



6. Keith collected a data set of size  $n = 5$ , namely,  $\{x_i, y_i\}_{i \in \{1,2,3,4,5\}}$ , where  $y_i$  denotes the  $i$ -th student's GPA while  $x_i$  denote his/her I.Q. score. Keith applied a simple linear regression model  $Y_i = \beta_0 + \beta_1 x_i + e_i$ . Suppose that he obtains  $\{\hat{e}_1, \dots, \hat{e}_4\} = \{0.1, -0.2, 0.3, -0.1\}$  and  $SY Y = 3.3$ . What is the value of  $R^2$ . (6 marks)
7. Which of the following is a stronger correlation than -0.42? (Circle the answer) (3 marks)  
a) 0, b) -0.1, c) -0.6, d) 2.
8. In simple linear regression, if the value of the predictor  $X$  is replaced by  $cX$ , where  $c$  is some non-zero constant, which of the following will be affected? (Circle the answer(s))  
(a)  $\widehat{\beta}_0$ , (b)  $\widehat{\beta}_1$ , (c)  $\widehat{\sigma}^2$ , (d)  $R^2$ , (e) t-test statistic of the null hypothesis  $H_0 : \beta_1 = 0$ . (6 marks)
9. If  $X$  is an  $n$  by  $r$  matrix and  $I_n$  is an  $n$  by  $n$  identity matrix, Find  $tr(I_n - X(X'X)^{-1}X')$ . (3 marks)
10. Let  $f(\beta) = Y^T X \beta - \beta^T X^T X \beta$  where  $\beta$  and  $Y$  are column vectors and  $X$  is an  $n$  by  $p$  matrix, find the minimum value of  $f(\beta)$ . (3 marks)
11. What is the difference between confidence interval and confidence band for  $E(Y|X)$ ? (3 marks)

12. Given  $Y = (-2, -1, 6, 9)$ ,  $X_1 = (1, 2, 3, 6)$  and  $X_2 = (-2, 0, 0, 2)$ .  
Fill in the ANOVA Table: (6 marks)

ANOVA Table				
Source	Sum of Squares	d.f.	Mean Square	F-statistics
Regression	_____	_____	_____	_____
Residuals	12	_____	_____	_____
Total	_____	_____	_____	_____

13. (5 marks). For the regression  $Y = X\beta + e$ ,  $e \sim N(0, \sigma^2)$ , let  $\hat{y}_i = X_i\hat{\beta}$  be the fitted value of the  $i$ -th observation,  $i = 1, \dots, n$ . Let  $X$  be a  $n \times p$  matrix. Find  $E(\sum_{i=1}^n \hat{y}_i^2)$  in terms of  $X$ ,  $\beta$ ,  $p$  and  $\sigma^2$ .

14. (30 marks) A certain metal discolours when exposed to air. To protect the metal against discoloration, it is coated with a chemical. In an experiment, coatings of varying thickness,  $x$  mm, of the chemical were applied to standard samples of the metal, and the times,  $t$  hours, for the metal to discolour were noted. The results are as shown.

x	1.8	3.0	4.0	5.7	7.2	8.4	10.3
t	3.4	5.9	7.0	8.7	9.5	10.4	11.1

- i) (4 marks) A researcher suggests that the theoretical relationship between  $t$  and  $x$  should be of the form

$$\exp(t) = Ax^B,$$

where  $A$  and  $B$  are constants. Show that this relationship may be expressed in the form

$$t = a + b \log x,$$

where  $a$  and  $b$  are functions of  $A$  and  $B$  respectively, which you should identify.

ii) (10 marks) The researcher found that

$$\sum \log x = 11.2476, \quad \sum t = 56, \quad \sum (\log x)^2 = 20.3687, \quad \sum t \log x = 100.101,$$

Use these results to calculate the least squares regression line of  $t$  on  $\log x$ . Plot this line and the data on a scatter diagram with values of  $\log x$  on the horizontal axis. Is this fitting appropriate?

iii) (8 marks) For the model in (ii), construct a 95% confidence band by joining the values of the band on  $x = 2, 4, 6, 8$ . Draw the band on the graph in (ii)

iv) (8 marks) Sketch the 95% confidence ellipse for the parameter  $(a, b)$ . (You may use 6 points to stretch the ellipse.)

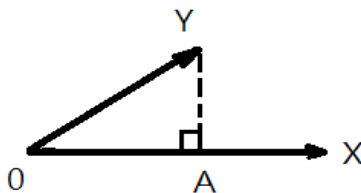
15. (10 marks). Given three variables  $Y = (y_1, \dots, y_n)$ ,  $X1 = (x_1, \dots, x_n)$ ,  $X2 = (1, 1, 1, \dots, 1)$ .

i) (5 marks) Find the slope of the added variable plot between  $Y$  and  $X1$ , with the effect of  $X2$  removed.

- ii) (5 marks) Find the slope of the added variable plot between  $Y$  and  $X_2$ , with the effect of  $X_1$  removed.

16. (5 marks) Let  $Y$  and  $X$  be two vectors as shown below.

Express the length of the segment  $OA$  in terms of  $X$  and  $Y$  in vector's notation. If you want to



represent the vector  $Y$  by  $cX$  for some constant  $c$ , which  $c$  should you use?