Kapitel 4:

Multiple lineare Regression: Inferenz

Moodle

Lehrbuch

Das klären wir in diesem Kapitel:

Stichprobenverteilung der OLS-Schätzer

Hypothesentests eines einzelnen Parameters: der *t*-Test

Konfidenzintervalle

Hypothesentest über eine Linearkombination der Parameter

Der F-Test: mehrere Linearkombinationen der Parameter

Appendix

Zusammenfassung Kapitel 3 – Annahmen:

MLR 1 **y** linear in den Parametern $\boldsymbol{\beta}$

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$$

MLR 3 lineare Unabhängigkeit der Regressoren

$$rk(X) = k + 1,\ n > k$$

MLR 4 strikte Exogenität der Regressoren

$$E[\mathbf{u}|X] = E[\mathbf{u}] = \mathbf{0}$$

MLR 5 serielle Unkorreliertheit und Homoskedastizität

$$\Sigma(\mathbf{u}) = \sigma^2 I_n \qquad \text{(OLS)}$$

Zusammenfassung Kapitel 3 – Resultate:

MLR 1, **MLR 3**, **MLR 4**:

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y},\ E[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta},\ \Sigma(\hat{\boldsymbol{\beta}}|X) = \sigma^2(X'X)^{-1}$$

MLR 5:

Behauptung: $\hat{\boldsymbol{\beta}}$ BLUE.

$$\hat{\sigma}^2 = \frac{1}{n-k-1}\mathbf{y}'\left[I - X(X'X)^{-1}X'\right]\mathbf{y},\ E[\hat{\sigma}^2] = \sigma^2$$

Wir benutzen nun in Kapitel 4 den Standardfehler $se(\hat{\boldsymbol{\beta}}|X)$ von $\hat{\boldsymbol{\beta}}$,

$$se(\hat{\boldsymbol{\beta}}|X) = \sqrt{\underbrace{\widehat{\Sigma}(\hat{\boldsymbol{\beta}}|X)}_{k+1 \times k+1}} = \hat{\sigma}\sqrt{(X'X)^{-1}}$$

um statistische Tests auf $\hat{\boldsymbol{\beta}}$ durchzuführen.

Stichprobenverteilung der OLS-Schätzer

Statistische Inferenz im Regressionsmodell

Wir stellen Hypothesen an die Parameter des Regressionsmodells auf und verwenden die OLS-Schätzer, um diese Hypothesen zu überprüfen.

Hierzu konstruieren wir Konfidenzintervalle. Dies sind zufällige Bereiche, die, basierend auf den OLS-Schätzern, mit gewissen Wahrscheinlichkeiten die festen aber unbekannten Parameter umschließen.

Dies ist aber nur möglich, wenn wir weitergehende Verteilungsannahmen an die Zufallsvariablen des Modells treffen.

Stichprobenverteilung von $\hat{\beta}$

Die Standardfehler können wir interpretieren, wenn wir über Informationen der Wahrscheinlichkeitsverteilung des Schätzers $\hat{\beta}$ verfügen.

Das versetzt uns dann in die Lage, Hypothesentests durchzuführen.

Es gibt hier zwei Möglichkeiten:

- Man trifft eine Annahme über die Verteilung der Fehlerterme, und leitet daraus die Verteilung des Schätzers ab.
- Man vertraut auf Resultate über die asymptotische Verteilung des Schätzers.

Die zweite Möglichkeit verfolgen wir weiter unten. Sie setzt eine große Stichprobe voraus.

In kleinen Stichproben bleibt uns nur die erste Möglichkeit, die wir hier betrachten.

Stichprobenverteilung von $\hat{\beta}$

Annahme MLR 6

"ist verteilt"

Die Fehler u_i sind unabhängig und identisch normalverteilt mit Erwartungswert 0 und Varianz σ^2:

$E[u_i]$ $\quad$ $var(u_i)$

$$u_i|X \sim \mathcal{N}(0, \sigma^2)$$

normal verteilt

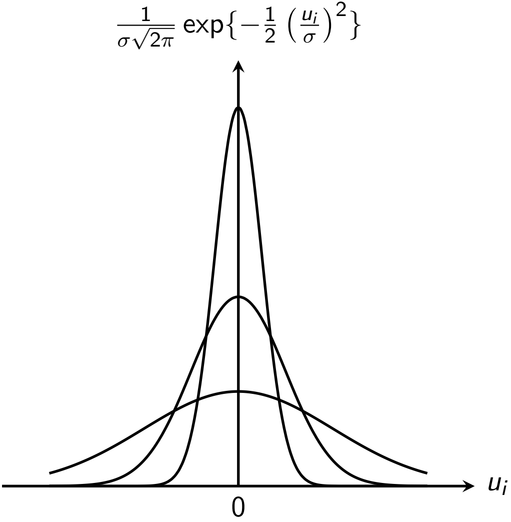
bzw.

$$\mathbf{u}|X \sim \mathcal{N}\left(\underset{\substack{\uparrow \\ \text{Vektor} \\ n\times 1}}{\mathbf{0}}, \underbrace{\sigma^2 I_n}_{\substack{\Sigma(u) \\ n\times n}}\right)$$

MLR 6 impliziert direkt:

- **MLR 4**: $E[\mathbf{u}|X] = E[\mathbf{u}] = \mathbf{0}$
  und
- **MLR 5**: $\Sigma(\mathbf{u}) = \sigma^2 I$

Dichtefunktion von $u_i \sim \mathcal{N}(0, \sigma^2)$



Diskussion der Normalverteilungsannahme

Der Fehlerterm setzt sich aus vielen unbeobachtbaren Faktoren zusammen.

Zentraler Grenzwertsatz:

Die Verteilung der Summe unabhängiger & identisch verteilter Zufallsvariablen konvergiert gegen die Normalverteilung.

Probleme:

- Ist die Anzahl verschiedener Faktoren groß genug?
- Sind die individuellen Faktoren zu heterogen?
- Sind die individuellen Faktoren unabhängig voneinander?

In manchen Fällen ist die Normalverteilung fraglich oder sogar per Definition unmöglich: Löhne – nicht negativ;
Verhaftungen – kleine, natürliche Zahlen; Arbeitslosigkeit – 0,1

Verteilung von $\hat{\beta}$ unter MLR 1 , MLR 3 und MLR 6

$y = X\beta + u$

Mit **MLR 1** gilt $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$ und

mit **MLR 3** gilt $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y} = \underline{\boldsymbol{\beta} + (X'X)^{-1}X'\mathbf{u}}$.

Da $\hat{\boldsymbol{\beta}}$ linear in $\mathbf{u}$ ist, ist mit **MLR 6** auch $\hat{\boldsymbol{\beta}}$ normalverteilt. (Linearkombinationen normalverteilter Variablen sind normalverteilt.)

Es gilt also

$E[\hat{\beta}]$ $\quad V(\hat{\beta})$

$$\hat{\boldsymbol{\beta}}|X \sim \underbrace{\mathcal{N}}_{\text{MLR 6}}\left(\boldsymbol{\beta}, \sigma^2(X'X)^{-1}\right)$$

Verteilung von $\hat{\beta}_j$

Wegen

$$\hat{\boldsymbol{\beta}}|X \sim \mathcal{N}\left(\boldsymbol{\beta}, \overbrace{\sigma^2(X'X)^{-1}}^{\text{Varianz – Kovarianzmatrix von dem } \underline{\text{Vektor}}\ \hat{\beta}}\right)$$

gilt für jeden einzelnen Schätzer

$$\hat{\beta}_j|X \sim \mathcal{N}\left(\beta_j, \overbrace{\sigma^2(X'X)^{-1}_{(j,j)}}^{\text{j. Element auf der Diagonale von } \Sigma(\hat{\beta}) = Var(\hat{\beta}_j)}\right) \text{, wobei}$$

$$\sigma^2(X'X)^{-1}_{(j,j)} = \sigma^2 \frac{1}{n} \frac{1}{\overline{x_j x_j} - \bar{x}_j^2} \frac{1}{1 - R_j^2} \quad \text{VIF}$$

Bemerkung:

Im Gegensatz zu **u**, welches mit **MLR 6** unabhängig normalverteilt ist, sind zwei einzelne Schätzer $\hat{\beta}_j$ und $\hat{\beta}_s$ nicht unabhängig. Da die Matrix $(X'X)^{-1}$ in den Einträgen, welche nicht auf der Diagonalen liegen, allgemein ungleich null ist, sind $\hat{\beta}_j$ und $\hat{\beta}_s$ in der Regel korreliert.

Standardisierung von $\hat{\beta}_j$

$\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}$ ist als Linearkombination von $\hat{\beta}_j$ normalverteilt.

Mit

$$E\left[\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}\middle| X\right] = \frac{E[\hat{\beta}_j|X] - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}} = 0$$

Konstante

und

$(\ )^2$

$$var\left[\frac{\hat{\beta}_j - \cancel{E[\hat{\beta}_j|X]}}{\sqrt{var(\hat{\beta}_j|X)}}\middle| X\right] = \frac{var(\hat{\beta}_j|X)}{var(\hat{\beta}_j|X)} = 1$$

ist $\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}$ also standard normal verteilt:

$$\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}} \sim \mathcal{N}(0,1)$$

Hypothesentests eines einzelnen Parameters: der t-Test

Hypothesentests eines einzelnen Parameters: der t-Test

Bei einer gegebenen Stichprobe ist die wahre Varianz σ^2 – und damit auch $var(\hat{\beta}_j)$ – unbekannt und muss geschätzt werden.

Wir ersetzen nun die wahre Varianz des Schätzers $\hat{\beta}_j$,

$$var(\hat{\beta}_j|X) = \sigma^2 / n(\overline{x_j x_j} - \bar{x}_j^2)(1 - R_j^2)$$

durch die geschätzte Varianz

$$\widehat{var}(\hat{\beta}_j|X) = \hat{\sigma}^2 / n(\overline{x_j x_j} - \bar{x}_j^2)(1 - R_j^2)$$

und definieren die Test-Statistik

$$t_j = \frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$$

Wie ist diese Zufallsvariable verteilt?

Ausflug Wahrscheinlichkeitsrechung

In **MLR 6** haben wir bereits die **Normalverteilung** angenommen.

Aus dieser Verteilung lassen sich weitere Verteilungen erzeugen:

- χ^2-Verteilung
- t-Verteilung
- F-Verteilung

Wir definieren die χ^2- und t-Verteilungen auf den folgenden Folien.
Die F-Verteilung folgt später.

Die χ^2-Verteilung

Seien x_i, $i = 1, \ldots, n$ unabhängig standardnormalverteilte Zufallsvariablen:

$$x_i \overset{iid}{\sim} \mathcal{N}(0, 1)$$

Eine Zufallsvariable z heißt dann χ^2_n-verteilt mit n Freiheitsgraden, falls:

$$z = \sum_{i=1}^{n} x_i^2 = \mathbf{x}'\mathbf{x}$$

$$Var(\tilde{u}_i) = Var\left(\frac{u_i}{\sigma}\right) = \frac{1}{\sigma^2} Var(u_i) = \frac{\sigma^2}{\sigma^2} = 1$$

Beispiel MLR 6 $\quad u_i \overset{iid}{\sim} N(0, \sigma^2) \Rightarrow \tilde{u}_i = \frac{u_i}{\sigma} \overset{iid}{\sim} N(0, 1)$

$$\Rightarrow \sum_{i=1}^{n} \tilde{u}_i^2 = \sum_{i=1}^{n} \left(\frac{u_i}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} u_i^2 \sim \chi^2_n$$

Die t-Verteilung

Seien $x \sim \mathcal{N}(0,1)$ und $y \sim \chi_n^2$ und x und y unabhängig.

Die Zufallsvariable z heißt t-verteilt mit n Freiheitsgraden, falls

$$z = \frac{x}{\sqrt{y/n}}$$

- Für $n \to \infty$ konvergiert die t_n-Verteilung gegen die $\mathcal{N}(0,1)$-Verteilung.
- Für hinreichend großes n kann man statt den Quantilen der t_n-Verteilung im Wesentlichen auch einfach die Quantile der $\mathcal{N}(0,1)$-Verteilung verwenden.

Verteilung der Test-Statistik t_j

$$t_j = \frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$$

Verteilung der Test-Statistik t_j

$$t_j = \frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$$

Kann konkret ausgerechnet werden, falls y & X beobachtet.

Wir multiplizieren den Nenner mit:

$$1 = \sqrt{\left(\frac{\sigma\sqrt{n-k-1}}{\sigma}\right)^2 /(n-k-1)}$$

Mit

$$\widehat{var}(\hat{\beta}_j|X) = \hat{\sigma}^2/n(\overline{x_j x_j} - \bar{x}_j^2)(1-R_j^2)$$

gilt dann:

$$t_j = \frac{\overbrace{\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}}^{\sim \mathcal{N}(0,1)}}{\sqrt{\underbrace{\left(\frac{\hat{\sigma}\sqrt{n-k-1}}{\sigma}\right)^2}_{\sim \chi^2_{n-k-1}} /(n-k-1)}}$$

unbekannt

Kann nicht berechnet werden.

unbekannt

Verteilung der Test-Statistik t_j

$$t_j = \frac{\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}}{\sqrt{\left(\hat{\sigma}\sqrt{n-k-1}/\sigma\right)^2/(n-k-1)}}$$

Wir wissen bereits:

$$\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}} \sim \mathcal{N}(0,1)$$

Appendix A zeigt, dass der Term aus dem Nenner,

$$\hat{\sigma}^2(n-k-1)/\sigma^2$$

χ^2_{n-k-1}-verteilt ist.

Appendix B zeigt, dass dieser Term unabhängig vom Zähler ist.

Hypothesentests eines einzelnen Parameters: der t-Test

Theorem

Unter den Annahmen **MLR 1**, **MLR 3** und **MLR 6** ist die Test-Statistik

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}} \sim t_{n-k-1}$$

t-verteilt mit $n - k - 1$ Freiheitsgraden.

Hypothesentests eines einzelnen Parameters: der t-Test

Da $\frac{\widehat{\beta}_j - \beta_j}{se(\widehat{\beta}_j)}$ t-verteilt ist, können wir einen **Hypothesentest** – den t-Test – konstruieren.

Nehmen wir an, wir hätten die Theorie, dass der Parameter β_j gleich dem Wert γ sein sollte.

Alternativ

→ Nullhypothese:

$$H_0 : \beta_j = \gamma$$

$H_0 : \beta_j \geq \gamma$ $H_0 : \beta_j \leq \gamma$

Sind die Daten „kompatibel" mit H_0?

Wir verwerfen H_0, wenn die Teststatistik Werte annimmt, die unter H_0 „unwahrscheinlich" sind.

Unter H_0 folgt: $E[\hat{\beta}_j] = \gamma$

Hypothesentests eines einzelnen Parameters: der t-Test

Signifikanzniveau α:

$\boldsymbol{\alpha}$ ist die Wahrscheinlichkeit, dass der Test die Hypothese H_0 ablehnt, obwohl sie wahr ist (Fehler 1. Art)

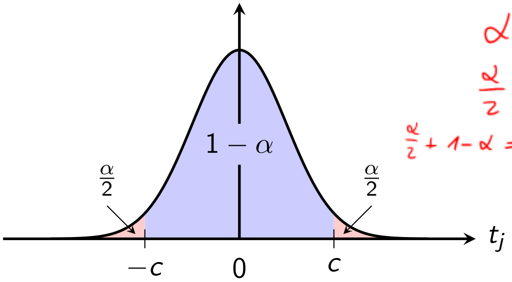
Gebräuchlich sind 10%, 5% oder 1% (also $\alpha = 0.1,\ 0.05,\ 0.01$)

Die Gegenwahrscheinlichkeit $1 - \boldsymbol{\alpha}$ nennen wir später **Konfidenzniveau**.

Zweiseitiger Test $H_0 : \beta_j = \gamma,\ H_1 : \beta_j \neq \gamma$

Teststatistik: $$t_j = \frac{\hat{\beta}_j - \gamma}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$$

Dichte der t-Verteilung unter H_0



$\alpha = 5\%$

$\frac{\alpha}{2} = 2{,}5\%$

$\frac{\alpha}{2} + 1 - \alpha = 1 - \frac{\alpha}{2} = 97{,}5\%$

Wie bestimmen wir den kritischen Wert c?

Tabelle $\quad \mathbb{P}(t_j \leq c) = \frac{\alpha}{2} \quad \Rightarrow \quad c = c_{t_{n-k-1}, 1-\frac{\alpha}{2}}$

Zweiseitiger Test $H_0 : \beta_j = \gamma,\ H_1 : \beta_j \neq \gamma$

Bestimme c, sodass (falls H_0 wahr) $prob(-c \leq t_j \leq c) = 1 - \alpha$.

Lehne H_0 ab, falls $t_j < -c$ oder $t_j > c$.

Wir nennen die Zahl c den **kritischen Wert**.

Der kritische Wert c hängt von der t-Verteilung bei $n - k - 1$ Freiheitsgraden und dem Signifikanzniveau α ab.

Da die t-Verteilung symmetrisch ist, wird $c_{t_{n-k-1},1-\frac{\alpha}{2}}$ implizit durch die Gleichung bestimmt:

$$F_{t_{n-k-1}}(t \leq c_{t_{n-k-1},1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$$

Die Werte von F_t werden tabellarisch in Handbüchern für Statistik und Ökonometrie aufgeführt.

Beispiel: Bestimmungsfaktoren von College GPA

Datensatz: GPA1.xls, $n = 141$ $\quad n-k-1 = 137$

$$\widehat{colGPA} = \underset{(0{,}33)}{1,390} + \underset{(0{,}094)}{0,412}hsGPA + \underset{(0{,}011)}{0,015}ACT - \underset{(0{,}026)}{0,083}skipped$$

skipped: verpasste Kurse pro Woche

In Klammern jeweils der Standardfehler: $\sqrt{\widehat{var}(\hat{\beta}_j|X)}$

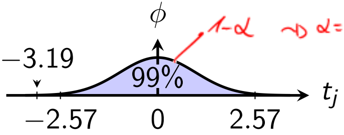
Fragestellung: trägt – falls für *hsGPA* und *ACT* kontrolliert wird – *skipped* zur Erklärung der Variation von *colGPA* bei?

Ist der mit *skipped* assoziierte Parameter signifikant von null verschieden?

$$t_j = \frac{\hat{\beta}_j - \gamma}{se(\hat{\beta}_j)}$$

$$t_{skipped} = \frac{-0{,}083-0}{0{,}026} = -3,19$$

$1-\alpha \rightarrow \alpha = 1\%$



$t_{137,\,0{,}995} = 2,626$

Eine nützliche Faustregel für zweiseitige t-Tests

Teststatistik für H_0 $\beta_j = 0$: $t_j = \frac{\hat{\beta}_j}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$

Kritischer Wert für große Stichproben bei $\alpha = 5\%$:
$c_{t_\infty,0.975} = 1.9645$

Grobe Faustregel:

Der Parameter β_j ist für große Stichproben signifikant von null verschieden (bei $\alpha = 5\%$), falls $|\hat{\beta}_j|$ mindestens doppelt so groß ist, wie $se(\hat{\beta}_j) = \sqrt{\widehat{var}(\hat{\beta}_j|X)}$.

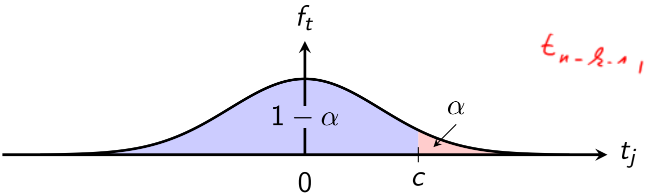
Einseitiger t-Test

Einseitige Nullhypothese

$$H_0: \ \beta_j \leq \gamma$$

Gegenhypothese $H_1: \ \beta_j > \gamma$

Lehne H_0 ab, wenn Test-Statistik $t_j = \frac{\hat{\beta}_j - \gamma}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}$ „zu groß“:



$t_{n-k-1,\,1-\alpha}$

Wähle kritischen Wert c so, dass $prob(t \leq c) = 1 - \alpha$.

Beispiel: Die Lohngleichung

wage1.xls, $n = 526$

Induziert eine höhere Arbeitserfahrung (unter Berücksichtigung für Bildung und Arbeitsalter) einen höheren Lohn?

$$\widehat{log}(wage) = \underset{(0{,}104)}{0,284} + \underset{(0{,}007)}{0,\overset{0}{0}92}educ + \underset{(0{,}0017)}{0.0041}exper + \underset{(0{,}003)}{0,022}tenure$$

$H_0 : \beta_{exper} \leq 0$, $H_1 : \beta_{exper} > 0$.

t-Statistik: $\frac{0{,}0041 - 0}{0{,}0017} \approx 2,41$

Kritische Werte für $n - k - 1 = 522$ Freiheitsgrade:

- $\alpha = 1\%$: $c_{\cancel{0,01}} = 2,3335$ $\quad c_{t_{522},0{,}99}$
- $\alpha = 5\%$: $c_{\cancel{0,05}} = 1,6478$ $\quad c_{t_{522},0.95}$
- $\alpha = 10\%$: $c_{\cancel{0,10}} = 1,2832$ $\quad c_{t_{522},0.9}$

Wir lehnen H_0 ab; wir verwerfen die Behauptung, dass eine höhere Arbeitserfahrung den Lohn reduziert, sogar für $\alpha = 1\%$.

p-Werte für t-Tests

Je kleiner das Signifikanzniveau α, desto unwahrscheinlicher wird H_0 verworfen – gegeben, dass sie wahr ist.

Der p-Wert oder das marginale Signifikanzniveau einer Teststatistik ist das kleinste Signifikanzniveau α_{min} zu dem H_0 bei einer gegebenen Stichprobe verworfen wird.

Betrachtet sei ein zweiseitiger Test. Gesucht ist die Wahrscheinlichkeit, dass H_0 $\beta_j = \gamma$ verworfen wird, obwohl sie wahr ist. Sei t^* der Wert der Test-Statistik, welcher aus den vorliegenden Daten mit $n - k - 1$ Freiheitsgraden berechnet wurde, zum Beispiel $t^* = 1.85$. Es bezeichne t_j die Zufallsvariable, welche t_{n-k-1}-verteilt ist. Dann ist der p-Wert gegeben durch

$$prob(t_j < -1.85) + prob(t_j > 1.85) = 0.0359 + 0.0359 = 0,0718$$

"p-Wert"

In diesem Fall verwerfen wir H_0 bei $\alpha = 10\%$, aber nicht bei $\alpha = 5\%$.

Konfidenzintervalle

Konfidenzintervalle

Eine Schätzung $\hat{\beta}_j$ beinhaltet keine Information über die Unsicherheit der Schätzung des Parameters β_j.

Welches Intervall beinhaltet den wahren, unbekannten Parameter β_j mit einer vorgegebenen Wahrscheinlichkeit?

Es bezeichne das Intervall $[V_u, V_o]$ ein Konfidenzintervall für den Parameter β_j zum Niveau $1 - \alpha$, falls

$$prob(V_u \leq \beta_j \leq V_o) = 1 - \alpha$$

Wichtig:

- Der Parameter β_j ist nicht stochastisch!
- Die Intervallgrenzen V_u und V_o hängen von der Stichprobe ab und sind zufällig.

Konfidenzintervalle

Sei $c := c_{t_{n-k-1},1-\alpha/2}$ der kritische Wert, sodass $H_0\ \beta_j = \gamma$ verworfen wird, falls $t_j = \frac{\hat{\beta}_j - \gamma}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}} < -c$ oder falls $t_j > c$.

Wir verwerfen $H_0\ \beta_j = \gamma$ also, falls

$$\gamma < \hat{\beta}_j - c \cdot \sqrt{\widehat{var}(\hat{\beta}_j|X)} \text{ oder } \gamma > \hat{\beta}_j + c \cdot \sqrt{\widehat{var}(\hat{\beta}_j|X)}$$

Das Konfidenzintervall für β_j zum Niveau $1 - \alpha$ ist demnach:

$$[V_u, V_o] = \left[\hat{\beta}_j - c \cdot \sqrt{\widehat{var}(\hat{\beta}_j|X)}, \hat{\beta}_j + c \cdot \sqrt{\widehat{var}(\hat{\beta}_j|X)}\right]$$

mit $c = c_{t_{n-k-1},1-\alpha/2}$.

Mit Wahrscheinlichkeit $1 - \alpha$ umschließt das Intervall $[V_u, V_o]$ den unbekannten Parameter β_j.

Beispiel: Ausgaben für Forschung und Entwicklung

rdchem.xls $n = 32$

$$\widehat{\log(rd)} = \underset{(0{,}47)}{-4,38} + \underset{(0{,}06)}{1,084} \log(sales) + \underset{(0{,}0128)}{0,0217} profmarg$$

Konstruktion von 95%-Konfidenzintervallen:

Freiheitsgrade: $n - k - 1 = 32 - 2 - 1 = 29$

Kritischer Wert $c_{t_{29};0{,}975} = 2,045$

Konfidenzintervall für $\beta_{\log(sales)}$:

$$[1,084 - 2,045 \cdot 0,06; 1,084 + 2,045 \cdot 0,06] = [0,9613; 1,2067]$$

Konfidenzintervall für $\beta_{\log(profmarg)}$:

$$[0,0217 - 2,045 \cdot 0,0128; 0,0217 + 2,045 \cdot 0,0128] = [-0,0045; 0,0479]$$

Für alle Zahlen $\gamma \in [V_u; V_o]$ gilt:

Wir können H_0 $\beta_j = \gamma$ nicht zum Niveau $\alpha = 5\%$ verwerfen.

Konfidenzintervalle für typische Signifikanzniveaus

$$prob\left(\hat{\beta}_j - c_{0.01}\cdot \widehat{var}(\hat{\beta}_j|X) \leq \beta_j \leq \hat{\beta}_j + c_{0.01}\cdot \widehat{var}(\hat{\beta}_j|X)\right) = 99\%$$

$$prob\left(\hat{\beta}_j - c_{0.05}\cdot \widehat{var}(\hat{\beta}_j|X) \leq \beta_j \leq \hat{\beta}_j + c_{0.05}\cdot \widehat{var}(\hat{\beta}_j|X)\right) = 95\%$$

$$prob\left(\hat{\beta}_j - c_{0.10}\cdot \widehat{var}(\hat{\beta}_j|X) \leq \beta_j \leq \hat{\beta}_j + c_{0.10}\cdot \widehat{var}(\hat{\beta}_j|X)\right) = 90\%$$

Daumenregeln: (für $n \to \infty$, also Normalverteilung)

$$c_{\underset{0.99}{\cancel{0.01}}} = 2.576,\ c_{\underset{0.95}{\cancel{0.05}}} = 1.96,\ c_{\underset{0.90}{\cancel{0.10}}} = 1.645$$

(Eigentlich hängt der kritische Wert c von den Freiheitsgraden ab.)

Zusammenhang von Konfidenzintervallen und Hypothesentests:

$$\gamma \notin [V_u, V_o] \Rightarrow \text{ verwerfe } H_0\ \beta_j = \gamma \text{ zugunsten } H_1\ \beta_j \neq \gamma$$

Hypothesentest über eine Linearkombination der Parameter

Hypothesentest über eine Linearkombination der Parameter

Bisher haben wir getestet, ob ein einzelner Parameter einen bestimmten Wert hat.

Nun testen wir, ob mehrere Parameter eine lineare Restriktion erfüllen.

Später werden wir testen, ob mehrere Paramater mehrere lineare Restriktionen erfüllen.

Hypothesentest über eine Linearkombination der Parameter

Eine Linearkombination der Parameter als Restriktion:

$$R_0 \cdot \beta_0 + R_1 \cdot \beta_1 + \ldots + R_k \cdot \beta_k = r \ ,$$

wobei $R_0, \ldots, R_k, r \in \mathbb{R}$.

Beispiel:

Sind β_1 und β_2 gleich?

$\Rightarrow \ R_0 = 0,\ R_1 = 1,\ R_2 = -1,\ R_3 = 0,\ \ldots\ R_k = 0,\ r = 0$

$$0 \cdot \beta_0 + 1 \cdot \beta_1 + (-1) \cdot \beta_2 + 0 \cdot \beta_3 + \ldots + 0 \cdot \beta_k = 0 \Leftrightarrow \beta_1 = \beta_2$$

Beispiel: Lohn bei verschiedenen Bildungsabschlüssen

twoyear.xls, $n = 6763$

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

$$\begin{aligned} jc &: \text{ Jahre junior college} \\ univ &: \text{ Jahre Uni} \\ exper &: \text{ Monate im Arbeitsleben} \end{aligned}$$

Fragestellung: Zählt (in Bezug auf den Lohn) ein Jahr junior college genau so viel wie ein Jahr Uni?

$$H_0 : \ \beta_1 = \beta_2$$

Schätzung:

$$\widehat{\log(wage)} = \underset{(0{,}021)}{1{,}472} + \underset{(0{,}0068)}{0{,}0667} jc + \underset{(0{,}0023)}{0{,}0769} univ + \underset{(0{,}0002)}{0{,}0049} exper$$

Beispiel: Lohn bei verschiedenen Bildungsabschlüssen

twoyear.xls, $n = 6763$

$H_0: \beta_1 = \beta_2 \Leftrightarrow \beta_1 - \beta_2 = 0$

Mögliche Teststatistik:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2 - 0}{\sqrt{\widehat{var}(\hat{\beta}_1 - \hat{\beta}_2|X)}}$$

Problem:

Die übliche Software liefert keine Information über $\widehat{var}(\hat{\beta}_1 - \hat{\beta}_2|X)$.

Es gilt: $var(\hat{\beta}_1 - \hat{\beta}_2|X) = var(\hat{\beta}_1|X) + var(\hat{\beta}_2|X) - 2cov(\hat{\beta}_1, \hat{\beta}_2|X)$.

$var(\hat{\beta}_1|X)$ und $var(\hat{\beta}_2|X)$ können durch $\widehat{var}(\hat{\beta}_1|X)$ bzw. $\widehat{var}(\hat{\beta}_2|X)$ erwartungstreu geschätzt werden, aber für die Schätzung von $cov(\hat{\beta}_1, \hat{\beta}_2|X)$ müsste $\hat{\sigma}(X'X)^{-1}_{(1,2)}$ bekannt sein. Dieser Term müsste separat ausgerechnet werden.

Es gibt aber einen einfacheren Weg!

Beispiel: Lohn bei verschiedenen Bildungsabschlüssen

twoyear.xls, $n = 6763$

$H_0: \; 0 \cdot \beta_0 + 1 \cdot \beta_1 - 1 \cdot \beta_2 + 0 \cdot \beta_3 = 0$

$\Leftrightarrow \quad \beta_1 - \beta_2 = 0$

Betrachte Modell:

$$\log(wage) = \beta_0 + \beta_1 jc + \beta_2 univ + \beta_3 exper + u$$

Subtrahiere und addiere $\beta_2 jc$: $-\beta_2 jc \quad +\beta_2 jc$

$$\log(wage) = \beta_0 + \underbrace{(\beta_1 - \beta_2)}_{\theta_1} jc + \beta_2 \underbrace{(jc + univ)}_{totcoll} + \beta_3 exper + u$$

Wir definieren einen neuen Parameter $\theta_1 = \beta_1 - \beta_2$ und einen neuen Regressor $totcoll = jc + univ$:

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 totcoll + \beta_3 exper + u$$

„Neue“ Hypothese: $H_0 \; \theta_1 = 0$

Beispiel: Lohn bei verschiedenen Bildungsabschlüssen

twoyear.xls, $n = 6763$

Die KQ-Schätzung von

$$\log(wage) = \beta_0 + \theta_1 jc + \beta_2 totcoll + \beta_3 exper + u$$

liefert

$$\widehat{\log(wage)} = \underset{(0{,}021)}{1.472} - \underset{(0{,}0069)}{0,0102}jc + \underset{(0{,}0023)}{0,0769}totcoll + \underset{(0{,}0002)}{0,0049}exper$$

Der Wert 0,0069 als Schätzer für die Standardabweichung von θ_1, also $\widehat{var}(\hat{\beta}_1 - \hat{\beta}_2|X)$, wird nun explizit von der Software ausgegeben!

t-Statistik: $t = \frac{-0{,}0102}{0{,}0069} = -1,478$

Kritischer Wert für zweiseitigen Test bei 6760 Freiheitsgraden und $\alpha = 5\%$: 1,96

Wegen $|t| < 1,96$ kann $H_0 : \theta_1 = 0$ bzw. $H_0 : \beta_1 = \beta_2$ nicht verworfen werden.

Hypothesentest über eine Linearkombination der Parameter

Betrachte Restriktion $R_0 \cdot \beta_0 + R_1 \cdot \beta_1 + \ldots + R_k \cdot \beta_k = r$

für ein gegebenes Modell $\mathbf{y} = \beta_0 + \beta_1 \cdot \mathbf{x}_1 + \ldots + \beta_k \cdot \mathbf{x}_k + \mathbf{u}$

Normiere die Restriktion für ein β_j mit $R_j \neq 0$

$$\underbrace{\frac{R_0}{R_j} \cdot \beta_0 + \ldots + \beta_j + \ldots + \frac{R_k}{R_j} \cdot \beta_k}_{=:\theta} = \frac{r}{R_j} =: \gamma$$

und forme das Modell um zu $\quad + \theta \cdot x_j \quad - \theta \cdot x_j$

$$\mathbf{y} = \beta_0 \left(1 - \frac{R_0}{R_j}\mathbf{x}_j\right) + \ldots + \underline{\theta \cdot x_j} + \ldots + \beta_k \cdot \left(\mathbf{x}_k - \frac{R_k}{R_j}\mathbf{x}_j\right) + \mathbf{u}$$

Berechne nun $\hat{\theta}$, $\widehat{var}(\hat{\theta}|X)$ und $t = \frac{\hat{\theta}-\gamma}{\sqrt{\widehat{var}(\hat{\theta}|X)}}$ und teste per t-Test:
H_0 $\theta = \gamma$.

Der F-Test: mehrere Linearkombinationen der Parameter

Der F-Test: mehrere Linearkombinationen der Parameter

Mit dem F-Test können wir mehrere simultane lineare Restriktionen an die Parameter überprüfen:

$$\begin{array}{ccccccccc} R_{10}\cdot\beta_0 & + & R_{11}\cdot\beta_1 & + & \dots & + & R_{1k}\cdot\beta_k & = & r_1 \\ R_{20}\cdot\beta_0 & + & R_{21}\cdot\beta_1 & + & \dots & + & R_{2k}\cdot\beta_k & = & r_2 \\ & & & & \vdots & & & & \\ R_{J0}\cdot\beta_0 & + & R_{J1}\cdot\beta_1 & + & \dots & + & R_{Jk}\cdot\beta_k & = & r_J \end{array} \quad \Leftrightarrow \underbrace{R}_{J\times k+1}\underbrace{\boldsymbol{\beta}}_{k+1\times 1} = \underbrace{\mathbf{r}}_{J\times 1}$$

Anzahl Restriktionen

mit

- R: Matrix der Ordnung $J \times k+1$
- $J \leq k$
- $rk(R) = J$ (voller Zeilenrang)
- $\mathbf{r} \in \mathbb{R}^J$

Beispiel: Determinanten der Gehälter von Baseball-Spielern

MLB1.xls $n = 353$, sum of squared residuals SSR=183,186, R^2=0,6278

$$
\begin{aligned}
\widehat{\log(salary)} &= \underset{(0{,}2888)}{11{,}1924} + \underset{(0{,}0121)}{0{,}0689}\, years + \underset{(0{,}0026)}{0{,}0126}\, gamesyr \\
&+ \underset{(0{,}0011)}{0{,}0010}\, bavg + \underset{(0{,}0161)}{0{,}0144}\, hrunsyr + \underset{(0{,}0072)}{0{,}0108}\, rbisyr
\end{aligned}
$$

- *salary*: Gehalt Baseballspieler
- *years*: Jahre in der Liga
- *gamesyr*: durchschnittliche Anzahl der Spiele pro Jahr
- *bavg*: batting average (Erfolgsw'keit eines Schlags)
- *hrunsyr*: Home runs pro Jahr
- *rbisyr*: runs batted in pro Jahr (Schlag führt zu Punktgewinn)

Beispiel: Determinanten der Gehälter von Baseball-Spielern

Die $t-$Quotienten für *bavg*, *hrunsyr* und *rbisyr* sind alle sehr niedrig, sodass wir für keinen der drei assoziierten Parameter die Hypothese $H_0 : \beta = 0$ mit einem t-Test verwerfen können.

Beachte aber die Aussage dieser drei Nullhypothesen:

Gegeben des Erklärungsgehaltes der restlichen vier Regressoren erklärt jeweils einer der drei Regressoren *bavg*, *hrunsyr* und *rbisyr* nichts von der Variation von $\log(sales)$.

Dies könnte zum Beispiel daran liegen, dass der Erklärungsgehalt *bavg* bereits in *hrunsyr* und *rbisyr* enthalten ist.

Beispiel: Determinanten der Gehälter von Baseball-Spielern

Wir möchten prüfen, ob die Gruppe von Regressoren *bavg*, *hrunsyr* und *rbisyr* keinen Erklärungsgehalt hat, wenn bereits für eine Konstante, *years* und *gamesyr* kontrolliert wird.

Drei Restriktionen: R r

$$0 \cdot \beta_0 + 0 \cdot \beta_{years} + 0 \cdot \beta_{gamesyr} + 1 \cdot \beta_{bavg} + 0 \cdot \beta_{hrunsyr} + 0 \cdot \beta_{rbisyr} = 0$$

$$0 \cdot \beta_0 + 0 \cdot \beta_{years} + 0 \cdot \beta_{gamesyr} + 0 \cdot \beta_{bavg} + 1 \cdot \beta_{hrunsyr} + 0 \cdot \beta_{rbisyr} = 0$$

$$0 \cdot \beta_0 + 0 \cdot \beta_{years} + 0 \cdot \beta_{gamesyr} + 0 \cdot \beta_{bavg} + 0 \cdot \beta_{hrunsyr} + 1 \cdot \beta_{rbisyr} = 0$$

$$\Leftrightarrow \begin{pmatrix} \beta_{bavg} = 0 \\ \beta_{hrunsyr} = 0 \\ \beta_{rbisyr} = 0 \end{pmatrix}$$

Beispiel: Determinanten der Gehälter von Baseball-Spielern

$\hat{\beta} \sim \mathcal{N} \Rightarrow R\hat{\beta} \sim \mathcal{N}$ $\quad H_0: R\beta = 0 \quad E[R\hat{\beta}] = RE[\hat{\beta}] = R\beta = 0$

Die drei Restriktionen in Matrixschreibweise: $\quad \Sigma(R\hat{\beta}) = R\,\Sigma(\hat{\beta})\,R'$

$$R = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, R\boldsymbol{\beta} = \mathbf{0}$$

Wir können diese Restriktion nun durch $R\hat{\boldsymbol{\beta}}$ schätzen.

Wir werden die statistischen Eigenschaften dieses Terms verwenden um später die Wald-Statistik daraus zu konstruieren.

Mit dieser können wir dann multiple lineare Restriktionen an die Parameter testen.

Zunächst führen wir aber eine Formel ein, welche für diesen Spezialfall einfacher zu handhaben ist.

Beispiel: Determinanten der Gehälter von Baseball-Spielern

mlb1.xls, $n = 353$, $SSR = 198,311$, $R^2 = 0,5971$

Wir vergleichen nun die Schätzung des unrestringierten Modells mit der Schätzung des unter H_0 geltenden restringierten Modells:

$$\widehat{\log(salary)} = \underset{(0{,}1083)}{11,2238} + \underset{(0{,}0125)}{0,0713} years + \underset{(0{,}0013)}{0,0292} gamesyr$$

Im Vergleich zum unrestringierten Modell ist die Summe der quadriertem Residuen (SSR) von 183,186 auf 198,311 gestiegen und R^2 ist von 0,6278 auf 0,5971 gefallen.

Beachte, dass bei der Wegnahme von Regressoren algebraisch SSR steigen und R^2 fallen **muss**.

Können wir das Ausmaß der Veränderung nutzen, um einen Test zu formulieren?

Restringierte und unrestringierte Modelle

Sei für ein beliebiges lineares Modell (mit Konstante) die Nullhypothese gegeben durch:

Restriktion

Falls H_0 wahr

$$H_0 : R\boldsymbol{\beta} = \mathbf{r}$$

$\Rightarrow SSR_r = SSR_{ur}$

Es bezeichne *ur* das unrestringierte und *r* das restringierte Modell.

$\Rightarrow F = 0$

Wir definieren die F-Statisik:

$$F = \frac{(\overbrace{SSR_r - SSR_{ur}}^{\geq 0})/J}{SSR_{ur}/(n-k-1)}$$

Die F-Statistik ist F-verteilt mit J und $n-k-1$ Freiheitsgraden. Wir werden die F-Verteilung später definieren.

Ist H_0 wahr, so unterscheiden sich SSR_r und SSR_{ur} kaum (wobei $SSR_r \geq SSR_{ur}$ gelten muss). Große Werte von F sprechen also gegen H_0.

Die F-Statistik

$$F = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)}$$

Wir können nun für gegebene Freiheitsgrade J, $n-k-1$ und für ein gegebenes Signifikanzniveau α kritische Werte c festlegen, sodass H_0 verworfen wird, falls $F > c$.

Im Beispiel: (mit $J = 3$ und $n-k-1 = 347$, $c_{\cancel{1\%}} = 3,78$)
38%

$$F = \frac{(SSR_r - SSR_{ur})/3}{SSR_{ur}/347} = \frac{(198,311 - 183,186)/3}{183,186/347} = 9,55 > 3,78$$

Wir können also H_0 mit einem Signifikanzniveau von $\alpha = 1\%$ verwerfen!

Die F-Verteilung

n. c. bis hier

Seien $x \sim \chi^2_m$ und $y \sim \chi^2_n$ und x und y unabhängig voneinander.

Dann heißt die Zufallsvariable $z = \frac{x/m}{y/n}$ F-**verteilt** mit m und n Freiheitsgraden.

Wir schreiben:

$$z \sim F_{m,n}$$

Standardisierung

$$v \sim \mathcal{N}(\overset{E[v]}{\mu}, \overset{Var(v)}{\sigma^2})$$

$$v - \mu \quad \leadsto \quad E[v-\mu] = E[v] - \mu = \mu - \mu = 0$$

$$\frac{v-\mu}{\sigma} \quad \leadsto \quad Var\left(\frac{v-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(v-\mu) = \frac{1}{\sigma^2} Var(v)$$

$$= \frac{\sigma^2}{\sigma^2} = 1$$

$$\frac{v-\mu}{\sigma} \sim \mathcal{N}(0, 1)$$

F - Test: Test auf Restriktion $\underbrace{R\beta = r}_{\substack{\text{Nullhypothese} \\ H_0}}$

Schätzer $R \cdot \hat{\beta}$

MLR 6: $u \sim \mathcal{N}(0, \sigma^2 I)$

$$\Rightarrow \hat{\beta} \sim \mathcal{N}\left(\beta, \sigma^2 (X'X)^{-1}\right)$$

$$\Rightarrow R\hat{\beta} \sim \mathcal{N}\Big(\underbrace{E[R\hat{\beta}]}_{r}, \overbrace{\Sigma(R\hat{\beta})}^{\sigma^2 R (X'X)^{-1} R'}\Big)$$

$$E[R\hat{\beta}] = R E[\hat{\beta}] = R \cdot \beta \overset{H_0}{=} r$$

$$\Sigma(R\hat{\beta}) = R\, \Sigma(\hat{\beta})\, R' = R \cdot \sigma^2 (X'X)^{-1} R'$$

$$R\hat{\beta} - r \rightsquigarrow E[R\hat{\beta} - r] = E[R\hat{\beta}] - r \overset{H_0}{=} r - r = 0 \checkmark$$

Die Wurzel einer Matrix

$$B = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad \text{mit} \quad a \cdot c \neq 0$$

$$\Rightarrow \begin{pmatrix} a \\ 0 \end{pmatrix} \;\&\; \begin{pmatrix} b \\ c \end{pmatrix} \quad \text{linear unabhängig}$$

$$\Rightarrow rg(B) = 2$$

$$\Rightarrow B^{-1} = \frac{1}{a \cdot c} \begin{pmatrix} c & -b \\ 0 & a \end{pmatrix}$$

$$B \cdot B' = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} = \begin{pmatrix} \underbrace{a^2}_{>0} + \underbrace{b^2}_{>0} & bc \\ cb & \underbrace{c^2}_{>0} \end{pmatrix} = A$$

positiv definit
symmetrisch
pds

$$\det(A) = (a^2 + b^2)c^2 - b \cdot c \cdot c \cdot b = \underbrace{a^2c^2}_{>0} + \cancel{b^2c^2} - \cancel{b^2c^2}$$

Eine Matrix A hat eine Wurzel B genau dann wenn A pds

- $B \cdot B' = A$
- $(B^{-1})' B^{-1} = A^{-1}$

$$R\hat{\beta} - r \sim \mathcal{N}\left(0, \sigma^2 R (X'X)^{-1} R'\right) \qquad \underset{J \times k+1}{R}$$

$rk(X) = k+1$ (MLR 3) $\Rightarrow X'X$ pds

$\Rightarrow (X'X)^{-1}$ pds

$rk(R) = J \Rightarrow \sigma^2 R (X'X)^{-1} R'$ pds

$\Rightarrow$ es gibt eine Matrix B mit $B \cdot B' = \sigma^2 R (X'X)^{-1} R'$

$(B^{-1})' \cdot B^{-1} = [\sigma^2 R (X'X)^{-1} R']^{-1}$

$$B^{-1}(R\hat\beta - r) \qquad E[B^{-1}(R\hat\beta - r)] = B^{-1}(R\underbrace{\underbrace{E[\hat\beta]}_{=\beta}}_{H_0:\ =r} - r) = B^{-1}(r-r) = 0 \checkmark$$

$$\Sigma(B^{-1}(R\hat\beta - r)) = B^{-1}\,\Sigma(R\hat\beta - r)\,(B^{-1})'$$

$$= B^{-1}\,\sigma^2 R(X'X)^{-1}R'\,(B^{-1})'$$

$$= \underbrace{B^{-1}\cdot B}_{I}\cdot B'\,(B^{-1})'$$

J: # Restriktionen

$$= I\cdot\underbrace{B'(B')^{-1}}_{I} = I$$

$$\overbrace{B^{-1}(R\hat\beta - r)}^{\text{Spaltenvektor mit } J \text{ Komponenten}} \sim N(0, I)$$

$$\Rightarrow [B^{-1}(R\hat\beta - r)]'\,B^{-1}(R\hat\beta - r) \sim \chi_J$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \rightsquigarrow (x_1, x_2, \dots, x_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + \dots + x_n^2$$

$$= \sum_{i=1}^{n} x_i^2$$

$$(R\hat{\beta} - r)' \underbrace{(B^{-1})' B^{-1}}_{[\sigma^2 R(X'X)^{-1}R']^{-1}} (R\hat{\beta} - r)$$

$$= \sigma^2 (R\hat{\beta} - r)'[R(X'X)^{-1}R]^{-1}(R\hat{\beta} - r)$$

$$\hat{u} = \underbrace{[I - X(X'X)^{-1}X']}_{rk() = tr() = n-k-1} u \quad , \quad u \sim N(0, \sigma^2 I)$$

Wald-Statistik

$\frac{\hat{\sigma}}{\sigma} \sim$

Für $R \in \mathbb{R}^{J \times (k+1)}$, $\mathbf{r} \in \mathbb{R}^J$, $J \leq k+1$ und $H_0: \; R\boldsymbol{\beta} = \mathbf{r}$:

$$W = \frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)}$$

Wir zeigen in Appendix C, dass:

$$\frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} \sim F_{J,n-k-1}$$

und wir zeigen in Appendix D, dass:

$$F = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)} = \frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} = W$$

Zusammenfassung *F*-Test

Nullhypothese

$$H_0: \; R\boldsymbol{\beta} = \mathbf{r}, \text{ mit } R \in \mathbb{R}^J \times \mathbb{R}^{k+1}, \; J \leq k, \; rk(R) = J, \; \mathbf{r} \in \mathbb{R}^J$$

Teststatistik

$$\begin{aligned} W &= \frac{\underbrace{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)'}_{=0,\text{ falls } H_0 \checkmark} \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} \\ &= \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)} = F \end{aligned}$$

ist F-verteilt mit J und $n-k-1$ Freiheitsgraden.

Die Anwendung des F-Tests

?

im Output enthalten

$$F = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)}$$

Der Nenner des Terms ist dem Output ökonometrischer Software zu entnehmen.

Im Allgemeinen gibt die Software allerdings den Zähler nicht aus.

In bestimmten Spezialfällen können wir den Zähler aber leicht herauslesen.

F-Test: Zwei Spezialfälle

$$\underline{R = (0, 0, \ldots, \overset{j}{1}, \ldots, 0),\ \mathbf{r} = 0 \text{ (mit } J = 1)}$$

$$H_0:\ \beta_j = 0$$

Dieser Fall verdeutlicht den Zusammenhang zwischen F- und t-Test.

Spalten mit nullen

$$\underline{R = (\mathbf{0}, I_k),\ \mathbf{r} = \mathbf{0} \text{ (mit } J = k)}$$

$$H_0:\ \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

Dieser Fall wird auch **Globaler F-Test** genannt.

$$R = (0, 0, \dots, \overset{j}{1}, \dots, 0), \; \mathbf{r} = 0 \text{ (mit } J = 1)$$

Der Ausdruck

$$W = \frac{\left(\overbrace{R\hat{\boldsymbol{\beta}}}^{=\hat\beta_j} - \overset{0}{\mathbf{r}}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(\overset{\hat\beta_j}{R\hat{\boldsymbol{\beta}}} - \mathbf{r}\right) \Big/ J}{\underbrace{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)}_{\hat\sigma^2}}$$

vereinfacht sich zu

t_j mit $H_0 \; \hat\beta_j = 0$

$$\frac{\hat\beta_j \left[(X'X)^{-1}_{j,j}\right]^{-1} \hat\beta_j}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)} = \frac{\hat\beta_j^2}{\hat\sigma^2 (X'X)^{-1}_{j,j}} = \left(\frac{\hat\beta_j}{\sqrt{\widehat{var}(\hat\beta_j|X)}}\right)^2$$

Die F-Statistik ist also das Quadrat der t-Statistik zur Nullhypothese $H_0: \; \beta_j = 0$.

SST = SSE + SSR

y $\hat{y}$ $\hat{u}$

Es gilt:

$$\hat{\mathbf{u}}'\hat{\mathbf{u}} = \sum_{i=1}^{n} \hat{u}_i^2 = SSR$$

Also:

$$F = \frac{\left(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}\right)/J}{\hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}/(n-k-1)} = \frac{\left(SSR_r - SSR_{ur}\right)/J}{SSR_{ur}/(n-k-1)}$$

Außerdem gilt (für alle Restriktionen)

$$SST_r = SST_{ur} = SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

da sich die Restriktion $R\boldsymbol{\beta} = \mathbf{r}$ nur auf die rechte Seite der Gleichung $\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}$ auswirkt.

$SST = SST_r = SST_{ur}$

$$F = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)}$$

$$= \frac{\left(\frac{SSR_r}{SST_r} - \frac{SSR_{ur}}{SST_{ur}}\right)/J}{\frac{SSR_{ur}}{SST_{ur}}/(n-k-1)}$$

Bestimmtheitsmaß

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

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

$$\Rightarrow F = \frac{\left(\cancel{1} - R_r^2 - (\cancel{1} - R_{ur}^2)\right) / J}{(1 - R_{ur}^2) / (n - k - 1)}$$

im Output enthalten

beim globalen F-Test ($\beta_1 = \beta_2 = \dots = \beta_k = 0$): $R_r^2 = 0$

$$F = \frac{(R_{ur}^2 - R_r^2) / J}{(1 - R_{ur}^2) / (n - k - 1)}$$

im Output enthalten

$$= \frac{(SSR_r - SSR_{ur}) / J}{SSR_{ur} / (n - k - 1)}$$

Beim globalen F-Test: (mit $J = k$)

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

$R = (\mathbf{0}, I_k)$, $\mathbf{r} = \mathbf{0}$ (mit $J = k$)

Globaler F-Test $H_0: \ \beta_1 = \beta_2 = \ldots = \beta_k = 0$

Für den Spezialfall der Restriktion

$$\beta_1 = \beta_2 = \ldots = \beta_k = 0 \ ,$$

also den globalen F-Test mit

$$R = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} \text{ und } \mathbf{r} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}$$

Schreiben wir die Formel der F-Statistik etwas um:

$$F = \frac{\left(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}\right)/J}{\hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}/(n-k-1)} = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)} = \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_{ur}}{SST}\right)/J}{\frac{SSR_{ur}}{SST}/(n-k-1)}$$

Unter der Restriktion

$$\beta_1 = \beta_2 = \ldots = \beta_k = 0$$

gilt:

$$y_i = \beta_0 + u_i \text{ für } i = 1, \ldots, n$$

Der OLS-Schätzer für β_0 ist schlicht $\hat{\beta}_0 = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Daher gilt für die Residuen: $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 = y_i - \bar{y}$.

Daher gilt unter dieser Restriktion:

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

Die F-Statistik reduziert sich daher zu

$$F = \frac{\left(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}\right)/J}{\hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}/(n-k-1)} = \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_{ur}}{SST}\right)/J}{\frac{SSR_{ur}}{SST}/(n-k-1)} = \frac{\left(1 - \frac{SSR_{ur}}{SST}\right)/J}{\frac{SSR_{ur}}{SST}/(n-k-1)}$$

Spezialfall: Globaler F-Test ($\beta_1 = \ldots = \beta_k = 0$)

Mit

$$SSR_r = SST$$

und

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

gilt für die F-Statistik des globalen F-Tests:

$$F = \frac{\left(\hat{\mathbf{u}}_r'\hat{\mathbf{u}}_r - \hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}\right)/J}{\hat{\mathbf{u}}_{ur}'\hat{\mathbf{u}}_{ur}/(n-k-1)} = \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_{ur}}{SST}\right)/J}{\frac{SSR_{ur}}{SST}/(n-k-1)} = \frac{R^2/J}{(1-R^2)/(n-k-1)}$$

Beispiel: Bildung der Eltern und Geburtsgewicht der Kinder

bwght.xls n=1191

$$bwght = \overbrace{\underbrace{\beta_0+\beta_1 cigs+\beta_2 parity+\beta_3 faminc\;{}_{+u}}_{\text{restringiertes Modell}}+\beta_4 motheduc+\beta_5 fatheduc+u}^{\text{unrestringiertes Modell}}$$

- *bwght*: Geburtsgewicht des Kindes, in Pfund
- *cigs*: #Zigaretten/Tag der Mutter während Schwangerschaft
- *parity*: Geburtsreihenfolge des Kindes
- *faminc*: jährliches Familieneinkommen
- *motheduc*: Schuljahre der Mutter
- *fatheduc*: Schuljahre des Vaters

Nullhypothese $H_0: \ \beta_4 = \beta_5 = 0$

Beispiel: Bildung der Eltern und Geburtsgewicht der Kinder

bwght.xls n=1191

Nichtrestringiertes Modell *ur*:

$$\begin{aligned}\widehat{bwght} \quad &= \quad \underset{(3{,}7285)}{114,524} - \underset{(0{,}11035)}{0,595936}\,\text{cigs} + \underset{(0{,}65941)}{1,78760}\,\text{parity} + \underset{(0{,}036562)}{0,0560414}\,\text{fami} \\ &\quad - \underset{(0{,}31986)}{0,370450}\,\text{motheduc} + \underset{(0{,}28264)}{0,472394}\,\text{fatheduc}\end{aligned}$$

$$\begin{aligned} SSR_{ur} &= 464041,1 \\ R^2_{ur} &= 0,038748 \\ F_{5,1185} &= 9,553500 \end{aligned}$$

↳ F-Statistik für globalen F-Test

Mit $c_{F_{5,1185},1\%} = 3,21$ und $c_{F_{5,1185},5\%} = 3,0$ können wir schon jetzt die Hypothese verwerfen, dass alle Regressoren gemeinsam insignifikant wären.

Beispiel: Bildung der Eltern und Geburtsgewicht der Kinder

bwght.xls n=1191

Restringiertes Modell r:

$$\begin{aligned}
\widehat{bwght} &= \underset{(1{,}6559)}{115,470} - \underset{(0{,}10877)}{0,597852}\,\text{cigs} + \underset{(0{,}65754)}{1,83227}\,\text{parity} + \underset{(0{,}032394)}{0,0670618}\,\text{famir} \\
SSR_r &= 465166,8 \\
R_r^2 &= 0,036416
\end{aligned}$$

Es gilt nun

$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/1185} = 1,436 = \frac{(R_{ur}^2 - R_r^2)/2}{(1 - R_{ur}^2)/1185}$$

Mit $c_{F_{5,1185},1\%} = 3,21$ und $c_{F_{5,1185},5\%} = 3,0$ können wir $H_0:\ \beta_{motheduc} = \beta_{fatheduc} = 0$ nicht verwerfen.

Test auf Strukturbruch: Chow-Test

Fragestellung:

Sind die Parameter über verschiedene Teilstichproben gleich?

Querschnitte:

Sind die Parameter für Beobachtungen aus zwei Regionen dieselben?

Zeitreihen:

Sind die Parameter vor und nach einembestimmten Zeitpunkt gleich geblieben?

Das kann man mit einem F - Test auf einen Strukturbruch im Modell überprüfen. In diesem Kontext bezeichnet man den F - Test nach seinem Erfinder als **Chow** - **Test**.

Test auf Strukturbruch: Chow-Test

Das Modell sei:

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + u_i$$

Zwei disjunkte Gruppen I_1 und I_2 mit $I_1 \cup I_2 = \{1, 2, \ldots, n\}$.

Nullhypothese: Parameter des Modells sind in beiden Gruppen gleich.

Gegenhypothese:
Die Parameter unterscheiden sich in beiden Gruppen strukturell („Strukturbruch")

Es ist also (for ein oder mehrere j)

$$\begin{aligned} H_0 \quad &: \quad \beta_j = \beta_j \text{ für alle } i = 1, \ldots, n \\ &\quad\quad \text{gegen} \\ H_1 \quad &: \quad \beta_j = \beta_j^{(1)} \text{ für } i \in I_1 \\ &\quad\quad \beta_j = \beta_j^{(2)} \neq \beta_j^{(1)} \text{ für } i \in I_2 \end{aligned}$$

Test auf Strukturbruch: Chow-Test

bis hier 18.6.

Man kann jede der Formen des F - Tests verwenden. Am einfachsten ist die Verwendung der ersten Form. Man definiert eine Dummy-Variable:

$$
\begin{aligned}
d_i &= 0 \text{ für } i \in I_1 \\
d_i &= 1 \text{ für } i \in I_2
\end{aligned}
$$

und schätzt dann das Modell mit zusätzlichem Parameter $\gamma \in \mathbb{R}^k$

schätze $\hat{\beta}$, $\hat{\gamma}$
berechne $\hat{u}_{ur} \rightarrow SSR_{ur}$

$$y_i = \underbrace{\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}}_{\text{restringiertes Modell}} + \cancel{\sum_{j=1}^{k} \gamma_j \left(d_i \cdot x_{ij}\right)} + u_i$$

schätze $\hat{\beta}_r$ berechne $\hat{u}_r \rightarrow SSR_r$

Natürlich kann sich der Test auch nur auf einzelne Parameter oder eine Teilmenge aller Parameter beziehen.

Strukturbruch, falls ein oder mehrere Parameter $\gamma_j \neq 0$

Test auf Strukturbruch: Chow-Test

Nullhypothese (Parameterstabilität)

$$H_0 : \gamma_1 = ... = \gamma_k = 0$$

oder

$$H_0 : \overbrace{\begin{pmatrix} \underset{\text{Nullmatrix}}{N_{k\times k+1}} & \underset{\text{Einheitsmatrix}}{I_k} \end{pmatrix}}^{R} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = \overset{r}{0_k}$$

Überprüfbar mit der ersten Form des F - Tests

Die Ablehnung der Nullhypothese besagt, dass eine signifikante Variation der Parameter (also ein Strukturbruch) vorliegt, und es daher ein Fehler wäre, über die gesamte Stichprobe unveränderte Parameter anzunehmen.

Der p-Wert der F-Statistik

Der p-Wert der F-Statistik hat die analoge Interpretation des p-Werts für die t-Statistik:

Der p-Wert ist das kleinste Signifikanzniveau α_{min} zu dem H_0 bei einer gegebenen Stichprobe verworfen wird.

Je kleiner der p-Wert, desto stärker die Evidenz gegen H_0.

Zusammenfassung Statistischer Test

- **Test-Statistik:**
  Aus Daten errechnete Zufallsgröße, die einer bestimmten Verteilung unterliegt
- **Statistischer Test:**
  Prüfung, ob Test-Statistik unter einer festgelegten Nullhypothese H_0 plausible Werte annimmt
- **Signifikanzniveau α:**
  Wahrscheinlichkeit, mit der H_0 abgelehnt wird, falls sie wahr ist
- ***p*-Wert:**
  Marginales/kleinstes Signifikanzniveau α_{min} zu dem H_0 verworfen werden muss.
- **Kritischer Wert c:**
  Falls die Test-Statistik (im Betrag) den kritischen Wert überschreitet, wird H_0 abgelehnt

Zusammenfassung t-Test

- $\hat{\beta}$ ist unter **MLR 6** normalverteilt.
- $(n-k-1)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\overbrace{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}^{SSR}$ ist unter **MLR 6** χ^2_{n-k-1}-verteilt.
- $H_0 : \beta_j = \gamma$

  Die t-**Statistik** $\underbrace{\frac{\hat{\beta}_j - \cancel{\beta_j}\,\gamma}{\sqrt{\widehat{var}(\hat{\beta}_j|X)}}}_{se(\hat{\beta}_j)}$ ist t_{n-k-1}-verteilt.
- **Zweiseitiger t-Test:**
  Prüfung an einzelne Parameter, ob $H_0 : \ \beta_j = \gamma$ verworfen werden muss
- **Einseitiger t-Test:**
  Prüfung an einzelne Parameter, ob $H_0 : \ \beta_j \geq, \leq \gamma$ verworfen werden muss
- **Konfidenzintervall:**
  Intervall dessen Intervallgrenzen von den (zufälligen) Daten abhängen. Parameterwerte außerhalb des Intervalls sind nicht mit den Daten vereinbar.

Zusammenfassung F-Test

- **F-Test:**
  Prüft, ob System von linearen Restriktionen $R\boldsymbol{\beta} = \mathbf{r}$ mit Daten vereinbar ist

- **Wald-Statistik:**
  $W = \dfrac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/(n-k-1)}$

- Die Wald-Statistik ist $F_{J,n-k-1}$-verteilt.

$\dfrac{(R^2_{ur} - R^2_r)/J}{(1 - R^2_{ur})/(n-k-1)}$

$\parallel$

- Für die F-Statistik gilt: $F = \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)} = W$

- **Globaler F-Test:**
  $H_0: \ \beta_1 = \ldots = \beta_k = 0 \Rightarrow F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{k,n-k-1}$

+ Chow-Test auf Strukturbruch

Die Appendizes A bis D enthalten Beweise, die nicht in der Klausur abgefragt werden.

Appendix A

Beweis für:

$$\hat{\sigma}^2(n-k-1)/\sigma^2$$

ist χ^2-verteilt mit $n-k-1$ Freiheitsgraden.

Konstruktion und Verteilung von $\hat{\sigma}^2(n-k-1)/\sigma^2$

Da $\mathbf{u} \sim \mathcal{N}(0, \sigma^2 I)$, gilt $\frac{1}{\sigma}\mathbf{u} \sim \mathcal{N}(0, I)$ mit u_i unabhängig.

Außerdem gilt mit $\mathbf{u} = y - X\boldsymbol{\beta}$:

$$
\begin{aligned}
\frac{\mathbf{u}'}{\sigma}[I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma} &= \frac{(\mathbf{y} - X\boldsymbol{\beta})'}{\sigma}[I - X(X'X)^{-1}X']\frac{(\mathbf{y} - X\boldsymbol{\beta})}{\sigma} \\
&= \frac{\mathbf{y}'}{\sigma}[I - X(X'X)^{-1}X']\frac{\mathbf{y}}{\sigma} \quad = \quad \frac{1}{\sigma^2}\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}
\end{aligned}
$$

Die Matrix $I - X(X'X)^{-1}X'$ ist idempotent und hat als solche die Eigenschaft:

$$
rk(I - X(X'X)^{-1}X') = tr(I - X(X'X)^{-1}X') = n - k - 1,
$$

wobei wir die zweite Gleichung in Kapitel 3 hergeleitet hatten.

Konstruktion und Verteilung von $(n-k-1)\hat{\sigma}^2/\sigma^2$

Für $(n-k-1)\hat{\sigma}^2/\sigma^2$ folgt:

$$(n-k-1)\hat{\sigma}^2/\sigma^2 = (n-k-1)\frac{1}{n-k-1}\frac{1}{\sigma^2}\hat{\mathbf{u}}'\hat{\mathbf{u}} = \frac{\mathbf{u}'}{\sigma}[I-X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}$$

Theorem B.8 (aus Greene 2012, p. 1084)

Wenn $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, I)$ und A idempotent,
dann ist $\mathbf{x}'A\mathbf{x}$ χ^2-verteilt mit $rk(A)$ Freiheitsgraden.

Es folgt also, dass $\hat{\sigma}^2(n-k-1)/\sigma^2$ χ^2-verteilt mit $n-k-1$ Freiheitsgraden ist.

Appendix B

Beweis für:

$$\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}}$$

und

$$(n-k-1)\hat{\sigma}^2/\sigma^2$$

sind unabhängig.

Statistische Unabhängigkeit

Schritt 1: Der Zähler ist eine lineare Funktion von $\frac{\mathbf{u}}{\sigma} \sim \mathcal{N}(0,1)$.

$$\frac{\hat{\beta}_j - E[\hat{\beta}_j|X]}{\sqrt{var(\hat{\beta}_j|X)}} = \frac{e_j'(X'X)^{-1}X'\mathbf{y} - \beta_j}{\sigma\sqrt{(X'X)^{-1}_{j,j}}} = \frac{e_j'(X'X)^{-1}X'\mathbf{u}}{\sigma\sqrt{(X'X)^{-1}_{j,j}}}$$

Schritt 2: Die quadratische Form der χ^2-Verteilten Zufallsvariable ist gegeben durch

$$\frac{\mathbf{u}'}{\sigma}[I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma} = \left([I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}\right)'\left([I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}\right)$$

Statistische Unabhängigkeit

Schritt 3

$\frac{e_j'(X'X)^{-1}X'\mathbf{u}}{\sigma\sqrt{(X'X)^{-1}_{j,j}}}$ und $[I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}$ sind unkorrelliert:

$$
\begin{aligned}
& cov\left(\frac{e_j'(X'X)^{-1}X'\mathbf{u}}{\sigma\sqrt{(X'X)^{-1}_{j,j}}}, [I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}\right) \\
= \quad & \frac{e_j'(X'X)^{-1}X'}{\sqrt{(X'X)^{-1}_{j,j}}} cov\left(\frac{\mathbf{u}}{\sigma}, \frac{\mathbf{u}}{\sigma}\right)[I - X(X'X)^{-1}X']' \\
= \quad & \frac{1}{\sqrt{(X'X)^{-1}_{j,j}}} e_j'(X'X)^{-1}X'[I - X(X'X)^{-1}X'] \\
= \quad & \frac{1}{\sqrt{(X'X)^{-1}_{j,j}}} e_j'(X'X)^{-1}[X' - \underbrace{X'X(X'X)^{-1}}_{I}X'] = N
\end{aligned}
$$

Normalverteilt & unkorrelliert $\Rightarrow$ statistisch unabhängig

Appendix C

Beweis für

$$W = \frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\left(\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}\right) / \left(n-k-1\right)} \sim F_{J,n-k-1}$$

Begründung für: die Wald-Statistik ist F-verteilt.

Wir hatten bereits gezeigt, dass $\frac{1}{\sigma^2}\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} \sim \chi^2_{n-k-1}$-verteilt ist.

Wir schreiben daher

$$F = \frac{\left(\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ \sigma^2\right) \Big/ J}{\left(\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/\sigma^2\right) / (n-k-1)}$$

und müssen nun zeigen, dass

$\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ \sigma^2 \; \chi^2_J$-verteilt ist.

$\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ \sigma^2\ \chi^2_J$-verteilt

Betrachte zunächst:

$$R\hat{\boldsymbol{\beta}} - \mathbf{r} = R(X'X)^{-1}X'\left(X\boldsymbol{\beta} + \mathbf{u}\right) - \mathbf{r} = R\boldsymbol{\beta} - \mathbf{r} + R(X'X)^{-1}X'\mathbf{u}$$

Mit $H_0 : R\boldsymbol{\beta} = \mathbf{r}$ können wir den Ausdruck im Titel umschreiben als:

$$\frac{\mathbf{u}'}{\sigma} \underbrace{X(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'}_{=:A} \frac{\mathbf{u}}{\sigma}$$

Mit $A'A = A$ (A idempotent symmetrisch) gilt:

$$\frac{\mathbf{u}'}{\sigma} A \frac{\mathbf{u}}{\sigma} \sim \chi^2_s\ ,$$

wobei $s = rk(A)$. Wegen $rk(X) = k + 1$ und $rk(R) = J$ und $rk(CD) = min\{rk(C), rk(D)\}$ gilt $rk(A) = \min\{J, k+1\} = J$.

Unabhängigkeit von $\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/\sigma^2$ und $\mathbf{u}'A\mathbf{u}/\sigma^2$

Mit $\hat{\boldsymbol{u}} = [I - X(X'X)^{-1}X']\mathbf{u}$ und $[I - X(X'X)^{-1}X']$ idempotent symmetrisch gilt:

$$\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/\sigma^2 = \left([I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}\right)'[I - X(X'X)^{-1}X']\frac{\mathbf{u}}{\sigma}$$

und mit A idempotent symmetrisch gilt:

$$\mathbf{u}'A\mathbf{u}/\sigma^2 = \left(A\frac{\mathbf{u}}{\sigma}\right)'A\frac{\mathbf{u}}{\sigma}$$

Die Kovarianz der beiden normalverteilten Zufallsvektoren $\hat{\boldsymbol{u}}/\sigma$ und $A\mathbf{u}/\sigma$ beträgt:

$$\begin{aligned}
&cov(\hat{\boldsymbol{u}}/\sigma, A\mathbf{u}/\sigma) \\
&\quad = [I - X(X'X)^{-1}X']\underbrace{cov(\mathbf{u}/\sigma, \mathbf{u}/\sigma)}_{I}A' \\
&\quad = \underbrace{[I - X(X'X)^{-1}X']X}_{=N}(X'X)^{-1}R'\left[R(X'X)^{-1}R'\right]^{-1}R(X'X)^{-1}X'
\end{aligned}$$

Appendix D

Beweis für

$$
\begin{aligned}
W &= \frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/(n-k-1)} \\
&= \frac{(SSR_r - SSR_{ur})/J}{SSR_{ur}/(n-k-1)} = F
\end{aligned}
$$

$$W = \frac{\left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right) \Big/ J}{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}/(n-k-1)}$$

Während im Nenner $SSR_{ur} = \sum_{i=1}^{n} \hat{u}_i^2 = \hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ übereinstimmt, ist die Beweisführung für die Gleichheit der Nenner aufwendiger.

$$SSR_r - SSR_{ur} = \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[R(X'X)^{-1}R'\right]^{-1} \left(R\hat{\boldsymbol{\beta}} - \mathbf{r}\right)$$

Im folgenden bezeichnet $\hat{\boldsymbol{\beta}}_{ur}$ den OLS-Schätzer im unrestringierten Modell, also $\hat{\boldsymbol{\beta}}_{ur} = \hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$, und $\hat{\boldsymbol{\beta}}_r$ den OLS-Schätzer im restringierten Modell.

Wir beginnen mit der Herleitung für:

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}}_{ur} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)$$

Der restringierte Schätzer $\hat{\boldsymbol{\beta}}_r$ löst folgendes Minimierungsproblem unter der Nebenbedingung $R\boldsymbol{\beta} = \mathbf{r}$:[1]

$$\max_{\mathbf{b}\in\mathbb{R}^{k+1}} (\mathbf{y} - X\mathbf{b})'(\mathbf{y} - X\mathbf{b}) \text{ u.d.B. } 2(\mathbf{r} - R\mathbf{b}) = \mathbf{0}$$

$$\mathcal{L}(\mathbf{b}, \boldsymbol{\lambda}) = \mathbf{y}'\mathbf{y} - 2\mathbf{b}'X'\mathbf{y} + \mathbf{b}'X'X\mathbf{b} - \boldsymbol{\lambda}' \cdot 2(\mathbf{r} - R\mathbf{b})$$

BEO:

$$\begin{aligned} -2X'\mathbf{y} + 2X'X\hat{\boldsymbol{\beta}}_r + 2R'\boldsymbol{\lambda} &= \mathbf{0} \\ R\hat{\boldsymbol{\beta}}_r &= \mathbf{r} \end{aligned}$$

$$\Leftrightarrow \begin{pmatrix} X'X & R' \\ R & N \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}}_r \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} X'\mathbf{y} \\ \mathbf{r} \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \hat{\boldsymbol{\beta}}_r \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & N \end{pmatrix}^{-1} \begin{pmatrix} X'\mathbf{y} \\ \mathbf{r} \end{pmatrix}$$

[1] Es gilt $R\mathbf{b} = \mathbf{r} \Leftrightarrow 2(\mathbf{r} - R\mathbf{b}) = 0$. Dies vereinfacht die Rechnung an späterer Stelle.

Die inverse Matrix

$$\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & N \end{pmatrix}^{-1}$$

erfüllt:

$$\begin{pmatrix} X'X & R' \\ R & N \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} I & N \\ N & I \end{pmatrix}$$

Also:

$$\begin{aligned} X'XB_{11} + R'B_{21} &= I && (1) \\ X'XB_{12} + R'B_{22} &= N && (2) \\ RB_{11} &= N && (3) \\ RB_{12} &= I && (4) \end{aligned}$$

Aus (1) folgt:

$$B_{11} + (X'X)^{-1}R'B_{21} = (X'X)^{-1}$$

und mit (3) folgt:

$$\underbrace{RB_{11}}_{=N} + R(X'X)^{-1}R'B_{21} = R(X'X)^{-1}$$

Da R vollen Zeilenrang hat, existiert die Inverse $[R(X'X)^{-1}R']^{-1}$ eindeutig und

$$B_{21} = [R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

Damit folgt wieder aus (1):

$$B_{11} = (X'X)^{-1} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}$$

Desweiteren lösen

$$
\begin{aligned}
B_{12} &= (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1} \\
B_{22} &= -[R(X'X)^{-1}R']^{-1}
\end{aligned}
$$

Die Gleichungen (2) und (4). Mit $M := [R(X'X)^{-1}R']^{-1}$ gilt also

$$
\begin{pmatrix} X'X & R' \\ R & N \end{pmatrix}^{-1} = \begin{pmatrix} (X'X)^{-1}\left[I - R'MR(X'X)^{-1}\right] & (X'X)^{-1}R'M \\ MR(X'X)^{-1} & -M \end{pmatrix}
$$

und wir können

$$
\begin{pmatrix} \hat{\boldsymbol{\beta}}_r \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} X'X & R' \\ R & N \end{pmatrix}^{-1} \begin{pmatrix} X'\mathbf{y} \\ \mathbf{r} \end{pmatrix}
$$

lösen.

$$
\begin{aligned}
&\begin{pmatrix} \hat{\boldsymbol{\beta}}_r \\ \boldsymbol{\lambda} \end{pmatrix} \\
= &\begin{pmatrix} (X'X)^{-1}\left[I - R'MR(X'X)^{-1}\right] & (X'X)^{-1}R'M \\ MR(X'X)^{-1} & -M \end{pmatrix} \begin{pmatrix} X'\mathbf{y} \\ \mathbf{r} \end{pmatrix} \\
= &\begin{pmatrix} (X'X)^{-1}X'\mathbf{y} - (X'X)^{-1}R'M\left(R(X'X)^{-1}X'\mathbf{y} - \mathbf{r}\right) \\ M\left(R(X'X)^{-1}X'\mathbf{y} - \mathbf{r}\right) \end{pmatrix}
\end{aligned}
$$

Mit $\hat{\boldsymbol{\beta}}_{ur} = (X'X)^{-1}X'\mathbf{y}$ gilt also

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_r &= \hat{\boldsymbol{\beta}}_{ur} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right) \\
\boldsymbol{\lambda} &= [R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)
\end{aligned}
$$

Die BEO des restringierten Optimierungsproblems sind nun leicht zu überprüfen.

Wir leiten nun SSR_r her:

$$\hat{\boldsymbol{u}}_r = \mathbf{y} - X\hat{\boldsymbol{\beta}}_r = \mathbf{y} - X\hat{\boldsymbol{\beta}}_{ur} - X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right) = \hat{\boldsymbol{u}}_{ur} - X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)$$

Daher gilt

$$\begin{aligned} \hat{\boldsymbol{u}}_r'\hat{\boldsymbol{u}}_r &= \left(\hat{\boldsymbol{u}}_{ur} - X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)\right)'\left(\hat{\boldsymbol{u}}_{ur} - X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)\right) \\ &= \hat{\boldsymbol{u}}_{ur}'\hat{\boldsymbol{u}}_{ur} - \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)' X'\hat{\boldsymbol{u}}_{ur} \\ &- \hat{\boldsymbol{u}}_{ur}'X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right) + \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)' X'X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right) \end{aligned}$$

und wegen $X'\hat{\boldsymbol{u}}_{ur} = \mathbf{0}$ gilt:

$$\hat{\boldsymbol{u}}_r'\hat{\boldsymbol{u}}_r - \hat{\boldsymbol{u}}_{ur}'\hat{\boldsymbol{u}}_{ur} = \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)' X'X\left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)$$

Wegen

$$\hat{\boldsymbol{\beta}}_r = \hat{\boldsymbol{\beta}}_{ur} - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)$$

gilt also

$$\begin{aligned}
&\hat{\boldsymbol{u}}_r'\hat{\boldsymbol{u}}_r - \hat{\boldsymbol{u}}_{ur}'\hat{\boldsymbol{u}}_{ur} \\
= \;& \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right)' X'X \left(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_{ur}\right) \\
= \;& \left((X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)\right)' \\
& \cdot X'X(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right) \\
= \;& \left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)' [R(X'X)^{-1}R']^{-1}R(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right) \\
= \;& \left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)' [R(X'X)^{-1}R']^{-1}\left(R\hat{\boldsymbol{\beta}}_{ur} - \mathbf{r}\right)
\end{aligned}$$