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

MODEL QUESTION PAPER - THEORY

STATISTICS

MG5DSCSTA300 – APPLIED REGRESSION ANALYSIS

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. Define simple linear regression. [U] [1]
2. Mention any three applications of regression analysis. [U] [1]
3. Write the general form of a multiple linear regression model. [U][2]
4. What is the role of $(X'X)^{-1}$ in OLS estimation? [An][2]
5. State the formula for R^2 and specify its range. [A][2]
6. What is meant by serial correlation in regression analysis? [A][2]
7. Define the Generalized Least Squares (GLS) method. [An][3]
8. What is nonparametric regression. [U][3]
9. Define Poisson regression. [U][4]
10. What is meant by classification in logistic regression? [A][4]
11. What is meant by a fixed effect in a regression model? [U][5]
12. State two causes of multicollinearity. [U][5]

Part B

Short Essay questions

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

13. Explain the assumptions of the simple linear regression model. [K][1]
14. Explain the OLS method for estimating coefficients in multiple linear regression. [A][2]
15. Explain the need for model selection in multiple regression. [A][2]
16. Explain Weighted Least Squares and discuss its purpose. [A][3]
17. Explain the role of the reference category in multinomial logistic regression. [An][4]
18. Explain mixed effect models and their components. [A][5]

Part C

Essay questions

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

19. Explain the role of regression analysis in statistical modeling and decision-making. [A][1]
20. Explain forward selection, backward elimination, and stepwise selection procedures in detail. [A][3]

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

MODEL QUESTION PAPER - THEORY

STATISTICS

MG5DSCSTA301 – SAMPLING TECHNIQUES

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. What do you mean by non-probability sampling ? [U] [1]
2. What is Quota sampling? [U] [1]
3. What is the probability of getting a specified sample in SRSWOR? [K][2]
4. Describe the procedure of preparing slips in the Lottery Method. What precautions should be taken? [K][2]
5. What is meant by an unbiased estimator of population mean under SRSWR? [U][2]
6. What is sampling fraction? [U][2]
7. Define stratum weight for the h^{th} stratum under stratified random sampling.[K][3]
8. What is proportional allocation? [U][3]
9. Explain how you will draw a linear systematic sample.[U][4]
10. Explain circular systematic sampling.[K][4]
11. What are the advantages of systematic sampling? [K][4]
12. What do you mean by linear trend in population values? [K][4]

Part B

Short Essay questions

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

13. Explain the merits and demerits of Probability Sampling ? [K][1]

14. Explain how you would select a random sample of 11 households from a list of 112 households in a village using random number table method. [A][2]
15. Explain complete procedure of selecting random sample with replacement.[K][2]
16. What is Strata? What are the principles of stratification?[U][3]
17. A flood affected district in Kerala is divided into three zones. The number of villages in each zone is given by 440, 405, and 100 respectively. In order to estimate the total number of houses affected by flood in the district, the number of villages selected from each zone are 25,15 and 8 respectively. If total number of houses affected in the sampled villages are 750, 600, 344 respectively, estimate the total number of houses affected by flood in the district. [A][3]
18. Give an unbiased estimator of population total based on cluster sampling where the clusters are of equal Size. [U][5]

Part C

Essay questions

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

19. Explain in detail the various types of Probability Sampling with suitable examples ? [U][1]

20. Carryout a comparison between systematic sampling and stratified random sampling in the case of a population with linear trend among observations.[U][4]

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

MODEL QUESTION PAPER - THEORY STATISTICS
MG5DSCSTA302 - ANALYTICAL TOOLS FOR STATISTICS-I

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 **10** questions. Each question carries **3 marks**.

1. Find the supremum and infimum of the set $A = \left\{\frac{1}{5n};\ n \in I,\ n \neq 0\right\}$. [A] [1].
2. Prove that every sub set of a bounded set is bounded. [U] [1].
3. Is it true that 'a set of real numbers that is not open is closed'? Justify your answer [An][1].
4. Define subsequence of a sequence. Give example[U] [1].
5. Find the value of $\sum_{n=1}^{\infty} \frac{1}{n(n+1)}$[An][2].
6. State Leibniz's test for the convergence of an alternating series[A][2].
7. Evaluate $\lim_{x\to 0} \frac{\sqrt{4+x}-2}{x}$[An][3].
8. Distinguish between continuity and uniform continuity of a function [An][3].
9. State Darboux's theorem for derivatives [U] [4].
10. Give an example of a function that is continuous at a point but not differentiable there. [A] [4].
11. Discuss the differentiability of $f(x) = |x| + |x-1|$ at $x = 0$ and $x = 1$[An][4].
12. Verify Lagrange's mean value theorem for the function $f(x) = 2x^2 - 10x + 29$ on $[2,7]$ [An][4].

$10 \times 3 = 30$ marks

Part B

Short Essay questions

Answer any **4** questions. Each question carries **7 marks**

13. What are supremum and infimum of a set? Find the infimum and supremum of the set $A = \left\{\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots\right\}$ [An][1].
14. Prove that the sequence $\{S_n\}$, where $S_n = \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{n+n} \quad \forall\ n \in N$ is convergent [An][1].

15. Test for convergence of the series $\frac{\alpha}{\beta}+\frac{1+\alpha}{1+\beta}+\frac{(1+\alpha)(2+\alpha)}{(1+\beta)(2+\beta)}+\cdots$ [A][2].

16. Show that the function $f(x)=x^2$ is uniformly continuous on $[-1,1]$[An][3].

17. Show that if $f_n(x)=\frac{n^2x}{1+n^4x^2}$, then $\{f_n\}$ converges non-uniformly on $[0,1]$.[An] [4].

18. Prove that if a function is differentiable at a point, then it is continuous at that point. Is the converse true? Justify your answer with a suitable example. [An] [4].

$4\times7=28$ marks

Essay questions

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

19. (a) State and prove Cauchy's Root Test for the convergence of a series of positive terms. [An] [2].

(b) Examine the series $\sum\left(\frac{n}{2n+1}\right)^n$ for convergence. [An] [2].

20. (a) Determine the constants and so that the function defined below is continuous everywhere:

$$f(x)=\begin{cases}1 & \text{for } x\le 3\\ ax+b & \text{for } 3<x<5\\ 7 & \text{for } x\ge 5\end{cases}$$ [An] [3].

(b) Discuss the continuity at $x=1,2$ and 3 of the following function:

$$f(x)=\begin{cases}x+4 & \text{if } x\le 1\\ x^2+3x+1 & \text{if } 1<x\le 2\\ 2x^2+x+1 & \text{if } 2<x<3\\ 8x-2 & \text{if } x\ge 3\end{cases}$$ [An] [3].

$1\times12=12$ marks

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

MODEL QUESTION PAPER - THEORY

STATISTICS

MG5DSESTA300 – INTRODUCTION TO 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. How to use Hotelling′s T^2 statistic to test a given mean vector. [U] [1]
2. "T^2 is invariant under non−singular transformation" Verify. [U] [1]
3. Define Canonical Correlation Analysis in multivariate statistics. [K][1]
4. What are the three basic questions considered in profile analysis? [K][1]
5. When it is preferred to consider correlation matrix over covariance matrix in the case of factor analysis. [U][3]
6. Define Confirmatory Factor Analysis. [K][3]
7. State how EFA helps to identify unobserved constructs. [K][3]
8. List any two characteristics of orthogonal factor model. [K][3]
9. Under what condition do we obtain a quadratic discriminant function? [U][4]
10. What is a hyperplane in SVM? [K][4]

Part B

Short Essay questions

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

11. Prove that the extraction of principal components from a dispersion matrix is the study of characteristic roots and vectors of the same matrix. [A][2]
12. Discuss the core structural characteristics of EFA. [U][3]
13. Compare and contrast Varimax and Equamax rotation methods in Exploratory Factor Analysis. [U][3]
14. Explain the geometric interpretation of factor rotation using factor axes. [U][3]
15. Explain the Maximum Likelihood method of estimating factor loadings. [U][3]
16. Explain the agglomerative approach to hierarchical clustering. [U][4]

Part C

Essay questions

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

17. Derive the connection between Hotelling′s T^2 statistic and Likelihood Ratio Test(LRT) for testing H0: $\mu=\mu 0$ in a p-variate normal distribution. [U][1]
18. Discuss the steps involved in performing correspondence analysis. [U][5]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS)
FOURTH SEMESTER EXAMINATION
(2024 ADMISSION ONWARDS)
MODEL QUESTION PAPER - THEORY
STATISTICS
MG5DSESTA302-STATISTICAL COMPUTING USING PYTHON

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. Distinguish between a Python interpreter and a Python interface. [U][1]
2. What is the difference between append() and extend() methods in a list? Illustrate each with an example. [U][1]
3. Describe the significance of indentation in Python with an example. [An][1]
4. What is the advantage of a NumPy array over a Python list? [U][2]
5. What is the difference between a Pandas Series and a Pandas DataFrame? [U][2]
6. Write a single line of code to detect missing values in a DataFrame. [A][2]
7. Define a scatter plot. Give the Matplotlib command to create a simple scatter plot. [U][2]
8. What is lmplot? How is it different from regplot? [U][2]
9. What is Pseudo-Random Number? [U][3]
10. What is a t-test? Give the Python command to perform a one sample t-test. [U][4]

Part B

Short Essay Questions

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

11. What is an Integrated Development Environment (IDE)? Explain the major components of an IDE. [U][1]
12. Explain any five built-in methods used with Python lists. [U][1]
13. Load the built-in titanic dataset, using seaborn.load_dataset('titanic'). [A][2]
    Write Python code to:
    a) Create a DataFrame with columns age, fare, sex and name it as data.
    b) Display data types and shape.

c) Display age of first 5 female passengers.

Show expected output structure.

14. Create a long-format sales DataFrame using dummy values, containing Region, Quarter and Sales, with duplicate Region-Quarter combinations. Use pivot_table to obtain total sales, melt the result back to long format, and draw a Matplotlib bar chart of total sales by region with title and axes labels. [A][2]

15. Load the tips dataset using sns.load_dataset("tips"). Create a catplot with kind="count" to show the number of observations for each sex, using day to distinguish the groups. Write the Python code and explain the output. [A][2]

16. Explain the procedure for testing the equality of two population proportions using the Z-test. State the test statistic clearly. [U][4]

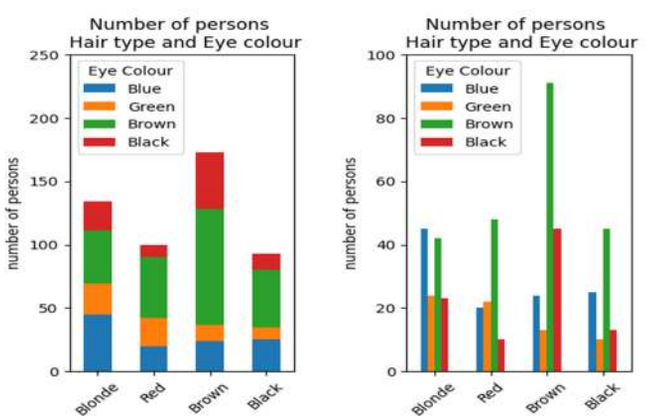
Part C

Essay questions

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

17. Write a Python programme to display the given data in sub-divided and multiple bar diagrams of the following forms. [A][2]

Hair Colours	**Eye Colours**			
	Blue	Green	Brown	Black
Blonde	45	24	42	23
Red	20	22	48	10
Brown	24	13	91	45
Black	25	10	45	13



18. Explain the complete procedure of a two-way ANOVA (with one observation per cell). Describe how the total variation is decomposed into the effects due to the two factors, their interaction, and error. Present the ANOVA table format used to summarize the results. Also give the Python codes for implementing two-way ANOVA. [U][4]

MAHATMA GANDHI UNIVERSITY, KOTTAYAM

MGU-UGP (HONOURS)

FIFTH SEMESTER EXAMINATION

(2024 ADMISSION ONWARDS)

MODEL QUESTION PAPER

STATISTICS

MG5SECSTA300 – Applied Statistics and Data Mining

Duration: 1.5 Hour Maximum Marks: 50

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. What are the different types of Index numbers? [R][1]
2. Define Kelly's index number. [R][1]
3. Distinguish between simple AM and simple GM index numbers. [U][1]
4. Define Chain base index number.[R][1]
5. What are the sources of Vital statistics in India. [R][2]
6. Write limitations of Total fertility rate. [U][2]
7. Define Gross reproduction rate and Net reproduction rate. [R][2]
8. What is Market basket analysis? [R][3]
9. What is noise in data, and how does it differ from an outlier? [An][3]
10. Distinguish between Additive and Multiplicative models of Time series. [An][3]

Part B

Short Essay questions
Answer any four questions. Each question carries 6 marks.

11. Define Factor reversal test. Examine whether Paasche's and Fisher's index number satisfies this test. [An][1]
12. Define Cost of living index number. Mention its uses and limitations. [U][1]

13. Briefly explain any 3 measures of Fertility rates.[U][2]
14. Explain the underlying assumptions of a life table and give a brief description of its construction. [U][2]
15. Explain the four components of a Time series with an example of each.[U][3]
16. Discuss how data mining is used in Healthcare, Finance, Marketing and Social Sciences. [U][4]

Part C

Essay questions

Answer any one question. Each question carries 12 marks.

17. Construct Laspeyre's, Paasche's and hence Fisher's index number for the following data: [A][1]

Commodity	Base year Price	Base year Expenditure	Current year Price	Current year Expenditure
A	6	48	10	60
B	4	20	6	120
C	10	60	12	96
D	8	200	14	336

18. For the distance matrix D , Apply complete linkage method and draw the corresponding dendrogram.[A][3]

$$D = \begin{bmatrix} 0 & 9 & 3 & 6 & 11 \\ 9 & 0 & 7 & 5 & 10 \\ 3 & 7 & 0 & 9 & 2 \\ 6 & 5 & 9 & 0 & 8 \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix}$$