

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS)
THIRD SEMESTER EXAMINATION
(2024 ADMISSION ONWARDS)

MODEL QUESTION PAPER - THEORY

STATISTICS

MG3DSCSTA200 – STATISTICAL DISTRIBUTIONS

Duration: 2 Hours

Maximum: 70 Marks

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Short answer questions

Answer any ten questions. Each question carries **3 marks**.

1. Distinguish between discrete and continuous random variables. [U] [1]

2. Let X be a random variable with p.d.f $f(x) = \begin{cases} \left(\frac{1}{2}\right)^x & x = 1, 2, \dots \\ 0 & \text{elsewhere} \end{cases}$

Find the p.d.f of $Y = X^2$. [An] [1]

3. Define MGF of a random variable. What is the MGF of geometric distribution? [K] [2]

4. Establish Additive Property of Poisson Distribution [A] [1,2]

5. Given that $f(x) = kx^4 e^{-x}$ $0 \leq x \leq \infty$ is a p.d.f. Determine k . [A] [1]

6. Find the m.g.f of X with p.d.f $f(x) = \frac{1}{\theta}$, $0 < x < \theta$ and 0 elsewhere.

[An] [2]

7. What are the properties of a normal distribution? [U] [1,4]

8. Obtain the characteristic function of $f(x) = \theta e^{-\theta x}$ $0 < x < \infty$. [A] [2]

9. Write down the probability density function of Snedecor's F-distribution with 'm' and 'n' degrees of freedom). [K] [3]

10. Discuss the relationship between the t-distribution and the normal distribution. [A] [3]
11. Find the value of k such that $P(X < k) = 0.95$ for a Chi-square distribution with 3 degrees of freedom [An] [3]
12. Explain the relationship between the t-distribution and the χ^2 distribution [An] [3]

Part B

Short Essay questions

Answer any four questions. Each question carries **7 marks**.

13. For the following p.m.f find the distribution function and sketch its graph. [A] [1]

X	0	1	2	3
$f(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

14. A r.v X has the p.d.f $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} - \infty < x < \infty$. Find the p.d.f of $Y = X^2$. [An] [1]
15. Derive the characteristic function of the binomial distribution and hence obtain its mean and variance [An] [2]
16. Define the exponential distribution and establish its lack of memory property. [A] [1,2]
17. Define a Chi-square variate with n d.f. and derive its distribution. [A] [3]
18. Describe the method to obtain the probability points of $F(n_2, n_1)$ from those of $F(n_1, n_2)$. [An] [3]

Part C

Essay questions

Answer any one question. Each question carries **12 marks**.

19. The following data is related to number of wrong calls received at a call centre during a year. Fit a Poisson distribution to this data and obtain the expected frequencies

x	0	1	2	3	4
Frequency	182	88	66	24	5

[An] [1,2]

20. State Lindeberg-Levy central limit theorem. Show that binomial distribution tends to normal distribution as $n \rightarrow \infty$ [An] [1,4].

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS) THIRD SEMESTER EXAMINATION (2024 ADMISSION ONWARDS)

MODEL QUESTION PAPER - THEORY

STATISTICS

MG3DSCSTA201-Analytical Tools for Multivariate Analysis

Duration: 1.5 Hours

Maximum: 50 Marks

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Short answer questions

Answer any seven questions. Each question carries 2 marks.

1. Define orthogonality of vectors and give an example. [K][1]
2. Define basis and dimension. [K][1]
3. Explain the concept of linear independence of vectors with an example. [U][1]
4. Explain orthogonal projection of a vector. [U][1]
5. Write the properties of an idempotent matrix? [A][2]
6. Explain how rank of a matrix is determined using row echelon form? [U][2]
7. Explain the significance of Sylvester's Inequality in matrix rank. [U][2]
8. What is a generalized inverse (G-inverse) of a matrix? [K][2]
9. Define diagonalizable matrix. [K][4]
10. State one application of characteristic roots and vectors in statistics [K][4]

Part B

Short Essay Questions

Answer any four questions. Each question carries 6 marks

11. Define linear independence of vectors and examine the linear independence of the following vectors – $(1, 1, -1)$, $(-1, 1, 2)$, $(1, 3, 1)$. [A][1]

12. Prove that union of two subspaces of a vector space is also a subspace iff one is contained in the other one. [A][1]
13. Briefly explain the multiplication and inverse of a matrix using partitioning. [K][2]
- 14.. If A and B are 2 square matrices of same order, then show that $\text{rank}(AB) \geq \text{rank}(A) + \text{rank}(B) - n$. [K][2]
15. Show that if G is the g-inverse of A and $GA = H$ then (1) H is an idempotent matrix, (2) $AH = A$, (3) $\text{rank}(A) = \text{rank}(H)$. [A][2]
16. Briefly explain the spectral decomposition of a square matrix. [U][4]

Part C

Essay questions

Answer any one question. Each question carries 12 marks.

17. Construct an orthonormal set of vectors from the following vectors $X = [1, 1, 0]$, $Y = [0, 1, 1]$, $Z = [1, 0, 1]$. [E][1]
18. State and prove Cayley Hamilton theorem. [An][4]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS)

THIRD SEMESTER EXAMINATION

(2024 ADMISSION ONWARDS)

MG2DSCSTA203- Statistical Research Techniques Using Softwares

Duration: 1Hour 30 Minutes

Maximum: 50 Marks

PART A (Short answer type questions)

Answer any SEVEN questions. Each question Carries TWO marks

1. Define research design. [U] [1]
2. Distinguish between primary data and secondary data. [K] [1]
3. What is meant by literature review in a research process? [U] [1]
4. Define statistic with example. [U] [1]
5. Distinguish between simple and composite hypothesis. [K] [1]
6. What are the two types of errors in a statistical test procedure? [K] [1]
7. What are the assumptions used in t-test? [U] [3]
8. Distinguish between parametric and non parametric tests. [U] [2]
9. What is meant by contingency table? [K] [3]
10. Give any two applications of statistical tests in real life? [K] [4]

PART B (Short answer type questions)

Answer any FOUR questions. Each question Carries SIX marks

11. Distinguish between qualitative and quantitative research with examples. [K] [1]
12. Explain any three sampling techniques. [U] [1]
13. Explain different steps in a statistical test procedure. [U] [1]
14. A sample of 6,400 Englishmen has a mean height of 170 cms and a standard deviation of 6.4 cms, while a simple sample of 1600 Australians has a mean height of 172 cms and a standard deviation of 6.3 cms. Do the data indicate that Australians are on the average taller than Englishmen at 5% level of significance? [A] [2]
15. Explain one way ANOVA. [U] [3]
16. (a) Explain Wilcoxon signed rank test. State clearly the assumptions.
(b) A high school track and field coach is interested to know whether the new training program improved his athletes 400 meter sprint time and decided to

conduct an analysis. He obtained the permission of 7 athletes to record their sprint times both before and after his new training program which are shown below. Use appropriate non-parametric test to check whether these scores provide evidence of an increase in median sprint time at 5% level of significance?

Athlete	1	2	3	4	5	6	7
Sprint Time (Before)	63	61	62	58	59	56	61
Sprint time (After)	58	57	59	57	58	55	55

[An] [4]

PART C (Short answer type questions)

Answer any ONE questions. Each question Carries TWELVE marks

17. Explain different steps in writing a research proposal. [U] [1]
18. Following data represents the experience and annual income of 400 employees in a company. Test whether there is any association between length of service of employees and their annual income.

Experience (in years)	Annual Income (in Dollars)		
	Less than 40000	40000-75000	More than 75000
Less than 5	50	75	25
5-10	25	50	25
10 or more	25	75	50

[An] [4]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS)

THIRD SEMESTER EXAMINATION

(2024 ADMISSION ONWARDS)

MODEL QUESTION PAPER- THEORY

STATISTICS

MG3DSCSTA204 – Business Data Analytics

Duration: 1.5 Hour

Maximum:50 marks

Students should attempt at least one question from each course outcomes to enhance their overall outcome attainability.

PART A

Short answer questions

Answer any seven questions. Each question carries 2 marks.

1. What is a data and explain data in various fields. [U] [C01]
2. State the difference between quality and quality assurance. [An] [C03]
3. Define Control chart. [An] [C03]
4. Mention any 2 Statistical Quality Assurance Techniques. [An] [C03]
5. What is the purpose of using control charts in statistical quality control? [An] [C03]
6. Define Decision theory . [An] [C04]
7. What is the difference between decision making under risk and decision making under uncertainty ? [An] [C04]
8. What is TQM and state one benefit of implementing TQM in an organization. [An] [C04]
9. Define quality assurance and give one example of a method used in quality assurance. [An] [C04]
10. Explain the key principles of total Quality Management. [An] [C03]

PART B

Answer any four questions. Each questions carries 6 marks.

11. Describe the different types of data and provide examples of each. [U] [C01]
12. How can you classify digital data? Explain. [U] [C01]
- 13 .Explain the concept of time series analysis and its components. [A] [C02]
14. Describe the decision making process under uncertainty and provide an example. [An] [C04]
15. Difference between single sampling and Double sampling with examples. [An] [C04]
16. Discuss the importance of statistical quality assurance in industry. [An] [C03]

PART C

Essay questions

Answer any one question. Each question carries 12 marks.

17. Briefly explain timeseries, its components and methods for finding components. [A] [C02]
18. Explain the concept of statistical quality control and its importance in ensuring product quality. Also, explain the role of leadership in implementing Total Quality Management in an organization. [An] [C03]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS) THIRD SEMESTER EXAMINATION (2024 ADMISSION ONWARDS)

MODEL QUESTION PAPER - THEORY

STATISTICS

MG3DSESTA200- Statistical Techniques for Data Science and Machine Learning (Data Analytics Specialization)

Duration: 1.5 Hours

Maximum: 50 Marks

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Short answer questions

Answer any seven questions. Each question carries 2 marks.

1. Define sampling distribution. [K][1]
2. Explain the two types of errors in testing a statistical hypothesis. [U][1]
3. Write any two uses of the chi-square test. [K][2]
4. What are the assumptions of one-way ANOVA? [U][2]
5. Write any two applications of Machine learning [K][3]
6. State the general form of a simple linear regression model. [K][3]
7. Define logistic regression. [K][3]
8. What is categorical data? Give an example. [U][4]
9. Name any two sources of data in data science. [K][4]
10. What is meant by structured data? [U][4]

Part B

Short Essay Questions

Answer any four questions. Each question carries 6 marks.

11. A factory was producing electric bulbs of average life length 2000 hours with S.D 300. A new process was introduced with the hope that the life length of bulbs would increase. A sample of 50 bulbs produced by the new process was found to have an average length of life 2200 hours.

Examine whether it is reasonable to think that the length of life of bulbs has increased Assuming that the standard deviation has not changed.

[A][1]

- 12.The following data was obtained in an investigation about the effect of vaccination for small pox:

	Vaccinated	Not vaccinated
Attacked by smallpox	3	12
Not Attacked	8	5

Examine whether the vaccination is effective in preventing smallpox. [A][2]

- 13.Differentiate between supervised and unsupervised learning with examples. [U][3]
- 14.Explain logistic regression with a real-world application. [U][3]
- 15.Discuss the types and classifications of digital data used in data science. [U][4]
- 16.Explain different types of data. [U][4]

Part C

Essay questions

Answer any one question. Each question carries 12 marks.

- 17.Explain the procedure of one-way ANOVA [U][2]
18. (i). What is multiple linear regression?
- (ii). Derive the normal equations used to estimate the regression coefficients and illustrate its practical use with an example. [A][3]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS)
THIRD SEMESTER EXAMINATION
(2024 ADMISSION ONWARDS)
MG3VACSTA200 - Applied Statistical Analysis: Ethical Data Collection, Interpretation
and Decision making in Society

Duration: 1.5 Hours

Maximum Marks: 50

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Short Answer Type Questions. Answer any **seven** Questions

Each question carries **2** marks

1. Define sampling. Name any two sampling methods. [K] [1]
2. Define tabulation of data. [K] [1]
3. What is positive, negative and zero correlation? [U] [1]
4. What is a null hypothesis? [K] [2]
5. Give the null hypothesis for testing the difference between two means in large samples. [U] [2]
6. Define p-value. [K] [2]
7. What is the test statistic for testing proportion in large samples? [U] [2]
8. Define a numeric variable in R. [K] [3]
9. What is R Commander? [K] [3]
10. What function is used to calculate mean in R? [U] [1]

[7x2=14]

Part B

Short Essay Type Questions. Answer any **four** Questions

Each question carries **6** marks

11. Describe the essential steps involved in designing a good questionnaire. [A] [3]
12. Find standard deviation for the following data and interpret the result: 5, 10, 15, 20, 25. [U] [1]
13. Distinguish between parameter and statistic with example. [U] [2]
14. What is a paired t-test for related samples? [U] [2]
15. Explain Yule's coefficient of association with an example. [U] [1]
16. Describe the process of performing a t-test for a single mean using R. [U] [3]

[4x6=24]

Part C

Essay Type Questions. Answer any **one** Question

Each question carries **12** marks

17. Calculate mean, median and mode from the following grouped data: [U] [1]

Class	0-10	10-20	20-30	30-40	40-50
Frequency	5	8	12	10	5

18. The following data relates to marital status and performance in an examination. Examine whether the performance depends on marital status [U] [2]

Marital Status	Performance	
	Good	Bad
Married	60	80
Unmarried	20	40

[1x12= 12]

Reg. No:.....

Name:.....

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS) REGULAR EXAMINATION

Third Semester

MODEL QUESTION PAPER - THEORY

Minor Course - MG3DSCSTA202–

DATA ANALYSIS IN INFERENTIAL STATISTICS USING R/PYTHON

(2024 ADMISSION)

Duration: 1.5 Hours

Maximum Marks: 50

**Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C),
Interest (I), Appreciation (Ap), and Skill (S)**

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

[Learning Domain][CO No(s)]

Part A

Short Answer Questions

*Answer any **seven** questions. Each question carries **2 marks**.*

1. Given the joint probability mass function, $f(x, y) = c(x+y)$, defined over the points (1,1), (2,1), (2,2), (3,1), determine the value of the constant c . [U][1]
2. What is the definition of a conditional probability density function (pdf)? [U][1]
3. State the Central Limit Theorem. [U][2]
4. If X follows a normal distribution with mean 2 and standard deviation 3, find the probability that the sample mean $\bar{X}$ is greater than 2.5, given a sample size of 16. [U][2]
5. State the sufficient conditions for consistency of a parameter. [K][3]
6. Obtain the moment estimate for μ in $N(\mu, \sigma)$, σ known. [A][3]
7. Briefly explain the two types of errors that can occur in statistical hypothesis testing. [U][4]
8. A sample of 100 students is taken from a population with heights having standard deviation 10 cm. The mean height of the sample of students is found to be 168.8cm. Can we accept the assumption that the mean height of the students is 170 cm? [A][4]
9. State any two conditions required for the valid application of the non-parametric Chi-square tests. [K][4]
10. A sample of 10 individuals is chosen at random from a normal population and their mean weight is found to be 67.8 Kg with standard deviation 2.826 Kg. Check whether the mean of the population is 66. [A][4]

(7x2=14 marks)

Part B

Short Essay Questions

Answer any **four** questions. Each question carries **6 marks**.

11. For (X, Y) , the joint *pdf* is given by $f(x, y) = 30xy^2$; $x = 1, 2, 3$; $y = 1, 2$. Check independence of X and Y . [U][1]
12. Define Student's t distribution. State any two of its properties. Mention any one Student's t statistic. [U][2]
13. Explain the concepts of efficiency and sufficiency as desirable properties of point estimators, and support your explanation with one example each. [U][3]
14. Obtain maximum likelihood estimators of the parameters of uniform distribution over the interval (a, b) . [A][3]
15. The average wage of a sample of 150 workers in factory A was Rs. 2.56 with standard deviation Rs. 1.08. A sample of 200 workers in factory B has average hourly wage of Rs. 2.87 with standard deviation Rs. 1.28. Can the applicant safely assume that the hourly wage paid by factory B is higher than that of factory A? [An][4]
16. Briefly discuss ANOVA. If the Between Sum of Squares (SSC) is 123.33 and the Within Sum of Squares (SSE) is 94 for data collected on the test scores of students taught using three different teaching methods, conduct a one-way ANOVA to test whether there is a significant difference in the average scores among the teaching methods. Assume that the data was collected from 5 students per group (total of 15 students), and use a 5% significance level for the test. [An][4]

(4x6=24 marks)

Part C

Essay Type Questions

Answer any **one** question. Each question carries **12 marks**.

17. Define chi-square distribution and state any four of its properties. For a sample of size 10 from a normal population with standard deviation 5, find $P(8.4 < S^2 < 42.3)$. [U][2]
18. Fit a Poisson distribution to the following frequency data and perform a goodness-of-fit test. [An][4]

x	0	1	2	3	4
f	17167	1861	124	2	1

(1x12=12 marks)