Building The Regression Model
Model Selection

AU STAT-615

Emil Hvitfeldt

2021-03-24

Model Selection

The goal of chapter 9 is to discuss methods for selecting predictor variables regarding an exploratory observational study

Criteria for Model Selection

In general for any set of $p - 1$ predictors, 2^{p-1} alternative models can be constructed

This becomes a very difficult talk as p increases

p	possibilities
1	2
2	4
3	8
4	16
5	32
10	1024
15	32768

Criteria for Model Selection

We will focus on six criteria for comparing the regression models

- R_p^2
- $R_{\alpha,p}^2$
- C_p
- AIC_p
- SBC_p
- $PRESS_p$

Notation

- We denote the number of potential X variables by $P-1$

- All regression models contain an intercept β_0

- The number of X variables in a subset will be denoted by $p-1$. Thus we have $1 \leq p \leq P$

- We assume that the number of observations exceeds the number of potential parameters

$$n > P$$

R_p^2 or SSE_p

The R_p^2 criterion is equivalent to using the error sum of squares SSE_p as the criterion

With the SSE_p criterion subsets for which SSE_p is small are considered "good"

The equivalence follows from

$$R_p^2 = 1 - \frac{SSE_p}{SSTO}$$

since $SSTO$ is constant for all possible regression models

R_p^2 or SSE_p

Note:

The intent in using the R_p^2 criterion is to find the point where adding more X variables is not worthwhile because it leads to a very small increase in R_p^2

$R^2_{\alpha,p}$ or MSE_p

The adjusted coefficient of multiple determination $R^2_{\alpha,p}$ can be suggested as an alternative of R^2_p

$$R^2_{\alpha,p} = 1 - \frac{n-1}{n-p} \cdot \frac{SSE_p}{SSTO} = 1 - \frac{MSE_p}{\frac{SSTO}{n-1}}$$

it can be shown that $R^2_{\alpha,p}$ increase if and only if MSE_p decreases since $\frac{SSTO}{n-1}$ is fixed for any number of predictors

Mallows' C_p Criterion

The criterion is concerned with the total mean squared error of the n fitted values for each subset regression model

The mean squared error concept involves the total error in each fitted value

$$\hat{Y}_i - \mu_i$$

Mallows' C_p Criterion

This total error is made up of a bias component and a random error component

Bias component

$$E\{\hat{Y}_i\} - \mu_i$$

Random error component

$$\hat{Y}_i - E\{\hat{Y}_i\}$$

Mallows' C_p Criterion

It can be shown that

$$E\left\{\hat{Y}_i - \mu_i\right\}^2 = \left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sigma^2\left\{\hat{Y}_i\right\}$$

therefore, the total mean squared error for all n fitted values is given by

$$\begin{aligned}\sum_{i=1}^{2}\{Y_i - \mu_i\}^2 &= \sum_{i=1}^{n}\left[\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sigma^2\left\{\hat{Y}_i\right\}\right] \\ &= \sum_{i=1}^{n}\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sum_{i=1}^{n}\sigma^2\left\{\hat{Y}_i\right\}\end{aligned}$$

Mallows' C_p Criterion

The criterion measure, denoted by Γ_p is simply

$$\Gamma_p = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}\left(E\{\hat{Y}_i\} - \mu_i\right)^2 + \sum_{i=1}^{n}\sigma^2\left\{\hat{Y}_i\right\}\right]$$

The model which includes all $P-1$ potential X variables is assumed to have been carefully chosen so that $MSE(X_1, X_2, \ldots, X_{P-1})$ is an unbiased estimator of σ^2

Mallows' C_p Criterion

It can be shown that C_p is an estimator of Γ_p

$$C_p = \frac{SSE_p}{MSE(X_1, \ldots, X_{P-1})} - (n - 2p)$$

Mallows' C_p Criterion

Notes:

When the C_p values for all possible regression models are plotted against p, those models with little bias will tend to fall near the line $C_p = p$

Models with substantial bias will tend to fall considerably above this line

In using the C_p criterion, we seek to identify subsets of X variables for which the C_p value is small **and** the C_p value is near p (Bias of regression model is small)

AIC_p and SBC_p

Alternative criteria that provide penalties for adding predictors are Akaike's Information Criterion (AIC_p) and Schwarz' Bayesian Criterion (SBC_p)

We search for models that have small values of AIC_p or SBC_p

$$AIC_p = n\ln(SSE_p) - n\ln(n) + 2p$$
$$SBC_p = n\ln(SSE_p) - n\ln(n) + \ln(n)p$$

- $n\ln(n)SSE_p$ decreases as p increases

$PRESS_p$ Criterion

The $PRESS_p$ criterion is a measure of how well the use of the fitted values for a subset model can predict the observed responses Y_i

Note:

The press measure differs from SSE in that each fitted value $\hat{Y}_i$ for the $PRESS$ criterion is obtained by deleting the ith case from the data set, estimating the regression function for the subset model from the remaining $n-1$ cases, and then using the fitted regression function to obtain the predicted value $\hat{Y}_{i(i)}$ for the ith case

$$PRESS_p = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_{i(i)}\right)^2$$

models with small press values are considered good candidate models