1. Suppose $\mathbf{X} = [X_1, X_2, X_3]'$ is multivariate normal with mean vector and covariance matrix

$$\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 7 & 6 & 3 \\ 6 & 9 & 4 \\ 3 & 4 & 6 \end{bmatrix}$$

(a) Let $Y_1 = X_1 + X_3$ and $Y_2 = X_3 - X_1$. What is the distribution of $\mathbf{Y} = [Y_1, Y_2]'$? Show your work.

First, note that the $X_1 + X_3$ and $X_3 - X_1$ can be defined jointly as $\mathbf{Y} = \mathbf{AX}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

so that $\mathbf{Y}$ is normal with mean

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

and variance

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & 6 & 3 \\ 6 & 9 & 4 \\ 3 & 4 & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 10 & 9 \\ -4 & -2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 19 & -1 \\ -1 & 7 \end{bmatrix}$$

(b) What is the distribution of $Y_1|Y_2 = 4$ (that is, the distribution of $X_1 + X_3$ given that $X_3 - X_1 = 4$)? Show your work.

From our work above in the first part (and our results from Lesson 6.1), $X_1 + X_3|X_3 - X_1 = 4$ is normal with mean

$$7 + (-1)\left(\frac{1}{7}\right)(4 - 3) = \frac{48}{7} = 6.857$$

and variance

$$19 - (-1)\left(\frac{1}{7}\right)(-1) = \frac{132}{7} = 18.857$$

2. From a sample of 30 multivariate data values, a correlation coefficient between X_1 and X_2 is found to be 0.234.

   (a) Compute a 95% confidence interval for the population correlation. Show your work.

   Using Fisher's transformation with $z = .5\log[(1+.234)/(1-.234)] = .2384$ and multiplier 1.96, the 95% confidence interval for the transformed correlation is

$$.2384 \pm 1.96\frac{1}{\sqrt{30-3}} = (-.1388, .6156)$$

   Applying the reverse transformation gives

$$\left(\frac{\exp\{2\times -.1388\}-1}{\exp\{2\times -.1388\}+1}, \frac{\exp\{2\times .6156\}-1}{\exp\{2\times .6156\}+1}\right) = (-.1379, .5481)$$

   (b) Is this significant evidence that these two variables are (linearly) related? Conduct an appropriate hypothesis test with significance level $\alpha = .05$. Show your work.

   A correlation of 0 corresponds to no linear relationship, and the interval above includes 0 as a possible value. So, we cannot say that the variables are linearly related.

   (c) State any assumptions you are making about these data.

   We are assuming that the 30 (X_1, X_2) pairs are independently sampled from a bivariate normal distribution.

3. At the start of a study to determine whether exercise or dietary supplements would slow bone loss in older women, an investigator measured the mineral content of bones. Measurements were recorded on 25 women for three bones on each of the dominant and non-dominant sides. The data was named "mineral" and has the following variables: dominant radius, radius, dominant humerus, humerus, dominant ulna, and ulna.

   SAS output is included on the following pages. Use it to answer the questions below.

   (a) What is the 90% confidence interval for the population correlation between dominant radius and dominant humerus? Report this from the output, and provide an interpretation of this interval. That is, what do we mean by "90% confidence" here?

   From the first page of the output, this interval is $(.462, .834)$. We're 90% confident that for all women in this age group, the correlation between dominant radius and dominant humerus mineral contents is between .462 and .834.

   That is, if similar samples of this size were repeatedly collected, about 90% of the resulting intervals would include this correlation.

   (b) What is the 90% confidence interval for the population partial correlation between dominant radius and dominant humerus conditioned on radius and humerus? How is the interpretation of this partial correlation different from that above in part a)? Also, how do you explain the difference in the numeric values between the two intervals?

   This interval is given as $(.329, .792)$. The difference is that this (partial) correlation measures the linear relationship between dominant radius and dominant humerus mineral values if both radius and humerus mineral values are held fixed. Since the conditional correlation is smaller, it suggests that the nondominant bone variables are related to their dominant counterparts.

   (c) Consider both intervals from parts a) and b). How confident are you that both intervals simultaneously cover their population parameters? *Hint: use the Bonferroni adjustment*

   To be $1-\alpha$ confident of both intervals simultaneously, the Bonferroni adjustment to use $\alpha/2$ for each individual interval. In this case,

$$.90 = 1 - \frac{\alpha}{2} \Rightarrow \alpha = .2$$

   So, we are 80% confident that both intervals cover their parameters simultaneously.

```
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum domulna;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum;
partial rad hum;
run;
proc corr data=mineral fisher(alpha=.10 biasadj=no);
var domrad domhum domulna;
partial rad hum ulna;
run;
```