

Überblick

In diesem Kapitel geht es um die Ausprägungen von Merkmalen und um die Skalen, auf denen sie dargestellt werden: Nominale, ordinale Skalen, Intervallskalen und Verhältnisskalen. Der zweite Abschnitt behandelt die Häufigkeiten, mit denen die Ausprägungen in einer Messreihe vorkommen, und ihre übersichtliche graphische Darstellung.

Zur Beschreibenden Statistik gehört auch das Kap. 3, für dessen Inhalt ein eigenes Kapitel angemessen ist, und für zwei miteinander zusammenhängenden Merkmale das Kap. 15.

2.1 Merkmale und ihre Ausprägungen

2.1.1 Grundgesamtheit

In der Biologie führt man Beobachtungen oder Versuch an *gleichartigen* Untersuchungsobjekten durch. Diese bilden die sogenannte *Grundgesamtheit*. Gleichartig bedeutet dabei, dass die Untersuchungsobjekte in einer Reihe von *Merkmalen* „hinreichend gut“ übereinstimmen. Was dies letztlich bedeutet, hängt vom Versuchsziel ab und muss in einer sorgfältigen Versuchsplanung festgelegt werden. Untersucht werden dann ein oder mehrere Merkmale, in denen sich die Untersuchungsobjekte in der Regel unterscheiden.

2.1.2 Beispiel

Eine gewisse Menge von Menschen wird als gleichartig angesehen. Beobachtet werden die Merkmale Haarfarbe, Augenfarbe oder die Körperlänge. Bei letzterem Merkmal wird man als gleichartig verlangen, dass die Menschen etwa alle das gleiche Alter haben oder dass es lauter Erwachsene sind.

Die Versuchsplanung wird nach biologischen, chemischen, technischen und anderen außermathematischen Gesichtspunkten durchgeführt. In der Mathematik befassen wir uns mit der Frage, wie das Beobachtungsmaterial sinnvoll verarbeitet wird. Der erste Punkt dabei ist die Aufstellung einer *Skala* für die Versuchsausgänge.

2.1.3 Nominale Skala

Ein Merkmal kommt bei den Versuchsobjekten in gewissen *Ausprägungen* vor, z. B.:

Versuchsobjekte: Menschen

Merkmal: Haarfarbe

Ausprägungen: blond, braun, schwarz, rot, grau

Die mathematische Forderung für die Aufstellung einer solchen Liste dabei ist, dass genau eine Ausprägung des Merkmales für jeden Versuchsausgang vorliegen muss. Nicht zulässig wäre in diesem Beispiel:

- blond, braun (für manche Objekte keine dieser Ausprägungen vorliegend)
- blond, hell, braun, schwarz, grau, rot, dunkel (für manche Objekte lägen mehrere vor)

Im übrigen ist der Experimentator frei, je nach Zweckmäßigkeit eine solche Liste aufzustellen und dabei etwa nach folgenden Gesichtspunkten vorzugehen:

- nach dem vorliegenden Datenmaterial oder
- nach dem ins Auge gefassten Ziel der Untersuchungen.

Will man z. B. die Menschen auf ihre Empfindlichkeit gegen UV-Strahlen testen, wäre auch folgende Skala der Haarfarbe denkbar: blond, nicht blond

Eine solche Liste von Ausprägungen heißt eine *namensmäßige* oder eine *nominale Skala*. Kennzeichen einer nominalen Skala ist also, dass *jedem Versuchsausgang in eindeutiger Weise* eine Marke, d. h. ein Name auf der Skala zugeordnet ist; durch diese werden die Beobachtungsergebnisse klassifiziert, wobei es auf die Reihenfolge der Messmarken auf der Skala nicht ankommt (nur beim Anschreiben der Liste muss man in irgendeiner Reihenfolge vorgehen). Man denke z. B. an die Klassifizierung der Wirbeltierarten durch die nominale Skala:

Vögel, Säugetiere, Fische, usw.

2.1.4 Ordinale Skala

In der Biologie spricht man jedoch statt von einer Tierart auch von einer Tierordnung und tatsächlich ist mit der Einteilung in die Tierarten auch eine Anordnung verbunden unter dem Gesichtspunkt von höheren und niederen Tieren:

Säugetiere, Vögel, Fische, usw.

Haben die Namen einer Skala eine ganz bestimmte Ordnung, so spricht man von einer *ordinalen* Skala.

2.1.5 Diskretes Merkmal

Ein Merkmal nennt man *diskret*, wenn es nur endlich viele Ausprägungen hat, oder wenn man die Ausprägungen mit den natürlichen Zahlen nummerieren kann. (In der Mathematik spricht man von *abzählbar unendlich vielen* Ausprägungen.) Die bisher betrachteten Merkmale waren also diskret.

2.1.6 Weiteres Beispiel

Eine Krankheit wird manchmal in verschiedene Stadien eingeteilt, z. B.

Stufe 1: A = Anfangsstadium

Stufe 2: F = Fortgeschrittene Erkrankung

Stufe 3: K = Kritischer Zustand

Diese ordinale Skala ist also geordnet nach der Schwere der Erkrankung.

Auch wenn man diese Stufen mit den Zahlen 1, 2, 3 kenntlich machen kann, darf man bei *nur* ordinalen Skalen nicht darauf schließen, dass dem Stufenunterschied eine Bedeutung zukommt. Die Stufenunterschiede können gekennzeichnet sein dadurch, dass gewisse neue Symptome dazukommen. Aber ein Maß dafür ist nicht gegeben, um wieviel in Stufe 2 die Krankheit schwerer geworden ist als in Stufe 1. Die Stufendifferenz $2 - 1$ ist kein Maß für die Erschwerung. Man kann nicht sagen, dass von Stufe 1 zu Stufe 2 die gleiche Erschwerung besteht wie von Stufe 2 zu Stufe 3, nur weil die Stufendifferenzen $2 - 1$ und $3 - 2$ übereinstimmen. Man hätte die Stadien auch bezeichnen können mit Stufe 1, Stufe 10 und Stufe 100, und hätte ganz andere Stufendifferenzen bekommen. Wenn die Marken einer *nur* ordinalen Skala überhaupt durch Zahlen ausgedrückt sind, dann haben die Differenzen von Skalenwerten keine Bedeutung. Statt Differenz sagt man hier auch *Intervall* (präziser: *Intervall-Länge*).

Nun gibt es viele Erscheinungen, bei denen man eine *Zahlenskala* einführen kann, bei der die *Differenz zweier Skalenwerte sinnvoll* ist. Solche Skalen heißen

2.1.7 Intervall-Skalen

Da jetzt die Intervalle ein sinnvolles Maß sind, gehören die Intervall-Skalen zu den *quantitativen* Skalen, während nominale und ordinale Skalen nur *qualitativ* sind.

2.1.8 Beispiel: Die Celsius-Skala der Temperatur

Um die Temperatur von einem Liter Wasser um 1°C zu erhöhen, muss man immer die gleiche Energiemenge dem Wasser zuführen, gleich ob man von 10 auf 11 oder von 47 auf 48°C erhöht. Temperaturdifferenzen in $^{\circ}\text{C}$ gemessen haben also eine strenge physikalische Bedeutung. Der Nullpunkt ist ziemlich willkürlich beim Gefrierpunkt des Wassers bei einer Atmosphärendruck festgelegt.

2.1.9 Beispiel: Die Zeit

Zeitpunkte werden auf einer Intervallskala dargestellt. Es gibt kein in der Natur der Zeit begründetes Kriterium, einen Nullpunkt für die Zeitrechnung festzulegen. Er wird also stets willkürlich festgelegt. Aber Zeitdifferenzen haben eine klare Bedeutung.

In einer zahlenmäßigen Skala, in der der Nullpunkt willkürlich gewählt wurde, macht es keinen Sinn, etwa zu sagen, dass ein Skalenwert viermal so groß sei wie ein anderer; z. B. dass es an einem Sommertag bei 20°C viermal so warm sei wie an einem Wintertag mit 5°C . Sondern es ist eben dann um die Differenz von 15°C wärmer. Hätte man den Nullpunkt in einer Skala D mit der gleichen Intervalleinteilung wie der C -Skala auf -10°C gelegt, dann hätte man an dem Sommertag also $30^{\circ}D$ und an dem Wintertag $15^{\circ}D$. Bezüglich der D -Skala käme also heraus, dass es zweimal wärmer ist, ein ganz anderer Faktor.

2.1.10 Merke

Ist eine Skala *nur* eine Intervallskala, so haben Verhältnisse von Maßzahlen keine Bedeutung.

2.1.11 Verhältnisskalen

Für manche Größen gibt es Skalen, für die zusätzlich Verhältnisse von Maßzahlen eine Bedeutung haben. Solche Skalen heißen *Verhältnisskalen*.

2.1.12 Beispiel: Gewicht eines Körpers

Beim Gewicht muss der Nullpunkt nicht willkürlich gewählt werden, sondern er ist von Natur aus gegeben. Dass ein Tier das doppelte Gewicht hat wie ein anderes, hat einen Sinn, ebenso, dass ein Tier um 2 % zugenommen hat. Jedoch macht es keinen Sinn, zu sagen, dass die Zeit um 2 % vorgerückt sei.

2.1.13 Beispiel: Kelvin-Skala der Temperatur

Die Physik lehrt, dass die Temperatur eine Verhältnisskala besitzt, die Kelvin-Skala, bei der die Temperatur in ° Kelvin gemessen wird. Für 1° K sagt man auch ° absolut. Der Nullpunkt liegt bei -273°C und die Intervalleinteilung ist wie bei der C-Skala.

2.1.14 Kontinuierliches Merkmal

Ein nicht diskretes Merkmal, dessen Ausprägungen durch reelle Zahlen auf einer Intervall- oder einer Verhältnis-Skala gekennzeichnet sind, heißt ein *kontinuierliches Merkmal*. Typische Beispiele sind: Größe, Länge, Durchmesser, Gewicht, kontinuierlich gemessene Zeit (vgl. 1.2.2) etc. Bei einem kontinuierlichen Merkmal besteht die Skala also nicht mehr aus einer endlichen Liste von Ausprägungen, sondern die Ausprägungen werden normalerweise ein Intervall reeller Zahlen ausfüllen. Als Beispiel könnte das Zeitintervall von 1900 bis 2000 betrachtet werden.

2.1.15 Zusammenstellung

Skala	nominal	ordinal	Intervall-	Verhältnis-
Zur Klassifizierung	x	x	x	x
Zur Anordnung		x	x	x
Differenzen bedeutungsvoll			x	x
Nullpunkt nicht willkürlich				x
Verhältnisse sinnvoll				x

2.1.16 Versuchsreihen und Messreihen

Unternimmt man einen Versuch, bei dem ein endliches Merkmal (vgl. 2.1.5) untersucht werden soll, dann stellt man wie in Abschn. 2.1.3 zuerst eine Liste von Ausprägungen

auf:

$$(a_1, a_2, \dots, a_k)$$

Beim k -tupel der Ausprägungen sind die einzelnen Elemente a_i für $i = 1, 2, \dots, k$ verschieden voneinander. Hat man nur eine nominale Skala, so wird die Nummerierung willkürlich vorgenommen, bei ordinaler Skala wird man die Ausprägungen ihrer Ordnung entsprechend nummerieren. Wir können auch die *Menge*

$$\Omega = \{a_1, a_2, \dots, a_k\}$$

aller k Ausprägungen betrachten. Für eine klare Protokollierung kann man in der Praxis die Ausprägungen auch in Tabellenform aufschreiben:

i	1	2	...	k
a_i	a_1	a_2	...	a_k

Wir befassen uns mit einer *Versuchsreihe*, bei welcher der Versuch n -mal durchgeführt wird. Nach jedem Versuch notieren wir, welche Ausprägung aufgetreten ist, und erhalten eine *Messreihe*, die auch *Urliste* genannt wird. n wird die *Länge* oder der *Umfang* der Messreihe genannt. Sie kann einfach als n -tupel von Elementen von Ω oder in Tabellenform dargestellt werden:

j	1	2	...	n	
x_j	x_1	x_2	...	x_n	$x_j \in \Omega$

x_1 ist also die Ausprägung, die man bei der ersten Durchführung des Versuches gefunden hat, x_2 diejenige beim zweiten Versuch, usw. Im Gegensatz zum k -tupel der Ausprägungen werden die x_j i. a. nicht mehr voneinander verschieden zu sein.

2.1.17 Beispiel für endliches Merkmal

In einem Aufzuchtbecken werden aus einem Gelege Fische gezüchtet.

Grundgesamtheit: Das gesamte Gelege

Untersuchungsmerkmal: Entwicklungsstadium zu einem bestimmten Zeitpunkt

i	1	2	3	4
a_i	Ei	Larve	frei schwimmend	abgestorben

oder kurz: (E,L,f,a)

Es ist also $k = 4$. Es seien $n = 100$ Eier abgelaicht worden. Einige Wochen nach der Eiablage könnte man etwa folgende Messreihe beobachten:

j	1	2	3	4	5	6	7	...	100	
x_j	E	E	f	L	L	f	a	...	f	kürzer: (E, E, f, L, L, f, a, ..., f)

2.1.18 Beispiel für kontinuierliches Merkmal

Grundgesamtheit: Eine Menge von erwachsenen Personen

Merkmal: Körperlänge

Ausprägungen: Alle reellen Zahlen oder
alle positiven reellen Zahlen oder
alle reellen Zahlen – sagen wir – zwischen 100 und 250

Messreihe: (Einheit 1 cm) (157, 168, 172, 157, 172, 197, 168, 172)

2.2 Empirische Häufigkeitsverteilung bei endlichem Merkmal

2.2.1 Absolute Häufigkeiten

Um Aussagen über eine Messreihe zu machen, kann man zunächst einmal zählen, wie oft jede der Ausprägungen a_i ($i = 1, 2, \dots, k$) unter den x_j ($j = 1, 2, \dots, n$) vorkommt; dies hält man zweckmäßigerweise zuerst in einer Strichliste fest und notiert anschließend die Häufigkeitszahlen $h(a_i)$ der Ausprägungen a_i , die also angeben, wie oft a_i unter den x_j , $j = 1, 2, \dots, n$ auftrat. $h(a_i)$ heißt *absolute Häufigkeit* der Ausprägung a_i in der Messreihe. Siehe die ersten drei Zeilen von [2.2.5](#).

Aus den Bedingungen [2.1.16](#) an eine Liste von Ausprägungen erhalten wir folgende

2.2.2 Eigenschaften der absoluten Häufigkeiten

1. $h(a_i) \in \mathbb{N}_0 := \{0, 1, 2, 3, \dots\}$
2. $0 \leq h(a_i) \leq n$
3. $h(a_1) + h(a_2) + \dots + h(a_k) = n$

2.2.3 Definition: Relative Häufigkeit

Die *relative Häufigkeit* von a_i wird definiert durch:

$$r(a_i) := \frac{h(a_i)}{n} \quad i = 1, 2, \dots, k$$

$r(a_i)$ wird oft in % angegeben (s. 2.2.5, letzte Zeile).

2.2.4 Eigenschaften der relativen Häufigkeiten

- 1. $r(a_i) \in \mathbb{Q} := \{\text{rationale Zahlen}\}$
- 2. $0 \leq r(a_i) \leq 1$
- 3. $r(a_1) + r(a_2) + \dots + r(a_k) = 1$

2.2.5 Auflistung

Wir verwenden im Folgenden den Dezimalpunkt an Stelle des Dezimalkommas.

Ausprägungen ($k = 5$)	a_1	a_2	a_3	a_4	a_5	$n = \sum_{i=1}^k h(a_i)$
Strichliste			/		/	
Absolute Häufigkeiten	1	4	5	3	7	20
Relative Häufigkeiten	0.05	0.20	0.25	0.15	0.35	
in %	5	20	25	15	35	

2.2.6 Graphische Darstellung

Trägt man die Häufigkeit jeder Ausprägung a_i als Stab auf, dessen Länge gleich $h(a_i)$ ist, so erhält man ein *Stabdiagramm* oder *Histogramm*. Die Histogramme der absoluten und relativen Häufigkeiten für die Daten von 2.2.5 sind in Abb. 2.1 gezeichnet. Zur Zeichnung des Histogramms der relativen Häufigkeiten braucht man am Histogramm der absoluten Häufigkeiten nur eine Skalenänderung an der vertikalen Achse vorzunehmen. Man gibt daher meistens beide Histogramme durch *eine* Zeichnung wieder mit zwei Skalen an der vertikalen Achse.

2.2.7 Beispiel: Häufigkeitspolygon einer zeitlichen Entwicklung

Zu 2.1.17: In der folgenden Tabelle von absoluten Häufigkeiten ist in der ersten Spalte die Zeit in Tagen nach der Eiablage angegeben. Für jeden Tag (vom 35. bis zum 42.) ist eine Messreihe aufgestellt worden.

Tage	Ei	Larve	freischwimmend	abgestorben
35	74	0	0	26
36	56	18	0	26
37	43	30	0	27
38	15	38	18	29
39	0	21	48	31
40	0	14	54	32
41	0	0	67	33
42	0	0	67	33

Dann kann die zeitliche Entwicklung etwa der Larven durch ein Histogramm graphisch dokumentiert werden, oder auch durch ein sogenanntes *Häufigkeitspolygon* (s. Abb. 2.2).

2.2.8 Balkendiagramm und Komponenten-Balkendiagramm

Zur deutlicheren Dokumentation der Messwerte wird ein *Balken-* oder *Säulendiagramm* verwendet. Die Messreihe des 38. Tages aus den obigen Daten ist als Balkendiagramm in Abb. 2.3 dargestellt. Die Balken eines Balkendiagramms übereinander gesetzt ergeben ein Komponenten-Balkendiagramm (s. ebenfalls Abb. 2.3).

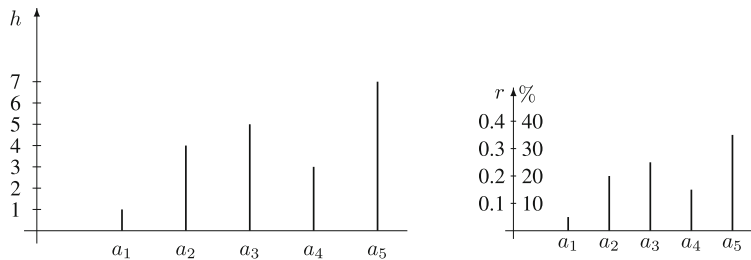


Abb. 2.1 Histogramme der absoluten und relativen Häufigkeiten

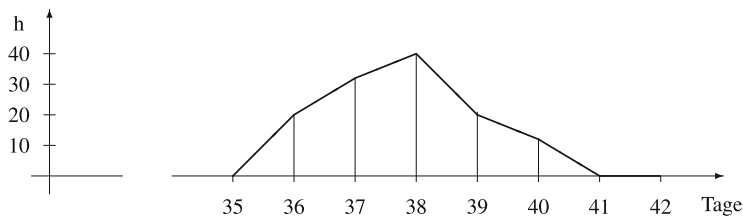


Abb. 2.2 Häufigkeitspolygon der Larvenentwicklung

Abb. 2.3 Balkendiagramm
und Komponenten-
Balkendiagramm

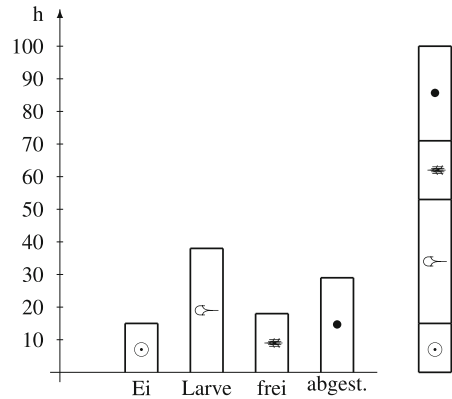
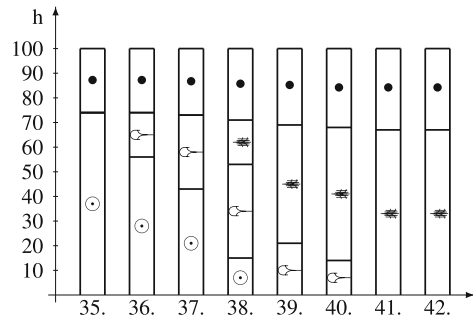


Abb. 2.4 Komponenten-
Balkendiagramme vom 35. bis
42. Tag



Durch mehrere Komponenten-Balkendiagramme für aufeinanderfolgende Zeitpunkte kann eine zeitliche Entwicklung dokumentiert werden (s. Abb. 2.4).

Statt eines Komponenten-Balken-Diagrammes für die relativen Häufigkeiten kann man auch ein „Kuchendiagramm“ verwenden, die man so häufig sieht, dass hier nicht weiter auf sie eingegangen wird. Als letzten Häufigkeitsbegriff behandeln wir die *Summenhäufigkeit*, die ein weiteres Werkzeug zur Beschreibung von Daten liefert.

2.2.9 Beispiel: Zeitliche Entwicklung von Larven

Aus einem Aufzuchtbecken werden aus einem Gelege die geschlüpften Larven jeden Tag herausgefangen, gezählt und in ein gesondertes Becken gesetzt. Es mögen sich folgende Zahlen ergeben haben: Vor dem 35. und nach dem 43. Tag seien keine Larven geschlüpft. Das beobachtete Merkmal ist jetzt der Zeitpunkt des Schlüpfens, hier gemessen in Tagen nach der Eiablage, die sogenannte *Zeitigungsdauer*.

a_i	35	36	37	38	39	40	41	42	43	$k = 9$
$h(a_i)$	0	6	18	36	36	20	8	4	0	$n = 128$

Mathematik für Biowissenschaftler

Grundlagen mit Schwerpunkt Statistik für den Bachelor

Riede, A.

2015, XII, 296 S., Softcover

ISBN: 978-3-658-03686-7