**Lab 1: Sampling distribution and simple random sampling**

**MSDS 6370**

**Objectives:**

* Introduce the student to the idea of a probability sample and the simple random sample.
* Give the student a preview of the idea of the sampling distribution via a demonstration.

**Introduction:**

We will use the rectangle population that you encountered in the Rectangle Exercise in your asynchronous session. The term *judgment sample* means that no random process was used to choose the sample; rather you used your own judgment to select a representative sample. Another name for such a sample is *quota sample*.

#### Mean of population versus mean of judgment samples

Calculate the mean of the 15 estimates of total area from the judgment samples selected by the class members and write it in Table 1 on the Results page of this lab along with the largest and smallest estimate. Compare the mean of 15 estimates of total area calculated from your judgment samples with the true total area for the population, which is 739. Then compare the 15 estimates calculated from your judgment samples with the true value for the population, How many of the estimates for the total area are too high and how many are lower?

#### Mean of population versus mean of simple random samples (SRSs)

Calculate the mean of the 15 estimates of total area from the SRSs selected by the class members and write it in Table 1 on the Results page of this lab along with the largest and smallest estimate. Compare the mean of 15 estimates of total area calculated from your SRSs with the true total area for the population, which is 739. Then compare the 15 estimates calculated from your SRS with the true value for the population. How many of the simple random sample means are higher and how many are lower than the true population area?

A *probability sample* is one in which the probability of selection of each frame element is known and non-zero. A probability sample removes judgment from the process of selecting a sample by using some random method for deciding which units from the population are chosen into the sample. A probability sample is a methodology that produces unbiased estimates of the mean and the total of the population. The simplest probability sample is a *simple random sample*.

**Sampling Distribution:**

When a probability sample design is used, the estimates of total area calculated from the samples have a predictable behavior that can be determined from the characteristics of the population. We are examining that behavior by performing an experiment with our 14 SRSs from the rectangle population. We calculatedthe total area from each, and now can see what the distribution of these estimates of total area looks like. The distribution of a statistic, such as an estimated mean or total, from a probability sample is called its *sampling distribution*. In our case, we are learning about the sampling distribution of the estimates of total area from an experiment and then comparing it to what theory says it should look like.

**Exercises:**

Below you have 2 sets of 15 estimates of the rectangle population’s total area. The 1st column contains the estimates from the 15 judgement samples and the 2nd column contains the estimates from 15 simple random samples. Previous class members submitted these estimates.

1. Provide your total estimate from SRS and judgment sample and explain how you selected the judgment sample and the SRS in breakout group.

Judgment: …………920…. SRS: ……………780……

2. Calculate the mean, maximum, and minimum of each set of 15 estimates of total area of the rectangles. Record all in Table 1.

|  |  |  |
| --- | --- | --- |
|  | judgment samples | simple random samples |
| 1 | 920 | 750 |
| 2 | 940 | 780 |
| 3 | 850 | 590 |
| 4 | 790 | 780 |
| 5 | 890 | 660 |
| 6 | 780 | 870 |
| 7 | 900 | 830 |
| 8 | 890 | 980 |
| 9 | 1270 | 720 |
| 10 | 810 | 650 |
| 11 | 1040 | 620 |
| 12 | 790 | 700 |
| 13 | 890 | 770 |
| 14 | 850 | 850 |
| 15 | 1320 | 780 |

Table 1. Mean, maximum, and minimum of estimates of total area from 15 judgment samples and from 15 simple random samples (SRS)

|  |  |  |
| --- | --- | --- |
|  | Judgment samples | SRS |
| Mean total area | 928.7 | 755.3 |
| Minimum total area | 780 | 590 |
| Maximum total area | 1320 | 980 |

3. Sketch (or use a software package of your choice and cut-&-paste) a histogram of the 15 judgment samples values of total area.

4. Sketch (or use a software package of your choice and cut-&-paste) a histogram of the 15 SRS values of total area.

5. The mean of estimates of total area from several SRSs is designed to be an estimate of population total area, the sum of the areas of all 100 rectangles. Comment on the performance of the mean of the estimates of total area from the class’s SRSs to that of mean of the estimates of total area from our judgment samples, as far as how they performs for estimating the true population total area.

The mean of the SRS samples performed better than the mean of the judgement samples. Human judgement, at least in my case, tried to choose a random sample of sizes rather than just a random sample of rectangles. This led me to underrepresent the smaller rectangles and over-estimate the population mean.

6. The 15 values of estimated total area the class obtained are not the only possible values of the total area that could be obtained from a SRS. What is the minimum possible value of total area that could be obtained from a SRS of size 10 from the rectangle population? \_\_10\_\_ . What is the maximum possible value of total area that could be obtained from a SRS of size 10 from the rectangle population? \_\_\_\_170\_\_

7. Is it possible that estimated total area from a SRS could be worse than your judgment sample estimated total area? Explain.

It is possible that estimated total area from an SRS could be worse than my judgement. The larger the n umber of SRS samples used to estimate the population mean, the better will be the estimate.

8. Is it likely that estimated total area from a SRS could be worse than your judgment sample estimated total area? Explain.

It is not likely that a SRS estimate would be farther from the true mean than my judgement estimate.