TECHNICAL UNIVERSITY OF MOMBASA

FACULTY OF APPLIED AND HEALTH SCIENCES

DEPARTMENT OF MATHEMATICS AND PHYSICS

UNIVERSITY EXAMINATION FOR:

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.**

Do not write on the question paper.

Question ONE (30 Marks)

a) State and briefly explain three limitations of sampling theory. (3 mks)

b) (i) Explain simple random sampling method. (2 mks)

(ii) Is the sample mean of the above design, a consistent estimator of the population mean? Show. (2mks)

(iii) Give the variance of the sample mean and also an unbiased estimator of this variance. (2mks)

(iv) What is the sample size requirement to estimate the population mean with a given standard error? (2mks)

c) (i) Explain how systematic sampling is performed. (3 mks)

(ii) Assuming the list (sampling frame) from which a systematic sample is taken is not a random ordering of the elements, with respect to the variable being measured, explain by means of an example how you would estimate the variance of a point estimator of the population mean ($\bar{y}$).You may use k=44, j = 1514 and N =15301, where each symbol retains its usual meaning. Give an expression for Var $(\bar{y})$. (6mks)

d) (i) Explain briefly five advantages of stratification. (5 mks)

(ii) Why is stratified random sampling regarded as a restriction over simple random sampling? (3mks)

(iii) Distinguish between the terms pre-stratification and post-stratification. (2mks)

Question TWO (20 marks)

Wildlife managers want to estimate the total number of wild beasts in the Maasai Mara herd located in Kenya. The density of wild beasts differs dramatically in different types of habitat. A preliminary aerial investigation has identified the area used by the herd, and divided it into six strata based on habitat type.

For the main survey, the organiser decides to divide the area into sub-areas called quadrats, each of size 4 km^2. The survey is conducted by selecting a simple random sample of quadrats from each stratum, and for each quadrat the area is searched by aircraft to locate and then photograph the animals; the number of wild beasts, *y*, in each quadrat is counted in the photographs.

The sample means and standard deviations of the measurements, *y*, in each stratum based on a sample of 211 quadrats are as follows.

Stratum (h)	Map Quadrants N_h	Sample Quadrants n_h	$\overline{y_h}$	s_h
1	400	98	24.1	74.7
2	40	10	25.6	63.7
3	100	37	267.6	589.5
4	40	6	179.0	151.0
5	70	39	293.7	351.5
6	120	21	33.2	99.0
Total	770			

(i) The sampling frame for this survey is a land map. Discuss briefly what problems are likely to be associated with this type of sample. (4mks)

(ii) Define the terms s_h , N_h and n_h as used above. (3mks)

(iii) Using the data above, estimate the total number of wild beasts in the herd and obtain an approximate 95% confidence interval for this total. (10mks)

(iv) For this survey discuss briefly the merits of using stratified sampling rather than simple random sampling. (3mks)

Question THREE (20 Marks)

a) Explain what is meant by non response in a sample survey and state the types of non response that may be encountered. (6 mks)

b) Explain clearly the difference between stratified random sampling and quota sampling. (2mks)

c) State what is meant by the term quota controls and list three quota controls commonly used. (4mks)

d) For stratum $i(i=1,2,.....,k)$, *let* N_i denote the total number of units, n_i the number of units in the sample, $y_{ij}(i=1,2,.....,N_i)$ the value of the measured characteristic for the j^{th} elementary unit in the i^{th} stratum, $\overline{Y_i}$ the stratum mean and $\overline{y_i}$ the sample mean for the i^{th} stratum. Assuming that sampling is done without replacement, obtain an unbiased estimator of the population mean $\overline{Y_i}$ and find the variance of this estimator. (8mks)

Question FOUR (20 Marks)

a) Discuss the main arguments for quota sampling and against quota sampling. (6mks)

b) A market research is allocated £20,000 to conduct a survey by means of a stratified random sample. The survey is to be conducted on three stratums: A, B and C. The population of stratum A is 40,000, B is 20,000 and C is 10,000. The set cost of administering the survey is £200 and the cost of sampling one unit are £2.25, £4.00, £1.00 for stratum A, B and C respectively. The standard deviation of the observations in stratum A is thought to be twice that of B and C. Determine the optimum and the proportional allocations assuming that all money is to be spent. (14 mks)

Question FIVE (20 Marks)

a) A student wants to estimate the average grade point average (GPA) in his dormitory. Instead of obtaining a listing of all students in the dorm and conducting an SRS, he notices that the dorm consists of 100 suites, each with four students; he chooses 5 of those suites at random, and asks every person in the 5 suites what her or his GPA is.
The results are as follows:

	Suite (Clusters)				
Person Number	**1**	**2**	**3**	**4**	**5**
1	3.08	2.36	2.00	3.00	2.68
2	2.60	3.04	2.56	2.88	1.92
3	3.44	3.28	2.52	3.44	3.28
4	3.04	2.68	1.88	3.64	3.20
Total	12.16	11.36	8.96	12.96	11.08

(i) Find the estimate of the population total (2 mks)

(ii) Find the average of suite totals, $\overline{t_i}$ (2 mks)

(iii) Find the sample variance, (ie $\overline{s_i}$ is the usual sample variance), hence find the standard error. (5 mks)

b) Give three situations in which a cluster sampling design might be used rather than simple random sampling. What might be the drawbacks of cluster sampling in such cases?

(8 mks)

c) In systematic sampling, the intraclass correlation coefficient can with the usual notation be given as

$$\rho_y = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{n}\sum_{i<j}(y_{ij}-\bar{y})(y_{ij}-\bar{y})}{nk(n-1)\sigma_y^2}$$

(i) What does ρ_y measure? (1mk)

(ii) It can be shown that $\frac{-1}{n-1} \le \rho_y < 1$ and that $Var(\bar{y}_{sy}) = \left(\frac{\sigma_y^2}{n}\right)\left[1+\rho_y(n-1)\right]$

Explain what happens as $Var(\bar{y}_{sy})$ as $\rho_y \to 1$, $\rho_y \to 0$ (2mks)