Last name, first name: ______________________________. Student #: ________________

STA 304H1 F FALL 2010, Second Term-test, November 19 (20%)
Duration: 1h. Allowed: hand-calculator, aid-sheet, one side, with theoretical formulas only

[38] 1) A city is divided into 600 blocks. A preliminary SRS of size 10 was selected from the city blocks and the following results were obtained (x - number of cars parked on the street, y - number of illegally parked cars on the street, z – total number of cars in the block, no matter where parked (could be in a garage)):

var	1	2	3	4	5	6	7	8	9	10	Σ var	Σ var^2
x	8	9	4	2	0	5	3	0	4	3	38	224
y	2	3	2	0	0	2	0	0	1	1	11	23
z	12	15	10	8	6	20	5	12	8	9	105	1283

Σxy = 68

(a) [12] Estimate (i) the total number of cars in the city, (ii) the total number of cars parked on the street in the city, (iii) the total number of illegally parked cars on the street in the city.

(b) [16] Estimate (i) the proportion of cars parked on the street (out of all cars), (ii) the proportion of cars illegally parked on the street (out of all cars parked on the street) and the standard deviation of that estimator.

(c) [10] Later you found that the total number of cars in the city is 6200.Use this information to again estimate the total number of cars parked on the street in the city. Do you expect this estimator be better than one in (a) (ii)? Explain. How would you check it? Just explain, don't do any calculation.

Solutions:

From the table Σxy = 68, $\bar{x} = 3.8$, $\bar{y} = 1.1$, $\bar{z} = 10.5$.

(a) (i) $\hat{\tau}_z = N\bar{z}$ = 600 x 10.5 = 6300, **[4]** (ii) $\hat{\tau}_x = N\bar{x}$ = 600 x 3.8 = 2280, **[4]**

(iii) $\hat{\tau}_y = N\bar{y}$ = 600 x 1.1 = 660. **[4]**

(b) (i) $\hat{R}_{x/z}$ = 38/105 = 0.362 = 36.2%, **[4]**

(ii) $\hat{R}_{y/x}$ = 11/38 = 0.289 = 28.9%, **[4]**

$S_r^2 = \sum (y_i - rx_i)^2/(n-1)$ = (23 – 2 x 0.289 x 68 + 0.289^2 x 224)/9 = 0.267, **[4]**

$\hat{Var}(\hat{R}) = \frac{N-n}{N} S_r^2/(n\bar{x}^2)$ = (600-10)/600 x 0.267/(10 x 3.8^2) = 1.82 x 10^{-3},

$\hat{S}d(\hat{R}) = \sqrt{1.82 \times 10^{-3}}$ = 0.043. **[4]**

(c) You may use the ratio estimator $\hat{\tau}_x = \hat{R}_{x/z}\tau_z$ = (38/105) x 6200 = 2243.8 = 2244. **[5]**

Yes, we expect this ratio estimator be better than one in (a)(ii), due to correlation between the number of cars and the block size. **[3]** To check it, we just may calculate their variances. **[2]**

[40] 2) A service company maintains three condominium buildings, known as North, South and West wing, having 250, 150, and 200 condominiums respectively. To investigate the quality of building maintenance, a stratified random sample of 120 condominiums (using proportional allocation) was selected from the buildings and the condominium owners were questioned about quality of maintenance. The following results were obtained:

Wing	Size	Sample size	Owners s	ns	no
North	250	50	27	14	9
South	150	30	18	8	4
West	200	40	22	12	6
total	600	120	67	34	19

s = satisfied, ns = not satisfied, no = no opinion

(a) [10] Estimate the following parameters related to the owners: (i) the total number of owners with an opinion, (ii) the proportion of owners who are satisfied with building maintenance.
(b) [10] Estimate the standard deviation of the estimator in (a-ii).
(c) [20] What would be the optimal allocation of another sample of size 120 if the parameter in (a-ii) should be estimated again? Use the obtained sample as a presample.

Solutions:

(a) (i) $\hat{p}_{a1} = \sum \frac{N_i}{N}\hat{p}_i$ = [250 x (50-9)/50 + 150 x (30-4)/30 + 200 x (40-6)/40]/600 = 505/600,

$\hat{\tau}_{a1} = N\ \hat{p}_{a1}$ = 600 x 505/600 =505 **[5]**

(ii) $\hat{p}_{a2} = \sum \frac{N_i}{N}\hat{p}_i$ = (250 x 27/50 + 150 x 18/30 + 200 x 22/40)/600 = 0.5583. **[5]**

(b) $\hat{V}ar(\hat{p}_{a2}) = \sum W_i^2 \frac{N_i - n_i}{N_i}\hat{p}_i\hat{q}_i/(n_i - 1) =$

$(\frac{250}{600})^2 \frac{250-50}{250}\frac{27}{50}\frac{23}{50}\frac{1}{49} + (\frac{150}{600})^2 \frac{150-30}{150}\frac{18}{30}\frac{12}{30}\frac{1}{29} + (\frac{200}{600})^2 \frac{200-40}{200}\frac{22}{40}\frac{18}{40}\frac{1}{39}$ = 1.682 x 10^{-3}. **[10]**

(c) Use estimates $\hat{p}_i, \hat{q}_i$ to calculate $n_i = n\frac{N_i\sqrt{p_i q_i}}{\sum N_i\sqrt{p_i q_i}}$, n = 120. **[3]**

$\sum N_i\sqrt{p_i q_i} = 250\sqrt{\frac{27}{50}\frac{23}{50}} + 150\sqrt{\frac{18}{30}\frac{12}{30}} + 200\sqrt{\frac{22}{40}\frac{18}{40}}$ = 124.60 + 73.48 + 99.50 = 297.58, **[8]**

n_1 = 120 x 124.60/297.58 = 50, n_2 = 120 x 73.48/297.58 = 30,
n_3 = 120 x 99.50/297.58 = 40. **[9]**

[22] 3) An auditor is confronted with a long list of accounts receivable for a firm. She must verify the amounts on 5% of these accounts and estimate the average difference between the audited and book values. In the following cases explain your choice.

(a) [4] Suppose the accounts are arranged chronologically (in calendar time), with the older accounts tending to have smaller values. Would you choose a systematic or a simple random sampling design to select the sample?

(b) [4] Suppose the accounts are arranged alphabetically. Would you choose a systematic or a simple random sampling design to select the sample?

(c) [5] Suppose the accounts are grouped by department and then listed chronologically within departments. Would you choose a systematic or a simple random sampling design to select the sample?

(d) [5] How would you estimate the variance of the sample mean according to your choice of the design in (b)?

(e) [4] Explain in detail how you would select a systematic sample from the list of accounts. Write down a list of first 6 accounts that would appear in your sample.

Solutions:

(a) Order in the population is highly correlated with the values of the variable of interest, and a systematic sample should be better that an SRS, i.e., the values in a systematic sample will be more heterogeneous that in SRS. **[4]**

(b) Both types of samples would give similar results, because the random order of elements can be assumed. If a systematic sample is selected, it can be used as if it were an SRS. Selecting a systematic sample may be simpler. **[4]**

(c) If there are only a few departments, the population can be considered as ordered, and a systematic sample could be better than an SRS. If the number of departments is large, each with a small number of accounts, the population could be periodic, and systematic sampling could produce biased results, depending on size and step. An SRS is, then, preferable. **[5]**

(d) In both cases, a method from SRS sampling can be used to estimate the variance of the sample mean, using sample variance S^2, that is $\hat{V}ar(\hat{\mu}) = \frac{N-n}{N}\frac{S^2}{n} \approx \frac{S^2}{n}$ (because N is large). **[5]**

(e) 5% of the accounts means an 1- in – 20 systematic sample with random start, that is, the first element should be selected at random out of the first 20 elements, and then every 20^{th} . If, e.g., the first element is 14, then the other elements are 34, 54, 74, 94, 114. **[4]**