Zane Alderfer

Dr. Bolton

IST 707

October 10, 2023

Homework 2

**Introduction**

Around the globe, the standards and values placed upon education vary. Not only does the quality of education fluctuate across countries but even within one country, a lack of uniformity with educational expectations can differ drastically. The importance of understanding why this takes place within a country, namely the United States, is the first step in planning and devising a course of action to develop solutions. In order to ensure the most effective form of education for everyone in a country, accurate record keeping of performance by every party involved is vital.

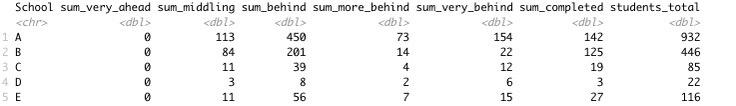
Parents are not the only ones who care if their child gets an A on their math assignment. Schools, private institutions, and government entities such as the Department of Education keep track of school performance. A number of variables are taken into account for data collection with regard to education. Demographics, type of area (rural or urban), number of teachers, number of schools, and plenty of other variables are carried out in data collection by the National Center for Education Statistics (NCES, 2020). A number of supplemental areas of study such as labor force, government finances, and economic trends are also of interest to the NCES (NCES,2020).

Historically, the federal government has been collecting data on the progression of American education since the 1870s (Snyder, 2018). Thanks to the history of this data collection, the overall growth and importance placed on education has been readily apparent with a steady increase of attendance at every level of education. As of recently, there have even been studies that involve a longitudinal type approach that observes a child from birth to kindergarten and how this may shape the kind of student they’ll end up being (Snyder, 2018). While every child is different and has different levels of ambition, continuing to compare the quality of education across schools is crucial in understanding how a student in a less favorable environment can unlock their full potential as a student.

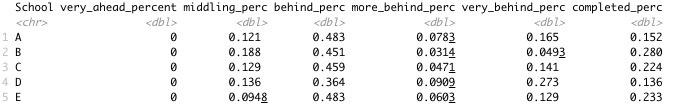
**Analysis and Models**

**About the Data**

Some schools are trying a new math course and to keep track of how well the students are doing in this course, they have been divided into subgroups to keep track of their progress. These subgroups are “very ahead,” “middling,” “behind,” “more behind,” “very behind,” and “completed.” The subgroups directly correlate to how many lessons the students have completed ¾ of the way into the semester. “Very ahead” would apply to any student 5 or more lessons ahead, “middling” is on pace, “behind” is 1-5 lessons behind, and so on. The dataset under review has 5 schools(A, B, C, D, E) with sections at each school totalling 30 sections of classes. The table below shows how many students in total are in each subgroup with a grand total in the last column for each school.



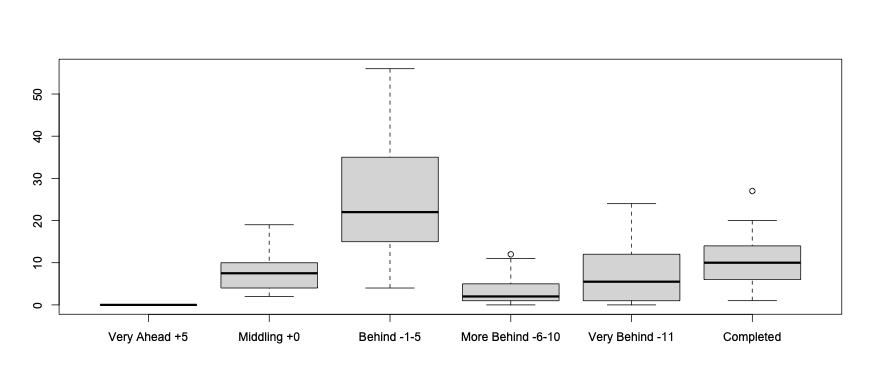
In terms of cleaning the data, the table above was developed by grouping each school’s sections together to get a full count for each school. To make the data relative, however, a second table was made finding the percentage of each subgroup to the total number of students at that school. This is shown in the table below.



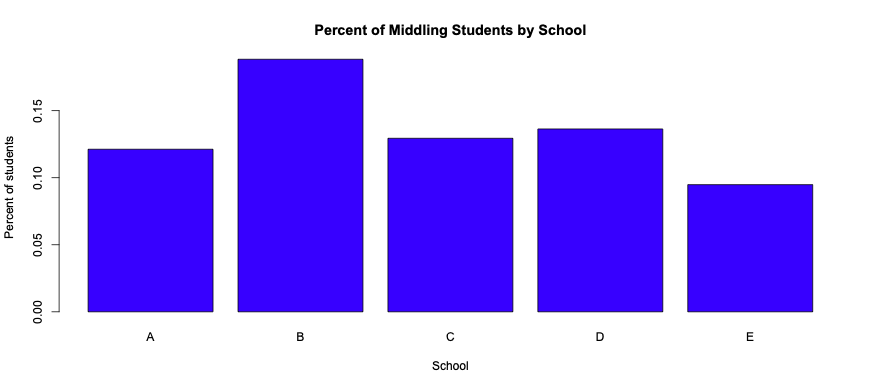
There was no missing information in the dataset or outliers that drastically changed the data so using the data as is was more than sufficient for this analysis.

**Analysis of the Models**

**Figure 1.** This boxplot presents a view of how many students in each section are in each subgroup. How many sections are in each school are irrelevant as each section is measured equally here.

****

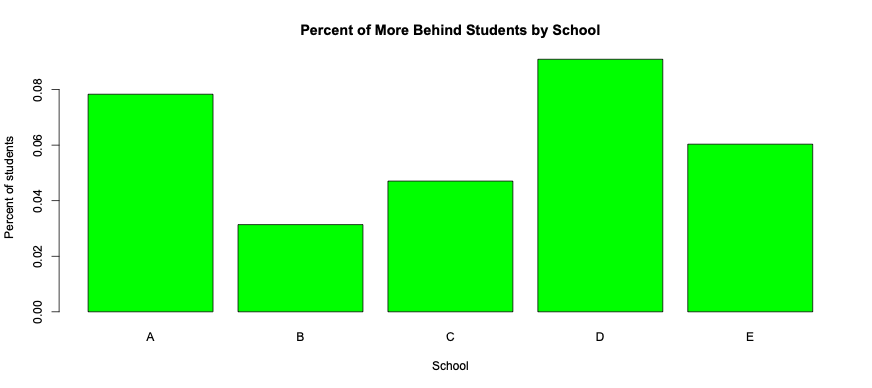
**Figure 2.** This barplot presents the percent of students in the “middling” subgroup. To see a more relative view for each school, the average number of students in each section for this subgroup was used in the calculation for each school instead of the total number of students at each school.

****

**Figure 3.** This barplot presents the percent of students in the “behind” subgroup. For the same reason as above, the average number of students for each section was used to present a more relative model.



**Figure 4.** This barplot presents the percent of students in the “more behind” subgroup. For the same reason as above, the average number of students for each section was used to present a more relative model.



**Figure 5.** This barplot presents the percent of students in the “very behind” subgroup. For the same reason as above, the average number of students for each section was used to present a more relative model.



**Figure 6.** This barplot presents the percent of students in the “completed” subgroup. For the same reason as above, the average number of students for each section was used to present a more relative model.



**Results**

There were no correlation methods or statistical methods such as z-score, t-stat, etc. The barplots and boxplot presented enough information and visuals to adequately explain the data. The dataset was also small so to suggest some kind of correlation between these schools and what subgroups their students fall into seems futile and non-truthful.

**Figure 1**

In figure 1, it can be noted that the biggest percentage of students lie within the “behind” subgroup. Interestingly, there is no student in the “very ahead” subgroup. A big takeaway from this model would be that there is a similar average number of students in each section who are “very behind” and who have “completed” the course.

**Figure 2**

It should be mentioned that schools A, B, and C all had multiple sections while D and E only had one section each. The larger sample size from the former 3 schools could lead to more accurate results and analysis. With that being said, school B and D had a larger percentage (12% +) of students who were on pace to complete the course on time. This is a very small margin, however, and shouldn’t be assumed to remain the same by the time the semester ends.

**Figure 3**

Figure 3 presents the most uniformity across the 5 schools in a subgroup. The “behind” subgroup also holds the highest percentage of students in each section on average.

**Figure 4**

The “more behind” subgroup has the smallest percentage of students on average per section. School D has the highest percent of students in the section with 8%+ of their students “more behind.” School B and C are both doing well in this subgroup as both schools average less than 5% of their students in this section.

**Figure 5**

Students who fall very behind tend to keep falling further behind as School D has the highest percentage of students (25%+) in their section who have fallen very behind. School B, however, has the lowest percentage of students at 5% or less on average in each section.

**Figure 6**

This is one of, if not, the most important models. School D also has the smallest percent (14%) of students who have completed the course. School B however has a respectable 25%+ with the other 3 schools falling in between them.

**Conclusion**

After looking at the boxplot in figure 1, the amount of students who have completed the course and the amount of students who are very behind are similar in percentage of average. This would suggest that for the most part for every student who falls very behind, there’s a student who overachieves. Due to no students being in the “very ahead” subgroup, those students who overachieve expectations tend to overachieve by a wide margin. After seeing the “middling” subgroup, there’s a large percentage of students from each section on average in this subgroup. This is interesting as it could suggest that these 5 schools could potentially be in a similar area or contain a similar demographic that would lead to similar results.

When looking at figure 4, there’s the smallest percent of students on average in each section in this subgroup. While this seems like an insignificant subgroup, this can be complementary to figure 5 as the lack of students in the “more behind” subgroup could suggest that the students who fall behind in the classes fall very behind and as a student reaches a point of “no return,” they tend to give up. That being said, School D also has the highest percentage of students in the “more behind” subgroup which would suggest that this section from School D could have more factors impacting their ability to finish the course such as teaching, outdated learning methods, etc.

Based on the average percentages of each section in each subgroup, School B presents as the most respectable among the 5 schools in terms of completing the course in a timely manner. School D, however, would appear to fall short of satisfactory as they have a large number of students who are “more behind” or “very behind.” While this data is interesting but small, it would be nice to have more information like demographics, school funding, number of tutors per school, etc. The dataset given is probably not applicable in terms of making concrete conclusions and analysis toward discovering correlating impacts.

**References**

National Center of Education Statistics (NCES). “Digest of Education Studies.” Institution of Education. 2020. Web. 20 October 2023. <https://nces.ed.gov/programs/digest/>

Snyder, Tom. “Celebrating 150 Years of Education Data.” Institution of Education. 2018. Web. 20 October 2023. <https://nces.ed.gov/blogs/nces/post/celebrating-150-years-of-education-data>