Assignment No. 2

Q1. An intelligence test was administreted on 8 students they obtaind followoing scores in the test Find out mean of intelligence scores.

80, 100, 120, 105, 90, 110, 115, 112

$n = 8$

$$\text{mean } (\bar{x}) = \frac{\sum x_i}{n}$$

$$= \frac{80+100+120+105+90+110+115+112}{8}$$

$$= \frac{832}{8}$$

$\therefore$ mean $(\bar{x}) = 104$

Mean of intelligence scores is 104.

Q2. Find out mean, median & mode for following scores.

(a) 22, 21, 24, 18, 19, 23, 12, 20

$n = 8$

$\rightarrow$ $\text{mean}(\bar{x}) = \frac{\sum x_i}{n}$

$$= \frac{22+21+24+18+19+23+12+20}{8}$$

$$= \frac{159}{8}$$

mean$(\bar{x}) = 19.87$

$\rightarrow$ To calculate meadian we are arranging data in asending order.

12, 18, 19, 20, 21, 22, 23, 24

Here $n = 8$ (Even)

using median formula,

→ median $= \dfrac{\left(\frac{n}{2}\right)^{th} + \left(\frac{n}{2}+1\right)^{th}}{2}$

$= \dfrac{\left(\frac{8}{2}\right)^{th} + \left(\frac{8}{2}+1\right)^{th}}{2}$

$= \dfrac{4^{th} \text{ term} + (4+1)^{th} \text{ term}}{2}$

$= \dfrac{4^{th} \text{ term} + 5^{th} \text{ term}}{2}$

$= \dfrac{20+21}{2} = \dfrac{41}{2}$

median $= 20.5$

→ Mode

mode is nothing but the most occuring number.

In above dataset there is no datapoint is appering more than once.

So, mode is ~~not~~ undefined for the same.

(b) 9, 8, 13, 10, 11, 10, 12, 10, 14

→ ~~n=8~~ $n = 9$

$$\text{mean}(\bar{x}) = \frac{\sum x_i}{n}$$

$$= \frac{9+8+13+10+11+10+12+10+14}{9}$$

$$= \frac{97}{9}$$

$$\text{mean}(\bar{x}) = 12.125$$

→ Arranging data in ascending order to calculate the median.

8, 9, 10, 10, 10, 11, 12, 13, 14.

Here, $n = 9$ (odd)

By using median Formula

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

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

$$= \left(\frac{9+1}{2}\right)^{th} \text{ term}$$

$$= \frac{10}{2} = 5^{th} \text{ term}$$

median = 10

→ Mode

To calculate mode we are ~~call~~ counting the number of occurences for each data points

x	No. of occurences
8	1
9	1
10	3
11	1
12	1
13	1
14	1

From the above observation its seems that 10 is appears three times, which is the higest frequency

∴ mode = 10

Q.3 Compute the mean, median & mode for the following frequency distribution.

CI	F	X	F(X)	CCI	CF
100-109	5	104.5	522.5	99.5-109.5	5
90-99	9	94.5	850.5	89.5-99.5	14
80-89	14	84.5	1183	79.5-89.5	28
70-79	19	74.5	1415.5	69.5-79.5	47
60-69	21	64.5	1354.5	59.5-69.5	68 = cf
50-59	f=30	54.5	1635	$\boxed{49.5-59.5}$	98
40-49	25	44.5	1112.5	39.5-49.5	123
30-39	15	34.5	517.5	29.5-39.5	138
20-29	10	24.5	245	19.5-29.5	148
10-19	8	14.5	116	9.5-19.5	156
0-9	6	4.5	27	0.5-9.5	162
	162		8979		

→ $Mean(\bar{x}) = \frac{\sum f(x)}{\sum f} = \frac{8979}{162} = 55.425$

→ To calculate mean we measured the midpoint (x) of classinterval.
After ~~that~~ calculating midpoint we calculated f(x) by multiplying midpoint and frequency (x * f).

The mean of the given grouped dataset is approximately 55.42.

→ median.
(1) Frist we converting class Interval to continous class Interval by subtracting 0.5 from lower limit & addition of 0.5 in upper limit of each class Interval.
(2) calculate cumulative frequency
(3) Find median class.

$N = 162$

$\frac{N}{2} = \frac{162}{2} = 81$

81 lies between 68-98 cf

median class (49.5-59.5)

$L_1 = 49.5, f = 30, cf = 68, L_2 = 59.5, \frac{N}{2} = 81$

(4) By using median formula,

$$median = L_1 + \left(\frac{\frac{N}{2} - cf}{f} \right) \times (L_2 - L_1)$$

$$= 49.5 + \left(\frac{81-68}{30}\right) \times 10$$

$$= 49.5 + \frac{13}{30} \times 10$$

$$= 49.5 + \frac{130}{30} = 49.5 + 4.33$$

median. $= 53.83$

→ mode

Highest frequency $= 30$

~~By~~

$L_1 = 49.5$, $L_2 = 59.5$, $F_0 = 21$, $f_1 = 30$, $F_2 = 25$

By using mode formula,

$$\text{mode} = L_1 + \left(\frac{f_1 - f_0}{(f_1 - f_0) + (f_1 - f_2)}\right) \times (L_2 - L_1)$$

$$= 49.5 + \left(\frac{30-21}{(30-21) + (30-25)}\right) \times 10$$

$$= 49.5 + \frac{9}{9+5} \times 10$$

$$= 49.5 + \frac{9}{14} \times 10$$

$$= 49.5 + 6.42$$

mode $= 55.92$

PAGE NO.:

Q4. In the situation describe below, what measure of central tendency would you like to compare & why?

(1) The average of intelligence class

→

(i) To compare the average intelligence level of class, we can use "mean".

(ii) The mean, also called the average, is the sum of all intelligence score divided by the number of student in the class

(2) The most popular dress of teenagers:

→

(i) To compare the popularity of dress among teenagers the most suitable measure is "mode"

(ii) It reprents that dress which is worn by the most number of teenagers.

(iii) By finding mode, we can easily determine the most commonly worn dress and understand its populcirity compard to other dresses.

(3) Determine the midpoint of the score of group in examination

→

(i) To compare the midpoint of the score of group in examination, the most appropriate measure is "Median".

(ii) median will represent the midpoint of scores.

(iii) To calcullate median, we need to arrange the individual scores in asending order. The process allow us to identify the exact middle value or average of two middle values (if the number of scores is even)

PAGE NO.:

Q.5 "Every measure of central Tendency has it own particular charcteristics. It is difficult to say which measure is best." Explain it with an Example.

→ The statement highlights the fact that each measure of central tendency (mean, median, mode) has its strengths and weaknesses. The choice of the most appropricde measure depends on the nature of the data and the specific objective of the analysis.

Example!- consider the following data representing the ages of a group of indivituals (in years):

12, 15, 18, 22, 25, 35, 40, 45, 60, 75.

① mean: The mean is the sum of all data points divided by the number of data points. It represents the average age of the group

mean = (12+15+18+22+25+35+40+45+60+75)/10

mean ≈ 33.2

Advantages: The mean takes into account all the data points, providing a comprehensive view of the central tendency. It is sensitive to small changes in data, giving a precise representation of the average age.

Disadvantages: The mean is highly influenced by extreme values (outliers). In this dataset, the value of 75 is an outlier, and it significantly affetcts the mean, making it higher than the majority of ages.

② median: The median is the middle value when the data is arranged in ascending or descending order. It seperate the higher half from the lower half of the data.

Arranging data in ascending order.

12, 15, 18, 22, 25, 35, 40, 45, 60, 75

median = 25

Advantages: The median is not affected by outliers. It represents the "typical" value and is especially useful when the data is skwed or contains outliers.

Disadvantages: The median does not ~~contain~~ consider the exact values of all data points, which can be a limitation when precise values are required.

③ mode: The mode is the value that appears most frequently in the dataset. In this dataset, there is no mode as all values occurs only once.

Advantages: The mode is useful when identifying the most common value or category in Categorical data.

Disadvantages: The mode might not provide a comprehensive view of the central tendency when data is continuous.

In this example, we see that each measure of central tendency has its pros and cons. The mean might not be best choice because of the outliers, while the median is a more robust representation in the presence of outliers. The choice of the best measure depends on the specific context and the characteristics of the data being analyzed.

Q.6 Find the Range and variance of following data
18, 12, 14, 16, 18, 13, 4, 4, 7, 12, 9, 12, 4, 7

→

* Range:
  Largest value $(x_n) = 18$
  Smallest value $(x_1) = 4$

  $\therefore$ Range $= x_n - x_1 = 18 - 4 = 14$

* Variance: spread of the data.

  No. of observations $(n) = 14$
  We need mean to calculate variance.

  $$\text{mean } (\bar{x}) = \frac{\Sigma x_i}{n} = \frac{150}{14} = 10.71$$

  mean $(\bar{x}) \approx 11$

PAGE NO.:

x	$x_i - \bar{x}_i$	$(x_i - \bar{x})^2$
18	7	49
12	1	1
14	3	9
16	5	25
18	7	49
13	2	4
4	-7	49
4	-7	49
7	-4	16
12	1	1
9	-2	4
12	1	1
4	-7	49
7	-4	16
		322

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

$$= \frac{322}{14}$$

$$\sigma^2 = 23$$

$$\boxed{\text{Variance} = 23}$$

PAGE NO.:

Q.7 write advantages & disadvantages of Range & variance.

⇒ Range

	Advantages	Disadvantages
(1)	Simple and easy to calculate	Highly sensitive to outliers, which can distort its value
(2)	Provides an intuitive representation of data spread	Ignores the distribution of data between extreme points
(3)	useful for quick comparison of variability in different datasets	Not suitable for continuous data with a large number of values
(4)	Requires only two data points (highest and lowest) to compute	Dependent on sample size, can be unreliable for small samples.

⇒ Variance.

	Advantages	Disadvantages
(1)	Provide a comprehensive measure of data dispertion and variability	Sensitive to outliers, which can lead to inflated variance values
(2)	Accounts for all data points, considering individual deviations from the mean	square unit make it less interpretable in the original scale of the data.

Advantages	Disadvantages
(3) Useful for comparing the spread of data across different datasets.	Assumes that data follows a normal distribution, which may not always be true.
(4) Basis for calculating standard deviation, a widely used measure in statistics	magnifies errors and deviations due to squaring of differences.
(5) Useful in inferential statistics for estimating population variance from sample variance	Complex calculations may be challenging

Q.8 Write limitations of mean, median & mode.

→

(A) Limitations of mean.

(i) Sensitive to outliers:

The mean is highly influenced by extreme values or outliers, which can significantly distort its value and may not represent the central tendency accurately.

(ii) Not suitable for skewed Data:

In the presence of skewed distributions, the mean may not be a representative measure as it can be pulled towards the ~~sample size~~. long tail of the distribution.

PAGE NO.:

(iii) Affected by sample size:
The mean can be affected by sample size. In small samples, It might not provide a reliable estimate of population mean.

(B) Limitations of median:
(i) Ignores the magnitude of Deviations:
The median only considers the middle value(s) and does not take into account the magnitude of deviations from the central point. It provides less information about spread of data

(ii) Not suitable for Continuous Data:
The median might not adequately represent the central tendency for continuous data, as it could lie between two data points, lossing precision.

(iii) Less sensitive to Data:
While the median is robust to outliers, it is less sensitive to changes in the data, which might be a limitation when analyzing subtle changes or patterns.

(C) Limitations of Mode:
(i) May Not Always Exist:
In some datasets, there may be no mode or multiple modes, making it less informative as a measure of central tendency

(ii) Not suitable for Numerical Data:
The mode is commonly used for categorical or nominal data but might not be applicable or meaningful for numerical data

(iii) Ignores magnitude:
The mode does not consider the actual values but only identifies the most frequent category or value, disregarding the magnitude or numerical information.