Last name, first name: _____ . Student #: _____

STA 304H1 F SUMMER 2011, First Test, June 2 (20%)
Duration: 60 min. Allowed: nonprogrammable hand-calculator, aid-sheet, one side, with theoretical formulas only; the test contains 3 pages, + an empty page; please check

[25] 1) The number of words in a large book is to be determined by selecting a sample of pages and counting the number of words on those pages.
(a) What is the target population? Explain it in some details. What is the sampling population?
(b) What is the frame in this sampling? Does the frame properly cover the population? Explain.
(c) What is the variable that is measured? Discus possible problems that might appear with the definition of that variable.
(d) What is the parameter which should be estimated?
(e) From a brief investigation it was found that there are about 400-500 words per page. What is an approximate sample size required to estimate the average number of words per page with a bound on the error of estimation of 10 words, assuming there is a large number of pages in the book?

Solutions:
(a) (5) Target population: All pages in the book that contain some text, except cover pages and title page, maybe contents, and some other pages that contain publishing information, etc. **[3]**
Sampling population: Same as target population. **[2]**

(b) (4) Frame: List of page numbers that are included in the book. **[2]**
Frame may not properly cover the population, if the pages that are excluded from the population are not excluded (or notified) from the frame. Or, some pages, such as Preface or Announcement, may be included in the population, but not numbered, and then not sampled. **[2]**

(c) (4) Variable **y**: Number of words on the page. **[2]** Here also may be some problem. Do we count figures, and how, graphs, tables, captions, do we count formulas, etc? **[2]**

(d) (4) Parameter which should be estimated: Total number of words in the book, i.e. the population total for variable **y**. **[4]**

(e) (8) Assuming large N (# of pages in the book), $Var(\bar{y}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} \approx \frac{\sigma^2}{n}$. **[2]** Use $\hat{\sigma}^2 = (Range/4)^2 = ((500\text{-}400)/4)^2 = 25^2$. **[1]** Then from $B_\mu = 10 = 2\sqrt{Var(\bar{y})}$, **[1]**

$Var(\bar{y}) = \frac{\sigma^2}{n} = (10/2)^2 = 25$, or $n = \frac{\sigma^2}{25} = \frac{25^2}{25} = 25$. Required sample size n = 25. **[4]**

[40] 2) A survey is conducted in a cottage country place in Muskoka region with a goal to investigate the age structure, occupation, and some other characteristics of the residents. The place has 1000 households. A preliminary, small SRS (without replacements) of size n = 12 households was selected from the place directory. Each household is either in the north, or south part of the place. At the completion of visiting the sampled households, the following results were obtained:

Household in the sample	1	2	3	4	5	6	7	8	9	10	11	12	total
Area (North, South)	N	N	N	N	S	S	N	S	S	N	N	S	
Size (residents)	6	4	5	7	3	2	8	4	4	5	4	6	58
Adults over 65	2	0	0	2	0	2	1	0	2	2	1	0	12

(a) Estimate the total number of residents in the place, and the total number of adults over 65 in the place.

(b) Use an unbiased estimator to estimate the variance of the size of the household.

(c) Place a bound on the error of estimation of the total number of residents.

(continued)

Solutions:

(a) $\hat{\tau}_y = N\overline{y}$ = 1000x58/12 = 1000x4.833 = 4833.3, **[3]**

$\hat{\tau}_x = N\overline{x}$ = 1000x12/12 = 1000x1.0 = 1000. **[3]**

(b) $\hat{\sigma}^2 = ((N-1)/N)S^2$ = 999/1000x2.879 = 2.876. ((N-1)/N factor should be used) **[6]**

(c) $B_\tau = 1000 \times 2 \times \sqrt{\frac{1000-12}{1000}\frac{2.879}{12}}$ = 973.7. **[7]**

(d) Estimate the total number of households in the North and in the South.
(e) Estimate the percentage of households with at least one adult over 65 in the place.
(f) Propose and calculate an estimate of the percentage of adults over 65 in the place.

Solutions:

(d) $\hat{\tau}_N = 1000 \times 7/12 = 583.3$, $\hat{\tau}_S = 1000 \times 5/12 = 416.7$. **[7]**

(e) $\hat{p} = 7/12 = 0.583 = 58.3\%$ (% of households with at least one adult over 65). **[7]**

(f) $\hat{p} = 1000/4833.3 = 0.2069 = 20.69\%$ (number of adults over 65/number of residents) **[7]**

[35] 3) Use the following table of random numbers
31624 76384 17403 53363 44167 64486 64758 75366 76554 31601 12614 33072 60332
92325 19474 23632 27889 47914 02584 37680 20801 72152 39339 34806 08930 25570
to choose a SRS (without replacement) of 6 even pages from a book with 650 pages.
(a) Which pages will be selected in the sample? Explain your procedure briefly.
(b) Assume that four pages in the sample were found with figures, containing 9 figures in total. Numbers of figures per page in the sample were 0, 2, 1, 3, 0, 3.
Estimate the total number of pages in the book with at least one figure and place a bound on the error of estimation. (hint: first estimate the proportion of pages with at least one figure)
(c) Estimate the total number of figures in the book. Place a bound on the error of estimation.
(d) Discuss the sampling method used in this question. Does it produce an SRS of pages? Explain. What should you assume to justify using the obtained estimates?

Solutions:
(a) (8) Population size N = 650; we have to use 3 digits from the table of random numbers; we have to ignore odd numbers and the numbers greater than 650, including 000. Using, e.g., consecutive groups of three digits, we get:
316, 247, 638, 417, 403, 533, 634, 416, 764, 486, 647, 587, 536, ...
so that the pages 316, 416, 468, 536, 634, 638 will be in the sample. **[8]** //student may use some other method, but with explanation//

(b) (12) $\hat{p} = \frac{4}{6}$, $\hat{\tau} = N\hat{p} = 650 \times 4/6 = 433.3$ **[6]**

$$\hat{Var}(\hat{p}) = \frac{\hat{p}\hat{q}}{n-1}\frac{N-n}{N} = \frac{4/6 \times 2/6}{6-1}\frac{650-6}{650} = 0.0440342$$

$B_\tau = 2N\sqrt{\hat{Var}(\hat{p})} = 2 \times 650 \times \sqrt{0.0440342}$ =272.8. **[6]**
//if students use N'=650/2=325 in this part, it can be accepted.

(c) (8) For the average number of figures per page, 1.5, for the total,
$\hat{\tau} = N\hat{\mu} = 650 \times 1.5 = 975$. **[5]**

$S^2 = \frac{1}{6-1}\sum (y_i - 1.5)^2 = 1.9$. $\hat{Var}(\bar{y}) = \frac{650-6}{650} \times \frac{1.9}{6}$ =0.3137,

$B_\tau = 2 \times 750 \times \sqrt{0.314} =$ $B_\tau = 2 \times 650 \times \sqrt{0.3137} = 728.46$ **[3]**

(d) (7) The sample of pages is not an SRS of pages from the book, because the odd pages are not included in sampling. **[4]** We can assume that figures are randomly placed over the book, so that sampling only from even pages is as representative as sampling from the whole book. **[3]**