

Musterlösung zur Klausur Statistik

WMKE18A

Oettinger 06.2019

Zeit: 60Min.

Insgesamt erreichbare Punktzahl: 50.

Aufgabe 1

- (a) Der Median entspricht dem 50%-Quantil - richtig, er ist so definiert.
- (b) Median und Modus sind verschiedene Mittelwerte, sie ehmen im Allgemeinen nicht denselben Wert an - falsch.
- (c) Median und Modus können bei einer vollständig symmetrischen Verteilung natürlich auch denselben Wert annehmen - falsch.
- (d) Das arithmetische Mittel kann auch negative Werte annehmen, beispielsweise bei negativen Merkmalswerten - falsch.

Aufgabe 2

6-stellige Matrikelnummern für die Hochschule, Berechnung unter der vereinfachenden Annahme, dass auch 000000 als Matrikelnummer zulässig sein soll:

- (a) Jede beliebige Antwort.
- (b) Anzahl A aller möglichen Kombinationen für 6-stellige Nummern: Jede Stelle kann mit $0 \dots 9$ besetzt werden. $A = 10^6$ (unter der Annahme, dass auch 000000 als ID-Nummer zählt).

- (c) Soll keine Nummer mit einer Null beginnen, gibt es für die erste Stelle nur 9 Möglichkeiten (1 . . . 9). $A = 9 \cdot 10^5$.
- (d) Keine Ziffer soll zweimal vorkommen (Berechnung unter der Annahme, dass eine führende Null erlaubt ist): für die erste Stelle gibt es 10 Möglichkeiten,
für die zweite 9,
für die dritte 8 usw.
 $A = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 = 151200$

Aufgabe 3

Benötigt werden zunächst die Mittelwerte (arithmetisches Mittel) der beiden Stichproben (Angaben in Jahren).

Gruppe 1:

$$\bar{x}_1 = \frac{1}{6}(3 \cdot 22 + 2 \cdot 23 + 28) = 23,33$$

Gruppe 2:

$$\bar{x}_2 = \frac{1}{6}(40 + 42 + 39 + 44 + 2 \cdot 45) = 42,5$$

Als Maß für die Streuung wählen wir die Standardabweichung

$$s = +\sqrt{s^2} \text{ mit } s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2,$$

also für die beiden Gruppen

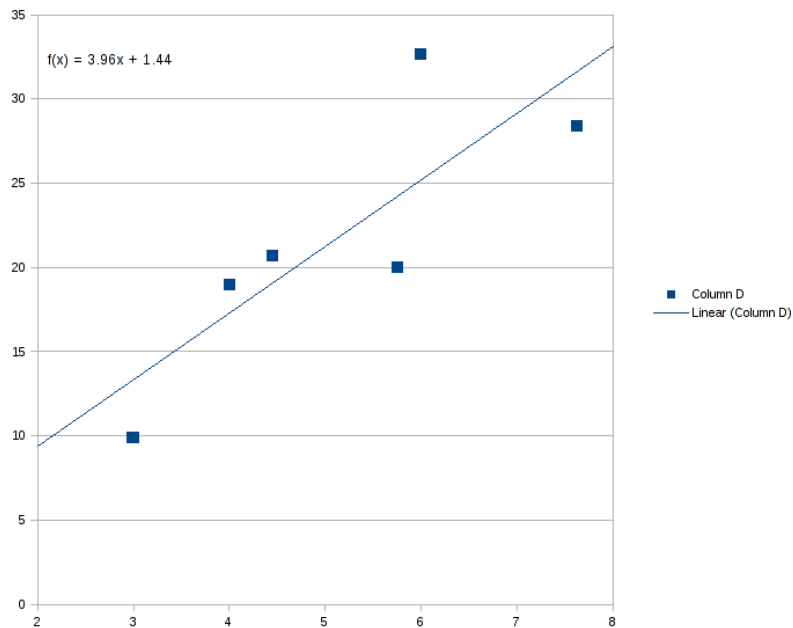
$$s_1 = \sqrt{\frac{1}{6}(3 \cdot (22 - 22,83)^2 + 2 \cdot (23 - 22,83)^2 + (28 - 22,83)^2)} = 2,13$$

$$s_2 = \sqrt{\frac{1}{6}((40 - 42,5)^2 + (42 - 42,5)^2 + (39 - 42,5)^2 + (44 - 42,5)^2 + 2 \cdot (45 - 42,5)^2)} = 2,36$$

Die Streuung der Einzelwerte um den Mittelwert ist in der ersten Gruppe etwas geringer, also sind die einzelnen Merkmalswerte typischerweise näher beieinander.

Aufgabe 4

- a) Alle Werte sind zulässig.
 b) Daten mit Ausgleichsgerade



- c) Anpassung einer Geraden $y = a \cdot x + b$ über lineare Regression: Tabelle benötigter Daten

City	Hamburger	Kinokarten				
	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
Toyo	5.99	32.66	1.1	12.64	13.9	1.21
London	7.62	28.41	2.73	8.39	22.9	7.45
New York	5.75	20	0.86	-0.02	-0.02	0.74
Sydney	4.45	20.71	-0.44	0.69	-0.3	0.19
Rio de Janeiro	2.99	9.9	-1.9	-10.12	19.23	3.61
Ravensburg	4	19	-0.89	-1.02	0.91	0.79

Die Steigung der Geraden ist (für die Daten oben)

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = 3,96$$

der Achsenabschnitt ist

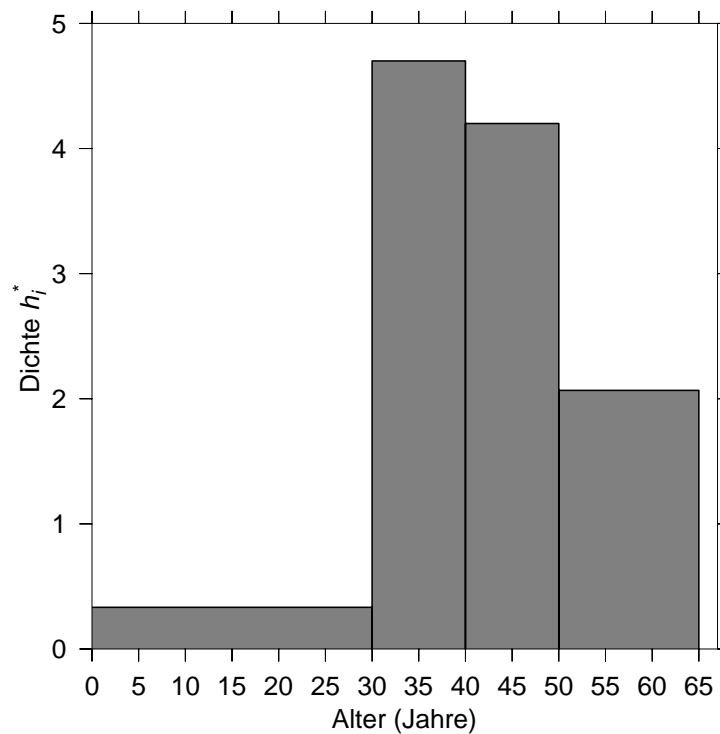
$$b = \bar{y} - a\bar{x} = 1,44$$

Aufgabe 5

Daten zur Altersverteilung:

Alter in Jahren von ... bis unter ...	Absolute Häufigkeit	relative Häufigkeit f_i	kumulierte rel. Häufigkeit F_i	Klassenbreite Δ_i	Dichte $h_i^* = h_i/\Delta_i$
bis 30	10	0,076	0,076	30	0,333
30 - 40	47	0,362	0,438	10	4,7
40 - 50	42	0,323	0,761	10	4,2
50 -65	31	0,238	1	15	2,07

(a) Histogramm der Altersverteilung:
Altersverteilung



(b) Berechnung des Durchschnittseinkommens (in tausend Credits):

$$\begin{aligned}\bar{y} &= \frac{1}{130}(10 \cdot 1,25 + 47 \cdot 3,35 + 42 \cdot 4,6 + 31 \cdot 4,95) \\ &= \frac{516,60}{130} = 3,97\end{aligned}$$

(c) Berechnung des Medians:

50% der befragten Personen werden in der 3.Klasse erreicht. Der Median lässt sich also wie folgt berechnen:

$$\begin{aligned}\bar{x}_Z &= x_3^u + (x_3^o - x_3^u) \frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = x_3^u + (x_3^o - x_3^u) \frac{F(0,5) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} \\ &= 40 + 10 \cdot \frac{0,5 - 57/130}{(99 - 57)/130} = 40 + 10 \cdot \frac{8}{42} = 41,9\end{aligned}$$

Aufgabe 6

Geeignete Mittelwerte.

1. Der Mittelwert berechnet sich nach

$$\bar{x} = \frac{1}{30}(10 \cdot 1 + 11 \cdot 2 + 8 \cdot 3 + 4) = \frac{60}{30} = 2,0.$$

Für die Kandidaten, die nicht bestanden haben, kann keine Note angegeben werden, sie werden zur Berechnung *nicht* herangezogen.

2. Insgesamt befragte Personen: $100 + 1000 = 1100$. Für die Abschaffung sind $60 + 380 = 440$. Also sind $440/1100 = 40\%$ dafür.