Homework: Line of Best Fit

From the notes on the line of best fit $y = \beta_0 + \beta_1 x$, we can determine β_0 if we know β_1:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

and we can determine β_1 by the fraction below (the sums runs from 1 to n if we have n points):

$$\beta_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i}{\sum x_i^2 - \bar{x} \sum x_i}$$

or, more simply:

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

1. The notes show that

   $$\sum (x_i - \bar{x})^2 = \sum x_i^2 - \bar{x} \sum x_i$$

   Use the same reasoning to show that the numerators are equal, too:

   $$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \bar{y} \sum x_i$$

2. Given the dataset below, use the least squares method to determine the equation of the line of best fit by hand with a calculator (probably best to think about the design matrix, then the system of equations for m, b).

   (1,2), (2,3), (3,5), (4,6)

   Show your computations.

3. From the previous problem, calculate the residuals and the sum of squared errors (SSE) by hand with a calculator.

4. Explain the meaning of $R^2 = 0.85$.

5. (i) What is the design matrix if we have the data:

   $$(1, 2), (2, 5), (3, 10), (4, 17), (5, 21)$$

   and we want to fit the model $y = \beta_0 + \beta_1 x + \beta_3 x^3$.

   (ii) Is this problem still linear? (Give a short reason)