

Week 6: Lecture

Violation of any of the assumptions underlying ANOVA becomes less important if the study was a true experiment, with manipulated independent variable(s), randomization, equal sample sizes and large samples. Even if F_{max} is statistically significant, meaning that the group to group variation is not homogeneous can essentially be ignored if our study had the following characteristics listed above. With the homogeneity of variance assumption, equal sample sizes have been shown to override the significant F_{max} .

The normality concerns can be addressed with the large sample size. Larger sample sizes, as stated by the Central Limit Theorem will tend to approximate the normal distribution as the sample size increases. So how do we find out if we have an adequately large sample size? The power analysis will tell us this. In the example worked in class, the sample size was shown to be large with a power equal to .99+. This would provide evidence that the sample size is large.

The other assumptions concerning the error can be addressed through the use of randomization.

In contrast to nonexperimental research, ANOVA methods can be applied to non-experimental data but the ANOVA will not necessarily be robust to the violations of the assumptions in these cases. Nonexperimental research is those done without randomization and/or manipulated independent variable. Note that had we included the attributive independent variable: Anxiety, we may not be so sure that the assumptions can be satisfied or ANOVA is robust to the violations of the assumptions. ANOVA appears to work best with controlled studies. Studies have been done where the assumptions were intentionally violated using artificially generated data. So long as the features of the study mentioned above are present, the ANOVA works fine. Some studies have even used a dependent variable that was ordinal in nature with no major problems. Some of these studies appeared in a 1947 issue of Biometrics that dealt primarily with ANOVA.

How do we know when to reject the null hypothesis? From the SPSS output, we can use the value listed under "Sig." If the value found under "Sig." is less than our chosen α level, we would reject the null hypothesis; otherwise we would not reject the null hypothesis. For the sake of completeness, we also need to see if the null hypothesis should be rejected if the "Sig." column is not given. If "Sig." were not given we would need to use a Table for the F-distribution. Such a table is given in the appendix of Kirk. To use the table we need to define a few terms. The degrees of freedom for the numerator would be our degrees of freedom for Between groups, or Treatment. The degrees of freedom for the denominator would correspond to our degrees of freedom for Within Groups. Also we will call the F-value computed in the ANOVA Summary Table "F-obtained" and the F-value found in the Table for the F-distribution "F-critical". Using these definitions we can write our decision rule as:

Reject H_0 if $F_{\text{obtained}} > F_{\text{critical}}$, otherwise do not reject H_0 .

For our example we would obtain an F-critical value of about 2.36. Our F-obtained was 17.434. Hence we would reject the null hypothesis because $17.434 > 2.36$.

With the Completely Randomized ANOVA with more than 2 levels of the independent variable, if we rejected the null hypothesis, we would need to use follow-up tests, called multiple comparisons, to determine where the differences are. With ANOVA we can only say that a difference exists and cannot specify which groups are different. We could however, state that we suspect that comparing the group with the largest and smallest means caused the difference. This is a very informal approach. Taking the difference between the highest and lowest group means do not really tell us these are statistically different. We would need to follow up with some formal methods of multiple comparisons.

The multiple comparison is Assignment 2. The order stated on the assignment is to first cover orthogonal contrasts followed by post hoc tests. We would probably benefit more to do this in reverse. On the computer output from SPSS on the one-way ANOVA: Completely Randomized Design, we requested also three multiple comparison tests: Scheffé, Tukey and Newman-Keuls. We will see next how these work and

how they are similar or different from each other. After covering these we will examine the orthogonal contrasts multiple comparisons.

The Scheffé test is by far the most conservative of all the post hoc tests. This says that it requires greater differences in order to detect a statistically significant difference. The approach that we will take in this course is to first form all possible non-redundant group mean differences and then develop the appropriate critical value for each test. Although SPSS can provide you with multiple comparison tests for the one-way CRD, independent groups ANOVA, later designs will not be so simple. So a full understanding of how multiple comparisons are accomplished is very important.

The Scheffé and Tukey tests usually require the researcher to compute only one critical value. The overall probability of Type I error has been taken care of when these types of comparisons are made. The Scheffé test can be used to test complex relations between means as well as simple pairs of means. The Tukey and Newman-Keuls is used for pairs of group means only. The Newman-Keuls, however, uses a range of critical values. The greater the separation between pairs of means, the greater the critical value.

Absolute Value of the Group Mean Differences

Using the descriptive information from an earlier analysis:

D1 = Group 1 versus Group 2:	25.6667 – 40.5833 = 14.917
D2 = Group 1 versus Group 3:	25.6667 – 38.0000 = 12.333
D3 = Group 1 versus Group 4:	25.6667 – 42.7500 = 17.083
D4 = Group 1 versus Group 5:	25.6667 – 52.5833 = 26.917
D5 = Group 1 versus Group 6:	25.6667 – 43.3333 = 17.667
D6 = Group 2 versus Group 3:	40.5833 – 38.0000 = 2.583
D7 = Group 2 versus Group 4:	40.5833 – 42.7500 = 2.167
D8 = Group 2 versus Group 5:	40.5833 – 52.5833 = 12.002
D9 = Group 2 versus Group 6:	40.5833 – 43.3333 = 2.750
D10 = Group 3 versus Group 4:	38.0000 – 42.7500 = 4.750
D11 = Group 3 versus Group 5:	38.0000 – 52.5833 = 14.583
D12 = Group 3 versus Group 6:	38.0000 – 43.3333 = 5.333
D13 = Group 4 versus Group 5:	42.7500 – 52.5833 = 9.833
D14 = Group 4 versus Group 6:	42.7500 – 43.3333 = 0.583
D15 = Group 5 versus Group 6:	52.5833 – 43.3333 = 9.25

The Scheffé Test.

If any of the differences above exceed the Scheffé test critical value we would say that the two groups are statistically different or significant. If the difference score is less than the Scheffé critical value, we would say that there is no statistical difference.

The general formula for the Scheffé critical value is

$$CV = \sqrt{(k-1)F} \sqrt{\frac{2MS_{within}}{n}}$$

where k = # groups or levels of the independent variable,
 F is a value from the F-distribution table for $k - 1$ and $N - k$ degrees of freedom
 n is the sample size per group.

Note that this test as well as the Tukey and Newman-Keuls assumes equal sample sizes per treatment group. If the researcher has unequal sample sizes that are not radically different. The n in the formula may

be substituted with the value n_h where n_h is the harmonic mean of the sample sizes. The formula for the harmonic mean is

$$n_h = \frac{k}{(1/n_1) + (1/n_2) + \dots + (1/n_k)}$$

For example if $n_1 = 6$, $n_2 = 5$ and $n_3 = 7$

$$n_h = \frac{3}{(1/6) + (1/5) + (1/7)} = \frac{3}{.5095} = 5.888$$

So we use n_h in place of n in the formula.

In our example the sample sizes are equal: $n = 12$. The value of k is 6, $F(5,66) = 2.36$ and $MS_{within} = 53.052$. Substituting these values into the formula we get:

$$CV = \sqrt{(5) 2.36} \sqrt{\frac{2(53.052)}{12}} = 3.435 \times 2.974 = 10.216$$

Comparing the differences presented above D1, D2, D3, D4, D5, D8 and D10 were significant at the $\alpha = .05$ level.

Tukey Test

For the Tukey test, the critical value is found using the following formula:

$$CV = q_r \sqrt{\frac{MS_{within}}{n}}$$

MS_{within} and n are the same as before. The q_r value is found in the Table for the Studentized Range. The r is the number of means to be compared. In this example, $r = 6$. Using $\alpha = .05$ with $df_{within} = 66$, the q_r value is 4.16. Inserting these values into the formula the CV is

$$CV = 4.16 \sqrt{\frac{53.052}{12}} = (4.16)(2.103) = 8.747.$$

This says that if any difference exceeds the value 8.747 is statistically significant. For our example, it would include: D1, D2, D3, D4, D5, D8, D10, D13 and D15. Note that the Tukey test “picked up” an additional 3 comparisons over the Scheffé test.

Newman-Keuls

For the Newman-Keuls, also known as S-N-K test, there are different critical values used among the comparisons. The S-N-K takes into consideration the distance between each pair of means in terms of ranks. The first step is to organize the group means in ascending order. Using data from our example:

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
25.667	38.000	40.5833	42.75	43.333	52.5833

When two group means are adjacent to one another, $r = 2$. If they are two ranks away, $r = 3$ and so forth. For our problem we will compute 5 critical values. The greater the rank difference, the larger the critical value.

For our example, the $r = 2$ comparisons would be Group 1 vs. Group 3, Group 3 versus Group 2, Group 2 versus Group 4, Group 4 versus Group 6 and Group 6 versus Group 5.

The $r = 3$ comparisons would be Group 1 versus Group 2, Group 3 versus Group 4, Group 2 versus Group 6 and Group 4 versus Group 5.

For $r = 4$, the comparisons are Group 1 versus Group 4, Group 3 versus Group 6 and Group 2 versus Group 5.

For $r = 5$, it would be Group 1 versus Group 6 and Group 3 versus Group 5.

Finally, for $r = 6$, the comparison is Group 1 versus Group 5.

The formula for computing the critical value is identical to the one presented for the Tukey test. The difference here is that there are five critical values instead of one.

$$CV1 \text{ (for } r = 2, q = 2.83) = 2.83 \sqrt{\frac{53.052}{12}} = 2.83 \times 2.103 = 5.95$$

$$CV2 \text{ (for } r = 3, q = 3.40) = 3.40 \times 2.103 = 7.15$$

$$CV3 \text{ (for } r = 4, q = 3.74) = 3.74 \times 2.103 = 7.865$$

$$CV4 \text{ (for } r = 5, q = 3.98) = 3.98 \times 2.103 = 8.370$$

$$CV5 \text{ (for } r = 5, q = 4.16) = 4.16 \times 2.103 = 8.748$$

The results of this test shows a significant difference between : D1, D2, D3, D4, D5, D8, D10, D13 and D15.

The Dunnett Test

The Dunnett test is different from the other three tests in that this one makes comparisons between one group (the control group) versus the other treatment groups. So if we had declared that Group 1 was the control group, the Dunnett test will compare Groups 2, 3, 4, 5, and 6 to Group 1. The formula for the critical value is

$$CV = d_r \sqrt{\frac{2 \times MS_{within}}{n}}$$

The MS_{within} and n are defined as before. The d_r is from a table created especially for the Dunnett test. Its use is similar to the studentized range table. We would enter the table with $\alpha = .05$, $r = k$ and $df_{within} = N - k$. For our example, the value of d_r will be 2.58 since $r = 6$, $df_{within} = 66$. Hence the critical value is

$$2.58 \sqrt{\frac{2 \times 53.052}{12}} = 2.58 \times 2.974 = 7.672 . \text{ If the difference score exceeds this value, we have a significant test. For our example, D1, D2, D3, D4 and D5 are all statistically significant.}$$