+1). Korrelationskoeffizienten unterschiedlicher Stichproben oder verschiedener Variablen können so miteinander verglichen werden, was mit der Kovarianz nicht möglich ist.

c. Eine Kovarianz (Korrelation) von null besagt, dass die untersuchten Merkmale voneinander unabhängig sind, d. h., Variable x steht in keinerlei Zusammenhang mit Variable y.

d. Mithilfe der Fishers Z-Transformation lassen sich Produkt-Moment-Korrelationen mitteln. Die nötigen Schritte sind:
   1. Transformation der einzelnen Werte in Fishers Z-Werte
   2. Bilden eines Z-Mittelwerts
   3. Rücktransformation des Z-Mittelwerts in eine Korrelation

e. Die lineare Regression dient der Merkmalsvorhersage. Aus der Ausprägung der unabhängigen Variable x, genannt Prädiktor, kann die zugehörige Ausprägung der abhängigen Variable, des Kriteriums, mithilfe der Regressionsgleichung vorhergesagt werden.

f. Der Determinationskoeffizient gibt denjenigen Varianzanteil der abhängigen Variable y wieder, der durch die unabhängige Variable aufgeklärt werden kann. Er ist somit ein prozentuales Maß für den gemeinsamen Varianzanteil zweier Variablen. Ebenso lässt er sich als Effektstärkenmaß begreifen, das die Größe des Einflusses der unabhängigen Variable auf die abhängige Variable angibt.

g. Bei der linearen Regression lässt sich die Gesamtvarianz der tatsächlichen Werte des Kriteriums in die Regressionsvarianz und die Residualvarianz aufspalten:

$$\hat{\sigma}^2_y = \hat{\sigma}^2_{\hat{y}} + \hat{\sigma}^2_{y/x}$$

Anwendungsaufgaben

Aufgabe 1

a. Anhand des Streudiagramms (Abb. 1) lässt sich ein linearer Zusammenhang vermuten.

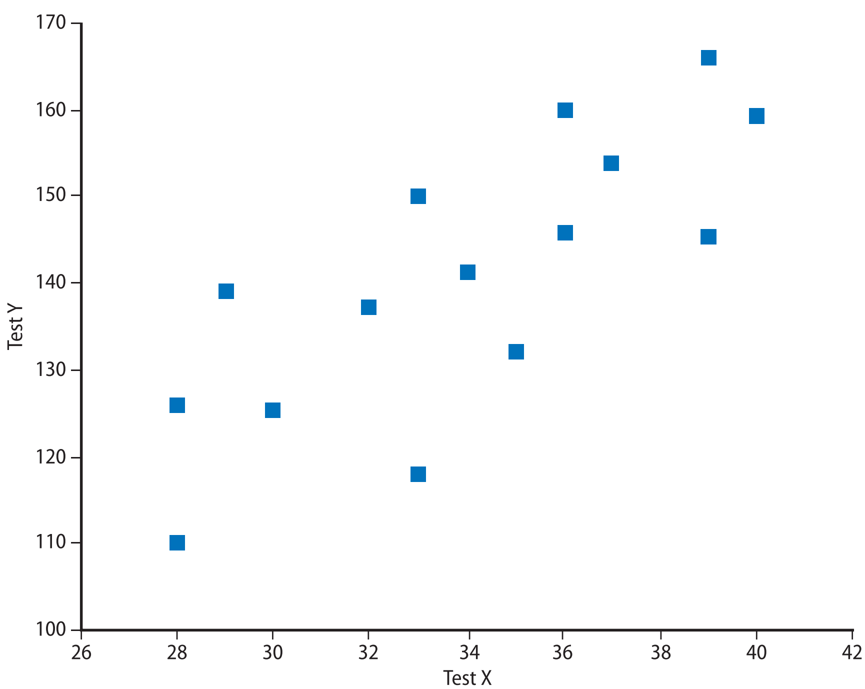


Abb. 1

b.

$$cov_{(x;y)} = \frac{\sum_{i=1}^{N}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{N-1}\ ;\ \bar{x} = 33{,}93\ ;\ \bar{y} = 140{,}53$$

$$cov_{(x;y)} = \frac{(34-33{,}93)\cdot(141-140{,}53)+\ldots+(36-33{,}93)\cdot(160-140{,}53)}{15-1}$$

$$= \frac{702{,}50}{14} = 50{,}18$$

Die Kovarianz (*x;y*) beträgt 50,18.

c.

$$r_{xy} = \frac{cov_{(x;y)}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} = \frac{50{,}18}{3{,}99 \cdot 16{,}19} = 0{,}78$$

Der Korrelationskoeffizient r_{xy} beträgt 0,78.

Aufgabe 2

a.

$$cov_{max} = \hat{\sigma}_x \cdot \hat{\sigma}_y = 3 \cdot 7 = 21$$

Der Wert, den die Kovarianz maximal annehmen kann, beträgt 21.

b.

$$r_{xy} = \frac{cov_{emp}}{cov_{max}} \Leftrightarrow cov_{emp} = r_{xy} \cdot cov_{max} = (0{,}80) \cdot 21 = 16{,}8$$

Die Kovarianz muss mindestens einen Wert von 16,8 annehmen, wenn $r_{xy} \geq 0{,}80$ sein soll.

Aufgabe 3

a.

$$\bar{y}_{(x=1)} = y_1 = 64{,}25\ \text{(m)};\ n_1 = 8$$
$$\bar{y}_{(x=0)} = \bar{y}_0 = 61{,}14\ \text{(w)};\ n_0 = 7$$
$$\hat{\sigma}_y = 3{,}91\ ;\ N = 15$$

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{\hat{\sigma}_y} \cdot \sqrt{\frac{n_1 \cdot n_0}{N^2}} \Leftrightarrow r_{pb} = \frac{(64{,}25 - 61{,}14)}{3{,}91} \cdot \sqrt{\frac{8 \cdot 7}{225}} = 0{,}40$$

Es besteht ein Zusammenhang zwischen dem Geschlecht und der Bereitschaft zu helfen (Stärke des Zusammenhangs: 0,40). Jungen helfen im Schnitt etwas mehr als Mädchen.

b.

$$t_{df=13} = \frac{r_{pb} \cdot \sqrt{N-2}}{\sqrt{1 - r_{pb}{}^2}} = \frac{0{,}40 \cdot \sqrt{13}}{\sqrt{0{,}84}} = 1{,}57$$

In der *t*-Verteilung nachgeschlagen ergibt sich ein kritischer *t*-Wert von 1,771. Somit ist das Ergebnis nicht signifikant (bei $\alpha = 0{,}05$). Wie immer bei nicht signifikanten Ergebnissen ist an dieser Stelle eine A-posteriori-Teststärkeanalyse nötig, sofern vor der Untersuchung keine Stichprobenumfangsplanung durchgeführt wurde.

Aufgabe 4

Gemäß der Formel aus ► Abschn. 4.1.11 ergibt sich eine Partialkorrelation zwischen Rotweinkonsum und Weisheit bei Auspartialisierung des Alters von:

$$r_{xy|z} = \frac{r_{xy} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2) \cdot (1 - r_{xz}^2)}} = \frac{0{,}45 - 0{,}68 \cdot 0{,}55}{\sqrt{(1 - 0{,}68^2) \cdot (1 - 0{,}55^2)}} = 0{,}124$$

Der Zusammenhang zwischen Rotweinkonsum und Weisheit entpuppt sich als Scheinkorrelation!

Aufgabe 5

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{240}{1716} = 1 - 0{,}14 = 0{,}86$$

Die Stärke des Zusammenhangs beträgt 0,86.

Aufgabe 6

a. Die Abb. 2 zeigt die Verteilung der Messwertepaare in einem Koordinatensystem (x-Achse: Zwilling 1; y-Achse: Zwilling 2).

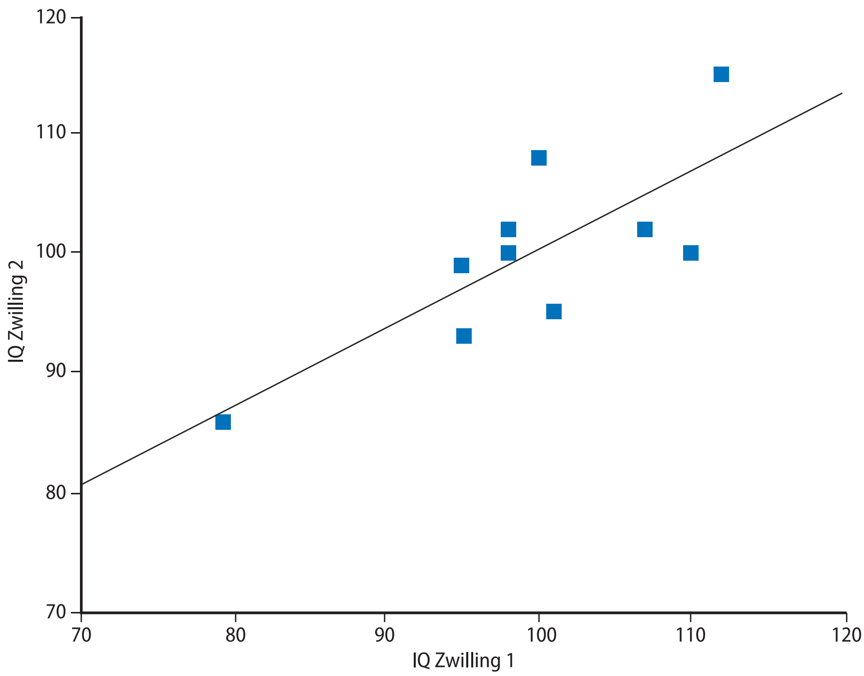


Abb. 2

b. Lineare Regressionsgleichung (für Vorhersage y aus x):

$$\hat{y}_i = b_{yx} x_i + a_{yx}$$

$$\bar{x} = 99{,}5\,;\ \bar{y} = 100\,;\ \hat{\sigma}_x^2 = 87{,}83\,;\ cov_{(x;y)} = 57{,}67$$

Berechnung des Regressionsgewichts b_{yx}: $b_{xy} = \frac{cov_{(x;y)}}{\hat{\sigma}_x^2} = \frac{57{,}67}{87{,}83} = 0{,}66$

Berechnung der Regressionskonstanten a_{yx}:

$$a_{yx} = \bar{y} - b_{yx} \cdot \bar{x} = 100 - (0{,}66 \cdot 99{,}5) = 34{,}33$$

⇒ Die Funktion der vorliegenden linearen Regressionsgleichung lautet:

$$\hat{y}_i = 0{,}66 \cdot x_i + 34{,}33$$

c. Die vorhergesagten Werte $\hat{y}_i$ sind in nebenstehender Tabelle aufgeführt.

x_i	$\hat{y}_i$
110	106,93
98	99,01
101	100,99
79	86,47
107	104,95
98	99,01
95	97,03
100	100,33
112	108,25
95	97,03

d. Die Formel zur Berechnung der Residualvarianz lautet:

$$\hat{\sigma}^2_{[y/x]} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N-1} = \frac{(100-106{,}93)^2 + \ldots + (99-97{,}03)^2}{10-1} = \frac{227{,}25}{9} = 25{,}25$$

Die Formel zur Berechnung des Standardschätzfehlers lautet: $\hat{\sigma}_{[y/x]} = \sqrt{\hat{\sigma}^2_{[y/x]}}$

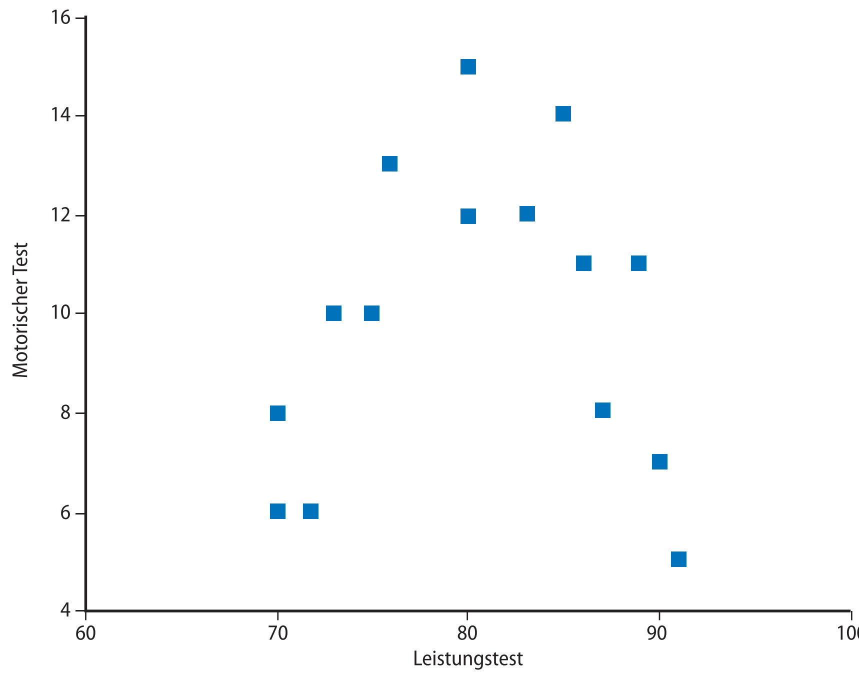
$$\Rightarrow\ \hat{\sigma}_{(y/x)} = 5{,}02$$

e. Der Anteil der Vererbung entspricht der durch den Effekt aufgeklärten Varianz. Die Fehlervarianz fasst den Anteil aller übrigen Umwelteinflüsse wie Erziehung oder soziales Umfeld zusammen.

Aufgabe 7

a. Abb. 3

Abb. 3



b. $$r_{xy} = \frac{cov_{(x;y)}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} = \frac{1{,}63}{7{,}42 \cdot 3{,}09} = 0{,}071$$

Der Korrelationskoeffizient r_{xy} beträgt 0,071.

c. Mit r_{xy} wird die Stärke des linearen Zusammenhangs zwischen x und y ermittelt. Sie beträgt 0,071. Man könnte also auf das Fehlen eines entscheidenden Zusammenhangs schließen. Ein linearer Zusammenhang existiert hier nicht, wie auch das Streuungsdiagramm zeigt (Abb. 3). Der Zusammenhang ist kurvilinear. Dessen Stärke kann jedoch durch r_{xy} nicht erfasst werden. An diesem Beispiel zeigt sich wieder, dass es immer besser ist, das Streudiagramm als Interpretationshilfe mit heranzuziehen.

Aufgabe 8

a. Streudiagramm (Abb. 4): Man sieht schon hier, dass die beiden Variablen **negativ** miteinander kovariieren, d. h., je höher die Angstneigung ist, desto geringer ist die Kreativität und umgekehrt. Die Stärke des Zusammenhangs ist nicht so ohne Weiteres zu ersehen, dürfte aber recht groß sein (die Linie beschreibt den Punkteschwarm ziemlich gut).
Zusätzlich ist in Abb. 4 die Regressionsgerade aus ► Aufgabe 8f eingezeichnet.

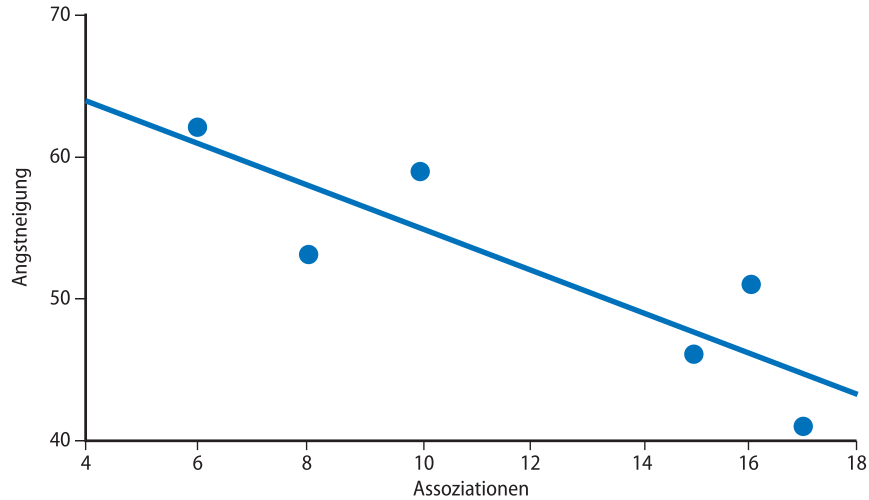


Abb. 4

b. Die Tabelle zeigt die univariaten Statistiken der Variablen.

	Angsttest	Assoziationen
Mittelwerte:	$\overline{y} = 52$	$\overline{x} = 12$
Varianzen:	$\hat{\sigma}^2 = 61{,}6$	$\hat{\sigma}^2 = 21{,}2$
Streuungen:	$\hat{\sigma} = 7{,}85$	$\hat{\sigma} = 4{,}6$

c. $$cov_{(x;y)} = \frac{\sum_{i=1}^{n}(x_i - \overline{x}) \cdot (y_i - \overline{y})}{N-1} \Rightarrow cov_{(\text{Angst, Assoz.})} = -31$$

d. $$r_{xy} = \frac{cov_{(x;y)}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} \Rightarrow r_{xy} = -0{,}86$$

e. Folgende Schritte sind erforderlich:
   1. Aus der Kovarianz und der Prädiktorvarianz das Regressionsgewicht b bestimmen.
   2. Aus den Mittelwerten und dem Regressionsgewicht b die Konstante a estimmen.
   3. Einsetzen in: [geschätzter Testwert] = b · [Assoziationen] + a

   Lineare Regressionsgleichung (für Vorhersage y aus x):

   $\hat{y}_i = b_{yx} x_i + a_{yx}$

   Berechnung des Regressionsgewichts b_{yx}: $b_{yx} = \frac{cov_{(x;y)}}{\hat{\sigma}_x^2}$

   Berechnung der Regressionskonstanten a_{yx}: $a_{yx} = \overline{y} - b_{yx} \cdot \overline{x}$

f. Regressionsgleichung:

   $\hat{y} = -1{,}46 \cdot x + 69{,}52$ (Regressionsgerade: Abb. 4)

g. $$r^2 = \frac{\hat{\sigma}_{\hat{y}}^2}{\hat{\sigma}_y^2} = \frac{\hat{\sigma}_{\hat{y}}^2}{\hat{\sigma}_{\hat{y}}^2 + \hat{\sigma}_{[y/x]}^2} \Rightarrow r^2 = 0{,}74$$

Der Determinationskoeffizient gibt an, wie viel Varianz der Kriteriumswerte (hier der Werte im Angsttest) durch die Regression über die Variable »Assoziationen« erklärt wird. Oder anders ausgedrückt: Wie viel Varianz haben die Variablen »Angsttest« und »Assoziation« gemeinsam, d. h., wie viel Varianz der einen Variable steckt auch in der anderen.

h. Der Standardschätzfehler $\hat{\sigma}_{[y|x]}$ ist die Wurzel aus der Residualvarianz.

Aufgrund der Beziehung [Kriteriumsvarianz] = [Residualvarianz] + [Regressionsvarianz] kann man ihn auch ohne mühsame Rechnung ermitteln, denn Kriteriumsvarianz und Regressionsvarianz sind ja bereits bekannt:

$$61{,}6 = \hat{\sigma}_{[y|x]}^2 + (0{,}74 \cdot 61{,}6)\ ;\ \hat{\sigma}_{[y|x]}^2 = 16{,}016 \Rightarrow \hat{\sigma}_{[y|x]} \approx 4$$

i. $\hat{\sigma}^2_{\hat{y}} = 0{,}74 \cdot 61{,}6 = 45{,}58$

j. Dies spielt für die Korrelation keine Rolle ($r = -0{,}86$).

k. Die Regressionsgeraden schneiden sich bei den Mittelwertskoordinaten.
Wenn also auf der x-Achse die Variable »Assoziationen« und auf der y-Achse die Variable »Angst« abgetragen wird, so lautet der konkrete Schnittpunkt (12;52).

l. Mithilfe der Fishers Z-Transformation erfolgen folgende Schritte:
   1. Transformation der einzelnen Korrelationswerte in Fishers Z-Werte
   2. Bildung des arithmetischen Mittels der Fishers Z-Werte
   3. Rücktransformation des arithmetischen Mittels der Fishers Z-Werte in eine Korrelation

Die Berechnungsvorschrift zur Transformation der Korrelationen in Fishers Z-Werte lautet:

$$Z = \frac{1}{2} \cdot \ln\left(\frac{1+r}{1-r}\right)$$

Rücktransformation nach Bildung eines Fishers Z-Mittelwerts findet statt nach folgender Formel:

$$\bar{r} = \frac{e^{2 \cdot \bar{Z}} - 1}{e^{2 \cdot \bar{Z}} + 1}$$

Es ergibt sich eine mittlere Korrelation von –0,745.

m.

$$r_{xy} = \frac{cov_{(x;y)}}{\hat{\sigma}_x \cdot \hat{\sigma}_y} \Rightarrow r = -0{,}11$$

n. Regressionsgleichung: [geschätzter Angstwert] = –0,21 · [Alter] + 57
   a. 53,22
   b. 48,6
   c. 36

Die Güte der Regression kann über den Determinationskoeffizienten beurteilt werden. Dieser beträgt hier $r^2 = 0{,}012$. Die Regression ist ungeeignet!

Die maximale Steigung ergibt sich, wenn man in die Formel zur Berechnung von b den maximalen Wert für die Kovarianz (= Produkt der geschätzten Populationsstreuungen) einsetzt, also $cov_{max} = -2{,}71$ (Achtung: inverser Zusammenhang, daher negative Kovarianz!). Man erhält eine maximale Steigung von $b_{max} = -1{,}88$.

Bei z-standardisierten Variablen entspricht die Steigung der Geraden der Korrelation. Die maximale Steigung zweier (negativ korrelierter) z-standardisierter Variablen beträgt also immer $b = -1$.

o.

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{\hat{\sigma}_y} \cdot \sqrt{\frac{n_0 \cdot n_1}{N^2}} \Rightarrow r_{pb} = 0{,}424$$

p. $r^2_{pb} = 0{,}179 \approx 18\%$

q. Der Psychologe sollte hier eine punktbiseriale Korrelation berechnen.