Statistik

Übungen, Lösungen, hilfreiche Tipps

von

Erhard Schulze

Oldenbourg Verlag München

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
oldenbourg.de

Lektorat: Rainer Berger
Herstellung: Anna Grosser
Coverentwurf: Kochan & Partner, München
Cover-Illustration: Hyde & Hyde, München
Gedruckt auf säure- und chlorfreiem Papier
Gesamtherstellung: Grafik + Druck GmbH, München

ISBN 978-3-486-70265-1

Vorwort

Dieses Projekt lief lange Zeit unter dem Arbeitstitel „Aufgabensammlung“. Es sollte also zunächst eine reine Aufgabensammlung für FH-Studenten, mit vielleicht einigen wenigen Erläuterungen werden. Wenn es dann doch mehr geworden ist, so ersetzt es nicht das Studium grundlegender Statistiklehrbücher. Aus meiner mehrjährigen Erfahrung als Lehrbeauftragter für Statistik an der Hochschule Augsburg weiß ich jedoch um die Schwierigkeiten, die viele Studenten mit allzu formalisierten Lehrbüchern haben. Deshalb soll dieses Buch als eine Art Untersetzung fungieren (etwa Langenscheidt „Statistikprofessorisch – Studentisch“), ohne dabei flach zu werden. Die Verwendung der wichtigsten Formeln ist an Hand ausführlicher (für einige vielleicht zu ausführlicher – niemand sollte sich beleidigt fühlen) Beispiele erläutert worden und vor allem gibt es nachvollziehbare Schritt-für-Schritt-Lösungen für alle Aufgaben im Anhang. Bei ernsthafter Bearbeitung des Buches sollte jeder Student in der Lage sein, alle Aufgaben – vielleicht bis auf die mit ° gekennzeichneten (erhöhter Schwierigkeitsgrad) zu lösen, und nach Bearbeitung des Buches sollte die Statistikprüfung ihren Schrecken zum großen Teil verloren haben. Wesentliche Grundlage dieses Buches ist das Lehrbuch „Statistik“ von Prof. Dr. Günter Bamberg, Dr. Franz Baur und Prof. Dr. Michael Krapp (Oldenbourg-Verlag), dessen Gebrauch ich Ihnen hiermit ans Herz lege, allein schon der Themen wegen, die ich aus den verschiedensten Gründen hier nicht behandeln konnte oder wollte. Es gibt natürlich auch jede Menge anderer geeigneter Lehrbücher, im Anhang finden Sie einige Vorschläge. Schließlich möchte ich Herrn Prof. Dr. Günter Bamberg sehr herzlich danken für die wirklich sehr gründliche Durchsicht des Manuskripts und seine hilfreichen Änderungsvorschläge.

Inhalt

1 Deskriptive Statistik

1.1 Grundbegriffe, Skalierung, Klassierung

1.1.1 Grundbegriffe

Folgende Grundbegriffe muss man kennen:

Merkmalsträger: statistische Einheit, an der die Merkmale (Alter, Geschlecht, usw.) erhoben wurden.

Merkmalsausprägung: entweder quantitative Merkmale (Zahlen) oder qualitative Merkmale (Geschlecht, Farbe, Beruf,..)

Grundgesamtheit: Gesamtheit der für die statistische Untersuchung relevanten Merkmalsträger

1.1.2 Skalierung

In der Statistik werden drei Skalen unterschieden:

Nominalskala: qualitative Merkmale werden nominal skaliert, die verschiedenen Namen können und werden meist quantifiziert.
Beispiel: Merkmal: Geschlecht, männlich = 0, weiblich = 1. Es kommt lediglich auf die Unterscheidung, nicht auf den Wert der Zahlen an. Genau so gut könnte man auch 17 und 23 wählen.

Ordinalskala: Es kommt nicht nur auf die Unterscheidung sondern auch auf die Rangfolge an.
Beispiel: 1., 2., 3. beim Pferderennen, ohne die konkreten Zeitabstände. (Auch hier könnte man theoretisch andere Zahlen zuordnen.)

Kardinalskala: Nicht nur die Rangordnung, sondern auch die Abstände zwischen den einzelnen Rängen sind wichtig.
Beispiel: Die jeweils erreichten Punkte bei einer Klausur.

Aufgaben:

1) Welche Skalierung liegt den folgenden Merkmalen zu Grunde?

   a) Produktionsdauer
   b) Schulnote
   c) Schwierigkeitsgrad einer Skipiste
   d) Gewicht von Schultaschen
   e) Zugehörigkeit zu einer bestimmten Konfession

2) Für 6 Romane wurde vom deutschen Buchhandel folgende Tabelle erstellt:

Roman	1	2	3	4	5	6
Sparte	Liebe	Krimi	Abenteuer	Biografie	Erotik	Historie
Verkaufserfolg	3	2	1	4	5	6
Preis	17,80	12,95	23,90	29,90	7.95	13,80

a) Was sind in diesem Beispiel die Begriffe
   - Merkmalsträger:
   - Merkmal:

b) Welches Skalenniveau haben die drei Merkmale?

1.1.3 Klassierung

Zunächst einmal ist zwischen diskreten und stetigen Merkmalen zu unterscheiden:

Diskrete Merkmale:

Salopp gesagt, alles, was einzeln zählbar ist (oder immer, wenn Sie im Englischen „many" sagen würden). Das heißt aber nicht, dass diskrete Merkmale auf ganze Zahlen reduziert werden können, auch die Ausprägungen 0,1; 0,2 usw. können diskret sein, wichtig ist, dass zwischen je zwei benachbarten Ausprägungen keine weiteren Werte zugelassen sind.

Stetige Merkmale:

Für je zwei Merkmalsausprägungen gelten auch alle Zwischenwerte. Typisch sind Zeit, Länge, Gewicht, die theoretisch beliebig genau gemessen werden können.

Stetige Merkmale müssen, diskrete Merkmale können in Klassen unterteilt werden.

Diskrete Merkmale mit sehr vielen Ausprägungen werden meist in Klassen unterteilt.

(Beispiel: Einkommen der Beschäftigten einer großen Firma oder Altersstruktur der Bevölkerung eines Landes)

Die Klassengrenzen werden meist gebildet *von ... bis unter*.

Mittelwerte und Streuungsmaße von klassierten Größen lassen sich erst ermitteln, wenn man die Klassenmitten ((Untergrenze + Obergrenze)/2) gebildet hat.

1.1.4 Häufigkeiten und graphische Darstellung

1.1.4.1 Absolute und relative Häufigkeit

Die folgenden Häufigkeiten sollte man kennen und unterscheiden können:

- absolute Häufigkeit h(a)
- absolute kumulierte Häufigkeit H(x)
- relative Häufigkeit f(a)
- relative kumulierte Häufigkeit F(x) (empirische Verteilungsfunktion)

Beispiel 1:

In einer Urliste werden die Augenzahlen bei 12-maligem Würfeln notiert:

3, 4, 2, 5, 5, 1, 3, 6, 5, 1, 5, 1

Die geordnete Urliste ergibt:

1	1	1	2	3	3	4	5	5	5	5	6
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}

Bitte beachten Sie die Unterscheidung x_i (für jedes einzelne Beobachtungsmerkmal) und a_j für die Merkmalsausprägungen.

Merkmalsausprägung:

1, 2, 3, 4, 5, 6

a_1, a_2, a_3, a_4, a_5, a_6

Als Tabelle:

Ausprägung (aj)	1	2	3	4	5	6
abs. Häufigkeit h(aj)	3	1	2	1	4	1
abs. Kum. Häufigkeit H(x)	3	4	6	7	11	12
rel. Häufigkeit f(aj)	3/12	1/12	2/12	1/12	4/12	1/12
rel. Kum. Häufigkeit F(x)	3/12	4/12	6/12	7/12	11/12	12/12

1.1.4.2 Graphische Darstellung

Es gibt eine Unmenge an Möglichkeiten, Daten graphisch darzustellen, von den wichtigsten

- Stabdiagramm,
- Kreisdiagramm und
- Histogramm

konzentrieren wir uns auf das Histogramm. Stabdiagramm und Kreisdiagramm erklären sich quasi von selbst, das Histogramm bietet dagegen einige Feinheiten, die Sie beachten sollten.
Histogramm:

Beispiel 2:

Bei einer Stichprobe von n = 20 Schülern einer 8. Klasse wurden folgende Sparguthaben in € festgestellt.

1000	580	520	350	620	800	120
600	550	420	470	200	560	480
1000	600	1150	800	250	650	

Erstellen Sie unter Verwendung der Klassengrenzen 0; 300; 500; 700; 1200 ein Histogramm!

Lösung:

Von .. bis unter:	absolute Häufigkeit	relative Häufigkeit
0 – 300	3	0,15
300 – 500	4	0,2
500 – 700	8	0,4
700 – 1200	5	0,25

Die x-Achse bildet die Merkmalsausprägung.

Achtung: Nicht die Höhe der Rechtecke, sondern deren Flächeninhalt ist proportional zur Häufigkeit.

Die Höhe der Rechtecke ist zu ermitteln aus: Flächeninhalt = Grundseite • Höhe

→ Höhe = Flächeninhalt/Grundseite

h_1: 0,15/3 = 0,05 · 20 = 1

h_2: 0,2 /2 = 0,10 · 20 = 2

h_3: 0,4 /2 = 0,20 · 20 = 4

h_4: 0,25/5 = 0,05 · 20 = 1

(Falls die Zahlen zu klein sind, kann man sie mit einem festen Wert λ multiplizieren. In diesem Fall: λ = 20)

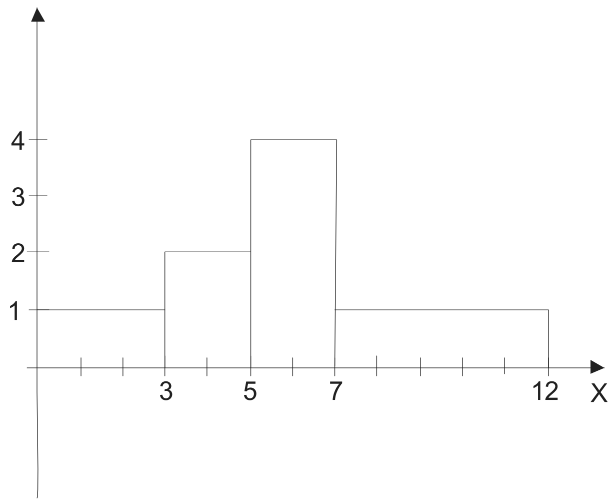

Abbildung 1: Histogramm

Aufgabe:

3) 120 erwachsene Personen wurden nach ihren monatlichen Ausgaben für Kino- bzw. Theaterbesuche befragt. Die Umfrageergebnisse wurden klassiert in folgender Tabelle zusammengestellt:

monatliche Ausgaben von: - bis unter	Anzahl
0 - 10	5
10 - 30	20
30 - 50	75
50 - 100	20

Zeichnen Sie ein Histogramm für das Merkmal „monatliche Ausgaben für Kino bzw. Theater“!

1.1.5 Lageparameter

Folgende Grundbegriffe sollte man kennen:

Modus (Modalwert): Häufigster Wert. Der Modus ist ein geeigneter Parameter für zahlenmäßig nicht erfassbare Merkmale (Farben, Einschaltquoten für Fernsehsendungen usw.)

Median (Zentralwert): Der mittlere Wert einer Datenreihe. Voraussetzung ist natürlich eine geordnete Reihe von Beobachtungswerten. Bei einer geraden Anzahl von Werten ist der Median das arithmetische Mittel der beiden mittleren Werte.

Arithmetisches Mittel:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1a}$$

Mit arithmetischem Mittel ist in aller Regel das gewogene arithmetische Mittel gemeint. Zum Unterschied:

Beispiel 3:

Eine Schülerin hat im ersten Schulhalbjahr folgende Noten im Fach Englisch erhalten: 2, 2, 3, 3, 3 und 5. Beim arithmetischen Mittel werden nach Formel 1a die einzelnen Noten mit deren Häufigkeit gewichtet (gewogen), also (2+2+3+3+3+5):6.

Beim nicht gewogenen Mittel würden dagegen nur die einzelnen Ausprägungen berücksichtigt, also (2+3+5):3.

Für r disjunkte statistische Massen, mit den jeweiligen arithmetischen Mitteln $\bar{x}_1;\bar{x}_2;...\bar{x}_r$ berechnet sich das Gesamtmittel:

$$\bar{x}_{Ges} = \frac{1}{r}\sum_{j=1}^{r} n_j \bar{x}_j \tag{1b}$$

– Geometrisches Mittel:

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot ..x_n} \tag{2}$$

Überall dort, wo man es mit Wachstumsraten oder prozentualer Zu- bzw. Abnahme zu tun hat, wird das geometrische Mittel berechnet.

Beispiel 4:

Sie zahlen am 1.1. eines Jahres bei einer Bank 1000 € ein. In den ersten beiden Jahren bekommen Sie 4% Zinsen, in den folgenden 3 Jahren 5,5 % und im sechsten Jahr 6% Zinsen.

a) Auf wie viel Euro ist Ihr Kapital nach 6 Jahren angewachsen?

b) Bei welcher Durchschnittsverzinsung hätten Sie das gleiche Endkapital?

Lösung:

a) $K_6 = 1000 \cdot 1.04^2 \cdot 1{,}055^3 \cdot 1{,}06$

$K_6 = 1346{,}26$

b) $1000 \cdot q^6 = 1346{,}26$

$$q^6 = 1{,}34626$$
$$q = \sqrt[6]{1{,}34626} = \sqrt[6]{1{,}04^2 \cdot 1{,}055^3 \cdot 1{,}06}$$
$$q = 1{,}0508$$

Damit ist q $\approx$ 1,051 die durchschnittliche Wachstumsrate ($\bar{x}_G$) des Kapitals für die 6 Jahre.

– Harmonisches Mittel:

$$\bar{x}_H = \frac{\sum_{j=1}^{m} h_j}{\sum_{j=1}^{m} \frac{h_j}{a_j}} = \frac{n}{\sum_{j=1}^{m} \frac{h_j}{a_j}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} \qquad (3)$$

Die Berechnung dieses Mittelwertes ist eigentlich auch ganz einfach, aber immer, wenn man etwas formalisieren oder allgemeingültig machen muss, wird es kompliziert. Das harmonische Mittel ist immer dann geeignet, wenn die Größe, um die es geht, indirekt proportional zu einer anderen Größe ist. Typisches Anwendungsgebiet für das harmonische Mittel ist die Durchschnittsgeschwindigkeit.

Die Geschwindigkeit ist umgekehrt proportional zur Zeit.

Beispiel 5:

Ein Pkw legt vier gleich lange Teilstrecken (40km) in folgenden Geschwindigkeiten zurück:

Teilstrecke sj	40	40	40	40
Geschwindigkeit vj	40	50	80	100

Die Geschwindigkeit v ist hier das Beobachtungsmerkmal, steht also für x, v_j ist die Ausprägung und steht in diesem Fall für a_j.

s_j steht hier für die Anzahl der Kilometer, also für h_j, die mit einer bestimmten Geschwindigkeit gefahren werden.

n ist die Gesamthäufigkeit der gefahrenen Kilometer, also 160.

Es gilt für die Geschwindigkeit: $v = \frac{s}{t}$ und für die Durchschnittsgeschwindigkeit:

$\bar{v} = \frac{\sum_{j=1}^{4} s_j}{\sum_{j=1}^{4} t_j}$ (Summe aller Strecken):(Summe aller Zeiten). Da die Zeiten nicht explizit angegeben sind, müssen sie ersetzt werden: $t = \frac{s}{v}$

$$\bar{v} = \frac{\sum_{j=1}^{4} s_j}{\sum_{j=1}^{4} \frac{s_j}{v_i}} = \frac{40+40+40+40}{\frac{40}{40}+\frac{40}{50}+\frac{40}{80}+\frac{40}{100}} = \frac{n}{\sum_{j=1}^{m} \frac{s_j}{v_j}}$$

Aufgaben:

4) Bestimmen Sie für die folgende Urliste Modus, Median und das gewogene arithmetische Mittel:

   3, 4, 2, 5, 5, 1, 3, 6, 5, 1, 5, 1

5) Jemand fuhr mit seinem Sportwagen die Strecke Stuttgart-München (einfache Entfernung = 220 km) und zurück. Auf den ersten 40 km erzielte er im Großraum Stuttgart nur eine Geschwindigkeit von 120 km/h. Dann konnte er 5 Minuten lang 240 km/h fahren. Anschließend hatte er bis München eine Durchschnittsgeschwindigkeit von 160 km/h. Auf der Rückfahrt fuhr er konstant 110 km/h.

   Berechnen Sie seine Durchschnittsgeschwindigkeit!

6) Die Schüler eines Gymnasiums mit 800 Schülern wurden in drei Gruppen eingeteilt und an einem normalen Schultag wurden die Gewichte ihrer Schultaschen ermittelt. Es ergaben sich für die drei Gruppen jeweils die folgenden Durchschnittsgewichte:

   Gruppe I (Klasse 5 – 7, 305 Schüler): 9,8 kg

   Gruppe II (Klasse 8 – 10, 290 Schüler): 7,3 kg

   Gruppe III (Klasse 11 – 12, 205 Schüler): 4,2 kg

   Berechnen Sie das durchschnittliche Gewicht der Schultaschen an dieser Schule! Berechnen Sie den durchschnittlichen Bruttomonatslohn!

7) Die jährlichen Ausgaben für Bildung und Erziehung der letzten 6 Jahre in Millionen € seien in der folgenden Tabelle zusammengefasst (fiktive Zahlen):

	Deutschland	Italien	Frankreich
2002	124,00	74,07	203,80
2003	132,6	80,53	235,10
2004	125,70	86,77	244,5
2005	126,40	89,85	240,00
2006	127,20	91,44	230,70
2007	127,20	95,42	206,70

   Bestimmen Sie die durchschnittlichen Wachstumsraten!

8) Für die Bruttomonatslöhne der Arbeiter eines Unternehmens wurde folgende gruppierte Häufigkeitstabelle aufgestellt:

Löhne	Zahl der Arbeiter
0 – 1800	350
1800 – 2200	960
2200 – 2400	1460
2400 – 2600	1890
2600 – 2800	1560
2800 – 3200	1760

Berechnen Sie den durchschnittlichen Bruttomonatslohn!

9) Die folgende ungeordnete Urliste enthält das Alter von 50 Mitarbeitern eines mittelständischen Unternehmens:

40 20 22 15 18 51 37 42 31 58 33 39 49 22 23 62 42 53 43 44 19 49 40 36 37 38 22 24 32 29 41 40 40 38 27 51 52 54 28 22 64 19 50 40 18 68 51 41 48 57

Bestimmen Sie den Modalwert, den Median sowie das arithmetische Mittel!

10) An 100 Tagen registriert jemand morgens und abends seine jeweiligen Wartezeiten auf den Bus:

$u \leq t \prec o$	abs. Häufigkeit
0 – 1	108
1 – 2	55
2 – 3	17
3 – 4	9
4 – 5	7
5 – 6	4

Berechnen Sie das gewogene arithmetische Mittel der Wartezeiten!

1.2 Streuungsparameter und Konzentrationsmaße

1.2.1 Streuungsparameter

Folgende Grundbegriffe sollte man kennen:

- **Spannweite** ($SP = x_{max} - x_{min}$) (4)
- **mittlere quadratische Abweichung (s^2):**
  Sie gibt Auskunft über die Abweichung der einzelnen Werte vom Mittelwert (großes s^2 = große Streuung, kleiner Wert: die einzelnen Beobachtungswerte sind relativ dicht um den Mittelwert konzentriert). Alle 6 angegebenen Formeln führen bei richtigem Rechnen zum gleichen Ergebnis. Erfahrung, Datenstruktur oder Belieben des Statisti-kers sind entscheidend, welche der Formeln er verwendet.

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (5a)$$

$$s^2 = \frac{1}{n}\sum_{j=1}^{k}a_j^2 \cdot h(a_j) - \bar{x}^2 \quad (5d)$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}^2 \quad (5b)$$

$$s^2 = \sum_{j=1}^{k}(a_j - \bar{x})^2 \cdot f(a_j) \quad (5e)$$

$$s^2 = \frac{1}{n}\sum_{j=1}^{k}(a_j - \bar{x})^2 \cdot h(a_j) \quad (5c)$$

$$s^2 = \sum_{j=1}^{k}a_j^2 \cdot f(a_j) - \bar{x}^2 \quad (5f)$$

- **Standardabweichung:**

$$s = \sqrt{s^2} \quad (6)$$

– **Variationskoeffizient:**

$$V = \frac{s}{\bar{x}} \tag{7}$$

Beispiel 6:

Von 20 Haushalten wurde ermittelt, wie viele Computer im jeweiligen Haushalt vorhanden sind. In ungeordneter Reihe erhielt man folgendes Ergebnis:

1	0	2	1	1	1	2	0	3	0	2	2	3	2	1	2	0	2	1	3
x_1	x_2	.	.	.	.	.	.	x_i	.	.	.	.	.	.	.	.	.	.	x_{20}

Merkmalsausprägungen: $a_j = 0, 1, 2, 3$

Spannweite: $3 - 0 = 3$

a_j	0	1	2	3
$h(a_j)$	4	6	7	3
$H(x_i)$	4	10	17	20
$F(a_j)$	0,2	0,3	0,35	0,15
$F(x_i)$	0,2	0,5	0,85	1

Arithmetisches Mittel: $\bar{x} = \frac{1}{20} \cdot (6 \cdot 1 + 7 \cdot 2 + 3 \cdot 3) = 1{,}45$

Mittlere quadratische Abweichung:

5a: $s^2 = \frac{1}{20} \cdot [(0-1{,}45)^2 + .. + (0-1{,}45)^2 + (1-1{,}45)^2 + .. + (3-1{,}45)^2] = 0{,}9475$

5b: $s^2 = \frac{1}{20} \cdot (0^2 + 0^2 + ..1^2 + .. + 3^2 + 3^2 + 3^2) - 1{,}45^2 = 0{,}9475$

5c: $s^2 = \frac{1}{20} \cdot [(0-1{,}45)^2 \cdot 4 + (1-1{,}45)^2 \cdot 6 + (2-1{,}45)^2 \cdot 7 + (3-1{,}45)^2 \cdot 3] = 0{,}9475$

5d: $s^2 = \frac{1}{20} \cdot (0^2 \cdot 4 + 1^2 \cdot 6 + 2^2 \cdot 7 + 3^2 \cdot 3) - 1{,}45^2 = 0{,}9475$

5e: $s^2 = (0-1{,}45)^2 \cdot 0{,}2 + (1-1{,}45)^2 \cdot 0{,}3 + ... + (3-1{,}45)^2 \cdot 0{,}15 = 0{,}9475$

5f: $s^2 = (0^2 \cdot 0{,}2 + 1^2 \cdot 0{,}3 + 2^2 \cdot 0{,}35 + 3^2 \cdot 0{,}15) - 1{,}45^2 = 0{,}9475$

Standardabweichung: $s = \sqrt{0{,}9475} \approx 0{,}9734$

Variationskoeffizient: $V = \frac{0{,}9734}{1{,}45} \approx 0{,}6713$