Part A. Exercises (Faraway book)

Ch.7 Ex.3  Read in data, fit the model and transform it.
> library(MASS)
> library(faraway)
> data(ozone)
> a <- boxcox(lm(O3~., data = ozone), lambda = seq(0, .5, by = .05))

![Graph of box-cox transformation]

Figure 1: This figure shows the box-cox transformation of the response.

A cube-root or fourth-root transformation ($\lambda = 0.33$ or $0.25$) would work fine according to the CI from the Box-Cox method.

Ch.8 Ex.1  The following code is used to determine the best model in each case.
Backward Elimination: model = lcavol + lweight + svi

> data(prostate)
> g <- lm(lpsa~., data = prostate)
> summary(g)

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 0.669337 | 1.296387 | 0.516 | 0.60693 |
| lcavol | 0.587022 | 0.087920 | 6.677 | 2.11e-09 *** |
| lweight | 0.454467 | 0.170012 | 2.673 | 0.00896 ** |
| age | -0.019637 | 0.011173 | -1.758 | 0.08229 . |
| lbph | 0.107054 | 0.058449 | 1.832 | 0.07040 . |
| svi | 0.766157 | 0.244309 | 3.136 | 0.00233 ** |
| lcp | -0.105474 | 0.091013 | -1.159 | 0.24964 . |
| gleason | 0.045142 | 0.157465 | 0.287 | 0.77503 |
| pgg45 | 0.004525 | 0.004421 | 1.024 | 0.30886 |
Remove gleason

```r
> summary(lm(lpsa ~ lcavol + lweight + age + lbph + svi + lcp + pgg45, data = prostate))
Coefficients:

| Estimate  | Std. Error | t value | Pr(>|t|) |
|-----------|------------|---------|----------|
| (Intercept) | 0.953926  | 0.829439 | 1.150   | 0.25319 |
| lcavol     | 0.591615  | 0.086001 | 6.879   | 8.07e-10 *** |
| lweight    | 0.448292  | 0.167771 | 2.672   | 0.00897 ** |
| age        | -0.019336 | 0.011066 | -1.747  | 0.08402 . |
| lbph       | 0.107671  | 0.058108 | 1.853   | 0.06720 . |
| svi        | 0.757734  | 0.241282 | 3.140   | 0.00229 ** |
| lcp        | -0.104482 | 0.090478 | -1.155  | 0.25127 |
| pgg45      | 0.005318  | 0.003433 | 1.549   | 0.12488 |

...continue this by removing variables until only significant variables (0.05 level) remain....

> summary(lm(lpsa ~ lcavol + lweight + svi, data = prostate))
Coefficients:

| Estimate  | Std. Error | t value | Pr(>|t|) |
|-----------|------------|---------|----------|
| (Intercept) | -0.26809  | 0.54350 | -0.493  | 0.62298 |
| lcavol     | 0.55164   | 0.07467 | 7.388   | 6.3e-11 *** |
| lweight    | 0.50854   | 0.15017 | 3.386   | 0.00104 ** |
| svi        | 0.66616   | 0.20978 | 3.176   | 0.00203 ** |

AIC: model = lcavol + lweight + age + lbph + svi

> step(g)
Start:  AIC=-58.32

```r
Coefficients:

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>lcavol</th>
<th>lweight</th>
<th>age</th>
<th>lbph</th>
<th>svi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95100</td>
<td>0.56561</td>
<td>0.42369</td>
<td>-0.01489</td>
<td>0.11184</td>
<td>0.72095</td>
</tr>
</tbody>
</table>
```

$R^2$: model = lcavol + lweight + age + lbph + svi + lcp + pgg45

$C_p$: model = lcavol + lweight + age + lbph + svi

> library(leaps)
> gs <- regsubsets(lpsa~., prostate)
> (gsum <- summary(gs))

<table>
<thead>
<tr>
<th>lcavol</th>
<th>lweight</th>
<th>age</th>
<th>lbph</th>
<th>svi</th>
<th>gleason</th>
<th>pgg45</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>2 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>3 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>4 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>5 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>6 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>7 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
<tr>
<td>8 (1)</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
<td>&quot;*&quot;</td>
</tr>
</tbody>
</table>

> (1:8)[which.max(gsum$adjr2)] ## Get best $R_a^2$
[1] 7
> min((1:8)[gsum$cp<c(2:9)]) ## Get smallest $C_p$ where $C_p < p$
[1] 5

Thus, maximum adjusted $R^2$ criterion suggests a model with 7 predictors, while the minimum $C_p$ criterion suggests a model with 5 predictors.
Part B. Exercises (S & Y book)

Ch. 5 Ex. 4  (a) Variance of estimates are calculated via

\[
X = \begin{pmatrix}
1 & x_{11} & x_{12} \\
1 & x_{21} & x_{22} \\
\vdots & \vdots & \vdots \\
1 & x_{n1} & x_{n2}
\end{pmatrix}, \quad X'X = \begin{pmatrix} n & 0 & 0 \\
0 & n & k \\
0 & k & n 
\end{pmatrix}, \quad (X'X)^{-1} = \frac{1}{n(n^2-k^2)} \begin{pmatrix} n^2-k^2 & 0 & 0 \\
0 & n^2 & -nk \\
0 & -nk & n^2 
\end{pmatrix}.
\]

In order to calculate VIFs, standardize the data matrix \(X\) further. Let \(Z = X/\sqrt{n}\) so that \(Z\) is standardized data matrix. VIFs are \((1, \frac{n^2}{n^2-k^2}, \frac{n^2}{n^2-k^2})\) from following result.

\[
(Z'Z)^{-1} = \frac{1}{1-k^2/n^2} \begin{pmatrix} 1-k^2/n^2 & 0 & 0 \\
0 & 1/k^2 & -k/n \\
0 & -k/n & 1 \end{pmatrix}.
\]

(b) From the equation (characteristic function) \(\text{det}(\lambda I - X'X) = (\lambda - n)(\lambda - n + k)(\lambda - n - k) = 0\), eigenvalues of \(X'X\) are \(n+k\), \(n\), and \(n-k\). Thus the ratio of the largest and the smallest eigenvalues is \(\frac{n+k}{n-k}\). Condition indices are \(\eta_1 = 1, \eta_2 = (\frac{n+k}{n-k})^1/2\) and \(\eta_3 = (\frac{n+k}{n-k})^{1/2}\).

(c) Write \(a = (a_1, a_2, a_3)'\), \(\beta = (\beta_0, \beta_1, \beta_2)'\). The problem is to find a vector \(a\) which maximizes \(\text{Var}(\sum a_j\beta_j) = \text{Var}(a'\beta) = a'\Sigma a\) where \(\Sigma = (X'X)^{-1}\sigma^2\) subject to \(|a| = 1\). Let \(\lambda\) be a Lagrange multiplier,

\[
\Gamma(a, \lambda) = a'\Sigma a - \lambda(a'a - 1)
\]

\[
d\Gamma/da = 2\Sigma a - 2\lambda a = 0 \iff \Sigma a = \lambda a
\]

Thus \(\lambda\) is an eigenvalue of \(\Sigma\) with corresponding eigenvector \(a\). Since \(\text{Var}(\sum a_j\beta_j) = a'\Sigma a = a'\lambda a = \lambda\), thus maximizing vector, \(a\), corresponds with the largest eigenvalue \(\lambda\). The largest eigenvalue of \(\Sigma\) is \(\sigma^2/(n-k)\) (among \(\sigma^2/n, \sigma^2/(n+k), \sigma^2/(n-k)\)) with eigenvector \(a = (0, 1/\sqrt{2}, -1/\sqrt{2})\). Variance is maximized as \(\sigma^2/(n-k)\). When \(k = 0\), \(\text{Var}(a'\beta) = \sigma^2/n\) so variance is \(n/(n-k)\) times inflated.

(d)

\[
X'X + cI = \begin{pmatrix} n+c & 0 & 0 \\
0 & n+c & k \\
0 & k & n+c \end{pmatrix},
\]

\[
(X'X + cI)^{-1} = \frac{1}{(n+c)((n+c)^2-k^2)} \begin{pmatrix} (n+c)^2-k^2 & 0 & 0 \\
0 & (n+c)^2 & -(n+c)k \\
0 & -(n+c)k & (n+c)^2 \end{pmatrix}.
\]

\[
\hat{\beta}_1^{(c)} = (X'X + cI)^{-1}X'Y = \frac{1}{(n+c)^2-k^2} \left( (n+c) \sum x_{1j}y_j - k \sum x_{2j}y_j \right),
\]

\[
\text{Var}(\hat{\beta}_1^{(c)}) = \frac{\sigma^2}{(n+c)^2-k^2} \left( (n+c)^2n + k^2n - 2(n+c)k^2 \right) \left( (n+c)^2-k^2 \right)^2 \sigma^2
\]

(e) From the fact that bias = \(-c(X'X + cI)^{-1}\beta\), the result is easily calculated.

(f) Evaluating mean squared error using results from (d) and (e), one two-step grid search will give \(c = 4.4\) as the optimal value with MSE 0.0799. Sample code for this is attached at the end of solution.

Ch. 5 Ex. 5  (a) Statistician fits the model \(Y = X_1\beta_1 + \epsilon\). LSE of \(\beta_1\) is \(\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y\), but the true model is \(Y = X_1\beta_1 + X_2\beta_2 + \epsilon\). So \(EY = (X_1 \ X_2) \begin{pmatrix} \beta_1 \\
\beta_2 \end{pmatrix} \)

\[
E(\hat{\beta}_1) = (X_1'X_1)^{-1}X_1' (X_1 \ X_2) \begin{pmatrix} \beta_1 \\
\beta_2 \end{pmatrix} = \beta_1 + (X_1'X_1)^{-1}X_1'X_2 \beta_2
\]
Thus, $\tilde{\beta}_1$ is biased with bias $(X'X)^{-1}X'\beta_2$.

(b) Let $H_1 = X_j(X_j'X_j)^{-1}X_j'$ then $(I - H_1)X_j = 0$ and $(I - H_1)$ is symmetric and idempotent. Now $Y - \tilde{Y} = (I - H_1)Y = (I - H_1)(X_j\beta_1 + X_2\beta_2 + \epsilon) = (I - H_1)X_2\beta_2 + (I - H_1)\epsilon$. So,

$$R = (Y - \tilde{Y})'(Y - \tilde{Y}) = 2\epsilon'(I - H_1)'((I - H_1)X_2\beta_2) + \beta_2'X_2'(I - H_1)X_2\beta_2 + \epsilon'(I - H_1)\epsilon$$

Thus, MSPE$_d = E(\epsilon^* - \epsilon)^2 + \epsilon'(I - H_1)\epsilon$. For the calculation of $\delta$, note that $\delta^2 = |(I - H_1)X_2\beta_2|^2 |(I - H_1)X_2\beta_2| = \beta_2'X_2'(I - H_1)X_2\beta_2 = \beta_2^2-C\beta_2$.

(c) Now let $X = (X_1 X_2)$, $\beta = (\beta_1' \beta_2')'$ so LSE $\hat{\beta} = (X'X)^{-1}X'Y$ and $\tilde{Y} = X(X'X)^{-1}X'Y = HY$.

Thus, MSPE$_e = E(\epsilon^* - \epsilon)^2 + \epsilon'(I - H)\epsilon$. For the calculation of $\delta$, note that $\delta^2 = |(I - H)X_2\beta_2|^2 |(I - H)X_2\beta_2| = \beta_2'X_2'(I - H)X_2\beta_2 = \beta_2^2-C\beta_2$.

Thus, MSPE$_d <$ MSPE$_e$ iff $\beta_2'X_2 < p_2\sigma^2$.

Ch.5 Ex.7 (a) Forward variable selection is the easiest and straightforward method in this case. Beginning with an empty model, one can choose and include one variable whose inclusion results in the largest reduction in SSE. If that reduction is not significant, then stop the selection step. (F test is used.)

1. When $p = 1$ (intercept), $SSE_1 = 185008830$, $d.f = 16 - 1$.

2. When $p = 2$, inclusion of $x2$ results in largest reduction in SSE. $SSE_2 = 6036140.1661$ with $d.f = 16 - 2$. With $F = \frac{(SSE_1 - SSE_2)/1}{SSE_2/(16 - 2)} = 415.1026 \sim F_{1,14}$ under $H_0$ : model in step 1. $p$-value $< 0.0001$.

3. $p = 3$, $x2, x3$, $SSE_3 = 3579064$, $d.f = 16 - 3$, $F = 8.9247$, $p$-value $= 0.0098$.

4. $p = 4$, $x2, x3, x4$, $SSE_4 = 2756711$, $d.f = 16 - 4$, $F = 3.5797$, $p$-value $= 0.079$.

5. $p = 5$, $x2, x3, x4, x6$, $SSE_5 = 858680$, $d.f = 16 - 5$, $F = 24.3$, $p$-value $= 0.0002$.

6. $p = 6$, $x2, x3, x4, x6, x5$, $SSE_6 = 839348$, $d.f = 16 - 6$, $F = 0.23$, $p$-value $= 0.6387$. Inclusion of $x5$ is not significant. Stop the procedure.
One would choose (const., $x_2, x_3, x_4, x_6$).

(b) According to the result of Table 5.25, all variables except $x_1$ are significant. But $x_1$ called GNP deflator is an important factor in predicting $Y$ (total derived employment) and we suspect multicollinearity between variables because GNP deflator ($x_1$) is highly relative to GNP ($x_2$) and year ($x_6$). We also observe high correlation between $x_1$, $x_2$ and $x_6$ and VIFs are the following:

$$
x_1 : 76.6414, \ x_2 : 546.8705, \ x_3 : 14.2896, \ x_4 : 3.4608, \ x_6 : 644.6264.
$$

Moreover, Fig 5.11 shows that two residual plots against year and fitted value are similar. It also suggests multicollinearity of $x_6$ and nonlinear evidence. We may need transformation of $x_6$ and remedies for multicollinearity such as ridge regression.

...There is no strict solution for this problem but you may mention the following:

- Statistical comment includes Normality, patterns on residual plots, insignificance of variable $x_1$ of the model.
- Explanation to the economist includes the relationship of variables based on regression performed, especially meaning of estimates.
- Further analysis might be multicollinearity checking, regressing more or less variables, model diagnostics and so on.

**Ch.5 Ex.9**  
(a) Linear regression fitting

Call:

```
lm(formula = sal ~ lagsal + fflow + period + year)
```

Coefficients:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -831.88289  448.54450  -1.855  0.0765 .
  lagsal     0.61639    0.11860   5.197  2.86e-05 ***
    fflow    -0.26339    0.10299  -2.557  0.0176 *
    period   0.06002    0.15987   0.375  0.7108
     year    0.42648    0.22733   1.876  0.0734 .
```

---

Residual standard error: 1.266 on 23 degrees of freedom  
Multiple R-squared: 0.8494, Adjusted R-squared: 0.8232  
F-statistic: 32.44 on 4 and 23 DF, p-value: 3.768e-09

(b) Model selection

```
lm(formula = sal ~ lagsal + fflow + year)
```

Coefficients:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -783.28581  421.70351  -1.857  0.07556 .
  lagsal     0.62204    0.11552   5.385  1.57e-05 ***
    fflow    -0.28218    0.08839  -3.192  0.00391 **
     year    0.40214    0.21395   1.880  0.07236 .
```

---

Residual standard error: 1.243 on 24 degrees of freedom  
Multiple R-squared: 0.8485, Adjusted R-squared: 0.8296  
F-statistic: 44.81 on 3 and 24 DF, p-value: 3.768e-09

(c) Transformation

At the result of applying boxcox transformation, we can consider the square root or no transformation. But $\lambda = 1$ is selected though we obtain the smaller AIC value when lambda is 0.5.
(d) Detecting influential values and outliers
(e) Fitted model (removing obs. 16 as outlier)

\[
\text{lm(formula = sal ~ lagsal + fflow, data = data)}
\]

Coefficients:

|                     | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | 16.24142 | 2.87288    | 5.653   | 8.04e-06 *** |
| lagsal              | 0.71311  | 0.07155    | 9.967   | 5.24e-10 *** |
| fflow               | -0.55841 | 0.10761    | -5.189  | 2.58e-05 *** |

***Attached is the sample code which answers to this question.

data <- read.table('D:/2007-Fall/664 - solution/Hw4-664/salinity.dat',header=F)
names(data) <-c("obs","sal","lagsal","fflow","period","year")
attach(data)
summary(lm(sal~lagsal+fflow+period+year))
n <-length(obs)

##Variable Selection
bm.step <- step(lm(sal~1), ~lagsal+fflow+period+year)
summary(bm.step)

##Checking any need of transformation (package "MASS" loaded)
library(MASS)
boxcox(bm.step,lambda = seq(-.5,3,0.1))

## candidate transform : sqrt or nothing

## Comparison between sqrt or nothing
gmean<-1;for(i in 1:n){gmean <- gmean*sal[i];gmean<-gmean^((1/n)
lambda <- 0.5;p<-3
h_sal <- (gmean^((1-lambda)) *((sal^lambda-1)/lambda)
SSE <- (summary(lm(h_sal ~ lagsal+fflow+year))$sigma)^2/(n-p)
AICsqrt<-n*log(SSE/n) +2*p
BICsqrt<-n*log(SSE/n) +p*log(n)

SSE <- (summary(bm.step)$sigma)^2/(n-p)
AICnull<-n*log(SSE/n) +2*p
BICnull<-n*log(SSE/n) +p*log(n)

## No transform (Though transformation gives less AIC)
lm<-lm(sal~lagsal + fflow + year)

## Diagnostics for influential values and outliers
sigma<-summary(lm)$sigma
inf<- lm.influence(lm)
shapiro.test(lm$residual)
par(mfrow=c(1,3))

## 1. internally standadized residual plot
Internally.Studentized.Residual<-(lm$residuals/
(sigma*sqrt(1-inf$hat))
plot(Internally.Studentized.Residual);title("Internally Standardized Residual")
abline(h=c(-2,2),lty="dotted")
for(i in 1:n){
  if(abs(Internally.Studentized.Residual[i])> 2)text(i+1,Internally.Studentized.Residual[i],i,cex=0.6)
## 2. externally studentized residual plot

```r
Externally.Studentized.Residual <- lm$residuals * 
  sqrt(((n-p-1)/((1-inf$hat)*(n-p)*sigma^2 
           - (lm$residuals)^2)) 
plot(Externally.Studentized.Residual); title("Externally Studentized Residual") 
abline(h=c(-2,2), lty="dotted") 
for(i in 1:n){ 
  if(abs(Externally.Studentized.Residual[i]) > 2) text(i+1, Externally.Studentized.Residual[i], i, cex=0.6) 
}
```

## 3. qq plot

```r
qqnorm(Externally.Studentized.Residual); abline(c(0,0), c(1,1)) 
```

## 4. leverage - hat matrix

```r
plot(inf$hat); title("leverage plot") 
abline(h=2*p/n, lty=3) ### high leverage points 
leverage <- c(inf$hat > 2*p/n) 
for(i in 1:n){ if(leverage[i] == T) text(i+1, inf$hat[i], i, cex=0.6) 
}
```

## 5. DFFITS

```r
DFFITS <- Externally.Studentized.Residual * sqrt(inf$hat / (1-inf$hat)) 
plot(DFFITS); abline(h=2*sqrt(p/n), lty=2); abline(h=-2*sqrt(p/n), lty=2); title("DFFITS") 
DF.detected <- c(abs(DFFITS) > 2*sqrt(p/n)) 
for(i in 1:n){ if(DF.detected[i] == T) text(i+1, DFFITS[i], i, cex=0.6) 
}
```

## 6. Cook's D

```r
plot(lm, which=4) 
```

## 7. COVRATIO

```r
first <- ((n-p-1) / (n-p)) + Externally.Studentized.Residual^2 / (n-p) 
COVRATIO <- first^-p / (1-inf$hat) 
plot(COVRATIO); abline(h=3*p/n+1, lty=2); abline(h=-3*p/n+1, lty=2); title("COVRATIO") 
CR.detected <- c(abs(COVRATIO-1) > 3*p/n) 
for(i in 1:n){ if(CR.detected[i] == T) text(i+1, COVRATIO[i], i, cex=0.6) 
}
```

## obs. 16 seems to be outlier
## Delete obs. 16 from data matrix

```r
data <- subset(data, obs != 16) 
summary(lm2 <- lm(sal ~ lagsal + fflow + year, data)) # year is not significant 
summary(lm3 <- lm(sal ~ lagsal + fflow, data)) 
```

## 7. COVRATIO

```r
first <- ((n-p-1) / (n-p)) + Externally.Studentized.Residual^2 / (n-p) 
COVRATIO <- first^-p / (1-inf$hat) 
plot(COVRATIO); abline(h=3*p/n+1, lty=2); abline(h=-3*p/n+1, lty=2); title("COVRATIO") 
CR.detected <- c(abs(COVRATIO-1) > 3*p/n) 
for(i in 1:n){ if(CR.detected[i] == T) text(i+1, COVRATIO[i], i, cex=0.6) 
}
```

## let \( y_i^* \) be new sample from same covariate
## then standard error of \( \hat{y}_i^* \) is \( s \sqrt{x'_i (X'X)^{-1} x_i +1} \)
## where \( s^2 \) is mean squared error

```r
X = cbind(data$lagsal, data$fflow) 
Xstar = cbind(lagsal, fflow) 
se = c(1) 
for(i in 1:n) { 
  se[i] = summary(lm3)$sigma * sqrt(t(Xstar[i,])%*%solve(t(X)%*%X)%*%Xstar[i,] +1) 
}
se
```
Appendix: Sample code for Ex.4

n<-20;k<-19;s<-1;beta1=0.5
MSE <- c(0:10)
for(c in 0:10){
bias <- -c*(n+c)*beta1/((n+c)^2-k^2)
variance <- ((n+c)^2*n+k^2*n-2*(n+c)*k^2)/((n+c)^2-k^2)^2
MSE[c+1] <- bias^2 +variance
}
c(MSE==min(MSE))
MSE <- c(0:10)
for(i in 0:10){
c<-4.0+i/5
bias <- -c*(n+c)*beta1/((n+c)^2-k^2)
variance <- ((n+c)^2*n+k^2*n-2*(n+c)*k^2)/((n+c)^2-k^2)^2
MSE[i+1] <- bias^2 +variance
}
c(MSE==min(MSE))