

1.1 EINFACHE REGRESSION - EINFÜHRUNG

Notation

- x Regressor, unabhängige variable
- y Zielvariable, abhängige variable

Visualisierung: Streudiagramm

- erstellen der Datenpaare (x_i, y_i)
- Erstelle Streudiagramm
- x: Abszisse y: Ordinate

Informell (-> 1.2)

- Stärke der Beziehung
- Richtung der Beziehung
- Ausreisser

1.2 ER - KOVARIANZ UND KORRELATION

Kovarianz

-> Beschreiben der linearen Beziehung von x und y

BV-Variante

$$COV(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

SP-Variante

$$COV = \frac{1}{n-1} \sum_{i=1}^n [(X_i - \bar{X})(Y_i - \bar{Y})]$$

POS. KOV. -> positive lineare Assoziation
NEG. KOV. -> negative lineare Assoziation
Problem: Wert der KOV. von Einheiten abhängig
-> kein absoluter Wert
Stärke der Bez. kann nicht beurteilt werden

Korrelation

-> beschreiben der linearen Beziehung

BV-Variante

$$COR(X,Y) = \frac{COV(X,Y)}{SA(X) SA(Y)}$$

SP-Variante

$$r = \frac{COV}{S_x S_y}$$

Nie blind Korrelation bilden

- bei nichtlinearer Beziehung
- Nicht robust gegen Ausreisser

-1 ≤ COR(X,Y) ≤ 1 -1 ≤ r ≤ 1

je näher Betrag an 1, desto stärker ist Beziehung

lineare Transformation ändern nichts

pos/neg kor. -> pos/neg lineare Assoziation

Regel für Varianzen

- var (x + y) = var (x) + var (y) + 2COV (x, y)
- var (x - y) = var (x) + var (y) - 2 COV (x, y)

wenn x und y unabhängig

-> keine lineare Beziehung und COV (x, y) = 0

Datenpaare

Bisher

$$SF = \frac{S_D}{\sqrt{n}} = \frac{\Sigma (X - Y)}{\sqrt{n}}$$

ASsoziation indirekt berücksichtigt

Jetzt

$$SF = \sqrt{\frac{S_x^2 + S_y^2 - 2COV}{n}} = \sqrt{\frac{S_x^2 + S_y^2 - 2rS_xS_y}{n}}$$

- COV = r * S_x S_y
- ASsoziation direkt berücksichtigt
- keine ASsoziation -> COV = 0

Kovarianz: Regeln

- COV (X, X) = var (X)
- COV (X, y₁ + y₂) = COV (X, y₁) + COV (X, y₂)
- COV (x₁ + x₂, y) = COV (x₁, y) + COV (x₂, y)
- COV (x, by) = b * COV (x, y)
- COV (X, a) = 0
- COV (X, a + by + c) = b COV (X, y) + c var (X)

E(a + bx) = a + bE(x) cov(a + bx, c + dy)

var(a + bx) = b² var(x) = b * a * cov(x, y)

SA(a + bx) = |b| SA(x) - Cor(a + bx, c + dy)

falls b > 0 d > 0 = Cor(x, y)

1.3 ER - SCHATZUNG

„Wahre Punkte“

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad i = 1, \dots, n$$

Annahmen

- Regressoren x₁, ..., x_n sind vorgegeben
- Fehler u₁, ..., u_n sind eine Zufall-Sp einer gemeinsamen Verteilung mit μ = 0 und unbek. σ

Notation

- β₀ = (Achsen-)Abschnitt = Intercept
- β₁ = Steigung
- misst den Effekt von x auf y
- misst wie stark y auf x reagiert
- σ = Fehler-Standardabweichung

Interpretation

- E(y_i) = β₀ + β₁ x_i
- wenn x = 0 -> im Durchschnitt y = β₀
- x steigt um 1 -> y steigt im Durchschnitt um β₁
- typischer Abstand von y von der Geraden: σ

-> i.d.R. β₀, β₁ und σ unbekannt!

Beobachtete Punkte

- Situation: Beobachtung von x_{neu} y_{neu} noch nicht bekannt
- Lösung: entsprechenden Punkt auf Regressions-Geraden nehmen

Idee

- kandidat einer Geraden durch paar (b₀, b₁) geg.
- Fehler, den der Kandidat am i-ten Punkt macht:

P_i = y_i - (b₀ + b₁ x_i) „Wahrheit - gewählte Gerade“

wähle unter allen Kandidaten denjenigen, den den kleinsten globalen Fehler macht

$$GF = \sum_{i=1}^n P_i^2$$

kleinste-Quadrate schätzer

- kleinste-Quadrate schätzer für β₀ und β₁

(β̂₀, β̂₁) = arg min_{β₀, β₁} Σ_{i=1}^n [y_i - (β₀ + β₁ x_i)]²

resultierende Formeln

$$\hat{\beta}_1 = r * \frac{S_y}{S_x} = \frac{COV}{S_x S_y} * \frac{S_y}{S_x} = \frac{COV}{S_x^2} = \frac{\Sigma [(x_i - \bar{x})(y_i - \bar{y})]}{\Sigma (x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

zugehöriger FQ-Schätzer für σ²

$$s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n-2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$$

geschätzter Fehler = Residuen

$$[1 - (\frac{\beta_1 * S_x}{S_y})^2] \quad (n-1) * s_y^2$$

SSR = [1 - r²] * SST

Erwartungswerte und Varianzen

- alle Schätzer β̂₀, β̂₁, s² sind erwartungstreu
- E(β̂₀) = β₀ E(β̂₁) = β₁ E(s²) = σ²
- Varianzen

$$\text{var}(\hat{\beta}_0) = \sigma^2 \frac{\Sigma x_i^2}{n \Sigma (x_i - \bar{x})^2}$$
$$\text{var}(\hat{\beta}_1) = \sigma^2 \frac{1}{\Sigma (x_i - \bar{x})^2}$$

Gauss-Markov Theorem

- β̂₀ und β̂₁ sind lineare Schätzer: sie sind gewichtete Durchschnitts der y-Daten von der Form Σ w_i y_i

$$\hat{\beta}_1 = \sum_{i=1}^n \left[\frac{(x_i - \bar{x})}{\Sigma (x_i - \bar{x})^2} \right] y_i$$

-> von allen lineare & erwartungstreuen Schätzern sind sie die besten (kleinste Varianz)

=> BLUE (Best Linear Unbiased Estimator)

Geschätztes Modell (Nutzen & Gefahr)

- β̂₀ und β̂₁ erlauben Beziehung von x und y zu quantifizieren und vorhersagen zu treffen
- beobachten von x_{neu} und vorhersage von y_{neu}

$$\hat{y}_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu}$$

- nie blind mit β̂₀ oder β̂₁ vorhersagen oder quantifizieren
- Nichtlineare Beziehung
- Ausreißer -> ggf. zuerst entfernen

=> Streudiagramm beachten

- Assoziation ≠ Verursachung
- > schlummernde Variablen können enthalten sein (andere Variablen, die Einfluss auf y haben, aber nicht in Regression sind)
- => Streudiagramm betrachtet nur ein x

Residuen-Diagramm

- ŷ - ũ Streudiagramm
- y-Achse: Residuen ũ_i = y_i - ŷ_i
- x-Achse: gefittete y-Werte ŷ_i = β̂₀ + β̂₁ x_i
- funktioniert für mehrere x-Variablen!

- lineares x, y Streudiagramm -> Chaos
- Nichtlin. x, y Str.-Dia. -> nichtlin. Residuen-Dia.

Einflussreiche Beobachtung für Ausreißer

-> kann „geschätzte Gerade“ an sich heranziehen

=> zugehöriges Residuum ist zu klein und sieht Ausreißer nicht mehr im Residuen-Diagramm

- mögl. x-Ausreißer: zugehöriger x-Wert ist ein Ausreißer innerhalb x-Beobachtungen

=> Ausreißer passt nicht in allgemeine Beziehung

Varianz-Zerlegung

- SST = sum of square total = Σ_{i=1}^n (y_i - ȳ)²

SST = SSR (Explained) + SSE (Residual)

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$
$$\sum_{i=1}^n \hat{u}_i^2$$

- SS vom Modell erklärt
- SS bleibt übrig (Fehler/ nicht erklärbar)

R² - Statistik, Bestimmtheits-Koeffizient

$$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} = 1 - \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y})^2} = \frac{\Sigma (\hat{y}_i - \bar{y})^2}{\Sigma (y_i - \bar{y})^2}$$

(n - k - 1) * s²

(n - 1) * s_y²

- SST = SSR + SSE -> 0 ≤ R² ≤ 1 je größer, desto besser
- Prozentsatz der Variation von y, welcher durch das Modell erklärt wird
- Für mehrere x-Variablen anwendbar
- eine x-Variable: R² = r² = Korrelation²

1.4 ER - INFERENCE

Konfidenzintervalle

- Standard-KI KI = Schätzer ± z_{1-α/2} * SF = Schätzer ± Fehler-Marge (FM)

- KI für β₀ KI = β̂₀ ± z_{1-α/2} * √(Σ x_i² / (n Σ (x_i - x̄)²))
- KI für β₁ KI = β̂₁ ± z_{1-α/2} * √(s² / Σ (x_i - x̄)²)

β₀ Hypothesentest

- H₀ = β₀ = β_{0,0} H₁ = { a) β₀ > β_{0,0}
b) β₀ < β_{0,0}
c) β₀ ≠ β_{0,0}

- Test-Statistik Z = (β̂₀ - β_{0,0}) / SF(β̂₀)
- p-Wert PW = a) P(Z ≥ z)
b) P(Z ≤ z)
c) 2 * P(Z ≥ |z|)

β₁ Hypothesentest

- H₀ = β₁ = β_{1,0} H₁ = { a) β₁ > β_{1,0}
b) β₁ < β_{1,0}
c) β₁ ≠ β_{1,0}

- Test-Statistik Z = (β̂₁ - β_{1,0}) / SF(β̂₁)
- p-Wert PW = a) P(Z ≥ z)
b) P(Z ≤ z)
c) 2 * P(Z ≥ |z|)

Achtung

- n ≥ 50
- Zufallsstichprobe -> keine Zeitreihen
- lineare Beziehung
- keine Ausreißer

Fehler u

- u_i müssen Zufallsstichproben darstellen
- u_i sind unabhängig voneinander cov = 0
- alle gleiche Verteilung -> gleiche SA
- i.d. Praxis aber oft verletzt!

- Erkennen
- Streudiagramm (eine Variable)
- Residendiagramm (beliebig viele Variablen) -> Fächer-Form

„Exakte“ Inferenz

Annahme

- Fehler u_i mit Normalverteilung -> u₁, ..., u_n ~ N(0, σ)
- benutzt t_{n-2}-Verteilung statt N(0,1) für KI & PW

-> KI = Schätzer ± t_{n-2, 1-α/2} * SF

-> PW = a) P(t_{n-2} ≥ z)

wenn Annahmen nicht zu treffen, nur bei großen n (auf N(0,1)) möglich.

Regression durch den Ursprung

Annahme: β₀ = 0 -> y_i = β₁ x_i + u_i i = 1, ..., n

-> unbekannter Parameter: β₁

- kleinste-Quadrate-Schätzer

$$\hat{\beta}_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}$$
$$s^2 = \frac{1}{n-1} \sum_{i=1}^n u_i^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \hat{\beta}_1 x_i]^2$$

- E(β̂₁) = β₁
- var(β̂₁) = σ² / Σ x_i² SF(β̂₁) = √(s² / Σ x_i²)
- KI = β̂₁ ± z_{1-α/2} * √(s² / Σ x_i²)
- Test-Statistik für Hypothesentest

$$Z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)} \leftarrow \text{„exakte“ Inferenz verwendet } t_{n-1} \text{ statt } N(0,1)$$

- Vorteile (β₀ = 0) -> einfaches Modell stimmt
- β̂₁ wird genauer geschätzt (kleinere Varianz)
- vorhersagen sind i.d.R. genauer
- Nachteile (β₀ ≠ 0) -> einfaches Modell stimmt nicht
- schätzen das Modell zu klein
- E(β̂₁) ≠ β₁
- vorhersagen ungenauer

WICHTIG (β₀ = 0)

- SST ≠ SSR + SSE
- R² ≠ SSR/SST
- R² < 0 möglich
- R² ≠ r²

Strategie

- theoretischer Grund für β₀ = 0
- Schätze das allgemeine Modell und teste H₀: β₀ = 0
- wenn Test nicht verwirft -> auf einfaches Modell umsteigen

1.5 ER - VORHERSAGE

Idealfall

- Annahme: β₀ und β₁ sind bekannt
- y_{neu} = β₀ + β₁ x_{neu} + u_{neu} -> ŷ_{neu} = β₀ + β₁ x_{neu}
- E(y_{neu}) = μ_{neu}
- vorhersage der Zufallsvariablen y_{neu} = μ_{neu} + u_{neu}
- wichtiger Unterschied
- E(ŷ_{neu}) kennen wir genau
- Realisierung von y_{neu} können wir nie genau vorhersagen, da es von u_{neu} abhängt

Realfall

- Annahme: β₀ und β₁ sind unbekannt
- y_{neu} = β₀ + β₁ x_{neu} + ũ_{neu} -> ŷ_{neu} = β̂₀ + β̂₁ x_{neu}
- E(ŷ_{neu}) = μ̂_{neu}
- vorhersage der Zufallsvariable y_{neu} = μ_{neu} + u_{neu}
- Realisierung von y_{neu} ungenauer, da noch Ungewissheit von μ_{neu} und u_{neu}

Voraussetzungen

- Für μ_{neu}
- Daten unabhängig voneinander
- Konstante Fehler-SA σ
- Stichprobe n ≥ 50
- VI für y_{neu}: Fehler u_i haben Normalverteilung

KI für μ_{neu}

- KI für den Mittelwert macht eine IGPR-Aussage
- Schätzer ist erwartungstreu

$$E(\hat{\mu}_{neu}) = E(\hat{\beta}_0 + \hat{\beta}_1 x_{neu}) = E(\hat{\beta}_0) + E(\hat{\beta}_1) x_{neu} = \beta_0 + \beta_1 x_{neu} = \mu_{neu}$$
$$\text{var}(\hat{\mu}_{neu}) = \sigma^2 \left(\frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\Sigma (x_i - \bar{x})^2} \right) \quad s^2 = \frac{s}{\sqrt{n}}$$

=> KI für μ_{neu} μ̂_{neu} ± z_{1-α/2} * √(1/n + (x_{neu} - x̄)² / Σ(x_i - x̄)²)

VI für y_{neu}

$$\text{var}(\hat{y}_{neu} - y_{neu}) = \text{var}(\hat{\mu}_{neu}) + \sigma^2 = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\Sigma (x_i - \bar{x})^2} \right)$$

- VI macht KZPR-Aussage für einen best. Fall
- Zusätzlich zur Variation μ̂_{neu} noch Variable u_{neu}
- VI für y_{neu} = √(residual-scale² + sse.pit²)

ŷ_{neu} ± z_{1-α/2} * s * √(1 + 1/n + (x_{neu} - x̄)² / Σ(x_i - x̄)²)

Beurteilung der Normalität (Normalverteilung)

- Normal-Quantil-Diagramm (N-Q-D)

heavy tail Light tail

Rechtsschief Linksschief

x²-Verteilung (-1) * x²-Verteilung

Normalverteilung

N-Q-D für (standardisierte) Residuen

- GO
- Gerade -> VI funktioniert „richtig“
- Light Tails -> VI ist „konservativ“
- NO-GO
- Schiefe -> VI oft „antikonservativ“
- Heavy Tails -> VI ist „antikonservativ“

Der Begriff „Regression“

- geschätztes Modell ŷ_i = β̂₀ + β̂₁ x_i
- äquivalent zu (y_i - ȳ) = β̂₁ (x_i - x̄)

=> Vorhersage von y_{neu} basierend auf x_{neu}

$$(\hat{y}_{neu} - \bar{y}) = \hat{\beta}_1 (x_{neu} - \bar{x}) = r \frac{S_y}{S_x} (x_{neu} - \bar{x})$$

=> (y_{neu} - ȳ) / s_y = r * (x_{neu} - x̄) / s_x

Interpretation

- wenn x_{neu} eine SA-Einheit von x̄ entfernt liegt, dann liegt ŷ_{neu} nur |r| SA-Einheiten von ȳ entfernt

-> (ŷ_{neu} - ȳ) < (x_{neu} - x̄), weil |r| < 1

=> Karl Pearson: „Regression to the mean“ / „Rückschritt zum Mittelwert“

1.6 ER - CAPM

Capital Asset Pricing Model

- x = Überschussrendite des Markt-Portfolios
- y = Überschussrendite eines Aktivums

-> Überschussrendite = Rendite - risikofr. Rendite

Theorie und Praxis

- erwartete Ü-Rendite des Aktivums ist proportional zur erwarteten Ü-Rendite des Marktes

$$E(y_i) = \beta E(x_i) \quad \text{oder} \quad y_i = \beta x_i + u_i$$

=> y_i = α + β x_i + u_i α ≠ β statt β₀ ≠ β₁

Steigung β

-> misst Risiko des Aktivums in Relation zum Markt

- „defensiv“: Aktivum mit β < 1 in Relation
- „aggressiv“: Aktivum mit β > 1 zum Markt

Achsenabschnitt α

-> durchschnittl. Ü-Rendite (Aktivum), wenn Ü-Rendite (Markt) = 0

- CAPM besagt, dass α = 0 (E(R_y) = r_p)
- wenn CAPM irrt: α ≠ 0

Schätzung

- Ü-Renditen über mehrere Perioden beobachten
- Regressions-Gerade (α und β) schätzen
- Frequenz: monatl. Daten Zeitraum: 5-10 J.

Inferieren

- n groß genug
- Daten stellen Zeitreihe dar, aber hier erlaubt, weil Aktien-Renditen und Zeit fast unabh.
- Residuen-Diagramm zeigt keine Fächer-Form

VI

- Aktienrenditen haben „Heavy Tails“ und nicht Normalverteilung -> VI nicht gültig

2.1 MULTIPLE REGRESSION - EINFÜHRUNG

- mehrere Regressoren: x₁, x₂, ..., x_n
- Vorteil: y besser modellieren und vorhersagen
- Nachteil: komplexer, mehr Gefahren

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i \quad i = 1, \dots, n$$

Hyper-Ebene Abweichung von Hyper-Ebene

Annahmen

- Regressoren sind deterministisch
- Fehler μ_i sind Zufallsstichproben von einer gemeinsamen Verteilung mit μ = 0 und (unbekannten) σ > 0

Notation

- β_j = j-te Steigung (j = 1, ..., k)

Vektor- und Matrizen-Notation

- y = (y₁, ..., y_n)' => n * 1 Vektor
- X = (x_{ij})_{1 ≤ i ≤ n, 0 ≤ j ≤ k} => n * (k+1) Vektor (Wobei x_{i0} = 1 stets) erste Spalte wegen β₀
- β̂ = (β̂₀, β̂₁, ..., β̂_k)' => (k+1) * 1 Vektor
- ũ = (u₁, ..., u_n)' => n * 1 Vektor

=> y = Xβ + ũ

Beurteilung der Beziehung

Punkte-Wolke in „Mengen“-R^{k+1} ein Chaos?

- Ja -> keine Beziehung
- Nein -> Beziehung

Falls nein, Formen Punkte grob eine Hyper-Ebene?

- Ja -> linear
- Nein -> nichtlinear

Streudiagramm mit multiplen Regressoren

- Für jede x-Variable einzeln ein Streudiagramm machen und „kombinieren“
- > mühsam und evtl. voll daneben
- alle x-Variablen in 1 Regressions-Modell aufstellen!

Multiples regressions-Modell

- $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i(1)} + U_i$
- wenn X_{i1} um 1 steigt und $X_{i(1)}$ konstant
-> Y_i erhöht sich um β_1
=> β_1 ist Effekt von X_{i1} „kontrolliert“ für $X_{i(1)}$
 - wenn $X_{i(1)}$ um 1 steigt und X_{i1} konstant
-> Y_i erhöht sich um β_2
=> β_2 ist Effekt von $X_{i(1)}$ „kontrolliert“ für X_{i1}

Gefahren

- separate Schätzung versagt, weil
 - evtl. negative Korrelation zw. X-Variablen
 - sachbezogene Konflikte
- Falls man nicht „kontrolliert“
 - unterschätzung der einzelnen X-Variablen Effekte (pos./neg. Verzerrung)
- wichtige Variablen gehören ins Modell auch wenn der Effekt nicht interessant ist

Beziehungen und Ausreißer

- Beziehung oder Ausreißer nicht mit individuellen Streudiagramm bestimmen
- Individuelle Bez., aber Gesamt-Bez. linear möglich
- Individuelle Bez. kann Ausreißer haben, aber Gesamt-Bez. nicht mehr

=> Residuen-Diagramm

2.2 SCHÄTZUNG

- wähle Hyper-Ebene, die GF minimiert
- Kandidat für Hyper-Ebene gegeben durch $\vec{b} = (b_0, b_1, \dots, b_k)'$
- Globaler Fehler gegeben als

$GF(\vec{b}) = \sum_{i=1}^n [Y_i - (b_0 + b_1 X_{i1} + \dots + b_k X_{ik})]^2 = [(Y - X\vec{b})']^2$

- kleinste-Quadrate-Schätzer**
 - $\hat{\vec{b}} = (X'X)^{-1} X' \vec{y} = \arg \min_{\vec{b}} GF(\vec{b})$
 - $\hat{U}_i = \vec{y}_i - \hat{\vec{y}}_i = \vec{y}_i - X \hat{\vec{b}}$
 - σ^2

Einige Eigenschaften

- „normal equations“ implizieren $0 = X'X\hat{\vec{b}} - X'\vec{y} = -X'(\vec{y} - X\hat{\vec{b}}) = -X'\hat{U}$

-> $\hat{U}$ ist orthogonal zu jeder Spalte von X („the residuals are orthogonal to the regressors“)

=> $\sum_{i=1}^n X_{ji} \hat{U}_i = 0$ für alle $j = 0, \dots, k$

- konstante β_0 im Modell (Abschnitt)
 - > erste Spalte von X ist ein Vektor von Einsen
 - => $\sum_{i=1}^n U_i^2 = 0$ $\beta_0 \neq 0$

Erwartungswert und Varianz von $\hat{U}$

- $E(\hat{U}) = \vec{0}$
- $var(\hat{U}) = \sigma^2 I$ $I = \text{Identitätsmatrix}$

EW, Varianz und Kovarianz-Matrix

$\vec{W}$ ist ein Zufallsvektor mit $(W_1, W_2, \dots, W_n)' \in \mathbb{R}^n$

EW $E(\vec{W}) = (E(W_1), \dots, E(W_n))' \in \mathbb{R}^n$

Varianz $var(\vec{W}) = \Sigma = (\sigma_{ij}) \in \mathbb{R}^{n \times n}$
 $\Sigma = \begin{pmatrix} var(W_1) & cov(W_1, W_2) \\ cov(W_2, W_1) & var(W_2) \end{pmatrix}$
 $\sigma_{ij} = var(W_j) \rightarrow$ Hauptdiagonale
 $\sigma_{ij} = cov(W_i, W_j) = cov(W_j, W_i)$ für $i \neq j$

=> (Varianz-)Kovarianz-Matrix

Eigenschaften der Schätzer

- $E(\hat{\vec{b}}) = \vec{b}$
- $E(S^2) = \sigma^2$
- $var(\hat{\vec{b}}) = \sigma^2 (X'X)^{-1}$ (Kovarianzmatrix von $\hat{\vec{b}}$)
 - ganze Matrix berechnen & einzelne (Ko-)Varianzen ablesen

Linearkombinationen

$\vec{a} \in \mathbb{R}^m, \vec{b} \in \mathbb{R}^{m \times n}$

$E(\vec{a} + \vec{b}\vec{W}) = \vec{a} + \vec{b} E(\vec{W})$

$var(\vec{a} + \vec{b}\vec{W}) = \vec{b} var(\vec{W}) \vec{b}'$
-> $\vec{b}' var(\vec{W}) \vec{b}$ nur sinnvoll für $n = m = 1$

Standardisierte Residuen

- Implikation $\hat{U}_i$ hat die gleiche SA -> gilt nicht für $\hat{U}_i$

KOV-Matrix von $\hat{U}_i$
 $var(\hat{U}_i) = \sigma^2 (I - R)$
 $R = X(X'X)^{-1} X'$

$SA(\hat{U}_i) = \sigma \sqrt{1 - r_{ii}} \Rightarrow SF(\hat{U}_i) = \sigma \sqrt{1 - r_{ii}}$

-> Verfeinerung: i-te standardisiertes Residuum

stand. Residuum $i = \frac{\hat{U}_i}{\sigma \sqrt{1 - r_{ii}}}$

N-Q-D in R benutzt die stand. Residuen

COOK'S Distance

-> Maß zur Erkennung einflussreicher Beobachtungen

- Erkennen der Gefahren
- Nichtlinearität: Res.-Diagramm hat ein Muster
- Ausreißer: Res.-Diagramm/COOK'S Distance

Für i-te Beobachtung und β -Schätzer
 $D_i = \frac{(\hat{\vec{b}} - \hat{\vec{b}}_i)' X' X (\hat{\vec{b}} - \hat{\vec{b}}_i)}{(k+1) S^2}$

- Mit $R = X(X'X)^{-1} X'$

$D = \{\text{Stand. Resid}_i\}^2 \cdot \left(\frac{r_{ii}}{1 - r_{ii}} \right) \cdot \frac{1}{k+1}$
großer Wert, wenn Ausreißer im Modell
großer Wert, wenn X_i weit weg von $\bar{X}$
 $\hookrightarrow \sum (X_i - \bar{X})^2$ groß

Beurteilung der konstanten Fehler-SA

- Plotte $\sqrt{|\text{Stand. Resid}_i|}$ gegen $\hat{y}_i$ in R-Software
- > wegen der Wurzel des Absolutwertes ist die halbe Fächerform schon Kennzeichen für Gefahr

Stärke der Beziehung

Bestimmtheit-Koeffizient $\in [0,1]$

$R^2 = 1 - \frac{SS_R}{SST} = \frac{SS_E}{SST} = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y})^2} = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}$

Bestimmt Stärke der Beziehung und erklärt wie viel der Variation von Y durch das Modell erklärt wird.

- eine X-Variable: $R^2 = r^2$
- In R-Software-Output

2.3 INFERENCE

- KI für β_j und $H_0: \beta_j = \beta_{j,0}$
- $H_0: R\vec{\beta} = \vec{r}$
- $H_1: \beta_1 = \dots = \beta_k = 0$

Konfidenzintervall β_j

- KI = Schätzer $\pm Z_{1-\alpha/2} \cdot SF$
- SF: Wurzel des j-ten Diagonalelemente aus Kovarianzmatrix aus 2.2 (Std. Error in Output)
- „exakte“ Inferenz: $t_{n-k-1, 1-\alpha/2}$

Hypothesen-Test β_j

- $H_0: \beta_j = \beta_{j,0}$ Software liefert Teststatistik (t-value) und PW für zweiseitigen Test
- $Z = \frac{\beta_j - \beta_{j,0}}{SF(\hat{\beta}_j)}$ (Pr(>|t|) zu einem Signifikanzniveau
- PW: wie immer

Erkennen der Gefahren

- $n \geq 50$
- Zufallsstichprobe: Daten unabhängig, konstante Fehler-Standardabweichung, lineare Beziehung und keine Ausreißer (2.2)

Allgemeiner Hypothesen-Test

- H_0 : Subvektor -> $\beta_1 = \beta_2 = 0$
lineare Kombination -> $\beta_1 + \beta_2 = 1$
- > jede Kombi stellt eine lineare Restriktion dar
=> Anzahl der Restriktionen = q

Formal $H_0: R\vec{\beta} = \vec{r}$
 $H_1: R\vec{\beta} \neq \vec{r}$
kleineres Modell -> weniger Regressoren
größeres Modell -> mehr Regressoren
=> genauer
 $R: [q, k+1]$
 $\vec{r}: [q, 1]$

Vergleich von H_0 und H_1 -> Schätzung beider Modelle
Alternative Modell nehmen, wenn es bedeutend besser ist, da es in jedem Fall mehr erklärt (kann sich den Daten besser anpassen)

$F = \frac{(R_1^2 - R_0^2) / q}{(1 - R_1^2) / (n - k - 1)} = \frac{(SS_R_0 - SS_R_1) / q}{SS_R_1 / (n - k - 1)}$
nur verwenden, wenn Zielvariablen gleich + R in Modell
Differenz zw. # Regressoren der zwei Modelle

$SS_R = \text{Freiheitsgrade} \cdot (\text{Residual Standard Error})^2$

-> $PW = P(F_{q, n-k-1} \geq F)$
kritischer Wert (wird angegeben)

$PW \leq \alpha$ Oder $F \geq F_{1-\alpha, q, n-k-1}$
-> H_0 verwerfen

- Vorzeichen eines Parameters ex ante bekannt & Modell schätzt dasselbe, ist PW (einseitig) = 1/2 PW (zweiseitig)

2.4 VORHERSAGE

Konfidenz-Intervall für $\mu_{\text{neu}} = E(\mu_{\text{neu}})$
Schätzer $\hat{\mu}_{\text{neu}} = \vec{X}'_{\text{neu}} \hat{\vec{\beta}}$

Eigenschaften $var(\hat{\mu}_{\text{neu}}) = \sigma^2 \vec{X}'_{\text{neu}} (X'X)^{-1} \vec{X}_{\text{neu}}$

$SF(\hat{\mu}_{\text{neu}}) = \sigma \sqrt{\vec{X}'_{\text{neu}} (X'X)^{-1} \vec{X}_{\text{neu}}}$
 $var(\hat{\vec{\beta}}) = S^2 (X'X)^{-1}$

=> $KI = \vec{X}'_{\text{neu}} \hat{\vec{\beta}} \pm Z_{1-\alpha/2} \cdot SF(\hat{\mu}_{\text{neu}})$
exakt: $t_{n-k-1, 1-\alpha/2}$
! Zufallsstichprobe und $n \geq 50$!

Vorhersageintervall für Y_{neu}

Vorhersage $\hat{Y} = \vec{X}'_{\text{neu}} \hat{\vec{\beta}}$

Eigenschaften $E(\hat{Y}_{\text{neu}} - Y_{\text{neu}}) = \mu_{\text{neu}} - \mu_{\text{neu}} = 0$

$var(\hat{Y}_{\text{neu}} - Y_{\text{neu}}) = \sigma^2 \vec{X}'_{\text{neu}} (X'X)^{-1} \vec{X}_{\text{neu}} + \sigma^2$

$SF(\hat{Y}_{\text{neu}} - Y_{\text{neu}}) = \sigma \sqrt{1 + \vec{X}'_{\text{neu}} (X'X)^{-1} \vec{X}_{\text{neu}}}$
 $= \sqrt{\$residual\text{-scale}^2 + \$se.\text{Fit}^2}$

2.5 DUMMY-VARIABLEN

- relevant für die Schätzung unters. Geraden (gemeinsame Gerade/Steigung/Abschnitt, total verschieden)
- > man nimmt das kleinste adäquate Modell

2.6 NICHTLINEARE BEZIEHUNGEN

definiere eine Variable D_i mit den Werten 0 & 1

Gemeinsame Steigung

$G_i = \beta_0 + \delta D_i + \beta_1 A_i + U_i$

- $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- $D_i = 1$: Abschnitt = $\beta_0 + \delta$, Steigung = β_1

- δ ist der geschätzte Unterschied der beiden Abschnitte
- PW für δ ist klein (signifikante Schätzung)
 - > „gemeinsame Gerade“ ist inadäquat
- $lm(\text{salary} \sim \text{gender} + \text{years})$, „gender“ = Dummy-v.

Gemeinsamer Abschnitt

$G_i = \beta_0 + \beta_1 A_i + \gamma (D_i * A_i) + U_i$

- $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- $D_i = 1$: Abschnitt = β_0 , Steigung = $\beta_1 + \gamma$
- γ ist der geschätzte Unterschied der beiden Steigungen
- PW für γ ist klein (signifikante Schätzung)
 - > „gemeinsame Gerade“ ist adäquat
- $lm(\text{salary} \sim \text{years} + I(\text{gender} * \text{years}))$

Total verschieden

$G_i = \beta_0 + \delta D_i + \beta_1 A_i + \gamma (D_i * A_i) + U_i$

- $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- $D_i = 1$: Abschnitt = $\beta_0 + \delta$, Steigung = $\beta_1 + \gamma$

- ergänzte Dummy-Variable ist signifikant (Abschnitt oder Steigung), sollte zu dem Modell gewechselt werden
- $lm(\text{salary} \sim \text{gender} + \text{years} + I(\text{gender} * \text{years}))$

Inferenz (Hypothesen-Test)

- Parameter für $D_i = 1$ und $D_i = 0$ ist derselbe, Antwort ist im Output Steigung in „gemeins. Steigung“
- Sonst ist Antwort anders zu berechnen
 - geschätzte COV zw. Parametern $\hat{\beta}$ & $\hat{\delta}$ in dies ist aber nicht im Output „gemeinsame St.“
 - > umschreiben des Modells

$G_i = \beta_0, D_i (1 - D_i) + \beta_0, D_i + \beta_1 A_i + U_i$

- $lm(\text{salary} \sim 1 + I(1 - \text{gender}) + \text{gender} + \text{years})$
 - „1“ ist die globale Konstante, die entfernt werden muss
 - im Output sind nun Werte enthalten, mit denen inferiert werden kann
 - „gemeinsamer Abschnitt“

$G_i = \beta_0 + \underbrace{\beta_0, D_i (1 - D_i) * A_i}_{\text{Steigung } D=0} + \underbrace{\beta_1, D_i (D_i * A_i)}_{\text{Steigung } D=1} + U_i$

Mehrere Gruppen

- mehr als 2 Gruppen
 - eine Basisgruppe, alle anderen mit eigener Dummy-Variable modelliert

$G_i = \beta_0 + \delta_1 D_{1i} + \delta_2 D_{2i} + \beta_1 A_i + U_i$
 $G_i = \beta_0 + \beta_1 A_i + \gamma_1 (D_{1i} * A_i) + \gamma_2 (D_{2i} * A_i) + U_i$
 $lm(\text{weight} \sim d1 + d2 + \text{weeks} + I(d1 * \text{weeks}) + I(d2 * \text{weeks}))$

(Total verschieden)

- Vergleich zweier Modelle -> F-Statistik
kleiner PW -> H_0 verwerfen

- Die Schätzung setzt sich aus dem Wert der Basisgruppe und der Parameter selbst zsm (wichtig für Intervalle und/ oder Vorhersage)
- kann für mehrere Gruppen beliebig oft wdh. werden

Chow Test

- multiple Regressions-Modell in g Gruppen
- H_0 : gemeinsames Modell H_1 : total verschieden

- SS_R_0 vom gemeinsamen Modell
- SS_R_1 von jeder einzelnen Gruppe schätzen und Σ

- # Restriktionen: $q = (g - 1) * (k + 1)$
- Gesamt-Freiheitsgrade: $F.G. = n - g * (k + 1)$

$F = \frac{(SS_R_0 - SS_R_1) / [(g-1)*(k+1)]}{SS_R_1 / [n-g*(k+1)]}$

- Alternative für SS_R_1 : Dummy-v. für jede Parameter einsetzen und Modell schätzen

2.6 NICHTLINEARE BEZIEHUNGEN

Polynomiale Beziehung

- Parameter werden zur linearen Schätzung auch mit einer höheren Potenz geschätzt

$Y_i = \beta_0 + \beta_1 X_i + [\beta_2 X_i^2] + [\beta_3 X_i^3] + U_i$

- Ausreißer müssen entfernt werden
- marginaler Nutzen einer zusätzlichen Potenz kann an seiner Signifikanz abgelesen werden
- Modelle vom Grad > 1 bedürfen anderer Methoden zur Interpretation

$\widehat{Effekt(X)} = \frac{d \hat{E}(y)}{d x}$
Residuen-Diagramm
≠ chaos, Polynom
-> Modell inadäquat

Nichtlineare Transformation

$h(Y_i) = \beta_0 + \beta_1 g(X_i) + U_i$

typische Transformationen

$\log(X), \exp(X), \sqrt{x}, x^2, 1/x$

Interpretation $\widehat{Effekt(X)} = \frac{d \hat{E}(y)}{d x}$

- höhere R^2 -Statistik lässt auf das neue Modell schließen
- Inferenz wie gewohnt, aber „naheliegende“ Interpretation mit Rücktransformation

Typ	$\beta_1 \rightarrow$ im Mittel!	η	Bedingung
Linear	$\beta_1 = \frac{\Delta y}{\Delta x}$ $x \uparrow \rightarrow \beta_1$	$\frac{\beta_1 x}{y}$	$y > 0$
Log-lin	$100\beta_1 = \frac{\% \Delta y}{\Delta x}$ $x \uparrow \rightarrow 100$	$\frac{\beta_1 x}{y}$	$y > 0$
Lin-log	$\frac{\beta_1}{100} = \frac{\Delta y}{\% \Delta x}$ $x \uparrow$ um $\Delta\%$	$\frac{\beta_1}{y}$	$x > 0$
Log-log	$\beta_1 = \frac{\% \Delta y}{\% \Delta x}$ $x \uparrow$ um $\Delta\%$	$\frac{\beta_1}{y}$	$x, y > 0$

Exakte Methode

- Ableitung von Modell nach ges. Regressor
- Ggf. Rücktransformation, wenn $h(y)$
- und „1“ rechnen

Interaktionen

- Interaktion, falls Effekt von β_j abhängig von anderen Variablen.

$Y_i = \beta_0 + \beta_1 G_i + \beta_2 A_i + \beta_3 (A_i * G_i) + U_i$

-> G und A hängen vom Niveau des anderen ab

- Gefahrendiagramm besser?
- Signifikanz vorhanden?

Vergleich (nicht) genesteter Modelle

- genestet = das eine Modell ist im anderen enthalten
- F-Test geht nur für getestete Modelle
- R^2 bevorzugt systematisch mehr Regressoren
- $\bar{R}^2$ (adjustiertes R) berücksichtigt #Regressoren

$\bar{R}^2 = 1 - \frac{\sum \hat{U}_i^2 / (n - k - 1)}{\sum (y_i - \bar{y})^2 / (n - 1)} = 1 - (1 - R^2) \left(\frac{n - 1}{n - k - 1} \right)$

- größeres k -> größeres Handicap, $\bar{R}^2$ kann steigen/fallen in k
- > genestetes Modell: wo der F-Test versagt (kleine Stichproben, Zeitreihen, nicht-konstante SA (U_i)) ist $\bar{R}^2$ praktikable Lösung.

2.7 HETEROSKEDASTIE

Homoskedastie
-> konstante Fehler-SA $var(U_i) = \sigma^2 \forall i = 1, \dots, n$

- allgemeine Formel für die Varianz von β_1 :
 $var(\beta_1) = (X'X)^{-1} X' \Sigma (X'X)^{-1}$
 $= \sigma^2 I = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$

- SF sind zu liberal (H_0 wird mit größerer W-keit verworfen als α), wenn keine Homoskedastie

Heterokedastie

-> individuelle Fehler $var(U_i) = \sigma^2 \forall i = 1, \dots, n$

Ansatz für $\hat{\Sigma}$ (konsistentere Fehler)

- $HCO: \hat{\sigma}_i^2 = \hat{U}_i^2$ nur für $n \geq 250$

$HC1: \hat{\sigma}_i^2 = \frac{n}{n - k - 1} * \hat{U}_i^2$ noch zu liberal

$HC2: \hat{\sigma}_i^2 = \frac{\hat{U}_i^2}{1 - r_{ii}} = S^2 (\text{stand-Resid}_i)^2$ „individuell aufgebläht“ manchmal zu liberal

$HC3: \hat{\sigma}_i^2 = \frac{\hat{U}_i^2}{(1 - r_{ii})^2}$ funktioniert i.d.R. gut bis sehr gut

=> Vergleich der Modelle: Std. Error etwas gleich, wenn Homoskedastie vorliegt

Test für Heteroskedastie

- Breusch-Pagan-Test testet H_0 : Homoskedastie H_1 : Heteroskedastie
-> ob die Fehler von den erklärten Variablen abhängen (nicht verlässlich bei kleinen n)

- Modell in standard-Inferenz schätzen und Residuen extrahieren
- Hilfsregression schätzen

- $\hat{U}_i^2 = \theta_0 + \theta_1 X_{i,1} + \dots + \theta_k X_{i,k} + \text{Fehler}_i$
Regression $H_0: \theta_1 = \dots = \theta_k = 0$
- Führe den Test mit „globalen“ F-Test in Hilfsregression aus
- Mit R-Software
 - Hypothesen-Test von Restriktionen
 - Software liefert PW

Allgemeiner F-Test unter Heteroskedastie

$F = \frac{(SS_R_0 - SS_R_1) / q}{SS_R_1 / (n - k - 1)} = \frac{(\hat{R}^2 - \bar{R}^2) / [R[S^2(X'X)^{-1} R]^{-1} (\hat{R}^2 - \bar{R}^2)]}{q}$

-> Schätzung von μ_A $var(\hat{\beta}_A)$ von H_3

2.8 Multikollinearität

Perfekte Multikollinearität

$X_{i,j} = \beta_$