

Tuition to Test Scores: A Statistical Analysis

ABSTRACT:

We performed an observational study to answer the question “Does the amount of tuition personally financed by students have an association with their average GPA?” Data was collected through responses to an anonymous survey sent out to college students. The observational units were students who responded to the survey, and the population of interest was all students at a particular college. Our null hypothesis was that there was no association between GPA and amount of tuition personally financed. The alternative hypothesis was that there was an association between student GPA and tuition personally financed. Overall we found that students in our survey who paid less of their tuition had higher GPAs on average (p value .021).

INTRODUCTION:

An observational study was performed on students to answer the research question, "Does the amount of tuition personally financed by students have an association with their average GPA?". The explanatory variable for this study was the amount of tuition that students paid for personally, a categorical variable. The response variable was the average GPA of students, a quantitative variable. We chose a significance level of .05 to measure the significance of our data.

METHODS:

Data was collected by creating an anonymous, online survey and sending it out to students. The questions of interest on the survey were "What is your GPA?" and "How much of your college education are you personally financing (outside of family assistance and scholarships?" We also added some extra questions to the survey to reduce bias in our results. The four categories of the amount of tuition personally financed by students were as follows: Zero; where they do not pay for any of their tuition, Small; where students pay 1%-25% of their own tuition, Medium; where students pay for 26%-99% and All; where students pay for 100% of their tuition. The variable of students' GPA was on a 4 point scale. Students input the average value of the grades they have received at college up until the survey was done. The results were analyzed using the multiple means applet. We used a simulation-based approach on shuffled data to analyze the Mean Absolute Difference (MAD) statistic from our collected data.

RESULTS:

Our results were as follows. For the zero group, the sample size was 14 with an average GPA of 3.49. For the small group, there was a sample size of 11 and an average GPA of 3.69. For the medium group the sample size was 4 students with an average GPA of 3.72. For the all group there was a sample size of 4 with a median GPA of 2.9. The Mean Absolute Difference statistic of the collected data was .44. We presented our results using a clustered bar graph created in excel.

The 95% confidence intervals of the differences in means between the groups were as follows: zero and all groups (.0638, 1.1084), the small and all groups (.2451, 1.328), the medium and all groups (.01636, 1.4664), zero and small group (-.5681, .1743), zero and medium groups (-.7512, .2934) and small and medium groups (-.5699, .5058). We ran a simulation of 1000 shuffles, which represented taking our data and assuming the null was true with no difference between the groups. The simulation represents taking 33 cards with all the GPA-values, separating them randomly into the 4 categories, and then finding a MAD statistic for this "fake" data. The p value was .021 which represents the probability that if the null was true that we would find the same results that we did with our sample data.



APPLET OUTPUT:

Sample data:

(explanatory,response) Unstacked

Amount tuition paid

Amount tuition paid	GPA
zero	2.6
small	4.00
small	3.7
small	3.4
medium	3.45
all	3.95
zero	3.48
small	3.2

Use Data Clear

Boxplots

Summary Statistics:

	n	Mean	SD
all	4	2.90	0.95
medi	4	3.72	0.20
smal	11	3.49	0.27
zero	14	3.49	0.42
pooled	33	3.51	0.45

Observed MAD=0.440

Statistic: MAD

Show ANOVA Table:

Source	df	SS	MS	F	p-value
Treatment	3	1.99	0.66	3.275	0.0351
Error	29	5.88	0.20		

95% CI(s) for difference in means

zero - small	(-0.5681, 0.1743)
zero - medium	(-0.7512, 0.2934)
zero - all	(0.0638, 1.1084)*
small - medium	(-0.5699, 0.5058)
small - all	(0.2451, 1.3208)*
medium - all	(0.1636, 1.4664)*

Show Shuffle Options:

Number of Shuffles: 1000

Hypothesized $\mu_2 - \mu_1$: 0

Shuffle Responses Data Plot

Most Recent Shuffle

Amount	tuition
zero	3.20
small	3.80
small	3.66
small	3.45
medium	3
all	3.28
zero	3.71
small	2.90
zero	3.70

Shuffled Summary Statistics:

	n	Mean	SD
all	4	3.32	0.52
medi	4	3.42	0.29
smal	11	3.68	0.31
zero	14	3.46	0.44
overall	33	3.51	0.50

Shuffled MAD=0.19

Most Recent Shuffled ANOVA Table

Source	df	SS	MS	F	p-value
Treatment	3	0.54	0.18	0.72	0.5493

Total Shuffles = 1000

Mean=0.220
SD=0.092

Count Samples: Greater Than \geq .44 Count

Count = 21/1000 (0.0210)

Discussion:

We reject our null hypothesis that there is no association between student GPA and the amount of their tuition that they personally finance. Our p-value was significant at .021, which is below the significance value of .05. In addition, there was a **significant difference** in means in confidence intervals between three groups: zero and all groups (.0638, 1.1084), the small and all groups (.2451, 1.328) and the medium and all groups (.01636, 1.4664). We are 95% confident that the average difference in GPA between pairs of these groups was contained in these intervals. For example, we are 95% confident that the difference in average GPA between students who paid none of their tuition vs those who paid all of it was between .0638 and 1.1084.

Generalization is not warranted because our sample was a convenience sample and not representative of the entire college's student body (there is potential for sampling bias). We also cannot claim causation because we did not randomly assign students to finance a certain amount of their college tuition, meaning there is potential for confounding variables. A possible confounding variable is students' major, because if students are paying for their degree they would be more likely to choose a more difficult major with more job security in the future. In addition, whether students work during the school year could also be a confounding variable because students who have to pay for their own tuition might work more often and therefore have less time to do their homework.

Our study could have been improved by obtaining a larger sample and collecting data from a random, more representative sample. If we had greater time and resources, we would have used them to better our sampling method in order to legitimize the study.