To: Professor Yuce

Applied Mathematics

New York City College of Technology

From: Brian Holliday

Applied Mathematics

New York City College of Technology

Subject: Lab 23: : Two Way ANOVA

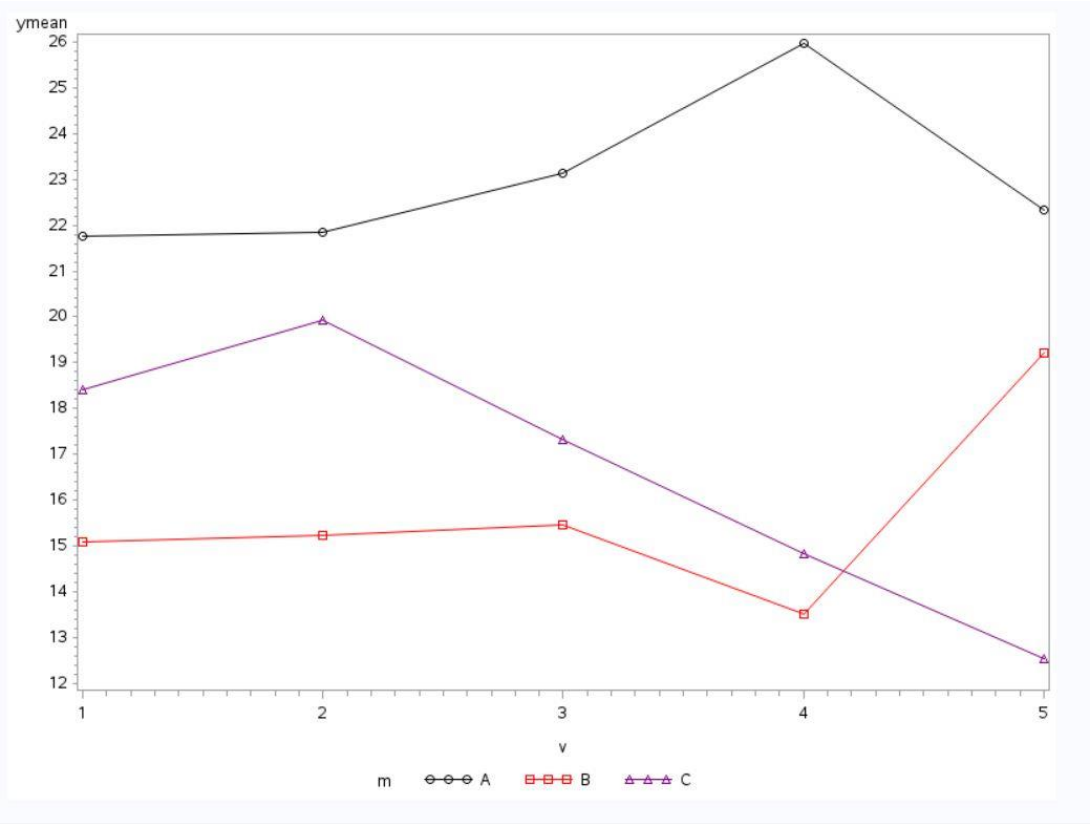
Figure:

Date: 5/5/20

Summary:

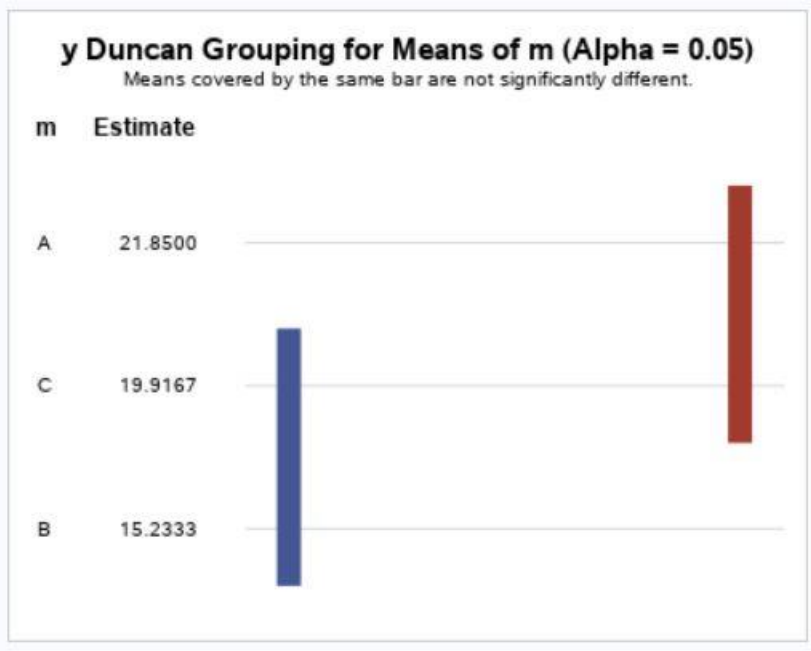
Two-way ANOVA is like one-way ANOVA. We can use it to check if there is a statistical difference in the means of multiple data sets at the same time. Two-way ANOVA is different in that it can account for more than one classification between groups. In this case we will be working with data on grass. We will be checking on the difference in the means between 3 different methods of growing seeds and 5 difference varieties of grass.

Figure 1: Variety vs Growth



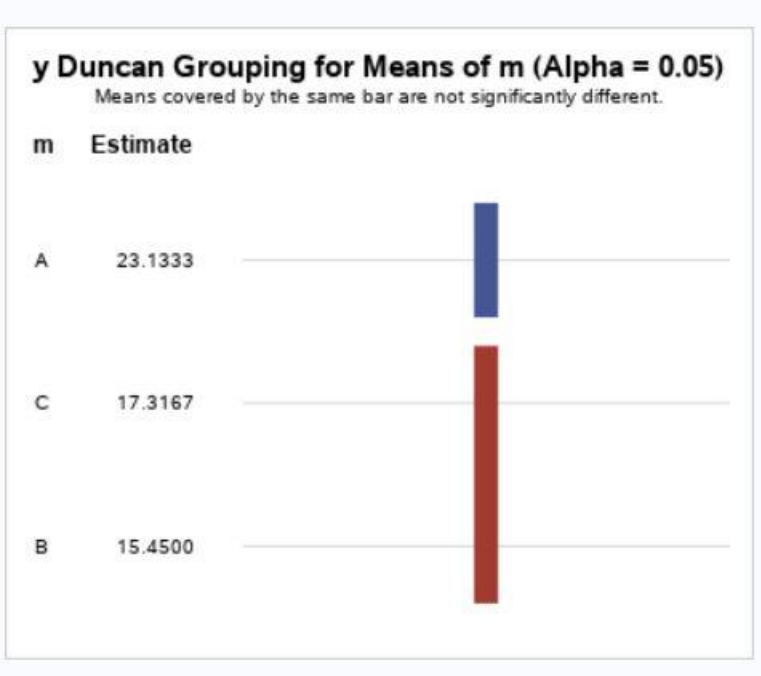
This is a graph of the means of the variety of the grass vs the method of growth. In the y-axis we are measuring the mean yield of the grass with respect to the method. This graph reveals that A is the best method for the highest yield, clearly staying above both methods B and C across all grass seed varieties. The results of the better method between B and C. The plots are not parallel. The yield seems to suggest that it depends on the variety and the type of seed.

Figure 2: Duncan Method Test Variety 2



The Duncan test is a test for the mean of the method by each variety. This chart is of variety 2 of grass seed. From this chart we can see the red bar shows that the mean of A and C are the statistical the same. The blue bar shows that B and C are statistically the same. Therefore, by the association rule we can say that the means of A and B are the same.

Figure 3: Duncan Method Test Variety 3



From this Duncan test in variety 3 of grass seed, we can see that the means of B and C are similar. Both B and C have different means than A. So, there is a clear difference in the yield of variety 3 with grass seed A and the other methods B and C with variety 3.

Figure 4: Variety Boxplot

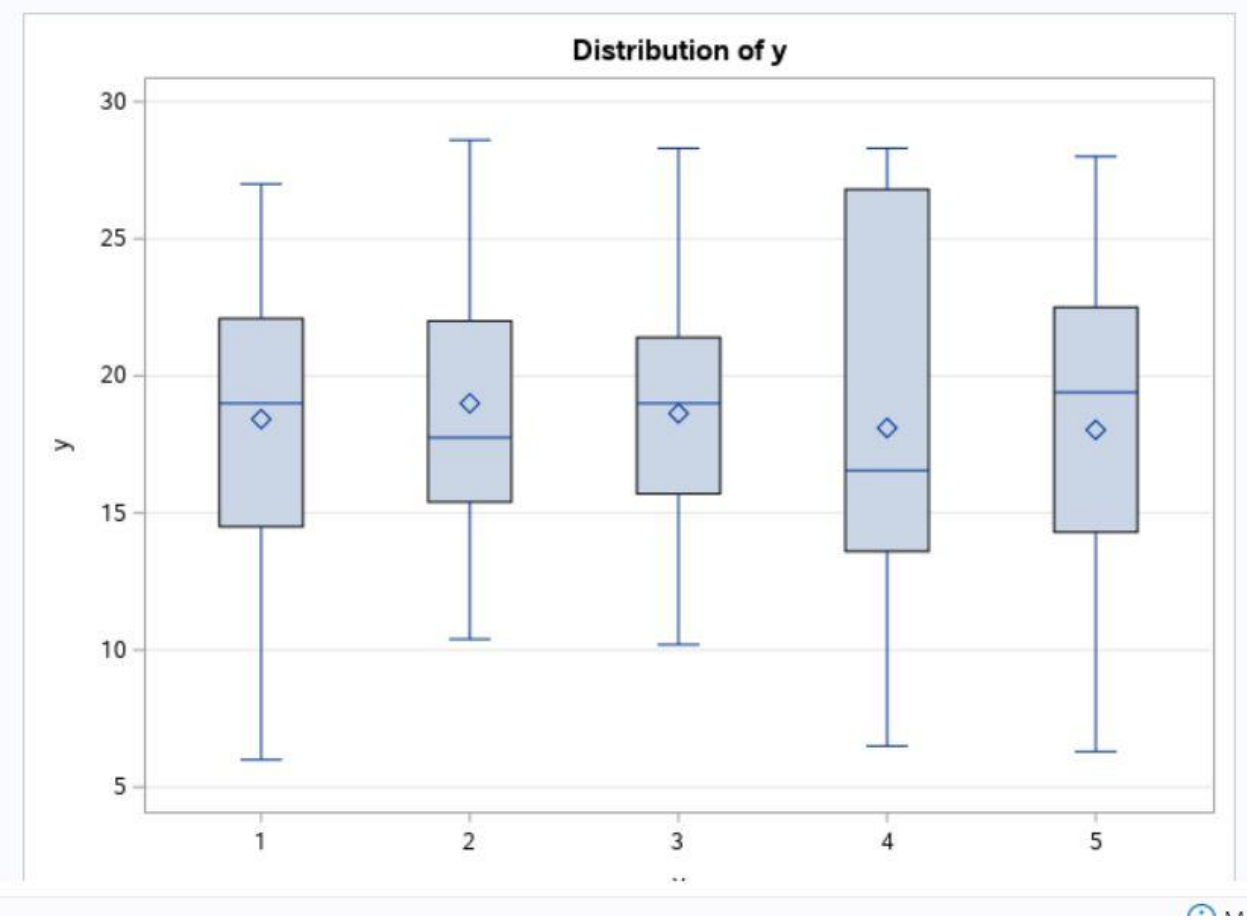
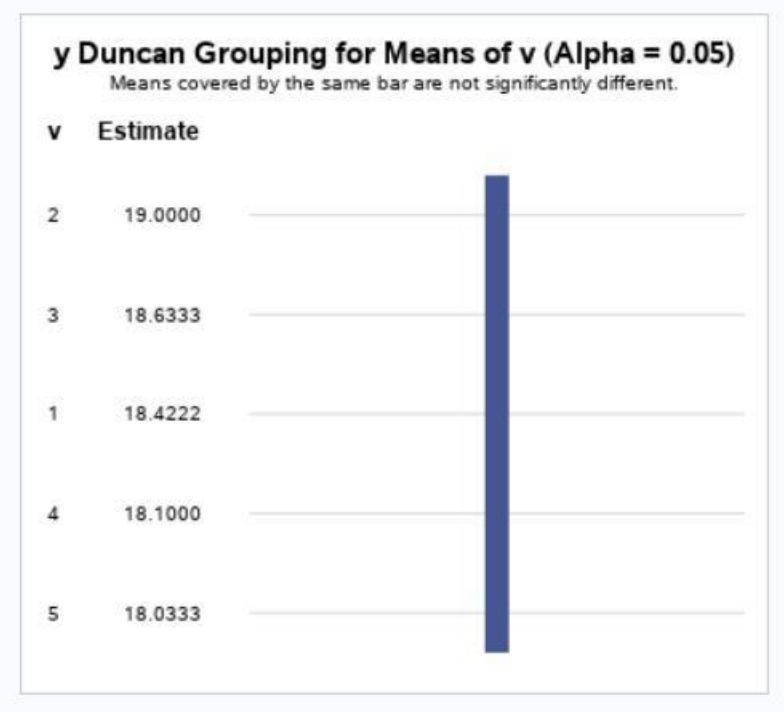


Figure 5: Duncan on Grass Variety



As we can see by the boxplot and Duncan test on the variety, there does not seem to be a statistical difference between which variety is used. But can the variety make a different when the method is isolated.

Figure 6: Mean by Method

The MEANS Procedure

m=A

Variable	Mean
v	3.0000000
TYPE	0
FREQ	6.0000000
ymean	23.0100000

m=B

Variable	Mean
v	3.0000000
TYPE	0
FREQ	6.0000000
ymean	15.6966667

m=C

Variable	Mean
v	3.0000000
TYPE	0
FREQ	6.0000000
ymean	16.6066667

From the PROC means procedure we can see that method B has the lowest mean but is method B salvageable or should we just throw it out completely. Let us see if there is a B method that works better than the others. We saw that A had the high overall mean. Let's check if there is a statistical difference between the highest B and the lowest A.

This is the code we will use to compare to the two:

Figure 7: B5 vs A5

```
PROC glm data = grass ;
    class m v ;
    model y = m v  m*v;
    means v / duncan;
    estimate 'B5' intercept 1 m 0 1 0 v 0 0 0 0 1 m*v 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0;
    estimate '.5B5 + .5A5' intercept 1 m .5 .5 0 v 0 0 0 0 1 m*v 0 0 0 0 .5 0 0 0 0 .5 0 0 0 0 0;
    estimate '.5B5 + .5A5 - C5' m .5 .5 -1 m*v 0 0 0 0 .5 0 0 0 0 .5 0 0 0 0 -1/ e;
run;
```

Figure 8: GLM A5 And B5

The GLM Procedure

Dependent Variable: y

Parameter	Estimate	Standard Error	t Value	Pr > |t|
B5	19.2166667	1.80970634	10.62	<.0001
.5B5 + .5A5	20.7750000	1.27965562	16.23	<.0001
.5B5 + .5A5 - C5	8.2250000	2.21642856	3.71	0.0004

For all estimates we reject the null hypothesis that the means are equal. Meaning that there is a significant difference between B5 and A5. This does not mean that B5 is not usable, we might need to run other tests.