Overall Health of Students

By: Abby Celander, Kelvin Monje and Shane Hauck

Researchers conducted a longitudinal study in 2015 to understand the potential effects of pushing back the start time for high school students. This study was conducted on high school students in New York during a 12-month interval, and the researchers observed 40 different variables that could be used to understand how this time change would affect students. We used this data to understand how mental and physical health is affected by using different predictors to predict a mental health score for a student as well as seeing if a student is physically healthy.

Davoren and Hwang found that student athletes who are highly motivated by athletics, tend to have episodes of anxiety and depression based on athletic performance. In a different study by DiClementi, Reese, and Borsa, they found that athletes during their offseason may be more stressed, depressed, and anxious than in season, as they find themselves feeling more isolated and essentially having too much free time. We decided to include a binary variable as a predictor which says whether a student is currently active in an athletic season or not.

Wheaton, Jones, Cooper and Croft found that with short sleep duration students can greatly affect their mental health and performance in school. This idea led us to want to investigate what time students go to bed and wake up and the affect it had on mental health scores.

We are going to be using DASS_sum as our response variable. The variable is the sum of Depression, Anxiety and Stress where higher scores equals more problems. Therefore, lower DASS_sum scores mean better mental health.

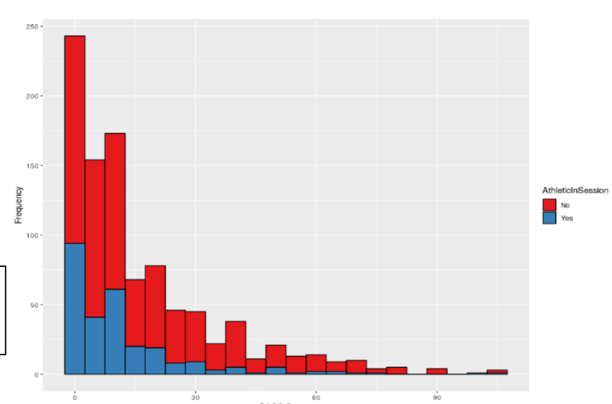

Figure 1: Looks at the frequency of individual observations of average DASS Sum.

We started our exploration by looking at the frequency of mental health scores to understand, at a broad scope, what mental health scores students are getting, categorized by whether it is an athlete in season. As seen in Figure 1, we have many zeros for mental health score.

Table 1: Athlete in season and average mental health score

In Athletic Season	Total	Mean DASS_sum
No	663	18.3
Yes	270	12.1

Table 1 shows that while students are not in season, their mental health score is larger than while students are in season. That being said, there are more observations of students who are not in season compared to students who are.

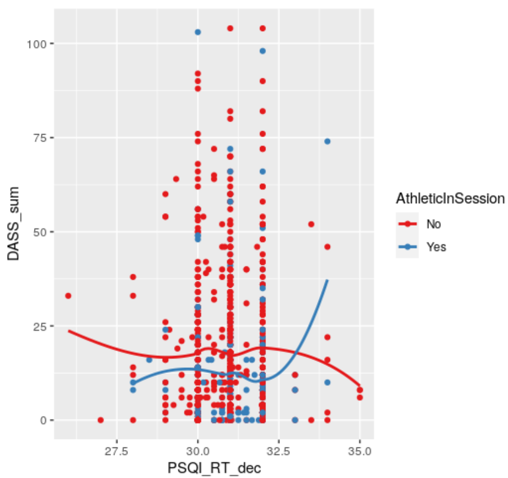

Figure 3: Rise time versus mental health score categorized by athlete in season.

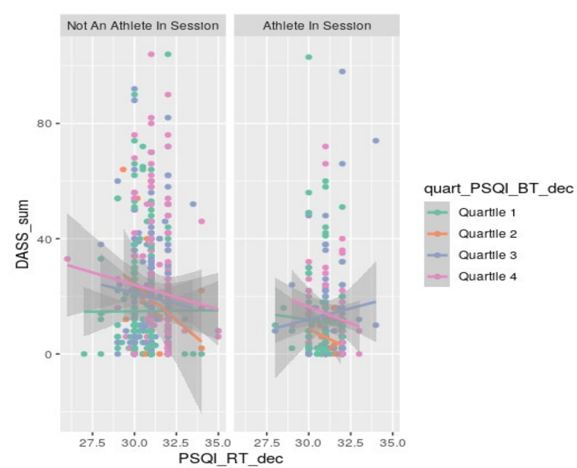

Figure 5: Rise time versus mental health score categorized by athlete in season and an interaction with bedtime.

Figures 3 shows our rise time for students versus their mental health score. A lot of these students have similar rise times which makes sense because they are all going to school at the same time each morning. We found that the plot for bedtime was very similar to rise time. Therefore, we created an interaction between them as seen in Figure 5. There appears to be a negative trend whereas bedtimes and risetimes get later, mental health gets worse.

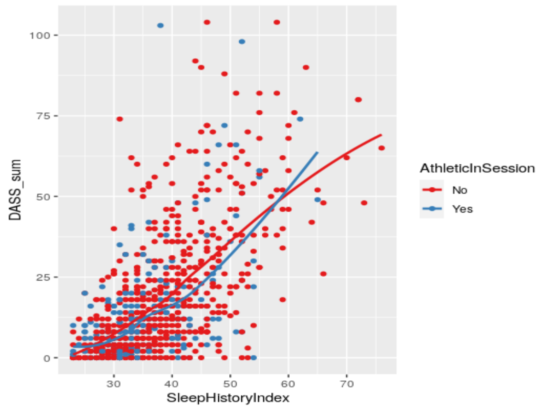

Figure 6: Sleep History Index versus mental health score characterized by athlete in season

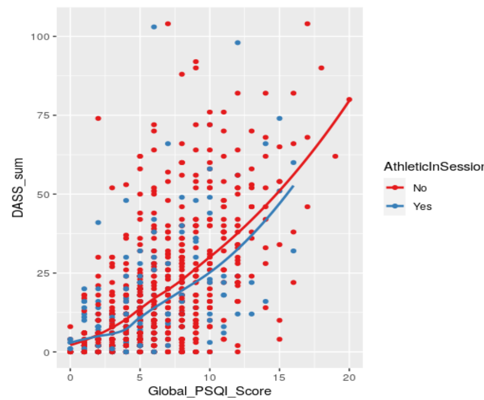

Figure 7: Global sleep quality versus mental health score.

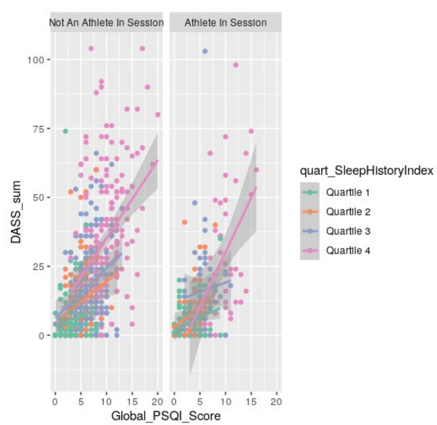

Figure 8: Global sleep quality versus mental health score, characterized by athlete in season and sleep history index.

Sleep History Index refers to a score of daytime sleepiness where a higher score means that there are more issues. Global sleep quality refers to a score of several components of quality sleep where a higher score means more problems. Both Figure 6 and Figure 7 showed similar trends, therefore, we decided to use Sleep History Index and Global sleep quality as an interaction. As shown in Figure 8, Sleep History Index score is divided into quartiles, quartile 1 being the lowest scores and quartile 4 being the highest scores. When plotting these against global sleep quality scores and mental health scores we see a positive trend where lower sleep history index score and global sleep quality scores have better mental health and vice versa.

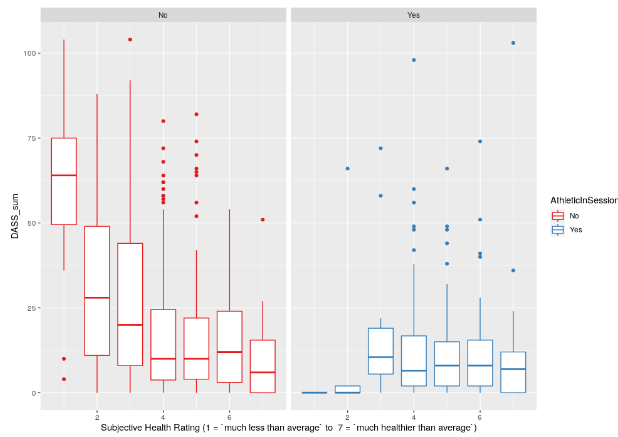

Figure 9: Subjective health rating versus mental health score and categorized by athlete in season.

When testing athlete in season, it was not a significant predictor but when creating an interaction term with Health as a factor it became significant. Figure 9 shows us the interaction between health and athlete in season and a big difference in the subjective health rating versus mental health score for athletes in season and athletes not in season. Students that are healthy and in season have relatively low mental health scores and students that are not healthy and in season have much higher scores. This clear interaction made us want to investigate the health variable as a response variable.

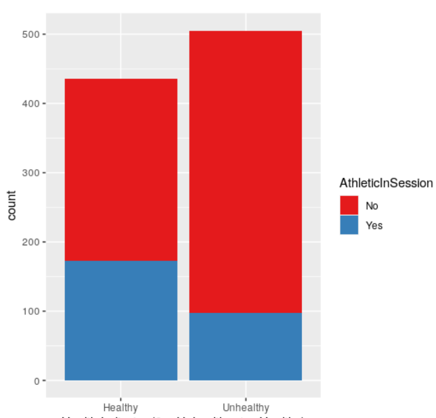

Figure 10: Health indicator versus count categorized by athlete in season.

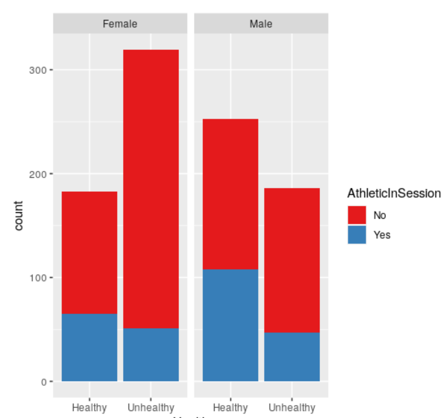

Figure 11: Health indicator versus count categorized by gender and athlete in season.

Due to the Health variable being a factor, we decided that to investigate Health as response it needed to be used as a binary predictor. We split scores up by if it had a score of 4 or less it was considered “Unhealthy”, and if it had a score of 5 or more it was considered “Healthy”. Figure 10 tells us that there is around 500 count of students that are unhealthy, and most of the unhealthy students are not in athletic season. There is a smaller total count for healthy students, but athletes not in season still has the higher count for healthy students. Figure 11 compares the healthy and unhealthy students by both whether they are in season and gender. There is a much higher count for unhealthy females than males, but we had more responses from females so this could be somewhat misleading.

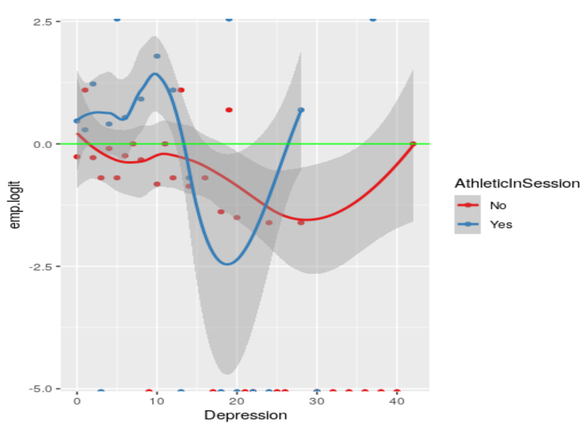

Figure 12: Depression versus empirical logit categorized by athlete in season

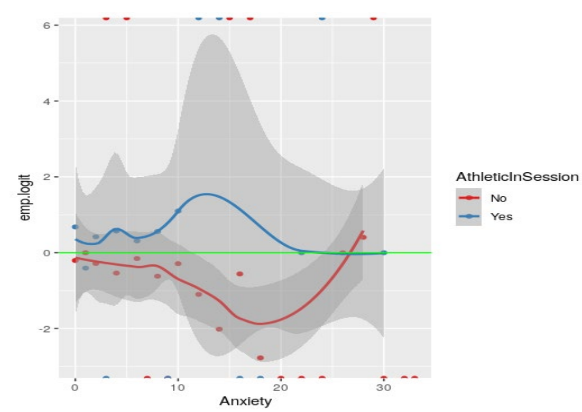

Figure 13: Anxiety versus empirical logit categorized by athlete in season

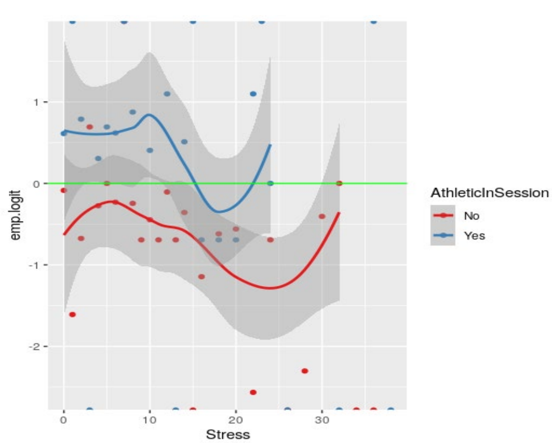

Figure 14: Stress versus empirical logit categorized by athlete in season

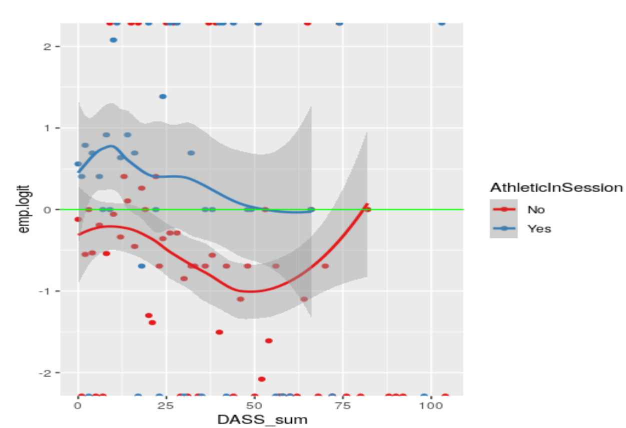

Figure 15: Mental health score (sum of depression, anxiety, and stress) versus empirical logit categorized by athlete in season

Figures 12, 13 and 14 all show similar trends. Students that are in season tend to have lower depression, anxiety, and stress scores respectively, therefore, they have a greater than 50% probability of being physically healthy. Meanwhile, students not in season tend to have higher depression, anxiety, and stress score respectively which means that they are more likely to be physically unhealthy. Although similar, there does seem to be a difference in Figure 14 regarding stress. There appears to be an increase of students who are in season that have higher stress scores causing some students to have a greater probability of being unhealthy. Figure 15 shows the total sum of the variables, and we again see similar results. Athletes in season have above a 50% probability of being healthy, and athletes not in season have below a 50% probability of being healthy. Again, this is due to students in season having lower mental health scores while students who are not in season tend to have higher scores.

These models show that there is a strong correlation between a student's mental health and their physical health. When either mental or physical health is unhealthy, the other is likely to also be unhealthy. On the other hand, students with good mental or physical health, the other aspect of their overall health seems to be in good shape. This positive relationship helps explain that students who are in athletic season, tend to have a higher chance at being healthy.

Our statistical methods we used to address our research questions, was to first decide what predictors to consider helping us predict our response variables. We looked at different studies to understand how mental health scores and if student is considered healthy are affected. We then created an unconditional means model then explored the data by looking at predictors that we thought would be useful and checked where we could put interactions. We used an exploratory data analysis to see what predictors significantly affect our response. This allowed us to understand what characteristics most affect our responses and build our final models.

There were a number of tricky decisions that had to be made throughout the process of building our models. The initial dataset had 1,372 observations, we then filtered that down into having 933 observations. Since our response variables were ultimately the most important things in our models, we filtered out any observation that was missing either score. Then when looking at the interaction between risetime and bedtime, the plot was being skewed by 3 major outliers. There was one point that has a risetime of greater than 40, which was very distant from all other points. Then there was two points which had a bedtime that was greater than a risetime. This makes no sense and it's likely something that could have been mistake when inputting the data. These points were having a large impact on the skewing of the interaction predictor.

The predictors that were used all provide different values to our model. To know whether the term is significant, the absolute value of the t-value has to be greater than 2. Most of the predictors in our model are significant. The output revealed that athlete in season, sleep history index, all the health factors and all the interactions with athlete in season and factors for Health are significant. Global sleep quality score by itself was very insignificant but when it was used as interaction with sleep history index, the interaction was significant. Bedtimes and rise times individually were both insignificant. We then tried their interaction together. Surprisingly, the interaction was also insignificant. Therefore, we decided to remove bedtimes and risetimes from our model. Since we found so much significance throughout this model, we believe that our model is a good predictor of mental health scores.

With the model for Health, due to the model having a binary response variable, we had to considered significance a little bit differently. Instead of using t-values that have absolute values greater than 2, we are now using p-values that are less than 0.05 to consider significance. All predictors, athlete in season; gender; depression; anxiety; stress; and mental health score are significant. Due to all predictors being significant, we have confidence in our model that it can predict whether a student is healthy or not.

We believe that a student's overall health as a human being is ultimately divided into 2 very important categories: mental health and physical health. These two variables go hand in hand, where when one is in a poor condition, so is the other. That being said, when it is in great condition, so is the other. Athletes being in season are more likely to be healthy whereas students who are not in season are more likely to be unhealthy. This relationship is shown to be a strong predictor of both mental and physical health. In contrary to our previous belief, bedtimes and risetimes were shown to not have a strong impact on predicting one's mental health. We also found there to be a strong interaction between Sleep History Index and Global sleep quality scores as those both positively predicted mental health scores. When using Health with whether a student is in season as an interaction, it was an extremely strong predictor. Our curiosity guided us into wanting to investigate Health as a response variable. During our investigation, we found that mental health has a very strong effect on predicting one's physical health. We feel that we have successfully created models that positively predict students mental and physical health respectively.

Although we do feel we were successful in building these models for this dataset, I do question the effectiveness of this data. The purpose of this data was to be a longitudinal study, but we found that many of the observations did not take all 3 tests. We believe, it would have been better if we were able to compare scores for the same students in and out of athletic season. We think that our study can potentially open up avenues for other future research in regard to this topic of the overall health of students.

The multilevel models were fit using the lme4 package (Bates et al., 2015) in R version 3.1 (RCore Team 2014). Graphs were made using the Tidyverse (Wichkam et al., 2019). Tables were made using the knitr package (Xie, 2021). Hmisc package was used to divide dataset into quartiles (Harell Jr, 2021).

References

Cooper, Adina., et al. "Short Sleep Duration among Middle School and High School Students - United States, 2015." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 25 Jan. 2018, https://www.cdc.gov/mmwr/volumes/67/wr/mm6703a1.htm?s_cid=mm6703a1_e.

Davoren, Ann and Hwang, Seunghyun. *NCAA.org - the Official Site of the NCAA*, https://www.ncaa.org/sport-science-institute/mind-body-and-sport-depression-and-anxiety-prevalence-student-athletes.

DiClementi, Jeannie D., et al. "Stress of the off-Season May Be Greater for the Student-Athlete than during Competition: A Pilot Study." *Journal of Behavioral and Brain Science*, Scientific Research Publishing, 12 July 2017, https://www.scirp.org/journal/paperinformation.aspx?paperid=77611.

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effect s Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067. i01.

Frank E Harrell Jr, with contributions from Charles Dupont and many others. (2021). Hmisc: Harrell Miscellaneous. R package version 4.5-0. https://CRAN.R-project.org/package=H misc.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686 , https://doi.org/10.21105/joss.01686

Yihui Xie (2021). knitr: A General-Purpose Package for Dynamic Report Generation in R. R pac kage version 1.33.