VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF ADVANCED SCIENCES

Winter Semester 2022-2023

Continuous Assessment Test –II

Programme Name & Branch: B. Tech (Common)

Slot: B2+TB2

Course Code & Name: BMAT202L - Probability and Statistics

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

General instruction(s): Answer ALL Questions

<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 sale of a product in lakhs of rupees(Y) is expected to be influenced by two variables namely the advertising expenditure(X_1)(in thousands of rupees) and the number of sales persons(X_2) in a region. Sample data on 8 regions of a state has given the following results:
<table>
<tr><td>Area</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></tr>
<tr><td>Y</td><td>105</td><td>95</td><td>80</td><td>115</td><td>135</td><td>100</td><td>95</td><td>125</td></tr>
<tr><td>X_1</td><td>20</td><td>33</td><td>38</td><td>25</td><td>28</td><td>24</td><td>27</td><td>33</td></tr>
<tr><td>X_2</td><td>13</td><td>15</td><td>7</td><td>9</td><td>6</td><td>12</td><td>14</td><td>11</td></tr>
</table>
Find the regression model.</td><td>10</td><td>CO3</td><td>BL5</td></tr>
<tr><td>2.</td><td>Let an irregular 6-faced die is thrown and the possibility that in 10 throws it will give 5 even numbers is twice the expectation that it will give 4 even numbers. Out of 10,000 sets of 10 throws, find the number of times you should expect to receive (i) no even numbers, (ii) at least 15 even numbers and (iii) at most 3 even numbers.</td><td>10</td><td>CO2</td><td>BL3</td></tr>
<tr><td>3.</td><td>Assume that there were 200 students participated in an annual examination in a school. As a result, their respective average score and standard deviation are 32 and 13.
(i) How many candidates can be expected to obtain marks between 35 and 75 assuming the normality of the distribution and
(ii) determine the limit of the marks of the central 60 % of the candidates.</td><td>10</td><td>CO2</td><td>BL2</td></tr>
</table>

4.	A company wants to improve the quality of products by reducing defects and monitoring the efficiency of assembly lines. In assembly line A, there were 22 defects reported out of 400 samples while in line B, 33 defects out of 800 samples were noted. Is there a difference in the procedures at α=2% LOS?	10	CO4	BL3
5.	a) A sample of 200 students is taken from a large population. The mean height of the students in this sample is 135cm. Can it be reasonably regarded that this sample is from a population of mean 135 cm and S.D 8 cm? Also find the 95% fiducial limits for the mean. b) A genetic experiment involving peas yielded one sample of offspring consisting of 334 green peas and 143 yellow peas. Use a 0.01 significance level to test the claim that under the same circumstances, 18% of offspring peas will be yellow. Identify the null hypothesis, alternative hypothesis, test statistic, P-value, conclusion about the null hypothesis, and final conclusion that addresses the original claim. Use the P-value method.	(5+5)	CO4	BL4

1) Let $Y = b_0 + b_1X_1 + b_2X_2$, then the normal equation are.

SLOT-B2
P&S

$\Sigma Y = Nb_0 + b_1\Sigma X_1 + b_2\Sigma X_2$,

$\Sigma YX_1 = b_0\Sigma X_1 + b_1\Sigma X_2^2 + b_2\Sigma X_1X_2$,

$\Sigma YX_2 = b_0\Sigma X_2 + b_1\Sigma X_1X_2 + b_2\Sigma X_2^2$.

Then we get

$850 = 8b_0 + 228b_1 + 87b_2$

$24020 = 228b_0 + 6736b_1 + 2443b_2$

$9100 = 87b_0 + 2443b_1 + 1021b_2$

Solving the above equations, we have $b_0 = 169.852029$; $b_1 = -1.249166$ and $b_2 = -2.528809$.

Hence, the regression model is

$Y = 169.852029 - 1.249166\,\cancel{X_1}\,X_1 - 2.528809X_2$

2) Let X be the random variable that denote the number of even numbers.

$P(\text{getting } x \text{ even numbers}) = P(X=x)$

$= nC_x\, p^x q^{n-x}$, $x = 0,1,2,\dots,n$.

Given $n = 10$

$P(X=x) = 10C_x\, p^x q^{10-x}$.

P(getting 5 even numbers)

= 2 P(getting 4 even numbers)

$P(X=5) = 2P(X=4)$.

$\Rightarrow {}^{10}C_5\, p^5 q^5 = 2\left({}^{10}C_4\, p^4 q^6\right)$

$\Rightarrow \frac{p}{5} = \frac{q}{3} \Rightarrow 3p = 5q = 5(1-p)$

$\Rightarrow p = \frac{5}{8}. \Rightarrow q = \frac{3}{8}$

$\therefore P(X=x) = 10C_x\left(\frac{5}{8}\right)^x\left(\frac{3}{8}\right)^{10-x}$, $x = 0,1,\dots,n$.

$\therefore$ The required number of times that in 10,000 sets of 10 throws each we get

i) no even numbers

$= 10000 \times P(X=0) = 10000 \times {}^{10}C_0\left(\frac{5}{8}\right)^0\left(\frac{3}{8}\right)^{10}$

$= 0.5499367$

$= 1$ (appox).

ii) at least 15 even numbers

$= 1000\ P(X \geq 15) = 1000 \times [1 - P(X < 15)]$

$= 1000[1 - P(X \leq 14)] = 1000 \times 0$.

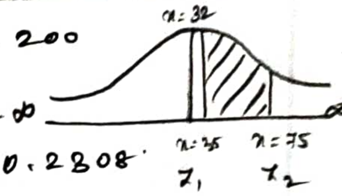
$= 0$.

iii) at most 3 even numbers.

$= 10000 \times P(X \leq 3) = 10000 \times 0.0383978$.

$\simeq 384$.

3) Given $\mu = 32$, $\sigma = 13$, $N = 200$



i) $P(35 \leq X \leq 75)$

$z_1 = \frac{x-\mu}{\sigma} = \frac{35-32}{13} = \frac{3}{13} = 0.2808$

$z_2 = \frac{75-32}{13} = \frac{43}{13} = 3.3077$.

$P(0.2308 \leq z \leq 3.3077)$

$= P(0 \leq z \leq 3.30769) - P(0 \leq z \leq 0.2308)$

$\simeq 4.9999 - 0.0910 \simeq 4.90899$.

$\therefore$ The number of students scoring between 35 and 75 ~~The~~

$= 4.90899 \times 200 \simeq 982$.

ii) Limit of central 60% of candidates:

Values of z_1 from the area label $= -0.84$.

Similarly for $z_2 = 0.84$.

$\therefore z_1 = \frac{X_1 - 32}{13} \Rightarrow -0.84 = \frac{X_1 - 32}{13} \Rightarrow X_1 = 21.08$

$z_2 = \frac{X_2 - 32}{13} \Rightarrow 0.84 = \frac{X_2 - 32}{13} \Rightarrow X_2 = 42.92$.

$\therefore$ 60% of the candidates score between 21.08 and 42.92

4) This is a two-tailed two proportion z-test.

H_0: The two proportions are the same.

H_1: The two proportions are not the same.

As this is a two-tailed test the alpha level needs to be divided by 2 to get 0.01.

Using this, the critical value from the z table is 2.33.

i.e., $Z_\alpha = 2.33$.

$n_1 = 400$; $n_2 = 800$

$p_1 = \frac{22}{400} = 0.055$

$p_2 = \frac{33}{800} = 0.0413$.

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$= \frac{22+33}{400+800}$$

$= 0.045833$.

$q = 1 - p = 0.954167$

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{0.01375}{\sqrt{0.045833 \times 0.954167 \left(\frac{1}{400}+\frac{1}{800}\right)}}$$

$$= \frac{0.01375}{0.01281}$$

$\simeq 1.07370$

As $2.33 > 1.0737$, the null hypothesis is accepted i.e., $z < z_\alpha$..

5) a) Given $n = 200$, $\bar{x} = 135$, $\sigma = 8$ and $\mu = 135$.

Null Hypothesis : $H_0 : \mu = 135$ ie., there is no difference between sample mean and population mean.

Alternative Hypothesis : $H_1 : \mu \neq 135$ (Two-tailed alternative)

The test statistics is given by

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{135-135}{8/\sqrt{200}} = 0 \text{ [calculated value].}$$

At 5% significance level the tabulated value for z_α is 1.96.

|Calculated value| $\leq$ Tabulated value then Accept H_0

$\Rightarrow 0 \leq 1.96$.

$\therefore$ We accept H_0.

95% fiducial limits (Confidence interval)

The confidence interval for the population mean is given by

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 135 \pm (1.96) \frac{8}{\sqrt{200}}$$

$$= 135 \pm 1.10874.$$

Thus, $x_1 = -133.89 \approx -134$. and $x_2 = 136.11 \approx 136$.

$\therefore -134 \leq P \leq 136$.

5) b) Given, The researcher wants to test the claim that 18% ($\tilde{p}$) of offspring peas will be yellow at the significance level 0.01.

* Total number of green peas (y) is 334.
* Total number of yellow peas (x) is 143.
* Total number of peas in the experiment (n) is $334 + 143 = 477$.

The estimated proportion of yellow peas (p) is given as $p = \frac{x}{n} = \frac{143}{477} = 0.29979$.

Let $\tilde{p}$ be the population proportion of yellow peas. Then the statistical hypothesis is, which implies that the test is two-tailed.

$H_0 : \tilde{p} = 0.18$ and $H_1 : \tilde{p} \neq 0.18$

Now the test statistic be. $z = \frac{p - \tilde{p}}{\sqrt{\frac{PQ}{n}}} = \frac{0.29979 - 0.18}{\sqrt{\frac{0.18 \times 0.82}{477}}} = \frac{0.11979}{\cancel{1.75907}\ 0.01758} \simeq \cancel{0.068098}\ 6.8098$

From the standard normal table,

$P\_value = 2P(Z > z) = 2P(Z > 0.068098) = 2(0.5 - P(Z < \cancel{0.068098}\ 6.8098))$

$= 2(0.5 - \cancel{0.0239}\ 0.5) = \cancel{0.9522}\ 0 < 0.01 > 0.01$

Since P-value is greater than significance level, H_0 is ~~accepted~~ rejected.