

## Topic 2: Multikollinearitaet

### Block 3: Komplexitaet, Kausalitaet & Generalisierung

January 25, 2026

# Was ist Multikollinearitaet?

- **Definition:** Praediktoren sind stark miteinander korreliert
- Keine eindeutige Trennung der Effekte einzelner Variablen
- **Symptome:**
  - Hohe  $R^2$ , aber keine signifikanten Koeffizienten
  - Instabile Koeffizienten bei kleinen Datenaenderungen
  - Grosse Standardfehler der Koeffizienten
  - Vorzeichen der Koeffizienten widersprechen der Theorie
- Beispiel: Wohnflaeche und Anzahl Zimmer in Immobilienpreismodell

---

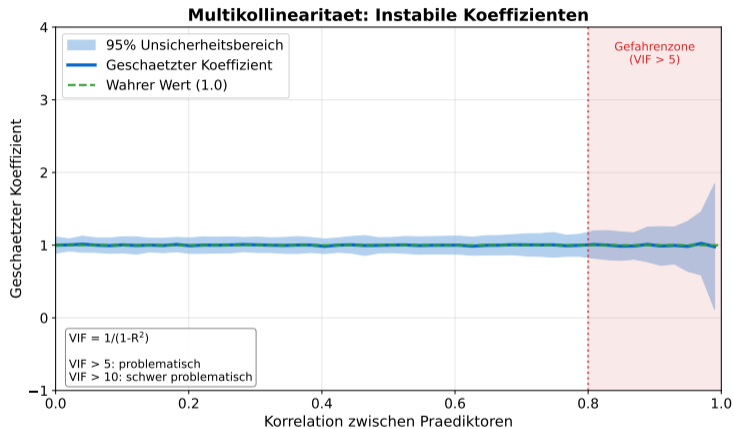
Multikollinearitaet erschwert die Interpretation einzelner Praediktoren

## Diagnose: Variance Inflation Factor (VIF)

- **VIF misst:** Wie stark Variable  $j$  mit anderen Prädiktoren korreliert
- **Formel:**  $VIF_j = \frac{1}{1-R_j^2}$ 
  - $R_j^2$ : Bestimmtheitsmass aus Regression von  $X_j$  auf alle anderen Prädiktoren
- **Interpretation:**
  - $VIF = 1$ : Keine Korrelation mit anderen Variablen
  - $VIF > 5$ : Problematische Multikollinearität
  - $VIF > 10$ : Schwerwiegende Multikollinearität
- Standardfehler wird um Faktor  $\sqrt{VIF_j}$  aufgebläht

---

VIF quantifiziert den Grad der Multikollinearität fuer jede Variable



Mit steigender Korrelation waechst VIF exponentiell

1. **Variable entfernen:** Eine der korrelierten Variablen weglassen
  - Einfachste Loesung bei redundanten Praediktoren
2. **Variablen kombinieren:** Index oder Durchschnitt bilden
  - Beispiel: Durchschnitt aus mehreren Leistungskennzahlen
3. **Regularisierung:** Ridge oder Lasso Regression
  - Schrumpft Koeffizienten, stabilisiert Schaetzungen
4. **Principal Component Analysis (PCA):** Orthogonale Komponenten
5. **Mehr Daten sammeln:** Erhoeht Variation, reduziert Korrelationen

---

Die Wahl der Loesung haengt von Forschungsfrage und Datenlage ab

## Interpretation:

- Multikollinearität ist **problematisch**
- Kausale Effekte nicht trennbar
- Koeffizienten instabil
- $\Rightarrow$  Variable entfernen oder kombinieren

**Faustregel:** Forschungsziel bestimmt Umgang mit Multikollinearität

## Vorhersage:

- Multikollinearität ist **weniger problematisch**
- Vorhersagegenauigkeit bleibt hoch
- Wichtig: Generalisierung (Overfitting vermeiden)
- $\Rightarrow$  Regularisierung bevorzugen

---

Multikollinearität schadet Interpretation, nicht unbedingt Vorhersage