# Regession interpretation 3 - 4

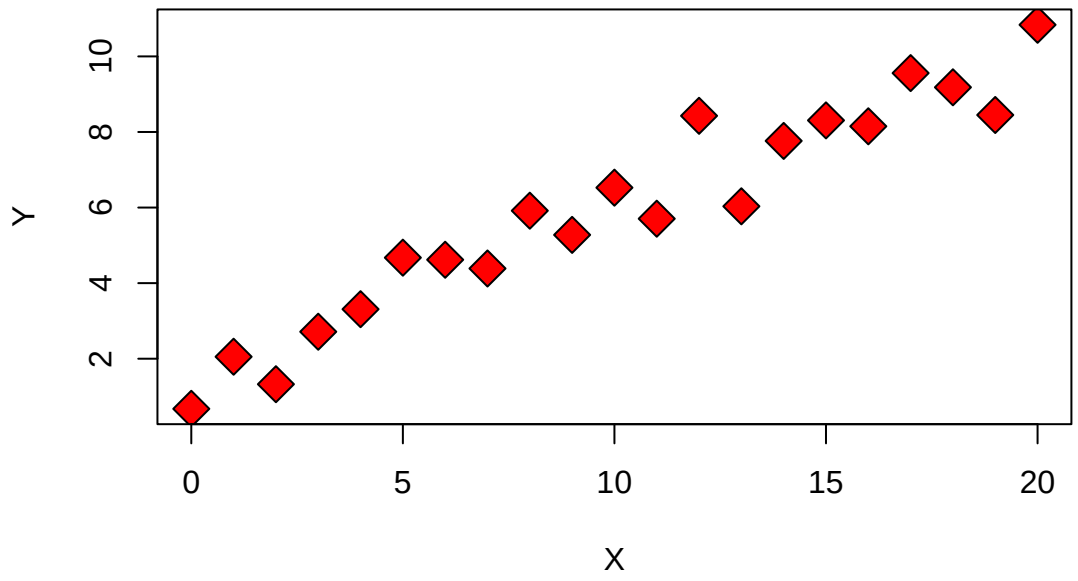
Oh SukJu

## Contents

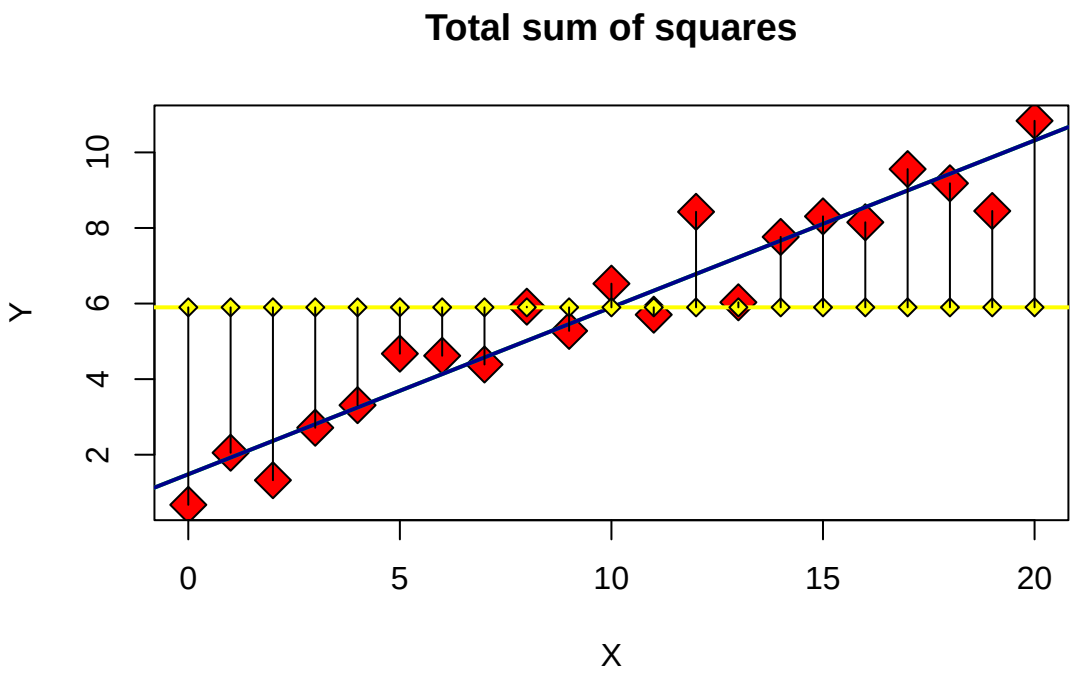


### 0.1 3. Diagnostics for Simple Linear Regression

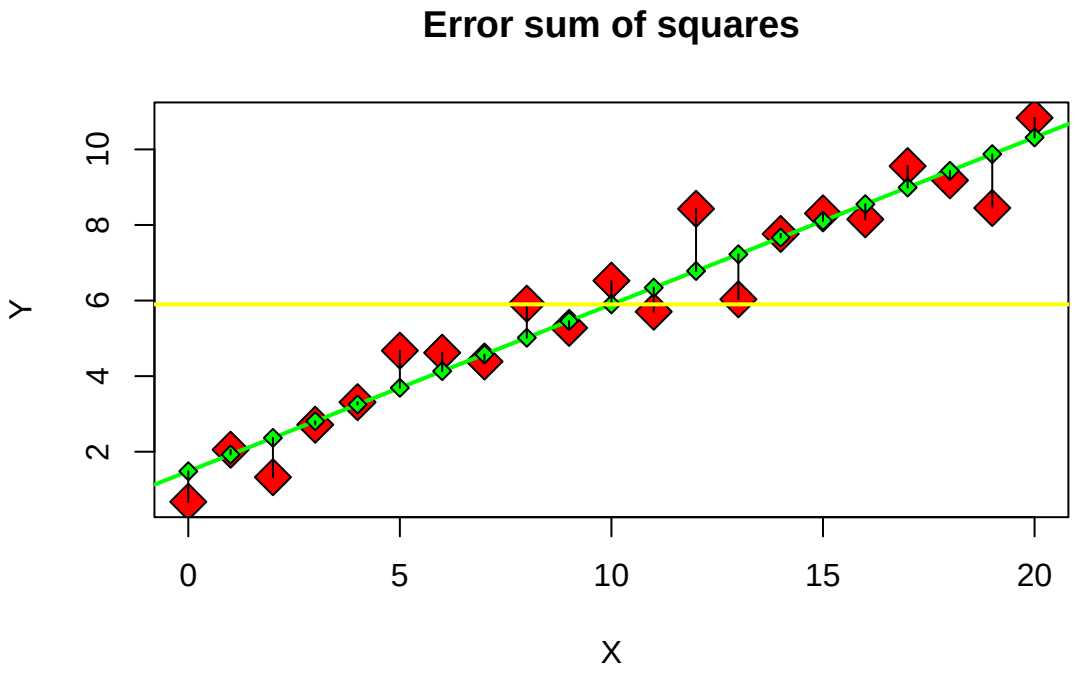
```
# page 10

set.seed(101)

X = seq(0, 20, length = 21)

Y = 0.5 * X + 1 + rnorm(21)

Y.lm = lm(Y ~ X)

meanY = mean(Y)

Yhat = predict(Y.lm)

plot(X,Y, pch=23, bg='red', cex=2)
```



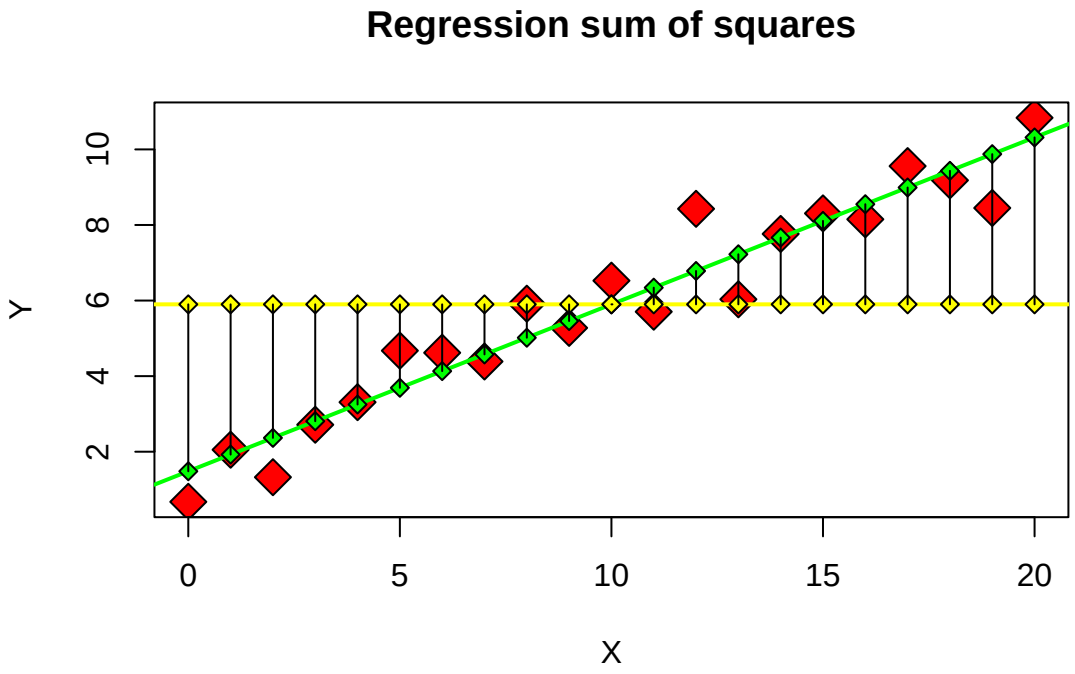

```r
# page 11

plot(X, Y, pch = 23, bg = "red", main='Total sum of squares', cex=2)

abline(Y.lm, pch=23, col='green', lwd=2)

abline(coef(Y.lm), pch=23, col='darkblue', lwd=2)


abline(h = meanY, col = "yellow", lwd = 2)

for (i in 1:21) {

  points(X[i], meanY, pch = 23, bg = "yellow")

  lines(c(X[i], X[i]), c(Y[i], meanY))

}
```




```
# page 12

plot(X, Y, pch = 23, bg = "red", main="Error sum of squares", cex=2)

abline(Y.lm, col = "green", lwd = 2)

for (i in 1:21) {

  points(X[i], Yhat[i], pch = 23, bg = "green")

  lines(c(X[i], X[i]), c(Y[i], Yhat[i]))

}

abline(h = meanY, col = "yellow", lwd = 2)
```




```
# page 13

plot(X, Y, pch = 23, bg = "red", main="Regression sum of squares", cex=2)

abline(Y.lm, col = "green", lwd = 2)

abline(h = meanY, col = "yellow", lwd = 2)

for (i in 1:21) {

  points(X[i], Yhat[i], pch = 23, bg = "green")

  points(X[i], meanY, pch = 23, bg = "yellow")

  lines(c(X[i], X[i]), c(meanY, Yhat[i]))

}
```




```
# page 14

load("/Users/ohsukju/Downloads/data/wages.Rdata")

wages.lm = lm(logwage ~ education, data=wages)


SSE = sum(resid(wages.lm)^2)

SST = sum((wages$logwage - mean(wages$logwage))^2)

SSR = sum((mean(wages$logwage) - predict(wages.lm))^2)

data.frame(SST, SSE + SSR)
```

```
##        SST SSE...SSR
## 1 410.2148  410.2148
```

```
F = (SSR / 1) / (SSE / wages.lm$df)

print(F)
```

```
## [1] 340.0297
```

```r
summary(wages.lm)
```

```
##
## Call:
## lm(formula = logwage ~ education, data = wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78239 -0.25265  0.01636  0.27965  1.61101
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.239194   0.054974   22.54   <2e-16 ***
## education   0.078600   0.004262   18.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4038 on 2176 degrees of freedom
## Multiple R-squared:  0.1351, Adjusted R-squared:  0.1347
## F-statistic:   340 on 1 and 2176 DF,  p-value: < 2.2e-16
```

```r
2176*0.1351 / (1 - 0.1351)
```

```
## [1] 339.8978
```

```r
cor(wages$education, wages$logwage)^2
```

```
## [1] 0.1351453
```

```r
# page 17

summary(wages.lm)
```

```
##
## Call:
## lm(formula = logwage ~ education, data = wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78239 -0.25265  0.01636  0.27965  1.61101
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.239194   0.054974   22.54   <2e-16 ***
## education   0.078600   0.004262   18.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4038 on 2176 degrees of freedom
## Multiple R-squared:  0.1351, Adjusted R-squared:  0.1347
## F-statistic:   340 on 1 and 2176 DF,  p-value: < 2.2e-16
```

```r
18.44**2
```

```
## [1] 340.0336
```

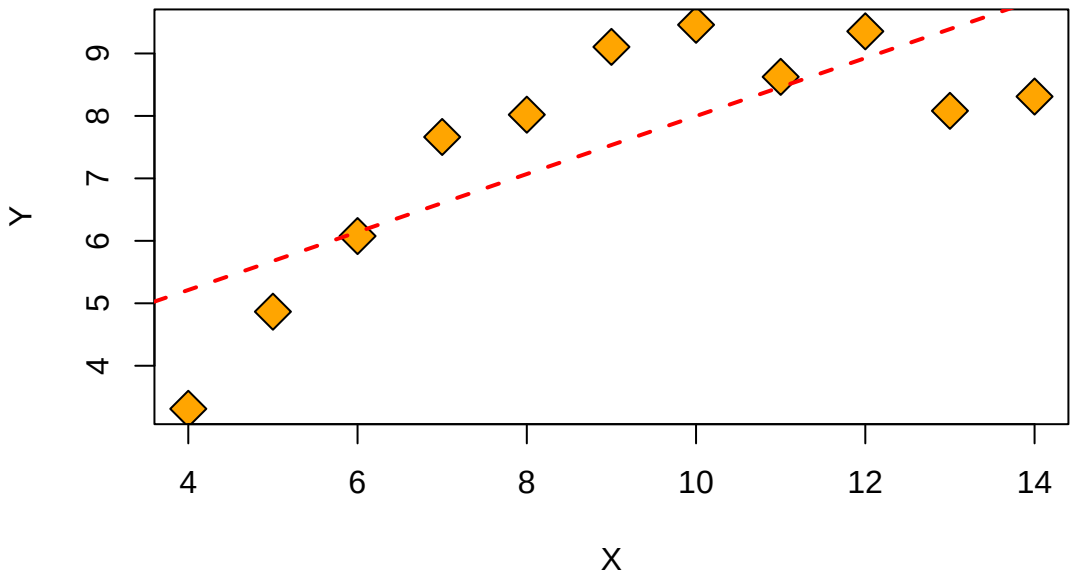
```r
# page 20

summary(wages.lm)
```

```
##
## Call:
## lm(formula = logwage ~ education, data = wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78239 -0.25265  0.01636  0.27965  1.61101
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.239194   0.054974   22.54   <2e-16 ***
## education   0.078600   0.004262   18.44   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4038 on 2176 degrees of freedom
## Multiple R-squared:  0.1351, Adjusted R-squared:  0.1347
## F-statistic:   340 on 1 and 2176 DF,  p-value: < 2.2e-16
```

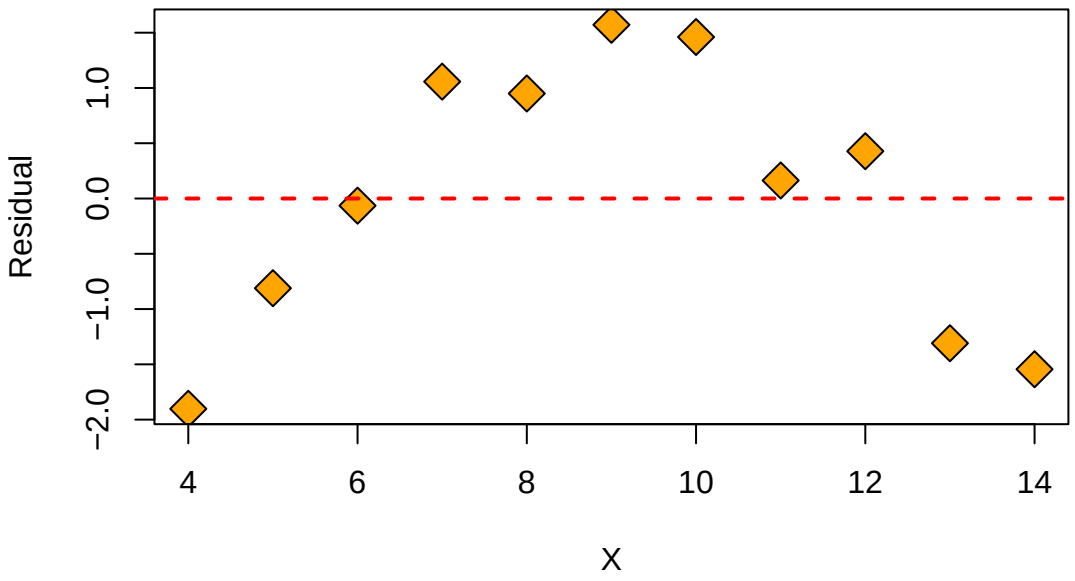
```r
qf(0.95, 1, 2176)
```

```
## [1] 3.845736
```

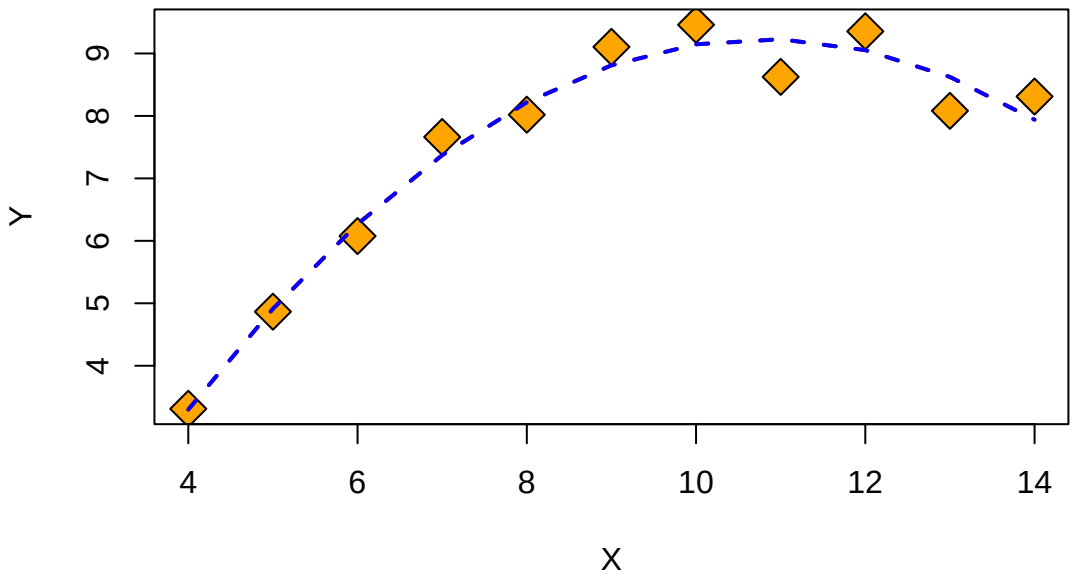
```r
# page 23

load("/Users/ohsukju/Downloads/data/anscombe.Rdata")

y = anscombe$y2

x = anscombe$x2

y = y + rnorm(length(y)) * 0.45

plot(x, y, pch = 23, bg = "orange", cex = 2, ylab = "Y", xlab = "X")

simple.lm = lm(y ~ x)

abline(simple.lm, lwd = 2, col = "red", lty = 2)
```



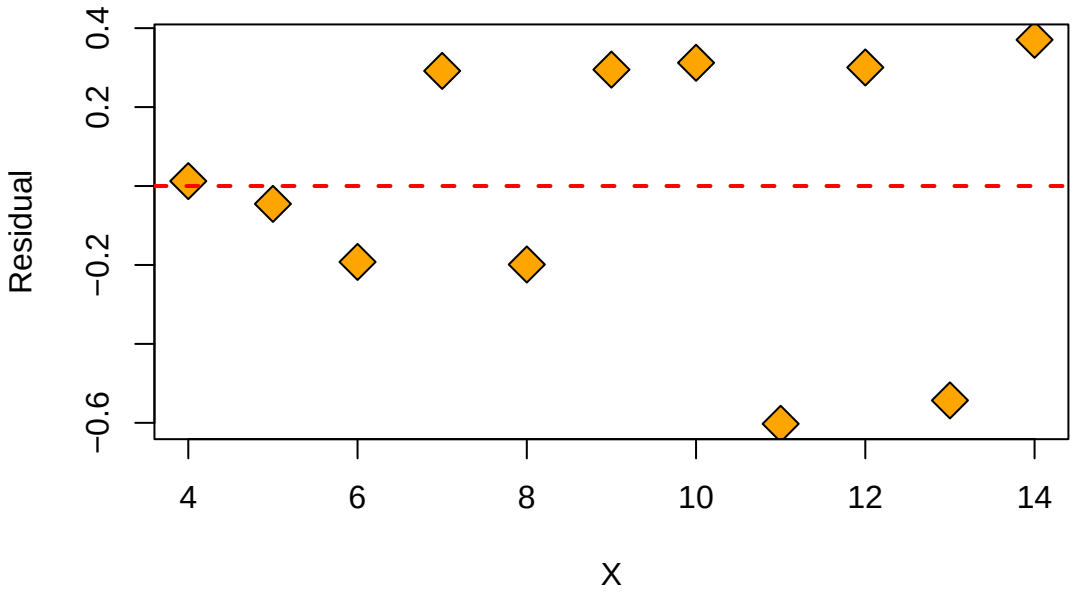

```
# page 24

plot(x, resid(simple.lm), ylab = "Residual", xlab = "X", pch = 23, bg = "orange", cex = 2)

abline(h = 0, lwd = 2, col = "red", lty = 2)
```



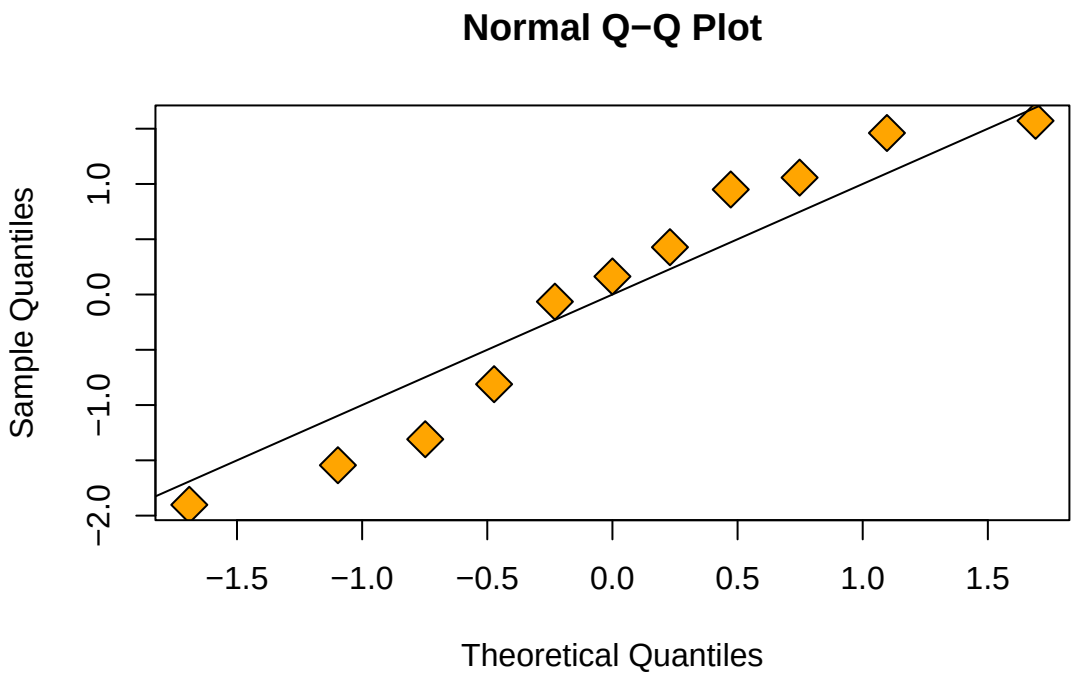

```
# page 25

x2<-x^2

quadratic.lm = lm(y ~ poly(x, 2))

quadratic.lm2 = lm(y ~ x+x2)

Xsort = sort(x)

plot(x, y, pch = 23, bg = "orange", cex = 2, ylab = "Y", xlab = "X")

lines(Xsort, predict(quadratic.lm, list(x = Xsort)), col = "red", lty = 2, lwd = 2)

lines(Xsort, predict(quadratic.lm2, list(x = Xsort, x2=Xsort^2)), col = "blue", lty = 2, lwd = 2)
```



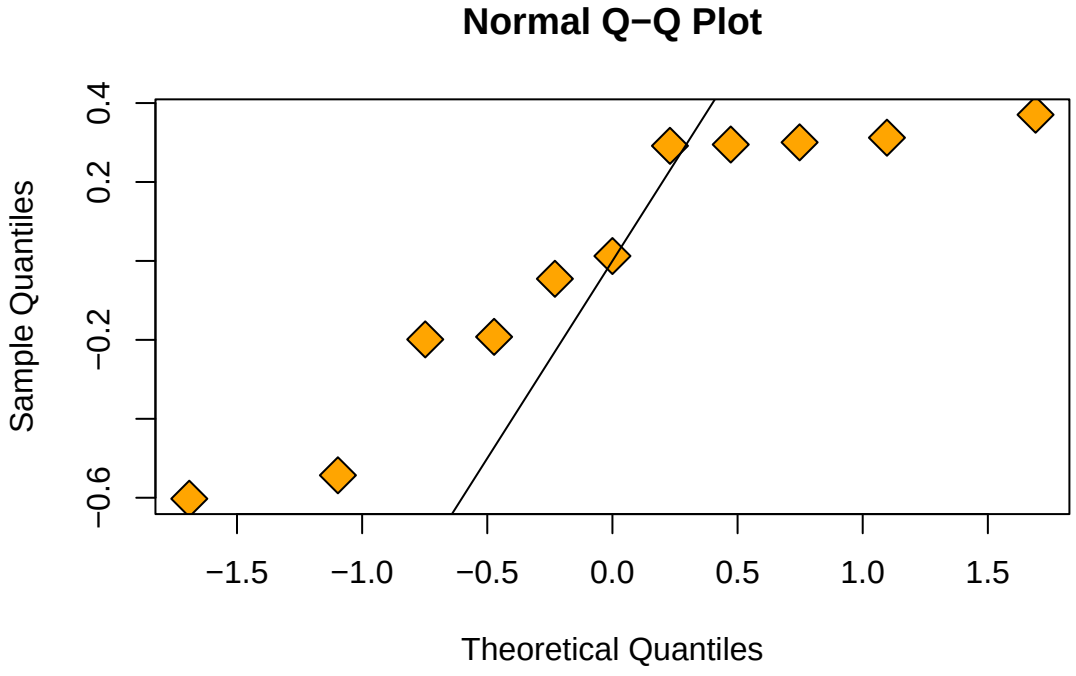

```
# page 26

plot(x, resid(quadratic.lm), ylab = "Residual", xlab = "X", pch = 23, bg = "orange", cex = 2)

abline(h = 0, lwd = 2, col = "red", lty = 2)
```



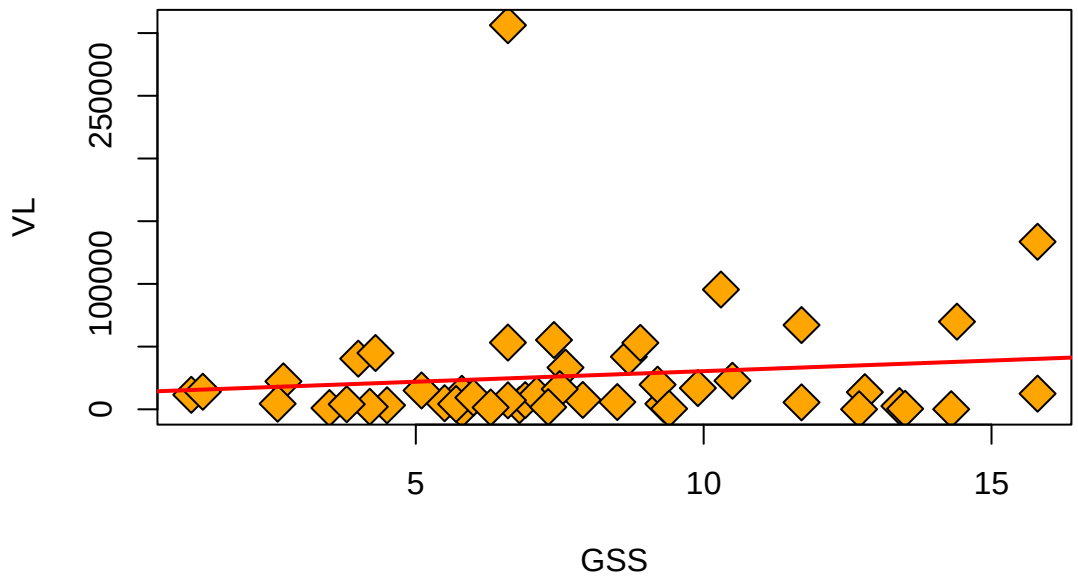

```
# page 28

qqnorm(resid(simple.lm), pch = 23, bg = "orange", cex = 2)

abline(0, 1)
```



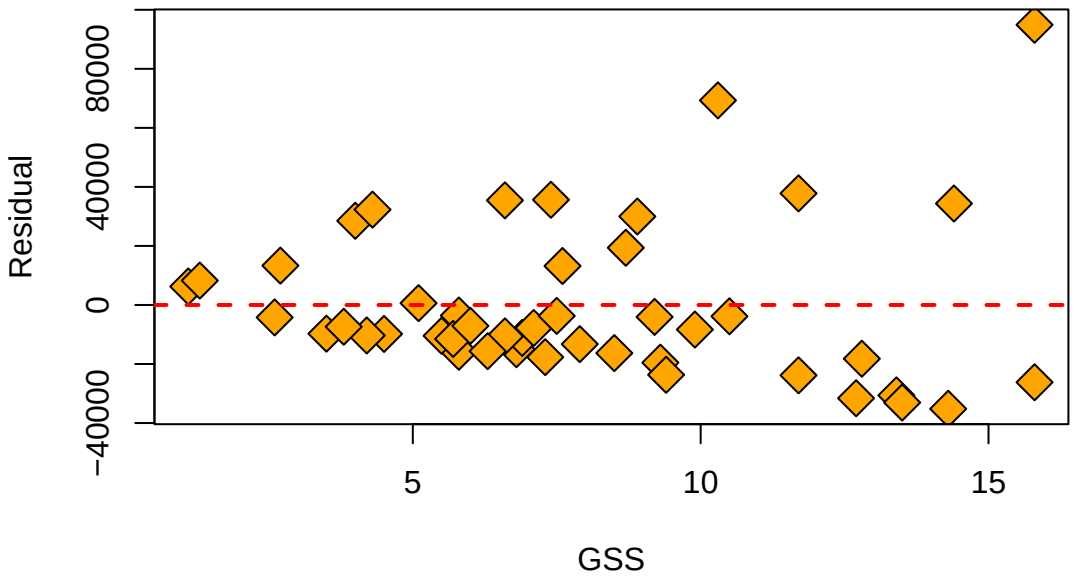

```
qqnorm(resid(quadratic.lm), pch = 23, bg = "orange", cex = 2)

abline(0, 1)
```



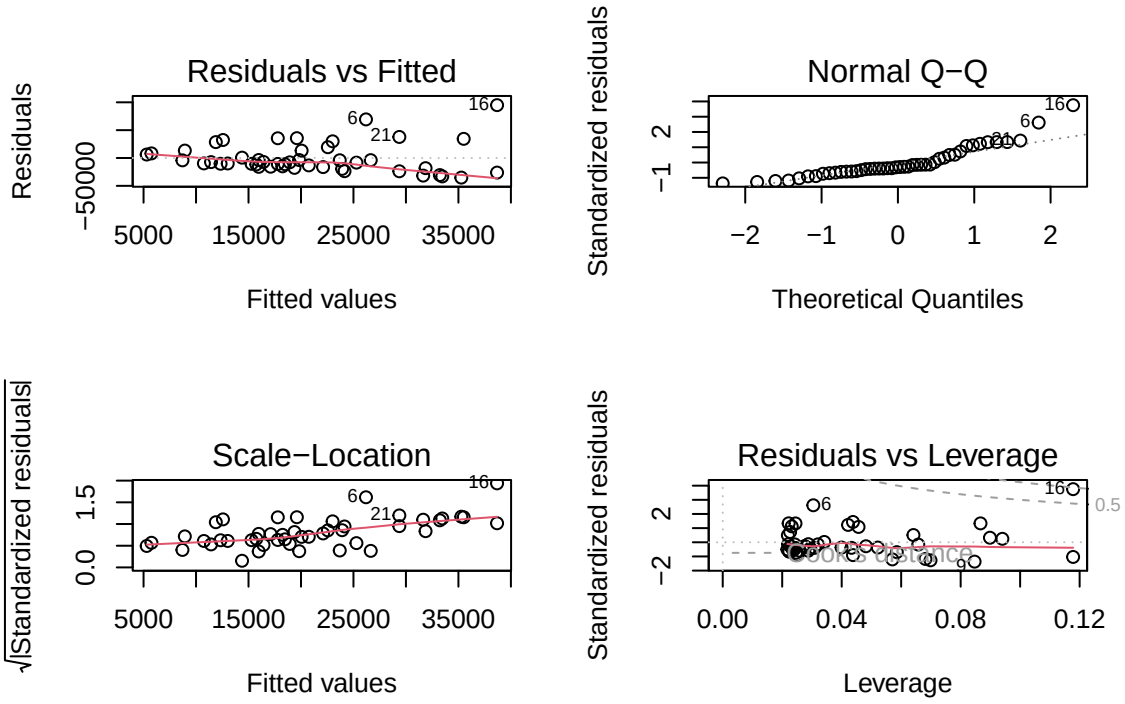
```r
# page 29

load("/Users/ohsukju/Downloads/data/HIV.VL.Rdata")

attach(viral.load)

plot(GSS, VL, pch=23, bg='orange', cex=2)

viral.lm = lm(VL ~ GSS)

abline(viral.lm, col='red', lwd=2)
```




```r
# page 30

good = (VL < 200000)

viral.lm.good = lm(VL ~ GSS, subset=good)

#abline(viral.lm.good, col='blue', lwd=2)


plot(GSS[good], resid(viral.lm.good), pch=23, bg='orange', cex=2, xlab='GSS', ylab='Residual')

abline(h=0, lwd=2, col='red', lty=2)
```

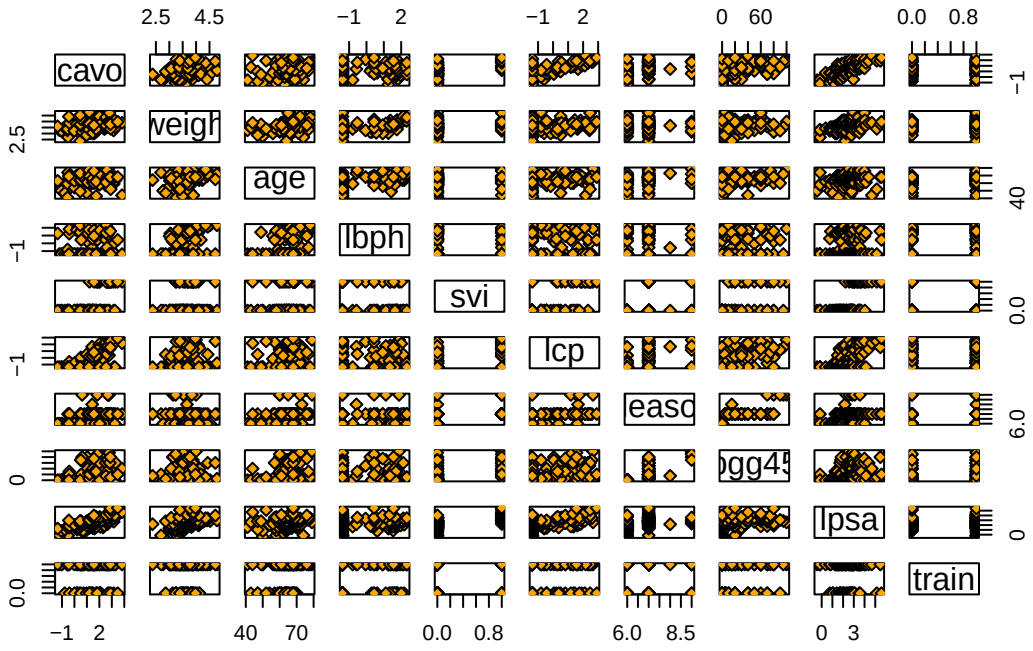



```
par(mfrow=c(2,2))

plot(viral.lm.good)
```



## 0.2 4. Multiple linear regression

```
# page 4

load("/Users/ohsukju/Downloads/data/prostate.Rdata")

pairs(prostate, pch=23, bg='orange', cex.labels=1.5)
```

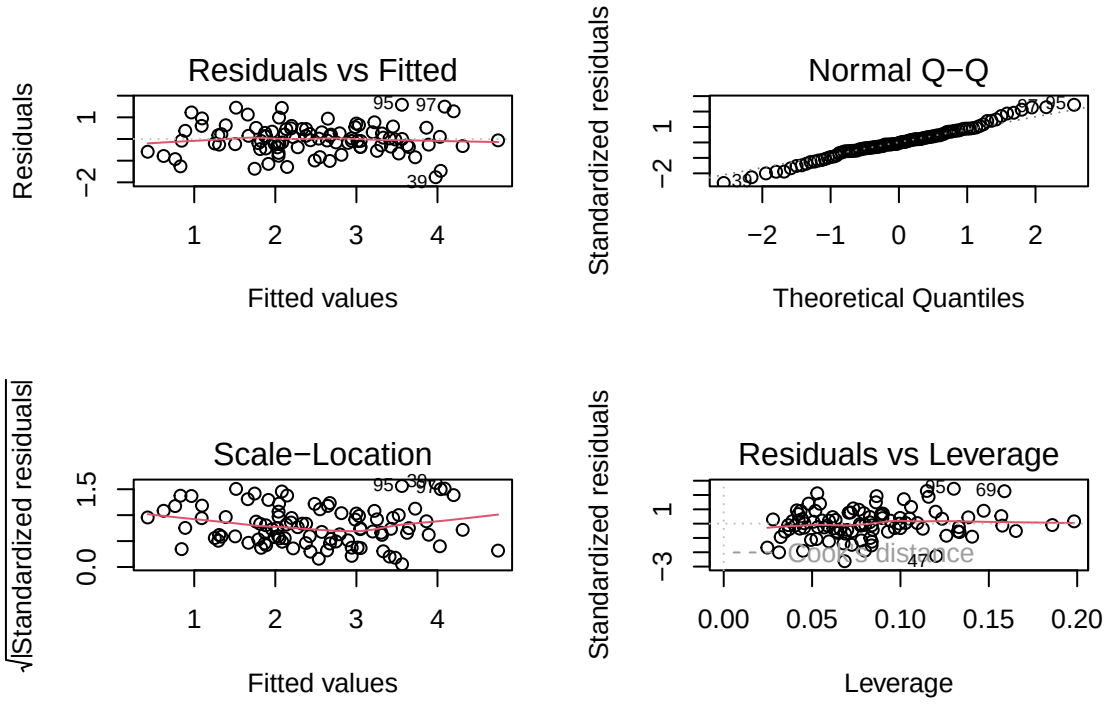



```
# page 7

prostate.lm = lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45, data=prostate)

summary(prostate.lm)
```

```
##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     pgg45, data = prostate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.76395 -0.35764 -0.02143  0.37762  1.58178
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.494155   0.873567   0.566  0.57304
## lcavol       0.569546   0.085847   6.634 2.46e-09 ***
```

```
## lweight      0.614420   0.198449   3.096  0.00262 **
## age         -0.020913   0.010978  -1.905  0.06000 .
## lbph         0.097353   0.057584   1.691  0.09441 .
## svi          0.752397   0.238180   3.159  0.00216 **
## lcp         -0.104959   0.089347  -1.175  0.24323
## pgg45        0.005324   0.003385   1.573  0.11923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.696 on 89 degrees of freedom
## Multiple R-squared:  0.663,  Adjusted R-squared:  0.6365
## F-statistic: 25.01 on 7 and 89 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))

plot(prostate.lm)
```




```
# Y_i=bt0+bt1lcavol_i+b2lweight_i+b3age_i+b4lbph_i+b5svi_i+b6lcp_i+b7pgg45_i+e_i
```

```
# page 8

print(prostate.lm$df.resid)
```

```
## [1] 89
```

```
sigma.hat = sqrt(sum(resid(prostate.lm)^2) / prostate.lm$df.resid)

sigma.hat
```

```
## [1] 0.6959559
```

```
summary(prostate.lm)$sigma
```

```
## [1] 0.6959559
```

```
# page 9

case1 = prostate[10,]

case2 = case1

case2['lcavol'] = case2['lcavol'] + 1



rbind(case1, case2)
```

```
##        lcavol  lweight age      lbph svi       lcp gleason pgg45     lpsa train
## 10  0.2231436 3.244544  63 -1.386294   0 -1.386294       6     0 1.047319 FALSE
## 101 1.2231436 3.244544  63 -1.386294   0 -1.386294       6     0 1.047319 FALSE
```

```
Yhat = predict(prostate.lm, rbind(case1, case2))

diff(Yhat)
```

```
##      101
## 0.569546
```

```
c(Yhat[2]-Yhat[1], coef(prostate.lm)['lcavol'])
```

```
##      101   lcavol
## 0.569546 0.569546
```

```
# page 10

partial_resid_lcavol = resid(lm(lcavol ~ lweight + age + lbph + svi + lcp + pgg45, data=prostate))

partial_resid_lpsa = resid(lm(lpsa ~ lweight + age + lbph + svi + lcp + pgg45, data=prostate))

summary(lm(partial_resid_lpsa ~ partial_resid_lcavol))
```

```
##
## Call:
## lm(formula = partial_resid_lpsa ~ partial_resid_lcavol)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.76395 -0.35764 -0.02143  0.37762  1.58178
##
```

```
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.956e-17  6.840e-02   0.000        1
## partial_resid_lcavol 5.695e-01  8.309e-02   6.854 7.15e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6736 on 95 degrees of freedom
## Multiple R-squared:  0.3309, Adjusted R-squared:  0.3239
## F-statistic: 46.98 on 1 and 95 DF,  p-value: 7.147e-10
```

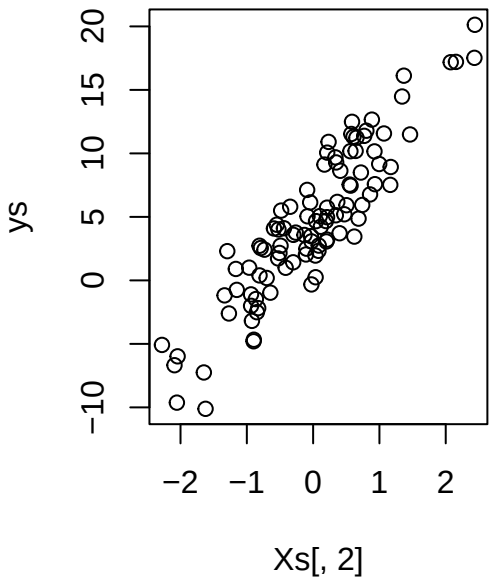
```
## simulation

# generative model

set.seed(101)

ns<-100

#install.packages('mvtnorm')

library(mvtnorm)
```

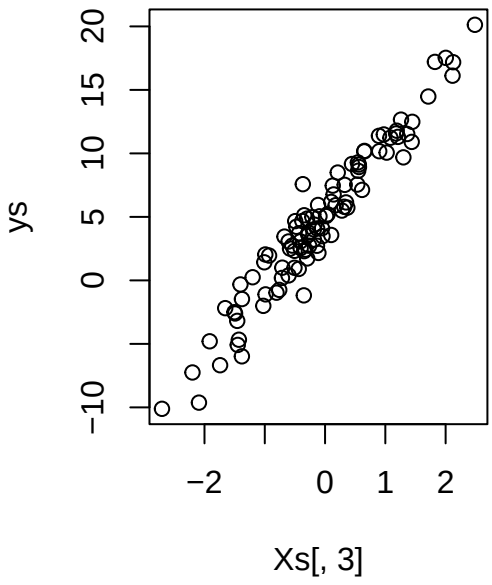
```
## Warning: package 'mvtnorm' was built under R version 4.2.3
```

```
rho.x<-0.8

xs<-rmvnorm(ns, mean=c(0, 0), sigma=matrix(c(1, rho.x, rho.x, 1), 2, 2))

Xs<-cbind(1, xs)
head(Xs)
```

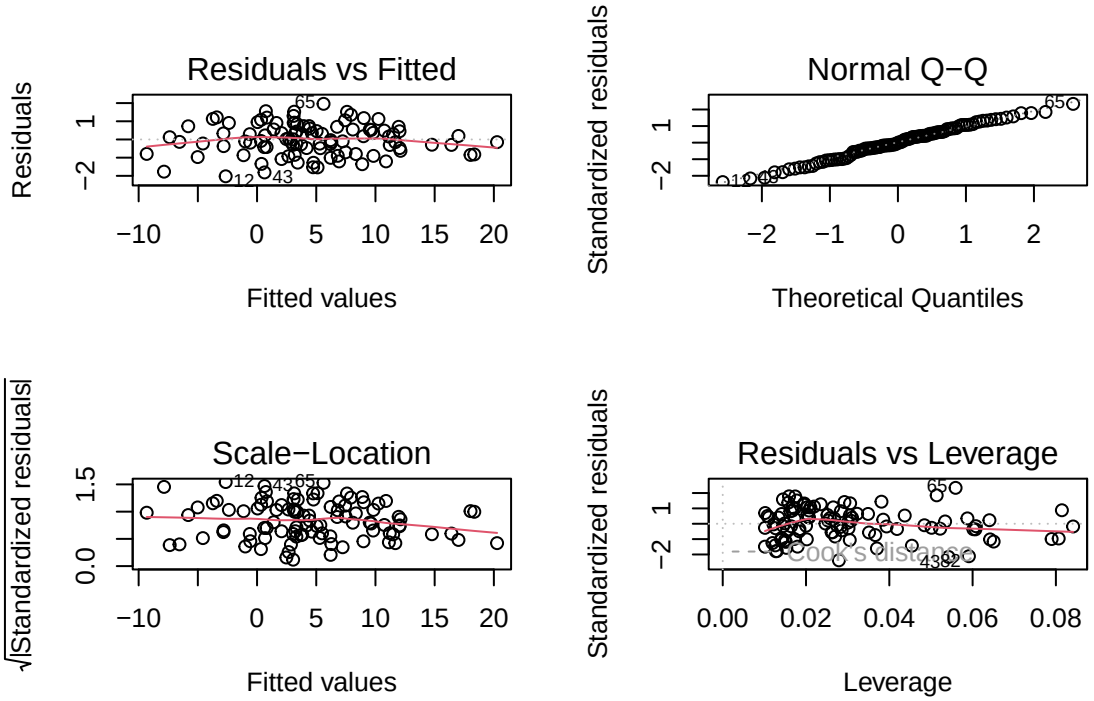
```
##      [,1]        [,2]       [,3]
## [1,]    1 -0.04454745  0.3483290
## [2,]    1 -0.50782366 -0.1101151
## [3,]    1  0.80297412  1.1890076
## [4,]    1  0.50304615  0.1758986
## [5,]    1  0.72037041  0.2104183
## [6,]    1  0.11540426 -0.4754957
```

```
tr.bt<-c(5, 2, 4)

tr.sigma<-1

ys<-Xs%*%tr.bt+rnorm(ns, 0, tr.sigma)



par(mfrow=c(1,2))

plot(Xs[,2], ys)

plot(Xs[,3], ys)
```




```r
datas<-data.frame(y=ys, x1=Xs[,2], x2=Xs[,3])



lm1<-lm(ys~x1+x2, data=datas)

par(mfrow=c(2,2))

plot(lm1)
```




```
summary(lm1)
```

```
##
## Call:
## lm(formula = ys ~ x1 + x2, data = datas)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.02519 -0.51758 -0.01226  0.56301  1.95904
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.00099    0.08714   57.39   <2e-16 ***
## x1           2.22736    0.15940   13.97   <2e-16 ***
## x2           3.95810    0.14676   26.97   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8628 on 97 degrees of freedom
## Multiple R-squared:  0.9791, Adjusted R-squared:  0.9787
## F-statistic:  2276 on 2 and 97 DF,  p-value: < 2.2e-16
```

```
lm0.1<-lm(ys~x1, data=datas)

lm0.2<-lm(ys~x2, data=datas)
```

```r
# results

cbind(rbind(summary(lm0.1)$coef[-1,c(1, 4)],

summary(lm0.2)$coef[-1,c(1, 4)]),

summary(lm1)$coef[-1,c(1, 4)])
```

```
##    Estimate     Pr(>|t|) Estimate     Pr(>|t|)
## x1 5.726837 1.360677e-38 2.227357 5.786541e-25
## x2 5.627316 1.089139e-60 3.958104 7.240269e-47
```

```r
# page 11

Y = prostate$lpsa

n = length(Y)

SST = sum((Y - mean(Y))^2)

MST = SST / (n - 1)

SSE = sum(resid(prostate.lm)^2)

MSE = SSE / prostate.lm$df.residual

SSR = SST - SSE

MSR = SSR / (n - 1 - prostate.lm$df.residual)

print(c(MST,MSE,MSR))
```

```
## [1]  1.3324756  0.4843546 12.1157287
```

```r
# page 13

print(summary(prostate.lm))
```

```
##
## Call:
## lm(formula = lpsa ~ lcavol + lweight + age + lbph + svi + lcp +
##     pgg45, data = prostate)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.76395 -0.35764 -0.02143  0.37762  1.58178
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.494155   0.873567   0.566  0.57304
## lcavol       0.569546   0.085847   6.634 2.46e-09 ***
## lweight      0.614420   0.198449   3.096  0.00262 **
```

```
## age         -0.020913   0.010978  -1.905  0.06000 .
## lbph         0.097353   0.057584   1.691  0.09441 .
## svi          0.752397   0.238180   3.159  0.00216 **
## lcp         -0.104959   0.089347  -1.175  0.24323
## pgg45        0.005324   0.003385   1.573  0.11923
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.696 on 89 degrees of freedom
## Multiple R-squared:  0.663,  Adjusted R-squared:  0.6365
## F-statistic: 25.01 on 7 and 89 DF,  p-value: < 2.2e-16
```

```
F = MSR / MSE

F
```

```
## [1] 25.01417
```

```
# page 17

n = length(Y)

attach(prostate)

X = cbind(rep(1,n), lcavol, lweight, age, lbph, svi, lcp, pgg45)

detach(prostate)

colnames(X)[1] = '(Intercept)'

head(X)
```

```
##      (Intercept)     lcavol  lweight age      lbph svi       lcp pgg45
## [1,]           1 -0.5798185 2.769459  50 -1.386294   0 -1.386294     0
## [2,]           1 -0.9942523 3.319626  58 -1.386294   0 -1.386294     0
## [3,]           1 -0.5108256 2.691243  74 -1.386294   0 -1.386294    20
## [4,]           1 -1.2039728 3.282789  58 -1.386294   0 -1.386294     0
## [5,]           1  0.7514161 3.432373  62 -1.386294   0 -1.386294     0
## [6,]           1 -1.0498221 3.228826  50 -1.386294   0 -1.386294     0
```

```
head(model.matrix(prostate.lm))
```

```
##   (Intercept)     lcavol  lweight age      lbph svi       lcp pgg45
## 1           1 -0.5798185 2.769459  50 -1.386294   0 -1.386294     0
## 2           1 -0.9942523 3.319626  58 -1.386294   0 -1.386294     0
## 3           1 -0.5108256 2.691243  74 -1.386294   0 -1.386294    20
## 4           1 -1.2039728 3.282789  58 -1.386294   0 -1.386294     0
## 5           1  0.7514161 3.432373  62 -1.386294   0 -1.386294     0
## 6           1 -1.0498221 3.228826  50 -1.386294   0 -1.386294     0
```

```r
# page 18

beta = as.numeric(solve(t(X) %*% X) %*% t(X) %*% Y)

names(beta) = colnames(X)

print(beta)
```

```
##  (Intercept)       lcavol      lweight          age         lbph          svi
##  0.494154754  0.569546032  0.614419817 -0.020913467  0.097352535  0.752397342
##          lcp        pgg45
## -0.104959408  0.005324465
```

```r
print(coef(prostate.lm))
```

```
##  (Intercept)       lcavol      lweight          age         lbph          svi
##  0.494154754  0.569546032  0.614419817 -0.020913467  0.097352535  0.752397342
##          lcp        pgg45
## -0.104959408  0.005324465
```

```r
# page 21

confint(prostate.lm, level=0.90)
```

```
##                        5 %         95 %
## (Intercept) -0.9578488958  1.946158404
## lcavol       0.4268548240  0.712237239
## lweight      0.2845659251  0.944273708
## age         -0.0391601782 -0.002666755
## lbph         0.0016386253  0.193066445
## svi          0.3565053323  1.148289353
## lcp         -0.2534678904  0.043549074
## pgg45       -0.0003011464  0.010950077
```

```r
# page 22

summary(prostate.lm)$coef
```

```
##                 Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)  0.494154754 0.873566764  0.5656749 5.730382e-01
## lcavol       0.569546032 0.085847096  6.6344240 2.461450e-09
## lweight      0.614419817 0.198449499  3.0961016 2.622121e-03
## age         -0.020913467 0.010977742 -1.9050792 5.999883e-02
## lbph         0.097352535 0.057584215  1.6906115 9.441057e-02
## svi          0.752397342 0.238179913  3.1589454 2.163265e-03
## lcp         -0.104959408 0.089346934 -1.1747399 2.432323e-01
## pgg45        0.005324465 0.003384528  1.5731780 1.192260e-01
```

```r
# H0: bt1=0 vs HA: bt1 != 0
```

```r
T1 = 0.56954 / 0.08584

P1 = 2 * (1 - pt(abs(T1), 89))

print(c(T1,P1))
```

```
## [1] 6.634902e+00 2.456119e-09
```

```r
# page 24

Y.hat = X %*% beta

sigma.hat = sqrt(sum((Y - Y.hat)^2) / (n - ncol(X)))

cov.beta = sigma.hat^2 * solve(t(X) %*% X)

cov.beta
```

```
##              (Intercept)        lcavol       lweight           age
## (Intercept)  0.763118892  9.968185e-03 -1.127145e-01 -5.757711e-03
## lcavol       0.009968185  7.369724e-03 -3.267921e-03 -1.303787e-04
## lweight     -0.112714488 -3.267921e-03  3.938220e-02 -4.088331e-04
## age         -0.005757711 -1.303787e-04 -4.088331e-04  1.205108e-04
## lbph         0.024821341  3.981644e-04 -4.540916e-03 -1.471652e-04
## svi          0.010342177 -2.643164e-03 -4.105065e-03 -7.340636e-05
## lcp          0.002077485 -3.502258e-03  6.198759e-05  1.429198e-04
## pgg45        0.000033477  7.077893e-06  6.876251e-05 -9.141775e-06
##                      lbph           svi           lcp         pgg45
## (Intercept)  2.482134e-02  1.034218e-02  2.077485e-03  3.347700e-05
## lcavol       3.981644e-04 -2.643164e-03 -3.502258e-03  7.077893e-06
## lweight     -4.540916e-03 -4.105065e-03  6.198759e-05  6.876251e-05
## age         -1.471652e-04 -7.340636e-05  1.429198e-04 -9.141775e-06
## lbph         3.315942e-03  2.136331e-03 -1.416010e-04 -1.238514e-05
## svi          2.136331e-03  5.672967e-02 -8.863127e-03 -5.227255e-05
## lcp         -1.416010e-04 -8.863127e-03  7.982875e-03 -1.368737e-04
## pgg45       -1.238514e-05 -5.227255e-05 -1.368737e-04  1.145503e-05
```

```r
sqrt(diag(cov.beta))
```

```
## (Intercept)      lcavol     lweight         age        lbph         svi
## 0.873566764 0.085847096 0.198449499 0.010977742 0.057584215 0.238179913
##         lcp       pgg45
## 0.089346934 0.003384528
```

```r
vcov(prostate.lm)
```

```
##              (Intercept)        lcavol       lweight           age
## (Intercept)  0.763118892  9.968185e-03 -1.127145e-01 -5.757711e-03
## lcavol       0.009968185  7.369724e-03 -3.267921e-03 -1.303787e-04
## lweight     -0.112714488 -3.267921e-03  3.938220e-02 -4.088331e-04
## age         -0.005757711 -1.303787e-04 -4.088331e-04  1.205108e-04
```

```
## lbph         0.024821341  3.981644e-04 -4.540916e-03 -1.471652e-04
## svi          0.010342177 -2.643164e-03 -4.105065e-03 -7.340636e-05
## lcp          0.002077485 -3.502258e-03  6.198759e-05  1.429198e-04
## pgg45        0.000033477  7.077893e-06  6.876251e-05 -9.141775e-06
##                      lbph           svi           lcp         pgg45
## (Intercept)  2.482134e-02  1.034218e-02  2.077485e-03  3.347700e-05
## lcavol       3.981644e-04 -2.643164e-03 -3.502258e-03  7.077893e-06
## lweight     -4.540916e-03 -4.105065e-03  6.198759e-05  6.876251e-05
## age         -1.471652e-04 -7.340636e-05  1.429198e-04 -9.141775e-06
## lbph         3.315942e-03  2.136331e-03 -1.416010e-04 -1.238514e-05
## svi          2.136331e-03  5.672967e-02 -8.863127e-03 -5.227255e-05
## lcp         -1.416010e-04 -8.863127e-03  7.982875e-03 -1.368737e-04
## pgg45       -1.238514e-05 -5.227255e-05 -1.368737e-04  1.145503e-05
```

```r
# page 25

CI.lm = function(cur.lm, a, level=0.95, extra=0) {

 # the center of the confidence interval

 center = sum(a*cur.lm$coef)

 # the estimate of sigma^2

 sigma.sq = sum(resid(cur.lm)^2) / cur.lm$df.resid

 # the standard error of sum(a*cur.lm$coef)

 se = sqrt(extra * sigma.sq + sum((a %*% vcov(cur.lm)) * a))

 # the degrees of freedom for the t-statistic

 df = cur.lm$df

 # the quantile used in the confidence interval

 q = qt((1 - level)/2, df, lower.tail=FALSE)

 # upper, lower limits

 upr = center + se * q

 lwr = center - se * q

 return(data.frame(center, lwr, upr))

}
```

```r
# 95% CI for E(y|xs) with xs=(lcavol=1.3, lweight=3.6,age=64,lbph=0.1,svi=0.2,lcp=-0.2,pgg45=25)

print(CI.lm(prostate.lm, c(1, 1.3, 3.6, 64, 0.1, 0.2, -0.2, 25)))
```

```
##     center      lwr      upr
## 1 2.422332 2.281218 2.563447
```

```r
predict(prostate.lm, list(lcavol=1.3, lweight=3.6,age=64,lbph=0.1,svi=0.2,lcp=-0.2,pgg45=25), interval=
```

```
##        fit      lwr      upr
## 1 2.422332 2.281218 2.563447
```

```r
# page 26

CI.lm(prostate.lm, c(0,0,0,2,0,0.5,0,0))
```

```
##      center        lwr       upr
## 1 0.3343717 0.09496226 0.5737812
```

```r
case1 = prostate[10,]

case2 = case1

case2['age'] = case2['age'] + 2

case2['svi'] = case2['svi'] + 0.5

rbind(case1, case2)
```

```
##        lcavol  lweight age      lbph svi       lcp gleason pgg45     lpsa train
## 10  0.2231436 3.244544  63 -1.386294 0.0 -1.386294       6     0 1.047319 FALSE
## 101 0.2231436 3.244544  65 -1.386294 0.5 -1.386294       6     0 1.047319 FALSE
```

```r
Yhat = predict(prostate.lm, rbind(case1, case2))

diff(Yhat)
```

```
##       101
## 0.3343717
```

```r
# page 29

f.test.lm = function(R.lm, F.lm) {

 SSE.R = sum(resid(R.lm)^2)

 SSE.F = sum(resid(F.lm)^2)

 df.num = R.lm$df - F.lm$df

 df.den = F.lm$df

 F = ((SSE.R - SSE.F) / df.num) / (SSE.F / df.den)

 p.value = 1 - pf(F, df.num, df.den)

 return(data.frame(F, df.num, df.den, p.value))
```

```
}



reduced.lm = lm(lpsa ~ lcavol + lweight + age + lbph + svi, data=prostate)

print(f.test.lm(reduced.lm, prostate.lm))
```

```
##          F df.num df.den   p.value
## 1 1.372057      2     89 0.2588958
```

```
anova(reduced.lm, prostate.lm)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ lcavol + lweight + age + lbph + svi
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     91 44.437
## 2     89 43.108  2    1.3291 1.3721 0.2589
```

```
anova(prostate.lm, reduced.lm)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     89 43.108
## 2     91 44.437 -2   -1.3291 1.3721 0.2589
```

```
# page 32

prostate$Z = prostate$lcavol + prostate$svi

equal.lm = lm(lpsa ~ Z + lweight + age + lbph + lcp + pgg45, data=prostate)

f.test.lm(equal.lm, prostate.lm)
```

```
##           F df.num df.den   p.value
## 1 0.4818657      1     89 0.4893864
```

```
anova(equal.lm, prostate.lm)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ Z + lweight + age + lbph + lcp + pgg45
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     90 43.341
## 2     89 43.108  1   0.23339 0.4819 0.4894
```

```r
anova(prostate.lm)
```

```
## Analysis of Variance Table
##
## Response: lpsa
##           Df Sum Sq Mean Sq  F value    Pr(>F)
## lcavol     1 69.003  69.003 142.4636 < 2.2e-16 ***
## lweight    1  7.173   7.173  14.8086 0.0002235 ***
## age        1  0.646   0.646   1.3330 0.2513641
## lbph       1  0.809   0.809   1.6702 0.1995770
## svi        1  5.851   5.851  12.0798 0.0007900 ***
## lcp        1  0.130   0.130   0.2692 0.6051404
## pgg45      1  1.199   1.199   2.4749 0.1192260
## Residuals 89 43.108   0.484
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# page 33

prostate$Z2 = prostate$lcavol - prostate$svi

prostate$Y2 = prostate$lpsa- prostate$svi

constrained.lm = lm(lpsa ~ Z2 + lweight + age + lbph + lcp + pgg45, data=prostate, offset=svi)

constrained.lm2 = lm(Y2 ~ Z2 + lweight + age + lbph + lcp + pgg45, data=prostate)

coef(constrained.lm)
```

```
##  (Intercept)           Z2      lweight          age         lbph          lcp
##  0.382975608  0.543672792  0.654779620 -0.019797945  0.083478678 -0.037271152
##        pgg45
##  0.005571861
```

```r
coef(constrained.lm2)
```

```
##  (Intercept)           Z2      lweight          age         lbph          lcp
##  0.382975608  0.543672792  0.654779620 -0.019797945  0.083478678 -0.037271152
##        pgg45
##  0.005571861
```

```r
anova(constrained.lm, prostate.lm)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ Z2 + lweight + age + lbph + lcp + pgg45
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     90 43.961
## 2     89 43.108  1   0.85359 1.7623 0.1877
```

```r
anova(constrained.lm2, prostate.lm)
```

```
## Warning in anova.lmlist(object, ...): models with response '"lpsa"' removed
## because response differs from model 1

## Analysis of Variance Table
##
## Response: Y2
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Z2         1 37.663  37.663 77.1051 9.831e-14 ***
## lweight    1  7.619   7.619 15.5975 0.0001554 ***
## age        1  0.569   0.569  1.1659 0.2831319
## lbph       1  1.129   1.129  2.3109 0.1319718
## lcp        1  0.161   0.161  0.3296 0.5673017
## pgg45      1  1.317   1.317  2.6956 0.1041138
## Residuals 90 43.961   0.488
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
C.mat = matrix(0, 2, 8)

C.mat[1,7] = 1

C.mat[2,8] = 1

print(C.mat)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
## [1,]    0    0    0    0    0    0    1    0
## [2,]    0    0    0    0    0    0    0    1
```

```r
#install.packages('car')

library(car)
```

```
## Loading required package: carData
```

```r
linearHypothesis(prostate.lm, C.mat, test='F')
```

```
## Linear hypothesis test
##
## Hypothesis:
## lcp = 0
## pgg45 = 0
##
## Model 1: restricted model
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     91 44.437
## 2     89 43.108  2    1.3291 1.3721 0.2589
```

```r
anova(reduced.lm, prostate.lm)
```

```
## Analysis of Variance Table
##
## Model 1: lpsa ~ lcavol + lweight + age + lbph + svi
## Model 2: lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     91 44.437
## 2     89 43.108  2    1.3291 1.3721 0.2589
```

```r
head(prostate)
```

```
##       lcavol  lweight age      lbph svi       lcp gleason pgg45       lpsa
## 1 -0.5798185 2.769459  50 -1.386294   0 -1.386294       6     0 -0.4307829
## 2 -0.9942523 3.319626  58 -1.386294   0 -1.386294       6     0 -0.1625189
## 3 -0.5108256 2.691243  74 -1.386294   0 -1.386294       7    20 -0.1625189
## 4 -1.2039728 3.282789  58 -1.386294   0 -1.386294       6     0 -0.1625189
## 5  0.7514161 3.432373  62 -1.386294   0 -1.386294       6     0  0.3715636
## 6 -1.0498221 3.228826  50 -1.386294   0 -1.386294       6     0  0.7654678
##   train          Z         Z2         Y2
## 1  TRUE -0.5798185 -0.5798185 -0.4307829
## 2  TRUE -0.9942523 -0.9942523 -0.1625189
## 3  TRUE -0.5108256 -0.5108256 -0.1625189
## 4  TRUE -1.2039728 -1.2039728 -0.1625189
## 5  TRUE  0.7514161  0.7514161  0.3715636
## 6  TRUE -1.0498221 -1.0498221  0.7654678
```

```r
prostate2<-prostate[,1:9]

head(prostate2)
```

```
##       lcavol  lweight age      lbph svi       lcp gleason pgg45       lpsa
## 1 -0.5798185 2.769459  50 -1.386294   0 -1.386294       6     0 -0.4307829
## 2 -0.9942523 3.319626  58 -1.386294   0 -1.386294       6     0 -0.1625189
## 3 -0.5108256 2.691243  74 -1.386294   0 -1.386294       7    20 -0.1625189
## 4 -1.2039728 3.282789  58 -1.386294   0 -1.386294       6     0 -0.1625189
## 5  0.7514161 3.432373  62 -1.386294   0 -1.386294       6     0  0.3715636
## 6 -1.0498221 3.228826  50 -1.386294   0 -1.386294       6     0  0.7654678
```

```r
prostate2<-scale(prostate2)

summary(prostate2)
```

```
##      lcavol            lweight              age               lbph
##  Min.   :-2.28833   Min.   :-2.92718   Min.   :-3.0713   Min.   :-1.0247
##  1st Qu.:-0.71031   1st Qu.:-0.59070   1st Qu.:-0.5193   1st Qu.:-1.0247
##  Median : 0.08222   Median :-0.01385   Median : 0.1523   Median : 0.1377
##  Mean   : 0.00000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.65927   3rd Qu.: 0.57761   3rd Qu.: 0.5553   3rd Qu.: 1.0048
##  Max.   : 2.09651   Max.   : 2.68770   Max.   : 2.0327   Max.   : 1.5343
##       svi               lcp             gleason            pgg45
```

```
##  Min.   :-0.5229   Min.   :-0.8632   Min.   :-1.0422   Min.   :-0.8645
##  1st Qu.:-0.5229   1st Qu.:-0.8632   1st Qu.:-1.0422   1st Qu.:-0.8645
##  Median :-0.5229   Median :-0.4428   Median : 0.3426   Median :-0.3326
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.:-0.5229   3rd Qu.: 0.9712   3rd Qu.: 0.3426   3rd Qu.: 0.5538
##  Max.   : 1.8925   Max.   : 2.2053   Max.   : 3.1122   Max.   : 2.6811
##       lpsa
##  Min.   :-2.5202
##  1st Qu.:-0.6469
##  Median : 0.0980
##  Mean   : 0.0000
##  3rd Qu.: 0.5007
##  Max.   : 2.6895
```

```
apply(prostate2, 2, sd)
```

```
##  lcavol lweight     age    lbph     svi     lcp gleason   pgg45    lpsa
##       1       1       1       1       1       1       1       1       1
```

```
prostate.lm2 = lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45, data=as.data.frame(prostate
```

```
summary(prostate.lm)$coef
```

```
##                Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)  0.494154754 0.873566764  0.5656749 5.730382e-01
## lcavol       0.569546032 0.085847096  6.6344240 2.461450e-09
## lweight      0.614419817 0.198449499  3.0961016 2.622121e-03
## age         -0.020913467 0.010977742 -1.9050792 5.999883e-02
## lbph         0.097352535 0.057584215  1.6906115 9.441057e-02
## svi          0.752397342 0.238179913  3.1589454 2.163265e-03
## lcp         -0.104959408 0.089346934 -1.1747399 2.432323e-01
## pgg45        0.005324465 0.003384528  1.5731780 1.192260e-01
```

```
summary(prostate.lm2)$coef
```

```
##                 Estimate Std. Error       t value     Pr(>|t|)
## (Intercept) -8.348482e-16 0.06121618 -1.363771e-14 1.000000e+00
## lcavol       5.815336e-01 0.08765397  6.634424e+00 2.461450e-09
## lweight      2.280323e-01 0.07365144  3.096102e+00 2.622121e-03
## age         -1.348863e-01 0.07080353 -1.905079e+00 5.999883e-02
## lbph         1.223565e-01 0.07237413  1.690611e+00 9.441057e-02
## svi          2.698439e-01 0.08542214  3.158945e+00 2.163265e-03
## lcp         -1.271383e-01 0.10822678 -1.174740e+00 2.432323e-01
## pgg45        1.300941e-01 0.08269510  1.573178e+00 1.192260e-01
```

```
#Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code t
```