

STAT 410

NAME:

STUDENT NO.:

You may use a calculator. Read all questions carefully. **SHOW YOUR WORK.**

NOTE: Throughout this exam, you may assume without proof that if \bar{y} is the mean of a srs of size n from a population of size N that $E(\bar{y}) = \bar{y}_u$, the population mean, that $V(\bar{y}) = (1 - f)S_y^2/n$, where $f = n/N$ and $S_y^2 = \sum_{i=1}^N (y_i - \bar{y}_u)^2 / (N - 1)$, and $E(s_y^2) = S_y^2$, where $s_y^2 = \sum_{i \in s} (y_i - \bar{y})^2 / (n - 1)$.

1. Please answer the following:

(a) [4] For each of the following surveys, describe the target population, sampling method, sampling frame, and the sampled population.

- i. The Math Faculty at UW wishes to know the percentage of its graduates who find a job within three months from their graduation for the last five years (1999-2003). The Faculty has separate mailing lists for each year's graduates, compiled during students' last year at UW. 50 students were randomly chosen from each year's list and data were collected through mailed questionnaire.
- ii. *The Globe Poll* is a public opinion survey conducted over Canadians 18 or older in the 10 provinces through telephone interviews. In a recent survey, each selected respondent was asked to choose one of the three answers, A (agree), D (disagree), and U (undecided), to the survey question "...*Unemployment is so bad that the federal government must increase government spending on programs that create more jobs*". The goal is to find the population distribution over the three possible opinions. Sample data were taken using separate lists of telephone numbers for each of the provinces.

(b) [2] What is *bootstrap*? When is it mainly used for?

(c) [2] Briefly describe the procedure of two-phase sampling.

(d) A sample of n units is to be drawn without replacement from a population of N units by some method. Let π_i = probability that the i th unit is drawn, π_{ij} = probability that that i th and j th units are both drawn. Show that

- i. [2] $\sum_{i=1}^N \pi_i = n$.
- ii. [2] $\sum_{i=1}^N \sum_{j \neq i}^N \pi_{ij} = (n - 1)\pi_i$.

2. Suppose a finite population has four strata with sizes $N_1 = 3374$, $N_2 = 1960$, $N_3 = 52$ and $N_4 = 5$. A stratified simple random sample is drawn from the population and the sample data are summarized in the following table:

Stratum h	1	2	3	4
n_h	100	50	45	5
\bar{y}_h	491	1,050	2,568	18,054
s_h^2	921	1,278	2,343	154,054

- (a) [3] Compute the stratified sample mean \bar{y}_{st} and its estimated variance.
- (b) [2] Construct an approximate 95% confidence interval for the population total t .
- (c) Suppose the unknown stratum quantities \bar{y}_{hU} and S_h^2 can be roughly approximated, using information from other sources, as follows:

Stratum h	1	2	3	4
N_h	100	50	45	5
\bar{y}_{hU}	488	1,025	2,636	18,054
S_h^2	880	1,434	2,665	154,054

Assume a stratified random sample of size $n = 300$ is being used for the estimation of \bar{y}_u .

- [2] Calculate the stratum sample sizes under proportional allocation.
- [2] Calculate the stratum sample sizes under optimal (Neyman) allocation.
- [3] What are the values of $V(\bar{y}_{st})$ with sample size allocations of i) and ii), respectively? compare these two variances and comment.

3. A home owner with a large library needs to estimate the purchase cost and replacement value of the book collection for insurance purposes. She has 44 shelves containing books and selects 12 shelves at random. To prepare the second stage of sampling, she counts the books on the selected shelves. She then generates five random numbers between 1 and M_i for each selected shelf to determine which specific books to examine more closely. She then looks up the replacement value for the sampled books in *Books in Print*. Summaries of the data are give in the table below.

Summary of sampled data				
Shelf	M_i	m_i	\bar{y}_i	s_i^2
2	26	5	9.6	17.80
4	52	5	6.2	1.70
11	70	5	9.2	18.70
14	47	5	7.2	3.20
20	5	5	41.8	2666.70
22	28	5	29.8	748.70
23	27	5	51.8	702.70
31	29	5	61.2	353.20
37	21	5	50.4	147.30
38	31	5	36.6	600.80
40	14	5	54.2	595.70
43	27	5	6.6	4.80
Sum	377			

- (a) [3] Estimate the total replacement cost for the library and find the standard error of your estimate. What is the estimated coefficient of variation?
- (b) [3] Estimate the average replacement cost per book, along with the standard error. What is the estimated coefficient of variation?
- (c) [2] Do you feel this was an efficient method of sampling? Discuss.
4. Suppose we have a population of N units. A simple random sample without replacement of n units is drawn and two characteristics, x_i and y_i are measured on the i -th unit for $i = 1, \dots, n$. Suppose we are interested in estimating the population characteristic

$$\theta = \bar{x}_u^2 \cdot \bar{y}_u.$$

Let our sample estimate be

$$\hat{\theta} = \bar{x}^2 \cdot \bar{y}.$$

- (a) [4] Obtain the leading term of $Bias(\hat{\theta})$.
- (b) [4] Derive the approximate variance of $\hat{\theta}$ and use it to give a variance estimator.

