

Name

Fall, 2009

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (20 Points) Briefly discuss the following statistics and describe how each is used in determining leverage and/or influence points.
 - (a) The diagonal elements of the hat matrix
 - (b) Cook's D
 - (c) DFBETAS
 - (d) DFFITS

2. (30 Points) A multiple regression model was used to relate a response variable to five continuous explanatory variables. There were 37 observations in the dataset. Use the following information to answer the questions below:

$$\hat{\sigma}^2 (X'X)^{-1} = \begin{bmatrix} 739.06 & -6.25 & -1.07 & -1.22 & -4.79 & -2.15 \\ -6.25 & 0.08 & -0.01 & -0.01 & 0.03 & 0.01 \\ -1.07 & -0.01 & 0.02 & 0.01 & 0.01 & 0.01 \\ -1.22 & -0.01 & 0.01 & 0.03 & 0.02 & -0.01 \\ -4.79 & 0.03 & 0.01 & 0.02 & 0.05 & 0.01 \\ -2.15 & 0.01 & 0.01 & -0.01 & 0.01 & 0.03 \end{bmatrix}$$

ANOVA for the Full Model

Source of Variation	df	SS	MS	F
Model	5	2524.77	504.96	45.97
Error	31	340.52	10.98	

The estimated parameter vector for the Full Model is
 $[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5] = [179.03, 0.58, 0.63, -0.01, -0.76, -0.10]$

ANOVA for the Reduced Model without X_3 , X_4 , and X_5

Source of Variation	df	SS	MS	F
Model	2	2365.70	1182.85	80.50
Error	34	499.60	14.69	

- (a) (10 Points) Using the full model, construct a 95% confidence interval for $\beta_1 - \beta_2$.
- (b) (10 Points) Test the hypothesis $H_o: \beta_3 = \beta_4 = \beta_5 = 0$ using a 0.05 level of significance.
- (c) (10 Points) Discuss the general problem with multicollinearity and three ways you would detect multicollinearity.

3. (30 Points) A company wanted to replace the machines used to make a certain component in one of its factories. Three different brands of machines were available, so the management designed an experiment to evaluate the productivity of the machines when operated by the company's own personnel. Six employees were randomly selected to participate in the experiment, each of whom was to operate each machine two different times. The response was an overall score, which took into account the number and quality of components produced. The partial ANOVA table is given below:

Source of Variation	df	Sum of Squares	Mean Square
Machine	X	1,755	X
Person	X	1,241	X
Machine*Person	X	462	X
Error	X	33	X

- (a) (10 Points) Write the mathematical model for this design. Be sure to describe each term in the model and state all model assumptions.
- (b) (10 Points) Fill in the rest of the partial ANOVA table by filling in the cells denoted with an X.
- (c) (10 Points) Determine the expected mean squares for each term in the ANOVA table.

4. (20 Points) Suppose we are interested in the effect of two different kinds of potting soil on the number of days to germination of three varieties of carrot seed. The following table displays the data.

Soil	Variety 1	Variety 2	Variety 3
1	6	13	14
	10	15	24
	11		
2	2	32	18
	5		9
	9		12
	8		

- (a) (10 Points) Discuss the presence/absence of interaction in the data and the effect of the interaction on the statistical analysis.
- (b) (10 Points) Calculate the least squares means for the two soil types.