10. Übungsblatt

Kovarianz und Korrelationskoeffizient, lineare Regression

1. Sei X eine Zufallsvariable, für die ran $X = \{-1, 0, 1\}$, $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{4}$ und $\mathbb{P}(X = 0) = \frac{1}{2}$ gelten. Sei noch $Y = |X|$. Zeigen wir, dass X und Y unkorreliert aber nicht unabhängig sind.

2. Es gibt drei Papierzettel mit Zahlen 1, 2 und 3 in einer Box. Wir ziehen zwei Zettel zufällig nacheinander (ohne Zurücklegen). Seien X die erste und Y die zweite gezogene Zahl.
   a) $\mathrm{cov}(X, Y) =?$ b) $\mathrm{cov}(X, X) =?$ c) $\mathrm{cov}(Y, Y) =?$ d) $\mathrm{corr}(X, Y) =?$ e) Sind X und Y unabhängig?

3. Zwei Würfel werden geworfen. Sei X die Anzahl der Augenzahlen 1, und sei Y die zweite Augenzahl. Bestimmen wir a) $\mathrm{cov}(X, Y)$, b) $\mathrm{corr}(X, Y)$.

4. Seien X und Y identisch verteilte Zufallsvariablen mit endlichen (und positiven) Varianzen. Zeigen wir, dass $X + Y$ und $X - Y$ unkorreliert sind.

5. Sei X eine Poisson-verteilte Zufallsvariable mit Parameter $\lambda > 0$, und sei noch $Y = 2X+1$. Bestimmen wir die Varianz von Y. Berechnen wir $\mathrm{corr}(X - 1, 3Y + 2)$.

6. Sei X eine Zufallsvariable, für die $\mathbb{E}(X) = 2$, $\mathbb{E}(X^2) = 5$ und $\mathbb{E}(X^3) = 14$ gelten. Berechnen wir $\mathrm{corr}(X, X^2 - 4X + 4)$. Sind $X - 2$ und $(X - 2)^2$ unabhängig?

7. In einer Urne gibt es N rote und M weisse Kugeln. Wir ziehen N Kugeln ohne Zurücklegen. Bestimmen wir die Varianz der Anzahl der gezogenen roten Bällen. (Hinweis: schreiben wir die Anzahl der roten Bällen als eine Summe von Indikatoren. Siehe auch Aufgabe 18 auf dem 3. Übungsblatt.)

---

8. Seien $X \sim U(0; 2\pi)$, $Y = \cos(X)$ und $Z = \sin(X)$. Geben wir $\mathrm{cov}(Y, Z)$ an. Sind Y und Z unabhängig?

9. Sei (X, Y) ein stetiger Zufallsvektor mit Dichtefunktion
$$f_{X,Y}(s,t) = \begin{cases} \dfrac{\alpha}{2s}, & \text{falls } 1 < s < 3 \text{ und } 0 < t < s, \\ 0 & \text{sonst,} \end{cases}$$
wo $\alpha \in \mathbb{R}$ eine geeignete reelle Zahl ist. Bestimmen wir $\mathrm{cov}(X, Y)$.

10. Die Dichtefunktion des Zufallsvektors (U, V) ist
$$f_{U,V} :\ (u, v) \mapsto \begin{cases} \frac{1}{\sqrt{v}} & \text{falls } 0 < u < 1 \text{ und } 0 < v < u^2, \\ 0 & \text{sonst.} \end{cases}$$
Bestimmen wir $\mathrm{cov}(U, V^2)$.

11. Sei (X, Y) ein stetiger Zufallsvektor mit der Dichtefunktion
$$f_{X,Y} : (s,t) \mapsto \begin{cases} \alpha s + \beta t & \text{falls } 0 < s < 2 \text{ und } 0 < t < 2, \\ 0 & \text{ansonsten.} \end{cases}$$
Nehmen wir an, dass $\mathrm{cov}(X, Y) = -\frac{1}{36}$ gilt. Bestimmen wir die Werte von α und β, und berechnen wir die Wahrscheinlichkeit $\mathbb{P}(X < 1)$.

---

12. Ein regelmäßiger Würfel wird zweimal geworfen. Sei X die Anzalh der Sechsen, und sei Y der Anzahl der geraden Ergebnisse. Bestimmen wir die Regressionsgerade von X auf Y und die Regressionsgerade von Y auf X. Berechnen wir den durchschnittlichen quadratischen Fehler der Approximation von Y durch die lineare Regression $\beta X + \alpha$ (also den Erwartungswert des Quadrats der Differenz von Y und $\beta X + \alpha$).

13. Seien X und Y die in der zweiten Aufgabe definierten Variablen. Bestimmen wir die Regressionsgerade von Y auf X. Berechnen wir den durchschnittlichen quadratischen Fehler der Approximation von Y durch die lineare Regression $\beta X + \alpha$.

14. Nehmen wir an, dass die Regressionsgerade von Y auf X durch $\left\{(x, y) \in \mathbb{R}^2 \mid y = x - 2\right\}$ angegeben ist, und dass die Regressionsgerade von X auf Y durch $\left\{(y, x) \in \mathbb{R}^2 \mid x = \frac{1}{4}y - \frac{11}{2}\right\}$ angegeben ist. Bestimmen wir die Werte $\mathbb{E}(X)$ und $\mathbb{E}(Y)$.

15. Nehmen wir an, dass X_1, X_2 und X_3 unkorrelierte Zufallsvariablen sind, und dass $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3) = 0$ und $\mathbb{D}(X_1) = \mathbb{D}(X_2) = \mathbb{D}(X_3) = 1$ gelten.

    a) Berechnen wir $\operatorname{cov}(X_1 + X_2, X_2 + X_3)$, $\mathbb{D}^2(X_1 + X_2)$ und $\mathbb{D}^2(X_2 + X_3)$.
    b) Bestimmen wir $\operatorname{corr}(X_1 + X_2, X_2 + X_3)$.
    c) Geben wir die Regressionsgerade von $X_2 + X_3$ auf $X_1 + X_2$ an.
    d) Was ist die Varianz des Fehlers der Approximation von $X_2 + X_3$ durch die obige Regressionsgerade?

16. Zwei reelle Zahlen werden uniform zufällig und unabhängig voneinander im Intervall $(0; 1)$ gewählt. Seien X das Maximum und Y das Minimum dieser Zahlen. Bestimmen wir die lineare Regression

    a) von X auf Y, b) von Y auf X.

17. Ein Experiment wird n-mal nacheinander durchgeführt, die einzelnen Experimente sind unabhängig. Sei X die Anzahl der Eintritte eines Ereignisses, das bei jedem Experiment mit Wahrscheinlichkeit p eintritt. Sei noch Y die Anzahl der Eintritte eines anderen Ereignisses mit Wahrscheinlichkeit q. Bestimmen wir die lineare Regression von Y auf X, wenn die Wahrscheinlichkeit dafür, dass beide Ereignisse gleichzeitig bei einem Experiment eintreten, r ist. Berechnen wir auch die Varianz des Fehlers der Approximation.

18. Geben wir die lineare Regression von $X^2 + X + 1$ auf X an, falls a) $X \sim \operatorname{Exp}(5)$, b) $X \sim N(0; 1)$, *c) $X \sim \operatorname{Bin}(n; p)$ gilt.