Case Study 4

WS 2020/2021

Deadline: December 13th, 2020@ 23:55

1 Data Description

The dataset `marketing.csv` contains $N = 3195$ observations. The response variable `rating` is the consumer rating for mineral water, on a scale between 1 and 20 (where 1 is the lowest possible score, and 20 stands for an excellent rating of the water). The following product properties are available as explanatory variables:

- `price`: Price of mineral water (in cents)
- `rq`, `ju`, `vo`, `wa`, `kr`: The brand of mineral water

as well as the following demographic variables relative to the consumers:

- `education`: Education (dummy variable: 1, if high school diploma or higher)
- `gender`: Gender (dummy variable: 1, if female)
- `income`: Income (dummy variable: 1, if income is above average)
- `age`: Age (in years)

2 Model

Analyze the effect of the product properties as well as the demographic variables on the `rating`, that is, the opinion of consumers have of the mineral water, using a multiple regression model. Specifically:

2.1 Model estimation

2.1.1

Model 1: Build a multiple linear regression model including all variables, perform OLS estimation and report the output. Use `ju` as a baseline for the brand effect.

2.1.2

Model 2: Again, estimate a linear regression model including all variables. This time, do not use a brand as a baseline and include all brand dummies in the model, while excluding the intercept. Report the output of the OLS regression. Compare the OLS estimates for all explanatory variables except for the brands.

2.2

Report the OLS estimates of the regression parameter associated with `kr` for Model 1 and Model 2, respectively. Provide its interpretation within the each of the models, respectively.

2.3

How can you calculate the regression parameter associated with `kr` in Model 1 from the regression parameters of Model 2?

2.4

In Model 1, test whether there is a difference in the average rating between the brands `ju` and `wa` at a significance level of $\alpha = 0.05$, *ceteris paribus*. In particular, report the corresponding null hypthesis and the p-value.

Bonus question (no points): Answer the same question in Model 2.

2.5

Now assess whether brand information is helpful to determine the rating of mineral water. Address this problem from the following two different angles:

2.5.1

Perform an appropriate test for Model 1.

Bonus question (no points): Perform an appropriate test for Model 2.

2.5.2

Compare Model 1 and a reduced Model 3 in terms of model selection criteria (AIC, BIC, R^2, adjusted R^2).

Bonus question (no points): Would the results for a comparison of Model 2 and 3 look any different?

2.6

Check whether one may assume that the error term is normally distributed in Model 1. To this end, consider appropriate histrograms, qqplots and tests.

2.7

Now amend Model 1 by incorporating interaction terms between some dummy variables and continuous explanatory variables which you might deem possibly important.

2.8

Whenever you add an interaction effect, check wether the extended model performs better than model 1 by monitoring appropriate goodness of fit criteria as well as appropriate tests. Document the procedure step by step.

2.8.1

Call the amended model "Model 4". Report the full output of the OLS estimate.

2.8.2

Provide an interpretation of the coefficient of one interaction term of your choice in your model.

2.9

Bonus question: Try to improve model 4 by omitting terms which are not helpful. Do so with appropriate tests and goodness of fit criteria.

Case Study 4

WS 2020/2021

Deadline: December 13th, 2020@ 23:55

3 Theory

Judge if the following assertions are correct. Provide theoretical arguments and or demonstrate your assessment with an appropriate data example / simulation example.

3.1

For a fixed number N of observations y_i, $i = 1, \ldots, N$, of a response variable Y, it is always possible to find regressors $x_{1i}, \ldots, x_{Ki}$, $i = 1, \ldots, N$ such that one can fit a multiple linear model that achieves a coefficient of determination R^2 of 1.

3.2

Consider the following quadratic model

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u,$$

where $\mathbb{E}(u|X) = 0$.

Provide the null hypothesis for a statistical test that the point where the effect of a marginal increase in X on the conditional expectation $\mathbb{E}(Y|X)$ changes its sign is 1. What test can you use to this end?

3.3

Bonus question: Illustrate such a test with an appropriate simulation study.