3.5 Ethische Bewertung einer Forschungsfrage
- 3.6 Zusammenfassung des Kapitels
- 3.7 Übungsbeispiele

## 4 Datenerhebung: Die schriftliche Befragung (Fragebogen)

- 4.1 Methoden der quantitativen Datenerhebung
- 4.2 Allgemeine inhaltliche Vorbemerkungen zur Fragebogenkonstruktion
- 4.3 Erste inhaltliche Schritte
- 4.4 Prinzipien der Konstruktion
  - 4.4.1 Fragenauswahl
  - 4.4.2 Einleitung, Instruktion und Anrede
  - 4.4.3 Richtlinien zur Formulierung der Items
  - 4.4.4 Antwortformate
- 4.5 Pretest
- 4.6 Negative Antworttendenzen
  - 4.6.1 Absichtliche Verstellung
  - 4.6.2 Soziale Erwünschtheit (Social Desirability)
  - 4.6.3 Akquieszenz oder „Ja-Sage-Bereitschaft“ 62
  - 4.6.4 Bevorzugung von extremen, unbestimmten oder besonders platzierten Antwortkategorien
  - 4.6.5 Wahl von Antwortmöglichkeiten, die eine bestimmte Länge, Wortfolge oder seriale Position aufweisen
  - 4.6.6 Verfälschung aufgrund der Tendenz, zu raten, oder aufgrund einer raschen Bearbeitung des Tests
  - 4.6.7 Tendenz zur ersten passenden Kategorie
  - 4.6.8 Beeinflussung durch motivationale Bedingungen
  - 4.6.9 „Mustermalen“
- 4.7 Zusammenfassung des Kapitels
- 4.8 Übungsbeispiele

## **5 Computerunterstützte Datenaufbereitung mittels SPSS**

- 5.1 Was ist SPSS?
- 5.2 Vom Fragebogen zur SPSS-Datei
  - 5.2.1 Wie rufe ich SPSS auf?
  - 5.2.2 Wichtige Anmerkungen vor der Dateneingabe
  - 5.2.3 Kodierung und Kodeplan
  - 5.2.4 Erstellung eines Datenfiles
  - 5.2.5 Datencheck/Data-Cleaning
  - 5.2.6 Weitere Datenaufbereitung
- 5.3 Zusammenfassung des Kapitels
- 5.4 Übungsbeispiele

## **6 Deskriptivstatistische Datenanalyse**

- 6.1 Tabellarische Darstellung der Daten
  - 6.1.1 Häufigkeitstabellen
  - 6.1.2 Kreuztabellen bzw. Kontingenztafeln
- 6.2 Grafische Darstellung der Daten
  - 6.2.1 Balkendiagramme
  - 6.2.2 Histogramme
  - 6.2.3 Boxplots
  - 6.2.4 Streudiagramme
- 6.3 Lagemaße - Lokalisationsparameter
  - 6.3.1 Normalverteilung

- 6.3.2 Das arithmetische Mittel – der Mittelwert
- 6.3.3 Der Median
- 6.3.4 Der Modus (Modalwert)
- 6.4 Dispersionsmaße (Streuungsmaße)
  - 6.4.1 Varianz
  - 6.4.2 Standardabweichung
  - 6.4.3 Der Quartilabstand
  - 6.4.4 Spannweite
  - 6.4.5 Perzentilwerte
- 6.5 Zusammenfassung des Kapitels
- 6.6 Übungsbeispiele

## **7 Schluss von der Stichprobe auf die Population**

- 7.1 Alltags- und statistische Hypothesen
- 7.2 Statistischer Test
- 7.3 Fehler erster und zweiter Art und die Macht eines Tests
- 7.4 Der Standardfehler des Mittelwerts
- 7.5 Zusammenfassung des Kapitels
- 7.6 Übungsbeispiele

## **8 Statistische Tests**

- 8.1 T-Test für unabhängige Stichproben
- 8.2 T-Test für abhängige Stichproben
- 8.3 U-Test nach Mann & Whitney
- 8.4 Wilcoxon-Test
- 8.5 Friedman-Test
- 8.6 Vierfelder-Chi-Quadrat-Test
- 8.7 Zusammenfassung des Kapitels
- 8.8 Übungsbeispiele

## **9 Korrelation und lineare Regression**

- 9.1 Produkt-Moment-Korrelation
- 9.2 Rangkorrelation nach Spearman
- 9.3 Vierfelderkorrelation
- 9.4 Partielle Korrelation
- 9.5 Biseriale Korrelation
- 9.6 Korrelation und Kausalität
- 9.7 Einfache lineare Regression
- 9.8 Multiple lineare Regression
- 9.9 Zusammenfassung des Kapitels
- 9.10 Übungsbeispiele

## **10 Varianzanalyse**

- 10.1 Grundlagen der Varianzanalyse
- 10.2 Einfaktorielle Varianzanalyse ohne Messwiederholung
- 10.3 Einfaktorielle Varianzanalyse mit Messwiederholung
- 10.4 Zusammenfassung des Kapitels
- 10.5 Übungsbeispiele

## **11 Der statistische Auswertungsbericht**

11.1 Der Theorieteil

11.2 Der Methodenteil

11.3 Der Ergebnisteil

11.4 Diskussion und Ausblick

11.5 Einige Zitierregeln

11.6 Das Literaturverzeichnis

11.7 Zusammenfassung des Kapitels

11.8 Übungsbeispiele

### **Anhang**

Lösungen zu den Übungsbeispielen

Beispiel: Fragebogen zur Studien- und Lebenssituation bei Studierenden

### **Literaturverzeichnis**

### **Stichwortverzeichnis**

# 1 Elementare Definitionen

## 1.1 Deskriptive Statistik und Inferenzstatistik

Grundsätzlich wird bei der Analyse quantitativer Beobachtungen bzw. Messungen und deren Beschreibung die **Inferenzstatistik** von der **Deskriptivstatistik** unterschieden. Diese beiden prinzipiellen Zugänge in der Statistik sollen im folgenden Kapitel in ihrer Unterschiedlichkeit und Anwendbarkeit genauer dargestellt werden. Diese Darstellungsform soll jedoch nicht den falschen Eindruck entstehen lassen, dass die beiden Zugänge konkurrierend auftreten. In der Praxis stellen sie einander ergänzende und inhaltlich bereichernde Zugänge dar.

### 1.1.1 Deskriptivstatistik (beschreibende Statistik)

„Ein wesentlicher Teil der Statistik ist die Datenbeschreibung einschließlich einer systematischen Suche nach aufschlussreichen Informationen über die Struktur eines Datenkörpers. Strukturen in den Daten und bedeutsame Abweichungen von diesen Strukturen sollen aufgedeckt werden.“ (Hedderich & Sachs, 2011, S. 11)

Es werden also bestimmte Charakteristika (Eigenschaften) einer Stichprobe beschrieben, allerdings noch ohne den Anspruch, etwas über die dahinterstehende

Grundgesamtheit (Population) auszusagen. Dies wäre der Ansatz, den die Inferenzstatistik verfolgt. Bei dieser Beschreibung interessieren im Grunde die Ableitungen von gewissen, in den Daten auffindbaren Gesetzmäßigkeiten, die auch die Basis für weitere inferenzstatistische Verwertungen darstellen.

Es handelt sich um einen summarischen Zugang zu quantitativen Informationen. Wenn wir z. B. etwas über eine Stichprobe von Studierenden ( $n = 127$ ) wissen möchten, müssen wir im ersten Schritt entscheiden, welche Eigenschaften dieser Stichprobe uns interessieren, und im nächsten Schritt, ob wir diese Eigenschaften zunächst grafisch veranschaulichen und/oder ob Maßzahlen wie Mittelwerte und Streuungen zur Beschreibung herangezogen werden (mehr dazu in [Kapitel 6](#)). Wir müssen also entscheiden, wie wir die wichtigsten „Eigenschaften“ der Stichprobe in geeigneter Form und gut überschaubar darstellen.

Nehmen wir an, uns interessiert die Geschlechterverteilung in der Stichprobe der 127 Studierenden. Für ihre Darstellung würde sich aufgrund der geringen Anzahl der Ausprägungen der Variable Geschlecht, nämlich männlich und weiblich, eine einfache Grafik wie das Kreisdiagramm ([Abb. 1.1](#)) anbieten:

Besteht diese Stichprobe von StudentInnen aus 70 männlichen und 57 weiblichen Personen, wird die Verteilung durch das Kreisdiagramm auf einfache und anschauliche Art und Weise grafisch dargestellt.

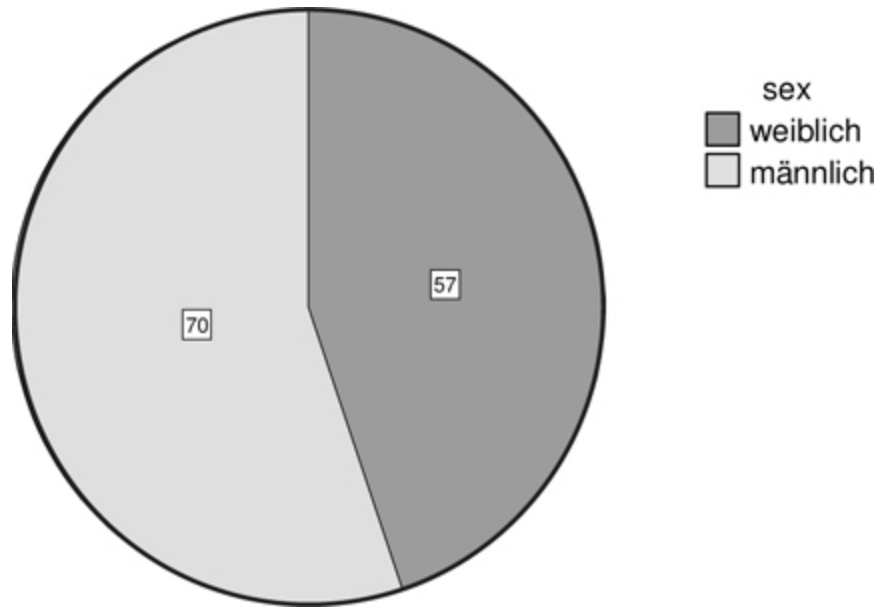


Abb. 1.1: Geschlechterverteilung/Angabe in absoluten Häufigkeiten

Eine weitere deskriptivstatistische Methode wäre die Darstellung einer einfachen Häufigkeitstabelle.

In [Tabelle 1.1](#) sind zusätzlich zu diesen absoluten Häufigkeiten von 70 und 57 die Prozente angegeben. Man berechnet sie, indem man die absoluten Zahlen jeder Gruppe durch die Stichprobengröße dividiert und anschließend mit 100 multipliziert ( $57/127 = 44,9\%$  und  $70/127 = 55,1\%$ ).

Tab. 1.1: Geschlechterverteilung/Häufigkeiten und Prozent

		Häufigkeit	Prozent
Gültig	weiblich	57	44,9%
	männlich	70	55,1%
Gesamt		127	100,0%

Natürlich ist Statistik mit Informationsreduktion verbunden, das ist eine ihrer Grundideen und bedeutet in

unserem Beispiel: Aus dem Kreisdiagramm ([Abb. 1.1](#)) oder der Tabelle ([Tab. 1.1](#)) ist nicht mehr ersichtlich, welches Individuum der Stichprobe männlich oder weiblich ist. Wir kennen nur noch die entsprechenden Anteile (45 % und 55 %) bzw. Häufigkeiten (57 und 70).

Eine weitere gängige Methode der Deskriptivstatistik, um Stichproben zu beschreiben, besteht darin, sogenannte deskriptivstatistische Maßzahlen zu berechnen. Die bekanntesten sind das arithmetische Mittel (meist nur „Mittelwert“ genannt) und die Standardabweichung (dazu eine ausführliche Beschreibung in [Kapitel 6](#)), die dazugehörige Streuung.

Den Mittelwert ( $\bar{x}$  = arithmetisches Mittel) erhält man, indem alle Messwerte (wie z. B. das Alter in Jahren) addiert werden und die resultierende Summe durch die Anzahl der Messwerte ( $n$  = Stichprobengröße) dividiert wird.

$\bar{x}$  = 24 Jahre für die männliche Stichprobe

Zieht man einen der sieben Studenten aus der Gruppe und erfragt sein Alter, so ist die Wahrscheinlichkeit hoch, dass es im Bereich um 24 Jahre liegt. Allerdings ist die Angabe des Mittelwertes praktisch sinnlos, wenn man nichts über die Verteilung der ursprünglichen Messwerte weiß. In einer Stichprobe von drei 20-jährigen Personen beträgt der Mittelwert zwanzig Jahre  $[(20 + 20 + 20) / 3 = 20]$ . Auch in einer Stichprobe mit einem 10-jährigen, einem 11-jährigen und einem 39-jährigen macht der Mittelwert zwanzig Jahre aus  $[(10 + 11 + 39) / 3 = 20]$ .

Dies führt uns zum nächsten Schritt - der Angabe der dazugehörigen Streuungsmaße (Dispersionsmaße), die Aufschluss über die „Differenzen“ in der Altersverteilung geben können, z. B. die Standardabweichung =  $s = 3$  Jahre.



Das heißt, in Kombination mit der Angabe des Mittelwerts von 24 Jahren kann unter der Annahme der Normalverteilung (dazu ebenfalls mehr in [Kapitel 6](#)) davon ausgegangen werden, dass rund 68 % der Studenten im Altersbereich von 21 bis 27 Jahren liegen (d. h. im Bereich 24 Jahre +/- 3 Jahre).

Durch die so durchgeführte Beschreibung der Stichprobe gewinnt man bereits einen guten Überblick über deren Charakteristika, also wesentliche Informationen über ihre Beschaffenheit: Wir wissen bis jetzt, dass die Stichprobe aus 57 weiblichen und 70 männlichen Studierenden besteht. Dies könnte mit einer Häufigkeitstabelle unter der zusätzlichen Angabe von Prozenten noch ergänzt werden. Der Altersdurchschnitt der männlichen Studierenden liegt bei 24 Jahren. Und rund 68 % der männlichen Studierenden liegen im Altersbereich von 21 bis 27 Jahren.

Statistische Methoden zur Beschreibung der Daten von Stichproben in Form von Grafiken, Tabellen oder einzelnen Kennwerten (Lagemaße bzw. Streuungsmaße) bezeichnen wir zusammenfassend als deskriptive (beschreibende) Statistik.

Sie gibt einen Überblick über die Merkmalsausprägungen einzelner Variablen und stellt oft eine fundierte Basis für weitere statistische Berechnungen dar.

### **1.1.2 Inferenzstatistik (beurteilende bzw. schließende Statistik)**

Auf Basis von Erfahrungen, Beobachtungen und Wissen ziehen wir Rückschlüsse - dieser Prozess wird als Inferenz

bezeichnet. Dabei können zwei Zugänge verfolgt werden: Vom „Allgemeinen“ auf das „Besondere“ zu schließen wird als **deduktiver Zugang** bezeichnet. Im Gegensatz dazu spricht man vom **induktiven Zugang**, wenn man vom „Besonderen“ auf das „Allgemeine“ schließen möchte. Durch diese Differenzierung werden auch grob qualitative von quantitativen Forschungszugängen unterschieden.

Während die Deskriptivstatistik eine Stichprobe beschreibt, ermöglicht die Inferenzbzw. analytische oder beurteilende Statistik, über diese Stichprobe hinaus etwas über die dahinterstehende Grundgesamtheit (Population) auszusagen, also Verallgemeinerungen zu treffen.

Die Inferenzstatistik bzw. „Beurteilende (Schließende) Statistik untersucht [...] nur einen Teil, der für die Grundgesamtheit, deren Eigenschaften uns interessieren, charakteristisch oder repräsentativ sein soll“ (Hedderich & Sachs, 2011, S. 9).

Die Grundidee liegt also darin, von einer kleinen Auswahl (Stichprobe) auf die dahinterliegende Grundgesamtheit zu schließen.

Es gibt neben den bereits erwähnten Begrifflichkeiten Inferenzstatistik, beschreibende bzw. analytische Statistik auch noch die Bezeichnung induktive (hinführende) Statistik. Alltagssprachlich wird eine solche Hinführung als logischer Schluss dargestellt. Die Verwendung dieser Begriffe verwirrt oft, bezeichnet aber den gleichen Zugang zur Statistik.

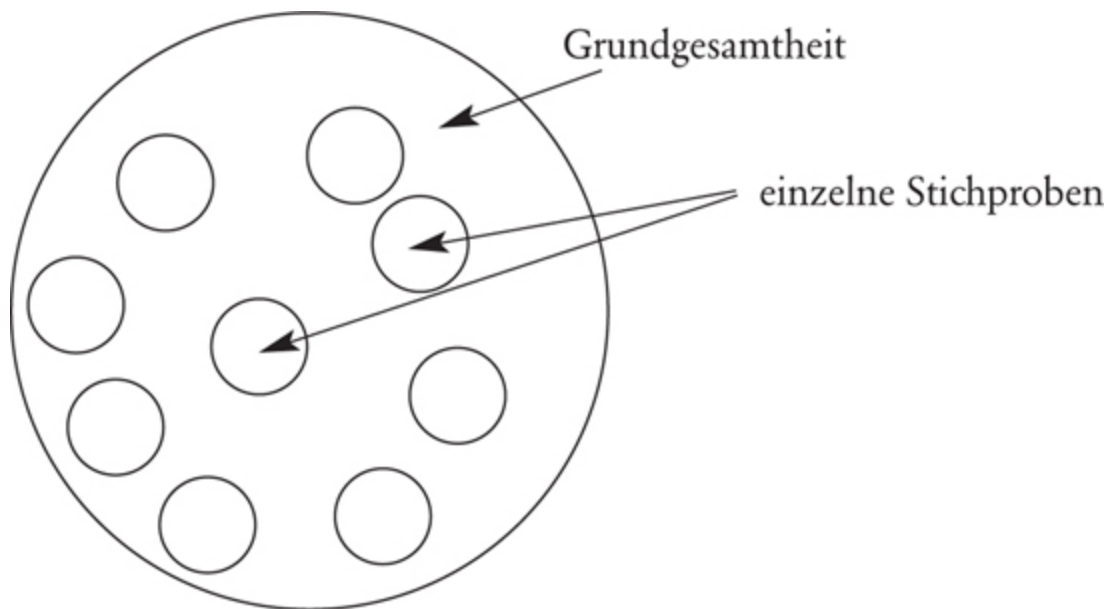
Wir stellen in der Stichprobe fest, dass sich die männlichen von den weiblichen Studierenden hinsichtlich des Lernaufwandes für eine bestimmte Prüfung unterscheiden, also eine Gruppe für dieselbe Prüfung

länger lernt. Mittels der Methoden der Inferenzstatistik, mit denen wir uns später beschäftigen werden, kann nun festgestellt werden, ob es sich in der Grundgesamtheit (alle StudentInnen dieser Studienrichtung an dieser Universität) genauso wie in der Stichprobe verhält. Dieses Schließen von der Stichprobe auf das Dahinterstehende – die Grundgesamtheit – ist allerdings nicht mit absoluter Sicherheit möglich, sondern nur mit einer bestimmten Wahrscheinlichkeit. Die Verallgemeinerung auf die Population ist stets unsicher. Wir können mithilfe statistischer Auswertungen prinzipiell nur Wahrscheinlichkeitsaussagen treffen – und dies mit unterschiedlicher Genauigkeit.

In sozialwissenschaftlichen Untersuchungen möchte man also meist über die Beschreibung einer ausgewählten (spezifischen) Gruppe von Untersuchungseinheiten (Stichproben) hinausgehen und allgemein gültige Aussagen treffen. Dazu ist die rein deskriptive Statistik, die Beschreibung der Daten in Form von Häufigkeitstabellen, Grafiken und einzelnen Kennwerten, in den wenigsten Fällen ausreichend.

Die Inferenzbzw. beurteilende Statistik nimmt sich des Problems an, wie man Ergebnisse, die an einer verhältnismäßig kleinen Zahl von Personen (Stichprobe) gewonnen wurden, auf die Grundgesamtheit (Population) umlegen, also allgemein gültige Aussagen treffen kann. Die allgemein gültige Aussage (über die Grundgesamtheit) wird als Hypothese formuliert, die anhand von Stichproben zu überprüfen ist. Hierin liegt ein wesentlicher Unterscheidungspunkt der zwei Zugänge. Die Inferenzstatistik stellt Hypothesen auf und ermöglicht deren Überprüfung.

Aus der Grundgesamtheit wird eine von vielen möglichen Stichproben gezogen. Die folgende [Abbildung 1.2](#) verdeutlicht dies. Aus einer größeren Grundgesamtheit (der große Kreis) gibt es nahezu unendlich viele Möglichkeiten, einzelne Stichproben (kleine Kreise) zu erhalten. Wichtig ist, dass diese gezogene Stichprobe „repräsentativ“ ist, also die wesentlichen Charakteristika der Grundgesamtheit widerspiegelt.



*Abb. 1.2: Grundgesamtheit mit verschiedenen Stichprobenziehungen*

Ein bekanntes Anwendungsgebiet ist die Hochrechnung vor Wahlen. Die Meinungsforschungsinstitute konkurrieren jeweils um die korrekteren Vorhersagen des Wahlausgangs. Sie gehen dabei so vor, dass sie eine kleine (aber repräsentative) Stichprobe von WählerInnen befragen, von der sie auf die Grundgesamtheit der Bevölkerung schließen können.

Repräsentativität bedeutet in diesem Beispiel, dass die „kleine“ ausgewählte Gruppe möglichst die reale Situation der „Grundgesamtheit“ beschreibt, also die Variablen (Eigenschaften), wie z. B. Geschlecht, Alter,

Ausbildungsstand, soziale Schicht usw., real abgebildet sind.

Natürlich sind Ergebnisse, die aufgrund von Daten einer Stichprobe gewonnen werden, mit Ungenauigkeiten behaftet. Das ist auch der Grund, weshalb bei einer Wahlprognose stets ein Bereich angegeben wird, z. B.  $\pm 2\%$ , in dem das „wahre“ Ergebnis (also der Anteil der WählerInnen an der Grundgesamtheit) mit gewisser Wahrscheinlichkeit liegt.

Neben den Ergebnissen, die durch analytische Verfahren gewonnen werden, können deskriptivstatistische zusätzlich zu einer übersichtlichen und anschaulichen Informationsaufbereitung beitragen.

Das Zusammenspiel der beiden Methoden kann sich gut ergänzen und zu einem Höchstmaß an Information führen.

Die Inferenzstatistik wird häufig auch als analytische Statistik oder schließende Statistik bezeichnet. Der wesentliche Unterschied zur deskriptiven Statistik liegt darin, dass es zur Überprüfung von Hypothesen, die sich auf die dahinterstehende Grundgesamtheit beziehen, kommt. Auf diese Weise sollen allgemein gültige Aussagen über die Stichprobe hinaus getroffen werden. Es wird ein deduktiver Zugang verfolgt.

## **1.2 Stichprobenarten**

In der Empirie (wissenschaftlich gewonnene Erfahrung) werden unterschiedliche Zugänge zur Auswahl einer repräsentativen Stichprobe verfolgt. Mittels eines Stichprobenplans wird das Zufallsverfahren festgelegt, um repräsentative Elemente zu ziehen.

Der Begriff „Stichprobe“ bezeichnet eine kleine Teilmenge der sogenannten Grundgesamtheit, deren Auswahl nach bestimmten Kriterien erfolgen sollte, um verallgemeinerbare Aussagen treffen zu können.

Die Ziehung einer Stichprobe hat einen sehr pragmatischen Ursprung, nämlich jenen, dass die Befragung der Grundgesamtheit (Vollerhebung, z. B. der österreichischen Gesamtbevölkerung) nicht (oder nur sehr aufwendig) möglich ist und den Rahmen einer Untersuchung meist sprengen würde. Allerdings ist bei sozialwissenschaftlichen Fragen anzunehmen, dass gezogene Stichproben auch unter sehr guten Überlegungen und Bedingungen die Verteilung der Merkmale in der Population nicht exakt abbilden. Man müsste im Vorfeld bereits exakte Angaben über Verteilungen und Merkmalsausprägungen haben, was in der Realität kaum gegeben ist. Nichtsdestotrotz ist der grundsätzliche Zugang bei der Ziehung von Stichproben das sogenannte Induktionsprinzip (vom lateinischen *inductio*, Hineinführen), bei dem vom besonderen Fall auf den allgemeinen geschlossen wird.

„Unter Grundgesamtheit ist diejenige Menge von Individuen, Fällen, Ereignisse zu verstehen, auf die sich die Aussagen der Untersuchung beziehen sollen und die im Hinblick auf die Fragestellung und Operationalisierung vorher eindeutig abgegrenzt werden muss“ (Kromrey, 2009, S. 255).

Beispiele für Grundgesamtheiten sind: alle BewohnerInnen von Wien, alle RaucherInnen einer Zigarettenmarke in Österreich, alle RechtshänderInnen, alle

ostösterreichischen StudentInnen einer bestimmten Studienrichtung etc.

Eine gezogene Stichprobe sollte die Grundgesamtheit möglichst genau abbilden. Je besser diese kleine Teilmenge die Grundgesamtheit abbildet, desto präzisere Aussagen können über sie gemacht werden. Dies stellt jedoch eine gewisse Herausforderung dar, denn die Repräsentativität in allen Merkmalen kann in den seltensten Fällen im statistischen Sinne erfüllt werden, besonders dann, wenn die Untersucherin/der Untersucher keinerlei Hinweise auf die Verteilung der relevanten einzelnen Variablen in der Stichprobe hat.

Neben der Art und Weise, wie die Stichprobe gezogen (zufallsgesteuert bzw. nicht zufallsgesteuert) wird, ist natürlich auch deren Größe von Bedeutung. Im Allgemeinen kann jedoch eine auch noch so große Stichprobe gravierende Fehler bei der Stichprobenziehung nicht wettmachen. Möchte man beispielsweise etwas über das Durchschnittseinkommen der StudentInnen wissen und befragt dazu fünftausend Studierende, wird man stark verzerrte Ergebnisse erhalten, wenn diese fünftausend Personen zum Großteil nebenberuflich studieren, also vollwertige Einkommen haben. Diese Stichprobe wäre nicht repräsentativ für „die StudentInnen“, wenn diese zum Großteil eben nicht nur nebenberuflich studieren. Es würde zu einem „Bias“ kommen, einer systematischen Verzerrung. Die Stichprobe müsste, um zu sinnvollen Schlussfolgerungen zu kommen, so gezogen werden, dass sie die realen Verhältnisse gut abbildet – eine Vorerhebung der Verteilungen wäre unerlässlich.

An dieser Stelle sollen nun die in den Sozialwissenschaften gängigen Stichprobenarten dargestellt werden. Ein besonderes Augenmerk liegt dabei auf der Zufallsstichprobe, welche die häufigste Variante darstellt.

### **1.2.1 Einfache Zufallsstichprobe (Random Sample)**

Liegen, wie oben erwähnt, keinerlei Hinweise auf die Verteilung relevanter Variablen in der Grundgesamtheit vor, empfiehlt sich die Ziehung einer **Zufallsstichprobe** (Random Samples), denn bei dieser Stichprobe hat dann jedes Merkmal die gleiche Wahrscheinlichkeit, in die relevante Stichprobe gezogen zu werden.

„Eine Zufallsstichprobe ist dadurch gekennzeichnet, dass jedes Element der Grundgesamtheit mit gleicher Wahrscheinlichkeit ausgewählt werden kann“ (Bortz, 2010, S. 87).

Man spricht in diesem Fall von einer reinen (einfachen) Zufallsstichprobe (Simple Random Sample). Die Wichtigkeit dieser Stichprobenziehung in den Sozialwissenschaften muss besondere Beachtung erhalten. Es ist dies oft die einzige Möglichkeit des Zugangs.

### **1.2.2 Geschichtete Zufallsstichprobe**

Eine weitere Möglichkeit wäre es, eine **geschichtete (stratifizierte) Zufallsstichprobe** zu ziehen. Dabei wird die Stichprobe anhand einer ausgewählten Schichtungsvariable in einander nicht überschneidende Schichten geteilt. Diese Schichten sollten in sich ziemlich homogen sein, untereinander aber sehr unterschiedlich. Aus diesen Segmenten zieht man dann eine Zufallsstichprobe. Diese Vorgangsweise macht natürlich nur dann Sinn, wenn die Schichtungsvariable einen hohen Zusammenhang mit dem eigentlich interessierenden Untersuchungsmerkmal hat. Man muss über die Verteilung



der Merkmale in der Grundgesamtheit Bescheid wissen, um eine repräsentative Stichprobe erzeugen zu können. Die so gezogene Stichprobe wird als geschichtet oder stratifiziert bezeichnet (vgl. ebd., S. 88). Ein Beispiel: Wenn das Freizeitverhalten Jugendlicher untersucht werden soll, muss bei der Ziehung der Stichprobe auf Alter, Taschengeldhöhe, Stadt/Land, Geschlecht etc. geachtet werden. Aus diesen einzelnen Schichten (Strata; Sg. Stratum) werden dann zufällig Jugendliche gezogen, also aus der Gruppe Stadt/Land, der Gruppe Taschengeldhöhe usw.

Bei der stratifizierten Zufallsstichprobe muss bekannt sein, welche Faktoren die Verteilung des untersuchten Merkmals beeinflussen, um eine geeignete Auswahl treffen zu können.

### **1.2.3 Klumpenstichprobe (Cluster Sample)**

In der praktischen Arbeit mit Daten kommt es immer wieder vor, dass vorgruppierte Teilmengen der Grundgesamtheit vorliegen. Man spricht in diesem Fall von sogenannten **Klumpenstichproben**, diese werden neben den geschichteten Stichproben ebenfalls den mehrstufigen Zufallsstichproben zugeordnet. Klumpenstichproben sind dann sinnvoll, wenn die Elemente der Grundgesamtheit nicht erfasst werden können, aber Informationen darüber vorhanden sind, wo diese Elemente gefunden werden können. Ein Beispiel: Es gibt keine Listen darüber, welche und wie viele Wiener PatientInnen an Bluthochdruck leiden. Aber Spitäler führen Aufzeichnungen über ihre eigenen PatientInnen, und so könnte man eine bestimmte Anzahl an Wiener Spitälern (das wären die „Klumpen“ oder

„Cluster“) auswählen und aus diesen Clustern Zufallsstichproben von Bluthochdruck-PatientInnen ziehen.

„Eine Klumpenstichprobe besteht aus allen Untersuchungsteilnehmern, die sich in mehreren, zufällig ausgewählten Klumpen befinden“ (Bortz, 2010, S. 87).

Diese Klumpenstichproben müssen allerdings von **Ad-hoc-Stichproben** (anfallenden Stichproben) differenziert werden - es müssen mehrere zufällig ausgewählte Klumpen vollständig untersucht werden. Ad-hoc-Stichproben wären eine Schulklasse, eine Seminargruppe, Kranke auf einer Station im Krankenhaus. Bei diesen anfallenden Stichproben wird ohne spezielle Planung und ohne genaue Kenntnis der Merkmalsausprägungen in der Population vorgegangen.

Bei einer Klumpenstichprobe wird die Grundgesamtheit in einzelne, sich ähnelnde Klumpen (Homogenität der Klumpen) zerlegt. Daraus wird eine Zufallsstichprobe genommen, z. B. werden zuerst einzelne Schulklassen (Klumpen) aus allen Klassen (Grundgesamtheit) gezogen und dann die SchülerInnen daraus befragt.

Das Grundproblem liegt darin, dass die Gefahr der nicht hinreichend gegebenen Repräsentativität sehr hoch ist.

Zu berechnende Kenngrößen unterscheiden sich in Abhängigkeit von der Klumpenauswahl. Je homogener die Gruppen sind, desto größer die Schwankungen zwischen den Auswahlen.

### **1.2.4 Zufall versus willkürliche Auswahl**

Dem Prinzip der Zufallsstichprobe steht die willkürliche Auswahl von Stichproben gegenüber. Dabei werden von der Befragerin/dem Befrager willkürliche Kategorien

eingezogen. Wahrscheinlichkeiten darüber, ob ein bestimmtes Element in die Stichprobe aufgenommen wird, können dabei nicht angegeben werden.

Es geht um eine bewusste Auswahl. Beispiele hierzu sind:

- eine rein willkürliche Auswahl - ein sehr unwissenschaftlicher Zugang, z. B. Befragungen auf der Straße, bei denen jeder zehnte Passant angesprochen wird;
- eine Schneeballauswahl - diese wird häufig als Methode für den Zugang zu kleinen bzw. schwer zugänglichen Gruppen genutzt. Eine Person dieser Gruppe gibt das Erhebungsinstrument (z.B. Fragebogen) an eine von ihr als relevant eingeschätzte Person weiter.
- eine Auswahl der Elemente, die als sehr typisch angesehen werden;
- eine Quotenauswahl - vorausgehende Festlegung der Gruppen, die gezogen werden müssen. Das setzt voraus, dass über die diesbezüglichen Informationen verfügt wird.

### **1.2.5 Abhängigkeit der Stichproben**

Ein sehr wesentlicher Punkt, falls es zu Gruppenvergleichen mittels analytisch-statistischer Verfahren kommen soll, ist die Frage nach der Abhängigkeit der Stichproben. Dabei muss die abhängige von der unabhängigen Stichprobe unterschieden werden:

- **Abhängige Stichproben:** Typisch für abhängige Stichproben ist das zwei- oder mehrmalige Untersuchen derselben Personen, also beispielsweise vor und nach einem Therapieprogramm. Bei einer Befragung derselben Personen zu zwei Zeitpunkten muss etwa durch entsprechende Probandencodes sichergestellt werden, dass die zweiten Messwerte eindeutig den ersten zugeordnet werden können.

- **Unabhängige Stichproben:** Die Stichproben bestehen aus Elementen, die voneinander unabhängig sind, d. h., wer zur Stichprobe A gehört, kann nicht Teil der Stichprobe B sein. Typisch für ein unabhängiges Design wäre die Befragung von männlichen und weiblichen SchülerInnen zu einem bestimmten Thema, um mögliche Geschlechtsunterschiede zu untersuchen. Wer männlich ist, kann nicht Teil der weiblichen Stichprobe sein und umgekehrt.

### **1.3 Schluss von der Stichprobe auf die Grundgesamtheit**

Die analytische Statistik (Inferenzstatistik) beschäftigt sich also mit dem Schluss von der Stichprobe auf die Grundgesamtheit. Wie bereits mehrfach erwähnt, ist es in den allermeisten Fällen allein schon aus organisatorischen Gründen nicht möglich, die gesamte Population (Grundgesamtheit) zu untersuchen.

Dabei stellt sich aber ein gravierendes Problem: Wie kann man etwas über eine Population aussagen, wenn nur Stichprobenresultate bekannt sind? Derartige Schlüsse sind nicht mit absoluter Sicherheit möglich, sondern nur als Wahrscheinlichkeitsaussagen formulierbar, was wir schon bei unserem Beispiel der Hochrechnung von Wahlen festgehalten haben.

Bei Wahlprognosen finden wir solche Unsicherheiten durch die Angabe eines Intervalls von zumeist „ $\pm 2\%$ “: Auf die Partei X werden  $38\%$  ( $\pm 2\%$ ) der Stimmen entfallen, womit ausgedrückt wird, dass mit einer bestimmten Wahrscheinlichkeit der „wahre“ Anteil der WählerInnen dieser Partei (also der Anteil der WählerInnen in der Population der Wahlberechtigten) hier im Bereich von  $36\%$  bis  $40\%$  liegt. Könnte man alle Wahlberechtigten befragen und nicht nur eine Stichprobe von zumeist rund

tausend Personen, bräuchte man nicht die Wahrscheinlichkeit bemühen, sondern könnte eine „sichere“ Aussage treffen. Die Situation ist vergleichbar mit dem Schwangerschaftstest: Zu Beginn der Schwangerschaft ist es nicht möglich, zu sehen, ob eine Frau schwanger ist oder nicht. Aber mit 99,9 %iger Wahrscheinlichkeit ist sie es nicht, sollte das Testergebnis (dieses steht hier beispielhaft für die Stichprobe) negativ sein, und mit vergleichbarer Wahrscheinlichkeit ist sie schwanger bei positivem Testergebnis. Beim Schluss von einer Stichprobe auf die dahinterliegende Grundgesamtheit muss auch der Stichprobenumfang beachtet werden. Nach dem „Gesetz der großen Zahlen“ nähern sich die Eigenschaften einer Stichprobe mit wachsendem Umfang den Eigenschaften der Grundgesamtheit an. Für die praktische Arbeit muss jedoch eine handhabbare Lösung gefunden werden und oft können aus Zeitbzw. Kostengründen Stichprobenumfänge nicht in optimalem Umfang erhoben werden.

Eine verbindliche Untergrenze kann auch hier nicht in eine Zahl gefasst werden, da dies von einigen Komponenten, wie z. B. der Streuung der untersuchten Variable oder dem relativen Anteil der Stichprobe an der Gesamtpopulation, abhängig ist und individuell entschieden werden muss.

Für die Repräsentativität einer Stichprobe und die Anwendbarkeit der meisten Test- und Schätzverfahren der analytischen Statistik sollte jedoch ein Mindestumfang von 30 Fällen pro Untergruppe (Merkmalsausprägung einer Variable), z. B. Geschlecht mit männlich und weiblich, gegeben sein. Dies ist allerdings nur eine Faustregel, die die Verteilungen der Variablen nicht berücksichtigt.

## **1.4 Zusammenfassung des Kapitels**

Grundsätzlich wird die Deskriptivstatistik von der Inferenzstatistik unterschieden. Deskriptiv heißt, in der Datenaufbereitung beschreibend vorzugehen. Die Darstellung der Ergebnisse erfolgt in Form von Grafiken, Tabellen und einzelnen statistischen Kennzahlen. Im Gegensatz dazu ermöglicht die Inferenzstatistik, über die bestehende Stichprobe hinaus Aussagen über die dahinterstehende Grundgesamtheit zu treffen. Es müssen dazu Hypothesen formuliert werden.

Als Stichprobe wird eine kleine Teilmenge der sogenannten Grundgesamtheit verstanden, die nach bestimmten Kriterien ausgewählt wird. Wir können dabei die einfache Zufallsstichprobe, die geschichtete Zufallsstichprobe, die Klumpenstichprobe und die Ad-hoc-Stichprobe unterscheiden. Selbstverständlich kann eine Stichprobe auch willkürlich gezogen werden, dies wäre z. B. das Quotaverfahren. Diese kleine Teilmenge soll repräsentativ sein, d. h. die Grundgesamtheit in ihren Eigenschaften gut abbilden. Um dies zu gewährleisten, ist in der sozialwissenschaftlichen Untersuchungsplanung die häufigste Art der Stichprobe die Zufallsstichprobe, in ihr hat jedes Element der Grundgesamtheit die gleiche Wahrscheinlichkeit, ausgewählt zu werden.

Ein weiterer wesentlicher Aspekt ist die Differenzierung zwischen abhängigen und unabhängigen Stichproben, vor allem wenn mittels analytisch-statistischer Verfahren Gruppenvergleiche angestellt werden sollen.

Wird von der repräsentativen Stichprobe auf die Grundgesamtheit geschlossen, kann dies nur mit einer gewissen Wahrscheinlichkeit getan werden. Eine absolute Aussage wäre nur durch eine Vollerhebung möglich.

## **1.5 Übungsbeispiele**