EASTERN UNIVERSITY, SRI LANKA

THIRD EXAMINATION IN SCIENCE - 2005/2006

(Mar./Apr.' 2008)

SECOND SEMESTER

MT 308 - STATISTICS

(Proper and Repeat)

Answer all questions

Time : Two hours

(a) Show that Spearman's rank correlation coefficient r_s is given by

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

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

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

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 $x_1, x_2, \cdots, x_n$ be the ranks of n individuals according to a characteristic A and $y_1, y_2, \cdots, y_n$ be the ranks of the same individuals according to other characteristic

B. Obviously, $(x_1, x_2, \cdots, x_n)$ and $(y_1, y_2, \cdots, y_n)$ are permutations of $1, 2, \cdots, n$. It is given that $x_i + y_i = 1 + n$, $i = 1, 2, \cdots, n$. Show that the value of the rank correlation, r, between the two characteristic A and B is -1.

Q2. A farmer wanted 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 fertilizers to grow corn. The following table gives the amount of fertilizer (in pounds) used and the yield (in bushels) of corns for each of the seven acres.

Fertilizer used	**Yield of corn**
120	138
80	112
100	129
70	96
88	119
75	104
110	134

(a) Construct a scatter diagram for these data. Does the scatter diagram show a linear relationship between fertilizer used and yield of corn?

(b) Fit the estimated regression line, with fertilizer used as an independent variable and yield of corn as a dependent variable.

(c) Give a brief interpretation of the estimated slope, β_1, calculated in Part (b).

(d) Test at 5% significance level if the slope, β_1, is different from zero.

(e) Compute the coefficient of determination.

(f) Construct 95% confidence interval for the slope and intercept.

(g) What is the estimated value of the yield of corn if the farmer uses 125 pounds of fertilizer to grow?

(a) A number of particular kind of small animals were classified according to their weights. After starvation of two weeks the same animals have again been weighed and similarly classified. It is known that the median weight in the first weightment was 25.8 oz, 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 out the values of 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

(b) If n_1, n_2 are sizes; $\bar{x}_1$, $\bar{x}_2$ are the means and σ_1, σ_2 are the standard deviations of two series, then show that the standard deviation σ of the combined series of size $n_1 + n_2$ is given by

$$\sigma^2 = \frac{1}{n_1 + n_2}[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)],$$

where $d_i = \bar{x}_i - \bar{x}$, $i = 1, 2$ and $\bar{x}$ is the combined mean.

(c) The population of a country increased by 20% in the first decade, 30% in the second decade and 45% in the third decade. What is the average rate of increase per decade in the population?

Q4. (a) The U.S. Department of the Interior releases figures on mineral production. Following are the data of the 10 leading states nonfuel mineral production,

State	Value($ millions)
California	3,350
Nevada	2,800
Arizona	2,550
Texas	2,050
Florida	1,920
Michigan	1,670
Georgia	1,660
Minnesota	1,570
Utah	1,420
Missouri	1,320

i. Calculate the mean, median, and mode.

ii. Calculate the range, interquartile range, mean absolute deviation, sample variance and sample standard deviation.

iii. Compute the Pearson coefficient of skewness for these data.

iv. Sketch a box and whisker plot.

(b) Given the following results relating to two groups containing 20 and 30 observations; calcualte the coefficient of variation of all the 50 observations by combining both the groups.

	Group	
	I	II
n	20	30
$\sum X$	45	55
$\sum X^2$	118	132