

CHUKA



UNIVERSITY

UNIVERSITY EXAMINATIONS

**FOURTH YEAR EXAMINATION FOR THE AWARD OF DEGREE OF BACHELOR
OF SCIENCE IN ECONOMICS AND STATISTICS AND BACHELOR OF
EDUCATION SCIENCE**

MATH 441: SAMPLING METHODS II**STREAMS: BSc Econ Stat****TIME: 2 HOURS****DAY/DATE: MONDAY 07/12/2018****11.30 A.M. – 1.30 P.M.****INSTRUCTIONS: Answer Question ONE and any other TWO Questions.****QUESTION ONE [30 MARKS]**

- a) State three properties of estimators and state how each is measured. (3 marks)
- b) State and explain three types of non-response in Sample Surveys. (6 marks)
- c) A Cider forest has got 280 trees of the same species and of similar ages. A preliminary estimate is required of the total weight of timber that trees will yield. A forestry expert claims to be able to make fairly accurate assessments of the yield from any tree merely by visual inspection and makes such assessments for all 280 trees. He assesses the total yield as 439.5 tonnes. Subsequently, 25 trees picked at random are felled and their timber yields

y_i and their corresponding assessed yields, x_i , provide the following summary results.

$$\sum_{i=1}^{25} x_i = 41.4, \quad \sum_{i=1}^{25} x_i^2 = 73.47, \quad \sum_{i=1}^n y_i = 39.8,$$

$$\sum_{i=1}^n y_i^2 = 69.08, \quad \sum_{i=1}^{25} x_i y_i = 70.64 \quad . \text{ Using the ratio estimator, estimate the:}$$

- i) Total yield, \hat{Y}_r

- ii) $\text{Var}(\hat{Y}_r)$ and
- iii) The confidence interval of \hat{Y}_r (10

marks)

- d) State three problems that the missing data causes. (3

marks)

- e) A firm is interested in estimating the average per capita income in a certain city. There is not an available list of resident adults. The city is marked off into rectangular blocks, except for two industrial areas and three parks which contain a few houses.

The researchers decide that each of the city blocks will be considered a cluster, the two industrial areas will be considered a cluster and, finally, the three parks will be considered a cluster. The clusters are numbered from 1 to 60 and there is budget for sampling $n = 20$ clusters and to interview every household within each cluster.

Number of Residents m_i	55	60	63	58	71	78	69	58	52	71
	73	64	69	58	63	75	78	51	67	70
Total Income per cluster y_i	2210	2390	2430	2380	2760	3110	2780	2370	1990	2810
	2930	2470	2830	2370	2390	2870	3210	2430	2730	2880

i) Obtain the mean estimate (\hat{Y}) per capita income in this city.

ii) $\text{Var}(\hat{Y})$

(8 marks)

QUESTION TWO (20 MARKS)

In studying lung function in a group of 560 workers in a coal mine, an estimate was required of the mean value of some relevant measure Y . A simple random sample of 10 workers was chosen and their Y values, y_i , determined by an appropriate test. A note was also made of

their heights, x_i . The results were as below:

MATH 441

y_i	3.0	3.5	3.3	3.1	4.1	3.2	3.7	2.9	3.9	3.4
x_i (c m)	173	183	170	175	160	157	168	180	178	163

From routine medical records the average height for the group of 560 workers is known to be

$$\bar{X} = 173.2 \text{ cm.}$$

- a) Obtain \hat{Y}^i of \bar{Y} from the data using:
- i) Ratio estimator
 - ii) Regression estimator
 - iii) Estimator based on SRSWOR (6 marks)
- b) Calculate the var (\hat{Y}^i) using the three methods in (a) above and hence (10 marks)
- c) Compare the efficiencies of ratio estimator, regression estimator and estimator based on SRSWOR. (3 marks)
- d) Under what circumstances does the ratio estimator fail? (1 mark)

QUESTION THREE (20 MARKS)

- a) Describe the basic steps in Bootstrap procedure of sampling, stating one advantage and one disadvantage of the method. (7 marks)
- b) Suppose that the following 16 observations were recorded during a laboratory experiment:
- 17.23 13.93 15.78 14.91 18.21 14.28 18.83 13.45
- 18.71 18.81 11.29 13.39 11.57 10.94 15.52 15.25
- i) Obtain the variance estimate (σ^2) of the data using Jackknife method and
 - ii) Find a 95% Jackknife confidence interval for $\log(\sigma^2)$ (13 marks)

QUESTION FOUR (20 Marks)

- a) State and explain four techniques used in imputation (4 marks)
- b) State two advantages and two disadvantages of imputation (4 marks)

MATH 441

- c) A manager is interested in estimating the total sale in thousands for all of its 360 branches. From last year record, the total sale in thousands for all the 360 branches is 31500. Careful check of this year records are obtained for a SRS of 15 branches with the following results:

Branch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Last year sale x	5	4	1	1	1	3	9	2	111	21	6	5	10	110	50
This year sale y	6	4	2	1	1	2	1	3	16	41	7	6	95	65	83

Using the regression estimator, estimate the:

- iii) Total sale (\hat{Y}_{reg}) in thousands for all the 360 branches.
- iv) Standard error for the total yield, $Se(\hat{Y}_{reg})$ (12 marks)

QUESTION FIVE (20 Marks)

- a) State four properties of cluster sampling method. (4 marks)
- b) State two advantages and two disadvantages of cluster sampling (4 marks)
- c) There are 36 departments in a college. One wants to estimate the average amount of money the students spent on textbooks last semester. Since the size of each department varies very much, a two-stage cluster sampling using probability proportional to size for the primary unit is carried out. The results are listed in the table below.

Department	Mi	mi	Textbook expenses in \$ for last semester
1	10	4	326, 400, 423, 443
2	20	8	278, 312, 450, 350, 227, 438, 512, 403
3	30	12	512, 256, 332, 402, 512, 309, 411, 610, 422, 630, 550, 470
4	15	6	426, 312, 512, 440, 342, 533

- v) Obtain an unbiased estimator (\hat{Y}) of the average amount of money the students spent on textbooks last semester.
- vi) Calculate the var (\hat{Y}) for above data. (12 marks)

