

STAT 350: 97-1

Midterm, 19 February 1997

Instructor: Richard Lockhart

Instructions: This is an open book test. You may use notes, text, other books and a calculator. Your presentations of statistical analysis will be marked for clarity of explanation. I expect you to explain what assumptions you are making and to comment if those assumptions seem unreasonable. The exam is out of 25.

1. Suppose two objects with weights α_1 and α_2 are weighed separately and then together. The resulting data points Y_1 , Y_2 and Y_3 satisfy $Y_1 = \alpha_1 + \epsilon_1$, $Y_2 = \alpha_2 + \epsilon_2$ and $Y_3 = \alpha_1 + \alpha_2 + \epsilon_3$.

(a) What is the design matrix of this linear model? [2 marks]

(b) If

$$(X^T X)^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

what is the hat matrix? [2 marks]

(c) Write $\hat{\alpha}_1$ in the form $a_1 Y_1 + a_2 Y_2 + a_3 Y_3$ giving specific numerical values for the a_i . [2 marks]

(d) What is the standard error of $\hat{\alpha}_1$? [2 marks]

(e) What is the variance of the residual corresponding to Y_1 ? [2 marks]

2. A company measures its annual sales Y in each of 26 regions, along with the values of 4 covariates, X_1 , the advertising expenditure in the region, X_2 , the number of active accounts in the region, X_3 , the number of competing brands, and X_4 , a measure of the potential for sales in the region. I attach some SAS code and an edited version of the output.

(a) Is the regression significant? [3 marks]

(b) Can advertising expenditure and sales potential be dropped from the full model? [3 marks]

(c) In a model which includes all 4 covariates test the hypothesis that the advertising expenditure is an unimportant predictor. [3 marks]

(d) What final fitted model seems best? (You will not be able to examine plots or diagnostics). [3 marks]

(e) Give a 95% confidence interval for the coefficient of X_3 . [3 marks]