Bohner
Kessler
Ott

Stochastik im Berufskolleg – Berufliches Gymnasium

Merkur
Verlag Rinteln

Wirtschaftswissenschaftliche Bücherei für Schule und Praxis
Begründet von Handelsschul-Direktor Dipl.-Hdl. Friedrich Hutkap †

Die Verfasser:

Kurt Bohner
Oberstudienrat

Dipl.-Math. Roland Kessler
Studienrat

Roland Ott
Oberstudienrat

* * * * *

1. Auflage 2011

Gesamtherstellung:
MERKUR VERLAG RINTELN Hutkap GmbH & Co. KG, 31735 Rinteln

E-Mail: info@merkur-verlag.de
lehrer-service@merkur-verlag.de
Internet: www.merkur-verlag.de

ISBN 978-3-8120-**0073-4**

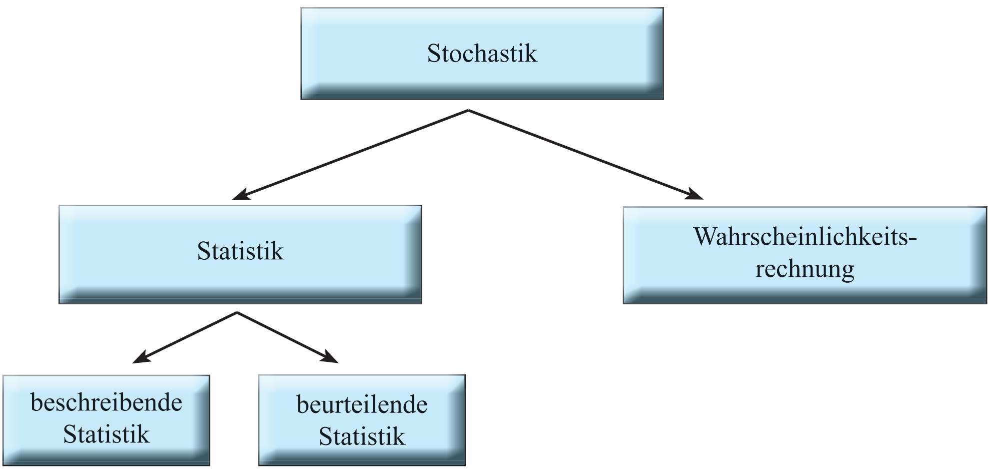
I. Beschreibende Statistik

„Mit Statistik kann man alles beweisen.
Ich glaube keiner Statistik, die ich nicht selbst gefälscht habe“. (Winston Churchill)

1 Einführung

„Die Statistik hat eine erhebliche Bedeutung für eine staatliche Politik, die den Prinzipien und Richtlinien des Grundgesetzes verpflichtet ist ...” (Volkszählungsurteil des BVerfG)

Schon ca. 3000 v. Chr. wurden im alten Ägypten statistische Erhebungen durchgeführt. Erfasst wurden z.B. die Anzahl der Felder, die Vorräte und Arbeitskräfte. Auch von den Römern sind statistische Erhebungen schon aus der Zeit um 550 v. Chr. bekannt. Ende des 18. Jahrhunderts besaßen alle wichtigen Gemeinwesen statistische Ämter. Entscheidend für den weiteren Ausbau der Wahrscheinlichkeitstheorie waren der Schweizer Leonhard Euler, der Deutsche Carl Friedrich Gauß und die Franzosen Poisson und Laplace. Die Zusammenfassung von Wahrscheinlichkeitsrechnung und mathematischer Statistik, bezeichnet man mit dem modernen Begriff Stochastik.



In der **beschreibenden Statistik** werden **Daten erhoben, aufbereitet und analysiert.**
Die erhobenen Daten werden geordnet und übersichtlich dargestellt.
Dadurch bekommt man einen ersten Überblick, erkennt Zusammenhänge und Strukturen.
Die Struktur einer Verteilung wird durch Lagemaße (z. B. Mittelwert) und Streumaße (z. B. Standardabweichung) beschrieben.
Zwei Verteilungen und deren Zusammenhänge untersucht man in einer Regressionsanalyse, die die Grundlage für Trends und Prognosen darstellt.

2 Häufigkeitsverteilung

Absolute und relative Häufigkeit

Eine Häufigkeitsverteilung dient zur statistischen Beschreibung von Daten.
An einer Kreuzung werden innerhalb einer halben Stunde 125 Fahrzeuge gezählt.
Davon sind 18 Fahrzeuge Lkws. Die **absolute Häufigkeit** der Lkws ist somit 18.
Dies sagt wenig darüber aus, wie groß der Anteil der Lkws am Verkehr auf dieser Kreuzung ist. Um ein brauchbares Maß für diesen Anteil zu bekommen, benötigt man die relative Häufigkeit. Die **relative Häufigkeit** ist der Quotient $\frac{18}{125} = 0{,}144 = 14{,}4\ \%$,
d. h., ca. 14 % der vorbeigefahrenen Fahrzeuge waren Lkws.

Beispiel

Ein Schüler erkundigt sich bei einer Zulassungsstelle nach der Anzahl der zugelassenen Autos, sortiert nach Automarken. Er erstellt eine **Häufigkeitstabelle.**

Marke	Ford	VW	Mercedes	andere	Summe n
abs. Häufigkeit n_i	2810	3211	1398	2081	9500
rel. Häufigkeit $h = \frac{n_i}{n}$	$\frac{2810}{9500} \approx 0{,}29$	$\frac{3211}{9500} \approx 0{,}34$	$\frac{1398}{9500} \approx 0{,}15$	$\frac{2081}{9500} \approx 0{,}22$	$\frac{9500}{9500} = 1$
rel. Häufigkeit in %	29 %	34 %	15 %	22 %	100 %

Festlegung: Unter der **absoluten Häufigkeit H** einer Merkmalsausprägung versteht man die Anzahl der Fälle, in denen die Ausprägung eintritt.
Ist n die Anzahl der Durchführungen (Stichprobenumfang),
so ist $\mathbf{h = \frac{H}{n}}$ die relative Häufigkeit der Merkmalsausprägung.

$$\textbf{Relative Häufigkeit} = \frac{\text{absolute Häufigkeit}}{\text{Stichprobenumfang}}$$

Eigenschaften der relativen Häufigkeit:

- Für die **relative Häufigkeit** gilt: $0 \le h \le 1$
  Die relative Häufigkeit h liegt zwischen 0 und 1.
- Die **Summe** der relativen Häufigkeiten ist 1 bzw. 100 %.

Beispiele

1) Im Laden liegt eine Lieferung von 12 Schachteln mit je 10 Eiern. Eine Aushilfskraft notiert sich für jede Schachtel die Anzahl der weißen Eier. Sie erhält folgende Liste: 1; 3; 0; 2; 5; 1; 4; 0; 4; 1; 2; 4.
   a) Berechnen Sie den Mittelwert für die Anzahl der weißen Eier pro Schachtel.
   b) Wie groß ist für diese Lieferung die relative Häufigkeit der weißen Eier?

Lösung

a) **Durchschnittliche Anzahl** von weißen Eiern pro Schachtel (Mittelwert):

$$\overline{x} = \frac{1+3+0+2+5+1+4+0+4+1+2+4}{12} = \frac{27}{12} = 2{,}25$$

Eine Schachtel enthält durchschnittlich 2,25 weiße Eier.

b) **Relative Häufigkeit** der weißen Eier:

$$h(\text{weiße Eier}) = \frac{\text{Anzahl der weißen Eier}}{\text{Gesamtzahl der Eier}} = \frac{27}{120} = 0{,}225$$

2) In einem voll besetzten Zug mit 6-Personen-Abteilen hat man in jedem Raucherabteil die Zahl der Nichtraucher ermittelt. Das Ergebnis gibt folgende Tabelle wieder:

n	0	1	2	3	4	5	6
a(n)	20	14	9	3	2	1	1

Dabei ist a(n) die Anzahl der Abteile, in denen man n Nichtraucher angetroffen hat.
   a) Bestimmen Sie für alle n die relative Häufigkeit h(n), mit der man n Nichtraucher in einem Raucherabteil angetroffen hat.
   b) Wie viel Prozent aller Fahrgäste waren Nichtraucher?
   c) In wie viel % der Raucherabteile befinden sich höchstens 2 Nichtraucher?

Lösung

a) Die Summe aller Abteile beträgt 50.

In 20 von 50 Abteilen hat man keinen Nichtraucher angetroffen, also $h(0) = \frac{20}{50} = 0{,}4$.

Allgemein gilt für die relative Häufigkeit für n [{1; 2; 3; 4; 5; 6}: $h(n) = \frac{a(n)}{50}$.

n	0	1	2	3	4	5	6
a(n)	20	14	9	3	2	1	1
h(n)	0,40	0,28	0,18	0,06	0,04	0,02	0,02

b) Gesamtzahl der Fahrgäste: $G = 50 \cdot 6 = 300$

Gesamtzahl der Nichtraucher: $N = 14 \cdot 1 + 9 \cdot 2 + 3 \cdot 3 + 2 \cdot 4 + 1 \cdot 5 + 1 \cdot 6 = 60$

Anteil der Nichtraucher: $\frac{N}{G} = \frac{60}{300} = 0{,}2 = 20\,\%$

20 % aller Fahrgäste waren Nichtraucher.

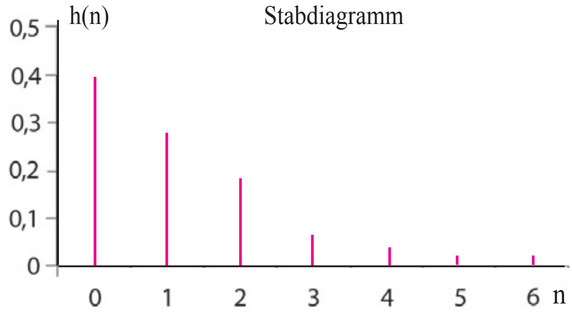
c) In 20 + 14 + 9 = 43 Raucherabteilen befinden sich höchstens 2 Nichtraucher.

Relative Häufigkeit: $h = \frac{43}{50} = 86\,\%$. In 86 % befinden sich höchstens 2 Nichtraucher.

Grafische Darstellungen einer Häufigkeitsverteilung

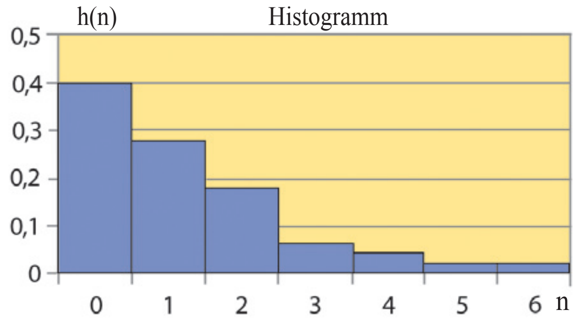
Die grafische Darstellung wird anhand der Häufigkeitsverteilung von Beispiel 2 (s. Seite 11) erläutert.

a) **Stabdiagramm**



Auf der y-Achse werden die relativen Häufigkeiten abgetragen;
auf der x-Achse die Anzahl der Nichtraucher.

b) **Histogramm**



Ein Histogramm besteht aus Rechtecken. Wählt man Rechtecke mit der Grundseite der Länge 1, entspricht der Inhalt eines Rechtecks der zugehörigen relativen Häufigkeit.
An der y-Achse wird die relative Häufigkeit abgetragen.
Bei einem Histogramm beträgt die **Summe der Inhalte der Rechtecksflächen 1.**

Beachten Sie: Ein Histogramm ist eine grafische Darstellung einer Häufigkeitsverteilung.
Es besteht aus mehreren, direkt aneinander angrenzenden Säulen.
Die Summe der Inhalte der Rechtecksflächen beträgt 1.

Bemerkung: Weitere grafische Darstellungen sind Punktdiagramm, Liniendiagramm, Balkendiagramm und Kreisdiagramm.

Aufgaben

1. In einer Gartenanlage sind alle Gärten einheitlich mit 6 Obstbäumen bepflanzt. Die Bäume wurden auf Krankheiten untersucht. Die Untersuchung erbrachte folgendes Ergebnis:

n	0	1	2	3	4	5	6
a(n)	1	3	6	7	5	2	1

Dabei ist a(n) die Anzahl der Gärten, in denen man n geschädigte Bäume gezählt hat.
a) Berechnen Sie für $n \in \{0; 1; 2; 3; 4; 5; 6\}$ die relative Häufigkeit h(n), mit der man n geschädigte Bäume gefunden hat.
b) Wie viel Prozent der Obstbäume in der gesamten Anlage sind geschädigt?

2. Bei einer Stichprobe werden in einem Wörterbuch 500 Wörter mit jeweils 6 Buchstaben auf die Anzahl der enthaltenen Vokale (A, E, I, O, U) untersucht.
Nachfolgende Tabelle zeigt das Ergebnis:

Vokale pro Wort	0	1	2	3	4	5	6
Häufigkeit	0	75	175	190	50	10	0

a) Berechnen Sie die relativen Häufigkeiten für die Anzahl der Vokale je Wort.
b) Wie viele Vokale enthält ein Wort durchschnittlich?
c) Wie viele Wörter mit 3 Vokalen müssten dazukommen, damit der Durchschnitt bei genau 2,5 Vokalen je Wort liegt?

3. Bei einer Mathematikklassenarbeit gab es folgende Noten:
3; 4; 3; 2; 3; 1; 5; 5; 4; 3; 3; 2; 1; 4; 2; 5; 4; 2; 4; 3
Erstellen Sie eine Häufigkeitstabelle. Stellen Sie die Verteilung grafisch dar.

4. Bei einer Aufnahmeprüfung sind von jedem Bewerber 5 Aufgaben zu bearbeiten. Das Ergebnis der Prüfung zeigt die folgende Tabelle, wobei n(k) die Anzahl der Bewerber angibt, die k Aufgaben richtig bearbeitet haben:

k	5	4	3	2	1	0
n(k)	4	7	14	11	8	6

a) Ermitteln Sie für $k \in \{0; 1; 2; 3; 4; 5\}$ die relative Häufigkeit dafür, dass ein Bewerber k Aufgaben richtig gelöst hat. Stellen Sie die Häufigkeitsverteilung grafisch dar.
b) Wie viele Aufgaben hat jeder Bewerber im Mittel richtig bearbeitet?
c) Wie viel Prozent der bearbeiteten Aufgaben wurden richtig gelöst?

5. Eine Fabrik produziert Stifte. Die Stifte werden auf Abweichungen im Durchmesser und in der Länge geprüft. Ein Stift ist fehlerhaft, wenn er im Durchmesser oder in der Länge abweicht. Von 2000 Stiften gab es 65 Abweichungen im Durchmesser, 87 Abweichungen in der Länge und 25 Abweichungen im Durchmesser und in der Länge.
Bestimmen Sie die relative Häufigkeit der fehlerhaften Stifte.

3 Lagemaße

In diesem Kapitel sollen zu einer Beobachtungsreihe charakteristische Größen bestimmt werden, die Aussagen über die Lage der Beobachtungswerte zulassen. Das bekannteste Lagemaß ist das arithmetische Mittel.

3.1 Arithmetisches Mittel

Beispiele

1) Die folgende Liste enthält das Körpergewicht in kg von 7 Schülern:
52; 67; 60; 55; 63; 63; 70
Bestimmen Sie das durchschnittliche Körpergewicht dieser Schüler.

Lösung

Anzahl der Schüler (Merkmalsträger): $n = 7$

Summe aller Körpergewichte: $52 + 67 + ... + 70 = 430$

Durchschnittswert $\overline{x}$ (Mittelwert) $\overline{x} = \frac{430}{7} = 61{,}4$

Das durchschnittliche Körpergewicht beträgt 61,4 kg.

Bemerkung: Der Mittelwert 61,4 kg bedeutet: Hätte jeder Schüler ein Körpergewicht von 61,4 kg, so ergäbe die Summe aller Körpergewichte 430 kg (gerundet).

Berechnung des (arithmetischen) Mittelwertes $\overline{x}$ aus den Beobachtungswerten x_i

$$\text{Mittelwert } \overline{x} = \frac{\text{Summe aller Beobachtungswerte } x_i}{\text{Anzahl } n \text{ der Beobachtungswerte } x_i}$$

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + ... x_n) = \sum_{i=1}^{n} x_i$$

$\sum_{i=1}^{n} x_i$ gelesen: Summe aller x_i mit i gleich 1 bis n

Beispiele für Mittelwerte

Pro-Kopf-Verbrauch von Wasser: 135 Liter pro Tag

Durchschnittseinkommen aller Arbeitnehmer: 28500 € pro Jahr

Durchschnittlicher Zigarettenverbrauch: 24,2 Zigaretten pro Tag

2) Eine Umfrage nach der Anzahl x_i der Geschwister hat ergeben:

x_i	0	1	2	3	4
relative Häufikeit h_i	0,48	0,3	0,14	0,07	0,01

Berechnen Sie das arithmetische Mittel.

Lösung

Mittelwert: $\overline{x} = 0 \cdot 0{,}48 + 1 \cdot 0{,}3 + 2 \cdot 0{,}14 + 3 \cdot 0{,}07 + 4 \cdot 0{,}01 = 0{,}83$

3) Ein Weingut bietet vier Sorten Weine aus verschiedenen Lagen an.
Die nachfolgende Liste gibt die verkauften Mengen für einen Jahrgang an.

Sorte	A	B	C	D
Verkaufspreis pro Flasche in € (x_i)	5	7	8	12
Verkaufte Flaschen (n_i)	150	600	250	300

Berechnen Sie den durchschnittlichen Verkaufspreis pro Flasche.

Lösung

Anzahl der verkauften Flaschen: $n = 1300$

Gesamteinnahmen (Erlös): $5 \cdot 150 + 7 \cdot 600 + 8 \cdot 250 + 12 \cdot 300 = 10550$

Erlös pro Flasche: $\frac{10550}{1300} = 8{,}12$

Der durchschnittliche Verkaufspreis pro Flasche beträgt 8,12 €.

Berechnung des arithmetischen Mittels $\overline{x}$ aus einer Häufigkeitstabelle

für **vier Ausprägungen**: $\overline{x} = \frac{1}{n}(x_1 n_1 + x_2 n_2 + x_3 n_3 + x_4 n_4) = \frac{1}{n}\sum_{i=1}^{4} x_i n_i$

für **k Ausprägungen**: $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i h_i \quad \text{mit } h_i = \frac{n_i}{n}$

Hierbei müssen die Häufigkeiten n_i aller k Merkmalsausprägungen x_i bekannt sein.
n ist die Summe der Häufigkeiten n_i.

Aufgaben

1. Berechnen Sie das arithmetische Mittel folgender Daten:
   a) 15,2 16,1 17,3 15,7 14,8 17,0 16,8 15,1

   b)

x_i	52	55	58	60	65
n_i	3	4	5	3	1

   c)

x_i	0	1	2	3
h_i	0,25	0,2	0,15	0,4

2. Die 32 Schüler in der Klasse haben ein Durchschnittsgewicht von 74 kg.
   Nach langer Krankheit hat ein Schüler 24 kg abgenommen.
   Um wie viel ändert sich der Mittelwert?
   Wie ändert sich der Mittelwert, wenn sich bei einer Datenreihe der Länge n ein Datenwert um a vergrößert (verkleinert)?

3.2 Modus und Median

Beim Modus wird nach nach der Merkmalsausprägung mit der größten Häufigkeit gesucht.

Der Modus x_{Mod} (Modalwert) ist der Merkmalswert, der am häufigsten vorkommt.

Beispiele

a) Merkmal Sportart
Fußball; Handball; Volleyball; Fußball; Fußball; Handball; Fußball
Modus x_{Mod} = Fußball. Dieser Wert kommt am häufigsten (4-mal) vor.

b) Merkmal Gewicht in kg
Liste: 44; 57; 32; 44; 32; 63; 66; 63; 99; 63
Der Modus x_{Mod} ist 63 (kg) (3-mal).

Bemerkungen:
1. Gibt es mehrere Merkmalswerte mit der gleichen maximalen Häufigkeit, existiert **kein** Modus.
2. Der Modus kann aus einem Säulendiagramm (Histogramm) direkt abgelesen werden.

Beispiel

Ein Bautrupp mit 9 Personen hat folgende monatliche Einkünfte (in €):

1160	1050	980	1200	970	1800	6600	1180	1090

Hierbei ist das arithmetische Mittel $\overline{x} = 1781$.
Dieser Durchschnitt liefert ein falsches Bild, weil die Mehrzahl
(sieben von neun Personen) höchstens 1200 € verdient.

Der Wert 6600 (Ausreißer) zieht den Mittelwert $\overline{x}$ nach oben.
Man sucht nach einem Wert, der die Verteilung der Einkünfte besser charakterisiert.

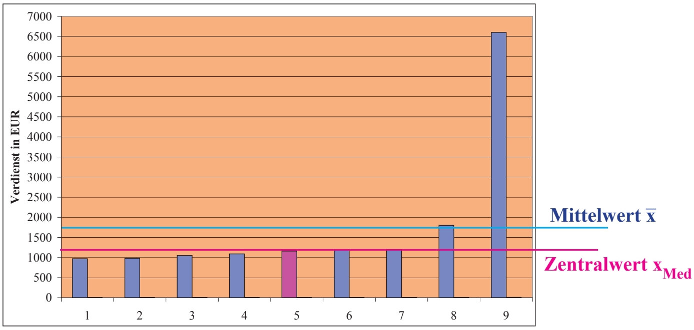
Dazu ordnen wir die Verdienste der Größe nach.

970	980	1050	1090	**1160**	1180	1200	1800	6600

Der in der Mitte liegende Wert 1160 € beschreibt die Verteilung besser als der Mittelwert $\overline{x}$.
Der Wert 1160 ist der Median (Zentralwert).

Der **Zentralwert x_{Med} (Median)** ist derjenige Wert, der in der Mitte steht, wenn alle Beobachtungswerte x_i der Größe nach geordnet sind.

Vergleich von Mittelwert $\overline{x}$ und Zentralwert x_{Med} anhand eines Diagramms.



Bemerkung: Zentralwert x_{Med} für eine **ungerade Anzahl n** von Beobachtungswerten:

$$x_{Med} = x_{\frac{n+1}{2}}$$

Zentralwert x_{Med} für eine **gerade Anzahl n** von Beobachtungswerten:

$$x_{Med} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$$

Was man wissen sollte ... über Lagemaße

Arithmetisches Mittel: Summe der Merkmalsausprägungen dividiert durch deren Anzahl: $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Modus (Modalwert): Merkmalswert, der am häufigsten vorkommt

Zentralwert (Median): Merkmalsausprägung desjenigen Wertes, der eine der Größe nach sortierte Reihe halbiert.

Aufgaben

1. In einem Unternehmen sind 10 Frauen in einer Putzkolonne auf 400-€-Basis beschäftigt. Der Chef stellt einen Vorarbeiter ein, der 2800 € pro Monat verdienen soll. Welche Auswirkungen ergeben sich dadurch auf den Modus, den Median und das arithmetische Mittel der Monatseinkommen aller Mitarbeiter?

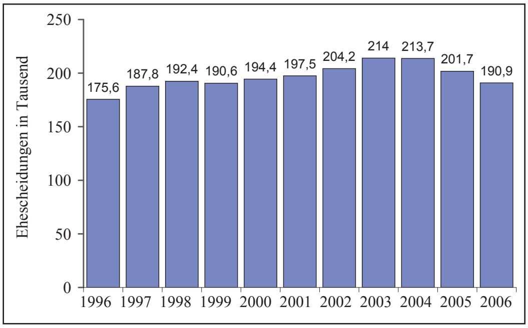
2. Die 13 Studenten in einem Kurs geben ihre monatlichen Ausgaben in €
   wie folgt an: 1300, 1200, 1400, 700, 200, 750, 1450, 1500, 800, 800, 950, 900, 3000.
   a) Berechnen Sie das arithmetische Mittel, den Median und den Modalwert.
      Interpretieren Sie diese Maße inhaltlich.
   b) Erklären Sie, warum sich die Lagemaße unterscheiden.
   c) Welche Maßzahl charakterisiert Ihrer Meinung nach die Stichprobe am besten?

3. Nach Angaben des Statistischen Bundesamtes kamen im Jahre 2007 auf je 100 000 Einwohner im Alter zwischen 18 und 25 Jahren 14,3 Verkehrstote.
   Inwiefern handelt es sich bei dieser Angabe um einen Mittelwert?

4. Die Grafik des statistischen Bundesamtes zeigt die Zahl der Ehescheidungen in Deutschland.

   Bestimmen Sie das arithmetische Mittel und den Zentralwert.

   Warum lässt sich kein Modalwert angeben?



5. Die Wetterstation liefert die Tagestemperaturen (in °C), gemessen um 12:00 Uhr, für die 30 Tage eines Monats:
   11,8 12,4 18,5 24,2 23,5 20,8 21,5 23,5 20,6 15,4 14,8 17,5 16,9 18,2 16,4
   17,9 20,3 19,5 17,9 18,5 24,0 23,5 25,2 23,6 22,2 20,7 21,0 20,4 18,9 21,8
   a) Berechnen Sie die durchschnittliche Tagestemperatur.
   b) Berechnen Sie den Median.
   c) Im langjährigen Mittel lagen die Durchschnittstemperaturen für diesen Monat bei 18,5 °C. Haben sich die klimatischen Verhältnisse geändert?

6. Die Häufigkeitstabelle zeigt die Anzahl der Kunden an der Kasse im Supermarkt in 30 aufeinanderfolgenden Zeitabschnitten von je 10 Minuten.

Anzahl der Kunden:	0	2	3	4	5	6	7	9
Häufigkeit:	1	3	4	5	8	3	2	4

   a) Stellen Sie die Verteilung in einem Säulendiagramm dar.
   b) Berechnen Sie den Zentralwert und den Mittelwert. Was fällt Ihnen auf?