Prof. Dr. Mitja Stadje
Björn Kriesche, M.Sc.

SS 2015
2. 6. 2015

Ökonometrie - Übungsblatt 4

Abgabe am 16. 6. **vor** Beginn der Übung

- Am Dienstag, dem 9. Juni findet statt der Übung eine Vorlesung statt.
- Am Dienstag, dem 16. Juni findet dann statt der Vorlesung die Besprechung des 4. Übungsblattes statt. Gleichzeitig wird das 5. Übungsblatt veröffentlicht und eine Woche später (am 23. Juni) besprochen.

Aufgabe 1 (1+4+1,5+1,5+3 Punkte)
Die folgende Aufgabe soll ohne die Verwendung von **R** bearbeitet werden.
Von 8 Angestellten wurden der Stundenlohn in € und die Berufserfahrung in Jahren dokumentiert. Zusätzlich gibt eine Dummy-Variable an, ob die betreffenden Angestellten überdurchschnittlich gut aussehen. Es wird angenommen, dass folgender Zusammenhang gilt: $log(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

Stundenlohn (y) in €	überdurchschnittliches Aussehen (x_1)	Berufserfahrung (x_2) in Jahren
11,4	1	30
8,5	0	28
16	1	35
23,1	0	38
22,8	0	27
7,8	0	20
17,6	0	12
9,2	1	20

(a) Welche der folgenden Fälle könnten dazu führen, dass die für Tests der Regressionsparameter betrachteten $t-$Statistiken nicht länger $t-$verteilt sind:
   (i) Auftreten von Heteroskedastizität,
   (ii) die Stichprobenkorrelation zwischen zwei unabhängigen Variablen beträgt 0,95,
   (iii) eine wichtige erklärende Variable wurde weggelassen.

(b) Bestimme die Regressionsparameter und interpretiere ihre Bedeutung in Worten.

(c) Schätze die Varianz der Störterme.

(d) Teste zum Niveau $\alpha = 0,01$ ob man durch überdurchschnittlich gutes Aussehen einen Gehaltsvorteil erreichen kann.

(e) Teste zum Niveau $\alpha = 0,05$ die allgemeine Signifikanz des Modells, d.h. die Nullhypothese $H_0 : \beta_1 = \beta_2 = 0$. Berechne die Testgröße auf zwei verschiedenen Wegen: (1) als Funktion des Bestimmtheitsmaßes und (2) als Spezialfall des Tests für mehrere Linearkombinationen der Parameter, d.h. ohne das Bestimmtheitsmaß oder SSR zu verwenden.

Hinweis: Für $A \in \mathbb{R}^{2\times 2}$ gilt, dass

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

und $\det A = ad - bc$.
Eine Formel für die Inverse einer Matrix $A \in \mathbb{R}^{3\times 3}$ kann auf Übungsblatt 3 gefunden werden.

Aufgabe 2 (1,5+2+1+2+1,5 Punkte)
Die Datei `ergebnisse.txt` enthält folgende Daten über die Ergebnisse von 680 Studenten in einer Vorlesung:

- y = erzielte Punkte in der Klausur (Ergebnisse)
- x_1 = bisherige Durchschnittsnote im Studium (Gesamtnote)
- x_2 = erzielte Punkte im Zulassungstest der Universität (Zulassung)
- x_3 = Anteil der Vorlesungen bei denen der Student anwesend war (Anteil)
- x_4 = Indikator, ob es sich um einen Studenten im ersten Studienjahr handelt (Anfaenger)

Es wird folgender linearer Zusammenhang vermutet:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 \log x_2 + \beta_4 x_3 + \beta_5 x_1 x_3 + \beta_6 x_4$$

(a) Schätze die Regressionsparameter in **R** ohne die Funktion `lm` zu verwenden.

(b) Beschreibe für alle unabhängigen Variablen, welchen Einfluss eine Änderung der Variablen auf die erzielten Klausurpunkte hat. Falls das für manche Variablen nicht pauschal möglich ist, begründe warum.

(c) Ein Student hat eine Durchschnittsnote von 2,2 und war bei 85% der Vorlesungen anwesend. Wie würde sich sein erwartetes Prüfungsergebnis ändern, wenn er noch vor der Klausur seine Durchschnittsnote auf 2,0 herabsenken könnte?

(d) Teste in **R** die Nullhypothese $H_0 : \beta_4 + 2\beta_5 = 0, \beta_1 = \beta_3, \beta_0 = 10$ zum Niveau $\alpha = 0,01$.

(e) Führe in **R** einen White-Test zum Niveau $\alpha = 0,05$ durch, um zu überprüfen, ob in diesem Modell Heteroskedastizität vorliegt. Dabei soll eine Regression von $\hat{u}_i^2$ auf $\hat{y}_i$, $\hat{y}_i^2$ und $\hat{y}_i^3$ durchgeführt werden. Es darf die **R**-Funktion `lm` verwendet werden.
Hinweis: Möchte man in `lm` eine Potenz (z.B. Quadrat) einer Variablen `x` oder ein Produkt zweier Variablen `x` und `y` betrachten, so muss man `I(x^2)` bzw. `I(x*y)` anstatt nur `x^2` bzw. `x*y` verwenden.

Aufgabe 3 (1+3,5+1,5+2 Punkt)

Die Datei `schule.txt` enthält folgende Daten über 1.692 Schulen in Michigan:

- y = Anteil der Viertklässler (in %) die eine ausreichende Note in Mathe erreicht haben (Bestehen)
- x_1 = Anteil der Schüler (in %) die ein kostenloses Mittagessen erhalten haben (Lunch)
- x_2 = Anzahl der Schüler (Schueler)
- x_3 = Ausgaben (in \$) der Schule pro Schüler (Ausgaben)

Es wird folgender linearer Zusammenhang vermutet:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 \log x_3$$

(a) Welche der folgenden Probleme können durch das Auftreten von Heteroskedastizität entstehen?

(i) Die Schätzer für die Regressionsparameter sind inkonsistent.

(ii) Die üblicherweise verwendeten $F-$Statistiken sind nicht länger $F-$verteilt.

(iii) Der Schätzer für die Regressionsparameter ist nicht länger "BLUE" (best linear unbiased estimator).

(b) Schätze die Regressionsparameter in **R** mit Hilfe von `lm` und interpretiere β_1, β_2 und β_4 in Worten. Wie sollte man das lineare Modell ändern um β_1 und β_2 besser interpretieren zu können.

(c) Teste mit Hilfe des Breusch-Pagan-Tests zum Niveau $\alpha = 0,01$, ob in diesem linearen Modell Heteroskedastizität vorliegen könnte. (Es soll keine **R**-Funktion verwendet werden, die den Test automatisch durchführt, `lm` kann aber verwendet werden.)

(d) Schätze die normale Kovarianzmatrix von $\hat{\beta}$ und die robuste White-Kovarianzmatrix von $\hat{\beta}$. Vergleiche die Standardfehler von $\hat{\beta}_0, \dots, \hat{\beta}_4$.

Aufgabe 4 (3 Punkte)

Es kann vorkommen, dass nicht nur Heteroskedastizität auftritt, sondern die Störterme auch noch beliebig korreliert sind, d.h. die (bedingte) Kovarianzmatrix der Störterme $u = (u_1, \dots, u_n)$ ist gegeben durch $V(u|X) = \sigma^2 V$, wobei $\sigma^2 > 0$ und V eine beliebige symmetrische und positiv definite Matrix ist. Wir betrachten den Schätzer $\hat{\beta} = (X^\top V^{-1} X)^{-1} X^\top V^{-1} Y$ für den Vektor β der Regressionsparameter. Zeige, dass $\mathbb{E}(\hat{\beta}|X) = \beta$ und berechne die Kovarianzmatrix $V(\hat{\beta}|X)$.

Hinweis: Dieses Verfahren wird auch als *generalized least squares (GLS)* bezeichnet und ist eine Verallgemeinerung des von uns genutzten *ordinary least squares (OLS)*.