Musterlösung zur Übungsklausur Statistik

TIT15 / Oettinger / 03.2017

Aufgabe 1

(15 Punkte)

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

(b) Falsch: die Varianz und damit auch die Standardabweichung kann natürlich Null werden (wenn alle Abweichungen vom arithmetischen Mittel zu Null verschwinden) - dann ist eigentlich keine Statistik nötig.

(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

(39 Punkte)
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

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

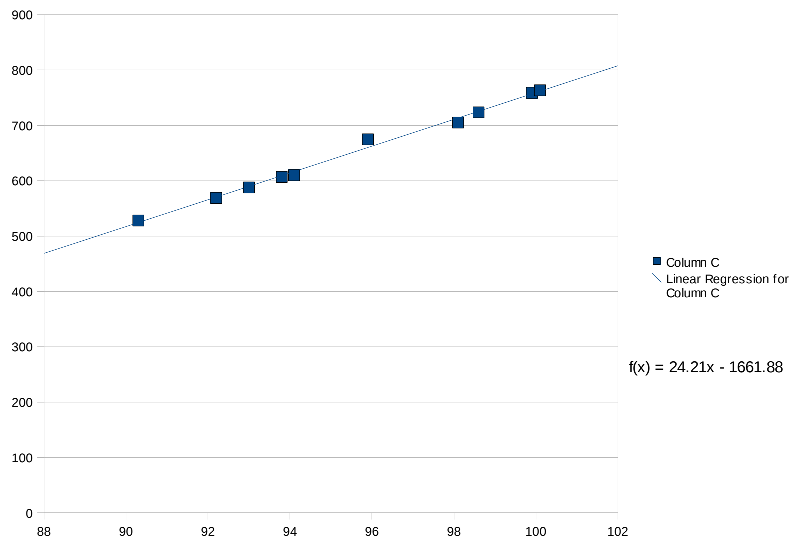
Lineare Regression: sei $y(x) = a \cdot x + b$:

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

und

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

Die Standardabweichungen der beiden Größen sind $s_x = 3,26$°C und $s_y = 79,05$ mm Hg. Plot der Daten:



Aufgabe 3

(13 Punkte)
Die Tabelle in Form relativer Häufigkeiten:

X (PKW) Verbrauch Y	1	2	3	4	$\sum$
6,1	0,182	0,2	0,182	0,194	0,185
6,2	0,545	0,6	0,545	0,516	0,546
6,3	0,273	0,2	0,273	0,290	0,269
$\sum$	1	1	1	1	1

(a) Die Spalten der Tabelle relativer Häufigkeiten unterscheiden sich $\Longrightarrow$ die Größen X und Y sind nicht stochastisch unabhängig.

(b) Benötigt wird der Mittelwert des Verbrauchs für PKW 2 (alle Angaben in l):

$$\bar{y}(X=2) = \frac{1}{20}\,(4 \cdot 6,1 + 12 \cdot 6,2 + 4 \cdot 6,3) = 6,2.$$

Die Varianz ist die mittlere quadratische Abweichung vom Mittelwert, also

$$s^2(Y|X=2) = \frac{1}{20}\left(4 \cdot (6,1-6,2)^2 + 12 \cdot (6,2-6,2) + 4 \cdot (6,3-6,2)\right)$$

$$= \frac{1}{20} 8 \cdot 0,1^2 = 4 \cdot 10^{-3}.$$

Aufgabe 4

(43 Punkte)

a) Die benötigten Daten:

Einkommen in Talern	Klassenbreite in Tausend Talern	Häufigkeit	relativ	kumuliert	Dichte
]0-500]	0,5	9	0,09	0,09	18
]500-1000]	0,5	13	0,13	0,22	26
]1000-1500]	0,5	32	0,32	0,54	64
]1500-2000]	0,5	41	0,41	0,95	82
]2000-3000]	1,0	3	0,03	0,98	3
]3000-5000]	2,0	2	0,02	1,0	1

Histogramm der Einkommen:

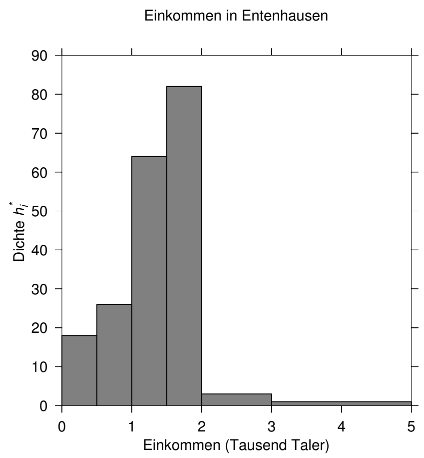


Abbildung 1: Einkommensverteilung in Entenhausen

b) arithmetisches Mittel (in Tausend Taler):

$$\bar{x} = \frac{1}{100}(0,25 \cdot 9 + 0,75 \cdot 13 + 1,25 \cdot 32 + 2,5 \cdot 3 + 4,0 \cdot 2) = \frac{139,25}{100} = 1,39$$

50% der befragten Personen werden in der 3.Klasse erreicht, der Median (in Tausend Taler) ist

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

$$1 + (1,5 - 1) \cdot \frac{0,5 - 0,22}{0,54 - 0,22} = 1,44.$$

Der Median ist etwas größer - die Verteilung scheint leicht rechtssteil.

c) Für die Lorenzkurve wird das gesamte Einkommen benötigt (in Tausend Talern):

$$g_{ges} = 0,25 \cdot 9 + 0,75 \cdot 13 + 1,25 \cdot 32 + +2,5 \cdot 3 + 4,0 \cdot 2 = 139,25$$

Jetzt kann der Anteil der Personen in jeder Klasse am Gesamteinkommen berechnet werden:

$$q_k = f_k \cdot (x_k^o + x_k^u)/2.$$

Einkommen in Talern	relative Häufigkeit f_k	kumuliert F_k	Anteil q_k	kumuliert Q_k
]0-500]	0,09	0,09	0,032	0,032
]500-1000]	0,13	0,22	0,070	0,102
]1000-1500]	0,32	0,54	0,287	0,389
]1500-2000]	0,41	0,95	0,515	0,904
]2000-3000]	0,03	0,98	0,054	0,958
]3000-5000]	0,02	1,0	0,057	1

Lorenz-Plot:

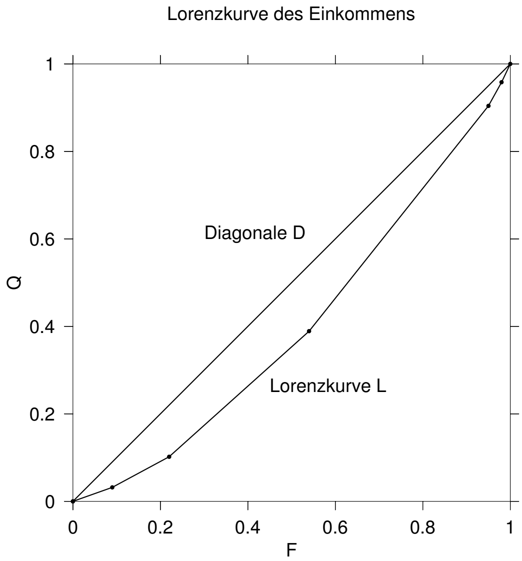


Abbildung 2: Lorenzkurve der Einkommensverteilung in Entenhausen

Der Gini-Koeffizient entspricht der doppelten Fläche zwischen den beiden Kurven. Man erkennt, dass die relative Konzentration (und damit der Gini-Koeffizient) nicht besonders groß ist.