CAB220 Portfolio 2 Task Brief

Overview

This portfolio accounts for 20% of your overall grade of CAB220. Full mark of this portfolio is 20. The tasks in this portfolio are designed to assess your knowledge and skills in

- Descriptive statistical data analysis and visualisation
- Statistical hypothesis testing
- Linear regression
- Logistic regression

Data

The fictitious data set for this portfolio includes the records of 2,550 first-year students of an Australian university in terms of case *ID, Attrition, Degree Type, Achieved Credit Points, Attendance Type, Age, Failed Credit Points, International student, First in family in university, Gender, GPA, OP Score, Socio Economic Status, Teaching Period Admitted,and Faculty*.

Tasks

1. Summarise the information in each variable (except case ID) using a table or an appropriate statistical graph (7 marks)

2. Compare average GPA between male and female students using a graph, conduct a statistical test, and interpret its results (2 marks)

3. Explore the relationship between OP Score and GPA using a graph, describe the relationship (2 marks)

4. Develop a linear regression model of GPA using the given data. You need to describe your choice of predictors, examine your model's assumptions, assess model fit, and interpret the final model's regression coefficients. (5 marks)

5. Develop a logistic regression model to predict Attrition. You need to describe your choice of predictors, assess model fit, and interpret the final model's regression coefficients. (4 marks)

All data analyses in this portfolio have to be performed using R packages.

Format

Please prepare your work of the tasks in a **PDF** document and submit it on CAB220 Blackboard site by **11:59pm Sunday 22nd September 2019**. When presenting your work for each task, please also **include the R codes you used**.