University of Waterloo STAT 332 – W. H. Cherry

Assignment 8

A8 - 1. All the tire dealers in a region were assigned to strata according to the number of new tires held at a previous census. Using equiprobable selecting within each stratum, 4,170 dealers were then obtained and their number of new tires ($\mathbf{Y}$) ascertained. The results, based on complete response, are shown in the table at the right.

Stratum boundaries	$\mathbf{N}_h$	n_h	$\bar{y}_h$	s_h
1 - 9	19,850	3,000	4.1	5.9
10 - 19	3,250	600	13.0	9.6
20 - 29	1,007	340	25.0	13.2
30 - 39	606	230	38.2	17.9

(a) Estimate the average number of new tires per dealer, and find the standard deviation of the estimate.

(b) Find an approximate 90% confidence interval for the total number of new tires stocked by all the dealers.

A8 - 2. A corporation wishes to estimate the total number of person-hours lost because of accidents to its employees; because of the different accident rates among different categories of employees, the population of employees was separated into three strata as shown in the table above; the stratum standard deviation (s.d.) estimates stem from an earlier sample survey.

Stratum number (h)	1	2	3
Stratum description	Workers	Technicians	Administrators
Stratum size	132	92	27
Estimated stratum s.d.	6.3	5.3	3.2

(a) Find the approximate Neyman allocation for a sample of size 30, and the corresponding approximation to the standard deviation for the estimate of the total person-hours lost because of accidents among the employees of the corporation.

(b) Suppose that, in practice, the sample shown at the right had been drawn. Use these data to estimate the total number of person-hours lost because of accidents, and find an estimate of the standard deviation of this estimate of the total.

n_h	Stratum sample values (*hours*)
18	8 0 6 7 9 18 24 16 0 4 5 2 0 32 16 4 8 0
10	4 0 8 3 1 5 24 12 2 8
2	1 8

(c) Explain briefly the reason(s) for any differences in the approximate standard deviation in (a) and the estimate of the standard deviation in (b).

A8 - 3. The farms of a large region were assigned to seven strata on the basis of their areas as reported in the latest census, and the yields ($\mathbf{Y}_h$) of the main cash crop were then estimated by means of a sample survey; the data were rounded to convenient numbers and are shown in the table at the right. A future sample survey of this population is being planned, using equiprobable selecting from these strata.

h	$\mathbf{N}_h$	$100\mathbf{W}_h$	$\bar{y}_h$	s_h	s_h^2
1	50,000	50	0.13	0.50	0.25
2	23,000	23	0.72	1.7	2.89
3	20,000	20	3.34	8.5	72.25
4	5,300	5.3	18.03	35	1,225
5	1,500	1.5	68.85	95	9,025
6	120	0.12	786	200	40,000
7	80	0.08	484	170	28,900

(a) If all 200 farms in strata 6 and 7 must be included in the sample, calculate the approximate Neyman allocation for a total sample size of 3,000 farms. Find the corresponding approximation for the standard deviation of $\overline{Y}_{st}$ for this allocation.

(b) Repeat the calculations in (a) but using *proportional* allocation; again include all 200 farms in strata 6 and 7 in the sample. Comment briefly on any differences between the standard deviation values in (a) and (b).

(c) Suppose that the costs per response in the strata are $c_1 = c_2 = \$2$, $c_3 = c_4 = c_5 = \$3$, and $c_6 = c_7 = \$6$. If all the farms in strata 6 and 7 must again be included in the sample, calculate the approximate sample allocation for minimum standard deviation and a total cost of \$10,000. Also, find an approximate value for this minimum standard deviation of $\overline{Y}_{st}$.

A8 - 4. A corporation wishes to obtain information on the effectiveness of a business machine. A number of division heads are to be interviewed by telephone and asked to rate the machine on a numerical scale. The divisions of the corporation are located in North America, Europe and Asia, so that equiprobable selecting is to be used, with the three regions as strata. The table at the right gives, for each stratum, the cost per response, the approximate value anticipated for $\mathbf{S}$ and the size (with $\mathbf{N}$ = 219).

	STRATUM	
1 (North America)	2 (Europe)	3 (Asia)
$c_1 = \$9$	$c_2 = \$25$	$c_3 = \$36$
$\mathbf{S}_1^2 = 1.5^2$	$\mathbf{S}_2^2 = 1.8^2$	$\mathbf{S}_3^2 = 1.8^2$
$\mathbf{N}_1 = 112$	$\mathbf{N}_2 = 68$	$\mathbf{N}_3 = 39$

(a) The corporation wishes to estimate the overall average rating of the machine to within $2\sqrt{0.1}$ units with 95% probability; find the least costly sample size (n) that will achieve this precision, and also find the appropriate allocation.

(b) If the average ratings of the machine for the respective strata for the allocation in (a) are found to be 2.6, 3.1 and 1.8 units, estimate the overall average rating and find its standard deviation. Explain briefly whether the precision of your estimate of the average rating should be expressed as a confidence interval.

(c) If the survey has fixed costs of \$450, find its total cost.

(*continued overleaf*)

A8 - 5. A stratified population has five strata. The stratum sizes ($\mathbf{N}_h$), averages ($\overline{\mathbf{Y}}_h$) and standard deviations ($\mathbf{S}_h$) of response $\mathbf{Y}$ are shown in the table at the right.

Stratum (h)	$\mathbf{N}_h$	$\overline{\mathbf{Y}}_h$	$\mathbf{S}_h$
1	117	7.3	1.14
2	98	6.9	1.43
3	74	11.2	1.06
4	41	9.1	1.40
5	45	9.6	1.32

(a) Calculate the population average ($\overline{\mathbf{Y}}$) and standard deviation ($\mathbf{S}$).

(b) For a *total* sample of size about 80, determine the *stratum* sample sizes under proportional allocation and under Neyman allocation.

(c) For the same total sample size as in (b), find the relative efficiency of the sample average ($\overline{Y}_{st}$) under equiprobable selecting (EPS) with*out* stratification, as an estimator of $\overline{\mathbf{Y}}$, relative to the averages for each of the two allocation schemes in (b) under EPS from the stratified population.

NOTE: In a survey sampling context, *relative efficiency* is usually taken as the *ratio* of the squares of the relevant standard deviations.

A8 - 6. Data for a sample obtained by ERS from a stratified population are shown in the table at the right.

Stratum (h)	$\mathbf{N}_h$	n_h	$\overline{y}_h$	s_h
1	1,000	50	50	20
2	7,000	350	1,000	2
3	2,000	100	20	10

(a) What allocation scheme was used? Explain briefly.

(b) Is stratification likely to give a more precise estimate of $\overline{\mathbf{Y}}$ in this situation than EPS from the *un*stratified population would have done? Explain briefly.

(c) If you were conducting a future sample survey of this population and measuring the same response, how could the allocation be improved? Is the improvement likely to be substantial? Justify your answers briefly.

NOTE: Extensive calculations should *not* be part of answers to this question.

A8 - 7. A stratified population has the characteristics shown in the table at the right, where c_h is the cost per response in stratum h. Assuming *complete* response, find:

Stratum (h)	$\mathbf{N}_h$	c_h	$\overline{\mathbf{Y}}_h$	$\mathbf{S}_h$
1	5,000	1	50	3.2
2	2,500	4	40	2.2
3	2,500	4	60	2.2

(a) the standard deviation of $\overline{Y}_{st}$ and the total cost for 500 units obtained by EPS from these strata using **(i)** proportional allocation; **(ii)** Neyman allocation;

(b) the standard deviation of $\overline{Y}_{st}$ and the total cost expected for 500 units obtained by EPS with*out* stratification.

A8 - 8. A sample survey is to be conducted to estimate the total number of books borrowed from the 217 public libraries in a region during a particular week. The libraries can be classified as small, medium or large in size, on the basis of their holdings of books. The average numbers of books borrowed from libraries in the three groups are thought to be roughly in the ratio 1 : 2 : 3; it is also anticipated that the within-group standard deviations squared of numbers of books borrowed will be proportional to the square root of the corresponding averages. There are 71 small, 126 medium, and 20 large libraries in the region. If a *total* sample size of 40 is required and if the cost per response in each group is the same, how should stratum sample sizes be allocated to the three groups under equiprobable selecting from these three groups taken as strata?

A8 - 9. A sample survey is to be conducted to estimate the total acres of land sown with corn on 1,000 farms in a region. The farms can be classified as small, medium or large on the basis of their sizes. The average corn acreages in the three groups are thought to be roughly in the ratio 1 : 3 : 8; it is also anticipated that the within-group standard deviations of corn acreages will be proportional to the corresponding averages. There are 500 small, 400 medium, and 100 large farms in the region. If a *total* sample size of 60 is required and if the cost per response in each group is the same, how should stratum sample sizes be allocated to the three groups under equiprobable selecting from these three groups taken as strata?

***A8 - 10.** Suppose that the sampling cost for obtaining, by equiprobable selecting from a stratified population, a sample of size n made up of n_h observations from stratum h, is given by the expression at the right; c_h is the cost per response in stratum h and c_o is the fixed costs. If the stratified sample average ($\overline{Y}_{st}$) is to be used to estimate the population average ($\overline{\mathbf{Y}}$), determine the (optimum) allocation of the n_h which minimizes the standard deviation squared (and, hence, the variance) of $\overline{Y}_{st}$ for a fixed total survey cost C, for the cases:

$$C = c_o + \sum_{h=1}^{k} c_h n_h^z$$

(a) $z = 1/2$; (b) $z = 1/3$.

*Optional question for enrichment