

Statistische Grafiken

Aufgabe 1

Um die Berufsaussichten von Absolvent(inn)en des Masterstudiengangs 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 zu den befragten Personen. Der in den folgenden beiden Tabellen abgedruckte Teildatensatz mit 36 Absolvent(inn)en und fünf Variablen soll für eine erste Analyse mit Hilfe graphischer 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				
M	Ausrichtung der Masterarbeit mit 4 Kategorien: 1 = empirisch - Primärerhebung 2 = empirisch - Sekundärerhebung 3 = empirisch - qualitativ 4 = Literaturarbeit				
N	Note Masterarbeit				

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

- (a) Erstellen Sie eine Häufigkeitstabelle für das Merkmal *Note*, die absolute und relative Häufigkeiten enthält.
- (b) Erstellen Sie ein Säulen- und ein Kreisdiagramm der relativen Häufigkeiten des Merkmals *Note*.

- (c) Unterteilen Sie die Stichprobe nach Geschlecht. Wie groß sind die beiden Stichproben? Erstellen Sie nun für beide Gruppen getrennt Häufigkeitstabellen mit relativen Häufigkeiten für das Merkmal *Note*. Zeichnen Sie jeweils ein Säulendiagramm der relativen Häufigkeiten und interpretieren Sie das Ergebnis.
- (d) Erstellen Sie für die beiden Geschlechter jeweils ein Stapeldiagramm bezogen auf die relativen Häufigkeiten des Merkmals *Note*. Welche Note war jeweils die schlechteste unter den besten 30%. Welche Note war jeweils die beste unter den schlechtesten 10%?

Aufgabe 2

25 Studenten der Universität X in X-hausen wurden im Juni 2018 nach dem Studiengang, der Zahl der Geschwister und nach dem Einkommen befragt. Das Ergebnis der Befragung ist in der folgenden Tabelle dargestellt:

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) Wie ist das Merkmal *Studiengang* skaliert? Ermitteln Sie aus den Daten die absoluten und relativen Häufigkeiten. Stellen Sie die relativen Häufigkeiten in einem Säulendiagramm dar.
- (b) 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.
- (c) Beantworten Sie folgende Fragen:
- Wieviele Studenten haben höchstens zwei Geschwister?
 - Wie groß ist der Anteil der Studenten, die mindestens zwei Geschwister haben?
 - Wie groß ist der Anteil der Studenten, die ein oder zwei Geschwister haben?
- (d) Wie ist das Merkmal *Einkommen* skaliert?

- (e) 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.

Aufgabe 3

Das Softwareprogramm R beinhaltet einen Datensatz zu den Eruptionen des Old Faithful Geisirs im Yellowstone National Park (Wyoming, USA). Der Datensatz heißt `faithful`, enthält zwei Variablen und kann direkt aufgerufen werden.

- (a) Stellen Sie die Variable `waiting` mit einem Histogramm dar. Verwenden Sie dazu einmal die Standardeinstellung von R und einmal die Option `breaks = 20`. Stellen Sie die empirische Dichte der Variable außerdem mit einem Kerndichteschätzer dar. Verwenden Sie einmal die Standardeinstellung von R und einmal die halbe Standardbandbreite.
- (b) Interpretieren Sie die erzeugten Graphiken kurz.

Hinweis: Die benötigten R-Funktionen sind `hist()` und `density()`.