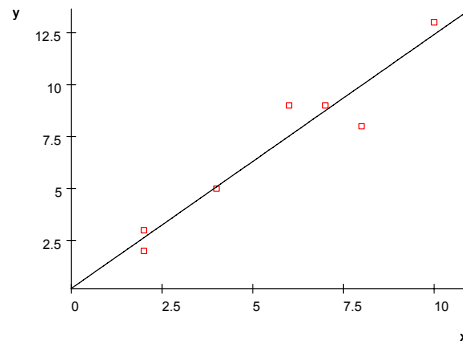


Lösungen zu Übung (10)

- (1) $r(X, Y) = 0.9603$, empirische Regressionsgerade $y = 1.2205x + 0.2$, Bild:



- (2) Ist der durch lineare Regression auf X erklärte Varianzanteil von Y gleich 0.9, so muss $\rho(X, Y)$ eine der Zahlen $\pm\sqrt{0.9}$ sein. Das ist etwa ± 0.949 .
- (3) (a) Es sei $Cov(X, Y) = 10$. Dann ist $Cov(-2X, 3Y) = -2 \cdot 3Cov(X, Y) = -60$.
 (b) Für die Variablen aus a. gelte $\sigma(X) = 4$, $\sigma(Y) = 3$. Dann hat man

$$\rho(X, Y) = \frac{10}{4 \cdot 3} = \frac{5}{6}.$$

$\rho(-2X, 3Y)$ ist also $-\frac{5}{6}$, weil

$$\rho(-2X, 3Y) = \frac{Cov(-2X, 3Y)}{\sigma(-2X)\sigma(3Y)} = \frac{-60}{6 \cdot \sigma(X)\sigma(Y)} = -\rho(X, Y).$$

Es folgt: Der erklärte Varianzanteil ist in beiden Fällen derselbe, nämlich $\rho^2(X, Y) = \rho^2(-2X, 3Y) = \frac{25}{36}$, das macht etwa 69.4%.

- (c) Mit denselben Daten und ferner $\mu(X) = 3$, $\mu(Y) = 10$ hat man:

$$a_{Y,X} = \frac{10}{16} = \frac{5}{8}, \quad b_{Y,X} = 10 - \frac{5}{8} \cdot 3 = \frac{65}{8}, \quad \text{dagegen}$$

$$a_{3Y,-2X} = -\frac{15}{16}, \quad b_{3Y,-2X} = 30 - \frac{-15}{16} \cdot (-6) = \frac{195}{8}$$

Die Gleichung der Regressionsgeraden, welche die beste lineare Darstellung der Y -Werte durch die X -Werte erbringt, lautet also

$$\widehat{Y} = \frac{5}{8}X + \frac{65}{8},$$

die Gleichung für die lineare Regression von $3Y$ auf $-2X$ ist

$$\widehat{3Y} = -\frac{15}{16}(-2X) + \frac{195}{8}. (*)$$

Man sollte einsehen, dass beide Gleichungen dasselbe besagen, dass eben nur die Koeffizienten umzurechnen sind. Man hätte auch allein mit Umformen der ersten Gleichung die zweite korrekt gewonnen, ohne über erneute Berechnung der Koeffizienten zu gehen. Nun könnte man denken, die Regressionsgerade für lineare Regression von $-2X$ auf $3Y$ durch Umformen der letzten Gleichung zu gewinnen, das ergäbe:

$$(\text{fälschlich!}) \widehat{-2X} = -\frac{16}{15}(3Y) + \frac{195}{8}.$$

Aber dies ist nicht die Regressionsgerade, es ist nämlich dieselbe wie (*). (Man beachte, dass unabhängige und abhängige Variable vertauscht auftreten.) **Vielmehr hat man einen Winkel zwischen beiden Regressionsgeraden (bei Vertauschen der Vorhersagerichtung), der nur dann Null ist, wenn der Korrelationskoeffizient ± 1 beträgt!** Rechnen wir die Koeffizienten richtig aus, so erhalten wir

$$a_{-2X.3Y} = \frac{-60}{81} = -\frac{20}{27} \text{ und } b_{-2X.3Y} = -6 - \frac{-20}{27} \cdot 30 = \frac{146}{9}.$$

. Also muss es richtig heißen:

$$\widehat{-2X} = -\frac{20}{27}(3Y) + \frac{146}{9}.$$

(d) Das gesuchte Vertrauensintervall lautet

$$3.8 \cdot \frac{5}{2} \pm 2.58 \cdot 3 \sqrt{1 - \left(\frac{5}{6}\right)^2}, \text{ das ist etwa } [5.221; 13.778].$$

(4) Man hat

$$\begin{aligned} \rho(X+Y, Z) &= \frac{\text{Cov}(X+Y, Z)}{\sigma(X+Y)\sigma(Z)} = \frac{\text{Cov}(X, Z) + \text{Cov}(Y, Z)}{\sqrt{\sigma^2(X) + \sigma^2(Y)}\sigma(Z)} \\ &= \frac{\text{Cov}(X, Z)}{\sqrt{\sigma^2(X) + \sigma^2(Y)}\sigma(Z)} + \frac{\text{Cov}(Y, Z)}{\sqrt{\sigma^2(X) + \sigma^2(Y)}\sigma(Z)} \\ &= \frac{\sigma(X)}{\sqrt{\sigma^2(X) + \sigma^2(Y)}}\rho(X, Z) + \frac{\sigma(Y)}{\sqrt{\sigma^2(X) + \sigma^2(Y)}}\rho(Y, Z). \end{aligned}$$

Wenn $\sigma(X)$ und $\sigma(Y)$ nicht allzu verschieden sind und ebenfalls $\rho(X, Z)$ und $\rho(Y, Z)$ vergleichbar (insbesondere sollten sie dasselbe Vorzeichen haben!), so wird mit linearer Regression von Z auf $X+Y$ mehr Varianz erklärt als durch jede einzelne der Variablen X, Y . Wenn aber $\rho(Y, Z)$ sehr klein ist gegen $\rho(X, Z)$ und $\sigma(Y)$ sehr groß gegen $\sigma(X)$, so wird auch bei gleichem Vorzeichen von $\rho(X, Z)$ und $\rho(Y, Z)$ die Korrelation $\rho(X+Y, Z)$ wesentlich kleiner sein als die Korrelation $\rho(X, Z)$. Beispiel: $\rho(X, Z) = 0.9$, $\rho(Y, Z) = 0.01$, $\sigma(X) = 0.1$, $\sigma(Y) = 10$, dann hat man $\rho(X+Y, Z) = \frac{0.1}{\sqrt{0.1^2+100}} \cdot 0.9 + \frac{10}{\sqrt{0.1^2+100}} \cdot 0.01$, das ist etwa 0.019.