Statistics 5301 — Exam 1 — solutions
April 22, 1996
Gary W. Oehlert

1. Did the data require a transformation in order to meet the usual ANOVA assumptions? If so, which transformation?
I thought that the variance increased slightly with the mean. The boxcox SS was smallest on the log scale, and I went ahead and used the log transformed data. (The boxcox SS for log was 38606 vs 41092 on natural scale; this is more or less at the edge of a 95% CI.) The rankit plot looked great.

2. Does growth temperature have any effect on the response?
The main effect of growth temperature was not significant, but growth temperature has a significant interaction with variety; variety 1 has a higher log activity at the first growth temperature, while variety 2 has a higher log activity for the second growth temperature.

```
Cmd> tabs(lact,gt,v,means:T)
(1,1)        5.921       5.7044
(2,1)       5.8476       5.7508
```

There is also some indication that growth temperature interacts with analysis temperature.

3. How would you describe the effect of analysis temperature on the response?
Analysis temperature is a quantitative variable running from 10 to 40 degrees. The response increases till about 30 degrees and then decreases. The decrease from 30 to 40 degrees is steeper than the increase from 20 to 30. Quantitatively, the linear, quadratic, and cubic terms in analysis temperature are signficant, but none of the higher order terms is. (The cubic term models the asymmetry of the shape.) There is also a slight growth temperature by analysis temperature interaction, which shows up as a slightly different linear term for analysis temperature for the two different growth temperatures. Specifically, growth temperature 2 gives higher activities than growth temperature 1 at lower analysis temperatures and vice versa at high analysis temperatures.

4. Which factor has the greatest effect on the response?
On the log scale, the total SS explained by the model was 3.8524. Of this total, the linear, quadratic, and cubic terms for analysis temperature explain 3.007 and variety explains 0.5896; the remaining 27 degrees of freedom in the model explain the remaining 0.2557 of the SS. Several of these remaining terms are statistically significant, but do not explain a very large fraction of the variability.

5. Which effects would you conclude are significant?
My model would include linear, quadratic, and cubic terms in analysis temperature, variety, growth temperature, the variety by growth temperature interaction, and growth temperature by linear analysis temperature terms.

6. The data below give women of child bearing age per birth in 1985 for 51 sampling areas in the US divided into 5 regions (data from Lunneborg).

	North East	Mid West	South Atlantic	South Central	West
$\bar{y}_{i\bullet}$	70.12	64.045	67.126	63.321	55.935
n_i	9	12	9	8	13
$\bar{y}_{\bullet\bullet}$	63.48				
MSE	31.53				

Just looking at the data, it would appear that the number of women per live birth is different in the west than in the other regions. Test the null hypothesis that the average number of women per live birth in the west is the same as the average of the other four regional averages.

We can investigate this difference with a contrast such as (1,1,1,1,-4). The main thing to notice is that the contrast is suggested by the data, so that we must evaluate its significance with the Scheffe method. The contrast value is

$$70.12 + 64.045 + 67.126 + 63.321 - 4 \times 55.935 = 40.87$$

The sum of squares for the contrast is

$$SS = \frac{40.87^2}{1/9 + 1/12 + 1/9 + 1/8 + 4^2/13} = 1005.6$$

Scheffe treats this contrast SS as if it had all 4 df between the 5 groups, so the Scheffe F is 1005.6/4/31.53 = 7.97 with 4,46 df. The p-value is very small (about 5e-5), so I reject the null hypothesis.

7. Consider the following MacAnova output. What would you conclude about the signficance of the main effects and interactions?

```
Cmd> anova("y=a*b")
Model used is y=a*b
WARNING: summaries are sequential
                DF           SS          MS
CONSTANT         1        20552       20552
a                1        60.44       60.44
b                2       140.67       70.34
a.b              2        46.14       23.07
ERROR1          41        450.4       10.99

Cmd> anova("y=b*a")
Model used is y=b*a
WARNING: summaries are sequential
                DF           SS          MS
CONSTANT         1        20552       20552
b                2       171.44       85.72
a                1        29.67       29.67
a.b              2        46.14       23.07
ERROR1          41        450.4       10.99
```

These data are unbalanced since the order of the terms changes the SS. For these data, the F for interaction is 23.07/10.99 = 2.1; p-value about .14, so I do not conclude that we need an interaction term. For the main effects, I look at each adjusted for the other. For A, the F is 29.67/10.99 = 2.7; p-value about .1, so I conclude I don't need an A main effect. For B, the F is 70.34/10.99 = 6.4; p-value about .004, so I reject the null hypothesis of no B effect.