R Notebook

Group Members: Katherine Bacon Kay Youngstrom Ali Floyd Bethany Newcomb

library(readr)  
sleep <- read.csv("https://raw.githubusercontent.com/JA-McLean/STOR455/master/data/SleepStudy.csv")

1. Interaction Terms

Create a model that predicts GPA based on Average Sleep.

mod1 = lm(GPA~ AverageSleep, data = sleep)  
summary(mod1)

##   
## Call:  
## lm(formula = GPA ~ AverageSleep, data = sleep)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.29018 -0.24395 0.03294 0.26317 0.80407   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.44618 0.21183 16.269 <2e-16 \*\*\*  
## AverageSleep -0.02541 0.02640 -0.962 0.337   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4043 on 251 degrees of freedom  
## Multiple R-squared: 0.003676, Adjusted R-squared: -0.000293   
## F-statistic: 0.9262 on 1 and 251 DF, p-value: 0.3368

Now create a model precting GPA based on Average Sleep, Classes Missed, and an interaction between these two variables.

mod2 = lm(GPA~ AverageSleep + ClassesMissed + AverageSleep\*ClassesMissed, data = sleep)  
summary(mod2)

##   
## Call:  
## lm(formula = GPA ~ AverageSleep + ClassesMissed + AverageSleep \*   
## ClassesMissed, data = sleep)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.33450 -0.23587 0.04836 0.23665 0.77052   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.5748832 0.2608322 13.706 <2e-16 \*\*\*  
## AverageSleep -0.0350662 0.0323790 -1.083 0.280   
## ClassesMissed -0.0310116 0.0454508 -0.682 0.496   
## AverageSleep:ClassesMissed 0.0009665 0.0057040 0.169 0.866   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3987 on 249 degrees of freedom  
## Multiple R-squared: 0.0388, Adjusted R-squared: 0.02722   
## F-statistic: 3.35 on 3 and 249 DF, p-value: 0.01968

Does the ClassesMissed term improve our model? Use Anova or another test we have learned in class. No need for a formal hypothesis.

anova(mod1, mod2)

## Analysis of Variance Table  
##   
## Model 1: GPA ~ AverageSleep  
## Model 2: GPA ~ AverageSleep + ClassesMissed + AverageSleep \* ClassesMissed  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 251 41.037   
## 2 249 39.590 2 1.4466 4.5492 0.01147 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The P-value is smaller than .05 which shows that it is significant. This means that at least one of the coefficients of ClassesMissed is non-zero and therefore makes our model better.