Problem Set - 3 (Simple Linear Regression)

ECO 204 (Section 6)
Instructor: Shaikh Tanvir Hossain

Due: 3rd Dec (before 10.00 PM), submit in Google Classroom

§

Please try to solve them all using R and avoid excel or calculator, so that you have a good practice. Submit a PDF file. You can type in Word or Rmarkdown.

1. Please solve following problems from the Chapter 14 of Anderson et al. (2020)
   (a) 3, 9, 13, 20, 25, 31, 36, 38, 39

2. Use `Auto.xlsx` data set to perform a simple linear regression with `mpg` as the response and `horsepower` as the predictor. Use the `summary ()` function to print the results, also answer following questions,
   (a) What is the intercept and the slope coefficient? Interpret them.
   (b) Write the equation of the estimated regression line.
   (c) What does the coefficient R^2 tell you? How is the fit?
   (d) Is there a statistically significant relationship between horsepower and mpg? Justify your answer. Does this make sense?
   (e) Construct a residual plot against the independent variable. Do the assumptions of the simple linear regression model, in particular linear model and homoskedasticity seem reasonable in light of the residual plot?
   (f) What is the predicted mpg associated with a horsepower of 98? What are the associated 95% confidence and prediction intervals?

3. In this problem you will create some simulated data and will fit simple linear regression models to it. Make sure to use `set.seed(last 3 digits of your id)` prior to starting part (a) to ensure consistent results.
   (a) Using the `rnorm()` function, create a vector, `x`, containing 100 observations drawn from a $\mathcal{N}(0, 1)$ distribution. This represents a feature, X.
   (b) Using the `rnorm()` function, create a vector, `eps`, containing 100 observations drawn from a $\mathcal{N}(0, 0.25)$ distribution - a normal distribution with mean zero and variance 0.25 .
   (c) Using `x` and eps, generate a vector `y` according to the model

$$Y = -2 + 0.7X + \epsilon \tag{1}$$

   (d) Find the length of the vector `y` ? What are the values of β_0 and β_1 in this linear model?
   (e) Create a scatterplot displaying the relationship between `x` and `y`. Comment on what you observe.
   (f) Fit a least squares linear model to predict `y` using `x`. Comment on the model obtained. How do $\hat{\beta}_0$ and $\hat{\beta}_1$ compare to β_0 and β_1 ?
   (g) Find 95% confidence interval for β_0 and β_1.
   (h) Display the least squares line on the scatterplot obtained in (d). Draw the population regression line on the plot, in a different color. Use the `legend ()` command to create an appropriate legend.
   (i) Repeat (a) - (g) except (d), after modifying the data generation process in such a way that there is less noise in the data. The model (1) should remain the same. You can do this by decreasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.
   (j) Repeat (a) - (g) except (d), after modifying the data generation process in such a way that there is more noise in the data. I The model (1) should remain the same. You can do this by increasing the variance of the normal distribution used to generate the error term ϵ in (b). Describe your results.
   (k) What are the confidence intervals for β_0 and β_1 based on the original data set, the noisier data set, and the less noisy data set? Comment on your results.

References:

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J. and Ohlmann, J. W. (2020), *Statistics for Business & Economics*, 14th edn, Cengage, Boston, MA.