

1. In order to study the length of time after billing for customers to settle the payments, a manager of a virtual bank took a random sample of 100 credit card invoices. A frequency table for the data on payment time, i.e., the number of days between the invoice date and the payment receipt date, was constructed as below.

| Payment time (in days) | Number of invoices |
|------------------------|--------------------|
| 0 – 9 | 7 |
| 10 – 19 | 20 |
| 20 – 29 | 48 |
| 30 – 39 | 19 |
| 40 – 59 | 6 |

- (a) Calculate the mean, mode, standard deviation and interquartile range. [8 marks]
- (b) Calculate the coefficient of skewness using results in (a) and interpret your result briefly. [2 marks]
- (c) In order to encourage early payment, the following scheme is introduced:
- For payments made before the mean payment time, the company refunds the customers for each invoice \$50 a day for the number of days before the mean payment time.
 - For payments made after the mean payment time, the company charges the customers for each invoice \$100 a day for the number of days after the mean payment time.

How much on average will the bank receive/spend per invoice on the above 100 invoices under this scheme? [6 marks]

- (d) Develop a 99% confidence interval for the mean payment time of all credit card invoices. [4 marks]

2. (a) In the annual recruitment exercise of a large commercial company, all applicants are required to go through two tests which include a written test and an oral interview in the screening stage. The failure rates for the written test and the oral interview are 85% and 80% respectively, and 75% applicants fail both tests. Only those applicants who pass both tests in the screening stage will be invited for a final interview. Offers of employment are made to 25% of those invited for the final interview.

i. Find the probability that a randomly selected applicant is invited for the final interview. [3 marks]

ii. Find the probability that a randomly selected applicant passes the written test or passes the oral interview but is not invited for the final interview. [2 marks]

iii. Find the probability that a randomly selected applicant passes the oral interview knowing that he/she has passed the written test. [2 marks]

iv. Find the probability that only one of ten randomly selected applicants is offered employment. [3 marks]

(b) There are three bags that each contains 100 marbles:

- Bag 1 has 75 red and 25 blue marbles;
- Bag 2 has 60 red and 40 blue marbles;
- Bag 3 has 45 red and 55 blue marbles.

One of the bags is selected at random and then pick a marble from the selected bag, also at random. Suppose that the chosen marble is red. What is the probability that Bag 1 was selected? [5 marks]

(c) From an ordinary deck of playing cards, cards are drawn successively at random and without replacement. What is the probability that the third heart appears on the seventh draw? [5 marks]

3. (a) The sales manager of a chain store studies the difference between the mean daily sales of two new stores. It is known that the population standard deviations of daily sales of the two new stores are \$500 and \$480 respectively. If he wants to produce a point estimate of the difference between the two population means with error not more than \$100 with 95% confidence, determine the required sample size when sample sizes are identical. [5 marks]

- (b) The number of customers visiting the website of an online shopping store has a Poisson distribution with a mean of 4.25 per hour.
- In a randomly selected hour, what is the probability that less than 3 customers visit the website? [3 marks]
 - In a randomly selected hour, what is the probability that 6 or more customers visit the website? [3 marks]
 - For five consecutive hours, what is the probability to have 2 consecutive hours with 6 or more customers and 3 other consecutive hours with less than 3 customers visit the website in each hour? [4 marks]
 - With suitable approximation, find the probability that the total number of customers visit the website for the next 48 consecutive hours is less than 200. [5 marks]
4. (a) A coffee dispensing machine is set to discharge on the average 245 ml of coffee into 250 ml cups. Suppose the amount of coffee dispensed is normally distributed with a standard deviation of 7 ml.
- Find the proportion of cups that will be overfilled. [3 marks]
 - Determine the size of the cup that should be used so that not more than 5% cups will overflow. [3 marks]
 - The coffee is sold at \$25 per cup. However, the dispenser owner promises \$5 refund if a cup contains less than 235 ml coffee. Suppose the cost of coffee is \$15 per cup, find the owner's expected profit per cup. [4 marks]
- (b) The Department of Environment did a study of automobile pollution at a particular location in the past ten years. The pollution measurements are found to be normally distributed with the variance of 10. It is also found that the mean pollution index was 112 during the months of April in previous years. During April this year the department took a random sample of seven measurements and the sample mean pollution index was 105. Test the hypothesis that the mean pollution index has been reduced at the 5% level of significance. [4 marks]
- (c) In a clinical test with 540 subjects, 63% showed improvement from the old treatment. While in a clinical test with another 630 subjects, 67% showed improvement from the new treatment. Test, at the 2% level of significance, on whether the new treatment performs better than the old treatment. State any assumption(s)/approximation(s) used. [6 marks]

5. (a) A statistician would like know if there is association between the number of parking fine notices a driver received and the age of the driver. The following table shows the data for the number of parking fine notices in a month and the age of the driver in a random sample of 300 drivers between 18 and 60.

| | | Age of drivers | | |
|-----------------------------|---|----------------|---------|---------|
| | | 18 – 30 | 31 – 45 | 46 – 60 |
| No. of parking fine notices | 0 | 51 | 39 | 35 |
| | 1 | 60 | 32 | 24 |
| | 2 | 22 | 12 | 25 |

Test, at the 5% level of significance, the hypothesis that there is no association between the number of parking fine notices and the driver's age. [7 marks]

- (b) A basketball game consists of four quarters of 10 minutes each. A random sample of 100 games of a certain team are investigated. In each game, the number of quarters that the team was in the lead by the end of the quarter are recorded below:

| Number of quarters in the lead | 0 | 1 | 2 | 3 | 4 |
|--------------------------------|---|----|----|----|---|
| Frequency | 8 | 17 | 41 | 30 | 4 |

Is there any evidence that the number of quarters the team is in the lead is binomially distributed, at the 5% level of significance? [7 marks]

- (c) Ten packets of milk chocolate were examined and the following net weight in gram were recorded:

103 104 93 101 105 95 101 102 109 102

Construct a 95% confidence interval for the mean net weight per packet. State any assumption(s)/approximation(s) used. [6 marks]

6. (a) A random sample of 6 houses is drawn. The area of the house (x , in m^2) and electricity consumption (y , in \$100) are recorded below:

| | | | | | | |
|-------------------|-----|-----|-----|-----|-----|----|
| x , (in m^2) | 120 | 100 | 140 | 110 | 120 | 90 |
| y , (in \$100) | 7 | 8 | 13 | 9 | 10 | 5 |

- Calculate the rank correlation coefficient, r_s , for the data. [3 marks]
 - Find the least squares regression line for predicting electricity consumption. [4 marks]
- (b) The admission office of a university would like to develop a regression model of student's first year GPA (y) on four admission test scores denoted by x_1, x_2, x_3 and x_4 , respectively. Twenty-five first year students are selected at random and a multiple regression equation is fitted with the following results:

ANOVA Table

| <i>Source</i> | <i>SS</i> | <i>df</i> | <i>MS</i> | <i>F</i> |
|---------------|-----------|-----------|-----------|----------|
| Regression | 5.6722 | b | d | f |
| Residual | a | c | e | |
| Total | 7.6453 | 24 | | |

| | <i>Coefficients</i> | <i>Standard Error</i> |
|-----------|---------------------|-----------------------|
| Intercept | -0.760594 | 0.641177 |
| x_1 | 0.001976 | 0.013542 |
| x_2 | 0.021241 | 0.009344 |
| x_3 | 0.017504 | 0.008856 |
| x_4 | 0.017965 | 0.008972 |

- Find the fitted multiple regression line from the above results. [1 mark]
- Calculate the missing values **a** to **f** in the ANOVA table. [3 marks]
- Calculate the adjusted coefficient of determination. [2 marks]
- Test the overall significance of the regression model at the 5% level of significance. [3 marks]
- The regression line will be further revised so as to exclude those independent variables which are not significant. Determine those independent variables, if there any, to be excluded at the 5% level of significance. [4 marks]