Statistically Significant Differences in Samples

Lab C Group 2: Simu Huang, Minqi Li, Jiachen Feng, Xijia Luo

1 Introduction

1.1 Project Introduction

Our client collected some carbonized seeds and charcoal samples from several different rooms of the site and she is trying to figure out all the different ways that people use these things and see if the way these plants are distributed in archaeological sites today still basically represents those different areas of activity.

Therefore, in this project, the client wants to compare whether all the different categories of corps from the 10 different sample sites are significantly different from each other.

We will use the Kruskal-Wallis test and post hoc test to find out whether there is a statistical significance to the difference between these samples.

1.2 Data Introduction

The data we get from the client contains 10 samples with multi-dimensional variables. Since the client wanted to study the relationship between each other under the variable: corps categories. We did the data cleaning first and kept the following variables.

Table1: Data Description

	Description	Type
Cereal	Main classification of corps	Continuous
Cereal indet (wt.)	subset of cereal	Continuous
Hordeum vulgare (wt)	subset of cereal	Continuous
cf. Hordeum vulgare (wt)	subset of cereal	Continuous
Triticum aestivum/durum (wt)	subset of cereal	Continuous
Pulses	Main classification of corps	Continuous
cf. Lathyrus sp. (wt)	subset of pulses	Continuous

Lathryus sp. (wt)	subset of pulses	Continuous
cf. Lens culinaris (wt)	subset of pulses	Continuous
cf. Pisum sp. (wt)	subset of pulses	Continuous
cf. Vicia sp. (wt)	subset of pulses	Continuous
Fabaceae/Pulse (wt)	subset of pulses	Continuous
Fruit, Nuts, etc	Main classification of corps	Continuous
Ficus carica (wt)	subset of Fruit, Nuts, etc	Continuous
Olea europaea (wt)	subset of Fruit, Nuts, etc	Continuous
cf. Pinus sp. (wt)	subset of Fruit, Nuts, etc	Continuous
Vitis vinifera (wt.)	subset of Fruit, Nuts, etc	Continuous
Endocarp (wt.)	subset of Fruit, Nuts, etc	Continuous

2 EDA

Our EDA includes a table summarizing samples' compositions and 4 different types of charts comparison-Cereals, Pulses, Fruit, Nuts, Etc.

2.1 Data Overview

- Cereals:

Table 2: Data Overview of Cereals

Case Sample #	Cereal indet (wt.)	Hordeum vulgare (wt)	cf. Hordeum vulgare (wt)	Triticum aestivum/durum (wt)	Total Weight
20656	0.000	0.000	0.000	0.000	0.000
20655	0.000	0.004	0.000	0.000	0.004
20654	0.019	0.020	0.007	0.026	0.072
20650	0.000	0.000	0.000	0.000	0.000
20648	0.009	0.000	0.000	0.007	0.016

20649	0.023	0.053	0.047	0.027	0.150
20651	0.000	0.000	0.000	0.000	0.000
20657	0.033	0.057	0.000	0.059	0.149
20653	0.035	0.000	0.000	0.000	0.035
20652	0.000	0.000	0.000	0.000	0.000

- Pulses:

Table 3: Data Overview of Pulses

Case Sample #	cf. Lathyrus sp. (wt)	Lathryus sp. (wt)	cf. Lens culinaris (wt)	cf. Pisum sp. (wt)	cf. Vicia sp. (wt)	Fabaceae/ Pulse (wt)	Total
20656	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20655	0.000	0.000	0.001	0.000	0.000	0.000	0.001
20654	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20650	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20648	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20649	0.000	0.028	0.000	0.000	0.005	0.020	0.053
20651	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20657	0.000	0.000	0.000	0.003	0.000	0.000	0.003
20653	0.000	0.000	0.000	0.000	0.000	0.000	0.000
20652	0.000	0.000	0.000	0.000	0.000	0.000	0.000

- Fruit, Nuts, etc.

Table 4: Data Overview of Fruit, Nuts, etc.

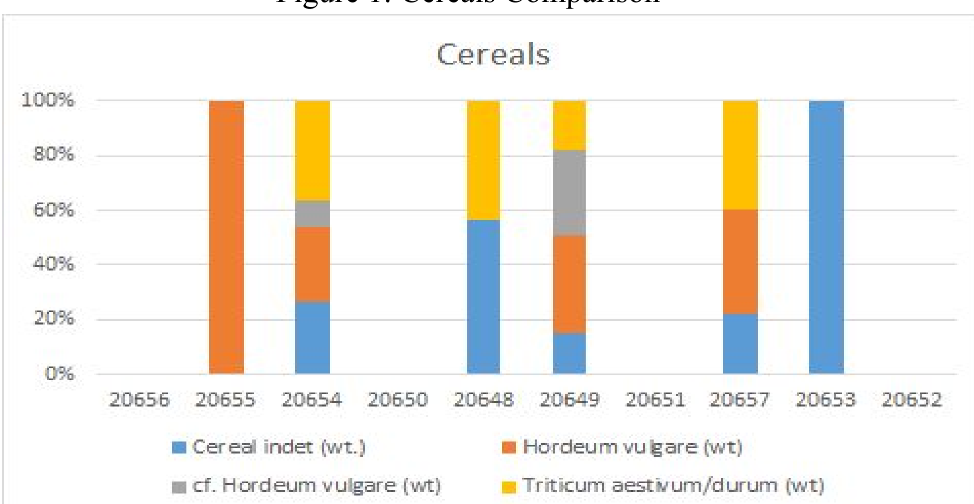
Case Sample #	Ficus carica (wt)	Olea europaea (wt)	cf. Pinus sp. (wt)	Vitis vinifera (wt)	Endocarp (wt.)	Total
20656	0.000	0.000	0.000	0.000	0.000	0.000

20655	0.000	0.060	0.000	0.007	0.000	0.067
20654	0.000	0.069	0.000	0.000	0.000	0.069
20650	0.000	0.044	0.000	0.000	0.000	0.044
20648	0.000	0.275	0.000	0.000	0.000	0.275
20649	0.000	0.229	0.000	0.000	0.000	0.229
20651	0.000	0.159	0.000	0.000	0.000	0.159
20657	0.000	0.288	0.000	0.000	0.000	0.288
20653	0.000	0.353	0.000	0.000	0.000	0.353
20652	0.000	0.120	0.000	0.000	0.007	0.127

2.2 Visualization

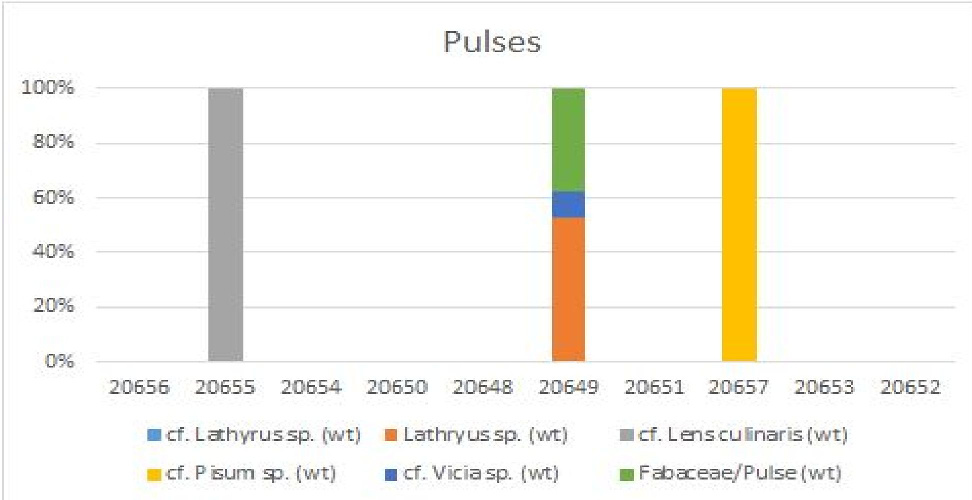
- Cereals Comparison:

Figure 1: Cereals Comparison



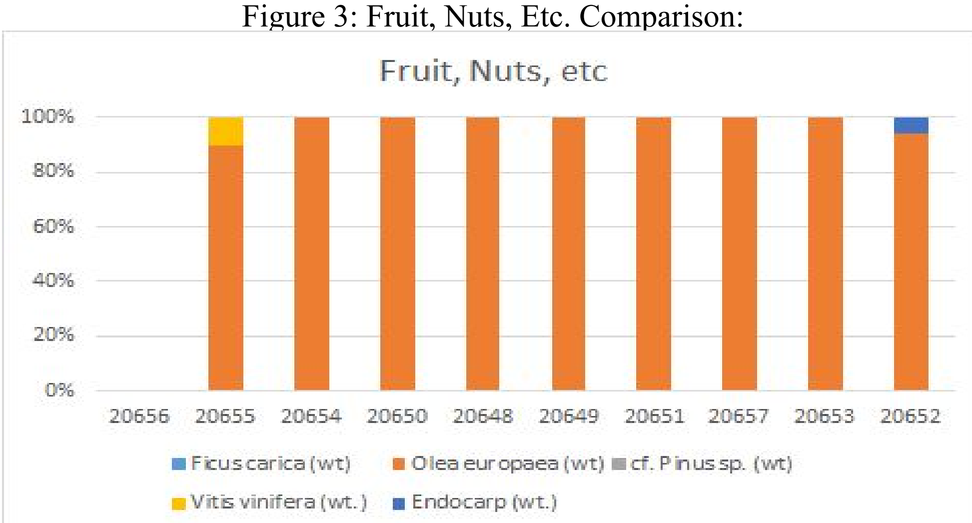
- Pulses Comparison：

Figure 2: Pulses Comparison



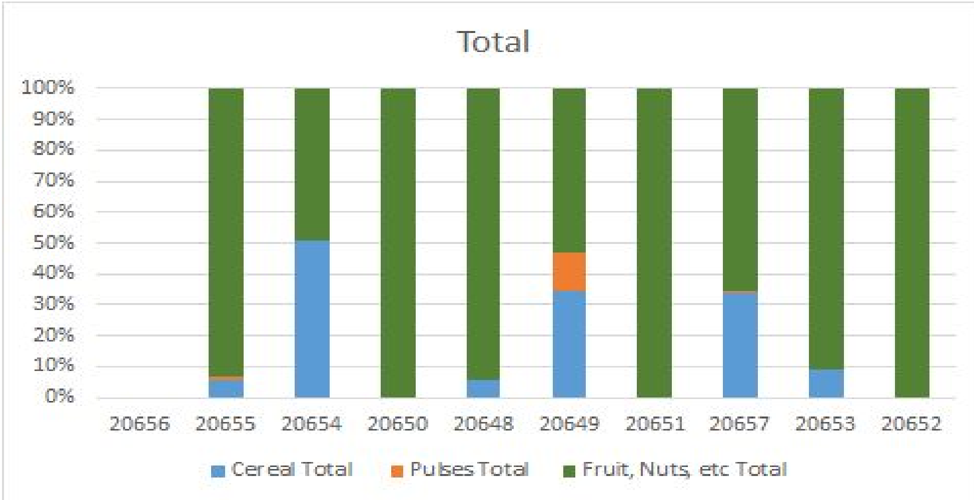
- Fruit, Nuts, Etc. Comparison:

Figure 3: Fruit, Nuts, Etc. Comparison:



- Total:

Figure 4: Total Comparison:



A preliminary review of the data reveals interesting compositions between the different groups. It appears, at least from the sample, that the proportion of fruit, nut, etc. group is the highest. Among them, Olea europaea is the most common component. This setup is conducive to directly observe and analyze whether there is a significant difference between each two samples.

Another noteworthy element of the EDA is that there may be errors in the original dataset. There are many blank cells in the excel document, we just auto filled all these blank cells as 0. The clients asked us to only analyze these data using the total weight. However, we found out that it is almost impossible to do any research with only total weight, thus we break them into parts which can help the clients better understand the analysis.

3 Hypothesis Test

3.1 Kruskal–Wallis test

Kruskal-Wallis test is used for comparing two or more independent samples of equal or different sample sizes. It extends the Mann-Whitney U test which is used for comparing only two groups. The null hypothesis of the K-W test is that the distribution is the same across different groups, and the alternative hypothesis is that at least one group's distribution is different from at least one other group's distribution.

3.2 Results

We use the K-W test to compare the difference of the weight of plant remains under each crop among ten samples. The following is the output.

3.2.1 Cereals

Table 5: Hypothesis Test Summary with Kruskal–Wallis Test

Null Hypothesis	Test	Sig.	Decision
The distribution of cereals is the same across categories of groups.	Independent-Samples Kruskal-Wallis Test	.002	Reject the null hypothesis.

3.2.2 Pulses

Table 6: Hypothesis Test Summary with Kruskal–Wallis Test

Null Hypothesis	Test	Sig.	Decision
The distribution of pulses is the same across categories of groups.	Independent-Samples Kruskal-Wallis Test	.025	Reject the null hypothesis.

3.2.3 Fruits, Nuts, etc.

Table 7: Hypothesis Test Summary with Kruskal–Wallis Test

Null Hypothesis	Test	Sig.	Decision
The distribution of fruits, nuts, etc. is the same across categories of groups.	Independent-Samples Kruskal-Wallis Test	.981	Retain the null hypothesis.

4 Post Hoc Test

4.1 Introduction

Post hoc tests perform two vital tasks. They tell us which group means are significantly different from other group means. Crucially, they also control the experiment-wise, or familywise, error rate.

For every hypothesis test we perform, there is a type I error rate—a false positive, which your significance level (alpha) defines. When we perform only one test, the type I error rate equals significance level, which is often 5%. However, when many comparisons are being made in an analysis, the error rate for each comparison can impact the other results, creating multiple false positives. If we perform enough tests, we're virtually guaranteed to get a false positive! The error rate for a family of tests is always higher than an individual test. Therefore, we need to use post hoc tests to control experiment-wise error rate.

4.2 Bonferroni Correction

The Bonferroni correction is one of several methods used to counteract the problem of multiple comparisons and is also applied as a post-hoc test in many statistical procedures. Bonferroni correction is calculated by taking the number of tests and dividing it into the alpha value. Using the 5% error rate from our example, two tests would yield an error rate of 0.025 or (.05/2) while four tests would therefore have an error rate of .0125 or (.05/4).

However, it has disadvantages, as well, since it is unnecessarily conservative (with weak statistical power). The adjusted p-value is often smaller than required, particularly if there are many tests or the test statistics are positively correlated. Therefore, this method often fails to detect real differences.

Therefore, we plan to use a less consevative method which is the Benjamini-Hochberg method.

4.3 Benjamini-Hochberg Method

Benjamini-Hochberg method controls the false discovery rate which is an alternative to control the family-wise error rate. The false discovery rate controls the expected rate of the null hypothesis that is incorrectly rejected (type I error) in the rejected hypothesis list. It is less conservative. By performing the comparison procedure with a greater power compared to family-wise error rate control, the probability that a type I error will occur can be increased.

Calculate each individual p-value's Benjamini-Hochberg critical value, using the formula $(i/m)Q$, where:

- i = the individual p-value's rank,
- m = total number of tests,
- Q = the false discovery rate (a percentage).

Compare original p-values to the critical B-H value; find the largest p value that is smaller than the critical value. The largest P value that has $P<(i/m)Q$ is significant, and all of the P values smaller than it are also significant, even the ones that aren't less than their Benjamini-Hochberg critical value.

In our report, we use the false discovery rate $Q = 0.1$.

4.4 Results

The following tables are the output which just displays statistically significant differences of samples. The complete output is in the appendix.

4.4.1 Cereals

Table 8: Pairwise Comparisons

Sample 1-Sample 2	Sig.	Adj. Sig. (Bonferroni Correction)	(i/m)Q (Benjamini-Hochberg Method)	Decision
20656-20649	0.002	0.088	0.002	reject
20652-20649	0.002	0.088	0.002	reject
20651-20649	0.002	0.088	0.002	reject
20650-20649	0.002	0.088	0.002	reject
20655-20649	0.008	0.369	0.011	reject
20650-20657	0.009	0.409	0.013	reject
20651-20657	0.009	0.409	0.013	reject
20652-20657	0.009	0.409	0.013	reject
20656-20657	0.009	0.409	0.013	reject
20650-20654	0.016	0.703	0.022	reject
20651-20654	0.016	0.703	0.022	reject
20652-20654	0.016	0.703	0.022	reject
20656-20654	0.016	0.703	0.022	reject
20653-20649	0.022	0.976	0.031	reject
20655-20657	0.031	1.000	0.033	reject
20648-20649	0.042	1.000	0.036	retain

4.4.2 Pulses

Table 9: Pairwise Comparisons

Sample 1-Sample 2	Sig.	Adj. Sig. (Bonferroni Correction)	(i/m)Q (Benjamini-Hochberg Method)	Decision
20648-20649	0.001	0.06	0.002	reject

20650-20649	0.001	0.06	0.002	reject
20651-20649	0.001	0.06	0.002	reject
20652-20649	0.001	0.06	0.002	reject
20653-20649	0.001	0.06	0.002	reject
20654-20649	0.001	0.06	0.002	reject
20656-20649	0.001	0.06	0.002	reject
20655-20649	0.025	1	0.018	retain

5 Discussion

Based on the results we provided above, we can conclude that:

- For cereals, we got a p-value of 0.002 through a Kruskal–Wallis Test, which means that there is a statistical significance to the difference. Even if there is a statistical significance between the overall samples, we still cannot exactly find those samples which stand out of all the ten samples. Therefore, we performed post hoc tests. Although the Bonferroni correction is widely used in this situation, we chose to use the Benjamini-Hochberg method because of its less conservative characteristics. The Bonferroni correction corresponds to an extreme low significance level, which leads to a low statistical power. And a study with a lower statistical power is less likely to detect the difference between groups.
  The results of post hoc tests are divided into two types, one is there is statistical significance to the two samples and the other is of no clear significance. Interestingly, we found that samples 20649, 20654 and 20657 are included in those groups that show statistical significance. Therefore, we conclude that these three samples are statistically significant.

- For pulses, we got a p-value of 0.025 through a Kruskal–Wallis Test, which means that there is a statistical significance to the difference. They encountered the same situation as the cereals group, so we also performed post hoc tests. Similarly, we got a sample 20649 included in groups that show statistical significance. Therefore, we conclude that sample 20649 is statistically significant.

- For fruits and nuts, we got a p-value of 0.981 through a Kruskal–Wallis Test, which means that there is no statistical significance to the difference.

6 References

[1] Wikipedia [online]. Available from:
https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance
[accessed 10 November 2020]

[2] Statistics By Jim [online]. Available from:
https://statisticsbyjim.com/anova/post-hoc-tests-anova/ [accessed 17 November 2020]

[3] Investopedia [online]. Available from:
https://www.investopedia.com/terms/b/bonferroni-test.asp [accessed 17 November 2020]

[4] Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test?. Korean journal of anesthesiology, 71(5), 353–360. https://doi.org/10.4097/kja.d.18.00242

[5] McDonald, J.H., 2009. Handbook of biological statistics (Vol. 2, pp. 6-59). Baltimore, MD: sparky house publishing.

[6] Statistics How to [online]. Available from:
https://www.statisticshowto.com/benjamini-hochberg-procedure/ [accessed 17 November 2020]

7 Appendix

- Cereals

Table: Pairwise Comparisons of Samples

Sample 1-Sample 2	Sig.	Rank (Sig.)	Adj. Sig. (Bonferroni Correction)	(i/m)Q (Benjamini-Hochberg Method)	Decision
20656-20649	0.002	1	0.088	0.002	reject
20652-20649	0.002	1	0.088	0.002	reject
20651-20649	0.002	1	0.088	0.002	reject
20650-20649	0.002	1	0.088	0.002	reject
20655-20649	0.008	5	0.369	0.011	reject
20650-20657	0.009	6	0.409	0.013	reject

20651-20657	0.009	6	0.409	0.013	reject
20652-20657	0.009	6	0.409	0.013	reject
20656-20657	0.009	6	0.409	0.013	reject
20650-20654	0.016	10	0.703	0.022	reject
20651-20654	0.016	10	0.703	0.022	reject
20652-20654	0.016	10	0.703	0.022	reject
20656-20654	0.016	10	0.703	0.022	reject
20653-20649	0.022	14	0.976	0.031	reject
20655-20657	0.031	15	1.000	0.033	reject
20648-20649	0.042	16	1.000	0.036	retain
20655-20654	0.049	17	1	0.038	retain
20653-20657	0.070	18	1	0.040	retain
20653-20654	0.106	19	1	0.042	retain
20648-20657	0.122	20	1	0.044	retain
20648-20654	0.175	21	1	0.047	retain
20650-20648	0.289	22	1	0.049	retain
20651-20648	0.289	22	1	0.049	retain
20652-20648	0.289	22	1	0.049	retain
20656-20648	0.289	22	1	0.049	retain
20651-20653	0.424	26	1	0.058	retain
20652-20653	0.424	26	1	0.058	retain
20656-20653	0.424	26	1	0.058	retain
20650-20653	0.424	26	1	0.058	retain
20654-20649	0.498	30	1	0.067	retain
20655-20648	0.543	31	1	0.069	retain
20657-20649	0.626	32	1	0.071	retain
20650-20655	0.651	33	1	0.073	retain
20651-20655	0.651	33	1	0.073	retain

20652-20655	0.651	33	1	0.073	retain
20656-20655	0.651	33	1	0.073	retain
20655-20653	0.728	37	1	0.082	retain
20653-20648	0.794	38	1	0.084	retain
20654-20657	0.848	39	1	0.087	retain
20650-20656	1.000	40	1	0.089	retain
20651-20652	1.000	40	1	0.089	retain
20651-20656	1.000	40	1	0.089	retain
20652-20656	1.000	40	1	0.089	retain
20650-20652	1.000	40	1	0.089	retain
20650-20651	1.000	40	1	0.089	retain

- Pulses

Table: Pairwise Comparisons of Samples

Sample 1-Sample 2	Sig.	Rank (Sig.)	Adj. Sig. (Bonferroni Correction)	(i/m)Q (Benjamini-Hochberg Method)	Decision
20648-20649	0.001	1	0.06	0.002	reject
20650-20649	0.001	1	0.06	0.002	reject
20651-20649	0.001	1	0.06	0.002	reject
20652-20649	0.001	1	0.06	0.002	reject
20653-20649	0.001	1	0.06	0.002	reject
20654-20649	0.001	1	0.06	0.002	reject
20656-20649	0.001	1	0.06	0.002	reject
20655-20649	0.025	8	1	0.018	retain
20657-20649	0.027	9	1.000	0.020	retain
20650-20657	0.317	10	1	0.022	retain
20651-20657	0.317	10	1	0.022	retain

20652-20657	0.317	10	1	0.022	retain
20653-20657	0.317	10	1	0.022	retain
20654-20657	0.317	10	1	0.022	retain
20656-20657	0.317	10	1	0.022	retain
20648-20657	0.317	10	1	0.022	retain
20651-20655	0.334	17	1	0.038	retain
20652-20655	0.334	17	1	0.038	retain
20653-20655	0.334	17	1	0.038	retain
20654-20655	0.334	17	1	0.038	retain
20656-20655	0.334	17	1	0.038	retain
20648-20655	0.334	17	1	0.038	retain
20650-20655	0.334	17	1	0.038	retain
20655-20657	0.972	24	1	0.053	retain
20650-20656	1	25	1	0.056	retain
20651-20652	1	25	1	0.056	retain
20651-20653	1	25	1	0.056	retain
20651-20654	1	25	1	0.056	retain
20651-20656	1	25	1	0.056	retain
20652-20653	1	25	1	0.056	retain
20652-20654	1	25	1	0.056	retain
20652-20656	1	25	1	0.056	retain
20653-20654	1	25	1	0.056	retain
20653-20656	1	25	1	0.056	retain
20654-20656	1	25	1	0.056	retain
20648-20650	1	25	1	0.056	retain
20648-20651	1	25	1	0.056	retain
20648-20652	1	25	1	0.056	retain
20648-20653	1	25	1	0.056	retain

20648-20654	1	25	1	0.056	retain
20648-20656	1	25	1	0.056	retain
20650-20651	1	25	1	0.056	retain
20650-20652	1	25	1	0.056	retain
20650-20653	1	25	1	0.056	retain
20650-20654	1	25	1	0.056	retain