Prof. Dr. A. Steland
N. Mause

RWTH Aachen, SS2023

Ausgabe: 13. Juni 2023 — Besprechung: 19. Juni 2023

Einführung in die angewandte Stochastik

Übungsblatt 8

Klassifikation von Merkmalen

Merkmal: Größe, die an einer statistischen Einheit erhoben wird, z.B. statistische Einheit „Student" mit Merkmal Alter, Geschlecht, Studiengang, ...

Merkmale haben **Ausprägungen**, z.B. Merkmal „Alter" mit Ausprägung 18,20, 25,...

Mathematisch: Merkmal ist eine Abbildung

$$X : G \mapsto M,$$

wobei G die Menge aller statistischen Einheiten und M die Menge aller Ausprägungen ist.

Klassifikation (nicht immer eindeutig!)

Unterscheidungsaspekt 1

- quantitativ: Ausprägungen sind Zahlen, z.B. 15,20,25,...
- qualitativ: Ausprägungen sind Texte, z.B. gut, schlecht, ja, nein,...

Unterscheidungsaspekt 2 (nur bei quantitativen Merkmalen)

- diskret: es gibt nur endlich viele Ausprägungen, d.h. $M \subset \mathbb{N}$, z.B. Anzahl von Äpfeln
- stetig: unendlich viele Ausprägungen, d.h. $M \subset \mathbb{R}$ bzw. $M = [a, b]$, z.B. Aktienkurse
- quasi-stetig: diskrete Merkmale, die aber eigentlich auch stetig sind

Unterscheidungsaspekt 3 (bei quantitativen und qualitativen Merkmalen)
Unterscheidung bezüglich des Skalenniveaus

- Nominalskala: Ausprägungen stehen in **keinem** Verhältnis zueinander, z.B. Äpfel, Bananen,...,
- Ordinalskala: Ausprägungen können direkt verglichen werden, z.B. gut, schlecht, ...
- metrische Skala
- Kardinalskala: Ausprägungen werden mit einem Mess-Stab gemessen und sind Vielfache einer Grundeinheit
  - Intervallskala: Nullpunkt willkürlich, d.h. Quotienten können nicht interpretiert werden, z.B. Temperatur
  - Verhältnisskala: Nullpunkt eindeutig bestimmt, d.h. Quotienten können sinnvoll interpretiert werden, z.B. Anzahlen, Längen, Gewichte,...

Aufgabe 31

Einer Firma können untere anderem die folgenden Merkmale zugeordnet werden:

Mitarbeiterzahl; Familienstand des Chefs; Produkte, die hergestellt werden; Qualität der Produkte; Umsatz; Rechtsform; Betriebsklima; Zufriedenheit der Kunden; Sitz der Zentrale; Wert der Immobilien; Gründungsjahr.

Bestimmen Sie jeweils den zugehörigen Merkmalstyp, und geben Sie zu den einzelnen Merkmalen jeweils eine mögliche Menge von Ausprägungen an.

Lösung:

Merkmal	Merkmalstyp	mögliche Menge von Ausprägungen
Mitarbeiterzahl	quantitativ diskret nominal	10, 15, 20, ... $\mathbb{N}$
Familienstand Chef	qualitativ nominal	ledig, verheiratet
Produkte	qualitativ nominal	Laptops, Drucker, Regal...
Qualität der Produkte	qualitativ ordinal	gut, schlecht
Umsatz	quantitativ, diskret ordinal (oder quasi-stetig)	100.000, 100.001, ... [0, 100.000], [900.000, 1000000]
Rechtsform	qualitativ, nominal	GmbH, OHG
Betriebsklima	qualitativ, ordinal	gut, schlecht

Zufriedenheit der Kunden	qualitativ ordinal	sehr zufrieden, nicht zufrieden
Sitz der Zentrale	qualitativ nominal	Aachen, Köln ...
Wert der Immobilien	quantitativ ordinal, diskret	10000, 100.000
Gründungsjahr	quantitativ diskret nominal	1980, 1981...

Aufgabe 32

Ordnen Sie den folgenden Merkmalen sinnvolle Skalen zu (Nominal-, Ordinal-, Intervall- und Verhältnisskala).

(a) Alter metrisches Merkmal → Verhältnisskala

(b) Familienstand Nominalskala, siehe A31

(c) Geschlecht Nominalskala

(d) Einkommen Ordinalskala oder Verhältnisskala

(e) Schulbildung Ordinalskala

(f) Beruf Nominalskala

(g) Schulnoten Ordinalskala oder Intervallskala

(h) Körpergewicht Verhältnisskala

(i) Intelligenzquotient Ordinalskala

Lösung:

Aufgabe 33 Stichprobenumfang $n = 50$

Bei einer Umfrage in einem Verein werden 50 Teilnehmer zu ihrem Familienstand befragt. Die Antwortmöglichkeiten sind *ledig(l)*, *verheiratet(vh)*, *geschieden(g)* sowie *verwitwet(vw)*. Es werden folgende Antworten gegeben:

vh, vw, l, vw, g, vh, vh, l, g, vw, vw, l, vh, vw, l, vw, l, vh, vh, vh, vh, l, g, vw, vh, vw, vh, vw, vh, vh, l, g, g, vh, l g, vw, vh, vw, vh, vw, vh, vw, g, vh, g, vh, vw, vh, vh

(a) Bestimmen Sie die absoluten und relativen Häufigkeiten der einzelnen Ausprägungen des Merkmals Familienstand.

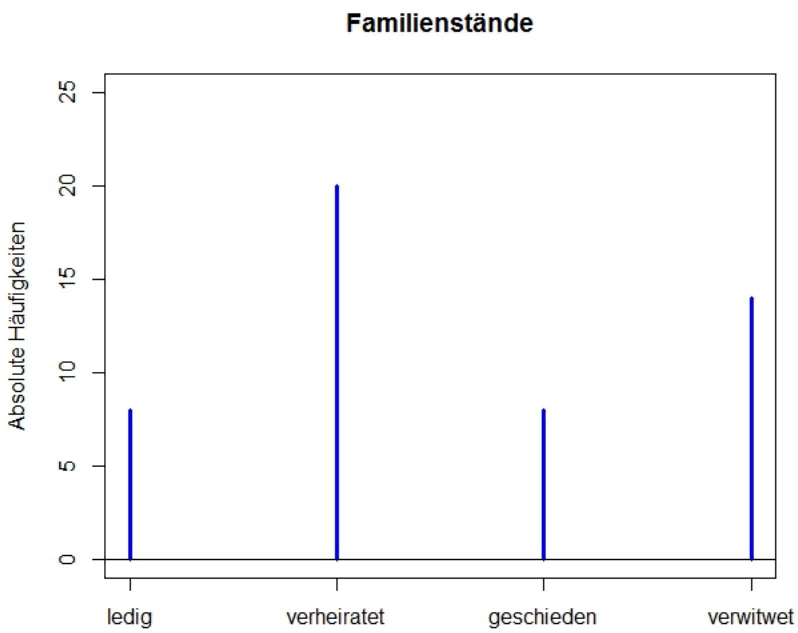
(b) Stellen Sie die absoluten Häufigkeiten in einem Stabdiagramm dar.

(c) Stellen Sie die relativen Häufigkeiten in einem Kreisdiagramm dar.

Lösung: nominales Merkmal Familienstand

Ausprägung	absolute Häufigkeit	relative Häufigkeit
ledig (l)	8	$\frac{\text{abs. H.}}{n} = \frac{8}{50} = 0{,}16$
verheiratet (vh)	20	$20/50 = 0{,}4$
geschieden (g)	8	$8/50 = 0{,}16$
verwitwet (vw)	14	$14/50 = 0{,}28$

b)



Ausprägungen x

c)

Berechnung der Winkel im Kreisdiagramm:

ledig: $0{,}16 \cdot 360° = 57{,}6°$

verheiratet: $0{,}4 \cdot 360° = 144°$

geschieden: $0{,}16 \cdot 360° = 57{,}6°$

verwitwet: $0{,}28 \cdot 360° = 100{,}8°$

Aufgabe 34 $n = 200$ Stichprobenumfang

Mit dem Ziel, Informationen über den Zeitaufwand im Bachelorstudium Informatik zu erhalten, wurden 200 Studierende der Informatik nach der Wochenstundenzahl der von ihnen im letzten Semester regelmäßig besuchten Veranstaltungen befragt. Die Antworten sind nach Klassen in folgender Tabelle zusammengefasst:

Semesterwochenstunden	[0,5]	(5,12]	(12,16]	(16,20]	(20,25]	(25,32]
absolute Häufigkeit	9	16	28	72	47	28

Stellen Sie die Daten in einem Histogramm graphisch dar und bestimmen Sie die zugehörige Häufigkeitsdichte.

Lösung:

j	Klasse K_j	abs. Häufigkeit $n(K_j)$	relative Häufigkeit $f(K_j) = \frac{n(K_j)}{n}$	Klassenbreite b_j	Höhe $l_j = \frac{f(K_j)}{b_j}$
1	$[0,5]$	9	$9/200 = 0{,}045$	$5-0=5$	$\frac{0{,}045}{5} = 0{,}009$
2	$(5,12]$	16	$16/200 = 0{,}08$	$12-5=7$	$\frac{0{,}08}{7} = 0{,}0114$
3	$(12,16]$	28	$28/200 = 0{,}14$	$16-12=4$	$\frac{0{,}14}{4} = 0{,}035$
4	$(16,20]$	72	$\frac{72}{200} = 0{,}36$	$20-16=4$	$\frac{0{,}36}{4} = 0{,}09$
5	$(20,25]$	47	$\frac{47}{200} = 0{,}235$	$25-20=5$	$\frac{0{,}235}{5} = 0{,}047$
6	$(25,32]$	28	$\frac{28}{200} = 0{,}14$	$32-25=7$	$\frac{0{,}14}{7} = 0{,}02$

Flächeninhalt Rechteck: $l_j \cdot b_j = f(K_j)$ = relative Häufigkeit oberer Rand des Histogramms definiert eine Treppenfunktion

$\rightarrow$ Häufigkeitsdichte: Klassen $[g_1, g_2], (g_2, g_3], \ldots$

$$f(x) = \begin{cases} 0, & x < g_1 \\ l_1, & x \in [g_1, g_2] \\ l_j, & x \in (g_j, g_{j+1}], \; j = 2, \ldots, 6 \\ 0, & x > g_7 \end{cases}$$

$$= \begin{cases} 0, & x < 0 \\ 0{,}009, & x \in [0, 5], \\ 0{,}0114, & x \in (5, 12], \\ 0{,}035, & x \in (12, 16], \\ 0{,}09, & x \in (16, 20], \\ 0{,}047, & x \in (20, 25], \\ 0{,}02, & x \in (25, 32], \\ 0, & x > 32 \end{cases}$$

Aufgabe 35

Bei einem Quiz haben 15 Leute 8 Fragen beantwortet. Für die Personen ergaben sich jeweils folgende Anzahlen richtiger Antworten:

$$x_1 = 3, \quad x_2 = 5, \quad x_3 = 5, \quad x_4 = 7, \quad x_5 = 4, \quad x_6 = 8, \quad x_7 = 2, \quad x_8 = 4,$$
$$x_9 = 4, \quad x_{10} = 6, \quad x_{11} = 2, \quad x_{12} = 7, \quad x_{13} = 6, \quad x_{14} = 4, \quad x_{15} = 6.$$

(a) Geben Sie die Ordnungsstatistik an.

(b) Geben Sie die minimale und die maximale Anzahl richtiger Antworten an.

(c) Bestimmen Sie das arithmetische Mittel, die Stichprobenvarianz und die Standardabweichung des Datensatzes.

(d) Bestimmen Sie den Median.

(e) Berechnen Sie das untere und obere Quartil sowie den Quartilsabstand.

(f) Berechnen Sie die zugehörigen p-Quantile für $p \in \{0.2, 0.5, 0.8\}$.

(g) Zeichnen Sie einen Boxplot zur graphischen Darstellung relevanter Kennzahlen.

Falls ein Quantil nicht eindeutig bestimmt ist wählen Sie die Intervallmitte der Kandidaten.

Lösung:

a) Ordnungsstatistik (Sortiere Daten aufsteigend)

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
x_i	3	5	5	7	4	8	2	4	4	6	2	7	6	4	6
$x_{(i)}$	2	2	3	4	4	4	4	5	5	6	6	6	7	7	8

b) $x_{min} = x_{(1)} = 2$

$x_{max} = x_{(15)} = 8$

c) arithm. Mittel

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \overset{n=15}{=} \frac{1}{15} \sum_{i=1}^{15} x_i = \ldots = \frac{1}{15} \cdot 73 \approx 4{,}87$$

Mit dem Verschiebungssatz gilt nun

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\bar{x})^2$$

$$\Rightarrow s_x^2 = \frac{1}{15}\cdot \underbrace{\sum_{i=1}^{15}(3^2+5^2+\ldots+4^2+6^2)}_{=401} - \left(\frac{73}{15}\right)^2 = \frac{686}{225} \approx 3{,}05$$

$\Rightarrow$ Standardabweichung

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{686}{225}} \approx 1{,}746$$

d) Da $n=15$ ungerade ist, ergibt sich der Median

$$x_{med} = x_{(\frac{n+1}{2})} \overset{n=15}{=} x_{(8)} = 5$$

[Wenn n gerade ist, dann ist jede Zahl aus dem Intervall $[x_{(\frac{n}{2})}, x_{(\frac{n}{2}+1)}]$ ein Median.
Übliche Konvention: Nehme die Intervallmitte: $x_{med} = \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)})$]

e) untere Quartil: 0,25-Quartil
obere Quartil: 0,75-Quartil

$15\cdot 0{,}25 = 3{,}75 \notin \mathbb{N}$ und $15\cdot 0{,}75 = 11{,}25 \notin \mathbb{N}$

$$\Rightarrow Q_1 = \tilde{x}_{0,25} = x_{(\lfloor 15\cdot 0{,}25\rfloor + 1)} = x_{(4)} = 4$$

untere Gaußklammer
$\rightarrow$ Abrundung

$$Q_3 = \tilde{x}_{0,75} = x_{(\lfloor 15\cdot 0{,}75\rfloor + 1)} = x_{(12)} = 6$$

Quartilabstand:

$$IQR = Q_3 - Q_1 = 6 - 4 = 2$$

f) Für $p \in \{0.2, 0.5, 0.8\}$ erhält man

$15\cdot 0{,}2 = 3 \in \mathbb{N}$, $15\cdot 0{,}5 = 7{,}5 \notin \mathbb{N}$, $15\cdot 0.8 = 12 \in \mathbb{N}$.

Für $p = 0{,}5$ folgt daher

$$\tilde{x}_{0.5} = x_{(\lfloor 15\cdot 0{,}5\rfloor + 1)} = x_{(8)} = 5$$

Dies entspricht der Median, d.h. $x_{med} = \tilde{x}_{0,5}$

Für $p \in \{0.2, 0.8\}$ ist jede Zahl des Intervalls

$$[x_{(np)}, x_{(np+1)}]$$

ein p-Quantil. Konvention: Nehme die Intervallmitte, d.h.

$$\tilde{x}_p = \frac{1}{2}\left(x_{(np)} + x_{(np+1)}\right).$$

Somit ist

$$\tilde{x}_{0,2} = \frac{1}{2}\left(x_{(3)} + x_{(4)}\right) = \frac{1}{2}(3+4) = 3{,}5$$

$$\tilde{x}_{0,8} = \frac{1}{2}\left(x_{(12)} + x_{(13)}\right) = \frac{1}{2}(6+7) = 6{,}5.$$

g) Boxplot (5-Punkt Zusammenfassung)