Musterlösung zur Übungsklausur Statistik

WMS15B

Oettinger 9/2016

Aufgabe 1

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

(b) Falsch: die beiden Größen sind voneinander unabhängig - sie können gleich sein, müssen aber nicht.

(c) Richtig: es handelt sich um eine Summe positiver Abweichungen.

(d) Richtig: es kann als mittlerer Wert nur kleiner als das Maximum sein.

(e) Ein Merkmal ist entweder metrisch oder stetig, d.h. es gibt kein Merkmal, das gleichzeitig metrisch und stetig ist - falsch, ein stetiges Merkmal ist immer auch metrisch (Zahlen!).

Aufgabe 2

(a) Jede Antwort ist zulässig

(b) Statistische Einheit sind Giovannis Kunden, Merkmale sind Speisen und Getränke. Die Merkmalsausprägungen sind die vom Kunden bestellten Gerichte und Getränke, beispielsweise 'Rotwein' oder 'Pizza'.

(c) Darstellung der Daten im Stab- oder Balkendiagramm. Grafik nicht gefordert!

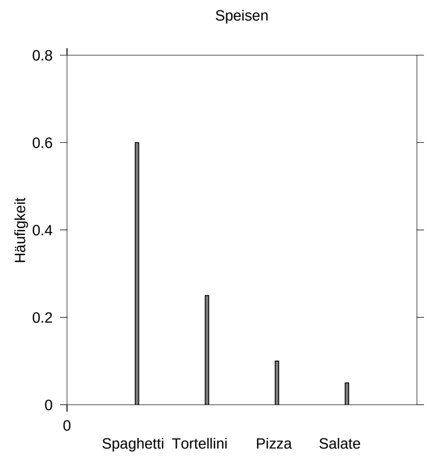


Abbildung 1: Stabdiagramm als Beispiel für die grafische Darstellung

(d) Ereignis A: Kunde bestellt Spaghetti, $P(A) = 0,6$.
Ereignis B: Kunde bestellt Rotwein, $P(B) = 0,7$.
Ereignis C: Kunde bestellt Spaghetti und Rotwein, $P(C) = P(A) \cdot P(B) = 0,6 \cdot 0,7 = 0,42 = 42\%$.
Ereignis D: Kunde bestellt Spaghetti, aber keinen Rotwein, $P(D) = P(A) \cdot (1 - P(B)) = 0,6 \cdot (1 - 0,7) = 0,6 \cdot 0,3 = 0,18 = 18\%$.

(e) Die relative Häufigkeit für eine Spaghettilänge größer als 80cm ist $0,15 + 0,05 = 0,2 = 20\%$.
Das arithmetische Mittel (in cm) muss mit den Klassenmitten gerechnet werden (wegen der Angabe relativer Häufigkeiten wird nicht durch die Gesamtzahl geteilt):

$$\bar{x} = 0,15 \cdot 30 + 25 \cdot 50 + 40 \cdot 70 + 15 \cdot 90 + 5 \cdot 115 = 64,25$$

(f) Der Median (in cm) entfällt auf die 3.Klasse $]60; 80]$:

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

$$= 60 + (80 - 60) \cdot \frac{0,5 - 0,4}{0,8 - 0,4} = 65.$$

(g) Da die Klassen unterschiedlich breit sind, wird die Dichte der relativen Häufigkeiten aufgetragen. Daten zur grafischen Darstellung:

Länge (cm)	Klasenbreite (in m) Δ_k	rel. Häufigkeit f_k	Dichte $f_k^* = \frac{f_k}{\Delta_k}$
$(20; 40]$	$0,20$	0,15	0,75
$(40; 60]$	$0,20$	0,25	1,25
$(60; 80]$	$0,20$	0,4	2,0
$(80; 100]$	$0,20$	0,15	0,75
$(100; 130]$	$0,30$	0,05	0,167

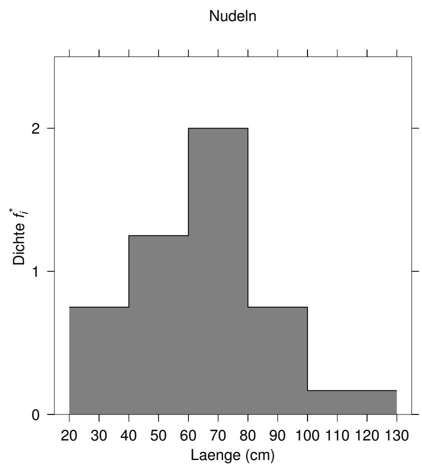


Abbildung 2: Histogramm der Verteilung der Nudellänge

Aufgabe 3

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 $f(x_i|Y = 6,2)$ ist die Zeile mit $Y = 6,2$, also $f(x_i|Y = 6,2) = \{6; 12; 90; 9\}$.

(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 4

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:

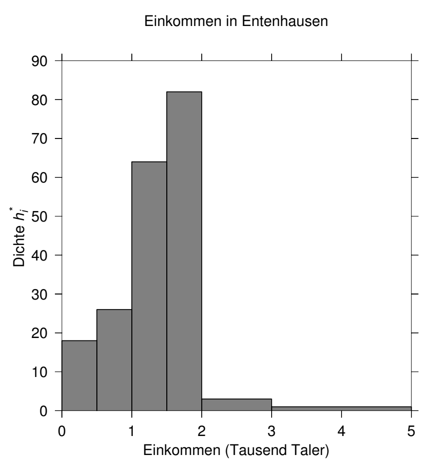


Abbildung 3: Einkommensverteilung in Entenhausen

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:

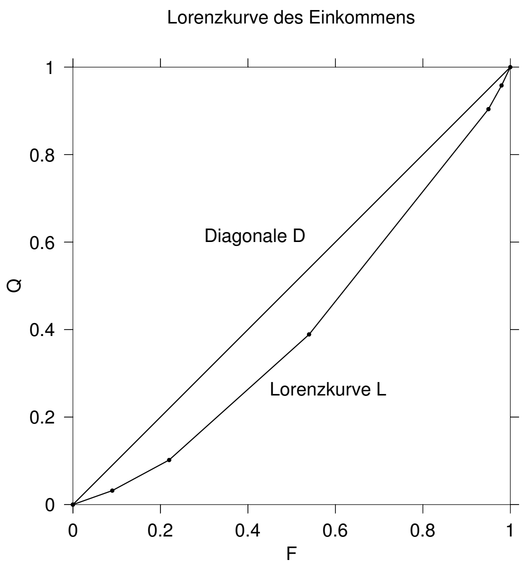


Abbildung 4: Lorenzkurve der Einkommensverteilung in Entenhausen

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 5

1. Harmonisches Mittel:

$$\bar{x}_H = \frac{7}{\frac{1}{15} + \frac{1}{16{,}5} + \frac{1}{17{,}5} + \frac{1}{18} + \frac{1}{18} + \frac{1}{20} + \frac{1}{22}} = 17,9037$$

2. Die Durchschnittsgeschwindigkeit erhält man als Quotienten der gesamten zurückgelegten Strecke und der gesamten benötigten Zeit, also

$$\frac{15km + 16,5km + 17,5km + 18km + 18km + 20km + 22km}{7h}$$

Die Anwendung des arithmetischen Mittels ist hier korrekt.

1. geometrisches Mittel:

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

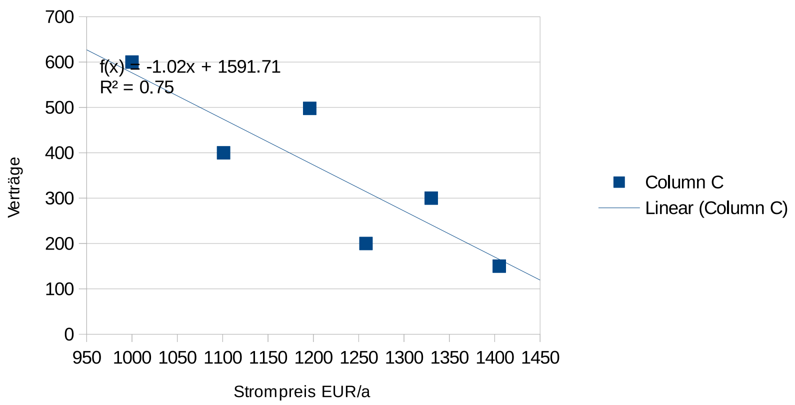
genau das ist ja die Aussage (verdoppelt sich jede Nacht!).

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

3. Eine Stunde 50 km/h, 1 Stunde und 15 Minuten 40 km/h.
   Die Gesamtzeit sind 2 Stunden und 15 Minuten, die zurückgelegte Strecke $s = 1$h$\cdot 50$km/h $+1,25$h$\cdot 40$km/h = 100km.
   Durchschnittsgeschwindigkeit in km/h:
$$\bar{v} = \frac{100}{2,25} = 44,\bar{4}.$$
   Das ist das harmonische Mittel der Geschwindigkeiten (in km/h):
$$\bar{v} = \frac{1}{1/2(\frac{1}{50} + \frac{1}{40})} = 44,\bar{4}$$

Aufgabe 6

a) Daten mit Ausgleichsgerade



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

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

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.