Model Selection Procedures

Statistics 135

Autumn 2005

Model Selection Procedures

Consider a regression setting with K potential predictor variables and you wish to explore the set of different possible models, assuming no polynomial or interaction effects. Then there are 2^K possible models. Now if K is small examining all of them isn't a problems but if K becomes even moderate in size 2^K can get large quickly. For example $2^{50} \approx 1 \times 10^{15}$. Trying all of these models will be prohibitive.

Another problem is what criteria do we want to use pick good potential models. One criteria would be to pick the model with the highest R^2 or equivalently the model with the smallest SSE since

$$R^2 = \frac{SSM}{SSTot} = 1 - \frac{SSE}{SSTot}$$

This will always choose the model with all the predictors, which will probably include variables we don't want as overfit models often have poor predictive properties.

So we need to find a different criterion for picking reasonable models, i.e. ones that match the data well and tend to have few predictor variables (parsimony).

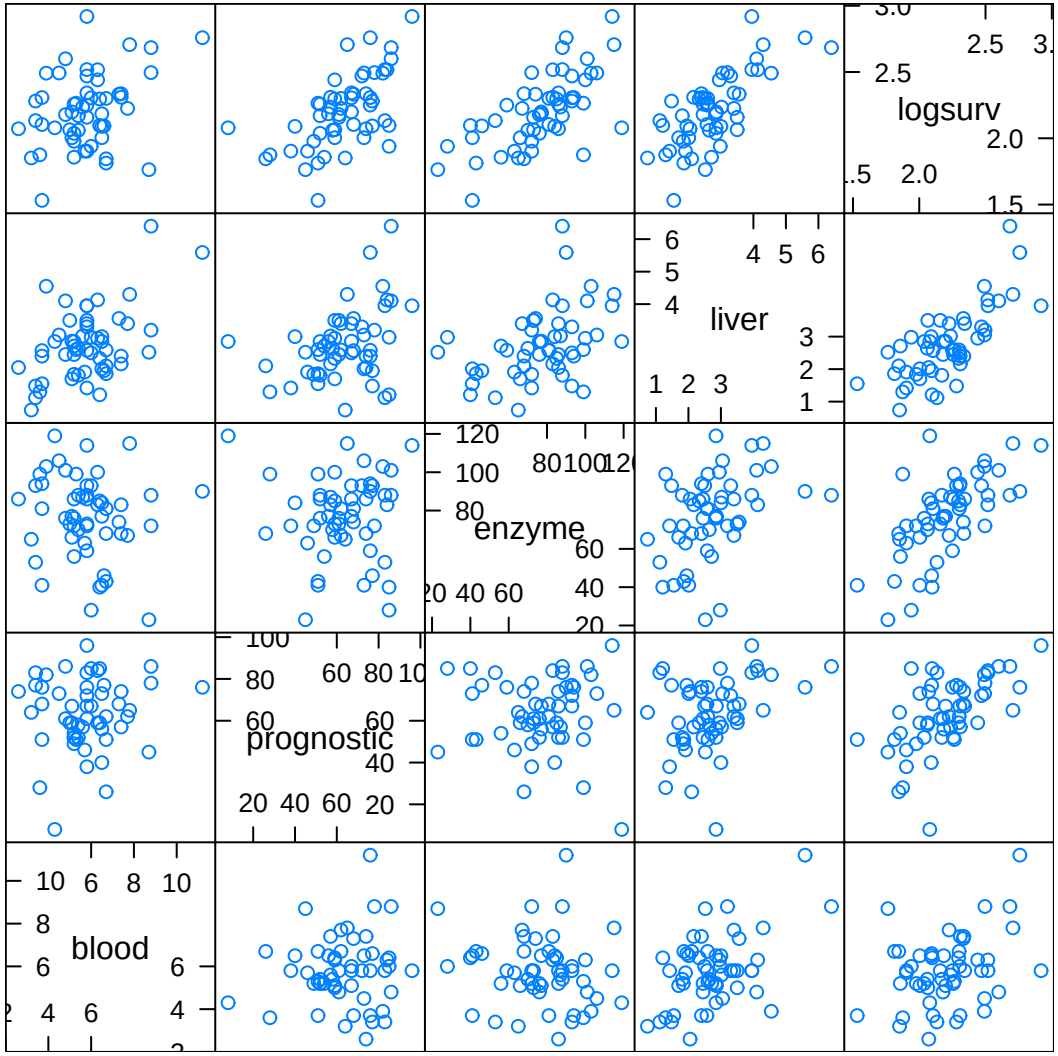
`PROC REG` has an option (`SELECTION`) that allows us to deal with these two problems.

Example: Surgery Survival Times

This dataset looks at survival times for 54 patients undergoing a particular type of liver operation.

The response variable is `logsurv`, the base 10 log of survival time of patients. The predictor variables are:

- `blood`: blood clotting score
- `prognostic`: prognostic index, which includes the age of the patient
- `enzyme`: enzyme function test score
- `liver`: liver function test score



Scatter Plot Matrix

Forward, Backward, and Stepwise Selection

One approach to the problem is to deal with building the model one variable at a time. There are three common related approaches for doing this, forward selection, backward deletion, and stepwise selection.

- Forward selection: This approach builds the model starting with no variables in the model and adds useful variables one by one. The general scheme is as follows

  1. For each predictor variable x_i not in the model, run a regression with this variable and those already in the model. Calculate the p-value (or F or $|t|$ statistic) to add this variable to the model.
  2. Choose the variable with the smallest p-value (or equivalently largest F or $|t|$). If this p-value $< p_{\text{add}}$ (or equivalently $F > F_{\text{add}}$ or $|t| > t_{\text{add}}$) add the variable to the model and go to step 1.
  3. Otherwise stop and declare variables already added as the model.

- Backward deletion: Instead of starting with no variables in the model, start with all predictor variable in the model and remove unhelpful variables from the model one by one. The algorithm for this scheme is

  1. Run the regression with all variables currently in the model and calculate the p-value (or F and $|t|$ statistics) for each.
  2. Choose the variable with the largest p-value (or equivalently smallest F or $|t|$). If this p-value $\geq p_{\text{drop}}$ (or equivalently $F > F_{\text{drop}}$ or $|t| > t_{\text{drop}}$) drop the variable from the model and go to step 1.
  3. Otherwise stop and declare the remaining variables as the model.

  Both of these approaches have problems. Forward selection can add variables early on that in the long run we don't want to include in the model. Similarly, backward deletion can remove variables that we should probably keep. An alternative that generally works better is

- Stepwise selection: This approach combines both forward selection and backward deletion. It allows variable added early on to be dropped out and variables that are dropped at one point to be added back in. The algorithm is as follows

1. For each predictor variable x_i not in the model, run a regression with this variable and those already in the model. Calculate the p-value (or F or $|t|$ statistic) to add this variable to the model.
2. Choose the variable with the smallest p-value (or equivalently largest F or $|t|$). If this p-value $< p_{\text{add}}$ (or equivalently $F > F_{\text{add}}$ or $|t| > t_{\text{add}}$) add the variable to the model.
3. Run the regression with all variables currently in the model and calculate the p-value (or F and $|t|$ statistics) for each.
4. Choose the variable with the largest p-value (or equivalently smallest F or $|t|$). If this p-value $\geq p_{\text{drop}}$ (or equivalently $F > F_{\text{drop}}$ or $|t| > t_{\text{drop}}$) drop the variable from the model.
5. If no variables are added in step 2 or dropped in step 4 stop and declare the variables currently selected as the model. Otherwise go to step 1.

For this algorithm to work it is necessary that $p_{add} \leq p_{drop}$ (or equivalently $F_{add} \geq F_{drop}$ and $t_{add} \geq t_{drop}$. If not, the algorithm will not terminate as it is possible to have a variable with $p_{drop} < p < p_{add}$ which will lead to the variable being added in and then immediately dropped.

For each of these algorithms values for p_{add} and p_{drop} need to be define (equivalently F_{add} and F_{drop} or t_{add} and t_{drop}). Common choices for these are

- $p_{add} = p_{drop} = 0.05$
- $F_{add} = F_{drop} = 4.0$
- $t_{add} = t_{drop} = 2.0$

Of course these can be changed to make it easier or harder to add or remove variables. The defaults for **SAS** are

- Forward selection: $p_{add} = 0.50$
- Backward deletion: $p_{drop} = 0.10$
- Stepwise selection: $p_{add} = p_{drop} = 0.15$

These procedures can be run from **SAS** using `PROC REG` using the `SELECTION` option. For example, the 3 schemes using the common cutoffs are done by

```
PROC REG DATA=surgery;

  Forward: MODEL logsurv = blood prognostic enzyme liver /
    SELECTION = FORWARD SLENTRY = 0.05;

  Backward: MODEL logsurv = blood prognostic enzyme liver /
    SELECTION = BACKWARD SLSTAY = 0.05;

  Stepwise: MODEL logsurv = blood prognostic enzyme liver /
    SELECTION = STEPWISE SLENTRY = 0.05 SLSTAY = 0.05;
```

As can be seen p_{add} is set by `SLENTRY` and p_{drop} is set by `SLSTAY`.

The output (edited) for these three possibilities are

- Forward selection:

Forward Selection: Step 1

Variable liver Entered: R-Square = 0.5274 and C(p) = 788.1481

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.09514	2.09514	58.02	<.0001
Error	52	1.87763	0.03611		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.69638	0.07174	20.18803	559.10	<.0001
liver	0.18575	0.02439	2.09514	58.02	<.0001

Forward Selection: Step 2

Variable enzyme Entered: R-Square = 0.6865 and C(p) = 507.8964

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.72744	1.36372	55.85	<.0001
Error	51	1.24533	0.02442		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.38878	0.08447	6.60079	270.32	<.0001
enzyme	0.00565	0.00111	0.63230	25.89	<.0001
liver	0.13901	0.02206	0.96994	39.72	<.0001

Forward Selection: Step 3

Variable prognostic Entered: R-Square = 0.8829 and C(p) = 161.6625

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.50756	1.16919	125.66	<.0001
Error	50	0.46521	0.00930		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.94226	0.07139	1.62089	174.21	<.0001
prognostic	0.00790	0.00086260	0.78012	83.85	<.0001
enzyme	0.00700	0.00070128	0.92694	99.63	<.0001
liver	0.08185	0.01498	0.27780	29.86	<.0001

Forward Selection: Step 4

Variable blood Entered: R-Square = 0.9724 and C(p) = 5.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3.86300	0.96575	431.10	<.0001
Error	49	0.10977	0.00224		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.48876	0.05023	0.21208	94.67	<.0001
blood	0.06852	0.00544	0.35544	158.66	<.0001
prognostic	0.00925	0.00043673	1.00583	448.99	<.0001
enzyme	0.00947	0.00039625	1.28075	571.71	<.0001
liver	0.00193	0.00971	0.00008809	0.04	0.8436

All variables have been entered into the model.

Summary of Forward Selection

Step	Variable Entered	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	liver	1	0.5274	0.5274	788.148	58.02	<.0001
2	enzyme	2	0.1592	0.6865	507.896	25.89	<.0001
3	prognostic	3	0.1964	0.8829	161.662	83.85	<.0001
4	blood	4	0.0895	0.9724	5.0000	158.66	<.0001

So in this example, all variables get added by forward selection.

- Backward elimination

```
Backward Elimination: Step 0
All Variables Entered: R-Square = 0.9724 and C(p) = 5.0000

                          Analysis of Variance
                                   Sum of           Mean
Source                     DF       Squares         Square    F Value    Pr > F

Model                       4       3.86300        0.96575     431.10    <.0001
Error                      49       0.10977        0.00224
Corrected Total            53       3.97277

              Parameter     Standard
Variable       Estimate        Error   Type II SS  F Value  Pr > F

Intercept       0.48876      0.05023      0.21208    94.67  <.0001
blood           0.06852      0.00544      0.35544   158.66  <.0001
prognostic      0.00925   0.00043673      1.00583   448.99  <.0001
enzyme          0.00947   0.00039625      1.28075   571.71  <.0001
liver           0.00193      0.00971   0.00008809     0.04  0.8436
```

Backward Elimination: Step 1

Variable liver Removed: R-Square = 0.9723 and C(p) = 3.0393

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.86291	1.28764	586.04	<.0001
Error	50	0.10986	0.00220		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.48362	0.04263	0.28279	128.71	<.0001
blood	0.06923	0.00408	0.63315	288.17	<.0001
prognostic	0.00929	0.00038250	1.29732	590.45	<.0001
enzyme	0.00952	0.00030641	2.12263	966.07	<.0001

```
All variables left in the model are significant at the 0.0500 level.
```

Summary of Backward Elimination

Step	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	liver	3	0.0000	0.9723	3.0393	0.04	0.8436

So for this example `blood`, `prognostic` and `enzyme` are kept and `liver` is dropped from the model.

- Stepwise selection:

Stepwise Selection: Step 1

Variable liver Entered: R-Square = 0.5274 and C(p) = 788.1481

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2.09514	2.09514	58.02	<.0001
Error	52	1.87763	0.03611		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.69638	0.07174	20.18803	559.10	<.0001
liver	0.18575	0.02439	2.09514	58.02	<.0001

Stepwise Selection: Step 2

Variable enzyme Entered: R-Square = 0.6865 and C(p) = 507.8964

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	2.72744	1.36372	55.85	<.0001
Error	51	1.24533	0.02442		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.38878	0.08447	6.60079	270.32	<.0001
enzyme	0.00565	0.00111	0.63230	25.89	<.0001
liver	0.13901	0.02206	0.96994	39.72	<.0001

Stepwise Selection: Step 3

Variable prognostic Entered: R-Square = 0.8829 and C(p) = 161.6625

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.50756	1.16919	125.66	<.0001
Error	50	0.46521	0.00930		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.94226	0.07139	1.62089	174.21	<.0001
prognostic	0.00790	0.00086260	0.78012	83.85	<.0001
enzyme	0.00700	0.00070128	0.92694	99.63	<.0001
liver	0.08185	0.01498	0.27780	29.86	<.0001

Stepwise Selection: Step 4

Variable blood Entered: R-Square = 0.9724 and C(p) = 5.0000

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	3.86300	0.96575	431.10	<.0001
Error	49	0.10977	0.00224		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.48876	0.05023	0.21208	94.67	<.0001
blood	0.06852	0.00544	0.35544	158.66	<.0001
prognostic	0.00925	0.00043673	1.00583	448.99	<.0001
enzyme	0.00947	0.00039625	1.28075	571.71	<.0001
liver	0.00193	0.00971	0.00008809	0.04	0.8436

Stepwise Selection: Step 5

Variable liver Removed: R-Square = 0.9723 and C(p) = 3.0393

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3.86291	1.28764	586.04	<.0001
Error	50	0.10986	0.00220		
Corrected Total	53	3.97277			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.48362	0.04263	0.28279	128.71	<.0001
blood	0.06923	0.00408	0.63315	288.17	<.0001
prognostic	0.00929	0.00038250	1.29732	590.45	<.0001
enzyme	0.00952	0.00030641	2.12263	966.07	<.0001

All variables left in the model are significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	liver		1	0.5274	0.5274	788.148	58.02	<.0001
2	enzyme		2	0.1592	0.6865	507.896	25.89	<.0001
3	prognostic		3	0.1964	0.8829	161.662	83.85	<.0001
4	blood		4	0.0895	0.9724	5.0000	158.66	<.0001
5		liver	3	0.0000	0.9723	3.0393	0.04	0.8436

So the result of stepwise is the same as backwards in this case.

Generally the three procedures can give different answers, particularly in datasets with many predictor variables. The relationship between the different results depends on the correlations between the different variables.

In this example, they can be gotten by

```
PROC CORR DATA=surgery;
  VAR logsurv blood prognostic enzyme liver;
```

The CORR Procedure

Simple Statistics

Variable	N	Mean	Std Dev	Sum
logsurv	54	2.20614	0.27378	119.13180
blood	54	5.78333	1.60303	312.30000
prognostic	54	63.24074	16.90253	3415
enzyme	54	77.11111	21.25378	4164
liver	54	2.74426	1.07036	148.19000

Simple Statistics

Variable	Minimum	Maximum
logsurv	1.53150	2.91910
blood	2.60000	11.20000
prognostic	8.00000	96.00000
enzyme	23.00000	119.00000
liver	0.74000	6.40000

Pearson Correlation Coefficients, N = 54
Prob > |r| under H0: Rho=0

	logsurv	blood	prognostic	enzyme	liver
logsurv	1.00000	0.34640 0.0103	0.59289 <.0001	0.66512 <.0001	0.72621 <.0001
blood	0.34640 0.0103	1.00000	0.09012 0.5169	-0.14963 0.2802	0.50242 0.0001
prognostic	0.59289 <.0001	0.09012 0.5169	1.00000	-0.02361 0.8655	0.36903 0.0060
enzyme	0.66512 <.0001	-0.14963 0.2802	-0.02361 0.8655	1.00000	0.41642 0.0017
liver	0.72621 <.0001	0.50242 0.0001	0.36903 0.0060	0.41642 0.0017	1.00000

Note that it is possible to force variables to be in model. For example assume that you want to have `prognostic` in every model examined.

You can force variables to be in the model by listing them first in the model statement and to add an `Include=`n option. For example

```
Stepwise_Force: MODEL logsurv = prognostic liver blood enzyme /
    SELECTION = STEPWISE SLENTRY = 0.05 SLSTAY = 0.05
    INCLUDE = 1         /* force prognostic into model */
    DETAILS = SUMMARY; /* only print summary info */
```

Summary of Stepwise Selection

Step	Variable Entered	Variable Removed	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	enzyme		2	0.4615	0.8130	283.669	125.83	<.0001
2	blood		3	0.1594	0.9723	3.0393	288.17	<.0001

Model Selection Criteria

Consider the situation with K potential predictor values $x_1, x_2, \ldots, x_K$. We want to find small, good models using these variables.

To do this we need criteria to decide what is a good model and what is a poor model. There are many available, some of which **SAS** will calculate. We will discuss two, Adjusted R^2 and Mallow's C_p.

For what follows, it is assumed that an intercept term will always be included in the model. Usually you will want to include it anyways, so this is no real constraint.

Adjusted R^2

As mentioned earlier, R^2 is not a good selection criterion as it must be maximized when all variables are in the model. In additional consider the nested models

$$\begin{aligned}
&\text{Model 1}: \quad y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \epsilon \\
&\text{Model 2}: \quad y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \ldots + \beta_{k+m} x_{k+m} \epsilon
\end{aligned}$$

So model 2 has an additional m predictor variables. It can be shown that $R^2(M2) \geq R^2(M1)$, i.e. adding variables can't decrease R^2.

Now assume that the true model is

$$y = \beta_0 + \epsilon$$

i.e. none of the predictors are associated with the response. Now assume that a model with k predictors is tried with a data set with n observations.

Then it can be shown that

$$E[R^2] \approx \frac{k}{n-1}$$

For these reasons a modified measure, the adjusted R^2

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k-1}\right)(1-R^2) = 1 - \left(\frac{n-1}{n-k-1}\right)\frac{SSE}{SSTot} = 1 - \frac{MSE}{\frac{SSTot}{n-1}}$$

is often used.

It can be shown that adding variables to a model won't necessarily increase $\bar{R}^2$. In fact $\bar{R}^2(M2) \geq \bar{R}^2(M1)$ only if the F statistic for examining

$$H_0 : \beta_{k+1} = \ldots = \beta_{k+m} = 0 \qquad \text{vs} \qquad H_A : \text{not all 0}$$

is greater than 1.

In addition is can be shown in the case where the true model is

$$y = \beta_0 + \epsilon$$

that when investigating model 1

$$E[\bar{R}^2] \approx 0$$

In the case when model 1 is the correct model, it can be shown that when investigating model 1 or any version of model 2 (one or more extra predictors) that

$$E[\bar{R}^2] > 0$$

(at least approximately)

Thus one approach to choosing a model is to select the one with the largest adjusted R^2.

To search for models based on adjusted R^2 in **SAS**, the following code can be used.

```
PROC REG DATA=surgery;

  Adjusted_Rsq: MODEL logsurv = blood prognostic enzyme liver /
    SELECTION = ADJRSQ;
```

This will print a summary for all possible models, ordered by adjusted R^2.

Number in Model	Adjusted R-Square	R-Square	Variables in Model
3	0.9707	0.9723	blood prognostic enzyme
4	0.9701	0.9724	blood prognostic enzyme liver
3	0.8759	0.8829	prognostic enzyme liver
2	0.8056	0.8130	prognostic enzyme
3	0.7023	0.7192	blood enzyme liver
2	0.6742	0.6865	enzyme liver
2	0.6358	0.6496	prognostic liver
2	0.6319	0.6458	blood enzyme
3	0.6290	0.6500	blood prognostic liver
1	0.5183	0.5274	liver
2	0.5093	0.5278	blood liver
1	0.4317	0.4424	enzyme
2	0.4160	0.4381	blood prognostic
1	0.3390	0.3515	prognostic
1	0.1031	0.1200	blood

Note that this selection criterion is the equivalent to minimizing MSE as

$$\bar{R}^2 = 1 - \frac{MSE}{\frac{SSTot}{n-1}}$$

$\bar{R}^2$ will increase when MSE decrease.

Mallow's C_p

Lets consider how well $\hat{y}_i$ does as an estimator of $\mu_i = E[y_i]$

$$\begin{aligned}\hat{y}_i - \mu_i &= (E[\hat{y}_i] - \mu_i) + (\hat{y}_i - E[\hat{y}_i]) \\ &= \text{Bias} + \text{Random error}\end{aligned}$$

$$\begin{aligned}E[(\hat{y}_i - \mu_i)^2] &= (E[\hat{y}_i] - \mu_i)^2 + \text{Var}(\hat{y}_i) \\ &= \text{Bias}^2 + \text{Variance}\end{aligned}$$

For a model with $p-1$ predictors (plus the intercept), let the total squared bias,$SSB(p)$, be defined as

$$SSB(p) = \sum_{i=1}^{n} (E[\hat{y}_i] - \mu_i)^2$$

and define the standardized total mean square error as

$$\begin{aligned}\Gamma_p &= \frac{1}{\sigma^2}\left\{\sum_{i=1}^{n}(E[\hat{y}_i] - \mu_i)^2 + \sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i)\right\} \\ &= \frac{SSB(p)}{\sigma^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i)\end{aligned}$$

It can be shown that

$$\sum_{i=1}^{n}\mathrm{Var}(\hat{y}_i) = p\sigma^2$$

and

$$E[SSE(p)] = SSB(p) + (n-p)\sigma^2$$

Plugging these in gives

$$\begin{aligned}\Gamma_p &= \frac{1}{\sigma^2}\left\{E[SSE(p)] - (n-p)\sigma^2 + p\sigma^2\right\} \\ &= \frac{E[SSE(p)]}{\sigma^2} - n + 2p\end{aligned}$$

So the idea is to find the model with a low value of Γ_p. Since we can't determine this, we need to estimate it. Suppose that $\hat{\sigma}^2$ is a good estimate of σ^2. Then if we replace $E[SSE(p)]$ by the observed value $SSE(p)$, then an estimate of Γ_p is

$$C_p = \frac{SSE(p)}{\hat{\sigma}^2} - n + 2p$$

If the p-term model has negligible bias, then $SSB(p) = 0$ which implies $E[SSE(p)] = (n-p)\sigma^2$ and

$$E[C_p|\text{Bias} = 0] \approx \frac{(n-p)\sigma^2}{\sigma^2} - n + 2p = p$$

The usual estimate used for $\hat{\sigma}^2$ is $MSE(K+1)$, the MSE from the model using all predictors, which gives

$$C_p = \frac{SSE(p)}{MSE(K+1)} - n + 2p$$

Note that with the definition, for the model with all predictors

$$C_p = p = K + 1$$

So usually we want to find a model with a small C_p and $C_p \approx p$.

One way to think of Mallow's C_p is to look at the SSE but then to add a penalty term based on the number of parameters in the model. So you want to keep adding terms while the SSE continues to drop at a fairly high rate.

To select models by C_p, the following **SAS** code can be used

```
PROC REG DATA=surgery;

  Cp: MODEL logsurv = blood prognostic enzyme liver /
    SELECTION = CP;
```

The resulting output is similar to what we saw earlier

Number in Model	C(p)	R-Square	Variables in Model
3	3.0393	0.9723	blood prognostic enzyme
4	5.0000	0.9724	blood prognostic enzyme liver
3	161.6625	0.8829	prognostic enzyme liver
2	283.6695	0.8130	prognostic enzyme
3	451.9870	0.7192	blood enzyme liver
2	507.8964	0.6865	enzyme liver
2	573.4372	0.6496	prognostic liver
3	574.7100	0.6500	blood prognostic liver
2	580.1453	0.6458	blood enzyme
1	788.1481	0.5274	liver
2	789.3404	0.5278	blood liver
1	938.8651	0.4424	enzyme
2	948.5500	0.4381	blood prognostic
1	1100.012	0.3515	prognostic
1	1510.590	0.1200	blood

So in this case C_p selects the same model as $\bar{R}^2$ (and stepwise and backward). However the ordering of the different models is slightly different. For larger datasets, these methods can find different models. For smaller examples, these methods will tend to find similar models.

In **SAS** there are a couple of other options for `SELECTION`. These are

- `RSQUARE`: Lists all models by R^2.
- `MAXR` and `MINR`: Stepwise type procedures which add variables based on changes in R^2. These approaches try to find the best 2 variable model, 3 variable model, 4 variable model, etc
- `NONE`: Fits the model with all variables.