where n_0 is the first approximation, which ignores the fpc, and n is the corrected value taking account of the fpc. In the development of these formulas, the factors $N_h/(N_h - 1)$ have been taken as unity.

These results apply to the estimate of a *proportion*. If it is preferable to think in terms of percentages, the same formulas apply if p_h, q_h, V, and so forth, are expressed as percentages. For the estimation of the total number in the population in class C, that is, of NP, all variances are multiplied by N^2.

EXERCISES

5.1 In a population with $N = 6$ and $L = 2$ the values of y_{hi} are 0, 1, 2 in stratum 1 and 4, 6, 11 in stratum 2. A sample with $n = 4$ is to be taken. (*a*) Show that the optimum n_h under Neyman allocation, when rounded to integers, are $n_h = 1$ in stratum 1 and $n_h = 3$ in stratum 2. (*b*) Compute the estimate $\bar{y}_{st}$ for every possible sample that can be drawn under optimum allocation and under proportional allocation. Verify that the estimates are unbiased. Hence find $V_{opt}(\bar{y}_{st})$ and $V_{prop}(\bar{y}_{st})$ directly. (*c*) Verify that $V_{opt}(\bar{y}_{st})$ agrees with the formula given in equation (5.6) and that $V_{prop}(\bar{y}_{st})$ agrees with the formula given in equation (5.8), page 93. (*d*) Use of formula (5.27), page 99, to compute $V_{opt}(\bar{y}_{st})$ is slightly incorrect because it does not allow for the fact that the n_h were rounded to integers. How well does it agree with the corrected value?

5.2 The households in a town are to be sampled in order to estimate the average amount of assets per household that are readily convertible into cash. The households are stratified into a high-rent and a low-rent stratum. A house in the high-rent stratum is thought to have about nine times as much assets as one in the low-rent stratum, and S_h is expected to be proportional to the square root of the stratum mean.

There are 4000 households in the high-rent stratum and 20,000 in the low-rent stratum. (*a*) How would you distribute a sample of 1000 households between the two strata? (*b*) If the object is to estimate the difference between assets per household in the two strata, how should the sample be distributed?

5.3 The following data show the stratification of all the farms in a county by farm size and the average acres of corn (maize) per farm in each stratum. For a sample of 100 farms, compute the same sizes in each stratum under (*a*) proportional allocation, (*b*) optimum allocation. Compare the precisions of these methods with that of simple random sampling.

Farm Size (acres)	Number of Farms N_h	Average Corn Acres $\bar{Y}_h$	Standard Deviation S_h
0–40	394	5.4	8.3
41–80	461	16.3	13.3
81–120	391	24.3	15.1
121–160	334	34.5	19.8
161–200	169	42.1	24.5
201–240	113	50.1	26.0
241–	148	63.8	35.2
Total or mean	2010	26.3	

5.4 Prove the result stated in formula (5.38), section 5.6:

$$V_{ran} = V_{prop} + \frac{(1-f)}{n(N-1)}\left[\sum N_h(\bar{Y}_h - \bar{Y})^2 - \frac{1}{N}\sum (N - N_h)S_h^2\right]$$

5.5 A sampler has two strata with relative sizes W_1, W_2. He believes that S_1, S_2 can be taken as equal but thinks that c_2 may be between $2c_1$ and $4c_1$. He would prefer to use proportional allocation but does not wish to incur a substantial increase in variance compared with optimum allocation. For a given cost $C = c_1n_1 + c_2n_2$, ignoring the fpc, show that

$$\frac{V_{prop}(\bar{y}_{st})}{V_{opt}(\bar{y}_{st})} = \frac{W_1c_1 + W_2c_2}{(W_1\sqrt{c_1} + W_2\sqrt{c_2})^2}$$

If $W_1 = W_2$, compute the relative increases in variance from using proportional allocation when $c_2/c_1 = 2, 4$.

5.6 A sampler proposes to take a stratified random sample. He expects that his field costs will be of the form $\sum c_hn_h$. His advance estimates of relevant quantities for the two strata are as follows.

Stratum	W_h	S_h	C_h
1	0.4	10	\$4
2	0.6	20	\$9

(*a*) Find the values of n_1/n and n_2/n that minimize the total field cost for a given value of $V(\bar{y}_{st})$. (*b*) Find the sample size required, under this optimum allocation, to make $V(\bar{y}_{st}) = 1$. Ignore the fpc. (*c*) How much will the total field cost be?

5.7 After the sample in exercise 5.6 is taken, the sampler finds that his field costs were actually \$2 per unit in stratum 1 and \$12 in stratum 2. (*a*) How much greater is the field cost than anticipated? (*b*) If he had known the correct field costs in advance, could he have attained $V(\bar{y}_{st}) = 1$ for the original estimated field cost in exercise 5.6? (*Hint.* The Cauchy–Schwarz inequality, page 97, with $V' = 1$, gives the answer to this question without finding the new allocation.)

5.8 In a stratification with two strata, the values of the W_h and S_h are as follows.

Stratum	W_h	S_h
1	0.8	2
2	0.2	4

Compute the sample sizes n_1, n_2 in the two strata needed to satisfy the following conditions. Each case requires a separate computation. (Ignore the fpc.) (*a*) The standard error of the estimated population mean $\bar{y}_{st}$ is to be 0.1 and the total sample size $n = n_1 + n_2$ is to be minimized. (*b*) The standard error of the estimated mean of each stratum is to be 0.1. (*c*) The standard error of the difference between the two estimated stratum means is to be 0.1, again minimizing the total size of sample.

5.9 With two strata, a sampler would like to have $n_1 = n_2$ for administrative convenience, instead of using the values given by the Neyman allocation. If $V(\bar{y}_{st})$, $V_{opt}(\bar{y}_{st})$ denote the variances given by the $n_1 = n_2$ and the Neyman allocations, respectively, show that the fractional increase in variance

$$\frac{V(\bar{y}_{st}) - V_{opt}(\bar{y}_{st})}{V_{opt}(\bar{y}_{st})} = \left(\frac{r-1}{r+1}\right)^2$$

where $r = n_1/n_2$ as given by Neyman allocation. For the strata in exercise 5.8, case a, what would the fractional increase in variance be by using $n_1 = n_2$ instead of the optimum?

5.10 If the cost function is of the form $C = c_0 + \sum t_h \sqrt{n_h}$, where c_0 and the t_h are known numbers, show that in order to minimize $V(\bar{y}_{st})$ for fixed total cost n_h must be proportional to

$$\left(\frac{W_h^2 S_h^2}{t_h}\right)^{2/3}$$

Find the n_h for a sample of size 1000 under the following conditions.

Stratum	W_h	S_h	t_h
1	0.4	4	1
2	0.3	5	2
3	0.2	6	4

5.11 If $V_{prop}(\bar{y}_{st})$ is the variance of the estimated mean from a stratified random sample of size n with proportional allocation and $V(\bar{y})$ is the variance of the mean of a simple random sample of size n, show that the ratio

$$\frac{V_{prop}(\bar{y}_{st})}{V(\bar{y})}$$

does not depend on the size of sample but that the ratio

$$\frac{V_{min}(\bar{y}_{st})}{V_{prop}(\bar{y}_{st})}$$

decreases as n increases. (This implies that optimum allocation for fixed n becomes more effective in relation to proportional allocation as n increases.) [Use formulas (5.8 and 5.27).]

5.12 Compare the values obtained for $V(p_{st})$ under proportional allocation and optimum allocation for fixed sample size in the following two populations. Each stratum is of equal size. The fpc may be ignored.

Population 1		Population 2	
Stratum	P_h	Stratum	P_h
1	0.1	1	0.01
2	0.5	2	0.05
3	0.9	3	0.10

What general result is illustrated by these two populations?

5.13 Show that in the estimation of proportions the results corresponding to theorem 5.8 are as follows.

$$V_{ran} = V_{prop} + \frac{(1-f)}{n} \sum W_h (P_h - P)^2$$

$$V_{prop} = V_{opt} + \frac{\sum W_h (\sqrt{P_h Q_h} - \overline{\sqrt{P_h Q_h}})^2}{n}$$

where

$$\overline{\sqrt{P_h Q_h}} = \sum W_h \sqrt{P_h Q_h}.$$

5.14 In a firm, 62% of the employees are skilled or unskilled males, 31% are clerical females, and 7% are supervisory. From a sample of 400 employees the firm wishes to estimate the proportion that uses certain recreational facilities. Rough guesses are that the facilities are used by 40 to 50% of the males, 20 to 30% of the females, and 5 to 10% of the supervisors. (*a*) How would you allocate the sample among the three groups? (*b*) If the true proportions of users were 48, 21, and 4%, respectively, what would the s.e. of the estimated proportion p_{st} be? (*c*) What would the s.e. of p be from a simple random sample with $n = 400$?

5.15 Formula (5.27) for the minimum variance of $\bar{y}_{st}$ under Neyman allocation reads as follows.

$$V_{min}(\bar{y}_{st}) = \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N}$$

A student comments: "Since $\sum W_h S_h^2 > (\sum W_h S_h)^2$ unless all the S_h are equal, the formula must be wrong because as n approaches N it will give a negative value for $V(\bar{y}_{st})$." Is the formula or the student wrong?

5.16 By formula (5.26) for Neyman allocation, the sampling fraction in stratum h is $f_h = n_h/N_h = nS_h/N \sum W_j S_j$. The situations in which this formula calls for more than 100% sampling in a stratum ($f_h > 1$) are therefore likely to be those in which the overall sampling fraction n/N is fairly substantial and one stratum has unusually high variability. The following is an example for a small population, with $N = 100$, $n = 40$.

Stratum	N_h	S_h	Optimum n_h
1	60	2	15
2	30	4	15
3	10	15	10
	100		40

(*a*) Verify that the optimum n_h are as shown in the right column. (*b*) Calculate $V(\bar{y}_{st})$ by formula (5.6) and by formula (5.42) and show that both give $V(\bar{y}_{st}) = 0.12$.