Revised Course work Structure and Syllabus of Ph.D. Statistics (Effective from 2021-22)

Structure: The course work consists of three Parts, **A, B** and **C,** totalling **18 credits** with the following details.

1. **Part A: Research Related Courses – 6 credits**
   a. STR RM : Research Methodology – 4 credits
   b. STR PE : Publication Ethics – 2 credits

2. **Part B: Subject Related Courses - 8 credits**
   a. STR MF : Mathematical Foundations for Research – 2 credits
   b. STR AP: Advanced Probability Concepts– 3 credits
   c. STR AI: Advanced Inference Concepts - 3 credits

3. **Part C: Research Related Activities - 4 credits**
   a. Prepare Proposals for Funding – 1 credit
   b. Seminar – Paper Presentation (2 National/International) - 2 credits
   c. Review of Research Papers (1 Theory + 1 Application) – 1 credit

Duration: A student is expected to complete PART A and PART B of the coursework within one year of the admission. However, under special circumstances, the Head of the Department, in consultation with the Department Research Council can extend the period for the completion of the coursework.

Equivalent Courses: The Departmental Research Council can decide the equivalent courses (from M.Sc. curriculum of Statistics/Mathematics/Computer Science/CMS/IDSSC Departments or NPTEL/SWAYAM courses with additional topics/exercises) for the courses in Parts A and B.

Evaluation: Concerned teaching faculty can opt for evaluation types such as tutorials, quizzes, seminars, assignments and tests (open/closed). However, for each course there shall be a minimum of three pieces of assessments. The final grade/percentage of marks with the details from three pieces should be submitted to the Ph.D. Course Coordinator/Head of the Department at the end of the course work.

Grades/Marks: For all the courses, the grading will be as follows:
O grade – 90 to 100 %; A grade – 75 to 89 %; B grade – 60 to 74 %; C grade – 50 to 59 %; Fail – Less than 50 %

Failed Candidates: Failed candidates will have to undergo the evaluation process again. The concerned faculty's decision would be final in this respect.

Part A: Research Related Courses

Course Code: STR RM
Course Name: Research Methodology
Credits: 4

Objectives:
The objective of the course is to introduce the meaning and scope of research. Also, this course is to expose the students to present day tools and techniques of handling problems related to statistical computations.

Learning Outcomes:
After completion of the course, students would be able to understand the meaning and scope of doing scientific research. Also, they would be able to use most of the advanced computational algorithms and tools used in modern statistical inference problems.

Detailed Syllabus:

1. **Scientific Research Methods**

Objectives and purpose of research, Philosophical foundation for knowledge creation and dissemination, Epistemological, Ontological and other issues in science research,

Qualitative and quantitative research, different methods

Role of statistics in scientific research, research design, statistical research project

Types of statistical research: empirical, field experiments, laboratory experiments, and secondary sources of data, exploratory and confirmatory research, planned and ad-hoc methods of data collection, non-response and methods of recovering the missing response

2. **Missing values and Imputations Methods:**

Missing values and types of missingness, imputations methods for missing values, single and multiple imputations. MCMC methods for missing values, EM algorithm and applications: EM algorithm for incomplete data, EM algorithm for mixture models, EM algorithm for missing values, stochastic EM algorithm.

3. **Bootstrap, Jackknife and Cross validation Methods:**

Bootstrap methods, estimation of sampling distribution, various types of confidence intervals, variance stabilizing transformation, jackknife and cross-validation, permutation tests. bagging and boosting methods with applications. cross validation

4. **Smoothing Methods:**

   Kernel estimators, nearest neighbor estimators, orthogonal and local polynomial estimators, wavelet estimators, splines, choice of bandwidth and other smoothing parameters.

5. **Optimization Methods:**

   Basic optimization problem, constraints: linear and nonlinear, local and global solutions, derivatives and gradients based methods, first and second order methods, stochastic methods genetic and particle swarm method, simulated annealing, dynamic programming

Reference Books:

1. Buuren, Stef van (2012). Flexible Imputation of Missing Data. Chapman and Hall.
2. Chihara, L. and Hesterberg, T. (2011) Mathematical Statistics with Resampling and R. Wiley.
3. Davison, A.C. and Hinkley, D.V. (1997) Bootstrap methods and their Applications. Chapman and Hall.
4. Effron, B and Hastie, T (2016). Computer-Age Statistical Inference-Algorithms, Evidence and Data Science, Cambridge University Press.
5. Engelbrecht, A. P. (2007). Computation Intelligence: An Introduction, Wiley
6. Gilks, W. R., Richardson, S., and Spiegelhalter, D. (eds.) (1995) Markov Chain Monte Carlo in Practice. Chapman and Hall.
7. Good, P. I. (2005) Resampling Methods: A Practical Guide to Data Analysis. Birkhauser Bosel.
8. Jim, A. (2009). Bayesian Computation with R, 2nd Edn, Springer.
9. Kochenderfer, M. J. and Wheeler, T. A. (2019). Algorithms for Optimization, MIT Press,
10. Kothari C.R (2014) - Research Methodology, Third Edition, Wiley Eastern Limited
11. McLachlan, G.Y. and Krishnan, T. (2008). The EM Algorithm and Extensions. Wiley.
12. Nocedal, J. and Wright, S. J. (2006), Numerical Optimization, Springer
13. Patten. M. L. and Newhart, M. (2017). Understanding Research Methods: An Overview of Essentials, 10th Ed., Routledge.

Course Code: STR PE
Course Name: Publication Ethics
Credits: 2

Objectives:

Publication is an important activity associated with any type of research. The objective of this course is to make the research student aware about the ethics associated with publication activity.

Learning Outcomes:

As an outcome of this course, the researcher is expected to follow good ethical practices in his/her future research publications.

Detailed Syllabus:

1. **Philosophy and Ethics:**
   Introduction to philosophy: definition, nature and scope, concept, branches. ethics: definition, moral philosophy, nature of moral judgements and reactions

2. **Scientific Conduct:**
   Ethics with respect to science and research, intellectual honesty and research integrity. scientific misconducts, falsification, fabrication, and plagiarism (FFP). Redundant publications: duplicate and overlapping publications, salami slicing. selective reporting and misrepresentation of data.

3. **Publication Ethics:**
   Publication ethics: definition, introduction and importance. Best practices/standard setting initiatives and guidelines: COPE, WAME, etc. conflicts of interest. Publication misconduct: definition, concept, problems that lead to unethical behavior and vice versa, types. Violation of publication ethics, authorship and contributorship, identification of publication misconduct, complaints and appeals. predatory publishers and journals.

4. **Open Access Publishing:**
   Open access publication and initiatives. SHERPA/ROMEO online resource to check publisher copyright & self-archiving policies. Software tool to identify predatory publications developed by SPPU. Journal finder/Journal suggestion tools viz. JANE. Elsevier journal finder, Springer journal suggester, etc.

5. **Publication Misconduct.**
   5.1 Group discussions: Subject specific ethical issues, FFP, authorship, conflicts of interest, complaints and appeals: examples and fraud from India and abroad.
   5.2 Software tools: use of plagiarism software like Turnitin, Urkund and other open-source software tools.

6. **Databases and Research Metrics:**
   6.1 Databases: (i) Indexing databases (ii) Citation databases: Web of Science, Scopus, etc.
   6.2 Research Metrics: (i) Impact Factor of Journal as per Journal Citation Report, SNIP, SJER, IPP, Cite Score. (ii) h-index, g index, i10 index, altmetrics

Books Recommended:

1. Beall, J. (2012). Predatory publishers are corrupting open access. Nature, 489 (7415), 179.

2. Bird, A. (2006). Philosophy of Science. Routledge.
3. Chaddah, P. (2018) Ethics in Competitive Research: Do not get scooped; do not get plagiarize.
4. Maclntyre, Alasdair (1967). A Short History of Ethics. Routledge, London.
5. Muralidhar, K., Ghosh, A. and Singhvi, A. K.(2019). Ethics in Science Education, Research and Governance, Indian National Science Academy, New Delhi.
6. National Academy of Sciences, National Academy of Engineering and Institute of Medicine. (2009). On Being a Scientist: A Guide to Responsible Conduct in Research: Third Edition. National Academies Press.
7. Resnik, D. B. (2011). What is ethics in research & why is it important. National Institute of Environmental Health Sciences, 1-10.

Video Resources:

1. Philosophy and Ethics: https://www.youtube.com/watch?v=ONdfTAOblJ8 ; https://www.youtube.com/watch?v=avv8x5FaSIA ; https://www.youtube.com/watch?v=K7S-Y7P3mN4
2. Scientific Conduct: https://www.youtube.com/watch?v=GA0faeZXP8w ; https://www.youtube.com/watch?v=sUEreDQpWeg&t=16s
3. Publication Ethics: https://www.youtube.com/watch?v=tFq5U-1nb8Q ; https://www.youtube.com/watch?v=URz1ewvc-xw; https://www.youtube.com/watch?v=fGgwNClHyCo ; https://www.youtube.com/watch?v=CxC6SJ5Q7FA&t=180s ; https://www.youtube.com/watch?v=e1UV1glccLU; https://www.youtube.com/watch?v=LmMDIBENHhU&t=78s; https://www.youtube.com/watch?v=S4WaifPcsl0; https://www.youtube.com/watch?v=fGgwNClHyCo&t=35s
4. Open Access Publishing: https://www.youtube.com/watch?v=CFa2QeMgk9k; https://www.youtube.com/watch?v=n1ZySivYQ3w; https://www.youtube.com/watch?v=msH-vW2tTio
5. Publication Misconduct: https://www.youtube.com/watch?v=Qcf5IZmRckY; https://www.youtube.com/watch?v=_OmEaqqI3-c; https://www.youtube.com/watch?v=goKHcO4JbGs;
6. Databases & Research Metrics: https://www.youtube.com/watch?v=cD9xVUGEzmM; https://www.youtube.com/watch?v=7HqQ7nk2Z-4; https://www.youtube.com/watch?v=WKDmC7q_scw

Part B: Subject Related Courses - 8 credits

Course Code: STR MF
Course Name: Mathematical Foundations for Research
Credits: 2

Objectives:

The objective of this course is to provide basic fundamental mathematical ideas of analysis, calculus and linear algebra needed for Statistics research.

Learning Outcomes:

Learning this course can help the student to create sharp logical arguments, clarity and correctness in statements and proofs related to the research.

Detailed Syllabus:

1. **Mathematical analysis**
   Real and complex number systems
   Basic notions of set theory
   Elements of point-set topology
   Limits and related theory
   Continuity and related theory
2. **Calculus of single and several variables**
   Calculus of single variable - Derivatives and related theory
   Calculus of several variables – Derivatives (partial, total, directional etc.) and related theory, Implicit functions and extremum problems
   Integrals and related theory - Riemann, Riemann-Stieltjes and Lebesque
   Multiple Integrals –Riemann, Lebesque, Line and Surface
3. **Linear algebra**
   Vector space and related theory, system of linear equations
   Linear Transformations
   Orthogonality, determinants, eigen value theory
   LU Factorization, Cholesky factorization, spectral decomposition, singular value decomposition
   Quadratic forms and related theory

Books Recommended

1. Apostol, T. M. (1975). Mathematical Analysis: A Modern Approach toAdvanced Calculus, Addison – Wesley

2. Ghorpade, S. R. and Limaye, B. V. (2006). A Course in Calculus and Real Analysis, Springer
3. Kreyszig, E. (1975). Advanced Engineering Mathematics, Wiley Eastern
4. Lay, D. C. Lay, S. R. and Mc Donald, J. J. (2016) .*Linear Algebra and Its Applications*, Fifth Edition, Pearson, Boston.
5. Ramachandra Rao, A. and Bhimasankaram, P. (2000). *Linear Algebra*. Hindustan Book Agency
6. Rao, C. R. (1995). *Linear Statistical Inference and Its Applications*, Wiley
7. Rudin, W. (1985). Principles of Mathematical Analysis, McGraw - Hill
8. Searle, S. R. and Khuri, A., (2017). *Matrix Algebra Useful for Statistics*, 2nd Edn., John Wiley, NewYork

Course Code: STR AP
Course name: Advanced Probability Concepts
Credits: 3

Objectives:

The objective of this course is to provide theoretical foundations of probability theory which are needed in statistics research.

Learning Outcomes:

After completion of the course, students would be able to (i) apply probability concepts in Statistics courses and (ii) apply probability and related concepts in theoretical development of research ideas wherever necessary.

Detailed Syllabus:

1. **Measure and Integration**

   Measures, Lebesgue-Stieltjes measures, Caratheodory extension theorem, Completeness of measures, Measurable transformations, Induced measures, distribution functions, Integration, Riemann and Lebesgue integrals

   The Lebesgue-Radon-Nikodym theorem, Signed measures, Functions of bounded variation, Absolutely continuous functions on R, Singular distributions, Jordan Decomposition of a cdf, Introduction to Product spaces and product measures, Convolutions

2. **Probability Spaces**

   Random variables, random vectors, random processes, Kolmogorov's consistency theorem

3. **Convergence and Laws of Large Numbers**

Review of different modes of convergence and their implications, Weak convergence, Skorohod's theorem and the continuous mapping theorem. Concept of tightness

Weak laws of large numbers, strong laws of large numbers (Kolmogorov's), Introduction to ergodic theorem

4. **Generating Functions**
   Generating functions, Laplace and Fourier transforms, Definition and examples, Characteristic functions, Inversion formulas, Levy-Cramer continuity theorem

5. **Central Limit Theorems**
   Lindeberg-Feller theorems, Stable distributions, infinitely divisible distributions, CLT for stationary ergodic sequences, Empirical process and Brownian bridge

6. **Martingales**
   Conditional Expectation and probability, Discrete Parameter Martingales, Martingale convergence theorems, SLLN, Central limit theorem, mixing sequences

Books Recommended:

1. Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer.
2. Billingsley, P. (1995), *Probability and Measure*, 3rd edn, John Wiley.
3. Chung, K. L. (2001). *A Course in Probability Theory*, Academic Press.
4. Klenke, A. (2014). *Probability Theory: A Comprehensive Course*. Springer.
5. Leadbetter, R., Cambanis, S. and Pipiras, V. (2014). *A Basic Course in Measure and Probability Theory for Applications*. CUP
6. Robert B. Ash (2000). *Probability & Measure Theory*, Academic Press
7. Royden, H. L. (1988), *Real Analysis*, 3rd edn, Macmillan.
8. Williams, D. (2004). *Probability with Martingales*, CUP.

Course Code: STR AI
Course Name: Advanced Inference Concepts
Credits: 3

Objectives:
The objective of this course is to provide theoretical foundations of statistical inference.

Learning Outcomes:
After completion of the course, students would be able to develop the inferential tools for their research and establish/investigate the properties of the suggested estimators and testing procedures.

Detailed Syllabus:

1. **Principles of Inference**
   Algorithms and inference, frequentist inference, frequentism in practice, frequentist Optimality, flaws in frequentist inference, Bayesian/frequentist comparison

   Likelihood principle, sufficiency principle, conditionality principle, implications

2. **Estimation Theory**
   A review of UMVU estimation, A review of CAN estimation theory.

   Maximum likelihood estimates, inconsistent MLEs , MLEs in the exponential family, more general cases and asymptotic normality, observed and expected Fisher information, asymptotic optimality of the MLE and super efficiency, loss of information and Efron's curvature.

3. **Asymptotic Theory in Testing of Hypothesis**
   Likelihood ratio tests and asymptotic theory of likelihood ratio test statistics, distribution under alternatives, asymptotic efficiency in testing, asymptotic distribution of Pearson's chi-square, asymptotic distribution under alternatives and consistency

4. **Large-Scale Hypothesis Testing and FDRs**
   Chi-square tests with many cells and sparse multinomials, regression models with many parameters: Multiple testing and false discovery definitions, Benjamini-Hochberg rule, distribution theory for false discoveries and Poisson and first-passage asymptotics, lower bounds on the number of false hypotheses, newer FDR

5. **Bayesian Methods**
   Fundamental elements, general form of Bayes rules, minimax, Bernstein-vonMises theorem, shrinkage and the James–Stein estimator, ridge regression, empirical Bayes, choice of prior distributions, hierarchical modeling, predictive distributions, posterior Consistency, introduction to nonparametric Bayesian methods

6. **Causality and Causal Inference**
   Introduction to causality and causal inference, theory of inferred causation, identification of causal effects, examples of causality and structural models in various disciplines

Books Recommended:

1. Anirban Das Gupta (2008). *Asymptotic Theory of Statistics and Probability*. Springer
2. Deshpande, J. V., Naik-Nimbalkar, U. N., Dewan, I. (2018). *Nonparametric Statistics: Theory and Methods*. Worl-Scientific.
3. Efron, B., Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, CUP.
4. Fergusson, T. S. (1968). *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press.
5. Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, Springer.
6. Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*, Springer
7. Pearl, J. (2009). *Causality: Models, Reasoning and Inference*.CUP
8. Young, G. A. and Smith, R. L. (2005). *Essentials of Statistical Inference*, CUP.