61A Exam Data Analysis: Measure of Academic Dishonesty

Kavi Gupta

April 15, 2017

1 General Strategy

The general strategy is to look at similarities between scores of students sitting next to each other against the similarities between scores of students sitting not next to each other.

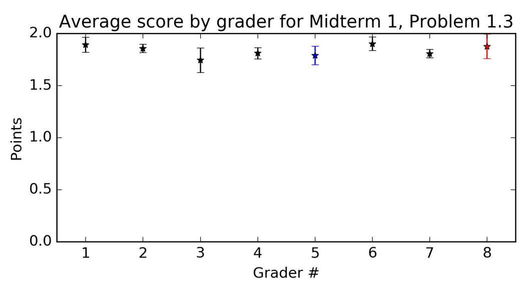
The result can be then used to estimate the number of students committing academic dishonesty using a model.

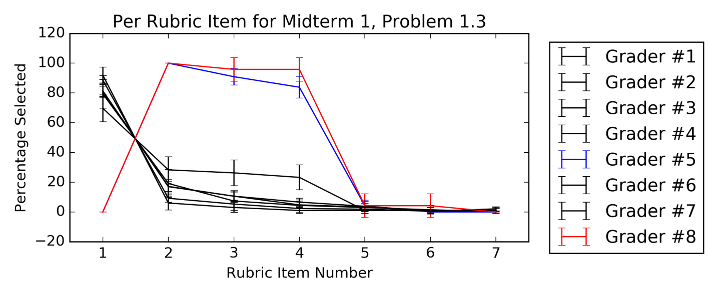
2 Confounding Factors

There are many potential confounding factors, which we control for in different ways.

2.1 Grader Differences

Some graders grade differently from each other. Taking Midterm 1 problem 1.3, we can see that graders 5 and 8 had fairly typical behavior in terms of average scores given. However, we can also see that they have a very different profile of rubric items. In this case, what happened was that these two graders were giving rubric items 2, 3, and 4, which corresponded to each of the three parts of the problem being correct, rather than giving rubric item 1, which corresponded to the entire problem being correct. While this is irrelevant to the student, it does confound analyses.





Since grading ranges tend to be at least somewhat associated with location, this might lead to students near each other having more similar profiles, artificially. We control for graders by subtracting out from each student's individual problem scores and rubric items the mean given by a grader.

Any student who had at least one problem that was either graded by a grader who had graded fewer than ten problems (not enough of a track record) or who had an abnormal pattern of rubric item grades[1].

If we do not use individual rubric items, but instead overall question scores, we can use more similar formulas but discard less data as graders' question scores are more similar to each other.

[1]Unusualness of a grader is defined as

$$u = \sum_k \left| \frac{x_k - \mu_k}{\sigma_k} \right|$$

where x_k is the mean for the given grader of the kth rubric item, and μ_k, σ_k are the mean and standard deviation for that rubric item in general

2.2 Sequencing

The Gambler's Fallacy could possibly lead to a negative correlation between consecutively graded exams. To control for this, we simply ignore pairs of exams that are graded near each other in time.

2.3 Regional Differences

Individual rooms might have different rubric profiles, due to, for example, different TAs in different rooms. Additionally, individual rows or locations within rows might have different profiles themselves. To compensate for this, we compare the similarities of students sitting next to each other with the similarities of students sitting 2 seats away from each other.

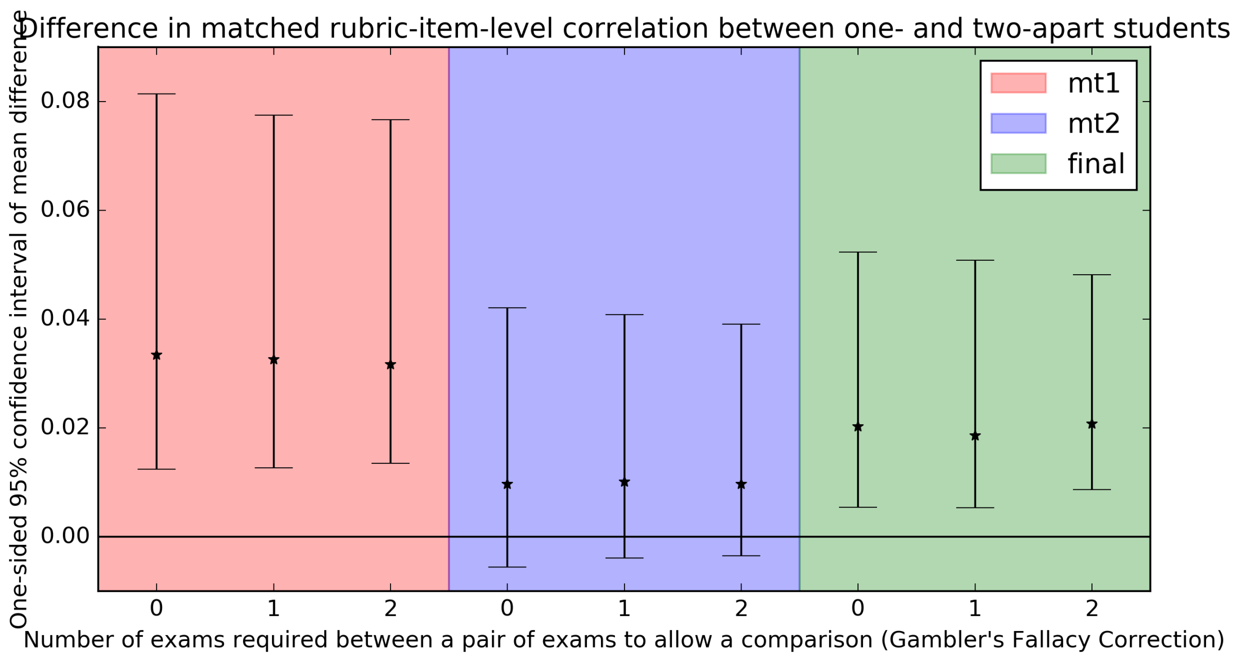
2.4 Seating Chart Inaccuracies

Not yet fixed. Have the data, but would require a lot of data input because OCR is not particularly effective at recognizing handwriting.

3 Measuring Similarity

3.1 Rubric-Item Level Correlation

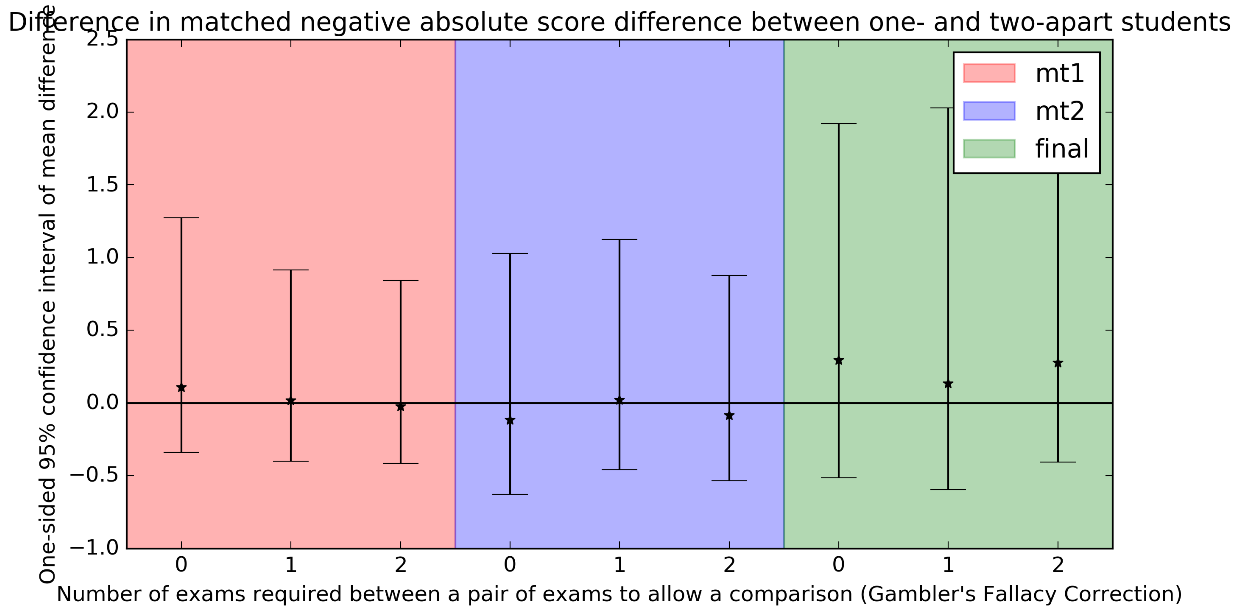
Using correlation between normalized scores as a measure of similarity, we can prove that there is a difference between the similarities in the set of adjacent pairs of students and the set of non-adjacent pairs of students for Midterm 1 and the final, but not for Midterm 2 (see controls above for more details). We can also see that the gambler's fallacy correction does not significantly impact it.



3.2 Absolute Difference in Overall Exam Score

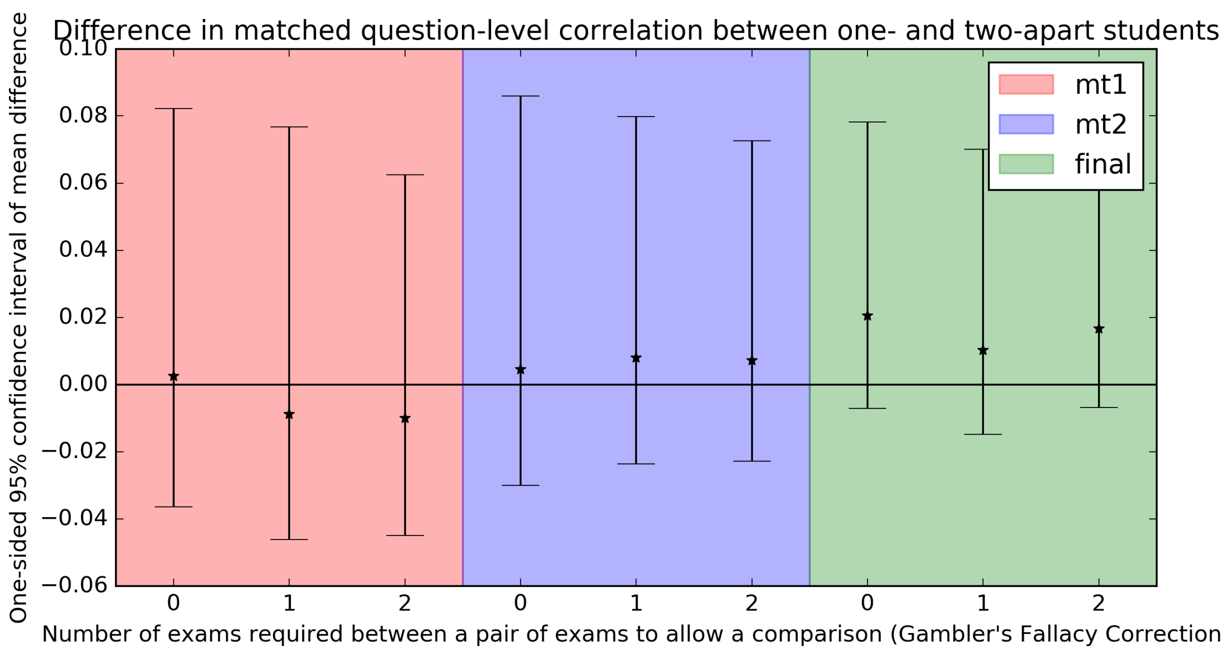
An alternate means of similarity is far simpler. We look at the average difference between the scores in the two exams. For this measure of similarity, we do not need to remove extreme graders.

Unfortunately, this does not seem to be granular enough to find cheaters.



3.3 Question Level Correlation

A compromise between the power of the rubric-item level correlation and the simplicity of the absolute difference in exam score is question-level correlation. Unfortunately, it still lacks the power to see beyond the noise.



4 Model For Academic Dishonesty

4.1 Score-Independent, Binary Cheater, Random Cheat-Selection, Model

Assume that some fixed fraction c of the students cheat by copying exactly k randomly selected points worth of material from a person sitting next to them (also selected randomly). Assume that non-cheating students get each point independently with probability p. Further assume that cheating students are located randomly.

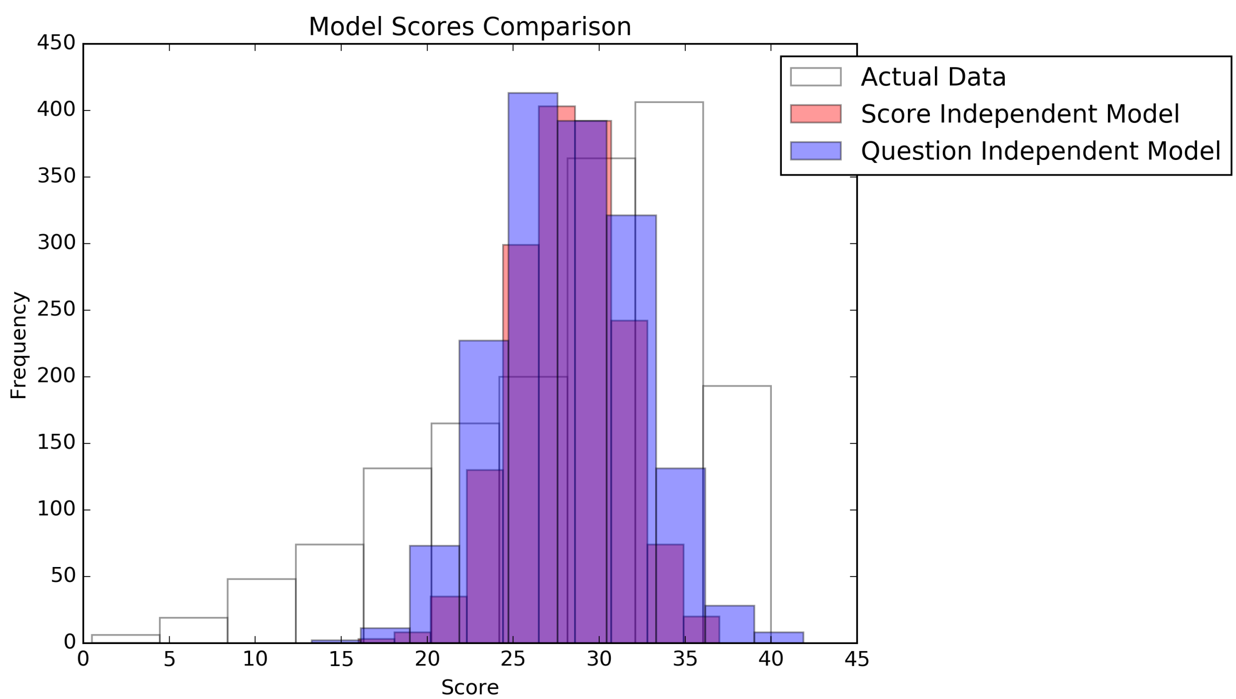
Using known exam data, we can approximate p simply by taking the average score divided by the total number of points.

Unfortunately, the grades this model produces do not approximate the true distribution of grades, as the graph below demonstrates.

4.2 Question-Independent, Binary Cheater, Random Cheat-Selection Model

This model is equivalent to the one above, but uses point chunks representing each question, which are assumed to be independent and take the sizes of each of the questions. A certain number of questions rather than points are copied. Each question is assigned a score based on a normal distribution from the mean and variance of the real data for that question.

Unfortunately, this strategy also produces a distribution of grades dissimilar to the original data.



4.3 Random Seating, Binary Cheater, Random Cheat-Selection Model

This is similar to above, but the initial non-cheater distribution is created by shuffling the students into random locations. The pitfall of this method is that some information about the cheating is encoded in this distribution.

However, if we assume that the amount of cheating is fairly low, it should be a fairly accurate model overall.