**Multiple Comparisons**

**If our test of the null hypothesis is rejected**, we conclude that not all the means are equal: that is, at least one mean is different from the other means. The ANOVA test itself provides only statistical evidence of a difference, but not any statistical evidence as to which mean or means are statistically different.

For instance, using the previous example for tar content, if the ANOVA test results in a significant difference in average tar content between the cigarette brands, a follow up analysis would be needed to determine which brand mean or means differ in tar content. Plus we would want to know if one brand or multiple brands were better/worse than another brand in average tar content. To complete this analysis we use a method called ***multiple comparisons.***

Multiple comparisons conduct an analysis of all possible pairwise means. For example, with three brands of cigarettes, A, B, and C, if the ANOVA test was significant, then multiple comparison methods would compare the three possible pairwise comparisons:

* Brand A to Brand B
* Brand A to Brand C
* Brand B to Brand C

These are essentially tests of two means similar to what we learned previously in our lesson for comparing two means. However, the methods here use an adjustment to account for the number of comparisons taking place. Minitab provides three adjustment choices. We will use the ***Tukey***adjustment which is an adjustment on the t-multiplier based on the number of comparisons.

**Note!** We don’t go in the theory behind the Tukey method. Just note that we only use a multiple comparison technique in ANOVA when we have a significant result.

Example:

20 young cows are assigned at random among 4 experimental groups. Each group is fed a different diet. (This design is a completely randomized design.) The data are the cow's weight, in kilograms, after being raised on these diets for 10 months ([cows\_weights.txt](https://online.stat.psu.edu/stat500/sites/stat500/files/pig_weights.txt)). We wish to determine whether the mean cows weights are the same for all 4 diets.

Feed 1 Feed 2 Feed 3 Feed 4

60.8 68.3 102.6 87.9

57.1 67.7 102.2 84.7

65 74 100.5 83.2

58.7 66.3 97.5 85.8

61.8 69.9 98.9 90.3

##### **Analysis of Variance**

| **Source** | **DF** | **Adj SS** | **Adj MS** | **F-Value** | **P-Value** |
| --- | --- | --- | --- | --- | --- |
| Factor | 3 | 4703.2 | 1567.73 | 206.72 | 0.000 |
| Error | 16 | 121.3 | 7.58 |   |   |
| Total | 19 | 4824.5 |   |   |   |

The p-value for the test is less than 0.001. With a significance level of 5%, we reject the null hypothesis. The data provide sufficient evidence to conclude that the mean weights of cows from the four feeds are not all the same.

With a rejection of the null hypothesis leading us to conclude that not all the means are equal (i.e., at least the mean cow weight or one diet differs from the mean cow weight from the other diets) some follow up questions are:

* "Which diet type results in different average cow weights?", and
* "Is there one particular diet type that produces the largest/smallest mean weight?"

To answer these questions we analyze the multiple comparison output (the grouping information)

#### ****Tukey Pairwise Comparisons****

##### **Grouping Information Using the Tukey Method and 95% Confidence**

| **Factor** | **N** | **Mean** | **Grouping** |
| --- | --- | --- | --- |
| Feed 3 | 5 | 100.340 | A |   |   |   |  |
| Feed 4 | 5 | 86.38 |   | B |   |   |  |
| Feed 2 | 5 | 69.24 |   |   | C |   |  |
| Feed 1 | 5 | 60.68 |   |   |   | D |  |

Means that do not share a letter are significantly different.

Each of these factor levels is associated with a grouping letter. If any factor levels have the same letter, then the multiple comparison method did not determine a significant difference between the mean responses. For any factor level that does not share a letter, a significant mean difference was identified. From the lettering we see each Diet Type has a different letter, i.e. no two groups share a letter. Therefore, we can conclude that all four diets resulted in statistically significant different mean cows weights.

Furthermore, with the order of the means also provided from highest to lowest, we can say that Feed 3 resulted in the highest mean weight followed by Feed 4, then Feed 2, then Feed 1.

Try yourself on the following data sets:

We want to see whether the tar contents (in milligrams) for three different brands of cigarettes are different. Two different labs took samples, Lab Precise and Lab Sloppy.

#### Lab Precise

Lab Precise took six samples from each of the three brands and got the following measurements:

| **Sample** | **Brand A** | **Brand B** | **Brand C** |
| --- | --- | --- | --- |
| **1** | 10.21 | 11.32 | 11.60 |
| **2** | 10.25 | 11.20 | 11.90 |
| **3** | 10.24 | 11.40 | 11.80 |
| **4** | 9.80 | 10.50 | 12.30 |
| **5** | 9.77 | 10.68 | 12.20 |
| **6** | 9.73 | 10.90 | 12.20 |
|  |  |  |  |

#### Lab Sloppy

Lab Sloppy also took six samples from each of the three brands and got the following measurements:

| **Sample** | **Brand A** | **Brand B** | **Brand C** |
| --- | --- | --- | --- |
| **1** | 9.03 | 9.56 | 10.45 |
| **2** | 10.26 | 13.40 | 9.64 |
| **3** | 11.60 | 10.68 | 9.59 |
| **4** | 11.40 | 11.32 | 13.40 |
| **5** | 8.01 | 10.68 | 14.50 |
| **6** | 9.70 | 10.36 | 14.42 |