STAT 514 Homework 2

Due May 31

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, "hw2.dat", from the course website.

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

   (a) Test the hypothesis that there is no difference across the treatments (use $\alpha = 0.05$).

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

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

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

   (e) Repeat (a) and (b) for the transformed response. You may need to use some 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 course website):

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

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

(b) Analyze the residuals from a). In particular, what do you think about the normality assumption? Use a 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 course 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 mean 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 that the average amount of insulin is significantly different from 3.5?

4. To study the effects of pesticides on birds, a scientist randomly and equally allocates $N = 65$ chicks to five diets (a control and four diets each 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 data:

		Pesticides			
	Control	1	2	3	4
Mean	11.54	11.00	11.42	11.44	11.28
Std Dev.	0.27	0.47	0.31	0.42	0.31

(a) Construct the ANOVA table (i.e. compute the between and within SS) and test if there appears to be any difference in the population means (use α=0.01).

(b) What is the estimate of the error variance from this data? (Bonus) Create a 95% lower confidence bound for the error variance. Hint: Use Question 7.

(c) If you knew that there is no difference in the population means, how would your answers for b) change? Recalculate the confidence bound for bonus points.

5. An experiment is conducted to study the impact of hormone on the liver of rats. Two types of hormones (I, II) each with two levels are involved. We consider the following four treatments: (A) Hormone I at high level; (a) Hormone I at low level; (B) Hormone II at high level; (b) Hormone II at low level. Each treatment is applied to six randomly selected rats. The response is the amount of glycogen (in mg) in the liver of a rat after a certain period of time.

Treatment	Response					
A	106	101	120	86	132	97
a	51	98	85	50	111	72
B	103	84	100	83	110	91
b	50	66	61	72	85	60

Suppose we are interested in the following three contrasts:

Comparison	A	a	B	b
Hormone I vs Hormone II	1	1	-1	-1
Low Level vs High Level	1	-1	1	-1
Equivalence of Levels	1	-1	-1	1

(a) Use ANOVA to check if there are differences between the treatments ($\alpha = 0.05$).

(b) Are these contrasts orthogonal? Why or why not?

(c) Compute the single degree of freedom sums of squares (i.e. SS for contrasts) and test each null hypothesis. Interpret the results. (NOTE: Be careful if you use a character string variable to denote the treatment levels. The order of the treatments SAS uses in the contrast statement is different from A, a, B, b. To avoid this, code the treatments as 1, 2, 3 and 4 respectively.)

6. With the standard notations for single factor ANOVA that were introduced in Lecture 3, prove that $\text{E}(\text{MS}_{\text{Treatment}}) = \sigma^2 + \sum n_i \tau_i^2/(a-1)$. Hint: Use the short-cut formula of $\text{SS}_{\text{Treatment}}$ from Page 10 and the distributions of $\bar{y}_{i.}$ from Page 9 of Lecture 3. You will have to assume that $\sum n_i \tau_i = 0$.

7. (Bonus) Derive formulas for $100(1-\alpha)\%$ CI, upper and lower confidence bounds for σ^2 in single factor ANOVA. Hint: Use the distribution of the point estimator of σ^2.