Abbie Pearson

June Suh

Jessica Warren

Nov. 22, 2016

STA4203

Blue code = right

Red code = wrong

Homework 11

1. Consider the seatpos dataset, with hipcenter as the dependent variable and the other variables as predictors. First, standardize all the predictors. Then divide the standardized dataset into seatpos0 and seatpos1,containing the even and odd observations respectively. We will use the Root Mean Square Error (RMSE) in this homework

where RSS is the residual sum of squares

between the actual values and the predicted values from one of the models below.

Code:

*proc import out=seatpos*

*datafile="/home/aep120/seatpos.csv"*

*dbms = csv replace;*

*run;*

*proc standard data=seatpos mean=0 std=1 out=seatposNew;*

*var age weight htshoes ht seated arm thigh leg;*

*run;*

*data seatpos0 seatpos1;*

*set seatposNew;*

*if mod(\_n\_,2) = 0 then*

*do;*

*output seatpos0;*

*end;*

*else*

*do;*

*output seatpos1;*

*end;*

*run;*

a) Perform Principal Component Regression with 5 principal directions, on the seatpos0 data with hipcenter as the dependent variable and the other variables as predictors. Report the regression equation.

Code:

*proc princomp data=seatpos0 outstat=PCA;*

*var age weight htshoes ht seated arm thigh leg;*

*run;*

*proc score data=seatpos0 score=PCA out=seat0Pca;*

*var age weight htshoes ht seated arm thigh leg;*

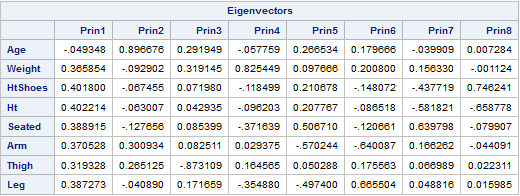
*run;*

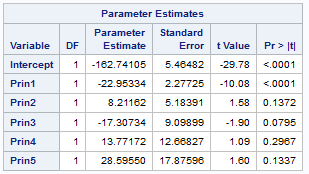
*proc reg data=seat0Pca outest=modelparams;*

*model hipcenter=prin1-prin5;*

*run;*

*quit;*





Equation:

b) Report the RMSE of the model from a) on seatpos0.

Code:

*proc score data=seat0pca score=modelparams out=seatpos0P residual type=parms;*

*var hipcenter prin1-prin5;*

*run;*

*proc univariate data=seatpos0P noprint;*

*var model1;*

*output out=seat0stat uss=ss1;*

*run;*

*data seat0stat;*

*set seat0stat;*

*rmse=sqrt(ss1/19);*

*run;*



RMSE is 19.70369

c) Report the RMSE of the model from a) on seatpos1.

Code:

*proc score data=seatpos1 score=PCA out=seat1Pca;*

*var age weight htshoes ht seated arm thigh leg;*

*run;*

*proc score data=seat1pca score=seatpos1pca out=seatpos1P residual type=parms;*

*var hipcenter prin1-prin5;*

*run;*

*proc univariate data=seatpos1P noprint;*

*var model1;*

*output out=seat1stat uss=ss1;*

*run;*

*data seat1stat;*

*set seat1stat;*

*rmse=sqrt(ss1/19);*

*run;*



RMSE is 46.95457

d) Perform an Ordinary Least Squares Regression on the seatpos0 data with hipcenter as the dependent variable and the other variables as predictors. Report the R2 of the obtained model.

Code:

*proc reg data=seatpos0;*

*model hipcenter=age weight htshoes ht seated arm thigh leg;*

*run;*

*quit;*

Screen Shot 2016-11-20 at 5.49.44 PM.png

R^2 is 0.9255.

e) On the model from d) and the seatpos0 data perform variable selection with the AIC criterion. Fit a model on the selected variables and the seatpos0 data and report the regression equation of the obtained model.

Code:

*Proc reg data=seatpos0;*

*model hipcenter=age weight htshoes ht seated arm thigh leg*

*/selection=adjrsq aic;*

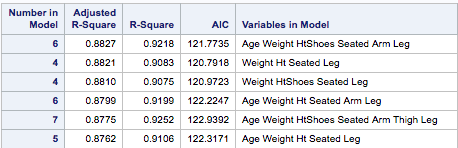
*run;*

*quit;*

*proc reg data=seatpos0 outest=estest;*

*model hipcenter = weight ht seated leg; //model hipcenter=leg;*

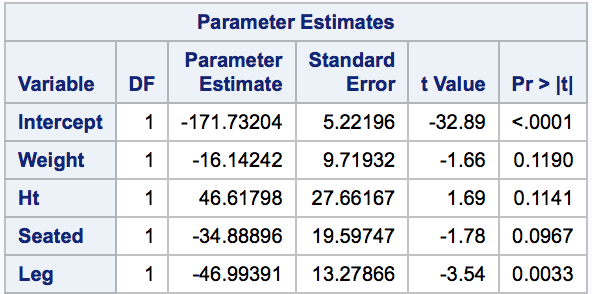
*run;*



Model variable selection: Weight, ht, seated, leg

Model variable selection: leg

You were supposed to look for the smallest AIC (=120.10 and everything else is wrong...)



Equation: hipcenter = -171.73204 - 16.14242\*weight + 46.61798\*ht - 34.88896\*seated - 46.99391\*leg

Hipcenter = -171.63148 - 49.99518\*leg

f) Report the RMSE of the model from e) on seatpos0.

Code:

*proc score data=seatpos0 score=estest out=seatpos0PP residual type=parms;*

*var hipcenter weight ht seated leg; // var hipcenter leg;*

*run;*

*proc univariate data=seatpos0PP noprint;*

*var model1;*

*output out=seat0stat1 uss=ss1;*

*run;*

*data seat0stat1;*

*set seat0stat1;*

*rmse=sqrt(ss1/19);*

*run;*



RMSE is 18.4594

RMSE is 21.23

g) Report the RMSE of the model from e) on seatpos1.

Code:

*proc score data=seatpos1 score=estest out=seatpos1PP residual type=parms;*

*var hipcenter weight ht seated leg; //var hipcenter leg;*

*run;*

*proc univariate data=seatpos1PP noprint;*

*var model1;*

*output out=seat1stat1 uss=ss1;*

*run;*

*data seat1stat1;*

*set seat1stat1;*

*rmse=sqrt(ss1/19);*

*run;*



RMSE is 54.3579

RMSE is 47.89