Alexander Martinez

Hw 2 Ch 3-5

3.4

c. Done in R

The only thing I find noteworthy is that the residuals are roughly closer to ∅ over the outliers.

d. Graphs done in R. These Graphs look exactly the Same. There seems to be no notable pattern, which is ideal. We can study departures such as whether or not

1. The Regression Function isnt linear
2. The error terms dont have constant variance
3. The error terms arent indep.
4. The model fits all but a few outliers
5. The error terms are not normall dist.
6. One or several important predictor variables have been omitted from the model)

with these plots.

e. The Graph was created in R. The Graph seems to be a straight, linear line, which indicates little to no departure from normality

g. BF Test p-value = .8699 which is greater than .05.

ie. .8699 > .05 → No significant departure from constant σ^2. (Done in R

3.5

c. The stem plot shows the residuals being pretty evenly distributed. There seems to be no crazy outliers

d. There seems to be a symmetrical pattern around the ∅ horizontal line, which means that a linear model may be inappropriate here (Done in R)

e. It doe seem to follow a line, which indicates no departure from normality. I do want to mention how almost all points have the same exact Y_i. It may be b/c the routes are 0-3, so there will be repeats (Done in R)

g. P-value = .3052 > .10. No significant departure from Constant Variance (Done in R)

3.13

a. LOF Test tells us whether a regression model is a poor model for the data. It may be b/c a poor choice of variables, or it may be b/c important terms were left out. It can also be b/c of poor experimental Design.

b. (Done in R)

$H_0: E(Y) = \beta_0 + \beta_1 X$

$H_a: E(Y) \neq \beta_0 + \beta_1 X$

Pval = .4766 > .05, there is no significant LOF.

3.17

a. Done in R
A linear relationship seems adequate.

b. Done in R
The Transformation Suggested is when $\lambda = .5$

c. Done in R, Transformed SLR is

$$\hat{Y}' = 10.261 + 1.076X$$

d. Yes, the linear model seems more appropriate than before. (Done in R)

e. The residual plot has no patterns, so it seems to not depart from anything. It is one cloud of randomness.

The NPP, on the other hand, seems to be somewhat linear. A bigger sample size would be helpful in concluding whether the linear model is appropriate.

f. $\hat{Y} = 91.56 + 32.5X$

4.3

b. $s\{b_0\} = 2.8083941$ (Done in R)

$s\{b_1\} = .4831$

$b_0 = -.5802$

$b_1 = 15.0352$

$B = t(1-(.05/4), 43) = 2.32262$

$$-7.092642 \leq \beta_0 \leq 5.93238$$

$$13.913 \leq \beta_1 \leq 16.157$$

c. Yes, 0 is in the β_0 CI
14.0 is in the β_1 CI

4.4

b. Found in R, Summary(linear Model)

$b_0 = 10.2$

$b_1 = 4$

$s\{b_0\} = .6633$

$s\{b_1\} = .4690$

$B = 3.832519 = t(1-(\frac{.01}{4}), 8)$

$7.65789 \leq \beta_0 \leq 12.74211$

$2.202549 \leq \beta_1 \leq 5.797451$

4.7

a. $\hat{Y} = -.5802 + 15.0352$

$X_h = 3$: 44.5256, $s\{\hat{Y}_h\} = 1.675012$

$X_h = 5$: 74.5961, $s\{\hat{Y}_h\} = 1.329831$

$X_h = 7$: 104.667, $s\{\hat{Y}_h\} = 1.6119$

In R, $W = 2.204725$

$X_h = 3$: $44.5256 \pm 2.204725(1.675012)$

$40.83266 \leq E\{Y_h\} \leq 48.21854$

$X_h = 5$: $74.5961 \pm 2.204725(1.329831)$

$71.66419 \leq E\{Y_h\} \leq 77.528$

$X_h = 7$: $104.667 \pm W(1.6119)$

$101.1132 \leq E\{Y_h\} \leq 108.22096$

4.8 All Math Done in R

a. $X_h = 0$; $16.2 \pm W(.6633)$
$X_h = 1$; $14.2 \pm W(.4690)$
$X_h = 2$; $18.2 \pm W(.6633)$

$W = 2.986292$

$X_h=0$: $8.219192 \leq E\{Y_h\} \leq 12.181$
$X_h=1$: $12.79943 \leq E\{Y_h\} \leq 15.60057$
$X_h=2$: $16.21919 \leq E\{Y_h\} \leq 20.18081$

b. $B = 3.015762$

Since $W < B$, we would be more efficient using W.

4.11 Sometimes regression in the origin makes sense. Lets say you are trying to predict the number of pages someone reads a minute. Lets say Y is # of pages read and B is the coeff representing the average pages per X minute. Let $B = .55$

$Y = .55X + \epsilon$ in the SLR. When read 0 minutes, the Model $= \emptyset$. This makes sense in real life. No pages can be read in $\emptyset$ Minutes.

4.16

a. $Y = 14.9472X$

b. $t(.95, 44) = 1.68023$

$s\{b_1\} = .226424$

$14.972 \pm 1.68023(.226424)$

$14.567 \leq \beta_1 \leq 15.328$

c. $\hat{Y}_h = 14.9472(6) = 89.6834$

$s\{pred\} = 8.92008$

$89.6834 \pm 1.68023(8.92008)$

$74.46433 \leq Y_{h(new)} \leq 104.7983$

4.17

a. Yes, it does look like a good fit (Graph done in R

b. Done in R, they dont add to $\emptyset$.
The Residual plot also done in R
Compared to the original plot, the residuals moved ever so slightly. Still no obvious patterns.

c. All Done in R

Pval = .5644 > .01, No significant Lack of fit

$H_0: E\{Y\} = \beta_1 X$ $F^* = .4648 \leq 2.96301$

$H_a: E\{Y\} \neq \beta_1 X$

Scanned with CamScanner

4.22

$P(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) \geq 1-\alpha-\alpha-\alpha = 1-3\alpha$

5.3

(2) $\sum x_i e_i$

$X = \begin{bmatrix} 1 & x_1 \\ \vdots & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} [e_1\ e_2 \ldots e_n] = 0$

5.4 (1) 503.77

(2) $\begin{bmatrix} 5 & 0 \\ 0 & 160 \end{bmatrix}$

(3) $\begin{bmatrix} 49.7 \\ -39.2 \end{bmatrix}$

Done in R

5.6

(1) 2194

(2) $\begin{bmatrix} 10 & 10 \\ 10 & 20 \end{bmatrix}$

(3) $\begin{bmatrix} 142 \\ 182 \end{bmatrix}$

5.12 $\begin{bmatrix} .02 & 0 \\ 0 & .00625 \end{bmatrix}$

5.19 $3Y_1^2 + 10Y_1Y_2 + 17Y_2^2$

$[Y_1 \quad Y_2] \begin{bmatrix} 3 & 5 \\ 5 & 17 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$

5.23 a. (1) $\begin{bmatrix} 9.94 \\ -.245 \end{bmatrix}$ in R

(2) $\begin{bmatrix} -.18 \\ .04 \\ .26 \\ .08 \\ -.2 \end{bmatrix}$ in R

(3) 9.604 in R

(4) .148 in R

(5) $MSE(X'X)^{-1} = \begin{bmatrix} .00592 & 0 \\ 0 & .000185 \end{bmatrix}$ in R

(6) 11.41 Using R

(7) .02097 Using R

b. Everything felt simpler, such as the answers being clear and concise. The temperature spacing showed when exactly things begin to deteriorate

C. Done in R

$$H = \begin{bmatrix} .6 & .4 & .2 & 0 & -.2 \\ .4 & .3 & .2 & .1 & 0 \\ .2 & .2 & .2 & .2 & .2 \\ 0 & .1 & .2 & .3 & .4 \\ -.2 & 0 & .2 & .4 & .6 \end{bmatrix}$$

5.25a.

(1) in R

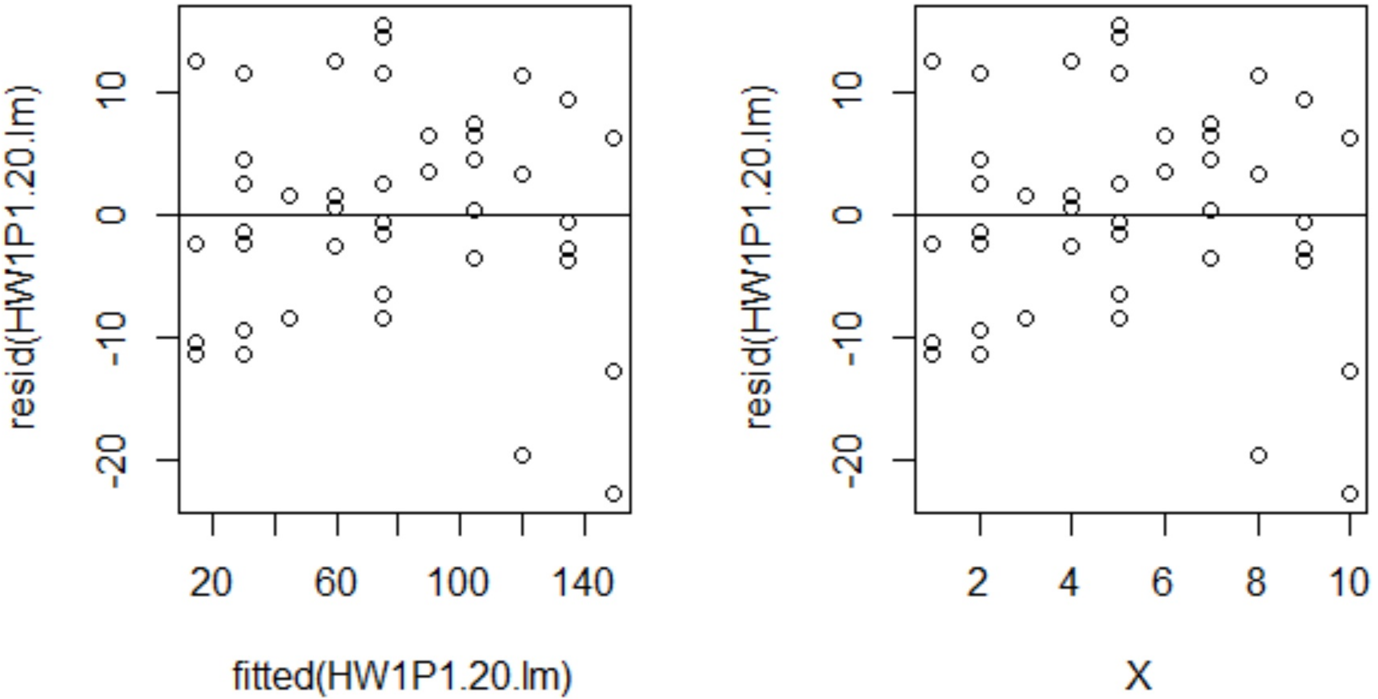
$$\begin{bmatrix} .2 & -.1 \\ -.1 & .1 \end{bmatrix}$$

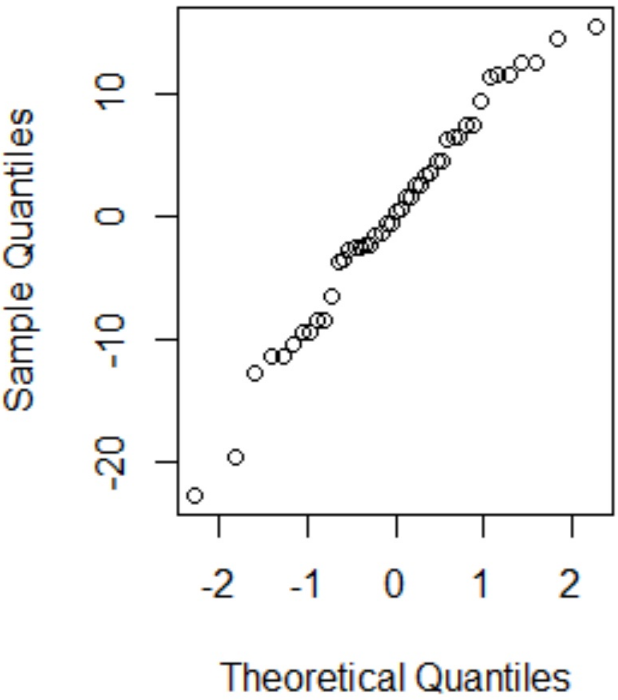
(2) $\begin{bmatrix} 10.2 \\ 4 \end{bmatrix}$

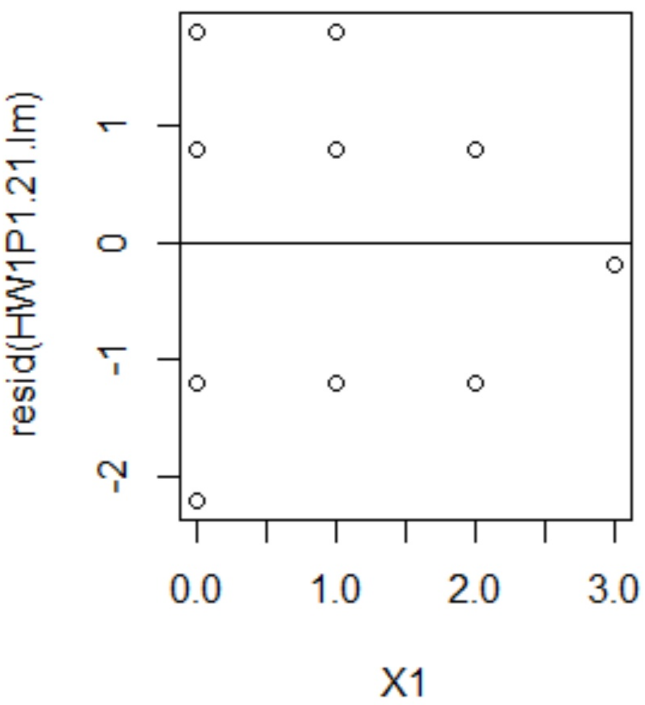
(3) $\begin{bmatrix} 1.8 \\ -1.2 \\ -1.2 \\ 1.8 \\ -.2 \\ -1.2 \\ .8 \\ .8 \\ .8 \end{bmatrix}$

(4) Large Matrix, Done in R

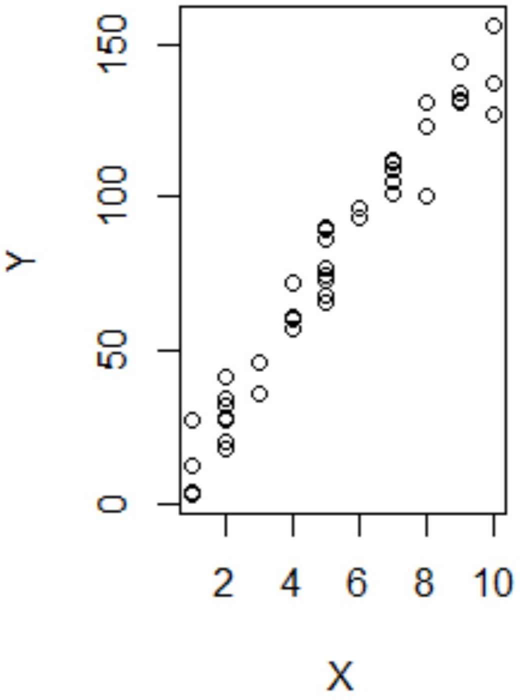
(5) 17.6, Done in R


resid(HW1P1.20.lm)
10
0
-10
-20
20
60
100
140
fitted(HW1P1.20.lm)
resid(HW1P1.20.lm)
10
0
-10
-20
2
4
6
8
10
X

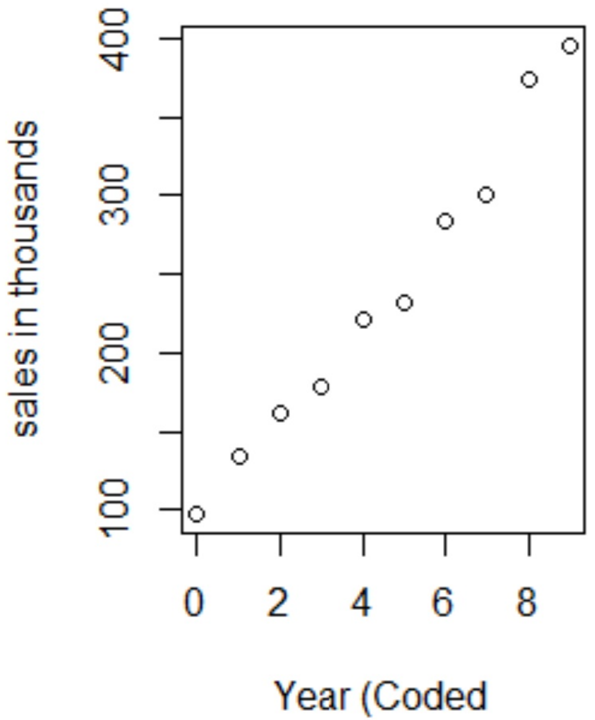

Sample Quantiles
Theoretical Quantiles
-20
-10
0
10
-2
-1
0
1
2



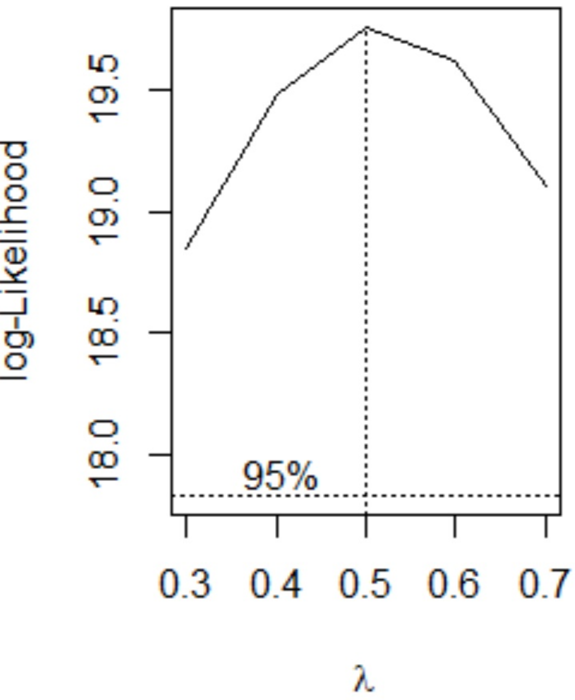
resid(HW1P1.21.lm)
1
0
-1
-2
0.0
1.0
2.0
3.0
X1



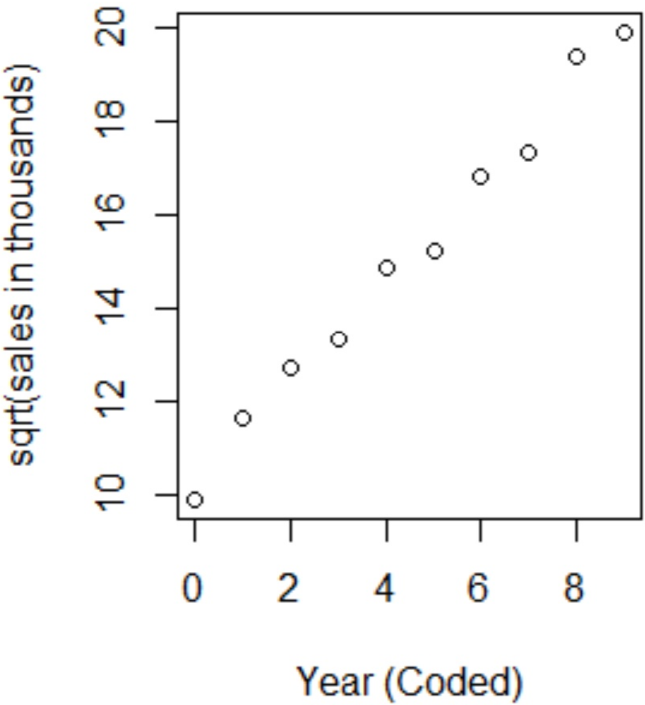
Y
0
50
100
150
X
2
4
6
8
10

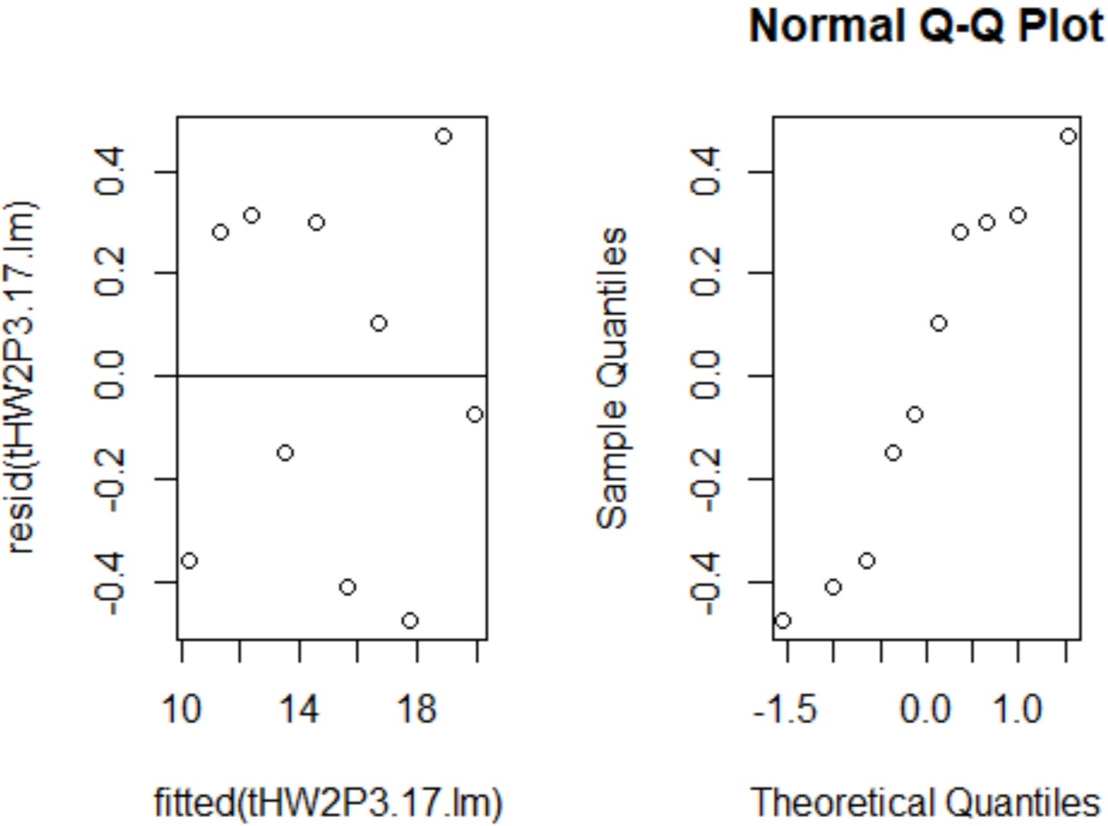


sales in thousands
100
200
300
400
0
2
4
6
8
Year (Coded

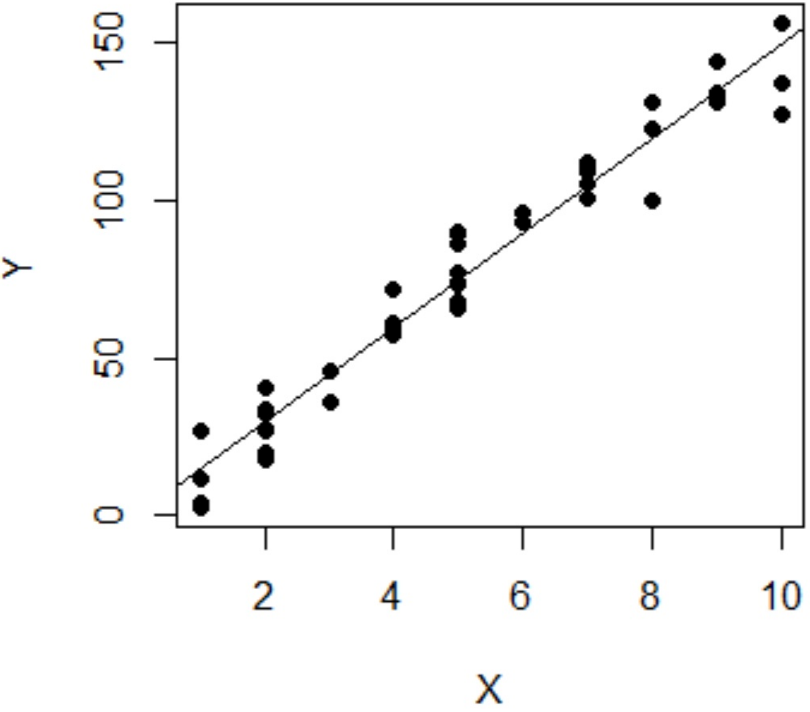


log-Likelihood
19.5
19.0
18.5
18.0
95%
0.3
0.4
0.5
0.6
0.7
λ

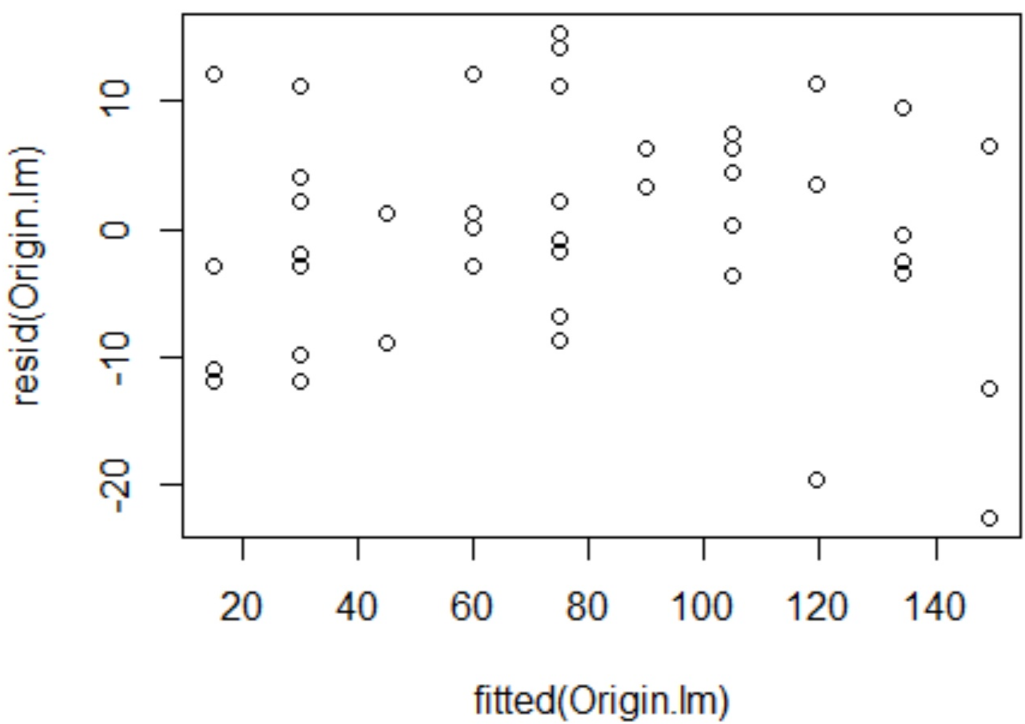

sqrt(sales in thousands)
10
12
14
16
18
20
0
2
4
6
8
Year (Coded)



Normal Q-Q Plot
resid(tHW2P3.17.lm)
fitted(tHW2P3.17.lm)
0.4
0.2
0.0
-0.2
-0.4
10
14
18
Sample Quantiles
Theoretical Quantiles
-1.5
0.0
1.0



Y
0
50
100
150
2
4
6
8
10
X


resid(Origin.lm)
fitted(Origin.lm)
10
0
-10
-20
20
40
60
80
100
120
140

```
R version 4.1.0 (2021-05-18) -- "Camp Pontanezen"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> ##########################################################################################
> ###3.4
> #c.
> ##getting data for the stem plot
> X <- c(2,4,3,2,1,10,5,5,1,2,9,10,6,3,4,8,7,8,10,4,5,7,7,5,
+       9,7,2,5,7,6,8,5,2,2,1,4,5,9,7,1,9,2,2,4,5)
> length(X)
[1] 45
> Y <- c(20,60,46,41,12,137,68,89,4,32,144,156,93,36,72,100,105,
+       131,127,57,66,101,109,74,134,112,18,73,111,96,123,90,20,
+       28,3,57,86,132,112,27,131,34,27,61,77)
> length(Y)
[1] 45
>
> HW1P1.20.lm <- lm(Y ~ X)
> HW1P1.20.lm

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)            X
    -0.5802      15.0352

>
> summary(HW1P1.20.lm)

Call:
lm(formula = Y ~ X)

Residuals:
```

```
    Min       1Q   Median       3Q      Max 
-22.7723  -3.7371   0.3334   6.3334  15.4039 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -0.5802     2.8039  -0.207    0.837    
X            15.0352     0.4831  31.123   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 8.914 on 43 degrees of freedom
Multiple R-squared:  0.9575,    Adjusted R-squared:  0.9565 
F-statistic: 968.7 on 1 and 43 DF,  p-value: < 2.2e-16 

> 
> R <- resid(HW1P1.20.lm)
> R
          1           2           3           4           5           6           7 
 -9.4903394   0.4391645   1.4744125  11.5096606  -2.4550914 -12.7723238  -6.5960836 
          8           9          10          11          12          13          14 
 14.4039164 -10.4550914   2.5096606   9.2629243   6.2276762   3.3686684  -8.5255875 
         15          16          17          18          19          20          21 
 12.4391645 -19.7018277   0.3334204  11.2981723 -22.7723238  -2.5608355  -8.5960836 
         22          23          24          25          26          27          28 
 -3.6665796   4.3334204  -0.5960836  -0.7370757   7.3334204 -11.4903394  -1.5960836 
         29          30          31          32          33          34          35 
  6.3334204   6.3686684   3.2981723  15.4039164  -9.4903394  -1.4903394 -11.4550914 
         36          37          38          39          40          41          42 
 -2.5608355  11.4039164  -2.7370757   7.3334204  12.5449086  -3.7370757   4.5096606 
         43          44          45 
 -2.4903394   1.4391645   2.4039164 
> stem(R,scale = 4)

  The decimal point is at the |

  -22 | 8
  -21 | 
  -20 | 
  -19 | 7
  -18 | 
  -17 | 
  -16 | 
  -15 | 
  -14 | 
  -13 | 
  -12 | 8
  -11 | 55
  -10 | 5
   -9 | 55
   -8 | 65
```

```
   -7 |
   -6 | 6
   -5 |
   -4 |
   -3 | 77
   -2 | 76655
   -1 | 65
   -0 | 76
    0 | 34
    1 | 45
    2 | 45
    3 | 34
    4 | 35
    5 |
    6 | 234
    7 | 33
    8 |
    9 | 3
   10 |
   11 | 345
   12 | 45
   13 |
   14 | 4
   15 | 4

>
> #d.
> ##against the fitted value
> par(mfrow = c(1,2))
> plot(resid(HW1P1.20.lm)~fitted(HW1P1.20.lm))
> abline(h=0)
>
> ##against X_i's
> plot(resid(HW1P1.20.lm) ~ X)
> abline(h = 0)
>
> #e.
> #normal probability plot
> qqnorm(resid(HW1P1.20.lm), main = "")
>
> #g.
> install.packages("lawstat")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into ‘C:/Users/alexa/OneDrive/Documents/R/win-library/4.1’
(as ‘lib’ is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/lawstat_3.4.zip'
Content type 'application/zip' length 251233 bytes (245 KB)
```

```
downloaded 245 KB

package 'lawstat' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
	C:\Users\alexa\AppData\Local\Temp\RtmpuYSzvp\downloaded_packages
> library("lawstat")
Warning message:
package 'lawstat' was built under R version 4.1.1
> ei <- resid(HW1P1.20.lm)
> require(lawstat)
> BF.htest <- levene.test(ei[order(X)], group = c(rep(1,22),rep(2,23)),
+                         location = "median")
> BF.htest

	Modified robust Brown-Forsythe Levene-type test based on the absolute deviations
	from the median

data:  ei[order(X)]
Test Statistic = 0.027134, p-value = 0.8699

>
> ################################################################################
> ###3.5
> #c.
> air <- read.delim("https://www.math.arizona.edu/~piegorsch/571A/Data/Chapter01/CH01PR21.txt",
+           header = F, sep ="")
> X1 <- air$V2
> Y1 <- air$V1
>
> HW1P1.21.lm <- lm(Y1~ X1)
> R1 <- resid(HW1P1.21.l)
Error in resid(HW1P1.21.l) : object 'HW1P1.21.l' not found
> R1
Error: object 'R1' not found
> stem(R1)
Error in stem(R1) : object 'R1' not found
>
> #d.
> ##plot against X
> plot(resid(HW1P1.21.lm) ~ X1)
> abline(h=0)
>
> #e.
> ## normal probability plot
> qqnorm(resid(HW1P1.21.lm), main = "")
>
> #d.
> #BF test
> library("lawstat")
```

```
> ei1 <- resid(HW1P1.21.lm)
> require(lawstat)
> BF.htest1<- levene.test(ei1[order(X1)], group = c(rep(1,5),rep(2,5)),
+                         location = "median")
> BF.htest1

        Modified robust Brown-Forsythe Levene-type test based on the absolute deviations
        from the median

data:  ei1[order(X1)]
Test Statistic = 1.2, p-value = 0.3052

>
> ##################################################################################
> ###3.13
> #b.
> plot(Y~X)
> rmHW1P1.20.lm = lm( Y~ X)
> rmHW1P1.20.lm

Call:
lm(formula = Y ~ X)

Coefficients:
(Intercept)            X
    -0.5802      15.0352

>
> factor(X)
 [1] 2  4  3  2  1  10 5  5  1  2  9  10 6  3  4  8  7  8  10 4  5  7  7  5  9  7  2  5  7
[30] 6  8  5  2  2  1  4  5  9  7  1  9  2  2  4  5
Levels: 1 2 3 4 5 6 7 8 9 10
> fmHW1P1.20.lm = lm(Y ~ factor(X))
> fmHW1P1.20.lm

Call:
lm(formula = Y ~ factor(X))

Coefficients:
(Intercept)   factor(X)2   factor(X)3   factor(X)4   factor(X)5   factor(X)6   factor(X)7
      11.50        16.00        29.50        49.90        66.37        83.00        96.83
 factor(X)8   factor(X)9  factor(X)10
     106.50       123.75       128.50

>
> anova(rmHW1P1.20.lm)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq F value    Pr(>F)
```

```
X          1  76960   76960  968.66 < 2.2e-16 ***
Residuals 43   3416      79
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> anova(rmHW1P1.20.lm,fmHW1P1.20.lm)
Analysis of Variance Table

Model 1: Y ~ X
Model 2: Y ~ factor(X)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     43 3416.4
2     35 2797.7  8    618.72 0.9676 0.4766
>
> ################################################################################
> ###3.17
> #a.
> sales <- read.delim("https://www.math.arizona.edu/~piegorsch/571A/Data/Chapter03/CH03PR17.txt",
+            header = F, sep = "")
> XS <- sales$V2
> YS <- sales$V1
> length(YS)
[1] 10
> length(XS)
[1] 10
> plot(YS ~XS, xlab = "Year (Coded)", ylab = "sales in thousands")
>
> #b.
> HW2P3.17.lm <- lm(YS ~ XS)
> library("MASS")
Warning message:
package ‘MASS’ was built under R version 4.1.1
> HW2P3.17.lm.bc = boxcox(HW2P3.17.lm,
+                     lambda = c(.3,.4,.5,.6,.7),interp = F)
> cbind(HW2P3.17.lm.bc$x, HW2P3.17.lm.bc$y)
     [,1]     [,2]
[1,]  0.3 18.84289
[2,]  0.4 19.48120
[3,]  0.5 19.75460
[4,]  0.6 19.61448
[5,]  0.7 19.10168
>
> #c.
> tHW2P3.17.lm <- lm(sqrt(YS) ~ XS )
> tHW2P3.17.lm

Call:
lm(formula = sqrt(YS) ~ XS)

Coefficients:
(Intercept)           XS
```

```
    10.261       1.076

>
> #d.
> plot(sqrt(YS) ~ XS, xlab = "Year (Coded)", ylab = "sqrt(sales in thousands)")
>
> #e.
> par(mfrow= c(1,2))
> resid(tHW2P3.17.lm)
          1           2           3           4           5           6           7 
-0.36143656  0.28172678  0.31440703 -0.14814273  0.29997018 -0.41084412  0.10392174 
          8           9          10 
-0.47446579  0.46781397 -0.07295049 
> plot(resid(tHW2P3.17.lm)~fitted(tHW2P3.17.lm))
> abline(h = 0)
> qqnorm(resid(tHW2P3.17.lm))
>
> #f
> HW2P3.17.lm <- lm(YS ~ XS)
> HW2P3.17.lm

Call:
lm(formula = YS ~ XS)

Coefficients:
(Intercept)           XS  
      91.56        32.50  

> ##########################################################################################
> ##4.3
> #b
> ##finding the standard error for beta_0 (i just wanted to do it manually
> ##to practice my R skills)
> Avg.squared <- (mean(X))^2
> Avg.squared
[1] 26.12346
> df4.3 <- data.frame(resid(HW1P1.20.lm))
> df4.3
   resid.HW1P1.20.lm.
1          -9.4903394
2           0.4391645
3           1.4744125
4          11.5096606
5          -2.4550914
6         -12.7723238
7          -6.5960836
8          14.4039164
9         -10.4550914
10          2.5096606
11          9.2629243
```

```
12          6.2276762
13          3.3686684
14         -8.5255875
15         12.4391645
16        -19.7018277
17          0.3334204
18         11.2981723
19        -22.7723238
20         -2.5608355
21         -8.5960836
22         -3.6665796
23          4.3334204
24         -0.5960836
25         -0.7370757
26          7.3334204
27        -11.4903394
28         -1.5960836
29          6.3334204
30          6.3686684
31          3.2981723
32         15.4039164
33         -9.4903394
34         -1.4903394
35        -11.4550914
36         -2.5608355
37         11.4039164
38         -2.7370757
39          7.3334204
40         12.5449086
41         -3.7370757
42          4.5096606
43         -2.4903394
44          1.4391645
45          2.4039164
> SSE4.3 <- sum((df4.3[1:45,])^2)
> SSE4.3
[1] 3416.377
> MSE4.3 <- SSE4.3/43
> MSE4.3
[1] 79.45063
>
> for (i in 1:43){
+   i <- X - mean(X)
+ }
> sum_X_squared <- sum((i)^2)
> SE.beta0 <- sqrt(MSE4.3*(1/45 +
+                          (Avg.squared/sum_X_squared)))
> SE.beta0
[1] 2.803941
> ## finding the standard error for beta_1
```

```
> SE.beta1 <- sqrt(MSE4.3/sum_X_squared)
> SE.beta1
[1] 0.4830872
>
> B <- qt(1-(.05/4), 43)
>
> CIupr <- -.580157 + B*(SE.beta0)
> CIlwr <- -.580157 - B*(SE.beta0)
>
> CIupr1 <- 15.0352 + B*(SE.beta1)
> CIlwr1 <- 15.0352 - B*(SE.beta1)
>
> ###############################################################################
> ##4.4
> #c.
> summary(HW1P1.21.lm)

Call:
lm(formula = Y1 ~ X1)

Residuals:
   Min     1Q Median     3Q    Max
  -2.2   -1.2    0.3    0.8    1.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.2000     0.6633  15.377 3.18e-07 ***
X1            4.0000     0.4690   8.528 2.75e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.483 on 8 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.8885
F-statistic: 72.73 on 1 and 8 DF,  p-value: 2.749e-05

>
> B1 <- qt(1-(.01/4), 8)
> B1
[1] 3.832519
>
> CIupr2 = 10.2 + B1*(.6633)
> CIlwr2 = 10.2 - B1*(.6633)
> CIupr2
[1] 12.74211
> CIlwr2
[1] 7.65789
>
> CIupr3 = 4 + B1*(.469)
> CIlwr3 = 4 - B1*(.469)
> CIupr3
```

```
[1] 5.797451
> CIlwr3
[1] 2.202549
>
> ##################################################################
> ##4.7
> #a.
> W <- sqrt(2*qf(.90,2,43))
> W
[1] 2.204725
> seYhat3 <- sqrt(MSE4.3*(1/45 +
+                          (3-mean(X))^2/sum_X_squared))
> seYhat3
[1] 1.675012
>
> seYhat5 <- sqrt(MSE4.3*(1/45 +
+                         (5-mean(X))^2/sum_X_squared))
> seYhat5
[1] 1.329831
>
> seYhat7 <- sqrt(MSE4.3*(1/45 +
+                         (7-mean(X))^2/sum_X_squared))
> seYhat7
[1] 1.6119
>
> ##math
> 44.5256+W*(1.675012)
[1] 48.21854
> 44.5256-W*(1.675012)
[1] 40.83266
>
> 74.5961 + W*(1.329831)
[1] 77.52801
> 74.5961 - W*(1.329831)
[1] 71.66419
>
> 104.667 + W*(1.6119)
[1] 108.2208
> 104.667 - W*(1.6119)
[1] 101.1132
>
> #############################################################
> ##4.8
> #a.
> W1 <- sqrt(2*qf(.95,2,8))
> W1
[1] 2.986292
> anova(HW1P1.21.lm)
Analysis of Variance Table
```

```
Response: Y1
          Df Sum Sq Mean Sq F value    Pr(>F)
X1         1  160.0   160.0  72.727 2.749e-05 ***
Residuals  8   17.6     2.2
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> MSE4.8 <-  17.6/8
> seYhat0 <- sqrt(MSE4.8*(1/10 +
+                       (0-mean(X1))^2/sum_X1_squared))
> seYhat0
[1] 0.663325
>
> seYhat1 <- sqrt(MSE4.8*(1/10 +
+                        (1-mean(X1))^2/sum_X1_squared))
> seYhat1
[1] 0.4690416
> seYhat2 <- sqrt(MSE4.8*(1/10 +
+                        (2-mean(X1))^2/sum_X1_squared))
> seYhat2
[1] 0.663325
>
> 10.2+W1*(.6633)
[1] 12.18081
> 10.2-W1*(.6633)
[1] 8.219192
>
> 14.2 + W1*(.469)
[1] 15.60057
> 14.2 - W1*(.469)
[1] 12.79943
>
> 18.2 + W1*(.6633)
[1] 20.18081
> 18.2 - W1*(.6633)
[1] 16.21919
>
> B2 <- qt(1-(.05/6),8)
> B2
[1] 3.015762
>
> ################################################################################
> ##4.16
> #b
> SE.beta1
[1] 0.4830872
> qt(.95,44)
[1] 1.68023
> 15.032 - 1.68023*(.4830872)
[1] 14.2203
> 15.032 + 1.68023*(.4830872)
```

```
[1] 15.8437
>
> #c.
> PI4.16 <- predict(HW1P1.20.lm, newdata= data.frame(X=6),
+                   interval = "pred", level =.90)
> PI4.16
       fit      lwr      upr
1 89.63133 74.46433 104.7983
>
> ##############################################################################
> ##4.17
> #a.
> Origin.lm = lm(Y ~ 0 + X)
> summary(Origin.lm)

Call:
lm(formula = Y ~ 0 + X)

Residuals:
     Min       1Q   Median       3Q      Max
-22.4723  -3.6306   0.2111   6.3694  15.2639

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
X  14.9472     0.2264   66.01   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.816 on 44 degrees of freedom
Multiple R-squared:   0.99,     Adjusted R-squared:  0.9898
F-statistic:  4358 on 1 and 44 DF,  p-value: < 2.2e-16

>
> confint(Origin.lm)
    2.5 %   97.5 %
X 14.4909 15.40356
>
> plot(Y ~ X , pch =19)
> abline(Origin.lm)
>
> #b.
> sum(resid(Origin.lm))
[1] -5.862797
> plot(resid(Origin.lm)~fitted(Origin.lm))
>
> #c.
> plot(Y~X,pch = 19)
>
> #### we now need MSLF and MSPE
>
```

```
> #fit reduced model
> rmOrigin.lm = lm(Y ~ 0 + X)
> rmOrigin.lm

Call:
lm(formula = Y ~ 0 + X)

Coefficients:
    X
14.95

>
> # fit full model\
> factor(X)
 [1] 2  4  3  2  1  10 5  5  1  2  9  10 6  3  4  8  7  8  10 4  5  7  7  5  9  7  2  5  7
[30] 6  8  5  2  2  1  4  5  9  7  1  9  2  2  4  5
Levels: 1 2 3 4 5 6 7 8 9 10
> fmOrigin.lm = lm(Y ~ 0 + factor(X))
> fmOrigin.lm

Call:
lm(formula = Y ~ 0 + factor(X))

Coefficients:
 factor(X)1   factor(X)2   factor(X)3   factor(X)4   factor(X)5   factor(X)6   factor(X)7
     11.50        27.50        41.00        61.40        77.87        94.50       108.33
 factor(X)8   factor(X)9  factor(X)10
    118.00       135.25       140.00

>
> anova(rmOrigin.lm,fmOrigin.lm)
Analysis of Variance Table

Model 1: Y ~ 0 + X
Model 2: Y ~ 0 + factor(X)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     44 3419.8
2     35 2797.7  9    622.12 0.8648 0.5644
> #####################################################################
> ##5.4
> #1
> MX <- matrix(c(1,1,1,1,1,8,4,0,-4,-8), ncol =2)
> MX
     [,1] [,2]
[1,]    1    8
[2,]    1    4
[3,]    1    0
[4,]    1   -4
[5,]    1   -8
>
```

```
> MY <- matrix(c(7.8,9,10.2,11,11.7))
> MY
     [,1]
[1,]  7.8
[2,]  9.0
[3,] 10.2
[4,] 11.0
[5,] 11.7
>
> tMY <- t(MY)
> tMY
     [,1] [,2] [,3] [,4] [,5]
[1,]  7.8    9 10.2   11 11.7
>
> tMY%*%MY
       [,1]
[1,] 503.77
>
> #2
> tMX <- t(MX)
>
> tMX%*%MX
     [,1] [,2]
[1,]    5    0
[2,]    0  160
>
> #3
> tMX%*%MY
      [,1]
[1,]  49.7
[2,] -39.2
>
> ##########################################################################
> ##5.6
> #a.
> MY1 <- matrix(Y1)
> MY1
      [,1]
 [1,]   16
 [2,]    9
 [3,]   17
 [4,]   12
 [5,]   22
 [6,]   13
 [7,]    8
 [8,]   15
 [9,]   19
[10,]   11
>
> tMY1 <- t(MY1)
```

```
>
> tMY1 %*% MY1
     [,1]
[1,] 2194
>
> #b.
> MX1 <- matrix(c(rep(1,10),X1), ncol = 2)
> MX1
      [,1] [,2]
 [1,]    1    1
 [2,]    1    0
 [3,]    1    2
 [4,]    1    0
 [5,]    1    3
 [6,]    1    1
 [7,]    1    0
 [8,]    1    1
 [9,]    1    2
[10,]    1    0
>
> tMX1 <- t(MX1)
> tMX1
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
[1,]    1    1    1    1    1    1    1    1    1     1
[2,]    1    0    2    0    3    1    0    1    2     0
>
> tMX1 %*% MX1
     [,1] [,2]
[1,]   10   10
[2,]   10   20
>
> #c.
> tMX1 %*% MY1
     [,1]
[1,]  142
[2,]  182
>
> ################################################################
> ##5.12
> solve(tMX%*%MX)
     [,1]    [,2]
[1,]  0.2 0.00000
[2,]  0.0 0.00625
>
> ##################################################################
> ##5.23
> #a (1)
> tXX <- solve(tMX %*%MX)
> tXY <- tMX %*% MY
> b <- tXX %*% tXY
```

```
>
> #(2)
> H <- MX%*%tXX%*%tMX
> H
     [,1] [,2] [,3] [,4] [,5]
[1,]  0.6  0.4  0.2  0.0 -0.2
[2,]  0.4  0.3  0.2  0.1  0.0
[3,]  0.2  0.2  0.2  0.2  0.2
[4,]  0.0  0.1  0.2  0.3  0.4
[5,] -0.2  0.0  0.2  0.4  0.6
> I <- diag(5)
> I
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    0    0    0    0
[2,]    0    1    0    0    0
[3,]    0    0    1    0    0
[4,]    0    0    0    1    0
[5,]    0    0    0    0    1
>
> e <- (I - H)%*%MY
> e
      [,1]
[1,] -0.18
[2,]  0.04
[3,]  0.26
[4,]  0.08
[5,] -0.20
>
> #(3)
> tb <- t(b)
> J <- matrix(rep(1,25), ncol = 5)
> J
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    1    1    1    1
[2,]    1    1    1    1    1
[3,]    1    1    1    1    1
[4,]    1    1    1    1    1
[5,]    1    1    1    1    1
> SSR <- tb%*%tMX%*%MY - (1/5)%*%tMY%*%J%*%MY
> SSR
      [,1]
[1,] 9.604
>
> #(4)
> SSE <- tMY %*% MY - tb%*%tMX%*%MY
> SSE
      [,1]
[1,] 0.148
>
> #(5)
```

```
> XF <- c(8,4,0,-4,-8)
> YF <-c(7.8,9,10.2,11,11.7)
> flavor.lm <- lm(YF ~ XF)
> flavor_summ <- summary(flavor.lm)
> mean(flavor_summ$residuals^2)
[1] 0.0296
> MSEF <- .04928
> Var_CovM <- MSEF*tXX
> Var_CovM
         [,1]     [,2]
[1,] 0.009856 0.000000
[2,] 0.000000 0.000308
>
> #(6)
> Xh <- matrix(c(1, -6))
> tXh <- t(Xh)
> Y_hat1 <- tXh %*% b
> Y_hat1
      [,1]
[1,] 11.41
>
> #(7)
> MSEF*(1 + tXh%*%tXX%*%Xh)
         [,1]
[1,] 0.070224
>
> #c.
> H <- MX%*%tXX%*%tMX
> H
     [,1] [,2] [,3] [,4] [,5]
[1,]  0.6  0.4  0.2  0.0 -0.2
[2,]  0.4  0.3  0.2  0.1  0.0
[3,]  0.2  0.2  0.2  0.2  0.2
[4,]  0.0  0.1  0.2  0.3  0.4
[5,] -0.2  0.0  0.2  0.4  0.6
>
> ################################################################################
> ##5.25
> #a
>
> #(1)
> tXX1 <- solve(tMX1 %*% MX1)
> tXX1
     [,1] [,2]
[1,]  0.2 -0.1
[2,] -0.1  0.1
>
> #(2)
> tXY1 <- tMX1 %*% MY1
> b1 <- tXX1 %*% tXY1
```

```
> b1
     [,1]
[1,] 10.2
[2,]  4.0
>
> #(3)
> H1 <- MX1%*%tXX1%*%tMX1
> H1
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [2,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [3,]  0.1  0.0  0.2  0.0  0.3  0.1  0.0  0.1  0.2   0.0
 [4,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [5,]  0.1 -0.1  0.3 -0.1  0.5  0.1 -0.1  0.1  0.3  -0.1
 [6,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [7,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [8,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [9,]  0.1  0.0  0.2  0.0  0.3  0.1  0.0  0.1  0.2   0.0
[10,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
> I1 <- diag(10)
> I1
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]    1    0    0    0    0    0    0    0    0     0
 [2,]    0    1    0    0    0    0    0    0    0     0
 [3,]    0    0    1    0    0    0    0    0    0     0
 [4,]    0    0    0    1    0    0    0    0    0     0
 [5,]    0    0    0    0    1    0    0    0    0     0
 [6,]    0    0    0    0    0    1    0    0    0     0
 [7,]    0    0    0    0    0    0    1    0    0     0
 [8,]    0    0    0    0    0    0    0    1    0     0
 [9,]    0    0    0    0    0    0    0    0    1     0
[10,]    0    0    0    0    0    0    0    0    0     1
>
> e1 <- (I1 - H1)%*%MY1
> e1
      [,1]
 [1,]  1.8
 [2,] -1.2
 [3,] -1.2
 [4,]  1.8
 [5,] -0.2
 [6,] -1.2
 [7,] -2.2
 [8,]  0.8
 [9,]  0.8
[10,]  0.8
> #(4)
> H1
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
 [1,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
```

```
 [2,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [3,]  0.1  0.0  0.2  0.0  0.3  0.1  0.0  0.1  0.2   0.0
 [4,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [5,]  0.1 -0.1  0.3 -0.1  0.5  0.1 -0.1  0.1  0.3  -0.1
 [6,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [7,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
 [8,]  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1   0.1
 [9,]  0.1  0.0  0.2  0.0  0.3  0.1  0.0  0.1  0.2   0.0
[10,]  0.1  0.2  0.0  0.2 -0.1  0.1  0.2  0.1  0.0   0.2
>
> #(5)
> tb1 <- t(b1)
> SSE1 <- tMY1 %*% MY1 - tb1%*%tMX1%*%MY1
> SSE1
     [,1]
[1,] 17.6
> capture.output()
character(0)
> save.image("~/ECON
MASTERS/MATH_STAT571A_AdvanceRegressionAnalysis/MATH571A_HW2output.R.RData")
Warning message:
In flat_str(content, breaks) : Coercing content to character
>
```