CMR
INSTITUTE OF
TECHNOLOGY

USN | | | | | | | | | | |

Internal Assessment Test - I

Sub:	Business Statistics							Code:	MBA104
Date:	**26.12.2025**	Duration:	**90 minutes**	Max Marks:	**50**	Sem:	**I**	Branch :	**MBA**
SET- II									

Part A

1. a. Outline the limitations of Statistics – 3 Marks

Deals Only with Quantitative Data:
Statistics considers only numerical data. Qualitative aspects such as honesty, motivation, intelligence, and attitudes cannot be measured accurately in numerical terms.

Does Not Study Individual Cases:
Statistical methods deal with aggregates or groups and not with individual items. Hence, individual variations may be ignored.

Results Are Approximate, Not Exact:
Statistical conclusions are based on averages and probabilities, so the results are only approximate and not 100% accurate.

Requires Adequate and Reliable Data:
If data collected are insufficient, biased, or inaccurate, statistical results will be misleading ("garbage in, garbage out").

Easily Misused or Misinterpreted:
Statistics can be manipulated to support preconceived conclusions if methods are improperly applied or results are selectively presented.

Cannot Establish Cause-and-Effect Relationships Alone:
Statistics can indicate association or correlation but cannot by itself prove causation.

Depends on Skilled Interpretation:
Statistical tools require proper knowledge and expertise. Incorrect analysis or interpretation can lead to wrong decisions.

b. Calculate the Mean, Median, and Mode from the following data – 7 Marks

X	10	20	30	40	50
f	4	7	10	6	3

Calculate Required Totals

$$\sum f = 4 + 7 + 10 + 6 + 3 = 30$$

$$\sum fx = (4 \times 10) + (7 \times 20) + (10 \times 30) + (6 \times 40) + (3 \times 50)$$

$$= 40 + 140 + 300 + 240 + 150 = 870$$

A. Arithmetic Mean

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{870}{30} = 29$$

Mean = 29

B. Median

First, find cumulative frequencies (cf):

X	f	cf
10	4	4
20	7	11
30	10	21
40	6	27
50	3	**30**

$$\frac{N}{2} = \frac{30}{2} = 15$$

The 15th item lies in the cumulative frequency **21**, corresponding to **X = 30**.

Hence, Median = 30

3. Mode

The value with the highest frequency is **30** (f = 10).

Hence, Mode = 30

c.(i). The average marks of 40 students in a class was calculated as 68. Later it was found that the marks of one student were entered as 52 instead of 82. Find the correct average marks. – 5 Marks

Given:
Average marks of 40 students (incorrect) = 68

Step 1: Find the total marks (incorrect)

$$\text{Total (wrong)} = 68 \times 40 = 2720$$

Step 2: Find the error in marks

Correct marks = 82
Wrongly entered marks = 52

Difference $= 82 - 52 = 30$

Step 3: Find the corrected total

$$\text{Correct total} = 2720 + 30 = 2750$$

Step 4: Find the correct average

$$\text{Correct average} = \frac{2750}{40} = 68.75$$

Correct Average Marks = 68.75

(ii). The mean height of a group of students is 162 cm. Find the missing frequency-5 Marks

Height (cm)	150–155	155–160	160–165	165–170	170–175
Frequency	6	14	?	10	5

Let the missing frequency for the class 160–165 be **x**.

Step 1: Find class mid-points (xi)

Height (cm)	Frequency (fi)	Mid-point (xi)	fi·xi
150–155	6	152.5	915
155–160	14	157.5	2205
160–165	x	162.5	162.5x
165–170	10	167.5	1675
170–175	5	172.5	862.5

Step 2: Apply the Mean Formula

$$\bar{X} = \frac{\sum fi\, xi}{\sum fi}$$

Given Mean $\bar{X} = 162$

$$162 = \frac{915 + 2205 + 162.5x + 1675 + 862.5}{6 + 14 + x + 10 + 5}$$

$$162 = \frac{5657.5 + 162.5x}{35 + x}$$

Step 3: Solve the Equation

$$162(35 + x) = 5657.5 + 162.5x$$

$$5670 + 162x = 5657.5 + 162.5x$$

$$162x - 162.5x = 5657.5 - 5670$$

$$-0.5x = -12.5$$

$$x = 25$$

The missing frequency = 25 students.

2. a. Mention two use cases each for arithmetic mean, geometric mean and harmonic mean

-3 Marks

A. **Arithmetic Mean (AM)-** best for simple, direct data.

- **Average Performance Measurement:** Used to find average marks, average sales, average income, or average production in business and education.
- **Business Planning & Reporting:** Helps management summarize large data sets such as average monthly sales, average costs, or average employee salary.

B. **Geometric Mean (GM) –** suitable for growth and multiplicative data.

- **Growth Rate Analysis:** Used to calculate average growth rates such as CAGR in sales, profits, population growth, or investment returns.
- **Index Numbers & Ratios:** Applied in constructing index numbers (price index, quantity index) and averaging ratios or percentages.

C. **Harmonic Mean (HM) –** appropriate for averaging rates and ratios.

- **Average of Rates and Ratios:** Used when averaging speeds (e.g., average speed over equal distances) or rates like cost per unit.
- **Financial & Operational Efficiency:** Applied in calculating average price per unit when total expenditure is fixed or in performance metrics like inventory turnover.

b.Calculate the Mode using the Grouping Method: - 7 marks

X	10	20	30	40	50
f	6	12	20	12	6

The Grouping Method is used to determine the mode when the maximum frequency is not unique or when there is a very small difference between the highest frequency and its neighbouring frequencies. This prevents "errors of sampling" where a single high value might be a fluke.

Step 1: Prepare the Grouping Table
6 columns created. In each column, we identify the maximum value (highlighted in bold).

X	Col I (f)	Col II (1+2)	Col III (2+3)	Col IV (1+2+3)	Col V (2+3+4)	Col VI (3+4+5)
10	6					
20	12	18		**38**		
30	20		**32**		**44**	
40	12	**32**				**38**
50	6		18			

Explanation of columns:

- **Col I:** Original f. Maximum is **20**.
- **Col II:** Sum in pairs (6+12, 20+12). Maximum is **32**.
- **Col III:** Sum in pairs, skipping the 1st (12+20, 12+6). Maximum is **32**.
- **Col IV:** Sum in threes (6+12+20). Maximum is **38**.
- **Col V:** Sum in threes, skipping the 1st (12+20+12). Maximum is **44**.
- **Col VI:** Sum in threes, skipping the 1st and 2nd (20+12+6). Maximum is **38**.

Step 2: Prepare the Analysis Table
Now, we tally how many times each value of X contributed to a maximum value in the columns above.

Column	10	20	30	40	50
Col I			1		
Col II			1	1	
Col III		1	1		
Col IV	1	1	1		
Col V		1	1	1	
Col VI			1	1	1
Total Tally	1	3	6	3	1

Final Result
The value X = 30 has the highest frequency of occurrences (**6**) across all grouping combinations.
Mode = 30

c.The following data shows the marks obtained by students in a test – 10 Marks

Marks	35	40	45	50	55	60	65
No. of Students	4	6	10	12	8	5	3

Find:

(a) Range
(b) Coefficient of Range
(c) Inter-Quartile Range
(d) Semi-Quartile Range
(e) Coefficient of Quartile Deviation
(f) Assess the dispersion of marks and justify whether the performance is uniform or varied.

Note: how to decide on frequency distribution positions

Type of Data	Use
Individual (raw) data	(N+1)
Discrete / frequency data	(N)
Grouped data	Formula with class limits

Solution:

$$N = 4 + 6 + 10 + 12 + 8 + 5 + 3 = 48$$

(a) Range

Highest value (H) = 65
Lowest value (L) = 35

$$\text{Range} = H - L = 65 - 35 = 30$$

(b) Coefficient of Range

$$\text{Coefficient of Range} = \frac{H - L}{H + L} = \frac{65 - 35}{65 + 35} = \frac{30}{100} = 0.30$$

(c) Inter-Quartile Range (IQR)

Step 1: Cumulative Frequency (cf)

Marks	f	cf
35	4	4
40	6	10
45	10	20
50	12	32
55	8	40
60	5	45
65	3	48

$$Q_1 \text{ position} = \frac{N}{4} = \frac{48}{4} = 12^{th} \text{ item}$$

12th item lies at **45**

$$Q_3 \text{ position} = \frac{3N}{4} = \frac{144}{4} = 36^{th} \text{ item}$$

36th item lies at **55**

$$\text{IQR} = Q_3 - Q_1 = 55 - 45 = 10$$

(d) Semi-Quartile Range (Quartile Deviation)

$$\text{QD} = \frac{Q_3 - Q_1}{2} = \frac{10}{2} = 5$$

(e) Coefficient of Quartile Deviation

$$\text{Coefficient of QD} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{55 - 45}{55 + 45} = \frac{10}{100} = 0.10$$

(f) Interpretation & Comment on Dispersion

- Range = 30 indicates a moderate spread between lowest and highest marks.
- Coefficient of Range (0.30) shows limited extreme variability.
- Inter-Quartile Range (10) and Quartile Deviation (5) indicate that the middle 50% of students' marks are closely clustered.
- Low Coefficient of Quartile Deviation (0.10) suggests high consistency.

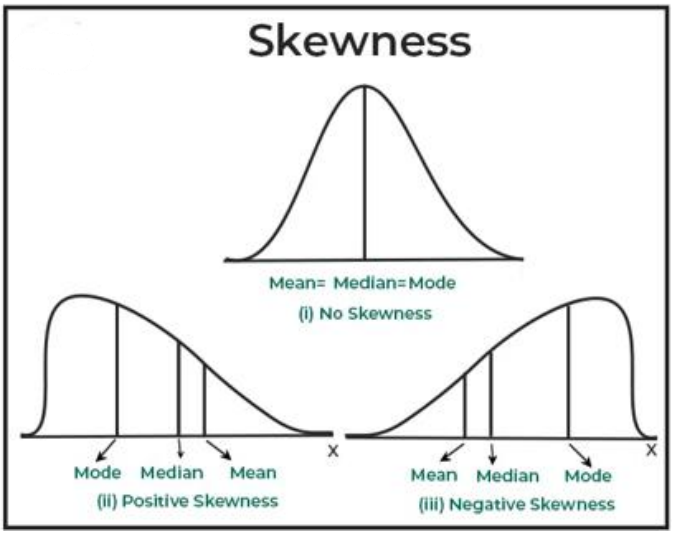
    - The dispersion of marks is low to moderate, and student performance is fairly uniform, with most students scoring near the central values.

3. a. Explain skewness diagrammatically.- 3 Marks

Skewness refers to the degree of asymmetry in a frequency distribution about its central value (mean). Skewness is a measure of the asymmetry of a probability distribution. In a perfectly symmetrical distribution (like the Normal Distribution), the left and right sides are mirror images of each other. When a distribution is "skewed," it is stretched more on one side than the other.

Feature	Negatively Skewed	Symmetrical	Positively Skewed
Tail Direction	Long tail to the left	Equal tails	Long tail to the right
Peak Position	Shifted to the right	Center	Shifted to the left
Comparison	Mean < Median < Mode	Mean = Median = Mode	Mean > Median > Mode

Skewness Value	Skewness < 0	Skewness = 0	$Skewness > 0



b.Outline the value of Median, Q_1, Q_3, D_7, $P_{50}$and P_{95}: - 7 marks

Income (₹ '000)	15	20	25	30	35	40	45
Frequency	8	14	22	36	28	16	6

Step 1: Cumulative Frequency (cf)

Income	f	cf
15	8	8
20	14	22
25	22	44
30	36	80
35	28	108
40	16	124
45	6	130

$$N = \sum f = 130$$

(a) Median

$$\text{Median position} = \frac{N}{2} = \frac{130}{2} = 65^{th} \text{ item}$$

65th item lies in cf **80**, corresponding to **Income = 30**

Median = 30

(b) First Quartile (Q_1)

$$Q_1 \text{ position} = \frac{N}{4} = \frac{130}{4} = 32.5^{th} \text{ item}$$

32.5th item lies in cf **44**, corresponding to **Income = 25**

$$Q_1 = 25$$

(c) Third Quartile (Q_3)

$$Q_3 \text{ position} = \frac{3N}{4} = \frac{390}{4} = 97.5^{th} \text{ item}$$

97.5th item lies in cf **108**, corresponding to **Income = 35**

$$Q_3 = 35$$

(d) Seventh Decile (D_7)

$$D_7 \text{ position} = \frac{7N}{10} = \frac{7 \times 130}{10} = 91^{st} \text{ item}$$

91st item lies in cf **108**, corresponding to **Income = 35**

$$D_7 = 35$$

(e) 50th Percentile (P_{50})

$$P_{50} \text{ position} = \frac{50N}{100} = \frac{50 \times 130}{100} = 65^{th} \text{ item}$$

Same as median → **Income = 30**

$$P_{50} = 30$$

(f) 95th Percentile (P_{95})

$$P_{95} \text{ position} = \frac{95N}{100} = \frac{95 \times 130}{100} = 123.5^{th} \text{ item}$$

123.5th item lies in cf **124**, corresponding to **Income = 40**

$$P_{95} = 40$$

Answer Summary

Measure	Value (₹'000)
Median	30
Q_1	25
Q_3	35
D_7	35

P_{50}	30
P_{95}	40

c. A production supervisor observed the daily output (in units) of a machine over several days. The data collected is: - 10 Marks

Output (Units)	50	60	70	80	90
No. of Days	4	6	10	6	4

To assess production stability, you are required to:

a) Compute the Variance of daily output.
b) Calculate the Standard Deviation.
c) Interpret the results with respect to production consistency.

Step 1: Compute Required Values

x	f	fx	x^2	$f x^2$
50	4	200	2500	10,000
60	6	360	3600	21,600
70	10	700	4900	49,000
80	6	480	6400	38,400
90	4	360	8100	32,400

$$\sum f = 30, \sum fx = 2100, \sum fx^2 = 151{,}400$$

Step 2: Calculate Mean

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{2100}{30} = 70$$

a) Variance

$$\sigma^2 = \frac{\sum fx^2}{\sum f} - (\bar{X})^2$$

$$\sigma^2 = \frac{151{,}400}{30} - (70)^2$$

$$= 5046.67 - 4900$$

$$= 146.67$$

Variance = 146.67

(b) Standard Deviation

$$\sigma = \sqrt{146.67} \approx 12.11$$

Standard Deviation ≈ 12.11 units

(c) Interpretation: Production Consistency

- The **mean daily output** is **70 units**.
- The **standard deviation (12.11 units)** is relatively small compared to the mean.
- Output values are symmetrically distributed around the mean.
- Most production days fall close to 70 units, with limited fluctuation.

- The machine's daily output shows **good consistency and stability**, indicating a **reliable production process** with moderate variability.

4. Case Study: 10 Marks

An HR analyst examined whether years of experience influence employee productivity scores in an organization.

Experience (Years) (X)	1	3	5	7	9
Productivity Score (Y)	40	48	60	65	70

a) Calculate Karl Pearson's coefficient of correlation between experience and productivity.
b) Interpret the correlation result.
c) Comment on the HR implications of the findings.

a) Karl Pearson's Coefficient of Correlation
(Actual Mean Method)
Step 1: Calculate actual means

$$\bar{X} = \frac{\sum X}{n} = \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5$$

$$\bar{Y} = \frac{\sum Y}{n} = \frac{40+48+60+65+70}{5} = \frac{283}{5} = 56.6$$

Step 2: Find deviations from actual means

$$x = X - \bar{X}, y = Y - \bar{Y}$$

X	Y	x = X−5	y = Y−56.6	x^2	y^2	xy
1	40	−4	−16.6	16	275.56	66.4
3	48	−2	−8.6	4	73.96	17.2
5	60	0	3.4	0	11.56	0
7	65	2	8.4	4	70.56	16.8
9	70	4	13.4	16	179.56	53.6
Σ		**0**	**0**	**40**	**611.2**	**154**

Step 3: Apply Karl Pearson's formula (Actual Mean Method)

$$r = \frac{\sum x\, y}{\sqrt{(\sum x^2)(\sum y^2)}}$$

Step 4: Substitute values

$$r = \frac{154}{\sqrt{(40)(611.2)}}$$

$$r = \frac{154}{\sqrt{24448}}$$

$$r = \frac{154}{156.36} \approx 0.985$$

b) Interpretation of the Correlation Result

•**Direction:** Positive

•**Degree:** Very high correlation

Interpretation:

There is a **strong positive relationship** between years of experience and employee productivity. As experience increases, productivity scores also increase.

c) HR Implications of the Findings

•Experienced employees tend to be **more productive**

•Retention strategies for experienced staff can improve organizational performance

•HR should invest in:

- Employee development
- Training and upskilling
- Career progression and retention policies

•Experience can be considered an important factor in **promotion and role assignment decisions**