Name Spring, 2013

Applied Statistics Comprehensive Examination
Regression & Linear Models

1. (30 points) An experiment was run to study the effects of two factors, A and B, on a response variable. Factor A and Factor B both have three levels. Assume that a means model will be fit to the data. Use the following information to answer the questions that follow.

   Use the following notation for the cell means:

	Factor B		
Factor A	1	2	3
1	μ_{11}	μ_{12}	μ_{13}
2	μ_{21}	μ_{22}	μ_{23}
3	μ_{31}	μ_{32}	μ_{33}

   The following table displays the sample mean and (sample size) for each cell:

	Factor B		
Factor A	1	2	3
1	10(4)	15(2)	20(2)
2	11(2)	9(2)	15(3)
3	15(3)	16(4)	22(2)

   The Mean Square Error resulting from the above model is 3.

   (a) (10 Points) Construct a complete set of mutually orthogonal contrasts for interaction.

   (b) (10 Points) Test the significance of the contrast $\mu_{11} + \mu_{12} - \mu_{21} - \mu_{22}$ at the 0.05 level.

   (c) (10 Points) Calculate the least squares means for the three levels of Factor B and briefly explain why least squares means should be used to compare column effects in this example.

2. (30 points) Suppose an effects model will be used to analyze the data in Problem 1 where the model will be $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$, where $i = 1, 2, 3$, $j = 1, 2, 3$, α_i and β_j are the main effects of Factor A and B, respectively, γ_{ij} is an interaction effect, and ϵ_{ijk} is an error term.

   (a) (20 Points) Construct a basis set of estimable functions.

   (b) (10 Points) Determine whether $\alpha_1 - \alpha_2 + \gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}$ is estimable.

3. (10 Points) In fitting a multiple regression model, you notice non-constant variance in the plot of the residuals vs. the predicted values. Two possible techniques to address the violation of the constant variance assumption are (1) use a weighted least squares model; (2) transform the response variable. Briefly describe when each of the two techniques should be used.

4. (30 Points) In fitting a multiple regression model, the full model is

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

where all four explanatory variables are continuous. The first page of SAS output on the pages that follow are from the full model and the remaining three pages of output are from the reduced model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$.

Use the SAS output on the following pages to answer the questions below.

(a) (10 Points) Using the output from the full model, test $H_o : \beta_2 = 0$ and interpret the results.

(b) (10 Points) Test $H_o : \beta_3 = \beta_4 = 0$.

(c) (10 Points) For the reduced model, state the model assumptions and comment on whether you believe that the assumptions are satisfied.

Full Model

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	25
Number of Observations Used	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	1073119	268280	30.51	<.0001
Error	20	175855	8792.75990		
Corrected Total	24	1248974			

Root MSE	93.76972	R-Square	0.8592
Dependent Mean	1269.02000	Adj R-Sq	0.8310
Coeff Var	7.38914		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-593.53745	259.19585	-2.29	0.0330
X1	1	2.51314	0.31428	8.00	<.0001
X2	1	1.90595	0.74239	2.57	0.0184
X3	1	2.65101	4.63566	0.57	0.5738
X4	1	-0.12073	0.37181	-0.32	0.7488

Reduced Model

The REG Procedure
Model: MODEL1
Dependent Variable: Y

Number of Observations Read	25
Number of Observations Used	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1067797	533899	64.83	<.0001
Error	22	181176	8235.29179		
Corrected Total	24	1248974			

Root MSE	90.74851	R-Square	0.8549
Dependent Mean	1269.02000	Adj R-Sq	0.8418
Coeff Var	7.15107		

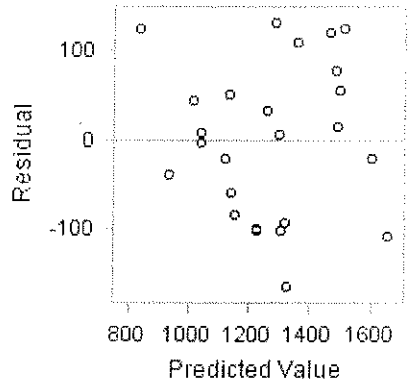
Parameter Estimates

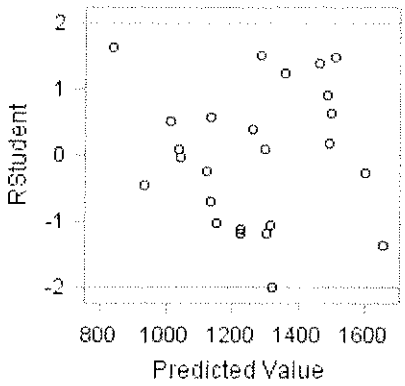
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-516.44428	189.87570	-2.72	0.0125
X1	1	2.47318	0.27531	8.98	<.0001
X2	1	1.85618	0.71573	2.59	0.0166

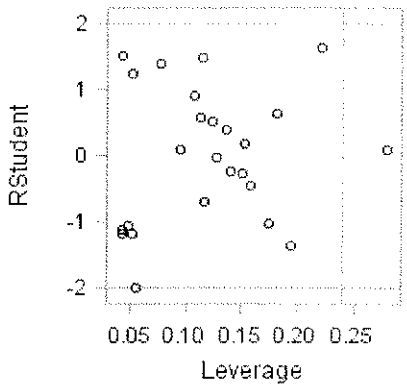
Reduced Model

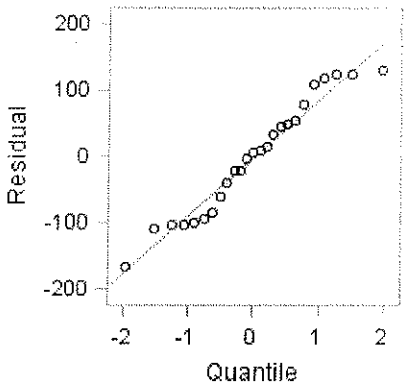
The REG Procedure
Model: MODEL1
Dependent Variable: Y

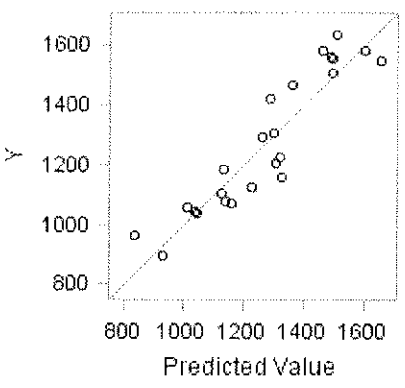
Fit Diagnostics for Y

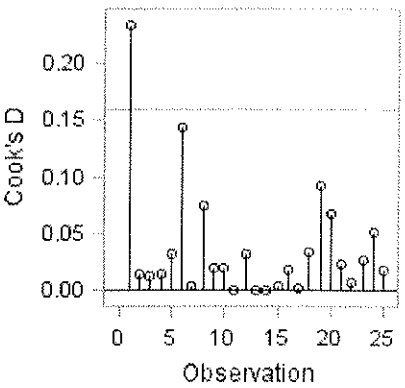


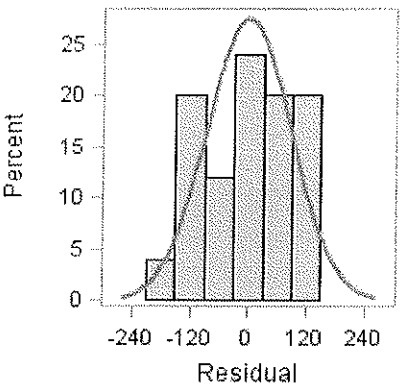


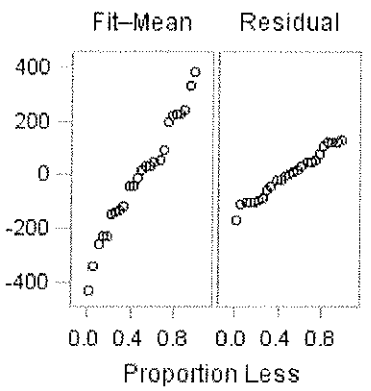




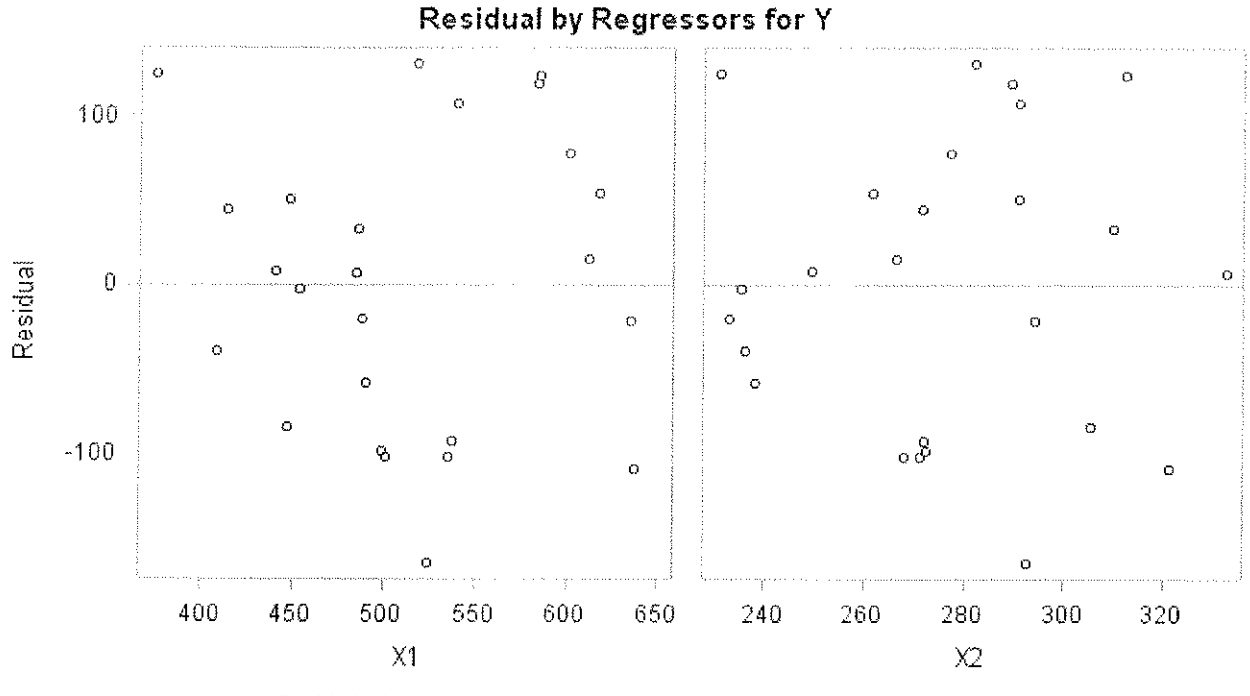








Observations	25
Parameters	3
Error DF	22
MSE	8235.3
R-Square	0.8549
Adj R-Square	0.8418


Residual by Regressors for Y
Residual
100
0
-100
400
450
500
550
600
650
X1
240
260
280
300
320
X2