

Lernzettel

Mehrdimensionale Daten: Kovarianz,
Korrelationsmatrix und Überblick über
multivariate Beschreibung

Universität: Technische Universität Berlin
Kurs/Modul: Statistik I für Wirtschaftswissenschaften
Erstellungsdatum: September 19, 2025



Zielorientierte Lerninhalte, kostenlos!
Entdecke zugeschnittene Materialien für deine Kurse:

<https://study.AllWeCanLearn.com>

Statistik I für Wirtschaftswissenschaften

Lernzettel: Mehrdimensionale Daten: Kovarianz, Korrelationsmatrix und Überblick über multivariate Beschreibung

(1) Begriffe und Notation.

Daten in mehreren Variablen werden als Vektor beobachtet:

$$\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p.$$

Die gemeinsame Verteilung wird durch die Dichte $f_{\mathbf{X}}(\mathbf{x})$ beschrieben. Die Erwartung von $\mathbf{X}$ ist der Vektor der Mittelwerte

$$\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_p])^\top.$$

(2) Kovarianz.

Die Kovarianz zwischen den Komponenten ist definiert als

$$\text{Cov}(X_i, X_j) = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)], \quad i, j = 1, \dots, p.$$

Der Vektor bzw. die Matrix der Kovarianzen heißt Kovarianzmatrix

$$\boldsymbol{\Sigma} = \text{Cov}(\mathbf{X}) = [\text{Cov}(X_i, X_j)]_{i,j=1}^p.$$

Eigenschaften:

- $\boldsymbol{\Sigma}$ ist symmetrisch.
- Für jeden Vektor $\mathbf{a} \in \mathbb{R}^p$ gilt $\mathbf{a}^\top \boldsymbol{\Sigma} \mathbf{a} \geq 0$ (positiv semidefinit).

(2a) Stichprobenkovarianz.

Aus Stichproben x_{ki} ($k=1, \dots, n$) ergibt sich die Schätzmatrix

$$S_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), \quad i, j = 1, \dots, p,$$

mit $\bar{x}_i = \frac{1}{n} \sum_{k=1}^n x_{ki}$.

(3) Korrelationsmatrix.

Die Korrelationen zwischen den Variablen sind definiert durch

$$\rho_{ij} = \text{Corr}(X_i, X_j) = \frac{\text{Cov}(X_i, X_j)}{\sigma_i \sigma_j}, \quad \sigma_i = \sqrt{\text{Var}(X_i)} = \sqrt{\sigma_{ii}}.$$

Die Korrelationsmatrix ist

$$\mathbf{R} = [\rho_{ij}]_{i,j=1}^p.$$

Eigenschaften:

- $\mathbf{R}$ ist symmetrisch und besitzt auf der Diagonalen die Einsen.
- $\mathbf{R} = \mathbf{D}^{-1} \boldsymbol{\Sigma} \mathbf{D}^{-1}$ mit $\mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_p)$.

(4) Überblick über die multivariate Beschreibung.

- Gemeinsame Beschreibung durch die Verteilung von $\mathbf{X}$ sowie Marginal- und Bedingungsverteilungen.
- Falls $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, besitzt $\mathbf{X}$ eine multivariate Normalverteilung mit Dichte

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

Lineare Transformationen:

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}, \quad \mathbf{Y} \sim \mathcal{N}_p(A\boldsymbol{\mu} + \mathbf{b}, A\boldsymbol{\Sigma}A^\top).$$

Standardisierung:

$$\mathbf{Z} = \mathbf{D}^{-1}(\mathbf{X} - \boldsymbol{\mu}), \quad \text{Cov}(\mathbf{Z}) = \mathbf{R}, \quad \mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_p).$$

Grafische Perspektiven: Streudiagramm-Paare, Interpretation von Korrelation und Kovarianz, Ellipsen-Darstellungen der Konturen.