

# Mathematik für Ingenieure C4: INF

## 1. Übung

16.04. - 20.04.2018  
Sommersemester 2018

Dr. Wigand Rathmann  
Lukas Pflug  
Department Mathematik  
Universität Erlangen-Nürnberg

---

**Hausaufgabe 4:****(10 Punkte)**

Wie „zuverlässig“ ist der *Old Faithful Geyser* im Yellowstone National Park? In der folgenden Tabelle sind die Zeiten (in Minuten) zwischen den Eruptionen des Geysirs angegeben (beobachtet im August 1985):

80	71	57	80	75	77	60	86	77	56
81	50	89	54	90	73	60	83	65	82
84	54	85	58	79	57	88	68	76	78
74	85	75	65	76	58	91	50	87	48
93	54	86	53	78	52	83	60	87	49
80	60	92	43	89	60	84	69	74	71
108	50	77	57	80	61	82	48	81	73
62	79	54	80	73	81	62	81	71	79
81	74	59	81	66	87	53	80	50	87
51	82	58	81	49	92	50	88	62	93

- Berechnen Sie den Median, die Quartile, die 8%/92%-Quantile, das arithmetische Mittel und die Standardabweichung.
- Zeichnen und beschriften Sie den einfachen Boxplot.
- Zeichnen Sie ein Histogramm, in dem Sie Klassen mit äquidistanter Klassenbreite 6 wählen. Beginnen Sie mit  $z_0 = 42$ .

---

**Lösung.**

Die Ordnungsstatistik lautet hier

43	48	48	49	49	50	50	50	50	50
51	52	53	53	54	54	54	54	56	57
57	57	58	58	58	59	60	60	60	60
60	61	62	62	62	65	65	66	68	69
71	71	71	73	73	73	74	74	74	75
75	76	76	77	77	77	78	78	79	79
79	80	80	80	80	80	80	81	81	81
81	81	81	81	82	82	82	83	83	84
84	85	85	86	86	87	87	87	87	88
88	89	89	90	91	92	92	93	93	108

□

# Lösungen

11

(a) Median:  $\tilde{x} = \frac{1}{2} (X_{\lceil \frac{n}{2} \rceil} + X_{\lfloor \frac{n}{2} \rfloor})$   
 $\uparrow$   
 $n = 100$

$$= \frac{1}{2} (75 + 75) = 75$$

Quartile:  $u_{25\%} = \frac{1}{2} (X_{\lceil \frac{n \cdot 25}{100} \rceil} + X_{\lfloor \frac{n \cdot 25}{100} \rfloor})$   
 $\uparrow$   
 $100 \cdot \frac{25}{100} = 25 \in \mathbb{Z}$

$$= \frac{1}{2} (58 + 59) = 58.5$$

$$u_{75\%} = \frac{1}{2} (X_{\lceil \frac{n \cdot 75}{100} \rceil} + X_{\lfloor \frac{n \cdot 75}{100} \rfloor})$$
  
 $\uparrow$   
 $100 \cdot \frac{75}{100} = 75 \in \mathbb{Z}$

$$= \frac{1}{2} (82 + 82) = 82$$

8% / 92% - Quantile:

$$u_{8\%} = \frac{1}{2} (X_{\lceil \frac{n \cdot 8}{100} \rceil} + X_{\lfloor \frac{n \cdot 8}{100} \rfloor})$$
  
 $\uparrow$   
 $100 \cdot \frac{8}{100} = 8 \in \mathbb{Z}$

$$= \frac{1}{2} (50 + 50) = 50$$

$$u_{92\%} = \frac{1}{2} (X_{\lceil \frac{n \cdot 92}{100} \rceil} + X_{\lfloor \frac{n \cdot 92}{100} \rfloor})$$
  
 $\uparrow$   
 $100 \cdot \frac{92}{100} = 92 \in \mathbb{Z}$

$$= \frac{1}{2} (89 + 89) = 89$$

Arithmetische Mittel:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{100} \sum_{i=1}^{100} x_i$$

$\uparrow$   
 $n=100$

$$= \frac{1}{100} \cdot 7162 = 71.62$$

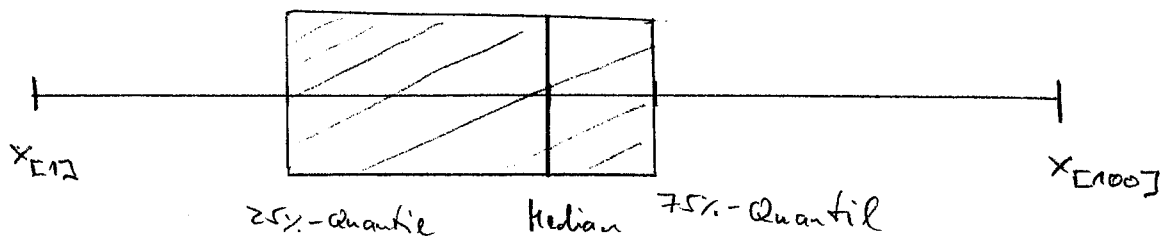
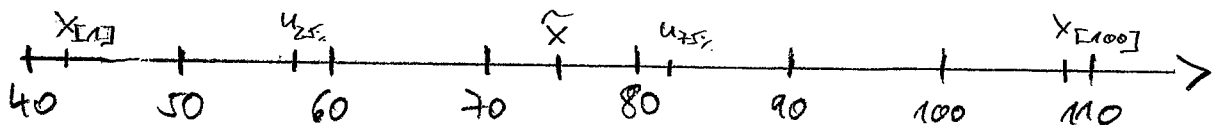
Standardabweichung

$$s_x^{100} = \left( \frac{1}{100} \cdot \sum_{i=1}^{100} x_i^2 - \bar{x}^2 \right)^{1/2}$$

$$\approx 14.15$$

(4)

(b)



25%-Quantil      Median      75%-Quantil

50% der Daten  
liegen in diesem Bereich

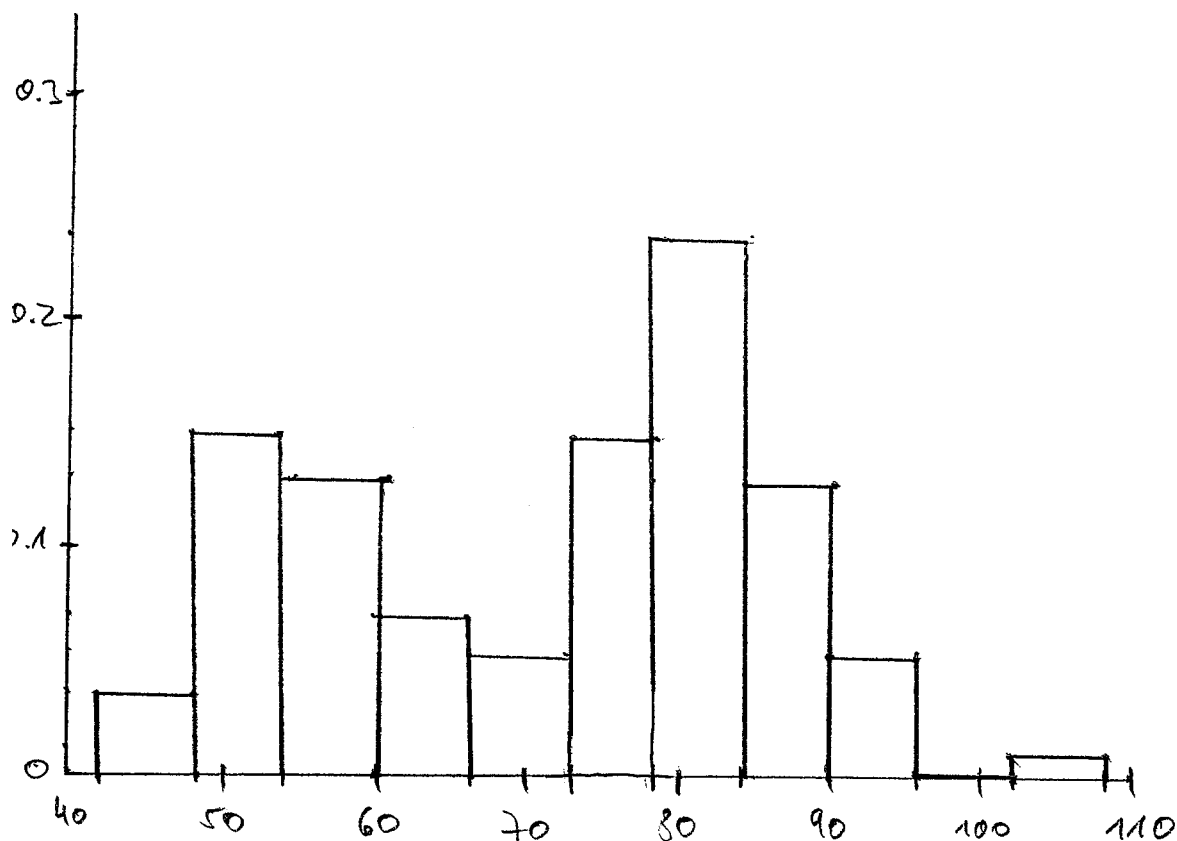
(2)

(c)

K<sub>2</sub>

Klasse	absolute Häufigkeit	rel. Häufigkeit
[42, 48]	3	0.03
(48, 54]	15	0.15
(54, 60]	13	0.13
(60, 66]	7	0.07
(66, 72]	5	0.05
(72, 78]	15	0.15
(78, 84]	23	0.23
(84, 90]	13	0.13
(90, 96]	5	0.05
(96, 102]	0	0.00
(102, 108]	1	0.01
zur Kontrolle	$\Sigma 100$ (✓)	$\Sigma 1$ (✓)

Histogramm (mit rel. Häufigkeiten)



(4)

---

**Hausaufgabe 5:****(10 Punkte)**

- a) Es sei  $\mathbf{x} \in \mathbb{R}^n$  gegeben. Zeigen Sie, dass das arithmetische Mittel die Summe der quadrierten Abweichungen minimiert:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \min_{a \in \mathbb{R}} \sum_{i=1}^n (x_i - a)^2$$

**Bem.** Aufgrund der Monotonie der Wurzelfunktion, folgt dann für die Standardabweichung:

$$\sigma_{\mathbf{x}}^n \leq \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - a)^2} \quad \forall a \in \mathbb{R}$$

- b) Es seien  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ,  $n > 1$ , gegeben. Zeigen Sie, dass die empirische Kovarianz  $s_{\mathbf{xy}}$  die folgende Identität erfüllt:

$$s_{\mathbf{xy}} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)$$

---

**Lösung.**

□

# Lösungen

(a) Für ein gegebenes  $x \in \mathbb{R}^n$  betrachte die Funktion  $f: \mathbb{R} \rightarrow \mathbb{R}$  mit

$$f(a) := \sum_{i=1}^n (x_i - a)^2.$$

Beachte:  $f$  ist eine Funktion in nur einer Variablen.  
L  $f$  ist differenzierbar.

$$\begin{aligned} \Rightarrow f'(a) &= 2 \sum_{i=1}^n (x_i - a) \cdot (-1) \\ &= -2 \sum_{i=1}^n x_i + 2 \sum_{i=1}^n a \\ &= -2n\bar{x} + 2na \end{aligned} \quad (2)$$

Notwendige Optimalitätsbedingung:

$$f'(a) \stackrel{!}{=} 0 \Leftrightarrow 2na = 2n\bar{x}$$

$$\Leftrightarrow a = \bar{x} \quad (2)$$

Hinreichende Optimalitätsbed. (min-Problem):

$$f''(a) \stackrel{!}{>} 0 \Leftrightarrow 2n > 0$$

d.h.  $f''(a) > 0 \forall a \in \mathbb{R}$ , jedoch ist  $a = \bar{x}$  einziger Kandidat für Minimalstelle. (2)

Fazit:  $a = \bar{x}$  ist die Minimalstelle des Problems

$$\min_{a \in \mathbb{R}} f(a)$$



1b) Seien  $x, y \in \mathbb{R}^n$ ,  $n > 1$ , bel. gewählt.

Dann gilt:

$$s_{xy} \stackrel{\text{def.}}{=} \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$= \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y})$$

$$= \frac{1}{n-1} \cdot \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} - n \bar{y} \bar{x} + n \bar{x} \bar{y} \right)$$

$$= \frac{1}{n-1} \cdot \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right) \quad (4)$$

Q



---

**Zusatzaufgabe 6:****(keine Punkte)**

- a) Stellen Sie den Datensatz  $(1, 0), (1, 1), (0, 1), (-1, 0), (-1, -1), (0, -1)$  grafisch dar.
- b) Bestimmen Sie die Regressionsgerade und tragen Sie diese in die Grafik ein. Erscheint die Regressionsgerade plausibel?
- 

**Lösung.**

Die Regressionsgerade lautet

$$g(x) = \frac{s_{\mathbf{xy}}}{s_{\mathbf{x}}^2} x + \left( \bar{y} - \frac{s_{\mathbf{xy}}}{s_{\mathbf{x}}^2} \bar{x} \right).$$

Hier ist:

- Empirische Kovarianz  $s_{\mathbf{xy}} = 0.4$ .
- Empirische Varianz x:  $s_{\mathbf{x}}^2 = 0.8$
- Arithmetisches Mittel y:  $\bar{y} = 0$ .
- Arithmetisches Mittel x:  $\bar{x} = 0$

Es folgt

$$g(x) = \frac{1}{2} x + 0.$$

Weitere interessante Werte (z.B. wenn die Rollen von  $x$  und  $y$  vertauscht werden):

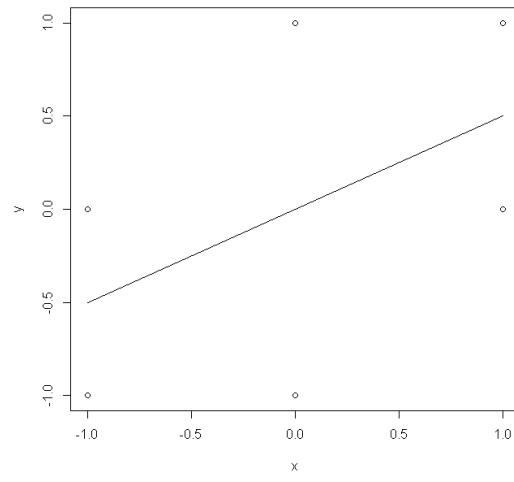
Empirische Streuung x:  $s_{\mathbf{x}} = 0.8944272$

Empirische Varianz y:  $s_{\mathbf{y}}^2 = 0.8$

Empirische Streuung y:  $s_{\mathbf{y}} = 0.8944272$

Empirische Korrelationskoeffizient  $r_{\mathbf{xy}} = 0.559017$

Korrelationskoeffizient für Regressionsgerade  $r_{g(\mathbf{x})\mathbf{y}} = 0.5$



---

**Zusatzaufgabe 7:****(keine Punkte)**

- a) Es seien  $x_1, \dots, x_n > 0$  beliebig. Weiter sei mit  $A(\mathbf{x}) := \bar{x}$  das *arithmetische Mittel*, mit  $G(\mathbf{x})$  das *geometrische Mittel* und mit  $H(\mathbf{x})$  das *harmonische Mittel* der  $x_i$  bezeichnet. Zeigen Sie, dass die Ungleichungskette

$$H(\mathbf{x}) \leq G(\mathbf{x}) \leq A(\mathbf{x})$$

erfüllt ist.

**Hinweis.** Für die Exponentialfunktion  $\exp : \mathbb{R} \rightarrow (0, \infty)$  und  $y_1, \dots, y_n \in \mathbb{R}$  gilt:

$$\exp \left( \sum_{i=1}^n g_i y_i \right) \leq \sum_{i=1}^n g_i \exp(y_i)$$

mit Gewichten  $g_1, \dots, g_n \geq 0$  und  $\sum_{i=1}^n g_i = 1$ .

- b) i) Stellen Sie den Datensatz

(6.2, 4.5), (3.7, 3.6), (8.5, 3.5), (3.5, 5.1), (4.8, 4.6), (6.7, 2.9), (7.8, 4.4),  
(5.4, 2.4), (4.8, 3.7), (9.8, 3.6), (2.5, 4.3), (2.6, 5.8), (6.8, 3.7), (5.2, 3.4),  
(7.4, 4.7)

grafisch dar.

- ii) Bestimmen Sie die Regressionsgerade und tragen Sie diese in die Grafik ein. Erscheint die Regressionsgerade plausibel?
- 

**Lösung.**

- a) Es seien  $x_1, \dots, x_n > 0$  beliebig gewählt.

*Schritt 1.* Zu zeigen:  $G(\mathbf{x}) \leq A(\mathbf{x})$ . Dazu:

$$G(\mathbf{x}) = \sqrt[n]{\prod_{i=1}^n x_i} = \exp \left( \ln \left( \sqrt[n]{\prod_{i=1}^n x_i} \right) \right) = \exp \left( \sum_{i=1}^n \frac{1}{n} \ln(x_i) \right) \leq \sum_{i=1}^n \frac{1}{n} \exp(\ln(x_i)) = \frac{1}{n} \sum_{i=1}^n x_i = A(\mathbf{x})$$

*Schritt 2.* Zu zeigen:  $H(\mathbf{x}) \leq G(\mathbf{x})$ . Setze dazu  $z_i := \frac{1}{x_i}$  für  $i = 1, \dots, n$ .

$$G(\mathbf{x}) \stackrel{\text{Schritt 1.}}{\leq} A(\mathbf{z}) \stackrel{z_i = x_i^{-1}}{\iff} \sqrt[n]{\prod_{i=1}^n \frac{1}{x_i}} \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \iff \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} \leq \frac{1}{\sqrt[n]{\prod_{i=1}^n \frac{1}{x_i}}} \iff H(\mathbf{x}) \leq G(\mathbf{x})$$

- b) Die Regressionsgerade lautet

$$g(x) = \frac{s_{\mathbf{xy}}}{s_{\mathbf{x}}^2} x + \left( \bar{y} - \frac{s_{\mathbf{xy}}}{s_{\mathbf{x}}^2} \bar{x} \right).$$

Hier ist:

- Empirische Kovarianz  $s_{\mathbf{xy}} = 0.7030476$ .
- Empirische Varianz  $\mathbf{x}$ :  $s_{\mathbf{x}}^2 = 4.646952$ .
- Arithmetisches Mittel  $\mathbf{x}$ :  $\bar{x} = 5.713333$ .

- Arithmetisches Mittel  $y$ :  $\bar{y} = 4.013333$ .

Es folgt

$$g(x) = -0.1512922x + 4.877716.$$

Weitere interessante Werte (z.B. wenn man die Rollen von  $x$  und  $y$  vertauscht werden):

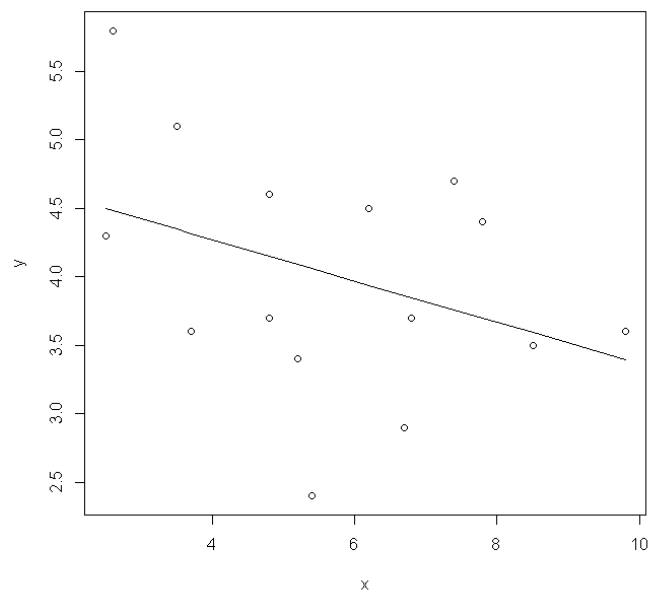
Empirische Streuung  $x$ :  $s_x = 2.155679$

Empirische Varianz  $y$ :  $s_y^2 = 0.7626667$

Empirische Streuung  $y$ :  $s_y = 0.8733079$

Empirische Korrelationskoeffizient  $r_{xy} = -0.4276277$

Korrelationskoeffizient für Regressionsgerade  $r_{g(x)y} = 0.3734507$



□