

Zusammenfassung Statistik FS14

VL1

Matrizen

Matrix = Container, der ähnliche Elemente kombiniert

- rechteckige Darstellung von Zahlen: spezifische Anordnung der Zahlen ist von Bedeutung

Order/Anordnung: Reihen vs. Spalten einer Matrix

→ 3 x 2 – Matrix = 3 Reihen und 2 Spalten

Elemente = jede einzelne Zahl in einer Matrix

a_{ij}
i = Reihe/Zeile
j = Spalte

Verschiedene Typen von Matrizen

Skalar:

- Matrix mit einem einzelnen Element
- 1 x 1 – Matrix

Square Matrix = Quadratische Matrix

- gleich viele Zeilen wie Spalten
- p x p – Matrix

Symmetrische Matrix

- Quadratische Matrix, wo jeweils gilt $a(ij) = a(ji)$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 3 \end{bmatrix}$$

Identity Matrix = Einheitsmatrix

- $I(p)$
- p x p – Matrix mit 1 auf der Diagonalen und 0 überall sonst
- Wird benutzt wie die Konstante „1“ in normaler Algebra

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Vektor

- Matrix mit nur einer Zeile/Spalte
- Beschriftung: Kleinbuchstabe
- Zeilenvektor $\mathbf{a} = [1 \ 2 \ 3]$

- Spaltenvektor (column) → Normalfall! $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

Blockmatrix

- Elemente der Matrizen sind selber Matrizen oder Vektoren

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}$$

Mathematische Operationen

Unterschied mathematische Operationen Matrizen vs. Ziffern

- gewisse Operationen gibt es nur für Matrizen
- Gewisse Operationen gibt es für Matrizen in verschiedenen Formen
- Addition/Multiplikation kann bei Matrizen nicht immer angewendet werden
- Reihenfolge bei Multiplikation spielt bei Matrizen eine Rolle

Transponieren

- $p \times q$ – Matrix umwandeln in $q \times p$ – Matrix
- Aus A wird A' (oder A^T)

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{A}^T = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

Addition

- nur möglich, wenn die beiden Matrizen dieselbe Ordnung haben
- alle Elemente des gleichen Ortes addieren

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 1+1 & 2+0 & 3+1 \\ 3+0 & 2+1 & 1+0 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 4 \\ 3 & 3 & 1 \end{bmatrix}$$

Subtraktion

- nur möglich, wenn die zwei Matrizen dieselbe Ordnung haben

Ausserdem:

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$

$$(\mathbf{A} - \mathbf{B})^T = \mathbf{A}^T - \mathbf{B}^T$$

Multiplizieren/Produkte:

1) Skalarmultiplikation: Multiplikation der Matrix mit einem Skalar (Konstante)

- $k \times A = B \rightarrow B$ hat gleiche Ordnung wie A , jedes Element wird mit k multipliziert

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

$$k \cdot \mathbf{A} = \begin{bmatrix} 2 \times 1 & 2 \times 2 & 2 \times 3 \\ 2 \times 3 & 2 \times 2 & 2 \times 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 \\ 6 & 4 & 2 \end{bmatrix}$$

2) Inneres Produkt zweier Vektoren

- zwei Vektoren müssen dieselbe Anzahl Elemente haben
 - man braucht aber einen Reihenvektor und einen Spaltenvektor
- Wenn dies nicht der Fall ist, muss der erste Vektor transponiert werden
- Ergebnis: Skalar; ein einzelner Wert

$$\mathbf{a} = [1 \ 2 \ 3] \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{a} \cdot \mathbf{b} = 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 = 4$$

3) Produkt zweier Matrizen

- A muss gleich viele Spalten haben wie B Zeilen hat
 - A: m x p; B: p x q; Produkt AB = m x q = C
- c(ij) = Inneres Produkt der i. Reihe in A mit der j. Spalte in B

→ Anzahl Spalten des ersten Faktors = Anzahl Reihen des zweiten Faktors!

Ergebnis: Anzahl Reihen des ersten Faktors * Anzahl Spalten des zweiten Faktors

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} 1 \cdot 1 + 2 \cdot 0 + 3 \cdot 1 & 1 \cdot 0 + 2 \cdot 1 + 3 \cdot 0 \\ 3 \cdot 1 + 2 \cdot 0 + 1 \cdot 1 & 3 \cdot 0 + 2 \cdot 1 + 1 \cdot 0 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 2 \\ 4 & 2 \end{bmatrix} \end{aligned}$$

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$$

Äusseres Produkt zweier Vektoren

- wenn ein Spaltenvektor mit einem Zeilenvektor multipliziert wird (nicht umgekehrt)
- Ergebnis wie bei Produkt zweier Matrizen

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} [1 \ 2 \ 3] = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

Produkt Matrix mit Einheitsmatrix

- Einheitsmatrix wirkt wie eine Konstante 1

$$\begin{aligned} \mathbf{AI} &= \mathbf{A} \\ \mathbf{IA} &= \mathbf{A} \end{aligned}$$

Divison

Gibt es nicht!

Inverse

Inverse einer Matrix $A \rightarrow A^{-1}$

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \quad \mathbf{I} = \text{Einheitsmatrix}$$

Bedingungen:

- A muss quadratische Matrix sein
- A muss vollrangig sein

Berechnung:

solve ex

Rechnungen mit Inversen:

- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
- $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$.
- $(k\mathbf{A})^{-1} = k^{-1}\mathbf{A}^{-1}$.
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Lösung Gleichungssysteme:

Bsp.

$$\begin{aligned} a + 2b - c &= -2 \\ -a + 0b + 3c &= 4 \\ 3a + 3b + c &= -2 \end{aligned}$$

Spaltenvektor p mit allen Unbekannten Spaltenvektor y mit allen „Ergebnissen“

$$\mathbf{p} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix}$$

Quadratische Matrix (3x3) mit allen Zahlen

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & -1 \\ -1 & 0 & 3 \\ 3 & 3 & 1 \end{bmatrix} \rightarrow \text{Gleichungssystem: } \mathbf{X}\mathbf{p} = \mathbf{y}$$

\rightarrow X ist Quadratische Matrix, wenn vollrangig kann Inverse gebildet werden

$$\mathbf{X}^{-1}\mathbf{X}\mathbf{p} = \mathbf{X}^{-1}\mathbf{y}$$

$$\mathbf{I}\mathbf{p} = \mathbf{X}^{-1}\mathbf{y}$$

$$\mathbf{p} = \mathbf{X}^{-1}\mathbf{y}$$

Rang einer Matrix

- Aussage über die Ausgeartetheit einer Matrix
- Je nach Anzahl Vektoren in der Matrix, die nicht linear abhängig sind
- Maximaler Rang: Anzahl Spalten/Reihen, die kleiner ist (z.B. bei 2x3-Matrix ist der maximale Rang = 2)

Vollrangigkeit

- Rang ist = maximaler Rang
- Zeilen des Gleichungssystems dürfen nicht linear abhängig sein
- Verschiedene Vektoren sind linear abhängig, wenn deren Summe = 0 ist

Quadratische Gleichung

$$y = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

$$\rightarrow y = \mathbf{x}^T \mathbf{A} \mathbf{x}$$

Ableitung von Skalaren nach Vektoren

$$y = \mathbf{a}^T \mathbf{x} = \sum_i a_i x_i.$$

→ Wenn nach einem Zeilenvektor abgeleitet wird, ist das Ergebnis ebenfalls ein Zeilenvektor. (umgekehrt bei Spaltenvektor)

$$\frac{\partial y}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \frac{\partial y}{\partial x_3} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \mathbf{a}$$

Ableitung der Quadratischen Form

$$y = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \frac{\partial y}{\partial \mathbf{x}} = \mathbf{A} \mathbf{x} + \mathbf{A}^T \mathbf{x}$$

Datenmatrizen

Daten können in Matrizen dargestellt werden.

Totalwerte der Daten:

- Spaltenvektor kreieren mit (1 1 1 1) (so viele Reihen wie benötigt für die Multiplikation mit der Datenmatrix)
- Multiplikation: (Datenmatrix) x (kreierter Spaltenvektor)

Bsp.

$$\begin{aligned} \mathbf{x}^T \mathbf{v} &= \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1.78 & 1.96 & 1.82 & 1.82 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 \times 1 + 1 \times 1 + 1 \times 1 + 0 \times 1 \\ 1.78 \times 1 + 1.96 \times 1 + 1.82 \times 1 + 1.82 \times 1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 7.38 \end{bmatrix} \end{aligned}$$

Mittelwerte der Daten:

Matrix Totalwerte mit Skalar (n^{-1}) Multiplizieren

Mittlere Abweichung / Standardabweichung

$D = x_i - \bar{x}$ (mittelwert)

$$\begin{aligned} \mathbf{X}^d &= \mathbf{X} - \iota \bar{x}^T \\ &= \begin{bmatrix} 0 & 1.78 \\ 1 & 1.96 \\ 1 & 1.82 \\ 0 & 1.82 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0.50 & 1.85 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1.78 \\ 1 & 1.96 \\ 1 & 1.82 \\ 0 & 1.82 \end{bmatrix} - \begin{bmatrix} 0.50 & 1.85 \\ 0.50 & 1.85 \\ 0.50 & 1.85 \\ 0.50 & 1.85 \end{bmatrix} = \begin{bmatrix} -0.50 & -0.07 \\ 0.50 & 0.11 \\ 0.50 & -0.03 \\ -0.50 & -0.03 \end{bmatrix} \end{aligned}$$

Varianz / Kovarianz

$$\begin{aligned} \mathbf{S} &= (n-1)^{-1} (\mathbf{X}^d)^T \mathbf{X}^d \\ &= (4-1)^{-1} \begin{bmatrix} -0.50 & 0.50 & 0.50 & -0.50 \\ -0.07 & 0.11 & -0.03 & -0.03 \end{bmatrix} \begin{bmatrix} -0.50 & -0.07 \\ 0.50 & 0.11 \\ 0.50 & -0.03 \\ -0.50 & -0.03 \end{bmatrix} \\ &= \begin{bmatrix} 0.33 & 0.03 \\ 0.03 & 0.10 \end{bmatrix} \end{aligned}$$

0.33 = Varianz im Geschlecht

0.10 = Varianz in Grösse

0.03 = Kovarianz

Korrelation

$$\mathbf{D}^{-.5} = \begin{bmatrix} \frac{1}{\sqrt{s_{11}}} & 0 \\ 0 & \frac{1}{\sqrt{s_{22}}} \end{bmatrix} = \begin{bmatrix} 1.73 & 0.00 \\ 0.00 & 12.67 \end{bmatrix}$$

$$\begin{aligned} \mathbf{R} &= \mathbf{D}^{-.5} \mathbf{S} \mathbf{D}^{-.5} \\ &= \begin{bmatrix} 1.73 & 0.00 \\ 0.00 & 12.67 \end{bmatrix} \begin{bmatrix} 0.33 & 0.03 \\ 0.03 & 0.10 \end{bmatrix} \begin{bmatrix} 1.73 & 0.00 \\ 0.00 & 12.67 \end{bmatrix} \\ &= \begin{bmatrix} 1.00 & 0.66 \\ 0.66 & 1.00 \end{bmatrix} \end{aligned}$$

→ Korrelation hier: 0.66

Vorlesung 4

Modelle

Modell = Vereinfachte Beschreibung einer Konzeption, die als Grundlage für ein theoretisches/empirisches Verständnis dienen kann.

→ Achtung: Essenz nicht verlieren

Merkmale Statistische Modelle:

- mathematisch
- stochastisch (vom Zufall abhängig)
- empirische Informationen werden dargestellt
- definiert für eine Population

„Zutaten“ statistische Modelle:

- Outcome Variable, **abhängige Variable**, die man erklären will
→ immer Zufallsvariable! Dies, weil das Modell eine Vereinfachung ist und immer bestimmte Faktoren vernachlässigt werden
- Predictor, erklärende/**unabhängige Variable**: kann unterschiedliche Formen annehmen
 - Kovariaten: Kontinuierliche/stetige Variablen
 - Faktoren: diskrete Variablen
- **Funktion**, die die erklärende mit der erklärten Variablen verknüpft
→ schliesst unbekannte Parameter mit ein

Lineares Regressionsmodell

→ oft gebraucht, wenn auch nicht mehr so aktuell

Unterschied zu normalen Modellen:

- abhängige Variable muss stetig sein (viele Ausprägungen, nicht zwingend unendlich)
- lineare Funktion bzgl. Parameter
($y = Bx$ ist ok, aber $y = x^B$ nicht)

Regressionslinie: y vorhergesagt anhand von x

→ kann auch wieder als Vorhersage benutzt werden

→ Erwartung aufgrund der erklärenden Variable

Residuum e:

Diskrepanz zwischen Vorhersage und eigentlichem Wert

→ **Populations-Regressionsmodell (PRM)**: generelle Darstellung für eine Population

Kennzeichen:

- Griechische Buchstaben

Ökonometrische Darstellung der PRM:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

b_0 = Intercept (Achsenabschnitt; Erwartung für y wenn $x = 0$)

b_1 = Steigung (Wenn x sich um eine Einheit ändert, um wieviel ändert sich dann der Erwartungswert von y)

e = Fehler:

- erklärende Variable, die ausserhalb des Modells sind
- Messfehler
- Spezifische Variation

y = Abhängige Variable

x = Unabhängige Variable

Annahmen:

- Bedingter Erwartungswert des Fehlers = 0
→ Fehler ist im Durchschnitt immer 0

$$E[\epsilon_i | x_i] = 0$$

→ Folglich: Populations-Regressionsfunktion (Erwartungswert)

$$E[y_i | x_i] = \beta_0 + \beta_1 x_i$$

→ Dies ist die **Funktion für die Regressionslinie!** Bedingte Erwartungswerte/Mittelwerte!

Alternative Formulierung für Regressionslinie: GLM
Annahme: Abhängige Verteilung ist normalverteilt

$$y_i \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 x_i$$

→ y ist Normalverteilt mit

- Mittelwert μ_i (μ_i ist Populations-Regressionsfunktion)
- Varianz, die man auf Regressionslinie ablesen kann (bestimmt, wie viele Fehler man eigentlich hat)

Stichproben-Regressionsfunktion

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad \text{Y(dach) = Vorhersage, Regressionslinie}$$

Stichproben-Regressionsmodell

Fehler miteinbezogen! Fehler = Residuum e

$$y_i = \hat{y}_i + e_i$$

Fit = Wie gut passt das Modell überhaupt?

Zwei Messmodelle:

RMSE:

- Schätzer der Standardabweichung
- Je grösser, desto weniger passt das Modell
- Auf gleicher Skala gemessen wie abhängige Variable

$$RMSE = \sqrt{\frac{\sum_i e_i^2}{n-2}} \quad \begin{array}{l} 0 = \text{alle Vorhersagen sind perfekt, liegen auf der} \\ \text{Regressionslinie} \end{array}$$

R²: Bestimmungskoeffizient:

- immer zwischen 0 und 1 (0 = schlecht, 1 = perfekt)
 - Frage: Was ist der Unterschied zwischen den Werten die man erhält mit Einbezug der predictors (Vorhersage, unabhängige Variable) und ohne?
 - Reduktion der Fehler durch Einbezug der Vorhersagen aus dem Regressionsmodell
 - E1 = Abweichung der Werte vom Mittelwert
 - E2 = Fehlervariation, wenn man Vorhersagen trifft mittels Modell
- wenn R = 0 → Ohne das Modell würde man gleich gute Voraussagen machen

$$\begin{aligned} R^2 &= \frac{E1 - E2}{E1} = \\ &= \frac{\sum_i (y_i - \bar{y})^2 - e_i^2}{\sum_i (y_i - \bar{y})^2} \\ &= 1 - \frac{\sum_i e_i^2}{\sum_i (y_i - \bar{y})^2} \\ &= 1 - \frac{\text{Residual Variation}}{\text{Total Variation}} \end{aligned}$$

17.03. VL4: **Multiple Regressionsanalyse**

mehrere erklärende Variablen

Populations-Regressionsmodell: Statt nur eine erklärende Variable hat man p erklärende Variablen

E = Fehlerkomponente; ist weniger gross als bei einfachem Regressionsmodell, da man jetzt mehrere erklärende Variablen hat.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

Populations-Regressionsfunktion: Bedingter Erwartungswert (Mittelwert) für y gegeben aller x

$$E[y_i | x_{i1}, x_{i2}, \dots, x_{ip}] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

Stichproben-Regressionsfunktion: aufgrund derer wird dann die Population geschätzt

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}$$

Stichproben-Regressionsmodell

$$y_i = \hat{y}_i + e_i$$

e = Residuum

Darstellung in Matrizen:

4 „Töpfe“:

abhängige Variable: n x 1 – Vektor

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Fehler: n x 1 – Vektor

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Beta: (p+1) x 1 – Vektor
(Koeffizienten)

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

Unabhängige Variable: n x (p+1) – Matrix:

Reihen/Zeilen: Verschiedene gemessene Einheiten

Spalten: verschiedene unabhängige Variablen

Erste Spalte: immer 1 -> Konstanten

$$\mathbf{X} = \begin{bmatrix} 1 & x^{u1} & x^{u5} & \dots & x^{ub} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x^{51} & x^{55} & \dots & x^{5b} \\ 1 & x^{T1} & x^{T5} & \dots & x^{Tb} \end{bmatrix}$$

Zusammenfassung dieser:

Populations-Regressionsmodell: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

Populations-Regressionsfunktion: $E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}$

Stichproben-Regressionsmodell: $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

Stichproben-Regressionsfunktion: $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$

Annahmen über erklärende Variablen:

1a: sind fix, keine Zufallsvariable! Nur abhängige Variable ist Zufallsvariable, weil sie von e bestimmt wird

→ Experiment: Man hat zwei Gruppen; Entweder x trifft zu oder nicht

1b: Vollrangigkeit ist bei X gegeben: kein Zusammenhang zwischen den erklärenden Variablen

Annahmen über Fehler

2a: Mittelwert/Erwartungswert der Fehler ist 0, haben keine systematische Tendenz

$$E[\boldsymbol{\epsilon}|\mathbf{X}] = \mathbf{0} \text{ so that } E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}.$$

2b: Fehler und unabhängige Variable korrelieren nicht

→ keine unberücksichtigten Variablen, die mit X korrelieren $E[\mathbf{X}^T \boldsymbol{\epsilon}] = \mathbf{0}.$

3a: Fehler haben alle die gleiche Varianz (homoskedastisch)

3b: Fehler korrelieren nicht untereinander

3c: Fehler sind normalverteilt

Folgerung aufgrund der Annahmen 2a und 3a-3c:

$$\boldsymbol{\epsilon}|\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

$$\sigma^2 \mathbf{I} = \mathbf{V}(\boldsymbol{\epsilon}|\mathbf{X})$$

$$= E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T|\mathbf{X}]$$

$$= \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Ceteris Paribus-Annahme

Bei mehreren erklärenden Variablen wird die Analyse mehrdimensional

→ deshalb Simplifikation:

Bei Aussagen über ein bestimmtes X geht man davon aus, dass alle anderen Faktoren konstant bleiben.

→ funktioniert nur, wenn die verschiedenen Variablen vollkommen unabhängig sind (was meist nicht der Fall ist)

marginaler Effekt:

$$\lim_{\Delta X_j \rightarrow 0} \frac{\Delta \mu}{\Delta X_j} = \frac{\partial \mu}{\partial X_j} = \beta_j$$

(Erwartete Änderung in Y bei kleinster Veränderung von Xj und Konstanthaltung aller anderer Parameter)

Standardisierung von Regressionskoeffizienten (damit diese vergleichbar werden):

$$\hat{\beta}_j^s = \hat{\beta}_j \frac{s_{x_j}}{s_y}$$

VL5, 24.3.

Schätzungsmethoden für partielle Regressionskoeffizienten

Kleinst-Quadrat-Schätzung (Ordinary least squares OLS)

Ziel: Regressionslinie erstellen, die am besten zu den erhobenen Daten passt

Kriterium: Residuen sollen auf irgendeine Weise minimiert werden

$$L = \sum_i (y_i - \mu_i)^2 = \sum_i \epsilon_i^2 \rightarrow \text{Total der quadrierten Fehler soll minimiert werden}$$

Warum quadriert und nicht absolute Werte?

- damit positive Fehler nicht von negativen aufgehoben werden
- grosse Fehler sollen bestraft werden

(siehe Übung 2)

→ Ableitung von L nach Beta und Gleichsetzung mit 0 ergibt das optimale Beta

1 Expand the fit criterion: $L = \sum_i (y_i - \mu_i)^2 = \sum_i (y_i - \beta_0 - \beta_1 x_i)^2 = \sum_i y_i^2 - 2\beta_0 \sum_i y_i - 2\beta_1 \sum_i x_i y_i + \sum_i \beta_0^2 + 2\beta_0 \beta_1 \sum_i x_i + \beta_1^2 \sum_i x_i^2$.

2 Partial derivatives:

$$\frac{\partial L}{\partial \beta_0} = 2n\beta_0 - 2 \sum_i y_i + 2\beta_1 \sum_i x_i$$

$$\frac{\partial L}{\partial \beta_1} = -2 \sum_i x_i y_i + 2\beta_0 \sum_i x_i + 2\beta_1 \sum_i x_i^2$$

→ beide Nullsetzen und auflösen:

Schätzer für Neigungen:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} \rightarrow \text{Kovarianz für x und y}$$

→ Varianz von x

Prinzip des kleinsten Quadrates kann auch mit Matrizen dargestellt werden

$$L = \epsilon^T \epsilon \quad \text{Epsilon = Fehler}$$

$$= (\mathbf{y} - \mathbf{X}\beta)^T (\mathbf{y} - \mathbf{X}\beta)$$

$$= \mathbf{y}^T \mathbf{y} - 2\beta^T \mathbf{X}^T \mathbf{y} + \beta^T \mathbf{X}^T \mathbf{X} \beta$$

→ Ableitung nach Beta und Nullsetzung

$$\frac{\partial L}{\partial \beta} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \beta \quad \begin{aligned} -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \beta &= \mathbf{0} \\ \mathbf{X}^T \mathbf{X} \beta &= \mathbf{X}^T \mathbf{y} \end{aligned}$$

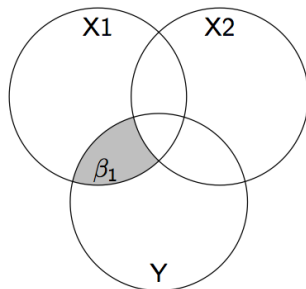
→ Normalvergleich (normal equations)

Erweiterung beider Seiten führt zur Lösung für beta (gemäss Annahme 1b):

$$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Schätzung von Y mittels X1 = beta1



1. Y auf X2 regressieren und die Residuen behalten (Anteil von Y, der nicht von X2 erklärt wird)
2. X1 auf X2 regressieren und Residuen behalten (Anteil von X1, der nicht X2 erklärt)
3. Residuen von 1 auf Residuen von 2 regressieren
→ B1 = Effekt des Teils von x1 auf y, das noch nicht durch x2 erklärt wurde.

(abhängige Variable wird immer auf unabhängige Variable regressiert)

Verzerrung des OLS-Schätzers (beta)

Wenn Erwartungswert von $\hat{\beta}$ nicht gleich ist wie β → Verzerrung (bias)

Ist OLS unverzerrt?

Ja, unter den Voraussetzungen von Annahmen 2a (Erwartungswert $\epsilon = 0$) und 2b (ϵ und erklärende Variable korrelieren nicht)

$$E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

(Ko)Varianz der Kleinst-Quadrat-Schätzer

$$\mathbf{V}[\hat{\boldsymbol{\beta}}] = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

Annahmen: Gleiche Varianz überall (3a)

Keine Korrelation der Fehler (3b)

→ Matrix der (Ko)varianz: $(k+1) \times (k+1)$

→ Diagonale Elemente: Quadrierte Standardfehler der Schätzer

→ restliche Elemente: Kovarianzen zwischen diesen Schätzern

Gauss-Markov-Theorem

Unter den Annahmen 2a, 2b, 3a und 3b ist der **OLS-Schätzer der beste lineare unverzerrte Schätzer**

(Der beste im Sinn von die kleinste Varianz, linear = nicht autokorreliert, unverzerrt = korrekt spezifiziert)

31.03.14

Teile der **Variation der abhängigen Variable:**

- Was man mittels Modell erklären kann (Stichproben-Regressionsfunktion)
- Was nicht erklärbar ist (Residuen)

ANOVA = Varianzanalyse: beide Teile der Varianz werden geschätzt

→ Dekomposition der Varianz

$$SST = SSR + SSE$$

SST = sum of squares total

SSR = sum of squares regression (Varianz, die durch das Modell erklärt wird)

SSE = sum of squares error (Residuen)

Bestimmungskoeffizient:

Unangepasst (unadjusted)

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

→ je höher, desto besser

→ immer zwischen 0 und 1

angepasst:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

→ kann auch negative Werte annehmen

angepasster Bestimmungskoeffizient ist immer kleiner als unangepasster!

→ Grund: Strafe für „overfitting“ (zu viele erklärende Variablen)

Höhe des Unterschieds hängt ab von Fallzahl und Anzahl Prädiktoren

Hypothesentest: 2 Arten von Tests

Einzelne Parameter (erklärende Variablen) testen

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$H_0 : \beta_1 = 0$$

Aussage: unabhängige Variable hat keinen Einfluss auf abhängige Variable

(→ Nullhypothese ist das Gegenteil der Hypothese, die man eigtl. beweisen will)

Erstmal Benutzung des Schätzers von Beta(1)

→ Kombiniert mit Nullhypothese:

$$\hat{\beta}_1 \sim \mathcal{N}(0, V[\hat{\beta}_1])$$

$$\frac{\hat{\beta}_1}{SE[\hat{\beta}_1]} \sim \mathcal{N}(0, 1)$$

Linke Seite: Test-Statistik; rechte Seite (Standardnormalverteilung): p-Wert

Problem 1: Standardfehler weiss man nicht, sondern nur den geschätzten Standardfehler

→ Folglich keine Normalverteilung, sondern eine Students-t-Verteilung mit $n-k-1$ Freiheitsgraden → „t-Test“

Problem 2: Fehler sind nicht normalverteilt

→ bei genügend grossem n ist die Verteilung des Schätzers normal

Generell: Hypothesentest mit einem Parameter:

$$H_0 : \beta_k = c, \quad c = \text{Konstante}$$

$$\frac{\hat{\beta}_k - c}{\widehat{SE}[\hat{\beta}_k]} \sim t_{n-k-1}$$

k = Anzahl erklärende Variablen

Mehrere Parameter testen

Man kann nicht selbes Verfahren (t-test) benutzen!

→ F-Test

Arbeiten mit Mean Squared Error

$$MSR = \frac{SSR}{k}$$

$$MSE = \frac{SSE}{n-k-1}$$

Teststatistik:

$$T = MSR / MSE$$

Wenn Nullhypothese korrekt ist (kein Einfluss der erklärenden Variable)

→ $MSR = 0$ und $T = 0$

Wenn Nullhypothese falsch ist (alle Varianz ist in abhängiger Variable erklärt)

→ $MSE = 0$, T strebt nach unendlich

→ Je niedriger die Teststatistik, desto besser für die Nullhypothese und desto schlechter für uns.

MSR und MSE: alternative Schätzer für die Standardabweichung

$$k \frac{MSR}{\sigma^2} = \frac{SSR}{\sigma^2} \sim \chi_k^2$$

$$(n-k-1) \frac{MSE}{\sigma^2} = \frac{SSE}{\sigma^2} \sim \chi_{n-k-1}^2$$

→ Verhältnis zweier unabhängigen Chi-Quadrat-Verteilung folgt einer F-Verteilung

$$\frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE} \sim F_{k, n-k-1}$$

Weiterer Test: Nur ein Teil der erklärenden Variablen hat einen Effekt von 0 auf die abhängige Variable

Nested Modelle: Wenn ein Modell Teil eines andern ist

$$M_1 \quad y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

$$M_2 \quad y = X_2\beta_2 + \epsilon$$

→ M2 ist in M1 enthalten

→ Waldtest

→ $F = t^2$

→ P-Wert ist derselbe wie im t-Test

14.4.

Diskrete Prädiktoren

- Lineares Regressionsmodell macht keine Annahmen über die Messebene der unabh. Var.
- Diskrete Prädiktoren kreieren Interpretationsprobleme
- Für Regressionskoeffizient: unklar, was die Änderung einer Einheit bedeutet

→ Lösung: Dummy-Variable

- Diskrete Variable wird nicht direkt eingebettet, sondern in Anzahl Dummy-Variablen (0,1) umgewandelt
- Bei Variablen mit M Ausprägungen kreiert man M-1 Dummy-Variablen

$$D_m = \begin{cases} 1 & \text{if } X = m \\ 0 & \text{otherwise} \end{cases}$$

- Die Ausprägung, für die keine Dummy-Variable kreiert wird: Baseline-Kategorie
→ trifft zu, wenn alle Dummies die Ausprägung 0 annehmen

$$y_i = \beta_0 + \sum_m \beta_m D_{im} + \mathbf{w}_i^T \boldsymbol{\gamma} + \epsilon_i$$

(R: lm-Befehl kreiert selber Dummy-Variablen, erste Kategorie ist Baseline)

Interpretation bei drei Ausprägungen:

Erwartungswert von y gegeben der beiden Dummies (PRF)

$$E[y | D_1, D_2] = \beta_0 + \beta_1 D_1 + \beta_2 D_2.$$

Wenn beide Dummies 0 sind: Erwartungswert für die Baseline = β_0

Baseline

- Ziel: perfekte Kollinearität / Multikollinearität vermeiden
- Welche der Kategorien die Baseline darstellt, ist egal (man erhält dieselben Zahlen)

Bsp. Interpretation:

$$\hat{y} = 67.9 - 4.8 \cdot \text{Midwest} - 10.5 \cdot \text{South} - 2.6 \cdot \text{West}$$

Baseline = Northeast

Northeast hat die grösste Stimmenanzahl für Obama, Süden die tiefste

Hypothesentest:

H0: $\beta(m) = 0$ → zwischen der Baseline und Kategorie m besteht kein signifikanter Unterschied

→ t-Test

$$T = \frac{\hat{\beta}_m}{SE[\hat{\beta}_m]} \sim t_{n-k-1}$$

lm-Output

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	67.869	2.361	28.748	< 2e-16	***
regionMidwest	-4.811	3.279	-1.467	0.14334	
regionSouth	-10.505	3.171	-3.313	0.00104	***
regionWest	-2.560	4.022	-0.636	0.52497	

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' ', 1

p-Wert für Unterschied zwischen Westen und Nordosten

→ keine Evidenz, dass der Unterschied zwischen Midwest und Northeast signifikant war (zu hoher p-Wert)

→ Nullhypothese kann nicht abgelehnt werden

Modell (Variablen) muss nicht zwingend linear sein. **Polynomiales** Regressionsmodell auch möglich:

$$y_i = \beta_0 + \sum_{k=1}^K \beta_k x_i^k + \mathbf{w}_i^T \boldsymbol{\gamma} + \epsilon_i$$

w = alle anderen unabhängigen Variablen

Interpretation **polynomiales** Regressionsmodell ist kompliziert

Benutzung: Wenn man das Gefühl hat, dass eine Variable nicht linear die abhängige Variable beeinflusst, sondern dass es einen Maximal/Minimalpunkt gibt

Hypothesentest:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i.$$

- es gibt eine lineare Beziehung zwischen X und Y: $\beta_2 = 0 \rightarrow$ t-Test
- Variable X hat gar keinen Effekt: $\beta_1 = \beta_2 = 0 \rightarrow$ F-Test

Moderate Variable

= Intervenierende Variable

- erhöht die Beziehung zwischen X und Y
- kann auch Y (AV) direkt beeinflussen

→ Interaktionen

- Fluktuierung des Effekts von X auf Y (Effekt ändert sich)
- Grund dafür: M

$$y_i = \beta_0 + \alpha_i x_i + \epsilon_i$$

Alpha bedeutet, dass der Koeffizient für jede Einheit von x unterschiedlich sein kann.

$$\alpha_i = \beta_1 + \beta_2 m_i \rightarrow \text{Modell für Alpha, wobei M die erklärende Variable ist}$$

Zusammenfügung der beiden Funktionen:

$$\begin{aligned}y_i &= \beta_0 + (\beta_1 + \beta_2 m_i)x_i + \epsilon_i \\ &= \beta_0 + \beta_1 x_i + \beta_2 x_i \cdot m_i + \epsilon_i\end{aligned}$$

$x_i \cdot m_i$ = **Interaktionsterm, Interaktionseffekt**; Produkt von X und M

Annahme: M hat auch einen eigenen Effekt auf Y:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 m_i + \beta_3 x_i \cdot m_i + \epsilon_i$$

Modul 3: Spezifikation und Spezifikationsfehler

- Modellspezifikation = Auswahl der Prädiktoren und Form für Regressionsmodell

Wichtig: Regressionsmodell an sich hat noch keine kausale Interpretation, sondern gibt nur Aussagen über die Korrelation

→ Damit Aussage über Kausalität gemacht werden kann, muss man das Modell spezifizieren

Arten von **Spezifikationsfehlern**:

Modell-Überspezifikation:

= unabhängige Variablen werden ins Modell miteinbezogen, die gar keinen Einfluss auf die abhängige Variable haben.

Probleme:

1. Statistische Effizienz geht verloren

- Freiheitsgrade gehen verloren
- Kollinearität kann vorkommen
- Statistische Aussagekraft wird tiefer

2. Prädiktoren werden schlechter

- v.a. bei Zeitreihenanalysen

3. Bias

- (Kontroll)variablen werden ins Modell eingefügt, die zwischen X und Y liegen
- man sollte also nichts kontrollieren, das eine Folge der unabhängigen Variable sein könnte

Vermeidung Überspezifikation:

- Überlegen: Welches ist Kernvariable?
- Variablen sollen rein genommen werden, die sowohl unabhängige, als auch abhängige Variable erklären können (ansonsten gibt es einen selection bias)
- Auch Variablen integrieren, die wesentlich abhängige Variablen erklären können, aber nichts mit unabhängiger Variable zu tun haben.

Modell-Unterspezifikation:

= Variablen, die einen Einfluss haben, fehlen im Modell (viel schlimmerer Fehler)

Probleme:

1. Verlust der statistischen Effizienz

2. Bias

- Verzerrung wird eingeführt; X wird ausgeklammert, X beeinflusst aber sowohl die unabhängige (D) als auch die abhängige (Y) Variable → Effekt der integrierten unabhängigen Variable wird überschätzt

→ Konfundierte Variable X, führt zu verzerrtem Schätzer

→ Annahme 2b wird verletzt

- Wenn Konfundierer ignoriert wird, wird die Annahme $Cov(D, \epsilon) = 0$ verletzt
 $\rightarrow E[\hat{\beta}] \neq \beta$.

Correlation between X and D	Effect of X on Y		
	< 0	0	> 0
< 0	+	0	-
0	0	0	0
> 0	-	0	+

X = konfundierende Variable, D = integrierte UV
 (+ = Überschätzung des Effekts, - = Unterschätzung des Effekts von D auf Y)

Wie testet man, ob es eine Unterspezifikation gibt?

- Ramsey's RESET Test: kein guter Test
- Es ist fast unmöglich, unentdeckte Variablen zu finden

Lösungen gegen Unterspezifikation:

- Effekt zwischen X und D zerstören
- Erreichung dieses z.B. durch Zufallsexperiment

Inklusion von konfundierenden Variablen:

- Hineinnehmen der konfundierenden Variablen als Kovariaten
- Annahme: Wenn alle konfundierenden Variablen integriert sind, ist die Korrelation zwischen unabhängiger Variable und Fehlerterm $Cov(x_i, e_i) = 0$.

Randomisierte Experimente

- „Treatment“ D ist zufällig zugeteilt (z.B. Erhaltung Medikament oder nicht); Einteilung in Experimentgruppe und Kontrollgruppe
- Annahme: Jede Einheit hat dieselbe Wahrscheinlichkeit, D zu erhalten oder nicht (unabhängig von X)
 \rightarrow Deshalb: Alle potenziellen erklärenden Variablen sind statistisch unabhängig von X

\rightarrow Goldstandard, um Konfusion zu verhindern

Probleme, die immer noch bestehen:

- Wenn Subjekte zwischen zwei Erhebungen verschwinden
- Wenn Leute der Versuchsgruppe z.B. das Medikament nicht nehmen und Leute der Kontrollgruppe es nehmen.
- Problem der Bias wird gelöst, aber anderes Problem (Problem der Präzision) bleibt bestehen

Randomisierung aber oft nicht möglich!

\rightarrow Instrumentvariablen-Analyse

- korrelieren mit Kern-UV, beeinflussen aber nicht direkt die AV und korrelieren nicht mit dem Fehlerterm
- Wenn eine UV teilweise mit dem Fehlerterm korreliert, kann durch die IV der Teil der UV herausgefiltert werden, der nicht mit dem Fehlerterm korreliert
- Hilfe auch wenn Experiment durchgeführt wird, aber nicht alle Einheiten die Anweisung befolgen (z.B. Medikament nehmen)

Endogenität: Wenn eine oder mehrerer UV mit dem Fehlerterm korrelieren (also z.B. mit anderen UVs korrelieren, die aber nicht im Modell integriert sind)

Ursachen Endogenität

- Unterspezifikation
- Simultanität (rückwirkender Einfluss von Y auf X)

→ Instrumentvariablen können in allen Fällen Endogenität verhindern

Annahmen für Instrumentanalyse (Bsp. Mit Brustkrebs): Randomisierung der Zuteilung als Instrumentvariable verwenden

- Randomisierung passiert völlig stochastisch
- Es gibt keine Individuen, die sich verhalten wie die Experimentgruppe, obwohl sie in Kontrollgruppe sind
- Instrumentanalyse kann helfen, Endogenität loszuwerden; Effekt wird aber nur auf die Individuen gemessen, die das Treatment befolgen

Annahmen, die vorhanden sein müssen für eine solche Analyse:

- Z (Instrumentvariable) hat Effekt auf X
- Ohne Randomisierung darf das Instrument nicht mit unbeachteten UV korrelieren
- Z darf keinen direkten Einfluss auf Y haben
- Zuteilung von Z darf keinen Einfluss auf das Verhalten der Individuen haben
- Monotonicity: Zuteilung von Z hat für alle Individuen denselben Effekt (nicht für die einen positiv, und für die anderen negativ)

→ Vorgehen für Instrumentvariablen-Analyse: Two-Stage Least Squares Approach

Wie findet man Instrumentvariablen ohne Randomisierung?

- Ausnutzen realer Randomisierung (die es schon gibt)
- Versuchen, Variablen zu finden, die verursachen, dass manche Individuen mit grösserer Wahrscheinlichkeit der Z ausgesetzt sind als andere (z.B. exogene historische/institutionelle Ursachen, Wetterschocks etc.)

Bsp.

Stiftung versucht, alle tiefen Einkommen in NYC zu identifizieren und per Lotterie Voucher auszuteilen, mit denen die Eltern ihre Kinder in Privatschule schicken konnten. Frage: Hat das Vorliegen eines Vouchers (X) einen positiven Effekt auf die schulischen Leistungen (post achievement) (Y)?

→ Stärkster Prädiktor dieser abhängigen Variable ist vermutlich die schulische Leistung vor diesem Experiment! Mit dem Voucher alleine kann die schulische Leistung nicht erklärt werden.

Weitere Spezifikationsfehler:

- Inkorrekte Spezifikation der funktionalen Form (Regressionslinie Gerade statt Kurve)
- Simultaneity: Abhängige Variable hat Rückwirkung auf unabhängige Variable

Heteroskedastizität

Definition: Varianz des Fehlerterms ist nicht konstant

- Fehler variiert zwischen den Einheiten
→ gewisse Einheiten (Werte von Y) können besser vorhergesehen werden als andere

klassischer Fall:
$$\sigma_i^2 = f(x_i)$$

- Konsequenzen weniger schlimm als Spezifikationsfehler

Heteroskedastizität in Matrix-Form:

$$\begin{aligned}
 E[\epsilon\epsilon^T] &= \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \\
 &= \mathbf{\Omega} \\
 &\neq \sigma^2 \mathbf{I}
 \end{aligned}$$

Auftreten Heteroskedastizität

- bei cross-sectional data (Erhebung der Einheiten zum selben Zeitpunkt)
- weniger bei Zeitreihen-Erhebungen
- bei Lerneffekten („neue“ machen noch viele Fehler, Routinierte weniger)

Konsequenzen Heteroskedastizität

- führt nicht zu Verzerrung des OLS-Schätzers! (ausser wenn Heteroskedastizität Folge eines Spezifikationsfehlers ist; Spezifikationsfehler führen zu Verzerrung)
- OLS-Standardfehler werden jedoch falsch geschätzt (zu niedrig oder zu hoch)
- Gauss-Markov-Theorem (BLUE) gilt nicht mehr: es gibt evt. bessere (kleinere Varianz), lineare, unverzerrte Schätzer
→ GLS (generalized least squares); hat weniger Varianz und ist effizienter

Design Effect

$$\text{deff} = \frac{\text{True Variance of } \hat{\beta} \text{ under Heteroskedasticity}}{\text{Variance of } \hat{\beta} \text{ According to OLS}}$$

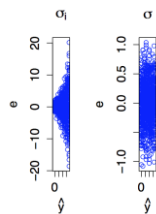
Wenn > 1: eigentliche Standardfehler sind grösser als was wir durch OLS herausfinden

Wenn < 1: eigentliche Standardfehler sind niedriger als die Voraussage durch OLS

Heteroskedastizität entdecken: Grafik

Links: Heteroskedastizität

Rechts: Homoskedastizität



Testen der Heteroskedastizität: Breusch-Pagan Test

Nullhypothese: Homoskedastizität ist vorhanden

→ Effekte, die mit X assoziiert sind, die Auswirkung auf e haben, sind alle 0

→ Wenn dies so ist, sollte Bestimmungskoeffizient (R^2) = 0 sein.

→ Wenn p-Wert niedrig ist: Nullhypothese wird abgelehnt, es besteht Heteroskedastizität