



## KAUTILYA SCHOOL OF PUBLIC POLICY

GITAM (Deemed to be University)  
Rudraram, Patancheru Mandal  
Hyderabad, Telangana 502329

<b>Course Code:</b> PPOL6731	<b>Course Title:</b> Statistics for Public Policy	
<b>Trimester:</b> 1	<b>Course Type:</b> Core	<b>Credits:</b> 3
<b>Home Program(s):</b> MPP	<b>Batch/Academic Year:</b> 2023-2025	
<b>Course Lead:</b> Dr. Vishnu S Pillai	<b>Assigned T/RA:</b> Ms. Srestha & Ms. Rashika	

### Course Description

Statistical thinking, comprehension, and competence are more crucial than ever in today's increasingly complex environment. Thus, understanding the mathematical and statistical concepts is necessary to understand the various policy frameworks. This course introduces students to statistical techniques used in Policy analysis and evaluation. It provides a solid understanding of the ideas and criteria for performing quantitative research. The emphasis will be on establishing a conceptual grasp of statistics and data analysis abilities and learning how to apply them to policy analysts' challenges.

This course will focus on Python as the programming language for data analysis. No prerequisite programming knowledge is required for this course, and students will receive valuable hands-on experience using Python for data analysis. The course content lays the groundwork for students to pursue more advanced topics in the discipline as per their career and research requirements. The techniques discussed in this course will significantly help students learn and understand theories and applications discussed in advanced courses on data analysis in the Public Policy domain.

### Learning Objectives

1. Engage students in sound statistical thinking and reasoning.
2. Explore multiple representations of topics, including graphical, symbolic, numerical, oral, and written.
3. Provide the students with the skills required to be proficient in statistics.
4. Provide the students with basic programming skills for data analysis using Python.

### Course Outcomes

On successful completion of this course:

1. Students will be equipped with the skills in describing, analysing, and interpreting policy data using statistical methods.

2. Develop research questions and test hypotheses for a research project.
3. Understand practical applications of statistical concepts that can be used in the public policy sector.
4. They can relate and apply statistical methodology to policy analysis using Python.
5. The students can produce policy papers and reports using statistical analysis.

### Assessment Criteria

The assessment criteria of the course are given in the table below:

Component	Weight	Remarks										
Attendance	15%	15 Marks if 22 -24 classes are attended										
		10 Marks if 19-21 classes are attended										
		5 Marks if 16-18 classes are attended										
		0 Marks if fewer than 16 classes are attended										
		A minimum of <b>65%</b> Attendance is required to qualify for the paper.										
Class participation	15%	<table border="1"> <thead> <tr> <th>Marks</th> <th>Criteria for the Sessions involving discussions</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>Absent/ non-participation</td> </tr> <tr> <td>1 to 5</td> <td>Responds to questions asked in class by the lecturer - offers straightforward information.</td> </tr> <tr> <td>6 to 10</td> <td>Contributes to the discussions in the class. Responds thoughtfully to other students' views.</td> </tr> <tr> <td>11 to 15</td> <td>Read and refer to materials other than the course content and provide valuable insights into the discussions in the class.</td> </tr> </tbody> </table>	Marks	Criteria for the Sessions involving discussions	0	Absent/ non-participation	1 to 5	Responds to questions asked in class by the lecturer - offers straightforward information.	6 to 10	Contributes to the discussions in the class. Responds thoughtfully to other students' views.	11 to 15	Read and refer to materials other than the course content and provide valuable insights into the discussions in the class.
		Marks	Criteria for the Sessions involving discussions									
		0	Absent/ non-participation									
		1 to 5	Responds to questions asked in class by the lecturer - offers straightforward information.									
		6 to 10	Contributes to the discussions in the class. Responds thoughtfully to other students' views.									
		11 to 15	Read and refer to materials other than the course content and provide valuable insights into the discussions in the class.									
		<b>Short coding exercises will be conducted during the class for Sessions that don't have class discussions.</b>										
<b>Sessions with class discussions – 2, 8, 11, 19 and 24</b>												
<b>Sessions with short coding exercises – 6, 16, 18, 21 and 23.</b>												
Each student's participation in class will be graded out of 15 points based on the above criteria for each session involving discussions or coding exercises. For the final evaluation of class participation, the average will be used based on the total number of sessions where the evaluation was conducted (here, 10).												
Assignment	20%	Problem sets based on Units 1 and 2. The problem sets will be shared after completing Session 6. Students should complete it and submit it in one week.										
Theory examination	20%	Examination – theory (“pen and paper” – no coding questions). Emphasis will be given to the contents taught till Session 18 (Unit 6). The exam duration is 90 minutes.										
Final examination (laboratory)	30%	Open-book examination – data analysis using Python after Session 24 (Unit 7). The exam duration is 120 minutes.										

## Classroom Etiquette

The consumption of food is not allowed in class. The use of laptops, tablets, smartphones, and other electronic devices is not allowed unless the Professor has specified any such device as a pre-requisite for a particular session. If you are unable to take notes by hand, please discuss an alternative arrangement with the teaching team. If you have been authorised to use a tablet or laptop, make sure you are disconnected from the Internet. Phones must be turned off at the start of class.

## Required Textbook

Alan Agresti. (2018). *Statistical Methods for the Social Sciences* by Alan Agresti., 5th ed. Pearson.

## Recommended Textbooks

Black, K. (2022). *Business statistics: for contemporary decision making*. John Wiley & Sons.

Macfie, B. P., & Nufrio, P. M. (2005). *Applied statistics for public policy*. Routledge.

Imai, Kosuke. (2018). *Quantitative social science: an introduction*. Princeton University Press, 2018.

Wes McKinney. (2017). *Python for Data Analysis*., 2nd ed. O'REILLY.

Daniel J. Denis. (2021). *Applied Univariate, Bivariate, and Multivariate Statistics Using Python*., 1st ed. John Wiley & Sons.

Charles R. Severance. 2016. *Python for Everybody: Exploring Data in Python 3*.

Gun, A.M., Gupta, M.K. and Dasgupta, B. (2013): *Fundamentals of Statistics, Vol. I*, The World Press, Kolkata.

Rohatgi, V. K. and Saleh, A.K. Md. E. (2008): *An Introduction to Probability and Statistics*. 2nd Edn. (Reprint) John Wiley and Sons.

## Course Schedule

Unit 1	Sessions: 2	Laying the foundation
<ol style="list-style-type: none"> <li>1. Why Statistics for Public Policy?</li> <li>2. Descriptive and inferential statistics.</li> <li>3. Types of variables and scales of measurement.</li> <li>4. Descriptive statistics - measures of central tendencies, measures of spread.</li> <li>5. Basics of Probability.</li> </ol>		
Unit 2	Sessions: 4	Introducing Python

<ol style="list-style-type: none"> <li>1. Introducing Python.</li> <li>2. Python – conditions, loops, and functions.</li> <li>3. Python - univariate descriptive statistics - measures of central tendency and spread.</li> <li>4. Python - NumPy package - creating and working with n-dimensional arrays - basics, filtering, slicing, and other array operations.</li> </ol>		
<b>Unit 3</b>	<b>Sessions: 4</b>	<b>Understanding probability distributions</b>
<ol style="list-style-type: none"> <li>1. Probability distributions.</li> <li>2. Python NumPy package – simulations and analysing distributions.</li> </ol>		
<b>Unit 4</b>	<b>Sessions: 2</b>	<b>Data collection</b>
<ol style="list-style-type: none"> <li>1. Methods of data collection – introducing experimental, quasi-experimental, and non-experimental methods.</li> <li>2. Various sampling methods, errors, and biases.</li> </ol>		
<b>Unit 5</b>	<b>Sessions: 4</b>	<b>Entering the world of data analysis with Python</b>
<ol style="list-style-type: none"> <li>1. Python Pandas introduction - data structures and basic functionalities.</li> <li>2. Data loading in Python.</li> <li>3. Data cleaning and dealing with missing data.</li> <li>4. Visualising data in Python (Matplotlib, Seaborn).</li> </ol>		
<b>Unit 6</b>	<b>Sessions: 2</b>	<b>Hypothesis testing</b>
<ol style="list-style-type: none"> <li>1. Hypothesis testing - basic steps.</li> <li>2. Hypothesis testing of means when the population standard deviation is known/unknown.</li> <li>3. Analysis of Variance – ANOVA.</li> </ol>		
<b>Unit 7</b>	<b>Sessions: 6</b>	<b>Prediction</b>
<ol style="list-style-type: none"> <li>1. Setting the context for linear regression.</li> <li>2. Linear regression model using Python.</li> <li>3. Multiple linear regression.</li> <li>4. Introducing the basics of Machine Learning.</li> </ol>		