Methods in Transportation Econometrics and Statistics (Master)

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

Solution to Problem 3.1: Booking rate of twelve hotels

(a) The first column of the system matrix $\underline{\underline{X}}$ contains the constants $x_{i0} = 1$ while the other columns contain the proper exogenous factors x_{i1}: quality proxy (# stars) of hotel i and x_{i2}: price single room/night [€] of hotel i. The endogenous data y_i, of course, contain the booking rate [%] for hotel i:

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

(b) We have

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

and with an inversion formula for 3×3-Matrizen (*not required in the examination!*),

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

the inverse matrix

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

Furthermore,

$$\underline{\underline{X}}'\vec{y} = \begin{pmatrix} 741 \\ 2087 \\ 50358 \end{pmatrix} \tag{2}$$

and finally

$$\hat{\vec{\beta}} = [\underline{\underline{X}}'\underline{\underline{X}}]^{-1}\underline{\underline{X}}'\vec{y} = \begin{pmatrix} 25.5 \\ 38.2 \\ -0.953 \end{pmatrix}. \tag{3}$$

Hint: No inversion of 3×3-matrices will be required in the examination.

(c)
- β_0=25.5: No useful meaning ("a free zero-star hotel has a booking rate of 25.5 %"). The limits of the application range of this linear model is exceeded (which would be also the case for factor combinations such as a very low price and five stars leading to a modelled bookig rate above 100 %)
- $\beta_1 = \partial y/\partial x_1 = 38.2$: Keeping the price constant (*ceteris paribus*), the booking rate increases by 38.2 % for each additional star.
- $\beta_2 = \partial y/\partial x_2 = -0.953$: Keeping the star rating constant, the booking rate will decrease by 0.953 % per additional Euro per night.
- The value of one star (in terms of € per night) can be determined by keeping the modelled booking rate y constant while changing both price and star rating:

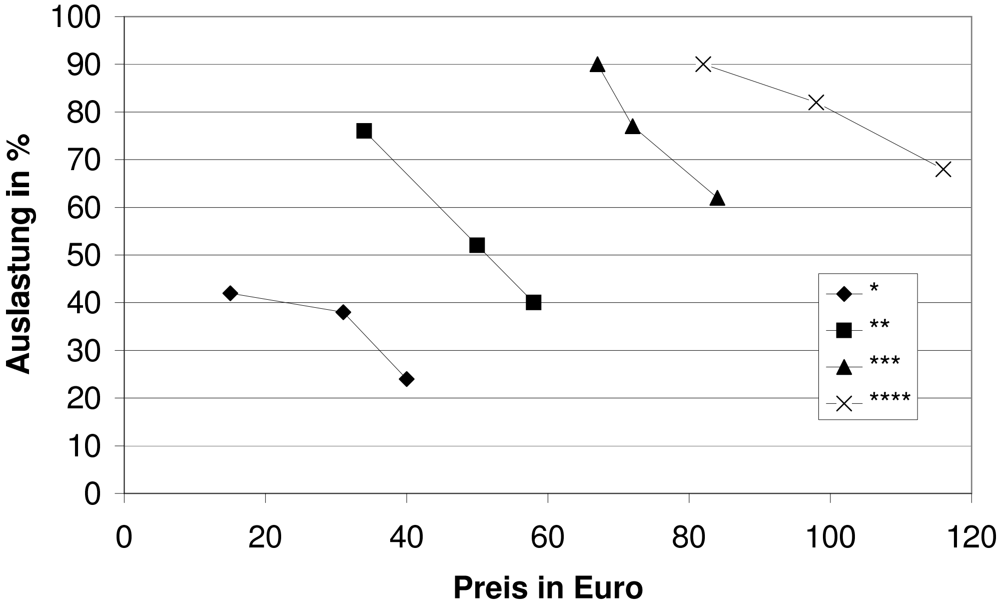
$$\Delta y = \beta_1 + \beta_2 \Delta x_2 \stackrel{!}{=} 0 \quad \Rightarrow \quad \Delta x_2 = -\beta_1/\beta_2 = 40.1 \text{ Euro}.$$

(d) The positive descriptive correlation between price and booking rate in the data may *not* be interpreted as a positive price sensitivity ("people love expensive hotels"). In fact, there is an indirect coupling via the quality (which is positively correlated with the price) and the quality appraisal of the guest: What one could say is that the quality appraisal prevails over the price sensitivity:

More expensive hotels tend to have a higher booking rate not because of, but in spite of, their higher price.

The multivariate model resolves this indirect dependence and allows for the correct conclusion. When ignoring price or quality, however, the most important of the functional specifications is violated and, consequently, one will have *junk in, junk out.*

Following figure clarifies this:



Each solid line links hotels of the same star category.

(e) (i) Null hypothesis $H_{01} : \beta_1 = 0$

(ii) Test function and their distribution if H_0 is true and additionally all the statistical Gauß-Markow requirements are satisfied:

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

(iii) Realisation:

$$t_{\text{data}} = \frac{38.2 - 0}{\sqrt{26.0}} = 7.49$$

(iv) Decision: H_{01} rejected at a significance level $\alpha = 5\,\%$ if

$$|t| > t^{(9)}_{1-\alpha/2} = t^{(9)}_{0.975} = 2.262 \;\Rightarrow\; \text{rejection} \tag{4}$$

Alternatively, the p value

$$p = \min\left(\alpha \mid H_0 \text{ can be rejected}\right)$$

can be read off from the given (cumulative) distribution function $F_{T(9)}(y)$ of the student distribution by observing that, for $\alpha = p$, Formula (4) is valid as an equation, and the quantile function is the inverse function of the distribution function:

$$\begin{aligned} |t| &= t^{(9)}_{1-p/2} \\ F_{T(9)}(|t|) &= 1 - \frac{p}{2} \\ p &= 2\left(1 - F_{T(9)}(|t|)\right) \ll 1\,\% \;\Rightarrow\; \beta_1 \text{ highly significant} \end{aligned}$$

Same procedure for the next test:

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

(ii) Test-Function

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

(iii) Realisation:

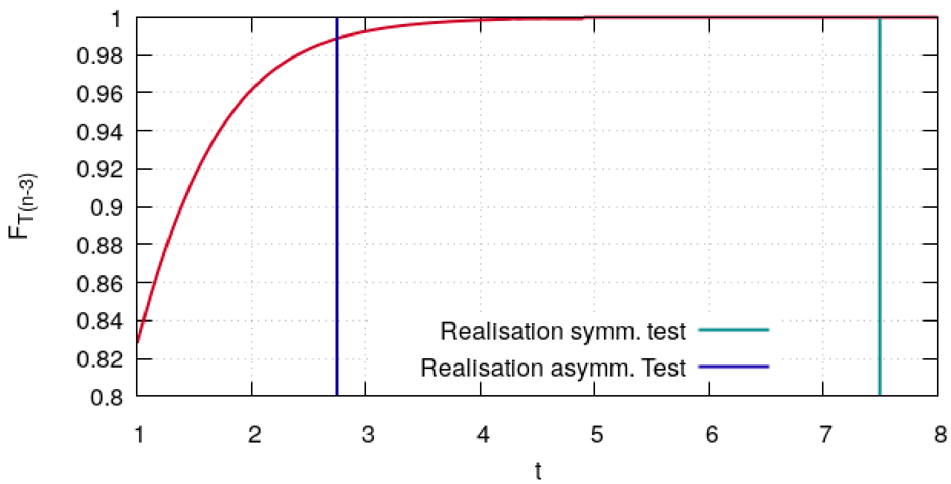
$$t = \frac{-0.953 - (-1.5)}{\sqrt{0.0397}} = 2.75$$

(iv) Decision: H_{02} can be rejected if

$$t > t_{1-\alpha}^{(9)} = 1.83 \;\Rightarrow\; H_{02} \text{ rejected.}$$

Or calculating the p value using the distribution function $F_{T(9)}$:

$$p = 1 - F_{T(9)}(t) \approx 1\,\% \;\Rightarrow\; \beta_2 \text{ significant}$$



(f) (i) Compound null hypothesis H_{03}: $\beta_1 = 30$ AND $\beta_2 = -0.5$.

(ii) Test function

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

where $F(2, n-3)$ is the Fisher-F distribution with 2 numerator and $n-3$ denominator degrees of freedom (df) and S_{restr} and S are the sum of errors (SSE) of the calibrated restrained and full model, respectively. The restraint model, given by

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

has two parameters less than the full model, hence the two numerator degrees of freedom.

(iii) Realisation from the data: $S_{\text{restr}} = \sum_i (y_i^{\text{restr}} - y_i)^2 = 878.6$, $S = 498.2$, so

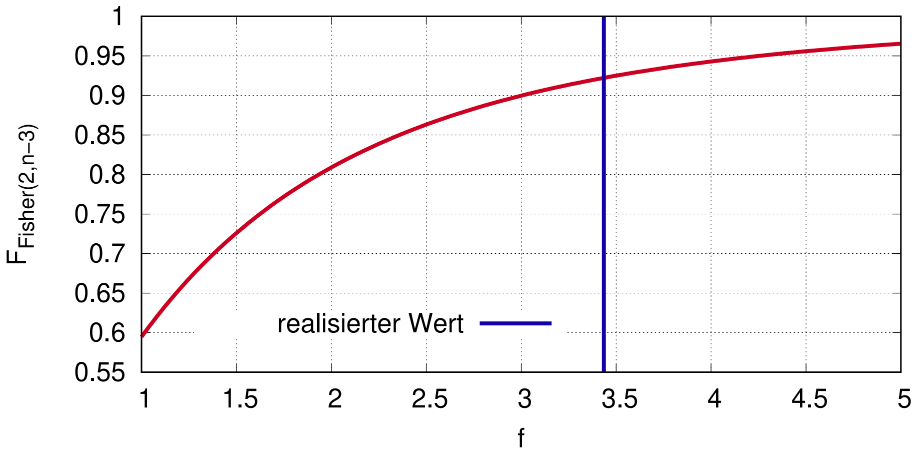
$$t_{\text{data}} = \frac{\frac{878.6-498.2}{2}}{\frac{498.2}{9}} = 3.44.$$

(iv) Decision: Rejection at α if t_{data} is greater than the $1-\alpha$ quantile:

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

This is not satisified, so no rejection is possible at $\alpha = 5\,\%$.

p value from the graph:



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

A p value above α confirms that no rejection is possible, here.

As an illustration, following figure shows the combined null hypothesis H_{03} together with the confidence *region* for the observation in β_1-β_2 parameter space, and also the isolated CIs for β_1 and β_2: