Date _____ Expt. No. 10

Expt. Name _____ Page No. 24

EXPERIMENT - 10

Aim:- Make use of One-way Anova and Two-way Anova techniques in R to explore the sources of variability among different groups.

Introduction:- The aim of this experiment is to utilize One-way ANOVA and Two-way ANOVA techniques in R to explore the sources of variability among different groups. The goal is to understand how these analysis of variance methods can be applied to assess the impact of categorical variables on the variability in a dataset.

Software Required:-

1. R Statistical Software
2. R Studio

Relevance of the Experiment:- Analysis of variance is a powerful statistical technique used to analyze the differences among group means in a sample. This experiment is relevant for individuals involved in experimental design, research and data analysis, as it provides hands-on experience in identifying and understanding the sources of ~~variance~~ variability among different groups.

Description:- The experiment involves applying One-way ANOVA and Two-way ANOVA techniques in R to explore the

Teacher's Signature: _____

Aim:- Make use of One-way ANOVA and Two-way ANOVA techniques in R to explore the sources of variability among different groups

```r
# Load required libraries
library(dplyr)
library(ggplot2)

# Load the iris dataset
data("iris")

# Function to perform sampling and estimate mean using a given sampling method
perform_sampling <- function(dataset, sample_size, sampling_method) {
  if (sampling_method == "simple_random") {
    sampled_data <- dataset[sample(nrow(dataset), sample_size, replace = FALSE),
  ] else if (sampling_method == "stratified") {
    sampled_data <- dataset %>%
      group_by(Species) %>%
      sample_n(sample_size, replace = FALSE)
  } else if (sampling_method == "systematic") {
    n <- nrow(dataset)
    k <- sample_size
    start <- sample(1:k, 1)
    indices <- seq(start, n, by = k)
    sampled_data <- dataset[indices, ]
  } else {
    stop("Invalid sampling method.")
  }

  return(mean(sampled_data$Sepal.Length))
}

# Define parameters
sample_size <- 30
num_iterations <- 100
sampling_methods <- c("simple_random", "stratified", "systematic")

# Perform sampling and estimate mean using each technique
sampling_results <- lapply(sampling_methods, function(method) {
  replicate(num_iterations, perform_sampling(iris, sample_size, method))
})

# Combine results into a data frame
results_df <- data.frame(
```

impact of one or two categorical variables on the variability within a dataset. Participants will learn to formulate hypotheses, conduct ANOVA tests, and interpret the results.

Pseudocode / Algorithm :-

1. Identify the categorical variables of interest.
2. Formulate the null and alternative hypothesis for the ANOVA tests.
3. Conduct One-way ANOVA or Two-way ANOVA using the relevant R functions.
4. Perform post-hoc tests if necessary to identify specific group differences.
5. Interpret the results and draw conclusions based on the ANOVA output.

Learning Outcomes :-

1. Understanding the principles of One-way and Two-way ANOVA.
2. Proficiency in formulating null and alternative hypotheses for ANOVA tests.
3. Ability to conduct One-way & Two-way ANOVA in R.

Teacher's Signature:

```r
# Perform sampling and estimate mean using each technique
sampling_results <- lapply(sampling_methods, function(method) {
  replicate(num_iterations, perform_sampling(iris, sample_size, method))
})

# Combine results into a data frame
results_df <- data.frame(
  sampling_method = rep(sampling_methods, each = num_iterations),
  mean = unlist(sampling_results)
)

# Plot histograms
histogram <- ggplot(results_df, aes(x = mean, fill = sampling_method)) +
  geom_histogram(position = "identity", alpha = 0.5, bins = 20) +
  labs(title = "Distribution of Sample Means", x = "Mean Sepal Length", y = "Frequency") +
  theme_minimal()

print(histogram)

# Calculate means
mean_results <- sapply(sampling_results, mean)
names(mean_results) <- sampling_methods

# Print results
print(mean_results)
```

```
> mean_results <- sapply(sampling_results, mean)
> names(mean_results) <- sampling_methods
>
> # Print results
> print(mean_results)
simple_random    stratified    systematic
     5.832267      5.838022      5.848600
>
```

Date

Expt. No.

Expt. Name

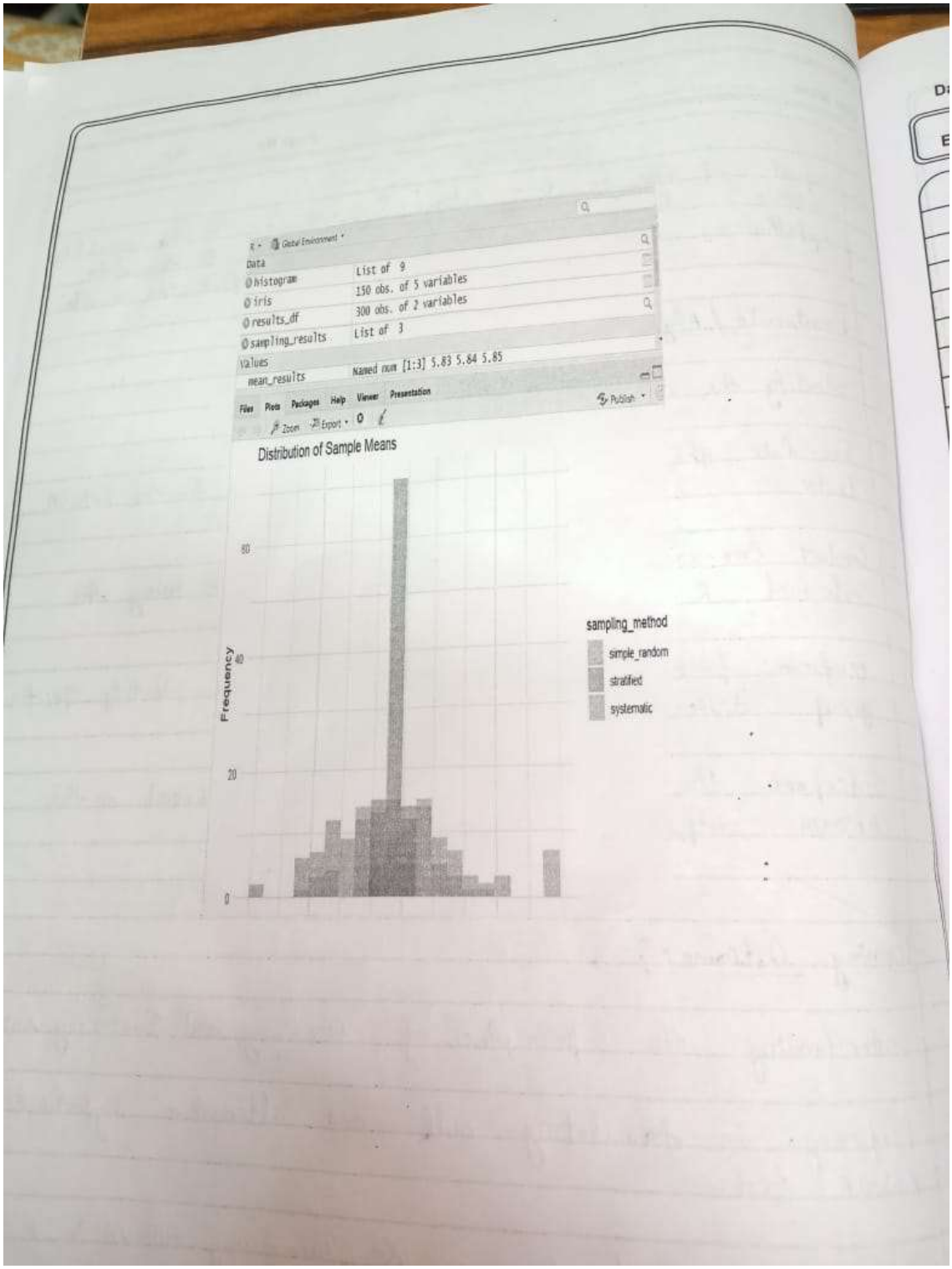
Page No. 26

4. Interpretation of ANOVA results, including F-statistics and p-values.

5. Practical experience in identifying sources of variability among different groups and assessing interaction effects between factors.

9/10/24

Teacher's Signature:



SCREENSHOTS:

```
> mean_results <- sapply(sampling_results, mean)
> names(mean_results) <- sampling_methods
>
> # Print results
> print(mean_results)
simple_random    stratified    systematic
     5.832267      5.838022      5.848600
>
```

```r
# Load required libraries
library(dplyr)
library(ggplot2)

# Load the iris dataset
data("iris")

# Function to perform sampling and estimate mean using a given sampling method
perform_sampling <- function(dataset, sample_size, sampling_method) {
  if (sampling_method == "simple_random") {
    sampled_data <- dataset[sample(nrow(dataset), sample_size, replace = FALSE), ]
  } else if (sampling_method == "stratified") {
    sampled_data <- dataset %>%
      group_by(Species) %>%
      sample_n(sample_size, replace = FALSE)
  } else if (sampling_method == "systematic") {
    n <- nrow(dataset)
    k <- sample_size
    start <- sample(1:k, 1)
    indices <- seq(start, n, by = k)
    sampled_data <- dataset[indices, ]
  } else {
    stop("Invalid sampling method.")
  }

  return(mean(sampled_data$Sepal.Length))
}

# Define parameters
sample_size <- 30
num_iterations <- 100
sampling_methods <- c("simple_random", "stratified", "systematic")

# Perform sampling and estimate mean using each technique
sampling_results <- lapply(sampling_methods, function(method) {
  replicate(num_iterations, perform_sampling(iris, sample_size, method))
})

# Combine results into a data frame
results_df <- data.frame(
```

```r
# Perform sampling and estimate mean using each technique
sampling_results <- lapply(sampling_methods, function(method) {
  replicate(num_iterations, perform_sampling(iris, sample_size, method))
})

# Combine results into a data frame
results_df <- data.frame(
  sampling_method = rep(sampling_methods, each = num_iterations),
  mean = unlist(sampling_results)
)

# Plot histograms
histogram <- ggplot(results_df, aes(x = mean, fill = sampling_method)) +
  geom_histogram(position = "identity", alpha = 0.5, bins = 20) +
  labs(title = "Distribution of Sample Means", x = "Mean Sepal Length", y = "Frequency") +
  theme_minimal()

print(histogram)

# Calculate means
mean_results <- sapply(sampling_results, mean)
names(mean_results) <- sampling_methods

# Print results
print(mean_results)
```