Summative Assessment 1

Vera Aguila

2024-10-07

A study was undertaken to compare the mean time spent on cell phones by male and female college students per week. Fifty male and 50 female students were selected from Midwestern University and the number of hours per week spent talking on their cell phones determined. The results in hours are shown in Table 10.6. It is desired to test $H_0: \mu_1 = \mu_2$ versus $H_1: \mu_1 \neq \mu_2$ based on these samples.

Table 10.6 Hours spent talking on cell phone for males and females at Midwestern University

Males	Females
12	11
7	10
7	11
10	10
8	11
10	12
11	12
9	10
9	9
13	9
4	9
9	10
12	8
11	7
9	12
9	9
7	7
12	8
10	9
13	8
11	7
10	7
6	9
12	9
11	12
9	10
10	9
12	13
8	9
9	9
13	10
10	9
9	6
7	12
10	8
7	11
10	8
8	8
11	11
10	12
11	9
7	10
15	11
8	14
9	12
9	7
11	11
13	10
10	9
13	11

Questions:

1. Formulate and present the rationale for a hypothesis test that the researcher could use to compare the mean time spent on cell phones by male and female college students per week.

Hypothesis Formulation:

The goal is to test if there is a significant difference in the mean time spent on cell phones by male and female college students. We can use a two-sample t-test (independent samples t-test) to compare the means of two independent groups: males and females.

- Null Hypothesis (H_0): There is no significant difference in the mean time spent on cell phones by males and females.

$H_0 : \mu_1 = \mu_2$ Where μ_1 is the mean time spent by males and μ_2 is the mean time spent by females.

- Alternative Hypothesis (H_1): There is a significant difference in the mean time spent on cell phones between males and females. $H_1 : \mu_1 \neq \mu_2$

This is a two-tailed test, as we are interested in any difference (whether males spend more or less time compared to females).

Rationale:

The rationale for the hypothesis test is to statistically evaluate whether any observed differences in the mean time spent on cell phones between males and females could have occurred by chance or if the differences are statistically significant, providing evidence that gender may influence cell phone use.

2. Analyze the data to provide the hypothesis testing conclusion. What is the p-value for your test? What is your recommendation for the researcher?

To analyze the data, we will perform a two-sample t-test assuming unequal variances (Welch's t-test), which is appropriate when the variance between groups is not assumed to be equal.

```
# Perform the two-sample t-test
t_test_result <- t.test(data$Males, data$Females, var.equal = FALSE)

t_test_result$p.value  # P-value
```

```
## [1] 0.7618345
```

```
t_test_result$statistic  # t-statistic
```

```
##         t
## 0.3039491
```

```
t_test_result$conf.int  # Confidence interval for the difference
```

```
## [1] -0.6638305  0.9038305
## attr(,"conf.level")
## [1] 0.95
```

Results:

- P-value: Based on the t-test, the p-value is calculated. If the p-value is less than the significance level (α = 0.05), we reject the null hypothesis.
- Recommendation: If p<0.05, the difference in means is statistically significant, and we recommend rejecting the null hypothesis, suggesting that there is a significant difference in the time males and females spend on their phones. If p ≥ 0.05, we fail to reject the null hypothesis, indicating no significant difference.

3. Provide descriptive statistical summaries of the data for each gender category.

We will compute the mean, standard deviation, and number of observations for both males and females:

```
# Descriptive statistics for Males
mean_male <- mean(data$Males)
sd_male <- sd(data$Males)
n_male <- length(data$Males)

# Descriptive statistics for Females
mean_female <- mean(data$Females)
sd_female <- sd(data$Females)
n_female <- length(data$Females)

# Output the results
list(mean_male = mean_male, sd_male = sd_male, n_male = n_male,
     mean_female = mean_female, sd_female = sd_female, n_female = n_female)
```

```
## $mean_male
## [1] 9.82
##
## $sd_male
## [1] 2.154161
##
## $n_male
## [1] 50
##
## $mean_female
## [1] 9.7
##
## $sd_female
## [1] 1.775686
##
## $n_female
## [1] 50
```

4. What is the 95% confidence interval for the population mean of each gender category, and what is the 95% confidence interval for the difference between the means of the two populations?

We can calculate the 95% confidence interval for each group and the difference between the means.

```
# 95% confidence intervals
ci_males <- t.test(data$Males)$conf.int
ci_females <- t.test(data$Females)$conf.int

# Confidence interval for the difference
ci_diff <- t_test_result$conf.int

list(ci_males = ci_males, ci_females = ci_females, ci_diff = ci_diff)
```

```
## $ci_males
## [1]  9.207794 10.432206
## attr(,"conf.level")
## [1] 0.95
##
## $ci_females
## [1]  9.195356 10.204644
## attr(,"conf.level")
## [1] 0.95
##
## $ci_diff
## [1] -0.6638305  0.9038305
## attr(,"conf.level")
## [1] 0.95
```

Results:

- 95% CI for Males: Lower and upper bounds of the confidence interval for the mean of males.
- 95% CI for Females: Lower and upper bounds of the confidence interval for the mean of females.
- 95% CI for the Difference: This gives the range within which the true difference between the population means lies with 95% confidence.

5. Do you see a need for larger sample sizes and more testing with the time spent on cell phones? Discuss.

Discussion:

- Sample Size: The current sample size of 50 males and 50 females provides a reasonable initial insight into the comparison. However, if the confidence intervals for the means and the difference are wide, it suggests that there may be uncertainty around the estimates, indicating the need for larger sample sizes.
- More Testing: Larger sample sizes could also reduce variability and increase the power of the hypothesis test, making it easier to detect smaller differences if they exist. Moreover, future tests could examine other factors that might influence phone usage (e.g., age, academic standing, major).

6. Make a report including the testing of the assumptions for two independent samples t-test.

Assumptions of the Two-Sample T-Test:

- Independence: The male and female groups are independent of each other.
- Normality: The data should be approximately normally distributed in both groups. We can check this using a normality test or visualize the distribution.
- Equality of Variances: The test assumes the variances of the two groups are equal. However, we use Welch's t-test here as it does not assume equal variances.

```
# Checking normality with Shapiro-Wilk test
shapiro.test(data$Males)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$Males
## W = 0.97465, p-value = 0.354
```

```
shapiro.test(data$Females)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data$Females
## W = 0.96388, p-value = 0.1292
```

```
# Checking equality of variances with Bartlett's test
bartlett.test(list(data$Males, data$Females))
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  list(data$Males, data$Females)
## Bartlett's K-squared = 1.7996, df = 1, p-value = 0.1798
```

Results:

- Normality: If the p-value of the Shapiro-Wilk test is greater than 0.05, the assumption of normality holds. If not, transformations or non-parametric tests (like the Mann-Whitney U test) may be needed.
- Equality of Variances: If the p-value of Bartlett's test is less than 0.05, the variances are unequal, justifying our use of Welch's t-test.