Human Cognitive Performance Data Analysis

Cachary Tolentino

Stockton University

April 12, 2025

# Abstract

Human cognition is a crucial portion of a well-functioning individual. Many factors play a role in the deterioration or nourishment of cognitive performance. Thus, it is important to understand and research what these factors are. In the growing age of technology, cognitive performance has been dramatically affected positively or negatively by constant screen time and media consumption (Stenberg et al., 2013). This research focuses on uncovering how habitual activities such as diet, exercising, and sleep may affect cognitive function. Furthermore, to analyze these issues, we approach the problem via performing statistical analysis on a real-world dataset, called “Human Cognitive Performance Analysis” by Samx\_sam from Kaggle, containing various information of individuals and their health statistics. Through this experiment, we can analyze and predict the possibility of cognitive performance as well as demonstrating the usage of statistical methods as predictors.

# Introduction

Human Cognitive Performance is a metric in evaluating the ability to process, understand, and respond to information. This metric varies from individual to individual primarily affected by their lifestyle. Lifestyles can greatly influence cognitive performance positively when monitored and balanced. But it can also deteriorate this performance.

A common problem in the modern age workforce is the damage in circadian rhythms due to overnight work (Chellappa et al., 2018). For example, according to Chellappa et al., “a cognitive slowing was observed under circadian misalignment with median reaction times of ~300 ms when assessed 11 h after scheduled awakening” (2018). This shows that even during work, especially for some, cognitive performance can greatly influence daily life. This is crucial, especially for growing children, as media consumption and screen time have overall increased which can affect their learning and attention. Lifestyles are only one of the few factors that can deteriorate cognitive performance, many other factors play a role such as eating habits, exercise, and sleep.

Through this study, we will analyze various health information of multiple individuals collected as a large dataset. We will perform multiple statistical methods such as: histograms to visualize an overall distribution of health information, probability of having certain cognitive performance score given a condition, and much more. This study’s goal is to perform as many statistical methods as possible in various question formats to understand human cognitive performance and the effects of certain factors in determining their scores.

# Methods

Note: Explain every detail on how the experiment was performed (so you can talk about the programs needed, the knowledge needed, etc..) to replicate the paper. Basically, here you can write about all the questions you are trying to answer and what you need for each one to be answered. REWORD FURTHER CAN’T FULLY COPY ONE TO ONE; MAKE IT MORE COGNITIVE RELATED

Questions List:

1. (CH1.1, 1.2) Construct a relative frequency histogram (PG.6, 1.3)
2. (CH1.3) Find the mean, media, mode, variance, and standard deviation (original)
3. (CH1.4) sleep duration for individuals are approximately distributed with mean (FIND VALUE) and standard deviation (FIND VALUE). What fraction of all individuals have sleep duration in the following intervals? (PG.11, 1.9)
   1. #.# to ##.# (measurement)
   2. #.# to ##.# (measurement)
   3. #.# to ##.# (measurement)
   4. Less than #.# or more than ##.# (measurement)
4. (CH2.1, 2.2, 2.3) An individual has noted their gender, diet type, and exercise frequency, Let F denote female, M for male. We can denote NVG for non-vegetarian, VGN for vegetarian, and VEN for vegan. Then we can also denote L for low exercise frequency, M for medium, and H for high. Construct a space S. Find A, the following subset of possibilities containing no vegetarians, B the subset containing two vegetarians, and C, the subset containing at least one vegetarian. List the element of A, B, C, A B, A B, A C, A C, B C, B C, and C . (PG.25, 2.1)
5. (CH2.4) The proportion of coffee intake, 0-99, 100-199, 200-299, 300-399, and 400-499, in the population are approximately 20.231, 19.903, 19.871, 20.000, and 19.995, respectively. A single individual is chosen at random from the population. (PG.32, 2.10)
   1. List the sample space for this experiment
   2. Make use of the information given above to assign the probabilities to each of the simple events
   3. What is the probability that the person chosen at random has either 0-99 or 200-299 coffee intake?
6. (CH2.5) A group of individuals contains six people. Two of the six are to be randomly selected to be classified as cognitively healthy or unhealthy.
   1. If two of the people are unhealthy, find the probability that at least one of the two people checked are unhealthy. Find the probability that both are unhealthy.
   2. If four of the people are unhealthy, find the probabilities indicated in part (a).
7. (CH2.6) Due to the influx of volunteers for the research, a raffle with 8000 tickets were handed, one per volunteer. There are only three positions needed to be filled. If four of the researchers were also given one ticket each, what is the probability that the four organizers win (PG.49, 2.51)
   1. All of the prizes?
   2. Exactly two of the prizes?
   3. Exactly one of the prizes?
   4. None of the prizes?
8. (CH2.7) An individual is randomly selected from the population of 8000. (PG.56, 2.75)
   1. If the first two individuals picked have a screen time duration of 7-9, what is the probability that the next three individuals also have the same screen time?
   2. If the first three individuals have a screen time duration of 7-9, what is the probability of the next two people having the same screen time
   3. If the first four people have a screen time duration of 7-9, what is the probability that the next person will also have the same screen time?
9. (CH2.8) If A, being the age between 18 to 19, and B, being the age between 20 to 21, are such that . Find the following: (PG.60, 2.96)
10. (CH2.9, 2.10) Male (and others) and Female were observed to have varying cognitive performance scores. It shows that 70% of females had high cognitive scores, 40% of males and others had high cognitive scores. A group of 20 people, 15 female, and 5 males, were subjected to a test to check their cognitive performance. A response picked at random from the 20 was negative. What is the probability that it was that of a male or other? (PG.74, 2.129)
11. (CH3.1, 3,2) A problem in the cognitive test was given to the volunteers. The task was to match three pictures of animals to the word identifying that animal. If a participant assigns the three words at random to the three pictures, find the probability distribution for Y, the number of correct matches. (PG.90, 3.5)
12. (CH3.3) Let Y be a random variable p(y) given in the accompanying table. Find E(Y), E(1/Y), E(Y^2 – 1), and V(Y) (PG.97, 3.12)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| y | 1 | 2 | 3 | 4 |
| P(y) | .4 | .3 | .2 | .1 |

1. (CH3.4) In an experiment with the volunteers, coffee was studied to see any correlation with cognitive performance and caffeine intake. Two types of coffee were presented: formula A (with a much lower caffeine concentration), and formula B (with a higher caffeine concentration). Four participants were selected, each given three cups of coffee in random order. Two contains formula A and the other containing formula B. Each participant was asked which of the cups made them focus more. Suppose that the two formulas are equally performant. Let Y be the number of participants stating a preference for formula B.
   1. Find the probability distribution function for Y
   2. What is the probability that at least three of the four participants state a preference for formula B.
   3. Find the expected value of Y
   4. Find the variance of Y
2. (CH3.5) Of the population of participants, 60% have high stress levels (8+). If a group of randomly selected participants is asked, what is the probability that exactly five people have to be interview to encounter the first participant who has a high stress level? At least five people? (PG.120, 3.78)
3. (CH3.6) During analysis, it was found that 25% of participants had a memory score from 400 to 500. (PG.124, 3.96)
   1. If you were to randomly ask a participant, what is the probability that they would have more or less of a memory score on the first try? The second try? The third try?
   2. If you wanted to speed up the analysis, you have yourself and your coworker ask a participant, what is the probability that a total of four tries will be necessary for the both of you to find someone not having a 400-500 memory score?
4. (CH3.7) At the current moment, ten participants have been fully examined. Four have been found to be Non-vegetarian. The researchers selects five participant records. What is the probability that all five participants were vegetarian or vegan? (PG.128, 3.103)
5. (CH3.8) The amount of sleep a participant gets has a Poisson distribution with an average of seven hours per day. If more than seven hours were slept in a day, the participant scores a higher cognitive performance. What is the probability that a randomly selected participant will not score a higher cognitive performance? (PG.136, 3.127)
6. (CH3.11) Amongst the 8000 participants, the average cognitive score was 60-70 with standard deviation of 10. Using tchebysheff’s theorem, find a lower bound for the number of participants from a 500 sample expected to have a cognitive score between 50 to 80. (PG.148, 3.170)
7. (CH4.2) Let Y be a random variable with p(y) given in the table below. (PG.166, 4.1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| y | 1 | 2 | 3 | 4 |
| P(y) | .4 | .3 | .2 | .1 |

* 1. Give the distribution function, F(y). Be sure to specify the value of F(y) for all y, .
  2. Sketch the distribution function given in part (a).

1. (CH4.3) If, Y has a density function

Find the mean and variance of Y (PG.172, 4.20)

1. (CH4.4) AI predicted scores is uniformly distributed over the interval 70 to 80 points. What is the probability that the score exceeds 75 points if it is known that scores exceed 72 points? (PG.177, 4.51)

# Results

Note: Here write all the answers to the questions above. Include any notable work, figure, etc…

According to the….

# Discussion

Note: Here is where you explain all the answers to from the above section. Make sure you tie in to the main point of the paper, maybe some limitation (if any) of the results.

Based on the results…

# Conclusion

Note: Here is kinda just use it as a reflection of some kind. Like what did you learn overall, etc…

# References

Chellappa, S. L., Morris, C. J., & Scheer, F. A. J. L. (2018). Daily circadian misalignment

impairs human cognitive performance task-dependently. *Scientific Reports,* 8(1), 3041–

11. <https://doi.org/10.1038/s41598-018-20707-4>

Samx\_sam. (2025). Human Cognitive Performance Analysis. *Kaggle.com.*

<https://www.kaggle.com/datasets/samxsam/human-cognitive-performance-analysis>

Sternberg, D. A., Ballard, K., Hardy, J. L., Katz, B., Doraiswamy, P. M., & Scanlon, M. (2013).

The largest human cognitive performance dataset reveals insights into the effects of

lifestyle factors and aging.*Frontiers in Human Neuroscience,*7.

<https://doi.org/10.3389/fnhum.2013.00292>

IMRAD (Introduction, Methods, Results and Discussion) | Science Writing Resources (new).

(n.d.). *Scwrl.ubc.ca.*<https://scwrl.ubc.ca/stem-writing-resources/features-of-academic-stem-research-writing/imrad/>

(2019). *Thevisualcommunicationguy.com*. [https://thevisualcommunicationguy.com/writing/how-](https://thevisualcommunicationguy.com/writing/how-to-organize-a-paper/how-to-organize-a-paper-the-imrad-format/)

[to-organize-a-paper/how-to-organize-a-paper-the-imrad-format/](https://thevisualcommunicationguy.com/writing/how-to-organize-a-paper/how-to-organize-a-paper-the-imrad-format/)