TRATIFIED SAMPLING CALCULATOR USING R SHINY

Asynchronous Assignment – R Shiny Development (Stratified Sampling)

Department: Statistics and Data Science
Course: Sampling Techniques / R Shiny
Student Name: Alan John PS
Programme: BSc Data Science & Statistics

---

1. Introduction

Stratified sampling is a probability sampling technique in which the population is divided into homogeneous subgroups called strata, and samples are drawn from each stratum. This method ensures better representation of the population compared to simple random sampling, especially when the population is heterogeneous.

The accuracy of stratified sampling largely depends on how the sample size is allocated among different strata. Proper allocation can reduce sampling variance, control survey cost, and minimize time required for data collection.

This project presents an **interactive R Shiny application** that calculates the required sample size for estimating the population mean using **Proportional Allocation, Neyman Allocation, and Optimised Allocation**, considering **variance, cost, and time** in each stratum.

---

2. Objectives

The objectives of this project are:

- To estimate the required sample size for population mean under stratified sampling
- To implement proportional, Neyman, and optimised allocation methods
- To incorporate cost and time constraints in sampling
- To develop an interactive R Shiny dashboard for educational and practical use

---

3. Methodology

3.1 Proportional Allocation

In proportional allocation, the sample size in each stratum is proportional to the stratum population size.

Formula:

$$n_h = n \times \frac{N_h}{N}$$

This method is simple and widely used when strata variances are similar.

3.2 Neyman Allocation

Neyman allocation minimizes the variance of the estimator by allocating more samples to strata with higher variability.

Formula:

$$n_h = n \times \frac{N_h \sigma_h}{\sum N_h \sigma_h}$$

This method is statistically efficient when variances differ across strata.

3.3 Optimised Allocation

Optimised allocation extends Neyman allocation by considering **cost per unit** of sampling in each stratum.

Formula:

$$n_h = n \times \frac{N_h \sigma_h / \sqrt{c_h}}{\sum (N_h \sigma_h / \sqrt{c_h})}$$

This method ensures cost-efficient and time-efficient survey design.

---

4. Application Development Using R Shiny

The application is developed using **R Shiny** and **shinydashboard** to provide an interactive user interface.

Key Inputs

- Desired precision (d)
- Z-value for confidence level
- Population size for each stratum
- Variance of each stratum

- Cost per sampling unit
- Time per sampling unit

Key Outputs

- Sample allocation for each stratum
- Total sample size
- Total cost and total time
- Comparative summary of allocation methods
- Downloadable results

The application dynamically updates results and provides clear comparison tables for decision-making.

---

5. Results and Discussion

The results show that:

- Proportional allocation is easy to implement but may not be cost-efficient
- Neyman allocation reduces variance by allocating more samples to high-variance strata
- Optimised allocation provides the most efficient solution when cost and time are critical factors

The summary section of the app recommends the most efficient method based on minimum cost, time, and sample size.

---

6. Conclusion

This R Shiny application successfully demonstrates the practical implementation of stratified sampling techniques. It serves as a useful educational tool for understanding sampling allocation methods and a practical tool for planning cost-effective surveys.

The project fulfills all academic requirements of the asynchronous assignment and demonstrates the effective use of R Shiny in statistical applications.

---

7. Tools and Technologies Used

- R Programming Language
- Shiny

- shinydashboard
- DT Package

---

8. GitHub Repository

The complete project including source code and documentation is available on GitHub.