# Applied Regression

## Multiple Linear Regression Model (MLR Model)

## Model Building - Part-2

Module 6 Lecture - 6-2

In the last lecture, we have evaluated all possible models and picked one out of that. But when number of variables is large then evaluating all possible regressions can be burdensome computationally, various methods have been developed for evaluating only a small number of subset regression models by either adding or deleting regressors one at a time. These methods are generally referred to as selection-type procedures. They can be classified into three broad categories: (1) forward selection, (2) backward elimination, and (3) stepwise regression.

**Forward Selection:**

This procedure begins with the assumption that there is no regressors in the model other than the intercept. At each step one variable is selected and added to the model if the criterion is met. To perform the forward selection we need to choose a significance level which will be used by the process to test the significance of the variable. Lets say that value is $\alpha$.

Step-1: The first $X_i$ selected to enter the model is the one that has the largest simple correlation with the response variable y. Suppose that this regressor is $x_1$. This is also the regressor that will produce the largest value of $r^2$ and hence the F statistic for testing significance of regression (which is same as testing the variable) . But this regressor is entered if the corresponding p-value is smaller than the $\alpha$.

Step-2: The second $X_i$ chosen for entry is the one that now has the largest correlation with y after adjusting for the effect of the first regressor ($X_1$) on y which has already entered the model. This type of correlations is called partial correlations. "Partial Correlations" are the simple correlations between the residuals from two separate regressions.

(1) Model: $Y = \beta_0 + \beta_1.X_1 + e \ \Rightarrow \ \hat{e}_i = y_i - \hat{\beta_0} + \hat{\beta_1}.X_{1i}$

(2) Model: $X_2 = \theta_0 + \theta_1.X_1 + e \ \Rightarrow \ \hat{e}_i = x_{2i} - \hat{\theta_0} + \hat{\theta_1}.X_{1i}$

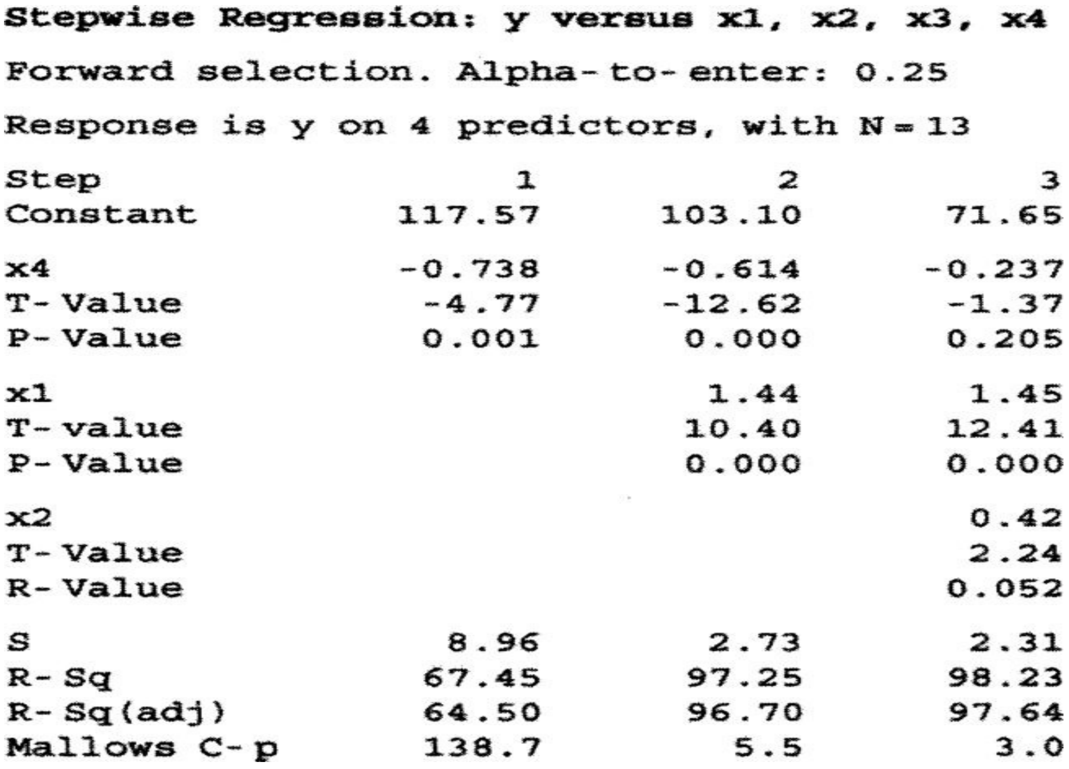
There is another way to look at the same situation. At this step that $X_j$ is chosen which has the highest t-stat value in the parameter estimate table from the model with two regressors ($X_1 \& X_j$). But also the p-value for $X_j$ has to be less than the $\alpha$ to enter the model. Else the selection procedure will stop.

Step-i - In general, at each step it picks one from the remaining candidates (say $X_k$) and fit a model and take a look at the t-stat and pick the one with highest t-stat but must have p-value less than the chosen $\alpha$. If there is none with p-value less than the chosen $\alpha$ then the selection procedure stops and declare the last step model as the final one.

Remark: Note that once a variable is picked, it never leaves the model even if its p-value may have increased beyond $\alpha$ after addition of the next variable. So we may end up with a model where few of the variables are insignificant at level $\alpha$.

Example: Lets look at the previous example (Data B21) using Forward Selection.

### Forward Selection - Example



```
Stepwise Regression: y versus x1, x2, x3, x4

Forward selection. Alpha-to-enter: 0.25

Response is y on 4 predictors, with N = 13

Step                   1          2          3
Constant          117.57     103.10      71.65

x4                -0.738     -0.614     -0.237
T-Value            -4.77     -12.62      -1.37
P-Value            0.001      0.000      0.205

x1                             1.44       1.45
T-value                       10.40      12.41
P-Value                       0.000      0.000

x2                                        0.42
T-Value                                   2.24
R-Value                                  0.052

S                   8.96       2.73       2.31
R-Sq               67.45      97.25      98.23
R-Sq(adj)          64.50      96.70      97.64
Mallows C-p        138.7        5.5        3.0
```

In this example,

Step-1 - $X_4$ is selected as it has the highest correlation with Y which also means that this $X_4$ has the g=highest correlation among the 4 SLR models.

Step - 2 AT this step 3 models are compared namely, $(X_4, X_1)$, $(X_4, X_2)$, and $(X_4, X_3)$, and for each of these models we look at line corresponding to $X_1, X_2, \& X_3$ in the parameter estimates tables. It seems $X_1$ has the highest t-stat value.

Step - 3 Now $(X_4 \& X_1)$ are both in the model so two new models are compared namely models with $(X_4, X_1, X_2)$ and $(X_4, X_1, X_3)$. Again By looking at the t-tables $X_2$ is added to the model.

Step - 4 Finally there is only one remaining to be added and so the model with all four is looked at but $X_3$ is not significant and hence the process stopped and final model is

$$Y = \beta_0 + \beta_1 . X_4 + \beta_2 . X_1 + \beta_3 . X_2 + \epsilon$$

## Backward Elimination

Forward selection begins with no regressors in the model and attempts to insert variables until a suitable model is obtained. Backward elimination attempts to find a good model by working in the opposite direction.

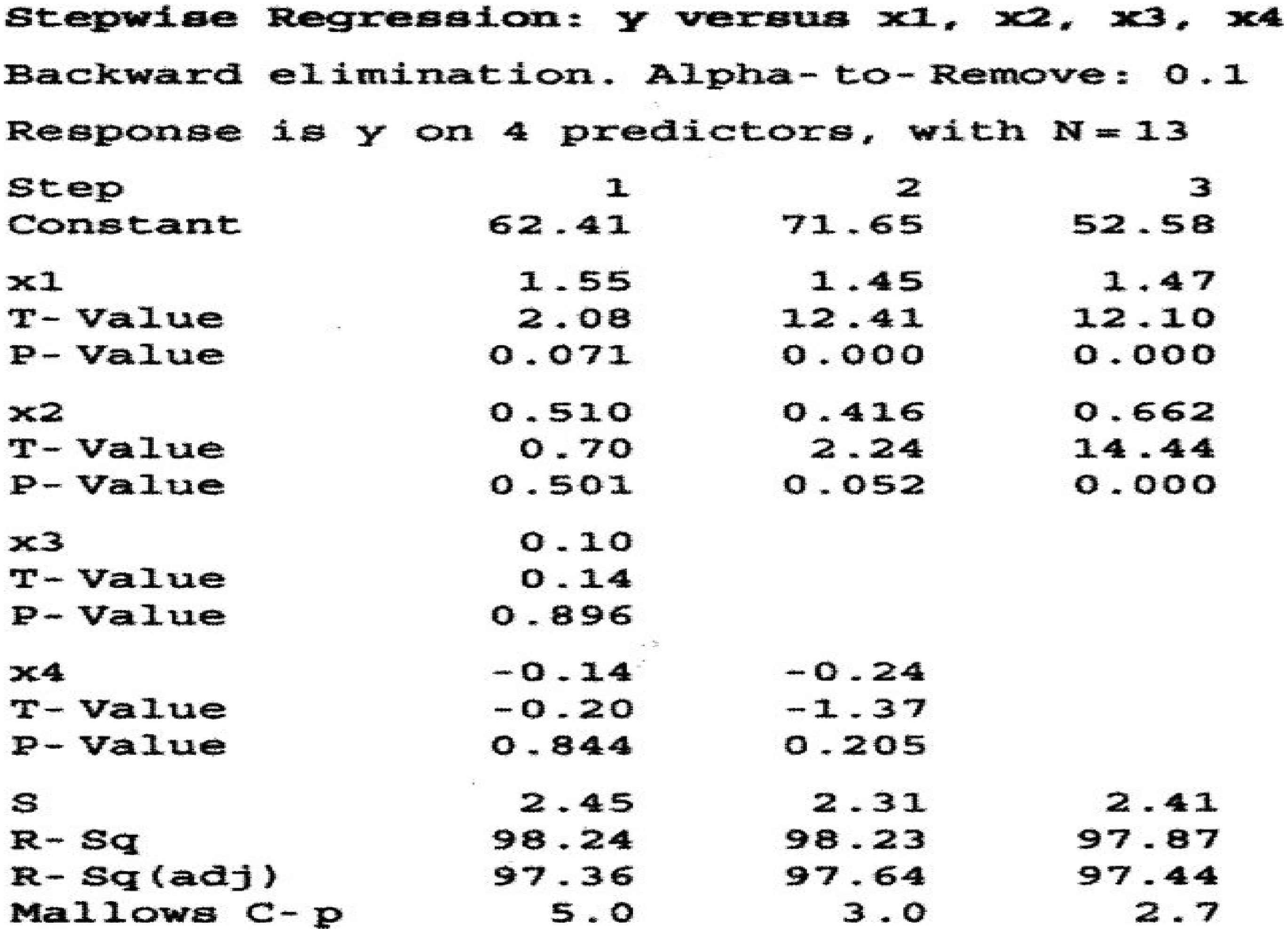
Step-1 We begin with a model that includes all K candidate regressors i.e the Full Model. We look at the parameter estimate table and select the variable with highest p-value (alternatively lowest value of $|t - stat|$) say $X_j$. If p-value corresponding to $X_j$ is smaller than the $\alpha$ then the process stops as it assures that all the variables are significant at level $\alpha$. But if highest p-value is greater than $\alpha$ then that variables is dropped and we rerun the model after dropping.

Step-i In general, it runs the model with current set of variables and variable with highest p-value is dropped if it is bigger than $\alpha$. Else process stops as no variable to drop.

Remark: Backward elimination is often a very good variable selection procedure. It is particularly favored by analysts who like to see the effect of including all the candidate regressors, just so that nothing obvious will be missed. Also it guaranties that all the variables are significant once process stops.

Example: Hald Cement Data - B21. Here Backward selection is applied with $\alpha = 0.10$

## Backward Elimination - Example



```
Stepwise Regression: y versus x1, x2, x3, x4

Backward elimination. Alpha-to-Remove: 0.1

Response is y on 4 predictors, with N=13

Step                  1        2        3
Constant          62.41    71.65    52.58

x1                 1.55     1.45     1.47
T-Value            2.08    12.41    12.10
P-Value           0.071    0.000    0.000

x2                0.510    0.416    0.662
T-Value            0.70     2.24    14.44
P-Value           0.501    0.052    0.000

x3                 0.10
T-Value            0.14
P-Value           0.896

x4                -0.14    -0.24
T-Value           -0.20    -1.37
P-Value           0.844    0.205

S                  2.45     2.31     2.41
R-Sq              98.24    98.23    97.87
R-Sq(adj)         97.36    97.64    97.44
Mallows C-p         5.0      3.0      2.7
```

Step-1 When full model has been run, it shows that highest p-value = 0.896 (the smallest t value is 0.14), and it is associated with $x_3$ and it is also bigger than $\alpha = 0.10$. Thus, $x_3$ is removed from the model.

Step-2 We see the results of fitting the three-variable model involving $(x_1, x_2, x_4)$. The highest p-value which is associated with $x_4$ is greater than 0.10. Hence it is dropped from the model.

Step-3, We see the results of fitting the two-variable model involving $(x_1, x_2)$. All the p-values are smaller than 0.10, hence no further regressors can be removed from the model. Therefore, backward elimination terminates, yielding the final model as

$$Y = \beta_0 + \beta_1.X_1 + \beta_1.X_2 + \epsilon$$

## Stepwise Regression:

One of the most popular process among these is the stepwise regression algorithm. It is a combination of Forward and Backward. It starts like forward selection method and as soon as it adds a variable it takes a look at the model like Backward does and tries to drop any of the variables in the whole set. In general, a regressor added at an earlier step may now be redundant because of the relationships between it and regressors now in the equation. As a result one step of Forward is followed by a step like Backward. When there is no variable to drop and no variable to add it stops.

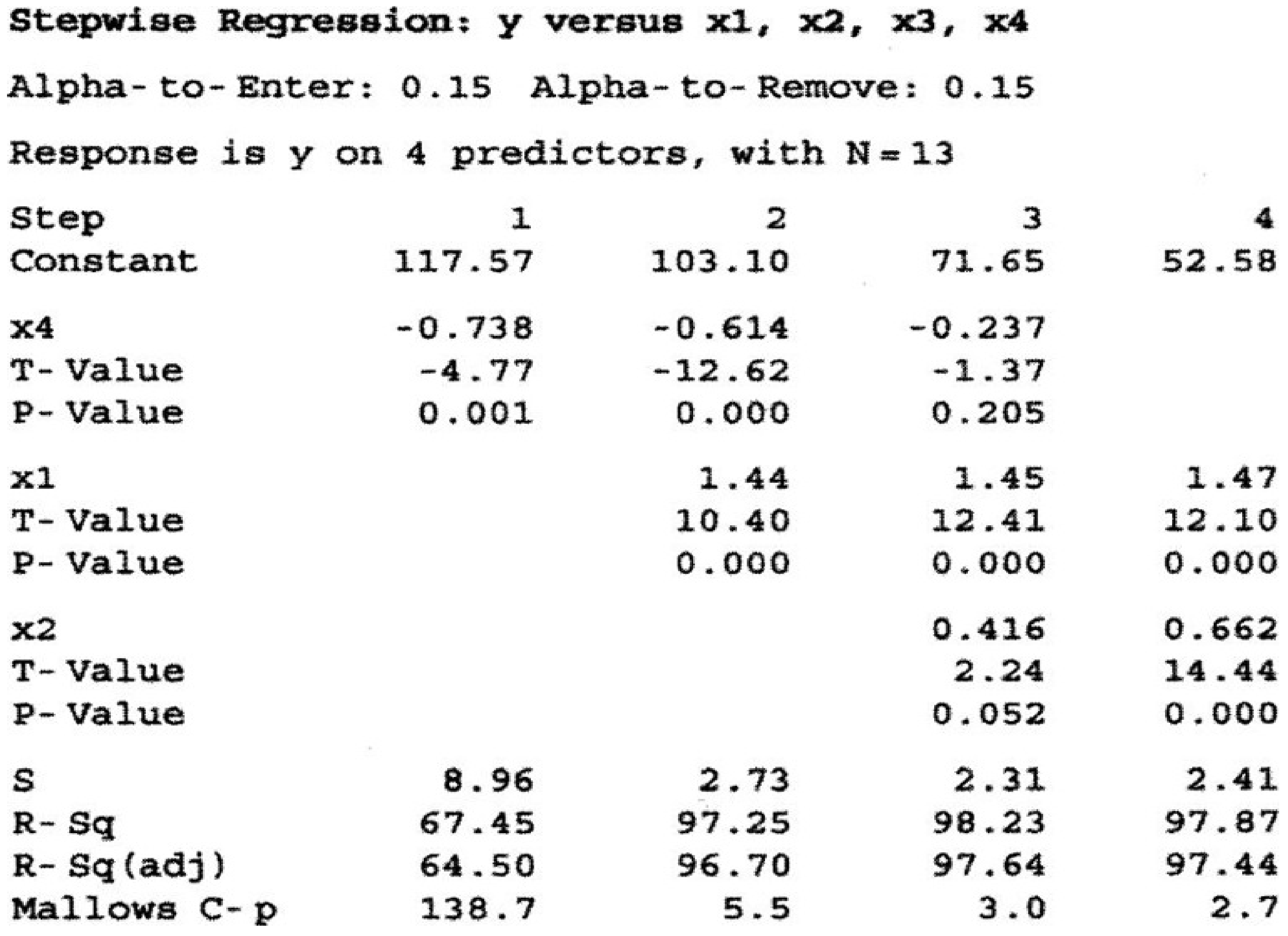
Stepwise regression requires two cutoff values, one for entering variables (like Forward) and one for dropping them (like Backward).

Previous table presents the results of using stepwise regression algorithm on the same data. Significance level ($\alpha$) for entry and stay is specified 0.15.

Step-1: Like Forward selection, the procedure begins with no variables in the model and adds $x_4$ as it has the highest correlation.

Step-2: Because Backward elimination will not drop anything, $x_1$ is added to the model as before.

## Stepwise Selection - Example



```
Stepwise Regression: y versus x1, x2, x3, x4

Alpha-to-Enter: 0.15  Alpha-to-Remove: 0.15

Response is y on 4 predictors, with N=13

Step                 1         2         3        4
Constant        117.57    103.10     71.65    52.58

x4              -0.738    -0.614    -0.237
T-Value          -4.77    -12.62     -1.37
P-Value          0.001     0.000     0.205

x1                          1.44      1.45     1.47
T-Value                    10.40     12.41    12.10
P-Value                    0.000     0.000    0.000

x2                                   0.416    0.662
T-Value                               2.24    14.44
P-Value                              0.052    0.000

S                 8.96      2.73      2.31     2.41
R-Sq             67.45     97.25     98.23    97.87
R-Sq(adj)        64.50     96.70     97.64    97.44
Mallows C-p      138.7       5.5       3.0      2.7
```

Step-3: Checking with Backward elimination shows that both the variables are significant and hence nothing is dropped.

Step-4 Again application of forward step adds $X_2$ to the model as its p-value is smaller than 0.15.

Step-5 Now looking at the parameter estimates table and using Backward elimination rule shows p-value for $X_4 = 0.205 > 0.15 = \alpha$. Hence $X_4$ is dropped.

At this point the only remaining candidate regressor is $x_3$, which cannot be added because its p-value is not small enough. Therefore, stepwise regression terminates with the model This is the same equation identified by the all-possible-regressions and backward elimination procedures.

**Example 4.1 (Revisit)** Consider the National Football League data.

a. Use the forward selection algorithm to select a subset regression model.
Solution:

**Forward Selection - Example 4.1** ($\alpha = 0.10$)

| Step | Variable | Estimate | StdErr | FValue | ProbF |
|---|---|---|---|---|---|
| 1 | Intercept | 21.78825 | 2.69623 | 65.3 | $<$ .0001 |
| 1 | x8 | -0.00703 | 0.00126 | 31.1 | $<$.0001 |
| 2 | Intercept | 14.71267 | 2.61753 | 31.59 | $<$.0001 |
| 2 | x2 | 0.00311 | 0.00070745 | 19.34 | 0.0002 |
| 2 | x8 | -0.00681 | 0.00096584 | 49.69 | $<$.0001 |
| 3 | Intercept | -1.80837 | 7.90086 | 0.05 | 0.8209 |
| 3 | x2 | 0.0036 | 0.000695 | 26.8 | $<$.0001 |
| 3 | x7 | 0.19396 | 0.08823 | 4.83 | 0.0378 |
| 3 | x8 | -0.00482 | 0.00128 | 14.22 | 0.0009 |

As we see in the above output that $x_8$ has entered first and then $X_2$ and then $x_7$. No other variables had small enough p-value (less than 0.10) to enter the model.

Summary of the step along with the partial r-sqaure is given below.

| Step | Entry | k | Partial RSq | Rsquare | Cp | FValue | ProbF |
|---|---|---|---|---|---|---|---|
| 1 | x8 | 1 | 0.5447 | 0.5447 | 20.4444 | 31.1 | $< .0001$ |
| 2 | x2 | 2 | 0.1986 | 0.7433 | 3.059 | 19.34 | 0.0002 |
| 3 | x7 | 3 | 0.043 | 0.7863 | 0.8591 | 4.83 | 0.0378 |

b. Use the backward elimination algorithm to select a subset regression model.

Solution: The Backward elimination process went through the following steps. The significance level used was also 0.10.

**Backward Elimination - Example 4.1 ($\alpha = 0.10$)**

| Step | Var Removed | Number In | Partial RSquare | Model Rsquare | Cp | FValue |
|---|---|---|---|---|---|---|
| 1 | x5 | 8 | 0 | 0.8156 | 8 | 0 |
| 2 | x1 | 7 | 0.0017 | 0.8139 | 6.165 | 0.17 |
| 3 | x6 | 6 | 0.0021 | 0.8118 | 4.3711 | 0.23 |
| 4 | x3 | 5 | 0.0053 | 0.8065 | 2.8862 | 0.59 |
| 5 | x4 | 4 | 0.0053 | 0.8012 | 1.4065 | 0.61 |
| 6 | x9 | 3 | 0.0149 | 0.7863 | 0.8591 | 1.72 |

It is easy to follow the reasoning behind backward elimination procedure as it runs the current set of regressors and look at the parameter estimates table. If you look at its output (in the last slide), it is very clear that why $x_5$ is dropped first and then followed by $x_1$ and on. The only thing look at the p-values of the variables and drop the variable with the highest p-value if it is above 0.10.

c. Use stepwise regression to select a subset regression model.

Solution: Stepwise only added exactly like forward and did not drop (like Backward) at any stage. So output is same as Forward.

d. Comment on the final model chosen by these three procedures.

In this problem, all three process ended up in the same model.

$$Y = \beta_0 + \beta_8.X_8 + \beta_2.X_2 + \beta_7.X_7 + e$$

Remark: But it is not necessary that all three process will provide the same answer. Then it is necessary to take a look at the model using other criterion like AIC, BIC, PRESS etc.

Example 4.1. Consider the National Football League data.

Restricting your attention to regressors $x_1$ (rushing yards), $x_2$ (passing yards), $x_4$ (field goal percentage), $x_7$ (percent rushing), $x_8$ (opponents rushing yards), and $x_9$ (opponents passing yards), apply the all-possible-regressions procedure.

Evaluate $R_p^2$, $C_p$ , and MSE for each model. Which subset of regressors do you recommend?

Solution: The possible models were run and sorted according to the $AdjR^2$, from highest to Lowest and then top 15 are presented in next slide.

It seems that the there are two models which are the two finalists. MDL-1 and MDL-4. These two are close in almost every respect. MDL-1 mainly beats MDL-4 in $Adj - R^2$ (and hence MSE) and marginally in AIC. The $C_p$ value is lower for MDL-4 but MDL-1 is closer to p=4. While MDL-4 has lower BIC and marginally lower in PRESS. But one thing is important that $x_9$ was not significant in the parameter estimates table and MDL-4 has smaller number of variables which is always preferred. Also if correlations are calculated then it shows that $x_9$ has a significant positive correlation with $x_8$ ($r_{89} = 0.4174$, p-value = 0.0271). Hence by looking at all things it looks like MDL-4 should be recommended.

**All Possible Models - Top 15 by $Adj - R^2$ - Problem 10.1**

| MDL | k | Adjrsq | RSquare | Cp | AIC | BIC | MSE | SSE | Var | Press |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 0.767 | 0.801 | 4.04 | 33.58 | 38.14 | 2.83 | 65.00 | x2 x7 x8 x9 | 87.66 |
| 2 | 5 | 0.763 | 0.807 | 5.41 | 34.77 | 40.27 | 2.87 | 63.14 | x1 x2 x7 x8 x9 | |
| 3 | 5 | 0.763 | 0.807 | 5.45 | 34.82 | 40.30 | 2.88 | 63.26 | x2 x4 x7 x8 x9 | |
| 4 | 3 | 0.760 | 0.786 | 3.69 | 33.60 | 36.99 | 2.91 | 69.87 | x2 x7 x8 | 87.46 |
| 5 | 4 | 0.759 | 0.795 | 4.73 | 34.45 | 38.66 | 2.91 | 67.04 | x1 x2 x8 x9 | 101.61 |
| 6 | 6 | 0.756 | 0.811 | 7.00 | 36.23 | 42.68 | 2.95 | 61.94 | x1 x2 x4 x7 x8 x9 | |
| 7 | 4 | 0.753 | 0.789 | 5.36 | 35.21 | 39.12 | 3.00 | 68.89 | x2 x4 x7 x8 | 94.50 |
| 8 | 4 | 0.752 | 0.789 | 5.38 | 35.24 | 39.14 | 3.00 | 68.97 | x1 x2 x7 x8 | |
| 9 | 3 | 0.750 | 0.778 | 4.66 | 34.73 | 37.78 | 3.03 | 72.75 | x1 x2 x8 | 100.00 |
| 10 | 5 | 0.749 | 0.796 | 6.66 | 36.36 | 41.08 | 3.04 | 66.84 | x1 x2 x4 x8 x9 | |
| 11 | 5 | 0.744 | 0.791 | 7.16 | 36.97 | 41.39 | 3.10 | 68.30 | x1 x2 x4 x7 x8 | |
| 12 | 4 | 0.739 | 0.778 | 6.65 | 36.72 | 40.04 | 3.16 | 72.70 | x1 x2 x4 x8 | |
| 13 | 2 | 0.723 | 0.743 | 6.46 | 36.74 | 38.64 | 3.36 | 83.94 | x2 x8 | |
| 14 | 3 | 0.718 | 0.750 | 7.77 | 38.06 | 40.12 | 3.41 | 81.91 | x2 x8 x9 | |
| 15 | 3 | 0.711 | 0.743 | 8.45 | 38.74 | 40.61 | 3.50 | 83.92 | x2 x4 x8 | |

# Backward Elimination - Ex 10.1

| Step | Variable | Estimate | StdErr | FValue | ProbF |
|---|---|---|---|---|---|
| 0 | Intercept | -7.2919 | 12.8128 | 0.32 | 0.5763 |
| 0 | x1 | 0.00081 | 0.00201 | 0.16 | 0.6903 |
| 0 | x2 | 0.00363 | 0.00084 | 18.64 | 0.0004 |
| 0 | x3 | 0.12217 | 0.25895 | 0.22 | 0.6427 |
| 0 | x4 | 0.03189 | 0.0416 | 0.59 | 0.4533 |
| 0 | x5 | 1.5E-05 | 0.04684 | 0 | 0.9997 |
| 0 | x6 | 0.00159 | 0.00325 | 0.24 | 0.6303 |
| 0 | x7 | 0.15435 | 0.15207 | 1.03 | 0.3235 |
| 0 | x8 | -0.0039 | 0.00205 | 3.6 | 0.0738 |
| 0 | x9 | -0.0018 | 0.00142 | 1.6 | 0.2225 |
| 1 | Intercept | -7.2937 | 11.3361 | 0.41 | 0.5277 |
| 1 | x1 | 0.00081 | 0.00195 | 0.17 | 0.6811 |
| 1 | x2 | 0.00363 | 0.00079 | 21.11 | 0.0002 |
| 1 | x3 | 0.12218 | 0.25137 | 0.24 | 0.6325 |
| 1 | x4 | 0.03189 | 0.03965 | 0.65 | 0.4311 |
| 1 | x6 | 0.00159 | 0.00313 | 0.26 | 0.6176 |
| 1 | x7 | 0.15437 | 0.13999 | 1.22 | 0.2839 |
| 1 | x8 | -0.0039 | 0.002 | 3.81 | 0.0659 |
| 1 | x9 | -0.0018 | 0.00138 | 1.69 | 0.2096 |
| 2 | Intercept | -9.1299 | 10.2296 | 0.8 | 0.3827 |
| 2 | x2 | 0.00363 | 0.00077 | 22.03 | 0.0001 |
| 2 | x3 | 0.16705 | 0.22247 | 0.56 | 0.4615 |
| 2 | x4 | 0.03699 | 0.03694 | 1 | 0.3286 |
| 2 | x6 | 0.00145 | 0.00305 | 0.23 | 0.639 |
| 2 | x7 | 0.18912 | 0.11019 | 2.95 | 0.1016 |
| 2 | x8 | -0.0042 | 0.0018 | 5.46 | 0.03 |
| 2 | x9 | -0.0017 | 0.00132 | 1.59 | 0.2212 |
| 3 | Intercept | -7.6949 | 9.59419 | 0.64 | 0.4315 |
| 3 | x2 | 0.00358 | 0.00075 | 22.68 | 0.0001 |
| 3 | x3 | 0.16754 | 0.21833 | 0.59 | 0.4514 |
| 3 | x4 | 0.035 | 0.03602 | 0.94 | 0.3423 |
| 3 | x7 | 0.19305 | 0.10784 | 3.2 | 0.0879 |
| 3 | x8 | -0.0044 | 0.00174 | 6.25 | 0.0208 |
| 3 | x9 | -0.0017 | 0.00129 | 1.65 | 0.2131 |
| 4 | Intercept | -4.6269 | 8.63964 | 0.29 | 0.5976 |
| 4 | x2 | 0.00371 | 0.00072 | 26.29 | <.0001 |
| 4 | x4 | 0.02639 | 0.03391 | 0.61 | 0.4446 |
| 4 | x7 | 0.23469 | 0.09232 | 6.46 | 0.0186 |
| 4 | x8 | -0.0037 | 0.00148 | 6.16 | 0.0212 |
| 4 | x9 | -0.0018 | 0.00127 | 1.96 | 0.1757 |
| 5 | Intercept | -1.8217 | 7.78471 | 0.05 | 0.817 |
| 5 | x2 | 0.00382 | 0.00071 | 29.33 | <.0001 |
| 5 | x7 | 0.21689 | 0.08868 | 5.98 | 0.0225 |
| 5 | x8 | -0.004 | 0.0014 | 8.24 | 0.0086 |
| 5 | x9 | -0.0016 | 0.00125 | 1.72 | 0.2024 |
| 6 | Intercept | -1.8084 | 7.90086 | 0.05 | 0.8209 |
| 6 | x2 | 0.0036 | 0.0007 | 26.8 | <.0001 |
| 6 | x7 | 0.19396 | 0.08823 | 4.83 | 0.0378 |
| 6 | x8 | -0.0048 | 0.00128 | 14.22 | 0.0009 |