

ON THE DEPENDENCE OF SAMPLING INSPECTION PLANS UPON POPULATION DISTRIBUTIONS

BY ALEXANDER M. MOOD

*University of Texas*¹

1. Introduction. The foundations of the science of quality control and quality determination have been laid by W. A. Shewhart [1, 2]. His ideas pervade what follows, but they are too well known to require discussion here. There is, however, one that should be specifically mentioned, that of *statistically controlled production*, because it provides the justification for the basic assumption of this paper: *When production is statistically controlled, there exists a probability, $P(N, X)$, that a lot of size N will contain X defective items.* Shewhart has given a complete discussion of assumptions of this nature.

Sampling inspection of lots may take one of two courses:

- (a) Item inspection, in which a lot is accepted or completely inspected on the basis of one or more samples drawn from the lot.
- (b) Lot inspection, in which a lot is accepted or rejected on the basis of one or more samples drawn from the lot.

The former has been extensively studied by Dodge and Romig [3, 4, 5]; the latter has received little attention, but some of the basic ideas of Dodge and Romig are applicable to this case also.

In this paper the approach to the general problem of lot inspection will be different from that of Dodge and Romig in one important respect: The role of the population distribution function will be emphasized, whereas they have directed their attention to methods which require no knowledge of the population distribution. Their techniques are particularly valuable when a probability distribution does not exist, that is, when production is not statistically controlled. The interest here will be in the inspection of lots which may be regarded as having been drawn from a statistical population. After the first sample from the first lot has been drawn, something is known of the distribution of that population, and as the inspection proceeds a great body of knowledge may be accumulated. Here, if ever, is a real opportunity to explore and to use a population distribution. The very nature of inspection supplies a continuous flow of information about it. To neglect this information would be wasteful indeed.

It is, therefore, the object of this paper to point the way to more efficient inspection procedures for situations in which production is statistically controlled. The inspection procedure will be considered to be an inferential process—on the basis of one or more samples, and with whatever information is available about the parent distribution, an inference will be made regarding the quality

¹ On leave to the War Department.

of those items which have not been examined. A distinction is made between the original lot and what remains of the lot after samples have been drawn. The latter is the appropriate subject of the inference, inasmuch as the quality of the sample is exactly known. The importance of this distinction will become clear in the third section of the paper.

The subject is, unhappily, very briefly developed. The paper contains a few fundamental results and some suggested procedures that may be used to obtain results of more immediate practical value. Time and facilities were not available for preparation of specific sampling plans.

2. Notation and formulae. The conventional notations $P(u)$, $P(u | v)$, $P(u, v)$ will be used to denote the probability of u , of u given v , of u and v , respectively. A lot will contain N items of which X are defective. A lot from which one sample has been drawn will be called an " x -lot;" after i samples have been drawn it will be referred to as an " x^i -lot." The number of items in the i -th sample will be n_i of which x_i are defective, except that the subscript will often be omitted when $i = 1$. The number of items in an x^k -lot will be:

$$N_k = N - \sum_{i=1}^k n_i$$

of which

$$X_k = X - \sum_{i=1}^k x_i$$

are defective.

The probability of x_i for a given x^{i-1} -lot is:

$$(1) \quad P(x_i | X_{i-1}) = \binom{n_i}{x_i} X_{i-1}^{(x_i)} (N_{i-1} - X_{i-1})^{(n_i - x_i)} / N_{i-1}^{(n_i)},$$

where $\binom{u}{v}$ is the binomial coefficient, and

$$u^{(v)} = u(u-1)(u-2) \cdots (u-v+1).$$

Under this conditional distribution, the m -th factorial moment of x_i is:

$$(2) \quad E(x_i^{(m)} | X_{i-1}) = n_i^{(m)} X_{i-1}^{(m)} / N_{i-1}^{(m)},$$

and the m -th factorial moment of X_i is:

$$(3) \quad E(X_i^{(m)} | X_{i-1}) = N_i^{(m)} X_{i-1}^{(m)} / N_{i-1}^{(m)}.$$

Repeated application of (3) to (2) results in:

$$(4) \quad E(x_i^{(m)}) = n_i^{(m)} E(X^{(m)}) / N^{(m)}.$$

In similar fashion it may be shown that:

$$(5) \quad E\left(\prod_{i=1}^k x_i^{(m_i)}\right) = \prod_{i=1}^k n_i^{(m_i)} E(X^{(\sum m_i)}) / N^{(\sum m_i)}.$$

3. Single sampling. Consider a population of lots of fixed size N such that the probability that a lot will contain X defective items is $P(X)$. If x is the number of defective items in a sample of size n drawn from one of these lots, then the joint probability of x and X is:

$$(6) \quad P(x, X) = \binom{n}{x} \frac{X^{(x)}(N - X)^{(n-x)}}{N^{(n)}} P(X).$$

The fundamental result of this paper is:

THEOREM 1. *The correlation between the number of defective items in the sample, x , and the number of defective items in the remainder of the lot, $X_1 = X - x$, is positive, zero, or negative according as the variance, σ_x^2 , of X is greater than, equal to, or less than $A - A^2/N$, where A represents the expected value of X .*

To prove this statement, one need merely compute the covariance between x and X_1 :

$$(7) \quad r_{xX_1} \sigma_x \sigma_{X_1} = \sum_{x, X} x(X - x)P(x, X) - E(x)(A - E(x)).$$

Summing first on x with the aid of (2):

$$r_{xX_1} \sigma_x \sigma_{X_1} = \sum_X \left(\frac{n}{N} X^2 - \frac{n^{(2)}}{N^{(2)}} X^{(2)} - \frac{n}{N} X \right) P(X) - E(x)(A - E(x))$$

which may be reduced to:

$$(8) \quad r_{xX_1} \sigma_x \sigma_{X_1} = \frac{n(N - n)}{N(N - 1)} \left[\sigma_x^2 - \left(A - \frac{A^2}{N} \right) \right]$$

by employing the definitions of A and σ_x^2 together with the relation,

$$E(x) = nA/N,$$

which follows from (4) on putting $m = 1$.

The fact that $A - A^2/N$ is the variance of a binomial distribution with mean A and range N , suggests:

THEOREM 2. *If X has the binomial distribution,*

$$(9) \quad P(X) = \binom{N}{X} p^X (1 - p)^{N-X},$$

then x and $X - x$ are independently distributed.

This statement is readily verified by substituting (9) in (6), and X_1 for $X - x$; a rearrangement of factors then gives:

$$P(x, X_1) = \left[\binom{n}{x} p^x (1 - p)^{n-x} \right] \left[\binom{N - n}{X_1} p^{X_1} (1 - p)^{N-n-X_1} \right].$$

It is clear that additional samples drawn from such lots will have the same property. Thus, sampling of lots drawn from a binomial population will provide no basis whatsoever for inferences concerning the remainder of the lot.

The question naturally arises as to whether distributions $P(X)$ exist for which $r_{xx_1} = \pm 1$.

THEOREM 3. *If*

$$(10) \quad \begin{aligned} P(X) &= 1, & X &= A, & A &\neq 0 \text{ or } N \\ &= 0, & X &\neq A, \end{aligned}$$

then $r_{xx_1} = -1$; if

$$(11) \quad \begin{aligned} P(X) &= p, & X &= 0 \\ &= 1 - p, & X &= N \\ &= 0, & X &= 1, 2, \dots, N - 1, \end{aligned}$$

then $r_{xx_1} = 1$. These are the only distributions which lead to these values of r_{xx_1} . It is first necessary to compute

$$(12) \quad \sigma_x^2 = \frac{n^{(2)}}{N^{(2)}} \left[\sigma_x^2 + \frac{N - n}{n - 1} \left(A - \frac{A^2}{N} \right) \right]$$

$$(13) \quad \sigma_{x_1}^2 = \frac{(N - n)^{(2)}}{N^{(2)}} \left[\sigma_x^2 + \frac{n}{N - n - 1} \left(A - \frac{A^2}{N} \right) \right]$$

by means of (2), (3), and (4). These, together with (8), may then be used to reduce the condition, $r_{xx_1}^2 = 1$, to the following condition on $P(X)$: either

$$(14) \quad \sum_x (X - A)^2 P(X) = 0,$$

or

$$(15) \quad \sum_x X(N - X)P(X) = 0,$$

whence the theorem follows at once. The distributions defined by (10) and (11) will be referred to hereafter as $P_-(X)$ and $P_+(X)$ respectively.

THEOREM 4. *The correlation, r_{xX} , between x and X is positive unless X is distributed by $P_-(X)$ in which case it is zero.*

Computing the covariance by means of (2), (3), and (4), one finds that

$$(16) \quad r_{xX} \sigma_x \sigma_X = n \sigma_x^2 / N.$$

The reason for so carefully distinguishing between the x -lot and the original lot is now apparent. While the number of defective items in the sample is always positively correlated with the number of defective items in the original lot (Theorem 4), it may be negatively correlated with the number of defective items in the x -lot (Theorem 1). The normal practice is to reject (or completely inspect) the x -lot if the sample has an excessive number of defectives, but when the distribution is sharper than a binomial distribution ($\sigma_x^2 < A - A^2/N$) just the reverse should be done. It is assumed, of course, that defective items would be removed from the sample during its inspection when the inspection was non-destructive.

It is clear that the basic rationale of a sampling inspection plan depends on the condition of Theorem 1. Having chosen a sample size n and an acceptance number a (defined by Dodge and Romig [1]), an x -lot would be

$$\begin{aligned} & \text{Accepted when } x \leq a && \text{if } \sigma_x^2 > A - A^2/N \\ & \text{Rejected when } x > a && \text{if } \sigma_x^2 > A - A^2/N \\ \text{but} & \text{Accepted when } x > a && \text{if } \sigma_x^2 < A - A^2/N \\ & \text{Rejected when } x \leq a && \text{if } \sigma_x^2 < A - A^2/N. \end{aligned}$$

Thus, it is essential that the first two moments of the population distribution be known accurately enough to determine the sign of $\sigma_x^2 - (A - A^2/N)$ before an efficient inspection plan can be devised.

4. Multiple sampling. In this section are given similar criteria for guidance in formulating more elaborate sampling plans. The actual computations are elementary and will be omitted.

THEOREM 5. *The mean and variance of the number of defective items in a sample drawn from an x^i -lot are:*

$$\begin{aligned} (17) \quad & E(x_i) = n_i A/N \\ (18) \quad & \sigma_{x_i}^2 = \frac{n_i^{(2)}}{N^{(2)}} \left[\sigma_x^2 + \frac{N - n_i}{n_i - 1} \left(A - \frac{A^2}{N} \right) \right]. \end{aligned}$$

THEOREM 6. *The mean and variance of the number of defective items in an x^i -lot are:*

$$\begin{aligned} (19) \quad & E(X_i) = N_i A/N \\ (20) \quad & \sigma_{X_i}^2 = \frac{N_i^{(2)}}{N^{(2)}} \left[\sigma_x^2 + \frac{N - N_i}{N_i - 1} \left(A - \frac{A^2}{N} \right) \right]. \end{aligned}$$

THEOREM 7. *The correlation between the numbers of defective items in the i -th and j -th samples is:*

$$(21) \quad r_{x_i x_j} = \frac{1}{\sigma_{x_i} \sigma_{x_j}} \frac{n_i n_j}{N^{(2)}} \left[\sigma_x^2 - \left(A - \frac{A^2}{N} \right) \right].$$

THEOREM 8. *The correlation between the numbers of defective items in the i -th sample and the x^j -lot is given by:*

$$(22) \quad r_{x_i x_j} \sigma_{x_i} \sigma_{x_j} = \frac{n_i(N_j - 1)}{N^{(2)}} \left[\sigma_x^2 + \frac{N - N_j}{N_j - 1} \left(A - \frac{A^2}{N} \right) \right], \quad i > j$$

$$(23) \quad = \frac{n_i N_j}{N^{(2)}} \left[\sigma_x^2 - \left(A - \frac{A^2}{N} \right) \right], \quad i \leq j.$$

Thus, the correlation is always positive if the sample is part of the lot even when X has the distribution $P_-(X)$, except only the case covered by Theorem 4 when $j = 0$. The correlations (21) and (23) will be positive or negative in accordance

with the condition of Theorem 1. The extreme values of all these correlations are again given by the distributions $P_-(X)$ and $P_+(X)$ defined in Theorem 3. When $P(X) = P_+(X)$, they all become plus one; when $P(X) = P_-(X)$, they become:

$$(24) \quad r_{x_i x_j} = -\sqrt{n_i n_j / (N - n_i)(N - n_j)},$$

$$(25) \quad r_{x_i x_j} = \sqrt{n_i(N - N_j) / N_j(N - n_i)}, \quad i > j$$

$$(26) \quad = -\sqrt{n_i N_j / (N - n_i)(N - N_j)}, \quad i \leq j$$

For $i = j = 1$, this last expression becomes minus one in accordance with Theorem 3.

5. Formulation of inspection plans. In practice, the formulation of specific sampling inspection plans would naturally begin with the examination of a preliminary sample (or samples) in order to estimate the first two moments of the population distribution. It would then be convenient to have some simple standard functional form which could be fitted to the distribution by means of these first two moments. Such a standard form must obviously contain two arbitrary parameters and should represent a discrete distribution with range N . The simplest function known to the author which satisfies these conditions is:

$$(27) \quad P_1(X) = \binom{N}{X} C^{(X)} D^{(N-X)} / (C + D)^{(N)}.$$

But it will be seen that this distribution is always sharper than the binomial distribution with the same range and mean. Hence a second form is suggested,

$$(28) \quad P_2(X) = \binom{N}{X} (C + X)^{(X)} (D + N - X)^{(N-X)} / (C + D + N + 1)^{(N)},$$

which, it turns out, is always flatter than the binomial distribution with the same range and mean. It is proposed that these two functions be used as standard forms in the belief that the simplicity of their functional form is a convenience which outweighs the inconvenience of having to study two separate functions.

The factorial moments of these distributions are:

$$(29) \quad \sum_0^N X^{(m)} P_1(X) = N^{(m)} C^{(m)} / (C + D)^{(m)}$$

$$(30) \quad \sum_0^N X^{(m)} P_2(X) = N^{(m)} (C + m)^{(m)} / (C + D + m + 1)^{(m)}$$

The variances are:

$$(31) \quad \sum_0^N (X - A)^2 P_1(X) = \frac{NCD(C + D - N)}{(C + D)^2(C + D - 1)}$$

$$(32) \quad \sum_0^N (X - A)^2 P_2(X) = \frac{N(C + 1)(D + 1)(N + C + D + 2)}{(C + D + 2)^2(C + D + 3)}$$

Examination of the expression, $\sigma_x^2 - (A - A^2/N)$, reveals that for $P_1(X)$ it is always negative, while for $P_2(X)$ it is always positive. Both $P_1(X)$ and $P_2(X)$ approach the binomial distribution when C and D become large in a fixed ratio. $P_1(X)$ becomes $P_-(X)$ when $C = A$ and $D = N - A$. As C and D become larger, the distribution becomes flatter until in the limit it is the binomial distribution. $P_2(X)$ becomes the rectangular distribution, $P(X) = 1/(N + 1)$, when $C = D = 0$, and becomes sharper as C and D increase.

The two distribution functions will not serve to approximate U -shaped distributions, and $P_1(X)$ has the disadvantage that C and D must be integers when they are less than N if negative probabilities are to be avoided, but since $C + D$ will be greater than or equal to N in any case, and much greater than N in most cases, this is not a serious limitation. The two functions are reproduced when the marginal distributions for samples are computed:

$$\begin{aligned}
 (33) \quad P_1(x_i) &= \sum_{x_1, \dots, x_{i-1}} P(x_1, \dots, x_i | X) P_1(X) \\
 &= \binom{n_i}{x_i} C^{(x_i)} D^{(n_i - x_i)} / (C + D)^{(n_i)}
 \end{aligned}$$

$$\begin{aligned}
 (34) \quad P_2(x_i) &= \sum_{x_1, \dots, x_{i-1}} P(x_1, \dots, x_i | X) P_2(X) \\
 &= \binom{n_i}{x_i} (C + x_i)^{(x_i)} (D + n_i - x_i)^{(n_i - x_i)} / (C + D + n_i + 1)^{(n_i)}.
 \end{aligned}$$

This is a most valuable property for two reasons. In the first place, it will appreciably facilitate the tedious machine calculations necessary in the work of providing specific optimum sampling plans. In the second place, it will simplify the study of the population distribution of lots by means of samples from those lots.

These two distributions should, then, provide an adequate basis for the formulation of sampling inspection plans in most circumstances.

6. Efficiency of sampling inspection. There are two aspects to the efficiency of an item inspection plan: the inspection aspect, which would be measured by the proportion of defective items eliminated, and the sampling aspect, which would be measured by the difference between the proportions of defective and good items examined. These two measures are primarily functions of the amount of inspection; the former will be large when the amount of inspection is large, and the latter will ordinarily be large when the amount of inspection is small. They will not, therefore, serve as useful criteria for excellence. The measure to be used here is:

$$(35) \quad E = R_B - R_G$$

where R_B is the proportion of defective items examined, and R_G is the proportion of good items examined. It will be zero when the inspection plan is not at all selective, and will be 100% when all of the defective items and none of the good

items are examined. It measures both aspects mentioned above, but has the disadvantage that it emphasizes one or the other for different amounts of inspection. It is not, therefore, a particularly good measure of efficiency, but it is a good criterion. It should ordinarily be maximized.

For single sampling with an acceptance number, a , and with a population distribution sharper than the binomial, the number of items inspected on the average per lot is:

$$(36) \quad I = n + (N - n) \sum_0^a P(x)$$

and the number of defective items inspected on the average per lot is:

$$(37) \quad B = E(x) + \sum_0^N \sum_0^a (X - x)P(x, X)$$

The efficiency will be:

$$(38) \quad E = B/A - (I - B)/(N - A)$$

which may be put in the form:

$$(39) \quad E = \frac{N(N - n)}{A(N - A)} \sum_0^N \sum_0^a \left(\frac{X - x}{N - n} - \frac{A}{N} \right) P(x, X)$$

after substituting (36) and (37). This may be further simplified to:

$$(40) \quad E = \frac{N(N - n)}{A(N - A)} \sum_0^a \left[\frac{x + 1}{n + 1} P_{n+1}(x) - \frac{A}{N} P_n(x) \right],$$

where $P_m(x)$ is the marginal distribution of x for samples of size m . For distributions flatter than the binomial, the limits of the summations on x would be $a + 1$ to n throughout, instead of 0 to a .

THEOREM 9. For a fixed value of n , the acceptance number which maximizes E is $a = E(x)$ when X is distributed by $P_1(X)$ or $P_2(X)$.

The expression in the brackets of (40) becomes:

$$(41) \quad \frac{E(x) - x}{C + D - n} P_n(x)$$

when (33) is substituted for $P(x)$, and becomes:

$$(42) \quad \frac{x - E(x)}{C + D + n + 2} P_n(x)$$

when (34) is substituted for $P(x)$. This theorem is true for a wider class of distribution functions, $P(X)$, but is not worth pursuing too deeply because its main value is in the light it throws on the general nature of inspection plans. It will be a rare case in practice when n is fixed and a is unrestricted. Some idea of the manner in which E depends on population distributions can be attained by computing it for some simple distributions, and by examination of equation (40).

E can be 100% only when all submitted items are defective, but it will obviously be very near 100% when the distribution is $P_+(X)$ if samples of one are used. However, a more reasonable maximum might be 50% which is the largest possible value when the distribution is rectangular (as is shown in the next section). As the distribution becomes sharper, the maximum efficiency decreases to zero when the binomial distribution is reached. As the distribution becomes still sharper, the efficiency increases until it again reaches 50% for the distribution $P_-(X)$. Thus the efficiency is limited, and, in fact, will ordinarily be further reduced by conditions (fixed amount of inspection, or fixed outgoing quality level, for example) which will not allow the unrestricted maximum efficiency to be used.

7. Sampling plans for the rectangular distribution. Excluding the extreme distributions, $P_-(X)$ and $P_+(X)$, the distribution which provides the simplest illustration of some of the ideas above is the rectangular one:

$$(43) \quad P(X) = 1/(N + 1), \quad X = 0, 1, 2, \dots, N,$$

the mean and variance of which are:

$$(44) \quad \begin{aligned} A &= N/2 \\ \sigma_x^2 &= N(N + 2)/12. \end{aligned}$$

The marginal distribution of x is:

$$(45) \quad P(x) = 1/(n + 1),$$

and the efficiency is:

$$(46) \quad E = 2 \frac{(N - n)(n - a)(a + 1)}{N(n + 1)(n + 2)}.$$

The values of n and a which maximize this expression are:

$$(47) \quad \begin{aligned} n &= \sqrt{N + 2} - 2 \\ a &= (\sqrt{N + 2} - 3)/2 \end{aligned}$$

whence

$$(48) \quad E_{\max} = \frac{1}{2} \left(1 - \frac{1}{\sqrt{N + 2}} \right) \left(1 - \frac{\sqrt{N + 2} - 2}{N} \right),$$

or nearly 50% for large N . This plan eliminates almost 75% of the defective items and entails examination of about 25% of the good items. 50% of all items will be inspected.

If the proportion of items to be inspected is fixed at r , then the maximization of E is subject to the restriction:

$$(49) \quad rN = n + (N - n)(n - a)/(n + 1)$$

and results in:

$$(50) \quad n = \frac{-nN(2-r) + \sqrt{n^2N^2(2-r)^2 + N(rN + 2r - 2)(N - Nr - 1)}}{N(1-r) - 1}$$

or for large N ,

$$(51) \quad \begin{aligned} n &= \sqrt{rN/(1-r)} \\ a &= \sqrt{r(1-r)N}. \end{aligned}$$

If the average outgoing quality (as defined by Dodge and Romig) is to be fixed at p (the proportion of defectives after inspection on the average), then the maximization of E is subject to the condition:

$$(52) \quad p = \frac{(N-n)(a+2)^{(2)}}{N(n+2)^{(2)} + (N-n)(a+2)^{(2)}}$$

and results in the relation:

$$(53) \quad (N-n)(n-a) = (a+1)(n+1)(n+2).$$

When N is large relative to $1/p$, the solution of these last two equations is approximately:

$$(54) \quad \begin{aligned} n &= \sqrt{N} \sqrt{\sqrt{\frac{1-p}{p}} - 1} \\ a &= \sqrt{\frac{p}{1-p}} n. \end{aligned}$$

The same result would have been obtained had the amount of inspection been minimized subject to (52).

8. Summary. Methods of sampling inspection in current use have been made independent of any population distribution that may exist. When production is statistically controlled, a population distribution may be postulated. In such circumstances it is to be expected that knowledge gained of the population by repeated sampling will be a valuable aid in specifying efficient sampling inspection techniques. This paper is a preliminary investigation of the relation of lot sampling inspection plans to population distributions.

Lots are assumed to be drawn from a population such that there is a unique probability the lot will contain a specified number of defective items. It is shown that:

1. The number of defective items in a sample from a lot is positively or negatively correlated with the number of defective items in the remainder of the lot according as the population distribution is "flatter" than or "sharper" than a binomial distribution. Distributions are found for which this correlation is plus or minus one.

2. If the distribution is the binomial one, the number of defective items in the sample is distributed independently of the number of defective items in the remainder of the lot. Thus a sample can furnish no basis for an inference concerning the remainder of the lot.
3. The correlation between the number of defective items in the sample and the number of defective items in the original lot is positive.

These results are generalized for repeated sampling of one lot.

There is discussed a standard functional form which can ordinarily be fitted to population distribution functions for purposes of constructing sampling inspection plans.

It is shown, for a class of distribution functions, that a single sampling plan for nondestructive inspection will be most efficient in a certain sense when the acceptance number is equal to the expected number of defective items in the sample.

Optimum single sampling plans for nondestructive inspection of lots with a rectangular probability distribution are determined for restricted amount of inspection and for restricted average outgoing quality.

REFERENCES

- [1] W. A. SHEWHART, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand, New York, 1931.
- [2] W. A. SHEWHART, *Statistical Method from the Viewpoint of Quality Control*, The Graduate School, U. S. Dept. of Agriculture, Washington, 1939.
- [3] H. F. DODGE AND H. G. ROMIG, "A method of sampling inspection," *Bell System Tech. Jour.*, Vol. VIII (1929) p. 613.
- [4] H. F. DODGE AND H. G. ROMIG, "Single sampling and double sampling inspection tables," *Bell System Tech. Jour.*, Vol. XX (1941) p. 1.
- [5] H. F. DODGE, "A sampling inspection plan for continuous production," *Annals of Math. Stat.*, Vol. XIV (1943) p. 264.