UPES

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Subject: Applied Statistical Analysis   **Semester:VII**
Programme: B. Tech CSE+BFSI   **Code   : CSIB 225**
 Max. Marks   : 100   **Duration   : 3 Hrs**
No. of page/s:4

 Instructions:

SECTION A
Section: A (Answer all questions, each question carries 4 Marks)   5X 4M=20M

S. No.		Marks	CO

Q 1	Explain the difference between following
	    a. Primary and Secondary Data
	    b. Descriptive Statistics Inferential Statistical
		4	CO1

Q 2	Explain different types of sampling technique with example.	4	CO1
Q 3	Aman visits campus every Thursday evening. However,some days the parking garage is full, often due to college events. There are academic events on 35% of evenings, sporting events on 20% of evenings, and no events on 45% of evenings. When there is an academic event, the garage fills up about 25% of the time, and it fills up 70% of evenings with sporting events. On evenings when there are no events, it only fills up about 5% of the time. If Aman comes to campus and finds the garage full, what is the probability that there is a sporting event?	4	CO2
Q 4	Prove that var(aX+b)=a^2var(X).Where X is a random variable.	4	CO2
Q 5	A poll conducted in 2013 found that 32% of Indian adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion. Construct a 99% confidence interval for the fraction of Indian adult Twitter users who get some news on Twitter, and interpret the confidence interval in context	4	CO3
	SECTION B **Answer all questions, each question carries 10 Marks) 4X 10M=40M**		
Q 6	A mail-order company has commissioned market research in an attempt to find whether there is a connection between the types of area that people live in and if they buy from mail-order companies. The following table gives the results from a random sample of 150 people:	10	CO3

Area	Urban	Suburban	Rural
Use-Mail-Order	14	33	31
Don't-Use-Mail-Order	26	37	9

Test the hypothesis that there is no association between the area people live in and their use of mail-order companies using a 1 per cent level of significance.

Q 7	Waiting times in a post office queue have a Poisson distribution with parameter α. Ten people had waiting times (in minutes) of: 1.6 , 0.9, 1.1, 2.1, 0.7, 1.5, 2.3, 1.7, 3.0, 3.4 A further six people had waiting times of more than 6 minutes. Based on these data find the maximum likelihood estimate of α. where $f(x)=\left(\alpha^x e^{-\alpha}\right)/x!$.	10	CO4
Q 8	The means of two large samples of 1000 and 2000 items are 67.5 cms and 68.0cms respectively. Can the samples be regarded as drawn from the population with standard deviation 2.5 cms. Test at 5% level of significance. OR Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kg) as follows: 50 49 52 44 45 48 46 45 49 45 Test if the average packing can be taken to be 50 kg.	10	CO3

Q 9	A car manufacturer runs tests to investigate the fuel consumption of cars using a newly developed fuel additive. Sixteen cars of the same make and age are used, eight with the new additive and eight as controls. The results, in miles per gallon over a test track under regulated conditions, are as follows:
	
	Control 27.0 32.2 30.4 28.0 26.5 25.5 29.6 27.2
	
	Additive 31.4 29.9 33.2 34.4 32.0 28.7 26.1 30.3
	
	If μ_C is the mean number of miles per gallon achieved by cars in the control group, and μ_A is the mean number of miles per gallon achieved by cars in the group with fuel additive, test:
	    a. $H_0:\mu_A=\mu_C$ Vs $H_1:\mu_A-\mu_C>0$
	    b. $H_0:\mu_A-\mu_C=8$ Vs $H_1:\mu_A-\mu_C<8$
		10	CO3

SECTION-C
Answer all questions, each question carries 20 Marks)   2X 20M=40M

Q 10	A university wishes to analyse the performance of its students on a particular degree course. It records the scores obtained by a sample of 12 students at entry to the course, and the scores obtained in their final examinations by the same students. The results are as follows:	10+5+5	CO5

Student	A	B	C	D	E	F	G	H	I	J	K	L
Entrance exam score x (%)	86	53	71	60	62	79	66	84	90	55	58	72
Finals paper score y (%)	75	60	74	68	70	75	78	90	85	60	62	70

$$\sum x = 836 \quad \sum y = 867 \quad \sum x^2 = 60,016 \quad \sum y^2 = 63,603 \quad \sum (x-\bar{x})(y-\bar{y}) = 1,122$$

    a. Calculate the fitted linear regression equation of y on x.
    b. Calculate finals paper score corresponding to an individual entrance score of 53.
    c. Find the correlation coefficient and coefficient determination and interpret the results.

Q 11	95% confidence interval for a population mean, μ, is given as (18.985, 21.015). This confidence interval is based on a simple random sample of 36 observations. Calculate the sample mean and standard deviation. Assume that all conditions necessary for inference are satisfied. Use the t-distribution in any calculations.	20	CO1, CO4

OR

One important factor in selecting software for word processing and database management systems is the time required to learn how to use a particular system. In order to evaluate three database management systems, a firm devised a test to see how many training hours were needed for five of its word processing operators to become proficient in each of three systems.

System A	16	19	14	13	18	hours
System B	16	17	13	12	17	hours
System C	24	22	19	18	22	hours

Using a 5% significance level, is there any difference between the training time needed for the three systems?

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
End Semester Examination, December 2018

Subject: Applied Statistical Analysis   **Semester:VII**
Programme: B. Tech CSE+BFSI   **Code   : CSIB 225**
Max. Marks   : 100   **Duration   : 3 Hrs**
No. of page/s:4

Instructions:

SECTION A
Section: A (Answer all questions, each question carries 4 Marks)   5X 4M=20M

S. No.		Marks	CO

Q 1	Define Followings
	    a. Types of Measurement
	    b. Point and interval Estimation
		4	CO1

Q 2	Discuss the nominal, ordinal, interval and ration scale data.	4	CO1
Q 3	Nine members of staff were selected at random from those using hire cars. The number of days he/she left the car unused in the garage in the previous week was: 2 6 2 4 1 4 3 1 1 Calculate Mean, median, mode and standard deviation.	4	CO2
Q 4	An auditor claims that 10% of invoices for a certain company are incorrect. To test this claim a random sample of 200 invoices are checked and 24 are found to be incorrect. Test at the 1% significant level to see if the auditor's claim is supported by the sample evidence.	4	CO1,CO2
Q 5	A company claims that its LED bulbs will last at least 8,000 hours. You sample 100 bulbs and find that $\bar{X}$ =7,800 hours and s=800 hours. Should the company's claim be rejected? Test at $\alpha = 0.05$.	4	CO3
	SECTION B **Answer all questions, each question carries 10 Marks) 4X 10M=40M**		
Q 6	Claims (in £000s) on a particular policy have a distribution with PDF given by:	10	CO3

	$f(x) = 2cx\, e^{-cx^2}\ x > 0$ Seven of the last ten claims are given below: 1.05, 3.38, 3.26, 3.22, 2.71, 2.37, 1.85 The three remaining claims were known to be greater than £6,000. Find the maximum likelihood estimate of c.		
Q 7	The null hypothesis is that 20 per cent of the passengers go in first class, but management recognizes the possibility that this percentage could be more or less. A random sample of 400 passengers includes 70 passengers holding first class tickets. Can the null hypothesis be rejected at 10 per cent level of significance?	**10**	**CO3**
Q 8	At a certain date in a large city 400 out of a random sample of 500 men were found to be smokers. After the tax on tobacco had been heavily increased, another random sample of 600 men in the same city included 400 smokers. Was the observed decrease in the proportion of smokers significant? Test at 5 per cent level of significance.	**10**	**CO2+ CO3**
Q 9	The table given below shows the data obtained during outbreak of smallpox: <table><tr><td></td><td>Attacked</td><td>Non Attacked</td></tr><tr><td>Vaccinated</td><td>31</td><td>469</td></tr><tr><td>Non Vaccinated</td><td>185</td><td>1315</td></tr></table> Test the effectiveness of vaccination in preventing the attack from smallpox. Test your result at 5 per cent level of significance. OR Certain pesticide is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kg) as follows: 50 49 52 44 45 48 46 45 49 45 Test if the average packing can be taken to be 50 kg.	**10**	**CO3**
	SECTION-C **Answer all questions, each question carries 20 Marks) 2X 20M=40M**		
Q 10	Two types of calculators are compared to determine the correlation in terms of the time required to perform a certain common statistical calculation. Twelve students chosen at random are given drills with both calculators so that they are familiar with the operation of each type. Then the time they take to complete the calculation on each device is measured in seconds. The data are as follows:	**20**	**CO2+ CO4**

With Ball No A	271	282	257	248	262
With Ball No B	252	275	302	268	276
~~With Ball No C~~	~~260~~	~~255~~	~~239~~	~~246~~	~~266~~
With Ball No D	279	242	297	270	258

Student	1	2	3	4	5	6	7	8	9	10	11	12
Calculator A	23	18	29	22	33	20	17	25	27	30	25	27
Calculator B	19	18	24	23	31	22	16	23	24	26	24	28

OR

At 5% level of significance to test the null hypothesis that a professional bowler performs equally well with the four bowling balls, given the following results:

Q 11	A university wishes to analyze the performance of its students on a particular degree course. It records the scores obtained by a sample of 12 students at entry to the course, and the scores obtained in their final examinations by the same students. The results are as follows:	

Student	A	B	C	D	E	F	G	H	I	J
Entrance exam score x (%)	86	53	71	60	62	79	66	84	90	55
Finals paper score y (%)	75	60	74	68	70	75	78	90	85	60

$$\sum x = 836 \quad \sum y = 867 \quad \sum x^2 = 60,016 \quad \sum y^2 = 63,603 \quad \sum (x - \bar{x})(y - \bar{y})$$

10+5+ 5

CO5

a. Calculate the fitted linear regression equation of y on x.
b. Calculate finals paper score corresponding to an individual entrance score of 53.
c. Find the correlation coefficient and coefficient determination and interpret the results.