2 Grundbegriffe

Motivation. Dieses Kapitel hat die paradigmatische (grch.: *paradeigma* → Beispiel) Erläuterung statistischer Grundbegriffe zum Gegenstand, ohne deren Kenntnis ein Verständnis der nachfolgend skizzierten statistischen Verfahren und Methoden sowie die darin eingeschlossene statistisch-methodische und sachlogische Ergebnisinterpretation nicht möglich ist.

Statistische Einheit
Eine statistische Einheit γ ist das kleinste Element in der Statistik. Eine statistische Einheit γ ist Träger von Informationen bzw. Eigenschaften, die für eine statistische Untersuchung von Interesse sind.

Anmerkungen zur statistischen Einheit
1. **Synonyme**. Merkmalsträger, statistisches Element, Beobachtungseinheit, Erhebungseinheit, engl.: *case* → Fall, Gegenstand
2. **Objekt** versus **Vorgang**. Eine statistische Einheit γ (lies: *Klein-Gamma*) kann ein reales *Objekt* (z.B. Person, Unternehmen, Kraftfahrzeug) oder ein *Vorgang* bzw. *Fall* (z.B. Verkehrsunfall, Theaterbesuch) sein. Bei Vorgangstatistiken ist stets zwischen dem einzelnen Vorgang und den daran beteiligten realen Objekten zu unterscheiden. ♦

Beispiel 2-1: Statistische Einheit als reales Objekt
Person. Im Sommersemester 2014 haben sich am Fachbereich Wirtschaftswissenschaften I der HTW Berlin in den Bachelor-Studiengängen 281 Studierende in die obligatorische Lehrveranstaltung „Statistik“ eingeschrieben. In dieser Einschreibestatistik repräsentiert eine Studentin bzw. ein Student die statistische Einheit γ, die erfassungsstatistisch als ein reales Objekt betrachtet wird. ♣

Beispiel 2-2: Statistische Einheit als Vorgang
Verkehrsunfall. In Deutschland wurden im Jahr 2012 insgesamt 299637 Verkehrsunfälle mit Personenschaden registriert. In dieser Unfallstatistik ist der einzelne Verkehrsunfall die statistische Einheit γ. Das Charakteristische am Vorgang eines Verkehrsunfalls ist, dass daran in der Regel mehrere reale Objekte (etwa verunglückte Personen und/oder beschädigte Fahrzeuge) beteiligt sind.

Museumsbesuch. Im Jahr 2011 wurden in den 537 Kunstmuseen Deutschlands 18,6 Millionen Besuche registriert. In dieser Besuchsstatistik ist der statistisch erfasste Vorgang eines einzelnen Museumsbesuches die kleinste statistische Einheit γ. Das Charakteristische am Vorgang eines Museumsbesuches ist, dass ein Museumsbesucher als ein und dieselbe Person durch Wiederholung des Vorganges eines Museumsbesuches mehrmals statistisch erfasst werden kann. In diesem Falle ist die Anzahl der Besucher bzw. der Objekte in der Regel kleiner als die Anzahl der erfassten Besuche bzw. Vorgänge. (Quelle: Statistisches Jahrbuch Deutschland und Internationales 2013, Statistisches Bundesamt, Seite 188, 591) ♣

P. P. Eckstein, *Repetitorium Statistik,* DOI 10.1007/978-3-658-05748-0_2,

Statistische Gesamtheit
Eine endliche Menge $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ wohl unterschiedener, sachlich, örtlich und zeitlich gleich abgegrenzter statistischer Einheiten γ_i heißt statistische Gesamtheit Γ vom Umfang n.

Anmerkungen zur statistischen Gesamtheit

1. **Synonyme**. Masse, Population, Kollektiv, Grundgesamtheit
2. **Abgrenzung**. Die Festlegung einer gleichen sachlichen (wer, was), örtlichen (wo) und zeitlichen (wann) Abgrenzung einer endlichen Menge $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von n Merkmalsträgern γ_i wird durch die Zielsetzung der statistischen Untersuchung bestimmt.
3. **Umfang**. Die Anzahl n der Elemente γ_i einer statistischen Gesamtheit Γ_n (lies: *Groß-Gamma*) heißt Umfang der Gesamtheit. Hinsichtlich des Umfangs einer Gesamtheit unterscheidet man zwischen endlichen und potentiell unendlichen statistischen Gesamtheiten. In der Deskriptiven Statistik werden stets nur endliche Gesamtheiten betrachtet.
4. **Spezialfälle**. Für die Bestandsanalyse ist die Unterscheidung von Bestands-, Bewegungs- und korrespondierenden Massen von Bedeutung (vgl. Kapitel 10). ♦

Beispiel 2-3: Statistische Gesamtheit
Gesamtheit. Gemäß Beispiel 2-1 bildet die Menge $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ der Studierenden γ_i, die sich am Fachbereich Wirtschaftswissenschaften I im Sommersemester 2014 in die Lehrveranstaltung „Statistik“ eingeschrieben haben, die statistische Gesamtheit Γ_n. Ihr Umfang umfasst n = 281 Studierende. Die statistische Gesamtheit ist wie folgt abgegrenzt: i) *sachlich*: Studierende, die sich in den Bachelor-Studiengängen in die Lehrveranstaltung „Statistik“ eingeschrieben haben, ii) *örtlich*: am Fachbereich Wirtschaftswissenschaften I der HTW Berlin und iii) *zeitlich*: im Sommersemester 2014. ♣

Statistisches Merkmal
Eine Eigenschaft einer statistischen Einheit γ, die Grundlage bzw. Gegenstand einer statistischen Untersuchung ist, heißt statistisches Merkmal.

Anmerkungen zum statistischen Merkmal

1. **Synonyme**. Variable, Erhebungs-, Erfassungs- bzw. Beobachtungsmerkmal
2. **Notation**. Statistische Merkmale werden in der Regel mit den großen lateinischen Endbuchstaben ... X, Y, Z bezeichnet. Die lateinischen Großbuchstaben fungieren dabei als Variablennamen.
3. **Unterscheidung**. Für die statistische Datenerfassung ist die Unterscheidung von Identifikations- und Erhebungsmerkmalen von Bedeutung.
4. **Identifikationsmerkmal**. Die eindeutige Definition und Abgrenzung (Identifikation) statistischer Einheiten erfordert die Festlegung mindestens eines sachlichen, örtlichen und zeitlichen Identifikationsmerkmals, das auf jeweils eine Ausprägung festgelegt ist.
5. **Erhebungsmerkmal**. Im Unterschied zu den Identifikationsmerkmalen variieren die Erhebungsmerkmale in ihren Merkmalsausprägungen. Erhebungsmerkmale sind der eigentliche Gegenstand einer statistischen Untersuchung, im Zuge derer die jeweiligen Ausprägungen eines oder mehrerer Erhebungsmerkmale statistisch erfasst werden. ♦

Merkmalsausprägung
Eine Aussage über ein Merkmal bzw. über eine Eigenschaft einer statistischen Einheit heißt Merkmalsausprägung.

Anmerkungen zur Merkmalsausprägung
1. **Synonyme**. Realisation, Datum (lat.: *datum* → das Gegebene), Beobachtung
2. **Notation**. Merkmalsausprägungen werden im Unterschied zu den Merkmalen in der Regel mit den jeweiligen kleinen lateinischen Endbuchstaben ... x, y, z bezeichnet. Bezeichnet z.B. X ein interessierendes Merkmal, das an n statistischen Einheiten $\gamma_i \in \Gamma_n$ einer statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,\ldots,n\}$ beobachtet wird, dann kann eine beobachtete Merkmalsausprägung formal durch die folgende Zuordnung beschrieben werden: Jeder statistischen Einheit $\gamma_i \in \Gamma_n$ der Ordnung i einer Gesamtheit Γ_n wird durch die Abbildung $X: \gamma_i \in \Gamma_n \rightarrow x_i = X(\gamma_i) \in \Xi$ eine Merkmalsausprägung x_i zugeordnet.
3. **Zustandsmenge**. Die Merkmalsausprägung $X(\gamma_i) = x_i$ ist ein Element bzw. eine Teilmenge der sogenannten Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\}$ (lies: *Groß-Xi*) aller m theoretisch möglichen bzw. aller m empirisch beobachteten und wohl voneinander unterschiedenen Merkmalsausprägungen ξ_j.
4. **Skalen**. Der Begriff einer statistisch beobachteten und mittels einer Zustandsmenge definierten Merkmalsausprägung führt unmittelbar zum statistischen Skalenbegriff. ♦

Beispiel 2-4: Identifikationsmerkmale
Gesamtheit. Gemäß Beispiel 2-3 bildet die Menge $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ der $n = 281$ Studierenden γ_i, die sich im Sommersemester 2014 in den Bachelor-Studiengängen in die Lehrveranstaltung „Statistik“ eingeschrieben haben, die statistische Gesamtheit Γ_n.

Abgrenzung. Die statistische Gesamtheit ist inhaltlich wie folgt abgegrenzt: i) *sachlich*: Studierende, die sich in den Bachelor-Studiengängen in die Lehrveranstaltung „Statistik“ eingeschrieben haben, ii) *örtlich*: Fachbereich Wirtschaftswissenschaften I der HTW Berlin, iii) *zeitlich*: Sommersemester 2014. Beachtenswert ist dabei, dass die Identifikationsmerkmale für eine statistische Gesamtheit in ihren Ausprägungen festgelegt sind und daher nicht variieren. ♣

Beispiel 2-5: Erhebungsmerkmale
Ist man in Anlehnung an das Beispiel 2-4 an der Analyse der geschlechtsspezifischen Altersstruktur der statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ der $n = 281$ eingeschriebenen Studierenden interessiert, dann bilden das Alter X und die Geschlechtszugehörigkeit Y die interessierenden Erhebungsmerkmale der statistischen Einheit γ_i „StudentIN“ der Ordnung i. Die jeweils beobachteten Ausprägungen $X(\gamma_i) = x_i$ bzw. $Y(\gamma_i) = y_i$ der Erhebungsmerkmale X bzw. Y sind im Zuge einer statistischen Untersuchung zu erfassen und zu analysieren. ♣

Statistische Skala
Eine relationstreue Abbildung von Merkmalsausprägungen eines Erhebungsmerkmals auf eine Zeichen- bzw. Zahlenmenge heißt statistische Skala.

Anmerkungen zur statistischen Skala

1. **Semantik**. Eine Skala (lat., ital.: *scala* → Treppe, Leiter) ist (vereinfacht ausgedrückt) eine Art „Messlatte" für die Ausprägungen eines statistischen Erhebungsmerkmals.
2. **Bedeutung**. Die Anwendung statistischer Analyseverfahren hängt entscheidend von der Skala ab, auf der die Ausprägungen eines statistischen Merkmals erfasst wurden.
3. **Typen**. In der Deskriptiven Statistik kommt vor allem den folgenden fünf hierarchisch (grch.: *hieros* → heilig + *archein* → herrschen) geordneten Skalentypen eine besondere praktische und theoretische Bedeutung zu: der nominalen, der ordinalen und der metrischen Skala in Gestalt einer Intervall-, Verhältnis- oder Absolutskala. ♦

Nominalskala
Eine Skala, mit der lediglich die Gleichartigkeit oder die Verschiedenartigkeit von Merkmalsausprägungen eines Erhebungsmerkmales zum Ausdruck gebracht werden kann, heißt Nominalskala.

Anmerkungen zur Nominalskala

1. **Hierarchie**. Die Nominalskala (lat.: *nominalis* → zum Namen gehörig, begrifflich) ist in der Statistik die niedrigstwertige Skala mit dem niedrigsten Informationsgehalt und der geringsten Fehlerempfindlichkeit.
2. **Adjektiv**. Ein statistisches Merkmal, dessen Ausprägungen mit Hilfe einer Nominalskala erfasst werden, heißt nominal skaliertes oder nominales Merkmal.
3. **Ausprägung**. Statistisch erfasste Ausprägungen $x_i = X(\gamma_i) \in \Xi$ eines nominalen Erhebungsmerkmals X, die ein Element der zugehörigen Zustandsmenge Ξ sind, werden auch als Kategorien oder Attribute bezeichnet.
4. **Dichotomie**. Ein nominales Merkmal, das nur zwei mögliche Ausprägungen besitzt, heißt dichotom (grch.: *dicha* → zweifach + *tome* → Schritt).
5. **Häufbarkeit**. Ein nominales Merkmal heißt häufbar, wenn sich auf ein und dieselbe statistische Einheit γ mehrere Ausprägungen eines Erhebungsmerkmals „häufen" können. Ansonsten heißt es nicht häufbar. Der Häufbarkeitsbegriff ist wohl zu unterschieden vom Häufigkeitsbegriff (vgl. Abschnitt 4.1). ♦

Beispiel 2-6: Nominale Merkmale

Merkmale. Das Geschlecht, der Familienstand, die Religionszugehörigkeit, die Nationalität oder der Beruf sind nominale Merkmale einer Person.

Zustandsmenge. Ist für eine statistische Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von n Personen γ_i das Erhebungsmerkmal X: *Familienstand* von Interesse, dann ergibt sich das folgende Bild: Die Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\} = \{\xi_1 = \text{ledig}, \xi_2 = \text{verheiratet}, \xi_3 = \text{geschieden}, \xi_4 = \text{verwitwet}\}$ (lies: *Groß-Xi*) für das Erhebungsmerkmal X ist in der amtlichen Statistik durch m = 4 voneinander verschiedene Merkmalsausprägungen ξ_j (lies: *Klein-Xi*) gegeben.

Merkmalsausprägung. Erfasst man das Merkmal X für alle statistischen Einheiten $\gamma_i \in \Gamma_n$ einer Gesamtheit Γ_n, so kann man mittels der n statistisch erfassten Merkmalsausprägungen $x_i \in \Xi$, wobei z.B. $X(\gamma_1) = x_1 =$ *verheiratet*, $X(\gamma_2) = x_2 =$ *ledig*, $X(\gamma_3) = x_3 =$ *verheiratet*, $X(\gamma_4) = x_4 =$ *geschieden*, ... gelten soll, lediglich eine Gleichartigkeit oder eine Verschiedenartigkeit der betrachteten Personen

$\gamma_i \in \Gamma_n$ bezüglich des statistischen Erhebungsmerkmals X: *Familienstand* mit seinen beobachteten Ausprägungen $x_i \in \Xi$ statistisch beschreiben. Während z.B. die statistischen Einheiten $\gamma_i \in \Gamma_n$ der Ordnung i = 1 und i = 3 bezüglich des Erhebungsmerkmals X gleichartig sind, kennzeichnet man im paarweisen Vergleich die statistischen Einheiten der Ordnung i = 1 und i = 2 bezüglich des erfassten Familienstands X als verschiedenartig. ♣

Beispiel 2-7: Dichotomes Merkmal

Geschlecht. Das Geschlecht ist ein nominales und dichotomes Merkmal X einer Person. Dies erklärt sich daraus, dass die Zustandsmenge $\Xi = \{\xi_j, j = 1, 2\} = \{\xi_1 = \text{männlich}, \xi_2 = \text{weiblich}\}$ lediglich aus den beiden (theoretisch und praktisch) möglichen Merkmalsausprägungen *männlich* oder *weiblich* besteht. Während eine Person als statistische Einheit $\gamma \in \Gamma_n$ fungiert, bildet die Menge aller für eine Erhebung interessierenden Personen eine statistische Gesamtheit Γ_n, die im konkreten Fall auch als dichotome Gesamtheit bezeichnet wird. ♣

Beispiel 2-8: Häufbares versus nicht häufbares Merkmal

Der Beruf ist ein häufbares nominales Merkmal einer Person. Dies erklärt sich daraus, dass eine Person mehrere Berufe besitzen bzw. ausüben kann. Dem gegenüber ist das Geschlecht ein nicht häufbares Merkmal einer Person. ♣

Ordinalskala

Eine Skala, mit der sowohl die Gleichartigkeit oder die Verschiedenartigkeit als auch eine natürliche Rangfolge von Merkmalsausprägungen eines Erhebungsmerkmals zum Ausdruck gebracht werden kann, heißt Ordinalskala.

Anmerkungen zur Ordinalskala

1. **Adjektiv**. Ein statistisches Merkmal, dessen Ausprägungen auf einer Ordinalskala erfasst werden, heißt ordinal skaliertes oder ordinales Merkmal. In praxi werden ordinale Ausprägungen meist begrifflich und sprachlich mittels des Komparativs, also der ersten Steigerungsstufe eines Adjektivs, dargestellt.
2. **Ordnungsprinzip**. Bei einem ordinalen (lat.: *ordinare* → ordnen) Merkmal, dessen Ausprägungen sich nach der Intensität unterscheiden, ist das Ordnungsprinzip die Stärke bzw. der Grad der Intensität. Dies ist ein Grund dafür, warum man (meist begriffliche) Ausprägungen $x_i \in \Xi$ eines ordinalen Merkmals X auch als *Intensitäten* oder *Prädikate* bezeichnet und diese in der Regel mit Hilfe von Rangzahlen kodiert.
3. **Rangzahl**. Mit Hilfe (in der Regel) natürlicher Zahlen kodierte (begriffliche) Ausprägungen eines ordinalen Merkmals heißen Rangzahlen. Rangzahlen, auch Rangwerte genannt, bilden z.B. die Grundlage des Rangkorrelationskoeffizienten nach SPEARMAN und des MANN-WHITNEY-Tests, die vor allem in der empirischen Wirtschafts- und Sozialforschung eine breite Anwendung erfahren (vgl. Abschnitt 6.2 und Abschnitt 20.3).
4. **Applikation**. Die Ordinalskala findet bei der statistischen Deskription und Analyse wirtschafts- und sozialwissenschaftlicher Sachverhalte (z.B. Qualitäts- und Leistungsmerkmale, Prädikate, sozialer Status) eine breite Anwendung. ♦

Beispiel 2-9: Ordinale Merkmale

Prädikat. Das Prädikat eines Studienabschlusses mit seinen in einer Rahmenprüfungsordnung festgelegten und die Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\}$ bildenden, m = 5 (theoretisch möglichen und) voneinander verschiedenen Ausprägungen ξ_1 = ausgezeichnet, ξ_2 = sehr gut, ξ_3 = gut, ξ_4 = befriedigend oder ξ_5 = bestanden ist ein ordinales Merkmal X eines Hochschulabsolventen $\gamma_i \in \Gamma_n$ einer (endlichen) statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von Hochschulabsolventen. $X(\gamma_i) = x_i \in \Xi$ bezeichnet eine statistisch erfasste Ausprägung des ordinalen Merkmals X: *Prädikat* für den Hochschulabsolventen $\gamma_i \in \Gamma_n$ der Ordnung i.

Konfektionsgröße. Die Konfektionsgröße X ist ein ordinales Merkmal einer statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von n Personen $\gamma_i \in \Gamma_n$. Die Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\}$ ist durch die m = 6 voneinander verschiedenen Ausprägungen ξ_1 = *eXtra Small*, ξ_2 = *Small*, ξ_3 = *Medium*, ξ_4 = *Large*, ξ_5 = *eXtra Large* und ξ_6 = *eXtra eXtra Large* gegeben. $X(\gamma_i) = x_i \in \Xi$ bezeichnet eine statistisch beobachtete Ausprägung des ordinalen Merkmals X: *Konfektionsgröße* für eine Person $\gamma_i \in \Gamma_n$ der Ordnung i.

Tabellenplatz. Der Tabellenplatz X nach einem Kegelabend ist ein ordinales Merkmal einer statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von n Kegelbrüdern und Kegelschwestern $\gamma_i \in \Gamma_n$. Die Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\}$ ist durch die $m \le n$ (in der Regel) wohl voneinander verschiedenen Ausprägungen ξ_1 = *Erster*, ξ_2 = *Zweiter*, ... gegeben. $X(\gamma_i) = x_i \in \Xi$ bezeichnet eine statistisch erfasste Ausprägung des ordinalen Merkmals X: *Tabellenplatz* für einen Kegelbruder bzw. für eine Kegelschwester $\gamma_i \in \Gamma_n$ der Ordnung i. ♣

Kardinalskala

Eine Skala, die mit Hilfe der Menge der reellen Zahlen sowohl die Gleich- oder die Verschiedenartigkeit und die Rangfolge als auch mess- und zählbare Unterschiede (Abstand, Vielfaches) für Merkmalsausprägungen eines Erhebungsmerkmals zum Ausdruck bringen kann, heißt Kardinalskala.

Anmerkungen zur Kardinalskala

1. **Synonyme**. metrische Skala, Hauptskala
2. **Adjektiv**. Ein statistisches Merkmal, dessen Ausprägungen auf einer Kardinalskala (lat.: *cardinalis* → im Angelpunkt stehend, hauptsächlich) gemessen werden, heißt kardinal skaliertes oder kardinales bzw. metrisch skaliertes oder metrisches (grch.: *metron* → Maß) Merkmal.
3. **Arten**. Eine kardinale bzw. metrische Skala wird in der statistischen Methodenlehre mitunter hierarchisch (grch.: *hieros* → heilig + *archein* → herrschen) in eine Intervall-, Verhältnis- oder Absolutskala gegliedert. ♦

Intervallskala

Eine Kardinalskala, die keinen natürlichen Nullpunkt und keine natürliche Maßeinheit besitzt, heißt Intervallskala.

Anmerkungen zur Intervallskala

1. **Hierarchie**. Die Intervallskala ist die niedrigstwertige Kardinalskala.
2. **Adjektiv**. Ein statistisches Merkmal, dessen Ausprägungen auf einer Intervallskala gemessen werden, heißt intervallskaliert.
3. **Operationen**. Für ein intervallskaliertes Merkmal ist es nur sinnvoll, zwischen den Merkmalswerten Abstände zu messen bzw. Differenzen zu berechnen und zu interpretieren. ♦

Beispiel 2-10: Intervallskaliertes Merkmal

Temperatur. In Wandlitz, Landkreis Barnim, wurden am Montag, dem 28. Oktober 2013 um 17 Uhr 12 °C gemessen. In diesem Fall ist die Temperatur ein intervallskaliertes Merkmal X einer Gemeinde γ mit dem beobachteten bzw. gemessenen Merkmalswert $X(\gamma) = x = 12$ °C entsprechend der Temperaturskala, die nach dem schwedischen Naturforscher Anders CELSIUS (*1701, †1744) benannt wurde und auf einem von CELSIUS künstlich festgelegten Nullpunkt von 0 °C beruht. Die Aussage „... heute ist es um 5 °C wärmer als gestern ...“ ist sinnvoll. Nicht sinnvoll ist die Aussage „... 12 °C sind doppelt so warm wie 6 °C“.

Zustandsmenge. Beachtenswert ist dabei, dass die Zustandsmenge Ξ des intervallskalierten Merkmals X: *Temperatur* mit Hilfe der Menge der reellen Zahlen $\mathbb{R}$ beschrieben werden kann, so dass allgemein für jede statistisch beobachtete Temperatur $X(\gamma) = x \in \mathbb{R}$ gilt. ♣

Verhältnisskala

Eine Kardinalskala, die einen natürlichen Nullpunkt, aber keine natürliche Maßeinheit besitzt, heißt Verhältnisskala.

Anmerkungen zur Verhältnisskala

1. **Adjektiv**. Ein statistisches Merkmal, das auf einer Verhältnisskala gemessen wurde, heißt verhältnisskaliert.
2. **Operationen**. Für die Merkmalswerte eines verhältnisskalierten Merkmals sind alle Vergleichs- und Rechenoperationen definiert. Die Bezeichnung selbst rührt daher, dass es für ein verhältnisskaliertes Merkmal sinnvoll ist, Verhältniszahlen (vgl. Abschnitt 9.1) zu berechnen und zu interpretieren.
3. **Applikation**. Messvorgänge basieren auf einer Verhältnisskala. Demnach sind z.B. Längen-, Flächen-, Volumen- und Gewichtsangaben in der Regel verhältnisskaliert. ♦

Beispiel 2-11: Verhältnisskalierte Merkmale

Fahrleistung. Die jährliche Fahrleistung (Angaben in km) ist ein verhältnisskaliertes Merkmal X eines Kraftfahrzeuges γ. Die Maßeinheit „Kilometer“ (grch.: *chilioi* → tausend + *metron* → Maß) ist eine durch das „Urmeter“ (ausgestellt in Sèvres bei Paris) künstlich festgelegte Maßeinheit. Der natürliche Nullpunkt wäre durch den Umstand gekennzeichnet, dass ein Kraftfahrzeug γ im Verlaufe eines Jahres keine Fahrleistung aufzuweisen hätte, für das man im konkreten Fall einen Merkmalswert $X(\gamma) = x = 0$ km statistisch beobachtet hätte. Sinnvoll ist z.B. die Aussage, dass sich im Jahr t im Vergleich zum Vorjahr t - 1 die Fahrleistung eines Kraftfahrzeuges von $x_{t-1} = 10000$ km auf $x_t = 15000$ km, absolut um $x_t - x_{t-1} = 5000$ km bzw. relativ auf das $x_t / x_{t-1} = 1{,}5$-Fache erhöht hat. Beachtenswert ist dabei, dass die Zustandsmenge Ξ des Merkmals X durch die Menge der positiven reellen Zahlen $\mathbb{R}^+$ gegeben ist, so dass für eine statistisch beobachtete Fahrleistung x eines Kraftfahrzeuges allgemein $X(\gamma) = x \in \mathbb{R}^+$ gilt.

Ausgaben. Die monatlichen Ausgaben (Angaben in €) für Theaterbesuche sind ein verhältnisskaliertes Merkmal X eines privaten Haushaltes γ. Der natürliche Nullpunkt ist dadurch charakterisiert, dass ein privater Haushalt γ im Verlaufe eines Monats keine Ausgaben für Theaterbesuche zu verzeichnen hat, also gleichsam $X(\gamma) = x = 0$ € gilt. Die Maßeinheit „Euro (€)“ ist keine natürliche, sondern eine künstlich festgelegte Geldeinheit, die offiziell mit Beginn des Jahres 2002 für die Staaten der Europäischen Union als Währungseinheit eingeführt wurde. Die Zustandsmenge Ξ des Merkmals X ist durch die Menge der reellen Zahlen $\mathbb{R}$ gegeben, so dass für statistisch beobachtete monatliche Ausgaben x für Theaterbesuche eines privaten Haushaltes allgemein $X(\gamma) = x \in \mathbb{R}$ gilt.

Preise. Der Preis ist ein verhältnisskaliertes Merkmal X eines Gutes $\gamma_i \in \Gamma_n$ eines Warenkorbes $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ (vgl. Kapitel 9). Es ist sinnvoll, einen Preisvergleich für zwei gleichartige Güter anzustellen, wenn die Güter z.B. wie folgt ausgepreist sind: $X(\gamma_1) = x_1 = 5{,}20$ € je kg und $X(\gamma_2) = x_2 = 4{,}80$ € je kg. Mit Hilfe der reellen Zahlen 5,20 und 4,80 kann ein Preisvergleich mittels der hierarchischen Skalen bewerkstelligt werden: i) *Nominalskala*. Mit der Aussage „... der Preis $X(\gamma_1) = x_1$ des Gutes γ_1 ist verschieden vom Preis $X(\gamma_2) = x_2$ des Gutes x_2 ...“ wird lediglich auf dem Niveau einer Nominalskala die Verschiedenartigkeit der Preise $x_1 \neq x_2$ zum Ausdruck gebracht. ii) *Ordinalskala*. Durch die Aussage „... das Gut γ_1 ist teurer als das Gut γ_2 ...“ wird auf dem Niveau einer Ordinalskala die Verschiedenartigkeit der Preisangaben $x_1 \neq x_2$ noch durch eine Rangfolge $x_1 > x_2$ ergänzt. Beachtenswert ist dabei, dass die Verwendung des Komparativs (lat.: *comparare* → vergleichen), also der ersten Steigerungsstufe eines Adjektivs, im Kontext eines Vergleichs stets ein Indiz für eine Ordinalskala ist. iii) *Intervallskala*. Die Aussage „... der Preis des Gutes γ_1 liegt um 0,40 € je kg über dem des Gutes γ_2 ...“ kennzeichnet auf dem Niveau einer Intervallskala die Preisdifferenz $x_1 - x_2 = 0{,}40$ € je kg. iv) *Verhältnisskala*. Der dimensionslose Quotient $x_1 / x_2 = 1{,}083$ aus den Güterpreisen x_1 und x_2, der als Preismesszahl bezeichnet wird (vgl. Abschnitt 9.3), lässt auf dem Niveau einer Verhältnisskala die folgende Aussage zu: „Der Preis $X(\gamma_1) = x_1$ des Gutes γ_1 macht das 1,083-Fache des Preises $X(\gamma_2) = x_2$ des Gutes γ_2 aus.“

Hierarchie. Aus den vier preisbezogenen Aussagen wird augenscheinlich, dass die vier verwendeten statistischen Skalen abgestuft und somit hierarchisch (grch.: *hieros* → heilig + *archein* → herrschen) sind. ♣

Absolutskala
Eine Kardinalskala, die einen natürlichen Nullpunkt und eine natürliche Maßeinheit besitzt, heißt Absolutskala.

Anmerkungen zur Absolutskala
1. **Hierarchie**. Die Absolutskala ist die höchstwertige statistische Skala.
2. **Adjektiv**. Ein statistisches Merkmal, das auf einer Absolutskala gemessen wird, heißt absolut skaliert.
3. **Applikation**. Sämtliche Zählvorgänge basieren auf einer Absolutskala. ♦

Beispiel 2-12: Absolutskala
Stückzahlen oder Anzahlen sind Ausprägungen absolut skalierter Merkmale. Die Mengenangabe 1 Stück ist im Unterschied etwa zu 1 €, 1 kg, 1 m etc. von keiner künstlich festgelegten Maßeinheit abhängig. Seit jeher benutzten die Menschen ihre (zehn) Finger als natürliche Maßeinheit beim Zählen. ♣

Diskretes Merkmal
Ein kardinales Merkmal, das in einem endlichen Intervall nur einzelne bzw. endlich viele Merkmalswerte annehmen kann, heißt diskretes Merkmal.

Anmerkungen zum diskreten Merkmal
1. **Synonyme**. diskontinuierliches oder ganzzahliges Merkmal
2. **Vorkommen**. Absolut skalierte Merkmale sind stets diskrete Merkmale. ♦

Beispiel 2-13: Diskrete Merkmale
Anzahl. Die Anzahl der Kinder ist ein absolut skaliertes und diskretes Merkmal X eines Arbeitnehmers $\gamma \in \Gamma_n$ einer Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$ von n Arbeitnehmern. Die Zustandsmenge Ξ des Merkmals X ist durch die Menge der natürlichen Zahlen $\mathbb{N}$ sowie der Zahl Null gegeben, so dass $\Xi = \{0\} \cup \mathbb{N} = \{0, 1, 2,...\}$ gilt. Für einen Arbeitnehmer $\gamma_i \in \Gamma_n$ der Ordnung i symbolisiert die Zuordnungsvorschrift $X(\gamma_i) = x_i \in \Xi$ eine statistisch beobachtete und erfasste Merkmalsausprägung, die als ein diskreter Merkmalswert $x_i \in \Xi$ definiert ist.

Gehalt. Das monatliche Nettogehalt X ist ein verhältnisskaliertes und diskretes Merkmal eines Arbeitnehmers $\gamma_i \in \Gamma_n$, weil es z.B. bei der europäischen Währung „auf Euro und Cent genau" eine (abzählbar endlich) kleinste Geldeinheit in Gestalt der kleinsten Scheidemünze „1 Cent" (lat.: *centum* → Hundert) als hundertster Teil eines Euro gibt. Die Zustandsmenge bzw. der Zustandsbereich Ξ des Merkmals X ist durch die Menge $\mathbb{R}^+$ der positiven reellen Zahlen gegeben. Für einen Arbeitnehmer $\gamma_i \in \Gamma_n$ der Ordnung i symbolisiert die Zuordnungsvorschrift $X(\gamma_i) = x_i \in \mathbb{R}^+$ eine statistisch beobachtete Merkmalsausprägung, die z.B. wegen x_i = 1234,56 € als ein diskreter Merkmalswert aufgefasst wird.

Punkte. Die in einer Statistikklausur erreichte Punktezahl X ist ein absolut skaliertes und diskretes Merkmal eines Studierenden γ. Die Zustandsmenge Ξ des

Merkmals X ist durch die Menge der natürlichen Zahlen $\mathbb{N}$ sowie durch die Zahl Null gegeben, so dass $\Xi = \{0\} \cup \mathbb{N} = \{0, 1, 2,...\}$ gilt und z.B. $X(\gamma_i) = x_i \in \Xi$ eine durch den Klausurteilnehmer $\gamma_i \in \Gamma_n$ mit der Nummer i erreichte und statistisch erfasste Punktezahl symbolisiert.

Note. Demgegenüber stellt eine erreichte Note Y ein ordinales Merkmal eines Klausurteilnehmers $\gamma \in \Gamma_n$ dar, deren Zustandsmenge $\Xi = \{\xi_j, j = 1,2,...,m\}$ durch die m = 5 ursprünglich begrifflichen und voneinander verschiedenen Ausprägungen ξ_1 = *sehr gut*, ξ_2 = *gut*, ..., ξ_5 = *ungenügend* auf die Menge der natürlichen Zahlen $\mathbb{N} = \{1, 2,..., 5\}$ abgebildet werden und z.B. wegen $X(\gamma) = x \in \mathbb{N}$ als Merkmalswerte eines diskreten Merkmals erscheinen. Dies gilt allgemein für die *Kodierung* von nominalen oder ordinalen Merkmalen, deren begriffliche Ausprägungen vor allem zum Zwecke der leichteren Erfassung und Auswertung auf ganze Zahlen abgebildet werden. ♣

Stetiges Merkmal
Ein kardinales Merkmal, das in einem geschlossenem Intervall jeden beliebigen aller theoretisch möglichen (und potenziell unendlich vielen) Merkmalswerte annehmen kann, heißt stetiges Merkmal.

Anmerkungen zum stetigen Merkmal

1. **Synonym**. kontinuierliches Merkmal, reellwertiges Merkmal
2. **Vorkommen**. Auf Messvorgängen basierende Gewichts-, Längen-, Flächen- und Volumenangaben sind Merkmalswerte verhältnisskalierter und stetiger Merkmale.
3. **Spezialfall**: Ein diskretes Merkmal, dessen Anzahl von Merkmalswerten in einem gegebenen Intervall sehr groß ist, wird in praxi wie ein stetiges Merkmal behandelt und daher als quasi-stetig bezeichnet. ♦

Beispiel 2-14: Stetige Merkmale

Zapfmenge. Die gezapfte Tagesmenge X (Angaben in Hektolitern) an Dieselkraftstoff ist ein verhältnisskaliertes und stetiges Merkmal einer Tankstelle γ. Die Zustandsmenge Ξ des Merkmals X ist durch die Menge der positiven reellen Zahlen $\mathbb{R}^+$ gegeben. Demnach kennzeichnet zum Beispiel $X(\gamma) = x = 28{,}1050$ hl eine gezapfte und statistisch erfasste Tagesmenge Dieselkraftstoff.

Wohnfläche. Die Wohnfläche X (Angaben in m²) einer Mietwohnung γ ist z.B. wegen $X(\gamma) = x = 76{,}54$ m² ein verhältnisskaliertes und stetiges Merkmal. Die Zustandsmenge Ξ des Merkmals X ist gleichsam durch $\mathbb{R}^+$ gegeben.

Gewinn. Obgleich der Gewinn X streng genommen ein verhältnisskaliertes und diskretes Merkmal eines Unternehmens γ ist, das auf „Euro und Cent" genau angegeben werden kann, wird ein statistisch erfasster Gewinn zum Beispiel von $X(\gamma) = x = 1{,}234$ Mio. € meist wegen seiner Darstellung in einer höheren Dimension als ein quasi-stetiges Merkmal behandelt. Die Zustandsmenge Ξ des Gewinns X ist durch die Menge $\mathbb{R}$ der reellen Zahlen gegeben. ♣

Zusammenfassung

In der Tabelle 2-1 sind der Übersichtlichkeit halber die Skalen und Merkmalsklassifikationen nochmals zusammengefasst, die in der Statistik üblich sind.

Tabelle 2-1: Skalen und Merkmalsklassifikationen

Skala					
Typ	Kategorial-		metrische bzw. Kardinal-		
Name	Nominal-	Ordinal-	Intervall-	Verhältnis-	Absolut-
Operation	$= \ \neq$	$= \ \neq \ > \ <$	$= \neq > < + -$	$= \ \neq \ > \ < \ + \ - \ \cdot \ /$	
Beispiel	Geschlecht	Prädikat	Temperatur	Umsatz	Anzahl
Merkmal					
Art	qualitativ		quantitativ		
Skalierung	nominal	ordinal	kardinal, metrisch		
Ausprägung	Kategorie		Wert		
	Begriff	Intensität	stetig	quasi-stetig	diskret
Beispiel	männlich	sehr gut	20,2 °C	1,2 Mio. €	20 Stück
Erfassbarkeit	direkt bzw. unmittelbar		indirekt bzw. mittelbar		
Beispiel	Körpergröße		Intelligenz		
Beispiel	männlich	sehr gut	12,34 °C	1,2 Mio. €	20 Stück

Anmerkungen zur Tabelle 2-1

1. **Operationen**. Die Symbole, die in der Rubrik „Operation“ aufgeführt wurden, kennzeichnen die für die jeweilige Skala definierten und aus statistisch-methodischer Sicht zulässigen und sinnvollen Vergleichs- und Rechenoperationen.

2. **Erfassbarkeit**. Im Blickwinkel der Erfassbarkeit von statistischen Erhebungsmerkmalen unterscheidet man in der Statistik zwischen direkt bzw. indirekt erfassbaren Erhebungsmerkmalen. Während zum Beispiel die Geschlechtszugehörigkeit einer Person direkt bzw. unmittelbar erfassbar ist, kann die Intelligenz einer Person (etwa mittels eines IQ-Tests) nur indirekt bzw. mittelbar erfasst bzw. gemessen werden.

3. **Extensiv** versus **intensiv**. Neben der Erfassbarkeit von statistischen Merkmalen, die skalenunabhängig auf die Klassifikation von „mittelbar erfassbar“ bzw. „unmittelbar erfassbar“ abstellt, erweist sich bei kardinalen Merkmalen die Unterscheidung zwischen extensiven und intensiven Merkmalen als bedeutungsvoll und nützlich. Extensive Merkmale bilden z.B. die Grundlage für statistische Konzentrationsanalysen (vgl. Kapitel 5).

4. **Häufbarkeit**. Vor allem im Blickwinkel der Analyse von Mehrfachantworten erweist sich die Betrachtung von häufbaren nominalen Erhebungsmerkmalen als substantiell. Ein nominales Erhebungsmerkmal X heißt häufbar, wenn sich mehr als eine der m wohl voneinander verschiedenen Merkmalsausprägungen $\xi_j \in \Xi = \{\xi_j, j = 1,2,...,m\}$ auf einen Merkmalsträger γ „häufen“ (vgl. Abschnitt 4.1).

5. **Häufigkeit**. Im Unterschied zum Häufbarkeitsbegriff subsumiert man in der Statistik die Anzahl $n(X = \xi_j)$ der Merkmalsträger γ_i einer statistischen Gesamtheit $\Gamma_n = \{\gamma_i, i = 1,2,...,n\}$, die sich auf eine zulässige Merkmalsausprägung $\xi_j \in \Xi$ „häufen“, unter dem Begriff einer absoluten Häufigkeit (vgl. Abschnitt 4.1). ♣

http://www.springer.com/978-3-658-05747-3

Repetitorium Statistik
Deskriptive Statistik - Stochastik - Induktive Statistik
Eckstein, P.P.
2014, XII, 408 S., Softcover
ISBN: 978-3-658-05747-3