Beleg 1

Beschreibende Statistik
Empirische Methoden für Informatiker

Nico Schramm
23INM-TZ

17. November 2024

Inhaltsverzeichnis

Aufgabe BI-1

Die Pizzeria Morte Dolce (des Eigentümers M.A. Fia) hat zwei Lokale (kurz L1 und L2 genannt), bei denen man Mittag- und Abendessen (kurz M und A) einnehmen kann, wobei es jedoch bei den Gerichten jeweils nur grob die Unterteilung zwischen Pizza, Spaghetti, Ravioli und Cannelloni gibt. Im letzten Monat (September) wurden in jener Pizzeria wie folgt Gericht bestellt, aufgetischt und verspeist:

	L1		**L2**		
	M	A	M	A	insgesamt
Pizza	400	600	600	800	2400
Sonstige	700	1100	400	400	2600
Summe	1100	1700	1000	1200	5000

(a) Wie viele Merkmale werden in dieser Tabelle dargestellt, wie heißen diese und welche Merkmalsausprägungen werden hierbei jeweils berücksichtigt?

In der Tabelle werden die folgenden drei Merkmale (inkl. Merkmalsausprägung) dargestellt:

Merkmal	Merkmalsausprägungen
Lokal	L1, L2
Essenszeit	Mittagessen (M), Abendessen (A)
Gericht	Pizza, Sonstige

(b) Was (Grundgesamtheit, statistische Einheit, Merkmal usw.) stellt im Falle dieser Datenerhebung jeweils das folgende dar:
- (b_1) die Angabe L2?
- (b_2) Herr M. Angione, der am 1. September mittags in L1 Cannelloni gegessen hat?
- (b_3) die Zahl 5000?
- (b_4) die 2400 Leute, denen eine Pizza aufgetischt wurde?

(b_1) Merkmalsausprägung (bzw. Beobachtungswert)

(b_2) Merkmalsträger (bzw. statistische Einheit)

(b_3) Stichprobenumfang

(b_4) Teilgesamtheit

Aufgabe BI-2

> Für die Bevölkerung gewisser Regionen der Erde wurden folgende Geschlechterverteilungen (das heißt Männer 100 Frauen)
>
> 96 101 98 96 101 98 105 106 101 104 88 97
> 100 96 101 92 98 104 102 97 98 93 100 94
>
> durch statistische Erhebungen erhalten.
>
> (a) Stellen Sie die Daten in einem Histogramm mit 5 Klassen gleicher Klassenbreite dar.

geordnete Urliste: 88 92 93 94 96 96 96 97 97 98 98 98 98 100 100 101 101 101 101 102 104 104 105 106

Stichprobenumfang $n = 24$

$\sqrt{n} = \sqrt{24} \approx 4.8990 \Rightarrow \ell = 5$ Klassen (nach Faustregel)

Einteilung der Klassen mit Breite $d_j = 5$ wie folgt:

j	1	2	3	4	5
Klasse	$[85, 90)$	$[90, 95)$	$[95, 100)$	$[100, 105)$	$[105, 110)$
absolute Häufigkeit h_j	1	3	9	9	2
relative Häufigkeit r_j	$0.041\overline{6}$	0.125	0.375	0.375	$0.08\overline{3}$

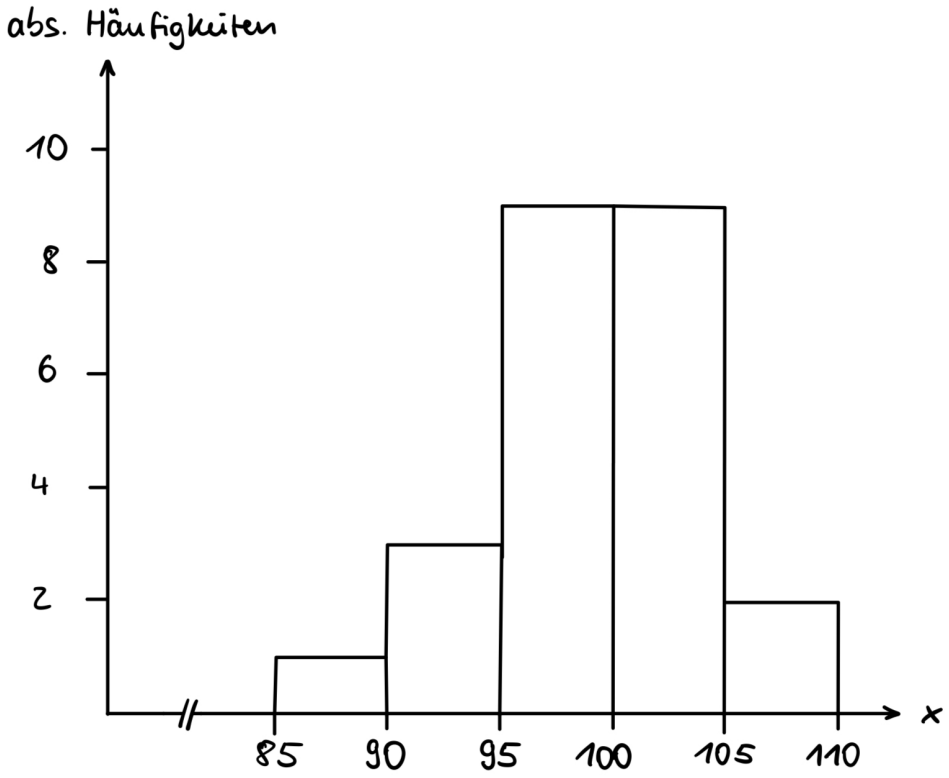


Abbildung 1: Histogramm für Aufgabe 2.a (handgezeichnet)

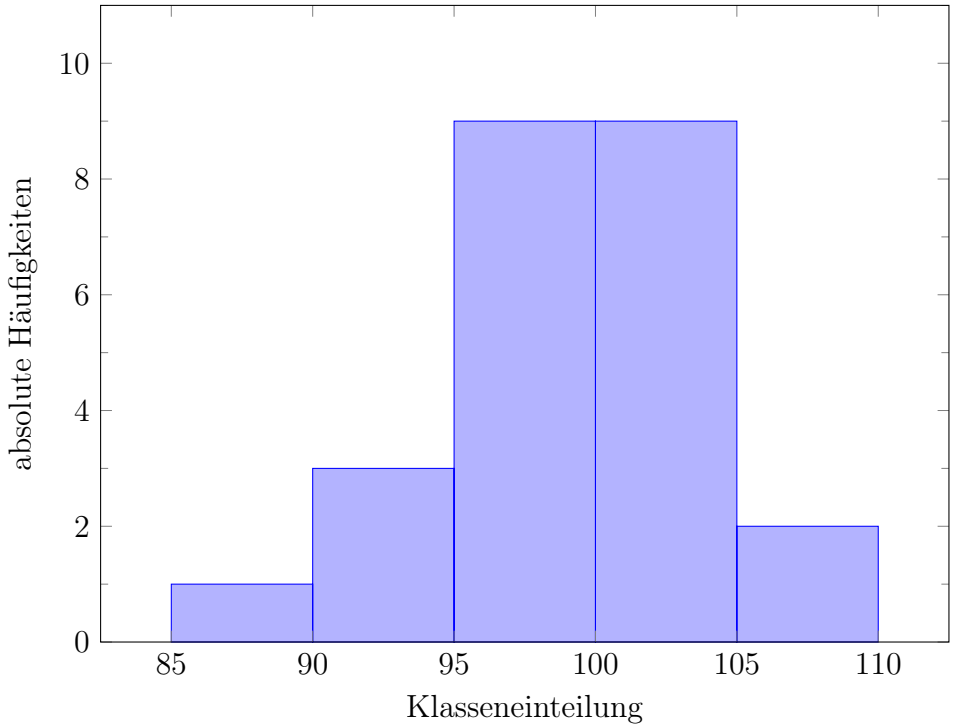


Abbildung 2: Histogramm für Aufgabe 2.a (PGF/TikZ)

(b) Berechnen Sie die Werte der (gewöhnlichen, nicht klassierten) empirischen Verteilungsfunktion F der Daten an den Stellen $x_1 = 95.5$ und $x_2 = 100$.

j	a_j	h_j	r_j	$F(x_j)$
1	88	1	$0.041\overline{6}$	$0.041\overline{6}$
2	92	1	$0.041\overline{6}$	$0.08\overline{3}$
3	93	1	$0.041\overline{6}$	0.125
4	94	1	$0.041\overline{6}$	$0.1\overline{6}$
5	96	3	0.125	$0.291\overline{6}$
6	97	2	$0.08\overline{3}$	0.375
7	98	4	$0.1\overline{6}$	$0.541\overline{6}$
8	100	2	$0.08\overline{3}$	0.625
9	101	4	$0.1\overline{6}$	$0.791\overline{6}$
⋮	⋮	⋮	⋮	⋮

$$F(x) = \begin{cases} 0 & \text{für } x < 88 \\ \vdots & \\ 0.1\overline{6} & \text{für } 94 \leq x < 96 \\ \vdots & \\ 0.625 & \text{für } 100 \leq x < 101 \\ \vdots & \\ 1 & \text{für } x \leq 106 \end{cases}$$

$\Rightarrow F(x_1) = F(95.5) = 0.1\overline{6}, \quad F(x_2) = F(100) = 0.625$

Die empirische Verteilungsfunktion nimmt an der Stelle $x_1 = 95.5$ den Wert $0.1\overline{6}$ und an der Stelle $x_2 = 100$ den Wert 0.625 an.

> (c) Erstellen Sie den zu den Beobachtungswerten gehörigen klassischen Box-Plot (mit Kennzeichnung von Ausreißern und Extremwerten, so vorhanden).

Fünf-Punkte-Zusammenfassung

$x_{(1)} = 88$

$q \cdot n = 0.25 \cdot 24 = 6 \in \mathbb{Z} \Rightarrow$ nicht eindeutig
$\Rightarrow \tilde{x}_{0.25} = \frac{1}{2}(x_{(6)} + x_{(7)}) = \frac{1}{2}(96 + 96) = 96$

$n = 24 \notin 2\mathbb{N}$
$\Rightarrow x_{\text{med}} = \frac{1}{2}(x_{(12)} + x_{(13)}) = \frac{1}{2}(98 + 98) = 98$

$q \cdot n = 0.75 \cdot 24 = 18 \in \mathbb{Z} \Rightarrow$ nicht eindeutig
$\Rightarrow \tilde{x}_{0.75} = \frac{1}{2}(x_{(18)} + x_{(19)}) = \frac{1}{2}(101 + 101) = 101$

$x_{(n)} = x_{(24)} = 106$

$\Rightarrow$ Quartilsabstand $d_Q = \tilde{x}_{0.75} - \tilde{x}_{0.25} = 101 - 96 = 5$

$\frac{3}{2} \cdot d_Q = \frac{3}{2} \cdot 5 = 7.5 \Rightarrow [88.5, 108.5]$
$\Rightarrow$ Ausreißer: $\{88\}$

$3 \cdot d_Q = 3 \cdot 5 = 15 \Rightarrow [81, 116]$
$\Rightarrow$ keine Extremwerte

linker Whisker: $\min\{x_j : x_j \geq \tilde{x}_{0.25} - \frac{3}{2}d_Q\} = 92$

rechter Whisker: $\max\{x_j : x_j \leq \tilde{x}_{0.75} + \frac{3}{2}d_Q\} = 106$

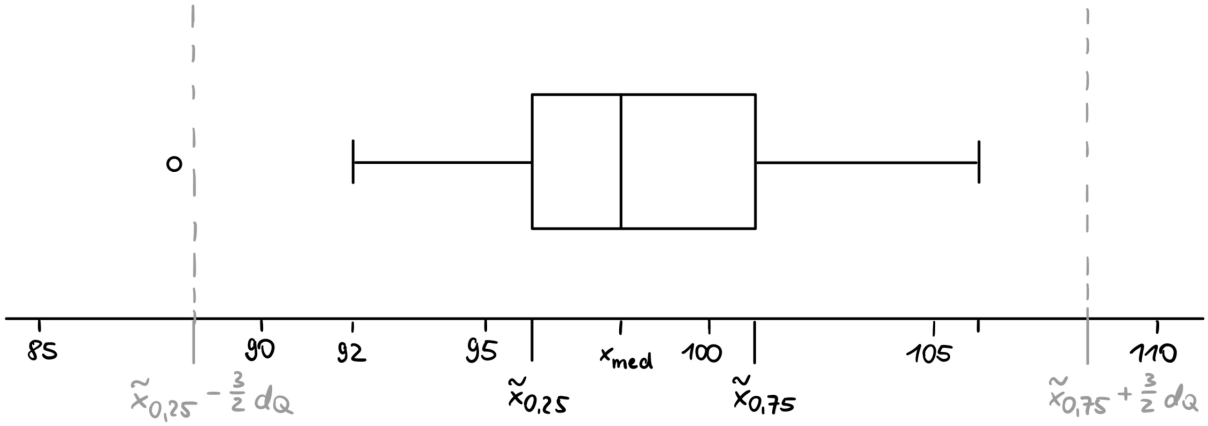


Abbildung 3: Klassischer Box-Plot zu Aufgabe 2.c (handgezeichnet)

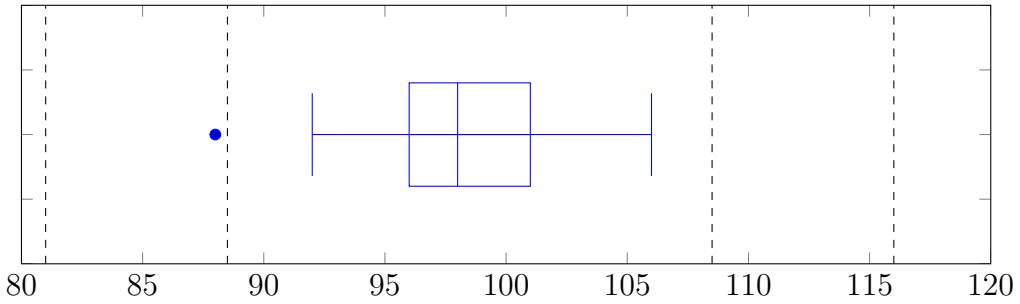
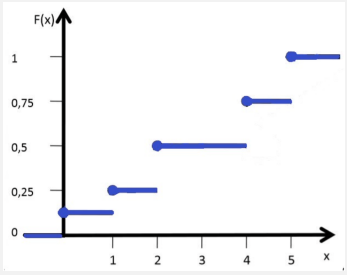


Abbildung 4: Klassischer Box-Plot zu Aufgabe 2.c (PGF/TikZ)

Aufgabe BI-3

Die folgende Graphik zeigt für $n = 200$ Beobachtungen eines Merkmals X die empirische Verteilungsfunktion:



(a) Welche Merkmalsausprägungen (verschieden, positiv) wurden für X beobachtet?

Aus der Graphik lassen sich die folgenden Sprünge ablesen, welche die beobachteten Merkmalsausprägungen von X darstellen: $\{0, 1, 2, 4, 5\}$.

(b) Bestimmen Sie die absoluten Häufigkeiten der Merkmalsausprägungen von X.

$$\forall j \in \{1, 2, \ldots, \ell\} : r_j = \begin{cases} F(x_j) & \text{für } j = 0 \\ F(x_j) - F(x_{j-1}) & \text{für } j > 0 \end{cases}$$

$$h_j = n \cdot r_j$$

j	a_j	$F(x_j)$	r_j	h_j
1	0	0.125	0.125	25
2	1	0.25	0.125	25
3	2	0.5	0.25	50
4	4	0.75	0.25	50
5	5	1	0.25	50

Die absoluten Häufigkeiten h_j können aus der oben dargestellten Tabelle abgelesen werden.

(c) Berechnen Sie das arithmetische Mittel $\overline{x}$ sowie die (korrigierte) Varianz s^2 der Daten.

$$\begin{aligned}\overline{x} &= \sum_{j=1}^{m} a_j \cdot r_j \\ &= 0 \cdot 0.125 + 1 \cdot 0.125 + 2 \cdot 0.25 + 4 \cdot 0.25 + 5 \cdot 0.25 \\ \Rightarrow \overline{x} &= 2.875\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{1}{n-1}\sum_{j=1}^{m}(x_j - \overline{x})^2 \\ &= \frac{1}{200-1}\big((0-2.875)^2 + (1-2.875)^2 + (2-2.875)^2 \\ &\qquad + (4-2.875)^2 + (5-2.875)^2\big) \\ &\approx \frac{1}{199} \cdot 18.3281 \\ \Rightarrow s^2 &\approx 0.0921\end{aligned}$$

Insgesamt ergibt sich ein arithmetisches Mittel von $\overline{x} = 2.875$ und eine (korrigierte) Varianz von $s^2 \approx 0.0921$.

(d) Es wird eine Stichprobe mit zehn weiteren Beobachtungen erhoben. Alle zehn Beobachtungen haben den Wert 3. Wie lauten dann die neuen relativen Häufigkeiten der Merkmalsausprägungen von X für die um jene Beobachtungswerte erweiterte Stichprobe?

neuer Stichprobenumfang $n' = 200 + 10 = 210$

j	a'_j	h'_j	r'_j
1	0	25	$25/210 \approx 0.1190$
2	1	25	$25/210 \approx 0.1190$
3	2	50	$50/210 \approx 0.2381$
4	3	10	$10/210 \approx 0.0476$
5	4	50	$50/210 \approx 0.2381$
6	5	50	$50/210 \approx 0.2381$

Die neuen relativen Häufigkeiten r'_j lassen sich aus der Tabelle ablesen.

Aufgabe BI-4

> In der Verwaltung einer bestimmten Hochschule sind 400 Personen beschäftigt. Jede Person ist entweder Arbeiter/in, angestellt oder beamtet. Die Aufteilung in Abhängigkeit vom Geschlecht ist in der Kontingenztabelle
>
> | | Arbeiter/in | angestellt | beamtet |
> |---|---|---|---|
> | weiblich | 5 | 160 | 42 |
> | männlich | 36 | 122 | 35 |
>
> zusammengestellt. Bestimmen Sie für jene Daten das Kontingenzmaß V nach Cramér und interpretieren Sie jenen Werte.

Vervollständigen zur folgenden Kontingenztabelle:

	Arbeiter/in	angestellt	beamtet	gesamt
weiblich	5	160	42	207
männlich	36	122	35	193
gesamt	41	282	77	400

Daraus Bestimmung von Zwischenschritten $\frac{1}{n}h_{j\bullet}h_{\bullet k}$

$\frac{1}{n}h_{j\bullet}h_{\bullet k}$	Arbeiter/in	angestellt	beamtet
weiblich	$\frac{41\cdot 207}{400} = 21.2175$	$\frac{282\cdot 207}{400} = 145.935$	$\frac{77\cdot 207}{400} = 39.8475$
männlich	$\frac{41\cdot 193}{400} = 19.7825$	$\frac{282\cdot 193}{400} = 136.065$	$\frac{77\cdot 193}{400} = 37.1525$

...sowie des χ^2-Koeffizienten

$$\begin{aligned}\chi^2 &= \sum_{j=1}^{m}\sum_{k=1}^{\ell}\frac{(h_{jk}-\frac{1}{n}h_{j\bullet}h_{\bullet k})^2}{\frac{1}{n}h_{j\bullet}h_{\bullet k}} \\ &= \frac{(5-21.2175)^2}{21.2175} + \frac{(160-145.935)^2}{145.935} + \frac{(42-39.8475)^2}{39.8475} \\ &\quad + \frac{(36-19.7825)^2}{19.7825} + \frac{(122-136.065)^2}{136.065} + \frac{(35-37.1525)^2}{37.1525} \\ \Rightarrow \chi^2 &\approx 28.7412\end{aligned}$$

Abschließende Bestimmung des Kontingenzmaßes nach Cramér V:

$$\begin{aligned} V &= \sqrt{\frac{\chi^2}{n(\min\{m,\ell\}-1)}} \\ &\approx \sqrt{\frac{28.7412}{400\cdot(\min\{2,3\}-1)}} = \sqrt{\frac{28.7412}{400\cdot 1}} \\ \Rightarrow V &\approx 0.2681 \end{aligned}$$

Es ergibt sich ein Kontingenzmaß nach Cramér von $V \approx 0.2581$. Somit gilt $0.2 < V \leq 0.6$, was für einen mittleren Zusammenhang zwischen beiden Merkmalen spricht [1].

Aufgabe BI-5

Eine Gesamtstichprobe von 20 Elementen, bei der man sich für zwei Merkmale X und Y interessiert, wurde zur Datenerhebung in zwei gleich große Teilstichproben aufgeteilt. Für die Teilstichproben ergaben sich dabei die folgenden Maßzahlen:

Teilstichprobe j	n_j	$\overline{x}_{n_j}^{(j)}$	$\overline{y}_{n_j}^{(j)}$	$\tilde{s}_{X,X}^{(j)}$	$\tilde{s}_{Y,Y}^{(j)}$	$r_{X,Y}^{(j)}$
1	10	12	0	36	9	1
2	10	0	6	9	36	1

Wie groß ist dann der Korrelationskoeffizient nach Bravais-Pearson $r_{X,Y}$ für die Gesamtstichprobe?

$$n = n_1 + n_2 = 10 + 10 = 20$$

$$\begin{aligned}\overline{x} &= \frac{1}{n}(n_1 \cdot \overline{x}^{(1)} + n_2 \cdot \overline{x}^{(2)}) \\ &= \frac{1}{20}(10 \cdot 12 + 10 \cdot 0) = \frac{1}{20} \cdot 120 \\ \Rightarrow \overline{x} &= 6\end{aligned}$$

$$\begin{aligned}\overline{y} &= \frac{1}{n}(n_1 \cdot \overline{y}^{(1)} + n_2 \cdot \overline{y}^{(2)}) \\ &= \frac{1}{20}(10 \cdot 0 + 10 \cdot 6) = \frac{1}{20} \cdot 60 \\ \Rightarrow \overline{y} &= 3\end{aligned}$$

$$\begin{aligned}\tilde{s}_{X,X}^2 &= \frac{1}{n}\Big(n_1 \cdot \big(\tilde{s}_X^{(1)}\big)^2 + n_2 \cdot \big(\tilde{s}_X^{(2)}\big)^2\Big) + \frac{1}{n}\Big(n_1\big(\overline{x}^{(1)} - \overline{x}\big)^2 + n_2\big(\overline{x}^{(2)} - \overline{x}\big)^2\Big) \\ &= \frac{1}{20}\big(10 \cdot 36^2 + 10 \cdot 9^2\big) + \frac{1}{20}\big(10 \cdot (12-6)^2 + 10 \cdot (0-6)^2\big) \\ \Rightarrow \tilde{s}_{X,X}^2 &= 724.5\end{aligned}$$

$$\begin{aligned}\tilde{s}_{Y,Y}^2 &= \frac{1}{n}\Big(n_1 \cdot \big(\tilde{s}_Y^{(1)}\big)^2 + n_2 \cdot \big(\tilde{s}_Y^{(2)}\big)^2\Big) + \frac{1}{n}\Big(n_1\big(\overline{y}^{(1)} - \overline{y}\big)^2 + n_2\big(\overline{y}^{(2)} - \overline{y}\big)^2\Big) \\ &= \frac{1}{20}\big((10 \cdot 9^2) + 10 \cdot 36^2\big) + \frac{1}{20}\big(10 \cdot (0-3)^2 + 10 \cdot (6-3)^2\big) \\ \Rightarrow \tilde{s}_{Y,Y}^2 &= 697.5\end{aligned}$$

$$
\begin{aligned}
\tilde{s}_{X,Y} &= \frac{1}{n}\big(n_1 \cdot r^{(1)}_{X,Y} \cdot \tilde{s}^{(1)}_{X,X} \cdot \tilde{s}^{(1)}_{Y,Y} + n_2 \cdot r^{(2)}_{X,Y} \cdot \tilde{s}^{(2)}_{X,X} \cdot \tilde{s}^{(2)}_{Y,Y}\big) \\
&\quad + \frac{1}{n}\big(n_1(\overline{x}^{(1)} - \overline{x})(\overline{y}^{(1)} - \overline{y}) + n_2(\overline{x}^{(2)} - \overline{x})(\overline{y}^{(2)} - \overline{y})\big) \\
&= \frac{1}{20}(10 \cdot 1 \cdot 36 \cdot 9 + 10 \cdot 1 \cdot 9 \cdot 36) + \frac{1}{20}\big(10(12-6)(0-3) + 10(0-6)(6-3)\big) \\
\Rightarrow \tilde{s}_{X,Y} &= 306
\end{aligned}
$$

$$
\begin{aligned}
r_{X,Y} &= \frac{\tilde{s}_{X,Y}}{\tilde{s}_X \tilde{s}_Y} = \frac{\tilde{s}_{X,Y}}{\sqrt{\tilde{s}^2_X \tilde{s}^2_Y}} \\
&= \frac{306}{\sqrt{724.5 \cdot 697.5}} \\
\Rightarrow r_{X,Y} &\approx 0.4305
\end{aligned}
$$

Insgesamt ergibt sich ein Korrelationskoeffizient nach Bravais-Pearson von $r_{X,Y} \approx 0.4305$. Da dieser Wert kleiner als 0.5 ist, ist von einer schwachen Korrelation zwischen Merkmal X und Y auszugehen.

Aufgabe BI-6

Unter 10 Studierenden wurde ein Wettlauf veranstaltet. Die folgende Tabelle enthält die Körpergröße (Merkmal X) und die Platzierung (Merkmal Y) der 10 Teilnehmer:

Körpergröße (in cm)	181	171	166	175	183	191	170	179	185	190
Platzierung	3	7	10	8	5	2	9	6	1	4

(a) Ermitteln Sie Schätzwerte $\hat{a}$ und $\hat{b}$ für die Parameter a und b aus dem Regressionsmodell $Y = a + bX$ mit $a, b \in \mathbb{R}$ nach der Methode der kleinsten Quadrate.

$$n = 10$$
$$\overline{x} = \frac{1}{10}(181 + 171 + 166 + 175 + 183 + 191 + 170 + 179 + 185 + 190) = 179.1$$
$$\overline{y} = \frac{1}{10}(3 + 7 + 10 + 8 + 5 + 2 + 9 + 6 + 1 + 4) = 5.5$$

$$\begin{aligned}\tilde{s}_{X,Y} &= \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})(y_j - \overline{y}) \\ &= \frac{1}{10}\big((181 - 179.1)(3 - 5.5) + (171 - 179.1)(7 - 5.5) + (166 - 179.1)(10 - 5.5) \\ &\quad + (175 - 179.1)(8 - 5.5) + (183 - 179.1)(5 - 5.5) + (191 - 179.1)(2 - 5.5) \\ &\quad + (170 - 179.1)(9 - 5.5) + (179 - 179.1)(6 - 5.5) + (185 - 179.1)(1 - 5.5) \\ &\quad + (190 - 179.1)(4 - 5.5)\big)\end{aligned}$$
$$\tilde{s}_{X,Y} = -20.45$$

$$\begin{aligned}\tilde{s}_X^2 &= \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2 \\ &= \frac{1}{10}\big((181 - 179.1)^2 + (171 - 179.1)^2 + (166 - 179.1)^2 + (175 - 179.1)^2 \\ &\quad + (183 - 179.1)^2 + (191 - 179.1)^2 + (170 - 179.1)^2 + (179 - 179.1)^2 \\ &\quad + (185 - 179.1)^2 + (190 - 179.1)^2\big)\end{aligned}$$
$$\tilde{s}_X^2 = 65.09 \text{ (positiv)}$$

$$\hat{\beta} = \frac{\tilde{s}_{X,Y}}{\tilde{s}_X^2} = \frac{-20.45}{65.09} \approx -0.3142$$
$$\hat{\alpha} = \overline{y} - \hat{\beta} \cdot \overline{x} \approx 5.5 + 0.3142 \cdot 179.1 \approx 61.7697$$
$$\Rightarrow Y = \hat{\alpha} + \hat{\beta} \cdot X \approx 61.7697 - 0.3142X$$

Insgesamt ergibt sich die geschätzte Regressionsgerade $Y \approx 61.7697 - 0.3142X$ nach der Methode der kleinsten Quadrate.

> (b) Berechnen Sie das Bestimmtheitsmaß R^2 für das lineare Regressionsmodell gemäß (a) sowie den empirischen Korrelationskoeffizient $r_{X,Y}$ und interpretieren Sie kurz Ihre Ergebnisse.

x	y	$\hat{y}$	$(\hat{y} - \overline{y})^2$	$(y - \overline{y})^2$
181	3	4.8995	0.3606	6.25
171	7	8.0415	6.4592	2.25
166	10	9.6125	16.9127	20.25
175	8	6.7847	1.6505	6.25
183	5	4.2711	1.5102	0.25
191	2	1.7575	14.0063	12.25
170	9	8.3557	8.155	12.25
179	6	5.5279	0.0008	0.25
185	1	3.6427	3.4496	20.25
190	4	2.0717	11.7532	2.25
			$\Sigma \approx 64.2580$	$\Sigma = 82.5$

$$\text{SQE} = \sum_{j=1}^{n} (\hat{y}_j - \overline{y})^2 \approx 64.2580$$

$$\text{SQT} = \sum_{j=1}^{n} (y_j - \overline{y})^2 = 82.5$$

$$R^2 = \frac{\text{SQE}}{\text{SQT}} = \frac{64.2580}{82.5} \approx 0.7789$$

$$\tilde{s}_Y^2 = \frac{1}{n} \sum_{j=1}^{n} (y_j - \overline{y})^2 = \frac{1}{10} \cdot 82.5 = 8.25$$

$$r_{X,Y} = \frac{\tilde{s}_{X,Y}}{\tilde{s}_X \tilde{s}_Y} = \frac{\tilde{s}_{X,Y}}{\sqrt{\tilde{s}_X^2 \tilde{s}_Y^2}} = \frac{-20.45}{\sqrt{65.09 \cdot 8.25}} \approx -0.8825$$

Insgesamt ergibt sich ein Bestimmtheitsmaß von $R^2 = 0.7789$ sowie ein empirischer Korrelationskoeffizient von $r_{X,Y} \approx -0.8825$. Somit ergibt sich $|r_{X,Y}| > 0.8$, was auf eine hohe Güte des Regressionsmodells sowie eine starke gegenläufige lineare Korrelation zwischen den beiden Merkmalen hinweist.

(c) Bestimmen Sie (zum Vergleich) den Rangkorrelationskoeffizienten $r^\star_{X,Y}$ nach Spearman (da ja im Grunde das Merkmal Y nur ordinalskaliert ist) und interpretieren Sie kurz Ihr Ergebnis.

i	1	2	3	4	5	6	7	8	9	10
x_i	181	171	166	175	183	191	170	179	185	190
$\mathrm{Rg}(x_i)$	6	3	1	4	7	10	2	5	8	9
y_i	3	7	10	8	5	2	9	6	1	4
$\mathrm{Rg}(y_i)$	3	7	10	8	5	2	9	6	1	4

Hier kommen *keine* Werte in den Stichproben $x_1, \ldots, x_n$ und $y_1, \ldots, y_n$ mehrfach vor. Daher kann die vereinfachte Formel zur Bestimmung des Rangkorrelationskoeffizienten $r^\star_{X,Y}$ verwendet werden.

$$
\begin{aligned}
r^\star_{X,Y} &= 1 - \frac{6}{n(n^2-1)} \sum_{j=1}^{n} \big(\mathrm{Rg}(x_j)^2 - \mathrm{Rg}(y_j)^2 \big) \\
&= 1 - \frac{6}{10 \cdot (10^2-1)} \sum_{j=1}^{10} \big(\mathrm{Rg}(x_j)^2 - \mathrm{Rg}(y_j)^2 \big) \\
&= 1 - \frac{6}{990} \big((6-3)^2 + (3-7)^2 + (1-10)^2 + (4-8)^2 + (7-5)^2 + (10-2)^2 \\
&\qquad + (2-9)^2 + (5-6)^2 + (8-1)^2 + (9-4)^2 \big) \\
&= 1 - \frac{1}{165} \cdot 314 \\
\Rightarrow r^\star_{X,Y} &\approx -0.9030
\end{aligned}
$$

Es ergibt sich ein Rangkorrelationskoeffizient (nach Spearman) von $r^\star_{X,Y} \approx -0.9030$. Dies weißt auf einen stark gegenläufigen monotonen Zusammenhang hin. Dies entspricht der Folgerung aus Teilaufgabe (b), aus dem ebenfalls eine stark gegenläufige lineare Korrelation gefolgert werden konnte.

Quellenverzeichnis

[1] IBM Corporation. *Cramér's V*. IBM Cognos Analytics 11.1.x. 29. Feb. 2024. URL: https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramrs-v (besucht am 08. 11. 2024).