
2.1 Überblick

Im Kap. 2 werden Sie die Bedeutung von statistischen Grundbegriffen wie Stichprobe oder Merkmal kennenlernen und verschiedene Skalenniveaus voneinander abgrenzen. Zunächst wird Ihnen die Betrachtung der Skalenniveaus abstrakt vorkommen, im weiteren Verlauf werden Sie sehen, dass die frühe Prüfung des Skalenniveaus bedeutsam dafür ist, welche statistischen Methoden Sie einsetzen dürfen und welche nicht.

2.2 Merkmale, Merkmalsträger, Merkmalsausprägung, Grundgesamtheit, Stichprobe

Statistik dient dazu, Ausschnitte der Welt¹ zu erfassen und auf Kennzahlen zu verdichten. Der damit verbundene Informationsverlust wird dadurch versüßt, dass entweder verschiedene Aspekte hinsichtlich dieser Kennzahlen² oder gleiche Aspekte zu verschiedenen Zeitpunkten³ verglichen werden können. Insbesondere wenn Menschen Fakten und Zusammenhänge beurteilen sollen⁴, ist eine Verdichtung zu Kennzahlen mit Hilfe der Statistik unerlässlich.

¹ Das sind insbesondere Ausschnitte der Wirklichkeit, die sich durch Zahlen beschreiben lassen, also z. B. die Anzahl von Waren in einem Lager.

² Z. B. der Vergleich verschiedener Warensorten im Lager.

³ Z. B. Anzahl der Warenart Klopapier zum Zeitpunkt 1. Januar und zum Zeitpunkt 1. Februar.

⁴ Wenn Sie Betriebswirt sind, kann Ihnen das bei der Unternehmensbewertung oder beim Lesen einer Bilanz passieren, wenn Sie Soziologe sind bei z. B. der Betrachtung von Arbeitslosenzahlen.

Merkmalsträger

Merkmalsträger sind die Objekte⁵, die hinsichtlich verschiedener Eigenschaften Gemeinsamkeiten aufweisen und daher zusammen betrachtet werden. Diese Eigenschaften dienen dazu, die Objekte vom Rest der Wirklichkeit, der nicht betrachtet werden soll, abzugrenzen.

Abgrenzung von Merkmalsträgern – sachlich, zeitlich, räumlich

Sachliche Abgrenzung: Beschreibung der Eigenschaften, die die betrachteten Merkmalsträger gemeinsam haben. Damit wird der Beobachtungsgegenstand spezifiziert und von Objekten, die nicht betrachtet werden sollen, abgegrenzt.

Zeitliche Abgrenzung: Der Zeitraum, der für die Beobachtung gilt, wird angegeben.

Räumliche Abgrenzung: Oft betrachtet man nur die Merkmalsträger, die innerhalb eines Areals⁶ zu finden sind.

Beispiel für die Abgrenzung von Merkmalsträgern

Es sollen alle männlichen Studierenden des Fachbereichs Sozialwesen an der Hochschule Niederrhein im Sommersemester 2013 betrachtet werden:

- Sachliche Abgrenzung: alle männlichen Studierenden des Fachbereichs Sozialwesen an der Hochschule⁷
- Zeitliche Abgrenzung: Sommersemester 2013
- Räumliche Abgrenzung: Hochschule in Neustadt

Oft werden die Merkmalsträger durchnummeriert und bekommen den Buchstaben i als Index.

Merkmal, Variable⁸

Merkmale bzw. Variablen sind Eigenschaften, die im Blickpunkt einer statistischen Untersuchung stehen und die den Merkmalsträgern zugeordnet werden können. Beispiele wären Haarfarbe oder die BAföG-Höhe. Merkmale werden mit Großbuchstaben gekennzeichnet, z. B. X .

⁵ Der Begriff Objekt steht hier nicht nur für Dinge. Häufig sind die betrachteten Objekte auch Menschen.

⁶ Areal kann ein Land, eine Stadt, aber auch eine Hochschule sein.

⁷ Damit werden dann die Studentinnen nicht betrachtet. Ebenso wenig ist es von Belang, ob ein Student arm oder reich bzw. groß oder klein ist.

⁸ In der deskriptiven Statistik, also wenn alle Daten bekannt sind, sprechen die meisten Autoren von Merkmalen. Wenn man über diese Merkmale noch etwas herausbekommen will, z. B. in der induktiven Statistik, wird eher von Variablen gesprochen.

Beispiel Merkmal und Merkmalsausprägung

Der Student Hansi Huster ist ein Merkmalsträger und wird z. B. mit der Nummer $i = 3$ gekennzeichnet. Die Höhe des BAföG-Anspruchs ist das Merkmal, das mit X bezeichnet wird. Dann schreibt man die Merkmalsausprägung für Hansi Huster als x_3 ⁹. Allgemein ist das BAföG des i -ten Studenten x_i .

Merkmalsausprägung

Die Merkmalsausprägung ist der jeweils konkrete Wert x_i eines Merkmals X für einen bestimmten Merkmalsträger i .

Es werden häufig und in diesem Buch folgende Formelzeichen verwendet:

Merkmalsträger: i

Merkmal: X

Merkmalsausprägung: x_i

Urliste

Die Urliste ist die Sammlung aller Merkmalsausprägungen x_i . Die Urliste gibt die Merkmalsausprägungen aller Merkmalsträger wieder, es können also auch mehrere gleiche Merkmalsausprägungen vorkommen. Wenn Sie die Merkmalsträger mit jeweils gleichen Merkmalsausprägungen zusammenfassen, erhalten Sie die weiter hinten beschriebenen Häufigkeitsverteilungen.

Grundgesamtheit

Von Grundgesamtheiten spricht man, wenn alle Merkmalsträger betrachtet werden, die aufgrund der sachlichen, zeitlichen und räumlichen Abgrenzung in Frage kommen.¹⁰

Stichprobe

Oft ist es kostspielig, alle Merkmalsträger zu untersuchen. Dann begnügt man sich mit einem Teil¹¹ der Merkmalsträger. Wenn nur ein Teil der Merkmalsträger untersucht wird, spricht man nicht mehr von der Grundgesamtheit, sondern von einer Stichprobe.

Viele Berechnungen in der Statistik unterscheiden sich je nachdem, ob eine Grundgesamtheit oder eine Stichprobe betrachtet wird.¹²

Vergleich mit Tabellenkalkulation

Zum Verständnis soll noch ein Vergleich zum Tabellenkalkulationsprogramm EXCEL gezogen werden. Wenn Sie eine Tabelle anlegen, entspricht das der Grundgesamtheit.¹³ Eine

⁹ Z. B. $x_3 = 300$ €, wenn Hansi Huster 300 € BAföG erhält.

¹⁰ Z. B. wenn Sie die Eigenschaften aller Einwohner von Neustadt untersuchen. Es darf dann keiner ausgelassen werden.

¹¹ Die intelligente Auswahl einer Stichprobe ist nicht trivial. Dies soll hier allerdings nicht Gegenstand sein. Wenn Sie mehr erfahren wollen, suchen Sie im Internet z. B. nach dem Stichwort Randomisierung.

¹² Schauen Sie sich in der Formelsammlung weiter hinten z. B. die Varianz an.

¹³ Ein horizontaler Teil der Tabelle mit einigen Zeilen aber allen Spalten wäre in diesem Bild die Stichprobe.

Spalte mit Spaltenüberschrift entspricht dem Merkmal X , eine Zeile entspricht dem Merkmalsträger oder Fall i . Der Inhalt einer Zelle ist die Merkmalsausprägung x_i .

2.3 Diskrete und stetige Merkmale

Diskret

Ein Merkmal heißt diskret, wenn es eine abzählbare Menge von Merkmalsausprägungen gibt und die Merkmalsausprägungen einen deutlichen Abstand voneinander haben. Ein Kennzeichen diskreter Merkmale ist, dass mehrere Merkmalsträger die gleiche Merkmalsausprägung haben können. Diskrete Merkmale können aus stetigen Merkmalen durch Runden der Werte oder durch Klassenbildung entstehen.

Stetig

Ein Merkmal heißt stetig, wenn es zwischen zwei Merkmalsausprägungen theoretisch beliebig viele Möglichkeiten für weitere Merkmalsausprägungen gibt. Die zwei Merkmalsausprägungen können dabei beliebig dicht beieinander liegen. In den meisten Fällen gibt es keine zwei Merkmalsträger, die genau die gleiche Merkmalsausprägung haben. Die Abgrenzung zu diskreten Merkmalen ist manchmal schwierig, da aus einem stetigen Merkmal (z. B. Temperatur) durch das Messverfahren (z. B. Digitalthermometer ohne Nachkommastellen) ein diskretes Merkmal wird.

Beispiel diskrete und stetige Merkmale

Die meisten Werte, die man in der Natur messen kann, sind stetig. Viele von Menschen geschaffene Merkmalsträger haben diskrete Merkmalsausprägungen, die man zählen kann. Die folgende Tabelle stellt einige Merkmale gegenüber.

Diskretes Merkmal	Stetiges Merkmal
Temperatur, wenn mit Digitalthermometer gemessen	Temperatur, wenn mit Quecksilberthermometer gemessen
Anzahl Brötchen in einer Brötchentüte	Gewicht eines Brötchens
Konfektionsgröße	Körpergröße

2.4 Skalen und Skalenniveaus

Qualitativ nominalskaliert

Ein Merkmal, das nicht in Form messbarer Zahlenwerte vorliegt und eher eine qualitative Eigenschaft der Merkmalsträger beschreibt, heißt nominalskaliert. Eine sinnvolle Operation auf nominalskalierte Merkmale ist das Ermitteln von Häufigkeiten. Hier das arithmetische Mittel¹⁴ anzugeben wäre unsinnig. Typische Beispiele für nominalskalierte

¹⁴ Was das arithmetische Mittel ist, wird in einem späteren Abschnitt erklärt (s. Abschn. 4.5). Viele werden es aus der Schule kennen.

Merkmale wären der Vorname oder die Haarfarbe von Personen. Man kann die Anzahl der Personen angeben, die z. B. blond sind oder Hans heißen, aber blond und Hans entspricht erst einmal nicht einem messbaren Zahlenwert.

Man erkennt ein quantitativ nominalskaliertes Merkmal daran, dass man es nicht nach Größe sortieren kann.

Quantitativ rangskaliert/ordinalskaliert

Ein Merkmal, das nach der Größe seiner Merkmalsausprägungen sortiert werden kann, heißt rangskaliert oder ordinalskaliert. Typische Beispiele hierfür wären die Platzierung bei einem Marathon oder Pferderennen, da man die Reihenfolge der Personen oder Pferde angeben kann.¹⁵ Schulnoten sind auch rangskaliert. Häufig wird zwar das arithmetische Mittel von Schulnoten berechnet. Streng genommen ist das unsinnig, da damit unterstellt wird, dass eine Zwei doppelt so gut wäre wie eine Vier und eine Eins dreimal so gut wie eine Drei. Eine Art Mittelwert für rangskalierte Merkmale wie Schulnoten ist der Median.¹⁶

Quantitativ metrisch

Ein Merkmal, dessen Merkmalsausprägung man messen kann und für das Grundrechenarten sinnvoll sind, heißt metrisch skaliert. Typische Beispiele für metrische Merkmale sind Länge, Gewicht, Laufleistung, Geschwindigkeit, Temperatur.¹⁷ Dabei ist noch die Intervallskalierung und die Verhältnisskalierung zu unterscheiden.

Verhältnisskaliert

Ein metrisches Merkmal heißt verhältnisskaliert, wenn

- Verhältnisse der Merkmalsausprägungen gebildet werden können: Merkmalsausprägung x_{13} ist doppelt so groß wie x_{24} .
- Abstände zwischen Merkmalsausprägungen gebildet werden können:
 $x_{i+1} - x_{i+1}$.
- es einen natürlichen Nullpunkt¹⁸ gibt.

¹⁵ Z. B. Altersklassenplätze: Hans ist erster in der Altersklasse Männer 30 Jahre, Karl ist zweiter usw.

¹⁶ Was der Median ist, wird in einem späteren Abschnitt erklärt (s. Abschn. 4.3). Hier nur kurz: Wenn Sie die Merkmale auf einer Perlenschnur sortieren und dann in der Mitte durchschneiden, ist der Wert an der Schnittstelle der Median.

¹⁷ Es kann beispielsweise gesagt werden, dass ein Bauteil doppelt so lang ist wie ein anderes, oder dass es drei Zentimeter kürzer ist.

¹⁸ Der natürliche Nullpunkt als Unterscheidungskriterium fällt Statistikanfängern schwer und Naturwissenschaftlern leicht: Die Temperaturskala in °C ist nicht verhältnisskaliert, 10 °C ist nicht doppelt so warm wie 5 °C. Der Nullpunkt ist willkürlich gewählt – Schmelzpunkt von Eis zu Wasser.

Intervallskaliert

Ein metrisches Merkmal heißt intervallskaliert, wenn Differenzen der Merkmalsausprägungen gebildet werden können: Merkmalsausprägung x_{13} ist um 534 größer als x_{24} .

Ein Merkmal, das verhältnisskaliert ist, ist auch gleichzeitig intervallskaliert. Die meisten metrischen Merkmale, die Ihnen begegnen werden, sind verhältnisskaliert.

Beispiel für Skalenniveaus

Typische Beispiele für metrische Merkmale sind Lebensdauer von Bauteilen, Geschwindigkeit von Zügen, Höhe von Bäumen oder Füllgewicht von Dosenuppen. Weitere Beispiele im Vergleich zeigt Abb. 2.1.

Merkmalstyp	Skala	Variable oder Merkmal	Merkmalsausprägung oder Wert
qualitativ	Nominal	Parteizugehörigkeit	CDU, SPD, Grüne,...
qualitativ	Nominal	Wahrheitswert einer Aussage	Wahr, falsch
qualitativ	Nominal	Spielausgang beim Toto	0,1,2
Rang	Ordinal	Schulnoten	1,2,3,4,5,6
Rang	Ordinal	Hausnummer	...,12,14,16 ,...
Rang	Ordinal	Dienstgrade bei der Bundeswehr	Gefreiter, ..., General
quantitativ	Metrisch, Intervallskala, stetig bzw. quasi-diskret	Uhrzeit	2:00, 4:00
quantitativ	Metrisch, Intervallskala, stetig bzw. quasi-diskret	Temperatur in Grad Celsius	...,12,13,14,...
quantitativ	Metrisch, Verhältnisskala, stetig	Entfernung zwischen Wohn- und Arbeitsstätte	1 km, 1,5 km, ...
quantitativ	Metrisch, Verhältnisskala, stetig	Alkoholgehalt im Blut	0, 0,1, ...0,8,...

Abb. 2.1 Gegenüberstellung von Skalentypen

2.5 Skalenniveaureduktion

Es ist immer möglich, das Niveau einer Skala zu reduzieren. Damit ist zwar immer ein Verlust von Informationen verbunden, oft gewinnt man allerdings einen besseren Überblick. So ist ein metrisches Merkmal zugleich rangskaliert. Aus sowohl einem metrischen als auch einem rangskalierten Merkmal lässt sich durch Gruppenbildung ein nominal skaliertes Merkmal konstruieren.

Beispiel Skalenniveaureduktion

Die Geschwindigkeit von Brieftauben, gemessen über die Ankunftszeit, ist zunächst metrisch skaliert. Bei einem Wettbewerb können Sie für jede Taube i die Zeit x_i notieren.

Hält man hingegen nur fest, welche Taube als erste, als zweite usw. eintrifft, erhält man ein rangskaliertes Merkmal. Es ist nun nicht mehr festzustellen, wie schnell eine Taube ist. Lediglich können Sie angeben, ob eine Taube schneller als andere war.

Halten Sie jedoch nur noch fest, an welchem Wochentag eine Taube eingeflogen ist, haben Sie ein prinzipiell nominalskaliertes Merkmal; Mittwoch ist nicht größer als Donnerstag.

Statistik von Null auf Hundert

Mit Kochrezepten schnell zum Statistik-Grundwissen

Brell, C.; Brell, J.; Kirsch, S.

2017, XIV, 162 S. 42 Abb., Softcover

ISBN: 978-3-662-53631-5