Übung zu Kapitel 4: Multiple lineare Regression: Inferenz

Aufgabe 1

Lineare Transformationen von normalverteilten Zufallsvariablen sind ebenfalls normalverteilt.

Zeige, dass unter den Annahmen MLR 1, MLR 3 und MLR 6 folgendes gilt:

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

$$\frac{1}{\sigma}\mathbf{u} \sim \mathcal{N}(0,I)$$

Aufgabe 2

Die Projektionsmatrix $I - X(X'X)^{-1}X'$ ist idempotent symmetrisch.

a) Zeige dies.

b) Zeige, dass $\frac{\hat{\sigma}^2(n-k-1)}{\sigma^2} = \frac{1}{\sigma}\mathbf{u}'[I - X(X'X)^{-1}X']\frac{1}{\sigma}\mathbf{u}$.

Aufgabe 3

Sei A eine quadratische, idempotente und symmetrische Matrix. Dann entspricht der Rang von A der Spur von A: $rk(A) = tr(A)$.

Sei zusätzlich der Zufallsvektor $\mathbf{v}$ standard normalverteilt: $\mathbf{v} \sim \mathcal{N}(0, I)$. Dann ist die quadratische Form $\mathbf{v}'A\mathbf{v}$ χ^2-verteilt mit $tr(A)$ Freiheitsgraden.

Zeige, dass $\frac{1}{\sigma}\mathbf{u}'[I - X(X'X)^{-1}X']\frac{1}{\sigma}\mathbf{u} \sim \chi^2_{(n-k-1)}$.

Aufgabe 4

Zeige schließlich, dass:

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

Welche Eigenschaft ist nun noch zu begründen, um zu zeigen, dass die Test-Statistik

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

t-verteilt mit $n-k-1$ Freiheitsgraden ist?

Aufgabe 5

Benutze *CEOSAL1.xls* für diese Aufgabe.

Beachte folgendes ökonometrisches Modell:

$$\ln(salary) = \beta_0 + \beta_1 \ln(sales) + \beta_2 roe + \beta_3 ros + u$$

wobei *ros* „return on firm's stock" ist.

1. Stelle die Nullhypothese H_0 dafür auf, dass *ros* keinen Effekt auf das Gehalt hat. ($\rightarrow$ zweiseitiger Test)
   Stelle die alternative Nullhypothese $\tilde{H}_0$ dafür auf, dass *ros* keinen positiven Effekt auf das Gehalt hat ($\rightarrow$ einseitiger Test)
2. Schätze nun das obige Modell. Um wie viel Prozent steigt das prognostizierte Gehalt, wenn *ros* um 50 Punkte steigt? Hat *ros* einen großen praktischen Effekt auf *salary*?
3. Teste H_0 und $\tilde{H}_0$ jeweils zum Signifikanz-niveau von 10%.
4. Würdest Du *ros* im finalen ökonometrischen Modell inkludieren?

$\boxed{H_0 : \beta_3 = 0}$

$\boxed{\tilde{H}_0 : \beta_3 \leq 0}$

$\hat{\beta}_3 \cdot 50 = 0{,}012$

$\Rightarrow$ Salary steigt um 0,012 %

$t_3 = \dfrac{\hat{\beta}_3 - 0}{se(\hat{\beta}_3)} = 0{,}4460$

$c_{t_{205};0{,}95} \approx 1{,}65$

$|t_3| < 1{,}65$

$\Rightarrow H_0$ nicht verwerfen.

t_3 (wie oben) $= 0{,}446$

$c_{t_{205};0{,}9} = 1{,}285$

$|t_3| < 1{,}285$

$\Rightarrow \tilde{H}_0$ nicht verwerfen

nein, β_3 ist nicht signifikant von null verschieden

Aufgabe 6

$X_i \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

Nehme an, es liegt eine Stichprobe von unabhängig und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$ vor. Gehe davon aus, dass es sich um eine Stichprobe von normalverteilten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 handelt.

Beschreibe detailliert die Vorgehensweise des t-Tests zur Überprüfung der Hypothese, dass der Erwartungswert gleich μ_0 ist und zwar *jeweils* für den zweiseitigen t-Test sowie für die beiden ein-seitigen t-Tests. Folgende Stichwörter sollten in der Beschreibung vorkommen:

Nullhypothese ✓ und Alternativhypothese ✓, Teststatistik ✓, kritischer Wert ✓, Stichprobenvarianz ✓, Testentscheidung ✓, Mittelwert ✓, Verteilung der Teststatistik unter der Nullhypothese ✓, Signifikanzniveau ✓

Zweiseitiger Test

Nullhypothese $H_0 : \mu = \mu_0$

Alternativhypothese $H_1 : \mu \neq \mu_0$

Stichprobe $(x_1, \dots, x_n)$

Teststatistik $$t = \frac{\bar{x} - \mu_0}{\sqrt{\widehat{Var}(\bar{x})}}$$ $Var(\bar{x}) = \frac{1}{n} \sigma^2$

Stichprobenmittel $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Stichprobenvarianz $\widehat{Var}(\bar{x}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Vert. von t unter H_0: t-Verteilung mit $n-1$ Freiheitsgraden

Signifikanzniveau $\alpha \in [0,1]$

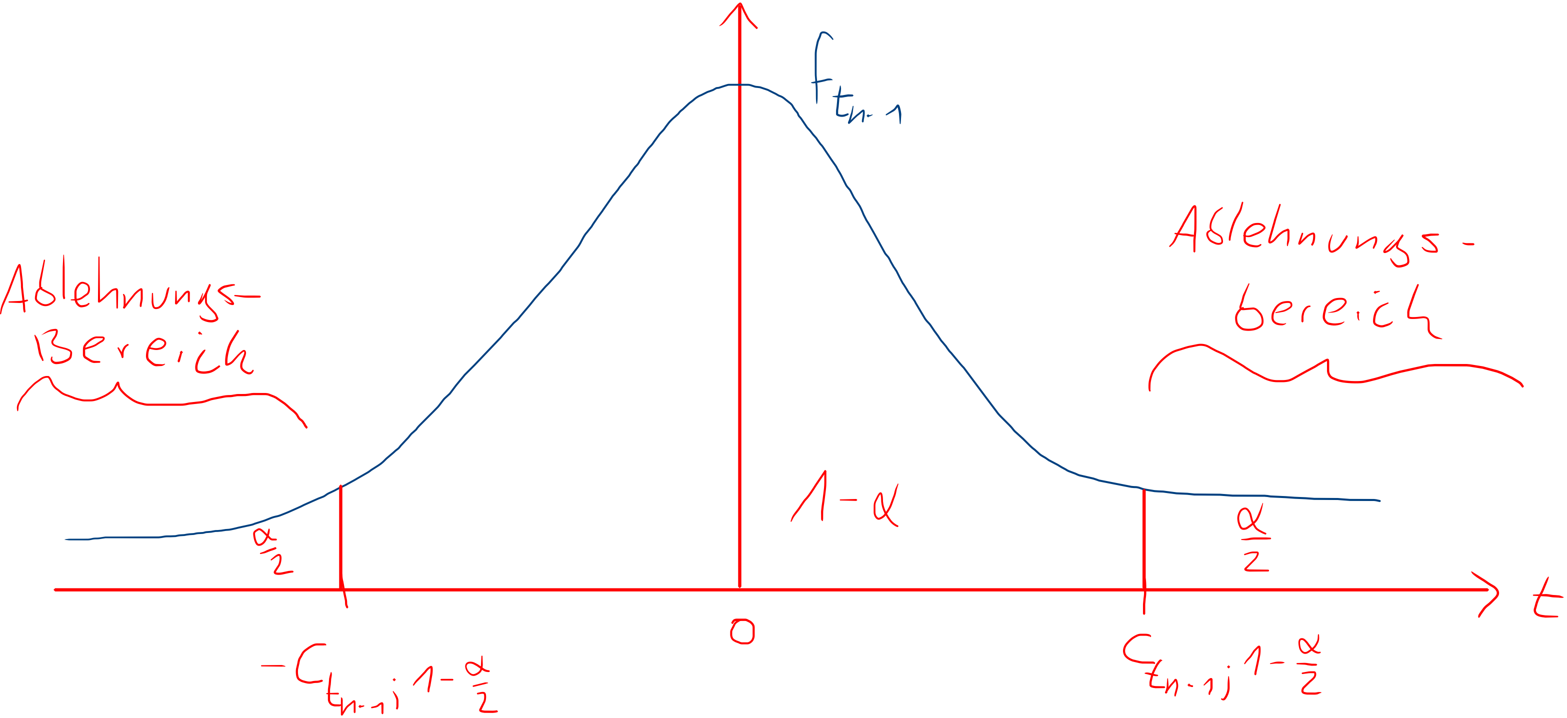
Kritischer Wert $c_{t_{n-1}; 1-\frac{\alpha}{2}}$ $\quad$ $\left(\mathbb{P}(t < c_{t_{n-1}; 1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \right.$

$\mathbb{P}(t > c_{t_{n-1}, 1-\frac{\alpha}{2}}) = \frac{\alpha}{2}$

Symmetrie $\left. \mathbb{P}(t < -c_{t_{n-1}, 1-\frac{\alpha}{2}}) = \frac{\alpha}{2} \right)$

Testentscheidung:

Lehne $H_0 : \mu = \mu_0$ ab, falls $|t| > c_{t_{n-1}; 1-\frac{\alpha}{2}}$.



$f_{t_{n-1}}$
Ablehnungs-Bereich
Ablehnungs-bereich
$1-\alpha$
$\frac{\alpha}{2}$
$\frac{\alpha}{2}$
t
$-c_{t_{n-1};1-\frac{\alpha}{2}}$
0
$c_{t_{n-1};1-\frac{\alpha}{2}}$

Einseitiger Test

Nullhypothese $H_0: \mu \leq \mu_0$

Teststatistik $t = \frac{\bar{x} - \mu_0}{\sqrt{\widehat{Var}(\bar{x})}}$ wie zuvor

Signifikanzniveau α wie zuvor

Vert. unter H_0: wie zuvor t_{n-1}

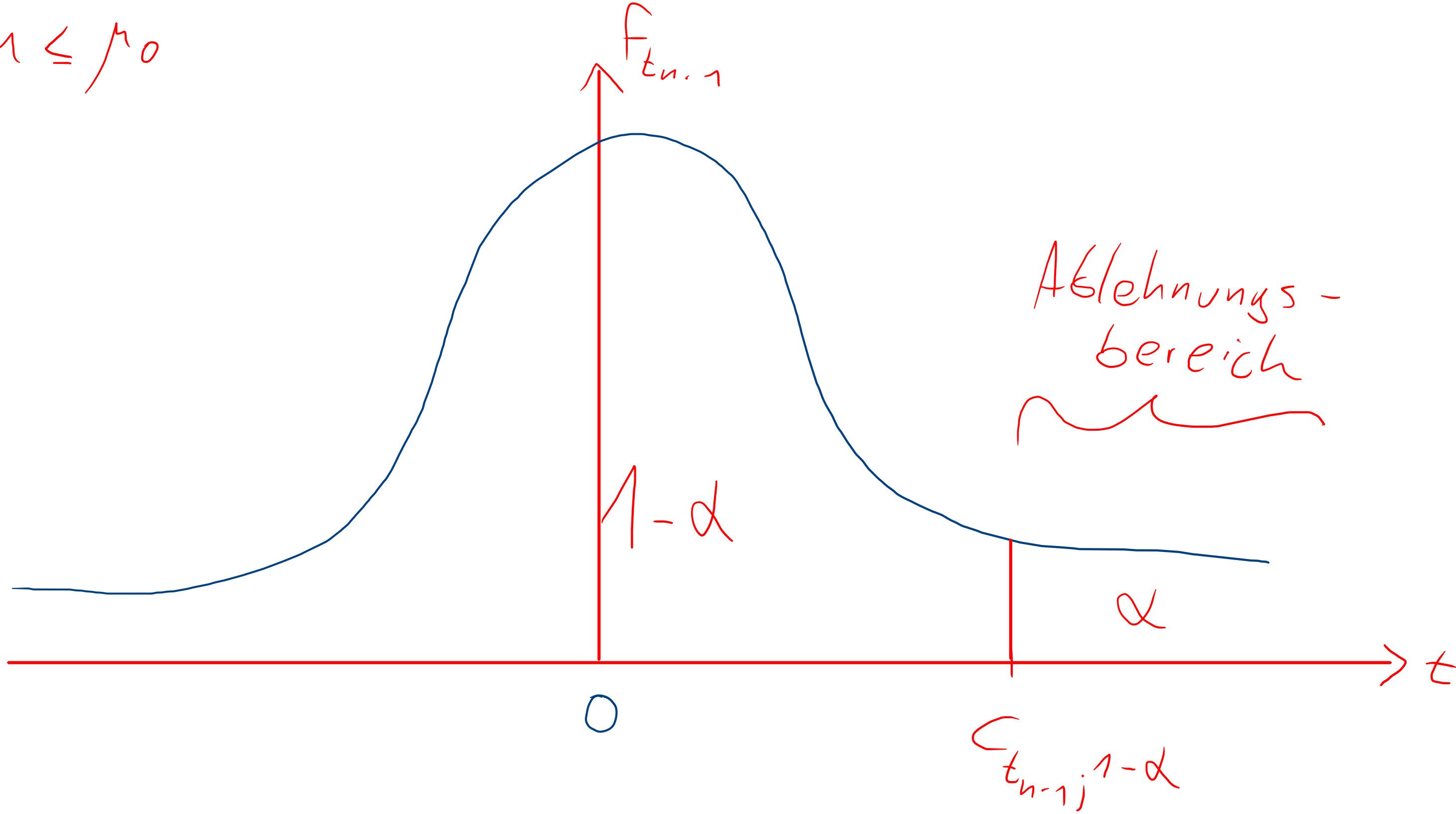
Kritischer Wert: $c_{t_{n-1};1-\alpha}$

$$\mathbb{P}(t < c_{t_{n-1};1-\alpha}) = 1-\alpha$$

$$\mathbb{P}(t > c_{t_{n-1};1-\alpha}) = \alpha$$

Testentscheidung: lehne H_0 ab, falls $t > c_{t_{n-1};1-\alpha}$

$H_0: \mu \leq \mu_0$



Einseitiger Test mit H_0 $\mu \geq \mu_0$

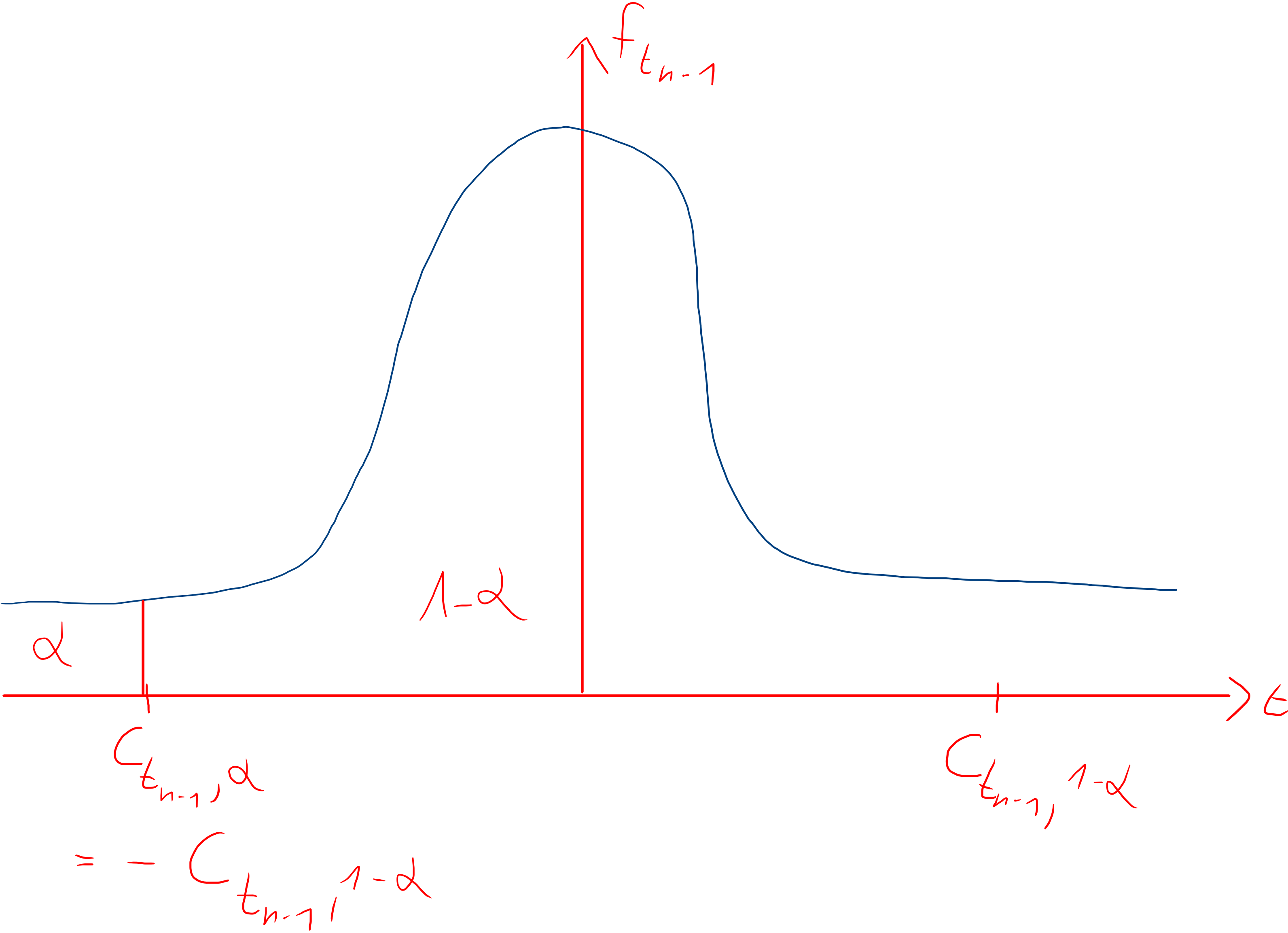
kritischer Wert $c_{t_{n-1};\alpha} = -c_{t_{n-1};1-\alpha}$

$$\mathbb{P}(t < c_{t_{n-1},\alpha}) = \alpha$$

$$\mathbb{P}(t > c_{t_{n-1},\alpha}) = 1-\alpha$$

Testentscheidung: Lehne H_0 ab, falls $t < c_{t_{n-1},\alpha}$

(falls $|t| > c_{t_{n-1},1-\alpha}$)


$f_{t_{n-1}}$
α
$1-\alpha$
t
$c_{t_{n-1},\alpha}$
$c_{t_{n-1},1-\alpha}$
$= -c_{t_{n-1},1-\alpha}$

Aufgabe 7

Die folgende Tabelle enthält die Jahresbestleistungen, die Leichtathletik Superstar Usain Bolt in den Jahren 2007 bis 2016 in der 100m-Disziplin erreicht hat:

Jahr:	2007	2008	2009	2010	2011
Zeit:	10,03s	9,69s	9,58s	9,82s	9,76s
Jahr:	2012	2013	2014	2015	2016
Zeit:	9,63s	9,77s	9,98s	9,79s	9,81s

$H_0: \mu \leq 9{,}75$

$H_1: \mu > 9{,}75$

$\alpha = 5\%$

Führe unter den obigen Annahmen einen einseitigen t-Test durch für die Nullhypothese, dass der Erwartungswert für die Jahresbestleistung unter 9,75 Sekunden liegt, gegen die Alternative, dass der Erwartungswert darüber liegt. Wähle dabei ein Signifikanzniveau von $\alpha = 5\%$.

$t = 0{,}8144 < c = 1{,}83$

Lehne H_0 nicht ab

$$\mathrm{Var}(\bar{x}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$

$$\overset{iid}{=} \frac{1}{n^2}\sum_{i=1}^{n} \underbrace{\mathrm{Var}(x_i)}_{=\sigma^2}$$

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

$$\hat{\sigma}^2 = \frac{1}{n-1}\hat{u}'\hat{u} = \frac{1}{n-1} SSR = \frac{1}{n-1}\sum_{i=1}^{n}\Big(\overbrace{x_i - \bar{x}}^{\hat{u}_i}\Big)^2$$

Aufgabe 8

Benutze für diese Aufgabe ~~die ersten 10~~ alle Beobachtungen des Datensatzes *sleep75.xls*.

Das zugehörige multiple lineare Regressionsmodell sei:

$$sleep = \beta_0 + \beta_1 age + \beta_2 earns74 + u_i.$$

$t_1 = \frac{\hat{\beta}_1 - 5}{se(\hat{\beta}_1)}$

$= \frac{3{,}77 - 5}{1{,}47} = -0{,}837$

$|t_1| < c$

$\Rightarrow$ lehne $H_0: \beta_1 = 5$ nicht ab.

a) Führe jeweils einen zweiseitigen sowie einen rechtsseitigen *t*-Test für die Nullhypothese

$$H_0: \beta_1 = 5 \quad \text{bzw.} \quad H_0: \beta_2 = 0$$

bei dem Signifikanzniveau $\alpha = 0.05$ (5 %) durch.

b) Überprüfe die Berechnungen in `gretl`.

$t_2 = -2{,}231$

$|t_2| > c$

$\Rightarrow$ Lehne $H_0: \beta_2 = 0$

kritischer Wert zweiseitiger Test $c_{t_{700}, 0.975} = 1{,}96$

$H_0: \beta_1 \geq 5$

$H_0 \; \beta_2 \geq 0$

$c_{t_{700}, 0{,}05} = -c_{t_{700}, 0.95} = -1{,}645$

$t > -1{,}645 \Rightarrow$ lehne $H_0 \; \beta_1 \geq 5$ nicht ab

$t < -1{,}645$

Lehne $H_0 \; \beta_2 \geq 0$ ab.

für Stichprobe 1, ..., 10:

$\hat{\beta}_1 = 3,23$ $\quad se(\hat{\beta}_1) = 12,2$

$\hat{\beta}_2 = -0,01$ $\quad se(\hat{\beta}_2) = 0,008$

$$se(\hat{\beta}_j) = \sqrt{\hat{var}(\hat{\beta}_j)}$$

$H_0: \beta_1 = 5$

$$t_1 = \frac{\hat{\beta}_1 - 5}{se(\hat{\beta}_1)} = \frac{3,23-5}{12,2} = \frac{-1,77}{12,2} = -0,145$$

$c_{t_7; 1-0,025} = 2,365$ $\quad |t_1| < c \Rightarrow H_0$ nicht verw.

$c_{t_7; 1-0.05} = 1,895$ $\quad t_1 < c \Rightarrow H_0: \beta_1 \geq 5$ nicht verw.

$t_1 > -c \Rightarrow H_0\ \beta_1 \leq 5$ nicht ver.

$\hat{\beta}_2 = -0.01 \qquad se(\hat{\beta}_2) = 0.008$

$H_0 \quad \beta_2 = 0 \qquad H_0 \quad \beta_2 \lessgtr 0$

$$t_2 = \frac{\hat{\beta}_2 - 0}{se(\hat{\beta}_2)} = \frac{-0.01}{0.008} = -\frac{10}{8} = -\frac{5}{4} = -1{,}25$$

$c^{II} = 2{,}3 \qquad c^{I} = 1{,}8$

H_0 nicht verwerfen

Aufgabe 9

Das Schulministerium beauftragt dich, herauszufinden, ob eine Verringerung der Klassengrößen in den Schulen des Landes zu besseren (durch entsprechende Tests gemessenen) schulischen Leistungen führt. Dir liegt eine Zufallsstichprobe im Umfang von $n = 184$ mit Daten über die klassendurchschnittliche Punktzahl bei Leistungstests y_i und die Schülerzahl in der Klasse x_i vor. Das OLS Regressionsergebnis ist $\hat{y}_i = 574.3 - 2.28x_i$, wobei der Standardfehler der Konstanten 10.4 und derjenige des Steigungsparameters 0.52 beträgt, sowie $R^2 = 0.0955$ ist.

a) Formuliere die der oben diskutierten Frage entsprechende Null- und Alternativhypothese.

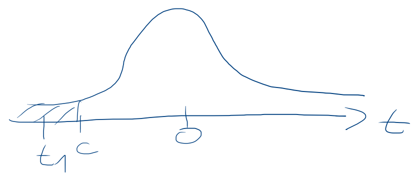
b) Führe den entsprechenden Test durch, wobei du alle Schritte und die dafür nötigen Annahmen darlegen solltest. Erläutere das Testergebnis.

c) Das Ministerium weist Deinen Bericht mit dem Hinweis zurück, wegen des geringen R^2 seien die Ergebnisse ohnehin nicht aussagekräftig. Nehme dazu Stellung.

$$\hat{Y}_i = \underset{(10{,}4)}{\overset{\hat{\beta}_0}{574{,}3}} - \underset{(0{,}52)}{\overset{\hat{\beta}_1}{2{,}28}}\, x_i \qquad n = 184 \quad R^2 = 0{,}0955$$

a) $H_0 : \beta_1 \geq 0 \qquad H_1 : \beta_1 < 0$

b) MLR 1 – 6



$$t_1 = \frac{\hat{\beta}_1 - 0}{se(\hat{\beta}_1)} \sim t_{182}$$

$$= \frac{-2{,}28}{0{,}52} = -4{,}38 \qquad \alpha = 1\%$$

$$c_{t_{182};1\%} = -c_{t_{182};99\%} \approx -2{,}35$$

$t_1 < c \Rightarrow H_0$ verwerfen

c) Das Testergebnis ist hoch signifikant, erklärt aber nur 10% der Variation von y. Omitted variable bias?

Aufgabe 10

Siehe Aufgabe 5

Benutze für diese Aufgabe den Datensatz *ceosal1.xls*.

Betrachte wieder das ökonometrische Modell

$$\ln(salary) = \beta_0 + \beta_1 \ln(sales) + \beta_2 roe + \beta_3 ros + u$$

Teste die Hypothese $H_0 : \beta_3 \leq 0$ zum Signifikanzniveau $\alpha = 10\%$.

Aufgabe 11

Folgende Gleichung ist das Resultat einer OLS-Schätzung auf Basis des Datensatzes *rdchem.xls*:

Level $\quad \hat\beta_0 \quad \hat\beta_1 \quad \hat\beta_2$ Log

$$\widehat{rdintens} = \underset{(1.676)}{0.470} + \underset{(0.046)}{0.050} \cdot profmarg + \underset{(0.216)}{0.321} \cdot lsales \; (n = 32)$$

%

$t_1 = \frac{0.05}{0.046} \approx 1 \qquad t_1 < c \qquad H_0: \beta_1 = 0$ nicht ablehnen

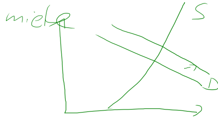
- Interpretiere den Schätzer für *lsales*. Falls *sales* um 10% ansteigt, was ist der prognostizierte Anstieg von *rdintens*? ⇝ $\Delta rdintens = 0,321 \cdot 10 = 3,21$
- Teste die Hypothese, dass *rdintens* in *lsales* sinkt zu den Signifikanzniveaus 5% und 10%. $\quad H_0: \beta_2 \le 0$
- Hat *profmarg* einen statistisch signifikanten Effekt auf *rdintens*?

$t_2 = \frac{0,321}{0,216} = 1,486$, $\quad c_{t_{29}, 95\%} = 1,7$

$c_{t_{29}; 90\%} = 1,3$

$t_2 < c_{t_{29}, 95\%} \Rightarrow$ Lehne $H_0: \beta_2 \le 0$ zu $\alpha = 5\%$ nicht ab

$t_2 > c_{t_{29}, 90\%} \Rightarrow$ Lehne H_0 $\beta_2 \le 0$ zu $\alpha = 10\%$ ab



Aufgabe 12

$t_3 = \frac{\hat{\beta}_3 - 0}{se(\hat{\beta}_3)}$

Benutze für diese Aufgabe den Datensatz *rental.xls*.

Betrachte das ökonometrische Modell

$= \frac{66}{12} = 5,5$

$$lrent = \beta_0 + \beta_1 lpop + \beta_2 lavginc + \beta_3 pctstu + u$$

$c = 2,617$

$t_3 > c \Rightarrow$ H_0 verwerfen

- Formuliere die Hypothese, dass der Anteil der Studierenden keinen c.p.-Effekt auf die Mieten hat. $H_0 \quad \beta_3 = 0$
- Welche Vorzeichen erwartest du für β_1 und β_2? $\beta_1 > 0, \ \beta_2 > 0$
- Die OLS-Schätzung liefert folgende Gleichung:

oder pcstu + 10%

$$\overset{\text{Log}}{\widehat{lrent}} = \underset{(0.46)}{-3.37} + \underset{(0.027)}{0.031} \cdot \overset{\text{Log}}{lpop} + \underset{(0.041)}{0.877} \cdot lavginc + \underset{(0.0012)}{0.0066} \cdot pctstu \ (n = 128)$$

rent $+ 10 \cdot 0,0066$

$= 0,066$

$= 6,6\%$

Was ist an folgender Aussage falsch: „Eine Steigerung von 10% der Bevölkerung (*pop*) ist verbunden mit einer Steigerung von 6.6% der Mieten“? Die 6,6 ist falsch; es sollte

- Teste die anfangs aufgestellte Hypothese zum Signifikanzniveau $\alpha = 1\%$.

Aufgabe 13

$t_2 = \frac{0,4118 - 0,4}{0,0937} = 0,126$

$c = 1,98 \Rightarrow H_0$ nicht verwerfen

Benutze für diese Aufgabe den Datensatz *GPA1.xls* und betrachte folgende OLS-Schätzung:

β_2

$$\widehat{\text{colGPA}} = \underset{(0,3316)}{1,3896} - \underset{(0,0260)}{0,0831}\,\text{skipped} + \underset{(0,0937)}{0,4118}\,\text{hsGPA} + \underset{(0,0106)}{0,0147}\,\text{ACT}$$

$$n = 141 \quad \bar{R}^2 = 0,2168 \quad F(3,137) = 13,919 \quad \hat{\sigma} = 0,32949$$

(Standardfehler in Klammern)

$t_2 = \frac{0,4118 - 1}{0,0937}$

$= -6,28$

$c = 1,98$

$|t_2| > c$

H_0 verwerfen

- Kannst du $H_0 : \beta_{hsGPA} = 0.4$ zugunsten $H_1 : \beta_{hsGPA} \neq 0.4$ zum Signifikanzniveau $\alpha = 5\%$ verwerfen?
- Kannst du $H_0 : \beta_{hsGPA} = 1$ zugunsten $H_1 : \beta_{hsGPA} \neq 1$ zum Signifikanzniveau $\alpha = 5\%$ verwerfen?

Aufgabe 14

Betrachte das lineare Regressionsmodell unter den Annahmen MLR 1 bis MLR 6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Du möchtest die Nullhypothese $H_0 : \beta_1 - 3\beta_2 = 1$ testen.

- Es seien $\hat{\beta}_1$ und $\hat{\beta}_2$ die OLS-Schätzer für β_1 und β_2. Bestimme $var(\hat{\beta}_1 - 3\hat{\beta}_2)$ in Ausdrücken von $var(\hat{\beta}_1)$, $var(\hat{\beta}_2)$ und $cov(\hat{\beta}_1, \hat{\beta}_2)$. Wie lautet der Standardfehler von $\hat{\beta}_1 - 3\hat{\beta}_2$?
- Bestimme die t-Statistik für die Hypothese $H_0 : \beta_1 - 3\beta_2 = 1$.
- Definiere $\theta_1 = \beta_1 - 3\beta_2$ und $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$. Stelle eine Regressionsgleichung mit β_0, θ_1, β_2 und β_3 auf, welches Dir erlaubt $\hat{\theta}_1$ und den Standardfehler von $\hat{\theta}_1$ abzulesen.

$$\mathrm{Var}(\hat\beta_1 - 3\hat\beta_2) = \mathrm{Var}(\hat\beta_1) + 2\,\mathrm{Cov}(\hat\beta_1, -3\hat\beta_2) + \mathrm{Var}(3\hat\beta_2)$$

^{2}

$$= \mathrm{Var}(\hat\beta_1) - 6\,\mathrm{Cov}(\hat\beta_1, \hat\beta_2) + 9\,\mathrm{Var}(\hat\beta_2)$$

$$t = \frac{\hat\beta_1 - 3\hat\beta_2 - 1}{se(\hat\beta_1 - 3\hat\beta_2)}$$

$$se(\hat\beta_1 - 3\hat\beta_2) = \sqrt{\widehat{\mathrm{Var}}(\hat\beta_1 - 3\hat\beta_2)}$$

$$\Theta_1 = \beta_1 - 3\beta_2$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$- 3\beta_2 x_1 + 3\beta_2 x_1$$

$$= \beta_0 + \underbrace{(\beta_1 - 3\beta_2)}_{\Theta_1} x_1 + \beta_2 \underbrace{(x_2 + 3x_1)}_{Z} + \beta_3 x_3 + u$$

$$Z = 3x_1 + x_2$$

$$y = \beta_0 + \Theta_1 x_1 + \beta_2 Z + \beta_3 x_3 + u$$

$$\Rightarrow \hat\Theta_1, \; se(\hat\Theta_1) \Rightarrow t_1 = \frac{\hat\Theta_1 - 1}{se(\hat\Theta_1)}$$

Aufgabe 15

Regressionsmodell

$x_i = \mu + u_i \quad ; \; u_i \sim N(0, \sigma^2)$

$\hat{\mu} = \hat{\beta}_0 = \overset{\beta_0}{\overline{x}}$

Nehme an, es liegt eine Stichprobe von unabhängig und identisch verteilten Zufallsvariablen $X_1, \ldots, X_n$ vor. Gehe davon aus, dass es sich um eine Stichprobe von normalverteilten Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 handelt.

normal

Stelle die zweiseitigen sowie einseitigen (auf dem arithmetischen Mittel $\hat{\mu}$ basierenden) Konfidenzintervalle für den Parameter μ auf und erläutere den Zusammenhang mit der Testentscheidung beim t-Test.

zweiseitiges $KI = [\overline{x} - c \cdot se(\overline{x}) ; \overline{x} + c \cdot se(\overline{x})]$

$se(\overline{x}) = \sqrt{\widehat{var}(\overline{x})}$

$Var(\overline{x}) = \frac{1}{n}\sigma^2 \Rightarrow \widehat{var}(\overline{x}) = \frac{1}{n}\hat{\sigma}^2$

$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ $\quad \left(\hat{\sigma}^2 = \frac{1}{n-k-1} SSR\right)$

$se(\overline{x}) = \sqrt{\frac{1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$

Lehne H_0 $\mu = \mu_0$ ab, falls $\mu_0 \in KI$

$c_{t_{n-1}, 1-\frac{\alpha}{2}} \leadsto c$

einseitiges $KI = (-\infty, \overline{x} + \tilde{c} \cdot se(\overline{x})]$

$\tilde{c}_{t_{n-1}; 1-\alpha} \leadsto \tilde{c}$

Aufgabe 16

Benutze für diese Aufgabe die ersten 10 Beobachtungen des Datensatzes *sleep*75.*xls*.

Das zugehörige multiple lineare Regressionsmodell sei:

$$sleep = \beta_0 + \beta_1 age + \beta_2 earns74 + u_i.$$

Im Folgenden ist das Signifikanzniveau $\alpha = 0.05$ zu wählen.

Berechne jeweils ein zweiseitiges symmetrisches sowie ein rechtsseitiges $(1-\alpha)$-Konfidenzintervall für den Parameter β_1 bzw. β_2 und formuliere die entsprechenden Null- und Alternativhypothesen.

Entscheide ohne einen t-Test durchzuführen anhand der zuvor berechneten Konfidenzintervalle, ob die Nullhypothese jeweils abgelehnt wird oder nicht. Erläutere in diesem Zusammenhang warum Konfidenzintervalle auch „Nichtverwerfungsregionen“ genannt werden.

Aufgabe 17

Benutze für diese Aufgabe den Datensatz *GPA1.xls* und betrachte folgende OLS-Schätzung:

$$\widehat{\text{colGPA}} = \underset{(0,3316)}{1,3896} - \underset{(0,0260)}{0,0831}\,\text{skipped} + \underset{(0,0937)}{0,4118}\,\text{hsGPA} + \underset{(0,0106)}{0,0147}\,\text{ACT}$$

$$n = 141 \quad \bar{R}^2 = 0,2168 \quad F(3,137) = 13,919 \quad \hat{\sigma} = 0,32949$$

(Standardfehler in Klammern)

Benutze die Standard-Normalverteilung und berechne ein 95%-Konfidenzintervall für β_{hsGPA}.

Aufgabe 18

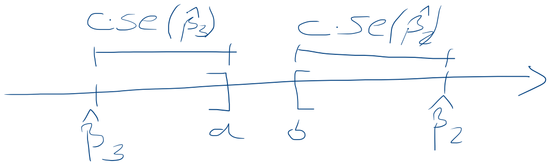
Benutze *wage2.xls* für diese Aufgabe.

Betrachte das ökonometrische Modell:

$$\ln(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

$$\beta_0 + \beta_1 educ + (\beta_2 - \beta_3) exper + \beta_3 (tenure + exper) + u$$

- Postuliere die Nullhypothese, dass ein Jahr Arbeitserfahrung bei einem beliebigen Arbeitgeber den gleichen Effekt auf den Lohn hat, wie ein Jahr Beschäftigungsdauer beim aktuellen Arbeitgeber. $H_0: \beta_2 = \beta_3 \iff \beta_2 - \beta_3 = 0$
- Teste diese Nullhypothese gegen eine zweiseitige Alternative zum 5%-Signifikanzniveau in dem du ein 95%-Konfidenzintervall bildest. Was ist Deine Schlussfolgerung?



$$a < b \iff \hat\beta_3 + c \cdot se(\hat\beta_3) < \hat\beta_2 - c\, se(\hat\beta_2)$$

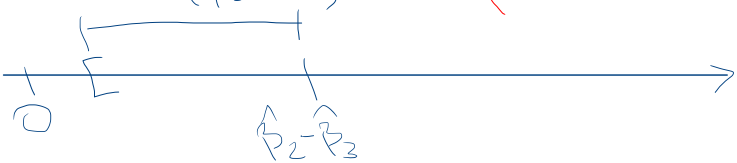
$$\iff 0 < \hat\beta_2 - \hat\beta_3 - c\left(se(\hat\beta_2) + se(\hat\beta_3)\right)$$

$$\implies 0 < \hat\beta_2 - \hat\beta_3 - c \cdot se(\hat\beta_2 - \hat\beta_3)$$

$$KI_\alpha = \left[\hat\beta_j - c \cdot se(\hat\beta_j)\,;\ \hat\beta_j + c \cdot se(\hat\beta_j)\right]$$

$$t_{FG;1-\frac{\alpha}{2}} =: c$$

$$c \cdot se(\hat\beta_2 - \hat\beta_3) < c \cdot \left(se(\hat\beta_2) + se(\hat\beta_3)\right)$$



$$se(\hat{\beta}_2 - \hat{\beta}_2) = \sqrt{\widehat{Var}(\hat{\beta}_2 - \hat{\beta}_3)}$$

$$Var(\hat{\beta}_2 - \hat{\beta}_3) = Var(\hat{\beta}_2) - 2cov(\hat{\beta}_2, \hat{\beta}_3) + Var(\hat{\beta}_3)$$

$$-\sqrt{Var(\hat{\beta}_2)Var(\hat{\beta}_3)} < cov(\hat{\beta}_2, \hat{\beta}_3) < \sqrt{Var(\hat{\beta}_2)Var(\hat{\beta}_3)}$$

Falls $cov(\hat{\beta}_2, \hat{\beta}_3) = -\sqrt{Var(\hat{\beta}_2)Var(\hat{\beta}_3)}$

$$\Rightarrow Var(\hat{\beta}_2 - \hat{\beta}_3) = \sqrt{Var(\hat{\beta}_2)}^2 - 2\left(-\sqrt{Var(\hat{\beta}_2)Var(\hat{\beta}_3)}\right) + \sqrt{Var(\hat{\beta}_3)}^2$$

$$= \left(\sqrt{Var(\hat{\beta}_2)} + \sqrt{Var(\hat{\beta}_3)}\right)^2$$

$$se(\hat{\beta}_2 - \hat{\beta}_3) = \sqrt{\widehat{Var}(\hat{\beta}_2 - \hat{\beta}_3)} = \sqrt{\widehat{Var}(\hat{\beta}_2)} + \sqrt{\widehat{Var}(\hat{\beta}_3)}$$

$$= se(\hat{\beta}_2) + se(\hat{\beta}_3)$$

$\hat{\beta}_2 = 0{,}0153285$

$\hat{\beta}_3 = 0{,}0133748$

$\hat{\beta}_2 - \hat{\beta}_3 = 0{,}0019537$

$se(\hat{\beta}_2) = 0{,}00336957$

$se(\hat{\beta}_3) = 0{,}00258720$

$FG = 935 - 4 = 931$

$\alpha = 5\%$

$c_{t_{931},1-\frac{\alpha}{2}} \approx 1{,}97$

KI obere Grenze für $\hat{\beta}_3$ V_o^3

untere Grenze für $\hat{\beta}_2$ V_u^2

$V_o^3 = \hat{\beta}_3 + c \cdot se(\hat{\beta}_3)$

$= 0{,}013 + 1{,}97 \cdot 0{,}0026$

$= 0{,}0185 > \hat{\beta}_2$

$V_u^2 = \hat{\beta}_2 - c\, se(\hat{\beta}_2)$

$= 0{,}015 + 1{,}97 \cdot 0{,}0034$

$= 0{,}00869 < V_o^3$

$\hat\beta_2 - \hat\beta_3 = 0{,}00195$ $\qquad H_0: \beta_2 - \beta_3 = 0$

$se(\hat\beta_2 - \hat\beta_3) = 0{,}004\cancel{7}4$

$$t = \frac{\hat\beta_2 - \hat\beta_3 - 0}{se(\hat\beta_2 - \hat\beta_3)} = 0{,}4119 < 1{,}96 = c$$

$\Rightarrow$ H_0 nicht verwerfen

Aufgabe 19

Benutze *gpa*1.*xls* für diese Aufgabe.

- Erkläre *colGPA* durch *PC* (PC-Besitz), *hsGPA* und *ACT* und bilde ein 95%-Konfidenzintervall für β_{PC}. Ist der geschätzte Koeffizient statistisch signifikant zum 5%-Niveau gegen die zweiseitige Alternative?

- Diskutiere die statistische Signifikanz der Schätzungen für $\beta_{hsGP\cancel{G}A}$ und β_{ACT}. Ist *hsGPA* oder *ACT* wichtiger für die Prognose von *colGPA*?

  hsGPA ist wichtig (***)
  ACT ist insignifikant (P=41%)

- Füge die beiden Indikatoren *fathcoll* und *mothcoll* zur Regression hinzu. Ist einer von beiden signifikant?

$$\hat{\beta}_{PC} = 0{,}157$$

$$se(\hat{\beta}_{PC}) = 0{,}057$$

$$c = 1{,}98$$

$$KI_{PC,\alpha} = [0{,}157 - 1{,}98 \cdot 0{,}057 ;\ 0{,}157 + 1{,}98 \cdot 0{,}057]$$

$$= [\underbrace{\cancel{0{,}270}\ 0{,}044}_{>0} ;\ 0{,}270]$$

Verwerfe $H_0: \beta_{PC} = 0$, da $0 \notin KI$.

$H_0 \quad \beta_{fcoll} = 0, \quad \beta_{mcoll} = 0$

F-Statistik

$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/135}$$

\# Gleichungen

\# FG

$$= \frac{(R^2_{ur} - R^2_r)/2}{(1 - R^2_{ur})/135}$$

$SSR_{ur} = 15{,}094 \qquad SSR_r = 15{,}14868$

$$F = \frac{(15{,}14868 - 15{,}094)/2}{15{,}094/135} = \frac{0{,}02734}{0{,}1118}$$

$$= 0{,}2445$$

$c_{F_{2,135},1-\alpha} \approx 3{,}07$

$F < c$

H_0 nicht verwerfen

Aufgabe 20

Betrachte erneut den Datensatz `sleep75`. Dieser enthält neben dem Alter (age) und der Schlafdauer (sleep) unter anderem auch Informationen über die Anzahl der Bildungsjahre (educ) und die wöchentliche Gesamtarbeitszeit, gemessen in Minuten (totwrk) von 706 Arbeitnehmern.

a) Schätze die beiden Modelle

$$\mathrm{sleep}_i = \beta_1 + \beta_2 \mathrm{totwrk}_i + \varepsilon_i \tag{1}$$

$$\mathrm{sleep}_i = \beta_1 + \beta_2 \mathrm{totwrk}_i + \beta_3 \mathrm{educ}_i + \beta_4 \mathrm{age}_i + \varepsilon_i \tag{2}$$

mittels OLS.

b) Teste in Modell (2) separat, ob educ bzw. age statistisch signifikant von Null verschieden sind. Verwende ein Signifikanzniveau von 5%.

c) Teste nun anhand eines F-Tests, ob educ und age <u>gemeinsam</u> signifikant (von Null verschieden) sind. Formuliere dazu H_0 und H_1 und verwende auch hier ein Signifikanzniveau von 5%.

Aufgabe 21

Du hast das Regressionsmodell

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

auf Basis von $n = 53$ Beobachtungen mittels OLS geschätzt und erhältst $\hat{\beta}_0 = 9.23$, $\hat{\beta}_1 = 1$, $\hat{\beta}_2 = -1$ sowie $SSR = 100$. Die geschätzte Varianz-Kovarianzmatrix von $\hat{\boldsymbol{\beta}}$ ist gegeben durch

$$\hat{\sigma}^2 (X'X)^{-1} = \begin{pmatrix} 5 & 0.5 & -1.5 \\ 0.5 & 1 & -1 \\ -1.5 & -1 & 2 \end{pmatrix}$$

Teste jeweils zum Signifikanzniveau $\alpha = 5\%$, ob $\hat{\beta}_1$ und $\hat{\beta}_2$ einzeln bzw. gemeinsam signifikant sind.

Aufgabe 22

Wie in der Vorlesung hergeleitet gilt

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

Im Folgenden wird ein einfaches lineares Regressionsmodell der Form $y = \beta_0 + \beta_1 x + u$ unterstellt.

- Gebe $\hat{\boldsymbol{\beta}}_r$ für $R = diag(1,1)$ und $\mathbf{r} = (1,2)'$ an.

Die Momentematrix sei gegeben durch:

$$X'X = \begin{pmatrix} 10 & 386 \\ 386 & 15440 \end{pmatrix}$$

- Gebe $\hat{\boldsymbol{\beta}}_r$ für $R = (0,1)$ und $r = 3$ in Abhängigkeit von $\hat{\boldsymbol{\beta}}_{ur}$ an.
- Bestimme zudem den Stichprobenumfang, den Mittelwert $\bar{x}$ und die Stichprobenvarianz $\overline{xx} - \bar{x}\bar{x}$.

Aufgabe 23

r-Modell

$\widehat{price} = 0 + 1 \cdot assess$

$u_r = price - \widehat{price}$

$= price - assess$

$SSR_r = \sum_{i=1}^{n} (price_i - assess_i)^2$

Benutze *hprice1.xls* für diese Aufgabe.

Im einfachen Regressionsmodell

$$price = \beta_0 + \beta_1 assess + u$$

ist die Bewertung *assess* rational, falls $\beta_0 = 0$ und $\beta_1 = 1$.

Die geschätzte Gleichung ist

Ur-Modell

$$\underset{(16{,}273)}{\widehat{price} = -14{,}4718} + \underset{(0{,}049365)}{0{,}975554} assess$$

$t_1 = \frac{\hat{\beta}_1 - 1}{se(\hat{\beta}_1)} = -0{,}495$

$|t_1| < 1$

H_0 $\beta_1 = 1$ nicht verwerfen

mit $n = 88$, $SSR_{ur} = 165644{,}5$ und $R^2_{ur} = 0{,}819531$. Ferner gilt $\sum_{i=1}^{n}(price_i - assess_i)^2 = 209448{,}99$. $= SSR_r$

- Teste H_0 $\beta_0 = 0$ und H_0 $\beta_1 = 1$ und H_0 $\beta_0 = 0, \beta_1 = 1$

$\left|\frac{\hat{\beta}_0}{se(\hat{\beta}_0)}\right| < 1$ H_0 nicht verwerfen

$H_0 \quad \beta_0 = 0 \, , \, \beta_1 = 1$

$$F = \frac{(209488{,}95 - 165644{,}5)/2}{165644{,}5/86}$$

$$= \frac{21922{,}245}{1926{,}10} = 11{,}382$$

$$C = F_{2,86,1-\alpha} \approx 4{,}85$$

$F > C \Rightarrow H_0 \quad \beta_0 = 0 \;\&\; \beta_1 = 1$

verwerfen!

Aufgabe 24

r-Model:

$\Rightarrow$ benutze R^2 aus A23

$R_r^2 = 0{,}819531$

$Price = \beta_0 + \beta_1 \, assess + u$

Benutze wieder *hprice1.xls* für diese Aufgabe.

Betrachte das Regressionsmodell

$$price = \beta_{\cancel{1}\,0} + \beta_1 assess + \beta_2 lotsize + \beta_3 sqrtft + \beta_4 bdrms + u$$

Das Bestimmtheitsmaß dieser Regression lautet $R^2_{ur} = 0,829205$.

- Teste $H_0: \ \beta_2 = \beta_3 = \beta_4 = 0$.
- Wie bewertest du diesen F-Test, falls die Varianz von *price* von einem oder mehreren der Regressoren abhängt?

\# Gleichungen

$$F = \frac{(R^2_{ur} - R^2_r)/3}{(1 - R^2_{ur})/83}$$

$88 - 5 = 76$

Aufgabe 25

Benutze *ceosal2.xls* mit $n = 177$ für diese Aufgabe.

In den folgenden drei Regressionsmodellen ist jeweils *lsalary* der Regressand:

	Modell 1	Modell 2	Modell 3
$\hat{\beta}_0$	$4,96108$ (0,19996)	$4,62069$ (0,25434)	$4,57198$ (0,25347)
lsales	$0,224279$ (0,027129)	$0,158483$ (0,039814)	$0,187787$ (0,040003)
lmktval	–	$0,112261$ (0,050393)	$0,0998716$ (0,049214)
profmarg	–	$-0,00225878$ (0,0021654)	$-0,00221093$ (0,0021054)
ceoten	–	–	$0,0171035$ (0,0055399)
comten	–	–	$-0,00923770$ (0,0033373)
R^2	0,280858	0,303494	0,352537

noch Aufgabe 25

Prüfe per F-Test, ob eine Reduktion des Modells

- von Modell 3 auf Modell 2
- von Modell 3 auf Modell 1
- von Modell 2 auf Modell 1

sinnvoll ist?

Welches Problem ergibt sich?

Sollte noch eine andere Reduktion geprüft werden?

Aufgabe 26

$$VoteA = \beta_0 + \underbrace{(\beta_1+\beta_2)}_{\theta} LexpA + \beta_2 (LexpB - LexpA) + \beta_3\, prtystrA + u$$

Benutze den Datensatz *vote*1.*xls* für diese Aufgabe.

Betrachte folgendes Regressionsmodell:

$$votaA = \beta_0 + \beta_1 \ln(expendA) + \beta_2 \ln(expendB) + \beta_3 prtystrA + u$$

$$+\beta_2 \ln(expendA) - \beta_2 \ln(expendA)$$

- Wie interpretierst du β_1? Falls expendA um 1% steigt, dann steigt voteA um β_1
- Stelle die Nullhypothese auf, dass eine 1% Steigerung der Ausgaben von A durch eine 1% Steigerung der Ausgaben von B in Bezug auf *votaA* ausgeglichen wird. $H_0: \beta_1 = -\beta_2 \Leftrightarrow \beta_1 + \beta_2 = 0$
- Schätze das Modell. Beeinflussen die Ausgaben von A das Ergebnis *voteA*? Und die Ausgaben von B? Kannst du diese Resultate~~t~~ benutzen um die Nullhypothese zu testen?
- Schätze ein Modell, welches die *t*-Statistik zu diesem Test direkt liefert. Wie lautet Deine Testentscheidung?

$$\mathrm{Var}(\hat\beta_1+\hat\beta_2) = \sigma_1^2 + 2\sigma_{12} + \sigma_2^2$$

$$\mathrm{Var}(\hat\beta_1) = \sigma_1^2$$

$$\mathrm{Var}(\hat\beta_2) = \sigma_2^2$$

$$\mathrm{Cor}(\hat\beta_1,\hat\beta_2) = \sigma_{12}$$

$$-\sigma_1\sigma_2 < \sigma_{12} < \sigma_1\sigma_2$$

$$\Rightarrow \mathrm{Var}(\hat\beta_1+\hat\beta_2) \begin{array}{l} > \sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2 = (\sigma_1-\sigma_2)^2 \\ < \sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2 = (\sigma_1+\sigma_2)^2 \end{array}$$

$$\sqrt{\mathrm{Var}(\hat\beta_1+\hat\beta_2)} < \sigma_1+\sigma_2$$

$$\sqrt{\widehat{\mathrm{Var}}(\hat\beta_1+\hat\beta_2)} = se(\hat\beta_1+\hat\beta_2) < se(\hat\beta_1)+se(\hat\beta_2)$$

$$> |se(\hat\beta_1)-se(\hat\beta_2)|$$

$H_0 \quad \beta_1 + \beta_2 = 0$

t- Statistik : $\frac{\hat{\beta}_1 + \hat{\beta}_2 - 0}{se(\hat{\beta}_1 + \hat{\beta}_2)} = t_{1+2}$

$|se(\hat{\beta}_1) - se(\hat{\beta}_2)| < se(\hat{\beta}_1 + \hat{\beta}_2) < se(\hat{\beta}_1) + se(\hat{\beta}_2)$

$t_{1+2} > \frac{\hat{\beta}_1 + \hat{\beta}_2}{se(\hat{\beta}_1) + se(\hat{\beta}_2)} = \frac{6{,}08 - 6{,}62}{0{,}38 + 0{,}38} = \frac{-0{,}54}{0{,}76} = -0{,}71$

$t_{1+2} < \frac{\hat{\beta}_1 + \hat{\beta}_2}{|se(\hat{\beta}_1) - se(\hat{\beta}_2)|} = \frac{6{,}08 - 6{,}62}{0{,}3822 - 0{,}3788} = \frac{-0{,}54}{0{,}0034} = -159$

$-159 < t_{1+2} < -0{,}71$

Abschätzungen sind nicht hilfreich

neues Modell mit $\theta = \beta_1 + \beta_2$

$H_0: \theta = 0$ lexpB - lexpA $\rightsquigarrow \beta_2$

t-Statistik: $t = \frac{\hat{\theta} - 0}{se(\hat{\theta})} = \frac{-0,5321}{0,533} = -0,9982$

$c_{t_{165, 1-\frac{\alpha}{2}}} = c_{t_{165, 0,95}} = 1,65 > |t|$

mit $\alpha = 10\%$

$\Rightarrow$ H_0 nicht verwerfen!

Aufgabe 27

Generiere mit Gretl folgende Zufallsdaten für 100 Beobachtungen:

- $x = randgen(u, 0, 100)$
- $d = randgen(b, .5, 1)$
- $x1 = d * x$
- $x2 = (1 - d) * x$
- $u = randgen(n, 0, 25)$
- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, wobei du $\beta_0, \beta_1, \beta_2$ mit $\beta_1 \neq \beta_2$ selbst aussuchst.

Zeige die Daten x und y in einem Streudiagramm.

Benutze die Daten x, y und d um auf einen Strukturbruch anhand eines Chow-Tests zu testen.

Restringierte Modell ($H_0: \beta_1 = \beta_2 \iff \beta_1 - \beta_2 = 0$)

$$y = \beta_0 + \beta_1 x + u \leadsto \begin{matrix} SSR_r \\ R_r^2 \end{matrix}$$

Unrestringiertes Modell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

Chow Test: $\leadsto \underbrace{1 \cdot d}_{=d}, \quad \underbrace{x \cdot d}_{= x_1}$

$$y = \beta_0 + \beta_1 x + \gamma_1 \cdot d + \gamma_2 x_1 + u \leadsto \begin{matrix} SSR_{ur} \\ R_{ur}^2 \end{matrix}$$

$H_0: \gamma_1 = \gamma_2 = 0$

$$F = \frac{(SSR_r - SSR_{ur})/2}{SSR_{ur}/(100-4)} = \frac{(2174000 - 47000)/2}{47000/96} = \frac{1063500}{490} = 2172$$

$$C_{F_{2,96;\,1-\alpha}} = 4{,}8 < F$$

α: 1%

$\Rightarrow H_0$ verwerfen