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

STA 304H1S SUMMER 2009, Second Test, June 11 (20%)
Duration: 50min. Allowed: hand-calculator, aid-sheet, one side, with theoretical formulas and definitions only

[50] 1) A survey was recently conducted using an SRS of 132 city blocks from a frame of 9460 city blocks in a city downtown (city, for short). Data on the number of households (x), the number of households parking their cars on their front loan (z), and the total number of parking places on households' loans (y) for the city blocks in the sample were obtained as follows:
$\sum x = 691$, $\sum y = 232$, $\sum z = 195$, $\sum x^2 = 4139$, $\sum y^2 = 816$, $\sum z^2 = 525$,
$\sum xy = 1435$, $\sum xz = 1195$.

[15] (a) Estimate the proportion of households in the city parking their cars on their front loan and the standard deviation of this estimate.
[10] (b) Estimate the total number of households and the total number of parking places on households' loans in the city. Are these estimators unbiased? Explain. **(continued)**

Solutions:

(a) N = 9460, n = 132, use ratio estimator

$\hat{p} = r = \frac{\Sigma z_i}{\Sigma x_i} = \frac{195}{691} = 0.2822$ **[6]**

$\hat{\mu}_x = \Sigma x_i / n = 691/132 = 5.235$, $S_r^2 = \Sigma(z_i - r x_i)^2/(n-1) = \Sigma(z_i^2 - 2r\Sigma x_i z_i + r^2 \Sigma x_i^2)/(n-1)$
$= (525 - 2\times 0.2822 \times 1195 + (0.2822)^2 \times 4139)/131 = 180.1589/131 = 1.375,$

$\hat{Var}(\hat{p}) = \frac{N-n}{N}\frac{1}{n}\frac{S_r^2}{\hat{\mu}_x^2} = \frac{9460-132}{9460} \times \frac{1.375}{132 \times (5.235)^2} = 3.74794 \times 10^{-4}$, $\hat{SD}(\hat{p}) = 0.01936$. **[9]**

(b) $\hat{\tau}_x = N\hat{\mu}_x = 9460 \times 5.235 = 49523.1$ **[4]**

$\hat{\tau}_y = N\hat{\mu}_y = 9460 \times 232/132 = 16626.7$ **[4]**

Both estimators are unbiased, because $\hat{\mu}_x = \bar{x}$, $\hat{\mu}_y = \bar{y}$ are unbiased estimators. **[2]**

[5] (c) It was found from the city office there were 56296 households in the city. Estimate again the total number of parking places on households' loans in the city, using this information and a ratio estimator. Is this estimator better than one from (b)? Discuss it without calculation.
[5] (d) Estimate again the total number of parking places on households' loans in the city using a difference estimator. Do you expect the difference estimator be better than one from (b)? Discuss it without calculation.
[10] (e) What would be an appropriate sample size if the proportion of the households parking their cars on their front loans has to be estimated with a bound on the error of estimation of 3%. Use the given sample as a presample study.
[5] (f) Comment on using the regression estimator in this problem, compared to the estimator used in (c).

Solutions:

(c) $\hat{\tau}_y = \hat{p}_y \tau_x = 232/691 \times 56296 = 0.3357 \times 56296 = 18898.6$ **[3]**
It is a ratio estimator. As the number of parking places is correlated with the number of households in city blocks, we expect that this ratio estimator is better than an SRS estimator (actual variances are close in this example) **[2]**
(d) Difference estimator: $\hat{\tau}_y = N\hat{\mu}_{y,D} = N(\overline{y} - (\overline{x} - \mu_x)) = 9460 \times (\frac{232}{132} - \frac{691}{132} + \frac{56296}{9460})$
= 9460 x 2.4736 = 23401. **[3]**
Difference estimator is more efficient than SRS if $\hat{\rho} > \frac{1}{2}\frac{S_x}{S_y}$. We could expect $S_x \approx S_y$, so that if $\hat{\rho} > \frac{1}{2}$, then difference estimator will be better than SRS (in this example $\hat{\rho} < \frac{1}{2}$) **[2]**
(e) $n = \dfrac{N\sigma_r^2}{N \times D + \sigma_r^2} = \dfrac{9460 \times 1.375}{9460 \times (0.03 \times 5.235/2)^2 + 1.375} = 218$ **[10]**.
$\hat{\sigma}_r^2 = S_r^2 = 1.375$.
(f) We may expect that the regression estimator would give similar result as the regression estimator used in (c), as the number of parking places is directly proportional to the number of household in the city block. The regression line should go through the origin. **[5]**

[50] 2) In order to estimate the total inventory of its products being held, a car company conducts a proportional stratified sample of its dealers, with these dealers being stratified according to their inventory held in previous year. For a total sample size of n = 100, the following data for the current inventory were obtained

Stratum (last year inventory)	Number of dealers	Sample results $\overline{y}_i$
I [0 – 100)	400	105
II [100-200)	1000	180
III [200 – 400)	540	370
IV [400 – 600)	60	590

[5] (a) Find the allocation of the sample used.
[12] (b) From this stratified sample, estimate the mean current inventory μ and the total current inventory.
[8] (c) Can you estimate the variance of the estimator $\hat{\mu}$ used in (b) using the sample? Is there any other information in the above table that might help to estimate that variance? Explain and use it to estimate the variance (you may ignore the finite population corrections). **(continued)**

Solutions:

[5] (a) Proportional allocation: $n_i = \frac{N_i}{N} n$, N = 400+1000+540+60 = 2000, or

n_1 = 100x400/2000 = 20,
n_2 = 100x1000/2000 = 50,
n_3 = 100x540/2000 = 27,
n_4 = 100x60/2000 = 3. **[5]**

[12] (b) $\hat{\mu} = \Sigma \frac{N_i}{N} \overline{y}_i$ = 0.2x105 + 0.5x180 + 0.27x370 + 0.03x590 = 228.6 **[7]**
$\hat{\tau} = N\hat{\mu}$ = 2000x228.6 = 457,200. **[5]**

[8] (c) Strata sample variances are missing from the table, so the variance of the sample mean cannot be estimated, using sample results only. **[2]**
Strata ranges from the last year inventory might be used to estimate the strata standard deviations for this year inventory, assuming they remain similar, even the mean values seem to be shifted up. Using ranges and the formula $\sigma_i = R_i / 4$ we obtain
$\sigma_1 = \sigma_2 = 25, \sigma_3 = \sigma_4 = 50$ **. [3]**
Then by ignoring the finite population corrections, we obtain (for proportional allocation)

$Var(\hat{\mu}) = \Sigma W_i^2 \sigma_i^2 / n_i = \Sigma W_i \sigma_i^2 / n = (0.2 \times 25^2 + 0.5 \times 25^2 + 0.27 \times 50^2 + 0.03 \times 50^2) / 100 =$ 11.875. **[3]**

[6] (d) Ignoring the sampling cost, do you expect that the stratified sample with the proportional allocation will produce significantly better results than an SRS? Explain.
[13] (e) If the cost of sampling one dealer is $20, and the total money that can be spent on sampling is $1500 (ignore presamping costs), calculate the allocation of the sample, which would minimize the error bound, using information obtained from the last year .
[6] (f) What was the main goal of this study? Do you think the stratification used in this study is advantageous for the goal of the study? Explain. Propose some other stratification that may be convenient in a different way.

Solutions:
[6] (d) It seems from the sample that the strata means are quite different (it is also obvious from the strata limits), which means that proportional allocation should produce an estimator with smaller variance than a SRS. **[6]**
[13] (e) The total sample size is n = 1500/20 = 75 **[2]**, due to equal costs of sampling from each stratum; we then can use the Neyman allocation, for given sample size. Relative allocation is then $\omega_i = N_i\sigma_i / \Sigma N_i\sigma_i$, and the allocation is $n_i = \omega_i n$, where we may use the strata ranges to estimate strata standard deviations (see (c)) Then
$\sigma_1 = \sigma_2 = 25, \sigma_3 = \sigma_4 = 50$ **[3].**
$\Sigma N_i\sigma_i$ = 400x25 + 1000x25 + 540x50 +60x50 = 65000, **[2]**
n_1 = 75x4x25/650 = 11.5=12,
n_2 = 75x10x25/650 = 28.8=29,
n_3 = 75x5.4x50/650 = 31.2=31,
n_4 = 75 – (12+29+31) = 3. **[6]**
[s.size values should be rounded reasonably so that the total is 75]

[6] (f) The main goal of the study was to estimate the mean and total inventory for this year. **[2]**
The stratification is very good, because the strata are homogeneous (stratified by the variable correlated with the variable of interest), that is, strata variances are small, and there are large differences between mean values in the strata. **[2]**
The other possible stratification would be by region, which might be convenient if the dealers should be visited, or if the goal of the study is also to estimate regional differences between dealers. **[2]**