

## MA329 Statistical linear models 23-24

**Assignment 1** (Due date: Oct 9, 11pm. For late submission, each day costs 10 percent)

1. (50 marks, don't use R outcome directly) The following table gives the data on daytime eruption of OLD Faithful Geyser in Yellowstone National Park on August 1, 1978. The variables are  $x$  =duration of an eruption and  $y$  =interval to the next eruption (both in minutes).

\$x\$	4.4	3.9	4.0	4.0	3.5	4.1
\$y\$	78	74	68	76	73	84

- (a) Construct and comment a scatterplot of the data.
  - (b) Find the least squares line from the data and plot it on your scatterplot.
  - (c) What is your linear regression model? State the necessary assumptions.
  - (d) Test the hypothesis that the duration of an eruption has no effect of the interval to the next eruption when a linear model is used (use  $\alpha = 0.05$ ). State the null and alternative hypotheses. Draw the appropriate test conclusions.
  - (e) Find a 95% confidence interval for  $\beta_1$  (the slope of the linear regression model). Interpret your results.
  - (f) Find the coefficient of determination for the linear regression model. Interpret your result.
  - (g) Find a prediction of the time to the next eruption when the Geyser eruption lasts for 4 minutes and its 95% interval.
2. (20 marks)
    - (a) Define a simple linear regression model and derive MLE (maximum likelihood estimation) for all the unknown parameters
    - (b) Comments on the difference between MLE and LSE (least square estimation)