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

Zusammenfassung Kapitel 3 – Annahmen:

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

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

MLR 3 lineare Unabhängigkeit der Regressoren

$$rk(\mathbf{X}) = K + 1,\; n > K$$

MLR 4 strikte Exogenität der Regressoren

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

MLR 5 serielle Unkorreliertheit und Homoskedastizität

$$\Sigma(\mathbf{u}) = \sigma^2 \mathbf{I}_n$$

Zusammenfassung Kapitel 3 – Resultate:

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

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

$+$ **MLR 5** $\Rightarrow$ Gauß-Markov: $\hat{\boldsymbol{\beta}}$ BLUE.

$$\hat{\sigma}^2 = \frac{1}{n-K-1}\hat{\mathbf{u}}'\hat{\mathbf{u}}, \; E[\hat{\sigma}^2] = \sigma^2$$

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

$$se(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sqrt{\widehat{\Sigma}(\hat{\boldsymbol{\beta}}|\mathbf{X})} = \hat{\sigma}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}}$$

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

Stichprobenverteilung der OLS-Schätzer

Statistische Inferenz im Regressionsmodell

- Hypothesen an Parameter des Regressionsmodells
- Benute OLS-Schätzer, um Hypothesen zu überprüfen
- t-Test
- Konfidenzintervalle
- F-Test
- χ^2-Test

Nur möglich bei Verteilungsannahmen an die Zufallsvariablen des Modells.

Stichprobenverteilung von $\hat{\boldsymbol{\beta}}$

Wahrscheinlichkeitsverteilung der OLS-Schätzer $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$?

Zwei Möglichkeiten:

- Annahme über die Verteilung der Fehlerterme $\rightarrow$ Verteilung des Schätzers
- Große Stichprobe $\rightarrow$ asymptotische Verteilung des Schätzers.

Die zweite Möglichkeit verfolgen wir in dieser Vorlesung nicht. Sie setzt eine große Stichprobe und eine asymptotische Betrachtung voraus.

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

Stichprobenverteilung von $\hat{\beta}$

Annahme MLR 6

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

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

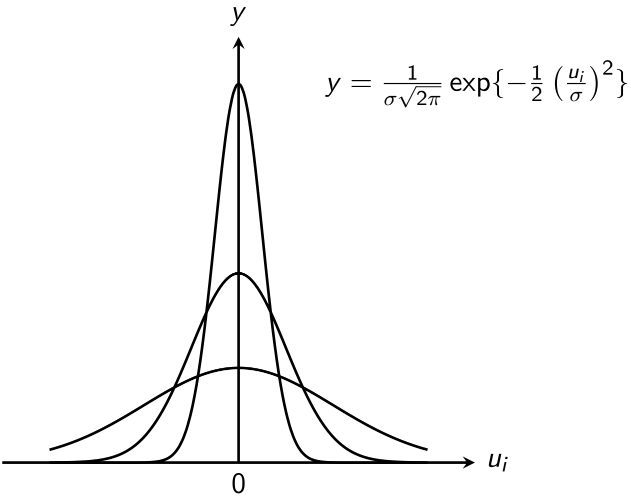
bzw.

$$\mathbf{u}|\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}_n\right)$$

MLR 6 impliziert direkt:

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

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



Diskussion der Normalverteilungsannahme

Fehlerterm: viele unbeobachtbaren Faktoren

Zentraler Grenzwertsatz:

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

Probleme:

- Anzahl verschiedener Faktoren groß genug?
- Individuelle Faktoren zu heterogen?
- 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 $\hat{\boldsymbol{\beta}}$ unter MLR 1 , MLR 3 und MLR 6

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

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

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

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

Verteilung von $\hat{\beta}_j$

Wegen

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

gilt für jeden einzelnen Schätzer

$$\hat{\beta}_j|\mathbf{X} \sim \mathcal{N}\left(\beta_j, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}_{(j,j)}\right) \text{ , wobei}$$

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

Bemerkung:

Im Gegensatz zu $\mathbf{u}$, welches mit **MLR 6** unabhängig normalverteilt ist, sind zwei einzelne Schätzer $\hat{\beta}_j$ und $\hat{\beta}_s$ nicht unabhängig. Da die Matrix $(\mathbf{X}'\mathbf{X})^{-1}$ 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 $\hat{\beta}_j$

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

Mit

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

und

$$Var\left[\left.\frac{\hat{\beta}_j - E[\hat{\beta}_j|\mathbf{X}]}{\sqrt{Var(\hat{\beta}_j|\mathbf{X})}}\right|\mathbf{X}\right] = \frac{Var(\hat{\beta}_j|\mathbf{X})}{Var(\hat{\beta}_j|\mathbf{X})} = 1$$

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

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

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 $\hat{\beta}_j$,

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

durch die geschätzte Varianz

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

und definieren die Test-Statistik

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

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:

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

Die t-Verteilung

Seien $x \sim \mathcal{N}(0,1)$ und $y \sim \chi^2_n$ 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 statt der Quantile der t_n-Verteilung die Quantile der $\mathcal{N}(0,1)$-Verteilung verwendet werden.

Die t-Statistik

Theorem

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

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

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

Ohne Beweis.

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.

$\rightarrow$ Nullhypothese:

$$H_0 : \beta_j = \gamma$$

Sind die Daten „kompatibel“ mit H_0?

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

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)

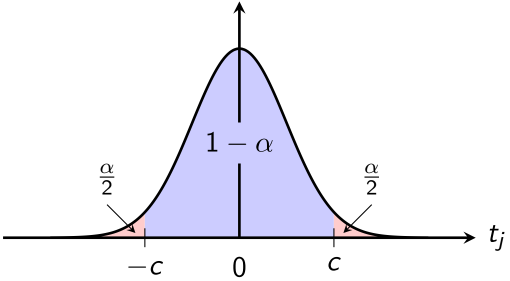
Gebräuchlich sind 10%, 5% oder 1% (also $\alpha = 0.1,\ 0.05,\ 0.01$)

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

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

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

Dichte der t-Verteilung unter H_0



Wie bestimmen wir den kritischen Wert c?

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.

Die entsprechende Tabelle ist auch der Klausur angehängt.

Beispiel: Bestimmungsfaktoren von College GPA

Datensatz: GPA1.xls, $n = 141$

Modell 1: KQ, benutze die Beobachtungen 1–141
Abhängige Variable: colGPA

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	1,38955	0,331554	4,191	0,0000
hsGPA	0,411816	0,0936742	4,396	0,0000
ACT	0,0147202	0,0105649	1,393	0,1658
skipped	−0,0831131	0,0259985	−3,197	0,0017

Mittel abhängige Var.	3,056738	Stdabw. abhängige Var.	0,372310
Summe quad. Residuen	14,87297	Stdfehler Regression	0,329487
R^2	0,233593	Korrigiertes R^2	0,216811
$F(3, 137)$	13,91875	P-Wert(F)	5,65e–08
Log-Likelihood	−41,50070	Akaike-Kriterium	91,00140
Schwarz-Kriterium	102,7964	Hannan–Quinn	95,79450

Beispiel: Bestimmungsfaktoren von College GPA

Datensatz: GPA1.xls, $n = 141$

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

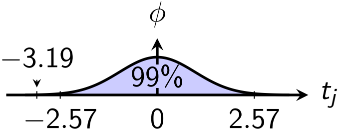
skipped: verpasste Kurse pro Woche

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

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?

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



Eine nützliche Faustregel für zweiseitige t-Tests

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

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 $se(\hat{\beta}_j) = \sqrt{\widehat{Var}(\hat{\beta}_j|\mathbf{X})}$.

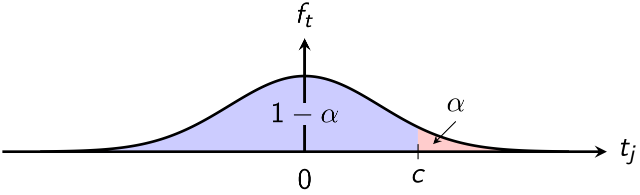
Einseitiger t-Test

Einseitige Nullhypothese

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

Gegenhypothese $H_1: \ \beta_j > \gamma$

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



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

Beispiel: Die Lohngleichung

wage1.xls, $n = 526$

Modell 2: KQ, benutze die Beobachtungen 1–526
Abhängige Variable: lwage

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	0,284359	0,104190	2,729	0,0066
educ	0,0920290	0,00732992	12,56	0,0000
exper	0,00412111	0,00172328	2,391	0,0171
tenure	0,0220672	0,00309365	7,133	0,0000

Mittel abhängige Var.	1,623268	Stdabw. abhängige Var.	0,531538
Summe quad. Residuen	101,4556	Stdfehler Regression	0,440862
R^2	0,316013	Korrigiertes R^2	0,312082
$F(3, 522)$	80,39092	P-Wert(F)	9,13e–43
Log-Likelihood	−313,5478	Akaike-Kriterium	635,0956
Schwarz-Kriterium	652,1568	Hannan–Quinn	641,7758

Beispiel: Die Lohngleichung

wage1.xls, $n = 526$

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

$$\widehat{log}(wage) = \underset{(0{,}104)}{0,284} + \underset{(0{,}007)}{0,092} educ + \underset{(0{,}0017)}{0.0041} exper + \underset{(0{,}003)}{0,022} tenure$$

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

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

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

- $\alpha = 1\%$: $c_{0{,}01} = 2,3335$
- $\alpha = 5\%$: $c_{0{,}05} = 1,6478$
- $\alpha = 10\%$: $c_{0{,}10} = 1,2832$

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

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

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 $t_j = \frac{\hat{\beta}_j - \gamma}{\sqrt{\widehat{Var}(\hat{\beta}_j|\mathbf{X})}} < -c$ oder falls $t_j > c$.

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

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

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

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

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 $n = 32$

Modell 3: KQ, benutze die Beobachtungen 1–32
Abhängige Variable: lrd

	Koeffizient	Std. Fehler	t-Quotient	p-Wert
const	$-4{,}37835$	0,468013	$-9{,}355$	0,0000
lsales	1,08423	0,0601941	18,01	0,0000
profmarg	0,0216594	0,0127820	1,695	0,1009

Mittel abhängige Var.	3,602825	Stdabw. abhängige Var.	1,734352
Summe quad. Residuen	7,650206	Stdfehler Regression	0,513615
R^2	0,917958	Korrigiertes R^2	0,912300
$F(2, 29)$	162,2384	P-Wert(F)	1,79e–16
Log-Likelihood	$-22{,}50998$	Akaike-Kriterium	51,01996
Schwarz-Kriterium	55,41717	Hannan–Quinn	52,47751

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:

$$c_{0.01} = 2.576,\ c_{0.05} = 1.96,\ c_{0.10} = 1.645$$

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

Zusammenhang von Konfidenzintervallen und Hypothesentests:

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

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, $n = 6763$

Mögliche Teststatistik:

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})}$$

Problem:

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

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

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

Restriktion $\beta_1 = \beta_2$, $\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2|\mathbf{X})$?

Der Term $\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2|\mathbf{X}) = \hat{\sigma}\left(\mathbf{X}'\mathbf{X}\right)^{-1}_{(1,2)}$ müsste eigentlich separat ausgerechnet werden, um die Teststatistik für $\hat{\beta}_1 - \hat{\beta}_2$ zu bestimmen.

Zwei alternative Lösungsansätze:

- Obere Schranke für $\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})$
  ($\rightarrow$ klappt nicht immer)
- Alternative Definition der Parameter & Regressoren
  ($\rightarrow$ klappt immer)

Obere Schranke für $\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})$

Für zwei beliebige Zufallsvariablen X und Y gilt:

$$Cov(X, Y) \geq -\sqrt{Var(X)}\sqrt{Var(Y)}$$

Daher folgt:

$$\begin{aligned} Var(X - Y) &= Var(X) - 2Cov(X, Y) + Var(Y) \\ &\leq Var(X) + 2\sqrt{Var(X)}\sqrt{Var(Y)} + Var(Y) \\ &= \left(\sqrt{Var(X)} + \sqrt{Var(Y)}\right)^2 \ , \end{aligned}$$

also:

$$\sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})} \leq \sqrt{\widehat{Var}(\hat{\beta}_1|\mathbf{X})} + \sqrt{\widehat{Var}(\hat{\beta}_2|\mathbf{X})}$$

$$\sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})} \leq \sqrt{\widehat{Var}(\hat{\beta}_1|\mathbf{X})} + \sqrt{\widehat{Var}(\hat{\beta}_2|\mathbf{X})}$$

Für die falsche und richtige Teststatistik $\tilde{t}$ und t gilt:

$$|\tilde{t}| = \frac{|\hat{\beta}_1 - \hat{\beta}_2|}{\sqrt{\widehat{Var}(\hat{\beta}_1|\mathbf{X})} + \sqrt{\widehat{Var}(\hat{\beta}_2|\mathbf{X})}} \leq \frac{|\hat{\beta}_1 - \hat{\beta}_2|}{\sqrt{\widehat{Var}(\hat{\beta}_1 - \hat{\beta}_2|\mathbf{X})}} = |t|$$

Falls $|\tilde{t}| > c$, falls also die Nullhypothese mit $\tilde{t}$ abgelehnt würde, dann gilt auch $|t| > c$.

H_0 kann also abgelehnt werden, ohne dass $\widehat{Cov}(\hat{\beta}_1, \hat{\beta}_2|\mathbf{X})$ berechnet werden muss.

Dieser Ansatz funktioniert aber nicht, falls $|\tilde{t}| < c$.

Alternative Definition der Parameter & Regressoren

Für diesen Ansatz ersetzen wir einen der beiden Regressoren von β_1 und β_2 durch die Summe dieser beiden Regressoren.

Hierdurch verändert sich die Interpretation der Parameter des Modells.

Konkret können wir einen Parameter des modifizierten Modells als $\beta_1 - \beta_2$ interpretieren.

Ein Test auf $\beta_1 - \beta_2 = 0$ ist dann direkt möglich.

Beispiel: Lohn bei verschiedenen Bildungsabschlüssen

twoyear.xls, $n = 6763$

Betrachte Modell:

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

Subtrahiere und addiere $\beta_2 jc$:

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

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

$$\mathbf{y} = \beta_0 \left(1 - \frac{R_0}{R_j}\mathbf{x}_j\right) + \ldots + \theta \cdot x_j + \ldots + \beta_K \cdot \left(\mathbf{x}_K - \frac{R_K}{R_j}\mathbf{x}_j\right) + \mathbf{u}$$

Berechne nun $\hat{\theta}$, $\widehat{Var}(\hat{\theta}|\mathbf{X})$ und $t = \frac{\hat{\theta}-\gamma}{\sqrt{\widehat{Var}(\hat{\theta}|\mathbf{X})}}$ 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:

$$\left.\begin{array}{ccccccccc} R_{10} \cdot \beta_0 & + & R_{11} \cdot \beta_1 & + & \dots & + & R_{1K} \cdot \beta_K & = & r_1 \\ R_{20} \cdot \beta_0 & + & R_{21} \cdot \beta_1 & + & \dots & + & R_{2K} \cdot \beta_K & = & r_2 \\ & & & & \vdots & & & & \\ R_{J0} \cdot \beta_0 & + & R_{J1} \cdot \beta_1 & + & \dots & + & R_{JK} \cdot \beta_K & = & r_J \end{array}\right. \Leftrightarrow \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

mit

- $\mathbf{R}$: Matrix der Ordnung $J \times K + 1$
- $J \leq K$
- $rk(\mathbf{R}) = J$ (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:

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

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

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

Beispiel: Determinanten der Gehälter von Baseball-Spielern

Die drei Restriktionen in Matrixschreibweise:

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

Wir können diese Restriktion nun durch $\mathbf{R}\hat{\boldsymbol{\beta}}$ 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, $n = 353$, $SSR = 198,311$, $R^2 = 0,5971$

Restringiertes Modell:

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

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:

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

Es bezeichne der Index u das unrestringierte und r das restringierte Modell.

Wir definieren die F-Statisik:

$$F = \frac{(SSR_r - SSR_u)/J}{SSR_u/(n - K - 1)}$$

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_u kaum (wobei $SSR_r \geq SSR_u$ gelten muss). Große Werte von F sprechen also gegen H_0.

Die F-Statistik

$$F = \frac{(SSR_r - SSR_u)/J}{SSR_u/(n-K-1)}$$

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 $n - K - 1 = 347$, $c_{1\%} = 3,78$)

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

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

Die F-Verteilung

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:

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

Wald-Statistik

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

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

Wir können zeigen, dass:

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

und wir können zeigen, dass:

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

Zusammenfassung F-Test

Nullhypothese

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

Teststatistik

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

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

Die Anwendung des F-Tests

$$F = \frac{(SSR_r - SSR_u)/J}{SSR_u/(n - K - 1)}$$

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

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

$H_0 :\ \beta_j = 0$

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

$$\underline{\mathbf{R} = (\mathbf{0}, \mathbf{I}_K),\ \mathbf{r} = \mathbf{0} \text{ (mit } J = K)}$$

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

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

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

Der Ausdruck

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

vereinfacht sich zu

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

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

Vorbereitung globaler F-Test

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

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

Wir schreiben die Formel der F-Statistik etwas um:

$$F = \frac{(SSR_r - SSR_u)/K}{SSR_u/(n-K-1)} = \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_u}{SST}\right)/K}{\frac{SSR_u}{SST}/(n-K-1)}$$

Teststatistik globaler F-Test

Unter der Restriktion

$$\beta_1 = \beta_2 = \ldots = \beta_K = 0$$

gilt $y_i = \beta_0 + u_i$ für $i = 1, \ldots, n$ mit $\hat{\beta}_0 = \bar{y}$.

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

Mit $\hat{u}_i^2 = (y_i - \bar{y})^2$ gilt dann $SSR_r = SST$.

Die F-Statistik des globalen F-Tests reduziert sich daher zu:

$$F = \frac{\left(\frac{SSR_r}{SST} - \frac{SSR_u}{SST}\right)/K}{\frac{SSR_u}{SST}/(n-K-1)} = \frac{\left(1 - \frac{SSR_u}{SST}\right)/K}{\frac{SSR_u}{SST}/(n-K-1)} = \frac{R_u^2/K}{(1-R_u^2)/(n-K-1)}$$

Beispiel: Bildung der Eltern und Geburtsgewicht der Kinder

bwght.xls n=1191

$$bwght = \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 faminc + \beta_4 motheduc + \beta_5 fatheduc + u$$

- *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

Unrestringiertes Modell:

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

$$\begin{aligned}
SSR_u &= 464041,1 \\
R_u^2 &= 0,038748 \\
F_{5,1185} &= 9,553500
\end{aligned}$$

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:

$$\widehat{bwght} = \underset{(1{,}6559)}{115,470} - \underset{(0{,}10877)}{0,597852}\,\text{cigs} + \underset{(0{,}65754)}{1,83227}\,\text{parity} + \underset{(0{,}032394)}{0,0670618}\,\text{famin}$$

$$SSR_r = 465166,8$$
$$R_r^2 = 0,036416$$

Es gilt nun

$$F = \frac{(SSR_r - SSR_u)/2}{SSR_u/1185} = 1,436$$

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:

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

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

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

$$\begin{aligned} H_1 \quad : \quad & \beta_j = \beta_j + 0 \text{ für } i \in I_1 \\ & \beta_j = \beta_j + \gamma_j \text{ für } i \in I_2 \text{ mit } \gamma_j \neq 0 \text{ für mindestens ein } j \end{aligned}$$

Nullhypothese:
Parameter des Modells sind in beiden Gruppen gleich.

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

Test auf Strukturbruch: Chow-Test

Zusätzliche Regressoren z_j für $j = 1, \ldots, K$:

$$
\begin{aligned}
z_{ij} &= 0 \text{ für } i \in I_1 \\
z_{ij} &= x_{ij} \text{ für } i \in I_2
\end{aligned}
$$

Zusätzliche Parameter $\gamma_1, \ldots, \gamma_K$.

Neues Modell:

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

Die Ablehnung der Nullhypothese $\gamma_1 = \ldots = \gamma_K = 0$ 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

- $\hat{\boldsymbol{\beta}}$ ist unter **MLR 6** normalverteilt.
- $(n-K-1)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}$ ist unter **MLR 6** χ^2_{n-K-1}-verteilt.
- Die t-**Statistik** $\frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}(\hat{\beta}_j|\mathbf{X})}}$ 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 $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ mit Daten vereinbar ist

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

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

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

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