

Stats 101B Final Report  
Experiment Design  
Team: ÷ & Conquer  
Maria Barrera, Shreya Singhal, Yu Zhang, Xin Jin, Yaxi Lou

## Contents

1	Abstract	2
2	Introduction	2
3	Research Questions	2
4	Experiment Design	2
5	Data & Variables	3
6	Analysis	5
6.1	Part I . . . . .	5
6.2	Part II . . . . .	7
7	Conclusion	8
8	Future Study	8
9	R code	9

# 1 Abstract

Since 1847, the year Joseph Fry invented the first chocolate bar, chocolate has become an ambrosial treat: “Americans will eat nearly 18 percent of the world’s chocolate confectionery by value in 2015 or around \$18.27 billion worth.”[1] Similarly, coffee has become a prominent part of society: “Americans crave their daily fix and drink an estimated 400 million cups a day.”[2] While there are several reasons for such mass consumption, the experiment presented here focuses on the potential benefits of such items on an individual’s mental performance. After all, it is a popular belief among the student population that chocolate can serve as an energy boost right before an exam, while coffee has the ability to enhance mental awareness. Therefore, the aim is to test whether these notions are constructed fallacies or if there is sufficient statistical evidence to support their validity.

## 2 Introduction

The objective was to determine whether an individual’s performance on a difficult four-minute mental arithmetic task could be improved using coffee and chocolate. Based on prior research, it was anticipated that the subject’s score would improve by at least ten percentage points after being administered the coffee and chocolate in conjunction with one another. In this light, data collection of the scores before and after administration of the treatment was completed in approximately two consecutive weeks.

## 3 Research Questions

Based on the experimental design utilized, which will be described in detail below, the following questions guided our research path:

1. If chocolate is a significant factor, is there a specific type of chocolate that gives uniformly higher test performance, regardless of coffee type?
2. If coffee is a significant factor, is there a specific type of coffee that gives uniformly higher test performance, regardless of chocolate type?

## 4 Experiment Design

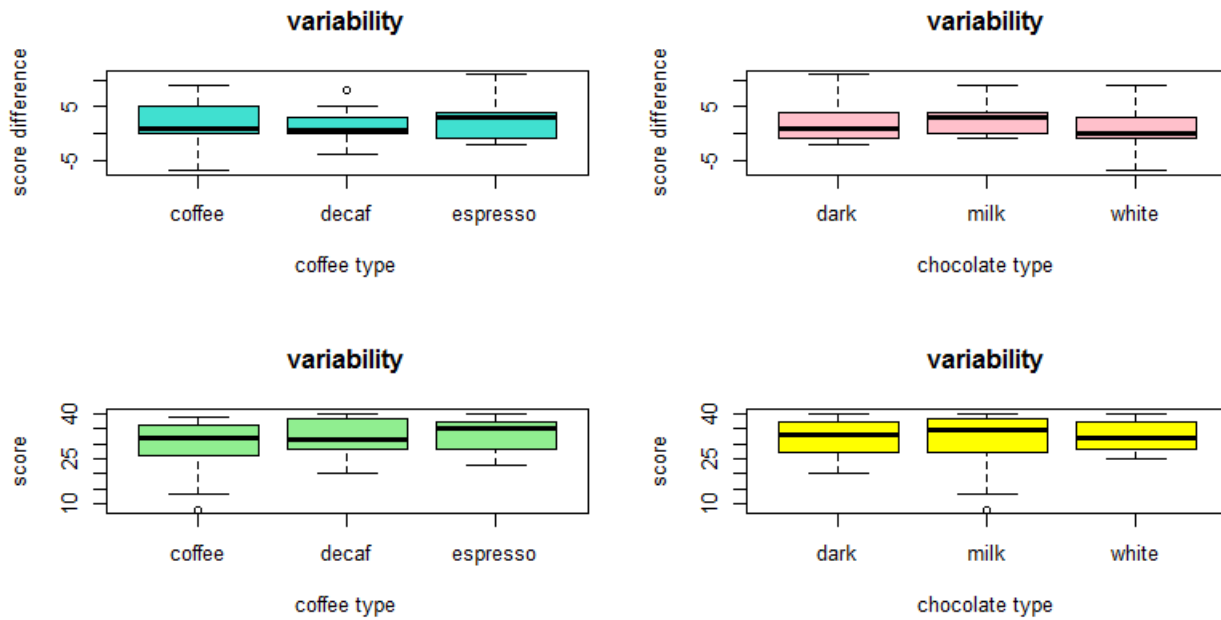
Because factorial designs allow an investigator to study the effects of each factor alone on the response variable and the effects of these factors simultaneously on the response variable, the use of a factorial design was deemed to be the ideal choice. Although there are alternative, more complex designs, such as the Latin Squares design, the restrictions

implemented during the randomization of our subjects made blocking unnecessary, and therefore, a simple design would suffice for our purposes. Additionally, a factorial design would be efficient in our case, as it would allow for the combination of a series of studies to be run, rather than several independent studies on their own. Since the aim was to examine the main effects of chocolate and coffee, as well as their interaction effects, on the task score, a  $2^3$  factorial design was chosen, where the two factors were coffee and chocolate, each comprised of three levels. Some uncontrolled factors such as the current health, drinking habits, smoking habits and intelligence levels of the subjects may have had an effect on the research results; however, in the name of simplicity and due to a lack of time and resources, these uncontrolled factors were not taken into consideration.

## 5 Data & Variables

In order to collect the data, 54 randomly chosen individuals from Vardo Island falling in the age range from 18 to 45 were required to take a difficult four-minute mental arithmetic test in the first phase of the experiment. The experimenter documented this score, which was equal to the amount of questions answered correctly out of the total forty questions. A corresponding treatment (one randomly chosen from below) was administered to the subject about 30 minutes later. Then, after an hour-long wait, the experimenter would readminister the test to the individual, in order to determine whether the treatment had a positive, negative or neutral effect on his/her score. Because we ultimately ended up incorporating two response variables into our experiment: `score_difference` (part I) and `score_after` (part II), our data analysis is split into two parts.

A factorial design requires that all possible combinations of the levels of the factors are investigated in each replicate. In order to maximize the reliability of the model, it was necessary to run multiple replicates for each treatment combination. Additionally, replication makes the detection of smaller effects more likely or increases the power in order to detect an effect of fixed size. In our case, we used six replicates, which equated to a total of 54 runs ( $9 \text{ treatments} \times 6 \text{ replicates}$ ). Moreover, the boxplots for the data (presented below) confirm that there is enough variability in the means of coffee type and chocolate type to use only six replicates. It is important to note that only 4 replicates were used initially; however, increasing the number of replicates to 6 did not lead to any significant differences in the results of the ANOVA table. In other words, in this specific case, further increasing the number of replicates, and thus the sample size, would not have altered our conclusions.



In this experiment, blocks were intentionally not utilized. For instance, while gender could have served as grounds to block off of, because a factorial design is already completely randomized, blocking by gender would not have produced any significant results. Furthermore, age could have also been a blocking variable, but the age restriction requirement imposed on each subject made blocking by age unnecessary.

Factors: chocolate and coffee (2)

Levels of chocolate: white, milk and dark (3)

Levels of coffee: decaf, regular and espresso (3)

Variables: chocolate\_type, coffee\_type, score\_difference (the difference in scores recorded before and after the treatment), and after\_score (score recorded after the treatment).

Treatments	Chocolate Type	Coffee Type
1	white	decaf
2	white	regular
3	white	espresso
4	milk	decaf
5	milk	regular
6	milk	espresso
7	dark	decaf
8	dark	regular
9	dark	espresso

Subject Name & House Number	Group Member	Treatment Number (1-9)	Coffee Type Administered	Chocolate Type Administered	Score Before Treatment	Score After Treatment (~1 hr)	Difference in Score
Elin (482)	Yaxi	6	Espresso 60 ml	milk 50g	32	32	0
Gerik (546)	Maria	2	Coffee 250 ml	white 50g	31	24	-7
Paul (138)	Shreya	2	Coffee 250 ml	white 50g	40	40	0
Halle (666)	Yaxi	2	Coffee 250 ml	white 50g	23	27	4
Leon (443)	Shreya	2	Coffee 250 ml	white 50g	37	39	2
Erika (70)	Yaxi	2	Coffee 250 ml	white 50g	31	30	-1
Sorena (138)	Maria	1	Decaf 250 ml	white 50g	29	29	0
Astrid (450)	Maria	1	Decaf 250 ml	white 50g	33	38	5
Hiro (45)	Shreya	2	Decaf 250 ml	white 50g	37	37	0
Torsten (296)	Yaxi	1	Decaf 250 ml	white 50g	32	32	0
Haruka (619)	Yaxi	1	Decaf 250 ml	white 50g	32	37	5
Akane (822)	Shreya	1	Decaf 250 ml	white 50g	39	39	0
Ingrid (559)	Xin	1	Decaf 250 ml	white 50g	37	33	-4
Jakob (226)	Yu Zhang	3	Espresso 60 ml	white 50g	20	23	3
Hallmar (700)	Yu Zhang	3	Espresso 60 ml	white 50g	39	38	-1
Brenda (854)	Xin	3	Espresso 60 ml	white 50g	38	36	-2

## 6 Analysis

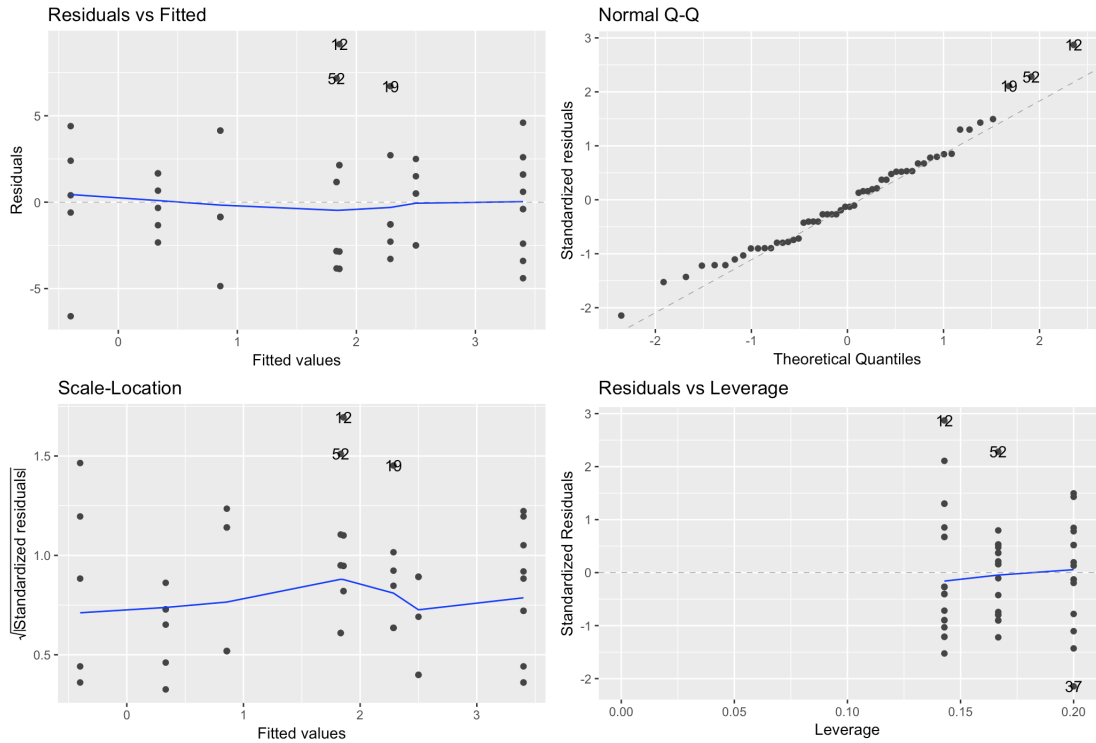
### 6.1 Part I

Table 1: treatment vs score\_difference  
ANOVA Table

	DF	Sum Sq	Mean Sq	F value	p value
coffee	2	4.2	2.087	0.176	0.839
chocolate	2	29.0	14.524	1.228	0.303
coffee:chocolate	4	40.2	10.060	0.850	0.501
Residual	45	532.4	11.831	—	—

Using the difference in score as our response variable, our ANOVA table suggests that coffee and chocolate alone do not have a significant effect on one's mental arithmetic score ( $p - \text{values} = 0.839, 0.303 > 0.05$ ). Our results also indicate that consuming coffee and chocolate simultaneously doesn't have a notable effect on one's mental arithmetic score ( $p - \text{value} = 0.501 > 0.05$ ).

Although the factor effects and interaction effect are statistically insignificant, the model satisfies the regression conditions.



The lack of pattern in the residual plot implies that the model satisfies the condition of linearity. Also, the Normal QQ plot shows that the residuals adhere to the straight line, thus satisfying the condition of normality.

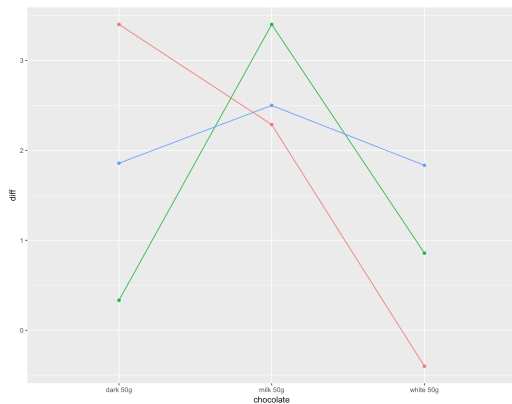


Figure 1: Interaction Plot 1

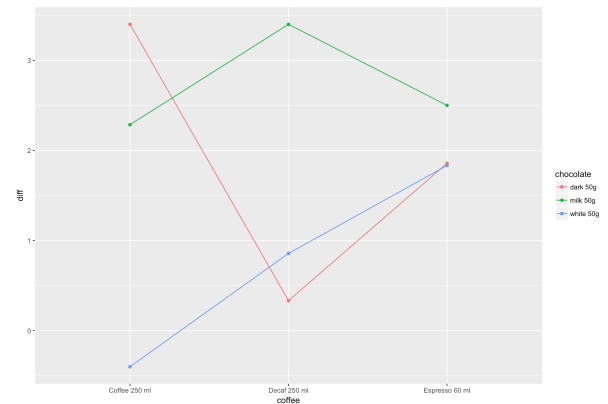


Figure 2: Interaction Plot 2

Nonetheless, despite the results from the ANOVA table, the interaction plots show that the interaction between chocolate type and coffee type is indeed significant due to the intersection of lines. Moreover, the highest mean score difference is yielded from dark chocolate with decaf coffee. Due to the discrepancy between the ANOVA table and interaction plots, the score after the treatment, instead of the difference in scores, was used as our response variable in part II.

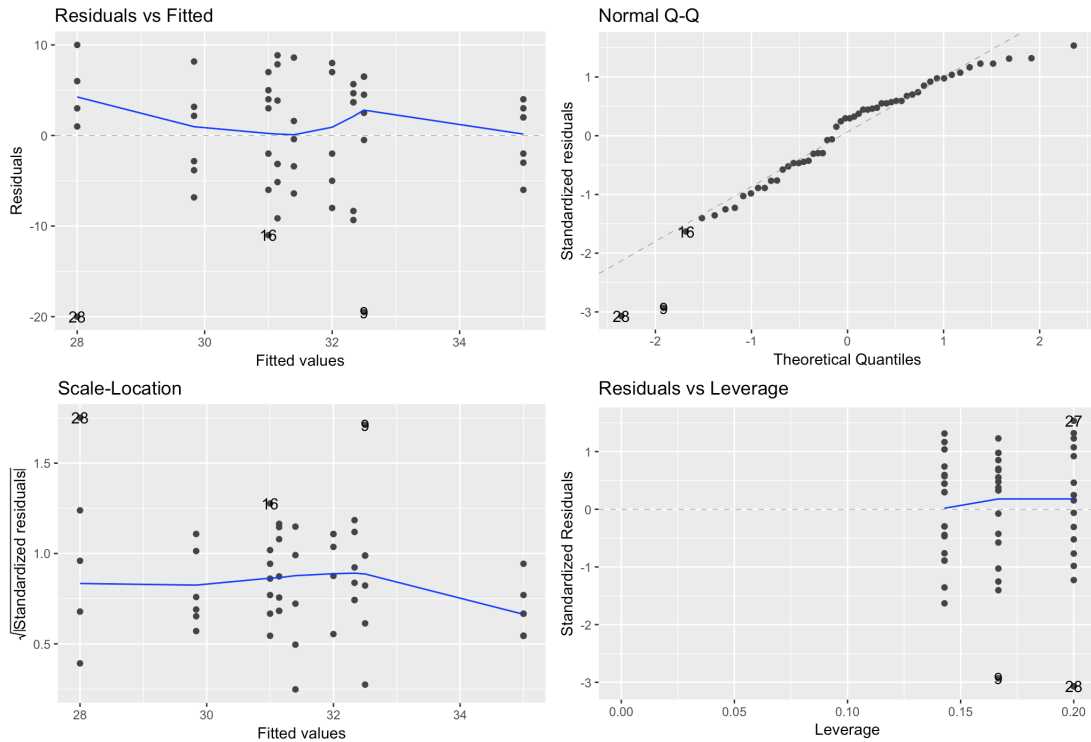
## 6.2 Part II

The new response variable is the score obtained on the difficult arithmetic task after the treatment was administered.

Table 2: treatments vs score\_after  
ANOVA Table

	DF	Sum Sq	Mean Sq	F value	p value
coffee	2	124.7	62.33	1.539	0.2258
chocolate	2	16.6	8.28	0.204	0.8159
coffee:chocolate	4	600.9	150.22	3.708	0.0108 *
Residual	45	1823.1	140.51	—	—

Using the after score as our response variable, our ANOVA table suggests that coffee and chocolate alone do not have a significant effect on one's mental arithmetic score ( $p - values = 0.2258, 0.8159 > 0.05$ ). On the other hand, the simultaneous consumption of coffee and chocolate has a significant effect ( $p - value = 0.0108 < 0.05$ ). As suspected, part II confirms the impression that the interaction effect between coffee and chocolate is indeed significant and thus affects, mostly positively, an individual's mental arithmetic task score.



The lack of pattern in the residual plot implies that the model satisfies the condition of linearity. Also, the Normal QQ plot shows that the residuals adhere to the straight line, thus satisfying the condition of normality.

## 7 Conclusion

The comparison of the two ANOVA tables confirms that they yield different results depending on the response variable used. Nonetheless, both models show that the interaction between coffee and chocolate is significant. If the score difference is used, there does not appear to be any significant interaction, according to the ANOVA table; however, the interaction plots reveal contradicting results. If we use the score after the treatment as our response variable instead, then the interaction between coffee and chocolate is deemed to be statistically significant, which was suspected in part I and verified in the ANOVA results of part II.

In this light, it is interesting to note that according to our data (parts I & II), the consumption of coffee and chocolate on their own do not affect individuals' mental capacity, but the combination of the two factors does influence an individual's score. In other words, among all three types of coffee and chocolate, consuming any one alone does not have a notable effect when it comes to altering one's test performance. However, a combination of both factors, particularly decaf coffee with milk chocolate and regular coffee with dark chocolate, causes an increase in one's arithmetic test score.

## 8 Future Study

Instead of using a mental arithmetic test as our only measurement of mental performance, we would like to use a more diverse and comprehensive range of response variables in order to collect a more precise measure of one's mental capacity. For example, we could utilize memory games and problem solving tasks, as it is possible that these activities are more holistic representations of an individuals mental capacity.

Due to the time constraints, the results of the experiment are based on the short term effects seen in the subjects. Thus, the experiment could have been improved by having the islander drink coffee or/and eat chocolate at the same more than once before performing the arithmetic task. Therefore, if the treatment was to be applied more often, it would make it possible for us to reconfirm our results with more conviction.

Additionally, although our subject pool comprised of 54 randomly chosen individuals from an island made up of 1000 individuals, all subjects originated from the same island. For this reason, another method of improvement could be the selection of participants from a larger pool, i.e. various islands. By doing so, our conclusions would have grounds for being more universal, as they would apply to a larger, more diverse population of islanders.



## 9 R code

```
library(ggplot2)
library(ggfortify)

Island <- read.csv("The Island Project Data Sheet - Sheet1.csv")
island <- Island[, c(1,3:8)]
names(island) <- c("names", "treatment number", "coffee", "chocolate", "score_before",
"score_after", "diff") lm1 <- lm(score_after ~ factor(coffee) + factor(chocolate) +

factor(coffee)*factor(chocolate), data = island)
autoplot(lm1)

lm2 <- lm(diff ~ factor(coffee) + factor(chocolate) + factor(coffee)*factor(chocolate), data
= island)
autoplot(lm2)

summary(aov(diff ~ factor(coffee)+factor(chocolate)+factor(coffee)*factor(chocolate),data
= island))
summary(aov(score_after ~ chocolate+coffee+chocolate*coffee, data = island))

ggplot(data = island, aes(x = chocolate, y = diff, color = coffee,group = coffee)) +
stat_summary(fun.y = mean, geom = "point") + stat_summary(fun.y = mean, geom =
"line")

ggplot(data = island, aes(x = coffee, y = diff, color = chocolate,group = chocolate)) +
stat_summary(fun.y = mean, geom = "point") + stat_summary(fun.y = mean, geom =
"line")
```

## References

- [1] Satioquia-Tan, Janine. *"Americans Eat HOW MUCH Chocolate?"* CNBC. 23 July 2015. Web.
- [2] Wapner, Scott. *"The Coffee Addiction."* CNBC. 29 Nov. 2012. Web.