



**EASTERN UNIVERSITY, SRI LANKA**

**THIRD EXAMINATION IN SCIENCE 2004/2005**

**SECOND SEMESTER (Aug./Sep.'2007)**

**SPECIAL REPEAT EXAMINATION**

**MT 308 - STATISTICS**

Answer all questions

Time: Two hours

- (a) In order to estimate the mean length of leaves from a certain tree a sample of 100 leaves was chosen and their lengths measured correct to the nearest cm. A grouped frequency table was set up and the results were as follows:

|                        |     |     |     |     |     |     |     |     |     |
|------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Mid-interval value(cm) | 2.2 | 2.7 | 3.2 | 3.7 | 4.2 | 4.7 | 5.2 | 5.7 | 6.2 |
| Frequency              | 3   | 5   | 8   | 12  | 18  | 24  | 20  | 8   | 2   |

- Find the boundary values of each of the mid-interval value.
  - Draw the histogram and frequency polygon curve for the above data.
  - Calculate mean, median, mode and standard deviation.
  - Comment on the shape of the distribution.
- (b) Two cricketers scored the following runs in 10 innings. Find who is a better run-getter and if the consistency is the criterion for awarding a prize, who should get the prize.

|             |    |    |    |    |    |     |    |    |    |    |
|-------------|----|----|----|----|----|-----|----|----|----|----|
| Cricketer A | 42 | 17 | 83 | 59 | 72 | 76  | 64 | 45 | 40 | 32 |
| Cricketer B | 28 | 70 | 31 | 0  | 59 | 108 | 82 | 14 | 3  | 95 |

2. (a) If  $X'$  and  $Y'$  are the deviations of the random variables  $X$  and  $Y$  from their respective arithmetic means. Show that

$$\text{i. } r = 1 - \frac{1}{2N} \sum_{i=1}^N \left( \frac{X'_i}{\sigma_x} - \frac{Y'_i}{\sigma_y} \right)^2$$

$$\text{ii. } r = -1 + \frac{1}{2N} \sum_{i=1}^N \left( \frac{X'_i}{\sigma_x} + \frac{Y'_i}{\sigma_y} \right)^2$$

where  $r$  is the correlation coefficient of  $X$  and  $Y$ .

Deduce also that,  $r$  lies between  $-1$  and  $1$ .

- (b) If  $Z = aX + bY$  and  $r$  is the correlation coefficient between the two variables  $X$  and  $Y$  then show that,

$$\sigma_z^2 = a^2 \sigma_x^2 + b^2 \sigma_y^2 + 2abr \sigma_x \sigma_y,$$

where  $\sigma_x$  and  $\sigma_y$  are the standard deviations of  $X$  and  $Y$  respectively.

Deduce that

$$r = \frac{(\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2)}{2\sigma_x \sigma_y}$$

where  $\sigma_{x-y}$  is the standard deviation of  $X - Y$ .

- (c) Let  $x_1, x_2, \dots, x_n$  be the ranks of  $n$  individuals according to a characteristic  $A$  and  $y_1, y_2, \dots, y_n$  be the ranks of the same individuals according to other characteristic  $B$ . It is given that  $x_i + y_i = 1 + n$  for  $i = 1, 2, \dots, n$ . Show that the value of the rank correlation,  $r$ , between the two characteristics  $A$  and  $B$  is  $-1$ .

3. (a) Briefly explain the concept of correlation.

- (b) Show that Spearman's rank correlation coefficient  $r_s$  is given by

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

where  $n$  is the number of observations and  $d_i$  is the difference between rank assigned to the  $i^{\text{th}}$  individual.

- (c) Ten competitors in a beauty contest are ranked by three judges in the following order:

|              |   |   |   |   |   |    |    |   |   |    |
|--------------|---|---|---|---|---|----|----|---|---|----|
| Competitor   | 1 | 2 | 3 | 4 | 5 | 6  | 7  | 8 | 9 | 10 |
| First Judge  | 1 | 5 | 4 | 8 | 9 | 6  | 10 | 7 | 3 | 2  |
| Second Judge | 4 | 8 | 7 | 6 | 5 | 9  | 10 | 3 | 2 | 1  |
| Third Judge  | 6 | 7 | 8 | 1 | 5 | 10 | 9  | 2 | 3 | 4  |



A farmer wants to find the relationship between the amount of fertilizer used and the yield of corn. He selected several acres of his land on which he used different amount of fertilizer to grow corn. The following table gives the amount of fertilizer (in pounds) used and yield of corn (in bushels) for each of the seven acres.

|                           |     |     |     |    |     |     |     |
|---------------------------|-----|-----|-----|----|-----|-----|-----|
| Amount of fertilizer used | 120 | 80  | 100 | 70 | 88  | 75  | 110 |
| Yield of corn             | 138 | 112 | 129 | 96 | 119 | 104 | 134 |

- Draw a scatter diagram for these data. Does the scatter diagram show a linear relationship between fertilizer and yield of corn?
- Fit the estimated regression line, giving the statistical model for the data.
- Give a brief interpretation of the estimated slope calculated in part (b).
- Test at 5% significant level if the true slope is different from zero.
- Find a 95% confidence interval for the true slope.
- What is the estimated value of the yield of corn if the farmer used 125 pounds of fertilizer to grow?