

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 less than or equal to five absences and those with more than five. 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 [1].

This project examines this question using an openly available dataset of student GPA 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, using Python's pandas library to ensure that each student had an equal probability of selection.

The 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

$$SE = \frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}}.$$

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

This section presents descriptive statistics, visual comparisons of GPA distributions for the two absence groups, and the results of the two-sample hypothesis test.

A. Descriptive Statistics

Table 1 summarizes the sample means, standard deviations, and sample sizes for each absence group. Students with five or fewer absences exhibit considerably higher GPAs on average, while those with more than five absences show notably lower academic performance. These preliminary trends motivate the hypothesis test conducted in the following subsection.

Table 1 Descriptive statistics for GPA by absence group.

| Group | <i>n</i> | Mean GPA | Standard Deviation |
|-------------------|----------|----------|--------------------|
| ≤ 5 Absences | 25 | 3.21298 | 0.37419 |
| > 5 Absences | 25 | 1.71087 | 0.73099 |

B. Distributional Comparisons

Two graphical analyses were produced to compare the GPA distributions across absence groups. Figure 1 displays the empirical cumulative distribution functions (CDFs). The high-absence group's CDF rises more sharply, indicating that a larger proportion of these students attain lower GPA values relative to the low-absence group. This visual trend is consistent with the lower mean GPA observed for high-absence students.

A complementary comparison is shown in Figure 2, which overlays histograms and kernel density estimates (KDEs) of GPA for both groups. The distribution for the low-absence group is concentrated towards higher GPAs, whereas the high-absence group displays a left-shifted and more dispersed distribution. These differences highlight both a shift in central tendency and wider variability among students with more absences.

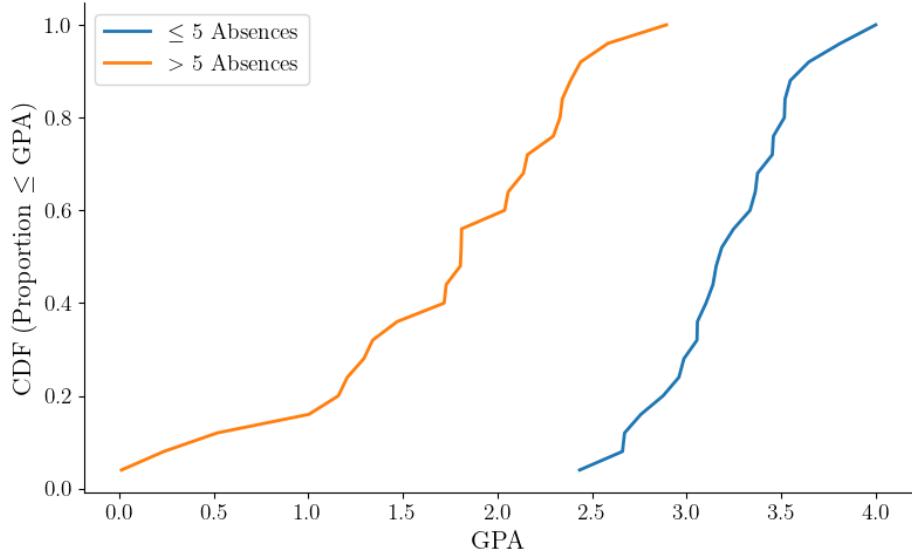


Fig. 1 Empirical cumulative distribution functions of GPA for students with ≤ 5 absences and > 5 absences.

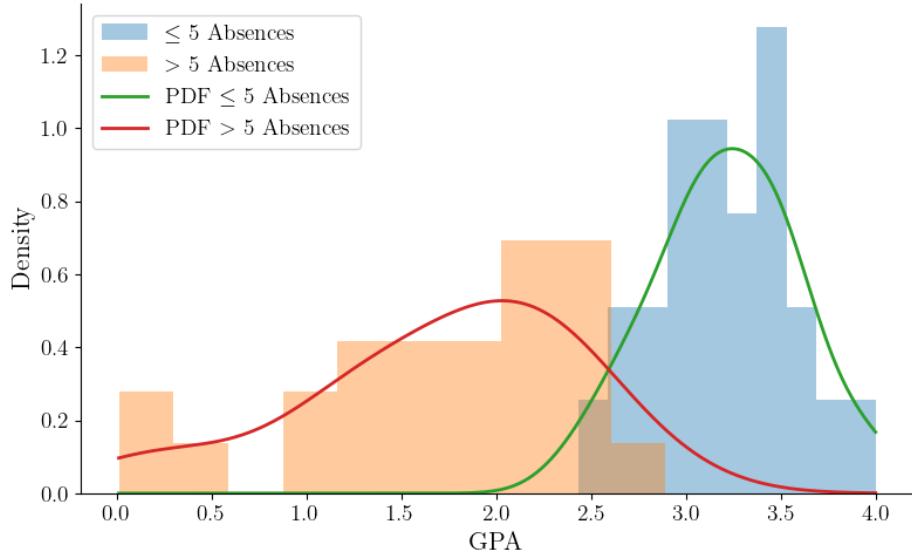


Fig. 2 Histogram and KDE-based probability density estimates of GPA for low-absence and high-absence groups.

C. Hypothesis Test

To formally assess whether the mean GPA differs between absence groups, a two-sample t-test was performed. The hypotheses tested were

$$H_0 : \mu_1 - \mu_2 = 0.0, \quad H_a : \mu_1 - \mu_2 \neq 0.0.$$

Using the sample means, standard deviations, and standard error described in the Methods section, the resulting two-tailed p-value was extremely small ($p < 4e - 11$), providing strong evidence against the null hypothesis. At the $\alpha = 0.05$ significance level, we reject H_0 and conclude that the mean GPAs of the two groups differ.

In addition, a 95% confidence interval was constructed to estimate the magnitude of this difference. The interval was found to be from 1.17 to 1.83 indicating that students with five or fewer absences score, on average, between 1.17

and 1.83 GPA points higher than students with more than five absences. Because the entire confidence interval lies well above zero, it provides further confirmation of a meaningful and statistically significant difference in academic performance.

IV. Conclusion

Across all analyses, the results consistently indicate that absenteeism is strongly associated with lower academic performance. Students with more than five absences not only have significantly lower mean GPAs but also exhibit distributional shifts toward poorer academic outcomes, as demonstrated in Figures 1 and 2. The hypothesis test and confidence interval confirm that these differences are statistically significant. Together, these findings suggest that frequent absences have a meaningful and detrimental impact on student performance.

V. References and Resources

References

- [1] Wolfson, A. R., and Carskadon, M. A., “Understanding adolescent’s sleep patterns and school performance: a critical appraisal,” *Sleep Medicine Reviews*, Vol. 7, No. 6, 2003, pp. 491–506. [https://doi.org/10.1016/S1087-0792\(03\)90003-7](https://doi.org/10.1016/S1087-0792(03)90003-7).