Musterlösung zur Testklausur Statistik

WMS17A (Oettinger, 2018)

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 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) Die bedingte Verteilung $f(x_i|Y = 6,2)$ besteht aus den relativen Häufigkeiten der Zeile mit $Y = 6,2$, also

$$f(x_i|Y = 6,2) = \left\{\frac{6}{124}; \frac{12}{124}; \frac{90}{124}; \frac{9}{124}\right\}.$$

(c) 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 3

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:

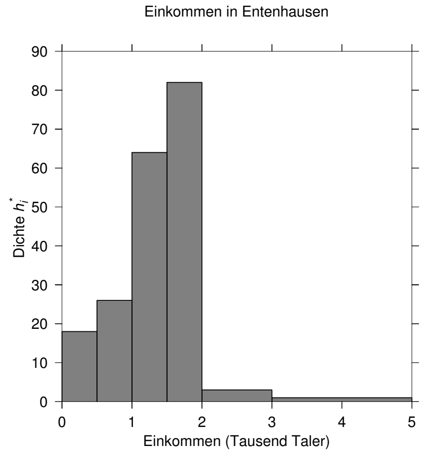


Abbildung 1: Einkommensverteilung in Entenhausen

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:

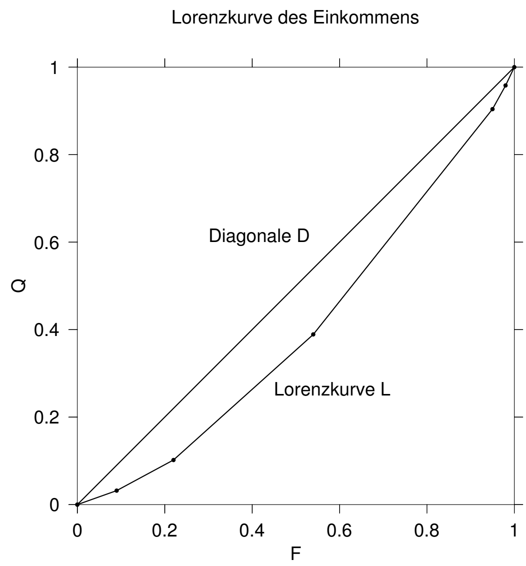


Abbildung 2: Lorenzkurve der Einkommensverteilung in Entenhausen

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.

Aufgabe 4

An einer neu gegründeten Universität sollen 6-stellige Matrikelnummern vergeben werden. Berechnung unter der vereinfachenden Annahme, dass auch 000000 als Matrikelnummer zulässig sein soll:

(a) Anzahl A aller möglichen Kombinationen für 6-stellige Matrikelnummern: Jede Stelle kann mit $0 \ldots 9$ besetzt werden. $A = 10^6$ (unter der Annahme, dass auch 000000 als Matrikelnummer zählt).

(b) Soll keine Nummer mit einer Null beginnen, gibt es für die erste Stelle nur 9 Möglichkeiten $(1 \ldots 9)$. $A = 9 \cdot 10^5$.

(c) 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 5

Die Aussage des Herrn Trumpet kann durch Vergleich der Variationskoeffizienten überprüft werden. Benötigt werden die arithmetischen Mittel (in %)

$$\bar{x}_A = \frac{1}{7}(5,6+6,3+6,6+6,9+7,1+7,6+6,1) = 6,6$$

$$\bar{x}_B = \frac{1}{7}(40,4+41,9+47,9+40,4+48,9+41,4+42,9) = 43,4$$

sowie die Varianzen und die daraus berechneten Standardabweichungen

$$s_A^2 = \frac{1}{7}\left(5,6^2+6,3^2+6,6^2+6,9^2+7,1^2+7,6^2+6,1^2\right) - 6,6^2$$

$$= 0,383$$

$$\Rightarrow s_A = \sqrt{s_A^2} = 0,619$$

$$s_B^2 = \frac{1}{7}\left(40,4^2+41,9^2+47,9^2+40,4^2+48,9^2+41,4^2+42,9^2\right) - 43,4^2$$

$$= 10,714$$

$$\Rightarrow s_B = \sqrt{s_B^2} = 3,273.$$

Die Variationskoeffizienten $v_x = \frac{s_x}{\bar{x}}$ für die beiden Stichproben sind

$$v_A = \frac{s_A}{\bar{x}_A} = \frac{0,619}{6,6} = 0,094$$

und

$$v_B = \frac{s_B}{\bar{x}_B} = \frac{3,273}{43,4} = 0,075.$$

Damit wird klar, dass Herr Trumpet falsch liegt, die Verteilung der Stimmenanteile für seine Partei A ist deutlich breiter.

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):

$$\bar{v} = \frac{1 \cdot 7 + 5}{1 + 1,25} = 5,\bar{3}$$

(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:

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

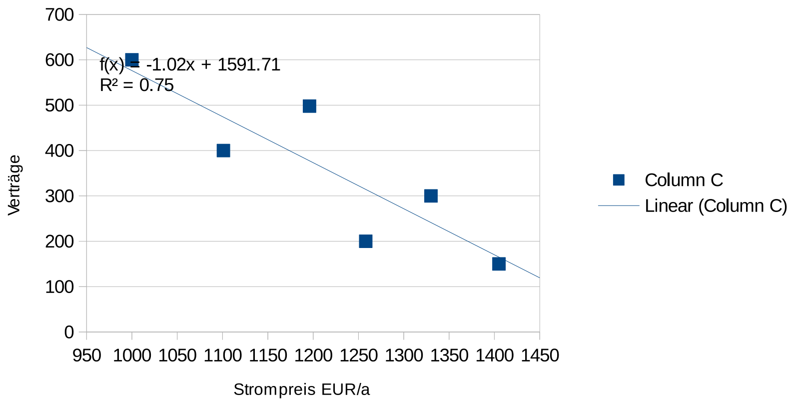
genau das ist ja die Aussage (verdoppelt sich jede Nacht!).

(e)

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

Aufgabe 7

a) Daten mit Ausgleichsgerade



Anpassung einer Geraden $y = a \cdot x + b$ über lineare Regression: die arithmetischen Mittel sind $\bar{x} = 1215$ und $\bar{y} = 358$, die Tabelle der benötigter Daten

Preis	Verträge	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1000	600	-215	242	-52030	46225	58564
1101	400	-114	42	-4788	12996	1764
1196	498	-19	140	-2660	361	19600
1258	200	43	-158	-6794	1849	24964
1330	300	115	-58	-6670	13225	3364
1405	150	190	-208	-39520	36100	43264

Die Steigung der Geraden ist

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = -1,02$$

der Achsenabschnitt ist

$$b = \bar{y} - a\bar{x} = 1591,71$$

b) Die erwartete Zahl der Verträge ist bei einem linearen Zusammenhang von $y = 1591,71 - 1,02 \cdot x$

$$y(1200) = 367,71.$$

Es werden also etwa 368 Verträge erwartet.

c) der Pearson-Koeffizient ist

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = -0,87 \quad (2)$$

Der Wert nahe eins legt einen relativ guten linearen Zusammenhang zwischen den Daten nahe.

Aufgabe 8

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.