STAT 504 **Solutions to Assignment 2** **Spring 2019**

PART I

1 (a) Since A is nonnegative definite (semipositive definite) matrix, we know by definition that

$$\underset{\sim}{y}^T A \underset{\sim}{y} \geq 0 \text{ for any vector } \underset{\sim}{y}.$$

For any vector $\underset{\sim}{z}$, let $\underset{\sim}{y} = P\underset{\sim}{z}$. Then,

$$\underset{\sim}{z}^T P^T A P \underset{\sim}{z} = (P\underset{\sim}{z})^T A\,(P\underset{\sim}{z}) = \underset{\sim}{y}^T A \underset{\sim}{y} \geq 0 \quad \text{for any vector} \quad \underset{\sim}{z}.$$

Consequently, $P^T AP$ satisfies the definition of a nonnegative definite matrix. This result holds even when P is a singualr matrix.

(b) For this result, we will make use of the condition that P is a nonsingular matrix. Since A is a postive definite matrix, we have by definition that $\underset{\sim}{y}^T A \underset{\sim}{y} > 0$ for every $\underset{\sim}{y} \neq 0$. For any vector $\underset{\sim}{z}$, let $\underset{\sim}{y} = P\underset{\sim}{z}$. Then,

$$\underset{\sim}{z}^T P^T A P \underset{\sim}{z} = (P\underset{\sim}{z})^T A\,(P\underset{\sim}{z}) = \underset{\sim}{y}^T A \underset{\sim}{y} > 0 \quad \text{for every vector} \quad \underset{\sim}{y} = P\underset{\sim}{z} \neq 0.$$

Since P is nonsingular, the columns of P are linearly independent and $\underset{\sim}{y} = P\underset{\sim}{z} = 0$ only if $\underset{\sim}{z} = 0$. Consequently, $P^T AP$ satisfies the definition of a positive definite matrix.

2.(a) For B to satisfy the definition of a generalized inverse of A, we must have $ABA = A$. Although this simple example could be done by hand, to review some Splus we can do the following:

```
> b <- matrix(c(1,0,0,0),ncol=4)
> a <- matrix(c(1,2,5,-2),ncol=1)
> a %*% b %*% a
     [,1]
[1,]    1
[2,]    2
[3,]    5
[4,]   -2
```

(b) Two other generalized inverses for A are the following

$$G_1 = [0,\ 0.5,\ 0,\ 0],\ \ G_2 = [0,\ 0,\ 0.2,\ 0]. \quad \text{Check in each case that } AG_iA = A.$$

3.(a) and **(b)** By the definition of an eigenvector, $\underset{\sim}{x}$ is an eigenvector for $I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T$ if there is a scalar λ such that

$$(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T)\underset{\sim}{x} \;=\; \lambda\underset{\sim}{x}$$

Note that

$$(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T)\underset{\sim}{x} \;=\; (\underset{\sim}{x} - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}(\underset{\sim}{x}^T\underset{\sim}{x}) \;=\; \underset{\sim}{x} - \underset{\sim}{x} \;=\; \underset{\sim}{0}$$

Consequently, $\underset{\sim}{x}$ is an eigenvector of $(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T)$ corresponding to the eigenvalue $\lambda = 0$.

(c) Since $\underset{\sim}{x}^T\underset{\sim}{u} = 0$, we have $(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T)\underset{\sim}{u} \;=\; \underset{\sim}{u}$.

Therefore, $\underset{\sim}{u}$ is an eigenvector of $(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T)$ corresponding to an eigenvalue of $\lambda = 1$.

(d) From part (b) the eigenvalue associated with $\underset{\sim}{x}$ is 0. Furthermore, $\underset{\sim}{x}$ is the basis for a one dimensional space (a line). The space orthogonal to $\underset{\sim}{x}$ has dimension $n-1$ and we can find a set of vectors $\underset{\sim 1}{u}, \ldots, \underset{\sim n-1}{u}$ that provide a basis for that space, are orthogonal to each other and are also orthogonal to $\underset{\sim}{x}$. Each of these vectors corresponds to and eigenvalue of 1. Hence, $I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T$ has one zero eigenvalue and the other $n-1$ eigenvalues are all 1. Furthermore,

$$rank(I - \underset{\sim}{x}(\underset{\sim}{x}^T\underset{\sim}{x})^{-1}\underset{\sim}{x}^T) \;=\; n-1,$$

the number of non-zero eigenvalues.

4. Since A is an $n \times n$ symmetric matrix with $rank(A) = r$, we can use the spectral decomposition to write A as

$$\begin{aligned}
\underset{n\times n}{A} &= \underset{n\times n}{L} \begin{bmatrix} \Delta_r & \underset{r\times(n-r)}{\mathbf{0}} \\ \underset{(n-r)\times r}{\mathbf{0}} & \underset{(n-r)\times(n-r)}{\mathbf{0}} \end{bmatrix} \underset{n\times n}{L^T} \\
&= \begin{bmatrix} \underset{n\times r}{L_1} & \underset{p\times(n-r)}{L_2} \end{bmatrix} \begin{bmatrix} \Delta_r & \underset{r\times(n-r)}{\mathbf{0}} \\ \underset{(n-r)\times r}{\mathbf{0}} & \underset{(n-r)\times(n-r)}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{r\times n}{L_1^T} \\ \underset{(n-r)\times n}{L_2^T} \end{bmatrix} \\
&= \underset{n\times r}{L_1}\, \Delta_r\, \underset{r\times n}{L_1^T}
\end{aligned}$$

Note that

$$\underset{n\times n}{G} = \underset{n\times n}{L} \begin{bmatrix} \Delta_r^{-1} & \underset{r\times(n-r)}{\mathbf{0}} \\ \underset{(n-r)\times r}{\mathbf{0}} & \underset{(n-r)\times(n-r)}{\mathbf{0}} \end{bmatrix} \underset{n\times n}{L^T}$$

$$= \begin{bmatrix} \underset{n\times r}{L_1} & \underset{n\times(n-r)}{L_2} \end{bmatrix} \begin{bmatrix} \Delta_r^{-1} & \underset{r\times(n-r)}{\mathbf{0}} \\ \underset{(n-r)\times r}{\mathbf{0}} & \underset{(n-r)\times(n-r)}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{r\times n}{L_1^T} \\ \underset{(n-r)\times n}{L_2^T} \end{bmatrix}$$

$$= \underset{n\times r}{L_1} \Delta_r^{-1} \underset{r\times n}{L_1^T}$$

Now show that the four properties of the Moore-Penrose inverse are satisfied.

(i) $AGA \;=\; A$.

$$\begin{aligned} AGA &= \left(\underset{n\times r}{L_1} \Delta_r \underset{r\times n}{L_1^T}\right) \left(\underset{n\times r}{L_1} \Delta_r^{-1} \underset{r\times n}{L_1^T}\right) \left(\underset{n\times r}{L_1} \Delta_r \underset{r\times n}{L_1^T}\right) \\ &= \underset{n\times r}{L_1} \Delta_r \left(\underset{r\times n}{L_1^T} \underset{n\times r}{L_1}\right) \Delta_r^{-1} \left(\underset{r\times n}{L_1^T} \underset{n\times r}{L_1}\right) \Delta_r \underset{r\times n}{L_1^T} \\ &= \underset{n\times r}{L_1} \Delta_r \Delta_r^{-1} \Delta_r \underset{r\times n}{L_1^T} \qquad \text{since } L_1^T L_1 = I_r \text{ and } L_1^T L_1 = I_r. \\ &= \underset{n\times r}{L_1} \Delta_r \underset{r\times n}{L_1^T} \qquad \text{since } \Delta_r \Delta_r^{-1} = I_r. \\ &= A \end{aligned}$$

(ii) $GAG \;=\; G$. The proof is similar to (i).

(iii) AG is symmetric. i.e. $(AG)^T \;=\; AG$.

$$\begin{aligned} AG &= \left(\underset{n\times r}{L_1} \Delta_r \underset{r\times n}{L_1^T}\right) \left(\underset{n\times r}{L_1} \Delta_r^{-1} \underset{r\times n}{L_1^T}\right) \\ &= \underset{n\times r}{L_1} \Delta_r \left(\underset{r\times n}{L_1^T} \underset{n\times r}{L_1}\right) \Delta_r^{-1} \underset{r\times n}{L_1^T} \\ &= \underset{n\times r}{L_1} \Delta_r \Delta_r^{-1} \underset{r\times n}{L_1^T} \quad \text{since } L_1^T L_1 = I_r. \\ &= \underset{n\times r}{L_1} \underset{r\times n}{L_1^T} \quad \text{since } \Delta_r \Delta_r^{-1} = I_r. \end{aligned}$$

Then

$$(AG)^T \;=\; \left(\underset{n\times r}{L_1} \underset{r\times n}{L_1^T}\right)^T \;=\; (\underset{r\times n}{L_1^T})^T (\underset{n\times r}{L_1})^T \;=\; \underset{n\times r}{L_1} \underset{r\times n}{L_1^T} \;=\; AG\,.$$

(iv) GA is symmetric. i.e. $(GA)^T \;=\; GA$. The proof is similar to (iii).

5.(a)
```
> V <- matrix(c(3,-1,1,-1,5,-1,1,-1,3),ncol=3,byrow=T)
> V
     [,1] [,2] [,3]
[1,]    3   -1    1
[2,]   -1    5   -1
[3,]    1   -1    3
> eigen(V)
$values:
[1] 6 3 2

$vectors
           [,1]      [,2]          [,3]
[1,]  0.4082483 0.5773503  7.071068e-01
[2,] -0.8164966 0.5773503  1.048966e-16
[3,]  0.4082483 0.5773503 -7.071068e-01
```

(b)
```
> ##########################################################
> # Function: spectral
> #    input: V = Symmetric matrix
> #           p = Power(e.g. -1   -> inverse matrix
> #                              -1/2 -> inv. squart root matrix
> #   output: inverse or inverse square root matrix
> ##########################################################
> spectral <- function(V,p)
+ {
+       eigen.V <- eigen(V)
+       eval    <- eigen.V$values
+       evec    <- eigen.V$vectors
+       spec.V  <- evec %*% diag(eval^p) %*% t(evec)
+       return(spec.V)
+ }
```

(c)
```
> VV <- spectral(V,-1/2)
> VV
            [,1]       [,2]        [,3]
[1,]  0.61404486 0.05636733 -0.09306192
[2,]  0.05636733 0.46461562  0.05636733
[3,] -0.09306192 0.05636733  0.61404486

#  The following should be an identity matrix
> VV %*% V %*% VV
```

```
                [,1]          [,2]         [,3]
[1,]  1.000000e+00 -4.931934e-17 1.254964e-17
[2,] -5.349013e-17  1.000000e+00 2.705084e-16
[3,] -1.240192e-16  3.081641e-16 1.000000e+00
```

(d) $V^{-1/2} = UD^{-1/2}U^T$, where $U = \begin{bmatrix} \underset{\sim}{u_1} \; \underset{\sim}{u_2} \; \underset{\sim}{u_3} \end{bmatrix}$ and $D = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}$.

Then, for $\underset{\sim}{Z} = V^{-1/2}\underset{\sim}{Y}$ we have

$$\begin{aligned} \mathrm{E}(\underset{\sim}{Z}) &= \mathrm{E}(V^{-\frac{1}{2}}\underset{\sim}{Y}) = V^{-\frac{1}{2}}\mathrm{E}(\underset{\sim}{Y}) = V^{-\frac{1}{2}}\underset{\sim}{0} = \underset{\sim}{0} \\ \mathrm{Var}(\underset{\sim}{Z}) &= \mathrm{Var}(V^{-\frac{1}{2}}\underset{\sim}{Y}) = V^{-\frac{1}{2}}\mathrm{Var}(\underset{\sim}{Y})[V^{-\frac{1}{2}}]^T = V^{-\frac{1}{2}}VV^{-\frac{1}{2}} = I \end{aligned}$$

6.(a) Let $\underline{b} = (X^TX)^-X^T\underline{Y}$ for some generalized inverse $(X^TX)^-$ of X^TX. Then $X^TX\underline{b} = X^TX(X^TX)^-X^T\underline{Y} = X^T\underline{Y}$ because $X(X^TX)^-X^TX = P_XX = X$.

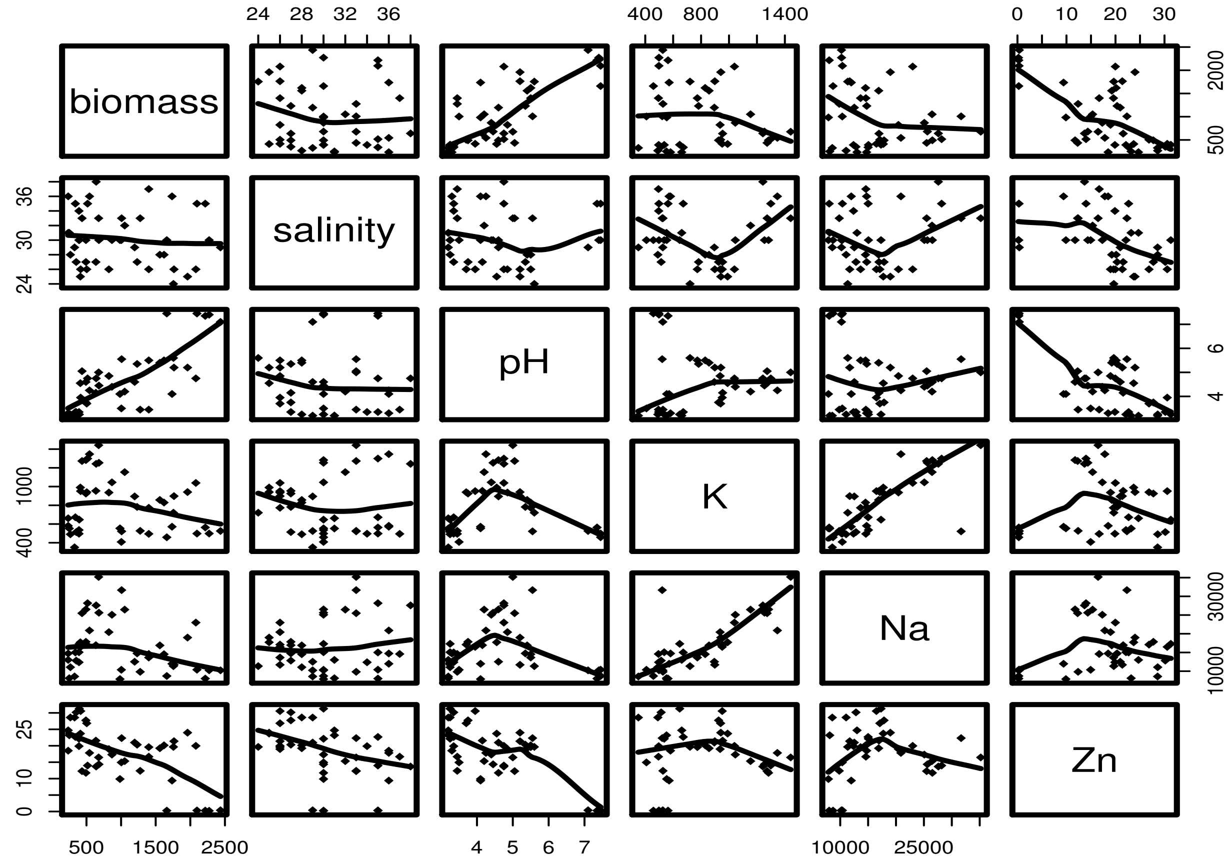
(b) The answer to the first question is yes if X has full column rank. The answer is no if X does not have full column rank. Let $\underline{b}^* = (X^TX)^-X^T\underline{Y} + \underline{a}$ for some generalized inverse $(X^TX)^-$ of X^TX. Then, $X^TX\underline{b}^* = X^TX(X^TX)^-X^T\underline{Y} + X^TX\underline{a} = X^T\underline{Y} + X^TX\underline{a}$. This is a solution to the normal equations if $X^TX\underline{a} = 0$. If X does not have full column rank, the columns of X are linearly dependent and there is at least one $\underline{a} \neq 0$ for which $X\underline{a} = 0$. Consequently, $X^TX\underline{a} = 0$ for that $\underline{a}$. If X has full column rank, then X^TX is a non-singular matrix and there is no $\underline{a} \neq \underline{0}$ for which $X^TX\underline{a} = 0$.

(c) It is easy to show that $b = (X^TX)^{-1}X^T\underline{Y}$ is a solution to the normal equations. Suppose there exits another solution to the normal equations, let's say $\underline{b}^*$. We can write $\underline{b}^* = \underline{b} + \underline{a} = (X^TX)^{-1}X^T\underline{Y} + \underline{a}$, for some $\underline{a} \neq 0$. Since, rank(X^TX)=rank(X) and X has full column rank, X^TX also has full column rank and the columns of X^TX are linearly independent. Therefore, it is impossible to have $X^TX\underline{a} = 0$, for some $\underline{a} \neq 0$. Consequently, $\underline{b} = (X^TX)^{-1}X^T\underline{Y}$ is the unique solution to the normal equations.

7.(a)

```
# Enter the data into a data frame.

>biomass <- read.table("biomass.txt",header=T)
>biomass

#  Compute correlations and round the results
#  to four significant digits
```

```
> round(cor(biomass[-(1:2)]),4)

          biomass salinity      pH       K      Na      Zn
biomass    1.0000  -0.1032  0.7742 -0.2051 -0.2721 -0.6244
salinity  -0.1032   1.0000 -0.0513 -0.0205  0.1623 -0.4208
pH         0.7742  -0.0513  1.0000  0.0187 -0.0378 -0.7222
K         -0.2051  -0.0205  0.0187  1.0000  0.7921  0.0740
Na        -0.2721   0.1623 -0.0378  0.7921  1.0000  0.1171
Zn        -0.6244  -0.4208 -0.7222  0.0740  0.1171  1.0000

#  Create a scatterplot matrix

> pairs(biomass[ ,3:8])

#  Create a scatterplot matrix with smooth  curves.

 points.lines <- function(x, y)
   {
    points(x, y)
    lines(loess.smooth(x, y, 0.90))
   }

 par(pch=18, mkh=.15, cex=1.2, lwd=3)
 pairs(biomass[ ,-(1:2)], panel=points.lines)
```



There seems to be a positive linear relationship between Y and pH and a negative linear relationship between Y and Zn. The other three variables appear to be at more weakly correlated with the response Y, but Y seems to be lower for larger values of K and Na. With the exception of one potentially influential case, it appears that K and Na have a strong positive correlation. There appears to be a negative association between levels of pH and Zn in the soil.

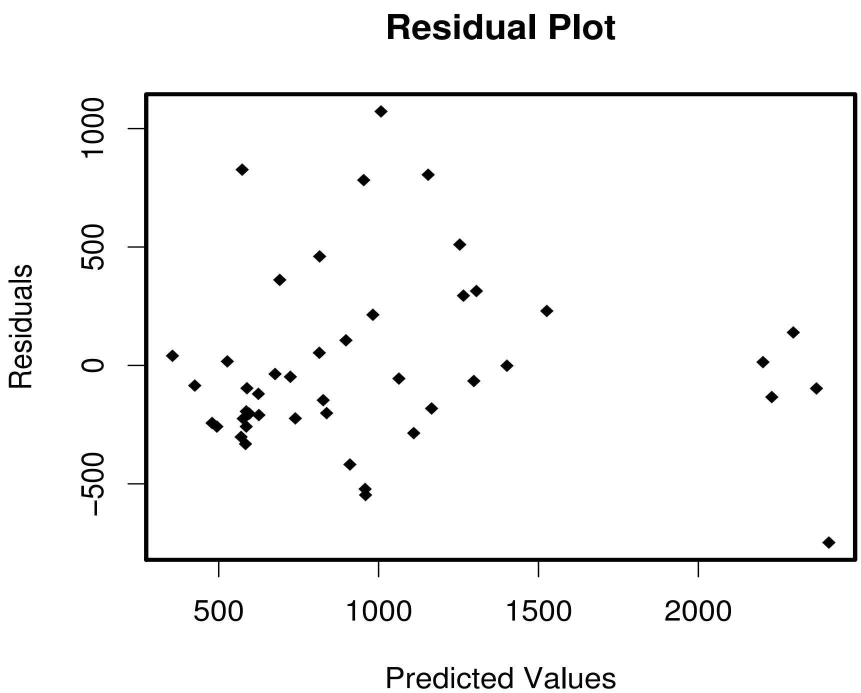
(b) The model matrix has full column rank.

```
# Construct the response vector and model matrix

> Y<-as.matrix(biomass[ ,3])
> X<-as.matrix(biomass[,4:8])
> X0 <- rep(1, length(Y))
> X<-cbind(X0,X)

# Check the rank of the model matrix

> qr(X)$rank
[1] 6
```

(c)
```
> b <- solve(t(X)%*%X) %*% t(X) %*% Y
> b
                    [,1]
      X0  1.252455e+003
salinity -3.028491e+001
      pH  3.054880e+002
       K -2.852645e-001
      Na -8.665809e-003
      Zn -2.067677e+001
```

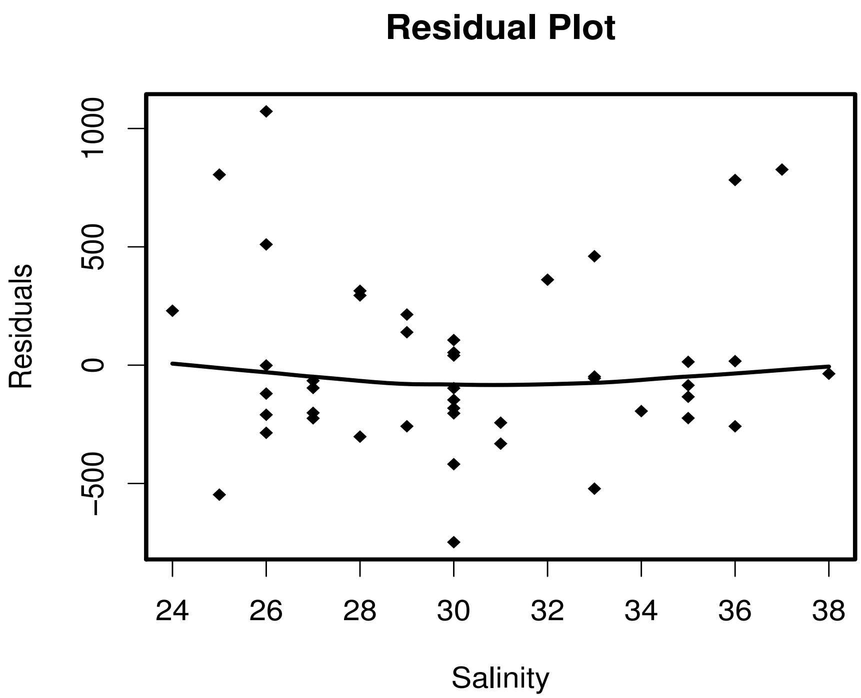
(d)
```
> yhat <- X %*% b
> yhat
        [,1]
 1  724.1509
 2  739.6501
 3  690.9358
 4  814.6608
 5 1063.5814
 6  957.8484
 7  527.0257
      .
      .
      .
42 1297.8306
43 1401.3330
44 1305.8814
45 1265.4053

> e <- Y - yhat
> e
           [,1]
 1   -48.150889
 2  -223.650127
 3   361.064216
 4    53.339248
 5   -55.581358
 6  -521.848358
 7    16.974252
          .
          .
          .
```

```
42   -65.830647
43    -1.332987
44   314.118581
45   294.594716


# Plot residuals against fitted values #

# Specify plotting symbol and size of graph in inches.
# pch=18  requests a filled diamond as a plotting
#          symbol
# mkh=b   requests plotting symbols that are  b
#          inches high
# cex=c   requests that the size of characters used to print
#         labels are c times the default for the printer
# mar     mar=c(5,5,4,2) defines the number of lines of
#          text allowed on each side of the figure,
#          starting on the bottom and moving clockwise.


> par(fin=c(8.0,8.0),pch=18,mkh=.1,cex=1.3,mar=c(5,5,4,2))

> plot(yhat,e,xlab="Predicted Values",
   ylab="Residuals", main="Residual Plot")
```

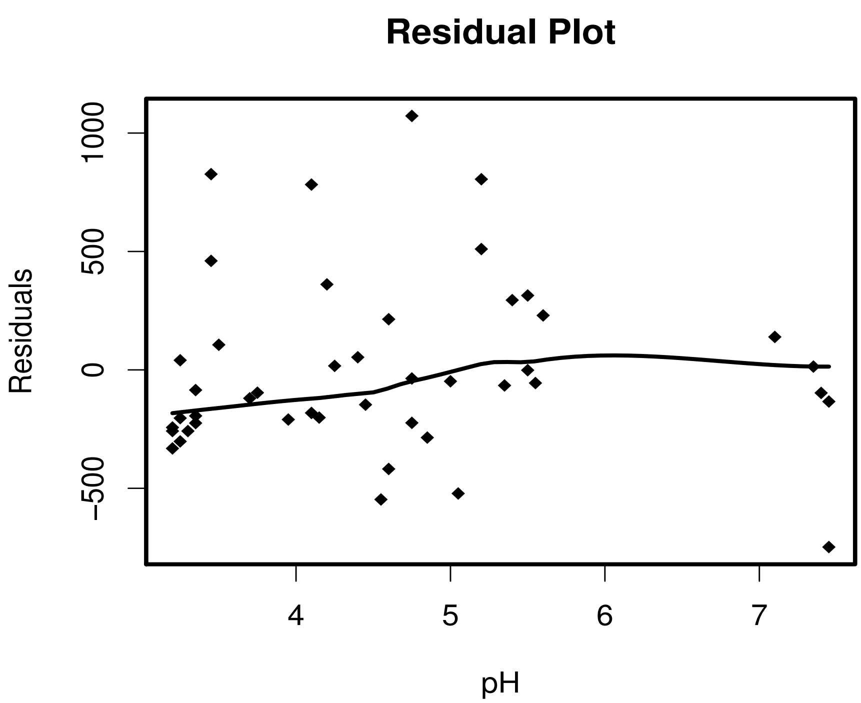


To some students this plot suggested that the assumption of constant variance does not hold. It is difficult to tell because there are only a few points on the right side of the plot. Another possibility is that the specified regression model is not a good approximation to the true model.

(e)
```
> plot(biomass$salinity,e,
+       xlab="Salinity",ylab="Residuals",main="Residual Plot")
> lines(loess.smooth(biomass$salinity, e, 0.90))
```

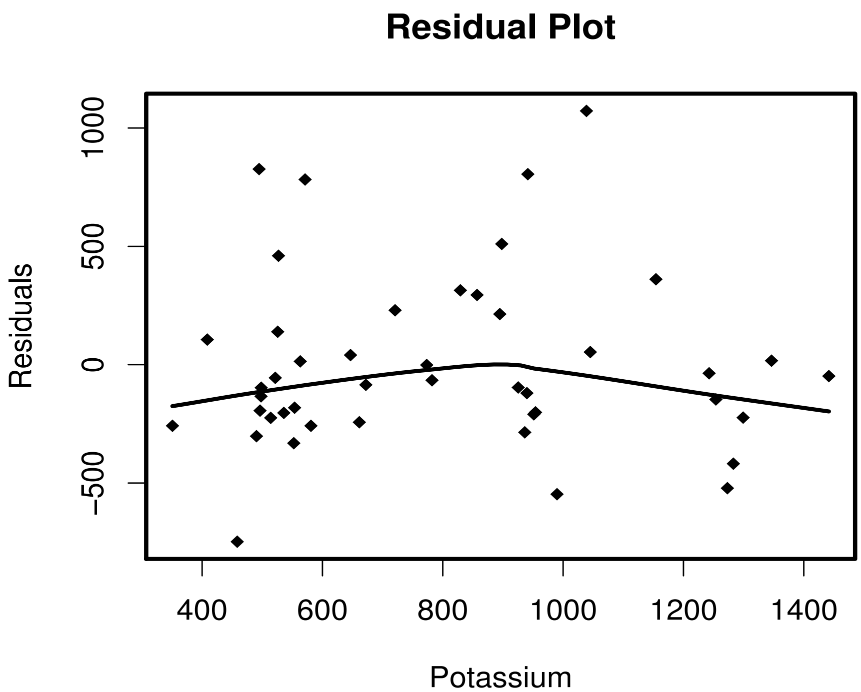



```
> plot(biomass$pH,e,
+       xlab="pH",ylab="Residuals",main="Residual Plot")
> lines(loess.smooth(biomass$pH, e, 0.90))
```

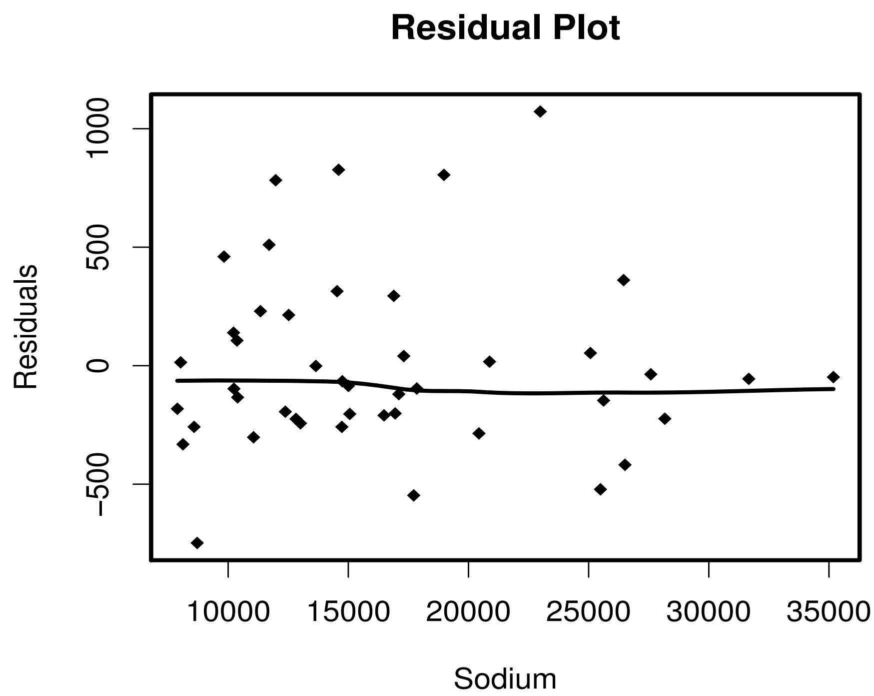



```
> plot(biomass$K,e,
```

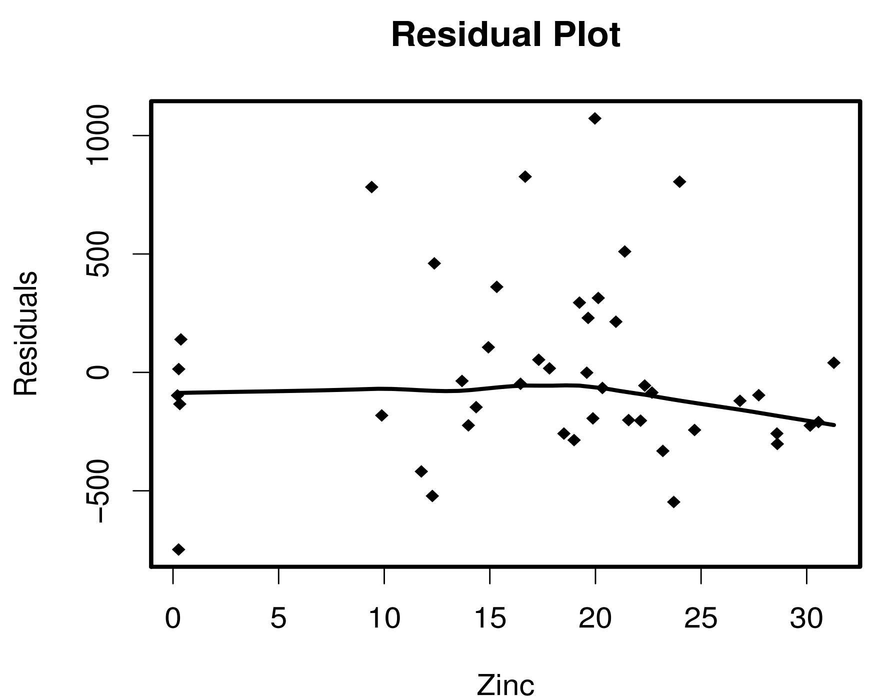
```
+      xlab="Potassium",ylab="Residuals",main="Residual Plot")
> lines(loess.smooth(biomass$K, e, 0.90))
```




```
> plot(biomass$Na,e,
+      xlab="Sodium",ylab="Residuals",main="Residual Plot")
> lines(loess.smooth(biomass$Na, e, 0.90))
```

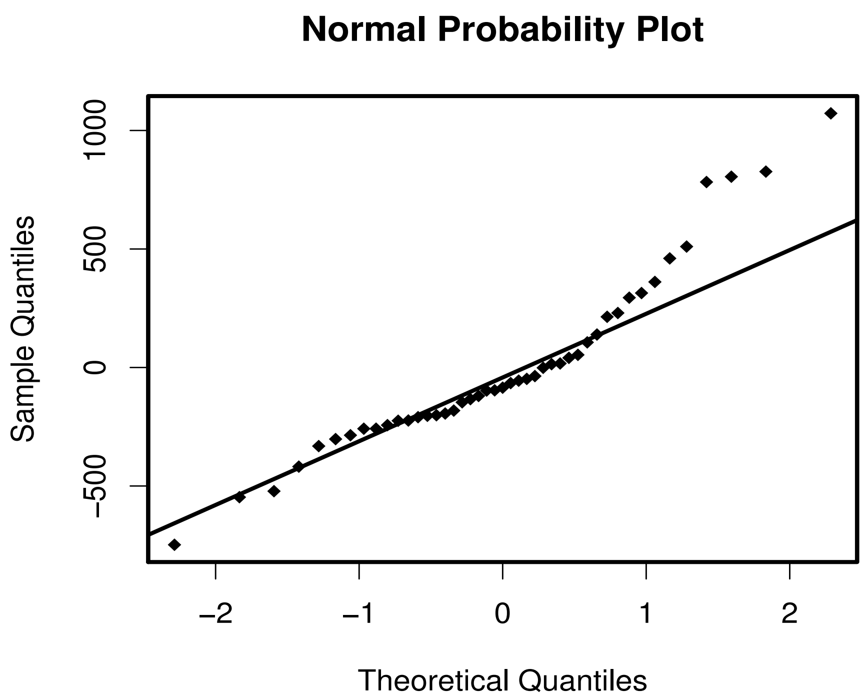



```
> plot(biomass$Zn,e,
+       xlab="Zinc",ylab="Residuals",main="Residual Plot")
> lines(loess.smooth(biomass$Zn, e, 0.90))
```



Based on the respective plots, the variance does not appear to be constant with respect to some of the explanatory variables, supporting our conclusion in part(d). There may be a curved effect of some of the variables on the mean response.

(f)
```
> qqnorm(e, main="Normal Probability Plot")
> qqline(e)
```



Based on this plot, the distribution of the random errors appears to be skewed to the right. This could be a result of non-homogeneous error variances or failure to incorporate curved relationships into the model.

(g)
```
> SSE <- crossprod(e,e)
> SSE
        [,1]
[1,] 6186015
> MSE <- SSE/(length(Y)-length(b))
> MSE
          [,1]
[1,] 158615.8
```

(h)
```
> Covb <- as.vector(MSE)*solve(t(X)%*%X)
> Covb
                   X0      salinity             pH
      X0  1524509.10834 -27237.9322767 -91209.1728455
salinity   -27237.93227    577.4385457   1276.4382277
      pH   -91209.17286   1276.4382280   7723.0703457
```

```
        K     -134.26839      2.5498036      3.7084102
       Na        7.40833     -0.1655202     -0.3417675
       Zn   -16373.54466    257.3248227   1104.7023266

                     K           Na              Zn
       X0 -134.268394262  7.408329862 -1.637354e+004
 salinity    2.549803557 -0.165520186  2.573248e+002
       pH    3.708410204 -0.341767495  1.104702e+003
        K    0.121317254 -0.004495673  1.000216e+000
       Na   -0.004495673  0.000253753 -8.119987e-002
       Zn    1.000215850 -0.081199866  2.266258e+002

#  Compute standard errors

> stderrb <- sqrt(diag(Covb))
> coef <- c("Intercept","Salinity", "pH", "K", "Na", "Zn")
> heading <- c("Estimate","Std. Error")
> tempb <-  cbind(b, stderrb)
> dimnames(tempb) <- list(coef, heading)
> round(tempb,4)
            Estimate Std. Error
Intercept  1252.4546  1234.7101
 Salinity   -30.2849    24.0300
       pH   305.4880    87.8810
        K    -0.2853     0.3483
       Na    -0.0087     0.0159
       Zn   -20.6768    15.0541
```

PART II

1. There is more than one way to approach the problem of deciding what linear combinations of parameters are estimable. One approach is to use the definition of estimability to directly show that $\mathbf{c}^T\boldsymbol{\beta}$ is estimable by finding a vector $\mathbf{a}$ such that $E(\mathbf{a}^T\mathbf{Y}) = \mathbf{c}^T\boldsymbol{\beta}$.To use this approach to show that $\mathbf{c}^T\boldsymbol{\beta}$ is not estimable you would have to provide an argument to show that there is no vector $\mathbf{a}$ for which $E(\mathbf{a}^T\mathbf{Y}) = \mathbf{c}^T\boldsymbol{\beta}$.

A second approach is based on Result 3.8(i) from the notes which says that $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if $\mathbf{c}^T$ is a linear combination of the rows of the model matrix. To use this approach you only have to work with the distinct rows of the model matrix. In this case, there are only six distinct rows. (In general you could work with any set of k linearly independent rows, where k is the rank of the model matrix.) Here the six distinct rows are

$$W = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

All vectors that are linear combinations of the rows of the model matrix are also linear combinations of these six rows, and all such vectors have the form

$$\begin{aligned} \mathbf{c}^T &= \mathbf{a}^T W \\ &= (a1, a2, a3, a4, a5, a6)W \\ &= (a1 + a2 + a3 + a4 + a5 + a6, a1 + a2 + a3, a4 + a5 + a6, a1 + a4, a2 + a5, a3 + a6, \\ &\quad a1, a2, a3, a4, a5, a6) \end{aligned}$$

Now you have a description of all possible $\mathbf{c}^T$ vectors that can provide an estimable function of the parameters. To show, for example, that

$$\gamma_{11} - \gamma_{13} - \gamma_{21}\gamma_{23} = (0\,0\,0\,0\,0\,0\,1\,0\ -1\ -1\,0\,1)\boldsymbol{\beta}$$

is estimable we can simply pick $a_1 = 1, a_2 = 0, a_3 = -1, a_4 = -1, a5 = 0, a6 = 1$ to get

$$\mathbf{c}^T = \mathbf{a}^T W = (0\,0\,0\,0\,0\,0\,1\,0\ -1\ -1\,0\,1)$$

To show that $\beta_1 - \beta_2$ is not estimable, we would note that we would have to have $a_1 = 0, a_2 = 0, a_3 = 0, a_4 = 0, a_5 = 0, a_6 = 0$ to prevent any γ_{ij} from appearing in the linear combination of parameters, but this would put zeros in the remaining positions of the $\mathbf{c}^T$ vector, and we are unable to produce $\beta_1 - \beta_2$

A third approach is to make use of Result 3.8 (ii) from the course notes. It may be convenient to use this result to show that a linear combination of the parameters is not estimable. This result says that $\mathbf{c}^T\boldsymbol{\beta}$ is estimable if and only if $\mathbf{c}^T\mathbf{d} = 0$ for every $\mathbf{d}$ for which $X\mathbf{d} = \mathbf{0}$. Since

$$\text{number of columns in } X - rank(X) = 12 - 6 = 6$$

you would have to find a set of 6 linearly independent $\mathbf{d}$'s that satisfy $X\mathbf{d} = \mathbf{0}$ to completely describe the linear combinations of parameters that are either estimable or not estimable. Showing that a particular linear combination of parameters is not estimable may be handled by identifying a single appropriate d vector. For this exercise, consider

$$\begin{aligned} \boldsymbol{\beta} &= (\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{22}, \gamma_{23})^T \\ \mathbf{d}_1 &= (3, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1)^T \\ \mathbf{d}_2 &= (0, 0, 0, 0, -1, 0, 0, 1, 0, 0, 1, 0)^T \end{aligned}$$

and

$$X = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Note that $X\mathbf{d}_1 = X\mathbf{d}_2 = \mathbf{0}$.

(a) Let $\mathbf{c}_1 = [1, 0, \cdots, 0]^T$. Note that $\mathbf{c}_1^T\boldsymbol{\beta} = \mu$ but $\mathbf{c}_1^T\mathbf{d}_1 \neq 0$. i.e., there exists a vector $\mathbf{d}$ such that $X\mathbf{d} = \mathbf{0}$ but $\mathbf{c}_1^T\mathbf{d}_1 \neq 0$. Thus, by Result 3.8 (ii), μ is not estimable.

(b) Let $\mathbf{c}_2 = [0, 0, 1, 0, \cdots, 0]^T$. Note that $\mathbf{c}_2^T\boldsymbol{\beta} = \alpha_2$ but $\mathbf{c}_2^T\mathbf{d}_1 \neq 0$. Thus, by Result 3.8 (ii), μ is not estimable.

(c) Let $\mathbf{c}_3 = [0, 0, 0, 0, 1, -1, 0, \cdots, 0]^T$. Note that $\mathbf{c}_3^T\boldsymbol{\beta} = \beta_2 - \beta_3$ but $\mathbf{c}_3^T\mathbf{d}_2 \neq 0$. Thus, by Result 3.9 (ii), $\beta_2 - \beta_3$ is not estimable.

(d) Let $\mathbf{c}_4 = [0, \cdots, 1]^T$. Note that $\mathbf{c}_4\boldsymbol{\beta} = \gamma_{23}$ but $\mathbf{c}_4\mathbf{d}_1 \neq 0$. Thus, by Result 3.9 (ii), γ_{23} is not estimable.

(e) Let $\mathbf{c}_5 = [1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 1]^T$ and $\mathbf{a}_5 = [0, \cdots, 0, 1]^T$. Then, we have $\mathbf{a}_5^T X = bfc_5^T$. Thus, by Result 3.9 (i), $\mu + \alpha_2 + \beta_3 + \gamma_{23}$ is estimable. $\mu + \alpha_2 + \beta_3 + \gamma_{23}$ is the mean volume when fat 2 is used with surfactant C and the OLS estimator is $\bar{y}_{23\cdot}$.

(f) Let $\mathbf{c}_6 = [0, 0, 0, 0, 0, 0, 1, -1, 0, 0, 0, 0]^T$. Then, $\mathbf{c}_6^T\boldsymbol{\beta} = \gamma_{11} - \gamma_{12}$ but $\mathbf{^T_6 d_2} \neq \mathbf{0}$. Thus, by Result 3.9 (ii), $\mathbf{c}_6^T\boldsymbol{\beta} = \gamma_{11} - \gamma_{12}$ is not estimable.

(g) Let $\mathbf{c}_7 = [0, 0, 0, 0, 0, 0, 1, 0, -1, -1, 0, 1]^T$ and
$\mathbf{a}_7 = [1/3, 1/3, 1/3, 0, 0, 0, -1/3, -1/3, -1/3, -1/3, -1/3, -1/3, 0, 0, 1/4, 1/4, 1/4, 1/4]^T$. Then, we have $\mathbf{a}^T X = \mathbf{c}_7$. Thus, by Result 3.9 (i), $\gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23}$ is estimable. $\gamma_{11} - \gamma_{13} - \gamma_{21} + \gamma_{23}$ = (difference in mean bread volumes for surfactants A and C when fat 1 is used) -(difference in mean bread volumes for surfactants A and C when fat 2 is used) This is an interaction contrast and the OLS estimator is $\bar{y}_{11\cdot} - \bar{y}_{13\cdot} - \bar{y}_{21\cdot} + \bar{y}_{23\cdot}$.

(h) Let $\mathbf{c}_8 = [0,0,0,0,1,-1,0,1,-1,0,1,-1]^T$ and $\mathbf{a}_8 = [0,0,0,1/6,1/6,1/6,-1/6,-1/6,-1/6,0,0,0,1/2,1/2,-1/4,-1/4,-1/4,-1/4]^T$. Then, we have $\mathbf{a}_8^T X = \mathbf{c}_8^T$. Thus, by Result 3.9 (i), $(\beta_2-\beta_3)+0.5(\gamma_{12}+\gamma_{22}-\gamma_{13}-\gamma_{23})$ is estimable. This is the difference between the mean bread volume when surfactant B is used and the mean bread volume when surfactant C is used, averaging across the fats giving equal weight to each fat. The OLS estimator is $0.5\sum_{i=1}^{2}\bar{y}_{i2\cdot} - 0.5\sum_{i=1}^{2}\bar{y}_{i3\cdot}$

(i) Let $\mathbf{c}_9 = [0,0,0,0,0,0,0,1,-1,0,1,-1]^T$. Note that $\mathbf{c}_9\boldsymbol{\beta} = \gamma_{12}+\gamma_{22}-\gamma_{13}-\gamma_{23}$ but $\mathbf{c}_9^T\mathbf{d}_2 \neq 0$. Thus, by Result 3.9 (ii), $\gamma_{12}+\gamma_{22}-\gamma_{13}-\gamma_{23}$ is not estimable.

2.(a)

$$X = \begin{bmatrix} 1 & 1 & 0 & 90 \\ 1 & 1 & 0 & 95 \\ 1 & 1 & 0 & 100 \\ 1 & 1 & 0 & 105 \\ 1 & 1 & 0 & 110 \\ 1 & 0 & 1 & 90 \\ 1 & 0 & 1 & 95 \\ 1 & 0 & 1 & 100 \\ 1 & 0 & 1 & 105 \\ 1 & 0 & 1 & 110 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_1 \\ \gamma \end{bmatrix}.$$

(b) Note that for the model matrix given in (a), $X\mathbf{d} = \mathbf{0}$ if and only if $\mathbf{d}^T = w[-1,1,1,0]$ for some scalar w.

	function	estimable	Explanation
(i)	μ	No	Here $\mathbf{c}^T = [1,0,0,0]$ and $w[-1111]\mathbf{c} = -w \neq 0$ for $w \neq 0$
(ii)	$\mu+\alpha_2$	Yes	use $\mathbf{a}^T = [0,0,0,0,0,10,0,-9,0,0]$
(iii)	β	Yes	use $\mathbf{a}^T = [-1/51/500000000]$
(iv)	$\alpha_1-\alpha_2$	Yes	use $\mathbf{a}^T = [10000-10000]$
(v)	$\mu+\beta T$	No	Here $\mathbf{c}^T = [100T]$ and $w[-1111]\mathbf{c} = -w+wT \neq 0$ for $w \neq 0$
(vi)	$\mu+\alpha_1+\beta(T-100)$	Yes	use $\mathbf{a}^T = [0.5-\beta(T-100)/20,0,0,0,0.5+\beta(T-100),0,0,0,0,0]$

Many students had great diculty with part (vi). You could reason in the following way to find a vector $\mathbf{a}$ such that $E(\mathbf{a}^T\mathbf{Y}) = \mu+\alpha_1+\beta(T-100)$. Since this quantity involves $\mu+\alpha_1$ and does not involve $\mu+\alpha_2$ only use observations from runs with catalyst A. To estimate the slope β you will need to use observations from runs with catalyst A at two different temperatures. You could the observations from runs with catalyst A, but we will consider the observations Y_{11} and Y_{15} from runs at 90 and 100 degrees. We need coefficients a_1 and a_5 such that

$$\mu+\alpha_1+\beta(T-100) = E(a_1Y_{11}+a_5Y_{15})$$
$$= a_1(\mu+\alpha_1+\beta(90-100)) + a_5(\mu+\alpha_1+\beta(110-100)) = (a_1+a_5)(\mu+\alpha_1)+\beta(10a_5-10a_1)$$

Consequently, we need $a_1+a_5=1$ and $10a_5-10a_1 = \beta(T-100)$. Solving these two equations, we obtain $a_1 = 0.5-\beta(T-100)/20$ and $a_5 = 0.5+\beta(T-100)/20$.

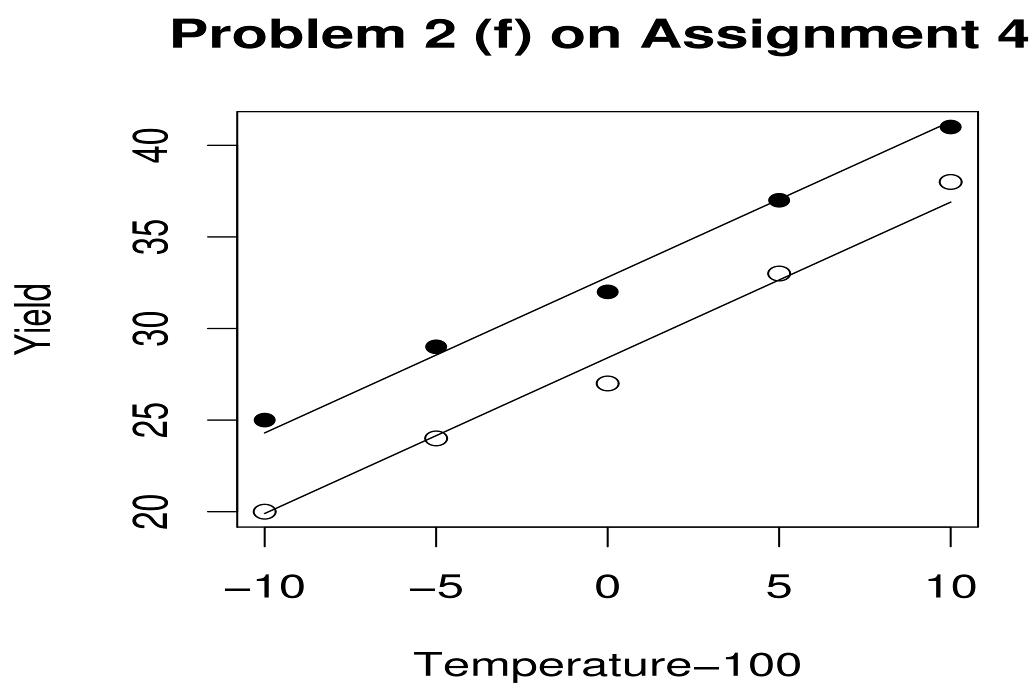
(c) The generalized inverse is shown in the solution to part (d).

(d)
```
> W <- read.table("c:/stat504/hw4p2.txt",header= T)
> Y <- as.matrix(W[ ,2 ])
> X <- as.matrix(cbind(rep(1,length(Y)),W[ ,3:5]))
> X
      rep(1, length(Y)) A B temp
 [1,]                 1 1 0  -10
 [2,]                 1 1 0   -5
 [3,]                 1 1 0    0
 [4,]                 1 1 0    5
 [5,]                 1 1 0   10
 [6,]                 1 0 1  -10
 [7,]                 1 0 1   -5
 [8,]                 1 0 1    0
 [9,]                 1 0 1    5
[10,]                 1 0 1   10

> library(MASS)
>
> A <- t(X)%*%X
> G <- ginv(A)
> G
           [,1]        [,2]        [,3]  [,4]
[1,] 0.04444444  0.02222222  0.02222222 0.000
[2,] 0.02222222  0.11111111 -0.08888889 0.000
[3,] 0.02222222 -0.08888889  0.11111111 0.000
[4,] 0.00000000  0.00000000  0.00000000 0.002

> # (i) AGA = A
> round( A %*% G %*% A - A, 5 )
                  rep(1, length(Y)) A B temp
rep(1, length(Y))                 0 0 0    0
A                                 0 0 0    0
B                                 0 0 0    0
temp                              0 0 0    0
>
> # (ii) GAG = G
> round( G %*% A %*% G - G, 5 )
     [,1] [,2] [,3] [,4]
[1,]    0    0    0    0
```

```
[2,]    0    0    0    0
[3,]    0    0    0    0
[4,]    0    0    0    0
>
> # (iii) (AG)^T = AG
> round( t(A %*% G) - A %*% G , 5 )
     rep(1, length(Y)) A B temp
[1,]                 0 0 0    0
[2,]                 0 0 0    0
[3,]                 0 0 0    0
[4,]                 0 0 0    0
>
> # (iv) (GA)^T = GA
> round( t(G %*% A) - G %*% A , 5 )
                  [,1] [,2] [,3] [,4]
rep(1, length(Y))    0    0    0    0
A                    0    0    0    0
B                    0    0    0    0
temp                 0    0    0    0
>
```

Thus, the $(X^T X)^{-1}$ generated from ginv() in R satisfies the four properties of the Moore- Penrose inverse.

(e)
```
> b <- ginv( t(X) %*% X ) %*% t(X) %*% Y
> b
      [,1]
[1,] 20.40
[2,]  8.00
[3,] 12.40
[4,]  0.85
```



Based on the figure above, we can see that the observed data conform nicely to the two parallel lines corresponding to catalysts A and B (Same $\hat{\gamma} = .55$). Consequently, the observed data seem to agree with the proposed model. You could also examine residual plots.

(g) No. Note that the projection matrix $P_X = X(X^TX)^- X^T$ is invariant to the choice of generalized inverse $(X^TX)^-$. Then, the estimates for the mean yield $\hat{\mathbf{Y}} = P_X\mathbf{Y}$ are also invariant and thus unchanged even if different solutions to the normal equations are used.

3.(a)

$$\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T\mathbf{Y} = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{5} Y_{ij} = \bar{Y}_{\cdot\cdot}$$

(b)

$$P_{\mathbf{1}}P_{\mathbf{1}} = \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T\mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T = \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T = P_{\mathbf{1}}$$

(c)

$$\hat{\mathbf{Y}} = P_{\mathbf{1}}\mathbf{Y} = \mathbf{1}\left(\mathbf{1}^T\mathbf{1}\right)^{-1}\mathbf{1}^T\mathbf{Y} = \mathbf{1}\bar{Y}_{\cdot\cdot}$$

(d)

$$
\begin{aligned}
\mathbf{Y}^T\mathbf{Y} &= \mathbf{Y}^T(P_{\mathbf{1}} + I - P_{\mathbf{1}})\mathbf{Y} \\
&= \mathbf{Y}^T P_{\mathbf{1}}\mathbf{Y} + (I - P_{\mathbf{1}})\mathbf{Y} \\
&= \mathbf{Y}^T P_{\mathbf{1}} P_{\mathbf{1}}\mathbf{Y} + \mathbf{Y}^T(I - P_{\mathbf{1}})(I - P_{\mathbf{1}})\mathbf{Y} \\
&= (P_{\mathbf{1}}\mathbf{Y})^T P_{\mathbf{1}}\mathbf{Y} + [(I - P_{\mathbf{1}})\mathbf{Y}]^T(I - P_{\mathbf{1}})\mathbf{Y} \\
&= \hat{\mathbf{Y}}^T\hat{\mathbf{Y}} + (\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{5}\hat{Y}_{ij}^2 + \sum_{i=1}^{2}\sum_{j-1}^{5}(Y_{ij} - \hat{Y}_{ij})^2 \\
&= SS_{model,uncorrected} + SS_{residuals}.
\end{aligned}
$$

(e)

$$
\begin{aligned}
\mathbf{Y}^T P_{\mathbf{1}}\mathbf{Y} &= \mathbf{Y}^T P_{\mathbf{1}} P_{\mathbf{1}}\mathbf{Y} \\
&= (P_{\mathbf{1}}\mathbf{Y})^T P_{\mathbf{1}}\mathbf{Y} \\
&= (\mathbf{1}\bar{Y}_{\cdot\cdot})^T\mathbf{1}\bar{Y}_{\cdot\cdot} \\
&= n\bar{Y}_{\cdot\cdot}^2.
\end{aligned}
$$

(f)

$$
\begin{aligned}
\mathbf{Y}^T\mathbf{Y} &= \mathbf{Y}^T(P_{\mathbf{1}} + P_X - P_{\mathbf{1}} + I - P_X)\mathbf{Y} \\
&= \mathbf{Y}^T P_{\mathbf{1}}\mathbf{Y} + \mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y} + \mathbf{Y}^T(I - P_X)\mathbf{Y}
\end{aligned}
$$

(g)

$$
\begin{aligned}
\mathbf{Y}^T(P_X - P_{\mathbf{1}})\mathbf{Y} &= \mathbf{Y}^T(I - P_{\mathbf{1}} + P_X - I)\mathbf{Y} \\
&= \mathbf{Y}^T(I - P_{\mathbf{1}})\mathbf{Y} - \mathbf{Y}^T(I - P_X)\mathbf{Y} \\
&= [(I - P_{\mathbf{1}})\mathbf{Y}]^T(I - P_{\mathbf{1}})\mathbf{Y} - [(I - P_X)\mathbf{Y}]^T(I - P_X)\mathbf{Y} \\
&= (\mathbf{Y} - \mathbf{1}\bar{Y}_{\cdot\cdot})^T(\mathbf{Y} - \mathbf{1}\bar{Y}_{\cdot\cdot}) - (\mathbf{Y} - \hat{\mathbf{Y}})^T(\mathbf{Y} - \hat{\mathbf{Y}}) \\
&= \sum_{i=1}^{2}\sum_{j=1}^{5}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2 - \sum_{i=1}^{2}\sum_{j-1}^{5}(Y_{ij} - \hat{Y}_{ij})^2 \\
&= SS_{residuals,\ common\ mean\ model} + SS_{residuals,\ problem\ 2}.
\end{aligned}
$$