M.Sc. Examination, 2018
Semester-I
Statistics
Course : MSC-11

(Linear Algebra and Linear Models)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. a) Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix, with $n \geq m$. Then show that the n eigenvalues of BA are m eigenvalues of AB with the extra eigenvalues being 0.

   b) Prove that if A is a symmetric matrix, then the eigenvalues of A are all real.

   c) Let A be a symmetric $n \times n$ matrix, and let $\lambda \neq \mu$ be eigenvalues of A with **x, y** as corresponding eigenvectors respectively. Then show that $\mathbf{x}'\mathbf{y}=\mathbf{0}$. 5+3+2=10

2. a) State and prove spectral decomposition theorem.

   b) Show that for any square matrix, the determinant is the product of all the eigenvalues and trace is the sum of all the eigenvalues. 5+5=10

3. a) State and prove a necessary condition for a linear parametric function $\lambda'\beta$ to be estimable under the Gauss-Markov linear model set up.

   b) Consider the model

$$E(y_1) = \beta_1 + 2\beta_2$$
$$E(y_2) = 2\beta_2$$
$$E(y_3) = \beta_1 + \beta_2,$$

   Find the residual sum of squares subject to the restriction $\beta_1 = \beta_2$. 5+5=10

4. a) With reference to the linear model

$$E(y_1) = \beta_1 + \beta_2,$$
$$E(y_2) = 2\beta_1 - \beta_2,$$
$$E(y_3) = \beta_1 - \beta_2,$$

   Where y_1, y_2, y_3 are uncorrelated with a common variance, answer the following questions:

   i. Find two different linear function of y_1, y_2, y_3 that are unbiased for β_1. Determine their variances and the covariance between the two.

   ii. Find two linear functions that are both unbiased for β_2 and are uncorrelated.

   iii. Write down the model in terms of the new parameters $\theta_1 = \beta_1 + 2\beta_2$, $\theta_2 = \beta_1 - 2\beta_2$.

b) Consider a linear model E(Y) = X β, where Y is the vector of n observations and β is the vector of n parameters. The (i, j)th element of the design matrix is of the form $1+ij(j+1), 1 \le i, j \le n$. Find the condition on n so that β will be estimable.

5. Consider the model

$$y_1 = \mu + \alpha_1 + \beta_1 + \epsilon_1$$
$$y_2 = \mu + \alpha_1 + \beta_2 + \epsilon_2$$
$$y_3 = \mu + \alpha_2 + \beta_1 + \epsilon_3$$
$$y_4 = \mu + \alpha_2 + \beta_2 + \epsilon_4$$
$$y_5 = \mu + \alpha_3 + \beta_1 + \epsilon_5$$
$$y_6 = \mu + \alpha_3 + \beta_2 + \epsilon_6$$

Answer the following questions with justification.

a) When is $\lambda_0\mu + \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3 + \lambda_4\beta_1 + \lambda_5\beta_2$ estimable?

b) Is $\alpha_1 + \alpha_2$ estimable?

c) Is β_1 / β_2 estimable

d) Is $\mu + \alpha_1$ estimable?

e) Is $\alpha_1 - 2\alpha_2 + \alpha_3$ estimable?

f) What is the covariance between the BLUEs of $\beta_1 - \beta_2$ and $\alpha_1 - \alpha_2$, if they are estimable.

g) Obtain any linear function of observations belonging to the error space.

h) What is the rank of the estimation space?

6. In case of multiple linear regression equation, discuss the testing of $H_0 : \beta_1 = \ldots = \beta_p = 0$. Express the test statistic in terms of the multiple correlation.

10

M.Sc. Examination, 2018
Semester-I
Statistics
Course : MSC-12

(Regression Analysis)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. a) Briefly write down the application of regression analysis.
   b) Find ols estimator of σ^2 for a simple linear centered regression model.
   c) In simple linear regression, find a 95% confidence interval of the mean response for a particular value of the regressor variable. 3+3+4=10

2. a) In usual notations for multiple linear regression, show that $SSR = y'[H = -\frac{1}{n}J]y$. Hence find the distribution of $\frac{SSR}{\sigma_2}$.
   b) Discuss the testing procedure for a set of regression coefficients equals to zero.
   c) Describe the unit length scaling procedure in multiple linear regression. (2+3)+3+2=10

3. a) Write notes on PRESS Residuals, R-Student residuals, Added variable plot.
   b) Show that $S_{(i)}^2$ can be written as $S_{(i)}^2 = \frac{(n-p)MS_{\text{Res}} - e_i^2/(1-h_{ii})}{n-p-1}$, where $S_{(i)}^2$ is the estimate of σ^2 after deleting the i^{th} observation, MS_{Res} is the estimate of σ^2, e_i is the i^{th} residual and h_{ii} is the i^{th} diagonal of the hat matrix. (2+2+2)+4=10

4. a) How orthogonal polynomials are constructed and used in regression setup?
   b) In multiple linear regression, show that the generalized least squares estimator of β is BLUE. 4+6=10

5. Discuss the criteria for evaluating subset regression models. 10

6. a) Discuss a parameter estimation method in non linear regression model.
   b) How do you interpret the parameters in a Logistic regression model. 5+5=10

M.Sc. Examination, 2018
Semester-I
Statistics
Course : MSC-13

(Stochastic Process and Distribution Theory)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Group-A (Stochastic Process)
Answer **any two** questions

1. a) Define covariance stationarity of a stochastic process. Cite an example where covariance stationarity does not imply strict stationarity.

   b) Let $\{y_n, n \geq 1\}$ be a sequence of independent random variables with $P(Y_n=1) = p = 1 - P(Y_n = -1)$. Let X_n be such that $X_0 = 0$

   $$X_{n+1} = X_n + Y_{n+1}.$$

   Examine if $\{X_n, n \geq 1\}$ is a Markov process.

   c) "If a markov Chain has limiting distribution it has stationary distribution as well" – Is the statement correct? Discuss with a suitable example. 3+4+3=10

2. a) What do you mean by an ergodic state? Suppose a Markov chain with state space S={0,1,2} has transition probability matrix.

   $$\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$$

   Check if the states are ergodic.

   b) Prove that in an irreducible Markov chain, all the states are of the same type. Then discuss on the states of a finite irreducible Markov Chain. (2+3)+(3+2)=10

3. a) Consider a gambler who at each play of the game either wins one unit of money with probability p or loses one unit of money with probability 1 – p. Let the game continues until the gambler's capital increases to a rupees or he goes broke. Find out the probability of his losing if at some point he has i unit of money.

   b) Consider a Poisson process $\{X(t), t \geq 0\}$. Let S_n denote the time of n occurrences of the event. Find the distribution of S_n. 6+4=10

4. a) Suppose illegal immigration to India occurs at a Poisson rate of 8 per week (say). The probability that an illegal immigrant is Bengali speaking is 1/6. Find the probability that no Bengali-speaking illegal immigrant will arrive in India during a period of 14 days. 5+5=10

P.T.O.

b) In the context of continuous time Markov Chain establish Feller-Kolnogorov baekward and forward equations. 5+5=10

Group – B (Distribution Theory)

Answer **any two** questions

5. a) Derive the expression of the MGF of non-central χ^2 distribution with d.f n and non-centrality parameter λ.

   b) Prove that $P(x_n^2(\lambda) \leq x) = P(X_1 - X_2 \geq \frac{n}{2})$, where $X_1 \sim$ Poisson $\left(\frac{x}{2}\right)$,

   $X_2 \sim$ Poisson $\left(\frac{\lambda}{2}\right)$, independently. 5+5=10

6. In the Gauss Markov setup, prove that the error sum of squares and sum of squares due to regression are independent x^2 random variables with degrees of freedom n-r and r respectively, r being the rank of the n × p design matrix X. You have to prove the necessary results. 10

7. Prove that the sample mean vector and dispersion matrix based on a normal data matrix distributed are independently distributed. Also find the sampling distribution of the sample dispersion matrix. 4+6=10

8. a) Find the sampling distribution of the sample Mahalanobis distance.

   b) Let $\underset{\sim}{X}_1, \underset{\sim}{X}_2 ... \underset{\sim}{X}_n$ $\underset{\sim}{i}.i.d$ $N_p(\underset{\sim}{\mu}, \Sigma)$. Derive the likelihood ratio test for testing $H_0 : R\underset{\sim}{\mu} = \underset{\sim}{r}$ with known R (matrix) and $\underset{\sim}{r}$ (vector). 5+5=10

M.Sc. Examination, 2018
Semester-I
Statistics
Course : MSC-14

(Statistical Inference-1)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. a) Define a minimal sufficient statistic. If $X_1, X_2, ...X_m$ are distributed as $N(\mu, \sigma_1^2)$ and $X_{m+1}, X_{m+2}, ..., X_{m+n}$ are distributed as $N(\mu, \sigma_2^2)$ independently, obtain a minimal sufficient statistic for $(\mu, \sigma_1^2, \sigma_2^2)$.

   b) State and prove the Rao-Blackwell theorem. 6+4=10

2. a) Let U_g and U_0 be, respectively, the class of all unbiased estimators of $\gamma(\theta)$ with finite variances and the class of all unbiased estimators of zero with finite variances. Then show that T is an uniformly minimum variance unbiased estimator (UMVUE) of $\gamma(\theta)$ if and only if cov(T, h) = 0 $\forall \theta \in \Omega, \forall h \in U_0$.

   b) Let the probability distribution of a random variable X be

   $P_\theta[X=-1]=\theta, P_0[X=x]=(1-\theta)^2\theta^x, x=0,1,2,...$ Show that any estimable function $\gamma(\theta)$ admits an UMVUE if and only if $\gamma(\theta)=\alpha+b(1-\theta)^2$ for some constants a, b. Also mention the UMVUE of $\gamma(\theta)$. 5+5=10

3. a) Describe Maximum likelihood method for parameter estimation. State its properties.

   b) Let $(X_1, X_2, \ldots, X_n)$ be a random sample from $R(\theta, \theta+1)$ with unknown parameter θ. Find the maximum likelihood estimator of θ. 6+4=10

4. a) Consider the exponential family of distributions

   $$P = \{f_\theta(x); \theta \in \Omega\},$$

   where $f_\theta(x) = k\theta e^{Q(\theta)t(x)} h(x)$.

   Prove that if P is of the above form, then T=t(x) is complete sufficient.

   b) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $R(\theta_1, \theta_2)$. Is T = $(X_{(1)}, X_{(n)})$ sufficient for $\theta = (\theta_1, \theta_2)$? Is it complete? Give reasons for your answers. 4+6=10

5. a) Find the expression for variance of U_n, a U-statistic based on n iid observations and a symmetric Kernel of size m (<n). also find the limit of [n Var (U_n)] as $n \to \infty$.

   b) State the result on asymptotic normality of $U_{n'}$. 8+2=10

6. a) Derive Bayes estimate of a real parametric function $\gamma(\theta)$ under squared error loss.

P.T.O.

b) suppose $X_i; i = 1(1)n$ follows Bernoulli with parameter θ. The prior distribution of θ is Beta with parameters α and β. Find Bayes estimate of θ under squared error loss. 4+6=10

7. Write short notes on **any two** of the following : 5+5=10

a) Bhattacharya's lower bound and its use

b) Minimum Chi-square method of estimation

c) U-statistic and Kendall's τ

d) Loss functions in Bayesian estimation

M.Sc Semester I Examination, 2018

Statistics

MSC-15(Practical)

Time: Four Hours · Full Marks: 40

(Computer Laboratory may be used, if necessary)

1. Following data represent a random sample of size from the Cauchy population with the probability density function $f(x,\theta)=\frac{1}{\pi}.\frac{1}{1+(x-\theta)^2}; -\infty < x,\theta < \infty$. Find out the MLE of θ.

The observations are 3.7807, 2.9957, 5.2043, 4.9893, 2.6874, 4.5957, 4.9367, 3.4996, 3.6174.

Without assuming any distribution, find out nonparametric estimate of mean and variance functional. 7

2. For double genetically data with some value of π, for both parents, following distribution is obtained:

	D1D2	D1R2	D2R1	R1R2
Frequency:	191	34	36	26
Probability:	$\frac{2+p}{4}$	$\frac{1-p}{4}$	$\frac{1-p}{4}$	$\frac{p}{4}$

where $p=(1-\pi)^2$. Find Maximum Likelihood Estimate of π and estimate its standard error. 7

3. Find the eigenvalues and eigenvectors of the matrix 6

$$\begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$$

4. We have the following football league data 15

Index	Y	X_1	X_2	X_3	X_4	Index	Y	X_1	X_2	X_3	X_4
1	10	1985	38.9	59.7	2205	15	6	2140	39	59.2	1901
2	11	2855	38.8	55	2096	16	5	1730	37	54.4	2288
3	11	1737	40.1	65.6	1847	17	5	2072	35	49.6	2072
4	13	2905	41.6	61.4	1903	18	5	2929	41	54.3	2861
5	10	1666	39.2	66.1	1457	19	6	2268	38	58.7	2411
6	11	2927	39.7	61	1848	20	4	1983	39	51.7	2289
7	10	2341	38.1	66.1	1564	21	3	1792	40	61.9	2203
8	11	2737	37	58	1821	22	3	1606	40	52.7	2592
9	4	1414	42.1	57	2577	23	4	1492	36	57.8	2053
10	2	1838	42.3	58.9	2476	24	10	2835	35	59.7	1979
11	7	1480	37.3	67.5	1984	25	6	2416	39	54.9	2048
12	10	2191	39.5	57.2	1917	26	8	1638	40	65.3	1786
13	9	2229	37.4	58.8	1761	27	2	2649	37	43.8	2876
14	9	2204	35.1	58.6	1709	28	0	1503	39	53.5	2560

where. Y: Games won (per 14 game season), X_1: Passing yards (season), X_2: Punting average (yards/punt), X_3: Percent rushing (rushing plays/total plays), X_4: Opponent's rushing yards (season).

Use this data to find the following with the help of 'R' programming.

a) Fit a multiple linear Regression model relating the number of games won to the team's passing yardage (X_1), the percentage of rushing plays (X_3), and the opponent's yard rushing (X_4).
b) Using the partial F test, determine the contribution of X_3 on the model.
c) Find a 95% confidence interval on β_3.
d) A 95% confidence interval on mean number of games won by a team when $X_1 = 2300$, $X_3 = 56$, and $X_4 = 2100$.
e) A 95% prediction interval on number of games won by a team when $X_1 = 2300$, $X_3 = 56$, and $X_4 = 2100$.
f) What conclusion can you draw from this problem about the consequence of omitting an important regressor from the model?
g) Test the hypothesis that $\rho(Y, X_3) = 0$
h) Draw the RVH of X_1, X_3.
i) Fit the model $Y = \beta 0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$ and test the general linear hypothesis $H0 : \beta_2 - \beta_3 = 0, \beta_1 = 0, \beta_1 + \beta_2 + \beta_4 = 0$.
j) Check for any possible multicollinearity of X2, X3 and X4.
k) Construct a normal probability plot of residuals. Does there seem to be any problem with the normality assumption?
l) Construct and interpret a plot of residuals versus the predicted response.
m) Construct plot of residuals versus each of the regressor variables. Do these plots imply that the regressor is correctly specified?
n) Compute the studentized residuals and the R-student residuals for this model. What information is conveyed by these scaled residuals?
o) Fit another model Y on X1. Compute the PRESS statistic for this model and the model in (a). Based on this statistic, which model is most likely to provide better predictions of new data?

5. Practical note book and Viva voce. 5

****

Visva Bharati University
M.Sc. Semester I Examination 2018
Subject: Statistics (Practical)
Paper: MSC-16

Full Marks: 40
Time: 4 Hrs.

1. Consider the following data matrix

$$X = \begin{pmatrix} 3.374 & 1.153 & 1.438 \\ 3.581 & 1.899 & 0.732 \\ 3.646 & 4.239 & 0.797 \\ 3.752 & 3.037 & 2.225 \\ 1.844 & 1.752 & 1.651 \\ 4.482 & 4.580 & 1.701 \\ 3.751 & 3.479 & 1.434 \\ 3.883 & 1.007 & -0.268 \\ 3.701 & 2.921 & 2.393 \\ 3.606 & 2.767 & 2.534 \end{pmatrix}.$$

Test whether the sample is coming from a trivariate normal distribution with mean vector

$$\underline{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$$

satisfying

$$\begin{aligned} \mu_1 - 2\mu_2 + \mu_3 &= 1 \\ \mu_2 + \mu_3 &= 4 \\ \mu_1 - \mu_3 &= 2. \end{aligned}$$

8

2. Load the data file named "mydata.csv" from the computer. It contains the measurements on 5 variable characteristics, viz. plant height (in cm), canopy (in cm), leaf length (in cm), leaf width (in cm) and leaf area (in sq. cm), for 10 green Tulsi plants (1st 10 rows) and 10 purple Tulsi plants (last 10 rows). Assuming multivariate normality of the data, test whether the mean morphological vector (consisting of these 5 variables) differ significantly among the purple and the green varieties.

(You can't use any package to perform the test. You can use R only for computational purpose. The details of null hypothesis, test statistic, critical region and the results of your analysis should be written in details.)

12

3. Below are given the yields in gm per plot for three varieties of seed cotton

$Variety1$	77	70	63	84	95	81	88	101
$Variety2$	109	106	137	79	134	78	126	98
$Variety3$	46	70	71	65	61	40	47	73

Using the Gauss-Markov Linear Model, test if the varieties differ significantly among themselves. 7

4. For the linear model, the normal equations are

$$\begin{pmatrix} 10 & -1 & -8 \\ -2 & 5 & -3 \\ -8 & -3 & 11 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} 12 \\ 16 \\ -28 \end{pmatrix}$$

(a) Obtain any solution of the normal equation.

(b) Find the maximum number of linearly independent estimable parameter functions.

(c) When is $\lambda_1\beta_1+\lambda_2\beta_2+\lambda_3\beta_3$ estimable? (3+2+3)

5. Practical Note book and Viva-voce. 5

M.Sc. Examination, 2018
Semester-III
Statistics
Course: MSC-31
(Real Analysis and Measure Theory)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. Let $\{A_n\}$ be a sequence of subsets of Ω. Find the $\limsup\{A_n\}$ and $\liminf\{A_n\}$ for the following
i) $A_n = [0, \frac{n}{n+1})$ ii) $A_n = [0, 1-\frac{1}{n}]$ iii) $A_n = [0, 1+\frac{1}{n})$ iv) $A_n = C$ if n is odd, $A_n = D$ if n is even. 10

2. State and prove monotone convergence theorem. Hence or otherwise prove Fatou's lemma. 10

3. If f and g are two integrable functions, show that $f \pm g$ is also integrable and $\int(f \pm g)d\mu = \int f d\mu \pm \int g d\mu$. 10

4. State and prove the continuity theorem of measure. 10

5 (a) Prove that the union of a finite number of closed sets is closed. Does this hold for an arbitrary number of closed sets.

(b) Exhibit an open cover of the set $\{\frac{1}{n} : n \in \mathrm{N}\}$ that has no finite subcover. Is the set compact? 5+5

6. (a) Let $f : S \to \boldsymbol{R}$ be continuous on a compact set S. Prove that $f(S)$ is a compact set.
(b)Find the radius of convergence of the power series $\sum_{n=0}^{\infty} \frac{(x+3)^n}{n+1}$ 5+5

7. (a) Let $\{f_n\}$ be a sequence of functions defined on a set E. Prove that $\{f_n\}$ converges uniformly to f on E if and only if $\sup_{x\in E} |f_n(x) - f(x)| \to 0$ as $n \to \infty$.
(b) Show that $\sum_{n=1}^{\infty} \frac{x}{n + n^2x^2}$ is uniformly convergent for all real values of x. 5+5

M.Sc. Examination, 2018
Semester-III
Statistics
Course : MSC-32

(Categorical Data Analysis and Advanced Data Analysis)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. Distinguish between case-control study and cohort study. Discuss about their relative merits and demerits. 5+5=10

2. a) Define a measure of association between two ordinal categorical variables in a 2×2 contingency table. Find the large sample standard error of your measure.

   b) Develop a test procedure for testing the independence of the variables. (2+5)+3=10

3. a) Discuss about three major components of the GLM.

   b) Show that the likelihood equations for a generalized linear model (with usual notation) can be written as

   $$\Sigma w(y-\mu)\frac{\partial \eta}{\partial \mu} xj = 0 \; i \text{ for each j.}$$

   c) Define Pseudo – R^2 measure for the GLM. 3+5+2=10

4. a) Stating the structural defect of linear probability model, discuss the genesis of logistic regression model.

   b) Interpret the parameters of the logistic regression model with one covariate.

   c) Discuss the fitting procedure of logistic regression model. 3+2+5=10

5. Discuss about the use of EM-algorithm to cluster data which can be modelled as a mixture of univariate normal populations. 10

6. How can you apply Gibbs sampling technique in one way ANOVA model to estimate the parameters and hyper parameters. 10

7. a) Write the bootsrap algorithm to find the standard error of a statistic. Also distinguish between bootstrap and jackknife techniques.

   b) Write a short note on Metropolis – Hastings algorithm. 5+5=10

M.Sc. Examination, 2018
Semester-III
Statistics
Course : MSC-33

(MSE-I Operations Research and Optimization Technique)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. What is an assignment problem? How is it different from traveling salesman problem? Show that if a constant be added to any row and/or any column of the cost matrix of an assignment problem, then the resulting assignment problem has the same optimal solution as the original problem. 3+3+4=10
2. For M/M/c queuing system find the expected number of customers in the system in the steady state and also the expected queue length. Find the cumulative distribution function for the waiting time of a customer who has to wait in a M/M/c queuing system. 5+5=10
3. Briefly describe the phases in Operations Research study. What do you mean by models in OR? Discuss the different types of models that are usually encountered in OR. 3+2+5=10
4. What is two-person zero-sum game? Transform this game to a Linear Programming Problem. Prove that the value of a two-person zero-sum game is unique. 2+3+5=10
5. a) Define an inventory. What are the advantages and disadvantages of having inventories?

   b) In a manufacturing situation the production is instantaneous and demand is D units per year. If no shortage is allowed, show that the optimal manufacturing quantity per run is

   $q = \sqrt{2C_p D / C_h(1 - D/K)}$, where C_p = set-up cost per run,

   C_h = holding cost per unit per year, K= manufacturing rate per unit of time (K > R). 4+6=10
6. What is replacement? Describe some important replacement situations. Discuss replacement policy of equipment that deteriorates gradually with (i) no change in time value of money and (ii) change in time value of money. 1+2+4+3=10
7. Write short notes on **any two** of the following : 5+5=10

   a) (s, S) inventory policy

   b) Duality problem in LPP

   c) Graphical solution of game problem

   d) Queue discipline

M.Sc. Examination, 2018
Semester-III
Statistics
Course : MSC-34 : MSS-3
(Time Series Analysis)

Time : 3 Hours **Full Marks : 40**

Questions are of value as indicated in the margin

Answer **any four** questions

1. Define Time Series. Explain the nature cyclical variations in a time series. How do seasonal variations differ from them? Give an outline of any method of measuring seasonal variations. 2+2+2+4=10

2. a) Explain the exponential smoothing method for forecasting time series. 4

   b) Suggest how the smoothing factor should be determined? 3

   c) What are the advantages and disadvantages of exponential smoothing? 3

3. a) Discuss auto correlation. For the model

   $Y_t = \beta X_t + u_t, u_t = \rho u_{t-1} + \varepsilon_1$

   where $E(\varepsilon_t) = 0$, and $Cov(\varepsilon_t, \varepsilon_{t'}) = \sigma^2 I$

   Find the estimate of β.

   b) Let $S_t = X_1 + X_2 + ... + X_t$, where X_t's are iid$(0, \sigma^2)$ variables. For h>0, find Cov (S_{t+h}, S_t) and comment on stationarity of the process $\{S_t\}$. 5

4. Write down an MA (2) process? Derive the auto-correlation functions and partial auto correlation functions of order 2 and 3 for the process. Which of the component of time series is mainly applicable in the following cases?

   a) Full of death due to advancement in sciences,

   b) Fire in a factory.

   c) Sales of New Year greetings cards. 1+6+3=10

5. What is a periodogram? Describe how the hidden periodicity of a Time Series can be estimated. 2+8=10

6. a) Deduce the power spectrum of the white noise process and justify name of the process.

   b) For the process $Z_t = 10 + a_t + a_{t-1}$ and $Z_t = 10 + a_t - a_{t-1}$ where $\{a_t\}$ is a sequence of uncorrelated N(0,1) variables. Find the covariance functions and spectral density functions of the processes. 4+3+3=10

Visva Bharati University
M.Sc. Semester III Examination 2018
Subject: Statistics (Practical)
Paper: MSC-35

Students are asked to save their codes and outputs in a folder with his/her roll number as name.

Full Marks: 40 Time: 4 Hrs.

1. Consider the following data

t	Y_t
1	0.08368
2	0.07456
3	0.11617
4	0.19309
5	0.21007
6	0.15519
7	0.10032
8	0.03708
9	0.02195
10	0.04116
11	0.02517

   (a) Fit a non-linear regression curve of the form $Y_t = ae^{-bt}t^c + \epsilon_t$.

   (b) Find the bootstrap estimate of the standard error of the parameters. Take the number of bootstrap replications (B) to be 50.

   (c) Draw the histogram of the bootstrap replications of each of the parameters. Also check graphically whether they can be regraded to come from a Normal distribution. 2+5+3=10

2. Consider the following sample from a Poisson distribution.

   7, 5, 6, 7, 3, 5, 7, 0, 9, 4

   Based on the sample, find the bootstrap and jackknife estimate of the standard error of (i) sample range and (ii) sample mode. Draw histograms of the replications in each case. 4+2=6

3. The following table gives the result of a study to compare radiation therapy with surgery in treating cancer in larynx. Use R to perform a suitable test for checking whether the population odds ratio equals unity. Discuss your findings.

	Cancer controlled	Cancer not controlled
Surgery	21	2
Radiation therapy	15	3

5

4. Draw a random sample of suitable size from the Beta (5, 7) distribution using acceptance-rejection sampling (Take your proposal density to be $U(0, 1)$). Construct the histogram of the sample and fit a Beta distribution using the same sample. Draw the actual density and the fitted density over the histogram (on the same graph). 7

5. Consider the following data set reporting the proportions of female children at various ages during adolescence who have reached menarche.

Age	No. of female children at that age	No. of children reaching menarche
9.21	376	0
10.21	200	0
10.58	93	0
10.83	120	2
11.08	90	2
11.33	88	5
11.58	105	10
11.83	111	17
12.08	100	16
12.33	93	29
12.58	100	39
12.83	108	51
13.08	99	47
13.33	106	67
13.58	105	81
13.83	117	88
14.08	98	79
14.33	97	90
14.58	120	113
14.83	102	95
15.08	122	117
15.33	111	107
15.58	94	92
15.83	114	112
17.58	1049	1049

Fit logit and probit models to this data and comment on your fit. 7

6. Practical Note Book and Viva-Voce 5

x——— Best of Luck———x

M.Sc.
Semester III Examination, 2018

Statistics

MSC-36(Practical)

Time: Four Hours Full Marks: 40

Answer all the questions. You may use computer laboratory whenever necessary, in that occasion you need to write the code in your main answer script.

1. Use the concept of dominance to convert the following game into 2×2 game and hence solve the game. 6

	Player B			
	I	II	III	IV
I	2	1	4	0
II	3	4	2	4
III	4	2	4	0
IV	0	4	0	8

2. Find out the optimal assignment and minimum cost for the assignment with the following cost matrix.

	I	II	III	IV	V
A	160	130	175	190	200
B	135	120	130	160	175
C	140	110	155	170	185
D	50	50	80	80	110
E	55	35	70	80	105

6

3. A supper market has two girls ringing up sales at the counters. If the service time for each counter is exponential with mean 4 minutes, and if people arrive in a Poisson fashion at the rate of 10 an hour, (a) what is the probability of having to wait for service? (b) What is the expected percentage of idle time for each girl? (c) Find the average queue length and the average number of units in the system. 3

4. A contractor has to supply 10,000 bearings per day to an automobile manufacturer. He finds that, when he starts a production run, he can produce 25,000 bearings per day. The cost of holding a bearing in stock for one year is 20

paise and the set-up cost of a production run is Rs. 180. How frequently should production run be made? 3

5. Using R-Code generate and plot 150 observations from AR(p), MA(q) and

ARMA(p,q) series. Take p, q at your own choice. Looking at the plots of the generated series comment on stationarity of each of them. 6

6. a) Comment on stationarity of the series

$$Y_t - 2.5Y_{t-1} + 0.5\ Y_{t-2} = e_t$$ 1.5

b) Examine and state whether the following series is invertible or not

$$Y_t = e_{t-1} + .2\ e_{t-1} + 0.3\ e_{t-2}$$ 1.5

7. Use R-code to plot the data available in your lab desktop namely MSC36_Problem in MS-Excel.

a) Fit an appropriate ARIMA model to the data.

b) How well does the model fit the data? Comment based on inspection of residuals.

c) How well does the fitted model predict? Comment based on values within the series (in-sample forecasting) and future forecasting.

d) Plot the original series, the forecasted series along with error bands. 8

8. Practical note book and Viva voce. 5

****