# Musterlösung zur Klausur Statistik

## TIT17 / TIM17

## Oettinger 03.2019

**Zeit: 90Min.**

Insgesamt erreichbare Punktzahl: 105, entspricht 105%.

## Aufgabe 1

(15 Punkte)
Welche der folgenden Aussagen sind richtig?

(a) Richtig, die Varianz ist eine Summe quadratischer Größen.

(b) Falsch, die Abweichung ordinaler Merkmale vom Median ist nicht definiert - also auch keine mittlere Abweichung.

(c) Richtig, die untersuchten Merkmale besitzen keine Lebensdauer.

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

(e) Richtig, der Quantilsabstand steigt mit der Breite der Verteilung an.

## Aufgabe 2

(28 Punkte)

(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!

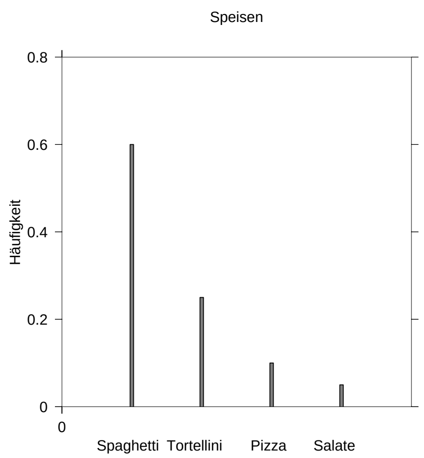



**Abbildung 1:** Stabdiagramm als Beispiel für die grafische Darstellung

(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):

$$\bar{x} = 0,15 \cdot 30 + 0,25 \cdot 50 + 0,40 \cdot 70 + 0,15 \cdot 90 + 0,05 \cdot 115 = 64,25$$

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

| Klasse | Klassenmitte $x_i$ | $f$ | $(x_i - \bar{x})^2$ |
|---|---|---|---|
| $(20; 40]$ | 30 | 0.15 | 175.959375 |
| $(40; 60]$ | 50 | 0.25 | 50.765625 |
| $(60; 80]$ | 70 | 0.4 | 13.225 |
| $(80; 100]$ | 90 | 0.15 | 99.459375 |
| $(100; 130]$115 | 0.05 | 128.778125 | |

**Tabelle 1:** Tabelle

| Länge (cm) | Klasenbreite (in m) $\Delta_k$ | 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 |

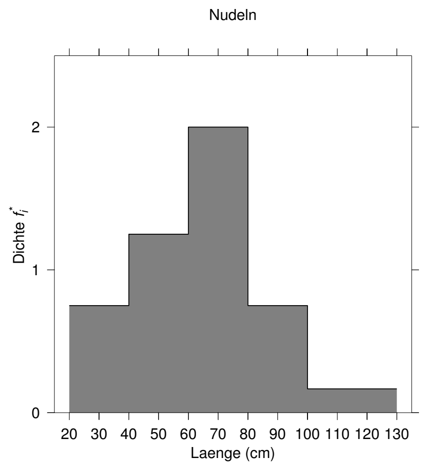



**Abbildung 2:** Histogramm der Verteilung der Nudellängen

(f) Die Varianz $s^2$ kann wie das arithmetische Mittel über die Klassenmitten genähert werden:

$$s^2 \approx \sum_{i=1}^{k} f_i \cdot \left( \frac{x_i^o + x_i^u}{2} - \bar{x} \right)^2 = 468,2\text{cm}^2.$$

Die Standardabweichung ist die positive Quadratwurzel daraus $s = 21,64$ cm.

## Aufgabe 3

(10 Punkte)

| Tag | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| km/h ($x$) | 15 | 16,5 | 17,5 | 18 | 18 | 20 | 22 |

a) Arithmetisches Mittel (in km/h):

$$\bar{x} = \frac{15 + 16,5 + 17,5 + 18 + 18 + 20 + 22}{7} = \frac{127}{7} = 18,1429$$

b) Harmonisches Mittel (in km/h):

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

c) Die Durchschnittsgeschwindigkeit erhält man als Quotient aus der gesamten zurückgelegten Strecke und der dafür benötigten Zeit, also

$$\frac{7 \cdot 10\text{km}}{\frac{10\text{km}}{15\text{km/h}} + \frac{10\text{km}}{16{,}5\text{km/h}} + \frac{10\text{km}}{17{,}5\text{km/h}} + \frac{10\text{km}}{18\text{km/h}} + \frac{10\text{km}}{18\text{km/h}} + \frac{10\text{km}}{20\text{km/h}} + \frac{10\text{km}}{22\text{km/h}}}$$

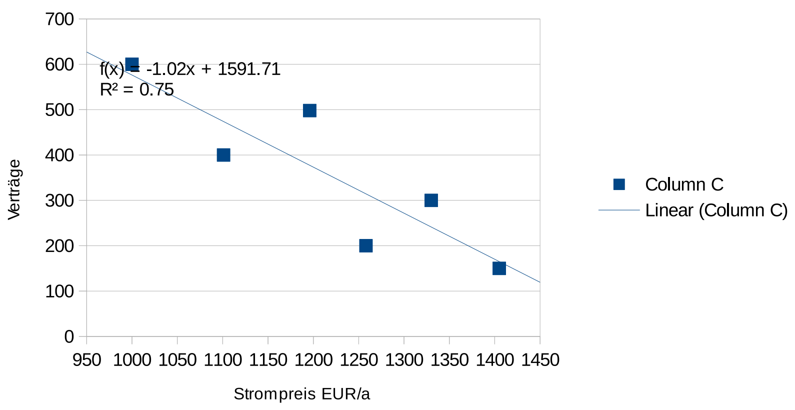
Die Anwendung des harmonischen Mittels ist hier korrekt.

d) Vielleicht!

## Aufgabe 4

(22 Punkte)

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

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

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

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

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

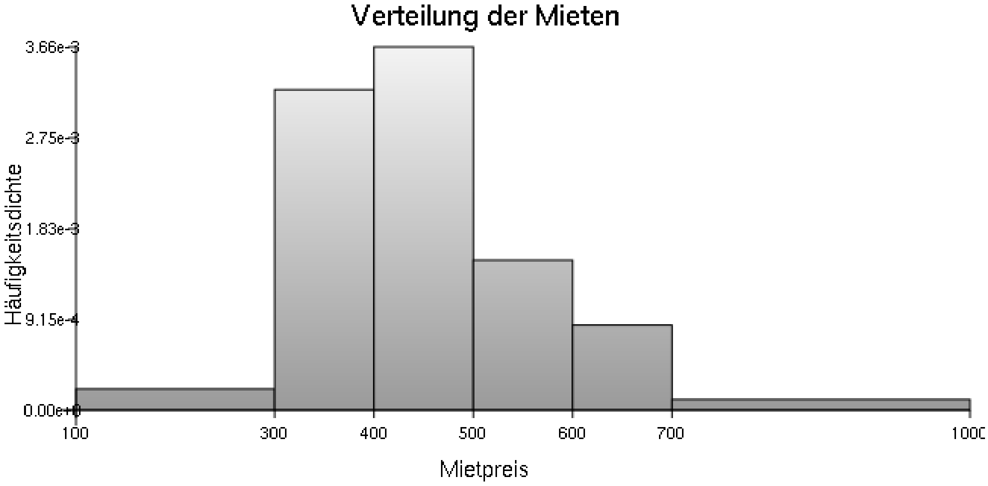
Der Wert nahe eins legt einen relativ guten linearen Zusammenhang zwischen den Daten nahe.

| $x^u$ | $x^o$ | $h$ | $f$ | $F$ | $\Delta$ | $f^*$ |
|---|---|---|---|---|---|---|
| 100 | 300 | 4 | 0.043 | 0.043 | 200 | 0.000215 |
| 300 | 400 | 30 | 0.323 | 0.366 | 100 | 0.00323 |
| 400 | 500 | 34 | 0.366 | 0.732 | 100 | 0.00366 |
| 500 | 600 | 14 | 0.151 | 0.883 | 100 | 0.00151 |
| 600 | 700 | 8 | 0.086 | 0.969 | 100 | 0.00086 |
| 700 | 1000 | 3 | 0.0323 | 1.0013 | 300 | 0.000108 |

## Aufgabe 5

(16 Punkte) Verteilung der Monatsmieten

a) Histogramm: Auftragung der Häufigkeitsdichten $h_k^* = h_k/\Delta_k$ mit der Klassenbreite $\Delta_k = (x_k^o + x_k^u)/2$, die Zahl der Klassen ist $k = 6$, der Stichprobenumfang $n = 93$.




b) arithmetisches Mittel:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} h_i \frac{x_i^u + x_i^o}{2} = 452,2$$

Median genähert in Klasse $3$:

$$\bar{x}_Z = x_3^u + (x_3^o - x_3^u)\frac{0,5 - F(x_3^u)}{F(x_3^o) - F(x_3^u)} = 400 + 100 \cdot \frac{0,5 - 0,366}{0,732 - 0,366} = 436,6.$$

## Aufgabe 6

(14 Punkte)

a) Geometrisches Mittel:
$\bar{x}_G = \sqrt[3]{(1+0,2)\cdot(1+0,15)\cdot(1-0,0005)} - 1 = 9,68\%$

b) Der Mittelwert berechnet sich nach

$$\bar{x} = \frac{1}{30}(10\cdot 1 + 11\cdot 2 + 8\cdot 3 + 4) = \frac{60}{30} = 2,0.$$

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

c) Es wurden insgesamt $n = 11$ Hotels untersucht, auch das Hotel ohne Stern kann mitgezählt werden (es gibt nur die eine Möglichkeit 0). also ist das arithmetische Mittel

$$\bar{x} = \frac{1}{11}(5\cdot 3 + 2\cdot 2 + 3\cdot 1 + 0) = \frac{22}{11} = 2,0.$$

d) Wenn Ted eine mittlere Geschwindigkeit von 60 km/h fahren will, benötigt er für die insgesamt 8 km Weg eine Zeit von 8/60 h = 8 min. Da er aber für den Rückweg von 4km bereits eine Zeit von 4/30 h = 8min einplant, kann er die geplante Durchschnittsgeschwindigkeit nicht erreichen.