

STAT 403 Project

Intermediate Sampling and Experimental Design

**Title: Analysis of effect of Coffee on
Memory Retention**

Presented to: Liangliang Wang

Prepared by: Sharjeel Ahmad

Table of Contents:

- 1. Abstract**
- 2. Problem**
- 3. Data Collection**
- 4. Randomness and Sample Size**
- 5. Research Question**
- 6. Basics of Experimental Design**
- 7. Histograms and Summary Statistics Table**
- 8. Initial Model and ANOVA (Before Transformation)**
- 9. Checking Conditions for ANOVA (Before Transformation)**
 - a. Independence**
 - b. Constant Variance**
 - i. Levene Test**
 - c. Normality**
 - i. Normal Quantile Plot**
- 10. Transformations**
- 11. Checking Conditions for ANOVA (After Transformation)**
 - a. Independence (After Transformation).**
 - b. Constant Variance (After Transformation).**
 - i. Levene test**
 - c. Normality (After Transformation)**
 - i. Normal Quantile plot**
- 12. Final Model and ANOVA**
- 13. Post Hoc Tukey HSD**
- 14. Conclusion**
- 15. Surprising Result**
- 16. References**

Abstract:

The presence of Caffeine in Coffee helps you stay awake, but it can also have some negative effects on your health. One of these negative effects is on the memory retention. For this experiment Adults living in different cities of North Island were given different amounts Coffee and a memory test was performed before they were given coffee and after they were given Coffee. We will use different statistical methods to check whether or not consumption of Coffee adversely effects on memory retention of adults on the North Island.

Key words: Coffee, Memory Retention, Treatment groups, North Island, Conditions

Problem:

Consumption of coffee can have a detrimental effect on memory retention. This study will explore truth of the aforementioned statement.

Data Collection:

For data collection adults living in different cities of North Island were chosen randomly using the website <https://theislands.umn.edu/login.php> and results were collected in csv file. Initially, they were made to look at the list of 20 words for one minute. After one minute they recalled how many words they remembered out of 20. The first group of adults were given 60 ml of coffee espresso, the second were given 120 ml of coffee espresso, and the third (control) group was given 60 ml of water. All the treatment groups were then shown a list of 20 words again for one minute and were told to remember as many words as possible. Finally, the difference in the words remembered before and after consumption of coffee were calculated. The data was collected in a csv file and will be uploaded separately on Dropbox link.

Randomness and Sample Size:

The sample was drawn using the Islands Sampler and selecting for only those individuals who lived on the North island. 150 samples were drawn however i only included the first 30 individuals to consent to our study in each treatment group. Our final sample size was 90 individuals (30 in each treatment group).

Research Question:

Does the Increase in coffee intake adversely affects memory in adults on the North Island?

Null Hypothesis: Consumption of Coffee has detrimental effect on the retention of memory.

Alternate Hypothesis: Consumption of Coffee does not have detrimental effect on the retention of the memory.

Target Population: Residents on the North Island

Study Population: Adults on the North Island that consent to be in my study

Experimental Unit: An Adult

Response Variable: Change in the number of words remembered after drinking coffee

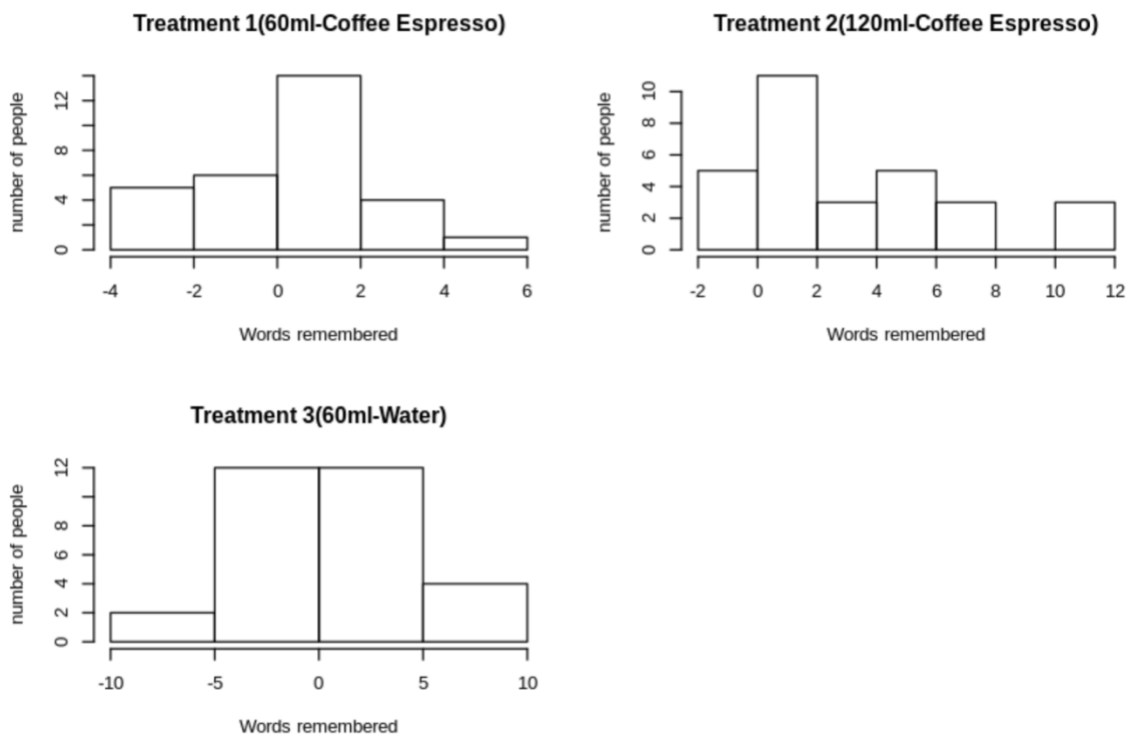
Factor: Over here factor serves as the Categorical explanatory variable that identifies the experimental treatment groups. We have 3 treatment groups.

1. Adults who have been give 60 ml of Coffee Espresso.
2. Adults who have been given 120 ml of Coffee Espresso.
3. Adults who have been give 60 ml of Water (Control).

Histograms and Summary Statistics Table:

Histograms for the treatment groups show that Treatment 3 kind of follows the Normal distribution but Treatment 1 and Treatment 2 do not follow the Normal distribution. This sort of distribution gives a hint that some transformation has to be applied to the data.

The summary statistics table below helps us understand the data with the help of measures of central tendency (Mean, Mode and Median) and gives an idea about Minimum, Maximum and Standard deviation values in each treatment group.



Summary Statistics Table:

	SampleSize	Mean	Median	SD	Min	Max
	<int>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
Coffee Espresso 120ml	30	3.7333333	2	3.657145	-2	12
Coffee Espresso 60ml	30	0.8333333	1	2.182743	-4	6
Water 120ml	30	0.8666667	1	3.692801	-7	10

Initial Model Fit and ANOVA:

We will fit the model initially and first verify if the conditions for the applying ANOVA are met or not.

Initial Model Fit:

The results obtained after fitting the model are given below

Call:

```
lm(formula = data$Change ~ data$DType, data = data)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-7.8667 -1.7333  0.1333  1.1667  9.1333
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      3.7333     0.5942   6.283 1.27e-08 ***
data$DTypeCoffee Espresso 60ml -2.9000     0.8403  -3.451 0.000864 ***
data$DTypeWater 120ml      -2.8667     0.8403  -3.411 0.000982 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.255 on 87 degrees of freedom

Multiple R-squared: 0.1529, Adjusted R-squared: 0.1334

F-statistic: 7.85 on 2 and 87 DF, p-value: 0.0007343

ANOVA Table:

The results obtained from ANOVA are given below.

```
              Df Sum Sq Mean Sq F value    Pr(>F)
data$DType     2  166.3    83.14    7.85 0.000734 ***
Residuals     87  921.5    10.59
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Checking Conditions:

Independence:

Independence is met for our model because our sample was a simple random sample and we assume that none of the individuals in our sample are related. Subjects were chosen using the Island Sampler and by randomly sorting the subjects into excel.

Constant Variance:

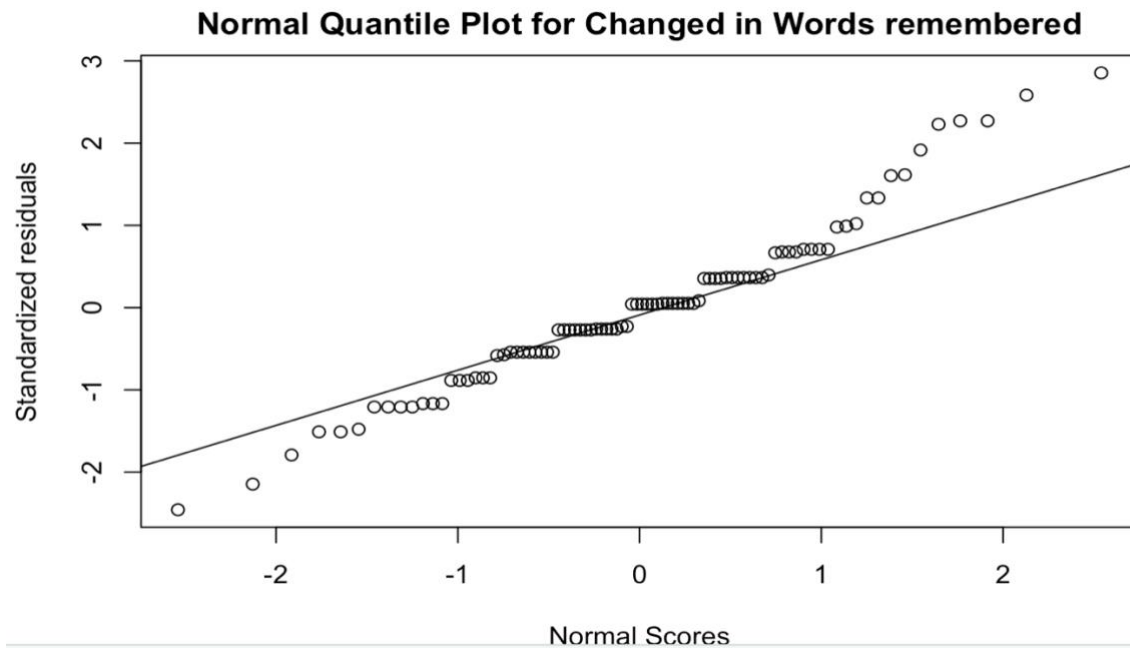
We use **Levene test** for checking the condition of Constant Variance. Since the p-value (0.1228) obtained from levene is greater than 0.05 then according to the levene test the condition of constant variance is met.

Levene's Test for Homogeneity of Variance (center = median)

```
              Df F value Pr(>F)
group     2    2.1482 0.1228
          87
```

Normality:

We use the Normal Quantile plots to check Normality. The Plot below shows that the points are close to the line but not close enough to declare it Normal. We might have to apply some suitable transformation to achieve Normality.

**Transformation:**

- The inverse transformation ($1/(\text{response-variable} + \text{constant})$) was applied to achieve both Normality and Constant-Variance.
- The constant term was added just because in our data sometimes the change in Time is 0. When we do $1/0$ then it leads to error as it cannot be calculated. i.e $1/0 = \text{Error}$. The addition of constant will make sure that we do not have zero terms in our data.

Checking Conditions (After Transformation):

In order to apply ANOVA, we need to make sure that all the conditions for ANOVA are met.

Independence (After Transformation):

The independence condition is already met.

Constant Variance (After Transformation):

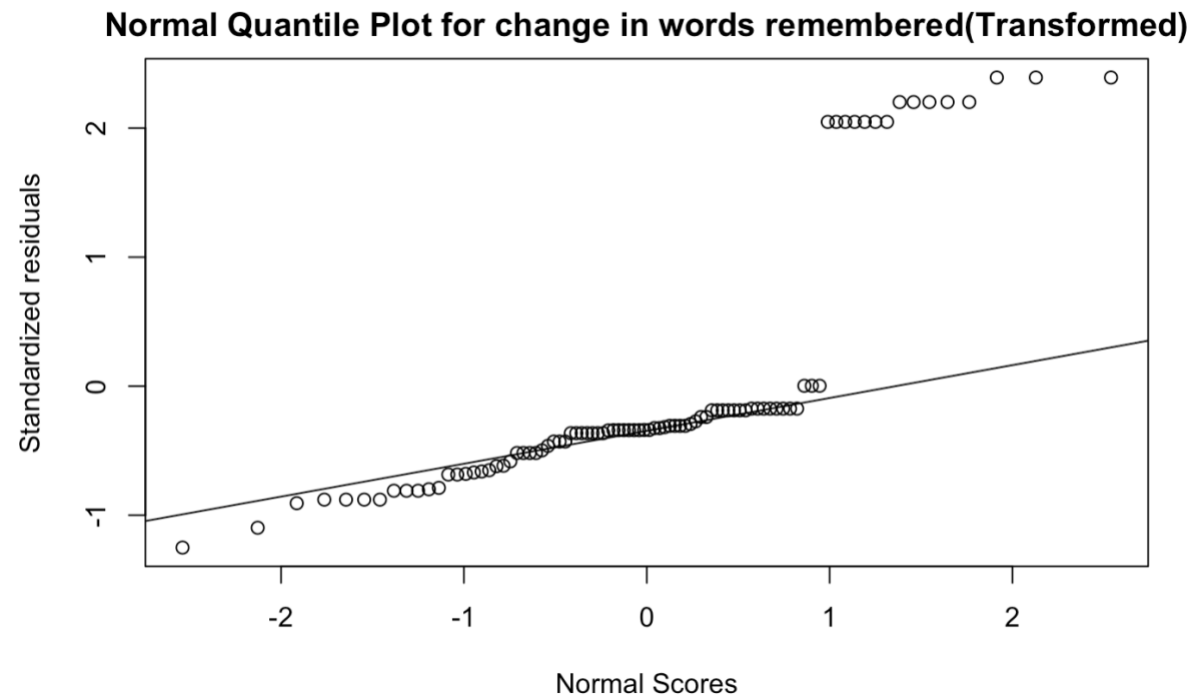
P-value (0.279) greater than 0.05 shows that the constant variance condition is met.

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	2	1.2953	0.279
	87		

Normality (After Transformation):

The points are clustered closer to the line but there are some major outliers. Since we have applied all the suitable transformations and chosen the best one, we are going ahead with the assumption that the data is fairly Normal.

**ANOVA Table (after transformation):**

Over here the p-value (0.426) is greater than 0.05 which that there is a difference in the means of the treatment groups and NULL hypothesis will not be rejected. More details about conclusion in the end of the report.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
data\$DType	2	9.1	4.529	0.861	0.426
Residuals	87	457.8	5.262		

Tukey HSD:

The ANOVA table only shows that there is difference in group means but it does not show which groups differ and for that we will use Tukey HSD. This will also help us to reduce family wise error rate.

```
[1] "Using tukey HSD to reduce family-wise error rate"
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
data$DType  2    9.1   4.529   0.861  0.426
Residuals  87 457.8   5.262
```

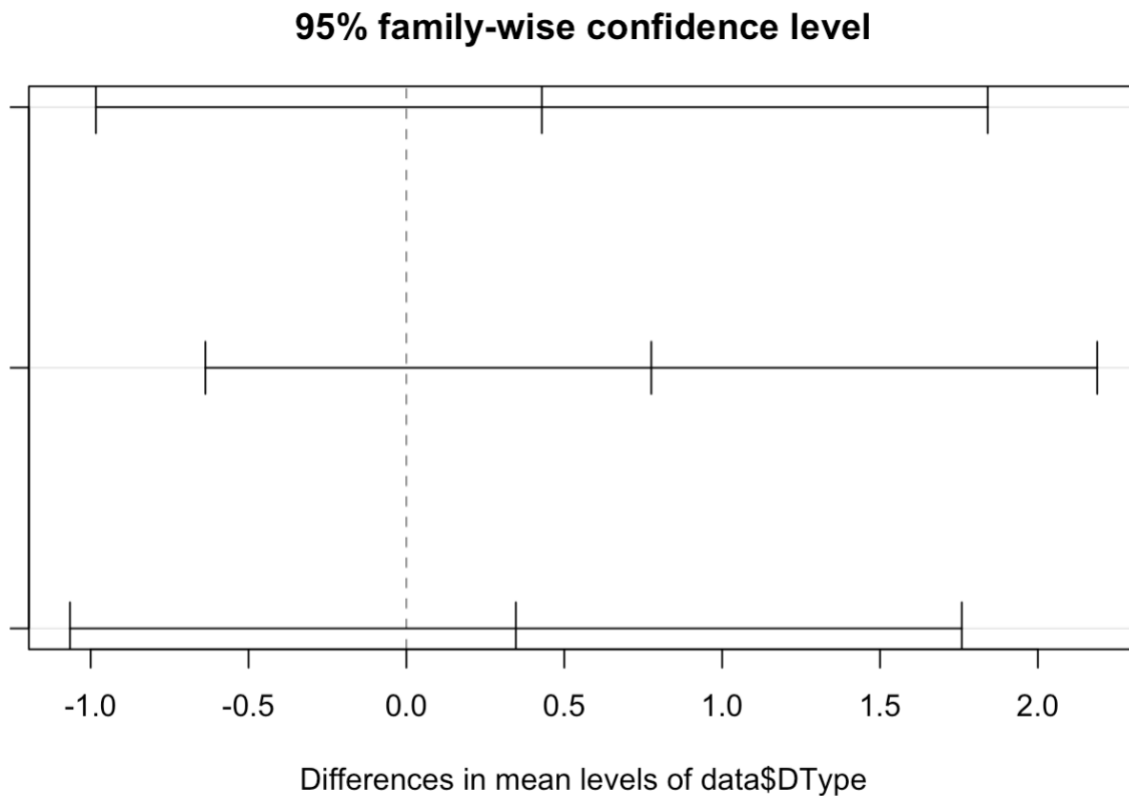
```
Tukey multiple comparisons of means
95% family-wise confidence level
```

```
Fit: aov(formula = 1/(data$Change + 0.16) ~ data$DType, data = data)
```

```
$`data$DType`
```

```
      diff      lwr      upr      p adj
Coffee Espresso 60ml-Coffee Espresso 120ml 0.4288985 -0.9834147 1.841212 0.7498963
Water 120ml-Coffee Espresso 120ml          0.7756371 -0.6366762 2.187950 0.3937384
Water 120ml-Coffee Espresso 60ml           0.3467385 -1.0655747 1.759052 0.8282836
```

Not significant if CI includes the zero reference line



All possible pairwise comparisons were performed, and p-values were adjusted using Tukey's HSD. The treatment group 1(Coffee Espresso-60ml) and the group 2(Coffee Espresso-120ml)) is not statistically significant as the p-value (0.74) is greater than the 0.05. Similarly, other comparisons of treatment groups are not statistically significant either. The figure above confirms that since the confidence interval for all the three comparisons include the zero-reference line, so they are not statistically significant.

Conclusion:

The ANOVA gives a p-value (0.426) greater than 0.05 so we do not reject the NULL Hypothesis and conclude that consumption of Coffee does not have detrimental effect on the retention of the memory.

Surprising Result:

Initially, when started the experiment I thought that result will prove that of Coffee has detrimental effect on the retention of the memory because it has large amount of Caffeine in it but turns out that it does not a detrimental effect on the memory. Moreover, it also shows that the results of giving Water and Coffee on memory retention were pretty much the same.

References:

- Caffeine's Effect on Memory, Cognition, and Alzheimer's. (2021). Retrieved 25 April 2021, from <https://www.caffeineinformer.com/bad-memory-drink-more-caffeine>
- Website used for Data Collection
 - The Islands. (2021). Retrieved 25 April 2021, from <https://theislands.umn.edu/login.php>