What is the process for finding the variance of grouped data, and can you provide an example?

Authored by

stats writer

June 30, 2024

The process for finding the variance of grouped data involves first determining the mean of the data set. Then, for each group, the difference between the group's midpoint and the mean is calculated and squared. These squared differences are then multiplied by the frequency of that group and added together. This sum is divided by the total frequency of all the groups minus 1, to get the variance.

For example, let's say we have a data set of the number of hours spent studying for a test by a group of students, grouped into 5 categories: 0-2 hours, 2-4 hours, 4-6 hours, 6-8 hours, and 8-10 hours. The data set also includes the frequency of students in each group.

First, we calculate the mean by adding up all the values and dividing by the total frequency. Let's say the mean is 5 hours.

Next, we calculate the squared differences for each group and multiply by the frequency. For the 0-2 hours group, the squared difference would be (1-5)^2 = 16, multiplied by the frequency of 5, giving us 80. We do this for each group and add up the values.

Finally, we divide this sum by the total frequency minus 1 (25-1=24), giving us a variance of 3.33 hours squared. This tells us that the data is spread out and there is a significant difference between the students' study hours.

Find the Variance of Grouped Data (With Example)

Often we may want to calculate the of a grouped frequency distribution.

For example, suppose we have the following grouped frequency distribution:

Range	Frequency
1-10	2
11-20	7
21-30	10
31-40	3
41-50	1

While it's not possible to calculate the exact variance since we don't know the , it is possible to estimate the variance using the following formula:

Variance: Σni(mi-μ)2 / (N-1)

where:

ni: The frequency of the ith groupmi: The midpoint of the ith groupμ: The meanN: The total sample size

Note: The for each group can be found by taking the average of the lower and upper value in the range. For example, the midpoint for the first group is calculated as: (1+10) / 2 = 5.5.

The following example shows how to use this formula in practice.

Example: Calculate the Variance of Grouped Data

Suppose we have the following grouped data:

Range	Frequency
1-10	2
11-20	7
21-30	10
31-40	3
41-50	1

Here's how we would use the formula mentioned earlier to calculate the variance of this grouped data:

Range	Frequency (n_i)	Midpoint (m_i)	$m_i * n_i$	μ	$m_i - \mu$	$(m_i - \mu)^2$	$n_i(m_i - \mu)^2$
1-10	2	5.5	11	22.89	-17.39	302.41	604.82
11-20	7	15.5	108.5	22.89	-7.39	54.61	382.28
21-30	10	25.5	255	22.89	2.61	6.81	68.12
31-40	3	35.5	106.5	22.89	12.61	159.01	477.04
41-50	1	45.5	45.5	22.89	22.61	511.21	511.21

We would then calculate the variance as:

Variance: Σni(mi-μ)2 / (N-1)Variance: (604.82 + 382.28 + 68.12 + 477.04 + 511.21) / (23-1)Variance: 92.885

The variance of the dataset turns out to be 92.885.

Additional Resources

The following tutorials explain how to calculate other metrics for grouped data: