

## Block 3: Zusammenfassung

Komplexität, Kausalität & Generalisierung

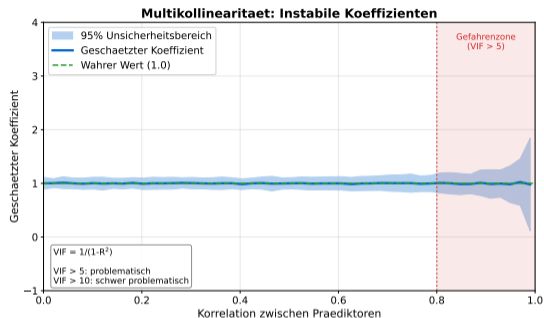
January 11, 2026

# 1. BLUE und Multikollinearität

## BLUE-Eigenschaften:

- Best Linear Unbiased Estimator
- Gilt unter idealen Bedingungen
- Multikollinearität verletzt dies

**VIF > 5 problematisch!**



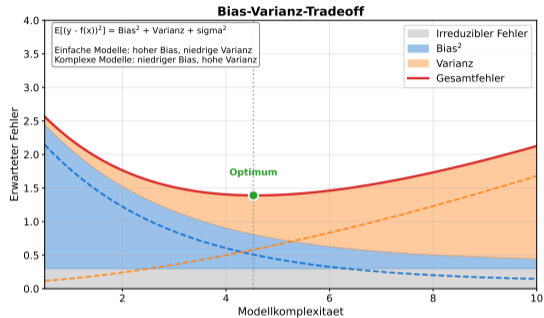
Multikollinearität = korrelierte Prädiktoren = instabile Koeffizienten.

## 2. Bias-Varianz-Tradeoff

**Bias:** Systematischer Fehler (Modell zu einfach)

**Varianz:** Zufällige Schwankung (Modell zu komplex)

**Optimum:** Minimaler Gesamtfehler



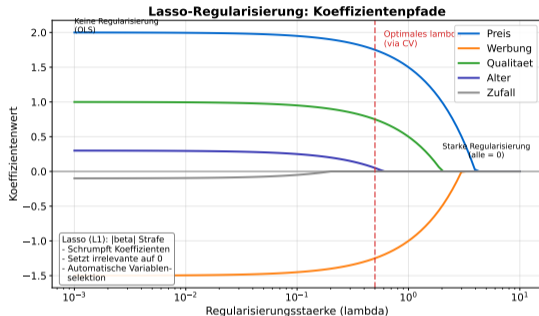
## 3. Regularisierung

### Lasso (L1):

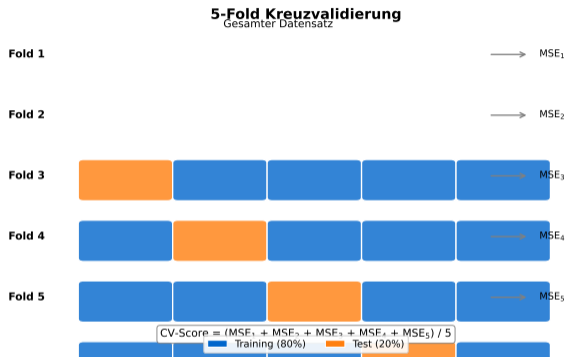
- Setzt Koeffizienten auf 0
- Automatische Variablenselektion
- Gut bei vielen irrelevanten Features

### Ridge (L2):

- Schrumpft, aber entfernt nicht
- Gut bei Multikollinearität

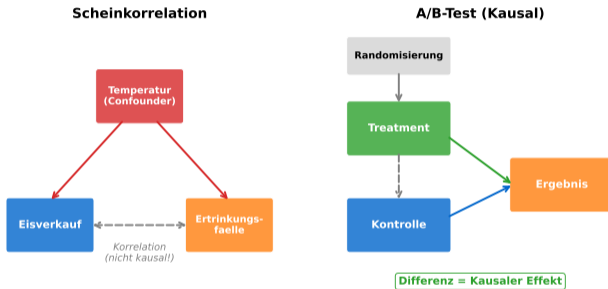


Lambda wird per Kreuzvalidierung gewaehlt.



**CV-Score = Durchschnitt über alle Folds – robuste Out-of-Sample Schätzung**

### Korrelation vs. Kausalität



**Nur Experimente (A/B-Tests) erlauben kausale Schlussfolgerungen!**

### **Randomisierung macht Gruppen vergleichbar:**

1. Zufällige Zuteilung zu A (Kontrolle) oder B (Treatment)
2. Intervention nur in B
3. Messung des Ergebnisses
4. Differenz = Kausaler Effekt

**Cohen's d:** Effektstärke – 0.2 klein, 0.5 mittel, 0.8 gross

---

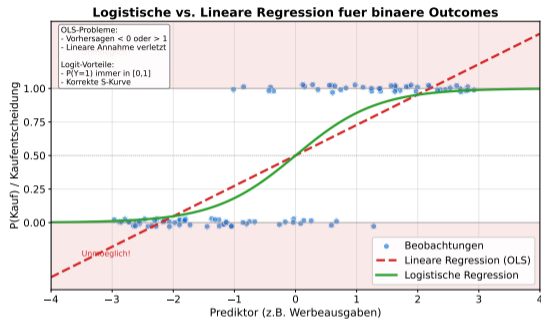
A/B-Tests sind Standard für kausale Fragen im digitalen Business.

## 7. Logistische Regression

Für binaere Outcomes:

$$P(Y = 1) = \frac{1}{1 + e^{-X\beta}}$$

- Garantiert P in [0,1]
- Koeffizienten als log-Odds
- $e^{\beta} = \text{Odds Ratio}$



## 8. Logit Anwendungen

### **Business Use Cases:**

- Churn Prediction (Kündigung ja/nein)
- Conversion (Kauf ja/nein)
- Credit Scoring (Rueckzahlung ja/nein)
- Fraud Detection (Betrug ja/nein)

**Metriken:** Accuracy, Precision, Recall, F1, AUC-ROC

---

Logit ist die Basis für viele Klassifikationsaufgaben.

### Konzepte verstanden?

- BLUE und Annahmen
- Multikollinearität (VIF)
- Bias-Varianz-Tradeoff
- Lasso vs. Ridge
- Kreuzvalidierung
- Korrelation vs. Kausalität
- A/B-Test Logik
- Logit vs. OLS

**Ausblick:** Modul ER017 – Machine Learning Algorithmen

### R-Skills anwendbar?

- VIF berechnen (car)
- glmnet für Lasso
- caret für CV
- A/B-Test auswerten
- glm() für Logit
- Odds Ratios berechnen