Assignment 4 (Due Thursday 9/19/2013)

1. To study the effect of pesticides on birds, a scientist randomly and equally allocates
   \( N = 65 \) chicks to five diets (a control and four with a different pesticide included). After a
   month each chick’s calcium content (mg) in 2 cm length of bone is measured resulting in the
   following:

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>pesticide 1</th>
<th>pesticide 2</th>
<th>pesticide 3</th>
<th>pesticide 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>11.54</td>
<td>11.00</td>
<td>11.42</td>
<td>11.44</td>
<td>11.28</td>
</tr>
<tr>
<td>St.Dev</td>
<td>.27</td>
<td>.47</td>
<td>.31</td>
<td>.42</td>
<td>.31</td>
</tr>
</tbody>
</table>

   Construct the ANOVA table (i.e. compute the between and within SS) and test if there
   appears to be any differences in means (use \( \alpha = 0.01 \))

2. 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 class website. 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). Assess whether the model assumptions are valid?

c). Construct 99% CIs for the average insulin amounts at the low, medium and high glucose
concentrations separately. The formula can be found in Montgomery Section 3.3.3 Equation
(3.12). Based on each confidence interval, does it appear the average amount of insulin is
significantly different than 3.5?

3. A factor with three levels was studied in an experiment. The data is given as fol-
lows, in which the first column includes the treatments and the second column includes the
responses. You can download the data, hw4.dat, from the class website.

\begin{verbatim}
1  2.23
1  3.04
   ...
2  3.65
   ...
3  8.53
   ...
3  8.12
\end{verbatim}

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 $\bar{y}_i$. plot (trans.sas) and estimate the possible transformation
for variance stabilization.
d). Use the formal Box-Cox procedure to identify the optimal transformation. You need use trans1.sas for this data set and generate proper output and plot to make the choice.

e). Repeat a) and b) for the transformed data. You need to use some sas functions in the data step to generate the new data.

4. 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 class website):

design  defect
1      7
1      2
1      4
1      7
1      2
2      10
2      6
2      9
2      7
2      5
3      16
3      13
3      11
3      13
3      13
4      5
4      5
4      2
a). Is the amount of defects present the same for all four designs? Use $\alpha = 0.05$.

b). Check the model assumptions. 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).