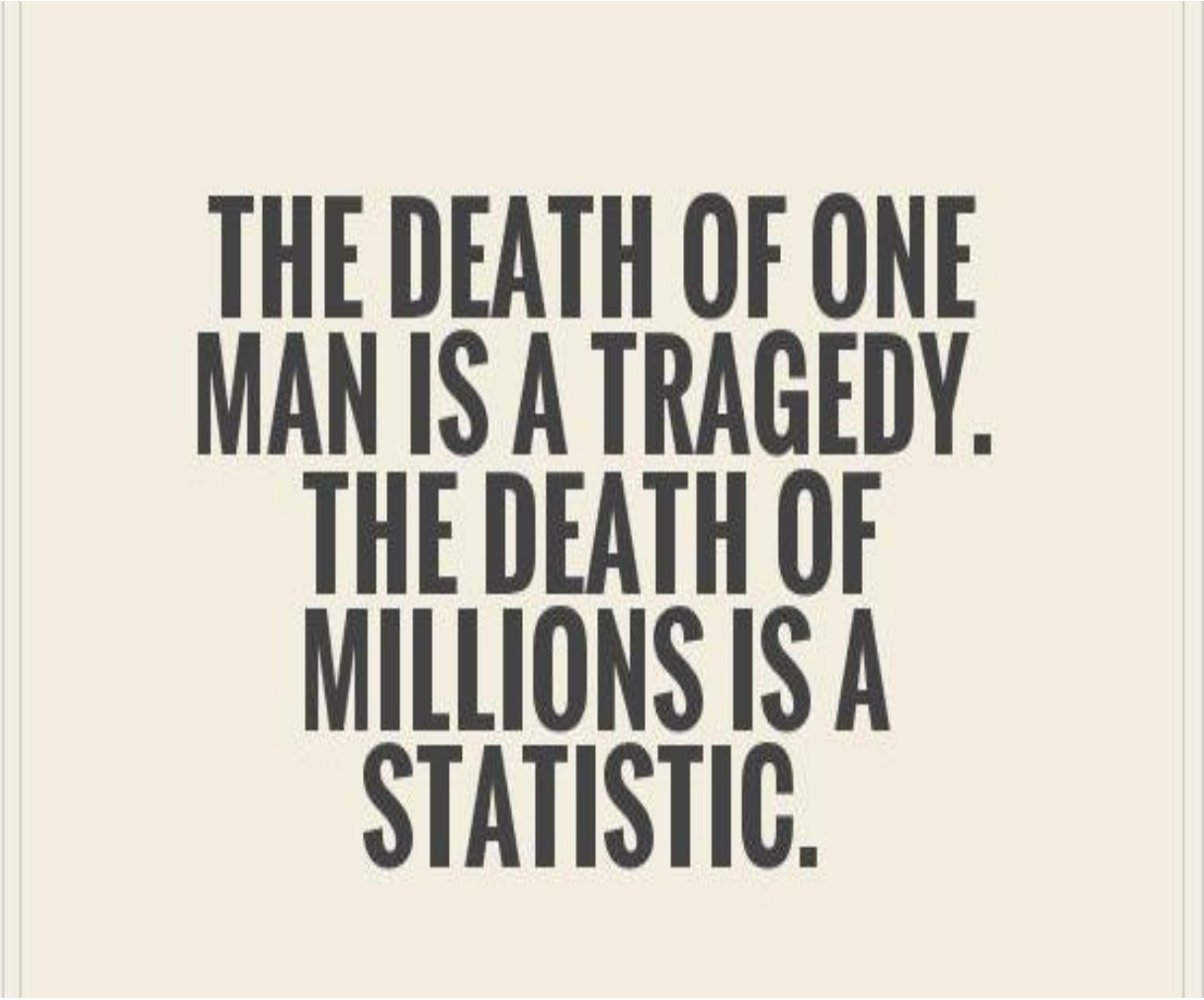
STATISTICS

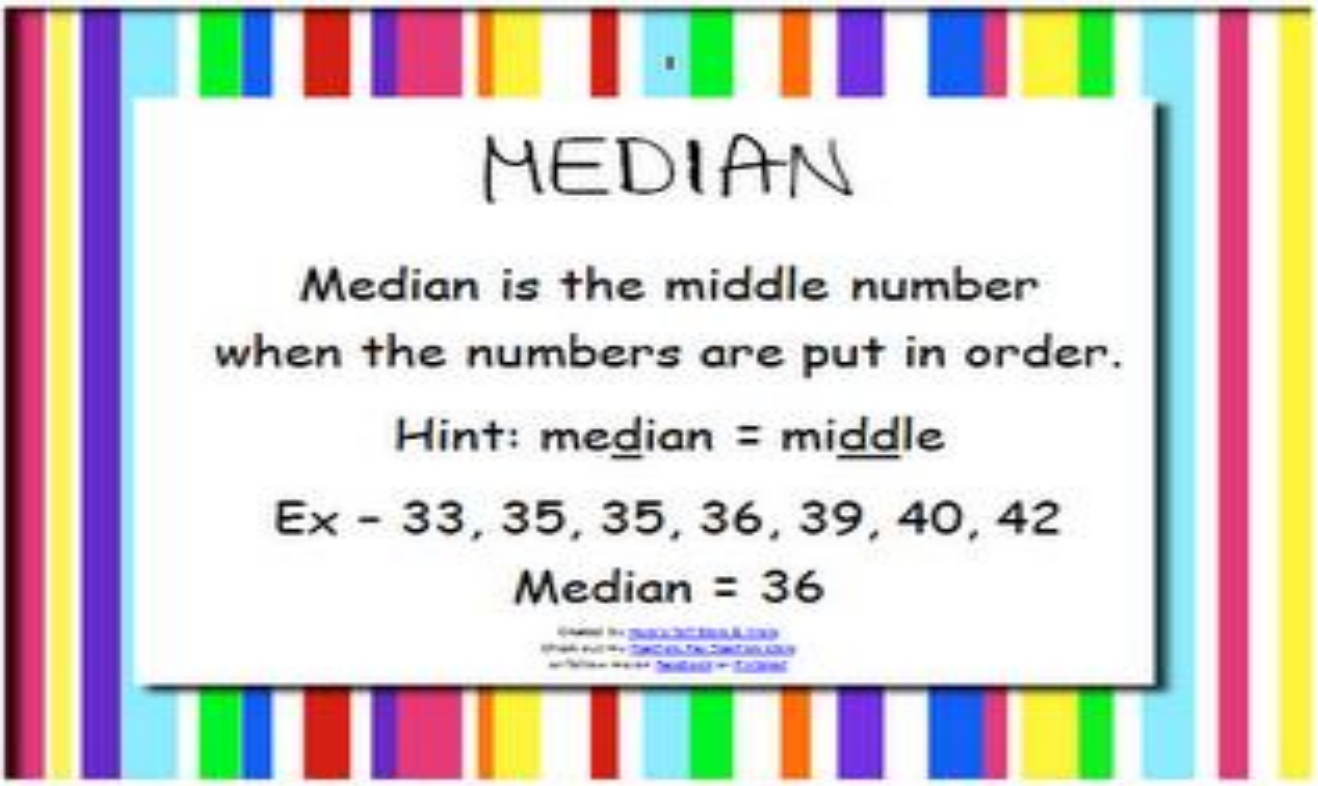

THE DEATH OF ONE MAN IS A TRAGEDY. THE DEATH OF MILLIONS IS A STATISTIC.

Mean Formula

$$\text{Mean} = \frac{\text{Sum of All Data Points}}{\text{Number of Data Points}}$$

$$\text{Mean} = \text{Assumed Mean} + \frac{\text{Sum of All Deviations}}{\text{Number of Data Points}}$$

Ex: Find the mean : { 100, 96,95,88,92,89, 91,93,85,87}



Median

First, arrange the observations in an ascending order.

If the number of observations (n) is odd: the median is the value at position

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

If the number of observations (n) is even:

1. Find the value at position $\left(\frac{n}{2}\right)$
2. Find the value at position $\left(\frac{n+1}{2}\right)$
3. Find the average of the two values to get the median.

Weight (kg) x	Frequency f	Cumulative Frequency $c.f$
10	4	4
20	7	11
30	12	23
40	15	38
50	13	51
60	5	56
70	4	60
Total	N = 60	

$$N = \sum f = 60$$

$$\frac{N+1}{2} = 30.5$$

Que1: The average of five positive integers is less than 20. What is the smallest possible median of this set?

(A) 19 (B) 10 (C) 4 (D) 3 (E) 1

Que2: Find the minimum value of a if median is 22.

Xi	7	12	19	22	30	35
Fi	8	13	9	a	7	6

Que3: A = (x,8,10,25,50) median of list A is 25

Quantity A	Quantity B
x	25

Mode

- The most frequently occurring number in a set of scores
- Ex: 60, 76, 82, 82, 93

- There can be more than one mode
- Ex: 60, 76, 76, 82, 93, 97, 97, 98,100

Relationship between Mean, Median and Mode

Mode = 3 Median – 2 Mean

Note: The above formula is not universal.

Ex:- { 1,2,3,4,5 }

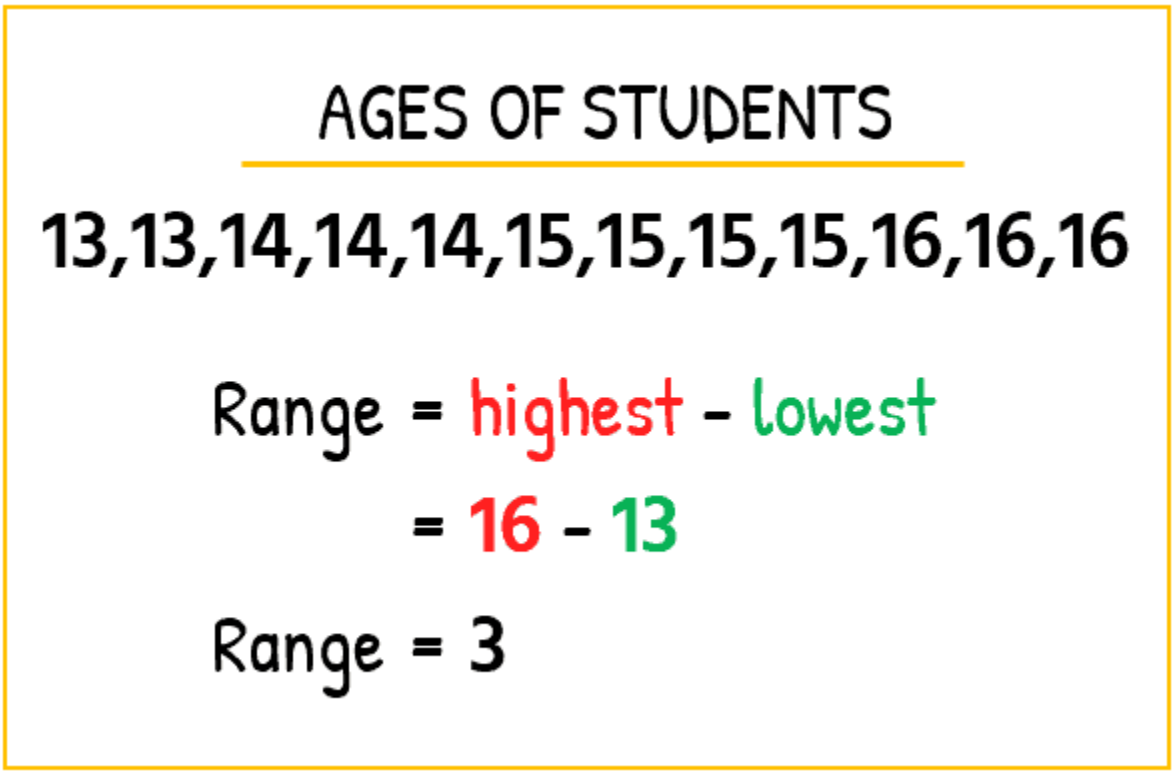
Ex:- { 3,3,3,3,3 }

Que4: Four positive integers have a mode of 4 and median of 3. What is their sum?

Que5: The greatest of the 41 positive integers in a certain list is 52. The Median of the 41 integers is 21. What is the least possible average of the 41 integers?

(A) 8 (B) 10 (C) 12 (D) 20 (E) 21

Range



Que6: . List A consist only prime numbers such that their range is even and it has at least 5 distinct prime numbers.

Quantity A	Quantity B
Any prime number in the list A	2

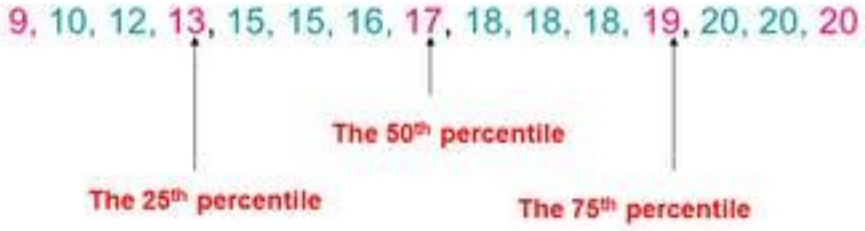
Que7: Quantity A	Quantity B
Range of first 100 positive multiples of 11	Range of 100 consecutive multiples of 11

Percentiles

- In statistics, a percentile is the value of a variable below which a certain percent of observations fall. For example, the 20th percentile is the value (or score) below which 20 percent of the observations may be found.
- The term percentile and the related term percentile rank are often used in the reporting of scores from norm-referenced tests. For example, if a score is in the 86th percentile, it is higher than 85% of the other scores.
- The median of a data set is also known as the 50^{th} percentile.

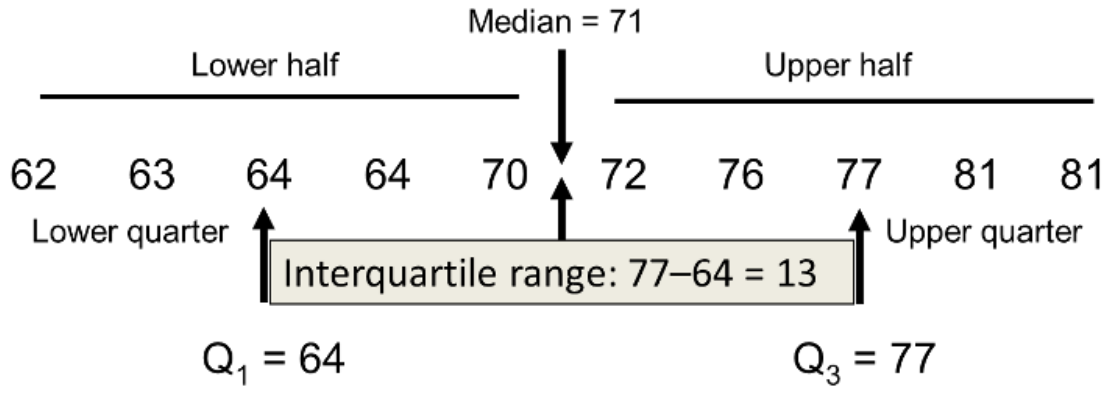
The quartiles divide the data set into 4 groups.

Each group contains approximately 25% of the data values, so the following terminology is also used.

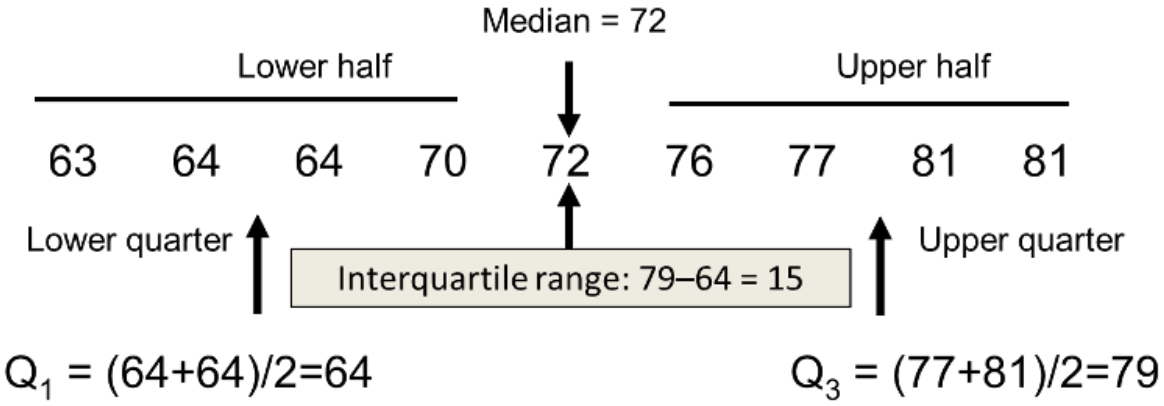


Interquartile Range

Case 1: When n is even

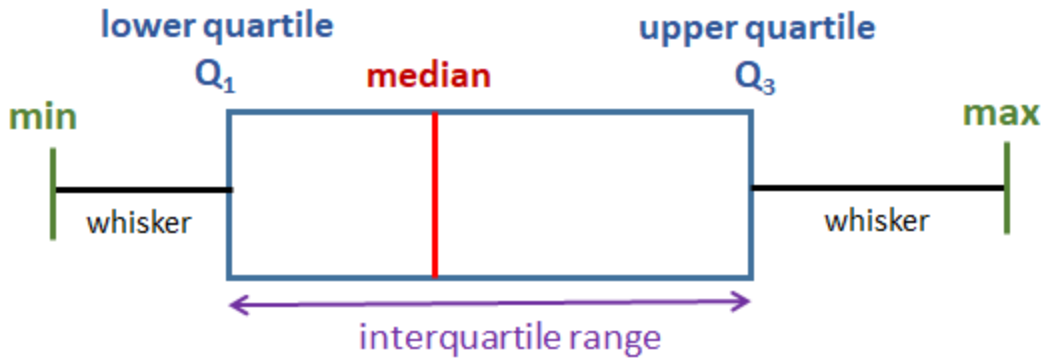


Case 2: When n is odd

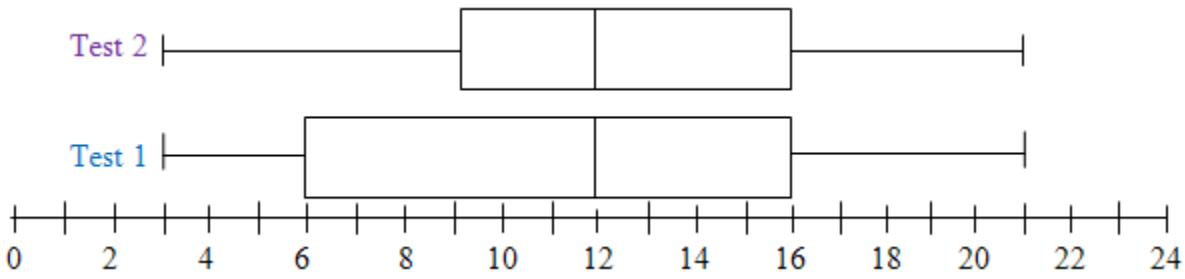


Box and Whisker Plot

A box and whisker plot (also called a box plot) shows the five-number summary of a set of data: **minimum**, **lower quartile**, **median**, **upper quartile**, and **maximum**.



Ex:- Here, there is some information we can conclude from the given plot about the students performance in these tests.

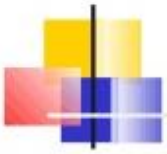


Standard Deviation and Variance

Definition

Standard Deviation shows the variation in data. If the data is close together, the standard deviation will be small. If the data is spread out, the standard deviation will be large.

Standard Deviation is often denoted by the lowercase Greek letter sigma, σ.



Variance

Variance is the average squared deviation from the mean of a set of data. It is used to find the **standard deviation**.

Formulas :

$$2,\ 4,\ 4,\ 4,\ 5,\ 5,\ 7,\ 9.$$

These eight data points have the mean (average) of 5:

$$\frac{2+4+4+4+5+5+7+9}{8} = 5.$$

First, calculate the deviations of each data point from the mean, and square the result of each:

$$\begin{aligned}
(2-5)^2 &= (-3)^2 = 9 \qquad & (5-5)^2 &= 0^2 = 0 \\
(4-5)^2 &= (-1)^2 = 1 \qquad & (5-5)^2 &= 0^2 = 0 \\
(4-5)^2 &= (-1)^2 = 1 \qquad & (7-5)^2 &= 2^2 = 4 \\
(4-5)^2 &= (-1)^2 = 1 \qquad & (9-5)^2 &= 4^2 = 16.
\end{aligned}$$

The variance is the mean of these values:

$$\frac{9+1+1+1+0+0+4+16}{8} = 4.$$

and the *population* standard deviation is equal to the square root of the variance:

$$\sqrt{4} = 2.$$

Que8: Quantity A Quantity B

SD of 1st 10 prime numbers SD of 1st 10 odd natural numbers

Que9: Quantity A Quantity B

SD of (24, 34, 44, 54, 64) SD of (89, 79, 69, 59, 49)

Que10: Quantity A Quantity B

SD of (10,20,50,80,90) SD of (10,30,50,70,90)

Change in standard deviation when an observation is added or removed

Ex:- List A = { 10,20,30,40,50}

Case1. 20 is removed

Case2. 50 is removed

Case3. 30 is added

Case4. 30 is removed

Que11: . M = {n, n+1, n+2, n+3, n+4} where n is greater than 2

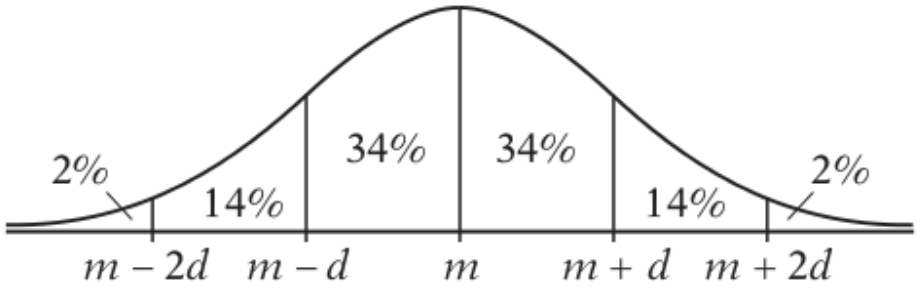
Quantity A	Quantity B
Standard deviation of the numbers in set M	n

Normal Distribution

Definition and Properties of a Normal Curve

A **normal curve** is a symmetric, bell-shaped curve. Any random variable whose graph has this characteristic shape is said to have a **normal distribution**.

On a normal curve, if the quantity shown on the horizontal axis is the number of standard deviations from the mean, rather than values of the random variable itself, then we call the curve the **standard normal curve**.



Que12: AIIMS conducted a survey to determine how many apples are eaten by 100,000 people during last year. Number of apples eaten by people is normally distributed with a mean of 29 apples and deviation of 4 apples. Approx. how many of the surveyed people ate more than 25 apples in the last year?

(A) 16000 (B) 68000 (C) 84000 (D) 98000

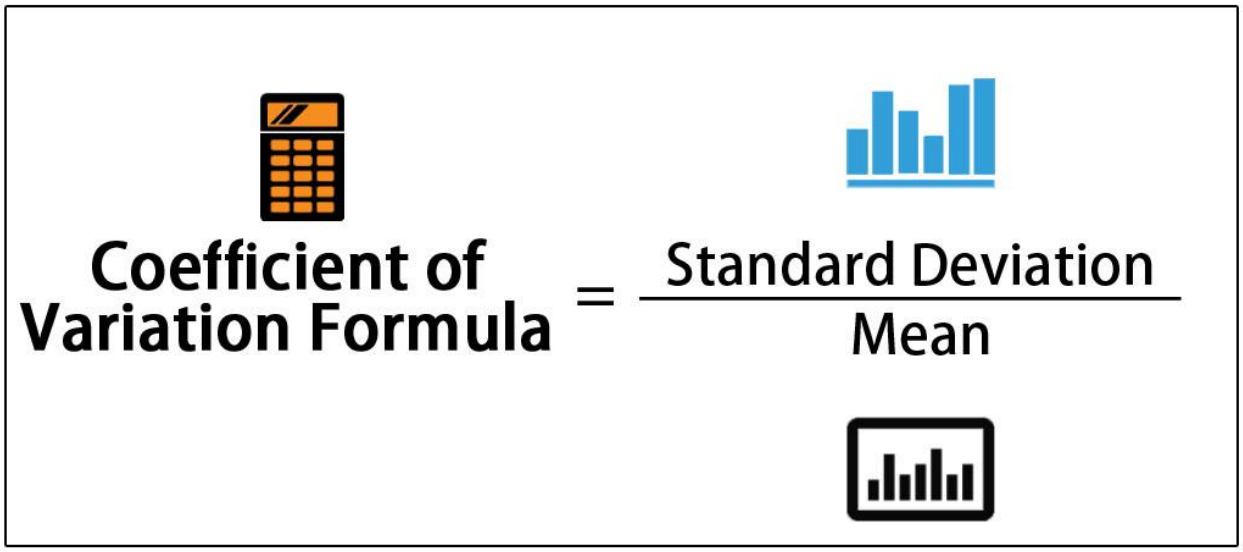
Que13: Standard deviation on a test was 12 points and the mean was 70 points. If the scores are normally distributed and student X scored 95 points then student X scored higher than approx. what % of the students?

(A) 84% (B) 98% (C) 68% (D) 50%

Que14: In a normally distributed set of data, the mean is 12 and the standard deviation is less than 3

Quantity A	Quantity B
Number of data points in the set located between 9 & 15	68% of the total data points

Coefficient of Variation

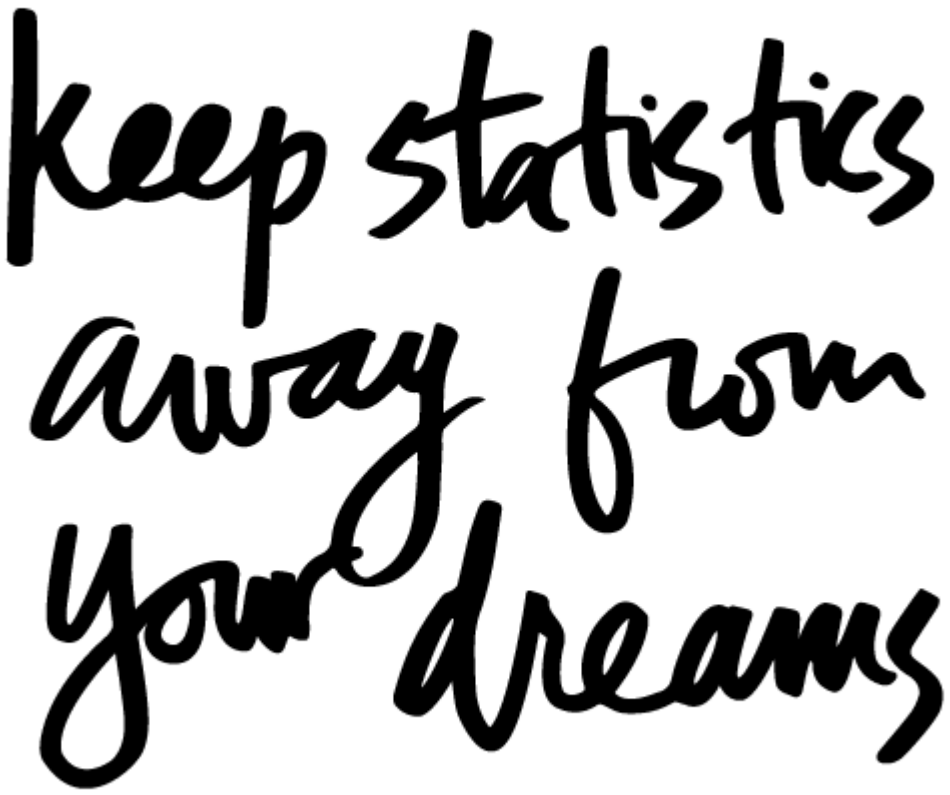


Que15: Mean and variance of 2 series are given below:

Series A: Mean = 54 and variance = 9

Series B: Mean = 100 and variance = 4 . Which series is more stable?

(A)Series A (B) Series B (C)Both are equally stable (D) cannot be determined


keep statistics
away from
your dreams

Statistics

1.] avg. of five +ve integers is less than 20.

The smallest possible median set will be 1.

1 1 1 1 1 , this can be possible.

2.] here median is 22.

total = 13+9+8 = 30 observation → 30 observation

22 observation

but here there is 13 observation only.

which is compensated by a.

now a = [18] Ans.

30 obs. [a] 30 obs.

↓ value

one has to be there

3.]

8 10 25 x 50

8 10 25 50 x

so [x ≥ 25] so Ans. [D]

4) Mode → 4 means it will be at least 2 times.

median = 3 means sum of 2nd and 3rd obs = 6

1 2 4 4

↳ this can't be 2, because then Mode will be 2 and 4 both.

5) Total 41 +ve integers.

Greatest is 52. Median = 21

20 obs. | 21 | 20 obs. | 52 |

↓ 20 obs. ↓ 19 obs.

these will all be 1, 1, 1 ... 20 times.

the remaining 19 obs. will be 21

So → $21 \times 20 + 20 + 52$

$= 420 + 72 = \frac{492}{41} = 12$

Avg = 12

6.] Range is even

Range = Max − Min

= Odd − Odd = even

E − E = even (this is not possible as all the nos are prime)

And only even prime no. is 2.

So in this list min prime no. will be 3.

So. Ans. [A]

7.] Take any example and solve it.

Ans. [C]

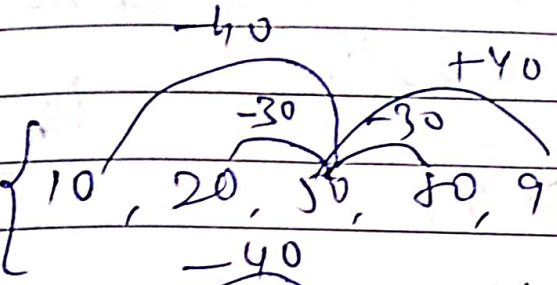
8.] First 10 prime nos will be more scattered than 1st 10 odd natural numbers.

So Ans. [A]

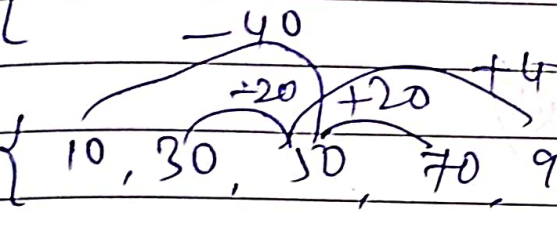
9.] Std. deviation depends on the deviation from the mean value, not on the observations.

So Ans. [C]

10) By solving we can see that Std. deviation of A > Std. deviation of B.



(A)



(B)

We can see that the deviation is more in case A.

So Ans. [A]

11) Maximum deviation in this set from mean is $(n+4)-(n+2) = \boxed{2}$

So Std. deviation < 2.

Ans. [B]

12) Ate more than 25 apples.

$m - d = 29 - 4 = \underline{25}$

more than [m-d] is [84%]

So. Ans - [B]

13.) We know that

$95 > (m+2d)$ ⟹ below $(m+2d)$ is 98%.

$95 > 70 + 2(12)$

$95 > 94$

So. Ans. [B]

14.) Ans. [A]

Let suppose std. deviation = 2.9

So $m-d = 12 - 2.9 = 9.1$

$m+d = 12 + 2.9 = 14.9$

So between 9.1 — 14.9 68% of data will be laid.

Now, there is more region going to be included in [9 - 15].

15)

$$\text{Cov } A = \frac{8\sigma}{|\bar{x}|} = \frac{8}{54} = \frac{1}{18}$$

$$\text{Cov } B = \frac{\sigma}{|\bar{x}|} = \frac{2}{100} = \frac{1}{50}$$

Cov A > Cov B So Ans. [B]