TECHNICAL UNIVERSITY OF MOMBASA

UNIVERSITY EXAMINATION FOR:

FACULTY OF APPLIED AND HEALTH SCIENCES

DEPARTMENT OF MATHEMATICS & PHYSICS

MASTER OF SCIENCE IN APPLIED STATISTICS

AMA 5107: DESIGN AND ANALYSIS OF SAMPLE SURVEYS

END OF SEMESTER EXAMINATION

SERIES: AUGUST

TIME: 3 HOURS

DATE: AUGUST 2019

Instructions to Candidates
You should have the following for this examination
-*Answer Booklet, examination pass and student ID*
This paper consists of **FIVE** questions. Attempt **Question ONE and any other TWO.**
Show all the working clearly
Do not write on the question paper.

QUESTION ONE (30 MARKS)

a) (i) State and explain types of sample surveys. (4 mks)
   (ii) What are the main types of non-response? (4 mks)

b) Consider a population consisting of 430 units by complete enumeration of the population. It was found that sample mean, $\bar{x}$ is 19 and S^2 is 85.6. There being true population values with simple random sampling, how many units must be taken to estimate $\bar{x}$ within 10% of $\bar{x}$, apart from a chance of 1 in 20? (5 mks)

c) In a population with N=6, the values of X_n are 8, 3, 1, 11, 4 and 7.
   (i) Compute the sample mean $\bar{x}$ for all possible simple random samples of size 2. (5mks)
   (ii) Verify that $\bar{x}$ is an unbiased estimator of $\bar{X}$. (3mks)

d) In the usual cluster notation for one-stage sampling (using a Simple random sample of clusters).

$$\hat{Y}_U = N\, \bar{y} \text{ and } \bar{y}_{cl} = \frac{\sum\limits_{i=1}^{n} y_i}{\sum\limits_{i=1}^{n} m_i} \quad \text{are estimators of the population total } Y$$

and the population mean $\bar{Y}_{CL}$ respectively.

   (i) Explain notations N, n, m_i, y_i, & $\bar{y}$ (4mks)

   (ii) Show that $\hat{Y}_U$ is an unbiased estimator of the population total Y. (3mks)

(iii)   Why is $\bar{y}_{cl}$, a biased estimator of the population mean $\bar{Y}_{CL}$ and when is its bias small? (2mks)

QUESTION TWO (20 MARKS)

a)   A population of size 800 is divided into 3 strata. Their sizes and standard deviations are given below,

Strata	1	2	3
Size	200	250	350
Std dev.	6	8	12

A stratified sample of size 120 is to be drawn from the population. Determine the sample sizes incase of

(i)   Proportional allocation (4mks)
(ii)   Optimal allocation. (8mks)

b)   An orange grower is to sell a truckload of oranges. The oranges are packed into 140 crates containing 120 oranges each. Before striking the deal, the buyer wants to estimate the quantity of juice in the oranges, and proposes to inspect a sample of oranges.

(i) *Convenience sampling* chooses the items which are most accessible while sampling is in progress. Suggest reasons why *cluster sampling* might be preferred to convenience sampling for this sampling inspection. How does one- and two-stage cluster sampling differ in the context of this example? Mention any practical difficulties that might arise in choosing genuinely random samples in this study. (8mks)

QUESTION THREE (20 MARKS)

a)   A large supermarket company plans to build a new store in a town which is situated in a popular retirement area. The local Chamber of Commerce has conducted a survey of local residents to determine their views on this proposed store.

The survey was conducted by questioning shoppers on the town's main street. The key question asked was "Do you approve of the proposed supermarket scheme?" The results for this question are summarised in the table below.

Sex	Population of each sex in the town	Achieved sample size	No. who approve of the supermarket scheme
Female	28000	100	32
Male	21000	100	65

(i)   Construct 95% confidence intervals for the proportions of residents of each sex who approve of the proposed supermarket scheme. From these confidence intervals, what might be said about the views of females compared to those of males?(9mks)
(ii)   Estimate the proportion of the total population of the town who approve of the supermarket scheme. Express this estimate as a 95% confidence interval. (6mks)

b)   A local academic has claimed that the extent of the difference identified between the views of the two sexes is misleading, and that much of the difference can be attributed to differences in opinions of people of different ages. The supermarket chain is also unhappy with the survey results, claiming that the method of selecting the sample was biased.

(i)   Explain what is meant by bias. (2mks)
(ii)   What assumptions is the academic making? To what extent do you think his claim is valid? (3mks)

QUESTION FOUR (20 MARKS)

a) A regional council wishes to assess the amount of hazardous waste produced by the 6231 manufacturing companies in its area. They are split into three strata:

(1) basic metal industries;
(2) food, textiles and mineral products;
(3) other manufacturing.

A simple random sample of companies was taken in each stratum, and for each company the total quantity of hazardous waste (in thousands of tonnes) produced in 2003 was measured.

stratum	Hazardous waste ('000's tonnes)			
	N_h	n_h	$\bar{y}_h$	s_h
1	92	11	166.7	207.7
2	1612	61	7.7	14.7
3	4527	292	0.3	4.5
Total	6231	364		

(i) Define the symbols N_h, n_h, $\bar{y}_h$, s_h as used above and show that

$$\bar{y}_{st} = \sum \frac{N_h}{N} \bar{y}_h$$

is an unbiased estimator for the mean hazardous waste produced per company. (8mks)

(ii) Estimate the mean hazardous waste produced per company and obtain an estimate of the standard error of your estimator. Give an approximate 95% confidence interval for the mean hazardous waste per company. (8mks)

(iii) A sampling frame is subject to several types of defects. Broadly classify these defects into four classes. (4mks)

QUESTION FIVE (20 MARKS)

a) (i) Write down the formula for calculating an unbiased estimate, s^2, of the variance of a large (but finite) population, based on a simple random sample of n items. Define any symbols you use.

Show also that, for a binary variable, $s^2 = n p (1 - p)/(n - 1)$, where n and p are to be defined. (8mks)

(ii) Explain the difference between *random* and *non-random* methods of sampling discussing both the construction of samples and the methods of analysing data collected by them. Suggest reasons why non-random samples may sometimes preferred. (8mks)

b) Explain the advantages of systematic sampling. (4mks)