1. (a) (4 points) Define \( \mathbf{c} = (3 \ 0 \ -2)' \). Then,

\[
\mathbf{Y} = (\mathbf{X} - 2\mathbf{X}) = \mathbf{c}' \mathbf{X} \sim N(\mathbf{c}' \mathbf{\mu}, \mathbf{c}' \Sigma \mathbf{c}), \quad \text{where} \quad \mathbf{c}' \mathbf{\mu} = 1 \quad \text{and} \quad \mathbf{c}' \Sigma \mathbf{c} = 85.
\]

(b) (6 points) First note that

\[
\mathbf{Y} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \begin{bmatrix} 3 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \sim N\left( \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \begin{bmatrix} 85 & 5 & 2 \\ 5 & 10 & 2 \\ -11 & 2 & 10 \end{bmatrix} \right)
\]

Then, the condition distribution of \( \mathbf{Y} \), given \( \mathbf{X}_2 = 2 \) and \( \mathbf{X}_3 = 5 \), is a normal distribution with mean

\[
1 + \begin{bmatrix} 5 & -11 \end{bmatrix} \begin{bmatrix} 10 & 2 \\ 2 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 2-1 \\ 5-4 \end{bmatrix} = 0.5
\]

and variance

\[
85 - \begin{bmatrix} 5 & -11 \end{bmatrix} \begin{bmatrix} 10 & 2 \\ 2 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 5 \end{bmatrix} = 67.5
\]

(c) (4 points) Yes, it satisfies the definition of an eigenvector because

\[
\begin{bmatrix} 3 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{8}{\sqrt{2}} \\ -\frac{8}{\sqrt{2}} \end{bmatrix} = 8 \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}
\]

(d) (4 points) Compute \( |\Sigma| = \lambda_1 \lambda_2 \lambda_3 = (15)(8)(6) = 720 \)

(e) (4 points) Yes, you could argue that rotations do not affect eigenvalues, or you could remember that the covariance matrix for \( \mathbf{Y} = \mathbf{A} \mathbf{X} \) is \( \mathbf{A} \Sigma \mathbf{A}' \) and \( \mathbf{A}' \mathbf{A} = \mathbf{I} \) when \( \mathbf{A} \) is an orthogonal matrix. Then using a property of determinants (see page 83 in the text), we have

\[
|\mathbf{A} \Sigma \mathbf{A}'| = |\mathbf{A}| |\Sigma| |\mathbf{A}'| = |\mathbf{A}'| |\mathbf{A}| |\Sigma| = |\mathbf{A}'| |\Sigma| = |\Sigma|
\]

(f) (4 points) For a tri-variate normal distribution, the volume of the smallest region that has probability \( 1 - \mathbf{\alpha} \) of containing a randomly selected observation is

\[
\frac{2\pi^{3/2}}{3 \Gamma(3/2)} |\mathbf{Z}_{(3),\mathbf{\alpha}}^{3/2} | \bigg| |\Sigma|^{1/2} \bigg|^2.
\]
2. (a) (4 points) \( \bar{X} \sim N(\mu, \frac{1}{n} \Sigma) \)

(b) (6 points) Since \( n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) = T^2 \sim \frac{p(p-1)}{n-p} F_{(p,p-n)} \), then

\[
(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) \sim \frac{4(n-1)}{n(n-4)} F_{(4,n-4)}.
\]

(c) (4 points) The smallest exact 95% region is defined by all vectors \( \mathbf{v} \) where

\[
n(\bar{X} - \mathbf{v})'S^{-1}(\bar{X} - \mathbf{v}) \leq \frac{4(n-1)}{n-4} F_{(4,n-4),0.05}
\]
and the hyper-volume of this region is

\[
\frac{\pi^2}{2} \left[ \frac{4(n-1)}{n-4} F_{(4,n-4),0.05} \right]^2 \left[ \frac{1}{n} S \right]^2.
\]

(d) (6 points) From the conservative \( T^2 \) method, simultaneous 95% confidence intervals are

\[
a' \bar{X} \pm \sqrt{\frac{4(n-1)}{n-4} F_{(4,n-4),0.05}} \sqrt{\frac{a'S^{-1}a}{n}}
\]
and

\[
b' \bar{X} \pm \sqrt{\frac{4(n-1)}{n-4} F_{(4,n-4),0.05}} \sqrt{\frac{b'S^{-1}b}{n}}
\]

(e) (6 points) 6 d.f.

3. (a) (8 points) A set of measurements taken on proteins A, B, and C before the chemotherapy is begun and a second set of measurements taken one week after treatment are taken on the same patient and should be considered as a set of six measurements taken on one patient. Use a one sample \( T^2 \) test is this problem.

Let \( X_j = \begin{bmatrix} X_{A1j} \\ X_{B1j} \\ X_{C1j} \\ X_{A2j} \\ X_{B2j} \\ X_{C2j} \end{bmatrix} \) denote the set of six measurements taken on the j-th patient.
Then, the vector of sample means is \( \bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j \), the sample covariance matrix is

\[
S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T
\]

and \( T^2 = n(C\bar{X} - 0)^T(CSC)^{-1}(C\bar{X} - 0) \),

where \( C = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \).

You could have obtained the same result by applying a one sample \( T^2 \) test to a sample of 3-dimensional vectors of differences.

(b) (4 points)

\[
F = \frac{(n - 3)}{(3)(n - 1)} T^2 = \frac{23}{(3)(25)} T^2 \sim F_{(3,23)}.\quad \text{Reject } H_0: C\mu = 0 \text{ if } \frac{23}{(3)(25)} T^2 > F_{(3,23), \alpha}.
\]

(c) (4 points) \( X_j \sim \text{NID}(\mu, \Sigma) \) for \( j = 1,2,\ldots,26 \), Note that this includes the assumption that each subject responds independently of any other subject.

4. (a) (5 points for each part) It is crucial to recognize that there are 8 measurements taken on each pig, 4 measurements on two increases in body fat and 4 measurement on two week increases in muscle tissue. There are three treatment groups, corresponding to the amount of riboflavin added to the diet, with 20 pigs in each group. PROC GLM in SAS would construct a model with the following parameter matrix:

\[
\beta = \begin{bmatrix}
\mu_{1f} & \mu_{2f} & \mu_{3f} & \mu_{4f} & \mu_{1m} & \mu_{2m} & \mu_{3m} & \mu_{4m} \\
\alpha_{1f1} & \alpha_{2f1} & \alpha_{3f1} & \alpha_{4f1} & \alpha_{1m1} & \alpha_{2m1} & \alpha_{3m1} & \alpha_{4m1} \\
\alpha_{1f2} & \alpha_{2f2} & \alpha_{3f2} & \alpha_{4f2} & \alpha_{1m2} & \alpha_{2m2} & \alpha_{3m2} & \alpha_{4m2}
\end{bmatrix}
\]

where the first four columns correspond to the bi-weekly gains in fat and the last four columns correspond to the biweekly gains in muscle tissue. The first row gives mean biweekly gains when riboflavin is added to the diet at a rate of 10 g/kg. The second and third row are deviations from these means for diets with 0 g/kg and 5g/kg of added riboflavin, respectively.
\[ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \]

\[ \begin{bmatrix} 1 \end{bmatrix} \]

\[ \begin{bmatrix} 1 \end{bmatrix} \]

\[ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

\[ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \]

\[ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \]

\[ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \]

(b) You get credit in this part if you made the correct calculations for your answer in part (a), regardless of whether or not your answers for part (a) were correct.

(i) (2 points) \( n=60 \ k=2 \ r=3 \ p=8 \ u=1 \) and \( a=57 \ b=1 \ c=0 \) and df are \( (2, 57) \)

(ii) (2 points) \( n=60 \ k=2 \ r=3 \ p=8 \ u=4 \) and \( a=55.5 \ b=2 \ c=3 \) and df are \( (8, 108) \)

C. (6 points) There are several answers depending on how complex you wish to make the covariance structure for the mixed model.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels of riboflavin (R)</td>
<td>2</td>
</tr>
<tr>
<td>pigs within levels of riboflavin</td>
<td>57</td>
</tr>
<tr>
<td>types of gain: fat vs. muscle (G)</td>
<td>1</td>
</tr>
<tr>
<td>time (T)</td>
<td>3</td>
</tr>
<tr>
<td>G x T interaction</td>
<td>3</td>
</tr>
<tr>
<td>R x G interaction</td>
<td>2</td>
</tr>
<tr>
<td>R x T interaction</td>
<td>6</td>
</tr>
<tr>
<td>R x G x T interaction</td>
<td>6</td>
</tr>
<tr>
<td>error</td>
<td>399</td>
</tr>
<tr>
<td>Corrected total</td>
<td>479</td>
</tr>
</tbody>
</table>
(d) (6 points) The mixed mode for the ANOVA table in part (c) is

\[ X_{ijk\ell} = \mu + R_i + \delta_{ij} + G_k + \tau_{i\ell} + G\tau_{k\ell} + RG_{ik} + RG\tau_{ik\ell} + \epsilon_{ijk\ell} \]

where
- \( i = 1,2,3 \) denotes the level of added riboflavin
- \( j = 1,2,...,20 \) denotes pigs within riboflavin levels
- \( k = 1,2 \) denotes the type of gain (\( k = 1 \) for fat, \( k = 2 \) for muscle)
- \( \ell = 1,2,3,4 \) denotes the two week time period

\[ \delta_{ij} \sim \text{NID}(0,\sigma_\delta^2) \]
\[ \epsilon_{ijk\ell} \sim \text{NID}(0,\sigma_e^2) \]

\[ \delta_{ij} \text{ is independent of any } \epsilon_{ijk\ell} \]

This model implies that all 8 measurements taken on a single pig have the same variances. The MANOVA approach allows these variances to be different. The mixed model also implies that the correlation between any two measurements taken on the same pig does not depend on the particular pair of measurements. The MANOVA model does not restrict these correlations in any way. The assumptions made by this mixed model may not be appropriate. Since muscle tissue gains tend to be larger than fat gains, for example, the variances for muscle tissue gains may be larger than the variances for fat gains. Also, the correlation between fat gains for different time periods may differ from the correlation between muscle and fat gains during the same time period.

The following mixed model includes more random terms to allow for a more complex covariance matrix. This model implies that all variances are equal, but it allows for different correlations for repeated measures of fat gains, repeated measures of muscle tissue gains, and correlations between fat and muscle tissue gains.

\[ X_{ijk\ell} = \mu + R_i + \delta_{ij} + G_k + RG_{ik} + \gamma_{ijk} + \tau_{i\ell} + G\tau_{k\ell} + \theta_{ij\ell} + RG\tau_{ik\ell} + \epsilon_{ijk\ell} \]

where
- \( i = 1,2,3 \) denotes the level of added riboflavin
- \( j = 1,2,...,20 \) denotes pigs within riboflavin levels
- \( k = 1,2 \) denotes the type of gain (\( k = 1 \) for fat, \( k = 2 \) for muscle)
- \( \ell = 1,2,3,4 \) denotes the two week time period

\[ \delta_{ij} \sim \text{NID}(0,\sigma_\delta^2) \]
\[ \gamma_{ijk} \sim \text{NID}(0,\sigma_\gamma^2) \]
\[ \theta_{ij\ell} \sim \text{NID}(0,\sigma_\theta^2) \]
\[ \epsilon_{ijk\ell} \sim \text{NID}(0,\sigma_e^2) \]

and \( \delta_{ij}, \gamma_{ijk}, \theta_{ij\ell}, \epsilon_{ijk\ell} \) are all independent of each other.
The ANOVA table for this is

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Levels of riboflavin (R)</td>
<td>2</td>
</tr>
<tr>
<td>error a (pigs within riboflavin levels)</td>
<td></td>
</tr>
<tr>
<td>types of gain: fat vs. muscle (G)</td>
<td>1</td>
</tr>
<tr>
<td>R x G interaction</td>
<td>2</td>
</tr>
<tr>
<td>error b</td>
<td>57</td>
</tr>
<tr>
<td>time (T)</td>
<td>3</td>
</tr>
<tr>
<td>RxT interaction</td>
<td>6</td>
</tr>
<tr>
<td>error c</td>
<td>171</td>
</tr>
<tr>
<td>G x T interaction</td>
<td>3</td>
</tr>
<tr>
<td>R x G x T interaction</td>
<td>6</td>
</tr>
<tr>
<td>error d</td>
<td>171</td>
</tr>
<tr>
<td>Corrected total</td>
<td>479</td>
</tr>
</tbody>
</table>

5. (6 points) A bootstrap confidence interval for the ratio of correlations can be constructed in the following way:

- Use sampling with replacement to select a new sample of 596 students from the original sample.
- Compute the sample correlation between the IQ score and the score for the mathematics exam.
- Compute the sample correlation between the IQ score and the score of the verbal ability exam.
- Compute the ratio of the correlations.
- Repeat the four steps listed above some large number, say B, times.
- Order the B bootstrap estimates of the correlation ratios from small to largest. Delete the smallest 2.5% and the largest 2.5% of the bootstrap estimates.
- The remaining interval of values is a 95% bootstrap confidence interval

**Exam Scores** (100 possible points):

<table>
<thead>
<tr>
<th>Score</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>133468</td>
</tr>
<tr>
<td>7</td>
<td>02346679</td>
</tr>
<tr>
<td>6</td>
<td>46789</td>
</tr>
<tr>
<td>5</td>
<td>2279</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>