

SONEK DATA SCHOOL

ATA ANALYSIS

Assessment Item 2 Research Report (2019 S1)

A researcher is interested in the factors that may explain the academic achievement of first year Business students. With the university's support, the researcher developed a standardized test that was to be undertaken by a random sample of students. The test contained questions related to the core business units of accountancy, marketing, management and statistics, and was administered towards the end of each student first year of study. In addition to answering the test, the students were asked questions on their gender, age, mother's highest level of education, whether the student was currently in a 'romantic' relationship. Students were also asked to report the number of instances where they did not attend either a lecture or a tutorial. In total, 649 students sat the test and a portion of the dataset^{1,2} is presented below:

Result	Gender	Age	Medu	Relationship	Lectures	Tutorials
55	F	20	PG	No	4	3
55	F	19	HS	No	2	3
65	M	18	PG	No	6	4
65	M	18	HS	No	0	0
65	F	18	U	No	0	1

Legend:

Result = Score 0 to 100

Gender = Male (M) or Female (F)

Age = Age in years when sitting test

Medu = Mother's highest level of educational obtainment, High School (HS), Undergraduate (U) or Post Graduate (PG).

Relationship = No (not in romantic relationship) and Yes (in romantic relationship).

Lectures = Number of lectures not attended year-to-date: self-reported

Tutorials = Number of tutorials not attended year-to-date; self-reported.

- (a) Test if there is any difference in Results on average between male and female students at the 5% level of significance.

(2 marks)

¹ The dataset is based on the UCI Machine Learning Repository. That data has been manipulated for the purposes of this case study. Source <http://archive.ics.uci.edu/ml/datasets/Student+Performance>.

² Original citation P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008)

SONEK DATA SCHOOL

One would expect that students that are involved in a romantic relationship may not perform as well as their peers. Test at the level of significance of 5% if students who are involved in a romantic relationship have a lower result than the students who are not.

(3 marks)

Note: In conducting the tests in (b) and (c), you should discuss briefly whether it is a one or two tail test, the test statistics, any assumption made and draw a conclusion based on Excel output.

Task 2 (Regression Analysis)

You plan to develop a regression model to further investigate how various factors influence students' Result.

- (a) Before you conduct any regression analysis, construct a correlation matrix of all the quantitative variables in the dataset. Based on the correlation matrix, comment briefly on the linear associations between Result and other quantitative variables (viz. Age, Lectures and Tutorials) and whether these variables are a predictor of result.

(2 marks)

- (b) Conduct a *simple regression* on:

- (i) Age is a predictor of Result?
- (ii) Lectures is a predictor of Result?
- (iii) Tutorials is a predictor of Result?

Does this support your answers to Task 2 (a)?

(3 mark)

- (c) For each of the independent variables contained in the regression model in Step 1, test their statistical significance. *In testing statistical significance of a regression coefficient, .*

(6 marks)

Task 3 (Summary Report)

Given your results you are asked to write a summary report (400 words) detailing *all* the findings from your data analysis. The issues you can discuss may include (but are not limited to):

- Do you think there is any gender difference in academic achievement of the students?
- Is there evidence that being in a romantic relationship affects results?
- In the Step 1 regression interpret how each variable is related to Results and report its significance.
- Is there a variable whose result is different to what you would general have expected?
- Based on this study, what is the final model you would recommend?
- Based on your final model, predict the Result of a female student who is 18, whose mother has post-graduate qualifications, is not involved in a romantic relationship and attended all classes. □ Are there any variables that influence academic performance that are omitted in this study?

(8 marks)

SONEK DATA SCHOOL

Notes:

- Use **1 & ½ spacing** and font size of **11**.
- You can and are encouraged to include relevant **charts** and **Excel objects** in your summary report (Task 3).
- **No referencing** is required in your summary report. However, if you wish to include, and refer to, additional information, you can use any referencing system as long as it is used consistently.
- There is no word limit for Tasks 1 and 2.
- The **word limit** of 400 (with a tolerance of 10%) applies only to the summary report (Task 3), and is exclusive of words in tables, appendices and reference list (*if any*).

Submission:

- You should submit your response to all three tasks as a single **pdf document** saved in the format:
BSB123 Report_StudentName.pdf
- After uploading your research report on Blackboard, it is your responsibility to go back to the Assignment Upload page to check that your report was properly uploaded.