Dunnett’s test

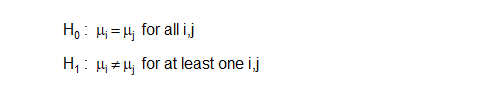
Steve Simon

2022-07-04

I’m a big fan of Analysis of Variance (ANOVA). I use it all the time. I learn a lot from it. But sometimes I want something a bit different. And the difference comes down to how you specify the hypothesis.

### The traditional hypothesis

Recall the way in which the hypothesis test is commonly set up when you are comparing three of more groups using analysis of variable.



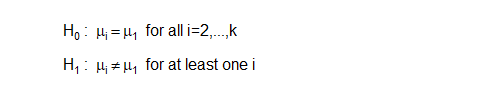
This is a very common approach in a comparison of k independent groups. Test for any deviation from the null hypothesis using an F test. If this is statistically significant, then use a Tukey follow-up test to see which pairs of means differ from one another.

This hypothesis specifies equality by specifying that every pair of means is equal. For four groups, this implies six equalities: 1=2, 1=3, 1=4, 2=3, 2=4, and 3=4. For six groups, you would have fifteen equalities; for ten you’d have forty five.

It gets messy very fast. That’s okay. If you have a lot of groups that you are comparing, you have to make it has to involve a lot of comparisons. Or do you?

### The placebo hypothesis

If one of the groups is a control or placebo, then you might consider an alternative formulation. Here’s what the hypothesis looks like, assuming that the control is group #1.



There’s a procedure for this, Dunnett’s test. It involves only three comparisons if you have four groups total, and only nine comparisons if you have ten groups total. You can quickly identify who is better than the control. This also gives you simplicity and a bit of extra power and precision.

Dunnett’s test is easy to implement. Calculate the traditional measures in analysis of variance, including mean squared error (MSE). You can skip the traditional F Test that you see in most ANOVA tables. It won’t hurt your overall alpha level, as long as you only look at each treatment versus the control. Simply compare each treatment mean minus the control mean to a cutoff value

Reject if or

where d is obtained from a table such as [this one](https://www.statology.org/dunnetts-table/) on the [Statology blog](https://www.statology.org/about/). The value of d depends on alpha (the desired Type I error rate for a two-sided test), g (the number of groups including the control group) and n (the number of data points in each group). Note that some tables define g as the number of groups excluding the control group.

Most statistical software packages will include Dunnett’s test as an option for ANOVA. Just remember that you don’t need to use the initial F test as a screen before jumping into the Dunnett’s test.

### Example-seed viability

In an experiment examining the effect of different disinfection protocols on orthodontic elastormeric ligatures, 120 ligatures were divided into 6 groups of 20. Each group received a different disinfection protocol, with the exception of the first group, which received no disinfection. This group served as a control group.



Leandro Berni Osorio, Luiz Makito Osawa Gutierrez, Eduardo Martinelli de Lima, Eduardo Gonçalves Mota, and Luciane Macedo de Menezes. Disinfection of Orthodontic Elastomers and Its Effects on Tensile Strength. Turkish Journal of Orthodontics, 2021, 35(1): 22-26.

The researchers were interested in whether any of the disinfection protocols reduced the tensile strength of the ligatures. Here are the summary statistics for the maximum strength measurement.

mean standard\_deviation  
1 21.39 3.35  
2 22.58 2.88  
3 16.94 3.50  
4 21.93 3.05  
5 19.93 1.86  
6 22.00 3.71

lower limit = - d(0.05, 6, 20)

= 21.39 - 2.73

= 18.7

upper limit = + d(0.05, 6, 20)

= 21.39 + 2.73

= 24.08

You can create a simple graph from these limits.



Only the third disinfection protocol has significantly less strength compared to the control group with no disinfection.

### Caveats

Modifications of Dunnett’s test are tricky when the group sizes are unequal.

Now you lose something when you simplify the hypothesis. Suppose you have six groups, a control and five different treatments. Now imagine that all of the treatments are significantly better than the control group. Jackpot! Every treatment is worth further study. But Dunnett’s test won’t allow you to see if some of the treatments are better than the others. There is no option for finding the best of the best.

Let’s consider as a reminder that there is no such thing as a free lunch. There are always trade-offs. No approach is superior in all settings.