

**Course Name – Statistics With R**  
**Course Code – STAR 3072**  
**Credits- 4**  
**Cross Elective**

**Faculty Name - Avanindra Nath Thakur**

**Faculty Biography** – Avanindra Thakur is a Professor of Economics trained as a development economist with more than 13 years of teaching experience. He did his MA, MPhil, and PhD from JNU. He has also completed a full 10-course specialisation in Data Science with R from Johns Hopkins University and a specialisation in statistics with R from Duke University.

**Email- [anthakur@jgu.edu.in](mailto:anthakur@jgu.edu.in)**

**Office Hours- TBD**

**Course Description-** This course is designed to equip students with the essential concepts and practical skills needed to conduct independent empirical research. A key focus of the course is to help students develop the ability to conduct preliminary data analysis using **R programming**, a powerful open-source statistical software.

Throughout the course, students will be introduced to core statistical concepts, including both descriptive and inferential statistics, and will learn how to apply these methods to different types of data using R. The emphasis is not just on understanding theory, but on applying it to real-world datasets.

Students will gain hands-on experience in using a variety of inferential statistical techniques. They will also learn to perform various statistical tests, understand their underlying assumptions, and identify potential violations that may compromise the validity of their analysis.

Another major component of the course is correlation and regression analysis, in which students will work with real datasets to explore relationships among variables, draw meaningful conclusions, and understand and evaluate relevant policies.

In addition, the course places strong emphasis on data visualisation. Students will learn how to create clear, effective, and advanced visualisations using R, enabling them to communicate their findings more effectively.

The course also highlights the importance of data exploration—helping students identify patterns, relationships, and outliers, as well as other critical insights that inform the data analysis process.

Finally, students will be introduced to working with spatial data, learning how to create maps and spatial visualisations in a clear and effective manner.

Overall, this course provides a comprehensive foundation for students who wish to develop strong analytical and research skills using modern data tools.

**Prerequisites:** The course requires a basic understanding of statistics, preferably from an introductory statistics course. This course is application-based, and it is better to have some knowledge of basic statistics to focus more on the application of statistics with R programming. No prior knowledge of R programming is required for this course, as it starts with a lecture on R programming from scratch.

### Learning Objective

Course Intended Learning Outcomes	Teaching and Learning Activities	Assessments/ Activities
1. Clean, analyse, and interpret data using R (Exploring Distributions)	Classroom lectures and data-driven exercises	Solve questions based on a dataset. In-class Exam. It will be assigned within the first three weeks.
2. Apply the measures of probability and simulate probability distributions	Classroom lectures and data-driven exercises simulation experiments for various distributions	In-class exam
3. Check assumptions, apply tools of regression and data analysis	Classroom lectures and data-driven exercises	In-class exam
4. Demonstrate different inferential statistics techniques on various datasets and communicate the results using advanced methods of visualising data.	Classroom lectures and data-driven exercises	Project and Presentation

**Scheme of Evaluation and Grading** – Since this course is primarily application-based, it will have continuous assessment. The assessment in this course will have three broad components: the first two assessments will each carry 30 per cent, and students will be asked to use basic

statistical concepts through R programming. The third and final assessment will involve a group project using secondary data, applying the concepts discussed throughout the period to establish results and explain them using a literature review. In other words, the last project will be a group project that addresses a research question using statistical concepts in R.

### Course Materials and Readings

- Agresti, A., Franklin, C. A., & Klingenberg, B. (2021). *Statistics: The art and science of learning from data* (5th ed.). Pearson.
- Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten* (2nd ed.). Analytics Press.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. SAGE Publications.
- Long, J. D., & Teetor, P. (2019). *R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics* (2nd ed.). O'Reilly Media.
- Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science* (2nd ed.). O'Reilly Media

### Session Plan

Session ( Weekly)	Broad Topic	Readings
Week 1	Setting up R, learning basic commands, getting started with statistical analysis	Chapters 1, 2, and 3: Agresti, Alan, and Christine Franklin. <i>Statistics: The Art and Science of Learning from Data</i> . 3rd ed. Pearson, 2013, Chapter 1-2, R Cookbook.
Week 2	Understanding Data through R, Descriptive analysis, data cleaning and importing data in R	Chapter 2-5 -Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). <i>R for data science</i> (2nd ed.). O'Reilly Media
Weeks 3	Making tables and graphs in R and doing descriptive statistical analysis	Chapters 12, 13: Few, Stephen. <i>Show Me the Numbers: Designing Tables and Graphs to Enlighten</i> . Analytic Press, 2012. Chapter 4: Agresti, Alan, and Christine Franklin. <i>Statistics: The Art and Science of Learning from Data</i> . 3rd ed., Pearson, 2013. Chapter 1 -Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). <i>R for data science</i> (2nd ed.).

		O'Reilly Media
Week 4	First Assessment/Quiz	
Week 5-8	<ul style="list-style-type: none"> <li>• Difference between sample population, sampling methods,</li> <li>• The weak law of large numbers, the central limit theorem, confidence intervals, and distributions.</li> <li>• Simulating probability distributions (binomial, normal, uniform, and Poisson distributions)</li> <li>• Introduction to hypothesis testing and estimation <i>Statistics</i></li> </ul>	Agresti, Alan, and Christine Franklin. <i>Statistics: The Art and Science of Learning from Data</i> . 3rd ed. Pearson, 2013. (Chapters 5, 6, 7 and 8)
Week 8	Second Assessment	
Week 9-12	<ul style="list-style-type: none"> <li>• Assumptions of parametric data</li> <li>• The assumption of normality: checking normality visually</li> <li>• Quantifying normality with numbers</li> <li>• Exploring groups of data</li> <li>• Testing whether a distribution is normal.</li> <li>• Doing the Shapiro–Wilk test in R</li> <li>• Testing for homogeneity of variance</li> <li>• Levene’s test</li> <li>• Reporting Levene’s</li> <li>• Hartley’s Fmax: the variance ratio</li> <li>• Correcting problems in the data</li> <li>• Dealing with outliers</li> <li>• Dealing with non-normality and unequal variances</li> </ul>	<i>Chapter 5, Discovering Statistics using R, Andy Field, Jeremy Miles and Zoe Field.</i>
Week 13-14	Regression Analysis	Agresti, Alan, and Christine Franklin. <i>Statistics: The Art and Science of Learning</i>

		from Data. 3rd ed., Pearson, 2013. (Chapters 13 and 14). Chapter 7, Discovering Statistics using R, Andy Field, Jeremy Miles and Zoe Field. Chapter 11, R Cookbook.
Week 15	Projects Presentation	