2021025081 DESQUITADO, Yuichi
APM1111 Statistical Theory – Summative Assessment 1

SIGNIFICANCE OF HOURS SPENT TALKING ON CELLPHONE BY THE MIDWESTERN UNIVERSITY STUDENTS BY GENDER

RATIONALE[1]

A study has been conducted to investigate whether the mean time spent on talking on cellphones has a significant difference between males and females in the Midwestern University. For the said study, 100 participants (50 for each gender) were gathered and asked for their time spent in hours on their devices as means of communication with others.

For this study, the hypothesis testing are as follows:

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

Null hypothesis suggests that there is no significant difference between the genders, while the Alternate hypothesis suggests otherwise.

RESULTS

The table below is the result of the said data gathering featuring the 100 students at Midwestern University, each half representing a distinct gender.

Male					Female				
12	4	11	13	11	11	9	7	10	9
7	9	10	10	7	10	10	7	9	10
7	12	6	9	15	11	8	9	6	11
10	11	12	7	8	10	7	9	12	14
8	9	11	10	9	11	12	12	8	12
10	9	9	7	9	12	9	10	11	7
11	7	10	10	11	12	7	9	8	11
9	12	12	8	13	10	8	13	8	10
9	10	8	11	10	9	9	9	11	9
13	13	9	10	13	9	8	9	12	11

Table 1. Hours spent talking on cellphones by Midwestern University students, N = 100

DATA PRE-ANALYSIS[2]

For this study, the researchers will be using Independent Samples T-Test (Student T-Test) since there are two means being compared sourcing from two different groups. As for the processing of the data, the writers of this paper will be using JASP. An alpha $\alpha = 0.05$ will be used (95% confidence level as the study did not mention any confidence interval, while 95% has been a standard/default for such test, implying that the researchers are 95% confident that the results are correct or 5% chance that it will be wrong[4]).

As we go through the dataset desiring to use Student's T-Test, we must do Assumption Checks 1-6 to determine whether the dataset is fit to go for the said test.

ASSUMPTION 1: You have one dependent variable that is measured at the continuous level.

Remark – For this study, the *Hours (Spent)* is the dependent variable measured at the continuous level.

ASSUMPTION 2: You have one independent variable that consists of three categorical, independent groups.

Remark – We have two categorical (independent) variables observed for the study, *Male* and *Female*.

ASSUMPTION 3: You should have independence of observations.

Remark – Each observation is independent of each other, wherein 50 students were gathered per gender, totaling 100 respondents.

1 – Answer for Question 1
2 – Answer for Question 2
4 – Answer for Question 4

ASSUMPTION 4: There should be no significant outliers in the three or more groups of your independent variable in terms of the dependent variable.

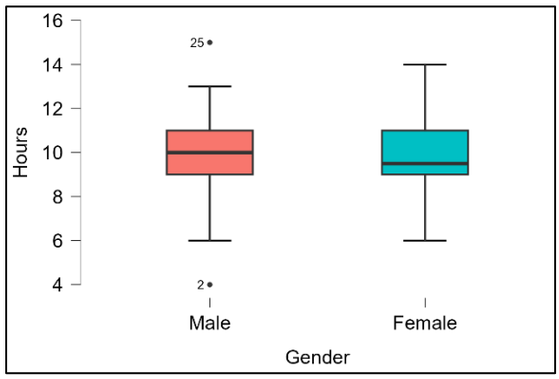

Figure 1. Boxplot of the gathered data

Remark – For this assumption check, we used visual techniques via JASP's boxplot to label data outliers. According to the boxplot, the group Male has the outliers in Row 2 (Hours = 4) and Row 25 (Hours = 15). Such outliers will not be removed as they are not extreme outliers for the data gathered.

ASSUMPTION 5: Your dependent variable should be approximately normally distributed for each group of the independent variable.

Test of Normality (Shapiro-Wilk)

		W	p
Hours	Male	0.975	0.354
	Female	0.964	0.129

Note. Significant results suggest a deviation from normality.

Table 2. Shapiro-Wilk's Test for Normality of the dataset

Remark – The Hours' normality will be checked using Shapiro-Wilk's Test via JASP. The aforementioned data gathered are normally distributed for each of the category,

$$p = 0.354, 0.129 > 0.05.$$

ASSUMPTION 6: You have homogeneity of variances (i.e., the variance of the dependent variable is equal in each group of your independent variable).

Test of Equality of Variances (Levene's)

	F	df_1	df_2	p
Hours	0.898	1	98	0.346

Table 3. Levene's Test of Equality of Variance of Hours

Remark – The variances between the two groups are homogenous as assessed by the Leven's Test, wherein $p = 0.346 > 0.05$.

COMPUTATION

Descriptive Statistics ▼

	Hours	
	Male	Female
Valid	50	50
Missing	0	0
Mean	9.820	9.700
Std. Deviation	2.154	1.776
Variance	4.640	3.153
Skewness	−0.115	0.134
Std. Error of Skewness	0.337	0.337
Kurtosis	0.177	−0.442
Std. Error of Kurtosis	0.662	0.662
Shapiro-Wilk	0.975	0.964
P-value of Shapiro-Wilk	0.354	0.129
Minimum	4.000	6.000
Maximum	15.000	14.000
25th percentile	9.000	9.000
50th percentile	10.000	9.500
75th percentile	11.000	11.000

Note. Excluded 1 rows from the analysis that correspond to the missing values of the split-by variable Gender

Table 4. Descriptive Statistics of the dataset[3]

Group Descriptives

	Group	N	Mean	SD	SE	Coefficient of variation
Hours	Male	50	9.820	2.154	0.305	0.219
	Female	50	9.700	1.776	0.251	0.183

Table 5. Group Description of Male and Female groups

Independent Samples T-Test

	t	df	p	Mean Difference	SE Difference
Hours	0.304	98	0.762	0.120	0.395

Note. Student's t-test.

Table 6. Independent T-Test (Student's T-Test) of Hours

3 – Answer for Question 3

REPORTING[6]

The researchers have investigated the time spent talking on cellphones by the Midwestern University students, which is divided by two categories according to their gender, Male and Female, each having 50 respondents (N = 100).

On Assumption 1-4, we have identified *Hours* as the dependent continuous variable, while *Male* and *Female* as independent categorical variables. By visual analysis using boxplot via JASP, Rows 2 (Male, 4 hours) and 25 (Male, 15 hours) were identified as outliers, but were not extreme outliers to be removed from the data, so they were retained.

For Assumption 5, Shapiro-Wilk's Test was used to check for the Normality of the dataset, which returned a *p-value* = *0.354*, *0.129*, which are both greater than $\alpha = 0.05$, thus implying their normality.

For Assumption 6, Levene's Test was utilized, which resulted to a *p-value* = 0.346, which is higher than $\alpha = 0.05$, signifying their homogeneity.

Since all the Assumption Checks were successful (the dataset fitted all the criteria), it is safe to use Independent T-Test (Student's T-Test) to get the *p-value* to determine the significance between the two categories. According to the computation via JASP,

$$p = 0.762 > 0.05.$$

The ***p-value* of the t-test is higher than the alpha** $\boldsymbol{\alpha = 0.05}$, which implies that the researchers were given insufficient evidence to reject the null hypothesis, thus the researchers ***fail to reject the null hypothesis*** in favor of the alternate hypothesis.

The data is presented as means $\pm$ standard deviation. It was found that the Male group ($\mu = 9.820,\ \sigma = 2.154,\ n = 50$) was not that significantly different from the Female group ($\mu = 9.7,\ \sigma = 1.776,\ n = 50$). $t(99) = 0.304,\ p > 0.05,\ d = 98$, 95% CI.

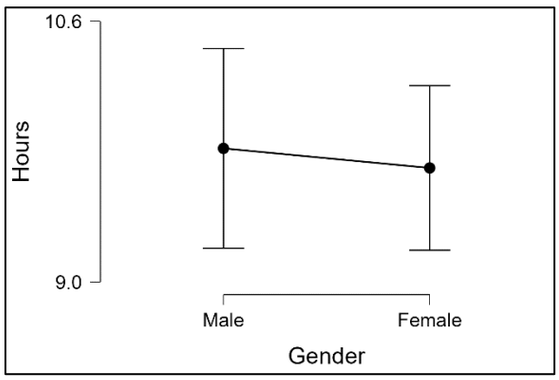

Figure 2. Descriptive Plot of the dataset

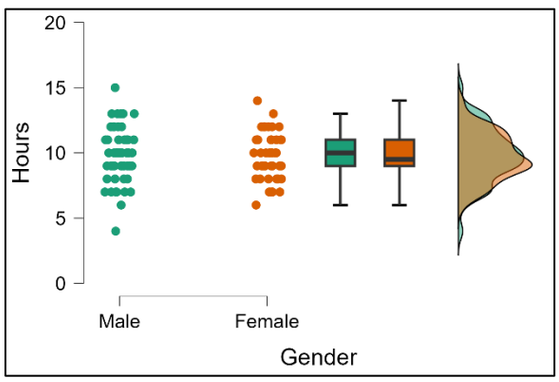

Figure 3. Raincloud plots of the dataset

RECOMMENDATIONS[5]

The researchers recommended that the study must be conducted with a larger sample size to reduce the chance of margin of error. Having a bigger sample size can also lower the risk of reporting false-positive/negative findings.

The higher the sample size, the higher the precision of the findings.

5 – Answer for Question 5
6 – Answer for Question 6