June Suh

Jessica Warren

Abbie Pearson

Nov. 15, 2016

STA4203

Homework 10

1. Use the *gala* data with *species* as the response and the other variables except *endemics* as predictors. Implement the following variable selection methods to determine the best model:

a) Forward selection with significance level 0.05. Report what variables were selected.

Code:

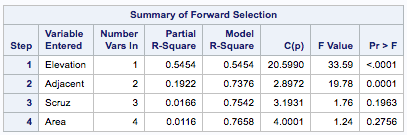
*Proc reg data = gala;*

*Model species = area elevation nearest scruz adjacent/*

*selection=forward;*

*Run;*

Results:



Elevation and Adjacent are significant to 0.05 level.

b) Backward elimination with significance level 0.05. Report what variables were selected.

Code:

*Proc reg data=gala;*

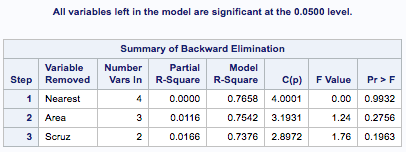
*model species= area elevation nearest scruz adjacent*

*/selection=backward slstay=0.05;*

*run;*

*quit;*

Results:



The variables not listed are significant to 0.05 level and are Elevation and Adjacent.

c) Adjusted R2 . Report what variables were selected and the Adjusted R2 of the model.

Code:

*Proc reg data=gala;*

*model species= area elevation nearest scruz adjacent/*

*selection=adjrsq;*

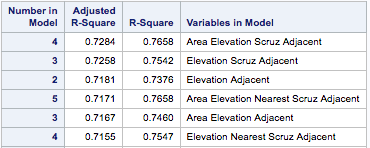
*run;*

*quit;*

Results:

Adjusted R-squared of the model: 0.7171

Screen Shot 2016-11-13 at 5.01.24 PM.png



Adjusted R2 shows that Area, Elevation, Scruz and Adjacent is the best combination.

d) AIC. Report what variables were selected and the AIC of the model.

Code:

*Proc reg data=gala;*

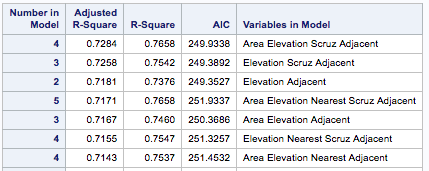
*model species= area elevation nearest scruz adjacent/*

*selection=adjrsq aic;*

*run;*

*quit;*

Results:



The minimum AIC is 249.3527 with the combination Elevation and Adjacent.

e) Mallows Cp. Report what variables were selected and the Cp of the model.

Code:

*Proc reg data=gala;*

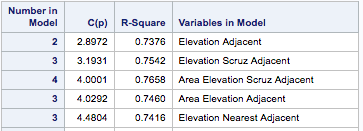
*model species= area elevation nearest scruz adjacent/*

*selection=cp;*

*run;*

*quit;*

Results:



Elevation and Adjacent is the best combination.

2. Using the *divusa* dataset with *divorce* as the response and the other variables as predictors, repeat the work of the first question.

a) Forward selection with significance level 0.05. Report what variables were selected.

Code:

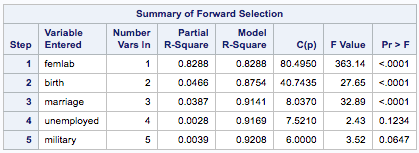
*Proc reg data = divusa;*

*Model divorce= unemployed femlab marriage birth military/*

*selection=forward;*

*Run;*

Results:



Femlab, birth, and marriage are all significant.

b) Backward elimination with significance level 0.05. Report what variables were selected.

Code:

*Proc reg data=divusa;*

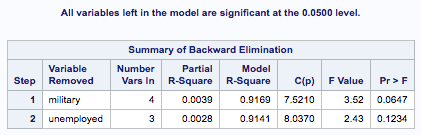
*model divorce= unemployed femlab marriage birth military*

*/selection=backward slstay=0.05;*

*run;*

*quit;*

Results:



The variables not listed are significant and include Femlab, birth, and marriage.

c) Adjusted R2 . Report what variables were selected and the Adjusted R2 of the model.

Code:

*Proc reg data=divusa;*

*model divorce= unemployed femlab marriage birth military*

*/selection=adjrsq;*

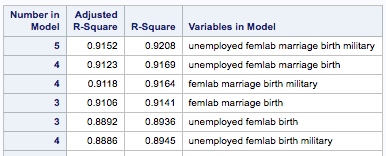
*run;*

*quit;*

Results:

Adjusted R-squared of model: 0.9152

Screen Shot 2016-11-14 at 6.03.56 PM.png



Adjusted R2 shows that unemployed, femlab, marriage, birth, and military is the best combination.

d) AIC. Report what variables were selected and the AIC of the model.

Code:

*Proc reg data=divusa;*

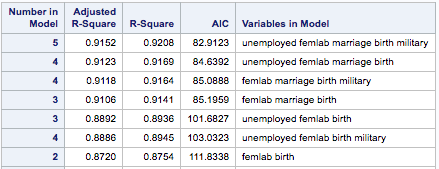
*model divorce= unemployed femlab marriage birth military*

*/selection=adjrsq aic;*

*run;*

*quit;*

Results:



Minimum AIC is 82.9123 with combination unemployed, femlab, marriage, birth, and military.

e) Mallows Cp. Report what variables were selected and the Cp of the model.

Code:

*Proc reg data=divusa;*

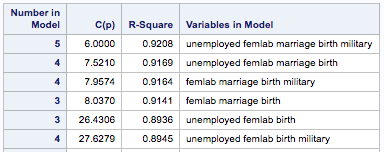
*model divorce= unemployed femlab marriage birth military*

*/selection=cp;*

*run;*

*Quit;*

Results:



Variables selected are unemployed, femlab, marriage, birth, and military.