# EASTERN UNIVERSITY, SRILANKA

## THIRD YEAR/ SECOND SEMESTER EXAMINATION IN SCIENCE (2002/03 & 2002/03 (A))

(April/May. 2004)

## MT 308 - STATISTICS

Answer all questions

Time allowed: Two hours

Q1. An insurance company wants to know how the amount of life insurance depends on income of persons. The research department at the company collected information on ten persons. The following table lists the annual income (in thousands of dollars) and amounts ( in thousands of dollars) of life insurance policies for these 10 persons.

| Annual income | Amount of Life Insurance |
|---|---|
| 47 | 250 |
| 54 | 300 |
| 25 | 100 |
| 37 | 150 |
| 62 | 500 |
| 18 | 75 |
| 39 | 160 |
| 70 | 400 |
| 29 | 95 |
| 48 | 264 |

(a) Construct a scatter diagram for these data. Does the scatter diagram show a linear relationship between annual income and amounts of life insurance policies?

(b) Fit the estimated regression line, with annual income as independent variable and amount of money as dependent variable. Interpret the slope of the regression line.

(c) Construct 95% confidence interval for the slope and test the hypothesis that the slope is different from zero at 5% significance level.

(d) Compute the coefficient of determination and compare your results given in (a)

(e) What is the estimated value of life insurance taken by a person with an annual income of $ 45,000?

Q2. a) If $X'$ and $Y'$ are the deviations of the random variables X and Y from their respective arithmetic means, Show that,

i) $r = 1 - \frac{1}{2N}\sum_{i=1}^{N}\left(\frac{X_i'}{\sigma_x} - \frac{Y_i'}{\sigma_y}\right)^2$ and

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,

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

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

Q3. a) Derive the equation that is used to calculate the Spearman's rank correlation coefficient with usual notations.

b) The following table shows the data on total costs in million Rs and output in million tons for a company over 10 time periods.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Cost: | 4.39 | 2.38 | 2.86 | 2.77 | 4.04 | 3.64 | 1.93 | 1.65 | 3.10 | 4.66 |
| Output: | 3.29 | 1.85 | 2.29 | 2.50 | 3.51 | 2.73 | 1.70 | 1.26 | 2.68 | 4.14 |

i) Compute the Pearson's correlation coefficient.
ii) Compute the Spearman's rank correlation coefficient.
iii) Comment your results on the basis of these two coefficients.



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

Q4. a) Show that the mean deviation from the mean and the standard deviation of the arithmetic progression $a, a+d, \ldots, a+2nd$ are

$$\frac{nd(n+1)}{2n+1} \quad \text{and} \quad d\sqrt{\frac{n(n+1)}{3}}$$ respectively. Verify that the latter is greater than the former.

b) A number of particular kind of small animals were classified according to their weights. After starvation of few weeks, the same animals have again been weighted and similarly classified. It is known that the median weight in the first weightment was 25.8 oz (ounce), while in the second weightment it was 19.8 oz. Some frequencies $f_1$ and $f_2$ in the first weightment and $f_3$ and $f_4$ in the second weightment are missing. It is known that $f_1 = \frac{1}{3} f_3$ and $f_2 = \frac{1}{4} f_4$. Find the missing frequencies.

| Class interval (oz) | Frequency ($1^{st}$ weightment) | Frequency ($2^{nd}$ weightment) |
|---|---|---|
| 0-6 | $f_1$ | $f_3$ |
| 6-12 | $f_2$ | $f_4$ |
| 12-18 | 11 | 40 |
| 18-24 | 52 | 65 |
| 24-30 | 75 | 28 |
| 30-36 | 23 | 13 |
| 36-42 | 14 | 6 |
| 42-48 | 5 | 2 |