THE CATHOLIC UNIVERSITY OF EASTERN AFRICA

A. M. E. C. E. A

P.O. Box 62157
00200 Nairobi - KENYA
Telephone: 891601-6
Fax: 254-20-891084
E-mail:academics@cuea.edu

MAIN EXAMINATION

AUGUST - DECEMBER 2018 TRIMESTER

FACULTY OF COMMERCE

DEPARTMENT OF ACCOUNTING AND FINANCE

ODEL PROGRAMME

CID 082: STATISTICAL METHODS IN DATA ANALYSIS

Date: DECEMBER 2018	Duration: 2 Hours
INSTRUCTIONS: Answer Question ONE and ANY OTHER TWO Questions	

Q1. a) Briefly explain the difference between the following terms:
   i) Primary data and secondary data
   ii) Mailed questionnaires and questionnaires by enumerators
   iii) Telephone interview and direct interview.

   b) State the advantages and disadvantages of stratified random sampling.

   c) Briefly explain the steps you would take while selecting a stratified random sample.   **(15 marks)**

   d) The table below shows the duration (in minutes) of an old faithful's eruptions and the times (in Minutes) until the next eruption.

DURATION X	TIME Y
1.8	56
1.82	58
1.88	60
1.9	62
1.92	60
1.93	56
1.98	57
2.03	60
2.05	57

ISO 9001:2008 Certified by the Kenya Bureau of Standards

2.13	60
2.3	57
2.35	57
2.37	61
2.82	73
3.13	76
3.27	77
3.65	77
3.7	82
3.78	79
3.83	85
3.87	81
3.88	80
4.1	89
4.27	90
4.3	84
4.3	89
4.43	84
4.43	89
4.47	86
4.47	80
4.53	89
4.55	86
4.6	88
4.6	92
4.63	91

Required:-
   a) Find the equation of the regression line for predicting the time unit in the next eruption. **(7 marks)**
   b) Explain in this context what the slope of this line means. **(4 marks)**
   c) Predict the time unit the next eruption when the duration is 4.61 minutes. **(4 marks)**

Q2.   The contingency table below shows the results of a random sample survey. One variable is the age group of the respondent and the other variable is the response to the statement "Julia Gillard will make a better PM"

ISO 9001:2008 Certified by the Kenya Bureau of Standards

AGE GROUP	RESPONSE TO STATEMENT			TOTAL
	AGREE	NO VIEW	DISAGREE	
< 25	45	16	21	82
25 to 44	55	7	16	78
45 to 64	15	6	61	82
> 65	11	12	35	58
TOTAL	126	41	133	300

Required:-

a) What is the probability that a respondent disagrees with the statement?

(2 marks)

b) What is the probability that a respondent is in the 25 to 44 age group? **(3 marks)**

c) What is the probability that a respondent disagrees with the statement and is in the 25 to 44 age group? **(3 marks)**

d) What is the probability that a respondent disagrees with the statement or is in the 25 to 44 age group? **(3 marks)**

e) What is the probability that a respondent disagrees with the statement given that the respondent is in the 25 to 44 age group? **(3 marks)**

f) Which two answers from questions a) to e) would you use to determine whether "disagreeing with the statement" and "being in the 25 to 44 age group" are independent? **(3 marks)**

g) Are the events "disagreeing with the statement" and "being in the 25 to 44 age group" independent? **(3 marks)**

Q3. a) Some forty (40) people were asked about their preferences as far as the daily papers are A, B and D. It was noted that those who buy Newspaper A do not buy Newspaper D and vice versa. Six (6) of them were found to buy news paper D only, seven (7) bought Newspaper A and B. Five (5) bought newspaper B only, while ten (10) bought Newspaper A only. Four (4) of them do not buy any single paper.

    i) Determine the number of persons who buy at least newspaper B.
(5 marks)

    ii) Identify the most popular Newspaper. **(5 marks)**

b) What are the major factors when deciding between a census and a sample?
(5 marks)

ISO 9001:2008 Certified by the Kenya Bureau of Standards

c)   What precautions are necessary when using secondary data? **(5 marks)**

Q4. Distinguish between the following:-
    i)   Mutually exclusive events and compound events.   **(4 Marks)**
    ii)   Systematic sampling and simple random sampling.   **(4 Marks)**
    iii)   Quota sampling and convenient sampling.   **(4 Marks)**
    iv)   Stratified random sampling and cluster sampling.   **(4 Marks)**
    v)   Sampling error and non-sampling error.   **(4 Marks)**

THE LIST OF FORMULAE

Covariance ($\text{cov}_{(xy)}$ or S_{xy}) $= \frac{1}{N}\sum(x_i - \ddot{X})(y_i - \dot{Y})$

$\text{cov}_{(xy)}$ or $S_{xy} = (\frac{1}{N}\sum x_i y_i) - \ddot{X}\dot{Y}$

Coefficient of Correlation (r_{xy}) $= \frac{Sxy}{SxSy}$

$$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right)-\left(\sum x\right)^2}\sqrt{n\left(\sum y^2\right)-\left(\sum y\right)^2}}$$

Rank correlation coefficient or spearman's rank correlation coefficient (r_s)

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Method of least squares
$\sum y \quad = na \quad + b\sum x_i$
$\sum y_i x_i \quad = a\sum x_i + b\sum x_i^2$

$\sum x_i \quad = na \quad + b\sum y_i$
$\sum x_i y_i \quad = a\sum y_i + b\sum y_i^2$

Kendall Rank correlation

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$$

Pearson r correlation

$$r = \frac{N\sum xy - \sum (x)(y)}{\sqrt{N\sum x^2 - \sum (x^2)][N\sum y^2 - \sum (y^2)]}}$$

$$b = \frac{n\left(\sum XY\right) - \left(\sum X\right)\left(\sum Y\right)}{n\left(\sum X^2\right) - \left(\sum X\right)^2}$$

$$a = \frac{\left(\sum Y\right) - b\left(\sum X\right)}{n}$$

$$LSMA = a + bX$$

$b = r_{xy} - \frac{Sy}{Sx}$

$a = \dot{Y} - b\ddot{X}$

$$Ƃ = \frac{\sum xiyi - n\ddot{X}\ddot{Y}}{\sum x^2 - n\ddot{X}^2}$$

$$b_{xy} = r\frac{\delta x}{\delta y}$$

$$b_{xy} = r\frac{\sum xy}{\sum y2}$$

$$b_{xy} = \frac{N\sum dxdy - \sum dx\sum dy}{N\sum dy2 - (\sum dy)^2}$$

$$b_{xy} = \frac{N\sum dxdy - \sum dx\sum dy}{N\sum dx2 - (\sum dx)^2}$$

$r = \sqrt{(b_{xy} \times b_{yx})}$

$$P(n, r) = {}^nP_r = {}_nP_r = \frac{n!}{(n-r)!}$$

$$C(n, r) = {}^nC_r = {}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

$$P(A \mid B) = \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

$$P(B \mid A) = P(A \mid B) \cdot \frac{P(B)}{P(A)}.$$

$P(A) = 1 - P(A')$

$P(A \cap B) = P(A) P(B|A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B))$

$P(A \cup B) = P(A) + P(B) - P(A)P(B \mid A)$

END

ISO 9001:2008 Certified by the Kenya Bureau of Standards