

Kovarianz: Beschreibung Lineare Beziehung zweier numerischer Variablen (X und Y) Bevölkerungs-Variante: Verteilung der Zufallsvariablen X und Y (in der Praxis nicht bekannt) $Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$ Stichproben-Variante: Die von den Daten berechnete Kovarianz $cov = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ x_i, y_i : Beobachtungen $\bar{x}, \bar{y}$: Mittelwerte Kovarianz für Mittelwerte: $Cov(\bar{X}, \bar{Y}) = \frac{1}{n} Cov(x_i, y_i)$ $Cov(X, Y) = \frac{Var(X+Y) - Var(X) - Var(Y)}{2}$ Interpretation • Eine positive (negative) Kovarianz bedeutet eine positive (negative) lineare Assoziation zwischen X und Y – Grosse Werte von X gehen mit grossen (kleinen) Werten von Y einher • Wichtig: Wert der Kovarianz von den Einheiten von X und Y abhängig - kein absoluter Wert! → Die Stärke der Beziehung kann mit der Kovarianz nicht beurteilt werden Korrelation Bevölkerungs-Variante: $Cor(X, Y) = \frac{Cov(X, Y)}{SA(X)SA(Y)}$ $SA(X) = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2}$ Stichproben-Variante: $r = \frac{cov}{s_x s_y} = \frac{s_{xy}}{s_x s_y}$ Eigenschaften der Korrelation: $-1 \leq Cor(X, Y) \leq 1$ sowie auch $-1 \leq r \leq 1$ • Linear-Transformationen von X und Y ändern den Wert nicht! aber ev. Vorzeichen • Positive (negative) Korrelation → positive (negative) Assoziation zwischen X und Y • Je näher Absolutwert der Korrelation an 1 ist, desto stärker die Beziehung • Korrelation: Richtung als auch Stärke der linearen Beziehung zwischen X und Y beurteilen → Korrelation unangebracht für nichtlineare Beziehungen : Korrelation kann negativ sein, obwohl diese eigentlich positiv ist ab einem Punkt! → Stichproben-Variante r ist nicht robust gegen Ausreisser Theorie zu Korrelation Untersuchung zweier numerischer Variablen mit $r = 0.97$ → Beziehung kann nichtlinear sein → Beziehung kann linear sein, aber einen extremen Ausreisser vorhanden → Eine nichtlineare Transformation einer der beiden Variablen wird in einer grösseren Stichproben-Korrelation resultieren → Die Residuen haben eine andere Korrelation und sollten i.d.R 0 sein Rechenregeln für Erwartungswert $E(X + Y) = E(X) + E(Y)$ Rechenregeln für Varianz $Var(X + Y) = Var(X - Y) = Var(X) + Var(Y)$ wenn unabhängig $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$ gilt immer $Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$ gilt immer → Wenn X und Y unabhängig sind, dann besteht keine Beziehung, so dass $Cov(X, Y) = 0$ → Wichtig: Obwohl $Cor(X, Y) = 0$ und $Cov(X, Y) = 0$, kann es sein, dass X und Y abhängig sind, wenn es sich etwa um eine nichtlineare Beziehung handelt. → Anwendung bei falschem Modell → systematischer Fehler Datenpaare Für (1) unabhängige und (2) abhängige Stichproben: (1) $SF_{Schätzer} = \sqrt{\frac{(s_x^2 + s_y^2)}{n}}$ (2) $SF_{Schätzer} = \sqrt{\frac{(s_x^2 + s_y^2) - 2cov}{n}}$ Korrelation innerhalb eines Paares = $\begin{cases} \text{Positiv} \rightarrow (2) \text{ besser} \\ \text{Keine} \rightarrow \text{gleich gut} \\ \text{Negativ} \rightarrow (1) \text{ besser} \end{cases}$ $SF = \sqrt{\frac{(s_x^2 + s_y^2) - 2cov}{n}} = \sqrt{\frac{(s_x^2 + s_y^2) - 2r s_x s_y}{n}}$ → Berücksichtigt Assoziation direkt! $s_D = \sqrt{(s_x^2 + s_y^2) - 2r s_x s_y} \rightarrow SF = \frac{s_D}{\sqrt{n}}$ für Inferenz Kovarianz: Rechenregeln $Cov(X, X) = Var(X)$ $Cov(X, Y_1 + Y_2) = Cov(X, Y_1) + Cov(X, Y_2)$ $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$ $Cov(X, bY) = bCov(X, Y)$ $Cov(X, a) = 0$ Korrelation: Rechenregeln: $Cor(a + bX, c + dY) = \frac{b \cdot d \cdot Cov(X, Y)}{ b \cdot d \cdot SA(X) \cdot SA(Y)} = cor(X, Y)$ Verändern! Lineare Transformation der Korrelation: Korrelation ändert sich nicht, wenn Einheiten einer der oder beider Variablen linear geändert werden falls b > 0 d > 0! Schätzung Regressions-Modell: $y_i = \beta_0 + \beta_1 x_i + u_i$ • β_0 : Achsenabschnitt, β_1 : Steigung, $\beta_0 + \beta_1 x_i$: Gerade, u_i : Fehler beschreibt Abweichung einer Beobachtung von der Gerade Annahmen für das Modell • Die Regressoren x_i sind deterministisch • Die Fehler u_i sind eine Zufalls-Stichprobe von einer gemeinsamen Verteilung mit: $\mu = 0$ (im Mittel sind die Fehler Null) und (unbekanntem) $\sigma > 0$ Vorhersage von x_{neu}: $\hat{y}_{neu} = \beta_0 + \beta_1 x_{neu}$ • Man beobachtet x_{neu} , eine neue Realisierung der X-Variable Kleinste-Quadrate Schätzung: $(\hat{\beta}_0, \hat{\beta}_1) = \argmin_{(\beta_0, \beta_1)} \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2$ Geschätzter Achsenabschnitt: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ Geschätzte Steigung: $\hat{\beta}_1 = r \frac{s_y}{s_x}$ Zugehöriger Kleinste-Quadrate Schätzer für $\sigma^2 s^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n-2} \sum_{i=1}^n [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2$ • Die geschätzten Fehler $\hat{u}_i$ heissen Residuen Erwartungswerte und Varianzen Alle drei Schätzer $\hat{\beta}_0, \hat{\beta}_1, s^2$ sind erwartungstreu (unbiased) • $E(\hat{\beta}_0) = \beta_0, E(\hat{\beta}_1) = \beta_1, E(s^2) = \sigma^2$ Varianzen von $\hat{\beta}_0$ & $\hat{\beta}_1$: $Var(\hat{\beta}_0) = \sigma^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}$ $Var(\hat{\beta}_1) = \sigma^2 \frac{1}{\sum (x_i - \bar{x})^2}$ → Formeln wichtig für Inferenz!	Wichtige Formeln für Varianz- Dekomposition: $\sum (x_i - \bar{x})^2 = (n-1) \cdot s_x^2$ $\sum (x_i - \bar{x})^2 = \sum x_i^2 - n \cdot \bar{x}^2$ $\sum x_i^2 = (n-1) s_x^2 + n \cdot \bar{x}^2$ Gauss-Markov Theorem Theorem: $\hat{\beta}_0$ und $\hat{\beta}_1$ sind die besten Schätzer → kleinsten Varianzen → $\hat{\beta}_0$ und $\hat{\beta}_1$ sind BLUE (Best Linear Unbiased Estimator) Extrapolation: Vorhersage für Wert x_{neu} weit ausserhalb der beobachteten X-Werte → Nicht dann ausgehen, dass immer linear ist für X-Werte weit ausserhalb von beobachteten Daten Schlummernde Variablen: Versteckte Variablen → Assoziationen bedeutet nicht unbedingt Verursachung → Kausalitäten sind in der Regressionsanalyse nicht zulässig Transformation Einheiten: $X: 1 \text{ inch} \rightarrow 2.54 \text{ cm} \rightarrow X^* = 2.54 \cdot X$ $Y: 1 \text{ pound} \rightarrow 0.45 \text{ kg} \rightarrow Y^* = 0.45 \cdot Y$ $\rightarrow \hat{\beta}_1 = \frac{r_{x^*y^*} s_{y^*}}{s_{x^*}} = \frac{2.54 \cdot 0.45 r_{xy}}{2.54 s_x} = 0.45 r_{xy}$ $\hat{\beta}_0 = 0.57 - 0.2 \cdot \hat{\beta}_1 \cdot 2.54 = 0.5(\bar{y} - \hat{\beta}_1 \bar{x})$ → R^2 : Keine Änderung, da einheitslos → $SER = 0.5 \cdot SER$ → Bezug auf Y-Einheiten! Residuen-Diagramm → Funktioniert auch wenn man mehrere X-Variablen hat! • $\hat{y} - \hat{u}$ Diagramm → Plot Residuen gegen gefittete (geschätzten) Y-Werte • Idealerweise Chaos → Punkte „zufällig“ in horizontalen Band → Lineare Beziehung • Nichtlineare Beziehungen ergeben nichtlineares Residuen-Diagramm • Ausreisser sind i.d.R auch erkenntlich: Entfernt vom Rest und Band für Rest nicht horizontal Bedingungen für Residuen: $\sum_{i=1}^n \hat{u}_i = 0$ → Summe Residuen 0 sein $\sum_{i=1}^n \hat{u}_i \cdot x_i = 0$ → Keine Korrelation zwischen $\hat{u}_i$ und x_i Interpretation des Residuen-Diagramms → Das Konfidenzintervall für die Steigung $\hat{\beta}_1$ ist nur dann bedeutend basierend auf der t-Verteilung anstatt der Normalverteilung, wenn $n \leq 50$? → Wenn eine Fächerform vorhanden ist, dann ist die Beziehung immer noch linear → Die Fächerform deutet lediglich auf Heteroskedastie hin nicht konstant Einflussreiche Beobachtungen • Wenn Ausreisser extrem: Geschätzte Gerade ganz nahe zu sich ziehen → Zugehörige Residuum ist klein und man erkennt den Ausreisser nicht mehr im Residuum-Diagramm Cooks Distance: Numerisches Mass, dass den Einfluss eines Datenpunktes misst • Kombination der Faktoren „Ausreisser“ und „Extremes X“ Varianz-Dekomposition TSS (Total) = ESS (Explained) + RSS (Residuals) bzw SST = SSE + SSR $\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n \hat{u}_i^2$ • ESS: Teil Variation, der von Modell erklärt wird • RSS: (residuale) Teil Variation, der von Modell nicht erklärt wird → Varianz-Zerlegung gilt nur, wenn die Konstante (d.h. Abschnitt) im Modell ist → Wenn Konstante nicht enthalten → R^2 nicht interpretierbar R^2-Statistik (=Bestimmtheitskoeffizient) $R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ • Prozentsatz der Variation von Y, der durch Modell erklärt wird • Wichtig: R^2 ist einheitslos • Funktioniert auch im multiplen Regressionsmodell • Im Falle einer X-Variable oder im linearen Modell gilt: $R^2 = r^2$ $R^2 = r^2 = (\frac{\hat{\beta}_1 s_x}{s_y})^2$ • Das R^2 von Y auf X ist das gleiche R^2 von der Regression X auf Y: $\frac{s_y^2}{r s_x^2} = \frac{r s_y^2}{r^2 s_x^2}$ Wichtige Umformungen: $TSS: \sum_{i=1}^n (y_i - \bar{y})^2 = (n-1) \cdot s_y^2 \rightarrow s_y = \sqrt{\frac{1}{n-1} \cdot TSS}$ SER: $\sqrt{\frac{RSS}{n-k-1}} \rightarrow RSS: (n-k-1) \cdot SER^2$ Inferenz Konfidenzintervall für β_0 : $KI = \hat{\beta}_0 \pm z_{1-\alpha/2} \cdot \sqrt{s^2 \frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$ Konfidenzintervall für β_1 : $KI = \hat{\beta}_1 \pm z_{1-\alpha/2} \cdot \sqrt{s^2 \frac{1}{\sum (x_i - \bar{x})^2}}$ Hypothesentest 1) Hypothesen aufstellen 2) Test-Statistik berechnen: $z = \frac{(\hat{\beta}_0 - \beta_{0,0})}{SF(\hat{\beta}_0)}$ 3) P-Wert: $PW = \begin{cases} (a) P(Z \geq z) \rightarrow 1 - \varphi(z) \\ (b) P(Z \leq z) \rightarrow \varphi(z) \\ (c) 2P(Z z) \rightarrow 2 \cdot (1 - \varphi(z)) \end{cases}$ Hypothesentest: Test-Statistik $H_0: \beta_1 = \beta_2 \rightarrow \beta_1 - \beta_2 = 0; H_1: \beta_1 \neq \beta_2 \rightarrow \beta_1 - \beta_2 \neq 0$ $t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - (\beta_{1,0} - \beta_{2,0})}{SF(\hat{\beta}_1 - \hat{\beta}_2)} = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{\sqrt{Var(\hat{\beta}_1) + Var(\hat{\beta}_2) - 2Cov(\hat{\beta}_1, \hat{\beta}_2)}}$ Wichtige Annahmen für Inferenz • Stichprobenumfang: $n \geq 50$, ansonsten t-Verteilung • Zufalls-Stichprobe: Methoden funktionieren i.d.R nicht für Zeitreihen • Gefahr der nichtlinearen Beziehung , Gefahr von Ausreissern Wichtige Annahmen für Fehler • Fehler $u_1, \dots, u_n$ müssen Zufalls-Stichprobe darstellen, u_i sind unabhängig voneinander • Die u_i haben alle die gleiche Standardabweichung: i.d.R. für grössere x-Werte grössere Abweichungen von der Gerade → Streudiagramm (für eine X-Variable) → Residuen-Diagramm → Fächerform (für beliebig viele X-Variablen) • Wichtig: u_i enthält Informationen über ausgelassene Variablen (OVb) • Wenn $E(u_i) = 0$ → Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ sind BLUE (Best linear unbiased estimators) Exakte Inferenz: Wenn angenommen wird, dass Fehler u_i von Normalverteilung kommen: $KI = \text{Schätzer} \pm t_{n-2, 1-\alpha/2} \cdot \text{Standardfehler}$ $(PW = P(t_{n-2} \geq z))$	P-Wert eines Chi-Quadrat Tests $\frac{PW_{alt}}{2} \rightarrow$ Vorzeichen der geschätzten Parameter mit a priori Vermutung übereinstimmt $1 - \frac{PW_{alt}}{2} \rightarrow$ Vorzeichen der geschätzten Parameter NICHT mit a priori Vermutung übereinstimmt Forscherin Aufgabe: Varianz-Minimierung des Schätzers Varianz von $\hat{\beta}_1$ minimiert, wenn $\begin{cases} 5x_1 = 0 \\ 5x_1 = 100 \rightarrow \bar{x} = (x_1 - \bar{x}) = 50 \end{cases}$ Standardfehler: $SF(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum (x_i - \bar{x})^2}} = \sqrt{\frac{s^2}{n s_x^2}}$ Only-Intercept Modell: $x_i = \beta_0 + u_i$ Interpretation β_0 : $E(x_i) = E(\beta_0 + u_i) = \beta_0 = \mu_x$ KQ-Schätzer $\hat{\beta}_0$: $\argmin \sum (x_i - \beta_0)^2 \rightarrow \frac{1}{n} \sum x_i = \bar{x}$ → Wenn $\beta_0 = 0$, dann ist auch $R^2 = 0$; Wenn $R^2 = 0$ heisst nicht, dass $\beta_0 = 0$ → Im Only-Intercept Modell ist IMMER $R^2 = 0$ Regression durch den Ursprung: $y_i = \beta_1 x_i + u_i$ • Annahme $\beta_0 = 0$ (Achsenabschnitt ist 0) • Regressionsgerade geht durch den Ursprung • Einziger unbekannter Parameter: Steigung β_1 Kleinste-Quadrate Schätzer: $\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$ und $s^2 = \frac{1}{n-1} \sum_{i=1}^n \hat{u}_i^2 = \frac{1}{n-1} \sum_{i=1}^n [y_i - \hat{\beta}_1 x_i]^2$ Erwartungswert und Varianz: $E(\hat{\beta}_1) = \beta_1$ und $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2} \rightarrow SF(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum x_i^2}}$ Konfidenzintervall: $\hat{\beta}_1 \pm z_{1-\alpha/2} \cdot \sqrt{\frac{s^2}{\sum x_i^2}}$ Teststatistik: $z = \frac{\hat{\beta}_1 - \beta_{1,0}}{SF(\hat{\beta}_1)}$ Vorteile, wenn das vereinfachte Modell stimmt ($\beta_0 = 0$): Parameter $\hat{\beta}_1$ genauer geschätzt (kleinere Varianz), Vorhersagen i.d.R. genauer (Aber für $n \geq 50$ Vorteile minimal) Nachteile, wenn das vereinfachte Modell nicht stimmt ($\beta_0 \neq 0$): Man schätzt falsches Modell (zu klein), $\hat{\beta}_1$ hat einen Bias: $E(\hat{\beta}_1) \neq \beta_1$, Vorhersagen i.d.R. ungenauer (auch für grosses n) Standard-Modell vs. Regression durch Ursprung 1. $y_i = \beta_0 + \beta_1 x_{i1} + u_i$ 2. $y_i = \gamma_1 x_{i1} + e_i$ (Regression durch Ursprung) R^2 : Das allgemeine Modell ist mind. so gut wie das vereinfachte Modell (meist besser) Varianz: Varianz im vereinfachten Modell ist stets kleiner als die Varianz im allg. Modell Unverzerrtheit: Wenn $\beta_0 = 0$, dann ist γ_1 ein unverzerrter Schätzer für β_1 Omitted Variable Bias (OVb) $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ und $y_i = \gamma_0 + \gamma_1 x_{i1} + e_i$ Formel für Verzerrung, wenn $n \rightarrow \infty$: $OVb = \beta_2 \frac{Cov(x_1, x_2)}{Var(x_1)}$? Unverzerrtheit, wenn: $y_i = \beta_1$ und/oder $\beta_2 = 0$ (γ_1 ist erwartungstreu und konsistent) und/oder $Cov(X_1, X_2) = 0$ Richtung der Verzerrung: β_2 , $Cov(X_1, X_2)$ Überschätzung, wenn OVb > 0 positive Verzerrung, $\gamma_1 > \beta_1$ Unterschätzung, wenn OVb < 0 negative Verzerrung, $\gamma_1 < \beta_1$ Omitted Variable Bias: Modellvergleich $y_i = \beta_0 + \beta_1 x_{i1} + e_i; y_i = \beta_{1,V} x_{i1} + e_i$ $R_A^2 \geq R_B^2$ → Das allgemeine Modell ist mind. so gut wie das vereinfachte Modell $Var(\hat{\beta}_{1,A}) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$ $Var(\hat{\beta}_{1,V}) = \frac{\sigma^2}{\sum (x_i - 0)^2} \rightarrow$ Varianz vereinfachtes Modell < Varianz allgemeines Modell $\hat{\beta}_0 = 0 \rightarrow \hat{\beta}_{1,V}$ unverzerrter Schätzer; $\beta_0 \neq 0 \rightarrow \hat{\beta}_{1,V}$ verzerrter Schätzer Vorhersage Vorhersage-Intervall: $\hat{y}_{neu} = \hat{\beta}_0 + \hat{\beta}_1 x_{neu}$ (1) Geschätzter Erwartungswert von $y_{neu} = \hat{\mu}_{neu}$ (2) Vorhersage der Zufallsvariablen: $y_{neu} = \mu_{neu} + u_{neu}$ Konfidenz-Intervall für $\mu_{neu} \rightarrow$ KI soll die Gerade einfangen! $KI = \hat{\mu}_{neu} \pm z_{1-\alpha/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(\hat{x}_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \cdot \sqrt{Var(\hat{\mu}_{neu})} = \sigma^2 \left(\frac{1}{n} + \frac{(\hat{x}_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)$ → In der Mitte wird Gerade besser geschätzt als am Rand. → Varianz steigt mit Abstand zwischen x_{neu} und $\bar{x}$ → Wichtig: Für $\hat{\mu}_{neu}$ nimmt man den Wert von y_{neu} Voraussetzungen: Zufallsstichprobe (Daten unabhängig voneinander, Konstante Fehler-Standardabweichung), Grosse Stichprobe ($n \geq 50$) Vorhersage-Intervall für $y_{neu} \rightarrow$ VI soll die Punkte einfangen! $VI = \hat{y}_{neu} \pm z_{1-\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(\hat{x}_{neu} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ Zusätzlich zur Variation von $\hat{\mu}_{neu}$ kommt nun noch die Variation von u_{neu} hinzu → Wichtig: Für y_{neu} nimmt man den Wert von $\hat{y}_{neu}$ → Wichtig: Exakte Inferenz IMMER mit $t_{n-2, 1-\alpha/2}$ wenn $n < 50$ Konfidenzintervalle und Vorhersage-Intervalle Konfidenzintervall für $\hat{\mu}_{neu}$: $\$fit \pm qt_{1-\alpha/2, df} \cdot \$se.fit$ Vorhersageintervall für $\hat{y}_{neu}$: $\$fit \pm qt_{1-\alpha/2, df} \cdot \sqrt{\$se.fit^2 + \$residual.scale^2}$ Wichtig: Für 90% KI/VI nimmt man $qt_{1-0.1, 2} = qt_{0.95} \rightarrow$ Tabelle 0.05
--	---	--

Schätzer-Berechnung im Software Output → Einfache Regression: P-Wert im globalen Test entspricht P-Wert für den Schätzer → t-Wert bestimmen mit 1. Freiheitsgrade und 2. Signifikanzniveau (Achtung: $1 - \alpha/2$) → Weiterhin gilt, dass: $\sqrt{F} = t $ entspricht → $\frac{\hat{\beta}_1}{SF(\hat{\beta}_1)} t \rightarrow$ Nach $\hat{\beta}_1$ auflösen! → Auch beim $q = 1$ gilt: $\sqrt{F} = t $	Normal-Quantil-Diagramm: Hat Verteilung eine Normalverteilung? Idealfall: Normal-Quantil-Diagramm für Fehler u_i erstellen → Gerade → Vorhersage-Intervall vertrauen Realfall: Residuen $\hat{u}_i$ verwenden statt unbekannten Fehler u_i (Am besten standard , Residuen) verwenden OK-Muster: Gerade – Vorhersage-Intervalle funktionieren richtig, Light-Tails – Vorhersage-Intervalle sind konservativ Nicht-OK-Muster: Heavy Tails – Vorhersage-Intervalle sind antinkonservativ, Schiefe – Vorhersage-Intervalle sind antinkonservativ Multipl. Regressionsmodell: $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i$ Vektor- oder Matrixschreibweise: $y = X\beta + e$ Verschiedene Parameter $Y = (y_1, \dots, y_n) \rightarrow n \times 1$ Vektor $X = (x_{ij})_{1 \leq i \leq n, 0 \leq j \leq k} \rightarrow n \times (k+1)$ Matrix (wobei $x_{i,0} = 1$ stets) $\beta = (\beta_0, \beta_1, \dots, \beta_k)' \rightarrow (k+1) \times 1$ Vektor $e = (e_1, \dots, e_n)' \rightarrow n \times 1$ Vektor Beau Counters: $G_i = \beta_0 + \beta_1 Q_i + \beta_2 A_i + u_i$ • Gemeinsames β_1 grösser als β_1 von einzelnen Regressionen → negative Korrelation! # Man unterschätzt den Effekt, wenn man nicht kontrolliert → Gemeinsames β_1 kleiner als β_1 von einzelnen Regressionen → positive Korrelation! # Man überschätzt den Effekt, wenn man nicht kontrolliert → OVb → Verzerrung durch Weglassen von Variablen → Wichtig: Wenn beide Variablen unkorreliert, dann gibt es kein OVb! Multipl. Modell & Gefahren • Wichtige Variablen gehören ins Modell, auch wenn Effekt eventuell gar nicht interessiert • Nur lineares Modell schätzen, wenn die Beziehung linear ist • Ausreisser → Können grossen Einfluss auf das geschätzte Modell Beurteilung, ob die Beziehung linear? • NICHT mit individuellen Streuungs-Diagrammen! • Individuelle Beziehung kann nichtlinear sein, auch wenn die Gesamt-Beziehung linear ist • Stattdessen: Residuen-Diagramm Beurteilung, ob es Ausreisser gibt? • NICHT mit individuellen Streuungs-Diagrammen! • Die individuellen Beziehungen können Ausreisser aufweisen , die in der Gesamtbeziehung keine mehr sind Schätzung einer multiplen Regression → Hyper-Ebene wählen, die den globalen Fehler minimiert → Kandidat für Hyper-Ebene gegeben durch $b = (b_0, b_1, \dots, b_k)'$ $GF(b) = \sum_{i=1}^n [y_i - (b_0 + b_1 x_{i1} + \dots + b_k x_{ik})]^2 = \ y - xb\ ^2$ Kleinste-Quadrate Schätzer für β : $\hat{\beta} = \argmin_b GF(b) = (X'X)^{-1}X'y$ Kleinste-Quadrate Schätzer für u : $\hat{u} = y - \hat{y} = y - X\hat{\beta}$ Kleinste-Quadrate Schätzer für σ^2 : $s^2 = \frac{1}{n-k-1} \ \hat{u} \ ^2 = \frac{1}{n-k-1} \sum_{i=1}^n \hat{u}_i^2$ Eigenschaften im multiplen Regressionsmodell → Normal equations implizieren: $0 = X'X\hat{\beta} - X'y = X'(y - X\hat{\beta}) = -X'\hat{u}$ → Das heisst $\hat{u}$ ist orthogonal zu jeder Spalte von X → Anders ausgedrückt: $\sum_{i=1}^n x_{ij} \hat{u}_i = 0$ für alle j Wichtige Annahmen: → I.d.R. haben Konstante im Modell (d.h. Abschnitt) → Dann ist erste Spalte von X ein Vektor von Einsen → Daraus folgt, dass: $\sum_{i=1}^n \hat{u}_i = 0$, wenn Konstante vorhanden → Wenn Konstante nicht vorhanden, dann gilt: $\sum_{i=1}^n \hat{u}_i \neq 0$ Erwartungswert und Varianz verallgemeinert Verallgemeinerung des Erwartungswertes: $E(W) = (E(W_1), \dots, E(W_n))' \in \mathbb{R}^n$ → Heisst immer noch Erwartungswert. Verallgemeinerung der Varianz: $Var(W) = \Sigma = (\sigma_{ij}) \in \mathbb{R}^{n \times n}$ $\sigma_{i,j} = Var(W_i) \rightarrow$ Diagonalelement $\sigma_{i,j} = Cov(W_i, W_j)$ für $i \neq j \rightarrow$ Oben & Links von Diagonalelement → Heisst Varianz-Kovarianz-Matrix oder nur Kovarianz-Matrix Modellvergleich Zielvariablen zu vergleichenden Modelle sind identisch JA NEIN Modelle sind gemittelt Modelle sind gleich gross JA NEIN JA NEIN Modelle sind gleich gross JA NEIN (BEDINGUNG (1): ungleicher Modell erhalten als im Intervall) F-Test R Vergleich t, R Vergleich RSS Vergleich max. RSS Vergleich
--	--

Regeln für Linear-Kombination Erwartungswert: $E(a + bW) = a + bE(W)$ Varianz: $Var(a + bW) = bVar(W)b'$ z.B. Varianz der Linearkombination $(\beta_1 - \beta_2 + 2\beta_3)$ ist $b = (0, 1, -1, 2)$ Eigenschaften der Schätzer: Schätzer β und s^2 sind erwartungstreu: $E(\hat{\beta}) = \beta$; $E(s^2) = \sigma^2$ Kovarianz-Matrix von $\hat{\beta}$: $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$ → Formel für gesamte Varianz! → Immer Diagonalelement verwenden, um $Var(\hat{\beta}_1)$ zu finden Bedingungen für Kovarianz-Matrix: Kovarianz-Matrix muss stets symmetrisch und positiv semidefinit sein $b'Sb \geq 0$ (1,1) Element ≥ 0 - Determinante $\geq 0 \rightarrow \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix} \rightarrow det = 1 - 4 < 0 \rightarrow$ Keine Kovarianz Matrix Rechnen mit der $(X'X)^{-1}$ Tabelle - Hypothesen: $H_0: Q_1 - Q_2 = 0$ und $H_0: Q_1 - Q_2 > 0$ - t-Statistik: $\frac{Q_1 - Q_2 - 0}{\sqrt{Var(Q_1) + Var(Q_2) - 2Cov(Q_1, Q_2)}}$ 3. Kovarianz-Matrix entspricht $(X'X)^{-1}$, deshalb: $Var(\hat{\beta}) = s^2(X'X)^{-1} \rightarrow Cov(Q_1, Q_2)$ muss noch mit SER^2 multipliziert werden Standardisierte Residuen Kovarianz-Matrix von $\hat{u}$: $R = X(X'X)^{-1} \rightarrow [Var(\hat{u}) = \sigma^2(I - R)]$ Obwohl $\hat{u}_i$ gleiche Standardabweichung haben, stimmt das für $\hat{u}_i$ nicht! $SA(\hat{u}_i) = \sigma\sqrt{1 - r_{ii}}$ und $SF(\hat{u}_i) = s\sqrt{1 - r_{ii}}$ Idealerweise: $\frac{\hat{u}_i}{s\sqrt{1-r_{ii}}} \rightarrow SA = 1 \rightarrow$ Man kennt aber σ nicht, deshalb muss man s nehmen Definition für das i-te standardisierte Residuum $\frac{\hat{u}_i}{s\sqrt{1-r_{ii}}}$ Erkennen der Gefahren - Gefahr: Beziehung ist nicht linear → Residuen-Diagramm → Chaos! linear - Gefahr: Ausreisser → Residuen-Diagramms → Zusätzlich Cook's Distance Cook's Distance: Mass zur Erkennung einflussreicher Beobachtungen - Kombination der Faktoren Ausreisser und Extremes X - Wie verändert sich das Modell, wenn eine Beobachtung weggelassen wird Äquivalente Formel: $R = X(X'X)^{-1}X' \rightarrow D_i = \left(\frac{\hat{u}_i}{s\sqrt{1-r_{ii}}} \right)^2 \left(\frac{1-r_{ii}}{1-r_{ii} + k+1} \right)$ → Stand. Residuum misst Ausreisser → Zweiter Faktor misst „Extremes x“: $r_{ii} = \frac{(x_i - \bar{x})^2 + n^{-1} \sum (x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$ Beurteilung der konstanten Fehler SA - Inferenz benötigt, dass die Fehler-Terme konstante Standardabweichungen haben - Im einfachen Regressions-Modell → Fächerform im Residuen-Diagramm Verfeinerte Methode nun: $\sqrt{Standard. Residuen}$ gegen $\hat{y}_i$ plotten → halbe Fächerform Standardisierte Residuen $\hat{e}_i$ ist aussergewöhnlich gross → Beobachtung ist Ausreisser r_{11} ist aussergewöhnlich klein → Miss extremes X → Wenn r_{11} klein ist, dann ist $(x_1 - \bar{x})^2$ auch klein, d.h. x_1 liegt nahe an $\bar{x}$ → Keine einflussreiche Beobachtung Inferenz: - Test von mehreren Hypothesen gleichzeitig - Subvektoren testen, d.h. $H_0: \beta_1 = \beta_2 = 0$ - Lineare Kombinationen kann getestet werden, d.h. $H_0: \beta_2 + \beta_4 = 1$ Formulierung der Hypothesen lautet: $H_0: R\beta = r$, $R: q \times (k+1)$ Matrix, $r: q \times 1$ Vektor → Man testet somit q lineare Restriktionen auf einmal Teststatistik für Hypothesentests im multiplen Regressionsmodell - Berechnung des Schätzers: $H_0: R\beta = r$ - Berechnung des Varianz: $Var(b'\beta) = \begin{pmatrix} b' & Var(\beta) & b \end{pmatrix}_{Kov-Matrix}$ Beispiel: $SF(2\beta_1 - \beta_2) \rightarrow b' = (0, 2, -1, 0)$ - Berechnung des Standardfehlers: $SF(b'\beta) = \sqrt{Var(b'\beta)}$ Hypothesentest-Transformation $H_0: \beta_2 + 80\beta_3 = 1 \rightarrow \beta_2 = 1 - 80\beta_3$ $y = \beta_0 + \beta_1x_1 + (1 - 80\beta_3) \cdot x_2 + \beta_3x_2^2 + e$ $y = \beta_0 + \beta_1x_1 + x_2 - 80\beta_3x_2 + \beta_3x_2^2 + e$ $y - x_2 = \beta_0 + \beta_1x_1 + \beta_3(x_2^2 - 80\beta_3x_2) + e$ F-Test: $F = \frac{(W_1^2 - R_1^2)/q}{(1-R_1^2)/(n-k-1)} = \frac{(SSR_0 - SSR_1)/q}{\frac{SSR_1}{DF \text{ von } H_0}}$ - Wenn Wert des F-Tests gross → Nullhypothese verwerfen Wichtig: Nur wenn gleiche Zielvariable, ansonsten $SS_{T0} \neq SS_{T1}$. - Beide Modelle müssen eine Konstante enthalten Implikationen: Je besser Gesamt-Modell relativ zum Null-Modell; desto grösser ist R_1^2 relativ zu R_0^2; Desto grösser ist SSR_0 relativ zu SSR_1; Desto grösser ist die F-Test-Statistik P-Wert & test zum Signifikanz-Niveau α: $P(F_{q,n-k-k-1} \geq F)$ Wichtig: $F_{q,n-k-k-1}$ bezieht sich IMMER auf die Alternativhypothese Test zum Signifikanz-Niveau α H_0 verwerfen, falls $PW \leq \alpha$ H_0 verwerfen, falls $F \geq F_{1-\alpha,q,n-k-k-1}$ (kritischer Wert in Tabelle) Wahl Test-Statistik Test-Statistik basierend auf R_0^2 und R_1^2 - Einfacher, da R^2 Werte direkt aus dem Software Output abgelesen werden können - Nur gültig, wenn beide Modelle die gleiche Zielvariable verwenden *β_0 im Modell Test-Statistik basierend auf SSR_0 und SSR_1 - Muhsamer, da SSR-Werte erst separat berechnet werden müssen IMMER gültig Test des Gesamt-Modells: Erklären alle Regressoren zusammen etwas über die Zielvariable? → Alle Variablen testen gegen 0: $H_0: \beta_1 = \dots = \beta_k = 0$, H_1: mindestens ein $\beta \neq 0$ Test-Statistik: $F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \rightarrow$ Zugehörige t-Wert & p-Wert im Software Output!	Vorhersage Konfidenz-Intervall: $KI = x'_{neu}\hat{\beta} \pm z_{1-\alpha/2} \times s \sqrt{x'_{neu}(X'X)^{-1}x_{neu}}$ Vorhersage-Intervall für y_{neu}: $x'_{neu}\hat{\beta} \pm z_{1-\alpha/2} \times s \sqrt{1 + x'_{neu}(X'X)^{-1}x_{neu}}$ $x'_{neu}\hat{\beta} \pm z_{1-\alpha/2} \times \sqrt{s^2 + (SF(\hat{u}_{neu}))^2}$ Voraussetzungen: Zufalls-Stichprobe, Daten unabhängig, konstante Fehler-SA: $\sigma, n \geq 50$ Vorhersage-Intervall im multiplen Regressions-Modell $\hat{y}_{neu} = x'_{neu} \cdot \hat{\beta}$ $SF(\hat{y}_{neu}) = s \sqrt{1 + x'_{neu}(X'X)^{-1}x_{neu}}$ ABER: $(X'X)^{-1}$ nicht vorhanden, d.h. eine Umformung nötig! → $SF(\hat{y}_{neu}) = \sqrt{\frac{s^2 + x'_{neu} \cdot s^2(X'X)^{-1} \cdot x_{neu}}{Kovarianz-Matrix}}$ $s^2 = \frac{RSS}{n-k-1} \rightarrow SF(\hat{y}_{neu}) = \sqrt{\frac{RSS}{n-k-1} + x'_{neu} \cdot \frac{s^2(X'X)^{-1}}{Kovarianz-Matrix} \cdot x_{neu}}$ Dummy-Variablen: $D_i = \begin{cases} 1 & \text{falls ite Person männlich} \\ 0 & \text{falls ite Person weiblich} \end{cases}$ Gemeinsame Steigung: $G_i = \beta_0 + \delta D_i + \beta_1 A_i + u_i$ Interpretation: δ ist der geschätzte Unterschied der beiden Abschnitte Gemeinsamer Abschnitt: $G_i = \beta_0 + \beta_1 A_i + \gamma(D_i \times A_i) + u_i$ Interpretation: γ ist der geschätzte Unterschied der beiden Steigungen Total verschieden: $G_i = \beta_0 + \delta D_i + \beta_1 A_i + \gamma(D_i \times A_i) + u_i$ Welches Modell nehmen: Wechsel zu Modell mit kleinstem P-Wert (einzelner Regressor und nicht P-Wert für gesamtes Modell betrachten). Auch anhand von R^2 entscheiden (Nur wenn Anzahl Freiheitsgrade gleich ist) Vergleich von Modellen mit Dummy-Variablen Gemeinsame Gerade (SSR0) → Gemeinsamer Abschnitt (SSR1) → F-Test Gemeinsamer Abschnitt (SSR0) → Total verschieden (SSR1) → F-Test → Wenn F-Wert > F-Kritischer Wert, auf SSR1-Modell umsteigen Gemeinsame Steigung → Gemeinsamer Abschnitt → Vergleich anhand F-Test nicht möglich, da nicht getestet → Wichtig: Wenn Freiheitsgrade (df) gleich, darf anhand von R^2 ein Vergleich gemacht werden → R^2 Annahmen: Konstante enthalten, gleiche Anzahl Regressoren → gleiche Anzahl df, gleiche Zielvariable Inferenz für Dummy Variablen Method 1: $G_i = \beta_0 + \delta D_i + \beta_1 A_i + u_i$ → Schätzer: $[\hat{\beta}_0 + \hat{\delta}] \rightarrow SF(\hat{\beta}_0 + \hat{\delta}) = \sqrt{Var(\hat{\beta}_0) + Var(\hat{\delta}) + 2Cov(\hat{\beta}_0, \hat{\delta})}$ Method 2: Das Modell muss zuerst umgeschrieben werden: Gemeinsame Steigung: $G_i = \beta_{0,F}(1 - D_i) + \beta_{0,M}D_i + \beta_1 A_i + u_i$ $\beta_{0,F}$: Abschnitt für die Frauen, $\beta_{0,M}$: Abschnitt für die Männer → Das Modell enthält keine globale Konstante mehr Schätzung: Gleiche Schätzer wie vorher Inferenz: Alle Standardfehler direkt im R Output Dummy Variablen bei mehreren Gruppen Bei $g > 2$ Gruppen: $g - 1$ Dummy Variablen nehmen - Immer eine Basis Gruppe: Problem: Welches Modell wählen → Lösung: Allgemeiner F-Test Nullhypothese besagt, dass alle $g - 1$ Parameter (Dummy Variablen) null sind Wichtig für F-Test: $q = \text{Anzahl Dummy Variablen}$ Konfidenzintervall für Dummy Variablen $\hat{y} = 0.5 + 1.2d + 2.0x$ Für 1: $\hat{y} = 1.7 + 2.0x$ KI: $1.7 \pm 1.96 \cdot SF(\hat{\beta}_0 + \hat{\delta}_d)$ $SF(\hat{\beta}_0 + \hat{\delta}_d) = \sqrt{Var(\hat{\beta}_0) + Var(\hat{\delta}_d) + 2 \cdot Cov(\hat{\beta}_0, \hat{\delta}_d)}$ Chow-Test: $F = \frac{(SSR_0 - SSR_{2,0})/(g-1)(k+1)}{SSR_{2,0}/(n-g \times (k+1))}$ - Multiplies Regressions-Modell und q Gruppen: H_0: Gemeinsames Modell, H_1: Total Verschieden - Gleich wie bei F-Test: SSR_0: Vom gemeinsamen Modell; SSR_k: Alle Modelle separat für Gruppen schätzt und SSRs aufaddiert Rechnen mit Chow Test SSR_0: Normales Modell SSR_1: $SSR_{1,0} + SSR_{2,0}$ → Beide Gruppen-Modelle berücksichtigen! Wichtig: Modell muss mind. 3 mal getestet werden (Alle, Vor, Nach) Nichtlineare Beziehungen Polynomiale Beziehung zwischen X und Y - Man verwendet höhere Potenzen von X als zusätzliche Regressoren. - Alle interessanten Regressoren werden vom zugrundeliegenden X generiert Quadratisches Modell: $y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i1}^2 + u_i$ Kubisches Modell: $y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i1}^2 + \beta_3x_{i1}^3 + u_i$ Interpretation des geschätzten Modells $E(f(x)) = \frac{df(x)}{dx}$ Wichtig: Effekt ist nicht mehr konstant, sondern ändert sich mit X - Wenn P-Werte klein sind der zusätzlichen Regressoren (+ R^2 erhöht) → nichtlineares Modell Wichtig: Nichtlineare Modelle können linear geschätzt ein hohes R^2 haben! Für die Beurteilung deshalb die Gefahren-Diagramme betrachten Nichtlineare Transformation von X oder Y - Die Transformation sollte theoretisch motiviert sein Ergebnisse zurücktransformieren für Interpretation in Original-Skala Rücktransformation Beispiel: $\log(Q) = 4.57 - 0.502 \cdot \log(3) = 4.018 \rightarrow \log - \text{Nachfrage}$ $Q = e^{4.018} = 55.6 \rightarrow \text{Original} - \text{Nachfrage} \rightarrow \text{Wert auf Kurve}$ Preiselastizität der Nachfrage (PEDn): $PEd_n = \frac{\% \text{Veränderung der erwarteten Nachfrage}}{\% \text{Veränderung des Preises}}$ → Im log-log Modell: die PEDn ist konstant mit Wert β_1 → D.h. wenn Preis um 1% fällt, dann steigt/sinkt die Nachfrage um $\beta_1\%$	90%-Vorhersage-Intervalle für Log-Log Modell: Für log-Nachfrage: $3.973 \pm 1.708\sqrt{0.027^2 + 0.060^2} = [3.860, 4.085]$ Für die Nachfrage: $[e^{3.860}, e^{4.085}] = [47.47, 59.44]$ Interaktion: Effekt von X_1 könnte von Wert anderer Variablen abhängen → Interaktion! $y = \beta_0 + \beta_1X_1 + \beta_2Z_1 + \beta_3(X_1 \times Z_1) + u_i$ Effekt von X_1: $dy/dX = \beta_1 + \beta_3Z_1$; Effekt von Z_1: $dy/dZ = \beta_2 + \beta_3X_1$ → Effekt einer Variablen hängt vom Wert der anderen ab Vergleich nicht generierter Modelle - Modelle nicht getestet, d.h. das eine ist nicht im anderen enthalten → kein F-Test möglich R^2/R^2 - Wenn neue Variable Modell hinzugefügt wird → Erhöhung R^2 - D.h., in einem Vergleich generierter Modelle: R^2 IMMER für das grössere Modell R^2 bevorzugt Modelle, die mehr Regressoren haben Adjustiertes $\bar{R}^2$: $\bar{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2/(n-k-1)}{\sum(y_i - \bar{y})^2/(n-1)} = 1 - (1 - R^2) \left(\frac{n-1}{n-k-1} \right)$ → erlaubt fairen Vergleich verschieden grosser Modelle! Neue Variablen im Modell Neue Variable im Modell: R^2 steigt IMMER an – $\bar{R}^2$ sowohl ansteigen als auch sinken → RSS sinkt i.d.R. mit zusätzlichen Regressoren, TSS unabhängig von Anzahl Regressoren Konstante im Modell → Wichtig: Wenn man mit R^2 oder $\bar{R}^2$ arbeitet, dann muss die Konstante immer im Modell enthalten sein R^2: Wenn von Regression durch den Ursprung auf allgemeine Regression → R^2 kann sinken Heteroskedastie Homoskedastie: Fehler-SA konstant Heteroskedastie: Fehler-SA nicht konstant sind, d.h. $[Var(u) = \sum = Diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)]$ → Diagonaleinträge in Varianz-Kovarianz-Matrix sind nun verschieden Schätzer: $E(\hat{\beta}) = \beta \rightarrow$ Auch unter Heteroskedastie unverzerrter Schätzer Varianz: $Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \rightarrow$ Benutzte Annahme der Homoskedastie → Problem Konsequenzen → Falsche Standardfehler für $\hat{\beta}_1 \rightarrow$ i.d.R. KI zu kurz, Hypothesentests liberal Liberaler Test: Wenn H_0 wahr, dann sollte sie höchstens mit W'keit α verworfen werden → Wenn Test liberal ist, dann ist diese W'keit $> \alpha$ Konsistente Inferenz HC0: $[\hat{\sigma}_i^2 = \hat{u}_i^2] \rightarrow$ Für $n \leq 250$ resultierende Inferenz oftmals liberal HC1: $[\hat{\sigma}_i^2 = \frac{\hat{u}_i^2}{k+1} \times \hat{u}_i^2] \rightarrow$ grösser als HC0 (v.a. für kleines n) - SF grösser, Inferenz konservativer, oftmals noch liberal HC2: $[\hat{\sigma}_i^2 = \frac{\hat{u}_i^2}{1-r_{ii}} = s^2 \cdot (stand. Resid.)^2] \rightarrow \frac{1}{n} \leq r_{ii} \leq 1 \rightarrow$ Grösser als HC0 - Unterschied zu HC1: Individuelle Faktoren, anstatt $(n - k - 1)$ - HC2 gibt einen höheren Wert als HC0 und i.d.R. auch als HC1 → Inferenz konservativer HC3: $[\hat{\sigma}_i^2 = \frac{\hat{u}_i^2}{(1-r_{ii}^2)}] \rightarrow$ HC3 wird grösser als HC2, da $(1 - r_{ii})^2 < (1 - r_{ii})$ - HC3 funktioniert i.d.R. sehr gut, auch unter Homoskedastie! Breusch-Pagan-Test: $[\hat{u}_i^2 = \theta_0 + \theta_1x_{i1} + \dots + \theta_kx_{ik} + error]$ H_0: Homoskedastie H_1: Heteroskedastie $H_0: \beta_1 = \dots = \theta_k = 0 \rightarrow$ Test ob Original-Regressoren 0 sind → Fehler-Varianz 0 - Wenn wir H_0 nicht verwerfen → OLS für Inferenz - Wenn wir H_0 verwerfen zugunsten von $H_1 \rightarrow$ HC-Inferenz (HC3) - BP-Test → nimmt nur Original-Regressoren, d.h. wenn Zusammenhang nichtlinear (z.B. quadratisch) ist, wird Test dies nicht feststellen, sogar für grosses n Empfehlungen von Long & Ervin - Für kleines und mittelgrosses n (Faustregel: $n \leq 250$): - Schwierig, Heteroskedastie mit BP-Tests festzustellen - Verdacht auf Heteroskedastie → Immer HC3 verwenden Für grosses n (Faustregel: $n > 250$): - Man kann Heteroskedastie mit BP-Tests feststellen Liegt Heteroskedastie vor → HC3 anwenden, aber auch OLS (Standard-Inferenz) anwenden Allgemeiner F-Test: Allgemeiner F-Test für $H_0: R\beta = r$ unter Heteroskedastie $F = \frac{SSR_0 - SSR_1/q}{SSR_1/(n-k-1)} = \frac{(R\hat{\beta} - r)'[R(s^2(X'X)^{-1}R')^{-1}R(\hat{\beta} - r)]}{q}$ <i>Funktioniert nicht unter Heteroskedastie</i> <i>Wie nahe ist der Vektor bei 0</i> → $\sigma^2(X'X)^{-1}$ durch HC3-Schätzer $Var(\hat{\beta})$ ersetzen Multikollinearität: Kleinste-Quadrate Schätzer für $\hat{\beta}: \hat{\beta} = (X'X)^{-1}X'Y$ Implizite Annahmen: $(X'X)^{-1}$ existiert; $X'X$ hat vollen Rang $k+1$; X vollen Spaltenrang $k+1$ → Keine Spalte kann mit X als Linearkombination der restlichen k Spalten dargestellt wird Perfekte Multikollinearität: Eine Spalte von X kann als Linearkombination der restlichen k Spalten dargestellt werden $x_{i1} = \gamma_0 + \gamma_1x_{i2} + \dots + \gamma_{j-1}x_{ij} + \gamma_{j+1}x_{i,j+1} + \dots + \gamma_kx_{ik}$ $(X'X)^{-1}$ existiert nicht; $X'X$ hat vollen Rang $< k+1$, d.h. $X'X$ ist singulär → Nicht invertierbar; X hat vollen Spaltenrang $< k+1$ Konsequenzen: Der OLS-Schätzer $\hat{\beta}$ existiert nicht → → Lösung: Überflüssige Variable entfernen ohne Verlust jeglicher Information Imperfekte Multikollinearität: Eine Spalte von X kann „annähernd“ als Linear-Kombination der restlichen k Spalten dargestellt werden $x_{i1} \approx \gamma_0 + \gamma_1x_{i2} + \dots + \gamma_{j-1}x_{ij} + \gamma_{j+1}x_{i,j+1} + \dots + \gamma_kx_{ik}$ $(X'X)^{-1}$ existiert, enthält aber einige sehr grosse Einträge: $X'X$ hat noch Rang $= k+1$, aber $X'X$ ist „annähernd“ singulär; X hat noch Spaltenrang $= k+1$ Konsequenzen: Der OLS-Schätzer $\hat{\beta}$ existiert → Jedoch: $Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \rightarrow$ Matrix existiert, aber sehr grosse Zahlen wenn annähernd singulär → Für gewisse Elemente $\hat{\beta}_j$ und für gewisse Linearkombinationen $a'\hat{\beta}$ ist die Varianz sehr gross Erläuterung anhand Homoskedastie: $Var(\hat{\beta}_1) = \frac{\sigma^2}{(1-r)^2(x_{i2} - R_2)^2}$ - Je grösser Stichprobenkorrelation r zwischen x_1 und x_2, umso ungenauer wird $\hat{\beta}_1$ geschätzt - Idealfall wäre hier $r = 0$, denn dann resultiert der kleinstmögliche Wert	Typisch für Multikollinearität - Einzelne Variablen sind individuell nicht signifikant, aber zusammen als Gruppe Grund: Die Variablen enthalten ähnliche Informationen über Y, und somit sind die individuellen geschätzten c.p. Effekte (kontrolliert für andere Regressoren) nicht mehr präzise Feststellen von Multikollinearität Alle paarweisen Korrelationen: Einfach und informativ; Findet keine komplexere Linearkombinationen, die M-K verursachen Alle „Eine-gegen-den-Rest“ Regressoren: $x_{i1} = \gamma_0 + \gamma_1x_{i2} + \dots + \gamma_{j-1}x_{ij} + \gamma_{j+1}x_{i,j+1} + \dots + \gamma_kx_{ik}$ - Jede Regression ergibt ein R_j^2 (Wie viel der Variable kann durch die Variablen erklärt werden?) Wichtig: Zielvariable nicht ins Modell nehmen Deren Maximum kann als ein Mass für M-K betrachtet werden Inverse der Konditionsnummern κ von $X'X$ Dieses Mass liegt in [0,1] und besagt, wie nahe an singulär (wie nahe an der 0, d.h. nicht invertierbar) eine quadratische Matrix ist - Werte unterhalb von 0.01 gelten auf jeden Fall als kritisch 0 bedeutet perfekte Multikollinearität Variance Inflation Factor: $VIF_j = \frac{1}{1-R_j^2}$ „Eine-gegen-den-Rest“ Regression: $x_{i1} = \gamma_0 + \gamma_1x_{i2} + \dots + \gamma_{j-1}x_{ij} + \gamma_{j+1}x_{i,j+1} + \dots + \gamma_kx_{ik}$ → Diese Regression ergibt ein R_j^2 Interpretation: VIF_j gibt an, um wieviel Varianz von $\hat{\beta}_j$ grösser ist im Vergleich zum Idealfall, wenn X_j komplett linear unabhängig von restlichen Regressoren - Bester Fall daher: $VIF_j = 1$ (nächlich falls $R_j^2 = 0$) Faustregel: Kritisch, wenn maximale VIF_j über 10 liegt Polynomiale Beziehungen: $y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i1}^2 + \dots + u_i$ Problem: Wenn alle $x_i > 0 \rightarrow$ Stichprobenkorrelation zwischen x und x^2 kann gross sein - Evt. $\hat{\beta}_1$ und/oder $\hat{\beta}_2$ nicht signifikant, obwohl theoretisch zu erwarten - Evtl. entfernt man zu Unrecht eine der Variablen Lösung: Um individuelle Signifikanz zu beurteilen → „orthogonalis“ Polynome verwenden (d.h. voneinander linear unabhängig) $VIF_j = 0$ Quick-and-dirty Ansatz: $(x_i - \bar{x})^2$ anstelle von x_i^2 - Obwohl Koeffizienten teilweise ändern → geschätzte Modell äquivalent im Sinne der $\hat{y}_i$. - Alle ergeben die gleiche geschätzte Hyperebene (d.h. die gleichen $\hat{y}_i$ und R^2, etc.) Änderung: Standardfehler und t-Werte! Wichtig: Die t-Statistik kann sich signifikant ändern, so dass diese nun signifikant verschieden von 0 ist! Berechnung von Elastizitäten Lin-Log Modell: $[\frac{\partial y}{\partial x}] \rightarrow$ Wichtig: $y = \beta_0 + \beta_1 \cdot \log(x)$ Lin-Log Modell: $[\frac{\partial y}{\partial x}] \rightarrow$ Wichtig: $y = \beta_0 + \beta_1 \cdot \log(x)$ Log-Log Modell: $e^{(\beta_0 + \beta_1x_1)} \cdot \beta_1 \cdot \frac{x_1}{y} \rightarrow$ Wichtig: $y = e^{\beta_0 + \beta_1x_1}$ Log-Log Modell: $[\frac{\partial y}{\partial x}]$ Interpretation der Elastizitäten: - Wenn x um 1 Einheit (bspw. 1000S) steigt, dann steigt y um β_1 Einheiten - Wenn x um 1% steigt, dann steigt y um $\beta_1 \cdot 0.01$ Einheiten - Wenn x um 1 Einheit (bspw. 1000S) steigt, dann steigt y um $\beta_1 \cdot 1000\%$ ($\beta_1 \cdot 100\%$) - Wenn x um 1 Einheit (bspw. 1000S) steigt, dann steigt y um $\beta_1\%$ Interpretation Software Output <pre> Call: lm(formula = "salary ~ qs + age", data = data) Residuals: Min 1Q Median 3Q Max -196.569 -10.711 -2.154 26.795 132.181 Coefficients: (Intercept) 62.51 104.206 -3.479 0.00172 ** qs 11.889 3.276 3.631 0.00116 ** age 6.411 1.074 0.968 0.30e+06 *** Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 68.79 on 27 degrees of freedom Multiple R-Squared: 0.5711, Adjusted R-Squared: 0.5393 F-statistic: 17.97 on 2 and 27 DF, p-value: 1.09e-05 </pre> $SS_{\text{Mod}} = 27^*(69.79) = 1881.63$ $R^2 = 0.5711$ $\text{adj. } R^2 = 0.5393$ $95\% \text{-KI für qs} = [11.889 \pm 1.96 * 3.276]$ (age (Modell ist konfident (p-Werte))) $(n-k-1) = 27$ $95\% \text{-KI für age} = [6.411 \pm 1.96 * 1.074]$ <pre> > fit.8.3 = lm(revems ~ price + ads + I(ads^2)) > predict(fit.8.3, data.frame(revems = 0, price = 2, ads = 40), se = T) [1] 181.7009 [1] 181.7009 See fit [1] 11.89446 ads [1] 74 95%-KI für mu = [181.7 + 1.993 * 11.89] 95%-VI für y = [181.7 - 1.993 * 6.92 + 11.89] * 0.5 Residual scale [1] 6.91807 > qt(0.975, 74) [1] 1.99245 </pre>
--	---	--	---