MAASAI MARA UNIVERSITY

REGULAR UNIVERSITY EXAMINATIONS

2023/2024 ACADEMIC YEAR

FOURTH YEAR SECOND SEMESTER
SCHOOL OF PURE, APPLIED AND HEALTH SCIENCES.
DEGREE IN APPLIED STATISTICS WITH COMPUTING.

COURSE CODE: STA 4244-1

COURSE TITLE: SAMPLING THEORY AND METHODS II

DATE: **TIME:**

<u>INSTRUCTIONS TO CANDIDATES</u>

Answer Question ONE and any other TWO questions

This paper consists of FOUR printed pages. Please turn over.

QUESTION ONE (20 MARKS)

Assuming that the random sample $(x_i, y_i), i = 1,2, \dots, n$ is drawn by SRSWOR and a known population mean $\overline{X}$ of X and $\overline{Y}$ of Y. Given also that

$\varepsilon_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}} \quad and\ \varepsilon_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$

a.

(i) Show that $E(\varepsilon_0^2) = \frac{f}{n} C_Y^2$

Where $f = \frac{N-n}{N}, \quad S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \overline{Y})^2 \quad and\ C_Y = \frac{S_Y}{\overline{Y}}$ (4 mks)

(ii) Show that $E(\varepsilon_0 \varepsilon_1) = \frac{f}{n}\rho\, C_X C_Y$ where $C_X = \frac{S_X}{\bar{X}}$ and ρ is the correlation coefficient between X and Y. (5 mks)

b. Show that the ratio estimate $\widetilde{\widetilde{Y_R}}$ is the best linear unbiased estimator of $\overline{Y}$ when:

(i) The relationship between y_i and x_i is linear passing through the origin, that is $y_i = \beta x_i + e_i$ where $e_i's$ are independent with $E(e_i|x_i) = 0$ and β is the slope parameter.

(ii) And the line is proportional to x_i, that is, $Var(y_i|x_i) = E(e_i^2) = Cx_i$ where C is constant. (8 mks)

c. Give the procedure of sampling by Lahiri's method. (3 mks)

QUESTION TWO (15 MARKS)

(a) Give the advantages and disadvantages of Lahiri's method of sampling procedure. (4 mks)

(b) Let

Y_i: Value of study variable for the i^{th} unit of the population, $i = 1, 2, \dots, N$.

X_i: Known value of an auxiliary variable (size) for the i^{th} unit of the population.

P_i: Probability of selection of i^{th} unit in the population at any given draw and is proportional to size X_i.

$Z_i = \frac{Y_i}{NP_i}, i = 1, 2, \dots, N$.

Consider the varying probability scheme and with replacement for a sample of size n. Letting $y_r =$ the value of r^{th} observation in the sample and Pr =initial probability of selecting yr. Using

$z_r = \frac{y_r}{Np_r}, \quad r = 1, 2, \ldots, n,$ then,

(i) $\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ is an unbiased estimator of the population mean $\overline{Y}$. Prove. (4 mks)

(ii) Show that $Var\left(\overline{Z}\right) = \frac{\sigma_z^2}{n}$ (7 mks)

QUESTION THREE (15 MARKS)

(a) Given that $\varepsilon_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}} \quad and\ \varepsilon_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$

and that $\widetilde{\widetilde{Y_R}} = \frac{\bar{y}}{\bar{x}}\bar{X}$

where $\bar{X}$ is the population mean of X

$\bar{Y}$ is the population mean of Y

$\bar{x}$ is the sample mean of x

$\bar{y}$ is the sample mean of y

Writing $\widetilde{\widetilde{Y_R}}$ in terms of $\varepsilon's$

(i) Show that $\widetilde{\widetilde{Y_R}} = (1+\varepsilon_0)(1+\varepsilon_1)^{-1}\overline{Y}$ (2 mks)

(ii) Show that $\widetilde{\widetilde{Y_R}} = \overline{Y}(1+\varepsilon_0-\varepsilon_1+\varepsilon_1^2-\varepsilon_1\varepsilon_0+\cdots)$ (4 mks)

(iii) Show that if we assume that the higher powers ε_0 and $\varepsilon_1$more than 2 are negligibly small, then the Bias of $\widetilde{\widetilde{Y_R}}$ is given by

$Bias\left(\widetilde{\widetilde{Y_R}}\right) = E\left(\widetilde{\widetilde{Y_R}} - \overline{Y}\right) = \frac{f}{n}\overline{Y}C_X(C_X - \rho C_Y)$ (4 mks)

(b) Given the Mean Squared Error of $\widehat{Y_R}$ is given by

$$MSE\left(\widehat{Y_R}\right) = \sum_{i=1}^{N}(Y_i - RX_i)^2$$

$= \sum_{i=1}^{N}(Y_i - \overline{Y}) + R^2\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 - 2R\sum_{i=1}^{N}(X_i - \overline{X})\,(Y_i - \overline{Y}).$

Prove by assuming that $\overline{Y} = R\overline{X}$ (5 mks)

QUESTION FOUR (15 MARKS)

Defining the product estimator of the population mean $\overline{Y}$ as

$\left(\widetilde{\overline{Y_p}}\right) = \frac{\overline{yx}}{\overline{X}}$ assuming that the population mean $\overline{X}$ is known and letting

$\varepsilon_0 = \frac{\bar{y}-\bar{Y}}{\bar{Y}} \quad and\ \varepsilon_1 = \frac{\bar{x}-\bar{X}}{\bar{X}}$

a.

(i) Show that the $Bias\left(\widetilde{\overline{Y_p}}\right)$ is given by

$Bias\left(\widetilde{\overline{Y_p}}\right) = \frac{f}{n\overline{X}} S_{XY}$ (6 mks)

(ii) Writing $\widetilde{\overline{Y_p}}$ in terms of ε_0 and ε_1,

Show that

$MSE\left(\widetilde{\overline{Y_p}}\right) = \overline{Y}^{\,2} E(\varepsilon_1^2 + \varepsilon_0^2 + 2\varepsilon_1\varepsilon_2$ (4 mks)

b. Compare the variances of sample mean under SRSWOR with that of the product estimator. (5 mks)