Group\_Review

library(readr)  
sleep = 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(sleep)

## # 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>

#Topic 4: ANOVA for simple linear regression  
#Question:Test the strength of the linear relationship between \_\_GPA\_\_ and \_\_AverageSleep\_\_ using ANOVA for regression. Include hypotheses and your conclusions in the context of the problem.

#Solution:  
mod1 = lm(GPA~AverageSleep, data = sleep)  
anova(mod1)

## Analysis of Variance Table  
##   
## Response: GPA  
## Df Sum Sq Mean Sq F value Pr(>F)  
## AverageSleep 1 0.151 0.15142 0.9262 0.3368  
## Residuals 251 41.037 0.16349

#Null Hypothesis: The coefficient for AverageSleep is 0.  
#Alternative Hypothesis: The coefficient for AverageSleep is not 0.  
#Conclusion: According to the table, p-value is 0.3368, which is a really large number. Therefore, we fail to reject the null hypothesis and come to the conclusion that the coefficient for AverageSleep is 0.

#Topic 17: Test of significance for nested models

#Question\_1:Construct a logistic model to predict if the student has an early class using Stress, AnxietyStatus, and the interaction between them as the predictor variables.

#Solution\_1:  
mod2 = glm(EarlyClass~Stress+AnxietyStatus+Stress\*AnxietyStatus,family = binomial, data = sleep)  
summary(mod2)

##   
## Call:  
## glm(formula = EarlyClass ~ Stress + AnxietyStatus + Stress \*   
## AnxietyStatus, family = binomial, data = sleep)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.6651 -1.4944 0.8906 0.8906 1.7941   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.51083 0.36515 1.399 0.1618   
## Stressnormal 0.58779 0.59628 0.986 0.3243   
## AnxietyStatusnormal -0.04082 0.67700 -0.060 0.9519   
## AnxietyStatussevere 0.47000 0.76920 0.611 0.5412   
## Stressnormal:AnxietyStatusnormal -0.33774 0.84118 -0.402 0.6881   
## Stressnormal:AnxietyStatussevere -2.95491 1.43662 -2.057 0.0397 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 322.99 on 252 degrees of freedom  
## Residual deviance: 317.01 on 247 degrees of freedom  
## AIC: 329.01  
##   
## Number of Fisher Scoring iterations: 4

#Question2:Conduct a drop in deviance hypothesis test to determine the effectiveness of the interaction terms in the model constructed in the previous question. Cite your hypotheses, p-value, and conclusion in context.

#Solution2:  
reduced = glm(EarlyClass~Stress+AnxietyStatus,family = binomial, data = sleep)  
anova(mod2, reduced,test = "Chisq")

## Analysis of Deviance Table  
##   
## Model 1: EarlyClass ~ Stress + AnxietyStatus + Stress \* AnxietyStatus  
## Model 2: EarlyClass ~ Stress + AnxietyStatus  
## Resid. Df Resid. Dev Df Deviance Pr(>Chi)   
## 1 247 317.01   
## 2 249 322.16 -2 -5.1531 0.07604 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#Null Hypothesis:The coefficient for the interaction terms is 0.  
#Alternative Hypothesis:The coefficient for the interaction terms is not 0.  
#Conclusion: According to the table, p-value is 0.07604, which is greater than 0.05. Therefore, we fail to reject the null hypothesis and come to the conclusion that the coefficient for interaction terms is 0.