Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES

MASTER OF SCIENCE IN STATISTICAL SCIENCE

END OF SEMESTER EXAMINATION

STA 8104 DESIGN AND ANALYSIS OF SURVEYS

DATE: 30th August, 2021 **TIME:** 3 HOURS

ANSWER QUESTION ONE AND ANY OTHER TWO QUESTIONS

QUESTION ONE (30 Marks)

a. Discuss any three types of repeated measurements (6 marks)

b. Define the following terms:
   a. Trueness (2 marks)
   b. Cluster Sampling (2 marks)

c. Factories in a certain region have the following number of workers: 65, 99, 70, 19 84, 72, 37, 93, 20, 25 respectively. Select 5 factories using the following random numbers: 501, 172, 138, 190, 485. (5 marks)

d. Show that in two stage sampling, show that $\bar{\bar{y}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}$ is an unbiased estimator for the population mean $\bar{Y}$. Where n is the number of sampled clustered and m is the cluster size. Assume that sampling is done without replacement (5 marks)

e. Suppose that the following summarized information is made available: n=25, N=275, $\bar{x} = 9.2, \bar{y} = 2.6, \sum_{i=1}^{25} x_i^2 = 2200, \sum_{i=1}^{25} x_i y_i = 500, \sum_{i=1}^{25} y_i^2 = 170.$ Estimate R and var(R) (4 marks)

f. Explain how systematic sampling can be viewed as a cluster sampling technique (3 marks)

g. 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)

h. What most increases the statistical power of a multi-stage cluster sample? (1 mark)

i. What 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)

Show that for a simple random sampling without replacement, $E(s^2) = S^2$, where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ (15 marks)

Question Three (15 marks)

a. Discuss the Full Information Maximum likelihood (FIML) technique for missing data (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 Four (15 marks)

A stratified population has 6 strata. The stratum sizes N_i and means $\bar{Y}_i$ *and* S_i^2 of some variable Y are as follows:

Stratum	N_i	$\bar{Y}_i$	S_i^2
1	117	7.3	1.31
2	98	6.9	2.03
3	74	11.2	1.13
4	41	9.1	1.96
5	45	9.6	1.74

i. Calculate the overall population mean and variance (5 marks)

ii. For a stratified simple random sample of size 80, determine the appropriate stratum sizes under

   a. Proportional allocation (4 marks)

   b. Neyman allocation (6 marks)

Question Five (15 Marks)

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.

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

a. Obtain the estimator r of the ratio R and calculate its sampling variance (7 marks)
b. Assuming that the population mean of X is 24, Obtain the estimate of the population mean of Y using regression estimation (8 marks).