Government of Karnataka
Department of Collegiate and Technical Education
Board of Technical Examinations, Bangalore

Course Code	20SC02P	**Semester**	I/II
Course Title	**STATISTICS AND ANALYTICS**	**Course Group**	Core
No. of Credits	4	**Type of Course**	Lecture and practice
Course Category	Practice	**Total Contact Hours**	6 Hrs Per Week
			78 Hrs Per Semester
Prerequisites	10TH LEVEL MATHEMATICS	**Teaching Scheme**	(L: T:P)-1:0:2
CIE Marks	60	**SEE Marks**	40

RATIONALE

Statistics and analytics help the learner to use the proper methods to collect the data, employ the correct analyses, effectively present the results and conduct research, to be able to read and evaluate journal articles, to further develop critical thinking and analytic skills, to act as an informed consumer and to know when you need to hire outside statistical help. The python language is one of the most accessible programming languages available because it has simplified syntax and not complicated, which gives more emphasis on natural language.

COURSE OUT COMES

At the end of the course, student will be able to

CO1	Understand the tools of data collection, classification and cleaning of data.
CO2	Able to summarize the given statistical data
CO3	Understand the measure of location and dispersion of data.
CO4	Learn the basics of Python programming.

DETAILS OF COURSE CONTENT

The following topics/subtopics is to be taught and assessed in order to develop Unit Skill Sets for achieving CO to attain identified skill sets.

UNIT NO	Unit skill set (In cognitive domain)	Topics/Subtopics	Hours L-T-P
UNIT-1 **STATISTICAL DATA COLLECTION AND TYPES**	➢ Able to collect statistical data. ➢ Able to distinguish the data types. ➢ Understands the usage of data collection tools ➢ Able to specify problem statement for data collection ➢ Able to collect data pointing the root cause of the problem statement.	a Definition of data and classification (qualitative quantitative discrete and continuous data). b Data collection tools iv) Questionnaires. v) Survey. vi) Interviews. vii) Focus group discussion. 1.3 Data cleaning.	4-0-8
UNIT-2 **SUMMARIZATION OF DATA**	➢ Sketches bar, pie and histograms on Microsoft Excel spread sheet. ➢ Sketches frequency curve and frequency polygon for the data set on Microsoft Excel spread sheet. ➢ Sketches bar, pie and histograms on Microsoft Excel spread sheet. ➢ Sketches frequency curve and frequency polygon for the data set on Microsoft Excel spread sheet.	a Descriptive statistics viii) Datatabulation(frequency table ix) Relative frequency table. b Grouped data x) Bar graph xi) Pie chart xii) Line graph xiii) Frequency polygon xiv) Frequency curve xv) Relative frequency polygon xvi) Histograms xvii) Box plot xviii) Leaf-stem plot To be done in Microsoft excel.	8-0-16
UNIT-3 **MEASURE OF LOCATION AND DISPERSION**	➢ Able to determine the descriptive statistical variables using Microsoft Excel.	a Determination of central tendencies Range, Mean, Mode and Median for the data in Microsoft excel. b Determination of absolute	6-0-12

	➢ Able to determine the absolute measures of dispersion of the given data set. ➢ Explain the symmetry and asymmetry of the distributed data.	measures of dispersion for data like range quartile deviation, mean deviation, standard deviation and variance in Microsoft Excel. c Skewness and kurtosis graphs in Microsoft excel and interpretations of results.	
UNIT-4 **INTRODUCTION TO** **PYTHON** **PROGRAMMING**	➢ Able Install and run the Python interpreter. Create and execute Python programs. ➢ Understand the concepts of file I/O. ➢ Able to read data from a text file using Python. ➢ Learn variable declarations in Python. ➢ Learn control structures. ➢ Learn loop constructs.	4.1 Introduction to PYTHON. 4.2 Syntax of PYTHON. 4.3 Comments of PYTHON. 4.4 Data types of PYTHON. 4.5 Variables of PYTHON. 4.6 If-else in PYTHON. 4.6 Loops in PYTHON. 4.7 Arrays and functions in PYTHON.	8-0-16

SL NO	Practical outcomes/Practical exercises	Unit no	PO	CO	L:T:P
1	Prepare a questionnaire (closed end) containing 25 questions for a specified problem statement: for example experience of an individual in a restaurant.	1	1,2,4,5,7	1	0:0:2
2	Prepare a Google form for a specified problem statement to collect the dataset. (for example questionnaire to conduct online quiz)	1	1,2,4,5,7	1	0:0:2
3	Send out a survey on your problem statement to number of 50 (By Google forms) and collect the data.	1	1,2,4,5,7	1	0:0:2
4	Remove duplicate or irrelevant observations. Remove unwanted observations from the dataset provided, including duplicate observations or irrelevant observations.	1	1,2,4,5,7	1	0:0:2
5	In Microsoft Excel spread sheet draw the frequency distribution table for the given data (data set should contain minimum 50 data).	2	1,2,4,5,7	2	0:0:2
6	In Microsoft Excel spread sheet draw the relative frequency distribution table for the given data (data set should contain minimum 50 data).	2	1,2,4,5,7	2	0:0:2
7	Using Microsoft Excel spread sheet plot bar graph for the data collected from 100 people(for example, conduct a survey on the favorite fruit of a person in your locality(restricting to 5 to 6 fruits). Explain the bar graph with minimum 30 words.	2	1,2,4,5,7	2	0:0:2
8	Using Microsoft Excel spread sheet plot pie chart for the data collected from 50 people(for example, conduct a survey on the smokers with respect to their ages in your locality. Explain the pie chart with minimum 30 words.	2	1,2,4,5,7	2	0:0:2
9	Using Microsoft Excel spread sheet draw a line graph for the given dataset.	2	1,2,4,5,7	2	0:0:2
10	Using Microsoft Excel spread sheet draw frequency polygon and frequency curve for the data collected from 50 people. (For example, marks obtained by the students in your class in 5 subjects in previous examination). Explain your observations from the graph in minimum 30 words.	2	1,2,4,5,7	2	0:0:2
11	Using Microsoft Excel spread sheet construct a box plot for the given dataset. (For example dataset can be the number of passengers in a flat form at different time in a day).	2	1,2,4,5,7	2	0:0:2
12	Using Microsoft Excel spread sheet construct a leaf plot for the given dataset. Explain the graph with minimum 30 words.	2	1,2,4,5,7	2	0:0:2

13	Using Microsoft Excel spread sheet find the Mean, Mode and Median for the data (univariate data) given and also represent them in a Histogram.	3	1,2,4,5,7	2	0:0:2
14	Generate a 50 random data sample (even and odd number dataset) using Microsoft Excel spread sheet and determine the range and Quartiles.	3	1,2,4,5,7	2	0:0:2
15	Collect the current yield of a crop from 50 different persons (problem statement can be changed according to priorities of the tutor) in your locality and determine mean deviation and Quartile deviation in Microsoft excel spread sheet and brief your inference with less than 30 words.	3	1,2,4,5,7	3	0:0:2
16	Collect the data of any 2 livestock population from 50 different houses in your locality (problem statement can be changed according to priorities of the tutor) and determine standard deviation for both the two separately in Microsoft excel spread sheet and brief your inference with less than 30 words.	3	1,2,4,5,7	3	0:0:2
17	Collect the data of two wheeler (with a rider and a pillion) crossing a busy junction in your locality in the peak hours (problem statement can be changed according to priorities of the tutor) and determine the variance of the data in Microsoft excel spread sheet and brief your inference with less than 30 words.	3	1,2,4,5,7	3	0:0:2
18	Using Microsoft Excel spread sheet draw a Skewness graph and kurtosis graph for randomly generated dataset.	3	1,2,4,5,7	3	0:0:2
20	Write a python program to add 2 integers and 2 strings and print the result.	4	1,2,4,5,7	4	0:0:2
21	Write a python program to find the sum of first 10 natural numbers.	4	1,2,4,5,7	4	0:0:2
22	Write a python program to find whether the number is odd or even.	4	1,2,4,5,7	4	0:0:2
23	Write a python program to find the variance and standard deviation for the given data..	4	1,2,4,5,7	4	0:0:2
24	Write a python program to display student marks from the record.	4	1,2,4,5,7	4	0:0:2
25	Write a python program to create a labeled bar graph using matpoltlib. pyplot.	4	1,2,4,5,7	4	0:0:2
26	Write a python program to create a labeled pie chart using matpoltlib. pyplot.	4	1,2,4,5,7	4	0:0:2
Total Hours					**0:0:52=52**

MAPPING OF CO WITH PO

CO	Course Outcome	PO Mapped	Experiment Linked	Cognitive Level R/U/A	Tutorial & Practical Sessions in Hrs.	TOTAL
CO1	Understand the tools of data collection, classification and cleaning of data.	1,2,4,5,7	1-4	A	12	12
CO2	Able to summarize the given statistical data	1,2,4,5,7	5-12	A	33	33
CO3	Understand the measure of location and dispersion of data.	1,2,4,5,7	13-18	A	12	12
CO4	Learn the basics of Python programming.	1,2,4,5,7	19-26	A	21	21
					78	**78**

Course	CO's	Programme Outcomes (PO's)						
		1	2	3	4	5	6	7
Statistics & Analytics	CO1	3	3	0	3	3	0	3
	CO2	3	3	0	3	3	0	3
	CO3	3	3	0	3	3	0	3
	CO4	3	3	0	3	3	0	3
Level 3- Highly Mapped, Level 2-Moderately Mapped, Level 1-Low Mapped, Level 0- Not Mapped								

SUGGESTED LEARNING RESOURCES:

1. Statistical Analysis with Excel For Dummies (For Dummies Series) Paperback Import, 9 April 2013 by Joseph Schmuller (Author)
2. https://www.brianheinold.net/python/A Practical Introduction to Python ProgrammingHeinold.pdf
3. http://www.bikeprof.com/uploads/9/0/6/5/9065192/excel stats handout npl.pdf
4. https://adminfinance.umw.edu/tess/files/2013/06/Excel-Manual1.pdf
5. https://www.brianheinold.net/python/A Practical Introduction to Python ProgrammingHeinold.pdf
6. Introduction to Python programming for beginners by Vivian Baily Kindle edition.
7. PYTHON PROGRAMMING: Python programming: the ultimate guide from a beginner to expert by Clive Campbell.
8. Open source for python: https://hub.gke2.mybinder.org/user/jupyterlab-jupyterlab-demo-zfkdwy4y/lab

SUGGESTED LIST OF STUDENT ACTIVITY

Note: The following activities or similar activities for assessing CIE (IA) for 10 marks (Any one)

1	Describe the data collection activity itself (interviews, surveys, library research, etc.) AND why this specific form of data collection was chosen. Be sure to explain why you think this kind of data will help you in your design process. Also be sure to provide details about the activity: how many interviews, how long they took, where they took place, how many questions asked in a survey, how many respondents, etc. Present the results of your data collection. You do not have to have completely analyzed all your data, but do make sure you present the results of your research. If you did a survey, please attach a copy of the survey as an appendix; if you did interviews, please attach a copy of the interview questions. Discuss any preliminary analysis of your data. What have you learned thus far from the data should be discussed from an analytical perspective (rather than a data dump). For example, if you surveyed people about their use of the local bus system, and 90% of your respondents said they take the bus when it is raining, and 60% of your respondents said they usually wait more than 10 minutes for a bus, think about what this teaches you rather than just the information itself. In this instance, you can see that people are generally waiting for several minutes in the rain for a bus, so a covered bus stop might be a good idea. Keep in mind that your findings from data should lead directly to the conclusions you make about your design recommendations. This is the time to begin thinking very specifically about your research in those terms. This is also an opportunity to think about your definition of "better" and how it applies to your design goals and your choice of research activities (for example, if you are choosing to make something better by making it cheaper, maybe you are interviewing people to see how much loss of functionality or decrease in features for a technology they are willing to tolerate).
2	https://ils.unc.edu/courses/2013_spring/inls541_001/Assignments.html#Assignment_9

	DOWNLOAD a dataset from the above link and use data visualization tools to analyze it.
3	Acquire the dataset from https://www.kaggle.com/datasets (For example acquire the data of IPL ball by ball scores and find the standard deviation and variance of score of a batsmen)and clean the data for the root cause of the problem statement and summarize the date and explain the inference.

COURSE ASSESSMENT AND EVALUATION CHART

Method	What		To whom	When/Where (Frequency in the course)	Max Marks	Evidence collected	Course outcomes
DIRECT ASSESSMENT	CIE (Continuous Internal Evaluation)	Models	Students	Two IA Tests (Written)	20	Blue Book	1,2,3.
				Three Skill tests	20	Model	1,2,3
				Student Activity	20	Model/Report	
				TOTAL	60		
	SEE (Semester End Examination)	End Exam		End of the course	100	Models	1,2,3
INDIRECT ASSESSMENT	Student Feedback on course		Students	Middle of the course		Feedback forms	1,2,3, Delivery of course
	End of Course Survey			End of the course		Questionnaires	1,2,3 Effectiveness of Demonstrations& Assessment Methods

Sl.No	Assessment	Duration	Max marks	Conversion
1	CIE Assessment 1 (Written Test -1-theory) - At the end of **3rd week**	60 minutes	20	Average of two written tests 20
2	CIE Assessment 2 (Written Test -2-theory) - At the end of **13th week**	60 minutes	20	
3	CIE Assessment 3 (Skill test) - At the end of **5th week**	3 Hrs	20	Average of three skill tests 20
4	CIE Assessment 4 (Skill test) - At the end of **7th week**	3 Hrs	20	
5	CIE Assessment 5 (Skill test) - At the end of **9th week**	3Hrs	20	
6	CIE Assessment 6 (Student activity) - At the end of **11th week**	-	20	20
7	Total Continuous Internal Evaluation (CIE) Assessment			60
8	Semester End Examination (SEE) Assessment (Practical Test)	3Hrs	100	40
		Total Marks		**100**

Note:

1. CIE written test is conducted for 20 marks (Two sections). Each section shall have two full questions of same CL, CO. Student shall answer one full question (10 marks) from each section.
2. CIE Skill test is conducted for 100 marks (3 Hours duration) as per scheme of evaluation and the obtained marks are scaled down to 20 marks.
3. SEE is conducted for 100 Marks (3 Hours duration) as per scheme of evaluation.

MODEL QUESTION PAPER

CIE, SKILL TEST AND SEMESTER END EXAMINATION

Course & Programme: Common to all Engineering Programmes.	**Semester:** II
Subject: Statistics and Analytics Practice	**Max Marks:** 100
Course Code : 20SC21P	**Duration** : 3Hrs

Instruction to the Candidate: Answer both questions

Qn.No	Question	CL	CO	PO	Marks
1	For the given ungrouped data set plot the bar graph by grouping the data in Microsoft excel spread sheet and interpret the obtained results. (Dataset. bar graphs and interpretation have to be entered in the answer script). OR Generate a random data set in Microsoft excel spread sheet containing 50 data and find the mean mode and median in Microsoft excel spread sheet and interpret the obtained results. (Dataset, bar graphs and interpretation have to be entered in the answer script).	A	2,3	1,2,4,5,7	50
2	Write the python program to enter two integers and two strings and to print the sum two integers and two strings.	A	4	1,2,4,5,7	50

Questions are not framed from Unit 1 in the final SEE. Short questions can only be asked from that unit.

SCHEME OF EVALUATION FOR BOTH CIE AND SEE

Sl. No	Particulars	Marks
1	Short questions from Unit 1	10
2	Observation	30

3	Conduction	20
4	Output and Interpretation of result	20
5	Viva-voce	20
Total		100

EQUIPMENT LIST

FOR STATISTICS AND DATA ANALYTICS LAB

2 laboratories. Each containing 30 computers (Desktop) with the following system requirements.

SYSTEM REQUIREMENTS			
SL NO	REQUIREMENTS	MINIMUM	RECOMMENDED
1	RAM	4GB FOR FREE RAM	8GB OF TOTAL SYSTEM RAM
2	DISK SPACE	2.5 GB AND 1 GB FOR CACHES	SSD DRIVE WITH AT LEAST 5 GB OF FREE SPACE
3	MONITOR RESOLUTION	1024x768	1920×1080
4	OS(OPERATING SYSTEM)	OFFICIALLY RELEASED 64-BIT VERSIONS OF THE FOLLOWING: MICROSOFT WINDOWS 8 OR LATER	LATEST 64-BIT VERSION OF WINDOWS