D 31218 (Pages : 2) Name.......................................

Reg. No..................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST3C09—APPLIED REGRESSION ANALYSIS

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries a weightage of 2.

1. Stating the assumptions, describe the multiple linear regression model.
2. What is meant by multicollinearity ?
3. Write the test procedure for testing significance of slope parameter in a simple linear regression model.
4. Explain dummy variables in regression model.
5. What are kernal smoothers in non-parametric regression ?
6. Write a note on stochastic regressors in a linear model.
7. What are orthogonal polynomials ?

($4 \times 2 = 8$ weightage)

Part B

Answer any **four** *questions.*
Each question carries a weightage of 3.

8. For the linear regression model $Y = X\beta + \varepsilon$, where β is an $p \times 1$ vector, obtain maximum likelihood estimator of β if ε is normally distributed.
9. Define auto correlation. Explain the Durbin-Watson Test.
10. What are the limitations of Nadaraya-Watson (Kernel regression) estimator ?

Turn over

11. Explain Mallows cp statistic.

12. How does the assumption of orthogonality simplify the problem of least squares in polynomial regression ?

13. Explain Poisson regression. Explain a method of estimation in this setup.

14. Explain the residual analysis in the generalized linear model.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries a weightage of 5.

15. For the linear regression model $Y = \beta_0 + X\beta_1 + \varepsilon, \varepsilon \sim N\left(0, \sigma^2 I\right)$, obtain the distribution of residual sum of squares. Find a $100(1-\alpha)\%$ confidence interval for σ^2.

16. Explain heteroscedasticity. What are its consequences ? How will you detect heteroscedasticity ?

17. Discuss logistic regression model and explain the estimation procedure of parameters.

18. State and prove Gauss Markov theorem.

(2 × 5 = 10 weightage)

D 31219 (Pages : 2) Name.................................

Reg. No..............................

THIRD SEMESTER M.Sc. (CBCSS) DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

Statistics

MST 3C 10—STOCHASTIC PROCESS

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries weightage 2.

1. Specify the domain and the range of a stochastic process.
2. State the properties of a transition probabiiity matrix. Write down the transition probability matrix of a random walk.
3. Define a non-homogeneous Poisson process.
4. Obtain any one relation of infinitesimal generator matrix with the transition probability matrix.
5. Define semi-Markov processes.
6. What do you mean by renewal reward processes ?
7. Examine the nature of stationarity of Wiener process.

($4 \times 2 = 8$ weightage)

Part B

Answer any **four** *questions.*
Each question carries weightage 3.

8. Explain different classification of states of a Markov chain.
9. Show that in a finite irreducible Markov chain all states are recurrent.
10. What do you mean by a continuous parameter homogeneous Markov chain ? Show that its transition probability matrix is a continuous function of time.
11. Explain insurer's ruin problem.
12. Obtain the relation of Poisson processes with uniform distribution. Also obtain the autocorrelation of a Poisson process.

Turn over

13. State and prove inspection paradox.

14. Describe pricing stock options.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries weightage 5.

15. Obtain the expressions of mean and variance of n^{th} generation size of a Galton-Watson branching process.

16. Obtain the expression for the ultimate ruin probability of a gambler in a Gambler's ruin problem.

17. State basic renewal theorem. Obtain the limiting distributions of current life and excess life using basic renewal theorem.

18. Carry out the stationary equilibrium analysis of a G/M/1 queue.

(2 × 5 = 10 weightage)

D 31220 (Pages : 3) Name.......................................

Reg. No...................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 01—OPERATION RESEARCH – I

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries a weightage 2.

1. Write the canonical form of an LPP. What are its characteristics ?
2. What is the need of introducing artificial variables in an LPP ? Which methods are used to solve it ?
3. Explain the Mathematical model of a Transportation problem. Give an algorithm to solve a transportation problem.
4. Write Johnson's algorithm for a 2 machine n-jobs sequencing problem, mentioning the assumptions to be followed.
5. What is meant by integer programing ? How is it different from an LPP ?
6. What are the characteristics of a two-person zero sum game ?
7. Find the ranges of p and q which will render the entry (2,2) a saddle point for the game with payoff matrix :

	Player B		
Player A	2	4	5
	10	7	q
	4	p	6

(4 × 2 = 8 weightage)

Part B

Answer any 4 questions.
Each question carries a weightage 3.

8. Write the simplex algorithm to solve an LPP.
9. Find the basic feasible solution and the degenerate solution of $2x_1 + x_2 - x_3 = 2$; $3x_1 + 2x_2 + x_3 = 3$.
10. Describe any method to solve a balanced transportation problem.
11. A department of a company has five employees with five jobs to be performed. The time (in hours) that each man takes to perform each job is given in the effectiveness matrix. How should the jobs be assigned ? What is the minimum total man-hours required for the performance ?

Employees / Jobs	A	B	C	D	E
I	10	5	13	15	16
II	3	9	18	13	6
III	10	7	2	2	2
IV	7	11	9	7	12
V	7	9	10	4	12

12. Explain the cutting plane algorithm in integer programming.
13. Discuss the Zero-One implicit enumeration technique.
14. Distinguish between pure strategy and mixed strategy. Demonstrate the determination of mixed strategies for a 2×2 payoff matrix.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries a weightage 5.

15. Solve by Simplex method :

    Maximum $Z = 2x_1 + 4x_2$,

    subject to $x_1 + 2x_2 \leq 5, x_1 + x_2 \leq 4, x_1 \geq 0, x_2 \geq 0$.

16. (a) Explain the traveling salesman problem.

(b) A salesman has to travel to 4 cities, including his home city. The travelling cost in rupees is given in the table. Find the least cost of his travelling through the shortest route in this task by visiting each city once only.

	A	B	C	D
A	0	25	75	45
B	35	0	150	25
C	35	40	0	15
D	65	75	130	0

17. Find the optimum integer solution of the problem :

Maximize $Z = 5x_1 + 8x_2$,

subject to $x_1 + 2x_2 \leq 8, 4x_1 + x_2 \leq 10$, $x_1 \geq 0, x_2 \geq 0$ and integers.

18. Define rectangular Games. Outline the matrix method for an $m \times n$ two person zero sum game that lead to at least one admissible solution. Use the matrix method to solve the game whose payoff matrix is :

$$\begin{bmatrix} 3 & 6 \\ 5 & 5 \\ 9 & 3 \end{bmatrix}.$$

(2 × 5 = 10 weightage)

D 31221 (Pages : 2) Name.......................................

Reg. No..................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 02—TIME SERIES ANALYSIS

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries a weightage of 2.

1. What are the important components of a time series model ?
2. Discuss exponential smoothing.
3. Define strict stationarity and wide sense stationarity.
4. Write the auto covariance function of a linear process.
5. Discuss l step forecast of auto regressive model.
6. Discuss the need of residual analysis of a time series data.
7. What is meant by saying: time series model is linear and non-linear ?

(4 × 2 = 8 weightage)

Part B

Answer any **four** *questions.*
Each question carries a weightage of 3.

8. Explain Holt method of smoothing for non seasonal time series model.
9. Consider the process $Z_t = Z_0 \cos ct$. Determine the value of c for which the process is stationary.
10. Show that the function $r(h) = \begin{cases} 1 & \text{if } h = 0 \\ 0.4 & \text{if } h = \pm 1 \\ 0 & \text{otherwise} \end{cases}$ is an autocovariance function.

Turn over

11. Describe the maximum likelihood estimation of parameters of AR(1) model.

12. Discuss unit root test for stationarity.

13. Describe the Box-Jenkins approach of analysis of a time series data.

14. Define ARCH and GARCH time series models.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries a weightage of 5.

15. Derive the condition for invertibility of MA(1) and MA(2) processes.

16. (a) Derive Yule-Walker equation for estimation of parameters of AR process.

(b) Discus the estimation of PACF using Levinsen-Durbin method.

17. Derive the spectral density of ARMA (p, q) process given by $\Phi(B) Z_t = \Theta(B) \varepsilon_t$. Identify the process with spectral density $s(\lambda) = \frac{\sigma^2}{2\pi}(2 - 2\cos\lambda)$.

18. Explain in detail various methods for determining trend in a time series.

(2 × 5 = 10 weightage)

D 31222 (Pages : 3) Name...

Reg. No....................................

THIRD SEMESTER M.Sc. DEGREE [REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

Statistics

MST 3E 07—STATISTICAL DECISION THEORY

(2019 Admission Onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries 2 weightage.

1. Define (i) Decision function, (ii) Loss function, (iii) Risk function, and (iv) Bayes decision rule.

2. Define prior and posterior distributions. Let X follows exponential distribution with mean $\frac{1}{\lambda}$ and prior distribution of λ is $\pi(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} e^{-\beta\lambda}\lambda^{\alpha-1}, 0 < x < \infty, \lambda > 0$. Find the posterior distribution of λ based on a random sample of size n from the population.

3. Prove that a Bayes estimator with constant risk is minimax.

4. Explain the theory of dominance in the solution of rectangular games.

5. Define (i) Pure strategy ; (ii) Mixed strategy ; (iii) Two person zero sum game and (iv) Payoff matrix.

6. Explain briefly (i) Noninformative priors ; and (ii) Maximum entropy priors.

7. Describe the graphical method of solving $2 \times n$ game.

(4 × 2 = 8 weightage)

Part B

Answer any **four** *questions.*
Each question carries 3 weightage each

8. Prove the following :
   (i) Bayes estimator is the mean of posterior distribution under squared error loss function.
   (ii) Family of beta distributions is a conjugate prior for parameter p of binomial distribution.

9. A random sample of size n is drawn from Uniform $(0,\theta)$ distribution. Estimate θ with the loss function $L(\theta,d) = \frac{(t-\theta)^2}{\theta^2}$. Assuming prior distribution of θ is uniformly distributed over (0,1), find the Bayes estimator of θ.

Turn over

10. If the loss function $L(\theta, d)$ is strictly convex, show that the class of all non-randomized decision rules for estimating the parameter θ is complete.

11. Define (i) complete and (ii) minimal complete decision rules. Prove that if a class of admissible decision rules is complete, it must be minimal complete.

12. Define admissible estimator. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, 1)$. Show that $\bar{X}$ is an admissible estimator of θ under squared error loss function.

13. Explain briefly (i) Games without saddle point ; and (ii) Minimax theorem and ; (iii) Solution of 2×2 matrix games.

14. Two competitors are competing for the market share of the similar product. The Payoff matrix in terms of their advertizing plan is shown below :

	Competitor B		
Competitor A	No advertizing	Medium advertizing	Heavy advertizing
No advertizing	10	5	–2
Medium advertizing	13	12	13
Heavy advertizing	16	14	10

Suggest optimal strategies for the two firms and the net outcome thereof.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries 5 *weightage each.*

15. Define minimax estimator. Show that $\dfrac{\sum_{i=1}^{n} X_i + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}$ is a minimax estimator of θ in sampling from the Bernoulli (θ) distribution under a squared-error loss function.

16. (i) What are competitive situations in game theory ? Explain with the help of an example ; (ii) Define saddle point. What are the rules to determine saddle point in game theory ?

(ii) Find the range of values of p and q which will render the entry (2, 2) a saddle point in the game with the following payoff matrix :

$$\begin{pmatrix} 2 & 4 & 5 \\ 10 & 7 & q \\ 4 & p & 6 \end{pmatrix}.$$

17. (i) Describe the importance of statistical decision theory to solve real problems.

(ii) Explain subjective, conjugate and Jeffrey's prior distributions.

18. (i) If $\theta(X)$ is a minimax estimate of θ, show that $f(X) = a\,\theta(X) + b$ is a minimax estimator of $a\theta + b, a \neq 0, b$ being constants and the loss function is assumed to be of the squared error type.

(ii) Write short notes on (a) Empirical Bayes analysis and (b) Bayesian robustness.

(2 × 5 = 10 weightage)

D 31223 (Pages : 3) Name.......................................

Reg. No..................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 10—STATISTICAL QUALITY CONTROL

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries a weightage 2.

1. What is the need for statistical quality control ?
2. Define total quality management.
3. Differentiate between specification limits and control limits.
4. Explain six-sigma concepts in quality control.
5. What is meant by process capability analysis ?
6. Explain AQL and LTPD.
7. What is the use of orthogonal arrays in Taguchi's optimization method ?

(4 × 2 = 8 weightage)

Part B

Answer any **four** *questions.*
Each question carries a weightage 3.

8. What is the difference between control charts for variables and control charts for attributes ?
9. Differentiate between Type A and Type B OC curves.
10. Explain Military Standard 105E.
11. Explain the V-mask procedure.
12. What are the salient features of the Dodge-Roming tables ?

Turn over

13. What is CSP-1 procedure ?

14. Establish the OC function of a *p*-chart.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries a weightage 5.

15. (i) Discuss the behavior of $\bar{X}$-charts in relation to R-charts.

(ii) The following data pertains to 6 samples whose values being taken every hour for 5 hours. Construct the X-bar and R-charts and comment on your answers :

Sample No.	Sample values				
1	42	65	75	78	87
2	42	45	68	72	90
3	19	24	80	81	81
4	16	54	69	77	84
5	42	51	57	59	78
6	51	74	75	78	60

16. (i) Discuss the control charts for fraction defectives.

(ii) Sixteen boxes of electric switches each containing 20 switches were randomly chosen from a lot and the following number of defects per box were noted. Draw a c-chart and make your conclusion :

Box No.	No. of defects	Box No.	No. of defects
1	12	9	11
2	15	10	12
3	9	11	16
4	14	12	13
5	18	13	19
6	26	14	18
7	8	15	14
8	6	16	21

17. (i) Write the procedure of a double sampling plan.

(ii) The values listed below are points on the OC curve of a sampling plan in which a sample of size 25 is accepted if there is at the most 1 defective and is rejected if otherwise (i.e., if there are 2 or more defectives). Plot the AOQ curve for the plan. What is the AOQL ?

Percentage defective	0	2	4	6	8	10	15	20
Probability of acceptance	1	0.91	0.74	0.55	0.40	0.27	0.09	0.03

18. Explain the Sequential Sampling Inspection Plan. Discuss the idea of SPRT.

(2 × 5 = 10 weightage)

D 31224 (Pages : 2) Name.......................................

Reg. No..................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 13—BIOSTATISTICS

(2019 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries weightage of 2.

1. Define survival function. Derive the survival function of an exponential distribution with mean 10.
2. Give an example of a distribution having bath-tub shape hazard function.
3. Distinguish between type I and type II progressive censoring.
4. What do you mean by odds ratio ? How do you compute it from a 2 × 2 contingency table ?
5. Explain the idea of permutation test.
6. Explain the planning and design of a clinical trial.
7. Explain the important ethics behind randomized studies involving human subjects.

(4 × 2 = 8 weightage)

Part B

Answer any **four** *questions.*
Each question carries weightage of 3.

8. Define Rayleigh distribution and derive its mean survival time, survival function and hazard function.
9. Explain Cox's F test for comparing two survival functions.
10. Give a brief explanation on simple and general epidemic models.
11. Explain maximum likelihood method for estimation of probabilities of death under competing risks.
12. Explain detection and estimation of linkage in heredity.
13. State Mendel's law and explain its importance in genetics.
14. Explain randomized dose-response study and its advantageous.

(4 × 3 = 12 weightage)

Turn over

Part C

Answer any **two** *questions.*
Each question carries weightage of 5.

15. (a) Establish the relationships between survival function, hazard function and density function. Also explain how do you obtain mean and median survival from the survival function.

    (b) Derive the survival function and hazard function of Gamma and Weibull distributions and give a brief discussion on the nature of these hazard functions.

16. (a) Explain Kaplan-Meier method estimating survival function.

    (b) Obtain Kaplan-Meier estimate of the survival function based on the following data :

    5, 6+, 8, 2, 10+, 3, 7+, 4, 8, 8+, 10, 9.

17. (a) Explain the basic model in logistic regression and estimation of its parameters.

    (b) Explain important diagnostic test related to logistic regression model.

18. Explain planning and design of clinical trials in phase I, II and III trials.

(2 × 5 = 10 weightage)

D 31225 (Pages : 2) Name.....................................

Reg. No.................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 19—STATISTICAL MACHINE LEARNING–I

(2020 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Weightage 2 for each question.

1. Outline model selection and the bias-variance trade-off.
2. What do you mean by Ridge Regression ?
3. Outline the working of Nonparametric Logistic Regression.
4. Explain Linear Basis Expansions.
5. Explain Parzen Density Estimate.
6. Explain "Local Linear Regression".
7. What do you mean by "Bagging" ?

(4 × 2 = 8 weightage)

Part B

Answer any **four** *questions.*
Weightage 3 for each question.

8. Write a note on Partial Least Squares.
9. Outline the working principle of Logistic Regression.
10. Write a note on smoothing Splines as a Spline Basis Method.
11. Discuss "Local Likelihoods".

Turn over

12. Discuss Mixture models for Density Estimation and Classification.

13. Discuss maximum likelihood inference.

14. Summarize Vapnik-Chervonenkis Dimension.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Weightage 5 *for each question.*

15. Formulate Linear Regression model and illustrate with a suitable example.

16. Give a detailed account of Reproducing Kernel Hilbert Spaces.

17. Discuss in detail Kernel Density Estimation for classification.

18. Discuss in detail Bias, Variance and Model Complexity.

(2 × 5 = 10 weightage)

D 31226 (Pages : 2) Name.....................................

Reg. No..................................

THIRD SEMESTER M.Sc. DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST3E 22—NON-PARAMETRIC STATISTICAL METHODS

(2020 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
Each question carries 2 *weightage each.*

1. Explain the difference between interval and ratio scales of measurements with examples.
2. Define empirical distribution function. Write down its probability distribution, mean and variance.
3. What is the use of Shapiro- Wilk test in non parametric inference ? Write down the null hypothesis and the test statistic.
4. Define run. Give the procedure of run test for randomness.
5. How to find confidence interval using the kernel density estimator ?
6. What is the purpose of resampling ? Mention different types of resampling techniques.
7. What is Crammer-Von mises criterion ? Define Crammer-Von mises test statistic for identical population.

($4 \times 2 = 8$ weightage)

Part B

Answer any **four** *questions.*
Each question carries 3 *weightage each.*

8. (i) What is QQ plot ? How to make a QQ plot ? Explain its uses.

   (ii) Define Phi coefficient for a 2×2 table. How to interpret Phi co-efficient ?
9. Describe Wald-Wolfowitz run test for two sample problem.
10. What is bias variance trade off ? Explain bias variance decomposition of mean squared error.
11. Describe Kruskal-Wallis test for one way analysis of variance.

Turn over

12. Write short notes on Lilliefors's test for normality.

13. What is a rug plot used in a density plot ? How do you make a drug plot in R ?

14. Distinguish between nominal and ordinal scales of measurements with examples.

(4 × 3 = 12 weightage)

Part C

Answer any **two** *questions.*
Each question carries 5 *weightage each.*

15. (i) Describe : (a) Crammers contingency co-efficient ; (b) Pearson contingency co-efficient ; and (c) Pearson's mean square co-efficient.

    (ii) Define Kaplan Meier estimator. What are its assumptions ?

16. Explain : (a) Anderson Darling test ; (b)Jarque bera test ; and (c) McNemar test.

17. (i) Write short notes on Friedman's two-way ANOVA by ranks.

    (ii) Describe bootstrap method for estimating variance of an estimator.

18. Explain histogram density and kernel density estimators. Find the bias of histogram density estimator. Analyze the estimation error of Kernel density estimation.

(2 × 5 = 10 weightage)

D 31227 (Pages : 2) Name.................................

Reg. No.............................

THIRD SEMESTER M.Sc. (CBCSS) DEGREE (REGULAR/SUPPLEMENTARY) EXAMINATION, NOVEMBER 2022

(CBCSS)

Statistics

MST 3E 24—APPLIED ALGORITHM AND ANALYSIS OF MULTITYPE AND BIG DATA

(2020 Admission onwards)

Time : Three Hours Maximum : 30 Weightage

Part A

Answer any **four** *questions.*
2 *weightages each.*

1. What is a hyperplane ?
2. Outline the three procedures commonly used to obtain respondents' perceptions of similarities.
3. Identify the purpose of a kernel in support vector machine.
4. What are key performance indicators ?
5. Identify the purpose of maximization step in EM algorithm.
6. Compare volume and velocity in big data.
7. Explain the concept of a mixture model.

(4 × 2 = 8 weightages)

Part B

Answer any **four** *questions.*
3 *weightages each.*

8. Compare and contrast Non metric Versus Metric Methods.
9. Outline the concept of data augmentation in EM algorithm.
10. Illustrate the procedure of one vs one classification in SVM.
11. Compare structured and semi structure data with example.
12. Explain the objectives of multidimensional scaling.

Turn over

13. Compare veracity and variety in big data.

14. Explain the concept of maximal margin classifier in SVM.

($4 \times 3 = 12$ weightages)

Part C

Answer any **two** *questions.*
5 *weightages each.*

15. Elaborate on EM algorithm with two component mixture model.

16. Explain Big Data Analytics Cycle.

17. Describe Support Vector Classifier with equations.

18. Compare the decompositional and compositional approaches in multidimensional scaling.

($2 \times 5 = 10$ weightages)