

ADL202	PYTHON AND STATISTICAL MODELLING LAB	Category	L	T	P	Credits	Year of introduction
			0	0	3		
		PCC				2	2019

Preamble: The Python and Statistical modelling course is intended to impart the elementary concepts of Python and apply various statistical techniques to a variety of data. This course provides the learners with hands-on experience in Python and statistical processes like measures of central tendency, measures of dispersion, probability distributions, graphical analysis, correlation analysis and use of statistical analysis software. The course enables the students to get an exposure to Python programming and use proper methods to analyze and interpret data effectively.

Prerequisite: A basic knowledge of Probability and Statistical Modelling.

Course Outcomes: After the completion of the course the student will be able to

CO#	Course Outcomes
CO 1	Experiment with concepts of iteration, function, string and list (Cognitive Knowledge Level: Apply)
CO 2	Identify the importance of tuples, dictionary traversal, dictionary methods, files and operations (Cognitive Knowledge Level: Apply)
CO 3	Model graphical representation of data, measures of central tendency and measures of dispersion (Cognitive Knowledge Level: Apply)
CO 4	Solve problems based on Binomial distribution, Poisson distribution, sampling and regression analysis (Cognitive Knowledge Level: Apply)
CO 5	Make use of various correlation tests and utilize statistical analysis software (Cognitive Knowledge Level: Apply)

Mapping of course outcomes with program outcomes

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	☒	☒	☒					☒				☒
CO2	☒	☒	☒	☒				☒				☒
CO3	☒	☒	☒	☒				☒				☒
CO4	☒	☒	☒	☒				☒				☒
CO5	☒	☒	☒	☒	☒			☒				☒

Abstract POs defined by National Board of Accreditation			
PO#	Broad PO	PO#	Broad PO
PO1	Engineering Knowledge	PO7	Environment and Sustainability
PO2	Problem Analysis	PO8	Ethics
PO3	Design/Development of solutions	PO9	Individual and team work
PO4	Conduct investigations of complex problems	PO10	Communication
PO5	Modern tool usage	PO11	Project Management and Finance
PO6	The Engineer and Society	PO12	Lifelong learning

Assessment Pattern

Bloom's Category	Continuous Assessment Tests	End Semester Examination (Percentage)
Remember	20	20
Understand	20	20
Apply	60	60
Analyze		
Evaluate		
Create		

Mark distribution

Total Marks	CIE Marks	ESE Marks	ESE Duration
150	75	75	3 hours

Continuous Internal Evaluation Pattern:

- Attendance : 15 marks
 Continuous Evaluation in Lab : 30 marks
 Continuous Assessment Test : 15 marks
 Viva Voce : 15 marks

The marks will be distributed as Design/Algorithm 30 marks, Implementation/Program 20 marks, Output 20 marks and Viva 30 marks. Total 100 marks which will be converted out of 15 while calculating Internal Evaluation marks.

End Semester Examination Pattern:

The marks will be distributed as Design/Algorithm 30 marks, Implementation/Program 20 marks, Output 20 marks and Viva 30 marks. Total 100 marks will be converted out of 75 for End Semester Examination.

Fair Lab Record:

All Students attending the Statistical Modelling Using Python Lab should have a Fair Record. The fair record should be produced in the University Lab Examination. Every experiment conducted in the lab should be noted in the fair record. For every experiment in the fair record, the right-hand page should contain Experiment Heading, Experiment Number, Date of experiment, Aim of the Experiment and the operations performed on them, Details of experiment including algorithm and result of Experiment. The left-hand page should contain a print out of the code used for experiment and sample output obtained for a set of input.

SYLLABUS**PYTHON AND STATISTICAL MODELLING LAB**

1. Familiarization of expressions, conditional and iteration statements.
2. Problems on function and function calls. **
3. String traversal and other important string methods. **
4. List traversal and list operations. **
5. Tuples, dictionary traversal and dictionary methods. **
6. Problems based on files and operations. **
7. Problems on graphical representation of data. **
8. Problems based on measures of central tendency and measures of dispersion using raw data and grouped data. **
9. Application problems based on Binomial and Poisson distribution. **
10. Implement Chi-square test for goodness of fit. **
11. Perform t-test for difference of means. **
12. Implement Correlation tests. (Karl Pearson correlation coefficient and Spearman rank correlation coefficient).
13. Estimation of gain in precision due to stratification. **
14. Analysis of a one way/ two-way ANOVA.
15. Problems on Lines of regression, regression coefficients, angle between regression lines.
16. Familiarization with statistical analysis software. (SPSS or similar) **

**mandatory

1. Write a program to find the largest of three numbers.
2. Write a program to print the multiplication table of a number n.
3. Write a program to find Surface area and volume of a cylinder using function.
4. Write a program to replace a word by another word in a sentence.
5. Write a program to confirm the validity of an email id by verifying its format.
6. Write a program to remove every occurrence of a number from a list.
7. Write a program to add two matrices.
8. Write a program to read a tuple of numbers and print even tuple and odd tuple.
9. Create a dictionary with a set of book title and corresponding stock. Write a program to update the stock and to add or delete books.
10. A set of numbers are stored in a file. Write a program to print the prime numbers among them.
11. Write a program to count the number of words, sentences, upper case letters, lowercase letters and special symbols in a text stored in file.
12. Plot a graph $y = f(x)$.
13. The areas of the various continents of the world (in millions of square miles) are as follows: 11.7 for Africa; 10.4 for Asia; 1.9 for Europe; 9.4 for North America; 3.3 Oceania; 6.9 South America; 7.9 Soviet Union. Draw a bar chart representing the given data.
14. Draw the histogram of the following data:

Height of student(m)	135 - 140	140 - 145	145 - 150	150 - 155
No. of students	4	12	16	8

15. Table contains population and murder rates (in units of murders per 100,000 people per year) for different states. Compute the mean, median and variance for the population.

State	Population	Murder
Alabama	4,779,736	5.7
Alaska	710,231	5.6
Arizona	6,392,017	4.7
Arkansas	2,915,918	5.6
California	37,253,956	4.4
Colorado	5,029,196	2.8
Connecticut	3,574,097	2.4
Delaware	897,934	5.8

16. Calculate the S.D. and coefficient of variation (C.V.) for the following table:

Class:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency:	5	10	20	40	30	20	10	5

17. If X is binomially distributed with 6 trials and a probability of success equal to 0.25 at each attempt, what is the probability of:
- exactly 4 successes
 - at least one success
18. If the random variable X follows a Poisson distribution with mean 3.4, find $P(X=6)$.
19. A random sample of 395 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table. Are gender and education level dependent at 5% level of significance?

	High School	Bachelors	Masters	Ph.D.	Total
Female	60	54	46	41	201
Male	40	44	53	57	194
Total	100	98	99	98	395

20. Calculate the correlation coefficient of the two variables shown in the table below.

Person	Hand	Height
A	17	150
B	15	154
C	19	169
D	17	172
E	21	175

21. Suppose a sample of 16 light trucks is randomly selected off the assembly line. The trucks are driven 1000 miles and the fuel mileage (MPG) of each truck is recorded. It is found that the mean MPG is 22 with a SD equal to 3. The previous model of the light truck got 20 MPG. Conduct a t-test of the null hypothesis at $p = 0.05$

22. The mean productivity rating for all employees at a company was 3.8 on a five-point scale last year. This year you get ratings from a representative sample of fifteen employees from the Human Research Management. Do the data from this sample provide evidence that employee productivity in the department of Human Resource Management is significantly higher than in the company as a whole? Write the null and alternative hypotheses for this problem. Use statistical analysis software to test the null hypothesis stated above.

23. Obtain the regression equation for predicting systolic blood pressure from job satisfaction with reference to the given data using statistical analysis software. If one knows that a subject in the future has a score on job satisfaction of 15, what is their systolic blood pressure predicted to be? What is the standard error of estimate?

Job Satisfaction	Systolic BP
34	124
23	128
19	157
43	133
56	116
47	125
32	147
16	167
55	110
25	156