Anschrift des Verfassers:
Dr. med. Volker Harms, In‘t Holt 37, 24214 Lindhöft
Zeichnungen:
Liane Pielke-Harms, In‘t Holt 37, 24214 Lindhöft
Lektorat: Bente Blasius, Plön

Titelbild: iStock by Getty Images
Die erwartungsvoll und fröhlich blickenden Menschen
könnten die zukünftigen Patienten darstellen, sie stellen aber auf jeden Fall
den Gegenstand der Medizinischen Statistik dar: die Vielfalt in Bezug auf
Alter, Geschlecht, Hautfarbe, Beruf und Lebensläufe.

1. Auflage: November 1976
2., völlig neu bearbeitete Auflage: Oktober 1977
3., überarbeitete Auflage: Oktober 1979
4., völlig neu bearbeitete Auflage: Oktober 1982
5., völlig neu bearbeitete Auflage: April 1988
6., völlig neu bearbeitete Auflage: November 1992
7., überarbeitete Auflage: Juni 1998
8., völlig neu bearbeitete Auflage: Juni 2012
9., völlig neu bearbeitete Auflage: Dezember 2019

ISBN 978-3-86026-241-2

Medizinische

Statistik

Originalfragen des IMPP

mit ausführlichen Kommentaren

Inhaltsverzeichnis

Aus den Vorworten des Lehrbuchs

Vorwort zur 8. Auflage

Das Titelbild zeigt eine einzelne Person umgeben vor einer sich in Bewegung befindlichen Menschenmenge. Dies mag für ein Statikbuch zunächst ungewöhnlich erscheinen, spiegelt aber das Selbstverständnis der medizinischen Statistik in der Ära der Evidence Based Medicine wider: Der einzelne Patient ist Dreh- und Angelpunkt aller medizinischen Bemühungen. Die gesunden und kranken Menschen um ihn herum und vor allem – zeitlich gesehen – vor ihm, haben den Fundus an medizinischem Wissen gelegt, aufgrund dessen der Einzelne behandelt wird, wenn er krank ist.

Dieses Bild entspricht aber auch der epidemiologischen Situation, in der sich der Einzelne befindet. Fast alle Erkrankungen ergeben sich aus der Interaktion mit den Mitmenschen. Bei Infektionskrankheiten ist dies offensichtlich, aber es gilt auch für verhaltensbedingte Erkrankungen, insbesondere für die sog. Zivilisationskrankheiten.

Bei der Neuauflage dieses Buches hat das Thema Epidemiologie eine große Rolle gespielt. Die Epidemiologie ist ebenso wie die medizinische Statistik ein Querschnittsfach, welches für fast alle Bereiche der Medizin von Bedeutung ist.

In den letzten Jahren hat die Evidence Based Medicine Einzug in den klinischen Alltag gehalten und damit den Transfer neuer Erkenntnisse von der Forschung zur Behandlung wesentlich beschleunigt. Die Regeln der ärztlichen Kunst werden heute durch Leitlinien definiert, die kontinuierlich dem medizinischen Fortschritt angepasst werden.

Die Medizin ist unzweifelhaft eine Naturwissenschaft: Die Pathophysiologie beruht auf den Prinzipien der Physiologie und Biochemie und auch die therapeutischen Eingriffe entfalten ihre Wirkungen auf naturwissenschaftlicher Grundlage.

Und dennoch gibt es im Bereich von Diagnostik, Therapie und Prognose niemals eine absolute Gewissheit. Die Variabilität des Patienten führt immer wieder zu Überraschungen. Dies entwertet keinesfalls die Regeln der Schulmedizin, aber es zwingt den behandelnden Arzt zur ständigen Wachsamkeit.

Auf dem Weg von der 7. zur 8. Auflage ist dieses Buch wesentlich umfangreicher geworden. Die Lesbarkeit hat darunter jedoch nicht gelitten, weil jedes Kapitel eine abgeschlossene Einheit bildet und in der Regel auch ohne Kenntnis der vorangegangenen Kapitel verständlich ist.

Ich hoffe, dass der Leser bei der Lektüre zur Erkenntnis kommt, dass medizinische Statistik heute weit mehr ist als die Auswertung medizinischer Datenreihen. Kapitel mit speziellem Bezug zur Medizin sind zum Beispiel *Entscheidungsfindung in der Medizin*, *Das Risiko*, *Fehler und ihre Vermeidung*, *Kausalität*, *Versuchsplanung*, *Der klinische Versuch*, *Epidemiologische Studien*, *Demographischer Wandel*, *Grundzüge der Epidemiologie*, *Systematic Reviews und Metaanalysen*, *Evidenzbasierte Medizin und Leitlinien* sowie *Literatursuche*. Schon diese Aufzählung zeigt, dass Statistik zu einem tieferen Verständnis medizinischer Zusammenhänge verhilft.

Zum Schluss möchte ich mich bei den engagierten Mitstreitern bedanken, ohne deren Hilfe dieses Buch nicht hätte entstehen können. Meine Frau hat alle Abbildungen neu angefertigt, Bente Blasius hat sich um Satz und Layout gekümmert, Ulf Schiefer um die Orthografie, Prof. Detlev Kraack um Stil und Verständlichkeit, und mein Sohn stand immer dann mit Rat und Tat zur Seite, wenn es Probleme gab, bei Abbildungen, Formeln, Tabellen usw.

Für inhaltliche Anregungen bin ich Prof. Gerd Antes, dem Direktor des Deutschen Cochrane Zentrum in Freiburg, für viele Denkanstöße und zum Teil auch Formulierungen dankbar, aber auch seinen Mitarbeiterinnen, Frau Dr. Christine Schmucker und Frau Edith Motschall.

Für Fehler hingegen – sollten doch noch welche vorhanden sein – bin ich selbst verantwortlich.

Vorwort zur 9. Auflage

Die Bearbeitung der 9. Auflage steht unter dem Motto, den Text zu verschlanken und auf das zum Verständnis Notwendige zu reduzieren. Die Prüfungsfragen wurden in ein separates Buch ausgegliedert.

Auch in den früheren Auflagen habe ich mich darum bemüht, den Stoff so darzustellen, dass er auch für den mathematisch weniger bewanderten Leser verständlich ist. Das ist manchmal leichter gesagt als getan. Wenn man sich seit Jahrzehnten mit der Materie beschäftigt, ist einem häufig gar nicht bewusst, wo die Stolpersteine des Verständnisses liegen. Meine jüngste Tochter Jana Julia hat gerade Abitur gemacht und war auf der Suche nach einem Ferienjob. Mit ihrer jugendlichen Unbefangenheit, die sich in der Haltung niederschlägt, dass es wohl am Text liegen müsse, wenn sie etwas nicht versteht, hat sie das Buch gelesen und überall markiert, wo es ihr schwer- oder sogar unverständlich erschien. Mir hat das anschließend viel Kopfzerbrechen bereitet, aber ich denke, die Mühen des Umformulierens haben sich gelohnt.

Eine unerwartete Nebenwirkung dieses Ferienjobs bestand darin, dass Jana Julia sich nach jahrelangen Überlegungen spontan entschlossen hat, Medizin zu studieren, was sie jetzt mit großem Eifer und Interesse macht.

Weiterhin hoffe ich auf zahlreiche Rückmeldungen, Anregungen, kritische Hinweise und Verbesserungsvorschläge für die nächste Auflage.

Algarobo, Spanien, 3.11.2019 Volker Harms

Übungsaufgaben

2.1 Begriffsbestimmung

1. Im Rahmen einer Untersuchung an adipösen Patienten über die Wirksamkeit einer neuen Diät wird täglich das Körpergewicht in kg gemessen. Welche der folgenden Aussagen trifft dann zu?

(A) Das Körpergewicht ist Merkmalsträger.
(B) Die Beobachtungseinheit ist kg.
(C) Die Diät ist eine Beobachtungseinheit.
(D) Merkmalsträger sind die adipösen Patienten.
(E) Das Körpergewicht ist eine Beobachtungseinheit.

2. Bei der Neuaufnahme ins Krankenhaus wird das Körpergewicht eines Patienten festgestellt. Welche der folgenden Aussagen ist/sind richtig?

(1) Der Patient ist der Merkmalsträger.
(2) Die wägende Krankenschwester ist die Beobachtungseinheit.
(3) Das Körpergewicht ist das (hier) interessierende Merkmal.

(A) nur 1 ist richtig
(B) nur 3 ist richtig
(C) nur 1 und 3 sind richtig
(D) nur 2 und 3 sind richtig
(E) 1–3 = alle sind richtig

3. Es wird ein Versuch geplant, bei dem die Wirkung eines Medikaments auf die Herzfrequenz bei der Bradykardie untersucht werden soll. Ordnen Sie den Begriffen die entsprechenden Oberbegriffe zu.

Begriff	**Oberbegriff**
(1) Patient Meier	(a) Beobachtungseinheit
(2) 45/min	(b) Grundgesamtheit
(3) Herzfrequenz	(c) Merkmal
(4) Menge aller bradykarden Patienten	(d) Ausprägung
(5) Bradykardie	

(A) 1–a, 2–c, 3–d, 4–b
(B) 1–a, 2–d, 3–c, 4–b
(C) 1–a, 2–d, 3–c, 5–b
(D) 1–c, 2–d, 3–a, 4–b
(E) 1–d, 2–c, 3–a, 5–b

4. Wir unterscheiden qualitative und quantitative Merkmale. Welche der folgenden, bei einer Patientengruppe erhobenen Merkmale sind qualitativ?

(1) Vitalkapazität
(2) Blutgruppe
(3) Serumglucose
(4) Kammerflimmern

(A) nur 1 und 2 sind richtig
(B) nur 1, 2 und 3 sind richtig
(C) nur 2, 3 und 4 sind richtig
(D) nur 2 und 4 sind richtig
(E) nur 1, 2 und 4 sind richtig

Geben Sie bitte zu jeder Merkmalsart in Liste 1 das zutreffende Merkmal aus Liste 2 an:

Liste 1
5. Quantitativ diskretes Merkmal
6. Qualitatives Merkmal

Liste 2
(A) Körpergewicht
(B) Nüchternblutzucker
(C) Blutgruppe
(D) spezifisches Gewicht des Urins
(E) Anzahl der Geschwister eines Patienten

7. Prüfen Sie, welche der folgenden Merkmale quantitativ sind:

(1) Blutgruppe
(2) Pulsfrequenz
(3) Raucher
(4) Teilnahme an der heutigen Prüfung
(5) Punktzahl der heutigen Prüfung

(A) nur 1 und 3 sind quantitativ
(B) nur 1 und 4 sind quantitativ
(C) nur 2 und 5 sind quantitativ
(D) nur 1, 3 und 4 sind quantitativ
(E) 1–5 = alle sind quantitativ

2.2 Graphische Darstellung

8. Mit Hilfe eines Histogramms

(A) kann die Stärke der Abhängigkeit zweier Zufallsvariablen graphisch dargestellt werden.
(B) können computertomographisch gewonnene Gewebequerschnitte varianzanalytisch ausgewertet werden.
(C) werden Messwerte auf Vorliegen einer Normalverteilung getestet.
(D) werden klassierte stetige Merkmale graphisch so dargestellt, dass die relativen Klassenhäufigkeiten als Rechtecke erscheinen.
(E) werden relative Häufigkeiten in Wahrscheinlichkeiten umgerechnet.

2.3 Beschreibung durch Maßzahlen

9. Darf man bei diskreten Merkmalen, z.B. Kinderzahl, Mittelwerte berechnen?

(A) Man darf Mittelwerte nur errechnen, wenn sich ganze Zahlen ergeben.
(B) Mittelwerte dürfen überhaupt nicht berechnet werden, weil sie beliebige Zwischenwerte annehmen können, die im Einzelfall real nicht vorkommen.
(C) Man darf Mittelwerte auch bei diskreten Merkmalen ohne Einschränkung berechnen.
(D) Man muss dabei die Formel für geometrische Mittel benutzen.
(E) Keine Antwort ist richtig.

10. Welche Aussage trifft **nicht** zu?

Der Mittelwert $\bar{x}$ ist

(A) stets größer als der Median
(B) ausreißerempfindlich
(C) ein Schätzwert
(D) ein Lagemaß
(E) ein Durchschnitt

Lösung der Übungsaufgaben

1. (D) Beobachtungseinheit oder Merkmalsträger sind die Patienten, das beobachtete Merkmal ist das Körpergewicht und die Maßeinheit dieses Merkmals ist Kilogramm.

2. (C) Die Krankenschwester würde man als Untersucher(in) bezeichnen. Beobachtungseinheit ist jedenfalls der Patient.

3. (B) Die Zuordnung 1–a (Patient Meier = Beobachtungseinheit), 2–d (45/min = Ausprägung) und 3–c (Herzfrequenz = Merkmal) dürfte nach den zu Frage 1 und 2 gemachten Kommentaren einleuchtend sein, so dass nur noch zwischen 4–b und 5–b zu unterscheiden ist. Die Grundgesamtheit besteht aus allen an Bradykardie leidenden Patienten, im Krankenhaus wird eine Stichprobe dieser Grundgesamtheit behandelt, Patient Meier ist Teil dieser Stichprobe. Für die Erkrankung Bradykardie gibt es in diesem Zusammenhang keinen statistischen Fachterminus.

4. (D) Blutgruppe und Kammerflimmern sind qualitative Merkmale, weil sich ihre Merkmalsausprägungen (Blutgruppe: A, B, AB, 0; Kammerflimmern: vorhanden/nicht vorhanden) nicht zahlenmäßig erfassen lassen. Es liegt eine sog. Nominalskala vor.

5. (E) Körpergewicht, Nüchternblutzucker und spezifisches Gewicht des Urins sind quantitativ stetige Merkmale, weil sie in einem bestimmten Bereich jeden beliebigen Zwischenwert annehmen können.

6. (C) Siehe Erläuterung zu Aufgabe 4.

7. (C) Die Merkmale Blutgruppe, Raucher und Prüfungsteilnahme lassen sich nur durch eine Verschlüsselung zahlenmäßig erfassen. Eine Verschlüsselung ist jedoch willkürlich, so dass den Schlüsselzahlen keine quantitative Bedeu-^{n}tung zukommt und etwa eine Mittelwertsberechnung sinnlos wäre. Mit oder ohne Verschlüsselung handelt es sich um rein qualitative Merkmale (Nominalskala).

8. (D) Bei diskreten Merkmalen ist eine Klassierung nicht erforderlich, manchmal jedoch sinnvoll, um die Zahl der Klassen und damit der Balken zu verringern und die Übersichtlichkeit zu erhöhen.

9. (C) Der Einwand, dass der arithmetische Mittelwert $\bar{x}$ bei diskreten Werten zu einem Ergebnis führen kann, z.B. 1,85 Kinder, welches im realen Fall nicht vorkommen kann, spricht nicht gegen die Berechnung, weil Mittelwerte keine Aussagen über Einzelfälle machen, sondern nur über die gesamte Gruppe, für die sie berechnet werden.
Das geometrische Mittel $\bar{x}_{\mathrm{g}}$ von n Werten errechnet sich als n-te Wurzel ihres Produktes:

$$\bar{x}_g = \sqrt[2]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Das geometrische Mittel wird für durchschnittliche Wachstumsquoten (Bakterienkultur, Zins und Zinseszins, Inflation) verwendet.
Beispiel: Die Inflationsrate beträgt im 1. Jahr 50 % und im 2. Jahr 10 %. Wie hoch ist die durchschnittlliche jährliche Inflationsrate?

$$\bar{x}_g = \sqrt[2]{1{,}5 \cdot 1{,}1} = \sqrt[2]{1{,}65} \approx 1{,}28$$

Die durchschnittliche jährliche Inflationsrate beträgt 28 %.

10. (A) Es gibt drei Lagemaße zur Kennzeichnung eines Durchschnitts: den arithmetischen Mittelwert $\bar{x}$, den Median und den Modalwert. In welcher Beziehung diese drei Werte zueinander stehen, hängt im Einzelfall von den Werten der Verteilung ab.

Übungsaufgaben

2.4 Lagemaße (Lokalisationsmaße)

11. Gegeben ist folgende Häufigkeitstabelle für Daten eines stetigen Merkmales ($(a, b]$ bedeutet: $a < x \leq b$)

Klasse	absolute Häufigkeit
(2,4]	20
(4,6]	10
(6,8]	10
(8,10]	10

Der arithmetische Mittelwert der klassierten Daten ist

(A) 5,4
(B) 6,0
(C) 6,2
(D) 7,1
(E) 7,2

12. Eine Beobachtungsreihe soll aus den 17 Messwerten $x_1, \ldots, x_{17}$ bestehen. Der empirische Median dieser Beobachtungen ist definiert als

(A) x_9
(B) $x_{8,5}$
(C) x_8
(D) $½(x_{18} + x_9)$
(E) $½(x_1 + x_{17})$

13. Bei 10 Patienten, die an Krebs einer bestimmten Art operiert wurden, betrug die Überlebensdauer 3, 4, 5, 5, 5, 7, 8, 9, 11, 12 Monate.
Welches ist der Median (in Monaten)?

(A) 5 (B) 5,5 (C) 6 (D) 6,5 (E) 7

14. In einer Schulklasse wird der Kariesbefall der Zähne bei Kindern untersucht. Es ergibt sich die folgende Häufigkeitstabelle:

Anzahl mit Karies befallener Zähne	0	1	2	3	4
absolute Häufigkeit	4	7	5	3	1

Der empirische Median der Daten ist

(A) 1 (B) 1,5 (C) 2 (D) 4 (E) 5

15. In einer Schulklasse wurden die Zähne von 10 Kindern auf Kariesbefall untersucht. Die Anzahl der mit Karies befallenen Zähne für die einzelnen Kinder ist:

Kind Nr.	1	2	3	4	5	6	7	8	9	10
Anzahl mit Karies befallener Zähne	0	1	1	2	2	1	0	4	1	1

Der empirische Median der Daten ist

(A) 1 (B) 1,3 (C) 1,5 (D) 4 (E) 5,5

16. Von 10 Frauen wurden Daten zur Fruchtlage erhoben. Diese Daten sind folgendermaßen verschlüsselt:

(1) normal
(2) Beckenendlage
(3) Querlage
(4) fehlende Angabe

Patient Nr.	1	2	3	4	5	6	7	8	9	10
Fruchtlage	1	4	1	3	1	1	2	1	4	2

Der empirische Median der Fruchtlage ist:

(A) 1
(B) 1,5
(C) 2
(D) 5,5
(E) Der empirische Median ist nur für quantitative Merkmale definiert.

17. In einer Stichprobe von 10 Beobachtungen 90, 92, 95, 96, 98, 102, 106, 107, 109, 118 ergab sich als Median der Wert 100 und als Mittelwert $\bar{x} = 101{,}3$. Nachträglich stellte man fest, dass die Messergebnisse 95, 96, 106 um jeweils 3 Einheiten zu „hoch“ abgelesen wurden. Wie verändern sich durch diese Verbesserungen Median und Mittelwert?

(A) Median und Mittelwert verkleinern sich um 0,3.
(B) Median und Mittelwert verkleinern sich um 0,9.
(C) Der Median bleibt gleich, der Mittelwert verkleinert sich um 0,3.
(D) Der Median bleibt gleich, der Mittelwert verkleinert sich um 0,9.
(E) Median und Mittelwert bleiben gleich.

18. Der Median einer Stichprobe ändert sich **nicht**, wenn man

(A) jeweils ein Datum mit dem größten und kleinsten Wert weglässt.
(B) alle Werte außerhalb der 2σ-Grenze weglässt.
(C) zu allen Beobachtungen eine Konstante hinzuaddiert.
(D) alle Beobachtungen mit der gleichen Zahl multipliziert.
(E) alle Beobachtungen logarithmiert.

Lösung der Übungsaufgaben

11. (A) Für die Berechnung des Mittelwertes der Tabelle gehen wir davon aus, dass der durchschnittliche Wert in jeder Klasse mit der Klassenmitte übereinstimmt, dass z.B. die 20 Werte in der Klasse (2,4] im Durchschnitt 3 betragen, zusammen also 60 ergeben. Diese Annahme ist spekulativ, muss jedoch gemacht werden, um die Aufgabe überhaupt lösen zu können. Folgende Indizien sprechen dafür, dass der hierdurch entstehende Fehler vermutlich nicht sehr groß ist:

(1) Die Klassenbreiten sind gleich groß.
(2) Die Klassenbesetzung mit 20 bzw. 10 Daten je Klasse ist relativ groß.
(3) Es ist ein stetiges Merkmal, das sich möglicherweise gleichmäßig innerhalb der Klasse verteilt.

Wenn es ein diskretes ganzzahliges Merkmal wäre, könnten in der Klasse (2,4] gemäß der in der Aufgabenstellung definierten Bedeutung der Klammern nur die Werte 3 und 4 vorkommen. Wir müssten den Durchschnitt innerhalb der Klasse dann als 3,5 annehmen.

Die Berechnung von $\overline{x}$ ergibt:

$$\begin{aligned} 20 \cdot 3 &= 60 \\ 10 \cdot 5 &= 50 \\ 10 \cdot 7 &= 70 \\ 10 \cdot 9 &= 90 \\ \Sigma &= \mathbf{270} \end{aligned}$$

$$\overline{x} = \frac{\Sigma x}{n} = 270/50 = 5{,}4$$

Der arithmetische Mittelwert der klassierten Daten beträgt 5,4. Die Berechnung aus den unklassierten Orginalwerten würde vermutlich zu einem geringfügig abweichenden Ergebnis führen.

12. (A) Zur Beantwortung dieser Frage muss man davon ausgehen, dass die mit den Indizes 1 bis 17 bezeichneten Messwerte bereits der Größe nach sortiert sind. Dann ist der Wert x_9 der Median, weil er in der Mitte aller Werte steht, denn acht sind kleiner/gleich und acht sind größer/gleich x_9.
Es ist möglich, dass einige der 16 verbleibenden Werte mit x identisch sind. Aber auch in diesem Fall steht x in der Mitte aller Werte, so dass die Definition des Medians erfüllt ist.

13. (C) Streng genommen erfüllt jede der unter (A) bis (E) genannten Zahlen die Definition des Medians, die besagt, dass nicht mehr als die Hälfte aller Werte kleiner und nicht mehr als die Hälfte der Werte größer als der Median sein dürfen. Aufgrund einer Konvention gilt der in der Mitte aller „Mediane" liegende Wert als Median.

Sachliche Anmerkung: Wenn die Operationen die letzten 4 Patienten völlig geheilt hätten, würde sich die mediane Überlebenszeit nicht verlängern.

14. (A) Von den insgesamt 20 Kindern haben 11 Kinder einen oder keinen mit Karies befallenen Zahn. „1 kariöser Zahn" ist sowohl Median (Kind „Nr. 10,5") als auch Modalwert (sieben Kinder mit einem kariösen Zahn).

15. (A) Zur Lösung dieser Aufgabe muss man die willkürlich durchnummerierten Kinder zunächst nach der Anzahl ihrer kariösen Zähne sortieren und dann „Kind Nr. 5,5" ermitteln. Zwei Kinder haben keinen und fünf Kinder haben einen kariösen Zahn. „Kind Nr. 5,5" hat deshalb ebenfalls einen kariösen Zahn.

16. (E) Die Bestimmung des Medians ist nur bei Daten sinnvoll, die einer Ordinal-(Rang)skala oder einem höheren Skalenniveau entsprechen. Bei der Verschlüsselung der Fruchtlage handelt es sich um eine Nominalskala, so dass höchstens die Bestimmung des Modalwertes sinnvoll wäre. Das wäre hier die normale Lage, die insgesamt fünfmal vorkommt.

17. (D) Der Median bleibt gleich, weil der Wert 106 auch nach der Subtraktion von 3 weiterhin größer als der Median ist. Deshalb beträgt auch nach der Berichtigung die Hälfte aller Werte 102 oder mehr, und die zweite Hälfte aller Daten beträgt 98 oder weniger. Der Median liegt deshalb auch nach der Datenberichtigung in der Mitte zwischen 98 und 102.

Die Gesamtsumme aller 10 Werte verringert sich um $3 \cdot 3 = 9$, wodurch der arithmetische Mittelwert um 9/10 = 0,9 kleiner wird.

18. (A) Siehe Erläuterungen zu den Aufgaben 12–17.

Übungsaufgaben

19. Welche Maßzahl einer Stichprobe aus einer stetigen Zufallsvariablen ist bei Vorliegen von Ausreißern besonders wenig störanfällig:

(A) geometrischer Mittelwert
(B) Spannweite (Range)
(C) $\tilde{x}$ = Median
(D) $\bar{x}$ = arithmetischer Mittelwert
(E) Varianz s^2

20. Welche Aussage trifft **nicht** zu?

Der Median ändert sich,

(A) wenn zum Mittelwert eine Zahl addiert wird.
(B) wenn zu jedem Wert eine Zahl addiert wird.
(C) wenn der höchste und niedrigste Wert weggelassen werden
(D) wenn m.an den höchsten Wert weglässt.
(E) wenn man den niedrigsten Wert weglässt.

21. In einer Stichprobe nehmen die relativen Häufigkeiten, mit denen die einzelnen Werte eines klassierten quantitativen Merkmals auftreten, mit wachsenden Merkmalswerten ab. Dies lässt vermuten:

(A) Die Beobachtungen sind voneinander abhängig.
(B) Der Median ist kleiner als der Mittelwert.
(C) Die Beobachtungen stammen nicht aus der gleichen Grundgesamtheit.
(D) Die Standardabweichung ist größer als der Erwartungswert.
(E) Es liegt eine Normalverteilung zugrunde.

2.5 Streuungs- oder Dispersionsmaße

22. Die Dauer eines Krankenhausaufenthaltes wegen einer bestimmten Krankheit ist eine Zufallsvariable. Sie besitzt stets dann eine große Varianz, wenn

(A) es sich um eine seltene Krankheit handelt.
(B) es sich um eine häufige Krankheit handelt.
(C) bei dieser Krankheit große Abweichungen von ihrer durchschnittlichen Dauer eine große Wahrscheinlichkeit haben.
(D) die durchschnittliche Dauer der Erkrankung groß ist.
(E) der Krankheitsverlauf von vielen, zum Teil unbekannten Faktoren abhängt.

Aus 11 zufällig gezogenen Krankenblättern einer Klinik ergeben sich als Aufenthaltsdauer, der Größe nach geordnet, in Tagen 4, 7, 8, 8, 9, 9, 10, 10, 10, 12, 12 (Σ = 99).

Ordnen sie jedem der in Liste 1 aufgeführten Maße den dafür richtigen Wert der Liste 2 zu.

Liste 1	Liste 2
23. Spannweite	(A) 8
24. Median	(B) 9
25. Mittelwert	(C) 10
	(D) 11
	(E) 12

26. Zwei Zahlengruppen liegen vor:

I: 1, 3, 5, 7
II: 2, 4, 6

Welche der folgenden Behauptungen trifft **nicht** zu?

(A) I und II haben gleiche Mediane.
(B) I und II sind beide symmetrisch bezüglich der Mittelwerte.
(C) I und II haben gleiche Mittelwerte.
(D) I und II haben gleiche empirische Standardabweichungen.
(E) I und II haben ungleiche Spannweiten.

27. Für welche der folgenden Maßzahlen ist die Summe der Abweichungsquadrate der Einzelwerte von der betreffenden Maßzahl am kleinsten?

(A) Standardabweichung
(B) Varianz
(C) Median
(D) Modus
(E) Mittelwert

28. Welche Aussage trifft **nicht** zu?
Als Maßzahlen für die Variabilität in einer Stichprobe sind geeignet ($\bar{x}$ = Mittelwert, $\tilde{x}$ = Median):

(A) $\Sigma\,(x_i - \bar{x})^2$
(B) $\Sigma\,(|x_i - \tilde{x}|)$
(C) $\Sigma\,(x_i - \bar{x})$
(D) $\Sigma\,(|x_i - \bar{x}|)$
(E) die Differenz zwischen größtem und kleinstem Wert der Stichprobe

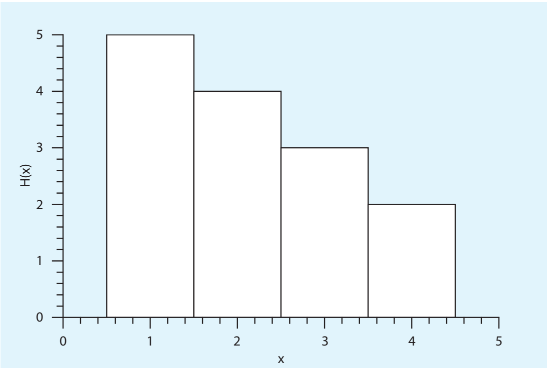
Lösung der Übungsaufgaben

19. (C) „Zufallsvariable" ist in diesem Zusammenhang ein zufallsabhängiges quantitatives Merkmal (genaue Definition der Zufallsvariablen s.S. 38). Der Modalwert wird von den Ausreißern überhaupt nicht beeinflusst, er ist jedoch nur sinnvoll für die Stichprobe eines diskreten oder eines klassierten stetigen Merkmals. Der Modalwert ist der häufigste Wert, ein Messwert muss also mindestens zweimal aufgetreten sein, um Modalwert zu sein. In der Stichprobe eines stetigen Merkmals wird jeder Messwert häufig nur ein einziges Mal auftreten, besonders wenn die Messwerte mit vielen Stellen hinter dem Komma versehen sind. Deshalb existiert in einer Stichprobe eines nicht klassierten stetigen Merkmals häufig kein Modalwert.

20. (C) Vgl. Erläuterung zu den Aufgaben 12–17.

21. (B) Untenstehend ist die in der Aufgabenstellung genannte Situation beispielhaft dargestellt.
Der Median beträgt 2, weil 5 Werte größer und 5 Werte kleiner als 2 sind.
Der arithmetische Mittelwert $\bar{x}$ errechnet sich als:

$$\bar{x} = \frac{5 \cdot 1 + 4 \cdot 2 + 3 \cdot 3 + 2 \cdot 4}{5 + 4 + 3 + 2} = 30/14 = 2{,}14$$



Tatsächlich ist der Median kleiner als der Mittelwert. Dies liegt daran, dass die hohen x-Werte den Mittelwert stärker beeinflussen (weil sie weiter weg sind) als die niedrigen x-Werte.

22. (C) Bei der Berechnung der Varianz werden die Abweichungen vom Mittelwert quadriert. Deshalb erhöhen besonders Einzelwerte mit großen Abweichungen vom Mittelwert, insbesondere auch Ausreißer, die Varianz. Wenn die Abweichung aus vielen einzelnen, entsprechend kleinen Komponenten besteht, kann dies die Varianz sogar reduzieren, wenn sich positive und negative Abweichungskomponenten gegenseitig kompensieren. Ob die einzelnen Abweichungskomponenten bekannt oder unbekannt sind, spielt keine Rolle, für die Varianz ist nur die Gesamtabweichung maßgeblich.

23. (A) Die Spannweite (range) ergibt sich als Differenz zwischen dem größten und dem kleinsten Wert: $12 - 4 = 8$.

24. (B) Der Median liegt in der Mitte der 11 Werte, so dass 5 Werte kleiner und 5 Werte größer als der Median sind.

25. (B) Der arithmetische Mittelwert errechnet sich als Summe aller Werte geteilt durch die Anzahl aller Werte:

$$\bar{x} = 99/11 = 9.$$

26. (D) Median und arithmetischer Mittelwert beider Gruppen beträgt 4. Beide Gruppen liegen symmetrisch zum Median. Die Spannweite von Gruppe I beträgt 7–1 = 6 und von Gruppe II 6 – 2 = 4.
Für die empirische Standardabweichung s ergibt sich:

Gruppe I:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{(-3)^2 + (-1)^2 + 1^2 + 3^2}{3}} = \sqrt{\frac{20}{3}} = 2{,}58$$

Gruppe II:

$$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}} = \sqrt{\frac{(-2)^2 + 2^2}{2}} = \sqrt{4} = 2$$

27. (E) Die Summe der Abweichungsquadrate wird bestimmt, um die

$$\text{Varianz } s^2 = \frac{1}{n-1} \sum (x-\bar{x})^2$$

zu errechnen. Wenn anstelle des arithmetischen Mittelwertes $\bar{x}$ eine andere Maßzahl eingesetzt wird, etwa der Median oder der Modalwert, ergibt sich bei der Summe der Abweichungsquadrate ein höherer Wert. Der arithmetische Mittelwert ist die Maßzahl, zu der die Summe der Abweichungsquadrate minimal ist.

28. (C) Der Ausdruck $\sum (x - \bar{x})$ ergibt dann, wenn die Summe für alle x-Werte gebildet wird, aus denen $\bar{x}$ errechnet wurde, stets Null.

Beispiel: Die Stichprobe sei 1, 3, 5, 8, 10. Hieraus ergibt sich $\bar{x} = (1 + 3 + 5 + 8 + 10)/5 = 5{,}4$.
$\sum (x - \bar{x}) = -4{,}4 + (-2{,}4) + (-0{,}4) + 2{,}6 + 4{,}6$
$= -7{,}2 + 7{,}2 = 0$
Die positiven und negativen Differenzen ergänzen sich gegenseitig zu Null.

Übungsaufgaben

29. Zur Berechnung der Varianz kann man von jedem einzelnen Messwert eine feste Größe a abziehen (bei der Körperlänge von Erwachsenen z. B. 100 cm), damit bei der Rechnung nicht zu große Zahlen auftreten. Dies muss am Schluss der Rechnung berücksichtigt werden durch

(A) Addition von a
(B) Multiplikation mit a
(C) Addition von a^2
(D) Multiplikation mit a^2
(E) a wird bei der Berechnung der Varianz nicht berücksichtigt

30. Bei einem Versuch werden drei Versuchsgruppen von je 10 Ratten gebildet und ihre Gewichte festgestellt. Dabei gilt stets Folgendes:

(A) Der arithmetische Mittelwert aller Daten ist gleich dem arithmetischen Mittelwert der Mittelwerte der drei Gruppen.
(B) Der Median aller Daten ist gleich dem Median der Mediane der drei Gruppen.
(C) Die Summe der Abweichungsquadrate aller 30 Daten vom Gesamtmittelwert ist gleich der Summe der Abweichungsquadrate in den drei Gruppen von den Gruppenmittelwerten.
(D) Die empirische Varianz aller Daten ist gleich dem Mittelwert der drei Varianzen in den drei Gruppen.
(E) Die Spannweite aller Daten ist gleich dem Maximum der Spannweite der drei Gruppen.

31. Welche der folgenden Kenngrößen einer Stichprobe wird (werden) mit zunehmendem Stichprobenumfang nie kleiner?

(1) Mittelwert
(2) empirische Varianz
(3) empirischer Median
(4) Spannweite
(5) Abweichungssumme vom Mittelwert

(A) nur 5 ist richtig
(B) nur 2 und 4 sind richtig
(C) nur 4 und 5 sind richtig
(D) nur 1, 2 und 5 sind richtig
E) nur 2, 3, 4 und 5 sind richtig

32. Welche Aussage trifft **nicht** zu?

Werden die Beobachtungen in Zentimetern statt in Millimetern ausgedrückt, ändert sich

(A) die Varianz
(B) das arithmetische Mittel
(C) die Standardabweichung
(D) der Median
(E) der Variationskoeffizient

33. Welche Aussage über Streuungsmaße trifft nicht zu?

(A) Für Merkmale mit positiven und negativen Ausprägungen ist es nicht sinnvoll, den Variationskoeffizienten zu verwenden.
(B) Die Spannweite ist ausreißerempfindlich.
(C) Die Varianz ist nur für quantitative Merkmale definiert.
(D) Die Varianz kann stets aus der Standardabweichung ermittelt werden.
(E) Die Varianz ist stets ein Maß für die Abweichung der Beobachtungen von ihrem Median.

2.6 Die empirische Verteilungsfunktion

34. Um die Häufigkeitssummen einer Stichprobe für ein quantitatives Merkmal festzustellen, ist es zweckmäßig, die Daten zunächst

(A) mit ihrer relativen Häufigkeit zu gewichten
(B) nach der Größe zu ordnen
(C) logarithmisch zu transformieren
(D) auf Mittelwert = 0 zu normieren
(E) es zu quadrieren

35. Die empirische Verteilungsfunktion $F(x)$ gibt zu jedem Wert x an:

(A) die Wahrscheinlichkeit, mit der die Beobachtungen kleiner oder gleich x sind
(B) die relative Häufigkeit von Beobachtungen kleiner oder gleich x
(C) die relative Häufigkeit von Beobachtungen größer oder gleich x
(D) die Besetzungszahl der Klasse, in die x fällt
(E) die relative Häufigkeit, mit der x beobachtet wurde

36. Die Verteilungsfunktion $F(x)$ gibt zu jedem Wert die Wahrscheinlichkeit an, dass die Zufallsvariable X einen Wert annimmt, der

(A) kleiner als x ist
(B) kleiner oder gleich x ist
(C) gleich x ist
(D) größer oder gleich x ist
(E) größer als x ist

37. Bei einer Untersuchung wurden die Körpergrößen von 300 Schulkindern gemessen. Dabei wurde festgestellt, dass 60 % der Kinder kleiner als 145 cm sind. Die relative Häufigkeit der Kinder, die mindestens 145 cm groß sind, ist

(A) 0,4
(B) 0,6
(C) 40
(D) 120
(E) 180

Alter in Jahren	bis 35	36–45	46–55	56–65	66–75
abs. Häufigkeit	17	116	493	545	186

38. Gegeben sei die obige Tabelle der absoluten Häufigkeit des Bronchialkarzinoms bei 1357 Männern verschiedener Altersklassen:
Die relative Häufigkeit höchstens 55 Jahre alter Männer unter diesen Patienten ist

(A) 493/1357
(B) (493 + 545 + 186)/1357
(C) (17 + 116 + 493)/1357
(D) (35 + 45 + 55)/1357
(E) 55/1357

Lösung der Übungsaufgaben

29. (E) Wenn von jedem x-Wert die konstante Größe a abgezogen wird, ist auch der arithmetische Mittelwert $\bar{x}$ um die Größe a kleiner: $x_i - a = x'_i$
daraus folgt: $\bar{x} - a = \bar{x}'$.

Die Abweichung vom Mittelwert $(x_i - \bar{x})$ bleibt jedoch konstant: $(x_i - \bar{x}) = (x'_i - \bar{x}')$. Entsprechend bleibt das Abweichungsquadrat und damit bleibt auch die Summe aller Abweichungsquadrate konstant.

30. (A) Aussage A ist richtig. Dies sei an einem vereinfachten Beispiel erläutert:

Gruppe	I:	4, 5, 9	$\bar{x} = 18/3 = 6$
Gruppe	II:	4, 7, 10	$\bar{x} = 21/3 = 7$
Gruppe	III:	12, 14, 16	$\bar{x} = 42/3 = 14$

Der arithmetische Mittelwert $\bar{x}_g$ aller Gruppen errechnet sich als

$$\bar{x}_g = (18 + 21 + 42)/(3 + 3 + 3) = 81/9 = 9$$

oder als $\bar{x}_g = 27/3 = 9$

Die Rechnung geht nur deshalb auf, weil alle Stichproben denselben Stichprobenumfang haben.

(B) ist falsch, in unserem Beispiel beträgt der Median aller Daten 9, der Median der Mediane ist 7.

Die unter (C) und (E) genannten Streuungsparameter sind für alle drei Gruppen zusammen in der Regel größer als innerhalb der einzelnen Gruppen.

31. (C) Wenn bei der Erweiterung einer Stichprobe Werte hinzukommen, die kleiner als der kleinste oder größer als der größte bisherige Wert sind, vergrößert sich die Spannweite. Die Abweichungsquadratsumme vom Mittelwert wird durch alle hinzukommenden Werte erhöht, die nicht mit dem Mittelwert identisch sind. Die Varianz hingegen wird nicht zwangsläufig erhöht, weil bei größer werdendem Stichprobenumfang n auch der Nenner der entsprechenden Formel größer wird.

32. (E) Der Variations- oder Variabilitätskoeffizient ist als $s/\bar{x} \cdot 100\,\%$ definiert. Wenn die Werte in Millimetern statt in Zentimetern angegeben werden, erhöht sich sowohl der Zähler als auch der Nenner um den Faktor 10, der Quotient bleibt unverändert.

33. (E) Die Varianz ist ein Maß für die Abweichungen vom arithmetischen Mittelwert $\bar{x}$.

34. (B) Die Häufigkeitssumme eines Merkmals x_i gibt an, wie häufig ein Merkmal vorgekommen ist, das kleiner oder genauso groß wie x_i ist. In Tabelle 2.3 des Lehrbuchs ist in der rechten Spalte die Häufigkeitssumme für alle Entlassungstage x angegeben. In dieser Tabelle sind die Liegezeiten ihrer Größe nach angeordnet.

35. (B) Die relative Häufigkeit kann als Schätzwert für die Wahrscheinlichkeit dienen, ist aber selbst noch keine Wahrscheinlichkeit.

36. (B) Zufallsvariable bedeutet hier ein zufallsabhängiges quantitatives Merkmal, genaue Definition s. S. 38 f.

37. (A) Die relative Häufigkeit beträgt 1 - 0,6 = 0,4, denn die Kinder sind entweder kleiner als 145 cm oder mindestens 145 cm groß.
Die absolute Häufigkeit beträgt 0,4 · 300 = 120.

38. (C) Insgesamt sind 17 + 116 + 493 = 626 Männer höchstens 55 Jahre alt. Die relative Häufigkeit ergibt sich als Division durch die Gesamtzahl aller Männer.