CAFFEINE CONSUMPTION HABITS of METU STUDENTS



İSMAİL EREN ÇAKIR

NURAY TAGHİYEVA

SİNEM SARICA

SU ALADAĞ

**TABLE OF CONTENT *EN BAŞA AL***

1. ABSTRACT ………………………………………….................…….…………………4

2. INTRODUCTION………………………………………...……..................................…4

3. LITERATURE………………………………………..................…………..….5

4. AIM OF RESEARCH…………………………………………………

5. SURVEY METHODOLOGY……………………………………………........................…….5

5.1 Survey Design

5.1.1 Sample Design

5.1.2 Data Collection

5.2 Questionnaire Design & Construction

5.3 Methods of Analysis

5.3.1 Descriptive Statistics

5.3.2 Statistical Tests

6. DATA ANALYSIS, FINDINGS AND DISCUSSION………......…………………...............10

5.1 Do male students behave more aware of climate change at METU?

5.2 Is there any relationship between parents' education level and students' behaviors

about climate change?

5.3 Does faculty is an essential factor in impress climate change awareness behavior?

5.4 How does social media using time per day affect students' awareness level and

awareness of government policy about climate change?

5.5 Does the period when the students learn about climate change affect their awareness

level and behaviors about climate change?

6. CONCLUSION AND RECOMMENDATIONS……………………………….............…….18

7. REFERENCES……………………………………………………………………..........…....19

8. APPENDICES……………………………………………………………………..........….....19

**1. Abstract**

Caffeine, a widely consumed stimulant, plays a significant role in enhancing physical and mental performance, especially among high-stress groups such as students. This research delves into caffeine consumption patterns through a structured survey, investigating its sources, cost, consumption levels, and perceived effects on academic performance and sleep quality. Key variables such as self-reported CGPA, daily caffeine intake, and primary caffeine sources are analyzed to explore the relationship between caffeine habits and sleep struggles, providing nuanced insights into how caffeine influences daily functioning and productivity of METU students. The findings aim to bridge gaps in understanding caffeine's dual role as a productivity enhancer and a potential disruptor of health and well-being. Simple Random Sampling and convenience sampling methods are used throughout the survey. FINDINGS…

Key words: Caffeine, METU, …

**2. Introduction**

Caffeine (C8H10N4O2), a natural stimulant found in coffee, tea, chocolate, and energy drinks, is one of the most widely consumed psychoactive substances universal. Its effects range from enhanced alertness and cognitive function to potential negative impacts such as sleep disruption and dependency. Understanding caffeine consumption patterns is critical in addressing its influence on health and performance, particularly among populations like students, who often rely on caffeine to manage demanding schedules. This study investigates the caffeine consumption habits of METU students, diverse sample group through a structured survey. Key areas of focus include the primary sources of caffeine, daily consumption levels, and the perceived effects on sleep and academic performance. Additionally, the research explores demographic factors, such as self-reported CGPA, to contextualize consumption behaviors.Using statistical methods such as Multiple Linear Regression, Two-Sample t-Test, ANOVA, this project seeks to identify significant predictors of caffeine-related outcomes, such as sleep struggles, academic success. The findings aim to inform both academic discourse and practical strategies for balancing caffeine use with overall wellness.

**3. Literature**

Caffeine consumption is a deeply ingrained part of modern life, recognized for its stimulating effects and its role in enhancing productivity. As one of the most widely used psychoactive substances globally, it is particularly popular among university students.The primary cause for its use is basically to increase alertness, decrease fatigue, and improve focus during periods of academic stres (McLellan et al., 2016). The most significant source of caffeine among all sources is coffee, followed by tea and energy drinks (Clark & Landolt, 2017). Studies indicate that moderate caffeine intake up to 400 mg per day, approximately four cups of brewed coffee is generally safe for healthy adults (FDA, 2021). However, exceeding this threshold can have negative side effects such as insomnia, anxiety, headaches, and heart palpitations (Juliano & Griffiths, 2004).

Research highlights the dual nature of caffeine's impact on academic performance. Moderate intake has been linked to cognitive benefits, including enhanced concentration, faster reaction times, and improved memory (Smith et al., 2020). However, overconsumption can negatively affect sleep quality, leading to decreased academic performance in the longer term (Clark & Landolt, 2017 Surveys often indicate that caffeine consumption varies by gender and academic year, with senior students frequently reporting higher intake levels due to greater academic pressures (Lopez et al., 2017). Other factors that contribute to caffeine intake in students are social habits and relief from stress (Lara, 2014). Although caffeine can be a useful tool for managing academic pressures, understanding the potential health risks and maintaining a balanced intake is crucial for long-term well-being and academic success.

**4. Aim of research**

Caffeine consumption is a widely prevalent behavior, influencing thousands of METU students. It plays a significant role in shaping individuals' daily routines, yet its impact on various aspects of physical well-being and academic success remains a subject of concern. The primary purpose of this research is to investigate the preferred sources of caffeine, consumption habits, assess their effects on health and academic success, and analyze how demographic factors influence caffeine consumption patterns. Through statistical analysis and measurable outcomes, this research aims to investigate the effects of caffeine consumption and how consumption habits vary among METU students based on different demographic and behavioral factors, offering reliable conclusions.

**4.1. Research Philosopy**

The research philosophy we're following in this study is called positivism. In positivist research, the focus is on gathering objective facts that can be observed, measured, and tested. This approach allows for the formulation and testing of hypotheses based on objective facts, such as frequency of caffeine intake, cognitive performance measures, and sleep quality assessments. In positivist research, the role of the researcher is to focus on empirical data, ensuring that findings are grounded in factual evidence that can be generalized.

**5. SURVEY METHODOLOGY**

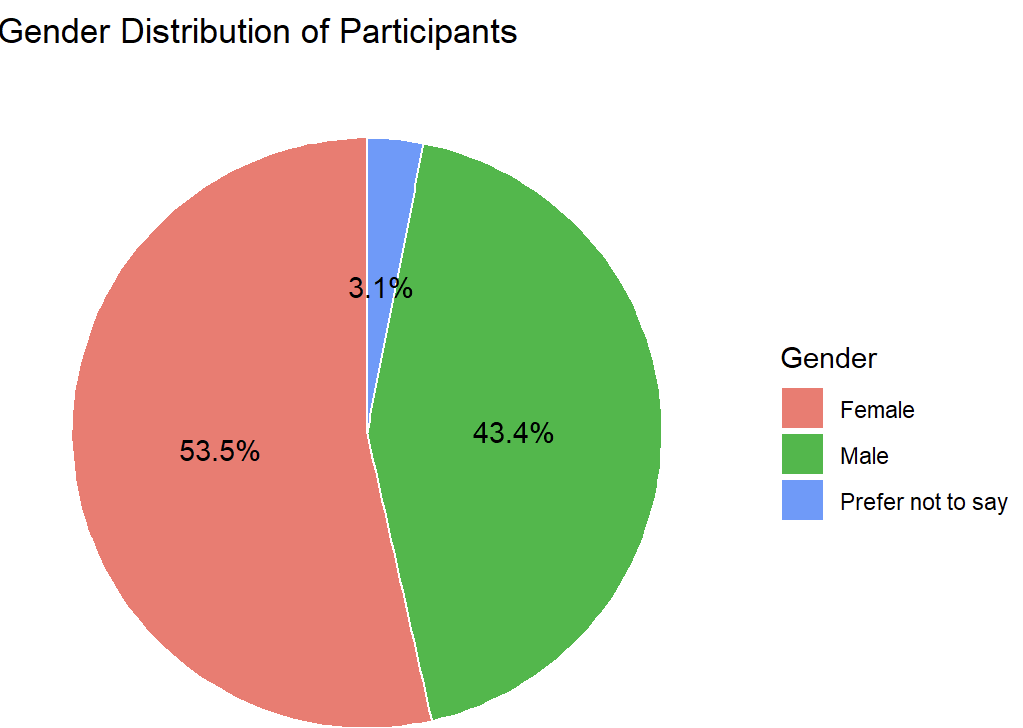
**5. Observations and Inferences**

During the survey, here are some highlights that caught our attention.

1. The third question of our survey, 'What year were you born?', is often misunderstood by participants, leading to some inaccurate age inputs
2. Preparatory students struggled to answer the questions due to their level of English.
3. Most of the larger groups declined to participate in the survey.
4. In restaurants and cafes, people who were waiting for their orders were more willing to participate, while others were not.
5. People refused to participate in the survey because they believed it was intended for certain groups.
6. Timing is important because people feel disturbed if asked to participae in the survey during active conversations.
7. Some participants asked to select more than one option in the last question of our survey, ‘What is your favourite coffeee type?’, which is restricted with only one selection.
8. Pepople with common interest were more willing to fill the form.
9. Some people dismiss the researcher with responses such as ‘I will complete it later.’
10. Open-ended questions received irrelevant and distruptive responses.
11. The survey is initially planned to be conducted both for METU and Bilkent students, but due to permission restrictions, it is conducted only with METU students.

**4.4 Descriptive Statistics**



Figure 1 visualizes the gender distribution of the survey participants. As shown in pie chart, the majority of respondents are female (53.5%), followed by 43.4% of male participants. A minority of the survey responses were from participants who identified as Prefer not to say (3.1%). Overall, this distribution clearly illustrates the gender representation among the respondents

A graph of a number of people

Description automatically generated

Figure 2 indicates that most survey participants were born between 1999 and 2005. The dataset is concentrated on a specific age group. The distribution is approximately symmetrical around the peak (2001–2002) and it gives a balanced representation within the main age group. Frequency of the respected years before 1995 or after 2005 are very few, which aligns with the expected demographic of university students.

|  |  |  |
| --- | --- | --- |
| **Faculty** | **Frequency** | **Percentage** |
| Architecture | 30 | 9% |
| Arts and Sciences | 82 | 25% |
| Economics and Administrative Sciences | 36 | 11% |
| Education | 36 | 11% |
| Engineering | 140 | 44% |

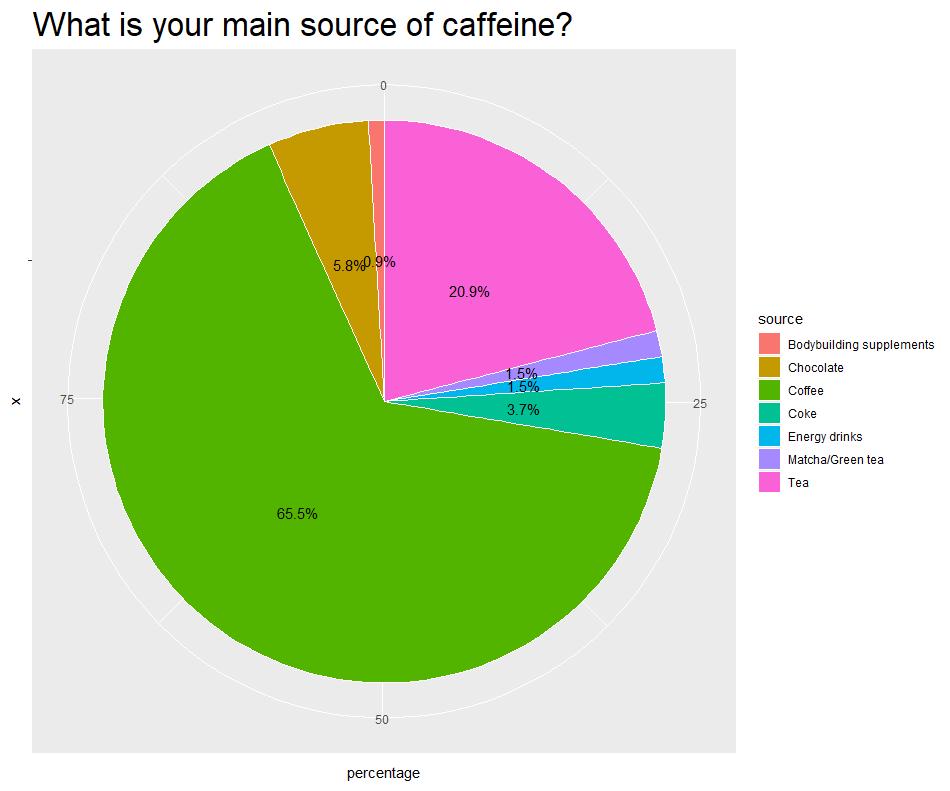
Forms response chart. Question title: What is your current year of study?  

. Number of responses: 332 responses.Figure 3 illustrates the distribution of survey participants based on their current year of study. The largest group of participants is from the 3rd year (25.3%), followed closely by the 4th year students (23.2%). Together, these two groups account for nearly half of the survey respondents. Preparatory and graduate students comprise smaller proportions, responsible for 11.4% and 7.1% of the responses. This distribution highlights a strong participation from 3rd and 4th year undergraduate students while graduate students and preparatory year participants are less represented in the survey.

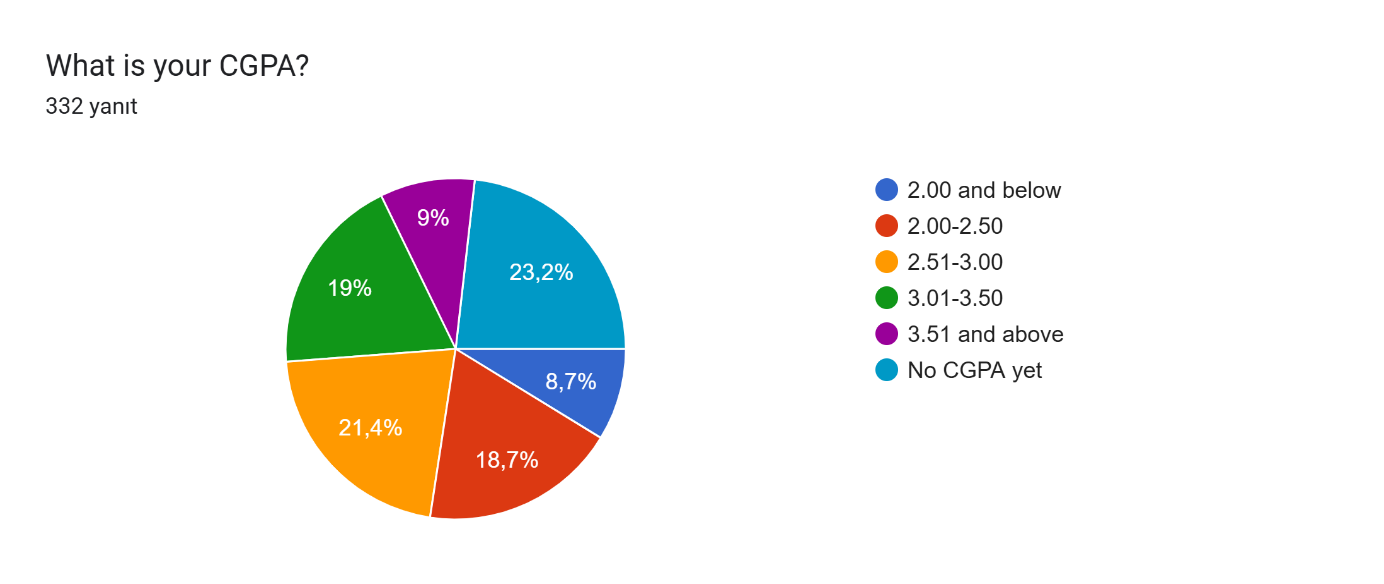
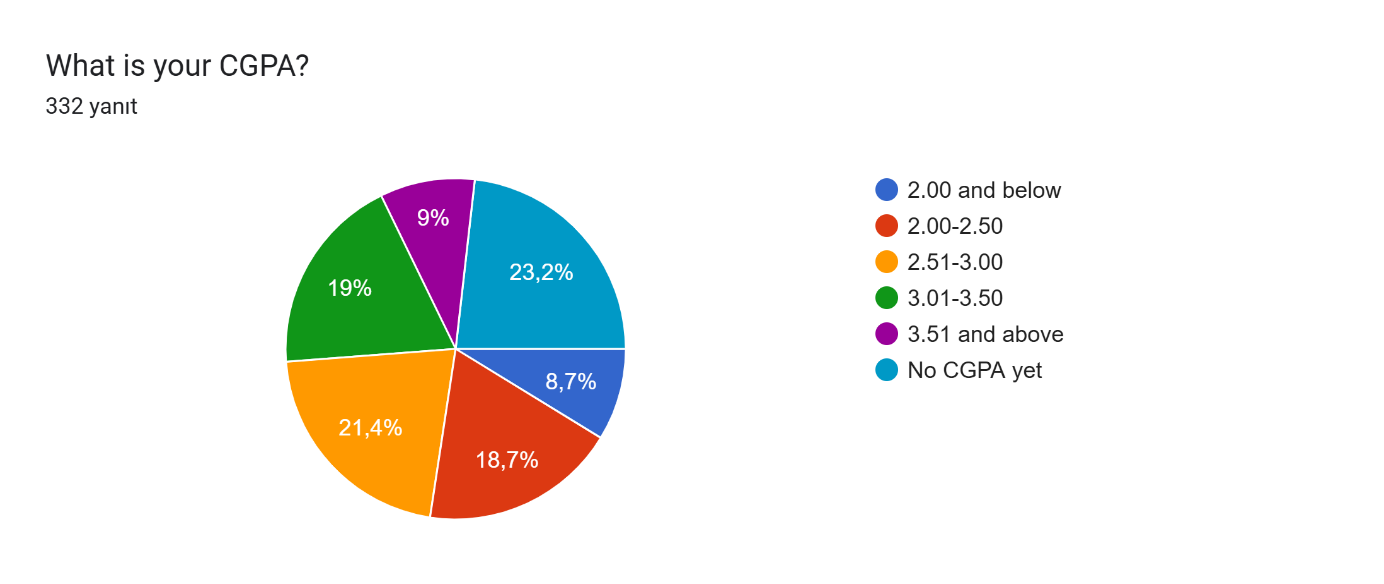
Tablo 1?? Faculty Distribution

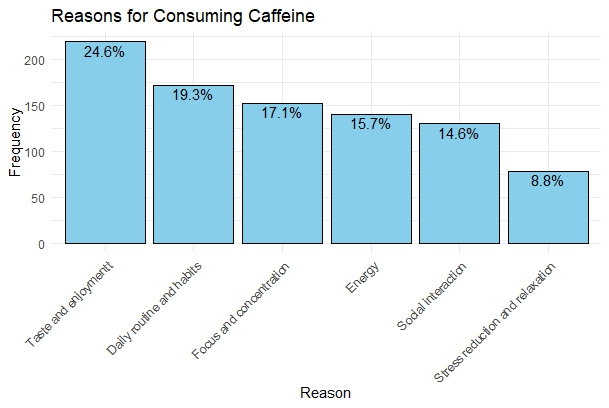
Table 1 represents the distribution of survey participants based on their faculties. The majority of survey participants are from the Faculty of Engineering, representing 44% of the total sample (140 students). The Faculty of Arts and Sciences accounts for the second-largest group with 25% (82 participants), followed by Economics and Administrative Sciences and Education at 11% each, and Architecture making up 9% of the total. This distribution indicates a strong representation of engineering students in the survey while smaller faculties like Architecture have a relatively limited contribution.

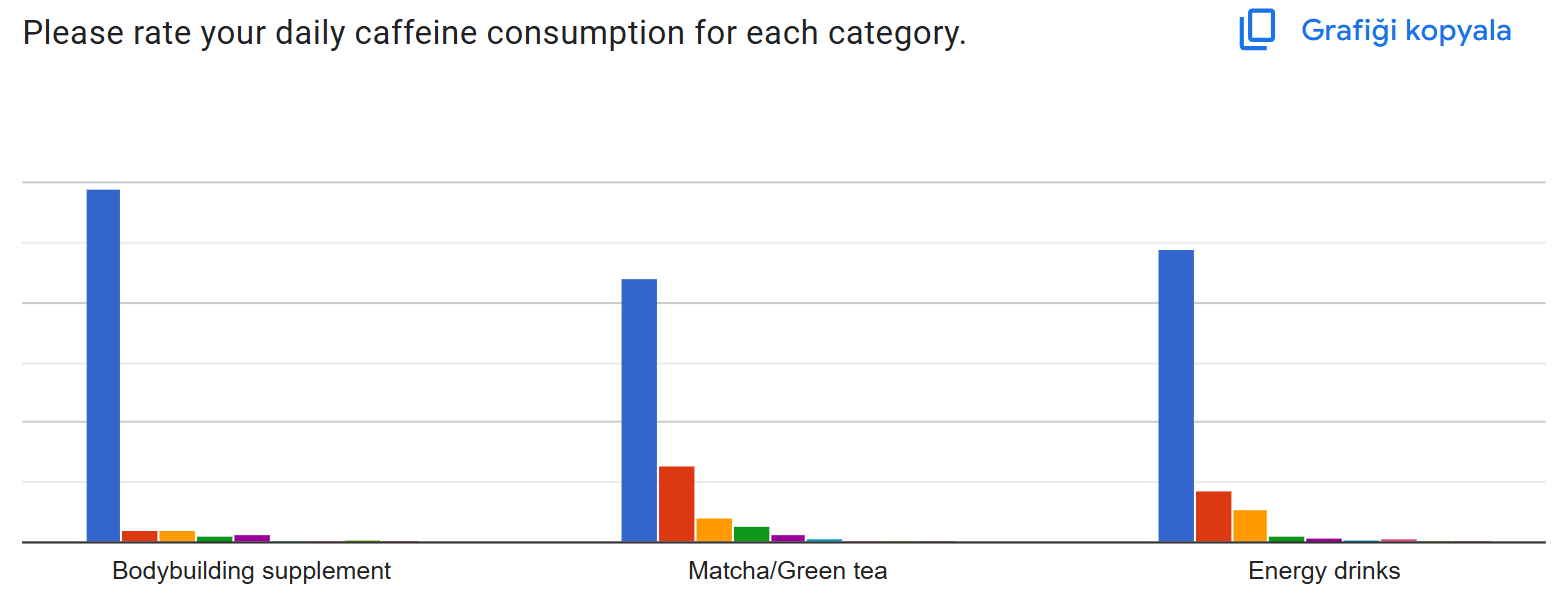
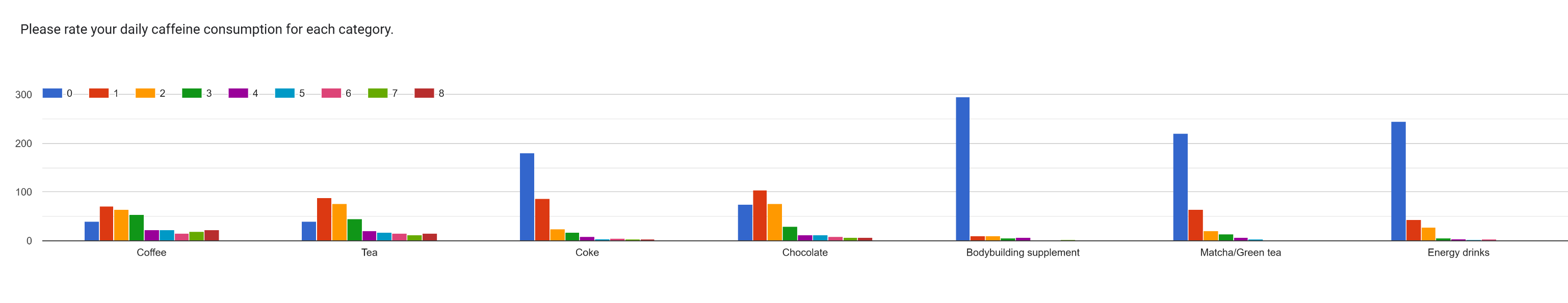
The distribution of participants across departments is shown below: The Statistics (STAT) department has the highest count with 32 participants, followed by Civil Engineering (CE) (25 participants), and Architecture (ARCH) and Electrical and Electronics Engineering (EEE) (21 participants each). Other participants are from Mechanical Engineering (ME), Chemical Engineering (CHE), and Political Science and Public Administration (PADM), each with 15. Lower representation is observed in departments like Sociology (SOC), Elementary Science Education (ESE), and Molecular Biology and Genetics (MOLE), with 2 participants each, while City and Regional Planning (CRP), History (HIST), and Childhood Education (CHED) had only 1 participant each.

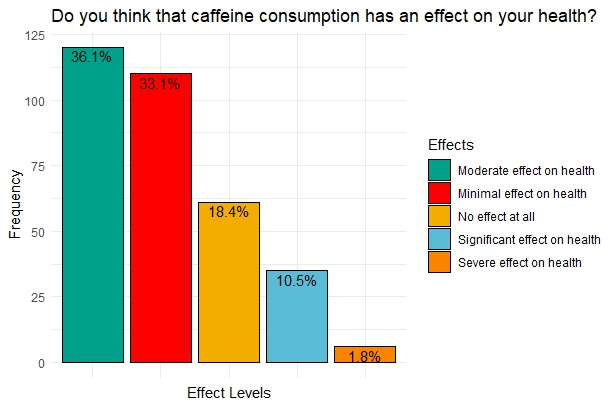
Coffee is the **dominant source** of caffeine, consumed by nearly two-thirds of participants. This reflects its popularity as a daily stimulant and a globally consumed beverage. Moreover, **tea (20.9%)** follows as a strong secondary choice, reflecting cultural or habitual consumption patterns. Chocolate contributes a moderate share to caffeine consumption with 5.8%. Other three sources Coke, Matcha/Green tea, and Energy drinks are not preferred as much as the other mentioned choices. Bodybuilding supplements choice is the **least significant source** of caffeine, reflecting its niche use.

Participants with the highest cgpa(3.51 and above) and the lowest cgpa (2.00 and belove) groups have the least percentages and close proportions to each other meaning they form the least crowded groups compared to other groups. Moreover, the most crowded groups are the preparatory students who do not have a cgpa yet, the group with 2.51-3.00 cgpa, and the group with 2.00-2.50 cgpa. These three groups form the majority of the participants all together.

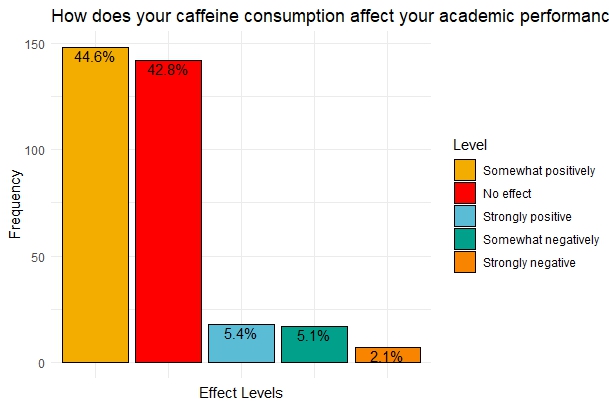


The barplot highlights that a quarter of the participants consume caffeine primarily for its enjoyable taste. Second most common reason is daily routine suggesting that caffeine consumption is of often embedded in participants’ dailty habits. With 17.1% portion, it is reflected that participants consume caffeinated products to maintain their focus. 14.6% of the participants use energy-boosting property of caffeine. Social perspective accounts for 15.7% of METU students. Lastly, a smaller group, (<10%) states that they use caffeine as a way of stress reduction.



The daily consumption graph for Coffee shows minimal skewness, indicating balanced and widespread consumption. Tea and Chocolate consumption is concentrated around 1–2 cups daily, with moderate right skewness, reflecting lower daily intake compared to Coffee. Coke, Matcha/Green Tea, and Energy Drinks show strong right skewness, with most participants consuming very little or none daily, highlighting their lower popularity. Bodybuilding supplements exhibit the strongest right skew, indicating they are rarely consumed and are the least popular caffeine source.

In the figure, more than one third of METU students state that the caffeine they take on a daily basis has moderate effect on their health. Slightly less proportion of the participants think it has a minimal effect. The people who claim that they are not affected by caffeine make up 18.4%. of the survey. While less than 50 students express that they are affected significantly, only 1.8% says caffeine has severe effect on their health.

According to the survey, nearly 90% of METU students say that their academic performance is not affected at all or somewhat positively affected due to their caffeine consumption habits. In addition, roughly equal number of students think that it has strongly positive (5.4%) or somewhat negative (5.1%) effect on academic aspect. However, only 7 individuals’ acedemic success is seriously adversely affected.

A graph of a number of blue rectangular objects

Description automatically generated

According to Figure 12, most METU students rarely experience sleeping struggles after caffeine consumption, with 96 responses. This is followed by 87 students selecting "sometimes." Additionally, 77 participants stated they "never" experience sleeping struggles. The remaining responses were "most of the time," with 46 participants, and "always," with 19 participants.

A graph of a bar graph

Description automatically generated with medium confidenceIn Figure 13, students were asked about the side effects they experienced after caffeine consumption. The most frequently selected responses were "No side effects at all" and "Heart throb," with 109 and 94 responses, respectively. Out of 326 total responses, 68 indicated addiction, 64 mentioned digestion problems, 67 reported anxiety triggers, 54 noted insomnia triggers, 49 experienced headaches, and 30 stated aggression. Accordingly, the most common response, with 109 mentions, was "No side effects at all," whereas the least reported side effect was aggression, with 30 mentions.

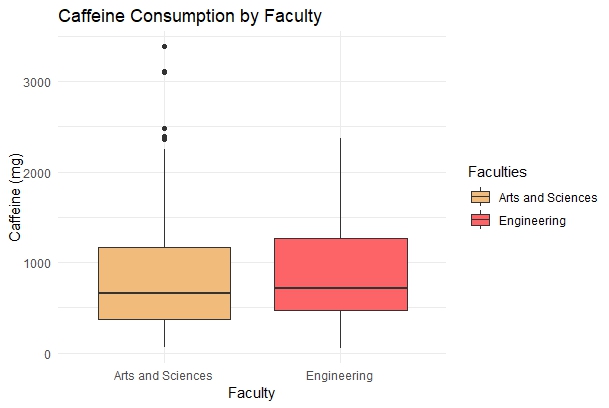
A pie chart of coffee

Description automatically generatedThe distribution of favorite coffee types has been shown in Figure 14. The most popular choice among students is espresso-based coffee, preferred by 36.4% of respondents. This is followed by coffee bean-based options at 31.62%, instant coffee at 12.7%, and Turkish coffee at 11.1%. Additionally, 8.1% of the 325 participants selected the option "I do not consume coffee."

**5. Data Analysis, Findings, and Discussions**

**5.1 Does total caffeine consumption amount differ among faculties?**

To ensure robust and reliable conclusions, the analysis focuses on the two most populated faculties(*Arts and Science (n=82) & Engineering (n=140)*). This approach leverages larger sample sizes to enhance statistical power and minimize variability.

To have an insight about the variances of the two faculties, we checked box-plots of each according to total caffein consumption of them.

The median caffeine intake is slightly higher in Engineering faculty, yet both faculties have close interquartile ranges, suggesting compartable variability in typical consumption. However, it is observed that there are individuals in Arts and Sciences facult with exceptionally high caffeine intake, while Engineering displays a broader overall range without outliers. In order to reach a robust result, some statistical methods are used.

Firstly, we need to check whether the variances or equal or not. To do so, we use Levene’s Test.

*H0* : σAS = σEng, *H1* : σAS ≠ σEng

|  |  |  |  |
| --- | --- | --- | --- |
| Levene’s Test for Homogenity | | | |
|  | Df | F value | Pr(>F) |
| group | 1 | 1.7924 | 0.182 |
|  | 220 |  |  |

With p-value = 0.182, we fail to reject the null hypothesis for α=0.05. Therefore, we conclude that data does not provide sufficient evidence to suggest that the variances of the two faculties differ significantly. Having found that, Two-Sample t-Test can be applied.

*H0* : µAS = µEng, *H1* : µAS ≠ µEng

|  |  |  |
| --- | --- | --- |
| Two-Sample t-Test | | |
| Df:  220 | T:  0.64278 | p-value:  0.521 |
| Confidence Interval:  (-120.3818, 236.9142) | | |
| Means | Arts and Sciences:  927.78 | Engineering:  869.51 |

Two-Sample t-Test shows that p-value (0.521) is not less than the significant level (0.05) which means the null hypothesis cannot be rejected. Moreover, the conficence interval includes zero suggesting no significant difference. Therefore, we conclude that there is no significant difference in mean caffeine consumption between Arts and Sciences and Engineering faculty students.

**5.2 Does the total caffeine consumption differ significantly among students with different CGPA ranges?**

We plan to investigate whether there is a significant difference in total caffeine consumption among students based on their CGPA. Start with setting hypotheses:

*H0* : µAS = σEng, *H1* : µAS ≠ σEng

There is no significant difference in the mean total caffeine consumption across CGPA categories. (1=2=3=…. =k)

**Alternative Hypothesis (H1) :** At least one CGPA category has a different mean total caffeine consumption.

We conducted a One-Way ANOVA to test the hypotheses. The dependent variable is Total Caffeine, and the independent variable is CGPA, which is a categorical variable with levels.

If the ANOVA test shows significant results (p<0.05), a Tukey HSD post-hoc test will be performed to identify exactly which groups differ.

To use ANOVA test, be sure that all assumptions are met:

1. Normality of Residuals:

To test the normality of residuals, both a Q-Q plot and the Shapiro-Wilk test were utilized:

1.1. Q-Q Plot

A graph with a line

Description automatically generated

The Q-Q plot (Figure 5) was examined to visualize the normality of residuals. As shown in the plot, particularly in the tails , the residuals initially deviated from normality.

1.2. Shapiro-Wilk Test:

To statistically confirm the normality, the Shapiro-Wilk test was performed.Hypotheses:

**Null Hypothesis(H0):** *H0* : µAS = σEng, *H1* : µAS ≠ σEng

**Alternative Hypothesis(Ha) METİN KUTUSUNA AL**

Residuals are normally distributed. Residuals are not normally distributed.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Phase** | **W statistic** | **p-value** | **Conclusion** |
| Initial Residuals | 0.90531 | 0,02135 | Not normally distributed |
| Residuals after Box-Cox | 0,99875 | 0.057 | Normally distributed |

The Shapiro-Wilk test resulted in p<0.05, therefore we reject null hypothesis, stating that the residuals were not normally distributed. To address the violation of normality, a Box-Cox transformation was applied to the dependent variable (Total Caffeine). After transformation, the residuals were re-tested, and the Shapiro-Wilk test resulted in p>0.05, confirming that the residuals were now normally distributed.

1. Homoscedasticity (Equal Variance):

To test the assumption of equal variances across CGPA groups, Levene’s Test was performed. Setting hypotheses:

**Null Hypothesis(H0): Alternative Hypothesis(H1)**

The variances across CGPA groups are equal. At least one CGPA group has a variance significantly different from the others. ?????????

|  |  |  |  |
| --- | --- | --- | --- |
| **LEVENEt** |  |  |  |
| **Source of Variation** | **DF** | **F-value** | **p-value** |
| **Group** | 5 | 1.1877 | 0.3149 |
| **Residuals** | 319 |  |  |

The results of Levene’s Test indicate that the variances are equal across CGPA groups (p=0.3149>0.05).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ANOVA** |  |  |  |  |  |
| **SOV** | **DF** | **SS** | **MS** | **F-value** | **p-value** |
| **CGPA** | 5 | 233 | 46.64 | 2.695 | 0.021 |
| **Residuals** | 319 | 5520 | 17.30 |  |  |

Using Box-cox transformation, we made sure that all required assumptions are met to conduct ANOVA test. The ANOVA test indicates that there is a statistically significant difference in Total Caffeine consumption among students with different CGPA categories (p=0.021). Conduct a post-hoc test to determine which CGPA categories differ significantly.

|  |  |  |  |
| --- | --- | --- | --- |
| Comparison | Mean Difference | 95% Confidence Interval | p-value |
| 2.00 and below - 2.00-2.50 | 0.79 | −1.96,3.55 | 0.963 |
| 2.51-3.00 - 2.00-2.50 | -0.34 | −2.43,1.75 | 0.997 |
| 3.01-3.50 - 2.00-2.50 | -0.91 | −3.06,1.24 | 0.832 |
| 3.51 and above - 2.00-2.50 | -0.77 | −3.46,1.92 | 0.964 |
| No CGPA yet - 2.00-2.50 | -2.01 | −4.06,0.04 | 0.057 |
| 2.51-3.00 - 2.00 and below | -1.13 | −3.83,1.57 | 0.837 |
| 3.01-3.50 - 2.00 and below | -1.70 | −4.45,1.05 | 0.484 |
| 3.51 and above - 2.00 and below | -1.56 | −4.75,1.63 | 0.725 |
| No CGPA yet - 2.00 and below | -2.81 | −5.48,−0.13 | **0.033\*** |
| 3.01-3.50 - 2.51-3.00 | -0.57 | −2.65,1.51 | 0.969 |
| 3.51 and above - 2.51-3.00 | -0.43 | −3.07,2.20 | 0.997 |
| No CGPA yet - 2.51-3.00 | -1.68 | −3.65,0.30 | 0.148 |
| 3.51 and above - 3.01-3.50 | 0.14 | −2.54,2.82 | 1.000 |
| No CGPA yet - 3.01-3.50 | -1.11 | −3.15,0.94 | 0.630 |
| No CGPA yet - 3.51 and above | -1.24 | −3.85,1.36 | 0.744 |

The Tukey's HSD test results confirm that students with "No CGPA yet" differ significantly in their caffeine consumption from those in the "2.00 and below" CGPA category. (mean difference=−2.81, p=0.033. No other pairwise comparisons were statistically significant due to the p-values

A graph of different colored rectangular shapes

Description automatically generatedThe box plot illustrates the total caffeine consumption across different CGPA categories. While median caffeine consumption appears relatively similar across most groups, students with "2.00 and below" CGPA exhibit slightly higher median values. The "No CGPA yet" group shows the greatest variability in caffeine consumption with a wide range. Statistical tests support this observation, like ANOVA stating that a significant difference in total caffeine consumption across CGPA categories (p=0.021). Post-hoc analysis showed that students with "No CGPA yet" consume significantly less caffeine than those with "2.00 and below" CGPA, while no other differences were statistically significant. This suggests caffeine consumption may vary slightly with academic progression or habits.

**5.3 Is There a Relationship Between Gender and Side Effects Caused by Caffeine Consumption?**  
This study aims to examine the potential relationship between the side effects experienced by university students due to caffeine consumption and the gender variable. Caffeine consumption is not only influenced by gender but also impacts how individuals experience side effects such as headaches, digestion problems, and heart palpitations. Research suggests that gender differences may play a role in the frequency and severity of these side effects due to physiological and hormonal variations (Lovallo et al., 2004). Females, for instance, may be more sensitive to caffeine’s cardiovascular effects, while males may report a higher frequency of gastrointestinal disturbances.An analysis of survey data on caffeine-related side effects across genders provides valuable insights into these variations and their implications for health management strategies. In the study, the distributions of side effects reported across different gender groups were statistically evaluated using the Chi-Square (χ²) test. The reason for selecting this test is that the data analyzed is categorical both for the independent variable (gender) and the dependent variable (caffeine consumption side effects). The Chi-Square test is an ideal statistical method for evaluating the relationship between such categorical variables.

A blue squares with white text

Description automatically generatedMoreover, the Chi-Square test is frequently used for analyzing relationships between variables with multiple categories. In this study, the gender variable is divided into two categories (female, male), and caffeine-related side effects are categorized into 8 different types (addiction, headache, digestive problems, etc.).

Graphical analyses indicate that side effects such as "Addiction" and "Digestive Problems" are reported more frequently among female participants, while "Heart Palpitations" and "Insomnia" are observed more frequently among male participants. However, the "No Side Effect" option has been observed at a notably high frequency in both groups. Furthermore, when the data distribution was examined, it was found that the frequency in each cell was greater than 5. This confirms that the assumption for the Chi-Square test has been met, allowing the hypothesis to be tested.

A purple rectangular sign with black text

Description automatically generated *H₀*: There is no relationship between gender and the side effects caused by caffeine consumption.

**Null Hypothesis (H₀):** There is no relationship between gender and the side effects caused by caffeine consumption.

**Alternative Hypothesis (H₁):** There is a significant relationship between gender and the side effects caused by caffeine consumption.

*H₁:* There is a significant relationship between gender and the side effects caused by caffeine consumption.

According to the Chi-Square test results, the P-value was found to be 0.5394, which is greater than the significance level of 0.05. Therefore, the null hypothesis (H₀) cannot be rejected. It can be concluded that there is no statistically significant relationship between gender and the side effects caused by caffeine consumption.

A graph of different colored bars

Description automatically generatedThe graph shows the distribution of side effects caused by caffeine consumption across gender groups. Digestive problems and addiction are more frequently reported among women, while insomnia and heart palpitations are more prevalent among male participants. However, these differences are not statistically significant and are likely to be random. Conducting separate tests for each side effect would require performing a large number of tests, which could lead to false positive results (Type I Error), and therefore, this approach was not preferred.

This analysis, while evaluating the side effects of caffeine consumption among university students, concludes that gender differences do not play a determining role in the side effects experienced.

**5.4 Do main caffeine consumption source and age affects total caffeine amount consumed by the METU students?**

The total amount of consumed caffeine differ significantly among individuals, often influenced by factors such as age or the caffeine source. At METU; understanding these preferences and their relationship with total caffeine consumption is essential to identify consumption patterns and behaviors. To explore the relationship between total caffeine consumption and the predictors (age and main caffeine source), a multiple linear regression model is conducted. Age is included in the model as a continuous variable to evaluate how caffeine consumption changes across different ages. The main caffeine source, a categorical variable, is set as factors to represent different preferences such as coffee, tea, energy drinks, or soft drinks. This categorization allows us to compare the influence of each source on total caffeine consumption. The model includes the dependent variable y (total caffeine amount), and independent variables x’s (age and main sources). To conduct the model, first assumptions should be checked.

For the linearity assumption Ramsey RESET test is conducted:

|  |  |  |  |
| --- | --- | --- | --- |
| RESET | Df1 | Df 2 | p-value |
| 4.3855 | 2 | 315 | 0.04323 |

Since the p-value of the RESET test is less than 0.05, the null hypothesis is rejected concluding the linearity assumption is violated. However, since it is a very close value, it will be assumed to satisfy the condition.

Secondly, for homoscedasticity, Breusch-Pagan test is conducted:

|  |  |  |
| --- | --- | --- |
| Breusch-Pagan | Df | p-value |
| 6.5487 | 7 | 0.4773 |

Since the p-value of the Breusch-Pagan test is lower than 0.05, the H0 is rejected concluding the homoscedasticity assumption is violated. However, since it is a very close value, it will be assumed to satisfy the condition.

For the uncorrelation and independence assumptions, Durbin-Watson test is conducted:

|  |  |  |
| --- | --- | --- |
| Autocorrelation | Durbin-Watson | p-value |
| 0.0184 | 1.95 | 0.61 |

Since the p-value of the Durbin-Watson test is greater than 0.05, the H0 is failed to reject concluding the uncorrelation and independence assumptions is satisfied.

For the normally distribution assumption, Shapiro-Wilks test is conducted:

|  |  |
| --- | --- |
| Shapiro-Wilks | p-value |
| 0.9453 | 0.048 |

Since the p-value of the Shapiro-Wilks test is lower than 0.05, the H0 is rejected concluding the normality assumption is violated. However, since it is a very close value, it will be assumed to satisfy the condition.

For the no multicollinearity assumption, the VIF function is used:

|  |  |  |
| --- | --- | --- |
|  | Df | VIF |
| Main Source of Caffeine | 6 | 1.0007 |
| Age | 1 | 1.0040 |

Since the VIF results are less than 5, there is no multicollinearity, concluding the assumption is satisfied.

The assumptions are checked, in next step the multiple linear regression model will be introduced.

Firstly, to observe the effect of the age variable on the total caffeine consumption, consider the following simple linear regression model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | SE | t-statistic | | p-value | |
| Intercept | 493.582 | 121.814 | | 4.052 | | 6.37e-05 |
| Age | 17.835 | 5.117 | | 3.485 | | 0.00056 |

The model to predict the total caffeine amount with the age variable is:

The model suggests that for every one unit increase in the age variable, the total caffeine consumption increases by 17.835 milligrams. This increase is incremental and cumulative, indicating that older individuals are consistently consuming more caffeine compared to younger individuals. The intercept estimate is 493.582, with a p-value 6.37e-05, and the coefficient for the age variable is 17.835, with a p-value 0.00056. The null hypothesis suggests that the model is not significant. However, as their p-values are less than 0.05 the null hypothesis is rejected. Therefore, concluded as both the intercept and the age variable are statistically significant in the model indicating that they contribute meaningfully to explaining variations in caffeine consumption.

Secondly, to observe the effect of main source of caffeine variable on the total caffeine consumption, consider the following linear regression model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Estimate | SE | t-statistic | | p-value |
| Intercept | 1466.7 | 374.1 | 3.921 | 0.000108 | |
| Chocolate | -991.1 | 402.5 | -2.462 | 0.014342 | |
| Coffee | -448.1 | 376.7 | -1.190 | 0.235121 | |
| Coke | -702.8 | 418.2 | -1.680 | 0.093861 | |
| Energy drinks | -656.9 | 473.2 | -1.388 | 0.166042 | |
| Matcha/Green tea | -1255.0 | 473.2 | -2.652 | 0.008397 | |
| Tea | -770.6 | 382.2 | -2.016 | 0.044638 | |

The model to predict the total caffeine amount with the main source of caffeine variable is:

The model is designed in a way that the participant’s main source of caffeine choice (only 1 of the above sources is selected by participants of the survey) predicts the total caffeine amount consumed by that participant, with the intercept and the coefficient of the source. For example, if one’s main caffeine source is coke, the resulted model equation becomes in this case. Moreover, the intercept is highly significant since its p-value (0.000108) is lower than 0.05. For Chocolate, the coefficient (-991.1) is statistically significant since its p-value (0.014342) is lower than 0.05, suggesting that participants whose main caffeine source is chocolate consume significantly less caffeine compared to most of the other groups. For Matcha/Green Tea, the coefficient (-1255.0) is also statistically significant since its p-value (0.008397) is lower than 0.05, indicating that participants whose main source of caffeine is Matcha/Green tea consume the least total amount of caffeine compared to other groups. However, the variables Coffee (p-value = 0.235121), Coke (p-value = 0.093861), and Energy Drinks (p-value = 0.166042) are not statistically significant since their p-values are greater than 0.05, meaning their effects on caffeine consumption are not strong enough to draw definitive conclusions. The significant coefficients for some categories suggest that the main caffeine source is an important factor in determining total caffeine consumption. The results of the regression model reveal an interesting pattern; some of the more popular caffeine sources are not statistically significant predictors of total caffeine consumption while some less preferred sources are found to be significant. For example, surprisingly, the most popular caffeine source, coffee, is found to be insignificant to predict the total caffeine amount. However, even though the mentioned situation may be unexpected for most people, the insignificant sources will be excluded to get the final regression model.

To ensure that the insignificant variables in the above model are also insignificant in the multiple regression model, observe the model which includes age and all the main sources of caffeine.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | SE | t-statistic | p-value |
| Intercept | 1088.515 | 386.283 | 2.818 | 0.00514 |
| Age | 16.207 | 4.963 | 3.266 | 0.00121 |
| Chocolate | -975.424 | 396.574 | -2.460 | 0.01444 |
| Coffee | -443.605 | 371.106 | -1.195 | 0.23284 |
| Coke | -652.822 | 412.299 | -1.583 | 0.11433 |
| Energy drinks | -644.982 | 466.156 | -1.384 | 0.16745 |
| Matcha/Green tea | -117.151 | 466.286 | -2.610 | 0.00948 |
| Tea | -752.567 | 376.599 | -1.998 | 0.04654 |

Similar to the previous results, Coffee, Coke and Energy drinks are not significant for the model since their p-values are greater than 0.05. In the next and final model, these variables will not be included.

Finally, the resulted multiple regression model includes both age and the significant main caffeine sources to predict the total caffeine amount consumed by the participants.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | SE | t-statistic | p-value |
| Intercept | 2171.41 | 528 | 4.106 | 8.85e-05 |
| Age | -30.20 | 18.23 | -2.657 | 0.01011 |
| Chocolate | -1020.20 | 338.52 | -3.014 | 0.003350 |
| Matcha/Green tea | -1325.44 | 399.66 | -3.316 | 0.00132 |
| Tea | -804.21 | 321.66 | -2.500 | 0.01423 |

The multiple linear regression model to predict the total caffeine amount with age and the significant main sources of caffeine variable is:

By excluding insignificant predictors like coffee and energy drinks, the final model is significant and provides meaningful insights into the key factors influencing caffeine consumption. The final multiple linear regression model provides a comprehensive analysis of total caffeine consumption, including age and three significant caffeine sources; Chocolate, Matcha/Green Tea, and Tea. The model indicates that caffeine consumption decreases with age and is lower among participants whose main sources are these three options. The intercept and the coefficients for all predictors are statistically significant since all having p-values lower than 0.05. In summary, this final model highlights that age and the choice of specific caffeine sources are important predictors of total caffeine consumption, with clear statistical significance of all included variables.

**REFERENCES**

Clark, I., & Landolt, H. P. (2017). Coffee, caffeine, and sleep: A systematic review of epidemiological studies and randomized controlled trials. *Sleep Medicine Reviews, 31*, 70–78.

Food and Drug Administration (FDA). (2021). *Spilling the Beans: How Much Caffeine is Too Much?* Retrieved from [FDA.gov](https://www.fda.gov)

Juliano, L. M., & Griffiths, R. R. (2004). A critical review of caffeine withdrawal: Empirical validation of symptoms and signs, incidence, severity, and associated features. *Psychopharmacology, 176*(1), 1–29.

Lara, D. R. (2014). Caffeine, mental health, and psychiatric disorders. *Journal of Caffeine Research, 4*(4), 151–158.

Lopez, A. A., et al. (2017). Gender differences in caffeine consumption patterns and effects on academic stress. *Gender and Behavioral Science, 10*(2), 43–58.

Lovallo, W. R., et al. (2004). Gender differences in caffeine's cardiovascular effects: The role of estrogen and individual sensitivity. *Journal of Psychopharmacology, 18*(4), 510–516.

McLellan, T. M., Caldwell, J. A., & Lieberman, H. R. (2016). A review of caffeine's effects on cognitive, physical, and occupational performance. *Neuroscience & Biobehavioral Reviews, 71*, 294–312.

Smith, A. P., et al. (2020). The effects of caffeine on cognitive function and performance. *Journal of Cognitive Enhancement, 4*(3), 215–230.  
 Lovallo, W. R., et al. (2004). Gender differences in caffeine's cardiovascular effects: The role of estrogen and individual sensitivity. *Journal of Psychopharmacology, 18*(4), 510–516.

,

4. What is your department?  
(Please type your department in lowercase and explicitly.)

5. What is your CGPA?  
2.00 and below  
2.00-2.50  
2.51-3.00  
3.01-3.50  
3.51 and above  
No CGPA yet

6. What is your main source of caffeine?  
Matcha/Green tea  
Energy drinks  
Coffee  
Coke  
Tea  
Bodybuilding supplements  
Chocolate  
Other:

9. What is your reason for consuming caffeine?  
(You can select multiple options.)  
Daily routine and habits  
Focus and concentration  
Stress reduction and relaxation  
Energy  
Social interaction  
Taste and enjoyment  
Other:

10. Do you think that caffeine consumption has an effect on your health?  
No effect at all  
Minimal effect on health  
Moderate effect on health  
Significant effect on health  
Severe effect on health

7. Please rate your daily caffeine consumption for each category  
(e.g., coffee as a cup, tea as a tea cup, Coke as a can, chocolate as a bar, supplement as a scoop)  
Options for each category: 0, 1, 2, 3, 4, 5, 6, 7, 8  
Coffee:   
Tea:   
Coke:   
Bodybuilding supplement:   
Energy drinks:   
Chocolate:   
Matcha/Green tea:

8. How would you rate your weekly caffeine consumption cost?  
Very Low (1) to Very High (5)

1. What is your gender?  
Female  
Male  
Prefer not to say  
Other: \_\_\_\_\_\_\_

2. What year were you born?  
(Please enter the year)

3. What is your current year of study?  
Preparatory  
1st year  
2nd year  
3rd year  
4th year  
Graduate

**Null Hypothesis (H₀):** There is no relationship between gender and the side effects caused by caffeine consumption.

**Alternative Hypothesis (H₁):** There is a significant relationship between gender and the side effects caused by caffeine consumption.

**Null Hypothesis (H₀):** There is no relationship between gender and the side effects caused by caffeine consumption.

**Alternative Hypothesis (H₁):** There is a significant relationship between gender and the side effects caused by caffeine consumption.

13. Which side effects do you experience after consuming caffeine?  
(You can select multiple options.)  
Insomnia trigger  
Headache  
Anxiety trigger  
Aggression  
Heart throb  
Digestion problems  
Addiction  
No side effects at all  
Other:

14. What is your favorite coffee type?  
(Percentage Distribution)  
Turkish Coffee  
I do not consume coffee.  
Coffee bean-based (Filter coffee)

Espresso-based (Americano, Latte...)  
Instant coffee

12. Do you struggle with sleeping after consuming caffeine?  
Never  
Rarely  
Sometimes  
Most of the time  
Always

11. How does your caffeine consumption affect your academic performance?  
Strongly negative  
Somewhat negatively  
No effect  
Somewhat positively  
Strongly positive

install.packages("readxl"); install.packages("writexl"); library(readxl); library(dplyr); library(writexl); data <- read\_excel("caffein.xlsx"); data$'What is your department?\n' <- tolower(data$'What is your department?\n'); data$dep <- tolower(data$'What is your department?\n'); dep\_map <- c("food engineering" = "fde", "physics" = "phys", "foreign language education" = "fle", "mechanical engineering" = "me", "elemantary mathematics educational" = "eme", "elementary mathematic education" = "eme", "elementary mathematics education" = "eme", "elemantry math education" = "eme", "politics" = "padm", "eee" = "eee", "early childhood education" = "ece", "elementry science education" = "ese", "political science and public administration" = "padm", "economics" = "econ", "psychology" = "psy", "petroleum and natural gas engineering" = "pete", "chemical engineering" = "che", "biology" = "bio", "political science and public adm" = "padm", "metallurgical and materials engineering" = "mete", "ie(industrial engineering)" = "ie", "international relations" = "ir", "molecular biology and genetics" = "mole", "civil engineering" = "ce", "endustriyel tasarim" = "id", "computer engineering" = "ceng", "business administration" = "ba", "philosophy" = "phil", "statistics" = "stat", "aee" = "aee", "architecture" = "arch", "math" = "math", "chem" = "chem", "uluslararasi iliskiler" = "ir", "political science" = "padm", "global international affairs" = "gia", "business administration suny" = "ba", "industrial design" = "id", "ched" = "ched", "computer education and instructional technologies" = "ceit", "industrial engineering" = "ie", "fde" = "fde", "mimarlD1k" = "arch", "che" = "che", "mining engineering" = "mine", "architecturr" = "arch", "electric and electronics engineerD1ng" = "eee", "fizik" = "phys", "metallD1rgical and material enginering" = "mete", "electrical and electronics engineering" = "eee", "electric electronics engineering" = "eee", "electical and electronics engineering" = "eee", "me" = "me", "industrial enginnering" = "ie", "mathematics" = "math", "D1ndustrial design" = "id", "computer education and instructional technology" = "ceit", "chemistry" = "chem", "environmental engineering" = "enve", "stat" = "stat", "ee" = "eee", "elementary math education" = "eme", "biol" = "bio", "electrics and electronics engineering" = "eee", "istatistik" = "stat", "mechanical engineer" = "me", "metalurji ve malzeme mC<hendislio?=\u009fi" = "mete", "bilgisayar mC<hendislio?=\u009fi" = "ceng", "geoe" = "geoe", "history" = "hist", "geological engineering" = "geoe", "elemantary mathematics education" = "eme", "mechanical enginering" = "me", "bussines administration" = "ba", "computer engineer" = "ceng", "aerospace engineering" = "ae", "adm" = "adm", "ce" = "ce", "sosyoloji" = "soc", "eme" = "eme", "elementary science education" = "ese", "city and regional planning" = "crp", "elektronik" = "eee", "sociology" = "soc", "civil engineerin" = "ce", "aerospace" = "ae", "pete" = "pete", "phil" = "phil", "aerospace engineer" = "ae", "metallurgical and material engineering" = "mete", "mhed" = "eme", "havacD1lD1k ve uzay mC<hendislio?=\u009fi" = "ae", "havacD1lD1k ve uzay mC<hendisliD i" = "ae", "Bilgisayar mC<hendisliD i" = "ceng", "Metalurji ve Malzeme MC<hendisliD i"= "mete"); data$dep <- dep\_map[match(data$dep, names(dep\_map))]; faculty\_mapping <- list(Engineering = c("aee", "ceng", "che", "enve", "geoe", "ie", "me", "mete", "mine", "pete", "fde", "ce", "eee"), Arts\_and\_Sciences = c("chem", "hist", "math", "phil", "phys", "psy", "soc", "stat", "bio", "mole"), Education = c("ceit", "ece", "ese", "fle", "ched", "eme"), Architecture = c("id", "arch", "crp"), Economics\_and\_Administrative\_Sciences = c("padm", "ba", "econ", "ir")); data <- data %>% mutate(Faculties = case\_when(dep %in% faculty\_mapping$Engineering ~ "Engineering", dep %in% faculty\_mapping$Arts\_and\_Sciences ~ "Arts and Sciences", dep %in% faculty\_mapping$Education ~ "Education", dep %in% faculty\_mapping$Architecture ~ "Architecture", dep %in% faculty\_mapping$Economics\_and\_Administrative\_Sciences ~ "Economics and Administrative Sciences", TRUE ~ NA\_character\_)); caffeine\_content <- c(Coffee = 202.5, Tea = 55, Green\_tea = 32, Energy\_Drinks = 80, Coke = 34, Chocolate = 10); data <- data %>% mutate(Total\_Caffeine = (Please rate your daily caffeine consumption for each category. [Coffee] \* caffeine\_content["Coffee"]) + (Please rate your daily caffeine consumption for each category. [Tea] \* caffeine\_content["Tea"]) + (Please rate your daily caffeine consumption for each category. [Matcha/Green tea] \* caffeine\_content["Green\_tea"]) + (Please rate your daily caffeine consumption for each category. [Energy drinks] \* caffeine\_content["Energy\_Drinks"])); write\_xlsx(data, "cleaned\_data4.xlsx"); install.packages("readxl"); library(readxl); library(ggplot2); library(dplyr); datafinal <- read\_excel("final\_cleaned2\_en\_son.xlsx"); gender\_data <- as.data.frame(table(datafinal$What is your gender?)); colnames(gender\_data) <- c("Gender", "Count"); gender\_data$Percentage <- round(100 \* gender\_data$Count / sum(gender\_data$Count), 1); ggplot(gender\_data, aes(x = "", y = Count, fill = Gender)) + geom\_bar(stat = "identity", width = 1, color = "white") + coord\_polar("y", start = 0) + geom\_text(aes(label = paste0(Percentage, "%")), position = position\_stack(vjust = 0.5)) + labs(title = "Gender Distribution of Participants", x = NULL, y = NULL, fill = "Gender") + theme\_minimal() + theme(axis.text = element\_blank(), axis.ticks = element\_blank(), panel.grid = element\_blank()); datafinal$What year were you born?[datafinal$What year were you born? == 1881] <- 2002; datafinal$What year were you born?[datafinal$What year were you born? == 2019] <- 2002; datafinal$What year were you born? <- as.numeric(datafinal$What year were you born?); ggplot(datafinal, aes(x = datafinal$What year were you born?)) + geom\_histogram(binwidth = 1, fill = "brown", color = "black", boundary = 1995) + labs(title = "Distribution of Participants' Birth Years", x = "Year of Birth", y = "Frequency") + theme\_minimal(); CGPA <- datafinal$What is your CGPA? <- as.factor(datafinal$What is your CGPA?); anova\_result <- aov(Total\_Caffeine ~ CGPA, data = datafinal); summary(anova\_result); plot(anova\_result, which = 2); residuals\_anova <- residuals(anova\_result); shapiro\_test <- shapiro.test(residuals\_anova); shapiro\_test; optimal\_lambda <- 0.25; if (optimal\_lambda == 0) { datafinal$BoxCox\_Total\_Caffeine <- log(datafinal$Total\_Caffeine) } else { datafinal$BoxCox\_Total\_Caffeine <- (datafinal$Total\_Caffeine^optimal\_lambda - 1) / optimal\_lambda }; anova\_boxcox <- aov(BoxCox\_Total\_Caffeine ~ CGPA, data = datafinal); summary(anova\_boxcox); residuals\_boxcox <- residuals(anova\_boxcox); shapiro\_test\_boxcox <- shapiro.test(residuals\_boxcox); shapiro\_test\_boxcox; posthoc <- TukeyHSD(anova\_boxcox); posthoc; install.packages("car"); library(car); leveneTest(Total\_Caffeine ~ CGPA, data = datafinal); ggplot(datafinal, aes(x = CGPA, y = BoxCox\_Total\_Caffeine, fill = CGPA)) + geom\_boxplot() + labs(title = "Total Caffeine Consumption by CGPA Categories", x = "CGPA Categories", y = "Total Caffeine (mg)") + theme\_minimal(); library(tidyverse); side\_effect\_data <- data %>% select(What is your gender?, Which side effects do you experience after consuming caffeine?) %>% rename(Gender = What is your gender?, Side\_Effect = Which side effects do you experience after consuming caffeine?); side\_effect\_data <- side\_effect\_data %>% mutate(Side\_Effect = case\_when(grepl("Headache", Side\_Effect) ~ "Headache", grepl("Addiction", Side\_Effect) ~ "Addiction", grepl("Insomnia", Side\_Effect) ~ "Insomnia", grepl("No side effects", Side\_Effect) ~ "No Side Effects", TRUE ~ "Other")); chi\_test <- chisq.test(table\_data); print(chi\_test); side\_effect\_data %>% group\_by(Gender, Side\_Effect) %>% summarise(Count = n()) %>% ggplot(aes(x = Gender, y = Count, fill = Side\_Effect)) + geom\_bar(stat = "identity", position = "dodge") + labs(title = "Gender vs Side Effects", x = "Gender", y = "Count") + theme\_minimal(); chisq.test(data)$expected; table\_data <- table(data$What is your gender?, data$Which side effects do you experience after consuming caffeine?); chisq.test(table\_data); print(table\_data);