

The Effects of Absences on Student Performance

Nathan R. Lehnhofer*
BYU Provo, UT

Student attendance is widely recognized as a critical factor influencing academic success. Absences can disrupt learning continuity, reduce exposure to instructional materials, and impair performance on assessments. In educational research, frequent absences have been linked to lower achievement, decreased engagement, and increased risk of academic difficulties. This report investigates whether patterns of student absences support this understanding and quantifies the relationship between absenteeism and academic performance.

Using a random sample of Kaggle's Students Performance Dataset, we compare test scores between students with fewer than five absences and those with five or more. A two-sample t-test is conducted to determine whether the difference in mean performance is statistically significant, followed by development of a confidence interval for the mean difference. The results allow us to evaluate commonly held beliefs about attendance and achievement.

I. Introduction

Regular classroom attendance plays an essential role in supporting academic success. Consistent presence allows students to receive instruction, participate in learning activities, and develop understanding through repeated engagement. Numerous studies indicate that absenteeism negatively affects learning outcomes and increases the likelihood of academic struggle.

However, students may not always recognize the long-term effects of missing class. Some may assume that occasional absences have little impact or that material can be made up independently. Although prior research suggests that absenteeism reduces performance, the magnitude of its effect can vary across students, schools, and subjects.

This project examines this question using an openly available dataset of student test scores. We analyze data from Kaggle's Student Performance Database to explore whether students with higher numbers of absences perform worse academically. By statistically comparing performance between students with fewer than five absences and those with five or more, we aim to determine whether this widely held belief is supported by evidence in a real-world sample.

The goal of this study is not only to evaluate a common assumption but also to quantify the performance differences associated with absenteeism. This analysis allows us to assess whether increased absence corresponds to meaningful differences in academic outcomes and whether the relationship is statistically significant under formal hypothesis testing.

II. Methods

This study investigates whether student absenteeism is associated with differences in academic performance. To test this hypothesis, we analyzed a subset of the Student Performance Dataset obtained from Kaggle. The dataset contains individual student records, including the number of absences and final grade point average (GPA). Two groups were constructed based on absence frequency: students with five or fewer absences and students with more than five. A simple random sample of 25 students was drawn from each group to form the two samples used in the analysis. Random sampling was performed using Python's pandas library to ensure that each student had an equal probability of selection.

Sample means and standard deviations of GPA for each group were computed to summarize central tendency and variability. To further visualize the distributional differences between groups, several graphical techniques were employed. Histograms were generated to show the empirical frequency distribution of GPA, and empirical cumulative distribution functions (CDFs) were constructed to compare the proportion of students achieving a given GPA or lower across groups. Additionally, probability density functions (PDFs) were estimated using kernel density estimation (KDE), a non-parametric technique that provides a smooth approximation of the underlying GPA distribution without assuming a theoretical distributional shape. These plots visually support the inferential analysis by illustrating shifts in distribution location and spread between absence groups.

To formally test the hypothesis that the two groups have equal mean GPA, a two-sample pooled-variance t-test was conducted. Let \bar{x}_1 and \bar{x}_2 denote the sample means of the low-absence and high-absence groups, respectively, with

*Mechanical Engineering

sample standard deviations s_1 and s_2 and sample sizes $n_1 = n_2 = 25$. The standard error for the difference in means is

$$s_e = \frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}},$$

Under the null hypothesis $H_0 : \mu_1 = \mu_2$, the test statistic is computed as

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE},$$

which follows a t-distribution with $n_1 + n_2 - 2 = 48$ degrees of freedom. A two-tailed p-value was calculated using the cumulative distribution function of the t-distribution. A significance level of $\alpha = 0.05$ was used to determine whether the observed difference in group means is statistically significant.

All computations, random sampling, graphics, and hypothesis testing procedures were conducted using Python, relying on the `pandas`, `numpy`, `matplotlib`, and `scipy.stats` libraries.

III. Results

IV. Conclusion

V. References and Resources