

## Biostatistics 537: Longitudinal Data Analysis - Fall 2009

Problem Set 1 - Due: September 10, 2009

There are two data sets for this problem set: `vt.dat` and `box.dat`. Both of these can be downloaded at the class website <http://www.uic.edu/classes/bstt/bstt513/> (towards the bottom of the page).

1. The data (in `vt.dat`) consist of a sample of 22 subjects from a study of affective facial expressions (Vasey and Thayer, 1987). In this study several pieces of music were played to each subject in an attempt to elicit selected affected states. Trial 1 was a baseline, relaxing music condition. Trial 2 was designed to produce positive effects, and trial 3 was designed to produce agitation. Each trial lasted 90 seconds and the response variable at each trial was the mean electromyographic (EMG) amplitude ( $\mu\text{V}$ ) from the left brow region. The variable order in the datafile is subject number followed by the three trial EMG measurements. Consider the question of whether each of the two non-baseline trials differed from the baseline trial in EMG. For this, choose an appropriate set of contrasts that relate to this question. Posit a reasonable model and interpret the results. In doing this, you should choose between the univariate and multivariate repeated measures analysis of variance approaches, and give reasons for the approach you choose. If you choose the univariate repeated measures model, estimate the intraclass correlation and comment on its magnitude. Also, to better normalize the outcome variable use a log transformation. Here's some SAS code that will help with this problem:

```
FILENAME in1 'C:\vt.dat';
TITLE1 'Vasey & Thayer Data - Left Brow EMG amplitudes in 22 subjects';

/* input the data in multivariate format */
DATA MultData; infile in1;
INPUT subject 2. (trial1 trial2 trial3) (4.) ;
    ltrial1 = log(trial1);
    ltrial2 = log(trial2);
    ltrial3 = log(trial3);

PROC MEANS;
RUN;

/* set up the data in univariate format */
DATA UNIDATA;
SET MultData;
    ARRAY T(3) ltrial1-ltrial3;
    DO TRIAL = 1 TO 3;
        logEMG = t(TRIAL);
    OUTPUT;
END;
DROP ltrial1-ltrial3;
RUN;

PROC MEANS;
RUN;
```

2. Box (1950) described an experiment in which 30 rats were randomly assigned to three treatment groups. Group 1 was a control group, group 2 had thiouracil added to their drinking water, and group 3 had thyroxin added to their drinking water. Thyroxin is one of the two hormones of the thyroid gland, and the primary action of thyroxin is to increase the metabolic rate. Thiouracil, on the other hand, is an antithyroid drug that blocks thyroid action. While both groups 1 and 2 had 10 rats each, group 3 only had 7 rats (due to an unspecified accident at the beginning of the experiment). The body weights of each of the 27 rats were recorded at the beginning and at weekly intervals for four weeks. The file `box.dat` contains: `GROUP`, `RAT`, `BW0`, `GAIN1` (`BW1-BW0`), `GAIN2` (`BW2-BW1`), `GAIN3` (`BW3-BW2`), and `GAIN4` (`BW4-BW3`). Convert the weekly `GAIN` values to `BW` values and analyze the five repeated measures (**`BW0`**, **`BW1`**, **`BW2`**, **`BW3`**, **`BW4`**) using either the univariate or multivariate analysis of variance approach. Use orthogonal polynomials to perform a growth curve analysis, and then examine whether the curves are parallel for the three groups. In terms of between-group comparisons, the investigators were primarily interested in contrasting each drug group to the control group. What degree of polynomial is required to express any group differences across time? Summarize your findings (note: a plot of the group means across time helps).