**MSDS 6379**

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

**Objectives:**

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

**Exercise 1**: Using student responses from the Rectangle Exercise in your asynchronous session, calculate the mean, minimum, and maximum of the sample estimates from all the class students for the judgement samples and SRSs. Compare the mean of all the sample estimates to the population mean which is 738. How many of the student estimates are too high and how many are too low? Enter your responses in the table below:

|  |  |  |
| --- | --- | --- |
| **Item** | **Judgement Samples** | **SRS** |
| Mean total area | 857.5 | 770 |
| Minimum total area | 650 | 550 |
| Maximum total area | 1340 | 1030 |
| Number < population mean | 3 | 9 |
| Number >= population mean | 3 | 9 |

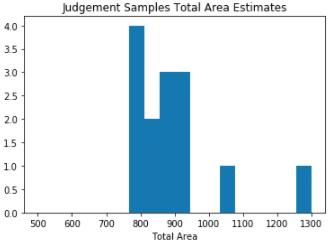
For the exercises below, use the table below to craft you responses. This table contains the responses of students to the Rectangle Exercise in a previous section of the Sampling class.

|  |  |  |
| --- | --- | --- |
| **Sample #** | **Judgement Sample** | **Simple Random Sample** |
| 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 |

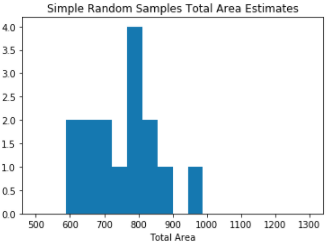
**Exercise 2**: Calculate the overall mean, maximum, and minimum using each of the total area estimates for the Judgment Samples and SRSs. Enter you responses in the table below.

|  |  |  |
| --- | --- | --- |
| **Item** | **Judgement Samples** | **SRS** |
| Mean total area | 928.67 | 755.33 |
| Minimum total area | 780 | 590 |
| Maximum total area | 1320 | 980 |

**Exercise 3:** Sketch or use software and create a histogram of the 15 judgement sample values of total area:

****

**Exercise 4:** Sketch or use software and create a histogram of the 15 SRS values of total area:



**Exercise 5:** The mean of estimates of the total area of SRSs is designed to be an estimate of the population total area, i.e., the sum of the areas of all 100 rectangles. Comment on the performance of the mean of the estimates of total area from SRS from the table above and compare it to the mean of the estimates from the judgement samples. Which is a better estimate of the true population total area and explain why one estimate should be better than another?

The better estimate of the true population total area is SRS. The overall mean is closer to the true population mean and the distribution more closely revolves around the true population mean, limiting its variance. The judgement sample yielded results much higher than the true population. The reasoning behind this being is the judgement sample was taken without equal probability of a lot being selected. The selection of lots using judgement unknowingly injects bias towards lots of a larger size since they cover more surface area. An individual would notice and select larger plots as opposed to smaller ones. The SRS method, however, gives each plot an equal chance of selection regardless of size. In turn, it produces an unbiased estimator of the true population mean.

**Exercise 6**: What is the minimum possible value of estimated total area that could be obtained from a SRS of size 10 from the rectangular population? What is the minimum possible value of estimated total area that could be obtained from a SRS of size 10?

Minimum possible area from sample of size 10: 100

Maximum possible area from sample of size 10: 1700

**Exercise 7**: Is it possible that the estimated total area from an SRS could be worse than your judgement sample estimated total area? Explain.

Yes, it is entirely possible that the SRS sample could yield an estimate further away from the judgement sample. Any estimate from the range of possible estimates in the answer noted above are entirely possible for either a judgement sample or a SRS sample.

**Exercise 8**: Is it likely that the estimated total area from an SRS could be worse than your judgement sample estimated total area? Explain.

No, even though the entire range of values is possible does not indicate that picking on either extreme is likely. SRS gives each lot an equal chance of selection so it is more likely to select same amount above the true population mean as below. In turn, producing a relatively close mean population estimate on average. The judgement sample, however, gives a highly probability of selecting more lots above to the true population mean resulting in potentially over representing larger plots. Therefore, it is more likely that the SRS method performs better than the judgement sample on average.

**Appendix:**

LAB01JackNelson.ipynb

