University of Ruhuna
Faculty of Agriculture
End-semester Examination – 2010 September
Applied Statistics – I (CC-1202)

Answer all questions   Time: 02 ½ Hrs

----

1. (a) Discuss the followings
    (i) Coefficient of regression
    (ii) Coefficient of determination

(b) Briefly explain the ordinary least square (OLS) method and assumptions underline it.

(c) Why is the random error term (ε_i) included in the regression model.

(d) Identify independent and dependent variables in following cases.
    (i) Size of the house and monthly rent
    (ii) Final GPA of the students and their AL 'Z' Score

(e) Explain the meanings of SST (Total sums of squares), SSR (Regression sums of squares) and SSE (Error sums of squares). You may use graphs for illustrations.

(f) A survey was conducted to develop a model to explain the expenditure on food by the person's income. Observations are given below.

Income (X) (1000 rupees)	Food expenditure (Y) (1000 rupees)
35	9
49	15
21	7
39	11
15	5
28	8
25	9

$\sum X = 212$, $\sum Y = 64$, $\sum X^2 = 7222$, $\sum Y^2 = 646$, $\sum XY = 13568$

(g) Determine the regression model

(h) Calculate the food expenditure when the income is 26000/=

(i) Write the hypothesis (H_0, H_1) to test the significance of the regression coefficients

(j) Test the hypothesis and interpret the results

(k) Fill the following ANOVA table for the regression model

Source of Variation	df	Sum of Squares	Mean Squares	F

(l) Calculate the coefficient of determination (R) and interpret it.

(m) Prove that, Coefficient of determination = Coefficient of regression x (S_{xy} / S_{yy})

2. (a) Discuss the use of Pearson Correlation coefficient in data analysis

(b) Calculate the Pearson correlation coefficient for the following data set

weight of root nodules (g)	Length of roots (cm)
27	59
12	30
34	62
57	70
74	32
17	22
45	64
80	88

(c) Write the hypothesis to test the significance of correlation coefficient

(d) Test the hypothesis

(e) Interpret the value of correlation coefficient using the results of hypothesis testing.

3. (a) What are the basic properties of standard normal distribution?

(b) Thousand (1000) candidates who applied for a job appeared in a IQ test and the average marks were calculated as 53 and the standard deviation was 12.

I. How many students have obtained marks less than 40?

II. How many students have obtained more than 75?

III. How many students are in the group of 40 – 60?

IV. If the top 10% are to be called for the interview, what is the cut-off mark to select the students?

4. (a) The average paddy yield in Matara district is 4000 kg ha^{-1}. Randomly selected 20 farmers were given a new variety and their average yield was 4100 kg ha^{-1} and the standard deviation (s) was 40. Test whether the new variety has significantly improved the paddy yield using an appropriate test.

(b) A researcher wanted to test whether there is a significant difference of general knowledge of students in Badulla district and Monaragala district. Two randomly selected samples of class ten students from two districts were given a general knowledge paper and marks are summarized below.

	Badulla District	Monaragala district
Sample size	200	200
Average marks	67	63
Standard deviation	23	17

Test whether there is a significant difference of general knowledge between two districts using an appropriate statistical test.

(c) To reduce the late attendance of the workers, tea factories of Matara district introduced a new motivational package for the factory workers. Average number of late attendance per day before and after the motivational package in ten tea factories is given below. Test whether there is a significant impact of the motivational package.

Factory	A	B	C	D	E	F	G	H	I	J
Late attendance (Before)	23	12	8	31	14	3	5	14	16	9
Late attendance (After)	15	9	8	27	15	3	6	10	12	5

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \qquad\qquad SSR = \frac{S_{xy}^2}{S_{xx}}$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}} \qquad\qquad t = \frac{X_i - \bar{X}}{s/\sqrt{n}}$$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\left(\ldots\right)} \qquad\qquad t = \frac{\bar{d}}{s_d/\ldots}$$