Name ________________________ Spring, 2015

Applied Statistics Comprehensive Examination
Regression Methods & Linear Models

Calculators are permitted on this part of the examination.

1. (50 points) Data were collected from Kelly Blue Book for several hundred 2005 used GM cars that were considered to be in excellent condition. The following variables were collected:

    Price: suggested retail price of the used 2005 GM car in excellent condition.
    Mileage: number of miles the car has been driven
    Make: manufacturer of the car – Cadillac, Saturn, Pontiac, Chevrolet, SAAB, or Buick
    Type: body type - sedan, coupe, wagon, hatchback, and convertible
    Cylinder: number of cylinders in the engine- 4, 6, or 8
    Cruise: indicator variable representing whether the car has cruise control (1 = cruise)
    Sound: indicator variable representing whether the car has upgraded speakers (1 = upgraded)
    Leather: indicator variable representing whether the car has leather seats (1 = leather)

    Complete the following questions:

    a. For the <u>reduced model</u>, complete the following.
        i. (10 points) State the assumptions for the model and, using the graphs provided, comment on whether the assumptions are reasonable.
        ii. (5 points) Interpret the parameter estimates for the *Mileage* and *Cadillac* variables.
    b. (5 points) The Box-Cox method suggested that a logarithmic transformation of *price* (the dependent variable) would be most appropriate. Which of the assumptions of the model are affected by this transformation?
    c. (15 points) Is there evidence that the addition of the independent variables to the full model improves the predictive ability of the model? State the appropriate hypotheses and conduct the appropriate hypothesis test.
    d. The highest variance inflation factor (VIF) from the full model was calculated to be 5.35.
        i. (5 points) What does this information tell you about the model?
        ii. (5 points) Explain why none of the VIFs for the reduced model can be higher than 5.35.
        iii. (5 points) Does transformation of the dependent variable have any effect on the VIFs? Justify your answer.

Analysis of Variance – Full Model

```
Source       DF       Adj SS      Adj MS  F-Value  P-Value
Regression   15  72118625487  4807908366   597.32    0.000
  Mileage     1   1842143485  1842143485   228.86    0.000
  Make        5  14314576324  2862915265   355.68    0.000
  Type        4   5264542021  1316135505   163.51    0.000
  Cylinder    2   9125078761  4562539381   566.83    0.000
  Cruise      1     32599504    32599504     4.05    0.045
  Sound       1     48523755    48523755     6.03    0.014
  Leather     1     36625254    36625254     4.55    0.033
Error       788   6342757374     8049184
Total       803  78461382861

      S    R-sq  R-sq(adj)
2837.11  91.92%     91.76%
```

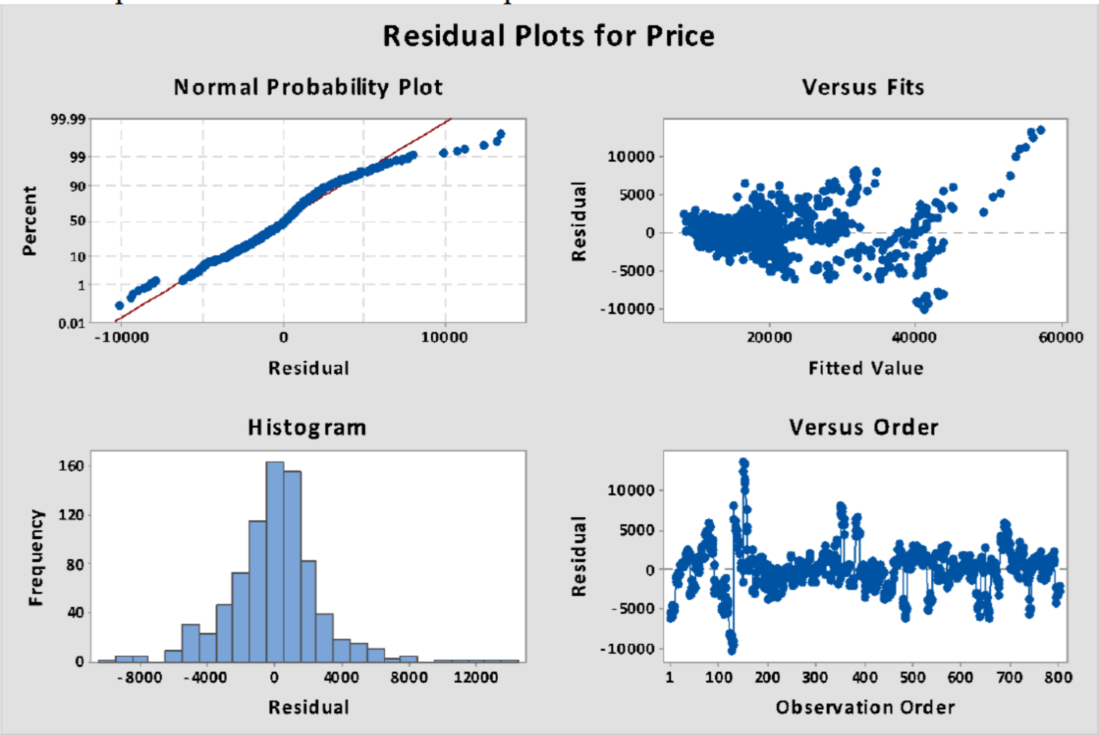
Analysis of Variance – Reduced Model

```
Source          DF       Adj SS       Adj MS  F-Value  P-Value
Regression       6  52149784493   8691630749   263.28    0.000
  Mileage        1   1566276900   1566276900    47.44    0.000
  Make           5  50544194118  10108838824   306.21    0.000
Error          797  26311598368     33013298
Total          803  78461382861

      S    R-sq  R-sq(adj)
5745.72  66.47%     66.21%

Term            Coef  SE Coef  T-Value  P-Value
Constant       24306      818    29.71    0.000
Mileage      -0.1709   0.0248    -6.89    0.000
Make
  Cadillac     19862      909    21.84    0.000
  Chevrolet    -4520      718    -6.29    0.000
  Pontiac      -2592      796    -3.26    0.001
  SAAB          8771      838    10.47    0.000
  Saturn       -6852      981    -6.98    0.000
```

Residual plots for the reduced model are presented below:



2. (30 points) The observations in the table below were obtained from an experiment that was run to study the impact of two factors on a response. There were two levels of factor A and three levels of factor B. The researchers decided to fit an effects model without interaction.

	B		
A	1	2	3
1	2, 4	3	6
2	1	2, 6	10

a. (10 points) Find the normal equations for these particular data.

b. (10 points) Let α_i be the effect associated with the ith level of factor A. Determine whether each of the following expressions is estimable, making sure to justify your answers. (*i*) α_1 (*ii*) $\alpha_1 - \alpha_2$

c. (10 points) Create a plot for assessing the presence or absence of interaction. Using this plot, comment on the appropriateness of the decision not to include an interaction term.

3. (20 points) An experiment was run to investigate the effect of two seed types and three fertilizer levels on crop yield. There were three observations for each combination of seed type and fertilizer level, and the cell means are given in the table below. A full model with interaction was fit.

	Fertilizer Level		
Seed Type	1	2	3
1	14	18	19
2	12	11	16

a. (10 points) Provide a complete set of orthogonal contrasts that could be used to obtain the sums of squares for seed type, fertilizer level, and the interaction.

b. (10 points) It was found that $SSE = 22$. Create a contrast that tests whether seed types 1 and 2 interact with fertilizer levels 1 and 2, and test the significance of this contrast at level 0.05.