Design and Analysis of Sample Surveys, final exam

7 May 2012

1. Suppose that, in a survey of 1000 people in a state, 400 say they voted in a recent primary election. Actually, though, the voter turnout was only 30%. Give an estimate of the probability that a nonvoter will falsely state that he or she voted. (Assume that all voters honestly report that they voted.)

2. Which of the following are useful goals in a pilot study? (Indicate all that apply.)

   (a) You can search for statistical significance, then from that decide what to look for in a confirmatory analysis of your full dataset.

   (b) You can see if you find statistical significance in a pre-chosen comparison of interest.

   (c) You can examine the direction (positive or negative, even if not statistically significant) of comparisons of interest.

   (d) With a small sample size, you cannot hope to learn anything conclusive, but you can get a crude estimate of effect size and standard deviation which will be useful in a power analysis to help you decide how large your full study needs to be.

   (e) You can talk with survey respondents and get a sense of how they perceived your questions.

   (f) You get a chance to learn about practical difficulties with sampling, nonresponse, and question wording.

   (g) You can check if your sample is approximately representative of your population.

3. We discussed in class the best currently available method for estimating the proportion of military servicemembers who are gay. What is that method? (Recall the problems with the direct approach: there is no simple way to survey servicemembers at random, nor is it likely that they would answer such a question honestly.)

4. Researchers have found that survey respondents overreport church attendance. Thus, naive estimates from surveys overstate the percentage of Americans who attend church regularly. Does this have a large impact on estimates of *time trends* in religious attendance?

5. Which of the following better describes changes in public opinion on most issues? (Choose only one.)

   (a) *Dynamic stability*: On any given issue, average opinion remains stable but liberals and conservatives move back and forth in opposite directions (the "accordion model")

   (b) *Uniform swing*: Average opinion on an issue can move but the liberals and conservatives don't move much relative to each other (the disribution of opinions is a "solid block of wood")

   (c) *Compensating tradeoffs*: When considering multiple survey questions on the same general topic, average opinion can move sharply to the left or right on individual questions while the average over all the questions remains stable (the "rubber band model")

6. A survey of New York City residents is performed using cluster sampling. The design effect is 3.0. From the survey, the estimated proportion who prefer the Mets to the Yankees is 0.42 with a standard error of 0.05. How many people were in the sample?

7. Which of the following statements accurately summarize claims made by Page and Shapiro in *The Rational Public* and their associated research articles? (Indicate all that apply.)

   (a) Americans' attitudes on policy alternatives are highly unstable over time, reflecting a rational response to unstable political conditions.

   (b) When studying public opinion, question-wording is less important than scholars have traditionally thought.

   (c) Attitudes about foreign policy change more abruptly than attitudes on domestic issues.

   (d) The contents of the mass media account for a high proportion of opinion changes on foreign policy.

   (e) Using the assumption of rationality, Page and Shapiro fit a hedonic regression to estimate the underlying utility function of survey respondents.

   (f) Page and Shapiro use the term "rational" ironically; their fundamental claim is that Americans are easily distracted and that rational-public models are seriously flawed.

8. Which of the following statements accurately characterize the National Election Studies? (Indicate all that apply.)

   (a) The NES began in 1960.

   (b) Since 1980, the NES has mostly relied on telephone interviews.

   (c) The NES typically has a sample size of about 1000–2000 people.

   (d) The NES uses a sampling design that ensures they get respondents from all fifty states and D.C.

9. Out of a population of 100 medical records, 40 are randomly sampled and then audited. 10 out of the 40 audits reveal fraud. From this information, give an estimate, standard error, and 95% confidence interval for the proportion of audits in the population with fraud.

10. Out of a random sample of 100 Americans, zero report having ever held political office. From this information, give a 95% confidence interval for the proportion of Americans who have ever held political office.

11. Here is the result of fitting a logistic regression to Republican vote in the 1972 NES.

```
glm(formula = vote ~ income, family=binomial(link="logit"))
            coef.est coef.se
(Intercept)   -1.40    0.19
income         0.33    0.06
  n = 1179, k = 2
  residual deviance = 1556.9, null deviance = 1591.2 (difference = 34.3)
```

Income is on a 1–5 scale. Approximately how much more likely is a person in income category 4 to vote Republican, compared to a person in income category 2? Give an approximate estimate, standard error, and 95% interval.

12. A researcher fits a regression model predicting some political behavior given predictors for demographics and several measures of economic ideology. The coefficients for the ideology measures are not statistically significant, and the researcher creates a new measure, adding up the ideology questions and creating a common score, and then fits a new regression including

the new score and removing the individual ideology questions from the model. Which of the following statements are basically true? (Indicate all that apply.)

(a) If the original ideology measures are close to 100% correlated with each other, there will be essentially no benefit from this approach.

(b) If the original ideology measures are not on a common scale, they should be rescaled before adding them up.

(c) If the original result was not statistically significant, the researcher should stop, so as to avoid data dredging and selection bias.

(d) Another reasonable option would be to perform a factor analysis on the ideology measures and create a common score in that way.

13. A survey of American adults is conducted that includes too many women and not enough men in the sample. In the resulting weighting, each female respondent is given a weight of 1 and each male respondent is given a weight of 1.5. The sample includes 600 women and 380 men, of whom 400 women and 100 men respond Yes to a particular question of interest. Give an estimate and standard error for the proportion of American adults who would answer Yes to this question if asked.

14. A public health survey of elderly Americans includes many questions, including "How many hours per week did you exercise in your most active years as a young adult?" and also several questions about current mobility and health status. Response rates are high for the questions about recent activitiies and status, but there is a lot of nonresponse for the question on past activity. You are considering imputing the missing values on the question, "How many hours per week did you exercise in your most active years as a young adult?" Which of the following statements are basically correct? (Indicate all that apply.)

(a) If done reasonably well, imputation is preferred to available-case and complete-case analysis.

(b) If you do impute, you should also present the available-case and complete-case analysis and analyze how the imputed estimates differ.

(c) It is OK to include current health status variables as predictors in a model imputing past activities: anything that adds information is good when imputing.

(d) It is probably not a good idea to include current health status variables as predictors in a model imputing past activities: current health is possibly influenced by past activities, and including a casual outcome can bias estimates of a treatment variable.

(e) If you fit a regression model and impute your best prediction for each person (rather than imputing random draws from the predictive distribution), you can have problems because you will be more likely to impute extreme values.

(f) It is a good idea to fit a logistic regression predicting response/nonresponse to the question of interest as a way to look for systematic differences between respondents and nonrespondents on this question.

15. A researcher conducts a random-digit-dial survey of individuals and married couples. The design is as follows: if only one person lives in a household, he or she is interviewed. If there are multiple adults in the household, one is selected at random: he or she is interviewed and, if he or she is married to one of the other adults in the household, the spouse is interviewed

as well. Come up with a scheme for inverse-probability weights (ignoring nonresponse and assuming there is exactly one phone line per household).

16. You are doing a survey in a war-torn country to estimate what percentage of unemployed men support the rebels in a civil war. Express this as a ratio estimation problem, where goal is to estimate $\bar{Y}/\bar{X}$. What are x and y here? Give the estimate and standard error for the percentage of unemployed men who support the rebels.

17. In a survey of n people, half are asked if they support "the health care law recently passed by Congress" and half are asked if they support "the law known as Obamacare." The goal is to estimate the effect of the wording on the proportion of Yes responses. How large must n be for the effect to be estimated within a standard error of 5 percentage points?

18. A survey is taken of 100 undergraduates, 100 graduate students, and 100 continuing education students at a university. Assume a simple random sample within each group. Each student is asked to rate his or her satisfaction (on a 1–10 scale) with his or her experiences. Write the estimate and standard error of the average satisfaction of all the students at the university. Introduce notation as necessary for all the information needed to solve the problem.

19. A survey is taken of students in a metropolitan area. At the first stage a school is sampled at random. The schools are divided into two strata: 20 private schools and 50 public schools are sampled. At the second stage, 5 classes are sampled within each sampled school. At the third stage, 10 students are sampled within each class. What is the probability that any given student is sampled? Express this in terms of the number of students in the class, number of classes in the school, and number of schools in the area. Define appropriate notation as needed.

20. Explain in two sentences why we expect survey respondents to be honest about vote preferences but possibly dishonest about reporting unhealty behaviors.

21. A country is divided into three regions with populations of 2 million, 2 million, and 0.5 million, respectively. A survey is done asking about foreign policy opinions. Somebody proposes taking a sample of 50 people from each region. Give a reason why this non-proportional sample would not usually be done, and also a reason why it might actually be a good idea.

22. A supermarket chain has 100 equally-sized stores. It is desired to estimate the proportion of vegetables that spoil before being sold. Three stores are selected at random and are checked: the percent of spoiled vegetables are 3%, 5%, and 10% in the three stores. Give an estimate and standard error for the percentage of spoiled vegetables for the entire chain.

23. Suppose you are conducting a survey in which people are asked about their health behaviors (how often they wash their hands, how often they go to the doctor, etc.). There is a concern that different interviewers will get different sorts of responses—that is, there may be important interviewer effects. Describe (in two sentences) how you could estimate the interviewer effects within your survey. Can the interviewer effects create problems of reliability of the survey responses? Explain (in one sentence). Can the interviewer effects create problems of validity of the survey responses? Explain (in one sentence).

24. A supermarket chain has 100 equally-sized stores. It is desired to estimate the proportion of vegetables that spoil before being sold. The following sampling designs are considered:

(a) Sample 10 stores, then sample half the vegetables within each of these stores; or

(b) Sample 20 stores, then sample one-quarter of the vegetables within each of these stores.

Which of these designs has the lowest variance? Why might the higher-variance design still be chosen?

25. You are using multilevel regression and poststratification (MRP) to a survey of 1500 people to estimate support for the space program, by state. The model is fit using, as a state-level predictor, the Republican presidential vote in the state, which turns out to have a low correlation with support for the space program. Which of the following statements are basically true? (Indicate all that apply.)

(a) For small states, the MRP estimates will be determined almost entirely by the demographic characteristics of the respondents in the sample from that state.

(b) For small states, the MRP estimates will be determined almost entirely by the demographic characteristics of the population in that state.

(c) Adding a predictor specifically for this model (for example, a measure of per-capita space-program spending in the state) could dramatically improve the estimates of state-level opinion.

(d) It would not be appropriate to add a predictor such as per-capita space-program spending in the state: by adding such a predictor to the model, you would essentially be assuming what you are trying to prove.

26. You have just graded an an exam with 28 questions and 15 students. You fit a logistic item-response model estimating ability, difficulty, and discrimination parameters. Which of the following statements are basically true? (Indicate all that apply.)

(a) If a question is answered correctly by students with very low and very high ability, but is missed by students in the middle, it will have a high value for its discrimination parameter.

(b) It is not possible to fit an item-response model when you have more questions than students. In order to fit the model, you either need to reduce the number of questions (for example, by discarding some questions or by putting together some questions into a combined score) or increase the number of students in the dataset.

(c) To keep the model identified, you can set one of the difficulty parameters or one of the ability parameters to zero and set one of the discrimination parameters to 1.

(d) If two students answer the same number of questions correctly, they will have the same estimated ability parameter.

(e) Under the model, if a student's ability parameter has the same value as a particular question's difficulty parameter, there is a 50% chance the student will get the question right.

27. Which of the following problems were identified with the Burnham et al. survey of Iraq mortality? (Indicate all that apply.)

(a) The survey used cluster sampling, which is inappropriate for estimating individual outcomes such as death.

(b) In their report, Burnham et al. did not identify their primary sampling units.

(c) The second-stage sampling was not a probability sample.

(d) Survey materials supplied by the authors are incomplete and inconsistent with published descriptions of the survey.

28. A telephone survey was conducted several years ago, asking people how often they were polled in the past year. I can't recall the responses, but suppose that 40% of the respondents said they participated in zero surveys in the previous year, 30% said they participated in one survey, 15% said two surveys, 10% said three, and 5% said four. From this it is easy to estimate an average, but there is a worry that this survey will itself overrepresent survey participants and thus overestimate the rate at which the average person is surveyed. Come up with a procedure to use these data to get an improved estimate of the average number of surveys that a randomly-sampled American is polled in a year.