

Subject Code: R232143

KALLAM HARANADHAREDDY INSTITUTE OF TECHNOLOGY (AUTONOMOUS)**II B.Tech. I Semester (R23)****STATISTICAL METHODS FOR DATA SCIENCE****Common to CSE(DS) & CSE (AI&DS)**

Q. No.		Questions	CO	BTL	Total Marks																		
Unit-I																							
1.	i	<p>Following are the heads of income of Railways during 2004 and 2005.</p> <table><tr><td></td><td>2004 (in crores of rupees)</td><td>2005 (in crores of rupees)</td></tr><tr><td>Coaching</td><td>26</td><td>31</td></tr><tr><td>Goods</td><td>40</td><td>39</td></tr><tr><td>Others</td><td>4.5</td><td>3.5</td></tr></table> <p>Represent the above data by a bar chart.</p>		2004 (in crores of rupees)	2005 (in crores of rupees)	Coaching	26	31	Goods	40	39	Others	4.5	3.5	1	2	10M						
		2004 (in crores of rupees)	2005 (in crores of rupees)																				
Coaching	26	31																					
Goods	40	39																					
Others	4.5	3.5																					
ii	<p>120 students of a college were asked to opt for different experiences. The details of these options are as under.</p> <table><tr><td>Areas of work experience</td><td>No. of students</td></tr><tr><td>Photography</td><td>6</td></tr><tr><td>Clay modelling</td><td>30</td></tr><tr><td>Kitchen gardening</td><td>48</td></tr><tr><td>Doll making</td><td>12</td></tr><tr><td>Book binding</td><td>24</td></tr></table> <p>Represent the above data through a pie diagram.</p>	Areas of work experience	No. of students	Photography	6	Clay modelling	30	Kitchen gardening	48	Doll making	12	Book binding	24	1	2								
Areas of work experience	No. of students																						
Photography	6																						
Clay modelling	30																						
Kitchen gardening	48																						
Doll making	12																						
Book binding	24																						
2.	i	<p>The following dataset shows the weight and height of players on a basketball team.</p> <table><tr><td>Height (inches)</td><td>70</td><td>72</td><td>75</td><td>78</td><td>83</td><td>72</td><td>72</td><td>64</td></tr><tr><td>Weight (pounds)</td><td>180</td><td>190</td><td>200</td><td>240</td><td>218</td><td>200</td><td>172</td><td>170</td></tr></table> <p>Represent the above data by a scatterplot.</p>	Height (inches)	70	72	75	78	83	72	72	64	Weight (pounds)	180	190	200	240	218	200	172	170	1	2	10M
	Height (inches)	70	72	75	78	83	72	72	64														
Weight (pounds)	180	190	200	240	218	200	172	170															
ii	Write briefly on principal component analysis.	1	1																				
3.	i	<p>Is the function defined by</p> $f(x) = \begin{cases} 0, & x < 2 \\ \frac{1}{18}(2x + 3), & 2 \leq x \leq 4 \\ 0, & x > 4 \end{cases}$ <p>a probability distribution function. Find the probability that a variate having $f(x)$ as probability distribution will fall in the interval $2 \leq x \leq 3$.</p>	1	1	10M																		
	ii	<p>Find the mean and variance of a random variable having the probability function $f(x)$ defined by</p> $f(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$	1	1																			
4.	i	Suppose 2% of the people on the average are left handed. Find the probability of finding i) 3 or more left handed ii) none or one left handed.	1	1	10M																		

	ii	20% of items produced from a factory are defective. Find the probability that in a sample of 5 chosen at random i) one is defective ii) $P(1 < x < 4)$.	1	1																																
5.	i	Given that the mean height of students in a class is 158cms with standard deviation of 20cms. Find how many students heights lie between 150cms and 170cms if there are 100 students in the class.	1	1	10M																															
	ii	The distribution of life, in hours, of a bulb is known to be exponential with mean life of 600hours. What is the probability that i) it will not last for more than 500hours ii) it will last for 700hours?	1	1																																
Unit-II																																				
6.	i	In a study of an automobile insurance a random sample of 80 body repair costs had a mean of Rs.472.36 and the S.D. of Rs.62.35. If \bar{X} is used as a point estimate to the true average repair costs, with what confidence we can assert that the maximum error doesn't exceed Rs.10.	2	2	10M																															
	ii	Find 95% confidence limits for the mean of a normally distributed population from which the following sample 15, 17, 10, 18, 16, 9, 7, 11, 13, 14 was taken.	2	1																																
7.	i	A sample of 64 students have a mean weight of 70kgs. Can this be regarded as a sample from a population with mean weight 56kgs and standard deviation 25kgs.	2	2	10M																															
	ii	Samples of students were drawn from two universities and from their weights in kilograms, mean and standard deviations are calculated and shown below. Make a large sample test to test the significance of the difference between the means.	2	4																																
		<table><tr><td></td><td>Mean</td><td>Standard deviation</td><td>Size of the sample</td></tr><tr><td>University A</td><td>55</td><td>10</td><td>400</td></tr><tr><td>University B</td><td>57</td><td>15</td><td>100</td></tr></table>					Mean	Standard deviation	Size of the sample	University A	55	10	400	University B	57	15	100																			
		Mean			Standard deviation	Size of the sample																														
University A	55	10	400																																	
University B	57	15	100																																	
8.	i	Write the procedure for testing of hypothesis.	2	1	10M																															
	ii	<table><tr><td colspan="11">The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data</td></tr><tr><td>Item</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Life in 1000hrs</td><td>1.2</td><td>4.6</td><td>3.9</td><td>4.1</td><td>5.2</td><td>3.8</td><td>3.9</td><td>4.3</td><td>4.4</td><td>5.6</td></tr></table> <p>Can we accept the hypothesis that the average life time of bulbs is 4000hrs. Use a 0.05 level of significance.</p>	The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data											Item	1	2	3	4	5	6	7	8	9	10	Life in 1000hrs	1.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6	2
The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data																																				
Item	1	2	3	4	5	6	7	8	9	10																										
Life in 1000hrs	1.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6																										
9.	i	The mean life of a sample of 10 electric bulbs was found to be 1456hours with standard deviation of 423hours. A second sample of 17 bulbs chosen from a different batch showed a mean life of 1280hours with standard deviation of 398hours. Is there a significant difference between the means of two batches?	2	1	10M																															
	ii	A random sample of size 16 values from a normal population showed a mean of 53 and a sum of squares of deviations from the mean equals to 150. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% confidence limits of the mean of the population.	2	4																																

10.	i	To examine the hypothesis that the husbands are more intelligent than the wives, an investigator took a sample of 10 couples and administered them a test which measures the I.Q. The results are as follows:										2	4	10M	
		Husbands	117	105	97	105	123	109	86	78	103	107			
		Wives	106	98	87	104	116	95	90	69	108	85			
		Test the hypothesis with a reasonable test at the level of significance of 0.05.													

Course Advisor

BoS Chairman

II B. Tech I Semester (R23)
STATISTICAL METHODS FOR DATA SCIENCE
Common to CSE(DS) & CSE (AI&DS)
Short Questions

Q. No.	Questions									CO	BTL	Total Marks
Unit-I												
1	Define the terms i) primary data ii) secondary data.									1	1	2M
2	Define the terms i) frequency distribution ii) probability distribution.									1	1	2M
3	Construct the frequency distribution table for the following data 8,10,15,19,20,15,8,19,16,8,15,10.									1	1	2M
4	Construct the probability distribution table for the number of heads on tossing a coin 2 times.									1	1	2M
5	Find mean of the following data 40,50,55,78,58,60,73,35,43,48.									1	1	2M
6	Find k of the following probability distribution									1	1	2M
	Value of x	1	2	3	4	5	6	7	8			
	$P(X=x)$	$2k$	$4k$	$6k$	$8k$	$10k$	$12k$	$14k$	$4k$			
7	Find mean of the following probability distribution $f(x) = \begin{cases} \frac{1}{4}(x+1), & -1 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$									1	1	2M
8	If mean of the binomial distribution is 4 and variance is 2 then find p .									1	1	2M
9	If a random variable has a poisson distribution such that $P(1)=P(2)$, then find mean of the distribution.									1	1	2M
10	Find the value of $A(1.54) + A(0.27)$.									1	1	2M
Unit-II												
11	Write the i) 99% and ii) 90% confidence limits for population mean.									2	1	2M
12	Write the formulae of sample size for estimating i) population mean and ii) population proportion.									2	2	2M
13	Determine a 95% confidence interval for the mean of a normal distribution with variance 0.25, using a sample of $n=100$ values with mean 212.3.									2	1	2M
14	State central limit theorem.									2	1	2M
15	Find $Z_{\alpha/2}$, where $\alpha = 99\%$ and $\alpha = 96\%$									2	1	2M
16	Define the terms i) null hypothesis ii) alternative hypothesis.									2	1	2M
17	Define the terms i) type I error ii) type II error.									2	1	2M
18	Write the test statistics for the test of significance of i) single mean ii) equality of two means.									2	1	2M
19	Write the assumptions for Student's t -test.									2	1	2M
20	Determine i) $t_{0.01}$ with $\nu = 18$ ii) $t_{0.05}$ with $\nu = 12$.									2	4	2M