Does Instructor Rank Matter? Grade Variation Among Math Courses at CSUN, 2005-2014

Author: Maria Verna Aquino **Advisors: Dr. Bruce Shapiro, Dr. Carol Shubin**
Department of Computer Science & Mathematics, California State University Northridge, CA, USA

Abstract

Grading patterns across the Department of Mathematics at the California State University, Northridge over a ten-year period (2005-2014) were examined to determine if there were differences in grading between instructors of different rank. The ranks were divided into Tenured, Tenure-track, Adjunct and Non-Faculty (TA/GA/Other) instructors. A chi-squared goodness of fit and chi-squared test for independence of the variables: "grade" and " type of instructor", were done for each Math course in order to determine how differently instructors graded from each other. The p-values resulting from these tests dictate how similar the grading distributions are from each other. Significant differences in grading were found in most of the Math classes taught which is apparent in the low p-values obtained.

Introduction

This study aims to determine if we can identify some of the root causes of grade variation in Math classes. Specifically we want to identify if instructor rank plays a role in grade variation.

Background

Through this study we can answer questions like:

- How do grading patterns vary across the Math department?
- Are some instructors particularly hard graders or easy graders?
- How do grading patterns vary within single Math classes? A histogram of grades reveal more than the mean and standard deviation. For example, two instructors may have identical mean and standard deviations, but one may choose to give only A, B, C, D and F grades, while the other splits the B grades equally into B+, B, and B-, giving very different histograms.

Methods

The **Chi-Squared test of independence** is used to determine if there is a significant relationship between two variables. The frequency of each category for one variable is compared across the categories of the second variable. In this study, we want to examine the relationship between instructor ranks (Adjunct , Tenured/Tenure-track , Non-faculty) and grading patterns for each Math class. The null hypothesis for this test is that there is no relationship between instructor rank and grading. The alternative hypothesis is that there is a relationship between instructor rank and grading.

Null hypothesis: Assumes that there is no association between the instructor ranks.

Alternative hypothesis: Assumes that there is an association between the instructor ranks.

Hypothesis testing: Hypothesis testing is used to find out whether or not we have meaningful results. To do hypothesis testing, first we compute from the observations (i.e. grades) the observed value of the test statistic chi2. Second we calculate the p-value. The p-value is the probability, under the null hypothesis, of sampling a test statistic (chi2) at least as extreme as that which was observed. We reject the null hypothesis in favor of the alternative hypothesis, if and only if the p-value is less than the significance level threshold of 0.05. The chi2 statistic and p-values were calculated using the chi2_contingency function from Python's scipy.stats library.

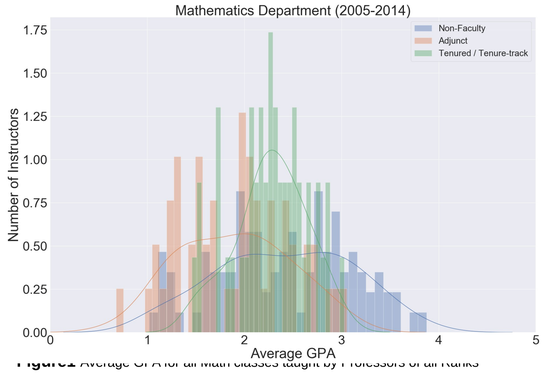


Figure 1 Average GPA for all Math classes taught by Professors of all Ranks

Adjunct VS Tenured/Tenure-track

Course Number	Adjunct Graded	Tenured/Tenure-track Graded	Chi2	P-val
102	4106	707	6.33	0.0967
103	3314	1913	0.0896	0.993
104	1944	192	21.7	0.0000747
105	355	371	13.6	0.00357
131	1178	483	6.92	0.0746
140	4102	5285	11.2	0.0105
150	930	2495	1.79	0.617
210	559	598	8.38	0.0388
250	658	725	9.29	0.0257
255	847	380	3.26	0.353
262	120	510	18.5	0.000344
280	279	641	24.5	0.0000199
310	896	454	6.99	0.0721
340	41	399	7.53	0.0567
341	35	282	0.966	0.809

Table 2 Adjunct and Tenured instructors grade differently for Math 104, Math 105, Math 140, Math 210, Math 250, Math 255, Math 262 and Math 280

Adjunct VS Tenured/Tenure-track VS Non-faculty

Course Number	Adjunct Graded	Tenured/Tenure-track Graded	Non-faculty Graded	Chi2	P-val
102	4106	707	1316	43	1.19e-7
103	3314	1913	217	5.72	0.456
104	1944	192	57	32	0.0000163
131	1178	483	3295	182	1.2200000000000002e-36
140	4102	5285	235	30.3	0.0000337

Table 1 Adjunct, Tenured and Non-faculty Instructors grade differently for Math 102 (Pre-Calculus I) , Math 104 (Trigonometry) , Math 131 (Mathematical Ideas) and Math 140 (Intro Statistics)

Results

Table 1 and Table 2 show the results of the Chi-squared test for independence on the grades given by different instructor ranks. The p-values in red refer to the statistically significant results (i.e. $p < 0.05$) which means that there is only a 5% likelihood that the grades from a given instructor rank could have been produced by normal random variation from another instructor rank. In other words, a significant (red) p-value means that the instructor ranks being compared in the Chi-squared test grade differently for that specific Math class. If the p-value is not significant (not red), this means that the test is inconclusive for that specific Math class.

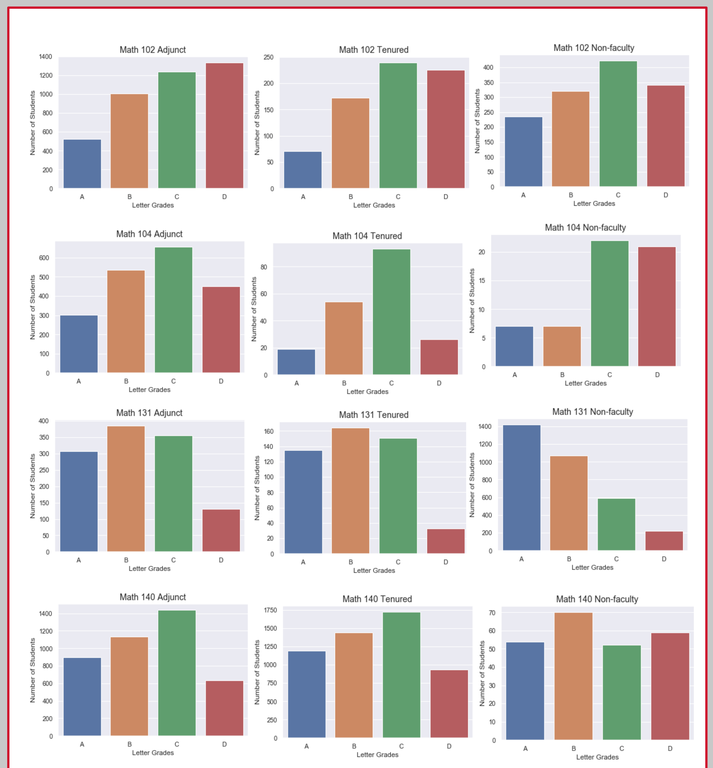
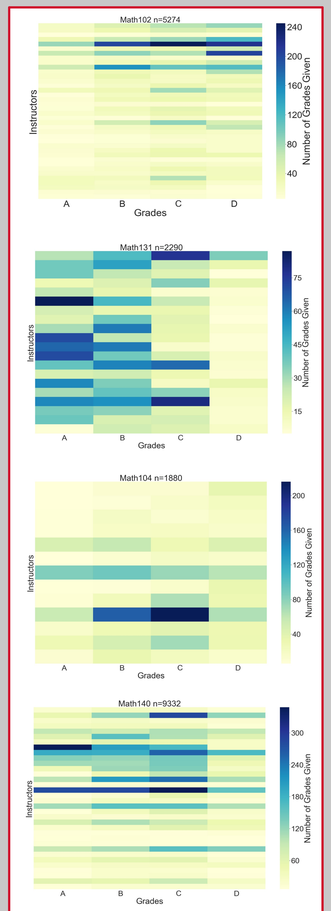


Figure 2 Shows the difference in grade distributions between Math 102 (Pre-Calculus I) , Math 104 (Trigonometry) , Math 131 (Mathematical Ideas) and Math 140 (Intro Statistics) classes.



Grades

Discussion and Conclusion

Figure 1 shows the average GPA given by instructors of different ranks for all Math classes that give out letter grades. Already, it can be seen that in general, the distributions are quite different.

Figure 2 illustrates why there are differences in p-values between Table 1 and Table 2. Notice that for Math 102 and Math 131, the distributions of the Adjunct VS Tenured instructors are quite similar. This image agrees with the inconclusive results we got for these classes in Table 2. However, when comparing Adjunct VS Tenured VS Non-faculty, a significant p-value is obtained for Math 102 (as seen in Table 1) due to the addition of the Non-faculty rank grading distribution.

Figure 3 shows different heatmaps for the significant results from Table 1. Each row in the heatmap corresponds to an instructor teaching a specific Math class. Each column corresponds to an aggregated letter grade (e.g. A+ and A- are treated the same as an A). The number of students given a certain letter grade by each instructor is shown using the color maps on the right. A light yellow means less students and dark blue means more students. The total student population given grades A to D for each specific class is listed as the n value in the graph's title.

From Figure 3 you can notice some trends in grading. For example, in the Math131 heatmap the left side of the map is considerably darker than the right side of the map (i.e. more students passing). This shows that instructors tend not to give D's for this particular class and can mostly be considered easy graders. You can also notice that certain instructors are hard graders. For example, a hard grader at the top of the Math102 heatmap has failed most of the students in their class. The number of D's they have given looks to be in the 200's range while the rest of the letter grades are in the 80s or less.

In conclusion, instruction rank does matter for some classes. Instructors of different ranks do grade differently for certain classes. This result can be seen by the significant p-values obtained from certain Math classes and can be illustrated by graphing the different grading distributions for these Math classes.

Acknowledgments

Supported by "US National Science Foundation Award 1842386: Data Science Program with Career Support and Connections to Industry"

Citations / References

- https://www.statisticssolutions.com/non-parametric-analysis-chi-square/
- https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

CSUN® COLLEGE OF ENGINEERING AND COMPUTER SCIENCE

NSF