Chapter 9

Selection of an Appropriate Model

Overview

- Possible procedures to find the best regression model
- All possible regression models
  - R^2
  - Mallow's C_p statistic
  - Others: MSE, AIC
- Stepwise regression
  - Partial F-test
- Forward selection
- Backward elimination
- Use of Statistical Software: SAS and R

9.1 Introduction

Two opposed criteria for selecting a model are usually involved:

1. To include as **many** predictors as possible so that reliable fitted values can be obtained.

2. To include as **few** predictors in the model as possible because of the cost in obtaining and monitoring the info.

The procedure to find a compromise between these two contradicting criteria is what is usually called selecting the best regression model.

Introduction (Continued)

Possible procedures to find the best regression model

- All possible regression models

- Stepwise regression
  - Forward selection
  - Backward elimination

9.2 All Possible Regression Models

- Suppose we have t predictor variables $x_1, \ldots, x_t$. Then there are 2^t possible models. They are

$$y = \beta_0 + \sum_{i=1}^{t} \delta_i \beta_i x_i + \epsilon$$

where $\delta_i = 0$ or 1.

- For example, with $t = 4$, the possible models are

$$y = \beta_0 + \epsilon, \qquad (\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0)$$

$$y = \beta_0 + \beta_1 x_1 + \epsilon, \qquad (\delta_1 = 1, \delta_2 = \delta_3 = \delta_4 = 0)$$

$$\vdots$$

$$y = \beta_0 + \beta_4 x_4 + \epsilon, \qquad (\delta_1 = \delta_2 = \delta_3 = 0, \delta_4 = 1)$$

(Continued)

All Possible Regression Models (Continued)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad (\delta_1 = \delta_2 = 1, \delta_3 = \delta_4 = 0)$$

$$\vdots$$

$$y = \beta_0 + \beta_3 x_3 + \beta_4 x_4 + \epsilon, \quad (\delta_1 = \delta_2 = 0, \delta_3 = \delta_4 = 1)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon,$$

$$(\delta_1 = \delta_2 = \delta_3 = 1, \delta_4 = 0)$$

$$\vdots$$

$$y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon,$$

$$(\delta_1 = 0, \delta_2 = \delta_3 = \delta_4 = 1)$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \epsilon,$$

$$(\delta_1 = \delta_2 = \delta_3 = \delta_4 = 1)$$

Criteria Used to Assess Models

Some criteria used to assess the models are

- The value of the coefficient of determination R^2 (= SSR/SST)
- The Mallows C_p statistic
- Others
  - Mean sum squares error (MSE)
  - Akaike Information Criterion (AIC)

9.2.1 Use of R^2

Consider the data set on the right hand side.

(It is known as the Hald data)

x_1	x_2	x_3	x_4	y
7	26	6	60	78.5
1	29	15	52	74.3
11	56	8	20	104.3
11	31	8	47	87.6
7	52	6	33	95.9
11	55	9	22	109.2
3	71	17	6	102.7
1	31	22	44	72.5
2	54	18	22	93.1
21	47	4	26	115.9
1	40	23	34	83.8
11	66	9	12	113.3
10	68	8	12	109.4

Use of R^2 (Continued)

Step 1:

Divide the possible models into sets based on the number of parameters in the model :

- Set A : only the constant parameter;
- Set B : four 1-predictor models;

  (i.e. $\mu_y = \beta_0 + \beta_i x_i, \quad i = 1, 2, 3, 4$)
- Set C : all the 2-predictor models;

  (i.e. $\mu_y = \beta_0 + \beta_i x_i + \beta_j x_j$)
- Set D : all the 3-predictor models and

  (i.e. $\mu_y = \beta_0 + \beta_i x_i + \beta_j x_j + \beta_k x_k$)
- Set E : the model with all 4 predictors.

Use of R^2 (Continued)

Step 2:

- Order the models within each set by the value of R^2.

Step 3:

- Examine the leaders and see if there is any consistent pattern in the predictors in the leading models in each set.

Use of R^2 (Continued)

- For Hald data, the leaders in each set are:

Set	Variables in Equation	$100R^2\%$
B	$\hat{y} = f(x_4)$	67.5%
C	$\hat{y} = f(x_1, x_2)$	97.9%
	$\hat{y} = f(x_1, x_4)$	97.2%
D	$\hat{y} = f(x_1, x_2, x_4)$	98.234%
E	$\hat{y} = f(x_1, x_2, x_3, x_4)$	98.237%

Use of R^2 (Continued)

- Note that R^2 increases when the number of predictors increases.

- The goal is to compare R^2 for alternative models so that we may determine when an introduction of an additional predictor does not result in a 'substantial' increase in R^2.

Use of R^2 (Continued)

- From the above results, we know that a two predictor model is already good enough.
- Examination of the correlation matrix for the data reveals that (x_1 and x_3) and (x_2 and x_4) are highly correlated, ($r_{13} = -0.824$ and $r_{24} = -0.973$).
- Hence it doesn't matter if we use $f(x_1, x_2)$ or $f(x_1, x_4)$.
- However, some people prefer $f(x_1, x_4)$ because it is consistent with the best single predictor model involving x_4.

9.2.2 Use of the Mallow's C_p Statistic

- The C_p statistic is defined as

$$C_p = \frac{SSE_p}{MSE_t} - (n - 2p)$$

where

- p = no. of parameters included in the model,
- SSE_p = error sum of squares of a model with p parameters
- t = total number of parameters to be considered for inclusion in the regression model,
- MSE_t = mean SSE with all t parameters,
- n = sample size.

Use of C_p Statistic (Continued)

- If a model with ***p* parameters** is adequate (i.e. no lack of fit), then

$$E(SSE_p) = (n - p)\sigma^2,$$

- Also $E(MSE_t) = \sigma^2$
- Therefore

$$E\left(\frac{SSE_p}{MSE_t}\right) \approx \frac{(n-p)\sigma^2}{\sigma^2} = n - p$$

- Hence if the model with parameters is adequate, then $E(C_p) \approx p$.
- The goal of the C_p criterion is to choose a model with a low C_p value about equal to p.

Use of C_p Statistic (Continued)

- For the Hald data, we have $n = 13$ and $MSE_t = 5.893$.
- For the model $y = \beta_0 + \beta_1 x_1 + \epsilon$, we have $\text{SSE}_1 = 1265.687$
- Hence $C_p = (1265.687/5.893) - (13 - 2(2)) = 202.5$.
- Since the C_p statistic for the model $y = f(x_1) + \epsilon$ is much larger than 2, therefore we suspect there is a lack of fit and hence the model is no good.

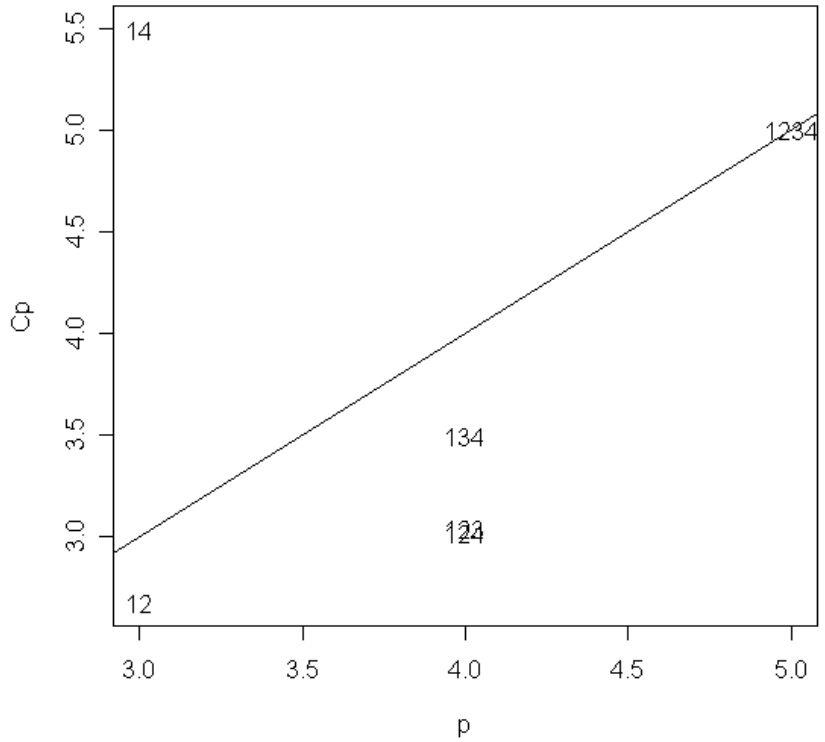
Use of C_p Statistic (Continued)

- All the C_p values for the Hald data are given as follows:

Variables in the equation	C_p	p
	443.2	1
1, 2, 3, 4	202.5, 142.5, 315.3, 138.7	2
(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)	2.7, 198.1, 5.5, 62.4, 138.2, 22.4	3
(1, 2, 3), (1, 2, 4), (1, 3, 4), (2, 3, 4)	3.0, 3.0, 3.5, 7.3	4
(1, 2, 3, 4)	5.0	5

Use of C_p Statistic (Continued)

- A plot of the smaller C_p against p is given below.



- On the basis of the C_p statistic, the model with x_1 and x_2 is preferred over all others.

9.2.3 SAS Program

- The following is a SAS program for all possible regressions:
- Procedure RSQUARE is used.

```
proc rsquare cp b;
  model y = x1 x2 x3 x4;
run;
```

Partial SAS Output

Number in Model	R-Square	C(p)	Parameter Estimates: Intercept	x1	x2	x3	x4
1	0.6745	138.7308	117.56793	.	.	.	-0.73816
1	0.6663	142.4864	57.42368	.	0.78912	.	.
1	0.5339	202.5488	81.47934	1.86875	.	.	.
1	0.2859	315.1543	110.20266	.	.	-1.25578	.
2	0.9787	2.6782	52.57735	1.46831	0.66225	.	.
2	0.9725	5.4959	103.09738	1.43996	.	.	-0.61395
2	0.9353	22.3731	131.28241	.	.	-1.19985	-0.72460
2	0.8470	62.4377	72.07467	.	0.73133	-1.00839	.
2	0.6801	138.2259	94.16007	.	0.31090	.	-0.45694
2	0.5482	198.0947	72.34899	2.31247	.	0.49447	.
3	0.9823	3.0182	71.64831	1.45194	0.41611	.	-0.23654
3	0.9823	3.0413	48.19363	1.69589	0.65691	0.25002	.
3	0.9813	3.4968	111.68441	1.05185	.	-0.41004	-0.64280
3	0.9728	7.3375	203.64196	.	-0.92342	-1.44797	-1.55704
4	0.9824	5.0000	62.40537	1.55110	0.51017	0.10191	-0.14406

9.2.4 R Program

- We need to install package “leaps”
- Click “Packages”
- Select a CRN mirror “Singapore”
- Select Packages “leaps”

R program

```
library(leaps)
Ch9ex1 <- read.table(“d:/ST3131/ch9ex1.txt”,header=T)
attach(ch9ex1)
Model1234 <- lm(y~x1+x2+x3+X4)
X <- model.matrix(model1234)[,-1]
```

R Program (Continued)

```
> # Get R^2 for all possible models
> Allrsq <- leaps(x,y,method="r2")
> # Select the max R^2 in each set (from one predictor
    to 4 predictors)
> for(i in 2:5){
+ maxr2 <- max(allrsq$r2[allrsq$size==i])
+ whichmodel <- allrsq$which[allrsq$r2==maxr2,]
+ namemodel <- names(whichmodel)[whichmodel==T]
+ cat(namemodel,"\n",maxr2,"\n")}
4
 0.674542
1 2
 0.9786784
1 2 4
 0.9823355
1 2 3 4
 0.9823756
```

R Program (Continued)

```
# Get Cp for all possible models
allCp <- leaps(x,y,method="Cp")
# Select the min Cp in each set (from one predictor to
# 4 predictors)
for(i in 2:5){
minCp <- min(allCp$Cp[allCp$size==i])
Whichmodel <- allCp$which[allCp$Cp==minCp,]
Namemodel <- names(whichmodel)[whichmodel==T]
cat(namemodel,"\n",minCp,"\n")}
4
 138.7308
1 2
 2.678242
1 2 4
 3.018233
1 2 3 4
 5
```

9.3 Stepwise Regression

- In this procedure, a predictor variable is added or deleted from the model at each step based on the observed values of the partial F-statistics.

- i.e. the test statistic used in testing the contribution of a predictor variable when some other variables are already included in the model.

Stepwise Regression (Continued)

Step 1

- Compute the correlations r_{y,x_i}, between y and x_i for all i's.

- Include the predictor, say, x_{a_1}, where $\left|r_{y,x_{a_1}}\right|$ is the largest among all the r_{y,x_i}'s

- Compute

$$F = \frac{MSR(x_{a_1})}{MSE(x_{a_1})}$$

Stepwise Regression (Continued)

Step 1 (Continued)

- If the largest of the observed F-value is greater than $F_{\alpha_1}(1, \nu_1)$, where ν_1 is the d.f. for the $MSE(x_{a_1})$, then retain x_{a_1} in the model, otherwise do not include any independent variable in the model.

- α_1 is called **the significance level for entry** into the model.

Stepwise Regression (Continued)

Step 2

- Compute

$$F_i = \frac{SSR(x_i | x_{a_1})}{MSE(x_i, x_{a_1})}$$

for each $i \neq a_1$

- If the largest of these observed F_i's is less than $F_{\alpha_1}(1, \nu_2)$, then stop and include only x_{a_1} in the model,
- otherwise include the predictor x_{a_2}, where F_{a_2} is the largest among all the F_i's.

Stepwise Regression (Continued)

Step 3

- Determine whether x_{a_1} can be deleted from the model by computing

$$F_{a_1} = \frac{SSR\left(x_{a_1} \middle| x_{a_2}\right)}{MSE\left(x_{a_1}, x_{a_2}\right)}$$

- If the observed $F_{a_1} < F_{\alpha_2}(1, \nu_2)$, then delete x_{a_1} from the model with x_{a_1} and x_{a_2},

- otherwise keep both x_{a_1} and x_{a_2} in the model.

Stepwise Regression (Continued)

Step 3 (Continued)

- α_2 is called **the significance level for staying** in the model.
- Note: $\alpha_1 = 0.15$ and $\alpha_2 = 0.15$ are used as default in the SAS Stepwise Procedure.

Step 4

- Repeat Steps 2 and 3 with F_i and F_{a_i} modified accordingly until the process ends or all the predictors have been added to the model.

Remarks

1. In Step 2, the general form of the partial F-test statistic should be

$$F = \frac{SSR\left(x_i \middle| x_{a_1}, \cdots, x_{a_k}\right)}{MSE\left(x_i, x_{a_1}, \cdots, x_{a_k}\right)}$$

where $x_{a_1}, \cdots, x_{a_k}$ are already included in the model from the previous steps, and x_i is not in the model. (i.e. $i \notin \{a_1, \cdots, a_k\}$.)

Remarks (Continued)

2. In Step 3, the partial F-test statistic should be

$$F = \frac{SSR\left(x_{a_i} \middle| x_{a_1}, \cdots, x_{a_{i-1}}, x_{a_i+1}, \cdots, x_{a_k}\right)}{MSE\left(x_{a_i}, x_{a_1}, \cdots, x_{a_{i-1}}, x_{a_i+1}, \cdots, x_{a_k}\right)}$$

where $x_{a_1}, \cdots, x_{a_k}$ and x_{a_i} are already included in the model from Step 2.

6.4 Example 1

- Use the stepwise regression method to find the 'best' model for the Hald data.

Solution

- Correlation coefficient matrix

$$
\begin{array}{c} \\ y \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{array}
\begin{array}{c}
\begin{array}{ccccc} y & x_1 & x_2 & x_3 & x_4 \end{array} \\
\begin{pmatrix}
1 & 0.73 & 0.82 & -0.53 & -0.82 \\
0.73 & 1 & 0.23 & -0.82 & -0.25 \\
0.82 & 0.23 & 1 & -0.14 & -0.97 \\
-0.53 & -0.82 & -0.14 & 1 & 0.03 \\
-0.82 & -0.25 & -0.97 & 0.03 & 1
\end{pmatrix}
\end{array}
$$

Example 1 (Continued)

Step 1:

- Since r_{y,x_4} $(= -0.82)$ is the largest among all r_{y,x_i}, therefore x_4 is the best candidate to be the first predictor to be included in the model.

- Also $F_{obs} = MSR(x_4)/MSE(x_4) = 22.8 > F_{0.10}(1, 11) = 3.23$. Hence x_4 should be included in the model.

Example 1 (Continued)

Step 2:

- To determine which of the remaining predictors x_1, x_2 and x_3 (if any) to be included in the model.

$$F_1 = \frac{SSR(x_1|x_4)}{MSE(x_1, x_4)} = \frac{809.10}{7.4762} = 108.22$$

$$F_2 = \frac{SSR(x_2|x_4)}{MSE(x_2, x_4)} = \frac{14.9868}{86.8880} = 0.172$$

$$F_3 = \frac{SSR(x_3|x_4)}{MSE(x_3, x_4)} = \frac{708.1289}{17.5738} = 40.29$$

Example 1 (Continued)

Step 2 (Continued)

- Since $F_1 = \max(F_1, F_2, F_3) > F_{0.10}(1, 10) = 3.29$, therefore we include x_1 in the model that has x_4.

- Note: Since

$$F_i = \frac{SSR(x_i, x_4) - SSR(x_4)}{\big(SST - SSR(x_i, x_4)\big)/df}$$

therefore F_i is an increasing function of $SSR(x_i, x_4)$

- Hence finding $\max_i F_i$ is equivalent to finding $\max_i SSR(x_i, x_4)$.

Example 1 (Continued)

Step 3

- To determine if x_1 or x_4 can be deleted from the model.

$$F_4 = \frac{SSR(x_4|x_1)}{MSE(x_1, x_4)} = \frac{1190.9246}{7.4762} = 159.30$$

- Since $F_4 = 159.30 > F_{0.05}(1,10) = 4.96$, therefore x_4 contributes significantly to the model even though x_1 is already there.

- Hence x_4 is retained.

Example 1 (Continued)

Step 4

- (Similar to Step 2) To determine which of the remaining predictor variables x_2 and x_3 (if any) to be included in the model

$$F_2' = \frac{SSR(x_2|x_1, x_4)}{MSE(x_1, x_2, x_4)} = \frac{26.7803}{5.3303} = 5.026$$

$$F_3' = \frac{SSR(x_3|x_1, x_4)}{MSE(x_1, x_3, x_4)} = \frac{26.7803}{5.6485} = 4.236$$

Example 1 (Continued)

Step 4 (Continued)

- Since $F_2' = \max(F_2', F_3') > F_{0.10}(1, 9) = 3.36$, we include x_2 in the model.

Note:

- Finding $\max(F_2', F_3')$ is equivalent to finding $\max_i SSR(x_i, x_1, x_4)$, $i = 2, 3$.

Example 1 (Continued)

Step 5

- (Similar to Step 3) To determine if x_1, x_2 or x_4 can be deleted from the model.

$$F_1'' = \frac{SSR(x_1|x_2, x_4)}{MSE(x_1, x_2, x_4)} = \frac{820.9074}{5.3303} = 154.01$$

$$F_4'' = \frac{SSR(x_4|x_1, x_2)}{MSE(x_1, x_2, x_4)} = \frac{9.9318}{5.3303} = 1.863$$

- Since $F_4'' = \min(F_1'', F_2'', F_4'') < F_{0.05}(1, 9) = 5.12$, therefore x_4 is deleted from the model.

Example 1 (Continued)

Step 5 (Continued)

Note: Since

$$F_i'' = \frac{SSR(x_1, x_2, x_4) - SSR(\{x_1, x_2, x_4\} \backslash \{x_i\})}{MSE(x_1, x_2, x_4)}$$

therefore F_i'' is a decreasing function of $SSR(\{x_1, x_2, x_4\} \backslash \{x_i\})$

- Hence finding $\min_i F_i''$ is equivalent to finding $\max_i SSR(\{x_1, x_2, x_4\} \backslash \{x_i\})$.

Example 1 (Continued)

Step 6

- (Similar to Step 2) To determine if x_3 or x_4 to be included in the model with x_1 and x_2 already in it.

$$F'''_3 = \frac{SSR(x_3|x_1, x_2)}{MSE(x_1, x_2, x_3)} = \frac{9.7939}{5.3456} = 1.832$$

$$F'''_4 = \frac{SSR(x_4|x_1, x_2)}{MSE(x_1, x_2, x_4)} = \frac{9.9318}{5.3303} = 1.863$$

- Since $F'''_4 = \max(F'''_3, F'''_4) < F_{0.10}(1, 9) = 3.36$, therefore we do not include either x_3 or x_4 in the model with x_1 and x_2 in it and the process stops.

Example 1 (Continued)

Conclusion

- Hence the 'best' regression equation for the Hald data using the stepwise method is .

$$\hat{y} = f(x_1, x_2)$$

- From the printout, the fitted equation is

$$\hat{y} = 52.5773 + 1.4683x_1 + 0.6623x_2$$

9.5 Forward and Backward Selections

9.5.1 Forward Selection

- Do Steps 1 and 2 in the stepwise regression procedure, but once a predictor is included, it will stay. (i.e. we can skip Step 3.)

- Repeat Step 2 until the process stops or all the predictor variables are included in the model.

9.5.2 Backward Selection

Step 1: A regression equation containing all predictor variables is computed.

Step 2: The partial F-test value is calculated for every predictor variable treated as though it was the last predictor variable to enter the regression equation.

9.5.2 Backward Selection

Step 3: The lowest partial F-test value, say, F_L, is compared with a critical value F_0 at a pre-selected level.

- If $F_L < F_0$, remove the variable x_L, which gives rise to F_L, from consideration and repeat Step 2.
- If $F_L > F_0$, process stops and we adopt the regression equation as the 'best' equation.

9.5.3 Example on Forward Selection

Refer to the Hald data on p9-8

- Use the forward selection to obtain the 'best' regression model.
- Steps 1 and 2 are the same as Steps 1 and 2 in the stepwise regression on p.9-33 to p.9-35.
  - x_4 is included in the model in Step 1
  - x_1 is included in the model with x_4 in Step 2
- Step 3 is similar to Step 4 in the stepwise regression on p.9-37 to p.9-38.
- Step 4: To determine if x_3 can be included into the model with x_1, x_2 and x_4 in it.

Example on Forward Selection (Continued)

$$F_3 = \frac{SSR(x_3|x_1, x_2, x_4)}{MSE(x_1, x_2, x_3, x_4)} = \frac{0.1091}{5.9830} = 0.018$$

- Since $F_3 < F_{0.10}(1, 8) = 3.46$, therefore x_3 is not included in the model.

- Hence the 'best' equation by using forward selection method is

$$\hat{y} = 71.6483 + 1.4519x_1 + 0.4161x_2 - 0.2365x_4$$

9.5.4 Example on Backward Selection

Refer to the Hald data on p9-8

- Use the backward elimination to obtain the 'best' model.

- **Step 1**: Fit the data using the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon.$$

$MSE(x_1, x_2, x_3, x_4)$ = 5.98295 with 8 d.f.

- **Step 2**:

Compute

$$F_i = \frac{SSR(x_i | \text{all except } x_i)}{MSE(x_1, x_2, x_3, x_4)}$$

Example on Backward Selection (Continued)

- Step 2 (Continued)

$$F_1 = \frac{SSR(x_1 | x_2, x_3, x_4)}{MSE(x_1, x_2, x_3, x_4)} = \frac{25.9509}{5.98295} = 4.337$$

$$F_2 = \frac{2.9724}{5.98295} = 0.4968$$

$$F_3 = \frac{0.1091}{5.98295} = 0.0182$$

$$F_4 = \frac{0.2470}{5.98295} = 0.0413$$

- Since $F_3 = \min(F_1, F_2, F_3, F_4) < F_{0.05}(1, 8) = 5.32$, therefore x_3 is removed from the model.

Example on Backward Selection (Continued)

- Step 3
- Compute $F_i' = \dfrac{SSR(x_i|\text{all except } x_i \text{ and } x_3)}{MSE(x_1,x_2,x_4)}$

$$F_1' = \frac{SSR(x_1|x_2 \text{ and } x_4)}{MSE(x_1,x_2,x_4)} = \frac{820.9074}{5.3303} = 154.01$$

$$F_2' = \frac{26.7894}{5.3303} = 5.03$$

$$F_4' = \frac{9.4318}{5.3303} = 1.86$$

- Since $F_4' = \min(F_1', F_2', F_4') < F_{0.05}(1, 9) = 5.12$, therefore x_4 is removed from the model.

Example on Backward Selection (Continued)

- Step 4
- Compute $F_i'' = \dfrac{SSR(x_i|x_1 \text{ and } x_2)}{MSE(x_1,x_2)}$

$$F_1'' = \frac{848.4319}{5.7904} = 146.52, \quad F_2'' = \frac{1207.7823}{5.7904} = 208.58$$

- Since $F_1'' = \min(F_1'', F_2'') > F_{0.05}(1, 10) = 4.96$, therefore we do not delete x_1 or x_2 from the model and the process stops.
- Hence the 'best' regression equation by the backward selection is

$$\hat{y} = 52.5773 + 1.4683x_1 + 0.6623x_2$$

9.6 SAS and R programs

9.6.1 SAS Program

```
proc stepwise data=ch9ex1;
  model y = x1 x2 x3 x4/stepwise forward
   backward sle = 0.10 sls=0.05;
run;
```

The output can be found in the file
"Ch9SASStepwiseOutput.doc" in IVLE.

9.6.2 R program

```
# Null model: y = beta_0 + epsilon
Nullmodel <- lm(y~1, data=ch9ex1)
# Full model: y = beta_0 + beta_1 x_1 + beta_2 x_2
# + beta_3 x_3 + beta_4 x_4 + epsilon
Fullmodel <- lm(y~., data=ch9ex1)

# Stepwise regression based on Akaike Information
# Criterion
step(fullmodel, data=ch9ex1, direction="both", k=2)
```

The output can be found in the file
"Ch9RStepwiseOutput.doc" in IVLE.

Akaike Information Criterion (AIC)

- AIC is defined as

$$-2\ln(L) + 2p$$

where L is the likelihood function of the model and p is the number of parameters in the model.

- A small AIC value implies a better model
- If errors follow a normal distribution, then

$$AIC = 2p + n\log\left(\frac{SSE}{n}\right) + \text{constant}$$

where SSE is the Error Sum of Squares and n is the sample size

- In R, the constant is ignored.

R program (Continued)

```
#Stepwise regression step by step
cor(ch9ex1)
Model4 <- lm(y~x4)
summary(model4)

Model14 <- lm(y~x4+x1)
model24 <- lm(y~x4+x2)
model34 <- lm(y~x4+x3)
anova(model4,model14)
anova(model4,model24)
anova(model4,model34)

#Check if x4 can be removed from the model with x1 and x4
model1 <- lm(y~x1)
anova(model1,model14)
```

R program (Continued)

```
# Check if x2 or x3 can be included in the model with
# x1 and x4.
model124 <- lm(y~x4+x1+x2); model134 <- lm(y~x4+x1+x3)
anova(model14,model124)
anova(model14,model134)

# Check if x1, x2 or x4 can be removed from the model
# with x1, x2 and x4
anova(model24,model124); anova(model14,model124)
model12 <- lm(y~x1+x2)
anova(model12,model124)
#Check if x3 or x4 can be included in the model
model123 <- lm(y~x1+x2+x3)
anova(model12,model123)
anova(model12,model124)
```