

# CSCI 1951a cs-number-1 Abstract

ble2

May 8, 2020

## Hypothesis

We wanted to investigate whether an increase in the number of hours spent on a course resulted in an increase in the expected grade.

At first, we noticed that the number of hours a student spent on a course was negatively correlated with their expected grade. This doesn't make intuitive sense, since one would expect to see a higher grade with more hours of study. We suspected that a third variable, *difficulty*, which was positively correlated with hours and negatively correlated with grade, could explain this relationship.

## Data

We scraped data from The Critical Review at <https://thecriticalreview.org/>. The resulting data contained a list of student reviews for each class offering at Brown. After preprocessing the data, we ended up with a table where each row represented a single review, and the columns represented various fields including the `course_offering_id`, `minhours`, `course_difficulty`, `grade`, `course_rating`, among many others. The fields that required the student to provide a rating such as `course_difficulty` were represented as a score on a five point likert scale. If a student did not provide a valid response for a field which we wanted to analyze, we would remove their entire response. The grade of a student was represented as  $A = 4.0$ ,  $B = 3.0$ ,  $C = 2.0$ . Responses which received other grades such as  $S$  were not considered. After cleaning the data removing responses from older versions of the Critical Review, we ended up with 41065 responses.

## Findings

### Claim #1:

The minimum study time ( $X$ : minhours) is positively correlated with the grade received ( $Y$ : grade) after controlling for difficulty level ( $Z$ : difficult). To test this hypothesis, we performed a statistical test on the partial correlation coefficient:

$$H_0 : \rho_{XY|Z} = 0$$

$$H_A : \rho_{XY|Z} \neq 0$$

### Support for Claim #1:

The correlation between minhours and grade was -0.29. But after controlling for difficulty, the correlation becomes 0.16, with a significant p-value. This shows that for a given difficulty of class, spending more hours results in a better grade on average, which is what we would expect.

The following table shows the various correlation coefficients, before and after controlling for difficulty.

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)
<b>minhours and grade</b>	<b>-0.29</b>	< 0.01	<b>-0.30</b>	< 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		Partial correlation coefficient	
	p-value (two-tailed)		p-value (two-tailed)	
<b>minhours and grade</b>	<b>+0.16</b>	< 0.01	<b>+0.16</b>	< 0.01

## Claim #2

An increase in number of hours spent corresponds to an increase in expected grade, after controlling for multiple variables, including difficulty of a course.

We represented the relationship between minimum study time (X: minhours), expected grade received (Y: grade), and other control variables (Z) with the following linear model:

$$\hat{Y} = \hat{\beta}_X X + \hat{\beta}_Z Z + \hat{\beta}_0$$

$$H_0 : \beta_X = 0$$

$$H_A : \beta_X \neq 0$$

## Support for Claim #2

Performing a t-test on the coefficients of the multiple regression showed that every increase of 1 hour spent on a course per week on average resulted in a 0.003 increase in expected grade ( $p < 0.001$ ), after controlling for many variables. Below are the coefficients and p-values for hours spent and other variables with large significant effects on grade.

Variable	Coefficient	95% confidence interval	p value
minhours	0.003	[0.002, 0.004]	<0.001
difficult	-0.1236	[-0.128, -0.119]	<0.001
loved	0.1006	[0.094, 0.107]	<0.001
concentrator	0.0725	[0.064, 0.081]	<0.001
constant	3.5138	[3.47, 3.557]	<0.001

In addition to these factors, the regression controlled for these other variables with small but still significant effects on grade: readings, class-materials, amount learned, grading-speed, grading-fairness, recommended to non-concentrators, efficient use of class time, encouraged discussion, passionate professor, receptive to student needs, available feedback, and STEM. It is important to note that concentrator and STEM are dummy variables that indicate if the student was a concentrator in the department or if the class was a STEM class respectively.