

Lösung zu den Übungsaufgaben Kapitel 15

(1) Sie fragen acht Kinder und Jugendliche in der Landauer Fußgängerzone, wie viele Stunden sie am Tag fernsehen und wie alt sie sind. Sie erhalten die folgenden Daten:

Person Nr.	Alter	Stunden TV pro Tag
1	12	2
2	13	1
3	15	4
4	16	3
5	18	6
6	20	5
7	22	3
8	24	4

Berechnen Sie die Kreuzproduktsumme, die Kovarianz und die Produkt-Moment-Korrelation.

Die Ergebnisse sind in der folgenden Tabelle angegeben. Die Variable Alter wird dabei mit X bezeichnet, die Variable »Stunden TV pro Tag« mit Y . In den beiden ersten Spalten stehen die Originalwerte, in den Spalten 3 und 4 die zentrierten Werte und in Spalte 5 die Kreuzprodukte.

	x_m	y_m	$x_m - \bar{x}$	$y_m - \bar{y}$	$(x_m - \bar{x}) \cdot (y_m - \bar{y})$
	12	2	-5,5	-1,5	8,25
	13	1	-4,5	-2,5	11,25
	15	4	-2,5	0,5	-1,25
	16	3	-1,5	-0,5	0,75
	18	6	0,5	2,5	1,25
	20	5	2,5	1,5	3,75
	22	3	4,5	-0,5	-2,25
	24	4	6,5	0,5	3,25
Summe:	140	28	0	0	25
Mittelwert:	17,5	3,5	0	0	3,125
Standardabweichung:	4	1,5			

In der Zeile »Summe« unterhalb der letzten Spalte findet sich die Kreuzproduktsumme ($KPS = 25$), in der Zeile »Mittelwert« darunter die Kovarianz ($s_{XY} = 3,125$). Teilt man die Kovarianz durch das Produkt der Streuungen ($s_X \cdot s_Y = 4 \cdot 1,5 = 6$), so ergibt sich die Korrelation $r_{XY} = 0,52$.

- (2) Die Korrelation zweier Variablen X und Y beträgt $r_{XY} = -0,88$. Die Streuungen betragen $s_x = 3,3$ und $s_y = 10,5$. Es wurden Daten von $n = 230$ Personen erhoben. Berechnen Sie die Kovarianz und die Kreuzproduktsumme.

Löst man Formel F 15.5 nach der Kovarianz s_{XY} auf, so ergibt sich:

$$r_{XY} = \frac{s_{XY}}{s_X \cdot s_Y}$$

$$s_{XY} = r_{XY} \cdot s_X \cdot s_Y$$

$$= -0,88 \cdot 3,3 \cdot 10,5 = -30,492$$

Setzt man Formel F15.1 in Formel F15.2 ein und löst nach der Kreuzproduktsumme auf, so ergibt sich:

$$s_{XY} = \frac{1}{n} \cdot \sum_{m=1}^n (x_m - \bar{x}) \cdot (y_m - \bar{y})$$

$$s_{XY} = \frac{1}{n} \cdot KPS$$

$$KPS = s_{XY} \cdot n$$

$$= -30,492 \cdot 230 = -7013,16$$

- (3) Welche der folgenden Aussagen ist richtig?

- a) Ist der Zusammenhang zweier Variablen positiv, so sind auch alle Kreuzprodukte positiv. richtig falsch
- b) Bei einem perfekten Zusammenhang zwischen X und Y liegen alle Punkte im Scatterplot auf einer Linie. richtig falsch
- c) Es gilt immer: Je steiler die »Linie« im Scatterplot, desto stärker ist der Zusammenhang zwischen X und Y . richtig falsch
- d) Wenn x -Werte, die oberhalb des Mittelwertes von X liegen, mit y -Werten einhergehen, die unterhalb des Mittelwerts von Y liegen (und umgekehrt), so liegt ein negativer Zusammenhang vor. richtig falsch
- e) Das Produkt-Moment kann niemals kleiner als Null werden. richtig falsch
- f) Die Kovarianz kann niemals kleiner als Null werden. richtig falsch
- g) Die Kovarianz kann niemals größer als Eins werden. richtig falsch
- h) Die Kovarianz kann maximal so groß werden wie das Produkt der Streuungen von X und Y . richtig falsch

- (4) Berechnen Sie die Rangkorrelation nach Spearman für die Rohdaten in Tabelle 15.7. Vergleichen Sie das Ergebnis mit dem Wert, den wir für die Rangkorrelation nach Kendall in unserem Rechenbeispiel bestimmt hatten. Was muss man aus dem Vergleich schließen?

Um Spearmans *Rho* zu berechnen, benötigt man zunächst die quadrierten paarweisen Differenzen zwischen den Rangplätzen auf beiden Variablen. Diese sind in der folgenden Tabelle abgetragen:

	R_X	R_Y	d	d^2
Uta	1	2	-1	1
Anna	2	3	-1	1
Elke	3	1	2	4
Ina	4	5	-1	1
Gabi	5	4	1	1
Summe:				8

Eingesetzt in Formel F 15.15 ergibt sich dann:

$$r_s = 1 - \frac{6 \cdot \sum_{m=1}^n d_m^2}{n \cdot (n^2 - 1)}$$

$$= 1 - \frac{6 \cdot 8}{5 \cdot (25 - 1)} = 1 - \frac{48}{120} = 0,60$$

Für die Rangkorrelation nach Spearman ergibt sich ein Wert von 0,60, während für die Rangkorrelation nach Kendall ein Wert von 0,40 errechnet worden war. Beide Koeffizienten führen also nicht zum gleichen Ergebnis.

(5) In einer Untersuchung zum Sonnenschutzverhalten wurden $n = 518$ in Bezug auf ihre Einstellung zur Hautbräune befragt (Variable X) und notiert, welcher Lichtschutzfaktor (Variable Y) die Sonnenmilchflasche aufwies, die die Personen aus einem Sortiment gewählt haben. Die Variable X hat die beiden Ausprägungen »negativ« (a_1) und »positiv« (a_2). Die Variable Y hat die beiden Ausprägungen »gering« (b_1) und »hoch« (b_2). Es ergaben sich folgende Zellhäufigkeiten: $n_{11} = 39$, $n_{12} = 75$, $n_{21} = 203$, $n_{22} = 201$. Berechnen Sie

(a) den $\hat{\phi}$ -Koeffizienten zum einen nach Formel F 15.17, zum anderen als Spezialfall von Cramérs V nach Formel F 15.25,

(b) das Odds-Ratio,

(c) Yules Q .

Überprüfen Sie die statistische Nullhypothese, dass es keinen Zusammenhang zwischen beiden Variablen gibt.

(a) $\hat{\phi} = -0,13$

(b) $OR = 0,51$

(c) $Q = -0,36$

Die statistische Nullhypothese, dass es keinen Zusammenhang zwischen beiden Variablen gibt, prüfen wir mit dem χ^2 -Test nach Formel F 11.48b. Der Wert der Prüfgröße beträgt:

$$\chi^2 = 9,19$$

Dieser Wert ist größer als der kritische Wert $\chi^2_{0,95;1} = 3,8415$, den wir in Tabelle A5 ablesen. Die Nullhypothese muss somit verworfen werden.