S.No. of Question Paper :

Unique Paper Code : 12273402_OC

Name of the Paper : Data Analysis

Name of the course : BA(H) Economics CBCS-SEC

Semester : IV

Duration : 2 Hours Maximum Marks: 75

Instructions for Candidates

1. Write your Roll No. on top immediately on receipt of the question paper.
2. Answer **any 4** questions.
3. All questions carry equal (18.75) marks.
4. Answers may be written either in English or in Hindi, but the same medium should be used throughout the paper.

परीक्षार्थियों के लिए निर्देश

1. प्रश्नपत्र मिलते ही निर्धारित स्थान पर अपना अनुक्रमांक लिखें |
2. किन्ही चार प्रश्नों का उत्तर दीजिये |
3. सभी प्रश्नों के अंक (18.75) समान है |
4. प्रश्न का उत्तर हिंदी अथवा अंग्रेजी माध्यम में लिखा जा सकता है परन्तु सभी प्रश्नों का उत्तर एक ही माध्यम में होना चाहिए |

Q1.a) Let $y_1, y_2, ..., y_n$ be a random sample from the pdf $f_y(y, \theta) = \frac{2y}{\theta^2}, 0 \leq y \leq \theta$.

Let two estimators be $\hat{\theta}_1 = \frac{3}{2}\bar{y}$ and $\hat{\theta}_2 = \frac{4}{5}\bar{y}$, be both unbiased. Which is more efficient?

b) The lifespan (in '000 hours) of four LED bulbs of 07 watts are as follows:

40,46,48, 50

(i) How many samples of size 2 can be formed in case when sampling is done with replacement and without replacement?

ii) Write all samples of size 2 with replacement from the above observations. Compute the sampling distribution and find mean and standard error of the distribution.

(iii) Write all samples of size 3 with replacement from the above observations. Compute the sampling distribution and find mean and standard error of the distribution.

(iv) Compare means and standard errors and draw the conclusion.

Q2.a) If there is perfect positive correlation between x and y. What can you say about correlation between:

i) Log x and Log y

ii) 2X+3, 5Y-6

b) For the following data on two variables, compute correlation coefficient between their levels and their Logs:

X	2	3	1	10	81
Y	8	9	4	16	100

(log 2 = 0.30103, log 3= 0.477121)

Q3. a) Which measure(s) will suit best in each of the following situation and Why?

i) In a school, students are graded on equal intervals except that all students scoring less than 40% are considered fail and clubbed as a single category. A visitor wishes to compute average marks a student obtains in this school given this data.

ii) In a survey, age has been categorized into various intervals as :

<15, 15-25,25-35,35-45,45 and above

Surveyor wishes to compute average age.

iii) A store outlet manager wishes to put apparels on sale, but wishes to exclude one size for each dress which is otherwise in high demand.

iv) It is known that two batsman have equal averages. But the coach wishes to know that if they are equally consistent too or one of them is more consistent.

v) The data on income of Indian individuals is given. The policymaker wishes to know the average income, and wants to know whether majority of people have lower levels of income or not.

vi) In a particular college 10,000 applicants applied in a specified course, where only 100 seats are available and admission is on merit basis. The dean needs to announce cut-off.

Q4. a) if variance of X = 9, regression equations : 8X-10Y+66 = 0, 40X-18Y=214. What are:

i) mean values of X and Y.

ii) correlation coefficient between X and Y.

iii) standard deviation of Y?

b) It was believed that NIFTY for IT companies depends on NIFTY 50 index, so a regression was run and following results were obtained:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.930152
R Square	0.865183
Adjusted R Square	0.862973
Standard Error	615.6311
Observations	63

ANOVA

	df	*SS*	*MS*	*F*
Regression	1	148366090.68	148366090.68	391.47
Residual	61	23119097.92	379001.61	
Total	62	171485188.6		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	5589.786	500.1509596	11.18	0.000
NIFTY 50	1.338237	0.067637312	19.79	0.000

Given above information answer the following:

i) Is it true that NIFTY 50 causes NIFTY IT? Explain why or why not?

ii) Compute coefficient of correlation between two.

iii) Is this model good or not? Explain which all values would you use and what do they show?

iv) Construct a 95% confidence interval for coefficient of NIFTY 50 and interpret it.

Q5. a) Suppose there are 3 coins in a bag. One of them is a fair coin, but the others are biased trick coins. When flipped, the three coins come up heads with probability 0.5, 0.6, 0.1 respectively. Suppose one of these three coins uniformly at random and flip it three times. What is P (HTT)? (That is, it comes up heads on the first flip and tails on the second)

b) There are 5 men and 8 women in a ballroom dancing class. If four men and four women are chosen and paired off, how many pairings are possible?

c) The volume in a set of wine bottles is known to follow a N(μ, 25) distribution. You take a sample of the bottles and measure their volumes. How many bottles do you have to sample to have a 95% confidence interval for μ with width 1?

Q6. a) For the following data, compute consumer price index for beverages using taking 2019 as base year:

Good	**2019**		**2020**		**2021**	
	Quantity	**Price**	**Quantity**	**Price**	**Quantity**	**Price**
Pepsi	10	6	12	8	5	10
Fanta	9	7	10	10	8	9
Coffee	12	10	6	6	12	15
Tea	6	5	5	6	10	7

Also, compute inflation rate between 2019-20 and 2020-21.

b) If price index increases from 100 to 150, what can you say about value of rupee during same period of time?