

What is the difference between within-group and between-group variation in ANOVA?

Authored by
stats writer

July 1, 2024

RECOMMENDED CITATION

stats writer (2024). *What is the difference between within-group and between-group variation in ANOVA?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=165320>

ANOVA (Analysis of Variance) is a statistical tool used to determine if there are significant differences between the means of two or more groups. Within-group variation refers to the differences in scores within each group, while between-group variation refers to the differences between the group means. In simpler terms, within-group variation measures the variability within a single group, while between-group variation measures the variability between multiple groups. ANOVA takes into account both within-group and between-group variation to determine if there are significant differences between the groups being compared. Understanding the difference between within-group and between-group variation is important in interpreting the results of ANOVA and can help identify the sources of variability in a dataset.

Within-Group vs. Between Group Variation in ANOVA

A is used to determine whether or not the means of three or more independent groups are equal.

A one-way ANOVA uses the following null and alternative :

H₀: All group means are equal. H_A: At least one group mean is different from the rest.

Whenever you perform a one-way ANOVA, you will end up with a summary table that looks like the following:

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Between Groups	207.2	2	103.6	7.6952	0.0023	3.3541
Within Groups	363.5	27	13.4630			
Total	570.7	29				

We can see that there are two different sources of variation that an ANOVA measures:

Between Group Variation: The total variation between each group mean and the overall mean.

Within-Group Variation: The total variation in the individual values in each group and their group mean.

If the Between group variation is high relative to the Within-group variation, then the F-statistic of the ANOVA will be higher and the corresponding p-value will be lower, which makes it more likely that we'll reject the null hypothesis that the group means are equal.

The following example shows how to calculate the Between group variation and Within-group variation for a one-way ANOVA in practice.

Example: Calculating Within-Group and Between Group Variation in ANOVA

Suppose we want to determine if three different studying methods lead to different mean exam scores. To test this, we recruit 30 students and 10 each to use a different studying method.

The exam scores for the students in each group are shown below:

Method 1	Method 2	Method 3
75	78	82
77	78	82
78	79	84
78	81	86
79	81	86
81	82	87
81	83	87
83	85	89
86	86	90
87	88	94

We can use the following formula to calculate the between group variation:

$$\text{Between Group Variation} = \sum nj(X_j - X_{..})^2$$

where:

n_j : the sample size of group j
 \sum : a symbol that means "sum"
 X_j : the mean of group j
 $X_{..}$: the overall mean

To calculate this value, we'll first calculate each group mean and the overall mean:

	Method 1	Method 2	Method 3
	75	78	82
	77	78	82
	78	79	84
	78	81	86
	79	81	86
	81	82	87
	81	83	87
	83	85	89
	86	86	90
	87	88	94
Group Mean	80.5	82.1	86.7
Overall Mean	83.1		

Then we calculate the between group variation to be:
 $10(80.5-83.1)^2 + 10(82.1-83.1)^2 + 10(86.7-83.1)^2 = 207.2$.

Next, we can use the following formula to calculate the within group variation:

Within Group Variation: $\sum(X_{ij} - X_j)^2$

where:

Σ : a symbol that means "sum"
 X_{ij} : the i th observation in group j
 X_j : the mean of group j

In our example, we calculate within group variation to be:

$$\text{Group 1: } (75-80.5)^2 + (77-80.5)^2 + (78-80.5)^2 + (78-80.5)^2 + (79-80.5)^2 + (81-80.5)^2 + (81-80.5)^2 + (83-80.5)^2 + (86-80.5)^2 + (87-80.5)^2 = 136.5$$

$$\text{Group 2: } (78-82.1)^2 + (78-82.1)^2 + (79-82.1)^2 + (81-82.1)^2 + (81-82.1)^2 + (82-82.1)^2 + (83-82.1)^2 + (85-82.1)^2 + (86-82.1)^2 + (88-82.1)^2 = 104.9$$

$$\text{Group 3: } (82-86.7)^2 + (82-86.7)^2 + (84-86.7)^2 + (86-86.7)^2 + (86-86.7)^2 + (87-86.7)^2 + (87-86.7)^2 + (89-86.7)^2 + (90-86.7)^2 + (94-86.7)^2 = 122.1$$

$$\text{Within Group Variation: } 136.5 + 104.9 + 122.1 = 363.5$$

If we use statistical software to perform a one-way ANOVA using this dataset, we'll end up with the following ANOVA table:

ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	207.2	2	103.6	7.6952	0.0023	3.3541
Within Groups	363.5	27	13.4630			
Total	570.7	29				

Notice that the between group and within-group variation values match the ones we calculated by hand.

The overall F-statistic in the table is a way to quantify the ratio of the between group variation compared to the within group variation.

The larger the F-statistic, the greater the variation between group means relative to the variation within the groups.

Thus, the larger the F-statistic, the greater the evidence that there is a difference between the group means.

We can see in this example that the p-value that corresponds to an F-statistic of 7.6952 is .0023.

Since this value is less than $\alpha = .05$, we reject the null hypothesis of the ANOVA and conclude that the three studying techniques do not lead to the same exam score.

Additional Resources

The following tutorials provide additional information about ANOVA models: