Queen Mary
University of London

B. Sc. Examination 2008 By Course Unit

MAS 234 Sampling, Surveys and Simulation

Duration: 1.5 hours

Date and time: Nov 5th 2007, 1310–1440

Suggested solutions and marking scheme: marks indicated in square brackets. Text in square brackets is explanatory but not necessary.

Question 1 The *sampling distribution* of an estimator is the collection of all[1] possible samples[1], together with the corresponding estimates[1] and their probabilities[1] under the sampling design[1] for a fixed population. [5]

Question 2 Under SRS size n for any population size $N > n$:

[**2**] $E[\bar{y}] = \bar{Y}$ [i.e. the sample mean is a UE of the population mean]

[**3**] $\text{Var}[\bar{y}] = \frac{1-f}{n} S^2$ where $f = \frac{n}{N}$

[**2**] $E[s^2] = S^2$ [i.e. the sample variance is a UE of the population variance]

[**3**] The key step in the proof is to write

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{N} t_1 Y_i,$$

where $t_i = 1$ if unit i is in the sample, zero otherwise [sample indicator r.v.].

[10]

Question 3 Initial sample is $\{123456\}$. Select new unit 7 with prob. 6/7: $u = .4\ldots < 6/7$ so select unit 7 and replace $|\underline{1\ 2\ 3\ 4\ 5\ 6}|$? $u = .8332\ldots < 5/6$ ($> 4/6$) so replace unit 5, sample is now $\{123467\}$. Select unit 8 with prob. 6/8: $u = .5\ldots < 6/8$ so select and replace $|\underline{1\ 2\ 3\ 4\ 6\ 7}|$? $u = .0\ldots, 1/6$ so replace unit 1, sample is now $\{234678\}$. Select unit 9 with prob. 6/9: $u = .7\ldots > 6/9$ so do not select, and final sample is $\{234678\}$. [3/10 for some idea, 5/10 if used 1 digit all the time, 7/10 if used 2 or 3 digits all the time]. [10]

© Queen Mary, University of London 2008

Question 4 [**4**] $\bar{y}_{st} = \sum_{h=1}^{H} W_h \bar{y}_h = \frac{325}{725} \times 15.67 + \frac{400}{725} \times 8.89 = 11.93$ [4s.f.]

[**6**] $v[\bar{y}_{st}] = \sum_{h=1}^{H} W_h^2 \frac{1-f_h}{n_h} s_h^2 = \left(\frac{325}{725}\right)^2 \left(\frac{1}{10} - \frac{1}{325}\right)(3.65)^2 + \left(\frac{400}{725}\right)^2 \left(\frac{1}{12} - \frac{1}{400}\right)(2.08)^2 =$ 0.36593 and thus s.e. is 0.6049 [4s.f.]

[**3**] with 95% C.I.: $\bar{y}_{st} \pm 1.96\sqrt{v[\bar{y}_{st}]} = (10.74, 13.12)$ [4s.f.]

[13]

Question 5 [**2**] optimal allocation (Neyman) is $n_h \propto N_h S_h$, and $N_1 S_1 = 1186.25$ $N_2 S_2 = 832$

[**6**] so $n_1 = 50 \times 1186.25/2018.25 = 29.38 \rightarrow n_1 = \underline{29}$
$n_2 = 50 \times 832/2018.25 = 20.61 \rightarrow n_2 = \underline{21}$

[**5**] achieved variance is $\underbrace{\frac{W_1^2 S_1^2}{n_1} + \frac{W_2 S_2^2}{n_2}}_{0.15503} - \underbrace{\frac{W_1 S_1^2 + W_2 S_2^2}{N}}_{0.01153} = 0.1435$ [4s.f.]

[13]

Question 6 (a) $\hat{\bar{Y}}_R = \bar{X}\frac{\bar{y}}{\bar{x}}$ [3]. No[1], [it is not unbiased under SRS]. [4]

(b) $\text{Var}[\hat{\bar{Y}}_R] = \frac{1-f}{n}\left(S_Y^2 - 2RS_{XY} + R^2 S_X^2\right)$ [3]. No[1], this formula is not exact. [4]

(c) $v[\hat{\bar{Y}}_R] = \frac{1-f}{n}(s_Y^2 - 2\hat{R}s_{XY} + \hat{R}^2 s_X^2)$. [4]

Question 7 (a) $\hat{R} = \bar{y}/\bar{x} = 0.5$[4], $v[\hat{R}] = \frac{1-4/10000}{4\times 2.2^2}(1.58\dot{3} - 2 \times 0.5 \times (-0.1\dot{6}) + (0.5)^2 \times 1.\dot{6}) = 0.11186$ [7] [11]

(b) $\hat{R}^* = \bar{y}/\bar{X} = \frac{5/4}{2.2} = 0.5682$ [3], $v[\hat{R}^*] = \frac{1-4/10000}{4\times 2.2^2} \times 1.58\dot{3} = 0.08175$ [4] [7]

(c) $\tilde{R} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i = \frac{1}{4}\left(\frac{1}{2} + \frac{1}{1} + \frac{3}{3} + \frac{0}{4}\right) = 0.625$ [4],
$v[\tilde{R}] = \frac{1}{n}\sum_{i=1}^{n}(\bar{y}_i - \tilde{R})^2/(n-1) = \frac{1}{4}\frac{1}{3}(0.5^2 + 1^2 + 1^2 + 0^2 - 4 \times 0.625^2) = 0.05729$ [7] [11]

(d) $p = 5/10 = 0.5$ [3], $v[p] = \frac{p(1-p)}{n-1}(1-f) = \frac{0.5^2}{9} \times (1 - 10/20000) = 0.02776$ [4] [7]

© Queen Mary, University of London 2008 END OF EXAMINATION