Analysis of Variance (ANOVA)

- analysis of variance - inferential stats technique used to test for significant relationships between 2 variables in **2 or more samples**. The logic is the same as t-tests, just extended to independent variables with two or more samples. You likely wouldn't do it with just 2 samples, in that case you'd use the t-test, using ANOVA with only 2 samples, you lose specified direction of the distribution & it also doesn't tell you which means is larger.

Understanding Analysis of Variance:

- One-way ANOVAs - An analysis of variance procedure using 1 dependent & 1 independent variable

- ANOVAs examine the differences between the samples as well as the differences within a single sample

- ANOVA compares differences between the samples & the differences within the sample. If the differences between the samples is greater than the number of differences within the single samples, the mean likely isn't equal

5 Steps in ANOVA Hypothesis Testing:

1. Make assumptions
2. State the Research & Null Hypothesis & Select the Alpha (Alpha Levels - .05, .01, .001)
3. Select the Sampling Distribution & Specify the Test Statistic (F-distribution & F-statistic)
4. Compute the Test Statistic
5. Make a Decision & Interpret the Results

The Structure of Hypothesis Testing with ANOVA: Assumptions:

1. Independent **random samples**are used. Independent means our choice of sample members from one population has no effect on the choice of members from subsequent populations. In other words, the members from one sample don't influence the way members of another sample are selected.
2. The dependent variable is measured at the **interval-ratio** level. Some researchers however, do apply ANOVA to ordinal level measurements. ANOVA could work with ordinal level data if there are enough categories.
3. The population is normally distributed or the variable we're looking at is normally distributed. We can never confirm this since we don't have the population data (remember, we're doing inferential statistics here, inferring about the population based on samples).
4. The population variances are equal (never really know since we don't have population data. Instead we use sample variances). If the variance of 1 group is more than double of the other group, we can't use ANOVA.

Stating Research & Null Hypothesis:

- H1 - At least 1 mean is different from the others
- H0 - All means are equal

Needed Calculations:

1. Calculate means for individual samples
2. Calculate overall mean (mean of all samples combined)

Between-Group Sum of Squares:

- tells us the differences **between** the groups

Formula:

- $$SSB = \sum N_k(\bar{Y}_k - \bar{Y})^2$$

  N_k = the number of cases in a sample (k represents the number of different samples)

  $\bar{Y}$ = the overall mean

  $\bar{Y}_k$ = the mean of a sample

Within-Group Sum of Squares:

- tells us the variations **within** our groups; it also tells us the amount of **unexplained** variance

Formula:

- $$SSW = \sum (Y_i - \bar{Y}_k)^2$$

  N_k = the number of cases in a sample (k represents the number of different samples)

  Y_i =each individual score in a sample

  $\bar{Y}_k$ = the mean of a sample

Alternative Formula for Calculating Within-Group Sum of Squares (SSW):

Formula:

- $$\sum Y_i^2 - \sum \frac{(\sum Y_k)^2}{n_k}$$

  where

  n_k = the total of each sample

  Y_i^2 =the squared scores from each sample

  $\sum Y_k$ = the sum of the scores of each sample

- Square all the individual scores in each sample & add up all the squared values

- Add up all the individual scores in each sample & square up each sum then divide by the number of scores in each sample (right-side of equation)
- Add up all the divided values
- Subtract!

Total Sum of Squares:

- SST = SSB + SSW

Mean Square Between:

- An estimate of the between-group variance obtained by dividing the **between-group variance** of squares by its **degrees of freedom**
- dfb = degrees of freedom between, aka **df1**
- dfb = k - 1 (where k is the number of samples)
- Mean Square Between (MSB) = SSB/dfb

Mean Square Within:

- An estimate of the **within-group variance** obtained by **dividing** the within-group sum of squares by its degrees of freedom
- dfw = degrees of freedom within, aka **df2**
- dfw = N - K (N = total number of cases, K = number of samples)
- Mean square within (MSW) = SSW/dfw

The F-Statistic:

- F-stat is the ratio of between-group variance to within-group variance (between-group variance : within-group variance)
- F = Mean Square Between/Mean Square Within
- Take F-obtained & compare to F-critical, if F-obtained is below critical value, we fail to reject the null hypothesis. If F-obtained is at or below the f-critical, we reject the null hypothesis.
- F ratio (aka F Statistic) - Used in an analysis of variance, the F statistic represents the ratio of between-group variance to within-group variance
- F obtained - The test statistic computed by the ratio for between-group to within-group variance
- F critical - The F score associated w/a particular **alpha level** & **degree of freedom**. This F score marks the beginning of the **region of rejection** for our null hypothesis