**QUESTION 1**

Since the variable ‘internet’ is not normal from the Shapiro test.

The p-value that we get is 2.2e-16 which is less than 0.05. Hence, we conclude that a non-parametric test needs to be conducted.

Here, the independent variable is continuous, and the dependent variable is binary and categorical, and the data is not normal. Hence, we perform Mann-Whitney U test.

After performing the Mann-Whitney U test, we get p-value as 0.0324 which is less than 0.05.

*Hence, we reject the null hypothesis and accept the alternate hypothesis. The alternate hypothesis states that there is a difference for G3 final grades between students with internet from students without internet.*

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**QUESTION 2**

Since, both the variables i.e., activities and higher education are binary and categorical, it is safe to assume that the data is not normal, and a non-parametric test needs to be conducted.

Here, both the variables are binary. Hence, we have to use chi square and fisher test to determine the outcome.

The Minimum Expected Frequency is 9.28 which is greater than 5.

Hence, we consider the p-value from chi squared test which is 0.05516458 which is greater than 0.05.

*Hence, we accept the null hypothesis and reject the alternative hypothesis. The null hypothesis states that there is no relationship between activities and higher education.*

Text

Description automatically generated

Table

Description automatically generated

Text

Description automatically generated

**QUESTION 3**

In this question, the data is continuous, and we check for the normality using the Shapiro test. It turns out that the p-value for G2 is 2.084e-07 and p-value for G3 is 8.836e-13. Both are less than 0.05 and hence the variables are not normal. It is safe to assume that a non-parametric paired test is required to test the before and after grades.

Since the data is not normal but paired at the same time, we use Wilcoxon Signed Rank Test.

The p-value that we get after Wilcoxon Signed Rank Test is 0.8748 which is greater than 0.05.

*Hence, we accept the null hypothesis and reject the alternative hypothesis. The null hypothesis states that there is no significant difference between G2 and G3.*

A picture containing text

Description automatically generated

Text

Description automatically generated

**QUESTION 4**

Since, the variables under consideration are binary, it is safe to assume that the data is not normal, and a non-parametric test is required to conduct hypothesis testing.

Since both the variables are not normal, binary, and paired, we use McNemar test for hypothesis testing.

After using McNemar test, the p-value that we get is 0.09609, which is greater than 0.05.

*Hence, we accept the null hypothesis and reject the alternative hypothesis. The null hypothesis states that there is no difference between G2\_cat and G3\_Cat.*

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated with medium confidence

Table

Description automatically generated

Text, letter

Description automatically generated with medium confidence

**QUESTION 5**

*It is quite evident from the output that even if the significance value changes from 95% to 50%, the probability of students receiving extra educational support will remain the same which is 0.1291139. The only thing that changes with the confidence level is the confidence interval. When the confidence level is 95%, the confidence interval is between 9.958378% to 16.579947%. But when we perform proportion test with confidence level as 50%, the confidence level is between 11.81591% and 14.09221%.*

A picture containing text

Description automatically generated

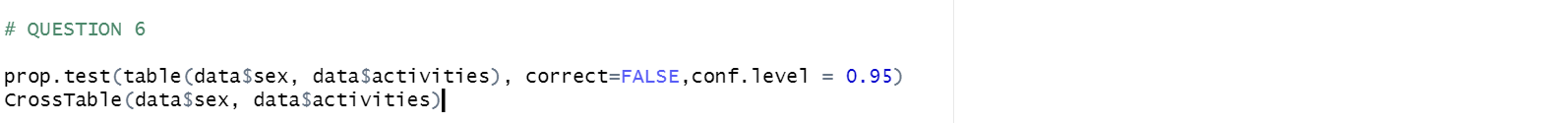
Graphical user interface, text, application, email

Description automatically generated

**QUESTION 6**

The table in the output shows the data for confidence interval of 95%.

Here, the proportion of Males in extracurricular activities is 26.6% and the proportion of Females in extracurricular activities is 24.3%.



Graphical user interface, text, application

Description automatically generated with medium confidence

Table

Description automatically generated

**QUESTION 7**

For data cleaning, I started with summary and describe function. Using these functions, I was able to deduce that there were no missing values in the table.

So, I checked with the NA or Null values using the is.na function. After using this function, the sum that was returned was 0. Hence, I concluded that the dataset was fairly clean and ready to use for further testing.

Text

Description automatically generated with medium confidence

**QUESTION 8**

Following are the limitations for these study questions:

* The study questions are fairly straightforward and do not dive deeper for any actual insights.
* The study questions only tells whether the variables depend on each other or not and fail to provide a quantity. For example, if internet affects the G3 grades, then by how much does it affect.
* The sample that is considered for this study is way too small and may lead to wrong conclusions.

**QUESTION 9**

From the answers to questions 1-6 we can conclude the following points:

* Having or not having internet has an impact on final grade of the student.
* A students involvement in extracurricular activity does not affect his/her higher education.
* There is not a significant improvement or deterioration, in the final grades of students when compared to their second grades.
* 12.9% students are receiving extra educational support.
* More males are involved in extracurricular activities than females.

**QUESTION 10**

The future work or additional research questions based upon findings of question 1-6 is as follows:

* If internet is affecting the grades of the students, then is it affecting in a good way or bad way?
* Does having internet means good final grades or bad final grades?
* If there is no relationship between activities and higher education, then are there any other variables that have a relationship with activities or higher education or both. If yes, then how do they affect each other?
* If there is no difference between G2 and G3 grade, then what are the factors that can bring a change in G2 and G3 grades?