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:

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

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

Aufgabe 2

a) Zeige, dass mit $\mathbf{M}_X = I - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

$$(n-k-1)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{\sigma^2}\mathbf{u}'\mathbf{M}_X\mathbf{u}$$

Sei $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ und $\mathbf{A}$ idempotent symmetrisch. Dann ist die quadratische Form $\mathbf{v}'\mathbf{A}\mathbf{v}$ χ^2-verteilt mit $tr(\mathbf{A})$ Freiheitsgraden.

b) Zeige:

$$\left(\frac{\hat{\sigma}\sqrt{n-K-1}}{\sigma}\right)^2 \sim \chi^2_{n-K-1}$$

Aufgabe 3

Zeige:

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

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

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\widehat{Var}(\hat{\beta}_j|X)}}$$

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

Aufgabe 4

Benutze *CEOSAL*1.*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?

Aufgabe 5

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

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 $\alpha = 5\%$.

Aufgabe 6

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

Das zugehörige multiple lineare Regressionsmodell sei:

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

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

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

bei dem Signifikanzniveau $\alpha = 0.05$ durch.

Aufgabe 7

Benutze für diese Aufgabe die ersten 10 Beobachtungen des Datensatzes *sleep75.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 8

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.

Aufgabe 9

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

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

- Interpretiere den Schätzer für *lsales*. Falls *sales* um 10% ansteigt, was ist der prognostizierte Anstieg von *rdintens*?
- Teste die Hypothese, dass *rdintens* in *lsales* sinkt zu den Signifikanzniveaus 5% und 10%.
- Hat *profmarg* einen statistisch signifikanten Effekt auf *rdintens*?

Aufgabe 10

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

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

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

(Standardfehler in Klammern)

- 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 11

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 12

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.

Aufgabe 13

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.

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.

Aufgabe 14

Benutze *wage2.xls* für diese Aufgabe.

Betrachte das ökonometrische Modell:

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

- 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.
- Teste diese Nullhypothese gegen eine zweiseitige Alternative zum 5%-Signifikanzniveau in dem du ein 95%-Konfidenzintervall bildest. Was ist Deine Schlussfolgerung?

Aufgabe 15

Benutze *gpa1.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 β_{hsGPA} und β_{ACT}. Ist *hsGPA* oder *ACT* wichtiger für die Prognose von *colGPA*?
- Füge die beiden Indikatoren *fathcoll* und *mothcoll* zur Regression hinzu. Ist einer von beiden signifikant?

Aufgabe 16

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 17

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 18

Wie in der Vorlesung hergeleitet gilt

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

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 $\hat{\boldsymbol{\beta}}_u$ an.
- Bestimme zudem den Stichprobenumfang, den Mittelwert $\bar{x}$ und die Stichprobenvarianz $\overline{xx} - \bar{x}\bar{x}$.

Aufgabe 19

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

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

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

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

Aufgabe 20

Benutze wieder *hprice*1*.xls* für diese Aufgabe.

Betrachte das Regressionsmodell

$$price = \beta_1 + \beta_1 assess + \beta_2 lotsize + \beta_3 sqrtft + \beta_4 bdrms + u$$

Das Bestimmtheitsmaß dieser Regression lautet $R^2 = 0,829205$.

- 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?

Aufgabe 21

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 21

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 22

Benutze den Datensatz *vote1.xls* für diese Aufgabe.

Betrachte folgendes Regressionsmodell:

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

- Wie interpretierst du 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.
- Schätze das Modell. Beeinflussen die Ausgaben von A das Ergebnis *voteA*? Und die Ausgaben von B? Kannst du diese Resultatet benutzen um die Nullhypothese zu testen?
- Schätze ein Modell, welches die t-Statistik zu diesem Test direkt liefert. Wie lautet Deine Testentscheidung?

Aufgabe 23

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.