An Analysis of Course Review Data at Brown

Min Jean Cho, James Ding, Brendan Le, John Zhou

INTRODUCTION

Introduction

Course reviews have been a tradition at Brown for many years, with the Critical Review. We wanted use this data to investigate whether an increase in the number of hours spent on a course resulted in an increase in the expected grade.

However, an initial correlation analysis revealed that there was an unintuitive negative correlation between hours spent and grade.

Hypothesis

If students are able to correctly rate the difficulty of a course, we should find that, **after controlling for difficulty** of a course, students who spend more time on the class get better grades. This is what we expect to find.

If students simply rate a course's difficulty using the grade they receive, we would not expect to find such a relationship.

METHODOLOGY

Data Source

The course data was scraped from The Critical Review at https://thecriticalreview.org

Preprocessing

Records without relevant features were removed, as well as invalid entries such as "a lot of time" in hours spent. This resulted in 2774 course records, and 41065 individual responses.

Data Usage

- In each feature in each course, the average student value was used
- A, B, C, NC were converted to 4, 3, 2, 0, and S was not used
- Students of all years were summed to get class size
- Courses were weighted by class size
- Correlation coefficients were estimated using Pearson sample correlation coefficient
- Significance level was set at 0.05 for a two-tailed test
- Calculations/visualizations created with Python
- Outlier were removed. Correlation hypothesis tested with both All-A classes included and not included.

MOTIVATIONS

Student-run class ratings are a fairly contentious topic. Brown has had The Critical Review for decades, but in many other institutions such rating systems have just been established or are nonexistent. In some cases, they have even been shut down.

The primary downside to these student reviews is that they may be biased by personal performance or otherwise not reflect the course correctly.

Our study could lend support to the idea that these student reviews contain valuable information, as long as students are able to rate hours and difficulty on a consistent scale

RESULTS

Variable pair	With all-A courses (n = 2606)		Without all-A courses (n = 2179)	
	Correlation coefficient	p-value (two-tailed)	Correlation coefficient	p-value (two-tailed)
minhours and grade	**-0.29**	< 0.01	**-0.30**	< 0.01
difficult and minhours	+0.69	< 0.01	+0.69	< 0.01
difficult and grade	-0.56	< 0.01	-0.57	< 0.01
Variable pair	Partial correlation coefficient	p-value (two-tailed)	Partial correlation coefficient	p-value (two-tailed)
minhours and grade	**+0.16**	< 0.01	**+0.16**	< 0.01

Table 1: Correlation analysis using outlier-removed dataset. Correlation coefficients between minhours, grade, and difficulty, and partial correlation between minhours and grade (while controlling for difficulty) are shown. p-values are based on t-distributions

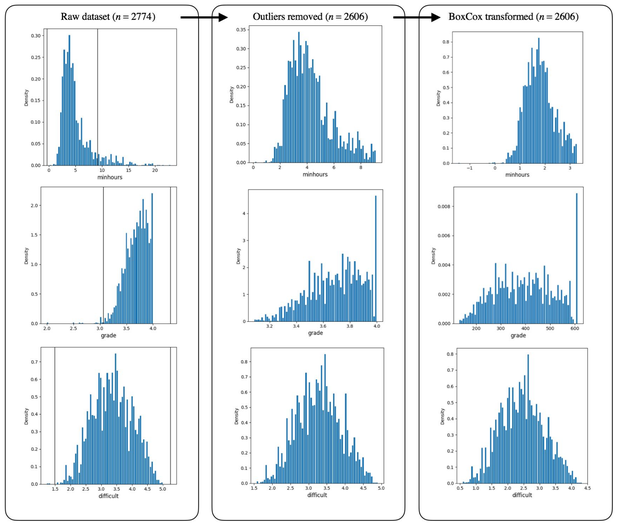

Figure 2: Histogram representation of distribution of raw, outlier-removed, and Box Cox transformed (after outliers removed) datasets. Class sizes were used for weighting each data point. Vertical lines for raw dataset indicate boundaries for outliers

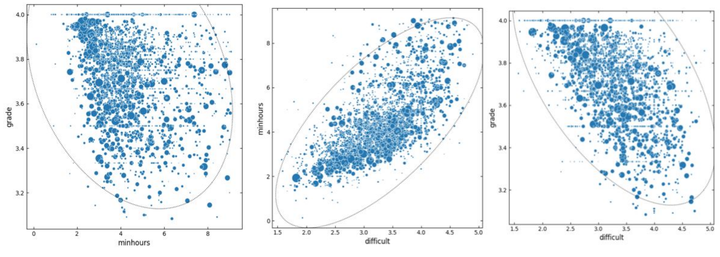

Figure 4: Scatter plot representations of pairwise relationships between minhours, grade, and difficulty, determined using outlier-removed dataset. All-A courses are included in the dataset. Sizes of markers are scaled relative to the class size. Dotted line indicates confidence ellipse (the radiuses of the ellipse was calculated using 3X standard deviation, which makes the ellipse enclose ca. 99.7% of the points

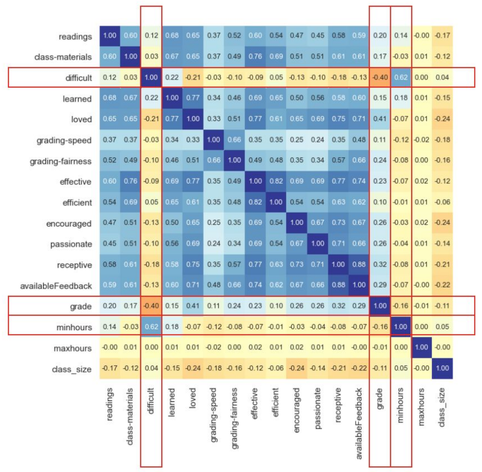

Figure 1: Correlation analysis of survey responses (n=2774). Each label represents a field on the survey. Heat map is based on simple correlation coefficients (Pearson r) calculated from raw data. Blue: positive correlation, red: negative correlation.

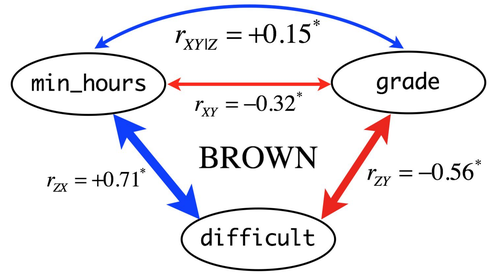

Figure 3: Schematic representation of the relationships between minhours, grade, and difficulty. r_{xy} and $r_{xy|z}$ indicate simple correlation and partial correlation coefficients respectively. Correlation coefficients are based on outlier-removed dataset with all-A courses. Asterisks indicate significant ($p < 0.05$) correlation between variables. Width of arrow lines (positive correlation: blue, negative correlation: red) were scaled approximately relative to correlation coefficients.

Independent Variable	Coefficient	95% Confidence Interval	p-value of T test
Number of hours	0.003	[0.002, 0.004]	<0.001
Difficulty	-0.1236	[-0.128, -0.199]	<0.001
Class Enjoyment	-0.1006	[0.094, 0.107]	<0.001
Concentrator	0.0725	[0.065, 0.081]	<0.001
Constant	3.5138	[3.47, 3.557]	<0.001

Table 2: Multiple regression performed on the survey responses, with the dependent variable being the GPA, and independent variables being all of the fields on the Critical Review survey.

DISCUSSION

As shown in Table 1, there is a statistically significant correlation between minhours and grade after controlling for difficulty

This is evidence students are to some degree able to rate difficulty independently of the grade they received. In particular, this relationship indicates that students who rate a course a given difficulty perform better with more hours put into the course. However, since this partial correlation does not show the size of the relationship, we then did multiple regression.

Linear Regression

An initial linear regression to predict grade received using hours studied yielded a statistically significant coefficient of -0.0325 on hours studied. This meant that every hour of additional study per week resulted in on average a decrease in the GPA of 0.0325.

We ran another test, controlling for other factors such as difficulty and class enjoyment by adding them as independent variables in a multiple regression. This test resulted in a statistically significant and positive coefficient on number of hours spent. **For every increase of 1 hour per week studying, we saw an average increase in GPA of 0.003, after controlling for all other factors on the survey.**

This result is interesting as while it further supports our hypothesis that an increase in the number of hours studied corresponds to an average increase in the GPA when controlling for other features, it also indicates that without controls there is a very large negative correlation between hours spent and grade received, more than 10 times the size of the controlled.

This is no paradox, however; without controlling variables, more hours spent generally indicates more effort, difficulty, and thus a lower grade overall.

Future Work

Our results indicated a definite positive relationship between minhours and grade controlling for difficulty. However, a low r-squared means that these features do not account for much of the variation in grades.

While students are able to rate course difficulty on a consistent scale, there are other factors that play a large role in determining grade, difficulty, and minhours.

This is likely due in part to the limitations of The Critical Review dataset. The surveys for The Critical Review are optional and there could exist significant response bias. Without having access to official grades of any class we cannot make any concrete comparisons, but the average grade reported on the Critical Review was high, with a GPA 3.72.

A superior source of information would be the course evaluations internal to Brown which strongly incentivize student completion and are less likely to have sampling bias. In any case, further data collection would be required to make more concrete statements on students' ability to rate classes.