UNIVERSITY EXAMINATIONS 2023/2024

FIRST YEAR SECOND SEMESTER EXAMINATION FOR DEGREE OF MASTERS OF SCIENCE IN APPLIED STATISTICS

SMS 5175: DESIGN AND ANALYSIS OF SAMPLE SURVEYS

DATE: APRIL 2023   **TIME: 2 HOURS**

INSTRUCTIONS: Answer Question ONE and any other Three questions.

QUESTION ONE (15 MARKS)

(a) Given that $Y \sim Niid(\beta, \sigma_1^2)$, show that $\bar{x}$ is a consistent estimator of β.   (3 marks)

(b) Show that the approximate for a simple random size is given as $n \geq \dfrac{z^2_{\frac{\alpha}{2}} s^2}{d^2}$

(4marks)

(c) For the population total X_T, the estimator is $N\bar{x}_{st}$ in a stratified random sample, find $Var(N\bar{x}_{st})$ (4 marks)

(d) Let $\bar{x}_{sys}$ be the estimate of $\bar{X}$ from a systematic sample, show that $E(\bar{x}_{sys}) = \bar{X}$.

(4 marks)

QUESTION TWO (15 MARKS)

a) If $\hat{\theta}$ is an estimator of the parameter θ, outline two conditions that must satisfied by var $(\hat{\theta})$ (2 marks)

b) Show that the variance of the simple random sample estimator, $\bar{x}$ is given by $var(\bar{x}) = \dfrac{s^2}{n}(1 - f)$ where f is the sampling fraction.   (6 marks)

(b) Signatures to a petition were collected on 676 sheets each sheet had enough space for 42 signatures, but on many sheets, a smaller number of signatures had been collected. The numbers of signatures per sheet were counted on a random sample of 50 sheets (about a 7% sample). The results are given in the table below

y_i	42	41	36	32	29	27	23	19	16	15	14	11	10	7	6	5	4	3
f_i	23	4	1	1	1	2	1	1	2	2	1	1	1	1	3	2	1	1

Estimate the total number of signatures to the petition. (7 marks)

QUESTION THREE (15 MARKS)

(a) Outline two merits of a stratified random sample. (2 marks)

(b) Show that a stratified random sample mean, $\bar{x}_{str}$ is more efficient than the simple random sample mean, $\bar{x}$ as the estimator of the population mean, $\bar{X}$. (6marks)

(c) A daily newspaper conducts a survey of food costs by taking a simple random sample of 48 basic food stuffs purchased in a large supermarket. Prices (in dollars) for these items are recorded in two separate occasions, three months apart, the earlier ones being denoted x_i and the later y_i. The sample ratio $r = \frac{\bar{y}}{\bar{x}}$ gives an indication of change of these basic food prices over three months period in the form of an estimate of the population ration R of the mean prices of food stuffs on the two occasions. The following results were obtained:

$\bar{x} = 11.41, \bar{y} = 12.07, \sum x_i^2 = 8431.7 \sum x_i y_i = 8564, \sum y_i^2 = 9270.6\,\text{m}$ and $n = 48$.

Estimate the 95% confidence interval for R (7 marks).

QUESTION FOUR (15 MARKS)

(a) Explain the types of non response in sample surveys. (2 marks)

(b) In two stage sampling with unequal first sampling units, let $\hat{\bar{y}} = \frac{1}{n}\sum_{i=1}^{n} y_i$ be an estimator of the population mean $\bar{Y}$ where $\bar{y}_i = \frac{1}{mi}\sum_{i=1}^{mi} y_{ij}$ with y_{ij} being the jth

unit in the ith cluster. Test $\dfrac{\wedge}{y}$ for the Unbiasedness. Suppose that we now define

another estimator, $\dfrac{\wedge}{y_1} = \dfrac{1}{n}\sum_{i=1}^{n}\dfrac{M_i}{m}\bar{y}_l$ with $\dfrac{m_i}{m}$ being the weight of the ith stratum.

Compare $\hat{\bar{y}}$ to $\widehat{y1}$ and choose the best estimator of the population mean $\bar{Y}$ using Unbiasedness criteria. (6 marks)

(c) All the farms in a country are stratified by farm size and the mean number of hectares of wheat per farm in each stratum, with the following results.

Farm size (ha)	No. of farms	Mean wheat (ha)	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 sample of 100 farms, compute the sizes in each stratum under stratified simple random sampling with Neyman allocation. (7 marks)

QUESTION FIVE (15 MARKS)

(a) Explain the limitation of the Horvitz - Thompson technique in estimating the population mean. (2 marks)

(b) Show that for a simple random sample, the regression estimator, $\bar{x}_L$ is atleast as efficient as the ratio estimator, $\bar{x}_R$ of the population mean, $\bar{X}$ (6 marks)

(c) As the estimator of $\bar{Y}$, consider $\bar{\bar{y}} = \dfrac{1}{n}\sum_{i=1}^{n}\bar{y}_i$ as the cluster sample mean. Show that is an unbiased estimator for $\bar{Y}$. (7 marks)