CHUKA UNIVERSITY

UNIVERSITY EXAMINATIONS

FIRST YEAR EXAMINATION FOR THE AWARD OF DEGREE OF MASTERS OF SCIENCE IN APPLIED STATISTICS

MATH 851: SAMPLE SURVEYS

STREAMS: MSC (APP STAT) **TIME: 3 HOURS**

DAY/DATE: FRIDAY 06/12/2019 **2.30 P.M – 5.30 P.M**

INSTRUCTIONS:

- Answer **ANY THREE** 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: (20 MARKS)

a) In a population of 4 000 people, we are interested in two proportions: P_1 = proportion of individuals owning a dishwasher, P_2 = proportion of individuals owning a laptop computer. According to 'reliable' information, we know a priori that:

$45\ \% \leq P_1 \leq 65\ \%$, and $5\ \% \leq P_2 \leq 10\ \%$.

What does the sample size n have to be within the framework of a simple random sample if we want to know at the same time P_1 near ± 2 % and P_2 near ± 1 %, with a confidence level of 95 %? (5 marks)

b) We want to estimate the surface area cultivated on the farms of a rural township. Of the N = 2010 farms that comprise the township, we select 100 using simple random sampling. We measure y_k, the surface area cultivated on the farm k in hectares, and we find $\sum_{k\in S} y_k$ = 2907 ha and $\sum_{k\in S} y_k^2$= 154593 ha^2. Give the value of the standard unbiased estimator of the mean $\bar{y}$ and 95 % confidence interval for $\bar{y}$. (5 marks)

c) We are interested in estimating the proportion of men P affected by an occupational sickness in a business of 1500 workers. In addition, we know that three out of 10 workers are usually affected by this sickness in businesses of the same type. We propose to select a sample by means of a simple random sample.

i) What sample size must be selected so that the total length of a confidence interval with a 0.95 confidence level is less than 0.02 for simple designs with replacement and without replacement ?

ii) What should we do if we do not know the proportion of men usually affected by the sickness (for the case of a design without replacement)?

(10 marks)

QUESTION TWO: (20 MARKS)

a) The H-T estimator is often studied under a model which represents the $y_i's$ are realized values according to the predictor model;

$E(y_i/x_i) = \beta x_i$

$var(Y_i/X_i = x_i) = \sigma_i x_i^2$

$cov(Y_i/X_i = x_i) = 0, i \neq j, \pi_i = kx_i, k = \frac{n}{NX}$

$\hat{\hat{Y}}_{HT} = \sum_{i=1}^{n} \frac{y_i}{N\pi_i} = \frac{\bar{X}}{n} \sum_{i=1}^{n} \frac{y_i}{x_i}$

Show that;

i) The H-T estimator is model unbiased under the prediction model given above.

ii) The optimal sample to use is the one with the largest values of x_s' so as to reduce the variance $var\,(\hat{\hat{Y}}_{HT} - \hat{Y})$ (10 marks)

b) The results from a simple random sample from a stratified population are given in the following table. The columns are the strata sizes, the strata sample sizes, the sample means and the sample variances.

Stratum	N_h	n_h	$\bar{y}_h$	S_h^2
1	100	20	115.7	57.6
2	300	20	147.2	46.9
3	400	30	133.6	75.3

i) Find a 95% confidence interval for the population mean.

ii) The person taking the sample claimed that they used optimal allocation to select the sample size based on good information about the likely sizes of the strata variances. Given their allocation can you determine their prior guess for the strata variances? Explain. Based on the results of the sample does it appear that they made good choices with their guesses for the strata variances assuming the observed sample variances are the true sample variances and the optimal allocation for a sample size of 70? (10 marks)

QUESTION THREE: (20 MARKS)

a) Distinguish between multiphase sampling and multi stage sampling (2 marks)

b) Explain briefly how double sampling procedure is performed (3 marks)

c) Consider the data below on number of trees per acre on a 1000- acre plantation.

Plot	**Actual number per acre *Y***	**Aerial estimate *X***
1	25	23
2	15	14
3	22	20
4	24	25
5	13	12
6	18	18
7	35	30
8	30	27
9	10	8
10	29	31

The investigator samples 10 one-acre plots by simple random sampling and counts the number of trees (y) on each plot. She also has aerial photographs of the plantation from which she can estimate the number of trees (x) on each plot of the entire plantation. Use the information to compute;

i) The ratio estimator

ii) The sampling variance of the ratio estimator

iii) The 95% confidence interval for the ratio estimator (10 marks)

c) In successive sampling on two occasions, suppose that $\bar{y}'_m$ and $\bar{y}'_u$ are two estimates for the mean for the second occasion (with usual notation). Show that

$var(\bar{y}'_m) = \frac{S^2}{np}[1 + (1 - 2\rho)q]$ (5 marks)

QUESTION FOUR: (20 MARKS)

a) Discuss the various imputation techniques. (3 marks)

b) Classify the types of non-response that may be encountered in sample surveys (3 marks)

c) Discuss clearly the procedure of obtaining a bootstrap estimate of a sampling distribution (3 marks)

d) Consider the following set of 12 observations.

17.23 13.93 15.78 14.91 18.21 14.28

18.71 18.81 11.29 13.39 11.57 10.94

Use the Jacknife resampling technique to obtain the estimate for σ^2 and its confidence interval for the data. (7 marks)

e) Discuss the relationship between the Jacknife and Bootstrap techniques (4 marks)
