

FYBA- Statistics I- Semester I

Correct the following if necessary-(2 marks each)

1. Median gets affected by extreme observations in data.
2. For a symmetric distribution $\beta_1 = 3$.
3. Population is subset of sample.
4. Primary data can be collected by oral interview of respondent.
5. Mean can be located by Histogram.
6. Mean deviation about mean is minimum.
7. Coefficient of variation is measure of skewness.
8. Gender, Religion are examples of quantitative data.
9. Intersection of two types of ogives occurs at mode.
10. Percentiles divide the data into 50 equal parts.
11. Colour of hair is an attribute.
12. SRSWR is a type of sampling.
13. Percentiles divide the data into 10 equal parts.
14. Mode can be determined using 'more than ogive'.
15. Standard deviation is a measure of central tendency.
16. Income of a person is an attribute.
17. Ultimate class frequencies are the frequencies of the classes of lowest order.
18. Median is the minimum value in the data.
19. Upper limit is included in case of inclusive type class intervals.
20. If graph of the frequency curve is symmetric then variable is negatively skewed.
21. Beauty is a variable
22. Discrete variable is a variable taking all possible values in a certain range.
23. If observations in a data set are arranged from highest to lowest, the observations lying in the middle is the median of the data set.
24. Mean absolute deviation is equal to the positive square root of the variance.
25. Third order central moments can be negative.
26. Religion is a variable.
27. Lower limit is excluded in case of exclusive type of class intervals.
28. Mode is the minimum value in the data.
29. Mean deviation is a measure of central tendency.
30. Standard deviation (S.D.) is not based on all observations.
31. For leptokurtic distribution $\beta_2 = 3$.
32. Range is measure of central tendency.
33. Weight, height are examples of attributes.
34. Histogram is used to locate median.
35. To find average speed of journey weighted arithmetic mean is used.
36. Monthly salary is an attribute.

37. For a symmetric distribution $\mu_1 = 3$
38. $AM \geq GM \geq HM$
39. Quartile deviation = $Q_1 - Q_3$
40. Data from Economic and Political weekly is primary data.
41. Rainfall for state of Maharashtra represents a discrete random variable.
42. Skewness is departure from symmetry.
43. Standard deviation is not affected by shift of origin and change in scale.
44. Stubs stands for column headings.
45. Median is always equidistant from first and third quartiles.
46. Variance can be negative.
47. The symmetric curve is always mesokurtic curve.
48. In Histogram, if all the class widths are of same size then height of rectangle is proportional to the frequency density.
49. Published data is primary data.
50. The age of students is an Attribute.
51. Mode is located using Ogive curves.
52. Mean deviation about mean is maximum.
53. Range is measure of central tendency.
54. Income of a person is an attribute.
55. Ultimate class frequencies are the frequencies of the classes of lowest order.
56. Median is the minimum value in the data.
57. Upper limit is included in case of inclusive type of class intervals.
58. If graph of the frequency curve is symmetric then variable is negatively skewed.

Answer in one sentence (2 marks each)

1. What is simple random sampling with replacement and simple random sampling without replacement?
2. Three of the 5 observations are 1,2,6. If mean is 4.4 and variance is 8.24, find remaining 2 observations.
3. If the number of observations in ungrouped data are odd then how to calculate median?
4. State coefficient of kurtosis based on moments. State how will you decide type of kurtosis using it.
5. Distinguish between variable and attribute.
6. What are the sources available for collection of secondary data?
7. Find geometric mean of $5, 5^2, 5^3, 5^4, 5^5$.
8. State any two components of statistical table.
9. Define measure of kurtosis.
10. What is simple random sampling?
11. State two requirements of a good questionnaire.
12. Define class marks and width of a class interval.
13. State the empirical relation between mean, mode and median.
14. Explain how two sets of data can be compared using standard deviation.

15. State a difference between simple random sampling without replacement and simple random sampling with replacement.
16. Define class boundaries and width of a class interval.
17. Define class boundaries and width of a class interval.
18. State any two sources of secondary data.
19. Explain kurtosis in a data.
20. The coefficient of variation for a variable is 35% and arithmetic mean is 35.2. Obtain variance.
21. State any two objectives of classification.
22. Define class limits and width of a class interval.
23. Explain how two sets of data can be compared using range.
24. Define weighted arithmetic mean.
25. What is meant by symmetric distribution?
26. State when median is appropriate measure of central tendency.
27. Explain the role of range as a measure of dispersion.
28. State any two requirements of a good statistical table.
29. What is association of attributes?
30. Explain any two methods of primary data collection.
31. For the data given below find mode
7, 10, 12, 10, 16, 15, 18, 16, 27, 20, 30, 16
32. Give one merit and one demerit of Mean.
33. What are types of skewness of data.

Attempt the following-(10 marks each)- Unit -I

1. Describe different scales of measurement with example.
2. Discuss requirements of a good questionnaire.
3. What is primary data? Explain in detail different methods of collecting primary data.
4. Find all ultimate class frequencies from the following data and tabulate them.

(A) = 224	(B) = 301	(C) = 150	(AB) = 125
(AC) = 72	(BC) = 60	(ABC) = 32	N = 800.
5. What is association in context with attributes? Define Yule's coefficient of association (Q) and Yule's coefficient of colligation (Y). Derive the relationship between them.
6. (i) Examine whether following data are consistent.

N = 100	(A) = 30	(B) = 80	(AB) = 40
---------	----------	----------	-----------

(ii) Present the following information in a tabular form.

Out of total 2000 candidates, interviewed for employment in a company in Nashik, 628 were from Pune and the rest from Nashik. Amongst the graduates from Pune, 350 were experienced and 80 were inexperienced. While the corresponding figures for undergraduates from Nashik were 615 and 52 respectively. The total number of inexperienced candidates from Pune and Nashik were 175 and 192 respectively.

7. Explain what is meant by secondary data. Discuss various sources of secondary data. State differences between primary data and secondary data.
8. Explain what is meant by secondary data. Discuss various sources of secondary data. State differences between primary data and secondary data.
9. (i) Describe association between attributes. When do you say that two attributes are positively and negatively associated? Define Yule's coefficient of association (Q) and Yule's coefficient of colligation (Y).
(ii) State the conditions for consistency of data based on two attributes.
10. (i) State the requirements of a good statistical table.
(ii) Give a blank table representing data on students of a college according to following characteristics in a table - (a) Faculty- Arts, Science and Commerce (b) Class- First year, Second year and Third year.
11. (i) Define Yule's coefficient of association (Q) and Yule's coefficient of colligation (Y). Derive the relationship between them.
(ii) State the conditions for consistency of data based on three attributes.
12. (i) Give a blank table representing data on students according to following characteristics in a same table - (a) Faculty- Arts, Commerce and Science (b) Gender- Boys and Girls.
(ii) State requirements of good statistical table.
13. (i) Define: questionnaire, schedule.
(ii) How questionnaire is different from schedule?
14. (i) Prepare a blank table with appropriate title to show the number of women in the age group 20-30, 30-40, 40-50 in Mumbai according to Marital status: Married, Unmarried and Employment status: Employed, Unemployed.
(ii) 'A secondary Source is not reliable as a primary source'.
Discuss the validity of the statement.
15. (i) What is classification? What are main objectives of classification?
(ii) Describe various parts of a statistical table.
16. (i) For two attributes A and B we have $(AB) = 8$, $(A) = 18$, $(\alpha\beta) = 5$, $N = 35$; find missing frequencies hence check consistency of the data also compute Yule's coefficient of association.
(ii) What is time series data and cross section data? Give one example of each.
17. Write a short note on classification and tabulation of data.
18. Explain the concept of consistency of data. Derive consistency conditions in case of three attributes.

19. Explain- (i) Time series data (ii) Cross section data
 (iii) Discrete data (iv) Continuous data
20. Given $N = 2500$, $(A) = 420$, $(B) = 670$ and $(AB) = 85$.
 Calculate the missing values and also calculate Yule's coefficient of association.

Attempt the following-(10 marks each)- Unit-II

1. (i) Write procedure of drawing - (a) Histogram when classes are of equal and of unequal width (b) Frequency Polygon.
- (ii) What are requisites of good measure of central tendency?
2. Discuss merits and demerits of (i) Arithmetic mean (ii) mode.
 (i) Define- (a) Decile (b) Percentiles (c) Quartiles
- (ii) How will you obtain median for ungrouped and grouped data.
3. (i) What are demerits of geometric mean.

(ii) Following data represent age distribution of 60 boys.

Age in years	10-12	12-14	14-16	16-18	18-20	20-22
No. of boys	5	9	15	17	10	4

- (a) Obtain cumulative frequency distribution less than type.
 (b) Obtain 7th decile
 (c) Thirty eighth percentile.

4. Discuss merits and demerits of (i) Arithmetic mean (ii) Median
5. Describe the procedure of drawing
 (i) Frequency curve and
 (ii) less than and greater than ogives.
6. (i) Write a short note on stem and leaf diagram.
 (ii) Discuss properties of arithmetic mean.
7. (i) Define harmonic mean. State demerits of harmonic mean.
 (ii) Sales of a company increased by 5% in first year, 7% in second year and 10% in the third year. Find average increase in percentage of sales using appropriate average.
8. (i) Explain how median can be obtained for - (a) Raw Data (b) Ungrouped frequency distribution (c) Grouped frequency distribution.
 (ii) What are the requirements of a good average?

9. Calculate median, 7th decile and 92nd percentile for the following data representing monthly food expenditure of 100 families.

Food expenditure in Rs.	1000-1300	1300-1600	1600-1900	1900-2200	2200-2500
No. of families	20	25	35	15	5

10. Write procedure of drawing –
 (i) Histogram when classes are of equal and unequal width
 (ii) Frequency polygon

11. (i) What are demerits of Harmonic mean?
 (ii) Following data gives marks of 100 students in a certain test.

Marks	10-12	12-14	14-16	16-18	18-20	20-22	22-24
No. of students	11	17	20	22	10	10	10

Obtain arithmetic mean and mode for the data.

12. (i) State merits and demerits of geometric mean.
 (ii) Prepare a frequency distribution of daily wages (in Rs.) by taking class intervals 30 – 40, 40 – 50, 50 – 60, 60 – 70, 70 – 80 etc.

85	69	72	77	48	50	99	90	70	32	76
75	68	31	39	61	89	71	82	78	87	41
88	40	66	42	93	53	97	92	43	95	71
36	56	54	60	59	80	52	67	55	47	37
47	86	46	44	68	81					

Hence find –(a) cumulative frequency less than, greater than or equal to type
 (b) relative frequency
 (c) harmonic mean.

13. (I) Define – (i) Arithmetic mean (ii) Combined mean (iii) Weighted Arithmetic mean.

(II) Derive relation between Arithmetic mean, Geometric mean, Harmonic mean for two positive numbers a, b.

14. (I) Explain how mode can be calculated from (i) raw data (ii) grouped frequency Distribution (iii) Histogram

(II) State the empirical relation between mean, mode and median. When do you use it?

15. Explain construction and use of (i) Histogram (ii) Stem and leaf diagram with illustration for each.

16. Discuss merits and demerits of any two measures of central tendency.
 17. Define – (i) Arithmetic mean (ii) Geometric mean (iii) Harmonic mean.
 For any two positive numbers a and b, show that $GM^2 = AM \times HM$

Attempt the following-(10 marks each)- Unit -III

1. (i) Define standard deviation. Discuss how it is affected by change of origin and Change of scale for grouped data
 (ii) Find first four central moments for the following observations 3, 8, 0, 7, 2
2. (i) What is dispersion? State all absolute and relative measures of dispersion.
 (ii) Suggest a suitable measure of dispersion if frequency distribution includes open end classes.
3. (i) Explain the need of relative measure of dispersion.
 (ii) State merits and demerits of variance as a measure of dispersion.
4. (i) Explain the term 'Kurtosis'. Explain different types of kurtosis.

(ii) Obtain Bowley's measure of skewness for the following data.

Height (in inches)	59-61	61-63	63-65	65-67	67-69
No.of students	4	30	45	15	6

5. (I) (i) Define:-raw moments, central moments.
 (ii) Derive expressions for first four central moments in terms of raw moments about origin zero
- (II) For a distribution , mean is 10, S.D. is 3, $\beta_1 = 1$, $\beta_2 = 3.5$.
 Obtain first four central moments.
6. (I) (i) Explain concept of kurtosis and discuss types of kurtosis.
 (ii) State measure of kurtosis.
- (II) First three raw moments about zero are 0.4, 1, 0.64.
 Find measure of skewness and comment.
- (i) Why corresponding to every absolute measures of dispersion there is need to define relative measures of dispersion?
 (ii) State expressions of various absolute and relative measures of dispersion.
 Calculate Karl Pearson's coefficient of skewness, Bowley's coefficient of skewness for the following data representing marks obtained by 200 students.

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No.of students	10	40	20	50	10	40	16	14

9. (i) Explain concept of skewness and discuss types of skewness. State various measures of skewness.
- (ii) For a certain distribution mean is 50, coefficient of variation is 24% and coefficient of skewness is 0.4. Find mode and median.
10. (i) Define raw moments and state their properties.
- (ii) Explain role of measures of dispersion in comparing two data sets.
- (iii) What are different relative measures of dispersion?
11. (i) Define standard deviation. Discuss the properties of standard deviation.
- (ii) First four raw moments for a variable are -1.5, 17, -30, 308. Find β_1 and β_2 .
12. The distribution of length of illness (in days) for a certain type of disease is given below:

Length of illness	0-14	14-28	28-42	42-56	56-70
No. of patients	5	17	17	7	7

Calculate

- (i) Average length of illness and corresponding mean deviation about Median
- (ii) Coefficient of quartile deviation.
- (iii) Bowley's measure of skewness.
13. . Explain the concept of dispersion. Define quartile deviation and standard deviation. State their merits and demerits.
14. . (i) What is kurtosis? Explain different types of kurtosis. State the measure of kurtosis.
- (ii) Given $\mu_1 = 0$, $\mu_2 = 40$, $\mu_3 = -100$, $\mu_4 = 200$, find β_1 and β_2 .
15. . Define raw moments about origin zero and central moments. Derive relationship between first four central moments and raw moments about origin zero.
16. . (i) Obtain mean deviation about mean and coefficient of mean deviation about mean for the following data on time required to complete typing test.

Duration in minutes	11	12	13	14	15	16
No. of candidates	7	19	25	23	15	11

- (ii) What is the effect of change of origin and scale on raw moments?

17. Discuss merits and demerits of (i) Quartile deviation
(ii) Standard deviation as measures of dispersion
18. . State expressions for all absolute and relative measures of dispersion. Explain utility
of relative measures of dispersion.
19. . Obtain Karl Pearson's coefficient of skewness for the following data.

Daily Wages	70 - 90	90 - 110	110 - 130	130 - 150	150 - 170
No. of workers	16	22	36	18	8

20. . Explain the concept of dispersion. Define quartile deviation and mean absolute
deviation. State their merits and demerits.
21. (i) Write a note on kurtosis in the data.
(ii) Find first three central moments for following observations: 1, 2, 3, 4, 5
22. . A survey conducted to determine distance travelled (in kms.) per litre of petrol by
newly introduced motorcycle gives the following distribution.

Distance (in kms.)	40 - 45	45 - 50	50 - 55	55 - 60	60 - 65
No. of motorcycles	10	17	23	40	10

Obtain standard deviation and coefficient of variation.

23. What is dispersion? Discuss different relative and absolute measures of dispersion.
24. Write note on raw and central moments and its relationship.

Attempt the following-(5 marks each)- Unit -I/II/III

1. Explain the terms (i) Population (ii) Sample (iii) Qualitative data
(iv) Time series data (v) Discrete data.
2. Distinguish between a questionnaire and a schedule.
3. Explain the construction of Box and Whisker plot.
4. Discuss the effect of change of origin and change of scale on standard deviation for raw data.
5. State the guidelines to decide number of classes and width of class interval.
6. Arithmetic mean and standard deviation of 25 observations was obtained as 45 and 7
respectively. Later it was understood that one observation was wrongly taken as 45 instead of
54. Obtain correct arithmetic mean and standard deviation.

7. An analysis of wages paid to workers, two companies from the same industry gave the following results-

	Company A	Company B
No. of workers		650
Average wages	550	450
S.D. of wages	9	10

- (i) Which company pays out larger amount as wage bill?
 (ii) Which company has greater variability in wages?

8. Explain the terms (i) cross sectional data (ii) simple random sampling (iii) Quantitative data (iv) Discrete data (v) Continuous data.
9. Explain the procedure to draw less than ogive for a data. State uses of ogive curves.
10. Discuss algebraic properties of arithmetic mean.
11. The arithmetic mean and standard deviation of a group of 50 observations were 47 and 4 respectively. Later it was understood that one observation was wrongly taken as 54 instead of 45. Obtain correct arithmetic mean and standard deviation.
12. Represent the following information in a suitable tabular form.
 The number of employees in 'ABC computer limited' in 2005 was 850 of which 40% were ladies and rest were gents. In 2006, the number of gents increased by 32 and the number of ladies increased by 46 with reference to 2005. In 2007, the total number of employees increased by 25% as compared 2006, while the increase in the number of ladies was 20 more than the increase in the number of gents.
13. Explain the terms (i) dichotomous classification (ii) simple random sampling (iii) Qualitative data (iv) time series data.
14. Write a note on stem and leaf diagram.
15. Discuss the effect of change of origin and scale on standard deviation.
16. If a and b are any two positive numbers then prove that
 Arithmetic mean \geq Geometric mean \geq Harmonic mean.
17. A survey in a certain locality revealed that that out of 300 persons considered, 110 were attacked by a disease. Out of 300 persons, 160 had been vaccinated and of those only 50 were attacked by the disease. Calculate the coefficient of association between vaccination and attack of the disease. Comment on the value of the coefficient of association.
18. In a market survey of 25000 women from Mumbai, exactly one fourth were non-Maharashtrians. 15 out of every 25 women were employed and 70% of Maharashtrians were employed. 60% of employed Maharashtrian women and 20% of employed non-Maharashtrian were married. The number of married Maharashtrian women who were unemployed was 3250 and 1435 unmarried non-Maharashtrian were unemployed. Tabulate the information and give a suitable title.

19. The following data pertain to workers of two factories.

	Factory A	Factory B
No. of workers	100	200
Mean time of completing the job	40	42
Standard deviation (minutes)	8	6

- (i) Obtain combined mean
- (ii) Workers of which company are more efficient?

20. Represent following set of 20 observations using stem and leaf diagram.
20, 30, 50, 42, 48, 37, 28, 32, 54, 36, 41, 53, 29, 27, 36, 30, 44, 33, 39, 51
21. Discuss the effect of change of origin and scale on central moments.
22. Define deciles and percentiles. Explain procedure to obtain percentiles for a grouped data. Obtain combined arithmetic mean and combined standard deviation for the following data.

23.

Section	No. of workers	Mean wages	Standard Deviation
A	60	120	7
B	90	115	8

24. What is primary data? In what respect it is different from secondary data?
25. How is association between two attributes measured? State conditions for consistency of three attributes.
26. The mean annual salary paid to all the employees in a company was Rs.50000. The mean annual salaries paid to male and female employees were Rs.52000 and Rs. 42000 respectively. Find the percentage of males and females employed in the company.
27. State any two merits and two demerits of median.
28. The mean and standard deviation of 100 values are 10 and 5.1 respectively. It was later discovered that one observation 40 was misread as 50. Find correct mean and S.D.
29. Draw Histogram for the following data. Hence calculate mode.

Sales I in (0,00 Rs.)	0-10	10-20	20-30	30-40	40-50
No. of days	18	9	33	6	15

30. Derive relation between Yule's coefficient of association and coefficient of colligation.
31. State any two merits and two demerits of mode.
32. Distinguish between (i) Ordinal scale and nominal scale (ii) Qualitative and quantitative data.

33. Out of 200 students in a class consisting of 60% boys, 60 wear glasses and 150 passed in a test. Of the boys, 80 passed in the test including 30 wearing glasses. In all 40 boys wear glasses and among the students wearing glasses 48 passed. Tabulate the information and give suitable title.

34. Find the missing frequencies if the median of the following 229 observations is 45.

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80
frequency	12	30	?	65	?	25	18

	Boys	Girls
Number	72	38
Mean Height (inches)	68	61
S.D. (inches)	2.4	3

Consider above information to find:

- (i) Combined standard deviation of heights
- (ii) Whose heights are more consistent?
