



KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

INSTITUTIONAL BASED PROGRAMME (IBP) – DECEMBER 2011 SESSION
EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE IN SOCIAL
STATISTICS

SMA/SS/562: SURVEY DATA ANALYSIS

DATE: Wednesday, 2nd May, 2012

TIME: 9.00 a.m. – 12.00 p.m.

INSTRUCTIONS:

Answer question **ONE** and any other **TWO** questions.

Question 1 (30 marks)

- a) Explain the difference between following terms as applied in statistics:
- i) Multiple and partial correlation in a trivariate distribution. (2 marks)
 - ii) Type I error and Type II error (2 marks)
 - iii) One tailed test and two tailed test (2 marks)
- b) A product is manufactured in two ways. A pilot test on 64 items from each method indicates that the product of method 1 have a sample mean tensile strength of 106 units and a standard deviation of 12 units whereas in method 2 the corresponding values of mean and standard deviation are 100 units and 10 units respectively. Greater tensile strength in the product is preferable. Use the Z-test at 5 percent level of significance to test whether or not method 1 is better for processing the product. (5 marks)
- c) In a tri-variate distribution, $r_{12} = 0.87$, $r_{13} = 0.62$ and $r_{23} = 0.52$, find
- i) The partial correlation coefficient $r_{23.1}$ (3 marks)
 - ii) Multiple correlation coefficient $R_{1.23}$ (4 marks)

- d) A detergent manufacturing company claims that its top quality detergent is good, on average for at least 240 months before expiry. A consumer protection agency tested 50 such detergents and found that on average they last for 242 months with a standard deviation of 8 months. Construct 97% level of significance for the mean. (6 marks)
- e) A soft drink company is studying the effects of its latest advertising campaign. People chosen at random were called and asked how many cans of the soft drink they had bought in the past week and how many soft drink advertisements they had either read or seen in the past one week. The data gathered is shown in the table below

Number of Adverts, X	3	7	4	2	0	4	1	2
Cans Purchased, Y	11	18	9	4	7	6	3	8

Determine the coefficient of correlation using least squares method

(6 marks)

Question 2 (20 marks)

- a) Let N be the number of individuals which are ranked to two different characters A and B. Let x and y be the ranks with respect to A and B respectively. Assuming that the ranks are not repeated in either series, show that the Spearman's rank

correlation coefficient is given by $1 - \frac{6}{N(N^2 - 1)} \sum_{i=1}^N d_i^2$. (11 marks)

- b) Ten competitors in beauty contest are ranked by three judges as shown in the table below:

Judge 1	1	6	5	10	3	2	4	9	7	8
Judge 2	3	5	8	4	7	10	2	1	6	9
Judge 3	6	4	9	8	1	2	3	10	5	7

Using the method of rank correlation, determine which pair of judges have the nearest approach to common judgment (9 marks)

Question 3 (20 marks)

- a) Explain Five general steps for testing a hypothesis (5 marks)
- b) A country's Association of Retired Persons reported that typical senior citizens living at high altitudes claim their blood pressure is lower than the average of 160. To test this claim, 16 senior citizens living at high altitudes were selected at random and their blood pressures checked. Their mean pressure was 151; the standard deviation of the sample was 12. Do senior citizens living at high altitudes have significantly lower blood pressure as claimed? Use 0.05 level of significance.
- i) State the null and alternative hypothesis symbolically (2 marks)
- ii) Give the test statistic (1 mark)
- iii) Determine the degrees of freedom (1 mark)
- iv) Compute the test statistic (2 marks)
- v) Arrive at a conclusion (2 marks)
- c) The following data shows different colours preferred by males and females

Colour	Males	Females	Total
Red	10	40	50
White	70	30	100
Green	30	20	50
TOTAL	110	90	200

Determine whether there is any relationship between sex and preference of colour.

(7 marks)

Question 4 (20 marks)

- a) A student lives in a home with a solar electric system. At the same time each day, she collected voltage readings from a meter connected to the system and the results are listed below:

Sunny days	Cloudy days	Rainy days
13.5	12.7	12.1
13.0	12.5	12.2
13.2	12.6	12.3
13.9	12.7	11.9
13.8	13.0	11.6
14.0	13.0	12.2

At 0.05 significance level

- i) test the claim that the mean voltage reading is the same under the three different types of day. (6 marks)
 - ii) Using Analysis of Variance technique, can we conclude that sunny day result in greater amounts of electricity energy? (4 marks)
- b) Given the following set of data

Y	6	10	9	14	7	5
X_1	1	3	2	-2	3	6
X_2	3	-1	4	7	2	-4

- i) Calculate the multiple regression of Y and X_1 and X_2 (8 marks)
- ii) Predict Y when $X_1 = -1$ and $X_2 = 4$ (2 marks)