

Topic 4: Einfache Lineare Regression

Grundformel der einfachen linearen Regression:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Komponenten:

- Y : Abhängige Variable (zu erklärende Variable)
- X : Unabhängige Variable (erklärende Variable)
- β_0 : Achsenabschnitt (Intercept) – Wert von Y wenn $X = 0$
- β_1 : Steigung (Slope) – Änderung von Y bei Erhöhung von X um 1
- ε : Fehlerterm (Residuum) – nicht erklärter Teil

Beispiel: Umsatz = 1000 + 50 × Werbung

- Ohne Werbung: 1000 Euro Umsatz
- Pro 1 Euro Werbung: 50 Euro mehr Umsatz

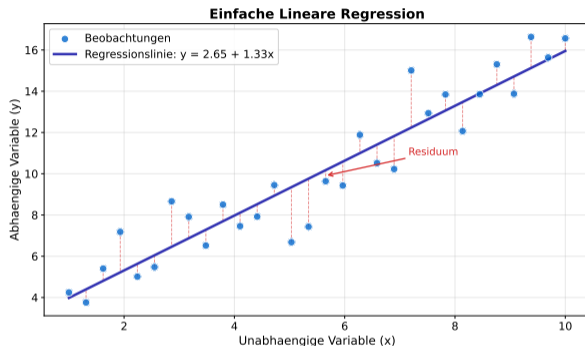
Lineare Regression modelliert den linearen Zusammenhang zwischen zwei Variablen

Methode der kleinsten Quadrate (OLS):

- Findet die beste Regressionsgerade
- Minimiert die Summe der quadrierten Residuen
- $\min \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- Residuum: Abstand vom Punkt zur Geraden

Eigenschaften:

- Eindeutige Lösung
- Minimiert Vorhersagefehler
- Standard in der Praxis



OLS findet die Gerade, die den Gesamtfehler minimiert

Definition:

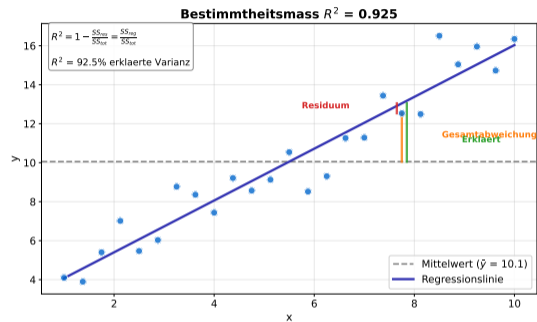
$$R^2 = \frac{\text{Erklärte Varianz}}{\text{Gesamtvarianz}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

Interpretation:

- Wertebereich: 0 bis 1
- $R^2 = 0$: Modell erklärt nichts
- $R^2 = 0.5$: 50% der Varianz erklärt
- $R^2 = 1$: Perfekte Anpassung

Anteil der durch das Modell erklärten Varianz

R^2 misst die Güte der Modellanpassung an die Daten



Was sagt R² aus?

- $R^2 = 0.70$ bedeutet: 70% der Varianz von Y wird durch X erklärt
- Kontextabhängige Bewertung notwendig

Typische Richtwerte (sehr kontextabhängig):

- **Sozialwissenschaften:** $R^2 > 0.3$ bereits gut
- **Business/Marketing:** $R^2 = 0.2 - 0.5$ typisch
- **Naturwissenschaften:** $R^2 > 0.8$ oft erwartet

Was sagt R² NICHT aus?

- **Kausalität:** Hohe Korrelation \neq Kausalität
- **Korrektheit:** Modell kann falsch spezifiziert sein
- **Praktische Relevanz:** Statistisch signifikant \neq praktisch wichtig

Warnung: Mehr Variablen erhöhen immer R^2 – verwenden Sie adjustiertes R^2 !

R^2 ist nützlich, aber nicht das einzige Gütekriterium eines Modells