

## Übungsblatt 3 — Einfaches Regressionsmodell

Zu diesem Übungsblatt empfehlen wir neben der Lektüre der Kapitel 4 des Lehrbuches *Introduction to Econometrics* von *Stock & Watson* eine Aufarbeitung mithilfe der Kapitel 2 und 3 in unserem Online-Companion *Introduction to Econometrics with R*.

### Aufgabe 1 – Regression mit Konstante

Betrachten Sie das Regressionsmodell

$$Y_i = \beta + u_i, \quad i = 1, \dots, n.$$

Zeigen Sie, dass der OLS-Schätzer  $\hat{\beta}$  von  $\beta$  dem Stichprobenmittelwert  $\bar{Y}$  entspricht.

### Aufgabe 2 – KQ-Annahmen

Nennen und erläutern sie *kurz* die Kleinst-Quadrate-Annahmen im einfachen Regressionsmodell

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n.$$

### Aufgabe 3 – KQ-Schätzer im einfachen Regressionsmodell

Betrachten Sie erneut das lineare Regressionsmodell

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, n.$$

- (a) Zeigen Sie, dass die Regressionsgerade durch den Punkt  $(\bar{X}, \bar{Y})$  verläuft.
- (b) Angenommen es gilt  $\beta_0 = 0$ . Leiten Sie eine Formel für den Kleinst-Quadrate-Schätzer  $\hat{\beta}_1$  von  $\beta_1$  her, d.h. lösen Sie das Minimierungsproblem

$$\min_{b_1} \sum_{i=1}^n [Y_i - b_1 X_i]^2.$$

### Aufgabe 4 – Regressionsbeispiel: Bremsweg

Bei Verkehrsunfällen spielt für die Klärung der Schuldfrage oft die Beziehung zwischen Geschwindigkeit und Bremsweg eine wichtige Rolle. Die folgende Tabelle enthält diesbezügliche Informationen für einen PKW, die im Rahmen eines Sicherheitstrainings gemessen wurden:

Geschwindigkeit (in km/h)	20	30	50	80	100
Bremsweg (in m)	5	10	25	60	10

- (a) Schätzen Sie den Zusammenhang des Bremswegs  $Y$  und der Geschwindigkeit  $X$  mit der KQ Methode, d.h. schätzen Sie  $\beta_0$  und  $\beta_1$  im Modell

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad i = 1, \dots, 5.$$

- (b) Berechnen Sie die geschätzten Werte  $\hat{Y}$  und die Residuen  $\hat{u}$  für alle fünf Messungen. Überprüfen Sie, dass sich die Residuen zu 0 aufsummieren.
- (c) Geben Sie eine Schätzung für den Bremsweg bei einer Geschwindigkeit von 60 km/h an.
- (d) Wie groß ist der Anteil der durch die Geschwindigkeit ( $X$ ) erklärten Variation des Bremsweges ( $Y$ )?
- (e) Stellen Sie die Daten und die geschätzte Regressionsgerade grafisch dar.
- (f) Lösen Sie die Teilaufgaben (a)–(e) mit R.

### Aufgabe 5 – Regressionsanalyse mit CASchools

- (a) Laden Sie den Datensatz **CASchools** (verfügbar in Moodle). Nutzen sie hierzu den Befehl `load()`.  
Wir setzen zunächst das Arbeitsverzeichnis und laden anschließend den Datensatz **CASchools.Rdata**. Dateien mit der Endung **.Rdata** enthalten *R-Objekte* (**CASchools.Rdata** enthält den `data.frame` **CASchools**) und können mit der Funktion `load()` eingelesen werden.

```
## setwd("Z:/")
load("CASchools.Rdata")
```

- (b) Erzeugen Sie eine statistische Zusammenfassung des Datensatzes mit `summary()`.  
Wir nutzen `summary()` für die statistische Zusammenfassung aller Variablen in **CASchools**.

```
summary(CASchools)
```

math	STR	income	computer
Min. :605.4	Min. :14.00	Min. : 5.335	Min. : 0.0
1st Qu.:639.4	1st Qu.:18.58	1st Qu.:10.639	1st Qu.: 46.0
Median :652.5	Median :19.72	Median :13.728	Median : 117.5
Mean :653.3	Mean :19.64	Mean :15.317	Mean : 303.4
3rd Qu.:665.9	3rd Qu.:20.87	3rd Qu.:17.629	3rd Qu.: 375.2
Max. :709.5	Max. :25.80	Max. :55.328	Max. :3324.0

expenditure
Min. :3926
1st Qu.:4906
Median :5215
Mean :5312
3rd Qu.:5601
Max. :7712

Stichprobenstandardabweichung und -varianz können mit `sd()` bzw. `var` für einzelnen Variablen in **CASchools** berechnet werden. Beachte, das `$` genutzt wird, um Variablen in **CASchools** zu referenzieren.

```
# Empirische Standardabweichungen für math und STR
sd(CASchools$math)
```

```
[1] 18.7542
```

```
sd(CASchools$STR)
```

```
[1] 1.891812
```

```
# Empirische Varianzen für math und STR
var(CASchools$math)
```

```
[1] 351.7201
```

```
var(CASchools$STR)
```

```
[1] 3.578952
```

- (c) Regressieren Sie die im Mathetest erreichte Punktzahl (**math**) auf die Klassengröße (**STR**). Lassen Sie sich mit **summary()** eine statistische Zusammenfassung des Modells anzeigen.

Wir übergeben den Datensatz sowie das zu schätzende Modell als Formel an **lm()** und erzeugen mit **summary()** eine statistische Zusammenfassung der Schätzung.

```
# Schätzen
mod_CASchools <- lm(math ~ STR, data = CASchools)
# Geschätzte Koeffizienten 'schnell' erhalten
mod_CASchools
```

Call:

```
lm(formula = math ~ STR, data = CASchools)
```

Coefficients:

```
(Intercept)      STR
   691.417      -1.939
```

```
# Statistische Zusammenfassung des Modells erzeugen
summary(mod_CASchools)
```

Call:

```
lm(formula = math ~ STR, data = CASchools)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-44.615 -13.374  -0.828  12.728  52.711
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  691.4174     9.3825   73.692 < 2e-16 ***
STR          -1.9386     0.4755   -4.077 5.47e-05 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 18.41 on 418 degrees of freedom

Multiple R-squared: 0.03824, Adjusted R-squared: 0.03594

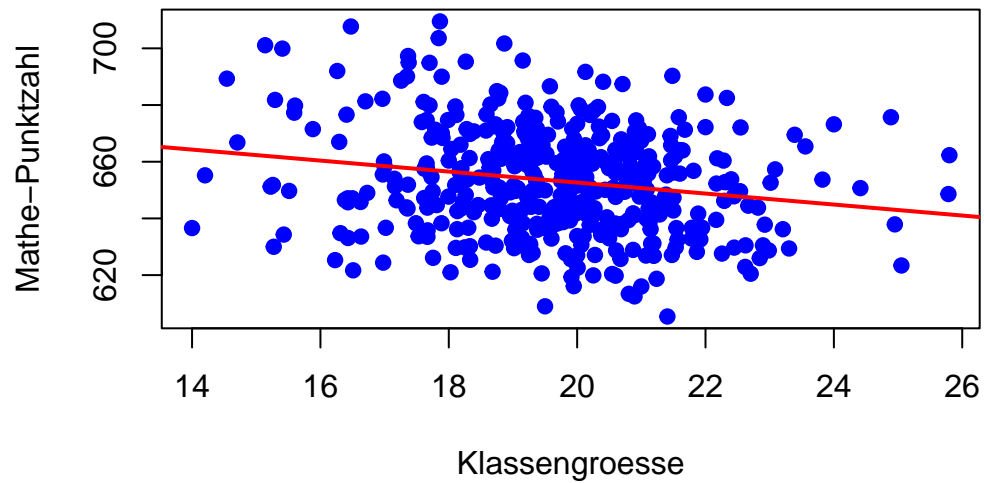
F-statistic: 16.62 on 1 and 418 DF, p-value: 5.467e-05

- (d) Plotten Sie die Daten zusammen mit der Regressionsgeraden. Interpretieren Sie *kurz* die Ergebnisse.

```

plot(x = CASchools$STR, # Daten
     y = CASchools$math,
     pch = 19, # ausgemahlte Kreise
     col = "blue", # Farbe
     xlab = "Klassengroesse", # Beschriftung X
     ylab = "Mathe-Punktzahl" # Beschriftung Y
     )
abline(mod_CASchools,
       col = "red",
       lwd = 2 # Dicke der Linie
       )

```



*Hinweis:* CASchools ist ein data.frame. Variablen können mit \$ ausgelesen werden. Beispiel: CASchools\$math