The validated data is used to test if assess if the fitting degree of the model is good enough and make a comparison between the top two models to select one from them. The statistic used here is MSE, the smaller the MSE is, the better the fitting degree will be. The value of covariances in the validated data was used to predict the response of the two models and calculate the value of MSE of each model. The result shows that the value of step model is slightly lower than the value of forward model, which seems to mean it is the better model. But it still depends on other conditions like whether it fits the assumptions or the value of R-square.

The 5-fold cross validation is used to test the accuracy of the algorithm. It divides the data set into 5 parts and take turns to take 4 of them as training data and one as test data and carry out the test. The correct rate will be obtained for each test. The average value of the accuracy of the results of 5 times is used as the estimation of the accuracy of the algorithm. The advantage of this method is to ensure that every subsample is trained and tested to reduce the generalization error.

The function was written to calculate the value of MSE of the two models, the result of which shows that the value of for model is much smaller.

|  |  | | |
| --- | --- | --- | --- |
|  | | formod | Stepmod | stepmod |
| 5-fold MSE | | 458.3237 | 525.4774 |
| validetion-dataset MSE | | 309.9346 | 306.8825 |