## MG221: Categorical Data Problems

**1.** Frequency distribution of the number of overseas projects completed by 80 programmers in a software company in the last calendar year is as follows:

No. of Projects	0	1	2	3	$\geq 4$
No. of Programmers	23	31	20	3	3

Can it be assumed that the number of overseas projects completed by a programmer in that company has a Poisson distribution?

**2.** Frequency of accidents faced by the assembly-line workers in a factory, classified according to the position of a worker in the assembly-line (Beginning, Middle or End) and the type of injury (Cut, Burn or Fracture) is as follows:

$\begin{array}{c} \text{Position} \rightarrow \\ \text{Type} \downarrow \end{array}$	Beginning	Middle	End
Cut	8	5	2
Burn	2	6	3
Fracture	2	3	5

Answer the following:

- **a.** Does the risk of an accident (irrespective of the type of injury) depend on the position of a worker in the assembly-line?
- **b.** Is there enough evidence suggesting the existence of an association between the type of injury faced by a worker and the worker's position in the assembly-line?

**3.** An FMCG company is interested in studying how the income level of consumers influence the choice of their purchase decision among four different brands of detergents that is currently available in a specific market. A random sample of 600 consumers covering the entire market yielded the following Cross Tabulation of Income versus Brand chosen (Figures in the cells represent number of consumers):

$\begin{array}{c} \text{Brands} \rightarrow \\ \text{Income} \downarrow \end{array}$	Brand1	Brand2	Brand3	Brand4	Total
Lower	25	15	55	65	160
Middle	30	25	35	30	120
Upper Middle	50	55	20	22	147
Upper	60	80	15	18	173
Total	165	175	125	135	600

Do the data provide any significant evidence of income level affecting the choice of brands of the consumers?

4. In three shifts in a manufacturing plant each producing 1000 units, the scrap rates are 4%, 2.5% and 4.5% respectively. Is there a significant difference in scrap rates among the three shifts?

5. Weekly number of break downs that had occurred in two soft drink bottling plants, one in Tambaram and the other one in Peenya, managed by the same company, last year, was as follows:

No. of Break Downs $\rightarrow$	0	1	2	3
Plant Location $\downarrow$	No.	of V	Veek	s↓
Tambaram	28	16	7	1
Peenya	14	24	11	3

Answer the following:

- a. Show that, based on point estimates, the Peenya Plant is more break down prone.
- **b.** Also show that this difference in the probability distributions of the weekly number of break downs between the two plants is statistically significant at 5% level.

6. Performances of two heuristic algorithms for solving a NP-hard optimization problem are to be compared. 1500 problem instances with known optima are chosen and both the algorithms are applied on them. It is found that algorithm 1 reached the optima 1000 times while algorithm 2 reached the optima 975 times, with 750 common cases, where both the algorithms have reached the optima. For i = 1, 2 let  $\pi_i = \text{Prob}(\text{Algorithm-}i)$  reaches the optima). Based on the above data, find a 95% confidence interval for  $\pi_1 - \pi_2$ , and use this confidence interval to decide whether there is any significant difference in performance of the two algorithms.

7. One way of measuring the effectiveness of an advertisement is its recall potential. An advertisement which is more likely to be remembered by a consumer is considered to be more effective. Two advertisements A and B for a certain product are under consideration for final launch in television. Both of these advertisements together with some others were shown to 50 volunteers and then they were again shown another set of advertisements the following day. Advertisements A and B were repeated on the second day as were a few others, while some of the advertisements were new on the second day. During the show on second day, each of the volunteers was asked whether they had seen the advertisement the previous day for each of the advertisements. The recall data of Advertisements A and B for these 50 volunteers are as follows:

No. of Volunteers Who Could	Recall A	Not Recall A
Recall B	15	20
Not Recall B	5	10

Is there sufficient evidence to conclude that the recall potential of one advertisement is superior to another?

8. The number of logical bugs found in the codes written by programmers A and B involved in a project are as follows:

Programmer A	0	0	0	0	0	1	1	1	1	1	1	1	1	2	2	2	2	2		
Programmer B	0	0	0	0	0	0	0	0	0	0	0	1	1	2	2	2	2	2	2	2

Is there a significant difference between the two distributions?

**9.** Number of PC's sold daily by 3 sales-persons on 20 randomly selected days sorted according to their values is as follows:

Sales-person	0	0	0	1	1	1	1	1	1	1
А	1	1	1	1	1	1	1	2	2	2
Sales-person	0	0	0	0	0	0	0	0	1	1
В	1	1	2	2	2	2	2	2	2	2
Sales-person	0	0	0	0	0	0	0	1	1	1
C	1	1	1	2	2	2	2	2	2	2

Answer the following:

- **a.** Is there a significant difference in the distributions of the number of PC's sold by 3 sales-persons?
- **b.** Is there a significant difference in the mean number of PC's sold by 3 sales-persons?
- c. Comment on the aptness of the above analysis and explain the appare nt anomaly, if you find any.

10. In a random sample of 150 married women in the age group of 25-39 years of age, it is found that 90 are homemakers. Among the remaining 60 working women 40 reported to have their dinner in a restaurant at least once a week. Of the 90 homemakers 45 reported to have their dinner in a restaurant at least once a week. Is there ample evidence to suggest that among the married women in the age group of 25-39 years of age, the working women are more likely to have their dinner in a restaurant at least once a week compared to the homemakers?

11. In a random sample of 60 managers in manufacturing industry, 25 have a degree in Engineering, 15 of whom feel it to be useful. Of the remaining 35 managers, 20 feel that Engineering knowledge would have been useful in their job. Is there a significant difference in perception about the utility of the knowledge of Engineering among the managers in manufacturing industry?

12. In a sample of 10 married men 6 reported that they were happy with their lives, while in a sample of 8 single men, who were otherwise similar to the married population with respect to their age, social status, occupation, income etc. only 4 reported that they were happy with their lives. From this observation would it be fair to say that in the stated group, in general, married men are happier than their single counterparts?

13. Among 12 scientists working in the R&D lab of an MNC, 4 are foreigners and the remaining 8 are Indians. The number of scientists holding at least one patent in his/her name is 3 each for both the groups of Indian and foreign scientists. Answer the following:

- **a.** Is there a significant difference between the two groups of scientists w.r.t. their patent holding pattern?
- b. For what kind of data (with the same marginal totals as above) you would have said that there is a significant difference in patent holding patterns of the two groups for  $\alpha = 0.1$ ?

14. The distribution of starting offers of the 9 students of the 2004-06 MBA batch of IISc along with the status of their work experience is as follows:

$\begin{array}{c} \text{Starting Offer} \rightarrow \\ \text{Work Experience } \downarrow \end{array}$	<7 l.p.a.	>7 l.p.a
Without	3	2
With	1	3

Is there any significant evidence that students with work-experience are more likely to obtain a starting offer of more than 7 l.p.a.?

15. A student from rural background in a small class of 15 students in BE first year found 6 others also from rural background, among whom 4 were his friends among his 6 friends in the class. Is there sufficient evidence to conclude that students from rural background are more likely to be his friends, than the ones from the urban background?