MA331 Homework 01

Problem 1. Complete the following problems for the sample of two variables (X, Y) by using R.

$(0.2, 1.1), (1.2, 2.3), (0.9, 1.1), (2.2, 3.6), (3.2, 0.1), (0.3, 1.0), (1.7, 6.9)$

$(3.1, 4.8), (2.3, 6.5), (1.5, 7.8), (2.5, 5.8), (3.0, 8.0), (2.6, 9.4), (9.0, 9.8).$

(i) Plot the histogram, pie chart of X and Y, respectively, and then describe their distributions. (Hint: Since histogram and pie chart are for categorical variables, numerical observation ought to be grouped before applying the two functions.)

(ii) For X and Y, build the box-plot, compute their five-number summaries and variances, respectively. Are there any outliers of X and Y, respectively?

(iii) Produce the scatter plot of (X, Y) and evaluate their correlation coefficient. Qualitatively describe the linear association between X and Y.

(iv) Are there any outliers of (X, Y)? If yes, remove them and compute the correlation coefficient again.

(v) What difference do you observed between the numerical results in (iii) and (iv)?

(vi) Produce the normal QQ plots for observations of X and Y, respectively, and then determine which one is more likely to be of normal distribution?

Problem 2. For $x_1, \cdots, x_n$, show that

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

and thus

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2.$$