IST 707: Homework 2

**Introduction**

In order to gauge the value of local and federal government investment in education, public schools are required to quantify the success of their curriculum. Schools often use metrics such as test score rankings to prove the progress students are making in retaining key curriculum concepts. Higher performance ratings can lead to additional funds being invested while underperformance can threaten a school’s funding. If test scores (or other pertinent measures of curriculum impact) are unsatisfactory, schools may explore revising certain courses to mitigate the risk of lost funding.

A recent study has shown that the mathematics scores of five Syracuse, New York, public high schools over the last ten years have been consistently lower than the majority of other New York public schools. Local parents are concerned that their children are receiving a sub-par education. Additionally, students are concerned that their chances of being admitted into colleges may be hurt by a perceived lack of quality in education compared to other New York students. Principals at the schools are concerned that government funding may be cut which may endanger the jobs of certain staff members.

In an attempt to ease the angst among the community, the schools have agreed to implement a common math course this semester. The schools hope that standardizing the curriculum will lead to a stabilization of test scores. Furthermore, standardization should expedite curriculum improvements with a wider audience for course feedback. Each school has been tracking student course completion statuses in hopes of finding any early indications about the performance of the new math course. At the time of this report, the semester is currently about ¾ complete.

**Analysis and Models**

**About the Data**

The data provided for the purpose of this report was supplied by all five schools. In order to avoid bias, each school was assigned an anonymous letter A-E for its identifier (i.e. School A, School B etc.). Although data was supplied from each school, some schools are offering more sections of the course this semester; therefore, school representation varies within the data. There is a total of 30 sections of the course being taught this semester.

The original dataset contained eight variables which were as follows:

|  |  |
| --- | --- |
| Variable Name | Variable Description |
| School | Letter A-E used to represent each of the schools |
| Section | Categorical number 1-13 representing each instance of a course being taught at a school |
| Very Ahead +5 | Number of students in a section that are more than 5 lessons ahead of the curriculum schedule. |
| Middling +0 | Number of students in a section that a 0-5 lessons ahead of the curriculum schedule |
| Behind -1-5 | Number of students in a section that are 1-5 lessons behind the curriculum schedule |
| More Behind -6-10 | Number of students in a section that are 6-10 lessons behind the curriculum schedule |
| Very Behind -11 | Number of students in a section that are more than 11 lessons behind the curriculum schedule |
| Completed | Number of students in a section who have completed the course |

In order to improve code readability and reduce issues with spaces and symbols in the original names, each of the variables were renamed to the following:

["School," "Section," "VeryAhead," "Middling," "Behind," "MoreBehind," VeryBehind," "Completed"]

Again, for increased readability, the columns were rearranged to reflect the proper ascending order of course completion status from “VeryBehind” to “Completed”:

[“School,” “Section, “VeryBehind,” “MoreBehind,” “Behind,” “Middling,” “VeryAhead,” “Completed”]

Initial observation of the data showed that there were currently no students who were “VeryAhead” in any of the sections being taught. Due to this lack of data, “VeryAhead” was removed.

As previously mentioned, there is an uneven amount of data from each of the schools. Likewise, the size of sections being taught has not been standardized—leading to a range in section size from as few as 13 students to as many as 116 students. To capture this dynamic in the data, the variable “StudentsPerSection” was added as a sum of all students per section.

**Course Completion Status Ratio Model**

**Section-level Ratio Model**

The disparity in sections taught and section size at each school make direct comparisons of the number of students with certain completion statuses imprecise. To offset this aspect of the data, a model using a series of ratios of student completion statuses to total students per section was developed. In support of this model, the following variables were created and added to the data:

|  |  |
| --- | --- |
| Variable Name | Variable Description |
| VeryBehindRatio | Ratio of “VeryBehind” students in a section to total number of students in a section.  Calculated as:  “VeryBehind” / “StudentsPerSection” \* 100, rounded to nearest .01 |
| MoreBehindRatio | Ratio of “MoreBehind” students in a section to total number of students in a section.  Calculated as:  “MoreBehind” / “StudentsPerSection” \* 100, rounded to nearest .01 |
| BehindRatio | Ratio of “Behind” students in a section to total number of students in a section.  Calculated as:  “Behind” / “StudentsPerSection” \* 100, rounded to nearest .01 |
| MiddlingRatio | Ratio of “Middling” students in a section to total number of students in a section.  Calculated as:  “Middling” / “StudentsPerSection” \* 100, rounded to nearest .01 |
| CompletedRatio | Ratio of “Completed” students in a section to total number of students in a section.  Calculated as:  “Completed” / “StudentsPerSection” \* 100, rounded to nearest .01 |

The primary ratio variable for this model is based on the ratios associated with a less than on-track course completion status. This variable (“ConcernRatio”) is a sum of “VeryBehindRatio,” “MoreBehindRatio,” and “BehindRatio” for each section. The “ConcernRatio” indicates the percentage of students who are one or more lessons behind the course curriculum schedule in a section with less than ¼ of the semester remaining. The ratio model as described can be used to compare percentiles of students, particularly those in the “ConcernRatio” range at the course section level; however, it cannot be used at the school level.

**School-level (Aggregate) Ratio Model**

To address the disproportionate school representation, an aggregation was performed on the existing section-level ratio model data. Each of the original values as well as the added ratios were averaged across course sections by school, resulting in one representative record for each school. Although this aggregation allows cross-school comparison, the data may be less representative of schools A, B, and C who have wide ranges of “StudentsPerSection” (48, 29, and 23 respectively). The limitation of this model is that there is not a weighted aspect to the aggregate averages, so higher or lower completion rates in smaller sections may skew the school-level values.

The school-level aggregate data can be used with the aforementioned ratio model to highlight “ConcernRatio” percentiles across schools. Additionally, the aggregate level data can be used to explore the impact of average students per section on completion rates.

**Results**

**Section-level Results**

As part of the section-level ratio model, each section was segmented into percentiles by course completion status. From a section perspective, the most differentiating element of each record-set within the data is the section size (“StudentsPerSection”).

**Figure 1**

**A close up of a map

Description automatically generated**

When comparing the “ConcernRatio” and “CompletedRatio” for each of the sections relative to the number of students per section, there is roughly a correlation between smaller section sizes and reduced concern/increased completion rates (see Figure 1). These results suggest that perhaps the schools should limit course section sizes in the future to promote student success.

To confirm the linkage between students per section and student success, this model should be reapplied to student grades or test score data. These results support this theory; however, they are not representative of an entire semester. Additionally, completion status rates may change at an exponential rate as opposed to a linear rate as students approach the end of the semester. Extending this model to grades and/or test scores should indicate the quality of completion statuses as a barometer for student performance.

**School-level (Aggregate) Results**

With the aggregate-level model, summary statistics resulted in the following averages for all five schools:

|  |  |  |  |
| --- | --- | --- | --- |
| Stat | StudentsPerSection | ConcernRatio | CompletedRatio |
| Min | 22.00 | 52.06 | 13.64 |
| 1st Quartile | 28.33 | 60.76 | 16.08 |
| Median | 37.17 | 67.24 | 23.28 |
| Mean | 55.04 | 64.96 | 21.51 |
| 3rd Quartile | 71.69 | 72.04 | 25.78 |
| Max | 116.00 | 72.72 | 28.74 |

This statistical summary helps to illustrate the previously described disparity in section size. Figure 2 (below) further supports the notion that higher “ConcernRatio” values may be associated with larger section sizes. The upper half of the “StudentsPerSection” data is rather elevated which may be inflating concern ratios.

**Figure 2**

**A screenshot of a cell phone

Description automatically generated**

In addition to comparison based on section size, the aggregate model can be used to compare concern ratios of completion status by school. The distribution of average concern ratio is shown in Figure 3. It does not appear that any school is having substantially lower average “ConcernRatio” values; however, School B does seem to have somewhat lower values. It seems that schools are having difficulty keeping students on pace with the current curriculum.

**Figure 3**

A screenshot of a computer

Description automatically generated

It would be beneficial to do a similar comparison to grades at end of term as recommended with the section-level model. If the schools wish to use this aggregate data for immediate decision-making, it indicates that reduced section sizes lead to lesser “ConcernRatio” values. School B has relatively smaller sections compared to the other schools which may be the cause of their slightly lower concern ratios.

**Conclusions**

In summation, five Syracuse public high schools have implemented a new standardized math course to help improve student performance. To measure course in-progress success, each of the schools has tracked how many lessons students have completed relative to the outlined schedule in the curriculum. By grouping students who are one or more lessons behind schedule into a “concerned” category, it can be calculated what percentage of students are falling behind.

Across all of the schools, it appears that something about the curriculum is causing students to learn at a pace slower than anticipated. On average, each school has at least half of its students in the “concerned” category. Based on the data available, the only clear differentiator between each school is the size of the course sections being offered.

School B has the best student performance of the schools at this point in the semester. Furthermore, School B also has relatively smaller section sizes. The data points to a linkage between reduced course section size and improved completion status. It is recommended that the schools confirm this relationship by re-using the models in this report with final grades and/or test score data. However, the actionable insight at this point in the semester is to reduce course section size going forward.