(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2017/2018 ACADEMIC YEAR

FIRST YEAR FIRST SEMESTER

SPECIAL/SUPPLEMENTARY EXAMINATION

FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

COURSE CODE: **STA 807**

COURSE TITLE: SAMPLING THEORY AND PRACTICE I

DATE: 17/10/18 **TIME:** 8 AM -11 AM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 3 Hours

SECTION I: Answer ALL the questions in this section

QUESTION ONE

(a) For a population of 4 given by (2, 5, 3, 6), consider all possible simple random sampling with replacement of n = 2. Show that $\bar{y}$, S^2 are unbiased estimators of $\bar{Y}$, σ^2 respectively. [6 mks]

(b) Show that a standard error of $\bar{y}$ is given by: [7 mks]

$$\frac{S}{\sqrt{n}}\sqrt{1-f}\text{: where } f = \frac{n}{N}$$

(c) Signatures to a petition were collected on 676 sheets. Each sheet has enough space for 42 signatures, but on many sheets a smaller number of signatures was collected. The number of signatures per sheet were counted on a random sample of 50 sheets. The results are given below:

Numberofsignatures	42	41	36	32	29	27	23	19	16	15
Frequency	23	4	1	1	1	2	1	1	2	1
Numberofsignatures	14	11	10	9	7	6	5	4	3	Total
Frequency	1	1	1	1	1	3	2	1	1	50

Give an 80% confidence interval for the total number of signatures. [5 mks]

(d) For a bed of silver maple seedlings 1 ft wide and 430 ft long, it was found by complete enumerationthat $\mu = \bar{Y} = 19$and $S^2 = 85.6$, these being the true population values. The sampling unit was 1 ftof the length of the bed, so that N = 430. With simple random sampling, how many units must betaken to estimate μwithin 10% of accuracy, apart from a chance of 1 in 20? [2 mks]

(e) We have the 1930 number of inhabitants, in thousands, of N = 64 large cities in the United States, see Table below. The cities are arranged in two strata, the first containing the 16 largest cities, and the second containing the remaining 48 cities. The total number of inhabitants in all 64 cities is to be estimated from a sample of size n = 24. Calculate the standard errors of the estimated total for a stratified random sample with proportional allocation **[10 mks]**

*Strt*1	*Strt*2		
900	364	209	113
822	317	183	115
781	328	163	123
805	302	253	154
670	288	232	140
238	291	260	119
573	253	201	130
634	291	147	127
578	308	292	100
487	272	164	107
442	284	143	114
451	255	169	111
459	270	139	163
464	214	170	116
400	195	150	122
366	260	143	134

SECTION II: Answer anyTWO questions from this section

QUESTION TWO

(a) Let $\bar{y}$, $\bar{Y}$, s^2 and S^2 denote sample mean, Population mean, sample variance and population variance respectively given by:

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$,

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$, $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$

Show that:

(i) $\bar{y}$ is unbiased estimator of $\bar{Y}$ **[6 mks]**

(ii) s^2 is unbiased estimator of S^2 **[9 mks]**

(b) From a list of 3042 names and address,a simple random sample of 200 shown on investigation tha 38 address were wrong. Estimate the total number of addresses that need correction in the list and find a standard error of the estimate . **[5 mks]**

QUESTION THREE

Given is a hypothetical population consisting of $N = 7$ units with values 3, 6, 6, 8, 15, 18, and 28. Population of N = 7 units divided into two strata of $N_1 = 4$ and $N_2 = 3$ units as shown below"

Stratum	*Values*
1	3, 6, 6, 15
2	8, 18, 28

(a) List all possible sample for the stratified sampling procedure.

(b) Show that $E(\bar{y}_h) = \bar{Y}_h$ and $\hat{Y}_h = Y_h$

(c) Show that a stratified random sampling with $n_1 = 3$ and $n_2 = 2$, is better than a simple random sampling with $n = 5$

QUESTION FOUR

(a) Let $\bar{y}_{st}$ be the estimate for the population in a stratified sampling. Show that in a negligible stratum sampling fraction, the $Var(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h}$

(b) If terms in $\frac{1}{N_h}$ are ingnored relative to unity, show that $V_{opt} \leq V_{prop} \leq V_{ran}$ where the optimum allocation is for fixed n, that is with $n \propto N_h S_h$

QUESTION FIVE

For a systematic sampling of a sample of size n and k cluster unites, the symbol y_{ij} denotes the j^{th} member of the i^{th}systematic sample so that $j = 1, 2, ..., n$ and $i = 1, 2, ..., k$. Let the population variance be defined by $S^2 = \sum_{j=1}^{n}\sum_{i=1}^{k}(y_{ij} - \bar{Y})^2$ and the variance among units lying within the same systematics sample is denoted by $S_{wsy}^2 = \frac{1}{k(n-1)}\sum_{j=1}^{n}\sum_{i=1}^{k}(y_{ij} - \bar{y}_{i.})^2$. Show that:

(a) $V(\bar{y}_{sy}) = \frac{N-1}{N}S^2 - \frac{k(n-1)}{N}S_{wsy}^2$

(b) The mean of a systematic is more precise than the mean of a simple random sample if and only if $S_{wsy}^2 > S^2$

(c) $V(\bar{y}_{sy}) = \frac{N-1}{nN}S^2[1 + (n-1)\rho_{wst}]$ where ρ_{wsy} is the inter-class correlation coefficient between pairs of units that are in the same systematic sample and is denoted by

$\rho_{wsy} = \frac{2}{(N-1)(n-1)S^2}\sum_{i=1}^{k}\sum_{j<j'=1}^{n}(y_{ij} - \bar{Y})(y_{ij'} - \bar{Y})$