Musterlösung zur Nachklausur Statistik

TIT/TIM/TIS 19 / Oettinger / 10.2021

Insgesamt erreichbare Punktzahl: 105, 100%: 100 Punkte.

Aufgabe 1

(15 Punkte)

(a) Das arithmetische Mittel wird durch die Ausreißer mit kleinem Betrag geringer, also nach unten verschoben - falsch.

(b) Die Standardabweichung ist als positiver Ast der Wurzel der Varianz definiert (die immer positiv oder Null ist) - richtig.

(c) Die Varianz kann natürlich Null werden - wenn alle Merkmalswerte gleich sind (auch wenn dann eigentlich keine Statistik notwendig ist) - falsch.

(d) Der Modus ist der am häufigsten auftretende Wert, das 100%-Quantil entspricht dem Maximum der Werte - falsch.

(e) Das geometrische Mittel kann bei negativer Veränderungen auch negative Werte annehmen - falsch.

Aufgabe 2

(24 Punkte)
Die arithmetischen Mittel sind $\bar{x} = 5000$ und $\bar{y} = 60000$ EUR. Die benötigten Daten in Tabellenform:

Monat i	1	2	3	4	5	6	(Summe)
Menge x_i	2.000	3.000	6.000	4.000	8.000	7.000	30000
Kosten y_i	30.000	35.000	75.000	55.000	85.000	80.000	360000
$(x_i - \bar{x})$	-300	-200	100	-100	300	200	0
$(y_i - \bar{y})$	-3000	-2500	1500	-500	2500	2000	0
$(x_i - \bar{x})(y_i - \bar{y})$	900000	500000	150000	50000	750000	400000	2750000
$(x_i - \bar{x})^2$	90000	40000	10000	10000	40000	90000	280000
$(y_i - \bar{y})^2$	9000000	6250000	2250000	250000	6250000	4000000	28000000

a) Die Ausgleichsgerade lautet $y = a \cdot x + b$, die Werte für die Steigung und den Achsenabschnitt sind

$$a = \frac{S_{xy}}{S_{xx}} = \frac{sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{2750000}{28000000} = 9,82$$

Achsenabschnitt $b = \bar{y} - m \cdot \bar{x} = 1089,3$

b) Die Gerade beschreibt den Zusammenhang zwischen der produzierten Menge und den Gesamtkosten des Produkts. Die Steigung gibt die Kosten je produzierter Einheit an, der Achsenabschnitt beschreibt Fixkosten, die von der hergestellten Menge unabhängig sind.

Aufgabe 3

(10 Punkte)
Redezeiten von Politikern, alle Angaben und Ergebnisse in Minuten:

Politiker	A	B	C	D	E	F
Redezeit	6	8	10	12	20	10

(a) Geordneter Vektor der Merkmalswerte: (6, 8, 10, 10, 12, 20). Median ist

$$\bar{x}_Z = \frac{10 + 10}{2} = 10.$$

(b) Der Modus ist 10, $\bar{x} = \frac{1}{6} \cdot (6 + 8 + 10 + 10 + 12 + 20) = 11$.

(c) $s_w =$max$(x_i)-$min$(x_i) = 20 - 6 = 14$.

(d) mittlere absolute Abweichung

$$d_{\bar{x}} = \frac{1}{6} \sum_{i=1}^{n} \mid x_i - \bar{x} \mid$$

$$= \frac{1}{6} (\mid 6 - 11 \mid + \mid 8 - 11 \mid + 2 \cdot \mid 10 - 11 \mid + \mid 12 - 11 \mid + \mid 20 - 11 \mid) = 10/3$$

Aufgabe 4

(17 Punkte)
Vollständige gemeinsame Häufigkeitstabelle (Kreuztabelle) der beiden Zufallsvariablen X und Y

Schuhgröße / Mathenote	39	40	41	Summe
1	1	2	1	4
2	2	4	2	8
3	1	2	1	4
Summe	4	8	4	16

und daraus berechnete relative Häufigkeiten:

Y / X	39	40	41	$f_X(X)$
1	1/4	1/4	1/4	1/4
2	1/2	1/2	1/2	1/2
3	1/4	1/4	1/4	1/4
$f_Y(Y)$	1	1	1	1

- Die Spalten sowie die Randverteilung der zweiten Tabelle sind identisch $\Longrightarrow$ die Zufallsvariablen X und Y sind stochastisch unabhängig.
- Häufigkeitsverteilung der Mathenote für Schuhgröße 40: $f(x_i|Y = 40) = (1/8, 1/4, 1/8)$
- Varianz $s^2(X|Y = 40) = \sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i$. Benötigt wird das arithmetische Mittel (Berechnung aus der Spalte mit $Y = 40$)

$$\bar{x} = \frac{1}{1/2}(1/8 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3) = 2 \cdot (\frac{1+4+3}{8}) = 2$$

$$s^2(X|Y = 2) = \sum_{i=1}^{3}(x_i - \bar{x})^2 \cdot f_i = (1-2)^2 \cdot \frac{1}{4} + (2-2)^2 \cdot \frac{1}{2} + (3-2)^2 \cdot \frac{1}{4}$$

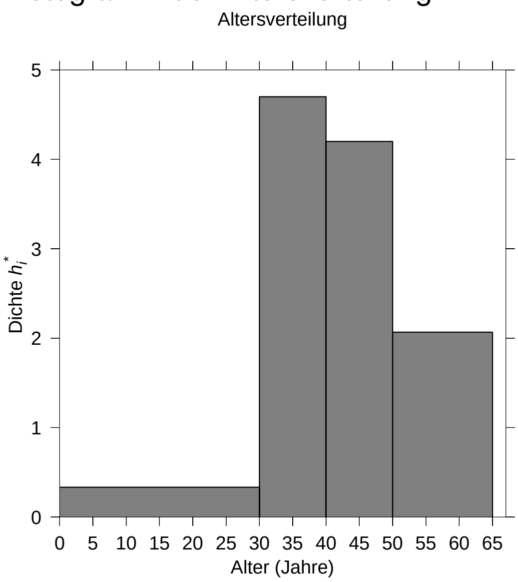
$$= \frac{1}{4} + 0 + \frac{1}{4} = \frac{1}{2}$$

Aufgabe 5

(25 Punkte)
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	2,067

(a) Histogramm der Altersverteilung:



(b) Berechnung des Durchschnittsalters (in Jahren):

$$\bar{x} = \frac{1}{130}(10 \cdot 15 + 47 \cdot 35 + 42 \cdot 45 + 31 \cdot 57,5) = 42,06$$

(c) Berechnung des Medians:
50% der befragten Personen werden in der 3.Klasse erreicht. Der Median lässt sich also wie folgt berechnen:

$$F(\bar{x}_Z) = x_3^u + (x_3^o - x_3^u)\frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = x_3^u + (x_3^o - x_3^u)\frac{F(0,5) - F(x_3^u)}{F(x_3^o) - F(x_3^u)}$$

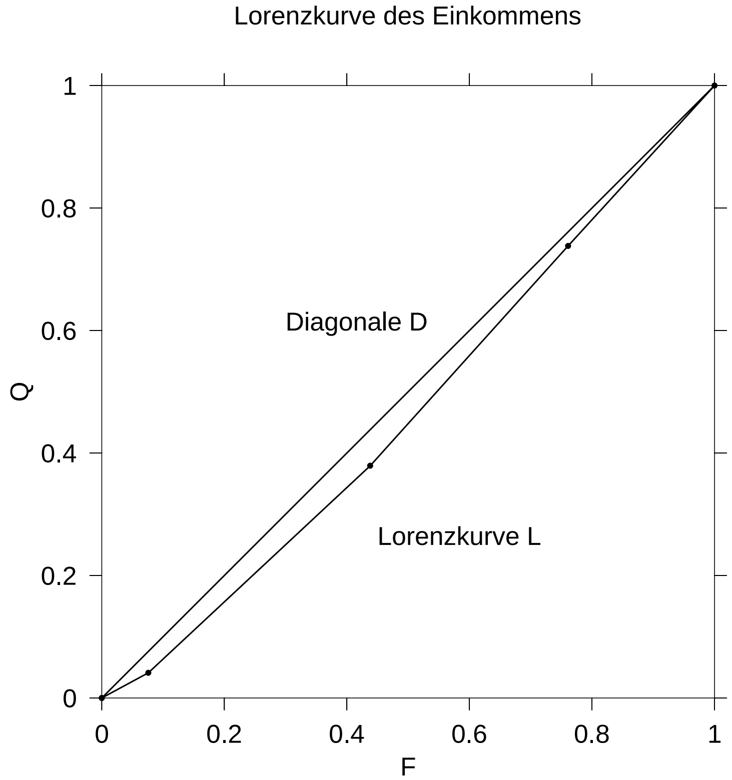
$$= 40 + 10 \cdot \frac{0,5 - 57/130}{(99 - 57)/130} = 40 + 10 \cdot \frac{8}{42} = 41,9$$

(d) Für die Darstellung der Lorenzkurve werden die relativen Anteile q_i an der Gesamtsumme der Merkmalswerte bzw deren kumulierte Werte Q_i benötigt:

$$\begin{aligned} q_{ges} &= 10 \cdot 2,5 + 47 \cdot 4,2 + 42 \cdot 5,0 + 31 \cdot 4,9 = 584,3 \\ q_1 &= \frac{10 \cdot 2,5}{584,3} = 0,0428 \\ &\vdots \end{aligned}$$

Alter in Jahren von ... bis unter ...	Absolute Häufigkeit	$\bar{y}$	relativer Anteil q_i	kumulierte rel. Anteile Q_i
bis 30	10	2,5	0,0428	0,0428
30 - 40	47	4,2	0,3438	0,381
40 - 50	42	5,0	0,359	0,740
50 -65	31	4,9	0,260	1

Der Wert des Gini-Koeffizienten entspricht der doppelten Fläche zwischen der Diagonalen und der Lorenzkurve. Der Gini-Koeffizient ist sehr klein, damit ist die Disparität (die relative Konzentration) klein, die Einkommen sind relativ gleichmäßig verteilt.



Aufgabe 6

(14 Punkte)
Daten zum Reis:

Jahr	2014	2015	2016	2017	2018	2019	2020
x_i	13	7	12	8	11	5	7

Tabelle 1: Daten zu Reisunfällen in China

Das arithmetische Mittel ist

$$\bar{x} = \frac{1}{7}(13 + 7 + 12 + 8 + 11 + 5 + 7) = 9,$$

der Median ist der zentrale Wert in der geordneten Stichprobe

$$\bar{x}_Z = 8.$$

Die Varianz ist die mittlere qadratische Abweichung vom arithmetischen Mittel

$$s^2 = \frac{1}{7}((5-9)^2 + 2 \cdot (7-9)^2 + (8-9)^2 + (11-9)^2 + (12-9)^2 + ({}^3-9)^2) = 7,71$$

und damit die Standardabweichung

$$s = \sqrt{s^2} = 2,78$$