**GANPAT UNIVERSITY- FACULTY OF ENGINEERING & TECHNOLOGY**

**Programme:** Bachelor of Technology

**Branch:** Computer Engineering (Artificial Intelligence)

**Semester:** V

**Effective from Academic Year: 2024-25**

**Effective for the Batch admitted in:** July 2022

**Course Code:** 2CEAI502

**Course Name:** Computational Data Analytics

**Assignment** **Questions:** (Any -3 Questions)

Batch 1 | Q-1 to Q-8 have to attempt

Batch 2 | Q-9 to Q-16 have to attempt

Batch 3 | Q-17 to Q-24 have to attempt

Exercise:

1. What is Chi-square text? Explain its significance in statistical analysis.
2. Write short notes on the following:
   1. Additive property of Chi-square;
   2. Chi-square as a test of ‘goodness of fit’;
   3. Precautions in applying Chi-square test;
   4. Conditions for applying Chi-square test.
3. An experiment was conducted to test the efficacy of chloromycetin in checking typhoid. In a certain hospital chloromycetin was given to 285 out of the 392 patients suffering from typhoid. The number of typhoid cases were as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| *Typhoid* | | *No Typhoid* | *Total* |
| Chloromycetin | 35 | 250 | 285 |
| No chloromycetin | 50 | 57 | 107 |
| Total | 85 | 307 | 392 |

With the help of 2 , test the effectiveness of chloromycetin in checking typhoid. (The 2 value at 5 per cent level of significance for one degree of freedom is 3.841).

(*M. Com., Rajasthan University, 1966*)

1. On the basis of information given below about the treatment of 200 patients suffering from a disease, state whether the new treatment is comparatively superior to the conventional treatment.

|  |  |  |
| --- | --- | --- |
| *Treatment* | *No. of patients* | |
| *Favourable*  *Response* | *No Response* |
| New Conventional | 60  70 | 20  50 |

For drawing your inference, use the value of 2 for one degree of freedom at the 5 per cent level of significance, viz., 3.84.

1. 200 digits were chosen at random from a set of tables. The frequencies of the digits were:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Digit | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Frequency | 18 | 19 | 23 | 21 | 16 | 25 | 22 | 20 | 21 | 15 |

Calculate 2 *.*

1. Five dice were thrown 96 times and the number of times 4, 5, or 6 was thrown were Number of dice throwing

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 4, 5 or 6 | 5 | 4 | 3 | 2 | 1 |  | 0 |
| Frequency  Find the value of Chi-square. | 8 | 18 | 35 | 24 | 10 |  | 1 |

1. Find Chi-square from the following information:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Condition of child* | *Condition*  *Clean* | *of* | *home*  *Dirty* | *Total* |
| Clean | 70 | 50 | | 120 |
| Fairly clean | 80 | 20 | | 100 |
| Dirty | 35 | 45 | | 80 |
| Total | 185 | 115 | | 300 |

State whether the two attributes viz., condition of home and condition of child are independent (Use Chi-square test for the purpose).

1. In a certain cross the types represented by *XY*, *Xy*, *xY* and *xy* are expected to occur in a 9 : 5 : 4 : 2 ratio. The actual frequencies were:

|  |  |  |  |
| --- | --- | --- | --- |
| *XY* | *Xy* | *xY* | *xy* |
| 180 | 110 | 60 | 50 |

Test the goodness of fit of observation to theory.

1. The normal rate of infection for a certain disease in cattle is known to be 50 per cent. In an experiment with seven animals injected with a new vaccine it was found that none of the animals caught infection. Can the evidence be regarded as conclusive (at 1 per cent level of significance) to prove the value of the new vaccine?
2. Result of throwing die were recorded as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Number falling upwards | 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | 27 | 33 | 31 | 29 | 30 | 24 |

Is the die unbiased? Answer on the basis of Chi-square test.

1. The Theory predicts the proportion of beans, in the four groups A, B, C and D should be 9 : 3 : 3 : 1. In an experiment among 1600 beans, the number in the four groups were 882, 313, 287 and 118. Does the

experimental result support the theory? Apply 2 test.

1. You are given a sample of 150 observations classified by two attributes *A* and *B* as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *A*1 | | *A*2 | *A*3 | Total |
| *B*1 | 40 | 25 | 15 | 80 |
| *B*2 | 11 | 26 | 8 | 45 |
| *B*3 | 9 | 9 | 7 | 25 |
| Total | 60 | 60 | 30 | 150 |

Use the 2 test to examine whether *A* and *B* are associated.

1. A survey of 320 families with five children each revealed the following distribution:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No. of boys | 5 | 4 | 3 | 2 | 1 | 0 |
| No. of girls | 0 | 1 | 2 | 3 | 4 | 5 |
| No. of families | 14 | 56 | 110 | 88 | 40 | 12 |

Is this distribution consistent with the hypothesis that male and female births are equally probable? Apply Chi-square test.

1. What is Yates’ correction? Find the value of Chi-square applying Yates’ correction to the following data:

|  |  |  |  |
| --- | --- | --- | --- |
| *Passed* | | *Failed* | *Total* |
| Day classes | 10 | 20 | 30 |
| Evening classes | 4 | 66 | 70 |
| Total | 14 | 86 | 100 |

Also state whether the association, if any, between passing in the examination and studying in day classes is significant using Chi-square test.

1. (a) 1000 babies were born during a certain week in a city of which 600 were boys and 400 girls. Use 2

test to examine the correctness of the hypothesis that the sex-ratio is 1 : 1 in newly born babies.

(b) The percentage of smokers in a certain city was 90. A random sample of 100 persons was selected in which 85 persons were found to be smokers. Is the sample proportion significantly different from the proportion of smokers in the city? Answer on the basis of Chi-square test.

1. A college is running post-graduate classes in five subjects with equal number of students. The total number of absentees in these five classes is 75. Test the hypothesis that these classes are alike in absenteeism if the actual absentees in each are as follows:

History = 19

Philosophy = 18

Economics = 15

Commerce = 12

Chemistry = 11

1. The number of automobile accidents per week in a certain community were as follows:

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in agreement with the belief that accident conditions were the same during the 10 week period under consideration?

1. A certain chemical plant processes sea water to collect sodium chloride and magnesium. From scientific analysis, sea water is known to contain sodium chloride, magnesium and other elements in the ratio of 62 : 4 : 34. A sample of 200 tons of sea water has resulted in 130 tons of sodium chloride and 6 tons of magnesium. Are these data consistent with the scientific model at 5 per cent level of significance?
2. An oil company has explored three different areas for possible oil reserves. The results of the test were as given below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *A* | *Area*  *B* | *C* | *Total* |
| Strikes | 7 | 10 | 8 | 25 |
| Dry holes | 10 | 18 | 9 | 37 |
| Total number of test wells | 17 | 28 | 17 | 62 |

Do the three areas have the same potential, at the 10 per cent level of significance?

1. While conducting an air traffic study, a record was made of the number of aircraft arrivals, at a certain airport, during 250 half hour time intervals. The following tables gives the observed number of periods in which there were 0, 1, 2, 3, 4, or more arrivals as well as the expected number of such periods if arrivals per

half hour have a Poisson distribution = 2. Does this Poisson distribution describe the observed arrivals at 5 per cent level of significance.

|  |  |  |  |
| --- | --- | --- | --- |
| *Number of observed*  *arrivals (per half hour)* | *Number of periods*  *observed* | *Number of perio*  *expected (Poisson,* | *ds*  *= 2)* |
| 0 | 47 | 34 | |
| 1 | 56 | 68 | |
| 2 | 71 | 68 | |
| 3 | 44 | 45 | |
| 4 or more | 32 | 35 | |

1. A marketing researcher interested in the business publication reading habits of purchasing agents has assembled the following data:

**Business Publication Preferences (First Choice Mentions)**

|  |  |
| --- | --- |
| *Business Publication* | *Frequency of first choice* |
| *A* | 35 |
| *B* | 30 |
| *C* | 45 |
| *D* | 55 |

* 1. Test the null hypothesis ( = 0.05) that there are no differences among frequencies of first choice of tested publications.
  2. If the choice of *A* and *C* and that of *B* and *D* are aggregated, test the null hypothesis at = 0.05 that there are no differences.

1. A group of 150 College students were asked to indicate their most liked film star from among six different well known film actors viz., *A*, *B*, *C*, *D*, *E* and *F* in order to ascertain their relative popularity. The observed frequency data were as follows:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Actors* | *A* | *B* | *C* | *D* | *E* | *F* | Total |
| *Frequencies* | 24 | 20 | 32 | 25 | 28 | 21 | 150 |

Test at 5 per cent whether all actors are equally popular.

1. For the data in question 12, find the coefficient of contingency to measure the magnitude of relationship between *A* and *B*.
2. (a) What purpose is served by calculating the Phi coefficient ( )? Explain.

(b) If 2 = 16 and *N* = 4, find the value of Phi coefficient.