Methods for Choosing Predictors

Alyssa Warnock, Helen Johnston, Ashley Wade, Hunter Barbee

11/28/2021

Consider the SleepStudy file described on Sakai. In this exercise, you will consider models to predict Grade Point Average (GPA) using any of the predictors.

sleep <- read.csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/SleepStudy.csv")  
  
source("https://raw.githubusercontent.com/JA-McLean/STOR455/master/scripts/ShowSubsets.R")  
  
Full=lm(GPA~factor(Gender)+factor(ClassYear)+factor(EarlyClass) + LarkOwl + NumEarlyClass + ClassesMissed + CognitionZscore + PoorSleepQuality + DepressionScore + AnxietyScore + StressScore + DepressionStatus + AnxietyStatus + Stress + DASScore + Happiness + AlcoholUse + Drinks + WeekdayBed + WeekdayRise + WeekdaySleep + WeekendBed + WeekendRise + WeekendSleep + AverageSleep + AllNighter, data=sleep)

Use forward selection until you have a model for GPA.

none=lm(GPA~1,data=sleep)  
  
MSE=(summary(Full)$sigma)^2  
 #Specify the direction  
step(none,scope=list(upper=Full),scale=MSE, direction= "forward", trace = FALSE)

##   
## Call:  
## lm(formula = GPA ~ factor(ClassYear) + Drinks + CognitionZscore +   
## StressScore + DASScore, data = sleep)  
##   
## Coefficients:  
## (Intercept) factor(ClassYear)2 factor(ClassYear)3 factor(ClassYear)4   
## 3.52431 -0.34090 -0.33396 -0.29771   
## Drinks CognitionZscore StressScore DASScore   
## -0.01598 0.12020 0.03682 -0.01366

Use backwards elimination until you have a model for GPA.

Full=lm(GPA~factor(Gender)+factor(ClassYear)+factor(EarlyClass) + LarkOwl + NumEarlyClass + ClassesMissed + CognitionZscore + PoorSleepQuality + DepressionScore + AnxietyScore + StressScore + DepressionStatus + AnxietyStatus + Stress + DASScore + Happiness + AlcoholUse + Drinks + WeekdayBed + WeekdayRise + WeekdaySleep + WeekendBed + WeekendRise + WeekendSleep + AverageSleep + AllNighter, data=sleep)  
# Find the MSE for the full model  
  
MSE=(summary(Full)$sigma)^2  
# Backward: use the step( ) command starting with the full model  
#MSE = variance of the residuals   
  
step(Full,scale=MSE, trace = FALSE)

##   
## Call:  
## lm(formula = GPA ~ factor(ClassYear) + CognitionZscore + DepressionScore +   
## AnxietyScore + StressScore + Drinks, data = sleep)  
##   
## Coefficients:  
## (Intercept) factor(ClassYear)2 factor(ClassYear)3 factor(ClassYear)4   
## 3.52494 -0.33946 -0.33128 -0.29517   
## CognitionZscore DepressionScore AnxietyScore StressScore   
## 0.12009 -0.01182 -0.01642 0.02354   
## Drinks   
## -0.01610

Use a “best subsets” procedure to determine which predictors would explain the most variability in GPA.

source("https://raw.githubusercontent.com/JA-McLean/STOR455/master/scripts/ShowSubsets.R")  
library(leaps)  
all=regsubsets(GPA~factor(Gender)+factor(ClassYear)+factor(EarlyClass) + LarkOwl + NumEarlyClass + ClassesMissed + CognitionZscore + PoorSleepQuality + DepressionScore + AnxietyScore + StressScore + DepressionStatus + AnxietyStatus + Stress + DASScore + Happiness + AlcoholUse + Drinks + WeekdayBed + WeekdayRise + WeekdaySleep + WeekendBed + WeekendRise + WeekendSleep + AverageSleep + AllNighter, data= sleep, nvmax=26)

## Warning in leaps.setup(x, y, wt = wt, nbest = nbest, nvmax = nvmax, force.in =  
## force.in, : 1 linear dependencies found

## Reordering variables and trying again:

ShowSubsets(all)

## factor(Gender)1 factor(ClassYear)2 factor(ClassYear)3  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 )   
## 5 ( 1 ) \* \*  
## 6 ( 1 ) \* \*  
## 7 ( 1 ) \* \*  
## 8 ( 1 ) \* \*  
## 9 ( 1 ) \* \*  
## 10 ( 1 ) \* \*  
## 11 ( 1 ) \* \*  
## 12 ( 1 ) \* \*  
## 13 ( 1 ) \* \*  
## 14 ( 1 ) \* \*  
## 15 ( 1 ) \* \*  
## 16 ( 1 ) \* \*  
## 17 ( 1 ) \* \*  
## 18 ( 1 ) \* \*  
## 19 ( 1 ) \* \*  
## 20 ( 1 ) \* \*  
## 21 ( 1 ) \* \*  
## 22 ( 1 ) \* \*  
## 23 ( 1 ) \* \*  
## 24 ( 1 ) \* \*  
## 25 ( 1 ) \* \*  
## 26 ( 1 ) \* \*  
## 27 ( 1 ) \* \*  
## factor(ClassYear)4 factor(EarlyClass)1 LarkOwlNeither LarkOwlOwl  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 )   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \* \*   
## 9 ( 1 ) \* \*   
## 10 ( 1 ) \* \*   
## 11 ( 1 ) \* \*   
## 12 ( 1 ) \* \*   
## 13 ( 1 ) \* \*   
## 14 ( 1 ) \* \*   
## 15 ( 1 ) \* \*   
## 16 ( 1 ) \* \*   
## 17 ( 1 ) \* \*   
## 18 ( 1 ) \* \*   
## 19 ( 1 ) \* \* \*   
## 20 ( 1 ) \* \* \*   
## 21 ( 1 ) \* \* \*   
## 22 ( 1 ) \* \* \*   
## 23 ( 1 ) \* \* \*   
## 24 ( 1 ) \* \* \*   
## 25 ( 1 ) \* \* \*   
## 26 ( 1 ) \* \* \*   
## 27 ( 1 ) \* \* \*   
## NumEarlyClass ClassesMissed CognitionZscore PoorSleepQuality  
## 1 ( 1 )   
## 2 ( 1 ) \*   
## 3 ( 1 ) \*   
## 4 ( 1 ) \*   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \* \*   
## 10 ( 1 ) \* \*   
## 11 ( 1 ) \* \*   
## 12 ( 1 ) \* \*   
## 13 ( 1 ) \* \*   
## 14 ( 1 ) \* \* \*  
## 15 ( 1 ) \* \*   
## 16 ( 1 ) \* \* \*  
## 17 ( 1 ) \* \* \*  
## 18 ( 1 ) \* \* \*  
## 19 ( 1 ) \* \* \* \*  
## 20 ( 1 ) \* \* \* \*  
## 21 ( 1 ) \* \* \* \*  
## 22 ( 1 ) \* \* \* \*  
## 23 ( 1 ) \* \* \* \*  
## 24 ( 1 ) \* \* \* \*  
## 25 ( 1 ) \* \* \* \*  
## 26 ( 1 ) \* \* \* \*  
## 27 ( 1 ) \* \* \* \*  
## DepressionScore AnxietyScore StressScore DepressionStatusnormal  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 ) \*   
## 4 ( 1 ) \*   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \*   
## 10 ( 1 ) \*   
## 11 ( 1 ) \*   
## 12 ( 1 ) \*   
## 13 ( 1 ) \*   
## 14 ( 1 ) \*   
## 15 ( 1 ) \*   
## 16 ( 1 ) \*   
## 17 ( 1 ) \* \*  
## 18 ( 1 ) \* \*   
## 19 ( 1 ) \* \*   
## 20 ( 1 ) \* \*   
## 21 ( 1 ) \* \*   
## 22 ( 1 ) \* \*   
## 23 ( 1 ) \* \*   
## 24 ( 1 ) \* \*   
## 25 ( 1 ) \* \*   
## 26 ( 1 ) \* \*   
## 27 ( 1 ) \* \* \*  
## DepressionStatussevere AnxietyStatusnormal AnxietyStatussevere  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 )   
## 5 ( 1 )   
## 6 ( 1 )   
## 7 ( 1 )   
## 8 ( 1 )   
## 9 ( 1 )   
## 10 ( 1 )   
## 11 ( 1 )   
## 12 ( 1 )   
## 13 ( 1 )   
## 14 ( 1 )   
## 15 ( 1 )   
## 16 ( 1 )   
## 17 ( 1 )   
## 18 ( 1 ) \*  
## 19 ( 1 ) \*  
## 20 ( 1 ) \*  
## 21 ( 1 ) \*  
## 22 ( 1 ) \*  
## 23 ( 1 ) \*  
## 24 ( 1 ) \*  
## 25 ( 1 ) \*  
## 26 ( 1 ) \* \*  
## 27 ( 1 ) \* \*  
## Stressnormal DASScore Happiness AlcoholUseHeavy AlcoholUseLight  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 ) \*   
## 5 ( 1 )   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \*   
## 10 ( 1 ) \*   
## 11 ( 1 ) \*   
## 12 ( 1 ) \*   
## 13 ( 1 ) \* \*   
## 14 ( 1 ) \* \*   
## 15 ( 1 ) \* \*   
## 16 ( 1 ) \* \*   
## 17 ( 1 ) \* \*   
## 18 ( 1 ) \* \*   
## 19 ( 1 ) \* \*   
## 20 ( 1 ) \* \*   
## 21 ( 1 ) \* \*   
## 22 ( 1 ) \* \* \*  
## 23 ( 1 ) \* \* \*  
## 24 ( 1 ) \* \* \* \*  
## 25 ( 1 ) \* \* \* \* \*  
## 26 ( 1 ) \* \* \* \* \*  
## 27 ( 1 ) \* \* \* \* \*  
## AlcoholUseModerate Drinks WeekdayBed WeekdayRise WeekdaySleep  
## 1 ( 1 ) \*   
## 2 ( 1 ) \*   
## 3 ( 1 ) \*   
## 4 ( 1 ) \*   
## 5 ( 1 )   
## 6 ( 1 )   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \*   
## 10 ( 1 ) \* \*  
## 11 ( 1 ) \*   
## 12 ( 1 ) \* \*  
## 13 ( 1 ) \* \*  
## 14 ( 1 ) \* \*  
## 15 ( 1 ) \* \*  
## 16 ( 1 ) \* \*  
## 17 ( 1 ) \*   
## 18 ( 1 ) \* \*  
## 19 ( 1 ) \* \*  
## 20 ( 1 ) \* \*  
## 21 ( 1 ) \* \* \*  
## 22 ( 1 ) \* \* \*  
## 23 ( 1 ) \* \* \* \*  
## 24 ( 1 ) \* \* \* \*  
## 25 ( 1 ) \* \* \* \*  
## 26 ( 1 ) \* \* \* \*  
## 27 ( 1 ) \* \* \* \*  
## WeekendBed WeekendRise WeekendSleep AverageSleep AllNighter Rsq  
## 1 ( 1 ) 7.25  
## 2 ( 1 ) 12.43  
## 3 ( 1 ) 16.16  
## 4 ( 1 ) 19.71  
## 5 ( 1 ) 22.22  
## 6 ( 1 ) 27.01  
## 7 ( 1 ) 29.49  
## 8 ( 1 ) 30.20  
## 9 ( 1 ) 30.76  
## 10 ( 1 ) 31.18  
## 11 ( 1 ) \* \* 31.57  
## 12 ( 1 ) \* \* 31.83  
## 13 ( 1 ) \* \* 32.07  
## 14 ( 1 ) \* \* 32.25  
## 15 ( 1 ) \* \* \* \* 32.44  
## 16 ( 1 ) \* \* \* \* 32.64  
## 17 ( 1 ) \* \* \* \* \* 32.80  
## 18 ( 1 ) \* \* \* \* 32.95  
## 19 ( 1 ) \* \* \* 33.12  
## 20 ( 1 ) \* \* \* \* 33.30  
## 21 ( 1 ) \* \* \* \* 33.43  
## 22 ( 1 ) \* \* \* \* 33.50  
## 23 ( 1 ) \* \* \* \* 33.55  
## 24 ( 1 ) \* \* \* \* 33.59  
## 25 ( 1 ) \* \* \* \* 33.62  
## 26 ( 1 ) \* \* \* \* 33.64  
## 27 ( 1 ) \* \* \* \* 33.65  
## adjRsq Cp  
## 1 ( 1 ) 6.88 57.19  
## 2 ( 1 ) 11.73 42.08  
## 3 ( 1 ) 15.15 31.79  
## 4 ( 1 ) 18.41 22.08  
## 5 ( 1 ) 20.64 15.79  
## 6 ( 1 ) 25.23 1.97  
## 7 ( 1 ) 27.47 -4.22  
## 8 ( 1 ) 27.91 -4.58  
## 9 ( 1 ) 28.20 -4.43  
## 10 ( 1 ) 28.34 -3.80  
## 11 ( 1 ) 28.45 -3.10  
## 12 ( 1 ) 28.42 -1.96  
## 13 ( 1 ) 28.38 -0.75  
## 14 ( 1 ) 28.27 0.65  
## 15 ( 1 ) 28.16 2.05  
## 16 ( 1 ) 28.07 3.39  
## 17 ( 1 ) 27.93 4.87  
## 18 ( 1 ) 27.79 6.35  
## 19 ( 1 ) 27.66 7.80  
## 20 ( 1 ) 27.55 9.19  
## 21 ( 1 ) 27.38 10.78  
## 22 ( 1 ) 27.14 12.54  
## 23 ( 1 ) 26.88 14.37  
## 24 ( 1 ) 26.60 16.25  
## 25 ( 1 ) 26.30 18.16  
## 26 ( 1 ) 26.00 20.08  
## 27 ( 1 ) 25.69 22.04

Predictors: factor(ClassYear), ClassesMissed, CognitionZscore, PoorSleepQuality, StressScore, DASScore, Drinks, WeekdaySleep, WeekendRise, AllNighter.