STAT 514 Homework 3

Due: Sep 19

1. A factor with three levels was studied in an experiment. The data is given as follows, in which the first column includes the treatments and the second column includes the responses. You can download the data, hw.dat, from the Blackboard Learning:

```
1 2.23
1 3.04
. ...
3 8.12
```

   (a) Test the hypothesis that there is no difference across the treatments (use $\alpha = .05$).

   (b) Use proper plots to check whether the constant variance assumption is valid. Can you use a formal test to support your conclusion?

   (c) Generate the $\log s_i$ vs. $\log \bar{y}_{i\cdot}$ plot (template code "trans.sas" is available on Blackboard) and estimate the possible transformation for variance stabilization.

   (d) Use the formal Box-Cox procedure to identify the optimal transformation. You need use template code "trans1.sas" for this data set and generate proper output and plot to make the choice.

   (e) Repeat (a) and (b) for the transformed response. You may need use some sas function in the data step to generate the new responses.

2. Four different designs for a digital computer circuit are being studied to compare the amount of defects. The following data have been obtained ("defects.dat" on the Blackboard):

```
design defect
1 7
1 2
1 4
1 7
1 2
. .
4 2
2 7
```

   (a) Is the amount of defects present the same for all four designs? (use $\alpha = 0.05$).

(b) Analyze the residuals from a). In particular, how do you think about the normality assumption? Can you use any formal test to support your conclusion?

(c) Use the Kruskal-Wallis test for the data and compare the results with a).

3. In a study of the effect of glucose on insulin release, identical specimens of pancreatic tissue were equally and randomly assigned to three different levels of glucose concentration (low, medium, high). The amount of insulin produced by each tissue after treatment was recorded. The data set, "insulin.dat", can be downloaded from the Blackboard Learning. In "insulin.dat", the first column contains the amounts at the low concentration, the second column the amounts at the medium concentration, and the third column the amounts at the high concentration. To read a data set like this, do the following in the data step to create a data set suitable for the glm procedure.

```
data insulin;
infile 'H:\dataset\insulin.dat';
input t1 t2 t3;
y=t1; trt=1; output;
y=t2; trt=2; output;
y=t3; trt=3; output;
drop t1 t2 t3;
```

This creates a treatment variable (trt) and a response variable (y).

(a) Test the hypothesis that there is no difference across treatments in the amount of insulin produced (use $\alpha = 0.01$).

(b) Diagnose whether the assumptions are valid?

(c) Construct 99% CIs for the average insulin amounts at the low, medium and high glucose concentrations separately (not simultaneously). In other words, compute the 99% CIs for $L = \mu_1$, $L = \mu_2$, and $L = \mu_3$ separately. Based on each confidence interval, does it appear the average amount of insulin is significantly different than 3.5?