STUDYDADDY

Get Homework Help From Expert Tutor

Get Help

Week 3 Lecture 7

We have so far seen how we can summarize data sets using descriptive statistics, showing several characteristics including mean and standard deviation. We also found that if our data comes from a random sample of a larger population, these descriptive statistics become inferential statistics, and can be used to make inferences about the population. These inferences can then be used in statistical tests to see if things have changed or not (equal to known standards or other data sets or not).

We have looked at one and two sample mean tests (with the t-test) and two sample comparisons of variance equality (with the F test). This week we will look at the Analysis of Variance (ANOVA) test for mean equality between three or more groups.

ANOVA

The first question often asked is why not just do multiple t-tests comparing three or more different group means? One answer involves efficiency. Conducting multiple t-tests can become somewhat tedious. Comparing just three groups (A, B, and C) requires us to compare A and B, B and C, and A and C (3 tests). With 4 groups (A, B, C, and D) we have A and B, A and C, A and D, B and C, B and D, C and D (6 tests)! So a single test can save us a lot of time and is much more efficient.

A second reason and much more important reason is that we lose confidence in our results when multiple tests are performed on the same data. With an alpha of 0.05, we are 95% certain we are right with each test, but being certain we are right for all the tests involves multiplying the results together, so for three tests we would be .95*.95*.95 or 86% certain; with six tests, our confidence drops to .95^6 = .74, a long way from our desired 95% confidence. So, a single test maintains our desired level of confidence in the outcome (Lind, Marchel, & Wathen, 2008).

Logic

A second question asked comes from the name itself, how can analyzing variance tell us anything about mean differences? The answer lies in how ANOVA works. The key assumptions for an ANOVA analysis are that each of the groups are normally distributed AND have equal variances. These mean that the distributions are shaped the same and, this allows for an easy comparison. Take a look at the following two sets of normal curves.
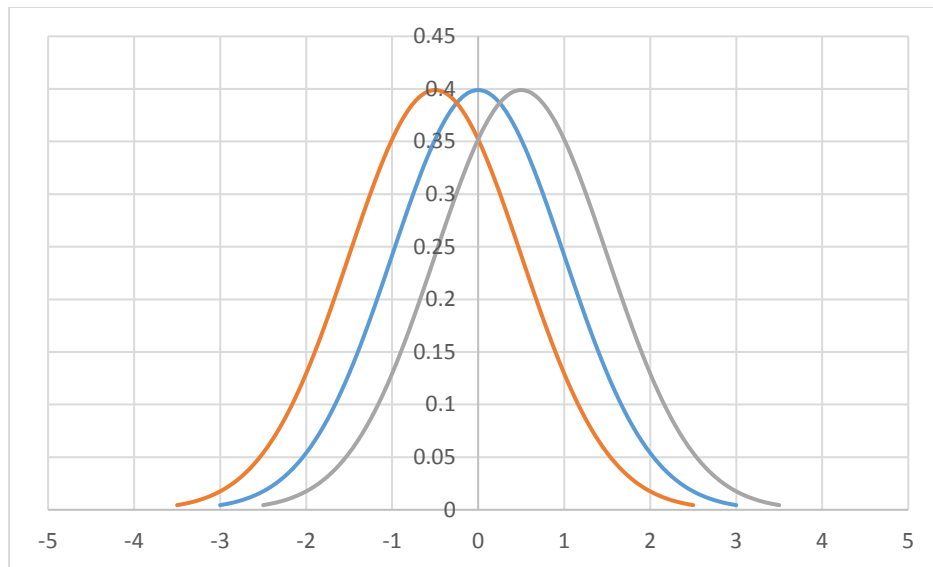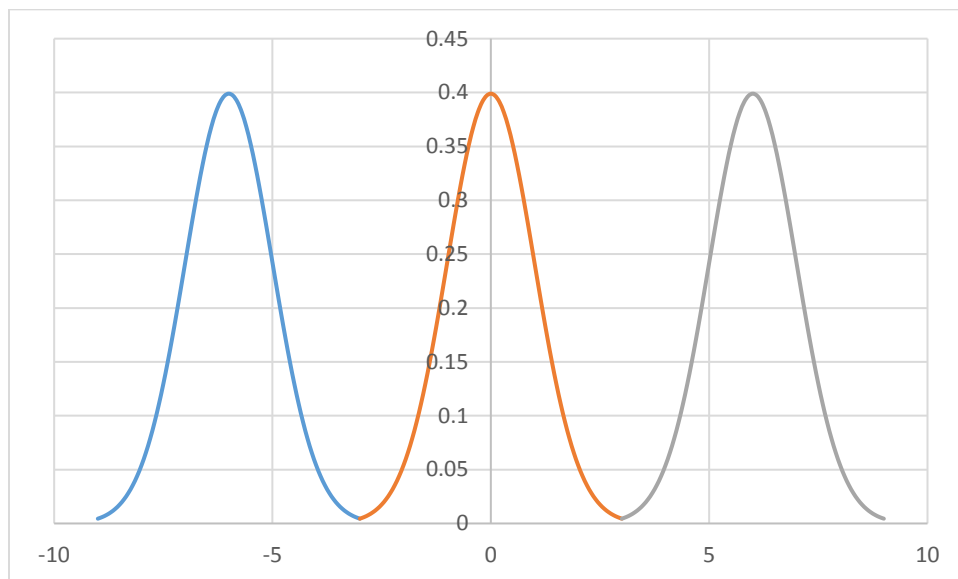
Exhibit A

Exhibit B

The means of the three sample groups in Exhibit A could clearly come from three populations that have the same mean, and the differences seen are merely sampling errors. However, we cannot say the same thing about the sample groups in Exhibit B.

ANOVA takes the variation of all of the data in the groups being tested (three in this case) and compares it with the average variation for each of the groups using the F-test (discussed last week). Since for the Exhibit A groups, the overall variation will be only slightly larger than the average of the three (which are assumed to be equal). Since the resulting F value

will not be statistically significant, we can say that the groups are closely distributed and the means are statistically equal.

In Exhibit B, however, the variation of the entire group would be around three times the variation of the average. Just by looking at the average variance for the individual groups and comparing it to the variance for the entire group, we can make a judgement on how close the distributions are, and with that a judgement on mean equality.

As with the t-test, ANOVA will let us know exactly how much difference in the population locations is enough to say means differ or not, we cannot just "eyeball" it.

Hypothesis

Stating the null and alternate hypothesis for an ANOVA test is simple, as they are always the same:

Ho: All means equal.

Ha: At least one mean differs (Tanner & Youssef-Morgan, 2013).

You might recall from last week that we said the alternate always states the opposite from the null statement. If so, why isn't our alternate: all means differ, which seems like the opposite? The reason is that the ANOVA test will reject the null hypothesis if even one mean from the groups being examined is statistically significant difference. So, the opposite of all means differ is actual at least one mean differs.

Data Set-up

Setting-up the data for an ANOVA analysis is just a bit more complicated than for a t-test. While with the T-test we just highlighted the column or portion of a column of data (sometimes after sorting it by a variable such as gender), for an ANOVA test, we need to create a table. For example, if we wanted to look at average salaries per grade (shown in the Week 3 Lecture 8 example), we would need a table looking like this.

D	**E**	**F**	**G**	**H**	**I**	**J**	**K**
A	B	C	D	E	F		
23	27	41	47	58	76		
22	34	42	57	66	77		
23	36	47	50	60	76		
24	34	40	49	69	75		
24	28	43	55	64	72		
24	28			56	77		
23	35			60			
24				65			
24				62			
24				65			
24				60			
23				66			
22							
25							
24							

Doing this is fairly simple. Copy the grade and salary columns (separately) and paste them onto a new Excel sheet (probably in Week 3 to the right of the questions). Then, highlight both columns – from labels to last value – and select *Data Sort*. Select sorting on the *grade* variable and click on OK. Both columns are now in grade order, and you can highlight and cut the salaries for each grade and paste them into a new table you create with the grade letter as the head. When finished, you will have the input table used in setting up an Excel ANOVA test.

References

Lind, D. A., Marchel, W. G., & Wathen, S. A. (2008). *Statistical Techniques in Business & Finance.* (13th Ed.) Boston: McGraw-Hill Irwin.

Tanner, D. E. & Youssef-Morgan, C. M. (2013). *Statistics for Managers.* San Diego, CA: Bridgeport Education.