Method to Exclude the Effect of Testing Error When Estimating the Percentage Defective of a Continuous Normal Population

RICHARD M. WEED AND WILLIAM E. STRAWDERMAN

The quality of a product is often characterized by the percentage of the population that falls outside specific limits. Although established methods for estimating this percentage defective are accurate as far as the overall distribution of test results is concerned, part of the variability of this distribution is due to the presence of testing error that causes the percentage defective of the product itself to be overestimated. A method is developed to overcome this problem and computer simulation is used to demonstrate that it is effective for situations in which the testing error is no larger than about one-half of the variability associated with the product. The results of several unsuccessful attempts to improve on this technique are also presented and described briefly.

Engineers and specification writers have found the concept of percentage defective (percentage of the total population outside specification limits) to be a particularly effective way to describe the quality of a variety of construction materials and products. The overall proportion within specification limits is believed to be strongly related to a product's performance, service life, or a combination of these qualities.

To implement a specification that uses this concept, an acceptable quality level (AQL), an acceptable level of percentage defective, is defined along with a sampling procedure and a means for estimating the percentage defective for prescribed quantities of product. A graduated pay schedule is usually employed to establish the appropriate reduction in payment when the production lot is found to be of less than AQL.

Established procedures for estimating percentage defective are effective in that they provide unbiased estimates of the quality of the populations to which they are applied. However, because any test result is affected by both product and testing variablity, the population to which the estimation procedure is applied is somewhat more dispersed and, consequently, its apparent percentage defective is somewhat larger than the true product percentage defective. This is illustrated in Figure 1.

In some cases, this effect is of little concern. If specification writers have defined an AQL that is based on historical data that included testing error, and the present testing error has not changed appreciably, there is no real need to modify the acceptance procedure. It will continue to accept product of the same quality considered acceptable in the past. Also, if the testing error is relatively small in comparison to product variability, or if the sampling procedure is such that the effect of the testing error is reduced by averaging several replicate tests together, then the estimate of percentage defective will be virtually unaffected. However, in other cases a method may be desired to estimate the true product percentage defective exclusive of the effects of testing error. It is toward this end that the efforts of this paper are directed.

DEVELOPMENT OF THE METHOD

In the discussion that follows, the total sample from any given quantity of product that is to be evaluated consists of N random samples, each of size n. This assumes that the samples are taken from N different portions of product and that, within each portion, the n individual tests are subject only to testing error. A test result is defined as the mean of n tests and, therefore, the estimate of the percentage defective is based on N test results.

The standard deviation method of Military Standard 414 (1) is generally recognized as the best method for estimating the percentage defective of a normal population. A quality index (Q) is calculated by Equation 1 and special tables are consulted to convert this to a percentage defective estimate. Although the number of samples (N) does not enter into the computation of Q, it is accounted for when entering the percentage defective tables. A typical table is shown in Figure 2.

$$Q = (\bar{X} - L)/S \quad (1)$$

where

- Q = quality index (for lower limit in this example),
- L = lower limit,
- $\bar{X}$ = sample mean (more specifically, an unbiased estimate of the population mean), and
- S = sample standard deviation (more specifically, the square root of the unbiased estimate of the population variance).

A careful look at the definitions of $\bar{X}$ and S in Equation 1 makes it possible to deduce what is required to obtain an estimate of the true product percentage defective. First, in place of $\bar{X}$, an unbiased estimate of the mean of the product distribution will be required. This is an easy matter since, assuming there is no testing bias, the sample mean is also an unbiased estimate of the product mean. Therefore, this term will remain unchanged in the method to be developed.

Second, an unbiased estimate of the variance of the product distribution is required. By using the well-known theorem that independent variances are additive and then transposing, Equation 2 can be written:

$$\sigma_p^2 = \sigma_N^2 - (\sigma_n^2/n) \quad (2)$$

where

- σ_p^2 = variance of the product population,
- σ_N^2 = variance of the population of sample means (so designated because there are N sample means per lot),
- σ_n^2 = variance associated with the n replicate tests for each of the N samples (i.e., the testing error), and
- n = number of replicate tests (i.e., size of each of the N samples).

For each of the N samples, n replicate tests are made. If the variance is calculated for each set of n values, this furnishes N independent unbiased estimates of σ_n^2. These are pooled to give a single un-

Figure 1. Illustration of effect of testing error on percentage defective.

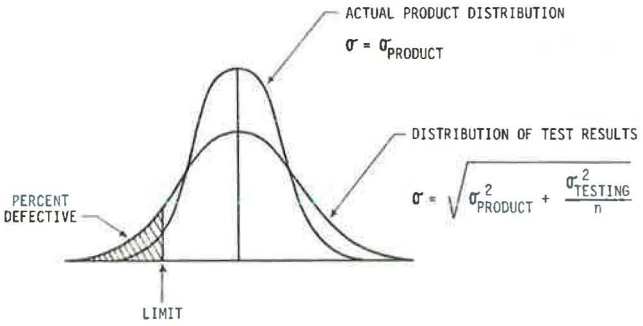

Table 1. Sample calculations required for the modified method.

	Test			Sample Statistics	
Sample	1	2	3	Mean	SD
1	103	100	103	102.0	1.732
2	101	103	106	103.3	2.517
3	102	98	99	99.7	2.082
4	105	107	104	105.3	1.528
5	106	105	109	106.7	2.082
6	103	100	98	100.3	2.517
7	105	103	106	103.7	1.528

Note: $\bar{X}$ of the means = 103.0, S_N of the means = 2.541, and pooled S_n for the standard deviations = 2.036.

Figure 2. Typical table used to estimate the percentage defective of a normal population: standard deviation method (N = 7).

Q	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	50.00	49.63	49.25	48.88	48.50	48.13	47.75	47.38	47.01	46.63
0.1	46.26	45.89	45.51	45.14	44.77	44.40	44.03	43.65	43.28	42.91
0.2	42.54	42.17	41.80	41.44	41.07	40.70	40.33	39.97	39.60	39.23
0.3	38.87	38.50	38.14	37.78	37.42	37.06	36.69	36.33	35.98	35.62
0.4	35.26	34.90	34.55	34.19	33.84	33.49	33.13	32.78	32.43	32.08
0.5	31.74	31.39	31.04	30.70	30.36	30.01	29.67	29.33	28.99	28.66
0.6	28.32	27.98	27.65	27.32	26.99	26.66	26.33	26.00	25.68	25.35
0.7	25.03	24.71	24.39	24.07	23.75	23.44	23.12	22.81	22.50	22.19
0.8	21.88	21.58	21.27	20.97	20.67	20.37	20.07	19.78	19.48	19.19
0.9	18.90	18.61	18.33	18.04	17.76	17.48	17.20	16.92	16.65	16.37
1.0	16.10	15.83	15.56	15.30	15.03	14.77	14.51	14.26	14.00	13.75
1.1	13.49	13.25	13.00	12.75	12.51	12.27	12.03	11.80	11.56	11.33
1.2	11.10	10.87	10.65	10.42	10.20	9.98	9.77	9.55	9.34	9.13
1.3	8.93	8.72	8.52	8.32	8.12	7.93	7.73	7.54	7.35	7.17
1.4	6.98	6.80	6.62	6.45	6.27	6.10	5.93	5.77	5.60	5.44
1.5	5.28	5.13	4.97	4.82	4.67	4.52	4.38	4.24	4.10	3.96
1.6	3.83	3.70	3.57	3.44	3.31	3.19	3.07	2.96	2.84	2.73
1.7	2.62	2.51	2.41	2.30	2.20	2.11	2.01	1.92	1.83	1.74
1.8	1.66	1.57	1.49	1.41	1.34	1.26	1.19	1.12	1.06	0.99
1.9	0.93	0.87	0.81	0.76	0.70	0.65	0.61	0.56	0.51	0.47
2.0	0.43	0.39	0.36	0.32	0.29	0.26	0.23	0.21	0.18	0.16
2.1	0.14	0.12	0.10	0.08	0.07	0.06	0.05	0.04	0.03	0.02
2.2	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.0	0.0	0.0

biased estimate, which will be designated S_n^2. Equation 3 can then be written in terms of sample standard deviations to obtain the value being sought, the standard deviation of the product distribution, which is the square root of the unbiased estimate of the product variance.

$$S_P = \sqrt{S_N^2 - (S_n^2/n)} \quad (3)$$

where

S_P = estimated standard deviation of the product population,
S_N = computed standard deviation of the N sample means,
S_n = pooled estimate of testing error, and
n = size of each of the N samples.

The next step is to substitute these results into Equation 1 to develop a modified Q-statistic for use in estimating the percentage defective of the product distribution only. This modified value, designated Q' to distinguish it from the standard Q value, is given by Equation 4.

$$Q' = (\bar{X} - L)/S_P = (\bar{X} - L)/\sqrt{S_N^2 - (S_n^2/n)} \quad (4)$$

Sample Calculation

For this example, there are N = 7 samples, each of size n = 3, and the lower limit is 100.0. The test results and calculations given in Table 1 are substituted into the appropriate equations to give the following results.

By using Equation 3 and the results from Table 1,

$$S_P = \sqrt{2.541^2 - (2.036^2/3)} = 2.253.$$

By using Equation 4 and the results from Table 1,

$$Q' = (103.0 - 100.0)/2.253 = 1.33.$$

From Figure 2, the estimated percentage defective by the modified method = 8.32.

By using Equation 1,

$$Q = (103.0 - 100.0)/2.541 = 1.18.$$

From Figure 2, the estimated percentage defective by the standard method = 11.56.

Although the difference in the results obtained by the two methods is not great in this example, there are situations in which this difference may be important. If the AQL were defined to be 10 percent, the modified method would produce a clearly acceptable result, but the standard approach would not. If a graduated pay schedule were in effect, a difference of a few percent might correspond to a substantial reduction in payment.

A Minor Problem

A problem that occasionally turns up in analysis of variance applications may occur with this procedure. When a variance component is estimated from the difference between two other estimates, on rare occasions this difference will be negative. This occurs because the two estimates are independent random variables and it is possible for two extreme values to combine to produce a negative result. If this should occur in the expression for S_P in Equation 3, the appropriate remedy is to set S_P equal to zero. This produces a zero denominator in the Q'-statistic in Equation 4, which is handled in the same manner as it is for a conventional Q-statistic. If the sample mean is greater than or equal to the lower limit, the percentage defective is considered to be zero. If the sample mean is less than the lower limit, the percentage defective is estimated to be 100 percent.

Computer Simulation Tests

In order to demonstrate the effectiveness of the modified Q-statistic given by Equation 4, a computer simulation program was written that permitted the testing of this approach with various combinations of product and testing variability, sample size, and percentage defective. Each simulation run involved 5000 replications of the sampling and testing procedure with a lower limit of L = 3000.0. The other values of interest are presented along with the simulation results in Table 2.

It can be seen from the results in this table

Table 2. Typical results of computer simulation tests.

Samples	Sample Size	Product SD	Testing SD	Product Mean	Product Percentage Defective	Average Estimated Percentage Defective	
						Standard Method	Modified Method
3	4	400.0	0.0	3658.08	5.0	5.06	5.06
3	4	400.0	0.0	3512.69	10.0	10.12	10.12
3	4	400.0	0.0	3336.58	20.0	20.06	20.06
3	4	400.0	0.0	3101.17	40.0	39.96	39.96
3	4	400.0	200.0	3658.08	5.0	5.59	5.29
3	4	400.0	200.0	3512.69	10.0	11.12	10.61
3	4	400.0	200.0	3336.58	20.0	20.60	19.79
3	4	400.0	200.0	3101.17	40.0	40.86	40.44
5	3	400.0	100.0	3658.08	5.0	5.35	5.21
5	3	400.0	100.0	3512.69	10.0	10.52	10.31
5	3	400.0	100.0	3336.58	20.0	20.07	19.79
5	3	400.0	100.0	3101.17	40.0	39.80	39.64
5	3	400.0	300.0	3658.08	5.0	6.38	5.29
5	3	400.0	300.0	3512.69	10.0	12.24	10.40
5	3	400.0	300.0	3336.58	20.0	22.35	20.00
5	3	400.0	300.0	3101.17	40.0	40.71	39.39
7	2	400.0	200.0	3658.08	5.0	5.95	5.11
7	2	400.0	200.0	3512.69	10.0	11.40	10.12
7	2	400.0	200.0	3336.58	20.0	21.04	19.38
7	2	400.0	200.0	3101.17	40.0	40.71	39.88
7	2	400.0	400.0	3658.08	5.0	8.81	5.83
7	2	400.0	400.0	3512.69	10.0	14.69	10.48
7	2	400.0	400.0	3336.58	20.0	24.24	18.71
7	2	400.0	400.0	3101.17	40.0	42.02	39.55

Note: The lower limit = 3000.0 and there were 5000 replications per run.

that, when the testing error is zero (or relatively small compared to product variability), the two methods both produce very accurate estimates of the product percentage defective. As the testing error becomes larger, the standard method begins to overestimate the product percentage defective; however, the modified method continues to be accurate within the degree of precision expected of the simulation experiment.

Limitations of the Modified Method

Although it would be very unusual in actual practice for the testing standard deviation to equal or exceed the product standard deviation, additional simulation tests were made to determine whether the modified method would produce accurate estimates under such extreme conditions. The results of these tests are presented in Table 3.

The tables used for estimating percentage defective do not exist for fewer than N = 3 samples and the modified approach requires a minimum sample size of n = 2 so that testing error can be distinguished from product variability. Therefore, the first group of results in Table 3 represents the smallest total sample and the most severe test of the modified method. Within this group, the results appear to be quite satisfactory up to and including a testing error equal to about one-half of the product standard deviation. As the testing error increases above this level, even the modified method begins to overestimate the product percentage defective, apparently the result of the occurrence of too many negative variance estimates.

The second group of tests in Table 3 demonstrates the mitigating effect of an increased sample size. This reduces the frequency of occurrence of negative variance estimates and lessens the tendency to overestimate the product percentage defective. In this case, the modified method may remain accurate for testing errors somewhat greater than one-half of the product standard deviation.

The inability of the modified method to remain unbiased for extremely large values of testing error is not considered to be a serious drawback. For most practical applications, the component of variability due to testing will be substantially less than the variability associated with the product, a condition under which the modified method is accurate. Furthermore, for those rare cases in which the modified method may exhibit some bias, it is still considerably less biased than the standard method.

In addition to accuracy (lack of bias), another important characteristic of any statistical estimator is its degree of precision (repeatability). Several tests were made that indicated that the precision of the modified method is essentially the same as that of the standard method.

Attempts to Improve the Modified Method

Several attempts were made to improve the modified method. Although none of these proved to be fruitful, each will be discussed briefly in the belief that this may be of some use to other researchers.

Table 4 presents typical results obtained when a total of six methods were applied with increasing levels of testing standard deviation. The first method (Standard Q) refers to the standard method based on Equation 1 and the second method (Q, S_p) is the modified method based on the estimate of the product standard deviation (S_p) from Equation 3 and the modified Q-statistic in Equation 4.

The third method (Q, S_p, EQUIV N) is based on the well-known facts that the variance of a sample drawn from a normal population is a χ^2-distributed variable and that the difference between two χ^2-distributed variables is not distributed as χ^2. Therefore, the value of S_p^2 derived in this paper is not distributed as χ^2 and, consequently, the modified Q-statistic given by Equation 4 must be distributed somewhat differently from the standard Q-statistic in Equation 1. However, if S_p^2 can be assumed to be approximately χ^2 distributed, it is possible to derive an expression that gives the equivalent degrees of freedom in much the same way that this is done for the approximate F-test (2, p. 247). Then, if we assume that the equivalent sample size ($\hat{N}$) is one more than the degrees of freedom, Equation 5 can be written. The estimate of percent-

Table 3. Tests that demonstrate the limitations of the modified method.

Samples	Sample Size	Testing SD	Average Estimated Percentage Defective: Standard Method	Average Estimated Percentage Defective: Modified Method	Relative Frequency of Negative Variance Estimates
3	2	100.0	10.38	10.12	0.029
3	2	200.0	11.29	10.36	0.106
3	2	300.0	13.25	11.29	0.178
3	2	400.0	14.87	11.96	0.259
3	2	500.0	16.37	12.55	0.319
3	2	600.0	19.12	14.67	0.369
3	2	700.0	20.70	15.36	0.405
3	2	800.0	22.99	17.71	0.424
3	4	100.0	10.12	9.85	0.012
3	4	200.0	10.38	9.86	0.054
3	4	300.0	11.48	10.24	0.117
3	4	400.0	12.33	10.57	0.180
3	4	500.0	13.89	11.20	0.236
3	4	600.0	15.00	11.47	0.287
3	4	700.0	17.00	12.61	0.358
3	4	800.0	17.86	13.13	0.371

Note: Lower limit = 3000.0; product mean = 3512.69; product SD = 400.0; product percentage defective = 10.0; replications per run = 5000.

Table 4. Results of attempts to improve the modified method.

Method	Testing SD	Estimated Percentage Defective: Mean	Estimated Percentage Defective: Min/Max	Estimated Percentage Defective: SD	Relative Frequency of Negative Variance Estimates
Standard Q	0.0	10.04	0.0/ 57.5	9.3	0.0
Q, S_P	0.0	10.04	0.0/ 57.5	9.3	0.0
Q, S_P, EQUIV N	0.0	10.04	0.0/ 57.5	9.3	0.0
Q, S_P, NEG VAR	0.0	10.04	0.0/ 57.5	9.3	0.0
STD NML, $S_{P,N-1}$	0.0	10.60	0.0/ 58.0	8.5	0.0
STD NML, $S_{P,N}$	0.0	9.27	0.0/ 58.6	8.2	0.0
Standard Q	100.0	10.38	0.0/ 57.8	9.4	0.0
Q, S_P	100.0	10.04	0.0/ 57.9	9.4	0.0
Q, S_P, EQUIV N	100.0	9.93	0.0/ 57.9	9.5	0.0
Q, S_P, NEG VAR	100.0	10.04	0.0/ 57.9	9.4	0.0
STD NML, $S_{P,N-1}$	100.0	10.60	0.0/ 58.4	8.6	0.0
STD NML, $S_{P,N}$	100.0	9.41	0.0/ 59.0	8.3	0.0
Standard Q	200.0	11.49	0.0/ 56.0	9.9	0.0
Q, S_P	200.0	10.22	0.0/ 56.1	9.9	0.008
Q, S_P, EQUIV N	200.0	9.73	-11.7/ 56.1	10.4	0.008
Q, S_P, NEG VAR	200.0	10.22	-0.1/ 56.1	9.9	0.008
STD NML, $S_{P,N-1}$	200.0	10.72	0.0/ 56.5	9.2	0.008
STD NML, $S_{P,N}$	200.0	9.91	0.0/ 57.0	8.8	0.002
Standard Q	300.0	12.93	0.0/ 53.1	10.4	0.0
Q, S_P	300.0	10.28	0.0/ 54.1	10.4	0.042
Q, S_P, EQUIV N	300.0	9.25	-15.7/ 52.4	11.6	0.042
Q, S_P, NEG VAR	300.0	10.25	-34.1/ 54.1	10.5	0.042
STD NML, $S_{P,N-1}$	300.0	10.69	0.0/ 54.4	9.7	0.042
STD NML, $S_{P,N}$	300.0	10.47	0.0/ 54.1	9.4	0.009
Standard Q	400.0	14.83	0.0/ 65.4	11.1	0.0
Q, S_P	400.0	10.55	0.0/100.0	11.2	0.101
Q, S_P, EQUIV N	400.0	9.14	-18.6/100.0	13.1	0.101
Q, S_P, NEG VAR	400.0	10.30	-39.8/138.5	11.8	0.101
STD NML, $S_{P,N-1}$	400.0	10.86	0.0/100.0	10.7	0.101
STD NML, $S_{P,N}$	400.0	11.31	0.0/ 68.9	10.3	0.031

Note: Number of samples = 7; sample size = 2; lower limit = 3000.0; product SD = 400.0; product percentage defective = 10.0; replications per run = 5000.

age defective can then be obtained by using $\hat{N}$ to interpolate between appropriate tables of the type shown in Figure 2.

$$N = [S_N^2 - (S_n^2/n)]^2/\{[S_N^4/(N-1)] + [S_n^4/n^2(n-1)]\} + 1 \quad (5)$$

Bounds for $\hat{N}$ can readily be derived and are given by Equation 6. Because $\hat{N}$ can take on values as low as one, but the tables for estimating percentage defective do not exist for N less than three, it is sometimes necessary to extrapolate to obtain the estimate of percentage defective. Unfortunately, this often produces negative values of percentage defective, as can be observed from the results in Table 4.

$$1 \leq \hat{N} \leq \text{Max}(N, n) \quad (6)$$

Although the concept of a negative value of percentage defective has no physical interpretation, it conceivably could have practical value if the distribution from which it came provided an unbiased estimate of the true percentage defective of the population being sampled. However, as can be seen in Table 4, this appears not to be the case. For this particular series of tests, the method of using an equivalent value of N is negatively biased to a slightly greater degree than the positive bias of the original modified method.

The fourth method (Q, S_P, NEG VAR) was planned to occasionally produce estimates of percentage defective outside the normal range of 0-100 percent. This procedure did not set negative variance estimates equal to 0 as was done with the previous methods. In this case, S_P was calculated as the square root of the absolute value of the variance. Then, if the variance was negative, a minus sign was attached to the estimate of percentage defective. It was clear that this approach would tend to reduce the positive bias, but to what extent was not known. As seen from the results in Table 4, the reduction was too slight to be of value.

Two additional methods were included, both of which involved the use of the standard normal distribution. For the fifth method (STD NML, $S_{P,N-1}$), a Z-score was computed by using the value of S_P obtained from Equation 3. The final method (STD NML, $S_{P,N}$) was identical except that all standard deviations were computed by using N instead of N-1 in the denominator under the radical. For both procedures, the percentage defective estimate was the area under the standard normal distribution that corresponds to the computed Z-score. As can be observed from the results in Table 4, these two methods were no more successful than the others.

Although the distributions of percentage defective estimates are extremely skewed for all of these methods, the standard deviations have been included in Table 4 as an indicator of precision. Since the standard deviations for all methods are about the same and the original modified method (Q, S_P) exhibits less bias while producing no percentage defective estimates outside the range of 0-100 percent, it is judged to be the best of the methods that were tested.

SUMMARY AND CLOSING REMARKS

The standard method for estimating the percentage defective is unbiased, but the population to which it is applied includes the component of variability due to the testing process and this causes the percentage defective of the actual product to be overestimated. For those situations in which it is deemed desirable to overcome this inaccuracy, a method based on analysis of variance techniques is presented that makes it possible to estimate the actual percentage defective of a continuous normal population exclusive of testing error.

An extensive series of computer simulation tests was conducted to demonstrate that the method is effective, provided the testing error does not exceed about one-half the product standard deviation--a condition that is easily met in most practical situations. Even for those cases in which the testing error is larger than this, the modified method is still considerably less biased than is the standard method.

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