MAASAI MARA UNIVERSITY
REGULAR UNIVERSITY EXAMINATIONS
2022/2023 ACADEMIC YEAR
FOURTH YEAR SECOND SEMESTER
SCHOOL OF PURE APPLIED AND HEALTH SCIENCES
THE DEGREE OF BACHELOR OF SCIENCE IN APPLIED STATISTICS

STA 4244 SAMPLING THEORY AND METHODS II

Instructions to candidates:

Answer Question 1. And any other TWO.
All Symbols have their usual meaning

DATE: TIME:

Question 1(30 Marks)

(a) Show what ratio estimator is a better estimate of $\bar{Y}$ than sample mean based on SRSWOR if

$$MSE(\hat{\bar{Y}}_R) = Var_{SRS}(\bar{y})$$

(5 Marks)

(b) Given that the mean squared errors (MSE) of the ratio estimate of the population mean $\bar{Y}(ie\hat{\bar{Y}}_R)$ is

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n}\bar{Y}^2(C_Y^2 + C_X^2 - 2_\rho C_X C_Y)$$

Show that

$$MSE(\hat{\bar{Y}}_R) = \frac{N-n}{nN(N-1)}\sum_{i=1}^{n}(Y_i - RX_i)^2$$

where $C_X = \frac{S_X}{\bar{X}}$ and $C_Y = \frac{S_Y}{\bar{Y}}$ **(5 Marks)**

(c) The ratio estimator $\hat{\bar{Y}}_R$ is the best linear unbiased estimator of $\bar{Y}$ when

(i) the relationship between y_i and x_i is linear passing through the origin, ie, $y_i = \beta x_i + e_i$, where $e_i's$ are independent with $E(e_i|x_i) = 0$ and β is the slope parameter. **(4 Marks)**

(ii) (ii) this line is proportional to x_i ie

$$Var(y_i|x_i) = E(e_i^2) = Cx_i$$

where C is a constant **(10 Marks)**
Prove the above

(d) (i) Give the difference between the methods of simple random sampling (SRS) and probability proportional to a given measure of size (PPS) **(2 Marks)**

(ii) Under PPS there is a method used to draw a sample with replacement called cumulative total method. List its procedure **(3 Marks)**

(iii) List the procedure of Lahiri's method of selecting a sample **(3 Marks)**

(iv) Prove that by Lahiri's method, the probability that no unit is selected at a trial is given by

$$1 - \frac{\bar{X}}{M}$$

, where $M = Max x_i, i = 1, 2, ..., N$ is the maximum of sizes of N units in the population or some convenient number greater than M

Question 2 (20 Marks)

(a) (i) Let y_r be the value of the r^{th} observation on the study variables in the sample and p be its initial probability of selection. Define

$$z_r = \frac{y_r}{Np_r}, r = 1, 2, ..., n$$

then

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$$

is an unbiased estimator of the population mean $\bar{Y}$ (**3 Marks**)

(ii) Prove also that $Var(\bar{z}) = \frac{\sigma_z^2}{n}$ (**10 Marks**)

(b) Show that $\frac{\sigma_z^2}{n}$ is an unbiased estimator of the variance of $\bar{z}$ given that

$$\frac{\sigma_z^2}{n} = \frac{1}{n-1}\sum_{r=1}^{n}(z_r - \bar{z})^2$$

is the unbiased estimator of variance of $\bar{z}$ (**7 Marks**)

Question 3 (20 Marks)

(a) (i) Given that the estimate of population total is

$$\hat{Y}_{tot} = \frac{1}{n}\sum_{r=1}^{n}\left(\frac{y_r}{p_r}\right) = N\bar{Z}$$

Prove that $\hat{Y}_{tot}$ is an unbiased estimator of population total,Y_{tot} (**3 Marks**)

(ii) Prove also that the variance $\hat{Y}_{tot}$ which is the unbiased estimator of the population total is given by

$$Var(\hat{Y}_{tot}) = \frac{1}{n}\left[\sum_{i=1}^{N}\frac{Y_i^2}{P_i} - Y_{tot}^2\right]$$

where p_i is the probability of selection of the i^{th} unit in the population at a given draw
Y_i is the value of study variable for the $i^{th}4$ unit of the population $i = 1, 2, ..., N$ and $Z_i = \frac{Y_i}{NP_i}$, $i = 1, 2, ..., N$ (**5 Marks**)

(b) (i) Describe briefly the varying probability scheme without replacement (PPSWOR) (**5 Marks**)

(ii) Let U_i be the i^{th} unit, P_i be the probability of selecting U_i at the first draw $i = 1, 2, ..., N, \sum_{i=1}^{N} P_i = 1$, $P_i(r)$ is the probability of selecting U_i at the r^{th} draw, show that $P_i(2)$ the probability of selecting U_i at 2^{nd} draw is

$$P_i(2) = P_i \left[\sum_{j=1}^{N} \frac{P_j}{1 - P_j} - \frac{P_i}{1 - P_i} \right]$$

(**7 Marks**)

Question 4 (20 Marks)

(a) (i) Given that $\varepsilon_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$ and $\varepsilon_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}$ and that when we follow SRSWOR so that $E(\varepsilon_0) = E(\varepsilon_1) = 0$, show that $E(\varepsilon_0^2) = \frac{f}{n} C_Y^2$, where

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

(**4 Marks**)

(ii) Show also that $E(\varepsilon_0 \varepsilon_1) = \frac{f}{n} \rho \, C_X C_Y$ where ρ is the population correlation coefficient between X and Y (**6 Marks**)

(b) (i) Starting $\hat{\bar{Y}}_R = (1 + \varepsilon_0)(1 + \varepsilon_1)^{-1} \bar{Y}$, show that for the purpose of getting estimation error $\hat{\bar{Y}}_R$, we have

$$\hat{\bar{Y}} - \bar{Y} = \bar{Y}(\varepsilon_0 - \varepsilon_1 + \varepsilon_1^2 + \varepsilon_1 \varepsilon_0 + ...)$$

(**4 Marks**)

(ii) Given that $MSE(\hat{\bar{Y}}_R) = E(\hat{\bar{Y}}_R - \bar{Y})$ and that

$$\hat{\bar{Y}}_R - \bar{Y} = \bar{Y}(\varepsilon_0 - \varepsilon_1 + \varepsilon_1^2 + \varepsilon_0 \varepsilon_1 + ...)$$

Prove that

$$MSE(\hat{\bar{Y}}_R) = \frac{\bar{Y} f}{n} \left[C_X^2 + C_Y^2 - 2\rho \, C_X C_Y \right]$$

where $C_X = \frac{S_X}{\bar{X}}$, $C_Y = \frac{S_Y}{\bar{Y}}$ and ρ is the correlation coefficient between X and Y (**6 Marks**)