Introduction

I have collected personalized data about myself on how I spend my days. I recorded data from January 14th, 2022, to April 13th, 2022, on these variables: number of hours I spent on zoom class, number of hours I spent studying, number of hours I spent sleeping, number of times I left the house, number of emails I received on that day, did I watch news that day, which fruit I ate that day and how did my dinner taste that day.

The first five variables listed above are quantitative. I am currently a student, so a lot of my time is spent on online zoom classes and studying. These variables are recorded by time in hours. The number of times I left the house is usually for my job and is counted as the number of times I go out of the house. The last quantitative variable is the number of emails I receive every day. Since every information is now shared online because of the pandemic, I had to keep checking my email to keep up with my assignments, class schedules, and job schedule so I tracked the number of emails I receive every day.

The last three variables that I collected are categorical. Did I watch the news that day and which fruit did I eat that day? Both these variables are nominal. I was interested in tracking whether I watch the news or not and which fruit I consume the most. I also recorded how my dinner tastes each day on a scale of 1 to 5, '1' tasting very bad and '5' tasting very good. Since I started living on my own and I am new to cooking I wanted to track if my cooking has improved. This is an ordinal categorical variable. Two of the main categorical variables that I used in this Analysis are the news and the fruit variables. The news variables are recorded using 2 parameters 'Yes' and 'No', if I watch the news that day, it is recorded as 'Yes' else it is recorded as 'No'. The fruit variable is recorded using the name of the fruits that I ate that day like 'Apple', Lemon, 'Banana', etc.

I am doing this analysis to figure out if watching the news affects how I spent my day and do I watch the news more than other people. To answer these questions, I will use probability analysis and sample distribution analysis. For probability analysis, I will use different probability rules such as the rule of compliment, addition rule, conditional probability rule, and multiplication rule. And for sample distribution analysis, I will use the central limit theorem, confidence interval, and hypothesis testing.

Data Analysis

Probability Analysis

Contingency table:

For this analysis, I had to create a contingency table to compare the news variable and the variable that may have a relationship with me watching the news. But none of the categorical variables seem to have a strong relationship with the news variables. So, I chose the fruit variable for this analysis because it would not matter since both the categorical variables have a weak relationship with the news variable. Table 1 shows the contingency table that I got in R for this analysis.

Table 1: Contingency Table comparing the fruit variable and the news variable

		News	
		No	**Yes**
Fruit	**Apple**	38.89%	5.56%
	Banana	27.78%	4.44%
	Grapes	5.56%	2.22%
	Lemon	5.56%	0%
	Orange	5.56%	1.11%
	Pineapple	2.22%	1.11%

Probability Rule Questions:

To test the relationship between the news and fruit variables, I formulated 4 questions using probability rules. I used the four rules, rule of compliment, addition rule for probability, conditional probability rule, and multiplication rule for probability.

The first question was "What is the probability that I do not watch the news?". For this, I calculated what was the probability of me watching the news. Then using the rule of compliment, I calculated the probability that I do not watch the news which was equal to 0.8556. This probability calculation shows that I often do not engage in news watching. The calculation is shown here,

Let, A = I watch the news, P(A) = 0.1444 (This can be calculated using the Table 1)

$P(A^c) = 1 – P(A) = 1 – 0.1444$

$P(A^c) = 0.8556$ □ *(Equation 1)*

Where, $P(A)$ = Probability of me watching the news,

$P(A^c)$ = Probability of me not watching the news

The second question was “What is the probability that I watch the news or eat an apple?”. For this evaluation, I used the non-disjoint addition rule for probability. Since there is a probability that both events occur at once I had to use the non-disjoint addition rule for probability. The probability that I watch the news or eat an apple is 0.5332. Since the probability of me watching the news is very low, the probability of me watching the news or eating any fruit would give a low probability, which is the case for this probability as well. The calculations are shown here,

Let, A = I watch the news, B = I ate an apple

P (A) = 0.1444, P (B) = 0.4444, P (A and B) = 0.0556 (From Equation 4)

P (A or B) = P (A) + P (B) – P (A and B) = 0.1444 + 0.4444 – 0.0556

P (A or B) = 0.5332 □ ***(****Equation 2)*

Where, P (A) = Probability of me watching the news,

P (B) = Probability of me eating an apple,

P (A and B) = Probability of me eating an apple and watching the news on the same day,

P (A or B) = Probability of me eating an apple or watching the news on the same day

The third question was “What is the probability of me eating an apple given I watched the news?”. For this calculation, I used the conditional probability rule. The probability of me eating an apple given that I watched the news is 0.3846. This calculation shows that if I have watched the news on a certain day, there is a 38.46% chance that I eat an apple on the same day. The calculation for the conditional probability is shown here, *Let, A = I watch the news, B = I ate an apple*

P (A) = 0.144444444, P (A and B) = 0.055555556 (From Equation 4)

P (B | A) = P (A and B) / P(A) = 0.0555555556 / 0.144444444

P (B | A) = 0.3846 *□ (Equation 3)*

Where, P (A) = Probability of me watching the news,

P (A and B) = Probability of me eating an apple and watching the news on the same day,

P (B | A) = Probability of me eating an apple given that I watched the news that day

The last question was "What is the probability of me eating an apple and watching the news?". To calculate this probability, I used the rule of multiplication for dependent events which gave the probability of me watching the news and eating an apple equal to 0.05555. The calculation is shown here,

Let, A = I watch the news, B = I ate an apple

P (A) = 0.1444, P (B | A) = 0.3846 (From Equation 3)

*P (A and B) = P (A) * P (B | A) = 0.1444 * 0.3846*

P (A and B) = 0.05555 *□ (Equation 4)*

Where, P (A) = Probability of me watching the news,

P (B | A) = Probability of me eating an apple given that I watched the news that day,

P (A and B) = Probability of me eating an apple and watching the news on the same day

This section calculated the probability of me watching the news with the fruit variable using different probability rules. And the calculations show that there is not a strong relationship between these two categorical variables. Since I would spend my time doing something else like studying or sleeping rather than watching the news the probability of me watching the news is very low. And since I did not focus on collecting my data based on how watching news affects my daily schedule, there are no categorical variables that have a very strong relationship with the news variable.

Disjoint Event:

During the calculation of the probability of the categorical variables, I found that me watching the news and eating a lemon are 2 disjoint events. This means that if I watch the news, I will not eat lemon on the same day and vice versa. The contingency table shows that there is a 'zero' occurrence when both events occur. This means *P (News and Lemon) = 0*. Since there is very little probability for both events to occur, the probability of me watching the news is 14.14% and the probability of me eating the lemon is 5.56%, I find these two events being disjointing just a coincidence.

Independent Events:

For the analysis of the contingency table, I tried to find out if my eating an apple and watching the news are independent events or not. The calculation for this is shown below,

Let, A = I watch the news, B = I ate an apple

P (A) = 0.1444, P (B) = 0.4444

P (A and B) = 0.05555 (From Equation 4)

*P (A) * P (B) = 0.1444 * 0.4444 = 0.0641*

*Here P (A and B) != P (A) * P (B) means not independent events.*

The calculation shows that P (A and B) and P (A) * P (B) are not equal, which means me watching the news and me eating an apple is not an independent event (i.e., They are dependent on each other). Even though I watch the news rarely, I eat a fruit every day so the probability of me watching the news and eating any type of fruit has a high chance of occurring on the same day. So, when they occur on the same day, they become dependent events.

Sample Distribution Analysis

Population Proportion:

According to Statista reports 59% of the general Canadian population access the news daily. And according to the data I collected, I only watch the news 14.44% of the time, and 85.56% of the time I don't watch the news. This information shows that I watch the news very less compared to the general Canadian population. This may be because I only watch the news from one source 'Last week tonight with John Oliver' and it only airs once a week. All the other sources of news I get are from Twitter where I can read the news written in very few words which are effective and quick.

Mean and Standard Deviation:

According to Statista reports 59% of the general Canadian population access the news daily. And according to the central limit theorem for a normal model, the mean of the sample distribution (p-hat) with categorical data will be p, which is equal to 0.59 in this case. And the standard deviation is 0.05184. The calculation for standard deviation is shown here,

p (mean) = 0.59

n (Sample Size) = 90

*SD (Standard Deviation) = (p * (1 – p) / n) ^ (1 / 2) = (0.59 * 0.41 / 90) (1 / 2) = (0.002687) ^ (1/2)*

SD = 0.05184 □ *(Equation 5)*

Sample Distribution:

For this calculation, I used the pnorm function in RStudio to calculate the probability of getting a value less than or equal to my sample population. And the value is equal to 4.19e-18 (4.19×10^{-18}). This means that if you randomly sample the Canadian population, the chance of getting a population proportion less than or equal to 14.44% (my sample proportion) is equal to 4.19×10^{-18}. The calculation for the is shown here,

My sample proportion (p.hat) = 0.1444

Population proportion (p) = 0.59

Standard Deviation = 0.05184 (from equation 5)

z-score = (p.hat – p) / SD = (0.1444 – 0.59) / 0.05184 = -8.59 > 3

z-score = -8.59 < -3

P (p.hat <= 0.14444) = 1 – 0.9999 (closest value) = 0.0001 = 0.01%

The calculation shows that the z-score is way below -3 which makes my sample distribution an outlier. This means that it is very unlikely that the population proportion of the Canadian population is less than or equal to my sample proportion.

Confidence Interval:

When the confidence interval for the days I watch the news is at 95%, the proportion of the days I watch the news is between the intervals of 7.18% to 21.71%. This value is calculated using the qnorm function in RStudio. ‘The probability of me watching the news is between 7.18% to 21.71%’, this statement has a 95% probability of being true. The calculation for the confidence interval is shown here,

p.hat = 0.1444, n = 90

q.hat = 1 – 0.1444 = 0.8556

*Standard Error (SE) = sqrt (p.hat * q.hat / n) = sqrt(0.1444 * 0.8556 / 90) = 0.0371*

Z = 1.96 (value for 95% From Normal Distribution Table)*

Lower-bound = p.hat – Z x SE = 0.1444 – 1.96 x 0.0371 = 0.0718 = 7.18%*

Upper-bound = p.hat + Z x SE = 0.1444 + 1.96 x 0.0371 = 0.2171 = 21.71%*

Confidence Interval = (7.18%, 21.71%)

From the calculation of the confidence interval, I watch the news less than 50% of the time. I can infer this because the upper bound of the interval is 21.71%.

Hypothesis Testing:

I used hypothesis testing to investigate if I watch the news more or less than the general population of Canadians. Statista reports that 59% of the Canadian population watch the news daily. And my sample proportion shows that I watch the news only 14.44% of the time. So, I cannot be watching the news more than the Canadian population. So, instead, I used hypothesis testing to investigate if I watch the news less than the Canadian population. For this analysis the hypotheses are,

Null Hypothesis (H_0): $p = 0.59$

Alternative Hypothesis (H_1): $p < 0.59$

I used the pnorm function in RStudio to calculate the p-value for these hypotheses, which is equal to 4.19e-18 (4.19×10^{-18}). This p-value is very less than the level of significance which is 0.01. So, the null hypothesis is rejected, and the alternative hypothesis is true. That means I watch the news less than the general Canadian population. The calculation for the p-value is shown here,

p.hat = 0.1444, p = 0.59, SD = 0.05184 (from equation 5)

z = (p.hat – p) / SD = (0.1444 – 0.59) / 0.05184 = -8.5957 < -3

p-value = 0.00000 = 4.19×10^{-18} (approximately equals to zero)

Conclusion

This analysis of my data set recorded from January 14th, 2022, to April 13th, 2022, has given me a lot of insights to figure out if watching the news affects how I spent my day and do I watch the news more than other people.

The probability analysis shows that the probability of me watching the news on a certain day is only 14.44% which is very low compared to the Canadian general population (59%). I dig deeper to figure out whether my watching the news affects my daily activities. From the analysis, I can conclude that I do not usually watch the news in a way that interferes with my daily activities. I watch the news usually when I have free time and it does not interfere with my daily schedule. So, to answer the question of whether watching the news affects how I spent my day, no it does not affect how I spend my day.

The sample distribution analysis using the central limit theorem shows that compared to the general population of Canada the proportion in which I watch the news is considered an outlier. At a 95% confidence interval, the proportion of the time that I will watch the news is between 7.18% and 21.71% which is way below the 59% mark. And finally, the hypothesis testing proved that I watch the news very less compared to the general Canadian population at a 0.01 level of significance. So, the sample distribution analysis confirmed that I watch less news compared to the general Canadian population.

This report has shown me that I do not engage myself with news watching like the rest of the Canadian population. Instead, I prefer to get news from other sources like Twitter where the information is written short and is faster. Since I don’t often watch the news, it does not interfere with my daily schedule.