EPSY 5244: Survey Design, Sampling, & Implementation Sampling Assignment

**Simple Probability Samples**

1. Discuss whether an SRS would be appropriate for the following situations. What other designs might be used?

a. For an e-mail survey of students, you have a sampling frame that contains a list of e-mail addresses for all students.

b. You want to estimate the percentage of topics in a medical website that have errors.

c. A county election official wants to assess the accuracy of the machines that counts the ballots by taking a sample of the paper ballots and comparing the estimated vote tallies for candidates from the sample to the machine counts.

2. The Special Census of Maricopa County, Arizona, gave 1995 populations for four cities:

City Population Corrected Sample Size *n*'

Buckeye 4,857 .

Gilbert 59,338 .

Gila Bend 1,724 .

Phoenix 1,149,417 .

Suppose that you are interested in estimating the percentage of persons who have been immunized against polio in each city and can take an SRS of persons. What should your sample size be in each of the 5 cities if you want the estimate from each city to have margin of error of 4 percentage points? For which cities does the finite population correction make a difference? *HINT: Use the sample size calculator.*

3. Estimating the size of an audience. A common method for estimating the size of an audience is to take an SRS of *n* of the *N* rows in an auditorium, count the number of people in each of the selected rows, then multiply the total number of people in your sample by *N*/*n*. Why is it important to take an SRS instead of a convenience sample of the first 10 rows?

**Stratified Sampling**

4. What stratification variable(s) would you use for each of the following situations:

a. A political poll to estimate the percentage of registered voters in Arizona that approve of the governor’s performance.

b. An e-mail survey of students at your university, to estimate the total amount of money students spend on textbooks in a term.

c. A sample of high schools in New York City to estimate what percentage of high schools offer one or more classes in computer programming.

d. A sample of public libraries in California to study the availability of computer resources, and the per capita expenditures.

e. A survey of anglers (people who fish) visiting a freshwater lake, to learn about which species of fish are preferred.

**Cluster Sampling**

5. A city council of a small city wants to know the proportion of eligible voters that oppose having an incinerator of Phoenix garbage opened just outside of the city limits. They randomly select 100 residential numbers from the city’s telephone book that contains 3000 such numbers. Each selected residence is then called and asked for (a) the total number of eligible voters and (b) the number of voters opposed to the incinerator. A total of 157 voters were surveyed; of these, 23 refused to answer the question. Of the remaining 134 voters, 112 opposed the incinerator, so the council estimates the proportion by

$$\hat{p}\_{SRS}=\frac{112}{134}=.83582$$

with

$$\hat{V}(\hat{p})=\frac{.83582(1-.83852)}{134}= .00102$$

Are these estimates valid? Why, or why not?

6. Senturia et al. (1994) describe a survey taken to study how many children have access to guns in their households. Questionnaires were distributed to all parents who attended selected clinics in the Chicago area during a one-week period for well or sick child visits.

* 1. Suppose that the quantity of interest is percentage of the households with guns. Describe why this is a cluster sample. What is the psu? The ssu? Is it a one-stage or two-stage cluster sample?
	2. What is the sampling population for this study? Do you think this sampling procedure results in a representative sample of households with children? Why, or why not?