Musterlösung zur Klausur Statistik

TIT11 03.2013 (Oettinger), 90 Min.

Aufgabe 1

(a) Richtig: das arithmetische Mittel reagiert im Gegensatz zum Median empfindlich auf Ausreißer.

(b) Falsch: ein ordinales Merkmal besitzt eine natürliche Rangfolge, ein nominales Merkmal keine.

(c) Richtig: es handelt sich um eine Summe quadrierter Größen.

(d) Falsch: der Modus ist die am häufigsten auftretende Merkmalsausprägung in einer Stichprobe.

(e) Falsch: es handelt sich um ein mittleres Wachstum, es kann auch negativ sein.

Aufgabe 2

Die arithmetischen Mittel für die Siedetemperatur T_S und den Druck sind $\bar{x} = 95,6$°C und $\bar{y} = 652,806$ mm Hg. Damit lassen sich die benötigten Daten berechnen:

T_S (°C)	Druck	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
90.3	528.07	-5.3	-124.736	28.09	15559.069696	661.1008
92.2	568.96	-3.4	-83.846	11.56	7030.151716	285.0764
93	588.01	-2.6	-64.796	6.76	4198.521616	168.4696
93.8	606.81	-1.8	-45.996	3.24	2115.632016	82.7928
94.1	610.11	-1.5	-42.696	2.25	1822.948416	64.044
95.9	674.88	0.3	22.074	0.09	487.261476	6.6222
98.6	723.65	3	70.844	9	5018.872336	212.532
98.1	705.1	2.5	52.294	6.25	2734.662436	130.735
99.9	758.95	4.3	106.144	18.49	11266.548736	456.4192
100.1	763.52	4.5	110.714	20.25	12257.589796	498.213

Die Kovarianz (nicht gefragt, in °C · mm Hg) ist

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 256,60$$

Lineare Regression sei $y(x) = a \cdot x + b$:

$$a = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 24,21$$

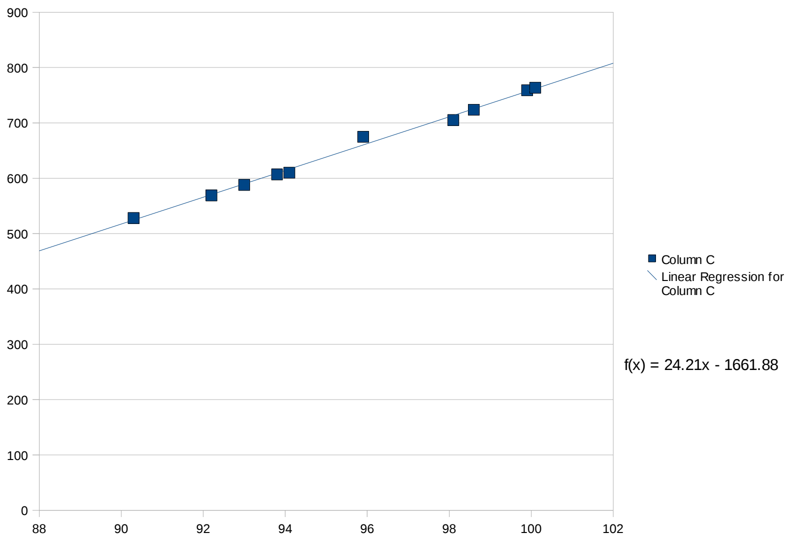
und

$$b = \bar{y} - a\bar{x} = -1661,88.$$

Die Standardabweichungen der beiden Größen sind $s_x = 3,26$°C und $s_y = 79,05$ mm Hg. Der Pearson-Koeffizient ist

$$r_{xy} = \frac{s_{xy}}{s_x \cdot s_y} = 0,997.$$

Er liegt sehr nahe bei 1, es besteht (rechnerisch) ein starker linearer Zusammenhang zwischen den Merkmalen. Plot der Daten:



Aufgabe 3

Alle Angaben in min:

(a) Geordneter Vektor der Verspätungen: (2, 6, 8, 8, 10, 10, 12, 20). Der Umfang der Stichprobe ist $n = 8$, das untere Quartil (das 0,25-Quantil) ist $p = \frac{x_{n/4}+x_{n/4+1}}{2} = \frac{6+8}{2} = 7$.

(b) Mittelwert $\bar{x} = \frac{1}{8}(2 + 6 + 2 \cdot 8 + 2 \cdot 10 + 12 + 20) = 9,5$.

(c) Die Spannweite ist $\max(\{x_i\}) - \min(\{x_i\}) = 20 - 2 = 18$.

(d) Die mittlere quadratische Abweichung (die Varianz) hat den Wert

$$s^2 = \big((2-9,5)^2 + (6-9,5)^2 + 2 \cdot (8-9,5)^2 + 2 \cdot (10-9,5)^2 + (12-9,5)^2 \\ +(20-9,5)^2\big) - 9,5^2 = 23,75$$

Aufgabe 4

Es handelt sich beim idealen Wüfel um ein Laplace-Experiment mit einer Wahrscheinlichkeit von $1/6$ für das Auftreten einer bestimmten Augenzahl.

(a) Wahrscheinlichkeit, zweimal hintereinander die zwei zu würfeln: $P = 1/6 \cdot 1/6 = 1/36$.

(b) Für den ersten Würfel sind alle Augenzahlen zulässig, für den zweiten nur die vom ersten Würfel vorgegebene: $P = 1 \cdot 1/6 = 1/6$.

(c) Die Wahrscheinlichkeit dafür, dass die Würfel eine Augensumme zeigen ist 1. Die Wahrscheinlichkeit für unterschiedliche Augenzahlen ist die Wahrscheinlichkeit einer Augensumme minus der Wahrscheinlichkeit dafür, dass beide Würfel dieselbe Augenzahl zeigen, also $P = 1 - 1/6 = 5/6$.

Aufgabe 5

Die benötigten Daten sind in der folgenden Tabelle zusammengefasst.

Alter (Jahre) ...	Absolute Häufigkeit	$\bar{y}_i$	$s^2_{Y,i}$	Klassen-breite Δ_i	h_i^*	rel. f_i	kumul. F_i	rel. q_i	kumul. Q_i
bis 30	10	2,5	1,8	30	0,37	0,0775	0,0775	0,04	0,04
30 - 40	47	4,5	2,9	10	4,7	0,364	0,4415	0,337	0,377
40 - 50	42	5,3	3,4	10	4,5	0,326	0,768	0,355	0,732
50 -65	30	5,6	3,6	15	0,2	0,233	1	0,268	1

Der Anteil der einzelnen Altersgruppen q_i am Gesamteinkommen kann aus dem mittleren Einkommen $\bar{y}_i$ berechnet werden, indem aus den absoluten Häufigkeiten und dem mittleren Einkommen die Summe des Einkommens in der Altersgruppe berechnet und durch die ebenso bestimmte Gesamtsumme geteilt wird.

(a) Berechnung des arithmetischen Mittels in Jahren: Histogramm:

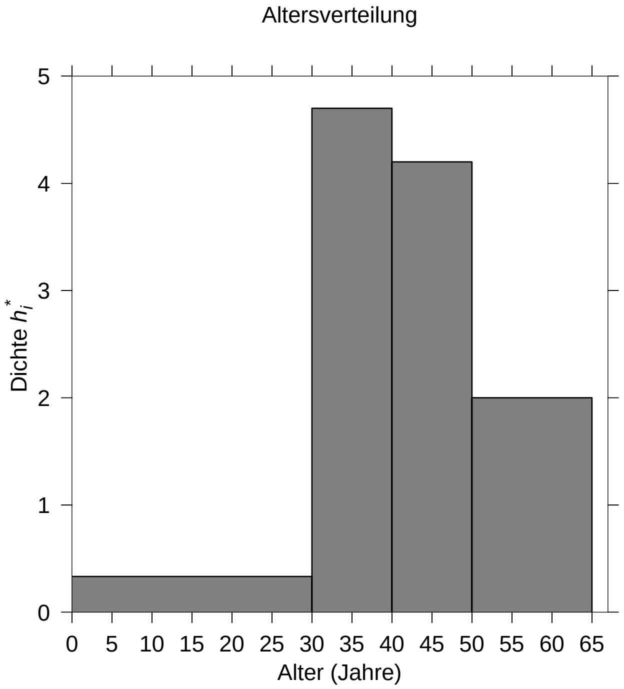


Abbildung 1: Histogramm der Altersverteilung

$$\bar{x} = \frac{1}{129}(15 \cdot 10 + 35 \cdot 47 + 45 \cdot 42 + 57,5 \cdot 30) = 41,94.$$

(b) Berechnung des Medians in Jahren unter Annahme von Gleichverteilung innerhalb der Klassen: der Wert von 0,5 wird in der dritten Klasse ($40 - 50$ Jahre) erreicht.

$$\begin{aligned}\bar{x}_Z &= x_3^u + (x_3^o - x_3^u) \cdot \frac{F(\bar{x}_Z) - F(x_3^u)}{F(x_3^o) - F(x_3^u)} \\ &= 40 + (50 - 40) \cdot \frac{0,5 - 0,4415}{0,768 - 0,4415} = 41,79\end{aligned}$$

(c) Lorenz-Plot:

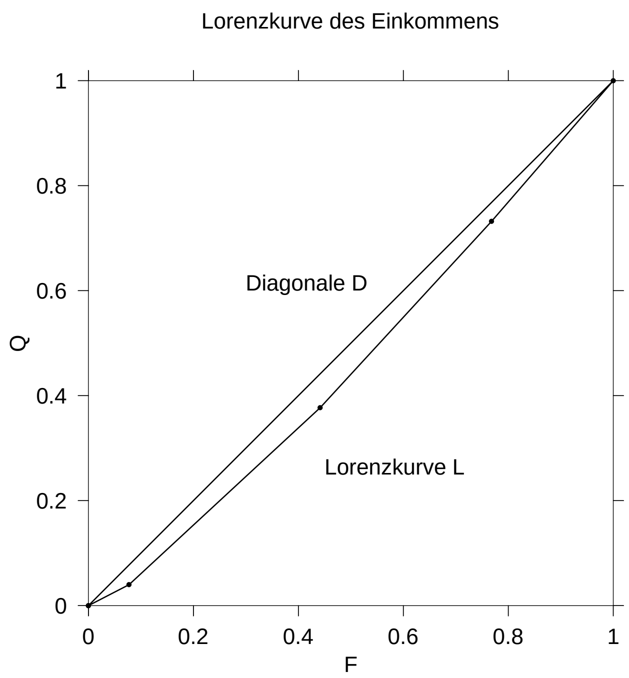


Abbildung 2: Lorenzkurve der Einkommensverteilung

Der Gini-Koeffizient berechnet sich nach

$$G = 1 - \sum_{i=1}^{n} f_i(Q_{i-1} + Q_i)$$

$$= 1 - (0,078 \cdot 0,04 + 0,4415(0,377 + 0,04) + 0,326(0,732 + 0,377) + 0,233(1 + 0,732))$$

$$= 0,048$$

Der kleine Gini-Koeffizient weist auf eine geringe relative Konzentration hin.

Aufgabe 6

(a) die mittlere erreichte Geschwindigkeit ergibt sich ganz einfach, indem die gesamte zurückgelegte Strecke durch die gesamte benötigte Zeit geteilt wird (Geschwindigkeit in km/h):

$$\bar{v} = \frac{1 \cdot 7 + 1,25 \cdot 5}{1 + 1,25} = 5,89$$

(b) Der Mittelwert berechnet sich nach

$$\bar{x} = \frac{1}{20}(5 \cdot 1 + 11 \cdot 2 + 3 \cdot 3 + 4) = 2,0$$

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

(c) Geometrisches Mittel:

$$\bar{x}_G = \sqrt{1,56 \cdot 1,04} - 1 = 0,274 \text{ oder } 27,4\%$$

(d) Wiederum geometrisches Mittel:

$$\bar{x}_G = \sqrt{2 \cdot 2} - 1 = 2 - 1 = 1 \text{ oder } 100\%,$$

genau das ist ja die Aussage (verdoppelt sich jede Nacht!).

(e)

$$\bar{x}_G = \sqrt{2 \cdot 8} - 1 = 4 - 1 = 3 \text{ oder } 300\%.$$

Aufgabe 7

1. Studierende der DHBW in Ravensburg: Bestandsmasse.
2. Todesfälle in einer Gemeinde: Bewegungsmasse.
3. Maschinenausfälle in einer Werkstatt: Bewegungsmasse.
4. Die Bevölkerung in Ravensburg zum 1.9.2010: Bestandsmasse.