Lab 6: Estimate Population Parameters (Mean, Variance) from Sample Data

Theory

1. **Population Parameters**:
   - The **mean** (μ) and **variance** (σ^2) are key parameters describing a population.
   - These parameters are often unknown and must be estimated from sample data.
2. **Sample Statistics**:
   - The **sample mean** ($\bar{X}$) is an unbiased estimator of μ:

     $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} Xi$

   - The **sample variance** (S^2) is an unbiased estimator of σ^2:

     $S^2 = \frac{1}{n} - \sum_{i=1}^{n} (X_i - \bar{X})^2$

Objective

Estimate the population mean and variance from sample data using point and interval estimation.

Pseudocode

1. Generate a sample of size n from a normal distribution with mean μ and variance σ^2.
2. Compute the sample mean $\bar{X}$ and sample variance S^2.
3. Compute the 95% confidence interval for the mean and variance.
4. Plot the histogram of the sample data with the mean and confidence intervals.

R Code

```
# Parameters
mu <- 5
sigma <- 2
n <- 30
N_sim <- 1000

# Generate sample data
set.seed(123)
sample_data <- rnorm(n, mean = mu, sd = sigma)

# Point estimates
sample_mean <- mean(sample_data)
sample_var <- var(sample_data)

# Confidence intervals
conf_int_mean <- t.test(sample_data)$conf.int
conf_int_var <- c((n-1)*sample_var / qchisq(0.975, df = n-1),
                  (n-1)*sample_var / qchisq(0.025, df = n-1))

# Output
print(paste("Sample Mean:", sample_mean))
print(paste("Sample Variance:", sample_var))
print(paste("95% CI for Mean:", conf_int_mean))
print(paste("95% CI for Variance:", conf_int_var))

# Graphical Output
par(mfrow = c(1, 2))

# Histogram with mean and CI
hist(sample_data, breaks = 30, col = "lightblue", main = "Sample Data with Mean", xlab =
"Value", border = "white")
abline(v = sample_mean, col = "red", lwd = 2)
abline(v = conf_int_mean, col = "blue", lty = 2)

# Boxplot
boxplot(sample_data, col = "lightgreen", main = "Boxplot of Sample Data", ylab = "Value"
)
```

Sample Input/Output

- **Input**: $\mu = 5$, $\sigma = 2$, $n = 30$
- **Output**:
    - Sample Mean: Close to 5.
    - Sample Variance: Close to 4.
    - Confidence Intervals: Include the true population parameters.