Stochastik

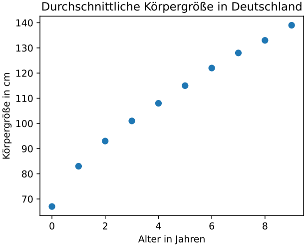
FH Aachen – Studienstandort Aachen
7. November 2025

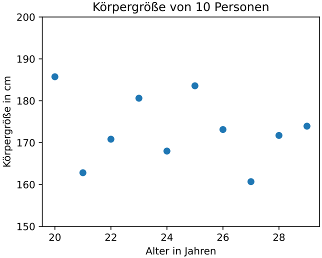
Prof. Dr. Florian Heinrichs

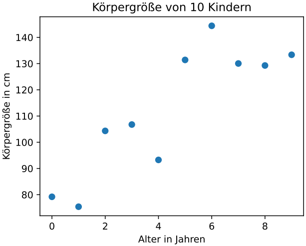
Kovarianz & Korrelation

Abhängigkeit

> Wie können wir Abhängigkeit von Zufallsvariablen messen?
> Bedingte Verteilung $\rightarrow$ kompliziert
> Besser: Eine Kennzahl
> Beispiel: Alter und Körpergröße







Kovarianz & Korrelation

Definition 49

Seien X, Y Zufallsvariablen mit existierenden Varianzen. Dann ist die *Kovarianz* $\mathsf{cov}(X, Y)$ definiert als

$$\mathsf{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

Falls die Varianzen positiv sind, ist der *Korrelationskoeffizient* $\rho_{X,Y}$ definiert als

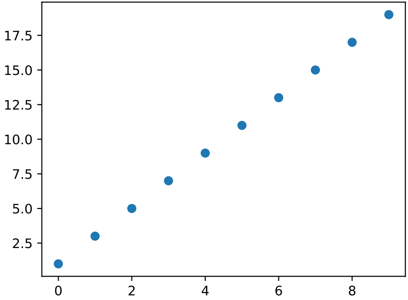
$$\rho_{X,Y} = \frac{\mathsf{cov}(X, Y)}{\sqrt{\mathsf{var}(X)}\sqrt{\mathsf{var}(Y)}}.$$

Falls $\rho_{X,Y} = 0$, heißen X und Y *unkorreliert*.

Kovarianz & Korrelation

Beispiel 118

- Sei $X \sim \mathcal{U}_{\{1,\dots,10\}}$ und $Y = 2X + 1$
- Dann ist $\mathbb{E}[X] = \frac{11}{2}$ und $\mathbb{E}[Y] = 2\mathbb{E}[X] + 1 = 12$
- Weiter gilt $\mathrm{var}(X) = \frac{10^2-1}{12} = \frac{33}{4}$ und $\mathrm{var}(Y) = 2^2\mathrm{var}(X) = 33$



Kovarianz & Korrelation

Beispiel 118: (Fortsetzung)

- Sei $X \sim \mathcal{U}_{\{1,\dots,10\}}$ und $Y = 2X + 1$
- Dann ist $\mathbb{E}[X] = \frac{11}{2}, \mathbb{E}[Y] = 12, \mathrm{var}(X) = \frac{33}{4}$ und $\mathrm{var}(Y) = 33$
- Damit folgt

$$\mathrm{cov}(X,Y) = \sum_{x,y} \big(x - \tfrac{11}{2}\big)\big(y - 12\big) p(x,y) = \tfrac{1}{10} \sum_{k=1}^{10} \big(k - \tfrac{11}{2}\big)\big(2k - 11\big)$$

$$= \tfrac{1}{5} \sum_{k=1}^{10} \big(k - \tfrac{11}{2}\big)^2 = \tfrac{1}{5} \sum_{k=1}^{10} k^2 - 11k + \big(\tfrac{11}{2}\big)^2$$

$$= \tfrac{1}{5}\big(\tfrac{10\cdot 11\cdot 21}{6} - \tfrac{11\cdot 10\cdot 11}{2} + \tfrac{10\cdot 11\cdot 11}{4}\big) = \tfrac{33}{2}$$

und

$$\rho_{X,Y} = \frac{33/2}{\sqrt{33/4}\sqrt{33}} = 1$$

Kovarianz & Korrelation

Übung 65

- Sei $X \sim \mathcal{U}_{\{-2,-1,0,1,2\}}$ und $Y = X^2$. Berechnen Sie $\text{cov}(X, Y)$.
- Sei $X \sim \mathcal{U}_{\{1,2,3,4\}}$ und $Y = X^2$. Berechnen Sie $\text{cov}(X, Y)$.

Kovarianz & Korrelation

Die Berechnung von Kovarianz & Korrelationskoeffizienten wird schnell kompliziert $\rightarrow$ Wir brauchen Rechenregeln!

Satz 28

Für Zufallsvariablen X und Y gilt:

1. $\text{cov}(X, X) = \text{var}(X)$
2. $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
3. X, Y unabhängig $\implies \text{cov}(X, Y) = 0$

Achtung: Die Rückrichtung in 3. ist im Allgemeinen falsch
Beweis:

1. Folgt direkt aus der Definition
2. Folgt analog zu $\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
3. X, Y unabhängig $\implies \mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \overset{2.}{\implies} \text{cov}(X, Y) = 0$

Kovarianz & Korrelation

Übung 66

Seien X, Y, Z Zufallsvariablen und $a \in \mathbb{R}$. Beweisen Sie die folgenden Rechenregeln

1. $\mathrm{cov}(X + a, Y) = \mathrm{cov}(X, Y)$
2. $\mathrm{cov}(X + Y, Z) = \mathrm{cov}(X, Z) + \mathrm{cov}(Y, Z)$

Kovarianz & Korrelation

Beispiel 119

Beim zweifachen Würfelwurf sei X die Augenzahl des ersten Würfels und Y die Summe der beiden Augenzahlen. Was ist $\text{cov}(X, Y)$?

> Wir wissen bereits: $\text{var}(X) = \frac{35}{12}$

> Sei Z die Augenzahl beim zweiten Wurf, dann gilt $Y = X + Z$

> $\text{cov}(X, Y) = \text{cov}(X, X + Z) = \text{cov}(X, X) + \text{cov}(X, Z) = \text{var}(X) = \frac{35}{12}$

Kovarianz & Korrelation

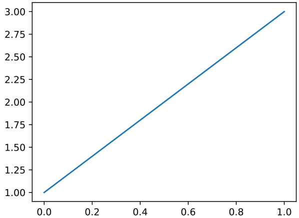
Beispiel 120

Sei $X \sim \mathcal{U}_{[0,1]}$ und $Y = X^2$. Es gilt

$$\mathbb{E}[X^k] = \int_0^1 x^k \mathrm{d}x = \left[\tfrac{1}{k+1} x^{k+1}\right]_0^1 = \tfrac{1}{k+1}$$

Damit folgt

$$\text{cov}(X,Y) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = \tfrac{1}{4} - \tfrac{1}{2} \cdot \tfrac{1}{3} = \tfrac{1}{12}$$



Kovarianz & Korrelation

Beispiel 121

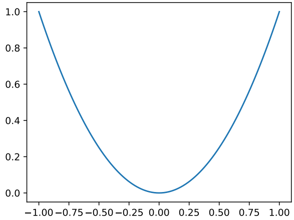
Sei $X \sim \mathcal{U}_{[-1,1]}$ und $Y = X^2$.
Für k ungerade gilt

$$\mathbb{E}[X^k] = \int_{-1}^{1} x^k \mathrm{d}x = 0$$

Damit folgt

$$\text{cov}(X,Y) = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0.$$

X und Y sind unkorreliert, aber offensichtlich **nicht** unabhängig.



Kovarianz & Korrelation

Übung 67

1. Sei X eine Zufallsvariable mit Varianz var$(X) = 1$. Weiter sei $Y = 2X + 1$. Berechnen Sie var(Y), cov(X, Y) und anschließend $\rho_{X,Y}$.
2. Sei X eine Zufallsvariable mit Varianz var$(X) = \sigma^2$. Weiter sei $Y = aX + b$ für $a, b \in \mathbb{R}$. Berechnen Sie var(Y), cov(X, Y) und anschließend $\rho_{X,Y}$.
3. Seien X und Y unabhängige Zufallsvariablen mit var$(X) =$ var$(Y) = 1$. Berechnen Sie cov$(X, X + Y)$ und var$(X + Y)$.

Kovarianz & Korrelation

Satz 29

Seien $X_1, \ldots, X_n$ Zufallsvariablen. Dann gilt

$$\text{var}(X_1 + ... + X_n) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

Falls die Zufallsvariablen unabhängig sind, folgt $\text{var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{var}(X_i)$.

Spezialfall $(n = 2)$: $\text{var}(X + Y) = \text{var}(X) + 2\text{cov}(X, Y) + \text{var}(Y)$

Beweis:

$$\text{var}\left(\sum_{i=1}^{n} X_i\right) = \text{cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \text{cov}(X_i, X_j)$$

$$= \sum_{i=1}^{n} \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j).$$

Kovarianz & Korrelation

Satz 30: (Cauchy-Schwarz Ungleichung)

Seien X und Y Zufallsvariablen mit $\mathbb{E}[X^2] < \infty$ und $\mathbb{E}[Y^2] < \infty$. Dann gilt

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}.$$

> Aus der Cauchy-Schwarz Ungleichung folgt

$$|\mathrm{cov}(X, Y)| \leq \sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}$$

und $-1 \leq \rho_{X,Y} \leq 1$.

> $\rho_{X,Y} = 1 \iff Y = aX + b$ für $a > 0, b \in \mathbb{R}$

> $\rho_{X,Y} = -1 \iff Y = aX + b$ für $a < 0, b \in \mathbb{R}$

Cauchy-Schwarz Ungleichung

Optional 7: (Beweis der Cauchy-Schwarz Ungleichung)

Falls $X = 0$ gilt $\mathbb{E}[XY] = 0 = \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$. Sei $X \neq 0$. Dann gilt

$$\begin{aligned}
0 &\leq \mathbb{E}\left[\left(Y - \tfrac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}X\right)^2\right] \\
&= \mathbb{E}[Y^2] - 2\mathbb{E}[XY]\tfrac{\mathbb{E}[XY]}{\mathbb{E}[X^2]} + \left(\tfrac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}\right)^2\mathbb{E}[X^2] \\
&= \mathbb{E}[Y^2] - 2\tfrac{(\mathbb{E}[XY])^2}{\mathbb{E}[X^2]} + \tfrac{\mathbb{E}[XY]^2}{\mathbb{E}[X^2]} \\
&= \mathbb{E}[Y^2] - \tfrac{\mathbb{E}[XY]^2}{\mathbb{E}[X^2]}.
\end{aligned}$$

Damit folgt $\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$.

Literatur I

Bundesministerium für Gesundheit (2024).
Infektionsradar.
https://infektionsradar.gesund.bund.de/de/covid/inzidenz.
Abgerufen: 2024-10-15.

Dehling, H. and Haupt, B. (2006).
Einführung in die Wahrscheinlichkeitstheorie und Statistik.
Springer-Verlag.

FH AACHEN
UNIVERSITY OF APPLIED SCIENCES

Literatur II

Dinnes, J., Sharma, P., Berhane, S., van Wyk, S., Nyaaba, N., Domen, J., Taylor, M., Cunningham, J., Davenport, C., Dittrich, S., Emperador, D., Hooft, L., Leeflang, M., McInnes, M., Spijker, R., Verbakel, J., Takwoingi, Y., Taylor-Phillips, S., Van den Bruel, A., and Deeks, J. (2022).
Rapid, point-of-care antigen tests for diagnosis of sars-cov-2 infection.
Cochrane Database of Systematic Reviews, (7).

Henze, N. et al. (1997).
Stochastik für Einsteiger, volume 4.
Springer.