



17BA102 STATISTICS FOR MANAGEMENT

Academic Year : 2018-19	QUESTION BANK	Programme : MBA
Year / Semester : I / I		Course Coordinator: Mrs. THARANI & Ms.R.SEETHA

Course Objectives	Course Outcomes
1. To enable the students to have an insight into basic statistical techniques. 2. To gain an understanding about hypothesis testing, using tools such as Z test, F test, ANOVA, chi- square tests, regression & correlation analysis and non-parametric tests. 3. To enable the students to draw conclusions from the analysis for better decision making.	On completion of the course, students will be able to CO1: Compare the collected information using descriptive and inferential statistics to interpret the collected data.(K2) CO2: Apply the sampling theory in accounting and auditing from business point of view.(K3) CO3: Use univariate and bivariate statistical tools for analysis of collected data. (K2) CO4: Apply non-parametric test for drawing meaningful conclusions.(K3) CO5: Apply correlation and simulation techniques for business decision making. (K3) CO6: Interpret the changes in economic phenomena of future events using time series (K2)

PART – A (2 MARK QUESTIONS WITH KEY) QUESTIONS

S.No	QUESTIONS	Mark	COs	BTL
UNIT I – INTRODUCTION				
1	Define Statistics? How is statistics used in everyday life? Statistics are usually defined as: A collection of numerical data that measure something, The science of recording, organizing, analyzing and reporting quantitative information. Statistics are everywhere, election predictions are statistics, anything food product that says they x% more or less of a certain ingredient is a statistic. Life expectancy is a statistic. If you play card games card counting is using statistics. There are tons of statistics everywhere you look.	1	1	K1
2	State Baye’s theorem on rule of inverse probability. If B_1, B_2, \dots, B_n be a set of exhaustive and mutually exclusive events and A be another event associated with B_i then $p(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A/B_i)}$	2	1	K1
3	Define mutually exclusive and exhaustive events. Two events are said to be mutually exclusive if the occurrence of any one of them excludes the occurrence of other in a single experiment. All possible outcomes of an experiment are called exhaustive events.	1	1	K1
4	The price of a selected stock over a five day period is shown as 170, 110, 130, 170 and 160. Compute Mean, median and mode $n=5, \text{ mean} = \frac{\sum x}{n} = \frac{740}{5} = 148$ median = $((n+1)/2)^{\text{th}}$ term 160, mode = repeated value=170	1	1	K2
5	If $P(A/B) = 0.2$ and $P(B)=0.4$ then find $P(A \cap B)$. $P(A \cap B) = P(A/B) \times P(B)$ $P(A/B) \times P(B) = 0.4 \times 0.2 = 0.08$	1	1	K2



6	What is the mean and variance of Binomial distribution? mean= np, variance = npq	1 1	1	K1
7	When do you say two events are independent? $P(A \cap B) = P(A) \times P(B)$	2	1	K1
8	What is a random experiment? An experiment is said to be a random experiment, if it's out-come can't be predicted with certainty.	2	1	K1
9	Define a sample space. The set of all possible out-comes of an experiment is called the sample space. It is denoted by 'S' and its number of elements are n(s).	2	1	K1
10	List the properties of Normal distribution. (i) The mean of a normally distributed population lies at the centre of the normal curve (ii) The mean, median and mode of the normal distribution coincide. (iii) The curve is asymptotic at x axis on both sides. (iv) The normal curve is bell shaped and symmetrical about the vertical line through the centre	2	1	K1
11	Give example of continuous and discrete variable. A discrete random variable is characterized by jumps and gaps between one value and the next. For example, the no. of tables in a Restaurant. These variables which can take all possible values in a given specified range are termed as continuous variables. For example, the age of students in a school, heights (in cm), weight (in kg) etc.	1 1	1	K2
12	A coin is tossed 3 times. Find the sample space and probability of two heads. $n(s) = 8$ $P(\text{two heads}) = 3/8$	1 1	1	K2
13	Given x is distributed in Poisson fashion. Find E (x) if $E(x^2) = 6$ $E(x^2) = \lambda + \lambda^2 = 6; \lambda = 2$ $E(x) = 2$	1 1	1	K2
14	How to organize a data in retrievable format? Qualitative, Quantitative, Chronological, Geographical	1	1	K1
15	From a pack of cards, one card is drawn . What is the probability that it is either a spade or a king? $P(\text{Spade card}) = 13/52, P(\text{king card}) = 4/52, P(\text{Spade} \cap \text{King}) = 1/52$ $P(\text{Either spade or king}) = P(A \cup B) = 4/13$	1 1	1	K2

UNIT II – SAMPLING DISTRIBUTION AND ESTIMATION

1	What is sampling distribution? Suppose that we draw all possible samples of size n from a given population. Suppose further that we compute a statistic (mean, proportion, standard deviation) for each sample. The probability distribution of this statistic is called Sampling Distribution.	2	2	K1
2	Give two rules for determining sample size Sample size $n = S^2 Z^2 / E^2$ if standard deviation is given Sample size $n = pq Z^2 / E^2$ if p, q values are given	1 1	2	K1
3	What are qualities of good estimator? Unbiasedness, Consistency, Efficiency, Sufficiency	2	2	K1
4	What are the advantages of Sampling? A Sample survey is cheaper than a census survey. More economical, Analysis of the results can be carried out speedily.	1 1	2	K1



5	What is point estimate? A point estimate is a single number , which is used to estimate an unknown population parameter. It is the common way of expressing an estimate. In other words, the point estimate does not give any idea about the reliability or precision of the method of estimation used.	2	2	K1
6	A automobile repair shop has taken a random sample of 40 services that the average service time on an automobile is 130 mins with a standard deviation of 26 min. Find the standard error of the mean. 16. $n = 40, \bar{x} = 130, \sigma = 26,$ $S.E = \sigma / \sqrt{n} = 4.11$	1	2	K2
7	Define Strata. Groups within a population formed in such a way that each group is relatively homogeneous but wider variability exists among the separate groups.	2	2	K1
8	List the various sampling techniques Simple random sampling, systematic sampling, stratified sampling, Cluster sampling, Judgment sampling and Quota sampling.	2	2	K1
9	What is cluster sampling? A sample design in which a cluster of elements is the primary sampling unit instead of individual element in the population	2	2	K1
10	What are non sampling errors? A non sampling error primarily arise at the stages of observation approximation and processing of data and are thus present in both the complete enumeration and the sample survey.	2	2	K1
11	Define unbiased estimator. An estimator is said to be unbiased if its expected value is equal to the population parameter it estimates.	2	2	K1
12	Define standard error of mean and confidence interval of the mean $S.E(\bar{x}) = \sigma / \sqrt{n}$ Confidence interval = $\bar{x} \pm S.E(\bar{x}) Z_{\alpha}$	1	2	K1
13	State central limit theorem Assumptions $\bar{x} \sim N(\mu, \sigma^2/n)$	1	2	K1
14	What are the ways to decrease the Standard error. Standard Error can be minimized by increasing the sample size Sampling error can be eliminated by increasing the efficiency of sampling	1	2	K1
15	Define interval estimate. The estimate of an interval in which an unknown population characteristic is expected to lie for a given level of significance.	2	2	K1

UNIT III – TESTING OF HYPOTHESIS - PARAMETRIC TESTS

1	Define Hypothesis A Hypothesis is a proposition or a statement that we would like to verify whether it is true or not.	2	3	K1
2	Differentiate one tailed and two tailed test. One tail test: A statistical hypothesis test in which the alternative hypothesis is specified such that only one direction of the [possible values is considered. Two tail test: A statistical hypothesis test in which the alternative hypothesis is specified such that it includes both the higher and the lower values of a parameter than the value specified in the null hypothesis.	1	3	K2
3	Define critical region. The set of values of the test statistic that will cause us to reject the null hypothesis.	2	3	K1



4	Define Type I and Type II error? Type I Error: An error caused by rejecting a null hypothesis that is true Type II Error: An error caused by accepting null hypothesis, that is false.	1 1	3	K1															
5	List out the different applications of t-test. T test is applied for small samples i.e n<30 It is used to test whether there is any significant difference between sample mean and population mean.	1 1	3	K1															
6	What is Null Hypothesis and Alternative Hypothesis? A Hypothesis of no difference is called Null Hypothesis. It is denoted by H ₀ Alternative hypothesis is opposite of null Hypothesis. It is denoted by H ₁	1 1	3	K1															
7	Write down the formula to test the difference of two means in case of large sample. $Z = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{(s_1^2/n_1 + s_2^2/n_2)}}$ Where \bar{x}_1 and \bar{x}_2 are sample mean, n ₁ , n ₂ are sample size and S ₁ , s ₂ are sample standard deviation.	1 1	3	K1															
8	What is ANOVA? Why is ANOVA helpful? Analysis of Variance (ANOVA) is a statistical technique used to test the equality of three or more sample means. It is helpful to test the homogeneity of several means.	1 1	3	K1															
9	What is level of significance? Maximum size of the type I error to face risk is called Level of significance. Type I error is reject H ₀ , when H ₀ is true.	1 1	3	K1															
10	Give test statistic for testing significance of difference between the variance of two small sample. $S_1^2 = \frac{\sum(x_1 - \bar{x}_1)^2}{n_1 - 1}$ $S_2^2 = \frac{\sum(x_2 - \bar{x}_2)^2}{n_2 - 1}$ F test : $F = \frac{S_2^2}{S_1^2}$	1 1	3	K1															
11	Write the table of one way ANOVA																		
	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Source</th> <th>Sum of squares</th> <th>DoF</th> <th>Mean</th> <th>F ratio</th> </tr> </thead> <tbody> <tr> <td>Betw sample</td> <td>SSR</td> <td>R-1</td> <td>MSR = SSR/R-1</td> <td>F = MSR/MSE</td> </tr> <tr> <td>Within Sample</td> <td>SSE</td> <td>N-R</td> <td>MSE = SSE/N-R</td> <td>(or) MSE/MSR</td> </tr> </tbody> </table>	Source	Sum of squares	DoF	Mean	F ratio	Betw sample	SSR	R-1	MSR = SSR/R-1	F = MSR/MSE	Within Sample	SSE	N-R	MSE = SSE/N-R	(or) MSE/MSR	2	3	K1
Source	Sum of squares	DoF	Mean	F ratio															
Betw sample	SSR	R-1	MSR = SSR/R-1	F = MSR/MSE															
Within Sample	SSE	N-R	MSE = SSE/N-R	(or) MSE/MSR															
12	State the basic principles of Experimental Design. Replication, Randomization and Local control.	2	3	K1															
13	What are the assumptions based on ANOVA? The observations are independent; The samples are drawn from normal population. Various treatment and environmental effects are additive in nature.	1 1	3	K1															
14	Define Degrees of freedom. The number of elements that can be chosen freely is called Degrees of freedom.	2	3	K1															
15	Define test statistic. The value of z or t calculated for a sample statistic such as the sample mean or the sample proportion is known as test statistic.	3		K1															

UNIT IV – NON-PARAMETRIC TESTS

1	What are the applications of chi- square test? Test of goodness of fit Test of independence of attributes, Test of Homogeneity	1 1	4	K1
2	Differentiate Kruskal Wallis test and Mann Whitney U test. Kruskal wallis is used test whether 3 or more populations are individual. Mann Whitney test is used to determine whether two different samples come from identical population or these populations have different means.	1 1	4	K1
3	What is non parametric test? Name any four non parametric test. Test that rely less on parameter estimation and /or assumptions about the shape of a population distribution. ch square, H test, U test, run test	1 1	4	K1



4	How do find the degrees o freedom in case of chi square test? Degrees of freedom = n-1 Degrees of freedom = (r-1) (c-1)	1 1	4	K1
5	Explain the term “run” with an example. A sequence of identical occurrences that may be preceded and followed by different occurrences. Example: The response from 10 school boys for the question “ whether the boys prefer their home near to school” - y y y n n y n y y	1 1	4	K2
6	Kruskal Wallis test is the non parametric ANOVA. Why? Kruskal wallis is used test whether 3 or more populations are individual. It is a non parametric test, whereas ANOVA is a parametric test used to test more than 2 samples.	2	4	K2
7	What are the major advantages of non parametric methods over parametric methods? No assumptions are required, more suitable when ranked or scaled. Do not take much time, simple calculations.	1 1	4	K1
8	Define Kolomogrov simirnnov test A non parametric test that is concerned with the degrees of agreement between a set of observed ranks and a theoretical frequency distribution.	2	4	K1
9	Two HR managers A and B ranked 5 candidates for a new position. Their rankings of the candidates are shown below. Compute the Spearman’s Rank Correlation Candidate name: Nancy Linda Oviya John Mary Rank by A: 2 1 3 5 4 Rank by B: 1 3 4 5 2 $\Sigma d^2 = 10$, $r = 1 - 6 (\Sigma d_i^2) / (n (n^2 - 1))$ $r = 0.5$ positive correlation	1 1	4	K2
10	Write the formula for chi – Square test of single standard deviation. Chi square = $\Sigma ((O_i - E_i)^2 / E_i)$	2	4	K1
11	Define one sample Runs test A non parametric test used for determining whether the items in a sample selected randomly.	2	4	K1
12	Write the standard error of the U statistic of the Mann Whitney U test. $\sigma = \sqrt{(n_1 n_2 (n_1 + n_2 + 1) / 12)}$	2	4	K1
13	A chi square value is never negative. Give reason. Differences between observed and expected frequencies are squared.	1	4	K2
14	A contingency table for a chi square test has 5 columns and 4 rows . how many degrees of freedom should be used? DoF = r-1 x c – 1 DoF = 12	1 1	4	K2
15	Disadvantages of non parametric test. Based on limited amount of information , it do not make use of all the available information. A good deal of information is lost.	2	4	K2

UNIT V – CORRELATION, REGRESSION AND TIME SERIES ANALYSIS

1	Differentiate Correlation and Regression. correlation coefficient are between X and Y is a measure of linear relationship between X and Y. It does not reflect upon the nature of variable Regression coefficients are mathematical measures expressing the average relationship between the two variables. It reflects on the nature of variable.	1 1	5	K2
2	What is cyclical and seasonal variation? Cyclical variations in a time series are the recurrent variations whose duration is more than one year. Seasonal variations are variations which occur with some degree of regularity within a specific period of one year or less.	1 1	5	K1



3	How will you test the accuracy of regression equation? If one of the regression coefficients is greater than unity the other must be less than unity. Both the regression lines pass through the point (\bar{x}, \bar{y}) are means of x and y.	1 1	5 5	K2 K2
4	What is non sense correlation? Correlating irrelevant things is a non sense correlation. When $r = 0$, the correlation between x and y is said to be non sense correlation.	2	5	K1
5	Find the mean values of the variables X & Y from the following regression lines $2y - x = 50$; $3y - 2x = 10$ solve equations $\bar{x} = 130, \bar{y} = 190$	1 1	5 5	K2 K2
6	Why do we use multiple regressions instead of simple regression in estimating a dependent variable? When we deal with more than one dependent variable we use multiple regression.	2	5	K2
7	What are the changes or variations involved in time series analysis? Secular trend, seasonal variations, cyclical variations and irregular variations.	2	5	K1
8	What is meant by forecasting errors? Mean absolute deviation $MAD = \sum y_i - \bar{y}_i / n$ where y_i = actual value \bar{y}_i = fitted value	1 1	5 5	K1 K1
9	State the major application of Time series analysis. Analysis of Time series is helpful in economics, Business and science It helps in understanding past behavior of a variable	1 1	5 5	K1 K1
10	What are the methods for measuring cyclical variations? Residual method, Direct method, Reference cycle analysis method, harmonic analysis method.	2	5	K1
11	Give some examples of time series. sales, production, import , export, population over the last 5 or 10 years.	2	5	K2
12	What is regression coefficient? $b_{xy} = (n\sum xy - \sum x \sum y) / (n\sum y^2 - (\sum y)^2)$, $b_{yx} = (n\sum xy - \sum x \sum y) / (n\sum x^2 - (\sum x)^2)$	1 1	5 5	K1 K1
13	What are the uses of Regression analysis? It can be useful to all natural, social and physical sciences where the data are in functional relationship. It helps in prediction and it can estimate the values of unknown quantities.	1 1	5 5	K1 K1
14	What are the time series forecast error measures? Root mean square error (RMSE) Mean absolute Deviation (MAD) Mean Absolute Percentage Error (MAPE)	1 1 1	5 5 5	K1 K1 K1
15	List the Merits and demerits of least Square method. Merits: this is mathematical method and is completely objective in character. This method gives trend value for the entire time period. Demerits: Time consuming, Tedious. The type of curve to be fitted has to be selected carefully.	1 1	5 5	K1 K1

PART – B (12 Mark Questions with Key)

S.No	Questions	Mark	COs	BTL
UNIT I – INTRODUCTION				
1	A sales representative can convert a customer as a potential buyer with the probability of 70% . If he is able to meet 10 customers in a day , Find the probability of (i) Atleast one customer (ii) not even a single customer (iii) exactly one customer	12	1	K2



Binomial distribution $P(X) = {}_n C_x p^x q^{n-x}$			
(i) $P(X \geq 1) = 0.9999$			4
(ii) $P(X = 0) = 0.0000059$			4
(iii) $P(X = 1) = 0.0001378$			4
2 In a bolt factory machines A, B and C manufactures respectively 25 %, 35% and 40%. Of the total of their output 5%, 4%, 2 % are defective bolts. A bolt is drawn at random from the product and found to be defective. What are the probability that it was manufactured by Machines A, B and C			12
Formula : $p(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A/B_i)}$ and assumptions			3
$P(E_1/A) = 25/69$			3
$P(E_2/A) = 28/69$			3
$P(E_3/A) = 16/69$			3
3 In a test of 2000 electric bulbs it was found that the life of a particular make was normally distributed with an average life of 2040 hours and SD of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours (ii) less than 1950 hours (iii) more than 1920 hours but less than 2160 hours			12
Normal distribution: Formula and given data $N=2000$			3
$P(X > 2150) = 0.0336, N = 2000 \times 0.0336 = 67$			3
$P(X < 1950) = 0.0668, N = 134$			3
$P(1920 < X < 2100) = 0.8185, N = 1637$			3
4 Marks of PG class are recorded and the head of the department need to do an analysis based on the student's performance on the particular paper for the improvement of result. The data are grouped in the following table. Find the mean and median. Also find the mark which maximum students obtained.			12
20-30 30-40 40-50 50-60 60-70			
3 5 20 10 5			
Assumed mean = 45, $\Sigma fd = 90$			
Mean $\bar{x} = A + \Sigma fd/n = 47.093$			
Median $m = l_1 + (l_2 - l_1)(m - c)/f = 47$			
Mode = $l_1 + [(f_1 - f_0) / ((f_1 - f_0) + (f_1 - f_2))] = 46.5$			
5 Four coins were tossed 150 times and the following results were obtained			
No of heads: 0 1 2 3 4			
Observed frequency: 28 62 46 10 4			
Fit a binomial distribution.			
Mean = $\Sigma fx / \Sigma f = 1.33$			
Mean = np therefore $p = 0.2667, q = 0.7333$			
$N = 150, P(X) = N \cdot {}_n C_x p^x q^{n-x}$			
0 1 2 3 4			
31.80 57.84 42.07 15.30 2.78			
6 The ages of a sample of 8 faculty members selected from the school of business administration are shown below. Compute average age, mode, median age and S.D			12
Faculty: 1 2 3 4 5 6 7 8			
Age: 42 30 73 50 51 37 42 59			
Mean = $\Sigma f/n = 48$			4
Median = $(n+1/2)$ th term = 46			4
Mode = 42, S.D = $\sqrt{(\Sigma(x - \bar{x})^2/n)} = 12.63$			4

UNIT II – SAMPLING DISTRIBUTION AND ESTIMATION

1 Describe the various technique of sampling.			12
Simple random sampling			3
systematic sampling		2	3
stratified sampling			3



	cluster sampling (any four)	3		
2	a) In a random selection 64 of 600 road crossing in a town, the mean number of automobile accident is 4.2 and sample standard deviation is 0.8 . Construct 95 % confidence interval for the mean number of accidents.			
	b) A cigarette manufacturer wishes to use the random sample to estimate average Nicotine content. Sampling error should be more than 1 milligram. Standard deviation is 4 milligrams. What sample size should the company take at 99 % confidence?	12		
	a) $n=64, N=600, \bar{x}=4.2, s=0.8$	1		
	S.E of mean = $s\sqrt{N-n}/\sqrt{(n-1)(N-1)} = 0.952$	2		
	Confidence interval at 95% = (4.0135, 4.3865)	3		
	(b) $s=4, E=1, Z= 1\% = 2.58$	3		
	$n= S^2z^2/E^2 = 107$	3	2	K3
3	a) A simple random sample of size 100 has mean 15, the population variance being 25. Find an interval estimate of the population mean with a confidence level of 99 % and 95 %			
	b) A random sample of 700 units from a large consignment showed that 200 were damaged. Find	12		
	(i) 95% and (ii) 99% confidence limits for the proportion of damaged units in the consignments.		2	K3
	(a) $n = 100, \bar{x}=15, SD =5, SE \text{ of mean} = s/\sqrt{n} = 0.5,$	3		
	Confidence interval at 99 % =(13.71, 16.29) , at 95 % = (14.02, 15.98)	3		
	(b) $P=x/n = 0.2857, 95\% \text{ confidence interval} = (0.2522, 0.3192)$	3		
	Confidence interval 99% = (0.2416, 0.3298)	3		
4	a) The quality control manager of a tyre company has sample of 100 tyres with mean 30214 km. Population standard deviation is 860 km. Construct 95% confidence interval of the mean life time of tyre.			
	b) A principal of a college wants to estimate the proportion of smokers among students. What size of sample size should be taken to have proportion not exceeding 10% with almost certainty. It is believed from previous records that the proportion of smokers was 0.30.	12		
	(a) $n =100, \bar{x}=30214, s=860, SE = s/ \sqrt{n}=86$	3		
	Confidence interval at 95% = (30045.44, 30382.56)	3		
	(b) $E =10\%, p=0.30, q=0.70$	3		
	$n = pqz^2/E^2 = 189$	3	2	K3
5	a) What do you mean by estimation? Differentiate point estimation and interval estimation.	12		
	b) Explain about estimation and qualities of good estimator.			
	Meaning of estimation	3	2	K3
	Point estimation	3		
	Interval estimation	3		
	Qualities of good estimator	3		
6	Estimate the population mean at 95% confidence interval from the following sample data drawn from normal distribution	12		
	x: 10 20 30			
	f: 6 10 4		2	K3
	$\bar{x} = \Sigma fx/n = 19, \sigma = \sqrt{(\Sigma fx^2 - \bar{x}^2)/n} = 7$	6		
	confidence interval = 11.08, 26.92	6		

UNIT III – TESTING OF HYPOTHESIS - PARAMETRIC TESTS

1	Ten individuals are chosen at random from a population and their heights are found to be in inches. 63,63,66,67,68,69,70,70,71,71. Find the mean heights in the universe is 66 inches	12	3	K2
---	---	----	---	----



	Hypothesis; $n=10, \mu = 66, \bar{x} = 67.8$	2	
	$\Sigma(x - \bar{x})^2 = 81.6, S^2 = \Sigma(x - \bar{x})^2 / n-1 = 9.07, S = 3.01$	4	
	T test = $ \bar{x} - \mu / (S / \sqrt{n}) = 0.18$	4	
	T.V = 2.262, Degrees of freedom = 9, LOS = 5%. Accept Ho	2	
2	The average number of articles produced by two machines per day are 200 and 250 with standard deviation 20 and 25 respectively on the basis of records of 25 days of production. Can you regard both the machines are equally efficient at 1% level of significance?	12	
	Hypothesis, Level of significance = 1%, Given value, $\bar{x}_1 = 200, \bar{x}_2 = 250, s_1=20,$	2	3 K2
	$s_2=25, n_1=25, n_2=25$	2	
	$S^2 = n_1 s_1^2 + n_2 s_2^2 / n_1 + n_2 - 2 = 533.85, S = 23.11$	4	
	T test = $ \bar{x}_1 - \bar{x}_2 / s \sqrt{(1/n_1 + 1/n_2)} = 7.645$	4	
	T.V = 2.576, Reject Ho	2	
3	Time taken by workers in performing a job are given below		
	Method I: 20 16 26 27 23 22 -		
	Method II: 27 33 42 35 32 34 38	12	
	Test at 1% level of significance whether there is any significant difference between the variance of time distribution.		3 K2
	Hypothesis, Level of significance = 5%, Given value, $\bar{x}_1 = 22.33, \bar{x}_2 = 34.43$	2	
	$\Sigma(x_1 - \bar{x}_1)^2 = 81.336, \Sigma(x_2 - \bar{x}_2)^2 = 133.70, S_1^2 = \Sigma(x_1 - \bar{x}_1)^2 / n_1 - 1 = 16.27, S_2^2 = \Sigma(x_2 - \bar{x}_2)^2 / n_2 - 1 = 22.28$	4	
	F test : $f = S_2^2 / S_1^2 = 1.3696,$	4	
	DOF = (6,5), T.V = 4.95 Accept Ho	2	
4	The mean population of a random sample of 400 villagers in Jaipur district was found to be 400 with standard deviation of 12. The mean population of a random sample of 400 villages in Meerut district was found to be 395 with a standard deviation 15 .	12	
	Is the difference between the two district means statistically significant?		3 K2
	$N_1 = 400, n_2 = 400, \bar{x}_1 = 400, \bar{x}_2 = 395, s_1 = 12, s_2 = 15$	2	
	Z test $ \bar{x}_1 - \bar{x}_2 / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$	4	
	$Z = 5.205, T.V = 1.96$ Reject Ho	6	
5	Before an increase in Exercise duty on Tea, 800 persons out of 1000 persons were found to be Tea drinkers and after an increase in exercise duty in Tea 800 people were Tea drinkers in a sample of 1200 people . Test whether there is any significant difference in tea consumption at 5 % level.	12	
	$n_1 = 1000, n_2 = 1200, \bar{x}_1 = 800, \bar{x}_2 = 800, p_1 = 0.8, p_2 = 0.667$	2	3 K2
	$P = 0.7291, q = 0.2709$	2	
	Z test = $ p_1 - p_2 / \sqrt{(pq/n_1 + pq/n_2)} = 6.8349 .$	6	
	T.V = 1.96. Accept Ho	2	
6	In a big city 325 men out of 600 men were found to be smokers. Does this information supports the conclusion that the majority of men in this city are smokers?	12	
	$N = 600, x = 325, p' = x/n = 0.54,$ Hypothesis	4	
	$P = 0.5, q = 0.5, z_{tset} = p' - p / (\sqrt{pq/n}) = 2$	6	
	T.V = 1.96, LOS = 5% . Reject Ho.	2	



UNIT IV – NON-PARAMETRIC TESTS

1 The following table gives the number of air crashes that occurred during various days of the week. Test whether accidents are uniformly distributed over the week.

Days:	Mon	Tue	Wed	Thur	Fri	Sat
No of aircrash :	14	18	12	11	15	14

Hypothesis, $Lo_s = 5\%$
 $E_i = \sum x / n = 14$,
 Chi square = $\sum ((O_i - E_i)^2 / E_i) = 2.13$
 DoF = $n-1 = 5$, T.V = 11.07. Accept H_0 .

12
 4 K3
 2
 2
 6
 2

2 A sample of 12 fathers and their sons height is given below

Father X	65	63	67	64	68	62	70	66	68	67	69	71
Son Y	68	66	68	65	69	66	68	65	71	67	68	70

Calculate Spearman's Rank correlation.
 Arrange data in ascending order and ranking
 $\sum d_i^2 = 72.5$
 $T_x =$ repeated value in X = $m(m^2 - 1) / 12 = 0.5$
 $T_y =$ repeated value in Y = $m(m^2 - 1) / 12 = 0.5$
 $r = 1 - 6 (\sum d_i^2 + T_x + T_y) / (n (n^2 - 1)) = 0.722$

12
 4 K3
 2
 4
 2
 4

3 Caffeine content in two brands of Coffee, Elite and Hot cafe are given below. Test the Hypothesis at 0.05 Level of Significance that the average caffeine contents of two brands are equal against that they are unequal.

Elite	2.1	4.0	6.3	5.4	4.8	3.7	6.1	3.3	-	-
Hotcafe	4.1	0.6	3.1	2.5	4.0	6.2	1.6	2.2	1.9	5.4

Hypothesis, $LOS = 5\%$
 Ascending order arranging and ranking
 $R_1 = 93$, $n_1=8$, $n_2=10$
 $U = n_1 n_2 + (n_1(n_1 + 1) / 2) - R_1 = 23$, $\mu = n_1 n_2 / 2 = 40$, $\sigma = \sqrt{(n_1 n_2 (n_1 + n_2 + 1) / 12)} = 11.25$
 $Z = (U - \mu) / \sigma = 1.51$, T.V = 1.96 Accept H_0

12
 4 K3
 2
 2
 6
 2

4 Use sign test to see if there is a difference between the number of days until collection of an account receivable before and after collection policy. Use 0.05 level of significance

Before	30	28	34	35	40	42	33	38	34	45	28	27	25	41	36
After	34	29	33	32	47	43	40	42	37	44	27	33	30	38	36

$n=14$, no of + =5, no of - =9, $p = 0.357$, $q = 0.643$,
 Hypothesis, $LoS = 5\%$
 Standard error = $\sigma_p = \sqrt{pq / n} = 0.13$
 Limits of acceptance region = $(P H_0 \pm 1.96 \sigma_p) = (0.24, 0.75)$. Accept H_0

12
 4 K3
 2
 2
 4
 4

5 Discuss the association for 1000 school boys between general and mathematical abilities

		General Ability		
		Good	Fair	Poor
Mathematical Ability	Good	44	22	4
	Fair	265	257	178
	Poor	41	91	98

Calculate Expected value.
 Hypothesis, $LoS = 5\%$
 Chi square = $\sum ((O_i - E_i)^2 / E_i) = 69.142$
 DoF = $(r-1)(c-1) = 4$, T.V = 3.841 Reject H_0

12
 4 K3
 2
 2
 4
 4



- 6 **The marks secured in History test of the students in public school and Primary schools are given below**
- Primary: 73 75 83 77 72 69 56 80 68 60 84**
- Public: 70 78 79 81 65 63 74 83 67 76 88 61 64 71 86 48** 12
- Test the Hypothesis at 0.05 LoS that there is no significant difference in the performance of test against the alternative that the performance are significantly different.** 4 K3
- Hypothesis, LOS = 5% 2
- Ascending order arranging and ranking 2
- $R_1 = 154.5, n_1 = 11, n_2 = 16$ 2
- $U = n_1 n_2 + (n_1(n_1 + 1) / 2) - R_1 = 87.5, \mu = n_1 n_2 / 2 = 88, \sigma = \sqrt{(n_1 n_2 (n_1 + n_2 + 1) / 12)} = 20.26$ 4
- $Z = (U - \mu) / \sigma = 0.02, T.V = 1.96$ Accept H_0 2

UNIT V – CORRELATION, REGRESSION AND TIME SERIES ANALYSIS

- 1 **Find the correlation coefficient between X and Y for the data**
- X: 12 9 8 10 11 13 7**
- Y: 14 8 6 9 11 12 3** 12
- (ii) Find the means of the variables from the regression equations**
- $10y = 8x + 66$ and $20x = 9y + 107$**
- $Cov(x,y) = 46, \sigma_y = 5.29, \sigma_x = 9.16$ 4
- $r = cov(x,y) / (\sigma_x \sigma_y), r = 0.949$ 4
- $\bar{X} = 13, \bar{y} = 17$ 4
- 2 **From the following data (i) Find two regression equation (ii) Calculate correlation coefficient (iii) Most likely mark of statistics mark when mark in economics is 30.** 12
- | | | | | | | | | | | |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Economics | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
| Statistics | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |
- $b_{xy} = (n\sum xy - \sum x \sum y) / (n\sum y^2 - (\sum y)^2), b_{yx} = (n\sum xy - \sum x \sum y) / (n\sum x^2 - (\sum x)^2), b_{xy} = 0.233,$ 2
- $b_{yx} = 0.66$ 5 K3
- Mean $\bar{x} = 32, \text{mean } y = 38$ 2
- Regression eqn x on y : $x = 0.233y + 23.14,$ Regression eqn y on x : $y = 0.666x + 16.88$ 4
- Correlation = $r = 0.392$ 2
- Statistics mark $y = 37$ 2
- 3 **Fit a trend line to the following data using free hand method and moving average method** 12
- | | | | | | | | | | | |
|------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| No. of students | 15 | 18 | 17 | 20 | 23 | 25 | 29 | 33 | 36 | 40 |
- 3 yearly moving average: 16.67 18.33 20 22.67 25.67 29 32 36.33 4
- Graph (free hand) 4
- Graph (moving average) 4
- 4 **A computer while calculating correlation r from 25 pairs of observation obtained the following. $n=25, \sum x = 125, \sum y = 100, \sum x^2 = 650, \sum y^2 = 460, \sum xy = 508.$ A recheck showed that he had copied two pairs (6,14) and (8,6) while correct values are (8,12) and (6,8). Obtain the correct value of correlation.** 12
- $\sum x = 125, \sum y = 100, \sum x^2 = 650, \sum y^2 = 436, \sum xy = 520$ 4
- $Cov(x,y) = 0.8, \sigma_x = 1, \sigma_y = 1.2$ 6
- $r = 0.67$ 2



5 Using the method calculate seasonal variations from the data given below

Year	I Quarter	II Quarter	III Quarter	IV Quarter
2011	72	68	80	70
2012	76	70	82	74
2013	74	66	84	80
2014	76	74	84	78
2014	78	74	86	82

5 K2

Total 376 352 414 384 2

Seasonal average 75.2 70.4 82.8 76.8 2

General average = 76.3, seasonal index = seasonal average x 100/ general average = 98.55, 92.26, 108.5, 100.65 4

6 What are the assumptions made by the regression model in estimating the parameters and in significance testing? 12

Analyze In what ways regression analysis can be used? 5 K2

Assumptions in regression model 6

Uses of regression analyzes 6

**PART – C (20 MARK QUESTIONS WITH KEY)
 QUESTIONS**

S.No Mark COs BTL

UNIT I – INTRODUCTION

1 A manufacturing firm produces steel pipes in three plants with daily production of 500, 1000, 2000 with respectively. According to past experience it is known that fraction of defective output is 0.005, 0.008, 0.010. If a pipe is selected from a days total production and found to be defective. What is the probability of that it come from I plant, II plant and III plant? 20

Formula : $p(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A/B_i)}$ and assumptions 2 1 K2

$P(A_1/B) = 0.0007$ 4

$P(A_2/B) = 0.0022$ 4

$P(A_3/B) = 0.0057$ 4

$P(\text{I plant}) = 0.0814, P(\text{II plant}) = 0.2558, P(\text{III plant}) = 0.6627$ 6

2 Global green company has the average early sales in an outlet is Rs. 25 lakhs and S.D is Rs. 5 Lakhs . If the sales follow normal distribution find

(i) Probability of sales greater than Rs. 30 Lakhs

(ii) Probability of sales between Rs. 20 Lakhs and 30 lakhs. 20

(iii) Since 1000 outlets the company owns, Find the number of outlets less than Rs. 15 Lakhs. 1 K2

$\mu = 25, S.D = 5, z = x - \mu / \sigma$ 3

$P(x \geq 30) = p(z \geq 1) = 0.1587$ 5

$P(20 \leq x \leq 30) = P(-1 \leq z \leq 1) = 0.6826$ 5

$P(X < 15) = P(z < -2) = 0.0228. \text{ No of outlets} = 23$ 7

UNIT II – SAMPLING DISTRIBUTION AND ESTIMATION

1 a)A movie maker sampled 55 fans who viewed his master piece movie and asked them whether they had planned to see it again. Only 10 of them believed that the movie was worthy of a second look. Find the Standard error of the population of fans who will view the film a second time. Construct a 90 % confidence interval for this population. 2 K3

20



b) The manager of a shop selling beverages wants to estimate the actual amount of beverages in one litre bottles from a nationally known manufacturer. As per manufacturer’s specification the standard deviation of the volume of the beverage is 0.02 litre . The average amount of beverage per one litre bottle is found to be 0.995 litre on checking 50 bottles. Set up 99 % confidence interval, estimate of the true population average amount of beverage in a one liter bottle. Check whether the manufacturer is genuine in filling the beverage.

(a) $n=55$, $p=0.18$, $q=0.82$ 3

$SE = \sqrt{pq/n} = 0.052$ 4

TV at 90% = 1.645, $P= (0.095, 0.266)$ 3

(b) $n=50$, $\bar{x}=0.995$, $\sigma = 0.02$ 4

T.V = 2.58, Confidence limit = (0.987, 1.003) Reject Ho. 6

2 **a)Television advertisers mistakenly believe that most viewers understand most of the advertisement that they see and hear . In this connection, a research study covering 2300 viewers above the age of 20 years was taken. Each viewer looked at 30 second television advertisement or a part of it. It was found that 1914 viewers misunderstood either the entire advertisement or a part of it. Determine 95% confidence interval for the proportion of all viewers that will misunderstand all are part of the television advertisement used in this study.**

b) a stock market analyst wants to estimate the average return on a certain stock. A random sample of 15 days yields an average return of $\bar{x}=10.37\%$ and a standard deviation of $s = 3.5\%$. assuming a normal population of returns. Give a 95% confidence interval for the average return on this stock.

20

2 K3

a) $\sigma = \sqrt{pq/n}$ 2

$p= 0.832$, $q = 0.1678$ 2

$\sigma = 0.00783$ 2

confidence interval at 95 % = (0.815, 0.845) 4

(b) $\bar{x}=10.37$, $s=3.5$, T.V= 2.145 4

Confidence interval = (8.43, 12.31) 6

UNIT III – TESTING OF HYPOTHESIS - PARAMETRIC TESTS

1 **Four doctors each test 4 treatments for a certain disease and observe the number of days each patient takes to recover. The recovery time in number of days are given as follows.**

Doctor	Treatment			
	A	B	C	D
1	10	14	19	20
2	11	51	17	21
3	9	12	16	19
4	8	13	17	20

20

3 K2

Find the Variance at 5 % level of significance to test whether the sales differ with respect (i) Treatment (ii) Doctor

Hypothesis, $CF = T^2/N = 4795.56$ 4

TSS = 1477.44, SSR = 321.69 4

SSC = 380.69, SSE = 775.06 4

F ratio =1.47(between treatment) $F= 1.24$ (Between doctor) 4

T.V = 3.86 (3,9) (both cases) 4

(i) Accept H_0 (ii) Accept H_0



- 2 To study the performance of 3 detergents and three different water temperatures, the following whiteness readings were obtained using specially designed equipment.

Perform two way ANOVA at 5% level.

Water Temperature	Detergents		
	A	B	C
Cold	57	55	67
Warm	49	52	68
Hot	54	46	58

20
 3 K2

Hypothesis, $CF = T^2/N = 28448.44$

TSS = 439.56, SSR = 73.56

SSC = 304.22, SSE = 61.78

F ratio = 2.38 (between temperature) F = 9.85 (Between detergents)

T.V = 6.94 (both cases)

(i) Accept H_0 (ii) Reject H_0

4
 4
 4
 4
 4

UNIT IV – NON-PARAMETRIC TESTS

- 1 A department store has three sales counters. The manager wants to compare the sales of the three stores over a six day week. The relevant data is as follows:

Counter A sales: 78 62 71 58 73

Counter B sales: 76 85 77 90 87

Counter C sales: 74 79 60 75 80

Compare the equality of mean sales in all the three counters using Kruskal Wallis method.

Hypothesis, LOS 5%, $n = n_1 + n_2 + n_3 = 15$

Arrange in ascending order and ranking

$R_1 = 23, R_2 = 59, R_3 = 38$

$H = 12/n(n+1) [R_1^2/n_1 + R_2^2/n_2 + R_3^2/n_3] - 3(n+1) = 6.54$

T.V = 5.991 Reject H_0

20
 4
 4
 4
 6
 2

- 2 A dietician wants to test three types of diet plans 1,2,3. He selected a homogenous group of 23 persons and placed them into 3 sub groups. Each subgroup trying a different diet plan. Each plan was tried for a period of 30 days. The following observation of weight loses in pounds were recorded for members of each group after this period of 30 days. Determine at 0.05 LoS whether there is a difference in weight reducing effects or not.

Diet plan 1: 4.0 3.8 3.7 6.2 5.6 4.2

Diet plan 2: 3.6 5.2 2.8 3.0 3.8 5.0 3.9 5.5

Diet plan 3: 6.5 7.2 5.9 5.5 6.8 7.7 8.0 8.2 7.0

Hypothesis, LOS 5%, $n = n_1 + n_2 + n_3 = 23$

Arrange in ascending order and ranking

$R_1 = 56.5, R_2 = 52, R_3 = 167.5$

$H = 12/n(n+1) [R_1^2/n_1 + R_2^2/n_2 + R_3^2/n_3] - 3(n+1) = 7.748$

T.V = 5.991 Reject H_0

20
 4
 K3

UNIT V – CORRELATION, REGRESSION AND TIME SERIES ANALYSIS

- 1 Mr. X owns a small company that manufactures portable message tables. Since he started the company, the number of tables he has sold is represented by this time series:

Year: 1987 1989 1990 1991 1992 1993 1996

Tables : 140 144 160 152 168 176 180

Find the linear equation using least square method that describes the trend in the number of tables sold by him. Also calculate trend value for the year 1988

20 5 K2



$\Sigma y = na + b\Sigma x, \Sigma xy = a\Sigma x + b\Sigma x^2,$	4		
$a = 159.29, b = 4.955$	4		
Trend value for 1988 = 144.425	2		
Trend values : 139.45, 149.37, 154.34, 159.29, 164.25, 169.21, 184.09	6		
Graph	4		
2 The manager of a company wants to analyze about the sales of a particular brand of television and he wants to forecast the sales of the television in future. Data from 1985 to 1989 was given. Fit a parabola of second degree to the following and analyse using the curve.	20		
Year: 1985 1986 1987 1988 1989		5	K3
Sales: 16 18 19 20 24			
$a = \Sigma y - c\Sigma x^2/n, b = \Sigma xy / \Sigma x^2, c = \Sigma x^2 y - a \Sigma x^2 / \Sigma x^4$	4		
$a = 18.83, b = 1.8, c = 0.286$	6		
Trend values: $y = a + bx + cx^2 = 16.35, 17.31, 18.83, 20.91, 23.5$	6		
Graph	4		