CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE · INDIA

Systematic Sampling – Estimation of Population Mean

Subject:

Sampling Techniques

Submitted By:

Ann Maria Anil

2341419

6BScDS

Submitted To:

Prof.Dibu

Date of Submission:

16th December, 2025

1. Introduction

Systematic sampling is a probability sampling technique used to estimate population parameters efficiently when the population units can be arranged in an ordered manner. Instead of dividing the population into subgroups, a fixed sampling interval is used to select units at regular positions. This assignment focuses on estimating the population mean using systematic sampling and visualizing the design of the experiment through an R Shiny application.

2. Objective

- To determine the required sample size for estimating the population mean
- To design a systematic sampling plan using allowable sampling bias and confidence level
- To visualize the design of the experiment using R Shiny
- To document the implementation for GitHub and LinkedIn submission

3. Methodology

Let:

- N be the population size
- Z be the confidence level (normal deviate)
- B be the allowable sampling bias (error)
- S be the population standard deviation (assumed)

The required sample size for estimating the population mean is given by:

$$n = (Z^2 S^2) / B^2$$

After computing the sample size, the sampling interval is calculated as:

$$k = \lfloor N / n \rfloor$$

A random start r is selected from the first interval (1 to k). The final sample consists of every k-th unit starting from r.

4. Design of Experiment

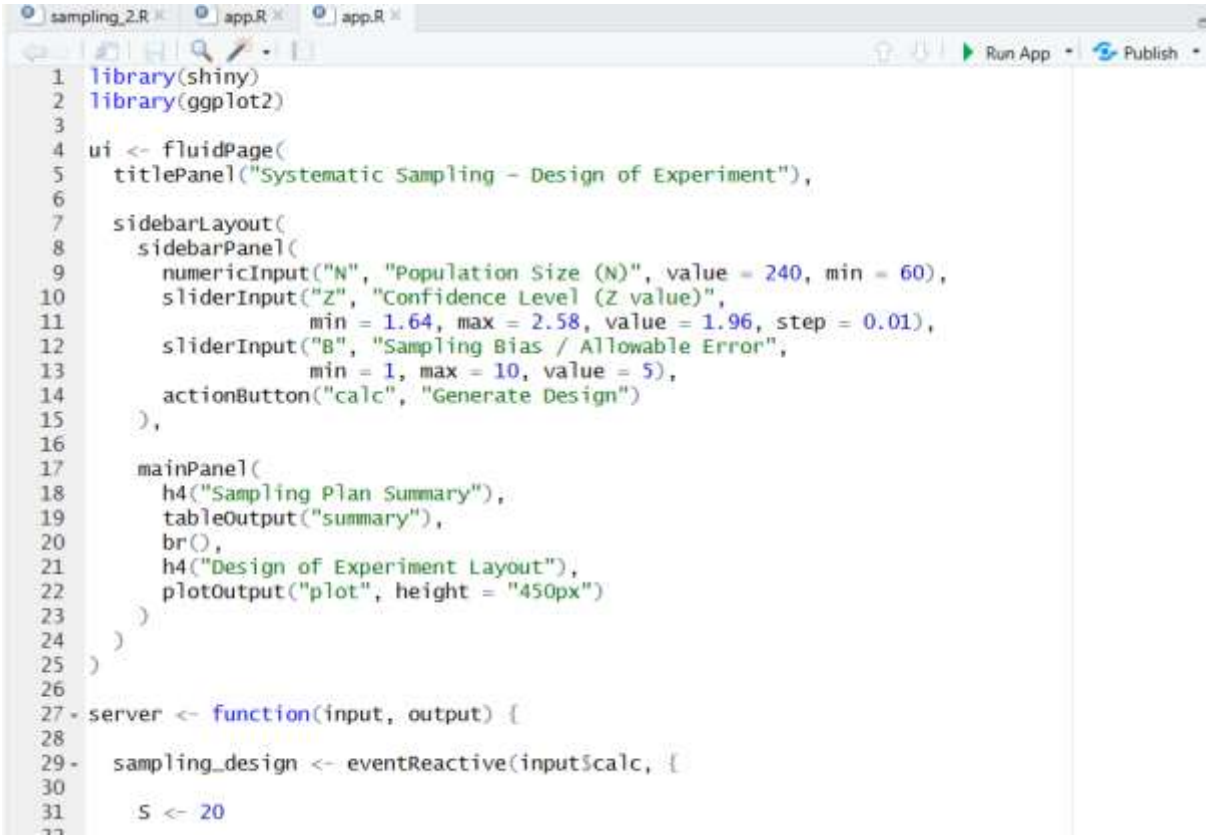
The systematic sampling design is represented by dividing the population into equal-sized blocks corresponding to the sampling interval. From each block, exactly one unit is selected. This block-wise visualization clearly demonstrates the systematic nature of selection and avoids overlap between sampled units.

An R Shiny application was developed to:

- Accept user inputs for population size, confidence level, and sampling bias
- Compute sample size, sampling interval, and random start
- Display a summary table of the sampling plan
- Plot the experimental design using block-based visualization

5. Implementation Details

- Software: RStudio
- Framework: R Shiny
- Visualization: ggplot2
- Output: Interactive sampling design and summary table



sampling_2.R | app.R | app.R

Run App | Publish

```
    n <- ceiling((input$Z^2 * S^2) / (input$B^2))
    k <- floor(input$N / n)
    r <- sample(1:k, 1)

    selected <- seq(r, input$N, by = k)

    blocks <- ceiling(input$N / k)

    df <- data.frame(
      Unit = 1:input$N,
      Block = ceiling((1:input$N) / k)
    )

    df$Selected <- ifelse(df$Unit %in% selected, "Selected", "Not Selected")

    list(df = df, N = input$N, n = n, k = k, r = r)
  })

  output$summary <- renderTable({
    d <- sampling_design()
    data.frame(
      Parameter = c("Population Size (N)",
                    "Sample Size (n)",
                    "Sampling Interval (k)",
                    "Random Start"),
      Value = c(d$N, d$n, d$k, d$r)
    )
  })

  output$plot <- renderPlot({
    d <- sampling_design()
```

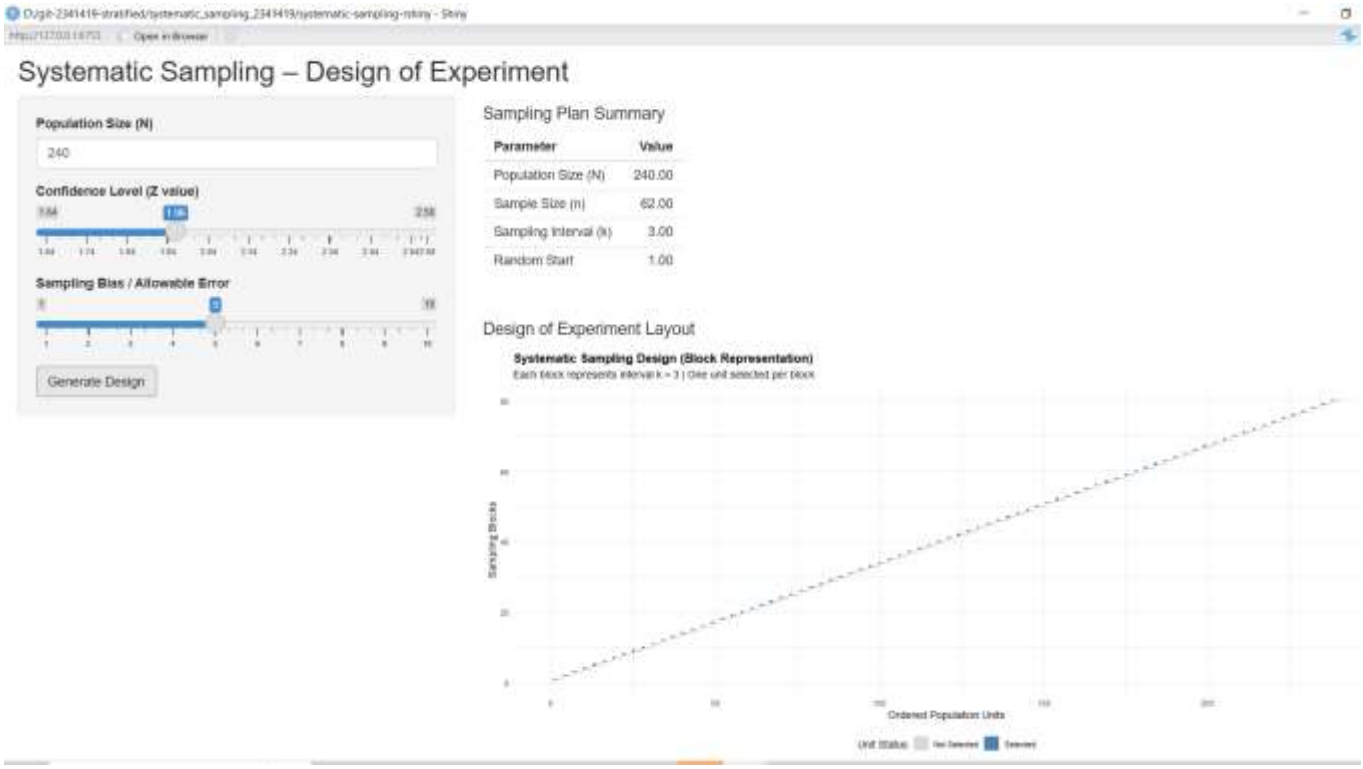
```
  output$plot <- renderPlot({
    d <- sampling_design()
    df <- d$df

    ggplot(df, aes(x = Unit, y = Block, fill = Selected)) +
      geom_tile(color = "white", height = 0.9) +
      scale_fill_manual(
        values = c("Not Selected" = "grey85",
                   "Selected" = "steelblue")
      ) +
      labs(
        title = "Systematic Sampling Design (Block Representation)",
        subtitle = paste("Each block represents interval k =", d$k,
                         "| One unit selected per block"),
        x = "Ordered Population Units",
        y = "Sampling Blocks",
        fill = "Unit Status"
      ) +
      theme_minimal() +
      theme(
        plot.title = element_text(face = "bold"),
        legend.position = "bottom"
      )
  })
}

shinyApp(ui, server)
```

89:1 (Top Level)

Console | Terminal | Background Jobs

6. Results and Interpretation

The visualization shows that sampled units are evenly distributed across the population. Compared to simple random sampling, systematic sampling ensures better coverage of the population. Unlike stratified sampling, no subgroups are formed; instead, regular spacing governs the selection process.



7. Conclusion

Systematic sampling provides a simple and effective method for estimating the population mean when the population is ordered. The developed R Shiny application successfully demonstrates both the computation and the design of the experiment in a clear and intuitive manner. This implementation satisfies the requirements of the Sampling–2 assignment and supports reproducible analysis through GitHub sharing.

8. Repository and Professional Link

The complete R Shiny application code and report have been uploaded to a GitHub repository. The repository link has been shared via LinkedIn as part of the assignment submission as well.

Git hub link: https://github.com/ann-maria-anil/systematic-sampling-r-shiny