Homework 7

Han Nguyen - TXN200004

11/13/2025

Question 1

```
# Data: Cola can fill amounts (ounces)
fill <- c(15.997, 16.005, 15.981, 15.954, 15.986, 16.021, 15.985, 16.001, 16.018, 16.056)
```

Step 1: State the null and alternative hypotheses

- H_0: $\mu = 16.00$ (The mean fill of cola cans is 16.00 ounces)
- H_a: $\mu < 16.00$ (The mean fill of cola cans is less than 16.00 ounces)

This is a one-sided (left-tailed) test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

Since we have a small sample (n = 10) and we don't know the population standard deviation, we use a one-sample t-test with the test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x}$ is the sample mean, $\mu_0 = 16.00$ is the hypothesized mean, s is the sample standard deviation, and $n = 10$ is the sample size.

Step 4: Formulate the decision rule

With $\alpha = 0.05$ and degrees of freedom $df = n - 1 = 9$, we will reject H_0 if:

- $t < -t_{\alpha,df}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

```
# Critical value for one-sided test
alpha <- 0.05
df <- length(fill) - 1
t_critical <- qt(alpha, df)
```

The critical value is $t_{0.05,9} =$ -1.8331.

Step 5: Collect data and calculate the test statistic

```
# Sample statistics
n <- length(fill)
x_bar <- mean(fill)
s <- sd(fill)
mu_0 <- 16.00
```

```r
# Calculate test statistic
t_stat <- (x_bar - mu_0) / (s / sqrt(n))

# Calculate p-value (one-sided, left-tailed)
p_value <- pt(t_stat, df)
```

Sample mean: $\bar{x} = 16.0004$ ounces

Sample standard deviation: $s = 0.0276$ ounces

Test statistic: $t = 0.0459$

Step 6: Make a decision

```r
# Decision based on critical value
reject_H0_critical <- t_stat < t_critical

# Decision based on p-value
reject_H0_pvalue <- p_value < alpha
```

P-value = 0.5178

Since $t = 0.0459$ is greater than the critical value of -1.8331, and the p-value (0.5178) is greater than $\alpha = 0.05$, we fail to reject H_0.

Step 7: Draw a conclusion

Based on the sample of 10 cola cans, we fail to reject the null hypothesis at the 0.05 significance level. There is not sufficient statistical evidence to conclude that the mean fill of cola cans is less than 16.00 ounces. Mark's data does not support the claim that the cans contain less than 16 ounces.

Question 2

```r
# Given data
n_2 <- 60
x_bar_2 <- 8.412
s_2 <- 1.512
mu_0_2 <- 8.2
```

Step 1: State the null and alternative hypotheses

- H_0: $\mu = 8.2$ (The mean app engagement time is 8.2 minutes)
- H_a: $\mu > 8.2$ (The mean app engagement time is more than 8.2 minutes)

This is a one-sided (right-tailed) test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

Since we have a large sample (n = 60) and we don't know the population standard deviation, we use a one-sample t-test with the test statistic:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x} = 8.412$ is the sample mean, $\mu_0 = 8.2$ is the hypothesized mean, $s = 1.512$ is the sample standard deviation, and $n = 60$ is the sample size.

Step 4: Formulate the decision rule

With $\alpha = 0.05$ and degrees of freedom $df = n - 1 = 59$, we will reject H_0 if:

- $t > t_{\alpha,df}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

```
# Critical value for one-sided test (right-tailed)
alpha_2 <- 0.05
df_2 <- n_2 - 1
t_critical_2 <- qt(1 - alpha_2, df_2)
```

The critical value is $t_{0.05,59} = 1.6711$.

Step 5: Collect data and calculate the test statistic

```
# Calculate test statistic
t_stat_2 <- (x_bar_2 - mu_0_2) / (s_2 / sqrt(n_2))

# Calculate p-value (one-sided, right-tailed)
p_value_2 <- 1 - pt(t_stat_2, df_2)
```

Sample mean: $\bar{x} = 8.412$ minutes

Sample standard deviation: $s = 1.512$ minutes

Sample size: $n = 60$

Test statistic: $t = 1.0861$

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_2 <- t_stat_2 > t_critical_2

# Decision based on p-value
reject_H0_pvalue_2 <- p_value_2 < alpha_2
```

P-value $= 0.1409$

Since $t = 1.0861$ is less than the critical value of 1.6711, and the p-value (0.1409) is greater than $\alpha = 0.05$, we fail to reject H_0.

Step 7: Draw a conclusion

Based on the sample of 60 tablet users, we fail to reject the null hypothesis at the 0.05 significance level. There is not sufficient statistical evidence to conclude that the mean app engagement time is more than 8.2 minutes. John's data does not provide sufficient evidence to prove Mary wrong about the 8.2 minute engagement time.

```
# We can't use t.test() directly since we don't have raw data,
# but we can verify our calculations manually
cat("Manual calculation verification:\n")
```

```
## Manual calculation verification:
```

```
cat("Test statistic t =", round(t_stat_2, 4), "\n")
```

```
## Test statistic t = 1.0861
```

```r
cat("P-value =", round(p_value_2, 4), "\n")
```

```
## P-value = 0.1409
```

```r
cat("Decision:", ifelse(reject_H0_pvalue_2, "Reject H0", "Fail to reject H0"), "\n")
```

```
## Decision: Fail to reject H0
```

Question 3

```r
# Given data
n_3 <- 200
x_3 <- 130
p_0_3 <- 0.70
```

Step 1: State the null and alternative hypotheses

- H_0: $p = 0.70$ (The proportion of stressed college students is 70%)
- H_a: $p \neq 0.70$ (The proportion of stressed college students is different from 70%)

This is a two-sided test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

For a proportion test with large sample size, we use the z-test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where $\hat{p} = \frac{x}{n}$ is the sample proportion, $p_0 = 0.70$ is the hypothesized proportion, and $n = 200$ is the sample size.

Step 4: Formulate the decision rule

With $\alpha = 0.05$ for a two-sided test, we will reject H_0 if:

- $|z| > z_{\alpha/2}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

```r
# Critical value for two-sided test
alpha_3 <- 0.05
z_critical_3 <- qnorm(1 - alpha_3/2)
```

The critical values are $\pm z_{0.025} = \pm$ 1.96.

Step 5: Collect data and calculate the test statistic

```r
# Sample proportion
p_hat_3 <- x_3 / n_3

# Standard error
se_3 <- sqrt(p_0_3 * (1 - p_0_3) / n_3)

# Test statistic
z_stat_3 <- (p_hat_3 - p_0_3) / se_3
```

```r
# P-value (two-sided)
p_value_3 <- 2 * (1 - pnorm(abs(z_stat_3)))
```

Sample proportion: $\hat{p} = \frac{130}{200} = 0.65$

Test statistic: $z =$ -1.543

Step 6: Make a decision

```r
# Decision based on critical value
reject_H0_critical_3 <- abs(z_stat_3) > z_critical_3

# Decision based on p-value
reject_H0_pvalue_3 <- p_value_3 < alpha_3
```

P-value = 0.1228

Since $|z| = 1.543$ is less than the critical value of 1.96, and the p-value (0.1228) is greater than $\alpha = 0.05$, we fail to reject H_0.

Step 7: Draw a conclusion

Based on the sample of 200 students, we fail to reject the null hypothesis at the 0.05 significance level. There is not sufficient statistical evidence to conclude that the proportion of stressed college students at this campus is different from 70%. The sample proportion of 0.65 (65%) is consistent with the claimed 70%.

```r
# Verification using prop.test function in R
prop_test_result_3 <- prop.test(x_3, n_3, p_0_3, alternative = "two.sided", correct = FALSE)
print(prop_test_result_3)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x_3 out of n_3, null probability p_0_3
## X-squared = 2.381, df = 1, p-value = 0.1228
## alternative hypothesis: true p is not equal to 0.7
## 95 percent confidence interval:
##  0.5816346 0.7127118
## sample estimates:
##    p
## 0.65
```

Question 4

```r
# Given data
n_total_4 <- 500
prop_18_50 <- 0.70
prop_over_50 <- 0.30

n1_4 <- n_total_4 * prop_18_50  # 18-50 age group
n2_4 <- n_total_4 * prop_over_50  # Over 50 age group

x_bar1_4 <- 2.00  # Mean for 18-50 group (Gb)
x_bar2_4 <- 1.85  # Mean for over 50 group (Gb)
```

```r
s1_4 <- 0.812  # SD for 18-50 group
s2_4 <- 0.837  # SD for over 50 group
```

Step 1: State the null and alternative hypotheses

Let μ_1 = mean data usage for people aged 18-50 years

Let μ_2 = mean data usage for people over 50 years

- H_0: $\mu_1 = \mu_2$ or equivalently $\mu_1 - \mu_2 = 0$ (Both age groups use the same amount of data)
- H_a: $\mu_1 > \mu_2$ or equivalently $\mu_1 - \mu_2 > 0$ (People aged 18-50 use more data than those over 50)

This is a one-sided (right-tailed) test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

For comparing two independent means with large samples, we use the two-sample t-test statistic:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\bar{x}_1 = 2.00$, $\bar{x}_2 = 1.85$, $s_1 = 0.812$, $s_2 = 0.837$, $n_1 = 350$, $n_2 = 150$.

Step 4: Formulate the decision rule

```r
# Calculate degrees of freedom using Welch's approximation
se1_4 <- s1_4^2 / n1_4
se2_4 <- s2_4^2 / n2_4
df_4 <- (se1_4 + se2_4)^2 / (se1_4^2/(n1_4-1) + se2_4^2/(n2_4-1))

alpha_4 <- 0.05
t_critical_4 <- qt(1 - alpha_4, df_4)
```

With $\alpha = 0.05$ and degrees of freedom $df \approx 274.38$, we will reject H_0 if:

- $t > t_{\alpha,df}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

The critical value is $t_{0.05,274.38} = 1.6504$.

Step 5: Collect data and calculate the test statistic

```r
# Calculate pooled standard error
se_pooled_4 <- sqrt(s1_4^2/n1_4 + s2_4^2/n2_4)

# Test statistic
t_stat_4 <- (x_bar1_4 - x_bar2_4) / se_pooled_4

# P-value (one-sided, right-tailed)
p_value_4 <- 1 - pt(t_stat_4, df_4)
```

Sample mean for 18-50 group: $\bar{x}_1 = 2$ Gb

Sample mean for over 50 group: $\bar{x}_2 = 1.85$ Gb

Difference in means: $\bar{x}_1 - \bar{x}_2 = 0.15$ Gb

Test statistic: $t = 1.8528$

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_4 <- t_stat_4 > t_critical_4

# Decision based on p-value
reject_H0_pvalue_4 <- p_value_4 < alpha_4
```

P-value = 0.0325

Since $t = 1.8528$ is greater than the critical value of 1.6504, and the p-value (0.0325) is less than $\alpha = 0.05$, we reject H_0.

Step 7: Draw a conclusion

Based on the sample of 500 customers, we reject the null hypothesis at the 0.05 significance level. There is sufficient statistical evidence to conclude that people older than 50 years use less cell phone data than people aged 18-50 years. The data supports that younger customers (18-50) use significantly more data (2 Gb) than older customers (1.85 Gb).

Question 5

```
# Given data
n_male_5 <- 30
n_female_5 <- 40
x_male_5 <- 22  # Males who play video games
x_female_5 <- 24  # Females who play video games
```

Step 1: State the null and alternative hypotheses

Let p_1 = proportion of males who play video games

Let p_2 = proportion of females who play video games

- H_0: $p_1 = p_2$ or equivalently $p_1 - p_2 = 0$ (Equal proportions of males and females play video games)
- H_a: $p_1 > p_2$ or equivalently $p_1 - p_2 > 0$ (More males play video games than females)

This is a one-sided (right-tailed) test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

For comparing two proportions, we use the two-proportion z-test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$, and $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$ is the pooled proportion.

Step 4: Formulate the decision rule

```
alpha_5 <- 0.05
z_critical_5 <- qnorm(1 - alpha_5)
```

With $\alpha = 0.05$ for a one-sided test, we will reject H_0 if:

- $z > z_\alpha$ (critical value approach), or

- p-value $< \alpha$ (p-value approach)

The critical value is $z_{0.05} = 1.6449$.

Step 5: Collect data and calculate the test statistic

```
# Sample proportions
p_hat1_5 <- x_male_5 / n_male_5
p_hat2_5 <- x_female_5 / n_female_5

# Pooled proportion
p_hat_pooled_5 <- (x_male_5 + x_female_5) / (n_male_5 + n_female_5)

# Standard error
se_5 <- sqrt(p_hat_pooled_5 * (1 - p_hat_pooled_5) * (1/n_male_5 + 1/n_female_5))

# Test statistic
z_stat_5 <- (p_hat1_5 - p_hat2_5) / se_5

# P-value (one-sided, right-tailed)
p_value_5 <- 1 - pnorm(z_stat_5)
```

Male proportion: $\hat{p}_1 = \frac{22}{30} = 0.7333$ (73.33%)

Female proportion: $\hat{p}_2 = \frac{24}{40} = 0.6$ (60%)

Pooled proportion: $\hat{p} = 0.6571$

Test statistic: $z = 1.163$

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_5 <- z_stat_5 > z_critical_5

# Decision based on p-value
reject_H0_pvalue_5 <- p_value_5 < alpha_5
```

P-value $= 0.1224$

Since $z = 1.163$ is less than the critical value of 1.6449, and the p-value (0.1224) is greater than $\alpha = 0.05$, we fail to reject H_0.

Step 7: Draw a conclusion

Based on the sample of 70 students (30 males, 40 females), we fail to reject the null hypothesis at the 0.05 significance level. There is not sufficient statistical evidence to conclude that more males play video games than females among STAT 3355 students. While the male proportion (73.33%) is higher than the female proportion (60%), this difference is not statistically significant.

Question 6

```
# Given data
n_6 <- 75
x_6 <- 30  # Number without insurance
p_0_6 <- 0.281  # Nationwide proportion
```

Step 1: State the null and alternative hypotheses

- H_0: $p = 0.281$ (The proportion of uninsured recent college graduates is 28.1%)

- H_a: $p \neq 0.281$ (The proportion of uninsured recent college graduates differs from 28.1%)

This is a two-sided test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

For a proportion test with large sample size, we use the z-test statistic:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where $\hat{p} = \frac{x}{n}$ is the sample proportion, $p_0 = 0.281$ is the hypothesized proportion, and $n = 75$ is the sample size.

Step 4: Formulate the decision rule

```
alpha_6 <- 0.05
z_critical_6 <- qnorm(1 - alpha_6/2)
```

With $\alpha = 0.05$ for a two-sided test, we will reject H_0 if:

- $|z| > z_{\alpha/2}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

The critical values are $\pm z_{0.025} = \pm\, 1.96$.

Step 5: Collect data and calculate the test statistic

```
# Sample proportion
p_hat_6 <- x_6 / n_6

# Standard error
se_6 <- sqrt(p_0_6 * (1 - p_0_6) / n_6)

# Test statistic
z_stat_6 <- (p_hat_6 - p_0_6) / se_6

# P-value (two-sided)
p_value_6 <- 2 * (1 - pnorm(abs(z_stat_6)))
```

Sample proportion: $\hat{p} = \frac{30}{75} = 0.4$ (40%)

Test statistic: $z = 2.2928$

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_6 <- abs(z_stat_6) > z_critical_6

# Decision based on p-value
reject_H0_pvalue_6 <- p_value_6 < alpha_6
```

P-value $= 0.0219$

Since $|z| = 2.2928$ is greater than the critical value of 1.96, and the p-value (0.0219) is less than $\alpha = 0.05$, we reject H_0.

Step 7: Draw a conclusion

Based on the sample of 75 recent college graduates, we reject the null hypothesis at the 0.05 significance level. There is sufficient statistical evidence to conclude that the proportion of uninsured recent college graduates differs from the nationwide proportion of 28.1%. The sample proportion (40%) is significantly different from the national rate.

```
# Verification using prop.test function in R
prop_test_result_6 <- prop.test(x_6, n_6, p_0_6, alternative = "two.sided", correct = FALSE)
print(prop_test_result_6)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x_6 out of n_6, null probability p_0_6
## X-squared = 5.2568, df = 1, p-value = 0.02186
## alternative hypothesis: true p is not equal to 0.281
## 95 percent confidence interval:
##  0.2966252 0.5131196
## sample estimates:
##   p
## 0.4
```

Question 7

```
# Given data
n_apple_7 <- 150
x_apple_7 <- 14  # Apple returns

n_samsung_7 <- 125
x_samsung_7 <- 15  # Samsung returns
```

Step 1: State the null and alternative hypotheses

Let p_1 = proportion of Apple iPhones returned

Let p_2 = proportion of Samsung Galaxy phones returned

- H_0: $p_1 = p_2$ or equivalently $p_1 - p_2 = 0$ (Both brands have the same return rate)
- H_a: $p_1 < p_2$ or equivalently $p_1 - p_2 < 0$ (Apple has a smaller return rate than Samsung)

This is a one-sided (left-tailed) test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

For comparing two proportions, we use the two-proportion z-test statistic:

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$, and $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$ is the pooled proportion.

Step 4: Formulate the decision rule

```
alpha_7 <- 0.05
z_critical_7 <- qnorm(alpha_7)
```

With $\alpha = 0.05$ for a one-sided (left-tailed) test, we will reject H_0 if:

- $z < -z_\alpha$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

The critical value is $-z_{0.05} =$ -1.6449.

Step 5: Collect data and calculate the test statistic

```
# Sample proportions
p_hat1_7 <- x_apple_7 / n_apple_7
p_hat2_7 <- x_samsung_7 / n_samsung_7

# Pooled proportion
p_hat_pooled_7 <- (x_apple_7 + x_samsung_7) / (n_apple_7 + n_samsung_7)

# Standard error
se_7 <- sqrt(p_hat_pooled_7 * (1 - p_hat_pooled_7) * (1/n_apple_7 + 1/n_samsung_7))

# Test statistic
z_stat_7 <- (p_hat1_7 - p_hat2_7) / se_7

# P-value (one-sided, left-tailed)
p_value_7 <- pnorm(z_stat_7)
```

Apple return rate: $\hat{p}_1 = \frac{14}{150} = 0.0933$ (9.33%)

Samsung return rate: $\hat{p}_2 = \frac{15}{125} = 0.12$ (12%)

Pooled proportion: $\hat{p} = 0.1055$

Test statistic: $z =$ -0.7169

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_7 <- z_stat_7 < z_critical_7

# Decision based on p-value
reject_H0_pvalue_7 <- p_value_7 < alpha_7
```

P-value $= 0.2367$

Since $z =$ -0.7169 is greater than the critical value of -1.6449, and the p-value (0.2367) is greater than $\alpha = 0.05$, we fail to reject H_0.

Step 7: Draw a conclusion

Based on the sample of 275 phones (150 Apple, 125 Samsung), we fail to reject the null hypothesis at the 0.05 significance level. There is not sufficient statistical evidence to conclude that Apple has a smaller return rate than Samsung. While Apple's return rate (9.33%) is numerically lower than Samsung's (12%), this difference is not statistically significant.

```
# Verification using prop.test function in R
prop_test_result_7 <- prop.test(c(x_apple_7, x_samsung_7), c(n_apple_7, n_samsung_7),
                                 alternative = "less", correct = FALSE)
print(prop_test_result_7)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
```

```
## data:  c(x_apple_7, x_samsung_7) out of c(n_apple_7, n_samsung_7)
## X-squared = 0.51397, df = 1, p-value = 0.2367
## alternative hypothesis: less
## 95 percent confidence interval:
##  -1.00000000  0.03507449
## sample estimates:
##     prop 1     prop 2
## 0.09333333 0.12000000
```

Question 8

```r
# Load the UsingR package and babies dataset
library(UsingR)
```

```
## Loading required package: MASS

## Loading required package: HistData

## Loading required package: Hmisc

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```r
data(babies)

# Extract mother's age and father's age
# Remove missing values (NA)
mothers_age <- babies$age
fathers_age <- babies$dage

# Create complete cases (pairs without missing values)
complete_cases <- complete.cases(mothers_age, fathers_age)
mothers_age_clean <- mothers_age[complete_cases]
fathers_age_clean <- fathers_age[complete_cases]
```

Step 1: State the null and alternative hypotheses

Let μ_d = mean difference in ages (father's age - mother's age)

- H_0: $\mu_d = 0$ (The mean ages of mothers and fathers are equal)
- H_a: $\mu_d \neq 0$ (The mean ages of mothers and fathers are different)

This is a two-sided paired t-test.

Step 2: Choose the significance level

$\alpha = 0.05$

Step 3: Determine the appropriate test statistic

Since we have paired data (mother and father ages for each baby), we use a paired t-test with the test statistic:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}$$

where $\bar{d}$ is the mean of the differences, s_d is the standard deviation of the differences, and n is the number of pairs.

Step 4: Formulate the decision rule

```
alpha_8 <- 0.05
n_8 <- length(mothers_age_clean)
df_8 <- n_8 - 1
t_critical_8 <- qt(1 - alpha_8/2, df_8)
```

With $\alpha = 0.05$ and degrees of freedom $df = n - 1 = 1235$, we will reject H_0 if:

- $|t| > t_{\alpha/2,df}$ (critical value approach), or
- p-value $< \alpha$ (p-value approach)

The critical values are $\pm t_{0.025,1235} = \pm$ 1.9619.

Step 5: Collect data and calculate the test statistic

```
# Calculate differences (father - mother)
differences <- fathers_age_clean - mothers_age_clean

# Calculate mean and standard deviation of differences
d_bar <- mean(differences)
s_d <- sd(differences)

# Test statistic
t_stat_8 <- d_bar / (s_d / sqrt(n_8))

# P-value (two-sided)
p_value_8 <- 2 * (1 - pt(abs(t_stat_8), df_8))
```

Sample size: $n = 1236$ pairs

Mean mother's age: $\bar{x}_{mother} = 27.3714$ years

Mean father's age: $\bar{x}_{father} = 30.7371$ years

Mean difference (father - mother): $\bar{d} = 3.3657$ years

Standard deviation of differences: $s_d = 6.8035$ years

Test statistic: $t = 17.3922$

Step 6: Make a decision

```
# Decision based on critical value
reject_H0_critical_8 <- abs(t_stat_8) > t_critical_8

# Decision based on p-value
reject_H0_pvalue_8 <- p_value_8 < alpha_8
```

P-value $=$ 0e+00

Since $|t| = 17.3922$ is greater than the critical value of 1.9619, and the p-value (0e+00) is less than $\alpha = 0.05$, we reject H_0.

Step 7: Draw a conclusion

Based on the sample of 1236 mother-father pairs from the babies dataset, we reject the null hypothesis at the 0.05 significance level. There is sufficient statistical evidence to conclude that the mean ages of mothers and fathers are different in the sampled population. On average, fathers are 3.37 years older than mothers, and this difference is statistically significant.

```r
# Verification using t.test function in R (paired test)
t_test_result_8 <- t.test(fathers_age_clean, mothers_age_clean, paired = TRUE, alternative = "two.sided
print(t_test_result_8)
```

```
##
##  Paired t-test
##
## data:  fathers_age_clean and mothers_age_clean
## t = 17.392, df = 1235, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  2.986035 3.745356
## sample estimates:
## mean difference
##        3.365696
```