100/100

see comments below

Assignment II.5. Pretest & Revise Questions

Spatial and Analytical Study of Student Housing at Carnegie Mellon

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

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.

ok

If reporting off campus addr to registrar is optional, what is different between those who report and those who do not? What would be different about their housing data, if you had it? Answer in report(s) on the project.

all good things to look for.

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 have a feasible amount of records and enough people to sort through the data so we do not need to have a random sampling scheme. 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.

good preliminary look

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.

for this I want you to treat the 891 + 4036 records you have as a stratified sample from the 2252 + 5769 possible off-campus residents.

L. QUESTIONNAIRE OR OBSERVATIONAL PROTOCOL:

Undergraduate/Graduate

Is the person an undergraduate student?

within the two strata (undergrads and grad students) it may be possible to further weight the data to take away some of the bias that might be there.

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

neat - anythnig else interesting to get from this database?

N. Pretest of a revised version of your questionnaire (or observation protocol) on a group of possible respondents/units.

We tested our questionnaire by picking a record randomly from the first 15 letters of the alphabet and only one record per letter, this would prevent randomly picking the same person who had multiple records. In order to condense the size of the table, we used simply "Undergrad, Masters, or PhD," "College" and "Department" instead of a column for each yes or no question about which year, college and department the student belonged to.

See Table I at the end.

O. Report on the specification of and results from the pretest, and any redesign of the questionnaire (or observation protocol) that may be required.

After pulling a couple of observations from our data set, we decided to include City as a variable in our observation protocol. We observed that some of the graduate students live outside of the city limits of Pittsburgh and including a City variable will improve these records. From the pretest, we also found a problem when calculating distance between residences and Carnegie Mellon's campus. Google Maps does not provide simple measures of distance (only distance on driving routes or walking paths), so we are going to look to the GIS software to provide distance measures between residences and campus. We also discussed how to measure time on public transit between each residence and CMU. We decided that the time it takes to walk to bus stops and from bus stops to campus should be included as part of the commuting time.

great. I would think GIS will give you "as the crow flies" distances. However, is that more relevant for locating one's residence than driving/walking routes?

Either way is OK, as long as you justify it in your report(s).

Table 1: Pretest Results

	(U)ndergrad, (M)asters, or (P)hD?	College (HSS, TSB, CIT, CFA, SCS, MCS, UC)	Major / Department	Type of Building (house or apartment)	Neighborhood	Distance to Campus (mi)	Time to campus by foot (min.)	Time to campus by car (min.)	Time to campus by public transit (min.)	Address	Letter
1	U	HSS	PSY	Apartment	Shadyside	1.2	21	5	14	617 Bellefonte	D
2	P	SCS	SE	Apartment	Squirrel Hill South	2.9	61	10	22	3245 Beechwood Blvd	I
3	U	CFA	MUS	House	Oakland	0.6	12	3	8	308 S. Neville St	N
4	M	TSB	IA	House	Shadyside	0.6	13	2	10	409 MOREWOOD AVE, APT 3	C
5	U	MCS	MCS	House	Squirrel Hill North	1	19	3	6	1712 WIGHTMAN ST	H
6	P	MCS	MCS	House	Shadyside	0.8	16	3	14	515 St James Place	M
7	P	HSS	ML	Apartment	Wilkinsburg	5	132	13	29	1000 Bryn Mawr Court West, Apt. 212 W	E
8	U	CFA	DRA	House	Shadyside	0.9	18	3	10	921 Ivy Street	J
9	M	TSB	IA	Apartment	Shadyside	0.9	20	3	21	545 South Aiken Ave, Apt. 14	O
10	M	CIT	INI	Apartment	Oakland	0.8	16	3	10	151 N. Craig Street Apt. 10C	A
11	U	CIT	CHE	Apartment	Squirrel Hill South	1.6	30	6	12	6432 Darlington Rd. Apt. 2	F
12	P	CIT	ECE	House	Shadyside	0.9	19	3	15	714 South Aiken Ave.	K
13	M	CIT	ECE	Apartment	Oakland	0.7	16	4	13	4 Bayard Road	B
14	M	TSB	IA	Apartment	Oakland	0.7	15	4	12	2 Bayard Road	G
15	P	SCS	ROB	Apartment	Squirrel Hill	0.9	18	4	8	1401 Wightman Street	L