

# TSTA602 Assignment 2

March 27, 2021

**Instructions.** Assignment 2 has a total of 40 marks and worths 40% of the final grade of TSTA602. The detailed mark allocations are provided at the beginning of each question. Unless otherwise specified, please calculate your answer to two decimal places if approximation is needed. This assignment is **due at 5pm on Friday, 28th May 2021 (Week 9)**. Please submit your assignment before the deadline, no late submission will be accepted unless a pre-approval from Lecturer is obtained. The assignment must be submitted via Moodle in a pdf format with the file name 'firstname\_surname\_studentID.pdf'. Any assignment submitted through email (without pre-approval) will not be marked. Although this is an individual assignment, you are encouraged to discuss the assignment with your peers, but must write down your own solution. Penalties will apply if two solutions have a high similarity.

**Scenario.** You, as a property investor, are interested in understanding which factor (or factors) drives the prices of investment properties. A dataset is collected which contains the prices (in thousand dollars, as denoted by `apart_price`) for 50 one-bedroom apartments in city X, their corresponding rents per week (in dollars, as denoted by `rent`) and the costs to hold each of these properties per week (in dollars, as denoted by `cost_of_property`). Following the procedures below to analyse the dataset 'assign2\_data.csv' by using Rstudio. Please only include relevant outputs from Rstudio in your solution and attach the R codes as appendice (**2 marks** for attach R codes).

- (a). (**3 marks**) Import the data into Rstudio, draw two scatter plots: `apart_price` versus `rent` and `apart_price` versus `cost`.
- (b). (**4 marks**) Fit the following two linear models:

$$\text{Model 1: } \text{apart\_price} = b_0 + b_1 \times \text{rent}$$

$$\text{Model 2: } \text{apart\_price} = c_0 + c_1 \times \text{cost}$$

Write down the equations of the two models with correct coefficients.

- (c). (**8 marks**) Written down the p-values from the output of your R codes. Comment on the significance of all coefficients obtained from (b) based on the p-values (from the outputs of Rstudio). The significance level is 0.05.
- (d). (**6 marks**) Produce residual plots for each model in (b), comment on each plot.
- (e). (**4 marks**) Produce normal qq plots for each model in (b), and comment on each plot.

(f). (3 marks) Fit the following linear model:

$$\text{Model 3: } \text{apart\_price} = d_0 + d_1\text{rent} + d_2\text{cost}$$

Write down the equation of the model with correct coefficients.

- (g). (6 marks) Write down the p-values from the output of your R codes. Comment on the significance of all coefficients obtained from (f) based on the p-values (from the outputs of Rstudio). The significance level is 0.05.
- (h). (2 marks) Compare Model 1 and Model 3, explain which one is better.
- (i). (2 marks) Given  $\text{rent} = 900$  and  $\text{cost} = 650$ , predict prices under Model 1 and Model 3.