Variablen, Ausprägungen und Skalen

In der beschreibenden Statistik legen wir für eine bestimmte Untersuchung jeweils eine Menge von Objekten als *Grundmenge* (oder Population, manchmal auch „Statistische Masse" genannt) fest. Wir bezeichnen alles als Objekt, was Eigenschaften hat. Ein wesentlicher Begriff der Statistik ist der Begriff der *statistischen Variable*. Beispiele für statistische Variablen sind die Variable „Geschlecht", die Variable „Einkommen", die Variable „Ausgaben für Unterhaltungselektronik", die Variable „Zufriedenheit mit einem Produkt". Variablen weisen sogenannte *Ausprägungen* auf, die auch „Werte" genannt werden. So weist die Variable „Geschlecht" die Ausprägungen „männlich" und „weiblich" auf. Die Variable „Zufriedenheit mit einem Produkt" kann z.B. die Ausprägungen „sehr zufrieden", „zufrieden", „müssig zufrieden" und „unzufrieden" aufweisen.

Formal gesehen sind statistische Variablen *Zerlegungen* der untersuchten Population. Eine Zerlegung (= Partition) einer Menge G ist eine Menge Z von nicht-leeren Mengen A_i (= Äquivalenzklassen) so dass gilt:

$$(i) : A_i \subset G$$
$$(ii) : \bigcup_{i=1}^{n} A_i = G$$
$$(iii) : A_i \cap A_j = \emptyset \text{ für } i \neq j.$$

Definition 0.1. ***„statistische Variable"***: *Sei eine Grundmenge G gegeben: Jede Zerlegung der Grundmenge G nennen wir „(statistische) Variable". Die (nichtleeren) Elemente (= Äquivalenzklassen) von Z nennen wir „Ausprägungen der Variable" oder „Werte der Variablen" (die Elemente von Z sind Mengen).* ♢

Variablen sowie deren Ausprägungen benennen wir gewöhnlich mit Namen.

Beispiel 0.2. *Population $G = \{Anton, Beat, Anna, Ruth, Christine\}$*
Mögliche Variable: $V = \{\{Anton, Beat\}, \{Anna, Ruth, Christine\}\}$ mit den Ausprägungen $\{Anton, Beat\}$ und $\{Anna, Ruth, Christine\}$. Wir benennen die Variable V mit „Geschlecht", die Ausprägung $\{Anton, Beat\}$ mit „männlich" und die Ausprägung $\{Anna, Ruth, Christine\}$ mit „weiblich". Die Grundmenge G wird im Beispiel in zwei Äquivalenzklassen zerlegt: die Menge der Männer und die Menge der Frauen. Wir beschreiben die Situation auch mit der folgenden Ausdrucksweise: „die Variable 'Geschlecht' weist zwei Ausprägungen auf, nämlich 'männlich' und 'weiblich'" oder mit „die Variable 'Geschlecht' weist zwei Werte auf, nämlich 'männlich' und 'weiblich'". (s. Abbildung 1)

♢

Der Variable V entspricht eine Äquivalenzrelation, die wir wie folgt erhalten:

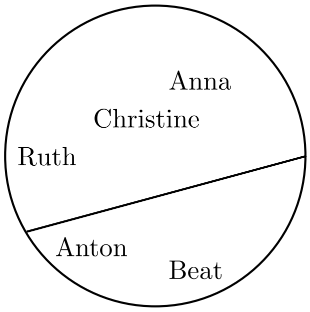
$$R_V := \bigcup_{i=1}^{n} (A_i \times A_i)$$

(A_i sind Äquivalenzklassen von Z, n ist die Anzahl der Ausprägungen der Variable). Dabei ist

$$A_i \times A_i = \{(x, y) : x, y \in A_i\}$$

das kartesische Produkt von A_i und A_i und (x, y) ein geordnetes Paar. Am obigen Beispiel erhalten wir:

R_V = {(Anton,Anton), (Beat, Beat), (Anton, Beat), (Beat, Anton), (Ruth, Ruth), (Anna, Anna), (Christine, Christine), (Ruth, Anna), (Anna, Ruth), (Anna, Christine), (Christine, Anna), (Ruth, Christine), (Christine, Ruth)}. Diese Äquivalenzrelation können wir sprachlich beschreiben, indem wir in „hat das (die, der) selbe ... wie" den Namen der Variable einsetzen. Im Beispiel erhalten wir: „hat das selbe Geschlecht wie".



statistische Variable ={{Beat, Anton},{ Ruth, Anna, Christine }}

Ausprägungen

Namen: „Geschlecht“ „männlich“ „weiblich“

Abbildung 1: Graphische und mengentheoretische Darstellung einer Variablen; Namengebung für die Variable und deren Ausprägungen

Umgekehrt erzeugt jede Äquivalenzrelation auf eine Menge G eine Zerlegung dieser Menge. Wir wählen der Reihe nach alle Komponente x in einem der geordneten Paare der Äquivalenzrelation und bilden $M_x := \{y : (x, y) \in R\}$. Die Menge der M_x bildet eine Zerlegung von G. Äquivalenzrelation und Zerlegung legen sich gegenseitig genau fest.

Beispiel 0.3. *Der Ausdruck „hat dieselbe Augenfarbe wie“ legt eine Äquivalenzrelation fest. Unterscheiden wir nur 4 Augenfarben (blau, grau, grün, braun), so erzeugt diese Äquivalenzrelation eine Zerlegung der betrachteten Grundmenge in 4 Äquivalenzklassen - sofern es Objekte gibt, die unter die vier Klassen fallen. Wir sagen in diesem Beispiel, dass die Variable „Augenfarbe“ vier Ausprägungen hat, nämlich, „blau“, „grau“, „grün“ und „braun“. Dabei müssen wir alle Augenfarben diesen vier Kategorien zuordnen können!* ♢

Für jede Variable muss gelten, da eine Variable eine Zerlegung ist: jedes Objekt der betrachteten Grundmenge muss unter genau eine der Ausprägungen der Variable fallen und kein Objekt darf unter zwei Ausprägungen der Variable fallen. So ist z.B. „ist männlich“ (als biologisches Kriterium) i.A. zweifelsfrei anwendbar. Demgegenüber wäre bezüglich „ist männlich“ (als Kriterium, dass manche Männer vor anderen „auszeichnet“) kaum zweifelsfrei entscheidbar, ob es auf ein Objekt anwendbar ist.

In der Praxis fehlen oft Daten und deshalb sind die entsprechenden Objekte bezüglich der betrachteten Variablen keinen Ausprägungen zuzuordnen. Dies ist kaum zu verhindern. Wichtig ist, dass vom Konzept her mit wirklichen Variablen gearbeitet wird.

Die Forderung, dass jedes Objekt der Grundmenge unter eine der Eigenschaften der Variable fallen muss, wird manchmal durch die Ausprägung „andere“ erfüllt (z.B. „ist verheiratet“, „ist ledig“, „andere“).

Die Grundmenge G und die Ausprägungen der Objekte von G werden gewöhnlich durch Tabellen dargestellt:

Beispiel 0.4. *Variable „Bezug von Überstunden“ mit den Ausprägungen „Bezug durch Geld“ und „Bezug durch Ferien“.*
Wir nehmen an, wir würden die folgenden Daten erhalten (s. Tabelle 1):
Eine solche Tabelle nennen wir „Urliste der Daten“. In der ersten Spalte werden die Objekte der Grundmenge durchnumeriert. In der Kopfzeile rechts der ersten Spalte stehen die Namen der Variablen. Im Beispiel gibt es nur einen Variablennamen. Jeder Variable entspricht eine Spalte. In der Spalte unter den Variablennamen stehen Namen der Ausprägungen, zu denen die Objekte bezüglich der jeweiligen Variable gehören.
Aus der Urliste der Daten kann man folgende Angaben herauslesen: die Gesamtpopulation ist:

Person	Bezug von Überstunden
1	Bezug durch Geld
2	Bezug durch Ferien
3	Bezug durch Ferien
4	Bezug durch Ferien
5	Bezug durch Geld
6	Bezug durch Geld
7	Bezug durch Ferien
8	Bezug durch Geld
9	Bezug durch Geld
10	Bezug durch Ferien
11	Bezug durch Ferien
12	Bezug durch Geld
13	Bezug durch Geld
14	Bezug durch Ferien
15	Bezug durch Ferien
16	Bezug durch Geld
17	Bezug durch Geld

Tabelle 1: Beispiel für eine Urliste von Daten

$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17\}$ *(statt uns mit Namen auf die Personen zu beziehen, verwenden wir hier Zahlen. Dabei sind nicht die Zahlen Elemente der Grundmenge, sondern die durch sie bezeichneten Personen).*
Für die Relation R_V „bezieht Überstunden auf dieselbe Art wie" erhalten wir die folgende Zerlegung: $V = \{\{1, 5, 6, 8, 9, 12, 13, 16, 17\}, \{2, 3, 4, 7, 10, 11, 14, 15\}\}$
Die Schnittmenge der Elemente der Zerlegung ist leer. Die Vereinigungsmenge dieser beiden Elemente ist mit der Gesamtpopulation identisch. ♢

Objekte können verschiedene Eigenschaften aufweisen und je nach Eigenschaften können sie bezügliche unterschiedlicher Variablen mit verschiedenen Methoden gemessen werden. Wir betrachten in der Folge die für die Statistik relevanten Messarten.

Nominal skalierte Variablen

Oft verwenden wir Ausdrücke, um Objekte zu klassifizieren, ohne weitere Relationen wie z.B. Grössenverhältnisse zwischen ihnen auszudrücken. Solche Ausdrücke stehen für sogenannte qualitative Merkmal oder Eigenschaft. Z.B. „ist weiblich", „ist männlich", „ist ledig", „ist verheiratet", „ist geschieden", „fährt mit dem Auto zur Arbeit", „geht zu Fuss zur Arbeit", „ist zuverlässig", „ist krank", „ist unberechenbar", „ist zerbrechlich", „ist teuer", „ist ein Kunde", „ist ein Mitarbeiter".

Eine Variable, welche mit ihren Ausprägungen eine Grundmenge nur klassifiziert, ohne dass die Objekte so gemessen werden können, dass weitere Relationen zwischen den Objekten der verschiedenen Ausprägungen berücksichtigt werden, wird in der Statistik „qualitative Variable", „nominal skalierte Variable" oder „kategoriale Variable" genannt, z.B. Variable „berufliche Tätigkeit" mit den Ausprägungen „ist berufstätigt", „ist nicht berufstätig".

Messen besteht im Zuordnen von Zahlen zu Objekten. Eine primitive und nicht sehr aussagekräftige Art des Messens wird oft bei qualitativen Variablen vorgenommen, indem den Objekten einer Ausprägung eine spezifische Zahl zugeordnet wird: jedem Objekt einer spezifischen Ausprägung wird dieselbe Zahl zugeordnet und Objekten verschiedener Ausprägungen werden andere Zahlen zugeordnet. Man spricht bei dieser primitiven Form des Messen oft von „Kodierung einer Variablen". z.B. In einer Tabelle wird statt „ist männlich" die Zahl 1 gesetzt, statt „ist weiblich" die Zahl 2. Die Zahlen drücken somit folgendes aus: weisen zwei Objekte dieselbe Zahl auf, so gehören sie zur selben Ausprägung. Weisen zwei Objekte verschiedene Zahlen auf, so gehören sie zu verschiedenen Ausprägungen. Somit spielt nur die mathematische Eigenschaft der Gleichheit und Ungleichheit für Zahlen bei nominalskalierten Variablen eine Rolle. Die übrigen mathematischen Eigenschaften von Zahlen spielen bei dieser Art des Messen keine Rolle: der Umstand, dass $2 > 1$, drückt nichts über das Verhältnis zweier Objekte aus, denen im Rahmen einer nominal skalierten Variable 2 und 1 zugeordnet ist. Messskalen mit den eben erwähnten Eigenschaften werden „Nominalskalen“ genannt.

Formaler können wir eine Nominalskala (bezüglich der Variable V) wie folgt definieren: Sei R_V die durch die Variable V definierte Äquivalenzrelation auf die Grundmenge G. Wir können nun definieren:

Definition 0.5. *Eine Abbildung $f : G \to \mathbb{R}$ ist eine Nominalskala auf die Grundmenge G bezüglich der Variable V genau dann, wenn:*
1. $f(x) = f(y)$ genau dann, wenn $(x, y) \in R_V$ ♢

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G. R_V ist eine Relation auf G genau dann, wenn alle Elemente von G mindestens einmal als Komponente in einem Element der Relation auftauchen).

Wir werden folgende Redewendungen verwenden: Sei f eine Skala auf G bezüglich V. Dann sagen wir, dass V eine $f-$skalierte Variable ist. Ist f eine Nominalskala, sprechen wir im Besonderen von einer „nominalskalierten Variable“.

Aus der Definition folgt, dass zwei Objekten x und y durch f dieselbe Zahl zugeordnet wird, wenn x und y zur selben Äquivalenzklasse (= Ausprägung) von V gehören.
x und y werden durch f verschiedene Zahlen zugeordnet, wenn x und y nicht zur selben Äquivalenzklasse von V gehören.

Beispiel 0.6. *Wir führen das obige Beispiel (Bezug von Überstunden) fort: Wir legen als Funktion, die die Bedingung „$f(x) = f(y)$ genau dann, wenn $(x, y) \in R_V$“, erfüllt fest:*
für $x \in \{1, 5, 6, 8, 9, 12, 13, 16, 17\} : f(x) = 1$
für $x \in \{2, 3, 4, 7, 10, 11, 14, 15\} : f(x) = 2$
In tabellarischer Form (1 = Bezug durch Geld, 2 = Bezug durch Freizeit, s. Tabelle 2):

Diese Tabelle ist mit der oben gelieferten eng verwandt. Es wurden nur die Bezeichnungen für die Ausprägungen der Variable durch die entsprechenden Skalenwerte ersetzt. Auch solche Tabellen werden wir „Urliste der Daten“ nennen. ♢

Ordinal skalierte Variablen

Während wir bei nominalskalierten Variablen auf Grund der Zahlen, die den Elementen einer Ausprägung zugeordnet sind, nur wissen können, ob zwei Objekte einer selben Ausprägung zugehören oder nicht, stellen wir bei ordinalskalierten Variablen zwischen den Objekten verschiedener Ausprägungen eine Ordnung (= ordo) her. Wir betrachten die Variable „Qualität eines Produktes“ mit den Ausprägungen „sehr gut“, „gut“, „zufriedenstellend“, „nicht zufriedenstellend“. Eine Grundmenge von Produkten würde den Ausprägungen entsprechend klassifiziert. Zwischen den klassifizierten Produkten besteht nun eine Ordnung: ein als sehr gut klassifiziertes Produkt ist besser als ein als gut, zufriedenstellend oder nicht zufriedenstellend klassifiziertes Produkt. Ein als gut klassifiziertes Produkt ist besser als ein als zufriedenstellend oder nicht zufriedenstellend klassifiziertes Produkt. Ein als zufriedenstellend klassifiziertes Produkt ist besser als ein als nicht zufriedenstellend klassifiziertes Produkt.

Auch bei ordinalskalierten Variablen muss die Zuordnung zu den Ausprägungen der Variable eindeutig sein. Zudem muss jedes Objekt der Grundmenge durch die Ausprägungen einer Variable klassifiziert werden.

Messen erfolgt bei ordinalskalierten Variablen ebenfalls durch die Zuordnung von Zahlen zu den Objekten der verschiedenen Ausprägungen. Wie bei nominalskalierten Variablen müssen die Zah-

Person	Bezug von Überstunden
1	1
2	2
3	2
4	2
5	1
6	1
7	2
8	1
9	1
10	2
11	2
12	1
13	1
14	2
15	2
16	1
17	1

Tabelle 2: Beispiel für eine Urliste von Daten mit Skalenwerten

len die Zugehörigkeit zu denselben Ausprägungen durch dieselbe Zahl ausdrücken. Verschiedene Zahlen drücken die Zugehörigkeit zu verschiedenen Ausprägungen aus. Im Gegensatz zu nominalskalierten Variablen müssen bei ordinalskalierten Variablen die Zahlen jedoch noch die Ordnung zwischen den Objekten der verschiedenen Ausprägungen ausdrücken. Im Beispiel könnten wir etwa den Objekten, die als sehr gut klassifiziert wurden, die Zahl 5 zuordnen, den Objekten, die als gut klassifiziert wurden, die Zahl 3, den Objekten, die als zufriedenstellend klassifiziert wurden, die Zahl 2.5, den Objekten, die als nicht zufriedenstellend klassifiziert wurden, die Zahl 1. Die mathematische Relation „>“ zwischen den Zahlen spiegelt nun die Relation „besser sein als “. Jede Zuordnung von Zahlen, die diese Relation spiegelt, stellt eine angemessene Messung der Objekte einer ordinalskalierten Variable dar. So wäre etwa die folgende Zuordnung zu den Objekten der oben verwendeten Ausprägungen ebenso angemessen: 1000, 100, 10, 1 oder 1000, 3, 2, 1, usw.

Wir stellen das eben gesagte nun noch etwas systematischer dar. Bei Objekten, die man mit einer Ordinalskala messen kann, muss eine Relation T definierbar sein, für die folgende Eigenschaften gelten:

Definition 0.7. *Die Abbildung $f : G \to \mathbb{R}$ ist eine Ordinalskala auf die Grundmenge G bezüglich der Variable V genau dann, wenn es eine Relation T auf G gibt, so dass gilt:*

1. *$(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$.*
2. *T ist transitiv und asymmetrisch.*
3. *$(x, y) \in T$ genau dann, wenn $f(x) > f(y)$*

$\Diamond$

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G).

Bemerkung 0.8.
- *eine Relation ist asymmetrisch, wenn für alle $(x, y) \in R$ gilt, $(y, x) \notin R$. Somit ist z.B. die Relation des Vaterseins oder des „Strikt-Grösser-Seins-Als-Jemand-Anders“ asymmetrisch*

- *Wird die obige Definition von einer Variable V erfüllt, sagen wir, dass V eine ordinalskalierte Variable ist.*
- *z.B. für „T" „ist besser als", für „R" „ist gleich gut wie". Wir legen eine beliebige Funktion f fest, für die gilt: f ordnet x genau dann eine höhere Zahl zu als y, wenn x in der Relation T zu y steht (z.B. wenn x besser als y ist). Die Zahl, die x zugeordnet ist, ist genau dann gleich der Zahl, die y zugeordnet ist, wenn x in der Relation R_V zu y steht (z.B. wenn x gleich gut ist wie y).*

$\diamondsuit$

Beispiel 0.9. *Die Angestellten einer Unternehmung werden befragt, ob Sie die Personalführung gut finden. Auf die Frage „Wie finden Sie die Personalführung" konnten Sie zwischen den folgenden Fragen wühlen: „Sehr gut", „Gut", „Recht gut", „Schlecht". Wir nehmen an, 20 Personen wären befragt worden. Wir ordnen den Personen, die die Personalführung sehr gut finden eine 4 zu, denen, die sie gut finden eine 3, denen die sie recht gut finden eine 2 und den restlichen eine 1. Wir nehmen an, wir hätten folgendes Resultat erhalten (s. Tabelle 3):* $\diamondsuit$

Personen	Personalführung (4 = „sehr gut"; 3 = „gut"; 2 = „recht gut"; 1 = „schlecht")
1	4
2	3
3	4
4	1
5	2
6	3
7	3
8	4
9	3
10	4
11	3
12	1
13	2
14	2
15	1
16	2
17	3
18	3
19	4
20	3

Tabelle 3: Mögliche Daten zu Beispiel 0.9 in Form einer Urliste.

Metrisch skalierte Variablen

Wenn die körperlichen Grössen von Personen in Metern gemessen wird, so stellt die Variable „Körpergrösse" eine metrisch skalierte Variable dar. Zeitmessung, Gewichtsmessung, Geschwindigkeitsmessung, usw. erfolgen im Allgemeinen mit metrischen Skalen. Metrische Skalen zeichnen sich durch folgende Eigenschaften aus:

Definition 0.10. *Die Abbildung $f : G \to \mathbb{R}$ ist eine metrische Skala auf die Grundmenge G bezüglich der Variable V genau dann, wenn es Relationen T und $\circ$ auf G gibt, so dass gilt:*

1. *$(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$.*
2. *T ist transitiv und asymmetrisch.*
3. *$(x, y) \in T$ genau dann, wenn $f(x) > f(y)$*
4. *$f(x \circ y) = f(x) + f(y)$ (Additionsprinzip oder Linearität)*

$\Diamond$

(R_V ist die durch die Variable V festgelegte Äquivalenzrelation auf G).

Bemerkung 0.11.

- *Wir sagen, dass V eine metrisch skalierte Variable ist.*
- *„$\circ$" ist irgend eine empirische Operation des Aneinanderfügens. Beim Messen von Längen besteht dieses etwa darin, auf einer Geraden zwei Gegenstände aneinanderzufügen. Beim Gewichtsmessen besteht dieses darin, auf eine Waage zu einem Gewicht ein weiteres hinzuzufügen. Für f muss dabei laut obiger Bedingung z.B. gelten: Wenn f dem Objekt x die Zahl 5 zuordnet und f dem Objekt y die Zahl 3 zuordnet, dann muss f den zusammengefügten Objekten $x \circ y$ die Zahl 8 zuordnen. Konkret: wenn ein Gegenstand 3 Kilo wiegt und ein weiterer Gegenstand 5 Kilo wiegt, so müssen sie zusammen 8 Kilo wiegen. Wenn ein Gegenstand 3 Meter lang ist und ein weiterer 5 Meter, so sind sie auf einer Geraden zusammengefügt 8 Meter lang.*
- *Offenbar kann man Geld metrisch messen. Da Geld etwas Abstraktes ist, stellt sich die Frage, wie hier eine entsprechende empirische Operation aussehen könnte. Wir können in der Tat ein empirisches Äquivalent für Geld finden, auf das eine solche Operation definiert werden kann. Wir können z.B. für Geldbeträge die entsprechende Menge an neuen Einfrankenstücken betrachten und dann die Gewichtsmessung verwenden. Für diese gibt es eine empirische Operation, welche die obige Anforderung erfüllt. Dies überträgt sich dann auf das Geld. Beachten Sie allerdings: obwohl man Geld metrisch messen kann, gilt dies nicht für den Nutzen, den Geld für eine Person darstellen kann!*
- *Bezüglich einer Variablen, für die eine metrische Skala definierbar ist, kann immer auch eine Ordinalskala definiert werden. Eine Ordinalskala ist jedoch weniger informativ. Nur die „$>$" Relation ist auf ihr sinnvoll interpretierbar. Bei metrischen Skalen ist jedoch zusätzlich die Operationen der Addition interpretiert, bei manchen metrischen Skalen zusätzlich die Operation der Multiplikation. So ist es sinnvoll zu sagen, eine Stange x sei dreifach so lang wie eine andere y (dies ist genau dann der Fall, wenn gilt: $f(x) = 3f(y)$). Haben wir eine ordinalskalierte Variable vor uns mit den Ausprägungen „sehr gut", „gut", „genügend", „ungenügend", so dass wir den entsprechenden Objekten 1000, 10, 5 und 1 zuordnen, hat es hingegen keinen Sinn zu sagen, ein als „sehr gut" klassifiziertes Objekt x sei 1000 mal besser als ein als ungenügend klassifiziertes Objekt y, weil $f(x) = 1000f(y)$.*
- *Metrisch skalierte Variablen können stetiger oder diskreter Natur sein. Stetige oder kontinuierliche Variablen kommen etwa bei der Längenmessung vor: in der Theorie - aber nicht in der Praxis - sind zwischen zwei Werten überabzählbar unendlich stetiger oder diskreter Natur viele Werte möglich (überabzählbar unendlich = gleichviele wie das entsprechende, beidseitig beschränkte Intervall reeller Zahlen enthält). Bei metrisch, stetigen Variablen tritt pro Ausprägung oft nur ein Objekt auf. Bei diskreten Variablen sind in einem beidseitig beschränkten Intervall nur endlich viele Skalenwerte oder abzählbar unendlich viele Werte verfügbar (abzählbar unendlich = gleichviele wie das entsprechende Intervall rationaler Zahlen). Werden nur natürliche Zahlen zugeordnet, sprechen wir von einer Absolutskala, die damit eine Unterart der diskret metrischen Skala darstellt. Absolutskalen kommen etwa vor,*

wenn man die Anzahl der Eigentumswohnungen oder die Anzahl der Kinder einer Population erfasst. Streng genommen gibt es nur diskrete Skalen, da wir in der Wirklichkeit nur auf endlich viele Stellen genau messen können. Bei Geld etwa messen wir höchstens zwei Stellen nach dem Komma. Trotzdem würden wir Einkommen als eine stetig metrisch skalierte Variable betrachten - ausser wir würden dieses z.B. in ganzen Tausendern messen. Allgemein betrachtet man eine metrisch skalierte Variable dann als stetig, wenn es viele Ausprägungen hat und die Anzahl der Skalenwerte fast oder ganz der Anzahl der Daten entspricht.

- *Bei stetig metrisch skalierten Variablen enthalten die Äquivalenzklassen jeweils nur ein Objekt, wenn wir genügend genau messen.*

- *Bei den metrischen Skalen sind weitere Unterscheidungen üblich (Intervallskalen, z.B. Temperaturmessung; Verhältnisskalen, z.B. Gewichtsmessung, Längenmessung. Verhältnisskalen haben einen absoluten Nullpunkt und die Multiplikation ist sinnvoll interpretierbar. z.B. ein Tisch ist dreimal so lang wie ein anderer). Es handelt sich um in unserem Rahmen überflüssige Unterscheidungen.*

- *Zu beachten ist, dass bezüglich der Skalenwerte die Operation der Addition und der Multiplikation unabhängig vom Skalenniveau möglich ist - sie ist aber nur bezüglich metrischer Skalenniveaus sinnvoll. So kann man zwar rechnerisch den Mittelwert bezüglich einer ordinalskalierten Variable berechnen. Dies macht aber wenig Sinn. Entscheidend ist deshalb nicht die Durchführbarkeit der Operationen, sondern das Auffinden einer empirischen Operation des Zusammenfügens, die den Operationen auf der Ebene der Zahlen entsprechen.*

- *Wer legen fest: die Ordinalskala ist eine höhere Skala als die Nominalskala und die metrische Skala ist eine höhere Skala als die Ordinalskala. Es gilt: Objekte, die auf einer höheren Skala messbar sind, sind auch auf einer tieferen Skala messbar. Dies ergibt sich durch den Umstand, dass jede höhere Skala alle Voraussetzungen von tieferen Skalen aufweist (siehe Zusammenstellung unten). Die Umkehrung gilt nicht. Nicht alle Objekte, die bezüglich einer Variable auf einer Nominalskala messbar sind, sind bezüglich dieser Variable ordinal oder metrisch messbar. Nicht alle Objekte, die bezüglich einer Variablen ordinal messbar sind, sind bezüglich dieser Variable metrisch messbar.(s. Tabelle 4)*

Eine Abbildung $f : G \to \mathbb{R}$ ist eine ..... Skala auf die
Grundgesamtheit G bezüglich der Variable V
genau dann, wenn es Relationen T und $\circ$ auf G gibt, so dass:

(1) $(x, y) \in R_V$ genau dann, wenn $f(x) = f(y)$ (R_V ist die durch die Variabale V festgelegte Äquivalenzrelation auf G).	**nominale**
(2) T ist transitiv und asymmetrisch. (3)$(x, y) \in T$ genau dann, wenn $f(x) > f(y)$. (4) $(x, y) \in T$ oder $(y, x) \in T$ genau dann, wenn $f(x) \neq f(y)$.	**ordinale**
(5) $f(x \circ y) = f(x) + f(y)$ (Additionsprinzip oder Linearität).	**metrische**

Tabelle 4: Zusammenstellung über den Zusammenhang zwischen den drei behandelten Skalen

- *Eine Messmethode wird objektiv genannt, wenn verschiedene Personen, die die Methode kennen, bei deren Anwendung zum selben Resultat gelangen. So ist etwa das Messen einer Tischlänge eine objektive Angelegenheit - wenn man eine nicht allzu grosse Genauigkeit verlangt. Die Benotung von Deutschaufsätzen oder von Kooperationsfähigkeit im Betrieb ist demgegenüber kaum objektiv, da verschiedene Spezialisten zu unterschiedlichen Beurteilungen kommen. Bezüglich der Bewertung von Deutschaufsätzen ist die mangelnde Objektivität durch empirische Untersuchungen gut bestätigt. Beachten Sie, dass Objektivität in der Statistik nicht als Übereinstimmung von Meinungen und Fakten interpretiert wird, sondern als die Produzierbarkeit identischer Resultate durch verschiedene Spezialisten (Kenner der Materie).*

- *Eine Messmethode wird valide genannt, wenn durch die Methode das gemessen wird, was man messen will. Die Messung der Schuhgrüsse in Zentimeter ist recht objektiv. Verwenden wir die Schuhgrüsse jedoch zur Messung der Intelligenz der Schuhträger, so ist die Methode kaum valide. Wir messen mit der Methode nicht das Gewünschte. Für ein weiteres Beispiel siehe in den Lösungen zu den Übungen.*

$\diamondsuit$

Beispiel 0.12. *Eine Unternehmung untersucht die Kundenstruktur. Dazu werden die Umsätze der Unternehmung mit den Kunden erhoben (in Franken). Wir erhalten z.B. die folgende Tabelle 5.*

Unternehmung	Umsatz in Fr.
1	54907.43
2	25187.51
3	61348.31
4	25001.79
5	84718.87
6	62441.16
7	48867.09
8	74686.72
9	36554.36
10	42621.76
11	46823.83
12	24619.05
13	14209.21
14	37737.41
15	71026.32
16	73230.15
17	119375.88
18	57365.19

Tabelle 5: Beispiel für eine metrisch skalierte Variable; Urliste von Daten

Die durch „hat gleichviel Umsatz wie" beschriebene Relation ist eine Äquivalenzrelation R_V auf G und sie induziert eine Zerlegung von G. Die Elemente dieser Zerlegung enthalten jeweils nur ein Element.
Die durch „hat mehr Umsatz als" beschriebene Relation T ist asymmetrisch und transitiv. Wir ordnen hier unmittelbar die Skalenwerte zu.
Zwei Kunden haben zusammen als Umsatz die Summe der Einzelumsätze. $\diamondsuit$

Datenvektoren und Datenmatrizen

Die Spalte der Urliste der Daten mit den Skalenwerten einer Variablen können wir auch als ein n-Tupel oder einen Vektor von Daten auffassen. Solche n-Tupel nennen wir künftig auch „Datenvektor“. Formellere Definition:

Definition 0.13. ***„Datenvektor der Grundmenge G“:*** *Sei $|G| = n$. Sei $y = (a_1, ..., a_n)$ ein n-Tupel der n Objekte der Grundmenge G so dass $a_i \neq a_j$ für $i \neq j$. $x = (f(a_1), ..., f(a_n))$ wird „Skalenwertvektor (der Grundmenge)“ genannt.*
(G = Grundmenge, a_i Elemente der Grundmenge. f irgend eine Messskala). ♢

Definition 0.14. ***Redewendungen:*** *Statt „Skalenwertvektor“ verwenden wir gewöhnlich „Datenvektor“ oder „Datensatz“. Die Skalenwerte $f(a_i)$ nennen wir gewöhnlich „Daten“. Statt „Skalenwerte einer nominal- oder ordinalskalierten oder metrisch skalierten Variable“ verwenden wir auch die kürzeren Redewendungen „nominal- oder ordinalskalierte Daten oder metrisch skalierte Daten“.* ♢

Betrachten wir eine Urliste von Daten mit mehreren Variablen, so können wir die Tabelle als eine Matrix von Daten auffassen. Wir nennen diese oft „Datenmatrix“.

Übungen

1. Wählen Sie zwei Variablen (eine mit zwei und eine mit drei Ausprägungen) und klassifizieren Sie damit die Menge der Mitstudentinnen und -studenten, indem Sie eine Urliste der Daten erstellen. Überprüfen Sie, ob die obigen Kriterien einer Variable beachtet wurden.

2. Überprüfen Sie, ob die Variable „Freizeitaktivitäten“ mit den vier Ausprägungen „führt Ski“, „führt Snowboard“, „führt Schlittschuh“, „andere“ die obigen Kriterien an eine Variable erfüllt: Wenn wir im Rahmen einer Marketingstudie Datenmaterial bezüglich des Freizeitverhaltens sammeln, wie müssten wir dann vorgehen, wenn wir den obigen Kriterien genügen möchten?

3. Definieren Sie für Variablen unter Übung .1 eine Nominalskala und stellen Sie die dadurch erreichten Zuordnungen in Tabellenform dar.

4. Legen Sie bezüglich Ihrer Klasse eine Variable fest, bezüglich der eine Ordinalskala definierbar ist. Bestimmen Sie die Ausprägungen der Variable. Bestimmen Sie die Elemente der Mengen, die den verschiedenen Ausprägungen zugeordnet sind. Bestimmen Sie die beiden Relationen T und R, die zur Festlegung der Ordinalskala dienen könnten. überprüfen sie, ob T transitiv und asymmetrisch ist. Überprüfen Sie, ob R eine Äquivalenzrelation ist. Legen Sie ein passende Skala f fest.

5. Legen Sie eine stetig und eine diskret metrisch skalierte Variable für die Studentinnen und Studenten Ihrer Klasse fest und überprüfen Sie die Voraussetzungen der Skala am Beispiel.

6. Wir überprüfen mit Hilfe eines Testes (Kenntnis von Wörtern), wer wie gut Französisch kann. Dabei möchten wir die Französischkenntnisse mit Hilfe der Anzahl gewussten Wörter messen. Ist diese Messmethode objektiv? Ist diese Messmethode valide? Die Ausprägungen der Variable „Französisch können“ sei die Anzahl Wörter, die jemand kann. Es ergeben sich die Daten der Tabelle 6.

7. Geben Sie bezüglich des Beispiels mit den Immobiliendaten an, welche Skalenniveaus bezüglich der einzelnen Variablen vorliegen.

8. Untersuchen Sie, ob Notenskalen in der Schule als metrische, ordinale oder nominale Skalen zu betrachten sind.

Person	Anzahl Wörter
1	5
2	1
3	8
4	2
5	9
6	10
7	3
8	4
9	8
10	1
11	2
12	11
13	3
14	4
15	2
16	8
17	7
18	5

Tabelle 6: Beispiel für Urliste der Daten einer (scheinbar) metrisch skalierten Variable.

Lösungen und Lösungshinweise

1. Sei $g(x)$ die Körpergrösse von x. Sei K die Menge der Studentinnen und Studenten x für die gilt: $g(x) \leq 1.70m$. Sei T die Menge der Studentinnen und Studenten x für die gilt: $g(x) > 1.70m$. Sei G die Gesamtmenge der Studentinnen und Studenten. Es gilt nun offenbar: $K \cup T = G$ und $K \cap T = \emptyset$.

   Sei
   C die Menge der Studentinnen und Studenten x für die gilt: x hat grüne Augen,
   B die Menge der Studentinnen und Studenten x für die gilt: x hat blaue Augen und
   A die Menge der Studentinnen und Studenten x für die gilt, x hat eine andere Augenfarbe.

   Es gilt nun offensichtlich: $C \cup B \cup A = G$ und $C \cap B = \emptyset, C \cap A = \emptyset, B \cap A = \emptyset$

   Wir nehmen an, wir würden folgende Daten erhalten, die wir in der Urliste 7 der Daten festhalten.

   Im Beispiel gibt es zwei Variablen, die je eine Spalte beanspruchen.
   Die Gesamtpopulation ist damit: $G = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$
   Der zur Relation R passende Ausdruck ist „hat dieselbe Augenfarbe wie“. Durch R_V wird die Zerlegung der Grundmenge in:
   $V = \{\{1, 3, 6, 10, 13\}, \{2, 7, 8\}, \{4, 5, 9, 11, 12\}\}$ festgelegt.

   Die Elemente der Zerlegung sind paarweise disjunkt und deren Vereinigungsmenge ist mit der Gesamtpopulation identisch.

2. Die Kriterien werden nicht erfüllt, da jemand Schlittschuhfahren und Snowboard fahren kann. Die Schnittmengen der beiden Ausprägungen wären somit nicht leer. Man könnte z.B. unterscheiden zwischen „Treibt Sport“, „Treibt nicht Sport“ oder z.B. „fährt nur Schlittschuh“, „fährt nur Snowboard“, „fährt nur Schlittschuh und Snowboard“, „fährt nur Ski“, „fährt nur Ski und Snowboard“, „fährt nur Ski und Schlittschuh“, „andere“.

Person	Augenfarbe	Körpergrösse
1	grün	< 1.70
2	blau	< 1.70
3	grün	< 1.70
4	andere	> 1.70
5	andere	> 1.70
6	grün	< 1.70
7	blau	> 1.70
8	blau	< 1.70
9	andere	< 1.70
10	grün	> 1.70
11	andere	< 1.70
12	andere	> 170
13	grün	> 1.70

Tabelle 7: Beispieldaten für eine Urliste von Daten zweier Variablen mit drei respektive zwei Ausprägungen.

3. Wir führen das Lösungsbeispiel mit der Augenfarbe und der Körpergrösse fort: Wir legen für die obigen Beispieldaten und die Variable „Augenfarbe" folgende Funktion, welche die Bedingung (1) der Definition 0.5 erfüllt, fest:

   für $x \in \{1, 3, 6, 10, 13\} : f(x) = 1$
   für $x \in \{2, 7, 8\} : f(x) = 2$
   für $x \in \{4, 5, 9, 11, 12\} : f(x) = 3$

   Durch die Relation R_V, zu der der sprachliche Ausdruck „hat dieselbe Körpergrösse wie" passt, wird die Zerlegung der Grundmenge in: $\{\{1, 2, 3, 6, 8, 9, 11\}, \{4, 5, 7, 10, 12, 13\}\}$ festgelegt. Die Elemente der Zerlegung sind paarweise disjunkt und deren Vereinigungsmenge ist mit der Gesamtpopulation identisch. Wir legen für die Variable „Körpergrösse" die folgende Funktion, welche die Bedingung (1) der Definition 0.5 erfüllt fest:

   für $x \in \{1, 2, 3, 6, 8, 9, 11\} : f(x) = 1$
   für $x \in \{4, 5, 7, 10, 12, 13\} : f(x) = 2$

   In tabellarischer Form erhält man die Tabelle 8.

   Am Beispiel sehen wir, dass dieselbe Grundmenge durch verschiedene Variablen zerlegt werden kann. In diesem Fall überlappen sich die Zerlegungen im Allgemeinen. (z.B. Variable Geschlecht mit den Ausprägungen „ist männlich" und „ist weiblich"; Variable Sportverhalten mit den Ausprägungen „treibt Sport" und „treibt keinen Sport". Die beiden Zerlegungen würden nur zusammenfallen, wenn alle Männer und nur die Männer Sport treiben, oder wenn alle Frauen und nur die Frauen Sport treiben).

4. Wir möchten wissen, wie zufrieden die Personen einer Klasse mit einer spezifischen Unterrichtsstunde sind. Dazu wird ein Fragebogen ausgeteilt mit den Ausprägungen „unzufrieden", „zufrieden", „sehr zufrieden". Wir sammeln den Fragebogen ein und erstellen z.B. die folgende Aufstellung (Urliste der Daten 9):

   Die Gesamtpopulation ist damit: $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
   Wir erhalten eine Zerlegung V durch die Relation R „hat denselben Zufriedenheitsgrad wie":
   $V = \{\{3, 4\}, \{1, 2, 9, 10\}, \{5, 6, 7, 8\}\}$.
   Für die Zerlegung gilt: die Elemente der Zerlegung sind paarweise disjunkt und deren Ver-

Person	Augenfarbe	Körpergrösse
1	1	1
2	2	1
3	1	1
4	3	2
5	3	2
6	1	1
7	2	2
8	2	1
9	3	1
10	1	2
11	3	1
12	3	2
13	1	2

Tabelle 8: Beispiel für eine Urliste von Daten mit zwei Variablen und Angabe der Ausprägungen durch Skalenwerte

Person	Zufriedenheitsgrad
1	zufrieden
2	zufrieden
3	unzufrieden
4	unzufrieden
5	sehr zufrieden
6	sehr zufrieden
7	sehr zufrieden
8	sehr zufrieden
9	zufrieden
10	zufrieden

Tabelle 9: Beispiel für Urliste der Daten einer ordinalskalierten Variable

einigungsmenge ist mit der Gesamtpopulation identisch.
Als T legen wir fest „ist zufriedener als“ (T ist transitiv und asymmetrisch).
Wir können nun die folgende Skala oder Funktion f festlegen:

für $x \in \{3, 4\} : f(x) = 1$
für $x \in \{1, 2, 9, 10\} : f(x) = 2$
für $x \in \{5, 6, 7, 8\} : f(x) = 3$

Offenbar erfüllt, wie man leicht kontrollieren kann, f die Anforderungen an eine Ordinalskala.
In tabellarischer Form ergibt sich die Tabelle 10.

5. stetig, metrisch skaliert. Wir untersuchen z.B. die Körpergrösse der Studentinnen und Studenten. Wir nehmen an, wir erhalten die folgenden Daten (s. Tabelle 11):
Wir haben die Relation „R“ mit „gleich gross wie“ und die Relation „T“ mit „grösser als“. R ist bei der vorgenommenen Interpretation symmetrisch, transitiv und reflexiv (Ist x gleichgross wie y und y gleichgross wie z, dann ist x gleichgross wie z. Ist x gleichgross wie y, so ist y gleichgross wie x. x ist gleichgross wie x). Es gibt keine zwei Personen, die genau dieselbe Grösse haben. Bei stetig, metrisch skalierten Variablen fallen bei genügend genauen

Person	Skalenwert
1	2
2	2
3	1
4	1
5	3
6	3
7	3
8	3
9	2
10	2

Tabelle 10: Beispiel für Urliste von Daten einer ordinalskalierten Variable (Angabe von Skalenwerten).

Person	Grösse
1	176.25
2	179.59
3	163.31
4	189.95
5	161.55
6	169.23
7	188.37
8	183.16
9	178.41
10	190.11
11	183.14
12	171.65
13	179.31
14	175.57
15	180.03

Tabelle 11: Beispieldaten für metrisch skalierte Variable. Urliste der Daten.

Messungen kaum je zwei Objekte in dieselbe Kategorie.
„grösser als" ist offensichtlich transitiv und asymmetrisch (wenn x grösser als y ist und y grösser als z, dann ist x grösser als z; wenn x grösser als y ist, dann ist y nicht grösser als x).
Durch das Messen der Grüssen ordnen wir den Personen unmittelbar Skalenwerte zu. Wir können entsprechend die Ausprägungen mit den Skalenwerten identifizieren. Für die Funktion f gilt offensichtlich: $f(x) = f(y)$ genau dann, wenn $(x, y) \in R$
$f(x) > f(y)$ genau dann, wenn $(x, y) \in T$
$f(x \circ y) = f(x) + f(y)$ (hat eine Person z.B. eine Grüsse von 1.76 und eine andere von 1.82, dann haben Sie zusammen - wenn man sie übereinanderstellt - eine Grüsse von 3.58).

6. 1 Wort können die Elemente der Menge $\{2, 10\}$
   2 Wörter können die Elemente der Menge $\{4, 11, 15\}$
   3 Wörter können die Elemente der Menge $\{7, 13\}$
   4 Wörter können die Elemente der Menge $\{8, 14\}$
   5 Wörter können die Elemente der Menge $\{1, 18\}$
   7 Wörter können die Elemente der Menge $\{17\}$
   8 Wörter können die Elemente der Menge $\{3, 9, 16\}$

9 Wörter können die Elemente der Menge $\{5\}$
10 Wörter können die Elemente der Menge $\{6\}$
11 Wörter können die Elemente der Menge $\{12\}$

Die Relationen sind: „T“ für „kann mehr Wörter als“ und „R“ für „kann ebenso viele Wörter wie“.
Die paarweisen Schnittmengen dieser Mengen sind leer. Die Vereinigungsmenge dieser Menge ergibt die Gesamtpopulation. Zudem ist R transitiv, symmetrisch und reflexiv (Kann x ebenso viele Wörter wie y und y ebenso viele Wörter wie z, so kann x ebenso viele Wörter wie z. Kann x ebenso viele Wörter wie y, so kann y ebenso viele Wörter wie x. x kann ebenso viele Wörter wie x).
T ist offenbar transitiv: Wenn eine Person mehr Wörter kann als eine zweite und diese zweite mehr Wörter kann als eine dritte, so kann die erste mehr Wörter als die dritte. T ist auch asymmetrisch: Es gibt keine Person, die mehr Wörter kann als eine andere, so dass letztere mehr Wörter kann als die erste.
Wir ordnen den Objekten der Elemente der Zerlegung als Skalenwerte die Anzahl der gekonnten Wörter zu:

für $x \in \{2, 10\} : f(x) = 1$	für $x \in \{4, 11, 15\} : f(x) = 2$
für $x \in \{7, 13\} : f(x) = 3$	für $x \in \{8, 14\} : f(x) = 4$
für $x \in \{1, 18\} : f(x) = 5$	für $x \in \{17\} : f(x) = 7$
für $x \in \{3, 9, 16\} : f(x) = 8$	für $x \in \{5\} : f(x) = 9$
für $x \in \{6\} : f(x) = 10$	für $x \in \{12\} : f(x) = 11$

Es gilt nun: ist $f(x) > f(y)$, dann gilt $(x, y) \in T$, d.h. x kann mehr Wörter als y.
und ist $f(x) = f(y)$, dann gilt $(x, y) \in R$, d.h. x kann ebenso viele Wörter wie y.

Sei $\circ$ die Operation des Zusammenzählens von gekonnten Wörtern. Dann gilt: $f(x) + f(y) = f(x \circ y)$

D.h. Hat eine Person 3 gekonnte Wörter und eine andere 4 gekonnte Wörter, haben sie zusammen 7 gekonnte Wörter. Eine Tabelle mit den Skalenwerten statt der Anzahl Wörter unterscheidet sich nicht von der obigen Tabelle.

Bezüglich des Beispiels kann man ein paar kritische Fragen stellen: ist wirklich die Kenntnis aller Wörter gleichgewichtig für die Messung der Französischkenntnisse. Sollten die Wörter unterschiedliches Gewicht haben, müsste man das Gewicht metrisch messen können, sonst sind die Französischkenntnisse mit Hilfe von Wörterkenntnissen nicht mal ordinal messbar (die Kenntnis der Wörter a, b und c könnten dann weniger wertvoll sein, als die Kenntnisse des Wortes d. Dann sagt jedoch die Anzahl der gekonnten Wörter als solche nichts mehr über die Französischkenntnisse aus). Das Zählen der gewussten Wörter ist deshalb nicht eine valide Methode, um die Französischkenntnisse zu messen - sie wäre allerdings objektiv.

7. 1.PRICE=Sellingprice($hundreds) (metrisch skaliert, und zwar diskret, da nur Hunderter berücksichtigt werden)
2.SQFT=Squarefeet of livingspace (stetig metrisch skaliert)
3.AGE=Age of home (years) (diskret, metrisch skaliert)
4.FEATS=Number out of 11 features (dishwasher, refrigerator, microwave, disposer, washer, intercom, skylight(s), compactor, dryer, handicapfit, cableTVaccess (diskret, metrisch skaliert, sofern man davon ausgeht, dass alle Maschinen oder Eigenschaften dasselbe Gewicht haben)
5.NE=Located in north east sector of city(1) or not(0) (nominal skaliert; wenn „located in north east sector of city“ einen höheren Wert ergibt als nicht dort zu liegen, kann man auch

von einer Ordinalskala sprechen).
6.COR=Cornerlocation(1)or not(0) (nominal skaliert; siehe jedoch Bemerkung zu 5)
7.TAX=Annualtaxes($) (stetig, metrisch skaliert).

8. Notenskalen stellen am ehesten Ordinalskalen dar. Für eine metrische Skala müsste man eine empirische Operation $\circ$ angeben können, so dass gilt: $f(x \circ y) = f(x) + f(y)$. Eine solche Operation scheint jedoch nicht angebbar zu sein. Anderseits drücken Notenskalen nicht nur eine Nominalskala aus, da im Allgemeinen eine Ordnungsrelation T definierbar ist, die transitiv und asymmetrisch ist. Wir interpretieren T als „ist besser als". und R als „ist gleich gut wie". Damit werden die Bedingungen bezüglich der Skala erfüllt. Andererseits ist zu beachten, dass bei Noten oft Punkte zusammengezählt werden, ohne zu garantieren, dass diese Punkte wirklich dasselbe Gewicht haben. Entsprechend wären Noten nicht einmal ordinalskaliert, wie ein Beispiel zeigen kann: Wir nehmen an, für Aufgaben werden je ein Punkt verteilt, die bezüglich der effektiven Leistung das folgende Gewicht hätten:

Aufgabe	1	2	3	4	5	Summen
Punkt	1	1	1	1	1	5
Gewicht	0.75	0.5	1.5	1.75	1	5.5
Student 1 Punkte	1	1	0	0	1	3
Student 1 Gewichte	0.75	0.5	0	0	1	2.25
Student 2 Punkte	0	0	1	1	0	2
Student 2 Gewichte	0	0	1.5	1.75	0	3.25

Student 1 würde eine bessere Note erhalten, obwohl er bezüglich effektiver Leistung weniger erreicht hätte. Es geht um ein Validitätsproblem (s. o.). Es gibt Methoden, um das Problem zu lösen (z.B. Rasch-Modell, IRT-Methoden).

Lernziele

- Bei konkreten Variablen unterscheiden können, ob sie nominal, ordinal oder metrisch (diskret oder stetig) skaliert sind.
- Die Begriffe „Variable", „Ausprägung einer Variable", „Skalenwert" korrekt verwenden können.
- Die Begriffe „Objektivität" und „Validität" bezüglich Messverfahren definieren können.
- Auf eine Grundmenge bezüglich einer spezifischen Variable eine angemessene Skala definieren können.

Bemerkung

Verfasst von Paul Ruppen, Statistiksupport der PH Wallis, 20. September 2021