

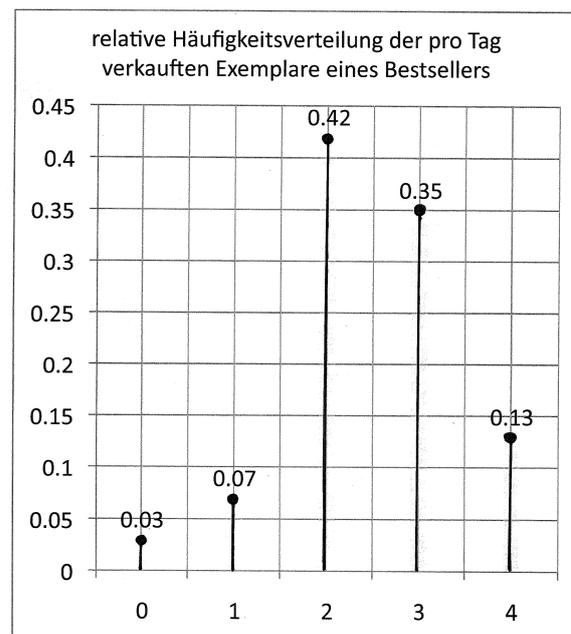
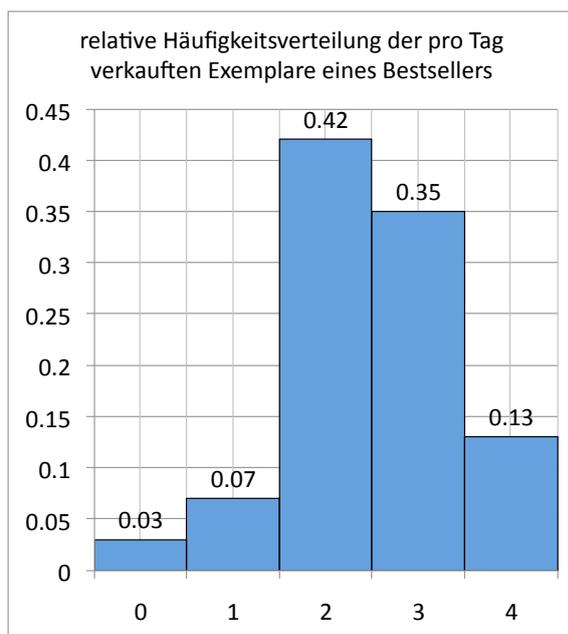


Angewandte Statistik (für HRG)

(Modul: „Modellieren und Anwendungen“)

Kerstin Hesse

Universität Paderborn, Sommersemester 2016



Dieses Skript wurde im Wintersemester 2014/15 und Sommersemester 2015 erstellt, basierend auf den handschriftlichen Notizen der Autorin von der Vorlesung „Angewandte Statistik (für HRGe)“ im Sommersemester 2014. Die Autorin dankt Frau Nurhan Sulak-Klute für ihre Mitarbeit beim Latexen des Skripts.

Paderborn, April 2016

Kerstin Hesse

Einleitung

Dieses Skript bietet eine **Einführung in die „Angewandte Statistik“** für Studierende des Hauptschul-, Realschul- und Gesamtschullehramts sowie des Grundschullehramts mit Vertiefung Mathematik. Die Vorlesung ist eine Wahlpflichtveranstaltung in dem **Modul „Modellieren und Anwendungen“**.

Obwohl diese Veranstaltung laut Studienplan im sechsten Semester des Bachelorstudiengangs gehört wird und somit nicht die erste Statistik/Stochastik-Vorlesung der Studierenden des Hauptschul-, Realschul- und Gesamtschullehramts ist, werden **keine** Statistik- bzw. Stochastik-Vorkenntnisse vorausgesetzt. Beginnend mit Merkmalen und Häufigkeitsverteilungen werden alle grundlegenden Themen eingeführt. Daher kann diese Vorlesung im Bachelor auch ohne Probleme bereits vorgezogen im vierten Semester des Bachelorstudiengangs gehört werden.

Die Vorlesung wird **teils mit dem Beamer, teils an der Tafel und teils mit Folien und Handouts** gehalten. Dabei werden insbesondere die Beispiele mit Folien und Handouts gemeinsam in der Vorlesung per Hand durchgerechnet. Die Vorlesung verzichtet bewusst auf den Einsatz eines Statistik-Programms, da es zum echten Verständnis statistischen Methoden erforderlich ist, diese zunächst einmal per Hand mit allen Schritten selber durchzuführen.

Einen ganz wichtigen Aspekt der Vorlesung stellen die **Übungszettel** dar, **deren Lösen einen zentralen und unverzichtbaren Bestandteil** des Lernens der statistischen Themen und Methoden darstellt. Diese werden durch eine zweistündige Übung unterstützt, in der einerseits ein betreutes Arbeiten an den Übungsaufgaben stattfindet und andererseits problematische Themen und Aufgaben an der Tafel besprochen werden. Für eine erfolgreiche Teilnahme an dieser Lehrveranstaltung ist es unabdingbar, die Übungsaufgaben selber zu lösen. Durch das Studium der Musterlösungen (ohne die eigenständige Lösung oder zumindest einen ernsthaften Lösungsversuch) kann man die Inhalte der Vorlesung nicht erfolgreich lernen.

Auf der nächsten Seite finden Sie zwei Lehrbücher, die ich als Literatur für diese Vorlesung empfehlen möchte. Das vorliegende Skript orientiert sich im Aufbau und in der Präsentation an diesen beiden Büchern, und viele Beispiele im Skript und viele Übungsaufgaben wurden von Beispielen und Aufgaben in diesen beiden Büchern inspiriert.

Ich freue mich auf Ihre Teilnahme an der Vorlesung und hoffe, dass Sie dabei ebenso viel Spaß haben wie ich!

Literatur:

- [B12] Josef Bleymüller: Statistik für Wirtschaftswissenschaftler (16. Auflage). Verlag Franz Vahlen, München, 2012. (Vorige Auflage: Josef Bleymüller, Günther Gehlert, Herbert Gülicher: Statistik für Wirtschaftswissenschaftler (15. Auflage). Verlag Franz Vahlen, München, 2008.)
- [W11] Max C. Wewel: Statistik im Bachelor-Studium der BWL und VWL. Pearson Studium, München, 2006.

Inhaltsverzeichnis

1	Einführung: Grundbegriffe und Fragestellungen	1
1.1	Deskriptive und induktive Statistik	1
1.2	Grundbegriffe der Statistik	2
1.3	Skalen von Merkmalen	3
1.4	Vorgehensweise bei einer statistischen Untersuchung	5
2	Empirische Verteilungen	7
2.1	Häufigkeitsverteilung eines diskreten metrisch skalierten Merkmals	7
2.2	Häufigkeitsverteilung eines metrisch skalierten Merkmals mit Klassenbildung	11
3	Mittelwerte	17
3.1	Arithmetisches Mittel	17
3.2	Median (oder Zentralwert)	23
3.3	Modus	25
3.4	Geometrisches Mittel	27
4	Streuungsmaße	29
4.1	Die Varianz und Standardabweichung	29
4.2	Spannweite, Quartile und Box-and-Whisker-Plot	40
5	Zweidimensionale Häufigkeitsverteilungen, Korrelation und Kovarianz	49
5.1	Zweidimensionale Häufigkeitsverteilung	50
5.2	Randverteilung und bedingte Häufigkeiten	52
5.3	Unabhängig verteilte Merkmale	57
5.4	Korrelation, Kovarianz und Korrelationskoeffizient	59
6	Lineare Regression	67
6.1	Problemstellung der Regression	67
6.2	Methode der kleinsten Quadrate	69
6.3	Streuungszerlegung und Bestimmtheitsmaß	79

7	Wahrscheinlichkeitsrechnung: Grundlagen	85
7.1	Zufallsexperimente und Ereignisse	85
7.2	Wahrscheinlichkeiten	90
7.3	Additionssatz für Wahrscheinlichkeiten	93
8	Wahrscheinlichkeitsrechnung: Weiterführende Resultate	97
8.1	Bedingte Wahrscheinlichkeiten, stochastisch unabhängige Ereignisse und der Multiplikationssatz	97
8.2	Theorem der totalen Wahrscheinlichkeit	106
8.3	Theorem von Bayes	108
9	Kombinatorik	111
9.1	Kombinationen n -ter Ordnung	111
9.2	Herleitungen der Formeln für die Kombinationen n -ter Ordnung von N Elementen	116
9.3	Berechnung von Wahrscheinlichkeiten mit Kombinatorik	119
10	Zufallsvariablen	123
10.1	Einführung: Zufallsvariablen	123
10.2	Wahrscheinlichkeitsfunktion und Verteilungsfunktion einer diskreten Zu- fallsvariablen	125
10.3	Verteilungsfunktion und Wahrscheinlichkeitsdichte einer stetigen Zufalls- variablen	130
10.4	Erwartungswert und Varianz einer diskreten Zufallsvariablen	139
10.5	Erwartungswert und Varianz einer stetigen Zufallsvariablen	141
A	Summen-Notation und Rechnen mit Summen	145

Einführung: Grundbegriffe und Fragestellungen

Diese Kapitel bietet eine kurze Einführung, in der die elementaren Grundbegriffe eingeführt und statistische Fragestellungen auf einem allgemeinen Niveau diskutiert werden.

1.1 Deskriptive und induktive Statistik

Die **Angewandte Statistik** gliedert sich in zwei Teilbereiche:

- **Deskriptive Statistik:** Darstellung von Daten in Tabellen und Diagrammen, Berechnung von Mittelwerten und Streuungsmaßen
- **Induktive Statistik:** Wahrscheinlichkeiten, Zufallsvariablen, Erwartungswerte, Verteilungen, statistisches Schätzen von der Stichprobe auf die Grundgesamtheit, Hypothesentests, Konfidenzintervalle, (lineare) Regression, statistische Modelle und Analyseverfahren, ...

Es ist ersichtlich, dass es sich bei der deskriptiven Statistik um den einfacheren Teil der Statistik handelt, der die Daten lediglich „beschreibt“ und auf einem sehr elementaren Niveau auswertet. Bei der induktiven Statistik versucht man dagegen, Schlussfolgerungen aus den Daten zu ziehen und Prognosen zu treffen. Solche Schlussfolgerungen und Prognosen können immer nur mit einer gewissen Wahrscheinlichkeit angegeben werden. Sichere Schlussfolgerungen und Prognosen sind in der Regel nicht möglich – dies ist das Wesen der Statistik: Wir betrachten **Ereignisse, die von Zufall abhängen**, und über die wir nur Prognosen mit einer gewissen **Wahrscheinlichkeit** machen können.

In der Vorlesung besprechen wir zunächst die wichtigsten Ideen der deskriptiven Statistik und werden danach die Anfänge der induktiven Statistik kennenlernen. Aufgrund der beschränkten Zeit von fünfzehn 90-minütigen Vorlesungen werden wir leider nicht alle interessanten grundlegenden Themen der induktiven Statistik besprechen können.

Es sei auch noch erwähnt, dass mit den oben angegebenen Themen die angewandte Statistik keineswegs vollständig beschrieben ist. In der empirischen Bildungsforschung, den Wirtschaftswissenschaften und vielen anderen Wissenschaften, in denen Versuchs- und Beobachtungsdaten analysiert und interpretiert werden müssen, kommen auch sogenannte „multivariate statistische Methoden“ zu Einsatz: Diese sind Methoden, mit denen der Zusammenhang zwischen verschiedenen Zufallsvariablen untersucht werden kann und mit denen Modelle über den funktionalen Zusammenhang zwischen verschiedenen Zufallsvariablen formuliert und getestet werden können. Ein einfaches solches Verfahren ist die (univariate) Regression, die wir im Verlaufe dieser Vorlesung besprechen werden.

1.2 Grundbegriffe der Statistik

In diesem Teilkapitel führen wir grundlegende statistische Begriffe ein.

Definition 1.1. (Grundgesamtheit, Element, Stichprobe, Merkmal)

- (1) Die **Grundgesamtheit** (oder **statistische Gesamtheit**) ist die Menge aller (für die Fragestellung betrachteten) **Elemente** (oder **Objekte** oder **Merkmalsträger**).
- (2) Eine **Stichprobe** ist eine nach einem Zufallsprinzip ausgewählte Teilmenge der Grundgesamtheit.
- (3) Ein **Merkmal** ist eine charakteristische Eigenschaft von Elementen der Grundgesamtheit. Man nennt die Elemente auch **Merkmalsträger**. Ein Merkmal kommt mit verschiedenen möglichen **Merkmalsausprägungen** vor.

Betrachten wir zwei Beispiele.

Beispiel 1.2. (Lebensdauer von einer Firma in 2013 produzierter Glühbirnen)

Grundgesamtheit: Menge aller in 2013 produzierten Glühbirnen einer Firma

Merkmal: Lebensdauer/Brenndauer der Glühbirne (in Stunden)

Das Merkmal Lebensdauer kann nur für eine Stichprobe getestet werden, da der Test die Glühbirne zerstört (Durchbrennen am Ende der Lebensdauer der Glühbirne).

Merkmalswerte: alle auftretenden Brenndauern in Stunden

Beispiel für eine *Stichprobe*: alle am 30.06.2013 produzierten Glühbirnen

Beispiel 1.3. (Merkmale von Einwohnern der BRD)

Grundgesamtheit: Einwohner der BRD

Merkmal 1: Geschlecht, mit den Merkmalsausprägungen: männlich, weiblich

Merkmal 2: Körpergröße (in Zentimetern), mit den Merkmalsausprägungen: alle auftretenden Körpergrößen in Zentimetern.

Beispiel einer *Stichprobe*: alle Einwohner Paderborns

Kommen wir noch einmal auf die Begriffe deskriptive Statistik und induktive Statistik zurück: Wir können **deskriptive Statistik** sowohl zur Beschreibung der Grundgesamtheit wie auch einer Stichprobe verwenden, indem wir diese durch ihre Häufigkeitsverteilung, Mittelwerte und Streuungsmaße charakterisieren. In der **induktiven Statistik** analysiert man dagegen Stichproben und schließt daraus auf z.B. die Verteilung von Merkmalswerten in der Grundgesamtheit. Aussagen über die Verteilung in der Grundgesamtheit kommen als **Aussagen mit Irrtumswahrscheinlichkeiten**.

1.3 Skalen von Merkmalen

Man kann jedem Merkmal einer Grundgesamtheit eine Skala zuordnen.

Definition 1.4. (Skala eines Merkmals)

- (1) **Merkmal auf einer Nominalskala** (auch **qualitatives Merkmal**): Die Ausprägungen des Merkmals besitzen **keine** Reihenfolge/natürliche Rangordnung.
- (2) **Merkmal auf einer Ordinalskala** (auch **komparatives Merkmal**): Die Merkmalsausprägungen haben eine **natürliche Rangordnung** mit einer „größer als“-Beziehung; aber die Abstände zwischen den Merkmalsausprägungen sind **nicht sinnvoll quantifizierbar**.
- (3) **metrisch skaliertes Merkmal auf einer Intervallskala**: Die Merkmalsausprägungen können in eine Rangordnung gebracht werden **und** die **Abstände** zwischen verschiedenen Merkmalsausprägungen sind **sinnvoll messbar**. Der Nullpunkt ist **willkürlich** festgelegt.
- (4) **metrisch skaliertes Merkmal auf einer Verhältnisskala**: Zusätzlich zu den Eigenschaften eines metrisch skalierten Merkmals auf einer Intervallskala gibt es einen **absoluten Nullpunkt**, so dass Verhältnisse von Merkmalsausprägungen (z.B. „doppelt so groß“) sinnvoll werden.

Bei metrisch skalierten Merkmalen spricht man auch von **Merkmalswerten** (statt Merkmalsausprägungen).

Wichtig ist, dass die **Skalen aufsteigend** sind, und man gibt für ein Merkmal immer nur die höchste Skala an. Mit „aufsteigend“ ist gemeint, dass z.B. ein metrisch skaliertes Merkmal auf einer Intervallskala (automatisch) auch ein Merkmal auf einer Nominalskala und ein Merkmal auf einer Ordinalskala ist.

Betrachten wir einige Beispiele.

Beispiel 1.5. (Skalen von Merkmalen)

- (a) Beispiele für Merkmale auf einer *Nominalskala*: Farbe, Geschlecht, Autokennzeichen
- (b) Beispiele für Merkmale auf einer *Ordinalskala*: Güteklassen bei Lebensmitteln, Rangplätze einer Fußballliga

- (c) Beispiele für metrisch skalierte Merkmale auf einer *Intervallskala*: Jahr der Ersteinstellung eines jeden Angestellten bei Firma X, Temperatur in Celsius
- (d) Beispiele für metrisch skalierte Merkmale auf einer *Verhältnisskala*: Körpergröße, Alter, Einkommen, Temperatur in Kelvin

Betrachten wir z.B. das Merkmal „Jahr der Ersteinstellung eines jeden Angestellten bei Firma X“. Da Jahresangaben angeordnet werden können und die Abstände in Jahren eine sinnvolle Abstandsmessung darstellen, handelt sich in jedem Fall um ein metrisch skaliertes Merkmal auf der Intervallskala. Damit handelt es sich automatisch auch um ein Merkmal auf der Ordinal- und Nominalskala; dieses muss man nicht weiter angeben. Da der Nullpunkt bei diesem Merkmal das „willkürlich“ festgelegte Jahr Null der christlichen Zeitrechnung ist (wir könnten eine andere Zeitrechnung mit einem anderen Jahr Null zugrunde legen) kann es sich nicht um ein metrisch skaliertes Merkmal auf der Verhältnisskala handeln.

Definition 1.6. (diskretes bzw. stetiges metrisch skaliertes Merkmal)

- (1) *diskretes metrisch skaliertes Merkmal*: Das Merkmal kann **nur bestimmte isolierte** (gegebenenfalls unendlich viele, aber abzählbar viele) Werte annehmen.
- (2) *stetiges (oder kontinuierliches) metrisch skaliertes Merkmal*: Das Merkmal kann **alle Werte in einem Intervall** (oder den reellen Zahlen) annehmen.

Betrachten wir auch hierzu einige Beispiele.

Beispiel 1.7. (diskretes bzw. stetiges metrisch skaliertes Merkmal)

- (a) Beispiele für *diskrete metrisch skalierte Merkmale*: Zahl der Studenten einer Universität, Jahresbruttogehalt gerundet auf ganze Euro, Anzahl der Kunden pro Tag in einem Supermarkt
- (b) Beispiele für *stetige metrisch skalierte Merkmale*: Körpergröße, Lebensalter, Temperatur

Ein stetiges Merkmal wird häufig nur **diskret erfasst**. Beispielsweise wird die Körpergröße in der Regel auf ganze Centimeter gerundet gemessen.

Betrachten wir noch einige der obigen Beispiele im Detail, und klassifizieren wir die Merkmale vollständig.

Beispiel 1.8. (Anzahl der Kunden pro Tag in einem Supermarkt)

Bei der Anzahl der Kunden pro Tag in einem Supermarkt handelt es sich um ein metrisch skaliertes Merkmal, weil die Merkmalswerte sich anordnen lassen und Abstände sinnvoll sind (z.B. „10 Kunden mehr als am Vortrag“). Es handelt sich um ein diskretes metrisch skaliertes Merkmal, weil als Werte nur natürliche Zahlen und Null vorkommen.

Wir müssen noch entscheiden, ob es sich um ein Merkmal auf einer Intervallskala oder auf einer Verhältnisskala handelt. Hier gibt es einen absoluten Nullpunkt („Keine Kunden waren an dem Tag in dem Supermarkt.“), und daher ist die Anzahl der Kunden pro Tag in einem Supermarkt ein metrisch skaliertes Merkmal auf einer Verhältnisskala.

Beispiel 1.9. (Temperatur in Grad Celsius)

Bei der Temperatur in Grad Celsius handelt es sich um ein metrisch skaliertes Merkmal, denn Abstände sind sinnvoll. Es handelt sich um ein metrisch skaliertes Merkmal auf der Intervallskala aber nicht auf der Verhältnisskala, denn der Nullpunkt ist willkürlich (es gibt auch negative Gradangaben in Celsius) und wurde als der Gefrierpunkt des Wassers unter bestimmten Bedingungen festgelegt. Das Merkmal ist stetig, denn als Temperaturangaben können beliebige Werte in einem Intervall auftreten. (Dass wir die Temperatur nicht beliebig genau messen können, spielt hierbei keine Rolle.)

1.4 Vorgehensweise bei einer statistischen Untersuchung

Zuletzt sollen noch kurz die Schritte einer statistischen Untersuchung vorgestellt werden.

- (1) **Planung:** Formulierung der Frage und des Untersuchungsziels, genaue Planung der Datenerhebung und der statistischen Auswertung
- (2) **Erhebung:** schriftliche Befragung (Fragebogen), mündliche Befragung (Interview), Beobachtung, Experiment, automatische Erfassung (z.B. durch Scannerkasse), ... oder Zugriff auf bereits vorliegende Daten (z.B. amtliche Statistiken)
- (3) **Aufbereitung des Datenmaterials:** Tabellen, Schaubilder, Verdichtung der Daten → **deskriptive Statistik**
- (4) **Analyse:** Auswertung und Analyse der Daten mit mathematischen/statistischen Verfahren → **induktive Statistik**
- (5) **Interpretation:** Ergebnisse der Analyse werden interpretiert und in Aussagen zusammengefasst.

Hier sind einige Beispiele für statistische Untersuchungen.

Beispiel 1.10. (Statistische Untersuchungen)

- (a) Bewertung eines neuen Produkts durch eine Kundenbefragung
- (b) Studienerfolg im Abhängigkeit vom Lernverhalten von Studierenden (durch Befragung zum Lernverhalten und Erhebung von Daten über den Studienerfolg)
- (c) Nachfrage nach den Absolventen der verschiedenen Studienfächer auf dem Arbeitsmarkt

Empirische Verteilungen

In diesem Kapitel betrachten wir (beobachtete) **Häufigkeitsverteilungen**, sogenannte **empirische Verteilungen**, weil hier nur die beobachteten Daten geeignet beschrieben werden. „Empirie“ bedeutet Erfahrung oder Erfahrungswissen. Wir postulieren in diesem Kapitel noch keine Verteilungstypen (z.B. eine angenäherte Normalverteilung oder Gleichverteilung), dieses wäre bereits induktive Statistik.

2.1 Häufigkeitsverteilung eines diskreten metrisch skalierten Merkmals

In diesem Teilkapitel betrachten wir immer:

- eine Grundgesamtheit mit N Elementen (Merkmalsträgern)
- ein **einziges diskretes metrisch skaliertes** Merkmal X mit k (möglichen) Merkmalswerten $x_1 < x_2 < \dots < x_k$ auf dieser Grundgesamtheit

Wir betrachten zunächst ein Beispiel, an dem wir neue statistische Begriffe kennenlernen.

Beispiel 2.1. (Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers.

Grundgesamtheit: 100 Beobachtungstage, also $N = 100$

Merkmal X = „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“

$x_i =$ Merkmalswert: Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers. (Hier treten 5 Merkmalswerte auf: $x_1 = 0$, $x_2 = 1$, $x_3 = 2$, $x_4 = 3$ und $x_5 = 4$.)

h_i = Anzahl der Tage, an denen genau x_i Exemplare des Bestsellers in der Buchhandlung verkauft wurden (*absolute Häufigkeit* des Merkmalswertes x_i)

i	x_i	h_i	f_i	Prozentual	H_i	F_i
1	0	3	0,03	3%	3	0,03
2	1	7	0,07	7%	10	0,1
3	2	42	0,42	42%	52	0,52
4	3	35	0,35	35%	87	0,87
5	4	13	0,13	13%	100	1,00
Σ		100	1,0	100%		

Wir berechnen aus diesen Daten:

$f_i = \frac{h_i}{N} = \frac{h_i}{100}$ = Anteil der Tage, an denen genau x_i Exemplare des Bestsellers in der Buchhandlung verkauft wurden (*relative Häufigkeit* des Merkmalswertes x_i)

$H_i = h_1 + h_2 + \dots + h_i$ = Anzahl der Tage, an denen höchstens x_i Exemplare des Bestsellers in der Buchhandlung verkauft wurden (*absolute Summenhäufigkeit*)

$F_i = f_1 + f_2 + \dots + f_i$ = Anteil der Tage, an denen höchstens x_i Exemplare des Bestsellers in der Buchhandlung verkauft wurden (*relative Summenhäufigkeit*). Es gilt $F_i = \frac{H_i}{N}$.

Wir zeichnen die relative (bzw. absolute) Häufigkeitsverteilung in einem *Histogramm* (siehe linkes Bild in Abbildung 2.1) bzw. *Stabdiagramm* (siehe rechtes Bild in Abbildung 2.1): Jeder Merkmalswert wird durch einen Balken gleicher Breite bzw. durch einen Stab mit der Höhe jeweils *proportional zur jeweiligen Häufigkeit* dargestellt. Somit sind die *Flächen* im Histogramm *proportional zur relativen (bzw. absoluten) Häufigkeit*. Wegen der gleichen Breite der Rechteckboxen im Histogramm sind auch die Höhen der Boxen proportional zur relativen (bzw. absoluten) Häufigkeit.

Wir definieren die *relative Summenhäufigkeitsfunktion*:

$$F(x) = \begin{cases} 0 & \text{wenn } x < x_1 = 0, \\ F_i & \text{wenn } x_i \leq x < x_{i+1}, \quad i = 1, 2, 3, 4, \\ 1 & \text{wenn } x \geq x_5 = 4. \end{cases}$$

Diese ist eine **Treppenfunktion** und hat die folgende Interpretation:

$F(x)$ = Anteil der Tage, an denen höchstens x Exemplare des Bestsellers in der Buchhandlung verkauft wurden.

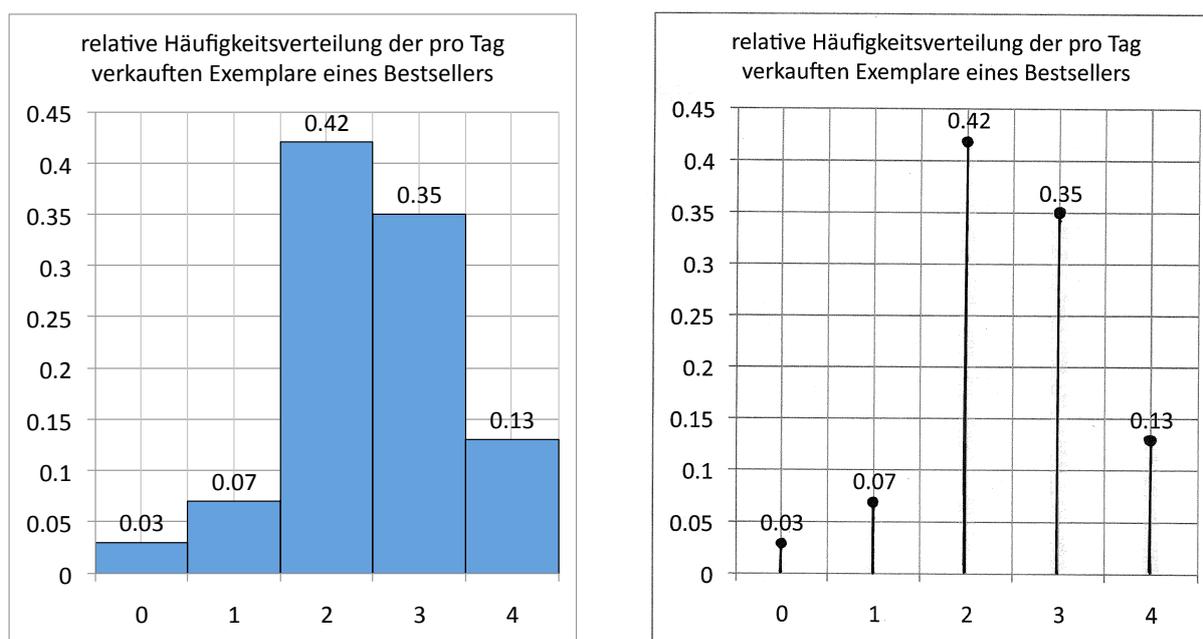


Abbildung 2.1: Links ein Histogramm und rechts ein Stabdiagramm der relativen Häufigkeitsverteilung aus Beispiel 2.1.

Für einen gebrochenen Wert, wie z.B. $x = 5/2$, ist also $F(5/2) = F(2)$, da der Anteil der Tage, an denen höchstens $x = 5/2$ Exemplare des Bestsellers in der Buchhandlung verkauft wurden, gleich dem Anteil der Tage ist, an denen höchstens 2 Exemplare des Bestsellers in der Buchhandlung verkauft wurden.

Wir zeichnen den Graphen der relativen Summenhäufigkeitsfunktion in Abbildung 2.2. Dabei verwenden wir die folgende Konvention an den Sprungstellen der „Treppenstufen“: Dort wo der Funktionswert liegt, (also z.B. bei $(2, 0,52)$) zeichnen wir einen ausgefüllten Punkt, und dort wo der Funktionswert nicht liegt (also z.B. bei $(2, 0,1)$), zeichnen wir einen leeren/unausgefüllten Punkt. An diesen „Sprüngen“ dürfen wir die beiden Werte natürlich nicht durch eine vertikale Linie verbinden, denn sonst hätten wir keine Funktion. Um besser zeichnen zu können ist es aber üblich den Sprung vertikal mit einer gepunkteten oder getrickelten Linie einzuzeichnen.

Wir halten die in dem Beispiel kennengelernten neuen statistischen Begriffe fest.

Definition 2.2. (absolute bzw. relative Häufigkeiten und Summenhäufigkeiten; Histogramm und Summenhäufigkeitsfunktion)

Wir betrachten ein *diskretes metrisch skaliertes Merkmal* X mit k (möglichen) Merkmalswerten $x_1 < x_2 < \dots < x_k$ auf einer Grundgesamtheit mit N Elementen.

(1) $h_i =$ **absolute Häufigkeit**, mit welcher der Merkmalswert x_i auftritt

(d.h. die Anzahl der Elemente mit Merkmalswert x_i)

(2) $f_i = \frac{h_i}{N} =$ **relative Häufigkeit**, mit welcher der Merkmalswert x_i auftritt

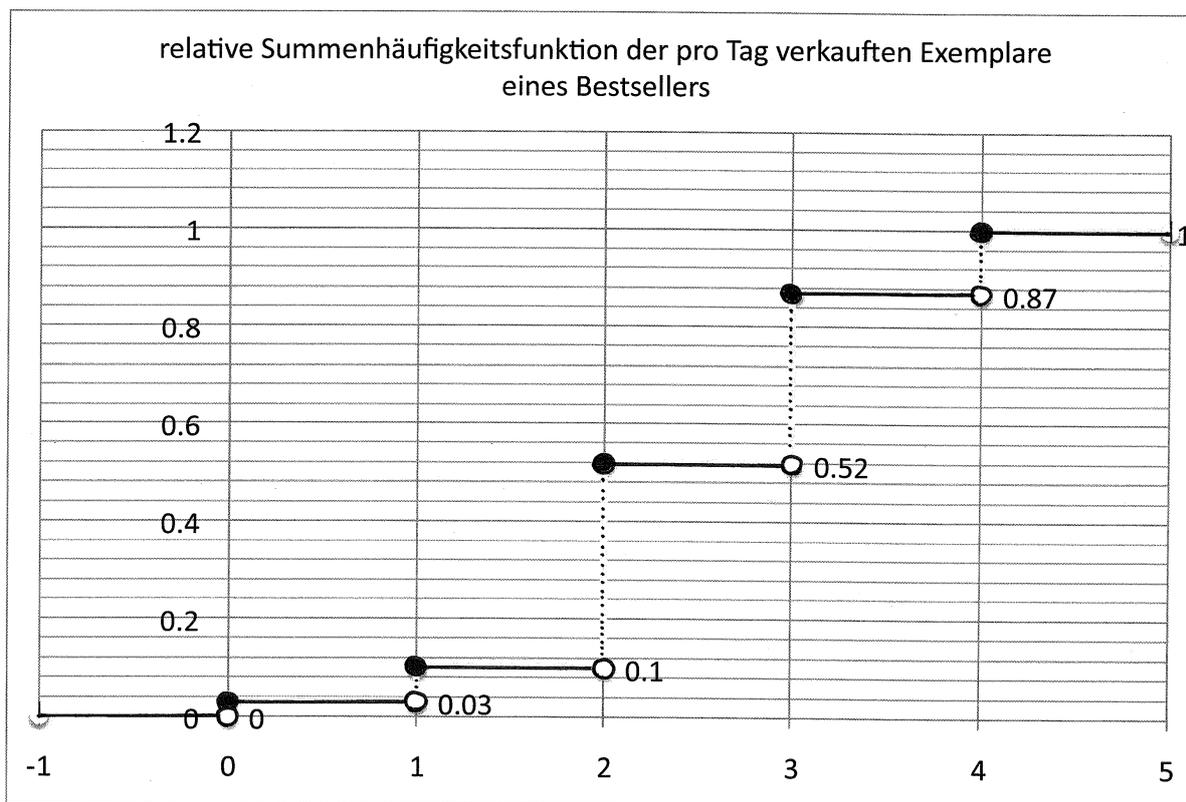


Abbildung 2.2: Graph der relativen Summenhäufigkeitsfunktion der Häufigkeitsverteilung aus Beispiel 2.1.

(d.h. der Anteil der Elemente mit Merkmalswert x_i)

$$(3) H_i = h_1 + h_2 + \dots + h_i = \sum_{j=1}^i h_j = \text{absolute Summenhäufigkeit}$$

(d.h. die Anzahl der Elemente mit Merkmalswert $\leq x_i$)

$$(4) F_i = \frac{H_i}{N} = \sum_{j=1}^i f_j = f_1 + \dots + f_i = \text{relative Summenhäufigkeit}$$

(d.h. der Anteil der Elemente mit Merkmalswert $\leq x_i$)

(5) **Histogramm der absoluten bzw. relativen Häufigkeitsverteilung:** Über den Werten x_i werden Balken gleicher Breite abgetragen, deren Flächen (und Höhen) jeweils proportional zu den Werten der absoluten bzw. relativen Häufigkeiten h_i bzw. f_i sind.

(6) **Stabdiagramm der absoluten bzw. relativen Häufigkeitsverteilung:** Über den Werten x_i werden „Stäbe“ abgetragen, deren Höhen jeweils proportional zu den Werten der absoluten bzw. relativen Häufigkeiten h_i bzw. f_i sind.

(7) **relative Summenhäufigkeitsfunktion:**

$F(x)$ = Anteil der Elemente mit Merkmalswert $\leq x$, d.h.

$$F(x) = \begin{cases} 0 & \text{für } x < x_1, \\ F_i & \text{für } x_i \leq x < x_{i+1} \quad \text{mit } i = 1, 2, \dots, k-1, \\ 1 & \text{für } x \geq x_k. \end{cases}$$

Zeichnet man die relative Summenhäufigkeitsfunktion, so erhält man eine Treppenfunktion.

2.2 Häufigkeitsverteilung eines metrisch skalierten Merkmals mit Klassenbildung

Wie geht man vor, wenn ein **stetiges** metrisch skaliertes Merkmal oder ein **diskretes** metrisch skaliertes Merkmal mit **sehr vielen verschiedenen Merkmalswerten** vorliegt? Wir bilden geeignete **Klassen** (in der Regel zwischen 5 und 20 Stück) **von Merkmalswerten**.

Auch hier betrachten wir zunächst ein Beispiel, an dem wir die neuen Ideen kennenlernen.

Beispiel 2.3. (Preisverteilung beim Autobestand eines Gebrauchtwagenhändlers mit Klassenbildung¹)

Ein Gebrauchtwagenhändler mit $N = 50$ Gebrauchtwagen nimmt eine Inventur seines Autobestands vor, wobei er die Autos in Preisklassen einteilt:

Merkmal X = „Preis des Gebrauchtwagens“

Die i -te Preisklasse enthält alle Autos, die mehr als x_i^u (untere Klassengrenze) aber höchstens x_i^o (obere Klassengrenze) in 1000 Euro kosten (also Autos mit Preisen $x \cdot 1000$ Euro mit $x_i^u < x \leq x_i^o$).

$\Delta x_i = x_i^o - x_i^u =$ Klassenbreite der i -ten Preisklasse

$h_i =$ Anzahl der Autos in der i -ten Preisklasse (*absolute Häufigkeit* der i -ten Preisklasse)

Die Klassen müssen den Wertebereich des Merkmals *überschneidungsfrei überdecken*, und insbesondere gilt somit $x_i^o = x_{i+1}^u$ für alle i .

Als repräsentativen Wert für die i -te Klasse nehmen wir die *Klassenmitte* $x_i' = \frac{1}{2}(x_i^u + x_i^o)$.

¹Diese Beispiel wurde von einem Beispiel in [B12, Teilkapitel 2.3 und 2.4] inspiriert.

i	Klasse (Preis in 1000 Euro)	x_i^u	x_i^o	x_i'	Δx_i	h_i	h_i^*	f_i	f_i^*	H_i	F_i
1	0 bis 2	0	2	1	2	2	1	0,04	0,02	2	0,04
2	über 2 bis 3	2	3	2,5	1	8	8	0,16	0,16	10	0,2
3	über 3 bis 4	3	4	3,5	1	7	7	0,14	0,14	17	0,34
4	über 4 bis 5	4	5	4,5	1	13	13	0,26	0,26	30	0,6
5	über 5 bis 10	5	10	7,5	5	15	3	0,3	0,06	45	0,9
6	über 10 bis 15	10	15	12,5	5	5	1	0,1	0,02	50	1,0
Σ						50		1,0			

Wir berechnen aus diesen Daten:

f_i = Anteil der Autos in der i -ten Preisklasse (*relative Häufigkeit* der i -ten Preisklasse)

$H_i = h_1 + h_2 + \dots + h_i$ = Anzahl der Autos mit Preis $\leq x_i^o \cdot 1000$ Euro (*absolute Summenhäufigkeit*)

$F_i = f_1 + f_2 + \dots + f_i = \frac{H_i}{N}$ = Anteil der Autos mit Preis $\leq x_i^o \cdot 1000$ Euro (*relative Summenhäufigkeit*)

Wir zeichnen die *absolute Häufigkeitsverteilung* in einem *Histogramm* (siehe Abbildung 2.3 unten).

Erklärung zum Histogramm: Die i -te Preisklasse wird durch einen Balken der Breite Δx_i und mit *Fläche proportional* zu h_i dargestellt. Die Grundseite des Balken für die i -te Klasse reicht dabei von x_i^u bis x_i^o . Die Höhen des Balken der i -ten Klasse ist damit proportional zu $h_i^* = h_i / \Delta x_i$ (da dann $h_i = h_i^* \cdot \Delta x_i$ = die Fläche des Rechtecks mit Breite Δx_i und Höhe h_i^* ist). Am Einfachsten wählt man den Proportionalitätsfaktor als 1, so dass h_i^* die Höhe der Balken angibt.

Achtung: Bei unterschiedlich breiten Klassen sind die Höhen der Balken und die h_i^* *nicht* proportional zu den h_i !

Analog kann man die *relative Häufigkeitsverteilung* in einem *Histogramm* zeichnen: Es ändern sich lediglich die Skala der y -Achse, und die Höhe der Box der i -ten Klasse ist nun proportional zu $f_i^* = f_i / \Delta x_i$ zu wählen, so dass die Fläche dieser Box proportional zu f_i ist.

Die *absolute bzw. relative Häufigkeitsverteilung im Histogramm* ordnet also den Wert h_i^* bzw. f_i^* allen x aus i -ter Klasse zu.

Wir definieren und zeichnen die *relative Summenhäufigkeitsfunktion* $F(x)$:

$$F(x) = \text{Anteil der Gebrauchtwagen mit Preis } \leq x$$

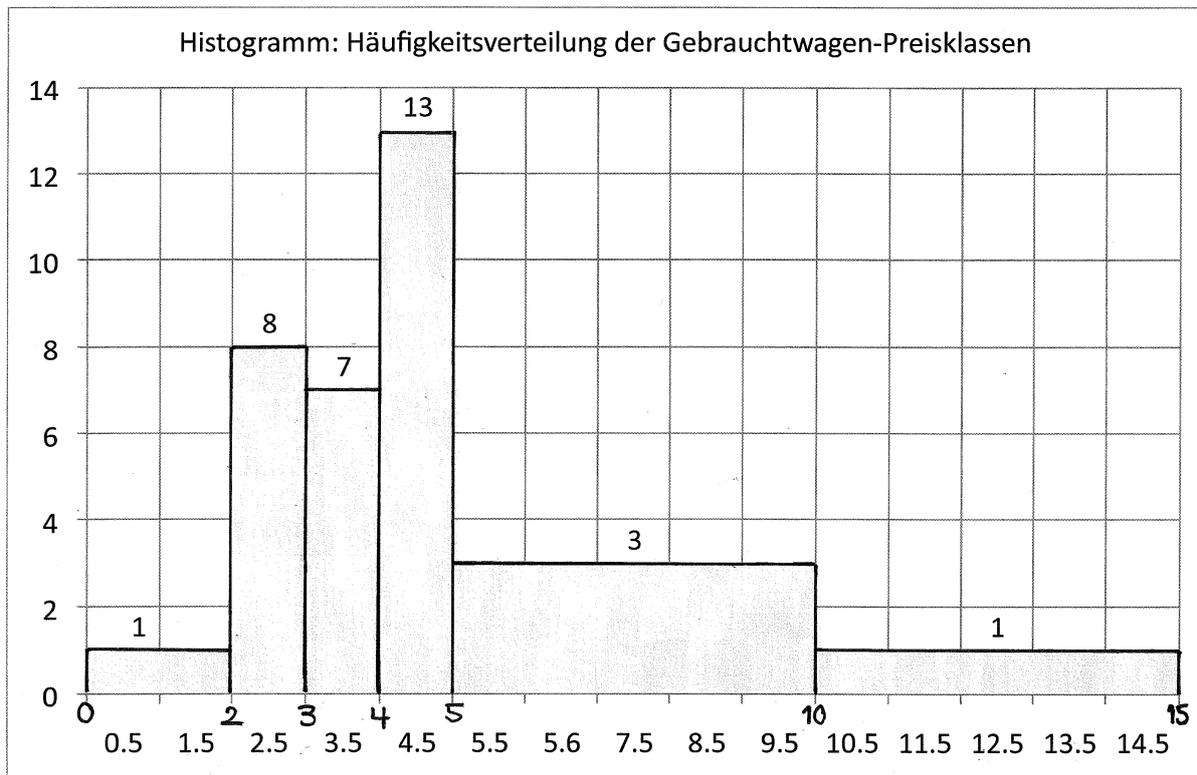


Abbildung 2.3: Histogramm der absoluten Häufigkeitsverteilung mit Klassenbildung der Gebrauchtwagenpreise aus Beispiel 2.3.

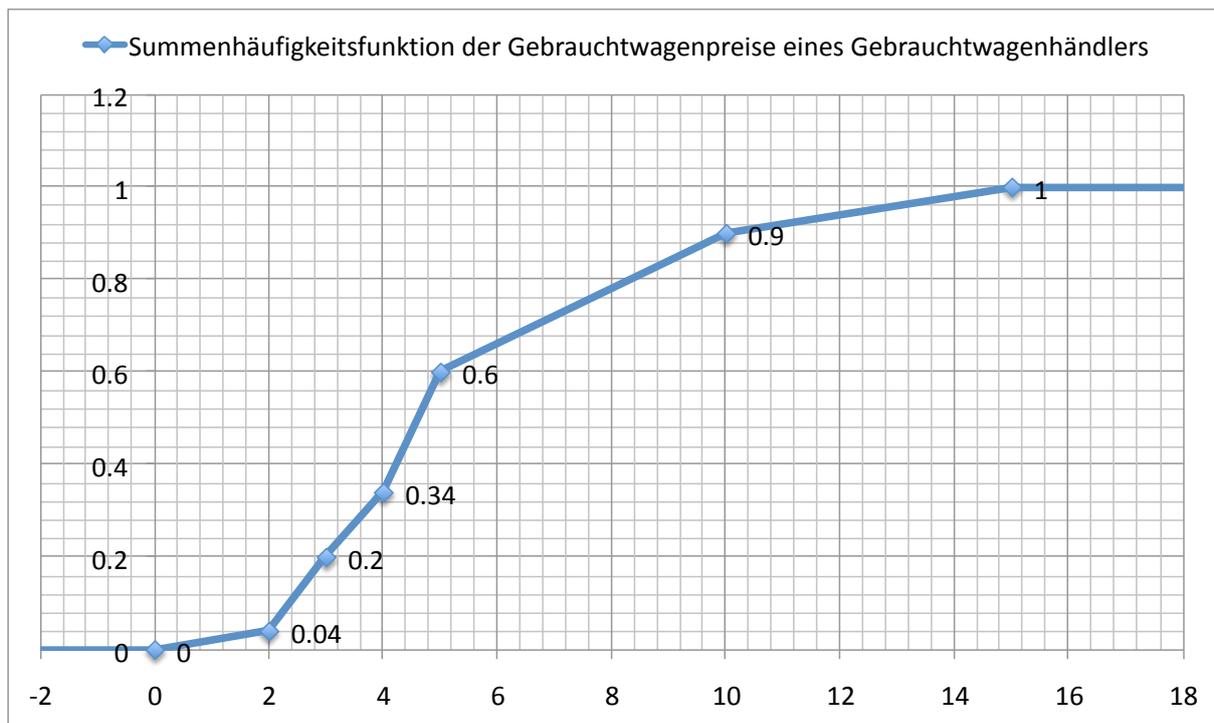


Abbildung 2.4: Graph der relativen Summenhäufigkeitsfunktion der Gebrauchtwagenpreise aus Beispiel 2.3. Durch die Annahme, dass sich die Preise gleichmäßig über die Klassen verteilen, erhalten wir eine stückweise lineare stetige Funktion.

Zeichnen der relative Summenhäufigkeitsfunktion: Wir tragen die Punkte $(x_1^u, 0)$, (x_i^o, F_i) , $i = 1, 2, \dots, 6$, in das Koordinatensystem ein und verbinden benachbarte Punkte durch Linien. Weiter gilt hier

$$F(x) = 0 \quad \text{wenn } x < x_1^u \quad \text{und} \quad F(x) = 1 \quad \text{wenn } x \geq x_6^o.$$

Erklärung der Grafik in Abbildung 2.4: Wir wissen, dass $F(x) = 0$ wenn $x < 0$, dass $F(0) = 0$ (es gibt keine Autos umsonst) und dass $F(x_i^o) = F_i$, $i = 1, 2, \dots, 6$ (z.B. $F(5) = 0,6$ – siehe Tabelle). Weiter gilt, dass $F(x) = 1$ wenn $x \geq x_6^o$, denn kein Auto kostet mehr als x_6^o . Wir zeichnen diese Punkte ein. Unter der Annahme, dass sich die Gebrauchtwagenpreise innerhalb einer Klasse *gleichmäßig verteilen*, verbinden wir die benachbarten Punkte durch eine Linie.

Wir halten die im vorigen Beispiel eingeführten neuen statistischen Begriffe fest.

Definition 2.4. (Häufigkeitsverteilung mit Klassen: absolute/relative Häufigkeiten und Summenhäufigkeiten, Histogramm, Summenhäufigkeitsfunktion)

*Wir betrachten ein (diskretes oder stetiges) **metrisch skaliertes Merkmal** X auf einer Grundgesamtheit mit N Elementen. Da sehr sehr viele Merkmalswerte auftreten, werden k **Klassen** (also Intervalle von Merkmalswerten) gebildet, die alle Merkmalswerte **überschneidungsfrei** überdecken. Es wird die Annahme gemacht, dass sich die Merkmalswerte in jeder Klasse über diese Klasse **gleichmäßig verteilen**.*

(1) ***i -te Klasse** von unterer Grenze x_i^u bis oberer Grenze x_i^o (wobei $x_i^o = x_{i+1}^u$ und $x_i^u = x_{i-1}^o$), und es muss klar sein, zu welcher Klasse x_i^o bzw. x_i^u gehören*

(2) $\Delta x_i = x_i^o - x_i^u =$ **Klassenbreite** der i -ten Klasse

(3) $x_i' = \frac{1}{2} (x_i^u + x_i^o) =$ **Klassenmitte** der i -ten Klasse

(Die Klassenmitte kann als repräsentativer Wert für die Klasse angesehen werden.)

(4) $h_i =$ **absolute Häufigkeit** der i -ten Klasse

(d.h. die Anzahl der Elemente in der i -ten Klasse)

(5) $f_i = \frac{h_i}{N} =$ **relative Häufigkeit** der i -ten Klasse

(d.h. der Anteil der Elemente in der i -ten Klasse)

(6) $H_i = h_1 + h_2 + \dots + h_i = \sum_{j=1}^i h_j =$ **absolute Summenhäufigkeit**

(d.h. die Anzahl der Elemente mit Merkmalswert $\leq x_i$ bzw. $< x_i$)

Alternativ: H_i ist die Anzahl der Elemente, die in der ersten oder der zweiten oder ... oder der i -ten Klasse liegen.

(7) $F_i = f_1 + f_2 + \dots + f_i = \sum_{j=1}^i f_j = \frac{H_i}{N} =$ **relative Summenhäufigkeit**

(d.h. der Anteil der Elemente mit Merkmalswert $\leq x_i$ bzw. $< x_i$)

Alternativ: F_i ist der Anteil der Elemente, die in der ersten oder der zweiten oder ... oder der i -ten Klasse liegen.

(8) **absolute Häufigkeitsverteilung im Histogramm:**

$$h(x) = \begin{cases} h_i^* & \text{wenn } x \text{ in } i\text{-ter Klasse für } i = 1, 2, \dots, k, \\ 0 & \text{sonst,} \end{cases}$$

wobei $h_i^* = h_i / \Delta x_i$

(9) **relative Häufigkeitsverteilung im Histogramm:**

$$f(x) = \begin{cases} f_i^* & \text{wenn } x \text{ in } i\text{-ter Klasse für } i = 1, 2, \dots, k, \\ 0 & \text{sonst,} \end{cases}$$

wobei $f_i^* = f_i / \Delta x_i$

(10) Die **relative Summenhäufigkeitsfunktion**

$F(x)$ = Anteil der Elemente mit Merkmalswert $\leq x$
erhält man, indem man

$$(x_1^u, 0), (x_1^o, F_1), (x_2^o, F_2), \dots, (x_{k-1}^o, F_{k-1}), (x_k^o, F_k)$$

einzeichnet und benachbarte Punkte durch eine Linie verbindet. Weiter gilt $F(x) = 0$ für $x \leq x_1^u$ und $F(x) = 1$ für $x \geq x_k^o$. (Die Summenhäufigkeitsfunktion ist also eine stückweise lineare und monoton wachsende stetige Funktion mit Werten in $[0, 1]$.)

Mittelwerte

In diesem und dem nachfolgenden Kapitel werden wir Häufigkeitsverteilungen durch geeignete Kennzahlen charakterisieren. Dabei verwenden wir:

Mittelwerte (diese kennzeichnen das Zentrum der Verteilung) und

Streuungsmaße (dies kennzeichnen die Streuung/Variabilität der Verteilung).

In diesem Kapitel lernen wir verschiedene Mittelwerte kennen.

Wir betrachten in diesem Kapitel also Häufigkeitsverteilungen **ohne Klassenbildung**. Bei Häufigkeitsverteilungen klassifizierter Daten (also mit Klassenbildung) nimmt man als repräsentativen Wert jeder Klasse die Klassenmitte

$$x'_i = \frac{1}{2}(x_i^u + x_i^o).$$

und behandelt die von den Klassenmitten mit den zugehörigen Häufigkeiten gebildete Häufigkeitsverteilung (ohne Klassenbildung) wie in diesem Kapitel diskutiert.

Liegen bei der Häufigkeitsverteilung eines Merkmals mit Klassenbildung allerdings auch die Einzelwerte vor, so sollte man unbedingt direkt mit diesen arbeiten, denn bei der Berechnung mit dem Klassenmitten als repräsentativen Werten der Klassen erhält man eine sehr viel ungenauere Berechnung von Mittelwerten und Streuungsmaßen.

3.1 Arithmetisches Mittel

Wir starten mit der Einführung des arithmetischen Mittels. Dieses ist der „gängigste“ Mittelwert und ist meist gemeint, wenn umgangssprachlich vom „Mittelwert“ oder „Durchschnitt(swert)“ geredet wird.

Definition 3.1. (arithmetisches Mittel für ein metrisch skaliertes Merkmal)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N , den k Merkmalswerten $x_1 < x_2 < \dots < x_k$ und den zugehörigen absoluten Häufigkeiten h_1, h_2, \dots, h_k . Das **arithmetische Mittel** \bar{x} von X ist definiert als

$$\bar{x} = \frac{1}{N} (a_1 + a_2 + \dots + a_N) = \frac{1}{N} \sum_{i=1}^N a_i. \quad (3.1)$$

Mit der Häufigkeitsverteilung von X können wir \bar{x} auch schreiben als

$$\bar{x} = \frac{1}{N} (h_1 x_1 + h_2 x_2 + \dots + h_k x_k) = \frac{1}{N} \sum_{i=1}^k h_i x_i = \sum_{i=1}^k \frac{h_i}{N} x_i = \sum_{i=1}^k f_i x_i, \quad (3.2)$$

wobei im letzten Schritt $f_i = h_i/N$ (relative Häufigkeiten) verwendet wurde.

In (3.1) und (3.2) taucht die Summennotation auf. Diese werden wir in der Vorlesung sehr häufig benötigen. Falls Sie mit dieser nicht sicher umgehen können, müssen Sie dieses üben! Hierzu gibt es eine Einführung in die Summennotation in Anhang A und einige Übungsaufgaben.

Warum liefern die beiden Formeln (3.1) und (3.2) denselben Wert? Dieses sieht man, wenn man sich klar macht, wie die Häufigkeitsverteilung und die Einzelwerte zusammenhängen. Wir untersuchen dieses auf einem Übungszettel.

Betrachten wir zwei Beispiele zur Berechnung des arithmetischen Mittels.

Beispiel 3.2. (arithmetisches Mittel – Zeit für den Weg zur Uni)

Ein Student stoppt seine Wegzeit (Merkmal X) für den Weg von der Wohnung zur Universität über $N = 10$ aufeinander folgende Tage und erhält dabei am i -ten Tag den Einzelwert a_i für die Wegzeit.

Nummer des Tages	i	1	2	3	4	5	6	7	8	9	10
Wegzeit in min	a_i	15	13	12	17	15	12	14	12	15	16

$$\text{Arithmetisches Mittel} = \bar{x} = \frac{1}{N} \sum_{i=1}^N a_i \stackrel{N=10}{=} \frac{1}{10} (a_1 + a_2 + \dots + a_{10})$$

Also erhalten wir

$$\begin{aligned} \bar{x} &= \frac{1}{10} \cdot (15 + 13 + 12 + 17 + 15 + 12 + 14 + 12 + 15 + 16) \\ &= \frac{1}{10} \cdot (45 + 36 + 60) = \frac{141}{10} = 14,1. \end{aligned}$$

Der Student ist durchschnittlich 14,1 Minuten von der Wohnung zur Universität unterwegs. In Abbildung 3.1 haben wir die Einzelwerte des Merkmal $X =$ „Wegzeit für den

Weg von der Wohnung zur Universität“ in einem Streudiagramm veranschaulicht und auch das *arithmetische Mittel* $\bar{x} = 14,1$ eingetragen. Man sieht gut, wie die Einzelwerte um das arithmetische Mittel streuen.

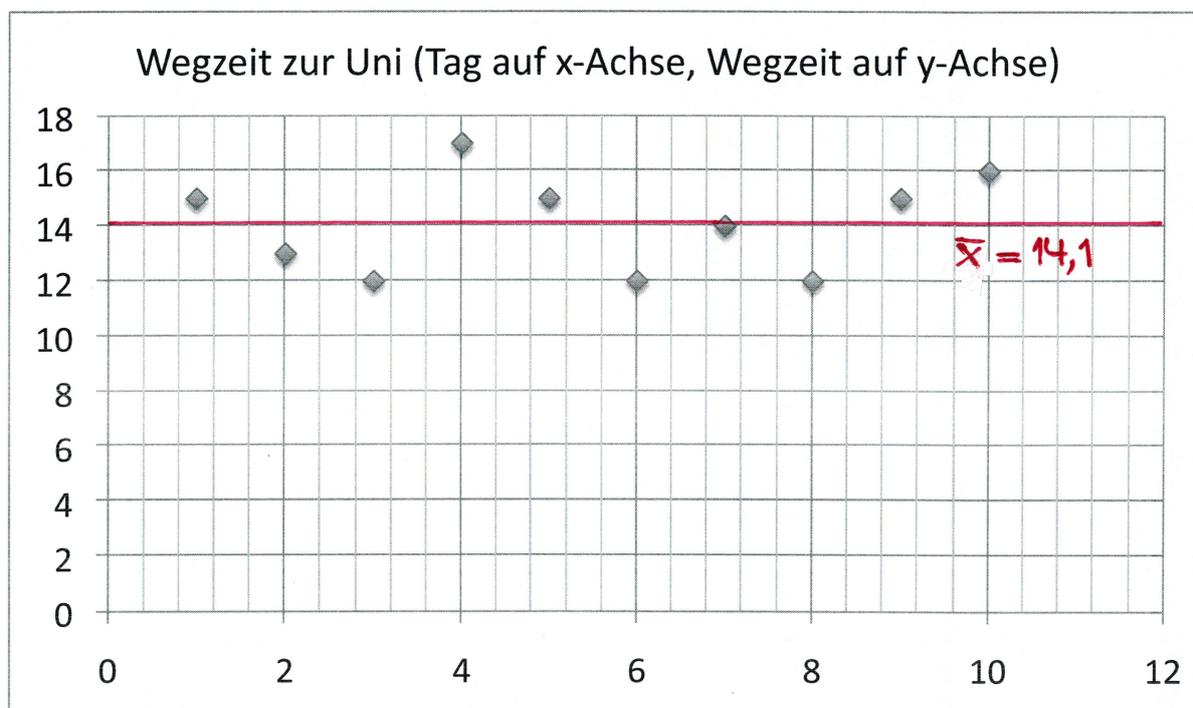


Abbildung 3.1: Streudiagramm der Einzelwerte des Merkmals $X =$ „Wegzeit für den Weg von der Wohnung zur Universität“ mit dem arithmetischen Mittel (siehe Beispiel 3.2).

Beispiel 3.3. (arithmetisches Mittel – Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiel 2.1):

Dabei sind:

i	x_i	h_i
1	0	3
2	1	7
3	2	42
4	3	35
5	4	13
Σ		100

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“
- $h_i =$ Anzahl der Tage, an denen genau x_i (Merkmalswert) Exemplare des Bestsellers verkauft wurden (absolute Häufigkeit)

Wir berechnen das *arithmetische Mittel*

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k h_i x_i \stackrel{N=100, k=5}{=} \frac{1}{100} (h_1 x_1 + h_2 x_2 + \dots + h_5 x_5).$$

mit Hilfe der erweiterten Berechnungstabelle:

i	x_i	h_i	$h_i x_i$
1	0	3	$3 \cdot 0 = 0$
2	1	7	$7 \cdot 1 = 7$
3	2	42	$42 \cdot 2 = 84$
4	3	35	$35 \cdot 3 = 105$
5	4	13	$13 \cdot 4 = 52$
\sum		100	248
$\frac{1}{100} \sum$			$\bar{x} = \frac{248}{100} = 2,48$

Das *arithmetische Mittel* ist $\bar{x} = 2,48$. Durchschnittlich wurden $\bar{x} = 2,48$ Exemplare des Bestsellers pro Tag in der Buchhandlung verkauft.

Nun untersuchen wir die Eigenschaften des arithmetischen Mittels.

Satz 3.4. (Eigenschaften des arithmetischen Mittels)

Seien die Voraussetzungen und Bezeichnungen wie in Definition 3.1. Dann gelten:

- (1) Die **Summe der Abweichungen** $a_i - \bar{x}$ der Einzelwerte a_i vom **arithmetischen Mittel** \bar{x} ist immer Null:

$$\sum_{i=1}^N (a_i - \bar{x}) = 0.$$

- (2) Die **Summe der quadrierten Abweichungen** $(a_i - \bar{x})^2$ der Einzelwerte a_i vom **arithmetischen Mittel** \bar{x} ist kleiner als die Summe der quadrierten Abweichungen $(a_i - M)^2$ der Einzelwerte a_i von einem beliebigen Wert M :

$$\sum_{i=1}^N (a_i - \bar{x})^2 < \sum_{i=1}^N (a_i - M)^2 \quad \text{für alle } M \neq \bar{x}.$$

- (3) Wendet man die **affin lineare Funktion** $y = \beta x + \alpha$ auf die Einzelwerte a_i an, also $a_i^* = \beta a_i + \alpha$, so gilt für das **arithmetische Mittel** \bar{x}^* der neuen Werte a_i^* :

$$\bar{x}^* = \beta \bar{x} + \alpha,$$

wobei \bar{x} das arithmetische Mittel von a_1, a_2, \dots, a_N ist.

Wir können uns die Eigenschaften aus Satz 3.4 (1) und aus Satz 3.4 (2) leicht **veranschaulichen**: Betrachten wir dazu noch einmal das Streudiagramm der Einzelwerte aus Beispiel 3.2 mit dem eingezeichneten arithmetischen Mittel. In den beiden nachfolgenden Abbildungen 3.2 bzw. 3.3 haben wir jeweils die Abweichungen $a_i - \bar{x}$ (aus Satz 3.4 (1)) bzw. die Abweichungsquadrate $(a_i - \bar{x})^2$ (aus Satz 3.4 (2)) eingezeichnet. Die Bildunterschrift erklärt jeweils Satz 3.4 (1) bzw. Satz 3.4 (2).

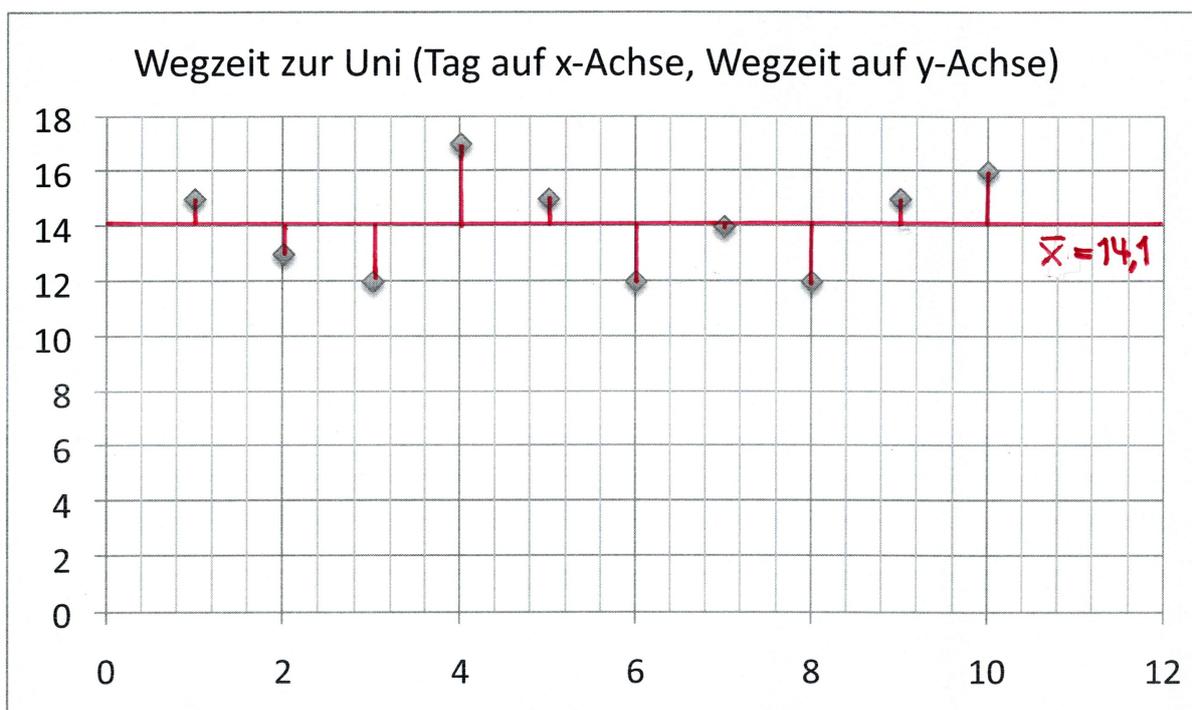


Abbildung 3.2: *Erklärung von Satz 3.4 (1)*: Die horizontale Gerade durch $y = 14,1$ kennzeichnet das arithmetische Mittel $\bar{x} = 14,1$ der Einzelwerte. Die Summe der in rot eingezeichneten Abweichungen $a_i - \bar{x}$ (mit Vorzeichen) ist nach Satz 3.4 (1) Null. Schiebt man die rote horizontale Gerade nach oben oder unten, so ändern sich diese Abweichungen, und ihre Summe ist nicht mehr Null.

Betrachten wir ein Beispiel für die Anwendung von Satz 3.4 (3).

Beispiel 3.5. (Zentrierung eines Merkmals)

Sei X ein Merkmal auf einer Grundgesamtheit mit N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N und dem arithmetischen Mittel \bar{x} . Dann liefert die affin lineare Funktion $y = x - \bar{x}$ (also hier $\beta = 1$ und $\alpha = -\bar{x}$) die neuen Einzelwerte

$$a_i^* = a_i - \bar{x}, \quad i = 1, 2, \dots, N,$$

eines neuen Merkmals X^* . Nach Satz 3.4 (3) gilt für das arithmetische Mittel \bar{x}^* der neuen Einzelwerte

$$\bar{x}^* = 1 \cdot \bar{x} - \bar{x} = 0,$$

d.h. das neue Merkmal X^* hat das arithmetische Mittel $\bar{x}^* = 0$. Wir sagen es ist **zentriert**, denn die Werte „streuen“ um das arithmetische Mittel $\bar{x}^* = 0$.

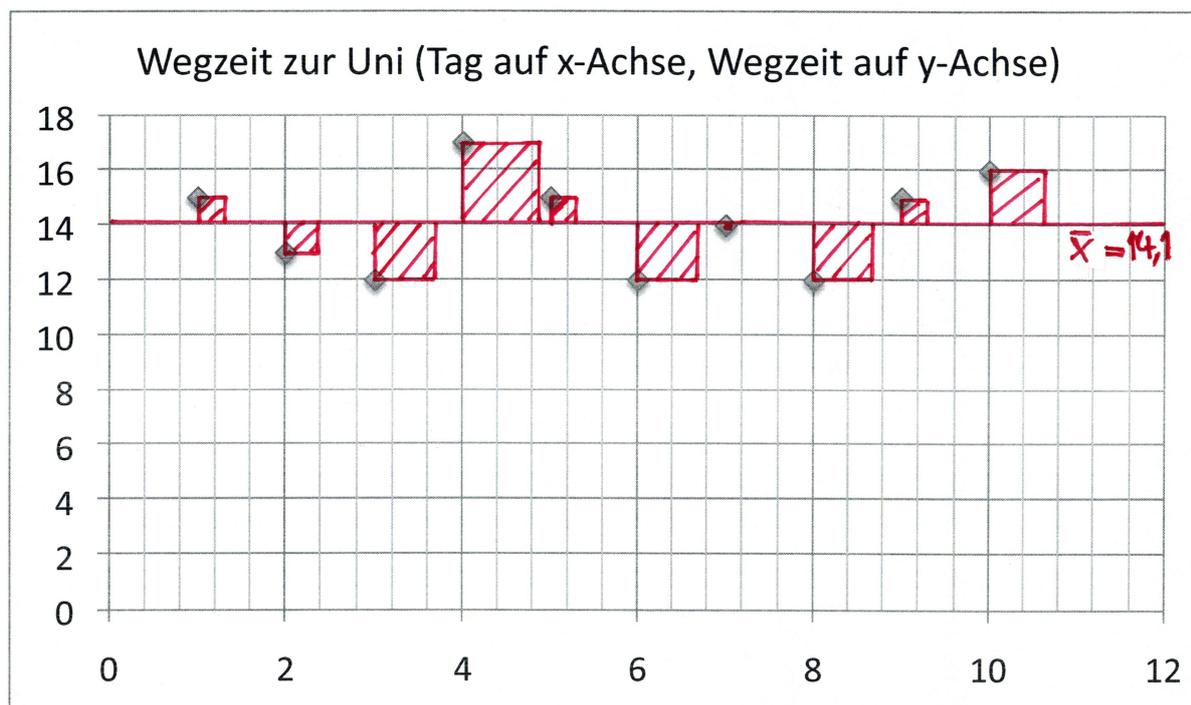


Abbildung 3.3: *Erklärung von Satz 3.4 (2)*: Die horizontale Gerade durch $y = 14,1$ kennzeichnet das arithmetische Mittel $\bar{x} = 14,1$ der Einzelwerte. Die Flächeninhalte der in rot eingezeichneten Quadrate sind gerade die Abweichungsquadrate $(a_i - \bar{x})^2$. Schiebt man die rote horizontale Gerade nach oben oder unten, so ändern sich diese Flächen (also die Abweichungsquadrate vom Wert der Geraden). Nach Satz 3.4 (2) ist die Summe der Flächeninhalte dieser Quadrate minimal, wenn die horizontale Gerade durch den Mittelwert geht.

Frage: Warum sollte man eine affin lineare Funktion auf ein Merkmal anwenden?

Für bestimmte statistische Verfahren kann es notwendig sein, ein Merkmal mit einer affin linearen Funktion so in ein neues Merkmal umzuwandeln, dass das neue Merkmal arithmetisches Mittel Null und Varianz Eins hat. (Die Varianz wird im nächsten Kapitel besprochen.)

Wir beweisen Satz 3.4 (1); Satz 3.4 (2) wird in Teilkapitel 4.1 bewiesen, und Satz 3.4 (3) dürfen Sie auf einem Übungszettel beweisen.

Beweis von Satz 3.4 (1): Wir formen geeignet um:

$$\sum_{i=1}^N (a_i - \bar{x}) = \sum_{i=1}^N a_i - \underbrace{\sum_{i=1}^N \bar{x}}_{= N \cdot \bar{x}} = N \cdot \underbrace{\left(\frac{1}{N} \sum_{i=1}^N a_i \right)}_{= \bar{x}} - N \cdot \bar{x} = N \cdot \bar{x} - N \cdot \bar{x} = 0,$$

wobei wir die Rechenregeln für Summen genutzt haben (vgl. Anhang A). \square

3.2 Median (oder Zentralwert)

Wir führen nun einen weiteren wichtigen Mittelwert ein, nämlich den Median.

Definition 3.6. (Median eines ordinalskalierten Merkmals)

Sei X ein (mindestens) ordinalskaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N . Ordnet man die Einzelwerte a_1, a_2, \dots, a_N der Größe nach, so dass

$$a_{[1]} \leq a_{[2]} \leq \dots \leq a_{[N]} \quad (3.3)$$

gilt, dann ist der **Median** (oder **Zentralwert**) Me definiert als

$$Me := \begin{cases} a_{[\frac{N+1}{2}]} & \text{bei ungeradem } N, \\ \frac{1}{2} (a_{[\frac{N}{2}]} + a_{[\frac{N}{2}+1]}) & \text{bei geradem } N. \end{cases}$$

In (3.3) bezeichnet also $a_{[i]}$ den Einzelwert an der i -ten Stelle, wenn die Einzelwerte in aufsteigender Größe sortiert sind.

Geometrische Anschauung des Medians: In (3.3) ist der Median bei ungeradem N der in der Mitte der Anordnung liegende Wert (siehe Abbildung 3.4). Bei geradem N gibt es keinen in der Mitte liegenden Wert. Hier nimmt man also das arithmetische Mittel der beiden Werte, die am dichtesten zur Mitte liegen. Bei N Werten, sind dieses der $N/2$ -te und der $(N/2 + 1)$ -te Wert (siehe Abbildung 3.5).

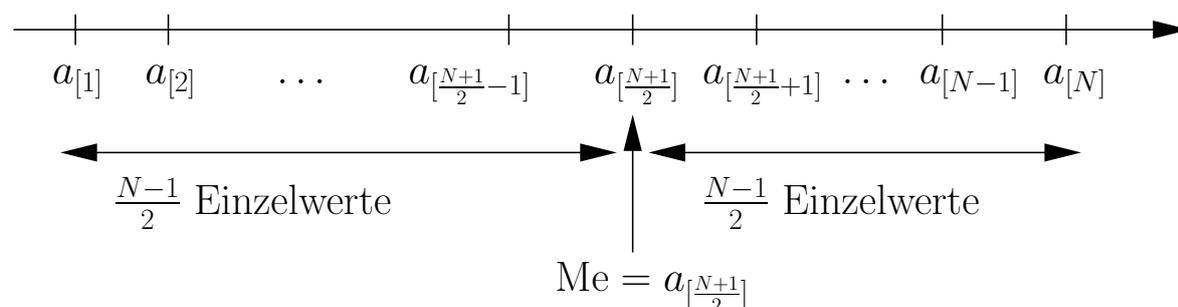


Abbildung 3.4: *Median bei ungeradem N* : Wir haben die sortierten Einzelwerte auf der reellen Zahlengeraden dargestellt. Bei ungeradem N gibt es einen mittleren Wert, nämlich den Einzelwert an der $\frac{N-1}{2} + 1 = \frac{N+1}{2}$ -ten Stelle. Also ist der Median $Me = a_{[\frac{N+1}{2}]}$. Links und rechts davon liegen jeweils $(N - 1)/2$ Einzelwerte.

Betrachten wir zwei Beispiele:

Beispiel 3.7. (Median – Beobachtung einer seltenen Vogelspezies)

Ein Naturforscher beobachtet eine seltene Vogelspezies über 5 Tage und zählt jeweils die Anzahl der am Tag gesichteten Vögel dieser Vogelspezies. Er findet für das Merkmal $X =$ „Anzahl der am Tag gesichteten Vögel“ die folgenden Einzelwerte:

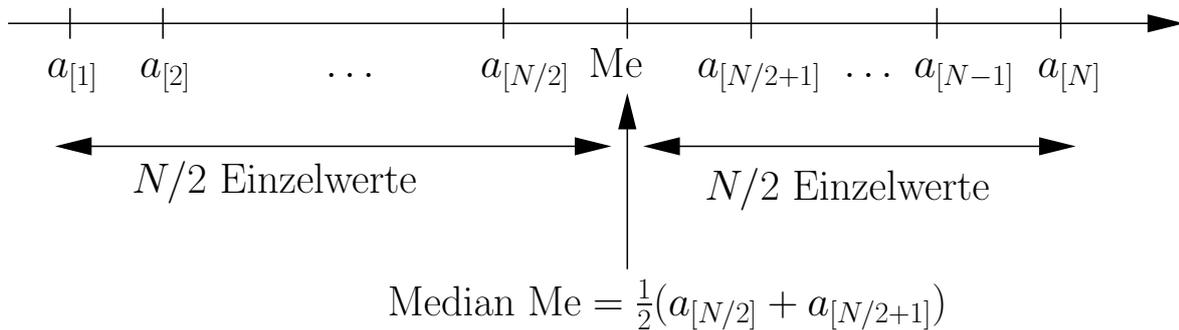


Abbildung 3.5: *Median bei geradem N* : Wir haben die sortierten Einzelwerte auf der reellen Zahlengeraden dargestellt. Bei geradem N gibt es keinen mittleren Wert. Am dichtesten an der Mitte liegen $a_{[N/2]}$ und $a_{[N/2+1]}$, und der Median ist das arithmetische Mittel dieser beiden Werte, also $Me = \frac{1}{2}(a_{[N/2]} + a_{[N/2+1]})$. Links und rechts vom Median liegen jeweils $N/2$ Einzelwerte.

i	1	2	3	4	5
a_i	2	4	1	0	2
	↑	↑	↑	↑	↑
	$a_{[4]}$	$a_{[5]}$	$a_{[2]}$	$a_{[1]}$	$a_{[3]}$

Wir sortieren die Einzelwerte in aufsteigenden Größe:

$$\underbrace{0}_{= a_{[1]}} \leq \underbrace{1}_{= a_{[2]}} \leq \underbrace{2}_{= a_{[3]} = a_{[\frac{5+1}{2}]}} \leq \underbrace{2}_{= a_{[4]}} \leq \underbrace{4}_{= a_{[5]}}$$

Bei $N = 5$ (ungerade) Einzelwerten ist der Wert in der Mitte $a_{[3]} = a_{[\frac{5+1}{2}]} = 2$. Also ist der *Median* $Me = a_{[\frac{5+1}{2}]} = 2$ (beobachtete Vögel pro Tag).

Beispiel 3.8. (Median – Zeit für den Weg zur Uni)

Ein Student stoppt seine Wegzeit für den Weg von der Wohnung zur Universität über $N = 10$ aufeinander folgende Tage und erhält dabei am i -ten Tag den Einzelwert a_i für die Wegzeit.

Nummer des Tages	i	1	2	3	4	5	6	7	8	9	10
Wegzeit in min	a_i	15	13	12	17	15	12	14	12	15	16

Wir ordnen die Einzelwerte der Größe nach:

$$12 \leq 12 \leq 12 \leq 13 \leq \underbrace{14}_{= a_{[5]} = a_{[\frac{10}{2}]}} \leq \underbrace{15}_{= a_{[6]} = a_{[\frac{10}{2}+1]}} \leq 15 \leq 15 \leq 16 \leq 17$$

Bei $N = 10$ (gerade) Einzelwerten ist der Median also das arithmetische Mittel von $a_{[\frac{10}{2}]} = a_{[5]}$ und $a_{[\frac{10}{2}+1]} = a_{[6]}$, also

$$\text{Me} = \frac{1}{2} (a_{[5]} + a_{[6]}) = \frac{1}{2} (14 + 15) = \frac{29}{2} = 14,5.$$

Der *Median* ist $\text{Me} = 14,5$ Minuten (für den Weg von der Wohnung zur Universität).

Was passiert wenn ein Merkmal in Form eine **Häufigkeitsverteilung** (anstatt in Form der Einzelwerte) vorliegt? An dieser kann man die Anordnung der Einzelwerte bequem ablesen und so den Median bestimmen. Betrachten wir auch hierzu ein Beispiel.

Beispiel 3.9. (Median – Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiele 2.1 und 3.3):

i	x_i	h_i	H_i
1	0	3	3
2	1	7	10
3	2	42	52
4	3	35	87
5	4	13	100
Σ		100	

Dabei sind:

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“
- $h_i =$ Anzahl der Tage, an denen x_i (Merkmalswert) Exemplare des Bestsellers verkauft wurden (absolute Häufigkeit)

Wir haben eine gerade Anzahl $N = 100$ von Einzelwerten und der Median ist somit

$$\text{Me} = \frac{1}{2} (a_{[\frac{100}{2}]} + a_{[\frac{100}{2}+1]}) = \frac{1}{2} (a_{[50]} + a_{[51]}).$$

Anhand der Summenhäufigkeiten lesen wir ab, dass $a_{[50]} = a_{[51]} = 2$. (Da $H_3 = 52$ und $H_2 = 10$ ist, muss $a_{[50]} = a_{[51]} = x_3 = 2$ sein.) Also ist der *Median*

$$\text{Me} = \frac{1}{2} (a_{[50]} + a_{[51]}) = \frac{1}{2} (2 + 2) = \frac{4}{2} = 2 \quad (\text{Exemplare des Bestsellers}).$$

3.3 Modus

Der Begriff des Modus ist der einfachste Mittelwert. Er kann für Merkmale auf allen Skalen gebildet werden.

Definition 3.10. (Modus für alle Skalen)

Ein **Modus** (oder **dichtester Wert**) Mo eines Merkmals auf einer Grundgesamtheit

ist eine Merkmalsausprägung, die **am häufigsten** vorkommt. Ein Merkmal kann **einen Modus oder mehrere Modi** besitzen. („Modi“ ist der Plural von „Modus“.)

Betrachten wir hierzu einige Beispiele.

Beispiel 3.11. (Modus – Beobachtung einer seltenen Vogelspezies)

Ein Naturforscher beobachtet eine seltene Vogelspezies über 5 Tage und zählt jeweils die Anzahl der am Tag gesichteten Vögel dieser Vogelspezies. Er findet für das Merkmal $X =$ „Anzahl der pro Tag gesichteten Vögel“ die folgenden Einzelwerte:

i	1	2	3	4	5
a_i	2	4	1	0	2

Da der Wert 2 am häufigsten auftritt, ist der *Modus* $Mo = 2$ (Vögel pro Tag).

Bei einer Häufigkeitsverteilung kann man den Modus bzw. die Modi direkt an den absoluten Häufigkeiten ablesen. Betrachten wir auch hierzu ein Beispiel.

Beispiel 3.12. (Modus – Anzahl der pro Tag verkauften Exemplare eines Bestellers)

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiele 2.1, 3.3 und 3.9):

i	x_i	h_i
1	0	3
2	1	7
3	2	42
4	3	35
5	4	13
Σ		100

Dabei sind:

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag verkauften in der Buchhandlung Exemplare des Bestellers“
- $h_i =$ Anzahl der Tage, an denen x_i (Merkmalswert) Exemplare des Bestellers verkauft wurden (absolute Häufigkeit)

Die größte Häufigkeit ist $h_3 = 42$. Also ist der *Modus*

$$Mo = x_3 = 2 \quad (\text{Exemplare des Bestellers}).$$

Betrachten wir noch ein Beispiel, bei dem mehrere Modi auftreten.

Beispiel 3.13. (Modi – Zeit für den Weg zur Uni)

Ein Student stoppt seine Wegzeit für den Weg von der Wohnung zur Universität über $N = 10$ aufeinander folgende Tage und erhält dabei am i -ten Tag den Einzelwert a_i für die Wegzeit.

Nummer des Tages	i	1	2	3	4	5	6	7	8	9	10
Wegzeit in min	a_i	15	13	12	17	15	12	14	12	15	16

Hier treten die Werte 12 und 15 beide mit der höchsten Häufigkeit 3 auf. Also hat die Verteilung *zwei Modi*: $Mo_1 = 12$ min und $Mo_2 = 15$ min. Eine solche Verteilung mit zwei Modi nennt man „*bimodal*“.

3.4 Geometrisches Mittel

Für verhältnisskalierte Merkmale kann es sinnvoll sein, statt des arithmetischen Mittels das geometrische Mittel zu betrachten. Ein klassisches Beispiel für solche verhältnisskalierten Merkmale sind (jährliche oder monatliche) Wachstumsfaktoren, z.B. bei Zinsen.

Definition 3.14. (geometrisches Mittel für verhältnisskaliertes Merkmal)

Sei X ein metrisch skaliertes Merkmal auf einer Verhältnisskala auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N , den k Merkmalswerten $x_1 < x_2 < \dots < x_k$ und den zugehörigen absoluten Häufigkeiten h_1, h_2, \dots, h_k . Das **geometrische Mittel** G von X ist definiert als

$$G = \sqrt[N]{a_1 \cdot a_2 \cdot \dots \cdot a_N} \quad (3.4)$$

bzw. über die Häufigkeitsverteilung

$$G = \sqrt[N]{x_1^{h_1} \cdot x_2^{h_2} \cdot \dots \cdot x_k^{h_k}} \quad (3.5)$$

Das geometrische Mittel ist sinnvoll bei **Wachstumsfaktoren** oder **Wachstumsraten**.

Frage: Warum liefern (3.4) und (3.5) den gleichen Wert?

Ausgeschrieben lautet (3.5)

$$G = \sqrt[N]{x_1^{h_1} \cdot x_2^{h_2} \cdot \dots \cdot x_k^{h_k}} = \sqrt[N]{\underbrace{(x_1 \cdot \dots \cdot x_1)}_{h_1\text{-mal}} \cdot \underbrace{(x_2 \cdot \dots \cdot x_2)}_{h_2\text{-mal}} \cdot \dots \cdot \underbrace{(x_k \cdot \dots \cdot x_k)}_{h_k\text{-mal}}},$$

und dabei werden alle Einzelwerte multipliziert. (*Begründung:* Für jedes $i = 1, 2, \dots, k$ treten jeweils h_i Einzelwerte a_ℓ mit Wert x_i auf.) Durch geeignetes Umsortieren der h_i erhält man gerade (3.4).

Betrachten wir ein Beispiel zum geometrischen Mittel

Beispiel 3.15. (Geldanlage mit Jahreszinns)

Eine Studentin hat einen festen Geldbetrag A (z.B. 1000 Euro) über drei Jahre fest angelegt mit einem festen jährlichen Zinssatz, den die Bank für jedes Jahr neu festlegt. Die Zinssätze für die drei Jahre sind

Jahr i	1	2	3
Zinssatz	5%	7%	4%
Wachstumsfaktor	1,05	1,07	1,04

Was ist der *durchschnittliche Zinssatz*?

Das arithmetische Mittel hilft uns hier nicht weiter, denn bei der Berechnung des Geldbetrags zum Ende des 3-Jahres-Zeitraums wird der ursprüngliche Geldbetrag nacheinander mit den Wachstumsfaktoren multipliziert:

$$\text{Geldbetrag am Ende der 3 Jahre} = \underbrace{A \cdot 1,05}_{\text{Geld nach Jahr 1}} \cdot 1,07 \cdot 1,04 = A \cdot 1,16844$$

$$\underbrace{\hspace{10em}}_{\text{Geld nach Jahr 2}}$$

$$\underbrace{\hspace{15em}}_{\text{Geld nach Jahr 3}}$$

Hier ist es nur sinnvoll das *geometrische Mittel* der jährlichen Wachstumsfaktoren zu berechnen, also

$$G = \sqrt[3]{1,05 \cdot 1,07 \cdot 1,04} = \sqrt[3]{1,16844} \approx 1,053.$$

Der durchschnittliche jährliche Zinssatz ist also 5,3%.

Achtung: Die Tatsache, dass man für das arithmetische Mittel mit Rundung auf drei Nachkommastellen den gleichen Wert erhält wie für das geometrische Mittel, bedeutet nicht, dass man das arithmetische Mittel verwenden kann!

Streuungsmaße

Aufbauend auf das vorige Kapitel über Mittelwerte lernen wir in diesem Kapitel verschiedene Streuungsmaße kennen, nämlich die **Varianz** und **Standardabweichung**, sowie die **Spannweite**, die **Quartile** und das **Box-and-Whisker-Plot**.

Wir betrachten in diesem Kapitel bis auf eine Ausnahme immer Häufigkeitsverteilungen **ohne Klassenbildung**.

4.1 Die Varianz und Standardabweichung

Die gängigsten Streuungsmaße sind die Varianz und die Standardabweichung.

Definition 4.1. (Varianz und Standardabweichung für ein metrisch skaliertes Merkmal)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N , den k Merkmalswerten $x_1 < x_2 < \dots < x_k$ und den zugehörigen absoluten Häufigkeiten h_1, h_2, \dots, h_k . Die **Varianz** $\text{Var} = \sigma^2$ ist als **arithmetisches Mittel der Abweichungsquadrate von dem arithmetischen Mittel** \bar{x} definiert: Mit den Einzelwerten

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})^2 \quad (4.1)$$

bzw. mit der Häufigkeitsverteilung

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^k h_i (x_i - \bar{x})^2, \quad (4.2)$$

wobei $\bar{x} = \frac{1}{N} \sum_{i=1}^N a_i = \frac{1}{N} \sum_{i=1}^k h_i x_i$ (arithmetisches Mittel des Merkmals X).

Die **Standardabweichung** ist definiert als

$$\sigma = \sqrt{\text{Var}} = \sqrt{\sigma^2}.$$

Dass (4.1) und (4.2) wirklich denselben Wert liefern, zeigen wir auf einem Übungszettel.

Betrachten wir nun einige Beispiele.

Beispiel 4.2. (Varianz und Standardabweichung – Beobachtung einer seltenen Vogelspezies)

Ein Naturforscher beobachtet eine seltene Vogelspezies über 5 Tage und zählt jeweils die Anzahl der am Tag gesichteten Vögel dieser Vogelspezies. Er findet für das Merkmal $X =$ „Anzahl der pro Tag gesichteten Vögel“ die folgenden Einzelwerte:

i	1	2	3	4	5
a_i	2	4	1	0	2

Wir berechnen zunächst das arithmetische Mittel und dann die Varianz und die Standardabweichung:

arithmetisches Mittel: $\bar{x} = \frac{1}{5} (2 + 4 + 1 + 0 + 2) = \frac{9}{5} = 1,8$

Varianz:
$$\begin{aligned} \sigma^2 &= \frac{1}{5} \cdot \left((2 - 1,8)^2 + (4 - 1,8)^2 + (1 - 1,8)^2 + (0 - 1,8)^2 + (2 - 1,8)^2 \right) \\ &= \frac{1}{5} \cdot \left((0,2)^2 + (2,2)^2 + (-0,8)^2 + (-1,8)^2 + (0,2)^2 \right) = \frac{1}{5} \cdot 8,8 = 1,76 \end{aligned}$$

Standardabweichung: $\sigma = \sqrt{1,76} \approx 1,33$

Wir können das arithmetische Mittel und die Varianz auch bequem mit einer Berechnungstabelle berechnen:

i	a_i	$(a_i - \bar{x})^2$
1	2	$(2 - 1,8)^2 = (0,2)^2 = 0,04$
2	4	$(4 - 1,8)^2 = (2,2)^2 = 4,84$
3	1	$(1 - 1,8)^2 = (-0,8)^2 = 0,64$
4	0	$(0 - 1,8)^2 = (-1,8)^2 = 3,24$
5	2	$(2 - 1,8)^2 = (0,2)^2 = 0,04$
Σ	9	8,80
$\frac{1}{5} \Sigma$	$\bar{x} = \frac{9}{5} = 1,8$	$\sigma^2 = \frac{1}{5} \cdot 8,8 = 1,76$

Das arithmetische Mittel ist also $\bar{x} = 1,8$ Vögel, die Varianz ist $\text{Var} = \sigma^2 = 1,76$ Vögel² und die Standardabweichung ist $\sigma = \sqrt{1,76} \approx 1,33$ Vögel.

Bevor wir ein Beispiel mit einer Häufigkeitsverteilung betrachten, lernen wir eine alternative Formel für die Varianz kennen

Lemma 4.3. (alternative Formel für die Varianz)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N , den k Merkmalswerten $x_1 < x_2 < \dots < x_k$ und den zugehörigen absoluten Häufigkeiten h_1, h_2, \dots, h_k . Dann gilt für die **Varianz** die folgende alternative Formel: Mit den Einzelwerten

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^N a_i^2 - \bar{x}^2$$

bzw. mit der Häufigkeitsverteilung

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^k h_i x_i^2 - \bar{x}^2.$$

Dabei sind die obigen Formeln so zu lesen, dass \bar{x}^2 von der Summe **nach** deren Berechnung abgezogen wird, also

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^N a_i^2 - \bar{x}^2 = \left(\frac{1}{N} \sum_{i=1}^N a_i^2 \right) - \bar{x}^2$$

und analog für die zweite Formel.

Beweis von Lemma 4.3: Wir zeigen nur den Fall der Einzelwerte. Der Fall einer Häufigkeitsverteilung wird als Übungsaufgabe behandelt.

Der Beweis erfolgt durch geschicktes Rechnen mit Summen:

$$\begin{aligned} \text{Var} = \sigma^2 &= \frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})^2 \\ &= \frac{1}{N} \sum_{i=1}^N (a_i^2 - 2\bar{x}a_i + \bar{x}^2) && \text{(2. binomische Formel)} \\ &= \frac{1}{N} \sum_{i=1}^N a_i^2 - \frac{1}{N} \sum_{i=1}^N 2\bar{x}a_i + \frac{1}{N} \underbrace{\sum_{i=1}^N \bar{x}^2}_{= N \cdot \bar{x}^2} && \text{(Rechnen mit Summen)} \\ &= \frac{1}{N} \sum_{i=1}^N a_i^2 - 2\bar{x} \cdot \underbrace{\frac{1}{N} \sum_{i=1}^N a_i}_{=\bar{x}} + \frac{1}{N} \cdot N\bar{x}^2 && \text{(Formel für arithmetisches Mittel)} \end{aligned}$$

$$= \frac{1}{N} \sum_{i=1}^N a_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{N} \sum_{i=1}^N a_i^2 - \bar{x}^2$$

(Falls Sie mit dem Rechnen mit Summen Schwierigkeiten haben sollten, schauen Sie bitte Anhang A an.) \square

Frage: Warum ist die Formel aus Lemma 4.3 nützlich?

Die Berechnung mit der Formel aus Lemma 4.3 ist bequemer bei **ganzzahligen Merkmalswerten** und nicht ganzzahligem \bar{x} ! (Es kann weitgehend ohne Brüche gerechnet werden.)

Machen wir uns dies noch einmal an dem Beispiel 4.2 der Vogelbeobachtung klar.

Beispiel 4.4. (Varianz und Standardabweichung – Beobachtung einer seltenen Vogelspezies – Beispiel 4.2 fortgesetzt)

Wir betrachten wieder Beispiel 4.2 und legen eine geeignete Berechnungstabelle an:

i	a_i	a_i^2
1	2	$2^2 = 4$
2	4	$4^2 = 16$
3	1	$1^2 = 1$
4	0	$0^2 = 0$
5	2	$2^2 = 4$
Σ	9	25
$\frac{1}{5}\Sigma$	$\bar{x} = \frac{9}{5} = 1,8$	$\frac{25}{5} = 5$

Also finden wir: $\bar{x} = 1,8$, $\bar{x}^2 = 3,24$,

$$\frac{1}{N} \sum_{i=1}^N a_i^2 \stackrel{N=5}{=} \frac{1}{5} \sum_{i=1}^5 a_i^2 = \frac{25}{5} = 5$$

und somit

$$\begin{aligned} \text{Var} = \sigma^2 &= \frac{1}{N} \sum_{i=1}^N a_i^2 - \bar{x}^2 \\ &\stackrel{N=5}{=} \frac{1}{5} \sum_{i=1}^5 a_i^2 - \bar{x}^2 = 5 - 3,24 = 1,76, \\ \sigma &= \sqrt{1,76} \approx 1,33. \end{aligned}$$

Kommen wir noch einmal auf die Definition der Varianz und der Standardabweichung zurück, und versuchen wir zu verstehen, warum diese so definiert sind.

Bemerkung 4.5. (Erklärung der Formeln für Varianz und Standardabweichung)

*Vorab halten wir fest, dass die Varianz ein **Maß für die Streuung** der Einzelwerte um das arithmetische Mittel sein soll. Wir untersuchen nun, warum die Varianz durch (4.1) bzw. (4.2) und nicht anders definiert ist.*

(1) Es macht **keinen** Sinn

$$\frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})$$

zu betrachten, denn nach Satz 3.4 (1) gilt

$$\sum_{i=1}^N (a_i - \bar{x}) = 0.$$

Daher ersetzt man $a_i - \bar{x}$ durch die **Quadrate** $(a_i - \bar{x})^2 \geq 0$ der Abweichungen $a_i - \bar{x}$ vom arithmetischen Mittel \bar{x} .

(2) Man könnte auch die gewichtete Summe der **Absolutbeträge** $|a_i - \bar{x}|$ der Abweichungen $a_i - \bar{x}$ vom arithmetischen Mittel \bar{x} betrachten, also

$$\text{MAD} := \frac{1}{N} \sum_{i=1}^N |a_i - \bar{x}|. \quad (\text{mittlere absolute Abweichung vom arithmetischen Mittel } \bar{x})$$

Dies wird auch gelegentlich getan. Allerdings hat MAD nicht so schöne mathematische und statistische Eigenschaften wie die Varianz Var und die Standardabweichung σ . (Der Absolutbetrag und damit auch MAD ist keine differenzierbare Funktion.) Daher arbeitet man in der Regel mit der Varianz Var und der Standardabweichung σ .

(3) σ^2 hat die (Einheit der Merkmalswerte)². Das eigentliche Abweichungsmaß ist daher die Standardabweichung σ in der (Einheit der Merkmalswerte).

Wir betrachten nun ein Beispiel mit einer Häufigkeitsverteilung.

Beispiel 4.6. (Varianz und Standardabweichung – Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

i	x_i	h_i
1	0	3
2	1	7
3	2	42
4	3	35
5	4	13
Σ		100

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiele 2.1, 3.3 und 3.12). Dabei sind:

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“
- $h_i =$ Anzahl der Tage, an denen genau x_i (Merkmalswert) Exemplare des Bestsellers verkauft wurden (absolute Häufigkeit)

Wir haben das *arithmetische Mittel* bereits in Beispiel 3.3 berechnet

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k h_i x_i \stackrel{N=100, k=5}{=} \frac{1}{100} (h_1 x_1 + h_2 x_2 + \dots + h_5 x_5) = 2,48 \quad (\text{Exemplare des Bestsellers}),$$

und wir ergänzen unsere Berechnungstabelle nun um eine weitere Spalte:

i	x_i	h_i	$h_i x_i$	$h_i x_i^2$
1	0	3	$3 \cdot 0 = 0$	$3 \cdot 0^2 = 0$
2	1	7	$7 \cdot 1 = 7$	$7 \cdot 1^2 = 7$
3	2	42	$42 \cdot 2 = 84$	$42 \cdot 2^2 = 168$
4	3	35	$35 \cdot 3 = 105$	$35 \cdot 3^2 = 315$
5	4	13	$13 \cdot 4 = 52$	$13 \cdot 4^2 = 208$
Σ		100	248	698
$\frac{1}{100} \Sigma$			$\bar{x} = \frac{248}{100} = 2,48$	$\frac{698}{100} = 6,98$

Nach Lemma 4.3 gilt berechnet sich die *Varianz* zu

$$\text{Var} = \sigma^2 = \frac{1}{N} \sum_{i=1}^k h_i x_i^2 - \bar{x}^2 \stackrel{\substack{N=100, \\ k=5}}{=} \frac{1}{100} \sum_{i=1}^5 h_i x_i^2 - \bar{x}^2 = 6,98 - 2,48^2 = 6,98 - 6,1504 = 0,8296,$$

also $\text{Var} = \sigma^2 = 0,8296$ (Exemplare des Bestsellers)², und die *Standardabweichung* ist $\sigma = \sqrt{0,8296} \approx 0,91$ Exemplare des Bestsellers.

Der nächste Satz hält eine wichtige Eigenschaft des arithmetischen Mittels und der Varianz fest, die wir bereits in Satz 3.4 (2) kennengelernt haben. Nun beweisen wir diese Eigenschaft.

Satz 4.7. (Minimaleigenschaft der Varianz)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N . Das arithmetische Mittel \bar{x} **minimiert** die **mittlere quadratische Abweichung**

$$\text{MQ}(M) := \frac{1}{N} \sum_{i=1}^N (a_i - M)^2, \quad M \in \mathbb{R}.$$

Das bedeutet

$$\text{Var} = \text{MQ}(\bar{x}) < \text{MQ}(M) \quad \text{für alle } M \neq \bar{x}.$$

Frage: Was sagt uns dieser Satz? Das arithmetische Mittel ist der „passende“ Mittelwert zum Streuungsmaß der quadratischen Abweichung von einem Mittelwert, denn dieses Streuungsmaß ist für das arithmetische Mittel minimal. Arithmetisches Mittel und Varianz passen also zusammen!

Beweis von Satz 4.7: Wir starten mit $\text{MQ}(M)$ und versuchen dieses geeignet umzuformen und in die Summe zweier Terme zu zerlegen, von denen einer $\text{MQ}(\bar{x}) = \text{Var}$ ist.

$$\begin{aligned}
 \text{MQ}(M) &= \frac{1}{N} \sum_{i=1}^N (a_i - M)^2 \\
 &= \frac{1}{N} \sum_{i=1}^N \left((a_i - \bar{x}) + (\bar{x} - M) \right)^2 && \text{(Nulladdition } -\bar{x} + \bar{x}\text{)} \\
 &= \frac{1}{N} \sum_{i=1}^N \left((a_i - \bar{x})^2 + 2(a_i - \bar{x})(\bar{x} - M) + (\bar{x} - M)^2 \right) && \text{(1. binomische Formel)} \\
 &= \frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})^2 + \frac{1}{N} \sum_{i=1}^N 2(a_i - \bar{x})(\bar{x} - M) + \frac{1}{N} \sum_{i=1}^N (\bar{x} - M)^2 && \text{(Rechnen mit Summen)} \\
 &= \underbrace{\frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})^2}_{=\text{MQ}(\bar{x})} + 2(\bar{x} - M) \underbrace{\frac{1}{N} \sum_{i=1}^N (a_i - \bar{x})}_{=0} + \frac{1}{N} \sum_{i=1}^N (\bar{x} - M)^2 && \text{(Rechnen mit Summen, Satz 3.4 (1))} \\
 &= \text{MQ}(\bar{x}) + 0 + \frac{1}{N} \cdot N \cdot (\bar{x} - M)^2 \\
 &= \text{MQ}(\bar{x}) + (\bar{x} - M)^2
 \end{aligned}$$

Wir haben also gezeigt, dass gilt

$$\text{MQ}(M) = \text{MQ}(\bar{x}) + \underbrace{(\bar{x} - M)^2}_{>0 \text{ wenn } M \neq \bar{x}}$$

Es folgt $\text{MQ}(M) > \text{MQ}(\bar{x}) = \text{Var}$ für alle $M \neq \bar{x}$. □

Der nächste Satz befasst sich damit, wie sich das arithmetische Mittel, die Varianz und die Standardabweichung ändern, wenn man auf die Einzelwerte eines metrisch skalierten Merkmals eine affin lineare Funktion anwendet. Wir erklären nachher, warum man dies manchmal tun sollte und weshalb die neuen Merkmalswerte nützlich sind.

Satz 4.8. (affin lineare Transformation eines Merkmals)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N . Wendet man die **affin lineare Funktion**

$$y = \beta x + \alpha$$

auf die Einzelwerte a_1, a_2, \dots, a_N an, also $a_i^* = \beta a_i + \alpha$, so gilt für das arithmetische Mittel \bar{x}^* und die Varianz $(\sigma^*)^2$ und die Standardabweichung σ^* der neuen Einzelwerte $a_1^*, a_2^*, \dots, a_N^*$:

$$\bar{x}^* = \frac{1}{N} \sum_{i=1}^N a_i^* = \beta \bar{x} + \alpha \quad \text{und} \quad (\sigma^*)^2 = \frac{1}{N} \sum_{i=1}^N (a_i^* - \bar{x}^*)^2 = \beta^2 \sigma^2, \quad \sigma^* = |\beta| \sigma.$$

Beweis von Satz 4.8: Wir beweisen Satz 4.8 auf einem Übungszettel. \square

Die wichtigste Anwendung von Satz 4.8 ist die Standardisierung eines Merkmals.

Lemma 4.9. (Standardisierung eines Merkmals)

Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten a_1, a_2, \dots, a_N mit arithmetischem Mittel \bar{x} und Standardabweichung σ_X . Wendet man die **affin lineare Funktion**

$$z = \frac{x - \bar{x}}{\sigma_X} = \frac{1}{\sigma_X} x - \frac{\bar{x}}{\sigma_X}$$

auf die Einzelwerte a_1, a_2, \dots, a_N an, also

$$a_i^* := \frac{a_i - \bar{x}}{\sigma_X},$$

so hat das **neue Merkmal** Z mit den Einzelwerten $a_1^*, a_2^*, \dots, a_N^*$ das arithmetische Mittel $\bar{z} = 0$, die Varianz $\sigma_Z^2 = 1$ und die Standardabweichung $\sigma_Z = 1$.

Das Lemma ist ein Sonderfall von Satz 4.8.

Betrachten wir zunächst ein Beispiel zur Standardisierung eines Merkmals.

Beispiel 4.10. (Standardisierung – Beobachtung einer seltenen Vogelspezies)

Wir betrachten wieder die Beobachtung einer seltenen Vogelspezies aus Beispielen 4.2 und 4.4 und erweitern unsere Berechnungstabelle:

i	a_i	a_i^2	$a_i^* = \frac{a_i - \bar{x}}{\sigma_X}$
1	2	$2^2 = 4$	$\frac{2-1,8}{\sqrt{1,76}} \approx 0,151$
2	4	$4^2 = 16$	$\frac{4-1,8}{\sqrt{1,76}} \approx 1,658$
3	1	$1^2 = 1$	$\frac{1-1,8}{\sqrt{1,76}} \approx -0,603$
4	0	$0^2 = 0$	$\frac{0-1,8}{\sqrt{1,76}} \approx -1,357$
5	2	$2^2 = 4$	$\frac{2-1,8}{\sqrt{1,76}} \approx 0,151$
Σ	9	25	0,000
$\frac{1}{5} \Sigma$	$\bar{x} = \frac{9}{5} = 1,8$	$\frac{25}{5} = 5$	0,000

Aus Beispiel 4.4 wissen wir bereits:

$$\bar{x} = 1,8, \quad \sigma_X = \sqrt{1,76} \approx 1,33.$$

Die standardisierten Werte berechnet man also mit:

$$a_i^* = \frac{a_i - \bar{x}}{\sigma_X} = \frac{a_i - 1,8}{\sqrt{1,76}}.$$

Zur Kontrolle haben wir noch die Summe der a_i^* berechnet. Diese muss Null sein (was der Fall ist), da die standardisierten Einzelwerte arithmetisches Mittel Null haben.

Nun kommen wir zum Sinn und Zweck der Standardisierung zurück.

Bemerkung 4.11. (Gründe für die Standardisierung eines Merkmals)

- (1) Für bestimmte statistische Verfahren ist es notwendig, die Merkmale erst zu standardisieren.
- (2) Durch die Standardisierung eines Merkmals ändern sich zwar die Einzelwerte bei einer Häufigkeitsverteilung, aber die **Form des Histogramms bleibt die gleiche**. Lediglich das arithmetische Mittel ist nach der Standardisierung 0, und die Varianz und Standardabweichung sind nach der Standardisierung 1. Bei der Standardisierung handelt es sich also um eine Art „Skalierung“. Wir demonstrieren dieses weiter unten für ein Beispiel.
- (3) Betrachtet man mehrere Merkmale auf der gleichen Grundgesamtheit, z.B. die Größe und das Gewicht aller Studierenden der Universität Paderborn, und möchte deren **Häufigkeitsverteilungen vergleichen**, so lohnt es sich diese zu standardisieren. Nach der „Skalierung“ durch Standardisierung bewegen sich diese in einer ähnlichen Größenordnung (gleiches arithmetisches Mittel 0 und gleiche Standardabweichung $\sigma = 1$) und lassen sich besser vergleichen. (Ein standardisiertes Merkmal ist **einheitenfrei**, denn bei der Standardisierung kürzen sich die Einheiten heraus.) Durch die Standardisierung der metrisch skalierten Merkmale Größe und Gewicht verlieren wir allerdings deren Verhältnisskala und haben nach der Standardisierung nur noch eine Intervallskala!

Betrachten wir noch einmal Beispiel 4.6 und standardisieren unser Merkmal. Danach betrachten wir die Häufigkeitsverteilung und ihr Histogramm, um uns klar zu machen, dass die Standardisierung eines Merkmals die Form der Verteilung nicht ändert.

Beispiel 4.12. (Standardisierung – Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

i	x_i	h_i
1	0	3
2	1	7
3	2	42
4	3	35
5	4	13
Σ		100

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiele 2.1, 3.3, 3.9, 3.12 und 4.6). Dabei sind:

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“
- $h_i =$ Anzahl der Tage, an denen genau x_i (Merkmalswert) Exemplare des Bestsellers verkauft wurden

Wir hatten bereits das *arithmetische Mittel* \bar{x} und die *Standardabweichung* σ_X in Beispiel 3.12 berechnet:

$$\bar{x} = 2,48 \quad \text{und} \quad \sigma_X = \sqrt{0,8296} \approx 0,91.$$

Wir berechnen nun mit einer Berechnungstabelle die Merkmalswerte z_i des standardisierten Merkmals Z mittels

$$z_i = \frac{x_i - \bar{x}}{\sigma_X} = \frac{x_i - 2,48}{\sqrt{0,8296}}.$$

Zur Kontrolle berechnen wir auch das arithmetische Mittel \bar{z} und die Varianz σ_Z^2 von Z , deren Werte $\bar{z} = 0$ bzw. $\sigma_Z^2 = 1$ sein müssen.

i	x_i	h_i	f_i	$z_i = \frac{x_i - \bar{x}_X}{\sigma_X}$	$h_i z_i$	$h_i z_i^2$
1	0	3	0,03	$\frac{0-2,48}{\sqrt{0,8296}} \approx -2,72$	-8,17	22,24
2	1	7	0,07	$\frac{1-2,48}{\sqrt{0,8296}} \approx -1,62$	-11,37	18,48
3	2	42	0,42	$\frac{2-2,48}{\sqrt{0,8296}} \approx -0,53$	-22,13	11,66
4	3	35	0,35	$\frac{3-2,48}{\sqrt{0,8296}} \approx 0,57$	19,98	11,41
5	4	13	0,13	$\frac{4-2,48}{\sqrt{0,8296}} \approx 1,67$	21,69	36,20
Σ		100	1,00		0,00	99,99
$\frac{1}{100} \Sigma$					$\bar{z} = 0,00$	$\sigma_Z^2 \approx 0,9999 \approx 1,00$

Dass wir bei der Berechnung für σ_Z^2 nicht genau 1 erhalten, liegt an Rundungsfehlern. Zu beachten ist auch, dass hier

$$\sigma_Z^2 = \frac{1}{N} \sum_{i=1}^k h_i z_i^2 - \underbrace{\bar{z}^2}_{=0} = \frac{1}{N} \sum_{i=1}^k h_i z_i^2$$

gilt, weil \bar{z} wegen der Standardisierung Null ist.

In Abbildung 4.1 haben wir die Histogramme des Merkmals $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“ und des zugehörigen standardisierten Merkmals Z gezeichnet. Wir sehen, dass die Verteilungen der beiden Merkmale genau die gleiche Form haben; nur die Beschriftung der horizontalen Achse hat sich geändert.

Als **Motivation für den Variationskoeffizient**, der als nächstes definiert wird, stellen wir uns zunächst das folgende Szenario vor:

Betrachtet werden die Verkaufszahlen pro Tag eines neuen Lebensmittelprodukts in 1000 Geschäften, die dieses Produkt führen. In der Erhebung sind sowohl die großen Lebensmittelgeschäfte von Supermarktketten, also auch die Supermärkte in Kaufhäusern und kleine „Tante-Emma-Läden“ vertreten. Wir betrachten das Merkmal $X =$ „Anzahl der pro Tag in dem Geschäft verkauften Exemplare des Lebensmittelprodukts“.

- Für jedes beobachtete Geschäft erhalten wir eine Häufigkeitsverteilung mit arithmetischem Mittel \bar{x} und Standardabweichung σ .
- Das arithmetische Mittel, also die durchschnittliche Verkaufszahl des Produkts, und die Standardabweichung, also die Streuung dieser Verkaufszahlen, wird in einem großen Supermarkt **viel größer** sein als in einem kleinen „Tante-Emma-Laden“. Es macht daher **keinen Sinn** diese Kennzahlen zu vergleichen!

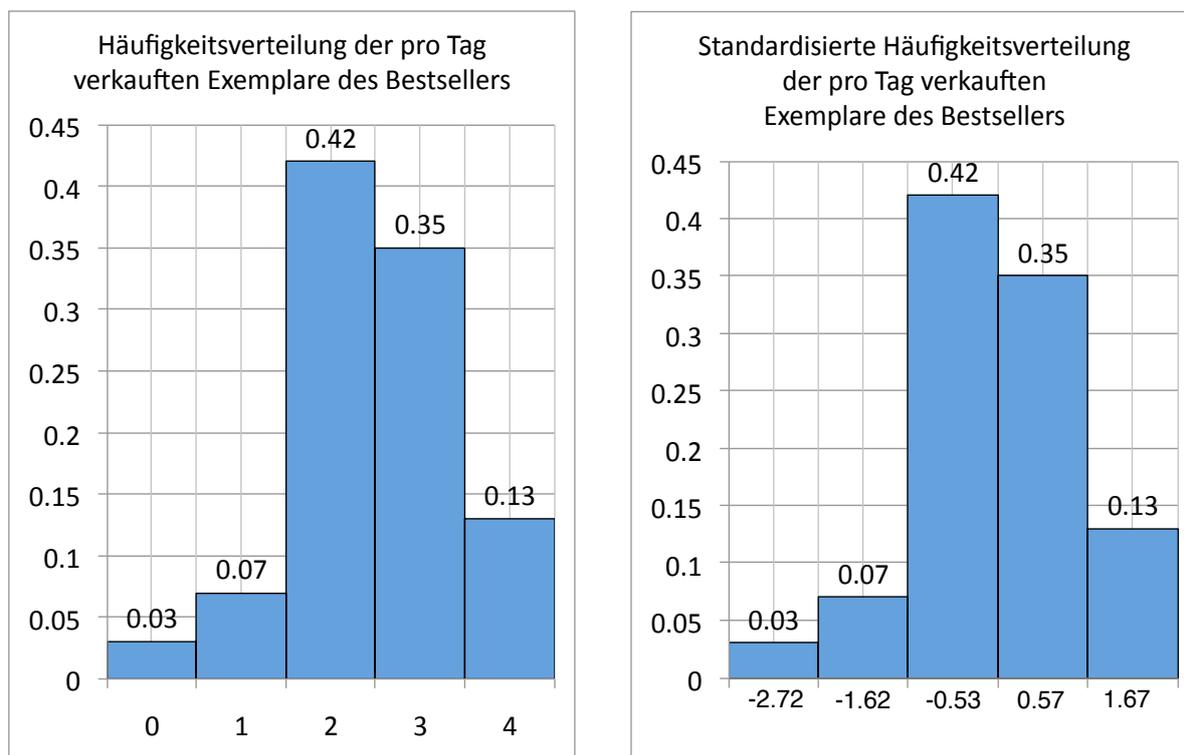


Abbildung 4.1: Links das Histogramm der relativen Häufigkeitsverteilung des Merkmals $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“ und rechts die Häufigkeitsverteilung des zugehörigen standardisierten Merkmals Z . Die „Form“ der beiden Verteilungen ist gleich; nur die Beschriftung der horizontalen Achse hat sich geändert.

- Trotzdem könnten die Histogramme der Verteilungen ähnlich sein! Wie können wir die Häufigkeitsverteilungen **einfach vergleichen**? Um die Verteilung der Verkaufszahlen in den verschiedenen Geschäften zu vergleichen, brauchen wir ein **relatives Streuungsmaß**!

Definition 4.13. (Variationskoeffizient für verhältnisskalierte Merkmale)

Sei X ein metrisch skaliertes Merkmal auf einer Verhältnisskala auf einer Grundgesamtheit von N Elementen mit dem arithmetischen Mittel \bar{x} und der Standardabweichung σ . Dann definieren wir den **Variationskoeffizient** VC als **relatives Streuungsmaß**:

$$VC := \frac{\sigma}{\bar{x}} \quad \left(= \frac{\text{Standardabweichung}}{\text{arithmetisches Mittel}} \right) \quad \text{bzw. in Prozent: } VC = \frac{\sigma}{\bar{x}} \cdot 100\%.$$

Interpretation: Je größer der Variationskoeffizient ist, desto größer ist auch die (relative) Streuung der Einzelwerte. Der Variationskoeffizient ist dimensionslos.

Betrachten wir auch hierzu zwei Beispiele.

Beispiel 4.14. (Variationskoeffizient – Beobachtung einer seltenen Vogelspezies – Bsp. 4.4 fortgesetzt)

Aus Beispiel 4.4 wissen wir bereits:

$$\bar{x} = 1,8, \quad \sigma = \sqrt{1,76} \approx 1,33.$$

Der Variationskoeffizient des Merkmals $X =$ „Anzahl der pro Tag gesichteten Vögel“ ist also

$$VC = \frac{\sigma}{\bar{x}} = \frac{\sqrt{1,76}}{1,8} \approx 0,737.$$

Beispiel 4.15. (Variationskoeffizient – Anzahl der pro Tag verkauften Exemplare eines Bestellers)

Wir setzen Beispiel 4.6 fort. Wir wissen bereits, dass für das Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestellers“ das arithmetische Mittel bzw. die Standardabweichung durch

$$\bar{x} = 2,48 \quad \text{bzw.} \quad \sigma = \sqrt{0,8296} \approx 0,91$$

gegeben sind. Also ist der Variationskoeffizient

$$VC = \frac{\sigma}{\bar{x}} = \frac{\sqrt{0,8296}}{2,48} \approx 0,367.$$

Vergleicht man Beispiel 4.14 und Beispiel 4.15 so sieht man man dass bei dem Merkmal „Anzahl der (pro Tag) gesichteten Vögel“ eine deutlich größere relative Streuung als bei dem Merkmal „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestellers“ vorliegt, da der Variationskoeffizient in Beispiel 4.14 ungefähr doppelt so groß ist wie der Variationskoeffizient in Beispiel 4.15.

Abschließend soll noch kurz auf den Fall eines **Merkmals mit Klassenbildung** eingegangen werden, bei dem die Einzelwerte **nicht** vorliegen. Hier berechnet man das arithmetische Mittel, die Varianz und die Standardabweichung, indem man den Häufigkeiten der Klassen die Klassenmitten

$$x'_i = \frac{1}{2} (x_i^o + x_i^u)$$

zuordnet und für diese „künstlichen Merkmalswerte“ die entsprechende Formel mit den Häufigkeiten verwendet. Dabei ignoriert man bei den Streuungsmaßen σ^2 und σ gezwungenermaßen die **Streuung innerhalb der einzelnen Klassen**; hierfür gibt es **Korrekturterme**.

4.2 Spannweite, Quartile und Box-and-Whisker-Plot

In diesem Teilkapitel betrachten wir ein mindestens ordinalskaliertes Merkmal und betrachten zusätzlich zu der Situation eines Merkmals mit Einzelwerten (bzw. deren Häufigkeitsverteilung) auch die Situation, wenn klassifizierte Daten vorliegen.

Wir führen zunächst die Spannweite eines metrisch skalierten Merkmals ein.

Definition 4.16. (Spannweite für ein metrisch skaliertes Merkmal)

- (1) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den nach aufsteigender Größe sortierten Einzelwerten

$$a_{[1]} \leq a_{[2]} \leq \dots \leq a_{[N-1]} \leq a_{[N]}.$$

Dann ist die **Spannweite** von X definiert als

$$R = a_{[N]} - a_{[1]}.$$

- (2) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den k verschiedenen Merkmalswerten $x_1 < x_2 < \dots < x_{k-1} < x_k$. Dann ist die **Spannweite** von X definiert als

$$R = x_k - x_1.$$

- (3) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit Klassenbildung mit k Klassen mit den Klassengrenzen

$$x_1^u < x_1^o = x_2^u < x_2^o = x_3^u < \dots < x_{k-1}^o = x_k^u < x_k^o.$$

Dann ist die **Spannweite** von X definiert als

$$R = x_k^o - x_1^u = \left(\begin{array}{c} \text{obere Klassengrenze der} \\ \text{obersten auftretenden Klasse} \end{array} \right) - \left(\begin{array}{c} \text{untere Klassengrenze der} \\ \text{untersten auftretenden Klasse} \end{array} \right).$$

Betrachten wir einige Beispiele.

Beispiel 4.17. (Spannweite eines metrisch skalierten Merkmals)

- (a) Bei dem Merkmal $X =$ „Anzahl der pro Tag gesichteten Vögel“ in Beispielen 3.7, 4.2, 4.4 und 4.14 ist die Spannweite

$$R = a_{[5]} - a_{[1]} = 4 - 0 = 4 \quad (\text{Vögel}).$$

- (b) Bei dem Merkmal $X =$ „Anzahl der pro Tag in der Buchhandlung verkauften Exemplare des Bestsellers“ in Beispielen 2.1, 3.3, 3.9, 3.12, 4.6 und 4.15 ist die Spannweite

$$R = x_5 - x_1 = 4 - 0 = 4 \quad (\text{Bücher pro Tag}).$$

- (c) Im Beispiel 2.3 mit dem Merkmal $X =$ „Preis des Gebrauchtwagens“ mit Klassenbildung ist die Spannweite

$$R = x_6^o - x_1^u = 15 - 0 = 15, \quad \text{also: } R = 15000 \text{ Euro.}$$

Als nächstes führen wir die Quartile ein. Dazu brauchen wir die „Aufrundungsfunktion“ $\lceil x \rceil$, definiert durch

$$\lceil x \rceil = \begin{cases} x & \text{wenn } x \in \mathbb{Z}, \\ k & \text{mit } k \in \mathbb{Z} \text{ und } k - 1 < x < k, \text{ wenn } x \notin \mathbb{Z}. \end{cases}$$

Definition 4.18. (Quartile für ein metrisch skaliertes Merkmal)

(1) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den nach aufsteigender Größe sortierten Einzelwerten

$$a_{[1]} \leq a_{[2]} \leq \dots \leq a_{[N-1]} \leq a_{[N]}.$$

Dann sind die Quartile wie folgt definiert:

- Das **untere Quartil** ist $Q_1 = a_{[\lceil N/4 \rceil]}$.
- Das **mittlere Quartil** ist der Median, also $Q_2 = \text{Me}$.
- Das **obere Quartil** ist $Q_3 = a_{[\lceil 3N/4 \rceil]}$.

(2) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den k verschiedenen Merkmalswerten $x_1 < x_2 < \dots < x_{k-1} < x_k$, und sei $F : \mathbb{R} \rightarrow [0, 1]$ die relative Summenhäufigkeitsfunktion von X . Dann sind die Quartile wie folgt definiert:

- Das **untere Quartil** Q_1 ist der kleinste Merkmalswert x_i für den gilt $F(x_i) \geq 0,25$.
- Das **mittlere Quartil** ist der Median, also $Q_2 = \text{Me}$.
- Das **obere Quartil** Q_3 ist der kleinste Merkmalswert x_i für den gilt $F(x_i) \geq 0,75$.

(3) Sei X ein metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit Klassenbildung mit k Klassen und der relativen Summenhäufigkeitsfunktion $F : \mathbb{R} \rightarrow [0, 1]$. Dann sind die Quartile wie folgt definiert:

- Das **untere Quartil** Q_1 ist der x -Wert $x = Q_1$ mit $F(Q_1) = 0,25$.
- Das **mittlere Quartil** Q_2 ist der x -Wert $x = Q_2$ mit $F(Q_2) = 0,5$.
- Das **obere Quartil** Q_3 ist der x -Wert $x = Q_3$ mit $F(Q_3) = 0,75$.

Die Quartile unterteilen die Grundgesamtheit in **vier (ungefähr) gleich große Teilgesamtheiten**. Der **Quartilabstand** $QA = Q_3 - Q_1$ gibt die Spannweite der 50% mittleren Werte an.

Die Bedeutung der Quartile ist in dem oberen Bild in Abbildung 4.2 **graphisch dargestellt**: Zwischen dem unteren Quartil Q_1 und dem oberen Quartial Q_3 liegen die mittleren 50% der Einzelwerte. Unterhalb (also links) von Q_1 liegen die unteren 25% der Einzelwerte, und oberhalb (also rechts) von Q_3 liegen die oberen 25% der Einzelwerte.

Graphisch kann man die Streuung einer Verteilung gut mit Hilfe der Spannweite und der Quartile mit einem **Box-and-Whisker-Plot** („Schachtel und Barthaar Schaubild“) darstellen (siehe das untere Bild in Abbildung 4.2 für den Fall mit Einzelwerten). Dabei gilt:

- Der linke Rand der Schachtel ist bei Q_1 .
- Der rechte Rand der Schachtel ist bei Q_3 .
- Die vertikale Linie in der Schachtel ist bei $Q_2 = \text{Me}$.
- Das linke Barthaar endet bei $a_{[1]}$ (kleinster Einzelwert) bzw. bei x_1 (kleinster Merkmalswert) bzw. bei x_1^u (untere Grenze der untersten Klasse bei klassifizierten Daten).
- Das rechte Barthaar endet bei $a_{[N]}$ (größter Einzelwert) bzw. bei x_k (größter Merkmalswert). bzw. bei x_k^o (obere Grenze der obersten Klasse bei klassifizierten Daten).

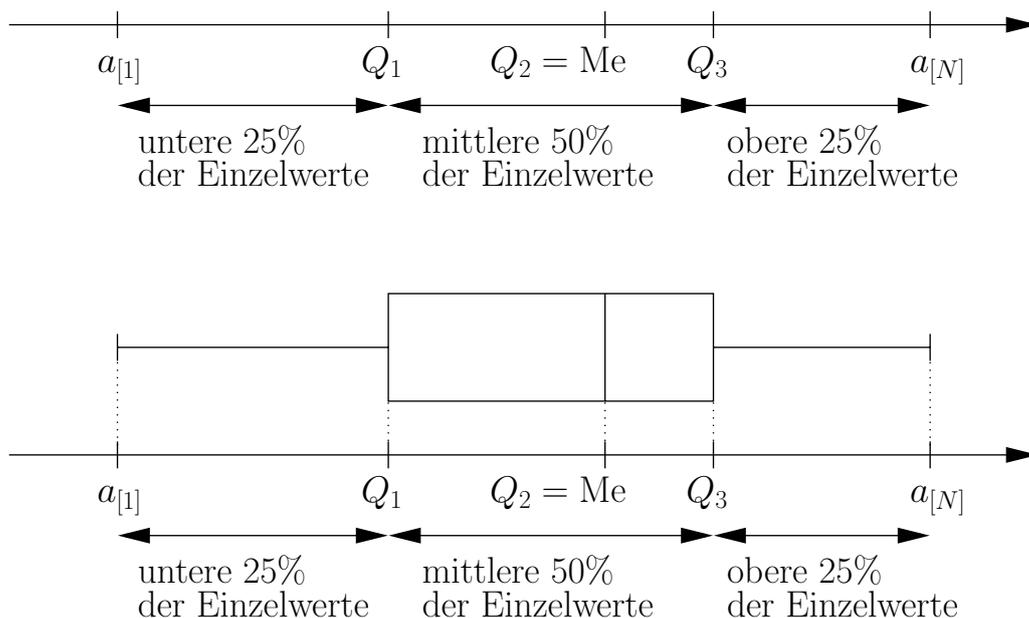


Abbildung 4.2: Veranschaulichung der Quartile und des Box-and-Whisker-Plots für den Fall von Einzelwerten.

In der Schachtel befinden sich die mittleren 50% der Einzelwerte. Links von der Schachtel in dem Bereich bis zum Ende des linken Barthaars befinden sich die unteren 25% der Einzelwerte, und rechts von der Schachtel in dem Bereich bis zum Ende des rechten Barthaars befinden sich die oberen 25% der Einzelwerte.

Betrachten wir einige Beispiele zu den Quartilen und dem Box-and-Whisker-Plot.

Beispiel 4.19. (Quartile, Box-and-Whisker-Plot – Beobachtung einer seltenen Vogelspezies)

Ein Naturforscher beobachtet eine seltene Vogelspezies über 5 Tage und zählt jeweils die Anzahl der am Tag gesichteten Vögel dieser Vogelspezies. Er findet für das Merkmal $X =$ „Anzahl der pro Tag gesichteten Vögel“ die folgenden Einzelwerte:

i	1	2	3	4	5
a_i	2	4	1	0	2

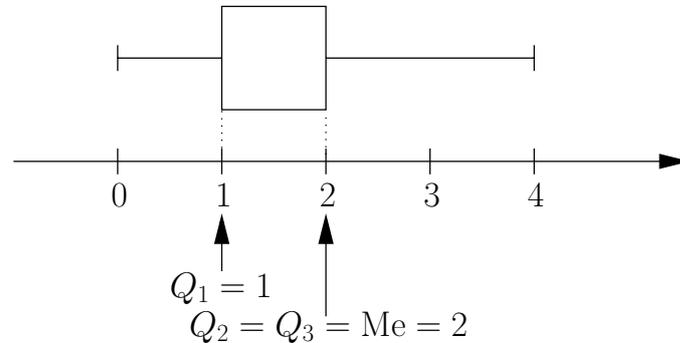
Wir sortieren die Einzelwerte der Größe nach:

$$\underbrace{0}_{=a_{[1]}} \leq \underbrace{1}_{=a_{[2]}} \leq \underbrace{2}_{=a_{[3]}} \leq \underbrace{2}_{=a_{[4]}} \leq \underbrace{4}_{=a_{[5]}}$$

Wir finden wegen $\lceil 5/4 \rceil = \lceil 1,25 \rceil = 2$ und $\lceil 3 \cdot 5/4 \rceil = \lceil 3,75 \rceil = 4$, dass

$$Q_1 = a_{[2]} = 1, \quad Q_2 = \text{Me} = a_{[3]} = 2 \quad Q_3 = a_{[4]} = 2.$$

Der kleinste Einzelwert ist $a_{[1]} = 0$ und der größte Einzelwert ist $a_{[5]} = 4$. Also sieht das Box-and-Whisker-Plot wie folgt aus:



Beispiel 4.20. (Quartile, Box-and-Whisker-Plot – Anzahl der pro Tag verkauften Exemplare eines Bestsellers)

Ein Buchhändler beobachtet 100 Tage lang die Anzahl der täglich verkauften Exemplare eines Bestsellers und erhält die folgende Häufigkeitsverteilung (vgl. auch Beispiele 2.1, 3.3, 3.9, 3.12, 4.6, 4.15 und 4.17). Dabei sind:

i	x_i	h_i	H_i	F_i
1	0	3	3	0,03
2	1	7	10	0,10
3	2	42	52	0,52
4	3	35	87	0,87
5	4	13	100	1,00
Σ		100		

- Grundgesamtheit: $N = 100$ Beobachtungstage
- Merkmal $X =$ „Anzahl der pro Tag in der in der Buchhandlung verkauften Exemplare des Bestsellers“
- $h_i =$ Anzahl der Tage, an denen genau x_i Exemplare des Bestsellers verkauft wurden

In Beispiel 3.9 hatten wir den Median bereits berechnet.

$$\text{Me} = \frac{1}{2} (a_{[50]} + a_{[51]}) = \frac{1}{2} (2 + 2) = 2.$$

Also gilt $Q_2 = \text{Me} = 2$.

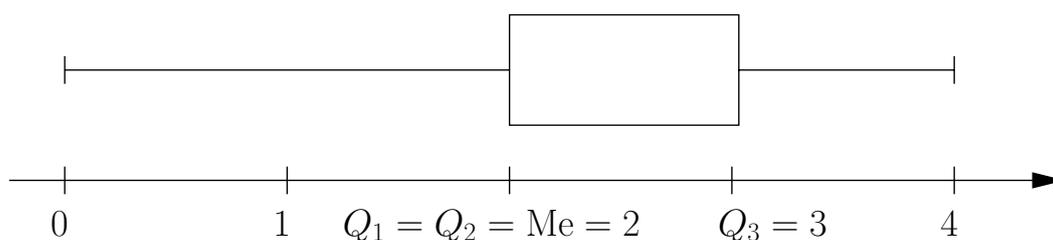
Der kleinste Merkmalswert x_i mit $F(x_i) \geq 0,25$ ist $x_3 = 2$. Also ist $Q_1 = x_3 = 2$.

Der kleinste Merkmalswert x_i mit $F(x_i) \geq 0,75$ ist $x_4 = 3$. Also ist $Q_3 = x_4 = 3$.

Wir finden also die Quartile

$$Q_1 = 2 \qquad Q_2 = \text{Me} = 2, \qquad Q_3 = 3.$$

Der kleinste Merkmalswert ist $x_1 = 0$ und der größte Merkmalswert ist $x_5 = 4$. Also sieht das Box-and-Whisker-Plot wie folgt aus:



Liegt ein Merkmal mit Klassenbildung vor, so hat man verschiedene Möglichkeiten die Quartile mit Definition 4.18 (3) zu bestimmen:

Zum einen kann man die relative Summenhäufigkeitsfunktion F als stückeweise affin lineare Funktion durch eine Formel beschreiben und dann die Gleichungen $F(x) = 0,25$, $F(x) = 0,5$ bzw. $F(x) = 0,75$ jeweils nach x auflösen.

Einfacher (wenn auch etwas ungenauer) ist die graphische Bestimmung der Quartile aus dem Graphen der Summenhäufigkeitsfunktion: Dazu zeichnen wir jeweils eine Parallele zur horizontalen Achse durch 0,25 bzw. 0,5 bzw. 0,75 und fällen im Schnittpunkt mit dem Graphen von $F(x)$ jeweils ein Lot auf die horizontale Achse. Die Schnittpunkte mit der horizontalen Achse liefern Q_1 bzw. Q_2 bzw. Q_3 .

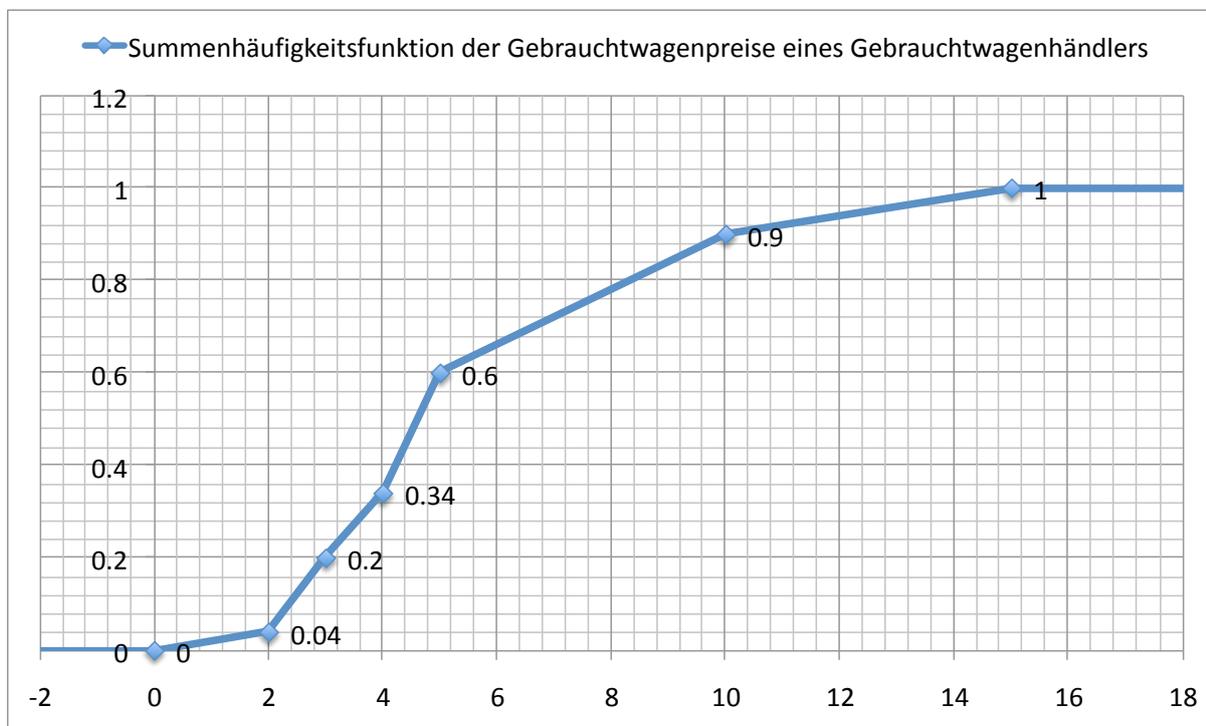
Wir üben beide Vorgehensweisen nun an einem Beispiel.

Beispiel 4.21. (Quartile, Box-and-Whisker-Plot – Preisverteilung beim Autobestand eines Gebrauchtwagenhändlers mit Klassenbildung)

Ein Gebrauchtwagenhändler mit $N = 50$ Gebrauchtwagen nimmt eine Inventur seines Autobestands vor, wobei er die Autos in Preisklassen einteilt und die folgende Häufigkeitsverteilung mit Klassenbildung für das Merkmal $X =$ „Preis des Gebrauchtwagens (in 1000 Euro)“ erhält (vgl. Beispiel 2.3):

i	Klasse (Preis in 1000 Euro)	x_i^u	x_i^o	Δx_i	h_i	h_i^*	f_i	f_i^*	H_i	F_i
1	0 bis 2	0	2	2	2	1	0,04	0,02	2	0,04
2	über 2 bis 3	2	3	1	8	8	0,16	0,16	10	0,2
3	über 3 bis 4	3	4	1	7	7	0,14	0,14	17	0,34
4	über 4 bis 5	4	5	1	13	13	0,26	0,26	30	0,6
5	über 5 bis 10	5	10	5	15	3	0,3	0,06	45	0,9
6	über 10 bis 15	10	15	5	5	1	0,1	0,02	50	1,0
Σ					50		1,0			

In der nachfolgenden Zeichnung ist der Graph der relativen Summenhäufigkeitsfunktion noch einmal abgebildet.



Am Graphen der relativen Summenhäufigkeitsfunktion oder an den relativen Summenhäufigkeiten F_i in der Tabelle lesen wir ab, dass F den Wert 0,25 im Intervall $[3, 4]$ annimmt, wo F als affin lineare Funktion $F_1(x) = \beta x + \alpha$ durch die beiden Punkte $(3, 0,2)$ und $(4, 0,34)$ eindeutig festgelegt ist. Wir berechnen die Steigung

$$\beta = \frac{0,34 - 0,2}{4 - 3} = 0,14$$

und den Achsenabschnitt

$$0,2 = 0,14 \cdot 3 + \alpha \quad \Leftrightarrow \quad \alpha = 0,2 - 0,14 \cdot 3 = 0,2 - 0,42 = -0,22.$$

Also ist

$$F(x) = F_1(x) = 0,14 \cdot x - 0,22 \quad \text{für } x \in [3, 4],$$

und Auflösen von $F_1(x) = 0,25$ liefert

$$0,25 = 0,14 \cdot x - 0,22 \quad \Leftrightarrow \quad 0,47 = 0,14 \cdot x \quad \Leftrightarrow \quad x = \frac{0,47}{0,14} \approx 3,36.$$

Also ist $Q_1 \approx 3,36$.

Am Graphen der relativen Summenhäufigkeitsfunktion oder an den relativen Summenhäufigkeiten F_i in der Tabelle lesen wir ab, dass F den Wert 0,5 im Intervall $[4, 5]$ annimmt,

wo F als affin lineare Funktion $F_2(x) = \beta x + \alpha$ durch die beiden Punkte $(4, 0,34)$ und $(5, 0,6)$ eindeutig festgelegt ist. Wir berechnen die Steigung

$$\beta = \frac{0,6 - 0,34}{5 - 4} = 0,26$$

und den Achsenabschnitt

$$0,6 = 0,26 \cdot 5 + \alpha \quad \Longleftrightarrow \quad \alpha = 0,6 - 0,26 \cdot 5 = 0,6 - 1,3 = -0,7.$$

Also ist

$$F(x) = F_2(x) = 0,26 \cdot x - 0,7 \quad \text{für } x \in [3, 4],$$

und Auflösen von $F_2(x) = 0,5$ liefert

$$0,5 = 0,26 \cdot x - 0,7 \quad \Longleftrightarrow \quad 1,2 = 0,26 \cdot x \quad \Longleftrightarrow \quad x = \frac{1,2}{0,26} \approx 4,62.$$

Also ist $Q_2 \approx 4,62$.

Am Graphen der relativen Summenhäufigkeitsfunktion oder an den relativen Summenhäufigkeiten F_i in der Tabelle lesen wir ab, dass F den Wert $0,75$ im Intervall $[5, 10]$ annimmt, wo F als affin lineare Funktion $F_3(x) = \beta x + \alpha$ durch die beiden Punkte $(5, 0,6)$ und $(10, 0,9)$ eindeutig festgelegt ist. Wir berechnen die Steigung

$$\beta = \frac{0,9 - 0,6}{10 - 5} = \frac{0,3}{5} = 0,06$$

und den Achsenabschnitt

$$0,9 = 0,06 \cdot 10 + \alpha \quad \Longleftrightarrow \quad \alpha = 0,9 - 0,06 \cdot 10 = 0,9 - 0,6 = 0,3.$$

Also ist

$$F(x) = F_3(x) = 0,06 \cdot x + 0,3 \quad \text{für } x \in [5, 10],$$

und Auflösen von $F_3(x) = 0,75$ liefert

$$0,75 = 0,06 \cdot x + 0,3 \quad \Longleftrightarrow \quad 0,45 = 0,06 \cdot x \quad \Longleftrightarrow \quad x = \frac{0,45}{0,06} = 7,5.$$

Also ist $Q_3 = 7,5$.

Alternativ bestimmen wir die Quartile mit einer etwas größeren Ungenauigkeit grafisch in Abbildung 4.3: Dazu zeichnen wir jeweils eine Parallele zur horizontalen Achse durch $0,25$ bzw. $0,5$ bzw. $0,75$ und fällen im Schnittpunkt mit dem Graphen von $F(x)$ jeweils ein Lot auf die horizontale Achse. Die Schnittpunkte mit der horizontalen Achse liefern Q_1 bzw. Q_2 bzw. Q_3 .

Wir lesen an der Zeichnung (siehe unten) ab, dass

$$Q_1 \approx 3,4, \quad Q_2 \approx 4,6 \quad \text{und} \quad Q_3 \approx 7,5.$$

Die untere Grenze der untersten Klasse ist $u_1^u = 0$ und die obere Grenze der obersten Klasse ist $x_6^o = 15$. Das Box-and-Whisker-Plot der Häufigkeitsverteilung ist in der nachfolgenden Zeichnung dargestellt:

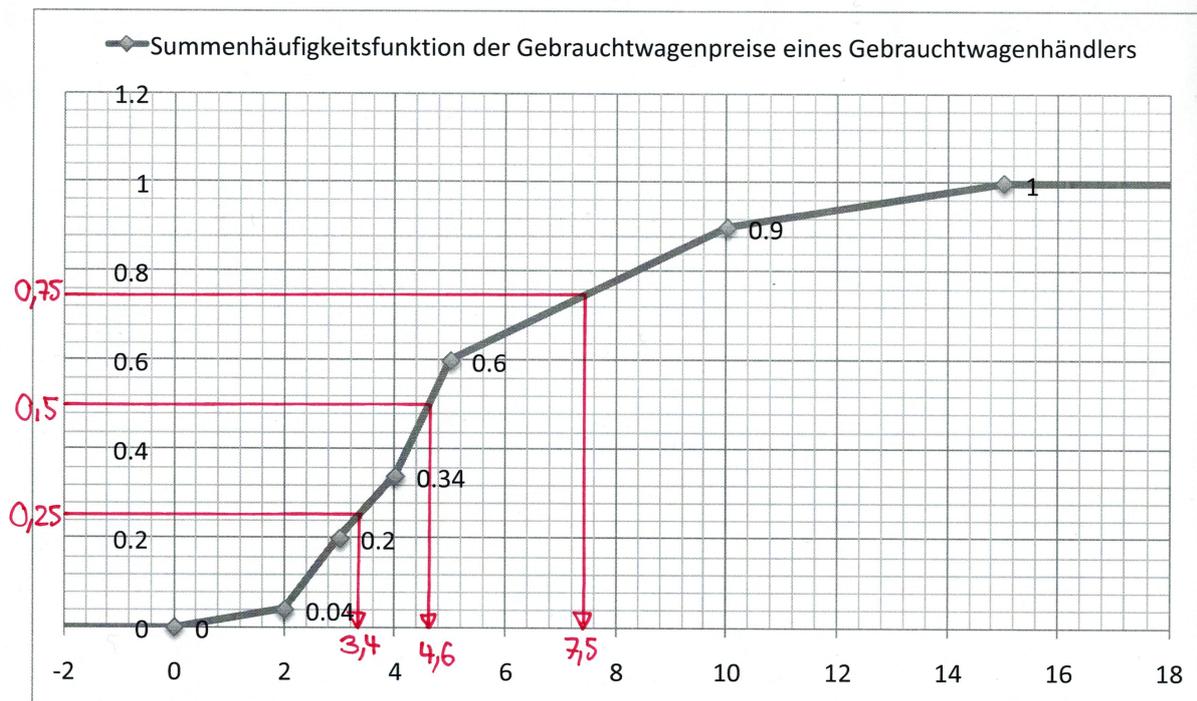
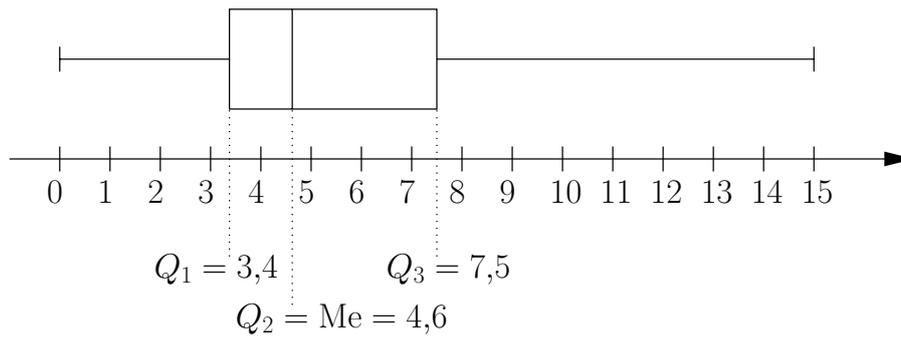


Abbildung 4.3: Grafische Bestimmung der Quartile im Beispiel 4.21.

Zweidimensionale Häufigkeitsverteilungen, Korrelation und Kovarianz

In diesem Teilkapitel betrachten wir nicht nur ein sondern zwei Merkmale auf einer (gemeinsamen) Grundgesamtheit. Wir können dann fragen, **ob es zwischen diesen Merkmalen eine (affin lineare) Zusammenhang gibt**. Betrachten wir ein Beispiel:

Nehmen wir die Menge der volljährigen Einwohner Paderborns als Grundgesamtheit und beobachten die Merkmale Körpergröße und Gewicht, so erwarten wir, dass zwischen diesen beiden Merkmalen ein (möglicherweise näherungsweise affin linearer) Zusammenhang besteht: Je größer eine Person ist, desto schwerer ist sie in der Regel auch. Natürlich finden wir keinen ganz klaren funktionalen Zusammenhang bedingt durch individuelle Schwankungen in Körperbau und Gewicht innerhalb der Gruppe aller Personen mit derselben Körpergröße. Trotzdem wird eine statistische Analyse der (zweidimensionalen) Verteilung der beiden Merkmale Körpergröße und Gewicht einen näherungsweise affin linearen funktionalen Zusammenhang dieser beiden Merkmale enthüllen.

An diesem Beispiel haben wir bereits die wesentlichen Ideen dieses Kapitels kennengelernt:

Wir betrachten also die (gemeinsame zweidimensionale) Verteilung zweier Merkmale X und Y auf einer Grundgesamtheit. Dieses ist eine sogenannte **zweidimensionale Häufigkeitsverteilung**, bei der jeder Kombination von Merkmalswerten (x_i, y_j) (wobei x_i ein Merkmalswert von X und y_j ein Merkmalswert von Y ist) eine absolute Häufigkeit $h_{i,j}$ zugeordnet wird. Natürlich könnten wir die beiden Merkmale auch separat betrachten, aber die beiden (eindimensionalen) Verteilungen der Merkmale würden keine Zusammenhänge zwischen den Merkmalen X und Y aufzeigen können. Wir erhalten die eindimensionalen Verteilungen als **Randverteilungen** (mit **Randhäufigkeiten**) und betrachten weiter sogenannte **bedingte Verteilungen** (mit **bedingten Häufigkeiten**). Stimmen alle bedingten Verteilungen jeweils mit der entsprechenden Randverteilung überein, so haben wir **unabhängig verteilte Merkmale**. Es gibt dann in den Daten keinen Hinweis auf einen Zusammenhang zwischen den beiden Merkmalen.

Liegt dagegen ein (**affin**) **linearer Zusammenhang** zwischen den beiden Merkmalen X und Y vor, so spricht man von „**Korrelation**“ – wir sagen X und Y sind (positiv oder negativ) korreliert. Wir messen eine solche Korrelation, indem wir die **Kovarianz** der beiden Merkmale aus der zweidimensionalen Häufigkeitsverteilung berechnen. Bewertet wird die Korrelation, also ein möglicher affin linearer Zusammenhang zweier Merkmale, mittels des **Korrelationskoeffizienten**, der sich aus der Kovarianz und den Standardabweichungen der beiden Merkmale berechnet.

5.1 Zweidimensionale Häufigkeitsverteilung

Wir beobachten **zwei verschiedene Merkmale** X und Y an den N Elementen der selben Grundgesamtheit. Dabei seien:

- Merkmalswerte von X : $x_1 < x_2 < \dots < x_k$
- Merkmalswerte von Y : $y_1 < y_2 < \dots < y_\ell$
- Einzelwerte (a_t, b_t) , $t = 1, 2, \dots, N$, wobei a_t bzw. b_t der Einzelwert von X bzw. Y am t -ten Element der Grundgesamtheit ist.

Will man die Merkmale X und Y nicht jeweils separat betrachten, so stellen sich die folgenden Fragen:

Frage: Wie oft tritt die Kombination (x_i, y_j) der Merkmalswerte x_i (von X) und y_j (von Y) auf? Dieses führt uns zu dem neuen Begriff einer **zweidimensionalen Häufigkeitsverteilung**.

Frage: Was für „Trends“ kann man an den zweidimensionalen Häufigkeiten ablesen? Beispielsweise: Wenn die Merkmalswerte von X groß werden, was passiert dann mit den Merkmalswerten von Y ? Oder hängen die Merkmale nicht voneinander ab? Die letzte Frage führt auf den Begriff „**unabhängig verteilt**“, den wir in Teilkapitel 5.3 kennenlernen werden. Fragen wir nach der Natur des (möglichen) Zusammenhangs zwischen den beiden Merkmalen so kommen wir zu den Begriffen der **Korrelation** und **Kovarianz** in Teilkapitel 5.4.

Wir führen zunächst den Begriff einer zweidimensionalen Häufigkeitsverteilung als direkte Verallgemeinerung des Begriffs einer (eindimensionalen) Häufigkeitsverteilung (siehe Teilkapitel 2.1) ein.

Definition 5.1. (zweidimensionale Häufigkeitsverteilung)

Seien X und Y zwei diskrete metrisch skalierte (oder ordinalskalierte) Merkmale auf einer Grundgesamtheit von N Elementen. Dabei habe X die Merkmalswerte $x_1 < x_2 < \dots < x_k$, und Y habe die Merkmalswerte $y_1 < y_2 < \dots < y_\ell$. Das **zweidimensionale Merkmal** (X, Y) hat eine **zweidimensionale Häufigkeitsverteilung** mit:

(1) **zweidimensionalen Merkmalswerten** (oder **Merkmalsausprägungen**):

$$(x_i, y_j), \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, \ell$$

(2) (a_t, b_t) , $t = 1, 2, \dots, N$, bezeichnen die **zweidimensionalen Einzelwerte** der N Elemente in der Grundgesamtheit. Genauer hat das t -te Element der Grundgesamtheit den Einzelwert a_t für X bzw. b_t für Y .

(3) $h_{i,j} =$ **absolute Häufigkeit**, mit der (x_i, y_j) auftritt

(4) $f_{i,j} = \frac{h_{i,j}}{N} =$ **relative Häufigkeit**, mit der (x_i, y_j) auftritt

(5) **absoluten Summenhäufigkeiten:**

$$H_{i,j} = \sum_{r=1}^i \sum_{s=1}^j h_{r,s} = h_{1,1} + \dots + h_{1,j} + \dots + h_{i,1} + \dots + h_{i,j}$$

= Anzahl der Elemente mit Merkmalswert (x, y) mit $x \leq x_i$ und $y \leq y_j$

(6) **relativen Summenhäufigkeiten:**

$$F_{i,j} = \sum_{r=1}^i \sum_{s=1}^j f_{r,s} = f_{1,1} + \dots + f_{1,j} + \dots + f_{i,1} + \dots + f_{i,j} = \frac{H_{i,j}}{N}$$

= Anteil der Elemente mit Merkmalswert (x, y) mit $x \leq x_i$ und $y \leq y_j$

Betrachten wir ein Beispiel.

Beispiel 5.2. (2-dim. Häufigkeitsverteilung – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Bei $N = 100$ Studienanfängern des Bachelorstudiengangs Mathematik wird beobachtet, wie viele Lehrbücher sie im ersten Semester jeweils für die Vorlesung Analysis 1 (Merkmal X) bzw. für die Vorlesung Lineare Algebra 1 (Merkmal Y) kaufen.

Wir haben also ein *zweidimensionales Merkmal* (X, Y) mit

- $X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“ mit den Merkmalswerten $x_1 = 0$, $x_2 = 1$ und $x_3 = 2$
- $Y =$ „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“ mit den Merkmalswerten $y_1 = 0$, $y_2 = 1$ und $y_3 = 2$
- *absoluten Häufigkeiten* $h_{i,j}$ bzw. *relativen Häufigkeiten* $f_{i,j}$ der *zweidimensionalen Merkmalswerte* (x_i, y_j)

Tabelle der *2-dimensionalen absoluten Häufigkeitsverteilung*

$h_{2,1} = 15$ bedeutet: „15 Studienanfänger kauften genau ein ($x_2 = 1$) Lehrbuch für die Analysis 1 und kein ($y_1 = 0$) Lehrbuch für die Lineare Algebra 1.“

$h_{3,2} = 5$ bedeutet: „5 Studienanfänger kauften genau zwei ($x_3 = 2$) Lehrbücher für die Analysis 1 und genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1.“

		j		
		1	2	3
i	$x_i \setminus y_j$	0	1	2
1	0	20	5	0
2	1	15	20	15
3	2	0	5	20

Zusammenhang zwischen der absoluten und relativen 2-dim. Verteilung: $f_{i,j} = \frac{h_{i,j}}{N} = \frac{h_{i,j}}{100}$

Tabelle der 2-dimensionalen relativen Häufigkeitsverteilung

$f_{2,1} = 0,15$ bedeutet: „Ein Anteil von 0,15 der Studienanfänger kauften genau ein ($x_2 = 1$) Lehrbuch für die Analysis 1 und kein ($y_1 = 0$) Lehrbuch für die Lineare Algebra 1.“

$f_{3,2} = 0,05$ bedeutet: „Ein Anteil von 0,05 der Studienanfänger kauften genau zwei ($x_3 = 2$) Lehrbücher für die Analysis 1 und genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1.“

		j		
		1	2	3
i	$x_i \setminus y_j$	0	1	2
1	0	0,2	0,05	0
2	1	0,15	0,2	0,15
3	2	0	0,05	0,2

Beobachtet man nur separat die Anzahl der gekauften Bücher für die Analysis 1 bzw. die Lineare Algebra 1, so gewinnt man **keine** Information über einen **möglichen Zusammenhang** zwischen den Buchkäufen für die beiden Vorlesungen. Wir vermuten aber, dass ein Student, der für eine der beiden Vorlesungen Lehrbücher kauft, dieses auch für die andere Vorlesung tut. Die Daten der zweidimensionalen Häufigkeitsverteilung scheinen dieses zu belegen, da längs der Diagonale (von links oben nach rechts unten) besonders große Häufigkeiten auftreten. Wie werden noch lernen, wie man einen solchen möglichen Zusammenhang untersucht.

5.2 Randverteilung und bedingte Häufigkeiten

Wir betrachten weiterhin ein zweidimensionales Merkmal (X, Y) mit den Merkmalswerten (x_i, y_j) , $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$ (wobei $x_1 < x_2 < \dots < x_k$ die Merkmalswerte von X und $y_1 < y_2 < \dots < y_\ell$ die Merkmalswerte von Y sind). Wie im vorigen Teilkapitel sei $h_{i,j}$ bzw. $f_{i,j}$ die absolute bzw. relative Häufigkeit des zweidimensionalen Merkmalswertes (x_i, y_j) .

Aus der zweidimensionalen Häufigkeitsverteilung kann man die **Häufigkeitsverteilung jedes einzelnen Merkmals** als sogenannte **Randverteilung** bestimmen. Betrachten wir hierzu zunächst wieder unser Beispiel 5.2 aus dem vorigen Teilkapitel.

Beispiel 5.3. (Randverteilungen – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Bei $N = 100$ Studienanfänger des Bachelorstudiengangs Mathematik wird beobachtet, wie viele Lehrbücher sie im ersten Semester jeweils für die Vorlesung Analysis 1 (Merkmal X) bzw. für die Vorlesung Lineare Algebra 1 (Merkmal Y) kaufen (vgl. Beispiel 5.2). Wir haben also ein *zweidimensionales Merkmal* (X, Y) mit

$X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“ mit den Merkmalswerte $x_1 = 0$, $x_2 = 1$ und $x_3 = 2$,

Y = „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“ mit den Merkmalswerte $y_1 = 0$, $y_2 = 1$ und $y_3 = 2$,

und den *absoluten Häufigkeiten* $h_{i,j}$ bzw. *relativen Häufigkeiten* $f_{i,j}$ der *zweidimensionalen Merkmalswerte* (x_i, y_j) .

Absolute und relative Randhäufigkeiten von X:

- $h_{i\cdot}$ gibt die absolute Häufigkeit an, mit der das Merkmal X den Wert x_i annimmt.
- $f_{i\cdot}$ gibt die relative Häufigkeit an, mit der das Merkmal X den Wert x_i annimmt.

In Formeln: $h_{i\cdot} = h_{i,1} + h_{i,2} + h_{i,3}$ und $f_{i\cdot} = f_{i,1} + f_{i,2} + f_{i,3}$.

Erklärung der Formeln für die absoluten/relativen Randhäufigkeiten von X: Bei der Berechnung von $h_{i\cdot}$ (bzw. $f_{i\cdot}$) wird jeweils über die absoluten (bzw. relativen) Häufigkeiten der zweidimensionalen Merkmalsausprägungen summiert, bei denen X den Wert x_i hat.

Absolute und relative Randhäufigkeiten von Y:

- $h_{\cdot j}$ gibt die absolute Häufigkeit an, mit der das Merkmal Y den Wert y_j annimmt.
- $f_{\cdot j}$ gibt die relative Häufigkeit an, mit der das Merkmal Y den Wert y_j annimmt.

In Formeln: $h_{\cdot j} = h_{1,j} + h_{2,j} + h_{3,j}$ und $f_{\cdot j} = f_{1,j} + f_{2,j} + f_{3,j}$.

Erklärung der Formeln für die absoluten/relativen Randhäufigkeiten von Y: Bei der Berechnung von $h_{\cdot j}$ (bzw. $f_{\cdot j}$) wird jeweils über die absoluten (bzw. relativen) Häufigkeiten der zweidimensionalen Merkmalsausprägungen summiert, bei denen Y den Wert y_j hat.

Tabelle der 2-dimensionalen absoluten Häufigkeitsverteilung mit den *absoluten Randhäufigkeiten*

$h_{1\cdot} = 25$ bedeutet: „25 Studienanfänger kauften kein ($x_1 = 0$) Lehrbuch für die Analysis 1.“

$h_{2\cdot} = 30$ bedeutet: „30 Studienanfänger kauften genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1.“

	j	1	2	3	
i	$x_i \setminus y_j$	0	1	2	$h_{i\cdot}$
1	0	20	5	0	25
2	1	15	20	15	50
3	2	0	5	20	25
	$h_{\cdot j}$	35	30	35	100

Interpretation der absoluten Randverteilungen:

$x_1 = 0$, $x_2 = 1$, $x_3 = 2$ mit den absoluten Randhäufigkeiten $h_{1\cdot} = 25$, $h_{2\cdot} = 50$, $h_{3\cdot} = 25$ gibt die *absolute Häufigkeitsverteilung des Merkmals X* = „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“ an (unabhängig von Merkmal Y).

$y_1 = 0$, $y_2 = 1$, $y_3 = 2$ mit den absoluten Randhäufigkeiten $h_{\cdot 1} = 35$, $h_{\cdot 2} = 30$, $h_{\cdot 3} = 35$ gibt die *absolute Häufigkeitsverteilung des Merkmals Y* „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“ an (unabhängig von Merkmal X).

Tabelle der 2-dimensionalen absoluten Häufigkeitsverteilung mit den *absoluten Randhäufigkeiten*

$f_{1,\cdot} = 0,25$ bedeutet: „Ein Anteil von 0,25 der Studienanfänger kaufte kein ($x_1 = 0$) Lehrbuch für die Analysis 1.“

$f_{\cdot,2} = 0,3$ bedeutet: „Ein Anteil von 0,3 der Studienanfänger kaufte genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1.“

	j	1	2	3	
i	$x_i \setminus y_j$	0	1	2	$f_{i,\cdot}$
1	0	0,2	0,05	0	0,25
2	1	0,15	0,2	0,15	0,5
3	2	0	0,05	0,2	0,25
	$f_{\cdot,j}$	0,35	0,3	0,35	1,0

Interpretation der relativen Randverteilungen:

$x_1 = 0, x_2 = 1, x_3 = 2$ mit den relativen Randhäufigkeiten $f_{1,\cdot} = 0,25, f_{2,\cdot} = 0,5, f_{3,\cdot} = 0,25$ gibt die *relative Häufigkeitsverteilung des Merkmals X* = „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“ an (unabhängig von Merkmal Y).

$y_1 = 0, y_2 = 1, y_3 = 2$ mit den relativen Randhäufigkeiten $f_{\cdot,1} = 0,35, f_{\cdot,2} = 0,3, f_{\cdot,3} = 0,35$ gibt die *relative Häufigkeitsverteilung des Merkmals Y* = „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“ an (unabhängig von Merkmal X).

Wir halten die neuen statistischen Begriffe in einer Definition fest.

Definition 5.4. (eindimensionale Randverteilungen mit Randhäufigkeiten)

Sei (X, Y) ein zweidimensionales metrisch skaliertes (oder ordinalskaliertes) Merkmal mit den Merkmalswerten (x_i, y_j) und den zugehörigen absoluten bzw. relativen Häufigkeiten $h_{i,j}$ bzw. $f_{i,j}$, wobei $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$ (und $x_1 < x_2 < \dots < x_k$ sind die Merkmalswerte von X und $y_1 < y_2 < \dots < y_\ell$ die Merkmalswerte von Y). Wir definieren für $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$:

$$h_{i,\cdot} = \sum_{j=1}^{\ell} h_{i,j} = \text{absolute Randhäufigkeit, mit der } X \text{ den Merkmalswert } x_i \text{ annimmt,}$$

$$h_{\cdot,j} = \sum_{i=1}^k h_{i,j} = \text{absolute Randhäufigkeit, mit der } Y \text{ den Merkmalswert } y_j \text{ annimmt,}$$

$$f_{i,\cdot} = \sum_{j=1}^{\ell} f_{i,j} = \text{relative Randhäufigkeit mit der } X \text{ den Merkmalswert } x_i \text{ annimmt,}$$

$$f_{\cdot,j} = \sum_{i=1}^k f_{i,j} = \text{relative Randhäufigkeit mit der } Y \text{ den Merkmalswert } y_j \text{ annimmt.}$$

Analog zu Summenhäufigkeiten kann man Randsummenhäufigkeiten erklären. Dieses wird auf einem Übungszettel behandelt.

Kommen wir nun noch einmal zu unserem Beispiel des Kauf von Lehrbüchern zurück und betrachten die folgenden **Frage**: Was ist die **relative** Häufigkeitsverteilung der Anzahl der für die Analysis 1 gekauften Lehrbücher, **wenn** für die Lineare Algebra 1 beispielsweise genau $y_2 = 1$ Lehrbuch gekauft wurde? Dies führt uns zu dem Begriff einer **bedingten Häufigkeitsverteilung**. Wir betrachten dieses zunächst für unser Beispiel 5.3.

Beispiel 5.5. (bedingte relative Häufigkeiten – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Sei (X, Y) das zweidimensionale Merkmal aus Beispielen 5.2 und 5.3, wobei $X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“,

$Y =$ „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“.

In Beispielen 5.2 und 5.3 hatten wir die relative Häufigkeitsverteilung mit den zugehörigen Randhäufigkeiten berechnet:

	j	1	2	3	
i	$x_i \setminus y_j$	0	1	2	$f_{i\cdot}$
1	0	0,2	0,05	0	0,25
2	1	0,15	0,2	0,15	0,5
3	2	0	0,05	0,2	0,25
	$f_{\cdot j}$	0,35	0,3	0,35	1,0

Frage 1: Was ist die *relative Häufigkeitsverteilung* der Anzahl der für die Analysis 1 gekauften Lehrbücher, *wenn* genau $y_2 = 1$ Lehrbuch für die Lineare Algebra 1 gekauft wurde?

Frage 2: Was ist die *relative Häufigkeitsverteilung* der Anzahl der für die Lineare Algebra 1 gekauften Lehrbücher, *wenn* genau $x_3 = 2$ Lehrbücher für die Analysis 1 gekauft wurden?

Antwort auf Frage 1: Wir beziehen die relativen Häufigkeiten $f_{i,2}$, $i = 1, 2, 3$, nun auf die Randhäufigkeit $f_{\cdot 2} = 0,3$ (relative Randhäufigkeit, mit der genau $y_2 = 1$ Lehrbücher für die Lineare Algebra 1 gekauft wurden), also

$$f_{i,(2)} = \frac{f_{i,2}}{f_{\cdot 2}}, \quad i = 1, 2, 3.$$

Man nennt die $f_{i,(2)}$, $i = 1, 2, 3$, *bedingte (relative) Häufigkeiten*, denn es sind Häufigkeiten unter der *Bedingung*, dass genau $y_2 = 1$ Lehrbücher für die Lineare Algebra 1 gekauft wurden.

Interpretation:

Unter den Studienanfängern, die genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1 gekauft haben, hat ein Anteil von $f_{1,(2)} \approx 0,167$ dieser Studienanfänger kein ($x_1 = 0$) Lehrbuch für die Analysis 1 gekauft.

Unter den Studienanfängern, die genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1 gekauft haben, hat ein Anteil von $f_{2,(2)}$ dieser Studienanfänger genau $x_2 = 1$ Lehrbücher für die Analysis 1 gekauft.

i	x_i	$f_{i,2}$	$f_{i,(2)}$
1	0	0,05	$\frac{0,05}{0,3} = \frac{1}{6} \approx 0,167$
2	1	0,2	$\frac{0,2}{0,3} = \frac{2}{3} \approx 0,667$
3	2	0,05	$\frac{0,05}{0,3} = \frac{1}{6} \approx 0,167$
Σ		0,3	1,0

Antwort auf Frage 2: Wir beziehen die relativen Häufigkeiten $f_{3,j}$, $j = 1, 2, 3$, nun auf die Randhäufigkeit $f_{3,\cdot} = 0,25$ (relative Randhäufigkeit, mit der genau $x_3 = 2$ Lehrbücher für die Analysis 1 gekauft wurden), also

$$f_{(3),j} = \frac{f_{3,j}}{f_{3,\cdot}}, \quad j = 1, 2, 3.$$

Man nennt die $f_{(3),j}$, $j = 1, 2, 3$, *bedingte (relative) Häufigkeiten*, denn es sind Häufigkeiten unter der *Bedingung*, dass genau $x_3 = 2$ Lehrbücher für die Analysis 1 gekauft wurden.

Interpretation:

Unter den Studienanfängern, die genau zwei ($x_3 = 2$) Lehrbücher für die Analysis 1 gekauft haben, hat ein Anteil von $f_{(3),2} = 0,2$ dieser Studienanfänger genau ein ($y_2 = 1$) Lehrbuch für die Lineare Algebra 1 gekauft.

Unter den Studienanfängern, die genau ein ($x_3 = 1$) Lehrbücher für die Analysis 1 gekauft haben, hat ein Anteil von $f_{(3),j}$ dieser Studienanfänger genau y_j Lehrbücher für die Lineare Algebra 1 gekauft.

j	y_j	$f_{3,j}$	$f_{(3),j}$
1	0	0	$\frac{0}{0,25} = 0$
2	1	0,05	$\frac{0,05}{0,25} = \frac{1}{5} = 0,2$
3	2	0,2	$\frac{0,2}{0,25} = \frac{4}{5} = 0,8$
Σ		0,25	1,0

Die restlichen vier bedingten Häufigkeitsverteilungen sollte man zur Übung selber ausrechnen.

Wir halten die neuen statistischen Begriffe in einer Definition fest.

Definition 5.6. (bedingte (relative) Häufigkeitsverteilungen)

Sei (X, Y) ein zweidimensionales metrisch skaliertes (oder ordinalskaliertes) Merkmal mit den Merkmalswerten (x_i, y_j) und den zugehörigen relativen Häufigkeiten $f_{i,j}$, wobei $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$ (und $x_1 < x_2 < \dots < x_k$ sind die Merkmalswerte von X und $y_1 < y_2 < \dots < y_\ell$ die Merkmalswerte von Y). Wir definieren für $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$:

(1) $f_{i,(j)} = \frac{f_{i,j}}{f_{\cdot,j}} =$ **bedingte (relative) Häufigkeit**, mit welcher der Merkmalswert x_i

bei Merkmal X auftritt **unter der Annahme/Bedingung**, dass das Merkmal Y den Wert y_j annimmt

(2) $f_{(i),j} = \frac{f_{i,j}}{f_{i,\cdot}} =$ **bedingte (relative) Häufigkeit**, mit welcher der Merkmalswert y_j

bei Merkmal Y auftritt **unter der Annahme/Bedingung**, dass das Merkmal X den Wert x_i annimmt.

5.3 Unabhängig verteilte Merkmale

Mit den relativen Randverteilungen und den bedingten relativen Häufigkeitsverteilungen können wir untersuchen, ob zwischen zwei Merkmalen ein funktionaler Zusammenhang besteht. Genauer definieren wir nun sogenannte **unabhängig verteilte Merkmale**. Sind zwei Merkmale (auf der gleichen Grundgesamtheit) unabhängig verteilt, so besteht zwischen ihnen kein funktionaler Zusammenhang.

Definition 5.7. (unabhängig verteilte Merkmale)

Seien die Voraussetzungen und Bezeichnungen wie in Definition 5.6. Die Merkmale X und Y heißen **unabhängig verteilt**, wenn alle bedingten (relativen) Häufigkeitsverteilungen von X bzw. Y mit den jeweiligen Randverteilungen übereinstimmen, d.h. wenn gilt:

$$f_{i,(j)} = f_{i,\cdot} \quad \text{und} \quad f_{(i),j} = f_{\cdot,j} \quad \text{für alle } i = 1, 2, \dots, k \text{ und alle } j = 1, 2, \dots, \ell.$$

Betrachten wir zwei Beispiele zu der neuen Definition.

Beispiel 5.8. (nicht unabhängig verteilte Merkmale – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Sei (X, Y) das zweidimensionale Merkmal aus Beispielen 5.2, 5.3 und 5.5, wobei $X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“, $Y =$ „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“. In Beispielen 5.3 und 5.5 haben wir berechnet, dass $f_{(3),1} = 0$ und $f_{\cdot,1} = 0,35$ ist. Also gilt

$$0 = f_{(3),1} \neq f_{\cdot,1} = 0,35,$$

und wir sehen, dass die Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher (Merkmal X) und die Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher (Merkmal Y) *nicht unabhängig verteilt* sind.

Beispiel 5.9. (unabhängig verteilte Merkmale)

Gegeben seien zwei Merkmale X und Y auf derselben Grundgesamtheit:

Das Merkmal X hat die Merkmalsausprägungen $x_1 = 0$ und $x_2 = 1$.

Das Merkmal Y hat die Merkmalsausprägungen $y_1 = 1$ und $y_2 = 2$.

Die zweidimensionale relative Häufigkeitsverteilung ist in der Tabelle angegeben, und wir berechnen zunächst die Randhäufigkeiten.

	j	1	2	
i	$x_i \setminus y_j$	1	2	$f_{i,\cdot}$
1	0	0,02	0,08	0,1
2	1	0,18	0,72	0,9
	$f_{\cdot,j}$	0,2	0,8	1,0

Um zu untersuchen, ob die Merkmale unabhängig verteilt sind, berechnen wir die bedingten Verteilungen:

i	x_i	$f_{i,1}$	$f_{i,(1)}$	$f_{i,2}$	$f_{i,(2)}$
1	0	0,02	0,1	0,08	0,1
2	1	0,18	0,9	0,72	0,9
	Σ	0,2	1,0	0,8	1,0

j	y_j	$f_{1,j}$	$f_{(1),j}$	$f_{2,j}$	$f_{(2),j}$
1	1	0,02	0,2	0,18	0,2
2	2	0,08	0,8	0,72	0,8
	Σ	0,1	1,0	0,9	1,0

Wir finden:

$$\begin{aligned} f_{1,(1)} = f_{1,(2)} = f_{1,\cdot} &= 0,1, & f_{(1),1} = f_{(2),1} = f_{\cdot,1} &= 0,2, \\ f_{2,(1)} = f_{2,(2)} = f_{2,\cdot} &= 0,9, & f_{(1),2} = f_{(2),2} = f_{\cdot,2} &= 0,8. \end{aligned}$$

Also sind die Merkmale X und Y *unabhängig verteilt*.

Bemerkung 5.10. (unabhängig verteilte Merkmale)

In der Praxis wird man selten gemäß der obigen Definition unabhängig verteilte Merkmale finden. Man wird daher in der praktischen Anwendung Merkmale **auch als unabhängig verteilt betrachten**, wenn gilt:

$$f_{i,(j)} \approx f_{i,\cdot} \quad \text{und} \quad f_{(i),j} \approx f_{\cdot,j} \quad \text{für alle } i = 1, 2, \dots, k \text{ und alle } j = 1, 2, \dots, \ell.$$

Zuletzt lernen wir einen Satz, mit dem man leichter untersuchen kann, ob zwei Merkmale unabhängig verteilt sind.

Satz 5.11. (alternative Charakterisierung unabhängig verteilter Merkmale)

Seien die Voraussetzungen und Bezeichnungen wie in Definition 5.6. Die Merkmale X und Y sind genau dann **unabhängig verteilt**, wenn gilt:

$$f_{i,j} = f_{i,\cdot} \cdot f_{\cdot,j} \quad \text{für alle } i = 1, 2, \dots, k \text{ und alle } j = 1, 2, \dots, \ell$$

(d.h. wenn die zweidimensionalen relativen Häufigkeiten das Produkt der jeweiligen relativen Randhäufigkeiten sind).

Beweis von Satz 5.11: Der Beweis folgt aus der Definition der bedingten Häufigkeiten und der Definition von unabhängig verteilten Merkmalen. Wir führen den Beweis auf einem Übungszettel durch. \square

Betrachten wir für die Anwendung von Satz 5.11 noch einmal Beispiel 5.9.

Beispiel 5.12. (unabhängig verteilter Merkmale)

Gegeben seien die zwei Merkmale X und Y auf derselben Grundgesamtheit aus Beispiel 5.9. In Beispiel 5.9 hatten wir bereits die Randhäufigkeiten berechnet.

Wir testen nun mit Satz 5.11, ob die beiden Merkmale unabhängig verteilt sind: Es gilt

	j	1	2	
i	$x_i \setminus y_j$	1	2	$f_{i\cdot}$
1	0	0,02	0,08	0,1
2	1	0,18	0,72	0,9
	$f_{\cdot j}$	0,2	0,8	1,0

$$0,02 = f_{1,1} = f_{1\cdot} \cdot f_{\cdot 1} = 0,1 \cdot 0,2 = 0,02,$$

$$0,18 = f_{2,1} = f_{2\cdot} \cdot f_{\cdot 1} = 0,9 \cdot 0,2 = 0,18,$$

$$0,08 = f_{1,2} = f_{1\cdot} \cdot f_{\cdot 2} = 0,1 \cdot 0,8 = 0,08,$$

$$0,72 = f_{2,2} = f_{2\cdot} \cdot f_{\cdot 2} = 0,9 \cdot 0,8 = 0,72,$$

und nach Satz 5.11 sind die Merkmale X und Y somit unabhängig verteilt.

5.4 Korrelation, Kovarianz und Korrelationskoeffizient

In diesem Teilkapitel werden wir nun den Zusammenhang zwischen den beiden (nicht unabhängig verteilten) metrisch skalierten Merkmalen einer zweidimensionalen Häufigkeitsverteilung untersuchen und klassifizieren.

Definition 5.13. (positive bzw. negative Korrelation)

- (1) Gilt für die Einzelwerte $(a_t, b_t), t = 1, \dots, N$, eines zweidimensionalen metrisch skalierten Merkmals (X, Y) , dass tendenziell b_t desto größer ist je größer a_t ist, dann nennt man X und Y **positiv korreliert**. Man spricht von **positiver Korrelation**.
- (2) Gilt für die Einzelwerte $(a_t, b_t), t = 1, 2, \dots, N$, eines zweidimensionalen metrisch skalierten Merkmals (X, Y) , dass tendenziell b_t desto kleiner ist je größer a_t ist, dann nennt man X und Y **negativ korreliert**. Man spricht von **negativer Korrelation**.
- (3) Mathematisch präziser: **Korrelation** misst **affin lineare Zusammenhänge**, also Zusammenhänge der Form

$$Y = \beta X + \alpha$$

zwischen zwei metrisch skalierten Merkmalen X und Y auf derselben Grundgesamtheit.

- (4) Streuen die Einzelwerte $(a_t, b_t), t = 1, 2, \dots, N$, eines zweidimensionalen metrisch skalierten Merkmals (X, Y) ohne erkenntlichen linearen Zusammenhang zwischen den Werten von X und von Y , so nennt man X und Y **unkorreliert**.

Betrachten wir jeweils ein Beispiel für positive Korrelation und ein Beispiel für negative Korrelation.

Beispiel 5.14. (positiv korrelierte Merkmale)

Es wurden die Körpergröße in Zentimeter (Merkmal X) und das Gewicht in Kilogramm (Merkmal Y) von 12 Personen mit einem Alter von 20 bis 30 Jahren erhoben.

Nummer	t	1	2	3	4	5	6	7	8	9	10	11	12
Größe in cm	a_t	170	180	170	160	160	175	173	185	190	182	187	193
Gewicht in kg	b_t	65	75	70	55	50	70	63	80	90	80	90	85

Die Einzelwerte sind in einem Streudiagramm in Abbildung 5.1 gezeichnet. Wir sehen an dem affin linearen Trend: Körpergröße und Gewicht sind *positiv korreliert*.

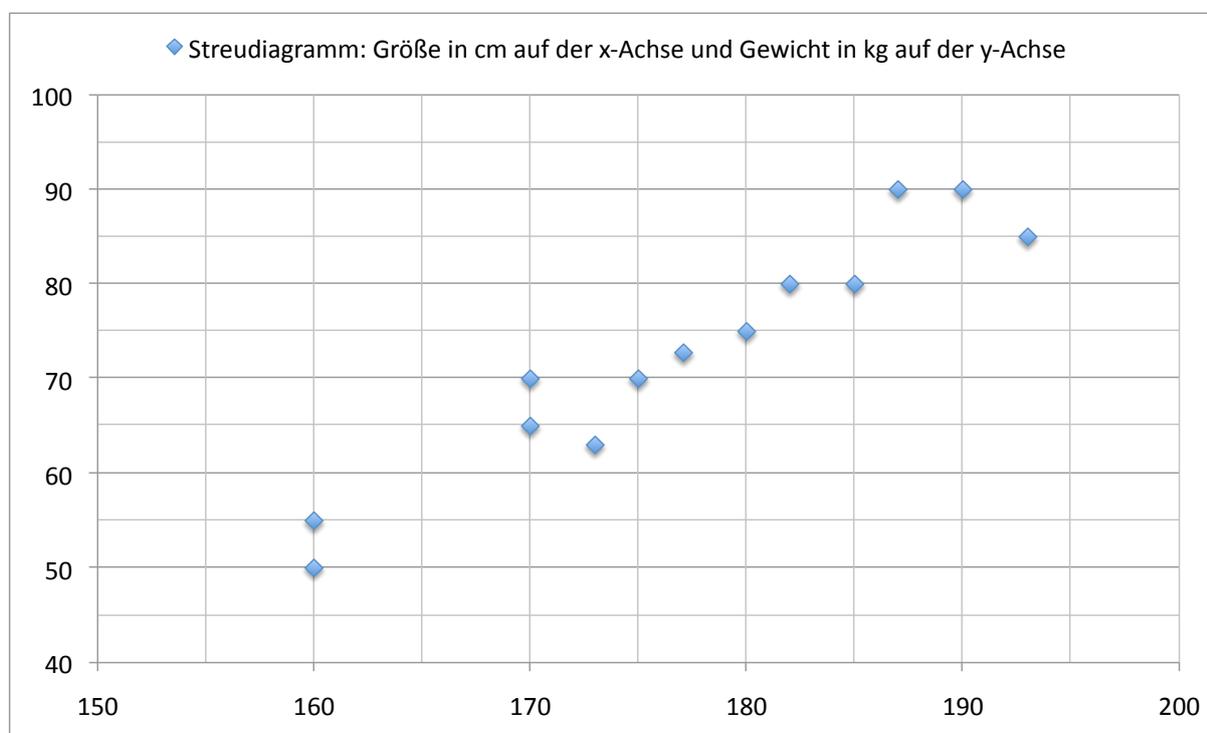


Abbildung 5.1: Streudiagramm des zweidimensionalen Merkmals aus Beispiel 5.14.

Beispiel 5.15. (negativ korrelierte Merkmale)

Es wurden die Trainingstunden pro Woche in Stunden (Merkmal X) und die 10 km-Laufzeiten in Minuten (Merkmal Y) von 8 männlichen Schülern erhoben.

Nummer	t	1	2	3	4	5	6	7	8
Lauftraining in h pro Woche	a_t	0	3	6	4	1	2	10	5
Laufzeit auf 10 km in min	b_t	70	55	40	45	65	60	38	45

Die Einzelwerte sind in einem Streudiagramm in Abbildung 5.2 gezeichnet. Wir sehen an dem affin linearen Trend: Stunden Lauftraining pro Woche und die Laufzeit über 10 km sind *negativ korreliert*.

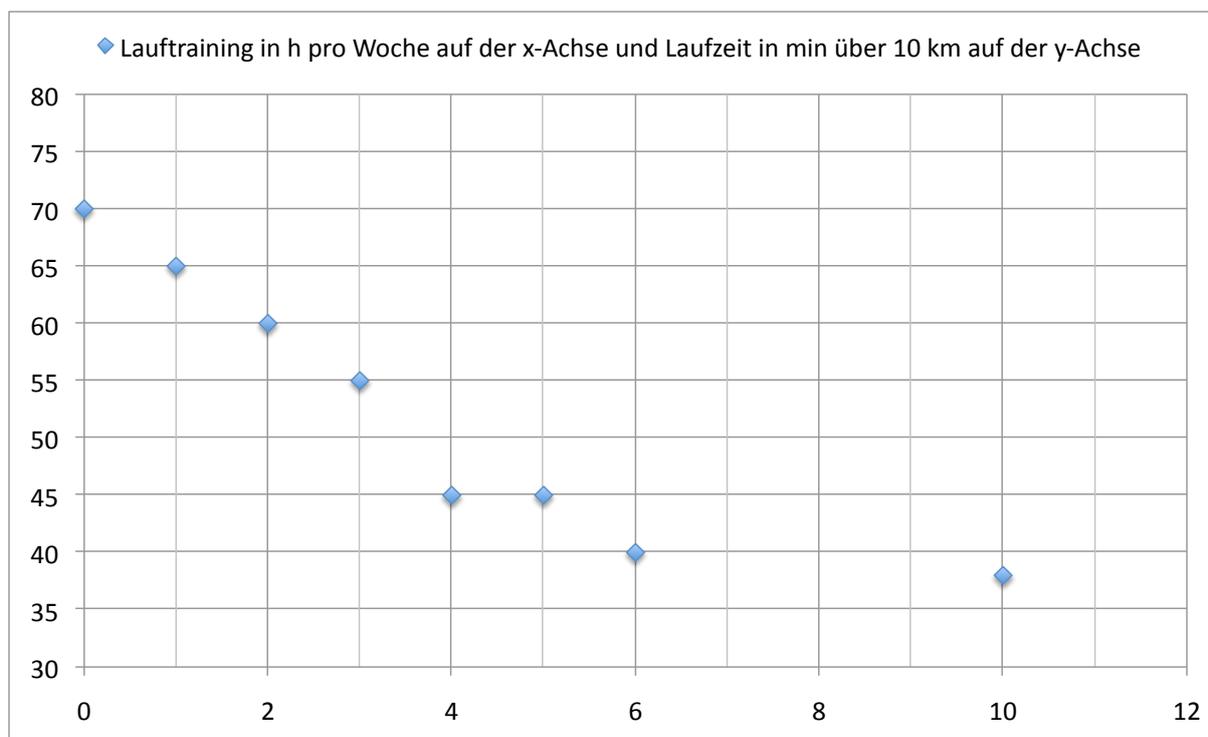


Abbildung 5.2: Streudiagramm des zweidimensionalen Merkmals aus Beispiel 5.15.

Um positive bzw. negative Korrelation mathematisch sauber messen zu können, führen wir nun die Kovarianz ein.

Definition 5.16. (Kovarianz und Korrelationskoeffizient)

Sei (X, Y) ein zweidimensionales metrisch skaliertes Merkmal auf einer Grundgesamtheit von N Elementen mit den Einzelwerten (a_t, b_t) , $t = 1, 2, \dots, N$, und Merkmalswerten (x_i, y_j) und den zugehörigen absoluten Häufigkeiten $h_{i,j}$, wobei $i = 1, 2, \dots, k$ und $j = 1, 2, \dots, \ell$ (und $x_1 < x_2 < \dots < x_k$ bzw. $y_1 < y_2 < \dots < y_\ell$ sind die Merkmalswerte für X bzw. Y). Seien \bar{x} bzw. \bar{y} das arithmetische Mittel des Merkmals X bzw. Y .

(1) Die **Kovarianz** σ_{XY} der Merkmale X und Y

$$\sigma_{XY} = \frac{1}{N} \sum_{t=1}^N (a_t - \bar{x})(b_t - \bar{y}) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} (x_i - \bar{x})(y_j - \bar{y})$$

ist ein Maß für die Korrelation bei einem **affin linearen** Zusammenhang zwischen X und Y .

(2) Zur **Bewertung der Korrelation** nutzt man den (dimensionslosen) **Korrelationskoeffizienten**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

wobei σ_X bzw. σ_Y die Standardabweichung des Merkmals X bzw. Y ist. $\rho_{X,Y}$ nimmt nur Werte im Intervall $[-1, 1]$ an, und man **bewertet die Korrelation wie folgt**:

$-1 \leq \rho_{XY} < -0,6$	\implies	<i>starke negative Korrelation</i>
$-0,6 \leq \rho_{XY} < 0$	\implies	<i>schwache negative Korrelation</i>
$\rho_{XY} = 0$	\implies	<i>keine Korrelation</i>
$0 < \rho_{XY} \leq 0,6$	\implies	<i>schwache positive Korrelation</i>
$0,6 < \rho_{XY} \leq 1$	\implies	<i>starke positive Korrelation</i>

Betrachten wir zwei Beispiele.

Beispiel 5.17. (Korrelation zwischen Stunden Lauftraining und Laufzeit)

Es wurden die Trainingstunden pro Woche in Stunden (Merkmal X) und die 10 km-Laufzeiten in Minuten (Merkmal Y) von 8 männlichen Schülern erhoben. Die Daten finden sich in der Tabelle in Beispiel 5.15. Zur Berechnung der arithmetischen Mittel \bar{x}, \bar{y} , der Standardabweichungen σ_X, σ_Y und der Korrelation σ_{XY} legen wir eine geeignete Berechnungstabelle an:

t	a_t	$(a_t - \bar{x})^2$	b_t	$(b_t - \bar{y})^2$	$(a_t - \bar{x})(b_t - \bar{y})$
1	0	15,015625	70	315,0625	-68,78125
2	3	0,765625	55	7,5625	-2,40625
3	6	4,515625	40	150,0625	-26,03125
4	4	0,015625	45	52,5625	-0,90625
5	1	8,265625	65	162,5625	-36,65625
6	2	3,515625	60	60,0625	-14,53125
7	10	37,515625	38	203,0625	-87,28125
8	5	1,265625	45	52,5625	-8,15625
Σ	31	70,875	418	1003,5	-244,75
$\frac{1}{8} \Sigma$	$\bar{x} = 3,875$	$\sigma_X^2 = 8,859375$	$\bar{y} = 52,25$	$\sigma_Y^2 = 125,4375$	$\sigma_{XY} = -30,59375$
$\sqrt{\frac{1}{8} \Sigma}$		$\sigma_X \approx 2,9765$		$\sigma_Y \approx 11,1999$	

Wir lesen ab:

Arithmetisches Mittel \bar{x} und *Standardabweichung* σ_X von Merkmal X (Trainingsstunden pro Woche in h):

$$\bar{x} = \frac{1}{8} \sum_{t=1}^8 a_t = 3,875 \text{ h} \quad \text{und} \quad \sigma_X = \sqrt{\frac{1}{8} \sum_{t=1}^8 (a_t - \bar{x})^2} \approx 2,9765 \text{ h.}$$

Arithmetisches Mittel \bar{y} und *Standardabweichung* σ_Y von Merkmal Y (Laufzeit in min auf 10 km):

$$\bar{y} = \frac{1}{8} \sum_{t=1}^8 b_t = 52,25 \text{ min} \quad \text{und} \quad \sigma_Y = \sqrt{\frac{1}{8} \sum_{t=1}^8 (b_t - \bar{y})^2} \approx 11,1999 \text{ min.}$$

Kovarianz von X und Y :
$$\sigma_{XY} = \frac{1}{8} \sum_{t=1}^8 (a_t - \bar{x})(b_t - \bar{y}) = -30,59375 \text{ h} \cdot \text{min}$$

Korrelationskoeffizient von X und Y :
$$\rho_{XY} = \frac{\sigma_{X,Y}}{\sigma_X \cdot \sigma_Y} \approx \frac{-30,59375 \text{ h} \cdot \text{min}}{2,9765 \text{ h} \cdot 11,1999 \text{ min}} \approx -0,918$$

Interpretation: Da $-1 \leq \rho_{XY} = -0,918 < -0,6$ ist, liegt eine *starke negative Korrelation* vor. Die Zahl der Stunden Lauftraining pro Woche ist *stark negativ korreliert* mit der über 10 km gelaufenen Zeit.

Im zweiten Beispiel haben wir eine zweidimensionale Häufigkeitsverteilung vorgegeben. Hier muss man mit den Randverteilungen arbeiten, um die arithmetischen Mittel und die Standardabweichungen der einzelnen Merkmale effizient zu berechnen.

Beispiel 5.18. (Kovarianz und Korrelationskoeffizient – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Sei (X, Y) das zweidimensionale Merkmal aus Beispielen 5.2, 5.3 and 5.5, wobei $X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“, $Y =$ „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“. In Beispiel 5.3 hatten wir aus der vorgegebenen zweidimensionalen absoluten Häufigkeitsverteilung die folgende relative Häufigkeitsverteilung mit den zugehörigen Randverteilungen berechnet:

	j	1	2	3	
i	$x_i \setminus y_j$	0	1	2	$h_{i.}$
1	0	20	5	0	25
2	1	15	20	15	50
3	2	0	5	20	25
	$h_{.j}$	35	30	35	100

Zur Berechnung der arithmetischen Mittel und Standardabweichungen der einzelnen Merkmale nutzen wir die Randverteilungen:

Für die *arithmetischen Mittel* finden wir

$$\begin{aligned}\bar{x} &= \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} x_i = \frac{1}{N} \sum_{i=1}^k \underbrace{\left(\sum_{j=1}^{\ell} h_{i,j} \right)}_{=h_{i,\cdot}} x_i = \frac{1}{N} \sum_{i=1}^k h_{i,\cdot} x_i = \sum_{i=1}^k f_{i,\cdot} x_i, \\ \bar{y} &= \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} y_j = \frac{1}{N} \sum_{j=1}^{\ell} \underbrace{\left(\sum_{i=1}^k h_{i,j} \right)}_{=h_{\cdot,j}} y_j = \frac{1}{N} \sum_{j=1}^{\ell} h_{\cdot,j} y_j = \sum_{j=1}^{\ell} f_{\cdot,j} y_j,\end{aligned}$$

also

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k h_{i,\cdot} x_i = \sum_{i=1}^k f_{i,\cdot} x_i \quad \text{und} \quad \bar{y} = \frac{1}{N} \sum_{j=1}^{\ell} h_{\cdot,j} y_j = \sum_{j=1}^{\ell} f_{\cdot,j} y_j,$$

wobei wir für die zweite Formel genutzt haben, dass

$$\frac{1}{N} h_{i,\cdot} = f_{i,\cdot} \quad \text{da} \quad \frac{1}{N} h_{i,\cdot} = \frac{1}{N} \sum_{j=1}^{\ell} h_{i,j} = \sum_{j=1}^{\ell} \underbrace{\frac{h_{i,j}}{N}}_{=f_{i,j}} = \sum_{j=1}^{\ell} f_{i,j} = f_{i,\cdot}, \quad (5.1)$$

$$\frac{1}{N} h_{\cdot,j} = f_{\cdot,j} \quad \text{da} \quad \frac{1}{N} h_{\cdot,j} = \frac{1}{N} \sum_{i=1}^k h_{i,j} = \sum_{i=1}^k \underbrace{\frac{h_{i,j}}{N}}_{=f_{i,j}} = \sum_{i=1}^k f_{i,j} = f_{\cdot,j}. \quad (5.2)$$

Also berechnen wir

$$\begin{aligned}\bar{x} &= \frac{1}{100} \cdot (h_{1,\cdot} x_1 + h_{2,\cdot} x_2 + h_{3,\cdot} x_3) = \frac{1}{100} \cdot (25 \cdot 0 + 50 \cdot 1 + 25 \cdot 2) = \frac{100}{100} = 1, \\ \bar{y} &= \frac{1}{100} \cdot (h_{\cdot,1} y_1 + h_{\cdot,2} y_2 + h_{\cdot,3} y_3) = \frac{1}{100} \cdot (35 \cdot 0 + 30 \cdot 1 + 35 \cdot 2) = \frac{100}{100} = 1.\end{aligned}$$

Für die *Standardabweichungen* finden wir

$$\begin{aligned}\sigma_X &= \sqrt{\frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} (x_i - \bar{x})^2} = \sqrt{\frac{1}{N} \sum_{i=1}^k \underbrace{\left(\sum_{j=1}^{\ell} h_{i,j} \right)}_{=h_{i,\cdot}} (x_i - \bar{x})^2} = \sqrt{\frac{1}{N} \sum_{i=1}^k h_{i,\cdot} (x_i - \bar{x})^2}, \\ \sigma_Y &= \sqrt{\frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} (y_j - \bar{y})^2} = \sqrt{\frac{1}{N} \sum_{j=1}^{\ell} \underbrace{\left(\sum_{i=1}^k h_{i,j} \right)}_{=h_{\cdot,j}} (y_j - \bar{y})^2} = \sqrt{\frac{1}{N} \sum_{j=1}^{\ell} h_{\cdot,j} (y_j - \bar{y})^2},\end{aligned}$$

also

$$\sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^k h_{i,\cdot} (x_i - \bar{x})^2} = \sqrt{\sum_{i=1}^k f_{i,\cdot} (x_i - \bar{x})^2},$$

$$\sigma_Y = \sqrt{\frac{1}{N} \sum_{j=1}^{\ell} h_{\cdot,j} (y_j - \bar{y})^2} = \sqrt{\sum_{j=1}^{\ell} f_{\cdot,j} (y_j - \bar{y})^2},$$

wobei wir wieder (5.1) und (5.2) genutzt haben. Also berechnen wir

$$\sigma_X = \sqrt{\frac{1}{100} \cdot (25 \cdot (0-1)^2 + 50 \cdot (1-1)^2 + 25 \cdot (2-1)^2)} = \sqrt{\frac{50}{100}} = \sqrt{0,5} \approx 0,707,$$

$$\sigma_Y = \sqrt{\frac{1}{100} \cdot (35 \cdot (0-1)^2 + 30 \cdot (1-1)^2 + 35 \cdot (2-1)^2)} = \sqrt{\frac{70}{100}} = \sqrt{0,7} \approx 0,837.$$

Nun können wir die *Kovarianz*

$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} (x_i - \bar{x}) (y_j - \bar{y})$$

berechnen:

$$\begin{aligned} \sigma_{XY} &= \frac{1}{100} \cdot (20 \cdot (0-1) \cdot (0-1) + 5 \cdot (0-1) \cdot (1-1) + 0 \cdot (0-1) \cdot (2-1) \\ &\quad + 15 \cdot (1-1) \cdot (0-1) + 20 \cdot (1-1) \cdot (1-1) + 15 \cdot (1-1) \cdot (2-1) \\ &\quad + 0 \cdot (2-1) \cdot (0-1) + 5 \cdot (2-1) \cdot (1-1) + 20 \cdot (2-1) \cdot (2-1)) \\ &= \frac{40}{100} = 0,4. \end{aligned}$$

Somit erhalten wir für den *Korrelationskoeffizienten*

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0,4}{\sqrt{0,5} \cdot \sqrt{0,7}} = \frac{0,4}{\sqrt{0,35}} \approx 0,676.$$

Interpretation: Da $0,6 < \rho_{XY} = 0,676 \leq 1$ ist, sind die Anzahl der für die Analysis 1 gekauften Lehrbücher (Merkmal X) und die Anzahl der für die Lineare Algebra 1 gekauften Lehrbücher (Merkmal Y) *stark positiv korreliert*.

Zuletzt halten wir zwei wichtige Aussagen über die Korrelation und Kovarianz fest.

Lemma 5.19. (unabhängig verteilt \implies unkorreliert)

*Sind zwei metrische skalierte Merkmale X und Y auf der gleichen Grundgesamtheit **unabhängig verteilt**, so sind sie **unkorreliert**, und es gilt $\sigma_{XY} = 0$ und $\rho_{XY} = 0$.*

Achtung: Aus $\sigma_{XY} = 0$ folgt **nicht**, dass die Merkmale X und Y unabhängig verteilt sind!

Beweis von Lemma 5.19: Dieses Lemma wird auf einem Übungszettel bewiesen. \square

Lemma 5.20. (alternative Formel zur Berechnung der Kovarianz)

Seien die Voraussetzungen und die Bezeichnungen wie in Definition 5.16. Die **Kovarianz** σ_{XY} kann auch mit der folgenden Formel berechnet werden:

$$\sigma_{XY} = \left(\frac{1}{N} \sum_{t=1}^N a_t b_t \right) - \bar{x} \bar{y} = \left(\frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{\ell} h_{i,j} x_i y_j \right) - \bar{x} \bar{y}$$

Beweis von Lemma 5.20: Der Beweis erfolgt ausgehend von der ursprünglichen Formel für σ_{XY} durch Nachrechnen, wobei man geeignet vereinfachen und die Formeln für die arithmetischen Mittel

$$\bar{x} = \frac{1}{N} \sum_{i=1}^k h_{i,\cdot} x_i \quad \text{und} \quad \bar{y} = \frac{1}{N} \sum_{j=1}^{\ell} h_{\cdot,j} y_j$$

ausnutzen muss. Wir beweisen das Lemma auf einem Übungszettel. □

Lineare Regression

In diesem Kapitel lernen wir die Methode der kleinsten Quadrate kennen und werden mit dieser eine „optimale“ affine lineare Regressionsfunktion berechnen.

6.1 Problemstellung der Regression

In diesem Kapitel betrachten wir die folgende **Problemstellung**: Für zwei (korrelierte) metrisch skalierte Merkmale X und Y (auf derselben Grundgesamtheit) werde angenommen, dass X unabhängig ist und dass Y **von X abhängig ist**. Der Zusammenhang zwischen X und Y wird durch die **Regressionsfunktion** $y = f(x)$, wobei $x =$ Merkmalswerte von X , $y =$ Merkmalswerte von Y sind, beschrieben. Wir nennen dabei X die **unabhängige/erklärende Variable** und Y die **abhängige/erklärte Variable**.

Wie sieht die Regressionsfunktion aus? Das hängt vom Problem, oder genauer den Einzelwerten der beiden Merkmale ab. Zeichnet man das Streudiagramm der Einzelwerte $(a_t, b_t), t = 1, \dots, N$, des zweidimensionalen Merkmals (X, Y) in der Grundgesamtheit, so gewinnt man einen ersten Eindruck und kann eine Vermutung über die Form der Regressionsfunktion aufstellen.

Wir betrachten in dieser Vorlesung **nur** den Fall **linearer Regression**, d.h. wenn

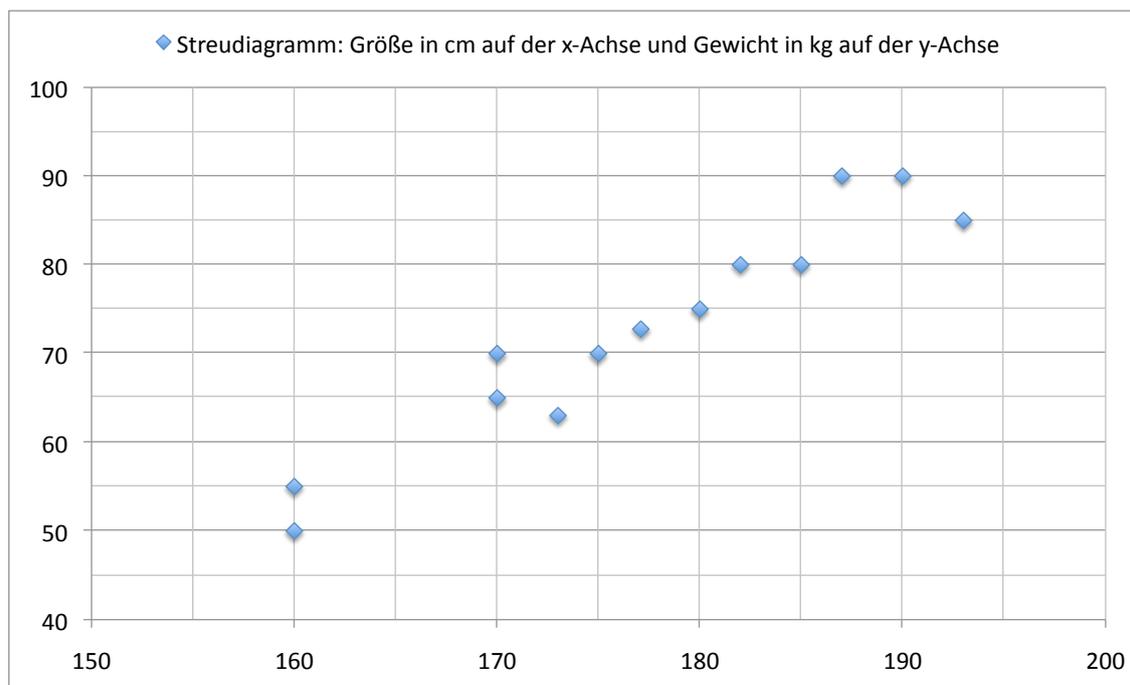
$$y = f(x) = \beta x + \alpha,$$

also wenn die Regressionsfunktion eine **affin lineare Funktion** ist. Wir werden lernen, wie man die Steigung β und den Achsenabschnitt α „optimal“ bestimmt. Anschließend werden wir lernen, wie man die Qualität der Annäherung (oder Approximation) der Daten durch die Regressionsfunktion bewertet. Bei der Berechnung der affin linearen Regressionsfunktion und der Bewertung deren Approximationsgüte spielen die bereits kennengelernten Begriffe der Standardabweichungen σ_X und σ_Y der beiden Merkmale X und Y und der Kovarianz σ_{XY} des zweidimensionalen Merkmals (X, Y) eine wichtige Rolle.

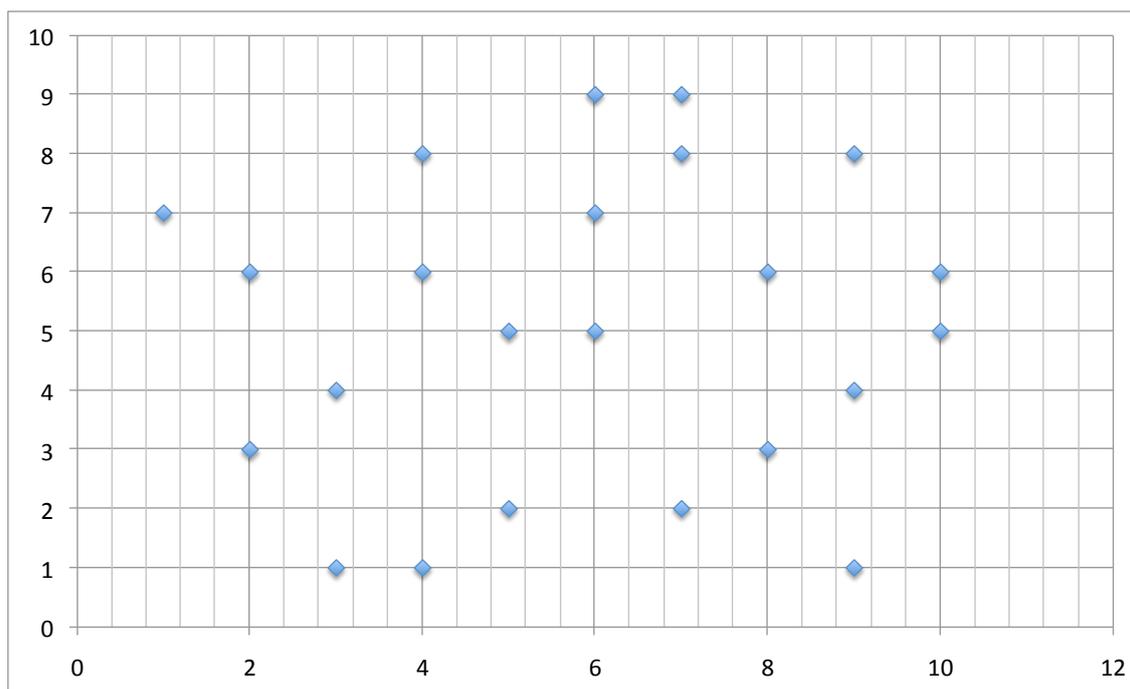
Betrachten wir als Abschluss der Einführung drei Beispiele von Streudiagrammen.

Beispiel 6.1. (Streudiagramme zweidimensionaler Merkmale)

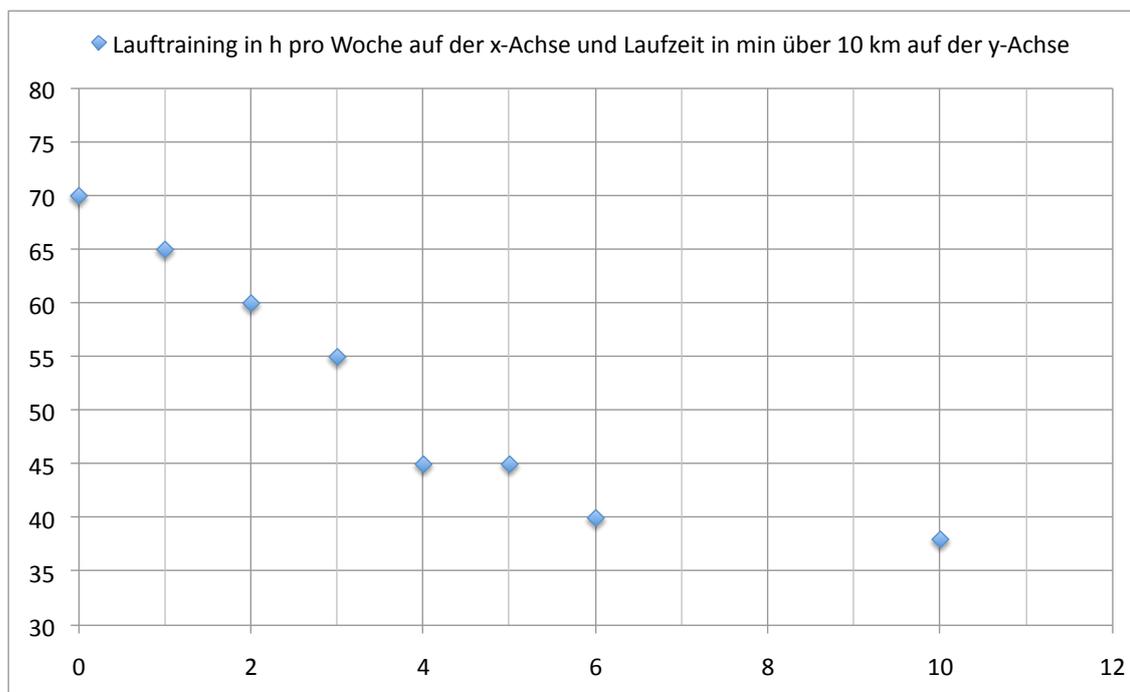
- (a) Für den bereits in Beispiel 5.14 betrachteten Datensatz des zweidimensionalen Merkmals (X, Y) mit $X =$ „Körpergröße“ und $Y =$ „Gewicht“ sieht man im Streudiagramm einen linearen Trend mit positiver Korrelation. Hier ist inhaltlich klar, dass das Merkmal $Y =$ „Gewicht“ vom Merkmal $X =$ „Körpergröße“ abhängen muss.



- (b) In dem nachfolgenden Streudiagramm eines zweidimensionalen Merkmals (X, Y) sieht man keinen (affin linearen) funktionalen Zusammenhang.



- (c) Für das in Beispielen 5.15 und 5.17 betrachtete zweidimensionale Merkmal (X, Y) mit $X =$ „Stunden Lauftraining pro Woche“ und $Y =$ „Laufzeit (in min) über 10 km“ sieht man in dem Streudiagramm einen linearen Trend mit negativer Korrelation. Hier ist inhaltlich klar, dass das Merkmal $Y =$ „Laufzeit (in min) über 10 km“ von dem Merkmal $X =$ „Stunden Lauftraining pro Woche“ abhängen muss.



Als Letztes weisen wir noch darauf hin, dass der hier betrachtete Fall der linearen Regression natürlich eine **Einschränkung** darstellt, denn die funktionale Abhängigkeit des Merkmals Y von dem Merkmal X kann ganz anders aussehen, z.B. könnten wir einen quadratischen Zusammenhang $y = f(x) = x^2$ haben. Hier wäre eine globale Annäherung durch eine affin lineare Regressionsfunktion völlig ungeeignet. **Lokal** kann aber auch hier eine Annäherung durch eine affin lineare Regressionsfunktion nützlich sein, denn wir wissen, dass lokal die Tangente einer differenzierbaren Funktion eine brauchbare Näherung der Funktion darstellt.

6.2 Methode der kleinsten Quadrate

Wir starten damit, dass wir für das Problem der linearen Regression einige Begriffe und Notation einführen

Problemstellung 6.2. (lineare Regression)

Wir betrachten in diesem Kapitel immer ein zweidimensionales metrisch skaliertes Merkmal (X, Y) mit den Einzelwerten (a_t, b_t) , $t = 1, 2, \dots, N$, auf einer Grundgesamtheit von N Elementen. Wir nehmen an, dass X das **unabhängige Merkmal** und Y das (von

X) **abhängige Merkmal** ist, und dass der funktionale Zusammenhang durch eine **affin lineare Funktion**

$$y = f(x) = \beta x + \alpha$$

beschrieben werden kann. Dabei sind die Parameter α und β noch geeignet zu bestimmen. Der funktionale Zusammenhang $y = \beta x + \alpha$ muss nicht für alle möglichen Merkmalswerte von X (angenähert) gelten sondern nur in einem Intervall für X , das die Einzelwerte von X enthält.

Unabhängig von der konkreten Wahl von β und α werden die Einzelwerte (a_t, b_t) in der Regel um die Regressionsfunktion **streuen**, d.h.

$$b_t \approx f(a_t) = \beta a_t + \alpha, \quad t = 1, 2, \dots, N.$$

Wir führen daher folgende Bezeichnungen ein:

- (1) $\hat{b}_t = f(a_t) = \beta a_t + \alpha$ ist der **geschätzte Wert** für b_t (für jedes $t = 1, 2, \dots, N$).
- (2) $e_t = b_t - \hat{b}_t = b_t - f(a_t)$ ist das **Residuum** von b_t (für jedes $t = 1, 2, \dots, N$), welches die Abweichung des geschätzten Y -Wertes \hat{b}_t von dem Einzelwert b_t misst.

Es gilt also

$$b_t = \underbrace{\beta a_t + \alpha}_{=\hat{b}_t} + e_t, \quad t = 1, 2, \dots, N.$$

Es ist üblich, durch ein statistisches Verfahren (also z.B. durch die Wahl einer Regressionsfunktion) **geschätzte Werte** mit einem $\hat{}$ über dem Buchstaben zu kennzeichnen. \hat{b}_t ist also eine Schätzung (durch die affin lineare Regressionsfunktion) für den Einzelwert b_t .

Wir wollen in diesem Kapitel die folgenden **Fragen** beantworten:

- (1) Wie soll man die **Steigung** β und den **Achsenabschnitt** α der **affin linearen** Regressionsfunktion $y = \beta x + \alpha$ „**optimal**“ wählen?
- (2) Wie **beurteilt** man die **Güte** der affin linearen Regressionsfunktion?

(Diese Frage ist wichtig, damit wir beurteilen können, ob unsere affin lineare Regressionsfunktion ein verlässliches Modell für den funktionalen Zusammenhang ist.)

In diesem Teilkapitel beschäftigen wir uns mit der ersten Frage. Die zweite beantworten wir in Teilkapitel 6.3.

Statistisches Verfahren 6.3. (Methode der kleinsten Quadrate)

Sei (X, Y) ein zweidimensionales metrisch skaliertes Merkmal mit den Einzelwerten (a_t, b_t) , $t = 1, 2, \dots, N$, auf einer Grundgesamtheit von N Elementen. Wir nehmen an, dass X das unabhängige Merkmal und Y das (von X) abhängige Merkmal ist und dass der funktionale Zusammenhang durch eine affin lineare Funktion $y = f(x) = \beta x + \alpha$ beschrieben werden kann. Bei der **Methode der kleinsten Quadrate** wählt man β und α so, dass die **Summe der quadrierten Residuen**

$$\text{SAQ}(\alpha, \beta) = \sum_{t=1}^N e_t^2 = \sum_{t=1}^N (b_t - \beta a_t - \alpha)^2 = \sum_{t=1}^N (b_t - \hat{b}_t)^2$$

minimiert wird. (SAQ steht für „Summe der Abweichungsquadrate“.)

Die Parameter α und β der affin linearen Regressionsfunktion werden also als Lösungen eines Optimierungsproblems bestimmt. Wir wollen dieses Optimierungsproblem nun lösen.

Bestimmung der Werte $\hat{\alpha}$ und $\hat{\beta}$ für α bzw. β , für die $\text{SAQ}(\alpha, \beta)$ minimiert wird:

- **notwendige Bedingung** für lokales Minimum: Die (ersten) partiellen Ableitungen sind Null.

Für die partielle Ableitung nach α finden wir

$$\begin{aligned} \frac{\partial \text{SAQ}}{\partial \alpha}(\alpha, \beta) &= \sum_{t=1}^N 2(b_t - \beta a_t - \alpha) \cdot (-1) \\ &= -2 \cdot \sum_{t=1}^N b_t + 2\beta \sum_{t=1}^N a_t + \underbrace{2 \sum_{t=1}^N \alpha}_{= 2N\alpha} \\ &= -2N \cdot \underbrace{\left(\frac{1}{N} \sum_{t=1}^N b_t\right)}_{=\bar{y}} + 2N\beta \underbrace{\left(\frac{1}{N} \sum_{t=1}^N a_t\right)}_{=\bar{x}} + 2N\alpha \\ &= -2N \cdot (\bar{y} - \beta \bar{x} - \alpha) \stackrel{!}{=} 0. \end{aligned}$$

Wir erhalten also die Bedingung

$$\bar{y} - \beta \bar{x} - \alpha = 0 \quad \iff \quad \bar{y} = \beta \bar{x} + \alpha \quad \iff \quad \alpha = \bar{y} - \beta \bar{x}. \quad (6.1)$$

Für die partielle Ableitung nach β finden wir

$$\begin{aligned} \frac{\partial \text{SAQ}}{\partial \beta}(\alpha, \beta) &= \sum_{t=1}^N 2(b_t - \beta a_t - \alpha) \cdot (-a_t) \\ &= -2 \sum_{t=1}^N a_t b_t + 2\beta \sum_{t=1}^N a_t^2 + 2\alpha \sum_{t=1}^N a_t \\ &= -2N \cdot \left(\frac{1}{N} \sum_{t=1}^N a_t b_t\right) + 2N\beta \left(\frac{1}{N} \sum_{t=1}^N a_t^2\right) + 2N\alpha \underbrace{\left(\frac{1}{N} \sum_{t=1}^N a_t\right)}_{=\bar{x}} \\ &= -2N \cdot \left(\frac{1}{N} \sum_{t=1}^N a_t b_t - \beta \frac{1}{N} \sum_{t=1}^N a_t^2 - \alpha \bar{x}\right) \stackrel{!}{=} 0. \end{aligned}$$

Wir erhalten also die Bedingung

$$\frac{1}{N} \sum_{t=1}^N a_t b_t - \beta \frac{1}{N} \sum_{t=1}^N a_t^2 - \alpha \bar{x} = 0, \quad (6.2)$$

und Ersetzen in (6.2) von α mittels (6.1) liefert:

$$\begin{aligned} & \frac{1}{N} \sum_{t=1}^N a_t b_t - \beta \frac{1}{N} \sum_{t=1}^N a_t^2 - (\bar{y} - \beta \bar{x}) \cdot \bar{x} = 0 \\ \Leftrightarrow & \underbrace{\frac{1}{N} \sum_{t=1}^N a_t b_t - \bar{x} \bar{y}}_{=\sigma_{XY}} - \beta \underbrace{\left(\frac{1}{N} \sum_{t=1}^N a_t^2 - \bar{x}^2 \right)}_{=\sigma_X^2} = 0 \\ \Leftrightarrow & \sigma_{XY} - \beta \sigma_X^2 = 0 \quad \Leftrightarrow \quad \beta = \frac{\sigma_{XY}}{\sigma_X^2} \end{aligned} \quad (6.3)$$

Aus (6.1) und (6.3) erhalten wir die folgenden beiden Gleichungen, welche die Parameter $\alpha = \hat{\alpha}$ und $\beta = \hat{\beta}$ eindeutig festlegen:

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{und} \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad (6.4)$$

- **hinreichende Zusatzbedingungen** für ein lokales Minimum:

(1) Für die zweite Ableitung nach α muss für $\hat{\alpha}$ und $\hat{\beta}$ aus (6.4) gelten $\frac{\partial^2 \text{SAQ}}{\partial \alpha^2}(\hat{\alpha}, \hat{\beta}) > 0$.

Wir finden

$$\frac{\partial^2 \text{SAQ}}{\partial \alpha^2}(\alpha, \beta) = 2N > 0 \quad \text{für alle } \alpha, \beta \in \mathbb{R}.$$

(2) Weiterhin muss für $\hat{\alpha}$ und $\hat{\beta}$ aus (6.4) gelten

$$\frac{\partial^2 \text{SAQ}}{\partial \alpha^2}(\hat{\alpha}, \hat{\beta}) \cdot \frac{\partial^2 \text{SAQ}}{\partial \beta^2}(\hat{\alpha}, \hat{\beta}) - \left(\frac{\partial^2 \text{SAQ}}{\partial \beta \partial \alpha}(\hat{\alpha}, \hat{\beta}) \right)^2 > 0.$$

Wir berechnen die gemischte zweite Ableitung

$$\frac{\partial^2 \text{SAQ}}{\partial \beta \partial \alpha}(\alpha, \beta) = 2N \bar{x}.$$

Somit erhalten wir

$$\begin{aligned} & \frac{\partial^2 \text{SAQ}}{\partial \alpha^2}(\alpha, \beta) \cdot \frac{\partial^2 \text{SAQ}}{\partial \beta^2}(\alpha, \beta) - \left(\frac{\partial^2 \text{SAQ}}{\partial \beta \partial \alpha}(\alpha, \beta) \right)^2 = 2N \cdot 2 \sum_{t=1}^N a_t^2 - (2N \bar{x})^2 \\ & = 4N^2 \underbrace{\left(\frac{1}{N} \sum_{t=1}^N a_t^2 - \bar{x}^2 \right)}_{=\sigma_X^2} = 4N^2 \sigma_X^2 > 0 \quad \text{für alle } \alpha, \beta \in \mathbb{R}, \text{ wenn } \sigma_X \neq 0. \end{aligned}$$

Die Bedingungen (1) und (2) gelten für alle α, β und damit insbesondere für $\alpha = \hat{\alpha}$ und $\beta = \hat{\beta}$, die (6.4) erfüllen.

- **Fazit:** Für die Parameter

$$\beta = \hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{und} \quad \alpha = \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x}$$

nimmt die Summe der Abweichungsquadrate $\text{SAQ}(\alpha, \beta)$ ein lokales (und auch globales) Minimum an!

Wir halten das Ergebnis unserer Untersuchung in einem Satz fest.

Satz 6.4. (Methode der kleinsten Quadrate)

Sei (X, Y) ein zweidimensionales metrisch skaliertes Merkmal mit den Einzelwerten (a_t, b_t) , $t = 1, 2, \dots, N$, auf einer Grundgesamtheit von N Elementen, wobei $\sigma_X^2 > 0$ ist. Wir nehmen an, dass X das unabhängige Merkmal und Y das (von X) abhängige Merkmal ist. Die mit der **Methode der kleinsten Quadrate** geschätzte **affin lineare Regressionsfunktion** lautet

$$\hat{y} = \hat{\beta}x + \hat{\alpha}$$

(bzw. $\hat{b}_t = \hat{\beta}a_t + \hat{\alpha}$, $t = 1, 2, \dots, N$, für die geschätzten Einzelwerte von Y) mit den **Regressionskoeffizienten**

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{und} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Das Symbol $\hat{}$ über α und β weist darauf hin, dass die Werte der Regressionskoeffizienten aus den vorgegebenen Daten **geschätzt** wurden.

Berechnen wir die affin lineare Regressionsfunktion für ein Beispiel.

Beispiel 6.5. (Lineare Regression – Abhängigkeit der Laufzeit über 10 km vom Lauftraining)

Es wurden die Trainingsstunden pro Woche in Stunden (Merkmal X) und die 10 km-Laufzeiten in Minuten (Merkmal Y) von 8 männlichen Schülern erhoben.

Nummer	t	1	2	3	4	5	6	7	8
$X = \text{„Lauftraining in h pro Woche“}$	a_t	0	3	6	4	1	2	10	5
$Y = \text{„Laufzeit auf 10 km in min“}$	b_t	70	55	40	45	65	60	38	45

In Beispiel 5.17 hatten wir bereits berechnet:

$$\sigma_{X,Y} = -30,59375 \text{ h} \cdot \text{min}, \quad \sigma_X^2 = 8,859375 \text{ h}^2, \quad \bar{x} = 3,875 \text{ h}, \quad \bar{y} = 52,25 \text{ min}.$$

Damit berechnen wir die Regressionskoeffizienten:

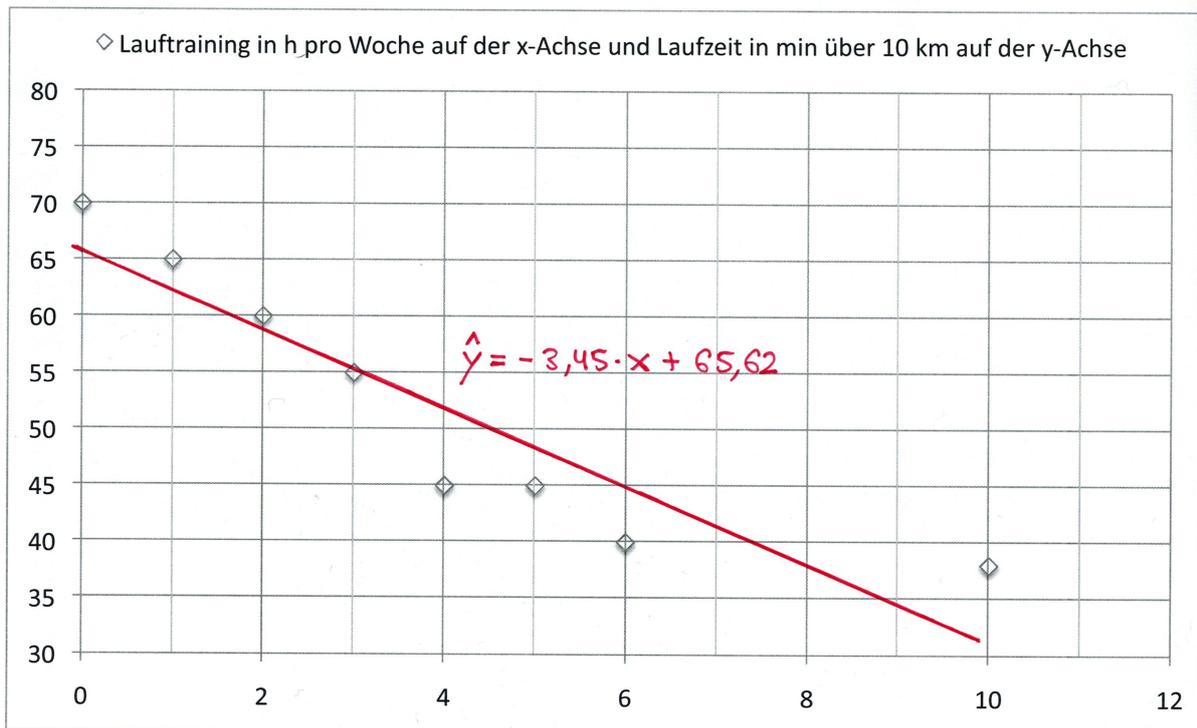
$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{-30,59375}{8,859375} \cdot \frac{\text{h} \cdot \text{min}}{\text{h}^2} \approx -3,45 \frac{\text{min}}{\text{h}},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \approx 52,25 \text{ min} + 3,45 \frac{\text{min}}{\text{h}} \cdot 3,875 \text{ h} \approx 65,62 \text{ min}.$$

Somit lautet die affin lineare Regressionsfunktion

$$\hat{y} = -3,45 \cdot x + 65,62. \quad (\text{Einheit: min}).$$

Wir haben diese im nachfolgenden Streudiagramm eingezeichnet.



Bevor wir die mit der Methode der kleinsten Quadrate berechnete affin lineare Regressionsfunktion weiter untersuchen, wollen wir uns noch eine **geometrische Anschauung** des Verfahrens verschaffen:

Die Regressionskoeffizienten $\hat{\alpha}$ und $\hat{\beta}$ werden durch die Methode der kleinsten Quadrate gerade so berechnet, dass für $\alpha = \hat{\alpha}$ und $\beta = \hat{\beta}$ die Summe der Abweichungsquadrate

$$\text{SAQ}(\alpha, \beta) = \sum_{t=1}^N e_t^2 = \sum_{t=1}^N (b_t - \beta a_t - \alpha)^2 = \sum_{t=1}^N \underbrace{(b_t - \hat{b}_t)}_{=e_t}^2 = \sum_{t=1}^N e_t^2$$

minimal wird (vgl. Statistisches Verfahren 6.3). Die Regressionsgerade wird also anschaulich gerade so in das Streudiagramm gelegt, dass die Summe der Residuenquadrate $e_t^2 = (b_t - (\beta a_t + \alpha))^2$, $t = 1, 2, \dots, N$, minimal wird. In Abbildung 6.1 ist dieses für den Datensatz aus Beispiel 6.5 veranschaulicht: Die Residuenquadrate $e_t^2 = (b_t - (\beta a_t + \alpha))^2$ sind die Flächen der in Abbildung 6.1 eingezeichneten Quadrate. Da hier die optimale Regressionsgerade eingezeichnet ist, ist die Summe der Flächen $e_t^2 = (b_t - (\beta a_t + \alpha))^2$, $t = 1, 2, \dots, N$, dieser Quadrate minimal. Verändert man die Gerade, so wird die Summe der Flächen der ebenfalls veränderten Quadrate größer.

Bemerkung 6.6. (Interpretation der Steigung der Regressionsgeraden)

Wegen

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} \quad \text{und} \quad \sigma_X^2 > 0,$$

sieht man

$$\hat{\beta} > 0 \quad \iff \quad \sigma_{XY} > 0$$

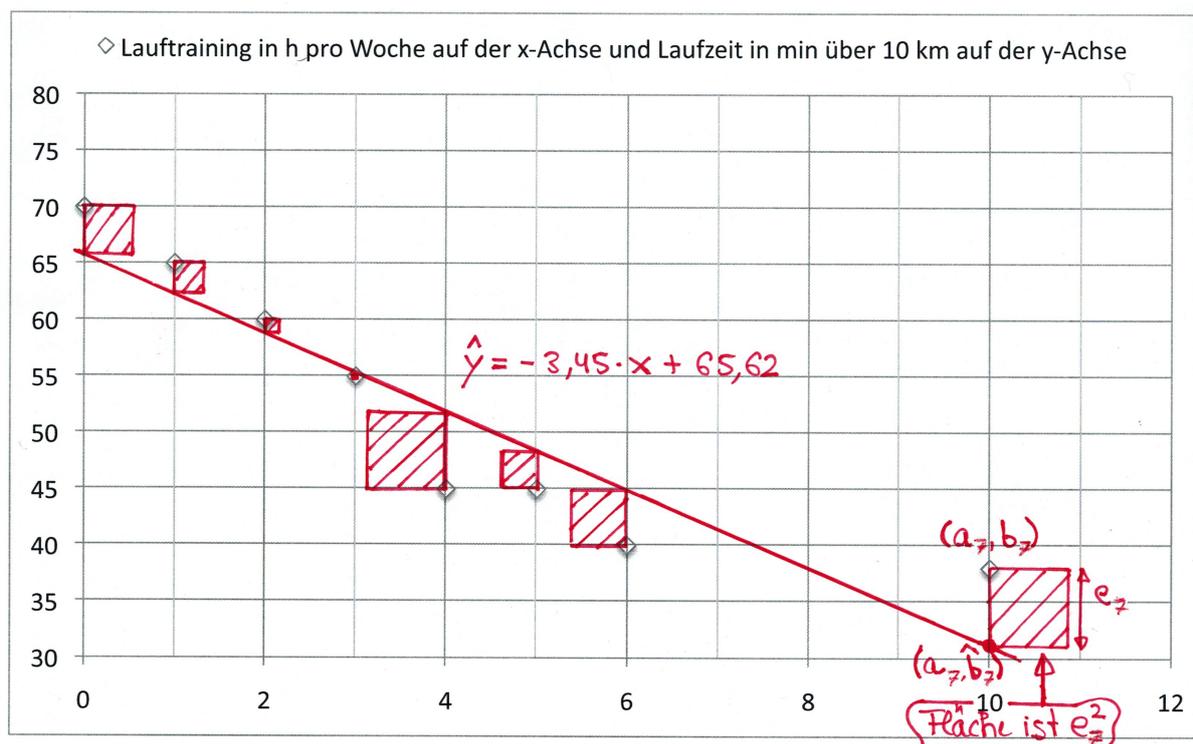


Abbildung 6.1: Veranschaulichung der Methode der kleinsten Quadrate: Die Summe der Flächen $e_t^2 = (b_t - (\beta a_t + \alpha))^2$, $t = 1, 2, \dots, N$, der eingezeichneten Abweichungsquadrate ist für die optimale Regressionsgerade minimal.

$$\hat{\beta} = 0 \quad \iff \quad \sigma_{XY} = 0$$

$$\hat{\beta} < 0 \quad \iff \quad \sigma_{XY} < 0$$

d.h. es gilt:

streng monoton wachsende Regressionsgerade \iff X und Y sind positiv korreliert

horizontale Regressionsgerade \iff X und Y sind unkorreliert

streng monoton fallende Regressionsgerade \iff X und Y sind negativ korreliert

Dieses entspricht genau unseren Erwartungen nach unserem Verständnis der Korrelation zweier Merkmale.

Wir halten nun wichtige Eigenschaften der affin linearen Regressionsfunktion fest.

Satz 6.7. (Eigenschaften der affin linearen Regressionsfunktion)

Die Voraussetzungen und Notation seien wie in Satz 6.4.

(1) Die **Summe der Residuen** $e_t = b_t - \hat{b}_t = b_t - (\hat{\beta} a_t + \hat{\alpha})$ ist Null, d.h.

$$\sum_{t=1}^N e_t = \sum_{t=1}^N \underbrace{(b_t - \hat{b}_t)}_{= e_t} = 0.$$

(2) Das **arithmetische Mittel** der (beobachteten) Einzelwerte b_t ,

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N b_t,$$

und das **arithmetische Mittel** der geschätzten Einzelwerte $\hat{b}_t = \hat{\beta} a_t + \hat{\alpha}$,

$$\bar{\hat{y}} = \frac{1}{N} \sum_{t=1}^N \hat{b}_t$$

sind gleich, also

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N b_t = \frac{1}{N} \sum_{t=1}^N \hat{b}_t = \bar{\hat{y}}.$$

(3) Die Regressionsgerade läuft durch den **Schwerpunkt**

$$(\bar{x}, \bar{y}) = \left(\frac{1}{N} \sum_{t=1}^N a_t, \frac{1}{N} \sum_{t=1}^N b_t \right)$$

der Punktwolke des Streudiagramms.

(4) Die Summe $\sum_{t=1}^N a_t e_t$ ist Null, also

$$\sum_{t=1}^N a_t e_t = a_1 e_1 + a_2 e_2 + \dots + a_N e_N = 0.$$

Bemerkung 6.8. (Erläuterung von Satz 6.7 (4))

Satz 6.7 (4) ist eine „technische“ Eigenschaft, die wir im **Beweis der Streuungszerlegung** brauchen werden. Satz 6.7 (4) besagt, dass der Vektor (a_1, a_2, \dots, a_N) der Einzelwerte für X und der Vektor der Residuen, (e_1, e_2, \dots, e_n) , **orthogonal** sind (d.h. einen 90° -Winkel zueinander haben).

Wir beweisen Satz 6.7.

Beweis von Satz 6.7 (3): Die Formel zur Berechnung von $\hat{\alpha}$ ist

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad \Longleftrightarrow \quad \bar{y} = \hat{\beta} \bar{x} + \hat{\alpha}.$$

Diese Formel besagt, dass (\bar{x}, \bar{y}) auf der Regressionsgeraden liegt. □

Beweis von Satz 6.7 (1) und (2): Wir starten mit der Formel zur Berechnung von $\hat{\alpha}$:

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad \Longleftrightarrow \quad \bar{y} = \hat{\beta} \bar{x} + \hat{\alpha}$$

$$\begin{aligned}
\implies \quad \underbrace{\frac{1}{N} \sum_{t=1}^N b_t}_{=\bar{y}} &= \hat{\beta} \cdot \underbrace{\frac{1}{N} \sum_{t=1}^N a_t}_{=\bar{x}} + \hat{\alpha} = \frac{1}{N} \sum_{t=1}^N \hat{\beta} a_t + \frac{1}{N} \cdot N \hat{\alpha} \\
&= \frac{1}{N} \left(\sum_{t=1}^N \hat{\beta} a_t + \underbrace{\sum_{t=1}^N \hat{\alpha}}_{=N\hat{\alpha}} \right) = \frac{1}{N} \sum_{t=1}^N \underbrace{(\hat{\beta} a_t + \hat{\alpha})}_{=\hat{b}_t} = \frac{1}{N} \underbrace{\sum_{t=1}^N \hat{b}_t}_{=\bar{\hat{y}}}
\end{aligned}$$

Also gilt

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N b_t = \frac{1}{N} \sum_{t=1}^N \hat{b}_t = \bar{\hat{y}}, \quad (6.5)$$

d.h. wir haben Satz 6.7 (2) bewiesen. Nach (6.5) gilt

$$\begin{aligned}
\frac{1}{N} \sum_{t=1}^N b_t = \frac{1}{N} \sum_{t=1}^N \hat{b}_t \quad \Big| \cdot N &\iff \sum_{t=1}^N b_t = \sum_{t=1}^N \hat{b}_t \iff \\
\sum_{t=1}^N b_t - \sum_{t=1}^N \hat{b}_t = 0 &\iff \sum_{t=1}^N \underbrace{(b_t - \hat{b}_t)}_{=e_t} = 0 \iff \sum_{t=1}^N e_t = 0,
\end{aligned}$$

d.h. wir haben Satz 6.7 (1) bewiesen. □

Beweis von Satz 6.7 (4): Formel für $\hat{\beta}$ ist

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} \iff \frac{\sigma_{XY}}{\sigma_X^2} - \hat{\beta} = 0 \iff \sigma_{XY} - \hat{\beta} \sigma_X^2 = 0.$$

Wir formen nun $\sigma_{XY} - \hat{\beta} \sigma_X^2 = 0$ geeignet um:

$$\begin{aligned}
0 = \sigma_{XY} - \hat{\beta} \sigma_X^2 &= \sigma_{XY} - \hat{\beta} \sigma_X^2 - \hat{\beta} \bar{x}^2 + \hat{\beta} \bar{x}^2 \\
&= \sigma_{XY} - \hat{\beta} (\sigma_X^2 + \bar{x}^2) + (\hat{\beta} \bar{x}) \bar{x} \\
&= \sigma_{XY} - \hat{\beta} (\sigma_X^2 + \bar{x}^2) + (\bar{y} - \hat{\alpha}) \bar{x},
\end{aligned}$$

wobei wir im letzten Schritt genutzt haben, dass die Formel für $\hat{\alpha}$ liefert

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \iff \hat{\beta} \bar{x} = \bar{y} - \hat{\alpha}.$$

Wir formen nun weiter um

$$0 = \sigma_{XY} - \hat{\beta} (\sigma_X^2 + \bar{x}^2) + (\bar{y} - \hat{\alpha}) \bar{x} = (\sigma_{XY} + \bar{x} \bar{y}) - \hat{\beta} (\sigma_X^2 + \bar{x}^2) - \hat{\alpha} \bar{x}. \quad (6.6)$$

Nach Lemma 5.20 gilt

$$\sigma_{XY} = \frac{1}{N} \sum_{t=1}^N a_t b_t - \bar{x} \bar{y} \iff \sigma_{XY} + \bar{x} \bar{y} = \frac{1}{N} \sum_{t=1}^N a_t b_t, \quad (6.7)$$

und nach Lemma 4.3 gilt

$$\sigma_X^2 = \frac{1}{N} \sum_{t=1}^N a_t^2 - \bar{x}^2 \quad \iff \quad \sigma_X^2 + \bar{x}^2 = \frac{1}{N} \sum_{t=1}^N a_t^2. \quad (6.8)$$

Einsetzen von (6.7) und (6.8), sowie der Formel

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N a_t$$

für das arithmetische Mittel \bar{x} , in (6.6) und Vereinfachen liefert:

$$\begin{aligned} 0 &= (\sigma_{XY} + \bar{x}\bar{y}) - \hat{\beta}(\sigma_X^2 + \bar{x}^2) - \hat{\alpha}\bar{x} \\ &= \frac{1}{N} \sum_{t=1}^N a_t b_t - \hat{\beta} \frac{1}{N} \sum_{t=1}^N a_t^2 - \hat{\alpha} \frac{1}{N} \sum_{t=1}^N a_t \\ &= \frac{1}{N} \left(\sum_{t=1}^N a_t b_t - \sum_{t=1}^N \hat{\beta} a_t^2 - \sum_{t=1}^N \hat{\alpha} a_t \right) \\ &= \frac{1}{N} \sum_{t=1}^N (a_t b_t - \hat{\beta} a_t^2 - \hat{\alpha} a_t) \\ &= \frac{1}{N} \sum_{t=1}^N a_t \underbrace{(b_t - (\hat{\beta} a_t + \hat{\alpha}))}_{=e_t} \\ &= \frac{1}{N} \sum_{t=1}^N a_t \cdot e_t. \end{aligned}$$

Also haben wir $\sum_{t=1}^N a_t e_t = 0$ bewiesen. □

Die Methode der kleinsten Quadrate ermöglicht **immer** die Bestimmung einer **optimalen** affin linearen Regressionsfunktion. Dieses ist auch dann der Fall, wenn man im Streudiagramm zwischen X und Y **keinen** Zusammenhang sehen kann. Daher braucht man ein **Kriterium zur Bewertung** der affin linearen Regressionsfunktion. Dieses werden wir im nächsten Teilkapitel untersuchen.

Bevor wir lernen, wie man die Güte einer Regressionsfunktion bewertet, kommen wir noch einmal auf den Nutzen einer Regressionsfunktion (mit einer guten Annäherung an die Daten) zurück.

Bemerkung 6.9. (Nutzen einer Regressionsfunktion für „Prognosezwecke“)

Liegt eine große und repräsentative Stichprobe aus einer Grundgesamtheit vor, so kann für ein zweidimensionales metrisch skaliertes Merkmal (X, Y) , bei dem Y von X affin linear abhängt, die affin lineare Abhängigkeit von Y von X mit der Regressionsfunktion

$y = f(x) = \beta x + \alpha$ modelliert werden. Wir nehmen nun an, dass diese Regressionsfunktion ein gutes Modell des funktionalen Zusammenhangs ist. Für Objekte der Grundgesamtheit, die nicht in der Stichprobe enthalten sind, kann nun unter Kenntnis des Einzelwerts x von Merkmal X ein **Referenzwert** (oder eine **Prognose**) $y = \beta x + \alpha$ für Y berechnet werden.

Betrachten wir ein Beispiel:

- Es werden die Körpergröße (Merkmal X) und das Gewicht (Merkmal Y) von 1000 gesunden volljährigen Personen mit einem „gesunden“ Körpergewicht (Inspektion durch Arzt) bestimmt.
- Mit diesem Datensatz wird eine affin lineare Regressionsfunktion bestimmt.
- Die affin lineare Regressionsfunktion kann zur Berechnung von Referenzwerten für das Gewicht für jede beliebige Körpergröße benutzt werden.

6.3 Streuungszerlegung und Bestimmtheitsmaß

Wir wollen nun ein Kriterium zur Bewertung der Güte der Annäherung (oder Approximation) der affin linearen Regressionsfunktion an die Daten herleiten. Unsere **Grundidee** ist dabei, dass die affin lineare Regressionsfunktion den Zusammenhang zwischen X und Y **um so besser** beschreibt **je kleiner** die Absolutbeträge $|e_t| = |y_t - \hat{y}_t|$ sind. Wir betrachten daher

$$\sigma_e^2 = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (= \text{Streuung/Varianz der Residuen}). \quad (6.9)$$

Bei (6.9) handelt es sich wirklich um die **Varianz der Residuen**, denn diese ist

$$\sigma_e^2 = \frac{1}{N} \sum_{t=1}^N (e_t - \underbrace{\mu_e}_{=0})^2 = \frac{1}{N} \sum_{t=1}^N e_t^2,$$

da nach Satz 6.7 (1) das arithmetische Mittel μ_e der Residuen $\mu_e = \frac{1}{N} \sum_{t=1}^N e_t = 0$ ist.

Je kleiner die Streuung der Residuen (6.9) ist, desto besser ist die Annäherung der affin linearen Regressionsfunktion an die Daten. Allerdings müssen wir die Streuung der Residuen noch in eine Beziehung zur Größenordnung von Y , genauer in die Beziehung zu der Streuung σ_Y^2 der Y -Daten setzen. Dabei hilft uns die **Streuungszerlegung** (siehe Satz 6.10 unten).

Satz 6.10. (Streuungszerlegung)

Die Voraussetzungen und Notation seien wie in Satz 6.4. Für die mit der Methode der kleinsten Quadrate berechnete affin lineare Regressionsfunktion gilt die **Streuungszerlegung**:

$$\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_e^2,$$

d.h. die **Varianz/Streuung der Einzelwerte**

$$\sigma_Y^2 = \frac{1}{N} \sum_{t=1}^N (b_t - \bar{y})^2 = \frac{1}{N} \sum_{t=1}^N b_t^2 - \bar{y}^2$$

kann additiv zerlegt werden in die **Varianz/Streuung der geschätzten Werte**

$$\sigma_{\hat{Y}}^2 = \frac{1}{N} \sum_{t=1}^N (\hat{b}_t - \bar{y})^2 = \frac{1}{N} \sum_{t=1}^N \hat{b}_t^2 - \bar{y}^2 \quad (\text{erklärte Streuung})$$

und die **Varianz/Streuung der Residuen**

$$\sigma_e^2 = \frac{1}{N} \sum_{t=1}^N e_t^2 \quad (\text{Reststreuung}).$$

Beweis von Satz 6.10: Der Beweis erfolgt durch Nachrechnen unter Ausnutzung von Satz 6.7 (4). Wir beweisen die Streuungsformel auf einem Übungszettel. \square

In der nächsten Bemerkung untersuchen wir die Streuungszerlegung noch einmal genauer.

Bemerkung 6.11. (Verständnis/Interpretation der Streuungszerlegung)

$$\sigma_Y^2 = \sigma_{\hat{Y}}^2 + \sigma_e^2 \quad (\text{Streuungszerlegung}) \quad (6.10)$$

oder in Worten

$$(\text{Streuung der Einzelwerte von } Y) = (\text{erklärte Streuung}) + (\text{Reststreuung}).$$

Um die Streuungszerlegung besser zu verstehen, betrachten wir nun zwei Grenzfälle:

(1) Wenn die **Reststreuung Null** ist, also wenn

$$\sigma_e^2 = \frac{1}{N} \sum_{t=1}^N e_t^2 = \frac{1}{N} \sum_{t=1}^N \underbrace{(b_t - \hat{b}_t)^2}_{\geq 0} = 0$$

gilt, dann folgt

$$\hat{b}_t = b_t \quad \text{für alle } t = 1, 2, \dots, N,$$

d.h. die geschätzten Werte \hat{b}_t sind gleich den Einzelwerten b_t und alle Einzelwerte liegen auf der Regressionsgeraden. Die Streuung der Einzelwerte wird **vollständig erklärt**, und es liegt ein **perfekter affin linearer Zusammenhang** vor.

(2) Wenn die **erklärte Streuung Null** ist, also wenn

$$\sigma_{\hat{Y}}^2 = \frac{1}{N} \sum_{t=1}^N (\hat{b}_t - \bar{y})^2 = 0$$

gilt, dann muss $\widehat{b}_t - \bar{y} = 0$ für $t = 1, 2, \dots, N$ gelten, d.h. $\widehat{b}_t = \bar{y}$ für $t = 1, 2, \dots, N$. Die Regressionsgerade ist **horizontal** und hat den Wert $y = \bar{y}$. Weiter folgt aus der Streuungszerlegung (6.10), dass $\sigma_e^2 = \sigma_Y^2$ ist. Die Reststreuung σ_e^2 ist also gleich der Streuung der Einzelwerte σ_Y^2 . Die affin lineare Regressionsfunktion **erklärt die Streuung der Einzelwerte (von Y) überhaupt nicht!**

Als **relatives Maß** zur Bewertung der erklärten Streuung nehmen wir $\sigma_{\widehat{Y}}^2/\sigma_Y^2$ und setzen somit die erklärte Streuung in Beziehung zu der Streuung in den Einzelwerte von Y .

Definition 6.12. (Bestimmtheitsmaß)

Die Voraussetzungen und Notation seien wie in Satz 6.4 und Satz 6.10. Das **Bestimmtheitsmaß** r^2 zur Bewertung der mit der Methode der kleinsten Quadrate berechneten affin linearen Regressionsfunktion ist definiert durch

$$r^2 = \frac{\sigma_{\widehat{Y}}^2}{\sigma_Y^2} \quad \left(= \frac{\text{erklärte Streuung}}{\text{Streuung der Einzelwerte}} \right).$$

Das **Bestimmtheitsmaß** r^2 misst den **Anteil** der erklärten Streuung an der Gesamtstreuung der Einzelwerte von Y .

Bevor wir zu der Interpretation des Bestimmtheitsmaßes r^2 kommen, halten wir in einem Lemma wichtige alternative Darstellungen von r^2 fest. Diese werden uns bei der Interpretation von r^2 helfen.

Lemma 6.13. (alternative Darstellungen des Bestimmtheitsmaßes)

Die Voraussetzungen und Notation seien wie in Definition 6.12. Das **Bestimmtheitsmaß** r^2 erfüllt $0 \leq r^2 \leq 1$, und es gilt

$$r^2 = \frac{\sigma_{\widehat{Y}}^2}{\sigma_Y^2} = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = 1 - \frac{\sigma_e^2}{\sigma_Y^2} = \varrho_{XY}^2. \quad (6.11)$$

Dabei ist ϱ_{XY} der Korrelationskoeffizient von X und Y (vgl. Definition 5.16).

Beweis von Lemma 6.13: Der Beweis erfolgt mit Hilfe der Streuungszerlegung und der Formel für das Verhalten der Varianz bei dem Anwenden einer affin linearen Funktion auf die Einzelwerte. Wir führen den Beweis auf einem Übungszettel durch \square

Bevor wir die Faustregel zur Bewertung der Regressionsgeraden mit Hilfe des Bestimmtheitsmaßes formulieren, schauen wir (6.11) genauer an. Da r^2 als $r^2 = \sigma_{\widehat{Y}}^2/\sigma_Y^2$ (Quotient zwei Quadrate) definiert ist, sieht man direkt, dass $r^2 \geq 0$ gelten muss. An der Darstellung

$$r^2 = 1 - \underbrace{\frac{\sigma_e^2}{\sigma_Y^2}}_{\geq 0}$$

in (6.11) sehen wir, dass $r^2 \leq 1$ gelten muss. Somit liegen die möglichen Werte für r^2 in dem Intervall $[0, 1]$. Weiter folgt aus (6.11), dass

$$r^2 = 1 - \frac{\sigma_e^2}{\sigma_Y^2} \iff \underbrace{\frac{\sigma_Y^2}{\sigma_Y^2}}_{=r^2} = 1 - \frac{\sigma_e^2}{\sigma_Y^2} \iff \underbrace{\frac{\sigma_Y^2}{\sigma_Y^2}}_{=r^2} + \frac{\sigma_e^2}{\sigma_Y^2} = 1 = \frac{\sigma_Y^2}{\sigma_Y^2}$$

An der letzten Formel sehen wir, dass r^2 , also der **Anteil der erklärten Streuung** (an der Streuung der Einzelwerte von Y), plus den **Anteil der (nicht erklärten) Reststreuung** (an der Streuung der Einzelwerte von Y) gerade Eins ergeben. Damit können wir $r^2 \cdot 100\%$ als den **Prozentsatz der durch die affin lineare Regressionsfunktion erklärten Streuung** auffassen.

Diese Überlegungen liefern den Hintergrund für unsere Faustregel zur Bewertung der Regressionsgeraden mit Hilfe des Bestimmtheitsmaßes.

Lemma 6.14. (Faustregel zur Interpretation von r^2)

Für eine mit der Methode der kleinsten Quadrate berechnete affin lineare Regressionsfunktion (Regressionsgerade) bewerten wir die **Güte der Regressionsgerade** mit dem Bestimmtheitsmaß r^2 wie folgt:

$$0 \leq r^2 < \frac{1}{3} \implies \text{kein linearer Zusammenhang zwischen } X \text{ und } Y \quad (6.12)$$

$$\frac{1}{3} \leq r^2 \leq \frac{2}{3} \implies \text{schwach ausgeprägter linearer Zusammenhang zwischen } X \text{ und } Y \quad (6.13)$$

$$\frac{2}{3} < r^2 \leq 1 \implies \text{stark ausgeprägter linearer Zusammenhang zwischen } X \text{ und } Y \quad (6.14)$$

Die Regressionsgerade als **Modell für den linearen Zusammenhang** zwischen X und Y **erklärt** $r^2 \cdot 100\%$ **der Streuung der Einzelwerte von** Y .

Im Fall (6.12) ist die berechnete Regressionsgerade ein Modell für einen „nicht vorhandenen“ linearen Zusammenhang, und hat daher keine Aussagekraft. Im Fall (6.13) ist die berechnete Regressionsgerade ein schwaches Modell für einen Zusammenhang zwischen X und Y . Im Fall (6.14) liegt ein stark ausgeprägter linearer Zusammenhang zwischen X und Y vor, und die Regressionsgerade stellt ein gutes Modell für diesen linearen Zusammenhang dar.

Betrachten wir hierzu zwei Beispiele.

Beispiel 6.15. (Bewertung der affin linearen Regressionsfunktion – Abhängigkeit der Laufzeit über 10 km vom Lauftraining)

In Beispiel 6.5 hatten wir die Regressionsgerade als Modell für die affin lineare Abhängigkeit der Laufzeit (in min) über 10 km (Merkmal Y) von den Stunden Lauftraining pro

Woche (Merkmal X) berechnet. Wir wollen nun mit dem Bestimmtheitsmaß die Qualität dieses Modells bewerten: Aus Beispiel 5.17 wissen wir bereits, dass

$$\sigma_{XY} = -30,59375 \text{ h} \cdot \text{min}, \quad \sigma_X^2 = 8,859375 \text{ h}^2, \quad \sigma_Y^2 = 125,4375 \text{ min}^2.$$

Nach Lemma 6.13 gilt

$$r^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \frac{(-30,59375 \text{ h} \cdot \text{min})^2}{8,859375 \text{ h}^2 \cdot 125,4375 \text{ min}^2} \approx 0,842 \quad (\text{und damit } \frac{2}{3} < r^2 \leq 1).$$

Also liegt ein *stark ausgeprägter linearer Zusammenhang* zwischen dem Merkmal $X =$ „Stunden Lauftraining pro Woche“ und $Y =$ „Laufzeit (in min) über 10 km“ vor. Die in Beispiel 6.5 berechnete Regressionsgerade liefert ein gutes Modell für diesen linearen Zusammenhang und erklärt 84,2% der Streuung der Laufzeiten auf 10 km.

Beispiel 6.16. (Bewertung der affin linearen Regressionsfunktion – Kauf von Lehrbüchern für Analysis 1 und Lineare Algebra 1)

Sei (X, Y) das zweidimensionale Merkmal aus Beispielen 5.2, 5.3, 5.5 und 5.18 wobei $X =$ „Anzahl der für die Vorlesung Analysis 1 gekauften Lehrbücher“, $Y =$ „Anzahl der für die Vorlesung Lineare Algebra 1 gekauften Lehrbücher“. Wir wollen nun ein lineares Regressionsmodell für die Abhängigkeit der Anzahl der für die Lineare Algebra 1 gekauften Lehrbücher von der Anzahl der für die Analysis 1 gekauften Lehrbücher berechnen und bewerten. Aus Beispiel 5.18 wissen wir

$$\bar{x} = 1, \quad \bar{y} = 1, \quad \sigma_X = \sqrt{0,5} = \sqrt{\frac{5}{10}}, \quad \sigma_Y = \sqrt{0,7} = \sqrt{\frac{7}{10}}, \quad \sigma_{XY} = 0,4 = \frac{4}{10}.$$

Damit folgt

$$\hat{\beta} = \frac{\sigma_{XY}}{\sigma_X^2} = \frac{0,4}{0,5} = \frac{4}{5} = 0,8 \quad \text{und} \quad \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = 1 - \frac{4}{5} \cdot 1 = \frac{1}{5} = 0,2,$$

und die affin lineare Regressionsfunktion ist

$$y = \hat{\beta} x + \hat{\alpha} = \frac{4}{5} \cdot x + \frac{1}{5}.$$

Das Bestimmtheitsmaß ist

$$r^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2} = \frac{\left(\frac{4}{10}\right)^2}{\frac{5}{10} \cdot \frac{7}{10}} = \frac{16}{35} \approx 0,46.$$

Da gilt $\frac{1}{3} \leq r^2 = 0,46 \leq \frac{2}{3}$ liegt ein *schwach ausgeprägter linearer Zusammenhang* vor.

Wahrscheinlichkeitsrechnung: Grundlagen

In diesem Kapitel lernen wir die Grundlagen der Wahrscheinlichkeitsrechnung kennen. Der Begriff der Wahrscheinlichkeit ist in einem gewissen Sinn analog zum Begriff der Häufigkeit: Man kann sich die Wahrscheinlichkeit eines Ereignisses als die „theoretisch erwartete“ Häufigkeit dieses Ereignisses vorstellen. So ist beispielsweise beim Würfeln mit einem perfekten Würfel die Wahrscheinlichkeit, eine 6 zu würfeln, gerade $1/6$.

7.1 Zufallsexperimente und Ereignisse

Wir beginnen mit der Einführung der klassischen Grundbegriffe der Wahrscheinlichkeitsrechnung, dem Zufallsexperiment und Ereignissen, bevor wir definieren können, was Wahrscheinlichkeit ist.

Definition 7.1. (Zufallsexperiment)

Ein **Zufallsexperiment** ist ein Vorgang, der nach einer bestimmten Vorschrift ausgeführt wird, der beliebig oft wiederholbar ist, und dessen Ergebnis **vom Zufall abhängt**.

Führt man ein Zufallsexperiment mehrmals durch, so setzt man voraus, dass sich die einzelnen Ergebnisse (der Durchführungen) nicht gegenseitig beeinflussen, also **unabhängig voneinander** sind.

Betrachten wir einige Beispiele.

Beispiel 7.2. (Zufallsexperiment)

- (a) einmaliges oder mehrmaliges Werfen einer Münze oder eines Würfels
- (b) Ziehung der Lottozahlen

- (c) Entnahme einer Zufallsstichprobe aus der laufenden Produktion eines Produkts zur Qualitätskontrolle (z.B. Bestimmung der Brenndauer von Glühbirnen)

Definition 7.3. (Elementarereignis, Ereignisraum und Ereignis)

- (1) Jedes mögliche Ergebnis eines Zufallsexperiments wird als **Elementarereignis** bezeichnet. Bei jeder Durchführung eines Zufallsexperiments tritt **genau ein** Elementarereignis ein. Wir bezeichnen die Elementarereignisse mit $\{e_1\}, \{e_2\}, \{e_3\}, \dots$
- (2) Die Menge aller Elementarereignisse, $S = \{e_1, e_2, e_3, \dots\}$, nennt man den **Ereignisraum** des Zufallsexperiments. Der Ereignisraum ist **endlich**, wenn er nur endlich viele Elementarereignisse umfasst; andernfalls ist er **unendlich**.
- (3) Jede beliebige Teilmenge des Ereignisraums ist ein **Ereignis**. In anderen Worten jedes Ereignis ist eine Menge von Elementarereignissen.

Die geschweiften Klammern $\{ \}$ in Definition 7.3 sind Mengenklammern. **Wir beschreiben Ereignisse mathematisch als Mengen von Elementarereignissen.**

Betrachten wir hierzu zwei Beispiele.

Beispiel 7.4. (einmaliges Werfen eines klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

Elementarereignisse: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ und $\{6\}$

Ereignisraum: $S = \{1, 2, 3, 4, 5, 6\}$

Ereignis: Ein Beispiel für ein Ereignis, das kein Elementarereignis ist, ist z.B. das Ereignis $A =$ „Werfen einer geraden Augenzahl“, also $A = \{2, 4, 6\}$.

Beispiel 7.5. (zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

Elementarereignisse: $\{(K, K)\}, \{(K, Z)\}, \{(Z, K)\}$ und $\{(Z, Z)\}$

Ereignisraum: $S = \{(K, K), (K, Z), (Z, K), (Z, Z)\}$

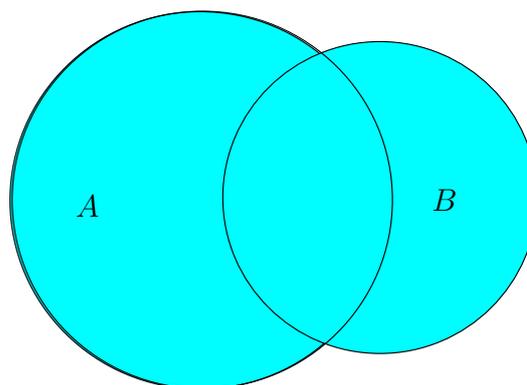
Ereignis: Ein Beispiel für ein Ereignis, das kein Elementarereignis ist, ist z.B. $B =$ „Im ersten Wurf wird eine Zahl(seite) geworfen.“, also $B = \{(Z, K), (Z, Z)\}$.

Wir sehen an Definition 7.3 und den beiden Beispielen, dass man ein Elementarereignis e_i üblicherweise als Menge $\{e_i\}$ schreibt, wenn es als eigenständiges Ereignis betrachtet.

Da Ereignisse Mengen vom Elementarereignissen sind, können wir für Ereignisse die üblichen Mengenoperationen ausführen. Wir führen diese nun ein und betrachten jeweils ein Beispiel dazu.

Definition 7.6. (Vereinigung zweier Ereignisse)

Die **Vereinigung** $A \cup B$ zweier Ereignisse A und B ist die Menge aller Elementarereignisse, die zu A **oder** zu B (oder zu A und B) gehören.



Es gilt immer $A \cup B = B \cup A$.

Betrachten wir zwei Beispiele.

Beispiel 7.7. (Vereinigung – einmaliges Werfen eines klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$$C = \text{„Werfen einer geraden Zahl oder einer Primzahl“} = \underbrace{\{2, 4, 6\} \cup \{2, 3, 5\}}_{= A \cup B} = \{2, 3, 4, 5, 6\}$$

mit $A = \text{„Werfen einer geraden Zahl“} = \{2, 4, 6\}$, $B = \text{„Werfen einer Primzahl“} = \{2, 3, 5\}$

Beachten Sie, dass das mathematische „oder“ in „Werfen einer Quadratzahl oder eine Primzahl“ und allgemeiner in Definition 7.6 nicht als ausschließendes „entweder oder“ zu lesen ist.

Beispiel 7.8. (Vereinigung – zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

$A = \text{„Im zweiten Wurf wird eine Zahl(seite) geworfen.“} = \{(K, Z), (Z, Z)\}$

$B = \text{„Im ersten Wurf wird eine Zahl(seite) geworfen.“} = \{(Z, K), (Z, Z)\}$

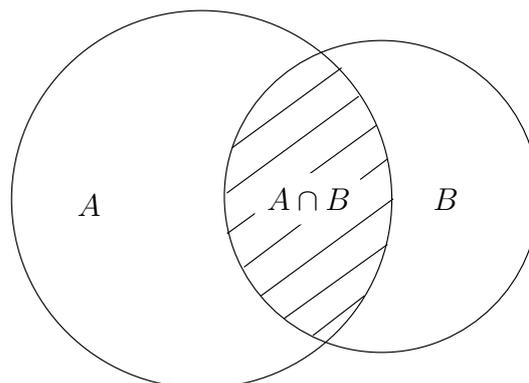
$$C = \text{„Im ersten oder im zweiten Wurf wird eine Zahl(seite) geworfen“} \\ = \underbrace{\{(K, Z), (Z, Z)\} \cup \{(Z, K), (Z, Z)\}}_{= A \cup B} = \{(K, Z), (Z, K), (Z, Z)\}$$

Definition 7.9. (Durchschnitt zweier Ereignisse)

Der **Durchschnitt** $A \cap B$ zweier Ereignisse A und B ist die Menge aller Elementarereignisse, die zu A **und** zu B gehören.

Es gilt immer $A \cap B = B \cap A$.

Betrachten wir auch hierzu zwei Beispiele.



Beispiel 7.10. (Durchschnitt – einmaliges Werfen eines klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$$D = \text{„Werfen einer geraden Primzahl“} = \underbrace{\{2, 4, 6\} \cap \{2, 3, 5\}}_{= A \cap B} = \{2\}$$

mit $A = \text{„Werfen einer geraden Zahl“} = \{2, 4, 6\}$, $B = \text{„Werfen einer Primzahl“} = \{2, 3, 5\}$

Beispiel 7.11. (Durchschnitt – zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

$$A = \text{„Im zweiten Wurf wird eine Zahl(seite) geworfen.“} = \{(K, Z), (Z, Z)\}$$

$$B = \text{„Im ersten Wurf wird eine Zahl(seite) geworfen.“} = \{(Z, K), (Z, Z)\}$$

$$D = \text{„Im ersten und im zweiten Wurf wird eine Zahl(seite) geworfen“}$$

$$= \underbrace{\{(K, Z), (Z, Z)\} \cap \{(Z, K), (Z, Z)\}}_{= A \cap B} = \{(Z, Z)\}$$

Definition 7.12. (sich ausschließende Ereignisse)

Zwei Ereignisse A und B (eines Zufallsexperiments) **schließen sich aus**, d.h. sie sind als Mengen **disjunkt**, wenn es kein Elementarereignis gibt, das sowohl zu A als auch zu B gehört, also $A \cap B = \emptyset$, wobei $\emptyset = \{\}$ die leere Menge (das „**unmögliche Ereignis**“) ist.

Betrachten wir zwei Beispiele für sich ausschließende Ereignisse.

Beispiel 7.13. (sich ausschließende Ereignisse – einmaliges Werfen eines klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$$A = \text{„Werfen einer geraden Zahl“} = \{2, 4, 6\}$$

$$E = \text{„Werfen einer ungeraden Zahl“} = \{1, 3, 5\}$$

Dann ist $A \cap E = \{2, 4, 6\} \cap \{1, 3, 5\} = \emptyset$. Die Ereignisse A und E schließen sich aus.

Beispiel 7.14. (sich ausschließende Ereignisse – zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

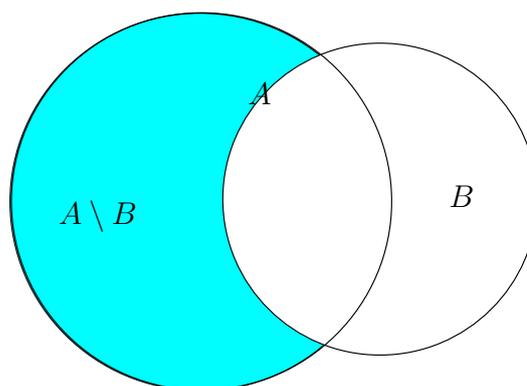
$$E = \text{„Im ersten Wurf wird eine Kopf(seite) geworfen.“} = \{(K, Z), (K, K)\}$$

$B =$ „Im ersten Wurf wird eine Zahl(seite) geworfen.“ $= \{(Z, K), (Z, Z)\}$

$E \cap B =$ „Im ersten Wurf wird eine Zahl(seite) und eine Kopf(seite) geworfen“
 $= \underbrace{\{(K, Z), (K, K)\} \cap \{(Z, K), (Z, Z)\}}_{= E \cap B} = \emptyset$. Die Ereignisse A und E schließen sich aus.

Definition 7.15. (Differenzmenge)

Die **Differenzmenge** $A \setminus B$ („ A ohne B “) zweier Ereignisse A und B enthält alle Elementarereignisse, die in A enthalten sind, aber die jedoch **nicht** in B enthalten sind.



Betrachten wir auch hierzu zwei Beispiele.

Beispiel 7.16. (Differenzmenge – einmaliges Werfen eines klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$F =$ „Werfen einer geraden Zahl, die keine Primzahl ist“ $\underbrace{\{2, 4, 6\} \setminus \{2, 3, 5\}}_{= A \setminus B} = \{4, 6\}$

mit $A =$ „Werfen einer geraden Zahl“ $= \{2, 4, 6\}$, $B =$ „Werfen einer Primzahl“ $= \{2, 3, 5\}$

Beispiel 7.17. (Differenzmenge – zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

$A =$ „Im zweiten Wurf wird eine Zahl(seite) geworfen.“ $= \{(K, Z), (Z, Z)\}$

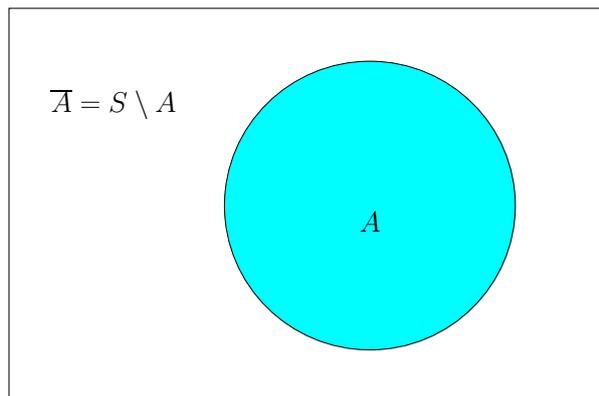
$B =$ „Im ersten Wurf wird eine Zahl(seite) geworfen.“ $= \{(Z, K), (Z, Z)\}$

$B \setminus A =$ „Im ersten Wurf wird eine Zahl(seite) und im zweiten Wurf wird keine Zahl(seite) geworfen“ $= \underbrace{\{(Z, K), (Z, Z)\} \setminus \{(K, Z), (Z, Z)\}}_{= B \setminus A} = \{(Z, K)\}$

Definition 7.18. (Komplementärereignis)

Das **Komplementärereignis** (oder **Gegenereignis**) \bar{A} zu einem Ereignis A ist die Menge sämtlicher Elementarereignisse des Ereignisraums S , die **nicht im Ereignis** A enthalten sind, also $\bar{A} = S \setminus A$.

Betrachten wir auch hierzu zwei Beispiele.

**Beispiel 7.19. (Komplementärereignis – einmaliges Werfen eines klassischen Würfels)**

Zufallsexperiment: einmaliges Werfen eines klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$$E = \text{„Werfen einer ungeraden Zahl“} = \{1, 3, 5\} = \underbrace{\{2, 4, 6\}}_{=\bar{A}}$$

mit $A = \text{„Werfen einer geraden Zahl“} = \{2, 4, 6\}$

Beispiel 7.20. (Komplementärereignis – zweifacher Münzwurf)

Zufallsexperiment: Eine Münze mit den Seiten Kopf (K) und Zahl (Z) wird zweimal geworfen, und wir notieren die Ergebnisse jeweils als

(obere Seite im ersten Wurf, obere Seite im zweiten Wurf).

$$A = \text{„Im zweiten Wurf wird eine Zahl(seite) geworfen.“} = \{(K, Z), (Z, Z)\}$$

$$\bar{A} = S \setminus A = \text{„Im zweiten Wurf wird keine Zahl(seite) geworfen.“}$$

$$= \{(K, K), (Z, K)\} = \overline{\{(K, Z), (Z, Z)\}}$$

7.2 Wahrscheinlichkeiten

In diesem Teilkapitel lernen wir zwei Definitionen des Wahrscheinlichkeitsbegriffs kennen. Die zweite Definition stellt eine Erweiterung und Verallgemeinerung der ersten dar. Wir starten mit der klassischen Wahrscheinlichkeitsdefinition von Laplace.

Definition 7.21. (klassische Definition der Wahrscheinlichkeit von Laplace)

Wir betrachten ein Zufallsexperiment mit einem **endlichen** Ereignisraum S mit **gleichwahrscheinlichen** Elementarereignissen: Die **Wahrscheinlichkeit** $W(A)$, dass bei einer Durchführung des Zufallsexperiments das Ereignis A eintritt, ist

$$W(A) = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller gleichmöglichen Fälle}} = \frac{h(A)}{h(S)},$$

wobei $h(A)$ bzw. $h(S)$ die absolute Häufigkeit (oder Anzahl) der Elementarereignisse in A bzw. S ist.

Betrachten wir zwei Beispiele

Beispiel 7.22. (Wahrscheinlichkeitsbegriff von Laplace – einmaliges Werfen eines perfekten klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines perfekten klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

Ein „perfekter“ Würfel ist ein Würfel, bei dem alle Seiten des Würfels gleichwahrscheinlich sind, d.h. alle Augenzahlen treten gleichwahrscheinlich auf. Also sind alle Elementarereignisse gleichwahrscheinlich, und wir können die Wahrscheinlichkeitsdefinition nach Laplace verwenden.

Ereignisraum: $S = \{1, 2, 3, 4, 5, 6\}$

$$A = \text{„Werfen einer geraden Zahl“} = \{2, 4, 6\}, \quad W(A) = \frac{h(A)}{h(S)} = \frac{3}{6} = \frac{1}{2} = 0,5$$

$$B = \text{„Werfen einer 1“} = \{1\}, \quad W(B) = \frac{h(\{1\})}{h(S)} = \frac{1}{6}$$

Beispiel 7.23. (Wahrscheinlichkeitsbegriff von Laplace – Urnenziehen)

Zufallsexperiment: Eine Urne enthält zwei blaue und drei rote Kugeln, die sich durch Befühlen nicht unterscheiden lassen. Es wird blind eine Kugel aus der Urne entnommen.

Elementarereignisse sind $b = \text{„Ziehen einer blauen Kugel“}$ und $r = \text{„Ziehen einer roten Kugel“}$. Der Ereignisraum ist $S = \{b, r\}$.

Da das Ziehen jeder einzelnen Kugel in der Urne gleichwahrscheinlich ist, ist für jede der fünf Kugeln in der Urne, die Wahrscheinlichkeit, dass diese gezogen wird, $1/5$. Nach der Wahrscheinlichkeitsdefinition von Laplace folgt:

$$W(\{b\}) = \frac{2}{5} \quad \text{und} \quad W(\{r\}) = \frac{3}{5}.$$

Hier haben wir **keine** gleichwahrscheinlichen Elementarereignisse. Wir konnten aber trotzdem die Wahrscheinlichkeitsdefinition von Laplace nutzen, da jede der fünf individuellen Kugeln mit der gleichen Wahrscheinlichkeit von $1/5$ gezogen wird. Zwei unter diesen fünf gleichwahrscheinlichen Fällen sind günstig für das Ziehen einer blauen Kugel, und drei unter diesen fünf gleichwahrscheinlichen Fällen sind günstig für das Ziehen einer roten Kugel.

Wir lernen nun die statistische Definition einer Wahrscheinlichkeit kennen, die Definition 7.21 erweitert und verallgemeinert.

Definition 7.24. (statistische Definition der Wahrscheinlichkeit)

Man geht von einem Zufallsexperiment aus, das aus einer langen Abfolge unabhängiger

Versuche/Durchführungen besteht. Die **Wahrscheinlichkeit** $W(A)$ des Ereignisses A ist der **Grenzwert der relativen Häufigkeit des Auftretens von A** , also

$$W(A) = \lim_{n \rightarrow \infty} f_n(A) = \lim_{n \rightarrow \infty} \frac{h_n(A)}{n}, \quad (7.1)$$

wobei $f_n(A)$ bzw. $h_n(A)$ die relative bzw. absolute Häufigkeit des Auftretens des Ereignisses A bei n Versuchen/Durchführungen ist.

Bei dem Grenzwert in (7.1) handelt es sich um **stochastische Konvergenz**; aufgrund des **Gesetzes der großen Zahlen** strebt die Folge der relativen Häufigkeiten $(f_n(A))_{n \in \mathbb{N}}$ gegen die Wahrscheinlichkeit $W(A)$.

Machen wir uns zunächst an dem Würfel-Beispiel klar, dass Definition 7.24 keinen anderen Wahrscheinlichkeitsbegriff liefert als die Laplacesche Definition der Wahrscheinlichkeit, wenn man diese anwenden kann.

Beispiel 7.25. (statistischer Wahrscheinlichkeitsbegriff – einmaliges Werfen eines klassischen perfekten Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen perfekten Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$A =$ „Werfen einer geraden Zahl“ $= \{2, 4, 6\}$; $B =$ „Werfen einer 1“ $= \{1\}$

Fangen wir mit Ereignis B an: Da alle Würfelseiten, also alle Augenzahlen mit der gleichen Wahrscheinlichkeit auftreten, erwarten wir, dass $f_n(B)$, die relative Häufigkeit der Würfe (unter n Würfeln), bei denen die Zahl 1 geworfen wurde, sich immer weiter an $1/6$ annähert, wenn n wächst. Also gilt

$$W(B) = \lim_{n \rightarrow \infty} f_n(B) = \frac{1}{6}.$$

Da alle Würfelseiten, also alle Augenzahlen mit der gleichen Wahrscheinlichkeit auftreten, erwarten wir, dass $f_n(A)$, die relative Häufigkeit der Würfe (unter n Würfeln), bei denen eine gerade Zahl geworfen wurde, sich immer weiter an $3/6 = 1/2$ annähert, wenn n wächst. Also gilt

$$W(A) = \lim_{n \rightarrow \infty} f_n(A) = \frac{3}{6} = \frac{1}{2}.$$

Wir sehen, dass der statistische Wahrscheinlichkeitsbegriff die gleichen Wahrscheinlichkeiten liefert wie der Laplacesche Wahrscheinlichkeitsbegriff, wenn man beide anwenden kann.

Der klassische und der statistische Wahrscheinlichkeitsbegriff erfüllen jeweils die **folgenden Eigenschaften**, die sogenannten **Axiome der Wahrscheinlichkeit**. (Ein „Axiom“ ist ein keines Beweises bedürftiger Grundsatz.)

Axiome 7.26. (der Wahrscheinlichkeit)

Axiom 1: Die Wahrscheinlichkeit $W(A)$ des Ereignisses A eines Zufallsexperiments ist eine eindeutig bestimmte nicht-negative reelle Zahl mit $0 \leq W(A) \leq 1$.

Axiom 2: Für den **Ereignisraum** S , also das Ereignis, das alle Elementarereignisse enthält, gilt $W(S) = 1$, und wir nennen S das „**sichere Ereignis**“.

Axiom 3: Schließen sich zwei Ereignisse A und B eines Zufallsexperiments gegenseitig aus, also wenn gilt $A \cap B = \emptyset$, dann gilt $W(A \cup B) = W(A) + W(B)$.

Wir ziehen einige Folgerungen aus den Axiomen der Wahrscheinlichkeit.

Folgerung 7.27. (aus den Axiomen der Wahrscheinlichkeit)

(1) Aus Axiom 3 folgt: Schließen sich m Ereignisse A_1, A_2, \dots, A_m paarweise aus, also wenn gilt $A_i \cap A_j = \emptyset$ für alle $i, j = 1, \dots, m$ mit $i \neq j$, dann gilt:

$$W(A_1 \cup A_2 \cup \dots \cup A_m) = W(A_1) + W(A_2) + \dots + W(A_m).$$

(2) Aus Axiomen 2 und 3 folgt: $W(\emptyset) = 0$.

(3) Aus Axiom 3 folgt: $W(\bar{A}) = 1 - W(A)$.

Betrachten wir ein Beispiel zu den Folgerungen aus den Axiomen des Wahrscheinlichkeitsbegriffs.

Beispiel 7.28. (einmaliges Werfen eines perfekten klassischen Würfels)

Zufallsexperiment: einmaliges Werfen eines perfekten klassischen Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$A =$ „Werfen einer geraden Zahl“ = $\{2, 4, 6\}$; $E =$ „Werfen einer ungeraden Zahl“

Da die Elementarereignisse $\{2\}$, $\{4\}$ und $\{6\}$ sich paarweise ausschließen und

$$A = \{2, 4, 6\} = \{2\} \cup \{4\} \cup \{6\}$$

gilt, folgt mit Folgerung 7.27 (1), dass

$$W(\{2, 4, 6\}) = W(\{2\}) + W(\{4\}) + W(\{6\}) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}.$$

Da E das Komplementärereignis von A ist, also $E = \bar{A}$, gilt mit Folgerung 7.27 (3), dass

$$W(E) = W(\bar{A}) = 1 - W(A) = 1 - \frac{1}{2} = \frac{1}{2}.$$

7.3 Additionssatz für Wahrscheinlichkeiten

Als letzte Grundlage lernen wir den sehr wichtigen Additionssatz für Wahrscheinlichkeiten kennen. Den relativ technischen Beweis dieses Satzes werden wir nicht besprechen. Wir betrachten aber zwei Anwendungsbeispiele.

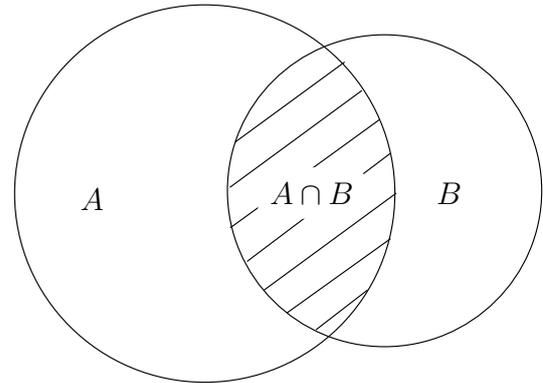
Satz 7.29. (Additionssatz für Wahrscheinlichkeiten)

Sind A und B zwei beliebige Ereignisse eines Zufallsexperiments, dann gilt

$$W(A \cup B) = W(A) + W(B) - W(A \cap B) \quad (7.2)$$

Ganz wichtig ist es, eine **Anschaung** für diesen Satz zu entwickeln: Sind A und B sich ausschließende Ereignisse, also $A \cap B = \emptyset$, so gilt in (7.2)

$$\begin{aligned} W(A \cup B) &= W(A) + W(B) - W(A \cap B) \\ &= W(A) + W(B) + \underbrace{W(\emptyset)}_{=0} \\ &= W(A) + W(B), \end{aligned}$$



und wir haben Axiom 3 einer Wahrscheinlichkeit.

Was ändert sich nun, wenn $A \cap B \neq \emptyset$ ist, also wenn $A \cap B$ Elementarereignisse enthält? Die Wahrscheinlichkeiten von Elementarereignissen in $A \cap B$ werden in $W(A) + W(B)$ „doppelt gezählt“. Daher muss $W(A \cap B)$ von $W(A) + W(B)$ subtrahiert werden.

Der Additionssatz wird in Anwendungen manchmal auch umgestellt als

$$W(A \cap B) = W(A) + W(B) - W(A \cup B)$$

benötigt (wir haben nach $W(A \cap B)$ aufgelöst).

Betrachten wir zwei Beispiele für die Anwendung des Additionssatzes. Das erste dieser Beispiele ist elementar, um plausibel zu machen, wie der Additionssatz funktioniert. Bei dem zweiten Beispiel sehen wir eine „raffiniertere“ Anwendung des Additionssatzes.

Beispiel 7.30. (Einmaliges Werfen eines klassischen perfekten Würfels)

Zufallsexperiment: einmaliges Werfen eines klassischen perfekten Würfels mit den Augenzahlen 1, 2, 3, 4, 5, 6

$A =$ „Werfen einer ungeraden Zahl“ = $\{1, 3, 5\}$

$B =$ „Werfen einer Quadratzahl“ = $\{1, 4\}$

Mit der Wahrscheinlichkeitsdefinition von Laplace finden wir

$$W(A) = \frac{3}{6} = \frac{1}{2} \quad \text{und} \quad W(B) = \frac{2}{6} = \frac{1}{3}.$$

Es gilt für den Durchschnitt

$$A \cap B = \{1\}, \quad W(A \cap B) = \frac{1}{6},$$

und nach dem Additionssatz für Wahrscheinlichkeiten folgt

$$W(A \cup B) = W(A) + W(B) - W(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{4}{6} = \frac{2}{3}.$$

Durch direkte Berechnung von $A \cup B = \{1, 3, 5\} \cup \{1, 4\} = \{1, 3, 4, 5\}$ finden wir mit der Laplaceschen Wahrscheinlichkeitsdefinition natürlich ebenfalls

$$W(A \cup B) = \frac{4}{6} = \frac{2}{3}.$$

Beispiel 7.31. (Wahrscheinlichkeiten für Buchverkaufszahlen)

Mit einer Wahrscheinlichkeit von 0,8 werden in einer großen Buchhandlung am Tag höchstens 20 Exemplare einer Neuerscheinung verkauft, und mit einer Wahrscheinlichkeit von 0,4 werden am Tag mindestens 10 Exemplare dieser Neuerscheinung verkauft. Mit welcher Wahrscheinlichkeit werden am Tag mindestens 10 und höchstens 20 Exemplare dieser Neuerscheinung verkauft?

Lösung: Wir definieren zunächst unsere Ereignisse:

$A =$ „Es werden am Tag höchstens 20 Exemplare verkauft.“

$B =$ „Es werden am Tag mindestens 10 Exemplare verkauft.“

Wir wissen bereits: $W(A) = 0,8$ und $W(B) = 0,4$

$C =$ „Es werden am Tag mindestens 10 **und** höchstens 20 Exemplare verkauft.“

(Anmerkung: „mindestens 10“ bedeutet „ ≥ 10 “; „höchstens 20“ bedeutet „ ≤ 20 “.)

Es gilt $A \cap B = C$, und durch Auflösen von (7.2) nach $A \cap B$ erhalten wir

$$W(A \cap B) = W(A) + W(B) - W(A \cup B).$$

$A \cup B =$ „Es werden am Tag mindestens 10 **oder** höchstens 20 Exemplare verkauft.“

Wir behaupten, dass $A \cup B = S$ ist, also dass $A \cup B$ der ganze Ereignisraum S (das sichere Ereignis) ist, und somit $W(A \cup B) = W(S) = 1$ (nach Axiom 2 einer Wahrscheinlichkeit). Dieses sieht man wie folgt: Sei $\{E_n\}$ (mit $n \in \mathbb{N}_0$) das Elementarereignis, dass am Tag genau n Bücher verkauft wurden. Ist $n \leq 20$, so ist E_n in A . Ist $n > 20$, so ist E_n in B . Also gehören alle Elementarereignisse zu $A \cup B$, und somit ist $A \cup B = S$.

Durch Einsetzen finden wir nun

$$W(C) = W(A \cap B) = W(A) + W(B) - W(A \cup B) = 0,8 + 0,4 - 1 = 0,2.$$

Mit einer Wahrscheinlichkeit von 0,2 (also 20%) werden am Tag mindestens 10 und höchstens 20 Exemplare der Neuerscheinung verkauft.

Wahrscheinlichkeitsrechnung: Weiterführende Resultate

Zunächst lernen wir in diesem Kapitel bedingte Wahrscheinlichkeiten, stochastisch unabhängige und stochastisch abhängige Ereignisse und den Multiplikationssatz für Wahrscheinlichkeiten kennen. Hier werden wir auch Baumdiagramme kennenlernen, mit denen man mehrstufige Zufallsexperimente gut veranschaulichen kann. Danach leiten wir zwei wichtige Sätze her: das Theorem der totalen Wahrscheinlichkeit und das Theorem von Bayes. Wir betrachten wie immer verschiedene Anwendungsbeispiele zu den neuen Inhalten kennen.

8.1 Bedingte Wahrscheinlichkeiten, stochastisch unabhängige Ereignisse und der Multiplikationssatz

Wir starten mit einem motivierenden Beispiel.¹

Beispiel 8.1. (Kugeln werden ohne Zurücklegen aus einer Urne gezogen)

Aus einer Urne mit 6 Kugeln, von denen 2 schwarz und 4 rot sind, werden *ohne Zurücklegen nacheinander* zwei Kugeln gezogen. Gesucht sind:

- die Wahrscheinlichkeit, in beiden Zügen eine rote Kugel zu ziehen
- die Wahrscheinlichkeit im zweiten Zug eine rote Kugel zu ziehen, *unter der Voraussetzung*, dass im ersten Zug bereits eine rote Kugel gezogen wurde

Modellierung des Problems:

$A =$ „Im ersten Zug wird eine rote Kugel gezogen.“

¹Das Beispiel in Beispielen 8.1, 8.4 und 8.7 und dessen Darstellung mit den Rechecktabellen und dem Baumdiagramm wurde von einem ähnlichen Beispiel in [B12, Kapitel 6] inspiriert.

$B =$ „Im zweiten Zug wird eine rote Kugel gezogen.“

R_1, R_2, R_3, R_4 seien die roten Kugeln; S_1, S_2 seien die schwarzen Kugeln.

Wir notieren die *Elementarereignisse* als

(im ersten Zug gezogene Kugel, im zweiten Zug gezogene Kugel).

Wir haben hier ein *zweistufiges Zufallsexperiment*, denn es wird zweimal gezogen. Ein solches kann man in einer „Rechteck-Tabelle“ darstellen:

2. Zug einer Kugel

		R_1	R_2	R_3	R_4	S_1	S_2
1. Zug einer Kugel	R_1		(R_1, R_2)	(R_1, R_3)	(R_1, R_4)	(R_1, S_1)	(R_1, S_2)
	R_2	(R_2, R_1)		(R_2, R_3)	(R_2, R_4)	(R_2, S_1)	(R_2, S_2)
	R_3	(R_3, R_1)	(R_3, R_2)		(R_3, R_4)	(R_3, S_1)	(R_3, S_2)
	R_4	(R_4, R_1)	(R_4, R_2)	(R_4, R_3)		(R_4, S_1)	(R_4, S_2)
	S_1	(S_1, R_1)	(S_1, R_2)	(S_1, R_3)	(S_1, R_4)		(S_1, S_2)
	S_2	(S_2, R_1)	(S_2, R_2)	(S_2, R_3)	(S_2, R_4)	(S_2, S_1)	

Jedes Feld stellt ein Elementarereignis dar. Die „Diagonale“ bleibt dabei frei, denn es wird *ohne Zurücklegen* gezogen (d.h. $(R_1, R_1), \dots, (R_4, R_4), (S_1, S_1), (S_2, S_2)$ können nicht auftreten).

Alle Elementarereignisse sind *gleichwahrscheinlich*, weil in jedem Zug alle noch in der Urne befindlichen Kugeln mit der gleichen Wahrscheinlichkeit gezogen werden! Daher können wir die Wahrscheinlichkeitsdefinition von Laplace anwenden.

Wir haben $h(S) = 6^2 - 6 = 36 - 6 = 30$ gleichwahrscheinliche Elementarereignisse.

Das Ereignis A entspricht den ersten vier Zeilen der Tabelle, d.h. es liegen $h(A) = 4 \cdot 6 - 4 = 20$ für A günstige Elementarereignisse vor. Die entsprechenden Einträge sind in der obersten Kopie der Tabelle in Abbildung 8.1 blau gekennzeichnet.

Das Ereignis B entspricht den ersten vier Spalten der Tabelle, d.h. es liegen $h(B) = 6 \cdot 4 - 4 = 20$ für B günstige Elementarereignisse vor. Die entsprechenden Einträge sind in der mittleren Kopie der Tabelle in Abbildung 8.1 violett gekennzeichnet.

Das Ereignis, dass im ersten und im zweiten Zug jeweils eine rote Kugel gezogen wird, ist $A \cap B$ und erhält daher alle Elementarereignisse, die zu A und B gehören. Wir haben $h(A \cap B) = 4 \cdot 4 - 4 = 12$ für $A \cap B$ günstige Elementarereignisse. Die entsprechenden Einträge sind in der untersten Kopie der Tabelle in Abbildung 8.1 rot gekennzeichnet.

2. Zug einer Kugel

	R_1	R_2	R_3	R_4	S_1	S_2
R_1		(R_1, R_2)	(R_1, R_3)	(R_1, R_4)	(R_1, S_1)	(R_1, S_2)
R_2	(R_2, R_1)		(R_2, R_3)	(R_2, R_4)	(R_2, S_1)	(R_2, S_2)
R_3	(R_3, R_1)	(R_3, R_2)		(R_3, R_4)	(R_3, S_1)	(R_3, S_2)
R_4	(R_4, R_1)	(R_4, R_2)	(R_4, R_3)		(R_4, S_1)	(R_4, S_2)
S_1	(S_1, R_1)	(S_1, R_2)	(S_1, R_3)	(S_1, R_4)		(S_1, S_2)
S_2	(S_2, R_1)	(S_2, R_2)	(S_2, R_3)	(S_2, R_4)	(S_2, S_1)	

2. Zug einer Kugel

	R_1	R_2	R_3	R_4	S_1	S_2
R_1		(R_1, R_2)	(R_1, R_3)	(R_1, R_4)	(R_1, S_1)	(R_1, S_2)
R_2	(R_2, R_1)		(R_2, R_3)	(R_2, R_4)	(R_2, S_1)	(R_2, S_2)
R_3	(R_3, R_1)	(R_3, R_2)		(R_3, R_4)	(R_3, S_1)	(R_3, S_2)
R_4	(R_4, R_1)	(R_4, R_2)	(R_4, R_3)		(R_4, S_1)	(R_4, S_2)
S_1	(S_1, R_1)	(S_1, R_2)	(S_1, R_3)	(S_1, R_4)		(S_1, S_2)
S_2	(S_2, R_1)	(S_2, R_2)	(S_2, R_3)	(S_2, R_4)	(S_2, S_1)	

2. Zug einer Kugel

	R_1	R_2	R_3	R_4	S_1	S_2
R_1		(R_1, R_2)	(R_1, R_3)	(R_1, R_4)	(R_1, S_1)	(R_1, S_2)
R_2	(R_2, R_1)		(R_2, R_3)	(R_2, R_4)	(R_2, S_1)	(R_2, S_2)
R_3	(R_3, R_1)	(R_3, R_2)		(R_3, R_4)	(R_3, S_1)	(R_3, S_2)
R_4	(R_4, R_1)	(R_4, R_2)	(R_4, R_3)		(R_4, S_1)	(R_4, S_2)
S_1	(S_1, R_1)	(S_1, R_2)	(S_1, R_3)	(S_1, R_4)		(S_1, S_2)
S_2	(S_2, R_1)	(S_2, R_2)	(S_2, R_3)	(S_2, R_4)	(S_2, S_1)	

Abbildung 8.1: „Recheck-Tabelle“ der Elementarereignisse in Beispiel 8.1.

Wir finden somit gemäß der Wahrscheinlichkeitsdefinition von Laplace:

$$\begin{aligned} W(A) &= \frac{h(A)}{h(S)} = \frac{20}{30} = \frac{2}{3}, \\ W(B) &= \frac{h(B)}{h(S)} = \frac{20}{30} = \frac{2}{3}, \\ W(A \cap B) &= \frac{h(A \cap B)}{h(S)} = \frac{12}{30} = \frac{2}{5}. \end{aligned}$$

Die Wahrscheinlichkeit, in beiden Zügen eine rote Kugel zu ziehen, ist also $W(A \cap B) = 2/5$.

Nimmt man an, dass im ersten Zug eine rote Kugel gezogen wurde (Ereignis A), so befinden sich beim zweiten Zug nur noch $h(A) = 20$ Elementarereignisse in dem neuen Ereignisraum.

Die *bedingte Wahrscheinlichkeit* für Ereignis B (zweite Kugel ist rot) unter der *Bedingung*, dass Ereignis A (erste Kugel ist rot) eingetreten ist, ist also

$$W(B|A) = \frac{h(A \cap B)}{h(A)} = \frac{h(A \cap B)/h(S)}{h(A)/h(S)} = \frac{W(A \cap B)}{W(A)} = \frac{12}{20} = \frac{3}{5}.$$

Wir halten den neuen Begriff einer bedingten Wahrscheinlichkeit als Definition fest.

Definition 8.2. (bedingte Wahrscheinlichkeiten)

Seien A und B Ereignisse bei dem gleichen Zufallsexperiment.

- (1) Die **bedingte Wahrscheinlichkeit** des Ereignisses B **unter der Bedingung** A mit $W(A) > 0$ ist

$$W(B|A) = \frac{W(A \cap B)}{W(A)}.$$

- (2) Die **bedingte Wahrscheinlichkeit** des Ereignisses A **unter der Bedingung** B mit $W(B) > 0$ ist

$$W(A|B) = \frac{W(A \cap B)}{W(B)}.$$

In unserem motivierenden Beispiel haben wir bereits ein Beispiel für bedingte Wahrscheinlichkeiten gesehen.

Zum Verständnis von Definition 8.2 ist es hilfreich sich die **Analogie von bedingten Wahrscheinlichkeiten und bedingten relativen Häufigkeiten** klar zu machen: Dazu schreiben wir die Definitionen mit Erklärung untereinander und kennzeichnen die entsprechenden Objekte farbig:

$$W(B|A) = \frac{W(A \cap B)}{W(A)} = \left(\begin{array}{l} \text{bedingte Wahrscheinlichkeit des Ereignisses } B \\ \text{unter der Bedingung, dass } A \text{ eingetreten ist} \end{array} \right)$$

$$f_{i,j}^{(i)} = \frac{f_{i,j}}{f_{i,\cdot}} = \left(\begin{array}{l} \text{bedingte relative Häufigkeit, mit der der Merkmalswert } y_j \text{ bei } Y \\ \text{auftritt unter der Annahme, dass } X \text{ den Merkmalswert } x_i \text{ hat} \end{array} \right)$$

$$W(A|B) = \frac{W(A \cap B)}{W(B)} = \left(\begin{array}{l} \text{bedingte Wahrscheinlichkeit des Ereignisses } A \\ \text{unter der Bedingung, dass } B \text{ eingetreten ist} \end{array} \right)$$

$$f_{i,j}^{(j)} = \frac{f_{i,j}}{f_{\cdot,j}} = \left(\begin{array}{l} \text{bedingte relative Häufigkeit, mit der der Merkmalswert } x_i \text{ bei } X \\ \text{auftritt unter der Annahme, dass } Y \text{ den Merkmalswert } y_j \text{ hat} \end{array} \right)$$

Als Nächstes lernen wir stochastisch unabhängige Ereignisse kennen.

Definition 8.3. (stochastisch unabhängige bzw. abhängige Ereignisse)

Seien A und B zwei Ereignisse in einem Zufallsexperiment mit $W(A) \neq 0$ und $W(B) \neq 0$.

- (1) Die Ereignisse A und B heißen **stochastisch unabhängig** (voneinander), wenn gilt $W(A|B) = W(A)$ und $W(B|A) = W(B)$.
- (2) Gilt dagegen $W(A|B) \neq W(A)$ oder $W(B|A) \neq W(B)$, so sind die Ereignisse A und B **stochastisch abhängig**.

Anschauung: Stochastische Unabhängigkeit bedeutet, dass das Eintreten (oder Nicht-Eintreten) des einen Ereignisses das Eintreten des anderen Ereignisses nicht beeinflusst. Also muss gelten: Wahrscheinlichkeit = bedingter Wahrscheinlichkeit.

Betrachten wir zwei Beispiele.

Beispiel 8.4. (Kugeln werden ohne Zurücklegen aus einer Urne gezogen)

Aus einer Urne mit 6 Kugeln, von denen 2 schwarz und 4 rot sind, werden *ohne Zurücklegen nacheinander* zwei Kugeln gezogen. Wir haben dieses Zufallsexperiment bereits in Beispiel 8.1 betrachtet und für die Ereignisse

A = „Ziehen einer roten Kugel im ersten Zug“

B = „Ziehen einer roten Kugel im zweiten Zug“

die Wahrscheinlichkeiten $W(A)$ und $W(B)$ und die bedingte Wahrscheinlichkeit $W(B|A)$ berechnet: Wir finden

$$\frac{2}{3} = W(B) \neq W(B|A) = \frac{3}{5}.$$

Also sind das Ziehen einer roten Kugel im ersten Zug (Ereignis A) und das Ziehen einer roten Kugel im zweiten Zug (Ereignis B) *stochastisch abhängig*. (Dieses ist intuitiv klar, denn beim Ziehen einer roten Kugel im ersten Zug ändert sich die Anzahl der roten Kugeln in der Schachtel, da *ohne Zurücklegen* gezogen wird.)

Beispiel 8.5. (Ziehen von Kugeln mit Zurücklegen)

Aus einer Urne mit 6 Kugeln, von denen 2 schwarz und 4 rot sind, werden *mit Zurücklegen nacheinander* zwei Kugeln gezogen. Wir betrachten wieder die folgenden beiden Ereignisse:

A = „Ziehen einer roten Kugel im ersten Zug“

$B =$ „Ziehen einer roten Kugel im zweiten Zug“

Nun sind A und B *stochastisch unabhängig*, denn nach dem ersten Zug wird die gezogene Kugel ja zurückgelegt. Also kann das Ergebnis des ersten Zuges den zweiten Zug nicht beeinflussen. Diese plausible Tatsache weisen wir auf einem Übungszettel mit der Definition von stochastisch unabhängigen Ereignissen nach.

Auch hier besteht eine **Analogie zwischen stochastisch unabhängigen Ereignissen und unabhängig verteilten Merkmalen**: Die Bedingungen

$$W(A|B) = W(A) \quad \text{und} \quad W(B|A) = W(B)$$

sind analog zu den Bedingungen (siehe auch die farbige Kennzeichnung)

$$f_{i,(j)} = f_{i,\cdot} \quad \text{für } i = 1, \dots, k; j = 1, \dots, \ell \quad \text{und} \quad f_{(i),j} = f_{\cdot,j} \quad \text{für } i = 1, \dots, k; j = 1, \dots, \ell.$$

Aus der Definition der bedingten Wahrscheinlichkeit

$$W(B|A) = \frac{W(A \cap B)}{W(A)} \quad \text{bzw.} \quad W(A|B) = \frac{W(A \cap B)}{W(B)}$$

folgt durch Multiplikation mit $W(A)$ bzw. $W(B)$ der Multiplikationssatz für Wahrscheinlichkeiten:

$$\begin{aligned} W(B|A) = \frac{W(A \cap B)}{W(A)} & \Big| \cdot W(A) & \implies & W(B|A) \cdot W(A) = W(A \cap B), \\ W(A|B) = \frac{W(A \cap B)}{W(B)} & \Big| \cdot W(B) & \implies & W(A|B) \cdot W(B) = W(A \cap B). \end{aligned}$$

Satz 8.6. (Multiplikationssatz für Wahrscheinlichkeiten)

Seien A und B Ereignisse in einem Zufallsexperiment. Die Wahrscheinlichkeit $W(A \cap B)$ dafür, dass sowohl das Ereignis A als auch das Ereignis B eintritt, kann mit dem **Multiplikationssatz (für Wahrscheinlichkeiten)** berechnet werden:

$$W(A \cap B) = W(A) \cdot W(B|A) \quad \text{bzw.} \quad W(A \cap B) = W(B) \cdot W(A|B).$$

Sonderfall von Satz 8.6: Sind A und B **stochastisch unabhängig**, also wenn gilt $W(A|B) = W(A)$ und $W(B|A) = W(B)$, so vereinfacht sich der Multiplikationssatz zu

$$W(A \cap B) = W(A) \cdot W(B).$$

Mit dem Multiplikationssatz kann man die Wahrscheinlichkeiten in einem zweistufigen Zufallsexperiment gut mit einem Baumdiagramm berechnen. Wir zeigen dieses für unser Beispiel „Ziehen zweier Kugeln ohne Zurücklegen“ (siehe Beispiel 8.1).

Beispiel 8.7. (Baumdiagramm: Kugeln werden ohne Zurücklegen aus einer Urne gezogen)

Aus einer Urne mit 6 Kugeln, von denen 2 schwarz und 4 rot sind, werden *ohne Zurücklegen nacheinander* zwei Kugeln gezogen. Wir betrachten die Ereignisse

A = „Im ersten Zug wird eine rote Kugel gezogen.“

B = „Im zweiten Zug wird eine rote Kugel gezogen.“

\bar{A} = „Im ersten Zug wird eine schwarze Kugel gezogen.“

\bar{B} = „Im zweiten Zug wird eine schwarze Kugel gezogen.“

Überlegungen zum Ausfüllen des Wahrscheinlichkeitsbaums:

- Am Anfang sind 6 Kugeln in der Urne; davon sind 4 rot und 2 schwarz.
- Wurde im ersten Zug eine rote Kugel gezogen (Ereignis A), so sind noch 5 Kugeln in der Schachtel. Davon sind 3 rot und 2 schwarz.
- Wurde im ersten Zug eine schwarze Kugel gezogen (Ereignis \bar{A}), so sind noch 5 Kugeln in der Schachtel. Davon sind 4 rot und 1 schwarz.

Das Baumdiagramm ist in Abbildung 8.2 gezeichnet.

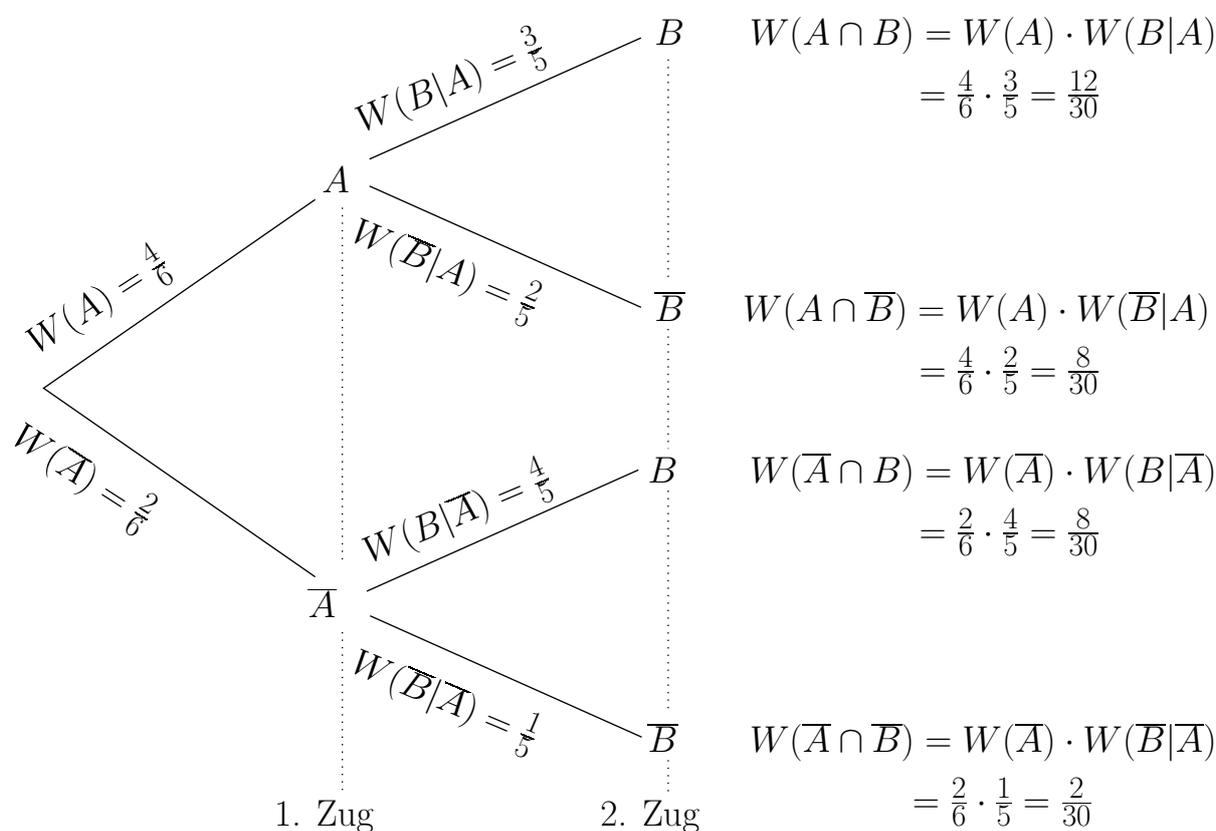


Abbildung 8.2: Baumdiagramm für das Ziehen zweier Kugeln aus einer Urne mit 4 roten und 2 schwarzen Kugeln ohne Zurücklegen: Längs jedes „kompletten“ Zweiges nutzt man den Multiplikationssatz.

Wir lernen nun zwei weitere Charakterisierungen von stochastischer Unabhängigkeit kennen. Zur Erinnerung (vgl. Definition 8.3): Per Definition sind zwei Ereignisse A und B (eines Zufallsexperiments) **stochastisch unabhängig**, wenn gilt:

$$W(A|B) = W(A) \quad \text{und} \quad W(B|A) = W(B).$$

Satz 8.8. (Alternative Charakterisierung 1 von „stochastisch unabhängig“)

Seien A und B zwei Ereignisse eines Zufallsexperiments mit $W(A) \neq 0$, $W(\bar{A}) \neq 0$, $W(B) \neq 0$ und $W(\bar{B}) \neq 0$. Die Ereignisse A und B sind genau dann **stochastisch unabhängig**, wenn gilt

$$W(A|B) = W(A|\bar{B}) \quad \text{und} \quad W(B|A) = W(B|\bar{A}).$$

Satz 8.9. (Alternative Charakterisierung 2 von „stochastisch unabhängig“)

Zwei Ereignisse A und B eines Zufallsexperiments mit $W(A) \neq 0$ und $W(B) \neq 0$ sind genau dann **stochastisch unabhängig**, wenn gilt

$$W(A \cap B) = W(A) \cdot W(B). \quad (8.1)$$

Beachten Sie die Analogie von (8.1) in Satz 8.9 zu der alternativen Charakterisierung unabhängig verteilter Merkmale (siehe Satz 5.11):

$$f_{i,j} = f_{i,\cdot} \cdot f_{\cdot,j} \quad \text{für alle } i = 1, 2, \dots, k \text{ und alle } j = 1, 2, \dots, \ell.$$

Beachten Sie in Sätzen 8.8 und 8.9, dass es sich um „genau dann, wenn“-Aussagen (also sogenannte Äquivalenzen) handelt. Dieses bedeutet, dass die Aussage zwei Folgerungsrichtungen hat. So muss man beim Beweis von Satz 8.8 zeigen, dass die folgenden beiden Aussagen gelten:

„ \Rightarrow “: Sind A und B stochastisch unabhängig, so gilt $W(A|B) = W(A|\bar{B})$ und $W(B|A) = W(B|\bar{A})$.

„ \Leftarrow “: Gilt $W(A|B) = W(A|\bar{B})$ und $W(B|A) = W(B|\bar{A})$, so sind A und B stochastisch unabhängig.

Satz 8.9 beweisen wir auf einem Übungszettel.

Satz 8.8 ist etwas schwieriger zu beweisen, und wir führen den Beweis nun durch.

Beweis von Satz 8.8:

Als Vorbereitung für den Beweis zeigen wir, dass die folgenden Aussagen für beliebige Ereignisse A und B gelten:

$$W(A) = W(A \cap B) + W(A \cap \bar{B}), \quad (8.2)$$

$$W(B) = W(B \cap A) + W(B \cap \bar{A}). \quad (8.3)$$

Wir brauchen nur eine der beiden Aussagen zu zeigen – die andere folgt dann durch Umbenennung von A in B und umgekehrt.

Beweis von (8.2) und (8.3): Es gelten $B \cap \bar{B} = \emptyset$ und $B \cup \bar{B} = S$. Daraus folgt

$$A = A \cap S = A \cap (B \cup \bar{B}) = (A \cap B) \cup (A \cap \bar{B})$$

und

$$(A \cap B) \cap (A \cap \bar{B}) \subseteq B \cap \bar{B} = \emptyset.$$

Nach dem Axiom 3 einer Wahrscheinlichkeit gilt somit

$$W(A) = W((A \cap B) \cup (A \cap \bar{B})) = W(A \cap B) + W(A \cap \bar{B}),$$

und wir haben (8.2) gezeigt. \square

Beweis von „ \Leftarrow “: Es gelte $W(A|B) = W(A|\bar{B})$ und $W(B|A) = W(B|\bar{A})$. Nach der Definition der bedingten Wahrscheinlichkeiten gilt

$$W(A|B) = W(A|\bar{B}) \quad \Longleftrightarrow \quad \frac{W(A \cap B)}{W(B)} = \frac{W(A \cap \bar{B})}{W(\bar{B})},$$

und wir multiplizieren auf beiden Seiten mit $W(B)$ und $W(\bar{B})$:

$$W(A \cap B) \cdot W(\bar{B}) = W(A \cap \bar{B}) \cdot W(B).$$

Nun ersetzen wir $W(\bar{B}) = 1 - W(B)$ und erhalten

$$\begin{aligned} & W(A \cap B) \cdot [1 - W(B)] = W(A \cap \bar{B}) \cdot W(B) \\ \Longleftrightarrow & W(A \cap B) - W(A \cap B) \cdot W(B) = W(A \cap \bar{B}) \cdot W(B) \\ \Longleftrightarrow & W(A \cap B) = W(A \cap B) \cdot W(B) + W(A \cap \bar{B}) \cdot W(B) \\ \Longleftrightarrow & W(A \cap B) = \underbrace{[W(A \cap B) + W(A \cap \bar{B})]}_{= W(A \cap S) = W(A)} \cdot W(B) \\ \Longleftrightarrow & W(A \cap B) = W(A) \cdot W(B), \end{aligned} \tag{8.4}$$

wobei wir im letzten Schritt (8.2) genutzt haben. Nun teilen wir in (8.4) durch $W(B)$ bzw. $W(A)$ und erhalten

$$\begin{aligned} \underbrace{\frac{W(A \cap B)}{W(B)}}_{= W(A|B)} = W(A) & \quad \Longleftrightarrow \quad W(A|B) = W(A), \\ \underbrace{\frac{W(A \cap B)}{W(A)}}_{= W(B|A)} = W(B) & \quad \Longleftrightarrow \quad W(B|A) = W(B), \end{aligned}$$

d.h. die Ereignisse A und B sind stochastisch unabhängig. \square

Beweis von „ \Leftarrow “: Seien A und B stochastisch unabhängig, d.h. es gelte $W(A) = W(A|B)$ und $W(B) = W(B|A)$. Nach der Definition der bedingten Wahrscheinlichkeit $W(A|B)$ folgt aus $W(A) = W(A|B)$, dass

$$W(A) = \frac{W(A \cap B)}{W(B)} \quad \Longleftrightarrow \quad W(A) \cdot W(B) = W(A \cap B).$$

Mittels Ersetzen von $W(B)$ durch (8.3) folgt

$$W(A) \cdot [W(B \cap A) + W(B \cap \bar{A})] = W(A \cap B)$$

$$\Leftrightarrow W(A) \cdot W(B \cap A) + W(A) \cdot W(B \cap \bar{A}) = W(A \cap B)$$

$$\Leftrightarrow W(A) \cdot W(B \cap \bar{A}) = W(A \cap B) - W(A) \cdot W(B \cap A)$$

$$\Leftrightarrow W(A) \cdot W(B \cap \bar{A}) = \underbrace{[1 - W(A)]}_{=W(\bar{A})} \cdot W(B \cap A)$$

$$\Leftrightarrow W(A) \cdot W(B \cap \bar{A}) = W(\bar{A}) \cdot W(B \cap A)$$

$$\Leftrightarrow \frac{W(B \cap \bar{A})}{W(\bar{A})} = \frac{W(B \cap A)}{W(A)}$$

$$\Leftrightarrow W(B|\bar{A}) = W(B|A).$$

Analog kann man herleiten, dass $W(A|\bar{B}) = W(A|B)$ gilt. □

8.2 Theorem der totalen Wahrscheinlichkeit

Wir leiten das sogenannte **Theorem der totalen Wahrscheinlichkeit** her:

A_1, A_2, \dots, A_n seien sich gegenseitig ausschließende Ereignisse, die den Ereignisraum S ganz ausfüllen, d.h.

$$A_i \cap A_j = \emptyset \quad \text{für alle } i \neq j \\ \text{mit } i, j = 1, 2, \dots, n,$$

$$\text{und } A_1 \cup A_2 \cup \dots \cup A_n = S.$$

Wir nennen A_1, A_2, \dots, A_n dann auch eine „**Einteilung**“ (oder „**Partition**“) des Ereignisraums S .

Eine Einteilung (oder Partition) von S ist in der Skizze rechts oben veranschaulicht. Dabei stellt der Inhalt des Rechtecks den ganzen Ereignisraum S dar.

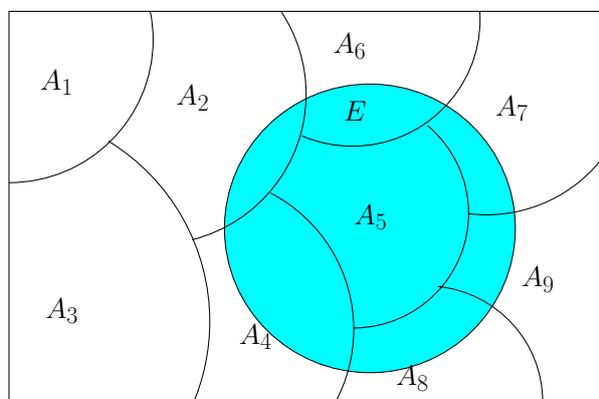
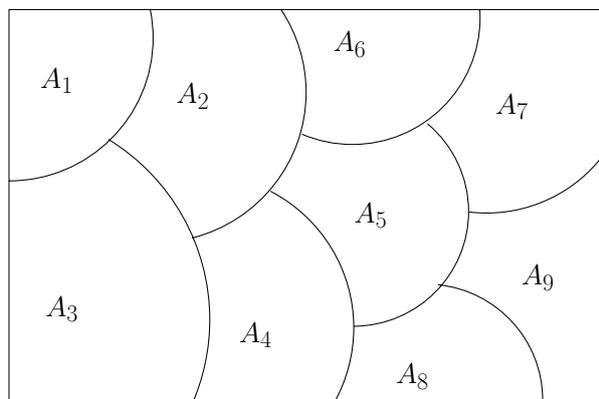
Für ein beliebiges Ereignis E gilt nun:

$$E = (E \cap A_1) \cup (E \cap A_2) \cup \dots \cup (E \cap A_n).$$

(Siehe die Skizze rechts zur Veranschaulichung.)

Es gilt für alle $i, j = 1, 2, \dots, n$

$$(E \cap A_i) \cap (E \cap A_j) \subseteq A_i \cap A_j = \emptyset \quad \text{wenn } i \neq j,$$



d.h. die Ereignisse $E \cap A_i$, $i = 1, 2, \dots, n$, schließen sich gegenseitig aus. Nach dem Additionssatz für die sich gegenseitig ausschließenden Ereignisse $E \cap A_i$, $i = 1, 2, \dots, n$, folgt nun

$$W(E) = W(E \cap A_1) + W(E \cap A_2) + \dots + W(E \cap A_n). \quad (8.5)$$

Nach dem Multiplikationssatz gilt

$$W(E \cap A_i) = W(A_i) \cdot W(E|A_i) \quad \text{für } i = 1, 2, \dots, n. \quad (8.6)$$

Einsetzen von (8.6) in (8.5) liefert

$$W(E) = W(A_1) \cdot W(E|A_1) + W(A_2) \cdot W(E|A_2) + \dots + W(A_n) \cdot W(E|A_n).$$

und wir haben das Theorem der totalen Wahrscheinlichkeit hergeleitet.

Satz 8.10. (Theorem der totalen Wahrscheinlichkeit)

Seien A_1, A_2, \dots, A_n sich gegenseitig ausschließende Ereignisse, die den Ereignisraum S ganz ausfüllen (d.h. $A_i \cap A_j = \emptyset$ für $i, j = 1, \dots, n$ mit $i \neq j$ und $A_1 \cup A_2 \cup \dots \cup A_n = S$). Dann gilt für ein beliebiges Ereignis E

$$\begin{aligned} W(E) &= \sum_{i=1}^n W(A_i) \cdot W(E|A_i) \\ &= W(A_1) \cdot W(E|A_1) + W(A_2) \cdot W(E|A_2) + \dots + W(A_n) \cdot W(E|A_n). \end{aligned}$$

Betrachten wir ein Beispiel für die Anwendung des Theorems der totalen Wahrscheinlichkeit.

Beispiel 8.11. (Qualitätskontrolle bei der Stoßstangenzulieferung)

Ein Automobilhersteller bezieht jeden Monat 1000 Stoßstangen über einer Zulieferer. Für diesen Zulieferer arbeiten drei Betriebe B_1 , B_2 und B_3 , in denen die Stoßstangen angefertigt werden. Die 1000 Stoßstangen teilen sich wie folgt auf die drei Betriebe auf:

- Betrieb B_1 produziert 300 Stück mit einem Ausschussanteil von 2%.
- Betrieb B_2 produziert 500 Stück mit einem Ausschussanteil von 4%.
- Betrieb B_3 produziert 200 Stück mit einem Ausschussanteil von 3%.

Aus der monatlichen Lieferung von 1000 Stoßstangen wird eine Stoßstange zufällig entnommen.

Frage: Wie groß ist die Wahrscheinlichkeit, dass die entnommene Stoßstange defekt ist?

Modellierung: Wir führen die Ereignisse A_1, A_2, A_3 und E ein:

E = „Die zufällig entnommene Stoßstange ist defekt.“

A_j = „Die zufällig ausgewählte Stoßstange wurde von Betrieb B_j produziert.“ (wobei $j = 1, 2, 3$)

Dann gilt: $A_1 \cup A_2 \cup A_3 = S$ und $A_1 \cap A_2 = A_2 \cap A_3 = A_1 \cap A_3 = \emptyset$

(Erklärung: Jede Stoßstange wurde von einem der drei Betriebe B_1, B_2, B_3 produziert, also $A_1 \cup A_2 \cup A_3 = S$. Eine Stoßstange kann immer nur von einem Betrieb produziert werden, also $A_1 \cap A_2 = A_2 \cap A_3 = A_1 \cap A_3 = \emptyset$.)

\implies Die Voraussetzungen des Theorems der totalen Wahrscheinlichkeit sind erfüllt.

Gesucht ist nun die Wahrscheinlichkeit $W(E)$.

Aus der Aufgabenstellung wissen wir:

$$W(A_1) = \frac{300}{1000} = \frac{3}{10}, \quad W(A_2) = \frac{500}{1000} = \frac{5}{10} = \frac{1}{2}, \quad W(A_3) = \frac{200}{1000} = \frac{2}{10} = \frac{1}{5},$$

$$W(E|A_1) = \frac{2}{100}, \quad W(E|A_2) = \frac{4}{100}, \quad W(E|A_3) = \frac{3}{100}.$$

Nach dem Theorem der totalen Wahrscheinlichkeit berechnen wir nun:

$$W(E) = W(A_1) \cdot W(E|A_1) + W(A_2) \cdot W(E|A_2) + W(A_3) \cdot W(E|A_3)$$

$$= \frac{3}{10} \cdot \frac{2}{100} + \frac{5}{10} \cdot \frac{4}{100} + \frac{2}{10} \cdot \frac{3}{100} = \frac{32}{1000} = 0,032.$$

Also beträgt die Wahrscheinlichkeit, dass die zufällig ausgewählte Stoßstange defekt ist, 0,032 (d.h. 3,2%).

8.3 Theorem von Bayes

Wir leiten das sogenannte **Theorems von Bayes** her:

A_1, A_2, \dots, A_n seien sich gegenseitig ausschließende Ereignisse, die den Ereignisraum S ganz ausfüllen, d.h.

$$A_i \cap A_j = \emptyset \text{ für alle } i, j = 1, 2, \dots, n \text{ mit } i \neq j \quad \text{und} \quad A_1 \cup A_2 \cup \dots \cup A_n = S.$$

Nach dem **Multiplikationssatz** gilt für jedes Ereignis E :

$$W(A_j \cap E) = W(E) \cdot W(A_j|E) \tag{8.7}$$

$$W(A_j \cap E) = W(E \cap A_j) = W(A_j) \cdot W(E|A_j) \tag{8.8}$$

Gleichsetzen von (8.7) und (8.8) ergibt

$$W(E) \cdot W(A_j|E) = W(A_j) \cdot W(E|A_j) \iff W(A_j|E) = \frac{W(A_j) \cdot W(E|A_j)}{W(E)}, \tag{8.9}$$

wobei wir die Umformung nur für $W(E) > 0$ durchführen können. Nach dem Theorem der totalen Wahrscheinlichkeit gilt

$$W(E) = \sum_{i=1}^n W(A_i) \cdot W(E|A_i). \tag{8.10}$$

Einsetzen von (8.10) in (8.9) liefert

$$W(A_j|E) = \frac{W(A_j) \cdot W(E|A_j)}{\sum_{i=1}^n W(A_i) \cdot W(E|A_i)},$$

und wir haben das Theorem von Bayes hergeleitet.

Satz 8.12. (Theorem von Bayes/Bayessche Regel)

Seien A_1, A_2, \dots, A_n sich gegenseitig ausschließende Ereignisse, die den Ereignisraum S ganz ausfüllen (d.h. $A_i \cap A_j = \emptyset$ für $i, j = 1, \dots, n$ mit $i \neq j$ und $A_1 \cup A_2 \cup \dots \cup A_n = S$). Dann gilt für ein beliebiges Ereignis E mit $W(E) > 0$

$$W(A_j|E) = \frac{W(A_j) \cdot W(E|A_j)}{\sum_{i=1}^n W(A_i) \cdot W(E|A_i)} \quad \text{für alle } j = 1, 2, \dots, n.$$

Verständnis/Interpretation des Theorem von Bayes: Ist ein Ereignis E eingetreten, so kann man mit dem Theorem von Bayes die Wahrscheinlichkeiten $W(A_j|E)$ berechnen, dass gleichzeitig mit E das Ereignis A_j eingetreten ist. Man nennt daher $W(A_j|E)$ auch die **a-posteriori-Wahrscheinlichkeit** und $W(A_j)$ die **a-priori-Wahrscheinlichkeit** (a-priori = im Voraus/vorher, a-posteriori = im Nachhinein/nachher).

Betrachten wir ein Beispiel zur Anwendung des Theorems von Bayes.

Beispiel 8.13. (Qualitätskontrolle bei der Stoßstangenzulieferung)

Ein Automobilhersteller bezieht jeden Monat 1000 Stoßstangen über einer Zulieferer. Für diesen Zulieferer arbeiten drei Betriebe B_1, B_2 und B_3 , in denen die Stoßstangen angefertigt werden. Die 1000 Stoßstangen teilen sich wie folgt auf die drei Betriebe auf:

- Betrieb B_1 produziert 300 Stück mit einem Ausschussanteil von 2%.
- Betrieb B_2 produziert 500 Stück mit einem Ausschussanteil von 4%.
- Betrieb B_3 produziert 200 Stück mit einem Ausschussanteil von 3%.

Aus der monatlichen Lieferung von 1000 Stoßstangen wird eine Stoßstange zufällig entnommen. Diese ist defekt.

Frage: Wie groß ist die Wahrscheinlichkeit, dass die defekte Stoßstange von Betrieb B_1, B_2 , bzw. B_3 gefertigt wurde?

Modellierung: Wir führen die Ereignisse A_1, A_2, A_3 und E ein:

E = „Die zufällig entnommene Stoßstange ist defekt.“

A_j = „Die zufällig ausgewählte Stoßstange wurde von Betrieb B_j produziert.“ (wobei $j = 1, 2, 3$)

Dann gilt: $A_1 \cup A_2 \cup A_3 = S$ und $A_1 \cap A_2 = A_2 \cap A_3 = A_1 \cap A_3 = \emptyset$

(Erklärung: Jede Stoßstange wurde von einem der drei Betriebe B_1, B_2, B_3 produziert, also $A_1 \cup A_2 \cup A_3 = S$. Eine Stoßstange kann immer nur von einem Betrieb produziert werden, also $A_1 \cap A_2 = A_2 \cap A_3 = A_1 \cap A_3 = \emptyset$.)

\implies Die Voraussetzungen des Theorems von Bayes sind erfüllt.

Gesucht sind nun die Wahrscheinlichkeiten $W(A_1|E)$, $W(A_2|E)$ bzw. $W(A_3|E)$.

Nach dem Theorem von Bayes gilt

$$\begin{aligned} W(A_j|E) &= \frac{W(A_j) \cdot W(E|A_j)}{W(A_1) \cdot W(E|A_1) + W(A_2) \cdot W(E|A_2) + W(A_3) \cdot W(E|A_3)} \\ &= \frac{W(A_j) \cdot W(E|A_j)}{W(E)}, \quad j = 1, 2, 3, \end{aligned} \quad (8.11)$$

wobei wir in der zweiten Zeile den Nenner mit dem Theorem der totalen Wahrscheinlichkeit vereinfacht haben.

$W(E)$ wurde bereits in Beispiel 8.11 mit dem Theorem der totalen Wahrscheinlichkeit berechnet:

$$W(E) = \sum_{i=1}^3 W(A_i) \cdot W(E|A_i) = \frac{32}{1000} = 0,032.$$

Aus der Aufgabenstellung wissen wir:

$$\begin{aligned} W(A_1) &= \frac{300}{1000} = \frac{3}{10}, & W(A_2) &= \frac{500}{1000} = \frac{5}{10} = \frac{1}{2}, & W(A_3) &= \frac{200}{1000} = \frac{2}{10} = \frac{1}{5}, \\ W(E|A_1) &= \frac{2}{100}, & W(E|A_2) &= \frac{4}{100}, & W(E|A_3) &= \frac{3}{100}. \end{aligned}$$

Einsetzen in (8.11) liefert

$$\begin{aligned} W(A_1|E) &= \frac{W(A_1) \cdot W(E|A_1)}{W(E)} = \frac{\frac{3}{10} \cdot \frac{2}{100}}{\frac{32}{1000}} = \frac{6}{32} = \frac{3}{16} = 0,1875, \\ W(A_2|E) &= \frac{W(A_2) \cdot W(E|A_2)}{W(E)} = \frac{\frac{5}{10} \cdot \frac{4}{100}}{\frac{32}{1000}} = \frac{20}{32} = \frac{5}{8} = 0,625, \\ W(A_3|E) &= \frac{W(A_3) \cdot W(E|A_3)}{W(E)} = \frac{\frac{2}{10} \cdot \frac{3}{100}}{\frac{32}{1000}} = \frac{6}{32} = \frac{3}{16} = 0,1875. \end{aligned}$$

Mit einer Wahrscheinlichkeit von 0,1875 bzw. 0,625 bzw. 0,1875 (also ca. 19% bzw. 63% bzw. 19%) wurde die defekte Stoßstange von Betrieb B_1 bzw. B_2 bzw. B_3 hergestellt.

Kombinatorik

In diesem Kapitel lernen wir die Grundlagen der Kombinatorik kennen und wenden diese an, um Wahrscheinlichkeiten zu berechnen.

9.1 Kombinationen n -ter Ordnung

Wir betrachten das folgende **Problem**: Von N Objekten/Elementen e_1, e_2, \dots, e_N werden n **mit oder ohne Zurücklegen** (d.h. **mit oder ohne Wiederholung**) zufällig ausgewählt. Wir erhalten eine sogenannte **Kombination n -ter Ordnung**.

Frage: Wie viele Kombinationen n -ter Ordnung gibt es jeweils, wenn wir die **Anordnung/Reihenfolge der Objekte** berücksichtigen bzw. ignorieren?

Betrachten wir zunächst ein Beispiel.

Beispiel 9.1. (Zufällige Auswahl von 2 von insgesamt 3 Objekten¹)

Wir betrachten drei Objekte e_1, e_2, e_3 und ziehen nacheinander zufällig zwei dieser drei Objekte. Ein gutes Modell ist hier immer eine Urne, in der sich die Kugeln e_1, e_2, e_3 befinden. Es macht einen Unterschied, ob wir nach dem ersten Zug aus der Urne die gezogene Kugel zurücklegen. Wird diese zurückgelegt, so spricht man von *Ziehen „mit Wiederholung“*. Weiter spielt es eine Rolle, ob wir nach dem Zug der zweiten Kugel berücksichtigen, *in welcher Reihenfolge*, die beiden Kugel gezogen worden sind oder nicht. Man spricht von Ziehen *„mit Berücksichtigung der Anordnung“* bzw. *„ohne Berücksichtigung der Anordnung“*. Wir untersuchen nun für die vier möglichen Fälle für die „Kombinationen 2-ter Ordnung aus 3 Elementen“:

Fall 1: Auswahl mit Zurücklegen/Wiederholung mit Berücksichtigung der Reihenfolge/Anordnung

¹Dieses Beispiel orientiert sich an einem Beispiel in [B12, Kapitel 9].

(e_1, e_1)	(e_2, e_1)	(e_3, e_1)
(e_1, e_2)	(e_2, e_2)	(e_3, e_2)
(e_1, e_3)	(e_2, e_3)	(e_3, e_3)

Wir schreiben die Elemente systematisch auf, indem wir die von den Indices gebildeten Zahlen der Größe nach aufsteigen lassen (spaltenweise zu lesen).

Wir erhalten $9 = 3^2$ Kombinationen 2-ter Ordnung mit Wiederholung und mit Berücksichtigung der Anordnung.

Fall 2: Auswahl ohne Zurücklegen/Wiederholung mit Berücksichtigung der Reihenfolge/Anordnung

	(e_2, e_1)	(e_3, e_1)
(e_1, e_2)		(e_3, e_2)
(e_1, e_3)	(e_2, e_3)	

Gegenüber dem Fall 1 mit Zurücklegen/Wiederholung werden alle Kombinationen mit zwei gleichen Objekten gestrichen. Diese sind im Fall 1 blau markiert.

Wir erhalten $6 = \frac{3!}{(3-2)!}$ Kombinationen 2-ter Ordnung ohne Wiederholung und mit Berücksichtigung der Anordnung.

Fall 3: Auswahl mit Zurücklegen/Wiederholung ohne Berücksichtigung der Reihenfolge/Anordnung

(e_1, e_1)		
(e_1, e_2)	(e_2, e_2)	
(e_1, e_3)	(e_2, e_3)	(e_3, e_3)

Wir schreiben die Elemente systematisch auf, indem wir die von den Indices gebildeten Zahlen der Größe nach aufsteigen lassen. Dabei verlangen wir, dass die Indices nicht abfallen, also nur (e_i, e_j) mit $i \leq j$, denn dieses gewährleistet, dass wir alle Kombinationen nur einmal auflisten. Im Fall 1 sind die wegfallenden Kombination violett markiert.

Wir erhalten $6 = \frac{4!}{2!2!} = \frac{(3+2-1)!}{2!(3-1)!}$ Kombinationen 2-ter Ordnung mit Wiederholung und ohne Berücksichtigung der Anordnung.

Fall 4: Auswahl ohne Zurücklegen/Wiederholung ohne Berücksichtigung der Reihenfolge/Anordnung

(e_1, e_2)	
(e_1, e_3)	(e_2, e_3)

Gegenüber dem Fall 3 mit Zurücklegen/Wiederholung werden alle Kombinationen mit zwei gleichen Objekten gestrichen. Alle in Fall 1 farbig (also in blau oder violett markierten Kombinationen) entfallen.

Wir erhalten $3 = \frac{3!}{2!(3-2)!} = \binom{3}{2}$ Kombinationen 2-ter Ordnung ohne Wiederholung und ohne Berücksichtigung der Anordnung.

Was wir im vorigen Beispiel bereits gesehen haben, funktioniert allgemein: Um alle Kombinationen n -ter Ordnung (von N Objekten) **ohne Anordnung** zu bestimmen, ordnet man zweckmäßigerweise die Indizes/Nummern der Objekte (in jeder Kombination) aufsteigend.

Wir wollen gleich allgemein bestimmen, wie viele Kombinationen n -ter Ordnung von N

Objekten es in jedem der vier möglichen Fälle gibt. Vorher betrachten wir zunächst einige Beispiele, um uns klar zu machen, warum diese Frage interessant ist.

Beispiel 9.2. (Kombinationen n -ter Ordnung aus N Elementen)

- (a) Dreimaliges Werfen mit einem Würfel, wobei berücksichtigt wird, in welcher Reihenfolge die Augenzahlen gewürfelt werden.

Modell: $N = 6$, $n = 3$, mit Wiederholung und mit Berücksichtigung der Anordnung

- (b) Dreimaliges Werfen mit einem Würfel, wobei nicht berücksichtigt wird, in welcher Reihenfolge die Augenzahlen gewürfelt werden.

Modell: $N = 6$, $n = 3$, mit Wiederholung und ohne Berücksichtigung der Anordnung

- (c) Lottospiel „6 aus 49“, d.h. es werden aus den natürlichen Zahlen von 1 bis 49 sechs Zahlen (ohne Wiederholungen) ausgewählt. Es spielt keine Rolle, in welcher Reihenfolge diese Lottozahlen gezogen werden.

Modell: $N = 49$, $n = 6$, ohne Wiederholung und ohne Berücksichtigung der Anordnung

- (d) Ein Post-Kurier-Service im Großraum Sydney erhält gleichzeitig 5 neue Aufträge. Er hat gerade 10 freie Kurierere, von denen jeder maximal einen Auftrag abholen soll. Wie viele Möglichkeiten gibt es, die 5 neuen Aufträge auf die 10 Kurierere zu verteilen, so dass jeder maximal einen Auftrag erhält?

Modell: $N = 10$, $n = 5$, ohne Wiederholung und mit Berücksichtigung der Anordnung (Die Anordnung wird berücksichtigt, da es eine Rolle spielt, welcher Kurierere welchen Auftrag bekommt.)

Der folgende Satz hält fest, wie viele Kombinationen n -ter Ordnung es für jeden der vier möglichen Fälle gibt.

Satz 9.3. (Anzahl der Kombinationen n -ter Ordnung aus N Elementen)

Die **Anzahl der Kombinationen n -ter Ordnung aus insgesamt N Elementen** ist für die verschiedenen Fälle **mit/ohne Wiederholung** und **mit/ohne Berücksichtigung der Anordnung** in der folgenden Tabelle angegeben:

	<i>mit Berücksichtigung der Anordnung</i>	<i>ohne Berücksichtigung der Anordnung</i>
<i>ohne Wiederholung</i>	$\frac{N!}{(N-n)!}$	$\binom{N}{n} = \frac{N!}{n!(N-n)!}$
<i>mit Wiederholung</i>	N^n	$\binom{N+n-1}{n} = \frac{(N+n-1)!}{n!(N-1)!}$

In den Formeln für die Kombinationen n -ter Ordnung aus N Elementen kommen **Fakultäten** und **Binomialkoeffizienten** vor.

Definition 9.4. (Fakultät und Binomialkoeffizient)

(1) Für $n \in \mathbb{N}_0$ ist $n!$, genannt ***n-Fakultät***, definiert durch

$$0! = 1 \quad \text{und} \quad n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot n \quad \text{für } n \in \mathbb{N}.$$

(2) Für $n, N \in \mathbb{N}_0$ mit $0 \leq n \leq N$ ist der ***Binomialkoeffizient*** $\binom{N}{n}$ (sprich „*N über n*“) definiert durch

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Betrachten wir eine Beispiel.

Beispiel 9.5. (Fakultäten und Binomialkoeffizienten)

(a) Fakultäten: $1! = 1$, $2! = 1 \cdot 2 = 2$, $3! = 1 \cdot 2 \cdot 3 = 6$,

(b) Binomialkoeffizienten:

$$\binom{2}{0} = \frac{2!}{0!(2-0)!} = 1, \quad \binom{2}{1} = \frac{2!}{1!(2-1)!} = 2, \quad \binom{2}{2} = \frac{2!}{2!(2-2)!} = 1.$$

Wir bemerken:

$$(a+b)^2 = a^2 + 2ab + b^2 = \binom{2}{0}a^2 + \binom{2}{1}ab + \binom{2}{2}b^2 = \sum_{n=0}^2 \binom{2}{k} a^{2-n} b^n$$

Wir halten einige wichtige Eigenschaften der Binomialkoeffizienten fest.

Lemma 9.6. (Eigenschaften der Binomialkoeffizienten)

(1) Die Binomialkoeffizienten sind die Koeffizienten in der ***binomischen Formel***:

$$(a+b)^N = \sum_{n=0}^N \binom{N}{n} a^{N-n} b^n \quad \text{für alle } N \in \mathbb{N}_0.$$

(2) Für $n, N \in \mathbb{N}_0$ mit $0 \leq n \leq N$ gelten

$$\binom{N}{n} = \binom{N}{N-n},$$

$$\binom{N}{n} = \frac{(N-n+1) \cdot (N-n+2) \cdot \dots \cdot (N-1) \cdot N}{1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n}.$$

(3) Für $n, N \in \mathbb{N}_0$ mit $0 \leq n < N$ gilt

$$\binom{N}{n} + \binom{N}{n+1} = \binom{N+1}{n+1}.$$

Lemma 9.6 (1) sollte bekannt sein. Lemma 9.6 (2) und Lemma 9.6 (3) beweisen wir auf einem Übungszettel.

Wir kommen nun noch einmal auf Beispiel 9.2 zurück und bestimmen die Anzahl der Kombinationen n -ter Ordnung aus N Elementen mit Hilfe von Satz 9.3.

Beispiel 9.7. (Anzahl der Kombinationen n -ter Ordnung aus N Elementen)

- (a) Dreimaliges Werfen mit einem Würfel, wobei berücksichtigt wird, in welcher Reihenfolge die Augenzahlen gewürfelt werden.

Modell: $N = 6$, $n = 3$, mit Wiederholung und mit Berücksichtigung der Anordnung

Anzahl: $N^n = 6^3 = 216$ Kombinationen 3-ter Ordnung aus 6 Elementen mit Wiederholung und mit Berücksichtigung der Anordnung

- (b) Dreimaliges Werfen mit einem Würfel, wobei nicht berücksichtigt wird, in welcher Reihenfolge die Augenzahlen gewürfelt werden.

Modell: $N = 6$, $n = 3$, mit Wiederholung und ohne Berücksichtigung der Anordnung

$$\text{Anzahl: } \binom{N+n-1}{n} = \frac{(N+n-1)!}{n!(N-1)!} = \frac{(6+3-1)!}{3!(6-1)!} = \frac{8!}{3!5!} = 56,$$

also 56 Kombinationen 3-ter Ordnung aus 6 Elementen mit Wiederholung und ohne Berücksichtigung der Anordnung

- (c) Im Lottospiel „6 aus 49“ werden aus den natürlichen Zahlen von 1 bis 49 sechs Zahlen (ohne Wiederholungen) ausgewählt. Es spielt keine Rolle, in welcher Reihenfolge diese Lottozahlen gezogen werden.

Modell: $N = 49$, $n = 6$, ohne Wiederholung und ohne Berücksichtigung der Anordnung

$$\text{Anzahl: } \binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{49!}{6!(49-6)!} = \frac{49!}{6!43!} = \frac{44 \cdot 45 \cdot \dots \cdot 49}{720} = 13.983.816,$$

also fast 14 Millionen Kombinationen 6-ter Ordnung aus 49 Elementen ohne Wiederholung und ohne Berücksichtigung der Anordnung

Bei fast 14 Millionen Kombinationsmöglichkeiten ist die Chance, im Lottospiel „6 aus 49“ den Jackpot zu gewinnen, extrem gering!

- (d) Ein Post-Kurier-Service im Großraum Sydney erhält gleichzeitig 5 neue Aufträge. Er hat gerade 10 freie Kurier, von denen jeder maximal einen Auftrag abholen soll. Wie viele Möglichkeiten gibt es, die 5 neuen Aufträge auf die 10 Kurier zu verteilen, so dass jeder maximal einen Auftrag erhält?

Modell: $N = 10$, $n = 5$, ohne Wiederholung und mit Berücksichtigung der Anordnung

$$\text{Anzahl: } \frac{10!}{5!} = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 = 30.240$$

Es gibt also 30.240 Möglichkeiten, die Aufträge so zu verteilen, dass jeder der 10 Kurier maximal einen Auftrag erhält.

9.2 Herleitungen der Formeln für die Kombinationen n -ter Ordnung von N Elementen

Wir werden nun die Formeln in Satz 9.3 herleiten bzw. beweisen. Dabei schreiben wir eine Kombination n -ter Ordnung als n -Zeilenvektor der N Elemente e_1, e_1, \dots, e_N , also

$$(e_{j_1}, e_{j_2}, \dots, e_{j_n}) \quad \text{mit } j_1, j_2, \dots, j_n \in \{1, 2, \dots, N\}. \quad (9.1)$$

Als **gedankliches Modell** kann man sich gut das Ziehen von Kugeln aus einer Urne vorstellen: In der Urne befinden sich N Kugeln e_1, e_2, \dots, e_N , die mit den den Ziffern $1, 2, \dots, N$ beschriftet sind (also e_i ist mit der Ziffer i beschriftet). Um die Notation zu vereinfachen, identifizieren wir dabei in den Beweisen die Kugel e_i mit ihrer Nummer i und schreiben somit statt (9.1)

$$(j_1, j_2, \dots, j_n) \quad \text{mit } j_1, j_2, \dots, j_n \in \{1, 2, \dots, N\}. \quad (9.2)$$

Wir sollten uns darüber klar sein, dass in (9.2) nicht die Kugeln sondern die Nummern der Kugeln notiert sind. Da die Kugeln und deren Nummern einander aber ein-eindeutig zugeordnet sind, dürfen wir hier diese Vereinfachung vornehmen.

Beweis der Formel in Satz 9.3 für den Fall „mit Wiederholung und mit Berücksichtigung der Anordnung“: Da mit Zurücklegen (also mit Wiederholung) gezogen wird, gibt es jeweils N Möglichkeiten beim Ziehen der ersten, der zweiten, ... und der n -ten Kugel. Weiter erfolgt die Auflistung (9.2) in der Reihenfolge, in der die Kugeln gezogen werden, da mit Berücksichtigung der Reihenfolge/Anordnung gezogen wird. Also finden wir

$$\begin{array}{c} (j_1, j_2, \dots, j_n), \quad \text{mit } j_1, j_2, \dots, j_n \in \{1, 2, \dots, N\}. \\ \uparrow \uparrow \dots \uparrow \\ \text{jeweils } N \text{ Möglichkeiten} \\ j_1, j_2, \dots, j_n \text{ zu wählen} \end{array}$$

Es kann jeweils jede der N Kugeln gezogen werden. Also gilt

$$\text{Anzahl der } n\text{-ten Kombinationen} = \underbrace{N \cdot N \cdot \dots \cdot N}_{n\text{-mal}} = N^n,$$

und wir haben die gesuchte Formel hergeleitet. \square

Beweis der der Formel in Satz 9.3 für den Fall „ohne Wiederholung und mit Berücksichtigung der Anordnung“: Die Kugeln bzw. deren Nummern werden auch hier in der Reihenfolge aufgelistet, in der sie gezogen werden (da „mit Berücksichtigung der Anordnung“). Allerdings wird nicht zurückgelegt (da „ohne Wiederholung“), d.h. die Nummer jeder Kugel kann nur einmal in der Auflistung vorkommen. Wir betrachten also

$$(j_1, j_2, \dots, j_n)$$

mit $j_1, j_2, \dots, j_n \in \{1, 2, \dots, N\}$, wobei die $e_{j_1}, e_{j_2}, \dots, e_{j_n}$ alle verschieden sind (da ohne Wiederholung).

Es gibt N Möglichkeiten, j_1 in $\{1, 2, \dots, N\}$ zu wählen.

Es gibt $N - 1$ Möglichkeiten, j_2 in $\{1, 2, \dots, N\} \setminus \{j_1\}$ zu wählen.

Es gibt $N - 2$ Möglichkeiten, j_3 in $\{1, 2, \dots, N\} \setminus \{j_1, j_2\}$ zu wählen.

Wir setzen den Prozess fort und finden als Letztes:

Es gibt $N - (n - 1)$ Möglichkeiten j_n in $\{1, 2, \dots, N\} \setminus \{j_1, j_2, \dots, j_{n-1}\}$ zu wählen.

Also folgt

$$\begin{aligned} \text{Anzahl der Kombinationen } n\text{-ter Ordnung} &= N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot (N - (n - 1)) \\ &= \frac{N \cdot (N - 1) \cdot (N - 2) \cdot \dots \cdot (N - (n - 1)) \cdot \overbrace{(N - n) \cdot (N - (n + 1)) \cdot \dots \cdot 2 \cdot 1}^{= (N-n)!}}{(N - n)!} \\ &= \frac{N!}{(N - n)!}, \end{aligned}$$

und wir haben die gesuchte Formel hergeleitet. \square

Wollen wir den Fall „ohne Wiederholung und ohne Berücksichtigung der Anordnung“ betrachten, so müssen wir in der Anzahl im vorigen Fall „ohne Wiederholung und mit Berücksichtigung der Anordnung“ noch alle bis auf die Anordnung identischen n -ten Kombinationen identifizieren. Als Vorbereitung dafür betrachten wir die folgende Frage:

Wie viele Möglichkeiten gibt es n Kugeln e_1, e_2, \dots, e_n bzw. deren Nummern $1, 2, \dots, n$ unterschiedlich anzuordnen?

Betrachten wir als Beispiel zunächst 3 Elemente e_1, e_2, e_3 und ordnen deren Nummern auf alle möglichen Varianten an:

$$\left. \begin{array}{l} (1, 2, 3) \\ (1, 3, 2) \\ (2, 1, 3) \\ (2, 3, 1) \\ (3, 1, 2) \\ (3, 2, 1) \end{array} \right\} \begin{array}{l} 6 = 3! \text{ Anordnungen} \\ \text{(oder } \mathbf{Permutationen}) \\ \text{der Elemente } e_1, e_2, e_3 \end{array}$$

Nun betrachten wir den allgemeinen Fall:

$$(j_1, j_2, \dots, j_n), \quad \text{wobei } j_1, j_2, \dots, j_n \in \{1, 2, \dots, n\} \text{ alle verschieden sind.}$$

j_1 kann beliebig in $\{1, 2, \dots, n\}$ gewählt werden (n Möglichkeiten).

j_2 kann beliebig in $\{1, 2, \dots, n\} \setminus \{j_1\}$ gewählt werden ($n - 1$ Möglichkeiten).

Wir setzen den Prozess fort:

j_{n-1} kann beliebig in $\{1, 2, \dots, n\} \setminus \{j_1, j_2, \dots, j_{n-2}\}$ gewählt werden ($n - (n - 2) = 2$ Möglichkeiten).

j_n wird das noch übrige Element in $\{1, 2, \dots, n\} \setminus \{j_1, j_2, \dots, j_{n-2}, j_{n-1}\}$ (1 Möglichkeit).

Es gibt also

$$n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 = n!$$

Anordnungen, d.h. $n!$ **Permutationen von** $1, 2, \dots, n$ bzw. der Kugeln e_1, e_2, \dots, e_n .

Nach dieser Vorbereitung können wir nun leicht den Fall „ohne Wiederholung und ohne Berücksichtigung der Anordnung“ behandeln.

Beweis der Formel in Satz 9.3 für den Fall „ohne Wiederholung und ohne Berücksichtigung der Anordnung“: Für den Fall „ohne Wiederholung und mit Berücksichtigung der Anordnung“ hatten wir

$$\frac{N!}{(N - n)!} \quad \text{Kombinationen } n\text{-ter Ordnung.}$$

Jede Kombination n -ter Ordnung im Fall „ohne Wiederholung und ohne Berücksichtigung der Anordnung“ kann auf $n!$ Arten angeordnet werden

Teilen die also die Anzahl n -ter Kombinationen

$$\frac{N!}{(N - n)! n!}$$

im Fall „ohne Wiederholung und **mit** Berücksichtigung der Anordnung“ durch $n!$, so identifizieren wir alle bis auf die Anordnung identischen Kombinationen n -ter Ordnung und erhalten somit

$$\text{Anzahl der } n\text{-ten Kombinationen} = \frac{N!}{(N - n)! n!} = \binom{N}{n}.$$

Damit ist die gesuchte Formel hergeleitet. □

Nun beweisen wir die Formel für den Fall „mit Wiederholung und ohne Berücksichtigung der Anordnung“ mit einem cleveren Trick.

Beweis der Formel in Satz 9.3 für den Fall „mit Wiederholung und ohne Berücksichtigung der Anordnung“: Da ohne Berücksichtigung der Anordnung gezogen wird, können wir unsere n gezogenen Kugeln bzw. deren Nummern in aufsteigender Reihenfolge notieren:

$$(j_1, j_2, \dots, j_n) \quad \text{mit} \quad j_1 \leq j_2 \leq \dots \leq j_n \quad \text{und} \quad j_1, j_2, \dots, j_n \in \{1, 2, \dots, n\}.$$

Wir definieren nun eine Funktion T , die unsere Kombination n -ter Ordnung von N Kugeln mit Wiederholung (und ohne Berücksichtigung der Anordnung) auf Kombinationen n -ter Ordnung von $N + n - 1$ Kugeln ohne Wiederholung (und ohne Berücksichtigung der Anordnung) **bijektiv** abbildet:

$$T(j_1, j_2, j_3, \dots, j_n) = (\underbrace{j_1 + 0}_{=k_1}, \underbrace{j_2 + 1}_{=k_2}, \underbrace{j_3 + 2}_{=k_3}, \dots, \underbrace{j_n + (n-1)}_{=k_n}).$$

Dann gilt

$$\underbrace{j_1 + 0}_{=k_1}, \underbrace{j_2 + 1}_{=k_2}, \underbrace{j_3 + 2}_{=k_3}, \dots, \underbrace{j_n + (n-1)}_{=k_n} \in \{1, 2, \dots, N, N+1, \dots, N+n-1\}$$

und

$$\underbrace{j_1 + 0}_{=k_1} < \underbrace{j_2 + 1}_{=k_2} < \underbrace{j_3 + 2}_{=k_3} < \dots < \underbrace{j_n + (n-1)}_{=k_n},$$

d.h. es liegen Kombinationen n -ter Ordnung aus $N + n - 1$ Elementen ohne Wiederholung und ohne Berücksichtigung der Anordnung vor. Mit der zugehörigen Umkehrabbildung

$$T^{-1}(k_1, k_2, k_3, \dots, k_n) = (\underbrace{k_1 - 0}_{=j_1}, \underbrace{k_2 - 1}_{=j_2}, \underbrace{k_3 - 2}_{=j_3}, \dots, \underbrace{k_n - (n-1)}_{=j_n})$$

wird umgekehrt jede Kombinationen $(k_1, k_2, k_3, \dots, k_n)$ n -ter Ordnung aus $N + n - 1$ Elementen ohne Wiederholung (und ohne Berücksichtigung der Anordnung) auf eine Kombinationen n -ter Ordnung aus N Elementen mit Wiederholung (und ohne Berücksichtigung der Anordnung) abgebildet. Also ist T in der Tat eine Bijektion von der Menge aller Kombinationen n -ter Ordnung aus N Elementen mit Wiederholung und ohne Berücksichtigung der Anordnung auf die Menge aller Kombinationen n -ter Ordnung aus $N + n - 1$ Elementen ohne Wiederholung und ohne Berücksichtigung der Anordnung. Beide Mengen haben also die gleiche Anzahl von Elementen und wir wissen aus dem vorigen Fall bereit, dass

$$\binom{N+n-1}{n} = \frac{(N+n-1)!}{n!(N-1)!}$$

die Anzahl Kombinationen n -ter Ordnung aus $N + n - 1$ Elementen ohne Wiederholung und ohne Berücksichtigung der Anordnung ist. \square

9.3 Berechnung von Wahrscheinlichkeiten mit Kombinatorik

Wir erinnern uns zunächst an die **Definition der Wahrscheinlichkeit nach Laplace** (siehe Definition 7.21):

Wir betrachten ein Zufallsexperiment mit einem **endlichen** Ereignisraum S mit **gleichwahrscheinlichen** Elementarereignissen. Die **Wahrscheinlichkeit** $W(A)$, dass bei diesem Zufallsexperiment das Ereignis A eintritt, ist

$$W(A) = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl aller gleichmöglichen Fälle}} = \frac{h(A)}{h(S)},$$

wobei $h(A)$ bzw. $h(S)$ die absolute Häufigkeit (oder Anzahl) der Elementarereignisse in A bzw. S ist.

Wir können nun unsere neu erworbenen Kombinatorik-Kenntnisse nutzen, um mit Hilfe der Laplaceschen Wahrscheinlichkeitsdefinition in vielen spannenden und überraschenden Anwendungsproblemen die Wahrscheinlichkeiten zu berechnen.

Achtung: Die Laplacesche Wahrscheinlichkeitsdefinition setzt gleichwahrscheinliche Elementarereignisse voraus! Kombinationen **mit Berücksichtigung der Anordnung** sind **gleichwahrscheinlich!** Kombinationen **ohne Berücksichtigung der Anordnung** sind im allgemeinen aber **nicht gleichwahrscheinlich!** (Ausnahme: Im Fall „ohne Wiederholung und ohne Berücksichtigung der Anordnung“ sind die n -ten Kombinationen auch gleichwahrscheinlich, weil jede n -te Kombination ohne Wiederholung genau $n!$ mögliche verschiedene Anordnungen hat.)

Betrachten wir drei Beispiele.

Beispiel 9.8. (Pizza-Lieferservice)

Ein Pizza-Lieferservice hat 20 neue Aufträge bekommen, die er auf seine zwei Auslieferer verteilen möchte. Dabei soll jeder der beiden Auslieferer 10 Aufträge übernehmen.

Frage 1: Wie viele Möglichkeiten hat der Pizza-Lieferservice seine Aufträge auf die beiden Auslieferer zu verteilen? Dabei wird nicht berücksichtigt, in welcher Reihenfolge jeder Auslieferer seine 10 Aufträge abarbeitet, d.h. die Reihenfolge der Zuordnung der Aufträge wird nicht berücksichtigt.

Wir überlegen uns zunächst, dass wir uns nur über einen der beiden Auslieferer Gedanken machen müssen, denn hat Auslieferer 1 seine 10 Aufträge bekommen, so erhält Auslieferer 2 automatisch die restlichen 10 Aufträge.

Modell: $N = 20$ (Aufträge insgesamt), $n = 10$ (Aufträge für Auslieferer 1)

Fall: „ohne Wiederholung“ (jeder Auftrag kann nur einmal vergeben werden) und „ohne Berücksichtigung der Anordnung“ (es spielt keine Rolle, in welcher Reihenfolge die Aufträge abgearbeitet werden).

Also gibt es

$$\begin{aligned} \binom{N}{n} &= \binom{20}{10} = \frac{20!}{10!(20-10)!} = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \\ &= \frac{19 \cdot 17 \cdot 14 \cdot 13 \cdot 12 \cdot 11}{7 \cdot 6} = 19 \cdot 17 \cdot 13 \cdot 4 \cdot 11 = 184.756 \end{aligned}$$

Möglichkeiten die Aufträge auf die beiden Zulieferer zu verteilen.

Frage 2: Wir betrachten nun 10 fest vorgegebene Aufträge, die einer der beiden Auslieferer bekommen hat. Darunter befinden sich zwei unabhängige Aufträge von Studentinnen der Uni Paderborn. Wie groß ist die Wahrscheinlichkeit, dass die beiden Studentinnen zuerst (also als erste und zweite Person) beliefert werden? Da uns keine geographischen Informationen über die Zustellorte vorliegen, nehmen wir an, dass jede mögliche Zustellreihenfolge gleichwahrscheinlich ist.

Seien e_1, e_2, \dots, e_{10} die 10 Aufträge. Jede Anordnung ohne Wiederholung (aber mit Berücksichtigung der Reihenfolge)

$$(e_{j_1}, e_{j_2}, e_{j_3}, \dots, e_{j_{10}})$$

stellt eine Reihenfolge der Zustellung dar. Wir haben $10!$ solcher Anordnungen, denn: Wir können für e_{j_1} jeden der 10 Aufträge wählen, aber für e_{j_2} dann nur noch jeden der 9 übrigen Aufträge, für e_{j_3} dann nur noch jeden der 8 übrigen Aufträge, usw.. Also haben wir

$$10 \cdot 9 \cdot 8 \cdot \dots \cdot 2 \cdot 1 = 10!$$

Möglichkeiten für die Reihenfolge der Zustellung der 10 Aufträge. Alle Reihenfolgen sind gleichwahrscheinlich.

Seien nun die Aufträge der beiden Studentinnen e_1 und e_2 . Dann werden die beiden Studentinnen zuerst beliefert, wenn die folgenden Zustellreihenfolgen vorliegen:

$$\begin{aligned} & \underbrace{(e_1, e_2, e_{j_3}, e_{j_4}, \dots, e_{j_{10}})}_{\substack{8! \text{ mögliche} \\ \text{Anordnungen}}} \quad \text{mit } e_{j_3}, \dots, e_{j_{10}} \in \{e_3, \dots, e_{10}\}, \\ & \underbrace{(e_2, e_1, e_{j_3}, e_{j_4}, \dots, e_{j_{10}})}_{\substack{8! \text{ mögliche} \\ \text{Anordnungen}}} \quad \text{mit } e_{j_3}, \dots, e_{j_{10}} \in \{e_3, \dots, e_{10}\}, \end{aligned}$$

wobei man sich die $8!$ möglichen Anordnungen wie oben im Fall der Gesamtzahl der Zustellreihenfolgen überlegt. Insgesamt finden wir also $2 \cdot 8!$ günstige Zustellreihenfolgen für den Fall, dass die beiden Studentinnen zuerst beliefert werden. Daher ist die Wahrscheinlichkeit (nach der Wahrscheinlichkeitsdefinition von Laplace) für das Ereignis

$E =$ „Die beiden Studentinnen werden als erste und zweite beliefert.“

$$W(E) = \frac{2 \cdot 8!}{10!} = \frac{2}{9 \cdot 10} = \frac{1}{45} \approx 0,022,$$

also ca. 2%.

Beispiel 9.9. (Geburtstage im gleichen Monat)

Frage: Wie groß ist die Wahrscheinlichkeit, dass unter 5 zufällig ausgewählten Personen mindestens 2 Personen im gleichen Monat Geburtstag haben? Wir nehmen dabei an, dass sich die Geburtstage gleichmäßig über die Monate verteilen.

Wir listen die Geburtsmonate der 5 Personen in der Stichprobe als geordnetes 5-Tupel auf:

$$(x_1, x_2, x_3, x_4, x_5) \quad \text{mit } x_i = \text{Geburtsmonat der } i\text{-ten Person.}$$

Unser Ereignisraum ist dann

$$S = \left\{ (x_1, x_2, x_3, x_4, x_5) \mid x_i \in \{\text{Jan, Feb, Mär, Apr, Mai, Jun, Jul, Aug, Sep, Okt, Nov, Dez}\} \right\}$$

mit $h(S) = N^n = 12^5 = 248.832$ gleichwahrscheinlichen Elementarereignissen (Fall: $N = 12$ (Monate), $n = 5$ (Personen), mit Wiederholung und mit Berücksichtigung der Anordnung).

A = „Mindestens zwei der fünf Personen haben im gleichen Monat Geburtstag.“

\bar{A} = „Keine zwei (oder mehr) der fünf Personen haben im gleichen Monat Geburtstag.“
 = „Alle fünf Personen haben in verschiedenen Monaten Geburtstag.“

Die Anzahl der für \bar{A} günstigen Fälle ist

$$h(\bar{A}) = \frac{N!}{(N-n)!} = \frac{12!}{(12-5)!} = \frac{12!}{7!} = 95.040.$$

(Fall: $N = 12$ (Monate), $n = 5$ (Personen), ohne Wiederholung und mit Berücksichtigung der Anordnung)

Nach der Laplaceschen Wahrscheinlichkeitsdefinition finden wir

$$W(\bar{A}) = \frac{h(\bar{A})}{h(S)} = \frac{95040}{248832} = \frac{12!/7!}{12^5} \approx 0,38.$$

Es folgt

$$W(A) = 1 - W(\bar{A}) = 1 - \frac{12!/7!}{12^5} \approx 0,62.$$

Mit einer Wahrscheinlichkeit von 0,62 (also 62%) haben mindestens 2 der 5 Personen im gleichen Monat Geburtstag.

Beispiel 9.10. (Lottospiel „6 aus 49“)

Wie groß ist die Wahrscheinlichkeit W beim Lottospiel „6 aus 49“, mit einmal Tippen zu gewinnen?

In Beispiel 9.7 (c) haben wir bereits das Lottospiel „6 aus 49“ betrachtet und uns überlegt, dass es

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{49!}{6!(49-6)!} = \frac{49!}{6!43!} = \frac{44 \cdot 45 \cdot \dots \cdot 49}{720} = 13.983.816,$$

mögliche Kombinationen 6-ter Ordnung gibt (Modell: $N = 49$, $n = 6$, ohne Wiederholung und ohne Berücksichtigung der Anordnung). Da hier zwar ein Fall ohne Berücksichtigung der Anordnung aber auch ohne Wiederholung vorliegt, sind auch hier alle auftretenden Kombinationen 6-ter Ordnung gleichwahrscheinlich. Die Wahrscheinlichkeit, mit einmal Tippen zu gewinnen, ist also

$$W = \frac{1}{13.983.816} \approx 0,000000072.$$

Zufallsvariablen

Bisher haben wir nur Experimente und deren Elementarereignisse betrachtet. für „kompliziertere“ Experimente, wie zum Beispiel die Summe der Augenzahlen zweier Würfe mit einem klassischen Würfel, haben wir dazu einen neuen Ereignisraum eingeführt, dessen Elementarereignisse in dem Beispiel die möglichen Summen der Augenzahlen waren, also $S = \{2, 3, \dots, 12\}$, waren. Mit dem in diesem Kapitel neu eingeführten Begriff einer **Zufallsvariablen** können wir solche Experimente einfacher beschreiben, indem wir einfach den „natürlichen“ Ereignisraum des Experiments betrachten und auf diesem das Ergebnis des Experiments als Funktion auf dem Ereignisraum definieren. Diese Funktion ist dann eine Zufallsvariable. Für das Beispiel der Summe der Augenzahlen zweier Würfe ist der „natürliche“ Ereignisraum

$$S = \{(i, j) \mid i, j = 1, 2, 3, 4, 5, 6\},$$

und wir beschreiben die Zufallsvariable $X =$ „Summe der Augenzahlen der beiden Würfe“ durch die Funktion

$$X : S \rightarrow \{2, 3, \dots, 12\}, \quad X(i, j) = i + j.$$

Aufbauend auf den Begriff einer Zufallsvariablen können wir deren **Verteilung** untersuchen (wie verteilt sich die Wahrscheinlichkeit auf die Werte der Zufallsvariablen?) und analog zum arithmetischen Mittel und der Varianz bei Häufigkeitsverteilungen den Erwartungswert und die Varianz einer Zufallsvariablen einführen.

10.1 Einführung: Zufallsvariablen

Sind die Ergebnisse eines Zufallsexperiments **numerische Größen**, so arbeitet man mit **Zufallsvariablen**.

Definition 10.1. (Zufallsvariable)

- (1) Eine **Zufallsvariable** ist eine Funktion X , die den Elementen des Ereignisraums $S = \{e_1, e_2, \dots\}$ eindeutig Werte aus einer reellen Bildmenge $B \subseteq \mathbb{R}$ zuordnet:

$$X : S \rightarrow B \subseteq \mathbb{R}, \quad X(e_i) = a_i$$

- (2) X ist eine **diskrete Zufallsvariable**, wenn X nur **endlich viele verschiedene** oder **abzählbar unendlich viele verschiedene** Werte (oder Realisationen oder Ausprägungen) annimmt.
- (3) X ist eine **stetige** (oder **kontinuierliche**) **Zufallsvariable**, wenn X in einem Bereich der reellen Zahlen (z.B. in einem Intervall) **jeden beliebigen** Wert annehmen kann.

Betrachten wir einige Beispiele für Zufallsvariablen.

Beispiel 10.2. (Zufallsvariablen)

- (a) Augenzahl beim Würfeln mit einem perfekten Würfel: Dann ist der Ereignisraum

$$S = \{i \mid i = 1, 2, 3, 4, 5, 6\},$$

und die *diskrete Zufallsvariable* ist die gewürfelte Augenzahl

$$X : S \rightarrow \{1, 2, 3, 4, 5, 6\}, \quad X(i) = i \quad \text{für } i = 1, 2, 3, 4, 5, 6.$$

- (b) Summe der Augenzahlen beim Wurf mit zwei perfekten Würfeln: Dann ist der (natürliche) Ereignisraum

$$S = \{(i, j) \mid i = 1, 2, \dots, 6 \text{ und } j = 1, \dots, 6\},$$

und die Zufallsvariable $X =$ „Summe der Augenzahlen der beiden Würfe“

$$X : S \rightarrow \{2, 3, \dots, 12\}, \quad X(i, j) = i + j,$$

ist eine *diskrete Zufallsvariable*.

- (c) Die Zufallsvariable $X =$ „Brenndauer einer Glühbirne“

$$X : S \rightarrow \mathbb{R}_0^+, \quad \text{wobei } S = \text{Menge der Glühbirnen,}$$

$$X(e_i) = \text{„Brenndauer der Glühbirne } e_i\text{“},$$

ist eine *stetige* (oder *kontinuierliche*) *Zufallsvariable*. Hier ist das Elementarereignis des Ziehens der Glühbirne e_i die gezogene Glühbirne e_i .

Betrachten wir ein weiteres Beispiel, das uns in diesem Kapitel begleiten wird.

Beispiel 10.3. (Zufallsvariable Anzahl Zahlseiten beim dreimaligen Werfen einer idealen Münze¹)

Unser Zufallsexperiment ist das dreimalige Werfen einer idealen Münze mit den Seiten Kopf (K) und Zahl (Z). Als *Zufallsvariable* X betrachten wir die „Anzahl der geworfenen Zahlseiten“.

Hier ist der Ereignisraum (mit Berücksichtigung der Reihenfolge der Münzwürfe)

$$S = \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\}.$$

Die Werte der Zufallsvariable für die einzelnen Elementarereignisse sind also

$$\begin{aligned} X(K, K, K) &= 0, & X(K, K, Z) &= 1, & X(K, Z, K) &= 1, & X(Z, K, K) &= 1, \\ X(K, Z, Z) &= 2, & X(Z, K, Z) &= 2, & X(Z, Z, K) &= 2, & X(Z, Z, Z) &= 3. \end{aligned}$$

Abschließend betrachten wir ein aufwendigeres Beispiel einer Zufallsvariable.

Beispiel 10.4. (Stichprobenmittelwert als Zufallsvariable)

Mit einer statistischen Erhebung soll das durchschnittliche Einkommen der Vollzeitbeschäftigten der BRD bestimmt werden. Dazu geht man wie folgt vor:

Nach einem System (Stichprobenplan) werden Stichproben aus der Menge der Vollzeitbeschäftigten gezogen. Die jeweilige Stichprobe ist das Ergebnis des Zufallsprozesses, und wir ordnen ihr als Zufallsvariable das arithmetische Mittel der Einkommen der Vollzeitbeschäftigten in der Stichprobe zu.

Aus den arithmetischen Mitteln verschiedener Stichproben kann dann mit dem Erwartungswert (siehe Teilkapitel 10.4) ein Wert für das durchschnittliche Einkommen der Vollzeitbeschäftigten der BRD bestimmt werden.

10.2 Wahrscheinlichkeitsfunktion und Verteilungsfunktion einer diskreten Zufallsvariablen

Wir betrachten zunächst diskrete Zufallsvariablen, denn hier sind alle neu einzuführenden Begriffe analog zu Häufigkeitsverteilungen, so dass beim Verständnis keine Hürde auftritt.

Definition 10.5. (Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen)

Sei $S = \{e_1, e_2, \dots\}$ der Ereignisraum eines Zufallsexperiments und $X : S \rightarrow B \subseteq \mathbb{R}$ eine **diskrete Zufallsvariable** mit den Werten $x_1 < x_2 < x_3 < \dots$.

(1) Die **Wahrscheinlichkeit**, mit der X den **Wert** x_i **annimmt**, ist

$$W(X = x_i) = \sum_{X(e_j)=x_i} W(e_j).$$

¹Dieses Beispiel ist durch ein Beispiel des dreifachen Münzwurfs in [B12, Kapitel 7] inspiriert und orientiert sich in seiner Darstellung an diesem Beispiel.

(Wir summieren also über die Wahrscheinlichkeiten aller Elementarereignisse e_j , denen von X der Wert x_i zugeordnet wird, d.h. $X(e_j) = x_i$.)

- (2) Die **Wahrscheinlichkeitsfunktion** (oder **Wahrscheinlichkeitsverteilung**) der Zufallsvariablen X mit den Werten (oder Ausprägungen) $x_1 < x_2 < x_3 < \dots$ ist

$$f: \mathbb{R} \rightarrow [0, 1], \quad f(x) = \begin{cases} W(X = x_i) & \text{wenn } x = x_i \text{ für } i = 1, 2, 3, \dots, \\ 0 & \text{wenn } x \notin \{x_1, x_2, x_3, \dots\}. \end{cases}$$

(Interpretation: $f(x)$ ist die Wahrscheinlichkeit, mit der X den Wert x annimmt.)

In der nächsten Bemerkung halten wir zwei wichtige Beobachtungen zu den neuen Begriffen fest.

Bemerkung 10.6. (Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen)

- (1) Da $f(x)$ eine Wahrscheinlichkeit ist (nämlich die Wahrscheinlichkeit, mit der X den Wert x annimmt), gilt immer

$$0 \leq f(x) \leq 1 \quad \text{für alle } x \in \mathbb{R}.$$

Weiter gilt

$$\sum_i f(x_i) = \sum_i W(X = x_i) = 1.$$

- (2) Die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen ist **analog zur relativen Häufigkeitsverteilung** bei nicht-klassifizierten Daten eines diskreten metrisch skalierten Merkmals.
- (3) Man kann eine Wahrscheinlichkeitsfunktion bequem mit einem **Stabdiagramm** (s.u. in Beispiel 10.7 und Abbildung 10.1) veranschaulichen.

Betrachten wir ein Beispiel.

Beispiel 10.7. (Wahrscheinlichkeitsfunktion – dreimaliges Werfen einer idealen Münze²)

Unser Zufallsexperiment ist das dreimalige Werfen einer idealen Münze mit den Seiten Kopf (K) und Zahl (Z). Als *Zufallsvariable* X betrachten wir die „Anzahl der geworfenen Zahlseiten“.

Hier ist der Ereignisraum (mit Berücksichtigung der Reihenfolge der Münzwürfe)

$$S = \{(K, K, K), (K, K, Z), (K, Z, K), (Z, K, K), (K, Z, Z), (Z, K, Z), (Z, Z, K), (Z, Z, Z)\},$$

und wir bestimmen zunächst für jedes Elementarereignis die Wahrscheinlichkeit. Dabei haben wir die Elementarereignisse bereits in Gruppen mit der gleichen Anzahl Zahlseiten

²Dieses Beispiel ist durch ein Beispiel des dreifachen Münzwurfs in [B12, Kapitel 7] inspiriert und orientiert sich in seiner Darstellung an diesem Beispiel.

gruppiert. Jedes dieser Elementarereignisse ist gleichwahrscheinlich mit der Wahrscheinlichkeit $(1/2)^3 = 1/8 = 1/2^3$ (Fall: „mit Wiederholung und mit Berücksichtigung der Anordnung“, $N = 2$ und $n = 3$).

Elementarereignis e	Wahrscheinlichkeit $W(e)$	„Anzahl Zahl“ x	Wahrscheinlichkeitsfunktion $f(x) = W(X = x)$
$e_1 = (K, K, K)$	$W(e_1) = 1/8$	$x_1 = 0$	$f(x_1) = \frac{1}{8} = 0,125$
$e_2 = (K, K, Z)$	$W(e_2) = 1/8$	$x_2 = 1$	$f(x_2) = 3 \cdot \frac{1}{8} = \frac{3}{8} = 0,375$
$e_3 = (K, Z, K)$	$W(e_3) = 1/8$		
$e_4 = (Z, K, K)$	$W(e_4) = 1/8$		
$e_5 = (K, Z, Z)$	$W(e_5) = 1/8$	$x_3 = 2$	$f(x_3) = 3 \cdot \frac{1}{8} = \frac{3}{8} = 0,375$
$e_6 = (Z, K, Z)$	$W(e_6) = 1/8$		
$e_7 = (Z, Z, K)$	$W(e_7) = 1/8$		
$e_8 = (Z, Z, Z)$	$W(e_8) = 1/8$	$x_4 = 3$	$f(x_4) = \frac{1}{8} = 0,125$

Damit erhalten wir die folgende *Wahrscheinlichkeitsfunktion*:

$$f : \mathbb{R} \rightarrow [0, 1], \quad f(x) = \begin{cases} \frac{1}{8} = 0,125 & \text{wenn } x = 0, \\ \frac{3}{8} = 0,375 & \text{wenn } x = 1, \\ \frac{3}{8} = 0,375 & \text{wenn } x = 2, \\ \frac{1}{8} = 0,125 & \text{wenn } x = 3, \\ 0 & \text{wenn } x \in \mathbb{R} \setminus \{0, 1, 2, 3\}. \end{cases}$$

Wir haben die Wahrscheinlichkeitsfunktion in Abbildung 10.1 als *Stabdiagramm* gezeichnet.

Nun kommen wir zu dem Begriff der Verteilungsfunktion einer diskreten Zufallsvariablen.

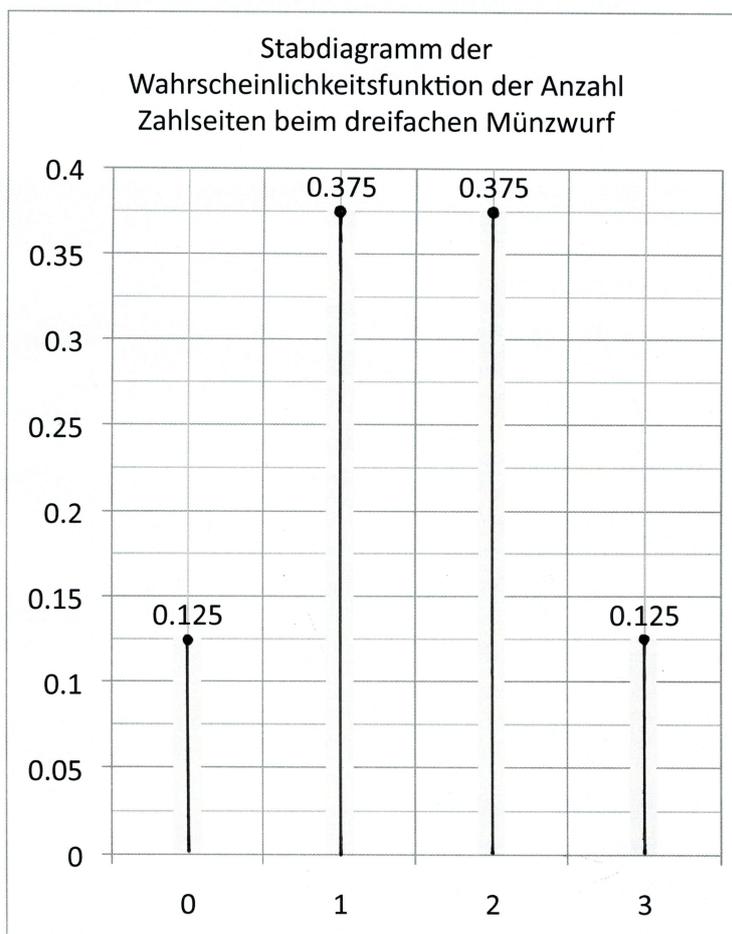


Abbildung 10.1: Stabdiagramm der Wahrscheinlichkeitsfunktion der Zufallsvariable $X =$ „Anzahl der geworfenen Zahlseiten“ beim dreifachen Münzwurf (siehe Beispiel 10.7).

Definition 10.8. (Verteilungsfunktion einer diskreten Zufallsvariablen)

Sei S der Ereignisraum eines Zufallsexperiments, und sei $X : S \rightarrow B \subseteq \mathbb{R}$ eine **diskrete Zufallsvariable** mit den Werten $x_1 < x_2 < x_3 < \dots$. Die **Verteilungsfunktion** von X ,

$$F : \mathbb{R} \rightarrow [0, 1], \quad F(x) = W(X \leq x) = \sum_{x_i \leq x} f(x_i),$$

gibt die **Wahrscheinlichkeit** $W(X \leq x)$ dafür an, dass X **höchstens** den Wert x annimmt.

In der nachfolgenden Bemerkung halten wir einige wichtige Eigenschaften der Verteilungsfunktion einer diskreten Zufallsvariablen fest.

Bemerkung 10.9. (Verteilungsfunktion einer diskreten Zufallsvariablen)

(1) F ist eine **monoton wachsende/steigende Treppenfunktion**, und es gilt immer

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

(2) Die **Wahrscheinlichkeit**, dass der Wert von X im Intervall $(a, b]$ liegt, ist

$$W(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} f(x_i).$$

(3) $f(x_{i+1}) = F(x_{i+1}) - F(x_i)$ für alle $i = 1, 2, \dots$ und $f(x_1) = F(x_1)$.

(4) Die Verteilungsfunktion ist **analog zur relativen Summenhäufigkeitsfunktion** (bei nicht-klassifizierten Daten) eines diskreten metrisch skalierten Merkmals.

Betrachten wir wieder unser Beispiel des dreifachen Münzwurfs.

Beispiel 10.10. (Verteilungsfunktion – dreimaliges Werfen einer idealen Münze³)

Wir betrachten wieder das Zufallsexperiment des dreimaligen Werfens einer idealen Münze mit den Seiten Kopf (K) und Zahl (Z). Unsere Zufallsvariable ist $X =$ „Anzahl der geworfenen Zahlseiten“. Die Verteilungsfunktion

$$\begin{aligned} F(x) = W(X \leq x) &= \left(\begin{array}{l} \text{Wahrscheinlichkeit, dass die Zufallsvari-} \\ \text{able } X \text{ höchstens den Wert } x \text{ annimmt} \end{array} \right) \\ &= \left(\begin{array}{l} \text{Wahrscheinlichkeit, dass höchstens} \\ x \text{ Zahlseiten geworfen werden} \end{array} \right) \end{aligned}$$

ist gegeben durch $F : \mathbb{R} \rightarrow [0, 1]$,

$$F(x) = \left\{ \begin{array}{ll} 0 & \text{wenn } x < 0, \\ W(X = 0) = \frac{1}{8} = 0,125 & \text{wenn } 0 \leq x < 1, \\ W(X = 0) + W(X = 1) = \frac{1}{8} + \frac{3}{8} = \frac{4}{8} = 0,5 & \text{wenn } 1 \leq x < 2, \\ \left. \begin{array}{l} W(X = 0) + W(X = 1) + W(X = 2) \\ = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} = \frac{7}{8} = 0,875 \end{array} \right\} & \text{wenn } 2 \leq x < 3, \\ \left. \begin{array}{l} W(X = 0) + W(X = 1) + W(X = 2) + W(X = 3) \\ = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{8}{8} = 1 \end{array} \right\} & \text{wenn } x \geq 3. \end{array} \right.$$

³Dieses Beispiel ist durch ein Beispiel des dreifachen Münzwurfs in [B12, Kapitel 7] inspiriert und orientiert sich in seiner Darstellung an diesem Beispiel.

Wir haben den Graphen der Verteilungsfunktion in Abbildung 10.2 gezeichnet.

Wir sehen an dem Beispiel: Die Verteilungsfunktion ist eine Treppenfunktion mit

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{und} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

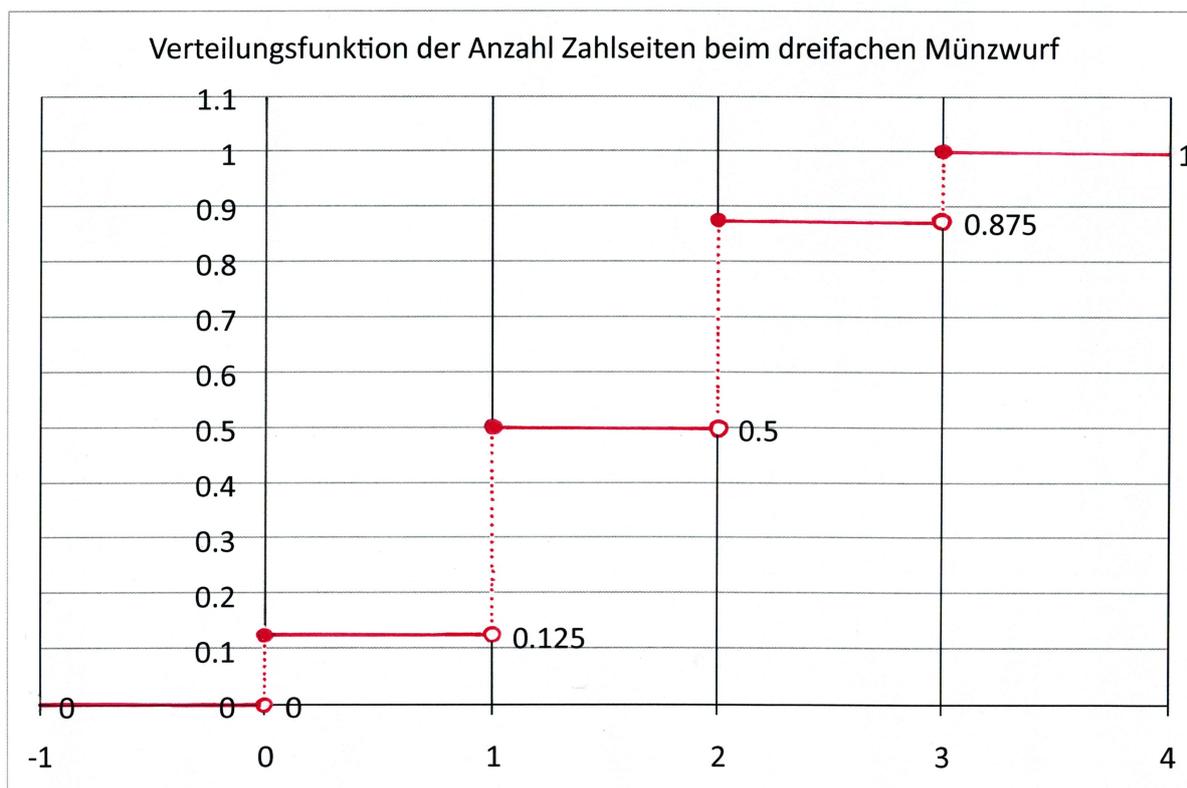


Abbildung 10.2: Graph der Verteilungsfunktion der Zufallsvariable $X =$ „Anzahl der geworfenen Zahlseiten“ beim dreifachen Münzwurf (siehe Beispiel 10.10).

10.3 Verteilungsfunktion und Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen

Nun betrachten wir stetige Zufallsvariable, die in einem Bereich B von \mathbb{R} (z.B. einem Intervall) alle reellen Zahlen als Ausprägungen annehmen können.

Definition 10.11. (Verteilungsfunktion einer stetigen Zufallsvariablen)

Sei S der Ereignisraum eines Zufallsexperiments, und sei $X : S \rightarrow B \subseteq \mathbb{R}$ eine **stetige Zufallsvariable** mit Werten in B . Die **Verteilungsfunktion** F von X ist definiert durch

$$F : \mathbb{R} \rightarrow [0, 1],$$

$F(x) = W(X \leq x) = \text{Wahrscheinlichkeit, dass } X \text{ höchstens den Wert } x \text{ annimmt.}$

(Erklärung: Die Werte der Verteilungsfunktion F müssen im Intervall $[0, 1]$ liegen, weil F Wahrscheinlichkeiten angibt.)

Betrachten wir zunächst ein Beispiel.

Beispiel 10.12. (Brenndauer einer Glühbirne)

Wir betrachten die stetige Zufallsvariable $X = \text{„Brenndauer einer Glühbirne“}$, also

$$X : S \rightarrow \mathbb{R}_0^+, \quad \text{wobei } S = \text{Menge der Glühbirnen,}$$

$$X(e_i) = \text{„Brenndauer der Glühbirne } e_i\text{“}.$$

Dann beschreibt die Verteilungsfunktion $F(x) = W(X \leq x)$ die Wahrscheinlichkeit, dass die Glühbirne höchstens eine Brenndauer von x hat.

In Abbildung 10.3 sind die Graphen dreier Beispiele für eine Verteilungsfunktion gezeichnet. Die gezeichneten Verteilungsfunktionen gehören jeweils zu einer (Gaußschen) Normalverteilung.

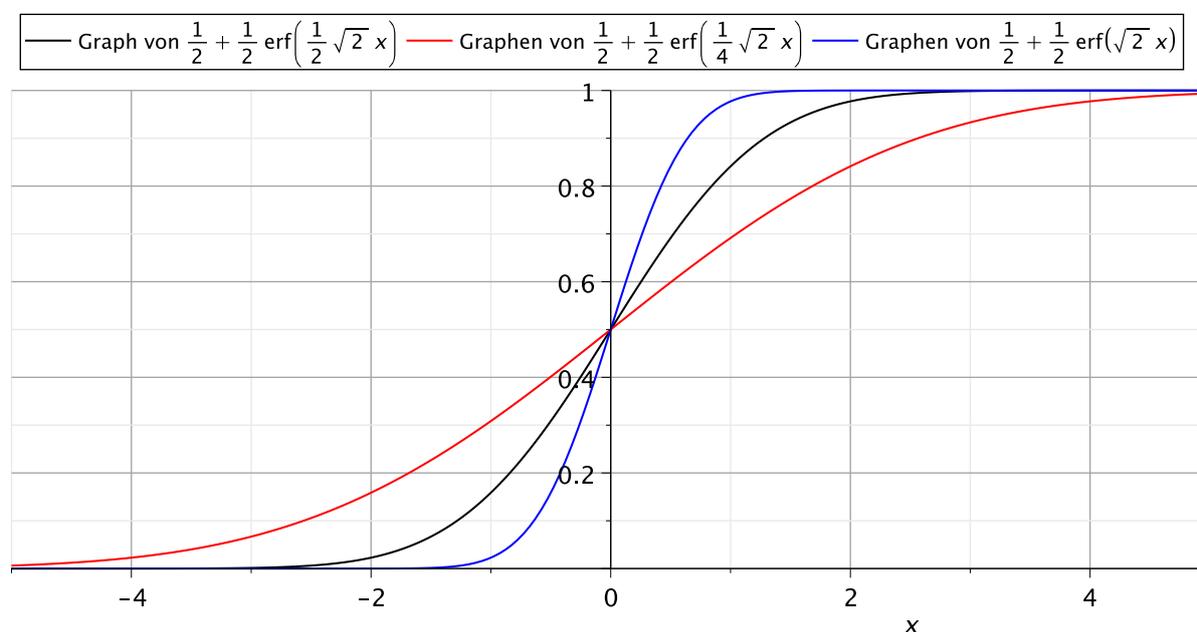


Abbildung 10.3: Graphen der Verteilungsfunktion einer normalverteilten stetigen Zufallsvariablen mit Erwartungswert 0 und Standardabweichung $\sigma = 1$ (schwarz), $\sigma = 2$ (rot) bzw. $\sigma = 0,5$ (blau).

Wir halten nun einige Eigenschaften der Verteilungsfunktion einer stetigen (oder kontinuierlichen) Zufallsvariablen fest.

Lemma 10.13. (Eigenschaften der Verteilungsfunktion einer stetigen Zufallsvariablen)

Die **Verteilungsfunktion** $F : \mathbb{R} \rightarrow [0, 1]$ einer **stetigen Zufallsvariablen** $X : S \rightarrow B \subseteq \mathbb{R}$ hat die folgenden Eigenschaften:

- (1) $F(x)$ ist **monoton wachsend/steigend**, d.h. für $x < \tilde{x}$ gilt $F(x) \leq F(\tilde{x})$.
- (2) Es gelten die **Grenzwerte** $\lim_{x \rightarrow -\infty} F(x) = 0$ und $\lim_{x \rightarrow \infty} F(x) = 1$.
- (3) $F(x)$ ist **überall stetig** und **bis auf höchstens endlich viele Stellen differenzierbar**.

Die Entsprechung der Wahrscheinlichkeitsfunktion im Fall diskreter Zufallsvariablen ist bei einer kontinuierlichen Zufallsvariablen die Wahrscheinlichkeitsdichte

Definition 10.14. (Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen)

Sei S der Ereignisraum eines Zufallsexperiments, und sei $X : S \rightarrow B \subseteq \mathbb{R}$ eine **stetige Zufallsvariable** mit der Verteilungsfunktion $F : \mathbb{R} \rightarrow [0, 1]$. Die **Wahrscheinlichkeitsdichte** (oder **Dichtefunktion**) f von X ist die Ableitung F' der Verteilungsfunktion F , also

$$f : \mathbb{R} \rightarrow \mathbb{R}_0^+ \quad , \quad f(x) = F'(x).$$

(Anmerkung: Wir bemerken, dass die Werte der Wahrscheinlichkeitsdichte ≥ 0 sein müssen, da F monoton wachsend/steigend ist und seine Ableitung $f = F'$ daher nur Werte ≥ 0 annimmt. Es gilt aber **nicht** $f(x) \in [0, 1]$.)

In Abbildung 10.4 sind die Graphen dreier Beispiele für eine Wahrscheinlichkeitsdichte gezeichnet. Es sind die zu den in Abbildung 10.3 gezeichneten Verteilungsfunktionen gehörigen Wahrscheinlichkeitsdichten (gleiche farbige Kennzeichnung). Bei allen drei Wahrscheinlichkeitsdichten handelt es sich um eine (Gaußsche) Normalverteilung.

In der nächsten Bemerkung halten wir wichtige Eigenschaften der Wahrscheinlichkeitsdichte fest.

Lemma 10.15. (Eigenschaften der Wahrscheinlichkeitsdichte)

Die **Wahrscheinlichkeitsdichte** $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$ einer **stetigen Zufallsvariablen** $X : S \rightarrow B \subseteq \mathbb{R}$ hat die folgenden Eigenschaften:

- (1) Die Verteilungsfunktion ist eine **Stammfunktion** der Wahrscheinlichkeitsdichte:

$$F(x) = \int_{-\infty}^x f(t) dt$$

- (2) Die Wahrscheinlichkeitsdichte hat nur **nicht-negative** Werte:

$$f(x) \geq 0 \quad \text{für alle } x \in \mathbb{R}.$$

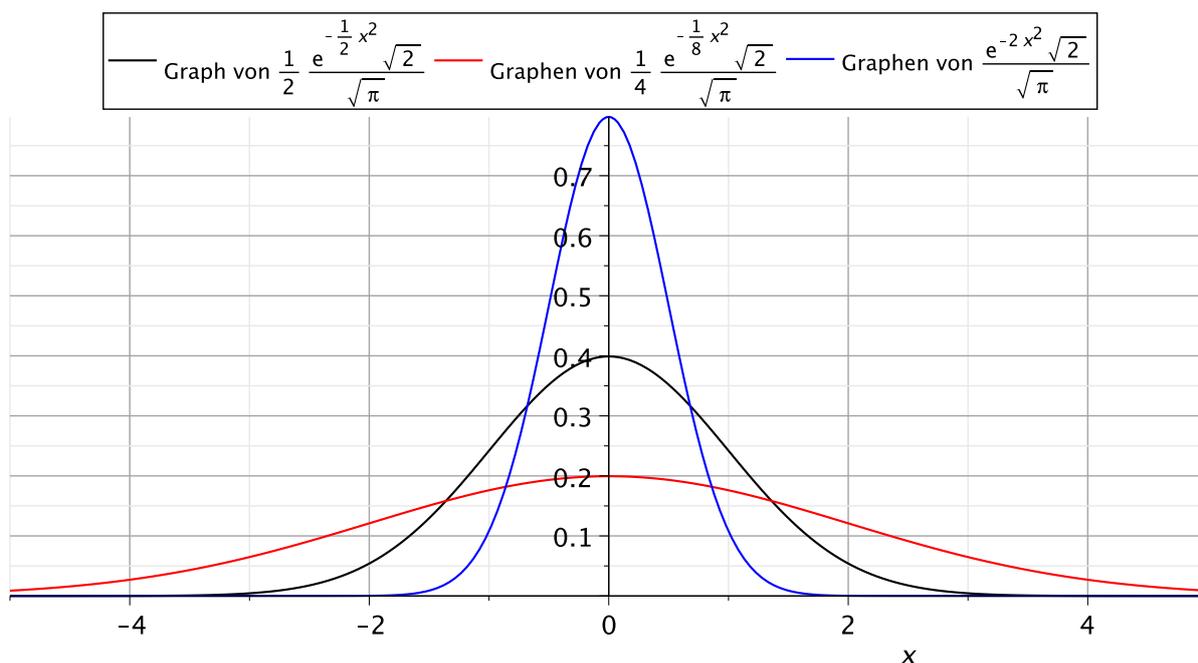


Abbildung 10.4: Graphen der Wahrscheinlichkeitsdichte einer normalverteilten stetigen Zufallsvariablen mit Erwartungswert 0 und Standardabweichung $\sigma = 1$ (schwarz), $\sigma = 2$ (rot) bzw. $\sigma = 0,5$ (blau).

(3) **Normierung:** $\int_{-\infty}^{\infty} f(x) dx = 1$ (Fläche unter dem Graphen von f ist 1.)

(4) $W(a \leq X \leq b) = \int_a^b f(x) dx$

= Wahrscheinlichkeit, dass X einen Wert im Intervall $[a, b]$ annimmt

= Fläche unter dem Graphen der Wahrscheinlichkeitsdichte von $x = a$ bis $x = b$ (siehe Abbildung 10.6)

(5) Für jeden Wert $x \in \mathbb{R}$ gilt $W(X = x) = 0$, d.h. die Wahrscheinlichkeit, dass ein **einzelner Wert** x angenommen wird, ist **immer Null**.

(Anschauung: Für einen einzelnen Wert a von X haben wir keinen Flächeninhalt, also $\int_a^a f(x) dx = 0$.)

Achtung: Die Werte $f(x)$ der Wahrscheinlichkeitsdichte sind **keine** Wahrscheinlichkeiten und können daher beliebig groß sein! Es gilt nur $f(x) \geq 0$ aber **nicht** $f(x) \leq 1$.

Betrachten wir nun ein Beispiel.

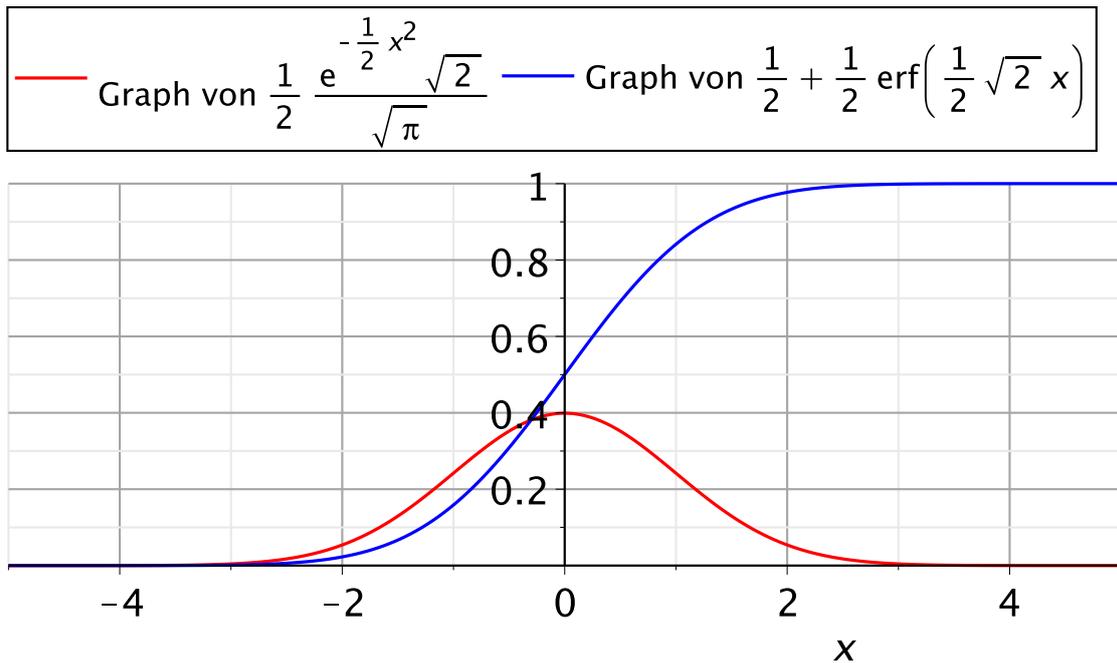


Abbildung 10.5: Graphen der Wahrscheinlichkeitsdichte (rot) und der zugehörigen Verteilungsfunktion (blau) einer normalverteilten Zufallsvariablen mit Erwartungswert 0 und Standardabweichung $\sigma = 1$.

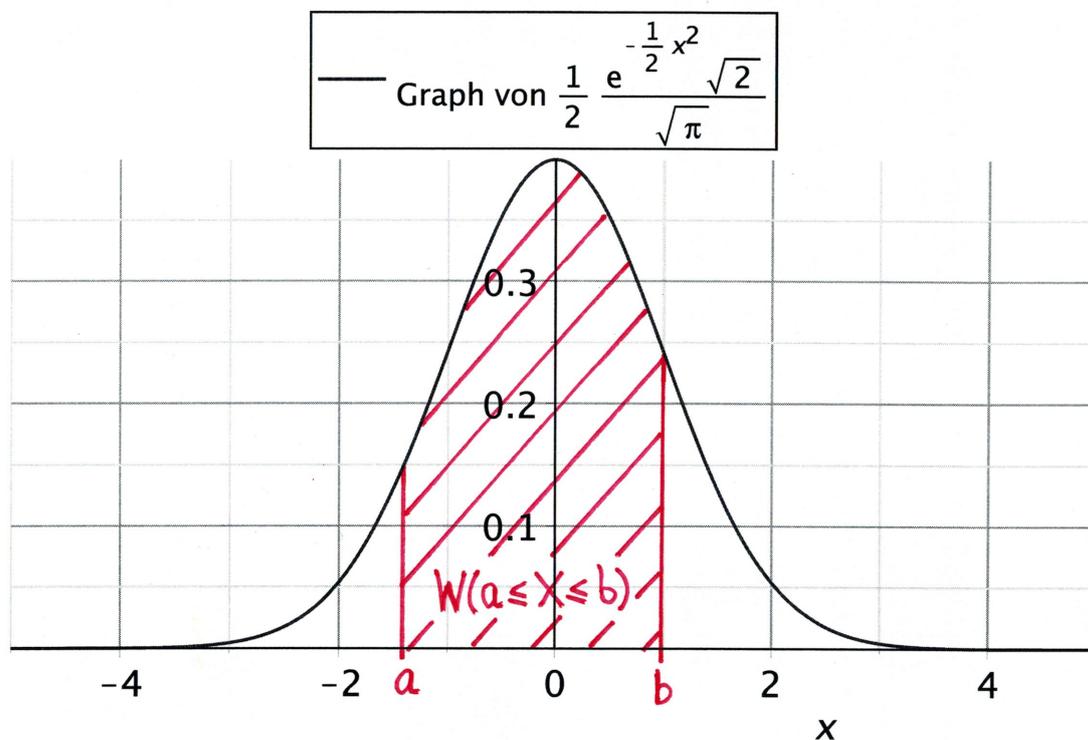


Abbildung 10.6: Die Wahrscheinlichkeit $W(a \leq X \leq b)$ als Fläche unter dem Graphen der Wahrscheinlichkeitsdichte.

Beispiel 10.16. (Wahrscheinlichkeitsdichte und Verteilungsfunktion – Abweichungen der Fahrzeit eines Airport-Shuttle-Trains)

Ein Airport-Shuttle-Train verbindet Terminal A und Terminal B eines Flughafens. Die vorgesehene Fahrzeit von einem Terminal zum nächsten ist 15 Minuten. Die *stetige Zufallsvariable* $X =$ „Abweichungen von der vorgesehenen Fahrzeit (in min)“ habe die folgende *Wahrscheinlichkeitsdichte*:

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{für } x < -1 \\ \frac{2}{3} + \frac{2}{3} \cdot x & \text{für } -1 \leq x < 0, \\ \frac{2}{3} - \frac{1}{3} \cdot x & \text{für } 0 \leq x < 2, \\ 0 & \text{für } x \geq 2. \end{cases}$$

Wir haben den Graphen der Wahrscheinlichkeitsdichte in Abbildung 10.7 gezeichnet. Wir

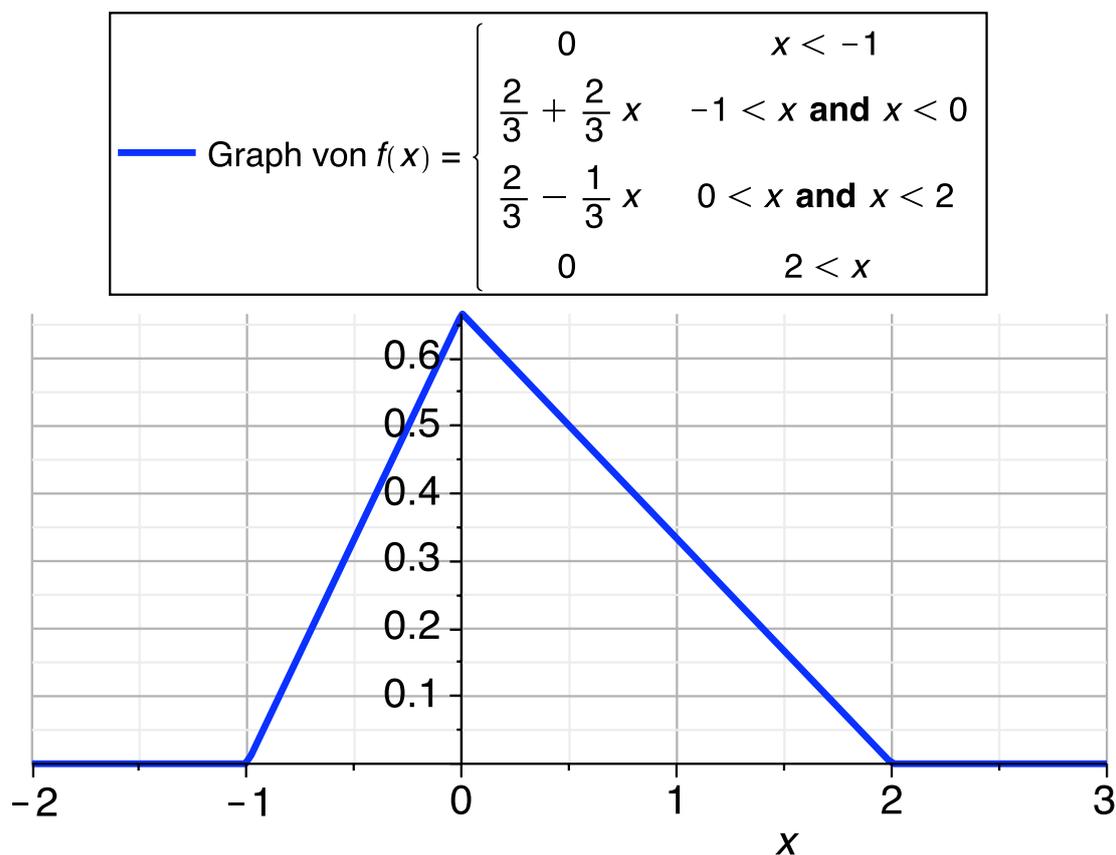


Abbildung 10.7: Wahrscheinlichkeitsdichte der Zufallsvariable $X =$ „Abweichungen von der vorgesehenen Fahrzeit (in min)“ aus Beispiel 10.16.

überprüfen, dass f eine zulässige Wahrscheinlichkeitsdichte ist, indem wir überprüfen, ob die Eigenschaften $f(x) \geq 0$ für alle $x \in \mathbb{R}$ und $\int_{-\infty}^{\infty} f(x) dx = 1$ gelten. Wir haben

$$f(x) \geq 0 \quad \text{für alle } x \in \mathbb{R},$$

$$\text{denn } \begin{cases} \frac{2}{3} + \frac{2}{3} \cdot x \geq \frac{2}{3} + \frac{2}{3} \cdot (-1) = 0 & \text{für } -1 \leq x < 0, \\ \frac{2}{3} - \frac{1}{3} \cdot x \geq \frac{2}{3} - \frac{1}{3} \cdot 2 = 0 & \text{für } 0 \leq x < 2, \\ f(x) = 0 & \text{für } x < -1 \text{ oder } x \geq 2, \end{cases}$$

$$\begin{aligned} \int_{-\infty}^{\infty} f(x) dx &= \int_{-1}^0 \left(\frac{2}{3} + \frac{2}{3} \cdot x \right) dx + \int_0^2 \left(\frac{2}{3} - \frac{1}{3} \cdot x \right) dx \\ &= \left[\frac{2}{3} \cdot x + \frac{1}{3} \cdot x^2 \right]_{-1}^0 + \left[\frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2 \right]_0^2 \\ &= 0 - \left[\frac{2}{3} \cdot (-1) + \frac{1}{3} \cdot (-1)^2 \right] + \left[\frac{2}{3} \cdot 2 - \frac{1}{6} \cdot 2^2 \right] - 0 \\ &= \frac{2}{3} - \frac{1}{3} + \frac{4}{3} - \frac{4}{6} = 1. \end{aligned}$$

Also ist $f(x)$ in der Tat eine Wahrscheinlichkeitsdichte.

Wir wollen nun die *Wahrscheinlichkeit* bestimmen, dass die Abweichung von der vorgesehenen Fahrzeit höchstens eine halbe Minute ist. Wir suchen also $W\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right)$

$$\begin{aligned} W\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right) &= \int_{-1/2}^{1/2} f(x) dx = \int_{-1/2}^0 \left(\frac{2}{3} + \frac{2}{3} \cdot x \right) dx + \int_0^{1/2} \left(\frac{2}{3} - \frac{1}{3} \cdot x \right) dx \\ &= \left[\frac{2}{3} \cdot x + \frac{1}{3} \cdot x^2 \right]_{-1/2}^0 + \left[\frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2 \right]_0^{1/2} \\ &= 0 - \left[\frac{2}{3} \cdot \left(-\frac{1}{2}\right) + \frac{1}{3} \cdot \left(-\frac{1}{2}\right)^2 \right] + \left[\frac{2}{3} \cdot \frac{1}{2} - \frac{1}{6} \cdot \left(\frac{1}{2}\right)^2 \right] - 0 \\ &= \frac{1}{3} - \frac{1}{12} + \frac{1}{3} - \frac{1}{24} = \frac{2}{3} - \frac{1}{8} = \frac{13}{24} \approx 0,542 \end{aligned}$$

Also beträgt die Wahrscheinlichkeit, dass die Abweichung von der vorgesehenen Fahrzeit höchstens eine halbe Minute ist, $W\left(-\frac{1}{2} \leq X \leq \frac{1}{2}\right) = \frac{13}{24} \approx 0,542$.

Wir wollen nun die *Wahrscheinlichkeit* bestimmen, dass die Abweichung von der vorgesehenen Fahrzeit mehr als eine Minute ist. Wir suchen also $W(X < -1) + W(X > 1)$. An der Dichtfunktion sieht man direkt, dass gilt $W(X < -1) = 0$ (denn es gilt $f(x) = 0$ für alle $x \leq -1$).

$$\begin{aligned} W(X > 1) &= \int_1^{\infty} f(x) dx = \int_1^2 \left(\frac{2}{3} - \frac{1}{3} \cdot x \right) dx = \left[\frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2 \right]_1^2 \\ &= \left[\frac{2}{3} \cdot 2 - \frac{1}{6} \cdot 2^2 \right] - \left[\frac{2}{3} \cdot 1 - \frac{1}{6} \cdot 1^2 \right] = \frac{4}{3} - \frac{4}{6} - \frac{2}{3} + \frac{1}{6} = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \approx 0,167 \end{aligned}$$

Also beträgt die Wahrscheinlichkeit, dass die Abweichung von der vorgesehenen Fahrzeit mehr als eine Minute ist, $W(X < -1) + W(X > 1) = 0 + \frac{1}{6} \approx 0,167$.

Wir berechnen nun die *Verteilungsfunktion* der Zufallsvariable $X =$ „Abweichungen von der vorgesehenen Fahrzeit (in min)“:

$$F(x) = \int_{-\infty}^x f(t) dt,$$

und wir müssen nun mehrere Fälle unterscheiden: Wir wissen bereits

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x 0 dt = 0 \quad \text{wenn } x < -1,$$

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-1}^2 f(x) dt = 1 \quad \text{wenn } x \geq 2.$$

Für $x \in [-1, 0)$ finden wir

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt = \int_{-1}^x \left(\frac{2}{3} + \frac{2}{3} \cdot t \right) dt = \left[\frac{2}{3} \cdot t + \frac{1}{3} \cdot t^2 \right]_{-1}^x \\ &= \left[\frac{2}{3} \cdot x + \frac{1}{3} \cdot x^2 \right] - \left[\frac{2}{3} \cdot (-1) + \frac{1}{3} \cdot (-1)^2 \right] = \frac{2}{3} \cdot x + \frac{1}{3} \cdot x^2 + \frac{1}{3}. \end{aligned}$$

Für $x \in [0, 2)$ finden wir

$$\begin{aligned} F(x) &= \int_{-\infty}^x f(t) dt = \int_{-1}^0 \left(\frac{2}{3} + \frac{2}{3} \cdot t \right) dt + \int_0^x \left(\frac{2}{3} - \frac{1}{3} \cdot t \right) dt \\ &= \left[\frac{2}{3} \cdot t + \frac{1}{3} \cdot t^2 \right]_{-1}^0 + \left[\frac{2}{3} \cdot t - \frac{1}{6} \cdot t^2 \right]_0^x \\ &= 0 - \left[\frac{2}{3} \cdot (-1) + \frac{1}{3} \cdot (-1)^2 \right] + \left[\frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2 \right] - 0 \\ &= \frac{1}{3} + \frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2. \end{aligned}$$

Also ist die *Verteilungsfunktion* gegeben durch

$$F : \mathbb{R} \rightarrow [0, 1], \quad F(x) = \begin{cases} 0 & \text{wenn } x < -1, \\ \frac{1}{3} + \frac{2}{3} \cdot x + \frac{1}{3} \cdot x^2 & \text{wenn } -1 \leq x < 0, \\ \frac{1}{3} + \frac{2}{3} \cdot x - \frac{1}{6} \cdot x^2 & \text{wenn } 0 \leq x < 2, \\ 1 & \text{wenn } x \geq 2. \end{cases}$$

Wir können nun die Wahrscheinlichkeit, dass die Abweichung von der vorgesehenen Fahrzeit höchstens eine halbe Minute ist, auch mit der Verteilungsfunktion berechnen:

$$W \left(-\frac{1}{2} \leq X \leq \frac{1}{2} \right) = \int_{-1/2}^{1/2} f(x) dx = F \left(\frac{1}{2} \right) - F \left(-\frac{1}{2} \right)$$

$$\begin{aligned}
&= \left[\frac{1}{3} + \frac{2}{3} \cdot \frac{1}{2} - \frac{1}{6} \cdot \left(\frac{1}{2}\right)^2 \right] - \left[\frac{1}{3} + \frac{2}{3} \cdot \left(-\frac{1}{2}\right) + \frac{1}{3} \cdot \left(-\frac{1}{2}\right)^2 \right] \\
&= \frac{1}{3} + \frac{1}{3} - \frac{1}{24} - \frac{1}{3} + \frac{1}{3} - \frac{1}{12} = \frac{2}{3} - \frac{1}{8} = \frac{13}{24} \approx 0,542.
\end{aligned}$$

Wir können nun die Wahrscheinlichkeit, dass die Abweichung von der vorgesehenen Fahrzeit mehr als eine Minute ist, auch mit der Verteilungsfunktion berechnen:

$$W(X < -1) = W(X \leq -1) = F(-1) = 0 \quad \text{und}$$

$$W(X > 1) = 1 - W(X \leq 1) = 1 - \left[\frac{1}{3} + \frac{2}{3} \cdot 1 - \frac{1}{6} \cdot 1^2 \right] = 1 - 1 + \frac{1}{6} = \frac{1}{6} \approx 0,167,$$

$$\text{also} \quad W(X < -1) + W(X > 1) = 0 + \frac{1}{6} = \frac{1}{6} \approx 0,167.$$

Der Graph der Verteilungsfunktion F ist in Abbildung 10.8 gezeichnet.

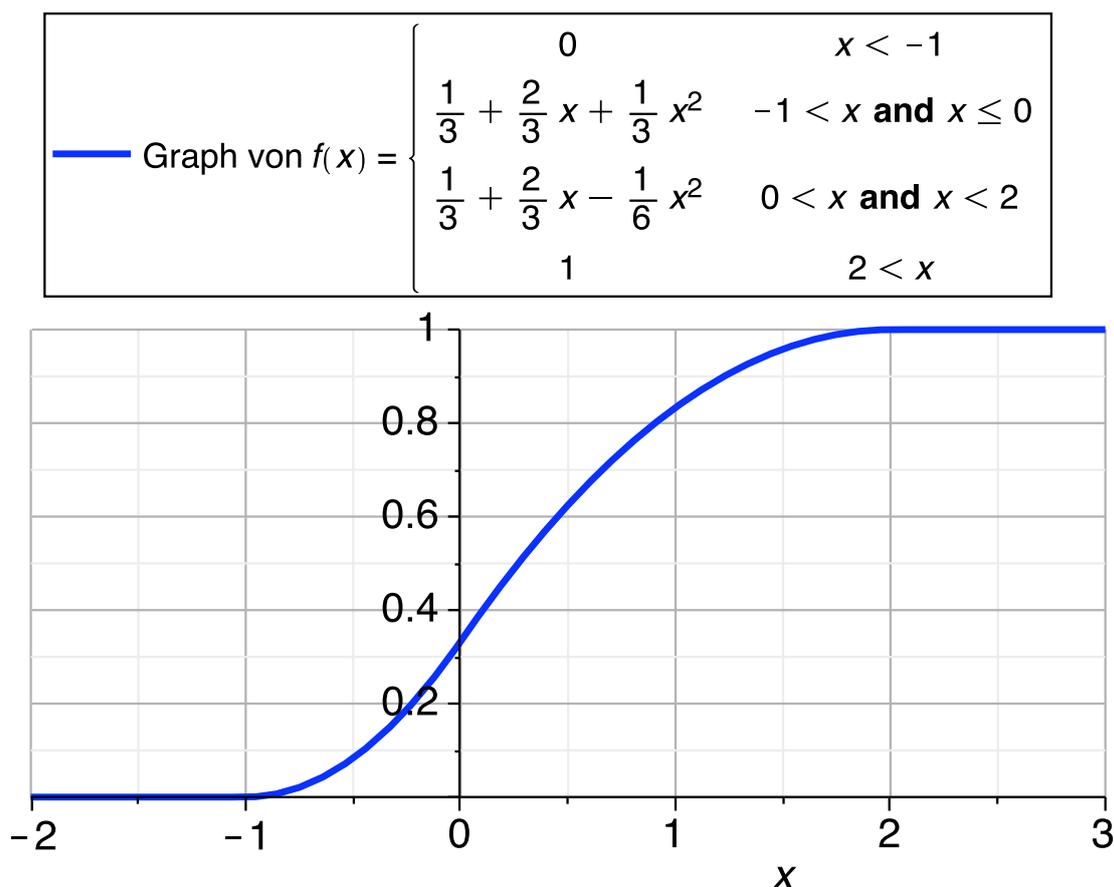


Abbildung 10.8: Verteilungsfunktion der Zufallsvariable $X =$ „Abweichungen von der vorgesehenen Fahrzeit (in min)“ aus Beispiel 10.16.

Wir halten noch wichtige Eigenschaften von stetigen Zufallsvariablen fest.

Bemerkung 10.17. (Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen)

Wollen wir überprüfen, ob eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ eine **Wahrscheinlichkeitsdichte** (einer geeigneten **stetigen Zufallsvariablen**) ist, so müssen wir die Eigenschaften

(1) $f(x) \geq 0$ für alle $x \in \mathbb{R}$, und

$$(2) \int_{-\infty}^{\infty} f(x) dx = 1$$

überprüfen. Sind diese beide erfüllt, so ist f eine **Wahrscheinlichkeitsdichte**.

Ist eine (oder sind beide) dieser Eigenschaften verletzt, so ist f **keine** Wahrscheinlichkeitsdichte.

Bemerkung 10.18. (Wahrscheinlichkeiten bei einer stetigen Zufallsvariablen)

Für eine stetige Zufallsvariable gilt immer

$$W(a \leq X \leq b) = W(a \leq X < b) = W(a < X \leq b) = W(a < X < b),$$

weil $W(X = a) = W(X = b) = 0$ ist. Dieses wird auf einem Übungszettel gezeigt.

10.4 Erwartungswert und Varianz einer diskreten Zufallsvariablen

Analog zu einer Häufigkeitsverteilung können wir auch Zufallsvariable durch geeignete Kennzahlen charakterisieren, nämlich dem **Erwartungswert** (dieser entspricht dem arithmetischen Mittel einer Häufigkeitsverteilung) und die **Varianz** (diese entspricht der Varianz einer Häufigkeitsverteilung) als Streuungsmaß.

Definition 10.19. (Erwartungswert und Varianz einer diskreten Zufallsvariablen)

Sei S der Ereignisraum eines Zufallsexperiments, und sei $X : S \rightarrow B \subseteq \mathbb{R}$ eine **diskrete Zufallsvariable** mit den Werten $x_1 < x_2 < x_3 < \dots$ und der Wahrscheinlichkeitsfunktion $f : \mathbb{R} \rightarrow [0, 1]$.

(1) Der **Erwartungswert** $E(X)$ von X ist definiert durch

$$E(X) = \sum_i x_i \cdot W(X = x_i) = \sum_i x_i \cdot f(x_i) = \sum_i f(x_i) \cdot x_i. \quad (10.1)$$

(2) Die **Varianz** $\text{Var}(X)$ von X ist definiert durch

$$\text{Var}(X) = \sum_i (x_i - E(X))^2 \cdot f(x_i) = \sum_i f(x_i) \cdot (x_i - E(X))^2 = \sum_i f(x_i) \cdot x_i^2 - (E(X))^2. \quad (10.2)$$

(3) Die **Standardabweichung** ist definiert durch $\sigma_X = \sqrt{\text{Var}(X)}$.

Dass die letzte Formel für die Varianz den gleichen Wert wie die beiden vorhergehenden liefert, wird auf einem Übungszettel nachgerechnet.

Wir beobachten die **Analogie zwischen dem Erwartungswert (10.1) und dem arithmetischen Mittel \bar{x} eines Merkmals X** , welche durch die Farben gekennzeichnet ist:

$$\bar{x} = \sum_i f_i \cdot x_i = \sum_i \frac{h_i}{N} \cdot x_i$$

Wir beobachten die **Analogie zwischen der Varianz (10.2) und der Varianz eines Merkmals X** , welche durch die Farben gekennzeichnet ist:

$$\text{Var} = \sigma^2 = \sum_i f_i \cdot (x_i - \bar{x})^2 = \sum_i f_i \cdot x_i^2 - \bar{x}^2$$

Dabei entsprechen die Wahrscheinlichkeiten $f(x_i) = W(X = x_i)$ jeweils den relativen Häufigkeiten $f_i = h_i/N$ der Merkmalswerte x_i .

Betrachten wir zunächst ein Beispiel.

Beispiel 10.20. (Erwartungswert und Varianz – dreimaliges Werfen einer idealen Münze⁴)

Unser Zufallsexperiment ist das dreimalige Werfen einer idealen Münze mit den Seiten Kopf (K) und Zahl (Z), und als Zufallsvariable X betrachten wir $X =$ „Anzahl der geworfenen Zahlseiten“. In Beispiel 10.7 hatten wir die Wahrscheinlichkeitsfunktion f von X berechnet:

$$f : \mathbb{R} \rightarrow [0, 1], \quad f(x) = \begin{cases} \frac{1}{8} = 0,125 & \text{wenn } x = x_1 = 0, \\ \frac{3}{8} = 0,375 & \text{wenn } x = x_2 = 1, \\ \frac{3}{8} = 0,375 & \text{wenn } x = x_3 = 2, \\ \frac{1}{8} = 0,125 & \text{wenn } x = x_4 = 3, \\ 0 & \text{wenn } x \in \mathbb{R} \setminus \{0, 1, 2, 3\}. \end{cases}$$

Damit finden wir für den *Erwartungswert*

$$\begin{aligned} E(X) &= \sum_{i=1}^4 x_i f(x_i) = 0 \cdot f(0) + 1 \cdot f(1) + 2 \cdot f(2) + 3 \cdot f(3) \\ &= 0 \cdot 0,125 + 1 \cdot 0,375 + 2 \cdot 0,375 + 3 \cdot 0,125 \\ &= 0 + 0,375 + 0,750 + 0,375 = 1,5. \end{aligned}$$

⁴Dieses Beispiel ist durch ein Beispiel des dreifachen Münzwurfs in [B12, Kapitel 7] inspiriert und orientiert sich in seiner Darstellung an diesem Beispiel.

Dieses Ergebnis ist nicht überraschend, denn wegen des symmetrischen Stabdiagramms (vgl. Abbildung 10.1) erwarten wir das arithmetische Mittel von $x_2 = 1$ und $x_3 = 2$, also 1,5 als Wert von $E(X)$.

Für die *Varianz* finden wir

$$\begin{aligned}\text{Var}(X) &= \sum_{i=1}^4 f(x_i) \cdot x_i^2 - (E(X))^2 \\ &= f(0) \cdot 0^2 + f(1) \cdot 1^2 + f(2) \cdot 2^2 + f(3) \cdot 3^2 - (1,5)^2 \\ &= 0,125 \cdot 0 + 0,375 \cdot 1 + 0,375 \cdot 4 + 0,125 \cdot 9 - 2,25 \\ &= 0 + 0,375 + 1,5 + 1,125 - 2,25 = 3 - 2,25 = 0,75.\end{aligned}$$

Der *Erwartungswert* für die Anzahl der Zahlseiten beim dreimaligen Werfen einer perfekten Münze ist also $E(X) = 1,5$ Zahlseiten, und die *Varianz* ist $\text{Var}(X) = 0,75$ Zahlseiten². Die *Standardabweichung* beträgt $\sigma = \sqrt{0,75} \approx 0,87$ Zahlseiten.

10.5 Erwartungswert und Varianz einer stetigen Zufallsvariablen

Bei einer stetigen (oder kontinuierlichen) Zufallsvariablen können wir ebenfalls den Erwartungswert und die Varianz definieren. Die Formeln sind analog zu der Situation bei einer diskreten Zufallsvariablen, wobei **aus den Summen nun Integrale werden** und die Wahrscheinlichkeitsfunktion durch die Wahrscheinlichkeitsdichte ersetzt werden muss.

Definition 10.21. (Erwartungswert und Varianz einer stetigen Zufallsvariablen)

Sei S der Ereignisraum eines Zufallsexperiments, und sei $X : S \rightarrow B \subseteq \mathbb{R}$ eine **stetige Zufallsvariable** mit der Wahrscheinlichkeitsdichte (oder Dichtefunktion) $f : \mathbb{R} \rightarrow \mathbb{R}_0^+$.

(1) Der **Erwartungswert** $E(X)$ von X ist definiert durch

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx,$$

und falls $f(x) = 0$ für $x \in \mathbb{R} \setminus [x_u, x_o]$ ist, gilt

$$E(X) = \int_{x_u}^{x_o} x \cdot f(x) \, dx.$$

(2) Die **Varianz** $\text{Var}(X)$ von (X) ist definiert durch

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 \cdot f(x) \, dx = \left(\int_{-\infty}^{\infty} x^2 \cdot f(x) \, dx \right) - (E(X))^2,$$

und falls $f(x) = 0$ für $x \in \mathbb{R} \setminus [x_u, x_o]$ ist, gilt

$$\text{Var}(X) = \int_{x_u}^{x_o} (x - E(X))^2 \cdot f(x) \, dx = \left(\int_{x_u}^{x_o} x^2 \cdot f(x) \, dx \right) - (E(X))^2.$$

(3) Die **Standardabweichung** ist definiert durch $\sigma_X = \sqrt{\text{Var}(X)}$.

Um die **Analogie zum Fall einer diskreten Zufallsvariablen** deutlich zu machen, sind die entsprechenden Terme analog zu Definition 10.19 eingefärbt.

Betrachten wir ein Beispiel.

Beispiel 10.22. (Erwartungswert und Varianz – Abweichungen der Fahrzeit eines Airport-Shuttle-Trains)

Ein Airport-Shuttle-Train verbindet Terminal A und Terminal B eines Flughafens. Die vorgesehene Fahrzeit von einem Terminal zum nächsten ist 15 Minuten. Die *stetige Zufallsvariable* $X =$ „Abweichungen von der vorgesehenen Fahrzeit (in min)“ habe die folgende *Wahrscheinlichkeitsdichte*:

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad f(x) = \begin{cases} 0 & \text{für } x < -1 \\ \frac{2}{3} + \frac{2}{3} \cdot x & \text{für } -1 \leq x < 0, \\ \frac{2}{3} - \frac{1}{3} \cdot x & \text{für } 0 \leq x < 2, \\ 0 & \text{für } x \geq 2. \end{cases}$$

Wir berechnen den *Erwartungswert*:

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^2 x f(x) dx \\ &= \int_{-1}^0 x \cdot \left(\frac{2}{3} + \frac{2}{3} \cdot x \right) dx + \int_0^2 x \cdot \left(\frac{2}{3} - \frac{1}{3} \cdot x \right) dx \\ &= \int_{-1}^0 \left(\frac{2}{3} \cdot x + \frac{2}{3} \cdot x^2 \right) dx + \int_0^2 \left(\frac{2}{3} \cdot x - \frac{1}{3} \cdot x^2 \right) dx \\ &= \left[\frac{1}{3} \cdot x^2 + \frac{2}{9} \cdot x^3 \right]_{-1}^0 + \left[\frac{1}{3} \cdot x^2 - \frac{1}{9} \cdot x^3 \right]_0^2 \\ &= 0 - \left[\frac{1}{3} \cdot (-1)^2 + \frac{2}{9} \cdot (-1)^3 \right] + \left[\frac{1}{3} \cdot 2^2 - \frac{1}{9} \cdot 2^3 \right] - 0 \\ &= -\frac{1}{3} + \frac{2}{9} + \frac{4}{3} - \frac{8}{9} = 1 - \frac{2}{3} = \frac{1}{3}. \end{aligned}$$

Der *Erwartungswert* der Abweichungen von der vorgesehenen Fahrzeit ist $E(X) = 1/3 \approx 0,333$ Minuten.

Nun berechnen wir die *Varianz*:

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2 = \int_{-1}^2 x^2 f(x) dx - \left(\frac{1}{3} \right)^2 \\ &= \int_{-1}^0 x^2 \cdot \left(\frac{2}{3} + \frac{2}{3} \cdot x \right) dx + \int_0^2 x^2 \cdot \left(\frac{2}{3} - \frac{1}{3} \cdot x \right) dx - \frac{1}{9} \end{aligned}$$

$$\begin{aligned} &= \int_{-1}^0 \left(\frac{2}{3} \cdot x^2 + \frac{2}{3} \cdot x^3 \right) dx + \int_0^2 \left(\frac{2}{3} \cdot x^2 - \frac{1}{3} \cdot x^3 \right) dx - \frac{1}{9} \\ &= \left[\frac{2}{9} \cdot x^3 + \frac{2}{12} \cdot x^4 \right]_{-1}^0 + \left[\frac{2}{9} \cdot x^3 - \frac{1}{12} \cdot x^4 \right]_0^2 - \frac{1}{9} \\ &= 0 - \left[\frac{2}{9} \cdot (-1)^3 + \frac{2}{12} \cdot (-1)^4 \right] + \left[\frac{2}{9} \cdot 2^3 - \frac{1}{12} \cdot 2^4 \right] - 0 - \frac{1}{9} \\ &= \frac{2}{9} - \frac{2}{12} + \frac{16}{9} - \frac{16}{12} - \frac{1}{9} = \frac{17}{9} - \frac{3}{2} = \frac{34 - 27}{18} = \frac{7}{18} \approx 0,389. \end{aligned}$$

Die *Varianz* (Streuung) beträgt $\text{Var}(X) = 7/18 \approx 0,389$ Minuten², und die *Standardabweichung* beträgt $\sigma_X = \sqrt{7/18} \approx 0,624$ Minuten.

Summen-Notation und Rechnen mit Summen

Hier erklären wir die Summen-Notation und das Rechnen mit Summen.

Definition A.1. (Summen-Notation)

Die Summe der Zahlen $a_1, a_2, \dots, a_n \in \mathbb{R}$ schreibt man mit dem **Summenzeichen**:

$$\sum_{k=1}^n a_k = a_1 + a_2 + \dots + a_n. \quad (\text{A.1})$$

Wir nennen k den sogenannten **Summationsindex**, und der kleinste Wert des Summationsindexes (also 1 in (A.1)) wird als **untere Grenze** des Summationsindexes und der größte Wert des Summationsindexes (also n in (A.1)) wird also **obere Grenze** des Summationsindexes bezeichnet. Der Summationsindex ist frei wählbar und hat keine Bedeutung für den Wert der Summe, d.h.

$$\sum_{k=1}^n a_k = \sum_{j=1}^n a_j.$$

Wir können mit dem Summenzeichen auch Teile der gewählten Zahlen aufsummieren:

$$\sum_{k=\ell}^m a_k = a_\ell + a_{\ell+1} + \dots + a_m \quad \text{mit } 1 \leq \ell \leq m \leq n.$$

Eine Summe, deren obere Grenze des Summationsindexes kleiner ist als deren untere Grenze, wird **leere Summe** genannt. Wir definieren die leere Summe als Summe ohne Summanden und setzen formal

$$\sum_{k=m}^n a_k = 0 \quad \text{für } n < m.$$

Verdeutlichen wir uns die Summennotation an zwei Beispielen.

Beispiel A.2. (Summen-Notation)

Seien $a_1 = 1, a_2 = 2, \dots, a_k = k, \dots, a_n = n$. Dann gilt:

$$\begin{aligned}\sum_{k=1}^n a_k &= \sum_{k=1}^n k = 1 + 2 + \dots + n, \\ \sum_{k=1}^4 a_k &= \sum_{k=1}^4 k = 1 + 2 + 3 + 4 = 10, \\ \sum_{k=2}^5 a_k &= \sum_{k=2}^5 k = 2 + 3 + 4 + 5 = 14.\end{aligned}$$

Beispiel A.3. (Summen-Notation)

Sei $a_k = k^2$ für alle $k \in \mathbb{N}$. Dann gilt:

$$\begin{aligned}\sum_{k=1}^5 a_k &= \sum_{k=1}^5 k^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 1 + 4 + 9 + 16 + 25 = 55, \\ \sum_{k=10}^{10} a_k &= a_{10} = 10^2 = 100.\end{aligned}$$

Indem man die Summen in dem nachfolgenden Lemma ausschreibt (und dies sollte man zur Übung wirklich tun!), erhält man die folgenden Rechenregeln für Summen.

Lemma A.4. (Rechenregeln für Summen)

Es gelten folgende Rechenregeln für Summen:

$$\sum_{k=m}^n a_k + \sum_{k=m}^n b_k = \sum_{k=m}^n (a_k + b_k), \quad (\text{A.2})$$

$$\sum_{k=m}^n a_k - \sum_{k=m}^n b_k = \sum_{k=m}^n (a_k - b_k), \quad (\text{A.3})$$

$$\sum_{k=m}^n c a_k = c \sum_{k=m}^n a_k \quad \text{für jedes } c \in \mathbb{R},$$

$$\sum_{k=m}^n a_k + \sum_{k=n+1}^p a_k = \sum_{k=m}^p a_k, \quad \text{wenn } m \leq n < p. \quad (\text{A.4})$$

Formel (A.4) besagt, dass wir die Summe in zwei Teilsummen zerlegen können. Man kann bei einer Summe auch den Summationsindex um p nach rechts oder links verschieben:

$$\sum_{k=m}^n a_k = \sum_{\ell=m+p}^{n+p} a_{\ell-p} \quad (\text{Indexverschiebung nach rechts}), \quad (\text{A.5})$$

$$\sum_{k=m}^n a_k = \sum_{\ell=m-p}^{n-p} a_{\ell+p} \quad (\text{Indexverschiebung nach links}). \quad (\text{A.6})$$

Formal werden die Indexverschiebungen (A.5) bzw. (A.6) durchgeführt, indem man den neuen Summationsindex $\ell = k + p$ (Indexverschiebung nach rechts) bzw. $\ell = k - p$ (Indexverschiebung nach links) einführt und damit $k = \ell - p$ (Indexverschiebung nach rechts) bzw. $k = \ell + p$ (Indexverschiebung nach links) erhält und entsprechend ersetzt. In (A.5) erhält man für den neuen Summationsindex $\ell = k + p$ die neue untere bzw. obere Grenze $m + p$ bzw. $n + p$, und der Index k in a_k wird durch $k = \ell - p$ ersetzt. Bei (A.6) geht man analog vor.

Betrachten wir zwei Beispiele, in denen die Rechenregeln für Summen aus Lemma A.4 angewendet werden.

Beispiel A.5. (Rechnen mit Summen)

$$\sum_{k=1}^n k - \sum_{k=1}^n (k-1) = \sum_{k=1}^n [k - (k-1)] = \sum_{k=1}^n 1 = n.$$

Wie man sieht, ist die Berechnung durch die Regel (A.3) für die Subtraktion von Summen erheblich vereinfacht worden.

Beispiel A.6. (Rechnen mit Summen) Beim Berechnen von

$$\sum_{k=1}^n k^2 - \sum_{k=1}^n (k+1)^2$$

bemerken wir zuerst, dass die Terme hinter dem jeweiligen Summenzeichen durch das Ersetzen von k durch $k+1$ ineinander überführt werden können. Daher führen wir in der zweiten Summe die Indexverschiebung $\ell = k+1$ (vgl. (A.5)) durch und erhalten die neue untere Grenze $1+1=2$ bzw. die neue obere Grenze $n+1$. Anschließend benennen wir ℓ wieder in k um.

$$\sum_{k=1}^n k^2 - \sum_{k=1}^n (k+1)^2 = \sum_{k=1}^n k^2 - \sum_{\ell=2}^{n+1} \ell^2 = \sum_{k=1}^n k^2 - \sum_{k=2}^{n+1} k^2.$$

Der Unterschied zwischen den beiden Summen besteht nun nur noch in den Grenzen für den Summationsindex. In der ersten Summe wird über $k = 1, 2, \dots, n$ summiert, und in der zweiten Summe wird über $k = 2, \dots, n, n+1$ summiert. Intuitiv ist damit klar, dass bei der Subtraktion beiden Summen genau der erste Term der ersten Summe und der letzte Term der zweiten Summe übrig bleiben. Wir nutzen (A.4), um aus der ersten Summe den Term für $k = 1$ und aus der zweiten Summe den Term mit $k = n+1$ herauszuziehen, und erhalten

$$\sum_{k=1}^n k^2 - \sum_{k=2}^{n+1} k^2 = \left(1^2 + \sum_{k=2}^n k^2\right) - \left(\sum_{k=2}^n k^2 + (n+1)^2\right)$$

$$\begin{aligned} &= 1^2 + \sum_{k=2}^n k^2 - \sum_{k=2}^n k^2 - (n+1)^2 \\ &= 1 - (n+1)^2 + \sum_{k=2}^n (k^2 - k^2) \\ &= 1 - (n+1)^2 \\ &= -n^2 - 2n. \end{aligned}$$

Insgesamt erhalten wir also

$$\sum_{k=1}^n k^2 - \sum_{k=1}^n (k+1)^2 = -n^2 - 2n.$$