

*ANEKANT EDUCATION SOCIETY'S
TULJARAM CHATURCHAND COLLEGE OF
ARTS, SCIENCE AND COMMERCE,
BARAMATI
AUTONOMOUS*

QUESTION BANK

FOR

F.Y.B.Sc(Computer Science)

STATISTICS

CSST- 1101: STATISTICAL METHODS - I

(With effect from June 2019)

Unit 1: Presentation of Data

Questions for 1 mark

MultipleChoice Questions

1. A qualitative characteristic like nationality, religion, beauty , blood group is called as _____
a) variable b) attribute c) constant d) none of these
2. A quantitative characteristic like weight of person, examination marks, population of country is called as _____
a) variable b) attribute c) constant d) none of these
3. A variable taking only particular values is called _____
a) continuous variable b) attribute c) constant d) discrete variable
4. A variable taking all possible values in a certain range is called _____
a) continuous variable b) attribute c) constant d) discrete variable
5. Frequency density is the ratio of _____
a) class frequency and class width b) class frequency and mid value
c) class width and class frequency d) class frequency and mid value
6. The difference between upper boundary and lower boundary is _____
a) class width b) class limit c) class mark d) class interval
7. The two numbers designating the class interval are called as _____
a) class width b) class limit c) class mark d) class interval
8. A class in which one of the limits is not specified is called an _____
a) class width b) class boundary c) open end class d) class interval
9. The number of observations belonging to a class intervals is called as
a) Cumulative frequency b) Class width
c) Class mark d) Class Frequency
10. Draw histogram, height of rectangle for unequal class width is proportional to
a) Class frequency c) Class width
b) Frequency density d) Mid values
11. Less than type cumulative frequencies are
a) Non increasing c) Non decreasing
b) Constant d) Not of any specific nature
12. In exclusive types of classification
a) Lower limits are excluded c) Upper limits are excluded
b) Both the limits are excluded d) None of limits is excluded
13. Less than cumulative frequency of the
a) First Class is 0. c) Last Class is 0.
b) First class is the total frequency. d) Last class is the total frequency.

14. If the classes are as follows: 0-9,10-19,20-29,... then the class widths are
 a)9 b)10 c)9.5 d)10.5
15. Histogram helps to get an idea
 a) about the spread of the distribution.
 b) about the shape of the distribution.
 c) about the pattern of the variation.
 d) all of the above.
16. A relative frequency distribution represents frequencies in terms of.....
 a) Proportions b) whole numbers c) percentages d) both (a) and (c)
17. Which of the following statements is true for less than cumulative frequency curve ?
 a) The curve is non-decreasing in nature.
 b) The curve is strictly decreasing in nature.
 c) The curve is non-increasing in nature.
 d) The curve is strictly increasing

Questions for 2marks.

Define the following terms with illustration

- | | |
|----------------------|------------------------------------|
| 1. Statistics | 8. Mid value |
| 2. Attribute | 9. Class boundaries |
| 3. Variable | 10. Class limits |
| 4. Discrete variable | 11. Less than cumulative frequency |
| 5. Class frequency | 12. More than cumulative frequency |
| 6. Open end classes | 13. Inclusive method |
| 7. Class width | 14. Exclusive method |

Questions for 4 marks.

1. Describe the procedure of histogram
2. Distinguish between – Inclusive classes and Exclusive classes
3. From the following data construct histogram.

Marks	0-20	20-40	40-60	60-80	80-100
No.of students	2	18	42	28	5

4. From the following data construct frequency curve.

Marks	0-10	10-20	20-30	30-40	40-50
No.of students	2	18	42	28	5

5. From the following data construct ogive curve.

class	0-100	100-200	200-300	300-400	400-500
frequency	20	180	420	280	50

6. From the following data construct frequency polygon.

class	100-200	200-300	300-400	400-500
frequency	150	250	200	50

7. Draw the histogram for the following data

class	10-20	20-25	25-40	40-50
frequency	15	25	20	5

8. Draw the histogram for the following data

class	10-19	20-29	30-39	40-49
frequency	10	20	15	3

9. Exhibit the following data by using Stem & leaf chart.

10.8 12.9 13.0 12.5 13.1 10.4 10.2 13.3 11.8 11.1
11.3 11.6 11.2 11.1 11.3 11.6

10. Exhibit the following data by using Stem & leaf chart.

52,45,78,69,32,56,45,78,25,69,45,74,54,66,85,25,52,34,75,50
Which stem has minimum leaves?

11. Exhibit the following data by using Stem & leaf chart.

320,351,326,345,3321,331,326,335,326,316,334,329,338,317,339,324,325,326
Which stem has maximum leaves?

12. Explain the Stem & leaf chart with illustration.

13. From the following stem and leaf plot find (i) the lowest observation, (ii) the highest observation.

Stem	Leaf
4	2, 3, 4
5	1, 1, 2, 2, 4, 5, 8, 9
6	2, 3, 8, 9
7	0, 5, 7

Key : "5 7" stands for 57

Also write the entire data set

Questions for 6 marks

- Explain the construction of following graphs along with the rough sketches.
i) Frequency curve ii) Frequency polygon
- Explain the construction of Ogives along with the rough sketch.

Unit No. 2: Measures of Central Tendency

Questions for 1 mark

A) Multiple Choice Questions

1. A measure of central tendency based on all observations is _____
a) mean b) mode c) median d) none of these
2. The median of values 27, 32, 18, 21 and 30 is _____
a) 27 b) 32 c) 21 d) 30
3. The second quartile is also known as _____
a) Mean b) Mode c) Median d) H.M.
4. The arithmetic mean of first 'n' natural number is _____
a) n b) $n(n+1)$ c) $(n+1)/2$ d) $2n$
5. The mean of values 8,16,13,14,9 and 11 is _____
a) 10.11 b) 11.8333 c) 10.8333 d) 11.10
6. The number of quartiles is _____
a) 2 b) 3 c) 4 d) 5
7. The number of percentiles is _____
a) 100 b) 9 c) 99 d) 10
8. The number of deciles is _____
a) 10 b) 9 c) 100 d) 99
9. The mode of values 15,14,11,13,16,11,12,16,17,11 _____
a) 11 b) 15 c) 13 d) 16
10. Median can be obtain graphically by means of _____ curve
a) less than ogive curve b) more than ogive curve
c) frequency curve d) frequency polygon
11. For deciding most favorite actor, which is the most appropriate average?
(a) Mean (c) Mode
(b) Median (d) Upper quartile
12. The third quartile divides the series in the ratio
a) 1 : 3 b) 3 : 1 c) 1 : 1 d) 4 : 1
13. Which one of the following is false?
a) $Q2 = \text{Median}$ b) $Q2 = 50\text{th Percentile}$
c) $Q2 = 25\text{th Percentile}$ d) $Q2 = 5\text{th decile}$
14. Median =
a) $Q1$ b) $D5$ c) $P10$ d) $Q3$
15. With the help of histogram, one can determine
a) Mode b) Deciles c) Percentile d) Median
16. Which of the following is used to show mode graphically?

12. Find A. M. given that $\sum(x - 10) = 230$ and $n = 50$
13. Find the A. M of the given observations: 10 , 12 , 13 , 10 , 11 , 13 , 10
14. Find the median of the given observations: 10, 12, 14, 12, 9, 8, 7, 10
15. If $\sum X_i = 525$ A. M. =25 then $n =$
16. What is mean by partition values?

Questions for 4 marks

1. State and prove the properties of A.M.
2. Find A.M. of first n natural numbers.
3. State requirements of a good measure of central tendency.
4. Write merits and demerits of A. M.
5. Discuss merits and demerits of Median.
6. Discuss merits and demerits of Mode.
7. Represent the following data using Box plot.
15, 11, 11, 28, 22, 14, 13, 18, 30, 12, 25
17. Define mode. Explain the procedure of locating mode graphically
18. Define median. Explain the procedure of locating median graphically
19. A variable takes values 1, 4, 9, ... n^2 . Find its A.M.
20. A variable takes values 1, 4, 9, ... n^2 with frequencies 1, 2, 3, ..., n respectively. Find its A.M.
21. A variable takes values 7, 10, 13, ... (50 terms), find its A.M. and median
22. Mean daily salary of 50 employees in a firm is Rs. 88.40 Frequency distribution of salaries of these employees in which some frequencies are missing, find it.

Salary	40 – 60	60 – 80	80 – 100	100 – 120	120 – 140
Frequency	6	-	17	-	5

23. If the Q_2 of the following distribution is 44 and total frequency is 100. Find the missing frequencies

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80
-------	-------	-------	-------	-------	-------	-------	-------

Frequency	5	12	--	20	--	10	4
-----------	---	----	----	----	----	----	---

24. The daily expenditure of 120 families is given below:-

Daily expenditure	20-29	30-39	40-49	50-59	60-69
No. of families	20	-	28	-	15

If the mode of the frequency distribution is 43. Find the missing frequencies.

25. Arithmetic mean of weight of 100 boys is 50 kg and arithmetic mean of 50 girls is 45 kg. Calculate A. M. of combined group of boys and girls.

26. Calculate the weighted mean of the first n natural numbers with the numbers themselves as weights.

27. Arithmetic mean of the following frequency distribution is 5, find the value of X .

Variable	Frequency
2	$X - 1$
4	$X + 1$
6	$X + 1$
8	$2X - 5$

28. Find the combined mean of the following data

Group I	$\bar{x}_1 = 2100$	$n_1 = 100$
Group II	$\bar{x}_2 = 1500$	$n_2 = 200$

29. The weekly salary paid to all employees in a certain company was Rs. 444. The arithmetic mean of weekly salary paid to the male and female employees were Rs. 480 and Rs. 360 respectively. Obtain the percentage of male and female employees in the company.

30. The values 3, 5 and 7 are assigned weights $(K - 4)$, $(K - 2)$ and $(K + 1)$ respectively. If the weighted arithmetic mean is 6, find K .

Questions for 6 marks

- Define A.M., Median, Mode and state the formula for each, in case of individual observations and frequency distribution
- Discuss merits and demerits of arithmetic mean, median, mode

- Define Quartiles, Deciles, and Percentiles. Give the formula each for frequency distribution.
- Find the arithmetic mean, median, mode, quartiles and quartile deviation of the following observations
61, 62, 63, 62, 63, 62, 64, 64, 60, 65, 60, 67, 66, 60
- Find arithmetic mean, median, mode, lower quartile and upper quartile of the following data

Weekly Wages(Rs)	50	55	60	65	70
No. of workers	3	7	20	6	4

- For the following frequency distribution

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	5	12	43	32	8

Find i) Mean ii) Mode iii) Median

- The monthly profit in rupees of 100 shops is distributed as follows :

Profit (in Rs.) per shop	Below100	100 -200	200 - 300	300 - 400	400 - 500	Above500
No. of Shops	12	18	27	20	17	6

- Calculate the median and mode for above data.
 - Find median and mode graphically.
- For the following frequency distribution find the median, seventh decile and eighty-fifth percentile.

Marks	0-100	100-200	200-300	300-400	400-500	500-600	600-700
No. of students	9	15	18	21	18	14	5

13. Which one of the following relation between standard deviation (S.D.) and variance $\text{Var}(X)$ of a variable X is true ?
- S.D. = $\text{Var}(X)$
 - $(\text{S.D.})^2 = \text{Var}(X)$
 - S.D. = $[\text{Var}(X)]^2$
 - S.D. = $[\text{Var}(X)]^3$
14. In standard deviation, the deviations of the items x_i are always taken from _____
- Mean
 - Median
 - Mode
 - Zero
15. _____ is used when only a rough measure of dispersion is required.
- Mean Deviation
 - Standard Deviation
 - Quartile Deviation
 - Range

B) Define the following terms

- Range
- variance
- Standard deviation
- C.V.
- Quartile deviation
- coefficient of quartile deviation

Questions for 2 marks

- State the properties of variance.
- State the Merit and demerit of range.
- State the Merit and demerit of quartile Deviation
- State the formula of variance for a grouped frequency distribution.
- State the formula of standard deviation for a grouped frequency distribution.
- State the formula of combined variance.
- State the formula of combined standard deviation.
- Find the variance for the given observations: 10, 8, 9, 10, 8, 7, 11
- Find the standard deviation of the following observation: 2, 3, 5, 2, 7, 5, 7, 6, 11, 12.
- Find the range of the following observation:
10, 100, 12, 1, 55, 20, 26, 111, 120, 12, 15, 14, 11, 10, 15, 15, 16, 12, 13, 130, 100, 52
- Find the coefficient of range of the following observation:
110, 10, 12, 11, 55, 20, 26, 1, 120, 12, 15, 14, 11, 10, 15, 15, 16, 12, 13, 10, 110, 520

Questions for 4 marks.

1. Explain the term dispersion. State different measures of dispersion.
2. State requirements of a good measure of dispersion
3. Write Merits and demerits of variance
4. Define quartile deviation .State its Merits and demerits.
5. Distinguish between absolute & relative measures of dispersion
6. A variable takes values $1, 2, \dots, n$. find its mean and variance
7. A variable takes values $a, a + d, a + 2d, \dots, a + (n - 1) d$; find its variance.
8. A variable takes values $1, 2, \dots, n$ with frequencies $1, 2, \dots, n$ respectively. Find the standard deviation.
9. Find the standard deviation and the coefficient of variation from the following data :

Marks	0-10	10-20	20-30	30-40	40-50
Frequency	10	16	30	32	12

10. If $n=100, \Sigma X = -20, \Sigma X^2 = 220$. Find S.D. & C.V.
11. Coefficients of variation of the two series are 60% & 80%. Their standard deviations are 20 & 16. What are their arithmetic means?
12. Two samples of sizes 40 & 50 have the same mean but different standard deviations 19 & 8 respectively. Find the standard deviation of the combined group.
13. Calculate the coefficient of variation (C.V.) for the following data :

Class	1-10	11-20	21-30	31-40	41-50	51-60
Frequency	5	9	15	21	6	4

14. Find combined standard deviation from the following data :

Workers	Number	Average Salary	Standard Deviation
Male	80	1520	06
Female	20	1420	05

15. The following results are available

	Factory	
	A	B
Mean wages (Rs./m)	3000	4000
S.D of wages (Rs./m)	27	56

In which factory wages show more variability? Justify your answer.

Questions for 6 marks

- State the merits and demerits of each of the following measures of dispersion: (i) range (ii) variance (iii) quartile Deviation
- Weight of group of students with weight (in k.g.) is given below
10, 15, 52, 81, 50, 41, 53, 14, 54, 57, 80, 57, 73

Compute -i) Range ii) Coefficient of Range iii) Standard Deviation

iv) Inter-Quartile Range v) Quartile Deviation

- Explain relative & absolute measures of dispersion.

- For the following frequency distribution,

size of item	2	4	6	8	10	12
frequency	6	10	20	24	12	7

Compute – i) range ii) Coefficient of Range iii) standard deviation iv) Quartile deviation

- For the following frequency distribution,

Marks	40-50	50-60	60-70	70-80	80-90	90-100
frequency	2	4	16	8	10	10

Compute –

- Range ii) Coefficient of Range iii) Standard Deviation iv) Quartile deviation
- Two automatic tea filling machines A and B tested for the performance. Machines are supposed to fill 500 gm tea in each packet. A random sample of 100 filled packets on each machine showed the following distribution.

Weight in gm	Frequency A	Frequency B
485-490	12	10
490-495	18	15
495-500	20	24
500-505	22	20
505-510	24	18
510-515	4	13

Which machine is more consistent? Why?

7. Information regarding daily salaries of two companies A and B is given below :

	Company A	Company B
No. of workers	600	400
Mean salary	Rs. 180	Rs. 200
S.D. of salary	Rs. 9	Rs. 10

- (i) Which company pays larger salary? Why?
- (ii) Which company has less variation in salaries? Why?
- (iii) Find combined mean and S.D. of two firms A and B.

Unit 4: Moments Skewness and Kurtosis

Questions for 1 marks

Multiple Choice Questions

- For a negatively skewed distribution, the relationship between mean, median and mode is _____
a) mean > median > mode b) mean > mode > median
c) mean = mode = median d) mean < median < mode
- Which one of the following is a raw moment?
a) Moment about a
b) Moment about 4
c) Moment about zero
d) Moment about mean
- The first order moment about origin is equal to _____
a) Zero b) One c) Three d) Mean
- For a positively skewed distribution the relationship between mean, median and mode is _____
a) mean > median = mode b) mode < median < mean
c) mean = mode < median d) mean < mode < median
- First central moment of data is _____
a) 7 b) 0 c) 1 d) 14
- The second central moment is also known as _____
a) variance b) mean c) mode d) second quartile
- For mesokurtic distribution, the value of β_2 is
a) Equal to 3 b) Greater than 3 c) Less than 3 d) Equal to 0
- The relation between β_2 and γ_2 is _____
a) $\beta_2 = \gamma_2$ b) $\beta_2 - 3 = \gamma_2$ c) $\beta_2 = \gamma_2 - 3$ d) $\lambda_2 = \sqrt{\beta_2}$

20. For a negatively skewed distribution

- a) left tail is elongated
- b) right tail is elongated
- c) both tails are equally elongated
- d) nothing can be said about tail

Questions for 2 marks

1. A distribution has mean 30, coefficient of variation 20% and coefficient of skewness 0.3, find its mode
2. Given that A.M. = 160, mode = 157, SD = 50, find Karl Pearson's coefficient of skewness.
3. The first two raw moments of a frequency distribution are 1 and 10 respectively. Find the variance of the distribution.

Questions for 4 marks.

1. Define r^{th} raw and r^{th} central moments and state the first four central moments in terms of the raw moments.
2. The first three moments of distribution about the value 1 are 2, 25 and 80 respectively. Find its mean, standard deviation and third central moment.
3. Define 'skewness'. What are the types of skewness?
4. Explain different measures of skewness.
5. Explain concept of kurtosis. State its types with help of frequency curve.
6. Find the C.V. of a frequency distribution given that its mean is 120, mode is 123 and Karl Pearson's coefficient of skewness is -0.3.
7. The standard deviation of a distribution is 5. What should be the value of fourth central moment so that distribution will be (i) mesokurtic (ii) leptokurtic?
8. Compute coefficient of skewness from the following information :
Upper quartile = 50, Lower quartile = 40, Median = 45. Comment on the result.
9. For a moderately skewed distribution the mean is 200, coefficient of variation is 8 and Pearson's coefficient of skewness is 0.3. Find the mode and median
10. Given that $X = 1$, $\mu_2 = 3$, $\mu_3 = 0$, $\mu_4 = 27$, find first four raw moments.
11. For a frequency, Bowley's coefficient of skewness is 0.6. The sum of first and third quartiles is 100 and the median is 38, Find the two quartiles.

Questions for 6 marks

1. The first four moments of a distribution about the value '5' are 2, 20, 40, and 200 respectively. Find the first four central moments.
2. A distribution has mean 30, coefficient of variation 20% and coefficient of skewness is 0.3. Find its mode.
3. In a certain frequency distribution the sum of upper and lower quartiles is 45 and the difference between them is 15 . If the median is 20, find the coefficient of skewness
4. Define Karl-Pearson's coefficients γ_1 and γ_2 and discuss their utility in Statistics.
5. Given that $n = 100$, $\Sigma x^2 = 140$, $\Sigma x^3 = -40$, $\Sigma x^4 = 560$. Find β_1 , β_2 and comment on the nature of skewness and kurtosis of the distribution.
6. Given that A.M. = 160 , mode = 157, $\sigma = 50$, find (i) Karl Pearson's coefficient of skewness (ii) median (iii) coefficient of variation

Unit 5: Correlation (for bivariate raw data)

Questions for 1marks

A) Multiple Choice Questions

- The extent of linear relation between the two variables is called as _____
a) Correlation b) covariance c) skewness d) kurtosis
- If the correlation between X and Y is 0.5, then the correlation between the variables X and (-Y) is _____
a) -0.5 b) 0.5 c) 0 d) 1
- Spearman's rank correlation coefficient lies between _____
a) 0 to 1 b) -1 to 1 c) 0 to ∞ d) $-\infty$ to ∞
- $\text{Cov}(X, X) =$ _____
a) $v(x)$ b) 0 c) 1 d) $-\infty$
- $\text{Cov}(X, \text{Constant}) =$ _____
a) $v(x)$ b) 0 c) 1 d) $-\infty$
- Karl Pearson's coefficient of correlation lies between
a) 0 to 1 b) -1 to 1 c) 0 to ∞ d) $-\infty$ to ∞
- If $\text{Corr}(X, Y) = 0$ then we conclude that
a) $\sigma_x = \sigma_y$
b) $\bar{X} = \bar{Y}$
c) There is no relationship between X and Y
d) None of the above.
- Correlation measures the extent of
a) Parabolic relation between two variables.
b) Linear relation between two variables.
c) Logarithmic relation between two variables.
d) Exponential relation between two variables

Questions for 2 marks.

1. Define the term correlation.
2. What is bivariate data?
3. State the merits of scatter diagram.
4. Discuss demerits of scatter diagram.
5. Which are measures of correlation?
6. Define covariance.
7. Prove that $\text{Cov}(X,X)=\text{Var}(X)$.
8. Given $X = 53$, $Y = 28$, $b_{yx} = -1.5$ & $b_{xy} = -0.2$. Find r
9. Explain the term :rank
10. Explain the term :Karl Pearson's coefficient of correlation
11. State any two properties of Karl Pearson's coefficient of correlation.
12. State the properties of covariance.
13. When two variables said to be correlated? Give two examples.
14. What do you mean by positive correlation? Give two examples.
15. What do you mean by negative correlation? Give two examples.
16. If $r = 0$. Comment on it.
17. If $r = 0 \rightarrow$ covariance = 0 comment.
18. If covariance = 0 $\rightarrow r = 0$ comment.
19. If the correlation co-efficient between two variables x & y is positive then
20. comment on the following –

The correlation coefficient between $-x$ & $-y$ is positive.

The correlation coefficient between $-x$ & y or x & $-y$ is positive.

Questions for 4 marks.

1. Explain the concept of correlation for a bivariate data and its types.
2. Define positive correlation and negative correlation with an illustration each.
3. Consider the following bivariate data :

X	6	2	10	4	8
Y	9	11	5	8	7

Compute Rank correlation coefficient between X and Y.

4. Find correlation coefficient (r) between X and Y given the following data
 $n = 7, \sum X = 119, \sum X^2 = 2833, \sum Y = 87, \sum Y^2 = 2385, \sum XY = 521$.
5. Find number of pairs of observations from the following data
 $r = -0.4, \sum X = 100, \sum X^2 = 2250, \sum Y = 100, \sum Y^2 = 2250, \sum XY = 1900$

Questions for 6 marks.

1. The following table shows ages (X) and the systolic blood pressures (Y) of 9 women.

Age (X)	45	52	53	61	54	48	49	50	51
Blood pressure (Y)	120	131	136	150	132	124	130	137	135

.Calculate the correlation coefficient between X and Y

2. Define Karl Pearson's coefficient of correlation & show that it lies between -1 & 1.