MSE's of the alternative forms of $\hat{Y}_{HT}$ and $\hat{Y}_M$, these forms doing almost as well as $\hat{Y}_{SRS}$. For both methods the $(\text{Bias})^2$ term was about 4% of the MSE.

EXERCISES

9A.1 Horvitz and Thompson (1952) give the following data for eye estimates M_i of the numbers of households and for the actual numbers y_i in 20 city blocks in Ames, Iowa. To assist in the calculations, values of $\bar{y}_i$ and $\bar{y}_i^2/M_i$ are also given. A sample of $n = 1$ block is chosen. Compute the variances of the total number of households Y, as obtained by (a) the unbiased estimate in sampling with equal probabilities, (b) the ratio estimate in sampling with equal probabilities, (c) sampling with probability proportional to M_i. (For the ratio estimate, compute the true mean square error, not the approximate formula.)

M_i	y_i	$\bar{y}_i$	y_i^2/M_i	M_i	y_i	$\bar{y}_i$	y_i^2/M_i
9	9	1.0000	9.000	19	19	1.0000	19.000
9	13	1.4444	18.778	21	25	1.1905	29.762
12	12	1.0000	12.000	23	27	1.1739	31.696
12	12	1.0000	12.000	24	21	0.8750	18.375
12	14	1.1667	16.333	24	35	1.4583	51.042
14	17	1.2143	20.643	25	22	0.8800	19.360
14	15	1.0714	16.071	26	25	0.9615	24.038
17	20	1.1765	23.529	27	27	1.0000	27.000
18	19	1.0556	20.056	30	47	1.5667	73.633
18	18	1.0000	18.000	40	37	0.9250	34.225

Do the results agree with the discussion in section 9A.5?

9A.2 A questionnaire is to be sent to a sample of high schools to find out which schools provide certain facilities, for example, a course in Russian or a swimming pool. If M_i is the number of students in the ith school, the quantity to be estimated for any given facility is the proportion P of high-school students who are in schools having the facility, that is,

$$P = \frac{\sum_{w} M_i}{\sum_{i=1}^{N} M_i}$$

where $\sum_w$ is a sum over those schools *with* the facility.

A sample of n schools is drawn with probability proportional to M_i with replacement. For one facility, a schools out of n are found to possess it. (a) Show that $\hat{P} = a/n$ is an unbiased estimate of P and that its true variance is $P(1-P)/n$. (*Hint.* In the corollary to theorem 9.4 let $y_i = M_i$ if the school has the facility and 0 otherwise.) (b) Show that an unbiased estimate of $V(\hat{P})$ is $v(\hat{P}) = \hat{P}(1-\hat{P})/(n-1)$.

9A.3 The large units in a population arrange themselves into a finite number of size classes: all units in class h contain M_h small units. (a) Under what conditions does sampling with *pps* give, on the average, the same distribution of the size classes in the sample as stratification by size of unit, with optimum allocation for fixed sample size? (b) If the variance among large units in class h is kM_h, where k is a constant for all classes, what system of probabilities of selection of the units gives a sample in which the sizes have approximately the same distribution as a stratified random sample with optimum allocation for fixed sample size?

9A.4 For a population with $N=3$, $z_i=\frac{1}{2},\frac{1}{3},\frac{1}{6}$ and $y_i=7, 5, 2$, two units are drawn without replacement, the first with probability proportional to z_i, the second with probability proportional to the remaining sizes. (*a*) Verify that $\pi_1=\frac{51}{60}$. $\pi_2=\frac{44}{60}$, $\pi_3=\frac{25}{60}$ and that $\pi_{12}=\frac{35}{60}$, $\pi_{13}=\frac{16}{60}$, $\pi_{23}=\frac{9}{60}$. (*b*) For this method of sample selection, compare the variances of $\hat{Y}_{HT}$ and $\hat{Y}_M$ and also compare them with the variance of $\hat{Y}_{RHC}$ using its method of sample selection. You may either construct all three possible estimates or use the variance formulas. (*c*) Show that the ratio of $V(\hat{Y}_M)$ to $V(\hat{Y}_{ppz})$ in sampling with replacement is close to the value $\frac{1}{2}$ that applies for equal probability sampling.

9A.5 For the population in exercise 9A.4, a second variable had values $y_{2i}=8, 5, 9$, not at all closely related to the z_i, so that with the sampling method used in exercise 9A.4, $\hat{Y}_{HT}$ and $\hat{Y}$ would be expected to perform poorly for this variable. For Rao's estimator $\hat{Y}^*_{HT}=1.5(y_{2i}+y_{2j})$, compare its MSE with the variance of $\hat{Y}_{SRS}=1.5(y_{2i}+y_{2j})$ in equal-probability sampling. How much does bias contribute to the MSE?

9A.6 For Brewer's method with $n=2$, section 9A.8 showed that

$$\pi_{ij}=\frac{4z_iz_j(1-z_i-z_j)}{(1-2z_i)(1-2z_j)}\bigg/\left(1+\sum_i^N\frac{z_i}{1-2z_i}\right)$$

(*a*) Show that if every $z_i<\frac{1}{2}$,

$$0<\pi_{ij}<4z_iz_j \qquad (\text{every } i\neq j)$$

(*b*) Show that this result makes the Yates–Grundy estimator of variance always positive for this method.

Hint. For (*a*) it is sufficient to show that

$$(1-z_i-z_j)=(1-2z_i)(1-2z_j)\left[1+\frac{z_i}{1-2z_i}+\frac{z_j}{1-2z_j}\right]$$

9A.7 (*a*) For Durbin's method with $n=2$, verify directly that the probability that the jth unit is drawn second is z_j as stated on p. 263.

(*b*) With $N=4$, $z_i=0.1$, 0.2, 0.3, 0.4, calculate the probability that with Brewer's method unit 1 is drawn first and the probability that it is drawn second. Verify that the two probabilities add to $0.2=2z_i$.

9A.8 In Madow's systematic method, a unit may be chosen more than once in the sample if $nz_i>1$ (i.e., $nM_i'>M_0'$). Show (as stated on p. 266) that for such units the average frequency of selection is nz_i, so that the Horvitz–Thompson estimator of Y remains unbiased for $nz_i>1$.