Statistical data analysis, Assignment 7

Problem 1. Let $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ be a sample of n observations on (X, Y)

(a) Define the following measures of association:

(i) population correlation coefficient ρ between X and Y.

(ii) sample correlation coefficient γ between X and Y among n observations.

(iii) coefficient of determination R^2 between X and Y among n observations.

(b) Suppose that X is used to predict Y using a simple linear regression:

(i) Write the statistical model that is the foundation for this analysis.

(ii) Describe the method of least squares for estimating the regression coefficient $\beta = (\beta_0, \beta_1)$. Also find the least square estimator $\hat{\beta}$ of β.

(iii) Write the ANOVA table for this regression, including the columns of Source, SS, DF, MS, and F.

(iv) What is the null hypothesis that is to be tested using the F statistic?

Problem 2. Textbook OpenIntro Statistics, 2019

- **Chapter 8:** Exercises 2,5,7,8,21,23, 25,35, 44
- **Chapter 9:** Exercises 3,7,8,9, 10l, 13, 19,

Problem 3. (R practice) The advertising dataset ”adv.csv” captures the sales(Y) revenue generated with respect to advertisement costs across multiple channels like radio, tv, and newspapers. It is required to understand the impact of ad budgets on the overall sales.

(a) Regress ”Sales” on ”TV.Ad.Budget + Radio.Ad.Budget + Newspaper.Ad.Budget”.

(b) From (a), do model selection via backward elimination process.

(c) By the model selected from (b), showing the residual plot and Q-Q plot. Are there any visual influential observations? Please remove the visual influential observations and then refit the appropriate model.

(d) Based on your model from (c), report (i) the estimated coefficients ; (ii) 95% confidence interval for these coefficients ; (iii) coefficient of determination : R^2 and mean square error: MSE.