**MSDS 6370 Statistical Sampling Midterm Exam -**

**Spring 2017**

**This exam is due at 11.00p.m. CT on Monday, February 27, 2017. You are to work independently on this exam. You may not consult other people. However, you may use course materials. (Please take SMU Honor code seriously)**

1. (10 pts) A SRS is chosen from a population of 1000 households. Results are shown below.

|  |  |
| --- | --- |
| income |  |
|  |  |
| Mean | 60199.95 |
| Standard Error | 6671.014 |
| Median | 47526.5 |
| Standard Deviation | 42191.2 |
| Minimum | 0 |
| Maximum | 215448 |
| Sum | 2407998 |
| Size | 40 |

a) (5 pts) Calculate **99% confidence interval** for the **population total income**.

Z value: **2.576**

60199.95 (2.576 \* 6671.014) = (43015.42, 77384.48)

(1000\*60199.95) 2.576\*(1000\*6671.014) = (43015417.94, 77,384,482.06)

**99% CI = (43015417.94, 77,384,482.06)**

b) (5 pts) Suppose you wanted to redo the survey above to achieve a MOE (margin of error) of $5000 with 90% confidence level. How large a sample size would be needed, if the population from which this sample has 1000 members?

Z value: **1.645**

MOE: **5000**

= (1.645\*42191.2)^2 / (5000)^2 = 193 (192.68)

**= 192.68 / (1 + 192.68 / 1000) = 162 (161.55)**

2. **True/False:** (20 pts; 2.5 points for each)

For each statement, select True or False. Make sure it is clear which one you are choosing.

1. The selection probability for each unit from simple random sample of size n is . (where N is the population size and n is the sample size.)

**True** False

1. Stratified sampling is a probability sampling technique.

**True** False

1. Measurement error is an example of a non-sampling error

**True** False

1. Based on Proportional allocated stratified design, your estimator will not have a higher variance than it would have from a SRS of the same size.

**True**  False

1. Based on Neyman allocated stratified design, your estimator will have a higher variance than it would have from a SRS of the same size.

True **False**

* Typically no, if we allocate SDs correctly and use correct proxy variables!!

1. A statistic is a numerical index that describes some feature of a sample.

**True** False

1. Sampling fraction for each strata is the same when we use stratified design with proportional allocation. (= sampling fraction in the stratum)

**True** False

1. A *nonprobability sample* is one in which the probability of selection for every member of the sample can be calculated.

True **False**

3. (20 pts) A population has 4 members: a,b,c,d. You would like to **estimate the total number of cars owned by members of this population from a SRSWOR of size 2**. The first member of the population has 2 cars, the second has 1 car, and the other two members of this population have 0 cars.

Find the sampling distribution of the usual estimate of total statistic

(i) (8 points) Present your work in two steps. First, complete this table showing all samples.

|  |  |  |
| --- | --- | --- |
| Sample id | sample members | statistic value |
| 1 | A,B | 1.5 |
| 2 | A,C | 1 |
| 3 | A,D | 1 |
| 4 | B,C | 0.5 |
| 5 | B,D | 0.5 |
| 6 | C,D | 0 |

(ii) (8 points) Then summarize the entire sampling distribution. You can do it either with a histogram (which you will paste below) OR display it in this (completed) table:

|  |  |  |
| --- | --- | --- |
| statistic value | # of samples with this statistic value | proportion of samples with this statistic value |
| 1.5 | 1 | .16 |
| 1 | 2 | .33 |
| 0.5 | 2 | .33 |
| 0 | 1 | .16 |

(iii) (4 points) Suppose you were to repeat this exercise, but to show the sampling distribution of a simple random sample without replacement of size 3. What would change about the sampling distribution of this estimate of total? \_\_ (enter a, b, c, or d)

(a) the mean of the sampling distribution

**(b) the variance of the sampling distribution**

(c) both mean and variance of the sampling distribution

(d) neither mean or variance of the sampling distribution

4. (10 pts) A population of 10 balls is made up of 4 white balls and 6 red balls as shown below:

R W W R W R W R R R

A **systematic sample of size 2** is selected from the population and **the proportion of red balls** in the population is estimated from the sample using as the estimator

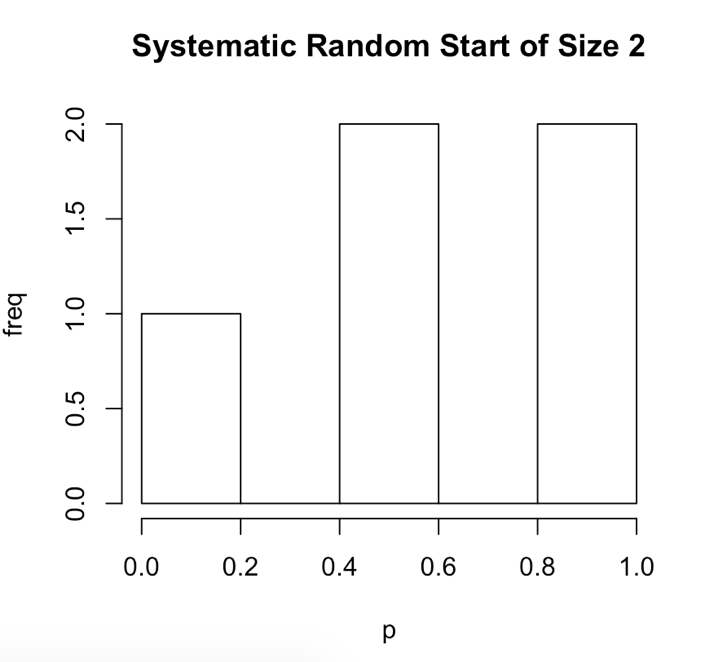
.

Find the sampling distribution of .

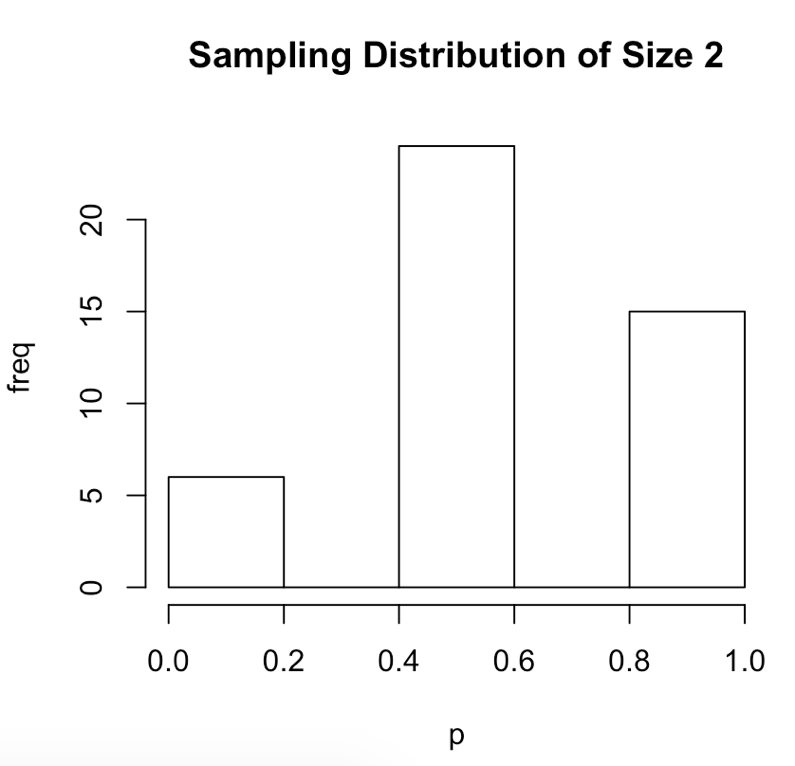
Since we are using a systematic sample, the number of outcomes is reduced as we select observations based on a fixed interval. The interval is 10/2 = 5, the population size of ten divided by our desired sample size of two.

Thus, we need to find the outcomes for all random starts between 1 and 5 with an interval of 5 as our sampling distribution. This reduces the number of outcomes versus a SRSWOR, as observations directly next to each other will not be chosen. Without ordering (implicitly stratifying), the sampling distribution of a systematic random sample looks as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| [RR | WW | WR | RR | WR] |



An SRS distribution would look as follows:



5. (10 pts) Estimate of the proportion of students who are graduates.

|  |  |  |
| --- | --- | --- |
| **Student** | **Stratum** | **Graduate?** |
| Ali | 1 | Y |
| Bucky | 1 | N |
| Judith | 2 | Y |
| Hal | 2 | N |
| Roy | 1 | Y |
| Gideon | 1 | N |
| John | 1 | Y |
| Yusun | 1 | Y |

Step 1. Select 1 student from each stratum.

Step 2. Enumerate all possible samples.

Step 3. Calculate for every sample.

Step 4. Specify the sampling distribution of sample proportions based on this stratified design.

Step 5. Calculate the mean of .

**Lab exam (30 pts)**

Consider data from all the public libraries in Texas. The column labeled *circ* gives the circulation (# of books checked out) for each library in a particular year. You are to select a stratified sample from this population using *circ* as a stratifying variable. You will be asked to construct a 95% confidence interval for the mean of a different characteristic (*inq* = # of inquiries answered) for the population of libraries from your chosen sample.

We sorted the libraries into small, medium, and large libraries and will regard those as strata. We show a column labeled S, M, L to denote which stratum each library belongs to.

Note that this data has been sorted by stratum. This sheet is provided in a csv file named ***libraryinfo.csv***.

Here are the steps you will follow:

(a) (6 points) You will select a stratified random sample of size 40, using proportional allocation. What is the allocation of sample for proportional allocation; i.e., how large a sample will you select from each stratum? *nS* = ; *nM* = ; *nL* =

Show your computations for determining this allocation.

(b) (6 pts) Using SAS PROC SURVEYSELECT, select a stratified random sample from the population, using allocation specified in (a). Paste your SAS code here, along with resulting output.

(c) (6 pts) **Print out your sample and copy and paste it below**. 3 pts for printing, 3 pts for including code to print.

(d) 12 pts Use SAS PROC SURVEYMEANS with the sample above to compute these statistics for the number of inquiries: the estimate of population mean, the standard error of the estimate, the 95% confidence interval for the mean. Paste your SAS code and output below. (This can be included above 5 pts for code, 5 pts for correct CI, mean, and sterr.) (Do not include the graphics in your paste.)