



KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2014/2015

DIGITAL SCHOOL OF VIRTUAL AND OPEN LEARNING

SECOND SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF EDUCATION SCIENCE AND BACHELOR OF SCIENCE

SMA 160: INTRODUCTION TO PROBABILITY AND STATISTICS I

DATE: Saturday 2nd May 2015

TIME: 11.00a.m-1.00p.m

INSTRUCTIONS

Attempt all parts of QUESTION 1 and any other TWO questions from the rest of the questions on the answer sheet given.

Question 1 (compulsory-30marks)

- a) Explain the four stages in statistics as defined by Croxton and Cowden (8marks)
- b) Describe any three functions of statistics. (6marks)
- c) Describe any two limitations of statistics. (4marks)
- d) Using appropriate examples differentiate between a population and a sample. (4marks)
- e) Identify the three axioms of probability (3marks)
- f) Define
 - i. Correlation analysis
 - ii. Regression analysis (2marks)
- g) Distinguish between nominal data and continuous data (2marks)
- h) Name any two branches of statistics (1mark)

Question 2 (20 marks)

Consider the following two sets of variables:

Y	10	7.5	5	2.5	0	-2.5	-5	-7.5	-10	12.5
X	0	5	10	15	20	25	30	35	40	45

- It is believed that there is a relationship between the two variables (Y and X). Present this data in a suitable diagram and discuss the type of relationship that exists. You may assume that the variable 'Y' is the dependent variable.
- Set up a suitable mathematical model to measure the linear strength of the relationship between the variables Y and X and use the model to comment upon the linear strength.
- Develop a suitable mathematical model to predict a value for the variable 'Y' given a value of the variable 'X'.
- Use the suitable mathematical model you have developed in part (iii) to forecast the value for the variable T when the variable 'X' is given a value of -5 (minus 5). How accurate do you think this result is? Justify your reasons for the accuracy.

(Hint for part (iv) consider your responses to part (ii) of this question as an aid).
- Suppose an additional observation of the 'Y' and 'X' variables are given as -15 (minus 15) and 50 respectively. Discuss the implications of your model with the inclusion of this extra information.

Question 3 (20 marks)

The following 40 observations of data represent the number of policy sales processed by an insurance company each day.

64	98	46	52	63	76	83	47	53	76
62	55	61	77	81	95	105	45	55	75
72	47	57	64	71	86	49	58	63	87
68	61	78	86	66	79	85	67	79	77

Visual inspection of the data suggest a maximum (minimum) value of 105 (45) sales. The variance between these two values is attributed to illnesses, weather conditions etc. affecting the promotion of the insurance policies.

The managing director delegates you to compile some basic statistics, to assist with decision making in the boardroom.

(for the 'raw' data)

- i) Find the mean, median and mode(s) and explain what they imply.
- ii) Identify (and explain) the five basic data measurements and use these to plot a Box and Whisker (Boxplot) diagram (to scale).
- iii) Organize these data into a frequency table with seven class intervals, commence with 41-50.

(for the 'grouped' data)

- iv) Add the following columns to your frequency table and complete them: Class midpoint (x), (x^2) , (fx) , $(f(x^2))$,
- v) Calculate the mean and the Standard Deviation of these data and explain what they imply.
- vi) Draw a histogram for these data and explain its implications.

Question 4 (20 marks)

The manager of the 'Sizzling Gastropod' bistro provides a delivery service for customers who telephone in an order. The manager would like to give callers an idea of the time it will take to deliver their order. This depends upon the distance of the customer from the bistro. The records of the last 10 deliveries are shown in the enclosed table:

Distance (miles)	Delivery Time (minutes)
2.3	5
6.7	13
7.5	10
3.1	5
4.6	9
3.9	8
8.7	15
9.8	20
10.1	18
6.5	13

- i) Plot these data on a graph with independent variable on the horizontal axis
- ii) Calculate the coefficient of correlation. Does the result obtained indicate that customer distance is a good predictor of delivery time?

- iii) Determine the regression equation which would enable the manager to estimate the time of delivery to the customer. Plot the line of the regression equation on the graph you drew in section (i).
- iv) If the delivery driver travels six miles to make a delivery, use your regression equation to determine the approximate time it will take to deliver the order. Comment upon the general predictive accuracy of your approximation.
- v) Consider the regression equation coefficients (a & b OR α & β). What do they tell us about the data? Do they supply any useful information?

Question 5 (20 marks)

The 120 employees at a factory were asked which would best improve their working life: better promotion prospects, higher pay or more respect from other staff. The results are tabulated below.

Response	Manual	Clerical	Managerial
Better promotion prospects:	12	12	3
Higher pay:	53	19	2
More respect:	7	6	6

- a) What is the probability that an employee selected more respect? (4marks)
- b) What is the probability that an employee is a clerical worker or selected better promotion prospects? (4marks)
- c) What is the probability that a manual employee selected higher pay? (4marks)
- d) What is the probability that an employee selected more respect and is a manager? (4marks)
- e) What is the probability that a managerial employee selected higher pay? (4 marks)