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

SS 2015
19. 5. 2015

Ökonometrie - Übungsblatt 3

Abgabe am 26. 5. **vor** Beginn der Übung

Aufgabe 1 (3,5+1+1+1,5+5 Punkte)

Die folgende Aufgabe soll ohne die Verwendung von **R** bearbeitet werden.
In neun verschiedenen amerikanischen Wintersportorten wurden während einer gewissen Beobachtungszeit die Anzahl der Besucher registriert. Es wird angenommen, dass diese linear von der Gesamtlänge der zur Verfügung stehenden Pisten sowie der Liftkapazität abhängen. Wir setzen voraus, dass die Annahmen MLR1-MLR6 erfüllt sind.

Pistenlänge	Liftkapazität	Besucherzahl
10,5	22	19929
2,5	10	5839
13,1	32,5	23696
4,0	14,75	9881
14,7	38	30011
3,6	12	7241
7,1	19	11634
17,0	42	36476
6,4	18,5	12068

(a) Bestimme und interpretiere die Regressionsparameter β_0,β_1 und β_2 und schätze die Varianz σ^2 der Störterme.

(b) Überlege dir zwei weitere Größen, die einen linearen Einfluss auf die Besucherzahl haben könnten.

(c) Prognostiziere die Anzahl der Besucher in einem Skigebiet mit 15,0 km Pistenlänge und einer Liftkapazität von 39,5.

(d) Bestimme und interpretiere das Bestimmtheitsmaß R^2.

(e) Prüfe die Hypothese H_0 gegen die Alternative H_1 zum Signifikanzniveau $\alpha = 0,05$, falls

  (i) H_0 : Es besteht keine Abhängigkeit der Besucherzahl von der Liftkapazität,
  H_1: Es besteht eine Abhängigkeit der Besucherzahl von der Liftkapazität

  (ii) H_0 : Es besteht keine Abhängigkeit der Besucherzahl von der Pistenlänge,
  H_1: Es besteht eine positive Abhängigkeit der Besucherzahl von der Pistenlänge.

  (iii) $H_0 : \beta_1 = \beta_2$, $H_1 : \beta_1 \neq \beta_2$.

Hinweis: Für $A \in \mathbb{R}^{3\times 3}$ gilt, dass

$$A^{-1} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}^{-1} = \frac{1}{\det A} \begin{pmatrix} a_{22}a_{33} - a_{23}a_{32} & a_{13}a_{32} - a_{12}a_{33} & a_{12}a_{23} - a_{13}a_{22} \\ a_{23}a_{31} - a_{21}a_{33} & a_{11}a_{33} - a_{13}a_{31} & a_{13}a_{21} - a_{11}a_{23} \\ a_{21}a_{32} - a_{22}a_{31} & a_{12}a_{31} - a_{11}a_{32} & a_{11}a_{22} - a_{12}a_{21} \end{pmatrix}$$

und $\det A = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$.

Aufgabe 2 (1+4+2 Punkte)

Die Anzahl der Autos pro 100 Einwohner in verschiedenen Ländern möge linear vom Pro-Kopf-Einkommen (in 10000 €) und dem Benzinpreis (in €) abhängen. Lade dazu die Datei `autos.txt` von der Vorlesungshomepage herunter. Wir betrachten ein multiples lineares Regressionsmodell, wobei wir MLR1-MLR6 als erfüllt annehmen.

(a) Begründe, welche Vorzeichen du für die Regressionsparameter β_1 und β_2 erwarten würdest.

(b) Führe in **R** die lineare Regression mit Hilfe des Befehls `lm()` durch (siehe Hinweis) und interpretiere ausführlich die Ausgabe.

(c) Berechne (mit Hilfe der Ausgabe aus (b)) symmetrische Konfidenzintervalle zum Niveau $\gamma = 0,98$ für die Regressionsparameter β_0, β_1 und β_2.

Hinweis: Sei `y` der Vektor der abhängigen Variablen und `x1`, `x2`, `x3` Vektoren der unabhängigen Variablen. Dann wird in **R** mittels `lm(y~1+x1+x2+x3)` ein lineares Regressionsmodell angepasst (das funktioniert natürlich auch mit einer beliebigen Anzahl unabhängiger Variablen). Um die komplette Modellanpassung zu erhalten wird der Befehl `print(summary(lm(...)))` verwendet.

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

Die folgende Aufgabe soll vollständig mit Hilfe von **R** bearbeitet werden.
Das Produktionsvolumen der USA zwischen 1932 und 1953 lässt sich mit Hilfe der Cobb-Douglas-Produktionsfunktion $Y_i = c \cdot K_i^{\beta_1} \cdot A_i^{\beta_2} \cdot u_i$ für $i = 1, \ldots, 22$ beschreiben. Dabei bezeichne Y_i die Produktion (in Mrd. Dollar), K_i den Kapitaleinsatz (in Mrd. Dollar) und A_i den Arbeitseinsatz (in Mio. Arbeitskräften) im Jahr i und c, β_1, β_2 sind unbekannte Konstanten. Die Daten für die Jahre 1932 bis 1953 sind in der Datei `produktion.txt` gespeichert, welche auf der Vorlesungshomepage heruntergeladen werden kann.

(a) Führe den Modellansatz in ein geeignetes lineares Modell über und schätze die Modellparameter $\log c, \beta_1$ und β_2 ohne die Funktion `lm()` zu verwenden (man nennt das Modell dann ein quasilineares Modell).
Hinweis: Für eine Matrix `A` lassen sich mit `t(A)` die Transponierte und mit `solve(A)` die Inverse berechnen. Beachte, wie die Matrixmultiplikation in **R** durchgeführt wird.

(b) Erstelle eine Abbildung, welche die tatsächliche Entwicklung des Produktionsvolumens Y über die Jahre 1932-1953 darstellt, sowie die Kurve des Produktionsvolumens $\hat{Y}$, die sich aus dem quasilinearen Modell ergibt.
Hinweis: Nutze für den ersten Graphen `plot()` und für den zweiten Graphen `points()` ohne den Parameter `add`. Mit dem Parameter `type` kannst du durchgezogene Linien einstellen.

(c) Bestimme ein symmetrisches Konfidenzintervall zum Niveau $\gamma = 0,95$ für β_1.

(d) Teste die Hypothese $H_0 : \beta_2 = 2$ gegen $H_1 : \beta_2 < 2$ zum Signifikanzniveau $\alpha = 0.01$.

Aufgabe 4 (1+2 Punkte)

Um die Modellspezifikationen im Rahmen der multiplen linearen Regression zu überprüfen (insbesondere die Linearität des Modells), wird häufig ein sogenannter RESET Test angewendet. Wenn MLR1-MLR6 erfüllt sind, dann sollten nichtlineare Kombinationen der erklärenden Variablen keinen Einfluss auf die abhängige Variable haben. Zunächst werden die Regressionsparameter $\beta_0, \ldots, \beta_k$ geschätzt und die Modellwerte $\hat{Y}_i$ berechnet. Anschließend wird das lineare Modell $Y_i = \hat{Y}_i + \gamma_2 \hat{Y}_i^2 + \ldots + \gamma_m \hat{Y}_i^m$ mit einem geeignet gewählten m betrachtet. Sollten $\hat{Y}_i^2, \ldots, \hat{Y}_i^m$ einen Einfluss haben, d.h. wenn der Test $H_0 : \gamma_2 = \ldots = \gamma_m = 0$ abgelehnt wird, dann ist wohl mindestens eine der Modellannahmen MLR1- MLR6 verletzt und das Modell sollte nochmal überdacht werden. (Wie die Testgröße dieses Tests aussieht und welche Verteilung sie hat soll uns hier nicht interessieren.)

(a) Im folgenden Beispiel wird der RESET Test in **R** durchgeführt. Der Dataframe `wohnungen` enthalte die Mietdaten, die schon auf Übungsblatt 1 und 2 betrachtet wurden.

```
library(lmtest)
resettest(wohnungen$nm ~ 1+wohnungen$wfl, power=2:4, type="fitted")

RESET test

data:  wohnungen$nm ~ 1 + wohnungen$wfl
RESET = 5.8799, df1 = 3, df2 = 2048, p-value = 0.0005398
```

Das package `lmtest` muss einmalig mittels `install.packages("lmtest")` installiert werden. Interpretiere das Ergebnis des Tests.

(b) Wende den RESET Test auf die lineare Regression in Aufgabe 2 an und interpretiere das Resultat. Es sollen dabei nur $\hat{Y}_i^2$ und $\hat{Y}_i^3$ betrachtet werden, d.h. $m = 3$.