STU22005 Applied Probability II
Professor Caroline Brophy
Continuous Assessment Sheet 3

Due date: submit before 4pm Friday 26th March 2021.

For this assignment, you must do question 1 (you should also do the remaining questions, but do not submit them). To submit the assignment, you must submit two files:

- File 1: Fill in the answer sheet that accompanies this assignment. You must fill in the answers on the pdf document, save the file, and upload to Blackboard.
- File 2: Take images (e.g., photographs on your phone) of your workings to show how you did the questions. Combine into a pdf document and upload to Blackboard.

NB: you must submit the two separate files, the answer sheet alone will not be accepted.

1. The reduction in blood pressure (Y) caused by a blood pressure drug was measured at each of a number of doses (x).

Reduction	4.2	6.8	5.2	8.4	5.9	10.4	9.1	12.4	9.4	12.8
Dose	10	20	30	40	50	60	70	80	90	100

Reduction	12.9	16.2	11.7	12.9	14.3
Dose	110	120	130	140	150

Answer the following questions about the dataset, but do not use R software to help, work out the calculations by hand.

(a) Sketch (by hand) a scatter plot of the data. Is there an indication that the mean value of Y depends on x?

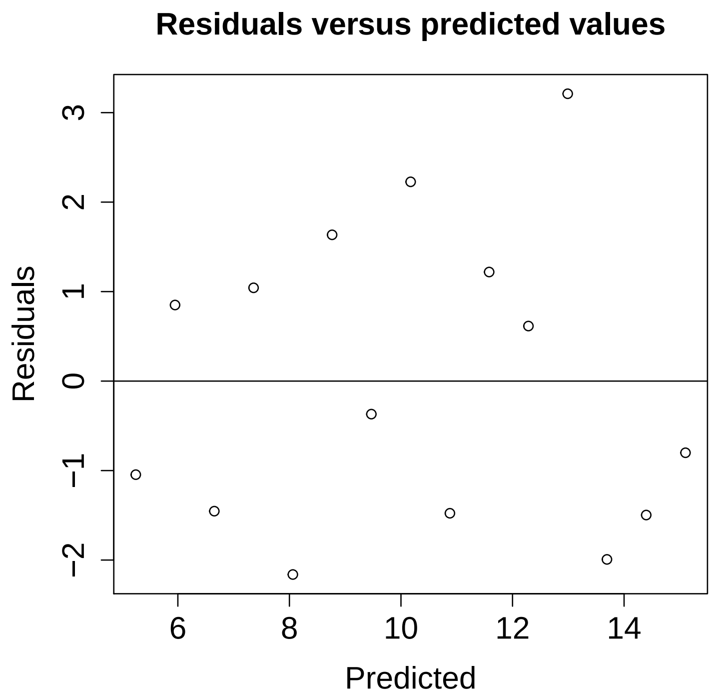
(b) Fit a simple linear regression model to these data. You may use without proof the formulas from least squares estimation.

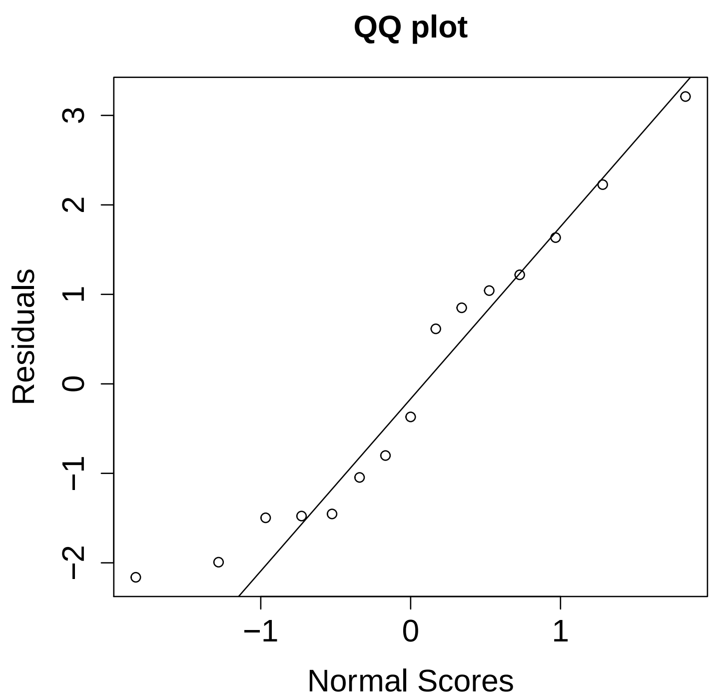
(c) Interpret the estimated parameters of the model in the context of the problem at hand.

(d) Compute an estimate of σ^2, the variance of the error term. Give a one-sentence practical interpretation of what this estimate represents.

(e) Based on the residual plots (see next page), are the assumptions for this model reasonable? In your workings, begin by stating the assumptions, and then refer to the plots in your assessment of how reasonable they are here.

Residual plots for question 1(e):





Below are additional questions that do not need to be handed up.

2. Repeat question 1, this time using R software. In part (e), generate the residual plots yourself. Compare your hand calculations in question 1 to your output.

   Hint: Lab session 4 will be useful in answering this question.

3. Suppose pairs of datapoints (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) are collected and a simple linear regression analysis will be carried out. Using algebraic manipulations, verify the following.

   (a) $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

   (b) $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i = \sum_{i=1}^{n} x_i(y_i - \bar{y})$.

   (c) $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

4. Consider the model

   $$y_i = \beta_0 + \epsilon_i, \qquad i = 1, ..., n$$

   where ϵ_i are assumed i.i.d. N(0, σ^2).

   (a) Show that the ordinary least squares estimate of β_0 is $\bar{y}$.

   (b) Find the E$[\hat{\beta}_0]$.

   (c) Find the Var$(\hat{\beta}_0)$.

   (d) Explain why $\hat{\beta}_0$ has a normal distribution.

   (e) Find an expression to estimate σ^2 in terms of the y_i values.

   Hint: For part (d), start with $\hat{\sigma}^2$ being equal to $\sum_{i=1}^{n} \hat{\epsilon}_i^{\,2}$ divided by the degrees of freedom.