

2 Theoretische Grundlagen

2.1 Grundgesamtheit und Stichprobe

Die Hypothesen, die in den Bio- und Sozialwissenschaften aufgestellt werden und zu verifizieren sind, beziehen sich meist auf eine sehr große Anzahl von Individuen. Es wäre aus organisatorischen und zeitlichen Gründen viel zu aufwendig, oft sogar vollkommen unmöglich, die gesamte Population zu untersuchen, auf die eine Hypothese zutreffen könnte. Dies ist im allgemeinen auch gar nicht notwendig. Die moderne Statistik stellt nämlich Methoden zur Verfügung, die es ermöglichen, basierend auf einer relativ kleinen Stichprobe allgemein gültige Aussagen bezüglich einer weitaus größeren Grundgesamtheit herzuleiten.

Eine **Total-** oder **Vollerhebung** wird daher nur in Ausnahmefällen durchgeführt. Beispielsweise beruhen die Todesursachenstatistiken, die im jährlich erscheinenden Statistischen Jahrbuch der Bundesrepublik Deutschland veröffentlicht werden, medizinische Register oder die Ergebnisse einer politischen Wahl auf einer Vollerhebung. Im allgemeinen beschränkt man sich jedoch – insbesondere in der medizinischen Forschung – auf die Untersuchung einer kleinen Teilmenge, nämlich der **Stichprobe**, und überträgt die daraus gewonnenen Erkenntnisse auf die **Grundgesamtheit**. Dies ist allerdings nur unter der Voraussetzung sinnvoll, dass die charakteristischen Eigenschaften der Stichprobe – abgesehen von zufällig bedingten Abweichungen – mit denen der Grundgesamtheit übereinstimmen. Eine solche Stichprobe heißt **repräsentativ**.

Bei vielen Untersuchungen ist man vor das Problem gestellt, aus einer konkret vorgegebenen Grundgesamtheit eine repräsentative Stichprobe zu wählen. Ein Beispiel hierfür stellt eine Umfrage vor einer politischen Wahl dar. Die Grundgesamtheit besteht in diesem Fall aus allen wahlberechtigten Bürgern. Um eine Prognose zu erstellen, beschränkt man sich auf eine Stichprobe von einigen tausend Personen. Diese Stichprobe muss repräsentativ und hinrei-

chend groß sein, damit sie das endgültige Wahlergebnis in brauchbarer Weise widerspiegelt.

Bei Untersuchungen in der Medizin ist die Problemstellung häufig umgekehrt: Gegeben sind eine (oder mehrere) konkrete Stichproben (beispielsweise Patienten, die im Rahmen einer klinischen Studie beobachtet werden). Danach ist zu klären, wie die dazugehörige Grundgesamtheit beschaffen ist und ob die Stichprobenergebnisse auf diese übertragbar sind. Eine Antwort auf diese Frage beruht mehr auf sachlogischen als auf statistischen Überlegungen und ist eng mit dem jeweiligen Forschungsvorhaben verknüpft. Oft lässt sich die entsprechende Grundgesamtheit gar nicht konkret angeben. Man sollte sich in jedem Fall davor hüten, allzu weitreichende Schlussfolgerungen zu ziehen, die sich im nachhinein als falsch herausstellen könnten.

Dieses Problem kann man zwar umgehen, indem man eine Untersuchung nur für einen speziellen, eng begrenzten Personenkreis durchführt und diesen als Grundgesamtheit auffasst. Allerdings gelten die dadurch gewonnenen Ergebnisse nur eingeschränkt auf die Menge der untersuchten Personen und lassen sich nicht verallgemeinern.

2.2 Die Aufgaben der deskriptiven Statistik

Aus dem obigen Abschnitt geht hervor, dass bei einer Stichprobenuntersuchung die statistische Analyse aus zwei Teilen besteht. Zunächst werden die Daten der Stichprobe ausgewertet mit dem Ziel, deren charakteristische Eigenschaften zu beschreiben. Dies ist das Aufgabengebiet der *deskriptiven Statistik*. Dazu zählen im einzelnen:

- das Zusammenfassen und Ordnen der Daten in übersichtlichen *Tabellen*,
- das Erstellen von *Diagrammen* und
- das Berechnen *charakteristischer Kenngrößen* oder Maßzahlen (z. B. Mittelwert, Standardabweichung, Korrelationskoeffizient).

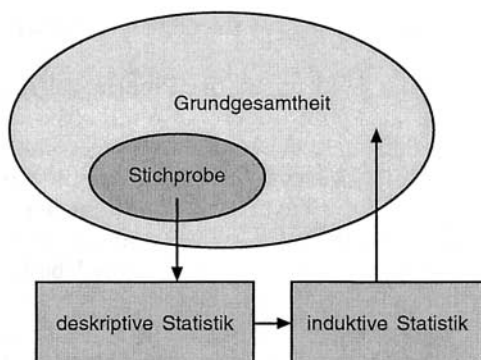


Abb. 2.1.
Grundgesamtheit und
Stichprobe

In einem zweiten Schritt versucht man dann, mit Methoden der **induktiven Statistik** zu allgemein gültigen Aussagen bezüglich der Grundgesamtheit zu gelangen. So gesehen, ist die deskriptive Statistik bei praktischen Anwendungen die Vorstufe zur induktiven Statistik. Beide Teilbereiche sind zur Datenanalyse notwendig und ergänzen sich.

Wenn anstelle einer Stichprobe die komplette Grundgesamtheit untersucht wird, werden die Daten nur mit Methoden der deskriptiven Statistik ausgewertet. Eine weitere Analyse mittels induktiver Statistik ist dabei nicht erforderlich.

2.3 Merkmale

2.3.1 Grundbegriffe

Untersuchungseinheiten. Die Personen oder Objekte einer Stichprobe werden als **Untersuchungseinheiten** (oder **Merkmalsträger**) bezeichnet. In der medizinischen Forschung handelt es sich dabei meist um Patienten, gesunde Probanden oder Versuchstiere.

Beobachtungseinheiten. Das sind die kleinsten Einheiten, an denen die einzelnen Beobachtungen registriert werden. Bei vielen Studien sind die Beobachtungseinheiten mit den Untersuchungsein-

heiten identisch. Oft ist es jedoch angebracht, die Beobachtungseinheiten näher zu spezifizieren. Wenn etwa von mehreren Patienten das rechte und das linke Auge untersucht wird, dann versteht man unter den Untersuchungseinheiten die Patienten und unter den Beobachtungseinheiten die einzelnen Augen. Wenn Patienten im Rahmen einer Studie mehrmals untersucht werden, dann ist eine Beobachtungseinheit identisch mit einem Patienten bezogen auf eine einzelne Untersuchung.

Merkmale. Die Beobachtungseinheiten sind durch bestimmte Merkmale charakterisiert – das sind Eigenschaften, die für die betreffende Studie relevant sind und statistisch ausgewertet werden. Andere Eigenschaften der Beobachtungseinheiten sind – zumindest im Rahmen der jeweiligen Studie – uninteressant. Die Werte, die ein bestimmtes Merkmal annehmen kann, heißen *Merkmalsausprägungen*. Die Art der Merkmale ist entscheidend für den weiteren Verlauf der Untersuchung, insbesondere für den erforderlichen Stichprobenumfang und die geeigneten statistischen Analysemethoden. Deshalb sind zu Beginn der Planungsphase die zu erfassenden Merkmale genau festzulegen und deren Eigenschaften zu spezifizieren.

Bei einer medizinischen Studie werden in der Regel mehrere Merkmale erhoben. Diese werden zunächst einzeln, also unabhängig voneinander ausgewertet. Dies ist das Aufgabengebiet der univariaten Datenbeschreibung, die in Kapitel 3 behandelt wird. In vielen Fällen ist es darüber hinaus interessant, den Zusammenhang zwischen zwei Merkmalen zu untersuchen. Dieses Thema behandelt die bivariate Statistik in Kapitel 4. Mit multivariaten Analysemethoden wird der Zusammenhang zwischen mehreren Merkmalen erforscht. Diese Methoden können in diesem Buch nicht ausführlich behandelt werden. In Abschnitt 9.6 werden einige kurz vorgestellt.

2.3.2 Ziel- und Einflussgrößen

Die Merkmale lassen sich grob einteilen in Ziel- und Einflussgrößen. Der eigentliche Zweck einer Studie besteht darin, Erkenntnisse über die *Zielgrößen* zu gewinnen.

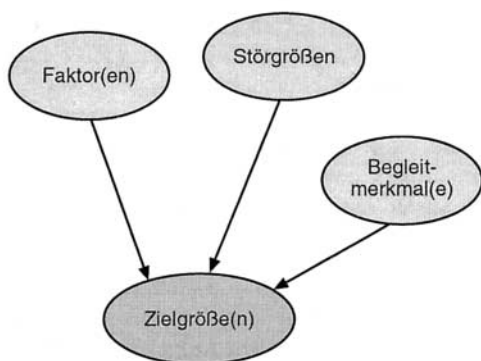


Abb. 2.2.
Einflussgrößen und
Zielgrößen

Die Merkmale, die in einem funktionalen Zusammenhang zu den Zielgrößen stehen und diese beeinflussen, heißen **Einflussgrößen**. Sie lassen sich unterteilen in:

- **Faktoren**, die erfasst und ausgewertet werden (im engeren Sinne versteht man unter den Einflussgrößen nur die Faktoren),
- **Begleitmerkmale**, die eventuell erfasst, aber im Rahmen der aktuellen Studie nicht ausgewertet werden,
- **Störgrößen**, die im Versuchsplan nicht berücksichtigt sind und deshalb nicht erfasst werden.

Beispiel 2.1

Aus der Hypothese „Zigarettenrauchen beeinflusst das Entstehen eines Lungenkarzinoms“ geht hervor, dass das Merkmal „Entstehen eines Lungenkarzinoms“ die Zielgröße ist, während „Zigarettenrauchen“ der zu untersuchende Faktor ist. Andere Einflussgrößen wie etwa das Alter der Untersuchungseinheiten wird man im allgemeinen auch erfassen und – falls diese Größen nicht explizit ausgewertet werden – als Begleitmerkmale behandeln. Zu den Störgrößen zählen genetische Veranlagungen, Umweltbelastungen etc. – also Merkmale, die ebenfalls das Entstehen eines Lungenkarzinoms beeinflussen, aber nicht explizit erfasst werden.

Anmerkung. Störgrößen können *nicht-verzerrend* oder *verzerrend* sein. Die nicht-verzerrenden sind verantwortlich für die zufallsbedingte Streuung der Versuchsergebnisse. Die verzerrenden sind gefährlicher: Sie werden fälschlicherweise als Faktoren angesehen und können dadurch ein Ergebnis verfälschen oder zu Fehlinterpretationen verleiten. Sie sind jedoch bei einer sorgfältigen Versuchsplanung vermeidbar (siehe Abschnitt 10.2).

2.3.3 Klassifikation nach Skalenniveaus

Jedes Merkmal lässt sich einem bestimmten *Skalenniveau* zuordnen. Dieses Niveau gibt Auskunft darüber, wie die entsprechenden Daten weiterverarbeitet werden können.

Nominalskala. Sie hat das niedrigste Niveau; die Ausprägungen unterscheiden sich nur begrifflich voneinander. Beispiele stellen die Augenfarbe, die Haarfarbe oder die Blutgruppe dar. Eine Spezialform bilden die *Alternativmerkmale* (die auch als *dichotome* oder *binäre Merkmale* bezeichnet werden) mit nur zwei Ausprägungen. So ist etwa das Geschlecht mit den Ausprägungen „männlich“ und „weiblich“ ein Alternativmerkmal, ebenso der Rhesusfaktor mit den Ausprägungen „positiv“ und „negativ“.

Ordinalskala (oder Rangskala). Sie besitzt ein höheres Niveau als die Nominalskala; die Ausprägungen dieser Merkmale lassen sich in einer Rangfolge anordnen. Ein bekanntes Beispiel bilden die Zensuren mit den Ausprägungen 1 bis 6. Auch medizinische Scores sind ordinalskaliert, ebenso das Merkmal Therapieerfolg mit den möglichen Abstufungen „vollständig geheilt“ bis hin zu „Patient verstorben“.

Nominal- und ordinalskalierte Merkmale werden zusammenfassend als *qualitative* (oder *kategoriale*) *Merkmale* bezeichnet. Es ist allgemein üblich, diese Merkmale zahlenmäßig zu codieren. So kann das Geschlecht einer Person durch die Zahlen 1 (für männlich) oder 2 (für weiblich) angegeben werden; der Therapieerfolg lässt sich mit natürlichen Zahlen 0, 1, 2 ... beschreiben. Diese Zahlen haben jedoch keine numerische Bedeutung. Man kann zwar zwei Ausprägungen A und B eines nominalen Merkmals durch $A = B$ oder $A \neq B$ miteinander in Beziehung setzen; bei einem ordinalen Merkmal lässt sich zusätzlich eine der Relationen $A = B$, $A < B$ oder $A > B$ angeben. Mathematische Operationen wie beispielsweise die Bildung einer Differenz oder eines Quotienten sind jedoch sinnlos. Es leuchtet ein, dass bei qualitativen Merkmalen weder der Abstand zwischen zwei Ausprägungen noch deren Verhältnis definiert ist.