R Notebook Group Project 3 - Topic 7: Methods for Choosing Predictors

library(readr)  
  
SleepData = read\_csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/SleepStudy.csv")

## Rows: 253 Columns: 27

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (5): LarkOwl, DepressionStatus, AnxietyStatus, Stress, AlcoholUse  
## dbl (22): Gender, ClassYear, NumEarlyClass, EarlyClass, GPA, ClassesMissed, ...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

head(SleepData)

## # A tibble: 6 x 27  
## Gender ClassYear LarkOwl NumEarlyClass EarlyClass GPA ClassesMissed  
## <dbl> <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 0 4 Neither 0 0 3.6 0  
## 2 0 4 Neither 2 1 3.24 0  
## 3 0 4 Owl 0 0 2.97 12  
## 4 0 1 Lark 5 1 3.76 0  
## 5 0 4 Owl 0 0 3.2 4  
## 6 1 4 Neither 0 0 3.5 0  
## # ... with 20 more variables: CognitionZscore <dbl>, PoorSleepQuality <dbl>,  
## # DepressionScore <dbl>, AnxietyScore <dbl>, StressScore <dbl>,  
## # DepressionStatus <chr>, AnxietyStatus <chr>, Stress <chr>, DASScore <dbl>,  
## # Happiness <dbl>, AlcoholUse <chr>, Drinks <dbl>, WeekdayBed <dbl>,  
## # WeekdayRise <dbl>, WeekdaySleep <dbl>, WeekendBed <dbl>, WeekendRise <dbl>,  
## # WeekendSleep <dbl>, AverageSleep <dbl>, AllNighter <dbl>

library(leaps)  
  
source("https://raw.githubusercontent.com/JA-McLean/STOR455/master/scripts/ShowSubsets.R")

Find the best multiple regression model predicting GPA using the following four method selections: all subsets, forward regression, backward regression, and stepwise regression.

1. All Subsets Method

all = regsubsets(GPA~., data=SleepData)

## 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)

## Gender ClassYear LarkOwlNeither LarkOwlOwl NumEarlyClass EarlyClass  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 )   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \* \*   
## 8 ( 1 ) \* \*   
## 9 ( 1 ) \* \*   
## ClassesMissed CognitionZscore PoorSleepQuality DepressionScore  
## 1 ( 1 )   
## 2 ( 1 ) \*   
## 3 ( 1 ) \*   
## 4 ( 1 ) \*   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \* \*  
## AnxietyScore StressScore DepressionStatusnormal DepressionStatussevere  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 ) \*   
## 4 ( 1 ) \*   
## 5 ( 1 ) \*   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \*   
## 9 ( 1 ) \*   
## AnxietyStatusnormal AnxietyStatussevere Stressnormal DASScore  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 ) \*  
## 5 ( 1 ) \*  
## 6 ( 1 ) \*  
## 7 ( 1 ) \*  
## 8 ( 1 ) \*  
## 9 ( 1 ) \*  
## Happiness AlcoholUseHeavy AlcoholUseLight AlcoholUseModerate Drinks  
## 1 ( 1 ) \*  
## 2 ( 1 ) \*  
## 3 ( 1 ) \*  
## 4 ( 1 ) \*  
## 5 ( 1 ) \*  
## 6 ( 1 ) \*  
## 7 ( 1 ) \*  
## 8 ( 1 ) \*  
## 9 ( 1 ) \*  
## WeekdayBed WeekdayRise WeekdaySleep WeekendBed WeekendRise  
## 1 ( 1 )   
## 2 ( 1 )   
## 3 ( 1 )   
## 4 ( 1 )   
## 5 ( 1 )   
## 6 ( 1 ) \*   
## 7 ( 1 ) \*   
## 8 ( 1 ) \* \*   
## 9 ( 1 ) \* \*  
## WeekendSleep AverageSleep AllNighter Rsq adjRsq Cp  
## 1 ( 1 ) 7.25 6.88 36.44  
## 2 ( 1 ) 12.43 11.73 22.49  
## 3 ( 1 ) 16.16 15.15 13.03  
## 4 ( 1 ) 19.71 18.41 4.11  
## 5 ( 1 ) 23.07 21.52 -4.25  
## 6 ( 1 ) 23.92 22.06 -4.85  
## 7 ( 1 ) 24.41 22.25 -4.37  
## 8 ( 1 ) 24.78 22.31 -3.50  
## 9 ( 1 ) 25.18 22.41 -2.73

ShowSubsetsModel = lm(GPA~Drinks+StressScore, data=SleepData)  
summary(ShowSubsetsModel)

##   
## Call:  
## lm(formula = GPA ~ Drinks + StressScore, data = SleepData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29843 -0.24307 0.03535 0.23565 0.89868   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.298430 0.052105 63.30 < 2e-16 \*\*\*  
## Drinks -0.024851 0.005946 -4.18 4.04e-05 \*\*\*  
## StressScore 0.008848 0.003062 2.89 0.00419 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3845 on 250 degrees of freedom  
## Multiple R-squared: 0.1025, Adjusted R-squared: 0.09533   
## F-statistic: 14.28 on 2 and 250 DF, p-value: 1.346e-06

1. Forwards Selection Method

full = lm(GPA~., data=SleepData)  
none = lm(GPA~1, data=SleepData)  
MSE = (summary(full)$sigma)^2  
  
step(none, scope=list(upper=full), scale=MSE, direction="forward", trace=FALSE)

##   
## Call:  
## lm(formula = GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
## ClassYear + WeekdayRise, data = SleepData)  
##   
## Coefficients:  
## (Intercept) Drinks CognitionZscore StressScore   
## 3.79938 -0.01700 0.13410 0.03478   
## DASScore ClassYear WeekdayRise   
## -0.01288 -0.07212 -0.04114

ForwardModel = lm(GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
 ClassYear + WeekdayRise, data = SleepData)  
summary(ForwardModel)

##   
## Call:  
## lm(formula = GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
## ClassYear + WeekdayRise, data = SleepData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.21505 -0.19065 0.01833 0.24037 0.84845   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.799377 0.223741 16.981 < 2e-16 \*\*\*  
## Drinks -0.017005 0.005692 -2.987 0.003098 \*\*   
## CognitionZscore 0.134102 0.032478 4.129 4.99e-05 \*\*\*  
## StressScore 0.034780 0.007691 4.522 9.53e-06 \*\*\*  
## DASScore -0.012883 0.003648 -3.531 0.000494 \*\*\*  
## ClassYear -0.072116 0.021940 -3.287 0.001160 \*\*   
## WeekdayRise -0.041143 0.024927 -1.651 0.100113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3569 on 246 degrees of freedom  
## Multiple R-squared: 0.2392, Adjusted R-squared: 0.2206   
## F-statistic: 12.89 on 6 and 246 DF, p-value: 1.17e-12

1. Backwards Regression Method

step(full, scale=MSE, trace=FALSE)

##   
## Call:  
## lm(formula = GPA ~ Gender + ClassYear + CognitionZscore + DepressionScore +   
## AnxietyScore + StressScore + Drinks + WeekdaySleep, data = SleepData)  
##   
## Coefficients:  
## (Intercept) Gender ClassYear CognitionZscore   
## 3.712435 -0.086172 -0.074506 0.128105   
## DepressionScore AnxietyScore StressScore Drinks   
## -0.009911 -0.018460 0.022651 -0.014895   
## WeekdaySleep   
## -0.029240

BackwardsModel = lm(GPA ~ Gender + ClassYear + CognitionZscore + DepressionScore +   
 AnxietyScore + StressScore + Drinks + WeekdaySleep, data = SleepData)  
summary(BackwardsModel)

##   
## Call:  
## lm(formula = GPA ~ Gender + ClassYear + CognitionZscore + DepressionScore +   
## AnxietyScore + StressScore + Drinks + WeekdaySleep, data = SleepData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.20958 -0.18108 0.02274 0.23235 0.79364   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.712435 0.175763 21.122 < 2e-16 \*\*\*  
## Gender -0.086172 0.052490 -1.642 0.101944   
## ClassYear -0.074506 0.022046 -3.380 0.000845 \*\*\*  
## CognitionZscore 0.128105 0.032871 3.897 0.000126 \*\*\*  
## DepressionScore -0.009911 0.004927 -2.012 0.045354 \*   
## AnxietyScore -0.018460 0.006133 -3.010 0.002890 \*\*   
## StressScore 0.022651 0.004556 4.972 1.25e-06 \*\*\*  
## Drinks -0.014895 0.006041 -2.466 0.014364 \*   
## WeekdaySleep -0.029240 0.019654 -1.488 0.138103   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3568 on 244 degrees of freedom  
## Multiple R-squared: 0.2457, Adjusted R-squared: 0.221   
## F-statistic: 9.937 on 8 and 244 DF, p-value: 5.801e-12

1. Stepwise Regression Method

step(none, scope=list(upper=full), scale=MSE, trace=FALSE)

##   
## Call:  
## lm(formula = GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
## ClassYear + WeekdayRise, data = SleepData)  
##   
## Coefficients:  
## (Intercept) Drinks CognitionZscore StressScore   
## 3.79938 -0.01700 0.13410 0.03478   
## DASScore ClassYear WeekdayRise   
## -0.01288 -0.07212 -0.04114

StepwiseModel = lm(GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
 ClassYear + WeekdayRise, data = SleepData)  
summary(StepwiseModel)

##   
## Call:  
## lm(formula = GPA ~ Drinks + CognitionZscore + StressScore + DASScore +   
## ClassYear + WeekdayRise, data = SleepData)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.21505 -0.19065 0.01833 0.24037 0.84845   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.799377 0.223741 16.981 < 2e-16 \*\*\*  
## Drinks -0.017005 0.005692 -2.987 0.003098 \*\*   
## CognitionZscore 0.134102 0.032478 4.129 4.99e-05 \*\*\*  
## StressScore 0.034780 0.007691 4.522 9.53e-06 \*\*\*  
## DASScore -0.012883 0.003648 -3.531 0.000494 \*\*\*  
## ClassYear -0.072116 0.021940 -3.287 0.001160 \*\*   
## WeekdayRise -0.041143 0.024927 -1.651 0.100113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3569 on 246 degrees of freedom  
## Multiple R-squared: 0.2392, Adjusted R-squared: 0.2206   
## F-statistic: 12.89 on 6 and 246 DF, p-value: 1.17e-12