STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES (SIMS)

MASTERS IN STATISTICAL SCIENCE

END OF SEMESTER EXAMINATION

STA 8104: DESIGN AND ANALYSIS OF SURVEYS

DATE: 8th December, 2023 Time: 3 Hours

Instructions

1. This examination consists of **4** questions.
2. Answer **Question 1 (COMPULSORY)** and any other **TWO** questions.

Question 1 (30 Marks)

(a) Show that in Simple Random Sampling Without Replacement (SRSWOR), the probability of a specified unit being chosen at any draw is equal to the probability of choosing it at the first draw (4 Marks)

(b) Define $a_i = \{1,$ if the i^{th} unit is in the sample, and 0, otherwise

Show that in Simple Random Sampling Without Replacement

(i) the sample mean is an unbiased estimator of the population mean, $E(\bar{y}) = \bar{Y}$ (4 Marks)

(ii) the sample variance is unbiased estimator of population variance, $E(s^2) = S^2$ (5 Marks)

(c) A sample of 30 students is to be drawn from a population of 300 students belonging to two colleges A, and B. The means and standard deviations of their marks are given below.

	Total number of students	$\bar{Y}_i$	S_i
College A	200	30	10
College B	100	60	40

Verify that Neyman's allocation is more efficient than proportional allocation (6 Marks)

(d) Suppose a recruitment drive to hire some auditors at the entry level is to be done. A decision is to be made on a competitive salary offer for these new auditors. From talking to some Human Resource professionals, a rough estimate obtained is that most new hires are getting starting salaries in the range $38,000-42,000 and the average is around $39,000. The standard deviation is estimated to be around $3000. Determine the sample size if this researcher is to be 95% confident about the average salary and is willing to tolerate an estimate that is within $500 (plus or minus) of the true estimate. (5 Marks)

(e) A simple random sample of 100 households located in a city recorded the number of people living in the household, X, and the weekly expenditure for food, Y. It is known that there are 100,000 households in the city. In the sample

$\sum X_i = 320$, $\sum Y_i = 10{,}000$, $\sum X_i^2 = 1250$, $\sum Y_i^2 = 1{,}100{,}000$,
$\sum X_i Y_i = 36{,}000$

(i) Estimate the ratio $r = \frac{\mu_y}{\mu_x}$ (3 Marks)

(ii) Estimate the variance of the ratio estimator, R (5 Marks)

Question 2 (15 Marks)

(a) Let Y_i be the value of the characteristic under study for the i^{th} unit of the population, X_i be the value of the auxiliary characteristic of the i^{th} unit of the population. Show that the ratio estimator is a biased estimator of the population mean, $\bar{Y}$ (3 Marks)

(b) Let $y_1, y_2, y_3, \ldots y_N$ be a sample drawn from a super population consisting of N independent random variables. Suppose the joint distribution is defined by

$$Y_i = \begin{cases} E(Y_i) = \beta x_i \\ Var(Y_i) = \sigma^2(x_i) \\ Cov(Y_i, Y_j) = 0, i \neq j \end{cases}$$

Determine that the best linear unbiased estimator of the population total T under the super population model (6 Marks)

(c) A company has 10 departments having 10, 20, 30, 15, 25, 35, 65, 55, 50, and 5 employees. It is desired to select a sample of size 4 (departments) with replacement and with probabilities proportional to the number of employees in each department. Explain the procedure on how to select the desired departments using the following extract of random numbers and provide the sampled departments for a survey.

(The random numbers chosen are: 065, 103, 302, 075, 009) (6 Marks)

Question 3 (15 Marks)

(a) Let π_i be the probability of including unit i in the sample. Define α_i = {1, if the ith unit is in the sample, and 0 otherwise. Further, let Y and X be the survey variable and auxiliary variable respectively. Prove that the Horvitz-Thompson estimator $\bar{y}_{HT} = \sum_{i \in s} \frac{y_i}{N\pi_i}$ is unbiased for $\bar{Y}$ (5 Marks)

(b) Derive the variance of the Horvitz –Thompson, $Var\left(\bar{y}_{HT}\right)$ (5 Marks)

(c) Suppose a distribution random variable $Y_i = \begin{cases} E(Y_i) = \beta x_i \\ Var(Y_i) = \sigma^2(x_i) \\ Cov(Y_i, Y_j) = 0, i \neq j \end{cases}$

where x_i is an auxiliary variable corresponding to the i^{th} survey variable. Show that the Horvitz-Thompson estimator $\bar{y}_{HT}$ is unbiased under the above model (5 Marks)

Question 4 (15 Marks)

(a) Consider a population of size N=3. The values y_1, y_2, y_3 are 0, 2, and 1 respectively. Consider the following two estimators of $\bar{Y}$. For a simple random sample without replacement of size 2 from the population,

$$t = \begin{cases} t_1 = \dfrac{y_1}{2} + \dfrac{y_1}{2} & , S_1 \quad occurs \\ t_2 = \dfrac{y_1}{2} + \dfrac{2y_3}{3} & , S_2 \quad occurs \\ t_3 = \dfrac{y_2}{2} + \dfrac{y_3}{3} & , S_3 \quad occurs \end{cases}$$

Where $S_1 = \{1,2\}, S_2 = \{1,3\}$ and $S_3 = \{2,3\}$

(i) Show that the sample mean $\bar{y}$ is unbiased for the population mean $\bar{Y}$ (2 Marks)

(ii) Show that the sample mean t is unbiased for the population mean $\bar{Y}$ (3 Marks)

(iii) Show that **t** is more efficient estimator than $\bar{y}$, the sample mean (5 Marks)

(b) Let Yi be the Y- value of the i^{th} unit of the population $i = 1,2,.., N$. y_i be the i^{th} unit of the sample, $i = 1,2,.., n$. p_i be the probability of selecting the i^{th} unit of population and $\bar{Z} = \dfrac{y_i}{Np_i}, i = 1,2,.., N.$ Suppose the samples are drawn with replacement and with unequal probabilities.

(i) Show that an unbiased estimator of the population mean is given by $\bar{Z} = \dfrac{1}{n}\sum_{1}^{n} {y_i}/{Np_i}$

(2 Marks)

(ii) Show that the estimator of var ($\bar{Z}$)=σ_z^2 / n (3 Marks)