Use the data file ***divusa.csv*** for this assignment. This dataset collects the information on divorce rates. This data frame includes 77 observations on 7 variables.

Variable	Description
year	the year from 1920-1996
divorce	divorce per 1000 women aged 15 or more
unemployed	unemployment rate
femlab	percent female participation in labor force aged 16+
marriage	marriages per 1000 unmarried women aged 16+
birth	births per 1000 women aged 15-44
military	military personnel per 1000 population

Use this data with divorce as the response and the other variables as predictors. Implement the following variable selection methods to determine the "best" model. Please also comment on how you reach your final decision for each method.

a. Backward Elimination
b. AIC
c. BIC
d. Adjusted R^2
e. Mallows C_p

To predict divorce rates, data was gathered on six predictors: year, unemployment rate, female participation in labor force, marriage rate, birth rate, and military personnel percentage. Several variable selection methods were implemented to try to select the best linear model. $\alpha_{criteria} = 0.05$ was used in all selection methods.

Backwards Elimination

A multiple linear regression analysis was performed on the full model with all six predictors. The largest p-value of 0.3622 belonged to unemployment rate, so that variable was eliminated. The new refitted model included the remaining five predictors who all had p-values less than the $\alpha_{criteria} = 0.05$ significant level. The final model selected from backwards elimination is

$$divorce = 405.6167 - 0.2179year + 0.8548femlab + 0.1593marriage - 0.1101birth - 0.0412military$$

Forward Selection Using AIC

The forward selection process using AIC began with the null model with no predictors. The smallest AIC of 134.28 belonged to female participation in labor force, so that variable was added to the null model. The subsequent refitted models had smallest AIC of 111.83 belonged to birth rate, 85.196 belonged to marriage rate, 76.691 belonged to year, and 69.33 belonged to military personnel percentage, respectively. The final model selected from forward selection using AIC is

$$divorce = 405.6167 - 0.2179year + 0.8548femlab + 0.1593marriage - 0.1101birth - 0.0412military$$

Forward Selection Using BIC

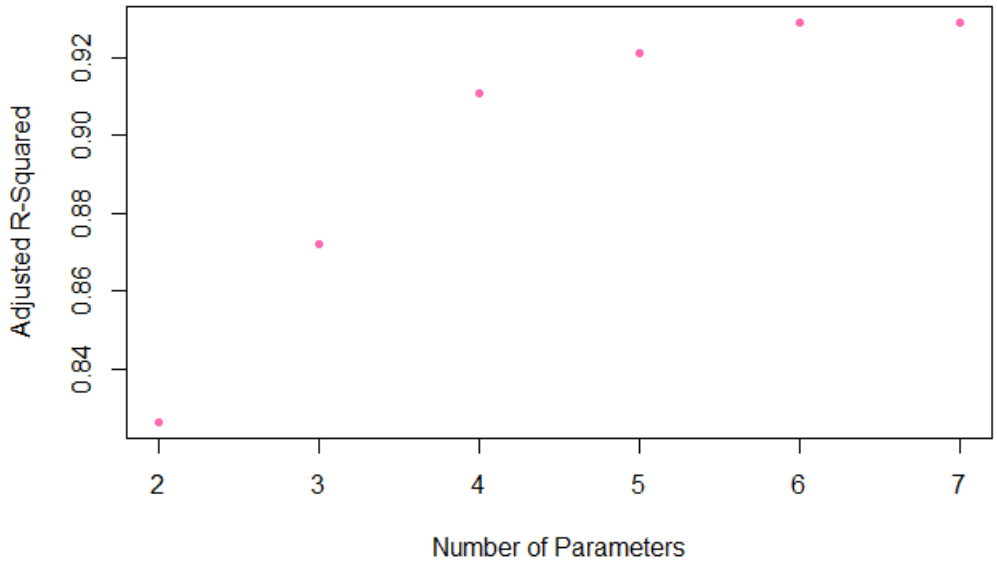
The forward selection process using BIC began with the null model with no predictors. The smallest BIC of 138.97 belonged to female participation in labor force, so that variable was added to the null model. The subsequent refitted models had smallest BIC of 118.86 belonged to birth rate, 94.571 belonged to marriage rate, 88.410 belonged to year, and 83.393 belonged to military personnel percentage, respectively. The final model selected from forward selection using BIC is

$$divorce = 405.6167 - 0.2179year + 0.8548femlab + 0.1593marriage - 0.1101birth - 0.0412military$$

Adjusted R^2

Adjusted R^2 values of linear models constructed from 2 to 6 parameters were plotted. The highest R^2 value of 0.92895 belongs to the model with 6 parameters including the intercept, so the final model selected using R^2 criteria includes all the predictors except for unemployment rate.

$$divorce = 405.6167 - 0.2179year + 0.8548femlab + 0.1593marriage - 0.1101birth - 0.0412military$$



Mallow's C_p Statistics

Mallow's C_p Statistics of linear models constructed from 2 to 6 predictors were plotted. The subset of 6 predictors with a Mallow's C_p statistic of 5.8413 was closest to the $C_p = p$ line, so the final model selected using Mallow's C_p statistic includes all the predictors except for unemployment rate.

$$divorce = 405.6167 - 0.2179year + 0.8548femlab + 0.1593marriage - 0.1101birth - 0.0412military$$