

SHYAMA PRASED MUKHERJI COLLEGE FOR WOMEN

TEACHING PLAN AUGUST 2022 TO DECEMBER 2022

COURSE AND YEAR: **B.A. (HONS.) II YEAR**

SEMESTER: **III**

TAUGHT INDIVIDUALLY OR SHARED: **SHARED**

PAPER: **DATA ANALYSIS**

FACULTY: **ITI TOMAR & RICHIKA RANA**

NUMBER of classes (per week): **4L**

Teaching plan

Unit 1

Introduction to the course: How can the representation and analysis of data help us study real world problems. Publicly available data sets.

Readings prescribed:

Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for Managers using Microsoft Excel*. (8th ed.). Pearson. pp. 25-35.

Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for Managers using Microsoft Excel*. (8th ed.). Pearson. Ch. 1, sections 1.1-1.3

Devore, J. L. (2012). *Probability and Statistics for Engineering and the Sciences*. (8th ed.). Cengage Learning. Ch. 1, pp. 32-33 (trimmed mean)

Data base of Indian Economy, World bank data set

RBI (n.d.). Retrieved from <https://dbie.rbi.org.in/DBIE/dbie.rbi?site=home>

World Bank (n.d.). Retrieved from <https://data.worldbank.org/>

On data and its representation and a large data set

https://www.ted.com/talks/hans_rosling_the_best_stats_you_ve_ever_seen (Gapminder.org)

Tattar, P., Ramaiah, S., & Manjunath, B. (2018). *A Course in Statistics with R*. Wiley. Ch. 1 pp. 6-7

Number of classes required: 8 LECTURES

Unit 2

Using Data: Available statistical software, steps in data storage, organisation and cleaning.

Readings prescribed:

Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for Managers using Microsoft Excel*. (8th ed.). Pearson. Ch. 1, Section 1.4 onwards; Ch. 2, Sections 2.1-2.2

Gardener, M. (2012). *Beginning R The Statistical Programming Language*. Wiley. Ch. 1, pp. 1-24; Ch. 2, pp. 25-52

Tattar, P., Ramaiah, S., & Manjunath, B. (2018). *A Course in Statistics with R*. Wiley. Ch. 2 pp. 15-18, pp. 41-46

Number of classes required: 14 LECTURES

Unit 3

Visualisation and Representation: Alternative forms of presenting summarising and presenting data.

Readings prescribed:

Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for Managers using Microsoft Excel*. (8th ed.). Pearson. Ch. 2, Section 2.3 onwards; Ch. 3, Ch. 4

Tattar, P., Ramaiah, S., & Manjunath, B. (2018). *A Course in Statistics with R*. Wiley. Ch. 5 pp. 105-109

Gardener, M. (2012). *Beginning R The Statistical Programming Language*. Wiley. Ch. 5, pp. 154-158, Ch. 7, pp. 215-217

Number of classes required: 21 LECTURES

Unit 4

Simple estimation techniques and tests for statistical inference

Readings prescribed:

Levine, D., Stephan, D. & Szabat, K. (2017). *Statistics for Managers using Microsoft Excel*. (8th ed.). Pearson. Ch 5 (pp. 199, 203-204); Ch 6 (pp. 225, 228-229); Ch. 7; Ch. 8, Sections 8.1-8.4, pp. 292-293; Ch. 9; Ch. 10, Sections 10.1, 10.4, pp. 361- Summary onwards, Relevant parts of Excel guide

Gardener, M. (2012). *Beginning R The Statistical Programming Language*. Wiley. Ch. 6, pp. 181-183

Number of classes required: 21 LECTURES

Course Learning Outcomes: The course will use data simulations and publicly available data sources to help students learn about data types, their organization and visual representation. They will learn how to compute summary statistics and do some basic statistical inference.

Methodology of Teaching: Interactive online Lectures and practical classes with actual big data sets, discussion on collection, visualization and making inferences about the real-world

problem. Students are familiarized to use Microsoft Office Excel and R software to analyse data.

Additional sources:

1. Gupta, S.C. (2019). Fundamentals of Statistics, 5th edition, Himalaya Publishing House
2. Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc. ".
3. <https://www.rstudio.com/resources/cheatsheets/>
4. Jaynal Abedin and Kishor Kumar Das: Data Manipulation with R, 2nd Edition (2015), PACKT Publishing
5. Winston W, (2019). Microsoft Excel 2019 Data Analysis and Business Modeling. Microsoft Press
6. Mark P. J. van der Loo and Edwin de Jonge: Data Validation Infrastructure for R, Journal of Statistical Software, Vol. 97, No. 10.
[https://cran.rproject.org/web/packages/validate/vignettes/JSS_3483.pdf]
7. https://stats.oecd.org/Index.aspx?DatasetCode=SNA_TABLE1

Utility of additional sources: Additional resources will provide better understanding of data, sampling techniques, data types, their organization and visual representation. The books can be referred to learn more about computing summary statistics and statistical inference. Also, these provide a relatively gentle yet informative exposure to the statistical software environment, alongside some common statistical analyses, so that students have a solid foundation.

Assessment: A hands-on project for the project the students are expected to use secondary sources of data available in public domain (eg. Indian economy data, World Bank data etc.) and analyse it using at least one of the software taught (Excel and R) preferably both(though for different parts of the project).

Criteria of Assessment: The students are assessed on comprehension and clarity of concepts based on learning and application.

We have adhered to the above-mentioned teaching plan and have completed the course on time.



ITI TOMAR



RICHIKA RANA