

Alex Hamedaninia

DSC530 - Final

Does study time affect exam performance?

We began this study using a data set from Kaggle that gave us 5 different variables: study, study_other, exam performance, other exam performance, and knowledge level. These variables tell us the degree of time spent studying for an exam, degree of time spent studying for other classes, exam performance for this class, exam performance in another class, and knowledge level for this class ranging from very low, low, medium and high.

We have learned quite a lot about this data set, and it brings us back to our original question -- how much of an effect does study time have on performance?

We saw that the correlation between just studying and exam performance was decently low, indicating not a strong relationship between these two variables. However, from our calculation of the correlation, we saw that there was a positive correlation, indicating there is a positive relationship between time spent studying and exam performance.

However, once we built our regression model and knowledge level was introduced, the variation jumped to 85% variability in the last regression model, from our R-squared statistic. This indicates that if we used this as a prediction model, we may be pretty accurate in our predictions as to how well (or how bad) a student will perform on the exam. In addition to our comparison, we saw on our comparison chart that the students who spent the most time studying had the highest knowledge level, though there were a

few who had high knowledge who didn't study much at all, but we can't all be as talented as they are.

This goes to say that studying more may lead to increasing knowledge level, and this could increase exam performance. More investigations could go further to refine the model to more accurately predict a student's exam performance based on other characteristics.

I do feel like afterwards looking through my analysis, I should have checked the relationship between knowledge level and exam performance, as we saw how much that affected the variability in the regression model. I think if I had checked the relationship either using the correlation calculator test, or checking graphically, there may have been a significant relationship there. I'm not sure if there were any assumptions I made in my analysis that were incorrect, but I was not fully sure in how I analyzed some of my results.

Some challenges I faced was finding a dataset to work with. Most of the datasets were in the sciences in fields I had no understanding of, so it would not make sense for me to analyze those. I found that every dataset I was interested in was qualitative/categorical in some kind of way. That was definitely the biggest challenge of this project.