MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS
2018/2019 ACADEMIC YEAR
FIRST YEAR SECOND SEMESTER EXAMINATION

SCHOOL OF SCIENCE AND INFORMATION SCIENCE
DEPARTMENT OF MATHEMATICS AND PHYSICAL SCIENCES
BACHELOR OF SCIENCE IN APPLIED STATISTICS WITH COMPUTING

COURSE CODE: STA 1208
COURSE TITLE: PRINCIPLES OF SAMPLE SURVEYS

DATE 18TH APRIL 2019 **TIME: 0830 -1030HRS**

INSTRUCTIONS:

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

SECTION ONE (30 MARKS)

a) Define the following terms as used in sample survey (6Marks)
   i) Sample and population
   ii) Sampling frame
   iii) Purposive sampling and systematic sampling
   iv) Parameter

b) Why do most researchers opt for using a sample in surveys as opposed to conducting a census 4marks)

c) Differentiate with examples and approaches given between probability sampling techniques from non-probability sampling? (6marks)

d) A Researcher has taken a small survey, using an SRS, for energy usage in houses. On the basis of the survey, each house is categorized as having electric heating or some other kind of heating. The January electricity consumption in kilowatt-hours for each house is recorded *(Yi)* and the results are given below: (8marks)

Type of Heating	Number of Houses	Sample Mean	Sample Variance
Electric	24	972	202,396
Nonelectric	36	463	96,721
Total	60		

From recording existing , it is known that 16,450 of the 35,000 houses have electric heating, and 18,550 have nonelectric heating.

i) Using the sample, give an estimate and its standard error of the proportion of houses with electric heating. Does your 95% CI include the true proportion?

ii) Give an estimate and its standard error of the average number of kilowatt-hours used by houses in the city. What type of estimator did you use, and why did you choose that estimator?

e) Explain when a design may be considered as a cluster sample. What are the first-stage and second-stage units in cluster sampling of a country like Kenya? (6marks)

QUESTION TWO (20 MARKS)

A simple random sample of 1 in 20 households in a small town provided the following data about the availability of cars and the number of adults in household3.

Number of cars *(Yi)* in the household	Adults in household *(Xi)* 1	2	3	4	5	Total
0	58	127	9	6	0	200
1	68	140	27	4	1	240
2	4	30	5	8	3	50
3	0	3	4	2	1	10
Total	130	300	45	20	5	500

a) Obtain point estimates, and approximate 95% confidence intervals for the following given that, $\sum \bar{x}\,\bar{y} = 795$):

i. the total number of cars in the town's households,

ii. the ratio of cars per adult in the town's households,

iii. the proportion of households with 1 or more cars per adult

b) A survey is to be conducted on the prevalence of the common diseases in a large population. For any disease that affects at least 1% of the individuals in the population, it is desired to estimate the total number of cases, with a coefficient of variation of not more than 20%. What size of a simple random sample is needed, assuming that the presence of the disease can be recognized without mistakes?

QUESTION THREE (20 MARKS)

(c) A campus population of size N = 9000 is to be surveyed by a stratified sample for the prevalence of a certain disease, based upon three strata of respective sizes Nh = 1000,3000,5000 for h 1,2,3. The costs of sampling individuals from these strata are estimated to be respectively 40, 20, and 10 USD per person. The campus health authorities believe that roughly 1% of stratum 1, 5% of stratum 2, and 12% of stratum 3 will test positive for the disease. I)

i) What is the optimal number of individuals to sample in each stratum if the total budget for data- collection in the survey is *USD20000.*

ii) Suppose that the same population were to be sampled by SRS. About how much would the SRS cost if you want to achieve the same MSE as in (a) in estimating the proportion of the population who have the disease ?

d) An opinion poll on Kenya's health concern was conducted by the Kenya National Aids Program between April 10-15, 2011. The survey reported that 89% of adults consider AIDS

as the most urgent health problem of the Kenya, with a margin of error of ±3%. The result was based on telephone interviews of 872 adults.

a. What was the target population?
b. What was the sample population?
c. How was the survey was conducted?
d. How was the sample selected?

QUESTION FOUR (20 MARKS)

a) Foresters want to estimate the average age of trees in a stand in Mau forest. Determining age is cumbersome because one needs to count the tree rings on a core taken from the tree. In general, though, the older the tree, the larger the diameter, and diameter is easy to measure. The foresters measure the diameter of all 1132 trees and find that the population mean equals 10.3. They then randomly select 20 trees for age measurement.

Tree No.	Diameter, x	Age, y	Tree No.	Diameter, x	Age, y
1	12.0	125	11	5.7	61
2	11.4	119	12	8.0	80
3	7.9	83	13	10.3	114
4	9.0	85	14	12.0	147
5	10.5	99	15	9.2	122
6	7.9	117	16	8.5	106
7	7.3	69	17	7.0	82
8	10.2	133	18	10.7	88
9	11.7	154	19	9.3	97
10	11.3	168	20	8.2	99

i) Estimate the population mean age of trees in the stand and give an approximate standard error for your estimate.

b) An accounting firm is interested in estimating the error rate in a compliance audit it is conducting. The population contains 828 claims, and the firm audits an SRS of 85 of those claims. In each of the 85 sampled claims, 215 fields are checked for errors. One claim has errors in 4 of the 215 fields, 1 claim has three errors, 4 claims have two errors, 22 claims have one error, and the remaining 57 claims have no errors.

i. Treating the claims as parameter 's and the observations for each field as sample's, estimate the error rate for all 828 claims. Give a standard error for your estimate.

ii. Estimate (with SE) the total number of errors in the 828 claims.

QUESTION FIVE

a) Suppose we want to estimate the average number of hours of TV watched in the previous week for all adults in some county. Suppose also that the populace of this county can be grouped naturally into 3 strata (Nairobi, Kisumu, Sayepei(rural)) as summarized in the table

Statistic	Nairobi	Kisumu	Sayepe(rural)
Nh	155	62	93
nh	20	8	12
Yh	33.90	25.12	19.00
Sh	5.95	15.24	9..36
Th	5254.5	1557.4	1767.0
Ch	2	2	3

(i) Compute a 95% confidence interval for the total number of hours of TV watched in the previous week for all adults in this county.

(ii) Estimate the total sample size needed to estimate the mean hours of TV watched in this particular county to within 1 hour with 99% probability using optimal allocation (unequal and equal costs).

(b) A local radio station carries out regular polls of its listeners on items of current interest. In one such poll listeners were asked to telephone the station and just answer "yes" or "no" to the following questions.

Do you think the government of Kenya is serious in the fight against corruption?

The poll was carried out between 8 am and 9 am one morning. At 8:30 am the announcer said the percentage of "yes" vote was 63%. When the poll closed at 9 am he announced that the percentage was 52%. List two problems associated with this method of polling and suggest why each problem might cause misleading conclusion to be drawn.

//END