|  |  |
| --- | --- |
| LOGO.jpg | **GEETHANJALIINSTITUTEOFSCIENCE&TECHNOLOGY**  (**ANAUTONOMOUSINSTITUTION**)  **(Approved by AICTE ,New Delhi & Affiliated to JNTUA, Ananthapuramu) (Accredited by NAAC with “A” Grade, NBA (EEE,ECE &ME) &**  **ISO9001:2008CertifiedInstitution)** |
| **QUESTIONBANK(DESCRIPTIVE)**  **Subject Name: STATISTICAL METHODS FOR DATA SCIENCE**  **Course & Branch: B. Tech & CSE (DS)**  **Year & Semester: II B. Tech & II Semester**  **Regulation:RG23** | |

# UNIT-I

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S.No** | **Question** | **[BTLevel][CO][**  **Marks]** | | |
| **2MarksQuestions (Short)** | | | | |
| **1.** | What is the formula of Mean and Variance of a discrete random variable? | | L1, CO1, 2M | |
| **2.** | Write formula to calculate mean and variance of a continuous random variable. | | L1, CO1 , 2M | |
| **3.** | If X is a continuous random variable and Y=a X +b, Prove that E(Y)=a E(X)+b. | | L2 ,CO1, 2M | |
| **4.** | The mean and variance of a binomial distribution are 4 and 4/3 respectively. Find n, p, q | | L2, CO1, 2M | |
| **5.** | write conditions of Poisson distribution | | L1, CO1, 2M | |
| **6** | Write probability of density function normal distribution | | L1, CO1, 2M | |
| **7** | Write the characteristics of a good estimator | | L1, CO1, 2M | |
| **8** | Define unbiasedness | | L1, CO1, 2M | |
| **9** | Define sufficient | | L1, CO1, 2M | |
| **10** | Define efficient estimator | | L1, CO1, 2M | |
| **Descriptive Questions(Long)** | | | | |
| **1.** | Two dice are thrown. Let X assign to each point (a,b) is S the maximum of its numbers I .e, X(a ,b)=max(a, b). Find the probability distribution. X is a random variable with X(s)={1,2,3,4,5,6}. Also, find the mean and variance of the distribution | | | L3, CO1,10M |
| **2.** | A random variable X has the following probability function  X=x 1 2 3 4 5 6 7 8  P(X=x) K 2K 3 K 4K 5K 6K 7K 8K  Find the value of k and (i) Mean (ii) Variance (iii) P(2≤x≤5) | | | L3, CO1,10M |
| **3.** | If the probability density of a random variable is given by  f(x) = k for x ≥ 0   1. otherwise   i)Find the value of k ii) mean and variance of the variable | | | L3, CO1,10M |
| **4.** | If the probability density of a random variable is given by  f(x)= ,  = 0,otherwise  Find the value of k and the probabilities that a random variable having this probability density will take on a value (i) between 0.1 and 0.2 (ii) greater than 0.5. | | | L3, CO1,10M |
| **5.** | Out of 800 families with 4 children each, how many families would be expected to have (a) 2 boys and 2 girls (b) at least one boy (c) no girl (d) at most 2 girls? Assume equal probabilities for boys and girls | | | L3, CO2,10M |
| **6.** | a)As only 3 students came to attend the class today, find the probability for exactly 4 students to attend the classes tomorrow.  b)If x is poisson variate such that p(x = 0) = p (x=1 ) and using recurrence relation formula find the probabilities at x= 1,2,3,4,5 | | | L3, CO1,10M |
| **7** | 1. If x is normal variate with mean 30 and standard deviation with 5 . find the probabilities (i) 26 ≤ x ≤ 40 (ii) x ≥ 45 2. A sales tax has reported that the average sales of the 500 business that has deal during with a year 36,000 with standard deviation 10,000   Assuming that the sales in these distributed in normal distribution  Find (i) The number of business as sales which RS 40,000  (ii)the percentage of business sales of which are likely range between  RS 30,000 and RS 40, 000 | | | L3, CO1,10M |
| **8** | In a Normal distribution, 7% of the items are under 35 and 89% are under 63. Determine the mean and variance of the distribution | | | L3, CO1,10M |
| **9** | Show that the sample variance of a sample drawn from population given by the expression  the is an unbiased estimator of the population variance | | | L3, CO1,10M |
| **10** | Let x1, x2, x3,….. xn is a random sample taken from a population with probability density function  , .Find a sufficient estimator for | | | L2, CO1,10M |

**UNIT-II**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No**  **.** | **Question** | | **[BTLevel][CO][**  **Marks]** |
| **2Marks Questions (Short)** | | | |
| **1.** | Define estimation, estimate and estimator | L1,CO2,2M | |
| **2.** | List the properties of the Maximum Likelihood Estimator. | L1,CO2,2M | |
| **3.** | Define normal equations for two-variable relationships Y on X1, X2 in the method of least squares | L1,CO2,2M | |
| **4.** | Define the Asymptotic Maximum Likelihood Estimator. | L1,CO2,2M | |
| **5.** | What is the formula for the modified minimum chi-square method? | L1,CO2,2M | |
| **6** | What are the good properties an estimator | L1,CO2,2M | |
| **7** | Write the formula unbiased estimator variance. | L1,CO2,2M | |
| **8** | Define estimator | L1,CO2,2M | |
| **9** | Define sufficient | L1,CO2,2M | |
| **Descriptive Questions(Long)** | | | |
| **11.** | Describe the method of maximum likelihood estimation. | | L2,CO2, 10M |
| **12.** | In a watch repair shop, the service time in minutes is 14,17,27,18,12,8,22,13,19 and 12. Give a maximum likelihood estimate of mean service time with the assumption that the service time follows an exponential distribution with parameter λ. | | L3,CO2, 10M |
| **13.** | A simple random sample of size n is taken from the probability density function  〖f(x)= 2θ,x>0,θ>0 is an unknown parameter.  Calculate the estimator of θ by the method of moments. | | L3, CO2,10M |
| **14.** | Explain the method of moment’s estimation. | | L3, CO2,10M |
| **15** | Let x1, x2, x3... xn be a random sample from the discrete distribution  P(X1=1)=(2(1-θ))/(2-θ) , P(X2=2)=θ/(2-θ) , where θ ϵ(0,1) is unknown.  Find the estimator θ by the method of moments | | L3, CO2,10M |
| **16** | A training data set of 9 different values for mid semester (say x) and end semester (say y) values are given by   |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | X | 10 | 7 | 3 | 16 | 9 | 11 | 7 | 10 | 8 | | Y | 42 | 39 | 32 | 50 | 44 | 55 | 43 | 37 | 43 |   Assuming a linear relationship y o x . Estimate the parameters by the method of least squares | | L3, CO2,10M |
| **17.** | Past experiences shown the following result of productivity per hector with respective use of fertilizers and seeds. Assuming the linear relationship of y on ,. Estimate the parameter from the given data   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Fertilizer | 45 | 30 | 70 | 75 | 65 | 80 | | Seeds | 2 | 18 | 3 | 2.5 | 2 | 3 | | Productivity y | 2000 | 2100 | 1800 | 1900 | 2400 | 2500 | | | L3, CO2,10M |
| **18.** | Consider a random sample of size n from an exponential distribution with parameter θ having p.d.f 〖f(x,θ)=θe〗^(-θx),x>0. Calculate the modified minimum chi-square estimator of θ based on the partitions A\_1={x,0≤x≤0.5} , A\_2={x,x>0.5} | | L3, CO2,10M |

|  |  |  |
| --- | --- | --- |
|  | **UNIT-III** |  |
| **S.No**  **.** | **Question** | **[BTLevel][CO][**  **Marks]** |
| **2Marks Questions (Short)** | | |
| **1.** | Define confidence interval | L2,CO3,2M |
| **2.** | What is the confidence interval for a single proportion for large sample? | L2,CO3,2M |
| **3.** | A sample of size 10 was taken from a population S.D of sample is 0.03.Find maximum error with 99%  confidence. | L2,CO3,2M |
| **4.** | Define the confidence interval for a single mean for large samples. | L2,CO3,2M |
| **5.** | Define the confidence interval for the ratio of population variance. | L1,CO3,2M |
| **6.** | Write formula the for confidence formula for mean | L1,CO3,2M |
| **7.** | Write formula the for confidence formula for two proportion | L1,CO3,2M |
| **8** | Write formula the for confidence formula for two means | L1,CO3,2M |
| **9** | Write test statistics two sample proportion test | L1,CO3,2M |
| **Descriptive Questions(Long)** | | |
| **1.** | A small poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favour of a particular candidate. Find a) 95% b)98% c) 99.73% confidence limits for the proportion of all the voters in favour of the candidate | L3,CO3,10M |
| **2.** | In a random sample of 400 adults and 600 teenagers who watched a certain television program 100 adults and 300 teenagers indicated that they liked it, construct (i) 95% (ii) 99% confidence intervals for the difference in proportion of all adults and all teenagers who watched for program and liked it | L3, CO3,10M |
| **3.** | A random sample of 16 items from the normal population showed a mean of 53 and sum of squares of deviation from this mean is equal to 150. Obtain 95% and 99% confidence limits for the mean of the population | L3,CO4,10M |
| **4.** | From the lots of transparent polyester film sheets produced by two machines sample are taken from both the machines and thickness values of the film sheets in mill microns are measured. The results are given below.   |  |  |  |  | | --- | --- | --- | --- | | Machine | Sample Size | Mean | Sample Variance | | 1  2 | 10  12 | 115  112 | 25  9 |  1. Obtain 95% and 99% confidence intervals for the difference of population’s means assuming that the unknown population variances are equal.   Obtain 95% and 99% confidence intervals for the difference of population’s means assuming that the unknown population variances are not equal. | L3, CO3,10M |
| **5.** | The average monthly electricity consumption for a sample of 100 families is 1250 units Assuming the S.D of electric consumption of all families is 150 units. Construct a 95% confidence interval estimate of the actual mean electric consumption | L3, CO3,10M |
| **6.** | A random sample of size 100 is taken from a population with Given that the sample mean is Cosnstruct 95% confidence interval for the population mean  A random sample of size 81 is taken from a population with Given that the sample mean is Cosnstruct 98% confidence interval for the population mean | L3,CO3,10M |
| **7.** | Suppose we have two independent random sample  Sample 1 :  Sample 2 :  Find 98% confidence interval for the difference of means | L3,CO3,10M |
| **8.** | The nicotine in milligrams of two samples of tobacco was found to be as follows. Find the standard errors and confidence limits for the difference between the means at a 0.05 level of significance. Assuming unknown population variances are equal.   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Sample A | 24 | 27 | 26 | 23 | 25 |  | | Sample B | 29 | 30 | 30 | 31 | 24 | 36 | | L3,CO3,10M |
|  | **UNIT-IV** |  |
| **S.No** | **Question** | **[BTLevel][CO][**  **Marks]** |
| **2MarksQuestions (Short)** | | |
| **1** | Define null and alternative hypotheses. | L1,CO4,2M |
| **2** | Define two types of Errors | L1,CO3,2M |
| **3** | State Neyman Pearson lemma | L1,CO3,2M |
| **4** | Describe the likelihood ratio test and state its properties | L1,CO3,2M |
| **5** | Write the formula for Type1 and Type2 Errors | L1,CO3,2M |
| **Descriptive Questions(Long)** | | |
| **11.** | State and prove Neyman Pearson Lemma for testing a simple hypothesis against a single alternative | L1,CO4,10M |
| **12.** | Suppose,, , ..... be a sample of size n from normal population with mean and variance are unknown. Test by the method of likelyhood ratio test. | L2,CO4,10M |
| **13.** | Let P be the probability that a coin will fall head in a single toss in order to test  Against. The coin is tossed 5 times and is rejected if more than 3 heads are obtained. Find the probability of type 1 error and power of the test. | L3,CO4,10M |
| **14.** | Suppose X is a single observation from a population with p.d.f for 0<x<1. Find the test with best critical region , find the most powerful test with significance level for testing the simple null hypothesis against the simple alternative hypothesis. | L3,CO4,10M |
| **15.** | Suppose ,…… is a random sample from a normal distribution with mean µ and variance 16. Find the best critical region with a sample size of n=16 and a significance level α = 0.05 to test the simple null hypothesis µ=10 against a simple alternative hypothesis µ=15. | L2,CO4,10M |
| **16.** | Explain about Likelihood Ratio Test and properties | L2,CO4,10M |
| **17.** | Let ,, , ..... be a random sample of 25 from N(. Find uniformly most powerful critical region of size for testing against | L2,CO4,10M |
| **18.** | Let a random variable X has Normal distribution with mean and variance of 2. Derive the LRT for testing the null hypotheisis against at a significance level of | L2,CO4,10M |
|  | **UNIT-V** |  |
| **S.No** | **Question** | **[BTLevel][CO][**  **Marks]** |
| **2MarksQuestions (Short)** | | |
| **1** | If =47.5, µ = 42.1, S = 8.4 and n=25 then find the test statistic `t’ | L2,CO5,2M |
| **2** | Explain briefly the paired t- test. | L2,CO5,2M |
| **3** | Explain briefly the variance ratio test (or) F-test. | L2,CO5,2M |
| **4** | Write the test statistic of χ2- test. | L2,CO5,2M |
| **5** | Define the contingency table. | L2,CO5,2M |
| **6** | Write the formula cell frequency (2x2) contingency table | L2,CO5,2M |
| **7.** | Write the formula test statistics for single mean in small samples test | L2,CO5,2M |
| **8 .** | Write the formula test statistics for two means in small samples test | L2,CO5,2M |
| **9.** | Write the formula test statistics for paired t – test | L2,CO5,2M |
| **Descriptive Questions(Long)** | | |
| **1** | A random of sample of six steel beams has a mean compressive strength of 58,392 p.s.i (pounds per square inch) with a standard deviation of 648 p.s.i. Use this information and the level of significance α = 0.05 to  test whether the true average compressive strength of the steel from which this sample came is 58,000 p.s.i. Assume normality | L3,CO5,10M |
| **2** | A random sample of 10 boys had the following I.Q’s: 70, 120, 110, 101, 88, 83, 95, 98, 107 and 100.  (a) Do these data support the assumption of a population mean I.Q of 100?  (b) Find a reasonable range in which most of the I.Q, values of samples of 10 boys | L3,CO5,10M |
| **3** | Two horses A and B were tested according to the time (in seconds) to run a particular track with the following results   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | Horse A | 28 | 30 | 32 | 33 | 33 | 29 | 34 | | Horse B | 29 | 30 | 30 | 24 | 27 | 29 | - | | L2,CO5,10M |
| **4** | 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:   |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 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. | L2,CO5,10M |
| **5** | The memory capacity of 10 students was tested before and after training. State whether the training was effective or not from the following scores   |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Before Training | 12 | 14 | 11 | 8 | | 7 | 10 | 3 | 0 | 5 | 6 | 66 | | After Training | 15 | 16 | 10 | | 7 | 5 | 12 | 10 | 2 | 3 | 8 |  | | L2,CO5,10M |
| **6** | The blood pressure of 5 women before and after intake of a certain drug are given below:   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Before | 110 | 120 | 125 | 132 | 125 | | After | 120 | 118 | 125 | 136 | 121 |          Test whether there is a significant change in blood pressure at 1% significance level | L2,CO5,10M |
| **7** | Two independent samples of 8 and 7 items respectively had the following values of the variables   |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Sample I | 9 | 11 | 13 | 11 | 16 | 10 | 12 | 14 | | SampleII | 11 | 13 | 11 | 14 | 10 | 8 | 10 | — |        Do the estimates of the population variance differ significantly? | L2,CO5,10M |
| **8** | From the following data find whether there is any significant liking in the habit of taking soft drinks  among the categories of employees.   |  |  |  |  | | --- | --- | --- | --- | | **Employees** | | | | | **Soft Drinks** | **Clerks** | **Teachers** | **Officers** | | Pepsi | 10 | 25 | 65 | | Thumsup | 15 | 30 | 65 | | Fanta | 50 | 60 | 30 | | L2,CO5,10M |