ST 732, Homework 1, Spring 2011
Due Wednesday, January 26, 2011.

The first few exercises are meant to familiarize you with some matrix operations that are the basis for
the analysis used throughout the course. Future homeworks will not involve matrix algebra problems
like these except where relevant.

1. Suppose $A$ and $B$ are both $2 \times 2$ matrices with

$A = \begin{pmatrix} 11 & 7 \\ -6 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & -3 \\ 9 & 1 \end{pmatrix}$.

(a) Verify that $\|A\| \|B\| = \|AB\|$.
(b) Verify that $\|A\| = 1/\|A^{-1}\|$.
(c) Verify that $\text{tr}(AB) = \text{tr}(BA)$.

2. Let $Y_1$ and $Y_2$ be random variables with means $\mu_1$ and $\mu_2$, variances $\sigma_1^2$ and $\sigma_2^2$, and covariance $\sigma_{12}$, respectively. Let $c_1$ and $c_2$ be constants.

(a) Starting with the definition of variance in equation (3.2) of the notes and results
given in Chapter 3, show that

$$\text{var}(c_1Y_1 + c_2Y_2) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + 2c_1c_2\sigma_{12}.$$ 

(b) Let

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$ 

Write down the covariance matrix $\Sigma$ of $Y$.

(c) Let

$$c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$ 

Verify that $\text{var}(c^T Y) = c^T \Sigma c$ (bottom of p. 45 of the notes) by evaluating each side of this
equation and showing that they are equivalent.

The covariance matrix of a linear combination of elements of a random vector is of routine interest in longitudinal analysis.

3. Let $Y_1$ and $Y_2$ be as in Problem 2. Let

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$$ 

(a) Find $CY$ and $\text{var}(CY)$ using results from Problem 1 and the definition of covariance.
(b) By doing the matrix multiplication $C\Sigma C^T$, show that $\text{var}(CY) = C\Sigma C^T$ as on p. 46 of
the notes, where $\Sigma$ is the matrix you found in Problem 1(b).

The covariance matrix of a general linear function of elements of a random vector is of routine interest in longitudinal analysis.
4. (a) Suppose we have the following statistical model:

\[ Y_{ij} = \mu + b_i + e_{ij}, \]

where \( i = 1, \ldots, m \) and \( j = 1, \ldots, n \), \( b_i \sim N(0, \sigma_b^2) \), \( e_{ij} \sim N(0, \sigma_e^2) \), \( b_i \) and \( e_{ij} \) are statistically independent for each \( i \) and \( j \), and \( e_{ij} \) and \( e_{ik} \) are statistically independent for any two values \( j, k = 1, \ldots, n \). Find the variance of \( Y_{ij} \) and the covariance and correlation between any two values \( Y_{ij} \) and \( Y_{ik}, j \neq k \).

(b) Let \( n = 3 \), and define

\[ Y_i = \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{pmatrix}. \]

Find the covariance matrix of \( Y_i \), and express it in terms of a \((3 \times 3)\) identity matrix and the \((3 \times 3)\) matrix \( J \) whose elements are all equal to 1.

(c) Show that the covariance matrix in (b) is of the form of one of the popular models in Section 4.4 of the notes by finding the associated correlation matrix and identifying to which popular model it corresponds.

We will see that this covariance matrix plays a special role.

5. In the file \texttt{insulin.dat} on the class web page you will find longitudinal data from an experiment conducted by a diabetes researcher who was interested in comparing patterns of blood sugar reduction brought about by the use of several different insulin mixtures. The study involved \( m = 36 \) rabbits, where 12 rabbits were randomly assigned to each of 3 groups: group 1 rabbits received the “standard” insulin mixture, group 2 rabbits received a mixture containing 1% less protamine than the standard, and group 3 rabbits received a mixture containing 5% less protamine. Rabbits were injected with the assigned mixture at time 0, and blood sugar measurements were taken on each rabbit at the time of injection (time 0) and 0.5, 1.0, 1.5, 2.5, and 3.0 hours post-injection.

Each data record in the file \texttt{insulin.dat} represents a single observation; the columns of the data set are (1) rabbit number, (2) hours (time), (3) response (blood sugar level), and (4) insulin group (1, 2, or 3).

(a) Write a SAS program to obtain the following:

(i) Read in the data in the form they appear in the data set and, if necessary, transform the data set into a form suitable for carrying out the analyses in (ii)–(v) below. (Hint: look at the program at the end of Ch. 4.)

(ii) Find the means for each insulin group at each day and plot them for each group on the same graph

(iii) Find the sample covariance and sample correlation matrix for each insulin group

(iv) Find the pooled sample covariance matrix and corresponding estimated correlation matrix under the assumption of a common covariance matrix. Hint:

\[
\begin{align*}
\text{proc discrim pccov pcorr data=insulin2;}
&\quad \text{class group;}
&\quad \text{var hour1 hour2 hour3 hour4 hour5 hour6 hour7;}
&\quad \text{run;}
\end{align*}
\]
(b) (optional) Use R, SAS proc insight or some other software to obtain a matrix of pairwise scatterplots for each insulin group. (In R use the pairs function.)

(c) Based on inspection of the results in (a) (and (b) if you did it), do you think the assumption of a common covariance matrix for each group is reasonable? Why or why not? Give details of your reasoning.

(d) Based on inspection of the results in (a) (and (b) if you did it), what covariance model(s) do you think is(are) appropriate for these data? Give details of your reasoning.