MATH 484/564 HOMEWORK #3

Due: October 17, Thursday

How to submit: submit in class

MATH 484 students: do problems 1, 3, 4.

MATH 564 students: do problems 1, 2, 3, 4.

Problem 1 Consider the multiple regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,\ i = 1, \ldots, n$$

where ϵ_i are independent with normal distribution $N(0, \sigma^2)$. State the least square criterion and derive the least square estimators of β_1 and β_2.

Problem 2 Consider the multiple regression model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_{11} x_{i1}^2 + \beta_2 x_{i2} + \epsilon_i,\ i = 1, \ldots, n$$

where ϵ_i are independent with normal distribution $N(0, \sigma^2)$. State the least square criterion and derive the least square estimators of all coefficients.

Problem 3 For a multiple regression model with 4 variables,

1) Show that $SSR(X1, X2, X3, X4) = SSR(X1) + SSR(X2, X3|X1) + SSR(X4, |X1, X2, X3)$

2) Show that $SSE(X1, X2) - SSE(X1, X2, X3, X4) = SSR(X3, X4|X1, X2)$

3) What are the degrees of freedom that are associated with each of the following:

3a) $SSR(X1|X2)$

3b) $SSR(X2|X1, X3)$

3c) $SSR(X1, X2|X3, X4)$

3d) $SSE(X1, X2)$

4) What is the relevant extra sum of squares for testing whether or not $\beta_4 = 0$? whether or not $\beta_2 = \beta_3 = 0$?

Problem 4 In the copier maintenance dataset (Copier.txt), X_1 (the second column) is the number of copiers serviced and Y (the first column) is the total number of minutes spent by the service person. X_2 (in Model.txt) is the binary predictor variable that indicates whether the copier model is small or large. It is coded as $X_2 = 1$ if small model is used and $X_2 = 0$ if large model is used.

1) Fit the regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$ and provide the estimated regression function.

2) Estimate the effect of copier model X_2 on mean service time μ_y with a 95% confidence interval.

3) Obtain the residuals from 1) and plot them agains $x_1 x_2$. Is there any indication that an interaction term in the regression model would be helpful?

4) Fit the regression model $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_{12} x_{i1} x_{i2} + \epsilon_i$ and provide the estimated regression function.

5) Based on 4), test whether the interaction term can be dropped from the model: let the significance level $\alpha = 10\%$, state the null, alternative hypotheses, the decision rule and your conclusion. What is the P-value of the test? If the interaction term cannot be dropped from the model, describe the nature of the interaction effect.