

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: Abzisse 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 - μ_X)(Y - μ_Y)] = E[XY] - μ_Xμ_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 $\text{cor}(X,Y) = \frac{\text{COV}(X,Y)}{\text{SA}(X)\text{SA}(Y)}$
- SP-Variante $r = \frac{\text{COV}}{s_x s_y}$

- 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 Transformationen 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 $SF = \frac{S_D}{\sqrt{n}} = \frac{\sum(X-Y)}{\sqrt{n}}$
- Assoziation indirekt berücksichtigt

Jetzt $SF = \sqrt{\frac{S_1^2 + S_2^2 - 2COV}{n}} = \sqrt{\frac{S_1^2 + S_2^2 - 2rS_xS_y}{n}}$

- 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 - X) = b COV(X, Y) - c var(X)

- E(a + bX) = a + bE(X) - cov(a + bX, c + dY)
- var(a + bX) = b^2 var(X) = b * a * cov(X, Y)
- SA(a + bX) = |b| SA(X) - cor(a + bX, c + dY) = cor(X, Y)

1.3 ER - SCHATZUNG

„Wahre Punkte“ Y₀ = β₀ + β₁X₁ + u_i i = 1, ..., 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

Ŷ_{neu} = β₀ + β₁ X_{neu} (β₀, β₁ unbekannt!)

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 $\sum_{i=1}^n [Y_i - (β_0 + β_1 X_i)]^2$

resultierende Formeln

β̂₁ = r * $\frac{S_y}{S_x}$ = $\frac{COV}{S_x S_x} * \frac{S_y}{S_x} = \frac{COV}{S_x^2} = \frac{\sum[(x_i - \bar{x})(y_i - \bar{y})]}{(x_i - \bar{x})^2}$
β̂₀ = $\bar{y} - \hat{\beta}_1 \bar{x}$
 $S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
 $S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{SST}{n-1}$

zugehöriger FQ-schätzer für σ²
 $s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n-2} \sum_{i=1}^n [Y_i - (β_0 + β_1 X_i)]^2$
↳ geschätzter Fehler = Residuen
 $[1 - (\frac{\beta_1 \cdot S_x}{S_y})^2] (n-1) \cdot S_y^2$
SSE = $[1 - r^2] \cdot SST$

Erwartungswerte und Varianzen

- alle schätzer β̂₀, β̂₁, s² sind erwartungstreu
E(β̂₀) = β₀ E(β̂₁) = β₁ E(s²) = σ²
- Varianzen
 $\text{var}(\hat{\beta}_0) = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}$
 $\text{var}(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum (x_i - \bar{x})^2}$

Gauss-Markov Theorem

- β̂₀ und β̂₁ sind lineare schätzer: sie sind gewichtete durchschnitts der Y-Daten von der Form $\sum w_i Y_i$
 $\hat{\beta}_1 = \sum_{i=1}^n \left[\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right] Y_i$

-> 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 - Ŷ streudiagramm
 - Y-Achse: Residuen $\hat{u}_i = y_i - \hat{y}_i$
 - X-Achse: gefittete Y-Werte $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- funktioniert für mehrere X-Variablen!

- lineares X,Y streudiagramm -> Chaos
- nichtlin. X,Y str.-diag. -> 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
- moGl. X-Ausreißer: zugehöriger X-Wert ist ein Ausreißer innerhalb X-Beobachtungen

=> Ausreißer passt nicht in allgemeine Beziehung

Varianz-Zerlegung

- SST = sum of square total = $\sum_{i=1}^n (y_i - \bar{y})^2$

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

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

R²-Statistik, Bestimmtheits-Koeffizient

$R^2 = 1 - \frac{SSE}{SST} = \frac{SSR}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{(n-k-1) \cdot s^2}{(n-1) \cdot s_y^2}$

- 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-Variablen: R² = r² = Korrelationskoeffizient

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} * $\sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$
- KI für β₁ KI = β̂₁ ± Z_{1-α/2} * $\sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$

β₀ Hypothesentest

- H₀ = β₀ = β_{0,0} H₁ = $\begin{cases} a) \beta_0 > \beta_{0,0} \\ b) \beta_0 < \beta_{0,0} \\ c) \beta_0 \neq \beta_{0,0} \end{cases}$
- Test-Statistik $Z = \frac{\hat{\beta}_0 - \beta_{0,0}}{SF(\hat{\beta}_0)}$
 - a) P(Z ≥ Z)
 - b) P(Z ≤ Z)
 - c) 2 * P(Z ≥ |Z|)
- p-Wert PW = $\begin{cases} a) P(Z \geq Z) \\ b) P(Z \leq Z) \\ c) 2 * P(Z \geq |Z|) \end{cases}$

β₁ Hypothesentest

- H₀ = β₁ = β_{1,0} H₁ = $\begin{cases} a) \beta_1 > \beta_{1,0} \\ b) \beta_1 < \beta_{1,0} \\ c) \beta_1 \neq \beta_{1,0} \end{cases}$
- Test-Statistik $Z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)}$
 - a) P(Z ≥ Z)
 - b) P(Z ≤ Z)
 - c) 2 * P(Z ≥ |Z|)
- p-Wert PW = $\begin{cases} a) P(Z \geq Z) \\ b) P(Z \leq Z) \\ c) 2 * P(Z \geq |Z|) \end{cases}$

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)
 - Residuen-Diagramm (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 = β₁x_i + u_i i = 1, ..., n

-> unbekannter Parameter: β₁

- kleinste-Quadrate-schätzer $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$

s² = $\frac{1}{n-1} \sum_{i=1}^n u_i^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \hat{\beta}_1 x_i]^2$

- E(β̂₁) = β₁
- var(β̂₁) = $\frac{\sigma^2}{\sum x_i^2}$ SF(β̂₁) = $\sqrt{\frac{s^2}{\sum x_i^2}}$
- KI = β̂₁ ± Z_{1-α/2} * $\sqrt{\frac{s^2}{\sum x_i^2}}$
- Test-Statistik für Hypothesentest $Z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)}$ -> „Exakte“ Inferenz verwendet t_{n-1} 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

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
 $y_{neu} = \beta_0 + \beta_1 x_{neu} + u_{neu}$
-> $\hat{y}_{neu} = \beta_0 + \beta_1 x_{neu}$
- E(y_{neu}) = μ_{neu}
- vorhersage der Zufallsvariablen y_{neu} = μ_{neu} + u_{neu}
- wichtiger Unterschied
 - E(μ_{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
 $y_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu} + \hat{u}_{neu}$
-> $\hat{y}_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu}$
- 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}
 - erwartete Ü-Rendite des Aktivums ist proportional zur erwarteten Ü-Rendite des Marktes
 - 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(μ̂_{neu}) = E(β̂₀ + β̂₁x_{neu}) = E(β̂₀) + E(β̂₁) x_{neu} = β₀ + β₁x_{neu} = μ_{neu}

var(μ̂_{neu}) = σ² $\left(\frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)$ SF = $\frac{s}{\sqrt{n}}$

=> KI für μ_{neu} $\hat{\mu}_{neu} \pm Z_{1-\alpha/2} * s \sqrt{\frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$

VI für y_{neu}

var(Ŷ_{neu} - y_{neu}) = var(μ̂_{neu}) + σ² = σ² $\left(1 + \frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)$

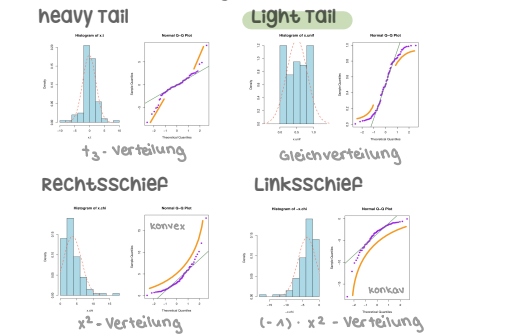
- VI macht KZPR-Aussage für einen best. Fall
- Zusätzlich zur Variation μ̂_{neu} noch Variable u_{neu}

- VI für y_{neu}

Ŷ_{neu} ± Z_{1-α/2} * s $\sqrt{1 + \frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ -> SST

Beurteilung der Normalität (Normalverteilung)

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



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 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- äquivalent zu $(y_i - \bar{y}) = \hat{\beta}_1 (x_i - \bar{x})$

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

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

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) = β E(X_i) oder y_i = β 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 = β₀ + β₁x_{i1} + ... + β_kx_{ik} + u_i i = 1, ..., n

- 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

Multiple regressions-Modell

- $$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i(i+1)} + u_i$$
- wenn X_{i1} um 1 steigt und $X_{i(i+1)}$ konstant
→ y_i erhöht sich um β_1
⇒ β_1 ist Effekt von X_{i1} „kontrolliert“ für $X_{i(i+1)}$
 - wenn $X_{i(i+1)}$ um 1 steigt und X_{i1} konstant
→ y_i erhöht sich um β_2
⇒ β_2 ist Effekt von $X_{i(i+1)}$ „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
- 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 $\hat{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$
- Globaler Fehler gegeben als

$$GF(\hat{\beta}) = \sum_{i=1}^n [Y_i - (\beta_0 + \beta_1 X_{i1} + \dots + \beta_k X_{ik})]^2 = \|(Y - X\hat{\beta})\|^2$$

Kleinste-Quadrate-Schätzer

- $\hat{\beta} = (X'X)^{-1}X'\vec{y} = \underset{\vec{b}}{\underset{\text{argmin}}{GF(\vec{b})}}$
- $\hat{u}_i = \vec{y} - \hat{\vec{y}} = \vec{y} - X\hat{\beta}$
- σ^2

$$s^2 = \frac{1}{n-k-1} \|\hat{u}\|^2 = \frac{1}{n-k-1} \sum_{i=1}^n \hat{u}_i^2$$

Einige Eigenschaften

- „normal equations“ implizieren

$$0 = X'X\hat{\beta} - X'\vec{y} = -X'(\vec{y} - X\hat{\beta}) = -X'\hat{u}$$

→ $\hat{u}$ 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 $\hat{u}$

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

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 $\text{var}(\vec{w}) = \Sigma = (\sigma_{ij}) \in \mathbb{R}^{n \times n}$

$$\Sigma = \begin{pmatrix} \text{var}(w_1) & \text{cov}(w_1, w_2) \\ \text{cov}(w_2, w_1) & \text{var}(w_2) \end{pmatrix}$$

$\sigma_{ii} = \text{var}(w_i) \rightarrow$ Hauptdiagonale
 $\sigma_{ij} = \text{cov}(w_i, w_j) = \text{cov}(w_j, w_i)$ für $i \neq j$

⇒ (Varianz-)Kovarianz-Matrix

Eigenschaften der Schätzer

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

Linearkombinationen

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

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

$\text{var}(\vec{a} + \vec{b}\vec{w}) = \vec{b} \text{var}(\vec{w}) \vec{b}'$
→ $\vec{b}' \text{var}(\vec{w})$ 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 $\hat{u}_i$
 $\text{var}(\hat{u}_i) = \sigma^2(I - R)$
 $R = X(X'X)^{-1}X'$

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

→ Verfeinerung: i-te standardisiertes Residuum

Stand. Residuum $_i = \frac{\hat{u}_i}{S \sqrt{1 - r_{ii}}}$

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

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})' X'X (\hat{\beta} - \hat{\beta}_{(i)})}{(k+1) S^2}$$

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

$D = (\text{Stand. Resid}_i)^2 \cdot \frac{r_{ii}}{1 - r_{ii}} \cdot \frac{1}{k+1}$

großer Wert, wenn Ausreißer im Modell $\rightarrow$ 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 $\in [0,1]$

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

- 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 Inferenz

- 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
- H_1 : wie immer (Prüfung) zu einem Signifikanzniveau
- $Z = \frac{\beta_j - \beta_{j,0}}{SF(\hat{\beta}_j)}$
- 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 $\rightarrow \beta_1 = \beta_2 = 0$
lineare Kombination $\rightarrow \beta_1 + \beta_2 = 1$
- jede Kombi stellt eine lineare Restriktion dar
⇒ Anzahl der Restriktionen = q

Formal $H_0: R\vec{\beta} = \vec{r}$ $R: [q, k+1]$
 $H_1: R\vec{\beta} \neq \vec{r}$ $\vec{r}: [q, 1]$

kleineres Modell → weniger Regressoren
größeres Modell → mehr Regressoren
⇒ genauer

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)

$$F = \frac{(R_1^2 - R_0^2) / q}{(1 - R_1^2) / (n - k - 1)} = \frac{(SSE_0 - SSE_1) / q}{SSE_1 / (n - k - 1)}$$

nur verwenden, wenn Zielvariablen gleich

$SSE = \text{Freiheitsgrade} \cdot (\text{Residual standard Error})^2$

→ $PW = P(F_{q, n-k-1} \geq F)$

- $PW \leq \alpha$ 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

Konfidenzintervall für $\mu_{\text{neu}} = E(\mu_{\text{neu}})$

- Schätzer $\mu_{\text{neu}} = \vec{x}_{\text{neu}}' \hat{\beta}$

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

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

⇒ $KI = \vec{x}_{\text{neu}}' \hat{\beta} \pm Z_{1-\alpha/2} \cdot SF(\hat{\mu}_{\text{neu}})$

exakt: $t_{n-k-1, 1-\alpha/2}$

! Zufallsstichprobe und $n \geq 50$!

Vorhersageintervall für y_{neu}

- Vorhersage $y = \vec{x}_{\text{neu}}' \hat{\beta}$
- Eigenschaften $E(\hat{y}_{\text{neu}} - y_{\text{neu}}) = \mu_{\text{neu}} - \mu_{\text{neu}} = 0$

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

$SF(\hat{y}_{\text{neu}} - y_{\text{neu}}) = S \sqrt{1 + \vec{x}_{\text{neu}}' (X'X)^{-1} \vec{x}_{\text{neu}}}$

$= \sqrt{\text{residual-scale}^2 + \text{sse.pit}^2}$

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

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

- definiere eine Variable D_i mit den Werten 0 & 1
- δ ist der geschätzte Unterschied der beiden Abschnitte
- PW für δ ist klein (signifikante Schätzung) → „gemeinsame Gerade“ ist inadäquat
- $\ln(\text{salary} \sim \text{gender} + \text{years})$, „gender“ = Dummy-v.

Gemeinsamer Abschnitt

$G_i = \beta_0 + \beta_1 A_i + \gamma (D_i \cdot A_i) + u_i$

- $D_i = 0$: 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
- $\ln(\text{salary} \sim \text{years} + I(\text{gender} * \text{years}))$

Total verschieden

$G_i = \beta_0 + \delta D_i + \beta_1 A_i + \gamma (D_i \cdot A_i) + u_i$

- $D_i = 0$: 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
- $\ln(\text{salary} \sim \text{gender} + \text{years} + I(\text{gender} * \text{years}))$

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 „gemeinsame St.“
 - dies ist aber nicht im Output
 - umschreiben des Modells

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

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

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

Mehrere Gruppen

- mehr als 2 Gruppen
 - eine Basisgruppe, alle anderen mit eigener Dummy-Variable modelliert
- $G_i = \beta_0 + \delta_1 D_{1i} + \delta_2 D_{2i} + \beta_1 A_i + u_i$ gemeins. Steigung
 $G_i = \beta_0 + \beta_1 A_i + \gamma_1 (D_{1i} \cdot A_i) + \gamma_2 (D_{2i} \cdot A_i) + u_i$ unters. Abschnitte, gemeins. Abs., unterschiedl. Steigung
- $\ln(\text{weight} \sim d1 + d2 * \text{weeks} + I(d1 * \text{weeks}) + I(d2 * \text{weeks}))$ (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

- SSE_0 vom gemeinsamen Modell
- SSE_1 von jeder einzelnen Gruppe schätzen und Σ

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

$$F = \frac{(SSE_0 - SSE_1) / [(g-1) \cdot (k+1)]}{SSE_1 / [n - g \cdot (k+1)]}$$

- Alternative für SSE_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

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

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

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

Typ	β_1	Approximation → im Mittel!	η	Bedingung
Linear	$\beta_1 = \frac{\Delta y}{\Delta x}$	$x \uparrow \rightarrow \beta_1$	$\frac{\beta_1 x}{y}$	
Log-lin	$100\beta_1 = \frac{\% \Delta y}{\Delta x}$	$x \uparrow \rightarrow 100 \cdot \beta_1 \%$	$\frac{\beta_1 x}{y}$	$y > 0$
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 $\Delta\%$	$\frac{\beta_1}{y}$	$x, y > 0$

→ $\beta_1 \%$ $\hookrightarrow \beta_1 \%$

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.
- $$Y_i = \beta_0 + \beta_1 G_i + \beta_2 A_i + \beta_3 (A_i \cdot G_i) + u_i$$
- 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
- R^2 (adjustiertes R^2) 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 R^2 praktikable Lösung.

2.7 Heteroskedastie

- Homoskedastie
- konstante Fehler-SA $\text{var}(u_i) = \sigma^2 \forall i = 1, \dots, n$
- allgemeine Formel für die Varianz von β_1 :
 $\text{var}(\beta_1) = (X'X)^{-1} X' \Sigma (X'X)^{-1}$
 $\hookrightarrow \sigma^2 I = \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix}$
- SF sind zu liberal (H_0 wird mit größerer W-keit verworfen als α), wenn keine Homoskedastie

Heteroskedastie

→ individuelle Fehler $\text{var}(u_i) = \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$

$= \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$

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

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

- $HCl: \hat{\sigma}_i^2 = \frac{n}{n - k - 1} \cdot \hat{u}_i^2$ noch zu liberal
- $Hc2: \sigma_i^2 = \frac{\hat{u}_i^2}{1 - r_{ii}} = s^2 (\text{stand-resid}_i)^2$ „individuell aufgebläht“ manchmal zu liberal
- $Hc3: \sigma_i^2 = \frac{\hat{u}_i^2}{(1 - r_{ii})^2}$ 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

$$\hat{u}_i^2 = \theta_0 + \theta_1 x_{i,1} + \dots + \theta_k x_{i,k} + \text{Fehler}_i$$

$H_0: \theta_1 = \dots = \theta_k = 0$

→ 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

$$F = \frac{(SSE_0 - SSE_1) / q}{SSE_1 / (n - k - 1)} = \frac{(\hat{R}^2 - \bar{R}^2) / q}{\bar{R}^2 / (n - k - 1)}$$

→ Schätzung von θ_1 $\text{var}(\hat{\beta}_1)$ von H_3

2.8 Multikollinearität

Perfekte Multikollinearität

- $x_{i,j} = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{j-1} x_{i,j-1} + \beta_{j+1} x_{i,j+1} + \dots + \beta_k x_{i,k}$
- (mind.) eine Spalte von X kann als Linearkombi. der restlichen k Spalten werden
 - X hat Spaltenrang $< k + 1$
 - $(X'X)$ hat Rang $< k + 1 \Rightarrow$ ist singular
 - $(X'X)$ existiert nicht
 - OLS-Schätzer $\hat{\beta}$ existiert nicht

Vorgehen