جامعة مصر للمعلوماتية
EGYPT UNIVERSITY OF INFORMATICS

Egypt University of Informatics
Computer and Information Systems
Data Analysis Course

# Relationship between Coffee Consumption among students and their CGPA

Submitted by: Nada Ashraf Saleh

9/3/2024

## Introduction

- Coffee is the universal companion of students everywhere. However, beyond its role as a morning routine or late-night study aid, there lies a question: does coffee consumption impact academic performance? This report explores the relationship between students' coffee intake and their cumulative GPA. Through a synthesis of research and analysis, we aim to uncover whether a meaningful correlation exists, shedding light on the potential implications for students' academic success. Join us as we delve into this intriguing connection, unravelling coffee's relationship with students' academic performance.

## Research Question

- Does coffee consumption affect students' cumulative GPA and their overall academic performance?

## Hypothesis

- We hypothesize that there is a positive correlation between coffee consumption among students and their cumulative GPA, suggesting that higher levels of coffee intake may be associated with higher academic performance.

## Population of Interest:

- Students enrolled in high schools, colleges, or universities.

## Sampling Method:

- My sampling method is "Simple Random Sampling Method".
- Online survey was sent on a group of college students.
- Advantages of this sampling method:
    - Ease of Implementation:
        - Simple random sampling is straightforward to implement, requiring minimal logistical effort.
        - It involves assigning each member of the population an equal chance of being selected.
    - Unbiased Representation:
        - Since every member of the population has an equal chance of being chosen, simple random sampling helps avoid potential biases that could arise from stratification or other sampling methods.
        - This ensures that the sample is representative of the entire population.

## Bias Identification:

- **Focus on Factual Information:**
    - Rather than asking opinion-based questions, I focused on gathering information, such as the frequency of coffee consumption and GPA scores, to minimize the influence of personal biases.

- **Confidentiality Assurance:**
    - I emphasized the confidentiality and anonymity of respondents' answers, reassuring them that their responses would be kept strictly confidential and used only for research purposes. This helped create a safe and non-judgmental environment for respondents to provide honest feedback.

## Survey Questions:

[On average, how many days per week do you consume coffee?]

[How many cups of coffee do you typically consume per day?]

[What is your CGPA?]

[During your exam periods, does your coffee consumption:]

[Do you believe there is a relationship between your coffee consumption and your academic performance?]

Online survey link: https://forms.gle/jRhJjEYNmfrSPX5N9

Number of samples collected: 34

## Analysis:

Here are some descriptive statistics such as mean, median, mode, and standard deviation:

| | coffee per day | CGPA |
|---|---|---|
| count | 34.000000 | 34.000000 |
| mean | 1.117647 | 2.808235 |
| std | 1.007989 | 0.606133 |
| min | 0.000000 | 1.700000 |
| 25% | 0.250000 | 2.425000 |
| 50% | 1.000000 | 2.840000 |
| 75% | 1.000000 | 3.075000 |
| max | 4.000000 | 4.000000 |

We will also create visual representations of the data using scatter plot and line of best fit to help identify any correlations.

| **Error! Filename not specified.** |
| --- |
| This scatter plot indicates that there is a positive yet weak correlation between coffee consumption per day and students' cumulative GPA.<br>Pearson's r correlation is 0.249.<br>I put my response variable (CGPA) on the y-axis and my explanatory variable (Cups of coffee per Day) on the x-axis. |
| **Error! Filename not specified.** |
| This bar char represents a one categorical variable which is the change in the consumption of coffee in students during exam seasons. As represented, during exam season, the consumption increases significantly. |
| **Error! Filename not specified.** |
| **Error! Filename not specified.** |
| This bar chart represents the categorical variable of days per week consumption of coffee. It is considered categorical because the categories (1-2, 3-4, 5-6, and every day) as distinct groups without any order or magnitude. Each category represents a discrete group, and there is no natural order or numerical relationship between the categories. |
| **Error! Filename not specified.** |
| Using the Five Number Summary that included the 1st quartile and 3rd quartile. I was able to conclude the box and whiskers plot by getting the upper and lower fences that highlights the fence where the outliers may lie. |

## Conclusion

- In summary, a Pearson correlation coefficient of (Pearson's r) = 0.249 suggests a weak positive linear relationship between coffee consumption per day and CGPA in the data.

## Any potential issues

- **Response Bias**: Respondents may provide biased responses based on their perceptions of the study's objectives or their desire to conform to societal norms. For

example, students may underreport their coffee consumption if they perceive it as unhealthy or overestimate their academic performance to align with societal expectations.

- **Confounding Variables**: There may be confounding variables that influence both coffee consumption per day and CGPA, leading to a spurious correlation. For example, factors such as sleep quality, study habits, or overall lifestyle could affect both variables independently.