1. The model is

$$y_{ijk} \;=\; \mu + \alpha_i + \beta_j + \gamma_{ij}\,, \quad \epsilon_{ijk} \sim NID(0, \sigma^2).$$

(a) Least squares estimates and interpretations for the quantities below are based the base line restrictions specified by options(contrasts=c("contr.treatment", "contr.ploy")) in the following code:

$$\alpha_1 = \beta_1 = \gamma_{1j} = \gamma_{i1} = 0; \qquad 1 \le i \le 4, 1 \le j \le 3.$$

specified by options(contrasts=c("contr.treatment", "contr.ploy")) in the following code:

```
> dogs <- read.table("stat511/HW7/dogs.dat", col.names=c("Drug","Disease","Y"))
> dogs$Drug <- as.factor(dogs$Drug)
> dogs$Disease <- as.factor(dogs$Disease)
> options( contrasts=c("contr.treatment", "contr.ploy") )
>
> lm.out1 <- lm( Y ~ Drug*Disease, data=dogs )
> lm.out1$coef
 (Intercept)     Drug2 Drug3     Drug4  Disease2  Disease3 Drug2Disease2
   29.33333 -3.333333   -13 -15.73333 -1.083333 -8.933333      8.583333

 Drug3Disease2 Drug4Disease2 Drug2Disease3 Drug3Disease3 Drug4Disease3
       -10.85     0.3166667     0.9333333           1.1      9.533333
> dummy.coef(lm.out1)
$"(Intercept)":
 (Intercept)
   29.33333
$Drug:
 1         2   3         4
 0 -3.333333 -13 -15.73333
$Disease:
 1         2         3
 0 -1.083333 -8.933333
$"Drug:Disease":
 11 21 31 41 12       22     32        42 13        23  33       43
  0  0  0  0  0 8.583333 -10.85 0.3166667  0 0.9333333 1.1 9.533333
```

(b) Least squares estimates of parameters resulting from using the restrictions to solve the normal equations are:

$$\begin{aligned}
\hat{\mu} &= 29.333 \\
\hat{\alpha}_1 &= 0 \\
\hat{\beta}_3 &= -8.933 \\
\hat{\gamma}_{23} &= 0.933 \\
\hat{\alpha}_2 - \hat{\alpha}_3 &= -3.333333 - (-13) \\
&= 9.667 \\
\hat{\gamma}_{22} - \hat{\gamma}_{23} - \hat{\gamma}_{32} + \hat{\gamma}_{33} &= 8.583333 - 0.9333333 - (-10.85) + 1.1 = 19.6 \\
\hat{\mu} + \hat{\alpha}_2 + \hat{\beta}_3 + \hat{\gamma}_{23} &= 29.33333 + (-3.333333) + (-8.933333) + 0.93333 = 18.0
\end{aligned}$$

$$\hat{\alpha}_2 - \hat{\alpha}_3 + \frac{1}{3}(\hat{\gamma}_{21} + \hat{\gamma}_{22} + \hat{\gamma}_{23}$$
$$-\hat{\gamma}_{31} - \hat{\gamma}_{32} - \hat{\gamma}_{33}) = 9.667 + \frac{1}{3}(0 + 8.583 + 0.933 - 0 + 10.85 - 1.1)$$
$$= 16.089$$

Note that by the restriction the cell means are represented in terms of the parameters as:

	Disease 1	Disease 2	Disease 3
Drug 1	$\mu_{11} = \mu$	$\mu_{12} = \mu + \beta_2$	$\mu_{13} = \mu + \beta_3$
Drug 2	$\mu_{21} = \mu + \alpha_2$	$\mu_{22} = \mu + \alpha_2 + \beta_2 + \gamma_{22}$	$\mu_{23} = \mu + \alpha_2 + \beta_3 + \gamma_{23}$
Drug 3	$\mu_{31} = \mu + \alpha_3$	$\mu_{32} = \mu + \alpha_3 + \beta_2 + \gamma_{32}$	$\mu_{33} = \mu + \alpha_3 + \beta_3 + \gamma_{33}$
Drug 4	$\mu_{41} = \mu + \alpha_4$	$\mu_{42} = \mu + \alpha_4 + \beta_2 + \gamma_{42}$	$\mu_{43} = \mu + \alpha_4 + \beta_3 + \gamma_{43}$

Then, an interpretation of each quantity for the restricted model is given as follows.

μ is the mean increase of systolic blood pressure with Disease 1 and Drug 1, i.e., $\mu = E(Y_{11k})$.
Here, the least squares estimate is $\hat{\mu} = \bar{y}_{11}$.

α_1 is restricted to be zero.

β_3 is the deviation of the mean increase in systolic blood pressure for Disease 3 treated with Drug 1 from that for Disease 3 treated with Drug 1., i.e., $\beta_3 = E(Y_{13k}) - E(Y_{11k})$.
The least squares estimate is $\hat{\beta}_3 = \bar{y}_{13\cdot} - \bar{y}_{11\cdot}$.

γ_{23} represents an interaction contrast. It is "the difference in the mean increases in systolic blood pressure for Disease 1 and Disease 3 when Drug 1 is given" minus "the difference in mean increases in systolic blood pressure difference Disease 1 and Disease 3 when Drug 3 is given," i.e., $\gamma_{23} = \gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23} = E(Y_{11k}) - E(Y_{13k}) - E(Y_{21k}) + E(Y_{23k})$
The least squares estimate is $\hat{\gamma}_{23} = \bar{y}_{11\cdot} - \bar{y}_{13\cdot} - \bar{y}_{21\cdot} + \bar{y}_{23\cdot}$.

$\alpha_2 - \alpha_3$ is the difference in mean increases of systolic blood pressure between treating Disease 1 with Drug 2 or Drug 3, i.e., $\alpha_2 - \alpha_3 = E(Y_{21k}) - E(Y_{31k})$.
The least squares estimate is $\hat{\alpha_2} - \hat{\alpha_3} = \bar{y}_{21\cdot} - \bar{y}_{31\cdot}$..

$\gamma_{22} - \gamma_{23}$ $-\gamma_{32} + \gamma_{33}$
is an interaction contrast. It is "the difference in mean blood pressure increases when Drug 2 is used to treat Disease 2 or Disease 3 " minus "the difference in mean blood pressure increases when Drug 3 is used to treat Disease 2 or Disease 3 " i.e., $\gamma_{22} - \gamma_{23}] - \gamma_{32} + \gamma_{33} = E(Y_{22k}) - E(Y_{23k}) - E(Y_{32k}) + E(Y_{33k})$
The least squares estimate is $\hat{\gamma}_{22} - \hat{\gamma}_{23} - \hat{\gamma}_{32} + \hat{\gamma}_{33} = \bar{y}_{22\cdot} - \bar{y}_{23\cdot} - \bar{y}_{32\cdot} + \bar{y}_{33\cdot}$.

$\mu + \alpha_2$ $+\beta_3 + \gamma_{23} = \mu_{23}$
is the mean increase in blood pressure when Drug 2 is used with Disease 3, and the OLS estimator is $\bar{y}_{23\cdot}$.

$\alpha_2 - \alpha_3$ $+\frac{1}{3}(\gamma_{21} + \gamma_{22} + \gamma_{23} - \gamma_{31} - \gamma_{32} - \gamma_{33}) = \frac{1}{3}\sum_{j=1}^{3} \mu_{2j} - \frac{1}{3}\sum_{j=1}^{3} \mu_{3j}$
is the difference between the mean increase in blood pressure when Disease 2 is treated and the mean increase in blood pressure when Disease 3 is used, averaging across the drugs giving equal weight to each drug.
The OLS estimator is $\frac{1}{3}\sum_{j=1}^{3} \bar{y}_{2j\cdot} - \frac{1}{3}\sum_{j=1}^{3} \bar{y}_{3j\cdot}$.

(c) Since μ, α_1, β_3, $\alpha_2 - \alpha_3$, and γ_{23} are not estimable for the unrestricted model, the interpretation of these parameters and the values of the corresponding OLS estimators depend on the particular restrictions placed on the model. On the other hand, $\gamma_{22} - \gamma_{23} - \gamma_{32} + \gamma_{33}$, $\mu + \alpha_2 + \beta_3 + \gamma_{23} = \mu_{23}$, and $\beta_2 - \beta_3 + \frac{1}{3}(\gamma_{12} + \gamma_{22} + \gamma_{32} - \gamma_{13} - \gamma_{23} - \gamma_{33})$ are estimable for the unrestricted model, and their interpretations and the values of the OLS estimators do not depend on the restrictions imposed on non-estimable functions of parameters (or equivalently, they do not depend on which generalized inverse is used to solve the normal equations).

You can directly show that the last three quantities are estimable by showing that they are expectations of linear combinations of observed bread volumes. For example,

$$\begin{aligned} E(\bar{y}_{23\cdot}) &= \mu_{23} = \mu + \alpha_2 + \beta_3 + \gamma_{23} \\ E(\bar{y}_{22\cdot} - \bar{y}_{23\cdot} - \bar{y}_{32\cdot} + \bar{y}_{33\cdot}) &= \mu_{22} - \mu_{23} - \mu_{32} + \mu_{33} = \gamma_{22} - \gamma_{23} - \gamma_{32} + \gamma_{33} \\ E\left(\frac{1}{3}\sum_j \bar{y}_{2j\cdot} - \frac{1}{3}\sum_j \bar{y}_{3j\cdot}\right) &= \frac{1}{3}\sum_j \mu_{2j} - \frac{1}{3}\sum_j \mu_{3j} \\ &= \left(\alpha_2 + \frac{1}{3}\sum_j \gamma_{2j}\right) - \left(\alpha_3 + \frac{1}{3}\sum_j \gamma_{3j}\right) \end{aligned}$$

This cannot be done for any of the first five quantities listed above in the unrestricted model and so they are not estimable. Using **Result 3.9**, this can be shown more formally by expressing each of those quantities in the form $c^T\underline{b}$ and noting that for each of the quantities a vector $\underline{d}$ can be found such that $X\underline{d} = 0$ for the unrestricted model, but $\underline{c^T d} \neq 0$.

$$\begin{aligned} \mu &\rightarrow \underline{d^T} = [\ 1\ -1\ -1\ -1\ -1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ] \\ \alpha_1 &\rightarrow \underline{d^T} = [\ 1\ -1\ -1\ -1\ -1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ] \\ \beta_3 &\rightarrow \underline{d^T} = [\ 1\ 0\ 0\ 0\ 0\ -1\ -1\ -1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ ] \\ \gamma_{23} &\rightarrow \underline{d^T} = [\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ -1\ ] \\ \alpha_2 - \alpha_3 &\rightarrow \underline{d^T} = [\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ -1\ -1\ -1\ 0\ 0\ 0\ 0\ 0\ 0\ ] \end{aligned}$$

(d) The ANOVA tables are computed as follows;

```
> # First compute the correction for the overall mean: R(mu)
> X       <- model.matrix(lm.out1)                          # model matrix
> ones    <- X[,1]                                          # vector of ones
> P1      <- ones%*%ginverse(t(ones)%*%ones)%*%t(ones)      # projection matrix
> y       <- dogs[,"Y"]                                     # responses
> R.mu    <- t(y) %*% P1 %*% y                              # R(mu)
> mse     <- deviance(lm.out1)                              # mse
> F.stat  <- R.mu / mse                                     # F statistics
> P.value <- 1 - pf(q=F.stat, df1=1, df2=lm.out1$df.resid) # P-value
>
> data.frame( SS=R.mu, df1=1, df2=lm.out1$df.resid, F=F.stat, Pval=P.value)
        SS df1 df2        F      Pval
1 20259.59   1  46 3.995987 0.0515372
>
> lm.out1 <- lm( Y ~ Drug*Disease, data=dogs )
> summary.aov( lm.out1, ssType=1 )
             Df Sum of Sq  Mean Sq  F Value     Pr(F)
        Drug  3  2992.830 997.6102 9.051325 0.0000805
     Disease  2   365.712 182.8558 1.659052 0.2015039
Drug:Disease  6   737.888 122.9814 1.115811 0.3680099
   Residuals 46  5069.983 110.2170

> lm.out2 <- lm(Y~Disease*Drug,data=dogs)
> summary.aov(lm.out2, ssType=1)
             Df Sum of Sq  Mean Sq  F Value     Pr(F)
     Disease  2   419.822 209.9108 1.904523 0.1604505
        Drug  3  2938.720 979.5735 8.887678 0.0000935
Disease:Drug  6   737.888 122.9814 1.115811 0.3680099
   Residuals 46  5069.983 110.2170
```

A useful inference is that the interaction between Drug and Disease is not significant. This does not prove that interaction does not exist, but it suggests that interaction effects may be small enough for the additive model to provide a good approximation. With respect to mean increases in systolic blood pressure, there appear to be substantial differences in drugs, but no large differences among diseases.

(e) We have seen in a previous homework assignment that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-rank(X)} = \chi^2_{46}$.
Furthermore, since we are assuming a normal theory Gauss-Markov model (i.e. independent errors and homogeneous variance), then $\overline{Y}_{11.} - \overline{Y}_{31.}$, $\overline{Y}_{12.} - \overline{Y}_{32.}$, and $\overline{Y}_{13.} - \overline{Y}_{33.}$ are mutually independent because each difference is obtained from a completely different set of observations. We also have

$$\overline{Y}_{1j.} - \overline{Y}_{3j.} = \sim N(\mu_{1j} - \mu_{3j}, \sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]) \quad \text{for} \quad j = 1, 2, 3.$$

Consequently,

$$\begin{bmatrix} \overline{Y}_{11.} - \overline{Y}_{31.} \\ \overline{Y}_{12.} - \overline{Y}_{32.} \\ \overline{Y}_{13.} - \overline{Y}_{33.} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{11.} - \mu_{31.} \\ \mu_{12.} - \mu_{32.} \\ \mu_{13.} - \mu_{33.} \end{bmatrix}, \sigma^2 \begin{bmatrix} {n_{11}}^{-1} + {n_{31}}^{-1} & 0 & 0 \\ 0 & {n_{12}}^{-1} + {n_{32}}^{-1} & 0 \\ 0 & 0 & {n_{13}}^{-1} + {n_{33}}^{-1} \end{bmatrix} \right)$$

Now express the numerator of the F-statisitc as a quadratic form, i.e.,

$$\begin{aligned} &\tfrac{1}{\sigma^2}\Sigma_{j=1}^3 (n_{1j}^{-1} + n_{3j}^{-1})^{-1}(\overline{Y}_{1j.} - \overline{Y}_{3j})^2 \\ &\quad = \begin{bmatrix} \overline{Y}_{11.} - \overline{Y}_{31.} \\ \overline{Y}_{12.} - \overline{Y}_{32.} \\ \overline{Y}_{13.} - \overline{Y}_{33.} \end{bmatrix}^T \begin{bmatrix} {n_{11}}^{-1} + {n_{31}}^{-1} & 0 & 0 \\ 0 & {n_{12}}^{-1} + {n_{32}}^{-1} & 0 \\ 0 & 0 & {n_{13}}^{-1} + {n_{33}}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \overline{Y}_{11.} - \overline{Y}_{31.} \\ \overline{Y}_{12.} - \overline{Y}_{32.} \\ \overline{Y}_{13.} - \overline{Y}_{33.} \end{bmatrix} \end{aligned}$$

Note that the matrix in the middle of this quadratic form is the inverse of the covariance matrix for $\begin{bmatrix} \overline{Y}_{11.} - \overline{Y}_{31.} \\ \overline{Y}_{12.} - \overline{Y}_{32.} \\ \overline{Y}_{13.} - \overline{Y}_{33.} \end{bmatrix}$. It follows from **Result 4.7** that

$$\frac{1}{\sigma^2}\Sigma_{j=1}^3 (n_{1j}^{-1} + n_{3j}^{-1})^{-1}(\overline{Y}_{1j.} - \overline{Y}_{3j})^2 \sim \chi^2_3(\sum_j \frac{[\mu_{1j} - \mu_{3j}]^2}{\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]})$$

Use

$$\begin{bmatrix} \overline{Y}_{11.} - \overline{Y}_{31.} \\ \overline{Y}_{12.} - \overline{Y}_{32.} \\ \overline{Y}_{13.} - \overline{Y}_{33.} \end{bmatrix} = B\underline{Y}$$

to express the quadratic form in the numerator of the F-statistic as a function of the vector of observations $\underline{Y}$. Then, $\frac{1}{\sigma^2}\Sigma_{j=1}^3 (n_{1j}^{-1} + n_{3j}^{-1})^{-1}(\overline{Y}_{1j.} - \overline{Y}_{3j})^2$

$$= \underline{Y}^T B^T \begin{bmatrix} {n_{11}}^{-1} + {n_{31}}^{-1} & 0 & 0 \\ 0 & {n_{12}}^{-1} + {n_{32}}^{-1} & 0 \\ 0 & 0 & {n_{13}}^{-1} + {n_{33}}^{-1} \end{bmatrix}^{-1} B\underline{Y},$$

Result 4.8 can be used to show that this quadratic form is distributed independently of S^2. Consequently, the statistic has an F-distribution with $(3, 46)$ degrees of freedom.

This result can be established in other ways. For example, let $A = (\frac{1}{\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]})$ be the $1x1$ matrix in the middle of the following quadratic form:

$$[\overline{Y}_{1j.} - \overline{Y}_{3j.}](\frac{1}{\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]})[\overline{Y}_{1j.} - \overline{Y}_{3j.}].$$

Now apply **Result 4.7** with $\Sigma = \sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]$. Then, $A\Sigma = (\frac{1}{\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]})\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]) = 1$ which is an idempotent $1x1$ matrix. Since both A and Σ are symmetric and positive definite, the conditions of **Result 4.7** are satisfied and

$$[\overline{Y}_{1j.} - \overline{Y}_{3j.}](\frac{1}{\sigma^2[{n_{1j}}^{-1} + {n_{3j}}^{-1}]})[\overline{Y}_{1j.} - \overline{Y}_{3j.}] \sim \chi^2_1(\delta_j^2),$$

where $\delta_j^2 = \mu_j^T A \mu_j = \frac{[\mu_{1j} - \mu_{3j}]^2}{\sigma^2[n_{1j}{}^{-1} + n_{3j}{}^{-1}])}$. Now apply **Theorem 5.3C** in Rencher to show that the sum of independent quadratic forms with chi-square distributions also has a chi-square distribution, i.e.,

$$\frac{1}{\sigma^2}\sum_{j=1}^{3}(n_{1j}^{-1} + n_{3j}^{-1})^{-1}(\overline{Y}_{1j.} - \overline{Y}_{3j.})^2 \;\sim\; \chi_3^2(\sum_j \delta_j^2)\ .$$

There are other ways to establish this result.

Next show that S^2 is independent of the quadratic form in the numerator of the F-statistic. One way to do this is to use **Result 4.8**. Write $\overline{Y}_{1j.} - \overline{Y}_{3j.}$ as a quadratic form in terms of $\underline{y}$.

This can be done by noting that (let j=1 for example):

$$\begin{aligned}
\overline{Y}_{11.} - \overline{Y}_{31.} &= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \overline{Y}_{11.} \\ \overline{Y}_{12.} \\ \overline{Y}_{13.} \\ \overline{Y}_{21.} \\ \overline{Y}_{22.} \\ \overline{Y}_{23.} \\ \overline{Y}_{31.} \\ \overline{Y}_{32.} \\ \overline{Y}_{33.} \\ \overline{Y}_{41.} \\ \overline{Y}_{42.} \\ \overline{Y}_{43.} \end{bmatrix} \\
&= \underline{a}^T \underline{y}
\end{aligned}$$

where

$$\underline{a}^T = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 0 & \cdots & 0 & 1/3 & 1/3 & 1/3 & 0 & \cdots & 0 \end{bmatrix}$$

Now we can write $[\overline{Y}_{11.} - \overline{Y}_{31.}]^2 = (\underline{a}^T\underline{y})(\underline{a}^T\underline{y}) = \underline{y}^T\underline{a}\underline{a}^T\underline{y} = \underline{y}^T A_1 \underline{y}$. The residual sum of squares is $(n - rank(X))S^2 = \underline{y}^T(I - P_X)\underline{y}$. Then, $A_1\Sigma(I - P_X) = \underline{a}\underline{a}^T(\sigma^2 I)(I - P_X) = \sigma^2\underline{a}\underline{a}^T(I - P_X) = \mathbf{0}$, because $\underline{a}$ is in the space spanned by the columns of the model matrix X. Verifying this last equality is easily done (you could use S-PLUS to numerically verify this, for example, but the details are omitted here to save space). Similarly, we can prove that $[\overline{Y}_{1j.} - \overline{Y}_{3j.}]^2$ is independent of the residual sum of squares for j=1,2,3. Consequently,

$$[\overline{Y}_{1j.} - \overline{Y}_{3j.}](\frac{1}{\sigma^2[n_{1j}{}^{-1} + n_{3j}{}^{-1}])})[\overline{Y}_{1j.} - \overline{Y}_{3j.}]$$

is independent of the residual sum of squares and we know shown that

$F = \frac{\sum_{j=1}^{3}(n_{1j}^{-1} + n_{3j}^{-1})^{-1}(\overline{Y}_{1j.} - \overline{Y}_{3j.})^2}{3S^2} \;\sim\; F_{(3,46)}(\sum_j \delta_j^2)$,

Note that the non-centrality parameter is $\delta^2 = \frac{1}{3}\sum_j \left[\mu_j^T A \mu_j\right] = \frac{1}{3}\sum_j \left[\frac{[\mu_{1j} - \mu_{3j}]^2}{\sigma^2[n_{1j}{}^{-1} + n_{3j}{}^{-1}])}\right]$, and this is zero if and only if $\mu_{1j} - \mu_{3j} = 0$, for j=1,2,3. It follows that the null hypothesis is

$$H_o : \mu_{1j} = \mu_{3j} \quad \text{for } j = 1, 2, 3,$$

i.e., there is no difference between the average increases of systolic blood pressure for Drugs 1 and 3 for any of the diseases.

(f) The F-statistic is 5.796863 with (3,46)d.f. and p-value=0.002. This is significant at the .05 level and the null hypothesis is rejected, i.e. for at least one disease, the average blood pressure increases are not equal for Drug 1 and Drug 3. The value for this F-statistic can be obtained by modifying the S-PLUS code used to compute F-tests for Type III sums of squares.

(g)
```
> summary.aov(lm.out1, ssType=3)
Type III Sum of Squares
              Df Sum of Sq  Mean Sq  F Value     Pr(F)
         Drug  3  2851.058 950.3527 8.622558 0.0001194
      Disease  2   371.711 185.8557 1.686270 0.1964555
Drug:Disease  6   737.888 122.9814 1.115811 0.3680099
   Residuals 46  5069.983 110.2170
```

i. $H_0:\ \mu_{ij} - \mu_{il} - \mu_{kj} + \mu_{kl} = 0$ for all (i,j) and (k,l).
$\Rightarrow$ Either interactions between factor1(Drug) and factor2(Disease) do not exist, or they are small relative to the error variance.

ii. The null hypothesis is

$$H_0 : \frac{1}{3}\sum_{j=1}^{3}\mu_{1j} = \frac{1}{3}\sum_{j=1}^{3}\mu_{2j} = \frac{1}{3}\sum_{j=1}^{3}\mu_{3j} = \frac{1}{3}\sum_{j=1}^{3}\mu_{4j}$$

Or, equivalently,

$$H_0 : \alpha_1 + \frac{1}{3}(\gamma_{11}+\gamma_{12}+\gamma_{13}) = \alpha_2 + \frac{1}{3}(\gamma_{21}+\gamma_{22}+\gamma_{23}) = \alpha_3 + \frac{1}{3}(\gamma_{31}+\gamma_{32}+\gamma_{33}) = \alpha_4 + \frac{1}{3}(\gamma_{41}+\gamma_{42}+\gamma_{43})$$

Then, the C matrix for $H_0 : C\underline{\beta} = 0$ may be specified as:

$$C = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

From the ANOVA table, the test statistic is 8.62 with degrees of freedom (3, 46) and p-value 0.0001. This indicates the mean increases in systolic blood pressure, averaging with equal weights across the Diseases, are not the same for all Drugs. Investigating further, for example, the mean increase in systolic blood pressure is lowest with Drug 3.

iii. The null hypothesis is

$$H_0 : \frac{1}{4}\sum_{i=1}^{4}\mu_{i1} = \frac{1}{4}\sum_{i=1}^{4}\mu_{i2} = \frac{1}{4}\sum_{i=1}^{4}\mu_{i3}.$$

Then, the C matrix for $H_0 : C\underline{\mu} = 0$ may be specified as:

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

From the ANOVA table, the test statistic is 1.686 with degrees of freedom (2, 46) and p-value 0.196. This indicates the mean increases in systolic blood pressure, averaging with equal weights across the Drugs, are nearly the same for all three Diseases (i.e., we cannot reject the null hypothesis).

2. (a)

$\mathrm{R}(\mu)$ $H_o : \mu + \sum_{i=1}^{a}\frac{n_{ij}}{n_{\cdot j}}\alpha_i + \sum_{j=1}^{b}\frac{n_{ij}}{n_{\cdot i}}\beta_j + \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}^2}\gamma_{ij} = 0$ or

$H_o : \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{n_{i\cdot}n_{\cdot j}}{n_{\cdot\cdot}^2}\mu_{ij} = 0$ $\rightarrow$ clearly not of interest since all $\mu_{ij} > 0$.

$\mathrm{R}(\underline{\alpha}|\mu) = \mathrm{R}(\underline{\alpha}|\mu,\underline{b})$ $H_o : \alpha_i + \sum_{j=1}^{b}\frac{n_{\cdot j}}{n_{\cdot\cdot}}(\beta_j + \gamma_{ij})$ are equal for all $i = 1, 2, \cdots, a$. or

$H_o : \sum_{j=1}^{b}\frac{n_{\cdot j}}{n_{\cdot\cdot}}\mu_{ij}$ are equal for all $i = 1, 2, \cdots, a$.

In this average, the weights assigned to the levels of the column factor are not necessarily equal, but the same set of weights is used at each level of the row factor. This is not the null hypothesis that is tested by the F-test for Type III sum of squares.

$R(\underline{b}|\mu) = R(\underline{b}|\mu, \underline{\alpha})$ $H_o : \beta_j + \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot\cdot}}(\alpha_i + \gamma_{ij})$ are equal for all $j = 1, 2, \cdots, b.$ or

$H_o : \sum_{i=1}^{a} \frac{n_{ij}}{n_{\cdot\cdot}}\mu_{ij}$ are equal for all $j = 1, 2, \cdots, b.$

$R(\underline{\gamma}|\mu, \underline{\alpha}, \underline{b})$ $H_o : \gamma_{ij} - \gamma_{kj} - \gamma_{il} + \gamma_{kl} = 0$ for all (i, j) and (k, l). or

$H_o : \mu_{ij} - \mu_{kj} - \mu_{il} + \mu_{kl} = 0$ for all (i, j) and (k, l).

The null hypothesis is that all interaction contrasts are zero.

(b) i.
```
> means <- tapply(battery$Y, list(battery$Temp,battery$Material),mean)
> means
          1     2     3
 15 153.00 173.5 142.5
 70 151.25 125.5  76.5
125 100.00  25.0  70.0
```

The plot indicates that there may be some interaction between Material and Temperature.

ii.
```
> lm.out1 <- lm(Y~Material*Temp,data=battery)
> summary.aov(lm.out1, ssType=1)
              Df Sum of Sq  Mean Sq  F Value       Pr(F)
     Material  2   5412.30 2706.150 11.66101 0.001914295
         Temp  2  17843.17 8921.587 38.44382 0.000010866
Material:Temp  4   6818.58 1704.644  7.34544 0.003915312
    Residuals 11   2552.75  232.068
> lm.out2 <- lm(Y~Temp*Material,data=battery)
> summary.aov(lm.out2, ssType=1)
              Df Sum of Sq  Mean Sq  F Value       Pr(F)
         Temp  2  17843.17 8921.587 38.44382 0.000010866
     Material  2   5412.30 2706.150 11.66101 0.001914295
Temp:Material  4   6818.58 1704.644  7.34544 0.003915312
    Residuals 11   2552.75  232.068
> # Compute Type III sums of squares and F-tests.
> summary.aov(lm.out1, ssType=3)
Type III Sum of Squares
              Df Sum of Sq  Mean Sq  F Value       Pr(F)
     Material  2   5083.78  2541.89 10.95321 0.002413331
         Temp  2  20306.43 10153.21 43.75100 0.000005804
Material:Temp  4   6818.57  1704.64  7.34544 0.003915312
    Residuals 11   2552.75   232.07
```

Conclusion: The Material effect, the Temp effect and the Material*Temp interaction are significant at the .05 level. Since the effects of Material and Temperature on mean life times of batteries are not additive, you should do further investigation of differences in mean life times for the three materials within each temperature, or compare trends in mean life times across temperatures for the three materials.

iii. Null hypotheses based on Type II sums of squares use cell means weighted with sample sizes in the cells, while null hypotheses based on Type III sums of squares use cell means weighted equally. Note that Type I SS and Type II SS are the same in this problem. With proportional counts, the order in which the factors are entered into the model does not matter.

iv. The null hypothesis for the F-test that the mean responses are the same for all three materials at the

temperature 15, say, is

$$H_0 : \mu_{11} = \mu_{21} = \mu_{31}$$

or under the unrestricted model,

$$H_0 : C\beta = 0$$

where

$$C^T = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix}$$

$$\beta = [\ \mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \gamma_{11} \quad \gamma_{12} \quad \gamma_{13} \quad \gamma_{21} \quad \gamma_{22} \quad \gamma_{23} \quad \gamma_{31} \quad \gamma_{32} \quad \gamma_{33}\ ]^T$$

The test statistic is

$$SS_{H_0} = (C^T\beta)^T(C(X^TX)^-C^T)^{-1}C^T\beta,$$

where X is the model matrix of the unrestricted model. Then, reject H_0 at the .05 level of significance if $(SS_{H_0}/2)/(SSE/11) > F_{(2,11),.05}$. The F-tests within other temperature levels can be conducted in a similar manner. //[.1in]
Alternatively, using the baseline restriction, as examined in Problem 1(a), the null hypothesis for the F-test is

$$H_0 : \alpha_2 = 0 \text{ and } \alpha_3 = 0$$

or

$$H_0 : C\beta = 0$$

where

$$C^T = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\beta = [\ \mu \quad \alpha_2 \quad \alpha_3 \quad \beta_2 \quad \beta_3 \quad \gamma_{22} \quad \gamma_{23} \quad \gamma_{32} \quad \gamma_{33}\ ]^T$$

The test statistic is as before with the exception that X is the model matrix for the restricted model. In the following S-plus codes, the F-tests are conducted with the baseline restricted model.

```
> # Function to test H0: C*beta = 0
> test.contr <- function( C, beta, X, sse, df1, df2 ){
 +     Cb <- C%*%beta
 +     SS.H0 <-  t(Cb)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%Cb
 +     F.stat <- (SS.H0/df1) / (sse/df2)
 +     pvalue <- 1 - pf(q=F.stat, df1=df1, df2=df2 )
 +     c( SS.H0=SS.H0, df1=df1, SSE=sse, df2=df1, F=F.stat, pvalue=pvalue)
 + }
>
> X <- model.matrix(lm.out1)
> sse <- deviance(lm.out1)
> df.resid <- lm.out1$df.residual
> beta <- lm.out1$coef
> # being careful of order of coefficients, specify C matrices
> c15 <- matrix( c( 0, 1, 0, 0, 0, 0, 0, 0, 0,
+                   0, 0, 1, 0, 0, 0, 0, 0, 0 ), ncol=9, byrow=T)
> c70 <- matrix( c( 0, 1, 0, 0, 0, 1, 0, 0, 0,
+                   0, 0, 1, 0, 0, 0, 1, 0, 0 ), ncol=9, byrow=T)
> c125 <- matrix( c( 0, 1, 0, 0, 0, 0, 0, 1, 0,
+                   0, 0, 1, 0, 0, 0, 0, 0, 1 ), ncol=9, byrow=T)
> rbind(
 +   Temp15=test.contr( C=c15, beta=beta, X=X, sse, df1=2, df2=df.resid),
 +   Temp70=test.contr( C=c70, beta=beta, X=X, sse, df1=2, df2=df.resid),
 +   Temp125=test.contr( C=c125, beta=beta, X=X, sse, df1=2, df2=df.resid) )
          SS.H0 df1     SSE df2         F       pvalue
```

```
 Temp15 1011.000   2 2552.75   2  2.178239 0.1596084206
 Temp70 7451.125   2 2552.75   2 16.053741 0.0005465377
Temp125 3768.750   2 2552.75   2  8.119920 0.0068239335
```

From above, we see that at Temp=70 and Temp=125, the null hypotheses of that the mean lifetimes are the same for all three materials are rejected.
Now, we use t-tests to determine which materials have different mean lifetimes at the temperatures 70 and 125. $t = \frac{(\overline{Y}_{1j} - \overline{Y}_{2j})}{\sqrt{\text{MSE}(\frac{1}{n_{1j}} + \frac{1}{n_{2j}})}} \sim t_{14} \quad j = 1, 2, 3$ can be used to test $H_o : \mu_{1j} = \mu_{2j}$ for each of temperature level. To control the Type I error rate using the Bonferoni Method (which is conservative) we will reject H_o if $t > t^*_{11_{(1-\frac{\alpha}{2(3)})}} = t^*_{11_{(.9917)}} = 2.822274$ at $\alpha = .05$.

(Temp 70)

- $(\mu_{12} = \mu_{22})$ $t = \frac{(151.25 - 125.50)}{\sqrt{(232.0682)(\frac{1}{4} + \frac{1}{2})}} = 1.952 \quad \rightarrow$ do not reject H_o.
- $(\mu_{12} = \mu_{32})$ $t = \frac{(151.25 - 76.50)}{\sqrt{(232.0682)(\frac{1}{4} + \frac{1}{2})}} = 5.666 \quad \rightarrow$ reject H_o.
- $(\mu_{22} = \mu_{32})$ $t = \frac{(125.50 - 76.50)}{\sqrt{(232.0682)(\frac{1}{2} + \frac{1}{2})}} = 3.217 \quad \rightarrow$ reject H_o.

(Temp 125)

- $(\mu_{13} = \mu_{23})$ $t = \frac{(100 - 25)}{\sqrt{(232.0682)(\frac{1}{1} + \frac{1}{1})}} = 4.020 \quad \rightarrow$ reject H_o.
- $(\mu_{13} = \mu_{33})$ $t = \frac{(100 - 70)}{\sqrt{(232.0682)(\frac{1}{2} + \frac{1}{1})}} = 1.608 \quad \rightarrow$ do not reject H_o.
- $(\mu_{23} = \mu_{33})$ $t = \frac{(25 - 70)}{\sqrt{(232.0682)(\frac{1}{2} + \frac{1}{1})}} = -2.089 \quad \rightarrow$ do not reject H_o.

So we can see from the above results that

- At Temp=70, there are significant differences for mean lifetimes of the batteries between Materials A and C (A>C) and between Materials B and C (B>C).
- At Temp=125, there are significant difference for mean lifetimes of the batteries between Materials A and B (A>B).

3. We will consider the following normal theory Gauss-Markov model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \quad \text{where } \epsilon_{ijk} \overset{\text{iid}}{\sim} N(0, \sigma^2) \quad i = 1, 2, 3, 4 \quad j = 1, 2, 3, 4.$$

and Y_{ijk} denotes the weight gain measured for the k-th female rat pup of i-th genotype of litter and j-th genotype of foster mother. Let A be the effect of litter genotype, and B be the effect of foster mother genotype.

```
filename rat "rats.dat"; data set1;
  infile rat;
  input A B y;
run; proc glm data=set1;
  class A B;
  model y = A B A*B / solution ss1 ss2 ss3 ss4 e e1 e2 e4 p;
  means A B A*B;
  lsmeans A*B / pdiff tdiff stderr;
  estimate 'B1-B2' B 1 -1 0 0 / e;
  estimate 'B1-B4' B -3 -1 1 3 / e;
  run;
```

```
 Source             DF         Squares     Mean Square    F Value    Pr > F
 Model              13     1485.995702      114.307362       2.01    0.0444
 Error              42     2391.089833       56.930710
 Corrected Total    55     3877.085536
```

Source	DF	Type I SS	Mean Square	F Value	Pr > F
A	3	127.3578913	42.4526304	0.75	0.5309
B	3	601.2304108	200.4101369	3.52	0.0230
A*B	7	757.4074003	108.2010572	1.90	0.0936

Source	DF	Type II SS	Mean Square	F Value	Pr > F
A	3	53.6870496	17.8956832	0.31	0.8149
B	3	601.2304108	200.4101369	3.52	0.0230
A*B	7	757.4074003	108.2010572	1.90	0.0936

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	3	21.3809445	7.1269815	0.13	0.9447
B	3	503.2284225	167.7428075	2.95	0.0437
A*B	7	757.4074003	108.2010572	1.90	0.0936

Source	DF	Type IV SS	Mean Square	F Value	Pr > F
A	3*	237.8674509	79.2891503	1.39	0.2583
B	3*	360.5134203	120.1711401	2.11	0.1132
A*B	7	757.4074003	108.2010572	1.90	0.0936

Since the test for interaction is not significant, the main effects may be approximately additive. Also, there is no evidence of a significant effect of litter genotype. Consequently, we can restrict our attention to foster mother genotype effects. We can examine some contrasts about foster mother genotype effects, but the results are not shown here. Fit a model with only foster mother genotype effects:

```
proc glm data=set1;
  class B;
  model y = B / solution ss1 ss2 ss3 e e1 e2 p;
  means B;
  run;
```

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	3	674.901252	224.967084	3.65	0.0182
Error	52	3202.184283	61.580467		
Corrected Total	55	3877.085536			

Source	DF	Type I SS	Mean Square	F Value	Pr > F
B	3	674.9012525	224.9670842	3.65	0.0182

Parameter		Estimate		Standard Error	t Value	Pr > |t|
Intercept		49.03076923	B	2.17645560	22.53	<.0001
B	1	6.24423077	B	2.93014303	2.13	0.0378
B	2	10.37832168	B	3.21483799	3.23	0.0022
B	3	4.33173077	B	2.93014303	1.48	0.1454
B	4	0.00000000	B	.	.	.

From the output above, we can see, mean weight gains of rat pups raised by foster mothers with genotype F is bigger than for foster mothers with other genotype J. This analysis only provides p-values for comparisons against foster mothers with genotype J. Other comparisons can be made with the LSMEANS option in PROC GLM or PROC MIXED in SAS. S-PLUs code for comparing pairs of means was illustrated in the code that was made available for assignment 8.

An alternative analysis could use a one-way ANOVA (the cell mean model) and compare all pairs of means for the 14 combinations of litter and mother genotypes.

```
data set2;
  set set1;
  C=10*A+B;
  run;
proc glm data=set2;
  class C;
  model y = C / solution e e2;
  estimate 'B1-B2' C .25 0 -.25 0 .25 0 -.25 .25 0 -.25 0 .25 -.25 0;
  lsmeans C / stderr tdiff pdiff;
  run;
```

Source	DF	Squares	Mean Square	F Value	Pr > F
Model	13	1485.995702	114.307362	2.01	0.0444
Error	42	2391.089833	56.930710		
Corrected Total	55	3877.085536			

Source	DF	Type II SS	Mean Square	F Value	Pr > F
C	13	1485.995702	114.307362	2.01	0.0444

Inspection of the t-tests and p-values for comparing all pairs of means for the 14 genotype combinations examined in this study reveals some significant t-tests for differences in litter genotypes within some foster mother genotypes. This seems to contradict the previous analysis which found no significant interaction among the litter and foster mother genotypes and no significant litter genotype effects. One possible reason why this occurs for these data may be that many interaction contrasts are close to zero, but a few are farther from zero. Then, part of the null hypothesis of no interaction may be essentially true and this dilutes the power of the F-test to find that a few interaction contrasts may not be zero (the test for interaction had a p-value of .0936). You should also consider that you did 91 t-tests to compare means for all pairs of factor combinations. If you perform each test at the .05 level, then each test has probability .05 of falsely rejecting the null hypothesis when there is essentially no difference between the mean response for two combinations of factors. If you use a conservative Bonferroni adjustment to achieve an experimentwise type one error level less than .05, the t-test must have p-value smaller than .05/91 to be significant.

The statement of this problem did not clearly describe the objectives of the study. If one objective was to determine whether or not rat pups tend to gain more weight when they are assigned to foster mothers of the same genotype, than it would be quite reasonable to compare the average of the weight gains for the four boxes on the main diagonal of the table to the average of the weight gains for the other boxes. Alternatively, within each litter genotype, one could compare mean weight gains for cases where the foster mothers had different genotypes to the case where the foster mother had the same genotype.

4. Consider a model with additive block effects

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \tau_k + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim NID(0, \sigma^2)$$

(a)

Source	DF	Type I SS	Mean Square	F Value	Pr > F
clinic	5	4634.894601	926.978920	41.99	<.0001
disease	2	663.918326	331.959163	15.04	<.0001
drug	3	2445.204851	815.068284	36.92	<.0001
drug*disease	6	517.280527	86.213421	3.91	0.0036
Error	41	905.115489	22.075988		
Corrected Total	57	9166.413793			

(i) F-test based on $R(\underline{\tau}|\mu)$. The null hypothesis for this f-test consists of setting to zero any contrast that can be constructed by choosing values for L2, L3, and L4 in the following:

```
Type I Estimable Functions : for clinic
   Effect              clinic
```

```
Intercept           0
clinic       1      L2
clinic       2      L3
clinic       3      L4
clinic       4      L5
clinic       5      L6
clinic       6      -L2-L3-L4-L5-L6
disease      1      -0.0303*L2-0.2386*L3+0.0364*L4+0.0808*L5-0.0909*L6
disease      2      -0.1414*L2+0.0114*L3+0.0364*L4-0.1414*L5
disease      3      0.1717*L2+0.2273*L3-0.0727*L4+0.0606*L5+0.0909*L6
drug         1      0.0606*L2-0.0227*L3-0.0727*L4-0.0505*L5
drug         2      -0.0505*L2-0.0227*L3+0.0273*L4-0.0505*L5
drug         3      -0.0707*L2+0.0682*L3+0.0182*L4+0.1515*L5
drug         4      0.0606*L2-0.0227*L3+0.0273*L4-0.0505*L5
drug*disease 1 1    0.0202*L2+0.0341*L3+0.0091*L4+0.0202*L5
drug*disease 1 2    0.0202*L2-0.0909*L3+0.0091*L4-0.0909*L5
drug*disease 1 3    0.0202*L2+0.0341*L3-0.0909*L4+0.0202*L5
drug*disease 2 1    0.0202*L2-0.0909*L3+0.0091*L4+0.0202*L5
drug*disease 2 2    -0.0909*L2+0.0341*L3+0.0091*L4-0.0909*L5
drug*disease 2 3    0.0202*L2+0.0341*L3+0.0091*L4+0.0202*L5
drug*disease 3 1    -0.0909*L2-0.0909*L3+0.0091*L4+0.0202*L5-0.0909*L6
drug*disease 3 2    -0.0909*L2+0.0341*L3+0.0091*L4+0.0202*L5
drug*disease 3 3    0.1111*L2+0.125*L3+0.1111*L5+0.0909*L6
drug*disease 4 1    0.0202*L2-0.0909*L3+0.0091*L4+0.0202*L5
drug*disease 4 2    0.0202*L2+0.0341*L3+0.0091*L4+0.0202*L5
drug*disease 4 3    0.0202*L2+0.0341*L3+0.0091*L4-0.0909*L5
```

$$
\begin{aligned}
H_0: \quad & (\text{Set } L2=1,\ L3=0,\ L4=0\ L5=0\ L6=0) \\
& \tau_1-\tau_6-.0303\beta_1+\cdots+.0202\gamma_{43} \ = \ 0 \\
& (\text{Set } L2=0,\ L3=1,\ L4=0\ L5=0\ L6=0) \\
& \tau_2-\tau_6-.2386\beta_1+\cdots+.0341\gamma_{43} \ = \ 0 \\
& (\text{Set } L2=0,\ L3=0,\ L4=1\ L5=0\ L6=0) \\
& \tau_3-\tau_6+.0364\beta_1+\cdots+0091\gamma_{43} \ = \ 0 \\
& (\text{Set } L2=0,\ L3=0,\ L4=0\ L5=1\ L6=0) \\
& \tau_4-\tau_6+.0808\beta_1+\cdots-.0909\gamma_{43} \ = \ 0 \\
& (\text{Set } L2=0,\ L3=0,\ L4=0\ L5=0\ L6=1) \\
& \tau_5-\tau_6-.0909\beta_1+\cdots+.0909\gamma_{43} \ = \ 0
\end{aligned}
$$

Due to the missing observations, block differences cannot be estimated free of the effects for diseases and drugs. For example, disease 2 is used only once in block II while it is used four times in blocks III, IV and VI. Drug 3 is used three times in block IV, but only once in block I. So we cannot exactly separate disease and drug effects from the effects of the blocks (different veterinary clinics). Nevertheless, the relatively large value for $R(\underline{\tau}|\mu)$ and the significant F-value suggest that there may be significant differences among clinics and we should adjust for such differences before examining possible effects of disease or drug. Note also that the MSE is now much smaller than it was in problem 1 where clinic effects were not included in the model. In the analysis in problem 1, a substantial part of the variation among clinics was put into the sum of squared residuals.

(ii) F-test based on $R(\underline{\alpha}|\mu,\underline{\tau})$.

```
Type I Estimable Functions for : disease
Effect              disease
```

```
Intercept           0
clinic       1      0
clinic       2      0
clinic       3      0
clinic       4      0
clinic       5      0
clinic       6      0
disease      1      L8
disease      2      L9
disease      3      -L8-L9
drug         1      0.08*L8-0.0286*L9
drug         2      -0.0444*L8-0.1034*L9
drug         3      -0.0486*L8+0.065*L9
drug         4      0.013*L8+0.067*L9
drug*disease 1 1    0.336*L8+0.0076*L9
drug*disease 1 2    -0.0085*L8+0.2083*L9
drug*disease 1 3    -0.2474*L8-0.2444*L9
drug*disease 2 1    0.2612*L8+0.0012*L9
drug*disease 2 2    0.0059*L8+0.2031*L9
drug*disease 2 3    -0.3116*L8-0.3078*L9
drug*disease 3 1    0.1416*L8-0.0101*L9
drug*disease 3 2    -0.0012*L8+0.262*L9
drug*disease 3 3    -0.189*L8-0.1869*L9
drug*disease 4 1    0.2612*L8+0.0012*L9
drug*disease 4 2    0.0037*L8+0.3267*L9
drug*disease 4 3    -0.252*L8-0.2609*L9
```

$$H_0: \quad (\text{Set } L8 = 1,\ L9 = 0)$$
$$\beta_1 - \beta_3 + .08\alpha_1 - .0444\alpha_2 + \cdots - .252\gamma_{43} = 0$$
$$(\text{Set } L8 = 0,\ L9 = 1)$$
$$\beta_2 - \beta_3 - .0286\alpha_1 - .1034\alpha_2; + \cdots - .2609\gamma_{43} = 0$$

Although this test is free of block effects, comparisons of mean volumes for different diseases use different sets of weights when averaging across drugs within individual diseases. This test is of limited practical interest. The large mean square, however, does suggest informally that the different disease may affect systolic blood pressure.

(iii) F-test based on $R(\underline{b}|\mu, \underline{\tau}, \underline{\alpha})$.

```
Type I Estimable Functions for : drug
Effect              drug
Intercept           0
clinic       1      0
clinic       2      0
clinic       3      0
clinic       4      0
clinic       5      0
clinic       6      0
disease      1      0
disease      2      0
disease      3      0
drug         1      L11
drug         2      L12
drug         3      L13
```

```
drug          4      -L11-L12-L13

drug*disease 1 1    0.3748*L11-0.0058*L12+0.014*L13
drug*disease 1 2    0.2951*L11+0.0271*L12+0.002*L13
drug*disease 1 3    0.3301*L11-0.0214*L12-0.016*L13
drug*disease 2 1    -0.0224*L11+0.3339*L12+0.0207*L13
drug*disease 2 2    0.0242*L11+0.2893*L12-0.0101*L13
drug*disease 2 3    -0.0018*L11+0.3767*L12-0.0106*L13
drug*disease 3 1    -0.0139*L11-0.0083*L12+0.2594*L13
drug*disease 3 2    0.0243*L11+0.024*L12+0.4052*L13
drug*disease 3 3    -0.0104*L11-0.0157*L12+0.3354*L13
drug*disease 4 1    -0.3385*L11-0.3199*L12-0.2941*L13
drug*disease 4 2    -0.3436*L11-0.3405*L12-0.3971*L13
drug*disease 4 3    -0.318*L11-0.3396*L12-0.3088*L13
```

$$
\begin{aligned}
H_0: \quad & (\text{Set } L11 = 1,\ L12 = 0,\ L13 = 0) \\
& \alpha_1 - \alpha_4 + .3748\gamma_{11} + \cdots - .318\gamma_{43} \ = \ 0 \\
& (\text{Set } L11 = 0,\ L12 = 1,\ L13 = 0) \\
& \alpha_2 - \alpha_4 - .0058\gamma_{11} + \cdots - .3396\gamma_{43} \ = \ 0 \\
& (\text{Set } L11 = 0,\ L12 = 0,\ L13 = 1) \\
& \alpha_3 - \alpha_4 + .014\gamma_{11} + \cdots - .3088\gamma_{43} \ = \ 0
\end{aligned}
$$

This test is free of both block and disease effects, but it uses different sets of weights when averaging across diseases for the various drugs. This test is also of little practical interest.

(iv) F-test based on $R(\underline{\gamma}|\mu, \underline{\tau}, \underline{\alpha}, \underline{b})$.

```
Type I Estimable Functions for: DISEASE*DRUG Effect
                Effect              drug*disease
               Intercept           0
               clinic       1      0
               clinic       2      0
               clinic       3      0
               clinic       4      0
               clinic       5      0
               clinic       6      0
               disease      1      0
               disease      2      0
               disease      3      0
               drug         1      0
               drug         2      0
               drug         3      0
               drug         4      0
               drug*disease 1 1    L15
               drug*disease 1 2    L16
               drug*disease 1 3    -L15-L16
               drug*disease 2 1    L18
               drug*disease 2 2    L19
               drug*disease 2 3    -L18-L19
               drug*disease 3 1    L21
               drug*disease 3 2    L22
               drug*disease 3 3    -L21-L22
               drug*disease 4 1    -L15-L18-L21
               drug*disease 4 2    -L16-L19-L22
               drug*disease 4 3    L15+L16+L18+L19+L21+L22
```

The null hypothesis consists of setting all interaction contrasts, between the effects of the diseases and drugs, equal to zero. Since there are six degrees of freedom for interaction, this is equivalent to setting six linearly independent interaction contrasts equal to zero. Every other interaction contrast is a linear function of these four contrasts.

$$
\begin{aligned}
H_0: \quad & (\text{Set } L15 = 1,\ L16 = 0,\ L18 = 0, L19 = 0,\ L21 = 0, \text{ and } L22 = 0) \\
& \gamma_{11} - \gamma_{13} - \gamma_{41} + \gamma_{43} = 0 \\
& (\text{Set } L15 = 0,\ L16 = 1,\ L18 = 0, L19 = 0,\ L21 = 0, \text{ and } L22 = 0) \\
& \gamma_{12} - \gamma_{13} - \gamma_{42} + \gamma_{43} = 0 \\
& (\text{Set } L15 = 0,\ L16 = 0,\ L18 = 1, L19 = 0,\ L21 = 0, \text{ and } L22 = 0) \\
& \gamma_{21} - \gamma_{23} - \gamma_{41} + \gamma_{43} = 0 \\
& (\text{Set } L15 = 0,\ L16 = 0,\ L18 = 0, L19 = 1,\ L21 = 0, \text{ and } L22 = 0) \\
& \gamma_{22} - \gamma_{23} - \gamma_{42} + \gamma_{43} = 0 \\
& (\text{Set } L15 = 0,\ L16 = 0,\ L18 = 0, L19 = 0,\ L21 = 1, \text{ and } L22 = 0) \\
& \gamma_{31} - \gamma_{33} - \gamma_{41} + \gamma_{43} = 0 \\
& (\text{Set } L15 = 0,\ L16 = 0,\ L18 = 0, L19 = 0,\ L21 = 0, \text{ and } L22 = 1) \\
& \gamma_{32} - \gamma_{33} - \gamma_{42} + \gamma_{43} = 0
\end{aligned}
$$

The F-test using $R(\gamma|\mu, \tau, \alpha, \beta)$ indicates that there is significant interaction between the effects of diseases and drugs. Given the significant interaction, tests for main effects for diseases or drugs would not be very meaningful, and the analysis should probably proceed by looking at contrasts of interest for available combinations of diseases and drugs.

(b)

Source	DF	Type III SS	Mean Square	F Value	Pr > F
clinic	5	4164.867844	832.973569	37.73	<.0001
disease	2	554.378429	277.189215	12.56	<.0001
drug	3	2405.711506	801.903835	36.32	<.0001
drug*disease	6	517.280527	86.213421	3.91	0.0036
Error	41	905.115489	22.075988		
Corrected Total	57	9166.413793			

(i) Type III Estimable Functions for: DISEASE

Effect		disease
Intercept		0
clinic	1	0
clinic	2	0
clinic	3	0
clinic	4	0
clinic	5	0
clinic	6	0
disease	1	L8
disease	2	L9
disease	3	-L8-L9
drug	1	0
drug	2	0
drug	3	0
drug	4	0

```
drug*disease 1 1     0.25*L8
drug*disease 1 2     0.25*L9
drug*disease 1 3     -0.25*L8-0.25*L9
drug*disease 2 1     0.25*L8
drug*disease 2 2     0.25*L9
drug*disease 2 3     -0.25*L8-0.25*L9
drug*disease 3 1     0.25*L8
drug*disease 3 2     0.25*L9
drug*disease 3 3     -0.25*L8-0.25*L9
drug*disease 4 1     0.25*L8
drug*disease 4 2     0.25*L9
drug*disease 4 3     -0.25*L8-0.25*L9
```

H_0 : **Set L8=1 and L9=0** to show that the null hypothesis includes

$$\beta_1 - \beta_3 + \tfrac{1}{4}(\gamma_{11} + \gamma_{21} + \gamma_{31} + \gamma_{41}) - \tfrac{1}{4}(\gamma_{13} + \gamma_{23} + \gamma_{33} + \gamma_{43}) \;=\; 0 \quad \text{which is equivalent to}$$

$$[\,\mu + \tfrac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta_1 + \tfrac{1}{4}(\gamma_{11} + \gamma_{21} + \gamma_{31} + \gamma_{41}) + \tau_k]$$

$$-[\,\mu + \tfrac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta_3 + \tfrac{1}{4}(\gamma_{13} + \gamma_{23} + \gamma_{33} + \gamma_{43}) + \tau_k] \;=\; 0$$

With respect to the cell means within any single block, this part of the null hypothesis is

$$\tfrac{1}{3}(\mu_{11k} + \mu_{21k} + \mu_{31k} + \mu_{41k}) - \tfrac{1}{3}(\mu_{13k} + \mu_{23k} + \mu_{33k} + \mu_{43k}) \;=\; 0$$

Set L8=0 and L9= 1 to show that the null hypothesis includes

$$\beta_2 - \beta_3 + \tfrac{1}{4}(\gamma_{12} + \gamma_{22} + \gamma_{32} + \gamma_{42}) - \tfrac{1}{4}(\gamma_{13} + \gamma_{23} + \gamma_{33} + \gamma_{43}) \;=\; 0 \quad \text{which is equivalent to}$$

$$[\,\mu + \tfrac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta_2 + \tfrac{1}{4}(\gamma_{12} + \gamma_{22} + \gamma_{32} + \gamma_{42}) + \tau_k]$$

$$-[\,\mu + \tfrac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta_3 + \tfrac{1}{4}(\gamma_{13} + \gamma_{23} + \gamma_{33} + \gamma_{43}) + \tau_k] \;=\; 0$$

With respect to the cell means within any single block, this part of the null hypothesis is

$$\tfrac{1}{3}(\mu_{12k} + \mu_{22k} + \mu_{32k} + \mu_{42k}) - \tfrac{1}{3}(\mu_{13k} + \mu_{23k} + \mu_{33k} + \mu_{43k}) \;=\; 0$$

Hence, this F-test provides a test of the null hypothesis

$$\frac{1}{4}\sum_{i=1}^{4}\mu_{i1k} \;=\; \frac{1}{4}\sum_{i=1}^{4}\mu_{i2k} \;=\; \frac{1}{4}\sum_{i=1}^{4}\mu_{i3k} \quad \text{for all blocks } k = 1,2,3,4,5,6. \tag{1}$$

(ii) F-test based on Type III SS for DRUG:

```
Type III Estimable Functions for: DRUG
          Effect               drug
          Intercept            0
          clinic       1       0
          clinic       2       0
          clinic       3       0
          clinic       4       0
          clinic       5       0
          clinic       6       0
          disease      1       0
          disease      2       0
          disease      3       0
          drug         1       L11
          drug         2       L12
          drug         3       L13
          drug         4       -L11-L12-L13
          drug*disease 1 1     0.3333*L11
          drug*disease 1 2     0.3333*L11
```

```
drug*disease 1 3     0.3333*L11
drug*disease 2 1     0.3333*L12
drug*disease 2 2     0.3333*L12
drug*disease 2 3     0.3333*L12
drug*disease 3 1     0.3333*L13
drug*disease 3 2     0.3333*L13
drug*disease 3 3     0.3333*L13
drug*disease 4 1     -0.3333*L11-0.3333*L12-0.3333*L13
drug*disease 4 2     -0.3333*L11-0.3333*L12-0.3333*L13
drug*disease 4 3     -0.3333*L11-0.3333*L12-0.3333*L13
```

H_0 : **Set L11=1, L10=0 and L11=0**to show that the null hypothesis includes

$$\alpha_1 - \alpha_4 + \tfrac{1}{3}(\gamma_{11} + \gamma_{12} + \gamma_{13}) - \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) \;=\; 0 \quad \text{which is equivalent to}$$

$$[\,\mu + \alpha_1 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{11} + \gamma_{12} + \gamma_{13}) + \tau_k]$$

$$-[\mu + +\alpha_4 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) + \tau_k] \;=\; 0$$

With respect to the cell means within any single block, this part of the null hypothesis is

$$\tfrac{1}{3}(\mu_{11k} + \mu_{12k} + \mu_{13k}) - \tfrac{1}{3}(\mu_{41k} + \mu_{42k} + \mu_{43k}) \;=\; 0$$

H_0 : **Set L11=0, L10=1 and L11=0**to show that the null hypothesis includes

$$\alpha_2 - \alpha_4 + \tfrac{1}{3}(\gamma_{21} + \gamma_{22} + \gamma_{23}) - \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) \;=\; 0 \quad \text{which is equivalent to}$$

$$[\,\mu + \alpha_2 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{21} + \gamma_{22} + \gamma_{23}) + \tau_k]$$

$$-[\mu + +\alpha_4 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) + \tau_k] \;=\; 0$$

With respect to the cell means within any single block, this part of the null hypothesis is

$$\tfrac{1}{3}(\mu_{21k} + \mu_{22k} + \mu_{23k}) - \tfrac{1}{3}(\mu_{41k} + \mu_{42k} + \mu_{43k}) \;=\; 0$$

H_0 : **Set L11=0, L10=0 and L11=1**to show that the null hypothesis includes

$$\alpha_3 - \alpha_4 + \tfrac{1}{3}(\gamma_{31} + \gamma_{32} + \gamma_{33}) - \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) \;=\; 0 \quad \text{which is equivalent to}$$

$$[\,\mu + \alpha_3 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{11} + \gamma_{12} + \gamma_{13}) + \tau_k]$$

$$-[\mu + +\alpha_4 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3) + \tfrac{1}{3}(\gamma_{41} + \gamma_{42} + \gamma_{43}) + \tau_k] \;=\; 0$$

With respect to the cell means within any single block, this part of the null hypothesis is

$$\tfrac{1}{3}(\mu_{31k} + \mu_{32k} + \mu_{33k}) - \tfrac{1}{3}(\mu_{41k} + \mu_{42k} + \mu_{43k}) \;=\; 0$$

Hence, this F-test provides a test of the null hypothesis

$$\frac{1}{3}\sum_{j=1}^{3}\mu_{1jk} \;=\; \frac{1}{3}\sum_{j=1}^{3}\mu_{2jk} \;=\; \frac{1}{3}\sum_{j=1}^{3}\mu_{3jk} \;=\; \frac{1}{3}\sum_{j=1}^{3}\mu_{4jk} \quad \text{for all blocks } k = 1, 2, 3, 4, 5, 6. \tag{2}$$

(iii) F-test based on Type III SS for DISEASE*DRUG:

```
Type III Estimable Functions for: DISEASE*DRUG
                Effect              drug*disease
                Intercept           0
                clinic      1       0
                clinic      2       0
                clinic      3       0
                clinic      4       0
                clinic      5       0
                clinic      6       0
                disease     1       0
                disease     2       0
                disease     3       0
                drug        1       0
```

```
drug          2       0
drug          3       0
drug          4       0
drug*disease 1 1      L15
drug*disease 1 2      L16
drug*disease 1 3      -L15-L16
drug*disease 2 1      L18
drug*disease 2 2      L19
drug*disease 2 3      -L18-L19
drug*disease 3 1      L21
drug*disease 3 2      L22
drug*disease 3 3      -L21-L22
drug*disease 4 1      -L15-L18-L21
drug*disease 4 2      -L16-L19-L22
drug*disease 4 3      L15+L16+L18+L19+L21+L22
```

This is the same test for interaction provided by the Type I (and Type II) sums of squares, so the null hypothesis and conclusions are the same as in part (a). The null hypothesis can also be written as $H_o : (\gamma_{ij} - \gamma_{kj}) - (\gamma_{il} - \gamma_{kl}) = 0 \qquad$ for all (i, j) and (k, l).

Given the significant interaction, we should go on to create a profile plot and compare cell means with appropriate t-tests or contrasts.

(c) The ANOVA table shows a significant difference exists in mean increase of systolic blood pressures among the 12 treatments. The t-tests between individual treatments within each Disease group suggest (controlling the type I error rate by Bonferoni's method with $t^*_{41(1-\frac{.05}{2(6)})} = 2.7723$)

- For Disease 1, Drug 4 has significantly less mean increase of systolic blood pressure than the other drugs.
- For Disease 2, Drug 1 and Drug 2 have significantly higher increase systolic blood pressure than Drug 3 and Drug 4.
- For Disease 3, Drug 1 have significantly higher increase systolic blood pressure than Drug 3 and Drug 4. The deviations Drug 2 from Drugs 3 and 4 are nearly significant. But, there is no significant differences between Drugs 1 and 2 and between Drugs 3 and 4.