Ökonometrie Grundlagen, Übung 5

Henry Haustein

Aufgabe 1

(A1) Positive und negative Fehler heben sich gegenseitig auf:

$$\mathbb{E}(u) = 0$$

(A2) Die Spalten von X sind linear unabhängig bzw. Abwesenheit von Multikollinearität.

$$\begin{aligned} \mathrm{rk}(X) &= k \\ \mathrm{rk}(X'X) &= k \\ |X'X| &\neq 0 \end{aligned}$$

(A2*) Die Regressormatrix X besitzt eine kleine Konditionszahl, ist also gut konditioniert.

(A3) Keine Autokorrelation und keine Heteroskedastizität.

(A4) Die Effekte von $x_{1t}, ..., x_{kt}$ und u_t auf y_t sind separabel.

$$\mathbb{E}(X'u) = 0$$

Ist bei deterministischem X erfüllt.

(A5) Der Fehlervektor soll einer multivariaten Normalverteilung folgen. Diese Annahme ist die Grundlage für die Konstruktion statistischer Signifikanztests.

Aufgabe 2

Ziel ist es die der quadrierten Fehler zu minimieren, in Matrixschreibweise $Q_T = (y - X\beta)'(y - X\beta)$. Wir werden für den Beweis einige Ableitungen von Matrizen brauchen[1]:

$$\begin{aligned} \frac{\partial(Av)}{\partial v} &= A' \\ \frac{\partial(v'A)}{\partial v} &= A \\ \frac{\partial(v'Av)}{\partial v} &= 2A'v \end{aligned}$$

[1] Hier kommt ziemlich komplizierte Mathematik ins Spiel, deswegen weiß ich nicht, ob dieser Lösungsweg überhaupt geplant war. Aber man kommt mit ihm recht schnell ans Ziel. Weitere Informationen hier: https://en.wikipedia.org/wiki/Matrix_calculus

Führen wir nun den Beweis:

$$
\begin{aligned}
Q_T &= y'y - y'X\hat{\beta} - \hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} \\
\frac{\partial Q_T}{\partial \hat{\beta}} &= 0 - \underbrace{(y'X)'}_{X'y} - X'y + 2(X'X)\hat{\beta} = 0 \\
0 &= -2X'y + 2(X'X)\hat{\beta} \\
(X'X)\hat{\beta} &= X'y \\
\hat{\beta} &= (X'X)^{-1}X'y
\end{aligned}
$$

Aufgabe 3

$$
\begin{aligned}
(1...1)\hat{u} &= (1...1)(y - X\hat{\beta}) \\
&= (1...1)(y - X(X'X)^{-1}X'y) \\
&= (1...1)(y - XX^{-1}(X')^{-1}X'y) \\
&= (1...1)(y - y) \\
&= 0
\end{aligned}
$$

Eigenschaft (A1)

Aufgabe 4

Der Beweis läuft sehr ähnlich zu Aufgabe 3

$$
\begin{aligned}
X'\hat{u} &= (1...1)(y - X\hat{\beta}) \\
&= X'(y - X(X'X)^{-1}X'y) \\
&= X'(y - XX^{-1}X'^{-1}X'y) \\
&= X'(y - y) \\
&= 0
\end{aligned}
$$

Eigenschaft (A4)

Aufgabe 5

Auf der Hauptdiagonalen der Varianz-Kovarianz-Matrix wird der Ausdruck Σ_z zu $\mathbb{E}((z_i - \theta_i)(z_i - \theta_i)') = \mathbb{E}((z_i - \theta_i)^2) = \mathrm{Var}(z_i)$ ausgewertet. Auf den Nebendiagonalen wird der Ausdruck zu $\mathbb{E}((z_i - \theta_i)(z_j - \theta_j)') = \mathbb{E}((z_i - \theta_i)(z_j - \theta_j)) = \mathrm{Cov}(z_i, z_j)$ ausgewertet.

Aufgabe 6

(a) Das ökonometrische Modell ist: $y_i = \beta_1 + \beta_2 x2i + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i$

(b) in R:

```r
datensatz = matrix(c(250,0,68,3500,1,850,1,49,1600,1,620,1,71,2000,
        0,160,0,23,800,0,600,0,45,3000,1,1420,1,33,2400,0,550,1,28,
        2100,1,1600,1,62,4500,0,2100,0,58,9000,0,1800,1,48,3300,1,
        420,1,24,400,1,300,0,22,480,1),ncol = 5, byrow = TRUE)
y = datensatz[,1]
X = cbind(rep(1,12),datensatz[,2],datensatz[,3],datensatz[,4],
        datensatz[,5])
beta = solve(t(X) %*% X) %*% t(X) %*%y
```

ergibt ein $\hat{\beta}$ von

$$\hat{\beta} = \begin{pmatrix} 186.9027199 \\ 659.2722612 \\ -9.2664317 \\ 0.2821646 \\ -85.8955146 \end{pmatrix}$$

(c) in R

```r
# Berechnen der Fehler u
error = y - X %*% beta
# Mean Squared Error
MSE = sum(error * error)/(12-(4+1))
# Matrix, die die Varianzen aller betas enthaelt
varMatrix = diag(MSE * solve(t(X) %*% X))
# Matrix, die die Standardabweichung aller betas enthaelt
sdMatrix = sqrt(varMatrix)
# Standardabweichung von beta4 = se(beta4)
SEbeta4 = sdMatrix[4]
# Teststatistik
beta[4,1]/SEbeta4
# 97.5 Quantil der t-Verteilung mit 5 df
qt(0.975, df=5)
```

Die Teststatistik hat einen Wert von 4.627785, aber der kritische Wert ist schon bei 2.570582. $\hat{\beta}_4$ ist also nicht signifikant von Null verschieden.

(d) in R

```r
summary(lm(y ~ X[,-1]))
```

Das bestätigt die vorher berechneten Werte. Obwohl $R^2_{adj} = 0.7239$ sind viele Parameter nicht signifikant von Null verschieden. Man sollte diese aus dem Modell entfernen.

Aufgabe 7

(a) ökonometrisches Modell: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + u_i$

(b) in R

```r
datensatz = read.csv2("Mikrozensus.csv")
head(datensatz)
modell = lm(Einkommen ~ ., data = datensatz)
summary(modell)
```

Schätzer für Alter: 6.8098, Konfidenzintervall: [5.488043; 8.131639]
Schätzer für Anstellungsdauer: 12.5050, Konfidenzintervall: [11.088655; 13.921416]
Schätzer für Teilzeit: -685.9728, Konfidenzintervall: [-719.082560; -652.862975]

(c) in R

```r
modell2 = lm(Einkommen ~ Alter + Geschlecht + Kinder +
         AnstDauer + Uni, data = datensatz)
summary(modell2)
```

Interessant ist, dass plötzlich die Variable Alter nicht mehr signifikant von Null verschieden ist.
Schätzer für Alter: 0.5623, Konfidenzintervall: [-0.8420126; 1.966635]
Schätzer für Anstellungsdauer: 17.7254, Konfidenzintevall: [16.2045149; 19.246345]