Post Graduate Diploma in Applied Statistics (PGDAS) - Digital Mode

I: Scope

Official statistics play a key role in evidence based policy making and policy research in all countries. A huge number of individuals are involved in the processes of generation, interpretation and management of such data and hence manpower training and skill building with respect to each of these exercises is of utmost importance. The contents and methodologies change and evolve as a response to the latest requirements and technology available. Dissemination of knowledge and applications of state of the art statistics in this domain to the producers and users of official data, many of whom are without a formal background in statistics, is a challenge. This course wishes to fill this gap, widening the accessibility and flexibility of convenience, by offering it in digital mode.

It is proposed that this course be developed by ISI and delivered through the e-platform of Coursera (a leading online course provider, which approached ISI soliciting online courses that ISI can develop). There will be a provision of offering the course without fee to a limited number of candidates who qualify through an admission test administered by ISI.

II: Broad Structure

While preparing the syllabus, it is kept in mind that the content of the course is useful to the users of official statistics as well as to those who are involved in the generation of official statistics. There are two levels of this course. The first level is Basic and the second level is Advanced. The advanced level would have two specializations, namely, Official Statistics and Data Analytics, catering to the needs of producers and users, respectively, of official statistics. A certificate will be issued after successful completion of the Basic Level and this will be converted into a diploma after successful completion of the Advanced Level.

III. Duration

The duration of each of the Basic and Advanced Levels will be one semester.

IV. Eligibility

A. Eligibility for Basic Level

1) *Channel with no course fee* (through admission test of ISI)
    a) Graduate Degree in any subject (BA/ BSc/ BCom/ BCA/ BTech/ BE).
    b) Mathematics at 10+2 Level.
    c) Indian Nationality.
    d) Certification of candidate's necessity to deal with official data (by any Central/State Govt department or Central/State Govt recognized Indian higher educational or research institution).
2) *Channel with course fee*

a) Graduate Degree in any subject (BA/ BSc/ BCom/ BCA/ BTech/ BE).
b) Successful completion of an online course on Mathematics (at a level equivalent to 10+2) offered by COURSERA.

B. Eligibility for Advanced Level

1) *Channel with no course fee*: Successful completion of the Basic Level through the channel with no course fee.
2) *Channel with course fee*: Successful completion of the Basic Level through the channel with course fee.

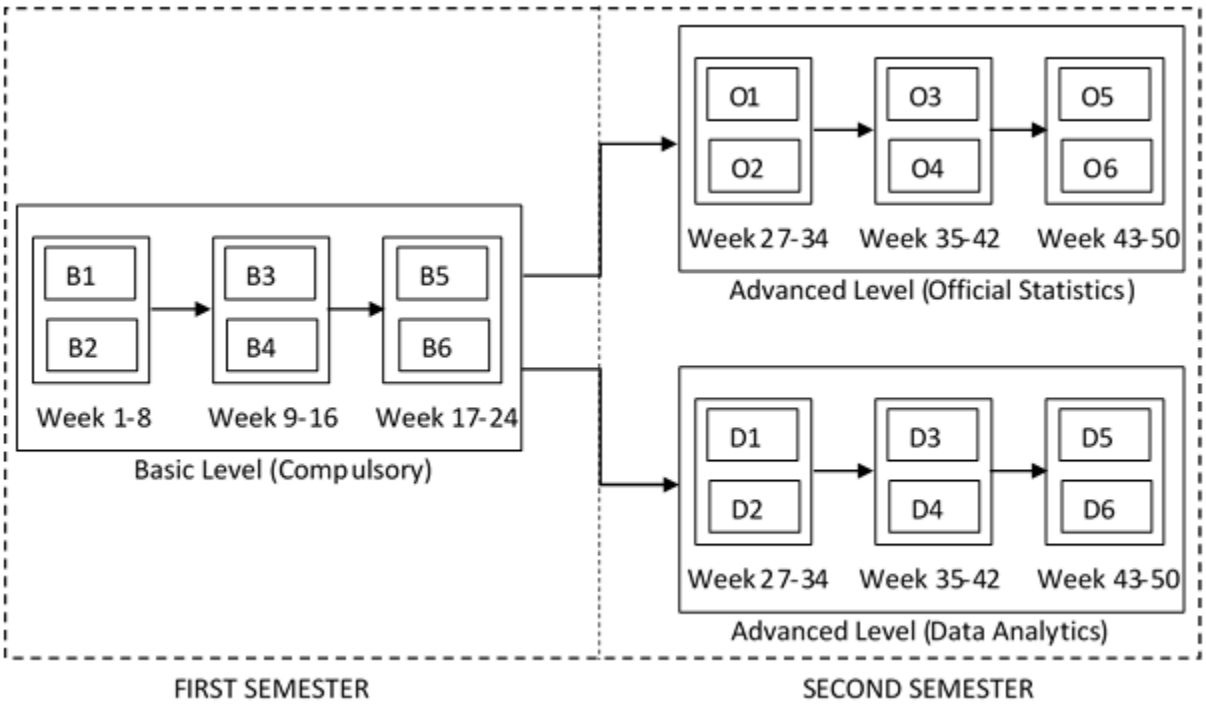
V. Admission

Admission to the Basic Level through the channel with no course fee will be through a merit list formed on the basis of performance in an annual admission test, which would coincide with the 10+2 level Mathematics test currently used for admission to the PGDSMA programme. The reservation policy of the Institute regarding admission will apply.

All other admissions (including admission to the Basic Level through the channel with course fee and to the Advanced Level through either channel) will be on a first-come-first served basis from the pool of eligible applicants.

VI. Detailed Structure

At each level there will be six subjects, with two subjects running in parallel for eight weeks, as shown in the diagram.



VII. Modality of Teaching

For each subject there will be approximately 15 hours of pre-recorded video, divided into six modules of 2.5 hours and further sub-modules of shorter duration. Each module is meant for one week, and would have accompanying notes and quiz/assignments accounting for about 4 hours of additional engagement on the part of the students. There will be a weekly interaction of one hour in real time for addressing queries and/or working out examples related to each weekly module of each subject, and a blog for exchange of additional queries and responses. The content of the weekly modules in a subject will be shared with the registered students sequentially, each after completion of the previous one (including submission of quiz/assignment). The technical knowhow of COURSERA will be utilized in developing the content for smooth running.

At any level, two subjects will run simultaneously for 6 weeks followed by 2 weeks for self-study and examination. A further two weeks will be allocated for a one-time supplementary examination at the end of a semester. Thus, each level will span approximately 26 weeks.

VIII. Modality of Evaluation

Students would be assessed periodically on the basis of an online MCQ type quiz or a programming assignment for each module, according to a weekly schedule. This will be graded automatically. Final examinations will also be held online. The facilities used in online courses of other internationally reputed universities for maintaining integrity of the evaluation, will be used. The weights for the two tests (periodic assessment and final examination) can be fixed at 40% and 60%. Any candidate will be declared “passed” if the aggregate grade in each subject is at least 50%. The overall score in the course will be the simple average of scores in all the subjects.

A student can make a second attempt to improve an assessment score in a module. The second attempt must be made 24-72 hours after the first attempt. A supplementary examination will be allowed as a second attempt of the final examination with a maximum allowable score of 50%. A student will be allowed to take the supplementary examination in at most two subjects in a given semester.

IX. Capacity

At each level, there will be up to 30 students admitted through the channel with no course fee. The number of students admitted through the channel with course fee, at each level, will be decided on the basis of available infrastructure from time to time.

X. Course Content and Syllabus

Basic Level: The six subjects to be offered in this level are:

1. **BASIC 1: Basic Statistics**

   Data types, Presentation of Data/ Measures of Central tendency, Measures of Dispersion, Moments, Measures of Association, Charts (Excel for all computations)

   References:

   (i) *Statistics*. David Freedman, Robert Pisani and Roger Purves; Norton (2007).

   (ii) *Probability and Statistics by Example: Vol 1 (Basic Probability and Statistics)*. Suhov Y. and M. Kelbert; Cambridge (2014).

   (iii) *Basic Statistics: Understanding Conventional Methods and Modern Insights*. Rand R. Wilcox; Oxford (2009).

   (iv) *Excel for Social Science Statistics - A Guide to Solving Practical Problems*. T. J. Quirk; Springer (2013).

2. **BASIC 2: Basic Probability**

   Basic definitions and results, Bayes rule, Random variables, distributions and expected values, Sampling distribution, sampling error, Simulation (Excel for all computation)

   References:

   (i) *Applied Statistics and Probability for Engineers*. Douglas C. Montgomery and George C. Runger; Wiley (2016).

   (ii) *Introduction to Probability and Statistics for Engineers and Scientists*. Sheldon M. Ross; Academic (2014).

   (iii) *Probability and Simulation*. Giray Otken; Springer (2020).

   (iv) *Simulation*. Sheldon M. Ross; Springer (2013).

3 **BASIC 3: Statistical Methods**

   Methods of estimation and testing, One and two sample t-test, Paired t test, ANOVA, Contingency tables, Regression, Times series methods (Excel for all computations)

   References:

   (i) *Introductory Statistics*. Sheldon M. Ross; Academic (2017).

(ii) *Probability and Statistics for Engineers and Scientists*. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye; Pearson (2006).

(iii) *Applied Business Statistics: Methods and Excel-Based Applications*. Trevor Wegner; Juta Academic (2012).

(iv) *Mathematical Statistics and Data Analysis*. John A. Rice; Cengage (2013).

4 BASIC 4: Census and Sample Surveys

Difference between census and sample survey, Effect of finite population, Inclusion probability, Non-sampling error/ SRSWR &WOR, Stratified sampling, PPSWOR & WR/ Cluster & Systematic Sampling, HTE

References:

(i) A First Course in Survey Sampling. T. Dalenius in *Handbook of Statistics* Vol. 6 (ed. P.R. Krishnaiah and C.R. Rao), Chapter 2; Wiley (1988).

(ii) *Practical Sampling Techniques*. Ranjan K. Som; Marcel Dekker (1996).

(iii) *Sampling Techniques*. Wiliam G Cochran; Wiley (1977).

5 BASIC 5: Introduction to Official Statistical System

Definition, Classification and Sources of Official Statistics; Statistical system and organisational setup; Sources of Social Statistics and data from Administrative sources/ Social Statistics: Important Censuses and the related concepts and definitions; Social Statistics: Important Surveys and their scope & coverage and general and specific features of survey design/ Economic Statistics: Based on administrative sources, Important Censuses and surveys/ Price Statistics; and Price & Production Indices.

References:

(i) *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

(ii) *Statistical System in India 2009*. Government of India, Ministry of Statistics and Programme Implementation.

6 BASIC 6: Statistics and Economy

Microeconomics - Basic Consumer and Producer behaviour models, Market models and concepts of equilibrium

Macroeconomics - Basic concepts and Definitions. Simple identities.

Basic Concepts and Definitions of National Accounts, Introduction to SNA Main Identities

References:

(i) *Microeconomic Theory: Basic Principles And Extensions*. Walter Nicholson and Christopher Snyder; S. Chand & Co (2000).

(ii) *Microeconomics*. P Krugman and R Wells, Worth (2018).

(iii) *Macroeconomics*. Greg Mankiw; Cengage Learning India (2017).

(iv) *Macroeconomics: Economic Growth, Fluctuations, and Policy*. Robert E. Hall and David H. Papell; Norton (2005).

(v) *National Accounts: A Practical Introduction*. United Nations (2003). https://unstats.un.org/unsd/publication/SeriesF/seriesF_85.pdf

There will be two alternative advanced level specializations, namely, in **Official Statistics** and in **Data Analytics.** There are six courses in each specialization.

Advanced Level A: Official Statistics Specialization

1 OFF STATISTICS 1: Data Storage and retrieval

Concepts of Data Integration, Introduction to Database Management System, Introduction to SQ, Methods of data validation.

References:

(i) *Database System Concepts, Sixth Edition*. Avi Silberschatz, Henry F. Korth and S. Sudarshan; McGraw-Hill (2013). http://www.db-book.com

(ii) *Database Management Systems, Third Edition*. Raghu Ramakrishnan and Johannes Gehrke; McGraw-Hill (2014). http://pages.cs.wisc.edu/~dbbook/

(iii) *Mining of Massive Datasets*. Jure Leskovec, Anand Rajaraman and Jeff Ullman; Cambridge University Press (2014). http://www.mmds.org

2 OFF STATISTICS 2: Survey design and concepts

Questionnaire, Pilot, Sampling Frame, Stratification, Selection Bias, Target and sampled population, perception survey, randomised response, network and adaptive sampling,quota sampling, introduction to planning and designing large scale sample surveys.

References:

(i) *Survey Methodology*. Robert M. Groves, Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer and Roger Tourangeau; Wiley (2009).

(ii) *Designing Household Survey Samples: Practical Guidelines*. United Nations (2005). https://unstats.un.org/unsd/demographic/sources/surveys/handbook23june05.pdf

(iii) *Sampling*. Steven K. Thompson , Wiley (2012)

(iv) *Survey Sampling*. Leslei Kish; Wiley (1995).

(v) *Sampling Techniques Third Edition*. Wiliam G Cochran; Wiley (1977).

(vi) *Survey Sampling: Theory and Methods, Second Edition*. Arijit Chaudhuri and Horst Stenger; CRC Press (2005).

(vii) *Randomized Response Theory and Techniques*. Arijit Chaudhuri and Rahul Mukerjee; Taylor and Francis (1988).

3 OFF STATISTICS 3: Population and Social Statistics

Vital Statistics, Consumer Expenditure Survey, Labour and Employment Statistics, Social Consumption, Health and Education Survey and Time-use survey.

References:

(i) *Sample Registration System Year Books, 1971 - 2007*. Office of Registrar General of India, Government of India.

(ii) *Handbook on Population and Housing Census*. United Nations (2009). https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1e.pdf

(iii) *The Follow-up Method in Demographic Sample Surveys*. United Nations Statistical Office (1992).

(iv) *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*. Ralf Hussmanns, Farhad Mehran and Vijay Verma; International Labour Office (1990). https://www.ilo.org/public/english/bureau/stat/download/lfs.pdf

4 OFF STATISTICS 4: Economic Statistics I

National Accounts Statistics - including compilation of Production-side, Income-side and Use-side estimates, Sectoral sequence of accounts; and Price Indices & Compilation of CPI - including Concept of simple & composite index, Consumption basket, Assigning weights, Reference periods and their implications, Index number formulas, Compilation stages and Selected specific issues of compiling CPI.

References:

(i) *Accounting for Production: Sources and Methods*. Handbook of National Accounting. United Nations (1986). https://unstats.un.org/unsd/publication/SeriesF/SeriesF_39E.pdf

(ii) *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

(ii) *Consumer Price Index Manual: Theory and Practice*, ILO (2004). https://www.ilo.org/public/english/bureau/stat/download/cpi/cpi_manual_en.pdf

(iii) *Producer Price Index Manua*l. IMF (2004). https://www.imf.org/external/pubs/ft/ppi/2010/manual/ppi.pdf

5 OFF STATISTICS 5: Economic Statistics II

Agriculture and Allied Sector Statistics - including Area and Yield estimation, Surveys for estimating output of livestock products, Surveys for estimating Marine fish catch and Inland fish production; Industrial Statistics, including - ASI and Unorganised sector surveys; Service Sector Statistics and other sectoral statistics.

References:

(i) *Food and Agricultural Organisation (FAO): Statistical Yearbooks*. FAO, Rome. http://www.fao.org/economic/ess/ess-publications/ess-yearbook/en/#.XbgCz9xhmM8

(ii) *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

(iii) *Sampling Methods for Agricultural Surveys.* FAO Statistical Development Series 3; FAO (1989)

http://www.fao.org/3/ca5865en/CA5865EN.pdf

(iv) *Sample-based Fishery Surveys: A Technical Handbook*. FAO (2002). http://www.fao.org/3/y2790e/y2790e.pdf

(v) *International Recommendations for Industrial Statistics 2008*. United Nations (2010). https://unstats.un.org/unsd/statcom/doc08/BG-IndustrialStats.pdf

(vi) *International Recommendations for Distributive Trade Statistics 2008*. United Nations (2009). https://unstats.un.org/unsd/trade/M89%20EnglishForWeb.pdf

6 OFF STATISTICS 6: Economic Statistics III

Government Financial Statistics (GFS), Banking and Financial Statistics (BFS), Foreign Trade Statistics and Balance of Payment (BoP) Statistics.

References:

(i) *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

(ii) *Government Finance Statistics Manual 2014.* International Monetary Fund; IMF (2015). https://www.imf.org/external/Pubs/FT/GFS/Manual/2014/gfsfinal.pdf

(iii) *Government Finance Statistics Guide*. European Central Bank (2019). https://www.ecb.europa.eu/pub/pdf/other/ecb.governmentfinancestatisticsguide1901.en.pdf

(iv) *Balance of Payments Manual*. International Monetary Fund; IMF (2005). https://www.imf.org/external/pubs/ft/bopman/bopman.pdf

Advanced Level B: Data Analytics Specialization

1 DATA ANALYTICS 1: Introduction to R and Python

Introduction to R: Objects, workspace, packages, operators and functions, user defined functions, sequences, vector and matrix computations,graphics, logical variables, conditional statements, loops and alternatives

Introduction to Python: Data types, data handling, printing, scanning, File handling, Loops and if-else conditioning, / libraries, text analysis, graphs, data analysis, data classification, sentiment analysis, functions, matrix operations, class and object

References:

(i) *Introductory Statistics with R*, Peter Dalgaard; Springer (2008).

(ii) *Learning Statistics with R*. Daniel Navarro (2015).

(ii) *Programming in Python 3: A Complete Introduction to the Python Language (CODE)*. Mark Summerfield; Pearson Addison-Wesley Professional (2010).

2 DATA ANALYTICS 2: Multiple Regression with R

Method and Interpretation, Multiple correlation and R-square,Prediction, Estimation versus prediction errors, Subset selection, Residual and leverage, Outliers.

References:

(i) *Regression Analysis by Example*. Samprit Chatterjee and Ali S. Hadi; Wiley (2013).

(ii) *Practical Regression and Anova using R*; J.J. Faraway (2002).

(iii) *An R Companion to Applied Regression Analysis*, John Fox and Sanford Weisberg; SAGE (2019).

3 DATA ANALYTICS 3: Advanced Regression with R:

Handling nonlinearity, heteroscedasticity, serial correlation and nonnormality in multiple linear regressions, Casewise diagnostics, GLM, Logistic regression, Kernel and Spline, Ridge regression and Lasso.

References:

(i) *Practical Regression and Anova using R*. J.J. Faraway (2002).

(ii) *Nonparametric Statistical Methods Using R*. J. Kloke and J.W. McKean; Chapman and Hall/CRC (2014).

(iii) *Advanced Regression Models with SAS and R*. Olga Korosteleva; Chapman and Hall/CRC (2020).

4 DATA ANALYTICS 4: Time Series Analysis and Forecasting with R

Trend, seasonality, stationarity, smoothing and differencing, ACF and PACF, SARIMA models, forecasting, ARCH/GARCH models, multivariate time series, VAR models

References:

(i) *Time Series Analysis and Forecasting by Example*. Soren Bisgaard and Murat Kulahci; Wiley (2011).

(ii) *Time Series Analysis and Its Applications: With R Examples*. Robert H. Shumway and David S. Stoffer; Springer (2017).

(iii) *Introductory Time Series with R*. Andrew V. Metcalfe and Paul S.P. Cowpertwait; Springer (2009).

(iv) *Basic Data Analysis for Time Series with R*. DeWayne R. Derryberry; Wiley (2014).

5 DATA ANALYTICS 5: Multivariate Statistical Methods with R

Visual representation of multivariate data, Principal Component Analysis,Factor Analysis, Multidimensional Scaling, Correspondence Analysis

References:

(i) *Univariate, Bivariate, and Multivariate Statistics Using R: Quantitative Tools for Data Analysis and Data Science*. Daniel J. Denis; Wiley (2020).

(ii) *Multivariate Statistics Made Simple: A Practical Approach*. K.V.S. Sarma and R. Vishnu Vardhan; Chapman and Hall/CRC (2019).

(iii) *Applied Multivariate Statistics with R*. Daniel Zelterman; Springer (2015).

6 DATA ANALYTICS 6: Introduction to Statistical Learning

Basic ideas of training and test data, validation, Discriminant analysis, Classification,Tree based methods, Clustering, SVM, Neural Networks

References:

(i) *Introduction to Machine Learning with Python: A Guide for Data Scientists.* Andreas C. Müller and Sarah Guido; O'Reilly (2016).

(ii) *The Elements of Statistical Learning*. Jerome H. Friedman, Robert Tibshirani and Trevor Hastie; Springer (2017).

(iii) *An Introduction to Statistical Learning: With Applications in R*. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani; Springer (2017).

(iv) *Mathematical Statistics and Data Analysis*. John A. Rice; Brooks/Cole (2006).

(v) *Data Analysis and Applications 1: Clustering and Regression, Modeling-estimating, Forecasting and Data Mining*. James R. Bozeman and Christos H. Skiadas (eds); Wiley (2019).

Anup Dewanji, Chair

Amit Biswas

Mausumi Bose

Nachiketa Chattopadhyay, Convener

Isha Dewan