VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF ADVANCED SCIENCES

Winter Semester 2023-2024

Continuous Assessment Test –II

Programme Name & Branch : B.Tech (All)

Slot: B1+TB1-Common

Course Name & code: Probability and Statistics & BMAT202L

Class Number (s): VL2023240501675, VL2023240504939, VL2023240501741, VL2023240501679, VL2023240501681, VL2023240501677, VL2023240501674, VL2023240501739, VL2023240501683, VL2023240501685

Exam Duration: 90 Min. **Maximum Marks: 50**

General instruction(s): Answer ALL Questions

(Use of statistical table is allowed)

<table>
<tr><th>Q.No.</th><th>Question</th><th>Max Marks</th><th>CO</th><th>BL</th></tr>
<tr><td>1.</td><td>The following table gives the test scores made by ten sales man on an intelligent test and the values of their weekly sales:
<table>
<tr><td>Salesman</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td><td>I</td><td>J</td></tr>
<tr><td>Test scores</td><td>40</td><td>70</td><td>50</td><td>60</td><td>80</td><td>50</td><td>90</td><td>40</td><td>60</td><td>60</td></tr>
<tr><td>Sales (in thousands)</td><td>2.5</td><td>6.0</td><td>4.5</td><td>5.0</td><td>4.5</td><td>2.0</td><td>5.5</td><td>3.0</td><td>4.5</td><td>3.0</td></tr>
</table>
Obtain the regression equation of sales on test scores. Also estimate the most probable value of weekly sales for a salesman whose test score is 70.</td><td>10</td><td>CO</td><td>BL</td></tr>
</table>

Solution.

Test Scores x	Sales y	$\frac{x-50}{10}=u$	$\frac{y-4.5}{(0.5)}=v$	u^2	v^2	uv
40	2·5	−1	−4	1	16	4
70	6·0	+2	+3	4	9	6
50	4·5	0	0	0	0	0
60	5·0	+1	+1	1	1	1
80	4·5	3	0	9	0	0
50	2·0	0	−5	0	25	0
90	5·5	4	+2	16	4	0
40	3·0	−1	−3	1	9	3
60	4·5	+1	0	1	0	0
60	3·0	+1	−3	1	9	−3
Total		10	−9	34	73	19

Thus,
$$\Sigma u=10,\quad \Sigma v=-9,\quad \Sigma u^2=34,\quad \Sigma v^2=73,\quad \Sigma uv=19,\quad n=10$$

$$\therefore \quad \text{Cov.}(u,v)=\frac{19}{10}-\left(\frac{10}{10}\right)\left(\frac{-9}{10}\right)=2.8$$

$$\sigma_u^2=\frac{34}{10}-\left(\frac{10}{10}\right)^2=2.4$$

and
$$\sigma_v^2=\frac{73}{10}-\left(\frac{-9}{10}\right)^2=6.49$$

Thus, $\sigma_u=1.549$ and $\sigma_v=2.547$

$$\therefore \quad r=\frac{2.8}{(1.549)(2.547)}=0.71$$

$$\sigma_x=1.459\times 10=15.49$$

$$\sigma_y=2.547\times\frac{5}{10}=1.27$$

$$\bar{x}=50+10\times 1=60$$

$$\bar{y}=4.5+\frac{5}{10}(-0.9)=4.05$$

Thus, the regression of sales (y) on test score (x) is

$$Y_e-\bar{y}=r\,\frac{\sigma_y}{\sigma_x}(X-\bar{x}),$$

which reduces to

$$Y_e-4.05=(0.71)\left(\frac{1.27}{15.49}\right)(X-60)$$

That is, $Y_e=0.558+0.582\,X$

Hence, an estimate of weekly sales for a test score of 70 is

$$Y_e=0.558+(0.0582)(70)$$
$$=4.632 \text{ thousand rupees}$$
$$=\text{Rs. } 4632.$$

2.	The probability of a man scoring a penalty in a Hockey match is $\frac{1}{3}$. How many times he should be given a chance so that the probability of scoring successfully at least once is greater than $\frac{3}{4}$.	10	CO	BL

	Solution. Let p be the probability of scoring a penalty corner and n be the number of chances given. Given $p = \frac{1}{3}$ $\therefore \quad q = 1 - \frac{1}{3} = \frac{2}{3}$ Probability of scoring successfully at least once is $1 - q^n = 1 - \left(\frac{2}{3}\right)^n$ We find n such that $1 - \left(\frac{2}{3}\right)^n > \frac{3}{4}$ or $\quad \frac{1}{4} > \left(\frac{2}{3}\right)^n$ or $\quad \log \frac{1}{4} > n \log \frac{2}{3}$ $\log 1 - \log 4 > n(\log 2 - \log 3)$ $0 - 2\log 2 > n(\log 2 - \log 3)$ $-0{\cdot}6021 > n(0{\cdot}3010 - 0{\cdot}4771)$ $0{\cdot}6021 < n(0{\cdot}4771 - 0{\cdot}3010)$ $n > \frac{0{\cdot}6021}{0{\cdot}1761}$ $n > 3{\cdot}4 \cong 4.$			
3.	If the number of kilometers that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 km and if the owner desires to take a 5000 km trip, what is the probability that he will be able to complete the trip without having to replace the car battery? Assume that the car has been used for some time. The given problem can be solved using the exponential distribution formula. Let X be the number of miles the car can run before its battery wears out, and λ be the rate parameter of the exponential distribution. The average value of X is given as 10,000 miles, so we have: E(X) = 10,000 We know that the mean of an exponential distribution is equal to the inverse of the rate parameter. Therefore, we can write: E(X) = 1/λ Substituting the value of E(X), we get: 10,000 = 1/λ	10	CO	BL

<table>
<tr>
<td></td>
<td>

The probability that the car can run for 5000 miles without having to replace the battery is given by:

P(X > 5000) = e^(-λx)

Substituting the values of λ and x, we get:

P(X > 5000) = e^(-1/10,000 * 5000)
= e^(-0.5)
= 0.6065 (approx.)

Therefore, the probability that the person will be able to complete the 5000-mile trip without having to replace the car battery is 0.6065 (approx.).

- Ex 5d. Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles. If a person desires to take a 5000-mile trip, what is the probability that he or she will be able to complete the trip without having to replace the battery? What can be said when the distribution is not exponential?

  The remaining lifetime of the battery is exponential with parameter 1/10 (in thousands of miles).

$$P(\text{remaining lifetime} > 5) = 1 - F(5) = e^{-5\lambda} = e^{-1/2} \approx .604$$

If the lifetime distribution is not exponential,

$$P(\text{lifetime} > t+5 \mid \text{lifetime} > t) = \frac{P(\text{lifetime} > t+5, \text{lifetime} > t)}{P(\text{lifetime} > t)} = \frac{1-F(t+5)}{1-F(t)}$$

</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>

A cool drinks manufacturing company claims that its brand A cool drinks outsells its brand B by 8%. It is found that 42 out of a sample of 200 people prefer brand A and 18 out of another sample of 100 people prefer brand B. Test at 10% level of significance, whether the 8% difference is a valid claim.

Ans: The test statistic of Z is $|z| = 1.021$ whereas $z_\alpha = 1.645$

Hence, the 8% difference in the sale of two brands of products is a valid claim by the firm.

</td>
<td>10</td>
<td>CO</td>
<td>BL</td>
</tr>
<tr>
<td>5.</td>
<td>

Two samples drawn from two different populations gave the following results:

Sample	Size	Mean	SD
I	400	124	14
II	250	120	12

</td>
<td>10</td>
<td>CO</td>
<td>BL</td>
</tr>
</table>

	Test the significance of the difference between the means of the samples and also find the 99% confidence limits for the difference of the population means.			

Answer:

Null Hypothesis (H_0): There is no significant difference between the means of the two populations.

Alternative Hypothesis (H_1): There is a significant difference between the means of the two populations.

We'll use the formula for the independent samples t-test:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{X}_1$ and $\bar{X}_2$ are the sample means of samples I and II respectively.
- s_1 and s_2 are the sample standard deviations of samples I and II respectively.
- n_1 and n_2 are the sample sizes of samples I and II respectively.

$$t = \frac{124 - 120}{\sqrt{\frac{14^2}{400} + \frac{12^2}{250}}}$$
$$t = \frac{4}{\sqrt{\frac{196}{400} + \frac{144}{250}}}$$
$$t \approx \frac{4}{\sqrt{0.49 + 0.576}}$$
$$t \approx \frac{4}{\sqrt{1.066}}$$
$$t \approx \frac{4}{1.032}$$
$$t \approx 3.876$$

Next, we need to find the critical value of t for a 99% confidence level with degrees of freedom $df = n_1 + n_2 - 2 = 400 + 250 - 2 = 648$.

Using a t-table or statistical software, we find that $t_{\text{critical}} \approx \pm 2.609$ for a two-tailed test.

Since $t_{\text{calculated}}$ is greater than t_{critical}, we reject the null hypothesis.

Now, let's find the 99% confidence interval for the difference of the population means:

$$\text{Confidence Interval} = (\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\text{Confidence Interval} = (124 - 120) \pm 2.609 \times \sqrt{\frac{14^2}{400} + \frac{12^2}{250}}$$
$$\text{Confidence Interval} = 4 \pm 2.609 \times \sqrt{0.49 + 0.576}$$
$$\text{Confidence Interval} = 4 \pm 2.609 \times \sqrt{1.066}$$
$$\text{Confidence Interval} = 4 \pm 2.609 \times 1.032$$
$$\text{Confidence Interval} = 4 \pm 2.694$$

So, the 99% confidence interval for the difference of the population means is approximately $(1.306, 6.694)$.