
Caffeine and Timed Up and Go Cognitive Test

Oliver Siu^{a,c}, Nicholas Cassol-Pawson^{a,c,d}, Emma Vidal^{a,b,c}, Romy Guo^{a,c}, Kanzah Jamil^{a,b,c}, Max Chalekson^{a,c}, Jesper White^{a,c,d}

^aUniversity of California, Los Angeles (UCLA); ^bDepartment of Mathematics; ^cDepartment of Statistics & Data Science; ^dDepartment of Economics

Group 4 | S24-LEC2-STATS 101B: Introduction to Design and Analysis of Experiment | Dr. Maria Cha

1. Introduction

Caffeine is one of the most widely consumed drugs worldwide, known for its stimulating effects on the central nervous system. It is commonly believed that caffeine enhances mental performance and physical energy, making it a popular choice in various consumer products, but research shows mixed evidence of caffeine's performance-boosting capabilities. Caffeine has varying effects on different individuals and may not always improve mental capabilities such as attention [2] (Anderson, 2000, p. 18). Depending on the task, caffeine may or may not improve mental performance [5] (Stephan et al., 2021, p. 1381).

Verifying the exact benefits and limitations of caffeine remains under study and is important for determining when caffeine consumption is most appropriate and effective.

1.1. Approach

Previous research suggests that caffeine may improve general alertness but does not have a pronounced effect on performance when it comes to more complicated and demanding tasks. So, we expect caffeine to improve performance when carrying out light physical and mental tasks.

To test this hypothesis, we used a Timed Up and Go Cognitive test to assess coordination and mental performance in working-age adults (between the ages of 20 and 65). We wished to see if taking caffeine could improve their performance in tasks that demand physical abilities and mental focus.

The Timed Up and Go Cognitive test seems perfect for this: it asks participants to rise from a chair, walk 3 meters, turn around, walk back to the chair, and sit down, all while counting backward by 3 from 72, which targets both their ability to concentrate on a somewhat complex mental task and their general coordination. Participants' times for completion are recorded in seconds, and a faster time indicates better performance. Tests were administered at 2:00 p.m. on a Monday, and subjects were chosen from the Islands Project, a population of virtual human subjects created by Dr. Michael Bulmer at the University of Queensland. Informed consent was obtained from all participants.

2. Methods

2.1. Design of Experiment

During our preliminary design of the experiment, we were concerned about two separate situations: the climate of the subjects' home islands impacting the rate of the caffeine intake of the subjects and the age of the subjects causing them to have different baseline scores on the tests and intake caffeine in different manners. We opted to block over these two variables.

There are three islands and we defined three age groups of 15 years each: 20-34, 35-49, and 50-65. Since we had two blocked factors of 3 levels each, we chose to run a replicated Latin square design. This meant that we would need three levels for our caffeine dosage. We chose to run the experiment with dosage levels 0, 100, and 200 mg of caffeine.

Since there are two blocking factors at three levels each, each replicate required nine runs of the experiment. We chose to do seven replicates of the experiment, so we generated seven random 3×3 Latin square designs. Our data and copies of the Latin squares are included in the Appendix A.

2.2. Sampling

We then randomly selected our subjects. Since we were blocking by island, we first randomly chose a village within the current island block with a random number generator. Then, within the village, we randomly selected a house and, if that house had a subject in the age category under investigation, we would ask them for consent to participate in the study. If there were multiple people who fell within the correct age bracket, we would randomly choose one of them. If we were denied consent, we would resample the house in the village until we got consent from someone in the age bracket.

This sampling technique is a limitation of our experimental design, as it is not simple random sampling. Since each village has a different number of people, people living in smaller villages have a higher probability of being drawn. This is an inherent limitation of the island system, as there was no centralized roster of people on the island from which to draw the blocked sample.

2.3. Sample Size and Power of the Test

Balanced one-way analysis
of variance power calculation

```
k = 3
n = 21
f = 0.3133475
sig.level = 0.05
power = 0.5751066
```

NOTE: n is number in each group

Figure 1. R Output of Power

After our experiment, we had data to calculate the power of our test (Figure 1). We computed an effect size of $f = 0.313$ by finding the difference between the largest and smallest sample means, which is the largest difference in them, by definition. We then estimated the standard deviation with the standard error, which we computed by taking the square root of the mean square error, giving a value of $\hat{\sigma} = 0.532$. Plugging all of these numbers into `pwr.test` in R, we found that a sample size of $n = 63$ gives a power of ≈ 0.58 . This is not satisfactory as it means that there is a $1 - \text{power} = \beta = 42\%$ chance of accepting a false null hypothesis (making a Type II error).

Balanced one-way analysis
of variance power calculation

```
k = 3
n = 33.72766
f = 0.3133475
sig.level = 0.05
power = 0.8
```

NOTE: n is number in each group

Figure 2. R Output of Recommended Observations

An optimal test power would be around 80%. To detect a maximum difference in mean test times of $f = 0.313$ seconds, we calculated a suggested $n = 33.7$ which recommends $n = 34$ (Figure 2). Given $n = 34$ observations per $k = 3$ treatment groups, a sample size of $N = 102$ is recommended.

2.4. Model and Hypotheses

The population model is

$$y_{ijk} = \mu + \alpha_{i(l)} + \tau_j + \beta_{k(l)} + \rho_l + \epsilon_{ijkl}$$

for $i, j, k \in 1, 2, 3$ and $l \in 1, 2, \dots, 7$

- y_{ijk} is the change in time to complete the Timed Up and Go Cognitive test for the i^{th} island and k^{th} age group for the j^{th} caffeine treatment;
- μ is the average difference in time to complete the cognitive test across all observations, before and after the caffeine intervention;
- $\alpha_{i(l)}$ is the j^{th} island block effect within replicate l ;
- τ_j is the j^{th} caffeine effect;
- $\beta_{k(l)}$ is the k^{th} age block effect within replicate l ;
- ρ_l is the l^{th} replicate effect; and
- ϵ_{ijkl} is the random error

The objective is to assess the impact of different dosages of caffeine on the Timed Up and Go Cognitive test scores of individuals blocked by their age groups and geographic locations.

Therefore, we wish to test the hypotheses:

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

$$H_1 : \tau_j \neq 0, \text{ at least one } j$$

3. Results and Interpretation

3.1. ANOVA Results

	Df	Sum Sq	Mean Sq
factor(caffeine)	2	0.341	0.1706
factor(age)	2	0.748	0.3740
factor(island)	2	0.177	0.0883
factor(user)	6	1.113	0.1855
Residuals	50	14.145	0.2829

	F value	Pr(>F)
factor(caffeine)	0.603	0.551
factor(age)	1.322	0.276
factor(island)	0.312	0.733
factor(user)	0.656	0.685
Residuals		

Figure 3. R output of ANOVA

Figure 3 shows $p = 0.551 > \alpha = 0.05$, so we fail to reject the null hypothesis at a 5% significance level. This suggests that caffeine does not significantly influence this measure of cognitive performance.

3.2. Diagnostic Plots

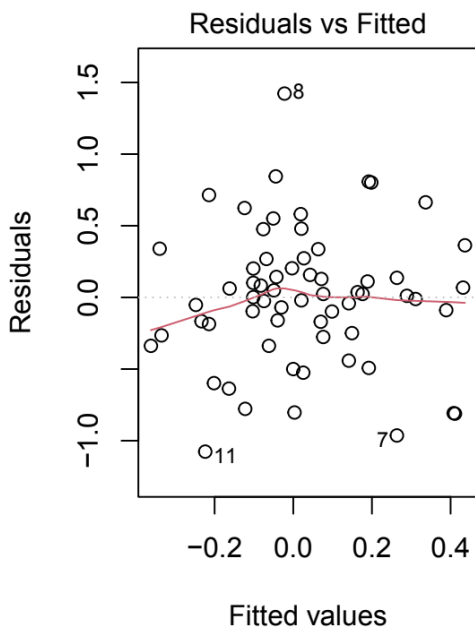


Figure 4. *Residuals vs. Fitted Diagnostic Plot*

In Figure 4, the Residuals vs. Fitted Values plot shows no obvious patterns, and the scattering along the horizontal axis is similar to the scattering along the vertical axis, indicating homoskedasticity, i.e., the constant variance assumption is satisfied.

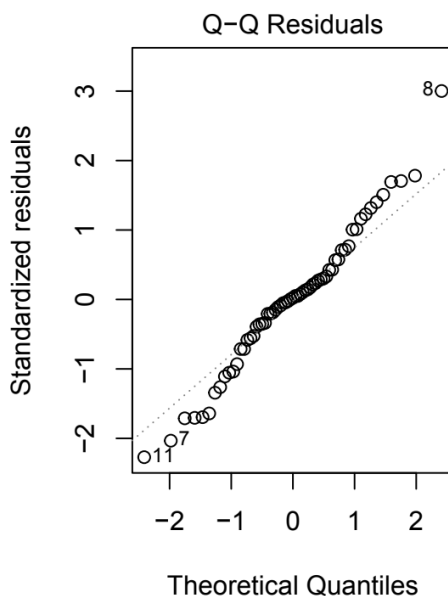


Figure 5. *Normal Q-Q Diagnostic Plot*

In the normal Q-Q plots for the residuals (Figure 5), the points lie approximately along the reference line, indicating that the residuals are normally distributed.

Both of the diagnostic plots suggest that the assumptions of the model are reasonably met.

3.3. Block Analysis

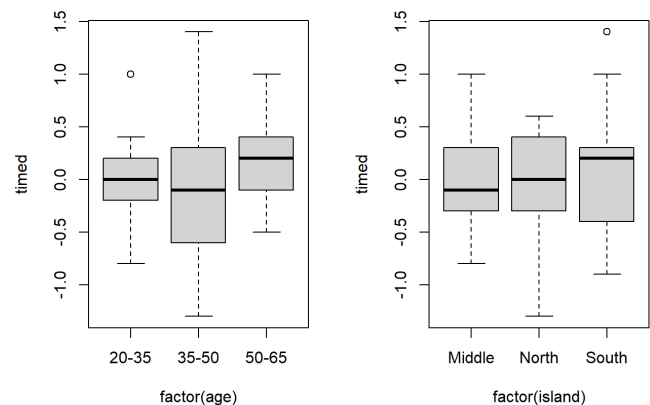


Figure 6. *Plots of Block Means*

The ANOVA results cannot be used to discuss the significance of the block effect, as the ANOVA process is not providing estimates for the blocking factors, rather it is considering them to be random variations. However, we can compare the mean values across the blocks to try to determine whether they were necessary.

We plotted the mean values across the 3 blocks for age group and the 3 blocks for the islands in paired box plots.

Looking at the age group results in Figure 6, we can see that the means are all very similar, with a range of 0.2: spanning from an average difference of -0.1 seconds in the 35-49 age group to 0.1 in the 50-65 age group. The total range of the observed differences is just under 3 seconds, with a minimum difference before and after the intervention in time to take the Timed Up and Go test of -1.4 and a maximum difference of 1.5. Since the mean range is only about 3% of the total range of the times, controlling for age group likely did not have much impact on our final results, so pooling the data would have been fine.

Similarly, we look at the island blocks in Figure 6. Here we see that there is a slightly larger range for the mean difference in time to take the Timed Up and Go test, with the highest average difference of 0.3 on the South island and the lowest average difference of -0.1 on the Middle island. This leads to a range of 0.4 seconds, just over one tenth of the total range of 3 seconds. This indicates that the island block probably also did not have much of an impact on the final results of our report.

3.4. Interpretation of Results

Our results suggest that caffeine intake does not significantly affect the time to complete the Timed Up and Go Cognitive test. Despite the expectation that caffeine might enhance cognitive performance, our study found a

p -value of 0.551 on this effect, indicating that we should accept the null hypothesis that there is no statistically significant effect of caffeine on the Timed Up and Go Cognitive coordination scores at a 5% level.

These findings highlight the complexity of factors that play into cognitive and motor skills and suggest that factors other than caffeine intake might play more important roles. Further research could explore different cognitive performance measures or consider other variables that might interact with caffeine consumption to affect cognitive outcomes. The diagnostic plots confirm that the ANOVA assumptions were met, which supports the validity of our findings.

4. Discussion and Conclusion

4.1. Summary

Our study focused on investigating the effects of caffeine on performance in light physical and mental tasks and allowed us to understand that caffeine products are not effective on cognition. To do this, we designed and tested subjects using a 3×3 Latin Square Design experiment, with caffeine dosages of 0 mg, 100 mg, or 200 mg of caffeine. We then tested their coordination and mental capabilities using the Timed Up and Go Cognitive test, both before and after administering the caffeine intervention.

In our analysis, we accounted for the variability introduced by island and age group by incorporating them as blocking factors. Our hypothesis sought to determine whether our treatment factor, caffeine dosage, had a significant effect on participants' cognitive and coordination abilities. The findings for this experiment show that caffeine does not significantly influence the performance of the tasks completed among young and middle-aged adults at the 5% significance level. This result challenges the belief that caffeine enhances cognitive performance and physical ability, requiring further investigation to fully understand the drug's effect on these parameters.

4.2. Applying Our Results to the Real World

This study utilizes the concept of optimizing the timing of caffeine consumption and its implications for an individual's cognitive and physical performance. Understanding such scenarios allows us to reap maximum benefits of this drug in real-world scenarios.

To start, considering the peak level of caffeine effectiveness occurs around the one-hour mark after consumption [1] (Liguori, 1997, p. 724), individuals can strategically time their caffeine intake to coincide with tasks that require heightened awareness and focus. By aligning caffeine consumption with the timing of a task that requires cognitive processing, individuals may capitalize on the alertness provided by caffeine, improving their performance on tasks that require one's attention and quick

reaction-time.

With the benefits of caffeine, it is important to understand that the effect of caffeine varies amongst individuals, influenced by factors such as, but not limited to, genetics, tolerance levels, and sensitivity to the drug [4] (James, 2014, p. 118). Therefore, customized approaches to caffeine consumption for each individual allows them to optimize the drug's benefits, while minimizing the adverse effects.

Our findings suggest caffeine's effect on cognitive performance may not be consistent across all tasks [2] (Anderson, 2000, p. 10), for complex tasks, it is deemed uncertain as to how caffeine directly aids in those tasks that have many integral parts. Thus, further research should be completed to uncover more of the effects of caffeine on different cognitive aspects, to allow individuals to make decisions on how to optimize their productivity.

Our findings reflect some real-world experiments. An experiment that looked at the effects of having either a 250 mg dose of caffeine or a placebo on a reaction test suggested that there was no significant impact of caffeine dosage in either direction on the participant's concentration [3] (Ruijter, 2000, pg. 29).

4.3. Limitations and Future Work

There were several limitations underlying our study that could have led to us not finding any significant causal relations between caffeine and mental performance. As we mentioned earlier, one of the main limitations is due to our sampling method. It was not simple random sampling, rather it involved choosing subjects randomly within blocks, which may have introduced experimental bias. We followed this technique as we were constrained by the Islands system's lack of a centralized population census, making it difficult to implement SRS within blocks. Individuals from smaller villages had a higher probability of being included in our sample, which affected how representative our sample was of island population as a whole.

We believe that there is a second limitation that arose from an inherent difficulty with the Islands system: that the Islanders may not be programmed for caffeine consumption to directly impact their performance in the Timed Up and Go test. This is entirely plausible: the block effects that we detected did not seem to be significant, although since we didn't test for this in a statistically significant manner, they may actually be. However, since the blocks had such a small effect, we are led to believe that regardless of which island or age group an Islander comes from, the amount of caffeine they consumed did not seem to affect their performance on the cognitive and motor skills test that we ran on them. This stands in contrast to the literature we reviewed when devising our original hypothesis, which indicated that we should

detect an improvement in these skills with an increase in caffeine dosage.

In the future, we can make a few adjustments to our experiment to try to detect some effect between caffeine consumption and mental and/or physical performance.

We could try to conduct a stronger form of sampling for the blocks to ensure that our sample population was chosen by SRS. This could increase the significance of the effects that we are looking for. However, as we noted before, even in our imperfectly sampled model, it is unlikely that the blocks are actually meaningful.

We could also try expanding our subject pool to cover a larger range of ages: our experiment only looked at adults in young and middle age. However, there is a chance that caffeine consumption actually has a statistically significant effect on the cognitive and motor skills of older adults or younger children, and just young and middle-aged adults are unaffected by it.

A third limitation was our limited sample size. The experiment utilized a replicated Latin square design with each of our group members as a replicate. This restricted our sample size to $N = 63$ which produced a test power below an adequate level. Optimally, the power of the test should be around 80% which would require a sample size of $N = 102$, as indicated in Figure 2. Due to time constraints, we were unable to increase the replicates or alter the experiment design to accommodate a larger sample size. Future tests should include more observations to increase test power and reduce the chance of making a Type II error. Since the block and replicate effects were not significant, a different design may also be recommended to increase the sample size to achieve optimal test power and improve the accuracy of future experiments.

Finally, we could try running different tests on our subjects than the Timed Up and Go Cognitive test. As we indicated before, there is literature that suggests that there may be an effect between caffeine consumption and mental and physical performance. It could be that the Timed Up and Go cognitive test is not ideal for this situation, but the Islands project offers a whole host of tests that can be run that seem to target the skills that we are testing for, including a variety of logic and focus tasks, many different forms of light and heavy exercise, and several coordination-targeted activities.

References

- [1] A. Liguori, J. R. Hughes, and J. A. Grass, "Absorption and subjective effects of caffeine from coffee, cola and capsules", *Pharmacology, Biochemistry and Behavior*, vol. 58, no. 3, pp. 721–726, 1997. DOI: [10.1016/S0091-3057\(97\)00003-8](https://doi.org/10.1016/S0091-3057(97)00003-8).
- [2] A. Anderson, "Caffeine effects on different levels of attention", Publication no. 11942, Honors thesis, Southern Illinois University, 2000. [Online]. Available: https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1007&context=uhp_theses.
- [3] J. Ruijter, M. Lorist, J. Snel, and M. De Ruiter, "The influence of caffeine on sustained attention: An ERP study", *Pharmacology Biochemistry and Behavior*, vol. 66, no. 1, pp. 29–37, 2000. DOI: [10.1016/S0091-3057\(00\)00229-X](https://doi.org/10.1016/S0091-3057(00)00229-X).
- [4] J. James, "Caffeine and cognitive performance: Persistent methodological challenges in caffeine research", *Pharmacology, Biochemistry, and Behavior*, vol. 124, pp. 117–122, 2014. DOI: [10.1016/j.pbb.2014.05.019](https://doi.org/10.1016/j.pbb.2014.05.019).
- [5] M. E. Stepan, E. M. Altmann, and K. M. Fenn, "Caffeine selectively mitigates cognitive deficits caused by sleep deprivation", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 47, no. 9, pp. 1371–1382, 2021. DOI: [10.1037/xlm0001023](https://doi.org/10.1037/xlm0001023).

A. Appendix

A.1. Data

[Here](#) is a Google Spreadsheet with the data we collected. This file was exported to a .csv to analyze the data in R.

A.2. Latin Squares

- A, B, and C correspond to the 0 mg, 100 mg, and 200 mg caffeine treatments, respectively.
- The numbers after the equal correspond to the measured change in test score for Timed Up and Go Cognitive.
- The rows correspond to the age range blocks (20-34, 35-49, and 50-65).
- The columns correspond to the island block (Ironbark, Providence, and Bonne Santé)

A = 0.2	B = 1	C = -0.7
B = 0	C = -0.5	A = 1.4
C = 0.4	A = -0.1	B = -0.4

Figure 7. *Latin Square 1*

A = -0.3	C = 0.4	B = 0.2
B = -1.3	A = -0.7	C = 0.5
C = 0.6	B = 0.2	A = -0.5

Figure 8. *Latin Square 2*

A = -0.4	C = 0.1	B = -0.2
C = 0.5	B = -0.1	A = -0.4
B = 0.1	A = -0.4	C = 0.2

Figure 9. *Latin Square 3*

A = 0.1	B = 1	C = -0.4
C = -0.3	A = 0.3	B = 0.3
B = 0.5	C = 0.8	A = 0.3

Figure 10. *Latin Square 4*

C = 0	B = -0.8	A = 0.2
A = -0.6	C = -0.8	B = 0.8
B = 0	A = -0.2	C = 1

Figure 11. *Latin Square 5*

B = 0.4	A = -0.1	C = 0.2
C = -0.2	B = 0.5	A = -0.9
A = 0.5	C = -0.3	B = 0.3

Figure 12. *Latin Square 6*

C = 0	B = -0.1	A = -0.1
B = -0.8	A = 0	C = 0
A = 0.1	C = -0.1	B = -0.3

Figure 13. *Latin Square 7*