

JK LAKSHMIPAT UNIVERSITY, JAIPUR
Pre- Ph. D EXAMINATION, YEAR 2016

Roll No.....

Course: Statistics for Research

Time: 2 hours

Max. Marks: 40

Instructions to students:

1. Do not write or mark anywhere other than your roll number on question paper.
2. Use of calculator and statistical tables is permitted.

Part A: Basic Concept

Q.1 Choose the correct answer for the following problems (10 questions @ 1 mark) 10

- i. If $P(A \cap B) = P(A) \cdot P(B)$, then A and B are
 - (a) Mutually exclusive events
 - (b) Independent events
 - (c) Mutually exhaustive events
 - (d) Dependent events
- ii. The lowest strength of association is reflected by which of the following correlation coefficients?
 - (a) 0.95
 - (b) - 0.60
 - (c) - 0.35
 - (d) 0.29
- iii. If a regression coefficient b_{yx} is negative, then b_{xy} is;
 - (a) negative
 - (b) positive
 - (c) zero
 - (d) none of these
- iv. Which of the following statement is incorrect?
 - (a) Standard deviation of a sample is a statistic
 - (b) Mean of a sample is a parameter
 - (c) Variance of a sample is a parameter
 - (d) Standard deviation of a sample is a parameter
- v. 'True hypothesis is rejected' is a;
 - (a) Type I error
 - (b) Type II error
 - (c) either (a) or (b)
 - (d) not an error
- vi. 't' statistic is used when;
 - (a) sample size is small and variance is known
 - (b) sample size is small and variance is not known
 - (c) sample size is large and variance is not known
 - (d) sample size is large and variance is known
- vii. χ^2 statistic is used for;
 - (a) test of goodness of fit
 - (b) test of hypothesis for single variance
 - (c) test of independence
 - (d) all of the above
- viii. Which of the following is the non-random method of selecting samples from a population?
 - (a) Multistage sampling
 - (b) Cluster sampling
 - (c) Quota sampling
 - (d) all of the above
- ix. The interval estimate of a population mean with large sample size and known standard deviation, is given by;
 - (a) $\bar{x} \pm z_{\alpha/2} \sigma_{\bar{x}}$
 - (b) $\bar{x} \pm z_{\alpha/2} s_{\bar{x}}$
 - (c) $\bar{x} \pm t_{\alpha/2} \sigma_{\bar{x}}$
 - (d) $\bar{x} \pm t_{\alpha/2} s_{\bar{x}}$
- x. Which of the following is the rank correlation coefficient;

- (a) Pearson (b) Spearman (c) Both (a) and (b) (d) none of these

Part B: Problems and Conceptual Understanding (Total Marks - 20)

Attempt any 4 problems (4 questions @ 5 marks)

Q.2A Bowler's scores for 6 games were: 182, 168, 184, 190, 170 and 174. Using this as a sample, compute the coefficient of variation. 2½

Q.2B The following times were recorded by the Quarter - Mile and Mile - Runners of a University track team (times are in minutes) 2½

Quarter - Mile times	.92	.98	1.04	.90	.99
Mile times	4.52	4.35	4.60	4.70	4.50

After viewing this sample of running times, one of the coaches commented that the quarter milers turned in the more consistent times. Use the standard deviation and coefficient of variation to summarize the variability in the data. Does the use of the coefficient of variation indicate that the coach's statement should be qualified?

Q.3 The manufacturers of tomato ketchup were interested in examining the relationship between the sales of the tomato ketchup and the shelf space allocated to that brand of tomato ketchup by various stores. Data was collected from 10 stores and part of the data is given below: 5

Store number	Sales (Rs. '000)	Shelf Space (Sqr. ft)
1	36	7
2	22	6.2
3	23	4.6
.	.	.
.	.	.
10	21	7.8

The output tables for the analysis of the given data are as follows:

Variables Entered / Removed ^b

Model	Variables Entered	Variables Removed	Method
Dimension	Shelf Space (Sqr. Ft) ^a	.	Enter

^a All requested variables entered.

^b Dependant variable: Sales (Rs. '000)

Model Summary

Model	R	R square	Adjusted R square	Std. Error of the Estimate
Dimension	.76 ^a	.645	.663	5.675

a. Predictors : (Constant), Shelf Space (Sqr. Ft)

Anova^b

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	241.098	1	241.098	18.744	.03 ^a
Residual	102.902	8	12.863		
Total	344.000	9			

^a Predictors : (Constant), Shelf Space (Sqr. Ft)

^b Dependant variable: Sales (Rs. '000)

Coefficients ^a					
Model	Unstandardized Coefficient		Standardized Coefficient	T	Sig.
	B	Std. Error	Beta		
1 (constant)	2.781	4.358	.837	.638	.541
Shelf Space (Sqr. Ft)	3.901	.901		4.329	.021

a. Dependant variable: Sales (Rs. '000)

Answer the following questions based on the output of regression analysis:

i) Give interpretation of the following from the model summary table

(a) R (b) R^2 (c) Adjusted R^2 (d) Standard error of estimate

ii) What does ANOVA table tells in regression model?

iii) Based on the output write down the regression equation

iv) Write the summary of your findings from the output given above

Q.4A Samples of two types of electric light bulbs were tested for length of life and the following data were obtained: 04

	Type I	Type II
Sample number	$n_1 = 8$	$n_2 = 7$
Sample means	$\bar{x}_1 = 1,234$ hours	$\bar{x}_1 = 1,036$ hours
Sample standard deviations	$s_1 = 36$ hours	$s_2 = 40$ hours

Is the difference in the variances significant to warrant that there is no difference in the length of life of both types? Test at $\alpha = 0.10$.

Q.4B Determine the value of $Z_{0.025}$ using normal distribution table. 01

Q.5A Calculate the coefficient of correlation between the industrial production (in crore tons) and export (in crore tons) using the following data: 03

Production	55	56	58	59	60	60	62
Export	3538	38	39	44	43	45	

Q.5B Explain the types of errors by taking a suitable example. 02

Q.6A The mean life of 10 electric motors was found to be 1450 hrs with a standard deviation of 423 hrs. A second sample of 17 motors chosen from a different batch showed a mean life of 1280 hrs with a standard deviation of 398 hrs. Is there a significant difference between the means of the two samples? 2½

Q.6B Suppose 26 cola drinkers are sampled randomly to determine whether they prefer brand A or brand B. The random sample contains 18 drinkers of brand A and 8 drinkers of brand B. Let C denotes brand A drinkers and D denotes brand B drinkers. Suppose the sequence of sampled cola drinkers is DCCCCDCCDCCCCDCDCCDDCCCC. Is this sequence of cola drinkers 2½

evidence that the sample is not random.

- Q.7A Write down the key difference between parametric and non-parametric tests. 2½
- Q.7B The average age of tourists who have come to Bharat is claimed to be 40 years. A random sample of 18 tourists gives the following ages: 24, 18, 37, 51, 56, 38, 45, 45, 29, 48, 39, 26, 38, 43, 62, 30, 66, 41. Test the hypothesis using $\alpha = 0.05$ level of significance. 2½

Part C: Analytical Ability and Situational Analysis (Total Marks - 10)

- Q.8 A company is interested in determining whether an association exists between the commuting time of their employees and the level of stress-related problems observed on the job. A study of 116 assembly-line workers reveals the following: 10

Commuting Time	Stress			
	High	Moderate	Low	Total
Under 20 minutes	9	5	18	32
20 - 50 minutes	17	8	28	53
Over 50 minutes	18	6	7	31
Total	44	19	53	116

At $\alpha = 0.01$ level of significance, is there any evidence of a significant relationship between commuting time and stress?

Or

- Q.8A The incidence of occupational diseases in an industry is such that the workers have twenty percent chance of suffering from it. What is the probability that out of six workers, four or more will contract the disease? 05
- Q.8B An article manufactured by a company comprises of 2 parts A and B. In the process of manufacturing part A, nine out of hundred are likely to be defective. Similarly, five out of hundred are likely to be defective in manufacturing of part B. Find the probability that the assembled parts will not be defective. 05