Last name, first name: ______________________ . Student #: ______________

STA 322H1 S WINTER 2008, Second Test, March 25 (20%)
Duration: 50min. Allowed: hand-calculator, aid-sheet, one side, with theoretical formulas and definitions only (check whether you have all 4 pages of the test)

[60] 1) A city hall economical study conducted in a city is interested in the age structure of the employees in small to medium stores (population). The file shows that 650 stores of this type are registered. An SRS of n = 12 stores was selected from the file. At the completion of the fieldwork the following results were obtained:

store	1	2	3	4	5	6	7	8	9	10	11	12	
# employed (size)	6	4	5	7	3	2	8	4	4	5	4	6	58
# employed, under 18 (juniors)	3	2	3	3	1	0	4	2	0	2	1	3	24
# employed, 18-65 (adults)	1	2	2	2	2	0	3	2	2	1	2	3	22
# employed, over 65 (seniors)	2	0	0	2	0	2	1	0	2	2	1	0	12

(a) [21] Estimate the following parameters:
   (1) The total number of employees in the population, and a bound on the error of the estimation,
   (2) The total number of employed nonjuniors in the population,
   (3) The total number of stores that employ at least one junior.

(b) [10] How large a sample should be taken in order to estimate the proportion of stores that employ at least one senior, with a bound of 0.1 on the error of estimation?

Solutions:

(a) [21] (1) y = size, τ_y = total # employed, $\bar{y} = 4.833,\ S^2 = 2.879$,

$\hat{\tau}_y = N\bar{y} = 650 \times 4.833 = 3141.7$, **[5]**

$B_\tau = 2N\sqrt{\hat{V}ar(\bar{y})} = 2N\sqrt{\frac{N-n}{N}\frac{S^2}{n}} = 1300\sqrt{\frac{650-12}{650}\frac{2.879}{12}} = 630.83$. **[4]**

(2) y = nonjuniors, τ_y = total # employed nonjuniors, $\bar{y} = 2.833$

$\hat{\tau}_y = N\bar{y} = 650 \times 2.833 = 1841.67$. **[6]**

(3) p = proportion of stores that employ at least one junior, τ_p = total # of stores that employ at least one junior, $\hat{\tau}_p = N\hat{p} = 650 \times \frac{10}{12} = 541.67$. **[6]**

(b) [10] $\hat{p} = \frac{7}{12} = 0.583,\ \hat{q} = \frac{5}{12} = 0.417$, **[3]**

$$n = \frac{Npq}{(N-1)D + pq} = \frac{650 \times 0.583 \times 0.417}{649 \times (0.1/2)^2 + 0.583 \times 0.417} = 86 \text{ . } \mathbf{[7]}$$

(c) [17] Estimate the proportion of employees who are seniors, and place a bound on the error of estimation. Is this estimator unbiased? Explain.

(d) [8] Assuming that the total # of employed is 3100, use the ratio estimator to estimate the total number of employed adults in the population.

(e) [4] Can you estimate the average age of employed using given sample? Explain.

Solutions:

(c) [17] $R = \dfrac{\text{total \# number of seniors}}{\text{total \# of employees}} = \dfrac{\tau_y}{\tau_x}$, $\hat{R} = r = \dfrac{12}{58} = 0.2069 = 20.68\%$, **[6]**

$\hat{V}ar(r) = \dfrac{N-n}{N\overline{x}^2}\dfrac{S_r^2}{n} = \dfrac{650-12}{650(4.83)^2}\dfrac{0.9571}{12} = 3.356\times10^{-3}$, $B_r = 2\sqrt{\hat{V}ar(r)} = 0.1159$. **[8]**

($\hat{\mu}_x = \overline{x} = 4.83$, $S_r^2 = \frac{1}{n-1}\Sigma(y_i - rx_i)^2 = 0.9571$)

This is a ratio estimator, and then is biased. **[3]**

(d) [8] $\hat{\tau}_y = \dfrac{\Sigma y_i}{\Sigma x_i}\tau_x = \dfrac{22}{58}3100 = 1175.86$. **[8]**

(e) [4] We cannot estimate the average age of employees, because the data includes only the counts for each age group, but not the actual ages. **[4]**

[40] 2) The city under the study in the previous question is actually divided into two city areas, the smaller south mostly with older buildings, including 250 registered stores, and the larger north mostly with new buildings and 400 registered stores.

(a) [8] Discuss some advantages or disadvantages, if any, of using the stratification by the area in the study of the age structure of employees.

(b) [16] Assume that the sample in the previous question was actually obtained as a stratified sample, with the stores # 5, 6, 8, 9, 12 from the south. Rearrange your sample to be more convenient for calculation. Estimate now

(1) The total number of employees in the population,

(2) The total number of stores that employ at least one senior.

Solutions:

(a) [8] We could expect that south and north are more homogeneous, with older population living in the south, and younger living in the north. This will likely reflect in employees in the smaller stores, as they mainly come from the neighborhood. This situation is convenient for stratified sampling. Cost of sampling could be also reduced. **[8]**

(c) [16]

	south					north						
store	5	6	8	9	12	1	2	3	4	7	10	11
y_1, size	3	2	4	4	6	6	4	5	7	8	5	4
y_2, seniors	0	2	0	2	0	2	0	0	2	1	2	1
$\bar{y}_1$	3.8					5.57						
$\bar{y}_2$	0.8					1.143						
N_i	250					400						

[2]

(1) $\hat{\tau}_{y_1} = N_1\bar{y}_1 + N_2\bar{y}_2 = 250\times 3.8 + 400\times 5.57 = 3178$, **[7]**

(2) $\hat{\tau}_{y_2} = N_1\hat{p}_1 + N_2\hat{p}_2 = 250\times 2/5 + 400\times 5/7 = 385.7$. **[7]**

(c) [16] What proportion of a sample should be taken from the south and from the north in order to estimate the total number of stores that employ at least one senior? Use the sample indicated in (b) as a presample. Assume that the cost of sampling from the north is twice the cost of sampling from the south and you want to use the optimal allocation.

Solution:

(c) [16] There are two strata with sampling costs $c_1 = c$ and $c_2 = 2c$ **[2]**. Then

$$\omega_1 = \frac{N_1\sqrt{p_1q_1/c}}{N_1\sqrt{p_1q_1/c} + N_2\sqrt{p_2q_2/(2c)}} = \frac{N_1\sqrt{p_1q_1}}{N_1\sqrt{p_1q_1} + N_2\sqrt{p_2q_2/2}}$$

$$= \frac{250\sqrt{(2/5)(3/5)}}{250\sqrt{(2/5)(3/5)} + 400\sqrt{(2/7)(5/7)/2}} = \frac{122.47}{122.47 + 127.76} = 0.489, \textbf{ [10]}$$

$\omega_2 = 1 - \omega_1 = 0.511$, **[4]**

i.e., of the total sample size 48.9% should be selected from the south, and 51.1% from the north.