Exercises: Multiple Linear Regression Part I

Valentina Zangirolami - valentina.zangirolami@unimib.it

December 5-14, 2023

(Thanks to Dr. Roberto Ascari for providing these exercises)
(Referring to the theoretical parts: 9, 10, 12, 13, 14, 15, 16, 17, 18, 19)

1 Exercise 1

Among 120 elementary school children, data about daily time spent in front of the TV (TV variable), gender (G variable) and time spent answering to a logic-mathematics question (T variable) were collected.

Exercise 1.1
Specify an appropriate regression model for the response variable T.

Exercise 1.2

The model specified at ex 1.1 provides the following values: $SST = 985$, $R^2 = 0.51$, $S.E.(\hat{\beta}_2) = 0.9$, $S.E.(\hat{\beta}_3) = 2.3$, where $\hat{\beta}_2$ and $\hat{\beta}_3$ are the maximum-likelihood estimators of the regression coefficients for TV variable and G variable, and $\hat{\rho}_{(\hat{\beta}_2,\hat{\beta}_3)} = 0.68$.

Perform a statistical test to check the goodness of fit of our model by employing the p-value (specify also the null hypothesis).

Exercise 1.3

Identify all the elements of the matrix $(X^T X)^{-1}$ which can be computed within the available data (specified in the above exercises).

2 Exercise 2

Among 100 households in northern Italy, the variables Y = monthly expenditures for foods (in hundreds of euros), X_1 = monthly household income (in hundreds of euros), X_2 = number of household members, and X_3 = type of diet (divided in *"vegetarian"*, *"vegan"* and *"other"*) were collected.

Exercise 2.1
Specify a multiple linear regression model for the response variable Y.

Exercise 2.2

Let $c_{j,h}$ be the elements of the matrix $(X^T X)^{-1}$, where $c_{2,2} = 0.02$, $c_{3,3} = 0.07$, $c_{2,3} = -0.02$. Let also consider that $\hat{\beta}_2 = 0.5$, $\hat{\beta}_3 = 0.8$ and $SSE = 300$.

Evaluate the significance of β_2 and try to interpret the value of $\hat{\beta}_2$.

Exercise 2.3

Find the probability distribution of $\hat{\beta}_2 - \hat{\beta}_3$ and build a statistical test based on the null hypothesis $H0 : \beta_2 = \beta_3$ (at 1% significance level).

Exercise 2.4

Knowing $SSE = 282$ for a model which includes an interaction between the dummy variable (referred to the type *"vegeterian"*) and X_1, compute and interpret the partial coefficient of determination and decide the best model through an appropriate test (specify hypothesis, test statistic and p-value).

3 Exercise 3

To assess the verbal skills of 33 children, a test was conducted by collecting: the final score, the number of books read monthly by each child, and the number of books read monthly by their parents.

Exercise 3.1

Choose an appropriate response variable together with an appropriate linear regression model. Then, specify the related assumptions and the dimension of the design matrix X.

Exercise 3.2

Complete the following table and provide an interpretation of the estimates of the significant regression coefficients.

	Estimates	**S.E.**	**t-value**	**p-value**
X_1	1.5	0.44		
X_2		0.22		0.01

Exercise 3.3

Knowing the SST is equal to 2980 and $R^2 = 0.59$, decide if one of the two below options are compatible with the previous data (considering that the below options are based on a regression model with just one independent variable X_1):

- $SSR = 1800$ and $SSE = 1180$
- $SSR = 1500$ and $SSE = 1500$

Justify your answer.

Exercise 3.4

Knowing the SSR is equal to 1440 (for a regression model which just includes X_1), try to evaluate if it is better to include the second independent variable through an appropriate statistical test (specify hypothesis, test statistic and p-value).

4 Exercise 4

Considering 84 business company in northern Italy, we estimated the following regression model

$$\hat{y} = 12.7 + 9.3x_1 + 1.9x_2 - 1.6x_3$$

where Y = monthly turnover (in thousands), X_1 = sector (1 = manufacturing, 0 = trade), X_2 = number of employees, and X_3 = decrease in investment advertising compared to the previous year (in hundreds of euros). Further, $SSE = 2308$ and $R^2 = 0.62$.

Exercise 4.1

Interpret the estimate of β_2 ($\hat{\beta}_2 = 9.3$).

Exercise 4.2

Complete the table below and show the formula we should use.

	Estimates	**S.E.**	**t-value**	**p-value**
X_3				0.02

Exercise 4.3

Evaluate the goodness of fit through a valid test (thus, using the p-value).

Exercise 4.4

We would like to assess the added explanatory contribution of the variable ”number of employees” compared with the model that does not contain it. Knowing that the regression sum of squares of the model without this variable is equal to 1978, compute an appropriate index/coefficient and interpret the result.