**Lab 3. Sampling distributions and stratified sampling**

**MSDS 6370**

**Objective:**

* For the student to learn about sampling distributions for simple random samples without replacement (SRSWOR).
* For the student to gain insight into stratified samples.

**Introduction**

The topic of the reading material for Asynchronous Lecture 3 was a discussion of sampling distributions. In this lab, we continue to study sampling distributions. In addition we consider a particular type of stratified sampling design.

**Estimating population mean and variance from a sample**

Two characteristics of a population *yi* of size *N* are its mean and its variance. The mean is the average of the population values and the variance indicates the spread of those values. Often we see the square root of the variance , called the standard deviation. The formulas we use today to calculate these are

and .

If we take a sample of size *n* from our population, we use the sample mean to estimate the population mean and also use the sample variance to estimate the population variance as follows:

and .

**Population data**

In this assignment, you will use a small population of 7 taxpayers. The data is in Table 1 (which will be available for download in class) has their actual income and their reported income. An audit is going to take a sample of size 2 and determine the actual income for the 2 taxpayers. The objective of the audit was to estimate the mean of difference between the actual and reported incomes of these 7 taxpayers.

**However, this assignment uses only the actual income**. We are going to examine the sampling distribution of the estimated mean of the actual income from a simple random sample without replacement of size 2.

**Table 1:** Actual and reported income for 7 taxpayers

|  |  |  |
| --- | --- | --- |
| Taxpayer number | Actual Income (thousands of dollars) | Reported Income (thousands of dollars) |
| 1 | 60 | 50 |
| 2 | 72 | 56 |
| 3 | 68 | 66 |
| 4 | 94 | 76 |
| 5 | 102 | 100 |
| 6 | 116 | 112 |
| 7 | 130 | 110 |

Next, we will add one taxpayer, No. 8, so now the population increases to 8 Taxpayers.

|  |  |  |
| --- | --- | --- |
| Taxpayer number | Actual Income (thousands of dollars) | Reported Income (thousands of dollars) |
| 8 | 200 | 175 |

The auditors decide that they have to include Taxpayer 8 has to be included in their sample because of the large income. In this case, we say that Taxpayer 8 is included with certainty or with a probability of selection equal to 1. We essentially have a stratified sample of size 3 where 2 units are selected from the original 7 and the third is selected with certainty. We will examine the effect on the estimator of accounting for the stratification in estimating the population mean.

**Exercise 1**

You may use the software of your choice in this exercise. The following describes the steps using Excel.

1.Systematically list all 21 possible samples of size n = 2 from the taxpayer population of size 7 in Table 1. For each sample, calculate estimates of the mean and variance and enter these into Table 2 on the Results page. Also calculate the means of sample estimates of mean and variance and enter these in Table 2 on the Results page.

2. Calculate the population mean and variance actual income of all 7 taxpayers used in creating Table 2 and compare them the mean of sample estimates of mean and variance

**Exercise 2**

1.Add Taxpayer 8 so that now the population size N = 8. Since the auditors want Taxpayer 8 to be in the sample with certainty, we essentially have 2 strata, Stratum 1 of size N1 =7 with the original 7 Taxpayers and Stratum 2 of size N2 =1 with a single taxpayer, Taxpayer 8. We can select a stratified sample of 3 where 2 are selected from Stratum 1 with the original 7 Taxpayers and 1 from Stratum 2 with only Taxpayer 8 whose actual income is 200. The mean of a sample of size 2 from Stratum 1 is and the mean of a sample of size 1 from Stratum 2 is , which equals 200 since Stratum 2 has only 1 Taxpayer. Therefore, the estimate of the population mean incorporating the stratification is

.

Use the estimates of for all possible samples of size 2 from the original 7 Taxpayers that you entered into Table 2 on the Results page to calculate all possible estimates . Enter these in the Table 3 on the Results page.

2. Calculate the population mean of the actual income for the population of all 8 taxpayers and compare it the mean of the 21 sample estimates of mean .

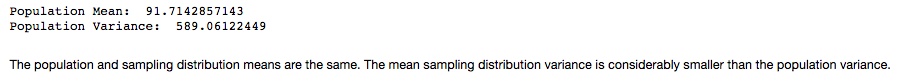
**Lab 3. Results and Exercises**

**Exercise 1**

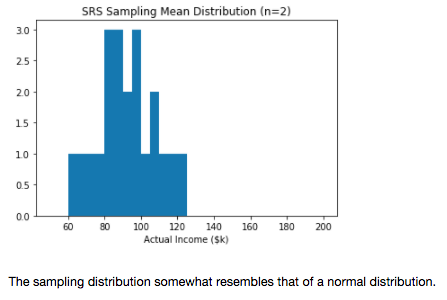
**Table 2. 8 Taxpayers**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample number** | **2 of 8 taxpayers**  **in sample** | **mean of actual income of 2 taxpayers** | **Estimate of variance** |
| 1 | 1,2 | 66 | 661.2245 |
| 2 | 1,3 | 64 | 768.0816 |
| 3 | 1,4 | 77 | 216.5102 |
| 4 | 1,5 | 81 | 114.7959 |
| 5 | 1,6 | 88 | 13.7959 |
| 6 | 1,7 | 95 | 10.7959 |
| 7 | 2,3 | 70 | 471.5102 |
| 8 | 2,4 | 83 | 75.9388 |
| 9 | 2,5 | 87 | 22.2245 |
| 10 | 2,6 | 94 | 5.2245 |
| 11 | 2,7 | 101 | 86.2245 |
| 12 | 3,4 | 81 | 114.7959 |
| 13 | 3,5 | 85 | 45.0816 |
| 14 | 3,6 | 92 | 0.0816 |
| 15 | 3,7 | 99 | 53.0816 |
| 16 | 4,5 | 98 | 39.5102 |
| 17 | 4,6 | 105 | 176.5102 |
| 18 | 4,7 | 112 | 411.5102 |
| 19 | 5,6 | 109 | 298.7959 |
| 20 | 5,7 | 116 | 589.7959 |
| 21 | 6,7 | 123 | 978.7959 |
|  |  |  |  |
| Sum |  | 1,926 | 5,154.2857 |
| Mean |  | 91.714 | 245.4422 |

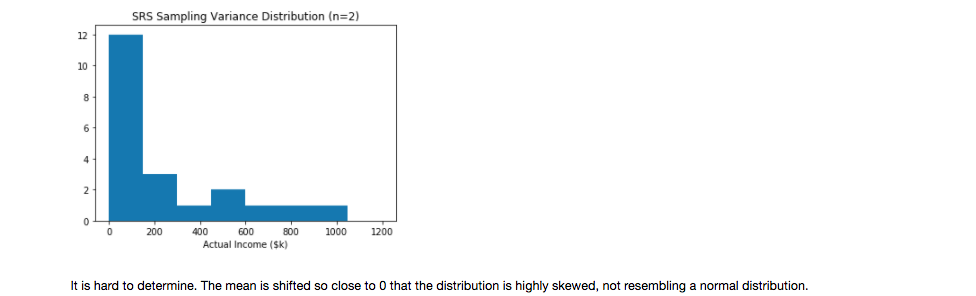
2. Calculate population mean and variance actual income of all 7 taxpayers used in creating Table 2. How does the mean of the sample estimates of mean and variance compare to the population mean and variance?



3. Construct a histogram of the 21 sample estimates of the mean , its *sampling distribution*. How does the shape of the histogram compare to the shape of a normal distribution?



4. Construct a histogram of the 21 sample estimates of the variance , its *sampling distribution*. How does the shape of the histogram compare to the shape of a normal distribution?

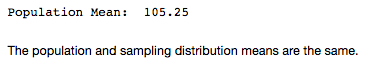


**Exercise 2**

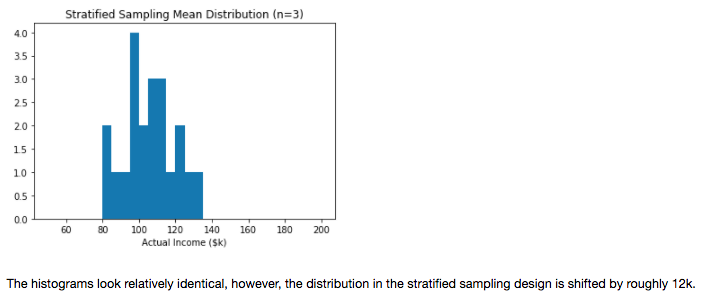
**Table 3. 8 Taxpayers**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample number** | **2 of 7 taxpayers**  **in Stratum 1** | **mean of actual income of 2 taxpayers from Stratum 1,** | **Estimate of population mean of actual income for 8 taxpayers** |
| 1 | 1,2 | 66 | 82.75 |
| 2 | 1,3 | 64 | 81 |
| 3 | 1,4 | 77 | 92.375 |
| 4 | 1,5 | 81 | 95.875 |
| 5 | 1,6 | 88 | 102 |
| 6 | 1,7 | 95 | 108.125 |
| 7 | 2,3 | 70 | 86.25 |
| 8 | 2,4 | 83 | 97.625 |
| 9 | 2,5 | 87 | 101.125 |
| 10 | 2,6 | 94 | 107.25 |
| 11 | 2,7 | 101 | 113.375 |
| 12 | 3,4 | 81 | 95.875 |
| 13 | 3,5 | 85 | 99.375 |
| 14 | 3,6 | 92 | 105.5 |
| 15 | 3,7 | 99 | 111.625 |
| 16 | 4,5 | 98 | 110.75 |
| 17 | 4,6 | 105 | 116.875 |
| 18 | 4,7 | 112 | 123 |
| 19 | 5,6 | 109 | 120.375 |
| 20 | 5,7 | 116 | 126.5 |
| 21 | 6,7 | 123 | 132.625 |
|  |  |  |  |
| Sum |  | 1,296 | 2,210.5 |
| Mean |  | 91.714 | 105.25 |

2. Calculate population mean of the actual income of all 8 taxpayers used in creating Table 3. How does the mean of the sample estimates of compare to the population mean?



3. Construct a histogram of the 21 sample estimates of How does this histogram compare to the histogram in Exercise 1 for ?



**Appendix:**

LAB03\_JackNelson.ipynb

