EASTERN UNIVERSITY, SRI LANKA

THIRD EXAMINATION IN SCIENCE 2004/2005

SECOND SEMESTER (Oct./Nov.'2006)

MT 308 - STATISTICS

Answer all questions | Time : Two hours

1. (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

   i. Find the boundary values of each of the mid-interval value.

   ii. Draw the histogram and frequency polygon curve for the above data.

   iii. Calculate mean, median, mode and standard deviation.

   iv. Comment on the shape of the distribution.

(b) The daily expenditure of 100 families is given below.

Expenditure	0 − 20	20 − 40	40 − 60	60 − 80	80 − 100
No. of Families	13	?	27	?	16

If the mode of the distribution is 44,

   i. find the missing number of families in 20 − 40 and 60 − 80.

   ii. calculate the Karl-Pearson's coefficient of skewness.

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

$$\frac{nd(n+1)}{2n+1} \quad \text{and} \quad d\sqrt{\frac{n(n+1)}{3}} \quad \text{respectively.}$$

(b) Two persons participate in 5 shooting competitions and were able to hit the target correctly out of 15 shots are given below.

Competitor A	6	12	12	10	7
Competitor B	12	15	7	7	4

If the consistency is the criterion for awarding a prize, who should get the prize?

(c) Consider the simple regression model

$$Y_i = \alpha + \beta\, X_i + \epsilon_i, \qquad \epsilon_i \sim NID(0, \sigma^2), \quad i = 1, 2, \cdots, n.$$

Show that $F-$test of testing null hypothesis $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ is given by reject H_0 if $\frac{r^2(n-2)}{1-r^2} > F^1_{n-2,\ \alpha}$, where r is the correlation coefficient between Y_i and X_i and $F^1_{n-2,\ \alpha}$ is the upper $100(1-\alpha)$ percentage point of a F distribution with 1 and $n-2$ degrees of freedom.

3. (a) 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^{th} individual.

(b) Show that $-1 \leq r \leq 1$, where r is the correlation coefficient.

(c) The number of goals scored by football teams and their positions in the league were recorded as follows for the top 12 teams.

Team	A	B	C	D	E	F	G	H	I	J	K	L
Goals	49	44	43	36	40	39	29	21	28	30	33	26
League position	1	2	3	4	5	6	7	8	9	10	11	12

Calculate Spearman's rank correlation coefficient for these data. What conclusion can be drawn from this result?

(a) Using the least squares criterion, derive the equations that are used to estimate the slop and intercept of a simple linear regression line.

(b) Raw materials used in the production of a synthetic fiber is stored in a place which has no humidity control. Measurements of the relative humidity in the storage place and the moisture content of a sample of the raw material (both in percentages) on 12 days yielded the following results.

Moisture Content(Y)	Humidity (X)
12	43
8	35
14	51
9	47
11	46
16	62
7	32
9	36
12	41
10	39
13	53
11	48

i. Find the estimated regression line.

ii. Construct the analysis of variance table and test the hypothesis $H_0 : \beta = 0$ with $\alpha = 0.05$, where β is the slope parameter.

iii. Find 99% confidence interval for β.

iv. Predict the moisture content when the relative humidity is 38 percent.