



# 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 PAPER

1

END OF SEMESTER EXAMINATION

**SERIES: 2015/2016**

**TIME: THREE HOURS**

**DATE: APRIL 2016**

## 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 ONE and any other two questions.

**Do not write on the question paper.**

## Question ONE (30 Marks)

- Briefly explain six principal steps in a sample survey. (3marks)
- A sampling frame is subject to several types of defects. Broadly classify these defects into four classes. (4 marks)
- Explain briefly four principles that should be followed in stratifying a population. (4 marks)
- A simple random sample of 2055 farms was selected from the 75 308 farms in a region, and the number of cattle ( $y$ ) and the total area ( $x$ ) were recorded for each farm. The results were as follows.

Sample total number of cattle,  $\sum y_i$  25751

Sample total area (hectares),  $\sum x_i$  62989

The sums of the squares are  $\sum y_i^2 = 596\ 737$  and  $\sum x_i^2 = 2\ 937\ 851$ , and the sum of the products is  $\sum x_i y_i = 1\ 146\ 391$ . The total area under cattle in this region is 2353365 hectares.

- (i) Estimate the total number of cattle in the region, and the standard error of your estimate, using
- (1) the simple random sample mean, (5 marks)
  - (2) the ratio estimator.

[Note. An estimate of the variance of a ratio estimator of the total is given by

$$\frac{N(N-n)}{n(n-1)} \left\{ \sum y_i^2 - 2r \sum x_i y_i + r^2 \sum x_i^2 \right\}, \text{ where } r = \frac{\sum y_i}{\sum x_i} \text{ and the other symbols have}$$

their usual meanings.] (5 marks)

- e) (i) Briefly explain how you can select a stratified random sampling. (3marks)
- (ii) Explain the advantages and limitations of sampling (6 marks)

### Question TWO

d) 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 tonnes)			
	$N_h$	$n_h$	$\bar{y}_h$	$S_h$
1	92	11	166.6	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.

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, and find the variance of  $\bar{y}_{st}$ . [Results from simple random sampling may be assumed without proof] (7 marks)

- (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. (7 marks)
- (iii) Compute the sample sizes in the strata if proportional allocation had been used for this survey. Give brief reasons why a stratified sample using proportional allocation would be more efficient than a simple random sample of 364 units. Explain briefly whether the allocation actually used has been effective in improving precision compared with a proportional allocation. (6 marks)

### Question THREE

a) A simple random sample of 10 hospitals was selected from a population of 33 hospitals that had received state funding to upgrade their emergency medical services. Within each of the selected hospitals, the records of all patients hospitalized in the past 12 months for traumatic injuries (i.e. accidents, poisonings, violence, burns, etc) were examined. The numbers of patients hospitalized for trauma conditions and the numbers discharged dead for the selected hospitals are given below.

Hospital	Number of patients hospitalized for trauma conditions	Number with trauma conditions discharged dead
1	560	4
2	190	4
3	260	2
4	370	4
5	190	4
6	130	0
7	170	9
8	170	2
9	60	0
10	110	1

- (i) Explain why this design may be considered as a cluster sample. What are the first-stage and second-stage units? (2 marks)
- (ii) Obtain a point estimate and an approximate 95% confidence interval for the total number of persons hospitalised for trauma conditions for the 33 hospitals. (11marks)
- b) Give reasons why, for this survey, cluster sampling might be preferred to stratified random sampling. What might be the drawbacks of cluster sampling? Discuss, with reasons, any improvements you might make if another survey was being planned on the same topic

### Question FOUR (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 in case of

(i) Proportional allocation (4 marks)

(ii) Optimum allocation (6 marks)

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.

(10marks)

### Question FIVE

a) 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 be preferred. (7 marks)

b) (i) Explain what is meant by a *quota sample*, a (linear) *systematic random sample*. Are these equal probability selection methods? Why or why not? (6marks)

(ii) Discuss any practical difficulties that might arise in planning a survey to study family income in a mixed urban and rural population. Explain how and why stratification and clustering might be used in such a survey. (5 marks)

(iii) What is meant by *post-stratification* (i.e. post-hoc stratification)?

(2marks)