

# Hwang\_Hw6

Charles Hwang

12/8/2021

## Problem 10.6

(a)

```
rm(list=ls())
y<-c(193,230,172,91,113,125)
x1<-c(1.6,15.5,22,43,33,40)
x2<-c(851,816,1058,1201,1357,1115)
w<-lm(y~x1+x2)
```

Model:  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$

$y$ : response variable (wear of bearing)

$\beta_0$ : intercept term

$\beta_1$ : intercept term for  $x_1$

$x_1$ : coded variable for oil viscosity

$\beta_2$ : intercept term for  $x_2$

$x_2$ : coded variable for load

$\epsilon$ : error term

(b)

```
anova(w)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: y
```

```
##          Df  Sum Sq Mean Sq F value  Pr(>F)
## x1         1 10240.4 10240.4 15.7510 0.02859 *
## x2         1  1921.2   1921.2  2.9551 0.18410
## Residuals  3   1950.4    650.1
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# We reject the null hypothesis at the alpha = 0.05 level for the x_1 variable and fail
# to reject the null hypothesis at the alpha = 0.05 level for the x_2 variable. There is
# sufficient evidence (p = 0.02859) that the x_1 variable is significant in the model and
# insufficient evidence (p = 0.1841) that the x_2 variable is significant in the model.
summary(w)
```

```
##
```

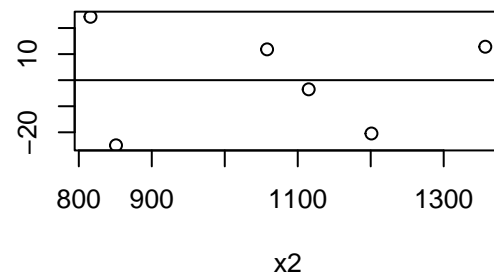
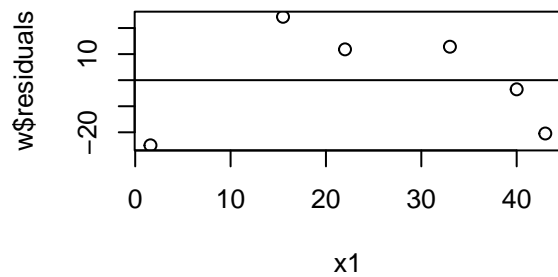
```
## Call:
```

```
## lm(formula = y ~ x1 + x2)
```

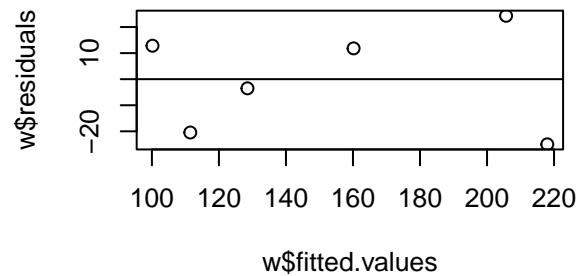
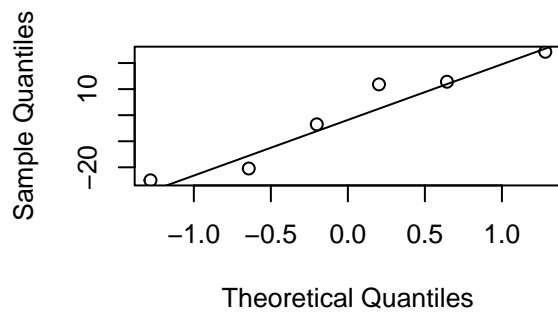
```
##
## Residuals:
##      1      2      3      4      5      6
## -24.987  24.307  11.820 -20.460  12.830  -3.511
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 350.99427   74.75307   4.695  0.0183 *
## x1          -1.27199    1.16914  -1.088  0.3562
## x2          -0.15390    0.08953  -1.719  0.1841
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.5 on 3 degrees of freedom
## Multiple R-squared:  0.8618, Adjusted R-squared:  0.7696
## F-statistic: 9.353 on 2 and 3 DF,  p-value: 0.05138
cat("We can see from the summary that adjusted-r^2 = ",summary(w)$adj.r.squared,". Approximately ",100*

## We can see from the summary that adjusted-r^2 = 0.7696497. Approximately 76.96497 percent
## of the variation in the data can be explained by the model.
shapiro.test(w$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  w$residuals
## W = 0.9214, p-value = 0.5155
# The null hypothesis was not rejected (p = 0.5155) at the alpha = 0.05 level, so the
par(mfrow=c(2,2)) # normality assumption appears to be met.
plot(x1,w$residuals)
abline(h=0)
plot(x2,w$residuals,ylab="")
abline(h=0)
qqnorm(w$residuals)
qqline(w$residuals)
plot(w$fitted.values,w$residuals)
abline(h=0)
```



### Normal Q-Q Plot



*# There appears to be a slight variation in the Q-Q plot. However, it is hard to accurately analyze and interpret the plots due to the small sample size.*

(c)

```
summary(w)$coefficients[,3:4]
```

```
##           t value   Pr(>|t|)
## (Intercept)  4.695382 0.01826948
## x1          -1.087974 0.35620034
## x2          -1.719030 0.18410102
```

We can see the intercept term is significant ( $p = 0.018269$ ) at the  $\alpha = 0.05$  level, but the two regression terms  $x_1$  and  $x_2$  are not ( $p = 0.3562$ ,  $p = 0.184101$ ). This could indicate that the model is not significant.

### Problem 10.12

```
rm(list=ls())
y<-c(26,24,175,160,163,55,62,100,26,30,70,71)
x1<-c(1,1,1.5,1.5,1.5,0.5,1.5,0.5,1,0.5,1,0.5)
x2<-c(1,1,4,4,4,2,2,3,1.5,1.5,2.5,2.5)
x1x1<-x1^2
x2x2<-x2^2
m<-lm(y~x1*x2+x1x1+x2x2)
anova(m)

## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## x1           1 11552.0 11552.0 316.3971 2.029e-06 ***
## x2           1 22950.3 22950.3 628.5848 2.651e-07 ***
## x1x1          1    21.9    21.9   0.5992 0.468283
```

```
## x2x2      1    520.8    520.8  14.2634  0.009216 **
## x1:x2      1     47.6     47.6   1.3048  0.296857
## Residuals  6    219.1    36.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# We reject the null hypothesis at the alpha = 0.05 level for the x_1, x_2, and x_2^2
# variables and fail to reject the null hypothesis at the alpha = 0.05 level for the
# x_1^2 variable and the interaction term (x_1 * x_2). There is sufficient evidence
# (p = 0.000002, p < 0.000001, p = 0.009216) that the x_1, x_2, and x_2^2 variables are
# significant in the model and insufficient evidence (p = 0.468283, p = 0.296857) that
summary(m) # the x_1^2 variable and the interaction term are significant in the model.

##
## Call:
## lm(formula = y ~ x1 * x2 + x1x1 + x2x2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.3504 -2.7354 -0.3533  2.7022  8.9329
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    24.410     26.592   0.918  0.3941
## x1             -38.033     40.452  -0.940  0.3834
## x2               0.720     11.687   0.062  0.9529
## x1x1            34.975     21.556   1.623  0.1558
## x2x2            11.066      3.158   3.504  0.0128 *
## x1:x2           -9.986      8.742  -1.142  0.2969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.042 on 6 degrees of freedom
## Multiple R-squared:  0.9938, Adjusted R-squared:  0.9886
## F-statistic: 192.2 on 5 and 6 DF,  p-value: 1.556e-06

cat("We can see from the summary that adjusted-r^2 = ",summary(m)$adj.r.squared,". Approximately ",100*

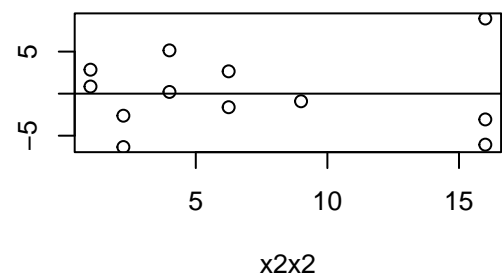
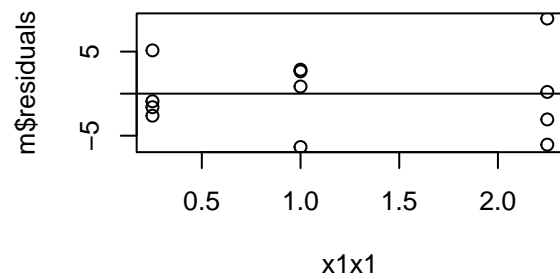
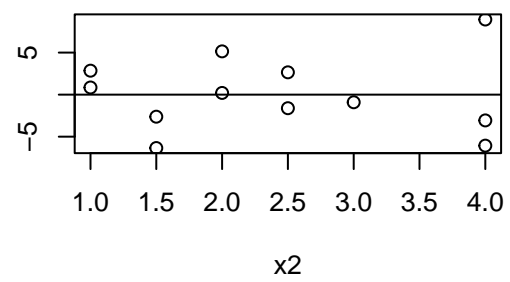
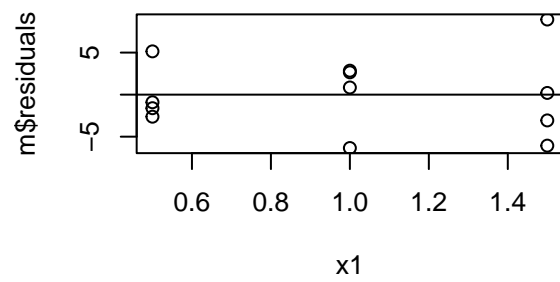
## We can see from the summary that adjusted-r^2 = 0.9886264. Approximately 98.86264 percent
## of the variation in the data can be explained by the model.

shapiro.test(m$residuals)

##
## Shapiro-Wilk normality test
##
## data:  m$residuals
## W = 0.97082, p-value = 0.9192

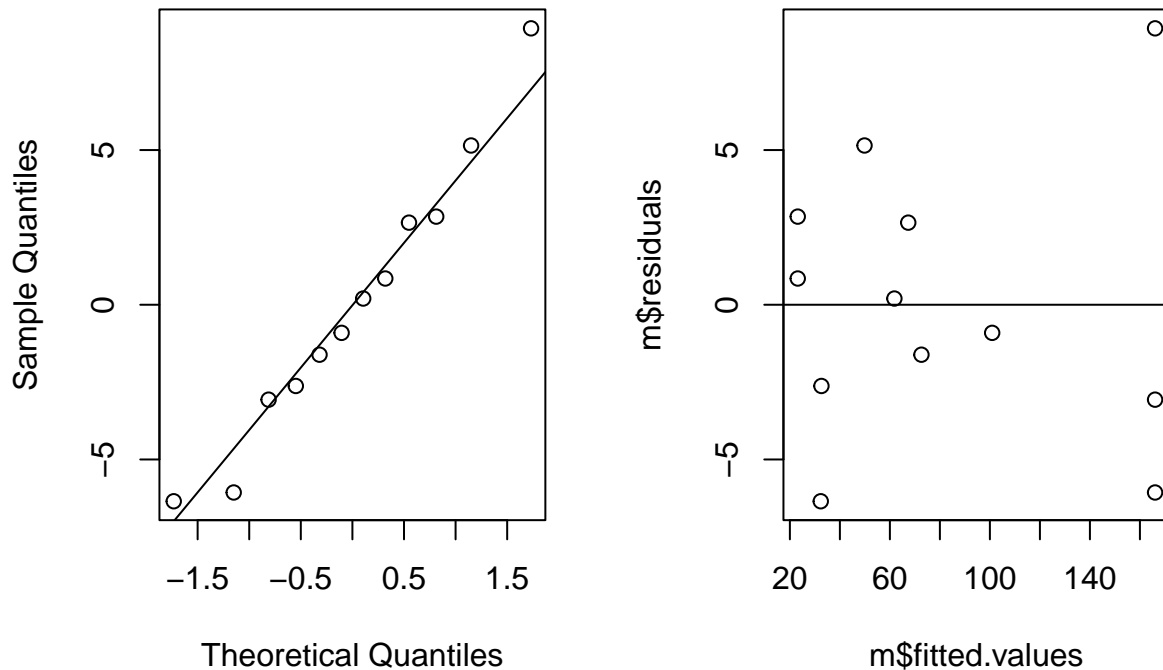
# The null hypothesis was not rejected (p = 0.9192) at the alpha = 0.05 level, so the
par(mfrow=c(2,2)) # normality assumption appears to be met.
plot(x1,m$residuals)
abline(h=0)
plot(x2,m$residuals,ylab="")
abline(h=0)
plot(x1x1,m$residuals)
abline(h=0)
```

```
plot(x2x2,m$residuals,ylab="")
abline(h=0)
```



```
par(mfrow=c(1,2))
qqnorm(m$residuals)
qqline(m$residuals)
plot(m$fitted.values,m$residuals)
abline(h=0)
```

## Normal Q-Q Plot



*# There does not appear to be any variation in the Q-Q plot or residuals vs. fitted plot.*

### Problem 15.15

```
rm(list=ls())
y<-c(46.5,45.9,49.8,46.1,44.3,48.7,49,50.1,48.5,45.2,46.3,47.1,48.9,48.2,50.3,44.7,43,51,48.1,48.6)
x<-c(13,14,12,12,14,12,10,11,12,14,15,14,11,11,10,16,15,10,12,11)
X<-x-mean(x)
glue<-as.factor(rep(1:4,each=5))
g<-aov(y~X+glue)
library(car)
Anova(g,type="III")
```

```
## Anova Table (Type III tests)
##
## Response: y
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 10800.8  1 7728.7307 < 2.2e-16 ***
## X           59.6   1   42.6236 9.543e-06 ***
## glue        1.8   3    0.4224  0.7397
## Residuals   21.0 15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

*# We reject the null hypothesis at the  $\alpha = 0.05$  level for the X variable and fail to reject the null hypothesis at the  $\alpha = 0.05$  level for the glue variable. There is sufficient evidence ( $p < 0.000001$ ) that the X variable is significant in the model and insufficient evidence ( $p = 0.7397$ ) that the glue variable is significant in the model.*

```
ybar.adj<-vector(length=4)
se_adj.mean<-vector(length=4)
```

```

for(i in 1:4)
{
  pred<-predict(lm(y~X+glue),new=data.frame(glue=as.factor(i),X=0),se.fit=T)
  ybar.adj[i]<-pred$fit
  se_adj.mean[i]<-pred$se.fit
}
tau_hat<-ybar.adj-mean(y)
prmatrix(cbind(tau_hat,ybar.adj,se_adj.mean))

##          tau_hat ybar.adj se_adj.mean
## [1,] -0.43953767 47.07546  0.5354766
## [2,]  0.12854452 47.64354  0.5381512
## [3,]  0.39251712 47.90752  0.5300869
## [4,] -0.08152397 47.43348  0.5314395

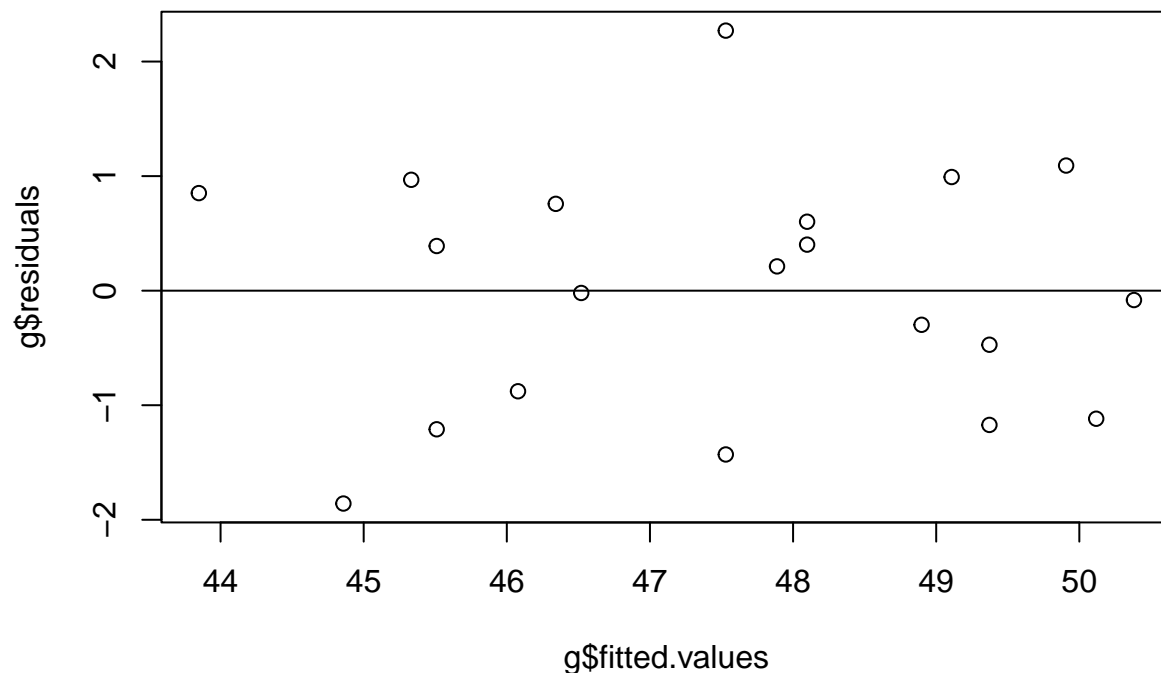
shapiro.test(g$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  g$residuals
## W = 0.96909, p-value = 0.7357

bartlett.test(g$residuals~glue)

##
##  Bartlett test of homogeneity of variances
##
## data:  g$residuals by glue
## Bartlett's K-squared = 1.2564, df = 3, p-value = 0.7395
# Neither of the null hypotheses were rejected, so the normality and equal
plot(g$fitted.values,g$residuals) # variance assumptions appear to be met.
abline(h=0)

```



```
# There does not appear to be any pattern in the residuals vs. fitted plot.
```

## Problem 15.17

```
rm(list=ls())
y<-c(68,90,98,77,88,112,94,65,74,85,118,82,73,92,80)
x<-c(120,140,150,125,136,165,140,120,125,133,175,132,124,141,130)
X<-x-mean(x)
cs<-as.factor(rep(1:3,each=5)) # The factors have to be set as 1-3 for the function to work
c<-aov(y~X+cs)
Anova(c,type="III")

## Anova Table (Type III tests)
##
## Response: y
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 37296  1 4298.0823 1.289e-15 ***
## X           3019  1  347.9561 1.127e-09 ***
## cs              2  2    0.1385  0.8721
## Residuals    95 11
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
# We reject the null hypothesis at the alpha = 0.05 level for the X variable and fail
# to reject the null hypothesis at the alpha = 0.05 level for the cutting speed
# variable. There is sufficient evidence (p < 0.000001) that the X variable is
# significant in the model and insufficient evidence (p = 0.8721) that the cutting
ybar.adj<-vector(length=3) # speed variable is significant in the model.
se_adj.mean<-vector(length=3)
for(i in 1:3)
{
  pred<-predict(lm(y~X+cs),new=data.frame(cs=as.factor(i),X=0),se.fit=T)
  ybar.adj[i]<-pred$fit
  se_adj.mean[i]<-pred$se.fit
}
tau_hat<-ybar.adj-mean(y)
prmatrix(cbind(tau_hat,ybar.adj,se_adj.mean))

##           tau_hat ybar.adj se_adj.mean
## [1,]  0.47821847 86.87822  1.325176
## [2,]  0.03598905 86.43599  1.317582
## [3,] -0.51420752 85.88579  1.327912

shapiro.test(c$residuals)

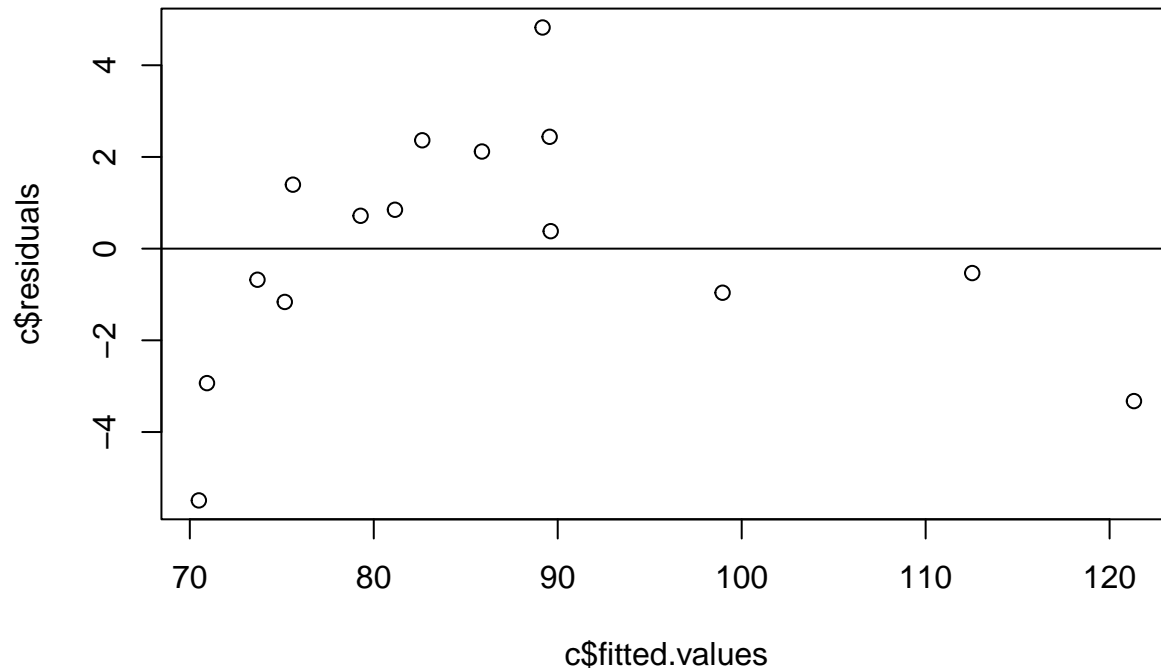
##
## Shapiro-Wilk normality test
##
## data:  c$residuals
## W = 0.97817, p-value = 0.9555

bartlett.test(c$residuals~cs)

##
## Bartlett test of homogeneity of variances
##
```



```
## data: c$residuals by cs
## Bartlett's K-squared = 2.0415, df = 2, p-value = 0.3603
# Neither of the null hypotheses were rejected, so the normality and equal
plot(c$fitted.values,c$residuals) # variance assumptions appear to be met.
abline(h=0)
```



*# There appears to be a slight negative quadratic pattern in the residuals vs. fitted plot.*

## Problem 10.9

(a)

```
rm(list=ls())
y<-c(81,89,83,91,79,87,84,90)
conc<-c(1,1,2,2,1,1,2,2) # This is the same as rep(1:2,2,each=2)
temp<-c(150,180,150,180,150,180,150,180) # This is the same as rep(c(150,180),4)
X<-matrix(c(conc,temp),ncol=2)
t(X)%*%X # X'X via matrix multiplication
```

```
##      [,1]  [,2]
## [1,]   20  1980
## [2,] 1980 219600
```

(b)

No, the matrix obtained in Problem 10.9(a) is not diagonal. We can see the  $(\mathbf{X}'\mathbf{X})_{ii}$  terms of this matrix,  $(\mathbf{X}'\mathbf{X})_{11}$  and  $(\mathbf{X}'\mathbf{X})_{22}$ , are not the only terms with nonzero values.

(c)

```
x1c<-(conc-1.5)/0.5
x2c<-(temp-165)/15
Xc<-matrix(c(x1c,x2c),ncol=2)
t(Xc)%*%Xc # X'X via matrix multiplication
```

```
##      [,1] [,2]
## [1,]    8    0
## [2,]    0    8
```

Yes, this matrix is diagonal. We can see the  $(\mathbf{X}'\mathbf{X})_{ii}$  terms of this matrix are the only terms with nonzero values. Since the concentration and temperature variables only have two levels each (levels 1 and 2 for concentration and levels 150 and 180 for temperature), we can standardize them such that the levels are instead set as -1 and 1. This way, when we multiply  $\mathbf{X}'$  and  $\mathbf{X}$  together, the resulting  $\mathbf{X}'\mathbf{X}$  matrix is diagonal.

(d)

```
x1d<-(conc-1)/1
x2d<-(temp-150)/30
Xd<-matrix(c(x1d,x2d),ncol=2)
t(Xd)%*%Xd # X'X via matrix multiplication
```

```
##      [,1] [,2]
## [1,]    4    2
## [2,]    2    4
```

No, this matrix is not diagonal. We can see the  $(\mathbf{X}'\mathbf{X})_{ii}$  terms of this matrix are not the only terms with nonzero values. The concentration and temperature variables were incorrectly standardized such that the levels are set as 0 and 1. Because 1 has a different numerical weight than 0 (i.e.,  $|1| \neq |0|$ ), the resulting  $\mathbf{X}'\mathbf{X}$  matrix is not diagonal.

(e)

We must exercise caution in properly coding independent variables in regression so they do not potentially have unintentional effects on the model, as evidenced in Problem 10.9(d).