**QMM1001 Case Study 2 [20%]**

**Goodnews Agbadu**

**A00238219**

**Due: December 17th, 2021, at 11:59 PM**

# Introduction

For this case study, personal data was about how I spent my day as an international student in Canada. To collect these data, I used a personalized data sheet to record my daily activities from **September 10th, 2021, to December 8th, 2021**. I assigned a total of nine (9) variables such as Date, hours spent on zoom for virtual learning, hours it takes to sleep each day, how often I left the house, time spent on news, time spent on playing games, how stress I am each day and the count of calls I receive each day. To record my variables, I used a scale of 5-1 to measure how stressed I was daily and for Zoom, Study, Sleep, and House, I used numbers in hours from 0-9 to measure the amount of time spent on each variable. While for Calls, I used a count from 0-7 to estimate how many times I spoke over the phone. Finally, I used “yes” and “No” to measure Game and News’s categorical variable.

Having a glance through my personalized data, I started with recording four (4) quantitative variables. Usually, I feel that my day seems to have ended after my Zoom class, and I want to understand what portion of my day was spent on classes as well as studying. Remembering I had to sleep each day gives me joy because it enables me to relax my mind and be prepared for the day ahead. One thing I notice about myself is that I do not leave my house unnecessarily, which has affected my social life. For the last quantitative variable, I collected data about how many calls I take each day. It is interesting to see if I spend more time daily talking over the phone.

Furthermore, to better understand my daily activities, I collected two nominal variables as well as an ordinal categorical variable. The nominal variables were to know if I listen to the news in order to familiarize myself with the political, economic, and social aspects of Canadian life and whether I play a game each day as I did when I was living in Nigeria. Finally, I am researching how stressed international students are during COVID, I used a categorical, ordinal variable to represent my stress level, and I am particularly interested to understand how stressed I am, maybe due to prolonged zoom classes or less sleeping time.

To answer the question what do you do in a day? from my dataset, I observed from the variables that some activities such as reading, sleeping, taking calls, and stress are all part of my daily life while some such as zoom, games, leaving the house are sometimes not part of what I do daily. This is surprising because I noticed I was stressed all day which may be because of not leaving my house often to catch some fun, and I would say I did not spend the day the way I expected it because I may plan to spend some time outside my house, but I ended up using up all the time for prolonged studies. At the end of this data analysis, I seek to have a better understanding of whether watching the news affects how I spend my day and whether I watch the news more than other people.

For this analysis, I will use the multiplication rules and conditional probability to formulate questions as it relates to watching the news and my level of stress. In addition, I will also identify if those formulated questions are disjoint or non-disjoint and if those questions are dependent or not as it relates to my daily activities. Furthermore, I will deploy the central limit theorem to determine my sample proportion and compare it with the population and the confidence interval to report and interpret findings. Finally, I will use a hypothesis to test proportions and determine whether to reject the null hypothesis or do not reject the null hypothesis.

# Data Analysis

**Part 1**

For this analysis, I am going to answer the question of whether watching the news affects how I spend my day. This was achieved by creating a contingency table for news variables and categorical variable stress from my personalized data set. The categorical variable stress will be important in this analysis because I want to understand whether consuming the news can activate the sympathetic nervous system, which can cause the body to release stress hormones like adrenaline when a crisis such as a new COVID-19 variant, earthquake, or tornado are projected to happen. In addition, because watching the news can have both negative and positive relationships, it is equally important to understand whether watching the news can reduce my stress level. For example, if the government of Canada should announce while I am watching the news an automatic permanent residency for international students in Sudbury, whether such positive news can reduce my level of stress.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Table 1: Contingency Table Comparing watching the news and my stress Level** | | | | | | | |
| **Did I watch the news?** | |  | | **Stress Level for game** | | | | |
| **1** | **2** | | **3** | **4** | **5** | **TOTAL** |
| **No** | | 3.3% | 12.2% | | 13.3% | 5.6% | 10.0% | **44.4%** |
| **Yes** | | 34.4% | 10.0% | | 3.3% | 6.7% | 1.1% | **55.6%** |
| **TOTAL** | | **37.8%** | **22.2%** | | **16.7%** | **12.2%** | **11.1%** | **100%** |

With the use of probability rules, four questions were formulated from the contingency table as shown in **Table 1** to capture how watching the news relates to stress.

The questions are as follows.

1. Using the rule of complement (NOT) to answer the question, what is the probability of not watching the news at all? This is denoted as **P (News! = Yes) = 44.4%,** which means that the chance of not watching the news at all each day is at **44.4%**; this result can be attributed to other factors such as the amount of time I spend in zoom and studying each day. This belief is evident from the contingency table above, which shows that the probability of not watching the news at **44.4%** shows that my stress level 5 is at **10%**. Further analysis shows that the event is disjoint because it is possible for not watching the news at all not equal to yes. These events can occur together.
2. Using an additional rule for probability (OR) to answer the question, what is the probability of being stressed at 4 or being stressed at 5? This is donated as **p (stress = 4 or stress = 5) = 23.3%** means that the probability of being stressed at level 4 or 5 is **23%**. This result could be attributed to the fact that I spend lesser time watching the news when my stress level is at the extreme; perhaps I used the time to sleep to relax the body for the task ahead. Further analysis shows that the event is non-disjoint because it is not possible to be stressed at level 4 and level 5 at the same time. These events cannot occur together.
3. Using a conditional probability (GIVEN) to answer the question, what is the probability of watching the news given my stress level 5? this is denoted as **P (News = YES | Stress = 5) = 10%,** which means that my chance of watching the news when my stress scale is at 5 is **10%.** These findings are not surprising because I would prefer to play a game or make a phone call when I am stressed on a scale of 5. Further analysis shows that the event is disjoint because it is possible to watch the news when my stress level is 5. These events can occur together.
4. Finally, using a multiplication rule for probability (AND) to answer the question, what is the probability that I watch the news and my stress level a 4? this is denoted as **P (News = YES AND Stress = 1) = 34.4%** means that there is a **34.4%** chance of watching the news and my stress level is 1. This simply shows that I watch more news at my lowest stress level; also, I can attribute this result to the fact that I brought a new Samsung television during the black Friday sale. Having a new television has motivated me to watch more news instead of leaving my house or talking over the phone. Further analysis shows that the event is disjoint because it is possible to watch the news and my stress level is 1. These events can occur together.

For this analysis, there are no disjoint events in my contingency table shown above; this is because none of the events in my contingency table is equal to zero. This is not surprising as the sample proportion of watching the news is 55.6%. To relate this to my daily activities, it is possible to watch the news while on the phone talking or to play a game while stressed, and it is equally possible to be stressed while attending zoom calls or studying.

Checking how events depend on others is crucial in this analysis; I used **P (A AND B) = P(A)\*P(B)** to check the independence between news and stress from my contingency table in **table 1**. If **P (A AND B) = P(A)\*P(B)** events are independence else **P (A AND B)! = P(A)\*P(B)** events are dependent. The test was carried out on three probability questions, and the results show that the probability of being stressed at 4 or being stressed at 5 is dependent with **P(A)\*P(B)= 0.01358025,**

when compared to **P (A AND B) = 0.01358025** shows a difference, which suggests that events are dependent on other variables in my personalized data set.This isbecause variables such as zoom class or studying can also lead to changes in the level of stress, I experience each day.

The second test was the probability of watching the news given my stress level 5; this was also dependent with **P(A)\*P(B)** = **0.0617284**. This isbecause they are also other variables in my personalized data that can increase my stress level or reduce it, such as spending more time in zoom call or having more sleeping time. In addition, looking at my personalized data shows that the chances of watching the news at stress level 5 is denoted as **P (A AND B) = 0.01111111 or 1.1%,** which supports why the events are dependent.

The final test was carried out on the probability that I watch the news and my stress level a 4, which shows events to be dependent on other factors with P**(A)\*P(B)** = **0.2098765** when compared to **P (A AND B) = 0.3444444** shows a difference which suggests that events are dependent on other variables in my personalized data set. For example, would I rather watch the news and my stress level is 4, or will I sleep? The chance of watching the news depends on whether I am not feeling asleep, then I may decide to watch the news.

**Part 2**

In this section of the analysis, I will be looking at whether I watch the news more than **59%** of Canadians daily. To achieve this, I used the news variable from my personalized data set collected for 90 days, and the results show that the number of days I watched the news was 50, which is **55.6%** of my sample proportion, while the number of days I did not watch the news was 40, which is **44.4%** of my sample proportion. Comparing my sample proportion with that of the Canadians shows that I watch the news less than the amount of the general Canadian population. This finding was not surprising to me because of other activities that make up my day, such as hours spent on zoom calls, time spent on sleeping and studying. In addition, it is worth noting that I spent more time watching the news during this winter period because, in extreme weather conditions, there is a possibility of not leaving my house, thereby increasing the time spent watching the news.

The mean of the normal model is the proportion of the Canadian population that access the news daily, which is 0.59 with the standard deviation of the normal model **0.07011895**. Furthermore, I calculated the probability of getting a value less than or equal to my sample proportion using **0.59** as the population proportion, which showed that the probability of getting a value less than or equal to the sample proportion of Canadians that watched the news daily is **31.1%**. In addition, I calculated a **95%** confidence interval for the proportion of days that I watched the news, which shows that We are **95%** confident that the true proportion of the time I watched the news is between **45%** and **66%**. It seemed that I watch the news more than **50%** of the time because my confidence interval is between **45%** and **66%**, while my sample proportion of the days I watched the news is **55.6%.**

A test of the hypothesis that I watch the news more or less than the general population of Canadians (that watch the news **59%** of the time) was conducted using both the null and alternate hypothesis as **H0: p = 0.59 and HA: p < 0.59** with an alpha value of **0.05**, shows a p-value of **0.2532207 or 25.3%** which means that p > level of significance of **0.05 or 5%.** The result means do not reject the null hypothesis. Findings show that there is no significant evidence that I watch the news less than the general population of Canadians.

# Conclusion

After analyzing the data collected from **September 10th to December 8th, 2021**, I can better understand how I spend my days.

In this analysis, I used the probability rules such as NOT, AND, OR and GIVEN, to answer the question of whether watching the news affects how I spend my day. The results show that watching the does not necessarily affect how I spend my day because they are other variables in my personalized data that have more impact on how I spend my day. Furthermore, it was surprising to know that watching the news may increase or decrease my stress level depending on if the news is good news or bad news.

Secondly, I used confidence intervals and a test of hypothesis to ascertain whether I watch the news more than other people. The confidence intervals showed that I watch the news more than **50%** of the time because my confidence interval is between **45%** and **66%**, while my sample proportion of the days I watched the news is **55.6%.** In addition, the test of hypothesis showed a p-value (p > level of significance, do not reject the null hypothesis), which means that there is no significant evidence that I watch the news less than the general population of Canadians.

By completing this report, I learned that I spend more time watching the news at stress levels 1 and 2 with the proportion **34.4%** and **10%**, respectively, while the chance of watching the news at stress level 5 is **1.1%.** In addition, I learned about other factors in my personalized data that have more impact on how I spend my day as a student. I have made some personal recommendations to improve how I spend my day for next semester:

1. I will reduce the number of hours I spend watching the news by **20%**
2. I will apportion more time towards my studying time as I will have more workload next semester.
3. I will improve my social and networking skills.