STA 322H1 S W 2007 FE $\boxed{1}$

$\boxed{14}$ 1) (a)^3 $\hat{p} = \frac{65}{100} = 0.65$ |1, $B_p = 2\sqrt{\hat{Var}(\hat{p})} = 2\sqrt{\frac{\hat{p}\hat{q}}{n-1}} =$

$= 2\sqrt{\frac{0.65 \times 0.35}{99}} = 0.096$ |1, $\frac{N-n}{N} \approx 1$, N large

(b)^3 $0.55 \le p \le 0.65$ $n = \frac{pq}{D^2} \le \frac{0.55 \times 0.45}{\left(\frac{0.03}{2}\right)^2} = 1100$

$D = \frac{B_p}{2}$

|1

(c)^2 Target population: ~~All~~ All women using lipstick.

sampling population: Women who come to the booth in the shopping mall. |1

(d)^2 Sample selected is not an SRS from target population because some women ~~s~~ don't come to shopping malls, or don't visit marketing boots. |1 If visiting shopping malls is not related to preference for lipsticks, then the sample may be considered as SRS. (i.e., representative of the population) |1 (e)[2] placing booths in all, or randomly selected shopping malls, selecting women at random from the mall, not only from the booth.

(f)[2] Women who ~~more~~ stay at home may prefer some more classical lipstick, and the sample may be biased. ~~Also~~, these differences may be small, and the bias may be small. 1

MRF can use telephone interviews, or ~~[illegible]~~ mailed questionaires, but it would be much more expensive. 1

16

2)

(a) 4

(i) $\hat{\tau}_{y+z} = N\,\overline{y+z} = 24 \times \frac{21+15}{5} = 172.8$ | 1

(ii) $\hat{\tau}_z = 24 \times \frac{15}{5} = 72$ | 1

(iii) $\hat{\mu}_{y+z} = \overline{y+z} = \frac{21+15}{5} = 7.2$ | 1

(iv) $\hat{\tau}_x = N\hat{\mu}_x = 24\,\frac{236}{5} = 1132.8$ | 1

(b) 4

(i) $30 \times \hat{\mu}_x = 30\,\frac{236}{5} = 1416.$ " average monthly income per household

the average monthly income per adult

$$30 \times \hat{r} = \frac{30 \times \hat{\mu}_x}{\hat{\mu}_y} = \frac{1416}{\left(\frac{21}{5}\right)} = 337.14 \quad \left(\hat{r} = \frac{\hat{\mu}_x}{\hat{\mu}_y} = \frac{47.2}{4.2} = 11.24\right)$$

(ii) $\hat{\mu}_z = \frac{15}{5} = 3$ | 1

(iii) $\hat{p} = \frac{\hat{\tau}_z}{\hat{\tau}_{y+z}} = \frac{\hat{\mu}_z}{\hat{\mu}_y + \hat{\mu}_z} = \frac{\sum z_i}{\sum(y_i + z_i)} = \frac{15}{21+15} = 41.6\%$ | 1

(iv) $\hat{\mu}_{\frac{z}{z+y}} = \frac{1}{5}\sum \frac{z_i}{z_i + y_i} = \frac{1}{5}\left(\frac{2}{5} + \frac{4}{8} + \frac{1}{4} + \frac{4}{9} + \frac{4}{10}\right) = 0.399$

$= 39.9\%$ | 1

2) (c) 4 (i) - unbiased and ratio |1 $\boxed{4}$

(ii) - unbiased |1

(iii) - ratio |1

(iv) - unbiased |1

(d) 4 $Var(30\hat{R}) = 30^2 \frac{N-n}{N} \frac{1}{\hat{\mu}_4^2} \frac{S_2^2}{n} = 30^2 \frac{24-5}{24} \frac{1}{(4.2)^2} \frac{S_2^2}{5}$

$\left(S_2^2 = \frac{1}{5-1}(12190 - 2\times 11.24\times 1067 + 11.24^2 \times 95) = \frac{205.912}{4} = \right.$

$\left. = 51.478 \right)$ |2

$Var(30\hat{R}) = 30^2 \times 0.4621$

$B = 2\times 30\sqrt{0.4621} = 40.78$ |2

3) $\boxed{20}$ (a) $\boxed{5}$

$N = 24,\ N_1 = 17,\ N_2 = 7$

$n = 5,\ n_1 = 3,\ n_2 = 2$

Total size $\bar{z}_1 = \frac{1}{3}(3+4+3) = 3.33,\ \bar{z}_2 = \frac{1}{2}(6+6) = 6$

$\hat{\tau}_z = N_1\bar{z}_1 + N_2\bar{z}_2 = 17\times 3.33 + 7\times 6 = \underline{98.61}$ |2

Total daily income

$\bar{y}_1 = \frac{1}{3}(33+39+35) = 35.67,\ \bar{y}_2 = \frac{1}{2}(62+61) = 61.5$

$\hat{\tau}_y = N_1\bar{y}_1 + N_2\bar{y}_2 = \underline{1036.84}$ |2

Total food cost

$\bar{x}_1 = 22.33,\ \bar{x}_2 = 30$

$\hat{\tau}_x = N_1\bar{x}_1 + N_2\bar{x}_2 = \underline{589.61}$ |2

$379.61 + 210$

(b) 3 $\hat{V}ar(\hat{\tau}_x) = N_1^2\hat{V}ar(\bar{x}_1) + N_2^2\hat{V}ar(\bar{x}_2) =$

$= N_1^2\frac{N_1-n_1}{N_1}\frac{s_1^2}{n_1} + N_2^2\frac{N_2-n_2}{N_2}\frac{s_2^2}{n_2} =$

$= 17(17-3)\frac{2.33}{3} + 7(7-2)\frac{2}{2} = 219.85$

$B_{\tau_x} = 2\sqrt{\hat{V}ar(\hat{\tau}_x)} = 29.65$

(c) 4 $\hat{r}_{y/z} = \frac{\hat{\tau}_y}{\hat{\tau}_z} = \frac{1036.84}{98.61} = 10.51$ — average daily income |2

percentage of food cost:

$\hat{r}_{x/y} = \frac{\hat{\tau}_x}{\hat{\tau}_y} = \frac{589.67}{1036.84} = 0.5687 = 56.87\%$ |2

Both estimators are ratio estim. — biased.

3) (d) 3 $\hat{\mu}_{x/y} = \frac{1}{N}(N_1 \hat{\mu}_1 + N_2 \hat{\mu}_2) =$ $\boxed{6}$

$= \frac{1}{24}(17 \times \frac{1}{3}(\frac{21}{33} + \frac{24}{39} + \frac{22}{35}) + 7 \times \frac{1}{2}(\frac{31}{62} + \frac{29}{61}))$

$= \frac{1}{24}(17 \times 0.6268 + 7 \times 0.4877) = 0.5862 = 58.62\%$ | 2

– unbiased, it is not a ratio estimator | 1

(e) 2 It would be, because of two different household sized in strata (1–5 person, and over 5 persons) and then different mean values of observed variables.

(f) 2 (i) ~~It~~ The stratification is good, because it is done by household size, which is highly correlated with variables of interest. | 1

(ii) It is likely that optimal all. will be better than proportional, because variances of variables such as income and expenditures ~~a~~ on food will be greater in greater households (i.e. different strata). | 1

4) $\boxed{14}$ $\boxed{7}$

(a) 3 (i) It is because the companies are ordered by their 2006 revenue ("decreasing" population), whic is a ~~good~~ case for systematic sampling. |1

(ii) Selection step is $240/10 = 24$. |1

(iii) It is not, because sample 1 starts with 13th company, and sample 3 with 28th company, which has smaller income than the 13th comp., and similar for 2, 3, ... company in both samples. |1

(b) 4 (i) $\hat{\tau} = \tau_1 + \hat{\tau}_2 = 40{,}126{,}556 + 240 \times \bar{y}_{50} =$

$= 40{,}126{,}556 + 240 \times \cancel{59{,}868.74}\ 50{,}362.78 =$ |2

$= 40{,}126{,}556 + \cancel{14{,}368{,}497.6}\ 12{,}087{,}067.2 = \cancel{54{,}495{,}053.6} = 52{,}213{,}623.2$

$\left(\bar{y}_{50} = \frac{1}{5}\sum_{1}^{5}\frac{y_i}{10} = \frac{1}{5}\left(\frac{847{,}881}{10} + \dots + \frac{477{,}650}{10}\right) = \cancel{59{,}868.74}\ 50{,}362.78\right)$

(ii) $\mu_1 = \frac{\tau_1}{10} = \frac{40{,}126{,}556}{10} = 4{,}012{,}655.6$ |1

(iii) $\hat{\mu}_2 = \bar{y}_{50} = \cancel{59{,}868.74}\ 50{,}362.78$ |1

~~(c) SD(μ_1) = 0 – no error (all values known)~~

~~SD($\hat{\mu}_2$) =~~

4) for (b)(iii) 8

(c) ~~each~~ mean revenues in the samples are

4) $\bar{y}_i$ 84,799.1, 55,849.1, 42,692.7, 20,708.0

47,765.0

This is repeated systematic sampling and from 24 possible samples, and

$$\hat{Var}(\hat{\mu}_2) = \frac{24-5}{24}\frac{s_y^2}{5} = \frac{19}{24}\frac{540,236,101.8}{5} =$$

$= 85,537,382.79$

$\hat{SD}(\hat{\mu}_2) = 9,248.64$ | 3

for (b)(ii) μ_2 - is actual value, because all companies (10) are included.

$Var(\mu_2) = 0$ | 1

(d) 3 $\hat{Var}(\bar{y}_{SRS}) = \frac{240-50}{240}\frac{\hat{\sigma}^2}{50} = \frac{19}{24}\frac{(250,000)^2}{50}$

$\hat{\sigma} = \frac{\text{range}}{4} = \frac{1000,000}{4} = 250,000$ | 1

$SD(\bar{y}_{SRS}) = 31,457.64 > \hat{SD}(\hat{\mu}_2) = 9,248.64$ | 2

almost 3 times, i.e. SRS is worse.

(e) - use difference metod with | 4

$d_i \approx \frac{1,000,000}{50} = 20,000$, $s_d^2 = \frac{1}{2(n-1)}\sum_1^{n-1} d_i^2 = \frac{(20,000)^2}{2}$

$Var(\bar{y}_{sy}) = \frac{240-50}{240}\frac{s_d^2}{50} = \frac{19}{24}\frac{(20,000)^2}{100} = (1779.51)^2$

BONUS

$\boxed{20}$

5) 5 (a) One stage cluster sampling, with places as sampling units.

$N=500, \quad n=8$

$\hat{\tau}_x = 500 \times \bar{x} = 500 \times \frac{53}{8} = 500 \times 6.625 = 3312.5$ | 3

$\hat{Var}(\hat{\tau}_x) = N^2 \hat{Var}(\bar{x}) = 500^2 \frac{500-8}{500} \times \frac{6.268}{8} = 192{,}741$ ($6.268 = S_x^2$)

$(S_x^2 = \frac{1}{n-1}(\Sigma x_i^2 - n\bar{x}^2) = \frac{1}{7}(395 - 8 \times 6.625^2) = 6.268)$

$B_\tau = 2 \times \sqrt{\hat{Var}(\hat{\tau}_x)} = 878$ | 2

2 (b) $\hat{\tau}_y = N\bar{y} = 500 \times \frac{28}{8} = 500 \times 3.5 = 1750$ | 2

(c) 5 $p = R = \frac{\tau_y}{\tau_x}$, $\hat{p} = \frac{\Sigma y_i}{\Sigma x_i} = \frac{28}{53} = 0.5283$ | 3

$\hat{Var}(\hat{p}) = \frac{N-n}{N} \frac{1}{\bar{x}^2} \frac{S_r^2}{n} = \frac{500-8}{500} \frac{1}{6.625^2} \frac{0.440}{8} = 1.233 \times 10^{-3}$

$B_p = 2\sqrt{\hat{Var}(\hat{p})} = 0.070$ | 2

$(S_r^2 = \frac{1}{7}(120 - 2 \times 0.5283 \times 215 + 0.5283^2 \times 395) = 0.440)$

$\Sigma x_i y_i = 215$

5) (d) 3 $\hat{\tau} = N \times \hat{\bar{\tau}} = N \times \hat{\mu}_t = 500 \times \frac{1}{n} \sum y_i =$

$= 500 \times \frac{51}{8} = 3187.5$ ($\times$ \$1000) |2

It is possible to calculate $\hat{Var}(\hat{\tau}) =$

$= N^2 \, Var(\hat{\mu}_t) = N^2 \frac{N-n}{N} \frac{s_t^2}{n}$,

$s_t^2 = \frac{1}{n-1} \sum (y_i - \hat{\mu}_t)^2$, y_i are given, so

s_t^2 can be calculated. |1

(e) 2 (i) the average daily revenue per franchise |1

(ii) percentage of franchises in shopping malls |1

and others, such as the average number of fran. per place. |2

(f) 3 $\tau_x = 3300$, $\hat{\tau}_y = r_{y/x} \times \tau_x = 3300 \times \frac{28}{53} = 1743.4$

This is a ratio estimator and should be better than unbiased (SRS), because x and y are correlated. |1

6) (a) 3 Two stage cluster sample $\boxed{11}$

$\boxed{16}$ Target population: all trees in the forest in the ~~county~~ country/region |1

sampling population: trees from selected county |1

M - # of plots, $N = 10$ - # of areas

$$\hat{M} = N \times \hat{\bar{M}} = 10 \times \frac{1}{3}(12+16+14) = 10 \times 14 = 140 \quad |1$$

(b) 5 Ratio estimator can be used:

$$\hat{\mu}_r = \frac{\sum M_i \bar{y}_i}{\sum M_i} = \frac{12 \times 3.67 + 16 \times 2.4 + 14 \times 2.5}{12+16+14} = 2.796 \quad |2$$

Biased estimator. |1

$$\hat{Var}(\hat{\mu}_r) = \frac{N-n}{N} \frac{1}{n\hat{\bar{M}}^2} s_r^2 + \frac{1}{nN\hat{\bar{M}}^2} \sum M_i^2 \frac{M_i - m_i}{M_i} \frac{s_i^2}{m_i}$$

Area	M_i	m_i	$\bar{x}_i$	s_i^2
1	12	3	3.67	2.33
2	16	5	2.4	2.80
3	14	4	2.5	1.67

from data

$$s_r^2 = \frac{1}{n-1} \sum M_i^2 (\bar{x}_i - \hat{\mu}_r)^2 = \frac{1}{3-1}[12^2(3.67-2.80)^2 + \dots] = 83.80 \quad |2$$

(calc. not required)

$$\hat{Var}(\hat{\mu}_r) = \frac{10-3}{10} \frac{83.80}{3 \times 14^2} + \frac{1}{3 \times 10 \times 14^2} \times$$

$$\times \left(12^2 \frac{12-3}{12} \frac{2.33}{3} + \dots\right) = 0.14073, \; B = 0.750$$

6) (c) 4 $\hat{\mu}_t = \frac{1}{n}\sum_{1}^{n} M_i \bar{x}_i = \frac{1}{3} 117.44$ $\boxed{12}$

$= \frac{1}{3}(12 \times 3.67 + 16 \times 2.4 + 14 \times 2.50) = 39.15$ ✓ /3

- unbiased estimator /1

same value will be obtained

from $\hat{M}\hat{\mu}_r = 14 \times 2.796$

(d) 4 $\hat{p} = \frac{\hat{\tau}_x}{\hat{\tau}_y} = \frac{\sum M_i \bar{x}_i}{\sum M_i \bar{y}_i} = \cancel{117.44}$ /2

$= \frac{\cancel{12 \times 3.67}\ 117.44}{12 \times \frac{53}{3} + 16 \times \frac{147}{5} + 14 \times \frac{54}{4}} = 0.135 = 13.5\%$ /2