SAMPLING STRATEGIES FOR DESTRUCTIVE TESTS

W. G. Warren
Environment Canada, Forestry Directorate, Western Forest Products Laboratory
Vancouver, British Columbia V6T 1X2

(Received 28 July 1975)

ABSTRACT

Two strategies are considered whereby the full sample need not be destroyed to obtain estimated of low-order percentiles. The properties are examined by Monte Carlo methods. One strategy, based on a proof-loading concept, appears to have practical value, particularly when used in conjunction with a non-destructively-determined concomitant variable. Even when the concomitant variable has a fairly low correlation with the property of interest, the number of samples destroyed is appreciably reduced. This number is, however, a random variable. The second strategy, which again uses a concomitant variable, fixes the number destroyed, but in other respects is less satisfactory; for reliable results the number destroyed will be uncomfortably high unless the variables are very highly correlated.

The use of a subjective ranking rather than a measured concomitant variable, to reduce the number of specimens destroyed, is considered. It is seen that, under some circumstances, subjective ranking, or even allocation into ordered groups, will be cost-efficient.

Additional keywords: Monte Carlo simulation, statistical methods, cost efficiency, concomitant variables.

INTRODUCTION

It is often the case that when the measurement of a certain characteristic requires the destruction of the specimen concerned, the parameter of interest is not the population mean but some low-order percentile. A familiar example concerns the determination of product strength for quality control purposes.

Perhaps the most reliable way of estimating a low-order percentile, such as the fifth, is to fit a smooth curve, by eye, in the region of the percentile, to the empirical distribution function constructed by ordering the observations of a large random sample. Such procedure is distribution-free, there being no possibility of bias from the fitting of an inappropriate function and, if the sample is sufficiently large, the precision should be high.

The drawback of such a system is that although the sample must be large, direct use is made of only a relatively small proportion of the observations that have been made on the entire sample for correct ordering. The larger observations are necessary only to ensure that if we say that we have the first M order statistics, i.e. the lowest M values properly ordered, then this is indeed the case. When the taking of an observation involves the destruction of a perhaps expensive specimen, the method tends to lose whatever attraction it might otherwise have.

There are situations, however, where the first M order statistics can be determined without having the actual observations on all specimens. For example, if the observations were time to failure, then individuals surviving after a certain test period need not be left to fail (provided the period has been chosen so that the minimum requisite number do, in fact, fail). Likewise, in the testing of lumber for some strength property, a certain "proof" load could be applied so that only a fraction of the sample would be broken, the remainder having strength greater than that corresponding to the prescribed load.

In practice, however, we do not necessarily have sufficient knowledge to determine in advance the suitable test period or proof load. In what follows we shall consider some strategies that embody the above philosophy, but overcome this difficulty. For convenience the methods will be described in the terminology of lumber-strength testing, but clearly can be related to other circumstances.

TABLE 1. *Monte Carlo generated expected numbers of specimens destroyed, D, under Strategy A for selected values of M (N = 200)*

	M			
	10	20	30	40
Average D	39.73	64.46	86.36	103.95
Variance of D	26.60	27.04	26.78	35.72
Std. error of D	0.52	0.52	0.52	0.62

STRATEGY A

Let the sample size be N and suppose that we wish to determine the lowest M values. Select M, at random, from the N specimens and test to destruction. Select another specimen at random and apply a load equivalent to the maximum of the M values obtained (the "current proof" load). If the specimen fails, it must be weaker than at least one of the M values. It then must join the set of lowest M values and the previous maximum must be removed from the set. The proof load is then reduced to the new maximum. If the specimen does not fail, it has strength greater than the proof load and no change is required. Another specimen is selected at random and the procedure repeated until all N specimens have been examined.

This strategy ensures that the lowest M out of the N will be found, but actual number D of specimens destroyed is not known in advance and is, in fact, a random variable.

Given N and M, it is a straightforward matter to examine the properties of D by Monte Carlo methods[1]. Since only order is relevant, it follows from the probability transformation (see, e.g. Kendall and Buckland 1971) that, with no loss of generality, the strengths can be assumed to be uniformly distributed; any convenient uniform pseudo-random number generator can then be used. The algorithm is outlined in Appendix 1.

For 100 trials with N = 200 at each level of M for M = 10, 20, 30, 40, the results given in Table 1 were obtained. Histograms are given in Fig. 1.

We see, for example, that if our objective is to determine the lowest 20 out of a sample of size 200 then under this strategy we would break on the average about 64 specimens, although we would also have about a 1-in-20 chance of breaking 73 or more. That is, if this procedure were repeated independently a large number of times, then on the average only one time in twenty would 73 (or more) specimens be broken.

[Note that although percentiles will often be closely approximated by mean $\pm$ K (standard deviation) with K obtained from standard normal tables, the distribution of the number broken is non-normal, and, in general, non-symmetric. If the direct estimate of a percentile, i.e. the appropriate point on the histogram, differs appreciably from the calculated approximation, the direct estimate, if based on a sufficiently large number of trials, would be preferred].

While strategy A could be applied to existing procedures, for example the distribution-free method of ASTM D2915, with the algorithm as here presented, such explicit usage is irrelevant to the immediate purpose. Strategy A is simply presented as a definitive method for determining the lowest M from a sample of size N.

STRATEGY A EXTENDED

It may happen that there is a quantity which can be measured non-destructively and which is correlated with the strength. It is clear that the number of specimens broken would be reduced if, instead of being taken in random order, the specimens were arranged in approximate order of magnitude by the use of such concomitant variable(s). For example, in determining modulus of rupture one might use a non-destructively-determined modulus of elas-

[1] Glick (in press) shows that $E(D) = \sum_{i=1}^{M} 1/i$, thus providing confirmation of our Monte Carlo derived values. The analytical procedure, however, does not readily extend to the case of prior ordering on a concomitant variable considered in the following section. We therefore elect to use Monte Carlo methods throughout.

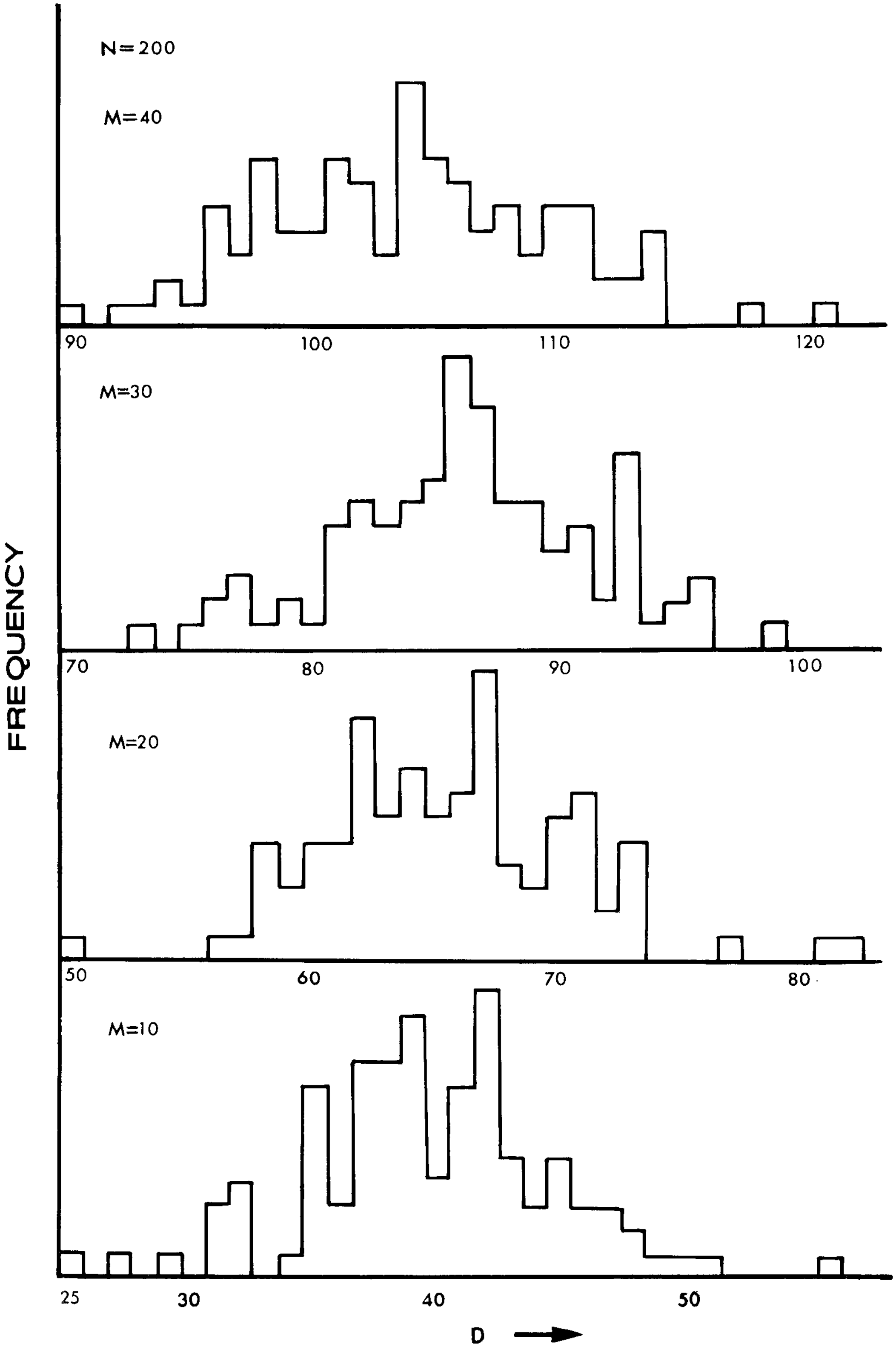

FIG. 1. Histograms of the number destroyed under strategy A for N = 200 and M = 10(10)40.

TABLE 2. *Monte Carlo generated expected numbers of specimens destroyed under Strategy A Extended for various degrees of correlation, ρ, on a concomitant variable and selected values of M (N = 200)*

	M							
	10		20		30		40	
ρ	Av. D	Var. D	Av. D	Var. D	Av. D	Var. D	Av. D	Var. D
0	39.73	26.60	64.46	27.04	86.36	26.78	103.95	35.72
0.1	36.48	36.48	60.92	27.35	79.63	24.38	98.30	27.42
0.2	35.13	32.58	56.52	23.55	76.02	23.09	93.06	27.75
0.3	30.57	22.21	52.88	22.69	72.22	21.12	88.73	34.95
0.4	29.43	29.32	50.14	23.92	67.54	24.76	83.68	26.46
0.5	26.36	17.69	46.46	19.67	63.36	24.72	79.11	20.02
0.6	24.41	16.55	42.64	17.47	58.97	19.38	73.08	24.94
0.7	22.74	19.63	39.00	12.26	55.10	19.97	68.23	14.83
0.8	19.92	13.25	34.86	12.55	49.93	14.39	62.38	15.55
0.9	16.86	9.21	30.06	10.87	43.41	10.24	55.67	10.47
0.925	-	-	28.69	7.89	-	-	-	-
0.95	-	-	26.91	6.75	-	-	-	-
0.975	-	-	25.41	5.25	-	-	-	-
0.99	-	-	23.03	2.74	-	-	-	-

ticity or specific gravity, or even a visual ranking.

To gain some idea of the degree of reduction possible, we shall extend the Monte Carlo study. Let us denote the strength by Y and assume a single concomitant variable X, and suppose that (X,Y) has a bivariate normal distribution. With no loss of generality, we may take the means of X and Y to be zero and the variances to be unity. The correlation between X and Y is denoted by ρ. For a given X, it follows that Y has a normal distribution with mean ρX and variance $1-\rho^2$. Accordingly, we may use a normal pseudo-random number generator to produce N values of X, which we then arrange in ascending order. We then proceed with the algorithm as before, but with the strengths obtained by a normal pseudo-random number generator scaled so as to have mean ρX_i and standard deviation $\sqrt{1-\rho^2}$, where X_i denotes the i^{th} value of X selected in order that $X_1 \leqslant X_2 \leqslant \ldots \leqslant X_N$.

If the correlation is perfect ($\rho = 1$), then D is identically equal to M. The case $\rho = 0$ corresponds to the situation already considered (i.e. random order). One hundred realizations were therefore carried out for N = 200, and each combination of M = 10, 20, 30, 40 and $\rho = 0.1, 0.2, \ldots, 0.9$ with results as given in Table 2 (some additional cases with $\rho > 0.9$ and M = 20 are included).

Plots of average D against ρ are given in Fig. 2. With the exception of values of ρ close to unity, the trend is approximately linear. To illustrate, with a correlation of 0.7, to obtain the lowest 20 out of 200 we would break, on the average, about 39 specimens; however only about one time in twenty would we break more than 45, compared with 64 and 73, respectively, for the same case with no concomitant variable. It would appear that appreciable reduction in the number destroyed can be achieved even by the use of an only moderately correlated concomitant variable.

In practice, the distribution of (X,Y) is unlikely to be bivariate normal (e.g. the marginal distribution of Y may well be better represented by a Weibull form). However, the average number destroyed at $\rho = 0$ and at $\rho = 1$ are independent of distributional form, and it seems reasonable to conjecture that the trend as indicated would not be departed from substantially for any reasonable distribution of the strength and concomitant variables.

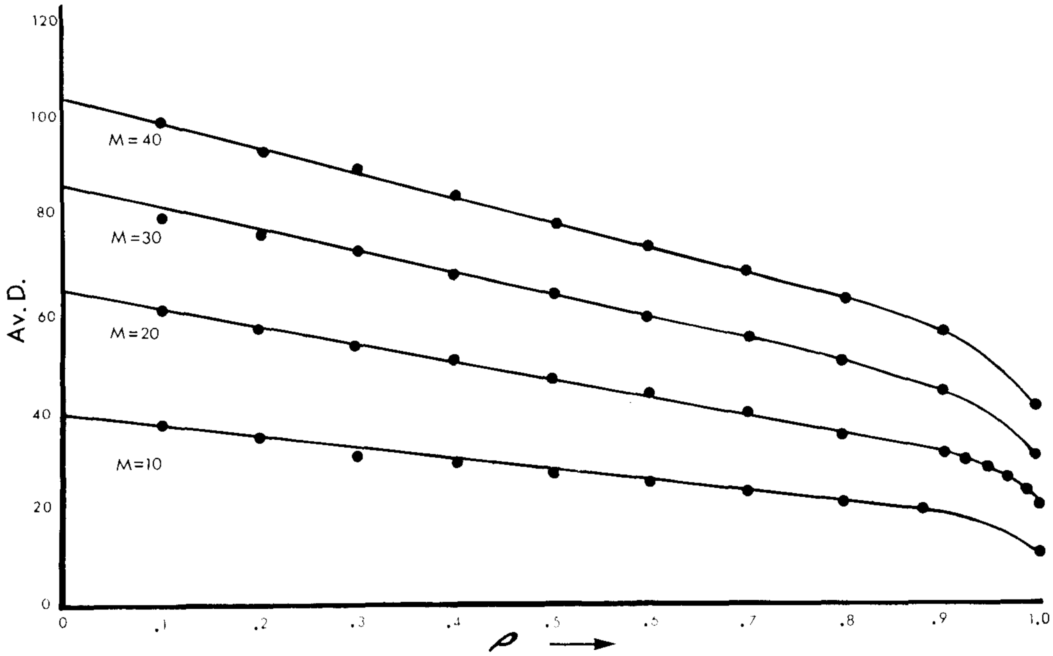

FIG. 2. Relationship between expected number destroyed and degree of correlation for strategy A with a concomitant variable.

STRATEGY B

The above method requires that all N specimens be tested to some degree. As an alternative, again suppose that the specimens have been ordered by means of the concomitant variable, but we test to destruction the lowest P ($\geqslant$ M) so indicated, the remainder being untested. The number P is fixed, and determined so that the probability is sufficiently high that the M lowest values (out of the N specimens) are contained within these P. The price paid for knowing in advance how many will be destroyed is some uncertainty that the actual M lowest values will be found.

Monte Carlo methods have again been employed to study this situation. Bivariate normal random variables have been generated as described in the previous section, but the remainder of the algorithm is now simpler. The P values obtained by ordering on the concomitant variable are then ordered by strength and the M^{th} strength value determined. The number of the remaining N-P strength values greater than the M^{th} are then counted. If the M lowest values are contained within the chosen P, then this count will be zero. Accordingly, the probability of this happening will be estimated by the proportion, q, of the Monte Carlo trials in which the count is indeed zero. From plots of q against P (see Fig. 3), we may estimate the number of tests required to achieve a specified value of q, say 95%. With only 100 realizations this cannot be done very precisely; however the trends are clearly indicated. The estimates, rounded to the nearest multiple of 5, are given in Table 3.

It would appear that this strategy is advantageous only when the correlation between strength and the concomitant variable(s) is very high. Of course, the "proof-loading" strategy could be applied within the lowest P specimens to reduce the number destroyed (which then becomes a random variable), but this would likely have little saving over proof loading the whole sample (of N) and, of course, does not

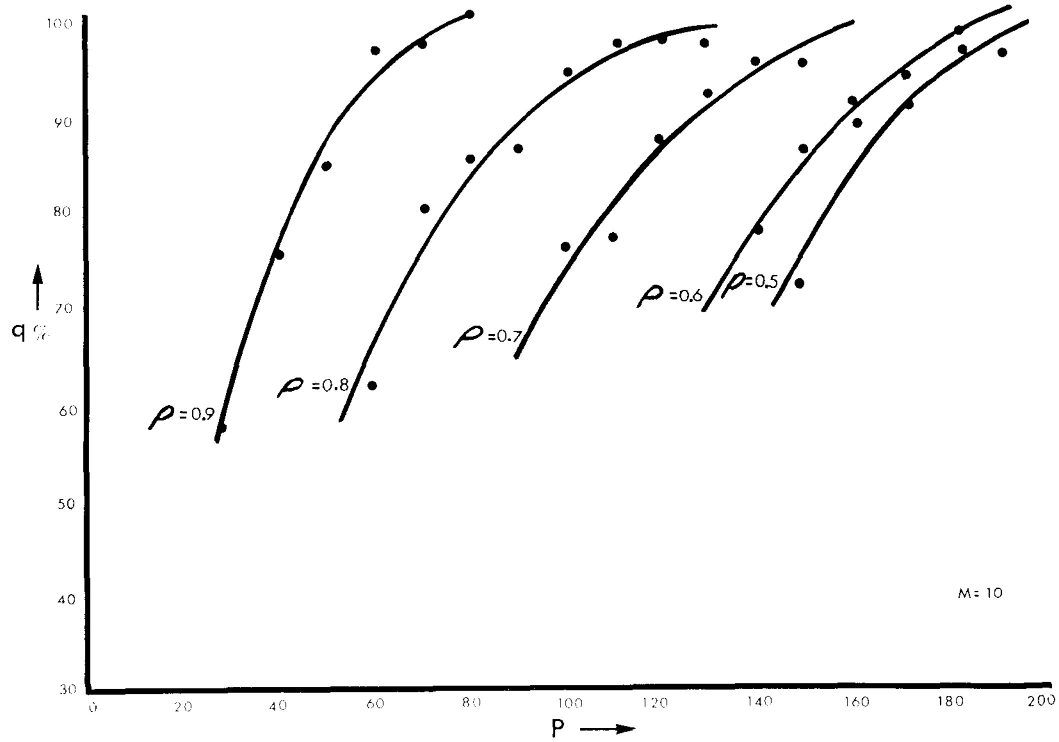

FIG. 3. Relationship between the probability of obtaining the lowest M (= 10) when P specimens destroyed (N = 200) for different levels of correlation on a concomitant variable; Strategy B.

change the risk of not obtaining the lowest M.

It may be possible to tolerate a small error in the determination of the lowest M, i.e. work with a set that contains at most one or two specimens that should not be in it. This would permit some reduction in P, but exploration of the technical consequences of inaccurately determining the lowest M relative to the value of any test specimens not destroyed is too complex a subject for consideration here.

A NOTE ON SUBJECTIVE RANKING

Let us return to the extension of strategy A, but where the prior ordering is based on a subjective, or visual, ranking rather than the actual measurement of an associated property. It is possible to compute the linear correlation between strength (or life span, etc.) and the rank order, the latter taking values 1, 2, 3, ... N, and from past work we may have reasonably good knowledge of the magnitude of this correlation. (It is assumed that the ranking is done by the same person, or by different people working to well-defined rules—that is, we are dealing with a stable situation). However, without further investigation, we can-

TABLE 3. *Approximate number of tests required to achieve q = 0.95 for selected values of M and ρ (N = 200)*

	M			
ρ	10	20	30	40
0.9	60	90	110	125
0.8	105	135	150	165
0.7	140	165	170	180
0.6	165	180	185	190
0.5	180	190	195	195

not assume that the expected number destroyed will be reasonably approximated by the computational procedures that led to Table 3, when this form of correlation is used in place of the correlation between continuous random variables. We here attempt, therefore, to obtain some idea of the magnitude of the error that would arise from such substitution.

It will be assumed that behind the ranking there exists a quantitative variable, the magnitude of which is not known explicitly. We shall call this the implied variable. Such assumption may not be unrealistic; when people rank objects they commonly make such quantification, although perhaps subconsciously.

The implied variable is then assumed to have a bivariate normal distribution with the strength (or life span, etc.). The problem is to determine the relationship between the correlation parameter (ρ) of the "implied" bivariate normal distribution and the correlation, η say, between strength and rank order, the latter obviously being a function of sample size (N). This is a problem in mathematical statistics, a solution to which is outlined in Appendix 2.

It turns out that, as N increases, η approaches ρ and, even for N as small as 20, η is approximately 0.95ρ. Accordingly, there appears to be no error of practical consequence if a well-determined value of η is used in place of ρ for the purpose of estimating the expected number of specimens destroyed by the extension of strategy A.

It should be noted, however, that whereas for a given sample, the value of the correlation between strength and a measured concomitant variable is unique (provided, of course, that the measurements are carefully made) even the same person will not necessarily place the members in the same order were he to carry out the subjective ranking more than once. Mathematical questions aside, it is clear that the estimates of η will exhibit greater variation than the estimates of ρ, so that, although a subjective ranking of the sample may well cost less than a ranking by measurement of a concomitant variable, a greater effort may be required to obtain an estimate of η which has precision equivalent to an estimate of ρ.

In practice it is not necessary to have a full ranking of the complete sample by either objective or subjective methods. It may be sufficient to allocate the material to one of two groups, one "low," the other "high." Our ability to group correctly is, of course, related to our ability to rank. Such allocation into groups can obviously reduce greatly the number of specimens that must be destroyed and provides a potentially comparable reduction in the cost of the testing program. Thus, while the introduction of η may be an artifact, it does lead to valid information on the possible advantage of employing strategy A with the material in other than random order.

DISCUSSION

Obviously a comprehensive study of the possible combinations of M and N by the methods here presented would be a mammoth undertaking, although the algorithms can be readily applied to any particular case of interest. The results from the few cases considered above, nevertheless give some feeling for the situation. It can be concluded that the "proof-loading" method can give substantial reduction in the number of specimens destroyed, especially if used in conjunction with an even moderately correlated concomitant variable or subjective ranking. The actual number destroyed is, however, a random variable. Unless it can be shown that some small level of inaccuracy in the determination of the set of lowest values is inconsequential, it would appear that the risk of such inaccuracy under the destructive testing of a fixed number of specimens, selected on the basis of an even fairly highly correlated concomitant variable, is unacceptable under most practical conditions.

In practice, one must also consider the question of cost, which would include components for labor, machine time, and related supplies as well as that of the test specimens. Each situation has to be examined on its own merits. Indeed, there may well be situations in which it is more economical to proof-load the whole sample than to measure the concomitant variable(s), i.e. the cost of obtaining the concomitant variables might exceed the value of the number of pieces saved from destruction. There are other situations where the concomitant variables will be obtained as an integral part of the study, and it is in these or in the cases of specimens of high value that the strategy will be most attractive.

Throughout the above, it has been tacitly assumed that no damage has resulted from proof-loading a specimen that has not visibly failed—in other words, that surviving specimens would withstand future application of the same load. This is not necessarily the case although definitive results on this topic appear to be lacking. Obviously there is little purpose to strategy A unless the unbroken material has some residual worth; for example, it might possibly be used safely in a lower grade application. Such considerations must be incorporated into the questions of cost.

REFERENCES

GLICK, F. P. A cheap sequential procedure for non-parametric tolerance limits or conservative estimation of small percentiles in stress studies. Am. Statist. Assoc. (in press).

KENDALL, M. G., AND W. R. BUCKLAND. 1971. A dictionary of statistical terms. 3rd ed. Oliver and Boyd, Edinburgh.

TEICHROEW, D. 1956. Tables of expected values of order statistics and products of order statistics from samples of size 20 and less from the normal distribution. Ann. Math. Statist. 27: 410–426.

APPENDIX 1

Outline of Algorithm for Strategy A.

1. Generate and sort into ascending order M uniform pseudo-random variables $X_1 \leqslant X_2 \leqslant X_3 \ldots \leqslant X_M$. It is convenient to define $X_0 = 0$ and X_{M+1} to which a value is not yet ascribed.
2. Set counters C = O, D = M
3. Set C = C + 1
4. If C > N-M, then terminate; total number of specimens destroyed is given by D.
5. Generate a new uniform pseudo-random variable, X.
6. If $X > X_M$, then go to step 3.
7. Set D = D + 1
8. Set j = M – 1
9. If $X < X_j$, then set j = j – 1 and repeat.
10. Let k = M + 1; (it is established that $X_j < X < X_{j+1}$)
11. If k < j + 2 then set $X_{k+1} = X$ and go to step 3.
12. Set $X_k = X_{k-1}$
13. Let k = k–1
14. Go to step 11

APPENDIX 2

Let $X_1, X_2 \ldots . X_N$ be the ordered values of the implied variable, i.e. $X_1 \leqslant X_2 \leqslant \ldots . \leqslant X_N$, and let $Y_1, Y_2 \ldots . Y_N$ be the associated strength values. Note that the Y_i will not, in general, be in ascending order.

We assume

$$(X,Y) \sim N(\mu_x, \mu_y, \sigma^2_x, \sigma^2_y, \rho)$$

With no loss of generality we may take

$$\mu_x = \mu_y = 0, \sigma^2_x = \sigma^2_y = 1,$$

whence

$$Y_i = \rho X_i + e_i$$

where $E(e_i) = 0$, $Var(e_i) = 1 - \rho^2$, $Cov(e_i,e_j) = 0$, and the X_i are independent of the e_i. The problem is then to compare ρ with $E(\eta)$ where

$$\eta = \Sigma(i-\bar{i})\ Y_i/[\Sigma(i-\bar{i})^2\ \Sigma(Y_i-\bar{Y})^2]^{1/2}$$
$$\bar{i} = \Sigma i/N = (N+1)/2, \bar{Y} = \Sigma Y_i/N.$$

Note that $\Sigma(Y_i - \bar{Y})$ is independent of the ordering of the Y_i.

Let us write $\eta = \eta(u,v) = u/C_N\sqrt{v}$

where $C_N = [\Sigma(i-(N+1)/2)^2]^{1/2}$

$= \sqrt{(N-1)N(N+1)/2}$

$u = \Sigma(i-(N+1)/2)Y_i$, $v = \Sigma(Y_i-\bar{Y})^2$.

Then $E(u) = \rho\, E(\Sigma i X_i)$ and $E(v) = n-1$.

As a first approximation to $E(\eta)$ we have

TABLE A1. *First order analytic approximation to $E(\eta)$*

n	3	4	5	6	7	8	9	10	15	20
E(η)	0.846ρ	0.874ρ	0.892ρ	0.905ρ	0.914ρ	0.921ρ	0.927ρ	0.932ρ	0.946ρ	0.954ρ

$$E(\eta) = E(\eta(u,v)) \simeq \eta(E(u), E(v)) = \rho\, E(\Sigma i X_i)/C_n \sqrt{N-1}$$

where $E(\Sigma i X_i) = \Sigma i E(X_i)$ can be readily determined from tables of the expected value of normal order statistics (e.g. Teichroew, 1956). Values of this approximation are given in Table A1 for N = 3(1)10,15,20.

The approximation can be improved by the addition of the term

$$\frac{1}{2}\left[\frac{\partial^2\eta}{\partial u^2}\sigma_u^2 + 2\frac{\partial^2\eta}{\partial u\partial v}\sigma_{uv} + \frac{\partial^2\eta}{\partial v^2}\sigma_v^2\right] \qquad [1]$$

where the second partials are evaluated at E(u),E(v). The adjustment is then

$$\frac{1}{2C_N}\left[\frac{3}{4}\frac{\rho E(\Sigma i X_i)}{(N-1)^{5/2}}\sigma_v^2 - \frac{\sigma_{uv}}{(N-1)^{3/2}}\right] \qquad [2]$$

where $\sigma^2_v = Var(v) = 2(N-1)$ and $\sigma_{uv} \simeq \rho(2-\rho^2)E(\Sigma i X_i)$.

(Verification of this approximation for σ_{uv} is given in a Western Forest Products Laboratory file report available from the author on request).

The adjustment is then, approximately,

$$\rho\,(\rho^2 - 1/2)E(\Sigma i X_i)/2C_N(N-1)^{3/2}$$

so that the refined estimate is

$$\frac{\rho E(\Sigma i X_i)}{C_N\sqrt{N-1}}\left[1 + \frac{\rho^2 - 1/2}{2(N-1)}\right] =$$

$$(\text{Initial Estimate})\left[1 + \frac{\rho^2 - 1/2}{2(N-1)}\right]. \qquad [3]$$

Values of the refined estimate for selected values of N and ρ are given in Table A2.

Monte Carlo methods have been used to investigate the adequacy of the above approximations. Full details are available in the Western Forest Products Laboratory file report. Satisfactory agreement between the analytical and Monte Carlo approximations was achieved; in particular the same trends on N and ρ were observed. For practical purposes the first-order analytical approximation would appear to be adequate, especially for ρ in the range where the reduction in the number of specimens destroyed would offset the expense of ranking, say $\rho > 0.6$.

TABLE A2. *Second order analytic approximation to $E(\eta)$*

ρ = 0.7	n =	3	4	5	6	7	8	9	10	15	20
	E(η) =	0.844ρ	0.873ρ	0.891ρ	0.904ρ	0.913ρ	0.921ρ	0.926ρ	0.931ρ	0.946ρ	0.951ρ
	ρ =	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
n = 3	E(η) =	0.743ρ	0.749ρ	0.760ρ	0.774ρ	0.793ρ	0.817ρ	0.844ρ	0.876ρ	0.912ρ	0.948ρ
n = 10	E(η) =	0.906ρ	0.908ρ	0.910ρ	0.914ρ	0.919ρ	0.924ρ	0.931ρ	0.939ρ	0.948ρ	0.957ρ