

One-way ANOVA

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 this 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 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 ($\bar{Y}_{..}$).

$$SS_A = n \sum (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

n is the number of observations in each group (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).

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

The total sum of squares can be computed by adding the SS_A and the $SS_{s/A}$, but 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 a 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 which 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:

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

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

Example of a Three-group ANOVA

The following is a hypothetical example comparing satisfaction ratings of teachers randomly assigned to public, charter, and private schools (e.g., Catholic schools).

Public	$(Y_{ij} - \bar{Y}_{.1})^2$	$(Y_{ij} - \bar{Y}_{..})^2$	Charter	$(Y_{ij} - \bar{Y}_{.2})^2$	$(Y_{ij} - \bar{Y}_{..})^2$	Private	$(Y_{ij} - \bar{Y}_{.3})^2$	$(Y_{ij} - \bar{Y}_{..})^2$
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
$\bar{Y}_{.1} = 6$	$\sum (Y_{ij} - \bar{Y}_{.1})^2 = 16$		$\bar{Y}_{.2} = 9$	$\sum (Y_{ij} - \bar{Y}_{.2})^2 = 4$		$\bar{Y}_{.3} = 6$	$\sum (Y_{ij} - \bar{Y}_{.3})^2 = 2$	

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

ANOVA Table

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

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

An Analysis of Variance tested whether there were differences in satisfaction among public, charter, and private school teachers. Charter school teachers had higher satisfaction ratings ($M = 9$) compared to public and private school teachers (both $M = 6$). Results indicated that these means differed significantly, $F(2,12) = 8.18, p < .01$.

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

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

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