

MTH211 Assignment 1 Part 3

Hall 4 Canteen Simulation

Arqam Patel, Kumar Kanishk Singh, Aditya Kumar Tiwari

Table of contents

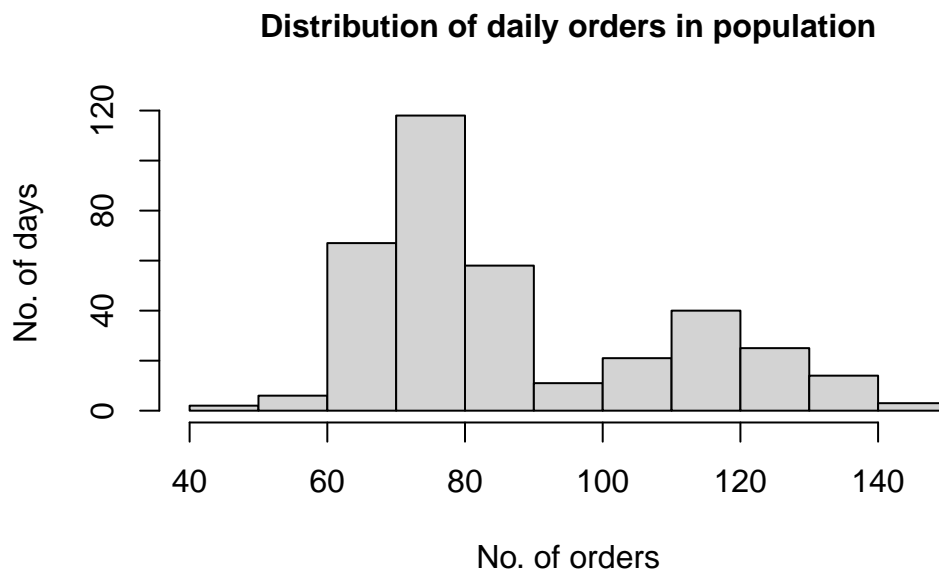
A. Generating population data	1
B. Taking sample means using SRSWR and SRSWOR	2
C. Comparing performance as sample size varies	4
D. Stratified random sampling method 1 (StRS1)	5
E. Stratified random sampling method 2 (StRS2)	6
F. Comparing performance of different sampling methods	7

A. Generating population data

First, we set the seed for reproducibility. Since there were 105 weekend days in 2022, we then simulate 105 draws from a Poisson distribution with mean 120. Then we simulate 260 draws, one for each weekday in 2022, from a Poisson distribution with mean 75. Combining the two, we get the dataset of the whole year i.e. our population data.

```
set.seed = 1
weekends <- rpois(n = 105, 120)
weekdays <- rpois(n = 260, 75)
year <- c(weekdays, weekends)
```

We can now look at the distribution of daily orders throughout the year.



We now calculate the population mean μ .

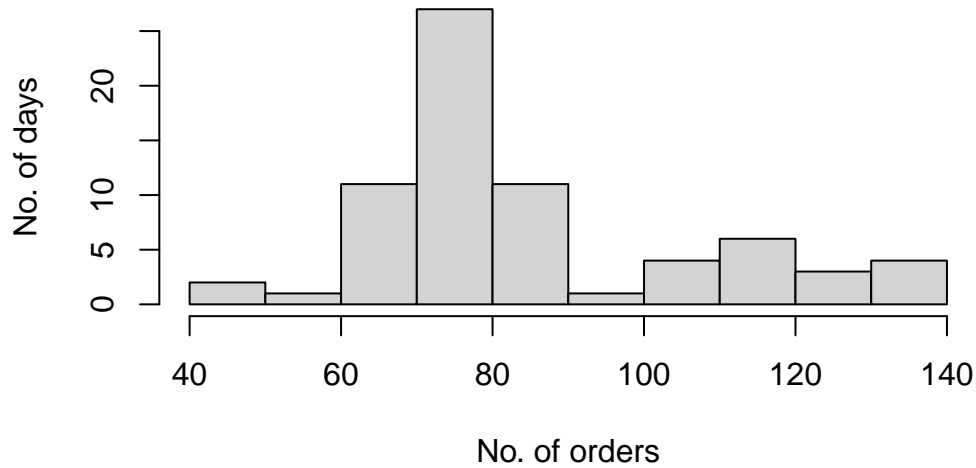
```
[1] "Population mean: 87.67"
```

B. Taking sample means using SRSWR and SRSWOR

We use Simple Random Sampling With Replacement (setting `replace = TRUE`) to get 70 samples from the population data.

```
srswr_sample <- sample(year, size = 70 , replace = TRUE)
```

Distribution of daily orders in SRSWR sample



We then find the SRSWR sample mean (\bar{Y}_{SRSWR}) and the percentage error relative to the population mean.

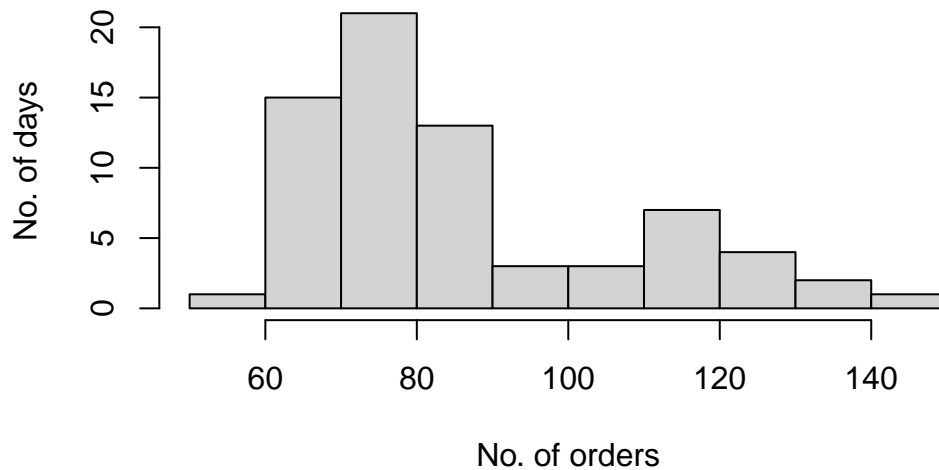
```
[1] "Sample mean using SRSR: 85.5"
```

```
[1] "SRSWR percentage error 2.47 %"
```

We use Simple Random Sampling Without Replacement (setting `replace = FALSE`) to get 70 samples from the population data.

```
srswor_sample <- sample(year, size = 70, replace = FALSE)
```

Distribution of daily orders in SRSWOR sample



We then find the SRSWOR sample mean (\bar{Y}_{SRSWOR}) and the percentage error relative to the population mean.

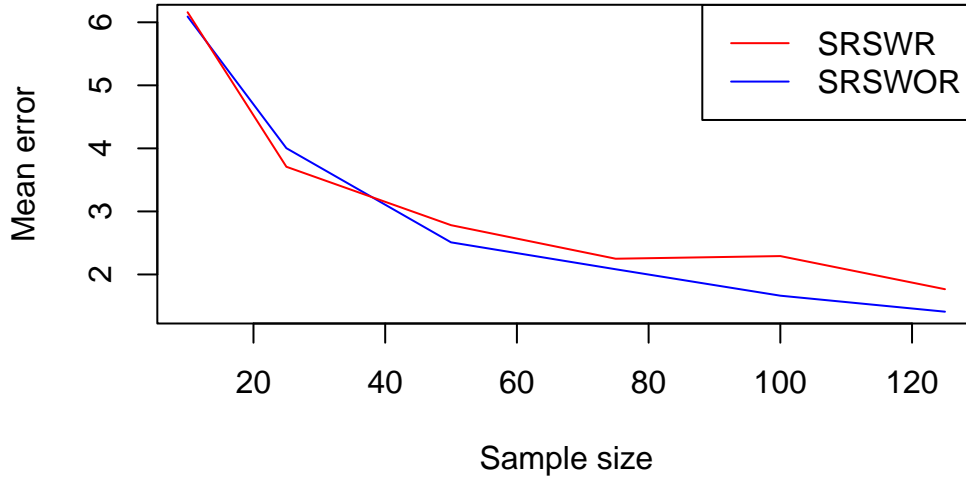
```
[1] "Sample mean using SRSWOR: 86.6"
```

```
[1] "SRSWOR percentage error 1.22 %"
```

We can observe that performance of SRSWOR is, in this case, better than that of SRSWR.

C. Comparing performance as sample size varies

To compare performances of sampling with different sample sizes, we take a large number of samples with each sample size and compute the average error corresponding to each sample size.



We can observe that in general, performance of both the sampling algorithms improves as we increase the sample size.

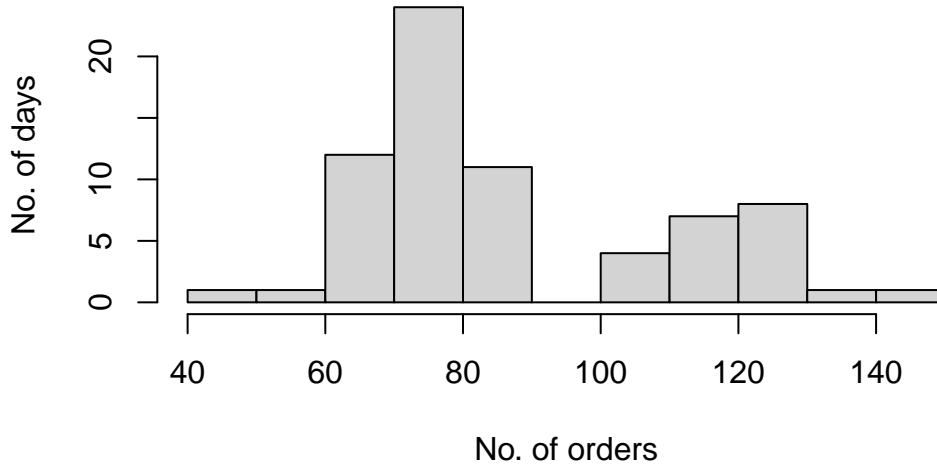
D. Stratified random sampling method 1 (StRS1)

$$n_i \propto N_i$$

In this technique, we consider weekends and weekdays as distinct strata in our population. We thus sample from each stratum independently, with the stratum sample size n_i proportional to the size of the stratum N_i we're sampling from. We then combine these stratum samples to get the final sample.

In this example we extract a sample of 70 applying this technique.

Distribution of daily orders in StRS1 sample



We then get the sample mean as well as the percentage error with respect to the population mean.

```
[1] "StRS1 sample mean 87.42"
```

```
[1] "StRS1 percentage error 0.27 %"
```

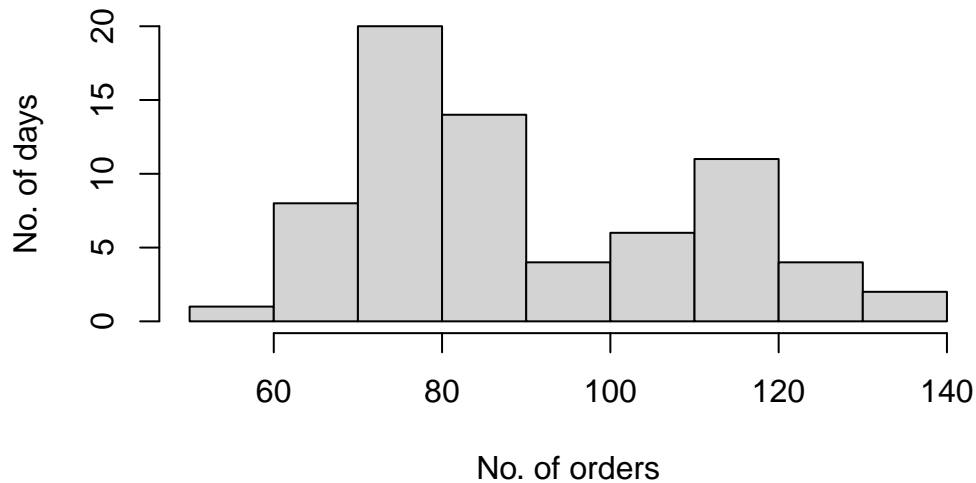
E. Stratified random sampling method 2 (StRS2)

$$n_i \propto N_i \cdot \sigma_i$$

In this technique, we consider weekends and weekdays as distinct strata in our population. We thus sample from each stratum independently, with the sample size proportional to the size of the stratum we're sampling from. We then combine these stratum samples to get the final sample.

In this example we extract a sample of 70 applying this technique.

Distribution of daily orders in StRS2 sample



We then get the sample mean as well as the percentage error with respect to the population mean.

```
[1] "StRS2 sample mean 90.21"
```

```
[1] "StRS2 percentage error 2.89 %"
```

F. Comparing performance of different sampling methods

For comparing SRSWR, SRSWOR, StRS1 and StRS2 let us take multiple samples and find the average error.

We take the same number of samples n with the same sample size for each method.

```
n <- 1e3  
sample_size <- 100
```

```
[1] "Average error of SRSWR 1.99 %"
```

```
[1] "Average error of SRSWOR 1.73 %"
```

```
[1] "Average error of StRS1 0.76 %"
```

```
[1] "Average error of StRS2 2.12 %"
```

We can observe that in this example, on average, StRS1 (with $n_i \propto N_i$) is the best sampling algorithm, followed by SRSWOR, then SRSWR. The worst performing algorithm is StRS2 (with $n_i \propto N_i \cdot \sigma_i$), possibly because the standard deviations of the strata are comparable.

[1] "SD of weekdays 8.66"

[1] "SD of weekends 10.72"