1. Products used to kill weeds generally consist of a small amount of an active ingredient and a large amount of inactive ingredients. The effects of inactive ingredients on the performance of a weed killer were examined in a study where

\[ Z_1 \] denotes the proportion of inactive ingredient A

\[ Z_2 \] denotes the proportion of inactive ingredient B

\[ Z_3 \] denotes the proportion of inactive ingredient C

in the mixture of inactive ingredients. Hence, \( Z_1 + Z_2 + Z_3 = 1 \) all cases. The same amount of active ingredient was added to each of the six mixtures of inactive ingredients considered in this study. Each of these six mixtures was applied to two test plots. After two weeks, each test plot was given a score (\( Y \)) reflecting the level of weed control in the plot. A higher score corresponds to a higher level of weed control (i.e., fewer weeds in the plot).

The data are shown below:

<table>
<thead>
<tr>
<th></th>
<th>( Y )</th>
<th>( Z_1 )</th>
<th>( Z_2 )</th>
<th>( Z_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( Y_1 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>( Y_2 )</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>( Y_3 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>( Y_4 )</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>( Y_5 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>( Y_6 )</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>( Y_7 )</td>
<td>.5</td>
<td>.5</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>( Y_8 )</td>
<td>.5</td>
<td>.5</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>( Y_9 )</td>
<td>.5</td>
<td>0</td>
<td>.5</td>
</tr>
<tr>
<td>10</td>
<td>( Y_{10} )</td>
<td>.5</td>
<td>0</td>
<td>.5</td>
</tr>
<tr>
<td>11</td>
<td>( Y_{11} )</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>12</td>
<td>( Y_{12} )</td>
<td>0</td>
<td>.5</td>
<td>5</td>
</tr>
</tbody>
</table>

Consider the following models:

Model 1: \[ Y_i = \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \epsilon_i \]

Model 2: \[ Y_i = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \epsilon_i \]

Model 3: \[ Y_i = \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \gamma_{12} Z_{1i} Z_{2i} + \gamma_{13} Z_{1i} Z_{3i} + \gamma_{23} Z_{2i} Z_{3i} + \epsilon_i \]
In matrix notation, we will write the \( i \)-th model as

\[
Y = X_i \beta_i + \epsilon.
\]

Thus, \( X_1 \) is a 12\times3 matrix
\( X_2 \) is a 12\times4 matrix
\( X_3 \) is a 12\times6 matrix

and \( \beta_1, \beta_2, \) and \( \beta_3 \) are 3\times1, 4\times1, 6\times1 vectors of regression coefficients, respectively.

(a) What conditions must \( Y \) satisfy to yield a Gauss-Markov model?

(b) Formulate what we mean by the statement that \( Y = X_1 \beta_1 + \epsilon \) is a reparameterization of \( Y = X_2 \beta_2 + \epsilon \). (For this part you should just give a definition. Do not try to determine if this statement is correct.)

(c) Is \( Y = X_1 \beta_1 + \epsilon \) a reparameterization of \( Y = X_2 \beta_2 + \epsilon \)? Explain.

(d) The parameter \( \alpha_i \) appears in \( \beta_1, \beta_2, \) and \( \beta_3 \). For each of these three models, determine if \( \alpha_i \) is estimable. Explain how you arrived at your conclusions.

(e) Suppose that \( C \beta_2 \) is estimable for model 2, \( Y = X_2 \beta_2 + \epsilon_2 \), and \( Y \) satisfies the Gauss-Markov condition from part (a). For which choice, or choices, of the matrix \( A \) is \( CAY \) a best linear unbiased estimator for \( C \beta_2 \)? Is the value of \( CAY \) unique?

(f) Show that \( H_0 : \alpha_1 = \alpha_2 = \alpha_3 \) is a testable hypothesis for model 2, \( Y = X_2 \beta_2 + \epsilon_2 \).

(g) Assuming that \( Y \sim N(X_2 \beta_2, \sigma^2 I) \), construct an F-test for the null hypothesis in part (f). In this part you should

(i) Give formulas for the numerator and denominator sums of squares used in the F statistics.

(ii) Show that those sums of squares are independent and have appropriate distributions.

(iii) Give the degrees of freedom and non-centrality parameter for the F-test.

You can write your answers in matrix form. Define any additional notation that you decide to use.
2. In class I stated without proof that \( \text{rank}(X) = \text{rank}(P_X) \), where \( X \) is any \( n \times k \) matrix and 
\[
P_X = X(X^TX)^{-1}X^T.
\] Suppose you wanted to check this numerically for \( X_2 \), the model matrix for model 2 in Problem 1. Assuming that the data are stored in a file called herb.dat, with four numbers per line as shown on page 1, give SPLUS code for

(a) entering the data from Problem 1 into a matrix,

(b) constructing \( X_2 \) and computing \( P_{X_2} \), and

(c) computing \( \text{rank}(X_2) \) and \( \text{rank}(P_{X_2}) \).

If you have forgotten the exact form of some of the functions, make your best guess. Anything close will be awarded some points.

3. Consider a 2-way crossed classification with factor A (with 3 levels) and factor B (with 3 levels). Let \( n_{ij} \) denote the number of observations with factor A at level i and factor B at level j. Consider the model

\[
Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}
\]

where \( \epsilon_{ijk} \sim \text{NID}(0, \sigma^2) \), \( i=1,2,3, \ j=1,2,3, \) and \( k=1,…,n_{ij} \). All other terms in the model are to be considered as fixed effects.

We will use the \( R(\cdot | \cdot) \) notation for sums of squares. For example, \( R(\alpha | \mu, \beta) \) stands for the sums of squares for the “main effects” of factor A corrected for \( \mu \) and the “main effects” of factor B. Equivalently, \( R(\alpha | \mu, \beta) \) is the reduction in the residual sums of squares achieved by adding the main effects for factor A to a model that already contains \( \mu \) and the main effects for factor B.

(a) State a condition on the \( n_{ij} \)'s, as general as possible, for which \( R(\alpha | \mu, \beta) \) and \( R(\alpha | \mu) \) are equal.

(b) List the restrictions that the GLM procedure in SAS would put on the parameters in this model to obtain a solution to the normal equations.

(c) Under the parameter restrictions in part (b), how would you interpret \( \hat{\alpha}_2 \) in the solution to the normal equations.
4. The sample sizes for the experiment described in Problem 3 are shown in the following table.

<table>
<thead>
<tr>
<th>Level of factor B</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>n_{11} = 1</td>
<td>n_{12} = 1</td>
<td>n_{13} = 0</td>
</tr>
<tr>
<td>of factor A</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>n_{21} = 2</td>
<td>n_{22} = 2</td>
<td>n_{23} = 2</td>
</tr>
<tr>
<td>3</td>
<td>n_{31} = 1</td>
<td>n_{32} = 0</td>
<td>n_{33} = 1</td>
</tr>
</tbody>
</table>

Note that there are no observations in the (1,3) and (3,2) cells.

(a) What are the degrees of freedom, for $R(\gamma | \mu, \alpha, \beta)$? What is the null hypothesis for the corresponding F-test?

(b) For the rest of this problem, assume an additive model for the cell means, i.e.,

$$\mu_{ij} = E(Y_{ijk}) = \mu + \alpha_i + \beta_j$$

for all (i,j) where $n_{ij} > 0$. As in Problem 3, assume that $\varepsilon_{ijk} \sim \text{NID}(0, \sigma^2)$. Show that $\alpha_1 - \alpha_2$, $\alpha_1 - \alpha_3$, and $\alpha_2 - \alpha_3$ are estimable for this model.

(c) Define $Y = (Y_{111} \ Y_{121} \ Y_{211} \ Y_{212} \ Y_{221} \ Y_{222} \ Y_{231} \ Y_{232} \ Y_{311} \ Y_{331})^T$

and

$$A = \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & \frac{-1}{2} & \frac{-1}{2} \\
0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{-1}{2} & \frac{-1}{2}
\end{bmatrix}$$

Then,

$$\begin{bmatrix}
Y_{1..} - Y_{3..} \\
Y_{2..} - Y_{3..}
\end{bmatrix} = AY$$

and define

$$T = (\bar{Y}_{1..} - \bar{Y}_{3..})^2 + (\bar{Y}_{2..} - \bar{Y}_{3..})^2 = Y^T A^T A Y.$$ 

Can you find positive scalar constants $C_1$ and $C_2$ so that $C_1 T$ has a non-central chi-square distribution with $C_2$ degrees of freedom? Explain.

(d) Give a formula for an F-test that could be used to test $H_0$: $\alpha_1 = \alpha_2 = \alpha_3$ versus the alternative the $\alpha_i \neq \alpha_j$ for some $i \neq j$. Give the degrees of freedom for this F-test.