36-303: Sampling, Surveys and Society
Exam 1
Thu Feb 18, 2010

- You have 80 minutes for this exam.
- The exam is closed-book, closed notes.
- A calculator is allowed.
- **A formula sheet is provided on the next page for your convenience.**
- Please write all your answers on the exam itself; your work must be your own.

Question	Points Possible	Points Earned
1	20	
2	18	
3	18	
4	20	
5	24	
Total	100	

Name: _____

Signature: _____

Some Useful Formulas From the Statistics of Survey Sampling

Equally-Likely Outcomes & Counting

- If K outcomes $O_1, \ldots, O_K$ are equally likely, then the probability of any one of them is $1/K$.
- Consider taking a sample of n objects from a population of N objects.
  - Sampling with replacement, there are N^n possible samples of size n; the probability of any one of them is $1/N^n$.
  - Sampling without replacement, there are $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ possible samples of size n [where $N! = N \cdot (N-1) \cdot (N-2) \cdots 3 \cdot 2 \cdot 1$], so the probability of any one of them is $1 \Big/ \binom{N}{n}$.

Discrete Random Variables

Let X and Y be random variables with sample spaces $\{x_1, \ldots, x_K\}$ and $\{y_1, \ldots, y_K\}$ and distributions

$$P[X = x_i, Y = y_j] = p_{ij} \quad , \quad P[X = x_i] = p_{i\cdot} = \sum_{j=1}^{K} p_{ij} \quad , \quad P[Y = y_j] = p_{\cdot j} = \sum_{i=1}^{K} p_{ij}$$

Then, for example

$$E[X] = \sum_{i=1}^{K} x_i p_{i\cdot} \quad , \quad Var(X) = \sum_{i=1}^{K} (x_i - E[X])^2 p_{i\cdot} \quad , \quad Cov(X,Y) = \sum_{i=1}^{K} (x_i - E[X])(y_i - E[Y]) p_{ij}$$

$$P[X = x_i | Y = y_j] = p_{ij}/p_{\cdot j} \quad , \quad E[X|Y = y_j] = \sum_{i=1}^{K} x_i P[X = x_i | Y = y_j] \quad , \quad E[aX + bY + c] = aE[X] + bE[Y] + c$$

Random Sampling From a Finite Population

Consider a population of size N and a sample of size n. Let y_i be the (fixed) values of some variable of interest in the population (such as a person's age, or whether they would vote for Obama). Let

$$Z_i = \begin{cases} 1, & \text{if } i \text{ is in the sample} \\ 0, & \text{else} \end{cases}$$

be the random sample inclusion indicators, and let Y_i be the random observations in the sample. Then the sample average can be written

$$\overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \frac{1}{n} \sum_{i=1}^{N} Z_i y_i$$

The Z_i's are Bernoulli random variables with

$$E[Z_i] = \frac{n}{N} \quad , \quad Var(Z_i) = \frac{n}{N}\left(1 - \frac{n}{N}\right) \quad , \quad Cov(Z_i, Z_j) = -\frac{1}{N-1}\frac{n}{N}\left(1 - \frac{n}{N}\right)$$

Confidence Intervals and Sample Size

(a) A CLT-based $100(1-\alpha)\%$ confidence interval for the population mean is $(\overline{Y} - z_{\alpha/2} SE \ , \ \overline{Y} + z_{\alpha/2} SE)$.

(b) For sampling with replacement from an infinite population, $SE = SD/\sqrt{n}$.

(c) For sampling without replacement from a finite population, the SE has to be multiplied by the finite population correction (FPC).

(d) For a given margin of error (ME, half the width of the CI) and confidence level $1-\alpha$, we can find the sample size by solving

$$z_{\alpha/2} SE < ME$$

for n. The same approach works for both SRS with replacement (using the SE in (b)) and SRS without replacement (using the SE in (c)).

February 18, 2010 Name: _____

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

   (a) [5 pts] Which of the following statements is **most correct**:

   i. If you do not have a random sample, it cannot be a representative sample.
   ii. The only valid random sampling for surveys is sampling without replacement.
   iii. It is possible to construct a representative sample without random sampling, but it is more difficult to argue that it is really representative.
   iv. If the sample is random, it is representative, regardless of the response rate.

   (b) [5 pts] Requiring researchers to obtain *Informed Consent* from participants in surveys and other human subjects research is considered an ethical obligation (and sometimes a legal one). Which one of the following is the **best** justification for informed consent for survey participants, according to our textbook:

   i. To protect respondents from harm.
   ii. To give respondents meaningful control over information about themselves.
   iii. To give respondents a chance to understand the research they will be participating in.
   iv. To make sure that the respondents benefit from the survey by getting a copy of the final report.

   (c) [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 with replacement from the target population?

   i. You can get a more representative sample by increasing n, regardless of the response rate.
   ii. To decrease variability in sample estimates, increase the sampling fraction.
   iii. To decrease possible bias in sample estimates, increase the response rate.
   iv. You can force the standard error of sample estimates to be zero by making the sampling fraction large enough.

   (d) [5 pts] When making a public report on a survey, which of the following is **not** required?

   i. Who sponsored it, who carried it out.
   ii. Sample size and precision (SE) of estimates.
   iii. The name of the statistical package used to do the analyses.
   iv. Target population, sampling frame, sampling method, response rates.

2. [18 pts] According to the blog `http://nullspace2.blogspot.com` (about economic issues in the greater Pittsburgh area), there are 2,505 full-time police officers employed by all Allegeny local governments (the county government, the city of Pittsburgh, and surrounding municipalities such as Monroeville, Mount Lebanon, Wilkinsburg, etc.). Your survey team is going to survey police attitudes toward their work. One of your survey questions is

   > *Should persons taken into custody always be told their Miranda rights ('You have the right to remain silent; anything you say can and will be used against you in a court of law [etc.]') Yes / No".*

   Answer the following questions (3 parts).

   (a) [6 pts] Let p be the proportion of police officers that would say yes. Assuming you are doing SRS without replacement, how large must your sample size be, to estimate p within a margin of error of ± 0.07, with a 95% confidence interval?

(b) [6 pts] It turns out you only have enough resources (time and money to collect data and process the responses, mainly) to do a phone survey of 50 police officers, chosen at random from a complete list of all 2,505 officers. Amazingly, all 50 officers reply, and of those, 20 respond "yes" to the question above. Compute a 95% confidence interval for p.

(c) [6 pts] Now suppose that you call a random sample of 200 officers from the complete list of 2,505 officers, but only 50 agree to be interviewed over the phone. Of those 50, 20 respond "yes". You calculate the same confidence interval as in part (b).

Which of the confidence intervals (this one, or the one in part (b)) is harder to believe as evidence about the whole population of Allegheny full-time police officers? Why?

3. [18 pts] *Simple Random Sampling (3 parts).*

   (a) [6 pts] Carefully define *Simple Random Sampling (SRS)* **with** *replacement*, and give an example. Give enough detail that it is obvious that this is a good example.

   (b) [6 pts] Carefully define *Simple Random Sampling (SRS)* **without** *replacement*, and give an example. Give enough detail that it is obvious that this is a good example.

(c) [6 pts] Suppose we take a SRS without replacement of n individuals from a population of N. Let y_i be the height of the i^{th} person. Let

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{N} Z_i y_i$$

be the sample average height, and let

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i$$

be the mean height in the population. Show that $\hat{Y}$ is an unbiased estimator for μ. In other words, show that

$$E[\hat{Y}] = \mu .$$

4. [20 pts] As part of their course assignment, undergraduate students in introductory psychology classes are sent to various areas in New York City to ask 1,520 passersby one of a variety of simple requests. They are told to ask for different kinds of help (e.g. asking for directions, asking for change for a $1 bill, etc.) and to ask for it in different ways. The varying responses to their requests provide a first approximation to answering questions about the prevalence of "altruistic compliance" (roughly, willingness to help with a stranger's requests) and the factors influencing it.

   Answer the following questions (3 parts).

   (a) [6 pts] What, if any, ethical principles do you think this study violates? (circle the roman numeral of the one best answer)

      i. Beneficence
      ii. Justice
      iii. Respect for Persons; Confidentiality
      iv. Informed Consent
      v. None of the above; it's just fine.
      vi. Not listed above; I think _____

      _____.

   (b) [8 pts] Do you think the results of this survey will be representative of the degree of altruistic compliance of all New Yorkers?

      ___ **Yes, representative.** If you select this answer, use the space below to explain why it is representative.

      ___ **No, not representative.** If you select this answer, use the space below to suggest one or more modifications of the survey design that would make it more representative.

   Be clear and complete, not just a word or two. Continue onto the next page if you need to.

(more room for your answer to (b))

(c) [6 pts] Sometimes a study that is not perfect can still be useful. List one or two possible advantages of doing the study as designed, rather than modifying it in any way. *Be clear and complete, not just a word or two.*

5. [24 pts] Below are several survey questions. For each question: (i) indicate a potential problem with the question *using specific ideas from the lecture notes on question writing*; (ii) suggest a way to rewrite it (as one or more questions, by providing more information, by improving grammar, etc.) that gets at the same thing while avoiding the problem you raised.

   (a) *"Most people think that abortion is immoral. Do you agree?"*

      i. [3 pts] A Potential Problem:

      ii. [3 pts] Suggestion(s) For Rewrite:

   (b) *"Do you agree or disagree that NCLB should be abolished and replaced with merit pay for teachers?"*

      i. [3 pts] A Potential Problem:

      ii. [3 pts] Suggestion(s) For Rewrite:

(c) *"How many times have you taken a PAT bus to go shopping since the start of classes at CMU in August? Answer here: _____"*

i. [3 pts] A Potential Problem:

ii. [3 pts] Suggestion(s) For Rewrite:

(d) *"What kind of exercise do you do?*

_ *Walking, jogging or running*

_ *Exercise machines or weights"*

i. [3 pts] A Potential Problem:

ii. [3 pts] Suggestion(s) For Rewrite: