

# Statistics By Jim

Making statistics intuitive

Basics Hypothesis Testing Regression ANOVA Fun Glossary Blog My Store

## Using Post Hoc Tests with ANOVA

By Jim Frost — 34 Comments

Post hoc tests are an integral part of ANOVA. When you use ANOVA to test the equality of at least three group means, statistically significant results indicate that not all of the group means are equal. However, ANOVA results do not identify which particular differences between pairs of means are significant. Use post hoc tests to explore differences between multiple group means while controlling the experiment-wise error rate.

In this post, I'll show you what post hoc analyses are, the critical benefits they provide, and help you choose the correct one for your study. Additionally, I'll show why failure to control the experiment-wise error rate will cause you to have severe doubts about your results.

### Starting with the ANOVA Omnibus Test

Typically, when you want to determine whether three or more means are different, you'll perform ANOVA. Statisticians refer to the ANOVA F-test as an omnibus test. Welch's ANOVA is another type of omnibus test.

An omnibus test provides overall results for your data. Collectively, are the differences between the means statistically significant—Yes or No?

If the p-value from your ANOVA F-test or Welch's test is less than your significance level, you can reject the null hypothesis.

- Null: All group means are equal.
- Alternative: Not all group means are equal.

However, ANOVA test results don't map out which groups are different from other groups. As you can see from the hypotheses above, if you can reject the null, you only know that not all of the means are equal. Sometimes you really need to know which groups are significantly different from other groups!

**Related posts:** [How F-tests Work in ANOVA](#) and [Welch's ANOVA](#)

### Example One-Way ANOVA to Use with Post Hoc Tests

We'll start with this one-way ANOVA example, and then use it as the basis for illustrating three different post hoc tests throughout this blog post. Imagine we are testing four materials that we're considering for making a product part. We want to determine whether the mean differences between the strengths of these four materials are statistically significant. We obtain the following one-way ANOVA results. To follow along with this example, download the CSV dataset: [PostHocTests](#).

#### One-way ANOVA: Strength versus Material

Method

Null hypothesis All means are equal  
Alternative hypothesis At least one mean is different  
Significance level  $\alpha = 0.05$

Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Material	4	A, B, C, D

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Material	3	281.7	93.90	6.02	0.004
Error	20	312.1	15.60		
Total	23	593.8			

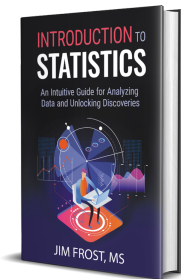
#### Meet Jim



I'll help you intuitively understand statistics by focusing on concepts and using plain English so you can concentrate on understanding your results.

[Read More...](#)

#### Buy My Introduction to Statistics eBook!



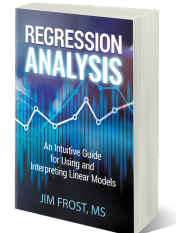
#### Introduction to Statistics: An Intuitive Guide [ebook]

\$9.00 USD

[Buy it now](#)



#### Buy My Regression eBook!



#### Regression Analysis: An Intuitive Guide [ebook]

\$14.00 USD

[Buy it now](#)

#### Means

Material	N	Mean	StDev	95% CI
A	6	37.73	3.36	(34.37, 41.10)
B	6	31.57	5.50	(28.20, 34.93)
C	6	35.98	3.73	(32.62, 39.35)
D	6	41.07	2.64	(37.70, 44.43)

The p-value of 0.004 indicates that we can reject the null hypothesis and conclude that the four means are not all equal. The Means table at the bottom displays the group means. However, we don't know which pairs of groups are significantly different.

To compare group means, we need to perform post hoc tests, also known as multiple comparisons. In Latin, post hoc means "after this." You conduct post hoc analyses after a statistically significant omnibus test (F-test or Welch's).

Before we get to these group comparisons, you need to learn about the experiment-wise error rate.

**Related posts:** [How to Interpret P-values Correctly](#) and [How to do One-Way ANOVA in Excel](#)

## What is the Experiment-wise Error Rate?

Post hoc tests perform two vital tasks. Yes, they tell you which group means are significantly different from other group means. Crucially, they also control the experiment-wise, or familywise, error rate. In this context, experiment-wise, family-wise, and family error rates are all synonyms that I'll use interchangeably.

What is this experiment-wise error rate? For every hypothesis test you perform, there is a type I error rate, which your significance level (alpha) defines. In other words, there's a chance that you'll reject a null hypothesis that is actually true—it's a false positive. When you perform only one test, the type I error rate equals your significance level, which is often 5%. However, as you conduct more and more tests, your chance of a false positive increases. If you perform enough tests, you're virtually guaranteed to get a false positive! The error rate for a family of tests is always higher than an individual test.

Imagine you're rolling a pair of dice and rolling two ones (known as snake eyes) represents a Type I error. The probability of snake eyes for a single roll is ~2.8% rather than 5%, but you get the idea. If you roll the dice just once, your chances of rolling snake eyes aren't too bad. However, the more times you roll the dice, the more likely you'll get two ones. With 25 rolls, snake eyes become more likely than not (50.8%). With enough rolls, it becomes inevitable.



**Related posts:** [Types of Errors in Hypothesis Testing](#) and [Significance Levels and P-values](#)

## Family Error Rates in ANOVA

In the ANOVA context, you want to compare the group means. The more groups you have, the more comparison tests you need to perform. For our example ANOVA with four groups (A B C D), we'll need to make the following six comparisons.

- A – B
- A – C
- A – D
- B – C
- B – D
- C – D

Our experiment includes this family of six comparisons. Each comparison represents a roll of the dice for obtaining a false positive. What's the error rate for six comparisons? Unfortunately, as you'll see next, the experiment-wise error rate snowballs based on the number of groups in your experiment.

## The Experiment-wise Error Rate Quickly Becomes Problematic!

The table below shows how increasing the number of groups in your study causes the number of comparisons to rise, which in turn raises the family-wise error rate. Notice how quickly the quantity of comparisons increases by adding just a few groups! Correspondingly, the experiment-wise error rate rapidly becomes problematic.



Search this website

## Subscribe by Email

Enter your email address to receive notifications of new posts by email.

Your email address

First Name

Subscribe

I won't send you spam. Unsubscribe at any time.

## Follow Me



### Popular

### Latest

[How To Interpret R-squared in Regression Analysis](#)

[How to Interpret P-values and Coefficients in Regression Analysis](#)

[Measures of Central Tendency: Mean, Median, and Mode](#)

[Normal Distribution in Statistics](#)

[Understanding Interaction Effects in Statistics](#)

[How to Interpret the F-test of Overall Significance in Regression Analysis](#)

[Multicollinearity in Regression Analysis: Problems, Detection, and Solutions](#)

## Recent Comments

Jim Frost on [Coronavirus Curves and Different Outcomes](#)

Sasidharan G on [Coronavirus Curves and Different Outcomes](#)

Jim Frost on [Understanding Probability Distributions](#)

The table starts with two groups, and the single comparison between them has an experiment-wise error rate that equals the significance level (0.05). Unfortunately, the family-wise error rate rapidly increases from there!

All Pairwise Comparisons Alpha = 0.05		
Groups	Comparisons	Experimentwise Error Rate
2	1	0.05
3	3	0.142625
4	6	0.264908109
5	10	0.401263061
6	15	0.53670877
7	21	0.659438374
8	28	0.762173115
9	36	0.842220785
10	45	0.900559743
11	55	0.940461445
12	66	0.966134464
13	78	0.981700416
14	91	0.990606054
15	105	0.995418807
From StatisticsByJim.com		

The formula for the maximum number of comparisons you can make for N groups is:  $(N*(N-1))/2$ . The total number of comparisons is the family of comparisons for your experiment when you want to compare all possible pairs of groups (i.e., all pairwise comparisons). Additionally, the formula for calculating the error rate for the entire set of comparisons is  $1 - (1 - \alpha)^C$ . Alpha is your significance level for a single comparison, and C equals the number of comparisons.

The experiment-wise error rate represents the probability of a **type I** error (false positive) over the total family of comparisons. Our ANOVA example has four groups, which produces six comparisons and a family-wise error rate of 0.26. If you increase the groups to five, the error rate jumps to 40%! When you have 15 groups, you are virtually guaranteed to have a false positive (99.5%)!

## Post Hoc Tests Control the Experiment-wise Error Rate

The table succinctly illustrates the problem that post hoc tests resolve. Typically, when performing a statistical analysis, you expect a false positive rate of 5%, or whatever value you set for the significance level. As the table shows, when you increase the number of groups from 2 to 3, the error rate nearly triples from 0.05 to 0.143. And, it quickly worsens from there!

These error rates are too high! Upon seeing a significant difference between groups, you would have severe doubts about whether it was a false positive rather than a real difference.

If you use 2-sample t-tests to compare all group means in your study systematically, you'll encounter this problem. You'd set the significance level for each test (e.g., 0.05), and then the number of comparisons will determine the experiment-wise error rate, as shown in the table.

Fortunately, post hoc tests use a different approach. For these tests, you set the experiment-wise error rate you want for the entire set of comparisons. Then, the post hoc test calculates the significance level for all individual comparisons that produces the familywise error rate you specify.

Understanding how post hoc tests work is much simpler when you see them in action. Let's get back to our one-way ANOVA example!

## Example of Using Tukey's Method with One-Way ANOVA

For our ANOVA example, we have four groups which require six comparisons to cover all combinations of groups. We'll use a post hoc test and specify that the family of six comparisons should collectively produce a familywise error rate of 0.05. The post hoc test I'll use is Tukey's method. There are a variety of post hoc tests you can choose from, but Tukey's method is the most common when you want to compare all possible group pairings.

There are two ways to present post hoc test results—adjusted p-values and simultaneous confidence intervals. I'll show both below.

### Adjusted P-values

The table below displays the six different comparisons in our study, the difference between group means, and the adjusted p-value for each comparison.

Jim Frost on [Repeated Measures Designs: Benefits and an ANOVA Example](#)

Elijah on [Repeated Measures Designs: Benefits and an ANOVA Example](#)

#### Tukey Simultaneous Tests for Differences of Means

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
B - A	-6.17	2.28	(-12.55, 0.22)	-2.70	0.061
C - A	-1.75	2.28	(-8.14, 4.64)	-0.77	0.868
D - A	3.33	2.28	(-3.05, 9.72)	1.46	0.478
C - B	4.42	2.28	(-1.97, 10.80)	1.94	0.245
D - B	9.50	2.28	(3.11, 15.89)	4.17	0.002
D - C	5.08	2.28	(-1.30, 11.47)	2.23	0.150

Individual confidence level = 98.89%

The adjusted p-value identifies the group comparisons that are significantly different while limiting the family error rate to your significance level. Simply compare the adjusted p-values to your significance level. When adjusted p-values are less than the significance level, the difference between those group means is statistically significant. Importantly, this process controls the family-wise error rate to your significance level. We can be confident that this entire set of comparisons collectively has an error rate of 0.05.

In the output above, only the D - B difference is statistically significant while using a family error rate of 0.05. The mean difference between these two groups is 9.5.

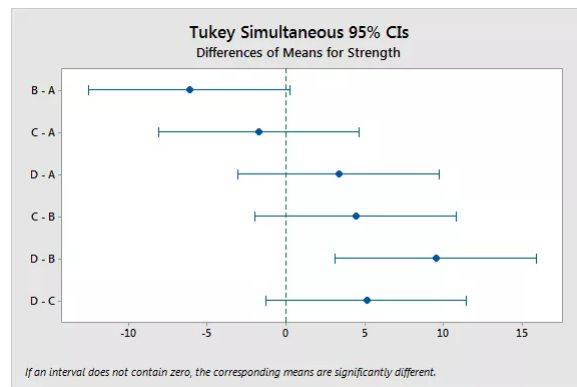
#### Simultaneous Confidence Intervals

The other way to present post hoc test results is by using simultaneous confidence intervals of the differences between means. In an individual test, the hypothesis test results using a significance level of  $\alpha$  are consistent with confidence intervals using a confidence level of  $1 - \alpha$ . For example, hypothesis tests with a significance level of 0.05 correspond to 95% confidence intervals.

In post hoc tests, we use a simultaneous confidence level rather than an individual confidence level. The simultaneous confidence level applies to the entire family of comparisons. With a 95% simultaneous confidence level, we can be 95% confident that *all* intervals in our set of comparisons contain the actual population difference between groups. A 5% experiment-wise error rate corresponds to 95% simultaneous confidence intervals.

#### Tukey Simultaneous CIs for our One-Way ANOVA Example

Let's get to the confidence intervals. While the table above displays these CIs numerically, I like the graph below because it allows for a simple visual assessment and it provides more information than the adjusted p-values.



Zero indicates that the group means are equal. When a confidence interval does not contain zero, the difference between that pair of groups is statistically significant. In the chart, only the difference between D - B is significant. These CI results match the hypothesis test results in the previous table. I prefer these CI results because they also provide additional information that the adjusted p-values do not convey.

These confidence intervals provide ranges of values that likely contain the actual population difference between pairs of groups. As with all CIs, the width of the interval for the difference reveals the precision of the estimate. Narrower intervals suggest a more precise estimate. And, you can assess whether the full range of values is practically significant. Remember, statistical significance doesn't necessarily indicate that the results are meaningful in the real world.

When the interval is too wide (imprecise) to be helpful and/or the range includes differences that are not practically significant, you have reason to hesitate before making decisions based on the results.

#### Post Hoc Tests and the Statistical Power Tradeoff

Post hoc tests are great for controlling the family-wise error rate. Many texts would stop at this point. However, a tradeoff occurs behind the scenes. You need to be aware of it because you might be able to manage it effectively. The tradeoff is the following:

Post hoc tests control the experiment-wise error rate by reducing the statistical power of the comparisons.

Here's how that works and what it means for your study.

To obtain a lower family error rate, the procedures must lower the significance level for all individual comparisons. For example, to end up with a family error rate of 5% for a set of comparisons, the procedure uses an even lower individual significance level.

As the number of comparisons increases, the post hoc analysis is forced to lower the individual significance level even further. For our six comparisons, Tukey's method uses an individual significance level of approximately 0.011 to produce the family-wise error rate of 0.05. If our ANOVA required more comparisons, it would be even lower.

What's the problem with using a lower individual significance level? Lower significance levels correspond to lower statistical power. If a difference between group means actually exists in the population, a study with lower power is less likely to detect it. You might miss important findings!

Avoiding this power reduction is why many studies use an individual significance level of 0.05 rather than 0.01. Yet, with just four groups, our example post hoc test is forced to use the lower significance level.

**Key Takeaway:** The more group comparisons you make, the lower the statistical power of those comparisons.

**Related post:** [Understanding Statistical Power](#)

## Managing the Power Tradeoff in Post Hoc Tests by Reducing the Number of Comparisons

One method to mitigate this tradeoff is by reducing the number of comparisons. If you reduce the number of comparisons, the procedure can use a larger individual error rate to achieve the family error rate that you specify—which increases the statistical power.

Throughout this article, I've written about performing all pairwise comparisons—which compares all possible group pairings. While this is the most common approach, the number of comparisons quickly piles up! However, depending on the nature and purpose of your study, you might not need to compare all possible groups.

Your study might need to compare only a subset of all possible comparisons for a variety of reasons. I'll cover two common reasons and show you which post hoc tests you can use. In the following examples, I'll display only the confidence interval graphs and not the hypothesis test results. Notice how these other methods make fewer comparisons (3 and 4) for our example dataset than Tukey's method (6).

While you're designing your study, it's crucial that you define in advance the multiple comparisons method that you will use. Don't try various methods, and then choose the one that produces the most favorable results. That's data dredging, and it can lead to spurious findings. I'm using multiple post hoc tests on a single dataset to show how they differ, but that's not an appropriate practice for a real study. [Define your methodology in advance](#), including one post hoc analysis, before analyzing the data, and stick to it!

**Key Takeaway:** When it's possible, compare a subset of groups to increase your statistical power.

## Example of Using Dunnett's Method to Compare Treatment Groups to a Control Group

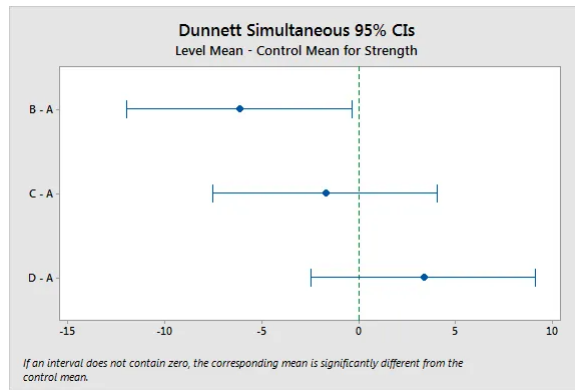
If your study has a control group and several treatment groups, you might need to compare the treatment groups only to the control group.

Use Dunnett's method when the following are true:

- You know in advance which group (control) you want to compare to all the other groups (treatments).
- You don't need to compare the treatment groups to each other.

Let's use Dunnett's method with our example one-way ANOVA, but we'll tweak the scenario slightly. Suppose we currently use Material A. We performed this experiment to

compare the alternative materials (B, C, and D) to it. Material A will be our control group while the other three are the treatments.



Using Dunnett's method, we see that only the B - A difference is statistically significant because the interval does not include zero. Using Tukey's method, this comparison was not significant. The additional power gained by making fewer comparisons came through for us. On the other hand, unlike Tukey's method, Dunnett's method does not find that the D - B difference is significant because it doesn't compare the treatment groups to each other.

### Example of Using Hsu's MCB to Find the Strongest Material

If the purpose of your study is to identify the best group, you might not need to compare all possible groups. Hsu's Multiple Comparisons to the Best (MCB) identifies the groups that are the best, insignificantly different from the best, and significantly different from the best.

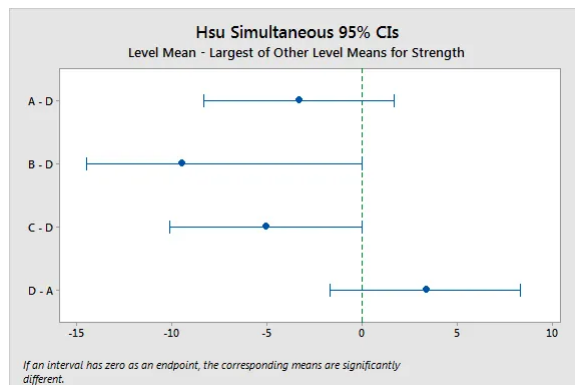
Use Hsu's MCB when you:

- Don't know in advance which group you want to compare to all the other groups.
- Don't need to compare groups that are not the best to other groups that are not the best.
- Can define "the best" as either the group with the highest mean or the lowest mean.

Hsu's MCB compares each group to the group with the best mean (highest or lowest). Using this procedure, you might end up with several groups that are not significantly different than the best group. Keep in mind that the group that is truly best in the entire population might not have the best sample mean due to sampling error. The groups that are not significantly different from the best group could potentially be as good as or even better than the group with the best sample mean.

### Simultaneous Confidence Intervals for Hsu's MCB

For our one-way ANOVA, we want to use the material that produces the strongest parts. Consequently, we'll use Hsu's MCB and define the highest mean as the best. We don't care about all of the other possible comparisons.



Group D is the best group overall because it has the highest mean (41.07). The procedure compares D to all of the other groups. For Hsu's MCB, a group is significantly better than another group when the confidence interval has zero as an endpoint. From

the graph, we can see that Material D is significantly better than B and C. However, the A-D comparison contains zero, which indicates that A is not significantly different from the best.

Hsu's MCB determines that the candidates for the best group are A and D. D has the highest sample mean, while A is not significantly different from D. On the other hand, the procedure effectively rules out B and C from being the best.

## Recap of Using Multiple Comparison Methods

In this blog post, you've seen how the omnibus ANOVA test determines whether means are different in general, but it does not identify specific group differences that are statistically significant.

If you obtain significant ANOVA results, use a post hoc test to explore the mean differences between pairs of groups.

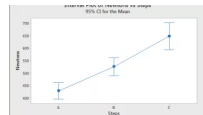
You've also learned how controlling the experiment-wise error rate is a crucial function of these post hoc tests. These family error rates grow at a surprising rate!

Finally, if you don't need to perform all pairwise comparisons, it's worthwhile comparing only a subset because you'll retain more statistical power.

### Share this:

Share 637   Share   Tweet   Save

### Related



Benefits of Welch's ANOVA Compared to the Classic One-Way ANOVA In "ANOVA"



Multivariate ANOVA (MANOVA) Benefits and When to Use It In "ANOVA"



How to do One-Way ANOVA in Excel In "ANOVA"

Filed Under: [ANOVA](#)

Tagged With: [analysis example](#), [choosing analysis](#), [conceptual](#), [graphs](#), [interpreting results](#)

## Comments



sara says

May 28, 2020 at 9:48 am

after a significant interaction, follow-up tests were done to explore the exact nature of the interaction. and i found an effects of one independent variable within one level of a second independent variable, so only one effect was significant and the other 2 effects were not, do i still conduct post hocs? since one simple main effect was significant?

★ Loading...

[Reply](#)



Jim Frost says

May 29, 2020 at 4:13 pm

Hi Sara, if you want to determine which pairs of groups specifically have significant differences, you'll need to perform a post hoc test.

★ Loading...

[Reply](#)



erika zafeiraki says

May 19, 2020 at 1:58 pm

Hey Jim, thank you for the info. However, i have a question regarding post-hoc. I have 13 groups (locations) with different size (from 2 concentrations to one group to 20 to another group). More specifically i have samples from different places and the concentration of metals in them. The number of samples is not the same for each location.

Elements concentrations were both normally distributed and homogeneous, so i further applied one-way ANOVA and a statistically significant difference was observed. So, i need to apply a post-hoc test but i don't know which one. I applied Scheffe, Bonferroni and LSD ...but i am not sure which one is the best. So, it would be really helpful if you could tell me which one to apply.

Thank you in advance

Erik

★ Loading...

[Reply](#)



Jim Frost says  
May 19, 2020 at 5:18 pm

Hi Erika,

That's an impossible question to answer in general. It depends on the specifics of what you want to learn. Do you want to compare all possible groups to each other, just compare treatment groups to a control group, or just find out which group is best and not significantly different from the best? It's your subject area knowledge combined with what you need to learn that determines which method is best.

You can rule out Fisher's LSD because that only maintains the **Type I** error rate at your significance level when you have three groups. Bonferroni compares all possible groups but is known to be conservative. If you want to compare all possible groups, I'd consider Tukey's method. Although, be aware that with 13 groups that's 78 comparisons if you compare all groups. That would really lower your statistical power as I describe in this article.

I know less about Scheffe's test than the others. But, I gather it's good when you want to make many comparisons and they don't have equal sizes. So, Scheffe's test might be a good choice for you if you want to compare all possible pairs.

My recommendation is to determine what you really need to compare to learn what you need to learn. With so many groups, you hopefully don't need to compare all possible combinations of groups.

★ Loading...

[Reply](#)



Agegnehu says  
May 10, 2020 at 4:02 am

Thanks Frost !

The way you put details using plain English and practical life experiences really helps much particularly for those reasonably far from the discipline of statistics as it is true for me.

could you have time to respond to this question, please ?

how can the statistical power of the post hoc tests be calculated to know how much we are underpowered ? The details of the role of sample sizes to increase the statistical power of the post hoc tests ?

Thanks ! Age

★ Loading...

[Reply](#)



Jim Frost says  
May 11, 2020 at 12:36 am

Hi Agegnehu,

I've never seen a statistical package that calculates the power of a statistical test. However, based on the properties of the test, we know that you lose power with more comparisons for the reasons I describe. One way to get an idea of how much



power you're losing is by looking at the individual confidence level for a set of comparisons. For example, in the section about Tukey's method, the output indicates that for the six comparison, the procedure uses an individual confidence level of 98.89%. Using that individual confidence level for each of the six comparisons collectively yields the 95% joint confidence.

If you convert that confidence level to the equivalent significance level, you'd see that it's as if you were using an alpha of  $1 - 0.9889 = 0.0111$  for each comparison. Any time you use a lower significance level, it decreases the power of the test.

I don't know if you could plug that information, along with other details, into statistical software for say a 2-sample t-test to get a valid power estimate for a single comparison or not. I've never looked into it. But, looking at that individual confidence level gives you an idea of how the procedure needs to effectively lower the significance level more and more for additional comparisons.

★ Loading...

[Reply](#)



Seble says

April 15, 2020 at 10:25 am

Thanks so much for a detailed response, Jim. That definitely helped.

★ Loading...

[Reply](#)



Seble says

April 13, 2020 at 9:47 pm

Hi Jim,

In what circumstances would it be acceptable to report a post-hoc multiple comparison while the main effect of ANOVA is not significant?

★ Loading...

[Reply](#)



Jim Frost says

April 14, 2020 at 10:54 pm

Hi Seble,

It is possible to obtain the situation you describe because the F-test for the main effect and the post hoc tests use different procedures and assess statistical significance differently.

This might be a bit of a controversial area. I'm not sure. In my mind, it is often valid to report post hoc multiple comparisons that are significant even when the main effect is not. However, be sure to report the full circumstances surrounding what is significant and what is not significant, along with the post hoc details. One caution, be sure you aren't just picking specific groups to compare because the overall main effect is not significant.

If you have a number of groups that are not very different but say a couple of groups that appear to have a large difference, it's not valid to intentionally choose a post hoc method that compares just those groups with larger differences. That's cherry picking your analysis to get the desired results, which gives misleading results. Choose your post hoc multiple comparison methodology at the beginning of your study and stick with it. Don't cherry pick the methodology later.

Another caution, in my experience when this happens, it's because the overall evidence is weak. You're probably just barely getting significant results, which represents fairly weak evidence. Pay particular attention to the CIs of the differences between group means to get an idea of the precision. CIs are often wide with weak evidence.

I hope this helps!

★ Loading...

[Reply](#)



Kumar C says

April 11, 2020 at 12:23 pm

Hi Jim, I have a question with respect to Tukey's HSD test. I have 3 groups lets say A, B, C and I have to prove that exactly one of the groups is significant. When I run the Tukey's HSD test after ANOVA, I am getting A-B, A-C and B-C are significantly different. Now how can I arrive at the correct conclusion that A is only significant or B is only Significant or C is only significant ?

★ Loading...

[Reply](#)



Jim Frost says

April 12, 2020 at 5:11 pm

Hi Kumar,

I'm not sure that I understand your question. But, if you're using Tukey's test, the number of pairs of groups that are significant different is determined by the data. You can't tell the test to find just one group. Tukey's will compare all possible group pairings and tell you which ones have differences that are statistically significant. That won't necessarily just be one pair of groups.

You might be thinking of something like Hsu's test, which I cover in this post. It takes the best group (defined by either the highest or lowest mean) and then compares that group to all other groups. Even then you might well find that the best group is significantly better than multiple other groups.

★ Loading...

[Reply](#)



Wittawat Chantkran says

April 9, 2020 at 2:42 pm

Dear Jim,

Using post hoc Tukey's HSD, I'm trying to reduce the number of comparisons by comparing A-B-C-D, A-B-E-F, and, A-B-G-H, separately, when A-B is mutual dataset of each experiment.

However, the result of A-B difference does not stay the same in each comparison (exactly the same of degree of freedom. In this case, although it reduces the statistical power, should I go back to A-B-C-D-E-F-G-H simultaneous comparison?

Many thanks

KK

★ Loading...

[Reply](#)



Wittawat Chantkran says

April 9, 2020 at 2:29 pm

Dear Jim,

Using post hoc Tukey's HSD, I'm trying to reduce the number of comparisons by comparing A-B-C-D, A-B-E-F, A-B-G-H, separately, when A-B is overlapped in every experiment.

Question : Why does the result of A-B difference not stay the same in each time of comparison ?

In this case, although it reduces the statistical power, should I go back to simultaneous A-B-C-D-E-F-G-H comparison ?

Many thanks,

KK

★ Loading...

[Reply](#)

0



Chris says

March 29, 2020 at 12:38 pm

Thanks for this overview. I was wondering why if you have a significant ANOVA and then run post-hoc test, in this case Tukey there was no significant difference between any groups? thanks

★ Loading...

[Reply](#)



Jim Frost says

March 29, 2020 at 7:59 pm

Hi Chris,

There are two primary reasons.

First, the F-test that ANOVA uses and the post hoc tests are assessing different things, which can lead to differing results. The F-test looks at all the differences between the means and determines whether they are collectively significant. In other words, is that entire set of differences statistically significant? The post hoc tests assess the difference between a specific pair of means. It's entirely possible for the F-test to conclude that the entire set of difference was unlikely to occur if there is no effect while the post hoc tests don't have sufficient evidence to conclude that the difference between specific pairs of means are statistically significant.

Additionally, with post hoc tests, you need to consider the fact that as the number of comparison increases, the power of the tests decrease. I explain that in the post so I won't retype it here. That power decrease doesn't apply to the F-test.

★ Loading...

[Reply](#)



Alex says

March 4, 2020 at 9:07 am

Why revealing your address is required to buy a book? Have anyone tried to buy and is there an actual book at the end of the process? 😊

★ Loading...

[Reply](#)



Jim Frost says

March 4, 2020 at 11:24 am

Hi Alex, the system requires an address to calculate taxes. I promise that I don't do anything with your address. Nothing at all.

Many people have bought both of my ebooks. If you want to see a free sample before buying, go to [My Store](#) and choose one of the free samples. No credit cards are required.

★ Loading...

[Reply](#)



Rebekah says

February 1, 2020 at 8:54 am

Hello I was wondering if you could help me with a question I have. What exactly is it meant by lower and higher order interaction. And are there any examples of this you can give me?

★ Loading...

[Reply](#)

0



Afnan says

November 23, 2019 at 2:20 am

Thank you for the information

but I have a question if the result for LSD post hoc test significance for negative mean difference is it ok or it means something different.

★ Loading...

[Reply](#)



Jim Frost says

November 23, 2019 at 5:20 pm

Hi Afnan,

I recommend that you don't use Fisher's LSD. It does not control the family error rate, which as I show in this post, can quickly get out of hand and lead to false positives.

That said I don't see the negative mean difference as a problem itself. Just be sure you understand which mean is higher than the other mean. It's just subtract and it must be subtraction the larger mean from the smaller mean. But, please don't use Fisher's LSD!

★ Loading...

[Reply](#)



Chloe says

October 18, 2019 at 6:37 am

Very helpful!

★ Loading...

[Reply](#)



Kami says

October 16, 2019 at 12:02 pm

Thanks for this really informative post. I have a question regarding using Benjamini-Hochberg method (BH) as a post-hoc method after ANOVA. Can we use BH as a post hoc test when we do NOT have many groups? (e.g the number of pair wise comparisons are less than 10). It seems that BH method for controlling FDR is developed for working with large data sets (genomic) when you have a large number of groups. But is there any limitation for using it for low number of groups?

What about using it for two-way-ANOVA?

Thanks

★ Loading...

[Reply](#)



M N WANA says

August 27, 2019 at 12:52 pm

That is great

★ Loading...

[Reply](#)



Dr M Kaladhar says  
August 19, 2019 at 5:29 am

Dear Jim Frost! really a good analysis and helps to laymen to understand without any ambiguity! Many thanks!!

★ Loading...

[Reply](#)



Jim Frost says  
August 19, 2019 at 2:10 pm

You're very welcome! I'm happy to hear that it was helpful!

★ Loading...

[Reply](#)



Manohar Lal says  
April 25, 2019 at 12:06 am

Hi sir,  
Good information, post hoc tests with some information about Strip plot design with three factor

★ Loading...

[Reply](#)



Kelly Papapavlou says  
February 1, 2019 at 2:46 pm

Thank you!! I went through all calculations steps again and MINITAB uses a different pathway to come up with the same result.

★ Loading...

[Reply](#)



Kelly Papapavlou says  
January 22, 2019 at 6:41 am

Thank you for the enlightening post. How did you calculate the standard error of difference for the Tukey simultaneous tests? I tried to repeat the calculations using the formula  $SE = \sqrt{ME/n}$  where ME is the ANOVA table variation within the groups (-15.6) and n=sample size per group (6). I get a standard error of 1.61, not of 2.28.....

On a different issue, what is the individual confidence level??

Thank you!!

★ Loading...

[Reply](#)



Jim Frost says  
January 22, 2019 at 11:11 am

Hi Kelly,

I received your multiple comments and contacts through the contact me page. As I note on the contact page, it takes some time for the comment approval process, particularly because I'm in a very different time zone than you! Patience please!

I'm using Minitab statistical software to calculate the Tukey's test. You can see their [Method's and Formula page for Tukey's method](#) to see how it is calculated in all of

its details.

The individual confidence level is the confidence that you have that an individual group comparison falls within that interval. The simultaneous confidence level applies to the entire set of comparisons while the individual level applies to an individual comparison.

★ Loading...

[Reply](#)



Jeremy says

January 14, 2019 at 6:28 pm

This is so superbly well-written. The Hsu's MCB test is new to me. And thanks for demystifying some of the terminology (experiment-wise, family-wise, etc). Would be nice to add a discussion of Bonferonni, LSD, and the others too.

★ Loading...

[Reply](#)



Janna Beckerman says

January 14, 2019 at 9:49 am

This is great. I don't know why, but I never thought of the Craps analogy. Thank you! And thank you for comparing and contrasting Dunnett's versus Tukey's.

★ Loading...

[Reply](#)



Jim Frost says

January 14, 2019 at 3:27 pm

Thanks, Janna! And, you're very welcome! The idea of the dice analogy just popped into my head. But, I really love linking additional comparisons to rolling the dice on a false positive. It's a crapshoot!

★ Loading...

[Reply](#)



Dr Eajaz Dar says

January 14, 2019 at 4:17 am

Nice information. I would suggest to discuss other post hoc tests like DMRT and LSD along with this, to gain a clear distinction between them.

★ Loading...

[Reply](#)



Jim Frost says

January 14, 2019 at 3:29 pm

Thanks for the good suggestion. I felt covering three post hoc tests in one blog post was about the maximum for a reasonably long blog post, but I might need to write another post about it!

★ Loading...

[Reply](#)

[Leave a Reply](#)

Enter your comment here...

0