

# Assignment-6

Ashish Verma

2024-02-25

## Section 8.2 Problem #18

Solution(a)

```
# Given values
x <- 72.3
null_mean <- 75
population_sd <- 9
number_of_samples = 25

# Calculate z-score
z_score <- (x - null_mean) / (population_sd/sqrt(number_of_samples))
cat("The standard deviations (of X) below the null value is x_bar 5 72.3?", z_score, "\n")
```

```
## The standard deviations (of X) below the null value is x_bar 5 72.3? -1.5
```

Solution(b)

```
# Given values
alpha <- 0.002
critical_z_value <- qnorm(alpha)
print(critical_z_value)
```

```
## [1] -2.878162
```

```
# Make a conclusion
# Compare z-score with critical z-value for a left-tailed test
if (z_score < critical_z_value) {
  cat("Reject the null hypothesis. There is enough evidence to suggest that the true mean is less than ")
} else {
  cat("Fail to reject the null hypothesis. There is not enough evidence to suggest that the true mean is less than ")
}
```

```
## Fail to reject the null hypothesis. There is not enough evidence to suggest that the true mean is less than 75
```

Solution(c)

```

# Given values
alpha <- 0.002
null_mean <- 75
alternative_mean <- 70
population_sd <- 9
number_of_samples <- 25

# Calculate critical z-value for alpha
critical_z_value <- qnorm(alpha)

# Calculate non-centrality parameter lambda
lambda <- sqrt(number_of_samples) * (null_mean - alternative_mean) / population_sd

# Calculate Type II error (beta)
beta <- pnorm(critical_z_value + lambda)

cat("Probability of Type II error (beta) for mu = 70 (left-tailed test):", beta, "\n")

```

```
## Probability of Type II error (beta) for mu = 70 (left-tailed test): 0.4600198
```

Solution(d)

```

population_sd <- 9
alpha <- 0.002
beta = 0.01
null_mean <-75
alternative_mean <-70

sample_size <-(population_sd*(qnorm(alpha)+qnorm(beta))/(null_mean - alternative_mean))^2
cat("Sample size is necessary to ensure that beta(70)", ceiling(sample_size), "\n")

```

```
## Sample size is necessary to ensure that beta(70) 88
```

Solution(e)

```

# Given values
alternative_mean_76 <- 76
sample_size_100 <- 100
alpha = 0.01
z_alpha =qnorm(alpha)
# Calculate z-score for mu = 76
new_null_mean <- z_alpha*(population_sd/sqrt(number_of_samples))+75

new_z_score <-(new_null_mean - alternative_mean_76) / (population_sd/sqrt(number_of_samples))
type_1 <- pnorm(new_z_score,lower.tail = TRUE)
cat("Type 1 error ", type_1, "\n")

```

```
## Type 1 error 0.001976404
```

## Section 8.2 Problem #23

Solution(a)

```
# ALD observations
ald_data <- c(1.38, 0.44, 1.09, 0.75, 0.66, 1.28, 0.51, 0.39, 0.70, 0.46,
             0.54, 0.83, 0.58, 0.64, 1.30, 0.57, 0.43, 0.62, 1.00, 1.05,
             0.82, 1.10, 0.65, 0.99, 0.56, 0.56, 0.64, 0.45, 0.82, 1.06,
             0.41, 0.58, 0.66, 0.54, 0.83, 0.59, 0.51, 1.04, 0.85, 0.45,
             0.52, 0.58, 1.11, 0.34, 1.25, 0.38, 1.44, 1.28, 0.51)

# Summary statistics
summary_stats <- summary(ald_data)
mean_ald <- mean(ald_data)
sd_ald <- sd(ald_data)

# Print summary statistics
cat("Summary Statistics:\n")
```

## Summary Statistics:

```
print(summary_stats)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3400  0.5200  0.6400  0.7498  1.0000  1.4400
```

```
cat("\nMean ALD:", mean_ald, "\n")
```

##

## Mean ALD: 0.7497959

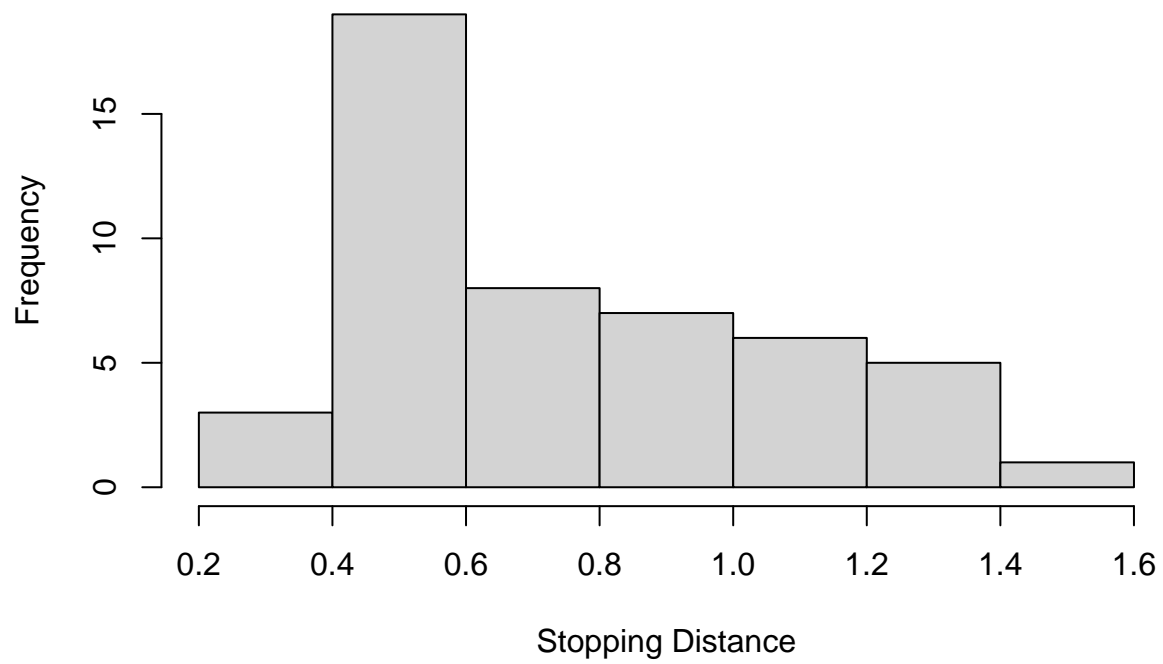
```
cat("Standard Deviation:", sd_ald, "\n")
```

## Standard Deviation: 0.3024655

Solution(b)

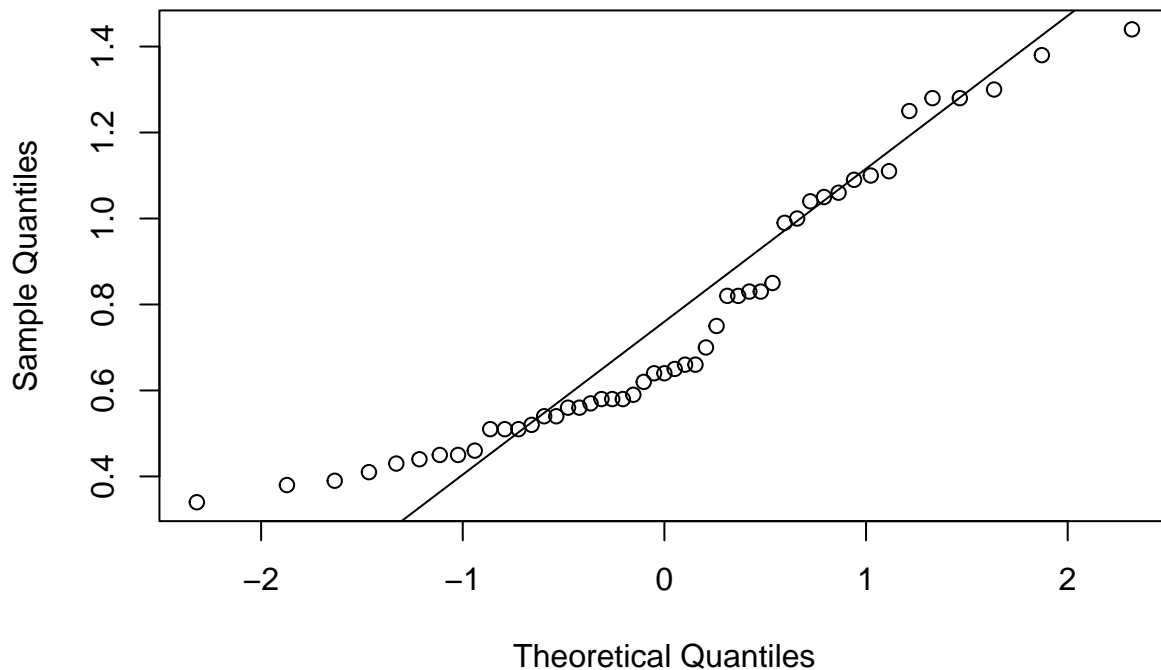
```
# Normality check using a normal probability plot
hist(ald_data, main="Histogram of ALD", xlab="Stopping Distance")
```

**Histogram of ALD**



```
qqnorm(ald_data)
qqline(ald_data)
```

## Normal Q-Q Plot



```
shapiro.test(ald_data)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  ald_data
## W = 0.90547, p-value = 0.0008329
```

The data is not normally distributed. Solution(c)

```
# Given data
ald_data <- c(1.38, 0.44, 1.09, 0.75, 0.66, 1.28, 0.51, 0.39, 0.70, 0.46,
             0.54, 0.83, 0.58, 0.64, 1.30, 0.57, 0.43, 0.62, 1.00, 1.05,
             0.82, 1.10, 0.65, 0.99, 0.56, 0.56, 0.64, 0.45, 0.82, 1.06,
             0.41, 0.58, 0.66, 0.54, 0.83, 0.59, 0.51, 1.04, 0.85, 0.45,
             0.52, 0.58, 1.11, 0.34, 1.25, 0.38, 1.44, 1.28, 0.51)

# Hypotheses
mu_0 <- 1.0 # hypothesized true average ALD
alternative <- "less" # one-sided test

# One-sample t-test
t_test_result <- t.test(ald_data, mu = mu_0, alternative = alternative)

# Print test result
cat("One-Sample t-Test Result:\n")
```

```
## One-Sample t-Test Result:
```

```
print(t_test_result)
```

```
##  
## One Sample t-test  
##  
## data: ald_data  
## t = -5.7905, df = 48, p-value = 2.615e-07  
## alternative hypothesis: true mean is less than 1  
## 95 percent confidence interval:  
##      -Inf 0.8222677  
## sample estimates:  
## mean of x  
## 0.7497959
```

```
# Interpretation based on p-value  
alpha <- 0.05 # significance level  
p_value <- t_test_result$p.value  
  
cat("\nInterpretation:\n")
```

```
##  
## Interpretation:
```

```
if (p_value < alpha) {  
  cat("The p-value is less than the significance level (", alpha, ")\n")  
  cat("we reject the null hypothesis and conclude that there is strong evidence\n")  
  cat("that the true average ALD is less than 1.0.\n")  
} else {  
  cat("The p-value is not less than the significance level (", alpha, ")\n")  
  cat("we do not reject the null hypothesis.\n")  
}
```

```
## The p-value is less than the significance level ( 0.05 ),  
## we reject the null hypothesis and conclude that there is strong evidence  
## that the true average ALD is less than 1.0.
```

Solution(d)

```
# Given data  
ald_data <- c(1.38, 0.44, 1.09, 0.75, 0.66, 1.28, 0.51, 0.39, 0.70, 0.46,  
             0.54, 0.83, 0.58, 0.64, 1.30, 0.57, 0.43, 0.62, 1.00, 1.05,  
             0.82, 1.10, 0.65, 0.99, 0.56, 0.56, 0.64, 0.45, 0.82, 1.06,  
             0.41, 0.58, 0.66, 0.54, 0.83, 0.59, 0.51, 1.04, 0.85, 0.45,  
             0.52, 0.58, 1.11, 0.34, 1.25, 0.38, 1.44, 1.28, 0.51)  
  
# Confidence level  
confidence_level <- 0.95
```

```

# Calculate upper confidence bound
upper_ci <- t.test(ald_data, conf.level = confidence_level)$conf.int[2]

# Print the result
cat("Upper Confidence Bound (", confidence_level * 100, "%):", upper_ci, "\n")

## Upper Confidence Bound ( 95 %): 0.8366742

# Interpretation
cat("\nInterpretation:\n")

##
## Interpretation:

cat("We are 95% confident that the true average ALD is less than", upper_ci, "\n")

## We are 95% confident that the true average ALD is less than 0.8366742

```

### Section 8.3 Problem #34

Solution(a)

```

# Stopping distance data
data <- c(32.1, 30.6, 31.4, 30.4, 31.0, 31.9)

# Hypothesized mean
mu_0 <- 30

# Significance level
alpha <- 0.01

# One-sample t-test
result <- t.test(data, mu = mu_0, alternative = "greater")

# Check if p-value is less than alpha
reject_null <- result$p.value < alpha

cat("a. Hypothesis Test Result:\n")

## a. Hypothesis Test Result:

cat("T-statistic:", result$statistic, "\n")

## T-statistic: 4.384921

cat("P-value:", result$p.value, "\n")

## P-value: 0.003560841

```

```
cat("Reject Null Hypothesis:", reject_null, "\n")
```

```
## Reject Null Hypothesis: TRUE
```

Conclusion: We conclude that the true average stopping distance exceeds the maximum value.

Solution(b)

```
# Given values for the Type II error calculation
```

```
mu_0 <- 30
```

```
mu_1 <- 31
```

```
mu_2 <- 32
```

```
sigma <- 0.65
```

```
n<-length(data)
```

```
df <-5
```

```
critical_t <- qt(1 - 0.01, df)
```

```
# Calculate the z-score for the actual mean (mu_1)
```

```
z <- (mu_1 - mu_0) / (sigma / sqrt(n))
```

```
# Calculate the t-statistic for the actual mean (mu_1)
```

```
t_stat_type_II <- (mu_1 - mu_0) / (sigma / sqrt(n))
```

```
# The probability of Type II error is the probability that the t-statistic is less than the critical t-  
# when the true mean is mu_1.
```

```
# This is the CDF of the non-central t-distribution with non-centrality parameter calculated above and
```

```
beta_1 <- pt(critical_t, df, ncp = sqrt(n) * (mu_1 - mu_0) / sigma)
```

```
beta_2 <- pt(critical_t, df, ncp = sqrt(n) * (mu_2 - mu_0) / sigma)
```

```
cat("b. Probability of Type II Error (Scenario 1):", beta_1, "\n")
```

```
## b. Probability of Type II Error (Scenario 1): 0.3402245
```

```
cat("b. Probability of Type II Error (Scenario 2):", beta_2, "\n")
```

```
## b. Probability of Type II Error (Scenario 2): 0.002079655
```

Solution(c)

```
# Given values for Scenario 2
```

```
mu_0 <- 30
```

```
mu_1 <- 31
```

```
mu_2 <- 32
```

```
sigma <- 0.80
```

```
n<-length(data)
```

```
df <-5
```

```
critical_t <- qt(1 - 0.01, df)
```



```

# Calculate the z-score for the actual mean (mu_1)
z <- (mu_1 - mu_0) / (sigma / sqrt(n))

# Calculate the t-statistic for the actual mean (mu_1)
t_stat_type_II <- (mu_1 - mu_0) / (sigma / sqrt(n))

# The probability of Type II error is the probability that the t-statistic is less than the critical t-
# when the true mean is mu_1.
# This is the CDF of the non-central t-distribution with non-centrality parameter calculated above and
beta_1_1 <- pt(critical_t, df, ncp = sqrt(n) * (mu_1 - mu_0) / sigma)

beta_2_1 <- pt(critical_t, df, ncp = sqrt(n) * (mu_2 - mu_0) / sigma)

cat("b. Probability of Type II Error (Scenario 1):", beta_1_1, "\n")

```

```
## b. Probability of Type II Error (Scenario 1): 0.5298098
```

```
cat("b. Probability of Type II Error (Scenario 2):", beta_2_1, "\n")
```

```
## b. Probability of Type II Error (Scenario 2): 0.02451232
```

The probability of Type 2 error for  $\mu = 31$  remains same but for  $\mu = 32$  increases.

Solution(d)

```

# Given values for Scenario 3
# Given values for sample size calculation
alpha <- 0.01
beta_desired <- 0.10
mu_1 <- 31

# Z-scores for alpha and power (1-beta)
z_alpha <- qnorm(1 - alpha)
z_beta <- qnorm(1 - beta_desired)

effect_size <- (mu_1 - mu_0) / sigma

# Sample size estimation using the formula for power of a test
sample_size <- ((z_alpha + z_beta)^2 * sigma^2) / effect_size^2

ceiling(sample_size) # Always round up to the next whole number

```

```
## [1] 6
```

### Section 8.3 Problem #36

Solution(a) We will not assume the renaming amount of toothpaste to be normally distributed because we don't know its population, and for applying the CLT we need sample of size greater than 40, but as the sample size is small we cannot assume it to be normal. Solution(b)

$$H_0 : \mu \geq 6$$

$$H_a : \mu < 6$$

```

sample_mean <- 0.502
sample_stddev <- 0.1023
sample_size <- 5
population_mean <- 0.6
alpha <- 0.05

# Calculate the standard error of the mean
standard_error <- sample_stddev / sqrt(sample_size)

# Calculate the t-statistic
t_statistic <- (sample_mean - population_mean) / standard_error

# Critical value for a one-tailed test
t_critical <- qt(alpha, df = sample_size - 1, lower.tail = TRUE)

# Compare the t-statistic with the critical value
decision <- ifelse(t_statistic < t_critical, "Reject H0", "Fail to Reject H0")

# Calculate the p-value for a one-tailed test
p_value <- pt(t_statistic, df = sample_size - 1, lower.tail = TRUE)

cat("T-statistic:", t_statistic, "\n")

```

```
## T-statistic: -2.142079
```

```
cat("Critical Value:", t_critical, "\n")
```

```
## Critical Value: -2.131847
```

```
cat("Decision:", decision, "\n")
```

```
## Decision: Reject H0
```

```
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.04942836
```

```

# Type I Error (False Positive)
type_I_error <- pnorm(t_critical, lower.tail = TRUE)

# Type II Error (False Negative)
type_II_error <- 1 - pnorm(t_critical, lower.tail = FALSE)

cat("Type I Error (False Positive):", type_I_error, "\n")

```

```
## Type I Error (False Positive): 0.01650972
```

```
cat("Type II Error (False Negative):", type_II_error, "\n")
```

```
## Type II Error (False Negative): 0.01650972
```

The rejection of the null hypothesis implies that there is evidence to suggest that the true population mean ( $\mu$ ) is less than 0.6. The risk of Type I Error is present, meaning there's a chance that the null hypothesis was true, but we incorrectly rejected it. If the null hypothesis were not rejected, there would be a risk of Type II Error, indicating a failure to reject a false null hypothesis

## Part 2

The null hypothesis  $H_0$  is that the new fertilizer does not have any effect on wheat production, and the alternative hypothesis  $H_a$  is that the new fertilizer will increase the production.

$$H_0 : \mu = 3$$

$$H_a : \mu > 3$$

```
# Provided data
data <- c(2.5, 3.0, 3.1, 4.0, 1.2, 5.0, 4.1, 3.9, 3.2, 3.3, 2.8, 4.1, 2.7, 2.9, 3.7)

# Null hypothesis: The mean yield is the same as the previous yield (3 bushels)
null_mean <- 3

# Sample mean and standard deviation
sample_mean <- mean(data)
sample_sd <- sd(data)

# Number of observations
n <- length(data)

# Calculate the t-statistic
t_statistic <- (sample_mean - null_mean) / (sample_sd / sqrt(n))

# Degrees of freedom
df <- n - 1

# Calculate the p-value
p_value <- pt(t_statistic, df, lower.tail = FALSE)

# Display the results
cat("Test Statistic:", t_statistic, "\n")
```

```
## Test Statistic: 1.302532
```

```
cat("P-value:", p_value, "\n")
```

```
## P-value: 0.1068774
```

```
# Check if the p-value is less than the significance level (e.g., 0.05)
if (p_value <= 0.05) {
  cat("Reject the null hypothesis. There is evidence that the new fertilizer increases wheat production")
} else {
  cat("Fail to reject the null hypothesis. There is not enough evidence to conclude that the new fertilizer")
}
```

```
## Fail to reject the null hypothesis. There is not enough evidence to conclude that the new fertilizer
```