

An overall comment:

*** see notes below about footnotes (I don't prefer) vs list of references and citations from the paper (I do prefer)**

Bassem Mikhael
Christopher Loncke
Abigail Daughtrey
Perna Agarwal
Yong-Gyun Choi
April 1, 2011

Final Report Rough Draft

Introduction:

Numerous health studies have demonstrated the damaging effects of excessive caffeine consumption on cardiac wellbeing in addition to psychological and mental health^{1,2}. Significant concern has been expressed that students today consume large amounts of caffeine to keep up with their academic workload or enhance their performance in sports activities. This problem is especially severe in schools which are more work-intensive, which Carnegie Mellon University certainly is. Therefore, we have studied patterns of caffeine consumption among undergraduate students at Carnegie Mellon with the hope of characterizing caffeine consumption, as well as gaining an understanding of students' perception of caffeine consumption. Through the former component, we hope to determine if there are any correlations between the amount/type of caffeine consumption with GPA, the college the student is in, their business level, etc. On the other hand, the latter component is designed to test whether students are comfortable with the amount of caffeine they consume, and to understand the more subjective perspective on this matter.

The summary of results will go here.

Methods:

The target population for our study is comprised of all current Carnegie Mellon undergraduate students on the Pittsburgh campus. The sampling frame we selected to administer the study is C-book, a published booklet listing all current students' names, years, and email addresses. A very small number of students actively choose to opt-out of being included in the C-book; these students are members of the target population but not the sampling frame, and therefore will not be included in the sample, resulting in coverage error. Due to the very small number of students in this category, however, we believe the sampling frame to be an excellent tool to use to sample the population.

The design of the questions proposed for the study is essential, as the results will allow for an understanding of the sample participants and a determination of the caffeine consumption patterns at Carnegie Mellon University. For this survey, we decided to keep the questions short, simple and concise to streamline the process of responding for survey participants. Also, we decided to make all of the questions multiple choice to maintain consistency between answers and to avoid any discrepancies which would arise from coding variable responses. The questions comprise five major categories - demographic information, academic performance, extracurricular involvement, general well-being, and student perception. The demographic questions asked for basic information such as class year, gender, college, etc. The academic performance questions determined the student's GPA, units taken, hours

¹Mayo Clinic "Caffeine": (<http://www.mayoclinic.com/health/caffeine/NU00600>)

²The Standard Online "Caffeine Addiction Problem" (<http://media.www.the-standard.org/media/storage/paper1059/news/2008/02/05/Features/The-Caffeine.Addiction.Problem-3189385.shtml>)

Instead of footnote citations I prefer a separate "references" section with citations in the text. Examples of both the citation format that I want and the reference list format that I want are at <http://www.library.cornell.edu/resrch/citmanage/apa>

spent on class work, etc. This can allow us to correlate caffeine consumption with academic performance. Extracurricular involvement, the third category, asks questions which aim to determine how busy the student is; questions ask the student how many hours he/she spends on leisure, extracurricular activities, employment, etc. This “business” score can be tested to determine correlation with caffeine consumption. The fourth category of questions takes a slightly different approach, and aims to determine the student’s general wellbeing through measures such as frequency of exercise and sleeping patterns. Lastly, the perception questions are designed to try to understand the students’ subjective feelings and comfort level with their present level of caffeine consumption.

We found that the easiest and most efficient manner to administer this survey is to administer it online. In order to do this, we chose Google Forms, a free service which is part of GoogleDocs, rather than the more traditional choice of SurveyMonkey. Google Forms offered automatic data entry into a spreadsheet at no cost, while SurveyMonkey lacked these features. Both to minimize expense and to minimize the probability of human error which would occur from manually copying data into a spreadsheet, we chose to use Google Forms. After composing the survey questions, we pretested the survey and discovered, in addition to minor errors, that we had to encode questions in a specific manner to allow for multiple responses. Certain questions in the survey necessitate the capability to choose multiple responses, so we modified the survey to allow for this.

The actual sampling method chosen was simple random sampling with stratification. First, the strata were decided to be the 5 possible class years, freshman, sophomore, junior, senior, and 5th year students, because we believed that this would allow us to obtain a smaller variance and that the classes would differ enough to make stratification effective. In order to calculate sample size for each stratum we used the mean error formula for stratified samples:

$$ME = (1.96) \times \frac{\sqrt{(\sum_{h=1}^H W_h^2)(1 - f_h)(s_h^2)}}{[(f \times N)_h]}$$

We used 0.1 for the ME to give us a small confidence interval, and 2 for s based off of pretesting and the responses of our own group to the question “How many servings of caffeine do you consume on a daily basis?” W_h and N_h were both easily calculable using admissions statistics. Using the class populations on

page 8 of the admissions statistics provided from CMU,³ we calculated f to be about 0.015. Using those same statistics, the respondents needed for each category were calculated to be 20 freshmen, 18 sophomores, 18 juniors, 17 seniors, and 3 fifth years. With a very conservative estimated response rate of 15%, the sample sizes for each class were 120, 108, 108, 102, and 18, respectively.

Year	Number Needed	Number Selected
Freshmen	20	120
Sophomores	18	108
Juniors	18	108
Seniors	17	102
5 th Years	3	18
Total	76	456

³ http://www.cmu.edu/ira/factbook/pdf/facts2011/4_enrollment-final-as-of-2_24_11.pdf

Once these figures were determined, we selected the sample. To do so, we used an Excel random number generator to generate pairs of random numbers (n1, n2), such that n1 was the page number and n2 was line number in C-book. That individual's email address was then recorded, and a master d-list composed. One of the biggest challenges we faced with respect to the survey creation was repetition. Sometimes after generating a pair of numbers, the corresponding name would either be a student we have already chosen or a student that wasn't supposed to be in our sample. In either case the solution was to generate new numbers and try again. Once the final distribution list was composed, we sent the survey out in email form. To provide an incentive to respond, we offered a \$50 Amazon gift card to one individual who was selected at random from a list of those who voluntarily provided their Andrew IDs. In order to maintain confidentiality, providing this ID was purely optional, so as to preclude any possibility of linking names with responses.

Results:

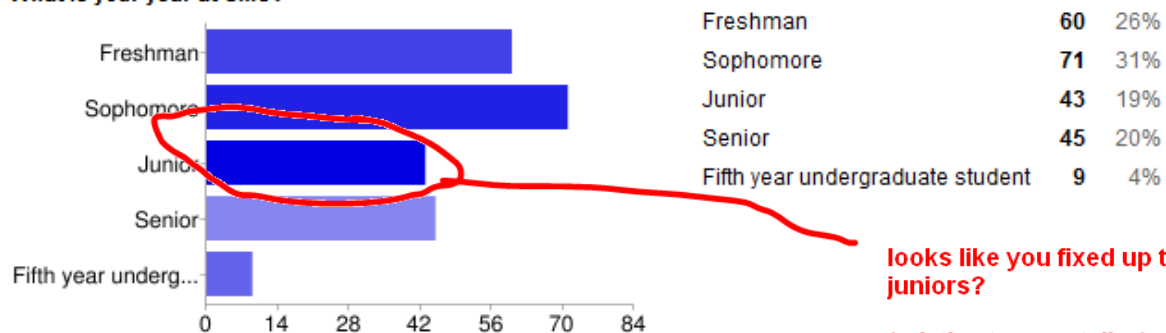
Full Results will go here, along with statistical analysis, conclusions, etc. Once we have compiled the data, we will use regression analysis to determine if there is a correlation between the "business" score or academic performance of students and caffeine consumption. We will also perform ANOVA testing to analyze the compositional factors against the different strata, to see if there are any trends to be seen. Further, based on the final results of the survey, we may find a need to do post-stratification or weighting of results, if certain strata were insufficiently represented. The statistical testing here would aim to characterize the strength of any correlation of a student factor with caffeine consumption. We will also discuss the inherent bias in our study, that those who are busy are less likely to respond to the survey, which will skew our estimate of just how busy students are.

looking forward to these analyses!

For now, we will include some preliminary results:

As can be seen here, the survey has so far achieved a total of 228 responses, which corresponds to a 50.0% response rate. We are extremely pleased with this response rate, as we had originally budgeted for only 15%!

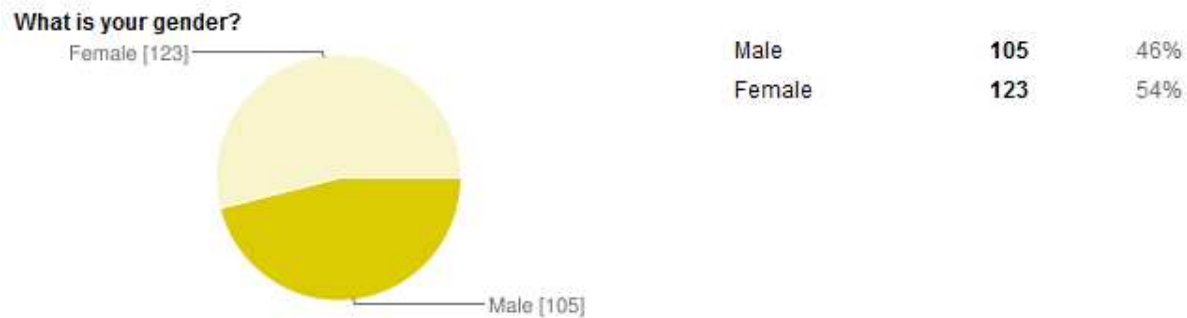
What is your year at CMU?



looks like you fixed up the juniors?

(relative to your talk...)

Interestingly, when we looked at the gender makeup of the respondents, we found that females outnumbered males. However, on campus, males outnumber females by a ratio of 3:2, so this observed respondent ratio is actually the opposite of what we expected. We will see after we perform the data analysis if perhaps there is an underlying reason for this unexpected trend.



Discussion:

This will be filled in after we have performed our data analysis and have conclusions from which to draw. We will give the final conclusions from the survey, and succinctly state all trends or correlations we uncovered between student attributes and caffeine consumption. We will also indicate any unusual or surprising conclusions from our results. After we have seen the project through to its entirety, we will be in a better position to advise future 303 students on what worked, what did not work, and what we would have done differently.

Appendix A: Full Questionnaire:

you can add summary statistics
for each question to this part
also....

1. What is your year at CMU?
 - a. Freshman
 - b. Sophomore
 - c. Junior
 - d. Senior
 - e. Fifth year undergraduate student
2. What college is your primary major in? (For interdisciplinary, select both)
 - a. CIT
 - b. H&SS
 - c. CFA
 - d. SCS
 - e. MCS
 - f. Heinz
 - g. Tepper
3. What is your GPA?
 - a. 0.0-2.5
 - b. 2.6-3.0
 - c. 3.0-3.5
 - d. 3.6-4.0
4. What is your gender?
 - a. Male
 - b. Female
5. How many units are you taking this semester?
 - a. 36 or less
 - b. 37-45
 - c. 45-55
 - d. More than 55
6. Do you have any additional majors or minors?
 - a. Yes
 - b. No
7. If so, which statement best describes your situation?
 - a. I have a minor
 - b. I have more than one minor
 - c. I have an additional major
 - d. I have an additional major and an additional minor
 - e. I have more than one additional major
 - f. I have more than one additional major and more than one minor
8. About how many hours a day, on average, do you spend on extracurricular activities such as clubs, student organizations, ROTC, or inter-mural sports (this includes attending and preparing for meetings or events)?
 - a. None
 - b. Less than 2 hours per day
 - c. 2-4 hours per day

- d. 4-6 hours per day
 - e. More than 6 hours per day
9. How much personal leisure time do you have in a given day? This includes all waking hours spent on activities you consider optional (i.e., NOT class, work, athletics, extracurricular obligations).
- a. None
 - b. Less than 2 hours per day
 - c. 2-4 hours per day
 - d. 4-6 hours per day
 - e. More than 6 hours per day
10. How many hours outside of class per day, on average, do you spend doing classwork?
- a. None
 - b. Less than 2 hours per day
 - c. 2-4 hours per day
 - d. 4-6 hours per day
 - e. More than 6 hours per day
11. How many hours do you sleep on an average week night?
- a. Less than 3 hours
 - b. 3-5 hours
 - c. 5-8 hours
 - d. More than 8 hours
12. Do you have a part-time job (employed on or off campus)?
- a. Yes; I work for less than 5 hours per week.
 - b. Yes; I work for less than 10 hours per week.
 - c. Yes; I work for less than 15 hours per week.
 - d. Yes; I work for more than 15 hours per week.
 - e. No; I do not have a job.
13. How often do you exercise in an average week?
- a. None
 - b. Once a week
 - c. 2-4 times per week
 - d. More than 4 times per week
14. What types of caffeine beverages do you consume? Select all which are relevant.
- a. Coffee
 - b. Tea
 - c. Espresso
 - d. Caffeine tablets
 - e. Redbull/energy drinks
 - f. Soda
 - g. Chocolate
 - h. I do not consume caffeine.
 - i. Other _____
15. How much caffeine do you consume on a daily basis, where 1 serving is one item (eg. One cup of coffee, one espresso shot, one can of soda, one bar of chocolate, etc..)
- a. None
 - b. Less than 2
 - c. 2-4

- d. 4-6
 - e. More than 6
16. What is your primary motivation for consuming caffeine?
- a. Pleasure/Leisure
 - b. To stay awake/maintain energy
 - c. Partying
 - d. Other _____
17. Are you consuming more or less caffeine than you are comfortable with?
- a. I am consuming more caffeine than I am comfortable with.
 - b. I am consuming a comfortable amount of caffeine.
 - c. I am drinking less caffeine than I am comfortable with
18. Do you feel that your personal caffeine consumption is more or less than the average consumption of your undergraduate peers at Carnegie Mellon?
- a. I believe I consume an above average amount of caffeine.
 - b. I believe I consume an average amount of caffeine.
 - c. I believe I consume a below average amount of caffeine.
19. Do you believe that consuming more caffeine positively impacts your performance in school work, employment, athletics, or extracurricular activities?
- a. Yes
 - b. No
20. Do you believe that consuming more caffeine negatively impacts your health?
- a. Yes
 - b. No
21. Do you experience any of the following on a regular basis (on most days)? Choose all that apply.
- a. Insomnia
 - b. Nervousness/anxiety throughout the day
 - c. Restlessness
 - d. Irritability
 - e. Nausea/gastrointestinal problems
 - f. Fast or irregular heartbeat
 - g. Muscle Tremors
 - h. Headaches
 - i. I do not experience any of these symptoms.
22. Do you smoke?
- a. Yes
 - b. No
 - c. I do not wish to disclose
23. How frequently do you consume alcohol?
- a. None
 - b. Less than twice a week
 - c. 2-4 times per week
 - d. More than 4 times per week

Appendix B: Informed Consent Statement:

Confidentiality Policy:

This survey tests student caffeine consumption at Carnegie Mellon University, and is a course requirement for course 36-303: Sampling, Surveys, and Society, taught by Professor Brian Junker. This research project guarantees respondent confidentiality. We conform to all aspects of the Ethics Code of the American Association of Public Opinion Researchers. Most importantly, we guarantee that we will honor clause D-2 of the Ethics Code, namely "Unless the respondent waives confidentiality for specified uses, we shall hold as privileged and confidential all information that might identify a respondent with his or her responses. We shall also not disclose or use the names of respondents for non-research purposes unless the respondent grants us permission to do so."

All data will be used in a form that will make it impossible to determine the identity of the individual responses. That is, the survey responses will not be integrated, analyzed, or reported in any way in which the confidentiality of the survey responses is not absolutely guaranteed. Access to raw data will be tightly restricted to only those individuals directly involved in data analysis. The survey administrators will retain the sole ownership of all raw data.

The final survey report will be made available by request to any interested party. Please email yonggyuc@andrew.cmu.edu or the course advisor, Brian Junker, at brian@stat.cmu.edu if you have any questions about this policy.

Appendix C: Survey Email Distributed by Email

Dear Student,

You have been randomly selected to participate in a campus-wide survey on caffeine usage among Carnegie Mellon undergraduate students. This very brief survey attempts to assess how much caffeine, and what types, students consume. Please take just a minute to complete the survey--all students who complete the survey will be entered into a raffle to win a \$50 Amazon gift card!

If you have trouble viewing or submitting this form, you can fill it out online:

<https://spreadsheets.google.com/viewform?formkey=dHY3cmhvZUVkQ2hQX205RjZVUI92bEE6MA>

Raffle Entry;

If you would like to be entered into the raffle to win the \$50 Amazon gift card, please click the survey link below AFTER YOU SUBMIT your survey responses. We guarantee the confidentiality of your responses, and listing your andrew ID here will not in any way link your name to your survey responses. Click here: https://spreadsheets0.google.com/viewform?hl=en&hl=en&formkey=dERubDhsT2lpWGhMeUhWT_EpCU2RGOUE6MQ#gid=0