KISII UNIVERSITY

UNIVERSITY EXAMINATIONS

SPECIAL EXAMINATION

FOURTH YEAR EXAMINATION FOR THE AWARD OF DEGREE IN DEGREE OF BACHELOR OF SCIENCE IN APPLIED STATISTICS

FIRST SEMESTER 2021/2022

(JULY, 2022)

STAT 446: SAMPLING THEORY II

STREAM: Y4 S1 **TIME: 2 HOURS**

DAY: WEDNESDAY, 10.30 AM – 1.30 PM **DATE: 20/07/2022**

INSTRUCTIONS:

1. Do not write anything on this question paper.
2. Answer Question ONE (Compulsory) and any other TWO Questions in section B.

QUESTION ONE (30 MARKS)

a. Differentiate between cluster sampling and two-stage sampling (4Marks)

b. A population of students whose standard deviation of their number of visits to the dispensary to seek medication is 1.2 is divided into 6 clusters of size 5 and their cluster mean responses $\bar{Y}_i$ per cluster are 8, 3, 1, 11, 4 and 7.

i. Verify that the mean of the cluster means $\bar{y}_{cl}$ for a sample of 5 clusters is an unbiased estimator of the population mean $\bar{Y}$. (8 Marks)

ii. Given further that the intra-class correlation between the elements within a cluster ρ is 0.6125, estimate the between cluster population variance $\left(S_b^2\right)$ and the variance of the means of the cluster means $(Var(\bar{y}_{cl}))$ (4 Marks)

iii. Assuming that the sample of 5 is large enough for the population of 6, estimate the relative efficiency of cluster sampling over simple random sampling without replacement. Hence state the condition to hold true for cluster sampling to be more efficient. (4 marks)

c. In a study of plant disease, the plants were grown in 160 small plots containing 9 plants each. A random sample of 40 plots was chosen and three random plants in each sampled plot were examined for the presence of disease. It was found that 22 pots had no diseased plant (out of three). 11 had one disease, 4 had two diseases and 3 had three diseases.
Estimate the proportion of diseased plants and its standard error. (10 Marks)

QUESTION TWO (20 MARKS)

a. In the implementation of cluster sampling, the whole population is divided into clusters (N) which are treated as sampling units from which a sample of clusters (n) is selected.

i. Show that its more precise while selecting the sample of n clusters to employ simple random sampling without replacement than simple random sampling with replacement. (10 marks)

ii. Show that the relative efficiency of cluster sampling over simple random sampling without replacement is:

$E = \frac{1}{NM-1}\left[\frac{N(M-1)}{M}\frac{\bar{S}_w^2}{S_b^2} + (N-1)\right]$

where :

$\bar{\mathrm{S}}_{\mathrm{w}}^2 = \text{mean sum of squares within a clister in the population}$

$\mathrm{S}_{\mathrm{b}}^2 = \text{mean sum of squares between cluster means in the population}$ (10 Marks)

QUESTION THREE (20 MARKS)

In two-stage sampling, n units and m subunits from each chosen unit are selected by simple random sampling. If $\bar{\bar{y}}$is the over − all sample mean per subunit, show that;

a. $\bar{\bar{y}}$ is an unbiased estimate of $\bar{\bar{Y}}$ (3 Marks)

b. $V(\bar{\bar{y}}) = \left(\frac{N-n}{N}\right)\frac{S_1^2}{n} + \left(\frac{M-m}{M}\right)\frac{S_2^2}{mn}$ (10 Marks)

c. The unbiased estimate of

$$V(\bar{\bar{y}}) = (1-f_1)\frac{S_1^2}{n} + (1-f_2)\frac{S_2^2}{mn}$$

is

$$v(\bar{\bar{y}}) = (1-f_1)\frac{s_1^2}{n} + f_1(1-f_2)\frac{s_2^2}{mn}$$

Where:

$f_1 = \frac{n}{N},\ f_2 = \frac{m}{M}, S_1^2 = \frac{\sum_{i=1}^{n}(\bar{y}_i-\bar{\bar{y}})^2}{n-1}$ and $S_2^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\bar{y}_i\right)^2}{n(m-1)}$ (7 Marks)

QUESTION FOUR (20 MARKS)

Under the regression method of estimation, the regression estimator of $\bar{Y}$ is given by:

$\hat{\bar{Y}}_{reg} = \bar{y} + \beta(\bar{X} - \bar{x})$. For pre assigned value of β say β_0

a) Show that $\hat{\bar{Y}}_{reg}$ is an unbiased estimator of $\bar{Y}$ (2 Marks)

b) Show that $Var\left(\hat{\bar{Y}}_{reg}\right) = \frac{f}{n}[S_Y^2 + \beta_0^2 S_X^2 - 2\beta_0 \rho S_X S_Y]$ (10 Marks)

Where;

$f = \frac{N-n}{N}$,

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

$$S_Y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$$

ρ: Correlation coefficient between X and Y

c) Compare $Var\left(\hat{\bar{Y}}_{reg}\right)$with $Var(\bar{y})$ and state the condition(s) under which the regression estimator is better than the sample mean under simple random sampling without replacement (8 Marks)

QUESTION FIVE (20MARKS)

a) An experienced farmer makes an eye estimate of the weight of peaches x_i on each tree in an orchard of N=200 trees. He finds a total weight of 11600kg. the peaches are picked and weighed on a simple random sample of 10 trees and the following results were obtained:

	Tree Number									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Actual wt($\boldsymbol{y_i}$)	**61**	**42**	**50**	**58**	**67**	**45**	**39**	**57**	**71**	**53**
Est. wt ($\boldsymbol{x_i}$)	**59**	**47**	**52**	**60**	**67**	**48**	**44**	**58**	**76**	**58**

As an estimate of the total actual weight Y, we take $\hat{Y} = N[\bar{X} + (\bar{y} - \bar{x})]$

Compute the estimate and find its standard error (7 Marks)

b) If a population is divided into N homogeneous clusters each having M elements with an intra-class correlation of ρ. Show that:

i. $S_b^2 = \frac{(MN-1)}{M^2(N-1)}[1 + (M-1)\rho]$ (9 Marks)

ii. for large N, $Var(\bar{y}_{cl}) \approx \frac{S^2}{nM}[1 + (M-1)\rho]$ (4 Marks)