Question Bank of Statistics and Probability

Date: / / Page no:

Short Notes :-

1. Collection and Classification of data :-

Data collection is a process of gathering information from all the relevant sources to find a solution to the research problem. It helps to evaluate the outcome of the problem. The data collection methods allow a person to conclude an answer to the relevant question.

The collection of data constitutes the starting point of any statistical investigation. Data may be collected for each and every unit of the whole lot (population), for it would ensure greater accuracy.
According to Coner, "The classification of data are done according to similarities and properties, so that special characteristics can be predicted."
The classification is done by dividing the raw data into a convenient number of groups according to the values of the variable and finding the frequency of the variable in each group.

2. Define Statistics and its Basic Objectives :-

Date: / / Page no:

Statistics is an important branch of mathematics. Statistics deals with the methods for collection, classification and analysis of numerical data for drawing valid conclusions and making reasonable decisions.

According to Bowley :- "Statistics is the science of the calculations and averages."

According to Lovit :- "Statistics is the science in which the collection, classification and tabulation of the numerical facts are done for the explanation, information and comparison of any event."

Basic Objectives :-

1. To make inference about a population from an analysis of information contained in the sample data.

2. To make assessments of the extent of uncertainity involved in these inferences.

3. To design the process and the extent of sampling so that the observations from a basis for drawing valid and accurate inferences

3. Discuss the applications of Statistics in various branches of Science :-

Statistics plays a vital role in every fields of human activity. Statistics has important role in determining the existing position of per capita income, unemployment, population growth rate, housing, schooling medical facilities etc. in a country.

(i) In Biology and Medical Sciences :-

A good understanding of biostatistics is essential for those who are from biological sciences, as the methods of bio-statistics are indispensable tools for the design and analysis of data and in the interpretation of experimental results for dependable conclusions.
In medical sciences also, the statistical tools for the collection, presentation and analysis of observed facts relating to the causes and incidence of diseases and the results obtained from the use of various drugs and medicines are of great importance.

(2). In Astronomy :- Astronomy is one of the most oldest branches of statistical study. It deals with the measurement of distance, sizes, masses and densities of heavenly bodies by means of observations. During these measurements errors are unavoidable, so most probable measurements are founded by using statistical methods

Date: / / Page no:

4. Write a Short note on : (i) Internal and External data
(ii) Primary & Secondary data.

(i) Internal and External Data :

Internal data is information generated from within the business, covering areas such as operations, maintenance, personal and finance.
External data comes from the market, including customers and competitors. It's things like statistics from surveys, questionnaires, research and customer feedback.

(ii) Primary and Secondary Data :

Primary data are those which are to be collected for the first time by the investigator (or on his behalf) and therefore, it is of original in nature, whereas Secondary data are those which do not originate from the investigator (or from the field of enquiry) but which are obtained from someone else's records.

Primary data may be used with greater confidence because the investigator will himself decide upon the coverage of the data, whereas secondary data is not so reliable. The secondary data may contain mistakes due to errors in transcription made when figures were copied.

5. Graphical Representation of Data :-

A convenient way of representing a sample frequency distribution is by means of graphs. It gives to the eyes the general run of the observations and at the same time makes the raw data readily intelligible.

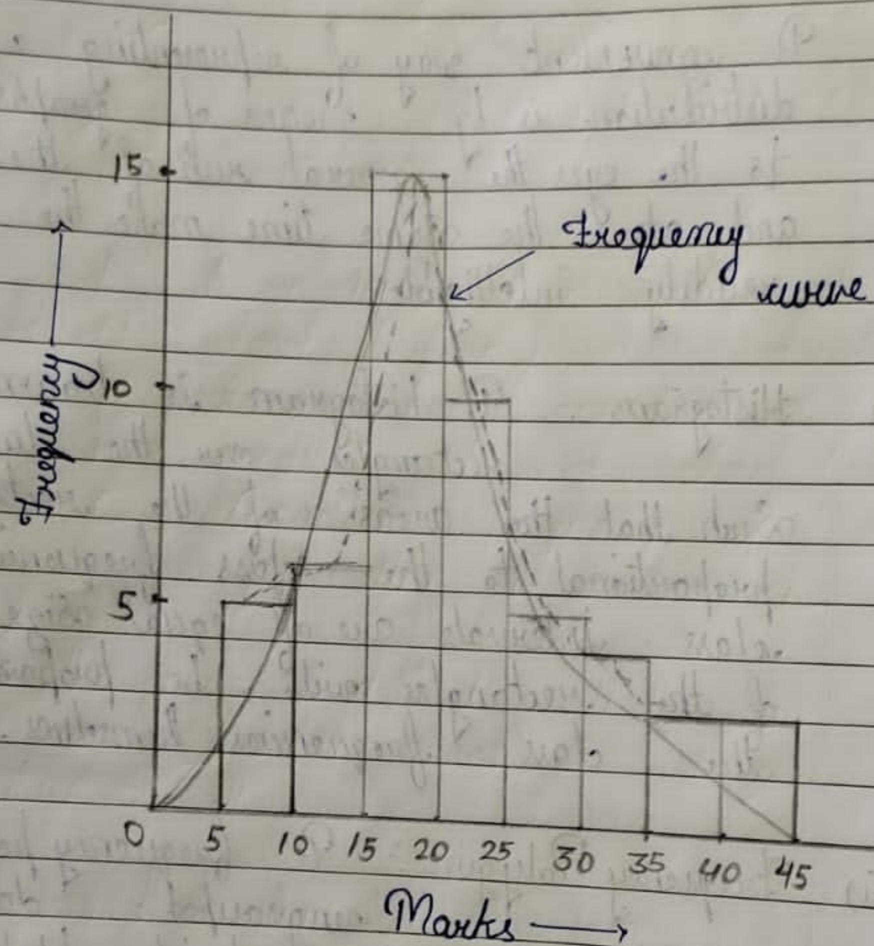
(i). Histogram :- A histogram is drawn by erecting rectangles over the class intervals, such that the areas of the rectangles are proportional to the class frequencies. If the class intervals are of equal size, the height of the rectangles will be proportional to the class frequencies themselves.

(ii). Frequency Polygon :- A frequency polygon for an ungrouped data can be obtained by joining points plotted with the variable values as the abscissa and the frequencies as the ordinates.

For a grouped-distributions, the abscissa of the points will be the mid-values of the class-intervals. In case the class intervals are equal, the frequency polygon can be obtained by joining the middle points of the upper sides of the rectangles of the histogram by straight lines. If the class intervals become very-very small, the frequency polygon takes

Date: / / Page no:

the form of a smooth curve called the frequency curve.



6. Mean, Median and Mode.

(i) Mean :- The mean is the average of data set.
(Average)

Formula :- Mean $= \frac{\text{Sum of observations}}{\text{Total no. of observations}}$

It is represented by $\overline{x}$.

$$\overline{x} = \frac{\Sigma x_i}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum f_i}$$

Assumed mean method : $\bar{x} = a + \frac{\sum f_i d_i}{\sum f_i}$

where $d_i = x_i - a$
a = assumed mean.

Step-deviation method : $\bar{x} = a + \frac{\sum f_i u_i}{\sum f_i} \times h$

where, $u_i = \left(\frac{x_i - a}{h}\right)$

$u_i = \frac{d_i}{h}$ or $d_i = u_i h$

h = class interval.

(ii) Median : If the values of a variable are arranged in the ascending order of magnitude, the median is the middle item if the number is odd and is the mean of the two middle items if the number is even. Thus, the median is equal to the mid-value, i.e., the value which divides the total frequency into two equal parts.

M (if data is odd) : $\left(\frac{n+1}{2}\right)^{th}$ term

Date: / / Page no:

M(if data is even) $= \left[\dfrac{\left(\frac{n}{2}+1\right)+\frac{n}{2}}{2}\right]^{th}$ term.

For the grouped data :- $M = L + \dfrac{\left(\frac{1}{2}N - c\right)}{f} \times h$

Where, L = lower limit of the median class.
N = total frequency $= \Sigma f_i$
f = frequency of the median class
h = width of median class (class size).
C = cumulative frequency of class preceeding the median class.

(iii). Mode : The mode is defined as that value of the variable which occurs most frequently, i.e., the value of the maximum frequency.

For a grouped distribution, Mode $= L + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} \times h$

where, L = lower limit of modal class
Δ_1 = excess of modal frequency over frequency of preceeding class $(f_1 - f_0)$
Δ_2 = excess of modal frequency over following class $(f_1 - f_2)$

h = size of modal class

7. Inter Quartile Range :

The inter-quartile range defines the difference between the third quartile and the first quartile. Quartiles are the partitioned values that divide the whole series into 4 equal parts. So, there are 3 quartiles. First is Q_1 and known as lower quartile, the second is Q_2 and third is Q_3 known as upper quartile.

Therefore, Inter-quartile range = Upper Quartile - Lower Quartile

$$= Q_3 - Q_1$$

8. Mean-Deviation, Standard Deviation and Variance :

(i) Mean-deviation : The mean deviation is the mean of the absolute differences of the values from the mean, median or mode.

Thus mean deviation (M.D.) $= \frac{1}{n} \Sigma f_i |x_i - A|$

where, A is either the mean or the median or the mode.

(ii). Standard-deviation : The most important and the most powerful measure of dispersion is the standard-deviation (S.D.) denoted by σ. It is computed as the square root of the mean of the squares of the differences of the

Date: / / Page no:

variate values from their mean.

$$S.D = \sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{N}}$$

where N is the total frequency Σf_i.

(iii). Variance : The square of the standard deviation is known as the variance. It is denoted by σ^2.

$$\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N}$$

9 Moments about a point and Moments about the mean:

Moments are a set of statistical parameters to measure a distribution.

(i). Moments about mean :

The moments about mean are the mean of deviations from the mean after raising them to integer powers. The r^{th} population moment about mean is denoted by μ_r is :

$$\mu_r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^r}{N} \quad \text{or} \quad \frac{1}{N} \Sigma f_i (x_i - \bar{x})^r$$

where, $r = 1, 2 \ldots$

Date: / / Page no:

The corresponding moment about any point a is defined as:

$$\mu'_r = \frac{1}{N}\Sigma f_i (x_i - a)^r$$

In particular, we have $\mu_0 = \mu'_0 = 1$

$$\mu_1 = \frac{1}{N}\Sigma f_i (x_i - \bar{x}) = 0$$

$$\mu'_1 = \frac{1}{N}\Sigma f_i (x_i - a) = \bar{x} - a = d$$

$$\mu_2 = \frac{1}{N}\Sigma f_i (x_i - \bar{x})^2 = \sigma^2$$

Hence, first moment is always zero.
Second moment is variance.

For grouped data the r^{th} sample moment about sample mean $\bar{x}$ is:

$$\mu_r = \frac{\Sigma_{i=1}^{n} f_i (x_i - \bar{x})^r}{\Sigma_{i=1}^{n} f_i}$$

where $\Sigma_{i=1}^{n} f_i = n$

$$\mu'_r = \frac{\Sigma_{i=1}^{n} f_i (x_i - a)^r}{\Sigma_{i=1}^{n} f} = \frac{\Sigma f_i d_i^r}{\Sigma f}$$

Moments about the mean in terms of moments about any point

Date: / / Page no:

* $\mu_r = \mu'_r - {}^rC_1 d\mu'_{r-1} + {}^rC_2 d^2\mu'_{r-2} - \ldots\ldots$

* $\mu_2 = \mu'_2 - \mu'^2_1$

* $\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu'^3_1$

* $\mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2\mu'^2_1 - 3\mu'^4_1$

10. Conditional Probability :-

Conditional Probability is defined as the likelihood of an event or outcome occurring, based on the occurrence of a previous event or outcome. Conditional probability is calculated by multiplying the probability of the preceding event by the updated probability of the succeeding, or conditional, event.

For eg. > Event A is that an individual applying for college will be accepted. There is an 80% chance that this individual will be accepted to college.

> Event B is that this individual will be given dormitory housing. Dormitory housing will only be provided for 60% of all of the accepted students.

> P(Accepted and dormitory housing) = P(Domitory housing | Accepted) P(Accepted)

$$= (0.60) \times (0.80) = 0.48$$

Conditional probability would look at these two events in relationship with one another, such as the probability that you are both accepted to college and you are provided with dormitory housing.

11. Addition and Multiplication rules of Probability:

(i) If the probability of an event A happening as a result of trial is P(A) and the probability of a mutually exclusive event B happening is P(B), then the probability of either of the events happening as a result of the trial is

$$P(A+B) \text{ or } P(A \cup B) = P(A) + P(B)$$

If A, B are any two events (not mutually exclusive), then

$$P(A+B) = P(A) + P(B) - P(AB)$$

OR

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(ii) If the probability of an event A happening as a result of trials is P(A) and after A has happened the probability of an event B happening as a result of another trial (i.e., conditional probability of B given A) is $P(B/A)$, then the probability of both the events A and B happening as a result of two trials is

$$P(AB) \text{ or } P(A \cap B) = P(A).P(B/A) = P(B).P(A/B)$$

If the events A and B are independent, i.e., if the happening of B does not depend on whether A has happend or not, then

$$P(B/A) = P(B) \text{ and } P(A/B) = P(A)$$

$$\therefore \; P(AB) \text{ or } P(A \cap B) = P(A).P(B).$$

12. Define Expectation:

Mathematical expectation, also known as the expected value, is the summation or integration of a possible values from a random variable. It is also known as the product of the probability of an event occurring, denoted $P(x)$, and the value corresponding with the actual observed occurrence of the event. The expected value is a useful property of any random variable. Usually notated as $E(X)$, the expect value can be computed by the summation overall the distinct values that the random variable can take. The mathematical expectation will be given by the mathematical formula as,

$$E(x) = \Sigma(x_1 p_1, x_2 p_2 \ldots\ldots, x_n p_n)$$

where, x is a random variable with the probability function, $f(x)$. P is the probability of the occurrence, and n is the number of all possible values.

13. Define Variance and moments about mean of the probability distribution :-

The variance of a probability distribution is the mean of the squared distance to the mean of the distribution. The variance is the second most important measure of a probability distribution, after the mean. It quantifies the spread of the outcomes of a probability distribution.

Variance of the distribution is given by:

$$\sigma^2 = \sum_i (x_i - \mu)^2 f(x_i) \quad \text{(Discrete distribution)}$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx \quad \text{(Continuous distribution)}$$

The r^{th} moment about the mean (denoted by μ_r) is defined by

$$\mu_r = \Sigma f \quad \Sigma(x_i - \mu)^r f(x_i) \quad \text{(discrete)}$$

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu)^r f(x)\, dx \quad \text{(continuous)}$$

14. Moment Generating Functions and its properties :-

The moment generating function (m.g.f) of the discrete probability distribution of the variate X about the value $x = a$ is defined as the expected value of $e^{t(x-a)}$ and is denoted by $M_a(t)$. Thus,

$$M_a(t) = \Sigma p_i e^{t(x_i - a)} \quad \text{—— (1)}.$$

which is a function of the parameter t only.

Expanding the exponential in (1),

$$M_a(t) = \Sigma p_i + t\Sigma p_i (x_i - a) + \frac{t^2}{2!} \Sigma p_i (x_i - a)^2 + \ldots\ldots + \frac{t^r}{r!} \Sigma x_i (x_i - a)^r + \ldots$$

$$= 1 + t\mu_1' + \frac{t^2}{2!}\mu_2' + \ldots + \frac{t^r}{r!}\mu_r' + \ldots$$

where μ_r' is the moment of order r about a. Thus $M_a(t)$ generates moments and that is why it is called the moment generating function.

15. Define Population and Sample :-

Population is the entire group that we want to draw conclusions about.

Population is the pool of individuals from which a Statistical Sample is drawn. Thus, any selection of individuals grouped together by a common feature can be said to be a population.

A sample refers to a smaller manageable version of a larger group. It is a subset containing the characteristics of a larger population. Samples are used in statistical testing when population sizes are too large for the test to include all possible members or observations.

16. Degree of freedom :-

Degree of freedom are the number of independent values that a statistical analysis can estimate. The ~~dof~~ degrees of freedom equals your sample size minus the number of parameters you need to calculate during an analysis. It is usually a positive whole number.

Degree of freedom $DF = N - P$

where, N = Sample Size, P = the number of parameters or relationships

17. Standard Error :-

The standard deviation of the sampling

Date: / / Page no:

distribution is called the Standard error (S.E.). Thus, the Standard error of the sampling distribution of means is called the Standard error of means. The Standard error is used to assess the difference between the expected and observed values. The reciprocal of the Standard error is called precision.

18. Type I and Type II errors :-

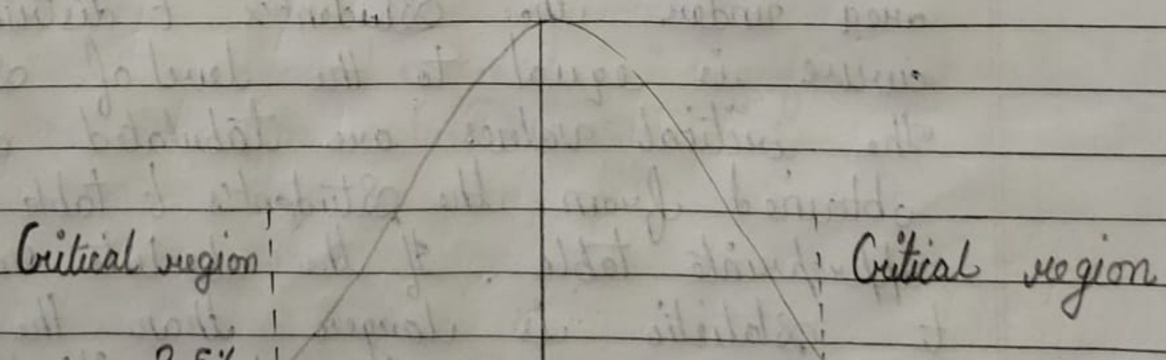
If a hypothesis is rejected while it should have been accepted, we say that a Type I error has been committed. On the other hand, if a hypothesis is accepted while it should have been rejected, we say that a Type II error has been made. The statistical testing of hypothesis aims at limiting the Type I error to a preassigned value (say: 5% or 1%) and to minimize the Type II error. The only way to reduce both types of errors is to increase the sample size, if possible.

19. Test of Significance :- The procedure which enables us to decide whether to accept or reject the hypothesis is called the test of significance. Here we test whether the differences between the sample values and the population values (or the

values given by two samples) are so large that they signify evidence against the hypothesis or these differences are so small as to account for fluctuations of sampling.

20. Level of Significance :

The probability level below which we reject the hypothesis is known as the level of significance. The region in which a sample value falling is rejected, is known as the critical region. We generally take two critical regions which cover 5% and 1% areas of the normal curve. The shaded portion, in the figure corresponds to 5% level of significance. Thus, the probability of the value of the variate falling in the critical region is the level of significance.



Date: / / Page no:

21. Null Hypothesis : The hypothesis formulated for the sake of rejecting it, under the assumption that it is true, is called the null hypothesis and is denoted by H_0. To test whether one procedure is better than another, we assume that there is no difference between the procedures. Similarly to test whether there is a relationship between two variates, we take H_0 that there is no relationship. By accepting a null ~~p~~ hypothesis, we mean that on the basis of the statistic calculated from the sample, we do not reject the hypothesis. It however, does not imply that the hypothesis is proved to be true. Nor its rejection implies that it is disproved.

22. Critical Region : The critical region is the region of values that corresponds to the rejection of the null hypothesis at some chosen probability level. The shaded area under the Student's t distribution curve is equal to the level of significance. The critical values are tabulated and thus obtained from the Student's t table or another appropriate table. If the absolute value of the t statistic is larger than the tabulated value, then t is in the critical region.

23. Goodness of fit :

The value of x^2 is used to test whether the deviations of the observed frequencies from the expected frequencies are significant ~~of~~ or not. It is also used to examine the validity of some hypothesis about an observed frequency distribution. As a test of goodness of fit, it can be used to study the correspondence between theory and fact.

Procedure to test significance and goodness of fit:

(i) Set up a "null hypothesis" and calculate,

$$x^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

(ii) Find the df and read the corresponding values of x^2 at a prescribed significance level from Table V.

(iii). From x^2-table, we can also find the probability P corresponding to the calculated values of x^2 for the given d.f.

(iv) If $P < 0.05$, the observed value of x^2 is significant at 5% level of sig. If $P < 0.01$, the value is significant at 1% level. If $P > 0.05$, it is a good fit and the value is not significant.

24. Random Sample: A random sample is a subset of a statistical population in which each member of the subset has an equal

probability of being chosen. A simple random sample is meant to be an unbiased representation of a group. A simple random sample takes a small, random portion of the entire population to represent the entire data set, where each member has an equal probability of being chosen.

25 Sample Size for Different distribution :-

Sample size measures the number of individual samples measured or observations used in a survey or experiment.

For eg. If you test 100 samples of soil for evidence of acid rain, your sample size is 100. If an online survey returned 30500 completed questionnaries, your sample size is 30500. It is represented by "n".

Q1. The mean of 200 items was 50. Later it was discovered that two items were missed as 92 and 8 instead of 192 and 88. Find out correct mean.

Sol. Given :- Mean of 200 items = 50.

Since, $\frac{\text{Sum of 200 items}}{\text{No. of items}} = \text{mean}$

Sum of 200 items = mean × no. of items

$= 50 \times 200$

$= 10,000$

So, wrong Sum of 200 items = 10,000

Correct Sum = $10,000 + 192 + 88 - 92 - 8$

$= 10280 - 100$

$= 10180.$

∴, Correct mean $= \frac{10180}{200} = 50.9$

Q2. Find mean of the first n natural numbers.

Sol. We know that, Sum of first n natural numbers is $S_n = \frac{n(n+1)}{2}$

So, mean of the first n natural numbers =

$$\frac{\text{Sum of natural numbers}}{n} = \frac{n(n+1)}{2} \Big/ n$$

$$= \frac{\cancel{n}(n+1)}{2\cancel{n}}$$

$$= \frac{n+1}{2}$$

Q3. Find the mean, median and mode of values:
2, 9, 11, 5, 1, 4, 15, 8, 10, 15, 15, 8, 15, 7.

Sol. (i) Mean $(\bar{x})$ = $\frac{\text{Sum of all observations}}{\text{No. of observations}}$

$$= \frac{2+9+11+5+1+4+15+8+10+15+15+8+15+7}{14}$$

$$= \frac{125}{14}$$

$$= 8.92$$

(ii). Now, Arranging the data in ascending order:
1, 2, 4, 5, 7, 8, 8, 9, 10, 11, 15, 15, 15, 15.

No. of observations n = 14 $\rightarrow$ even.

So, Median (M) $= \left[\dfrac{\left(\frac{n}{2}+1\right)+\frac{n}{2}}{2}\right]^{th}$ term

$= \left(\dfrac{\left(\frac{14}{2}+1\right)+\frac{14}{2}}{2}\right)^{th}$ term.

$= \left(\dfrac{(7+1)+7}{2}\right)^{th}$ term.

= Mean of 8^{th} and 7^{th} term

$= \dfrac{8+9}{2}$

$= \dfrac{17}{2}$

$= 8.5$.

(iii). No. of most frequent term = 15.

15 occurs most frequent time, So,

15 is the mode of data.

Q.4. Calculate the mean and median for the following distribution:

X:	1	2	3	4	5	6	7	8	9
Y:	8	10	11	16	20	25	15	9	6

Date: / / Page no:

Sol. Given :-

Class interval	x_i	f_i	$x_i - a = d_i$	$f_i d_i$	cf
0.5-1.5	1	8	-4	-32	8
1.5-2.5	2	10	-3	-30	18
2.5-3.5	3	11	-2	-22	29
3.5-4.5	4	16	-1	-16	45
4.5-5.5	$\boxed{5}$	20	0	0	$\boxed{65}$
5.5-6.5	6	25	1	25	90
6.5-7.5	7	15	2	30	105
7.5-8.5	8	9	3	27	114
8.5-9.5	9	6	4	24	120
		$\Sigma f_i = 120$		$\Sigma f_i d_i = 6$	

$$\text{Mean } \bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

$$= 5 + \frac{6}{120}$$

$$= 5 + \frac{1}{20}$$

$$= \frac{100+1}{20}$$

$$= \frac{101}{20}$$

$$= 5.05$$

For finding median :- $N = \Sigma f_i = 120$

$N/2 = 60$.

So, Median class = 4.5 – 5.5.

$l = 4.5$
$cf = 45$
$f = 20$
$h = 1$
$N/2 = 60$.

$$\text{Median} = l + \left(\frac{N/2 - cf}{f}\right) \times h$$

$$= 4.5 + \left(\frac{60 - 45}{20}\right) \times 1$$

$$= 4.5 + \left(\frac{3}{4}\right) \times 1$$

$$= 4.5 + 0.75$$

$$= 5.25.$$

Q.5. Find mean, median and mode of the following data :-

Marks :	0-10	10-20	20-30	30-40	40-50
No. of Students :	22	38	46	35	20

Sol.

x_i	Marks (Class interval)	No. of Students (f_i)	$x_i - a = d_i$	$u_i = d_i/h$	$f_i u_i$	cf
5	0-10	22	-20	-2	-44	22
15	10-20	38	-10	-1	-38	60
$\boxed{25}$	$\boxed{20}$-30	$\boxed{46}$	0	0	0	$\boxed{106}$
35	30-40	35	10	1	35	141
45	40-50	20	20	2	40	161
		$\Sigma f_i = 161$		Σf_i	$\Sigma f_i u_i = -7$	

$$\text{Mean } \bar{x} = a + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h$$

$$= 25 + \left(\frac{-7}{161}\right) \times 10$$

$$= 25 - \frac{70}{161}$$

$$= \frac{4025 - 70}{161}$$

$$= \frac{3955}{161}$$

$$= 24.5$$

(ii) For finding median : $N = 161$

$$= N/2 = \frac{161}{2} = 80.5$$

(iii).

So, median class : 20 - 30

$l = 20$

$cf = 60$

$f = 46$

$N/2 = 80.5$

$h = 10$

Median $M = 20 + \left(\frac{80.5 - 60}{46}\right) \times 10$

$= 20 + \left(\frac{20.5}{46}\right) \times 10$

$= 20 + \frac{205}{46}$

$= \frac{920 + 205}{46}$

$= \frac{1125}{46}$

$= 24.45.$

(iii). Modal class = 20 - 30 (because it has maximum frequency)

$L = 20$

$\Delta_1 = f_1 - f_0$

$= 46 - 38$

$= 8$

$\Delta_2 = f_1 - f_2$

$= 46 - 35 = 11$

$h = 10$

Date: / / Page no:

$$\text{Mode} = L + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right) \times h$$

$$= 20 + \left(\frac{8}{8+11}\right) \times 10$$

$$= 20 + \frac{80}{19}$$

$$= 20 + 4.21$$

$$= 24.21$$

Q6. An incomplete frequency distribution is given below:

Variable	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	12	30	?	65	?	25	18

Given that the total frequency is 229 and median is 46, find the missing frequencies.

Sol.

Variable	Frequency	cf
10-20	12	12
20-30	30	42
30-40	f_1	$42 + f_1$
40-50	65	$107 + f_1$
50-60	f_2	$107 + f_1 + f_2$
60-70	25	$132 + f_1 + f_2$
70-80	18	$150 + f_1 + f_2$
	$\Sigma f_i = 229$	

$$150 + f_1 + f_2 = 229$$

$$f_1 + f_2 = 229 - 150$$

$$\boxed{f_1 + f_2 = 79} \longrightarrow (1)$$

Given : Median $= 46$.

For median, median class will be 40-50

because, $N/2 = \frac{229}{2} = 114.5$.

$$l = 40$$
$$cf = 42 + f_1$$
$$f = 65$$
$$h = 10.$$

$$\text{Median} = l + \left(\frac{N/2 - cf}{f}\right) \times h$$

$$46 = 40 + \left(\frac{114.5 - (42 + f_1)}{\cancel{65}_{13}}\right) \times \cancel{10}^{2}$$

$$46 = 40 + \left(\frac{114.5 - 42 - f_1}{13}\right) \times 2$$

$$46 = 40 + \left(\frac{(72.5 - f_1)2}{13}\right)$$

$$46 = \frac{(40 \times 13) + (145 - 2f_1)}{13}$$

Date: / / Page no:

$$598 = 520 + 145 - 2f_1$$

$$598 - 520 - 145 = -2f_1$$

$$78 - 145 = -2f_1$$

$$\cancel{-}67 = \cancel{-}2f_1$$

$$f_1 = \frac{67}{2}$$

$$\boxed{f_1 = 33.5}$$ → which can be taken as 34.

Putting value of f_1 in eq^n (1).

$$f_1 + f_2 = 79$$

$$33.5 + f_2 = 79$$

$$f_2 = 79.0 - 33.5$$

$$\boxed{f_2 = 45.5}$$ → 45. (round figure)

Hence, the missing frequencies are 33.5 and 45.5
34 and 45

Q.7. Find mean and mode of the following frequency distribution :-

Variable :	0-9	10-19	20-29	30-39	40-49	50-59	60-69
Frequency :	2	18	30	45	35	20	6

Date: / / Page no:

Sol.

Variable	f_i	x_i	$d_i = x_i - a$	$u_i = d_i/h$	$f_i u_i$
-0.5 - 9.5	2	4.5	-30	-3	-6
9.5 - 19.5	18	14.5	-20	-2	-36
19.5 - 29.5	30	24.5	-10	-1	-30
29.5 - 39.5	45	$\boxed{34.5}$	0	0	0
39.5 - 49.5	35	44.5	10	1	35
49.5 - 59.5	20	54.5	20	2	40
59.5 - 69.5	6	64.5	30	3	18
	156				$\Sigma f_i u_i = 21$

$$\text{Mean } \bar{x} = a + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times h$$

$$= 34.5 + \left(\frac{21}{156}\right) \times 10$$

$$= 34.5 + \frac{210}{156}$$

$$= 34.5 + 1.34$$

$$= 35.84$$

Mode :- Modal class = 29.5 - 39.5 (because it has maximum frequency)

$f_0 = 30, \quad f_1 = 45, \quad f_2 = 35$

$l = 29.5, \quad \Delta_1 = f_1 - f_0 = 45 - 30 = 15$

$\Delta_2 = f_1 - f_2 = 45 - 35 = 10, \quad h = 10$

Date: / / Page no:

$$\text{Mode} = l + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$

$$= 29.5 + \left(\frac{15}{15+10}\right) \times 10$$

$$= 29.5 + \left(\frac{\cancel{15}^{3}}{\cancel{25}_{\cancel{5}\,1}}\right) \times \cancel{10}\,2$$

$$= 29.5 + 6$$

$$= 35.5$$

Ans. Hence, mean = 35.84 and mode = 35.5

Q.8 Find the median and lower quartile from the following table.

Marks	No. of students
below 10	15
below 20	35
below 30	60
below 40	84
below 50	94
below 60	127
below 70	198
below 80	249

Date: / / Page no:

Sol. The given table is cumulative frequency distribution.

Marks	Frequency	Cf	LCB
0 - 10	15 - 0 = 15	15	5
10-20	35-15 = 20	35	15
20-30	60-35 = 25	60	25
30-40	84-60 = 24	84	35
40-50	94-84 = 10	94	45
50-60	127-94 = 33	127	55
60-70	198-127 = 71	198	65
70-80	249-198 = 51	249	75
	$\Sigma f_i = 249$		

Here, $N/2 = \frac{249}{2} = 124.5$.

So, median class will be = 50-60

$l = 50,\ h = 10,\ N/2 = 124.5$

$cf = 94,\ f = 33$

$$\text{Median} = l + \left(\frac{N/2 - cf}{f}\right) \times h$$

$$= 50 + \left(\frac{124.5 - 94}{33}\right) \times 10$$

$$= 50 + \left(\frac{30.5 \times 10}{33}\right)$$

$$= 50 + \frac{305}{33}$$

Date: / / Page no:

$= 50 + 9.24$

$= 59.24.$

Now, for lower quartile : $N/4 = \frac{249}{4} = 62.5.$

So, lower quartile class will be $\rightarrow$ 30 − 40.

$l = 30, \; N/4 = 62.5, \; cf = 60, \; f = 24$

$h = 10, \; f_c =$

$$Q_1 = l + \left(\frac{N/4 - cf}{f}\right) \times h$$

$$= 30 + \left(\frac{62.5 - 60}{24}\right) \times 10$$

$$= 30 + \left(\frac{2.5 \times 10}{24}\right)$$

$$= 30 + \frac{25}{24}$$

$$= 30 + 1.04$$

$$= 31.04.$$

Q.9. Calculate mean, median and mode of the following data relating to weight of 120 articles.

Date: / / Page no:

Weight (in gm)	0-10	10-20	20-30	30-40	40-50	50-60
No. of articles	14	17	22	26	23	18

Sol.

Weight	f_i	x_i	$d_i = x_i - a$	$u_i = d_i/h$	f_iu_i	cf
0-10	14	5	-20	-2	-28	14
10-20	17	15	-10	-1	-17	31
20-30	22	$\boxed{25}$	0	0	0	53
30-40	26	35	10	1	26	79
40-50	23	45	20	2	46	102
50-60	18	55	30	3	54	120
	120				81	

$$\text{Mean } (\bar{x}) = a + \left(\frac{\Sigma f_iu_i}{\Sigma f_i}\right) \times h$$

$$= 25 + \left(\frac{81}{120}\right) \times 10$$

$$= 25 + \frac{81}{12}$$

$$= 25 + 6.75$$

$$= 31.675$$

$$\boxed{\bar{x} = 31.675}$$

Now, $N/2 = 120/2 = 60$

So, Median class is 30-40.

$l = 30, \quad N/2 = 60, \quad cf = 53, \quad f = 26,$

$h = 10$

Date: / / Page no:

$$\text{Median} = L + \left(\frac{N/2 - cf}{f}\right) \times h$$

$$= 30 + \left(\frac{60 - 53}{26}\right) \times 10$$

$$= 30 + \left(\frac{7 \times 10}{26}\right)$$

$$= 30 + \frac{70}{26}$$

$$= 30 + 2.6$$

$$\boxed{M = 32.6}$$

Now, Modal class is 30–40, because it has maximum frequency.

$l = 30$, $\Delta_1 = f_1 - f_0 = 26 - 22 = 4$

$h = 10$, $\Delta_2 = f_1 - f_2 = 26 - 23 = 3$.

$$\text{Mode} = l + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$

$$= 30 + \left(\frac{4}{4+3}\right) \times 10$$

$$= 30 + \left(\frac{40}{7}\right)$$

$$= 30 + 5.7$$

$$\boxed{\text{Mode} = 35.7}$$

Q.10 Calculate mean and standard deviation for the following:

Size of items	6	7	8	9	10	11	12
frequency	3	6	9	13	8	5	4

Sol.

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
6	3	18	$6-9 = -3$	9	27
7	6	42	$7-9 = -2$	4	24
8	9	72	$8-9 = -1$	1	9
9	13	117	$9-9 = 0$	0	0
10	8	80	$10-9 = 1$	1	8
11	5	55	$11-9 = 2$	4	20
12	4	48	$12-9 = 3$	9	36
	48	432			124

$$\text{Mean } (\bar{x}) = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{432}{48} = 9$$

$$\text{Standard deviation } (\sigma) = \sqrt{\Sigma f_i (x_i - \bar{x})^2 / \Sigma f_i}$$

$$= \sqrt{\frac{124}{48}}$$

$$= 1.607$$

Q.11 Show the variance of the first n positive integers is $\frac{1}{12}(n^2 - 1)$.

Sol. Variance of first n natural positive integers are:

$$\sigma^2 = \frac{\Sigma n^2}{n} - \left(\frac{\Sigma n}{n}\right)^2$$

$$= \frac{\cancel{n}(n+1)(2n+1)}{6\cancel{n}} - \frac{(\cancel{n}(n+1))^2}{4n^{\cancel{2}}}$$

$$= (n+1)\left[\frac{(2n+1)}{6} - \frac{(n+1)}{4n}\right]$$

$$= (n+1)\left[\frac{2(2n+1) - 3(n+1)}{12}\right]$$

$$= \frac{(n+1)}{12}\left[4n+2-3n-3\right]$$

$$= \frac{(n+1)}{12}\left[n-1\right]$$

$$= \frac{(n+1)(n-1)}{12}$$

$$\sigma^2 = \frac{1(n^2-1)}{12}$$

Hence Proved....

Q.12 The following are scores of two basketball players "A" and "B" in a series of 10 matches.

A:	11	12	15	6	8	5	14	9	7	13
B:	12	6	5	4	20	8	9	11	0	16

Who is better score getter and who is more consistent?

Date: / / Page no:

Sol. Let x denote score of A and y that of B. Taking "20" as the origin.

x	$d\ (x-20)$	d^2	y	$\delta\ (y-20)$	δ^2
11	-9	81	12	-8	64
12	-8	64	6	-14	196
15	-5	25	5	-15	225
6	-14	196	4	-16	256
8	-12	144	20	0	0
5	-15	225	8	-12	144
14	-6	36	9	-11	121
9	-11	121	11	-9	81
7	-13	169	0	-20	400
13	-7	49	16	-4	16
	-100	1110		-109	1503

For A, $\bar{X} = a + \left(\frac{\Sigma d}{n}\right) = 20 + \left(\frac{-10\cancel{0}}{1\cancel{0}}\right)$

$$= 20 - 10$$

$$= 10.$$

$$\sigma_1 = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \sqrt{\left(\frac{111\cancel{0}}{1\cancel{0}}\right) - \left(\frac{-10\cancel{0}}{1\cancel{0}}\right)^2}$$

$$= \sqrt{111 - 100}$$

$$= \sqrt{11}$$

$$= 3.31.$$

$$\text{Coefficient of variation} = \frac{\sigma_1}{\bar{x}} \times 100$$

$$= \frac{3.31}{10} \times 100$$

$$= 33.1\ \%$$

For B,
$$\bar{y} = a + \left(\frac{\Sigma\delta}{n}\right)$$

$$= 20 + \left(\frac{-109}{10}\right)$$

$$= \frac{200 - 109}{10}$$

$$= \frac{91}{10}$$

$$= 9.1$$

$$\sigma_2 = \sqrt{\frac{\Sigma\delta^2}{n} - \left(\frac{\Sigma\delta}{n}\right)^2}$$

$$= \sqrt{\frac{1503}{10} - \left(\frac{-109}{10}\right)^2}$$

$$= \sqrt{150.3 - \left(\frac{11881}{100}\right)}$$

$$= \sqrt{150.3 - 118.81}$$

$$= \sqrt{31.49}$$

$$= 5.61$$

$\therefore$ Coefficient of variation $= \frac{\sigma_2}{\bar{y}} \times 100$

$= \frac{5.61}{9.1} \times 100$

$= 0.61 \times 100$

$= 61.66\%$

Since, Arithmetic mean of A > B,

So A is a better score getter (i.e. more efficient) than B.

Since, the coefficient of variation of B > Coefficient of variation of A. So, A is more consistent than B.

Q.13. The index number of prices of two articles A and B for six consecutive weeks are given below:-

A:	314	326	336	368	404	412
B:	330	331	320	318	321	330

For which has more variable prices?

Sol. Let x denote prices of article A and y that of B.

Taking "320" as the origin.

Date: / / Page no:

x	$d(x-320)$	d^2	y	$\delta(y-320)$	δ^2
314	-6	36	330	10	100
326	6	36	331	11	121
336	16	256	320	0	0
368	48	2304	318	-2	4
404	84	7056	321	1	1
412	92	8464	330	10	100
	240	18152		30	326

For A, $\bar{x} = a + \left(\frac{\Sigma d}{n}\right)$

$= 320 + \left(\frac{240}{6}\right)$

$= 360$

$\sigma_1 = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$

$= \sqrt{\frac{18152}{6} - \left(\frac{240}{6}\right)^2}$ 40

$= \sqrt{3025.33 - 1600}$

$= \sqrt{1425.33}$

$= 37.75$

Coefficient of variation $= \frac{\sigma_1}{\bar{x}} \times 100 = \frac{37.75}{360} \times 100$

$= 10.4\%$

For B, $\bar{Y} = 320 + \left(\frac{30}{6}\right)$

$= 320 + 5$

$= 325.$

$$\sigma_2 = \sqrt{\frac{\Sigma \delta^2}{n} - \left(\frac{\Sigma \delta}{n}\right)^2}$$

$$= \sqrt{\frac{326}{6} - \left(\frac{30}{6}\right)^2}$$

$$= \sqrt{54.33 - 25}$$

$$= \sqrt{29.33}$$

$$= 5.41$$

Coefficient of variation $= \frac{\sigma_2}{\bar{Y}} \times 100$

$= \frac{5.41}{325} \times 100$

$= \frac{541}{325}$

$= 1.66\%$

Since, coefficient of variation of $A > B$.

So, A has more variable prices.

Date: / / Page no:

Q.14. The first three moments about the value 2 of the variable are 1, 16 and -40. Find the three moments about the mean.

Sol. Given : First three moments about the value 2 :

$$\mu_1' = 1, \quad \mu_2' = 16 \quad \text{and} \quad \mu_3' = -40, \quad a = 2$$

find : Three moments about mean, μ_1, μ_2 and μ_3

$$\mu_1' = \frac{1}{N} \Sigma f_i (x_i - a)$$

$$\mu_1' = \bar{x} - a$$

$$1 = \bar{x} - 2$$

$$\boxed{\bar{x} = 3}$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$= 16 - (1)^2$$

$$= 15$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= -40 - 3(16)(1) + 2(1)^3$$

$$= -40 - 48 + 2$$

$$= -86.$$

Date: / / Page no.

Q.15. The first four moments about the value 5 are 2, 20, 40 and 50. Find moments about the mean.

Sol. Given: First four moments about the value 5 is:

$\mu_1' = 2,\ \mu_2' = 20,\ \mu_3' = 40,\ \mu_4' = 50$ and $a = 5$.

find: Moments about mean = ?

$$\mu_1' = \frac{1}{N}\Sigma f_i(x_i - a)$$

$$\mu_1' = \bar{x} - a$$

$$2 = \bar{x} - 5$$

$$\boxed{\bar{x} = 7}$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$= 20 - (2)^2$$

$$= 20 - 4$$

$$= 16$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= 50 - 4(40)(2) + 6(20)(2)^2 - 3(2)^4$$

$$= 50 - 320 + 480 - 48$$

$$= 162.$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= 40 - 3(20)(2) + 2(2)^3$$

$$= 40 - 120 + 16$$

$$= -64.$$

Date: / / Page no:

Q.16 Calculate quartile coefficient of skewness for the following distribution.

Variable :	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency :	12	30	45	65	36	25	18

Sol.

Variable	Frequency	cf
10-20	12	12
20-30	30	42
30-40	45	87
40-50	65	152
50-60	36	188
60-70	25	213
70-80	18	231

Quartile coefficient of Skewness $= \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

$N = 231$, $N/2 = \dfrac{231}{2} = 115.5$

So, median class will be 40-50.

$l = 40$, $h = 10$, $N/2 = 115.5$, $cf = 87$, $f = 65$.

$$Q_2 = l + \left(\frac{N/2 - cf}{f}\right) \times h$$

$$= 40 + \left(\frac{115.5 - 87}{65}\right) \times 10$$

$$Q_2 = 40 + \left(\frac{28.5 \times 10}{65}\right)$$

$$= 40 + \left(\frac{285}{65}\right)$$

$$= 40 + 4.38$$

$$= 44.38.$$

Now, $N/4 = 57.75.$

So, lower quartile class = 30-40.

$l = 30,\ N/4 = 57.75,\ cf = 42,\ f = 45,\ h = 10.$

$$Q_1 = l + \left(\frac{N/4 - cf}{f}\right) \times h$$

$$= 30 + \left(\frac{57.75 - 42}{45}\right) \times 10$$

$$= 30 + \left(\frac{15.75 \times 10}{45}\right)$$

$$= 30 + \left(\frac{157.5}{45}\right)$$

$$= 30 + 3.5$$

$$= 33.5.$$

Now, $3N/4 = \frac{3(231)}{4} = 173.25$

So, Upper quartile class = 50-60

Date: / / Page no:

$l = 50,\ 3N/4 = 173.25,\ cf = 152,\ f = 36,\ h = 10$

$$Q_3 = l + \left(\frac{3N/4 - cf}{f}\right) \times h$$

$$= 50 + \left(\frac{173.25 - 152}{36}\right) \times 10$$

$$= 50 + \left(\frac{21.25 \times 10}{36}\right)$$

$$= 50 + \left(\frac{212.5}{36}\right)$$

$$= 50 + 5.9$$

$$= 55.9.$$

So, Quartile coefficient of Skewness $= \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

$$= \frac{55.9 + 33.5 - 2(44.38)}{55.9 - 33.5}$$

$$= \frac{89.4 - 88.76}{22.4}$$

$$= \frac{0.64}{22.4}$$

$$= 0.02 \text{ (approx)}.$$

Date / / Page no.

Q.17. Find the probability of drawing an ace or a spade or both from a pack of 52 cards.

Sol. Total no. of possible outcomes = 52.

Probability of drawing an ace or a spade or both from a deck of card is:

The total number of spades = 13.

Probability of drawing spades $P(A) = \frac{13}{52}$.

Now, the total number of aces = 4

Probability of drawing aces $P(B) = \frac{4}{52} = \frac{1}{13}$

The total number of aces of spade is 1.

$$P(A \cap B) = \frac{1}{52}.$$

Probability of drawing an ace or a spade or both from a deck of card is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \frac{13}{52} + \frac{1}{13} - \frac{1}{52}$$

$$= \frac{13 + 4 - 1}{52}$$

$$= \frac{16}{52} = \frac{4}{13}.$$

Date: / / Page no:

Q.18. A bag contains 8 white and 6 red balls. Find the probability of drawing two balls of the same colour.

Sol. A bag contains 8 white and 6 red balls.
Total no. of possible outcomes = 14.

then, sample space is $^{14}C_2 = 91$.

From 8 white balls drawing 2 is $^8C_2 = 28$

Similarly, from 6 red balls drawing 2 is $^6C_2 = 15$.

$$\left[{}^8C_2 + {}^6C_2 / {}^{14}C_2 \right]$$

$= 28 + 15/91$

$= \frac{43}{91}$.

So, the probability of drawing two balls of same colour is $43/91$.

Q.19. A problem is given to three student A, B and C whose chance of solving it are $\frac{1}{2}, \frac{1}{3}$ and $\frac{1}{4}$. respectively. What is the probability that the problem will be solved?

Sol. Given :- The probability of A, B and C solving a problem are :-

$P(A) = \frac{1}{2}$, $P(B) = \frac{1}{3}$ and $P(C) = \frac{1}{4}$

Find :- The probability of problem be solved = ?

Sol : let $\bar{A}, \bar{B}$ and $\bar{C}$ be the representative events of not solving the problem.

Then, A, B and C are independent event

$\therefore$, $\bar{A}, \bar{B}$ and $\bar{C}$ are also independent event.

Now, $P(\bar{A}) = 1 - P(A)$

$= 1 - \frac{1}{2}$

$= \frac{1}{2}$

$P(\bar{B}) = 1 - P(B)$

$= 1 - \frac{1}{3}$

$= \frac{2}{3}$

$P(\bar{C}) = 1 - P(C)$

$= 1 - \frac{1}{4}$

$= \frac{3}{4}$

$\therefore$, P(not solving the problem) $= P(\bar{A} \cap \bar{B} \cap \bar{C})$

$= P(\bar{A})\, P(\bar{B})\, P(\bar{C})$

[$\because$, $\bar{A}, \bar{B}, \bar{C}$ are independent events]

Date: / / Page no:

$$= \left(\frac{1}{2}\right)\left(\frac{2}{3}\right)\left(\frac{3}{4}\right)$$

$$= \frac{1}{4}$$

So, P(the problem will be solved) = 1 − P(not solving the problem)

$$= 1 - \frac{1}{4}$$

$$= \frac{3}{4}$$

Ans. Hence, the probability of problem of will be solved is $\frac{3}{4}$.

Q.20. In a lottery, m tikets are drawn at a time out of n tickets numbered from 1 to n. Find the expected value of the sum of the numbers on the tickets drawn.

Sol. Let $x_1, x_2 \ldots x_n$ be the variables representing the numbers on the first, second, ..., n^{th} ticket.

The probability of drawing a ticket out of n tickets being in each case $1/n$, we have

$$E(x_i) = 1.\frac{1}{n} + 2.\frac{1}{n} + 3.\frac{1}{n} + \ldots + n.\frac{1}{n}$$

$$= \frac{1}{2}(n+1)$$

∴, expected value of the sum of the numbers on the tickets drawn.

Solving problem)

$$= E(x_1 + x_2 + \dots + x_m)$$

$$= E(x_1) + E(x_2) + \dots + E(x_m)$$

$$= m E(x_i)$$

$$= \frac{1}{2} m(n+1).$$

Q21. A variate X has the probability distribution :

x :	-3	6	9
$P(X=x)$:	$1/6$	$1/2$	$1/3$.

Find $E(X)$ and $E(X^2)$. Hence evaluate $E(2X+1)^2$.

Sol. $E(X) = -3 \times \frac{1}{6} + 6 \times \frac{1}{2} + 9 \times \frac{1}{3}$

$$= \frac{11}{2}$$

$$E(X^2) = 9 \times \frac{1}{6} + 36 \times \frac{1}{2} + 81 \times \frac{1}{3}$$

$$= \frac{93}{2}$$

$\therefore$, $E(2X+1)^2 = E(4X^2 + 4X + 1)$

$$= 4E(X^2) + 4E(X) + 1$$

$$= 4\left(\frac{93}{2}\right) + 4\left(\frac{11}{2}\right) + 1 = 209.$$

Date: / / Page no:

Q.22. Prove that if $E[X] = 0$, then $Var[X] = E[X^2]$

Sol. Given : $E[X] = 0$

To prove : $Var[X] = E[X^2]$

We know that, $Var[X] = E(X^2) - [E[X]]^2$

If $E[X] = 0$

$$Var[X] = E[X^2] - 0$$

$$Var[X] = E[X^2].$$

From the definition of the variance of X also, we can write :

$$Var[X] = \Sigma [x - E(x)]^2 p(x)$$

If $E[X] = 0$.

So, $Var[X] = \Sigma x^2 p(x)$

$$Var[X] = E[X^2]$$

Hence, proved.

Q.23. Prove that if X and Y are independent then $E[XY] = E[X] \times E[Y]$

Date: / / Page no:

Sol. For discrete random variables X and Y.

$$E(XY) = \sum_i \sum_j x_i y_j f_{xy}(x_i, y_j) = \sum_i \sum_j x_i y_j f_x(x_i) f_y(y_j)$$

$$= \left(\sum_i x_i f_x(x_i)\right)\left(\sum_j y_j f_y(y_j)\right)$$

$$= E[X]\,E[Y]$$

Hence, Proved.

Q24. In 256 sets of 12 tosses of a coin, in how many cases one can ~~acc~~ expect 8 heads and 4 tails.

Sol. $P(\text{head}) = \frac{1}{2}$ and $P(\text{tail}) = \frac{1}{2}$

By binomial distribution, probability of 8 heads and 4 tails in 12 trials is.

$$P(X=8) = {}^{12}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^4$$

$$= \frac{12!}{8!\,4!} \cdot \frac{1}{2^{12}} = \frac{495}{4096}$$

∴, the expected number of such cases in 256 sets

$$= 256 \times P(X=8)$$

Date: / / Page no:

$= 256 \times \frac{495}{4056}$

$= 30.9$

$= 31.$

Q.25. The probability that a pen manufactured by a company will be defective is 0.1. If 12 such pens are manufactured, find the probability that.

(i) Exactly two will be defective.

(ii) At least two will be defective.

(iii) None will be defective.

Sol. The probability of a defective pen is 0.1.

∴, The probability of a non-defective pen is $= 1 - 0.1$
$= 0.9.$

(a). The probability that exactly two will be defective

$= {}^{12}C_2 (0.1)^2 (0.9)^{10}$

$= 0.2301$

(b). The probability that at least two will be defective.

$= 1 - ($ prob. that either none or one is non-defective)

$= 1 - \left[{}^{12}C_0 (0.9)^{12} + {}^{12}C_1 (0.1)(0.9)^{11}\right]$

$= 0.3412$.

(c). The probability that none will be defective.

$= {}^{12}C_{12} (0.9)^{12}$

$= 0.2833$.

Q.27. The probability that a bomb dropped from a plane will strike the target is $1/5$. If six bombs are dropped, find the probability that,

(i) Exactly two will strike the target.

(ii). At least two will strike the target.

Sol. (a). $P(X=2) = {}^6C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^4$

$= \frac{\cancel{6}^{3} \times 5}{\cancel{2} \times 1} \times \left(\frac{1}{\cancel{25}_{5}}\right)\left(\frac{256}{625}\right)$

$= \frac{768}{3125}$

$= 0.246$.

(b). $1 -$ [at most one will strike target]

$\Rightarrow 1 - [P(X=0) + P(X=1)]$

$= 1 - \left[{}^6C_0 \times \left(\frac{1}{5}\right)^0 \times \left(\frac{4}{5}\right)^6 + {}^6C_1 \times \left(\frac{1}{5}\right)^1 \times \left(\frac{4}{5}\right)^5\right]$

$$= 1 - \left[\frac{4096}{15625} + 6 \times \frac{1024}{15625}\right]$$

$$= 1 - \left[\frac{4096}{15625} + \frac{6144}{15625}\right]$$

$$= 1 - \left[\frac{10240}{15625}\right]$$

$$= \frac{15625 - 10240}{15625}$$

$$= \frac{5385}{15625}$$

$$= +0.345$$

Q.28. If the probability of a bad reaction from a certain infection is 0.001, determine the chance that out of 2,000 individuals more than two will get a bad reaction.

Sol. ~~If~~ It follows a Poisson distribution as the probability of occurrence is very small.

Mean $m = np = 2000(0.001) = 2$.

Probability that more than 2 will get a bad reaction

= 1 − [Prob. that no one gets a bad reaction +

Date: / / Page no:

prob. that one gets a bad reaction + prob. that two get bad reaction]

$$= 1 - \left[e^{-m} + \frac{m^1 e^{-m}}{1!} + \frac{m^2 e^{-m}}{2!} \right]$$

$$= 1 - \left[\frac{1}{e^2} + \frac{2}{e^2} + \frac{2}{e^2} \right] \qquad [\because, m = 2]$$

$$= 1 - \frac{5}{e^2}$$

$$= 0.32.$$

Q.29. Fit a Binomial distribution to the following data.

x:	0	1	2	3	4	5
Y:	2	14	20	34	22	8

Sol. Here, $n = 5$ and $N = \Sigma f_i = 100$

$$\text{mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{0 + 14 + 40 + 102 + 88 + 40}{100}$$

$$= \frac{284}{100}$$

$$= 2.84.$$

Mean of binomial distribution $= np = 5p$
$= 2.84$

$$5p = 2.84$$

$$p = 2.84/5 = 0.568$$

$q = 1 - p = 1 - 0.568 = 0.432$.

i.e., Hence, the binomial distribution to be fitted is:

$N(q+p)^n = 100(0.568 + 0.432)^5$ Ans.

$= 100 \times {}^5C_0 (0.432)^5 + 100\, {}^5C_1 (0.432)^4 (0.568)$

$+ 100\, {}^5C_2 (0.432)^3 (0.568)^2 + \dots$

$100\, {}^5C_5 (0.568)^5$.

$= 1.5 + 9.89 + 26.01 + 34.19 + 22.48 + 5.91$

$= 99.98$.

Q.30. A certain screw making machine produces on average 2 defective screws out of 100, and packs them in boxes of 500. Find the probability that a box contains 15 defective screws.

Sol. Given, out of 100 screws an average of 2 defective screws is produced by a certain screw making machine. So the probability will be

$2/100 = 0.02$.

So, we have poisson's distribution that is events occurring in a given time period.

Now, $n = 500$ and $p = 0.02$.

$\therefore$ mean $= \lambda = np = 500 \times 0.02$
$= 10$.

Now, the probability that a box contains 15 defective screws is :

$= \lambda^{15}/15! \times e^{-\lambda}$

$= 10^{15}/15! \times e^{-10}$.

$= 0.035$.

Q.31. Fit a poisson distribution to the following data :

x :	0	1	2	3	4
Y :	192	100	24	3	1

Sol. Mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = \dfrac{0+100+48+9+4}{192+100+24+3+1}$

$= \dfrac{161}{320} = 0.5$.

$\therefore$ mean of Poisson distribution, $m = 0.5$

Hence, the theoritical frequency for r successes is

Date: / / Page no:

$$\frac{N e^{-m} (m)^r}{r!} = \frac{320 \times e^{-0.5} (0.5)^r}{r!}$$

where $r = 0, 1, 2, 3, 4$.

∴, the theoritical frequencies are:

x:	0	1	2	3	4
y:	192	96	24	4	0.5

Q.32. The mean height of 500 male students in a certain college is 151 cm and the standard deviation is 15 cm. Assuming the heights are normally distributed, find how many students have heights between 120 and 155 cm?

Sol. Here, $\mu = 151$ and $\sigma = 15$.

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 151}{15}$$

When, $X = 120$, $Z = \frac{120 - 151}{15} = -2.067$.

When $X = 155$, $Z = \frac{155 - 151}{15} = 0.2667$.

When $X = 185$, $Z = \frac{185 - 151}{15} = \frac{34}{15} = 2.2667$

$$P(120 < X < 155) = P(-2.07 < Z < 0.27)$$

$$= \int_{-2.067}^{0.2667} \phi(z)\,dz$$

$$= \int_{0}^{0.2667} \phi(z)\,dz + \int_{0}^{2.0667} \phi(z)\,dz$$

$$= 0.4803 + 0.1026$$

$$= 0.5829$$

Here, $N = 500$

$\therefore$ The number of students whose weight is between 120 and 155 Pounds is

$$0.5872 \times 500 = 291.$$

Q33. A set of five similar coin is tossed 320 times and the result is :-

No. of heads :	0	1	2	3	4	5
Frequencies :	6	27	72	112	71	32

Test the hypothesis that the data follow a binomial distribution?

Sol.

For $v = 5$, we have $\chi^2_{0.05} = 11.07$

p, probability of getting head $= \frac{1}{2}$;

q, probability of getting tail $= \frac{1}{2}$

Hence, the theoritical frequencies of getting 0, 1, 2, 3, 4, 5 heads are the successive terms of the binomial expansion :

$$320(p+q)^5$$

$$= 320\left(p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5\right)$$

$$= 320\left[\frac{1}{32} + \frac{5}{32} + \frac{10}{32} + \frac{10}{32} + \frac{5}{32} + \frac{1}{32}\right]$$

$$\cancel{= 320\left(\frac{32}{3}\right.} = \overset{10}{\cancel{320}}\left(\frac{1+5+10+10+5+1}{\underset{1}{\cancel{32}}}\right)$$

$$= 10 + 50 + 100 + 100 + 50 + 10$$

Thus, the theoritical frequencies are 10, 50, 100, 100, 50, 10

Hence,

$$\chi^2 = \frac{(6-10)^2}{10} + \frac{(27-50)^2}{50} + \frac{(72-100)^2}{100} + \frac{(112-100)^2}{100} + \frac{(71-50)^2}{50} + \frac{(32-10)^2}{10}$$

$$= \frac{1}{100}\left(160 + 1058 + 784 + 144 + 882 + 4840\right)$$

$$= \frac{7868}{100}$$

$$= 78.68$$

and df $v = 6 - 1 = 5$

Since, the calculated value of χ^2 is much greater than $\chi^2_{0.05}$, the hypothesis that the data follow the binomial law is rejected.

Q.34 A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased at 5% level of significance.

Sol. Suppose the coin is unbiased.

Then, the probability of getting the head in a toss $= \frac{1}{2}$

$\therefore$ expected number of successes $= \frac{1}{2} \times 400 = 200$.

and the observed value of successes $= 216$.

Thus, the excess of observed value over expected value $= 216 - 200 = 16$.

Also, S.D. of simple sampling $= \sqrt{npq}$

$$= \sqrt{\left(400 \times \frac{1}{2} \times \frac{1}{2}\right)}$$

$$= 10.$$

Hence, $z = \frac{x - np}{\sqrt{npq}} = \frac{16}{10} = 1.6$

As $z < 1.96$, the hypothesis is accepted at 5%.

level of Significance i.e, we conclude that the coin is ~~x~~ unbiased at 5% level of significance

Q35. A sample of 900 members is found to have a mean of 3.4 cm. Can it be reasonably regarded as a truly random sample from a large population which mean 3.25 cm and S.D. 1.61 cm.

Sol. Here $\bar{x} = 3.4$ cm, $n = 900$, $\mu = 3.25$ and $\sigma = 1.61$ cm

$$\therefore z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{3.4 - 3.25}{1.61 / \sqrt{900}}$$

$$= \frac{0.15}{1.61 / \sqrt{30}}$$

$$= 2.8$$

As, $z > 1.96$, the deviation of the sample mean from the mean of the population is significant at 5% level of significance. Hence, it cannot be regarded as a random sample.