1. **Details of Module and its structure**

Module Detail	
Subject Name	Mathematics
Course Name	Mathematics 02 (Class XI, Semester - 2)
Module Name/Title	Statistics: Variance and Standard Deviation- Part 2
Module I	kemh_21502
Pre-requisites	Measures of Central Tendency
Objectives	After going through this module, the learner will be able to: 1. Calculate variance of the data 2. Calculate standard deviation of the data
Keywords	Mean, Variance, Standard Deviation, Frequency, Data

2. **Development Team**

Role	Name	Affiliation
National MOOC Coordinator (NMC)	Prof. Amarendra P. Behera	CIET, NCERT, New Delhi
Program Coordinator	Dr. Rejaul Karim Barbhuiya	CIET, NCERT, New Delhi
Course Coordinator (CC) / PI	Prof. Til Prasad Sarma	DESM, NCERT, New Delhi
Course Co-Coordinator / Co-PI	Ms. Anjali Khurana	CIET, NCERT, New Delhi
Subject Matter Expert (SME)	Ms. Neenu Gupta	PGT Mathematics Ahlcon International School, New Delhi
Review Team	Prof. V.P. Singh (Retd.)	DESM, NCERT, New Delhi

Table of Contents:

1. Introduction

Statistics has many tools and techniques to analyze data. If one of the tool has any limitations, then another can be used to suit the needs of the purpose. You have learnt about some measures of dispersion like range, mean deviation about mean and median. Let us learn about some more measures of dispersion in this module and before that let us learn why do we need them.

2. Limitations of Mean Deviations

There can be some limitations of mean deviations. Let's discuss them.

- If a data is highly variable then median is not a reliable measure of central tendency and so is mean deviation about median.
- The sum of absolute values of deviations from the mean is always more than the sum of the deviations from the median. So mean deviation is not so reliable.
- Since computation of mean deviation involves absolute values of deviations. So, it cannot be used for further calculations and hence for analysis.

Due to all these limitations, we need some more measures of dispersion called variance and standard deviation. Standard deviation has same units as the observations. so, as a measure of dispersion, standard deviation is more reliable than mean deviation.
Let us discuss them in this module.

3. Variance and Standard Deviation

Mean deviation about mean or median uses absolute values of deviations to avoid signs of deviations. One more way to avoid signs of deviations is to square them. Let us define variance.

Variance is the arithmetic mean of the squares of the deviations from the mean. It is denoted by σ^2 or Var(X).

Being mean of squares of deviation, unit of variance is different from that of observations and their mean. Therefore, positive square-root of variance is considered to be more accurate measure of dispersion, which is termed as standard deviation.

Standard deviation is the positive square-root of the variance and is denoted by σ or S.D., i.e., S.D. = $\sigma = \sqrt{\text{variance}}$

4. Variance and Standard Deviation of Ungrouped Data

Consider an ungrouped data $x_1, x_2, x_3, \dots, x_n$. Then

Variance = $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$

And Standard Deviation $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$

Algorithm to calculate variance and standard deviation

Step 1: Calculate mean of the observations.
Step 2: Calculate the deviations of observations from the mean, i.e., $x_i - \bar{x}$
Step 3: Compute squares of all the deviations, i.e., $(x_i - \bar{x})^2$
Step 4: Compute variance by using the formula $\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$, where n is the total number of observations.
Step 5: Compute standard deviation by using the formula $\sqrt{\text{Variance}}$

Example 1: Compute the variance and standard deviation of the following data:
3, 5, 7, 9, 11

Solution: Consider data in tabular form:

x_i	$x_i - \bar{x} = x_i - 7$	$(x_i - \bar{x})^2$
3	– 4	16
5	– 2	4
7	0	0
9	2	4
11	4	16
Total		40

$$\bar{x} = \frac{3+5+7+9+11}{5} = \frac{35}{5} = 7$$

Variance = $\sigma^2 = \frac{\sum_{i=1}^{5}(x_i-\bar{x})^2}{5} = \frac{40}{5} = 8$

Standard Deviation $= \sigma = \sqrt{\text{Variance}} = \sqrt{8} = 2\sqrt{2}$

Example 2: Compute the variance and standard deviation of the following data:
5, 10, 15, 20, 25, 30, 35

Solution: Consider data in tabular form:

x_i	$x_i - \bar{x} = x_i - 20$	$(x_i - \bar{x})^2$
5	– 15	225
10	– 10	100
15	– 5	25
20	0	0
25	5	25
30	10	100
35	15	225
Total		700

$$\bar{x} = \frac{5+10+15+20+25+30+35}{7} = \frac{140}{7} = 20$$

Variance = $\sigma^2 = \frac{\sum_{i=1}^{5}(x_i-\bar{x})^2}{5} = \frac{700}{7} = 100$

Standard Deviation $= \sigma = \sqrt{\text{Variance}} = \sqrt{100} = 10$

5. Variance and Standard Deviation of Discrete Frequency Distribution

A discrete frequency distribution is a data which gives us the frequency for a specific observation; which means we can easily analyze that which particular observation is being repeated for how many number of times.
Let us learn how to calculate mean deviation about mean for discrete frequency distribution.
Consider the following discrete frequency distribution:

x:	x_1	x_2	x_3	…….	x_n
f:	f_1	f_2	f_3	…….	f_n

Since each observation has some frequency, deviations are also to be multiplied by the frequency to calculate the variance.
In this case, we use the following formulae to calculate variance and standard deviation.

Variance = $\sigma^2 = \frac{\sum_{i=1}^{n} f_i(x_i-\bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\left[\sum_{i=1}^{n} f_i(x_i-\bar{x})^2\right]$, where N = $\sum_{i=1}^{n} f_i$

and Standard Deviation $=\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N}}$

Another Formula for Standard Deviation:

As stated above,

Variance = $\sigma^2 = \frac{1}{N}\left[\sum_{i=1}^{n} f_i (x_i - \bar{x})^2\right]$

$= \frac{1}{N}\sum_{i=1}^{n} f_i ({x_i}^2 - 2x_i\bar{x} + \bar{x}^2)$

$= \frac{1}{N}\left[\sum_{i=1}^{n} f_i {x_i}^2 - \sum_{i=1}^{n} 2\bar{x} f_i x_i + \sum_{i=1}^{n} f_i \bar{x}^2\right]$

$= \frac{1}{N}\left[\sum_{i=1}^{n} f_i {x_i}^2 - 2\bar{x}\sum_{i=1}^{n} f_i x_i + \bar{x}^2\sum_{i=1}^{n} f_i\right]$

$= \frac{1}{N}\left[\sum_{i=1}^{n} f_i {x_i}^2 - 2\bar{x}\,.N\bar{x} + \bar{x}^2 N\right]\left[N\,\bar{x} = \sum_{i=1}^{n} f_i x_i\right]$

$= \frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - 2\bar{x}^2 + \bar{x}^2$

$= \frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - \bar{x}^2$

$= \frac{1}{N}\sum_{i=1}^{n} f_i {x_i}^2 - \left(\frac{\sum_{i=1}^{n} f_i x_i}{N}\right)^2$

$$\sigma^2 = \frac{1}{N^2}\left[N\sum_{i=1}^{n} f_i {x_i}^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2\right]$$

and $\sigma = \frac{1}{N}\sqrt{N\sum_{i=1}^{n} f_i {x_i}^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$

Let us take some examples to learn how to apply these formulae.

Example 3: Compute variance and standard deviation for the following data:

x_i	5	10	15	20	25	30
f_i	2	3	2	3	2	3

Solution: Consider the data in the following tabular format:

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
5	2	10	– 13	169	338
10	3	30	– 8	64	192
15	2	30	– 3	9	18
20	3	60	2	4	12
25	2	50	7	49	98
30	3	90	17	289	867
Total	15	270			1525

Here, mean = $\bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{\sum_{i=1}^{6} f_i} = \frac{270}{15} = 18$

Variance = $\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{1525}{15} = 715.8$

Standard Deviation = $\sigma = \sqrt{\text{Variance}} = \sqrt{715.8} = 26.75$

Example 4: Compute standard deviation for the following data:

x_i	2	4	6	8	10
f_i	6	4	4	6	5

Solution: Consider the given data in the following tabular form:

x_i	f_i	$f_i . x_i$	x_i^2	$f_i x_i^2$
2	6	12	4	24
4	4	16	16	64
6	4	24	36	144
8	6	48	64	384
10	5	50	100	500
Total	N = 25	150		1124

$$\sigma = \frac{1}{N}\sqrt{N\sum_{i=1}^{n} f_i {x_i}^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$$

$$= \frac{1}{25}\sqrt{25 \times 1124 - (150)^2}$$

$$= \frac{1}{25}\sqrt{28100 - 22500}$$

$$= \frac{1}{25}\sqrt{5600}$$

$$= \frac{1}{25} \times 2 \times 10 \times \sqrt{14}$$

$$= \frac{4}{5}\sqrt{14} = 2.992 \text{ (approx.)}$$

6. Variance and Standard Deviation of Continuous Frequency Distribution

A continuous data is a series in which data is being classified in the form of class – intervals without any gap along with their respective frequencies.

Sometimes due to large values of x_i, calculation of mean becomes tedious so, step-deviation method is used.

Let us derive the formula for variance and standard deviation by step deviation method.

Let the assumed mean be 'A' and class interval be of size 'h'.
Let the step-deviations be y_i ,
i.e., $y_i = \frac{x_i - A}{h} \implies x_i = A + hy_i$ ……. (1)

We know, $\bar{x} = \frac{1}{N}\sum_{i=1}^{n} f_i . x_i$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i[A + hy_i] \quad \text{[Using (1)]}$$

$$= \frac{1}{N}\left[\sum_{i=1}^{n} f_i A + \sum_{i=1}^{n} hf_i y_i\right]$$

$$= A\frac{\sum_{i=1}^{n} f_i}{N} + h\frac{\sum_{i=1}^{n} f_i y_i}{N}$$

$$= A.\frac{N}{N} + h.\bar{y}$$

$$\implies \bar{x} = A + h.\bar{y} \quad \text{……. (2)}$$

Variance = ${\sigma_x}^2 = \frac{1}{N}\sum_{i=1}^{n} f_i(x_i - \bar{x})^2$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i(A + hy_i - A - h\bar{y})^2 \quad \text{[Using (1) and (2)]}$$

$$= \frac{1}{N}\sum_{i=1}^{n} f_i h^2 (y_i - \bar{y})^2$$

$$= \frac{h^2}{N}\sum_{i=1}^{n} f_i (y_i - \bar{y})^2$$

$$= h^2 . {\sigma_y}^2$$

$$\implies {\sigma_x}^2 = h^2 . {\sigma_y}^2$$

$$\implies \sigma_x = h . \sigma_y$$

$$i.e., \boldsymbol{\sigma_x} = \frac{\boldsymbol{h}}{\boldsymbol{N}} \sqrt{\boldsymbol{N}. \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{f_i y_i}^{\boldsymbol{2}} - \left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{f_i y_i}\right)^{\boldsymbol{2}}}$$

Now, let us take some examples to use this formula.

Example 5: Using step-deviation method, calculate Mean, Variance and Standard Deviation for the following data:

Class	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	6	7	15	16	4	2

Solution: Consider the data in tabular form:

Class	x_i	f_i	$y_i = \frac{x_i - 25}{10}$	y_i^2	$f_i y_i$	$f_i . y_i^2$
0-10	5	6	– 2	4	– 12	24
10-20	15	7	– 1	1	– 7	7
20-30	25(=A)	15	0	0	0	0
30-40	35	16	1	1	16	256
40-50	45	4	4	4	8	32
50-60	55	2	9	9	6	12
Total		N = 50			11	331

$$\bar{x} = A + \frac{\sum f_i . y_i}{N} \times h = 25 + \frac{11}{50} \times 10 = 25 + 2.2 = 27.2$$

$$\sigma^2 = \frac{h^2}{N^2}\left[N. \sum_{i=1}^{n} f_i y_i^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2\right]$$

$$= \frac{100}{50 \times 50}[50 \times 331 - 11 \times 11]$$

$$= \frac{1}{25}[16550 - 121]$$

$$= \frac{1}{25} \times 16429 = 657.16$$

$$S.D. = \sigma = \sqrt{\text{variance}} = \sqrt{657.16} = 25.63$$

Example 6: Calculate mean and standard deviation of first 'n' natural numbers.

[NCERT]

Solution: $\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$

$$\sigma^2 = \frac{1}{n}\sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2$$

$$= \frac{1^2 + 2^2 + 3^2 + \cdots .. + n^2}{n} - (\bar{x})^2$$

$$= \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n+1}{2}\left[\frac{(2n+1)}{3} - \frac{n+1}{2}\right]$$

$$= \frac{n+1}{2}\left(\frac{4n+2-3n-3}{6}\right)$$

$$= \frac{n+1}{2}\left(\frac{n-1}{6}\right)$$

$$= \frac{n^2-1}{12}$$

$$\sigma = \sqrt{\frac{n^2-1}{12}}$$

7. Real-life Applications of Variance and Standard Deviation

- Standard deviation is used for demographic analysis to get an idea what is normal trend in the population and what is needed as per the requirements to stat new programs by the government.

Source: https://scholarlykitchen.sspnet.org/2018/11/28/readership-survey-demographic-analysis/

- Variance and standard deviation are used for control system in industries. If standard or quality of any product is not being controlled it will show high rate of variance or standard deviation.

Source: https://fasttqmsoftware.com/quality-control/

- If value of the variance is low then a company is closer to its estimate during the budgeting period.

Source: https://www.cleverism.com/budgeting-process-complete-guide/

- During clinical testing of a new drug, if variance of the test is minimum then it means that the drug has more universal acceptance.


Source: https://indianexpress.com/article/business/economy/centre-plans-to-clip-wings-of-drug-regulator/

8. Summary

Let us summarize the concepts and formulas that we have learnt in this module:

- Variance is the arithmetic mean of the squares of the deviations from the mean. It is denoted by σ^2 or Var(X).
- Standard deviation is the positive square-root of the variance and is denoted by σ or S.D., i.e., S.D. = $\sigma = \sqrt{\text{variance}}$
- For ungrouped data, Variance = $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$
- For ungrouped data, Standard Deviation $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$
- For discrete data, Variance = $\sigma^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{n} f_i} = \frac{1}{N}\left[\sum_{i=1}^{n} f_i (x_i - \bar{x})^2\right]$, where N = $\sum_{i=1}^{n} f_i$
- For discrete data, Standard Deviation = $\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N}}$
- For discrete data, $\sigma = \frac{1}{N}\sqrt{N \sum_{i=1}^{n} f_i {x_i}^2 - \left(\sum_{i=1}^{n} f_i x_i\right)^2}$
- For continuous data, ${\sigma_x}^2 = \frac{h^2}{N^2}\left[N.\sum_{i=1}^{n} f_i {y_i}^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2\right]$
- For continuous data, $\sigma_x = \frac{h}{N}\sqrt{N.\sum_{i=1}^{n} f_i {y_i}^2 - \left(\sum_{i=1}^{n} f_i y_i\right)^2}$