

Statistics & Data Science

@

Indian Institute of Technology, Kanpur

Beginning academic year 2021-2022, the Department of Mathematics & Statistics at the Indian Institute of Technology, Kanpur will offer a new BS program in “Statistics and Data Science”. Like any other four-year BS or BTech program offered by the Institute, admissions to this new BS program will also be made through the Joint Entrance Examination (Advanced) conducted by IITs. Salient features of this program are described below:

- The program is credit-based offering flexibility in the course curriculum. Students admitted to this program, like any other four-year Bachelor’s program offered by the Institute will have the option to earn a Master’s degree in “Statistics and Data Science” by spending one more year over and above the Bachelor’s program (normal duration to earn BS-MS or BTech-MS dual degree is 4+1=5 years). There will be no direct admissions to a dual-degree (BS-MS or BTech-MS) program in “Statistics and Data Science”.
- Students admitted to any other four-year Bachelor’s program (BS or BTech) offered by the institute will have the option to earn a second major in “Statistics and Data Science” by spending one year extra (normal duration to earn two majors is 4+1=5 years).
- In addition, students admitted to any other four-year Bachelor’s program offered by the Institute will have the option to earn a minor in “Statistics and Data Science” by appropriately choosing open elective courses of the first major.

This new BS program in “Statistics and Data Science” will be in addition to the following existing programs offered by the Department, viz., (i) Four-year BS program in Mathematics and Scientific Computing; (ii) Two-year MSc program in Mathematics; (iii) Two-year MSc program in Statistics, (iv) PhD in Mathematics, (v) PhD in Statistics.

About the new program in Statistics and Data Science

The interface of Statistics and Data Science is a utopian combination of theory and implementation of modern data, with immeasurable use in industry, government, and

academia. Statistical models are grounded in mathematical validity and rigour, yielding stable inference. Adding to that, the sophisticated computational framework afforded by Data Science methods, yields a combination that can bring significant improvement to solutions of diverse real-world problems. This tight walk between theory and usability requires training specific to a unique set of skills. Based on this realization, the Department of Mathematics and Statistics at IIT Kanpur (IITK) is starting the BS program in Statistics and Data Science.

On a global level, in the past ten years, there has been a significant change in the research and teaching principles of Statistics. Theoretical and mathematical properties of data are no longer studied in isolation, but in conjunction with practical considerations of real-world implementations. There is an ever-increasing focus on the computational aspects of statistical methodologies. Somewhat parallelly, the larger data science and AI community has realized the importance of interpretable AI, where the need of the hour is explainable statistical models. Research in computational statistics in the next few decades will focus on developing mathematically rooted models that can be implemented efficiently. With this vision, our goal is to make IIT Kanpur the incubator for future thought leaders in this area with the new BS and BS-MS program in Statistics and Data Science.

The program focusses on three wide and relevant areas of coursework, viz., (a) Fundamental Statistical and Mathematical Courses, (b) Statistics and Data Science Courses, and (c) Computational Courses. One of the major aims of the program is to provide skilled professionals in the rapidly growing areas related to big data analytics. Students of this program will be exposed to various types of structured and unstructured data and will work on solving real world problems. These would include data related to health, biomedicine, bioinformatics and digital health, banking and financial data; speech, signal and image processing data; seismological data; social media and social network analysis; energy sector data, etc. For the students, there will be abundant avenues and opportunities to analyse and infer from huge repository of data available from various sources including the Institute itself. In particular, they would effectively work on health-related data and digital health, in collaboration with the proposed School of Medical Research and Technology to give a much-needed boost to research and analytics in the related areas in present times.

To keep the students acquainted with new challenges with data and cutting-edge technology, the Institute will invite industry personnel for guest lectures, real data analysis projects, seminar presentations, and possible internships for students.

The program, which has the promise of being eminent and unique in the landscape of courses offered in India and Asia, will be a star attraction for students interested in the study and

analysis of data. Students graduating with a degree in Statistics and Data Science will not only be well suited to build flourishing careers in industry and the new entrepreneurial India, but also to pursue higher studies in classical and modern statistics and data science.

Admission Process:

- Students for the BS program in “Statistics and Data Science” would be admitted through the JEE (Advanced) examination conducted by IITs. There will not be any direct admissions to five-year dual-degree BS-MS program in “Statistics and Data Science”. The BS students, during BS program, will have the option to convert their BS program to 5-year dual degree BS-MS program subject to satisfying certain eligibility requirements as laid down in the Undergraduate Manual (UG Manual) and Ordinances of IIT Kanpur. Students of the BS-MS program, in addition to the requirements of BS program, will have the opportunity to do advanced courses in Statistics and Data Science. They would also be required to do 36 credits of PG Project, which would involve students in handling applied / theoretical projects directed towards solving real-life problems.
- Eligibility criteria would be same as that of other four-year programs offering admissions through JEE (Advanced).
- For details on JEE (Advanced) and Eligibility criteria visit: <http://jeeadv.ac.in>

Course Templates:

Like any other UG program, the BS program in “Statistics & Data Science” is credit based with following credit requirements:

Minimum credit requirement for graduation

Bracket	Credits	Percentage
Institute Core (IC)	124	28.8
Department Compulsory (DC)	111	25.8
Department Elective (DE)	36	8.4
Open elective / UGP (OE / UGP)	63	14.7
ESO/SO	47	10.9
HSS	49	11.4
Total	430	

Although the BS, BS-MS, BTech-MS and Double Major programs in “Statistics and Data Science” are credit based with full freedom to students to earn requisite credits as per their convenience, suggestive templates are provided below:

BS Statistics & Data Science Course template

I Semester			Credits			II Semester			Credits		
IC: MTH 101A: Mathematics I	3-1-0-0	11	IC:MTH 102A: Mathematics II	3-1-0-0	11						
IC:PHY 103A: Physics II	3-1-0-0	11	IC:PHY 102A: Physics I	3-1-0-0	11						
IC:CHM 101A: Chemistry Laboratory	0-0-3-0	03	IC:PHY 101A: Physics Laboratory	0-0-3-0	03						
IC:ESC 101A: Fundamental of Computing	3-1-3-0	14	IC:LIF 101A: Introduction to Biology	2-0-0-0	06						
HSS: ENG 112A/HSS-1 (Level-1)	3-1-0-0	11	IC:CHM 102A: General Chemistry	2-1-0-0	08						
IC:PE 101A: Morning Exercise	0-0-3-0	03	IC:PE 102A: Evening Exercise	0-0-3-0	03						
			IC:TA 101A: Engineering Graphics	2-0-3-0	09						
	Total	53		Total	51						
III Semester			Credits			IV Semester			Credits		
DC: MTH 301A - Analysis I	3-1-0-0	11	DC: MTH xxxA - Theory of Statistics	3-1-0-0	11						
ESO/SO-1: MSO xxxA - Introduction to Probability Theory	3-1-0-0	11	DC: MTH xxxA - Statistical Computing	3-0-1-0	10						
ESO/SO-2: MSO 202A – Introduction to Complex Analysis (Modular 1 st half)	3-1-0-0	06	DC: MTH xxxA - (Modular 1st half) Elementary Stochastic Processes	3-1-0-0	06						
DC: MTH xxxA - (Modular 2nd half) Matrix Theory-II (Rename)	3-1-0-0	06	DC: MTH xxxA - Data Science Lab II	1-0-2-0	05						
DC: MTH xxxA - Data Science Lab I	0-0-3-2	05	HSS: HSS-2 (Level-1)	3-1-0-0	11						
IC: ESC 201A: Introduction to Electronics	3-1-3-0	14	IC: TA 202A: Manufacturing Process II	1-0-3-0	06						
IC: TA 201A: Manufacturing Process I	1-0-3-0	06	ESO/SO-3:	3-0-0-0	09						
	Total	59		Total	58						
V Semester			Credits			VI Semester			Credits		
DC: MTH 517A - Time Series Analysis	3-0-1-0	10	DC: MTH 535A - Bayesian Analysis	3-0-1-0	10						
DC: MTH xxxA - Linear Regression and ANOVA	3-0-1-0	10	DC: MTH 514A - Multivariate Analysis	3-0-1-0	10						
DC: MTH 399A - Communication Skills	0-0-2-0	02	DC: MTH xxxA - Data Science Lab III	1-0-2-0	05						
ESO/SO-4: ESO 207A - Data Structures and Algorithms	3-0-3-0	12	ESO/SO-5:	3-0-0-0	09						
IC: Com 200: Communication Skills: Composition	1-1-0-0	05	HSS: HSS-4 (Level-2)	3-0-0-0	09						
HSS: HSS-3 (Level-2)	3-0-0-0	09	OE-2/UGP - 2	0-0-9-0	09						
OE-1:	3-0-0-0	09									
UGP -1 (Extra Credit)	0-0-4-0	04									
	Total	57/61		Total	52						

VII Semester			VIII Semester		
		Credits			Credits
DC: MTH 552A - Statistical and AI Techniques in Data Mining	3-0-1-0	10	DE – 3		09
DE - 1		09	DE – 4		09
DE - 2		09	OE – 6		09
OE – 3/UGP-3		09	OE – 7		09
OE – 4		09	HSS: HSS-5 (Level-2)		09
OE - 5		09	UGP4 (Extra Credit)		09
	Total	55		Total	45/54

**Up to 45 credits of internships in lieu of open electives can be taken. This can be done through the courses MTHxxxA Internship I, MTHxxxA Internship II, MTHxxxA Internship III, MTHxxxA Internship IV, MTHxxxA Internship V, of 9 credits each. One would have an option to earn 45 credits of OE through internship courses by spending a full semester in an industry or may do online internships (under one or more OEs) from industry, spread across different semesters. The process for enrolling in the internship courses is as follows: the student identifies a viable internship opportunity in the general realm of statistics and data science and identifies a supervisor in the MTH department. The student, in consultation with the host industry/organization submits a proposal to the Department Undergraduate Committee (DUGC) with the approval of the industry liaison and the departmental supervisor, upon which it will be evaluated for approval and requisite number of credits (in multiples of 9) will be decided. The grading scheme for the internship courses will be S/X.

BS-MS Statistics & Data Science course templates

BS-MS PG (Part – Category A) (from the same program)					
COURSES					
IX Semester		Credits	X Semester		Credits
MS Thesis - I	6-0-0-0	18	MS Thesis - II	6-0-0-0	18
DE PG - I	3-0-0-0	09	DE PG-II	3-0-0-0	09
OE PG - I	3-0-0-0	09	OE PG-III	3-0-0-0	09
OE PG - II	3-0-0-0	09	OE PG - IV	3-0-0-0	09
Total		45			45

Minimum credit requirement in MS part for graduation

Bracket	Credits
PG Component	90
Total	90

Note: A maximum of 36 OE credits may be waived from the department under graduation requirement to be used for the PG requirement of dual degree students.

BS-MS or BTech-MS PG (Part – Category B) (from other programs)					
UG Pre-Requisites					
Odd Semester		Credits	Even Semester		Credits
MTH 301A – Analysis I	3-1-0-0	11	MTH xxxA – Theory of Statistics	3-1-0-0	11
			MTH xxxA – Statistical Computing	3-0-1-0	10
MTH xxxA – (Modular) Matrix Theory II	3-1-0-0	06	MTH xxxA – (Modular)Elementary Stochastic Processes	3-1-0-0	06
MTH 517A – Time Series Analysis	3-0-1-0	10	MTH 535A – Bayesian Analysis	3-0-1-0	10
MTH xxxA – Linear Regression and ANOVA	3-0-1-0	10	MTH 514A – Multivariate Analysis	3-0-1-0	10
MTH 552A – Statistical and AI Techniques in Data Mining	3-0-1-0	10	MTH xxx A – Data Science Lab II	1-0-2-0	05
DC: MTH xxxA - Data Science Lab I	0-0-3-2	05	MTH xxxA – Data Science Lab III	1-0-2-0	05
Total		52			57
PG Requirement					
Odd Semester			Even Semester		
MS Thesis - I	6-0-0-0	18	MS Thesis - II	6-0-0-0	18
DE PG - I	3-0-0-0	09	DE PG-II	3-0-0-0	09
OE PG - I	3-0-0-0	09	OE PG-III	3-0-0-0	09
OE PG - II	3-0-0-0	09	OE PG - IV	3-0-0-0	09
Total		45			45

Note: A maximum of 36 OE credits may be waived from the parent department graduation requirement to be used for the PG requirement of dual degree students.

Double Major in Statistics & Data Science

Double Major					
Odd Semester		Even Semester			
Pre-Requisites					
MSO xxxA – Introduction to Probability Theory or HSO 201A - Applied Probability and Statistics or CS 203A: Mathematics for Computer Science - III	3-1-0-0	11			
MSO 202A – Introduction to Complex Analysis (modular) or MTH403A – Complex Analysis		06			
ESO 207A – Data Structures and Algorithms	3-0-3-0	12			
Mandatory MTH Courses					
MTH 301A – Analysis I	3-1-0-0	11	MTH xxxA – Theory of Statistics	3-1-0-0	11
MTH xxxA – (Modular) Matrix Theory-II	3-1-0-0	06	MTH xxxA – Statistical Computing	3-0-1-0	10
MTH 517A – Time Series Analysis	3-0-1-0	10	MTH xxxA – (Modular)Elementary Stochastic Processes	3-1-0-0	06
MTH xxxA – Linear Regression and ANOVA	3-0-1-0	10	MTH 535A – Bayesian Analysis	3-0-1-0	10
MTH 552A – Statistical and AI Techniques in Data Mining	3-0-1-0	10	MTH 514A – Multivariate Analysis	3-0-1-0	10
DC: MTH xxxA - Data Science Lab I	0-0-3-2	05	MTH xxxA – Data Science Lab II	1-0-2-0	05
			MTH xxxA – Data Science Lab III	1-0-2-0	05
Total		52			57

Note: A maximum of 36 OE credits may be waived from the parent department graduation requirement to be used for the second major requirement of double major students.

Note: “Mathematics and Scientific Computing” and “Statistics and Data Science” students cannot take department compulsory (DC) courses of the other program as Department Electives (DE). That is, “Mathematics and Scientific Computing” students cannot take “Statistics and Data Science” DC courses as DE; and vice-versa.

Minor in Statistics & Data Science

Students pursuing UG programs of other departments/programs will also have option for a minor in Statistics and Data Science. Courses required to obtain a minor would be:

1. MSO xxxA - Introduction to Probability Theory **or**

HSO 201A - Applied Probability and Statistics **or**

CS 203A - Mathematics for Computer Science - III

2. MTH xxxA - Theory of Statistics
3. MTH xxxA - Linear Regression and ANOVA
4. MTH xxxA – Data Science Lab I

The incoming strength would be 5% of the sanctioned strength of the program.

Department Elective (DE) and Open Elective (OE) Courses:

During the course of the BS/BS-MS program the students will have the option of choosing appropriate DE/OE courses as per their desired area of interest. The following gives a list of such possible DE/OE courses.

Suggested Department Electives

- Mathematics and Statistics
 - MTH 202A – Set Theory and Discrete Mathematics
 - MTH 309A - Probability Theory
 - MTH404A – Analysis II
 - MTH 417A - Sampling Theory
 - MTH 418A - Inference I
 - MTH 421A – Ordinary Differential Equations
 - MTH 424A – Partial Differential Equations
 - MTH 515A - Inference II
 - MTH 620A - Measure Theory
 - MTH 673A - Robust Statistical Methods
 - MTH 676A - Econometrics
 - MTH 681A - Statistical Decision Theory
 - MTH 686A - Nonlinear Regression
 - MTH 682A - Order Statistics
 - MTH 657A – Graph Theory
 - MTH 695A - Empirical Processes
 - MTH 516A - Non-Parametric Inference
 - MTH 754A - Probability Theory

- MTH 755A - Statistical Inference
- MTH 784A - Statistical Reliability Theory
- Statistics, Computation, and Data Science
 - MTH 6xxx - ANN/ML Approach for Differential Equations
 - MTH 781A - Statistical Pattern Recognition
 - MTH 707A - Markov Chain Monte Carlo

Suggested Open Electives

- Statistics, Computation, and Data Science
 - CS 633A - Parallel Computing
 - CS 676A - Computer Vision and Image Processing
 - CS 685A - Data Mining
 - CS 698C - Sketching and Sampling for Big Data Analysis
 - CS 698X - Topics in Probabilistic Modeling and Inference
 - CS 771A - Introduction to Machine Learning
 - CS 772A - Probabilistic Machine Learning
 - CS 779A - Statistical Natural Language Processing
 - EE 602A - Statistical Signal Processing
 - EE 604A - Digital Image Processing
 - EE 609A - Convex Optimization in SP/COM
 - EE 623A - Detection & Estimation Theory
 - EE 627A - Speech Signal Processing
 - EE 659A - Computational Intelligence for Machine Vision,
Automation and Control
 - EE 671A - Neural Networks
 - EE 698V - Machine Learning for Signal Processing
 - EE 698U - Optimization for Big Data
 - EE 698L - Artificial Intelligence, Machine Learning & its Applications
 - EE 698R - Advanced Topics in Machine Learning

Career Prospects:

Students graduating with a degree in Statistics and Data Science will not only be well suited to build flourishing careers in analytics and the new entrepreneurial India, but also to pursue higher studies in classical and modern Statistics, and Data Science.

Departmental compulsory courses of the program would focus on three core topics, viz., (a) Fundamental Statistical and Mathematical Courses, (b) Statistics and Data Science Courses, and (c) Computational Courses. Courses of the program have been designed to integrate theoretical and practical aspects with discussions on data applications and analysis, wherever feasible. The new curriculum focuses on bridging the gap between theory and practice. During the course of the program, students will be exposed to various types of structured and unstructured data and will work on solving real world problems. These would include data related to health, biomedicine, bioinformatics and digital health, banking and financial data; speech, signal and image processing data; seismological data; social media and social network analysis; energy sector data, etc. Data Science Lab courses and the lab component of Applied Statistics and Data Science courses would focus on hands-on training with raw data and development of open-source software. Additionally, to keep the students acquainted with new challenges of data and cutting edge technology, the Institute will invite experienced industry personnel for guest lectures, real data analysis projects, seminar presentations, and possible internships for students. Further, the curriculum allows for up to 45 credits of open electives to be taken in the form of an internship in industry. All these would equip the students with necessary knowledge, expertise and experience to have flourishing careers in analytics industry and the new entrepreneurial India.

Students who aspire to go for higher studies in classical statistics/modern statistics/ data science, would have the opportunity to prepare themselves for the challenges of higher studies by choosing appropriate DE/OE courses (in addition to the compulsory DC courses) and research topics in UGP/PGP in that direction during the course of the program.