Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE IN STATISTICAL SCIENCE

END OF SEMESTER EXAM 2021

STA 8104 DESIGN AND ANALYSIS OF SURVEYS

DATE: 17th December 2021 TIME: 3 HOURS

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 Marks)

a. Briefly define the following terms (6 marks)
    a. Sampling frame
    b. Deductive Imputation
    c. Available Case analysis

b. The following are amount of cotton in metric tonnes produces by provinces 1-10 in Zambia: 65, 99, 70, 19 84, 72, 37, 93, 20, 25 respectively. Using the random numbers: 401, 170, 128, 110, 485, select 5 provinces (5 marks)

c. The National Cereals and Produce Board has 130 silos in Kitale, each of the same size. The amount of maize is 15 silos selected at random were found to be: 28,23,25,33,31,18,22,29,30,22,26,20,21,28 and 25. Estimate the total amount of maize in the silos and calculate an approximate 95% CI for this total (8 marks)

d. Show that $Var_{opt}(\bar{y}_{st}) \leq Var_{prop}(\bar{y}_{st})$ where: (5 marks)

$$Var_{prop}(\bar{y}_{st}) = \frac{N-n}{Nn}\sum_{i=1}^{k}\frac{N_i S_i^2}{N}$$

And

$$Var_{opt}(\bar{y}_{st}) = \frac{1}{n}\left(\sum_{i=1}^{k} w_i S_i\right)^2 - \frac{1}{N}\sum_{i=1}^{k} w_i S_i^2$$

Where $Var_{prop}(\bar{y}_{st})$ and $Var_{opt}(\bar{y}_{st})$ are respectively: variances of the sample mean under simple random sampling with stratified mean under proportional allocation and optimal allocation.

e. The women's and men's questionnaires ask respondents how often they watch TV and listen to the Radio in the last week. Here is a research question that you wish to answer: Is exposure to TV and Radio media associated with lower fertility in Kenya, adjusting for other factors? Which datasets should you request access to? (1 mark)
f. What most increases the statistical power of a multi-stage cluster sample? (1 mark)

g. This document provides a list of all variables that are contained in a specific DHS file (e.g. the 2010 Rwanda Birth Recode File) with a summary definition of each variable (1 mark)

Question Two (15 marks)

Let $\bar{y}_{sy}$ be the estimate for the population mean $\bar{Y}$ from a systematic sample. Show that

$$var(\bar{y}_{sy}) = E(\bar{y}_{sy} - \bar{Y})^2 = \frac{(N-1)S^2}{N} - \frac{(n-1)}{n} S_{wsy}^2$$

Where $S_{wsy}^2 = \frac{S_1^2 + \cdots + S_k^2}{k}$

Question Three (15 Marks)

a. Discuss the ratio estimation method (5 Marks)

b. A research firm conducts a survey of food cost pre and during corona by taking a simple random sample of 6 basic food stuffs purchased in a large supermarket. Prices are recorded in two separate occasions, 6 months apart, the earlier ones being denoted x_i and the later one y_i. Assume that the population mean of the earlier price was 28.

X	Y
19	24
28	13
23	20
30	35
21	16
30	38

Assuming that the population mean of X is 24, Obtain the estimate of the population mean and variance of Y using regression estimation (10 marks).

Question Four (15 marks)

a. Discuss any three methods of dealing with missing data and their potential problems (7marks)

b. The following data contains some missing values. Using the FIML technique and the data, write down the equation to compute the first iteration values for the likelihood function -2LL in row 8 and row 16. **NB. Leave your answer in equation form as I do not expect you to find inverse of matrices.** (8 marks)

read	science	math	medicine
12	27	19	44
11	25	13	
25	56	23	60
15	37		52
10	31	21	25
	40	22	43
15	50	28	
21	25		46
20	39	25	51
25	36	23	56
	27	21	39
14	32		46
22	40	25	
25	31	25	44
25			25
16	23	21	34
27	26	23	59
17	25		35
13	21	20	
11	51	18	54

	read	science	math	medicine
read	33.41	8.61	14.39	21.94
science	8.61	103.06	15.69	53.39
math	14.39	15.69	12.74	6.02
medicine	21.94	53.39	6.03	118

Question Five (15 marks)

a. Consider the following data on wheat farms in Narok.

Farm size (hectares)	No. of farms	Mean wheat (hectares)	standard deviation
0-20	368	2.7	2.1
21-40	425	8.1	3.6
41-60	389	12.1	3.9
61-80	316	16.9	5.1
81-100	174	20.8	6.1
101-120	98	25.2	6.5
121+	138	31.8	9.1

For a stratified simple random sample of size 100, determine the appropriate stratum sizes under

a. Proportional allocation (5 marks)
b. Neyman allocation (10 marks)