Aufgaben zum Exkurs Signifikanztests über einzelne Regressionskoeffizienten

E1. Wie groß ist der Standardfehler der Regression des Gasverbrauchs (GASV) auf das verfügbare Einkommen (VEINKR) ($\hat{=}$ Energiemodell Ib),

$$\underbrace{\widehat{GASV}_t}_{\hat{y}_t} = -0{,}433 + 0{,}119 \cdot \underbrace{VEINK_t}_{x_t},$$

(mit Interpretation)?

Lösung:

Der **Standardfehler der Regression** (SER) ist gleich der **Wurzel aus der Residualvarianz**, bei der die Summe der quadrierten Residuen auf die Anzahl der Freiheitsgrade bezogen ist:

$$SER = \hat{\sigma} = \sqrt{\frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-k}}$$

Arbeitstabelle:

t	y_t	$\hat{y}_t$	$\hat{u}_t$	$\hat{u}_t^2$
1	1,0	0,995	0,005	0,000025
2	0,8	0,995	-0,195	0,038025
3	0,7	0,638	0,062	0,003844
4	0,9	0,876	0,024	0,000576
5	1,2	1,114	0,086	0,007396
6	1,1	1,114	-0,014	0,000196
7	1,4	1,352	0,048	0,002304
Σ	7,1	≈7,1		0,051826

Residualvarianz:

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-k} = \frac{0{,}051826}{7-2} = \frac{0{,}051826}{5} = 0{,}010365$$

Standardfehler der Regression (SER)**:**

$$SER = \hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{0{,}010365} = 0{,}102$$

<u>Interpretation:</u>

Der Standardfehler der Regression ist die Wurzel aus der erwartungstreu geschätzten Varianz der Störgröße. Er gibt die mittlere Abweichung der y-Werte von den Regressionswerten $\hat{y}_t$ an. Bei der Regressionsschätzung des Gasverbrauchs in Abhängigkeit vom verfügbaren Einkommen wird danach ein durchschnittlicher Fehler von 0,102 Einheiten gemacht.

E2. Welche Aussage über die Güte der Anpassung des Energiemodells Ib (s. Aufg. E1) an den Rändern, d.h. das erste und letzte Jahr des Stützbereichs, lässt sich anhand der Residuen treffen? Wie groß ist die Güte der Anpassung im Stützbereich insgesamt?

Lösung:

Die Güte der Anpassung *in einer einzelnen Periode t* wird über die prozentuale Verhältniszahl

$$\frac{|\hat{u}_t|}{|y_t|} \cdot 100\%$$

ermittelt. Die Verhältniszahl ist bei einer optimalen Anpassung des Datenpunktes null, da dann der y-Wert mit dem Regressionswert $\hat{y}_t$ übereinstimmen würde. Je geringer der Wert, desto besser ist die Anpassung.

Güte der Anpassung bei t = 1:

$$\frac{|\hat{u}_1|}{|y_1|} \cdot 100\% = \frac{|0{,}005|}{|1{,}0|} \cdot 100\% = 0{,}005 \cdot 100\% = 0{,}5\%$$

Der relative Fehler in der ersten Periode berägt 0,5%.

Güte der Anpassung bei t = 7:

$$\frac{|\hat{u}_7|}{|y_7|} \cdot 100\% = \frac{|0{,}048|}{|1{,}4|} \cdot 100\% = 0{,}034 \cdot 100\% = 3{,}4\%$$

Der relative Fehler in der letzten Periode beträgt 3,4%.

Die Anpassung ist in der ersten Periode deutlich besser als in der letzten Periode, d.h. der Gasverbrauch lässt sich in der ersten Periode erheblich besser durch das verfügbare Einkommen erklären als in der letzten Periode.

Als Maß für die Güte der Anpassung *im gesamten Stützzeitraum* dient der Determinationskoeffizient bzw. das Bestimmtheitsmaß R^2, das das Verhältnis der Streuung von $\hat{y}_t$ zur Streuung von y_t angibt.

Arbeitstabelle:

t	y_t^2	$\hat{y}_t^2$
1	1,000	0,990
2	0,640	0,990
3	0,490	0,407
4	0,810	0,767
5	1,440	1,241
6	1,210	1,241
7	1,960	1,828
$\sum$	7,550	7,464

Arithmetisches Mittel (s. Arbeitstabelle zu E1.):

$$\bar{y} = \frac{1}{n} \cdot \sum_{t=1}^{n} y_t = \frac{1}{7} \cdot 7{,}1 = 1{,}014$$

Determinationskoeffizient (Bestimmtheitsmaß):

$$R^2 = \frac{\sum_{t=1}^{n} \hat{y}_t^2 - n \cdot \bar{y}^2}{\sum_{t=1}^{n} y_t^2 - n \cdot \bar{y}^2} = \frac{\sum_{t=1}^{7} \hat{y}_t^2 - 7 \cdot 1{,}014^2}{\sum_{t=1}^{7} y_t^2 - 7 \cdot 1{,}014^2} = \frac{7{,}464 - 7 \cdot 1{,}014^2}{7{,}550 - 7 \cdot 1{,}014^2} = \frac{0{,}267}{0{,}353} = 0{,}756$$

Das lineare Modell für den Gasverbrauch in Abhängigkeit vom verfügbaren Einkommen erklärt damit die Streuung des Gasverbrauchs im betrachteten 7-Jahres-Zeitraum zu 75,6%.

E3. Die OLS-Schätzer der Regressionskoeffizienten β_1 und β_2 der Nachfragefunktion für Erdgas (≙ Energiemodell Ia (Gasverbrauch in Abhängigkeit vom Gaspreis)) lauten (n=7): $\hat{\beta}_1 = 5{,}809$ und $\hat{\beta}_2 = -0{,}368$. Die Summe der quadrierten Residuen beträgt 0,067, die Summe der Gaspreise 91,2 und die Summe der quadrierten Gaspreise 1190,28.

Testen Sie die OLS-Schätzer der Regressionskoeffizienten auf Signifikanz (α=0,01)!

Lösung:

A. Signifikanztests (zweiseitig) über $\hat{\beta}_1$

1. Schritt	Hypothesenformulierung: H_0: $ß_1 = 0$, H_1: $ß_1 \neq 0$
2. Schritt	Festlegung des Signikanzniveaus: $\alpha = 0{,}01$
3. Schritt	Berechnen der Prüfgröße: Residualvarianz: $\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-k} = \frac{0{,}067}{7-2} = 0{,}0134$ Geschätzte Varianz von $\hat{\beta}_1$: $\hat{Var}(\hat{\beta}_1) = \hat{\sigma}^2 \frac{\sum x_t^2}{n\sum x_t^2 - (\sum x_t)^2} = 0{,}0134 \cdot \frac{1190{,}28}{7 \cdot 1190{,}28 - 91{,}2^2}$ $= 0{,}0134 \cdot 81{,}975 = 1{,}0985$ Prüfgröße: $t = \frac{\hat{\beta}_1}{\sqrt{\hat{Var}(\hat{\beta}_1)}} = \frac{5{,}809}{\sqrt{1{,}0985}} = \frac{5{,}809}{1{,}048} = 5{,}543$

4. Schritt	Kritischer Wert: $t_{n-k;1-\alpha/2} = t_{7-2;0,995} = t_{5;0,995} = 4{,}032$
5. Schritt	Testentscheidung: $\lvert t \rvert = 5{,}543 > t_{5;0,995} = 4{,}032$ → Nullhypothese ablehnen

B. Signifikanztest (zweiseitig) über $\hat{\beta}_2$:

1. Schritt	Hypothesenformulierung: H_0: $ß_2 = 0$, H_1: $ß_2 \neq 0$
2. Schritt	Festlegung des Signikanzniveaus: $\alpha = 0{,}01$
3. Schritt	Berechnen der Prüfgröße: Residualvarianz: $\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} \hat{u}_t^2}{n-k} = \frac{0{,}067}{7-2} = 0{,}0134$ Geschätzte Varianz von $\hat{\beta}_2$: $\hat{Var}(\hat{\beta}_2) = \hat{\sigma}^2 \frac{n}{n\sum x_t^2 - (\sum x_t)^2} = 0{,}0134 \cdot \frac{7}{7 \cdot 1190{,}28 - 91{,}2^2}$ $= 0{,}0134 \cdot 0{,}4821 = 0{,}00646$ Prüfgröße: $t = \frac{\hat{\beta}_2}{\sqrt{\hat{Var}(\hat{\beta}_2)}} = \frac{-0{,}368}{\sqrt{0{,}00646}} = \frac{-0{,}368}{0{,}0804} = -4{,}577$
4. Schritt	Kritischer Wert: $t_{n-k;1-\alpha/2} = t_{7-2;0,995} = t_{5;0,995} = 4{,}032$
5. Schritt	Testentscheidung: $\lvert t \rvert = 4{,}577 > t_{5;0,995} = 4{,}032$ → Nullhypothese ablehnen