Analysis and Design of CMU Alumni Survey

Final Report

Yi Jiang & Zhanwu Liu

May 4, 2007

Abstract

The goal of this project is to provide some suggestions for future Carnegie Mellon alumni panel survey based on the information in the alumni database and survey results. Analysis suggests one must be very cautious in designing a email-based survey to avoid any selection effect. We also identified some important predictors (variables) for alumni activity, including *Email indicator, Graduation Year, Undergraduate study at CMU, Number of Degrees from CMU, the living region* and *alumni spouse*. In contrast, the *studying school* and *the Field of Work* are not important predictors of activity. We believe that these results would be helpful in survey design.

In addition we also provided some general considerations based on literature review on panel survey. The considerations including *interview spacing*, *Mode of the interview* and *respondent selection* are given in the last part of the writing.

Contents

1 Introduction

The client of this project is Judy Cole and Christian Krohn from Carnegie Mellon Office of Alumni Relations. Their primary interest is to ensure that alumni participation and concerns are considered through the university's advancement efforts. More specifically, in this project they are interested in designing a survey panel that is representative of the whole CMU alumni population. In this project we analyzed information from CMU alumni database, and provide some suggestions for future panel survey design.

Based on the interest of the client, we formulated several problems that can be answered in one semester:

1. Analyze the bias of email based surveys. Nowadays, email is used widely in surveys, but one problem is the sampling bias of email based surveys. For example, some professionals are more likely to use email than other respondents. In our alumni database, only half of the records contain email addresses. We did some exploratory data analysis to determine what is the difference between those records with email addresses and those records without email addresses.

2. Analyze the predictor for *active* alumni. Alumni are categorized as either *active* or *inactive* based on their involvement in university advancement efforts, including donation, participation in university related activities, etc.

3. Present some suggestions on panel survey based on literature review.

4. **Planned but not done:** Estimate the active alumni population. We did not work on this part. We planned to use the respondent data from previous surveys as well as other indicators as response, to identify the characteristics of the alumni who like to take part in events related with the university. In addition, we will use capture/recapture methods to estimate the population size of all active members.

2 The Data

The office of alumni relations maintains a database with $70,985$ alumni records. The database contains the following information:

- Basic biographical information: name, address, employment information, personal relationships, degree information, affiliations, etc.
- Additional biographical information: activities/events, awards, committees, interests, philanthropic affinities, publications, student activities, sports, etc.
- Prospect information: wealth, prospect rating, etc.
- Giving summary: appeals, gift clubs, total giving, largest gift, matching claims, etc.

We have signed the confidentiality agreement and obtained most of the data on Feb 28. The client has provided us the records from the alumni database. In addition, the data of one email based survey in 2006 was also provided but we did not get chance to work on it.

2.1 Missing Data

In the basic biographic information, there are some data point missing, as shown in table 1. Missing data is unavoidable in data collection, and there is no doubt that the missing data cause some trouble in analysis. For example, more than half of the *FLD OF WORK CODE* data was missing, which made it hard to infer any properties related with *FLD OF WORK*.

Other data files contain donation, committee, activity and event participation information. According to the client, the donation information is up-to-date and accurate, while the event participation information has much lower quality(a lot of missing data during data collection). Thus when we consider the active state of particular alumnus, we pay more attention to donation information.

Variable	Percent Missing
Address	$665(0.94\%)$
Geo Code	$17,656(24.87\%)$
Job Title	$31,752(44.73\%)$
Fld Of Work Code	$37,850(53.32\%)$
School	$20(0.03\%)$

Table 1: Data Missing in the basic biographical information

3 Exploratory Data Analysis

The basic biographical data is the most complete set of data in the database. It contains record for all the alumni in the database, and contains most of the strata information(geographical, year/school, etc.) that can be used in stratified sampling. Thus the first thing we did is some EDA on basic biographical data.

3.1 Gender

In the alumni database, there are $70,985$ records, including $21,913(30.87\%)$ female and $49,071(69.13\%)$ male. The gender information is missing in one record.

3.2 Field of Work

Out of total 70985 records, only 33135 contains the field of work information. There are totally 92 fields of work coded in the records, and the top 10 fields are:

```
Code          Number     Percentage  Description
---------------------------------
BA            12320       37.18   Business/Industrial Administration
EO             3964       11.96   Engineering, Other
EH             3643       10.99   Education, Higher
CU             1657        5.00   Consulting
```

CP	1431	4.32	Computing/Programming
AT	1196	3.61	Art
AR	955	2.88	Architecture
OT	860	2.60	Other?
NP	845	2.55	Non-Profit Organization
PK	644	1.94	Public Works

3.3 Degree

Of all the records, 42903 have studied as undergraduate student at CMU. The others only did there graduate study here. The total number of degrees including undergraduate degree are listed in table 2.

Number of Degrees	Percent
1	88%
2	10.73%
3	1.24%
4	0.03%
5	1 count

Table 2: Number of Degrees from CMU

3.4 Geographical distribution of the alumni

Out of all records, $6,045$ records have international addresses, and $64,275$ have US addresses (including Puerto Rico). Address information is missing for 665 records (no zip code, and no country name).

The distribution of international students are plotted in Figure 1. From the pie chart, it is very clear that a large portion of the international alumni are in Asian countries/regions. One should note that this number is not the same as the number of international-origin alumni

because the alumni may work in a country different from the one he came from.

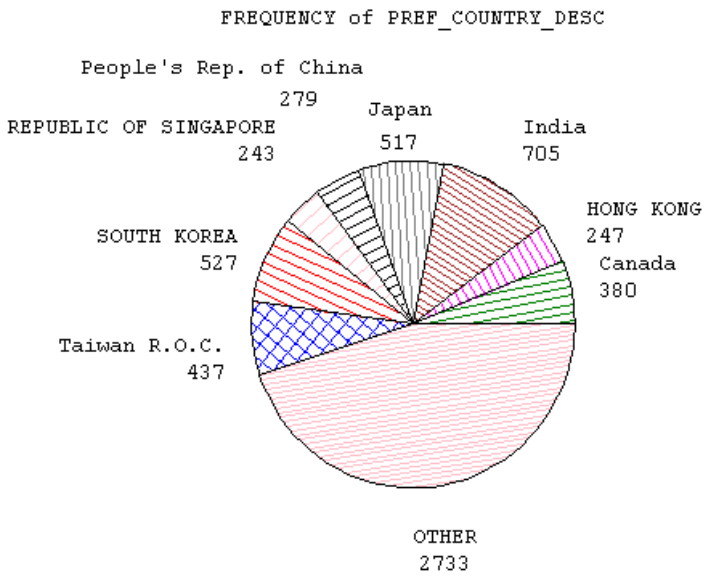

Figure 1: Distribution of alumni not in the US. It is clear that at least half of the $6,045$ international alumni are in Asian countries. Countries/regions with less alumni are grouped in "Other".

For the alumni in the US, Figure 2 shows the distribution of alumni in the states. It is natural to expect that a great fraction of the alumni will be in PA, and indeed this is the case. CA comes second, followed by east coast states (NY, NJ, MA and VA).

3.5 Graduation year and school

The number of alumni graduated in each year is plotted below. There is a clear increasing tendency over the years. We assume that the increase in number is due to the increase in number of students recruited. However, from the history of the database, the number of alumni from earlier years maybe underestimated when all the records were input manually.

The number of records from each school is also plotted. This plot was made using the *School* variable in the database directly. Note that since there is some change in the school

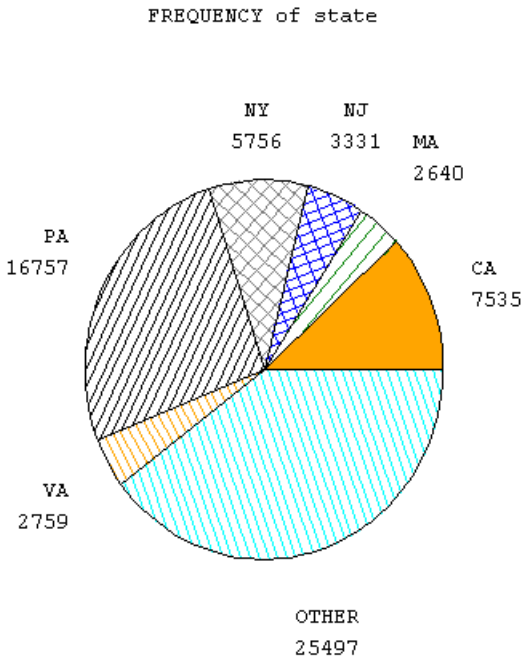

Figure 2: Number of alumni records in each state. States with less alumni are grouped in "Other"

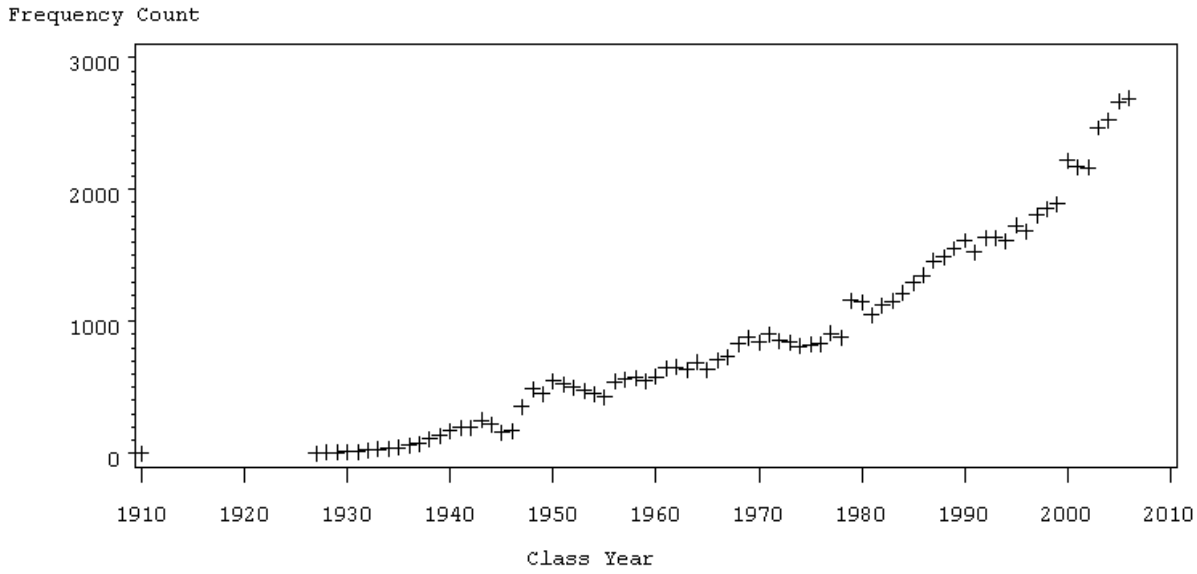

Figure 3: Number of alumni records from each graduation year

names, for example GSIA changed to TSB in 2004; there were merging and splitting events, and also some schools just disappear (such as Margaret Morrison), and even some alumni has the school name "CMU", which is hard to interpret.

school	Frequency	Percent
AM	134	0.19
CFA	10601	14.94
CIT	20405	28.75
CMU	1625	2.29
CS	2892	4.08
GSIA	8121	11.44
H&SS	7952	11.21
HNZ	4205	5.93
I	16	0.02

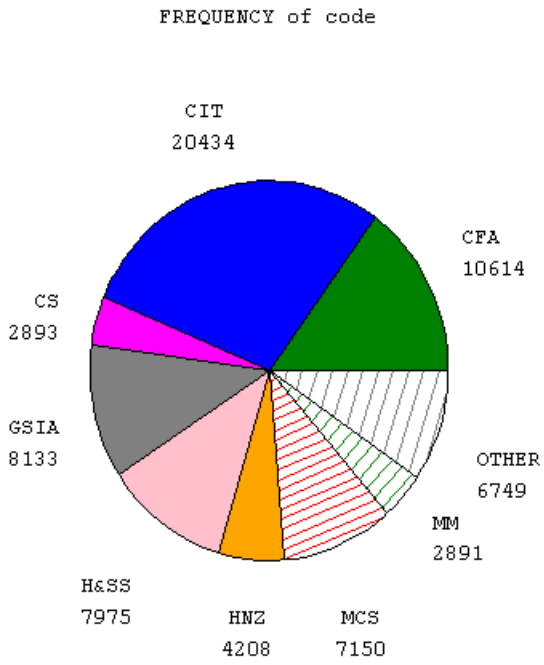

Figure 4: Number of Alumni records from each school

IM	2550	3.59
L	292	0.41
MCS	7144	10.07
MM	2896	4.08
PM	488	0.69
TSB	1642	2.31
UNSP	2	0.00

3.6 Family information

In 5889 records (8.30%), the spouse is also CMU alumni. We can expect that these alumni be more loyal to the university.

The number of children information is also included in the database. It is natural to expect that fresh graduates have none or small number of kids, and older alumni have more kids. In the database, there is a huge portion of the alumni have no kids (78%). Certainly this cannot be interpreted as CMU graduates do not like kids, rather this reflect that a lot of alumni records have not been updated for long time.

Num Child	Frequency	Percent
0	55755	78.54
1	4174	5.88
2	6004	8.46
3	3011	4.24
4	1252	1.76
5	410	0.58
6	221	0.31
7	74	0.10
8	42	0.06
9	13	0.02
10	13	0.02
11	7	0.01
12	3	0.00
13	1	0.00
14	3	0.00

4 The possible bias of email based survey

More than half of the records ($36,583,$ or 51.54%) contain email information. Since nowadays, more and more surveys are conducted using email, it is necessary to analyze if there is

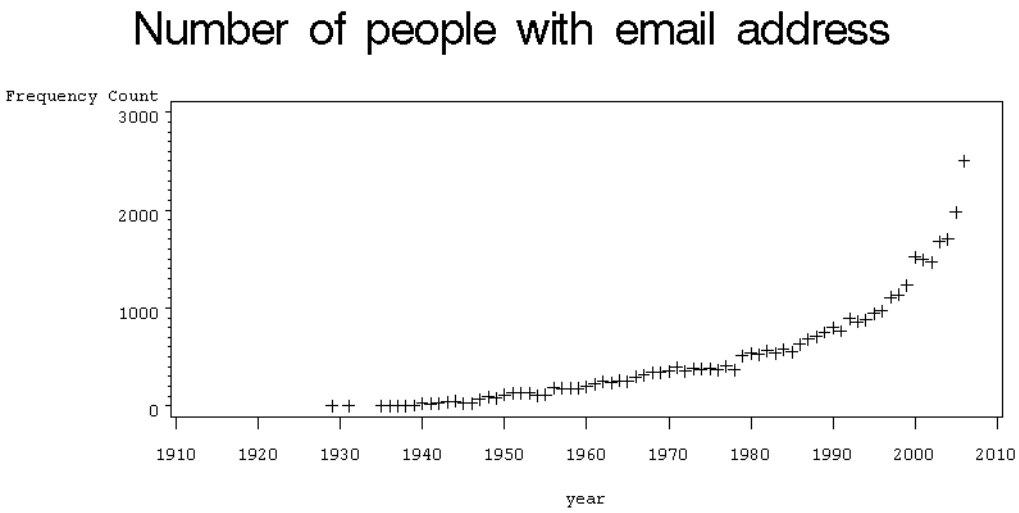

Figure 5: Number of people with email addresses from each graduation year

any selection effect exists or not.

The number of alumni with email addresses in records is plotted in Figure 5. We already knew that the number of alumni records increase with time, thus it is not surprising to see the same trend here.

The proportion of people with email addresses in records is also plotted in Figure 6. The proportion also increases with time, as we may have expected. The problem from this increase is that if we do survey based on email, there is more chance to sample alumni graduated in recent years, e.g. it is biased against alumni graduated earlier.

In addition the records show that the proportion with email is very different among different fields of works. As shown in table 4, those taking computer programming jobs have almost twice the probability of having email in record than those work in arts.

The *School* is very related with *the fields of work*, as shown in table 4.

Table 4 shows the number of people with email addresses in each school. Only the current seven schools are listed.

The non-uniform distribution of those with and without email will inevitably cause some

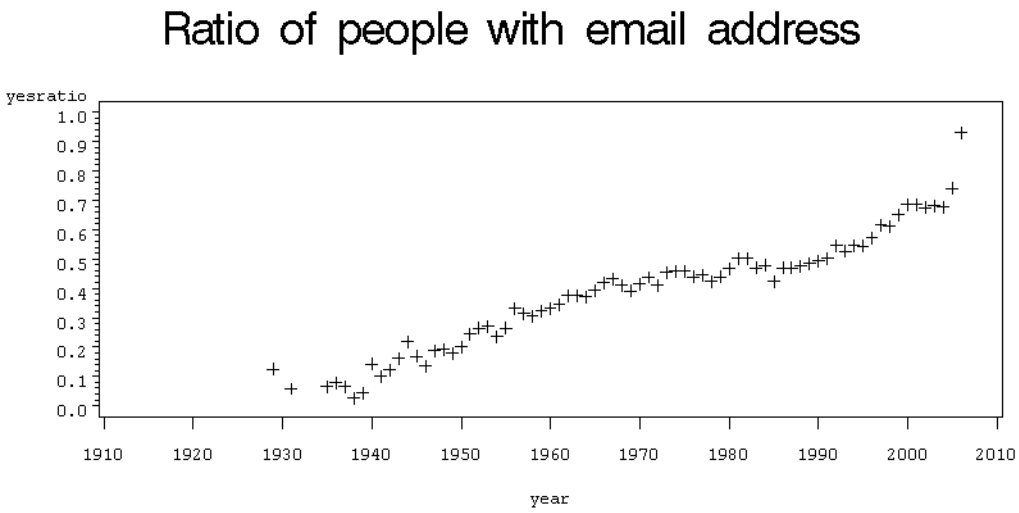

Figure 6: Fraction of alumni with email address in each graduation year

bias in the email based-survey. Thus we must be very cautious when we design a survey based solely on email questionnaires.

5 Activity analysis of all alumni

Identifying the predictors for active alumni is an important goal of this study. To do this, the very first thing we need to do is to define who is active and who is nonactive. Unfortunately, there is no golden rule to do this, and all categorizing method are arbitrary to some extent. The easiest quantity related with activity is the monetary donation amount, but still this quantity is affected by many factors, such as how wealthy one person is, etc. Because the missing data in event participation cannot be estimated, in the analysis, we used binary response “active” and “nonactive”.

The data files related with alumni activity are listed below.

- activity.tsv, remove the records with “Expired”, “Declined”, “Refused”.
- committee.tsv, treat all as active

Field Code	Field Description	Percent with email
BA	Business/Industrial Administration	60.44%
EO	Engineering, Other	42.99%
EH	Education, Higher	52.48%
CU	Consulting	68.92%
CP	Computing/Programming	77.22%
AT	Art	39.38%

Table 3: Percentage of person with email in several most populated fields

School	Number with email	Number without email	Percent with email
CFA	4386	6215	41.37%
CIT	9627	10778	47.18%
CS	2118	774	73.24%
GSIA + TSB + IM	7836	4477	63.64%
H&SS	4264	3688	53.62%
HNZ	2561	1644	60.90%
MCS	3601	3543	50.41%

Table 4: Percent with email in each school

- gift_clubs.tsv, treat all as active
- largest_gifts.tsv, treat all as active
- lifetime_year.tsv, treat all as active

To simplify the analysis, the address of the alumni in the US are coded as Census Bureau-designated areas, as shown in table 5. All the internationals are coded as “Foreign”.

Region	Division	States
NorthEast	New England	CT, ME, MA, NH, RI, VT
NorthEast	Middle Atlantic	NY, NJ, PA
MidWest	East North Central	IL, IN, MI, OH, WI
MidWest	West North Central	IA, KS, MN, MO, NE, ND, SD
South	South Atlantic	FL, GA, NC, SC, VA, MD, DE, WV
South	East South Central	AL, KY, MS, TN, AR, LA, OK, TX
West	Mountain	AZ, CO, ID, MT, NV, NM, UT, WY
West	Pacific	AK, CA, HI, OR, WA

Table 5: Census Bureau-designated US areas

5.1 Logistic regression 1:

Define active member as those who have donated

In the first logistic regression analysis, we define an ID as active if the same ID is found in the file *largest_gifts.tsv*. This file contains all IDs who has contributed to the university financially.

It turned out that *EMAIL_IND, YEAR, UnderDegree, NumDegree, Division* and *Alspouse* play important rules, and *School* and *Field of Work* are not significant predictors. The fitted data is shown in table 6.

5.2 Logistic regression 2:

Define active member as those who have donated or participated in events

For the second logistic regression, "active" is defined based on donation (largest_gifts.tsv), activity (activity.tsv with records with "Expired", "Declined", "Refused" removed), committee (committee.tsv) and gift club (gift_clubs.tsv) information. Thus totally 50520 of the 70985 were labeled as "active".

Parameter	*DF*	*Estimate*	*Standard Error*	*Wald Chi-Square*	*Pr > ChiSq*
EMAIL_IND Y	1	0.5092	0.00946	2895.5600	<.0001
year	1	-0.0571	0.000706	6547.8671	<.0001
UnderDegree Y	1	0.1041	0.0110	90.2578	<.0001
NumDegrees	1	0.1337	0.0241	30.7203	<.0001
Division East North Central	1	0.2611	0.0296	77.7701	<.0001
Division East South Central	1	0.1038	0.0717	2.0976	0.1475
Division Foreign	1	-1.2619	0.0318	1575.3907	<.0001
Division Middle Atlantic	1	0.1502	0.0187	64.2243	<.0001
Division Mountain	1	-0.0132	0.0449	0.0867	0.7684
Division New England	1	0.3075	0.0320	92.2533	<.0001
Division Pacific	1	0.00300	0.0247	0.0148	0.9033
Division South Atlantic	1	0.1913	0.0239	63.9900	<.0001
Division West North Central	1	0.1293	0.0621	4.3384	0.0373
ALSPOUSE Alumni Spouse	1	0.3951	0.0169	544.3944	<.0001

Table 6: Logistic regression results 1: *EMAIL_IND, YEAR, UnderDegree, Number of Degrees, Division* and *Alspouse* are significant predictors. Positive estimate indicates more active, and negative means less active.

The logistic regression results are very similar to the previous one, and the results are in table 7.

6 Discussion

6.1 Email based survey

In recent years, email is used more frequently in different kinds of survey because it is more convenient and cheaper. But there are two major problems related with email based survey:

Parameter	*DF*	*Estimate*	*Standard Error*	*Wald Chi-Square*	*Pr > ChiSq*
EMAIL_IND Y	1	0.5782	0.0106	2959.0770	<.0001
year	1	-0.0722	0.000932	6003.7310	<.0001
UnderDegree Y	1	0.3991	0.0118	1151.3026	<.0001
NumDegrees	1	0.1527	0.0284	28.9814	<.0001
Division East North Central	1	0.2867	0.0347	68.2828	<.0001
Division East South Central	1	-0.2451	0.0802	9.3358	0.0022
Division Foreign	1	-1.0614	0.0294	1301.5267	<.0001
Division Middle Atlantic	1	0.1391	0.0213	42.6616	<.0001
Division Mountain	1	0.1852	0.0539	11.8075	0.0006
Division New England	1	0.3645	0.0377	93.2218	<.0001
Division Pacific	1	0.1967	0.0281	49.1110	<.0001
Division South Atlantic	1	0.1849	0.0278	44.0716	<.0001
Division West North Central	1	-0.1224	0.0688	3.1675	0.0751
ALSPOUSE Alumni Spouse	1	0.4982	0.0239	435.6728	<.0001

Table 7: Logistic regression results 2: *EMAIL_IND, YEAR, UnderDegree, Number of Degrees, Division* and *Alspouse* are significant predictors. Positive estimate indicates more active, and negative means less active.

(1). Selection effect, e.g. the population use email may not be representative of the whole target population; (2). Validity of email address, e.g. even you have the email address, the address may be outdated and the information may never reach the target recipient, this will decrease the response rate.

In this study we worked on problem (1). We analyzed the data and showed that the population with email indeed is not representative of the whole alumni population. Alumni from different years, schools or work in different fields are have different proportion to have email address in the database. So we must be very careful in using email as the only means to do survey.

Problem (2) is not the target of our study, but we would like to give some general comments. One possible consideration is when selecting email addresses, we can treat differently for free email addresses (yahoo, hotmail, gmail, etc) and other more formal email addresses. How to deal with them warrants further study in this direction.

6.2 Strata

When sub-population vary considerably, stratification by sampling from each subpopulation independently has the advantage of improving representativeness of the sample and decreasing the variance of estimation.

Several possible strategies can be used in stratification, such as

1. Proportional allocation uses a sampling fraction in each strata which is equal to that of the total population. Then the obtained sample can closely represent the population.

2. Equal allocation take same numbers from strata varying widely in size, it may be used to equate the statistical power of tests of differences between strata.

3. Optimum allocation uses a sampling fraction proportionate to the standard deviation of the distribution of the variable, larger sample are taken in the strata withe the greater variability to generate smaller sampling variance.

In this study, we found that the activity of alumni differ significantly between those differ in the following predictors:

1. Email indicator: those with email in record are more active.

2. Class Year: recent graduates are less active. This seems counter-intuitive but one possible explanation is that new graduates tends to connect with friends rather than directly with the university; they are less likely to donate because of financial limitations.

3. Undergraduate Study: alumni studied as undergraduate are more likely to be active.

4. Number of Degrees from CMU: the alumni with more degrees from CMU are more likely to be active.

5. Alumni Spouse: If both of one couple are CMU alumni, they are more likely to be active.

6. Geographic location: those in foreign countries are less active. Also different area in the US are different, generally those in east or west coast are more active than those in the middle part.

Based on the analysis, I would suggest to consider class year, geographical region and undergraduate study in CMU as strata. Number of degrees from CMU and alumni spouse can also be considered, but not recommended because the number of alumni in each stratum based on this would be very unbalanced.

7 Suggestions for the Design of the Panel Survey(literature review)

Panel surveys are a kind of survey that measures the same sample at different points in time. They offer attractive features that cross-sectional surveys do not have. To better design the panel survey, its good to first review the purposes of a survey for different kinds of information needs.

1. understand the characteristics, behavior, hobbies, and attitudes

2. understand the external factors that are associated with or influencing the characteristics, behavior, and hobbies

3. estimate the changes of these characteristics over time

4. estimate the relationships among characteristics

5. estimate the frequency of occurrence for specific kinds of events

6. estimate the size of a specific kind of group of people

7. estimate the influences of surveys on the characteristics

8. estimate the causal effects of a specific characteristic

9. estimate the group characteristics by the individual characteristics

In designing a survey to understand and then improve the relationship between the Carnegie Mellon alumni and Carnegie Mellon, estimates of the changes of the alumnus's relationship with the school with time, and the changes due to the external factors, such as the school activities for the alumni, are certainly of great importance to the development of the strategies for the improving the relationship. In summary, purposes 2), 3), 6), 8), 9) from the above list are to be investigated in this project. And for these purposes, panel surveys are the right choice, where we can look at how the exogenous and endogenous factors interrelate with each other over a short or long time.

However, there are difficulties in conducting panel surveys which need a great deal of care, since they could affect the survey design, such as the basic sample structure, administration of the survey, and the database structure, estimation, and analysis of the panel data. (Kasprzyk, Duncan, Kalton, Singh, 1989)

According to (Kasprzyk, Duncan, Kalton, Singh, 1989), these difficulties mainly arise from three features in conducting panel surveys. They are 1) interview spacing, 2) the mode of the interview, and 3) respondent selection. In the sections below, we describe these features and further infer their influences and analyze possible solutions, particularly in our project.

7.1 Interview Spacing

Interview spacing is the most important factor that affects the design and the response characteristics. Proper selection of the interview spacing should take the following four facts into account:

1. If the interview spacing is too short, meaningful changes may not have taken place and its not meaningful to take the next survey.

2. If the interview spacing is too long, there will possibly be telescoping and omissions of events in the panel data, since the respondents are not likely to remember those that happened too long ago. For example, under some sociological topics such as crime victimization rates survey, short recall periods are shown to have significant higher rates than long recall periods.

3. There are panel effects that will affect the survey results. Panel effects are the results of behavioral effects which is a common problem in many types of evaluation studies where peoples behavior might be changed after participating in the survey. Take the election for example, people who took a survey relating to an election may pay more attention to the election after taking the survey and this will certainly affect their final decision on the vote.

There are implications of these facts on the interview spacing in our project.

Since a majority of the panel survey is expected to be done by emails, we want to be able to track the person through the same email address as before. Therefore, interview spacing for certain topics or events should not be longer than the expected time that people are going to change their email addresses. For example, people who have graduated for a short period of time have the tendency to change their email addresses more frequently. Surely, the difficulty on choosing the interview spacing when considering this fact can be eased by sending surveys to more reliable email addresses if we can get them.

The other thing is just as in general surveys, if the spacing is too long, it is going to introduce bias when asking the respondents to answer questions that require the alumni to recall events.

A good example for the importance of the survey spacing is the Annual Victimization Rates study by recall interval (Kasprzyk, Duncan, Kalton, Singh, 1989) in Table 8. The recall interval is equivalent to the survey spacing here. We can see a systematic increase of the victimization rate for the shorter recall interval: 3 months. Almost all of their differences are significant at either 10% or 5% level.

Type of Crime	Recall	Interval	Difference
	6 Months	3 Months	
Total personal crimes	12.85	15.49	-2.64**
Crimes of violence	3.46	4.29	-0.83**
Crimes of theft	9.39	11.20	-1.81**
Total household crimes	23.00	26.38	-3.83**
Burglary	8.53	9.68	-1.15*
Larceny	12.70	15.09	-2.39**
Auto theft	1.78	2.07	-0.29

Source: Bushery, 1981.

* Significant at the 10% level.

** Significant at the 5% level.

Table 8: Annual victimization rates by recall interval (per 100 persons 12+ years old).

7.2 Mode of the Interview

The mode of the interview has been shown to be related to the response rate in many cross-sectional surveys (Dillman, 1978; Oksenberg et all, 1986). Particularly, the mode of the first interview in the panel survey will affect both the response rate of itself and that of the

subsequent interview waves. Some studies showed that in-person surveys are good modes for the interview although it might be costly. However, partly for this reason, in-person surveys are conducted in the initial interview of the panel survey to obtain the best tradeoff between high response rates and low costs. Yet still, this is still too costly for the alumni surveys. A compromising method is to use the phone interview, particularly it is good for the short introduction of the survey or short surveys themselves.

Particularly, this is important in our project because surveying by email can actually be better than surveys by mail or phone interviews if its advantages are taken and disadvantages are avoided. Specifically, its advantages over mail or phone interview methods include faster delivery, lower cost, wider coverage, and easier maintenance. Its disadvantages include the reduced deliverability due to spam filtering, rarely used email addresses, and the limited number of techniques to express the information. Thus, it's worth attractive design and wording of the email title and the first few paragraphs. Particularly, highlights of the benefits of the survey for the participants in those locations are of great help in increasing the response rate in both the first interview and the subsequent interviews.

Another advantage in using emails in our project that is worth mentioning is that panel surveys by email make it easier to track the respondent if the interview spacing is properly chosen. Thus, it's the combined consideration of the first two features we just discussed.

Therefore, we suggest that we can use mail or phone interview in the first round of survey and use email in the following surveys. This would probably achieve the best tradeoff between high response rates and low costs.

A table of the advantages and disadvantages of these delivery methods are given in Table 9 below.

7.3 Respondent Selection

Certain types of panel survey may use a variety of types of respondents. These different types of respondents have different effects on the response quality. For example, bias can be

Type of Delivery Method	Advantage	Disadvantage
In person	Not feasible in alumni survey	
Phone interview	High response rate	Time sensitive
	Wide coverage	High cost for long distance calls
	Good for alumni who do not use email	Limited number of ways to express
Email	Fast delivery	Reduced deliverability due to spam filtering
	Low cost	Rarely used email addresses
	Wide coverage	Limited number of techniques to convey the information
Mail	Higher response rate than email	Higher cost than email
	Good for alumni who do not use email	

Table 9: Advantages and Disadvantages of a few survey delivery methods.

introduced if the respondents do not very well represent the whole population.

Thus, we can first get the estimate of the total number of the "active" alumni and try to maintain the same proportion of "active" alumni in the panel to be surveyed such that the panel represent the whole population of all the CMU alumni. This is also important in our project since we need to rotate a fraction of people out of the panel after they have been surveyed a number of times. Here we need to make sure the new selected people can also represent the whole population with the same proportion of the "active" alumni.

Also, we may want to keep track of the significant changes of this proportion of "active" alumni and other distributions of the alumni over different properties to adjust our panel composition accordingly.

8 Conclusions

For the data analysis, we suggest to consider class year, geographical region and undergraduate study in CMU as strata. When we design an email-based survey, we must be very cautious about the possible bias.

In designing the panel survey, survey spacing, the mode of the survey, and respondent selection are the three key factors that we should consider carefully.

9 Suggestions for Future Work and Collection of the Data

9.1 Future Work

It would be better to use the categorical data analysis methods such as loglinear models to analyze this database since most of the variables are very categorical. Clustering methods are good for both continuous and discrete variables, so it is also good the stratification of the alumni.

Review of the literature is a good way to find similar examples, such as in customer behavior analysis, to help decide the survey spacing in developing the panel survey.

9.2 Future Collection of the Data

Based on our analysis, we have some suggestions for future collection of the data that we think is helpful for further statistical analysis:

1. Complement the missing data as much as possible.

2. Group small categories of some variable into big categories if possible. And add this information to that variable for fast analysis in using some statistical techniques that desire low dimensions. For example, in the Job Title variable, there are still too many

small categories that make it very time-consuming to perform statistical analysis when the dataset is large.

3. Provide the information about the times when the email/mailing addresses are changed. Therefore, we can track how often do the alumni change their addresses change on average and then we can better estimate the optimal survey spacing.

4. Class and department size information for every individual alumni. The larger their network is with Carnegie Mellon, the higher the chances they are in touch with the school.

5. The scholarship, fellowship, assistantship, or other financial aid information of the alumni if it is available but not recorded in the variables, such as honor.

6. Through the coorperation with www.linkedin.com, we can record which alumni are in linkedin. These alumni are expected to have more chances to be in touch with the school.

References

Dillman, D.A. (1978), Mail and Telephone Surveys: The Total Design Method, John Wiley & Sons, New York.

Kasprzyk, D., G.J. Duncan, G. Kalton, M.P. Singh (1989), Panel Surveys, John Wiley & Sons, New York.

Oksenberg, L., Coleman, L., & Cannell, C.F. (1986), "Interviewer Voices and Refusal Rates in Telephone Surveys," Public Opinion Quarterly, 50, 97-111.