36-303
Oliver Lam
Mike Len
Go Okumura
Rita Wang
Wentian Zhu

I.5: Sampling Scheme and Question Design

K. Decide on a sampling scheme (e.g., SRS, Stratified random sample, etc.) and explain why you chose it.

We have decided that sending out questionnaires from the C-book is the most unbiased and error-free procedure in conducting this survey. Right now our concern is if we are able to collect data from each individual class level. Because we feel that the questions asked in the survey are only adequate to persons in grades sophomore and up, it is important to exclude freshmen samples. If the C-book does list the students' email addresses along with their class level, we will use stratified random sample without replacement as our sampling scheme. We will assume that simple random sample will distribute the population evenly in each grade automatically making this plan a stratified one.

If there is a possibility in which C-book does not indicate the grade level of the students, we will overcome this issue by asking participants to indicate their grade levels and excluding responses from freshmen respondents. This does decrease our effective sample size, but we were not conservative in increasing sample size to account for non-response and believe this should cover the potential unused observations of freshmen. With use of google-documents this is easily done.

By using stratified random sampling without replacement we are able to separate our sample population by grade level to isolate effects that may be present due to school or major experience. It also allows us to later fix our sample's representativeness if our initial sample does not represent our target population well. That is to say, if our initial sample turns out to have a non-representative percentage of sophomores, we can sample more sophomores. By using sampling without replacement (as opposed to sampling with replacement) we prevent any double counting of responses, which could skew our results negatively.

Although we would like to rely on a respondent naturally not taking the survey twice to ensure sampling without replacement, if we go down this route we cannot say for sure if each respondent only took the survey once. Alternatively, we can give the respondent the chance to provide us with their Andrew ID so that we do not contact them when we send out follow-up emails, which will ensure to a greater degree that our final sample was taken using sampling without replacement.

all you would have to do is strike out respondents from the first sample that appear in the second sample, before contacting people in the second sample (if you decide you need a second sample).

If any issues come up after the survey regarding response rate such as higher response rate in one of the grade level than the others, we will adjust the result by weighing accordingly.

L. Write a questionnaire with 20–30 questions. Up to approximately 1/3 of these can be background or demographic questions and the rest should be directly related to the research questions you will try to answer with your survey. NOTE: If your survey involves observations instead of asking people questions, then instead you should carefully describe your observation protocol. That is, list 20–30 things you will always look for when observing each unit in your survey.

Look at some of the articles you collected: what sorts of demographics seem to matter? Are you collecting that info?

What is your gender?
What year are you? (multiple choice)
What school are you in? (multiple choice)
What is your GPA Range (multiple choice, intervals of .5)?

might just ask for GPA!

If Undeclared
What majors are you considering? What attracts you to these majors?

If Declared
What's your current major?
Do you like your major?
What led you to choose your current major?
Would you pursue the field of your major in future professional career?

These are very open ended and will be difficult to code and summarize for a report. Try some closed or closed plus blank formats to see if you can get the same information (and have an easier time coding).

If you feel you need to stick with these, fine, but it will take time at the back end when you are coding and analyzing results!

same question. why ask twice?

Have you changed majors?

If yes:
From what major to what major?
Why did you choose your original major?
What factors influenced your choice?
What year did you change your major?

If No:
Have you considered changing your major?
Please indicate your level of satisfaction with your current major from 1-10 (1-strongly dislike, 10-strongly like).

some "whys" would be useful here

If Considered this isn't a possible response above - how will respondents get to this block?
What major were you thinking about switching to?

What encouraged you to change?
Why didn't you change?
Are you thinking about doubling or minoring in a second major?

again more open-ended. think of ways of making these questions more codable!

If Not Considered
What kept you interested in this major?
What possible changes in your current major will lead you to reconsider?

What would you suggest the University to do in the future regarding major declaring/changing policies?

If you do not wish to be contacted with further emails reminding non-respondents to take our survey, please provide us with your Andrew ID so we can place you on our do-not-contact list.

M. Give some idea of the sample size you will require and how you arrived at this number (talk about the margin of error for inferences you want to make)

Considering that the total undergraduate population of CMU is 5619, assuming that students are evenly distributed across the four grade levels, the sample target will be 75% of this population (sophomores, juniors, seniors) which is 4214 students.

Simple random survey without replacement
$n_0 = (z^2)(SD^2)/(ME^2) = (1.96^2)(.5^2)/(.05^2) = 384.16$
$n > (N*n_0)/(N + n_0)$; $n > (4214 * 384)/(4214 + 384) = 352.06$

According to our calculations, assuming a standard deviation of ½ with 95% confidence interval and 5% margin or error, we would need to sample at least 353 students. We decided on restricting the margin to 5% because this is recommended in the context of the book, and here our inference about the target population with 5% margin will be convincing enough.

Assuming we use as our medium of distribution, which has a maximum 20% response rate, we would need our sample size to be 353*5=1765 students in order to account for this low percentage. Just to simplify we choose our sample size to be 1800. This sample size seems large, but with our source of C-book and computer sending out email to 1800 students would just take a moderate amount of effort.

Note:

For your sake it is better to invite respondents to go to a site like surveymonkey or google forms to fill out the survey, than to put the survey right in the email.

The reason is that most of the data will be coded for you automatically then, rather than spending time transcribing responses from emails to excel files or something of the sort.

-BJ