**Atal Bihari Vajpayee-**
**Indian Institute of Information Technology and Management Gwalior**

Multivariate Data Analysis

Major Examination

Time: 3 Hours MM: 40

**All the questions are compulsory**
**Calculators and Statistical tables are allowed**
**(NO EXTRA SHEET/COPY PLEASE)**

---

1. Two sample polls of votes for two candidates A and B for a public office are taken, one from among the residents of rural areas. The results are given in the table. Examine whether the nature of the area is related to voting preference in this election at 5% level of significance. [6]

| Votes for area | A | B | Total |
|---|---|---|---|
| Rural | 620 | 380 | 1000 |
| Urban | 550 | 450 | 1000 |
| Total | 1170 | 830 | 2000 |

2. Suppose that during 400 five-minute intervals the air-traffic control of an airport received 0, 1, 2, $\cdots$, or 13 radio messages with respective frequencies of 3, 15, 47, 76, 68, 74, 46, 39, 15, 9, 5, 2, 0, and 1. Suppose, furthermore, that we want to check whether these data substantiate the claim that the number of radio messages which they receive during a 5-minute interval may be looked upon as a random variable having the Poisson distribution. Test the goodness of fit at 5% level of significance. [6]

3. Consider the following data:

| $x$ | 0 | 1 | 2 | 2 | 4 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 25 | 20 | 30 | 40 | 45 | 50 | 60 | 50 |

If $x$ = percentage of the new component present in the metal and $y$ = cooling rate during a heat-treatment stage in $F^0$ per hour, find out if there is any linear relationship between them. Also, find out the error involved in the process. [7]

4. Suppose 3 drying formulas for curing a glue are studied and the following times observed.

| | | | | | | |
|---|---|---|---|---|---|---|
| Formula A: | 13 | 10 | 8 | 11 | 8 | |
| Formula B: | 13 | 11 | 14 | 14 | | |
| Formula C: | 4 | 1 | 3 | 4 | 2 | 4 |

Construct an analysis of variance table, and test the equality of mean weights with 5% level of significance. [7]

5. A compound is produced for a coating process. It is added to an otherwise fixed recipe and the coating process is completed. Adhesion is then measured. The following data concern the amount of adhesion and its relation to the amount of an additive and temperature of a reaction. [7]

| Additive $y$ | Temperature $x_1$ | Adhesion $x_2$ |
|---|---|---|
| 0 | 100 | 10 |
| 70 | 100 | 48 |
| 35 | 140 | 41 |
| 0 | 180 | 40 |
| 70 | 180 | 39 |
| 70 | 140 | 44 |
| 0 | 140 | 24 |
| 35 | 100 | 31 |
| 35 | 180 | 44 |

Fit an equation of the form $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ to the given data and use it to estimate the amount of adhesion when the amount of additive is 40 and the temperature is 130.

6. Let us consider the scenario and assume the following features of dimensions as $F_1$, $F_2$, $F_3$, and $F_4$. The data has been presented in the following table. [7]

| | | | | | |
|---|---|---|---|---|---|
| Large size apples ($F_1$) | 1 | 4 | 1 | 4 | 5 |
| Rotten apples ($F_2$) | 5 | 2 | 4 | 4 | 5 |
| Damaged apples ($F_3$) | 3 | 6 | 3 | 1 | 2 |
| Small apples ($F_4$) | 1 | 3 | 2 | 1 | 3 |

Calculate the Mean and Standard Deviation for each feature and then find out the most important feature by using the concept of principle component analysis.

BEST OF LUCK