Summary of Statistical Analysis

Bobga-Herman Gwanvoma

DSC 530

Statistical/Hypothetical Question:

The primary question driving this analysis was, “Does the number of hours spent studying per week affect college students' final exam scores, considering their class attendance, participation in extracurricular activities, amount of sleep, and socioeconomic background?” This inquiry aims to identify the most significant factors influencing academic performance, focusing particularly on study habits while accounting for various other lifestyle and background variables.

Outcome of Exploratory Data Analysis (EDA):

The exploratory data analysis revealed several key insights. The distribution of study hours showed moderate variation among students, with a mean around 15 hours per week. The final exam scores also varied, with most students scoring between 60 and 85. Histograms of these variables indicated that while most students had regular study habits, a few outliers studied either very little or an excessive amount. The analysis of class attendance suggested that more frequent attendance might correlate with higher exam scores, although this relationship was not strictly linear. Participation in extracurricular activities and socioeconomic status showed weaker correlations with exam performance.

A comparison using probability mass functions (PMFs) indicated that students who frequently attended classes generally had higher exam scores compared to those who attended less often. Additionally, the cumulative distribution function (CDF) of study hours highlighted that a significant portion of students studied less than the average, potentially explaining some of the lower exam scores.

Missed Variables and Assumptions:

One limitation of the analysis was the exclusion of variables such as students’ prior academic performance (e.g., GPA), access to learning resources (e.g., tutoring, online materials), and psychological factors like stress and motivation levels. These factors could significantly influence academic outcomes and might have provided a more comprehensive understanding of the determinants of final exam scores.

Moreover, assumptions about linear relationships between study hours and exam scores might oversimplify the complexity of learning and performance. For instance, diminishing returns on study time, where additional study hours yield progressively smaller improvements in exam scores, was not accounted for in this analysis.

Challenges and Areas of Uncertainty:

One challenge encountered was ensuring that the dataset accurately reflected real-world scenarios, especially since this was a hypothetical dataset. This limitation makes it difficult to generalize the findings to a broader population. Additionally, understanding the implications of various statistical tests, such as hypothesis testing and regression analysis, posed some difficulties. In particular, differentiating between correlation and causation was challenging, as variables like class attendance might be correlated with exam performance without directly causing it.

In conclusion, while the analysis provided useful insights into the factors affecting student performance, it highlighted the need for a more nuanced approach, considering additional variables and accounting for more complex relationships. The process also underscored the importance of critical thinking in interpreting statistical results, particularly in distinguishing between correlation and causation.