

STA 113



ALUPE UNIVERSITY
COLLEGE
Bastion of Knowledge...

P.O.Box 845-50400 Busta(K)
principal@auc.ac.ke
Tel: +254 741 217 185
+254 736 044 469
off Busta-Malaba road

OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2017 /2018 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER REGULAR EXAMINATION

FOR THE DEGREE OF BACHELOR OF APPLIED STATISTICS WITH COMPUTING

COURSE CODE: STA 113

COURSE TITLE: PRINCIPLES OF SAMPLE SURVEY

DATE: 26TH APRIL, 2018

TIME: 9AM – 12.00 NOON

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 7 PRINTED PAGES

PLEASE TURN OVER

STA 113: PRINCIPLES OF SAMPLE SURVEY

STREAM: ASC

DURATION: 3 Hours

INSTRUCTIONS TO CANDIDATES

- i. Answer Question **TWO** questions in section A and any other **THREE** questions in section B.
- ii. Maps and diagrams should be used whenever they serve to illustrate the answer.
- iii. Do not write on the question paper.

SECTION A (31 marks): Answer ALL questions.**QUESTION ONE (16 Marks)**

- a) Distinguish between
 - i) Census and sample survey [2Marks]
 - ii) Population and a sample [2Marks]
 - iii) Target and sampled populations [2Marks]
- b) Explain the main reasons of collecting a sample instead of census [1Marks]
- c) A forester wishes to estimate the population of trees in a lowland forest that have met wood infections. These infections can be detected by the presence of a particular bacteria in the juice extracted from the boreholes. It is known that there are about 5000 trees in the forest. Borehole test performed on a simple random sample of 500 trees indicate that 100 trees were affected.
 - i) Estimate total number of affected trees in the forest. What is the standard error of your estimate? [2Marks]
 - ii) Construct 99% confidence interval for population proportion and population total of affected trees. [3Marks]
- d) In a psychological experiment, individuals are permitted to react to a stimulus in one of the two ways, say A or B. That experimenter wishes to estimate the proportion of persons exhibiting reaction A, which he denoted by p . How many persons should be 95% confidence that the error of estimation is within 0.05 if he wanted to be 95% confident?
 - i) Knows that $p=0.6$ [2Marks]
 - ii) Has no idea about the value of p . [2Marks]

QUESTION TWO (15 Marks)

- a) Distinguish between
 - i) Systematic and stratified random sampling methods. [2Marks]



- ii) Cluster and quotas sampling methods. [2Marks]
- iii) Simple random and two stage sampling [2Marks]

b) In a sample survey the following data were collected.

Strata		
1	2	3
3	6	5
5	4	7
2	5	2
2		4
		2

Calculate

- i) Sample mean \bar{y}_{st} [1Mark]
 - ii) An unbiased estimation of variance for sample mean, $V(\bar{y}_{st})$ [3Marks]
- c) Suppose the proportion of units with desired attributes in stratum 1, 0.5 in stratum 2 and 0.8 in stratum 3. If the cost of obtaining observations are £10, £15 and £20 in stratum 1, 2 and 3 respectively and strata sizes are respectively, $N_1 = 1000, N_2 = 1200,$ and $N_3 = 800,$
- i) Estimate the population proportion P. [1Marks]
 - ii) Find the sample size n and the strata sample sizes n_1, n_2 and n_3 that will give the desired bound of 0.1 at, minimum cost. [4Marks]

SECTION B (39 marks): Answer any THREE questions. All Questions carry equal marks

QUESTION THREE (13 Marks)

- a) In a survey of the prevalence of drug use, randomly sampled high-school students were asked if they had ever used any hard drugs or not. Their responses were completely confidential.
- i) How many students should be included in the sample if the investigator wishes to be 95% certain that the error of estimation of the proportion of hard-drug users does not exceed 0.02, if the investigator feels that about 10% of the students had tried hard-drug at least one time? [2Marks]

- ii) Suppose that a simple random sample of 100 students was actually taken and that 5 of them reported that they had used hard-drugs. Based on this information construct a 99% confidence interval for the proportion of hard-drug users among high-school students. [2Marks]
- b) From a list of 2300 names and addresses, simple random of 200 names showed investigation 20 wrong addresses. Estimate the total number of addresses that need correction in the list and find the standard error of your estimate. [3Marks]
- c) A zoologist wishes to estimate the mean blood sugar level of species of animals when injected with a specified dosage of adrenaline. A sample of 50 animals of common breed were injected with adrenaline and their blood-sugar measurements were recorded in units of milligrams per 100 millimeters of blood the mean and the standard deviation of these measurements were found to be 125 and 10 respectively. Calculate error bound of your sample mean and 95% confidence interval for population means. [3Marks]
- d) A simple random sample of 30 households was drawn from a certain village in Kenya and the number of persons per house hold counted in the sample as follows;

No. of house hold	4	8	6	5	3	2	1	1
No of persons	1	2	3	4	5	6	7	8

- i) Estimate the total number of people in the village if there are 24600 households in the village. [1Mark]
- ii) Ignoring fpc, calculate the 95% confidence interval for the population means. [2Marks]

QUESTION FOUR (13 Marks)

- a) i) Explain the main sources of errors in a sample survey. What can you do to minimize these errors? [3Marks]
- ii) Distinguish between sampling and non-sampling errors. What are the remedies? [2Marks]
- b) i) Distinguish between random and non-random sampling methods? [3Marks]
- ii) What are the main advantages and disadvantages of random sampling methods? [3Marks]



iii) Non-random sampling methods are commonly used by non-mathematicians and non-statisticians. Why? [2Marks]

QUESTION FIVE (13 Marks)

a) For a finite population consisting of elements 3, 6, 0, and 5, a simple random sample of size 2 was drawn.

i) Enumerate all possible samples. [1Marks]

ii) Calculate sample means and an estimate of population means, \bar{Y} [2Marks]

iii) Show that \bar{y} is an unbiased estimator of population mean, \bar{Y} [2Marks]

iv) Calculate sample variances and estimator for population variance, S^2 [2Marks]

v) Show that s^2 is unbiased estimator for population variance, S^2 [1Mark]

b)

i) Explain how you can collect a systematic random sample of size $N = nk$ using every k^{th} sampling method. [2Marks]

ii) What are the main advantages and disadvantages of a systematic sampling method?

[1Mark]

c) Interest was expressed in the amount of money students spend each month for housing. A simple random sample of 200 students from Alupe university college of 35000 provides the statistics mean $\mu=2000$, variance, $s^2 = 144$. Compute; Total amount spent by Alupe university students for housing. What is the standard error of your estimate? [2Marks]

QUESTION SIX (13 Marks)

a) In a population with $N=12$ and $K=3$, the values of y_i are 2,3,4 in stratum 1,3,5,6,8 in stratum 2 and 0,1,4,6,10 in stratum 3. If a random sample of 6 were drawn.

i) Find the optimum allocation, n_i , under Neyman allocation. [3Marks]

ii) If proportion allocation was used to collect the sample calculate n_1, n_2 , and n_3 ?

[2Marks]

- iii) Compute the estimates \bar{y}_{st} , for every possible sample that can be drawn under optimum and proportional allocations. [3Marks]
- iv) Show that $E(\bar{y}_{st}) = \bar{Y}$, population mean. [2Marks]

b) A sample collected the following data in a sample survey.

Observations	Stratum		
	1	2	3
	5	8	6
	4	5	12
	2	10	15
	6		8
			5

From a population with strata sizes $N_1 = 800$, $N_2 = 1200$ and $N_3 = 1000$

- i) Calculate (\bar{y}_{st}) [1Marks]
- ii) Compute 95% confidence interval for population mean. [2Marks]

QUESTION SEVEN (13 Marks)

a) A sample survey was conducted and the following observation made.

$p_1 = 0.6, p_2 = 0.5, p_3 = 0.4$ and $p_4 = 0.8$

The cost of selecting an observation is £9 for either strata 1 and 2 and £4 for either strata 3 and 4. If $N_1=100, N_2=200,$ and $N_4=80.$ ~~$N_3=100$~~ $N_3=120$

- i) Estimate the population proportion P, with a bound error of estimation equal to 0.1 [2Marks]
 - ii) Calculate optimal allocation and strata sample size n_1, n_2, n_3 and n_4 that will give the desired bound at minimum cost. [3Marks]
- b) A sampler estimated in advance that his field statistics would be

	Stratum		
	1	2	3
w_i	0.4	0.5	0.1
s	5	8	2
$c_i(\text{£})$	9	16	4

If a random sample of size $n = 300$ was collected, find

- i) n_1, n_2 and n_3 that minimize variance for fixed cost. [2Marks]
- ii) Calculate the minimum variance that can be achieved with the allocations. [3Marks]
- iii) If the overhead cost was £100, calculate the cost of the survey. Let cost, $C = C_0 + \sum c_i n_i$ [3Marks]

