To summarize, *BRR* consistently performs best in Table 11.8. Except for multiple R's, it can be regarded as adequate for practical use if one has the view that in data analysis a tabular 5% tail value represents an actual tail value somewhere between 3 and 8%. J does slightly better than Taylor. Except for *BRR* with ratios and simple regressions, all methods give actual tail frequencies higher than the t-tables, so that confidence probabilities are overstated. A puzzling feature is that for correlation coefficients the increase in sample size from 12 to 60 has not brought a corresponding improvement in the closeness of the actual to the t tail frequencies.

This study opens up a wide area for investigation of the methods with different survey plans and different types of estimator $f(\mathbf{Y})$.

In a Monte Carlo study of a larger, more complex sample (two-stage *pps* sampling with replacement, including both stratification and poststratification) Bean (1975) compared the Taylor and *BRR* methods for estimators of the ratio type. Both methods gave satisfactory variance estimates and adequate two-sided confidence probabilities calculated from the normal distribution. Sufficient skewness remained, however, so that one-sided confidence intervals could not be trusted.

EXERCISES

11.1 By working out the estimates for all possible samples that can be drawn from the artificial population in Table 11.1, by methods Ia, Ib, II, and III, verify the total MSE's given in Table 11.2.

11.2 For methods II (equal probabilities, unbiased estimate) and III(pps selection), recompute the variances of $\hat{\bar{Y}}$ for the example in Table 11.1 when $m_i = 1$. Show that the precision of method III in relation to method II is lower for $m_i = 1$ than for $m_i = 2$. What general result does this illustrate?

11.3 For the population in Table 11.1, if the estimated sizes z_i are 0.1, 0.3 and 0.6, with $m_i = 2$, show that the unbiased estimate (method IV) gives a smaller variance than *pps* sampling. What is the explanation of this result?

11.4 The elements in a population with three primary units are classified into two classes. The unit sizes M_i and the proportions P_i of elements that belong to the first class are as follows.

$$M_1 = 100, \quad M_2 = 200, \quad M_3 = 300, \qquad P_1 = 0.40, \quad P_2 = 0.45, \quad P_3 = 0.35$$

For a sample consisting of 50 elements from one primary unit, compare the MSE's of methods Ia, II, and III for estimating the proportion of elements in the first class in the population. (In the variance formulas in section 11.2, S_i^2 is approximately P_iQ_i.)

11.5 A sample of n primary units is selected with equal probabilities. From each chosen unit, a constant fraction f_2 of the subunits is taken. If a_i out of the m_i subunits in the ith unit fall in class C, show that the ratio-to-size estimate (section 11.8) of the population proportion in class C is $\bar{p} = \Sigma a_i/\Sigma m_i$. From formula (11.36), show that an estimate of MSE($\bar{p}$) is

$$v(\bar{p}) = \frac{1-f_1}{n\bar{M}^2} \frac{\sum^n M_i^2(p_i - \bar{p})^2}{n-1} + \frac{f_1(1-f_2)}{n^2\bar{m}\bar{M}} \sum^n \frac{M_i m_i}{m_1 - 1} p_i q_i$$

where $p_i = a_i/m_i$.

11.6 A firm with 36 factories decides to check the condition of some equipment of which $M_0 = 25{,}012$ pieces are in use. A random sample of 12 factories is taken, a 10% subsample being checked in each selected factory. The numbers of pieces checked (m_i) and the numbers found with signs of deterioration (a_i) are as follows.

Factory	m_i	a_i	$p_i = \frac{a_i}{m_i}$	Factory	m_i	a_i	$p_i = \frac{a_i}{m_i}$
1	65	8	0.123	7	85	18	0.212
2	82	21	0.256	8	73	11	0.151
3	52	4	0.077	9	50	7	0.140
4	91	12	0.132	10	76	9	0.118
5	62	1	0.016	11	64	20	0.312
6	69	3	0.043	12	50	2	0.040

Estimate the percentage and the total number of defective pieces in use and give estimates of their standard errors.

Note. Since $M_i/\bar{M} \doteq m_i/\bar{m}$, the between-units component of $v(\bar{p})$ may be computed as

$$\frac{1-f_1}{n\bar{m}^2(n-1)}\left(\sum a_i^2 - 2\bar{p}\sum a_i m_i + \bar{p}^2 \sum m_i^2\right)$$

and, since the m_i are fairly large, the within-units component as

$$\frac{f_1(1-f_2)}{(n\bar{m})^2}\sum a_i q_i$$

11.7 If primary units are selected with equal probabilities and f_2 is constant, show that in the notation of exercise 11.5 the unbiased estimate of a population proportion is $p = N\Sigma a_i/nM_0 f_2$ and that, if terms in $1/m_i$ are negligible, its variance may be computed as

$$v(p) = \frac{1-f_1}{n(n-1)\bar{m}^2}\sum^n (a_i - \bar{a})^2 + \frac{f_1(1-f_2)}{(n\bar{m})^2}\sum^n a_i q_i$$

Calculate p and its standard error for the data in exercise 11.6.

11.8 A sample of n primary units is chosen with probabilities proportional to estimated sizes z_i (with replacement) and with a constant expected over-all sampling fraction f_0. Show that the unbiased and the ratio-to-size estimates of the population total are, respectively, T/f_0 and $TM_o/\sum^n m_i$, where T is the sample total. (It follows that if M_0 is not known the unbiased estimate can be used, but not the ratio to size. For estimating the population mean per subunit, the situation is reversed.)

11.9 In a study of overcrowding in a large city one stratum contained 100 blocks of which 10 were chosen with probabilities proportional to estimated size (with replacement). An expected over-all sampling fraction $f_o = 2\%$ was used. Estimate the total number of persons and the average persons per room and their s.e.'s from the data below.

Block	1	2	3	4	5	6	7	8	9	10
Rooms	60	52	58	56	62	51	72	48	71	58
Persons	115	80	82	93	105	109	130	93	109	95

11.10 For Durbin's method (section 11.10) of simplifying variance estimation in *ppz* sampling without replacement, a simple method of sample selection, due essentially to Kish

(1965), is as follows. The subscript h to denote the stratum will be omitted and the number of primary units is assumed to be even.

Arrange the units in order of increasing z_i and mark them off in pairs. The method is exact only if $z_i = z_j$ for members of the same pair; this will be assumed here. Select two units *ppz with replacement*. If two different units are drawn, accept both. If the same unit is drawn twice, let the sample consist of the two members of the pair to which this unit belongs. Show that for this method: (a) $\pi_i = 2z_i$, (b) for units not in the same pair, $\pi_{ij} = 2z_iz_j = \pi_i\pi_j/2$, so that $\pi_i\pi_j\pi_{ij}^{-1} - 1 = 1$, and ($c$) for units in the same pair, $\pi_{ij} = 4z_iz_j = \pi_i\pi_j$, so that $\pi_i\pi_j\pi_{ij}^{-1} - 1 = 0$.

11.11 In section 11.9, formula (11.33) for $V(\hat{Y}_{ppz})$ in sampling with replacement was proved under the plan that whenever the ith unit was selected, an independent simple random subsample of size m_i was drawn from the whole of the unit. Prove the following results for two alternative plans.

(a) When the ith unit is selected t_i times, a simple random subsample of size m_it_i is drawn from it (assume $m_it_i \leq M_i$). Under this plan, $V(\hat{Y}_{ppz})$ in (11.41) is reduced by $(n-1)\sum^N M_iS_{2i}^2/n$ (Sukhatme, 1954).

(b) When the ith unit is selected t_i times, a simple random subsample of size m_i is drawn. Then $V(\hat{Y}_{ppz})$ in (11.41) is increased by

$$\frac{(n-1)}{n}\sum^N M_i^2(1-f_{2i})S_{2i}^2/m_i$$

In both (a) and (b), $\hat{Y}_{ppz} = \sum^N t_iM_i\bar{y}_i/nz_i$, the ith unit receiving weight t_i.