CHAPTER - 21
STATISTICS

The definitions given by famous statisticians are quoted below

Sir R.A. Fisher

" The science of statistics is essentially a branch of applied mathematics and may be regarded as mathematics applied to observational data".

Lovitt

"Statistics is the science which deals with the collecting, classifying, presenting, comparing and interpreting numerical data collected to throw light on any sphere of enquiry"

Averages – Measures of central tendency

Averages may be defined as a representative value and it stands for all the items in the population. The following are important averages

1. Arithmetic mean - (A.M)
2. Geometric mean - (G.M)
3. Harmonic mean - (H.M)
4. mode
5. median

Arithmetic mean A.M

$\bar{x} = \frac{\Sigma x_i}{n}$; for a raw data

$\bar{x} = \frac{\Sigma f_i x_i}{N}$; for a frequency distribution, where $N = \Sigma f_i$

Properties of A.M

1. The algebraic sum of deviations of a set of observations from their mean is zero. i.e $\sum\left(x_i-\bar{x}\right)=0$

2. The sum of squares of deviations from the A.M is the minimum. i.e $\Sigma\left(x_i-\bar{x}\right)^2$ is the minimum

3. The A.M is dependent of change of scale and origin

   AM (x_i) = $\bar{x}$

   A.M ($x_i \pm k$) = $\bar{x} \pm k$ (change of origin)

   A.M ($k\, x_i$) = $\bar{x}\,k$ (change of scale)

   $\text{A.M}\left(\frac{ax+b}{c}\right)=\frac{a\bar{x}+b}{c}$ (change of scale and origin)

4. A.M of an A.P $=\frac{\text{first term}+\text{last term}}{2}$

5. **Combined A.M**

   If $\bar{x}_1, \bar{x}_2 ----\bar{x}_k$ are means of k samples of size n_1, n_2 -----n_k then combined A.M is given by

   $$\bar{x}=\frac{\sum_{i=1}^{k} n_i\bar{x}_i}{\sum_{i=1}^{k} n_i}$$

6. **Weighted A.M**

   In statistics, weight means the importance given to be observations.

   Let x_1, x_2 ------x_n be the observations with corresponding weights w_1, w_2 , -------w_n , then weighted A.M is given by,

   $$\bar{X}_w=\frac{\Sigma w_i x_i}{\Sigma w_i}$$

Geometric mean (G.M)

G.M of x_1, x_2, x_3 ....x_n= n^{th} root of product

$$GM = \left(x_1 . x_2 . x_3 ....x_n\right)^{1/n}$$
$$= \left(\Pi x_i\right)^{1/n}$$

$$\log \ G.M = \frac{1}{n} \log(\Pi x_i) \quad = \frac{1}{n} \log\left(x_1 . \ x_2 ... x_n\right) \quad = \frac{1}{n} \Sigma \log \ x_i$$

$$\log G.M = \frac{\Sigma \log x_i}{n}$$

Result : log of G. M= A. M of log of observations

Result : G. M of a $GP = \sqrt{\text{first term} \times \text{ last term}}$

Harmonic mean - H . M

x_1, x_2, x_3 ------- x_n – observations

$$\frac{1}{x_1}, \frac{1}{x_2} ------ \frac{1}{x_n} \quad \text{reciprocals}$$

$$\frac{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + ------ \frac{1}{x_n}}{n} - A.M$$

Reciprocal of A.M of the reciprocals of observations

$$H.M = \frac{n}{\sum\left(\frac{1}{x_i}\right)}; \quad H.M \text{ of } a, b = \frac{2}{\frac{1}{a} + \frac{1}{b}} \quad = \frac{2ab}{a+b} \quad \text{H. M of a, b, c} = \frac{3abc}{ab + bc + ac}$$

Use of H.M

To find the average speed if equal distances are covered with different speeds

Mode

The value of the variable which occurs most frequently in a distribution is mode

4, 7, 9, 7, 8, 9, 6 $\rightarrow$ modes = 7 & 9 – bimodal

More than one mode $\equiv$ multi - modal

Median

It is the middle most item when the set is in ascending order $\text{median} = \frac{(n+1)^{th}}{2}$ item

Mean Deviation

Let $x_1, x_2 \dots x_n$ be n - observations. Let A be any one of the averages. Then the mean deviation about A is given by

M.D = $\frac{\sum |x_i - A|}{n}$, For a frequency distribution, $\text{M.D} = \frac{\sum f_i |x_i - A|}{\sum f_i}$

Absolute deviations are minimum when the average becomes the most central one. But we know that median is the most central value. Hence M.D is minimum when deviations are taken from median

Root mean square deviation (RMSD)

It is the square root of the A.M of the squares of deviations of observation from an average. A

$$\text{RMSD} = \sqrt{\frac{\Sigma (x_i - A)^2}{n}}$$

Note: The minimum value of R.M.S.D is S.D

The R.M.S.D from the A.M is called the standard deviation

ie if A = $\bar{x}$, R.M.S.D is S.D

For a raw data, standard deviation can be calculated by

$$\sigma = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}} \quad = \sqrt{\frac{\Sigma x_i^2}{n} - (\bar{x})^2}$$

In a freq. distribution, $\sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{N}}$

Note 1: S.D is independent of change of origin

Note 2: S.D is dependent of change of scale

Variance : It is the square of S.D **Variance =s²**

Coefficient of variance (C.V)

$$\text{C.V} = \frac{\text{S.D}}{\text{A.M}} \times 100 = \frac{\sigma}{\bar{x}} \times 100$$

The C.V is used to compare the variability among different sets of observations. The set having the least C.V is said to be the most consistent set.

Combined variance

$$\sigma^2 = \frac{n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)}{n_1 + n_2} \qquad d_1 = \bar{x}_1 - \bar{x} \qquad d_2 = \bar{x}_2 - \bar{x}$$

PART I - (JEEMAIN)

1. The mean age of 25 teachers in a school is 40 years. A teacher retires at the age of 60 years and a new teacher is appointed in his place. If the mean age of the teachers in this school now is 39 years, then the age (in years) of the newly appointed teacher is

   1) 35 2) 36 3) 37 4) 38

2. The mean value of the mean and median of the odd divisor of 360 is

   1) 12.2 2) 7 3) 13 4) 10

3. Let the median and the mean deviation about the median of 7 observation 170, 125, 230, 190, 210, a,b be 170 and $\frac{205}{7}$ respectively. Then the mean deviation about the mean of these 7 observations is

   1) 31 2) 28 3) 30 4) 32

4. Consider the following frequency distribution

class	10-20	20-30	30-40	40-50	50-60
Frequency	α	110	54	30	β

   If the sum of all frequencies is 584 and median is 45, then $|\alpha-\beta|$ is equal to

   1) 164 2) 184 3) 204 4) 144

5. The mean square deviation about –2 and +2 of a set of observations are 18 and 10 respectively. The S.D of the set is

   1) 3 2) 2 3) 1 4) 0

6. If the data $x_1, x_2, ..., x_{10}$ is such that the mean of first four of these is 11, the mean of the remaining six is 16 and the sum of squares of all of these is 2,000, then the standard deviation of this data is

   1) 4 2) 2 3) $\sqrt{2}$ 4) $2\sqrt{2}$

7. If the mean of the numbers a, b 8, 5, 10 is 6 and their variance is 6.8 then ab is equal to

   1) 6 2) 7 3) 12 4) 14

8. Let $\bar{x}, M$ and σ^2 be respectively, the mean, mode and variance of n observations $x_1, x_2, ...x_n$ and $d_i = -x_i - a, i = 1, 2, ...n$ where a is any number

   Statement I: Variance of $d_1, d_2, ...d_n$ is σ^2

   Statement II: Mean and mode of $d_1, d_2....d_n$ are $\bar{x} - a$ and M-a respectively

   1) Statement I and statement II are both true
   2) Statement I and statement II are both false
   3) Statement I is true and statement II is false
   4) Statement I is false and statement II is true

9. The outcome of each 30 items was observed. 10 items gave an outcome $\frac{1}{2}$-d each, 10 items gave outcome $\frac{1}{2}$ each and the remaining 10 items gave outcome $\frac{1}{2}+d$ each. If the variance of this outcome data is $\frac{4}{3}$ then $|d|$ equals

1) 2 2) $\frac{\sqrt{5}}{2}$ 3) $\frac{2}{3}$ 4) $\sqrt{2}$

10. Let 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

1) 425 2) 650 3) 250 4) 925

11. The mean and variance of the marks obtained by the students in a test are 10 and 4 respectively. Later the marks of one of the students is increased from 8 to 12. If the new mean of the marks is 10.2, then their new variance is equal to

1) 4.04 2) 4.08 3) 3.96 4) 3.92

12. Let the six numbers a_1, a_2, a_3, a_4, a_5, a_6 be in AP and $a_1 + a_3 = 10$. If the mean of these six numbers is $\frac{19}{2}$ and their variance is σ^2; then $8\sigma^2$ is equal to

1) 220 2) 210 3) 200 4) 105

13. The mean and the standard deviation (s.d) of 10 observations are 20 and 2 respectively. Each of these 10 observations is multiplied by p and then reduced by q, where $p \neq 0$ and $q \neq 0$. If the new mean and new standard deviation become half of their original values, then $|q|$ is

1) 10 2) 20 3) 30 4) 40

14. Let the observations $x_i \left(1 \le i \le 10\right)$ satisfy the equations $\sum_{i=1}^{10}[x_i - 5] = 10$ and $\sum_{i=1}^{10}[x_i - 5]^2 = 40$. If μ and λ are the mean and the variance of the observations, $x_1 - 3, x_2 - 3, ..., x_{10} - 3$ then the ordered pair (μ, λ) is equal to

1)(3,3) 2) (6,6) 3) (6,3) 4) (3,6)

15. The mean and standard deviation of 15 observations were found to be 12 and 3 respectively. On rechecking it was found that an observation was read as 10 in place of 12. If μ and σ^2 denote the mean and variance of the correct observations respectively, then $15\left(\mu + \mu^2 + \sigma^2\right)$ is equal to.....

1) 2521 2) 3681 3) 7011 4) 8169

16. For two data sets, each of size 5, the variances are given to be 4 and 5 and the corresponding means are given to be 2 and 4 respectively. The variance of the combined data set is

 1) 5.5 2) 6.7 3) 3.1 4) 1.9

17. The first of the two samples in a group has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and standard deviation $\sqrt{13.44}$, then the standard deviation of the second sample is

 1) 8 2) 6 3) 4 4) 5

18. The sum of deviations of ten observations about 50 is zero and the sum of squares of deviations of observations about 50 is 250. The coefficient of variance is

 1) 30 2) 60 3) 10 4) 20

19. A data consists of n observations: $x_1, x_2, ..., x_n$. If $\sum_{i=1}^{n}(x_i+1)^2 = 9n$ and $\sum_{i=1}^{n}(x_i-1)^2 = 5n$, then the standard deviation of this data is

 1) 5 2) $\sqrt{5}$ 3) $\sqrt{7}$ 4) 2

20. The mean and variance of 5 observations are 5 and 8 respectively. If 3 observations are 1,3,5 then the sum of cubes of the remaining two observations is

 1) 1072 2) 1082 3) 1092 4) 1002

SECTION - II

Numerical Type Questions

21. If both the mean and the standard deviation of 50 observations $x_1, x_2 ....., x_{50}$ are equal to 16, then the mean of $(x_1-4)^2, (x_2-4)^2, ...(x_{50}-4)^2$ is

22. Consider the statistics of two sets of observations as follows

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

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

23. The mean and variance of 7 observations are 8 and 16 respectively. If one observation 14 is omitted then a and b are respectively the mean and variance of remaining 6 observations, then $a+3b-5$ is equal to

24. Let the mean and standard deviation of marks of class A of 100 students be respectively 40 and $\alpha(>0)$, and the mean and standard deviation of marks of class B of n students be respectively 55 and $30-\alpha$. If the mean and variance of the marks of the combined class of 100+n students are respectively 50 and 350; then the sum of variances of classes A and B is

25. Let $x_1, x_2, ...x_{18}$ be eighteen observations such that $\sum_{i=1}^{18}(x_1-\alpha)=36$ and $\sum_{i=1}^{18}(x_1-\beta)^2=90$,

where α and β are distinct real numbers. If the standard deviation of these observations is 1, then the value of $|\alpha-\beta|$ is

PART - II (JEE ADVANCED)

SECTION - III (Only one option correct type)

26. Consider the data 12,3,18,17,4,9,17,19,20,15,8,17,2,3,16,11,3,1,0,5. Find the number of items between $\bar{x}-MD_{(\bar{x})}$ and $\bar{x}+MD_{(\bar{x})}$

A) 6 B) 7 C) 8 D) 9

27. Consider the following frequency distribution

Class :	0–6	6–12	12-18	18–24	24–30
Frequency	a	b	12	9	5

If mean = $\frac{309}{22}$ and median = 14, then the value $(a-b)^2$ is equal to

A) 8 B) 6 C) 4 D) 10

28. Let the mean and variance of the frequency distribution

x :	$x_1=2$	$x_2=6$	$x_3=8$	$x_4=9$
f :	4	4	α	β

be 6 and 6.8 respectively. If x_3 is changed from 8 to 7, then the mean for the new data will be

A) 4 B) 5 C) $\frac{17}{3}$ D) $\frac{16}{3}$

29. Let $a_1, a_2, ...a_{10}$ be 10 observations such that $\sum_{k=1}^{10}a_k=50$ $\sum_{\forall k<j}a_k.a_j=1100$. Then the standard deviation of $a_1, a_2,....a_{10}$ is equal to

A) 5 B) $\sqrt{5}$ C) 10 D) $\sqrt{115}$

30. Consider 10 observation $x_1, x_2, \ldots\ldots x_{10}$ such that $\sum_{i=1}^{10}(x_i-\alpha)=2$ and $\sum_{i=1}^{10}(x_i-\beta)^2=40$, where α,β are positive integers. Let the mean and the variance of the observations be $\frac{6}{5}$ and $\frac{84}{25}$ respectively. The $\frac{\beta}{\alpha}$ is equal to

A) 2 B) $\frac{3}{2}$ C) $\frac{5}{2}$ D) 1

31. If a variable takes values 0,1,2,.....n with frequencies $q^n, \frac{n}{1}q^{n-1}p, \frac{n(n-1)}{1.2}q^{n-2}p^2, \ldots, p^n$, where $p+q=1$, then the mean is

A) np B) nq C) n (p+q) D) none of these

32. If $\sum_{i=1}^{n}(x_i-a)=n$ and $\sum_{i=1}^{n}(x_i-a)^2=na,(n,a>1)$ then the standard deviation of 'n' observations $x_1, x_2, \ldots, x_n$ is

A) $n\sqrt{a-1}$ B) $\sqrt{a-1}$ C) $a-1$ D) $\sqrt{n(a-1)}$

33. If the mean and the standard deviation of the data 3,5,7,a,b are 5 and 2 respectively, then 'a' and 'b' are the roots of the equation

A) $2x^2-20x+19=0$ B) $x^2-10x+19=0$

C) $x^2-10x+18=0$ D) $x^2-20x+18=0$

34. Let $X=\{x \in N : 1 \le x \le 17\}$ and $Y=\{ax+b : x \in X \text{ and } a,b \in R, a>0\}$. If mean and variance of elements of Y are 17 and 216 respectively, then $|a+b|$ is equal to

A) 5 B) 6 C) 7 D) 8

35. Consider the data of 'x' taking the values 0,2,4,8,....,2^n with frequencies ${}^nC_0, {}^nC_1, {}^nC_2, \ldots, {}^nC_n$ respectively. If the mean of this data is $\frac{728}{2^n}$, then 'n' is equal to......

A) 6 B) 3 C) 1 D) 8