ID- 14080 | MSMC | UROOJ AFRIDI |

Question: 1

Find averages (A.M, G.M, H.M) of the following tables also justify their logical relationships.

a.

No. of Children/family	No. of families
1	4
2	13
3	9
4	4
5	1

i) ARITHEMETIC MEAN

No. of Children Per family (x)	No. of families (f)	fx
1	4	4
2	13	26
3	9	27
4	4	16
5	1	5
	$\Sigma f = 31$	$\Sigma fx = 78$

$$AM = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{78}{31}$$

$$= \boxed{2.51}$$

Scanned with CamScanner

a) GEOMETRIC MEAN

S.NO	x	$\log x$
1	4	$\log 4 = 0.6020$
2	13	$\log 13 = 1.1139$
3	9	$\log 9 = 0.9542$
4	4	$\log 4 = 0.6020$
5	1	$\log 1 = 0$
		$\sum \log x = 3.2721$

$$G.M = \text{Anti-log} \frac{\sum \log x}{n}$$

$$= \text{Anti log} \frac{3.2721}{5}$$

$$= \text{Antilog } 0.6544$$

$$G.M = \boxed{4.512}$$

a) HARMONIC MEAN

S.NO	x	$1/x$
1	4	$1/4 = 0.25$
2	13	$1/13 = 0.076$
3	9	$1/9 = 0.111$
4	4	$1/4 = 0.25$
5	1	$1/1 = 1$
		$\sum 1/x = 1.687$

$$H.M = \frac{n}{\sum (1/x)}$$

$$= \frac{5}{1.681} = \boxed{2.964}$$

Scanned with CamScanner

Q1. Part (.b)

Marks	Frequency
0 – 9	2
10 – 19	31
20 – 29	73
30 – 39	85
40 – 49	28.

b) ARITHEMATIC MEAN

Marks	Frequency	Mid-value (x)	fx.
0–9	2	4.5	9
10–19	31	14.5	449.5
20–29	73	24.5	1788.5
30–39	85	34.5	2932.5
40–49	28	44.5	1246
	$\Sigma f = 219$		$\Sigma fx = 6425.5$

$$A \cdot M = \frac{\Sigma fx}{\Sigma f}.$$

$$A \cdot M = \frac{6425.5}{219} = \boxed{29.3}$$

b) GEOMETRIC MEAN

Marks	Frequency (f)	x	log x	f log x
0–9	2	4.5	0.6532	2.9394
10–19	31.	14.5	1.1613	16.8388
20–29	73	24.5	1.3891	34.0329
30–39	85	34.5	1.5378	53.0541
40–49	28	44.5	1.6483	73.3493
	$\Sigma f = 219$			$\Sigma f \log x =$ 180.2145

$$G.M = \text{Anti-log } \frac{\sum f \log x}{\sum f}$$

$$= \text{Anti-log } \frac{180.2145}{219}$$

$$= \text{Anti-log } 0.8228$$

$$= \boxed{6.649}$$

b) HARMONIC MEAN:

Marks	frequence	x	f/x
0-9	2	4.5	0.4444
10-19	31	14.5	2.1379
20-29	73	24.5	2.9795
30-39	85	34.5	2.4637
40-49	28	44.5	0.6292
	$\sum f = 219$		$\sum f/x = 8.6547$

$$H.M = \frac{\sum f}{\sum (f/x)}$$

$$H.M = \frac{219}{8.6547}$$

$$H.M = \boxed{25.30}$$

LOGICAL RELATIONSHIP:

a) $G.M \geq H.M \geq A.M$

$4.512 \geq 2.96 \geq 2.51$

b) $A.M \geq H.M \geq G.M$

$29.3 \geq 25.30 \geq 6.64$

Scanned with CamScanner

Question : 2.

Find Median & Mode of the following tables.

a.

No. of Children	No. of families
1	4
2	13
3	9
4	4
5	1.

Finding Median

No. of Children	No. of families
1	4
2	3
3	9
4	4
5	1

Arrange in Ascending order $\rightarrow$

S.no	x
1	1
2	4
3	4
4	9
5	13.

No. of item is 5 (odd).

So, median = Size of $\left(\frac{n+1}{2}\right)^{th}$ item

$= \left(\frac{5+1}{2}\right)^{th}$ item

$= \frac{6^{th}}{2}$ item $= 3^{rd}$ item

Median $= \boxed{4}$

Scanned with CamScanner

Mode (a).

No. of Children	No. of families
1	4
2	13
3	9
4	4
5	1

$\longrightarrow$ Ascending order.

S.no	x
1	1
2	4
3	4
4	9
5	13.

In the above data, 4 is more frequent value.
So here, Mode $\boxed{= 4}$

b)

Marks	Frequency.
0-9	2
10-19	31
20-29	73
30-39	85
40-49.	28

Finding Median.

Marks	Frequency	C-B	C.F
0-9	2	0-8.5	2
10-19	31	9.5-18.5	33
20-29	73	19.5-28.5	106
30-39	85	29.5-38.5	191
40-49.	28	39.5-48.5	219.
	$\Sigma f = 219.$		

Median = Size of $\left(\frac{n+1}{2}\right)$+2 item

~~DA~~ $= \frac{219+1}{2}^{th} = 110^{th}$ item

$$\text{Median} = l + \frac{h}{f}\left(\frac{n}{2} - CF\right)$$

where.

$l = 29.5$

$h = 9$

$f = 85$

$C = 106$.

Now putting values, we get.

$$= 29.5 + \frac{9}{85}(109.5 - 106)$$

$$= 29.5 + \frac{9}{85}(3.5)$$

$$= 29.5 + 0.3705$$

$$= \boxed{10.9323}.$$

⇒ Finding MODE:

The mode lies in the group 30-39, so by form

$$\text{Mode} = l + \frac{(fm - fo)}{2fm - fo - fi} \times h.$$

$l = 30$

$fm = 85$

$f_0 = 73$

$f_1 = 28$.

$h = 9$

Putting values, we get,

$$\text{Mode} = 30 + \frac{(85-73)}{2(85)-73-28} \times 9$$

$$= 30 + \left(\frac{12}{69}\right) \times 9.$$

$$= 30 + 1.5652$$

$$\text{Mode} \boxed{= 31.5652}.$$

Scanned with CamScanner

Question : 3.

i) Find semi-quartile range and Semi-Inter Quartile Range of Q2(a).

S.no	x
1	4
2	13
3	9
4	4
5	1

$Q_2 = (n+1) \times 0.5$

$= (5+1) \times 0.5$

$Q_2 = \boxed{4}$

$Q_1 = (n+1) \times 0.25$

$= (5+1) \times 0.25$

$= \boxed{1.5}$

$Q_3 = (n+1) \times 0.75$

$= (5+1) \times 0.75$

$= \boxed{4.5}$

Scanned with CamScanner

Semi Inter Quartile Range.

$$\frac{Q_3 - Q_1}{2}$$

$$= \frac{4.5 - 1.5}{2}$$

$$= 1.5.$$

Q3. (b).

S.n	x
1	4
2	13
3	9
4	4
5	1.

First we find AM.

$$AM = \bar{x} = \frac{4+13+9+4+1}{5}$$

$$= \boxed{6.2.}$$

x	$x - \bar{x}$	$(x-\bar{x})^2$.
6	0.2	0.04
8	1.8	3.24
10	3.8.	14.44
12	5.8.	33.64
14	7.8.	60.84
		$\sum = 112.2.$

Scanned with CamScanner

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

$$= \frac{112.2}{5}$$

$$= 22.44.$$

Standard Deviation.

$$S = \sqrt{\text{variance}}$$

$$= \sqrt{22.44}.$$

$$= 4.73.$$

Now Co-efficient of Variance

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

$$= \frac{4.73}{6.2} \times 100$$

$$\boxed{= 76.29.}$$

Scanned with CamScanner

Question : 4.
Write short notes on following.

RANGE :
The range R is defined as the difference between the largest and the smallest observations in a set of data.

Symbolically : $R = x_m - x_0$

Where x_m stands for the largest observation and x_0 denotes the smallest one. When the ~~are~~ data are grouped into a frequency distribution, the range is estimated by finding the difference between the upper boundary of the highest class and lower boundary of the lowest class. The range can not be computed if there are any open end classes in distribution.

The range is a simple concept and is easy to compute. It has however two serious disadvantages. First, it ignores all the information available from the intermediate observations and second, as its value is based only on two extreme (usually large or small) observations it might give a misleading picture of the spread in the data. It is therefore unsatisfactory measure of ~~data~~ dispersion. However it is appropriately used in statistical quality control charts of manufactured products, daily temperatures, stock prices etc. This is an absolute measure of dispersion. Its relative measure known as the co-efficient of dispersion.

Example: We have a data as; 6, 8, 9, 10, 12 - highest
lowest

so Range $= X_m - X_0$
$= 12 - 6 = \boxed{6} \rightarrow$ Range.

2. QUARTILE RANGE :

Quartile and the interquartile range can be used to group and analyze data sets.

A Quartile is a group of values and/or means that divide a data set into quarters or groups four. A quartile is a value and not a group of numbers. Think of a quartile as a cutt off point for each group. A group has to start and stop somewhere and that's exactly what a quartile does.

The interquartile range is a value that is difference between the upper quartile value and lower quartile value

Symbollically ; $IQR = Q_3 - Q_1$.

Example :

Median = 71

lower half. | Upper Half.

62 63 64 64 70 | 72 76 77 81 81

Upper quartile

$\boxed{\text{Inter Quartile Rang : } 77-64=13}$

$Q_1 = 64$.

$Q_3 = 77$.

There are 5 values below the median (lower half), the middle value is 64 which is the first quartile. There are 5 values above the median (upper half), the middle value is 77 which is the third quartile. The IQR is 77-64 = 13; the IQR is the range of the middle 50% of the data.

3. SEMI INTER QUARTILE RANGE:

The interquartile range is a measure of dispersion, defined by the difference between the third and first quartiles, and half of this range is called Semi-interquartile Range (S.I.Q.R) or the quartile deviation (Q.D).

Symbollically;

$$Q.D = \frac{Q_3 - Q_1}{2}.$$

Where Q_1 and Q_3 are the first and the third quartiles of the data. The QD has an attractive feature that the range "Median ± QD" contains approximately 50% of the data. The quartile deviation is superior to range as it is not affected by extremely large and small observations. It is simple to understand and easy to calculate. It however give no information about the position of observations lying outside the two quartiles, is amenable to mathematical treatment and is greatly affected by sampling variability. The quartile deviation is not as widely used as other measures of dispersion. It is however, used in situations where extreme observations are thought to be ~~un~~ unrepresentative.

4. VARIANCE:

The variance of a set of observations is defined as the mean of the squares of deviations of all the observations from their mean. When it is calculated from the entire population, the variance is called the population variance traditionally denoted by σ^2 (σ is the greek lowercase "sigma"). If instead, the data from the sample is used to calculate the variance, it is then called as sample variance and is denoted by S^2 inorder to distinguish between the two. The symbolic definition of variance is;

$$\sigma^2 = \frac{\Sigma(x_i - \mu)^2}{N}, \text{ for population data.}$$

$$S^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}, \text{ for sample data}$$

5. STANDARD DEVIATION:

The positive square root of the variance is called standard deviation.

Symbolically,

$$\sigma = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}, \text{ for population dat}$$

$$S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}, \text{ for sample data.}$$

The standard deviation is expressed in the same units as the observation themselves and is a measure of the average spread around the mean.

6. CO-EFFICIENT OF VARIATION:

The variability of two or more than two sets of data can not be compared unless we have relative measure of dispersion. For this purpose, Karl Pearson (1857–1936) introduced a relative measure of variation, known as the co-efficient of variation, abbreviated C.V, which expresses the standard deviation as a percentage of the arthmetic mean of a data set.

Symbolically ; $C.V = \frac{S}{\bar{x}} \times 100$, for sample data.

$$C.V = \frac{\sigma}{\mu} \times 100, \text{ for population data.}$$

As the coefficient of variance is a pure number without units, it is therefore used to compare the variation in two or more data sets or distributions that are measured in different units e.g one may be measured in hours and other in kg or rupees. A large value of CV indicates that the variability is great and a small value of CV indicates less variability.

The coefficient of variation is also used to compare the performance of two candidates or of two players given their scores in various papers or games. the smaller co-efficient of variance the more consistent is the performance. It should be noted that the coefficient of variance is unreliable when the arthmetic mean is very small.

Scanned with CamScanner