

Topic 5: Multiple Regression

Block 1: Data Science Grundlagen

Erweitertes Regressionsmodell:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Warum mehrere Variablen?

- Mehr Varianz erklären: Einzelne Prädiktoren oft unzureichend
- Störvariablen kontrollieren: Konfundierende Effekte isolieren
- Realitätsnähe: Komplexe Phänomene haben viele Einflussfaktoren

Ceteris-paribus-Konzept:

Jeder Koeffizient β_j misst den Effekt von X_j , wenn alle anderen Variablen konstant gehalten werden.

Multiple Regression erlaubt die simultane Berücksichtigung mehrerer Einflussfaktoren

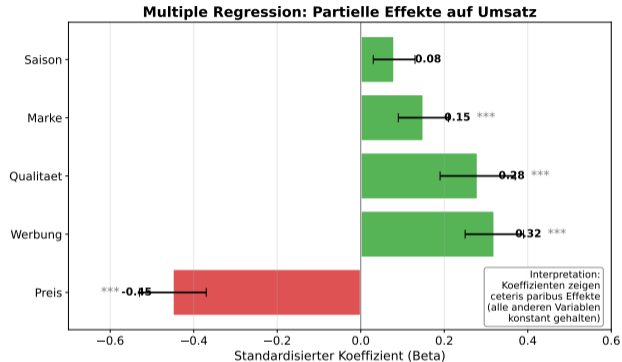
Interpretation der Koeffizienten:

- β_j : Effekt einer Variable, wenn alle anderen konstant
- Partielle Steigung: Isolierter Einfluss
- Kontrolle für Konfundierung

Beispiel:

$$\begin{aligned} \text{Umsatz} &= \beta_0 + \beta_1 \text{Preis} \\ &+ \beta_2 \text{Werbung} \\ &+ \beta_3 \text{Qualität} + \varepsilon \end{aligned}$$

β_1 : Preiseffekt bei gleicher Werbung und Qualität



Partielle Effekte isolieren den Einfluss einzelner Variablen unter Kontrolle aller anderen

Modell mit Interaktion:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \varepsilon$$

Interpretation:

- β_3 : Interaktionseffekt – der Effekt von X_1 hängt von X_2 ab
- Additive Modelle unterstellen unabhängige Effekte
- Interaktionen modellieren Synergien oder Abschwächungen

Beispiel: Preiselastizität nach Marke

- **Premium-Marke**: Preis hat weniger negativen Effekt
- **Budget-Marke**: Preis hat stärkeren negativen Effekt
- Interaktion: Preis \times Marke

Interaktionsterme erfassen nicht-additive Effekte zwischen Variablen

Wichtige Statistiken im Überblick:

Statistik	Bedeutung
Koeffizient (β_j)	Effektstärke (partielle Steigung)
Std. Error	Unsicherheit der Schätzung
t-Wert	Koeffizient / Std. Error
p-Wert	Signifikanz (H: $\beta_j = 0$)
R^2	Anteil erklärter Varianz
F-Statistik	Globale Modellgüte

Typische Fragen:

- Welche Variablen haben signifikante Effekte? (p-Wert)
- Wie stark ist der Effekt? (Koeffizient, nicht nur Signifikanz!)
- Wie gut erklärt das Modell die Daten? (R^2 , Residualplot)

Wichtig: Signifikanz \neq praktische Relevanz!

Ein vollständiger Regressionsoutput liefert Effektstärken, Unsicherheit und Modellgüte