**Q1. We perform best subset, forward stepwise, and backward stepwise selection on a single data set. For each approach, we obtain p+1 models containing 0,1,2,⋯,p predictors. Explain your answers :**

1. **Which of the three models with *k* predictors has the smallest training RSS ?**

Best subset selection: The model with k predictors is the model with the smallest RSS among all the models with k predictors.   
Forward stepwise selection: The model with k predictors is the model with the smallest RSS among the p−k models which augment the predictors in with one additional predictor.   
Backward stepwise selection: The model with k predictors is the model with the smallest RSS among the k models which contains all but one of the predictors in .   
So, the model with k predictors which has the smallest training RSS is the one obtained from best subset selection as it is the one selected among all k predictors models.

1. **Which of the three models with *k* predictors has the smallest test RSS ?**

It is hard to say, the best subset selection may have the smallest test RSS because it takes into account more models than the other methods. However, the other methods might also pick a model with smaller test RSS by sheer luck.

1. **True or False :**
2. **The predictors in the *k*-variable model identified by forward stepwise are a subset of the predictors in the (*k*+1)-variable model identified by forward stepwise selection.**

True. The model with (k+1) predictors is obtained by augmenting the predictors in the model with k predictors with one additional predictor.

1. **The predictors in the *k*-variable model identified by backward stepwise are a subset of the predictors in the (*k*+1)-variable model identified by backward stepwise selection.**

True. The model with k predictors is obtained by removing one predictor from the model with (k+1) predictors.

1. **The predictors in the *k*-variable model identified by backward stepwise are a subset of the predictors in the (*k*+1)-variable model identified by forward stepwise selection.**

False. There is no direct link between the models obtained from forward and backward selection.

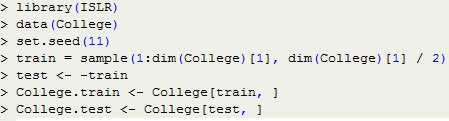
1. **The predictors in the *k*-variable model identified by forward stepwise are a subset of the predictors in the (*k*+1)-variable model identified by backward stepwise selection.**

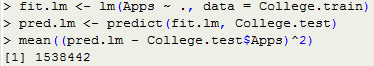
False. There is no direct link between the models obtained from forward and backward selection.

1. **The predictors in the *k*-variable model identified by best subset are a subset of the predictors in the (*k*+1)-variable model identified by best subset selection.**

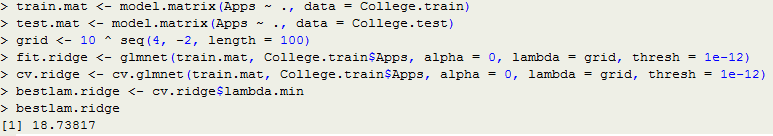
False. The model with (k+1) predictors is obtained by selecting among all possible models with (k+1) predictors, and so does not necessarily contain all the predictors selected for the k-variable model.

**Q4 In this exercise, we will predict the number of applications received using the other variables in the “College” data set.**

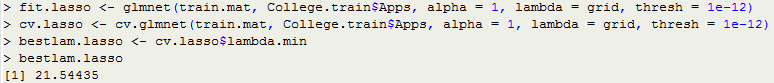
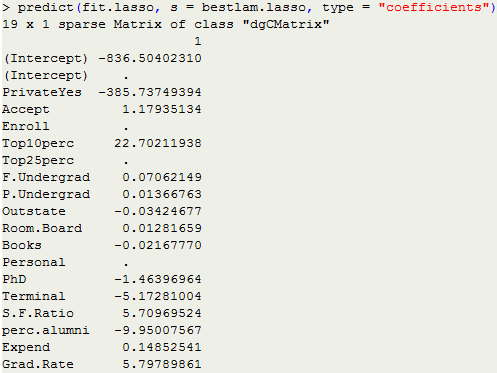
1. **Split the data set into a training and a test set.**
2. **Fit a linear model using least squares on the training set, and report the test error obtained.**

****The test MSE is 1.538442\*10^6

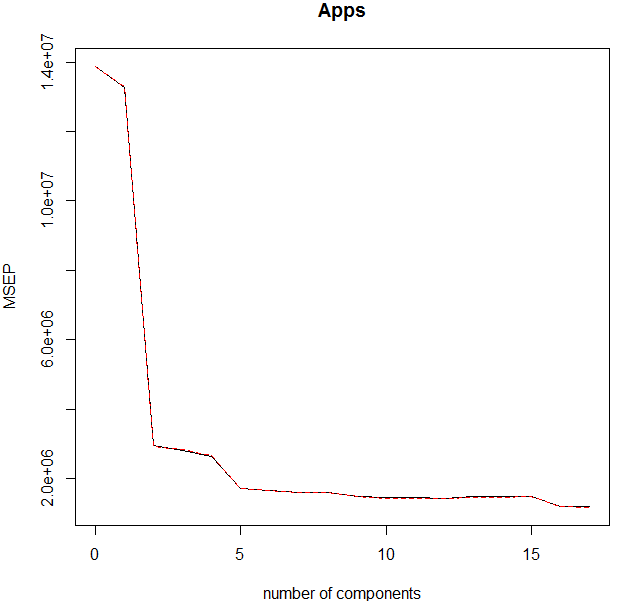
1. **Fit a ridge regression model on the training set, with *λ* chosen by cross-validation. Report the test error obtained.**

**  
**The test MSE is 1.608859\*10^6

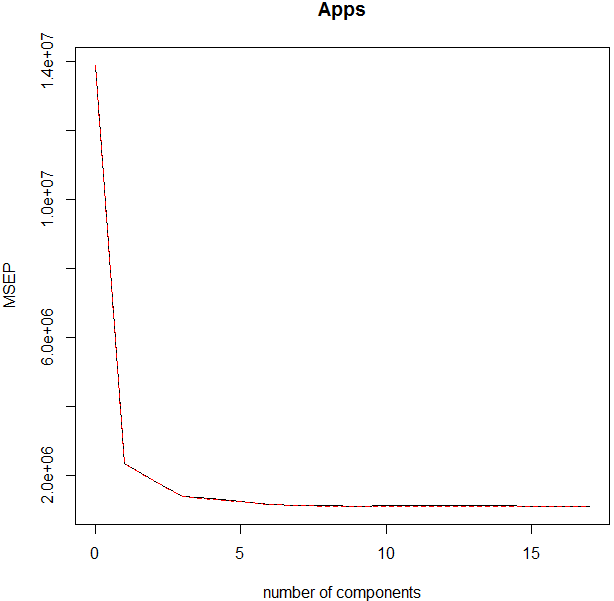
1. **Fit a lasso model on the training set, with *λ* chosen by cross-validation. Report the test error obtained, along with the number of non-zero coefficient estimates.**

**  
**The test MSE is also higher for ridge regression than for least squares. ****

1. **Fit a PCR model on the training set, with *M* chosen by cross-validation. Report the test error obtained, along with the value of *M* selected by cross-validation.**

**  
**The test MSE is also higher for PCR than for least squares.

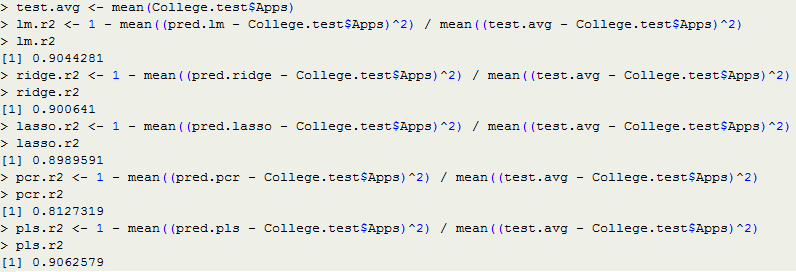
1. **Fit a PLS model on the training set, with *M* chosen by cross-validation. Report the test error obtained, along with the value of *M* selected by cross-validation.**

**  
**

****The test MSE is lower for PLS than for least squares.

1. **Comment on the results obtained. How accurately can we predict the number of college applications received? Is there much difference among the test errors resulting from these five approaches ?**

We compute the test *R^2* for all models.

****

As shown above,the test *R2* for least squares is 0.9044281, the test *R2* for ridge is 0.9000641, the test *R2* for lasso is 0.8989591, the test *R2* for pcr is 0.8127319 and the test *R2* for pls is 0.9062579. All models, except PCR, predict college applications with high accuracy.