LONGITUDINAL DATA ANALYSIS
Homework I, 2005

1. Suppose $\mathbf{A}$ and $\mathbf{B}$ are both 2×2 matrices with

$$\mathbf{A} = \begin{pmatrix} 6 & 3 \\ -2 & 5 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} -4 & 10 \\ 7 & 6 \end{pmatrix}$$

   (a) Verify that $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$.

   (b) Verify that $|\mathbf{A}| = 1/|\mathbf{A}^{-1}|$.

   (c) Verify that $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$.

2. Suppose that a_1 and a_2 are constants, and y_1 and y_2 are (possibly correlated) random variables with means μ_1 and μ_2 respectively. Show that $cov(a_1y_1, a_2y_2) = a_1a_2cov(y_1, y_2)$ **by using the definition of covariance**.

3. **Exploratory Data Analysis**

   A clinical trial for treatment of back pain was conducted to compare two treatments. The trial contained 27 subjects randomized to either intercostal/epidural analgesic (14 subjects group 1) and morphine infusion analgesic (13 subjects group 2). Patients were asked to give an assessment of their state of disturbance or disconfort on four different scales, one verbal rating scales (VRS) and three visual analogue scales (VASs). This was done at each of four times following treatment.

   The data file `back.raw` is on the web. The variables appearing in columns are case number, then treatment group. These are followed by four sets of variables, each comprising the five values: pain VRS (scores 1 to 5), pain VAS (scores 0 to 100), anxiety VAS (score 0 to 100), aleterness VAS (scores 0 to 100), and time since treatment (minutes). Missing values are coded as -9. *Look at the data sets readme file on the course web site.*

   Scientific interest in the study is to compare the two treatments with respect to each of several outomes: pain, anxiety, and alterness.

   (a) Download these data from the course web site and read them into Stata. Make sure that you have them in "long" format. To do this, you will probably have to use the `reshape` commands. Make sure you set the values -9 to missing when constructing the data set.

   (b) Describe the structure of the data set, including distribution of observation times, and a list of baseline variables and time-varying variables. Are the data balanced? Equally spaced?

   (c) Choose one of the three VAS response scores. Graphically explore this response with respect to time and treatment. Write a few sentences summarizing your findings to a non-statistical audience.

   (d) Explore the correlation structure of the variable you studied in (c) using the autocorrelation function or the variogram. Justify your choice. Make sure to remove covariate effects as you see appropriate. Describe your results.

4. **Confirmatory Data Analysis**

   Use the data from the clinical trial of back pain to assess the effects of treatment on alertness. From preliminary discussions with the investigators, you learn that it is possible that the two analgesics may behave differently from one another in their effects on alertness over time. I.e. the contrast between the two treatments may be just in the overall level of alertness, or it may be in both the overall level and the trend with respect to time.

(a) Supposing the response variable to be alterness, write down a regression model for the mean response that captures the scientific questions about the relative effects of the two analgesics.

(b) Fit the model (a) with ordinary least squares (OLS)

(c) Using your model (a), explore the correlation structure of the data using the variogram *(hint: remove the effect of time, treatment, and their interaction).*

(d) Based on your results of (c), choose a model for the correlation structure that can be used in model estimation and testing for the mean model in (a). Explain why you choose the correlation structure you did.