

STA 425



MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS 2019/2020 ACADEMIC YEAR FOURTH YEAR SECOND SEMESTER

SCHOOL OF SCIENCE AND INFORMATION SCIENCES BACHELOR OF SCIENCE & EDUCATION

COURSE CODE: STA 425

COURSE TITLE: SAMPLING THEORY & METHODS II

DATE: 29/04/2019

TIME: 8:30 AM – 10:30 AM

INSTRUCTIONS TO CANDIDATES

1. Answer **Question ONE** and any other **Two** questions.
 2. Show all the workings clearly
 3. Do not write on the question paper
 4. All Examination Rules Apply.
-

Question One (30 Marks)

a) Explain three conditions ratio estimation method is preferred to simple random sampling **(4 Marks)**

b) Each person in a population of adults is interviewed, an identification number is assigned to each one and gender and age of each individual recorded. The population size is 1000 of which 550 are female and 450 are male. The population age is 45 years. From the list of identification numbers a simple random sample of size 50 is obtained and the following measurements are recorded: x_i is age in years and y_i is systolic blood pressure in mg/100ml of the i^{th} subject. The results are;

$$\sum_{i=1}^{50} x_i = 2248 \quad \sum_{i=1}^{50} x_i^2 = 104384 \quad \sum_{i=1}^{50} y_i = 6744 \quad \sum_{i=1}^{50} y_i^2 = 928436 \quad \sum_{i=1}^{50} x_i y_i = 305125$$

Calculate the estimate of population mean systolic blood pressure and its error of bound using Ratio estimation. **(7 Marks)**

c) A statistics aptitude test was given to 400 students prior to entering a certain university. From this students a simple random sample of 10 students was selected and their progress in calculus was observed and the final calculus results were as follows;

Student	1	2	3	4	5	6	7	8	9	10
Aptitude test(X)	40	45	20	66	60	45	30	75	35	50
Final Score(Y)	65	80	50	80	90	85	75	99	60	75

If it's known that $\mu_x = 52$, using regression estimation method, estimate:

i) The estimated population mean μ_y and the error of estimation. **(5 Marks)**

ii) The estimated total T_y and the error of estimation **(5 Marks)**

d) Describe three conditions under which cluster sampling is used. **(3 Marks)**

e) A production company wishes to estimate the ratio of change from previous year to current year in the number of machine-hours lost due to sickness. A preliminary study of $n' = 10$ employees is made and the results are as follows:

Employees	1	2	3	4	5	6	7	8	9	10
Machine-hrs lost last yr(X)	14	26	17	32	34	28	12	17	2	16
Machine-hrs lost in current yr(Y)	15	27	17	34	38	26	14	18	4	14

A production company has 1000 machines and a record shows that the total number of machine-hours lost because of breakup last year was $T_x = 16000$.

Determine the sample size required to estimate R , with a bound of error of estimation $B = 0.02$. **(6 Marks)**

Question Two (20 Marks)

a) Explain how a two stage cluster sample can be picked from a population. Give an example. **(4 Marks)**

b) An experimenter was investigating a new food additive for cattle. Midway through the two-month, she was interested in estimating the average weight for the entire herd of $N = 500$ steers. A simple random sample of $n = 12$ steers was selected from the herd and weighed. These data and pre-study weights are presented below:

Steer	Pre-study weight (kg)	Present weight (kg)
1	815	897
2	919	992
3	690	752
4	984	1093
5	200	768

STA 425

6	260	828
7	1323	1428
8	1067	1152
9	789	875
10	573	642
11	834	909
12	1049	1122

Assume that μ_x , pre-study average weight was 880kg, using regression estimation method, estimate the present weight average, μ_y for these steers and its variance. **(8 Marks)**

- c) An industry is considering revision of its retirement policy and wants to estimate the proportion of employees that favor the new policy. The industry consists of 87 separate plants located throughout Kenya. Since results must be obtained quickly and with little cost, the industry decides to use cluster sampling with each plant as a cluster. A simple random sample of 15 plants is selected and the opinions of the employees in these plants are obtained by questionnaires. The results are shown below.

plant	Number of Employee	Number Favoring Policy	plant	Number of Employee	Number Favoring Policy
1.	51	42	9.	73	54
2.	62	53	10.	61	45
3.	49	40	11.	58	51
4.	73	45	12.	52	29
5.	101	63	13.	65	46

STA 425

6.	48	31	14.	49	37
7.	65	38	15.	55	42
8.	49	30			

Estimate the proportion of employees in the industry who favor the new retirement policy and place a bound on error of estimation. **(8 Marks)**

Question Three (20 Marks)

a) A forester wishes to estimate the average height of trees on a plantation. The plantation is divided into quarter-acre plots. A simple random sample of 20 is selected from the 386 plots on the plantation. All trees on the sampled plots are measured, with results are shown below.

Number of trees	Average Height(ft)	Number of trees	Average Height(ft)
42	6.2	60	6.3
51	5.8	52	6.7
49	6.7	61	5.9
55	4.9	49	6.1
47	5.2	57	6.0
58	6.9	63	4.9
43	4.3	45	5.3
59	5.2	46	6.7
48	5.7	62	6.1
41	6.1	58	7.0

Estimate the average height of trees on the plantation and place a bound on the error of estimation. **(10 Marks)**

b) A manufacturing company supplies a product that is packaged under two brand names for marketing purposes. These two brands serve as strata for estimating potential sales volume for the next quarter. A simple random sample of customers for each brand is contacted and asked to provide a potential sales figure y (in number of units) for the upcoming quarter. Last year's sales figure x for the same quarter is also available for each sampled customer. For Brand 1 and Brand 2 there are $N_1 = 120$ and $N_2 = 180$ customers with total sales in the same quarter last year of $X_1 = 21500$ and $X_2 = 21200$ units.

Brand 1	Brand 2
X: 204 143 82 256 275 198	X: 137 189 119 63 103 107 159 63 87
Y: 210 160 75 280 300 190	Y: 150 200 125 60 110 100 180 75 90

Based on the stratified sample, estimate the potential mean sales per customer for next quarter using separate and combined ratio estimation

(10 Marks)

Question Four (20 Marks)

a) A garment manufacturer has 90 plants located throughout the East Africa and wants to estimate the average number of hours that the sewing machines were down for repairs in the past months. Because the plants are widely scattered, she decides to use cluster sampling, specifying each plant as a cluster of machines. Each plant contains many machines and checking the repair record for each machine would be the time consuming. Therefore she uses two-stage sampling. Enough time and money are available to sample 10 plants and approximately 20% of the machines in each plant.

STA 425

<i>Plant</i>	M_i	m_i	\bar{y}_i	s_i
1	50	10	5.40	11.38
2	65	13	4.00	10.67
3	45	9	5.67	16.75
4	48	10	4.80	13.29
5	52	10	4.30	11.12
6	58	12	3.83	14.88
7	42	8	5.00	5.14
8	66	13	3.85	4.31
9	40	8	4.88	6.13
10	56	11	5.00	11.80

The manufacturer knows she has a combined total of 4500 machines in all plants. Estimate the average downtime per machine and place a bound on the error of estimation. **(12 Marks)**

b) Derive the relative efficiency of two stage cluster sampling over simple random sampling in terms of ρ , the intra class correlation coefficient.

(8 Marks)

//END