Methods in Transportation Econometrics and Statistics (Master)

Winter semester 2023/24, Tutorial No. 3

Problem 3.1: Booking Demand in Hotels

For twelve hotels, the quality (number of stars), the price (Euros per night), and the relative booking rate (%) is given according to following table:

quality [#stars]	1	1	1	2	2	2	3	3	3	4	4	4
price [€/night]	15	31	40	34	50	58	67	72	84	82	98	116
booking rate [%]	42	38	24	76	52	40	90	77	62	90	82	68

The dependence of the booking rate y on the quality x_1 and the price x_2 is to be analyzed by the linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad \epsilon \sim i.i.d. N(0, \sigma^2).$$

(a) Specify the vector $\vec{y}$ and the data matrix $\underline{\underline{X}}$ of the corrrsponding system equations.

(b) determine the OLS estimator $\hat{\vec{\beta}}$ for the parameter vector and the estimate for the booking rate.

Solution: $\hat{\beta}_0 = 25.5$, $\hat{\beta}_1 = 38.2$ und $\hat{\beta}_2 = -0.953$.

(c) Discuss the values for $\hat{\beta}_1$ und $\hat{\beta}_2$. Are their signs plausible?

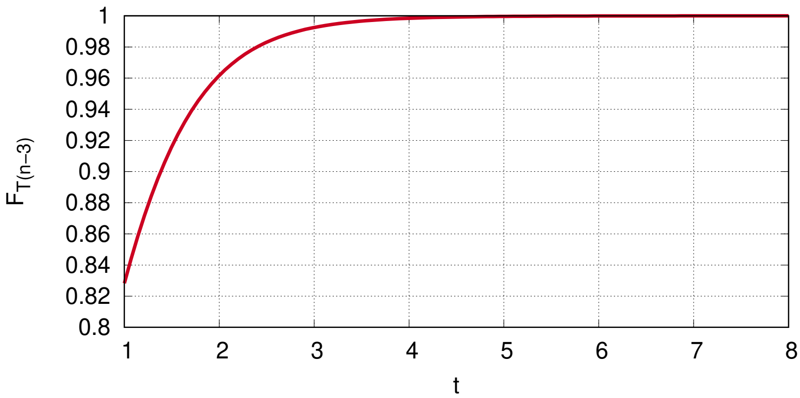
(d) Price and booking rate correlate positively with the descriptive correlation coefficient $r_{2y} = 0.58$. Naively, one could conclude that the attractiveness increases with the price which is inconsistent with the sign of $\hat{\beta}_2$ (and with the expectation). Discuss why this would be an erroneous conclusion. How could one resolve this contradiction, i.e., in which sense are more expensive hotels really more attractive?

(e) A statistics software calculated following estimate of the variance-covariance matrix of the parameter estimators:

$$\underline{\underline{\hat{V}}}(\hat{\vec{\beta}}) = \begin{pmatrix} 28.0 & -6.4 & -0.11 \\ -6.4 & 26.0 & -0.94 \\ -0.12 & -0.94 & 0.0397 \end{pmatrix}$$

Test at a significance level of $\alpha = 5\,\%$ if the appraisal for quality is significantly different from zero. Is it possible to reject the null hypothesis that the proce sensitivity is below $\beta_2 = -1.5$?

Hints: You may use the t-test using the included table for the quantile values or you may directly use the included plot for the student-t distribution below. Give also the p values for both tests.



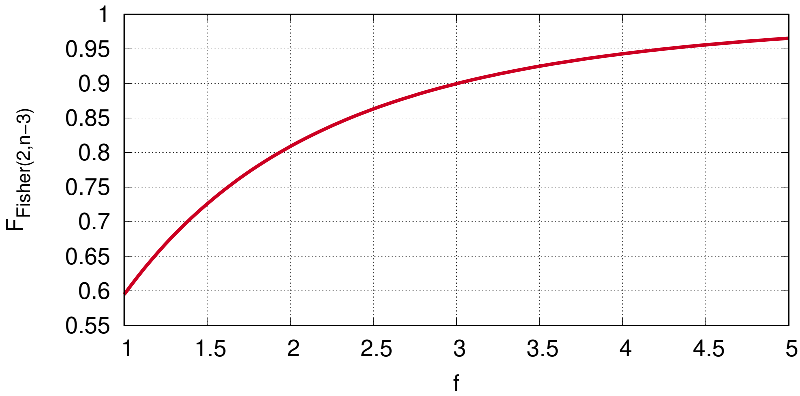
Quantiles $t_{n,p}$ of the Student t distribution with n degrees of freedom

n	$p = 0.60$	0.70	0.80	0.90	0.95	0.975	0.990	0.995	0.999	0.9995
1	0.325	0.727	1.376	3.078	6.315	12.706	31.821	63.657	318.31	636.62
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.327	31.598
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.153	3.707	5.208	5.959
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169	4.154	4.587
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576	3.090	3.291

(f) Use the F-test to test following compound null hypothesis at a significance level of $\alpha = 5\,\%$:

$$H_0 : (\beta_1 = 30) \cap (\beta_2 = -0.5).$$

Use the numerical values $S_{\text{restr}} = 878.6$ and $S_{\text{min}} = 498.2$ for the sum of squared errors (SSE) of the restrained and full models, respectively. Again, you can apply quantile tables or the plot of the distribution function below. Also give the p value


1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
F_Fisher(2,n−3)
1
1.5
2
2.5
3
3.5
4
4.5
5
f