CHUKA UNIVERSITY

UNIVERSITY EXAMINATIONS

FOURTH YEAR EXAMINATION FOR THE AWARD OF DEGREE OF BACHELOR OF SCIENCE IN MATHEMATICS

MATH 441: SAMPLING METHODS II

STREAMS: BSC (MATH) TIME: 2 HOURS

DAY/DATE: FRIDAY 26/03/2021 8.30 A.M. – 10.30 A.M.

INSTRUCTIONS:

- **Answer question ONE (Compulsory) and any other TWO questions**
- **Sketch maps and diagrams may be used whenever they help to illustrate your answer**
- **Do not write on the question paper**
- **This is a closed book exam, no reference materials are allowed in the examination room**
- **There will be No use of mobile phones or any other unauthorized materials**
- **Write your answers legibly and use your time wisely**

QUESTION ONE (30 MARKS)

(a) Distinguish between an estimator and an estimate using relevant examples [4 marks]

(b) Define cluster sampling and explain the various types of cluster samples [6 marks]

(c) Discuss three properties of estimators [6 marks]

(d) How is multistage sampling procedure carried out? [2 marks]

(e) Derive the condition when regression estimator and ratio estimator are equally efficient. [4 marks]

(f) Explain when the Jack knife re-sampling technique is likely to fail. [1 marks]

(g) Consider the following data where 3 clusters out of 10 clusters are sampled (n = 3) with replacement.

$y_1 = 420, y_2 = 1785, y_3 = 2198$

$M_1 = 650, M_2 = 2840, M_3 = 3200$

Find the Hansen-Hurwitz estimator for the population mean and also find the variance of the estimator [7 marks]

QUESTION TWO (20 MARKS)

A mathematical achievement test was given to 486 students prior to entering a certain college who then took a calculus class. A simple random sampling of 10 students is selected and their calculus score recorded. It is known that the average achievement test score for the 486 students was 52.

Student	Achievement test Score X	Calculus score Y
1	39	65
2	43	78
3	21	52
4	64	82
5	57	92
6	47	89
7	28	73
8	75	98
9	34	56
10	52	75

Obtain the following:

(i) The regression estimate [2 marks]

(ii) Variance of the regression estimate [4 marks]

(iii) An approximate 95% CI for μ using regression estimation [4 marks]

(iv) The value for the ratio estimate [1 mark]

(v) Variance of the ratio estimate [2 marks]

(vi) An approximate 95% CI for μ using ratio estimation [7 marks]

QUESTION THREE (20 MARKS)

(a) Explain three main problems caused by missing data [3 marks]

(b) Explain any three imputation methods [3 marks]

(c) A sociologist wants to estimate the average yearly vacation budget for each household in a certain city. It is given that there are 3,100 households in the city. The sociologist marked off the city into 400 blocks and treated them as 400 clusters. He then randomly sampled 24 clusters interviewing every household living in that cluster. The data are given in the table below:

1	7	12,000
2	9	15,000
3	5	8,000
4	8	13,000
5	12	18,000
6	5	7,000
7	4	6,000
8	8	13,000
9	14	22,000
10	6	9,800
11	3	7,000
12	13	18,000
13	8	12,340
14	4	5,000
15	6	8,900
16	9	14,000
17	3	4,000
18	10	11,400
19	4	5,000
20	7	13,000
21	6	8,900
22	5	8,700
23	7	10,000
24	6	9,200
	169	259,240

Find the

(i) Ratio estimator for the mean yearly vacation budget for each household in that city.

(ii) Estimated variance for the ratio estimator

(iii) Unbiased estimator for the mean yearly vacation budget for each household in that city

(iv) Estimated variance for the unbiased estimator

(v) Comment on the variances for the two estimators [14 marks]

QUESTION FOUR (20 MARKS)

(a) A population consists of MN elements grouped into N first stage units (fsu's) of M second stage units (ssu's) each. Let n be the number of fsu's in the sample and m be the number of ssu's to be selected from each of the fsu. Show that

$$var(\bar{y}) = \frac{N-n}{N}\frac{S_b^2}{n} + \frac{M-m}{M}\frac{S_w^2}{mn}$$

$S_b^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - Y)^2$ and

$S_w^2 = \frac{1}{N(M-1)}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(Y_{ij} - Y_i\right)^2$ [6 marks]

(b) A restaurant chain wants to estimate the average employee satisfaction with their job (the scale is from 1 to 7). They have 120 restaurants the total number of employees in the chain is 6860. They use simple random sampling to sample 10 restaurants. They then use simple random sampling to sample and interview about 20% of the employees in those restaurants. The data given as follows.

Restaurant	M_i	m_i	Employee satisfaction
1	54	10	5 , 7, 6, 5, 4, 7, 6, 6, 4, 5
2	48	10	7, 7, 7, 6, 5, 4, 7, 7, 6, 6
3	68	14	5, 6, 5, 6, 4, 5, 6, 5, 4, 5, 4, 6, 5, 6
4	70	14	6, 5, 7, 6, 7, 6, 5, 7, 5, 7, 6, 5, 7, 6
5	52	10	4, 5, 4, 5, 5, 5, 6, 5, 4, 4, 4
6	62	12	5, 7, 6, 7, 4, 3, 1, 5, 4, 6, 4, 5
7	41	8	7, 6, 7, 7, 6, 6, 5, 7
8	53	11	6, 6, 5, 4, 6, 7, 5, 5, 7, 6, 5
9	64	12	7, 6, 5, 4, 6, 5, 7, 4, 3, 6, 5, 7
10	43	9	7, 6, 6, 5, 7, 3 ,5, 4, 5

(i) The unbiased estimator of the average employee satisfaction and compute the standard error of your estimate

(ii) Find the estimated variance of the unbiased estimator [14 marks]

QUESTION FIVE (20 MARKS)

(a) Briefly discuss what is meant by bootstrapping in statistics [4 marks]

(b) What is the main difference between Jackknife and Bootstrap method [4 marks]

(c) Given the set of values {1, 3, 9, 15, 20}, determine the jackknife estimate for both the mean and the standard deviation of the mean. [12 marks]
