What is the formula for calculating the sum of squares in ANOVA, and can you provide an example?

Authored by

stats writer

June 28, 2024

ANOVA, or Analysis of Variance, is a statistical method used to compare the means of three or more groups. The sum of squares is a key component in ANOVA, as it measures the variability within and between groups. The formula for calculating the sum of squares in ANOVA is SS = Σ (X - X?)^2, where X represents the individual data points and X? represents the overall mean. This formula involves squaring the difference between each data point and the mean, and then summing up all the squared values. This process is repeated for each group, and the total sum of squares is then calculated by adding up the individual sums of squares. An example of calculating the sum of squares in ANOVA would be finding the sum of squares for three groups with the following data points: Group A (2, 4, 6), Group B (3, 5, 7), and Group C (1, 3, 5). The overall mean for these data points would be 4. The sum of squares for Group A would be: (2-4)^2 + (4-4)^2 + (6-4)^2 = 8. The sum of squares for Group B would be: (3-4)^2 + (5-4)^2 + (7-4)^2 = 6. The sum of squares for Group C would be: (1-4)^2 + (3-4)^2 + (5-4)^2 = 6. Therefore, the total sum of squares would be 8 + 6 + 6 = 20. This process can then be used to calculate the other components of ANOVA, such as the degrees of freedom and F-statistic, to determine if there is a significant difference between the means of the three groups.

Calculate Sum of Squares in ANOVA (With Example)

In statistics, a is used to compare the means of three or more independent groups to determine if there is a statistically significant difference between the corresponding population means.

Whenever you perform a one-way ANOVA, you will always compute three sum of squares values:

1. Sum of Squares Regression (SSR)

This is the sum of the squared differences between each group mean and the .

2. Sum of Squares Error (SSE)

This is the sum of the squared differences between each individual observation and the group mean of that observation.

3. Sum of Squares Total (SST)

This is the sum of the squared differences between each individual observation and the grand mean.

Each of these three values are placed in the final ANOVA table, which we use to determine whether or not there is a statistically significant difference between the group means.

The following example shows how to calculate each of these sum of squares values for a one-way ANOVA in practice.

Example: How to Calculate Sum of Squares in ANOVA

Suppose we want to know whether or not three different exam prep programs lead to different mean scores on a certain exam. To test this, we recruit 30 students to participate in a study and split them into three groups.

The students in each group are randomly assigned to use one of the three exam prep programs for the next three weeks to prepare for an exam. At the end of the three weeks, all of the students take the same exam.

The exam scores for each group are shown below:

Group 1	Group 2	Group 3
85	91	79
86	92	78
88	93	88
75	85	94
78	87	92
94	84	85
98	82	83
79	88	85
71	95	82
80	96	81

The following steps show how to calculate the sum of squares values for this one-way ANOVA.

Step 1: Calculate the group means and the grand mean.

	Group 1	Group 2	Group 3
	85	91	79
	86	92	78
	88	93	88
	75	85	94
	78	87	92
	94	84	85
	98	82	83
	79	88	85
	71	95	82
	80	96	81
Group Means	**83.4**	**89.3**	**84.7**
Overall Mean	**85.8**		

Step 2: Calculate SSR.

Next, we will calculate the sum of squares regression (SSR) using the following formula:

nΣ(Xj - X..)2

where:

n: the sample size of group jΣ: a greek symbol that means "sum"Xj: the mean of group jX..: the overall mean

In our example, we calculate that SSR = 10(83.4-85.8)2 + 10(89.3-85.8)2 + 10(84.7-85.8)2 = 192.2

Step 3: Calculate SSE.

Next, we will calculate the sum of squares error (SSE) using the following formula:

Σ(Xij - Xj)2

where:

Σ: a greek symbol that means "sum"Xij: the ith observation in group jXj: the mean of group j

In our example, we calculate SSE as follows:

Group 1: (85-83.4)2 + (86-83.4)2 + (88-83.4)2 + (75-83.4)2 + (78-83.4)2 + (94-83.4)2 + (98-83.4)2 + (79-83.4)2 + (71-83.4)2 + (80-83.4)2 = 640.4

Group 2: (91-89.3)2 + (92-89.3)2 + (93-89.3)2 + (85-89.3)2 + (87-89.3)2 + (84-89.3)2 + (82-89.3)2 + (88-89.3)2 + (95-89.3)2 + (96-89.3)2 = 208.1

Group 3: (79-84.7)2 + (78-84.7)2 + (88-84.7)2 + (94-84.7)2 + (92-84.7)2 + (85-84.7)2 + (83-84.7)2 + (85-84.7)2 + (82-84.7)2 + (81-84.7)2 = 252.1

SSE: 640.4 + 208.1 + 252.1 = 1100.6

Step 4: Calculate SST.

Next, we will calculate the sum of squares total (SST) using the following formula:

SST = SSR + SSE

In our example, SST = 192.2 + 1100.6 = 1292.8

Once we have calculated the values for SSR, SSE, and SST, each of these values will eventually be placed in the ANOVA table:

Source	Sum of Squares (SS)	df	Mean Squares (MS)	F-value	p-value
Regression	192.2	2	96.1	2.358	0.1138
Error	1100.6	27	40.8		
Total	1292.8	29			

Here is how we calculated the various numbers in the table:

df regression: k-1 = 3-1 = 2df error: n-k = 30-3 = 27df total: n-1 = 30-1 = 29MS treatment: SST / df treatment = 192.2 / 2 = 96.1MS error: SSE / df error = 1100.6 / 27 = 40.8F-value: MS treatment / MS error = 96.1 / 40.8 = 2.358p-value: p-value that corresponds to F value.

Note: n = total observations, k = number of groups

Check out for how to interpret the F-Value and p-value in the ANOVA table.