Welcome to

Aakash + BYJU'S LIVE

Statistics

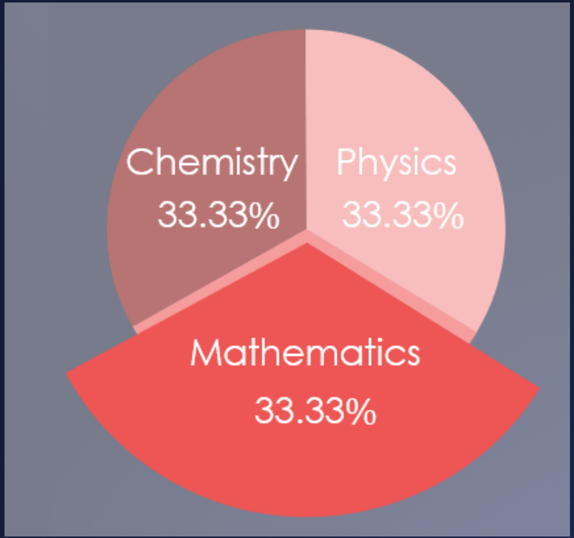
Table of contents

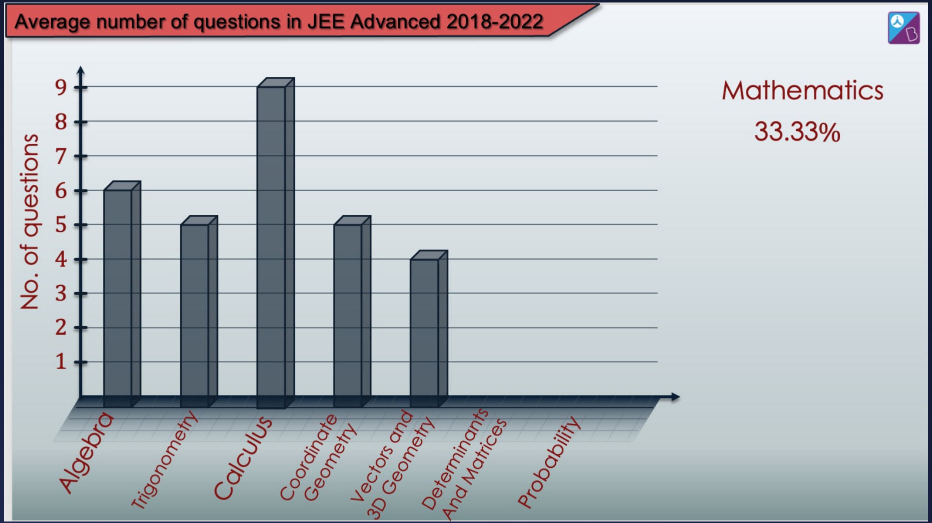
Session 01

Introduction to Statistics

Return To Top

Key Takeaways





Statistics:

Statistics is the branch of mathematics which deals with collection, organisation, analysis, interpretation and presentation of numerical data.

Return To Top

Key Takeaways

Types of Data:

- Types of Data
  - Ungrouped Data (Raw data)
  - Grouped Data
    - Discrete Frequency Distribution
    - Continuous Frequency Distribution

Example :

Data of marks of a student in 10 different tests:

$61, 48, 54, 49, 61, 61, 61, 49, 54, 52$

Discrete Frequency Distribution

Marks	f
49	2
52	1
54	2
61	4
48	1

Continuous Frequency Distribution

Class	f_i
$45-50$	3
$50-55$	3
$55-60$	0
$60-65$	4

Return To Top

Measures of Central Tendency:

An average value or central value of a distribution is the value of a variable which is representative of the entire distribution, this representative value is called the measure of central tendency.

It can be of following types:

- Mean
- Median
- Mode

Return To Top

Measures of Central Tendency:

For Ungrouped data:

Mean: It is the average value of all the observations.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example :

Data of marks of a student in 10 different tests:

$61, 48, 54, 49, 61, 61, 61, 49, 54, 52$

$$\bar{x} = \frac{61 + 48 + 54 + 49 + 61 + 61 + 61 + 49 + 54 + 52}{10} = 55$$

Return To Top

Measures of Central Tendency:

For Discrete Frequency Distributions:

Mean:

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \cdots + f_nx_n}{f_1 + f_2 + f_3 + \cdots + f_n} = \frac{\sum_{i=1}^{n} f_ix_i}{\sum_{i=1}^{n} f_i}$$

Example :

x_i	f_i	f_ix_i
2	1	2
3	2	6
5	3	15
7	4	28
9	5	45
	15	96

$$\bar{x} = \frac{96}{15} = 6.4$$

Return To Top

If for some $x \in \mathbb{R}$, the frequency distribution of the marks obtained by 20 students in a test is:

Marks	2	3	5	7
Frequency	$(x+1)^2$	$2x-5$	x^2-3x	x

then the mean of the marks is:

JEE Main 2019

Solution:

Given there are 20 students.

$$\therefore \sum f_i = (x+1)^2 + (2x-5) + x^2 - 3x + x = 20$$

$$\Rightarrow 2x^2 + 2x - 24 = 0$$

$$\Rightarrow x^2 + x - 12 = 0 \Rightarrow x = 3, -4 \text{ (rejected)}$$

$$\because \text{ Mean } (\overline{x}) = \frac{\sum f_i x_i}{\sum f_i} = \frac{2\cdot(x+1)^2 + 3\cdot(2x-5) + 5\cdot(x^2-3x) + 7\cdot x}{20} = \frac{2\cdot(4)^2 + 3\cdot 1 + 5\cdot 0 + 7\cdot 3}{20} = \frac{56}{20} = 2.8$$

A 3.0

B 2.5

C 2.8

D 3.2

Return To Top

Key Takeaways

Measures of Central Tendency:

For Ungrouped data:

Median: It is the middle value when the observations are arranged in increasing or decreasing order.

Case 1: When n is odd

$$\boxed{M = \left(\frac{n+1}{2}\right)^{th} \text{ term}}$$

Case 2: When n is even

$$\boxed{M = \text{Average of } \left(\frac{n}{2}\right)^{th} \text{term } \& \left(\frac{n}{2}+1\right)^{th} \text{ term}}$$

Return To Top

Key Takeaways

Measures of Central Tendency:

For Ungrouped data:

Example:

Given: Data of marks of a student in 10 different tests:

$61, 48, 54, 49, 61, 61, 61, 49, 54, 52$

After arranging in ascending order,

$48, 49, 49, 52, 54, 54, 61, 61, 61, 61$

$$M = \frac{\left(\frac{n}{2}\right)^{th} \text{term} + \left(\frac{n}{2}+1\right)^{th} \text{term}}{2} = \frac{5^{th}\text{term} + 6^{th}\text{term}}{2} = \frac{54+54}{2} = 54$$

Return To Top

Key Takeaways

Measures of Central Tendency:

For Discrete Frequency Distributions:

Median:

Case 1: When n is odd

$$\boxed{M = \left(\frac{n+1}{2}\right)^{th} \text{ term}}$$

Case 2: When n is even

$$\boxed{M = \text{Average of } \left(\frac{n}{2}\right)^{th} \text{term \& } \left(\frac{n}{2}+1\right)^{th} \text{ term}}$$

Where n is the Cumulative frequency.

Return To Top

Key Takeaways

Measures of Central Tendency:

For Discrete Frequency Distributions:

Example :

x_i	f_i	$C.F.$
2	1	1
3	2	3
5	3	6
7	4	10
9	5	15

$$M = \left(\frac{n+1}{2}\right)^{th} \text{ term} = \frac{15+1}{2} = 8^{th} \text{ term}$$

The closest C.F. greater than 8 is 10.

The term corresponding to C.F. 10 is 7.

$\therefore$ Median $= 7$

Return To Top

The mean and the median of the following ten numbers in increasing order $10, 22, 26, 29, 34, x, 42, 67, 70, y$ are 42 and 35 respectively, then $\frac{y}{x}$ is equal to:

JEE Main 2019

A $\frac{7}{2}$

B $\frac{8}{3}$

C $\frac{7}{3}$

D $\frac{9}{4}$

Solution:

Mean $= 42$

$$\Rightarrow \frac{10+22+26+29+34+x+42+67+70+y}{10} = 42$$

$$\Rightarrow x + y = 120$$

Since there are 10 terms, median will be mean of middle two terms.

Median $= \frac{34+x}{2} = 35 \Rightarrow x = 36$

$$\therefore y = 84$$

Hence, $\frac{y}{x} = \frac{7}{3}$

Return To Top

Mean & Median for Continuous Frequency Distribution

$$\text{Mean} = A + \frac{1}{N}\sum f_i d_i$$

$$\text{Median}(M) = l + \left[\frac{\left(\frac{N}{2}\right) - c}{f}\right] \times h$$

Where N = total frequency $= \sum f$

A = assumed mean

d_i = deviation of A from each of x_i i.e. $d_i = x_i - A$

l = lower limit of median class

f = frequency of the median class

c = cumulative frequency of the class preceding the median class

h = class interval (width) of the median class

Return To Top

Let the assumed mean $A = 15$

Class Interval	x_i	f_i	$x_i f_i$	d_i	$f_i d_i$	C.F.
$0 - 6$	3	a	$3a$	-12	$-12a$	a
$6 - 12$	9	b	$9b$	-6	$-6b$	$a + b$
$12 - 18$	15	12	180	0	0	$12 + a + b$
$18 - 24$	21	9	189	6	54	$21 + a + b$
$24 - 30$	27	5	135	12	60	$26 + a + b$
		$a + b + 26$	$3a + 9b + 504$		$114 - 12a - 6b$	

$$\text{Mean} = A + \frac{1}{N}\sum f_i d_i$$

$$\text{Mean} = 15 + \frac{114 - 12a - 6b}{a + b + 26} = \frac{309}{22}$$

$$\Rightarrow 81a + 37b = 1018 \cdots (i)$$

Return To Top

Let the assumed mean $A = 15$

Class Interval	x_i	f_i	$x_i f_i$	d_i	$f_i d_i$	C.F.
$0-6$	3	a	$3a$	-12	$-12a$	a
$6-12$	9	b	$9b$	-6	$-6b$	$a+b$
$12-18$	15	12	180	0	0	$12+a+b$
$18-24$	21	9	189	6	54	$21+a+b$
$24-30$	27	5	135	12	60	$26+a+b$
		$a+b+26$	$3a+9b+504$		$114-12a-6b$	

Median Class

$$\Rightarrow 81a + 37b = 1018 \cdots (i)$$

$$\text{Median} = 12 + \frac{13 + \frac{a+b}{2} - (a+b)}{12} \times 6 = 14 \Rightarrow a + b = 18 \cdots (ii)$$

From (i) and (ii), $a = 8$ and $b = 10$

$$\therefore (a-b)^2 = 4$$

$$\text{Median}(M) = l + \left[\frac{\left(\frac{N}{2}\right) - c}{f}\right] \times h$$

Return To Top

Key Takeaways

Measures of Central Tendency:

For Ungrouped data:

Mode: It is the observation with maximum frequency.

Example :

Given: Marks of a student in 10 different tests are

$61, 48, 54, 49, 61, 61, 61, 49, 54, 52$

Mode of the given data is 61 since it has maximum frequency.

Return To Top

Key Takeaways

Measures of Central Tendency:

For Discrete Frequency Distributions:

For discrete frequency distributions, mode is simply the distribution with highest frequency.

Example:

x_i	f_i
2	1
3	2
5	3
7	4
9	5

Mode : 9

Return To Top

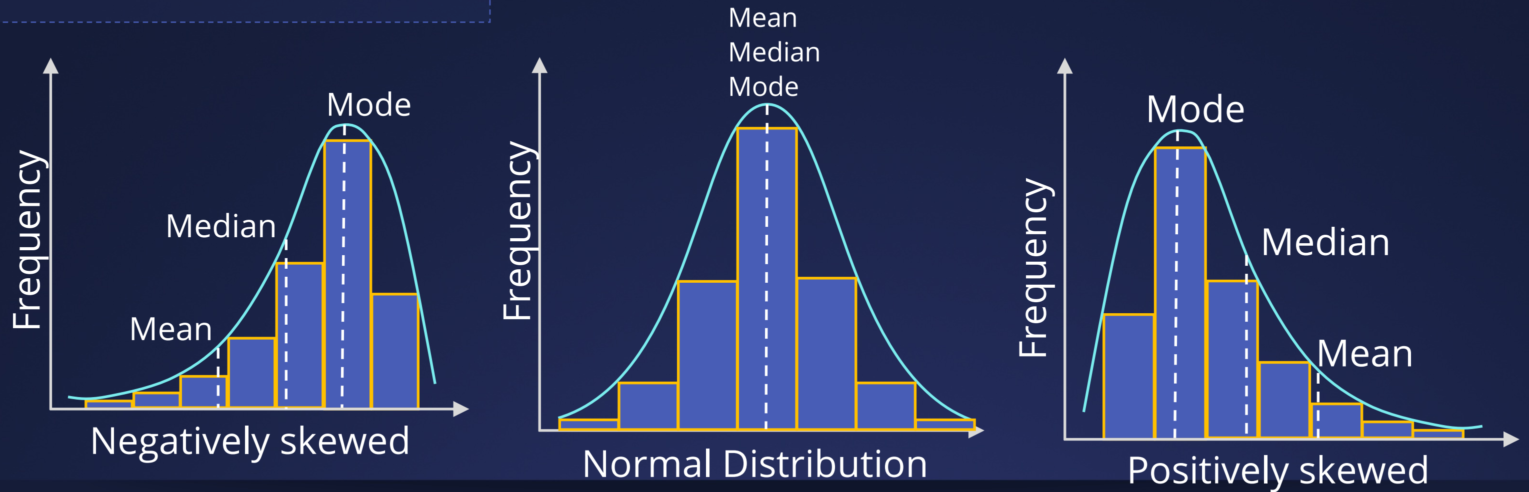
Key Takeaways

Measures of Central Tendency:

Empirical Relationship between Mean, Median and Mode

The formula to define the relation between mean, median and mode in a moderately skewed distribution is

$$\boxed{\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}}$$



Return To Top

In a frequency distribution, the mean and median are 21 & 22 respectively, then its mode is approximately.

Solution:

Mode = 3 median – 2 mean

$$= 3\times22 - 2\times21$$

$$= 3(22 - 14)$$

$$= 3\times8$$

$$= 24$$

A 20.5

B 24.0

C 22.0

D 25.5

Return To Top

Measure of Dispersion

It is measure of deviation of its value about their central values. It gives an idea of scatteredness of different values from the central values.

It has four types:

- Range
- Mean deviation
- Variance
- Standard deviation

Return To Top

Measure of Dispersion

Range

Consider $x_1, x_2, x_3, \cdots, x_n$ be data, then

$$\boxed{\text{Range} = x_{max} - x_{min}}$$

$$\boxed{\text{Coefficient of Range} = \frac{x_{max} - x_{min}}{x_{max} + x_{min}}}$$

Example

Let $20, 15, 12, 11, 10$ & 9 be data,

$\text{Range} = 20 - 9 = 11$

$$\text{Coefficient of Range} = \frac{20 - 9}{20 + 9} = \frac{11}{29}$$

Return To Top

Key Takeaways

Mean Deviation:

Mean deviation of a distribution is the mean of absolute value of deviation of variate from their statistical average (median, mean or mode).

If A is any statistical average the mean deviation about A is defined as

$$\text{Mean deviation} = \frac{1}{N}\sum_{i=1}^{n}|x_i - A|$$

Return To Top

Key Takeaways

Mean Deviation:

If A is any statistical average the mean deviation about A is defined as

$$\text{Mean deviation} = \frac{1}{N}\sum_{i=1}^{n}|x_i - A|$$

Example :

Given a data set $\{5, 3, 7, 8, 4, 9\}$, what is the mean deviation about the mean?

Solution: $$\text{Mean} = \frac{5+3+7+8+4+9}{6} = \frac{36}{6} = 6$$

$$\text{Mean Deviation} = \frac{|5-6|+|3-6|+|7-6|+|8-6|+|4-6|+|9-6|}{6} = \frac{1+3+1+2+2+3}{6}$$

$$= \frac{12}{6} = 2$$

Return To Top

Key Takeaways

Mean Deviation:

If A is any statistical average the mean deviation about A is defined as

$$\text{Mean deviation} = \frac{1}{N}\sum_{i=1}^{n}|x_i - A|$$

For frequency distributions,

$$\text{Mean deviation} = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - A|$$

Return To Top

If the mean deviation about the median of the numbers $a, 2a, \cdots, 50a$ is 50, then $|a|$ equals:

AIEEE 2011

Solution:

$a, 2a, \cdots, 50a \longrightarrow$ 50 terms

When n is even, Median $= \dfrac{\left(\frac{n}{2}\right)^{th}\text{term} + \left(\frac{n}{2}+1\right)^{th}\text{ term}}{2}$

Median: $M = \dfrac{25a + 26a}{2} = (25.5)a$

A 2

B 3

C 4

D 5

Return To Top

If the mean deviation about the median of the numbers $a, 2a, \cdots, 50a$ is 50, then $|a|$ equals:

AIEEE 2011

Solution:

Mean deviation $= \dfrac{1}{N}\sum_{i=1}^{n}|x_i - A|$

Median: $M = (25.5)a$

$$\text{Mean deviation} = \frac{\sum_{i=1}^{50}|x_i - M|}{50} = 50$$

$$\frac{|a-(25.5)a| \quad +|2a-(25.5)a| \quad +\cdots \quad +|50a-(25.5)a|}{50} = 50$$

$$2\times\frac{|(0.5)a| + |(1.5)a| + \cdots + |(24.5)a|}{50} = 50$$

$$\frac{2}{50}\times\frac{25}{2}(0.5|a| + 24.5|a|) = 50$$

$$\Rightarrow 625|a| = 2500 \qquad \Rightarrow |a| = 4$$

A 2

B 3

C 4

D 5

Return To Top

Session 02

Variance and its Properties

Return To Top

Property of Mean deviation

Mean deviation is independent of change of origin, but dependent on change of scale.

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X

Mean deviation(M.D.) $= \dfrac{1}{N}\displaystyle\sum_{i=1}^{n}|x_i - A|$

After multiplying with constant a and adding another constant b in each observation,

i.e. $ax_1 + b\,,\ ax_2 + b\,,\ \cdots\cdots\ ,\ ax_n + b$

New Mean deviation $= a \times$ M.D.

Return To Top

The mean deviation of an ungrouped data is 50. If each observation is increased by 2%, then the new mean deviation is:

A 50

B 51

C 49

D 50.5

Solution:

$$x'_i = \left(1 + \frac{2}{100}\right) x_i = 1.02x_i$$

M.D. $\rightarrow a \times$ M.D.

Thus, new mean deviation $= 1.02 \times 50 = 51$

Return To Top

Key Takeaways

Variance

It is the mean of squares of deviation of variate from their mean.

It is denoted by σ^2 or $\text{var}\,(x)$.

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X, then

$$\sigma^2 = \frac{1}{n} \cdot \sum (x_i - \overline{x})^2$$

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

$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 + \frac{(\overline{x})^2}{n} \cdot \sum_{i=1}^{n} 1 - \frac{2}{n} \cdot \overline{x} \sum_{i=1}^{n} x_i$$

Return To Top

Key Takeaways

Variance

It is the mean of squares of deviation of variate from their mean.

It is denoted by σ^2 or $\text{var}\,(x)$.

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X, then

$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 + \frac{(\overline{x})^2}{n} \cdot \sum_{i=1}^{n} 1 - \frac{2}{n} \cdot \overline{x} \sum_{i=1}^{n} x_i$$

$$\sigma^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 + (\overline{x})^2 - 2(\overline{x})^2 = \underbrace{\boxed{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}}_{\text{Mean of Squares}} - \underbrace{\boxed{(\overline{x})^2}}_{\text{Square of Mean}}$$

Return To Top

Key Takeaways

Variance

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

Note

Variance of discrete frequency distribution:

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X and corresponding frequencies of them are $f_1, f_2 \cdots\cdots, f_n$.

$$\mathrm{Var}(X) = \frac{1}{N}\left(\sum f_i(x_i - \bar{x})^2\right) = \frac{1}{N}\sum f_i x_i^2 - \left(\frac{1}{N}\sum f_i x_i\right)^2$$

Return To Top

The variance of first n natural numbers is ______ .

A. $\dfrac{n^2+1}{6}$

B. $\dfrac{n^2+1}{12}$

C. $\dfrac{n^2-1}{6}$

D. $\dfrac{n^2-1}{12}$

Solution:

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

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

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

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

Return To Top

Standard deviation

The standard deviation(s OR σ) is defined as the positive square root of the variance.

Return To Top

The mean and standard deviation of 15 observations are found to be 8 and 3 respectively. On rechecking it was found that, in the observations, 20 was misread as 5. Then, the correct variance is equal to ______.

JEE Main 2022

Solution: Let the observations be $x_1, x_2, \dots, x_{15}$

$$\frac{\sum_{i=1}^{14} x_i + 5}{15} = 8 \quad \Rightarrow \sum_{i=1}^{14} x_i = 115$$

Corrected Mean : $\overline{X_c} = \dfrac{\sum_{i=1}^{14} x_i + 20}{15} = \dfrac{115 + 20}{15} = 9$

Variance $= (S.D.)^2 = 9$

Return To Top

The mean and standard deviation of 15 observations are found to be 8 and 3 respectively. On rechecking it was found that, in the observations, 20 was misread as 5. Then, the correct variance is equal to ______.

JEE Main 2022

Solution: Let the observations be $x_1, x_2, \ldots, x_{15}$

Corrected Mean : $\overline{X_c} = 9$

Variance $= (S.D.)^2 = 9$

$$\frac{\sum_{i=1}^{14} x_i^2 + 5^2}{15} - 8^2 = 9$$

$$\sum_{i=1}^{14} x_i^2 = 1070$$

Return To Top

The mean and standard deviation of 15 observations are found to be 8 and 3 respectively. On rechecking it was found that, in the observations, 20 was misread as 5. Then, the correct variance is equal to ______.

JEE Main 2022

Solution: Let the observations be $x_1, x_2, \ldots, x_{15}$

Corrected Mean : $\overline{X_c} = 9$

Variance $= (S.D.)^2 = 9$ $\quad \sum_{i=1}^{14} x_i^2 = 1070$

Correct Variance:

$$\sigma^2 = \frac{\sum_{i=1}^{14} x_i^2 + 20^2}{15} - 9^2 = \frac{1070 + 400}{15} - 81 = 98 - 81 = 17$$

Return To Top

Key Takeaways

Variance

Property: Variance is independent of change of origin, but dependent on change of scale.

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X,

$$\boxed{\operatorname{var}(x_i + b) = (\operatorname{var} x_i)}$$

Example : Variance of $(1,2,3,4) = \frac{4^2-1}{12}$

On adding 3 to each term,

Variance of $(4,5,6,7) = \frac{4^2+5^2+6^2+7^2}{4} - \left(\frac{4+5+6+7}{4}\right)^2 = \frac{126-121}{4} = \frac{5}{4}$

Return To Top

Key Takeaways

Variance

Property: Variance is independent of change of origin, but dependent on change of scale.

If $x_1, x_2, \cdots\cdots, x_n$ are n values of a variable X,

$$\boxed{\operatorname{var}(x_i + b) = (\operatorname{var} x_i)}$$

$$\boxed{\operatorname{var}(ax_i + b) = a^2(\operatorname{var} x_i)}$$

Return To Top

If variance of first n natural numbers is 10 and variance of first m even natural numbers is 16, then $m+n$ is equal to:

JEE Main 2020

Solution:

Variance of first n natural numbers $= \frac{n^2-1}{12}$

$\mathrm{var}(ax_i + b) = a^2(\mathrm{var}\, x_i)$

$$\frac{n^2-1}{12} = 10$$

$$\Rightarrow n = 11$$

Variance of $(2,4,6\ldots) = 4\times$ variance of $(1,2,3,4\ldots)$

$$= 4\times\left(\frac{m^2-1}{12}\right)$$

$$\Rightarrow \frac{m^2-1}{3} = 16 \Rightarrow m = 7$$

$$\therefore n+m = 11+7 = 18$$

Return To Top

In a series of $2n$ observations, half of them are equal to a and remaining half are equal to $-a$. Also, by adding a constant b in each of these observations, the mean and standard deviation of new set become 5 and 20, respectively. Then the value of $a^2 + b^2$ is equal to:

JEE Main 2021

A 250

B 925

C 650

D 425

Solution:

Given series is $a(n$ times$), -a(n$ times$)$

$\overline{x} = 0$

If b is added to each of them, then the new mean is

$\overline{x'} = \overline{x} + b \Rightarrow b = 5$

Standard deviation doesn't change when same number is added to all, so

$$\sqrt{\frac{na^2 + na^2}{2n} - 0} = 20 \Rightarrow a^2 = 400 \therefore a^2 + b^2 = 425$$

Return To Top

Key Takeaways

Combined Variance/Standard Deviation

If there are two sets of observations containing n_1 & n_2 items.

Mean: $\overline{x_1}$ & $\overline{x_2}$

$$\boxed{\bar{x} = \frac{n_1\overline{x_1} + n_2\overline{x_2}}{n_1 + n_2}}$$

Standard deviation: σ_1 & σ_2

$$\boxed{\sigma^2 = \frac{1}{n_1 + n_2}[n_1({\sigma_1}^2 + d_1^2) + n_2({\sigma_2}^2 + d_2^2)]}$$

Where $d_1 = \bar{x} - \overline{x_1}$, $d_2 = \bar{x} - \overline{x_2}$

Coefficient of variation

$$\boxed{C.V. = \frac{\sigma}{\overline{X}} \times 100}$$

Return To Top

Co-efficient of variation of two series are 75 % and 90 % and their standard deviations 15 and 18. Find their mean.

Solution:

For 1^{st} series $75 = \frac{15}{\bar{x}} \times 100 \quad \Rightarrow \bar{x} = 20$

For 2^{nd} series $90 = \frac{18}{\bar{x}} \times 100 \quad \Rightarrow \bar{x} = 20$

Thus, both the series have same mean i.e. 20

Return To Top

	Size	Mean	Variance
Observation I	10	2	2
Observation II	n	3	1

If the variance of the combined set of these two observation is $\frac{17}{9}$, then the value of n is equal to:

Solution:

For Observation I:

$$\frac{\sum x_i}{10} = 2 \Rightarrow \sum x_i = 20$$

$$\Rightarrow \frac{\sum x_i^2}{10} - (2)^2 = 2$$

$$\Rightarrow \sum x_i^2 = 60$$

For Observation II:

$$\frac{\sum y_i^2}{n} - 3^2 = 1 \Rightarrow \sum y_i^2 = 10n$$

Return To Top

	Size	Mean	Variance
Observation I	10	2	2
Observation II	n	3	1

If the variance of the combined set of these two observation is $\frac{17}{9}$, then the value of n is equal to:

Solution:

$$\sigma^2 = \frac{\sum(x_i^2 + y_i^2)}{10+n} - \left(\frac{\sum(x_i + y_i)}{10+n}\right)^2$$

$$\sum x_i^2 = 60$$

$$\sum y_i^2 = 10n$$

$$\Rightarrow \frac{17}{9} = \frac{60+10n}{10+n} - \frac{(20+3n)^2}{(10+n)^2}$$

$$\Rightarrow 17(n^2 + 20n + 100) = 9(n^2 + 40n + 200)$$

$$\Rightarrow 8n^2 - 20n - 100 = 0$$

$$\Rightarrow 2n^2 - 5n - 25 = 0$$

$$\Rightarrow n = 5$$

Return To Top

Thank You

Return To Top