

Empirische Wirtschaftsforschung	Andreas Puccio	09.12.2005;	vervollständigt
	und bereinigt:	Philippe Ruh	und Martin
	Rapp	27.01.2007,	20.09.2008
		using	I ⁹ T ⁸ X

1. INFERENZ IN KONTINGENZTABELLEN

Notationen für Tabellen. h_{ij} = Anzahl in Zelle (i, j) $h_{i.}$ = Anzahl in Zeile i (total) $h_{.j}$ = Anzahl in Spalte j (total) n = Gesamtanzahl

Kategorische Variablen, formales Problem. - Zwei kategorische Variablen X und Y - Besteht Beziehung?
--

● Zugehöriger Hypothesentest H_0 : X und Y sind unabhängig H_A : X und Y sind abhängig ● Chi-Quadrat-Test-Statistik - Direkte Berechnung der Zellen-Anzahl: $e_{ij} = (h_{i.} \times h_{.j})/n$ - Destillieren in einen globalen Vergleich: $\chi^2 = \sum_{\text{Zellen}} \frac{(\text{Beobachtet} - \text{Erwartet})^2}{\text{Erwartet}}$

● p-Wert p-Wert unter H_0. $\chi^2 \sim \chi^2_{(r-1)(c-1)}$ r: Anzahl Zeilen c: Anzahl Spalten

χ^2-Verteilung. - stetige Verteilung χ^2_k k: Anzahl Freiheitsgrade: $(r - 1) \times (c - 1)$ $\mu = k$ $\sigma^2 = 2k$

Test zum Signifikanzniveau (krit. Wert) α. $p = P(\chi^2_{(r-1)(c-1)} \leq X^2)$ - Zwei Methoden - Verwerfe H_0 falls $p \leq \alpha$ - Verwerfe H_0 falls $X^2 \geq \chi^2_{1-\alpha, (r-1)(c-1)}$
--

Bedingungen: $n \geq 50$ - alle erwarteten Zellen-Einträge ≥ 5 - Stichprobe ist zufällig
--

Hinweis: Anz. Freiheitsgrade zeigen, wo in der Tabelle zu schauen.

2.1 EINFACHE REGRESSION - EINFÜHRUNG

Situation. - numerische Variablen X, Y - Frage nach Beziehung zw. X und Y - Wollen Y mit X vorhersagen Fragen: - Besteht eine Beziehung? - Wie sieht sie quantitativ aus? - Gegeben X, wie Y vorhersagen?
--

Notation. X: Regressor, unabh. Var. Y: Zielvariable, Regressand, abh. Var.
--

Visualisierung: Streuungsdiagramm. - Erstellen Datenpaare (x_i, y_i) - Erstelle Streuungsdiagramm X: Abszisse Y: Ordinate Beziehungen können grob ausgesehen werden: - keine $\rightarrow$ Chaos - Linear $\rightarrow$ Punkte formen grob eine Li-nie/Gerade - Nichtlinear $\rightarrow$ Punkte formen grob eine Kur-ve informell: - Stärke der Beziehung ($\rightarrow 2.2$) - Richtung der Beziehung ($\rightarrow 2.2$) - Ausreisser
--

2.2 EINFACHE REGRESSION - KOVARIANZ UND KORRELATION

Kovarians. - Beschreiben lineare Bez. zw. Variablen - Bevölkerungsvariante: $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ - Stichprobenvariante: $cov = \frac{1}{n-1} \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]$ - pos. Kov. $\rightarrow$ positive lin. Assoziation - neg. Kov. $\rightarrow$ negative lin. Assoziation - Problem: von Einheiten abhängig $\rightarrow$ nicht 'abh-solut' - Stärke der Beziehung kann nicht beurteilt wer-den.
--

Korrelation. - Beschreibt lineare Beziehung. - Bevölkerungsvariante: $Cor(X, Y) = \frac{Cov(X, Y)}{SA(X)SA(Y)}$ - Stichprobenvariante: $r = \frac{cov}{sx sy}$ $-1 \leq Cor(X, Y) \leq 1$ $-1 \leq r \leq 1$ - lin. Transformationen ändern nichts - lin. Transformationen ändern nichts - pos. Kor. $\rightarrow$ positive lin. Assoziation - neg. Kor. $\rightarrow$ negative lin. Assoziation - Je näher Betrag an 1, desto stärkere Bez. - empfindlich gegen Ausreisser - nicht blind Korrelationen bilden: - bei nichtlin. Beziehungen nicht - nicht robust gegen Ausreisser
--

Regel für Varianzen. $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$ $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$ Wenn X und Y unabhängig: - besteht insbesondere keine lineare Beziehung $Cov(X, Y) = 0$.

Datenpaare. - Bisher: $SF = \frac{s_D}{\sqrt{n}}$ - Assoziation indirekt berücksichtigt - Jetzt: $SF = \sqrt{\frac{s_1^2 + s_2^2 - 2cov}{n}} = \sqrt{\frac{s_1^2 + s_2^2 - 2rs_{XY}}{n}}$ - wobei $cov = r \times s_X s_Y$ - Assoziation direkt berücksichtigt
--

Kovarians: Regeln. $Cov(X, X) = Var(X)$ $Cov(X, Y_1 + Y_2) = Cov(X, Y_1) + Cov(X, Y_2)$ $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$ $Cov(X, bY) = bCov(X, Y)$ $Cov(X, a) = 0$ $Cov(X, a+bY - cX) = bCov(X, Y) - cVar(X)$
--

2.3 EINFACHE REGRESSION - SCHÄTZUNG
--

'Wahre Punkte'. $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ $i = 1, \dots, n$ Annahmen: - Die Regressoren $x_1, \dots, x_n$ sind vorgegeben - Fehler $\epsilon_1, \dots, \epsilon_n$ sind eine Zufall-S. von einer gemein. Verteilung mit $\mu = 0$ und unbekanntem σ Notation: β_0 = (Achsen-)Abschnitt β_1 = Steigung - misst den 'Effekt von X auf Y' - misst wie stark Y auf X 'reagiert' σ = Fehler-SA Folglich: $E(y_i) = \beta_0 + \beta_1 x_i$ - Wenn $x = 0$, dann im D'schnitt $y = \beta_0$ - Wenn x 1 rauf, dann y im D'schnitt um β_1 rauf - Typischer Abstand von y von der Geraden: σ
--

Aber: Wir kennen β_0, β_1 und σ nicht!
--

Beobachtete Punkte. - Situation: Wir beobachten x_{neu}, aber y_{neu} ist noch nicht bekannt. - Lösung: Wir nehmen den entsprechenden Punkt auf der Regr.-Geraden. $\hat{y}_{neu} = \beta_0 + \beta_1 x_{neu}$ - Aber: Wir kennen β_0, β_1 nicht!
--

Idee: - Kandidat einer Geraden durch Paar (b_0, b_1) gegeben. - Fehler, den dieser Kandidat am i-ten Punkt macht: $f_i = y_i - (b_0 + b_1 x_i)$ - Wähle unter allen Kand. denjenigen, der den kleinsten globalen Fehler macht - Achtung: e_i bei Bevölkerung, f_i bei Zufalls-Stichprobe! $GF = \sum_{i=1}^n f_i^2$
--

Kleinste-Quadrate Schätzer. - Kleinste-Quadrate Schätzer für β_0 und β_1: $(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2$ - Resultierende Formeln: $\hat{\beta}_1 = r \frac{s_y}{s_x}$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ - Zugehöriger KQ Schätzer für σ^2: $s^2 = \frac{1}{n-2} \sum_{i=1}^n \epsilon_i^2$ - Die geschätzten Fehler $\hat{\epsilon}_i$ heißen Residuen. $s^2 = \frac{1}{n-2} \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$

Erwartungswerte und Varianzen. - Alle drei Schätzer $\hat{\beta}_0, \hat{\beta}_1, s^2$ sind erwartungs-treu (unbiased): $E(\hat{\beta}_0) = \beta_0$ $E(\hat{\beta}_1) = \beta_1$ $E(s^2) = \sigma^2$ - Die Varianzen von $\hat{\beta}_0$ und $\hat{\beta}_1$ sind: $Var(\hat{\beta}_0) = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}$ $Var(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum (x_i - \bar{x})^2}$

Gauss-Markov Theorem. $\hat{\beta}_0$ und $\hat{\beta}_1$ sind lineare Schätzer: - Sie sind gewichtete Durchschnitte der Y-Daten von der Form $\sum w_i y_i$ $\hat{\beta}_1 = \sum_{i=1}^n \left[\frac{(x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \right] y_i$

Geschätztes Modell. $\hat{\beta}_0$ und $\hat{\beta}_1$ erlauben es uns, die Beziehung zwischen X und Y zu quantifizieren und vorher-sagen zu treffen. - Wie beobachten x_{neu} und sagen y_{neu} vorher: $\hat{y}_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu}$ - Nie blind mit Hilfe von $\hat{\beta}_0$ oder $\hat{\beta}_1$ vorhersagen oder quantifizieren: - wegen nichtlinearen Bez. - Ausreisser $\Rightarrow$ ggf. zuerst entfernen $\Rightarrow$ Steuungsdiagramm.
--

Kontext. - Assoziation $\neq$ Verursachung $\rightarrow$ Schlummernde Variablen Residuendiagramm. - Dies ist ein $\hat{y} - \hat{e}$ Streuungsdiagramm. - Residuen auf der 'Y'-Achse, gefitteten Y-Werte auf der 'X'-Achse. - lin. X,Y St.dia. $\rightarrow$ Chaos im Res.dia. - Nichtlin. X,Y St.dia. $\rightarrow$ nichtlin. Res.dia. - Auch für mehrere X-Variablen anwendbar!

Extreme Ausreisser. - Wenn Ausreisser extrem, können sie manchmal 'geschätzte Gerade' an sich heranziehen. $\Rightarrow$ Dann ist das zugehörige Residuum zu klein und sieht Ausreisser nicht mehr im Residuen-Diagramm - Möglicher X-Ausreisser: zugehöriger X-Wert ist ein Ausreisser innerhalb der X
--

Varianz-Dekomposition. - Der Summand $\sum_{i=1}^n (y_i - \bar{y})$ heisst SST = Sum of Squares Total - SST aufteilen (dekomponieren) in: $\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 (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \epsilon_i^2$ Erläuterung SS_{Tot}, SS_{Exp}, SS_{Res} SS_{Exp} vom Modell erklärt. SS_{Res} bleibt übrig.

R^2-Statistik, Bestimmtheits-Koeffizient. - Definiert als $R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{SS_{Exp}}{SS_{Tot}}$ - Da $SS_{Tot} = SS_{Exp} + SS_{Res}$ - Daher gilt: $0 \leq R^2 \leq 1$ - Prozentsatz der Variation von Y, welcher durch das Modell erklärt wird. - Auch für mehrere X-Variablen anwendbar! - Bei einer X-Variablen gilt: $R^2 = r^2$

2.4 EINFACHE REGRESSION - INFERENZ

Konfidenzintervalle. - Standard KI: $KI = \text{Schätzer} \pm z_{1-\frac{\alpha}{2}} \times St.fehler(SF)$ $KI = \text{Schätzer} \pm Fehler - Marge(FM)$ - KI für β_0 $KI = \hat{\beta}_0 \pm z_{1-\frac{\alpha}{2}} \times \sqrt{s^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$ - KI für β_1 $KI = \hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$
--

β_0 Hypothesentest. H_0: $\beta_0 = \beta_{0,0}$ H_A : $\begin{cases} (a) & \beta_0 > \beta_{0,0} \text{ (einseitig)} \\ (b) & \beta_0 < \beta_{0,0} \text{ (einseitig)} \\ (c) & \beta_0 \neq \beta_{0,0} \text{ (zweiseitig)} \end{cases}$ - Test-Statistik berechnen: $z = \frac{\hat{\beta}_0 - \beta_{0,0}}{SF(\hat{\beta}_0)}$ wobei $SF(\hat{\beta}_0) = \sqrt{s^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$ - p-Wert berechnen: $PW = \begin{cases} (a) & P(Z \geq z) \\ (b) & P(Z \leq z) \\ (c) & 2P(Z \geq z) \end{cases}$
--

β_1 Hypothesentest. H_0: $\beta_1 = \beta_{1,0}$ H_A : $\begin{cases} (a) & \beta_1 > \beta_{1,0} \text{ (einseitig)} \\ (b) & \beta_1 < \beta_{1,0} \text{ (einseitig)} \\ (c) & \beta_1 \neq \beta_{1,0} \text{ (zweiseitig)} \end{cases}$ - Test-Statistik berechnen: $z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)}$ wobei $SF(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}}$

- p-Wert berechnen: $PW = \begin{cases} (a) & P(Z \geq z) \\ (b) & P(Z \leq z) \\ (c) & 2P(Z \geq z) \end{cases}$ Achtung: - Stichprobenumfang: $n \geq 50$ - Zufallstichprobe $\rightarrow$ nicht Zeitreihen - Nicht nichtlineare Beziehungen - Nicht Ausreisser
--

e_i Details. e_i müssen eine Zufallstichprobe darstellen: e_i sind unabhängig voneinander - Die e_i alle gleiche Verteilung $\rightarrow$ gleiche SA $\rightarrow$ i.d. Praxis aber oft verletzt! - Erkennen: - Streuungsdiagramm - Residuendiagramm $\Rightarrow$ Fächerform
--

'Exakte' Inferenz. - Annahme: Fehler e_i mit Normalverteilung $e_n \sim N(0, \sigma)$; beliebiges n $\Rightarrow$ benutzt t_{n-2}-Verteilung statt $N(0, 1)$ für KI und FW.. $\Rightarrow$ KI = Schätzer $\pm t_{n-2, 1-\frac{\alpha}{2}} \times SF$ $\Rightarrow$ PW = (a) $P(t_{(n-2)} \geq z)$ - Wenn Annahmen nicht zu treffen, nur bei gros-sem n [auf $N(0, 1)$ Bats] möglich
--

Regression durch den Ursprung. - Annahme: $\beta_0 = 0$ $y_i = \beta_1 x_i + \epsilon_i$ $i = 1, \dots, n$ - unbekannter Parameter: β_1 - Kleinste-Quadrate Schätzer: $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$
--

$s^2 = \frac{1}{n-1} \sum_{i=1}^n \epsilon_i^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \hat{\beta}_1 x_i]^2$ Erwartungswert und Varianz: $E(\hat{\beta}_1) = \beta_1$ $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$ $SF(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum x_i^2}}$

- KI: $\hat{\beta}_1 \pm z_{1-\frac{\alpha}{2}} \times \sqrt{\frac{s^2}{\sum x_i^2}}$ - Test-Statistik für Hypothesentest: $z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)} \quad \text{wobei: } SF(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum x_i^2}}$ - Die 'exakte' Inferenz benutzt t_{n-1} anstelle von $N(0, 1)$ - Vorteile ($\beta_0 = 0$) $\hat{\beta}_1$ wir genauer geschätzt (kleinere Vari-anz) - Vorhersagen sind i.d.R. genauer - Nachteile ($\beta_0 \neq 0$) - Schätzen das Modell zu klein $E(\hat{\beta}_1) \neq \beta_1$ - Vorhersagen ungenauer
--

Strategie: - Theoretischer Grund für $\beta_0 = 0$ - Schätze das allg. Modell und teste H_0 : $\beta_0 = 0$ - wenn der Test nicht verwirft, dann steige um auf das vereinfachte Modell
--

2.5 EINFACHE REGRESSION - VORHERSAGE

Idealfall. - Annahme: Wir kennen β_0 und β_1 $y_{neu} = \beta_0 + \beta_1 x_{neu} + \epsilon_{neu}$ sonit : $\hat{y}_{neu} = \beta_0 + \beta_1 x_{neu}$ $E(y_{neu}) = \mu_{neu}$ - Vorhersage der Zufallsvariablen $y_{neu} = \mu_{neu} + \epsilon_{neu}$ - Wichtiger Unterschied: $E(\mu_{neu})$ kennen wir genau - Realisierung von y_{neu} können wir nie ge-nau vorhersagen, da von ϵ_{neu} abhängt
--

Realfall. - Annahme: Wir kennen i.d.R. β_0 und β_1 nicht $y_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu} + \epsilon_{neu}$ sonit : $\hat{y}_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu}$ $E(y_{neu}) = \mu_{neu}$ - Vorhersage der Zufallsvariablen $y_{neu} = \mu_{neu} + \epsilon_{neu}$ - Realisierung von y_{neu} ungenauer, da noch Un-gewissheit von μ_{neu}
--

KI für μ_{neu}. - Der KI für den Mittelwert macht eine langfri-stige Aussage. - Der Schätzer ist erwartungstreu: $E(\hat{\mu}_{neu}) = E(\hat{\beta}_0 + \hat{\beta}_1 x_{neu}) = E(\hat{\beta}_0) + E(\hat{\beta}_1) x_{neu} = \beta_0 + \beta_1 x_{neu} = \mu_{neu}$ - KI für μ_{neu} daher: $Var(\hat{\mu}_{neu}) = \sigma^2 \left(\frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)$ - KI für μ_{neu} daher: $\hat{\mu}_{neu} \pm z_{1-\frac{\alpha}{2}} \times s \sqrt{\frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ - Die exakte Inferenz benutzt $t_{n-2, 1-\frac{\alpha}{2}}$
--

VI für y_{neu}. $Var(\hat{y}_{neu} - y_{neu}) = Var(\hat{\mu}_{neu}) + \sigma^2 = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)$ - Der VI macht eine kurzfristige Aussage für einen bestimmten Fall. - Zusätzlich zur Variation $\hat{\mu}_{neu}$ noch Variation von ϵ_{neu} - VI für y_{neu}: $\hat{y}_{neu} \pm z_{1-\frac{\alpha}{2}} \times s \sqrt{1 + \frac{1}{n} + \frac{(x_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$
--

Voraussetzungen. - Für μ_{neu} - Daten unabhängig voneinander - Konstante Fehler-SA σ - Stichprobe $n \geq 50$ - Zusätzlich für VI für y_{neu} - Fehler ϵ_i haben eine Normalverteilung
--

Beurteilung Normalität. - Normal-Quantil-Diagramm (N-Q-D) - Heavy Tails: Links Punkte unterhalb der Geraden und rechts oberhalb - Light Tails: Links Punkte oberhalb der Geraden und rechts unterhalb - Rechtschief: beide oberhalb - Linksschief: beide unterhalb

N-Q-D für (standardisierte) Residuen. - Go: - Gerade $\Rightarrow$ VI funktionieren 'richtig' - Light Tails $\Rightarrow$ VI sind 'konservativ' - No-Go: - Schiefe $\Rightarrow$ VI sind oft 'antikonservativ' - Heavy Tails $\Rightarrow$ VI sind 'antikonservativ'
--

2.6 EINFACHE REGRESSION - CAPM

Capital Asset Pricing Model. X = Risiko-Prämium des Markt-Portfolios Y = Risiko-Prämium eines Aktivums - Wobei Risiko-Prämie = Rendite - risikofreie Rendite

Theorie und Praxis.

- Die erwartete Risiko-Prämie des Aktivums ist proportional zu der erwarteten Risiko-Prämie des Marktes:

$$E(y_i) = \beta E(x_i) \quad \text{anders formuliert: } y_i = \beta x_i + e_i$$

- statt β_0, β_1 α und β
- $y_i = \alpha + \beta x_i + e_i$

Steigung

- β misst daher das Risiko des Aktivums in Relation zum Markt
- Aktivum mit $\beta < 1$ 'defensiv'; Aktivum mit $\beta > 1$ 'aggressiv' (jeweils in Relation zum Markt)

Achsenabschnitt

- α ist die durchsch. Risiko-Prämie des Akt. wenn die Risiko-Prämie des Marktes = 0 ist
- CAPM besagt: $\alpha = 0$
- Wenn CAPM irrt: $\alpha \neq 0$

Schätzung

- Wir schätzen α und β in der gewohnten Weise

Inferieren

- n genug gross
- Obwohl Daten eine Zeitreihe darstellen, hier erlaubt, da Aktien-Renditen und Zeit fast unabhängig voneinander.
- Residuen-Diagramm zeigt keine Fächerform

VI

- Da Aktienrenditen 'Heavy Tails' und nicht Normalverteilung $\rightarrow$ VI nicht gültig

3.1 MULTIPLE REGRESSION - EINFÜHRUNG

- Mehrere Regressoren: $X_1, X_2, \dots, X_n$
- Vorteil:** Können Y besser modellieren und vorhersagen
- Nachteil:** komplexer, neue Gefahren

$$y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + e_i \quad i = 1, \dots, n$$

- $\beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k}$ beschreibt die *Hyper-Ebene*
- Der **Fehler** e_i beschreibt die Abweichung der Hyper-Ebene
- Annahmen: siehe 2.3
- Notation: $\beta_j = j$ -te Steigung ($j = 1, \dots, k$)

Vektor- und Matrizen-Notation.

- $\mathbf{Y} = (y_1, \dots, y_n)'$ $\Rightarrow n \times 1$ Vektor
- $\mathbf{X} = (x_{i,j})_{1 \leq i \leq n, 0 \leq j \leq k} \Rightarrow n \times (k+1)$ Matrix (wobei $x_{i,0} = 1$ stets)
- $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$ $\Rightarrow (k+1) \times 1$ Vektor
- $\mathbf{e} = (e_1, \dots, e_n)'$ $\Rightarrow n \times 1$ Vektor
- Das Modell ist dann:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Bewertung der Beziehung.

- Ist die Punkte-Wolke in 'Mengen'- R^{k+1} ein Chaos?
- ja $\rightarrow$ keine Beziehung
- nein $\rightarrow$ Beziehung
- Falls nein, formen doe Punkte grob eine Hyper-Ebene?
- ja $\rightarrow$ linear
- nein $\rightarrow$ nichtlinear

Streudiagramm mit multiplen Regressoren.

- Für jede X-Variable einzeln ein Streudiagramm zu machen und dann 'kombinieren' $\rightarrow$ mühsam + evtl voll daneben
- Alle X-Variablen in 1 Regressionsmodell aufstellen!

Multiplies Regressions-Modell.

- $y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,(+1)} + e_i$
- Wenn $x_{i,1}$ um 1 steigt und $x_{i,(+1)}$ konstant, dann erhöht sich y_i um β_1
- D.h. β_1 ist der Effekt von $x_{i,1}$ 'kontrolliert' für $x_{i,(+1)}$
- Wenn $x_{i,(+1)}$ um 1 steigt und $x_{i,1}$ konstant, dann erhöht sich y_i um β_2
- D.h. β_2 ist der Effekt von $x_{i,(+1)}$ 'kontrolliert' für $x_{i,1}$
- Dies sind die Effekte, die wir suchen.

Geschätztes multiples Regressions-Modell.

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,(+1)}$$

Gefahren.

- Separate Schätzung versagt, weil
 - evtl negative Korrelation zwischen X-Variablen
 - Sachbezogene Konflikte
- Falls wir nicht 'kontrollieren'
 - Unterschätzung der einzelnen X-Variablen Effekte.
- Wichtige Variablen gehören ins Modell an auch wenn uns deren Effekt gar nicht interessiert!

Beziehungen und Ausreisser.

- Lin. oder Ausreisser nicht mit individ. Streuungs-Dia.
- Individ. Bez. nichtlinear, aber Gesamt-Bez. linear möglich
- Individ. Bez. können Ausreisser haben, Gesamt-Bez. aber nicht mehr
- $\Rightarrow$ Stattdessen Residuen-Diagramm

3.2 SCHÄTZUNG

- Wähle Hyper-Ebene, die GF minimiert.
- Kandidat für Hyper-Ebene gegeben durch $\mathbf{b} = (b_0, b_1, \dots, b_k)'$
- Der Globale Fehler ist gegeben als

$$GF(\mathbf{b}) = \sum_{i=1}^n [(y_i - (b_0 + b_1 x_{i,1} + \dots + b_k x_{i,k}))]^2 = ||\mathbf{Y} - \mathbf{X}\mathbf{b}||^2$$

- Kleinste-Quadrate-Schätzer für $\boldsymbol{\beta}$ ist: $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$
- für $\mathbf{e}$ ist: $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$

für σ^2 ist:

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

EW, Varianz und Kovarianz-Matrix

- Sei $\mathbf{W}$ ein Zufallsvektor mit $(W_1, W_2, \dots, W_n)' \in \mathbb{R}^n$
- $E(\mathbf{W}) = (E(W_1), \dots, E(W_n))' \in \mathbb{R}^n$
- $\text{Var}(\mathbf{W}) = \sum (\sigma_{i,j}) \in \mathbb{R}^{n \times n}$
- $\sum = \begin{pmatrix} \text{Var}(W_1) & \text{Cov}(W_1, W_2) \\ \text{Cov}(W_2, W_1) & \text{Var}(W_2) \end{pmatrix}$
- $\sigma_{i,i} = \text{Var}(W_i)$ (Hauptdiagonale)
- $\sigma_{i,j} = \text{Cov}(W_i, W_j) = \text{Cov}(W_j, W_i)$ für $i \neq j$

Linearkombinationen

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

- $E(\mathbf{a} + \mathbf{b}\mathbf{W}) = \mathbf{a} + \mathbf{b}E(\mathbf{W})$
- $\text{Var}(\mathbf{a} + \mathbf{b}\mathbf{W}) = \mathbf{b}\text{Var}(\mathbf{W})\mathbf{b}'$
- $\rightarrow \mathbf{b}^2 \text{Var}(\mathbf{W})$ nur sinnvoll für $n=m=1$.

Eigenschaften der Schätzer

- $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$
- $E(s^2) = \sigma^2$
- $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ (Kovar.-matrix von $\hat{\boldsymbol{\beta}}$)

Standardisierte Residuen Die $\hat{e}_i$ haben nicht die gleiche Standardabweichung:

- Kov.-Matrix von $\hat{\mathbf{e}}$ ist: $\mathbf{R} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ und $\text{Var}(\hat{\mathbf{e}}) = \sigma^2(\mathbf{I} - \mathbf{R})$
- $SA(\hat{e}_i) = \sigma\sqrt{1 - r_{ii}}$ und $SF(\hat{e}_i) = s\sqrt{1 - r_{ii}}$
- Verfeinerung ergibt:
- $\text{Stand.} - \text{Resid.} = \frac{\hat{e}_i}{s\sqrt{1 - r_{ii}}}$

Das Normal-Quantil Diagramm in R benutzt die stand. Residuen.

Erkennen der Gefahren.

- Nichtlinearität: Das Residuen-Diagramm ist ein Muster und kein Chaos
- Ausreisser: Oft anhand des Residuen-Diagramms, besser: Cook's Distance

Cook's Distance

- für die i -te Beobachtung und β -Schätzer
- $D_i = \frac{(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i)' \mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_i)}{(k+1)s^2}$

- Äquivalent: $\mathbf{R} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ und dann: $D_i = (\text{Stand.} - \text{Resid.})^2 / (\frac{1}{1 - r_{ii}}) \frac{1}{k+1}$
- $(\text{Stand.} - \text{Resid.})^2$ misst Ausreisser; 2. Faktor extremes x für k

Für Inferenz benötigen wir konstante Standardabweichungen. R kann $\sqrt{|\text{Stand.} - \text{Resid.}|}$ gegen die $\hat{y}_i$ plotten. Wegen der Wurzel des Absolutwertes ist lediglich eine halbe Fächerform Kennzeichen für eine Gefahr.

Stärke der Beziehung

- Die Stärke der linearen Beziehung wird durch
- $R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{SS_{Exp}}{SS_{Tot}} \in [0, 1]$
- und erklärt wie viel der Variation von Y durch das Modell erklärt wird.
- Für nur eine X-Variable gilt: $R^2 = r^2$
- Der Wert für R ist im Software-Output zu finden.

3.3 INFERENZ

- Konf.intervall für β_j und $H_0: \beta_j = \beta_{j,0}$
- $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$
- $H_0: \beta_1 = \dots = \beta_k = 0$

Konfidenzintervall β_j .

- KI = Schätzer $\pm z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{F}}$
- SF: Wurzel des j -ten Diagonalelementes aus der Kovarianzmatrix aus 3.2. Die Software liefert diesen Wert (Std. Error)
- Anstelle von $z_{1-\frac{\alpha}{2}}$ benutzt 'exakte' Inferenz $t_{n-k-1, 1-\frac{\alpha}{2}}$.

Hypothesentest β_j .

- $H_0: \beta_j = \beta_{j,0}$
- H_A : wie immer
- $z = \frac{\hat{\beta}_j - \beta_{j,0}}{SF(\hat{\beta}_j)}$
- p-Wert: wie immer

Die Software liefert uns die Teststatistik ('t value') und den p-Wert für einen zweiseitigen Test ('P>r(> |t|)') zu einem Signifikanzniveau.

Erkennen der Gefahren.

- $n \geq 50$
- Zufallsstichprobe: Daten unabhängig, konstante Fehler-Standardabweichung
- Beziehung sollte linear sein und keine Ausreisser enthalten (Gefahren wie 3.2)

alg. Hypothesentest.

- Tests wie $H_0: \beta_1 = \beta_2 = 0$ oder $H_0: \beta_2 + \beta_3 = 1$ sind auch möglich.
- Jede Kombination stellt eine lineare Restriktion dar. Die Anzahl wird mit q bezeichnet. Formal: $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$; Typen: $R: [q, k+1]$; $\mathbf{r}: [q, 1]$
- Für einen Vergleich von H_0 und H_A werden beide Modelle geschätzt. Das alternative Modell wird nur angenommen, wenn es bedeutend besser ist, da es in jedem Fall mehr erklärt (kann sich den Daten besser anpassen).

$$F = \frac{(R_0^2 - R_1^2)/q}{(1 - R_1^2)/(n - k - 1)} = \frac{(SS_{Res,0} - SS_{Res,A})/q}{SS_{Res,A}/(n - k - 1)}$$

SS_{Res} = Freiheitsgrade $\times$ (Freiheits-Standard Error)²

- p-Wert: $P(F_{q, n-k-1} > F)$
- Falls $PW \leq \alpha$ oder $F \geq F_{1-\alpha, q, n-k-1}$ wird H_0 verworfen. (Kritische Werte von F werden an der Prüfung angegeben.)
- Für $H_0: \beta_1 = \dots = \beta_k = 0$ vereinfacht sich die Formel:

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

Wenn wir das Vorzeichen eines Parameters ex ante kennen und das Modell dasselbe schätzt, ist der PW des einseitigen Tests die Hälfte des zweiseitigen des Software-Outputs.
VORSICHT: Sollte die Zielvariable unter H_A anders sein als unter H_0 , darf F nicht mit R^2 berechnet werden! (Weil $SS_{Tot,0} \neq SS_{Tot,A}$)

3.4 VOHERSAGE

Konfidenzintervall für $\mu_{new} = E(\mu_{new})$.

- Schätzer: $\hat{\mu}_{new} = \mathbf{a}'_{new}\hat{\boldsymbol{\beta}}$
- Eigenschaften
- $\text{Var}(\hat{\mu}_{new}) = \sigma^2 \mathbf{a}'_{new} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}_{new}$
- $SF(\hat{\mu}_{new}) = s\sqrt{\mathbf{a}'_{new} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}_{new}}$
- KI = $\hat{\mu}_{new} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{F}} \approx s\sqrt{\mathbf{a}'_{new} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}_{new}}$
- exakter: $t_{n-k-1, 1-\frac{\alpha}{2}}$
- in R: $SF = \mathbf{sso.f}$ it

! Zufallsstichprobe und $n \geq 50$!

Vorhersageintervall für y_{new} .

- Vorhersage: $\hat{y} = \mathbf{a}'_{new}\hat{\boldsymbol{\beta}}$
- Eigenschaften:

$$\begin{aligned} E(\hat{y}_{new} - y_{new}) &= \mu_{new} - \mu_{new} = 0 \\ \text{Var}(\hat{y}_{new} - y_{new}) &= \sigma^2 \mathbf{a}'_{new} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}_{new} + \sigma^2 \\ SF(\hat{y}_{new} - y_{new}) &= s\sqrt{1 + \mathbf{a}'_{new} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{a}_{new}} \end{aligned}$$

- VI unter Normalität
- in R: $SF = \sqrt{\mathbf{sresidual.scale}^2 + \mathbf{sso.f}it^2}$

! Zusätzlich zu KI: e_i müssen normalverteilt sein! (Beachte das R-Diagramm mit den standardisierten Residuen)

3.5 DUMMY-VARIABLEN

- relevant für die Schätzung unterschiedlicher Geraden (gemeinsame Gerade, gemeinsame Steigung, gemeinsamer Abschnitt, total verschieden)
- wir nehmen immer das kleinste adäquate Modell

gemeinsame Steigung.

- Definiere eine Variable D_i mit den Werten 0 oder 1.
- Das Modell lautet:

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

- Für $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- Für $D_i = 1$: Abschnitt = $\beta_0 + \delta$, Steigung = β_1
- δ ist der geschätzte Unterschied der beiden Abschnitte
- Ist der p-Wert für δ klein (die Schätzung also signifikant), ist 'gemeinsame Gerade' inadäquat.
- Im(salary ~ gender + years) 'gender' = Dummy-Variable

gemeinsamer Abschnitt.

- Definiere eine Variable D_i mit den Werten 0 oder 1.
- Das Modell lautet:

$$G_i = \beta_0 + \beta_1 A_i + \gamma(D_i \times A_i) + e_i$$

- Für $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- Für $D_i = 1$: Abschnitt = β_0 , Steigung = $\beta_1 + \gamma$
- γ ist der geschätzte Unterschied der beiden Steigungen
- Ist der p-Wert für γ klein (die Schätzung also signifikant), ist 'gemeinsame Gerade' inadäquat.
- Im(salary ~ years + I(gender * years))

total verschieden.

- Definiere eine Variable D_i mit den Werten 0 oder 1.
- Das Modell

$$G_i = \beta_0 + \delta D_i + \beta_1 A_i + \gamma(D_i \times A_i) + e_i$$

- Für $D_i = 0$: Abschnitt = β_0 , Steigung = β_1
- Für $D_i = 1$: Abschnitt = $\beta_0 + \delta$, Steigung = $\beta_1 + \gamma$
- Ist die ergänzte Dummy-Variable (Abschnitt oder Steigung) signifikant, sollte zu diesem Modell gewechselt werden.
- Im(salary ~ gender + years + I(gender * years))

Inferenz.

- Ist der Parameter für $D_i = 1$ und $D_i = 0$ derselbe (z.B. Steigung bei 'gemeinsame Steigung'), ist die Antwort im Output.
- Sonst ist die Antwort anders zu berechnen.
 - Geschätzte Kovarianz zwischen den Parametern ('gemeinsame Steigung': $\hat{\beta}$ und $\hat{\delta}$). Diese ist aber nicht im Output enthalten. Von Hand rechnen müssen.
 - Umschreiben des Modells.

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

- Im(salary ~ 1 + I(1 - gender) + gender + years)
 - '1' ist die globale Konstante, die entfernt werden muss
 - im Output sind nun Werte enthalten, mit denen inferiert werden kann.
 - Für 'gemeinsamer Abschnitt':

$$G_i = \beta_0 + \beta_1, D_i([1 - D_i] \times A_i) + \beta_1, D_i(D_i \times A_i) + e_i$$

mehrere Gruppen.

- Für mehr als 2 Gruppen gibt es eine Basisgruppe, die anderen werden mit eigenen Dummy-Variablen modelliert.
- $G_i = \beta_0 + \delta_1 D_{1,i} + \delta_2 D_{2,i} + \beta_1 A_i + e_i$
- $G_i = \beta_0 + \beta_1 A_i + \gamma_1(D_{1,i} \times A_i) + \gamma_2(D_{2,i} \times A_i) + e_i$
- Im(weight ~ t1 + d2 + weeks + I(d1 * weeks) + I(d2 * weeks)) (total verschieden)
- Um zwei Modelle zu vergleichen, benutzen wir die F-Statistik (siehe oben). Ist der p-Wert klein, verwerfen wir H_0 .
- Die Schätzung z.B. für eine Steigung setzt sich aus dem Wert der Basisgruppe und dem Wert des Parameters selbst zusammen. (wichtig für Intervalle und/oder Vorhersage)
- Das Spiel kann für mehrere Gruppen beliebig fortgesetzt werden.

Chow-Test.

- Multiplies Regressions-Modell in g Gruppen
- H_0 : Gemeinsames Modell
- H_A : Total Verschieden
- $SS_{Res,0}$ vom gemeinsamen Modell
- $SS_{Res,A}$ von jeder einzelnen Gruppe schätzen und aufaddieren.
- Anzahl Restriktionen: $p = (g - 1) \times (k + 1)$
- Gesamt-Freiheitsgrade: $FC = n - g \times (k + 1)$
- Alternative für $SS_{Res,A}$: Dummy-Variablen für jeden Parameter einsetzen und dann das eine Modell schätzen.

3.6 NICHTLINEARE BEZIEHUNGEN

Polynomiale Beziehung.

- Parameter werden zur linearen Schätzung auch noch mit einer höheren Potenz geschätzt
- $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + [\beta_3 x_i^3] + e_i$
- Ausreisser müssen entfernt werden
- Der marginale Nutzen einer zusätzlichen Potenz kann an seiner Signifikanz abgelesen werden.
- Modelle vom Grad > 1 bedürfen anderer Methoden zur Interpretation.

$$\text{Effekt}(x) = \frac{dE(y)}{dx}$$

- ACHTUNG: Residuen-Diagramm $\neq$ Chaos, polynom. Modell inadäquat.

Nichtlineare Transformation.

- Transformation theoretisch motiviert

$$h(y_i) = \beta_0 + \beta_1 g(y_i) + e_i$$

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

- Interpretation

$$\text{Effekt}(x) = \frac{dE(y)}{dx}$$

- Eine höhere R^2 -Statistik lässt auf das neue Modell schliessen.
- Inferenz wie gewohnt, aber eine 'naheliegende' Interpretation bedarf einer Rücktransformation!

Typ	β_1	η	Bedingung
linear	$\beta_1 = \Delta y / \Delta x$	$\beta_1 x / y$	
log - lin	$100\beta_1 \approx \% \Delta y / \Delta x$	$\beta_1 x$	$y > 0$
lin - log	$\beta_1 / 100 \approx \Delta y / \% \Delta x$	β_1 / y	$x > 0$
log - log	$\beta_1 = \% \Delta y / \% \Delta x$	β_1	$x, y > 0$

Interaktion.

- Interaktion, falls effekt von β_j nicht unabhängig von anderen Variablen.
- $y_i = \beta_0 + \beta_1 G_i + \beta_2 A_i + \beta_3(A_i \times G_i) + e_i$
- Sowohl G, als auch A hängen nun vom Niveau des anderen ab.
- Gefahren-Diagramm besser? Signifikanz vorhanden?

Vergl. (nicht) genesteter Modelle.

- genestet = Das eine Modell ist im anderen enthalten.
- F-Test geht nur für genestete Modelle
- R^2 bevorzugt systematisch mehr Regressoren
- R^2 (adjustiertes R) berücksichtigt Anzahl Regressoren

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

- grösseres k, grösseres Handicap, R^2 kann steigen oder fallen in