
Besprechung der Lösungen: 10./11. März 2020 in den Übungsstunden

Aufgabe 1

- (a) Der Ticketpreis für einen Flug von Basel nach London hat in den letzten drei Jahren zweimal aufgeschlagen, das erste Mal um 2,5 % und das zweite Mal um 3,2 %. Um wieviele Prozent hat sich der Ticketpreis insgesamt erhöht?
- (b) In den USA leben 4 % der Weltbevölkerung. In den USA werden aber 40 % der weltweit produzierten Pharmazeutika konsumiert. In welchem Verhältnis steht der Pharmakonsum eines Einwohners der USA zum Pharmakonsum eines anderen Weltenbürgers?

Aufgabe 2

- (a) Skizzieren Sie das Streudiagramm und bestimmen Sie den Korrelationskoeffizienten r_{xy} für die folgenden Wertepaare (x_i, y_i) (für (i) von Hand, für (ii) und (iii) mit Excel o.Ä.).
- (i) (2, 2), (3, 1), (5, 3), (6, 4)
- (ii) (1, 5), (2, 3), (3, 4), (5, 2)
- (iii) (1, 5), (2, 2), (3, 1), (4, 2), (5, 5)
- (b) Berechnen Sie die Regressionsgerade für die Wertepaare (i).
- (c) Vertauschen Sie x_i und y_i in (i), d.h. betrachten Sie die Wertepaare (2, 2), (1, 3), (3, 5), (4, 6). Ändert sich der Korrelationskoeffizient? Ändert sich die Regressionsgerade? Erklären Sie, warum ja, bzw. nein.

Aufgabe 3

Gegeben seien die x -Werte $x_1 = -2$, $x_2 = -1$, $x_3 = 0$, $x_4 = 1$ und $x_5 = 2$

- (a) Bestimmen Sie je zwei Reihen von y -Werten $y_1, \dots, y_5$ so, dass der Korrelationskoeffizient der Wertepaare $(x_1, y_1), \dots, (x_5, y_5)$ gleich
- (i) +1 (ii) -1 bzw. (iii) 0 ist.
- (b) Die y -Werte seien durch die Funktion $y_i = x_i^2 + 1$ definiert. Es besteht also ein funktionaler Zusammenhang. Berechnen Sie den Korrelationskoeffizienten für die Wertepaare $(x_1, y_1), \dots, (x_5, y_5)$.

Aufgabe 4

Die folgende Tabelle zeigt die Prüfungsergebnisse von zehn Pharmaziestudenten in den Fächern Mathematik und Physik. Dabei ist x_i Punktzahl in der Mathematikprüfung von Student i und y_i Punktzahl in der Physikprüfung von Student i . Die maximal erreichbare Punktzahl war jeweils 50.

x_i	44	45	29	35	50	8	45	41	32	38
y_i	49	28	9	26	32	20	28	26	23	28

Ordnen Sie den Punktzahlen Ränge zu und bestimmen Sie den Rangkorrelationskoeffizienten.

Zusatzaufgaben

Aufgabe 5

Gegeben sind die folgenden Wertepaare (x_i, y_i) :

x_i	1	2	2	3	3	4	4	5
y_i	2	1	3	2	4	3	5	4

- (a) Skizzieren Sie das Streudiagramm.
- (b) Geben Sie eine Vermutung für die Gleichung der Regressionsgeraden an und überprüfen Sie sie rechnerisch (von Hand oder mit Excel o.Ä.).

Aufgabe 6

Der Luftdruck nimmt mit zunehmender Höhe über dem Meeresspiegel ab. Auf Grund theoretischer Überlegungen weiss man, dass bei Voraussetzung einer konstanten Temperatur bis in die Höhe von ca. 100 km zwischen der Höhe h (in km) und dem Luftdruck p (in Hektopascal hPa) die *barometrische Höhenformel*

$$p = c e^{-ah}$$

mit Konstanten $a > 0$ und $c > 0$ gilt.

Sechs Messungen haben die folgenden Werte ergeben:

h_i	0	1	2	3	4	5
p_i	1013	899	795	701	616	540

- (a) Bestimmen Sie mit Hilfe einer linearen Regression die Konstanten a und c (so dass die barometrische Höhenformel die Messungen gut approximiert).
- (b) Wie gross ungefähr ist der Luftdruck in 6 km Höhe? In welcher Höhe beträgt der Luftdruck nur noch etwa 200hPa?

Aufgabe 7

- (a) Zwei Personen müssen 4 Nahrungsmitteln, nämlich Schokolade, Pfannkuchen, Rösti und Fenchel, gemäss ihrer Beliebtheit Ränge zuordnen. Person A verteilt die Ränge in der Reihenfolge 1, 2, 3, 4; Person B verteilt die Ränge genau in umgekehrter Reihenfolge 4, 3, 2, 1. Berechnen Sie den Rangkorrelationskoeffizienten r_S .
- (b) Zwei Personen müssen n Nahrungsmitteln gemäss ihrer Beliebtheit Ränge zuordnen. Person A verteilt die Ränge in der Reihenfolge $1, 2, \dots, n$; Person B verteilt die Ränge genau in umgekehrter Reihenfolge $n, n-1, \dots, 1$. Berechnen Sie den Rangkorrelationskoeffizienten r_S .

Aufgabe 8

Beweisen Sie die Gleichung für die Kovarianz

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

Lösungshinweise

Aufgabe 1

- (a) Analog zum 3. Beispiel auf Seite 16 des Skripts.
- (b) Am einfachsten denkt man sich eine Modellwelt mit zum Beispiel 100 Einwohnern und 100 Pillen. Da es nur um Verhältnisse geht, spielt die Wahl der Einwohner- und Pillenzahlen keine Rolle.

Aufgabe 2

Analog zum 1. Beispiel auf den Seiten 20–21 und 27 des Skripts.

Aufgabe 3

- (a) Für $r_{xy} = \pm 1$ mit Hilfe von Satz 2.1. Wie gross muss die Steigung der Regressionsgeraden sein, damit $r_{xy} = 0$?
- (b) Ist der Zusammenhang linear?

Aufgabe 4

Analog zum Beispiel auf Seite 23 des Skripts.

Aufgabe 5 (Zusatzaufgabe)

Analog zu den Beispielen auf den Seiten 27–28 des Skripts.

Aufgabe 6 (Zusatzaufgabe)

- (a) Analog zum Beispiel auf den Seiten 28–29 des Skripts.
- (b) Einsetzen in die unter (a) gefundene Gleichung.

Aufgabe 7 (Zusatzaufgabe)

- (b) Zum i -ten Rang der Person A gehört der $(n-i+1)$ -te Rang der Person B. Zur Umformung der Summe der Rangdifferenzen im Quadrat sind die folgenden Summen hilfreich:

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{i=1}^n i = \frac{n(n+1)}{2}$$

Aufgabe 8 (Zusatzaufgabe)

Analog zur Herleitung der Gleichung für die Varianz (Satz 1.2) auf Seite 14 des Skripts.

Ergebnisse

Aufgabe 1

- (a) Um 5,78 % (nicht um $5,7\% = 2,5\% + 3,2\%$!)

Denn ist x_0 der ursprüngliche Ticketpreis, dann ist der Ticketpreis nach den beiden Erhöhungen gegeben durch $1,032 \cdot 1,025 \cdot x_0 = 1,0578x_0$.

- (b) Im Verhältnis 16 : 1.

Rechnung mit Modellwelt (100 Einwohner, 100 Pillen):

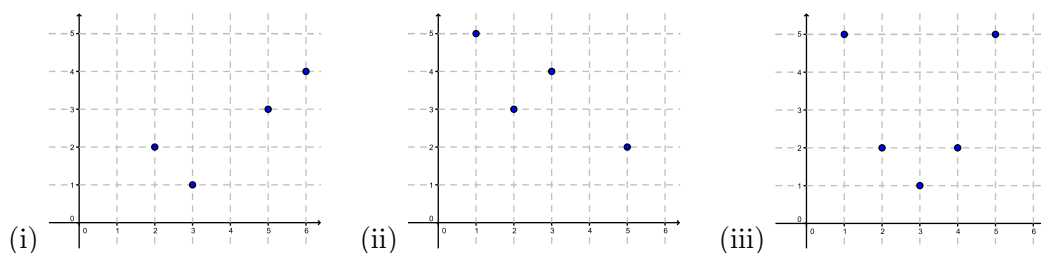
In den USA leben 4 Personen, die 40 Pillen konsumieren, also 10 pro Einwohner. In der restlichen Welt leben 96 Einwohner, die 60 Pillen konsumieren, also $\frac{60}{96} = 0,625$ pro Einwohner. Das Verhältnis ist demnach $\frac{10}{0,625} = 16$.

Rechnung ohne Modellwelt:

Seien e die Anzahl Einwohner und p die Anzahl Pillen. Dann konsumieren in den USA $0,04e$ Einwohner $0,4p$ Pillen, also $\frac{0,4p}{0,04e} = 10\frac{p}{e}$ Pillen pro Einwohner. In der restlichen Welt werden $\frac{0,6p}{0,96e} = 0,625\frac{p}{e}$ Pillen pro Einwohner konsumiert. Wir erhalten das Verhältnis $10\frac{p}{e} : 0,625\frac{p}{e} = 16$.

Aufgabe 2

- (a)



- (i) $r_{xy} = 0,849$ (ii) $r_{xy} = -0,832$ (iii) $r_{xy} = 0$

- (b) $y = \frac{3}{5}x + \frac{1}{10}$

- (c) r_{xy} bleibt unverändert, da die Formel für r_{xy} unverändert bleibt bei einer Vertauschung der x_i mit den y_i .

Das Streudiagramm für die vertauschten x - und y -Werte erhält man durch Spiegeln des ursprünglichen Streudiagramms an der Geraden $y = x$. Berechnet man jedoch die Regressionsgerade für die vertauschten Daten, so erhält man die Gerade $y = \frac{6}{5}x + 1$. Dies ist nicht die ursprüngliche an der Geraden $y = x$ gespiegelte Regressionsgerade (diese hat nämlich die Gleichung $y = \frac{5}{3}x - \frac{1}{6}$). Das ist auch so zu erwarten, da mit den ursprünglichen Daten die Abweichungen der y_i minimiert werden, nach Vertauschen jedoch die Abweichungen der x_i .

Aufgabe 3

- (a) (i) Zum Beispiel $y_i = x_i$ oder $y_i = 2x_i + 1$ für $i = 1, \dots, 5$.

Allgemein gilt $r_{xy} = +1 \iff y_i = ax_i + b$ mit $a > 0$

- (ii) Zum Beispiel $y_i = -x_i$ oder $y_i = -2x_i + 1$ für $i = 1, \dots, 5$.

Allgemein gilt $r_{xy} = -1 \iff y_i = ax_i + b$ mit $a < 0$

(iii) Zum Beispiel $y_1 = y_3 = y_5 = -1$, $y_2 = y_4 = 1$ oder $y_1 = y_3 = y_5 = 2$, $y_2 = y_4 = 0$
 Man muss $y_1, \dots, y_5$ so wählen, dass die Regressionsgerade parallel zur x -Achse verläuft, also Steigung Null hat (denn für x_i nicht alle gleich und y_i nicht alle gleich ist $r_{xy} = 0 \Leftrightarrow c_{xy} = 0 \Leftrightarrow$ (Steigung der Regressionsgerade) $= 0$)

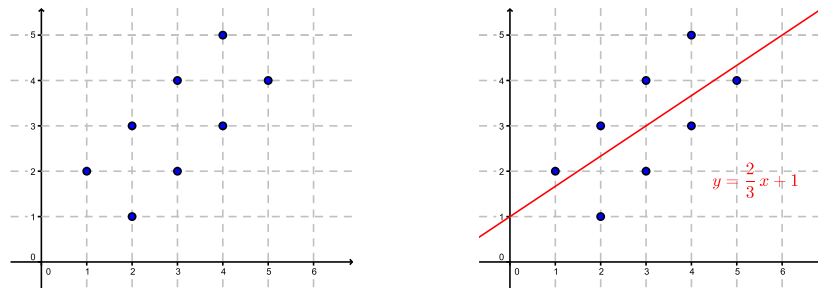
- (b) $r_{xy} = 0$
 (Der Korrelationskoeffizient gibt nur Auskunft über einen *linearen* Zusammenhang!)

Aufgabe 4

$$r_s = 1 - \frac{6 \cdot 23}{10 \cdot (100 - 1)} = 0,861$$

Aufgabe 5 (Zusatzaufgabe)

(a)



- (b) Die Punktwolke ist spiegelsymmetrisch zur Geraden $y = x$. Die Regressionsgerade ist jedoch nicht diese Gerade, sondern die Gerade $y = \frac{2}{3}x + 1$. Dies kommt daher, da die Formel zur Berechnung der Regressionsgeraden (d.h. die Steigung $a = \frac{c_{xy}}{s_x^2}$) nicht symmetrisch in x und y ist.

Aufgabe 6 (Zusatzaufgabe)

(a) Logarithmieren der barometrischen Höhenformel ergibt die Gleichung

$$\ln(p) = -ah + \ln(c) .$$

Es besteht also ein linearer Zusammenhang zwischen den Werten h_i und $\ln(p_i)$.
 Für die Messwertpaare

h_i	0	1	2	3	4	5
$\ln(p_i)$	6,9207	6,8013	6,6783	6,5525	6,4232	6,2916

erhalten wir die Regressionsgerade

$$\ln(p) = -0,1259 h + 6,9260 .$$

Damit gilt $a = 0,1259$ und $c = e^{6,9260} = 1018,4$.

- (b) $p = c e^{-ah} \approx 478,5 \text{ hPa}$ und $h = \frac{1}{a} \ln\left(\frac{c}{p}\right) \approx 12,9 \text{ km}$

Aufgabe 7 (Zusatzaufgabe)

(a) $r_S = 1 - \frac{6}{4 \cdot 15} \cdot 20 = -1.$

- (b) Wie im Hinweis bemerkt, gehört zum i -ten Rang der Person A der $(n - i + 1)$ -te Rang der Person B. Die Rangdifferenz ist damit $d_i = i - (n - i + 1) = 2i - (n + 1)$. Es folgt

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \sum_{i=1}^n (2i - (n + 1))^2 = \sum_{i=1}^n (4i^2 - 4i(n + 1) + (n + 1)^2) \\ &= 4 \sum_{i=1}^n i^2 - 4(n + 1) \sum_{i=1}^n i + (n + 1)^2 \sum_{i=1}^n 1. \end{aligned}$$

Durch Einsetzen der im Hinweis gegebenen Summen erhalten wir:

$$\begin{aligned} \sum_{i=1}^n d_i^2 &= \frac{2}{3}n(n + 1)(2n + 1) - 2n(n + 1)^2 + n(n + 1)^2 = \frac{1}{3}n(n^2 - 1) \\ \Rightarrow r_S &= 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2 = 1 - \frac{6}{n(n^2 - 1)} \cdot \frac{n(n^2 - 1)}{3} = -1 \end{aligned}$$

Aufgabe 8 (Zusatzaufgabe)

Durch Ersetzen der Summen $\sum_{i=1}^n x_i = n\bar{x}$ und $\sum_{i=1}^n y_i = n\bar{y}$ erhalten wir

$$\begin{aligned} (n - 1) c_{xy} &= \sum_{i=1}^n (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x} \bar{y}) \\ &= \sum_{i=1}^n x_i y_i - \bar{y} \sum_{i=1}^n x_i - \bar{x} \sum_{i=1}^n y_i + \bar{x} \bar{y} \sum_{i=1}^n 1 \\ &= \sum_{i=1}^n x_i y_i - \bar{y} n \bar{x} - \bar{x} n \bar{y} + \bar{x} \bar{y} n \\ &= \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}. \end{aligned}$$

Division durch $n - 1$ ergibt die gewünschte Gleichung.