Tutorial 3

1. Read the data from the file `Colleges.txt`. Consider a simple linear regression of percentage of applicants accepted (`Acceptance`) on the median combined math and verbal SAT score of students (`SAT`), called Model M1.

   (a) Write your own function in R, name the function as **simple**, to derive the intercept β_0 and the slope β_1 of Model M1.

   Hint: Use the formula of the estimated coefficients, $\hat{\beta}_1$ and $\hat{\beta}_0$, given in slide 31/52 of Topic 3.

   (b) Use function `lm()` in R to derive the coefficients of Model M1. Compare with your answer in part (a).

2. Consider a dataset about HDB resale flats in Singapore give in `hdbresale_reg.csv`. Consider a simple model (Model M2) where the resale price is the response and the floor area in square meters is the only regressor.

   (a) Use function **simple** you formed in the question above to find the coefficients of Model M2.

   (b) Use function `lm()` in R to derive the coefficients of Model M2.

3. Consider data set given in the file `hdbresale_reg.csv` on Canvas, which has the information of 6055 HDB resale flats in Singapore. We would want to form a linear model that helps to predict the resale price of HDB flats, **based on the floor area in square meters and the type of the flats**.

   (a) Consider the resale price, plot a histogram of it and give your comments. Is it suitable to fit a linear model for this response variable? Explain.

   (b) Consider the resale price, plot a histogram of log_e of it and give your comments. Is it more suitable to fit a linear model for this response variable than the original resale price?

   (c) Derive a scatter plot of the log_e of the resale price against the floor area in square meters. Give your comments.

   (d) Fit a linear model where the log of the resale price be the response. Write down the fitted equation.

   (e) Report the coefficient of the floor area in square meters and interpret it.

   (f) Predict the resale price of a 4-room HDB flat that is of 100 square meters.

   (g) Report R^2 of the model and interpret it.