

## Methoden Verkehrsökometrie für Master-Studierende

Winter semester 2021/22, Solutions to Tutorial No. 3

### Lösungsvorschlag zu Aufgabe 3.1: Hotelauslastung (Mehrfachregression mit Matrizen und Vektoren)

- (a) Die Matrix  $\underline{\underline{X}}$  beinhaltet in der 1. Spalte die Konstanten 1, während die 2. und 3. Spalte mit den Beobachtungen der exogenen, erklärenden Größen gefüllt wird:  $x_{i1}$  = Anzahl der Sterne bzw.  $x_{i2}$  = Preis im Hotel  $i$ . Die Beobachtungen der abhängigen Größe, der Hotelauslastung, werden durch den Vektor  $\vec{y}$  beschrieben. Es ergeben sich

$$\underline{\underline{X}} = \begin{pmatrix} 1 & 1 & 15 \\ 1 & 1 & 31 \\ 1 & 1 & 40 \\ 1 & 2 & 34 \\ 1 & 2 & 50 \\ 1 & 2 & 58 \\ 1 & 3 & 67 \\ 1 & 3 & 72 \\ 1 & 3 & 84 \\ 1 & 4 & 82 \\ 1 & 4 & 98 \\ 1 & 4 & 116 \end{pmatrix} \quad \text{und} \quad \vec{y} = \begin{pmatrix} 42 \\ 38 \\ 24 \\ 76 \\ 52 \\ 40 \\ 90 \\ 77 \\ 62 \\ 90 \\ 82 \\ 68 \end{pmatrix}. \quad (1)$$

- (b) Es ergeben sich

$$\underline{\underline{X}}'\underline{\underline{X}} = \begin{pmatrix} 12 & 30 & 747 \\ 30 & 90 & 2223 \\ 747 & 2223 & 56319 \end{pmatrix}$$

und mit der Inversionsformel für  $3 \times 3$ -Matrizen,

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}^{-1} = \frac{1}{aei + bfg + cdh - afh - bdi - ceg} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix},$$

die Inverse

$$[\underline{\underline{X}}'\underline{\underline{X}}]^{-1} = 10^{-3} \cdot \begin{pmatrix} 506.5 & -115.6 & -2.154 \\ -115.6 & 469.9 & -17.01 \\ -2.154 & -17.02 & 0.718 \end{pmatrix}$$

und

$$\underline{\underline{X}}'\underline{\underline{y}} = \begin{pmatrix} 741 \\ 2087 \\ 50358 \end{pmatrix} \quad (2)$$

und letztlich

$$\hat{\underline{\underline{\beta}}} = [\underline{\underline{X}}'\underline{\underline{X}}]^{-1}\underline{\underline{X}}'\underline{\underline{y}} \quad (3)$$

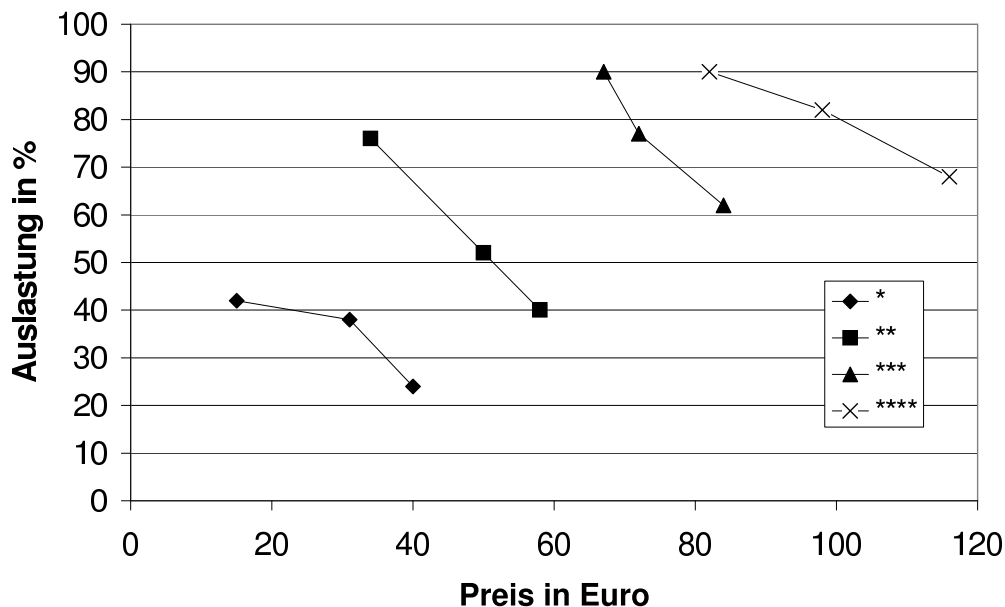
$$= \begin{pmatrix} 25.5 \\ 38.2 \\ -0.953 \end{pmatrix}. \quad (4)$$

*Hinweis:* In der Klausur müssen Sie keine  $3 \times 3$ -Matrix invertieren.

- (c)
- $\beta_0=25.5$ : Keine sinnvolle Aussage ("ein kostenloses Hotel mit null Sternen ist zu 25.5% ausgelastet"). Hier ist die Anwendungsgrenze des linearen Modells überschritten.
  - $\beta_1 = \partial y / \partial x_1 = 38.2$ : Bei gleichbleibendem Hotelpreis steigt die Hotelauslastung je zusätzlichen Stern um 38,2%.
  - $\beta_2=-0.953$ : Bei unveränderter Sternenkategorie sinkt die Hotelauslastung um 0.953%, wenn zusätzlich 1 Euro mehr pro Nacht verlangt wird.
  - Schließlich kann man noch den erwarteten Wert eines Sterns abschätzen, indem man eine die Auslastungsänderung durch eine Erhöhung der Sternezahl um  $\Delta x_1 = 1$  durch eine entsprechende Preisänderung  $\Delta x_2$  ausgleicht:

$$\Delta y \stackrel{!}{=} 0 = \beta_1 + \beta_2 \Delta x_2 \quad \Rightarrow \quad \Delta x_2 = -\beta_1 / \beta_2 = 40.1 \text{ Euro.}$$

- (d) Die Korrelation zwischen Auslastung und Preis darf nicht als Preisempfindlichkeit interpretiert werden, da eine indirekte Abhängigkeit zwischen Sternenzahl und Preis besteht. Das multiple Modell löst diese indirekte Abhängigkeit auf und erlaubt den plausiblen Schluss: Innerhalb einer Sternenkategorie sinkt die Auslastung mit steigendem Preis. Nicht die Preise wirken anziehend, sondern die Sternezahl. Der höhere Preis wird in Kauf genommen, aber nicht angestrebt. Folgende Abbildung visualisiert den Sachverhalt:



Jede durchgezogene Kurve entspricht dabei eine Sternenkategorie.

(e) Test von  $\beta_1$  auf  $\beta_1 = 0$ :

(i) Nullhypothese:  $H_{01} : \beta_1 = 0$

(ii) Test-Funktion und ihre Verteilung (falls alle statistischen Gauß-Markow-Annahmen zutreffen und die Nullhypothese (ggf grenzwertig) erfüllt ist):

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{\sqrt{\hat{V}(\hat{\beta}_1)}} \sim T(n - 3) = T(9)$$

(iii) Realisierung:

$$t = \frac{38.2 - 0}{\sqrt{26.0}} = 7.49$$

(iv) Entscheidung des Tests auf Gleichheit:  $H_{01}$  unter einer Fehlerwahrscheinlichkeit von  $\alpha = 5\%$  ablehnbar, falls

$$|t| > t_{1-\alpha/2}^{(9)} = t_{0.975}^{(9)} = 2.262 \Rightarrow \text{ist erfüllt!} \quad (5)$$

Alternativ kann der  $p$ -Wert

$$p = \min(\alpha \mid H_0 \text{ kann bei Fehlerwahrscheinlichkeit } \alpha \text{ abgelehnt werden})$$

direkt durch Ablesen der Verteilungsfunktion  $F_{T(9)}(y)$  der Student-t-Verteilung  $T(9)$  an einer geeigneten Stelle  $y$  abgelesen werden, indem man ausnutzt, dass die Quantilsfunktion die Umkehrfunktion der Verteilungsfunktion ist:

$$F_{T(9)}\left(t_y^{(9)}\right) = y$$

(die Quantilsfunktion ist die Umkehrfunktion der Verteilungsfunktion!) Die Grenzbedingung für  $p$  ergibt sich, indem man Gl. (5) als Gleichung auffasst:

$$\begin{aligned} |t| &= t_{1-p/2}^{(9)} \\ F_{T(9)}(|t|) &= 1 - \frac{p}{2} \\ p &= 2(1 - F_{T(9)}(|t|)) \ll 1\% \Rightarrow \text{ist erfüllt!} \end{aligned}$$

Analog Test von  $\beta_2$  auf den Wert  $-1.5$ :

(i) Nullhypothese:  $H_{02} : \beta_2 < -1.5$

(ii) Test-Funktion

$$T \sim T(n-3) = T(9)$$

(iii) Realisierung:

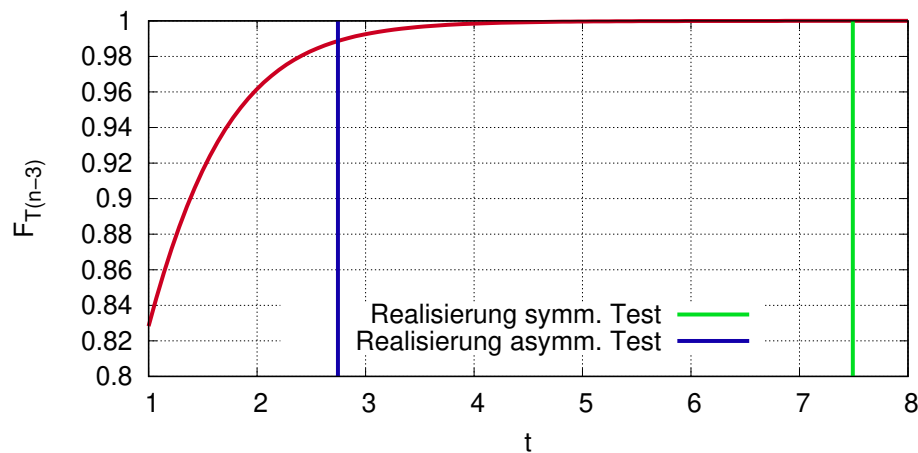
$$t = \frac{-0.953 - (-1.5)}{\sqrt{0.0397}} = 2.75$$

(iv) Entscheidung des Tests auf "Kleiner-Gleich":  $H_{02}$  ablehnbar, falls

$$t > t_{1-\alpha}^{(9)} = 1.83 \Rightarrow \text{ist erfüllt!}$$

Alternativ  $p$ -Wert durch Umformung und Ablesen aus der Graphik:

$$p = 1 - F_{T(9)}(t) \approx 1\% \Rightarrow \text{ist erfüllt!}$$



(f) (i) Nullhypothese:  $\beta_1 = 30$  und  $\beta_2 = -0.5$ .

(ii) Testfunktion und ihre Verteilung: Das zur Nullhypothese gehörige restringierte Modell ist

$$y_{\text{restr}}(\vec{x}, \beta_0) = \beta_0 + 30x_1 - 0.5x_2 + \epsilon$$

Diese Restriktion hat zwei Parameter weniger als das volle Modell. Die Testfunktion lautet also

$$S = \frac{\frac{S_{\text{restr}} - S_{\text{min}}}{3-1}}{\frac{S_{\text{min}}}{n-3}} \sim F(2, n-3).$$

- (iii) Realisierung: Schätzung des einzigen verbleibenden Parameters  $\beta_0$  ergibt die bereits angegebenen Fehlerquadratsumme  $S_{\text{restr}} = \sum_i (y_i^{\text{restr}} - y_i)^2 = 878.6$ . Mit der ebenfalls gegebenen Fehlerquadratsumme

$$S_{\text{min}} = \sum_i (\hat{y}_i - y_i)^2 = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} - y_i)^2 = 498.2$$

des vollen Modells lautet die Realisierung

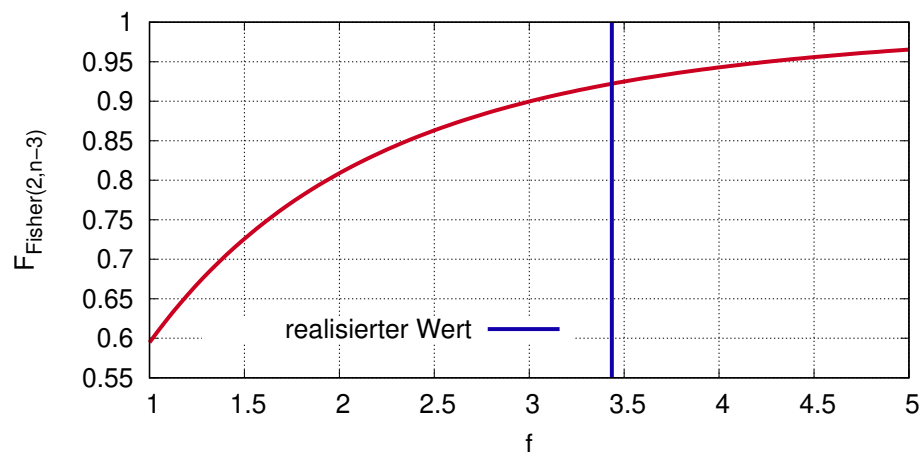
$$f = \frac{\frac{878.6 - 498.2}{2}}{\frac{498.2}{9}} = 3.44$$

- (iv) Entscheidung: Die verbundene Punkt-Nullhypothese ist beim F-Test bei einer Fehlerwahrscheinlichkeit  $\alpha = 5\%$  ablehnbar, falls

$$f = 3.44 > f_{1-\alpha}^{(2,9)} = 4.25.$$

Dies ist nicht erfüllt, sodass der Test nicht ablehnbar ist.

Dies sieht man auch durch den aus der Verteilungsfunktion der Fisher-Verteilung ablesbaren  $p$ -Wert aus der Abbildung:



$$p = 1 - F_{\text{Fisher}(2,9)}(3.44) \approx 8\% > \alpha$$

Folgende Abbildung veranschaulicht die Nullhypothese des verbundenen Tests und zeigt, zum Vergleich, die verbundene Konfidenzregion sowie die Konfidenzintervalle von  $\beta_1$  und  $\beta_2$ , jeweils getrennt betrachtet:

