Chapter 9: One-Way Analysis of Variance (ANOVA)

I. Overview

a. One-way ANOVA is used when you have one categorical independent variable and one continuous (i.e., intervally scaled) dependent variable.
   i. Usually, the independent variable has at least three categories, but it can have as few as two.
   ii. When the independent variable has only two categories, the results will be the same as those obtained in an independent samples *t* test, except the *F* value from the ANOVA will be the *t* value squared.

b. The purpose of an ANOVA is to determine whether there is a statistically significant difference between the means of the groups of the independent variable on the dependent variable.
   i. e.g., to examine whether 5th graders from Iowa, Michigan, and Arizona differ in their average scores on a test of math aptitude.

II. One-way ANOVA vs. Multiple *t* tests

a. To compare 3 or more independent sample means, would could just perform a series of independent *t* tests comparing each group mean to every other group mean. But this is problematic.
   i. When we run multiple *t* tests, we increase the probability of making a Type I error
      1. Rejecting the null hypothesis and concluding that the means are significantly different from each other when they are not.
   ii. The more groups we compare to each other using independent *t* tests, the greater our chances of making a Type I error become.

b. ANOVA solves this problem by taking the number of groups into account and controlling the Type I error level to be whatever our alpha level is (e.g., .05).
   i. If you look at Appendix C, you can see that the critical *F* values depend on the number of groups being compared.

c. Therefore, it is better to perform a one-way ANOVA rather than multiple independent *t* tests when we are comparing three or more group means.

III. Partitioning the Variance in a One-Way ANOVA

a. When you have different groups of cases (i.e., different samples), each group will have its own mean, but there will also be an overall mean for all of the groups combined, called the grand mean.

b. The difference between any individual score and the grand mean is the sum of the difference between the individual score and the mean for the group that individual belongs to plus the difference between the group mean and the grand mean.

c. The difference between the individual scores within a group and the mean for that group is considered *random sampling error*, or just *error*. In contrast, the difference between the individual group means and the grand mean is potentially important because it is the amount of difference that is attributed to belonging in one group or another.

d. The statistics of interest in an ANOVA is the *F* value. This *F* value is a measure of the average difference between the group means and the grand mean divided by the average difference between the individual scores and the group mean.
    i. In other words, the *F* value is a ratio of the differences attributable to group membership divided by random sampling error.

IV. Calculating an *F* value

a. When calculating differences, or *variation*, from a mean, all of the differences must be squared before they are added together, or the sum of the deviations will equal zero.

b. To calculate the average amount of difference between individual scores and their group means, the group mean is subtracted from each score in the group, those differences are each squared, and then all of these squared deviations are added together. This produces the sum of squared deviations within groups (*SSw*), also known as the sum of squared deviations error (*SSe*).
    i. This sum is then divided by the degrees of freedom within groups ($N - K$) to produce the mean square within (*MSw*), also known as the mean square error (*MSe*). This is the denominator of the *F* value, and is considered random sampling error.
        1. *N* refers to the number cases across all samples combined.
        2. *K* refers to the number of groups being compared.

c. To calculate the average difference between the group means and the grand mean, the grand mean is subtracted from each group mean, this value is squared, and then the squared value is multiplied by the number of cases in the group. This process is repeated for each group, and then these values are all added together to produce the sum of squared deviations between groups (*SSb*).
    i. This sum is then divided by the degrees of freedom between groups ($K - 1$) to produce the mean squares between groups (*MSb*). This is the numerator of the *F* value.

d. To calculate the *F* value, the *MSb* is divided by the *MSw*. In other words, $F = \frac{MS_b}{MS_e}$

V. Interpreting the *F* value

a. Once you have calculated the *F* value you can look in Appendix C to determine whether it is statistically significant.
    i. Use the degrees of freedom between groups (the numerator; $K - 1$) and the degrees of freedom error (the denominator; $N - K$) to find the critical *F* value in Appendix C.
    ii. If the *F* value that you calculated, known as the *observed F* value, is not statistically significant (i.e., $Fo < Fc$), you conclude there are not differences between the population means that your samples represent.
    iii. If the *F* value is statistically significant (i.e., $Fo > Fc$), then you conclude that the population means do differ somehow, but you are not yet sure which specific means differ from each other.

VI. Performing the post-hoc Tukey tests

a. When you find a statistically significant *F* value, you must do some sort of post-hoc analysis to determine which of the group means are significantly different from each other.
    i. There are many types of post-hoc tests. Different post-hoc test are better than others in particular circumstances (e.g., if the sample sizes are unequal, the variances between the groups are unequal, you want a particularly conservative or liberal test of statistical significance, etc.)
        1. In this textbook we only discuss Tukey HSD tests.

b. The Tukey test is sort of like a series of independent *t* tests in which each group in the ANOVA is compared to each other group.

c. To control for the Type I error rate, which is increased when doing multiple comparisons of group means, the critical Tukey value takes into account the number of groups being compared.
    i. Notice how the critical Tukey values in Appendix D depend upon specifying the number of groups being compared.

d. The Tukey formula simply has the difference between the two group means that are being compared as the numerator and a standard error in the denominator.
    i. The standard error is calculated by dividing the *MSe* by the sample size *for a single sample* and finding the square root.

$$Tukey\ HSD = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}}}$$

where $s_{\bar{x}} = \sqrt{\frac{MS_e}{n_g}}$

and n_g = the number of cases *in each group*

e. The results of the Tukey post-hoc comparisons reveal which population means are significant different from each other, and for which groups there are not significant differences between the means.
    i. If the observed Tukey value (i.e., the one you calculated) is larger than the critical Tukey value (i.e., the one you found in Appendix D), the two group means being compared are statistically significantly different from each other.

VII. Effect Size

a. The effect size used in ANOVA is the eta squared statistic.

b. The eta squared reveals the percentage of variance in the dependent variable that is explained by the independent variable.

VIII. Summary

a. A one-way ANOVA is performed to compare the means of multiple groups to each other, usually more than two.
    i. It is a very common statistical technique in the social sciences.

b. The *F* value represents the amount of variance attributable to the groups relative to the amount of random sampling error found within the groups.

c. A significant *F* value indicates that there are meaningful differences between the group means, but it does not indicate which group means are significantly different from each other.

d. Post-hoc tests like the Tukey HSD are needed to determine which group means differ from each other significantly.