



OFFICE OF THE DEPUTY PRINCIPAL  
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2020 /2021 ACADEMIC YEAR

FOURTH YEAR SECOND SEMESTER REGULAR EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE (APPLIED STATISTICS WITH  
COMPUTING)

COURSE CODE: STA 425

COURSE TITLE: SAMPLING THEORY AND METHODS II

DATE: 14/7/2021

TIME: 1300-1600HRS

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 4 PRINTED PAGES

PLEASE TURN OVER

**REGULAR – MAIN EXAM****STA 425: SAMPLING THEORY AND METHODS II****STREAM: ASC****DURATION: 3 Hours****INSTRUCTION TO CANDIDATES**Answer **ALL** questions from section A and any **THREE** from section B.**SECTION A [31 Marks]. Answer ALL questions.****QUESTION ONE (16 MARKS)**

- (a) Define the following terms as used in sample surveys
- (i) Sampling bias (2 marks)
  - (ii) Response bias (2 marks)
  - (iii) Under coverage error (2 marks)
- (b) Under what important situations are multiphase designs most appropriate to be used? (4 marks)
- (c) In a study to investigate the effect of a new food of supposedly higher nutritional value for chicken, a total of 100 chicken were weighed before the new food was introduced. A few weeks later after the food was introduced, the farm manager wanted to check if there was any gain in weight before continuing the study. A random sample of  $n=10$  chicken was taken and weighed. The total weight of the 100 chicken at the beginning of the study was  $X=250\text{kg}$ .

pre(X)	1.50	1.80	2.12	1.64	1.52	1.82	1.94	1.73	1.83	1.82
post (Y)	2.80	2.20	3.46	3.14	2.67	2.79	3.10	3.62	3.2	3.54

- (i) Find the value of the ratio estimator (2 marks)
- (ii) Obtain the regression estimator for the population mean (4 marks)

**QUESTION TWO: (15 MARKS)**

- (a) Consider a population of  $N = 1,000$  sampling units where you want to obtain a systematic random sample of size  $n = 100$
- (i) How many systematic random samples are there and show using a diagram what sampling units they consist of (2 marks)
  - (ii) Use this example to explain the relationship between cluster sampling and systematic random sampling (2 marks)
  - (iii) Explain why variances are difficult to calculate for systematic random sampling (2 marks)
- (b) What are the advantages of multi-stage sampling in national demographic surveys? (3 marks)
- (i) Discuss how non-coverage errors can be minimized in sample survey designs with special emphasis on rare events or conditions in the target population (6 marks)

**SECTION B: (39 MARKS)****QUESTION THREE (13 MARKS)**

- (a) Write down the equation of the population mean (2 marks)
- (b) What is the model regression equation relating  $y$  and  $x$  that is the basis of this estimator? (2 marks)
- (c) When should one decide that the ratio estimator should be used instead of the regression estimator? (2 marks)
- (d) Suppose a sociologist wants to estimate the total number of home-schooled children in a town based on two stage sampling. First a small pilot study is to be done. At the first stage 4 blocks are sampled at random out of the total of 300 blocks existing in the town. At the second stage 4 households are sampled at random out of each sampled block. (It is also known that there is a total of 3950 households in the town). The data obtained is shown below

Blocks	Number of households in block	Number of households in sample	Number of home school children
1	18	4	1,0,1,2
2	14	4	0,4,4,1
3	9	4	1,0,2,6
4	12	4	0,2,5,8

- (i) Estimate the average number of home-schooled children in the town using each of the two methods available to you. (5 marks)
- (ii) Evaluate the value of the ratio estimator (2 marks)

**QUESTION FOUR (13 MARKS)**

A restaurant chain wants to estimate the average employee satisfaction with their job (the scale is from 1 to 7). They have 100 restaurants and the total number of employees in the chain is 5260. They use simple random sampling to sample 6 restaurants. They then used simple random sampling to sample and interview about 10% of the employees in those restaurants. The data are given as follows.

Restaurant	$M_i$	$m_i$	Employee Satisfaction	$\bar{y}_i$	$s_i$
1	54	10	5, 7, 6, 5, 4, 7, 6, 6, 4, 5		
2	48	10	7, 7, 7, 6, 5, 4, 7, 7, 6, 6		
3	68	14	5, 6, 5, 6, 4, 5, 6, 5, 4, 5, 4, 6, 5, 6		
4	70	14	6, 5, 7, 6, 7, 6, 5, 7, 5, 7, 6, 5, 7, 6		0.83
5	52	10	4, 5, 4, 5, 5, 6, 5, 4, 4, 4		
6	62	12	5, 7, 6, 7, 4, 3, 1, 5, 4, 6, 4, 5		1.71

- a) Fill in the missing values for  $\bar{y}_i$  and  $s_i$  (6 marks)
- b) Find the unbiased estimator for the population total (4 marks)
- c) Find the ratio estimator for the population mean (3 marks)

**QUESTION FIVE (13 MARKS)**

- (a) Discuss major sources of non sampling errors in surveys (7 marks)  
 (b) How can you improve these errors as a researcher? Illustrate using examples (6 marks)

**QUESTION SIX (13 MARKS)**

a) A sociologist wants to estimate the average yearly vacation budget for each household in a certain city. It is given that there are 2,100 households in the city. The sociologist marked off the city into 200 blocks and treated them as 200 clusters. He then randomly sampled 8 clusters interviewing every household living in that cluster. The data are given in the table below:

Clusters	Number of Households ( $M_i$ )	Total vocational budget per cluster $y_i$
1	7 ✓	12000
2	9 ✓	15000
3	5 ✓	8000
4	8 ✓	13000
5	12 ✓	18000
6	5 ✓	7000
7	4	6000
8	8	13000

- (i) Find the ratio estimator for the average yearly vacation budget for each household in that city (4 marks)  
 (ii) Also estimated variance for the ratio estimator (5 marks)
- b) Give two reasons why cluster sampling is an effective design (4 marks)

**QUESTION SEVEN (13 MARKS)**

- (a) Define the term double sampling technique (2 marks)  
 (b) Suppose it is desired to estimate the total biomass of vegetation and average biomass ( $gm/m^2$ ) on a  $1000 m^2$  area. A systematic sample of twenty  $1m^2$  plots is selected and a visual estimate of the total grams of biomass is recorded. Five of the twenty plots are then randomly selected, and the total biomass is carefully determined on these 5 plots. The visual and actual measurements of total biomass are given below. We are given that

$$N = 1000 \text{ and } \sum_{i=1}^{n=20} x_i = 1570.$$

Plots	1	2	3	4	5
Visual ( $x_i$ )	20	80	150	40	80
Actual ( $y_i$ )	14	62	155	36	71

- (i) Find the ratio estimator for double sampling (2 marks)  
 (ii) Compute ratio estimate of the population total (3 marks)  
 (iii) Find estimated variance of the ratio estimator (6 marks)

\*\*\*\*\*