1. Details of Module and its structure

Module Detail	
Subject Name	Mathematics
Course Name	Mathematics 02 (Class XI, Semester - 2)
Module Name/Title	Statistics: Coefficient of Variance- Part 3
Module I	kemh_21503
Pre-requisites	Measures of Central Tendency
Objectives	After going through this module, the learner will be able to: 1. Compare two data by using coefficient of variance 2. Solve practical problems using measures of variation
Keywords	Coefficient of Variance, dispersion, frequency distribution

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 is also used to analyze data and draw many inferences. Two or more data can also be compared by using measures of dispersion. Variability or consistency of any data can be compared by coefficient of variation which is calculated by using standard deviation and mean.

In this module, we will learn to use coefficient of variation to compare data.

2. Analysis of Frequency Distributions

Suppose two candidates appear for an interview. Both have got same average score in their post-graduation.

Source: https://www.insperity.com/blog/professional-recruiters-reveal-16-of-the-best-interview-questions-to-ask/

Interviewer was in dilemma as to whom is to be selected as both have impressed her with their skills. Then she took help of statistics and rechecked their academic records in a new way. After her analysis, the candidate who got selection is the one having more consistent record.

Amazed! How can statistics check the consistency? To know this tool of statistics which is used to check the consistency or variability of data, let us go through this module.

Mean deviation and standard deviation are measured in same units in which the data is given. Thus, measures of dispersion are not reliable to compare two or more data which are measured in different units. So, to compare data, we must have some tools of statistics which are independent of units. One such tool is coefficient of variation, which is defined as

Coefficient of Variation = C.V. = $\boldsymbol{\frac{\sigma}{\bar{x}} \times 100;\ \bar{x} \neq 0}$
where, $\boldsymbol{\sigma}$ **is standard deviation and** $\boldsymbol{\bar{x}}$ **is mean of the data.**

The data having greater C.V. has more variability or less consistency. The data having lesser C.V. is more consistent and homogeneous.

Suppose two data have same value of mean but different values of standard deviations then how can C.V. be applied to compare the data?

Consider, C.V. (1^{st} data) = $\frac{\sigma_1}{\bar{x}} \times 100$
And C.V. (2^{nd} data) = $\frac{\sigma_2}{\bar{x}} \times 100$

Ratio of C.V.s = $\frac{\frac{\sigma_1}{\bar{x}} \times 100}{\frac{\sigma_2}{\bar{x}} \times 100} = \frac{\sigma_1}{\sigma_2}$

Thus, two data can be compared on the basis of their standard deviations also, provided they have same value of their means. Here, data having more value of standard deviation is more variable or less consistent and vice-versa.

Let us take some problems based on this concept and learn how to compare two data.

Example 1: Following are the marks obtained, out of 100, by two students Ravi and Hashina in 10 tests:

Ravi	25	50	45	30	70	42	36	48	35	60
Hashina	10	70	50	20	95	55	42	60	48	80

Who is more intelligent and who is more consistent?

[NCERT Exemplar]

Solution:

For Ravi			**For Hashina**		
x_i	$d_i = x_i - 45$	d_i^2	x_i	$d_i = x_i - 55$	d_i^2
25	– 20	400	10	– 45	2025
50	5	25	70	25	625
45	0	0	50	– 5	25
30	– 15	225	20	– 35	1225
70	25	625	95	40	1600
42	– 3	9	55	0	0
36	– 9	81	42	– 13	169
48	3	9	60	5	25

35	– 10	100	48	– 7	49
60	15	225	80	25	625
Total	– 14	1699	Total	0	6368

For Ravi,

$$\bar{x} = A + \frac{\sum d_i}{\sum f_i}$$

$$= 45 - \frac{14}{10}$$

$$= 45 - 1.4 = 43.6$$

$$\sigma = \sqrt{\frac{\sum d_i^2}{\sum f_i} - \left(\frac{\sum d_i}{\sum f_i}\right)^2}$$

$$= \sqrt{\frac{1699}{10} - \left(\frac{-14}{10}\right)^2} = \sqrt{169.9 - 0.0196} = \sqrt{169.88} = 13.03$$

C.V. = $\frac{\sigma}{\bar{x}} \times 100$

= $\frac{13.03}{43.6} \times 100 = 29.88$

For Hashina,

$$\bar{x} = A + \frac{\sum d_i}{\sum f_i}$$

$$= 55 - \frac{0}{10}$$

$$= 55$$

$$\sigma = \sqrt{\frac{\sum d_i^2}{\sum f_i} - \left(\frac{\sum d_i}{\sum f_i}\right)^2}$$

$$= \sqrt{\frac{6368}{10} - 0}$$

$$= \sqrt{636.8} = 25.2$$

C.V. = $\frac{\sigma}{\bar{x}} \times 100$

= $\frac{25.2}{55} \times 100 = 45.89$

Since, C.V. for Ravi is less; so Ravi is more consistent and intelligent.

Example 2: Life of bulbs produced by two factories A and B are given below:

Length of Life (in hours)	550 – 650	650 – 750	750 – 850	850 – 950	950 – 1050
Factory A (Number of bulbs)	10	22	52	20	16
Factory B (Number of bulbs)	8	60	24	16	12

The bulbs of which factory are more consistent from the point of view of length of life?

[NCERT Exemplar]

Solution: For factory A

Class	x_i	f_i	$y_i = \frac{x_i - 800}{100}$	y_i^2	$f_i y_i$	$f_i . y_i^2$
550-650	600	10	– 2	4	– 20	40
650-750	700	22	– 1	1	– 22	22
750-850	800 (=A)	52	0	0	0	0
850-950	900	20	1	1	20	20
950-1050	1000	16	2	4	32	64
Total		120			10	146

$$\bar{x} = A + \frac{\sum f_i . y_i}{N} \times h$$

$$= 800 + \frac{10}{120} \times 100 = 800 + 8.33 = 808.33$$

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

$$= \frac{100}{120}\sqrt{120 \times 146 - 10^2}$$

$$= \frac{5}{6}\sqrt{17520 - 100}$$

$$= \frac{5}{6}\sqrt{17420}$$

$$= \frac{5}{6} \times 131.98$$

$$= 0.83 \times 131.98 = 109.54$$

$$C.V\ (A) = \frac{\sigma}{\bar{x}} \times 100 = \frac{808.33}{109.54} \times 100 = 737.93$$

<u>For factory B</u>

Class	x_i	f_i	$y_i = \frac{x_i - 800}{100}$	y_i^2	$f_i y_i$	$f_i . y_i^2$
550-650	600	8	– 2	4	– 16	32
650-750	700	60	– 1	1	– 60	60
750-850	800 (=A)	24	0	0	0	0
850-950	900	16	1	1	16	16
950-1050	1000	12	2	4	24	48
Total		120			– 36	156

$$\bar{x} = A + \frac{\sum f_i . y_i}{N} \times h$$

$$= 800 - \frac{36}{120} \times 100 = 800 - 30 = 770$$

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

$$= \frac{100}{120} \sqrt{120 \times 156 - (-36)^2}$$

$$= \frac{5}{6} \sqrt{18720 - 1296}$$

$$= \frac{5}{6} \sqrt{17424}$$

$$= \frac{5}{6} \times 132$$

$$= \frac{660}{6} = 110$$

$$C.V\ (A) = \frac{\sigma}{\bar{x}} \times 100 = \frac{770}{110} \times 100 = 700$$

Since, C.V. of factory A is greater than that of B, therefore, factory B is more consistent.

Example 3: The mean and variance of cumulative marks scored by four studentsin examinations are given below:

Students	Student A	Student B	Student C	Student D
Mean	82	84	98	78
Variance	121	289	81	25

Which of the students shows highest consistency?

Solution:Consider the given data:

Students	Student A	Student B	Student C	Student D
Mean	82	84	98	78
Variance	121	289	81	25
Standard Deviation	11	17	9	5

Consistency can be checked by using C.V.

C.V. (Student A) = $\frac{\sigma}{\bar{x}} \times 100 = \frac{11}{82} \times 100 = 13.41$

C.V. (Student B) = $\frac{17}{84} \times 100 = 20.23$

C.V. (Student C) = $\frac{9}{98} \times 100 = 9.18$

C.V. (Student D) = $\frac{5}{78} \times 100 = 6.41$

Since, C.V of student D is least, therefore, he shows the highest consistency.

Example 4: Given below are details of heights (in cm) of boys and girls of Class XI:

	Boys	Girls
Mean (of height)	165	152
Variance (of height)	25	36

Which of the distributions is more variable?

Solution: Consider the given data:

	Boys	Girls
Mean (of height)	165	152
Variance (of height)	25	36
Standard Deviation	5	6

Variability can be checked by using C.V.

C.V. (Boys) = $\frac{\sigma}{\bar{x}} \times 100 = \frac{5}{165} \times 100 = 3.03$

C.V. (Girls) = $\frac{6}{152} \times 100 = 3.94$

Since, C.V. for girls is more, therefore, girls shows more variability.

3. Miscellaneous Problems on Measure of Dispersion

Let us take some miscellaneous problems on measures of dispersion.

Example 5: Let $x_1, x_2, x_3, \ldots, x_n$ be 'n' values of a variable X. If these values are changed to $x_1 + a, x_2 + a, x_3 + a, \ldots, x_n + a$; for $a \in R$, show that the variance remains unchanged. [NCERT]

Solution: Let $y_i = x_i + a; i = 1, 2, 3, \ldots, n$

$$\bar{y} = \frac{\sum y_i}{n}$$

$$= \frac{\sum (x_i + a)}{n}$$

$$= \frac{(\sum x_i) + na}{n}$$

$$= \frac{\sum x_i}{n} + a$$

$$\Rightarrow \bar{y} = \bar{x} + a$$

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$$

$$= \frac{\sum [(x_i + a) - (\bar{x} + a)]^2}{n}$$

$$= \frac{\sum (x_i + a - \bar{x} - a)^2}{n}$$

$$= \frac{\sum (x_i - \bar{x})^2}{n}$$

$$= \sigma_x^2$$

Thus, variance of y is same as variance of x.

Hence, variance remains unchanged if value of each observations is increased by the same unit.

Example 6: Let $x_1, x_2, x_3, \ldots, x_n$ be 'n' values of a variable X and let 'a' be a non-zero real number. Then prove that the variance of the observations $ax_1, ax_2, ax_3, \ldots, ax_n$ is $a^2\sigma^2$. [NCERT]

Solution: Let $y_1, y_2, y_3, \ldots, y_n$ be 'n' values of y such that $y_i = a.x_i$; i = 1, 2, 3,…, n

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum a.x_i}{n} = a.\frac{\sum x_i}{n} = a.\bar{x}$$

Consider, $\sigma_y^2 = \frac{\sum(y_i - \bar{y})^2}{n}$

$$= \frac{\sum(a.x_i - a.\bar{x})^2}{n}$$

$$= \frac{a^2 \sum(x_i - \bar{x})^2}{n}$$

$$= a^2.\sigma_x^2$$

Hence, variance of the observations $ax_1, ax_2, ax_3, \ldots, ax_n$ is $a^2\sigma_x^2$

Example 7: If the mean and standard deviation of 100 observations are 50 and 4 respectively, find the sum of the observations and sum of their squares.

[NCERT Exemplar]

Solution: $x_1, x_2, x_3, \ldots, x_{100}$ be 100 observations having mean $\bar{x}$ and standard deviation σ.

$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{100}$$

$$\Rightarrow 50 = \frac{\sum_{i=1}^{100} x_i}{100}$$

$$\Rightarrow 50 \times 100 = \sum_{i=1}^{100} x_i$$

$$\Rightarrow \sum_{i=1}^{100} x_i = 5000$$

Given that, $\sigma = 4$

$\Rightarrow \sigma^2 = 16$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{100} x_i - (\bar{x})^2 = 16$$

$$\Rightarrow \frac{1}{100}\sum_{i=1}^{100} x_i - 50^2 = 16$$

$\Rightarrow \frac{1}{100}\sum_{i=1}^{100} x_i$ = 16 + 2500 = 2516

$$\Rightarrow \sum_{i=1}^{100} x_i = 2516 \times 100 = 251600$$

Example 8: If for a distribution of 18 observations, $\sum(x_i - 5) = 3$ and $\sum(x_i - 5)^2 = 43$, find mean and standard deviation.

[NCERT Exemplar]

Solution: Given, $\sum_{i=1}^{18}(x_i - 5) = 3$

$$\Rightarrow \sum_{i=1}^{18} x_i - \sum_{i=1}^{18} 5 = 3$$

$$\Rightarrow \sum_{i=1}^{18} x_i - 18 \times 5 = 3$$

$$\Rightarrow \sum_{i=1}^{18} x_i = 3 + 90 = 93$$

$$\Rightarrow \bar{x} = \frac{\sum_{i=1}^{18} x_i}{18} = \frac{93}{18} = 5.17$$

Consider, $\sum_{i=1}^{18}(x_i - 5)^2 = 43$

$$\Rightarrow \sum_{i=1}^{18}(x_i^2 - 10x_i + 25) = 43$$

$$\Rightarrow \sum_{i=1}^{18} x_i^2 - 10\sum_{i=1}^{18} x_i + \sum_{i=1}^{18} 25 = 43$$

$$\Rightarrow \sum_{i=1}^{18} x_i^2 - 10 \times 93 + 18 \times 25 = 43$$

$$\Rightarrow \sum_{i=1}^{18} x_i^2 - 930 + 450 = 43$$

$$\Rightarrow \sum_{i=1}^{18} x_i^2 = 43 + 480 = 523$$

$$\sigma = \sqrt{\frac{1}{18}\sum_{i=1}^{18} x_i^2 - \left(\frac{1}{18}\sum_{i=1}^{18} x_i\right)^2}$$

$$= \sqrt{\frac{523}{18} - \left(\frac{93}{18}\right)^2}$$

$$= \sqrt{\frac{9414 - 8649}{324}}$$

$$= \frac{\sqrt{765}}{18}$$

$$= \frac{27.65}{18}$$

$$= 1.536$$

Example 9: While calculating the mean and variance of 10 readings, a student wrongly used the reading of 52 for the correct reading 25. He obtained the mean and variance as 45 and 16 respectively. Find the correct mean and variance.

[NCERT Exemplar]

Solution: We have, n = 10

Incorrect Mean = 45 and Incorrect Variance = 16

$$\Rightarrow \frac{\sum x_i}{10} = 45$$

$\Rightarrow \sum x_i = 45 \times 10 = 450$ ………… (1)

Correct $\sum x_i$ = Incorrect $\sum x_i$ – Incorrect Value + Correct Value

$\Rightarrow$ Correct $\sum x_i$ = 450 – 52 + 25

= 450 – 27 = 423

$\Rightarrow$ Correct $\bar{x} = \frac{\text{Correct} \sum x_i}{10} = \frac{423}{10} = 42.3$

Given, incorrect variance = 16

$$\Rightarrow \frac{\text{Incorrect} \sum x_i^2}{10} - (\text{Incorrect } \bar{x})^2 = 16$$

$$\Rightarrow \frac{\text{Incorrect} \sum x_i^2}{10} - 45^2 = 16$$

$$\Rightarrow \frac{\text{Incorrect} \sum x_i^2}{10} = 16 + 2025 = 2041$$

$$\Rightarrow \text{Incorrect} \sum x_i^2 = 2041 \times 10 = 20410$$

Correct $\sum x_i^2$ = Incorrect $\sum x_i^2$ – Square of Incorrect Value + Square of Correct Value

$\Rightarrow$Correct $\sum x_i^2$ = 20410 $-52^2 + 25^2$

= 20410 – 2704 + 625

= 20410 – 2079

= 18331

Correct Variance = $\frac{\text{Correct } \sum x_i^2}{10} - (\text{Correct } \bar{x})^2$

= $\frac{18331}{10} - (42.3)^2$

= 1833.1 – 1789.29

= 43.81

4. **Real life Applications of Coefficient of Variance**

We know that statistics helps us a good deal in our daily lives.

- In stock market or investing business, the coefficient of variance gives a fair idea as to how much risk can be taken to invest in a stock. The lower the value of C.V., more is the consistency of the stock.

Source: https://economictimes.indiatimes.com/markets/stocks/news/d-street-week-ahead-market-wont-see-directional-move-defensive-plays-to-be-in-focus/articleshow/70616769.cms

- To compare the consistency of players while selecting them for a tournament can also be done by using C.V. as it gives a fair idea about the scores and the consistency.

Source: https://www.noted.co.nz/archive/archive-listener-nz-2013/jane-clifton-always-the-bridesmaid

- To select a candidate for a course admission / al-rounder prize / job on the basis of the achievements throughout the session can also be done by using C.V. for comparing achievements of two or more candidates.

Source: https:a//www.flipkart.com/box-18-185-worlds-best-all-rounder trophy/p/itmes7t8whehfftz

5. Summary

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

- Variability or consistency of any data can be compared by using standard deviation.
- Coefficient of Variation = C.V. = $\frac{\sigma}{\bar{x}} \times 100$; $\bar{x} \neq 0$; where, σ is standard deviation and $\bar{x}$ is mean of the data.
- The data having greater C.V. has more variability or less consistency. The data having lesser C.V. is more consistent and homogeneous.
- Two data can be compared on the basis of their standard deviations also, provided they have same value of their means.
- Ratio of C.V.s = $\frac{\frac{\sigma_1}{\bar{x}}\times 100}{\frac{\sigma_2}{\bar{x}}\times 100} = \frac{\sigma_1}{\sigma_2}$, provided mean of the distributions is equal.