You Snooze You Lose: A Correlational Study of GPA and Number of Absences

jonathanferrari.com/projects/snooze

by Jonathan Ferrari, Adrienna Cao, and Ryan Cao

Awarded 3rd Place in Contra Costa County Science and Engineering Fair

Abstract

This research paper aims to address the issue of chronic absenteeism in the American public school system. To do so, we perform elementary data analysis on real-world data from one California high school. In our research, we found that there was a negative correlation between GPA and number of classes missed per school year. We represent the graphs visually in both smoothed and unsmoothed format. All data has been anonymized for the purpose of privacy protection of students whose information is represented in this report.

Media

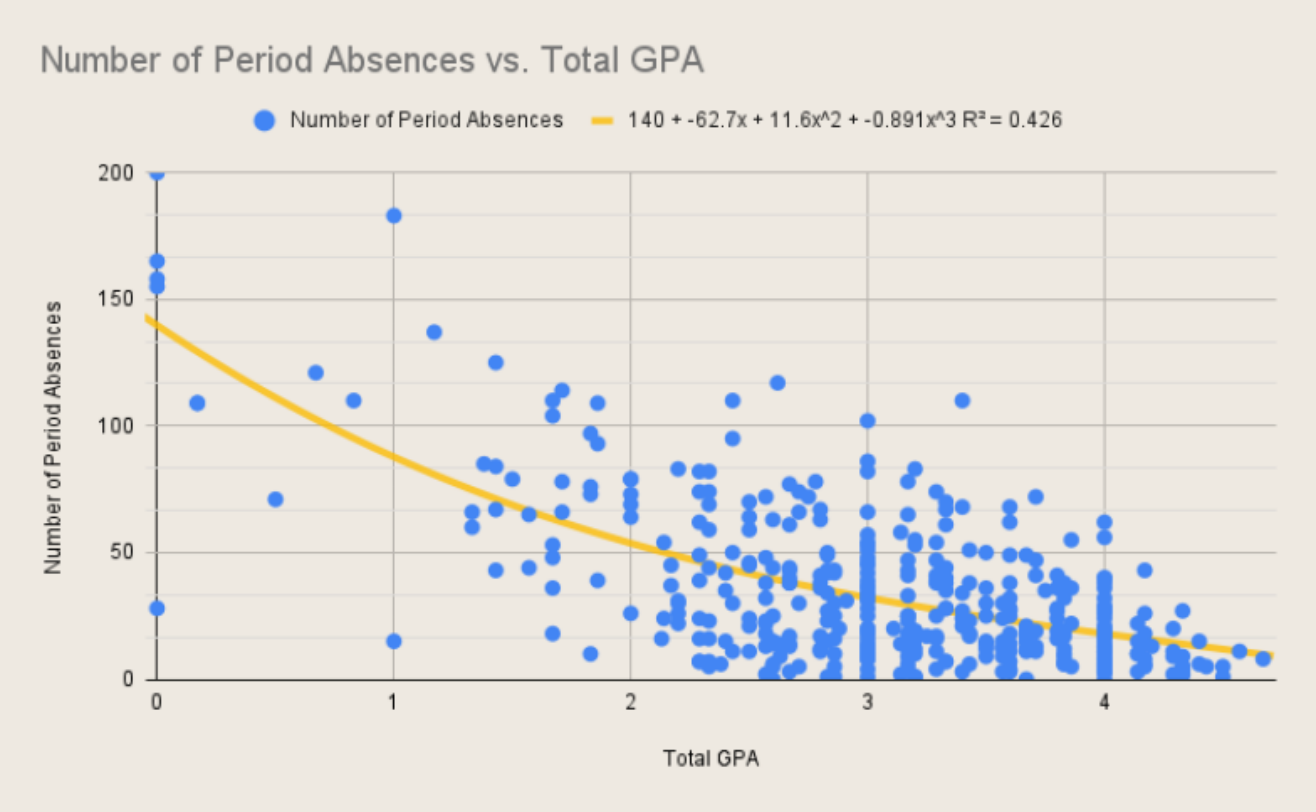

Figure 1

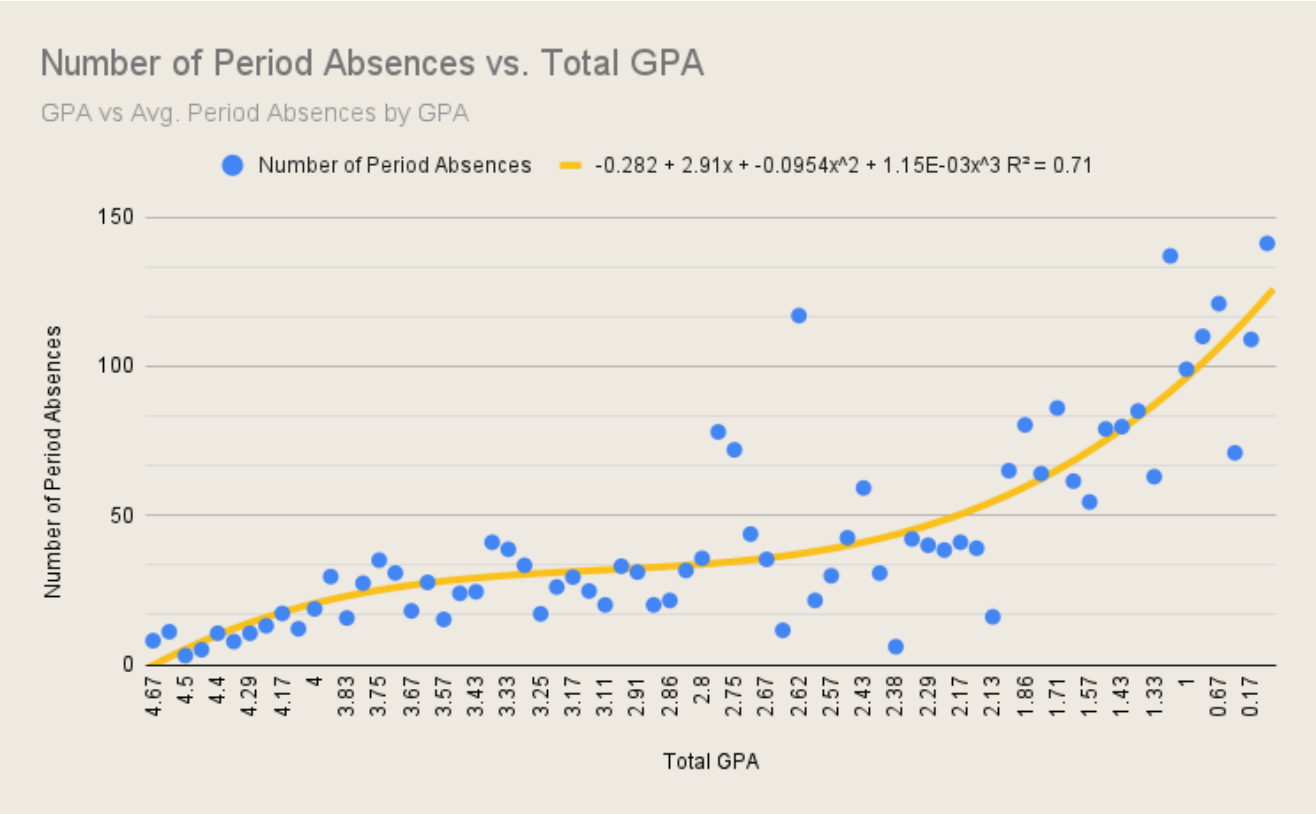

Figure 2

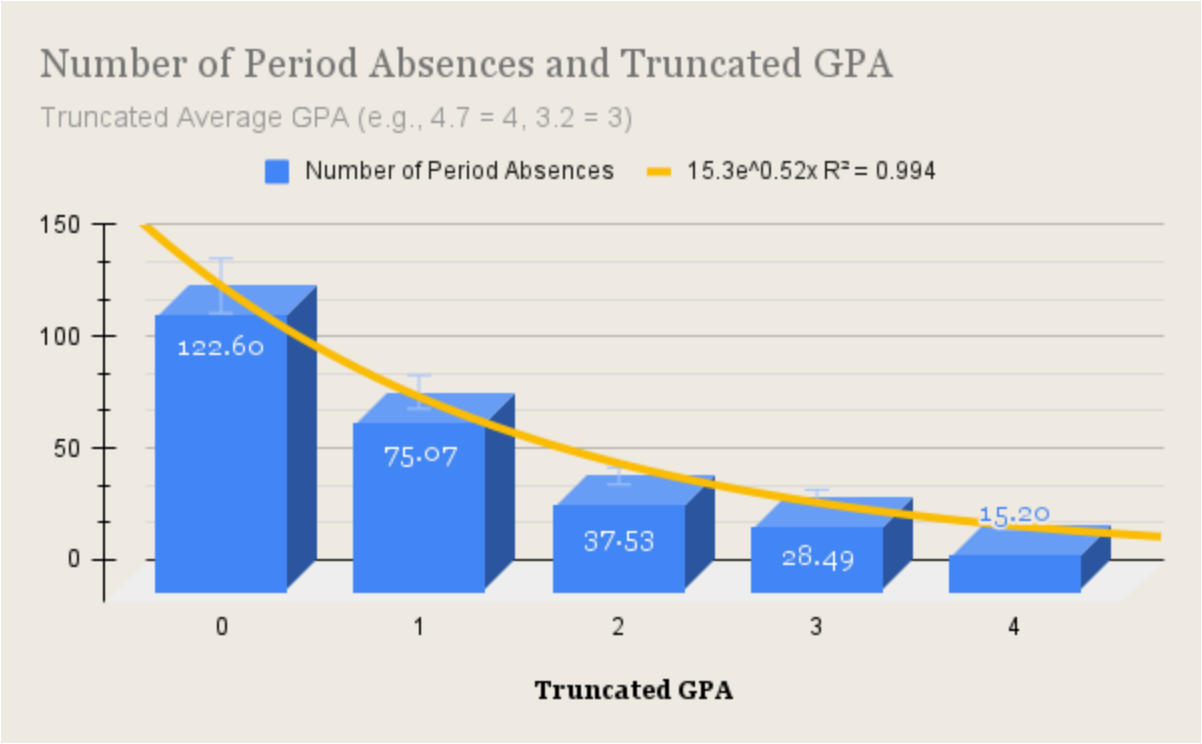

Figure 3

Hypothesis

As the number of period absences decreases, so will the GPA of a student.

Purpose

The purpose of our correlational study is to show how much of a problem absenteeism is, not just in the public school system, but in our society as a whole. This has been a problem evident in past generations of students, but today, it is a greater problem than ever. This can be an especially important issue in the Senior class. Senior students across the nation have long been diagnosed with "Senioritis", or the lack of effort in their last year, and especially last semester of high school. This can be very harmful to the students' educational experience, as many of them head off to various colleges, and they can be wildly unprepared for the coming task. They have missed out on much of one of the most important years of learning, and this leaves a large gap for the students to cover in college. This can discourage many people from further pursuing higher education; and although it may be too difficult, it may also be too difficult for them due to their lack of instruction.

When a student misses school, not only do they miss educational information, but they miss out on the learning environment that the school provides. Sure, a student could do their work from home, however, they will likely get distracted more and will be less likely to remember the information once they are in school, for example during a test. This is due to the psychological concept of context effects, which states that you are more likely to remember something in the place that you studied it. In our project, we plan to further explore this as we delve into the correlation between GPA and the number of absences. As mentioned previously, one hypothesis on why they could relate is that GPA drops because absences rise. However, another explanation for this could be that a student becomes discouraged and disillusioned with their academic career, and thus lose interest in school, if their grades drop for some other reason. This could include: doing poorly on an exam, forgetting about homework, etc. This would then cause them to be absent more. Due to the aforementioned hypothesis, these two causes could work hand in hand to create a vicious cycle, which harms our student population. Thus the logical thing to do would be to try and stop this before it happens.

With our study, we plan to find the relationship between the two variables, as well as come up with a mathematical model to predict one variable given the input of one variable. Furthermore, we will show how average absences change over a large GPA group, and a much smaller group as well. We hope to be able to show that with a large sample size, in an average socioeconomic population, that these results are accurate, and can be generalized, at least to other populations at similar socioeconomic levels. Additionally, we hope to show how big of a problem absenteeism is, and how much it can affect students' grades and their overall experience in the American education system.

Results

The results of our study were nearly exactly what we had hypothesized. We found a significant negative correlation between GPA and Number of Period Absences. This means that as GPA increased, the number of absences decreased. We modeled this relationship with a polynomial equation, and we got an R^2 value of .426 (Figure 1). This means that there was a noticeable relationship between the variables, but not an extremely strong one. Next, we graphed the average number of absences of each student with a specific GPA. For this scatter plot, we modeled the data with a three-degree-polynomial (i.e. a cubic function), in which the correlation was a bit stronger with an R^2 value of .71; here we saw an even more direct relationship (Figure 2). Finally, we used a bar graph to show the average number of absences per full GPA point, i.e. 1-2, 2-3, etc. For this relationship, we modeled the data with an exponential function. For this relationship, we got an R^2 value of .994 (Figure 3). These correlations may not seem especially significant, however, when dealing with humans, and in the social sciences, data often does not conform to a formula. So while these values may seem low for a study of physics, or chemistry, or biology, in social sciences, these are rather significant. However, we must be careful to avoid any incorrect conclusions. As the name suggests, this study determines correlation, *not causation*, between two variables. While we see significant correlations, there could be an uninvestigated confounding variable causing this relationship.

Background

Correlational studies are a type of study that is used in psychological observational studies or experiments. It is used as the first method of gathering information from a random sample size. Correlational studies are used to find out if there is a relationship between two or more variables. While you may find out if there is a relationship between two or more variables, it does not imply that there is a cause-and-effect relationship between the variables. The advantage of archival research is that the researcher can't introduce changes in the participants' behavior. The larger the sample size the better you can observe trends, relationships, and outcomes. Archival studies tend to be less expensive as well because the sample sizes can be accessed from free archives or record databases. The kind of correlational study that we're doing is archival research where we are looking at the records of GPA and absences of the senior class of 2020. 1 in 10 students were chronically absent in the state of California from 2017-2018, which would mean students would miss 10% of the school year, per records released by the California Department of Education. 1 in 6 students also missed 15 or more days of school which is 8.3% of the school year. The numbers of absences are on an upward trend. Gathered in the 2018-2019 school year, the department of education in California released records that pointed at a 12.1 percent chronic absenteeism rate state wide. This is one percent higher than the previous 11.1 percent in the previous 2017-2018 school year. Nearly 703,000 students have been recorded to have chronic absenteeism in the 2017-2018 school year and the number has been rising every year.

Procedure

The procedures of our correlational study include the retrieval of the GPA and attendance of the class of 2020 at Deer Valley High School and the subsequent calculations and interpretation of that data. After procuring the data, we performed a mathematical analysis of it using computer software. In this analysis we were able to generate multiple r correlation values along with their matching scatter plots — effectively creating mathematical models of # of period absences versus GPA, average # of period absences versus GPA, and average # of absences in GPA brackets.

References

Washburn, David. "Chronic Absenteeism in California Schools up Slightly, New Data Show." *EdSource*, EdSource, 5 June 2019, edsource.org/2019/chronic-absenteeism-in-california-schools-up-slightly-new-data-show/607993.

Cherry, Kendra. "Why Correlational Studies Are Used in Psychological Research." *Verywell Mind*, Verywell Mind, 4 Feb. 2019, www.verywellmind.com/correlational-research-2795774.

"2018-19 Chronic Absenteeism Rate." *Chronic Absenteeism Rate – State (CA Dept of Education)*, 2019, https://dq.cde.ca.gov/dataquest/DQCensus/AttChrAbsRate.aspx?agglevel=State&cds=00&year=2018-19.

"Child Welfare & Attendance." *Child Welfare & Attendance – Attendance Improvement (CA Dept of Education)*, www.cde.ca.gov/ls/ai/cw/.

"Chronic Absenteeism Calculation." *Chronic Absenteeism Calculation – California School Dashboard and System of Support (CA Dept of Education)*, www.cde.ca.gov/ta/ac/cm/chronabscal.asp.

"Chronic Absenteeism in the Nation's Schools." *Chronic Absenteeism in the Nation's Schools*, www2.ed.gov/datastory/chronicabsenteeism.html.

"Chronic Absenteeism." *American Federation of Teachers*, 4 Aug. 2016, www.aft.org/childrens-health-safety-and-well-being/childrens-access-care/chronic-absenteeism.

"Chronic Absenteeism: Healthy Schools Campaign." *Healthy Schools Campaign | Helping Children Learn and Thrive*, healthyschoolscampaign.org/issues/chronic-absenteeism/.

"The Problem." *Attendance Works*, www.attendanceworks.org/chronic-absence/the-problem/.

Materials

Computers including but not limited to iMac, MacBook Pro, and MacBook Air

Computer Software, including Jupyter environments, Python 3, Visual Studio Code, G Suite, and MS office

GPA and Attendance Data from Deer Valley High School with all P.I.D. removed