

2 Univariate deskriptive Statistik

zugehörige Seiten in Fahrmeir et al. (2007): Kap. 2.1

Aufgabe 7

Der Besitzer des Kinos *Cinemanía* macht sich Gedanken über die Wirtschaftlichkeit seines Hauses. An 100 Tagen zählt er daher die Anzahl der Besucher.

Besucherzahl (a_j)	41	42	43	44	45	46	47	48	49	50	51
Anzahl der Tage (h_j)	1	9	13	13	20	15	10	7	5	4	3

- (a) Wie heißt das untersuchte Merkmal und wie ist es skaliert?
- (b) Berechnen Sie die relativen und kumulierten relativen Häufigkeiten.
- (c) Stellen Sie die relativen Häufigkeiten in einem Stabdiagramm dar.
- (d) Zeichnen Sie die empirische Verteilungsfunktion.
- (e) Zur Existenzerhaltung reicht es aus, wenn an 90% der Tage mindestens 48 Besucher kommen. Hat das Kino eine Überlebenschance?
- (f) Wie groß ist der Anteil der Tage, an denen der Kinobesitzer weniger als 270 Euro einnimmt? Dabei wird ein Einheitspreis von 6 Euro erhoben.
- (g) An wieviel Prozent der Tage kommen maximal 50, mindestens aber 45 Besucher?

Aufgabe 8

In einer Umfrage wurden die Passagiere der Fluggesellschaft *Austrian Airlines* gebeten, die Serviceleistungen während des Fluges auf einer Punkteskala von 0 bis 9 zu bewerten. Nach Klassenbildung ergaben sich die folgenden absoluten Häufigkeiten (fiktive Daten):

Klasse	Wertung	h_j
schlechte Serviceleistungen	0,1	229
mittelmäßige Serviceleistungen	2,3,4	3491
gute Serviceleistungen	5,6	9933
sehr gute Serviceleistungen	7,8,9	6347

- (a) Stellen Sie die Serviceleistungen in einem Kreisdiagramm dar.
- (b) Zeichnen Sie das zugehörige Histogramm.

Aufgabe 9

25 Mitglieder eines Scottie-Fan-Clubs wurden zur Anzahl der von ihnen gesehenen Folgen der Serie *Star Trek* befragt. Die Befragten machten folgende Angaben:

180 190 202 175 199 200 205 185 190 209 165 203 177
205 175 207 200 195 172 200 184 180 193 195 201

- (a) Erstellen Sie ein Stamm-Blatt-Diagramm, in dem alle Ziffern der gesammelten Daten berücksichtigt werden. Leiten Sie daraus mögliche Interpretationen ab.

- (b) Überführen Sie obige Darstellung in ein Stamm-Blatt-Diagramm, in dem nur die jeweils führenden beiden Ziffern der Daten Berücksichtigung finden. Vergleichen Sie das Ergebnis hinsichtlich seiner Aussagefähigkeit mit dem Diagramm aus (a).

Aufgabe 10

25 Studenten der Universität X in X-hausen wurden im Juni 2004 nach dem Studiengang, der Zahl der Geschwister und nach dem Einkommen befragt. Das Ergebnis war:

i	Name	Studiengang	Anzahl der Geschwister	Einkommen (in Euro)
1	Martin A.	Mathematik	0	472
2	Ute A.	Informatik	1	403
3	Wilhelm A.	Statistik	0	698
4	Kurt B.	Statistik	1	349
5	Sven B.	Physik	1	380
6	Mike D.	Physik	2	327
7	Klaus D.	Informatik	2	323
8	Theo E.	Mathematik	1	416
9	Jean F.	Physik	1	398
10	Elvira G.	Statistik	0	543
11	Karl H.	Statistik	0	629
12	Andreas K.	Mathematik	1	350
13	Thomas K.	Statistik	0	435
14	Chris L.	Informatik	3	328
15	Uwe L.	Physik	2	327
16	Axel M.	Statistik	0	435
17	Maria M.	Statistik	1	349
18	Ruth M.	Informatik	0	315
19	Bärbel N.	Statistik	1	349
20	Armin R.	Statistik	2	349
21	Christa R.	Mathematik	1	337
22	Bernd S.	Statistik	1	736
23	Claudia S.	Informatik	3	406
24	Edmund T.	Mathematik	0	337
25	Claudia W.	Physik	1	327

- (a) Was ist in dieser Befragung die statistische Grundgesamtheit, was sind die einzelnen statistischen Einheiten? Durch welche Identifikationsmerkmale ist die Grundgesamtheit festgelegt?
- (b) Wie ist das Merkmal „Studiengang“ skaliert? Ermitteln Sie aus der Urliste die absoluten und relativen Häufigkeiten. Stellen Sie die relativen Häufigkeiten in einem Säulendiagramm dar.
- (c) Wie ist das Merkmal „Anzahl der Geschwister“ skaliert? Ermitteln Sie die absoluten und relativen Häufigkeiten sowie die kumulierten Häufigkeiten. Stellen Sie die relativen Häufigkeiten in einem Kreisdiagramm dar.
- (d) Beantworten Sie folgende Fragen:
- Wieviele Studenten haben höchstens zwei Geschwister?
 - Wieviel Prozent der Studenten haben mindestens zwei Geschwister?
 - Wieviel Prozent der Studenten haben ein oder zwei Geschwister?
- (e) Wie ist das Merkmal „Einkommen“ skaliert? Stellen Sie die Häufigkeitsverteilung in einem Stamm-Blatt-Diagramm dar. Interpretieren Sie das Diagramm.
- (f) Ermitteln Sie die absoluten und relativen Häufigkeiten für das Merkmal „Einkommen“ unter Verwendung der Einkommensklassen $[300, 330)$, $[330, 350)$, $[350, 450)$, $[450, 600)$, $[600, 750)$. Berechnen

Sie die absoluten und relativen Häufigkeiten für die gruppierten Daten. Stellen Sie die Häufigkeitsverteilung der gruppierten Daten graphisch in einem Histogramm dar. Vergleichen Sie das Histogramm mit dem Stamm-Blatt-Diagramm aus Aufgabe (e).

(g) Berechnen Sie ausgehend von der Gruppierung unter (f):

- Welcher Anteil der Studenten hat ein Einkommen von mindestens 375 Euro und höchstens 700 Euro?
- Welcher Anteil der Studenten hat ein Einkommen von mehr als 400 Euro?
- Welches maximale Einkommen trat unter den 50 Prozent einkommenschwächsten Studenten auf?

(Hinweis: Verwenden Sie lineare Interpolation.)

Aufgabe 11

Für die Urliste x_1, \dots, x_n soll ein Histogramm betrachtet werden mit k Klassen und Klassengrenzen $c_0 < c_1 < \dots < c_k$, $c_0 \leq \min_i x_i$, $\max_i x_i \leq c_k$. Für die abzutragende Höhe \tilde{f}_j der j -ten Klasse ($j = 1, \dots, k$) mit Klassenbreite $d_j = d$ für alle j gilt

$$\tilde{f}_j(x) = \begin{cases} f_j/d, & x \in [c_{j-1}, c_j), \\ 0, & \text{sonst.} \end{cases}$$

(a) Zeigen Sie, dass unter Berücksichtigung aller k Klassen für die abzutragende Höhe an einer beliebigen Stelle $x \in \mathbb{R}$ gilt

$$\tilde{f}(x) = \frac{1}{nd} \sum_{i=1}^n \sum_{j=1}^k 1_{\{x_i \in [c_{j-1}, c_j)\}} 1_{\{x \in [c_{j-1}, c_j)\}}, \text{ mit } 1_{\{x \in A\}} = \begin{cases} 1, & x \in A, \\ 0, & \text{sonst.} \end{cases}$$

(b) Zeigen Sie, dass

$$\int_{-\infty}^{\infty} \tilde{f}(x) dx = 1.$$

Aufgabe 12* (16 Punkte)

Um die Berufsaussichten von Absolvent(inn)en des Diplomstudiengangs Soziologie einschätzen zu können, wurde am Institut für Soziologie der LMU ein spezieller Fragebogen entwickelt, der insgesamt 82 Fragen umfasst. Der Fragebogen deckt zahlreiche inhaltliche Aspekte ab, wie etwa Studienverlauf, Studienschwerpunkt, mögliche Zusatzqualifikationen, aber auch Angaben zur Person. Der in den folgenden beiden Tabellen abgedruckte Teildatensatz mit 36 Absolvent(inn)en und 5 Variablen soll nun für eine erste Analyse mit Hilfe von graphischen Verfahren dargestellt werden.

G	Geschlecht 1 = weiblich 2 = männlich
S	Studiendauer in Semestern
E	Engagement im Studium mit 5 Kategorien: 1 = sehr engagiert ⋮ 5 = gar nicht engagiert
D	Ausrichtung der Diplomarbeit mit 4 Kategorien: 1 = empirisch - Primärerhebung 2 = empirisch - Sekundärerhebung 3 = empirisch - qualitativ 4 = Literaturarbeit
N	Note in der Diplomprüfung

Person i	G	S	E	D	N	Person i	G	S	E	D	N
1	1	12	1	3	2	19	2	12	2	2	2
2	1	13	3	4	2	20	1	15	2	3	3
3	1	12	5	4	3	21	1	13	3	4	2
4	1	12	2	3	3	22	2	13	4	3	3
5	1	9	3	4	2	23	1	15	1	4	2
6	1	12	2	1	1	24	1	13	3	3	2
7	2	14	5	3	5	25	2	15	4	4	3
8	2	10	1	4	2	26	1	12	2	4	2
9	1	18	3	3	1	27	1	14	1	3	2
10	2	10	3	4	3	28	1	10	2	4	2
11	1	13	4	4	3	29	1	12	3	3	2
12	1	15	4	3	2	30	1	17	2	3	2
13	2	13	2	2	2	31	1	11	1	4	2
14	1	16	3	3	2	32	1	14	3	2	3
15	1	14	3	4	2	33	1	11	2	1	2
16	1	13	2	3	2	34	2	13	2	4	3
17	1	13	2	4	2	35	2	11	3	4	3
18	1	17	1	4	3	36	2	7	1	4	2

- (a) Erstellen Sie eine Häufigkeitstabelle für das Merkmal *Note*, bestehend aus den absoluten, relativen und kumulierten (relativen) Häufigkeiten.
- (b) Erstellen Sie ein Säulen- und ein Kreisdiagramm des Merkmals *Note*.
- (c) Unterteilen Sie die Stichprobe in Absolvent(inn)en mit Prädikatsexamen (Note 1 oder 2) und Absolvent(inn)en ohne Prädikatsexamen (Note 3 oder schlechter). Zeichnen Sie nun für beide Gruppen getrennt ein Säulendiagramm der Studiendauer und interpretieren Sie das Ergebnis.
- (d) Erstellen Sie die empirischen Verteilungsfunktionen der jeweiligen Studiendauer der Absolvent(inn)en mit und ohne Prädikatsexamen. Wie viele Semester benötigten die 25% schnellsten Student(inn)en in jeder Teilstichprobe höchstens? Wie viele Semester brauchten die 25% langsamsten Student(inn)en mindestens?

Aufgabe 13* (freiwillig, bis zu 8 Zusatzpunkte)

Suchen Sie aus Zeitungen und Zeitschriften positive und negative Beispiele für die grafische Darstellung von Zahlenmaterial. Geben Sie diese bitte bis zum **1.12.2009** in den Übungen ab oder werfen diese in den Übungskasten oder senden Sie einen Scan per E-Mail an Andreas Groll. Begründen Sie kurz, warum Sie die gewählten Darstellungen für geeignet bzw. ungeeignet halten. Die Aufgabe wird anschließend in den nächsten Übungen / Tutorien besprochen.