

1 Grundlagen der Wahrscheinlichkeitsrechnung

Immer wieder begegnen uns im Alltag Dinge, die die Frage aufwerfen: Ist das nun Zufall, oder gibt es einen Grund dafür, dass das, was wir gerade erlebt haben, genau so passiert ist? Folgende Beispiele sollen dies verdeutlichen:

- Obwohl eine Universität von mehreren Tausend Studentinnen und Studenten besucht wird, gibt es 10 Tage im Jahr, an denen niemand Geburtstag hat.
Und: In einem Seminar hat keiner im August Geburtstag, aber zwei am 1. Mai.
Zufall oder nicht?
- Beim Mensch-ärgere-dich-nicht-Spielen muss manch einer 10-mal würfeln, um eine 6 zu erhalten.
Zufall oder Ungeschicklichkeit?
- Noch kurz vor den Landtagswahlen in Hessen 2008 hatten Meinungsforscher den Sieg der Regierungspartei vorausgesagt. Tatsächlich reichte das Wahlergebnis aber nicht aus.
Stimmungswechsel in letzter Minute oder haben die Befragungsmethoden versagt?

Bevor man auf diese Fragen Antworten geben kann, wird untersucht, ob und inwieweit sich die jeweiligen Ereignisse mathematisch beschreiben lassen. Es können Modelle entwickelt und Aussagen vorgenommen werden. Als Erstes werden Zufälle an Zufallsversuchen einfachster Art untersucht, wie z. B. Münzwürfen oder Würfel-ergebnissen.

„Der Zufall hat kein Gedächtnis“, formulierte einst der Mathematiker Bertrand (1889). Es gibt aber eine Reihe von Gesetzmäßigkeiten, die für Zufallsvorgänge gelten.

Definition

Ein Vorgang, der theoretisch beliebig oft wiederholt werden kann und dessen Ergebnis nicht mit Sicherheit vorauszusagen ist, heißt **Zufallsversuch**.

Die Anzahl der Stufen (Durchgänge) des Zufallsversuchs wird **Stichprobenumfang** genannt.

Die Menge aller möglichen Ergebnisse eines Zufallsversuchs wird als **Ergebnismenge des Zufallsversuchs S** bezeichnet.

Ergebnismenge eines Zufallsversuchs – Beispiele

- (1) Ein Würfel wird geworfen
- (2) Eine Münze wird geworfen
- (3) Ein Würfel und eine Münze werden gleichzeitig geworfen

Lösung:

zu (1)

Es können die Augenzahlen 1, 2, 3, 4, 5 oder 6 auftreten. Somit ist $S = [1; 2; 3; 4; 5; 6]$. Interessiert uns aber nur, ob die Augenzahl 3 fällt oder nicht, begrenzt sich die Ergebnismenge auf $S = [3, \text{nicht } 3]$.

zu (2)

Es kann Wappen oder Zahl oben liegen, somit ist die Ergebnismenge $S = [W; Z]$.

zu (3)

Die Ergebnisse des Münzwurfs ($W; Z$) können in beliebiger Kombination mit den Ergebnissen des Würfelversuchs auftreten:

$S = [W; 1], [W; 2], [W; 3], [W; 4], [W; 5], [W; 6], [Z; 1], [Z; 2], [Z; 3], [Z; 4], [Z; 5], [Z; 6]$.

1.1 Laplace-Versuche

Bei Zufallsversuchen sind vor allem folgende Fragen interessant:

- Wie häufig treten die einzelnen Ergebnisse auf?
- Lohnt es sich darauf zu wetten, dass ein bestimmtes Ergebnis eher auftritt als ein anderes?
- Welche Prognosen sind für bevorstehende Zufallsversuche möglich?

Definition

Haben bei einem Zufallsversuch mit n möglichen Ergebnissen all diese Ergebnisse dieselbe Chance aufzutreten, dann ordnet man jedem dieser Ergebnisse die Wahrscheinlichkeit $p = \frac{1}{n}$ zu. Solche Versuche werden als **Laplace-Versuche** bezeichnet.

Dem Münzwurf ordnen wir die Wahrscheinlichkeit $p = \frac{1}{2}$ zu, beim Würfeln ist die Wahrscheinlichkeit für jede Augenzahl $p = \frac{1}{6}$, bei einem Glücksrad mit 12 gleichen Sektoren ist die Wahrscheinlichkeit für jeden Sektor $p = \frac{1}{12}$.

Beispiel

Zwei gleichartige Münzen werden gleichzeitig geworfen. Welche Ergebnisse sind möglich?

Auch wenn zwei Münzen gleichartig sind, können sie durch eine Markierung unterscheidbar gemacht werden. So wird deutlich, dass es in diesem Fall vier verschiedene, gleich wahrscheinliche Ergebnisse gibt:

$S = [W; W], [Z; Z], [W; Z], [Z; W]$.

Für die Wahrscheinlichkeiten gilt:

$P(WW) = P(ZZ) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Wahrscheinlichkeiten gleichzeitig auftretender Ereignisse werden also miteinander multipliziert.

Für die Wahrscheinlichkeit „einmal Wappen, einmal Zahl“ gilt:

$P(WZ; ZW) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{2}{4} = \frac{1}{2}$. Da es zwei verschiedene Kombinationsmöglichkeiten gibt, ist die Wahrscheinlichkeit doppelt so hoch.

Bei diesem Beispiel wird deutlich, dass es nötig ist, zwischen dem Ergebnis „einmal Wappen und einmal Zahl“ und den Ergebnissen (*WZ*; *ZW*) zu unterscheiden. Da sich „einmal Wappen, einmal Zahl“ aus zwei Ergebnissen zusammensetzt, werden diese als **Ereignis E** bezeichnet.

Laplace-Regel

Eine Summe von Ergebnissen fasst man als Ereignis zusammen. Ereignisse sind somit Teilmengen der Ergebnismenge S eines Zufallsversuchs.

Bei Laplace-Versuchen wird die **Wahrscheinlichkeit $P(E)$** eines Ereignisses E wie folgt berechnet:

$$P(E) = \frac{\text{Anzahl der zum Ereignis } E \text{ gehörenden Ergebnisse}}{\text{Anzahl aller möglichen Ergebnisse}} = \frac{|E|}{|S|}$$

Beispiel 1

Zwei unterscheidbare Würfel werden gleichzeitig geworfen. Man berechne die Wahrscheinlichkeit für folgende Ereignisse:

- die Augensumme beträgt 7
- die Augensumme ist gleich 12
- die Augensumme ist ungerade.

Lösung:

Mit zwei Würfeln sind 36 Kombinationen möglich, also $|S| = 36$.

$$(a) |E| = [(1; 6), (2; 5), (3; 4), (4; 3), (5; 2), (6; 1)] = 6$$

$$\text{Somit ist } P(E) = \frac{|E|}{|S|} = \frac{6}{36} = \frac{1}{6}$$

$$(b) |E| = [(6; 6)] = 1 \text{ und somit } P(E) = \frac{1}{36}$$

$$(c) |E| = [(1;2), (1;4), (1;6), (2;1), (2;3), (2;5), (3;2), (3;4), (3;6), (4;1), (4;3), (4;5), (5;2), (5;4), (5;6), (6;1), (6;3), (6;5)] = 18$$

$$\text{Somit ist } P(E) = \frac{18}{36} = \frac{1}{2}$$

Beispiel 2

In einem Flugzeug befinden sich 24 Deutsche, 12 Italiener und 18 Spanier. Beim Zoll wird zufällig eine Person ausgewählt. Wie groß ist die Wahrscheinlichkeit dafür, dass diese Person

- Deutscher
- Italiener
- Spanier
- Deutscher oder Italiener ist?

Lösung:

$$|S| = 54$$

$$(a) P(E) = \frac{24}{54} = \frac{4}{9}$$

$$(b) P(E) = \frac{12}{54} = \frac{2}{9}$$

$$(c) P(E) = \frac{18}{54} = \frac{1}{3}$$

$$(d) P(E) = \frac{36}{54} = \frac{2}{3}$$

1.2 Rechenregeln für Wahrscheinlichkeiten

In Kapitel 1.1 wurde die Laplace-Regel behandelt. Diese besagt für den Fall der Gleichverteilung, dass sich die Wahrscheinlichkeiten für ein Ereignis aus der Anzahl der dazugehörigen Ergebnisse bestimmen lassen. Das bedeutet, die Wahrscheinlichkeiten für die einzelnen Ergebnisse müssen nur addiert werden. Eine solche Summenregel gilt auch für Nicht-Laplace-Versuche.

Um z. B. die Wahrscheinlichkeit für die Fächerbelegung von Studenten und Studentinnen an Universitäten zu errechnen, gibt es in der Mathematik die Wahrscheinlichkeitsgrundsätze, die sich aus der allgemeinen Summenregel und der Komplementärregel zusammensetzen. Im Folgenden werden zuerst die Regeln erklärt und dann anhand von Beispielen verdeutlicht.

Man bezeichnet Ereignisse, deren Ergebnisse zu einem **oder** zu einem anderen Ereignis gehören als **Oder-Ereignisse**. Ereignisse, deren Ergebnisse zum einen **und** zum anderen Ereignis gehören, werden als **Und-Ereignisse** bezeichnet. Das Symbol \cup steht für Oder-Ereignisse, das Symbol \cap für Und-Ereignisse.

1.2.1 Allgemeine Summenregel

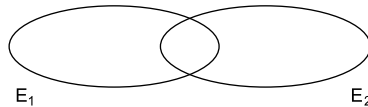
Setzt sich ein Ereignis E aus zwei Ereignissen E_1 und E_2 zusammen, die sich überschneiden, bzw. gemeinsame Ergebnisse enthalten, dann muss darauf geachtet werden, dass die gemeinsamen Ergebnisse nicht doppelt gewertet werden.

Allgemeine Summenregel

Für $E = E_1 \cup E_2$ gilt $P(E) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$

anders ausgedrückt:

Die Wahrscheinlichkeit eines Oder-Ereignisses $E_1 \cup E_2$ ist gleich der Summe der Wahrscheinlichkeiten der Ereignisse E_1 und E_2 , vermindert um die Wahrscheinlichkeit des Und-Ereignisses $E_1 \cap E_2$.



$E_1 \cap E_2$ ist das UND-Ereignis, $E_1 \cup E_2$ das Oder-Ereignis der Ereignisse E_1 und E_2 .

1.2.2 Komplementärregel

Schließen sich zwei Ereignisse E_1 und E_2 gegenseitig aus und ergänzen sich so, dass es kein Ergebnis des Zufallsversuchs gibt, welches zu E_1 oder E_2 gehört, addieren sich die Wahrscheinlichkeiten der beiden Ergebnisse zu 1.

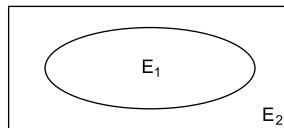
Komplementärregel

Sind $E_1 \cap E_2 = \{ \}$ (unmögliches Ereignis) und $E_1 \cup E_2 = S$ (sicheres Ereignis), dann gilt:

$$P(E_1) + P(E_2) = 1$$

anders ausgedrückt:

Die Wahrscheinlichkeit eines Ereignisses E und die Wahrscheinlichkeit des Gegenereignisses \bar{E} ergänzen sich zu 1.



Sind $E_1 \cap E_2 = \{ \}$ und $E_1 \cup E_2 = S$, so kann man auch schreiben $E_1 = \bar{E}_2$ oder $E_2 = \bar{E}_1$.

Das bedeutet, dass E_1 und E_2 zueinander Gegenereignisse sind.

Beispiel

In einer Urne befinden sich 40 gleichartige Kugeln, die von 1 bis 40 nummeriert sind. Eine Kugel wird zufällig gezogen.

Man bestimme die Wahrscheinlichkeit des Ereignisses $E_1 \cup E_2 \cup E_3$, wenn:

E_1 : die Zahl durch 2 teilbar ist, E_2 : die Zahl durch 3 teilbar ist, E_3 : die Zahl durch 5 teilbar ist.

$$P(X = 1) = \binom{6}{1} \cdot 0,4^1 \cdot 0,6^5 \approx 0,1866 = 18,66\%$$

$$P(X \geq 2) = 1 - P(X \leq 1) = 100\% - 23,33\% = 76,67\%$$

2.3 Der Erwartungswert einer Zufallsgröße

Den Erwartungswert der Zufallsgröße X bezeichnet man mit $E(X)$ oder auch μ . Der Erwartungswert einer Zufallsgröße ist ein Wert, den man bei häufiger Versuchsdurchführung als Mittelwert erhält.

Oft ist es sinnvoll, den Erwartungswert mithilfe einer Verteilungstabelle zu berechnen.

Definition 1

Eine Zuordnung X , die jedem Ausgang eines Zufallsversuches eine Zahl (z. B. Auszahlungsbetrag, Anzahl, ...) zuordnet, nennt man **Zufallsgröße**.

Die Zufallsgröße X kann die Werte x_1, x_2, \dots, x_n annehmen.

Beispiel 1

Bei einer Semesterfeier muss jeder der 50 Teilnehmer ein Los kaufen. Der erste Preis hat einen Wert von 15,00 €, der zweite von 10,00 €, der dritte von 4,00 €. Außerdem gibt es noch Trostpreise im Wert von je 0,50 € für alle Teilnehmer.

Wie viel muss ein Los kosten, damit Einnahmen und Ausgaben übereinstimmen?

Lösung:

In die erste Spalte der Tabelle werden alle Werte x_i eingegeben, die die Zufallsgröße X annehmen kann.

Nun werden die Wahrscheinlichkeiten $P(X = x_i)$ berechnet, indem man alle Wahrscheinlichkeiten der möglichen Ereignisse addiert. Diese werden in die zweite Spalte der Tabelle eingetragen.

In die dritte Spalte trägt man nun das Produkt aus x_i und $P(X = x_i)$ ein.

Die Summe der Zahlen der dritten Spalte ergeben den Erwartungswert $E(X)$.

x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
€ 15,00	$\frac{1}{50}$	$\frac{15}{50}$
€ 10,00	$\frac{1}{50}$	$\frac{10}{50}$
€ 4,00	$\frac{1}{50}$	$\frac{4}{50}$
€ 0,50	$\frac{47}{50}$	$\frac{47}{100}$

$$E(X) = = \frac{15}{50} + \frac{10}{50} + \frac{4}{50} + \frac{47}{100} = \frac{105}{100} = 1,05 \text{ €}$$

Der Lospreis für dieses Spiel müsste bei Übereinstimmung von Einnahmen und Ausgaben bei 1,05 € liegen.

Definition 2

Kann eine Zufallsgröße X die Werte x_1, x_2, \dots, x_n , annehmen, so heißt

$$E(X) = \mu = x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + \dots + x_n \cdot P(X = x_n)$$

Erwartungswert der Zufallsgröße.

Beispiel 2

Bei einem Glücksspiel setzt ein Spieler 20,- €. Er wirft einen Würfel. Liegt eine gerade Zahl (g) oben, so halbiert sich sein Einsatz, liegt eine ungerade Zahl (u) oben, so verdoppelt er sich. Welchen Gewinn kann der Spieler erwarten, wenn er insgesamt dreimal würfelt und nach jedem Wurf der noch vorhandene Einsatz halbiert oder verdoppelt wird?

Lösung:

Versuchskombinationen	Geld (x_i) des Spielers	Wahrscheinlichkeit
u u u	160,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
u u g	40,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
u g u	40,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
u g g	10,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
g u u	40,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
g u g	10,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
g g u	10,00 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$
g g g	2,50 €	$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$

Geld (x_i) des Spielers	zugehörige Wahrscheinlichkeit $P(X = x_i)$	$x_i \cdot P(X = x_i)$
160,00 €	$\frac{1}{8}$	20,- €
40,00 €	$\frac{3}{8}$	15,- €
10,00 €	$\frac{3}{8}$	3,75 €
2,50 €	$\frac{1}{8}$	0,3125 €

$$E(X) = \mu = 20 + 15 + 3,75 + 0,3125 = 39,0625 \approx 39,06 \text{ €}$$

Dies bedeutet: Wird das Spiel sehr oft ausgeführt, kann man erwarten, dass der Spieler durchschnittlich 39,06 € pro Spiel gewinnt. Somit ist das Spiel für den Spieler günstig, da er im Mittel 19,06 € pro Spiel verdient.

Ist bei einem Glücksspiel der „Gewinn“-Erwartungswert Null, so wird man auf Dauer weder Geld gewinnen noch verlieren. Ein solches Spiel wird als fair bezeichnet.

2.4 Varianz und Standardabweichung

Wie im vorherigen Kapitel beschrieben, gibt der Erwartungswert $E(X)$ den mittleren zu erwartenden Wert von X an. Der Erwartungswert gibt jedoch keinen Hinweis darauf, wie stark die einzelnen Werte von X bei häufigem Durchführen des Experiments um den Erwartungswert $E(X)$ abweichen („variieren“). Die **Varianz $V(X)$** drückt dies aus.

Beispiel

Es werden zwei verschiedene Glücksspiele angeboten:

Spiel 1: In einer Urne befinden sich drei rote, drei gelbe und vier schwarze Kugeln. Zieht man eine rote Kugel, gewinnt man 6,00 €, zieht man eine gelbe Kugel, verliert man 5,00 €. Beim Ziehen einer schwarzen Kugel gewinnt und verliert man nichts.

Spiel 2: In einer Urne befinden sich drei rote, sechs gelbe und eine schwarze Kugel. Zieht man eine rote Kugel, gewinnt man 11,00 €, zieht man eine gelbe Kugel, verliert man 5,00 €. Beim Ziehen der schwarzen Kugel gewinnt und verliert man nichts.

Der Erwartungswert $E(X)$ und die jeweilige Gewinnwahrscheinlichkeit berechnen sich wie folgt:

Spiel 1:		Spiel 2:	
x_i	$P(X = x_i)$	x_i	$P(X = x_i)$
6,00 €	$\frac{3}{10} = 0,3$	11,00 €	$\frac{3}{10} = 0,3$
- 5,00 €	$\frac{3}{10} = 0,3$	- 5,00 €	$\frac{6}{10} = 0,6$
0,00 €	$\frac{4}{10} = 0,4$	0,00 €	$\frac{1}{10} = 0,1$

Der Erwartungswert für das 1. Spiel berechnet sich aus:

$$E_1(X) = 6 \cdot \frac{3}{10} + (-5) \cdot \frac{3}{10} + 0 \cdot \frac{4}{10} = 0,3$$

Der Erwartungswert für das 2. Spiel berechnet sich wie folgt:

$$E_2(X) = 11 \cdot \frac{3}{10} + (-5) \cdot \frac{6}{10} + 0 \cdot \frac{1}{10} = 0,3$$

Der Erwartungswert ist für beide Spielvarianten gleich.

Um nun die Abweichungen zu berechnen, bietet es sich an, die Differenz zwischen dem Gewinn im einzelnen Spiel, also X , und dem durchschnittlichen Gewinn, also dem Erwartungswert $E(X)$, zu berechnen: $(X - E(X))$

Spiel 1:

Bei einer großen Serie von Spielen wird in
 30 % aller Fälle das Ergebnis um $(6 - 0,3) \text{ €}$
 30 % aller Fälle das Ergebnis um $(-5 - 0,3) \text{ €}$
 40 % aller Fälle das Ergebnis um $(0 - 0,3) \text{ €}$
 vom Erwartungswert $E(X)$ abweichen.

Spiel 2:

Bei einer großen Serie von Spielen wird in
 30 % aller Fälle das Ergebnis um $(11 - 0,3) \text{ €}$
 60 % aller Fälle das Ergebnis um $(-5 - 0,3) \text{ €}$
 10 % aller Fälle das Ergebnis um $(0 - 0,3) \text{ €}$
 von $E(X)$ abweichen.

In einigen Fällen – wie auch in diesem Beispiel – würden sich die Abweichungen größer Null und kleiner Null jedoch teilweise gegenseitig aufheben. Deshalb rechnet man mit den Quadraten der Abweichungen: $(X - E(X))^2$.

Das Maß für die „Abweichung“, also die Varianz, stellt sich somit für beide Spiele wie folgt dar:

$$V_1(X) = (6 - 0,3)^2 \cdot 0,3 + (-5 - 0,3)^2 \cdot 0,3 + (0 - 0,3)^2 \cdot 0,4 = 18,2$$

$$V_2(X) = (11 - 0,3)^2 \cdot 0,3 + (-5 - 0,3)^2 \cdot 0,6 + (0 - 0,3)^2 \cdot 0,1 = 51,2$$

Man kann erkennen, dass bei Spiel 2 die zu erwartende „Abweichung“ um den Wert 0,30 € größer ist als bei Spiel 1. Dies bedeutet, dass die Gewinn- und Verlustschwankungen bei Spiel 2 größer sind.

Wichtig: Beschreibt eine Zufallsgröße X z. B. bei einem Glücksspiel den Gewinn in €, so ist $V(X)$ ein Maß bezüglich €². Somit ist es sinnvoll, die Standardabweichung von X mit $\sqrt{V(X)}$ einzuführen.

Definition

Kann eine Zufallsgröße X die Werte x_1, x_2, \dots, x_n annehmen, so heißt $V(X)$ mit

$$V(X) = (x_1 - E(X))^2 \cdot P(X=x_1) + (x_2 - E(X))^2 \cdot P(X=x_2) + \dots + (x_n - E(X))^2 \cdot P(X=x_n)$$

Varianz von X .

$\sqrt{V(X)}$ heißt **Standardabweichung von X** .

Anstelle von $V(X)$ sagt man auch σ^2 , und anstelle von $\sqrt{V(X)}$ sagt man auch σ .

Die Standardabweichung ist ein Maß dafür, wie groß die Aussagekraft des Mittelwertes ist. Eine kleine Standardabweichung bedeutet, alle Beobachtungswerte liegen nahe am Mittelwert. Eine große Standardabweichung bedeutet, die Beobachtungswerte sind weit um den Mittelwert „gestreut“.

Beispiel 1

Zwei Tetraeder werden gleichzeitig geworfen. Nach jedem Wurf wird die Augensumme notiert. Man bestimme die Standardabweichung dieses Versuchs.

Lösung:

Die Zufallsgröße X gibt die jeweilige Augensumme der Tetraeder an. X kann also die Werte 2, 3, 4, 5, 6, 7 und 8 annehmen.

Die Wahrscheinlichkeiten für die Augensumme ergeben sich wie folgt:

Augensumme (x_i)	Wahrscheinlichkeit $P(X = x_i)$
2	$\frac{1}{16}$
3	$\frac{2}{16}$
4	$\frac{3}{16}$
5	$\frac{4}{16}$
6	$\frac{3}{16}$
7	$\frac{2}{16}$
8	$\frac{1}{16}$

Der Erwartungswert μ errechnet sich somit zu:

$$\mu = 2 \cdot \frac{1}{16} + 3 \cdot \frac{2}{16} + 4 \cdot \frac{3}{16} + 5 \cdot \frac{4}{16} + 6 \cdot \frac{3}{16} + 7 \cdot \frac{2}{16} + 8 \cdot \frac{1}{16} = 5$$

Die Varianz ergibt sich aus:

$$V(X) = \sigma^2 = (2-5)^2 \cdot \frac{1}{16} + (3-5)^2 \cdot \frac{2}{16} + (4-5)^2 \cdot \frac{3}{16} + (5-5)^2 \cdot \frac{4}{16} \\ + (6-5)^2 \cdot \frac{3}{16} + (7-5)^2 \cdot \frac{2}{16} + (8-5)^2 \cdot \frac{1}{16} = 2,3125.$$

Für die Standardabweichung gilt: $\sigma = \sqrt{V(X)} = \sqrt{2,3125} \approx 1,52$.

Beispiel 2

Auf einer Abi-Fete soll bei einem Glücksspiel Geld für den bevorstehenden Abi-Ball gesammelt werden. In einem Kasten befinden sich 5 schwarze und 3 goldene Kugeln.

Ein Schüler zieht eine Kugel. Es stehen zwei Spielvarianten zur Verfügung: Bei Variante 1 gewinnt der Spieler 7,00 €, wenn es eine goldene Kugel ist, ansonsten verliert er 5,00 €. Bei Variante 2 gewinnt der Spieler 2,00 €, wenn die Kugel golden ist, ansonsten verliert er 2,00 €.

(a) Wie groß sind die zu erwartenden Einnahmen pro Spiel?

(b) Man bestimme die entsprechenden Standardabweichungen.

Lösung:

X beschreibt die Einnahmen pro Spiel.

$$(a) \text{ Variante 1: } E(X) = \mu = 5 \cdot \frac{5}{8} + (-7) \cdot \frac{3}{8} = \frac{1}{2}$$

$$\text{Variante 2: } E(X) = \mu = 2 \cdot \frac{5}{8} + (-2) \cdot \frac{3}{8} = \frac{1}{2}$$

$$(b) \text{ Variante 1: } V(X) = \left(5 - \frac{1}{2}\right)^2 \cdot \frac{5}{8} + \left(-7 - \frac{1}{2}\right)^2 \cdot \frac{3}{8} = 33\frac{3}{4}$$

$$\sigma = \sqrt{V(X)} = \sqrt{33\frac{3}{4}} \approx 5,8$$

$$\text{Variante 2: } V(X) = \left(2 - \frac{1}{2}\right)^2 \cdot \frac{5}{8} + \left(-2 - \frac{1}{2}\right)^2 \cdot \frac{3}{8} = 3\frac{3}{4}$$

$$\sigma = \sqrt{V(X)} = \sqrt{3\frac{3}{4}} \approx 1,9$$

Bei beiden Varianten ist der zu erwartende Gewinn pro Spiel gleich groß, bei Variante 1 sind die „Gewinnschwankungen“ jedoch größer.

5 Grafische Darstellung statistischer Daten

Statistische Daten begegnen uns im Alltag immer wieder. Nur allzu bekannt ist die berühmte Sonntagsfrage: „Wen würden Sie wählen, wenn am nächsten Sonntag Bundestagswahlen wären?“. Die Antwort auf diese Frage, die einem repräsentativen Querschnitt der Bevölkerung gestellt wurde, wird uns allerdings nicht als Zahlenkolonne präsentiert, sondern als übersichtliches Kreis- oder Säulendiagramm. Die Ergebnisse von Umfragen oder Zufallsexperimenten können in Form eines Diagramms übersichtlich und anschaulich präsentiert werden und dadurch auch das Interesse des mathematisch weniger Aufgeschlossenen erwecken.

Heutzutage erleichtern diverse Computerprogramme das Zeichnen solcher Diagramme aus einem Datensatz. Bei der Vielfalt der Möglichkeiten, die solche Programme bieten, stellt sich aber auch schnell die Frage, welches Diagramm für welchen Zweck sinnvoll ist. Dazu soll dieses Kapitel einen Überblick geben. Grundsätzlich gilt, dass innerhalb einer Publikation nicht zu viele unterschiedliche Diagrammtypen verwendet werden sollten. Nicht alles, was technisch möglich ist, macht auch Sinn. Oberstes Gebot sollte sein, dass ein Diagramm einfach, klar und aussagekräftig ist und dem Leser einen leichten und einprägsamen Zugang zur Thematik erlaubt, anstatt ihn durch unnötige Kompliziertheit abzuschrecken.

5.1 Linien- und Kurvendiagramme

Liniendiagramme können Trends und Tendenzen gut darstellen. Sie kommen dann infrage, wenn eine große Menge an Daten vorliegt, die in einem z. B. zeitlichen Zusammenhang stehen. Somit wird durch Liniendiagramme der Zusammenhang zwischen zwei Merkmalen visualisiert. Kurvendiagramme werden beispielsweise verwendet, um den Gewinn eines Unternehmens über einen bestimmten Zeitraum

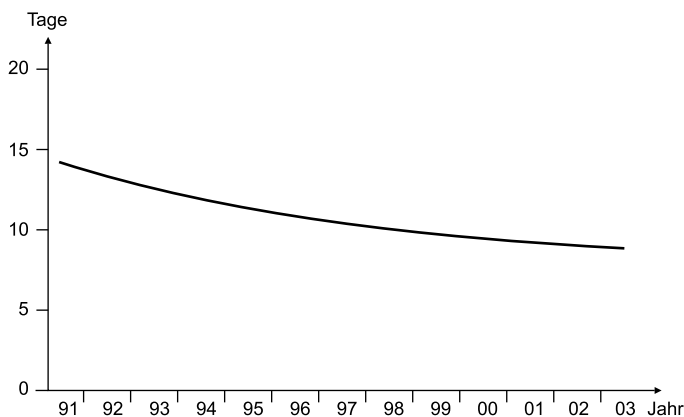


Abb. 5.1:

Durchschnittliche Verweildauer der Patienten im Krankenhaus (in Tagen).

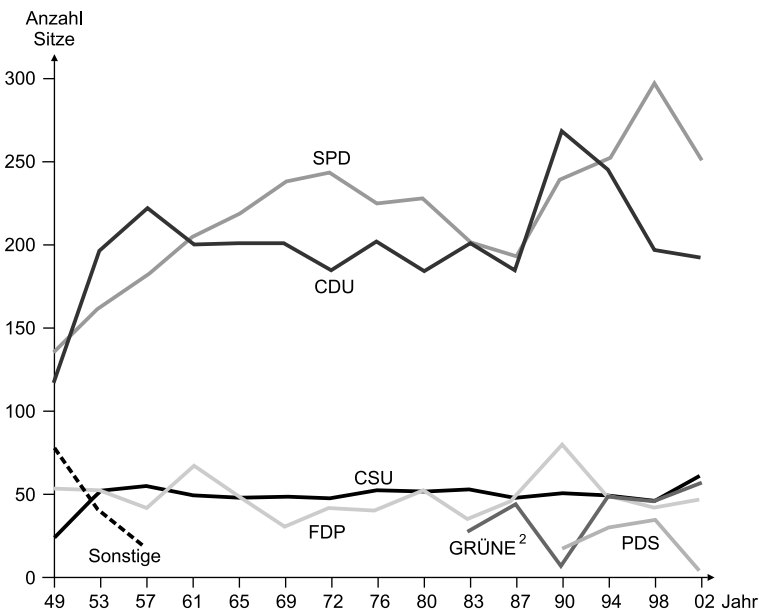
Quelle: Statistisches Bundesamt: Statistisch gesehen 2005.

darzustellen, oder den Wechselkurs des Britischen Pfunds zum Euro, die Entwicklung der Arbeitslosenzahlen oder die durchschnittliche Verweildauer von Patienten im Krankenhaus wie in Abbildung 5.1.

Diese Abbildung zeigt zugleich die richtige Beschriftung eines Diagramms: Ein kurzer knapper Text gibt Auskunft über das Thema der Grafik und über die Maßeinheit, in diesem Fall Tage. Die Abszisse (x -Achse) ist die horizontale Achse, die Ordinate (y -Achse) die vertikale Achse. Wichtig ist die Beschriftung der Achsen. Diagramme ohne Achsenbeschriftung sind wertlos!

Diese einfache Form des Liniendiagramms kann für viele Datenreihen verwendet werden. Wichtig ist, dass die Werte dicht aufeinander folgen, sodass bestimmte Entwicklungen nicht wegfallen und ein kontinuierlicher Verlauf suggeriert wird, der womöglich nicht besteht. Die Achseneinteilung muss proportional sein, damit keine Verzerrungen entstehen, d. h. wenn Werte anfangs im jährlichen Rhythmus angegeben werden, können sie am Ende der Grafik nicht auf einen monatlichen Rhythmus wechseln. Ebenfalls ist auf sinnvolle Abstände der Werte zu achten, sodass die Kurven nicht verzerrt oder gestaucht werden.

Außer solchen einfachen Liniendiagrammen kann man auch mehrfache Kurvendiagramme nutzen, um Vergleiche ziehen zu können. In Abbildung 5.2 ist die Sitzverteilung in der Bundesrepublik Deutschland seit 1949 dargestellt. Gut zu erkennen



1 Bis 1987 früheres Bundesgebiet einschl. der Abgeordneten von Berlin-West, ab 1990 Deutschland.
2 Bis 1987: GRÜNE; Bündnis 90/Grüne (nur neue Bundesländer); 1994: BÜNDNIS 90/DIE GRÜNEN

Abb. 5.2:
Sitzverteilung im Deutschen Bundestag.

bei dieser Grafik ist die nahezu konstante Anzahl der Sitze der CSU, während bei der FDP und den Grünen deutliche Schwankungen zu erkennen sind. Bei solchen Mehrfach-Liniendiagrammen gilt es, darauf zu achten, dass die verschiedenen Kurven gut voneinander zu unterscheiden sind, etwa durch Farben oder unterschiedliche Linienformen wie gepunktete und gestrichelte Linien. Dies ist besonders dann wichtig, wenn sich die Kurven mehrfach überschneiden. Zu viele Informationen können eine solche Darstellung deutlich überfrachten. Würde man z. B. den Stimmenanteil der einzelnen Parteien der Weimarer Republik über einen Zeitraum von 1919 bis 1933 darstellen, erhielte man aufgrund der vielen sich zur Wahl stellenden Parteien eine sehr unübersichtliche Grafik.

Mehrfach-Kurvendiagramme können, wenn keine Überschneidungen der einzelnen Kurven vorhanden sind, auch als Flächendiagramme dargestellt werden (Abb. 5.3). Solche Flächendiagramme wirken optisch attraktiver, verdecken aber eventuell vorhandene Gitternetzlinien, wodurch das Ablesen von Daten erschwert wird.

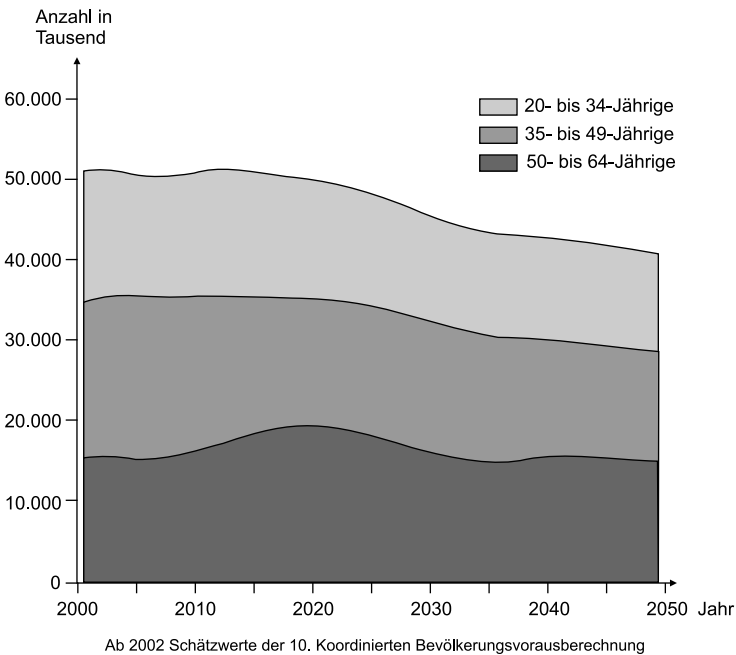


Abb. 5.3:
Altersaufbau der Bevölkerung im Erwerbsalter.

Grundsätzlich sollte das Koordinatenkreuz bei Null beginnen. Sind die Werte jedoch sehr hoch und ist der untere Wertebereich nicht besetzt, bietet es sich an, die Achsen mit höheren Werten beginnen zu lassen. Dies muss durch eine deutliche Unterbrechung der Achse gekennzeichnet werden (Abb. 5.4).

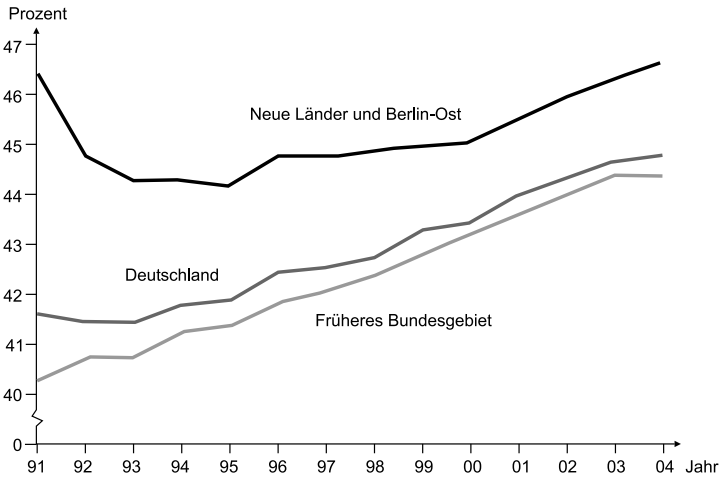


Abb. 5.4: Anteil der erwerbstätigen Frauen in Prozent. Statistisches Bundesamt 2005-15-0161.

Abweichungen um den Mittelwert können durch den Fehlerbalken sichtbar gemacht werden. Dieser wird als Standardabweichung vom Mittelwert nach oben und unten an den Messpunkt eingetragen (Abb. 5.5).

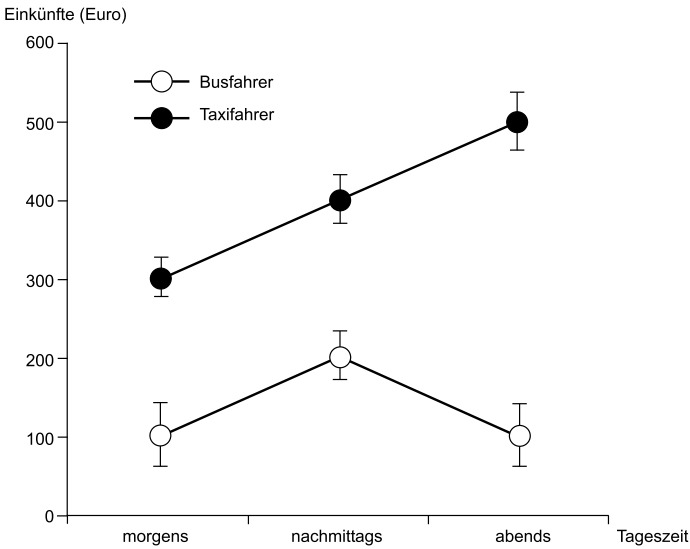


Abb. 5.5: Einkünfte von Bus- und Taxifahrern.

Linien- oder Kurvendiagramme

- Sie eignen sich besonders zur Darstellung von Entwicklungen.
- Aus ihnen sind Tendenzen ablesbar.
- Unterschiede und Ähnlichkeiten sowie parallele und gegenläufige Tendenzen sind leicht ablesbar.
- Sie werden unübersichtlich, wenn die Zahl der Datenreihen zu groß ist.

5.2 Balken- und Säulendiagramme

Säulen- und Balkendiagramme sind einander sehr ähnlich; im Prinzip sind sie nur um 90° zueinander gedreht, bzw. die Anordnung von x- und y-Achse ist vertauscht. Gut geeignet sind sie, um zwei Datenreihen miteinander zu vergleichen. Auch Entwicklungen sind insbesondere bei den Säulendiagrammen abzulesen. Will man den Eindruck einer Tendenz vermeiden, entscheidet man sich eher für das Balkendiagramm.

5.2.1 Balkendiagramme

In Balkendiagrammen sind die einzelnen Balken horizontal orientiert. Die x-Achse ist bei diesem Diagrammtyp die Größenachse. Auf der y-Achse kann ein Zeitraum angegeben sein, aber auch verschiedene Kategorien, wie Staaten, Bundesländer oder Antwortmöglichkeiten einer Befragung. Ein Beispiel für eine nicht zeitlich bezogene Anordnung zeigt Abbildung 5.6. Balkendiagramme ermöglichen den direkten Vergleich von Werten auch bei nur kleinen Unterschieden. Diese Abbildung macht deutlich, dass die Ausgaben pro Schüler bei Sonderschulen mit über 10.000 Euro deutlich über dem Schnitt aller Schularten zusammen liegen, was in diesem Fall dadurch zu begründen ist, dass pro Lehrer weniger Schüler zu betreuen sind. Unter dem Mittelwert aller Schularten liegen berufliche Schulen und Grundschulen.

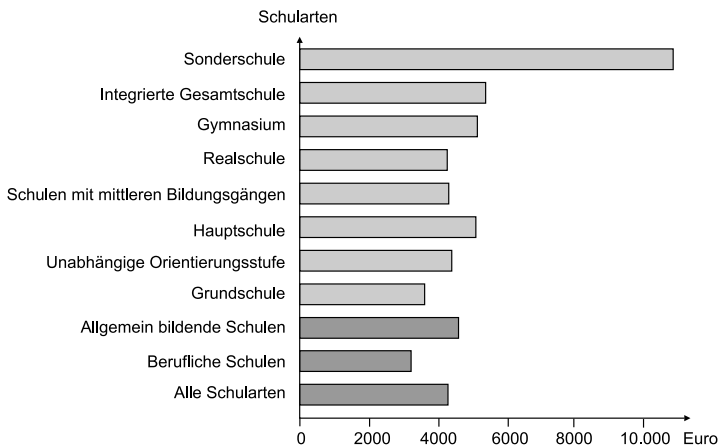


Abb. 5.6: Ausgaben pro Schüler/Schülerin an öffentlichen Schulen 2002 in Euro.

Nun kann man Balkendiagramme auch noch mit mehreren Unterkategorien versehen und so noch mehr Informationen visualisieren. Ein solches Beispiel zeigt Abbildung 5.7. Hier wurde die Gesamtzahl der Personen jeder Kategorie nochmals in Männer und Frauen unterteilt. Außerdem wurde durch Heranziehung weiterer Kategorien auf der rechten Seite die Informationsfülle der Grafik erhöht. Im Prinzip sind hier mehrere Grafiken zu einer zusammengefügt worden; man hätte auch drei Einzelgrafiken mit den Themen Erwerbstätigkeit, Arbeitslosengeld, -hilfe, Renten usw. sowie Unterstützung durch Angehörige erstellen können. In diesem Fall sind als Zusatzinformation die absoluten Zahlen in Tausend noch neben den Balken angegeben. Aus solchen Grafiken könnten natürlich auch Aufgabenstellungen formuliert werden, wie etwa: Wie groß ist die Wahrscheinlichkeit, dass Frau M., die in Erfurt wohnt, erwerbstätig ist? Für eine solche Fragestellung müsste man sich lediglich der Daten der neuen Länder ansehen und zwar die der Erwerbstätigen, der Arbeitslosengeld- bzw. Rentenempfänger und derer, die von der Unterstützung durch Angehörige leben.

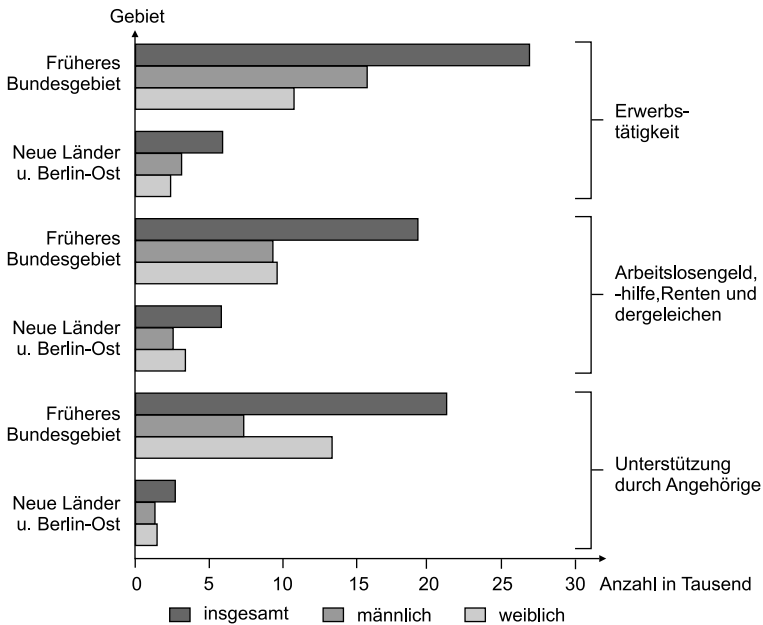


Abb. 5.7:

Bevölkerung nach überwiegendem Lebensunterhalt im Mai 2003 in Tausend. Ergebnisse des Mikrozensus.

Balkendiagramme können außerdem als gestapelte Diagramme angelegt werden und Veränderungen um einen bestimmten Wert anzeigen. Da dieses ebenso bei den Säulendiagrammen möglich ist, wird es im Kapitel 5.2.2 behandelt.

Wichtig bei der Erstellung von Balken- und Säulendiagrammen ist, dass die Größenachse bei Null beginnt (s. Kap. 5.5). Die gewählten Balken und Säulen sollten leicht voneinander zu unterscheiden sein, wobei die gewählte Breite der Elemente immer gleich sein muss. Auch der Abstand zwischen den einzelnen Balken und Säulen sollte gleich bleiben. Sinnvoll ist ein Abstand von einer halben bis zu einer Balkenbreite. Ein größerer Abstand sollte vermieden werden. Die Zahl der Balken und Säulen sollte nicht zu groß sein, empfohlen werden bis höchstens 15 Balken bzw. Säulen. Bei größeren Datenmengen sind Kurvendiagramme geeigneter. Bei der Farbwahl sollte auf eine klare Unterscheidbarkeit der einzelnen Kategorien geachtet werden, ohne dass die Farben aufdringlich wirken. Schraffuren sind oftmals nicht geeignet, weil sie einen Flimmereffekt hervorrufen oder im Fall von Längsstreifen die Säulen schlanker und im Fall von Querstreifen die Säulen breiter wirken lassen. Dreidimensionalität mag ein Balken- oder Säulendiagramm optisch beeindruckender machen, geht aber immer auf Kosten der Ablesbarkeit.

5.2.2 Säulendiagramme

Wie Balkendiagramme werden auch Säulendiagramme oft benutzt, um Vergleiche zwischen Gruppen anzustellen (vgl. Abb. 5.7). Ebenso können zeitliche Verläufe angegeben werden. Sie gehören zu den am häufigsten verwendeten Diagrammen. Einfache Säulendiagramme entsprechen im Wesentlichen den einfachen Balkendiagrammen.

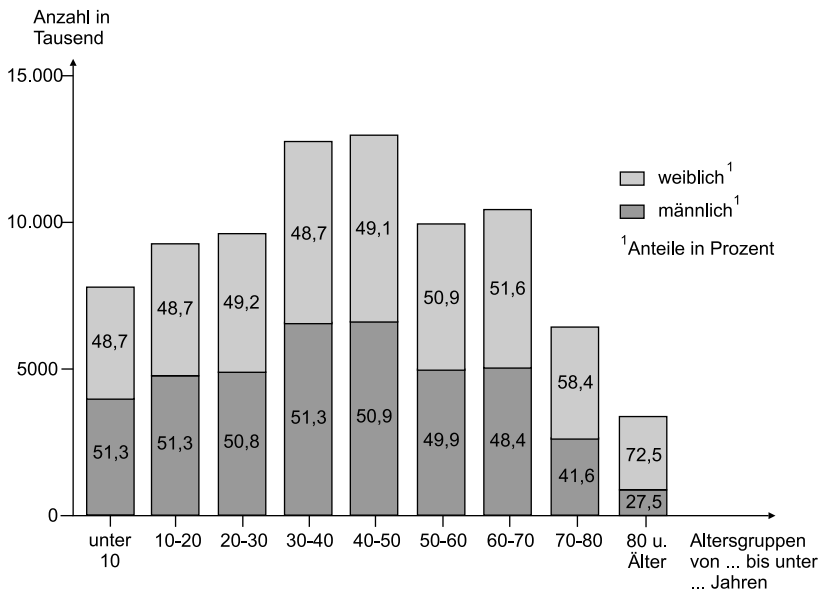


Abb. 5.8: Bevölkerung Deutschlands am 31.12.2003 nach Altersgruppen und Geschlecht.

grammen (s. Kap. 5.2.1) mit dem Unterscheid, dass die Grafik um 90° gedreht ist. Neben solchen einfachen Säulendiagrammen sind auch mehrfache möglich, die ebenfalls den in Kapitel 5.2.1 besprochenen Balkendiagrammen entsprechen. Außerdem kann man Säulen- wie auch Balkendiagramme stapeln. Dabei werden die Säulen aufeinander gestapelt. In Abbildung 5.8 ist so die Aussagekraft der Grafik vergrößert worden, indem neben den absoluten Zahlen pro Altersgruppe auch noch der prozentuale Anteil der Geschlechter an der jeweiligen Altersgruppe angegeben ist. Ein gestapeltes Diagramm kann durchaus in mehr als nur zwei Kategorien unterteilt sein, wie Abbildung 5.9 zeigt. Auch bei prozentualen Angaben ist diese Art der Darstellung beliebt.

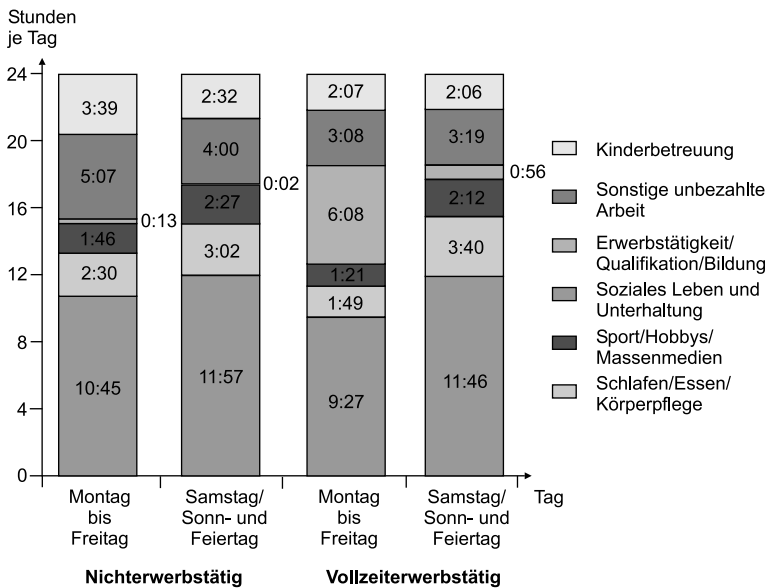


Abb. 5.9: Zeitverwendung von Frauen in Paarhaushalten mit Kindern unter 6 Jahren 2001/2002.

Die Möglichkeiten, die Säulendiagramme bieten, sind aber durchaus noch vielfältiger. So können auch Veränderungen einzelner Werte verdeutlicht werden (Abweichungs-Säulendiagramm). Am bekanntesten ist eine solche Darstellung wohl als Gewinn- und Verlustrechnung bei Wahlen. Dabei werden für die einzelnen Parteien die Gewinne bzw. Verluste gegenüber den vorangegangenen Wahlen dargestellt, indem Gewinne als positive Werte oberhalb der Abszisse aufgetragen werden und Verluste als nach unten gerichtete Balken eingezeichnet werden. Dass eine Präsentation im Fernsehen auch nicht immer zum gewünschten Erfolg führt, zeigt Abbildung 5.10. Hier wird die Veränderung der Umfrageergebnisse der späteren US-Präsidenten nach dem letzten Fernsehduell mit ihrem jeweiligen Konkurrenten

anschaulich gemacht. Offenbar hatten die Herren Carter, Clinton (1992) und Bush (2004) nicht eben ihren besten Tag erwischt, wenngleich es dennoch zum Sieg ge-
reicht hat.

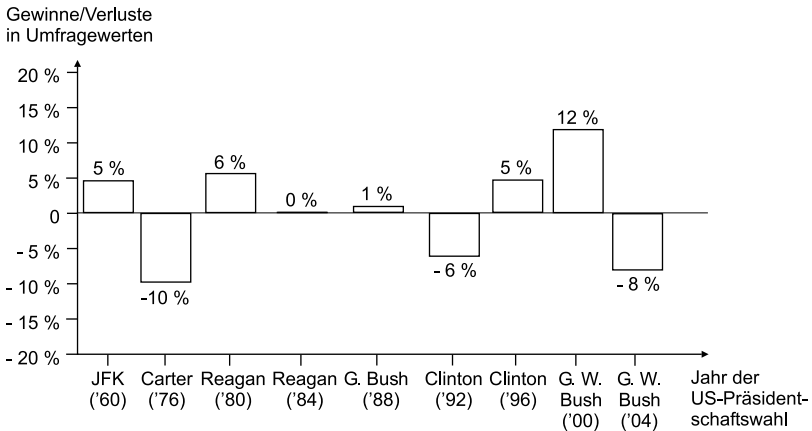


Abb. 5.10:

Veränderungen der Umfrageergebnisse für spätere US-Präsidenten nach dem letzten Fernsehduell.

Wie auch bei den Kurvendiagrammen können die einzelnen Werte mit einem Fehlerbalken für die Standardabweichung vom Mittelwert angegeben werden (Abb. 5.11). Die Höhe des Balkens entspricht hier dem Mittelwert. Beim hier auf-
gezeigten Beispiel ist besonders die Skalierung der y-Achse zu beachten, die im log-
arithmischen Maßstab vorliegt. Solche logarithmischen Skalen bereiten bei der
Interpretation manchmal Schwierigkeiten, weil man nicht beachtet, dass es sich
hierbei um einen extremen Anstieg handelt. Eine Bakterienkultur z. B. zeigt ein ex-
ponentielles Wachstum.

Noch besser geeignet, um die Schwankungen eines Wertes darzustellen, ist das
Spannen-Säulendiagramm oder der **Boxplot**. Hierbei kann man darstellen, um
welche Extremwerte die Daten schwanken. Solche Spannen-Säulendiagramme kö-
nnen z. B. die Schwankungsbreite der Noten bei verschiedenen Klassenarbeiten ver-
deutlichen. Gerne werden Boxplots benutzt, um die Streubreite darzustellen.
In der Mitte der Box (die rechteckige Säule) ist der **Median** angegeben, und von
dort aus nach oben und unten das **Quartil**, also 25 % der Werte, sodass die Streu-
ung von 50 % um den Median durch die Box repräsentiert wird. Minimum und Maxi-
mum der Messwerte werden durch Striche nach oben und unten eingetragen. Somit
wird durch den Abstand der Striche von Minimum und Maximum die Spannweite
der Messwerte bestimmt. Allerdings sind die Striche nach oben und unten, die
auch als Whisker bezeichnet werden, auf den 1,5-fachen Abstand von der Box be-
schränkt. Größere Abweichungen werden als **Ausreißer** betrachtet und gehen nicht
in die Berechnungen ein. Als Ausreißer definiert sind Werte, die 1,5 bis 3 Kasten-

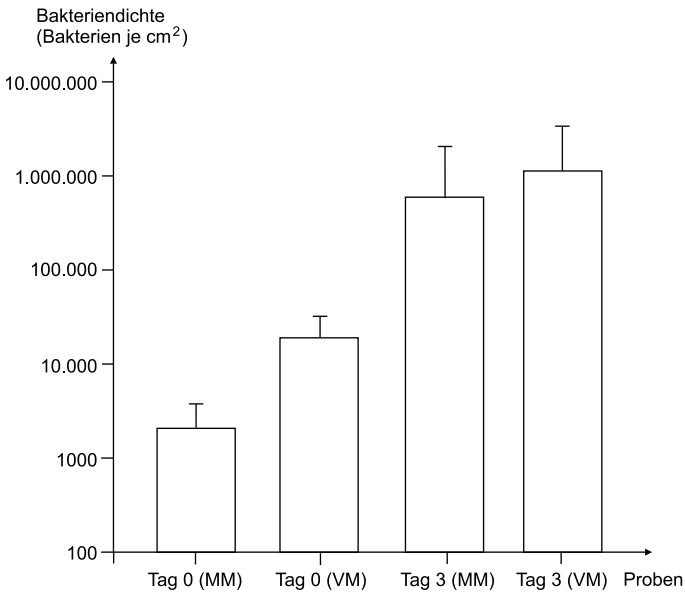


Abb. 5.11:

Wachstum einer Bakterienkultur auf einem Medium, bei dem der Nährboden in nur geringer Konzentration vorhanden ist (MM: Mangelmedium; VM: Vollmedium).

längen vom oberen oder unteren Kastenrand entfernt liegen. Werte, die mehr als 3 Kastenlängen entfernt sind, werden als **Extremwerte** bezeichnet. Die Boxplot-Darstellung gibt Auskunft über die symmetrische Verteilung der Messwerte um den Median. Nur bei einer symmetrischen Verteilung liegt der Median auch in der Mitte der Box.

In Abbildung 5.12 ist ein Beispiel für diese Darstellungsart gegeben. Hier ist klar zu erkennen, dass die Verteilung asymmetrisch ist. Die Kreise ober- und unterhalb der Striche sind dabei die Extremwerte, die aus der Berechnung herausfallen.

Balken- und Säulendiagramme

- geben zeitliche Verläufe wieder.
- können Abweichungen von einem bestimmten Wert darstellen (Abweichungs-Balken- und Säulendiagramm).
- können verschiedene Kategorien miteinander vergleichen (mehrfache Balken- und Säulendiagramme).
- stellen Relationen zu bestimmten Zeitpunkten dar.
- stellen weniger eine Tendenz dar, betonen dafür mehr den einzelnen Wert.
- sollten nicht zu viele Kategorien und Unterteilungen enthalten, um nicht unübersichtlich zu werden.

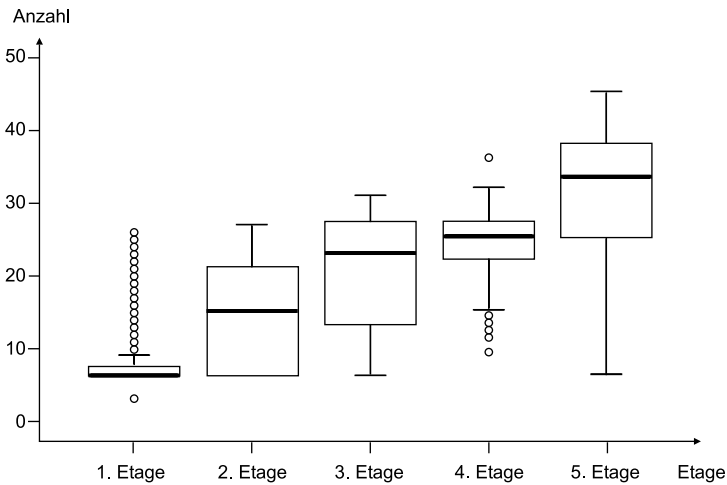


Abb. 5.12:
Nutzung des Aufzugs durch Hotelgäste je nach bewohnter Etage pro Aufenthalt.

5.3 Histogramme

Histogramme, die in diesem Buch häufiger vorkommen, zeigen die Häufigkeitsverteilung von Messwerten. Im Histogramm sind auf der x -Achse die Daten größenmäßig angeordnet und auf der y -Achse ist deren Häufigkeit angegeben. Die benachbarten Balken berühren einander. Die Größen auf der x -Achse sind dabei in Klassen angeordnet. Solche Klassen müssen nicht zwangsläufig gleich breit sein. Jeder Klasse wird ein Rechteck zugeordnet, das seiner Häufigkeit entspricht. Dabei ist die Fläche des Rechtecks proportional zur Häufigkeit der Klasse, es sind also flächentreue Darstellungen. Die Höhe der Elemente wird auch als Häufigkeitsdichte benannt. Sie errechnet sich folgendermaßen:

$$\text{Häufigkeitsdichte} = \text{Häufigkeit} / \text{Klassenbreite}$$

Abbildung 5.13 gibt ein Beispiel eines solchen Histogramms. Da in diesem Fall die einzelnen Säulen die gleiche Breite haben, kann über die Höhe der Säule direkt auf die Häufigkeit geschlossen werden. Anders dagegen bei Abbildung 5.14. Hier ist die Häufigkeitsverteilung proportional zur Fläche und kann nicht direkt über die Höhe der Säulen abgelesen werden.