

Wir schauen erneut auf Aufgabe 3 von Übungsblatt 2. Die Rohdaten dort waren:

	Pizzatyp				
	Margherita	Salami	Hawaii	Tonno	Funghi
Deutschland	42	20	16	18	14
Italien	30	25	0	22	23
Frankreich	35	40	12	10	33

Die erwarteten Häufigkeiten v_{jk} wurden berechnet als:

erwartete Häufigkeiten	Pizzatyp					Σ
	Margherita	Salami	Hawaii	Tonno	Funghi	
Deutschland	$\frac{110 \cdot 107}{340} = 34.62$	27.5	9.06	16.18	22.65	110
Italien	31.47	25.0	8.24	14.71	20.59	100
Frankreich	40.91	32.5	10.71	19.12	26.76	130
Σ	107	85	28	50	70	340

a) (2 Punkte) Berechnen Sie den χ^2 -Kontingenzkoeffizienten.

$$\chi^2 = \sum_{j=1}^J \sum_{k=1}^K \frac{(N_{jk} - v_{jk})^2}{v_{jk}} = \frac{(42 - 34.62)^2}{34.62} + \frac{(20 - 27.5)^2}{27.5} + \dots$$

$$= 33.19 \quad (\text{Excel})$$

jeder Term:				
1.57	2.05	5.32	0.21	3.30
0.07	0.00	8.24	3.62	0.28
0.85	1.73	0.16	4.35	1.45

b) (2 Punkte) Berechnen Sie den korrigierten Kontingenzkoeffizienten nach Pearson. Interpretieren Sie das Ergebnis.

$$\min \{J, K\} = 3$$

$$c = \sqrt{\frac{\chi^2}{\chi^2 + N} \cdot \frac{3}{3-1}} = \sqrt{\frac{33.19}{33.19 + 340} \cdot \frac{3}{2}} = 0.365$$

→ mittlere schwache Abhängigkeit

Aufgabe 2

(8 Punkte)

Die Hobbystatistikerin B. Rechnen übt ihre Statistikkenntnisse gerne an selbsterhobenen Datensätzen. Erst gestern hat sie die Körpergrößen und Alter der Kinder in der Nachbarschaft erhoben. Die Daten lauten:

Alter in Jahren	9	6	12	10	8	9
Körpergröße in cm	130	125	145	140	130	140

$$N=6$$

Sie ist daran interessiert, die Körpergröße anhand des Alters vorherzusagen.

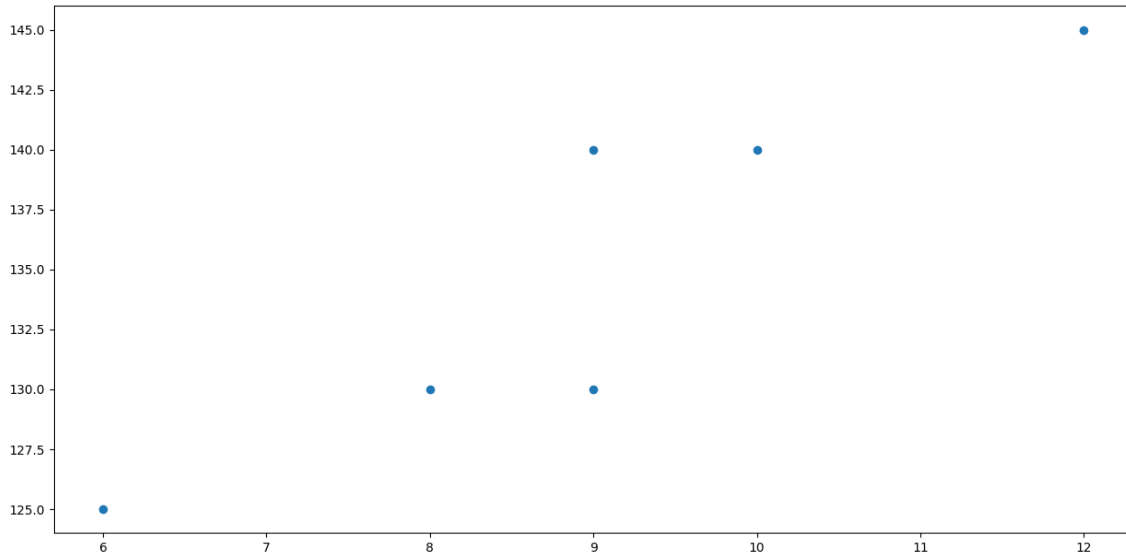
a) (1 Punkt) Erstellen Sie anhand der Rohdaten ein Streudiagramm. Achten Sie auf

Alter in Jahren	9	6	12	10	8	9	$N=6$
Körpergröße in cm	130	125	145	140	130	140	

Sie ist daran interessiert, die Körpergröße anhand des Alters vorherzusagen.

- a) (1 Punkt) Erstellen Sie anhand der Rohdaten ein Streudiagramm. Achten Sie auf die Wahl der richtigen Achsen für die Regression.

Anforderung für die ganze Aufgabe 2: matplotlib, numpy, scipy



- b) (1 Punkt) Berechnen Sie die Kovarianz zwischen Alter und Körpergröße.

$$\bar{x} = \frac{1}{6} (9+6+12+10+8+9) = 9$$

$$\bar{y} = \frac{1}{6} (130+125+145+140+130+140) = 135$$

$$s_{xy} = \frac{1}{5} \sum_{i=1}^6 (x_i - \bar{x})(y_i - \bar{y}) = 14$$

- c) (2 Punkte) Berechnen Sie den Korrelationskoeffizienten nach Bravais-Pearson zwischen Alter und Körpergröße.

$$\text{Var}_x = \frac{1}{5} (0^2 + (-3)^2 + 3^2 + 1^2 + (-1)^2 + 0^2) = \frac{1}{5} (20) = 4 \rightarrow s_x = 2$$

$$\text{Var}_y = \frac{1}{5} ((-5)^2 + (-10)^2 + 10^2 + 5^2 + (-5)^2 + 5^2) = \frac{1}{5} (300) = 60 \rightarrow s_y = 7,74$$

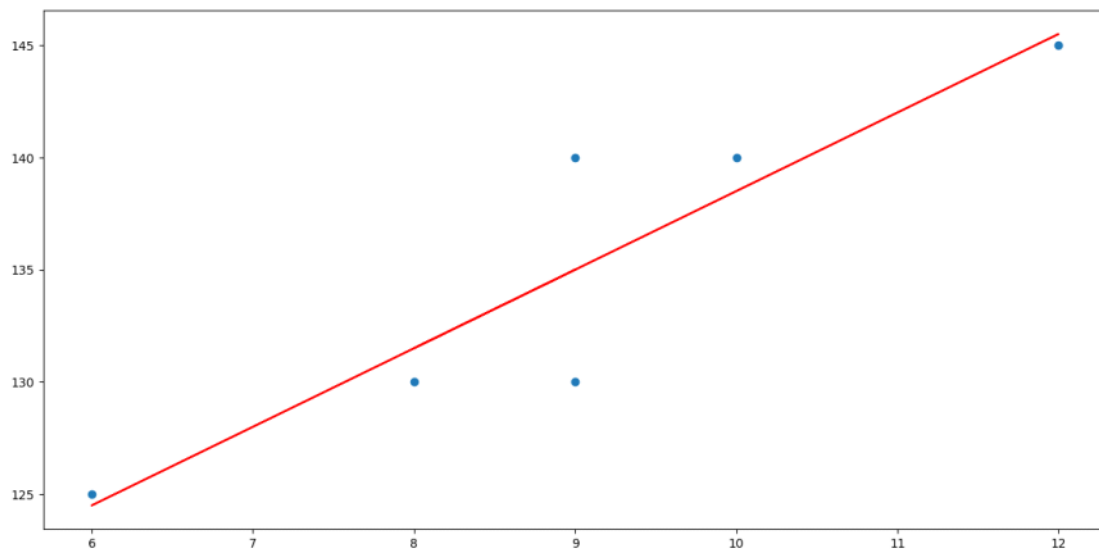
$$\rightarrow r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{14}{2 \cdot 7,74} = 0,904$$

- d) (2 Punkte) Führen Sie eine lineare Regression durch, d.h. berechnen Sie d und c und geben Sie die vollständige Regressionsgleichung an. Zeichnen Sie die Regressionsgerade in Ihr Streudiagramm aus a) ein.

$$d = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{\text{var}_x} = \frac{14}{4} = 3,5$$

$$c = \bar{y} - d\bar{x} = 135 - 3,5 \cdot 9 = 103,5$$

$$\Rightarrow y = 3,5x + 103,5$$



- e) (2 Punkte) Bestimmen Sie den Rangkorrelationskoeffizienten nach Spearman für die Variablen Alter und Körpergröße. Vergeben Sie bei geteilten Rängen den jeweils mittleren Rang. Beispiel: (5, 4, 5, 6) hat die Ränge (2.5, 1, 2.5, 4).

Alter	9	6	12	10	8	9
Alter gerankt	3.5	1	6	5	2	3.5
Körpergröße	130	125	145	140	130	140
Körpergröße gerankt	2.5	1	6	4.5	2.5	4.5

$$\begin{aligned}
 r_{xy}^{\text{Sp.}} &= 1 - \frac{6}{N(N^2-1)} \cdot \sum_{i=1}^N (R(x_i) - R(y_i))^2 \\
 &= 1 - \frac{6}{6(6^2-1)} (1^2 + 0^2 + 0^2 + 0.5^2 + (-0.5)^2 + (-1)^2) \\
 &= 1 - \frac{1}{35} (2.5) = 0.928
 \end{aligned}$$

Aufgabe 3

(3 Punkte)

Beweisen Sie die Gleichung der χ^2 -Größe der Vorlesung:

$$\sum_{j=1}^J \sum_{k=1}^K \frac{(N_{jk} - v_{jk})^2}{v_{jk}} = N \left(\sum_{j=1}^J \sum_{k=1}^K \frac{N_{jk}^2}{N_{j \cdot} \cdot N_{\cdot k}} - 1 \right)$$

Sei $p_{jk} = \frac{N_{j \cdot} \cdot N_{\cdot k}}{N} \rightarrow v_{jk} = \frac{p_{jk}}{N}$

Abgekürzt $\sum_{j=1}^J \sum_{k=1}^K f(j,k) = S(f(j,k))$

Rechte Seite = $N \cdot \left(S\left(\frac{N_{jk}^2}{p_{jk}}\right) - 1 \right) = N \cdot S\left(\frac{N_{jk}^2}{p_{jk}}\right) - N$

$$= S\left(\frac{N \cdot N_{jk}^2}{p_{jk}}\right) - N$$

$$\frac{(N_{jk} - v_{jk})^2}{v_{jk}} = \frac{N_{jk}^2 - 2 \cdot N_{jk} \cdot v_{jk} + v_{jk}^2}{v_{jk}} = \frac{N_{jk}^2}{v_{jk}} - 2N_{jk} + v_{jk}$$

$$= \frac{N_{jk}^2}{\frac{p_{jk}}{N}} - 2N_{jk} + v_{jk} = \frac{N \cdot N_{jk}^2}{p_{jk}} - 2N_{jk} + v_{jk}$$

$$\Rightarrow \text{Linke Seite} = S\left(\frac{(N_{jk} - v_{jk})^2}{v_{jk}}\right) = S\left(\frac{N \cdot N_{jk}^2}{p_{jk}} - 2N_{jk} + v_{jk}\right)$$

$$= S\left(\frac{N \cdot N_{jk}^2}{p_{jk}}\right) - 2 \cdot S(N_{jk}) + S(v_{jk})$$

Summe ganzen Tabelle
= N

$$= S\left(\frac{N \cdot N_{jk}^2}{p_{jk}}\right) - 2N + N$$

Linke Seite = $S\left(\frac{N \cdot N_{jk}^2}{p_{jk}}\right) - N$

□