Q1. Consider the following partial ANOVA table:

Source of Variation	***SS***	***df***	***MS***	***F***
Treatment	**75**	*	**25**	*
Error	*	*	*	
Total	**135**	**19**		

a. Determine the df for treatment

Solution : $MS_{Treatment} = \frac{SS_{Treatment}}{df_{Treatment}}$

Total of Degrees for Freedom = 19

Total of Sum for Squares = 135

$MS_{Treatment}$ = 25 and $SS_{Treatment}$ = 75

Calculating the values for $df_{Treatment}$, we get

$$MS_{Treatment} = \frac{SS_{Treatment}}{df_{Treatment}}$$

$$25 = \frac{75}{df_{Treatment}}, df_{Treatment} = 3$$

b.Determine the df for error

Solution : We know the values of $df_{Treatments}$ from the above question,

equating that to the formula we get,

Now, we also know that the Total Degrees of Freedom = 19.

Now, a * n-1 = 19, So, a * n = 20

We know that a -1 = 3, So, a = 4

As a * n = 20 and a = 4, so n = 5.

So, Degrees of Freedom for Error = a * (n-1) = 4 * (5-1) = 16

$df_T = df_{Treatments} + df_E$

$19 = 3 + df_E$

$df_E = 16$

c. Determine the SS for error:

Solution: We have the value for $SS_{Treatments}$ and SS_T, equating the values in the formula,

$SS_T = SS_{Treatments} + SS_E$

$135 = 75 + SS_E$

$SS_E = 60$

d. Determine the MS for error

Solution : We have the values as given for SS_E and df_E

$MS_E = SS_E/df_E$

$MS_E = SS_E/df_E$

$MS_E = 60/16$

$MS_E = 3.75$

e. Determine the F-statistic

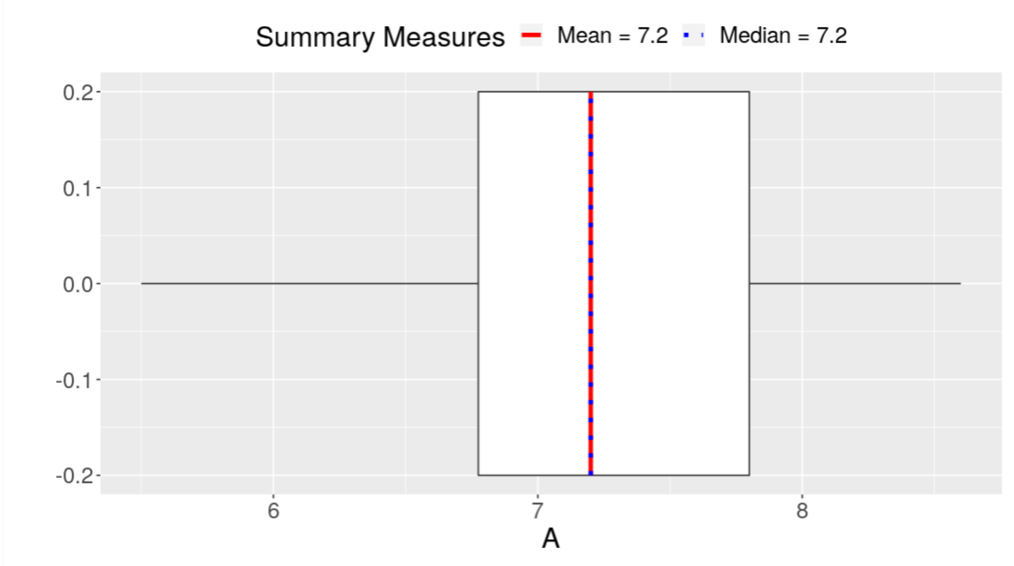
Solution : We have the values given,

$F_0 = \frac{MS_{treatment}}{MS_E}$ = 25 / 3.75 = 6.667

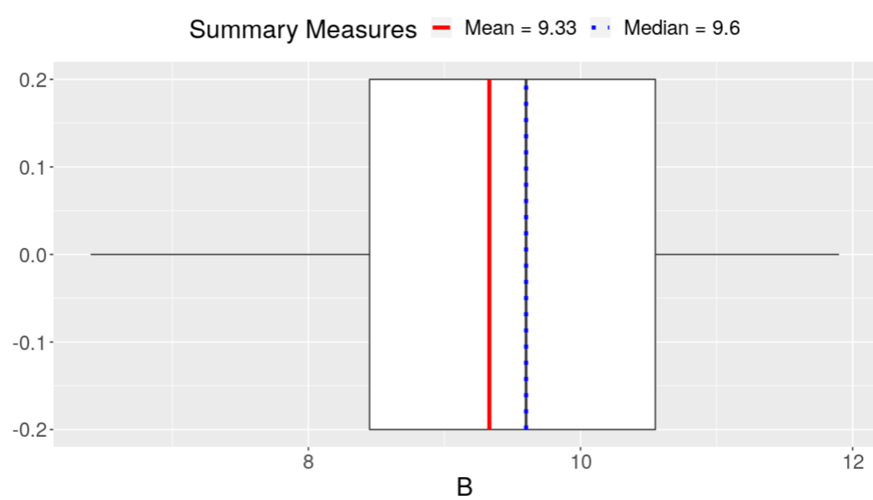
2. The data in the table below shows the measurements of hemoglobin (g per 100 ml) in the blood of brown trout. The trout were placed at random in four different troughs. The fish food added to the troughs contained, respectively, 0, 5, 10, and 15 grams of sulfamerazine per 100 pounds of fish (coded A, B, C, and D). The measurements were made on 10 fish randomly selected from each trough after 35 days.

Code	Hemoglobin (g per 100 ml)									
A	6.7	7.8	5.5	8.4	7	7.8	8.6	7.4	5.8	7
B	9.9	8.4	10.4	9.3	10.7	11.9	7.1	6.4	8.6	10.6
C	10.4	8.1	10.6	8.7	10.7	9.1	8.8	8.1	7.8	8
D	9.3	9.3	7.2	7.8	9.3	10.2	8.7	8.6	9.3	7.2

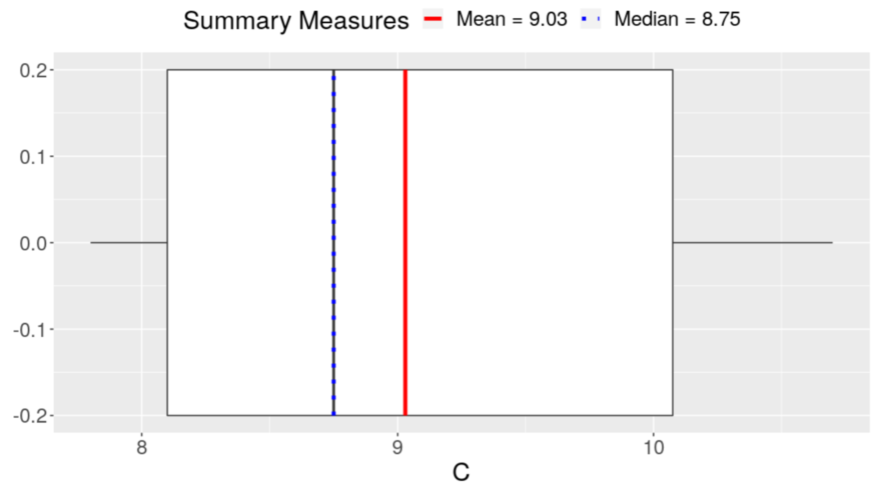
a. **Create the summary plot of the data and comment:**
   Summary plot for code A: Here, we can see that mean = 7.2 and median are also the same. There are no outliers present.



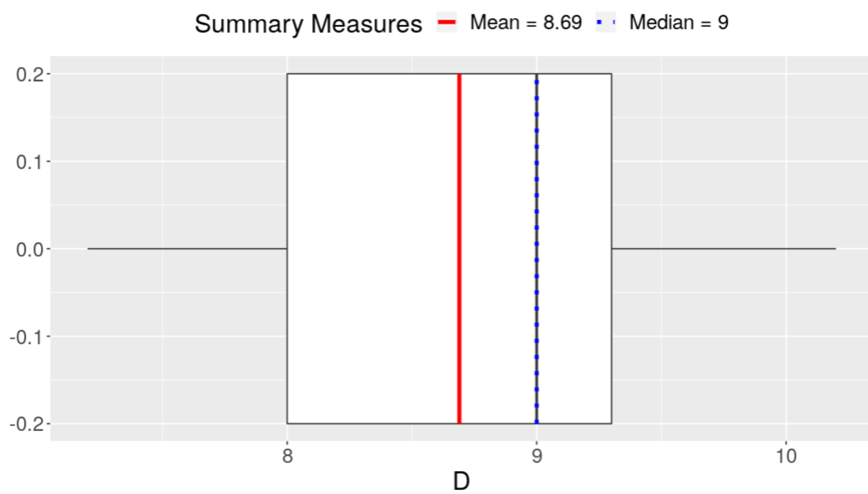
Summary plot for code B:The mean and median here are different. The mean is 9.33 and the median is 9.6. Similar to above there are no outliers present.



Summary plot for code C: Here the mean =9.03 and the median is 8.75. There are no outliers present



Summary plot for code D:Here, the mean is 8.69 and the median is 9, and similar to all above, no outliers are present.



b. Write down the null and alternative hypothesis for the test to see if there are any differences among the mean amounts of sulfamerazine added.

Since we have more than 2 levels for the input variable, we cannot see if all the means are different from each other with one test. By rejecting the null hypothesis, that at least one pair is different. So the alternative hypothesis is that at least one mean is different from another.

The null, in this case, will be that all of the means are equal.
The hypothesis becomes:
Ho:$\mu1=\mu2=\mu3=\mu4$
HA:$\mu i\neq\mu j$ for at least one (i,j) pair where i<j

c. Test this hypothesis using an ANOVA and a=0.05 and draw conclusions.
We know the formula for Fstat.
$F_{stat}=MS_{treat}/MS_E$
We have the data:
Sum of data:

SS_T=26.80
SS_M=56.47
SS_E=83.27

Degree of freedom:
Df_T=3
Df_M=36
Df_E=39

Mean square:
MS_M=8.93
MS_E=1.56

Now, F_{stat}= MS_{treat}/MS_E
F_{stat}= 8.93/1.56
`F_{stat}= 5.72

Now finding p-value:
p-value=$P(Fdf_{treat,}df_E \geq F_{stat})$
p-value=$1-P(Fdf_{treat,}df_E \leq F_{stat})$
p-value= 0.0027

In this case, the p-value is less than α,
hence we reject the null hypothesis
There is enough evidence to support the alternate hypothesis.
Hence we can conclude that there is at least one code type that has a different mean than another one.

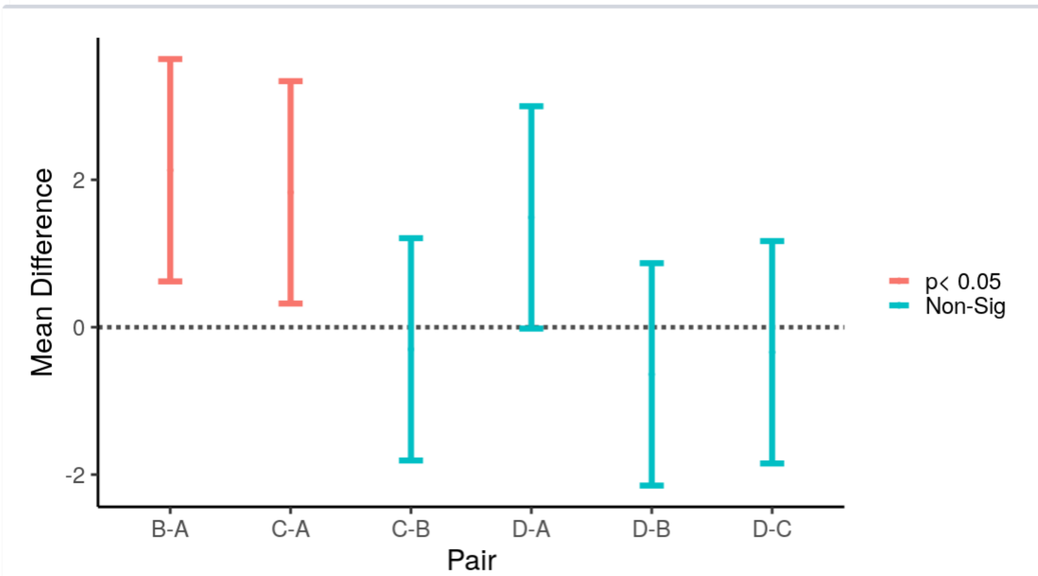
d. Determine which means are different

We look at Tukey's test results:
R^2=$(1 - SSE/SST) * 100$
R^2=32.186

Tukey's Test Results Table

Pairs	Mean Difference	Lower CB	Upper CB	P-Value
B-A	2.13	0.6215	3.639	0.0029
C-A	1.83	0.3215	3.339	0.0122
D-A	1.49	-0.0185	2.999	0.0539
C-B	-0.3	-1.809	1.209	0.9498
D-B	-0.64	-2.149	0.8685	0.666
D-C	-0.34	-1.849	1.169	0.9292

Tukey's Test Results Graph



Looking at the above graph, we see that there are 3 lines that cross the horizontal line at 0, and these pairs of code are (C-B, D-B, D-C, D-A).
We can conclude that these are the 4 pairs (C-B, D-B, D-C, D-A) that do not have a significant difference in means.
The rest of the pairs: (B-A, C-A) have a major significant difference in their means.

e. **Draw conclusions to the real-world**
Since we rejected the null hypothesis, we can conclude that there is at least one Code that has a different mean than any other code. We see that there are 4 pairs that show no significant change in the amount of hemoglobin content in the trouts, whereas there are 2 pairs that show that there is a significant change in the hemoglobin content in the trouts.

f. **Did the outcome match what you thought would happen based on your plot, explain**

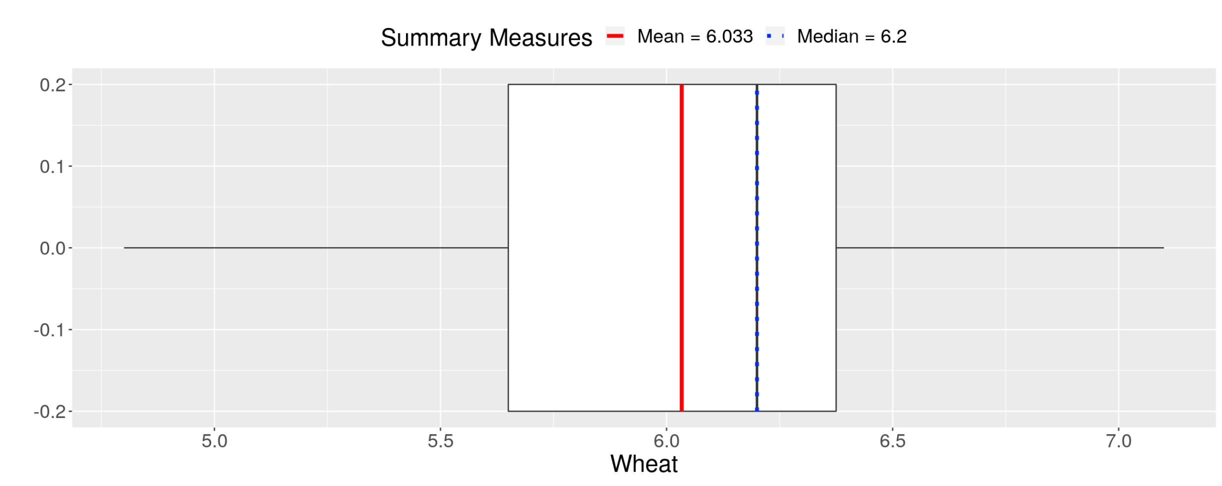
When we rejected the null hypothesis, we meant to say that there the means would have differences between them and won't be similar. But when we look at Tukey's test, we can see only 2 pairs having a significant difference in their means, the rest of the pairs do not have a significant difference in their mean values. Hence we can say that the outcome did not match what we had thought would probably happen.

3.

Solution:

a) **Create a summary plot of the data and comment**

Wheat:

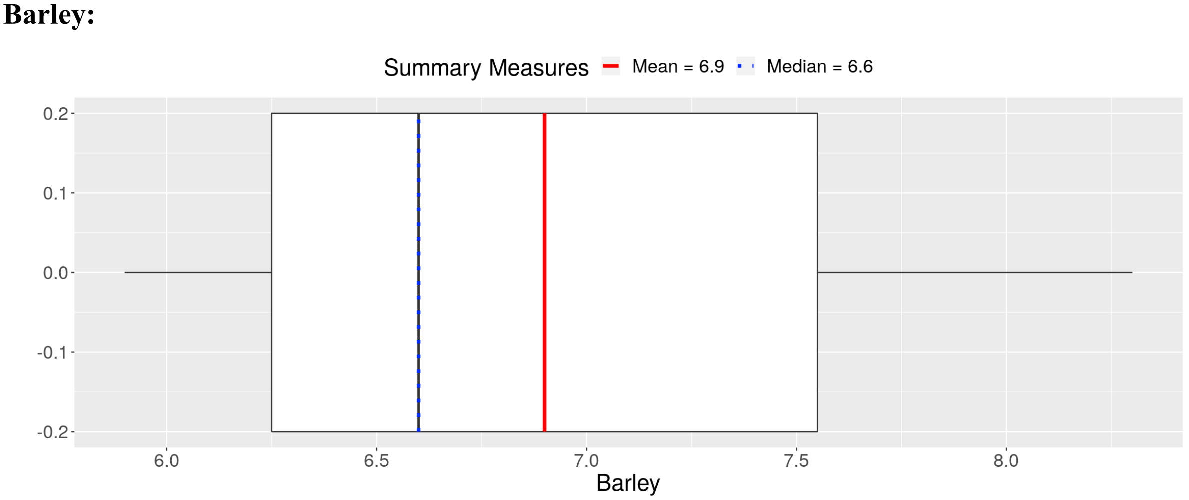


The boxplot of the Wheat shows that the mean of the data lies at 6.033 and the median is 6.2. As we can see from the box plot there are no outliers present.

Maximum and minimum values in the data are 7.1 and 4.8 respectively.

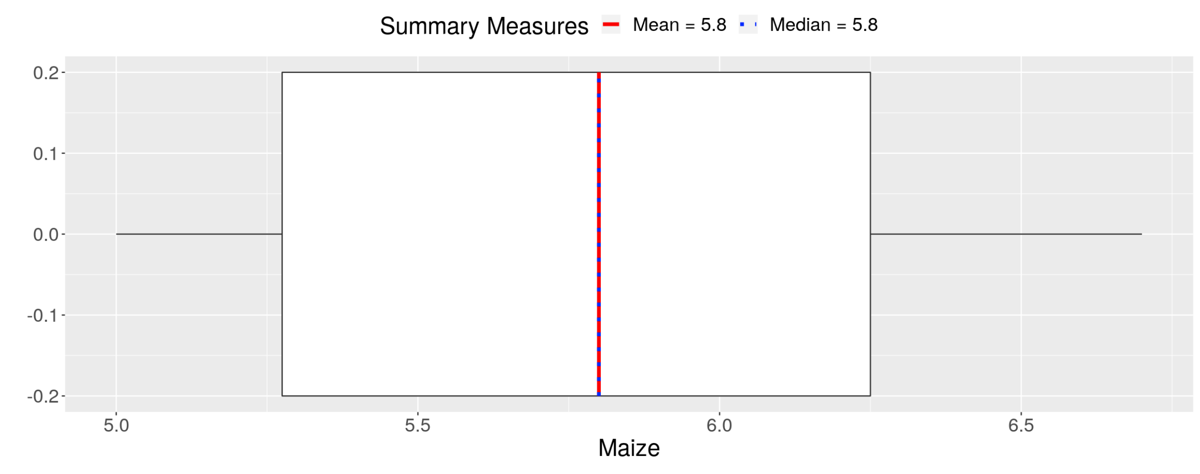
IQR = 1.25

Barley:



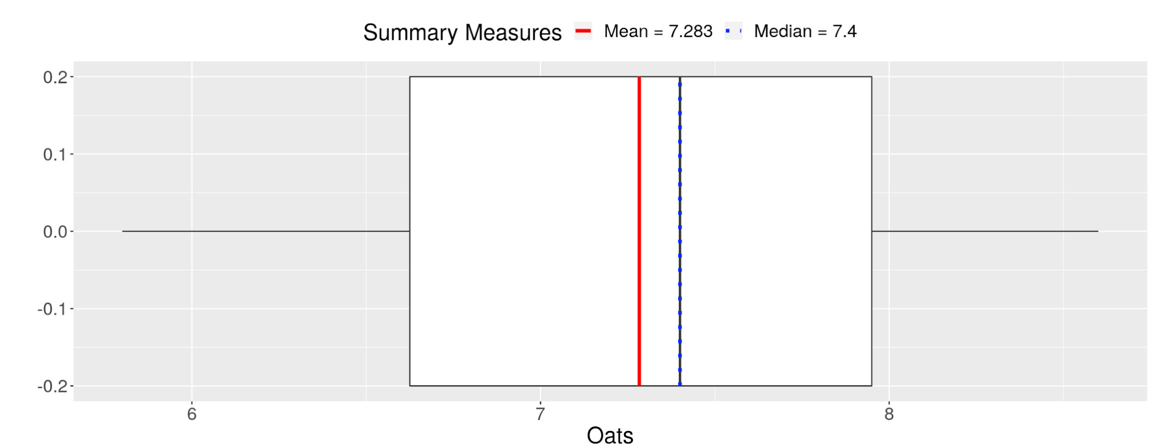
The boxplot of the Barley shows that the mean of the data lies at 6.9 and the median is 6.6. As we can see from the box plot there are no outliers present.
Maximum and minimum values in the data are 5.9 and 8.3 respectively.
IQR = 1.8

Maize:



The boxplot of the wheat shows that the mean of the data lies at 5.8 and the median is 5.8. As we can see from the box plot there are no outliers present.
Maximum and minimum values in the data are 5 and 6.7 respectively.
IQR = 1.25

Oats:



The boxplot of the wheat shows that the mean of the data lies at 7.283 and the median is 7.4. As we can see from the box plot there are no outliers present.
Maximum and minimum values in the data are 5.8 and 8.6 respectively.
IQR = 1.975

b) Write down the null and alternative hypothesis for the test to see if there are any differences among the mean amounts of sulfamerazine added.

We would like to see if the stated population means are equal.
The null, in this case, will be that all of the means will be equal.

The null and alternative hypothesis becomes,
$H_o: \mu 1=\mu 2=\mu 3=\mu 4$

$H_A: \mu i \neq \mu j$ for at least one (i,j) pair where $i<j$

c) Test this hypothesis using an ANOVA and a=0.05 and draw conclusions.

ANOVA Table

	SS	df	MS	F	P-value
X	8.89	3.00	2.96	3.86	0.0249
Error	15.34	20.00	0.77	NA	NA
Total	24.23	23.00	NA	NA	NA

Fstat=$MStreat/MSE$

Sum of squares
SS_T=24.23
SS_M=8.888
SS_E=15.342

Degree of freedom
dfT=23
dfM=3
dfE=20

Mean Square

MS_M=2.963

MS_E=0.7671

Now, we will put this value in the below equation,
Fstat=$MStreat/MSE$
Fstat=2.963/0.7671
Fstat=3.862

Now we will find a p-value with the formula below for a greater than a test.
p-value=P($F_{df_{treat}, df_E} \geq F_{stat}$)
p-value=1-P($F_{df_{treat}, df_E} \leq F_{stat}$)
p-value=0.0249

We will reject H0 if p-value ≤ α

In this case:

P-value ≤ α i.e 0.0249 < 0.05 hence *we reject null hypothesis.*
There is enough evidence to support alternative hypotheses.
We can conclude that there is at least one grain type that has a different mean than another grain type.

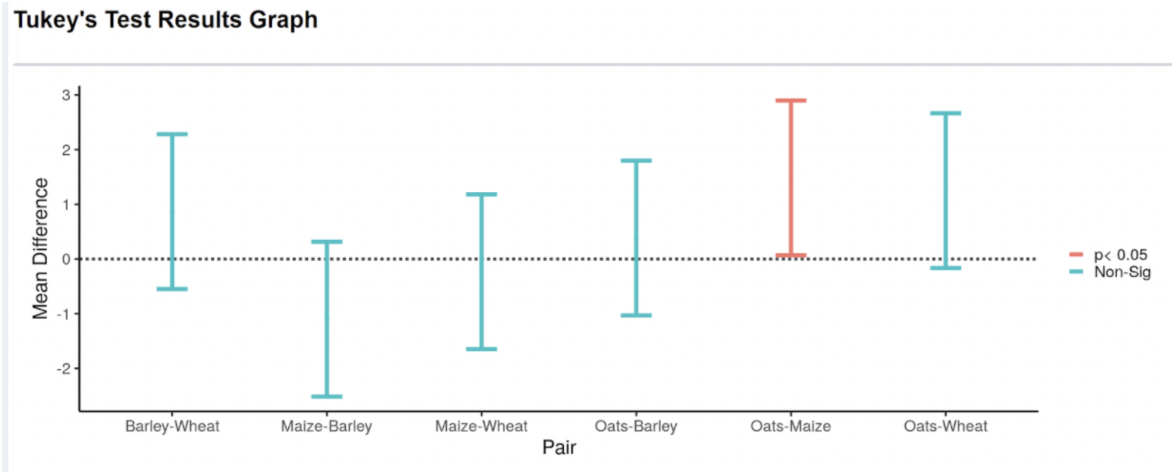
d) Determine which means are different

Tukey's Test Results Table

Pairs	Mean Difference	Lower CB	Upper CB	P-Value
Barley-Wheat	0.8667	-0.5487	2.282	0.3429
Maize-Wheat	-0.2333	-1.649	1.182	0.9666
Oats-Wheat	1.25	-0.1653	2.665	0.0955
Maize-Barley	-1.1	-2.515	0.3153	0.1644
Oats-Barley	0.3833	-1.032	1.799	0.8721
Oats-Maize	1.483	0.068	2.899	0.0379

R^2=$(1 - SSE/SST) * 100$
$= (1 - \frac{15.342}{24.23}) * 100$
R^2=36.682
Using the $\alpha = 0.05$ew can see that all the pairs are significant except Oats-Maize.

Since we have found the significance now, we will determine which pairs are different.
We will get this result using the pairwise Tukey test.



We can see from this plot that the only pair Oats-Maize does not cross the horizontal line at 0. So we can conclude that the only pair that does have a significant difference in the mean is Oats-Maize.
All the other pairs do not have a significant difference in mean as they are crossing the horizontal line at 0.

e) Draw conclusions to the real-world

Since we have rejected the null hypothesis, we can conclude that the thiamine content of the several grain growth in different regions are not significantly similar and the place or the region of growth does affect the mineral content of the grains.

f) Did the outcome match what you thought would happen based on your plot, explain

The outcome did not match what I thought would happen based on the plot because using the plot we can infer that one pair has a significant difference of mean in the thiamin content amongst 6 pairs of grains. By rejecting the null hypothesis, we infer that the means won't be equal but with the help of Tukey test results and the plot of the Tukey test we can see that the means are different but they don't have any significant difference among them.