36-303: Sampling, Surveys and Society
Exam 1 SOLUTIONS
Thu Feb 17, 2011

1. *[20 pts]* Multiple Choice (4 parts). For each part, circle the roman numeral of the one best answer.

   (a) *[5 pts] Following are some of the most important decisions that can be made as part of designing a survey. Which one* **doesn't belong***?*

   v. None of the above (I.e., all are important decisions in survey design).

   (b) *[5 pts] It is possible to construct a representative sample without random sampling, but it is more difficult to argue that it is really representative.*

   i. True

   (c) *[5 pts]* Beneficence *is a basic ethical principle for research with human subjects. Which of the following is* **not** *an aspect of beneficence?*

   iii. Make sure the subjects get the benefit of some compensation for their participation.

   (d) *[5 pts] Two important fractions in sample surveys are the* sampling fraction n/N *and the* response rate r/n *(where N is the population size, n is the intended sample size, and r is the number in the sample that actually responded). Which of the following is* **not** *true, for a simple random sample without replacement from the target population?*

   iv. You can get a more representative sample by increasing n, regardless of the response rate.

2. *[18 pts] Time-wasting survey...*

   (a) *[6 pts] What is the* **Target Population** *for this survey?*

   The target population appears to be American workers.

   NOTE: Other target populations can get full credit, but the target population should not be confined to people who access AOL and/or salary.com websites.

   (b) *[6 pts] Do you think the* **coverage error** *for this sampling method, for the target population in part (a), is high or low? Give at least one reason for your answer.*

   The coverage error is likely ot be large. The set of workers who access AOL and salary.com websites is not the same as the set of all American workers. For example perhaps these kinds of respondents are already more unhappy with their jobs (which is why they are checking a salary website, e.g.).

   NOTE: Any well-reasoned answer that is consistent with your answer in (a) will get credit.

   (c) *[6 pts] Give two other potential problems with the sampling method chosen for this survey. Use words from class to name the problem, and then use specific ideas from class, and/or evidence from the survey methodology paragraph, to explain what you mean.*

   Problem 1:

   Nonrandom sample. Therefore, there is no guarantee that any of the statistical formulae—about sample size, confidence intervals, etc. that we use in analysis of surveys—are valid.

NOTE: Other answers can get full credit as long as they are well-reasoned.

<u>*Problem 2*</u>*:*

$\boxed{\text{Self-selection bias.}}$ Workers only come to the salary.com website, for example, if they are curious about salaries. In addition, they may waste time on the internet more than other workers (they voluntarily are looking at AOL or salary.com instead of working, right now!).

NOTE: Other answers can get full credit as long as they are well-reasoned.

3. *[18 pts]* [continuation] *Replicate study among 4732 employees (faculty and staff) on the Pittsburgh campus of CMU.*

   (a) *[6 pts] Suppose you are able to get a simple random sample of 100 CMU Pittsburgh employees, and (remarkably) all of them respond to your survey. You find that 40% of the respondents list web surfing as their #1 distraction. Construct a confidence interval for p, the true proportion of Pittsburgh Campus CMU employees for whom web surfing is the #1 distraction.* (Hint: The SD is $\sqrt{p(1-p)}$.)

   NOTE: This should have been for a 95% CI, and a simple random sample without replacement. These corrections were made in class during the exam.

   The CI is $(\hat{p} - ME, \hat{p} + ME)$, where $\hat{p} = 0.40$ and

$$\begin{aligned} ME &= (1.96) \times FPC \times \frac{SD}{\sqrt{n-1}} \\ &= (1.96) \times \sqrt{1 - 100/4732} \times \frac{\sqrt{.5(1-.5)}}{\sqrt{100-1}} \\ &= 0.097 \end{aligned}$$

   where we have taken the "worst case" $SD = \sqrt{p(1-p)} = \sqrt{.5(1-.5)} = .5$.

   Therefore the CI is (0.40 - 0.097, 0.40 + 0.097) or $\boxed{(0.303, 0.497)}$.

   NOTE: Small variations, such as using $2 \approx 1.96$, or using $p = .40$ insteaad of $p = .5$ in the SD calcuation, are fine.

   (b) *[6 pts] Suppose you wanted to get a confidence interval in part (a) with a margin of error of 0.05 (5%). What would the sample size need to be, if you were doing simple random sampling* with replacement*?*

   NOTE: This should have been for a 95% CI. This was corrected in class during the exam.

   The sample size should be at least

$$\begin{aligned} n_0 &= \frac{(1.96)^2 (SD)^2}{ME^2} \\ &= \frac{(1.96)^2 (\sqrt{.5(1-.5)})^2}{(.05)^2} \\ &= 384.16 \end{aligned}$$

   So the sample size should be $\boxed{n = 385}$ (rounding up to the next whole person).

   (c) *[6 pts] Suppose you wanted to get a confidence interval in part (a) with a margin of error of 0.05 (5%). What would the sample size need to be, if you were doing simple random sampling* <u>without</u> replacement*?*

NOTE: This should have been for a 95% CI. This was corrected in class during the exam.

The sample size should be at least

$$\frac{Nn_0}{N+n_0} = \frac{(4732)(384.16)}{4732+384.16} = 355.31$$

so the sample size should be at least $\boxed{n = 356}$, again rounding up to the nearest whole person.

4. *[20 pts] Informed consent situations...*

NOTE: For each problem below I give a possible "good" answer. Other "good" answers could get full credit, if they are well-reasoned, well-written, and include all relevant plausible assumptions, etc.

(a) *[5 pts] As a part of a face-to-face survey, the interviewer is asked to observe how many books are visible in the respondent's living room.*

__ Approve.
X Disapprove. *Please Change the Following:*

Please obtain informed consent before counting the books, and make sure the respondent knows that the book count will be part of the data collected.

Reason:

Informed consent should be obtained before collecting *any* data from a respondent.

NOTE: If you assumed informed consent had already been obtained prior to counting the books, it would be OK to approve this study.

(b) *[5 pts] As part of a telephone survey, the first few moments of the interaction between the interviewer and the potential respondent are recorded, for use as examples in teaching other interviewers.*

__ Approve.
X Disapprove. *Please Change the Following:*

Recording should not begin until informed consent is obtained (and the informed consent request should include the information that the rest of the interview will be recorded).

Training tapes would need to be obtained in some other way.

Reason:

Informed consent should be obtained before collecting *any* data from a respondent.

(c) *[5 pts] Before approaching a parent-child pair for a face-to-face interview to study parent-child interations, the interviewer makes structured observations of how the parent and child interact in a public park. The observations are used to select parent-child pairs for the sample, and become part of the data collected for each parent-child pair in the sample.*

__ Approve.

X Disapprove. *Please Change the Following:*

Parent-child pairs should be selected for the sample, using a method involving a prior rule that is not based on any information about the individual respondents. For example: “choose every 5^{th} parent-child pair”.

Reason:

Again, informed consent should be obtained before collecting *any* data from a respondent.

In addition, a weakness of this design is that there can be interviewer bias in selecting respondents if a prior rule is not employed (this is not an informed consent problem, but it is a weakness in the study design).

(d) [5 pts] As part of a study of whether voter intentions are affected by the sponsor of the survey, some respondents are told that the sponsor is an organization different from the one actually conducting the survey.

NOTE: Here are two possible full-credit answers. (Other well-reasoned and well-grounded answers would also be acceptable.)

__ Approve.
X Disapprove. *Please Change the Following:*

The study cannot be done as proposed.

Reason:

Informed consent cannot be obtained if potential respondents are misinformed.

__ Approve.
X Disapprove. *Please Change the Following:*

You must debrief the respondent after the interview, revealing the misdirection and the purpose of the study.

Reason:

If the goal of the study is to measure whether voter intentions are affected by the sponsor of the survey, then this misdirection is the only way to get to the goal. We can minimize the risks of misdirection, relative to the benefits, by informing the respondents after the fact. [However, note that misdirection studies are typically much harder to get approved by IRB’s!]

5. *[24 pts] Revising survey questions...*

(a) “What was your income in the past year?
Answer here: ____________________”

(b) [3 pts] A Potential Problem:
Any one of the following (or other well-reasoned answers) could receive credit:

- The term “income” is not well-defined.
- The term “in the past year” is not well-defined.

(c) [3 pts] Suggestion(s) For Rewrite:

Any one of the following (or other well-reasoned answers) could receive credit:

- Rewrite to specify more clearly what counts as “income”
- Rewrite to clarify which dates should be included in “in the past year”.
- Rewrite as a series of “bounded recall” questions:
  - Are you paid an annaul salary or houry?
  - If hourly, how much did you earn in the past week? Was that typical? If not, what are your typical weekly earnings? [then the survey worker can multiply by 50 or 52 weeks if desired.]
  - If salary, what is your base annual salary? Are you also elegible for bonuses or other increases? If so, did you receive any since Feb 1, 2010?

  etc.

6. “Do you favor or oppose universal health care in the United States?”

(a) [3 pts] A Potential Problem:

One possible answer: this question is too general. we don’t know exactly what is going through respondent’s mind when he/she answers.

(b) [3 pts] Suggestion(s) For Rewrite:

One possible fix: focus on a specific universal health care option: “As you know, the *Affordable Health Care For America Act* which was passed by Congress in 2009 and signed into law by President Obama makes the following provisions

- prohibiting health insurers from refusing, or charging differently for, coverage based on patients’ medical histories or gender
- requiring most employers to provide, and most individuals to carry, coverage meeting certain minimum standards, or pay a surtax for non-compliance

Do you generally support or oppose this approach to universal health care?”

7. “I am happier than usual these days. (Circle the best answer below)

1 2 3 4 5

Strongly agree Strongly disagree.”

(a) [3 pts] A Potential Problem:

Two possible problems (either alone, or some other well-reasoned answer, could get full credit):

- We usually don’t know what is going through respondents’ minds with general, as opposed to specific, questions.
- Middle categories (#3 above) in Likert items gives respondents a chance to “weasel out” of answering the question.

(b) [3 pts] Suggestion(s) For Rewrite:

Some possible fixes (either alone, or some other well-reasoned answer, could get full credit):

- Precede the general question with some specific ones: “I am generally happier in my work these days”... “I am generally happier at home these days”... “I am generally happier with my spouse/significant other/family these days”... etc.

- Focus the time period: "I am generally happier in the past month than I have been since the beginning of 2010"... etc.
- Take out the middle category (#3) so respondents are forced to take an agree or disagree choice.

8. "What kind of music do you listen to when you study?

__ Rap.	__ Classic Rock.
__ Classical.	__ Industrial Music.

(a) [3 pts] A Potential Problem:

Two possible problems (either alone, or some other well-reasoned answer, could get full credit):

- Not enough different categories of music.
- Closed questions may fail to get the "true" response if it is missing among the many categories listed.

(b) [3 pts] Suggestion(s) For Rewrite:

Two possible fixes (either alone, or some other well-reasoned answer, could get full credit):

- Add more categories of music.
- Add an open-ended "__ Other, please write type of music here: ________" option at the end.