

[Total No. of CO's 05--]

Seat No:

[Total No. of Pages: -02--]

G. H. Raisoni College of Engineering and Management, Pune.
(An Autonomous Institution affiliated to Savitribai Phule , Pune University)
F.Y B. Tech/ (Branch *AIIT*) (Term- I)
ESE Winter 2020. (2020 Pattern)
Subject Name Foundtation of Data Analytics(UCOL102)

[Time:-- 2 Hours]

[Max. Marks50]

COURSE OUTCOME:

- CO1: Analyze the dataset and perform Descriptive Statistics*
CO2: Analyze the dataset and perform an Inferential Statistics
CO3: Apply linear regression on the given dataset
CO4: Apply the logistic regression on the given dataset
CO5: Create an interactive data visualization

Instructions to the candidates:

- 1) (CO1/CO2/CO....)at the beginning of question/sub question indicates the course outcome related to the question.
- 2) All questions compulsory.
- 3) Neat diagrams must be drawn wherever necessary.
- 4) Figures to the right indicate full marks.
- 5) Assume suitable data, if necessary.
- 6) Other Instructions, if any.

CO	Sub		Marks	BL
	Question			

CO1	a)	Is data analysis and data analytics are same concept? Justify your answer.	[3]	L4
-----	----	-------------------------------------------------------------------------------	-----	----

CO1	b)	Compare correlation Verses covariance.	[4]	L4
-----	----	----------------------------------------	-----	----

CO1	c)		[3]	L4
-----	----	--	-----	----

Price table of branded ball pen's is shown here;

Sr.No.	Ball Pen Brand	Price (Rs)
1	Classmate	12
2	Reynolds	15
3	Cello	18
4	Techno tip	15
5	Pilot	20

Calculate central (avg) price of a ballpen using Central tendency measures.

CO1	d)	Compare and contrast qualitative data verses quantitative data.	[4]	L4
-----	----	-----------------------------------------------------------------	-----	----

CO1	e)	List the characteristics of Standard Deviation	[3]	L4
-----	----	------------------------------------------------	-----	----

CO2	a)	Describe the properties of Normal Distribution Curve and List the types of distributions.	[5]	L4
-----	----	-------------------------------------------------------------------------------------------	-----	----

CO2	b)	The height of students studying at a Engineering college follows a normal distribution with a mean of 1.62 m and a standard deviation of 0.12. What is the probability that the mean of a random sample of 100 students will be taller than 1.60 m?	[3]	L4
-----	----	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----	----

CO2	c)	In a sample of 400 selected at random, a sample mean of 50 was	[3]	L4
-----	----	----------------------------------------------------------------	-----	----

obtained. Determine the confidence interval with a confidence level of 97% for the average population.

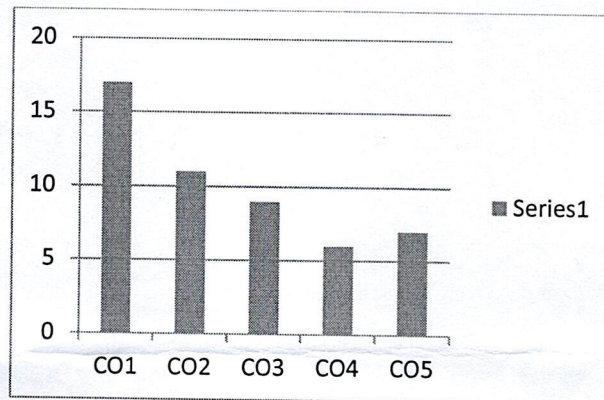
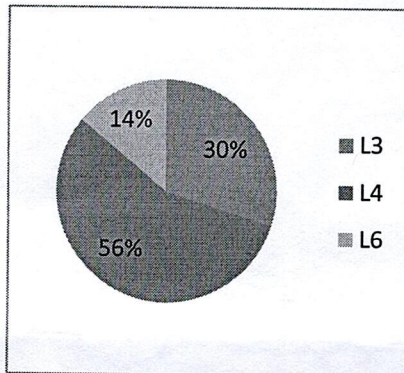
CO3	a)	Relate the terms sum of squares total (SST) , sum of squares due to regression (SSR) and sum of squares error (SSE) with mathematical expression.	[5]	L3
CO3	b)	What is P-Value? How do you interpret it?	[4]	L3
CO4	a)	What is logistic regression?	[3]	L3
CO4	b)	Illustrate the range properties of logistic regression functions.	[3]	L3
CO5	a)	Elaborate various graphical ways of data representation.	[3]	L6
CO5	b)	The number of hours spent by a school student on various activities on a working day, is given below.	[4]	L6

Construct a pie chart using the angle measurement.

Activity	Sleep	School	Play	Homework	Others
Number of Hours	8	6	3	3	4

So on.....

NOTE:



- BL – Bloom's Taxonomy Levels (1- Remembering, 2- Understanding, 3 – Applying, 4 – Analysing, 5 –Evaluating, 6 - Creating).