Ökonometrie Grundlagen, Übung 3

Henry Haustein

Aufgabe 1

Wir diesen Beweis brauchen wir noch mal die Formeln zur Schätzung der Parameter β_0 und β_1 für lineare Modelle:

$$\begin{aligned}
\beta_0 &= \bar{y} - \beta_1\bar{x} \\
\beta_1 &= \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}
\end{aligned}$$

Ferner brauchen wir die Definition von $\hat{y} = \beta_0 + \beta_1 x$. Jetzt wenden wir uns dem eigentlichen Beweis zu. Wir werden zeigen, dass $R^2 = r_{xy}^2$, was gleichbedeutend mit $r_{xy} = \sqrt{R^2}$ ist.

$$\begin{aligned}
R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \frac{\sum_{i=1}^n (\beta_0 + \beta_1 x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \frac{\sum_{i=1}^n (\bar{y} - \beta_1\bar{x} + \beta_1 x_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \beta_1^2 \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \left(\frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)^2 \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \frac{\left[\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})\right]^2}{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right]^2} \frac{\sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (y_i - \bar{y})^2} \\
&= \frac{\left[\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})\right]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2} \\
&= r_{xy}^2
\end{aligned}$$

Aufgabe 2

Interpretation des theoretischen KI's: Der wahre Parameter der GG liegt mit einer Wahrscheinlichkeit von $1-\alpha$ innerhalb des Intervalls.

Interpretation des numerischen KI's: Da das numerische bzw. realisierte KI eine Ausprägung der Zufallsvariable Konfidenzintervallschätzer ist (und damit aus 2 Zahlen, der linken und der rechten Grenze, besteht), liegt der Parameter innerhalb dieses Intervalls oder nicht. Die Wahrscheinlichkeit ist demnach 1 (innerhalb) oder 0 (außerhalb). Je höher der Stichprobenumfang T, desto näher liegt die Überdeckungwahrscheinlichkeit der Konfidenzintervalle beim Konfidenzniveau $(1-\alpha)100$.

Aufgabe 3

(a) Bei einer Lageverschiebung verschiebt sich das Konfidenzintervall von β_i mit; das Konfidenzintervall von σ_u^2 sollte sich nicht verändern, da sich die Varianz der Daten nicht ändert

(b) Das Konfidenzintervall von β_i wird größer, wenn α kleiner wird. Selbiges gilt auch für σ_u^2.

(c) Das Konfidenzintervall von β_i wird größer, wenn der standard error größer wird. Auch dies gilt für σ_u^2.

Aufgabe 4

In allen 3 Fällen muss zuerst die Teststatistik berechnet werden. Wenn σ_u^2 unbekannt ist, macht man einen sogenannten t-Test mit der folgenden Teststatistik:

$$t = \frac{\beta - \beta^*}{\mathrm{se}(\beta)} \sim t_{T-2}$$

Die Berechnung der kritischen Werte und Interpretation ist dann fallabhängig:

(a) zweiseitiger Test: Man berechnet $t_{T-2;1-\frac{\alpha}{2}}$ und wenn $|t| > t_{T-2;1-\frac{\alpha}{2}}$ gilt, dann lehnt man die Nullhypothese ab.

(b) + (c) einseitiger Test: Man berechnet $t_{T-2;1-\alpha}$ und wenn $|t| > t_{T-2;1-\alpha}$ gilt, dann lehnt man die Nullhypothese ab.

Aufgabe 5

Man berechnet das Konfidenzintervall zum gewünschten Signifikanzniveau und wenn β^* außerhalb dieses Intervalls liegt, lehnt man die Nullhypothese ab.

Aufgabe 6

Der p-value ist die minimale Irrtumswahrscheinlichkeit α^* bei gegebenem Wert der Teststatistik, bei der H_0 gerade noch abgelehnt werden kann.

Aufgabe 7

(a) ökonomisches Modell: $\log(y) = \beta_0 + \beta_1 \log(x)$
ökonometrisches Modell: $\log(y_t) = \beta_0 + \beta_1 \log(x_t) + u_t$

(b) in R

```
datensatz = read.csv2("Zigaretten.csv")
head(datensatz)
modell = lm(log(Zigarettenkonsum) ~ log(Realpreis), data = datensatz)
modell$coefficients
confint(modell)
```

$\beta_0 \in [4.975343, 5.2289130]$, $\hat{\beta}_0 = 5.102128$
$\log(\beta_1) \in [-1.796936, -0.6622263]$, $\widehat{\log(\beta)}_1 = -1.229581$

(c) Die Konfidenzintervalle werden größer

```
confint(modell,level=0.99)
```

$\beta_0 \in [4.932759, 5.2714968]$
$\log(\beta_1) \in [-1.987496, -0.4716662]$

(d) In R

```
summary(modell)
```

Der p-Value für β_0 ist $< 2 \cdot 10^{-16}$ und für β_1 $7.53 \cdot 10^{-5}$. Das bedeutet, dass die zugehörigen Nullhypothesen $\beta_0 = 0$ und $\beta_1 = 0$ abgelehnt werden. $R^2 = 0.3024$ was auf kein gutes Modell für diese Daten hindeutet.

Aufgabe 8

(a) ökonomisches Modell: Mineralölkonsum $= \beta_0 + \beta_1 \cdot$ Realeinkommen
ökonometrisches Modell: Mineralölkonsum$_t = \beta_0 + \beta_1 \cdot$ Realeinkommen$_t + u_t$

(b) In R

```
datensatz2 = read.csv2("Mineral.csv")
head(datensatz2)
plot(datensatz2$Realeinkommen,datensatz2$Mineraloelkonsum)
modell2 = lm(Mineraloelkonsum ~ Realeinkommen, data = datensatz2)
summary(modell2)
confint(modell2)
```

$\beta_0 \in [13.29841484, 18.58500742]$, $\hat{\beta}_0 = 15.94$
$\beta_1 \in [0.01987696, 0.02198638]$, $\hat{\beta}_1 = 2.093 \cdot 10^{-2}$