Optimizing Productivity: How Caffeine, Physical Activity, and Sleep Quality Shape Daily Performance

This research study investigates the relationship between habits (such as caffeine intake, frequency of physical activity, and sleep hours) and university students' productivity. Using a correlational study design with a target sample of 30-50 students, anonymous surveys will collect data on variables like caffeine consumption patterns, exercise frequency, sleep duration, and daily productivity ratings (on a 0-100 scale). The original proposal included tracking over time. However, based on feedback, the study will focus on data collection "on a given day" rather than weekly measurements due to survey anonymity requirements. The analysis will primarily use multiple regression and correlation analyses to examine relationships between these variables.

Research question: How do caffeine intake, physical activity, and sleep quality relate to self-reported productivity levels among university students on a daily basis?

- The population for this study consists of university students who are likely to consume caffeine daily, engage in physical activity, and experience varying sleep habits. This demographic is ideal because they often manage their productivity in academic or work environments.
- Recruitment will target students from the University of Southern California (USC), where the study is being conducted. We will share the survey through Email, Student organization platforms, and word of mouth.
- A sample of 30-50 participants is a reasonable size for detecting patterns in correlational research while maintaining manageable data collection and analysis efforts for our study.

Type of Analysis Conducted and Results:

a. Multiple regression analysis:

A multiple regression analysis was conducted to predict daily productivity based on work/study hours, sleep hours, energy fluctuations, age, consistent sleep, and exercise frequency. The overall model was significant $[R^2=0.597,\,F(6,\,37)=9.12,\,p<.001]$. Significant predictors included work/study hours (B = 0.075, SE = 0.028, t(37) = 2.67, p = .011), sleep hours (B = -0.309, SE = 0.093, t(37) = -3.33, p = .002), and energy fluctuations (B = 0.374, SE = 0.144, t(37) = 2.60, p = .013). These findings indicate that these predictors significantly contribute to daily productivity while holding all other variables constant. Non-significant predictors included age (B = -0.031, SE = 0.020, t(37) = -1.56, p = .131), consistent sleep (B = 0.164, SE = 0.090, t(37) = 1.82, p = .078), and exercise frequency (B = 0.017, SE = 0.051, t(37) = 0.35, p = .729).

For productivity consistency, the regression model was also significant [R^2 = 0.541, F(6, 37) = 7.26, p < .001]. Energy fluctuations significantly predicted productivity consistency (B = 0.829, SE = 0.170, t(37) = 4.88, p < .001), and exercise frequency showed a marginally significant effect (B = 0.111, SE = 0.058, t(37) = 1.93, p = .064). Age (B = -0.024, SE = 0.022, t(37) = -1.03, p = .308), work/study hours (B = -0.013, SE = 0.033, t(37) = -0.41, p=.684), sleep hours (B = -0.177, SE = 0.111, t(37) = -1.59, p = .112, and consistent sleep (B = -0.036, SE = 0.106, t(37) = -0.34, p = .737) were not significant predictors above and beyond other variables.

b. Pearson correlation analysis:

A Pearson correlation analysis revealed significant positive correlations between energy fluctuations and daily productivity (r = .543, p < .001), energy fluctuations and productivity consistency (r = .674, p < .001), and work/study hours and daily productivity (r = .529, p < .001). Consistent sleep was positively correlated with daily productivity (r = .375, p = .038) and productivity consistency (r = .311, p = .048). Sleep hours are negatively associated with daily productivity (r = -0.436, p = .004).

These results indicate that while energy fluctuations, work/study hours, and sleep hours were significant predictors of daily productivity in the regression model, energy fluctuations were the strongest predictor of productivity consistency. Other predictors, while showing significant correlations, were not significant in the regression model when holding all other variables constant.

Support for Hypothesis:

- 1. Regression Analysis:
 - a. Daily Productivity:
 - i. The regression analysis provided partial support for the hypothesis. Significant predictors included:
 - 1. Work/Study Hours (B = 0.075, SE = 0.028, t(37) = 2.67, p = .011)
 - 2. Sleep Hours (B = -0.309, SE = 0.093, t(37) = -3.33, p = .002)
 - 3. Energy Fluctuations (B=0.374,SE=0.144,t(37)=2.60,p=.013)
 - ii. These results allow for the rejection of the null hypothesis for these predictors, indicating that they contribute significantly to daily productivity above and beyond other variables.
 - iii. Non-significant predictors in the regression model included:
 - 1. Age (p=.131)
 - 2. Consistent Sleep (p=.078)
 - 3. Exercise Frequency (p=.729)
 - b. Productivity Consistency:
 - i. The regression model identified Energy Fluctuations as a significant predictor (B = 0.829, SE = 0.170, t(37) = 4.88, p < .001), providing strong evidence to support the hypothesis.
 - ii. Exercise Frequency was marginally significant (B = 0.111, SE = 0.058, t(37) = 1.93, p = .064), suggesting weak evidence.
 - iii. Non-significant predictors included:
 - 1. Age (p=.308)
 - 2. Work/Study Hours (p=.684)
 - 3. Sleep Hours (p=.112)
 - 4. Consistent Sleep (p=.737)
 - c. Pearson Correlation Analysis:
 - Pearson correlation analysis revealed significant bivariate relationships, supporting the hypothesis for several predictors:

- ii. Daily Productivity:
 - 1. Energy Fluctuations (r = .543, p < .001)
 - 2. Work/Study Hours (r = .529, p < .001)
 - 3. Sleep Hours (r = -.436, p = .004)
 - 4. Consistent Sleep (r = .375, p = .038)
 - a. Productivity Consistency:
 - i. Energy Fluctuations (r = .674, p < .001)
 - ii. Consistent Sleep (r = .311, p = .048)

2. Type of Error:

- 1. Type I Error (False Positive):
 - a. A Type I error occurs when the null hypothesis is wrongly rejected.
 - b. The probability of a Type I error is controlled by the significance level (α = 0.05), which applies to both the regression and correlation analyses.
 - c. For significant results, such as work/study hours, sleep hours, energy fluctuations, and their corresponding correlations, there is a 5% chance of making a Type I error.
- 2. Type II Error (False Negative):
 - a. A Type II error occurs when a false null hypothesis is not rejected.
 - b. For non-significant predictors in the regression analysis (e.g., age, consistent sleep, exercise frequency), the likelihood of a Type II error increases due to limited statistical power (1β) , small effect sizes, and modest sample size (n = 37).

3. Chance of Error:

- 1. Type I Error:
 - a. The probability of a Type I error is set at α = 0.05 for all tests conducted in both the regression and correlation analyses.
- 2. Type II Error:
 - a. Type II errors are more likely for predictors with non-significant results in the regression model, such as consistent sleep and exercise frequency.
 - b. These errors may occur due to insufficient power or small effect sizes, as indicated by higher p-values, such as p = 0.131.

How It Was Determined:

- 1. Regression Analysis:
 - a. The significance level (α = 0.05) was used to evaluate each predictor in the regression model.
 - b. Significant results indicated a low likelihood of Type I error for predictors such as work/study hours, sleep hours, and energy fluctuations.
- 2. Pearson Correlation Analysis:
 - a. Bivariate relationships were explored using Pearson r, which provided additional insights into the predictors' relationships with the outcomes.

- b. Significant correlations confirmed the relationships for predictors like energy fluctuations and consistent sleep that were not consistently significant in the regression model.
- 3. Power of the Test $(1-\beta)$:
 - a. Power was not explicitly computed but inferred from the sample size and observed effect sizes.
 - b. Larger p-values (e.g., p=.308) and marginal significance (p=.064) suggest potential Type II errors due to insufficient power.

Conducting a power analysis to see the likelihood of finding an effect as above: Power Analysis:

- 1. Main Effect of Work/Study Hours on Daily Productivity:
 - Power: The power analysis for the main effect of work/study hours indicated a high power of 0.99, suggesting a robust likelihood of detecting the observed effect if it truly exists.
 - Effect Size (Cohen's d): d = 1.96
 - Effect Size (Pearson's r): r = 0.529
 - o Common Language Effect Size: 83.54%
- 2. Main Effect of Sleep Hours on Daily Productivity:
 - Power: The power for the main effect of sleep hours was high at 0.97, indicating a strong likelihood of detecting the observed effect if it exists.
 - Effect Size (Cohen's d): d = 1.96
 - Effect Size (Pearson's r): r = -0.436
 - o Common Language Effect Size: 78.14%
- 3. Main Effect of Energy Fluctuations on Productivity Consistency:
 - Power: The power for the main effect of energy fluctuations was substantial at 1.00, indicating a very low risk of Type II error.
 - Effect Size (Cohen's d): d = 2.31
 - Effect Size (Pearson's r): r = 0.674
 - o Common Language Effect Size: 89.91%
- 4. Interaction Effect of Evening Caffeine and Exercise Frequency on Daily Productivity:
 - Power: The power for the interaction effect of evening caffeine and exercise frequency was moderate at 0.78, indicating a 22% chance of a Type II error (1-Power = 0.221).
 - Effect Size (Cohen's d): d = 0.90
 - Effect Size (Pearson's r): r = 0.418
 - Common Language Effect Size: 81.19%

Interpretation:

- The high power for the main effects of work/study hours, sleep hours, and energy fluctuations indicates a strong ability to detect the observed effects, and the effect sizes suggest substantial practical significance (all Cohen's d > 1.0).
- For the interaction effect of evening caffeine and exercise frequency, the moderate power of 0.78 indicates a small risk of failing to detect a true effect. The effect size (d = 0.90) suggests it is moderately related to the interaction on daily productivity.
- The Common Language Effect Size adds an intuitive understanding of the practical significance, with values above 80% for the main effects and interaction indicating a meaningful proportion of non-overlapping distributions.
- The findings emphasize the importance of work/study hours, stable energy, and targeted caffeine use for productivity. Larger sample sizes are needed for better exploration of interaction effects.

Discussion:

The results of our study provide valuable insights into the factors influencing daily productivity and productivity consistency. The main finding reveals a significant effect of energy fluctuations on both daily productivity and productivity consistency, with a substantial effect size (d = 2.31). This suggests that stable energy levels are strongly associated with improved productivity outcomes. The observed effect size implies a large practical significance, indicating a meaningful contribution of energy fluctuations to consistent performance.

Work/study hours also showed a significant positive effect on daily productivity, with a large effect size (d = 1.96). This finding highlights the importance of structured time allocation in enhancing productivity. Conversely, sleep hours are negatively related to daily productivity (d = 1.96), suggesting a potential short-term trade-off between sleep and productivity, though this raises concerns about long-term sustainability.

Comparison Between Regression and Correlation Findings:

Notably, consistent sleep and exercise frequency did not emerge as significant predictors in the regression model, despite showing significant moderate correlations with productivity metrics in the Pearson correlation analysis. For instance, consistent sleep was positively correlated with daily productivity (r = .375, p = .038) and productivity consistency (r = .311, p = .048). This discrepancy suggests that while consistent sleep may play a role when examined in isolation, its impact is less pronounced when accounting for other variables in the regression model. This distinction highlights the added value of multivariate analyses in identifying independent effects of predictors.

Similarly, exercise frequency was marginally significant (p = .064) in predicting productivity consistency in the regression model but did not reach statistical significance in the correlation analysis. These differences underscore the importance of considering both bivariate and multivariate contexts when interpreting predictors' effects.

Limitations:

Several limitations of our study should be noted. First, the sample size (n = 37) was relatively small, potentially limiting the generalizability of the findings to broader populations. The reliance on self-reported data introduces biases, such as over- or under-reporting productivity and related behaviors. Additionally, the cross-sectional design prevents us from establishing causal relationships, making it unclear whether the predictors directly drive productivity or are consequences of it.

Furthermore, potential confounding variables, such as stress levels, dietary habits, and work environment, were not accounted for and may have influenced the results. Future studies should address these factors by including them as covariates in the analysis to better isolate the effects of the predictors.

Power Analysis and Interpretation:

The power analysis underscores the importance of interpreting non-significant results with caution. For most predictors, the power was high (Power > 0.95), suggesting that the lack of significance reflects a true absence of independent effects rather than insufficient sensitivity. However, the moderate power (Power = 0.78) for the interaction effect of evening caffeine and exercise frequency leaves a 22% chance of a Type II error, highlighting the need for further investigation with larger samples.

The significant predictors, such as energy fluctuations, work/study hours, and sleep hours, demonstrated robust effects in both the regression model and the correlation analysis, reinforcing their critical roles in productivity outcomes. In contrast, the lack of significance for consistent sleep and exercise frequency in the regression model suggests their contributions may be more context-dependent.

Future Directions:

To address these limitations, future research should:

- 1. Incorporate Objective Measures: Use wearable devices to track sleep patterns, energy levels, and physical activity to reduce self-report biases.
- 2. Account for Confounders: Include variables like stress levels, dietary habits, and workload as covariates to better isolate the predictors' effects.
- Adopt Longitudinal Designs: Conduct studies over time to establish causal relationships and assess the long-term sustainability of trade-offs between factors like sleep and productivity.
- 4. Increase Sample Diversity: Recruit larger and more diverse samples to improve generalizability and ensure that findings are representative of various populations.

Conclusion:

In conclusion, our study highlights the critical role of energy fluctuations and structured work/study hours in enhancing productivity. While some predictors, such as consistent sleep and exercise frequency, were not significant in the regression model, their potential contributions should not be dismissed, as suggested by their moderate correlations with productivity metrics. Future research should aim to overcome the limitations identified here to provide a more comprehensive understanding of the

complex dynamics underlying productivity and consistency. By doing so, researchers can better inform interventions to enhance productivity in both personal and professional contexts.

Type of Variables:

Independent Variables (IVs):

1. Continuous Variables:

a. Caffeine consumption frequency:

Participants will track the exact number of servings they consume each day.

b. Caffeine consumption timing:

Participants will track caffeine timing in hours.

c. Physical activity frequency in Hours:

This variable measures the frequency of Activity, measured as the total hours of exercise per week.

d. Physical activity timing:

Like caffeine consumption timing, this variable captures physical activity timing in hours after activity for a continuous scale.

e. Energy level Rating:

Energy Level Variations will be measured on a continuous scale from 0 to 100.

f. Sleep schedule consistency:

This variable assesses how consistent participants are with their sleep schedules; it tracks the number of days per week that the sleep schedule is consistent.

g. Hours of sleep per night:

This variable measures the average hours of sleep participants get each night.

2. Categorical Variables:

a. None (As we are working only with Multiple Regression Analysis and Pearson Correlation Analysis)

Dependent Variables (DVs):

1. Continuous Variables:

a. Daily productivity score:

Participants will rate their overall productivity for the day on a continuous scale from 0 to 100.

b. Time-specific productivity levels:

Measures productivity at different times (morning, afternoon, evening) on a continuous scale from 0 to 100.

c. Productivity Consistency:

Record productivity consistency as the number of days the productivity level is steady

2. Categorical Variables:

a. None (As we are working only with Multiple Regression Analysis and Pearson Correlation Analysis)

Several data pre-processing steps will be necessary to ensure the dataset is clean and ready for analysis. These steps are critical to ensure the integrity and quality of the data, ultimately affecting the validity of the results.

Preprocessing Steps:

1. Data Cleaning

Data cleaning involves identifying and correcting errors or inconsistencies in the dataset. For this study, potential issues might include:

- Outliers: Check for outliers in continuous variables such as hours of sleep or physical activity
 frequency. Outliers could result from participants entering incorrect or exaggerated values (e.g.,
 reporting 20 hours of sleep per night or 15 days of physical activity in a week). These values should
 be flagged and either corrected (if possible) or removed if they are erroneous.
- **Inconsistent Responses:** Ensure that all responses align with the continuous nature of the variables. For example, if a participant provides a categorical response for caffeine consumption (e.g., "Never"), this needs to be converted into a continuous measure (e.g., 0 cups of caffeine per day).
- Duplicate Entries: Check for duplicate responses from participants, ensuring no participant has submitted multiple surveys.

2. Handling Missing Data

Missing data is common in survey-based studies. It is important to decide how to handle missing values to avoid bias in the analysis. There are several strategies for dealing with missing data:

- **Imputation:** If there are missing values for continuous variables such as hours of sleep or physical activity frequency, imputation can be used. A common approach is:
 - Mean Imputation: Replace missing values with the mean value of that variable across all
 participants. This method works well when the amount of missing data is small and when the
 data is normally distributed.
 - Median Imputation: If the data is skewed (e.g., sleep hours), using the median instead of the mean may be more appropriate.
 - Mode Imputation: For categorical variables like caffeine consumption timing or physical activity timing, missing values can be replaced with the most frequent category (mode).
- **Listwise Deletion:** If a participant has too many missing responses (e.g., more than 50% of their survey is incomplete), their entire response may be removed from the dataset. This method should be used sparingly to avoid reducing the sample size too much.

3. Bucketing and Transformation Adjustments

a. Bucketing:

i. Caffeine Intake: To simplify analysis and capture usage patterns, caffeine intake can be bucketed into categories such as "Low," "Moderate," and "High," based on daily consumption frequency. These buckets will help in understanding productivity correlations for different intake levels.

- **ii. Sleep Duration:** Bucket hours of sleep into ranges, such as "Short" (0-4 hours), "Moderate" (5-7 hours), and "Optimal" (8+ hours), to analyze sleep's impact on productivity more distinctly.
- **iii. Physical Activity Frequency:** Physical activity could be categorized into "Sedentary," "Moderate," and "Active" based on weekly exercise hours. These buckets provide an easier way to examine productivity variations related to activity levels.

b. Transformation:

- i. Scaling: Normalize continuous variables like hours of sleep, caffeine intake, and physical activity using min-max scaling. This approach brings all variables to a common scale (0-1), ensuring regression and correlation analysis consistency.
- **ii. Log Transformation:** If data for caffeine intake frequency or productivity is skewed, a log transformation can be applied to reduce skewness and stabilize variance, which enhances the model's robustness.

c. Handling Outliers:

i. Bucketing as an Outlier Solution: For variables with extreme values, such as high caffeine consumption, bucketing also contains outliers within a specific range.

4. Converting Categorical Variables to Continuous

Since both Multiple Regression and Pearson Correlation require continuous variables, any categorical variables need to be transformed into continuous measures:

- Caffeine Intake: Instead of using categories like "Never," "Rarely," or "Daily," caffeine intake should be recorded as a continuous variable based on the number of caffeinated beverages consumed per day (e.g., cups of coffee, tea, energy drinks). If participants have provided categorical responses, these can be converted into approximate daily consumption values (e.g., "Never" = 0 cups/day, "Rarely" = 0.5 cups/day, etc.).
- Physical Activity Frequency: Convert physical activity frequency into total hours per week rather
 than days per week. This provides a more precise measure of physical activity that can be treated as
 continuous.
- **Sleep Consistency:** If sleep schedule consistency was originally measured on an ordinal scale (e.g., "Very Consistent" to "Very Inconsistent"), this needs to be transformed into a continuous metric. For instance, participants could report how many days per week they maintain a consistent sleep schedule (0–7 days).

5. Normalization/Scaling

For continuous variables such as hours of sleep and physical activity frequency, normalizing or scaling the data may be necessary before conducting regression analysis, especially if these variables have widely different ranges.

• Min-Max Scaling: This technique scales all continuous variables to a range between 0 and 1, ensuring that no variable is disproportionately related to the results due to its scale.

• **Z-score Normalization:** This method standardizes continuous variables by subtracting the mean and dividing by the standard deviation, resulting in a distribution with a mean of 0 and a standard deviation of 1.

Normalization is particularly important when running algorithms like regression that assume all predictors are on comparable scales.

6. Addressing Survey Response Bias

Since this study relies on self-reported data, there may be biases such as social desirability bias or recall bias. While these biases cannot be eliminated through pre-processing, it's important to:

- Check for Response Patterns: Identify any participants who may have given identical responses across all questions (e.g., selecting "Neutral" for every item), which could indicate careless responding.
- **Cross-check Responses:** For questions that should logically align (e.g., caffeine intake frequency and timing), cross-check responses for consistency.

Conclusion

In summary, the pre-processing steps will include:

- 1. Cleaning data by addressing outliers, inconsistencies, and duplicates.
- 2. Handling missing data through imputation or listwise deletion.
- 3. Bucketing and Transformation Adjustments
- 4. Converting categorical variables into continuous measures where necessary.
- 5. Normalizing/scaling continuous variables if necessary.
- 6. Checking for response patterns to mitigate survey biases.

These steps will ensure that the dataset is clean and ready for robust statistical analysis, allowing us to accurately explore relationships between habits like caffeine intake, physical activity, sleep quality, and productivity among university students.

Multiple Regression Analysis and Pearson Correlation Analysis:

Hypotheses for Test Statistics

With the help of t-tests, we will isolate each of the relations in the regression model, determining the strength of each of the independent variables about the dependent one, as well as employing multiple regression analysis and Pearson correlation analysis to determine how three independent variables related to the dependent variable.

Multiple Regression Analysis and T-tests Hypotheses

Multiple regression analysis shall be carried out assisted with t-tests for each of the independent variables in order to test whether or not each of them is significantly related to productivity provided that other variables are already controlled.

1. Caffeine Intake (IV1)

- a. Null Hypothesis (H0): under this hypothesis, the regression coefficient for caffeine intake $\beta 1$ is equal to zero which means that caffeine intake does not significantly contribute towards enhancing daily productivity in relation to physical activity and sleep quality.
 - i. $H0:\beta 1=0$
- b. Alternative Hypothesis (H1): under this hypothesis, the regression coefficient for caffeine intake $\beta 1$ is not equal to zero which means that caffeine intake significantly contributes towards augmenting one's productivity levels on a daily basis in relation to physical activity and sleep quality.
 - i. H1:β1!=0
- 2. Physical Activity (IV2)
 - a. Null Hypothesis (H0): the regression coefficient for physical activity, $\beta 2$, is equal to a zero which implies that physical activity is not viewed as an important determinant of daily productivity level after the caffeine dosage and the quality of sleep has been considered.
 - i. H0:β2=0
 - b. Alternative Hypothesis (H1): the regression coefficient for physical activity, β 2, is not equal to a zero and this implies that physical activity is practiced on all days as a contributor to the daily productivity level even with the consideration of the caffeine dosage and the quality of sleep.
 - i. H1:β2!=0
- 3. Sleep Quality (IV3)
 - a. Null Hypothesis (H0): Sleep quality has a coefficient, β 3, that is equal to zero. This means that when we factor in the caffeine dosage and physical activity, sleep quality is not one of the key factors explaining the daily productivity level.
 - i. H0:β3=0
 - b. Alternative Hypothesis (H1): The regression coefficient for sleep quality (β 3)is not equal to zero, meaning sleep quality has a significant effect on daily productivity after accounting for caffeine intake and physical activity.
 - i. $H1:\beta 3!=0$

Pearson Correlation Analysis Hypotheses

We will also conduct a Pearson correlation analysis to assess the relationship between each independent variable (IV) and the dependent variable (DV).

- 1. Correlation between Caffeine Intake and Productivity
 - a. Null Hypothesis (H0): There is no significant correlation between caffeine intake and productivity.
 - i. H0:r=0
 - b. Alternative Hypothesis (H1): There is a significant correlation (either positive or negative) between caffeine intake and productivity.
 - i. H1:r!=0
- 2. Correlation between Physical Activity and Productivity

- a. Null Hypothesis (H0): There is no significant correlation between physical activity and productivity.
 - i. H0:r=0
- b. Alternative Hypothesis (H1): There is a significant correlation between physical activity and productivity (either positive or negative).
 - i. H1:r!=0
- 3. Correlation between Sleep Quality and Productivity
 - a. Null Hypothesis (H0): There is no significant correlation between sleep quality and productivity.
 - i. H0:r=0
 - b. Alternative Hypothesis (H1): There is a significant correlation between sleep quality and productivity (either positive or negative).
 - c. H1:r!=0

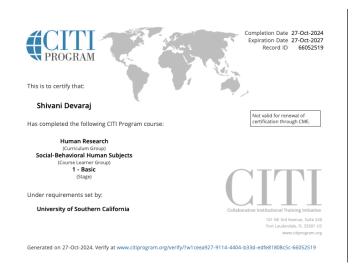
The analysis focuses on two key components:

- 1. Overall predictive relationships through multiple regression
- 2. Individual correlations between continuous variables

This comprehensive analysis approach will allow us to:

- Identify which habits have the strongest relationships with productivity
- Understand how the timing of different activities relates to productivity patterns
- Determine the role of consistency in habits and its relationship with consistent productivity
- Account for both continuous and categorical variables in our dataset

By examining these relationships through the above-mentioned statistical approaches, we can provide robust insights into how caffeine intake, physical activity, and sleep quality correlate with productivity among university students.



Screenshot of G*power:

