Assignment II.6. Project Plan

Spatial and Analytical Study of Student Housing at Carnegie Mellon

Ariel Liu, Sam Lavery, Alejandra Munoz, Terra Mack, Shannon Lauricella

A. TOPIC

Carnegie Mellon is an urban university with many students living off-campus. Finding housing off-campus is generally left up to individual students, who take into account many variables when choosing a house or apartment. Many students list their off-campus addresses in the C-Book directory published by Alpha Phi Omega (APhiO). We are interested in investigating the possibility of a correlation between where students choose to live and what they choose to study. The results of the survey will be a valuable tool that would be useful to the university for the planning of shuttle routes, campus police coverage, and future housing projects. Students would also be able to use the survey results to find neighborhoods in the city that are popular with other students like themselves. We are seeking to answer questions about the dynamics of student housing at CMU. An example of this is: Is there a correlation between address (either on-campus building or off-campus neighborhood) and major? Do students in certain majors cluster together?

B. Questions of the study

We are seeking to answer questions about the dynamics of student housing at CMU. Some questions that we want to investigate include:

- Is there a correlation between address (either on-campus building or off-campus neighborhood) and major? Do students in certain majors cluster together?
- What is the distribution of CMU students by neighborhood?
- What off-campus areas are most popular with undergraduates? With graduate students?
- What percent of off-campus students live within X miles of a shuttle or bus stop.
- What is the average commute distance for undergraduates? For graduate students?

C. Research

1. Title: The Causal Effect of Campus Residency on College Student Retention
   Author name(s): Lauren T. Schuddle
   Date: Summer 2011
   Source: *The Review of Higher Education*, Volume 34, No. 4, pp. 581-610
   http://muse.jhu.edu/journals/review_of_higher_education/v034/34.4.schudde.html
   Summary: Assessing demographics and certain characteristics among those college students who live on and off campus, particularly how campus residency affects retention of students. Our survey could drawn on some of the insight for possible survey measures and execution in looking at campus residency and characteristics among CMU's student populaton.
   Team member: Terra Mack

2. Title: Campus Housing Construction and Renovation
   Author name(s): James C. Grimm, Norbert W. Dunkel
   Date: June 1999
   Source: Full-length book published by *The Association of College and University Housing*
   Summary: This book investigates the relationship between college students and their physical environments by specifically looking at 42 projects in 36 different universities.

Housing characteristics studied in the book could be incorporated into the questions asked in our survey.
Team member: Shannon Lauricella

3. Title: Housing: A Financial Look
Author name(s): Michael R. Fitzgerald
Date: September 25, 2006
Source: CMU's *The Tartan*
http://thetartan.org/2006/9/25/news/housing_realestate
Summary: This article, published in Carnegie Mellon's own newspaper, specifically highlights some of the points we want to survey students on-campus. It talks about the price differential in living on-campus versus off-campus as well as the nature of housing prices in close proximity to Carnegie Mellon. Most importantly, this article shows that CMU students have an interest in their housing options, making our suggested survey more relevant.
Team member: Alejandra Munoz Munoz

4. Title: A Quasi-Experimental Approach to Estimating the Impact of Collegiate Housing
Author name(s): Ryan Yeung
Date: 2010
Source: http://student.maxwell.syr.edu/ryyeung/college.doc
Summary: Yeung attempts to identify the relationship between on/off-campus housing with GPA and social and academic integration. This is relevant to studying how areas of academic study play into college housing and proximity to campus.
Team member: Ariel Liu

5. Title: The Disengaged Commuter Student: Fact or Fiction?
Author name(s): George D. Kuh, Robert M. Gonyea, Megan Palmer
Date: 2001
Source:http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.9974&rep=rep1&type=pdf
Summary: This paper present the findings of a survey that wanted to answer if students that live off-campus are less engaged with studies compared to students that live on campus. Engagement is defined as an important part of the academic experience and it is believed that off-campus housing interferes with this process.
Team member: Sam Lavery

D. What is the sampling frame? What population or populations do you plan to sample from? (This is the question many tend to miss).

The sampling frames comes from the administrative records of the university registrar, John Papinchak. We have been in communication with John Papinchak who has agreed to provide the administrative records as long as we maintain the confidentiality of students' information. We plan on looking at housing information from undergraduate and graduate students enrolled at Carnegie Mellon.

E. What is the target population? To what population(s) do you wish to make inferences?

The target population is undergraduate and graduate students enrolled at CMU. It is the same population that we are looking to make inferences about from our survey. The target population differs from the sampling frame in that the registrar can only provide a sample of students

enrolled at CMU. Thus, we will not have access to information for the entire population enrolled. Therefore, our sampling frame will include only those students who comply with the registrar's office or volunteered their information to CMU.

What possible sampling and non-sampling errors could arise in the survey that you plan to conduct? Explain each possible error, how it could occur, and how you suggest tackling it.

This survey could encounter coverage error because the registrar's records are incomplete. The target population coverage depends on the completeness of the registrar office records. When a student leaves on-campus housing, they are asked to update their address on SIO but many probably fail to do so. Additionally, some people may change addresses again and not update this information. One solution to this problem would be to find the ratio of current students living in on-campus housing and weight our sample to account for any discrepancies. We could easily find the correct ratio by dividing the number of students living in dorms by the total student body.

F. What is the mode of data collection? How do you plan to carry out the survey (e.g., by telephone, e-mail) and why?

We plan to collect the data from administrative records provided by the office of the registrar. We believe surveying data records is a more accurate and reliable method in comparison to directly asking students. This mode of collection and survey can help reduce high non-response and coverage errors.

G. VARIABLES
We want to study the following variables:
School (MCS, CIT, HSS, etc)
Major
Location (address)
Class Year / level (graduate, undergraduate)

I.J. IRB/CONFIDENTIALITY OF RESPONDENTS
This is not applicable as we are not sampling human subjects.

K. SAMPLING SCHEME
We were successfully able to attain off-campus housing records from the University registrar. The records have 891 undergraduate records and 4,036 graduate records. The registrar provided us with all the records that they had. According to the CMU Factbook, (found at http://www.cmu.edu/ira/factbook/pdf/facts2012/11_campus-space-section-final.pdf), there are 2,252 undergraduates living off-campus and 5,769 graduates living off-campus. Clearly, the ratio of undergraduate records to graduate records is not the same as the population ratio, but there could be response errors that affect undergraduates more than graduate students. Most undergraduates start their CMU careers living on-campus so changing their address to an off-campus location will probably be less likely reported to the registrar (especially if they still use their SMC mailboxes to get mail from the university). Other sources of bias in the collection of data could also be the limit of one major and one department per student. When looking at clusters of students off-campus according to their major, a student could have more than one major, but the records only indicate one major and one affiliated department per student. Another possible bias is that students may not have reported accurate addresses of zip codes, such as using abbreviations or interchangeable zip codes. We will have to sort through the data

to locate any of these issues, as part of the data cleaning process.

Given that we have obtained all of the records from the registrar for students living off-campus that provided responses, we think that including all of the records we have (after cleaning the data, there will definitely be fewer records for graduate students) would be most appropriate for our analysis. We are going to treat our records as a stratified sample of the possible off-campus residents. As part of cleaning the data, we noticed that graduate students have a duplicate entry for their offices, therefore, we will have to make sure we only report their residences in our results. Other issues we will need to consider when cleaning the data are duplicate records, response missingness, and incorrect forms of address format. As an exercise, we looked at 438 records of the list. We found the following problems: one duplicated record (same person listed twice with different addresses), five records without college data, one record with a PO Box as an address, one record with no address, and 12 records with CMU offices address (all of the students that listed CMU offices addresses also reported off-campus housing). Thus, from 438 records, 24 presented problems to be included in the sample (5.4%). Minor issues were found with addresses' format in 10 records, but all information is present to correctly locate the addresses.

We think that students' decisions of where to live vary based on their needs and expectations. While graduate students might search for quiet places, closer to groceries stores, larger apartments/houses and more affordable rentals to accommodate living with a family, undergraduate students might search for more active places, closer to restaurants and smaller in size because a higher percent of them are singles. We will use a weighting scheme to take into account the differences in the needs of undergraduate versus graduate students.

L. QUESTIONNAIRE OR OBSERVATIONAL PROTOCOL:

Undergraduate/Graduate
Is the person an undergraduate student?
Is the person a graduate (Master) student?
Is the person a graduate (PhD) student?
College
Is the person a member of Marianna Brown Dietrich College of Humanities and Social Sciences (DC) (ex -HSS)?
Which department?
Is the person a member of Carnegie Institute of Technology (CIT)?
Which department?
Is the person a member of David A. Tepper School of Business (TSB)?
Which department?
Is the person a member of School of Computer Science (SCS)?
Which department?
Is the person a member of College of Fine Arts (CFA)?
Which department?
Is the person a member of H. John Heinz III College at Carnegie Mellon University (HC)?
Which department?
Is the person a member of Mellon College of Science (MCS)?
Which department?
Housing
Residence address
Type of Building - a house? An apartment? Number of stories *
Neighborhood

City
Distances
Distance to Campus
Time to campus by foot
Time to campus by driving
Time to campus by public transportation
Are students from some colleges or majors more likely to live off campus?
Do undergrads/grads, colleges or majors cluster together? If so, where?
We plan to further develop this question in looking at each college (HSS, Tepper, SCS, etc) and then majors within each college. This question will ultimately have many different results.
* This information is obtained from Department of city Planning of Pittsburgh - GIS database

M. SAMPLE SIZE:

To calculate the sample size we select the following question: Is this person a member of CIT?. Then, We used the source of the factbook from February 2012 which provided a head count of students in each college in the Fall Semester 2011 (only for Pittsburgh, PA campus) to calculate a value for p.
The total head count of students: 10,957
The head count for CIT students: 3,217
The proportion of CIT students out of total students: p : 3,217/10,957 = .293 = 29.3%
Total population size of students living off-campus: 2,252+5,769 = 8,021
p = .293
n = 8021
z/2= 1.96
SD = sqrt (p(1-p)) = sqrt (.293(1-.293)) = .4551
The first that we considered was 0.05
ME = z/2SDn = 0.05
With this value we calculate n for a SRS with replacement.
1.96 (.4551/n) = 0.05 ---> n = 318

Because the sample size is small we tried smaller MOE values.
Second ME = 0.001
1.96 (.4551/n) = 0.01 ---> n = 7,957
Third ME= 0.012
1.96 (.4551/n) = 0.012 ---> n = 5,525
Fourth ME = 0.011
1.96 (.4596/n) = 0.011 ---> n = 7,800

We estimated the sample size using a MOE of 0.05 as our first reference point. Then we decided that our sample of data available to us was much larger (in our database given by the registrar) and we could use a lower MOE. Thus, we estimated the sample size again to match our sample size available to us using an MOE of 0.01. However this sample was larger and closer to the sampling frame size we have available. We also took into account that we must perform data-cleaning and this may leave us a lower number of records. Therefore, we estimated the sample size using a slightly larger MOE of 0.0012 to obtain a sample size of 5,525 which seems more reasonable for the data available to us. Table 1 contains the different MOE values used and the n values obtained.

Table 1. MOE selected and n values obtained for defining a sample size

MOE	sqrt(n)=(1.96*SD)/MOE	n
0.010	90.08	7957
0.011	89.19	7800
0.012	75.07	5525
0.015	60.05	3607
0.050	18.02	325

Then we calculated the adjustment needed for sizing a sample without replacement:

Adjustment calculation for SRS without replacement:
$n >= (N*n_0) / (N+n_0) = (10,957)(7,955) / (10,957+7,955) = 3,672.9$

Seeing the size of the data obtained, we reconsidered the MOE and n used for the SRS without replacement. In this case we used a n value of 7957.

$n >= (N*n_0) / (N+n_0) = (10,957)(7,955) / (10,957+7,955) = 4,609$

This results gave us a larger size of the sample. Because we have a larger sampling frame, we think a n size of 4,609 is reasonable. However, after cleaning the data we could use a more accurate MOE value and sample size number..

Reference:
Carnegie Mellon University Factbook Volume 26. Headcount Enrollment by Location of Study, Home College, Level, and Status. Fall Semester 2011. Office of Institutional Research and Analysis. http://www.cmu.edu/ira/factbook/pdf/facts2012/entire-fb-for-web-as-of-3-1-121.pdf