TAMIL NADU PUBLIC SERVICE COMMISSION
STATISTICS
(POST GRADUATE DEGREE STANDARD)

CODE: 410

Unit - I: Probability and Random Variables

Introduction to probability: Random experiments, sample space and events, definition of probability - classical, empirical and axiomatic approaches to probability; addition and multiplication theorem, conditional probability and Bayes' theorem. Random variables and distribution function, distribution function of a random vector – Mathematical expectation and conditional expectation - Markov inequality - Chebyshev's inequality - Convergence in probability – Convergence in distributions - Weak and strong laws of large numbers – Central limit theorems (Lindeberg-Levy, Liapunov's and Lindeberg-Feller).

UNIT – II: Distribution Theory

Introduction to distributions: Marginal and conditional distributions – Generating functions: MGF, PGF and CGF – Characteristic functions.
Discrete distributions: Binomial, Poisson, Negative binomial, Uniform and Hyper geometric distribution.
Continuous distributions: Uniform, Normal, Cauchy, Beta, Gamma, Exponential, Weibull, Pareto and Log-Normal.
Sampling distributions: t, F and Chi-square distributions and their properties.

UNIT- III: Estimation Theory

Introduction to estimation theory: Consistency, unbiasedness, sufficiency, efficiency and completeness.
Theorems and Inequalities: Cramer-Rao inequality, Chapman-Robbins inequality, Rao-Blackwell, Lehman-Scheffe, Neyman-Fisher factorization theorems with examples.
Methods of Estimation: Methods of moments, maximum likelihood, minimum chi-square and least squares, Bayesian estimation (with examples) – Confidence intervals for large and small samples.

UNIT – IV: Testing of Hypothesis and Non-Parametric tests

Introduction to Testing of Hypothesis: Null and alternative hypotheses, simple and composite hypotheses, two type of errors – Critical region – Size and level of a test - Power function.
Tests: Most powerful test, Neyman-Pearson lemma: UMP and unbiased tests - MLR property and its uses for construction of UMP test.

Non–Parametric tests: Run test, Median test, Sign test, Mann–Whitney test, Wilcoxon test, Kolmogorov-Smirnov test (one and two sample test procedures), Kruskal-Wallis test and SPRT Test.

UNIT – V: Regression Analysis

Simple and Multiple regression models: Description of data model – Estimation and testing of hypothesis on regression coefficients – Adequacy measures – Predicted values and standard error – Evaluation of fit – Analysis of residuals.
Multicollinearity and its effects on inference and forecasting – Selection of variables – Forward selection and backward elimination procedures (step-wise method).
Introduction to Generalized Linear Models: Components of GLM - Logistic regression model - Fitting and interpretation.

UNIT – VI: Sampling Theory

Introduction to the theory of Sampling: Sampling designs – estimation procedures – properties of estimators – SRSWR and SRSWOR and their properties.
Systematic and Stratified Sampling methods, Ratio and Regression estimators, estimation under double sampling – Sampling and non-sampling errors – Cluster sampling – Two stage and Multistage sampling – Randomized response techniques - Sample survey organizations – CSO and NSO.

UNIT – VII: Design of Experiments

Contrasts: Linear and orthogonal contrasts – Linear models: Fixed, random and mixed effect models.
Principles of experimental designs – Construction and analysis of CRD, RBD, LSD, 2^n and 3^n factorial experiments - Partial and complete confounding – BIBD – PBIBD - Split plot design - Youden Square design – Lattice design.

UNIT – VIII: Statistical Quality Control (SQC)

Introduction to statistical process and product control: Control charts – Shewhart control charts for variables and attributes: X-bar, R, S, p, np, c and u charts – CUSUM charts – OC curve for control charts.
Acceptance sampling: Sampling inspection – AOQL, LTPD, producers' and consumers' risks - Single, double, multiple and sequential sampling plans for attributes and variables – OC, ASN, ATI and AOQ curves. Implementing six sigma – over view and implementations.

UNIT – IX: Time series and Index numbers

Concepts of time series, additive and multiplicative models, decomposition into components, determination of trend by free hand method, moving averages, fitting of mathematical curves, seasonal indices and the estimate of the variance for random components, autoregressive, moving averages and ARIMA models.

Definition, construction, interpretation and limitations of index numbers - Laspeyre's, Paasche's, Marshall–Edgeworth, Dorbish-Bowley, Fisher's index numbers and their comparisons for ideal index number - Construction for cost of living index and wholesale price index.

UNIT – X: Machine Learning Techniques through R and Python Languages

Overview of R Language – Defining the R project – Objects and data structures – Graphics using R language – Calculation of measures of central tendency, dispersion, correlation and fitting of regression lines (Linear and logistic).
Overview of Python Language – Regular expressions – Scientific libraries: Numpy, Scipy, Matplotlib and Pandas.
Machine Learning: Supervised learning – Classification (KNN and Naive Bayes) and Regression (Linear and Logistic) techniques – Unsupervised learning (Clustering methods).

Note: Medium of Instruction is **English only**