20IT401

Hall Ticket Number:

II/IV B.Tech. (Regular/Supplementary) DEGREE EXAMINATION

July, 2025 | **Information Technology**

Fourth Semester | **Probability And Statistics**

Time: Three Hours | **Maximum:** 70 Marks

Answer question 1 compulsorily. **(14X1 = 14 Marks)**

Answer one question from each unit. **(4X14 = 56 Marks)**

			CO	BL	M
1	a)	Define mean and variance of a probability density function f(x).	CO1	L1	1M
	b)	What is the probability density function of Normal distribution.	CO1	L1	1M
	c)	Find $z_{0.005}$	CO1	L2	1M
	d)	Find the mean and variance of gamma distribution with parameters α=2 and β=2.	CO1	L2	1M
	e)	Find the value of finite population correction factor for n =50 and N=1000.	CO2	L2	1M
	f)	Find the value of $F_{0.99}$ for 12 and 15 degrees of freedom.	CO2	L2	1M
	g)	What is the small sample confidence interval formula for the population parameter μ?	CO2	L1	1M
	h)	Write the test statistic for one mean in simple sample case.	CO2	L1	1M
	i)	What is (1-α)100% confidence interval for difference of two means in large sample case?	CO3	L1	1M
	j)	What is the critical region for testing one variance?	CO3	L1	1M
	k)	Find the value of $\chi^2_{0.975}$ at 19 degrees of freedom.	CO3	L2	1M
	l)	Define correlation.	CO4	L1	1M
	m)	Write normal equations to fit a parabola of the form $y = a+ bx+ cx^2$ by the method of least squares.	CO4	L2	1M
	n)	The range of the correlation coefficient is [-1,1]	CO4	L2	1M
		Unit-I			
2	a)	If a random variable has the probability density $f(x)=\begin{cases}2e^{-2x} & \text{for } x>0\\ 0 & \text{for } x\le 0\end{cases}$ Find the probabilities that it will take on a value (a). between 1 and 3 (b). greater than 0.5 (c). less than or equal to 1.	CO1	L3	7M
	b)	The time to microwave a bag of popcorn using the automatic setting can be treated as a random variable having a normal distribution with standard deviation 10 seconds. If the probability is 0.8212 that the bag will take less than 282.5 seconds to pop, find the probability that it will take longer than 258.3 seconds to pop.	CO1	L3	7M
		(OR)			
3	a)	The probability that an electric component will fail in less than 1000 hours of continuous use is 0.25. Use the normal approximation to find the probability that among 200 such components fewer than 45 will fail in less than 1000 hours of continuous use.	CO1	L3	7M
	b)	Derive mean and variance of uniform distribution.	CO1	L3	7M
		Unit-II			
4	a)	Hard disks for computers must spin evenly, and one departure from level is called roll. The roll for any disk can be modelled as a random variable having mean 0.2250 mm and standard deviation 0.0042 mm. The sample mean roll $\bar{x}$ will be obtained from a random sample of 40 disks. What is the probability that $\bar{x}$ will lie between 0.2245 and 0.2260 mm.	CO2	L3	7M
	b)	A random sample of 10 observations is taken from a normal population having the variance σ^2 = 42.5. Find the approximate probability of obtaining a sample standard deviation between 3.14 and 8.94.	CO2	L3	7M
		(OR)			
5	a)	In 64 randomly selected hours of production, the mean and the standard deviation of the number of acceptable pieces produced by a automatic stamping machine are $\bar{x}=1{,}038$ and $s=146$. At the 0.05 level of significance, does this enable us to reject the null	CO2	L3	7M

		hypothesis $\mu = 1{,}000$ against the alternative hypothesis $\mu > 1{,}000$.			
	b)	A manufacturer claims that the average tar content of a certain kind of cigarette is μ=14.0. In an attempt to show that it differs from this value, five measurements are made of the tar content (mg per cigarette): 14.5, 14.2, 14.4, 14.3, 14.6 Show that the differences between the mean of this sample $\bar{x} = 14.4$, and the average tar claimed by the manufacturer μ=14.0 is significant at α=0.05. Assume normality.	CO2	L3	7M
		Unit-III			
6	a)	The dynamic modulus of concrete is obtained for two different concrete mixes. For the first mix, n_1 = 33, $\bar{x}$ = 115.1 and s_1 = 0.47 psi. For the second mix, n_2 = 31, $\bar{y}$ = 114.6 and s_2 = 0.38. Test with α = 0.05, the null hypothesis of equality of mean dynamic modulus versus the two-sided alternative. Also construct a 95% confidence interval of the difference in mean dynamic modulus.	CO3	L3	14M
		(OR)			
7	a)	To compare two kinds of bumper guards, 6 of each kind, were mounted on a certain kind of compact car. Then each car was run into a concrete wall at 5 miles per hour and the following are the costs of the repairs (in rupees)			

Bumper Guard 1	407	448	423	465	402	419
Bumper Guard 2	434	415	412	451	433	429

		Use the 0.01 level of significance to test whether the difference between the two sample means is significant.	CO3	L3	7M
	b)	Playing 10 rounds of gold on his home course, a golf professional averaged 71.3 with a standard deviation of 1.32. Test the null hypothesis that the consistency of his game on his home course is actually measured by σ=1.20, against the alternative hypothesis that he is less consistent. Use the level of significance α=0.05.	CO3	L3	7M
		Unit-IV			
8	a)	The following are the weight losses of certain machine parts (in milligrams) due to friction when three different lubricants were used under controlled conditions:			

Lub A	12.2	11.8	13.1	11.0	3.9	4.1	10.9	8.4
Lub B	10.9	5.7	13.5	9.4	11.4	15.7	10.8	14.0
Lub C	12.7	19.9	13.6	11.7	18.3	14.3	22.8	20.4

		Test at the 0.01 level of significance whether the differences among the means can be attributed to chance. Also estimate the parameters of the model used in the analysis of experiment.	CO4	L3	14M
		(OR)			
9	a)	Fit a straight line of the form y = a+bx for the following data by the method of least squares:	CO4	L3	7M

x	20	60	100	140	180	220	260	300
y	0.18	0.37	0.35	0.78	0.56	0.75	1.18	1.36

	b)	Calculate the coefficient of correlation between x and y for the data	CO4	L3	7M

x	23	27	28	29	30	31	33	35	36	39
y	18	22	23	24	25	26	28	29	30	32

II/IV B.Tech. (Regular/Supplementary) DEGREE EXAMINATION	20IT401
July, 2025	**Information Technology**
Fourth Semester	**Probability And Statistics**

Scheme Of Valuation

<table>
<tr><th></th><th></th><th></th><th>M</th></tr>
<tr><td>1</td><td>a)</td><td>The mean of the probability density is $\mu = \int_{-\infty}^{\infty} x\ f(x)\,\mathrm{d}\,x$

And variance is $\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$</td><td>1M</td></tr>
<tr><td></td><td>b)</td><td>The Probability density function of Normal distribution is
$f\left(x;\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\ ; -\infty < x < \infty$</td><td>1M</td></tr>
<tr><td></td><td>c)</td><td>$z_{0.005}$ =2.575</td><td>1M</td></tr>
<tr><td></td><td>d)</td><td>Mean of gamma distribution is 4 and variance is 8</td><td>1M</td></tr>
<tr><td></td><td>e)</td><td>Population correction factor is $\frac{N-n}{N-1} = \frac{950}{999} = 0.9510$</td><td>1M</td></tr>
<tr><td></td><td>f)</td><td>$F_{0.99}$ (12,15)=1/4.01=0.2494</td><td>1M</td></tr>
<tr><td></td><td>g)</td><td>The small sample confidence interval formula for the population parameter μ is
$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$</td><td>1M</td></tr>
<tr><td></td><td>h)</td><td>The test statistic for one mean in simple sample case is, $t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$</td><td>1M</td></tr>
<tr><td></td><td>i)</td><td>(1-α)100% confidence interval for difference of two means in large sample case is
$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$</td><td>1M</td></tr>
<tr><td></td><td>j)</td><td>The critical region for testing $\sigma^2 = \sigma_0^2$
<table>
<tr><th>Alternative hypothesis</th><th>Reject the null hypothesis if</th></tr>
<tr><td>$\sigma^2 < \sigma_0^2$</td><td>$\chi^2 < \chi_{1-\alpha}^2$</td></tr>
<tr><td>$\sigma^2 > \sigma_0^2$</td><td>$\chi^2 > \chi_{\alpha}^2$</td></tr>
<tr><td>$\sigma^2 \neq \sigma_0^2$</td><td>$\chi^2 < \chi_{1-\alpha/2}^2$ or $\chi^2 > \chi_{\alpha/2}^2$</td></tr>
</table></td><td>1M</td></tr>
<tr><td></td><td>k)</td><td>The value of $\chi^2_{0.975}$ at 19 degrees of freedom=8.907</td><td>1M</td></tr>
<tr><td></td><td>l)</td><td>It is defined as when the changes in the values of one variable are associated with the changes in the values of the other variable is called correlation.</td><td>1M</td></tr>
<tr><td></td><td>m)</td><td>By the method of least squares, the normal equations to fit a parabola of the form
y = a+ bx+ cx^2 are
$\sum y = na + b\sum x + b\sum x^2$
$\sum xy = a\sum x + b\sum x^2 + b\sum x^3$
$\sum x^2 y = a\sum x^2 + b\sum x^3 + b\sum x^4$</td><td>1M</td></tr>
<tr><td></td><td>n)</td><td>The range of the correlation coefficient is [-1,1]</td><td>1M</td></tr>
</table>

2	a)	Given probability density function is $f(x)=\begin{cases}2e^{-2x} & \text{for } x>0\\ 0 & \text{for } x\le 0\end{cases}$ $P(1<X<3)=\int_1^3 f(x)\,\mathrm{d}x = \int_1^3 2e^{-2x}\,dx = 2\left(\frac{e^{-2x}}{-2}\right)_1^3 = e^{-2}-e^{-6}=0.1328$ (b). $P(X>0.5)=\int_{0.5}^{\infty} f(x)\,\mathrm{d}x = \int_{0.5}^{\infty} 2e^{-2x}\,dx = 2\left(\frac{e^{-2x}}{-2}\right)_{0.5}^{\infty} = e^{-1}$=0.3678 (c). $P(X\le 1)=\int_{-\infty}^{1} f(x)\,\mathrm{d}x = \int_{-\infty}^{0} f(x)\,\mathrm{d}x+\int_0^1 f(x)\,\mathrm{d}x = 0+\int_0^1 2e^{-2x}\,dx$ $= 2\left(\frac{e^{-2x}}{-2}\right)_0^1 = 1-e^{-2}=0.8646$	7M
	b)	Given the time to microwave a bag of popcorn using the automatic setting can be treated as a random variable having a normal distribution. Let it be X. Also given standard deviation, σ=10 seconds. And the probability that the bag will take less than 282.5 seconds to pop is 0.8212. *i.e,* $P(X\le 282.5)=0.8212$ $\Rightarrow P\left(\frac{X-\mu}{\sigma}\le\frac{282.5-\mu}{\sigma}\right)=0.8212$ $\Rightarrow P\left(Z\le\frac{282.5-\mu}{\sigma}\right)=0.8212$ $\Rightarrow F\left(\frac{282.5-\mu}{\sigma}\right)=0.8212=F(0.92)$ $\Rightarrow \frac{282.5-\mu}{\sigma}=0.92$ $\Rightarrow \mu=273.3\,\text{sec}$ The probability that the bag will take longer than 258.3 seconds to pop is $P(X\ge 258.3)=P\left(\frac{X-\mu}{\sigma}\ge\frac{258.3-\mu}{\sigma}\right)$ $=P\left(Z\ge\frac{258.3-273.3}{10}\right)$ $=P(Z\ge -1.5)$ = 1-F(-1.5) = F(1.5) = 0.9332	7M
3	a)	Here the number of electric components that will fail in less than 1000 hours of continuous use is the random variable. Given n=200, p=0.25, q=1-p =1-0.25 =0.75 Mean, μ=np =50 Standard deviation, $\sigma=\sqrt{npq}=\sqrt{200(0.2)(0.75)}=\sqrt{37.5}=6.1237$ The probability that the component fewer than 45 will fail in less than 1000 hours of continuous use is $P(X<45)=P(X\le 44.5)$ $=P\left(\frac{X-\mu}{\sigma}\le\frac{44.5-\mu}{\sigma}\right)$ $=P\left(Z\le\frac{44.5-50}{6.1237}\right)$ $=P(Z\le -0.8981)$ = F(-0.90) = 0.1841	7M

	b)	For the uniform distribution, the probability density function is $f(x)=\begin{cases}\frac{1}{\beta-\alpha} & for\ \ \alpha<x<\beta \\ 0 & elsewhere\end{cases}$ Mean, $\mu=\int_{-\infty}^{\infty} x\, f(x)\, dx$ $=\int_{-\infty}^{\alpha} x\, f(x)\, dx+\int_{\alpha}^{\beta} x\, f(x)\, dx+\int_{\beta}^{\infty} x\, f(x)\, dx$ $=\int_{\alpha}^{\beta} x\left(\frac{1}{\beta-\alpha}\right) dx=\left(\frac{1}{\beta-\alpha}\right)\left(\frac{x^2}{2}\right)_{\alpha}^{\beta}$ $=\left(\frac{1}{\beta-\alpha}\right)\left(\frac{\beta^2-\alpha^2}{2}\right)=\frac{\alpha+\beta}{2}$ Variance, $\sigma^2=\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\, dx$ $=\int_{-\infty}^{\infty} x^2 f(x)\, dx-\mu^2$ $=\int_{\alpha}^{\beta} x^2\frac{1}{\beta-\alpha}\, dx-\mu^2$ $=\frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta}-\left(\frac{\beta+\alpha}{2}\right)^2$ $=\frac{\beta^3-\alpha^3}{3(\beta-\alpha)}-\left(\frac{\beta+\alpha}{2}\right)^2$ $=\frac{(\beta-\alpha)(\beta^2+\alpha^2+\alpha\beta)}{3(\beta-\alpha)}-\left(\frac{\beta+\alpha}{2}\right)^2$ $=\frac{(\beta^2+\alpha^2+\alpha\beta)}{3}-\frac{\beta^2+\alpha^2+2\alpha\beta}{4}$ $=\frac{\beta^2+\alpha^2-2\alpha\beta}{12}=\frac{(\beta-\alpha)^2}{12}$ Variance, $\sigma^2=\frac{(\beta-\alpha)^2}{12}$	7M
4	a)	Given μ = 0.2250, n = 40, σ = 0.0042 We have to find out P(0.2245 < $\overline{X}$ < 0.2260). We will use Central limit theorem. Now, P(0.2245 < $\overline{X}$ < 0.2260) = P($\frac{0.2245-0.2250}{0.0042/\sqrt{40}}$ < $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ < $\frac{0.2260-0.2250}{0.00425/\sqrt{40}}$) = P(-0.753 < Z < 1.506) = F(1.506) – F(-0.753) = F(1.506) – (1-F(0.753)) = F(1.506) – 1+F(0.753) (average of .75 and .76 is .755) = 0.9332 – 1 + 0.7734 = 0.7066	7M
	b)	Given n = 10, σ^2 = 42.5. We have to find out P(3.14 < S < 8.94)	7M

<table>
<tr><td></td><td></td><td>But $P(3.14^2 < S^2 < 8.94^2) = P(9(3.14)^2/ 42.5< ((n-1)S^2/\sigma^2)< 9(8.94)^2/42.5)$
$= P(2.0879 < \chi^2 < 16.9249)$
$= P(2.088 < \chi^2 < 16.925)$
$= P(\chi^2 > 2.088) - P(\chi^2 > 16.925)$
$= 0.95 – 0.01$ (from Table 5 with $\upsilon = 9$) $= 0.94$</td><td></td></tr>
<tr><td>5</td><td>a)</td><td>Null hypothesis, H_0: $\mu = 1000$
Alternative hypothesis, H_1: $\mu > 1000$
Level of Significance $\alpha = 0.05$
Sample size, n = 64
$\bar{x} = 1,038$
$s = 146$ (when $n \geq 30$ we will replace s by σ)
Test statistic, $Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$
$= \frac{1,038 - 1000}{146/\sqrt{64}} = 2.08$
$Z_{0.05} = 1.645$
Since $Z = 2.08$ is positive and is greater than 1.645, the null hypothesis must be rejected at level $\alpha = 0.05$. Hence accept the alternative hypothesis.</td><td>7M</td></tr>
<tr><td></td><td>b)</td><td>Null hypothesis, H_0: $\mu = 14.0$ psi
Alternative hypothesis, H_1: $\mu \neq 14.0$ psi
Level of Significance, $\alpha = 0.05$
Sample size, n = 5
Sample mean $\bar{X} = 14.4$
Sample variance $s^2 = 0.025$
s = 0.1581
Test statistic, $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{14.4 - 14.0}{0.1581/\sqrt{5}} = 5.6574$
$t_{0.025}$ (4 d.f) = 2.776
Since $t = 5.6574$ is greater than 2.776, the null hypothesis must be rejected at level $\alpha = 0.05$. So, accept the alternative hypothesis.</td><td>7M</td></tr>
<tr><td>6</td><td>a)</td><td>Null Hypothesis, H_0 ; $\mu_1 - \mu_2 = 0$
Alternate Hypothesis, H_1: $\mu_1 - \mu_2 \neq 0$
Level of significance: $\alpha = 0.05$
$\bar{x} = 115.1$ $\quad \bar{y} = 114.6$
$s_1 = 0.47$ $\quad s_2 = 0.38$
$n_1 = 33$ $\quad n_2 = 31$
Test Statistic, $Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(115.1 - 114.6)}{\sqrt{\frac{0.47^2}{33} + \frac{0.38^2}{31}}} = 4.69$
$Z_{0.025} = 1.96$
Decision: Since Z = 4.69 > $Z_{0.025}$ = 1.96, we have to reject the null hypothesis. So, accept the alternate hypothesis
A 95% confidence interval of the difference in mean dynamic modulus is given by
$$(\bar{x} - \bar{y}) \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$\Rightarrow \quad (115.1 - 114.6) \pm z_{0.025}\sqrt{\frac{0.47^2}{33} + \frac{0.38^2}{31}}$$</td><td>14M</td></tr>
</table>

		$\Rightarrow \quad (115.1-114.6)\pm 1.96\sqrt{\frac{0.47^2}{33}+\frac{0.38^2}{31}}$ $\Rightarrow \quad 0.5 \pm 1.96*0.10655$ $\Rightarrow \quad 0.5 - 0.208838 < \mu1 - \mu2 < 0.5 + 0.208838$ $\Rightarrow \quad 0.291162 < \mu_1 - \mu_2 < 0.708838$	
7	a)	Null Hypothesis, H_0 ; $\mu_1 - \mu_2 = 0$ Alternate Hypothesis, H_1: $\mu_1 - \mu_2 \neq 0$ Level of significance: $\alpha = 0.01$ $\bar{x} = 427.33$ $\quad \bar{y} = 429$ $s_1^2 = 597.86$ $\quad s_2^2 = 202$ $n_1 = 6$ $\quad n_2 = 6$ Test Statistic, $t = \frac{(\bar{X}-\bar{Y})-\delta}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $where\ S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ $= \frac{(427.33-429)}{19.998\sqrt{\frac{1}{6}+\frac{1}{6}}}$ $= -0.1446$ $t_{0.005}$ 10 d.f = 3.169 **Decision: Since t = - 0.1446 > -$t_{0.005}$ = - 3.169, we have to accept the null hypothesis.**	7M
	b)	Null hypothesis: $\sigma = 1.20$ Alternative hypothesis: $\sigma > 1.20$ Level of significance $\alpha = 0.05$ n = 10, σ = 1.20 s = 1.32 Test statistic, $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ $= \frac{(10-1)(1.32)^2}{(1.20)^2}$ $= 10.89$ **Decision**: Since $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} = 10.89$ does not exceed 16.919, the value of $\chi^2_{0.05}$ for 9 degrees of freedom, the null hypothesis cannot be rejected. So, accept the null hypothesis.	7M
8	a)	Given k = 3 samples and the sizes are n_1 = 8, n_2 =8 and n_3 = 8 T_1 = 74.8 (the sum of all the values under Lub A) T_2 = 91.4 (the sum of all the values under Lub B) T_3 = 133.7 (the sum of all the values under Lub C) The grand total, $T\bullet = T_1 + T_2 + T_3 = 299.9$ $N = n_1 + n_2 + n_3 = 8 + 8 + 8 = 24$ $C = \frac{T\bullet^2}{N} = \frac{(299.9)^2}{24} = 3747.5$ $SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - C$ $= (12.2)^2 + (11.8)^2 + (13.1)^2 + \ldots + (22.8)^2 + (20.4)^2 - 3747.5$ $= 507.3896$	14M

$$SS(Tr)=\sum_{i=1}^{k}\frac{T_i^2}{n_i}-C$$

$= 230.5858$

SSE = SST – SS(Tr) = 276.8038

$F_{0.01}(3\text{-}1,24\text{-}3) = F_{0.01}(2,21) = 5.78$

Analysis of variance table:

Source of variation	Degrees of freedom	Sum of squares	Mean square	F
Treatments	k – 1 **2**	SS(Tr) **230.5858**	MS(Tr) = SS(Tr)/(k – 1) **115.2929**	MS(Tr)/MSE **8.7468**
Error	N - k **21**	SSE **276.8038**	MSE = SSE/(N – k) **13.1811**	
Total	N – 1 **23**	SST **507.3896**		

Decision: Since F = 8.7468 exceed the value of $F_{0.01}(2,21) = 5.78$, the null hypothesis must be rejected. So, accept the alternative hypothesis. That is the weight losses are not the same for the three lubricants.

9 a) By the method of least square, the normal equations to fit a straight line of the form y = a+ bx are

$$na+b\sum x=\sum y$$

$$a\sum x+b\sum x^2=\sum xy$$

Form the given data, n =8, $\sum x=1280$, $\sum x^2=2{,}72{,}000$ $\sum xy=1150.6$

Then y = 0.0584+0.0040x 7M

b) The coefficient of correlation

$$r=\frac{n\sum xy-\sum x\sum y}{\sqrt{\left(n\sum x^2-\left(\sum x\right)^2\right)\left(n\sum y^2-\left(\sum y\right)^2\right)}}$$

For the give data n = 10 ,

$\sum x=311, \sum y=257, \sum xy=8171, \sum x^2=9875, \sum y^2=6763$

The correlation coefficient r =0.9955 7M

N. Karunakar

Scheme Prepared by
N. Karunakar
Assistant Professor
Department of Mathematics
BEC, Bapatla
9247450048

Forwarded
23/7/2025