Lab 2 **Statistics for Laboratory Scientists II**

Due: 5 May 2006

In this lab you should apply what you learned about analysis of variance (ANOVA) in class. *Please include an edited version of your R code as an appendix.*

1. Below are the descriptions of two experiments. The resulting data of those experiments can be analyzed using ANOVA—your choices are one-way ANOVA, nested ANOVA, and two-way ANOVA. Your task is to analyze the data.

   Clearly state your assumptions. Check your model assumptions, and use a data transformation if necessary. Summarize your findings for each of the data sets in a sentence or two. (Include the ANOVA table.)

   You must be careful to ensure that certain treatment columns are properly treated as *factors* (rather than *numeric* vectors). You may need to use code such as the following:
   `ferm$oxygen <- factor(ferm$oxygen)`

   (a) Patients with advanced cancers of the stomach, bronchus, colon, ovary or breast were treated with ascorbate. The purpose of the study was to determine if patient survival differed with respect to the organ affected by the cancer (Reference: Cameron, E. and Pauling, L. (1978), PNAS, 75.) The data are in the file `cancer.csv`. Column one is the survival time in days, column two is the organ affected by the cancer.

   (b) The effect of oxygen level on fermentation end products was examined in the article "Effects of Oxygen on Pyruvate FormateLyase in Situ and Sugar Metabolism of Streptococcusmutans and Streptococcus samguis" (Infection and Immunity 1985, p129-134). Four oxygen concentrations (0, 46, 92, 138 microM) and two sugar types (galactose and glucose) were used. The amount of ethanol was measured for each oxygen-sugar combination. The data are in the file `fermentation.csv`.

2. In this experiment the effect of caffeine on the performance of a simple task was examined (data from Cliff Lunneborg, University of Washington). Thirty volunteer subjects (male college students) were trained at a finger tapping task. The subjects were then randomly divided into three treatment groups. One group drank a beverage containing 0 mg caffeine, a second group drank a beverage containing 100 mg caffeine, the third group drank a beverage containing 200 mg caffeine. Two hours after drinking, their taps per minute were measured. The data are available in `caffeine.csv`.

   Note that the computer code for lecture 9 will be useful to you.
   `http://www.biostat.jhsph.edu/~kbroman/labstat/fourth/comp09.R`

   (a) Use the appropriate ANOVA to confirm that there are differences among the treatment groups.

   (b) Carry out pairwise comparisons using Bonferroni's method. What can you conclude?

   (c) Find Tukey's confidence intervals (HSD). Again, what do you conclude?