MM 225 : AI and Data Science

Problems for Practice 5

1. Consider simple linear regression: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ for\ i = 1, 2, \dots, n.$
   a. Derive the least squares estimates for β_0 and β_1.
   b. What assumptions are made on $\epsilon_i, i = 1,2, \dots, n$?
   c. Additionally, assume that $\epsilon_i \sim N(0, \sigma^2)$, and derive MLE for β_0 and β_1.
2. For the simple linear regression model in question 1 above, show that

$$SST = SSR + SSE$$

3. Refer to the slides for Day 23 slide no.22. Calculate the correlation coefficient for Anscombe's four data sets. You may use calculator, Python or Excel.
   a. What does Anscombe's quartet signify?
4. In the process of setting up a regression model summary quantity were calculated from the data as follows:

   n = 18,

$$\sum x_i = 1950, \sum x_i^2 = 251970, \sum y_i = 47.92,$$
$$\sum y_i^2 = 130.6074, and \quad \sum x_i y_i = 5530.92$$

   a. Determine S_{xx}, S_{xy} and S_{yy}

   b. Determine the coefficient determination

   c. Determine the regression model

5. An investigation of a die-casting process resulted in the data on x_1 = furnace temperature and x_2 = die close time, and y = temperature difference on the die surface ("A Multiple-Objective Decision-Making Approach for Assessing Simultaneous Improvement in Die Life and Casting Quality in a Die Casting Process," Quality Engineering 1994: 371–383). The regression model fitting resulted in the following tables:

ANOVA

	df	*SS*	*MS*	*F*
Regression	2	715.5	357.75	?
Residual	6	6.722222	1.12037	
Total	8	722.2222		

Table of Coefficients

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-199.556	11.64056	-17.1431	2.52E-06
x1	0.21	0.008642	24.29876	3.19E-07
x2	3	0.432121	6.942503	0.000443

   a. State the estimated regression model.

   b. What is the F statistic for this estimated regression model?

6. What is the value of coefficient of determination?