**MGQ 301: Group 5 Project 2 (Option B)**

**Members**: Sidhant Mohanty, Shrirang Dabir, Alexander McLoughin, Afreen Tanisha, Alperen Bay,

**Introduction:**

The study in the article shows the daily caffeine intake of students participating. It indicates the following:

| **Part 1: Daily Intake Frequency of Caffeine by Source(2020)** | |
| --- | --- |
| **Variable \*** | **Participants, n = 727** |
| Intake frequency of regular coffee |  |
| None | 174 (23.92%) |
| 1–2 Units | 439 (60.39%) |
| 3–4 Units | 94 (12.93%) |
| ≥5 Units | 20 (2.76%) |

**Claim Source:**<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7230284/>

We chose to survey college students regarding their caffeine consumption. Here is the data:

| Observed Data | |
| --- | --- |
| Caffeine Consumption | College students |
| 1-2 Cups | 44 |
| 3-4 Cups | 34 |
| >=5 Cups | 17 |
| None | 17 |
| Total | 112 |

**So our null and alternative hypothesis would be:**

* **Ho: pNone = .2392, p1-2 Units=.6039, p3-4 Units=.1293, p>=5 units=.0276**
* **Ha: at least one of the claimed probabilities is different**

We chose this research article because it is something all the group members could collectively relate to. College students rely on caffeine a lot to get through the day, it provides the essential burst of energy and clears our mind to start the day. Hence, this made us wonder how much caffeine college students consume in 2021. We chose to sample a population of college students to gather our own data and make conclusions in comparison to the claim source.

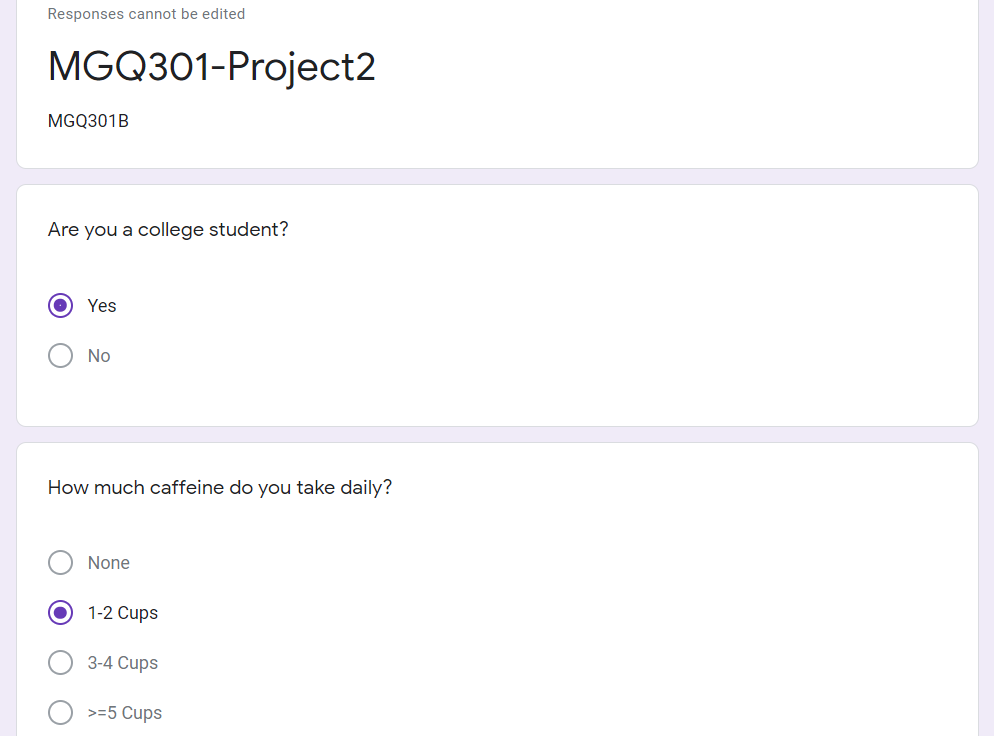
**Data collection & Sampling Procedure**:

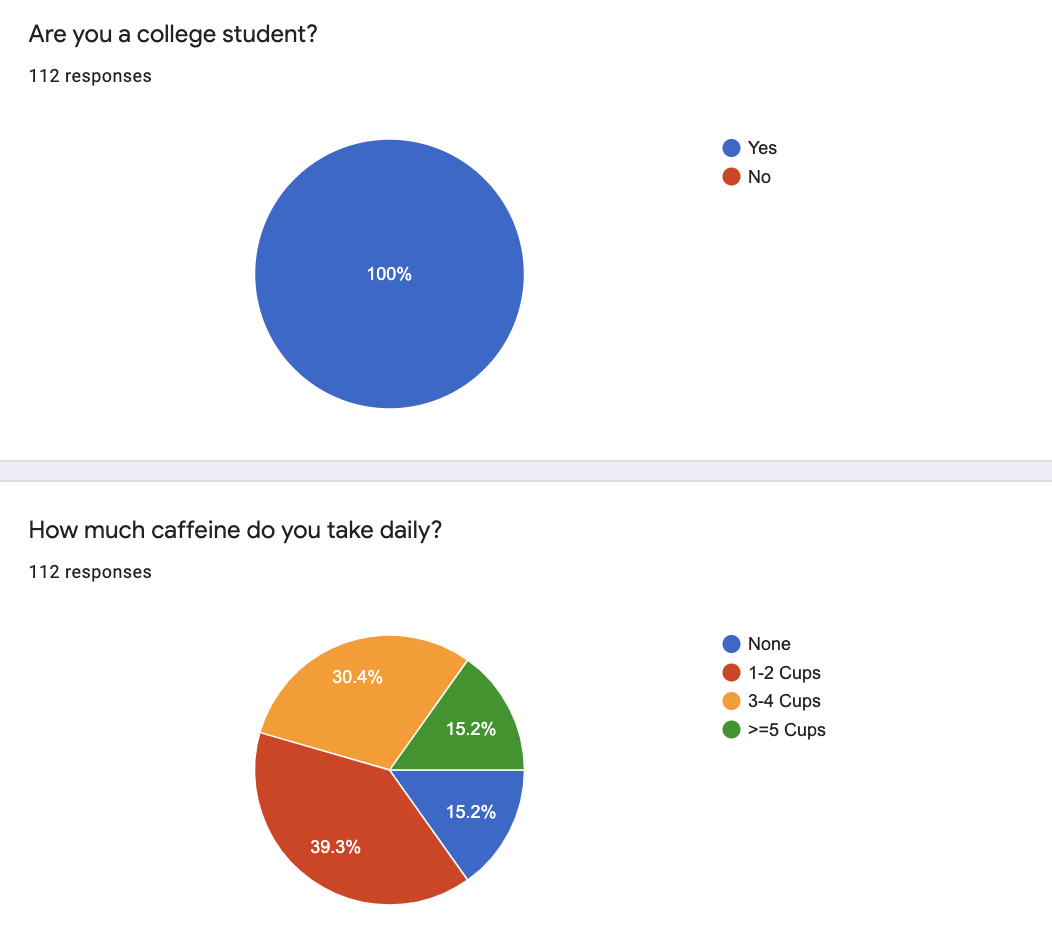
We chose the **Convenience Sampling** method. The source of our data came from surveying students at University at Buffalo’s Silverman Library. We walked around and explained to students our project and had them complete our survey. We believed this would be a great source of information because students tend to use some sort of caffeine on a regular basis. Our random variable was how much caffeine was consumed on a daily basis.

​​

**Survey Link:**

<https://docs.google.com/forms/d/e/1FAIpQLSd-UQHG8ysdTyO2BbGpM8kH1uqd5A7YdddVyCE6hreUg4rsIA/viewform?usp=sf_link>





**The descriptive statistics:**

| Descriptive Statistics | |
| --- | --- |
| Degree Of Freedom | 3 |
| Alpha | 0.05 |
| P value | 7.6563E-23 |
| N (Sample Size) | 112 |
| Test Statistic (Chi-Square) | 106.0790538 |
| Critical Value) | 7.814727903 |

**Hypothesis Testing Procedure:**

We chose our null hypothesis to be- “The probability of consuming zero unit of caffeine is 0.2392, the probability of consuming 1-2 Units of caffeine is 0.6039, the probability of consuming 3-4 units of caffeine is 0.1293, probability of consuming greater than equal to 5 units of caffeine is 0.0276”. And our alternative hypothesis is that “At Least one of our claimed probabilities is different”. The reason we chose this null or alternative hypothesis is because college students depend heavily on coffee and we wanted to test out daily consumption of coffee by college students in 2021. We are using the chi-square test to evaluate our claim. The chi-square test allows us to test the significance of the population variance. The chi square test is used to determine the difference between the observed and expected frequencies. The ability to test the variance with the mean allows for a more complete picture of what is happening in a distribution. We chose the alpha level to be 0.05. We are performing a right-tailed test.

**Hypothesis Testing Results:**

The **degree of freedom** for our data is **3** and we chose our **alpha level** to be **0.05.**

The **p-value** we calculated for the right-tailed test by using **CHISQ.TEST** is **​​7.653E-23 (Less than 0.05)**

**Test statistic** we calculated by using **CHISQ.INV.RT** is **106.0791**

**The critical value** we calculated by using alpha level and degree of freedom in **CHISQ.INV.RT** is **7.81473**.

| Observed Data | |
| --- | --- |
| Caffeine Consumption | College students |
| 1-2 Cups | 44 |
| 3-4 Cups | 34 |
| >=5 Cups | 17 |
| None | 17 |
| Total | 112 |

| Count of Data by percentage(observed) |  |
| --- | --- |
| 1-2 Cups | 39.3 |
| 3-4 Cups | 30.4 |
| >=5 Cups | 15.2 |
| None | 15.2 |
| Total | 100.0 |

| Caffeine intake of college students(expected) |  |
| --- | --- |
| 1-2 Cups | 60.3900% |
| 3-4 Cups | 12.9300% |
| >=5 Cups | 2.760% |
| None | 23.920% |
| Total | 100% |

| Expected Data | |
| --- | --- |
| Caffeine Consumption | College students |
| 1-2 Cups | 68 |
| 3-4 Cups | 14 |
| >=5 Cups | 3 |
| None | 27 |
| Total | 112 |

| Caffeine Consumption | Observed(fo) | Expected(fe) |
| --- | --- | --- |
| 1-2 Cups | 44 | 68 |
| 3-4 Cups | 34 | 14 |
| >=5 Cups | 17 | 3 |
| None | 17 | 27 |
| Total | 112 | 112 |

**Calculations:**

| fo-fe | (fo-fe)^2 | (fo-fe)^2/fe |
| --- | --- | --- |
| -24 | 576 | 8.470588235 |
| 20 | 400 | 28.57142857 |
| 14 | 196 | 65.33333333 |
| -10 | 100 | 3.703703704 |
| 0 |  | **106.0790538** |
|  |  | Test Statistic |

​​

**Descriptive Statistics:**

| **Descriptive Statistics** | |
| --- | --- |
| **Degree Of Freedom** | **3** |
| **Alpha** | **0.05** |
| **P value** | **7.6563E-23** |
| **N (Sample Size)** | **112** |
| **Test Statistic (Chi-Square)** | **106.0790538** |
| **Critical Value)** | **7.814727903** |

A p-value is the probability of obtaining a chi-square as large or larger than that in the current experiment. Since p-value is less than alpha we reject the null hypothesis and support the alternative hypothesis. Hence, the data provides sufficient evidence to claim that at least one of the claimed probabilities is different

**Conclusion:**

We have gathered enough data and run a hypothesis test in which we rejected the null hypothesis and supported the alternative. We tested the data using a Chi-Square test because we were testing a sample of the population. As we calculated the data for the hypothesis test we found that at least one of the claimed probabilities is different. The data provide sufficient evidence to claim that the actual distribution differs from what we expected. This led to a possibility of a type I error because we have rejected the null hypothesis because the p-value for Chi-Square test is less than alpha. Consequently, a type 1 error will **bring in a false positive**. This means that you will wrongfully assume that your hypothesis testing has worked even though it hasn't. In real life situations, this could potentially mean losing possible sales due to a faulty assumption caused by the test.