

A Major Analysis of Twicers

Maddy Cosgriff, Evan Olawsky, Thomas Redding

Introduction

Every spring, Carleton's sophomore class must undergo the arduous process of selecting a major. Well-rounded liberal arts students are often very torn between multiple interesting and impactful areas of study. Faced with such a difficult decision, many students simply choose not to choose by becoming double majors. In this paper, we examine the distribution of double majors across the different academic areas. Specifically, we compare the decisions of single majors with those of double majors.

Methodology

For all Carleton graduates from the classes of 2011 through 2014, data was collected on their major or majors. All data was drawn from the Carleton Alumni Directory. For the purposes of the study, majors fall into one of four academic areas – Arts, Humanities, Qualitative, and Sciences. See the Appendix for a list of which majors were sorted into each academic area. Students were then sorted into three groups. The first group is composed of students with only one major. The second group (referred to as “twicers”) is composed of students who had two majors in the same academic area. The third group includes students who had two majors in different academic areas. The latter group was thrown out of the data set used for our analysis.

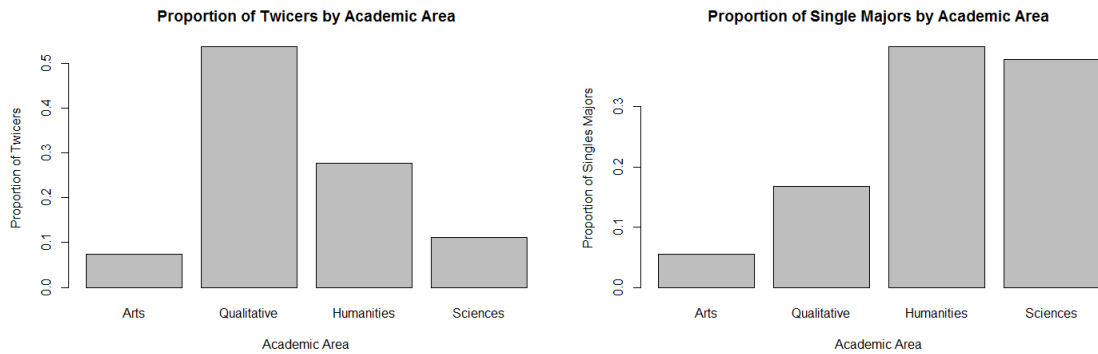
Using this data, we ask whether single majors and twicers behave similarly in choosing their major(s). Our null hypothesis is that the distribution of single majors across all academic areas is the same as the corresponding distribution of twicers, and our alternative is that it is not the same.

Results

First, we performed some explanatory data analysis to get a grasp of the data. It is clear that twicers are much more likely to major in qualitative majors (Economics, Mathematics, Statistics, and Computer Science) than single majors are. On the other hand, twicers are much less likely to choose a science major than single majors are.

Counts	Arts	Qualitative	Humanities	Sciences	Total
Singles	97	292	695	658	1742
Twicers	4	29	15	6	54

Proportions	Arts	Qualitative	Humanities	Sciences	Total
Singles	5.6%	16.8%	39.9%	37.8%	100%
Twicers	7.4%	53.7%	27.8%	11.1%	100%



We performed a two-sample permutation test to determine whether this difference in distributions is significant or not. The hypotheses are as follows:

H_0 : The distribution of twicers across the academic areas is the same as the distribution of single majors across the academic areas.

H_A : The distribution of twicers across the academic areas is not the same as the distribution of single majors

To test these hypotheses, we used a two-sample permutation test. The observed sample of twicers, when compared to observed sample of single majors, yielded a chi-square goodness-of-fit statistic of 50.75.

Single majors and twicers were then pooled and 54 majors were randomly chosen. A chi-square goodness-of-fit statistic was then calculated for this group of 54 against the distribution of single majors. In 9,999,999 resamples, no chi-square statistic was as large as the observed statistic. This test yields a p-value of only 1 in 10 million (0.0000001). We reject the null hypothesis and conclude that the distribution of twicers across the academic areas is not the same as the distribution of single majors.

Discussion

Our results indicate that an abnormal amount of the total twicers are Qualitative twicers, much more than we would expect if they followed the singles distribution. In the Sciences, it is relatively more likely to find single majors than twicers. A similar but less dramatic discrepancy was observed in the Humanities. The Arts show a fairly consistent proportion of both singles and twicers out of the total sample.

Although we cannot conclusively draw inferences, the fact that our study encompasses all graduates in the past 4 years allows us to reasonably predict similar results for the next few years of single major and twicer rates. We chose these years to analyze because they contain the same majors that are offered today, whereas those offered 5 and more years ago are not identical. We are limited to making inferences to times that offer the same majors as those in our study. Another limitation is that these data only reflect Carleton graduates, and may or may not be similar to other institutions.

We ran into trouble sorting a portion of graduates into double versus single majors because there are some majors that could also be concentrations, but there is no indicator of whether they were majors or concentrations in our data. This led us to throw out data that could not be differentiated into the two groups with certainty (see “Unknown” in Appendix).

We initially intended to perform a chi-square goodness-of-fit test on the data, but multiple expected counts were below 5, forcing us to change our design. In sorting our data we found that there was only one Language twicer, so we combined Languages and Humanities because we felt that the datum may not accurately reflect the proportion of Language twicers.

R Code

```
permutationPopulation <- c(datum$V1, data$V1) # combine the twicers and singles

N <- 9999999 # select the number of permutations to do
expectedArts <- length(which(permutationPopulation == 1))/length(permutationPopulation)*54 #
compute expected number of Arts twicers
expectedComputing <- length(which(permutationPopulation == 2))/length(permutationPopulation)*54 #
compute expected number of Computing twicers
expectedHumanities <- length(which(permutationPopulation == 3))/length(permutationPopulation)*54
# compute expected number of Humanities twicers
expectedSciences <- length(which(permutationPopulation == 4))/length(permutationPopulation)*54 #
compute expected number of Sciences twicers

observedArts <- length(which(data$V1 == "Arts")) # compute observed Art twicers
observedComputing <- length(which(data$V1 == "Computing")) # compute observed Computing twicers
observedHumanities <- length(which(data$V1 == "Humanities")) # compute observed Humanities
twicers
observedSciences <- length(which(data$V1 == "Sciences")) # compute observed Sciences twicers

# Compute observed Chi-Squared statistic
observedChisq <- (observedHumanities-expectedHumanities)^2/expectedHumanities + (observedArts-
expectedArts)^2/expectedArts + (observedComputing-expectedComputing)^2/expectedComputing +
(observedSciences-expectedSciences)^2/expectedSciences

# start counter at 1 to represent the observed data's extremity
counter <- 1

for(i in c(1:N)) {
  # permute the cmbined population
  permutedSample <- sample(permutationPopulation, 48, replace = FALSE)

  # note these hypothetical observed twicers
  observedArts <- length(which(permutedSample == 1))
  observedComputing <- length(which(permutedSample == 2))
  observedHumanities <- length(which(permutedSample == 3))
  observedSciences <- length(which(permutedSample == 4))
}
```

```

# compute the Chi-Squared statistic
chisq <- (observedHumanities-expectedHumanities)^2/expectedHumanities + (observedArts-
expectedArts)^2/expectedArts + (observedComputing-expectedComputing)^2/expectedComputing +
(observedSciences-expectedSciences)^2/expectedSciences

# if the result is more extreme increment "counter"
if(chisq >= observedChisq) {
  counter <- counter+1
}
}

# print results
counter/(N+1)

```

R Output

```
[1] 1e-06
```

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
≥0.1	

“If all else fails, use ‘significant at a p>0.05 level’ and hope no one notices”

List of Values	Area	Type
African-African American Studies	Humanities	Unknown
American Studies	Humanities	Major
Applied Logic	Humanities	Concentration
Archaeology	Sciences	Concentration
Art History	Arts	Major
Asian Studies	Humanities	Major
Biochemistry	Sciences	Concentration
Biology	Sciences	Major
Chemistry	Sciences	Major
Chinese	Languages	Major
Cinema and Media Studies	Humanities	Major
Classical Languages	Languages	Major
Classics - Carleton 4-track	Humanities	Major
Cognitive Science	Sciences	Concentration
Computer Science	Computing	Major
Cross-Cultural Studies	Humanities	Concentration
Dance	Arts	Major
East Asian Studies	Humanities	Concentration
Economics	Computing	Major
Educational Studies	Humanities	Concentration
English	Humanities	Major
Environmental Studies	Sciences	Major
European Studies	Humanities	Concentration
French and Francophone Studies	Languages	Unknown
Geology	Sciences	Major
German	Languages	Major
History	Humanities	Major
International Development	Humanities	Concentration
Islamic Studies	Humanities	Concentration
Japanese	Languages	Major
Latin American Studies	Humanities	Unknown
Linguistics	Languages	Major
Mathematics	Computing	Major
Mathematics-Statistics	Computing	Major
Medieval & Renaissance Studies	Humanities	Concentration
Music	Arts	Major
Neuroscience	Sciences	Concentration
Philosophy	Humanities	Major
Physics	Sciences	Major
Political Economy	Humanities	Concentration

Political Science	Humanities	Major
Psychology	Sciences	Major
Religion	Humanities	Major
Romance Languages & Literature	Languages	Concentration
Russian	Languages	Major
Self, Community & Interpretation	Humanities	Concentration
Sociology-Anthropology	Humanities	Major
South Asian Studies	Humanities	Concentration
Spanish	Languages	Major
Special Major	Unknown	Concentration
Studio Art	Arts	Major
Theater Arts	Arts	Unknown
Women's and Gender Studies	Humanities	Unknown

