

INDIAN STATISTICAL INSTITUTE

Students' Brochure

Post-Graduate Diploma in Applied Statistics (PGDAS)

6 August 2024



203 BARRACKPORE TRUNK ROAD
KOLKATA 700108
INDIA

Table of Contents

I. General Information	4
1. Scope	4
2. Eligibility	4
3. Duration	4
4. Selection	4
5. Registration and tuition Fees	5
Payment delays and enrollment	5
Cancellation and refund policy	5
Waiver of tuition fee	6
6. Programme Structure	7
7. Deferment Policy	8
8. Course Enrollment Policy	8
9. Grading	9
a. Grading of individual courses	9
b. Back-paper examination	9
c. Timeline for quizzes, assignments and examinations	9
d. Assignment and examination deadline policy	9
e. Grade dispute	9
10. Conduct	10
11. Final Result	10
12. Award of Certificate	10
13. Change of Rules	11
14. Legal Disputes	11
II. Syllabi of Courses	12
1. Compulsory Courses	12
F1: Basic Statistics	12
F2: Basic Probability	12
F3: Statistical Methods (Prerequisites: F1-F2)	12
F4: Survey Sampling (Prerequisite: F1-F2)	13
2. Specialisations	13
a. Official Statistics Specialisation (Prerequisites: F1-F4)	13
OS1: Management of Data	13
OS2: Advanced Survey Sampling	14
OS3: Population and Social Statistics (Prerequisites: OS7-OS8)	14
OS4: National Accounts Statistics and Price Indices (Prerequisites: OS7-OS8)	15
OS5: Sectoral Statistics (Prerequisites: OS7 and OS8)	15
OS6: Monetary, Financial System and Foreign Trade Statistics (Prerequisites: OS7-OS8)	16
OS7: Introduction to Official Statistical System	16
OS8: Statistics and Economy	17

b. Data Analytics Specialisation (Prerequisites: F1-F4)	17
DA1: Introduction to R and Python	17
DA2: Multiple Regression	18
DA3: Advanced Regression (Prerequisites: DA1 and DA2)	18
DA4: Time Series Analysis and Forecasting (Prerequisite: DA1)	19
DA5: Multivariate Statistics (Prerequisite: DA1)	19
DA6: Statistical Machine Learning (Prerequisite: DA1-DA3)	19
DA7: Advanced Statistical Methods	20
DA8: Specialised Models and Methods	21

I. General Information

1. Scope

This postgraduate diploma programme, delivered online¹, aims at providing the student the statistical tools and concepts necessary to make data-driven decisions and advance their career in the fields of applied statistics and quantitative analytics. The programme emphasises the use of real-world data, including data from governments and international organisations that are available in the public domain.

2. Eligibility

The student should have an undergraduate degree from any recognised university in India or abroad, and basic knowledge of high school-level mathematics.

3. Duration

The duration of this course is two semesters.

4. Selection

There are two channels of selection of candidates for this programme. Admission through the regular channel includes a review of academic transcripts submitted by the applicants. Applicants who qualify for the next stage are required to take a quiz for final selection. Admission through the tuition waived channel consists of an annually conducted written test, a review of academic transcripts, and a personal interview.

Once every year, a limited number of students will be selected for complete tuition fee waiver (in other words, no course fee) through the ISI admission test and interview. The number of seats and the selection process are available on the ISI admission site:

<https://www.isical.ac.in/~admission>

Usually, the application window for the ISI admission test starts towards the end of February and remains open till the end of March. Interested candidates are advised to visit the admission site towards the end of February or beginning of March for details of the admission test and application process for the term starting in August (Fall term). Note that the admission test and interviews, which are not online, are held in India only.

The tuition fee waiver is open to everyone. Foreign nationals are treated as general category candidates (no reservation) in the selection process. However, the ISI admission test is conducted in centres which are located only in India. All candidates applying through this channel will have to appear in the admission test in one of the designated centres in India and subsequently in an interview in ISI Kolkata only. There is no provision of appearing online for the admission test and interview.

¹ Further information on the programme and its delivery is available in the [PGDAS webpage at Coursera](#).

5. Registration and tuition Fees²

The total fee for the course is USD 6,000 or INR 450,000. The payment must be made according to the following schedule.

Item	Amount	Payment due
Registration Fee	USD 267 or INR 20,000	Within 15 days of receiving the admission letter
Semester 1 Term 1 Tuition Fee	USD 733 or INR 55,000	Within 1 week of the beginning of the semester
Semester 1 Term 2 Tuition Fee	USD 1000 or INR 75,000	Three days before the beginning of the term
Semester 1 Term 3 Tuition Fee	USD 1000 or INR 75,000	Three days before the beginning of the term
Semester 2 Term 1 Tuition Fee	USD 1000 or INR 75,000	One week before the commencement of the second semester.
Semester 2 Term 2 Tuition Fee	USD 1000 or INR 75,000	Three days before the beginning of the term
Semester 2 Term 3 Tuition Fee	USD 1000 or INR 75,000	Three days before the beginning of the term
Total	USD 6,000 or INR 450,000 ³	

Payment delays and enrollment

The deadlines for enrolling for a course and making the applicable payment are strict. If a student does not complete the initial payment due in a semester (USD 500 / INR 38,000) by the stipulated date, he/she will not be able to enrol in the corresponding semester.

If a student fails to make any instalment of the payment by the due date, his/her enrollment for the semester will be cancelled. However, the partial payment already made in that semester can be carried forward for enrollment in a subsequent semester.

² A different set of fees and the payment schedule may be applicable for students joining the programme after September 2024.

³ Additional payments at the rate of USD 1,000 or INR 75,000 will be payable for every pair of additional courses taken for credit (for the purpose of second specialization or any other purpose). In the case of Term 1 courses, this payment will be due one week before the commencement of the semester. In the case of Term 2 and Term 3 courses, this payment will be due three days before the commencement of the term.

Cancellation and refund policy

It is expected that, once enrolled, every student will complete the programme as planned initially. If a student contemplates withdrawing from the programme, it is recommended that he/she discuss his/her issues with the *Student Success Advisor* to figure out if the issues can be solved.

If a student still decides to withdraw from the programme before completion, he/she will have to inform the *PGDAS programme Coordinator*, Indian Statistical Institute. After the formal withdrawal, the student's access to the courses will be revoked.

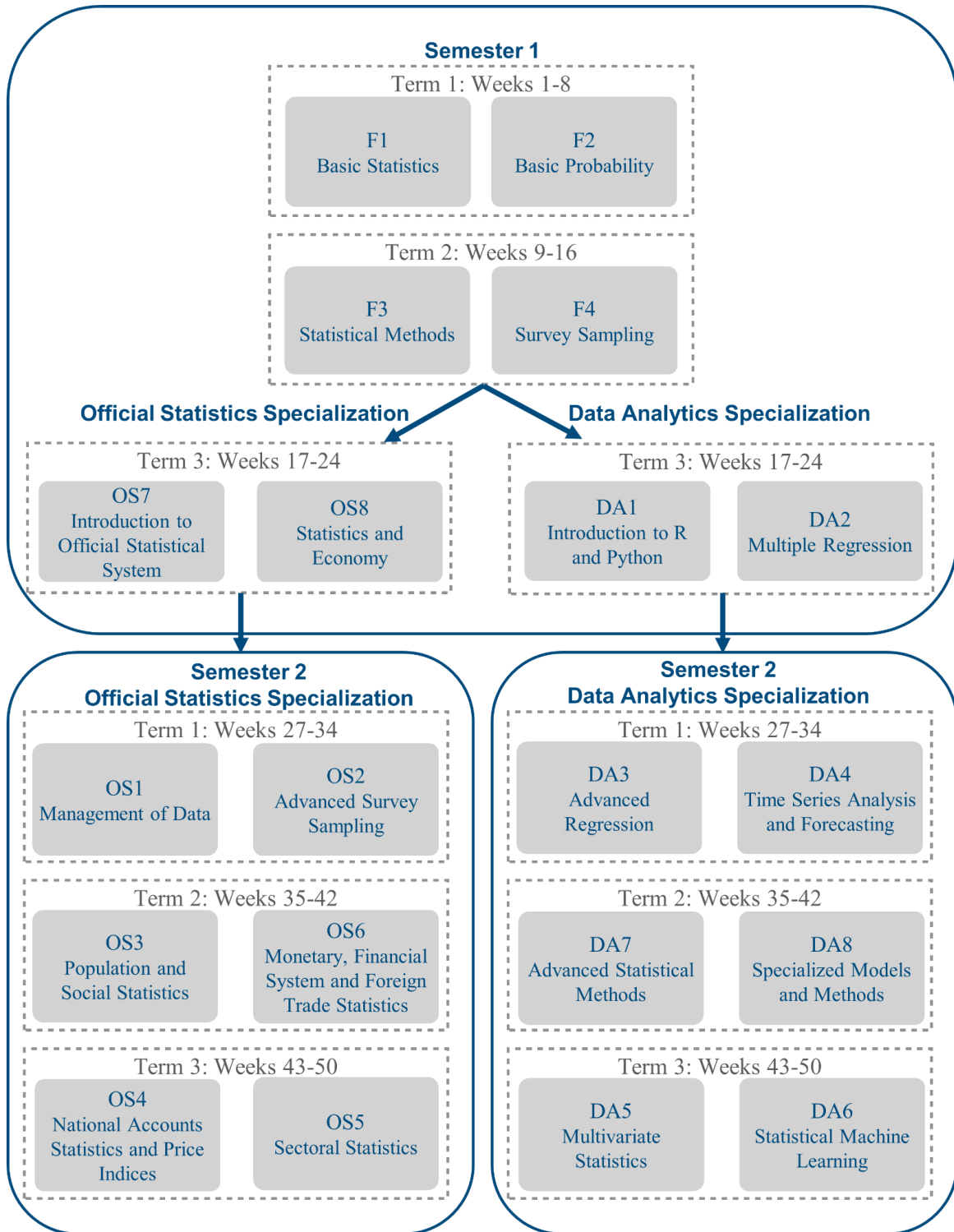
All the fees are non-refundable.

Waiver of tuition fee

There is no partial waiver or financial aid on a case by case basis. One has to be admitted to the programme through the tuition waived channel in order to be eligible for full tuition waiver. However, tuition waived students opting for any one specialization will have to pass all the courses within two semesters of registration. Tuition waived students opting for both the specializations will have to pass all the courses within the minimum possible time as per the regular schedule of courses, without any deferral. A tuition waived student, who fails a course or defers enrollment in a course may continue in the programme as a regular student (without tuition fee waiver) from the first term after not passing a course as per regular schedule. All rules for regular students (without tuition fee waiver) will be applicable to such a student, including tuition fees on a pro rata basis for pairs of courses.

6. Programme Structure

The courses offered in this two-semester online programme are outlined in the following diagram.⁴



⁴ Updated as per a decision taken in the 82nd meeting of the Academic Council held on 3 July 2024.

Each course runs for a term consisting of eight weeks, out of which six weeks are for lectures, quizzes, assignments and interaction sessions. During these weeks, the student can expect a weekly workload of 8-10 hours per week for each course. The last two weeks are for the final examination, back-paper examination (refer to the section on course grading policy to understand the rules of the back-paper), grading and publication of results. The results are expected to be declared in the eighth week.

At any point of time, two courses run in parallel. The sequence of courses within a semester may be altered from the indicative sequence shown in the diagram, while maintaining the requisite order between a course and its prerequisite.

The first two terms of the first semester consists of four compulsory courses. In the third term of the first semester, a student needs to choose one of the specialisation tracks from *Official Statistics* and *Data Analytics*. The chosen track continues in the second semester.

7. Deferral Policy

This programme offers substantial flexibility in meeting the needs of students who wish to take more than one year to complete it for some reason.

One can take up to three years (six semesters) starting from the semester which he/she initially applied for to complete the programme. A student may also choose to take a break after the first semester and enrol for the second semester courses in one of the subsequent semesters.

It is also possible to defer enrollment in the middle of a semester. Every semester consists of six courses, with two running at any point of time. After completing a few of those courses, one may defer the enrollment in the other courses to another semester.

Deferred enrollment will attract no extra fee. The payments to be made are always in advance, before the beginning of a semester, and non-refundable. However the total amount one will have to pay is still the same (USD 6,000 or INR 450,000).

The option of deferral is not available to tuition fee waived students admitted through the ISI test channel.

8. Course Enrollment Policy

- Some courses can be taken only after successful completion of the respective prerequisite courses, as mentioned in the syllabi.
- The student has to complete all four courses of the first two terms of the first semester before enrolling for any course belonging to a specialisation / track.
- The student has to express an intent to choose one of the two tracks three weeks before the end of the second term of the first semester.
- After the completion of the first semester, the student has to enrol for the next semester by paying the appropriate fees in case he/she chooses to continue the programme in a specialisation.

The courses for which the student has enrolled will be added to his/her PGDAS Home Page on Coursera approximately 1 day before the term begins.

9. Grading

a. Grading of individual courses

The grade for each individual course will be given as an absolute score, out of 100. The score will be a weighted sum of the scores obtained in the graded assessments (assignments, quizzes and the final examination).

The graded quizzes and assignments will have a total of 60% weightage and the final examination will have 40% weightage.

For most courses, a major part of the final examination will be staff graded.

To pass a course, one must get a score of at least 50 out of 100.

b. Back-paper examination

If someone does not pass in a course, he/she will have to appear for another examination, henceforth called the *back-paper*, which will be held immediately after the declaration of the results.

- The back-paper examination acts as a replacement for the final examination only.
- Irrespective of the performance in the back-paper examination, the maximum composite score that a student can obtain in a course via the back-paper channel will be capped at 50%.

c. Timeline for quizzes, assignments and examinations

All quizzes, assignments and examinations will have a 24-48 hours window, within which the student can start taking it. Once started, the student will have to finish within a stipulated time as specified in the quiz / assignment / examination.

d. Assignment and examination deadline policy

For all the assignments / quizzes except the final examination, a late submission will attract a penalty of 25% per day beyond the deadline. For the final or back-paper examination, a penalty of 50% per day will be applicable. In other words, if one submits more than 24 hours after the deadline, there will be a 100% penalty.

e. Grade dispute

In case of a grading dispute, the student may request a review of the respective module / assignment / examination and the instructor will review the disputed issue.

10. Conduct

If at any point of time documents provided by a candidate for establishing his/her eligibility are found not to be authentic, he/she will be disqualified. Any diploma provided to a disqualified student will stand revoked.

Submitting an assignment that is not a student's own may result in permanent failure in a course.

11. Final Result

The *composite score* of a student in the programme will be calculated as a simple average of his/her scores in all the courses and this composite score will determine the final grade of the student as per the following table.

Average Composite score range (%)	Grade	Grade point
90-100	A	10
80-89	B	9
70-79	C	8
60-69	D	7
50-59	P	6
Below 50	F	0

12. Award of Certificate

A student passing all the requisite courses for any specialisation is given the *Post Graduate Diploma in Applied Statistics* along with a final mark sheet which includes the list of all courses taken along with the respective composite scores and letter grades. The diploma mentions the specialisation and the overall letter grade in the programme.

A student passing all the requisite courses for both the specialisations is given the *Post Graduate Diploma in Applied Statistics* along with a final mark sheet which includes the list of all courses taken along with the respective composite scores and letter grades. The diploma mentions both the specialisations and the letter grade in the programme.

A student passing all compulsory courses and either of the pairs of the specialisation courses in the first semester but not intending to complete the diploma programme is given a *Post Graduate Certificate in Basic Statistics* along with a final mark sheet which includes the list of all courses taken along with the respective composite scores and letter grades. The certificate mentions the letter grade in the programme.

The diploma is awarded in the Annual Convocation of the Institute following the attainment of qualification. If a student is unable to be present in the Convocation, he/she will receive the degree online.

13. Change of Rules

The Institute reserves the right to make changes in the above rules, programme structure and the curriculum as and when needed. The modified rules will be applicable prospectively to students joining the programme after the rules are modified.

14. Legal Disputes

Any legal dispute regarding grades or otherwise will be dealt under the jurisdiction of the Hon'ble High Court at Calcutta only.

II. Syllabi of Courses

1. Compulsory Courses

F1: Basic Statistics

- Different types of data and ways to load/save/export data using LibreOffice.
- Graphical presentation of data: theory and practice
- Textual representation of data, histogram, contingency tables: theory and practice.
- Measures of central tendency: mean, median, mode etc. Concept of robustness.
- Measures of dispersion: Range, mean absolute deviation, semi-interquartile range and boxplot
- Bivariate data and measures of association. Skewness and kurtosis.

(LibreOffice Calc for all computations)

References

1. *Statistics* (11th edition) by R S Witte and J S Witte
2. *Statistics* (4th edition) by D Freedman, R Pisani and R Purves
3. *How to Lie with Statistics* by D Huff.

F2: Basic Probability

- Randomness, basic definitions, elementary probability based on counting
- Conditional probability, statistical independence; Bayes rule
- Random variables, distributions, expected values, variance and standard deviation, correlation
- Normal approximations for different distributions
- Sampling distribution, sampling error

References

1. *Applied Statistics and Probability for Engineers*. Douglas C. Montgomery and George C. Runger; Wiley (2016).
2. *Introduction to Probability and Statistics for Engineers and Scientists*. Sheldon M. Ross; Academic (2014).
3. *Probability and Simulation*. Giray Otken; Springer (2020).
4. *Simulation*. Sheldon M. Ross; Springer (2013).

F3: Statistical Methods (Prerequisites: F1-F2)

- Concept of population, sample, sampling distributions and estimation of mean, proportion and dispersion.
- Concept of test of statistical hypotheses, one-sample, two-sample and paired t-tests.
- Analysis of variance: the concept and methodology.

- Chi-square tests for goodness of fit and independence.
- Regression and least squares.
- Introduction to time series

(LibreOffice Calc for all computations)

References

1. *Mathematical Statistics with Applications* by K M Ramachandran and C P Tsokos (2009).

F4: Survey Sampling (Prerequisite: F1-F2)

- Terms and concepts in survey sampling
- Simple random sampling with and without replacement
- Systematic sampling
- Stratified random sampling
- Sampling with probability proportional to size
- Ratio and regression estimators

References

1. "A First Course in Survey Sampling" by T. Dalenius in *Handbook of Statistics* Vol. 6 (ed. P.R. Krishnaiah and C.R. Rao), Chapter 2; Wiley (1988).
2. *Practical Sampling Techniques*. Ranjan K. Som; Marcel Dekker (1996).
3. *Sampling Techniques*. William G Cochran; Wiley (1977).

2. Specialisations

a. Official Statistics Specialisation (Prerequisites: F1-F4)

OS1: Management of Data

- Introduction to databases, data abstraction, data integration, data validation, data indexing
- Introduction to database models, ER data model, ER diagram, normalisation, relational algebra
- Basics of SQL programming I: Database creation with SQL, database modification with SQL, database manipulation with SQL
- Basics of SQL programming II: Dealing with nullity and duplicity in SQL, database view in SQL, integrity control with SQL, efficient querying in SQL
- Spatial data and their uses: Introduction to spatial data, Representation of geographic features, Survey Techniques, Storage and presentation.
- Spatial data analytics: Analysis of Remote Sensing data, Analysis of vector data, Advanced analytical techniques for spatial data, Concepts of map making.

References

1. *Database System Concepts, Sixth Edition*. Avi Silberschatz, Henry F. Korth and S. Sudarshan; McGraw-Hill (2013). <http://www.db-book.com>
2. *Database Management Systems, Third Edition*. Raghu Ramakrishnan and Johannes Gehrke; McGraw-Hill (2014). <http://pages.cs.wisc.edu/~dbbook/>
3. *Mining of Massive Datasets*. Jure Leskovec, Anand Rajaraman and Jeff Ullman; Cambridge University Press (2014). <http://www.mmds.org>
4. *Geographic Information Science and Systems, Fourth Edition*. Longley, Goodchild, Maguire, Rhind, 2015, Wiley.
5. *Principles of Geographical Information Systems*. Burrough, McDonnell, Lloyd, 2015, Oxford University Press.
6. *GIS Fundamentals: A First Text on Geographic Information Systems*. Bolstad, 2016, XanEdu.
7. *Introduction to Remote Sensing*. Campbell and Wynne, 2011. Guilford Press.
8. *GIS Technology Applications in Environmental and Earth Sciences*. B. Tian, 2017, Routledge.
9. *Intro to GIS and Spatial Analysis*. Gimond, 2022, ebook. <https://mgimond.github.io/Spatial/>
10. *Introductory Digital Image Processing: A Remote Sensing Perspective, Fourth Edition*. Jensen, 2017, Pearson.
11. *Remote Sensing and Image Interpretation, Seventh Edition*. Lillesand, Kiefer and Chipman, 2015, Wiley.

OS2: Advanced Survey Sampling

- Cluster sampling, multistage sampling
- Planning and designing of large scale sample surveys
- Small area estimation Methods and associated estimators
- Randomised response technique for surveys on sensitive issues
- Adaptive sampling and network sampling
- Some other sampling methods

References

1. *Survey Methodology*. Robert M. Groves, Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau; Wiley (2009).
2. *Designing Household Survey Samples: Practical Guidelines*. United Nations (2005). <https://unstats.un.org/unsd/demographic/sources/surveys/handbook23june05.pdf>
3. *Sampling*. Steven K. Thompson, Wiley (2012)
4. *Survey Sampling*. Leslei Kish; Wiley (1995). [Some Part II chapters]
5. *Sampling Techniques Third Edition*. Wiliam G Cochran; Wiley (1977).
6. *Survey Sampling: Theory and Methods, Second Edition*. Arijit Chaudhuri and Horst Stenger; CRC Press (2005).
7. *Randomized Response Theory and Techniques*. Arijit Chaudhuri and Rahul Mukerjee; Taylor and Francis (1988).

OS3: Population and Social Statistics (Prerequisites: OS7-OS8)

- Population statistics

- Living standards statistics
- Labour statistics
- Health statistics
- Education statistics
- Other social statistics

References

1. *Sample Registration System Year Books, 1971 - 2007*. Office of Registrar General of India, Government of India.
2. *Handbook on Population and Housing Census*. United Nations (2009).
https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1e.pdf
3. *The Follow-up Method in Demographic Sample Surveys*. United Nations Statistical Office (1992).
4. *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*. RaIf Hussmanns, Farhad Mehran and Vijay Verma; International Labour Office (1990).
<https://www.ilo.org/public/english/bureau/stat/download/lfs.pdf>

OS4: National Accounts Statistics and Price Indices (Prerequisites: OS7-OS8)

- National accounts statistics
- Price indices
- Compilation of CPI
- Compilation of PPI
- SDG related indicators

References

1. Accounting for Production: Sources and Methods. Handbook of National Accounting. United Nations (1986). https://unstats.un.org/unsd/publication/SeriesF/SeriesF_39E.pdf
2. Financial Production, Flows and Stocks in the System of National Accounts. Handbook on National Accounting. United Nations (2015).
<https://www.ecb.europa.eu/pub/pdf/other/handbookofnationalaccounting2014en.pdf>
3. Consumer Price Index Manual: Theory and Practice, ILO (2004).
https://www.ilo.org/public/english/bureau/stat/download/cpi/cpi_manual_en.pdf
4. Producer Price Index Manual. IMF (2004).
<https://www.imf.org/external/pubs/ft/ppi/2010/manual/ppi.pdf>

OS5: Sectoral Statistics (Prerequisites: OS7 and OS8)

- Agriculture and allied sector statistics
- Animal husbandry
- Fishery statistics
- Industrial statistics
- Services sector statistics

- Environment statistics

References

1. Food and Agricultural Organisation (FAO): Statistical Yearbooks. FAO, Rome.
<http://www.fao.org/economic/ess/ess-publications/ess-yearbook/en/#.Xbg>
2. 2000 World Census of Agriculture. FAO Statistical Development, Series 12; FAO (2010).
http://www.fao.org/fileadmin/templates/ess/ess_test_folder/World_Census_Agriculture/Publications/Census14_v16.pdf
3. Sampling Methods for Agricultural Surveys. FAO Statistical Development Series 3; FAO (1989)
<http://www.fao.org/3/ca5865en/CA5865EN.pdf>
4. Sample-based Fishery Surveys: A Technical Handbook. FAO (2002).
<http://www.fao.org/3/y2790e/y2790e.pdf>
5. International Recommendations for Industrial Statistics 2008. United Nations (2010).
<https://unstats.un.org/unsd/statcom/doc08/BG-IndustrialStats.pdf>
6. Framework for the Development of Environment Statistics (FDES2013).
<https://unstats.un.org/unsd/environment/fdes>

OS6: Monetary, Financial System and Foreign Trade Statistics (Prerequisites: OS7-OS8)

- Monetary and financial system statistics
- Government finance statistics
- Foreign trade statistics
- Balance of payments statistics
- Uses of financial data for analysis

References

1. Monetary and Financial Statistics Manual and Compilation Guide. Cartas Jose and Harutyunyan Artak, International Monetary Fund; IMF (2017).
<https://www.imf.org//media/Files/Data/Guides/mfsmcg-final.ashx>
2. Government Finance Statistics Manual 2014. International Monetary Fund; IMF (2015).
<https://www.imf.org/external/Pubs/FT/GFS/Manual/2014/gfsfinal.pdf>
3. Government Finance Statistics Guide. European Central Bank (2019).
https://www.ecb.europa.eu/pub/pdf/other/ecb_governmentfinancesstatistics_guide1901.en.pdf
4. Balance of Payments Manual. International Monetary Fund; IMF (2005).
<https://www.imf.org/external/pubs/ft/bopman/bopman.pdf>
5. International Recommendations for Distributive Trade Statistics 2008. United Nations (2009).
<https://unstats.un.org/unsd/trade/M89%20EnglishForWeb.pdf>

OS7: Introduction to Official Statistical System

- An introduction to official statistics in general

- Social statistics
- Economic statistics
- Price statistics
- Environment statistics
- Process of collection, compilation and dissemination of official statistics and their quality

References

1. *Handbook of Statistical Organization: The Operation and Organization of a Statistical Agency* (Third Edition). United Nations (2003).
https://unstats.un.org/unsd/publication/SeriesF/SeriesF_88E.pdf
<https://www.imf.org//media/Files/Data/Guides/mfsmcg-final.ashx>
2. *Statistical System in India 2009*. Government of India, Ministry of Statistics and Programme Implementation.

OS8: Statistics and Economy

- Basic framework of an economy and national income accounting
- Theories of growth, trade cycle and inflation
- Models of income determination
- Theories of money supply
- Determination of price in a competitive market
- Market equilibrium

References

1. *Microeconomic Theory: Basic Principles And Extensions*. Walter Nicholson and Christopher Snyder; S. Chand & Co (2000).
2. *Microeconomics*. P Krugman and R Wells, Worth (2018).
3. *Macroeconomics*. Greg Mankiw; Cengage Learning India (2017).
4. *Macroeconomics: Economic Growth, Fluctuations, and Policy*. Robert E. Hall and David H. Papell; Norton (2005).
5. *National Accounts: A Practical Introduction*. United Nations (2003).
https://unstats.un.org/unsd/publication/SeriesF/seriesF_85.pdf

b. Data Analytics Specialisation (Prerequisites: F1-F4)

DA1: Introduction to R and Python

- Introduction and basics of R
- Basic data analysis using R
- Working with R

- Introduction to Python
- Introduction to Numpy
- Basic text analysis using Python

References

1. *Introductory Statistics with R*, Peter Dalgaard; Springer (2008).
2. *Learning Statistics with R*. Daniel Navarro (2015).
3. *Programming in Python 3: A Complete Introduction to the Python Language (CODE)*. Mark Summerfield; Pearson Addison-Wesley Professional (2010).

DA2: Multiple Regression

- Probabilistic basis of simple and multiple linear regression models
- Simple linear regression
- Multiple regression analysis
- Some tricks of the trade
- Prediction through regression modelling
- Multicollinearity and variable selection in regression

References

1. *Introduction to Linear Regression Analysis*. Douglas C. Montgomery, Elizabeth A. Peck and G. Geoffrey Vining; Wiley (2021).
2. *Practical Regression and Anova using R*; J.J. Faraway (2002).
3. *An R Companion to Applied Regression Analysis*, John Fox and Sanford Weisberg; SAGE (2019).

DA3: Advanced Regression (Prerequisites: DA1 and DA2)

- Detect nonlinearity, heteroscedasticity, serial correlation and non-normality in multiple linear regression through plots and tests
- Deal with above violations of assumptions of the multiple linear regression model
- Understand and make use of casewise diagnostics in regression
- Understand and deal with the issue of multicollinearity, Use penalised linear regression (ridge regression, lasso and elastic net)
- Understand and use Generalised Linear Models including Logistic Regression
- Understand the importance of nonparametric regression. Splines and Kernel smoothing

References

1. *Practical Regression and Anova using R*. J.J. Faraway (2002).
2. *Nonparametric Statistical Methods Using R*. J. Kloeke and J.W. McKean; Chapman and Hall/CRC (2014).
3. *Advanced Regression Models with SAS and R*. Olga Korosteleva; Chapman and Hall/CRC (2020).

DA4: Time Series Analysis and Forecasting (Prerequisite: DA1)

- Descriptive plots, summary and decomposition
- ARMA modelling
- Fitting of stationary and non-stationary time series models
- Forecasting
- Model selection issues
- Advanced models

References

1. *Time Series Analysis and Forecasting by Example*. Soren Bisgaard and Murat KulaHCI; Wiley (2011).
2. *Time Series Analysis and Its Applications: With R Examples*. Robert H. Shumway and David S. Stoffer; Springer (2017).
3. *Introductory Time Series with R*. Andrew V. Metcalfe and Paul S.P. Cowpertwait; Springer (2009).
4. *Basic Data Analysis for Time Series with R*. DeWayne R. Derryberry; Wiley (2014).

DA5: Multivariate Statistics (Prerequisite: DA1)

- Visualisation of multivariate data: problems and solutions.
- Concept of dimension
- Principal Component Analysis
- Factor Analysis
- Multidimensional Scaling
- Correspondence analysis

References

1. *Univariate, Bivariate, and Multivariate Statistics Using R: Quantitative Tools for Data Analysis and Data Science*. Daniel J. Denis; Wiley (2020).
2. *Multivariate Statistics Made Simple: A Practical Approach*. K.V.S. Sarma and R. Vishnu Vardhan; Chapman and Hall/CRC (2019).
3. *Applied Multivariate Statistics with R*. Daniel Zelterman; Springer (2015).

DA6: Statistical Machine Learning (Prerequisite: DA1-DA3)

- Basic ideas of supervised and unsupervised classification, General recipes for classifier construction.
- Classification based on Gaussian models: LDA and QDA. Logistic regression and GLMNET.
- Classification using kernel density estimates and kernel regression, nearest neighbour density estimation, and nearest neighbour classification.

- Classification tree and random forest, Linear Support Vector Machines (SVMs): separable and non-separable case, Nonlinear SVM using the kernel trick
- Neural network, single-layer and multi-layer perceptrons, backpropagation algorithm, brief ideas of deep learning, Bagging, boosting, and other ensemble methods
- A brief introduction to clustering. Different types of similarity and dissimilarity measures, Hierarchical methods for clustering, k -means clustering algorithm

References

1. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. Andreas C. Müller and Sarah Guido; O'Reilly (2016).
2. *The Elements of Statistical Learning*. Jerome H. Friedman, Robert Tibshirani and Trevor Hastie; Springer (2017).
3. *An Introduction to Statistical Learning: With Applications in R*. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani; Springer (2017).
4. *Mathematical Statistics and Data Analysis*. John A. Rice; Brooks/Cole (2006).
5. *Data Analysis and Applications 1: Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*. James R. Bozeman and Christos H. Skiadas (eds); Wiley (2019).

DA7: Advanced Statistical Methods

- Good estimation: Bias, variance, trade off, consistency, efficiency, robustness (concepts only)
- General methods of model fitting/estimation: Method of Moments, Maximum Likelihood Estimation, some common examples, Assessing methods through simulations
- Model fitting: Common distributions, mixture and compound distributions, QQ plot, Goodness of Fit chi-square, Kolmogorov-Smirnov and other tests
- Nonparametric inference: Histogram, Kernel density estimator, simulation from KDE, important nonparametric tests: 2 sample, paired, independence
- Basics of Bayesian statistics: Bayesian vs classical paradigm, Prior and posterior distribution, Credible intervals, Model selection using Bayes factor
- Estimation: Conjugate prior, Metropolis Hastings algorithm, MCMC, R demonstrations

References

1. *All of Statistics: A Concise Course in Statistical Inference*. Larry Wasserman; Springer (2013).
2. *Statistics, 4th Edition*. David Freedman, Robert Pisani and Roger Purves; W.W. Norton & Co (1998).
3. *The Art of Statistics. How to Learn from Data*. David Spiegelhalter; Basic Books (2021).
4. *Bayesian Essentials with R*. J-M. Marin and C.P. Robert; Springer (2013).
5. *Bayesian Data Analysis*. A. Gelman, J.B. Karlin, Hal S. Stern, D.B. Dunson, A. Vehtari and D.B. Rubin; Chapman and Hall/CRC (2013).

DA8: Specialised Models and Methods

- Bivariate and multivariate data: Modeling through conditioning, modelling through Copula, multivariate normal distribution
- Time to event data: Lifetime distributions, censored data likelihood, life tables, KM, two sample test, Cox regression (mention only)
- Categorical data: Visualisation of categorical data, measures of association, log-linear models
- Multiple event occurrence data: Examples, Markov chain (including estimation), Poisson Process (HPP and NHPP), Compound Poisson process
- Text data: Parsing, text mining, sentiment analysis, Natural Language Processing
- Image data: Representations, feature extraction, statistical models, storage and retrieval, basic operations

References

1. *Elements of Copula Modeling with R*. Marius Hofert, Ivan Kojadinovic, Martin Mächler, Jun Yan; Springer (2018).
2. *Applied Survival Analysis Using R*. Dirk F. Moore; Springer (2016).
3. *Analysis of Categorical Data with R*. Christopher R. Bilder, Thomas M. Loughin; CRC Press (2015).
4. *Introduction to Stochastic Processes*. Paul G. Hoel, Sidney C. Port and Charles J. Stone; Houghton Mifflin (1972).
5. *Mining Text Data*, Charu C. Aggarwal and Cheng Xiang Zhai; Springer (2012).
6. *Natural Language Processing with Python—Analyzing Text with the Natural Language Toolkit*, Steven Bird, Ewan Klein, Edward Loper, O'Reilly; (2009).
7. *Digital Image Processing*, Rafael Gonzalez and Richard Woods; Pearson (2017).