CSE 518 HW2 Report

Kaustav Sarkar
SBU ID: 112505000

Task 1:

The purpose of this task is to perform Analysis of Variance (ANOVA) and Pairwise t-test on the given dataset, comprising of user id, type of menu and time. There are 4 different types of menu in the dataset, with 40 users, 10 each for a distinctive menu type. Hence, there are 4 groups of users and it is a between-group design.

Analysis of Variance (ANOVA) Description:

The Analysis of Variance (ANOVA) is a statistical technique that allows us to compare mean differences of one outcome (dependent) variable across two or more groups (levels) of one independent variable (factor). If there are only two levels of the independent (predictor) variable the results are analogous to t-test. ANOVA test is centered around the different sources of variation (variation between and within-group) in a typical variable.

ANOVA test hypotheses:

- **Null hypothesis**: This is represented as H0 - the means of the different groups are the same. H0 hypothesis implies that there is not enough evidence to prove the means of the groups (factors) are different from another.
- **Alternative hypothesis**: Represented as H3 - At least one sample mean is not equal to the others.

Our Scenario and Data:

The scenario in our case can be summarized this way. Imagine that we are interested in understanding whether knowing a certain menu type can help us predict whether it is easier to navigate or not, in other words, performance of navigation. We have a sample of 40 users using 4 different menu types. While we expect variation across our sample we are interested in whether the differences between the menu types (groups) is significantly different than what we would expect in random variation within the groups.

Our research or testable (alternative) hypothesis (H3) is then described as

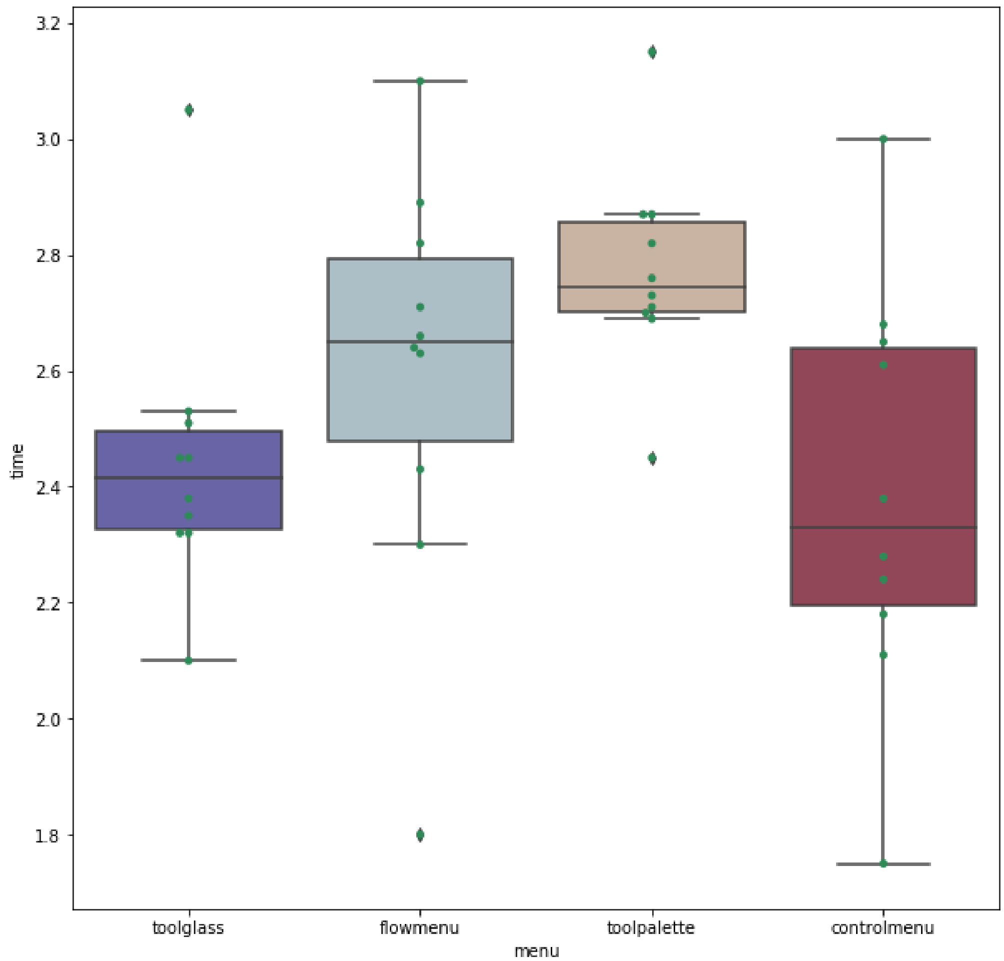

Figure 1: Boxplot for Task 1 - Menu vs Time

$\mu_{\text{toolglass}} \neq \mu_{\text{flowmenu}} \neq \mu_{\text{toolpalette}} \neq \mu_{\text{controlmenu}}$

Or at least one of the menu type is different than the other three. Our null hypothesis is basically the means of all the above menus are the same.

Implementation Steps

- We first load the required packages like dplyr and ggplot2 (for visualization).
- Then, we load the PATH of the input file, read the CSV and convert it into a dataframe.
- We can then summarise the dataframe to get different statistical measures like min, first quartile, mean, median, third quartile and max values present before performing the tests. This is visualized with a boxplot in Figure 1.

- Next, use the inbuilt function **aov** to compute one way ANOVA on time~menu, over the dataframe. The dependent variable goes to the left of the tilde and our independent or predictor variable to the right.
- Finally, use the **summary()** method to get the results obtained from ANOVA.

ANOVA Result and Analysis:
The results obtained from the ANOVA one-way test of dataset1 are:

```
> summary(anova_one_way)
            Df Sum Sq Mean Sq F value Pr(>F)
menu         3 0.8998 0.29992   3.455 0.0264 *
Residuals   36 3.1252 0.08681
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can then summarise the results of the ANOVA test as follows:

$P(3, 36) = 3.455$, $p<0.05$
where,
Menu has 3 degrees of freedom (since we had 4 different types of menus)
and Residual as 36 which is obtained from 40 (total rows in our dataset) - 3(DOF) - 1
Mean Sq = Sum Sq / Degree of Freedom
F value = Mean Sq of Menu / Mean Sq of Residual

In other words, we can reject the null hypothesis (H_0) that the means of all the different types of menus are the same, as the p-value is lower than the standard threshold of 0.05. Hence, we are confident to say that there is a significant **statistical difference between the groups**, indicated by the "*".

The data provide support for the hypothesis that the means aren't all equal – that's called the omnibus test. We have support for rejecting

$\mu_{toolglass} = \mu_{flowmenu} = \mu_{toolpalette} = \mu_{controlmenu}$

but at this point we can't state with any authority which specific pairs are different, all we can say is that at least one is different.

To analyze this we perform the pairwise t-test.

Pairwise t-test:

The pairwise t-test is used to calculate pairwise comparisons between group levels with corrections for multiple testing.

We use the built-in R method pairwise.t.test with time, menu as parameters and "bonf" as the p.adjust.method. The p.adjust.method is the method for adjusting the p-values, in other words, to control the error rate among comparisons. We use Bonferroni in our case.The Bonferroni correction is a method that is used to counteract the problem of inflated type I errors(A Type 1 error occurs when the null hypothesis is true, but we reject it because of usual sample result), while engaging in multiple pairwise comparisons between subgroups. The results obtained are:

```
        Pairwise comparisons using t tests with pooled SD

data:  df$time and df$menu

            controlmenu flowmenu toolglass
flowmenu    0.718       -        -
toolglass   1.000       1.000    -
toolpalette 0.034       1.000    0.103

P value adjustment method: bonferroni
```

Analysis of pairwise.t.test

From the above results, it is clear that the only statistically significant difference is between **toolpalette** and **controlmenu**, as the result is below 0.05, the standard value. The rest of the value pairs are above the standard p-value of 0.05.

This is the reason we get a single * in the ANOVA test summary, indicating there is only a slight statistically significant difference between the menus.

Task 2:

The purpose of this task is to perform Analysis of Variance (ANOVA) and Pairwise t-test on the given dataset, comprising of user id, type of menu and time. There are 4 different types of menu in the dataset, with 40 users, 10 each for a distinctive menu type. Hence, there are 4 groups of users and it is a between-group design.

Multivariate Analysis of Variance (MANOVA) Description:

The Multivariate Analysis of Variance (MANOVA) is simply an ANOVA with several dependent variables. That is to say, **ANOVA** tests for the difference in means between two or more groups, while **MANOVA** tests for the difference in two or more vectors of means.

In ANOVA, the variances are single values. In MANOVA, these variances are contained in a matrix.

- hypothesis SSCP: the matrix that represents the systematic variance and is called *hypothesis sum of squares and cross-products matrix*, denoted by H.
- error SSCP: the matrix that represents the unsystematic variance and is called *error sum of squares and cross-products matrix*, denoted by E.
- total SSCP: the total amount of variance, denoted by T.

If the global multivariate test is significant, we conclude that the corresponding effect is significant. In that case, we want to identify the specific dependent variables that contributed to the significant global effect.
To answer this question, we can use one-way ANOVA (or univariate ANOVA) to examine separately each dependent variable.

The test hypothesis of MANOVA are same as that of ANOVA.

Pillai-Trace:
Pillai's trace is used as a test statistic in MANOVA. It is a positively valued statistic ranging from 0 to 1. Increasing values means that effects are contributing more to the model; you should reject the null hypothesis for large values.

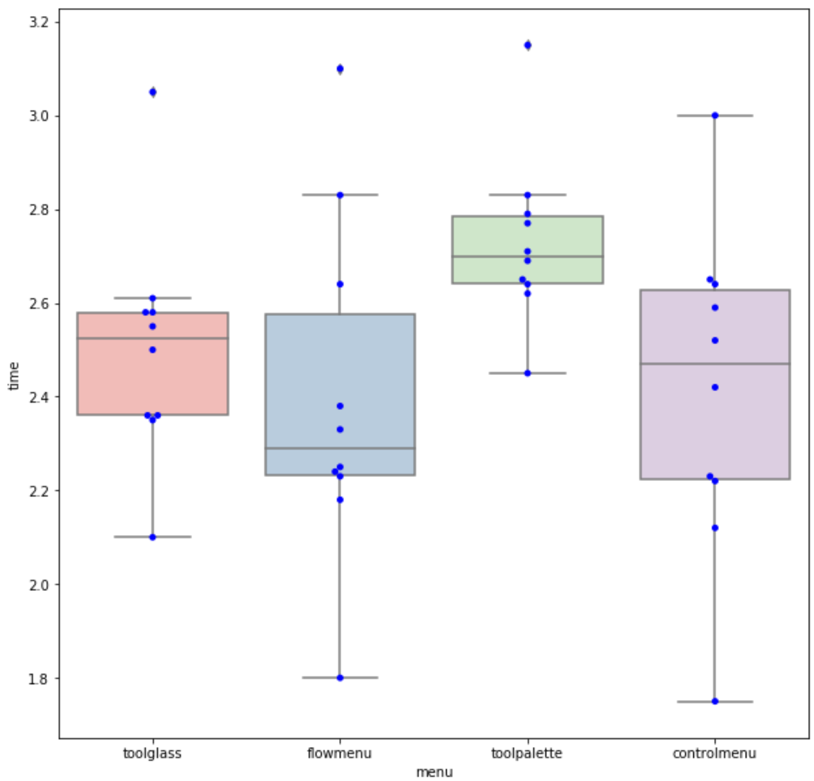

Figure 2: Boxplot for Task 2 - Menu vs Time

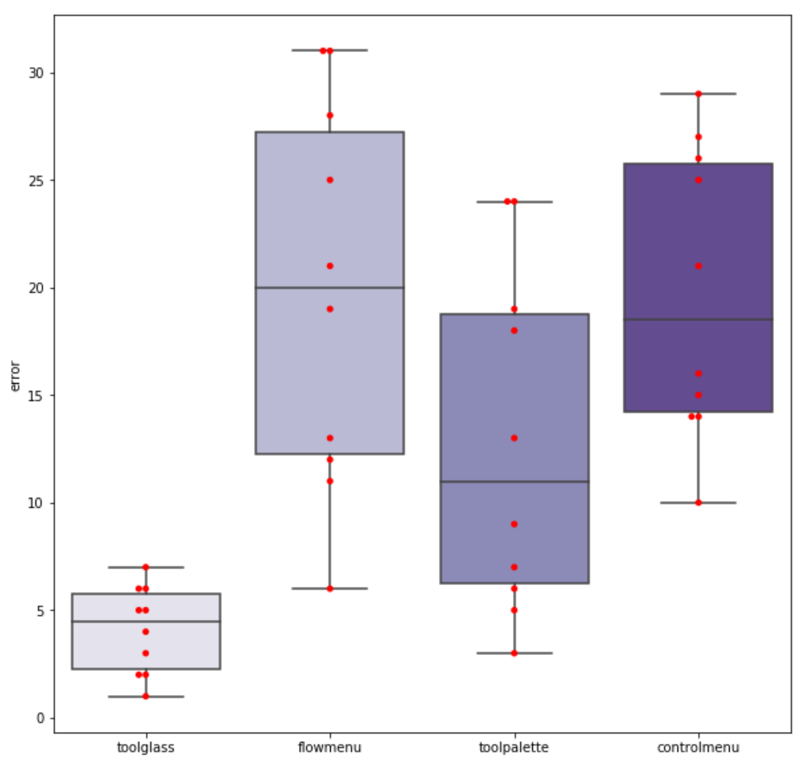

Figure 3: Boxplot for Task 2 - Menu vs Error

Implementation Steps:

- Load and summarise the data as done previously.
- We can then summarise the dataframe to get different statistical measures like min, first quartile, mean, median, third quartile and max values present before performing the tests. This is visualized with boxplots in Figure 2 and Figure 3.
- In MANOVA, we have several outcomes, so we need to bind them into one variable using cbind method.
- Then use the manova() function on the cbind method.
- Next, we use the summary() method to get the result.
- Finally, we can also perform anova on the above result to check which of the factors is differing to a greater extent.

MANOVA Result and Analysis:

The results obtained from the MANOVA test of dataset2 are:

```
> summary(manova_test)
          Df Pillai approx F num Df den Df    Pr(>F)
menu       3 0.6502   5.7804      6     72 5.918e-05 ***
Residuals 36
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The resultant MANOVA model reports a Pillai test statistic of 0.6502 and a p-value of 5.918e-05 which is below 0.05, thus H0 is rejected and it is concluded that there are significant differences in the means.

When we apply ANOVA on the above, we get:

```
> summary.aov(manova_test)
 Response tm :
            Df Sum Sq  Mean Sq F value  Pr(>F)
menu         3 0.7019 0.233957  2.6616 0.06267 .
Residuals   36 3.1644 0.087901
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response err :
            Df Sum Sq Mean Sq F value    Pr(>F)
menu         3 1644.1  548.03  11.399 2.109e-05 ***
Residuals   36 1730.7   48.07
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

On applying ANOVA, we see that it is the **error (and not time)** variable that is contributing to the significant global effect.

Pairwise t-test:

First, we perform pairwise t-test on time and menu. The results we get are:

```
> pairwise.t.test(df$time, df$menu, p.adj = "bonf")

        Pairwise comparisons using t tests with pooled SD

data:  df$time and df$menu

            controlmenu flowmenu toolglass
flowmenu    1.00        -        -
toolglass   1.00        1.00     -
toolpalette 0.14        0.10     0.58

P value adjustment method: bonferroni
```

Then, we perform pairwise t-test on error and menu. The results we get are:

```
> pairwise.t.test(df$error, df$menu, p.adj= "bonf")

        Pairwise comparisons using t tests with pooled SD

data:  df$error and df$menu

            controlmenu flowmenu toolglass
flowmenu    1.000       -        -
toolglass   8.2e-05     8.2e-05  -
toolpalette 0.195       0.195    0.048

P value adjustment method: bonferroni
```

Analysis of pairwise.t.test
From the above results, it is clear that although in the case of menu vs time, we see no statistical significance but in the case of menu vs error, there is statistically significant difference. The statistically significant difference is between **toolglass** and **controlmenu**, **toolglass** and **flowmenu**, and **toolpalette** and **toolglass**, as the result is below 0.05, the standard value. The rest of the value pairs are above the standard p-value of 0.05.

Visualizations:

Apart from the visualizations used in this report, I have also made some other visualizations, that will be included in the Jupyter Notebook file.

References:

1. Oneway ANOVA Explanation and Example in R; Part 1 | DataScience+
2. R ANOVA Tutorial: One way & Two way (with Examples)
3. MANOVA Test in R: Multivariate Analysis of Variance - Easy Guides - Wiki - STHDA