1. Lösungen zu Kapitel 5

Übungsaufgabe 5.1

a) *Falsch!*

Die Varianz des geschätzten Koeffizienten $\hat{\beta}_1$ berechnet sich folgendermaßen:

$$\widehat{Var(\beta_1)} = \frac{\hat{\sigma}^2}{\hat{\sigma}_X^2} = \frac{\left(\frac{\sum \hat{\varepsilon}_t{}^2}{N-(K+1)}\right)}{\overline{X_t^2} - \overline{X}^2}$$
$$= \frac{\frac{15}{5-2}}{10-8} = \frac{5}{2} = 2.5$$

b) *Falsch!*

Ist $E[\varepsilon_t \varepsilon_s] \neq 0$, so sind die Störterme korreliert (z.B. Autokorrelation bei Zeitreihendaten). Die geschätzten Koeffizienten sind weiterhin unverzerrt, jedoch sind die Standardfehler inkonsistent.

c) *Falsch!*

$$\hat{\beta}_1 = \frac{\overline{XY} - \overline{Y}\ \overline{X}}{\overline{X^2} - \overline{X}^2}$$
$$= \frac{18 - 5 \cdot 3}{11 - 3^2} = \frac{3}{2} = 1.5$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$
$$= 5 - 1.5 \times 3 = 0.5$$

d) *Falsch!*

Die Formel für die Varianz des Schätzers $\hat{\beta}$ verdeutlicht den Zusammenhang zwischen der Präzision der Schätzung und der Variation in X:

$$\hat{\sigma}_{\hat{\beta}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N} (X_i - \overline{X})^2}.$$

Die Präzision des geschätzten $\hat{\beta}$ ist folglich umso höher, je geringer die Varianz des Störterms und je größer die Variation der exogenen Variablen X ist.

e) *Falsch!*

Aus den zwei OLS-Annahmen, dass X_i *nicht* stochastisch ist und dass der Erwartungswert der Störterme gleich Null ist ergibt sich:

$$E(X_t \varepsilon_t) = X_t E(\varepsilon_t) = X_t \cdot 0 = 0.$$

Übungsaufgabe 5.2

Aus den beiden Normalgleichungen (5.6) und (5.7), die hier zur Wiederholung noch einmal aufgeführt sind

$$-2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \tag{1.1}$$

$$-2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0, \tag{1.2}$$

erhält man durch folgende Umformungen die Schätzformeln (5.8) und (5.9) (bzw. (5.10)),

$$\begin{aligned} \hat{\beta}_0 &= \overline{Y} - \hat{\beta}_1 \overline{X} \\ \hat{\beta}_1 &= \frac{\overline{YX} - \overline{Y}\ \overline{X}}{\overline{X_i^2} - \overline{X}^2} \end{aligned}$$

.

Aus Gleichung (1.1) ergibt sich durch Multiplikation mit $-\frac{1}{2N}$ und Aufspalten der Summe

$$\frac{1}{N} \sum_{i=1}^{N} Y_i - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_0 - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_1 X_i = 0. \tag{1.3}$$

Da $\hat{\beta}_0$ und $\hat{\beta}_1$ Konstanten sind, entspricht dies

$$\frac{1}{N} \sum_{i=1}^{N} Y_i - \frac{1}{N} \cdot N \cdot \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{N} \sum_{i=1}^{N} X_i = \overline{Y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{X} = 0, \tag{1.4}$$

bzw. nach $\hat{\beta}_0$ aufgelöst

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} = 0. \tag{1.5}$$

Dies bedeutet, dass der Schätzer $\hat{\beta}_0$ vom Schätzer $\hat{\beta}_1$ abhängt. Gleichung (1.5) legt nahe, zuerst einen Schätzwert für $\hat{\beta}_1$ zu ermitteln, um dann auf Basis von (1.5) den Schätzwert für $\hat{\beta}_0$ zu errechnen.

Aus Gleichung (1.2) ergibt sich wiederum durch Multiplikation mit $-\frac{1}{2N}$ und Aufspalten der Summe

$$\frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_0 X_i - \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}_1 X_i^2 = 0. \tag{1.6}$$

Da $\hat{\beta}_0$ und $\hat{\beta}_1$ Konstanten sind und für $\hat{\beta}_0$ Gleichung (1.5) gilt, ergibt sich

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \frac{1}{N}\sum_{i=1}^{N} (\overline{Y} - \hat{\beta}_1 \overline{X}) X_i - \hat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} X_i^2 = \\
&\frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \overline{Y}\frac{1}{N}\sum_{i=1}^{N} X_i + \hat{\beta}_1 \overline{X}\frac{1}{N}\sum_{i=1}^{N} X_i - \hat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} X_i^2 = \\
&\overline{YX} - \overline{Y}\ \overline{X} + \hat{\beta}_1 \overline{X}^2 - \hat{\beta}_1 \overline{X^2} = 0.
\end{aligned} \tag{1.7}$$

Löst man diese Gleichung nach $\hat{\beta}_1$ auf, erhält man

$$\hat{\beta}_1 \left(\overline{X^2} - \overline{X}^2\right) = \overline{YX} - \overline{Y}\ \ \overline{X} \tag{1.8}$$

bzw.

$$\hat{\beta}_1 = \frac{\overline{YX} - \overline{Y}\ \overline{X}}{\overline{X_i^2} - \overline{X}^2}. \tag{1.9}$$

Mit dieser Gleichung könnte man bereits arbeiten, denn sie enthält nur beobachtbare Werte und Konstanten. Sie lässt sich aber noch etwas vereinfachen und in eine übersichtlichere Schreibweise bringen. Hierzu vergegenwärtige man sich, dass für die geschätzte Kovarianz von Y und X gilt

$$\begin{aligned}
\hat{\sigma}_{XY} &= \frac{1}{N}\sum_{i=1}^{N} (Y_i - \overline{Y})(X_i - \overline{X}) \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \frac{1}{N}\sum_{i=1}^{N} \overline{Y} X_i - \frac{1}{N}\sum_{i=1}^{N} \overline{X} Y_i + \frac{1}{N}\sum_{i=1}^{N} \overline{X}\ \overline{Y} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \overline{Y}\ \overline{X} - \overline{X}\ \overline{Y} + \overline{X}\ \overline{Y} \\
&= \frac{1}{N}\sum_{i=1}^{N} Y_i X_i - \overline{Y}\ \overline{X} \\
&= \overline{YX} - \overline{Y}\ \overline{X}.
\end{aligned} \tag{1.10}$$

Gleichung (1.10) entspricht dem Zähler von $\hat{\beta}_1$ in Gleichung (1.9). Die geschätzte Kovarianz von X mit sich selbst entspricht natürlich der geschätzten Varianz von X. Für diese gilt dann analog zu Gleichung (1.10)

$$\begin{aligned} \hat{\sigma}_{XX} &= \frac{1}{N}\sum_{i=1}^{N} X_i X_i - \overline{X}\;\overline{X} \;=\; \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \overline{X}\;\overline{X} \\ &= \overline{X_i^2} - \overline{X}^2 \;=\; \hat{\sigma}_X^2 . \end{aligned} \tag{1.11}$$

Gleichung (1.11) entspricht dem Nenner von $\hat{\beta}_1$ in Gleichung (5.50). Der Schätzer für $\hat{\beta}_1$ kann daher auch als

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{N}(X_i - \overline{X})^2} = \frac{\hat{\sigma}_{XY}}{\hat{\sigma}_X^2} \tag{1.12}$$

geschrieben werden. Die Tatsache, dass in diesen Herleitungen N statt $N-1$ zur Normierung benutzt wurde, schränkt die Gültigkeit in keiner Weise ein, denn dieser Normierungsfaktor kürzt sich in der Schätzformel für $\hat{\beta}_1$ ohnehin heraus.

Häufig findet man als Formel für den Schätzer $\hat{\beta}_1$ auch die Gleichungen

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})Y_i}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \tag{1.13}$$

bzw.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})X_i}{\sum_{i=1}^{N}(X_i - \overline{X})^2} \tag{1.14}$$

Diese Gleichung erhält man, indem man den Zähler von Gleichung (1.12) entsprechend auflöst. Dies sei für Gleichung (1.14) exemplarisch dargestellt.

Löst man die Summe im Zähler von Gleichung (1.12) auf und verwendet man, dass $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ bzw. dass $\sum_{i=1}^{N} Y_i = N\overline{Y}$, erhält man den Zähler von Gleichung (1.14):

$$\begin{aligned}\sum_{i=1}^{N}(X_i-\overline{X})(Y_i-\overline{Y}) &= \sum_{i=1}^{N} Y_iX_i - \sum_{i=1}^{N}\overline{Y}X_i - \sum_{i=1}^{N}\overline{X}Y_i + \sum_{i=1}^{N}\overline{X}\,\overline{Y} \\ &= \sum_{i=1}^{N} Y_iX_i - \sum_{i=1}^{N}\overline{Y}X_i - \overline{X}\sum_{i=1}^{N}Y_i + \sum_{i=1}^{N}\overline{X}\,\overline{Y} \\ &= \sum_{i=1}^{N} Y_iX_i - \sum_{i=1}^{N}\overline{Y}X_i - \overline{X}N\overline{Y} + N\overline{X}\,\overline{Y} \\ &= \sum_{i=1}^{N} Y_iX_i - \sum_{i=1}^{N}\overline{Y}X_i \\ &= \sum_{i=1}^{N} Y_iX_i - \overline{Y}X_i = \sum_{i=1}^{N}(Y_i-\overline{Y})X_i.\end{aligned}$$

Übungsaufgabe 5.3

Der OLS-Schätzer ergibt sich aus der Minimierung der Summe der quadratischen Residuen, i.e.:

$$\min_{\hat{\beta}_0}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 = \sum_{i=1}^{N}(Y_i-\hat{\beta}_0)^2.$$

Durch Ableitung nach $\hat{\beta}_0$ gelangt man zur Bedingung erster Ordnung:

$$\frac{\partial\sum_{i=1}^{N}(\hat{\varepsilon}_i^2)}{\partial\hat{\beta}_0} = -2\sum_{i=1}^{N}(Y_i-\hat{\beta}_0)=0.$$

Durch Multiplikation mit $-\frac{1}{2N}$ erhält man:

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}(Y_i-\hat{\beta}_0) &= \frac{1}{N}\sum_{i=1}^{N}Y_i - \frac{1}{N}\cdot N\cdot\hat{\beta}_0 = 0 \\ \Leftrightarrow \hat{\beta}_0 &= \frac{1}{N}\sum_{i=1}^{N}Y_i = \bar{Y}\end{aligned}$$

Der geschätzte Koeffizient $\hat{\beta}_0$ in dem Regressionsmodell $Y=\beta_0+\varepsilon$ entspricht foglich dem Stichprobenmittel.

Übungsaufgabe 5.4

Ein Schätzer ist unverzerrt, wenn der Erwartungswert des Schätzers dem tatsächlichen Wert in der Population entspricht. Es muss gelten $E(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$.

$$\begin{aligned}
E(\hat{\boldsymbol{\beta}}) &= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}] \\
&= E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'}(\boldsymbol{X\beta}+\boldsymbol{\varepsilon})] \\
&= E[\underbrace{(\boldsymbol{X'X})^{-1}\boldsymbol{X'X}}_{=\boldsymbol{I}}\boldsymbol{\beta}] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\varepsilon}] \\
&= E[\boldsymbol{\beta}] + E[(\boldsymbol{X'X})^{-1}\boldsymbol{X'\varepsilon}] \\
&= \boldsymbol{\beta} + E[(\boldsymbol{X'X})^{-1}\underbrace{\boldsymbol{X'\varepsilon}}_{=\boldsymbol{0}}] \\
&= \boldsymbol{\beta}.
\end{aligned}$$

Übungsaufgabe 5.5

Die Erwartungswertbildung von $E(\boldsymbol{X'}\hat{\boldsymbol{\varepsilon}})$ im multivariaten Regressionsmodell findet analog zum bivariaten Modell statt:

$$E(\boldsymbol{X'}\hat{\boldsymbol{\varepsilon}}) = \boldsymbol{X'}\underbrace{E(\hat{\boldsymbol{\varepsilon}})}_{=0} = 0.$$

Da die erklärende Variable $\boldsymbol{X}$ nicht stochastisch ist, kann diese Umformung vorgenommen werden. Sind $\boldsymbol{X}$ und $\hat{\boldsymbol{\varepsilon}}$ unkorreliert, so wird $\boldsymbol{X}$ als *exogene erklärende Variable* bezeichnet.

Übungsaufgabe 5.6

Eine Koeffizient kann statistisch signifikant sein und dennoch ökonomisch insignifikant.

Ein Beispiel ist der Einfluss der Schulausbildung auf die Löhne. Dieser kann zwar statistisch signifikant sein, dennoch aber so gering, dass er ökonomisch nicht von großer Bedeutung ist. Führen z.B. 20 Schuljahre zu einer Erhöhung der Löhne um eine Einheit, so ist dieser Effekt zu vernachlässigen.

Übungsaufgabe 5.7

Das Bestimmtheitsmaß R^2 ist definiert als der Anteil der vom Modell erklärten Variation an der gesamten Variation in Y_i, d.h.

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS},$$

wobei *TSS* (*Total Sum of Squares*) die gesamte Variation in der abhängigen Variablen Y_i, *RSS*(*Regression Sum of Squares*) die vom Modell erklärte Variation in Y_i und *ESS* die *Error Sum of Squares*, also die verbleibende, nicht vom Modell erklärte Variation der abhängigen Variablen ist.

Am besten lässt sich der Einfluss eines weiteren Regressors über folgende Dekomposition betrachten:

$$\underbrace{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}_{TSS} = \underbrace{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2}_{ESS} + \underbrace{\sum_{i=1}^{N}(\hat{Y}_i - \overline{Y})^2}_{RSS}.$$

Die Hinzunahme eines weiteren Regressors führt nie dazu, dass RSS, sprich die vom Modell erklärte Variation der abhängigen Variablen, sinkt. Somit wird auch das R^2 durch Hinzunahme eines Regressors nicht sinken - unabhängig davon ob er einen Erklärungsbeitrag zu dem Model liefert oder nicht.

Da ESS durch die zusätzliche erklärende Variable sinken kann, ist es sogar möglich, dass das R^2 steigt.

Übungsaufgabe 5.8

a) Zunächst wird getestet, ob die Koeffizienten statistisch signifikant verschieden von Null sind. Die Hypothesen werden mit Hilfe des t-Tests überprüft. Die t-Statistiken lauten:

$$|t| = \left|\frac{-12,894 - 0}{3,177}\right| = 4,059 \sim t_{104-2-1}$$
$$|t| = \left|\frac{1,397 - 0}{0,229}\right| = 6,100 \sim t_{104-2-1}.$$

In beiden Fällen kann die Nullhypothese $H_0 : \beta_n = 0$ bzw. $H_0 : \beta_K = 0$ verworfen werden, da die t-Statistik größer als der kritische Wert der t-Verteilung ist ($1,984$ bei einer 5% Irrtumswahrscheinlichkeit).

Die Schätzung ergibt einen negativen Effekt des Bevölkerungswachstums auf das Bruttoinlandsprodukt pro Kopf und einen positiven Einfluss der Sparrate auf das BIP pro Kopf. Steigt die durchschnittliche Rate des Bevölkerungswachstums um eine Einheit, so sinkt das BIP pro Kopf um 12,894 Einheiten. Steigt der durchschnittliche Anteil der Investitionen am BIP von 1960 bis 1990 um eine Einheit, steigt auch das BIP um 1,379 Einheiten.

Während der positive Effekt der Sparrate den Erwartungen bzw. den Vorhersagen des neoklassischen Wachstumsmodells entspricht, ist der negative Effekt der Wachstumsrate der Bevölkerung überraschend. In der neoklassischen Wachstumstheorie erhöht sich durch Bevölkerungswachstum das Angebot an Arbeitskräften und somit die gesamtwirtschaftliche Produktion.

b) Die Variable „durchschnittliche Ausbildung in Jahren für 1985“ hat einen positiven und signifikanten Einfluss auf das Bruttoinlandsprodukt. Steigt die durchschnittliche Ausbildung der Bevölkerung um ein Jahr, erhöht sich das BIP um $0,055$ Einheiten.
Durch Hinzunahme dieser Variablen wird der Koeffizient der Wachstumsrate der Bevölkerung mehr als halbiert. Auch der positive Effekt der Sparrate verringert sich. Der Koeffizient ist zwar weiterhin auf dem 95%-Niveau signifikant, jedoch nicht mehr dem 99%-Signifikanzlevel.
Das R^2 des geschätzten Models steigt von $0,621$ auf $0,775$, sprich der Anteil der erklärten Varianz steigt um $15,4$ Prozentpunkte.
Eine mögliche Erklärung für dieses Ergebnis könnte sein, dass ein hohes Bevölkerungswachstum die durchschnittliche Ausbildung in Jahren pro Person reduziert hat, da die Bildungsinstitutionen überlastet sind.
c) Setzt man die Werte für Brasilien in die geschätzte Regression ein, so ergibt sich ein BIP pro Kopf von $0,24$. Das BIP wird für Brasilien folglich unterschätzt. Eine Verdopplung der Ausgaben würde zu einem geschätzten BIP von $0,43$ führen.

Übungsaufgabe 5.9

a) Die Schätzergebnisse besagen, dass ein Student ohne Lernen im Durchschnitt $25,87$ Punkte in der Klausur erreichen würde (β_0). Jede zusätzliche Stunde erhöht die Punktzahl im Schnitt um $3,45$ Punkte (β_1).
b) Die Teststatistik für den Test, ob der Lernaufwand einen signifikanten Effekt auf das Klausurergebnis hat ($H_0 : \beta_1 = 0$, $H_1 : \beta_1 \neq 0$), lautet:

$$t = \left| \frac{3,45 - 0}{0,521} \right| = |6,622|.$$

Bei einem Signifikanzniveau von 95% (99%) beträgt der kritische t-Wert bei 28 Freiheitsgraden ($N-(K+1)$) $1,701$ ($2,763$). Da unsere Teststatistik größer als der kritische Wert ist, können wir mit einer Irrtumswahrscheinlichkeit von 5% (1%) die Nullhypothese, dass Lernen das Testergebnis nicht beeinflusst, verwerfen.
c) Die t-Statistik für die Nullhypothese, dass das durchschnittliche Klausurergebnis der Studenten in Wirklichkeit größer ist ($H_0 : \beta_0 = 50$), gegen die Alternativhypothese $H_1 : \beta_0 \neq 50$ sieht folgendermaßen aus:

$$t = \left| \frac{25,87 - 50}{6,196} \right| = |-3,894|.$$

Da die Teststatistik größer als der kritische Wert der t-Verteilung ist ($3,894 > 2,763$) kann die Nullhypothese mit einer Irrtumswahrscheinlichkeit von 1% verworfen werden, β_0 ist folglich signifikant von 50 verschieden.

d) Die erwartete Anzahl von Punkten in der Klausur für einen Lernaufwand von 9,5 Stunden beträgt:

$$E(Y|X = 9,5) = 25,87 + 3,45 \times 9,5 = 58,645.$$

e) Das R^2 beträgt 0,15. Dies bedeutet, dass 15% der Variation der Klausurergebnisse durch den Lernaufwand der Studenten in Stunden erklärt werden können.

Übungsaufgabe 5.10

a) Die Schätzergebnisse besagen, dass sich die Größe eines Studenten für jeden *cm*, den die Eltern im Durchschnitt größer sind, um 0,73 *cm* erhöht. Die Konstante hat keine Interpretation.

b) (i) Wenn man erwarten würde, dass Kinder im Durchschnitt die gleiche Größe hätten wie ihre Eltern, dann wäre die Nullhypothese für den Achsenabschnitt, dass die Konstante gleich Null ist ($H_0 : \beta_0 = 0$). Die Alternativhypothese ist $H_1 : \beta_0 \neq 0$.
Die Teststatistik sieht folgendermaßen aus:

$$|t| = \left| \frac{49,9 - 0}{7,2} \right| = 6,931.$$

Die Teststatistik überschreitet den kritischen t-Wert ($t^*_\infty = 2,579$) und die Nullhypothese, dass die Konstante gleich Null, ist kann mit einer Irrtumswahrscheinlichkeit von 1% verworfen werden.

(ii) Die Nullhypothese für den Steigungskoeffizienten lautet $H_0 : \beta_1 = 1$. Man erwartet, dass wenn die Eltern im Schnitt einen *cm* größer sind, auch die Kinder einen *cm* größer sind. Dies wird gegen die Alternativhypothese $H_1 : \beta_1 \neq 1$ getestet.

$$|t| = \left| \frac{0,73 - 1}{0,10} \right| = 2,7$$

Bei einem Signifikanzniveau von 95% ist der kritische t-Wert bei 498 Freiheitsgraden $t^*_\infty = 1,960$. Die Nullhypothese, dass Kinder im Durchschnitt die gleiche Größe haben wie ihre Eltern, kann mit einer Irrtumswahrscheinlichkeit von 5% abgelehnt werden.

c) Die vorhergesagte Größe für ein Kind, dessen Eltern eine durchschnittliche Größe von 170 *cm* haben, beträgt 174 *cm*.

$$Student = 49,9 + 0,73 \cdot 170 = 174$$

d) Der Steigungskoeffizient gibt an, wie sehr die Größe der Eltern die Größe der Kinder beeinflusst. Je größer der Absolutwert des Steigungskoeffizienten ist, desto stärker ist der Einfluss.

Die geschätzte Gleichung besagt, dass bei einer durchschnittlichen Größe der Eltern von $184,81\ cm$, Eltern und Kinder gleich groß sind ($Student = Eltern$). Bis zu dieser Größe sind die Kinder im Schnitt größer als ihre Eltern und ab diesem Punkt sind die Kinder im Schnitt kleiner. Diese bedeutet, dass Kinder die relativ große Eltern haben, im Schnitt kleiner sind als ihre Eltern und Kinder mit relativ kleinen Eltern im Schnitt größer sind als ihre Eltern.
Große Eltern werden dennoch im Durchschnitt große Kinder haben (wenn diese auch nicht so groß sein werden wie ihre Eltern) und kleine Eltern werden im Durchschnitt kleine Kinder haben (wobei diese größer sein werden als ihre Eltern).

Übungsaufgabe 5.11

a)

$$\min_{\beta_1} \sum_{i=1}^{N} \hat{\epsilon}_i^2 = \sum_{i=1}^{N} [Y_i - (\hat{\beta}_1 X_i)]^2$$

$$\frac{\partial \sum_{i=1}^{N} \hat{\epsilon}_i^2}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{N} (Y_i - \hat{\beta}_1 X_i) X_i = 0$$

Durch Multiplikation mit $(-\frac{1}{2N})$ erhält man:

$$\frac{1}{N} \sum_{i=1}^{N} Y_i X_i - \hat{\beta}_1 \sum_{i=1}^{N} X_i X_i = 0$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} Y_i X_i}{\sum_{i=1}^{N} X_i X_i}.$$

b) Ist die Varianz des Störterms konstant, so spricht man von *Homoskedastizität*. Variiert des Störmerms mit der Höhe der Variablen X_i, liegt *Heteroskedastizität* vor.

c) Im Falle von Heteroskedastizität ist eine der fünf „klassischen Annahmen“ des bivariaten linearen Regressionsmodells verletzt. Die Kleinst-Quadrate-Methode würde in diesem Fall nicht zu effizienten Schätzern führen.

d)

$$E(Y_i|X_i = 2) = E(0,5 \times 2 + \epsilon_i) = E(1) + E(\epsilon_i) = 1 + E(\epsilon_i) = 1$$

e) $\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} \hat{\epsilon}_i^2}{N-2}$ ist ein erwartungstreuer Schätzer für σ^2
Die quadrierten Residuen berechnen sich als

$$\begin{aligned} \hat{\varepsilon}_i &= y_i - \hat{y}_i \\ &= y_i - \hat{\beta}_1 x_i \\ &= \beta_1 x_i + \eta_i - \hat{\beta}_1 x_i \\ &= (\beta_1 - \hat{\beta}_1) x_i + \eta_i. \end{aligned} \tag{1.15}$$

Quadriert man Gleichung (1.16), erhält man

$$\begin{aligned}\hat{\varepsilon}_i^2 &= (\beta_1 - \hat{\beta}_1)^2 x_i^2 + \eta_i^2 + 2(\beta_1 - \hat{\beta}_1) x_i \eta_i \\ &= (\hat{\beta}_1 - \beta_1)^2 x_i^2 + \eta_i^2 - 2(\hat{\beta}_1 - \beta_1) x_i \eta_i.\end{aligned} \tag{1.16}$$

Aufsummieren über alle N Beobachtungen ergibt

$$\sum_{i=1}^{N} \varepsilon_i^2 = (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N} \eta_i^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{N} x_i \eta_i. \tag{1.17}$$

Gleichung (1.17) ausgedrückt in Erwartungen lautet

$$\begin{aligned}E\left[\sum_{i=1}^{N} \hat{\varepsilon}_i^2\right] &= E\left[(\hat{\beta}_1 - \beta_1)^2\right] \sum_{i=1}^{N} x_i^2 + E\left[\sum_{i=1}^{N} \eta_i^2\right] \\ &\quad -2E\left[(\hat{\beta}_1 - \beta_1) \sum_{i=1}^{N} x_i \eta_i\right].\end{aligned} \tag{1.18}$$

Hierbei gilt

$$E\left[(\hat{\beta}_1 - \beta_1)^2\right] = Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \tag{1.19}$$

und (vgl. Kapitel 2)

$$E\left[\sum_{i=1}^{N} \eta_i^2\right] = E\left[\sum_{i=1}^{N} (\varepsilon_i - \overline{\varepsilon})^2\right] = \sigma^2 (N-1) \tag{1.20}$$

sowie

$$\begin{aligned}(\hat{\beta}_1 - \beta_1) &= \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} - \beta_1 \\ &= \frac{\sum_{i=1}^{N} x_i (\beta_1 x_i + \eta_i)}{\sum_{i=1}^{N} x_i^2} - \beta_1 \\ &= \frac{\beta_1 \sum_{i=1}^{N} x_i^2}{\sum_{i=1}^{N} x_i^2} + \frac{\sum_{i=1}^{N} x_i \eta_i}{\sum_{i=1}^{N} x_i^2} - \beta_1 \\ &= \frac{\sum_{i=1}^{N} x_i \eta_i}{\sum_{i=1}^{N} x_i^2}\end{aligned}$$

bzw.

$$\sum_{i=1}^{N} x_i \eta_i = (\hat{\beta}_1 - \beta_1) \sum_{i=1}^{N} x_i^2. \tag{1.21}$$

Setzt man die Gleichungen (1.19) - (1.21) in Gleichung (1.18) ein, erhält man

$$\begin{aligned}
E\left[\sum_{i=1}^{N} \hat{\varepsilon}_i^2\right] &= \frac{\sigma^2}{\sum_{i=1}^{N} x_i^2} \sum_{i=1}^{N} x_i^2 + \sigma^2(N-1) - 2E\left[(\hat{\beta}_1-\beta_1)(\hat{\beta}_1-\beta_1)\sum_{i=1}^{N} x_i^2\right] \\
&= \sigma^2 + \sigma^2(N-1) - 2E\left[(\hat{\beta}_1-\beta_1)^2 \sum_{i=1}^{N} x_i^2\right] \\
&= \sigma^2 + \sigma^2(N-1) - 2\sum_{i=1}^{N} x_i^2 \underbrace{E\left[(\hat{\beta}_1-\beta_1)^2\right]}_{Var(\hat{\beta}_1)=\frac{\sigma^2}{\sum_{i=1}^{N} x_i^2}} \\
&= \sigma^2 + \sigma^2(N-1) - 2\sigma^2 \\
&= -\sigma^2 + \sigma^2(N-1) \\
&= \sigma^2(N-2). \qquad (1.22)
\end{aligned}$$

f) Angenommen der Erwartungswert der Störgröße sei nicht Null, sondern würde den Wert einer Konstanten k annehmen ($E(\varepsilon_i) = k$). In diesem Fall lässt sich das Modell so transformieren, dass $E(\varepsilon_i^*) = 0$ wieder gilt.

$$\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + \varepsilon_i \\
&= \underbrace{\beta_0 + k}_{\beta_0^*} + \beta_1 X_i + \underbrace{\varepsilon_i - k}_{\beta_0^*} \\
&= \beta_0^* + \beta_1 X_i + \varepsilon_i^*.
\end{aligned}$$

Lediglich der Achsenabschnitt β_0 nicht aber der Steigungskoeffizient β_1 wäre betroffen.
Die Annahme, dass der Erwartungswert der Störterme ε_i Null ist besagt, dass die Störgröße keinen systematischen Beitrag zur Erklärung der abhängigen Variablen besitzt. Wichtig ist, dass der Erwartungswert existiert und konstant ist. Die Annahme, dass der Erwartungswert Null ist, ist lediglich eine Möglichkeit.

Übungsaufgabe 5.12

a) Eine mögliche kontrafaktische Fragestellung des Modells ist, wie die Brutto-Jahreslöhne mit einem Jahr mehr oder weniger Ausbildung bei gegebenem Alter variieren. Andersherum könnte man fragen, was für einen Effekt eine gegebene Anzahl von Ausbildungsjahren in verschiedenen Altersstufen hat.
b) Von „statistischer Signifikanz“ spricht man, wenn die Wahrscheinlichkeit, dass eine Variable einen bestimmten Wert nur durch Zufall annimmt, gering ist. Als Wahrscheinlichkeitsgrenze werden zumeist 1%, 5% oder 10% verwendet.

c) Um Signifikanz zu testen, braucht man eine Vermutung darüber, wie groß der Koeffizient tatsächlich ist.
Es wird fast immer getestet, ob der Koeffizient signifikant verschieden von Null ist. Das heißt man testet bei gegebener Irrtumswahrscheinlichkeit, ob der geschätzte (in diesem Falle positive) Einfluss nur Zufall ist.
Die Teststatistiken für einen zweiseitigen Test der jeweiligen Koeffizienten lauten:

$$|t| = \left| \frac{0,06}{0,007} \right| = 8,571$$
$$|t| = \left| \frac{0,07}{0,009} \right| = 7,778$$

In beiden Fällen ist die Teststatistik größer als der kritische t-Wert ($t^*_{2764} = 1,96$) bei einem Signifikanzniveau von 95%. Die Wahrscheinlichkeit, die geschätzten Koeffizienten zu erhalten, obwohl der tatsächliche Wert Null ist, ist unter 5%. Dies bedeutet, dass man sich mit einer 5% Wahrscheinlichkeit irrt, wenn man die Nullhypothese dass der wahre Koeffizient gleich Null ist, verwirft.

d) Das Bestimmtheitsmaß R^2 misst den Anteil der durch das Modell erklärten Variation an der gesamten Variation in Y.
Im vorliegenden Modell lassen sich 20% der Variation in den logarithmierten Brutto-Jahreslöhnen durch Unterschiede in der Ausbildungsdauer und Unterschiede im Alter der Individuen erklären.

e)

$$\ln(w_i) = 6,5 + 0,06 \times 18 + 0,07 \times 30 = 9,68$$
$$w_i = \exp(\ln(w_i)) = \exp(9,68) = 15.994,50$$

Laut Model würde man für eine 30-jährige Person mit 18 Jahren Schulausbildung einen Brutto-Jahreslohn von $15.994,50,-$ schätzen.

f) Die Homoskedastizitätsannahme besagt, dass der Erwartungswert des Störterms für alle Beobachtungseinheiten konstant ist. Sie wäre in dem vorliegenden Modell verletzt, wenn die Varianz von der Länge der Schulausbildung oder vom Alter abhängen würde.
Wenn z.B. die Varianz der Löhne für ältere Menschen viel größer wäre, läge Heteroskedastizität vor.

g) Unter *Erwartungstreue* (oder auch Unverzerrtheit) versteht man die Tatsache, dass die Erwartungswerte der Schätzer ($\hat{\beta}$) den tatsächlichen Populationsparametern (β) entsprechen, d.h. $E(\hat{\beta}) = \beta$. Man kann sich also vorstellen, dass man, würde man die Schätzung sehr häufig wiederholen, zumindest im Schnitt richtig liegen würde.
Das Konzept von *Konsistenz* liegt sehr nahe an dem der Erwartungstreue. Die Idee hinter Konsistenz ist, dass eine Erhöhung des Stichprobenumfangs wenigstens irgendwann zu einer Annäherung an den tatsächlichen Parameter führen sollte und die Schätzunsicherheit vernachlässigbar werden könnte. Dies lässt sich auch als $plim \hat{\beta}_N = \beta$ schreiben.

Effizienz bedeutet, dass die Varianz des Schätzers $\hat{\beta}$ möglichst gering ist. Je geringer die Varianz ist, desto höher ist die Präzision mit der β geschätzt werden kann und desto geringer ist die Varianz σ^2 des Störterms.