

Statistik I für Human- und Sozialwissenschaften

1. Übung Lösungsvorschlag

Gruppenübung

G 1 Auf einer Touristeninsel in der Karibik wurden in den letzten beiden Juliwochen morgens zur gleichen Zeit die folgenden Lufttemperaturen (in Grad Fahrenheit) gemessen:

78, 82, 81, 82, 80, 83, 77, 81, 79, 79, 83, 78, 78, 79

- (a) Geben Sie den Modalwert an, falls er eindeutig ist.
- (b) Bestimmen Sie den Median. Wie lautet der Median, falls der letzte Messwert weglassen wird?
- (c) Berechnen Sie das Arithmetische Mittel, d.h. die Durchschnittstemperatur der gemessenen Temperaturen.
- (d) Berechnen Sie die Varianz einmal mit der ursprünglichen Varianzformel und einmal mit der Alternativen Berechnungsformel.
- (e) Berechnen Sie die Standardabweichung.

(a) Die Meßwerte 78 und 79 kommen beide mit der höchsten vorhandenen absoluten Häufigkeit vor, nämlich 3. Daher gibt es in dieser Meßreihe keinen eindeutig bestimmbar Modalwert.

(b) Die Meßreihe besteht aus 14 Meßwerten. Ordnet man dieses, dann ergibt sich für die Berechnung des Medians:

$$Md = \frac{79 + 80}{2} = 79.5.$$

Fällt der letzte Meßwert 79 weg, dann besteht die Meßreihe nur noch aus 13 Werten und der Median lautet $Md = 80$.

(c)

$$\bar{x} = \frac{1}{14} \sum_{i=1}^{14} x_i = 80.$$

(d)

$$s^2 = \frac{1}{14} \sum_{i=1}^{14} (x_i - \bar{x})^2 \approx 3.7143. \quad s^2 = \frac{1}{14} \sum_{i=1}^{14} x_i^2 - \bar{x}^2 \approx 3.7143.$$

(e) $s = \sqrt{s^2} \approx 1.9273$.

G 2 Bei einer klinischen Studie zur Ursachenforschung von Herz-/Kreislaufkrankungen wurde das Gewicht (in kg) von 90 Patienten erhoben.

Es liegt folgende Urliste vor:

131,8	106,7	116,4	84,3	118,5	93,4	65,3	113,8	140,3
119,2	129,9	75,7	105,4	123,4	64,9	80,7	124,2	110,9
86,7	112,7	96,7	110,2	135,2	134,7	146,5	144,8	113,4
128,6	142,0	106,0	98,0	148,2	106,2	122,7	70,0	73,9
78,8	103,4	112,9	126,6	119,9	62,6	116,6	84,6	101,0
68,1	95,9	119,7	122,0	127,3	109,3	95,1	103,1	92,4
103,0	90,2	136,1	109,6	99,2	76,1	93,9	81,5	100,4
114,3	125,5	121,0	137,0	107,7	69,0	79,0	111,7	98,8
124,3	84,9	108,1	128,5	87,9	102,4	103,7	131,7	139,4
108,0	109,4	97,8	112,2	75,6	143,1	72,4	120,6	95,2

- (a) Sortieren Sie die Messwerte der Urliste in 9 gleichgroße Intervalle (auch Klassen oder Kategorien genannt): 60,0 - 69,9 ; 70,0 - 79,9 ; ... ; 140,0 - 149,9 mit Hilfe einer Strichliste.
- (b) Erstellen Sie aus obiger Strichliste für die angegebenen Intervalle eine Häufigkeitstabelle mit den Spalten: absolute Häufigkeit - kumulierte absolute Häufigkeit - relative Häufigkeit in % - kumulierte relative Häufigkeit in %.
- (c) Erstellen Sie für das erhobene Merkmal ein Histogramm, ein Stabdiagramm und ein Häufigkeitspolygon jeweils einmal für die absoluten und einmal für die relativen Häufigkeiten.

(a)

60.0 - 69.9	5
70.0 - 79.9	8
80.0 - 89.9	7
90.0 - 99.9	12
100.0 - 109.9	17
110.0 - 119.9	15
120.0 - 129.9	13
130.0 - 139.9	7
140.0 - 149.9	6

(b)

Intervall	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
60.0 - 69.9	5	5	5.6	5.6
70.0 - 79.9	8	13	8.9	14.4
80.0 - 89.9	7	20	7.8	22.2
90.0 - 99.9	12	32	13.3	35.6
100.0 - 109.9	17	49	18.9	54.4
110.0 - 119.9	15	64	16.7	71.1
120.0 - 129.9	13	77	14.4	85.6
130.0 - 139.9	7	84	7.8	93.4
140.0 - 149.9	6	90	6.7	100

G 3 Eine Messreihe x_1, \dots, x_n (mit dem arithmetischen Mittel \bar{x}) werde gemäß der Vorschrift $y_i = a \cdot x_i + b$ ($i = 1, \dots, n$) linear transformiert.

- (a) Zeigen Sie, dass für das arithmetische Mittel \bar{y} der transformierten Werte folgende Formel gilt:

$$\bar{y} = a \cdot \bar{x} + b.$$

- (b) In Brighton an der Südküste Englands wurden während der Weihnachtsferien die folgenden Tagestieftemperaturen x_i ($i = 1, \dots, 10$) in Grad Fahrenheit ($^{\circ}F$) gemessen:

31 27 28 26 30 36 35 34 31 30

Man berechne die mittlere Tagestieftemperatur, d.h. das arithmetische Mittel der einzelnen Tagestieftemperaturen, in $^{\circ}F$ und in Grad Celsius ($^{\circ}C$).

Hinweis: x [$^{\circ}F$] entsprechen $y = \frac{5}{9}(x - 32)$ [$^{\circ}C$].

$$\begin{aligned} \text{(a)} \quad \bar{y} &= \frac{1}{n} \cdot \sum_{i=1}^n y_i = \frac{1}{n} \cdot \sum_{i=1}^n (a \cdot x_i + b) \\ &= \frac{1}{n} \cdot \left(\sum_{i=1}^n a \cdot x_i + \sum_{i=1}^n b \right) = \frac{1}{n} \cdot \left(a \cdot \sum_{i=1}^n x_i + n \cdot b \right) \\ &= \frac{1}{n} \cdot a \cdot \sum_{i=1}^n x_i + \frac{1}{n} \cdot n \cdot b = a \cdot \frac{1}{n} \cdot \sum_{i=1}^n x_i + b \\ &= a \cdot \bar{x} + b. \end{aligned}$$

- (b) Die mittlere Tagestieftemperatur beträgt $\bar{x} = (31 + 27 + \dots + 30)/10 = 308/10 = 30.8^{\circ}F$. Laut Hinweis gilt für die Umrechnung von $^{\circ}F$ nach $^{\circ}C$:

$$y_i = \frac{5}{9}(x_i - 32) = \underbrace{\frac{5}{9}}_a x_i - \underbrace{\frac{5}{9} \cdot 32}_b, \quad i = 1, \dots, 10$$

Mit Hilfe dieser Umrechnungsformel und der Rechenregel aus a) erhält man als mittlere Tagestieftemperatur in $^{\circ}C$:

$$\bar{y} = a \cdot \bar{x} + b = \frac{5}{9} \cdot \bar{x} - \frac{5}{9} \cdot 32 = \frac{5}{9} \cdot (30.8 - 32) = -2/3 \approx -0.6667.$$

Hausübung

H 1 Beim Auszählen von Zellen in 50 Quadranten eines Hämazytometers ergaben sich die folgenden Werte:

1 2 2 2 4 4 4 5 5 5 2 1 2 2 7 6 7 4 4 4 4 4 4 4 4
4 4 4 4 5 5 6 6 2 3 3 3 3 3 6 7 7 7 5 2 2 2 7 9 9

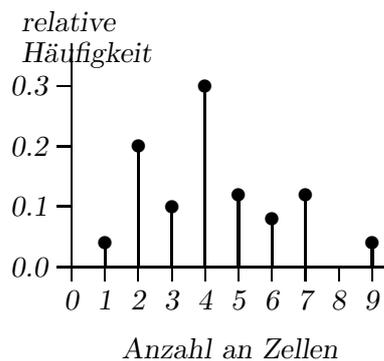
- a) Fertigen Sie ein Stabdiagramm zu den relativen Häufigkeiten dieser Messwerte an und zeichnen Sie die empirische Verteilungsfunktion.
- b) Lesen Sie das p -Quantil für $p = 0.2$ sowie $p = 0.76$ an der empirischen Verteilungsfunktion ab. Bestimmen Sie an Hand der geordneten Messreihe das empirische p -Quantil für $p = 0.25$, $p = 0.5$ und $p = 0.84$.

c) Berechnen Sie die empirische Standardabweichung der Messreihe.

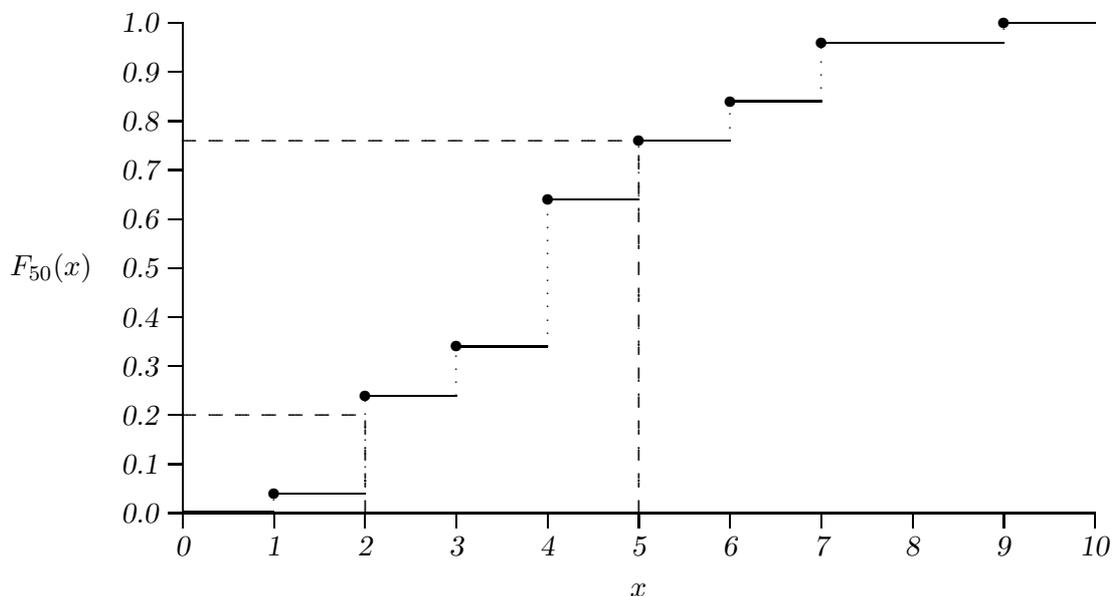
a) Es ergibt sich zunächst die folgende Häufigkeitstabelle:

i	1	2	3	4	5	6	7	8
Anzahl der Zellen w_i	1	2	3	4	5	6	7	9
absolute Häufigkeit	2	10	5	15	6	4	6	2
relative Häufigkeit $p_{50}(w_i)$	0.04	0.2	0.1	0.3	0.12	0.08	0.12	0.04

und somit folgendes Stabdiagramm:



Skizze der empirischen Verteilungsfunktion:



b) In die Skizze der empirischen Verteilungsfunktion in a) sind bereits die für die Quantilsbestimmung relevanten Linien eingezeichnet. Für $p = 0.2$ wird auf der Höhe 0.2 eine waagerechte Linie gezeichnet, bis man auf die empirische Verteilungsfunktion trifft. Da dies gerade bei einer Stufe der Treppe der Fall ist, zieht man an dieser Stelle lediglich einen senkrechten Strich, um den Wert $x_{0,2} = 2$ auf der x -Achse abzulesen. Für $p = 0.76$ geht man genauso vor. Nun trifft die waagerechte Linie die empirische Verteilungsfunktion am oberen Ende einer Treppenstufe. Auch hier fallen wir direkt das Lot und lesen den Wert $x_{0,76} = 5$ auf der x -Achse ab. Zur Bestimmung der restlichen Quantile verwenden wir wie gewohnt die geordnete Messreihe:

1 1 2 2 2 2 2 2 2 2 2 3 3 3 3 3 4 4 4 4 4 4 4 4
 4 4 4 4 4 4 4 5 5 5 5 5 6 6 6 6 7 7 7 7 7 7 9 9.

Es ergibt sich:

p	np	$i(p)$	x_p
0.25	12.5	13	$x_{(13)} = 3$
0.5	25	25	$x_{(25)} = 4$
0.84	42	42	$x_{(42)} = 6$

c) Wir bestimmen zunächst das arithmetische Mittel. Mit Hilfe der Häufigkeitstabelle ergibt sich (vgl. Formel im Skript)

$$\begin{aligned} \bar{x}_n &= \frac{1}{n} \sum_{i=1}^n x_i \\ &= \sum_{i=1}^k p_n(w_i) \cdot w_i = \sum_{i=1}^8 p_{50}(w_i) \cdot w_i \\ &= 0.04 \cdot 1 + 0.2 \cdot 2 + 0.1 \cdot 3 + 0.3 \cdot 4 + 0.12 \cdot 5 + 0.08 \cdot 6 + 0.12 \cdot 7 + 0.04 \cdot 9 = 4.22 . \end{aligned}$$

Für die empirische Varianz gilt damit:

$$\begin{aligned} s_x^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}_n^2 \\ &= \sum_{i=1}^k p_n(w_i) \cdot w_i^2 - \bar{x}_n^2 \\ &= 0.04 \cdot 1^2 + 0.2 \cdot 2^2 + \dots + 0.08 \cdot 6^2 + 0.12 \cdot 7^2 + 0.04 \cdot 9^2 - 4.22^2 = 3.7316, \end{aligned}$$

und schließlich für die empirische Standardabweichung

$$s_x = \sqrt{s_x^2} \approx 1.9317 .$$

H 2 Aus einer Studie zur Beurteilung der Berufsaussichten von Studienabgängern mit sozialwissenschaftlicher Ausrichtung an der Universität München 1995 sind uns die Befragungsergebnisse der folgenden 5 Merkmale bekannt:

G: Geschlecht (1=weiblich, 2=männlich);

S: Studiendauer in Semester;

E: fachliches Engagement im Studium (1=sehr engagiert, 2=engagiert, 3=mäßig engagiert, 4=kaum engagiert, 5=nicht engagiert);

D: Ausrichtung der Diplomarbeit (1=Primärerhebung, 2=Sekundärerhebung, 3=empirisch-qualitativ, 4=Literaturarbeit);

N: Gesamtnote der Diplomprüfung.

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

- (a) Geben Sie die Skalenart jedes Merkmals an.
- (b) Erstellen Sie für jedes einzelne Merkmal eine Häufigkeitstabelle mit den Spalten: absolute Häufigkeit - kumulierte absolute Häufigkeit - relative Häufigkeit in % - kumulierte relative Häufigkeit in %.
- (c) Erstellen Sie für jedes einzelne Merkmal ein Stabdiagramm für die relativen Häufigkeiten. Zeichnen Sie in die Stabdiagramme den jeweiligen Polygonzug dazu.

- (a) *G - Geschlecht - Nominalskala*
- S - Studiendauer - Intervallskala*
- E - Engagement - Ordinalskala*
- D - Diplomarbeit - Nominalskala*
- N - Gesamtnote - Intervallskala*

(b)

Geschlecht	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
weiblich	17	17	47.2	47.2
männlich	19	36	52.8	100

Diplomarbeit	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
1	9	9	25	25
2	11	20	30.6	55.6
3	9	29	25	80.6
4	7	36	19.4	100

Studiendauer	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
9	2	2	5.6	5.6
10	8	10	22.2	27.8
11	5	15	13.9	41.7
12	7	22	19.4	61.1
13	6	28	16.7	77.8
14	2	30	5.6	83.4
15	4	34	11.1	94.5
16	2	36	5.6	100

Engagement	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
1	8	8	22.2	22.2
2	13	21	36.1	58.3
3	8	29	22.2	80.5
4	5	34	13.9	94.4
5	2	36	5.6	100

Gesamtnote	abs. H	kum. abs. H	rel. H (%)	kum. rel. H (%)
1	9	9	25	25
2	12	21	33.3	58.3
3	9	30	25	83.3
4	4	34	11.1	94.4
5	2	36	5.6	100