

**University of Ruhuna**  
**Bachelor of Science Special Degree**  
**(Level II) Semester II Examination - December 2016**

Subject: Mathematics

Course unit: MSP4273 (Sampling Theory and Categorical Data Analysis)

Time: Two (02) Hours

Answer 04 Questions only by selecting 02 Questions from each section

**Section A**

1. a) Write down the principal steps involved in a sample survey.
- b) For a simple random sample, show that  $s^2$  is an unbiased estimate of  $S^2$ ; where  $s^2$  and  $S^2$  have the usual meanings.
- A dentist, wanted to investigate the state of the teeth of 200 children in a village. He selected a simple random sample of 20 children and counted the number decayed teeth for each child, with the following results.

Number of decayed teeth	0	1	2	3	4	5	6	7	8	9	10
Number of children	8	4	2	2	1	1	0	0	0	1	1

Estimate the total number of decayed teeth of the children and 90% confidence limit for it.

- c) In a plant nursery that produces young trees for sale it is interested of the number of healthy trees,  $Y$ . In the complete enumeration of the population, with the usual notation it was found that  $\bar{Y} = 19$ ,  $S^2 = 85.6$  and  $N = 430$ .  
How many units must be taken to estimate  $\bar{Y}$  with in 10% apart from its actual value, with the chance of 19 in 20, with a simple random sample?
2. Explain the use of stratified random sampling technique, clearly indicating its advantages and disadvantages.

In the usual notation  $\bar{y}_{st} = \frac{\sum_{i=1}^h \bar{y}_i N_i}{N}$ .

Show that  $\bar{y}_{st}$  is an unbiased estimate of the population mean.

Find  $V(\bar{y}_{st})$ .

For a stratified random sample taken from  $h$  strata, where  $i^{th}$  strata is of size  $N_i$  with stratum variance  $S_i^2$ , ( $i = 1, 2, \dots, h$ ), obtain the sample size  $n_i$  by minimizing  $V(\bar{y}_{st})$  for a fixed total sample size  $n$ .

The households in a town are to be sampled in order to estimate the average amount of asset per household. The households are stratified into two strata as high-rent and low-rent. A house in the high-rent stratum is thought to have about nine times as much asset as one in the low-rent stratum, and  $S_i$ ,  $i=1,2$  is expected to be proportional to the square root of the stratum mean.

There are 4000 households in the high-rent stratum and 20000 in the low-rent stratum. Explain how would you distribute 1000 households between two strata, by minimizing  $V(\bar{y}_{st})$  for a fixed total sample size  $n = 1000$ .

3. a) Explain briefly how you would draw
- (i) a linear systematic sample
  - (ii) a circular systematic sample

Consider a linear systematic sample.

Show that in the usual notation,  $\bar{y}_r = \frac{\sum_{j=1}^n y_{r+(j-1)k}}{n}$ , at any random start  $r$  is an unbiased estimate of the population mean.

Show also that  $V(\bar{y}_r) = \frac{(N-1)}{N} S^2 - \frac{k(n-1)}{N} S_{wsy}^2$ ;

where  $S_{wsy}^2 = \frac{1}{k(n-1)} \sum_{r=1}^k \sum_{j=1}^n (y_{rj} - \bar{y}_r)^2$  with the usual notation.

The number of books issued to members of a particular library on 16 working days is recorded as follows:

22	53	32	55	46	43	41	48	57	55	49	43	44	54	58	48
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Find an unbiased estimate for mean number books issued, drawing a linear systematic sample of size 4. With the usual notation, find  $S^2$  and,  $S_{wsy}^2$  and hence find the variance of your estimate.

- b) Explain the use of cluster sampling technique.

A simple random sample of  $n$  clusters, each containing  $M$  elements is drawn from a population of  $N$  clusters.

In the usual notation show that  $V(\bar{y}) = (\frac{1-f}{nM^2}) S^2 (1 + (M-1)\rho) (\frac{NM-1}{N-1})$  ; where  $\rho$  = intra-cluster correlation coefficient.

For large population of clusters deduce that  $V(\bar{y}) = (\frac{1-f}{nM}) S^2 (1 + (M-1)\rho)$ .

A simple random sample of 50 clusters each containing 25 elements is drawn from a large population of clusters. Find  $V(\bar{y})$  if  $S^2 = 2$  and  $\rho = 0.2$ .

## Section B

1.

(a) Let  $\pi$  be a parameter for a distribution of dichotomous categorical data and  $p$  be its maximum likelihood estimate. Write the following large sample test statistics and their asymptotic distributions under  $H_0 : \pi = \pi_0$  with the usual notation, using

- (i) the Wald method and
- (ii) the score method

(b) In a recent poll, 120 registered voters from a random sample of 200 indicated that they plan to vote for Candidate X. Based on these results, a researcher would like to determine whether it is fair to say that Candidate X will receive the majority of votes.

- (i) What are the null and alternative hypotheses?
- (ii) Use the score test to compute the p-value of the hypothesis test.
- (iii) Use Wald test to compute the p-value of the hypothesis test.
- (iv) State the conclusions

2.

(a) Explain the following sampling models in the categorical data analysis with the usual notation:

- (i) Poisson sampling
- (ii) Binomial sampling
- (iii) Multinomial sampling.

(b) Explain the following studies.

- (i) Observational study
- (ii) Experimental study.

(c) Explain the following studies in the categorical data analysis and their sampling models with the usual notation:

- (i) Prospective study
- (ii) Retrospective study
- (iii) Cross-sectional study.

3.

- (a) There are three categorical variables, X, Y and Z. The data are stratified into  $k$  subgroups according to confounding variable (Z) of a 2 x 2 contingency table as shown in the following table for the  $i$ th stratum.

		Y		
Z=i	X	Y=1	Y=2	
	X=1	$a_i$	$b_i$	$a_i + b_i$
	X=2	$c_i$	$d_i$	$c_i + d_i$
	Total	$a_i + c_i$	$b_i + d_i$	$n_i$

where  $i = 1, 2, \dots, k$ .

- (i) Write the null hypothesis for the Mantel-Haenszel test.  
 (ii) Write the Mantel-Haenszel test statistic.  
 (iii) What is the distribution of Mantel-Haenszel test statistic, under the null hypothesis?
- (b) Answer the following questions using the data in the following table, obtained from General Social Survey.

		Type of employment	
Gender	Type of work	Work for someone else	Self-employed
Males	Work full-time	1051	199
	Work part-time	108	38
Females	Work full-time	978	87
	Work part-time	238	60

- (i) What is the marginal association between type of work and type of employment? Interpret the value of the appropriate marginal odds ratio.  
 (ii) What are the conditional associations (odds ratios) between type of work and type of employment (by Gender)?  
 (iii) Is there evidence for homogeneous association between type of work and type of employment (conditioned on Gender)? Write the appropriate test statistic to support your answer.