Mental Health Survey for University Students

Assignment 1

MANJIMA BANERJEE - 1006926230

September 28th, 2022

Part 1

Goal

The goal of the survey is to understand how university students are affected by stress due to several reasons and what impacts may be caused in their body. High levels of stress for students in university has become a very concerning and emerging issue to address nowadays. Every student is experiencing some or the other amount of burnout, anxiety, fear or other issue that may be occurring due to stress, that some of the students may also not be able to handle. Due to these uncontrollable levels, a lot of students are also dropping out of university or in very extreme cases having 'suicidal thoughts' or depression. Therefore, with this survey, we try to get a glimpse of how undergraduate students at the University of Toronto are tackling their stress levels. We try to find the reason of their stress, what symptoms it may cause them physically or emotionally, how stressed they think they are and how healthy do they feel their body is from a day to day basis. Moreover, we may also be able to find other correlations of stress levels with factors such as hours of sleep, age, gender, level of healthiness, academic load (total credits of the year), any specific physical or mental symptom that may occur with higher stress for a majority of individuals, the main cause for students that relates to higher stress and so on.

Procedure

In order to implement this survey, the medium of usage is social media handles such as Instagram, Whatsapp, Snapchat and LinkedIn. To spread this survey across maximum students possible, I plan to use these handles to get responses. Moreover, through social media, users can also circulate the survey further to their known friends such that we get maximum respondents possible.Due to this, the simulation of our testing will be better and conclusions derived will be more accurate. Here, we see that the target population is all university students as the conclusions formulated from this survey will be applicable and generalized to all university students. The frame population is all the students studying at the University of Toronto and the sampling population is all the students studying at the University of Toronto who took part in the survey as their data will actually be worked on and used for further analysis.

Showcasing the survey.

Link to the survey: https://forms.gle/3dTqLjcPTtCf7ufV9 .

Choosing Questions from the survey -

Question 6 Which of the following are the Physical symptoms that you may be experiencing due to stress? (select all that apply)

For this question, we have focused on what physiological and physical problems may be faced by students undergoing some level of stress It is given as a 'check all that apply' category of question which has the most common stress symptoms stated by data from medical experts and researchers in patients. These symptoms include any uneasiness in the body that may be experienced on a day to day basis. The responses collected

from this question can help us understand the most common form of way that stress affects students' body in university. We may also be able to draw a correlation in a way such that high stress levels are generalized to have a specific symptom. Lastly, knowing the most common symptom can also be used for better treatment for the area of impact in the body along with other stress management techniques. However, I feel that the question gets slightly lengthy with so many options to refer to. For the audience, they would have to read every statement and contemplate on whether they are experiencing that for themselves. This becomes quite time consuming for them. For some people, it may also get confusing to see these options as a few sensations in the body may not be defined into a specific category. Moreover, as everyone has different ways of reaction, it is quite hard to include every small to major symptom that one may go through.

Question 8 How healthy do you feel your body is? (emotionally and physically)

This is a Linear scale question, 1 being the lowest in feeling of 'healthiness' and 5 being the highest. The aim of this question is for the audience to understand and interpret on how well they think their body is doing emotionally and physically. Through this scale, we can assess how intense the impact of stress levels are on your body and its functions. This is also a medium for the audience to reflect on how their stress may in some way be hindering their body physically and mentally. We may also be able to derive a correlation of the healthiness of the body with the level of stress as a conclusion. However, we see that this question cannot specify the details of what the level of 'healthiness' actually means to the person. For someone, healthiness can just be related to the fact that they may be able to get their work done for the day while for another student, they may believe that if they have even the slightest hint of anxiety, they are not healthy. The question therefore becomes too broad in its approach. Lastly, I also feel that as this question is placed in the later part of the survey after answering the symptoms on body, the students may get biased and be of the opinion that symptoms checked before relates to a lower level of healthiness.

Question 9 What are the most pressing stress factors in your current academic year?

The goal of this question is to understand the most significant reasons as to why students at the University of Toronto may experience increased stress. Through this, we may also be able to understand how or why students experience difficulties in that aspect of life. At the same time, if we know the cause of stress, we can also put in a larger proportion of effort into fixing that segment of life. The drawback of this question is that the information given in this question is very limited as there can be a variety of reasons that may affect a student. There can also be a situation wherein aspects not mentioned in the survey may play a role a large role in the student's overall stress level. This question could be improved by adding a sub question that may take into account other life factors contributing to negative feelings in a student.

Part 2

Data

The survey data includes the students' age, gender, course load, hours of sleep as well as other factors that may be related or are affected by stress. This survey data was collected using online social platforms as well as from circulation by posting the link and also sending it specifically to several students who are respondents. There were 71 responses for the survey, all of which were from University of Toronto students. Even though the combination of methods that I used gave me quite many responses, but it may still occur that some people may not have read the survey description stating that it is only for 'University of Toronto' students leading to an error in the study of target population of the survey.Moreover, due to the responses being anonymous, it was not quite possible to filter out and remove such responses from the dataset. Moreover, I feel a lot of people may not take part in the survey thinking it may consume quite a bit of their time. Due to this, there may also be a lot of ignored responses to the survey leading to lesser data.

In order to clean the dataset, I first removed both the variable columns 'Other.Physical.symptoms' and 'Other.Emotional.Symptoms' with data entries where respondents were asked to specify any other physical or emotional symptoms in Question 6 and Question 7 that were not included in the given options to check. This was done as all the responses included 'NA' as a response adding no additional information or support to the study of the survey. Moreover, the data exported from the survey also included a 'Timestamp' column that specified the date and time of each response being submitted. This was also not necessary for our study and hence all these columns were removed used the subset function. The survey data that had to be used was saved in a specific parameter called 'survey_data' to avoid repetitive calling out of the entire dataset. Moreover, in order to refine and work with the data which was an object, I used the as.data.frame function to modify it into a proper data frame with rows and columns. In addition to this, I also added a unique identifier for each data entry as 'index' using the 'rowid_to_column' function. This id is just like serial numbers given to each row of data entry in order to make it distinctive while analyzing.

Important Variables

Important variables here would be considered those that are in some way or the other related to stress levels for university students. The variables and their descriptions are as follows:

1. Level of Stress - This is a linear scale question where respondents have to select one number from 1 to 5 where 1 signifies 'least stressed' and 5 is 'most stressed'. This number is the students' judgment as to what they believe their stress levels are. The numerical value from this question gives us an overall perspective of the intensity of stress that University of Toronto students experience.
2. Hours of Sleep - This is a numerical input question which helps us understand the number of hours each respondent sleeps. This is an integral question as hours of sleep can be correlated to other factors in the survey such as Level of Stress, Level of Healthiness and Course Load, helping us draw valuable interpretations.
3. Level of Healthiness - This is a linear scale question where respondents have to select one number from 1 to 5 where 1 signifies 'least healthy' and 5 is 'most healthy' according to them. This number is the students' judgment as to how healthy their physical and mental body state is on a day to day basis. The numerical value from this question gives us an overall perspective of whether students are doing well in health terms and if they are not, then is this is any way related to their level of stress.
4. Course Load - This multiple choice question gives us an idea of how heavy the academic course load is for students in general for a majority of students. By understanding the course load for students, we may be able to link greater course load with higher levels of stress or lower level of healthiness and many others as hypotheses statements to find insightful correlations.

Numerical Summary

```
##      index           Age           Gender          Course.Load
##  Min.   : 1.0   Min.   :18.00   Length:71          Length:71
##  1st Qu.:18.5   1st Qu.:19.00   Class :character   Class :character
```

```
##  Median :36.0   Median :20.00   Mode  :character   Mode  :character
##  Mean   :36.0   Mean   :20.24
##  3rd Qu.:53.5   3rd Qu.:21.00
##  Max.   :71.0   Max.   :39.00
##  Physical.symptoms  Emotional.or.Behavioural.symptoms Level.of.Healthines
##  Length:71          Length:71                         Min.   :1
##  Class :character   Class :character                  1st Qu.:3
##  Mode  :character   Mode  :character                  Median :3
##                                                       Mean   :3
##                                                       3rd Qu.:3
##                                                       Max.   :4
##  Causes.of.Stress   Level.of.Stress Hours.of.Sleep
##  Length:71          Min.   :1.000   Min.   :4.500
##  Class :character   1st Qu.:3.000   1st Qu.:5.500
##  Mode  :character   Median :4.000   Median :6.000
##                     Mean   :3.676   Mean   :6.176
##                     3rd Qu.:4.000   3rd Qu.:6.500
##                     Max.   :5.000   Max.   :9.000
```

For our summaries, we see that the variable columns of Gender, Course Load, Physical Symptoms, Causes of Stress and Emotional or Behavioral Symptoms are character variables so there are no numerical summaries for them. We see that mean age is 20.24 as most of the students were of age 20. The minimum age among the respondents is 18 years while the maximum age is 39. The respondent of age 39 years probably may be a post graduation student at this university. His/her response becomes an outlier for our study as there is a vast deviation from the age of students' who filled up the survey compared to his age. For the hours of sleep, we see that the minimum hours of sleep a student gets from the respondents is 4.5 hours while the maximum is 9 hours. The average for the hours of sleep for all respondents is 6.176 hours. Lastly, for the linear scale questions for Level of Stress and Level of Healthiness, the average levels for all the respondents is 3.6 and 3 respectively. The minimum levels for both the variables are 1 (least). However, for maximum levels, we see that respondents also experience level 5 for stress which is the highest level, whereas no respondent feels fully healthy as the maximum response for level of healthiness was 4 and not 5.

Graphs

Figure 1: Bar Garph of the Hours of sleep for students at the University of Toronto

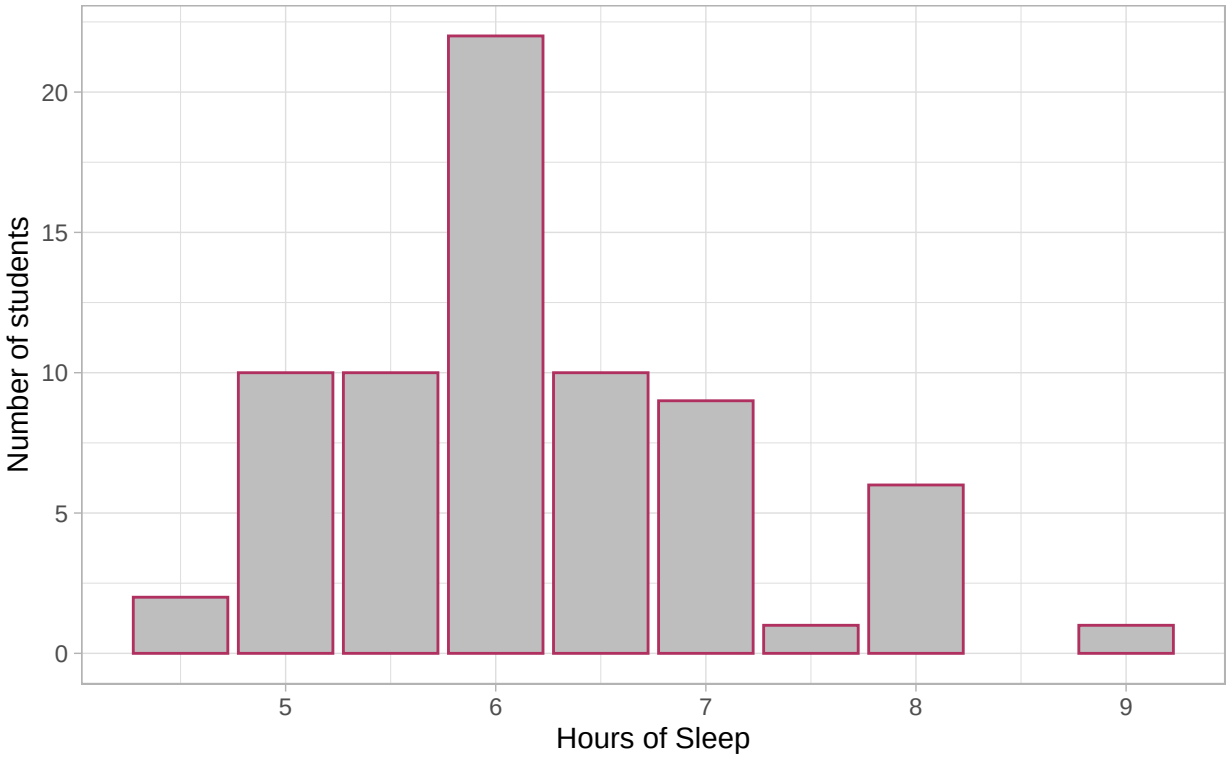
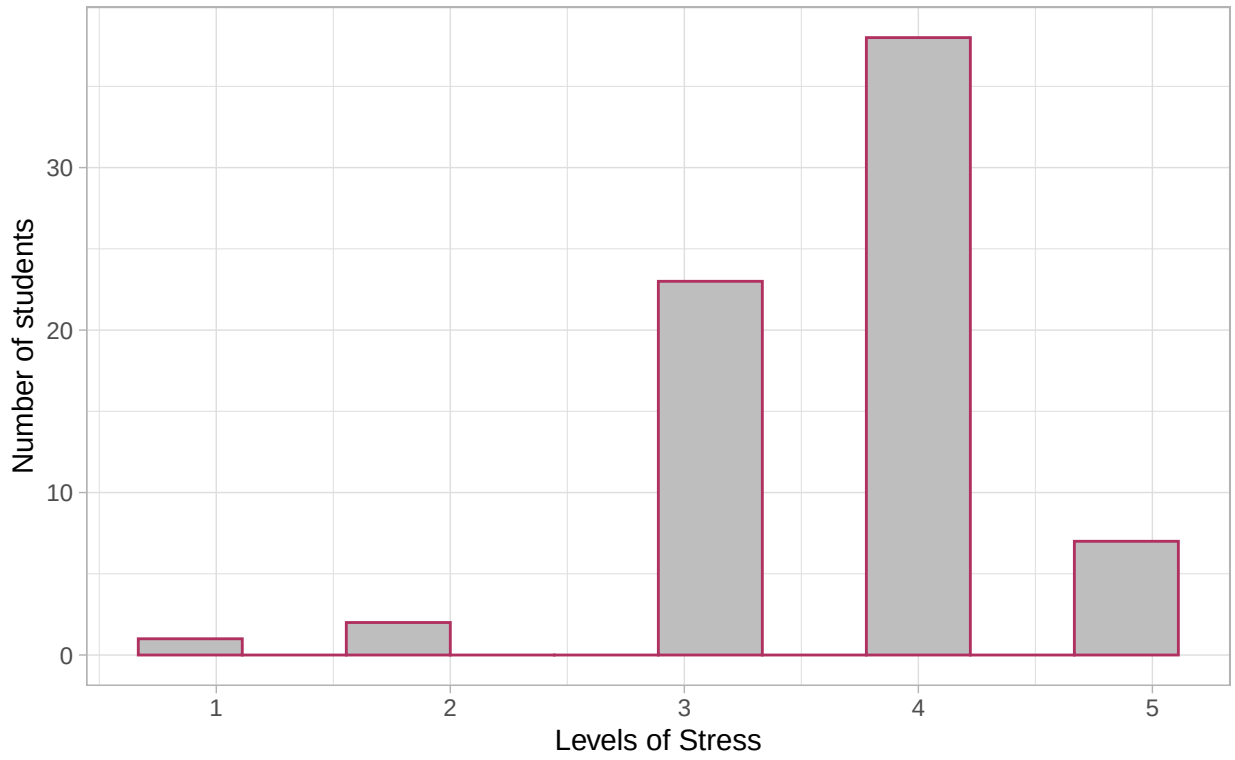


Figure 1: Histogram of the Level of Stress for students at the University of Toronto



Bar Graph for the hours of sleep for students

The x axis states the hours of sleep and the y axis has the number of students for different categories of hours. We see that the maximum number of students (around 22 students) get 6 hours of sleep on average while very few students get 7.5 hours of sleep on a daily basis. There are no students who sleep for 8.5 hours. Moreover, we also see that the graph is 'skewed to the right', signifying a few outliers with higher sleep hours on the extreme. In addition to this, we also observe that the mean for the data set is around 6 hours as well.

Histogram for the hours of sleep for students

The x axis states the different levels of stress and the y axis has the number of students for those levels. We see that the maximum number of students (around 48 students) are stressed on level 4 out of 5 (being the highest), which signifies a very high level of stress for a majority of students. Very few students had selected 1 level (signifying least stressed) implying that most of the students have some or the other intensity of stress in their day to day lives . There are no students being in the middle ground of 2.5 level of stress, suggesting that majority of students have opted for the higher extreme level(from 4 to 5) in the scale. We may also conclude from the graph that the mean level of stress is around level 3.

All analysis for this report was programmed using `R version 4.0.2`.

Methods

Hypothesis Testing

In order to draw some interpretations or conclusions from the survey, we will use Hypothesis Testing. Hypothesis is a supposition made for our population that can be proved or disproved using adequate evidence. The null hypothesis (H_0) is the statement that is positive and assumed to be true and the alternative hypothesis (H_A), is the opposite of the null hypothesis which signifies that if we are not able to collect enough evidence for the null hypothesis to be true, this would suggest that the alternative hypothesis stands true

for the study. For our survey, we have used the variable Hours of Sleep for students and specidied the null hypothesis as 'The mean ($\mu_{student.sleep.hours}$) of the number of hours of sleep for students at The University of Toronto is EQUAL TO 6' and the alternative hypothesis as 'The mean of the number of hours of sleep for students at The University of Toronto is NOT EQUAL TO 6'. The mean (μ) would signify the average of all the values of the hours of sleep for students that we get. In order to find enough evidence for the null hypothesis to be true, we use the p value. P value is how likely you will find observations for your null hypothesis to be true (probability of the null hypothesis to be true). The greater the p value (more than 0.05 at least), the more evidence you have for the null hypothesis being true. Our null and alternative hypotheses are as follows:

$$H_0 : \mu_{student.sleep.hours} = 6$$
$$H_1 : \mu_{student.sleep.hours} \neq 6$$

Confidence Intervals

Confidence interval shows how 'confident' you are to find the value of any parameter(mean,median,variance etc) of study for the population between a range of values calculated by the confidence interval. The higher the confidence value, the more confident you are. A larger sample size also makes the accuracy of our confidence intervals better and a higher confidence interval signifies a wider interval. To calculate confidence interval from our survey, we use the variable Level of Stress. I will invoke a parametric t-test confidence interval to derive the 95% confidence interval (CI) for the mean level of stress for students at the University of Toronto. A 95% confidence interval signifies a chance of being 5% wrong. To calculate the interval, we first find out the mean and standard deviation of the values of Level of stress from our survey and store them into named parameters. Then, we also need to find the error that indicates the uncertainty around the estimate of the mean measurement (how different is the population mean for the level of stress from our sample mean). The standard error is calculated using the t distribution as it helps us to compare the positive values of our sample to the standard mean value. Lastly, the lower value of our interval is generated by subtracting the error found from the mean value of sample and the upper value is found by adding the error to the mean value.

Results

```
One Sample t-test
```

data: survey_data_1$Hours.of.Sleep t = 1.582, df = 70, p-value = 0.1181 alternative hypothesis: true mean is not equal to 6 95 percent confidence interval: 5.954107 6.398006 sample estimates: mean of x 6.176056

lower_conf_interval_stress upper_conf_interval_stress 1 3.499425 3.852687

Results from Hypothesis Testing of mean hours of sleep

From the hypothesis testing using t test, the p value suggests that our null hypothesis, 'The mean ($\mu_{student.sleep.hours}$) of the number of hours of sleep for students at The University of Toronto is EQUAL TO 6'. We see from our summary table that our p-value comes out to be 0.1181 which is quite large and is greater than 0.05 signifying sufficient evidence collected for our null hypothesis. Along with this, we also observe the 95% confidence interval from 5.954 hours of sleep to 6.398 hours of sleep and our sample mean is also falling in this interval. We may say that we are 95% confident that the mean hours of sleep for our population can be from 5.954 hours to 6.398 hours of sleep for students.

Confidence Interval for mean level of stress

From our analysis, we may state that the 95% confidence interval for the level of stress for the students at the University of Toronto lies anywhere from 3.499 to 3.852 level and the mean level of stress for our sample is 3.67, which is also falling in this interval. We may state that we are 95% confident that the mean level of stress for our population can be from 3.499 to 3.852 level out of 5 levels for students.

Conclusion from our Analysis

In conclusion, we may conclude that the true mean for the number of hours of sleep for the student population at the University of Toronto is equal to 6 from the survey data and that we are 95% confident that the mean level of stress of the population of students would lie anywhere from 3.499 to 3.852 level out of 5 levels.

Bibliography

1. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html. (Last Accessed: May 5, 2021)
2. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.
3. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/. (Last Accessed: May 5, 2021)
4. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.
5. *Student Health and Wellbeing At the University of Toronto*, a report on the findings of the National College Health Assessment et. al. (2017)
6. Cairney et al. (2013) *The Elaborated Environmental Stress Hypothesis* Front. Psychol., Developmental Psychology

Appendix

Here is a glimpse of the data set simulated/surveyed:

```
## Rows: 71
## Columns: 10
## $ index                              <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 1~
## $ Age                                <int> 20, 20, 20, 19, 20, 21, 20, 20, 19, ~
## $ Gender                             <chr> "Male", "Female", "Male", "Female", ~
## $ Course.Load                        <chr> "5.0 credits", "6.0 credits", "6.0 c~
## $ Physical.symptoms                  <chr> "Increased or Decreased appetite, Fa~
## $ Emotional.or.Behavioural.symptoms <chr> "Decreased efficiency and effectiven~
## $ Level.of.Healthines                <int> 3, 3, 3, 3, 3, 2, 4, 1, 2, 3, 3, 3, ~
## $ Causes.of.Stress                   <chr> "Grades, Work-Life Balance", "Grades~
## $ Level.of.Stress                    <int> 3, 4, 4, 3, 4, 4, 3, 5, 4, 4, 3, 4, ~
## $ Hours.of.Sleep                     <dbl> 7.0, 5.5, 7.5, 8.0, 5.0, 4.5, 5.0, 5~
```

2. Mean of the hours of sleep for the student sampling population was calculated to be -

```
## [1] 6.176056
```

3. Mean level of stress for the student sampling population was calculated to be -

```
## [1] 3.676056
```