Name:	
Enrolment No:	

UPES
End Semester Examination, May 2025

Course: Advanced Statistics   Semester: IV
Program: BBA-ABD   Time : 03 hrs.
Course Code: DSQT 2004   Max. Marks: 100
Instructions:
1. Attempt all questions.
2. Simple/Scientific both calculators are allowed.

SECTION A
10Qx2M=20Marks

S. No.		Marks	CO
Q 1			
(i)	Which one of the following index numbers is based on Geometric Mean? (a) Lasperyres index number (b) Fischer's index number (c) Paasches index number (d) Bowley's index number	2	CO1
(ii)	The major purpose of price index is to measure change in the ……… (a) standard of living (b) cold content of money (c) Buying power of money (d) Capacity to produce	2	CO1
(iii)	There are several advantages of sampling because__. (a) sampling saves energy, money, and time in the collection of data (b) sampling help in lessening the data volume (c) help in getting higher accuracy if there is a homogeneous population (d) all the above-mentioned options	2	CO1
(iv)	With the increase in sample size, the error also____ (a) decreases (b) increases (c) remains same (d) all of the above	2	CO1
(v)	A teacher conducts an ethnographic probe for the issues that were being faced by the tribe members and the tribal students. The sampling method used here will be? (a) cluster sampling (b) systematic sampling (c) all of the above-mentioned options (d) none of these mentioned options	2	CO1
(vi)	The range of multiple correlation coefficient R is:	2	CO1

	(a) -1 to 1 (b) 0 to 1 (c) -1 to 0 (d) 0 to ∞		
(vii)	The regression equation $y = x^2 + 5$ represents: (a) A linear relationship between two variables (b) A quadratic relationship between two variables (c) An exponential relationship between two variables (d) No relationship between two variables	2	CO1
(viii)	The mean of the Binomial Distribution with n=5 and p= 14 is (a) 4/5 (b) 7/8 (c) 5/4 (d) 4/7	2	CO1
(ix)	Fishers Index number is based on (a) Base year quantities (b) Current Year quantities (c) Base & Current year quantities (d) none of the above	2	CO1
(x)	The Shape of Normal Curve (a) Bell Shaped (b) Rectangular (c) Triangular (d) Circular	2	CO1
SECTION B **4Qx5M= 20 Marks**			
Q 2.			
(i)	In customer segmentation for a retail analytics project, why might a simple random sample yield more generalizable insights compared to convenience sampling?	5	CO2
(ii)	What do you mean by random variable?	5	CO2
(iii)	Differentiate between correlation and causation. Give a real-life business example to support your answer	5	CO2
(iv)	Distinguish between the Laspeyres, Paasche, and Fisher's Ideal Index. Under what managerial conditions would each be preferred?	5	CO2
SECTION-C **3Qx10M=30 Marks**			
Q 3.			
(i)	(a) A fair coin is tossed 8 times. Find the probability that it shows heads i) exactly 5 times ii) at least once? (b) The Probability that a person who undergoes a kidney operation will be recovered is 0.5. find the probability that out of 6 patients who undergo similar operation i) none will recover ii) half of them recover.	5+5	CO3

(ii)	Given below are the price and quantity data for two years relating to three commodities.	10	CO3

Commodity	p_0	q_0	p_1	q_1
X	10	40	15	60
Y	15	80	20	100
Z	20	20	25	40

Calculate the Paasche's and Laspeyres price index number for the above given data.

(iii)	A quality manager is analyzing the impact of the number of customer service calls handled per day by different service centers on their average customer satisfaction score (out of 100). The data for 8 service centers is as follows:	5+3+2	CO3

Service Center	A	B	C	D	E	F	G	H
Calls Handled (X)	20	35	30	40	50	60	55	65
Satisfaction Score (Y)	65	75	70	78	85	90	88	95

i) Calculate the regression equation of Y on X. (Show step-by-step calculations)

ii) Estimate the expected customer satisfaction score if a service center handles 58 calls per day.

iii) Briefly explain any two assumptions of linear regression that should be considered while interpreting this model.

SECTION-D
2Qx15M= 30 Marks

Q 4.			
(i)	Suppose X_1, X_2, and X_3 represent performance metrics (e.g., quality scores from three machines).	15	CO4

X_1	50	54	50	56	50	55	52	50	52	51
X_2	42	46	45	44	40	45	43	42	41	42
X_3	72	71	73	70	72	72	70	71	75	71

Calculate the multiple correlation coefficient between $R_{1.23}$, $R_{2.13}$ and $R_{3.12}$.

(ii)	**Case Study: Assessing Student Satisfaction at a University** A university wants to evaluate student satisfaction with its new online learning platform. To ensure fair representation across different faculties and study levels, the administration uses stratified random sampling.	15	CO4

The student population is divided into four strata based on faculty: Arts, Science, Business, and Engineering. Each stratum is further divided into undergraduate and postgraduate groups. From each subgroup, 50 students are randomly selected, resulting in a sample of 400 students.

A structured online survey is administered, asking about ease of platform use, quality of resources, and overall satisfaction. The results show that 72% of students found the platform user-friendly, 65% were satisfied with the quality of resources, and 78% were generally satisfied with the platform.

The university plans to use these findings to improve the platform and inform future digital initiatives. Because the sampling method ensures each subgroup is proportionately represented, the results are considered generalizable to the entire student population.

(i)   Why is stratified random sampling appropriate for this case?

(ii)   What are the benefits of using probability sampling in student surveys?

(iii) How might the results differ if convenience sampling were used instead?

(iv)   What challenges might arise when forming strata in a diverse student population?