# 36-303: Sampling, Surveys and Society
# Exam 1
# Thu March 1, 2012

- 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.** You may tear it out of the exam if you like.
- Please write all your answers on the exam itself; your work must be your own.

| Question | Points Possible | Points Earned |
|---|---|---|
| 1 | 20 | |
| 2 | 20 | |
| 3 | 20 | |
| 4 | 25 | |
| 5 | 15 | |
| 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)).

March 1, 2012 Name: _____

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

   (a) [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.

   (b) [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.

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

      i. Maximize possible benefits and minimize possible harms, to subjects. the
      ii. Decide when to do a study because the benefits outweigh the risks to subjects; and when to forego a study because the risks outweigh the benefits.
      iii. Make sure the subjects get the benefit of some compensation for their participation.
      iv. None of the above (I.e., all are aspects of beneficence).

   (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?

      i. You can force the standard error of sample estimates to be zero by making the sampling fraction large enough, as long as the response rate is equal to 1.
      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 get a more representative sample by increasing $n$, regardless of the response rate.

2. [20 pts] You are going to test the following survey items in cognitive interviews with one person outside your project team. For each item below, indicate (by marking the appropriate blank) whether you think the item is OK as-is, and if not, what you would ask the person about the item, or about the things the person is thinking as he/she answers the item (to give you ideas to improve it).

   (a) [5 pts] During the past four weeks, beginning on [date 4 weeks ago] and ending today, have you done any exercise, including sports, physically active hobbies, and aerobic exercises, but not including any activities carried out as part of your job or in the course of ordinary housework?

   ___ This item is OK as-is.

   ___ This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:

   (b) [5 pts] How many times a week do you have milk, butter or other dairy products?

   ___ This item is OK as-is.

   ___ This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:

*[continued on next page. . . ]*

(c) [5 pts] Some people feel the US federal government should take action to reduce the national debt, even if it means unemployment would go up a lot. Others feel the government should take action to reduce the rate of unemployment even if it means the national debt would go up a lot. Where would you place yourself on the following 7-point scale?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Reduce Debt | | | | | | Reduce Unemployment |

___ This item is OK as-is.

___ This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:

(d) [5 pts] During the past 12 months, since [date], about how many days did illness or injury keep you in bed more than half the day? Include days while you were an overnight patient in a hospital.

___ This item is OK as-is.

___ This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:

3. [20 pts] The American Hospital Association (AHA) conducted a survey on drug shortages in hospitals, in summer 2011 In a powerpoint press release (see http://www.aha.org/content/11/drug-shortagesurvey.pdf), the survey methodology was described as follows:

   - Survey was sent to all community hospital CEOs on June 1, 2011 via fax and e-mail.
   - Data were collected through June 22, 2011.
   - Responses from 820 hospitals are included in analysis.
   - Respondents were broadly representative of the universe of community hospitals.
   - Survey questions were designed to assess the impact of drug shortages on patients and hospitals.
   - Nationally, there are about 2,800 urban hospitals, 1,300 critical access hospitals and 1,000 other rural hospitals.

   Please answer the following questions about this survey. If the information cannot be inferred from the bullets above, say so.

   (a) [2 pts] What is the *target population* for this survey?

   (b) [2 pts] What are the *units* in the target population?

   (c) [2 pts] What is the *sampling frame*?

   (d) [2 pts] What *sampling method* was used?

   (e) [2 pts] What *method of data collection* was used to contact units in the sample and collect data from them?

*[continued on next page. . . ]*

(f) The press release reports that "99.5% of hospitals reported experiencing one or more drug shortage in the last six months and nearly half of the hospitals reported 21 or more drug shortages."

i. [5 pts] Do you think this is useful information for the AHA? Why or why not?

ii. [5 pts] Do you think this is a scientifically valid inference about the target population? Explain carefully why or why not, using *at least two* concepts from the lecture notes or textbook.

4. [25 pts] Now let's assume that the AHA survey is a *simple random sample (SRS) without replacement*, from a population of size 5,100 (regardless of any answer you gave in problem 3).

   (a) [10 pts] 17% of hospitals in the sample reported that patient treatment was delayed because of a drug shortage. Calculate a 95% confidence interval for the same percentage in the population, based on the 820 responses in the sample.

*[continued on next page. . . ]*

(b) [5 pts] What is the margin of error (ME) in part (a)?

(c) [10 pts] Suppose we only want a ME of 0.01. What should the sample size be?

5. [15 pts] Figure 1 (page 9 of this test) displays an informed-consent form for a web survey.

   (a) [12 pts] Identify (using the paragraph numbers in Figure 1) any four of the six essential elements for an informed consent form (by filling out the table below):

| **Name of Element** | **Number of Paragraph Where it Appears** |
|---|---|
| | |
| | |
| | |
| | |

   (b) [3 pts] Identify an element of informed consent forms that is missing from Figure 1.

**Welcome to the CAREGIVING: STRESS AND SUPPORT Survey (HUM00001234)**

**(1)** Dr. John Jones and Dr. Sarah Smith, of the University of Michigan Department of Psychology, invite you to be a part of a research study that looks at the stresses that people experience when they are providing care to a seriously ill family member. The purpose of the study is to design better support programs for caregivers. We are asking you to attend because you recently attended a meeting of a UM Caregivers Support Group.

**(2)** If you agree to be part of the research study, you will be asked to complete an online survey about your experiences as a caregiver. We expect this survey to take 30 to 45 minutes to complete.

**(3)** Although you may not receive any direct benefit from participating, we hope that this study will contribute to the improvement of social support systems for those who provide care to others.

**(4)** Your responses to this survey are anonymous, meaning that the researchers will not be able to link your survey resposes to you. The survey software does not collect identifying information about you or your computer. We plan to publish the results of this study, but we will not include any information that would identify you.

**(5)** Participating in this survey is completely voluntary. Even if you decide to paticipate now, you may change your mind and stop at any time. You may choose not to answer a question, or skip any part of the survey. Simply click "Next" at the bottom of the survey page to move on to the next question.

**(6)** The University of Michigan Behavioral Sciences Institutional Review Board has determined that this study is exempt from IRB oversight.

**(7)** By clicking on the link below, you are consenting to participate in this research study.

**www.caregivingsurvey.net**

**If you do not wish to participate, click the "x" in the top right corner of your browser to exit.**

Figure 1: Informed consent form for problem 5.