SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY: PUTTUR

(AUTONOMOUS)

Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: STATISTICAL METHODS FOR DATA SCIENCE (**23HS0839**)

Branch: **CAD**

Year & Sem: II-B. Tech & II-Sem **Regulation:** R23

UNIT-I
BASIC CONCEPTS

<table>
<tr><td rowspan="5">1</td><td>a) Define types of random variables with illustrations.</td><td>[L2][CO1]</td><td>[2M]</td></tr>
<tr><td>b) State the properties of random variables.</td><td>[L1][CO1]</td><td>[2M]</td></tr>
<tr><td>c) The mean and variance of a binomial distribution are 4 and 4/3 respectively. Find n.p,q.</td><td>[L1][CO1]</td><td>[2M]</td></tr>
<tr><td>d) Define Population and Sample with illustrations</td><td>[L1][CO1]</td><td>[2M]</td></tr>
<tr><td>e) State the invariance property of Consistent estimator</td><td>[L1][CO1]</td><td>[2M]</td></tr>
<tr><td>2</td><td>A random variable X has the following probability function
<table>
<tr><td>X</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr>
<tr><td>P(x)</td><td>0</td><td>K</td><td>2K</td><td>2K</td><td>3K</td><td>K^2</td><td>$2K^2$</td><td>$7K^2+K$</td></tr>
</table>
Determine (i) K (ii) Evaluate P(X≥6) and P(0<X<5) (iii) if P(X≤K)>1/2, find the minimum value of K (iv) Mean.</td><td>[L5][CO1]</td><td>[10M]</td></tr>
<tr><td>3</td><td>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.</td><td>[L5][CO1]</td><td>[10M]</td></tr>
<tr><td>4</td><td>Probability density function of a random variable X is $f(x) = \begin{cases} \frac{1}{2} sin\, x\, , for 0 \leq x \leq \pi \\ 0, \quad elsewhere \end{cases}$. Find the mean, mode and median of the distribution and also find the probability between 0 and $^{\pi}/_{2}$.</td><td>[L1][CO1]</td><td>[10M]</td></tr>
<tr><td rowspan="2">5</td><td>a) Derive mean and variance of Binomial distribution.</td><td>[L3][CO1]</td><td>[5M]</td></tr>
<tr><td>b) Out of 800 families with 5 children each, how many would you expect to have (i) 3 boys (ii) 5 girls (iii) either 2 or 3 boys. Assume equal probabilities for boys and girls.</td><td>[L3][CO1]</td><td>[5M]</td></tr>
<tr><td rowspan="2">6</td><td>a) Derive mean and variance of Poisson distribution.</td><td>[L5][CO1]</td><td>[5M]</td></tr>
<tr><td>b) If 2% of light bulbs are defective. Find the probability that (i) 2 defective items (ii) at least 3 defective items (iii) $P(2 < x < 5)$in a sample of 100.</td><td>[L1][CO1]</td><td>[5M]</td></tr>
<tr><td rowspan="2">7</td><td>a) If X is a Poisson variate such that $3P(X = 4) = \frac{1}{2}P(X = 2) + P(X = 0)$, find (i) Mean (ii) $P(X \leq 2)$.</td><td>[L1][CO1]</td><td>[5M]</td></tr>
<tr><td>b) If X is a normal variate with mean 30 and standard deviation 5. Find $P(26 \leq X \leq 40)$</td><td>[L1][CO1]</td><td>[5M]</td></tr>
<tr><td>8</td><td>In a competition of 1000 cases, the mean of a certain test is 14 and standard deviation 2.5. Assuming the distribution to be normal, find
(i) How many students score between 12 and 15
(ii) How many students score above 18
(iii) How many students score below 8.</td><td>[L1][CO1]</td><td>[10M]</td></tr>
<tr><td>9</td><td>In a Normal distribution, 7% of the items are under 35 and 89% are under 63. Determine the mean and variance of the distribution</td><td>[L5][CO1]</td><td>[10M]</td></tr>
<tr><td>10</td><td>Explain the characteristics of a good estimator</td><td>[L2][CO1]</td><td>[10M]</td></tr>
<tr><td>11</td><td>If $X_{1,}X_{2,}X_3$ be independent random variables from the population with mean μ and variance σ^2. Find which of the following estimators for μ are unbiased and which is most efficient estimator.
$T_1 = \frac{X1+X2+X3}{3}, T_1 = \frac{X1+2X2}{3}, T_1 = \frac{X1+2X2+3X3}{3}$</td><td>[L1][CO1]</td><td>[10M]</td></tr>
</table>

UNIT-II

POINT ESTIMATION

<table>
<tr><td rowspan="5">1</td><td>a) Define estimation, estimate and estimator</td><td>[L1][CO2]</td><td>[2M]</td></tr>
<tr><td>b) Define method of moments.</td><td>[L1][CO2]</td><td>[2M]</td></tr>
<tr><td>c) Define normal equations for two-variable relationships Y on X1, X2 in the method of least squares</td><td>[L1][CO2]</td><td>[2M]</td></tr>
<tr><td>d) What is the formula for the modified minimum chi-square method?</td><td>[L1][CO2]</td><td>[2M]</td></tr>
<tr><td>e) Define the Asymptotic Maximum Likelihood Estimator.</td><td></td><td></td></tr>
<tr><td rowspan="2">2</td><td>a) Describe the method of maximum likelihood estimation.</td><td>[L1][CO2]</td><td>[5M]</td></tr>
<tr><td>b) List the properties of maximum likelihood estimation.</td><td>[L1][CO2]</td><td>[5M]</td></tr>
<tr><td>3</td><td>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.</td><td>[L1][CO2]</td><td>[10M]</td></tr>
<tr><td>4</td><td>Suppose that the random sample has a normal distribution $N(\mu, \sigma^2)$.
Determine the Maximum Likelihood estimator
(i) for μ when $\sigma^2 = 1$. (ii) for σ^2 when μ=0 is known.</td><td>[L3][CO2]</td><td>[10M]</td></tr>
<tr><td rowspan="2">5</td><td>a) A simple random sample of size n is taken from the probability density function $f(x) = 2\theta x e^{-\theta x^2}, x > 0, \theta > 0$ is an unknown parameter. Calculate the estimator of θ by the method of moments.</td><td>[L3][CO2]</td><td>[5M]</td></tr>
<tr><td>b) Let $x_1, x_2, x_3, \dots x_n$ be a random sample from the discrete distribution $P(X_1 = 1) = \frac{2(1-\theta)}{2-\theta}$, $(X_2 = 2) = \frac{\theta}{2-\theta}$, where $\theta \epsilon (0,1)$ is unknown.
Find the estimator θ by the method of moments.</td><td>[L3][CO2]</td><td>[5M]</td></tr>
<tr><td rowspan="2">6</td><td>a) Let $x_1, x_2, x_3, \dots x_n$ be a random sample of size n from population given by
$$f(x) = \begin{cases} \frac{2(\theta - x)}{\theta^2}, & 0 < x < \theta \\ 0 & elsewhere \end{cases}$$
Find an estimator of θ by the method of moments.</td><td>[L3][CO2]</td><td>[6M]</td></tr>
<tr><td>b) Let $x_1, x_2, x_3, \dots x_n$ be a random sample of size n from a poisson population with parameter λ. Obtain the estimator of λ by the method of moments.</td><td>[L3][CO2]</td><td>[4M]</td></tr>
<tr><td>7</td><td>A training data set of 9 different values for mid semester (say x) and end semester (say y) values are given by
<table>
<tr><td>X</td><td>10</td><td>7</td><td>3</td><td>16</td><td>9</td><td>1</td><td>7</td><td>10</td><td>8</td></tr>
<tr><td>Y</td><td>42</td><td>39</td><td>32</td><td>50</td><td>44</td><td>55</td><td>43</td><td>37</td><td>43</td></tr>
</table>
Assuming a linear relationship y o x. Estimate the parameters by the method of least squares.</td><td>[L2][CO2]</td><td>[10M]</td></tr>
<tr><td>8</td><td>The following data represents the body weight (Y)kg, body length (X1)cm and body breadth (X1)cm of 12 randomly selected sea fish.
<table>
<tr><td>X1</td><td>12</td><td>20</td><td>14</td><td>25</td><td>18</td><td>16</td><td>10</td><td>18</td><td>18</td><td>20</td><td>16</td><td>12</td></tr>
<tr><td>X2</td><td>4</td><td>7</td><td>6</td><td>12</td><td>10</td><td>8</td><td>4</td><td>8</td><td>9</td><td>10</td><td>7</td><td>4</td></tr>
<tr><td>Y</td><td>0.5</td><td>0.8</td><td>0.7</td><td>2</td><td>1.2</td><td>0.9</td><td>0.4</td><td>0.9</td><td>1.4</td><td>1.5</td><td>0.8</td><td>0.6</td></tr>
</table>
Assume the linear model of Y, X1 and X2 as Y = a + bX1 + c X2. Estimate the parameters a, b & c by the method of least squares</td><td>[L2][CO2]</td><td>[10M]</td></tr>
<tr><td>9</td><td>If $p(x, \theta) = \theta^x (1-\theta)^{1-x}$ for random variable x=0, 1. Find the modified minimum chi-square estimator of θ based on the partitions $A_1 = \{x = 0\}$, $A_2 = \{x = 1\}$</td><td>[L3][CO2]</td><td>[10M]</td></tr>
<tr><td>10</td><td>Consider a random sample of size n from an exponential distribution with parameter θ having p.d.f $f(x, \theta) = \theta e^{-\theta x}, x > 0$. Calculate the modified minimum chi-square estimator of θ based on the partitions
$A_1 = \{x, 0 \le x \le 0.5\}$, $A_2 = \{x, x > 0.5\}$</td><td>[L3][CO2]</td><td>[10M]</td></tr>
</table>

11	Let $x_1, x_2, x_3, \dots x_n$ be a random sample of Bernoulli Distribution $f(x,p) = p^x(1-p)^{1-x}$ Show that the Bernoulli distribution is asymptotically normal.	**[L3][CO2]**	**[10M]**

UNIT-III

INTERVAL ESTIMATION

1	a) Define Interval Estimation and give one example	**[L1][CO3]**	**[2M]**
	b) Define confidence interval and confidence coefficient	**[L1][CO3]**	**[2M]**
	c) Find the lower, upper confidence limits and also confidence coefficient for P[0≤θ≤1.5] =0.90	**[L3][CO3]**	**[2M]**
	d) What are the confidence intervals for Population mean when population Variance is known and population variance is unknown	**[L1][CO4]**	**[2M]**
	e) Write the confidence interval for Population proportion population Variance is known and population variance is unknown	**[L1][CO4]**	**[2M]**
2	a) A random sample of 800 units from a large consignment showed that 200 were damaged. Construct 95% confidence limits for the population proportion of damaged units in the consignment.	**[L3][CO3]**	**[5M]**
	b) A factory is producing 50,000 pairs of shoes daily from a sample of 500 pairs 2% were found to be substandard quality. Estimate the no. of pairs that can be reasonably expected to be spoiled in the daily production and assign limits at 95% confidence level.	**[L2][CO3]**	**[5M]**
3	a) Out of 300 households in a town 123 have T.V sets, find 95% confidence limits to the true value of proportion of households with T.V sets in the whole town.	**[L3][CO3]**	**[5M]**
	b) Out of 20000 customers ledger accounts a sample of 600 accounts was taken to test the accuracy of posting and balancing where in 45 mistakes were found. Assign limits within which the number of defective cases can be expected at 5% level of significance.	**[L3][CO3]**	**[5M]**
4	A small poll of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favor of a particular candidate. Construct a) 95% b) 98% c) 99.73% confidence limits for the proportion of all the voters in favor of the candidate.	**[L3][CO3]**	**[10M]**
5	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 a) 95% b) Construct 99% confidence intervals for the difference in the proportion of all adults and all teenagers who watched for program and liked it.	**[L3][CO3]**	**[10M]**
6	a) In a certain newspaper, it was given that 5% of Canadians are illiterate and that 7% of Americans are illiterate. In a sample of 1005 Canadians and 1015 Americans, what should be the difference be at 95% level of significance.	**[L1][CO3]**	**[5M]**
	b) From a lot of units produced by Machine A, a sample of 500 is drawn and tested for a quality characteristic. It is found that 16 units are not meeting specification. Another sample of size 100 is drawn from a lot of similar units produced by Machine B and tested. In this case only 3 units are found to be not meeting the specification. Obtain 99% confidence limits for the difference of proportions of detective units produced by two Machines.	**[L1][CO3]**	**[5M]**
7	a) A random sample of size n=100 is taken from a population with σ=5.1. Given that the sample mean $\bar{x} = 21.6$. Construct 95% confidence interval for population mean μ.	**[L3][CO4]**	**[5M]**
	b) 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][CO4]**	**[5M]**

<table>
<tr><td>8</td><td>a) In an air pollution study, the following amounts of suspended benzene solvable organic matter(in micrograms per cubic meter) were obtained at an experiment station for eight different samples of air:2.2, 1.8, 3.1, 2.0, 2.4, 2.0, 2.1 and 1.2 .Construct α = 0.05 confidence interval for the corresponding true mean.</td><td>[L3][CO4]</td><td>[5M]</td></tr>
<tr><td></td><td>b) A sample of 7 boxes of a certain cereals with the following weights: 775, 780, 781, 795, 803, 810, 823. Find 95% confidence interval for variance.</td><td>[L3][CO4]</td><td>[5M]</td></tr>
<tr><td>9</td><td>a) Find the 95% confidence interval for the variance and S.D of the nicotine content of cigarettes manufactured. If a random sample of 20 cigarettes has a S.D of 1.6mg. Assume that the variance is normally distributed.</td><td>[L3][CO4]</td><td>[5M]</td></tr>
<tr><td></td><td>b) Find 90% confidence interval for the variance for the number of storms per year in the Atlantic basin. A random sample of 10 years has been used. They are: 10, 5, 12, 11, 13, 15, 19, 18, 14, 16. Assume the distribution is approximately normal.</td><td>[L3][CO4]</td><td>[5M]</td></tr>
<tr><td>10</td><td>From the lots of transparent polyester film sheets produced by two machines samples are taken from both the machines and thickness values of the film sheets in millimicrons are measured. The results are given below
<table>
<tr><th>Machine</th><th>Sample size</th><th>Mean</th><th>Sample variance</th></tr>
<tr><td>1</td><td>10</td><td>115</td><td>25</td></tr>
<tr><td>2</td><td>12</td><td>112</td><td>9</td></tr>
</table>
(i) Compute 95% confidence intervals for the difference of population means assuming that the unknown population variances are equal.
(ii) Compute 99% confidence intervals for the difference of population means assuming that the unknown population variances are not equal.</td><td>[L5][CO4]</td><td>[10M]</td></tr>
<tr><td>11</td><td>a) Given n1 = 15,n2 = 12,S1 = 3.07 and S2 = 0.80. Compute 98% confidence interval for the ratio of the population variances.</td><td>[L5][CO4]</td><td>[5M]</td></tr>
<tr><td></td><td>b) Among 11 patients in a certain study, the S.D of the property of interest was 5.8. In another group of 4 patients the S.D was 3.4. Construct a 95% confidence interval for the ratio of the variances of these two populations.</td><td>[L5][CO4]</td><td>[5M]</td></tr>
</table>

UNIT-IV

TESTING OF HYPOTHESES

1	a) Define Null hypothesis.	**[L1][CO5]**	**[2M]**
	b) Define Alternative hypothesis.	**[L1][CO5]**	**[2M]**
	c) What is Type-I error?	**[L1][CO5]**	**[2M]**
	d) What is Type-II error?	**[L1][CO5]**	**[2M]**
	e) Define the contingency table.	**[L1][CO5]**	**[2M]**
2	a) Explain about errors in sampling.	**[L2][CO5]**	**[5M]**
	b) Let p be the probability that a coin will fall head in a single toss in order to test $\mathrm{H_0: p = 0.5}$ against $\mathrm{H_1: p = 0.75}$ .The coin is tossed 5 times and $\mathrm{H_0}$ is rejected if more than 3 heads are obtained. Calculate the probability of type-l error and power of test.	**[L3][CO5]**	**[5M]**
3	Given the probability function f(x, ɵ)=$\begin{cases} \frac{1}{\theta}, 0 \leq x \leq \theta \\ 0\,, otherwise \end{cases}$ and test the hypothesis $H_0: \theta = 1$ against $H_1: \theta = 2$ by means of single observed value of x what would be the sizes of type-I and type-II errors. If you choose the interval (i)X ≥ 0.5 (ii)1 ≤ X ≤ 1.5 as the critical region? Also, Compute the power function of the test.	**[L3][CO5]**	**[10M]**

4	a) If $x \geq 1$ is the critical region for testing $H_0: \theta = 2$ against the alternative $H_1: \theta = 1$ on the basis of the single observation from the population $f(x, \theta) = \theta e^{-\theta x}, 0 \leq x < \infty$, Obtain the value of type-I and type-II errors.	**[L2][CO5]**	**[5M]**
	b) If $x \geq 0$ is the critical region for testing $H_0: \theta = 5$ against the alternative $H_1: \theta = 10$ on the basis of the single observation from the population $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, 0 \leq x < \infty$, Obtain the value of type-I and type-II errors.	**[L1][CO5]**	**[5M]**
5	State and prove Neyman-Pearson Fundamental Lemma.	**[L3][CO5]**	**[10M]**
6	Suppose $x_{1,}x_2,\ldots\ldots x_n$ 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 H_0: μ=10 against a simple alternative hypothesis H_1: μ=15.	**[L3][CO5]**	**[10M]**
7	a) Suppose X is a single observation from a population with p.d.f $f(x, \theta) = \theta x^{\theta-1}$for 0<x<1. Find the test with best critical region, find the most powerful test with significance level $\alpha = 0.05$ for testing the simple null hypothesis $H_0: \theta = 3$ against the simple alternative hypothesis$H_{01}: \theta = 2$	**[L3][CO5]**	**[5M]**
	b) Obtain best critical region for testing null hypothesis $H_0: \lambda = \lambda_0$against $H_1: \lambda = \lambda_1$ for an exponential distribution $f(x, \lambda) = \lambda e^{-\lambda x}$	**[L1][CO5]**	**[5M]**
8	Let $x_1, x_2, x_3, \ldots.. x_{25}$ be a random sample of 25 from $N(\theta, 100)$. Find uniformly most powerful critical region of size $\alpha = 10\%$ for testing $H_0: \theta = \theta_0 = 75$ against $H_1: \theta = \theta_1 > 75$	**[L3][CO5]**	**[10M]**
9	Explain about Likelihood Ratio Test.	**[L2][CO5]**	**[5M]**
	Write the properties Likelihood Ratio Test	**[L1][CO5]**	**[5M]**
10	Let $x_{1,}x_2,\ldots\ldots x_n$ be a random sample of size n from the normal population with mean μ and variance $\sigma^2$where μ and σ^2 are unknown. Test H_0: μ=μ_0 by the method of likelihood ratio test.	**[L4][CO5]**	**[10M]**
11	Let $x_{1,}x_2,\ldots\ldots x_n$be a random sample of n observations from $N(\mu, \sigma^2)$ we need to test the null hypothesis $H_0: \sigma^2 = {\sigma^2}_0$ against $H_1: \sigma^2 \neq {\sigma^2}_0$ by the method of likelihood ratio test	**[L4][CO5]**	**[10M]**

UNIT-V

SMALL SAMPLE TESTS

<table>
<tr><td>1</td><td>a) Define t-test for single mean.</td><td>[L1][CO6]</td><td>[2M]</td></tr>
<tr><td></td><td>b) Write the formula for F-test.</td><td>[L1][CO6]</td><td>[2M]</td></tr>
<tr><td></td><td>c) What is the Null hypothesis and Alternative hypothesis for F-test ?</td><td>[L1][CO6]</td><td>[2M]</td></tr>
<tr><td></td><td>d) What is the difference between F-test and t-tests?</td><td>[L1][CO6]</td><td>[2M]</td></tr>
<tr><td></td><td>e) Write the formula for Chi-square test for goodness of fit.</td><td>[L1][CO6]</td><td>[2M]</td></tr>
<tr><td>2</td><td>A sample of 26 bulbs gives a mean life of 990 hours with a S.D of 20 hours. The manufacturer claims that the mean life of bulbs is 1000 hours. Is the sample not up to the standard.</td><td>[L4][CO6]</td><td>[5M]</td></tr>
<tr><td></td><td>Find the maximum difference that we can expect with probability 0.95 between the mean of samples of sizes 10 and 12 from a normal population if their standard deviations are found to be 2 and 3 respectively.</td><td>[L1][CO6]</td><td>[5M]</td></tr>
<tr><td>3</td><td>Two independent samples of 8 and 7 items had the following values
<table>
<tr><td>Sample-I</td><td>11</td><td>11</td><td>13</td><td>11</td><td>15</td><td>9</td><td>12</td><td>14</td></tr>
<tr><td>Sample-II</td><td>9</td><td>11</td><td>10</td><td>13</td><td>9</td><td>8</td><td>10</td><td>-</td></tr>
</table>
Is the difference between the means of samples significant?</td><td>[L2][CO6]</td><td>[10M]</td></tr>
</table>

4 Samples of two types of electrical light blubs were tested for length of life and following data were obtained **[L2][CO6]** **[10M]**

	Type I	Type II
Sample numbers	8	7
Sample mean	1234 hrs	1036 hrs
Sample S.D	36 hrs	40 hrs

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life

5 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: **[L4][CO6]** **[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 significant of 0.05 and also calculate F-test.

6 Scores obtained in a shooting competition by 10 soldiers before and after intensive training are given below: **[L4][CO6]** **[10M]**

Before	67	24	57	55	63	54	56	68	33	43
After	70	38	58	58	56	67	68	75	42	38

Test whether the intensive training is useful at 0.05 level of significance.

7 Blood pressure of 5 women before and after intake of a certain drug are given below **[L4][CO6]** **[10M]**

Before	110	120	125	132	125
After	120	118	125	136	121

Test whether the significant change in blood pressure at 1% level of significance.

8 The nicotine in milligrams of two samples of tobacco were found to be as follows. **[L2][CO6]** **[10M]**

Sample A	24	27	26	21	25	---
Sample B	27	30	28	31	22	36

Can it be said that the two samples have come from the same normal population.

9 A pair of dice are thrown 360 times and the frequency of each sum is indicated below: **[L4][CO6]** **[10M]**

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	8	24	35	37	44	65	51	42	26	14	14

Would you say that the dice are fair on the basis of the chi-square test at 0.05 level of significant?

10 The following table gives the classification of 100 workers according to gender and nature of work. Test whether the nature of work is independent of the gender of the worker ($\psi^2 = 3.84$ at 1d.f) **[L4][CO6]** **[10M]**

	Stable	Unstable	Total
Males	40	20	60
Females	10	30	40
Total	50	50	100

11 A random sample of size 20 from a normal population gives a mean of 42 and a variance of 25. Test the hypothesis that the population S.D is 8 at 5% level of significance. **[L4][CO6]** **[10M]**