Statistics 512: Homework#9
Due April 11, 2014 BEFORE CLASS

1. For this problem, the idea is to demonstrate the similarity between regression with dummy variables and ANOVA. To do this run the SAS code `stat512prob8.sas`.

   (a) Compare the ANOVA table and parameter results from the GLM analysis and Parameterization #1. What do the coefficients associated with X_1 and X_2 (i.e. b_1 and b_2) estimate in terms of treatment means? What constraint system does this parameterization correspond to?

   (b) Compare the ANOVA table and parameter results from the GLM analysis and Parameterization #2. What do the coefficients associated with X_1 and X_2 (i.e. b_1 and b_2) estimate in terms of treatment means? What constraint system does this parameterization correspond to?

   (c) Calculate $b_0 + b_1$ for both the parameterizations and show that the answers are the same. What does this quantity estimate in terms of the treatment means (i.e. why are they the same)?

   (d) Calculate $b_1 - b_2$ for both the parameterizations and show that the answers are the same. What does this quantity estimate in terms of the treatment means (i.e. why are they the same)?

The next three problems use the dataset from Problem 16.11 described on page 725 of KNNL, and continue the analysis begun on Problem Set 8.

2. Use the Tukey multiple comparison method to determine which pairs of machines differ significantly. Summarize the results.

3. Suppose you want to compare the average of the first two machines with the average of the last four. Use the `estimate` and `contrast` statements in `proc glm` to test the appropriate hypothesis. Report the estimated value of this contrast with its standard error; state the null and alternative hypotheses, the test statistic with degrees of freedom, the p-value and your conclusion.

4. Check assumptions using the residuals. Turn in the plots/output you used to check the assumptions and state your conclusions.

The remaining problems use the dataset from Problem 18.15 on page 804 of KNNL.

5. KNNL 18.15 (Omit part e). Please do not print out all 80 values for part a); it is sufficient to plot them in part b).

6. A rather simple approximation of the Box-Cox procedure is the following:

   (a) Compute the mean and standard deviation for each treatment factor level.

   (b) Take the log of both the mean and standard deviation.

(c) Fit the regression model $\log(\sigma_i) = \beta_0 + \beta_1 \log(\mu_i) + \epsilon$ using the observed means and standard deviations as the data for μ_i and σ_i respectively (there are 4 "observations" in this dataset).

(d) Set $\hat{\lambda} = 1 - b_1$ where b_1 is the estimate for β_1 obtained in part c.

Use the Helicopter service data to perform this approximation. What value of λ appears reasonable according to this method?

7. Define a new response variable by adding 1 to the original response. (This will avoid 0's which mess up the log and reciprocal transformations.) Then use SAS's Box-Cox procedure to determine an appropriate transformation. `Proc transreg` can be used to perform ANOVA if we tell it `shift` is a `class` variable, as in the following:

```
proc transreg data=helicopter;
  model boxcox(usesplus1) = class(shift);
```

8. KNNL 18.16 (Omit the coefficient of correlation in part b).

9. Use the Tukey multiple comparison method for differences in means on both the untransformed and transformed Helicopter service data to determine which shifts differ significantly. Summarize and compare the results.