

# STA304/1003 H1F - Summer 2014: Surveys, Sampling, and Observational Data

## Supplementary Exercises # 2

1. Let  $\mathcal{S}$  be a SRS of size  $n$  from a population of size  $N$  and let

$$Z_i = \begin{cases} 1, & i \in \mathcal{S} \\ 0, & i \notin \mathcal{S} \end{cases}$$

- (a) How many samples contain  $i$  ?
  - (b) Find  $P(Z_i = 1)$ .
  - (c) How many samples contain both  $i$  and  $j$  (where  $i \neq j$ )?
  - (d) Find the probability that both  $i$  and  $j$  are in  $\mathcal{S}$ .
  - (e) Derive  $E(Z_i)$  ,  $V(Z_i)$  , and  $Cov(Z_i, Z_j)$  for  $i \neq j$  .
  - (f) Are  $Z_i$  and  $Z_j$  independent? Explain. Find the distribution of  $Z_i$ .
  - (g) Show that  $\bar{y}$  is unbiased for  $\bar{y}_U$  and  $\hat{t}$  is unbiased for  $t$  .
  - (h) Derive  $V(\bar{y})$  and  $V(\hat{t})$ .
  - (i) Show that  $s^2$  is unbiased for  $S^2$ .
2. Consider the following example in which one wishes to take a sample of size  $n = 2$  from population with  $N = 8$ :

$i$	1	2	3	4	5	6	7	8
$y_i$	1	2	8	2	8	1	5	1

- (a) Find the sampling distribution of  $\hat{t}$ .
- (b) Find the sampling distribution of the  $\bar{y}$ .

(c) Use the sampling distributions to verify:

- (i) the unbiasedness of  $\hat{t}$  and  $\bar{y}$
- (ii) formulae for the variances of  $\hat{t}$  and  $\bar{y}$  as derived in 1.(h) above

3. Consider the population of undergraduate students who are currently registered in the Statistics program at University of Toronto. Let

$$y_i = \begin{cases} 1, & \text{if student } i \text{ wishes to pursue graduate studies in Statistics} \\ 0, & \text{otherwise} \end{cases}$$

(a) Give an expression for the true proportion of students in this population who wish to pursue graduate studies in Statistics.

(b) Prove that for 0-1 data such as this, the following is true:

$$S^2 = \frac{N}{N-1} \bar{y}_U (1 - \bar{y}_U)$$

(c) Does a similar formula hold for  $s^2$ ?

(d) Suppose the population has 4000 students and we take a SRS of 200. 125 students say they wish to pursue graduate studies in Statistics.

- (i) Find a 95% CI for the true proportion of students in this population who wish to pursue graduate studies in Statistics.
- (ii) Find a 95% CI for the population proportion of students who do not wish to pursue graduate studies in Statistics.
- (iii) Find a 95% CI for the population percentage of students who wish to pursue graduate studies in Statistics.
- (iv) Find a 95% CI for the population total number of students who wish to pursue graduate studies in Statistics.

4. Give an example of a scenario for sampling in which each unit has equal probability of selection but it is not a SRS.