UNIVERSITY OF CALICUT

Entrance Examination for Ph. D Programme in Statistics

MODEL QUESTION PAPER

RESEARCH METHODOLOGY AND STATISTICS

Time : 2 hours **Max. Marks: 100**

Section A: Research Methodology

Part I: Multiple Choice Type Questions.

Answer **ALL** the questions. Each question carries **TWO** marks.

(10 x 2=20 marks)

1. The purpose of Research is to
   - A. To gain familiarity with a or to achieve new insights into it a phenomenon
   - B. To portray accurately the characteristics of a particular individual, situation or a group
   - C. To test a hypothesis of a causal relationship between variables
   - D. All the above

2. Which of the following is the first step in a research process?
   - A. Collecting data
   - B. Formulation of the research problem
   - C. Sampling design
   - D. Literature survey

3. MS word is used for
   - A. Design paint
   - B. Design videos
   - C. Design texts
   - D. Design pictures

4. Which of the following is not a source of literature review?
   - A. Books
   - B. Educational Research Information Center
   - C. Magazines
   - D. Questionnaire

5. If a simple random sample of size 10 is drawn without replacement from a population of 100 units, the probability that a specified unit of the population is included in the sample is:
   - A. $\frac{1}{100}$
   - B. $\frac{1}{10}$

C. $\frac{1}{\binom{100}{10}}$
D. 1

6. For estimating the population proportion P in a class of a population having N units, the variance of the estimator p pf P based on a sample of size n is:
A. $\frac{N}{N-1}\frac{PQ}{n}$
B. $\frac{N}{N-1}\frac{PQ}{N}$
C. $\frac{N-n}{N-1}\frac{PQ}{n}$
D. $\frac{N-1}{N-n}\frac{PQ}{n}$

7. If $f(x) = \begin{cases} \sin\frac{1}{x} \ when\ x \neq 0 \\ 0\ when\ x = 0 \end{cases}$, which of the following is true?
A. f(x) is continuous at all x
B. f(x) is continuous at only x=0
C. f(x) is continuous at $x \neq 0$
D. f(x) is discontinuous at all points

8. The series $\sum_{n=1}^{\infty}\frac{1}{n^2+x^2}$ on $(0,\infty)$ is
A. Uniformly convergent
B. Convergent but not uniformly
C. Not convergent
D. None of the above is true

9. Which of the following vectors are linearly dependent?
A. V_1 =(1,4,1), V_2 =(1,1,1)
B. V_1 =(1,2,1), V_2 =(1,-1,1)
C. V_1 =(-1,2,1), V_2 =(1,1,0), V_3 =(-1,5,2)
D. V_1 =(1,4,2), V_2 =(2,8,4)

10. If C and D are symmetric matrices of same order, then
A. CD is always symmetric
B. CD is skew symmetric
C. CD is symmetric if and only if CD=DC
D. CD is never symmetric

Part II: Descriptive Type Questions

Answer any **FIVE** questions. Each question carries **SIX** marks.

(5x 6=30 marks)

11. Briefly describe the different steps involved in a research process.
12. Explain the different measurement and scaling techniques.
13. In stratified random sampling, explain the various methods of allocation and state the variances of the estimators of the population mean, in such cases.
14. State Rolles theorem. If Rolles theorem is satisfied by the function $f(x) = ax^2 - 3x + 4$ in [2,3] find the value of a.
15. If $P = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is a partition matrix and A is non-singular matrix, prove that $|P|$=$|AD - CB|$, provided AC=CA.
16. Create orthonormal system of vectors for R^3, based on the vectors $U_1 = (1,1,1), U_2 = (1,0,-1), U_3 = (1,-2,1)$.
17. Discuss about the convergence of the geometric series $1 + x + x^2 +$ ....
18. Show that the characteristic roots of an idempotent matrix are either 0 or 1.

Section B: Statistics (Subject of Study)

Part I: Multiple Choice Type Questions.

Answer **ALL** the questions. Each question carries **TWO** marks.

(10 x 2=20 marks)

19. Which of the following is true about $f(x, y) = x^3 - 3x^2 - 4y^2 + 1$.
    A. f(x,y) has stationary points (0,2) and (2,0) and (0,0).
    B. f(x,y) has neither maxima nor minima
    C. f(x,y) has maximum at (2,0) and minimum at (0,0) and (0,2).
    D. f(x,y) has maximum at (0,0), but neither maxima nor minima at (2,0)

20. For a random variable X and its characteristic function $\emptyset_X(t)$, which of the following statements is not true?
    A. $\emptyset_X(0) = 1$
    B. $|\emptyset_X(t)| \leq 1 \; for \; all \; t \in R$
    C. $\emptyset_X(-t) = \overline{\emptyset_X(t)} \; for \; all \; t \in R,$ where $\overline{z}$ denotes the complex conjugate of z
    D. $\frac{d^n}{dt^n} \emptyset_X(0) = \frac{1}{i^n} E(X^n)$ for *n* positive integer

21. If X and Y are independent geometric random variables , then the conditional distribution of $X/(X+Y=t)$ is
   A. Uniform over t+1 points
   B. Negative Binomial.
   C. Uniform over [0,1]
   D. Triangular over [0,1]

22. Let $X_1, X_2, \ldots, X_n$ be a random sample from the p.d.f
   $f(x)=\frac{1}{2}e^{-|x-\theta|}, -\infty < x < \infty, \theta$ unknown.
   Then the mle of θ is,

   A. $\bar{X}=\frac{1}{11}\sum_{i=1}^{n} X_i$
   B. The median of $X_1, X_2, \ldots, X_n$
   C. Max $X_i, 1 \le i \le 11$
   D. $\frac{1}{2}$(Max X_i + Min X_i)

23. Let X be a random variable with probability mass function f_0 under H_0 and f_1 under H_1 defined as follows:

X	:	1	2	3	4	5	6
f_0(x)	:	0.01	0.01	0.01	0.01	0.01	0.95
f_1(x)	:	0.05	0.04	0.03	0.02	0.01	0.85

   Then, which of the following test is a most powerful test of size 0.03?

   A. $\phi(x)=\begin{cases}1, if \quad f_1(x)/f_0(x) \ge 2 \\ 0, \quad otherwise\end{cases}$
   B. $\phi(x)=\begin{cases}1, \quad if\ f_1(x)/f_0(x) \le 4 \\ 0, \quad otherwise\end{cases}$
   C. $\phi(x)=\begin{cases}1, if\ f_1(x)/f_0(x) \ge 3 \\ 0, \quad otherwise\end{cases}$
   D. $\phi(x)=\begin{cases}1, if\ f_1(x)/f_0(x) \le 2 \\ 0, \quad otherwise\end{cases}$

24. In a factorial design with two blocks of 8 plots each in a replication, it was decided to confound the effect ABCD. If two blocks in a replication were incompletely constructed with the following treatment combinations:
   Block I : ab,cd, ac, ad, bc, bd

   Block II: a, b, abc, acd, abd, bcd

   The remaining treatments in Block I and Block II respectively are :

A. (1, c) and (d, abcd)

B. (1, d) and (c, abcd)

C. (c, abcd) and (d, abcd)

D. (1, abcd) and (c, d)

25. Which of the following statements are correct?

(i) Cook's distance D_i always lies between 0 and 1.
(ii) In polynomial regression, the matrix X'X becomes ill-conditioned, if there are small numbers of knots.
(iii) In a linear regression $Y = X\beta + \epsilon$ with usual assumptions having $p = 4$ and $n = 20$, if the third diagonal element of the hat matrix is 0.845, then it can be considered as a leverage point.
(iv) When the regressors are orthogonal, the multicollinearity problem doesn't exist.

A. (i), (ii) and (iii) only
B. (ii), (iii) and (iv) only
C. (i), (iii) and (iv) only
D. (i), (ii) and (iv) only.

26. Let Y_1 and Y_2 be independent random variables with means β and 2β respectively with a finite variance σ^2. If the realized values of Y_1 and Y_2 are 1 and 2 respectively, then the least squares estimate of β and the residual sum of squares are respectively.

A. 1, 0
B. 1, 1
C. 0, 1
D. 1, 2

27. If $X \sim N_p(\mu, \Sigma)$, which of the following is true?
A. The distribution of $(X-\mu)'\Sigma^{-1}(X-\mu)$ is Wishart.
B. The mle of $\mu \; is \; \bar{X}$, and Σ is $S = \frac{A}{N-1}$
C. The Hotelling's T^2 statistic is used for testing the mean when Σ is unknown
D. The distribution of Y=CX is, $N_p(C\mu, C\Sigma)$.

28. Choose the wrong statement
A. The state space of a finite Markov chain contains at least one persistent state.
B. If k is a transient state, $\sum p_{jk}{}^n$ converges to zero.
C. State j is persistent iff $\sum p_{jj}{}^n = \infty$
D. In a finite irreducible markov chain all states are transient.

Part II: Descriptive Type Questions

Answer any **FIVE** questions. Each question carries **SIX** marks.

(5x 6=30 marks)

29. Define limit of a multivariable function and find $\lim_{(x,y)\to(0,0)} \frac{\sin(x^2+y^2)}{(x^2+y^2)}$

30. Let $X_1, X_2, \ldots, X_n$ be i.i.d $P(\lambda)$ random variables. Let $S = S_{100} = \sum_{i=1}^{100} X_i$, $\lambda = 0.02$. Using central limit theorem evaluate $P(S \geq 3)$.

31. Let X and Y be i.i.d random variables with $P(X = k) = 2^{-k}$ for $k = 1,2,3$ .... Find $P(X > Y$ and $P(X > 2Y)$.

32. Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with common distribution specified by $(x, \theta) = \frac{\theta}{x^2}$, $x > \theta$. Construct a UMP size α test for testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. Show that likelihood ratio test coincides with UMP test.

33. Describe a BIBD and establish the parametric relations (i) $vr = bk$ and

    (ii) $\lambda(v-1) = v(k-1)$.

34. If $Y_1 = \alpha_1 + \alpha_2 + \epsilon_1$; $Y_2 = 2\alpha_2 + \epsilon_2$ and $Y_3 = -\alpha_1 + \alpha_2 + \epsilon_3$ where $\epsilon_i \sim N(0, \sigma^2)$ $(i = 1,2,3)$, derive the least square estimates of the parameters α_1, α_2 and σ^2.

35. For the three state Markov chain with transition probability matrix,

$$\begin{bmatrix} 0 & 2/3 & 1/3 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}$$

obtain the invariant distribution.

36. If$\sim N_3(\mu, \Sigma)$, with $\mu' = [3, -3, 5]$ and $\Sigma = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$, find the distribution of

    $3X_1 - 2X_2 + X_3$.

UNIVERSITY OF CALICUT
SYLLABUS

Entrance Examination for Ph. D. programme in Statistics

Section A: RESEARCH METHODOLOGY

Definition, meaning and motivation of Research, Objectives and significance, Types of research, Research approaches, Research methods versus methodology, Criteria of good research, Defining a research problem: Necessity, technique. Research design, Sampling design, Measurement and scaling techniques.

Importance of literature review, Reporting and thesis writing: Structure and components of scientific reports-Types of report-Technical reports and thesis. Bibliography and referencing. Communication practices: Importance of effective communication-Oral presentation- Use of visual aids- Scientific word processing with MS word.

Sampling Methods: Census and sampling, Probability sampling, and non-Probability sampling. SRSWOR and SRSWR. Estimation of population mean, total and proportion. Variance of the estimates and standard error. Determination of sample size. Stratified random sampling, PPS sampling with and without replacement, Ratio estimators and Regression estimators, Systematic sampling, Cluster sampling with equal clusters. Multi stage and multiphase sampling, Non-sampling errors.

Basic Calculus: Real Number system-Mathematical induction, order properties of real number, Archimedean property, supremum & infimum. Limits, continuity, uniform continuity and differentiability of functions, Rolle's theorem, Mean value theorem.

Convergence of sequences and series: Point wise convergence versus uniform convergence of sequences and series, Uniform convergence and continuity. Cauchy condition for uniform convergence.

Vector Algebra: Vector spaces. Subspaces. Linear independence. Basis and dimension. Linear equations. Vector spaces with an inner product.

Algebra of Matrices: Theory of matrices and determinants: Matrix operations. Elementary matrices. Determinants. Generalized inverse of a matrix. Eigen values and reduction of matrices: Classification of quadratic forms. Cayley Hamilton theorem and its applications. Canonical reduction of matrices.

Section B: STATISTICS (Subject of study)

Module 1

Mathematical Analysis:

Riemann – Stieltjes Integral - Definition, Linear properties. Integration by parts. Change of variable. Step functions as integrators. Reduction to a finite sum. Monotonically increasing integrators. Riemanns conditions. Comparison theorems. Functions of bounded variations (concepts only). Sufficient conditions for the existence of Riemann Stieltjes integral. Necessary conditions for the existence of Riemann Stieltjes integral. Mean-value theorems.

Multivariable Calculus - Limit and continuity of multivariable functions. Derivatives. Directional derivatives and continuity. Total derivative in terms of partial derivatives. Taylor's theorem. Inverse and implicit functions. Optima of multivariable functions.

Module 2

Probability Theory:

Sets and classes of events - Events. Algebra of sets (set operations, sequences of sets and limits). Fields. Sigma fields. Minimal field. Partition. Borel field. Random variables: Functions and inverse functions. Limits of random variables: Sigma fields induced by random variables. Limits of random variables. Discrete Probability space. General Probability space. Induced probability space.

Distribution functions of random variables. Decomposition of distribution functions. Expectation and moments, Properties of expectations. Moments and inequalities (Cr, Holder's, Jensen's, Basic and Markov inequalities), Moment generating functions. Characteristic functions: Definition, Properties, Inversion formula.

Independence: Independence of events and classes of events. Independence of random variables.

Convergence of random variables: Convergence in probability, almost sure, in distribution, in rth mean – their mutual implications. Weak and strong laws of large numbers, Central limit theorem- Lindeberge-Levy, Liapounov's, Lindeberg-Feller forms.

Module 3

Distribution Theory:

Standard discrete and continuous univariate distributions and their properties, bivariate Normal distribution and its properties, vector of random variables and their properties, transformation technique. Sampling distributions- chi-square, t, F, non-central chi-square, non-central t, non-central F and their properties.

Module 4

Statistical Inference:

Methods of estimation: methods of maximum likelihood, moments, and percentiles. Bayesian method of estimation.

Sufficient statistics, Factorisation theorem for sufficiency. Exponential family. Minimal sufficient statistics, Ancillary statistics, Complete statistics, Basu's theorem.

Unbiasedness, Best Linear Unbiased estimator(BLUE), Minimum variance unbiased estimator(MVUE), Rao-Blackwell and Lehman-Scheffe theorems. Fisher Information, Cramer Rao inequality and its applications. Confidence intervals-shortest expected length confidence interval, unbiased confidence interval, large sample confidence intervals.

Testing of hypothesis-Most powerful tests, Neyman-Pearson lemma, Likelihood ratio tests, Unbiased and invariant tests, similar tests and locality most powerful tests, SPRT.

Module 5

Design of Experiments:

Basic principles, guidelines of design of experiments. Experiments with single factor. ANOVA. Analysis of fixed effect models – comparison of individual treatment means. Random effect models. Model adequacy checking. Choice of sample size. Regression approach to ANOVA completely randomized block design, randomized block design, Latin square design. Greaco-Latin square design. BIBD, PBIBD, Youden square, Lattice design. Factorial design- two factor factorial design, 2^k factorial experiments, confounding techniques.

Module 6

Regression Analysis:

Simple and Multiple linear regression models, least square estimation, maximum likelihood estimation, Gauss-Markov theorem, estimation with linear restrictions, design matrix of less than full rank-generalized least squares, hypothesis testing, multiple correlation coefficient, confidence intervals and regions, simultaneous interval estimation, polynomials in one variable, piecewise polynomial fitting, polynomial regression in several variables, effect of outliers, detecting and dealing with outliers, non-constant variance and serial correlations, departures from normality, diagnosing collinearity, ridge regression.

Module 7

Multivariate Analysis:

Multivariate Normal distribution and properties, Distribution of quadratic forms, independence of a linear form and quadratic form, independence of two quadratic forms. Partial and multiple correlation coefficients, partial regression coefficient, estimation of mean vector and covariance vector, distribution of sample mean vector, Wishart distribution, generalized variance, Mahalnobis D^2 and Hotelling's T^2 statistics, testing the equality of mean vector and equality of dispersion matrices, classification problem, principal component analysis.

Module 8

Stochastic Process:

Concept of stochastic processes, Markov chain, Chapman Kolmogorov equations, classification of states, limiting probabilities, Gamblers ruin problem, branching processes, Counting process, Properties of Poisson processes, non –homogenous Poisson process, compound Poisson process, conditional mixed Poisson process, continuous time Markov Chains, birth and death processes, renewal processes, renewal reward process, regenerative processes, semi-Markov process, insurers ruin problem, basic characteristics of queues, M/M/1, M/M/s, M/G/I, and G/M/I models.
