One-Way ANOVA

I have already described the basic concepts and approach of analysis of variance (ANOVA). One of the main barriers to reading about and understanding ANOVA is getting through the notation, so I start with that here. Below, I present the definitional formulas for ANOVA. Many textbooks present the computational formulas which are simpler to use for larger problems. Definitional formulas have a very clear tie to the concepts behind the analysis, however. Nowhere is it more clear than in the ANOVA formulas, which quantify between and within-group variation. To simplify matters, I also use an equal-n version of the formulas, but ANOVA also can be used with unequal group sizes.

Notation

It is easy to get lost or bogged down in the notation used in the ANOVA formulas. There are lots of subscripts which can be confusing. The notation used is a classic notation, but it is difficult for many students to penetrate. With ANOVA, we now have to keep track of multiple groups, so a subscript, *j*, is used to denote a specific group. A single score is now represented by Y_{ij}, indicating the score is for an individual, i, within a particular group, j. There are now different means to refer to also. The mean of the full sample is now referred to as $\bar{Y}_{..}$, because it is calculated across all individuals and all groups. So, the "." refers to "computing across" that element—either individuals or groups. Then, $\bar{Y}_{.j}$ represents the mean of a particular group (e.g., $\bar{Y}_{.2}$ would be for the mean of the second group). The "." is used in place of the i because the mean is calculated using all the i's for a particular group.

Sum of Squares Components

There are three possible sums of squares—between-group some of squares (SS_A)[1], within-group or error sum of squares ($SS_{s/A}$), and total sum of squares (SS_T). Total sum of squares can be partitioned into between sum of squares and within sum of squares, representing the variation due to treatment (or the independent variable) and variation due to individual differences in the score respectively:

$$SS_T = SS_A + SS_{s/A}$$

Sum of squares between-groups examines the differences among the group means by calculating the variation of each mean ($\bar{Y}_{.j}$) around the grand mean ($\bar{Y}_{..}$): $SS_A = n\sum\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)^2$. n is the number of observations in each group (i.e., each cell or level of factor A).

Sum of squares within-groups examines error variation or variation of individual scores around each group mean. This is variation in the scores that is not due to the treatment (or independent variable):

$$SS_{s/A} = \sum\sum\left(Y_{ij} - \bar{Y}_{.j}\right)^2$$

The total sum of squares can be computed by adding the SS_A and the $SS_{s/A}$, but they can also be computed the same way we would for computing the numerator in the formula for sample variance—by simply subtracting each score from the grand mean, squaring, and then summing across all cases.

Degrees of Freedom

Each SS has different degrees of freedom associated with it: $df_A = a - 1$, $df_{s/A} = a(n-1) = N - a$, and $df_T = an - 1 = N - 1$. Here, a equals the number of groups (or "levels" of the independent variable), n is the number of observations in each group (assuming they are equal), and N is the total number of observations in the study (which is equal to *a* multiplied by n for equal group size).

Mean Squares and F

The mean squares are computed by dividing the SS by the df. This is akin to the computation of the sample variance that divides the sum of squares by degrees of freedom. In fact, $MS_T = s^2$. The F ratio is then computed by creating a ratio of the between-groups variance to the within-groups variance:

[1] SS_A is referred to as the *method sums of squares* in the example from Myers, Well, & Lorch (2010), because different levels of the memorization method were compared.

$$F = \frac{MS_A}{MS_{s/A}}$$

Magnitude of Effect

Significant differences among the groups indicate that it is unlikely that the differences among the means is due to random sampling chance, but that leaves unanswered the question as to how large the differences are. The most commonly used gauge of the magnitude of effect is η^2 (eta-squared), which can be defined as the proportion of variance accounted for in the dependent variable by the independent variable. It is simply the proportion of between-group variation, as measured by the sum of squares between groups (SS_A) relative to the total variation (SS_T).

$$\eta^2 = \frac{SS_A}{SS_T}$$

You may hear about two other magnitude of effect measures, ω^2 (omega-squared) and f. ω^2 is essentially a correction for bias in η^2 in estimating the population value (more on this later).

$$\omega^2 = \frac{(a-1)(F-1)}{(a-1)(F-1)+an}$$

f is Cohen's (1988) effect size measure and is equal to:

$$\hat{f} = \sqrt{\frac{(a-1)\left(MS_A - MS_{s/A}\right)}{an\left(MS_{s/A}\right)}}$$

f (the caret symbol above signifies a sample estimate) is used in power analysis and can be interpreted in terms of Cohen's suggested standards for small (.1), medium (.25), and large (.4) as a general guide.

Assumptions

The assumptions for ANOVA follow those we discussed for the t-test: normal distribution of the dependent variable in the population, independence of observations, and equal variance (homogeneity) among groups in the population. As was seen with the t-test, the normality and equal variance assumptions are of greater concern with smaller sample sizes in ANOVA. And as we know from the central limit theorem and what we saw earlier with the simulation demonstration in class, the population distribution can be very nonnormal and the sampling distribution for the mean is still quite normal. The primary exception is when the population distribution is pretty extremely skewed or kurtotic and the same size is quite small. Heterogeneity of variances across groups is perhaps of more concern, but as with the t-test, the worst performance is when the group sample sizes are small, unequal in size, and the variances are very unequal (e.g., 4:1 ratio, Myers, Well, & Lorch, 2013, p. 137). In these circumstances, researchers should explore one of the corrective tests, of which, the James and the Welch tests appear to perform fairly well (Algina, Oshima, & Lin, 1994; Lix, Keselman, & Keselman, 1996). Remember that tests of unequal variances to investigate heterogeneity may fail to find significance in these more critical circumstances and may identify relatively trivial violations as significant in large samples, so they may be of limited utility, at least in deciding conclusively that there is a problem with heterogeneity.

References

Algina, J., Oshima, T. C., & Lin, W. Y. (1994). Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. *Journal of Educational Statistics, 19*, 275-291.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd Edition*. New York: Routledge.
Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of educational research*, *66*, 579-619.
Myers, J.L., & Well, A.D., & Lorch, R.F.,Jr. (2010). *Research design and statistical analysis (3rd Edition)*. Mahwah, NJ: Erlbaum.

Example of a Three-group ANOVA

The following builds on the earlier hypothetical example by comparing the means of three learning strategies, adding a third group, concept mapping, to our earlier reading only and retrieval practice groups. Concept mapping involves drawing a diagram with nodes to illustrate links between concepts.[2]

Reading Only	$\left(Y_{ij}-\bar{Y}_{.1}\right)^2$	$\left(Y_{ij}-\bar{Y}_{..}\right)^2$	**Retrieval Practice**	$\left(Y_{ij}-\bar{Y}_{.2}\right)^2$	$\left(Y_{ij}-\bar{Y}_{..}\right)^2$	**Concept Mapping**	$\left(Y_{ij}-\bar{Y}_{.3}\right)^2$	$\left(Y_{ij}-\bar{Y}_{..}\right)^2$
4	4	9	9	0	4	6	0	1
4	4	9	8	1	1	6	0	1
6	0	1	10	1	9	6	0	1
8	4	1	8	1	1	7	1	0
8	4	1	10	1	9	5	1	4
$\bar{Y}_{.1}=6$	$\sum\left(Y_{ij}-\bar{Y}_{.1}\right)^2=16$		$\bar{Y}_{.2}=9$	$\sum\left(Y_{ij}-\bar{Y}_{.2}\right)^2=4$		$\bar{Y}_{.3}=6$	$\sum\left(Y_{ij}-\bar{Y}_{.3}\right)^2=2$	

$\bar{Y}_{..}=7$

ANOVA Table

$SS_A=n\sum\left(\bar{Y}_{.j}-\bar{Y}_{..}\right)^2=5\left[(6-7)^2+(9-7)^2+(6-7)^2\right]=30$	$df=a-1=3-1=2$	$MS_A=\frac{SS_A}{df_A}=\frac{30}{2}=15$	$F=\frac{MS_A}{MS_{s/A}}=\frac{15}{1.83}=8.182$
$SS_{s/A}=\sum\sum\left(Y_{ij}-\bar{Y}_{.j}\right)^2=16+4+2=22$	$df=N-a=15-3=12$	$MS_{s/A}=\frac{SS_{s/A}}{df_{s/A}}=\frac{22}{12}=1.833$	
$SS_T=\sum\left(Y_{ij}-\bar{Y}_{..}\right)^2=52$			

F_{crit} with df of 2 and 12 is 3.88, so the calculated F is significant.

One can then estimate the magnitude of the effect of the independent variable by computing η^2 or ω^2:

$$\eta^2=\frac{SS_A}{SS_T}=\frac{30}{52}=.58 \text{ or } 58\%$$

$$\omega^2=\frac{SS_A-(a-1)\left(MS_{s/A}\right)}{SS_T+MS_{s/A}}=\frac{30-(3-1)1.83}{52+1.83}=.49 \text{ or } 49\%$$

Write-up. See the subsequent handout illustrating computer software analysis for the write-up example.

[2] Although these are artificial data, they are based on actual research studies and the results do mirror findings in some learning strategy studies, such as results reported in Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772-775.