Stat 771, Fall 2011: Homework 1

Due Wednesday, February 6

1. Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \text{ and } \mathbf{c} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

   (a) Find $3\mathbf{B}$.

   (b) Find $\mathbf{A} - \mathbf{B}$.

   (c) Find $\mathbf{AB}$.

   (d) Find $|\mathbf{A}|$.

   (e) Find $\mathbf{A}^{-1}$.

   (f) Is $\mathbf{A}$ full rank? Why or why not?

   (g) Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Show

$$\mathbf{x}'\mathbf{A}\mathbf{x} = 2(x_1^2 + x_1x_2 + x_2^2).$$

   (h) Use $\mathbf{A}^{-1}$ to solve the system of two equations in two unknowns

$$\left\{ \begin{array}{rcrcl} 2x_1 & + & x_2 & = & 1 \\ x_1 & + & 2x_2 & = & 3 \end{array} \right\}.$$

   That is, solve the system

$$\mathbf{A}\mathbf{x} = \mathbf{c}.$$

   (i) Show $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$.

   (j) Let $\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$ be a *random vector* with mean $\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ and covariance matrix $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$. Use results on p. 46 to find $E(\mathbf{AY} + \mathbf{c})$ and $var(\mathbf{AY} + \mathbf{c})$.

2. Say Y_{ij} is the j^{th} measurement on subject i, where $i = 1, \ldots, n$ and $j = 1, \ldots, 4$. All n individuals have the same mean vector (multivariate one-sample problem). Define

$$\mathbf{Y}_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \\ Y_{i4} \end{bmatrix} \text{ and } E(\mathbf{Y}_i) = \boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}.$$

   Say we want to show that the mean changes over time. The null hypothesis is that *this doesn't happen* $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. Find a 3×4 matrix $\mathbf{C}$ such that

$$\mathbf{C}\boldsymbol{\mu} = \mathbf{0} \Leftrightarrow \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

   Hint: all four means are equal if and only if three pairs of means are *simultaneously* equal.

3. Let

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \ var(\mathbf{Y}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}, \text{ and } \mathbf{c} = \begin{bmatrix} c_1 & c_2 \end{bmatrix}.$$

   Use $var(c_1Y_1 + c_2Y_2) = c_1^2 var(Y_1) + c_2^2 var(Y_2) + 2c_1c_2 cov(Y_1, Y_2)$ to show

$$var(\mathbf{cY}) = \mathbf{c}\boldsymbol{\Sigma}\mathbf{c}'.$$

4. Consider the dental data of Example 1 (pp. 3–4 in the notes). Let Y_{ij} be the jth measurement on child i, where $i = 1, \ldots, 27$. Let $a_1 = 8$, $a_2 = 10$, $a_3 = 12$, and $a_4 = 14$ be the ages at the four measurements. Let $x_i = 0$ if child i is a girl and $x_i = 1$ if a boy.

   (a) Obtain profile plots (or "spaghetti" plots) of the boys and girls separately, but on the same scale. Do there appear to be differences between boys and girls? Elaborate. I posted some sample SAS code to aid you in this.

   (b) In your favorite SAS procedure (e.g. GLM or REG, GLM would be easier here) or some other package, fit the regression model

$$Y_{ij} = \beta_{00} + \beta_{01}x_i + \beta_{10}a_j + \beta_{11}a_j x_i + e_{ij}.$$

   Think about this model; in words what does it assume? Draw a hypothetical mean functions versus age for boys and girls separately.

   Formally test that boys and girls have the exact same mean growth over time $H_0 : \beta_{01} = \beta_{11} = 0$. Formally test that boys and girls grow at the same rate, but have different intercepts $H_0 : \beta_{11} = 0$.

   An assumption going into this model is that $e_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$; this will be relaxed later on when we allow $\mathbf{e}_i = (e_{i1}, e_{i2}, e_{i3}, e_{i4})'$ to be *correlated*.

   (c) Now let's **not** assume that the overall population means follow a line, but rather are unstructured:

$$Y_{ij} = \mu_j + \delta_j x_i + e_{ij}.$$

   Formally test no difference between boys and girls $H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0$. This is very easily carried out in SAS PROC GLM using the CLASS statement for both GENDER and AGE, and the appropriate CONTRAST statement. It may help to rewrite the model as

$$Y_{ijk} = \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk},$$

   where $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are age effects and β_1, β_2 are gender effects. Then you are formally testing

$$H_0 : \left\{ \begin{array}{l} \mu_{11} - \mu_{12} = \beta_1 + (\alpha\beta)_{11} - \beta_2 - (\alpha\beta)_{12} = 0 \\ \mu_{21} - \mu_{22} = \beta_1 + (\alpha\beta)_{21} - \beta_2 - (\alpha\beta)_{22} = 0 \\ \mu_{31} - \mu_{32} = \beta_1 + (\alpha\beta)_{31} - \beta_2 - (\alpha\beta)_{32} = 0 \\ \mu_{41} - \mu_{42} = \beta_1 + (\alpha\beta)_{41} - \beta_2 - (\alpha\beta)_{42} = 0 \end{array} \right\}.$$

   My code looks like

```
contrast 'H0: no difference' gender 1 -1 age*gender 1 -1  0  0  0  0  0  0,
                             gender 1 -1 age*gender 0  0  1 -1  0  0  0  0,
                             gender 1 -1 age*gender 0  0  0  0  1 -1  0  0,
                             gender 1 -1 age*gender 0  0  0  0  0  0  1 -1;
```