STU22005 Applied Probability II
Professor Caroline Brophy
Continuous Assessment Sheet 3

Due date: submit before 4pm Wednesday 16th March 2022.

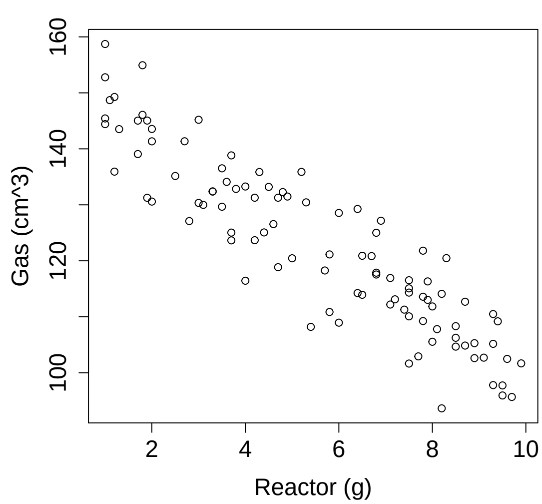
For this assignment, you must do and submit question 1. All other questions should also be done for practice and preparation for the live quiz during the week following submission, but are not handed up.

- Question 1 is to be done on your own and not in consultation with anyone else, do not discuss or show your answers to anyone else and do not look at answers belonging to anyone else. You are welcome to discuss your answers to the additional questions with other students.

To submit the assignment, you will submit one single pdf file:

- **File to be submitted**: Fill in the answer sheet that accompanies this assignment. Fill in the answers on the pdf document, save the file, and upload to Blackboard.

1. An experiment was carried out to determine the relationship between a response variable, y = the amount of a gas released, and a predictor variable, x which was the amount of reactor added prior to the gas measurement. The gas was measured in cm^3 (centimetres cubed) and the reactor was measured in g (grammes). The paired x, y data was recorded ($n = 100$) and is plotted below and summary statistics are provided.



The summary statistics are:

$$\sum_{i=1}^{n} x_i = 555.4, \sum_{i=1}^{n} x_i^2 = 3795.7,$$

$$\bar{x} = 5.554, \sum_{i=1}^{n} y_i = 12264,$$

$$\sum_{i=1}^{n} y_i^2 = 1526952.756, \bar{y} = 122.64,$$

$$\sum_{i=1}^{n} x_i y_i = 64445.499$$

(a) Fit a simple linear regression model to these data and provide the equation of the estimated line in your answer. You may use the least squares formulae without proof.

(b) Give the interpretation of the estimated slope.

(c) The estimate of σ^2 is equal to 40.5. Give a one-sentence practical interpretation of what this estimate represents.

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

2. 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?

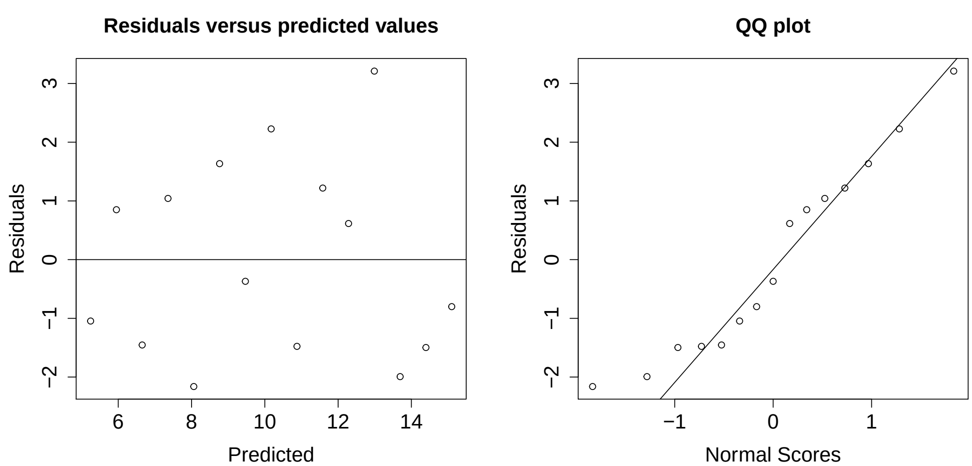
(b) Fit a simple linear regression model to these data. You may use without proof the formulas from least squares estimation.

(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):



3. Repeat the previous question, this time using R software. Carefully go through each part of the output and make sure you understand what each bit represents and how to interpret it. In part (e), generate the residual plots yourself. Compare your hand calculations in the previous question to your output in R.

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

4. Suppose pairs of datapoints (x_1, y_1), (x_2, y_2), $\ldots$, (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) $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n}$

5. 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 (e), start with $\hat{\sigma}^2$ being equal to $\sum_{i=1}^{n} \hat{\epsilon_i}^2$ divided by the degrees of freedom.