AD-A083 902 MISSOURI UNIV-COLUMBIA DEPT OF STATISTICS F/G 12/1
A SELECTION PROCEDURE USING A SCREENING VARIATE.(U)
MAR 80 R W MADSEN N00014-76-C-0789
UNCLASSIFIED TR-90 NL

1 OF 1
AD
A083902

END
DATE
FILMED
6-80
DTIC

(12)

LEVEL II

ADA083902

University of Missouri-Columbia

A Selection Procedure Using a Screening Variate

by

Richard W. Madsen

Technical Report No. 90
Department of Statistics

March 1980

Mathematical Sciences

80 5 5 018

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: (14) TR-90	2. GOVT ACCESSION NO.: AD-A083 902
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): (6) A Selection Procedure Using a Screening Variate	5. TYPE OF REPORT & PERIOD COVERED: (9) Technical rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) Richard W. Madsen	8. CONTRACT OR GRANT NUMBER(s): (15) ONR-N00014-76-C-0789
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Missouri-Columbia, Columbia, Missouri 65211	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-353
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, Virginia 22217	12. REPORT DATE: (11) March 1980
	13. NUMBER OF PAGES: 24
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 26	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Screening variables, Bivariate normal model, Selection procedures using correlated variates.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Consider N objects on which two correlated measurements X and Y can be made. Assume that the probability that the Y measurement meets a certain specification is γ. We present a method whereby a maximal subset of m out of the N objects can be chosen, based on the observed X measurements, so that there is a high probability that a large proportion of the selected subset will meet the desired specification related to the Y measurement. The method we propose uses the conditional probabilities that Y will meet the specification given by the observed value of the x's. The procedure compared favorably (continued on back of sheet)

gamma

next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

402660

20. **Continued.**

with one suggested by Owen, Chen, and Li but has the advantage that no special tables are needed.

A Selection Procedure Using a Screening Variate

Richard W. Madsen*

University of Missouri - Columbia

1. Introduction and Background

Consider N objects on which two correlated measurements X and Y can be made. Assume that the probability that the Y measurement meets a certain specification (e.g. $Y \leq u$) is γ. We present a method whereby a maximal subset of m out of the N objects can be chosen, based on the observed X measurements, so that there is a high probability (ζ) that a large proportion $(\Pi > \gamma)$ of the selected subset will meet the desired specification related to the Y measurement. In general such a selection procedure would be used when Y is based on a measurement which is difficult or expensive to make and X is based on one which is easier or less expensive to make. For example measuring Y may actually destroy the item being tested whereas measuring X will not. In another situation Y might be a student's grade point average after

---

*Research supported in part by the Office of Naval Research, Contract ONR-N00014-76-C-0789.

a number of years in college, while X is a score made on an entrance or qualifying examination.

A related problem considers an infinite population where we assume that a certain proportion, say γ , of the observed Y variates satisfy some specification. By screening on the observed value of the correlated variable X , it may be possible to raise the proportion of Y variates which satisfy the specification to a higher value, say δ . We will assume, as is generally done, that X and Y have a bivariate normal distribution with correlation coefficient ρ .

This problem has been studied for quite some time with the work of Taylor and Russell [8] in 1939 being among the earliest. More recently D. B. Owen and various co-researchers ([3], [4], [7], and [9]) have studied other aspects of this problem. For example, Thomas, Owen, and Gunst [9] considered two screening variables X_1 and X_2 ; Li and Owen [3] considered two sided screening procedures; Owen and Boddie [4] considered screening methods with some parameters unknown. In these cases, sharp cut-off scores are found such that if the X score is in a given range, say $X \leq \mu_X + k\delta_X$, the corresponding item is selected.

Much of the work done in this area has been to table values of k corresponding to values of γ , ρ , etc. to meet certain specifications. Hence one potential deterrent to implementation of these screening procedures is the need for specialized tables. A second point to consider is that with the usual procedures the precise value of X is not used, rather only the fact that X is above or below a given cut-off score is used. The procedure presented here has the advantage of not needing special tables (other than standard normal tables). It also makes use of the precise observed value of X, not simply whether or not the score is above a given cut-off. We assume that there are a finite number N of items available for screening.

2. The Selection Procedure

Consider a finite collection of objects, say N objects, on which it is possible to make measurements X and Y which come from a bivariate normal distribution with correlation coefficient $\rho > 0$. Assume that an item is acceptable if $Y \leq u$ and that the overall proportion of such acceptable items is to be raised from γ (before screening) to δ (after screening). Following the procedure of Owen, Chen, and Li [5], we might find a value k such that an item is selected if $X \leq \mu_X + k\sigma_X$. The value k, of course, is a function of the parameters. For this value of k,

$$P[Y \leq u \mid X \leq \mu_X + k\sigma_X] = \delta . \qquad (1)$$

While in an exceedingly large population (which we might take to be "infinite"), the proportion of selected items which are accepted will be δ, in a finite set of selected items the actual proportion of acceptable items will be a random variable. Specifically, if m items are selected, then the actual number of those items for which $Y \leq u$, will be a binomial random variable, say V, with parameters m and δ. If we want the proportion of acceptable items in the finite set of selected items to be at least Π with probability at least ζ, i.e. if we want

$$P[V \geq \Pi m] = \sum_{j=\ell}^{m} \binom{m}{j} \delta^j (1 - \delta)^{m-j} \geq \zeta , \qquad (2)$$

where $\ell = \ell(m)$ is the smallest integer greater than or equal to Πm , then δ must be chosen suitably large. By using the interrelationships among the binomial, beta, and F distributions it can be shown that a suitable choice for δ is given by

$$\delta = \ell/[\ell + (m - \ell + 1)F_{\zeta,2m-2\ell+2,2\ell}]$$

where $F_{\zeta,a,b}$ is the $(1 - \zeta) \cdot 100\%$ upper tail percentage point for an F distribution having a and b degrees of freedom for the numerator and denominator. This value of δ is then used in (1) and the value of k to satisfy the equality in (1) can be found by using tables given in Owen, McIntire, and Seymour [6]. Note that the value of m must be specified in advance and hence is a fixed quantity.

In the procedure we propose, we assume that a large lot of N items is available for screening. The values of X , call them $X_1, X_2, \cdots , X_N$ are found for each item. Using the conditional distribution of Y given $X = x$, calculate p_i by

$$p_i = P[Y \leq u \mid X_i = x_i]$$

$$= \Phi[(u - \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}(x_i - \mu_X))/(\sigma_y\sqrt{1 - \rho^2})]$$

where $\Phi[\cdot]$ represents the CDF of a standard normal random variable. Since we assume the parameters of the bivariate normal distribution are known, we can, wlog, take them to be $\mu_X = \mu_Y = 0$, $\sigma_X^2 = \sigma_Y^2 = 1$. In this case we have

$$p_i = \Phi[(u - \rho x_i)/\sqrt{1 - \rho^2}] \ . \tag{3}$$

By first ordering the x_i's , we can assume that $p_1 \geq p_2 \geq \cdots \geq p_N$. Now for each value of m , define $\ell(m)$ to be the smallest integer greater than or equal to $\Pi \cdot m$. The selection procedure is to select the m^* items having the largest p_i values, where m^* is the largest integer satisfying a relationship like (2), namely

$$P[V(m^*) \geq \ell(m^*)] \geq \zeta \ . \tag{4}$$

In so doing we select as many items as possible subject to satisfying the constraint given in (4). (Note that this kind of situation might be desirable for a manufacturer who produces lots of N items and wishes to sell a sub-lot of size m^*, as large as possible, such that a proportion Π or more (say a guaranteed proportion) of the screened items are satisfactory with probability at least ζ .)

In order to calculate the probability in (4), it is necessary to note that since we first order the X values and then consider the conditional distribution of the corresponding Y values given the X's , we are dealing with what are known as the concomitants of order statistics. (See David, O'Connell, and Yang [2].) It follows from Bhattacharya's work [1] that the Y_i values, conditional on the ordered X_i values, are independent with conditional distributions which are normal with $\mu_{Y|x_i} = \rho x_i$ and $\sigma^2_{Y|x_i} = 1 - \rho^2$. Now let the x_i values be given and define

$$W_i = \begin{cases} 1 & \text{if } Y_i \leq u \\ 0 & \text{otherwise .} \end{cases}$$

then the W_i are (conditionally) independent Bernoulli random variables and $P(W_i = 1) = p_i$. Consequently

$$P[V(m) \geq \ell(m)] = P[\sum_{i=1}^{m} W_i \geq \ell(m)]$$

$$= \sum \prod_{i=1}^{m} p_i^{\alpha_i}(1 - p_i)^{1-\alpha_i} \equiv \zeta_m \qquad (5)$$

where the sum is taken over all vectors $\underline{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_m)$ satisfying $\alpha_i = 0$ or 1, $\Sigma\alpha_i \geq \ell(m)$. We then take m^* to be the largest value of m such that $\zeta_m \geq \zeta$.

In trying to fin⅃ m^*, one could systematically calculate ζ_m for $m = N, N-1, N-2, \cdots$ stopping as soon as $\zeta_m \geq \zeta$. However it is not necessary to check all values of m because some values are inadmissible. Specifically, if $\ell(m) = \ell(m+1)$, then the value m is inadmissible. (For example if $\Pi = .8$, then $\ell(4) =$ (smallest integer $\geq (.8)(4) = 3.2) = 4$ and $\ell(5) = 4$. Since

$$P[V(4) \geq \ell(4) = 4] \leq P[V(5) \geq \ell(5) = 4] ,$$

it follows that $\zeta_4 \leq \zeta_5$. We would never take m^* to be 4 since if $\zeta_4 \geq \zeta$, it must also be true that $\zeta_5 \geq \zeta$. Hence we say that when $\Pi = .8$, the value $m = 4$ is inadmissible.) Since N is the size of the lot, take N to be admissible. Since the p_i are in decreasing order, it might appear that for admissible m's, the quantities ζ_m are strictly decreasing. However because of the rounding upwards that is done in calculating $\ell(m)$, this need not be the case. Consequently by following the given algorithm we can be sure to find m^*.

(1) Place the observed x_i in increasing order and relabel the x's so that $x_1 \leq x_2 \leq \cdots \leq x_N$.

(2) Find $p_i = P[Y_i \leq u \mid X_i = x_i]$

$$= \Phi[u - \rho x_i)/\sqrt{1 - \rho^2}] , \quad i = 1, 2, \cdots N$$

(3) Find the admissible

m_i , $m_1 < m_2 < m_3 < \cdots < m_e = N$.

(4) Set $j = e$ and find ζ_{m_j} by using equation (5).

(5) If $\zeta_{m_j} \geq \zeta$, set $m^* = m_j$. Otherwise reduce j by 1 and calculate the next ζ_{m_j} .

Note that while the sequence of $\{\zeta_{m_j}\}$ is not strictly decreasing empirical studies indicate that the size of any increase in successive terms is quite small relative to the typical amount of decrease. From a practical viewpoint then, one might use a different algorithm. For instance one might choose a middle admissible m_j and increase or decrease j depending on the value of ζ_{m_j} .

If m_j is sufficiently large, the value of ζ_{m_j} can be approximated by using a normal distribution. The use of this approximation can be justified by using a central limit theorem for independent but not identically distributed random variables. In particular

$$P[V(m_j) \geq \ell(m_j)]$$

$$\approx 1 - \Phi[(\ell(m_j) - .5 - \sum_{i=1}^{m_j} p_i)/(\sum_{i=1}^{m_j} p_i(1 - p_i))^{1/2}] \quad (6)$$

If N is relatively small so that ζ_m is to be found exactly by using (5) rather than being approximated by a normal distribution, the calculations can be quite tedious. One possible means of eliminating some of the calculations is to use Chbychev's inequality. In particular we have

$$E(V(m)) = \sum_1^m p_i \ , \quad \sigma_{V(m)} = (\sum_1^m p_i q_i)^{1/2}$$

so if $(\Pi m - \sum_1^m p_i) > 0$, then

$$P[V(m) \geq \Pi m] = P[V(m) - E(V(m)) \geq (\Pi m - \sum_1^m p_i)]$$

$$\leq P[\ |\ V(m) - E(V(m))\ |\ \geq (\Pi m - \sum_1^m p_i)] \leq \frac{1}{k^2} \ ,$$

where $k = (\Pi m - \sum_1^m p_i)/\sigma_{V(m)}$. It follows that

$$P[V(m) \geq \Pi m] < \zeta$$

provided that $(^1/_k2) < \zeta$, i.e. $\dfrac{\sum_1^m p_i q_i}{(\Pi m - \sum_1^m p_i)^2} < \zeta$. (7)

Consequently for any admissible value of m for which the inequality in (7) holds, the value ζ_m will be less than ζ, hence need not be calculated explicitly. If a normal approximation is to be used, the computations are quite simple and shortcut methods are not quite so necessary.

3. An Example

In this example we take $N = 10$. The data shown in Table 1 was generated from a bivariate normal distribution with $\mu_X = \mu_Y = 0$, $\sigma_X^2 = \sigma_Y^2 = 1$, and $\rho = .90$.

Table 1. Data for Example 1.

x_i	y_i	p_i
-1.8772	-1.3569	.9995
- .8058	- .7349	.8606
- .6222	- .8524	.7592
- 4457	- .0962	.6327
- .0152	- .8580	.2912
.3443	.5514	.0982
.5310	.5349	.0468
.5431	- .0648	.0444
1.2019	1.6161	.0011
1.7573	1.3359	.0000+

For convenience the x values have been placed in increasing order. The y values correspond to the appropriate x's . (That is the (x,y) pairs are ordered by the first element.) We will take $\gamma = .4$, $\Pi = .6$, and $\zeta = .90$. That is in the unscreened population $\gamma = 40\%$ of the items are acceptable. We wish to choose a subset of the $N = 10$ items available such that at least $\Pi = 60\%$ of the items in the screened subset are acceptable with probability $\zeta = .90$.

From this information we find the value of u . Since

$$P[Y \leq u] = .40 = \gamma$$

and since Y has a standard normal distribution, it follows that $u = -.2533$. Following the steps of the algorithm, we next find the values

$$p_i = \Phi[(u - \rho x_i)/(1 - \rho^2)^{1/2}]$$
$$= \Phi[(-.2533 - .90x_i)/.4359] .$$

These values are shown in the third column of Table 1.

Next find the admissible m's. With $\Pi = .6$, the admissible values of m are 1, 3, 5, 6, 8, and 10. Direct calculations show that for $m = 8$ and 10, the inequality (7) holds, so ζ_8 and ζ_{10} are both less than .90. Using (5) we find

$$\zeta_6 = .5751 , \quad \zeta_5 = .8909 , \quad \zeta_3 = .9663$$

so we would take $m^* = .3$. From Table 1 it can be seen that for the top three x values, each of the corresponding y's turned out to be acceptable (i.e. $y_i \leq u = -.2533$). In this sample all screened items happened to be satisfactory. In general, by following this procedure, at least 60% of the screened sample would be acceptable at least 90% of the time.

4. Comparison with Another Procedure

The procedure that we have proposed for screening is most similar to the one discussed in Owen, Chen, and Li [5]. For convenience we will refer to their procedure as the OCL procedure and will refer to ours as the Sigma procedure. The OCL procedure is to find a single cutoff score k_o so that any item having an X score below k_o is accepted In order to find the value of k_o from tables it is necessary to know the value of m_o, the total number of items to be accepted. This implies then that the number of items to be accepted is determined before inspection starts. It would be logical then to inspect the items one at a time, say as they become available. The inspection process would terminate when m_o items have been accepted. One advantage of such a procedure is that it is immediately known whether or not an item is to be accepted or rejected. Of course it is possible that the pool of items which are being inspected is too small to be able to find m_o acceptable items, in which case new screening criteria must be set forth resulting in a new cutoff score, etc.

With the Sigma procedure, all of N items available are inspected and the number which will ultimately be accepted, say M, is a random variable. Likewise the cutoff score is a random variable, say K. Since the value of K is not known until all N items have been inspected, even though the value of X and the corresponding value of p are known for a particular item, it may not be immediately known whether or not that item will ultimately be accepted. (In some cases, however, if the value of p is high enough, one can be virtually certain that the item will be acceptable. See Appendix 1.)

Monte Carlo studies were performed to compare the OCL and Sigma procedures quantitatively. Since there are some qualitative differences (e.g. in the OCL method m_o is fixed while in the Sigma method M is random), some reasonable basis for comparison had to be made. We took m_o to be 100 for the OCL method so that items were screened sequentially until 100 acceptable ones were found. We denote the random number which had to be screened by N_{OCL} . In the Sigma method N was determined empirically so that, in 500 Monte Carlo trials, the sample average value of M was also approximately 100 . (we were satisfied if $\bar{m}$ was within 100 ± 1.) Some results of the Monte Carlo studies are shown in Table 2. There are four comparisons that can be made here:

(1) Observed sample proportion of satisfactory items.

(2) Estimated variance of proportion of satisfactory items.

(3) Estimated average number screened to get 100 acceptable.

(4) Observed number of times that $V \geq \Pi m$.

Table 2

ζ	ρ	γ	π	Observed Prop. Successful $\hat{\mu}$	$\hat{\sigma}$	$\hat{\mu}$	$\hat{\sigma}$	# Screened to get 100 accepted OCL	Sigma	Observed #($V \le \pi m$) OCL	Sigma	# times all N accepted
.90	.90	.4	.4	.4582	.0503	.4454	.0346	114.64	111.37	447	464	85
			.6	.6507	.0482	.6483	.0392	169.78	169.31	445	445	—
			.8	.8385	.0372	.8396	.0341	263.66	264.67	434	439	—
			.9	.9280	.0256	.9294	.0252	373.69	379.21	447	447	—
		.8	.8	.8442	.0368	.8396	.0295	105.52	105.05	455	447	96
			.9	.9311	.0262	.9326	.0247	122.43	122.97	459	464	—
		.9	.9	.9316	.0256	.9309	.0223	103.44	103.67	465	470	89
.90	.99	.4	.4	.4598	.0498	.4342	.0262	114.73	108.39	452	477	150
			.6	.6546	.0464	.6306	.0246	164.19	157.93	450	455	—
			.8	.8435	.0376	.8335	.0261	212.74	210.47	442	449	—
			.9	.9387	.0242	.9290	.0232	246.15	241.77	475	460	—
		.8	.8	.8390	.0367	.8273	.0209	105.26	103.74	433	471	146
			.9	.9294	.0269	.9238	.0190	116.55	115.87	438	463	—
		.9	.9	.9281	.0270	.9225	.0179	103.24	102.64	438	479	164
.99	.90	.4	.4	.5105	.0486	.4809	.0369	128.02	120.59	495	495	11
			.6	.7010	.0467	.6859	.0371	186.67	181.67	496	493	—
			.8	.8677	.0344	.8741	.0329	282.37	289.24	490	495	—
			.9	.9496	.0211	.9477	.0227	439.73	429.85	496	492	—
		.8	.8	.8765	.0347	.8635	.0278	110.46	108.63	492	495	16
			.9	.9531	.0217	.9492	.0217	130.14	129.14	498	495	—
		.9	.9	.9506	.0213	.9457	.0199	107.55	106.96	498	498	6
.99	.99	.4	.4	.5099	.0549	.4485	.0246	128.21	112.87	492	496	99
			.6	.6982	.0453	.6538	.0254	176.23	164.79	492	492	—
			.8	.8794	.0316	.8600	.0254	223.79	218.15	495	497	—
			.9	.9455	.0221	.9464	.0213	251.19	251.57	492	498	—
		.8	.8	.8749	.0322	.8421	.0204	109.81	105.71	494	497	96
			.9	.9498	.0213	.9386	.0176	119.21	117.85	494	497	—
		.9	.9	.9511	.0209	.9322	.0158	105.69	103.62	498	498	86

The first and third of these are probably of greatest interest. In the first case, in order to have a high probability that $V \geq \Pi m$, the actual proportion of satisfactory items must exceed Π. However it is advantageous to a manufacturer, for example, to exceed Π by as little as possible. In almost every case the observed proportion is closer to Π for the Sigma procedure, especially when $\Pi = \gamma$. To compare average numbers screened to get 100 acceptable items, we used an empirical determination of N so that $\bar{m}$ was within 100 ± 1. Then the estimated average number under the Sigma procedure was taken to be $100N/\bar{m}$. This number is compared with $\bar{n}_{OCL}$, the average number screened under the OCL procedure to get 100 acceptable items. In 22 out of 28 cases investigated the average is smaller for the Sigma procedure.

In 500 trials, the expected number of times that V will be at least $\Pi \cdot m$ should be $500 \cdot \zeta$ with a variance of $500\zeta(1 - \zeta)$. There are some situations where the observed number is higher than this expected number for the Sigma procedure. However the cases where this happens all correspond to cases where $\gamma = \Pi$. Closer examination of the Monte Carlo output revealed that in these cases there were several times when m^*, the number accepted, was equal to N, i.e. all N items were "accepted." (See the last column of Table 2.) The implication of this is that $P[V(N) \geq \Pi \cdot N] \geq \zeta$, and in fact the probability most likely exceeds ζ by some amount. This leads to a higher expected value for the number of.times that V will be at least $\Pi \cdot m$. In each of the four quantitative comparisons the Sigma procedure compares quite favorably with the OCL procedure.

Another comparison that can be made using Monte Carlo studies is with the fixed cutoff score (k_o) of the OCL procedure and the random cutoff score K of the Sigma procedure. Recall that in the OCL procedure an item is accepted if $X \leq \mu_X + k_o \sigma_X = k_o$ (if $\mu_X = 0$ and $\sigma_X = 1$). In Table 3 the values k_o and $\bar{k}$, the average cutoff score based on 500 trials, are compared. Since K is random, there are times when it is smaller than k_o and other times when it is larger. If it is, on the average, larger than k_o, this indicates that the acceptance cirterion is less stringent. The observed value of $\bar{K}$ is larger than k_o in 23 out of 28 cases.

The fact that K is random and not fixed has the following implication. In repeated trials, a score of $x = k_o + \varepsilon$ would never be accepted under the OCL procedure while a score of $x = k_o - \varepsilon$ would always be accepted. However in repeated trials using the Sigma procedure, the probability that a score of x_o will be accepted is a non-increasing function of x_o. This is illustrated in Figure 1. This figure is based on the empirical Monte Carlo results and does not give the exact distribution of K . From this we see that, rather than there being a strict cutoff (as in the OCL procedure), the lower (better) the x score attained by an item (or individual) the higher the probability that the item will be accepted.

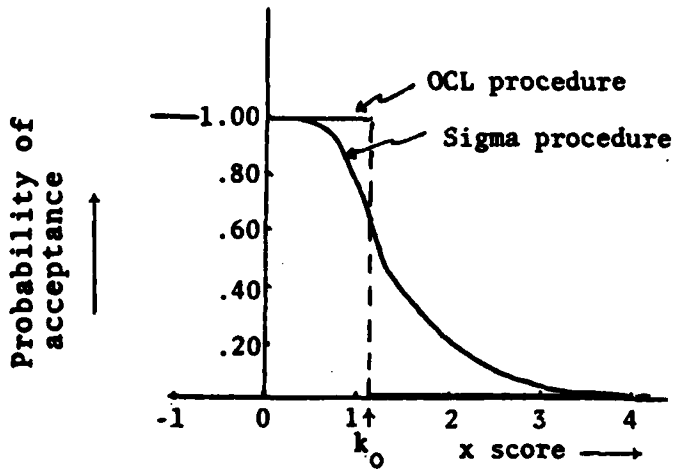

Figure 1. Comparison of Cut-off scores.

$\zeta = .9$, $\rho = .9$, $\gamma = .4$, $\Pi = .4$

5. Conclusions

There are various situations under which screening may be done. Some examples are admissions tests for educational placement, competency tests for employment, or quality tests for a product. Based on Monte Carlo studies, the Sigma procedure that we have described compares favorably with the procedure suggested by Owen, Chen, and Li in terms of meeting the conditions given in (4),

$$P[V(m^*) \geq \ell(m^*)] \geq \zeta \ .$$

The comparison is also favorable in terms of the number of items N which need to be screened in order to obtain a specified number of acceptable items. The Sigma procedure may offer a disadvantage in that the X measurements need to be made on the entire population of N items before it can be determined which items can be accepted. One of the major advantages is that the Sigma procedure does not require any specialized tables to implement.

Appendix

Determination of a "Guaranteed" Acceptable Score

In Section 4 we pointed out that the X measurements need to be taken on the entire population of N objects before determining which items in the population will be in the acceptable subset. In the OCL procedure, as soon as the x score for an item is found, it is immediately known whether or not x is less than or equal to k_o , and hence whether or not the item is acceptable. However if we assume that we know approximately how many items out of N , say m , will ultimately be acceptable, and if m is large enough to use a normal approximation, then we can find an x *score (or equivalently a* value of p) for which the corresponding item is almost certain to be accepted. The most difficult case would be when all items are essentially at the same minimal score, so we will let p denote this common conditional probability. Using (6) with all $p_i = p$ we obtain

$$P[V(m) \geq \ell(m)] \approx 1 - \Phi[(\ell(m) - .5 - mp)/(mp(1 - p))^{1/2}] \; . \tag{8}$$

If we let $\ell = \ell(m) - .5$, then the right hand side of (8) will be at least ζ if

$$(\ell - mp)/(mp(1 - p))^{1/2} \leq z \tag{9}$$

where $z = z_{1-\zeta}$ denotes the $100(1 - \zeta)\%$ point of the standard normal distribution. Solving the inequality in (9) for p via the quadratic formula yields

$$p \geq \frac{\ell + (z^2/2) + |z|\ell[(1 - \ell/m) + z^2/4]^{1/2}}{m + z^2}$$

The corresponding x values can be found by solving (3) to get

$$x \leq \frac{u - \Phi^{-1}(p)[1 - \rho^2]^{1/2}}{\rho}$$

From the empirical Monte Carlo studies these values of p and x appear to be highly conservative.

As an example, if $\zeta = .9$, $\rho = .9$, $\gamma = .4$, and $\Pi = .4$, and if we assume that m will be about 100 then we will find

$$p \geq \frac{39.5 + (\frac{(-1.282)^2}{2} + 1.282[39.5(1 - \frac{39.5}{100} + \frac{(-1.282)^2}{4}]^{1/2}}{100 + (-1.282)^2}$$

$$= .4589$$

and the corresponding values of x would be

$$x \leq \frac{-.2533 - \Phi^{-1}(.4589)[1 - (.9)^2]^{1/2}}{.9} = -.2315 .$$

References

1. Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *Annals of Statistics*, 2, 1034-1039.

2. David, H. A., O'Connell, M. J., and Yang, S. S. (1977). Distribution and expected value of the rank of a concomitant of an order statistic. *Annals of Statistics* 5, 216-223.

3. Li, L. and Owen, D. B. (1979). Two-sided screening procedures in the bivariate case. *Technometrics* 21, 79-85.

4. Owen, D. B. and Boddie, J. W. (1976). A screening method for increasing acceptable product with some parameters unknown. *Technometrics*, 18, 195-199.

5. Owen, D. B., Chen, Y. M., and Li, L. (1979). Prediction intervals applied to screening variates. *Technical Report* No. 130. Department of Statistics, Southern Methodist University.

6. Owen, D. B., McIntire, D., and Seymour, E. (1975). Tables using one or two screening variables to increase acceptable product under one-sided specifications. *J. Qual. Tech.*, 7, 127-138.

7. Owen, D. B., and Su, Y. H. (1977). Screening based on normal variables. *Technometrics*, 19, 65-68.

8. Taylor, H. C., and Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *J. Applied Psychology*, 23, 565-578.

9. Thomas, J. G., Owen, D. B., and Gunst, R. F. (1977). Improving the use of educational tests as selection tools. *J. Educational Stat.*, 2, 55-77.