REFERENCES

1. Sampling Procedures and Tables for Inspection by Variables for Percent Defective. U.S. Department of Defense, Military Standard 414, 1957.
2. H. Scheffé. The Analysis of Variance. Wiley, New York, 1967.

Stratified Random Sampling from a Discrete Population

RICHARD M. WEED

In the development of statistical acceptance procedures for products whose quantity is measured on a continuous scale by using units such as length, area, volume, or weight, quality-assurance engineers usually specify stratified random sampling plans to ensure a more uniform coverage of the product than is often achieved by pure random sampling. Stratified plans divide the total quantity of the product into an appropriate number of equal-sized sublots and require that a single random sample be taken from each. Not only is it desirable to develop an equivalent procedure for products that are measured in discrete units, but in many cases, such a procedure will prove to be more convenient for continuous products that are produced or delivered in discrete units such as batches or truckloads. However, the development of such a procedure is not as straightforward as might be expected. Weaknesses of some of the more obvious approaches are discussed and then a method is presented that achieves the desired result.

With pure random sampling, all possible sample combinations are equally probable. Although the theory associated with most statistical acceptance procedures is based on the concept of pure random sampling, this approach has the disadvantage that, on occasion, the samples may tend to be clustered within a small segment of the population. In the development of acceptance procedures for products whose quantity is measured in continuous units such as length, area, volume, or weight, it has become common practice to avoid this drawback by specifying stratified random sampling plans. These plans divide the total quantity of the product into an appropriate number of equal-sized sublots and require that a single random sample be taken from each.

Some construction products are measured only in discrete units such as pieces, and others that are measured in continuous units are produced or delivered in discrete units such as batches or truckloads. For both of these cases, it will be desirable to develop a stratified sampling procedure suitable for discrete populations. However, the stratification method described in the preceding paragraph cannot be applied directly unless the sample size happens to be an exact divisor of the lot size. Since this occurs only rarely, a modification of this procedure is required that will spread the samples throughout the entire population in a manner that produces the same degree of randomness as that provided by continuous stratified plans.

Whereas all possible combinations of individual samples may occur with pure random sampling, this obviously is not the case with stratified sampling since only one portion of the population is selected from each subgroup. However, computation of the probability of any particular portion being included in the sample is not difficult, and it can be shown that this probability is equal for all portions. It follows that the degree of randomness achieved by stratified random sampling is such that each item of the population has an equal chance of appearing in the sample.

This is a necessary but insufficient condition for pure random sampling and emphasizes that stratified random sampling produces a more restricted degree of randomness. Since the theory associated with statistical acceptance procedures is based primarily on pure random sampling, one might wonder about the extent to which the validity of these procedures is compromised by stratified sampling. By their silence on this subject, most authors have implied that there is no serious problem. Based on a few brief tests with computer simulation, this appears to be a correct assumption, although this is an area that might warrant further study. For purposes of this paper, however, assume that stratified sampling is a valid and practical approach, and attention will now be directed toward the development of a method for selecting stratified random samples from discrete populations.

UNSATISFACTORY METHODS

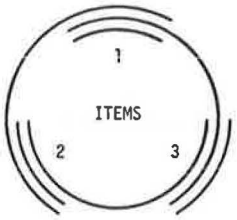
The objectives of the method to be developed are to guarantee that the samples will be distributed throughout the entire population and to do this in a manner that produces the same degree of randomness as that provided by continuous stratified plans. It is a simple matter to accomplish the first objective, but care must be exercised to ensure that the second objective is achieved. In several of the more obvious approaches, the probability of being included in the sample is not equal for all items of the population.

One method that produces an imperfect result consists of stratification by quantity, selection of the sample location by quantity, determination of the discrete batch or load within which this random location occurs, and then random sampling from that batch or load. For example, if a construction material is normally measured in tons, a lot could be defined as 1000 tons, each lot could be divided into five sublots of 200 tons each, and specific tonnage values would designate the random sampling locations within each sublot. The discrete sampling locations would then be the particular trucks within which these random tonnage values occur. Although this method works reasonably well when the total number of trucks represented by each sublot is large, it has a minor flaw that can become pronounced when the number of trucks is small. If the random sampling locations for two successive sublots both fall close to the boundary between these two sublots, they may both occur within the same truckload. When this happens, the theoretically correct approach is to take two samples from the same truck. However, from a practical standpoint, it is usually considered to be

Figure 1. A possible stratified sampling scheme.

TRUCK:	1	(2)	3	4	5	(6)	7	8	(9)	10	11	12	13	(14)	15	16	(17)	18
SUBGROUPING:																		
SUBGROUP SIZE:		3				4				4				3				4

Figure 2. Possible subgroup arrangements when circular array concept is used.



more useful to sample two successive trucks or to make some other similar adjustment. Either way, this distorts the randomness because not all trucks in the population are equally likely to experience this effect. This distortion increases as the number of trucks within each subgroup decreases and, in some cases, can become quite severe.

The method to be described next includes a useful procedure for stratifying a discrete population but, because of the steps that follow, the desired degree of randomness is not achieved. Since the sample size usually will not be an exact divisor of the population size, the best that can be done is to divide the population into subgroups of two sizes that differ by one unit. This is accomplished by the following equations:

$$S_1 = [N_P/N_S] \quad (1)$$

$$S_2 = S_1 + 1 \quad (2)$$

$$N_1 = N_S S_2 - N_P \quad (3)$$

$$N_2 = N_S - N_1 \quad (4)$$

where

N_P = population size,
N_S = sample size,
S_1 = size of smaller subgroup,
S_2 = size of larger subgroup,
N_1 = number of smaller subgroups,
N_2 = number of larger subgroups, and
[X] = largest integer in X.

Once these computations have been made, Equation 5 can be used to check that they have been performed properly:

$$N_1 S_1 + N_2 S_2 = N_P \quad (5)$$

For example, suppose the population consists of $N_P = 18$ trucks of which $N_S = 5$ are to be sampled. Equations 1-5 can be used to develop a stratification plan as follows:

$$S_1 = [N_P/N_S] = [18/5] = [3.6] = 3 \quad (6)$$

$$S_2 = S_1 + 1 = 3 + 1 = 4 \quad (7)$$

$$N_1 = N_S S_2 - N_P = (5)(4) - 18 = 2 \quad (8)$$

$$N_2 = N_S - N_1 = 5 - 2 = 3 \quad (9)$$

$$N_1 S_1 + N_2 S_2 = (2)(3) + (3)(4) = 18 = N_P \quad (10)$$

Once the numbers (N_1, N_2) and sizes (S_1, S_2) of the subgroups have been determined, the subgroups are arranged in random order. Then, to determine the items to be sampled, a random selection within each subgroup is made. For the case in which $N_P = 18$ and $N_S = 5$, one possible outcome of this procedure is shown in the schematic diagram in Figure 1, in which the horizontal lines define the separate subgroups and the circled numbers are the trucks that have been randomly selected for sampling.

To demonstrate that this is a satisfactory approach, it would be necessary to prove that, for any combination of values of N_P and N_S, each item in the population has an equal chance of being included in the sample. Conversely, to disprove this method, it is only necessary to show by counterexample that some particular combination of N_P and N_S produces an unsatisfactory result. This is a problem in combinatorial analysis that leads to very complex calculations except for those cases in which the sample size is only slightly smaller than the population size. Consequently, the following two cases have been selected to demonstrate that not all of the items in the population have an equal chance of being included in the sample:

	Probability	
Item	Case 1, $N_P = 7$, $N_S = 6$	Case 2, $N_P = 8$, $N_S = 6$
1	0.917	0.833
2	0.833	0.700
3	0.833	0.734
4	0.833	0.734
5	0.833	0.734
6	0.833	0.734
7	0.917	0.700
8		0.833
Total	5.999	6.002

Several interesting observations can be made from these computations. First, the sum of the probabilities equals the sample size, which is the mathematical expectation of this procedure. Second, there is a distinct departure from equal probability and an apparent tendency for the first and last items of the population to have a greater likelihood of being included in the sample. Finally, the departure from equal probability increases as the population size (N_P) increases from seven to eight, which suggests that this is a problem that will not diminish rapidly for larger populations. Subsequent tests by computer simulation indicate that this tendency persists even for much larger population sizes.

It should be emphasized that this problem is not the result of the stratification method given by Equations 1-4 but, rather, was caused by the manner in which the subgroups were randomly distributed throughout the population. The next section shows that this problem can be overcome by a simple refinement of this procedure.

DEVELOPMENT OF A SATISFACTORY METHOD

Although the subgroups were arranged in random order in the method that was just discussed, this produces different conditions for items in different positions in the population. This is best

illustrated by a simple example for which $N_P = 3$ and $N_S = 2$, as shown in the following schematic diagram:

	Items:	1	2	3
Possible subgroup arrangements:				

In this trivial example, only two possible subgroup arrangements exist; both are equally likely (P = 0.5) because they are selected at random. If an item happens to fall in the subgroup of size n = 1, it is certain that it will be selected as the sample from that subgroup and the probability of this event is P = 1.0. Similarly, if the item happens to fall in the subgroup of size n = 2, its probability of being selected is P = 0.5. By using routine probability theory, each item's probability of being included in the sample can be calculated as follows:

$$P_1 = (0.5)(1.0) + (0.5)(0.5) = 0.75 \quad (11)$$

$$P_2 = (0.0)(1.0) + (1.0)(0.5) = 0.50 \quad (12)$$

$$P_3 = (0.5)(1.0) + (0.5)(0.5) = 0.75 \quad (13)$$

Total 2.00

As before, the sum of the probabilities is equal to the sample size. The second item has a lower probability of being selected than does the first or third item because it can occur only in a subgroup of size n = 2. Similar results occur with larger populations and samples although the probability computations become much more complex.

Only a very slight conceptual change is required to correct this condition. If the items are thought of as being arranged in a circular array, the end effects are avoided and the individual probabilities will be equal. This is demonstrated by the diagram in Figure 2.

Now there are three possible arrangements of subgroups, and each item occurs once in a subgroup of size n = 1 and twice in a subgroup of size n = 2. The probabilities are equal and compute to be

$$P_1 = P_2 = P_3 = (0.333)(1.0) + (0.667)(0.5) = 0.667 \quad (14)$$

and, as a check, it is seen once again that their sum is equal to the sample size.

Figure 3. Sampling scheme resulting from the circular array concept.

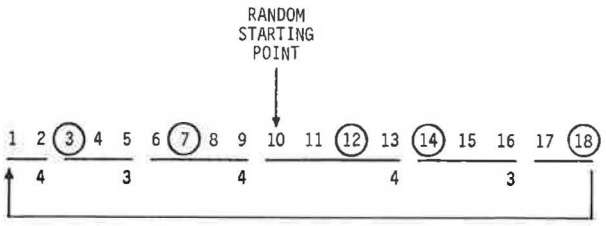


Figure 4. Typical work sheet outlining steps of discrete random selection procedure.

STRATIFICATION PROCEDURE

POPULATION SIZE = N = 23 (MAXIMUM 50)

SAMPLE SIZE = 6 (SAMPLE SIZE = N WHEN N < 6)

SIZE OF SMALLER SUBGROUP = $S_1 = \left[\frac{N}{6}\right] = \left[\frac{23}{6}\right] = 3$ (DISCARDING FRACTIONAL REMAINDER)

SIZE OF LARGER SUBGROUP = $S_2 = S_1 + 1 = 4$

NUMBER OF SMALLER SUBGROUPS = $N_1 = (6)(S_2) - N = (6)(4) - 23 = 1$

NUMBER OF LARGER SUBGROUPS = $N_2 = 6 - N_1 = 5$

CHECK THAT $(N_1)(S_1) + (N_2)(S_2) = N$: (1)(3) + (5)(4) = 23

ARRANGE SUBGROUPS IN ANY ORDER: 3, 4, 4, 4, 4, 4.

RANDOM STARTING POINT (TABLE R1) = 11

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 ~~24 25~~
4 4 4 3 4 4
~~26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50~~

RANDOM SELECTION PROCEDURE

LIST OF ITEMS IN SUBGROUP	RANDOM SELECTION (TABLE R2)	ITEM NUMBER
11, 12, 13	2	12
14, 15, 16, 17	3	16
18, 19, 20, 21	2	19
22, 23, 1, 2	4	2
3, 4, 5, 6	2	4
7, 8, 9, 10	3	9

ORDERED LIST OF ITEMS TO BE SAMPLED: 2, 4, 9, 12, 16, 19.

Figure 5. Special random number tables for use with discrete random selection procedure.

TABLE R1

21	47	18	47	37	10	44	6	37	21	29	25	15	20	37	6	18	43	10	6	1	31	37	47	37	43	25	46	35	21
11	46	10	35	9	24	50	13	10	9	3	14	2	15	43	45	2	31	11	44	30	43	2	46	16	5	15	28	15	6
12	20	49	7	6	20	42	34	11	42	39	32	32	21	27	47	3	27	2	12	20	50	37	45	34	33	36	14	36	35
8	49	2	9	48	25	14	33	48	15	22	7	18	25	41	13	7	35	15	34	43	32	5	49	18	39	2	46	45	3
45	38	7	28	39	5	34	5	6	38	19	12	22	43	9	33	40	10	37	2	50	40	36	8	43	32	42	32	26	1
40	27	33	45	15	4	31	29	48	38	26	30	3	1	41	41	33	31	6	20	30	44	23	4	20	42	43	40	47	9
41	14	33	30	16	35	42	16	5	5	3	11	9	29	17	48	37	4	23	25	44	24	17	32	38	37	42	46	46	21
39	23	34	22	44	19	5	9	4	15	50	15	2	28	8	24	36	32	2	8	7	46	45	1	10	12	38	3	27	8
13	49	20	43	49	14	26	28	17	44	26	7	14	47	19	36	42	26	26	49	29	15	50	30	49	20	7	6	11	3
27	19	14	1	17	39	45	27	34	13	32	23	42	11	26	50	5	46	6	14	11	26	10	48	41	22	13	4	19	33
9	48	41	16	23	8	23	40	36	13	18	24	33	9	17	47	11	24	16	28	5	47	18	20	10	12	37	17	10	13
25	13	35	16	25	47	11	48	23	46	21	28	48	19	22	18	7	48	22	4	41	41	35	1	18	49	22	35	36	37
50	43	33	15	24	4	23	25	23	19	1	31	30	35	1	34	35	8	34	16	40	49	30	17	44	31	30	50	50	2
41	23	19	4	7	40	36	11	28	26	5	29	25	49	35	38	29	29	38	4	27	28	24	44	38	1	44	9	39	12
36	13	40	42	19	8	21	4	35	44	1	29	41	3	40	17	27	1	14	36	17	45	6	23	15	13	39	48	19	46
8	2	34	41	41	6	30	34	3	9	1	45	12	7	10	31	5	22	42	28	22	40	24	26	13	44	38	40	27	16
22	42	11	20	42	26	18	12	8	5	37	27	18	9	48	3	28	25	17	45	12	12	31	44	6	28	43	25	32	50
46	4	27	24	16	29	18	30	32	16	2	38	10	27	39	33	29	17	21	19	19	21	46	39	20	47	36	23	39	40
3	8	13	22	29	22	32	24	34	20	38	45	21	34	12	29	33	17	14	30	39	30	24	31	14	49	25	8	43	31
12	50	18	14	7	28	7	10	24	32	38	11	21	36	26	48	47	21	49	39	33	4	16	50	47	3	16	31	45	31

TABLE R2

7	8	3	7	8	3	6	9	6	2	4	5	8	3	8	9	1	2	4	1	4	1	3	7	1	8	5	4	3	2	7	4	9	7	5	7	7	1	2	9	6	3	1	9	1
8	8	4	8	9	4	2	5	2	9	8	9	2	2	5	9	4	3	6	8	1	9	7	8	7	8	8	1	9	2	8	6	9	7	3	2	3	5	8	7	9	8	6	4	5
5	4	9	9	4	3	8	5	4	6	5	9	8	2	3	9	8	3	4	6	2	2	1	7	6	1	6	4	2	9	4	1	7	9	9	2	2	7	6	2	7	9	8	6	1
6	8	5	2	6	1	2	1	4	5	6	1	1	6	5	3	3	5	8	7	3	1	9	8	6	7	7	9	7	4	2	4	6	7	7	5	6	1	2	7	3	3	2	3	4
5	7	4	1	6	8	3	6	2	7	6	9	3	6	6	9	2	5	1	4	7	7	1	9	2	6	7	1	6	9	4	5	8	2	1	9	9	4	9	9	7	6	3	7	4	2
2	6	8	9	7	3	8	8	3	5	3	6	6	9	4	9	9	1	1	6	1	1	6	8	8	4	6	8	1	3	1	1	7	4	8	5	9	4	5	7	2	3	2	1	4
7	4	6	8	1	5	7	6	5	1	5	8	3	9	4	4	8	5	3	2	4	1	2	6	7	7	3	6	9	2	3	6	9	2	8	3	9	9	2	3	7	9	4	8	7
2	8	5	7	4	3	9	8	7	2	3	7	5	8	7	8	8	2	3	3	4	3	8	4	2	9	8	9	5	2	5	4	6	6	6	8	9	2	7	3	9	6	9	4	4
2	7	9	6	1	6	5	6	5	2	3	8	5	8	7	5	1	5	8	1	4	8	1	7	4	3	1	4	8	8	6	8	1	1	5	1	4	3	1	2	6	4	2	9	7
4	3	6	2	9	3	9	5	1	7	1	8	3	7	6	5	6	3	7	9	7	2	2	3	5	4	6	7	6	4	9	7	3	7	8	6	1	4	7	3	6	6	2	9	9
8	5	9	6	3	4	8	9	1	1	5	4	1	8	7	9	3	6	9	6	6	4	9	3	7	7	9	7	2	2	3	8	3	7	5	4	1	7	5	3	3	5	9	2	6
5	8	8	2	6	4	9	1	4	8	9	5	2	2	6	9	2	3	2	1	4	7	8	2	4	1	4	5	5	2	4	5	6	9	5	8	2	7	9	9	2	5	7	5	4
9	1	2	1	7	4	1	6	4	4	7	6	3	4	5	3	9	1	7	5	1	7	2	6	4	8	9	1	6	6	1	4	7	6	2	7	6	2	4	8	9	1	5	9	8
1	3	7	7	7	2	5	2	6	1	4	8	2	2	2	3	3	2	4	7	4	8	2	8	2	6	1	2	9	4	3	1	4	5	3	6	3	2	8	6	8	3	5	4	7
7	3	6	5	4	5	2	9	6	5	7	3	6	1	4	1	9	9	8	7	1	7	9	3	3	5	5	1	5	8	7	3	9	6	6	7	1	7	5	9	6	9	3	1	7
9	3	3	6	6	7	4	5	6	7	9	4	4	8	8	1	9	2	4	8	8	2	1	9	5	9	1	3	2	3	7	8	3	8	5	1	3	4	1	6	8	2	6	7	1
3	3	2	3	6	5	9	2	2	2	7	4	1	8	4	4	8	8	9	2	9	9	2	8	1	8	9	2	6	1	3	4	3	7	3	1	7	3	5	6	5	6	7	8	1
1	5	3	7	7	4	1	8	4	6	2	7	2	9	6	3	7	9	8	9	4	5	8	1	4	2	2	4	5	2	1	5	1	3	5	1	3	3	4	9	3	6	6	1	4
6	5	3	6	4	8	3	7	5	2	8	8	8	9	9	2	6	2	9	3	8	5	4	3	8	8	4	8	2	4	9	9	6	9	5	5	5	4	8	7	7	4	2	9	2
8	8	9	1	1	8	6	8	8	4	2	5	5	9	6	9	5	1	3	9	1	7	3	8	8	9	4	9	1	9	9	2	3	4	6	3	1	3	8	2	8	3	7	1	5
5	9	5	5	8	9	4	8	3	5	3	4	3	6	6	6	7	5	1	2	3	8	5	6	8	4	2	2	3	4	6	3	5	7	5	5	3	3	1	5	3	7	3	9	6
9	4	6	3	7	9	8	9	4	8	5	3	2	4	8	2	2	5	5	1	5	9	1	4	8	3	2	4	1	5	2	2	5	6	8	5	2	8	6	5	1	2	9	9	3
2	2	2	4	7	4	4	6	7	7	1	7	7	7	3	2	2	7	2	4	1	7	2	4	3	6	9	1	7	3	7	3	3	7	5	4	1	8	4	5	8	5	3	5	8
5	6	7	5	6	6	7	1	6	1	9	3	6	8	7	7	6	2	6	3	5	5	9	8	2	3	2	7	7	9	4	2	6	7	6	9	3	1	2	5	8	9	9	2	8
7	5	8	8	1	9	1	3	1	4	2	2	3	5	7	3	5	7	5	4	3	3	1	2	2	4	7	4	1	1	7	1	2	5	7	8	3	2	4	4	8	7	3	7	1
5	6	7	6	5	8	8	7	5	8	1	9	1	4	4	8	6	5	3	5	2	7	6	8	1	4	9	7	3	6	8	6	5	1	5	8	5	5	8	2	8	7	1	9	7
4	8	5	7	1	6	7	9	6	9	3	3	5	7	9	3	7	2	9	2	3	5	1	4	6	6	5	8	3	3	8	5	3	5	3	1	1	5	4	7	5	4	2	7	7
5	2	4	1	7	1	6	4	9	4	2	8	5	7	6	9	4	7	4	7	9	6	8	7	6	8	7	6	9	9	3	9	1	1	1	7	9	5	6	5	2	3	5	1	9
7	3	3	4	5	8	1	7	6	2	7	3	2	6	6	4	3	6	8	2	8	3	3	1	3	2	1	2	4	4	6	4	2	1	8	2	4	8	9	1	8	6	4	2	9
3	4	8	3	6	4	8	8	5	2	7	2	3	9	1	1	9	9	4	8	6	8	3	6	4	5	7	6	8	2	5	7	5	4	3	7	9	6	3	5	1	1	5	5	5
4	6	1	9	5	6	6	9	3	5	9	4	2	3	8	8	7	2	7	4	1	5	1	5	2	2	1	5	6	5	4	5	5	2	1	8	2	2	4	8	1	8	4	6	3
9	1	4	6	2	5	5	8	2	3	6	7	6	7	4	7	2	9	1	2	7	4	8	4	5	4	6	2	1	9	1	4	9	3	6	8	5	1	2	8	5	3	9	4	3
1	1	6	9	2	4	7	2	9	3	8	5	5	9	2	1	6	2	1	5	4	4	5	3	4	6	4	7	6	7	4	6	9	8	6	7	7	8	3	8	9	3	6	6	1
6	4	9	4	8	7	6	3	5	9	3	6	5	4	6	8	6	7	1	9	4	6	5	2	2	4	6	7	8	4	8	5	9	1	9	7	6	2	3	1	9	9	1	7	1
8	8	9	1	5	9	6	3	1	8	9	8	5	6	9	4	9	2	4	3	4	1	3	8	6	7	8	6	4	4	4	7	2	4	5	1	8	5	6	9	8	9	5	8	3
7	7	5	2	1	7	7	5	7	8	5	4	8	8	2	3	6	6	8	2	6	5	9	8	2	4	5	1	5	6	1	2	5	9	2	6	9	8	6	9	5	1	5	2	4
9	6	3	9	8	9	3	6	1	6	2	5	4	7	1	9	2	7	3	4	2	6	5	1	2	3	2	1	3	6	3	2	4	1	8	7	8	8	7	4	7	1	7	2	7
7	6	4	9	9	9	3	7	6	7	1	8	2	1	6	2	9	4	3	8	6	9	7	3	4	1	9	4	7	6	4	4	2	2	6	6	9	8	4	2	4	6	2	1	1
1	6	7	6	8	9	9	2	6	9	5	5	5	9	3	4	5	3	2	3	7	3	4	2	3	4	3	9	1	6	3	3	7	8	1	8	1	5	2	1	7	9	6	9	3	1
4	2	7	6	2	9	5	1	4	2	5	3	7	6	8	2	2	9	5	9	2	4	8	2	3	3	9	6	6	4	5	4	1	6	5	3	1	4	6	6	1	7	9	1	1
3	4	1	2	1	1	2	3	9	8	5	4	3	5	4	1	4	1	9	2	4	1	9	5	8	1	3	5	1	6	9	9	8	3	7	8	4	9	1	4	1	7	3	3	2
4	4	1	8	2	3	8	7	4	6	8	5	6	1	8	7	3	8	7	3	1	9	7	6	4	2	4	8	5	3	1	5	7	2	6	6	5	9	5	4	3	4	2	4	5
2	5	1	9	8	7	9	1	7	7	7	9	1	8	6	7	8	7	1	7	2	9	3	8	4	9	5	5	3	1	4	1	6	4	3	2	3	8	5	8	1	9	4	8	3
5	6	1	9	5	9	2	6	1	9	5	7	3	3	7	3	1	5	1	4	1	7	4	3	2	1	2	8	2	3	7	3	4	5	8	3	3	6	1	2	5	4	9	1	8
9	7	9	8	4	6	5	6	2	7	6	7	7	2	7	1	4	4	1	7	4	2	9	3	9	4	7	9	5	3	5	2	1	8	6	5	3	5	5	7	9	3	2	3	8
7	2	3	9	1	7	9	5	5	7	5	6	8	5	5	5	7	3	1	2	4	3	8	5	8	6	6	2	3	9	7	8	7	5	8	8	1	8	9	5	2	1	5	2	5
1	6	2	9	3	2	2	5	8	6	9	7	1	7	4	5	2	9	5	6	4	6	8	1	2	6	1	3	6	3	6	1	8	7	8	8	3	8	5	4	9	7	7	9	6	2

A somewhat simpler procedure can be devised that will produce the same result and be much easier to implement. As long as the array is considered to be circular, any arbitrary arrangement of the subgroups will be satisfactory provided that the starting point is selected at random. To see that this is so, consider the earlier example of selecting 5 out of 18 trucks. Since the concept of the circular array simply allows a subgroup to extend from the tail end to the front end of the array, it will be more convenient to list the array in the usual linear fashion and arrange the subgroups accordingly. Suppose the $N_1 = 2$ subgroups of size $S_1 = 3$ and the $N_2 = 3$ subgroups of size $S_2 = 4$ have been arbitrarily arranged in the order 4, 3, 4, 3, 4 and that item 10 has been determined to be the random starting point. The diagram in Figure 3 illustrates a sampling scheme that might result when the circular array concept is used.

To show that this procedure produces the desired degree of randomness, it is necessary to prove that every item in the population has an equal probability of being included in the sample. Although the particular random selections within the subgroups have not yet been made at the time the random starting point is chosen, it is known that five such selections will be made. Regardless of the actual combination that is ultimately chosen,

only 5 of the 18 possible starting points will result in any particular item being included in the sample. For example, with the selections shown in Figure 3, random starting points of 1, 4, 8, 10, and 15 result in item 12 being included in the sample; however, all other starting points exclude it. Since all 18 starting points are equally likely, the probability that item 12 will be included in the sample is 5/18. Similarly, this same probability holds for all other items in the population, and this result can be generalized to apply to any size of population and sample.

This result greatly simplifies the implementation of this procedure, since only a single random starting point is required in place of a random arrangement of several subgroups. It is still necessary to make a random selection within each subgroup but, with the aid of special random number tables, this method is extremely easy to apply. Figure 4 illustrates a typical work sheet that was used to select a stratified random sample of size $N_S = 6$ from a population size of $N_P = 23$; Figure 5 shows the special random number tables used with this procedure. The user obtains the starting point for the stratification arrangement by entering Table R1 at a random location and then moving in any predetermined direction until a number less than or equal to the population size is obtained. After underlining the subgroups on the work sheet, the user then enters Table R2 and, again moving in any predetermined direction, obtains a total of $N_S = 6$ numbers that are less than or equal to the respective subgroup sizes. The process is completed by converting these to actual item numbers as shown at the bottom of the work sheet. For convenience, the outline of the procedure and the special random number tables can be printed back to back on single sheets of paper. In this way, the documentation for the random selection process for each lot will be contained on a single piece of paper.

GENERATION OF SPECIAL RANDOM NUMBER TABLES

Although standard random number tables can be used for the sampling procedure just described, it is preferable to generate special tables such as those shown in Figure 5. For this particular application, in which the maximum population size is $N_P = 50$ and the sample size is specified to be $N_S = 6$, the largest numbers required in Tables R1 and R2 are 50 and 9, respectively. To generate tables of this type by computer, a one-dimensional array is first filled with equal quantities of all numbers from one up to the largest number that is to appear in the table. These values are then shuffled into random order by using a uniform random number generator (1,2) and a suitable shuffling algorithm (3, p. 125). Because each number appears with equal frequency but in random order, the table can be used repeatedly without the introduction of bias. This is not necessarily true for all random number tables that have been published although, for practical purposes, any bias of this type that might occur is so small that it would be of little consequence.

One other consideration regarding the use of these tables should be mentioned. For the selections to be strictly independent, a new random entering point should be chosen for each selection that is to be made. However, in the example illustrated in Figure 5, it will be observed that six selections were made from Table R2 by using only one random entry point. This is a practical expedient and is justified by the large size of this table. Since each digit appears a total of 235 times in Table R2, the selection of any particular digit has almost no effect on the probability of obtaining the same digit again on a subsequent selection. Table R1 has been designed to be smaller because only one selection at a time is required from this table.

SUMMARY AND CLOSING REMARKS

Stratified random sampling has gained wide acceptance as a practical method for sampling products whose quantity is measured in continuous units of various types. This approach can be equally useful for products that are measured in discrete units as well as for continuous products that are produced or delivered in discrete units. It was demonstrated that some of the potential methods for applying stratified sampling to discrete populations do not produce the desired degree of randomness, but this problem can be overcome with a minor refinement. A satisfactory method was then developed that, with the aid of a work sheet and special random number tables, is extremely easy to apply.

REFERENCES

1. R.E. Shannon. Systems Simulation, the Art and Science. Prentice-Hall, Englewood Cliffs, NJ, 1975.
2. R.M. Weed. An Introduction to Computer Simulation. Federal Highway Administration, 1976.
3. D.E. Knuth. The Art of Computer Programming. Addison-Wesley, Reading, MA, Vol. 2, 1969.