Testing Response Bias in Agree/Disagree Questions

Jordan Zoellmer, Branden McGarrity, Thomas Redding
Carleton College, Professor Katie St. Clair
Mathematics 255: Introduction to Sampling Techniques
November 14, 2015

I. ABSTRACT

This study tests the effectiveness of agree/disagree style survey questions and the impact that question wording has on response accuracy by sampling the population of students attending Carleton College. One survey was distributed with statements phrased as a comparison of two mutually exclusive alternatives, thereby presenting a distinct opinion and asking for an answer of agree or disagree. A second, alternative survey was distributed to a different sample of students with essentially the same questions, but with the comparisons flipped, such that the opinions expressed in this alternative survey were logically opposite to those in the initial survey. This study tests the null hypothesis that the alternative survey would have an opposite proportion of agree and disagree answers, consistent with the fact that the survey statements were logically opposite. Findings for the majority of questions indicate that the proportion of agrees in one survey significantly differs from the expected proportion of disagrees in another survey. Furthermore, results indicate a tendency for respondents to disagree regardless of the question wording, perhaps resulting from the absence of a neutral option.

II. INTRODUCTION

Many statistical surveys of people attempt to measure an individual's character, personality, or attitude. A common form of survey question that can help measure such responses involves making statements with which a respondent must agree or disagree. Pioneering research on such surveys was done by psychologist Rensis Likert in 1932[1]. His "Likert Scale" is in common use today, and involves responses of the form "Strongly Agree/Agree/Neutral/Disagree/Strongly Disagree", which can easily be interpreted mathematically using his 5-point methodology[2]. Some evidence suggests that questions with agree and disagree style options should be replaced entirely with item specific response options, eliminating the need for an opinionated statement to precede the responses[3], but sacrificing the convenient and streamlined Agree/Disagree format that is so pervasive in the sampling world today.

In addition to the 5-point Likert style of Agree/Disagree, some surveys employ a simple binary response system, forcing respondents to either "Agree" or "Disagree" with an opinionated statement. This study attempts to explore the accuracy of such a binary system, and the susceptibility a system like that may have to the phrasing of statements. Such a survey does not provide an exhaustive response list, forcing respondents to have an opinion. The analysis that follows ultimately calls into question the reliability of the binary response system. The results provide evidence that the binary system may introduce some bias with the wording of statements, and that the absence of a complete response list likely leads to inaccurate results. Studies such as this are necessary to provide concrete guidance as statisticians attempt to perfect the art of questionnaire design.

The population in this study is students at Carleton College, a convenient population from which to sample. Two samples were selected, and each was given a different version of questionnaire, with five questions asked under the guise of an "Educational Survey". The questions were formulated as "Agree/Disagree". Both versions essentially asked the same questions, with wording flipped such that a response of agree to one survey would correspond to a response of disagree in the alternative survey. As we will see in the upcoming analysis, the responses did not flip in the way expected from the survey design, but the results will provide valuable insight into the accuracy and flaws associated with such a binary response system.

III. METHODS

We sampled the population of Carleton students using a stratified design, separating the students into four groups based on graduating class year. While it is not the case that every student graduates in four years, we figured that almost every current student would identify with one of the next four graduating classes, and therefore fall into one of our four strata.

After developing two survey versions, nearly identical but for flipping the opinionated statement order, we issued each type to half of the sampled population in each class. Anticipating a 50% response rate and desiring a margin of error of 10% or less, we calculated that we needed to sample 200 students from each class - 100 for each version-class pairing.

The response rate for our survey was in reality much lower than anticipated, even after sending a reminder email. Overall, 23.5% of students responded to version A of the survey and 31.25% responded to version B of the survey. The response rate differed for the different classes and for the different surveys, meaning the sample sizes within each of our strata were not quite equal. This inequality was not an issue in our analysis.

The surveys were completed in October of 2015. The surveys asked each student to answer five agree/disagree questions along with a few demographic questions. The demographic questions allowed us to see if different groups of people responded differently to the questions, while the agree/disagree questions were our main point of focus. In each question of the survey, we asked the student to compare two things. Is one thing more important than another? Is one thing more valuable than another? In the two versions of the survey, we flipped the wording of each question, changing which concept we presented first. For example, in one survey the statement is I learn more from working with a group than from working individually, and in the other survey the statement is, I learn more from working individually than working with a group. For each survey, we measured the responses (proportion of agree and proportion of disagree) for each question. We expected that the number of students who agreed with a given question for the first survey would be the same as the number of students who disagreed with the corresponding question in the other survey. For each of the five questions we tested this hypothesis using the survey data.

We assigned a value of 1 to a response of agree and a value of 0 to a response of disagree. For each question we examined the mean response for both surveys. We compared the mean response for one survey with 1 minus the mean response for the other survey. We conducted t-tests to determine whether a difference in the means in fact existed.

IV. RESULTS

We begin with relevant preliminary exploratory data analysis, summarizing some of the collected data, including responses to the five questions of interest and the categorical questions, exploring the demographics of respondents. Then we display results of hypothesis testing, determining if the proportion of agrees in version A varies significantly from the proportion of disagrees in version B, thereby testing the null hypothesis that the two versions should yield the same amount of support for the ideas presented in the questions. Afterwards, we conduct a brief analysis at the non-meta level by analyzing how major interacts with individuals' responses.

A. Exploratory Data Analysis

Overall, 94 out of 400 individuals responded to version A, a response rate of 23.5%. For version B, 125 out of 400 individuals responded, yielding a response rate of 31.25%. We collected data on the proportion of agrees and disagrees for each of the five questions in both survey version A and survey version B. These proportions are displayed in Table I. Keep in mind the null hypothesis is that the proportion of agrees matches the proportion of disagrees between versions.

	Question 1		Question 2		Question 3		Question 4		Question 5	
	A	B	A	B	A	B	A	B	A	B
% Agree	0.670	0.152	0.181	0.224	0.553	0.288	0.585	0.416	0.457	0.408
% Disagree	0.330	0.848	0.819	0.776	0.447	0.712	0.415	0.584	0.543	0.592

TABLE I. Agree and disagree proportions of responses categorized by question and survey version.

The information in the table can also be displayed graphically. Figure 1 displays the proportion of agree responses in version A and the proportion of disagree responses in version B. Doing so allows us to directly visualize the testing of our null hypothesis, where bars of equal height agree with the null while bars of unequal height tend to disagree with the null hypothesis. For example, we notice a large proportion of disagrees in question 2 for both versions of the survey.

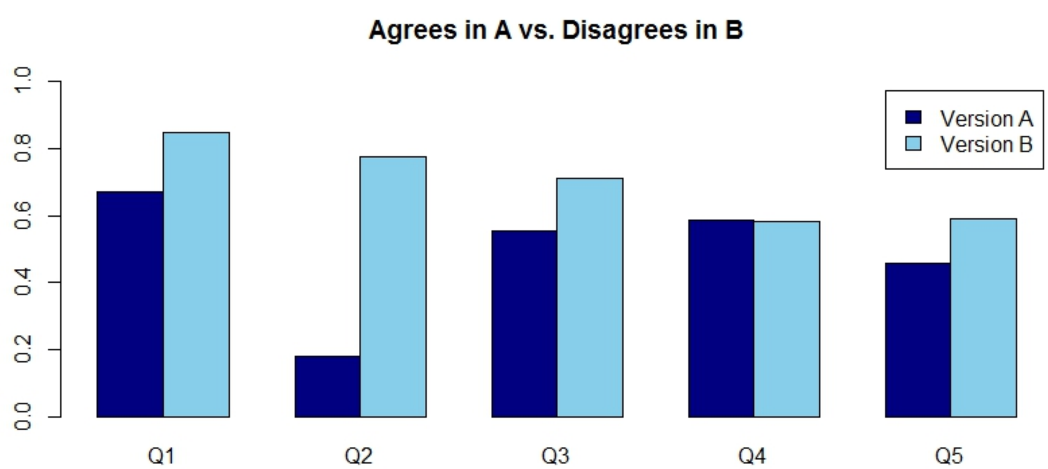

FIG. 1. Displayed here are the proportion of agree responses in version A and the proportion of disagree responses in version B.

In addition to responses to the five questions of interest, we analyze the demographic distributions, including the class years and sex of respondents. Although this information is not directly pertinent to our hypothesis of interest, our survey is voluntary, and it is important to analyze the respondents for potential confounding variables and bias. This demographic information is displayed for class year in Figure 2 and for sex in Figure 3.

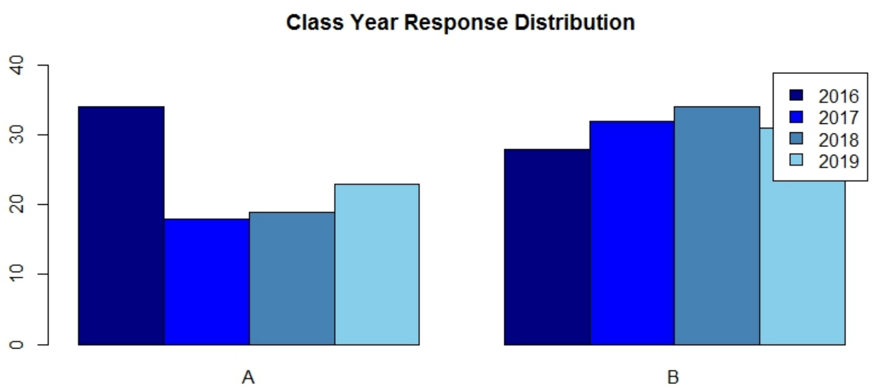

FIG. 2. Class year responses.

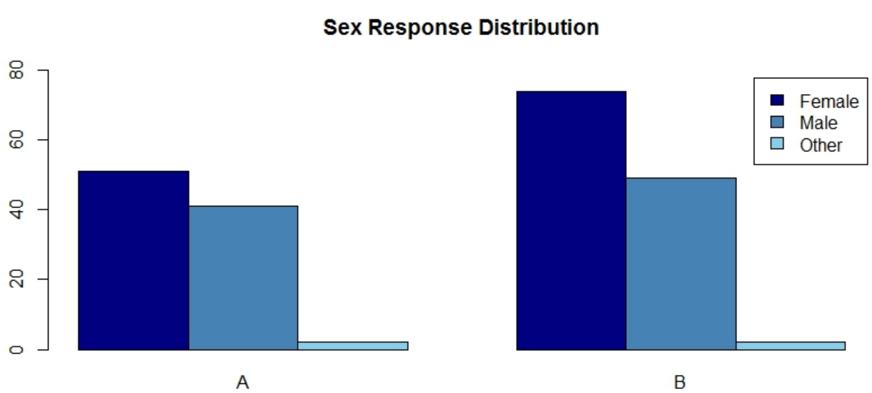

FIG. 3. Responses by sex.

Finally, Table II displays the distribution of responses by area of study (major). Again, this information may not be pertinent but is relevant to check for confounding variables. The distributions are, in general, very similar.

	Humanities	Math/Science	Music/Arts	Social Science	Undecided	Other	TOTAL
Version A	12	52	3	17	4	6	94
Version B	31	54	6	21	9	4	125

TABLE II. The distribution of responses by area of study/major.

B. Significance Testing

The main results of interest test the hypothesis that support for an idea should be equivalent regardless of the version of question asked. Therefore, the proportion of agrees in one version should equal the proportion of disagrees in the other. We proceed to test this hypothesis for each of the five questions, using a t-test with a significance level of 0.05.

Question 1: We find that the difference in proportions (between agree in A and disagree in B) is 0.182, leading to a p-value of 0.0011. We reject the null hypothesis and conclude that there did not exist equivalent support across versions A and B for the ideas in question 1 on going to college to get a job vs. going to college to learn. The overall proportion of agrees was only 37.4%, indicating that respondents tended to disagree more than agree. In other words, the proportion of disagrees in version B was more dramatic than the proportion of agrees in version A.

Question 2: We find that the difference in proportions is 0.588, leading to a p-value of essentially zero. We reject the null hypothesis and conclude that there is not equivalent support across versions for the ideas in question 2 on STEM fields vs. humanities. The overall proportion of agrees was only 20.5%, indicating that individuals tended to disagree more than agree.

Question 3: We find that the difference in proportions is 0.159, leading to a p-value of 0.012. We therefore reject the null hypothesis and conclude that there is not equivalent support across versions for the ideas in question 3 on discussion in class vs. lecture. The overall proportion of agrees was 40.2%, indicating again that individuals tended to disagree more than agree.

Question 4: We find that the difference in proportions is only 0.003, leading to a p-value of 0.960, making it evident that the proportions are very much equal. In this case, the proportion of agrees in version A is equal to the proportion of disagrees in version B meaning that equivalent support between versions was expressed for the ideas in question 4 on learning in a group vs. learning individually. The overall proportion of agrees was 49%, lending no evidence toward a conclusion that people tend to agree or disagree more for this question.

Question 5: We find a difference in means of 0.119, leading to a p-value of 0.067, forcing us to conclude that there is little (perhaps moderate) evidence that the proportions are different. So there is little evidence that the support is different between versions for the ideas expressed in question 5 on learning in class vs. learning outside of class. The overall proportion of agrees was 42.9%, indicating perhaps that individuals tend to disagree more as a whole for this question.

We also analyzed whether one sex was more likely to agree in general than another, regardless of which version they filled out. We found that question 2 had a statistically significant difference at the 5% level (p value of 2.5%), where males were more likely to agree than females. However, given that (1) we analyzed 5 questions and (2) that the data for this question was remarkable in other ways (i.e. a high level of disagreement), this is not particulary compelling evidence that a true difference in agreeableness exists.

C. Object Level Results

By averaging the answers given to both versions of each question, we can also analyze the object-level questions pertaining to students' educational experiences. Although slightly tangential from our primary purpose, it is none-the-less interesting, giving us persepective into the actual opinions of Carleton students on this topic. The object level responses are displayed graphically in Figure 4.

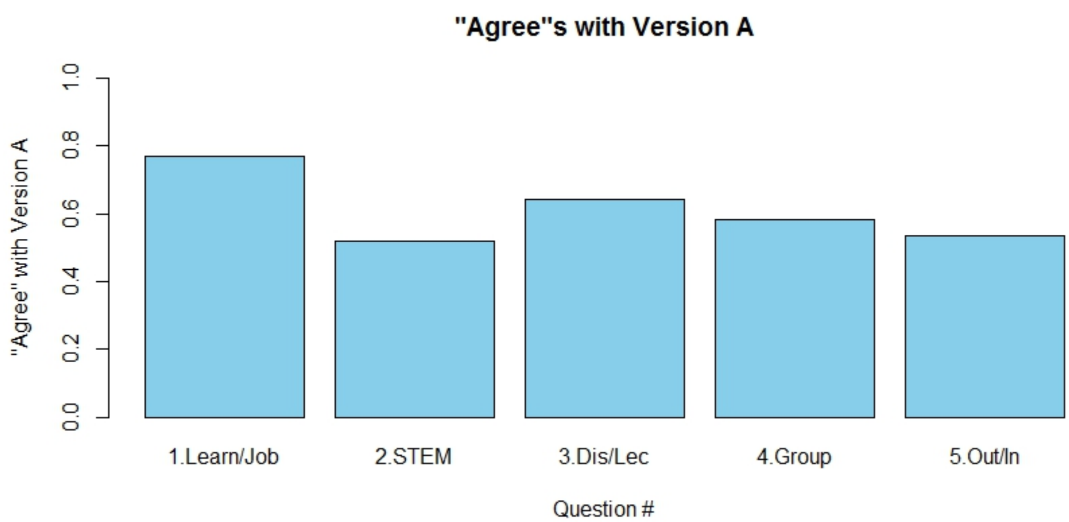

FIG. 4. The proportion of students who would agree with the statements presented in Version A.

Our null hypothesis in each case was that there was equal agreement to both sides of each question and the each answer was an independent Bernoulli random variable. We tested this hypothesis at the 5% significance level for each question, the results of which are listed below.

Question 1: We rejected the null hypothesis and concluded that we have evidence at the 95% confidence level for the claim that a majority of students at Carleton College believe going to college to learn is more important than going to college to get a job.

Question 2: We failed to reject the null-hypothesis, indicating that we have no evidence at the 95% confidence level against the hypothesis that an equal number of students show preference for STEM and non-STEM fields.

Question 3: We rejected the null hypothesis and concluded that we have evidence at the 95% confidence level for the claim that a majority of students at Carleton College believe that discussion is more valuable than lecture in class.

Question 4: We rejected the null hypothesis and concluded that we have evidence at the 95% confidence level for the claim that students at Carleton College believe they learn more form group-work than individual-work.

Question 5: We failed to reject the null-hypothesis, indicating that we have no evidence at the 95% confidence level against the hypothesis that an equal number of students believe they learn more in class than out of class.

Dividing the population into STEM-majors and non-STEM-majors reveals potential differences between these groups' beliefs. Again, by question, we have

Question 2: With 95% confidence, we found that STEM majors are more likely than non-STEM majors to believe that STEM fields are more important than the humanities.

Question 3: With 95% confidence, we found that STEM majors are more likely than non-STEM majors to believe that lecture was more valuable to their learning than inclass-discussion.

Questions 1, 4, and 5: We found no evidence at the 95% confidence level that STEM majors and non-STEM majors differened in their beliefs with regard to the relative importance of learning versus getting a job in going to college, the amount they learn from group versus individual work, and the degree to which most of their learning is done in versus outside of class.

D. Assumptions Made

Our primary assumption is that whether a student responded or not was independent of any of the parameters we wanted to study. To use more statistical terminology, we assume the responding students were a representative sample.

We did take steps to mitigate the problems posed by this assumption by stratifying by class year and using sampling weighing to give classes with lower response rates greater weight per respondent.

The main other assumption, if we want to infer to the broader population of college students, is that the Carleton student body is representative of college students in general.

V. DISCUSSION

We set out to determine the extent to which people respond agree and disagree in binary style questions, and if respondents tend to answer in one way more frequently than another. We found statistically significant evidence that respondents tended to disagree with questions more than agree in Questions 1, 2, and 3. Question 5 lent moderate support for this conclusion, while Question 4 demonstrated essentially no evidence either way.

Taken together, it seems likely that people tend to disagree more often with such binary questions. We believe a likely cause of this phenomenon is that when people tend to disagree with both alternatives, they choose "disagree" to indicate a lack of support towards the first option rather than agreement with the second. In any case, this provides strong evidence that surveys with binary options need to be careful to account for this negative bias.

One way to measure the strength of this negative bias is by adding up the proportion of disagrees in A to the proportion of disagrees in B. A naive view would hold that these should add up 1. If this sum is greater than 1, then it would appear that people disagree more than a "fair" question would cause. Figure 5 graphically displays this bias, demonstrating that most questions have a higher proportion of disagrees. For example, about 80% of students disagreed with both versions of question 2, meaning that the sum is about 60% greater than the assumed sum of 1. This indicates that respondents are disagreeing regardless of version, and Figure 5 shows this is the case for almost every question.

On the object-level, we found that a majority of Carleton College students believe that learning is more important than getting a job when going to college. Furthermore, we found that students claimed to learn more in groups than individually and found discussion to be more valuable than lecture. Interestingly, we found that this discussion-versus-lecture preference was moderated among students majoring in STEM fields. Unsurprisingly, we also found that STEM majors are more likely than non-STEM majors to believe that that STEM fields should be emphasized more than the humanities.

Overall, our results are not all that surprising. It is generally considered good practice to offer a neutral option on questionnaires, and the 5-point Likert scale is an even better method, giving a better range of opinions. It is not surprising that individuals would tend to have a neutral opinion, and would respond with disagree if not given a neutral option.

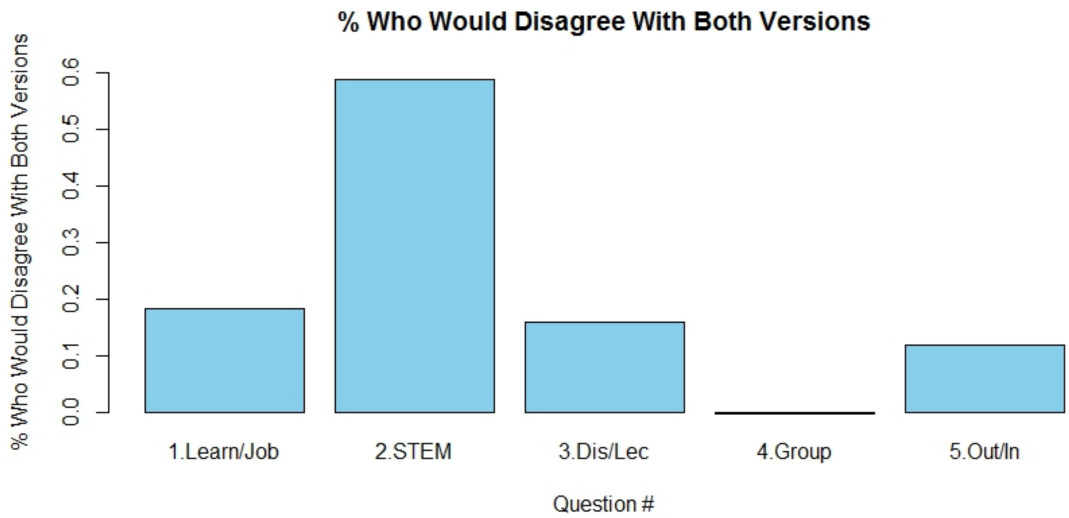

FIG. 5. The proportion of students who would disagree regardless of the version presented.

After all, they do not truly agree with the statement presented. However, this study raises quite a few questions about the exact effects of a neutral option. Further research could certainly compare questionnaires with neutral options to those without. This would perhaps confirm our suspicion that those with neutral opinions tend to answer disagree. More than anything, this study calls into question the validity of binary agree/disagree style surveys and gives researchers reason to be cautious of any results obtained from such a survey.

The implementation of the survey was fairly straightforward, and Survey Monkey made it easy to distribute and obtain the information in a reasonable manner. However, our proposed sample size and estimated response rate were not very accurate; the response rate was much less than predicted. Even so, we sampled a very large portion of the population and ended up with significant results despite these issues.

VI. CONCLUSION

This report investigated characteristics of agree/disagree style surveys, particularly using the binary response system with no neutral option. Two surveys were distributed expressing the same ideas but with alternative wording such that support for an idea in version A should be met with an answer of agree while support for that same idea in version B should be met with an answer of disagree. Our results indicate that such support was not equally distributed between surveys. As a whole, respondents to binary agree/disagree style questions tended to disagree far more than agree. This phenomenon most likely stems from the lack of a neutral option, prompting respondents to disagree when they feel that the implicit neutral option is the correct choice. These findings imply that there most likely exists some inherent bias in this style of question. Further research into the style of questions that exhibit this bias more than others would be helpful. However, researches should all be wary that such bias could exist when no neutral option is present and should not incorporate binary agree/disagree style questions into surveys without such considerations.

VII. APPENDIX I

When asked to provide their gender, 4 people chose "Other (please specify)". One of them elaborated "Decline to identify" while another wrote "Agender". The other two did not provide more information.

When asked to provide their major, 9 people chose "Other" and elaborated:

- "Environmental Studies: Science & Humanities"
- "Interdisciplinary? (Ents. Could count as science for some)"
- "Math and Music"
- "STEM and social science"
- "Linguistics"
- "Philosophy (both quantitative and humanities)"
- "History"
- "Geology"
- "Linguistics (The math of language)"

VIII. APPENDIX II

We wrote two versions of our survey:

Version A

1. Do you agree or disagree with following statement: Going to college to learn is more important than going to college to get a job.
2. Do you agree or disagree with the following statement: STEM fields should be emphasized more than the humanities.
3. Do you agree or disagree with the following statement: Discussion in class is generally more valuable to my learning than lecture.
4. Do you agree or disagree with the following statement: I learn more from working with a group than from working individually.
5. Do you agree or disagree with the following statement: the learning I do outside of class is generally more valuable than the learning I do in class.
6. What is your class year?
7. Which category best fits your intended or declared primary major?
8. What is your gender?

Version B

1. Do you agree or disagree with following statement: Going to college to get a job is more important than going to college to get a learn.
2. Do you agree or disagree with the following statement: The humanities should be emphasized more than the STEM fields.
3. Do you agree or disagree with the following statement: Lecture in class is generally more valuable to my learning than discussion.
4. Do you agree or disagree with the following statement: I learn more from working individually than working with a group.
5. Do you agree or disagree with the following statement: the learning I do in class is generally more valuable than the learning I do outside of class.
6. What is your class year?
7. Which category best fits your intended or declared primary major?
8. What is your gender?

IX. R Code

```
# Intro to Sampling Techniques
# Final Project
# JZ, TR, BM
# R Script

# Load data and survey library
# 1=Agree, 0=Disagree, regardless of version
data <- read.csv("~/Dropbox/Survey Sampling Main Project/data.csv")

# 1=Agree for vA & Disagree for vB, 0=Disagree for vA & Agree for vB
# Used for running ttest against our null
DataAlt = read.csv("~/Dropbox/Survey Sampling Main Project/DataAlt.csv")
library(survey)

# Setting Up Stratified Design
nrow(data)
N <- rep(NA, 219)
N[data$Class.Year == 2016] = 495
N[data$Class.Year == 2017] = 514
N[data$Class.Year == 2018] = 527
N[data$Class.Year == 2019] = 500
```

```
n <- rep(NA, 219)
n[data$Class.Year == 2016] = length(which(data$Class.Year == 2016))
n[data$Class.Year == 2017] = length(which(data$Class.Year == 2017))
n[data$Class.Year == 2018] = length(which(data$Class.Year == 2018))
n[data$Class.Year == 2019] = length(which(data$Class.Year == 2019))
wts <- N/n # 2016: 7.98, 2017: 10.28, 2018: 9.94, 2019: 9.26
design.strat <- svydesign(id=~1, fpc=N, weights=wts, strata=~Class.Year, data=
    data)
summary(design.strat)

# Alternate design same, different data
design.stratAlt = svydesign(id=~1, fpc=N, weights=wts, strata=~Class.Year, data=
    DataAlt)
summary(design.stratAlt)


# Analysis of Responses
# Question 1
mean(data$Q1)
svyby(~Q1, ~Version, design.strat, svymean, vartype = c("se", "ci"))
svyby(~Q1, ~Version, design.stratAlt, svymean, vartype = c("se", "ci"))
# Notice flipped means of vB in above summaries
# We want to test null that flipped prop are equal
svyttest(Q1 ~ Version, design.stratAlt) # stat. sig

# Question 2 (same process as above)
mean(data$Q2)
svyby(~Q2, ~Version, design.strat, svymean, vartype = c("se", "ci"))
svyby(~Q2, ~Version, design.stratAlt, svymean, vartype = c("se", "ci"))
# clearly lots of disagrees regardless of version
svyttest(Q2 ~ Version, design.stratAlt) # stat. sig

# Question 3(same process as above)
mean(data$Q3)
svyby(~Q3, ~Version, design.strat, svymean, vartype = c("se", "ci"))
svyby(~Q3, ~Version, design.stratAlt, svymean, vartype = c("se", "ci"))
svyttest(Q3 ~ Version, design.stratAlt) # stat. sig
```

```
# Question 4 (same process as above)
mean(data$Q4)
svyby(~Q4, ~Version, design.strat, svymean, vartype = c("se", "ci"))
svyby(~Q4, ~Version, design.stratAlt, svymean, vartype = c("se", "ci"))
# This seems like flipped means are equalish
svyttest(Q4 ~ Version, design.stratAlt) # not stat. sig

# Question 5 (same process as above)
mean(data$Q5)
svyby(~Q5, ~Version, design.strat, svymean, vartype = c("se", "ci"))
svyby(~Q5, ~Version, design.stratAlt, svymean, vartype = c("se", "ci"))
svyttest(Q5 ~ Version, design.stratAlt) # borderline but not stat. sig

# Questions by sex
svyttest(Q1 ~ Sex, design.strat)
svyttest(Q2 ~ Sex, design.strat)
svyttest(Q3 ~ Sex, design.strat)
svyttest(Q4 ~ Sex, design.strat)
svyttest(Q5 ~ Sex, design.strat)

# Jordan's EDA
# Barplot of class year
table(data$Class.Year, data$Version)
barplot(table(data$Class.Year, data$Version), beside = T,
       main="Class Year Response Distribution", col=c("navy","blue",
       "steelblue","skyblue"), ylim=c(0,40), legend=c("2016","2017",
       "2018","2019"))

# Barplot of sex
table(data$Sex, data$Version)
barplot(table(data$Sex, data$Version), beside = T,
       main="Sex Response Distribution", col=c("navy","steelblue","skyblue"),
       ylim=c(0,80), legend=c("Female","Male","Other"))

# Major Distribution
table(data$Major, data$Version)

# Barplot 1
```

```
M = matrix(c(mean(data$Q1[data$Version=="A"]),mean(data$Q1[data$Version=="B"]),
       mean(data$Q2[data$Version=="A"]),mean(data$Q2[data$Version=="B"]),
         mean(data$Q3[data$Version=="A"]),mean(data$Q3[data$Version=="B"]),
           mean(data$Q4[data$Version=="A"]),mean(data$Q4[data$Version=="B"]),
             mean(data$Q5[data$Version=="A"]),mean(data$Q5[data$Version=="B"])),
                 nr=2)
colnames(M) = c("Q1","Q2","Q3","Q4","Q5")
rownames(M) = c("A", "B")
M
barplot(M, beside=T, main="Percentage of Agrees by Question and Version",
       col=c("navy", "skyblue"),ylim=c(0,1), legend=c("Version A", "Version B"))

# Barplot 2 (Alternate)
MAlt = matrix(c(mean(DataAlt$Q1[DataAlt$Version=="A"]),mean(DataAlt$Q1[
    DataAlt$Version=="B"]),
            mean(DataAlt$Q2[DataAlt$Version=="A"]),mean(DataAlt$Q2[
                DataAlt$Version=="B"]),
            mean(DataAlt$Q3[DataAlt$Version=="A"]),mean(DataAlt$Q3[
                DataAlt$Version=="B"]),
            mean(DataAlt$Q4[DataAlt$Version=="A"]),mean(DataAlt$Q4[
                DataAlt$Version=="B"]),
            mean(DataAlt$Q5[DataAlt$Version=="A"]),mean(DataAlt$Q5[
                DataAlt$Version=="B"])), nr=2)
colnames(MAlt) = c("Q1","Q2","Q3","Q4","Q5")
rownames(MAlt) = c("A", "B")
MAlt
barplot(MAlt, beside=T, main="Agrees in A vs. Disagrees in B",
       col=c("navy", "skyblue"), ylim=c(0,1), legend=c("Version A", "Version B"))

# Means & sd by Version (Put these in a table?)
tapply(data$Q1, data$Version, mean)
tapply(data$Q2, data$Version, mean)
tapply(data$Q3, data$Version, mean)
tapply(data$Q4, data$Version, mean)
tapply(data$Q5, data$Version, mean)
tapply(data$Q1, data$Version, sd)
tapply(data$Q2, data$Version, sd)
tapply(data$Q3, data$Version, sd)
```

```
tapply(data$Q4, data$Version, sd)
tapply(data$Q5, data$Version, sd)

# Means by Version Alt
tapply(DataAlt$Q1, DataAlt$Version, mean)
tapply(DataAlt$Q2, DataAlt$Version, mean)
tapply(DataAlt$Q3, DataAlt$Version, mean)
tapply(DataAlt$Q4, DataAlt$Version, mean)
tapply(DataAlt$Q5, DataAlt$Version, mean)



# EDA
summary(data$Major)
summary(data$Sex)
hist(data$Class.Year, c(2016,2016.99,2017.99,2018.99,2019.99))
plot(data$Sex)

# "Non-Meta" Results
data$Ver <- ifelse(data$Version == "A", 1, -1)
data$Q1.adj <- ifelse(data$Q1 == 1, 1, -1) * data$Ver
data$Q2.adj <- ifelse(data$Q2 == 1, 1, -1) * data$Ver
data$Q3.adj <- ifelse(data$Q3 == 1, 1, -1) * data$Ver
data$Q4.adj <- ifelse(data$Q4 == 1, 1, -1) * data$Ver
data$Q5.adj <- ifelse(data$Q5 == 1, 1, -1) * data$Ver
data$Q1.adj <- ifelse(data$Q1.adj == 1, 1, 0)
data$Q2.adj <- ifelse(data$Q2.adj == 1, 1, 0)
data$Q3.adj <- ifelse(data$Q3.adj == 1, 1, 0)
data$Q4.adj <- ifelse(data$Q4.adj == 1, 1, 0)
data$Q5.adj <- ifelse(data$Q5.adj == 1, 1, 0)
df <- length(data$Major)-1
qt(0.975, df)
svymean(data$Q1.adj, design.strat)
svymean(data$Q2.adj, design.strat)
svymean(data$Q3.adj, design.strat)
svymean(data$Q4.adj, design.strat)
svymean(data$Q5.adj, design.strat)
svyratio(ifelse(data$Major == "Math, Science, CS", data$Q1.adj, 0), ifelse(
```

```
   data$Major == "Math, Science, CS", 1, 0), design.strat)
svyratio(ifelse(data$Major != "Math, Science, CS", data$Q1.adj, 0), ifelse(
   data$Major != "Math, Science, CS", 1, 0), design.strat)
(0.7222541 - 0.8208767)/sqrt(0.0412369^2 + 0.03431432^2)
svyratio(ifelse(data$Major == "Math, Science, CS", data$Q2.adj, 0), ifelse(
   data$Major == "Math, Science, CS", 1, 0), design.strat)
svyratio(ifelse(data$Major != "Math, Science, CS", data$Q2.adj, 0), ifelse(
   data$Major != "Math, Science, CS", 1, 0), design.strat)
(0.6226387 - 0.4271712)/sqrt(0.04410017^2 + 0.04430372^2)
svyratio(ifelse(data$Major == "Math, Science, CS", data$Q3.adj, 0), ifelse(
   data$Major == "Math, Science, CS", 1, 0), design.strat)
svyratio(ifelse(data$Major != "Math, Science, CS", data$Q3.adj, 0), ifelse(
   data$Major != "Math, Science, CS", 1, 0), design.strat)
(0.5043733 - 0.7786505)/sqrt(0.04533446^2 + 0.03739934^2)
svyratio(ifelse(data$Major == "Math, Science, CS", data$Q4.adj, 0), ifelse(
   data$Major == "Math, Science, CS", 1, 0), design.strat)
svyratio(ifelse(data$Major != "Math, Science, CS", data$Q4.adj, 0), ifelse(
   data$Major != "Math, Science, CS", 1, 0), design.strat)
(0.5562172 - 0.6242508)/sqrt(0.04566804^2 + 0.04339797^2)
svyratio(ifelse(data$Major == "Math, Science, CS", data$Q5.adj, 0), ifelse(
   data$Major == "Math, Science, CS", 1, 0), design.strat)
svyratio(ifelse(data$Major != "Math, Science, CS", data$Q5.adj, 0), ifelse(
   data$Major != "Math, Science, CS", 1, 0), design.strat)
(0.5250691 - 0.550078)/sqrt(0.04627317^2+0.0448263^2)


# "Meta" Analysis
Q1.est <- svyby(~Q1, ~Version, design.strat, svymean)
Q2.est <- svyby(~Q2, ~Version, design.strat, svymean)
Q3.est <- svyby(~Q3, ~Version, design.strat, svymean)
Q4.est <- svyby(~Q4, ~Version, design.strat, svymean)
Q5.est <- svyby(~Q5, ~Version, design.strat, svymean)

# Graph for Poster
disagreeWithBothTheory <- c(1 - (0.6677372 + 0.1504173), 1 - (0.1818114 +
   0.2296405), 1 - (0.5535361 + 0.2878656), 1 - (0.5930444 + 0.4102032), 1 -
   (0.4693410 + 0.4120988))
names(disagreeWithBothTheory) <- c("1.Learn/Job","2.STEM","3.Dis/Lec","4.Group
```

```
    ","5.Out/In")
barplot(disagreeWithBothTheory, col="skyblue", ylim=c(0,0.6), xlab="Question #",
    ylab="% Who Would Disagree With Both Versions", main="% Who Would Disagree
    With Both Versions")

avgs <- c(mean(data$Q1.adj), mean(data$Q2.adj), mean(data$Q3.adj), mean(data$Q4.
    adj), mean(data$Q5.adj))
names(avgs) <- c("1.Learn/Job","2.STEM","3.Dis/Lec","4.Group","5.Out/In")
barplot(avgs, xlab="Question #", ylim=c(0,1), col="skyblue", ylab="\"Agree\" with
     Version A", main="\"Agree\"s with Version A")
```

[1] R. Likert, "A Technique for the Measurement of Attitudes", Archives of Psychology **140**, 1932.

[2] H. Boone et al., "Analyzing Likert Data", Journal of Extension **50** (2), 2012.

[3] W. E. Saris et al., "Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options", Survey Research Methods, **4** (1), 61-79, 2010