



OFFICE OF THE DEPUTY PRINCIPAL
ACADEMICS, STUDENT AFFAIRS AND RESEARCH

UNIVERSITY EXAMINATIONS

2021 /2022 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: 6TH JUNE, 2022 **TIME: 1400 – 1700 HRS**

INSTRUCTION TO CANDIDATES

- SEE INSIDE

THIS PAPER CONSISTS OF 5 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 (15 MARKS)

- a) State any two advantages of multi-stage sampling (2 marks)
- b) Discuss any four the sources of non-sampling errors (8 marks)
- c) Show the efficiency of ratio estimator in comparison to SRSWOR (5 marks)

QUESTION TWO (15 MARKS)

- a) Define the following terms as used in sample surveys
 - i) Sampling bias (2 marks)
 - ii) Non-sampling errors (2marks)
 - iii) Voluntary response bias (2 marks)
- b) In a study to estimate the total sugar content of a truckload of oranges a SRS of $n = 10$ oranges was juiced and weighed as shown below. The total weight of all the oranges obtained by first weighing the truck loaded and unloaded was found to be 1800 pounds. You are required to;

Obs	1	2	3	4	5	6	7	8	9	10
Weight of orange (X)	1.54	1.78	2	1.64	1.51	1.79	1.92	1.73	1.54	1.8
Sugar content (Y)	2.81	2.2	3.4	3.14	2.68	2.79	3.11	3.6	3.2	3.5

- i) Plot the scatter plot and comment on the nature of relation (1 mark)
- ii) Estimate the ratio estimator (3 marks)
- iii) Estimate Y, the total sugar content (3 marks)
- iv) Provide the standard error of the estimate (3 marks)

SECTION B: (39 MARKS)

QUESTION THREE (13 MARKS)

- a) State any two sources of demographic data (2 marks)
- b) Discuss any three approaches of measurement and control of errors (6 marks)
- c) Large mammals in open habitat are often surveyed from aircraft. As the aircraft flies over a selected strip, all animals of the species within a prescribed distance of the aircraft path are counted, the distance sometimes being determined by markers on the wing struts of the aircraft. Because of the irregularities in the shape of the study area, the strips to be flown may be of varying lengths. One may select units (strips) with probability proportional to their lengths by randomly selecting n points on a map of the study region and including in the sample any strip that contains a selected point. The draw-by-draw selection probability for any strip equals its length times its width divided by the area of the study region. The Sample Observations is given in the table below:

y_i	Length	p_i
60	5	0.05
60	5	0.05
14	2	0.02
1	1	0.01

Compute:

- i) Probability-proportional- to-size (PPS) estimate (2 marks)
- ii) The estimated variance (3 marks)

QUESTION FOUR (13 MARKS)

- a) Discuss any three practical cases for the use of Multistage designs (6 marks)
- c) Consider the following data set of 10 number of workers in the factory and its output.

Factory no.	Number of workers (X) (in thousands)	Industrial production (in metric tons) (Y)
1	2	30
2	5	60
3	10	12
4	4	6
5	7	8
6	2	13
7	3	4
8	14	17
9	11	13
10	6	8

Illustrate the selection of units (at least 2 units) using the cumulative total method. (7 marks)

QUESTION FIVE (13 MARKS)

For a juice company, the price they paid for apples in large shipments is based on the amount of apple juice from the load. 15 apples selected by simple random samples were weighed and also juiced. The total weight of the apple shipment was found to be 2000 pounds.

Apple	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	0.16	0.15	0.2	0.25	0.16	0.27	0.28	0.16	0.11	0.16	0.17	0.24	0.21	0.11	0.22
X	0.22	0.26	0.31	0.37	0.28	0.38	0.4	0.21	0.18	0.29	0.26	0.32	0.33	0.16	0.35

- a) From the table above:
- i) Obtain a point estimate of the total weight of the juice for the shipment of apples (3 marks)
 - ii) Calculate the 95% confidence interval. (5 marks)
- b) Write R code to
- i) Input the data and obtain the ratio estimate (5 marks)

QUESTION SIX (13 MARKS)

- a) State two reasons why double sampling is mainly used (2 marks)
- b) A sociologist wants to estimate the average yearly vacation budget for each household in a certain city. It is given that there are 3,100 households in the city. The sociologist marked off the city into 400 blocks and treated them as 400 clusters interviewing every household living in that cluster. The data are given in the table below:

Cluster	Number households	Total vacation budget per cluster	Cluster	Number households	Total vacation budget per cluster
1	7	12,000	13	8	12,340
2	9	15,000	14	4	5,000
3	5	8,000	15	6	8,900
4	8	13,000	16	9	14,000
5	12	18,000	17	3	4,000
6	5	7,000	18	10	11,400
7	4	6,000	19	4	5,000
8	8	13,000	20	7	13,000
9	14	22,000	21	6	8,900
10	6	9,800	22	5	8,700
11	3	7,000	23	7	10,000
12	13	18,000	24	6	9,200

- i) Find the ratio estimator for the average yearly vacation budget for each household in that city and estimated variance for the ratio estimator (5 marks)
- ii) Find the unbiased estimator for the average yearly vacation budget for each household in that city (3 marks)
- iii) Find the estimated variance for the unbiased estimator (3 marks)

QUESTION SEVEN (13 MARKS)

Consider a population consists of $N = 20$ plots of land, and we take an SRS of $n = 7$ plots, counting the number of animals and number of females on these 7 plots. In addition, we also count the number of animals on all 20 plots, without knowing what sex they are because it may be easy to count the number of animals on a plot, but hard to identify which are females. Assume all the plots are equal in size. Let the response variable: $y_i =$ the number of female animals on plot i , $i = 1, \dots, N$ and auxiliary variable: $x_i =$ the total number of animals on plot i , $i = 1, \dots, N$. The data for the SRS is given below:

y_i	7	12	4	6	19	7	5
x_i	10	18	10	12	25	15	10

Write R codes to:

- i) Input the data (3 marks)
- ii) Plot the number of animals against the number of females (2 marks)
- iii) Estimation of proportion of females (5 marks)
- iv) Estimation of total number of females (3 marks)
