PART I

Reading assignment Rencher, Chapters 1-3. We will not cover all of the matrix algebra in Chapter 2 in the first few lectures of this course. You can review Chapter 2, however, to improve you knowledge of matrix algebra. In the lectures, we will move on to a discussion of linear models (Chapters 7 and 12) and return to results in Chapter 2 as we need them.

The Excel file containing the bone mineral density (bmd) data is available under the name **bone2.density.xls**. If you are a SAS user, there is also a file under the name **bone2.density.sas** that contains SAS code for importing an Excel file. You will have to change the name of the file in the DATAFILE= option of PROC IMPORT to the name of the Excel file on your PC. For those who wish to use R or SAS, the data have also been stored as a text file under the name **bone2.density.txt**. This file also has ten columns but the first row does not contain the variable names. The variable names you should use are **id, group, bmd, weight, age, caffeine, calcium, growmilk, teenmilk, twenmilk** respectively.
A description of each variable is given below in the order in which the columns appear on the data file:

id An identification number for each women in the study.

group Indicates the exercise group and is coded as follows:
1=sedentary: no regular exercise program,
2=walker: at least 135 minutes per week,
3=aerobic dancer: at least 135 minutes per week.

bmd Bone mineral density (g/cm2)

weight Weight (kg)

age Age (years)

caffeine Weekly caffeine consumption

calcium Weekly calcium consumption

growmilk Weekly milk consumption as a child

teenmilk Weekly milk consumption as a teenager

twenmilk Weekly milk consumption between 20 and 30 years of age.

Recent studies suggest that peak bone mass in women is achieved at some time between the ages of 15 and 40. After about age 45, bone loss is gradual in both males and females, but a sharp decrease in ovarian estrogen production has been identified as the cause of more rapid bone loss in women during the first ten years after menopause. Since osteoporosis is essentially irreversible, prevention is more practical than treatment. A combination of estrogen progestogen (hormone replacement therapy, or HRT) is the current treatment of choice for postmenopausal women, but HRT may be accompanied by side effects, including increased risk of cardiovascular disease, liver disease, and certain types of cancer. A second approach is to increase intake of calcium, other nutrients, and dietary fiber in developmental years in an attempt to increase peak bmd by age 45. A third approach hypothesized to increase peak bmd is the use of a reasonable amount of physical exercise. One objective of this study was to determine if physically active, non-athletic, premenopausal women engaged in regular moderate exercise achieve higher average bone mineral density than non-exercisers. A second objective was to examine effects of other factors (weight, age, caffeine intake, calcium intake, and milk consumption at various developmental levels) that might be associated with differences in bone mineral density.

Use the data to answer the following questions.

1. Plot the bmd measurements against the weights of the women. Submit the graph. What does this graph suggest about the association between bmd and weight?

2. Compute the least squares estimates of the coefficients in the model

$$\text{bmd} = \beta_0 + \beta_1 \text{weight} + \text{error}.$$

   Report the values of the estimates and their standard errors. State your conclusion about the correlation between bone mineral density and weight.

3. Consider the entire set of variables in the data file.

   (a) Report a formula for the model that you think best describes variation in bone density measurements among the women in this data set. Report estimates for all of the parameters in your model and the corresponding sums of squares.

   (b) Report the R^2 value and the AIC value for the model you presented in part (a).

   (c) Briefly summarize your conclusions about associations among the explanatory variables and associations between the non-exercise explanatory variables and bmd.

   (d) Carefully state your conclusions about the potential benefits of a regular program of moderate exercise (dancing or walking) on increasing bmd in premenopausal women

PART II

The first three problems on this part of assignment ask you to provide mathematical proofs of some matrix properties. Problems 4 and 6 ask you to use R to numerically evaluate some quantities. Problem 5 asks you to write an R function to convert a covariance matrix into a correlation matrix. In problem 7, you are asked to use formulas for the inverse of a matrix and numerical evaluation of inverses of larger matrices to find a formula for the inverse of a special type of covariance matrix.

1. Use the definition of orthogonal and idempotent matrices and properties of determinants in the lecture notes to prove the following results:

   (a) If A is an orthogonal matrix, then $|A|$ is either 1 or -1. (Hint: use the definition of an orthogonal matrix and consider the determinant of an identity matrix.)

   (b) If W is an idempotent matrix, then $|W|$ is either 0 or 1.

2. Let A be a $p \times p$ positive definite matrix and let B be a $k \times p$ matrix of rank $k \leq p$. Show that BAB^T is positive definite. (This is Corollary 1 on page 26 of Rencher's book.)

3. Show that a symmetric matrix A is positive definite if and only if there exists a nonsingular matrix P such that $A = P^T P$. (This is Theorem 2.6C on page 26 of Rencher's book.)

4. Consider the matrix

$$V = \begin{bmatrix} 5.0 & 4.0 & 3.2 \\ 4.0 & 5.0 & 4.0 \\ 3.2 & 4.0 & 5.0 \end{bmatrix}$$

   Use R to find values for

   (a) the eigenvalues and eigenvectors of V

   (b) trace(V)

   (c) det(V)

   (d) the inverse of$V$

5. Suppose $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_n)^T$ is a random vector with mean vector $E(\mathbf{Y}) = \mu$ and covariance matrix $Var(\mathbf{Y}) = \Sigma$.

   (a) Describe a matrix B such that $B\Sigma B$ is the correlation matrix for $\mathbf{Y}$

   (b) Write an R function to compute the correlation matrix corresponding to any covariance matrix. Submit a listing of your function.

(c) Use the R function you created in part (b) of this problem to compute the correlation matrix when $Var(\mathbf{Y}) = V$, the matrix from problem 4.

6. Consider the matrices

$$A = \begin{bmatrix} 4 & 4.001 \\ 4.001 & 4.002 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 4 & 4.001 \\ 4.001 & 4.002001 \end{bmatrix}$$

These matrices are nearly identical. There is only a small difference in the (2,2) elements of these matrices. Moreover, the columns of A (and the columns of B) are nearly linearly dependent. Use R to compute the following quantities:

(a) The determinant of A and the determinant of B.

(b) A^{-1} and B^{-1}. (Note that A^{-1} is approximately $-3B^{-1}$, even though A is nearly identical to B. This shows that small changes in elements of matrices that are nearly singular, can have big effects on some matrix operations. Consequently, failure to retain enough significant digits can result in substantial computational errors.)

7. Software such as R or the IML procedure in SAS can be useful in the development of statistical theory and mathematical results. You should take advantage of it in your theory courses and your research. If you are asked to prove a result and you are not sure if it is true, for example, use R to evaluate a few numerical examples. You only need to find one example where the result does not hold to show it is false. If the result holds true for every numerical example that you try, you may begin to believe that it is indeed true and try to develop a proof. Computational tools can also be useful in discovering patterns.

Here we will use R to help derive a formula for the inverse of an $n \times n$ matrix of the following form:

$$V = \sigma^2 \begin{bmatrix} 1 & b & b^2 & b^3 & \cdots & b^{n-1} \\ b & 1 & b & b^2 & \cdots & b^{n-2} \\ b^2 & b & 1 & b & \cdots & b^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b^{n-1} & b^{n-2} & b^{n-3} & b^{n-4} & \cdots & 1 \end{bmatrix}$$

Covariance matrices corresponding to a first order auto-correlation structure have this form. In that case, b would be a first order auto-correlation. You are first asked to drive a formula for the 2×2 and 3×3 cases. Next you are asked to use R to evaluate the inverse for several special cases and then infer the general result. You may have to evaluate some additional examples to recognize the pattern in the results. Once a formula for V^{-1} is identified, you can easily prove that it is the inverse of V by directly checking that $V^{-1}V = VV^{-1} = I$. Begin your investigation by finding a formula for the inverse of the correlation matrix R associated with V.

(a) First use Result 1.4 (i) in the lecture notes to obtain the formula for the inverse of

$$R = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}$$

(b) Now use Result 1.4 (ii) in the lecture notes to obtain the formula for the inverse of

$$R = \begin{bmatrix} 1 & b & b^2 \\ b & 1 & b \\ b^2 & b & 1 \end{bmatrix}$$

(c) Now use R to evaluate the inverse of

$$R = \begin{bmatrix} 1 & b & b^2 & b^3 \\ b & 1 & b & b^2 \\ b^2 & b & 1 & b \\ b^3 & b^2 & b & 1 \end{bmatrix}$$

for $b = 0.1$ and $b = 0.5$. Evaluate the inverse of the 5×5 version of R for $b = 0.1$ and $b = 0.5$. You may need to evaluate some additional cases with other values of b to see the pattern. Do not report numerical results for the examples, but use the numerical results and the results from parts (a) and (b) to guess a formula for the inverse of R. Use R to numerically verify that your guess is correct. When you are satisfied with your solution to the inverse for R, use it to obtain a formula for the inverse of V.