Übung (10)

(1) Bestimmen Sie zu folgenden $(X,Y)-$ Wertepaaren den empirischen Korrelationskoeffizienten $r(X,Y)$ sowie die empirische Regressionsgerade $y = \widehat{a}x + \widehat{b}$: $(2,3)$, $(2,2)$, $(4,5)$, $(6,8)$, $(7,10)$, $(8,8)$, $(10,13)$. Tun Sie das aber nur, wenn Sie das mittels eines Taschenrechners mit L/R-Modus oder vornehmeren Gerätes tun können. Zeichnen Sie die zugehörigen Punkte jedenfalls in ein Koordinatensystem ein, und ziehen Sie intuitiv eine passende Gerade dadurch, und stellen Sie die Gleichung zu dieser Geraden auf. Schauen Sie dann nach, wenn Sie auch die berechnete haben, wie gut das übereinstimmt.

(2) Welche Werte kann $\rho(X,Y)$ nur haben, wenn der Anteil der Varianz von Y, welche durch lineare Regression auf X erklärt wird, 0.9 beträgt?

(3) (a) Es sei $Cov(X,Y) = 10$. Welchen Wert hat $Cov(-2X, 3Y)$?

(b) Für die Variablen aus a. gelte $\sigma(X) = 4$, $\sigma(Y) = 3$. Welchen Wert hat $\rho(X,Y)$? Welchen Wert hat $\rho(-2X, 3Y)$? Was ist der Anteil der Varianz der $Y-$ Werte, der durch Regression auf X erklärt wird? Was ist der Anteil der Varianz der Werte von $3Y$, der durch lineare Regression auf $-2X$ erklärt wird?

(c) Mit den Daten aus a. und b. und ferner $\mu(X) = 3$, $\mu(Y) = 10$: Wie lautet Gleichung der Regressionsgeraden, welche die beste lineare Darstellung der $Y-$Werte durch die $X-$ Werte erbringt? Wie lautet die Gleichung der Regressionsgeraden, welche die Werte von $3Y$ durch die Werte von $-2X$ erklärt? Wie lautet die Gleichung der Regressionsgeraden, welche die Werte von $-2X$ durch die Werte von $3Y$ erklärt?

(d) Unter der idealen Voraussetzungen, dass das Paar (X,Y) einer zweidimensionalen Normalverteilung genügt: Geben Sie das zweiseitige 99%-Vertrauensintervall für den $Y-$ Wert eines Individuums, dessen $X-$ Wert $x_0 = 3.8$ beträgt.

(4) Es seien X, Y linear unabhängige Variablen. Sie kennen $\sigma(X)$ und $\sigma(Y)$ sowie $\rho(X,Z)$ und $\rho(Y,Z)$. Drücken Sie damit $\rho(X+Y,Z)$ aus. Machen Sie sich Gedanken darüber, in welchen Fällen $\rho(X+Y,Z)$ einen größeren Betrag hat als jede der Korrelationen $\rho(X,Z)$ und $\rho(Y,Z)$, so dass Regression von Z auf $X+Y$ einen größeren Anteil der Varianz von Z erklärt als Regression von Z auf jede der Variablen X, Y allein.