MIT Arts, Commerce and Science College, Alandi, Pune

DEPARTMENT OF STATISTICS

Question Bank

Statistical Methods-I

Questions for 2 marks

Q1 Define the following terms:

a. Class limits
b. Class mark
c. Class width
d. Class boundaries
e. Frequency of a class
e. Relative frequency
f. Exclusive class interval
g. Attribute

Q2 Explain the following terms:

a. Raw data
b. Variable
c. Discrete variable
d. Continuous variable
d. Cumulative frequency
e. Open End Class

Q3 Explain the concept of central tendency with an illustration.

Q4 State the empirical relation among mean mode and median. Hence find mode if mean is 156 and median is 150.

Q5 Define coefficient of variation.

Q6 In each of the following, state relationship among mean, mode and median:

i. Positively skew distribution
ii. Negatively skew distribution

Q7 In each of the following, state relationship among quartiles:

i. Positively skew distribution
ii. Negatively skew distribution

Q8 If the mean and standard deviation of a distribution are 60 and 5 respectively find coefficient of variation.

Q9 State the probability mass function of each of the following distribution.

a. Discrete uniform distribution
b. Binomial distribution
c. Poisson distribution
d. Geometric distribution

Q10 State the additive property of binomial/ Poisson distribution.

Q11 State the type of correlation for the following:

i. Sale of woolen garments and day temperature.

ii. Supply and price of onion.

iii. Speed of a vehicle and distance required to stop the vehicle after applying the brakes.

iv. Income and expenditure of a family.

v. Amount of yield of grain and length of leaves.

vi. Height and blood pressure of a person.

Q12 Interpret the following values of Karl Pearson's coefficient of correlation:

i. $r = 0$	ii. $r = -1$	iii. $r = -0.8$
iv. $r= 0.6$	v. $r = +1$	vi. $r = 0.2$

Q13 The regression coefficient of Y on X is -1.5 and the regression coefficient of X on Y is -0.2, find corr(X,Y).

Q14 In a trivariate data the total correlation coefficients are:

$r_{12} = 0.58$, $r_{13} = 0.6$, $r_{23} = -0.95$. Are these values consistent?

Q15 Define Time Series.

Q16 State additive/ multiplicative model of Time Series.

Q17 State the equations of AR(1)/ AR(2) model.

Questions for 4 marks

Q1 Write a note on Stem and leaf plot.

Q2 The following are the marks obtained by 40 students in a test:

19, 50, 47, 34, 34, 19, 21, 24, 22, 23, 13, 37, 33, 36, 28, 33, 32, 26, 28, 12, 37, 31, 41, 42, 38, 12, 22, 29, 15, 17, 15, 32, 21, 31, 48, 17, 22, 49, 33, 28

Construct the stem and leaf chart.

Q3 Describe the stepwise procedure for drawing each of the following graphs:

a. Histogram b. Frequency curve c. Cumulative frequency curves

Q4 Describe the stepwise procedure for determining the value of:

a. Mode b. Median c. Upper Quartile/ Lower Quartile

Q5 What is mean by central tendency? What are the requisites of a good measure of central tendency?

Q6 State the merits and demerits of each of the following

a. Arithmetic men b. Median c. Mode

Q7 Find median for the following frequency distribution.

Expenditure (in '000 Rs.)	20-29	30-39	40-49	50-59	60-69
No. of Families	14	21	27	15	12

Q8 Consider frequency distribution of 100 families:

Expenditure	0-20	20-40	40-60	60-80	80-100
Number of families	14	a	27	b	15

Find the values of a and b if the median is 50.

Q9 The median and mode of the following distribution are known to be 27 and 26.

Classes	0-10	10-20	20-30	30-40	40-50
Frequency	3	a	20	12	b

Find the values of **a** and **b**.

Q10 Define the following measures and State the formulae of each in case of ungrouped / grouped frequency distributions:

a. Arithmetic mean b. Median

c. Mode d. Quartiles

e. Standard deviation f. Variance

Q11 What is Box Plot? Draw box plot for the data set whose Q1 = 11.89 , Q2 = 15.87, Q3 = 31.25. Minimum of values = 5.02, Maximum of values = 46.02.

Q12 What are relative measures of dispersion? Explain how they are superior to the Absolute measures of dispersion.

Q13 Find the mean and variance of first n natural numbers.

Q14 For a set of 20 observations the mean and variance are found to be 60 and 25 respectively. While checking it was found that one of the observation was wrongly taken as 10 instead of 8. Find correct mean and standard deviation.

Q15 Consider the following information:

	Boys	**Girls**
Number	24	36

Mean height (in inches)	67	60
Variance	8	5

i. Which group boys or girls on an average are more heighted?

ii. Which group has more consistent heights? Justify your answer.

Q16 From the following data:

	Factory A	Factory B
Number of workers	350	380
Average daily wages (in Rs.)	72	69.50

i. Which factory pays out larger amount as wage?

ii. Calculate combined wage of workers of the two factories taken together.

Q17 Define raw and central moments of a frequency distribution.

Q18 Express the first four central moments in terms of raw moments.

Q19 Given that, mean =1, $\mu_2 = 3$, $\mu_3 = 0$, obtain the first three raw moments.

Q20 The first four raw moments of a distribution are 2, 20, 40 and 200 respectively. Find first four central moments.

Q21 Explain the concept of skewness of a frequency distribution. Discuss types of skewness with a help of diagram.

Q22 Write a short note on kurtosis.

Q23 In a frequency distribution, the coefficient of skewness based on the quartiles is 0.6. If the sum of the upper and lower quartile is 100 and median is 38, find the value of upper and lower quartile.

Q24 For a grouped frequency distribution mean, coefficient of variation and Karl Pearson's coefficient of skewness are 40, 45% and -0.5 respectively. Find the standard deviation, mode and median.

Q25 Given that β_1= 0.19,β_2 = 2.6, μ_2 = 1.2. Find the values of μ_3 and μ_4.

Q26 The standard deviation of a distribution is 3. What should be the value of fourth central moment if the distribution is:

i. Mesokurtic

ii. Platykurtic

Q27 Define distribution function of a discrete random variable. Also state its properties.

Q28 A discrete random variable has the following probability distribution:

X	0	1	2	3	4	5	6
P(X=x)	K	3K	5K	7K	9K	11K	13K

Find the value of K. Also find $P(X \geq 4)$.

Q29 The probability distribution of a random variable is:

$\mathbf{X_i}$	0	1	2	3
$\mathbf{P(X_i)}$	0.1	0.2	0.5	0.2

If Y = 3X +4, find

i. Probability distribution of Y
ii. Cumulative distribution function of Y.
iii. $P[(Y \leq 10)/(Y>4)]$

Q30 Define the following terms:

i. Expectation of a discrete random variable
ii. Variance of a discrete random variable

Q31 If X represents the total number of heads obtained, when a fair coin is tossed 3 times. Find :

i. The probability distribution of X.
ii. Mean of X.
iii. Median of X.

Q32 Define discrete uniform distribution. State its mean and variance.

Q33 Define binomial distribution. State its mean and variance. State the Poisson approximation to binomial distribution. Also state a real life situation.

Q34 Define Geometric distribution and state its mean and variance. Also give a real life situation where it can be used.

Q35 Suppose X~ B(n, p). If E(X) = 5 and Var(X) = 2.5, find the values of n and p. Also find $P(X \leq 1)$.

Q36 A hospital switchboard receives on an average of 4 emergency calls in a 10 minute interval. Find the probability that

i. There are at most two calls in a 10 minute interval.
ii. There is exactly one call in a 10 minute interval.

Q37 Let $X \rightarrow B(n=10, p=0.3)$. Find mean and standard deviation of X. Also find $P(X \leq 1)$.

Q38 A manufacturer produces IC chips, 1 % of which are defective. Find probability that in a box containing 100 IC chips, no defectives are found.(Use Poisson approximation to Binomial distribution)

Q39 Products produced by a machine has 3% defective rate. What is the probability that the first defective occurs in the fifth item inspected? Also find the expected number of defectives produced by a machine.

Q40 If 5% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 10 bulbs :

i. none is defective
ii. 5 bulbs will be defective.

Q41 Explain the concept of correlation for a bivariate data. Explain its types.

Q42 What is scatter diagram? Discuss how it is useful in deciding correlation between two variables in a bivariate data.

Q43 State the properties of correlation coefficient and regression coefficients.

Q44 Compute correlation coefficient between X and Y for the following data and interpret its value.

Statistics paper-I (X)	65	45	50	60	40
Statistics paper-II (Y)	70	35	60	50	40

Q45 Describe the least squares procedure of fitting of a regression line of Y on X.

Q46 Given the following information:

$n=20$, $\Sigma X=80$, $\Sigma Y=40$, $\Sigma X^2=1680$, $\Sigma Y^2=320$, $\Sigma XY=480$.

Obtain the equation of regression line of X on Y.

Q47 Given $X - 4Y = 5$ and $X - 16Y = -64$ are the regression lines, find

i. Correlation coefficient.
ii. The means of X and Y.

Q48 In regression study, two regression lines are $2X - 3Y + 6=0$ and $4Y - 5X +8 =0$. Find

i. The means of X and Y.
ii. The regression coefficient of X on Y.
iii. The regression coefficient of Y on X.

Q49 State the properties of correlation coefficient and regression coefficient

Q50 State the similarities and differences between correlation and regression of a bivariate data.

Q51 A certain population has logistic growth with equation

$$Y = \frac{60}{1+e^{180-0.09X}}$$ where Y= Population in lacs and X = Year

i. Estimate the population in the year 2014.

ii. Find the year when the population was half of the carrying capacity.

Q52 Describe the least squares procedure of fitting of an exponential curve of type $Y=ab^X$.

Q53 Explain the concept of multiple correlation coefficient in a trivariate data. State the expression for the multiple correlation coefficient $R_{2.13}$ in terms of total correlation coefficients.

Q54 Explain the concept of partial correlation coefficient in a trivariate data. State the expression for the multiple correlation coefficient $r_{23.1}$ in terms of total correlation coefficients.

Q55 For a trivariate data, $r_{12} = r_{13} = r_{23} = 0.3$. Find the values of $r_{13.2}$ and $R_{1.23}$.

Q56 For a trivariate data, $r_{12} = 0.59$, $r_{13} = 0.46$, $r_{23} = 0.77$. Compute the values of $r_{23.1}$ and $R_{3.12}$.

Questions for 8 marks

Q1 The first four raw moments of a distribution are 2, 5, 12 and 48 respectively. Find first four central moments. Also find the coefficient of skewness and kurtosis based upon moments and interpret their values.

Q2 **W**rite the stepwise procedure of fitting the curve $Y= a + b X + c X^2$ using the method of least square.

Q3 Describe the least squares method for fitting of multiple regression plane of X_1 on X_2 and X_3.

Q4 Following are the data on height (X_1) in cm, weight (X_2) in Kg, and age (X_3) in years for a group of 50 students

$\bar{X}_1 = 125$	$\bar{X}_2 = 42,$	$\bar{X}_3 = 15,$
$\sigma_1 = 3.2,$	$\sigma_2 = 4.1,$	$\sigma_3 = 1.7,$
$r_{12} = 0.72,$	$r_{13} = 0.45,$	$r_{23} = 0.32,$

Obtain the equation of plane of regression of X_2 on X_1 and X_3. Also estimate X_2 when

$X_1 = 120$ and $X_3 = 13$.

Q5 What is time series? Discuss the components of time series. Give one example for each.

Q6 Find the centered 4 yearly moving average from the following time series:

Years	**Sales (in Rs '00)**
2000	15
2001	20
2002	18
2003	17
2004	17
2005	26
2006	25
2007	22
2008	20
2009	29
2010	27
2011	24

Q7 Find three yearly and five yearly moving averages for the following data:

Year	1980	1981	1982	1983	1984	1985	1986
Sales	75	60	55	60	65	70	70

MAEER'S, MIT ARTS, COMMERCE AND SCIENCE COLLEGE, ALANDI (D), PUNE.

F.Y.B.Sc. (Comp. Sci.) Statistical Method –II

Question Bank

Chapter 1: Theory of Probability

1) Explain principles of counting.

2) Explain Deterministic and Non-Deterministic experiments.

3) Explain the following terms with one illustration each:

i) Mutually Exclusive Events. ii) Independence of two events.

iii) Sample space iv) Union of two events

v) Probability of an Event vi) Exhaustive Event

4) Define Probability model. State axioms of probability. If A and B are two events defined on sample space 'S' then, Prove that, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

5) Explain concept of conditional probability of an event. Also state Baye's theorem.

6) Let A and B be the two events defined on 'S' . If P(A) = 0.6 P(B) = 0.5 $P(A\cap B)$ = 0.2 Then find the probability of occurrence of i) at least one of the events , ii) none of the events, iii) only event A iv) not event A

7) What is the probability that a leap year will contain 53 Thursdays or Fridays?

8) From 6 positive and 8 negative numbers, two numbers are selected at random without replacement and they are multiplied. Find the probability that the product is positive. Also find the probability that the product is zero.

9) Let A, B, and c are MEE and exhaustive event defined for a random experiment. Find P(A) given that P(B) = 1.5P(A) and P(C) = 0.5 P(B).

10) Probability that a man will be alive 25 years hence is 0.3 and the probability that his wife will be alive 25 years hence is 0.4, Find the probability that i) Both will be alive ii) only the man will be alive iii) only the women will be alive iv) none will be alive.

11) Two events A & B are MEE with P(A) = 1/5 & P(B) = 1/3 find probability that

a) Either A or B occurs ii) both A or B occurs iii) Neither A nor B occurs

Chapter 2: Continuous Random Variable

1) Define the following terms :

a) Continuous Random Variable

b) Probability Density Function

c) Expectation of a Continuous Random Variable (CRV).

d) Expectation of a function of a continuous Random Variable.

e) Variance of a CRV

2) Define cumulative distribution function of a continuous random variable. State its properties.

3) A continuous random variable X has distribution function given by

$$F(X) = \begin{cases} 1 - (4/X^2) & ; \quad X > 2 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Find: **a)** Probability Density Function **b)** P(X >3) **c)** E (2X + 3) **d)** V(2X + 3).

4) Let X be a Continuous random variable with p. d. f.

$$f(X) = \begin{cases} 3X^2 & ; \quad 0 \le X \le 1 \\ 0 & ; \quad \text{otherwise} \end{cases}$$

Find E (3X + 2) and V(3X +2).

5) The distribution function of a Continuous Random Variable X is:

$$F(X) = \begin{cases} 0\ ; & X < -1 \\ (X3 + 1)/9\ ; & -1 \le X \le 2 \\ 1\ ; & X > 2 \end{cases}$$

Obtain p. d. f. of X. Also find $P(0 < X \le 1)$.

Chapter 3: Standard Continuous probability distributions.

1) Define uniform distribution over an interval [a,b]. Find its mean, variance and distribution function. Also, state its applications.

2) Define Exponential Distribution with parameter θ. State its mean and variance. Also, state lack of memory property and its interpretation.

3) a) Define Normal distribution. State its properties.

b) State the normal approximation to binomial and Poisson distribution.

4) Define Pareto distribution. State its mean and variance. State three real life situations where Pareto distribution is used. If mean and variance of Pareto distribution is 1.2 and 0.06 respectively, what is the value of α?

5) The amount of time, in hours that a computer functions before breaking down has Exponential distribution with probability density function (p. d. f.) given by,

$$f(X) = 0.005\, e^{-0.005X}; \qquad X \geq 0$$

$$= 0; \qquad \text{otherwise}$$

Find mean of above distribution. Also, write distribution function.

6) The quantity of oil per tin filled by a machine is uniformly distributed over 240 ml and 260 ml. What is the probability that an oil tin is filled with more than 250 ml of oil. Further, if 25 tins are filled with oil, what is the expected quantity of oil consumed?

7) The life time in hours of a certain electrical component follows exponential distribution with distribution function:

$$F(X) = 1 - e^{-0.01X}; \qquad X \geq 0$$

What is the probability that the component will survive 75 hours? Also, what is the probability that it will fail during 100 to 150 hours?

.

8) If mean and variance of a U[a, b] r.v. are 6 and 4 respectively, determine the values of 'a' and 'b' ii) Suppose $X \rightarrow U[0, 10]$. Find mean, variance, $P(X > 4)$, $P(X \leq 3)$ and $P(2 \leq X \leq 6)$.

9) On a route, the first bus is at 8:00 am and after every 30 minutes, there is a bus. A passenger arrives at the stop at time X, which is uniformly distributed over the interval [8:15 am, 8:45 am]. What is the probability that the passenger will have to wait for more than 15 minutes for a bus?

10) A random variable X has an exponential distribution with mean 5. Find:

i) $P[(X > 8)/(X > 4)]$ ii) $P[X > 5]$

11) Suppose that the life of a electrical component is exponentially distributed with a mean life of 1600 hrs. What is the probability that the electrical component will work : i) up to 2400 hours ii) after 1000 hours iii) between 1500 and 2000 hours. Also find the distribution function.

12) Suppose a r. v. X follows normal distribution with mean 3 and variance 16. Find:

i) $P(X > 5)$ ii) $P(X < 1)$ iii) $P(X > 0)$ iv) $P(X < 6)$ v) $(2 < X < 6)$ vi) $P(|X-3|<3.92)$ vii) $P(|X|> 4)$.

13) There are 1000 students in the university of a certain age group and it is known that their weights are normally distributed with mean 55 kg and standard deviation 4.5kg. Find the number of students having weight:

i) less than 48 kg ii) between 50 kg and 58 kg and iii) above 65

14) If $X \rightarrow N(10, 36)$ $Y \rightarrow N(20,49)$ and if X and Y be independent, then state the distribution of $(X+Y)$ and distribution of $(2X + 4Y+3)$.

15) Let $X \rightarrow U(-1,3)$ and Y follows an exponential distribution with mean θ. Find the value of θ such that Variance (X) = Variance(Y).

Chapter 4: Large Sample Test

1) Define the following terms:

i) population ii) Sample iii) parameter iv) statistic v)hypothesis vi) Null and alternative hypothesis vii) critical region vii) types of errors viii) level of significance and xi) p- value x) one sided and two sided tests.

2) Explain SRSWR and SRSWOR methods of sampling. Also define sampling distribution of a statistic with one illustration.

3) Describe the test procedure for testing Ho : μ1 = μ2 against H1 : μ1 ≠ μ2 at α% L.O.S. for n>30

4) Describe the test procedure for testing the null hypothesis that population mean (μ) has some specified value μo for n >30.

5) Describe the test procedure for testing Ho : P1 = P2 against H1 : P1 ≠ P2 for a large sample test at α% L.O.S.

6) A sample of 100 students is taken from a large population. The mean height of the student in this sample is 160 cm. Can it be reasonably regarded that, in the population, the mean height is 165 cm and the Standard deviation is 10 cm? Use 5% level of significance

7) A man buy's 50 electric bulbs of 'Philips' and another 50 of 'HMT'. He finds the Philips bulbs give an average life of 1500 Hrs with a Standard deviation of 60 Hrs and the HMT bulbs give an average of 1512 Hrs with standard deviations 80 Hrs. Is there significant difference between the 'Philips' and 'HMT'. (ά = 10%)

8) On 384 out of 600 randomly selected farms it was discovered that the farm operator was also the owner. Is there any reason to believe that 60% of the operators are also owners? Use 1% LOS.

9) The result of reading test given to the group of 100 girls and another group of 100 boys are as follows. Test at the 5 % level of significance whether girls read at faster rate than boys? [Use α=1%]

	Mean Rate	Variance
Girls	206	450
boys	191	450

Chapter 5: t- Test

1) Explain stepwise procedure of paired t – test.

2) Describe the test procedure for testing the null hypothesis that $\mu1 = \mu2$ against H1 : $\mu1 \neq \mu2$ at α% LOS [Assume that $\sigma12 = \sigma22 = \sigma$(unknown), $(n1,n2) \leq 30$].

3) A machine is supposed to produce washers of mean thickness 0.12c.m. A sample of 10 washers was found to have a mean thickness of 0.128 c.m. & S.D. of 0.008c.m.. Test whether the machine is working properly at 5 % LOS.

4) A certain stimulus administered to each of the 12 patients resulted in the following increases of blood pressure: 5 , 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6.Can it be conclude that the stimulus will, in general, be accompanied by an increase in blood pressure? Use 5 % LOS.

5) Twenty patients on a certain diet made the following weight gains(in pounds) 7, -6, 3 ,1, 6, 4, 9, -5, 9, -7, -3, 7, -9, 8, 6, -4, 4, 9, -6, 1. Test the hypothesis that the median weight gain is zero against it is not. (5% LOS)

Chapter 6: Chi-square Test

1) Explain Chi-Square test to test independence of two attributes.

2) Write stepwise procedure for testing a goodness of fit.

3) A survey was conducted to investigate whether alcohol drinking and smoking are related. The following information was compiled for 600 persons.

	Smoker	Non-Smoker
Drinker	193	165
Non-Drinker	89	153

Test the hypothesis whether alcohol drinking and smoking are related at 5% LOS.

4) A random sample of size 20 from a population gave the standard deviation as 7.3. Test at 5% level of significance the hypothesis that the standard deviation of population is 8.

5) A certain drug is claimed to be effective in curing colds. In an experiment on 500 people with colds, half of them were given the drug and half of them were given the sugar pills. The patients reaction to the treatment are recorded in the following table:

	Helped	Harmed	No effect
Drug	150	30	70
Sugar Pill	130	40	80

On the basis of data, can it be concluded that there is a significant difference in the effect of the drug and sugar pills?

Chapter 7: Non- Parametric Test

1) Explain the procedure of run test for testing the randomness of a sample.

2) Explain how sign test can be used in testing symmetry of the sample.

3) Describe Mann – Whitney test.

4) Given below is a sample of size 12 of length of rods (in cm) produced on lathe. **4**

0.39, 0.37, 0.62, 0.47, 0.52, 0.55, 0.64, 0.43, 0.82, 0.68, 0.35, 0.42

The population median length is claimed to be 0.5. Test this claim using sign test. Use 5% level of significance.

5) A sequence of small glass sculptures was inspected for shipping damage. The sequence of acceptable (A) and damaged (D) pieces was as follows:

D A A A D D D A A A A A D A D A A D A A .

Test for randomness of the damage to the shipment using 5% los.

Chapter 8: Simulation

1) Describe the stepwise procedure to generate random sample of size n from an exponential distribution with mean θ
2) Describe the stepwise procedure to generate random sample of size n from Uniform distribution (a,b)
3) Describe the stepwise procedure to generate random sample of size n from Normal distribution with mean μ and variance σ^2 by using Box-Muller Transformation.