DEVELOPMENT AND ANALYSIS OF A MODIFIED SCREENING
PROCEDURE TO INCREASE ACCEPTABLE PRODUCT

by

Youn-Min Chou and D. B. Owen

Technical Report No. 148
Department of Statistics ONR Contract

August, 1981

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

DEVELOPMENT AND ANALYSIS OF A MODIFIED SCREENING PROCEDURE

TO INCREASE ACCEPTABLE PRODUCT

Youn-Min Chou

Division of Mathematics, Computer
Science and Systems Design
The University of Texas at San Antonio
San Antonio, Texas 78285

D. B. Owen

Department of Statistics
Southern Methodist University
Dallas, Texas 75275

Much work has been done on screening procedures under the
assumption of a bivariate normal distribution. However, very
little effort has been expended on data which are from a trun-
cated bivariate normal distribution. Methods are developed for
a screening procedure to increase acceptable product from a
truncated distribution. An acceptance criterion on a linear
combination of the largest order statistics from a truncated
normal population with a given truncation point is discussed.
This paper also uses the approximate distribution of the sample
correlation coefficient in random samples of any size drawn from
a singly truncated bivariate normal distribution to obtain a
lower confidence limit on the population correlation coefficient
ρ. The screening procedure discussed here is based on knowledge
of the truncation point, the sample size and the lower confidence
limit for ρ.

KEY WORDS AND PHRASES: Singly truncated bivariate normal distri-
bution; performance variable; screening
variable; acceptance sampling; truncated
normal distribution.

1. INTRODUCTION

In developing screening procedures, many methods can be utilized, depending on the data we have and the nature of the problem. Most of the previous work done in the area of screening procedures, e.g., Owen-Boddie(1976), Owen et. al.(1975 and 1977) , is based on data from a bivariate normal distribution, which is utilized to calculate the proportion successful after selection. In this paper we will consider screening procedures when the data available are from a truncated bivariate normal distribution.

A performance variable with a one-sided specification cannot be measured directly, but a related variable (called a screening variable) can be measured. In the language of acceptance sampling the performance variable may be lifetime or some other variable for which the act of measuring would degrade the item. If the quality control engineer only keeps records of those values of the performance variable and of the screening variable for the acceptable product, then the values of the performance variable must exceed some lower limit, say w_0. Hence, past data may only be available on a singly truncated bivariate normal distribution.

Let Y be a future performance variable and X be a future screening variable having a joint bivariate normal distribution with parameters $(\mu_x, \sigma_x^2, \mu_y, \sigma_y^2; \rho)$. Let W be the past screening variable which exceeds w_0, i.e., W has a truncated normal distribution with parameter (μ_x, σ_x^2). Let Z be the past performance variable and then (W,Z) follow a singly truncated bivariate normal distribution where only W is truncated. A one-sided lower specification limit, L, is given on the future performance

variable Y, i.e., all items with Y values above L are acceptable and those with Y values below L are not acceptable. Suppose that the proportion of acceptable items in the future is γ before screening. That is, the proportion above the lower specification before screening is γ. The screening procedure is set up to raise the proportion successful from γ to δ, where $\delta > \gamma$. Our procedure will be to accept all items for which $X \geqslant aW_{(n)} + bw_o$, where $b = 1 - a$ and $W_{(n)}$ is the largest order statistic from a truncated normal distribution (μ_x, σ_x^2) based on a sample of size n. After screening we want to be $100\eta\%$ sure that $P_{Y|X,W_{(n)}, \rho}\{Y \geqslant L | X \geqslant aW_{(n)} + bw_o\} \geqslant \delta$, i.e., the proportion of Y's greater than L is at least δ in the screened population. The reason for using $aW_{(n)} + bw_o$ instead of a linear combination of the sample mean $\bar{W}_n = \sum_{i=1}^{n} W_i/n$ and sample standard deviation S_W is that after truncating a normal distribution, many nice properties no longer exist and to our knowledge no manageable expression for the joint density function of $\bar{W}_n$ and S_W has yet been derived.

2. PROBABILITY EXPRESSION

Consider the case where the parameters μ_x, σ_x^2, μ_y and σ_y^2 are known and ρ is unknown. Let $P\{Y \geqslant L\} = \gamma$ and $P\{X \geqslant w_o\} = p$ be given. We make the transformations $Z_1 = (X - \mu_x)/\sigma_x$, $Z_2 = (Y - \mu_y)/\sigma_y$, $U = (W_{(n)} - w_o)/\sigma_x$, $V = Z_1 - aU$. Let $(L - \mu_y)/\sigma_y = -K_\gamma$ and $(w_o - \mu_x)/\sigma_x = -K_p$. The problem then becomes one of finding "a" such that $P[P\{Z_2 > -K_\gamma | V > -K_p\} \geqslant \delta] = \eta$ where the outer probability is with respect to the estimator of ρ and the inner probability is the conditional normal given the screening procedure and the correlation. Then

$\Delta = P\{Z_2 > -K_\gamma | V > -K_p\}$ can be written as

$$\Delta = \frac{a\int_0^\infty [G(x - K_p) - 1 + p]^n G'(-ax + K_p) G\left(\dfrac{a\rho x + K_\gamma - \rho K_p}{\sqrt{1-\rho^2}}\right) dx}{n\int_0^\infty [G(x - K_p) - 1 + p]^{n-1} G'(x - K_p) G(-ax + K_p) dx} . \qquad (2.1)$$

where $G'(z) = (2\pi)^{-\frac{1}{2}} \exp(-z^2/2)$ and $G(z) = \int_{-\infty}^z G'(t)dt$ for $-\infty < z < \infty$, are the univariate normal density and cumulative distribution, respectively.

<u>Theorem 2.1</u>   Under the assumptions given in sections 1 and 2, $\Delta = P_{Y|X, W_{(n)}, \rho}\{Y \geq L | X \geq aW_{(n)} + bw_o\}$ is an increasing function of ρ.

The proof of this theorem is obtained by showing that the numerator of equation (2.1) is a monotonically increasing function of ρ since the denominator is free of ρ and positive.

Suppose that we want to be $100\eta\%$ sure that $P_{Y|X,W_{(n)},\rho}\{Y \geq L \; X \geq aW_{(n)} + bw_o\} \geq \delta$, that is, that $P\{\Delta \geq \delta\} = \eta$. This is equivalent to $P\{\rho \geq \rho*\} \geq \eta$ for some $\rho*$. Our goal is to find a and $b(= -a + 1)$ such that

$$P_{Y|X,W_{(n)},\rho*}\{Y \geq L | X \geq aW_{(n)} + bw_o\} = \delta .$$

In order to solve for a, $\rho*$ has to be known. Once $\rho*$ is known, the problem which remains is to solve the following equation for a:

$$\frac{a\int_0^\infty [G(x - K_p) + p - 1]^n G'(-ax + K_p) G\left(\dfrac{a\rho*x + K_\gamma - \rho*K_p}{\sqrt{1 - \rho*^2}}\right) dx}{n\int_0^\infty [G(x - K_p) + p - 1]^{n-1} G'(x - K_p) G(-ax + K_p) dx} = \delta . \qquad (2.2)$$

Solutions to this equation will be discussed in section 5.

3. DISTRIBUTION OF THE SAMPLE CORRELATION COEFFICIENT

Let $(W_i, Z_i)(i = 1, \ldots, n)$ be a random sample of size n from the past record which follows a singly truncated bivariate normal distribution, where $W_i \geq w_o$, $i = 1, \ldots, n$. In the following discussion we consider the standardized singly truncated bivariate normal distribution, in which only W is truncated. Extension to the non-standardized case is straight-forward by using new variables $(\mu_x + \sigma_x W, \mu_y + \sigma_y Z)$ instead of (W,Z). Let R_T be the sample correlation coefficient. Let $F_{R_T}(.;\rho)$ and $F_R(.;\rho)$ be the respective distribution functions of R_T and R. Applying equation (32) of Gayen (1951), the c.d.f. of R_T is given by

$$F_{R_T}(r;\rho_T) \doteq F_R(r;\rho_T) + (C_1 L_{41} + C_2 L_{61}) \frac{\partial}{\partial \rho_T} F_R(r;\rho_T)$$

$$+ (C_1 L_{42} + C_2 L_{62}) \frac{\partial^2}{\partial \rho_T^2} F_R(r;\rho_T) + C_2 L_{63} \frac{\partial^3}{\partial \rho_T^3} F_R(r; \rho_T), \quad (3.1)$$

where $C_1 = \frac{n-1}{8n(n+1)}$ and $C_2 = \frac{n-2}{12n(n+1)(n+3)}$;

$$L_{41} = 3\rho_T(\lambda_{40} + \lambda_{04}) - 4(\lambda_{31} + \lambda_{13}) + 2\rho_T \lambda_{22},$$

$$L_{42} = \rho_T^2(\lambda_{40} + \lambda_{04}) - 4\rho_T(\lambda_{31} + \lambda_{13}) + 2(2 + \rho_T^2)\lambda_{22},$$

$$L_{61} = -15\rho_T(\lambda_{30}^2 + \lambda_{03}^2) - 9\rho_T(\lambda_{21}^2 + \lambda_{12}^2) + 12\lambda_{12}\lambda_{21}$$

$$+ 18(\lambda_{30}\lambda_{21} + \lambda_{03}\lambda_{12}),$$

$$L_{62} = -9\rho_T^2(\lambda_{30}^2 + \lambda_{03}^2) - 3(8 + 5\rho_T^2)(\lambda_{21}^2 + \lambda_{12}^2) + 36\rho_T \lambda_{21}\lambda_{12}$$

$$+ 30\rho_T(\lambda_{30}\lambda_{21} + \lambda_{03}\lambda_{12}),$$

$$L_{63} = -\rho_T^3(\lambda_{30}^2 + \lambda_{03}^2) - 3\rho_T(4 + \rho_T^2)(\lambda_{21}^2 + \lambda_{12}^2) + 4(2 + 3\rho_T^2)\lambda_{21}\lambda_{12}$$

$$+ 6\rho_T^2(\lambda_{30}\lambda_{21} + \lambda_{03}\lambda_{12});$$

ρ_T is the correlation coefficient between W and Z and the λ_{ij} are the semi-invariants of the singly truncated bivariate normal

distribution.

The results of the theoretical distribution of R_T were checked by comparison with a Monte Carlo simulation. For each of the sample sizes $n = 3, 15, 50$, truncation points $w_o = 0.$ $(-.5)$ $-3.$ and $\rho = -.90 (.10) .90$, 4000 values of R_T were generated. We made comparisons between the empirical and the theoretical c.d.f. of R_T based upon the Kolmogorov-Smirnov test and conclude that the approximation holds well.

4. A LOWER CONFIDENCE LIMIT ON ρ

As we have seen from Section 2, a $100\eta\%$ lower confidence limit ρ^* on ρ is required for our acceptance criterion $X \geq aW_{(n)} + (1 - a)w_o$ Let $F_{R_T}(.;\rho)$ be the c.d.f. of R_T. By the probability integral transformation theorem, $F_{R_T}(R_T;\rho)$ follows a uniform $(0,1)$ distribution. It follows that $P\{F_{R_T}(R_T;\rho) \leq \eta\} = \eta$. Let $g(\rho) = F_{R_T}(R_T;\rho) - \eta$; then $P\{g(\rho) \leq 0.\} = \eta$. From the inequality $g(\rho) \leq 0$, we would like to get an inequality $\rho \geq \rho^*$ so that ρ^* is a $100\eta\%$ lower confidence limit on ρ. To do this, we need R_T, n and w_o. Since the exact distribution of R_T is not known, we will use the approximate distribution of R_T given by equation (3.1). It can be shown that the function $g(\rho)$ is a decreasing function of ρ. For each given confidence coefficient η, ρ^* can be obtained by examining the root of the equation $g(\rho) = 0$.

Since $-1 < \rho^* < 1$, we can use the IMSL (1979) subroutine ZFALSE, i.e., the false position method, to find the root of $g(\rho) = 0$. Tables $1 - 4$ give the result of this computation for $n = 15, 50$ and $R_T = .40, .50.$

5. SCREENING CRITERION

For given η, n, R_T, γ, $\delta (> \gamma)$ and $w_o (= -K_p)$, a 100η% lower confidence limit ρ^* can be found using the method in Section 4. Thus a is the only unknown in equation (2.2). We use the Gauss Laguerre quadrature to approximate integrals of the form $\int_0^\infty e^{-x} h(x)\, dx$ with h(x) a function of x. The solutions can be found iteratively by using an algorithm due to Miller (1956), which finds the zeros of nonlinear functions. The screening criterion is $X \geq aW_{(n)} + (1-a)w_o$.

6. EXAMPLE

Let Y be the performance variable for some device which is expensive to measure. Assume X and Y have a joint standard bivariate normal distribution with unknown correlation coefficient. A sample of size n = 50 is taken from the past record in which all the performance scores are at least 0, and the highest performance score in this sample is $W_{(n)} = 2.6$. The sample correlation coefficient R_T is found to be .50. Suppose we want to be 95% sure that the proportion of acceptable items will be raised from $\gamma = .70$ to $\delta = .90$ after screening. We use linear interpolation in Table 4 and find $\rho^* = .45814$. Then we compute a from equation (2.2) and it is a = .37280. Thus our screening criterion is to accept all items for which $X \geq .96928$.

7. CONCLUSION

It has always been necessary to solve screening problems first by assuming all parameters known. Then estimates of the parameters based on a training set are used. In this paper we have assumed that μ_x, σ_x, μ_y, σ_y and w_o, the truncation point were known. Obviously there still exists the unsolved problem of what to do when any of these parameters are unknown.

REFERENCES

Gayen, A. K. (1951). "The Frequency Distribution of the Product-Moment Correlation Coefficient in Random Samples of any Size Drawn From Non-Normal Universes," *Biometrika*, 38, 219-247.

IMSL Library, Mathematical-Statistical Computer Subroutines, (1979). "Reference Manual," International Mathematical and Statistical Libraries.

Muller, D. E. (1956). A method for solving algebraic equations using an automatic computer. *Mathematical Tables and Computations*, 10, 208-215.

Owen, D. B. and Boddie, J. W. (1976). "A Screening Method for Increasing Acceptable Product with Some Parameters Unknown," *Technometrics*, 18, 195-199.

Owen, D. B. and McIntire, D. and Seymour, E. (1975). "Tables Using One or Two Screening Variables to Increase Acceptable Product Under One-Sided Specifications," *Journal of Quality Technology*, 7, 127-138.

Owen, D. B. and Su, Y. H. (1977). "Screening Based on Normal Variables," *Technometrics*, 19, 65-68.

TABLE 1

CONFIDENCE COEFFICIENT ETA WHEN N= 15 AND R_T= .40

RHO* \ W0	0.000	-0.500	-1.000	-1.500	-2.000	-2.500	-3.000
-.65	.99930	.99973	.99990	.99996	.99998	.99998	.99999
-.60	.99871	.99942	.99975	.99989	.99993	.99995	.99995
-.55	.99783	.99890	.99947	.99972	.99983	.99986	.99987
-.50	.99658	.99807	.99896	.99940	.99960	.99967	.99969
-.45	.99488	.99685	.99812	.99882	.99916	.99929	.99933
-.40	.99262	.99509	.99682	.99785	.99837	.99859	.99866
-.35	.98971	.99266	.99487	.99628	.99705	.99739	.99751
-.30	.98600	.98937	.99205	.99388	.99493	.99542	.99560
-.25	.98136	.98502	.98809	.99031	.99166	.99231	.99255
-.20	.97562	.97937	.98266	.98517	.98676	.98757	.98788
-.15	.96859	.97213	.97538	.97796	.97967	.98057	.98092
-.10	.96006	.96298	.96577	.96809	.96967	.97053	.97088
-.05	.94977	.95156	.95333	.95485	.95592	.95651	.95676
.00	.93745	.93745	.93745	.93745	.93745	.93745	.93745
.05	.92276	.92017	.91747	.91503	.91322	.91218	.91173
.10	.90533	.89921	.89269	.88667	.88214	.87950	.87835
.15	.88474	.87398	.86234	.85145	.84316	.83829	.83616
.20	.86049	.84387	.82568	.80852	.79539	.78765	.78426
.25	.83205	.80823	.78200	.75720	.73823	.72709	.72221
.30	.79880	.76639	.73071	.69710	.67157	.65669	.65023
.35	.76008	.71775	.67144	.62828	.59596	.57738	.56941
.40	.71518	.66179	.60415	.55147	.51285	.49109	.48194
.45	.66340	.59821	.52937	.46821	.42472	.40089	.39109
.50	.60409	.52709	.44833	.38108	.33517	.31089	.30121
.55	.53681	.44910	.36327	.29376	.24874	.22598	.21725
.60	.46151	.36579	.27756	.21090	.17053	.15122	.14415
.65	.37891	.28000	.19582	.13769	.10533	.09087	.08588
.70	.29103	.19616	.12357	.07892	.05655	.04734	.04439
.75	.20200	.12045	.06632	.03764	.02501	.02030	.01892
.80	.11897	.06007	.02780	.01367	.00834	.00657	.00611
.85	.05228	.02099	.00776	.00320	.00177	.00135	.00125
.90	.01252	.00362	.00098	.00031	.00015	.00011	.00010
.95	.00052	.00008	.00001	.00000	.00000	.00000	.00000

TABLE 2

CONFIDENCE COEFFICIENT ETA WHEN N= 15 AND R_T=.50

WØ RHO*	0.000	-0.500	-1.000	-1.500	-2.000	-2.500	-3.000
-.60	.99968	.99987	.99995	.99998	.99999	.99999	.99999
-.55	.99943	.99973	.99988	.99995	.99997	.99997	.99998
-.50	.99906	.99951	.99975	.99987	.99992	.99993	.99994
-.45	.99853	.99915	.99953	.99972	.99981	.99984	.99985
-.40	.99780	.99861	.99915	.99945	.99960	.99966	.99968
-.35	.99682	.99783	.99855	.99899	.99922	.99932	.99935
-.30	.99553	.99672	.99763	.99823	.99857	.99872	.99877
-.25	.99384	.99519	.99628	.99705	.99751	.99772	.99780
-.20	.99167	.99311	.99434	.99525	.99582	.99610	.99621
-.15	.98891	.99032	.99159	.99259	.99323	.99357	.99370
-.10	.98543	.98664	.98779	.98872	.98936	.98970	.98983
-.05	.98105	.98182	.98259	.98324	.98369	.98394	.98405
.00	.97558	.97558	.97558	.97558	.97558	.97558	.97558
.05	.96877	.96754	.96625	.96507	.96419	.96369	.96347
.10	.96032	.95726	.95396	.95088	.94852	.94714	.94653
.15	.94985	.94421	.93796	.93200	.92737	.92461	.92338
.20	.93690	.92771	.91734	.90727	.89935	.89459	.89247
.25	.92091	.90698	.89104	.87538	.86298	.85550	.85216
.30	.90117	.88106	.85785	.83493	.81677	.80582	.80094
.35	.87681	.84886	.81647	.78454	.75938	.74433	.73767
.40	.84679	.80908	.76555	.72302	.68995	.67044	.66192
.45	.80980	.76033	.70386	.64966	.60843	.58460	.57438
.50	.76432	.70115	.63052	.56463	.51605	.48878	.47738
.55	.7C858	.63025	.54541	.46946	.41585	.38688	.37515
.60	.64068	.54680	.44974	.36768	.31299	.28485	.27393
.65	.55885	.45111	.34681	.26519	.21472	.19033	.18134
.70	.46204	.34561	.24277	.17022	.12953	.11133	.10505
.75	.35124	.23630	.14697	.09215	.06500	.05398	.05049
.80	.23196	.13419	.07076	.03845	.02480	.01988	.01849
.85	.11814	.05482	.02310	.01049	.00612	.00474	.00440
.90	.03442	.01150	.00357	.00128	.00065	.00049	.00045
.95	.00197	.00037	.00006	.00001	.00000	.00000	.00000

TABLE 3

CONFIDENCE COEFFICIENT ETA WHEN N= 50 AND R_T=.40

WO RHO*	0.000	-0.500	-1.000	-1.500	-2.000	-2.500	-3.000
-.15	.99978	.99985	.99990	.99993	.99994	.99995	.99995
-.10	.99954	.99964	.99971	.99977	.99980	.99982	.99983
-.05	.99907	.99917	.99925	.99933	.99937	.99940	.99941
.00	.99819	.99819	.99819	.99819	.99819	.99819	.99819
.05	.99658	.99624	.99585	.99548	.99519	.99502	.99494
.10	.99375	.99252	.99104	.98952	.98827	.98749	.98714
.15	.98890	.98571	.98168	.97734	.97363	.97125	.97015
.20	.98079	.97376	.96452	.95424	.94525	.93945	.93674
.25	.96762	.95366	.93488	.91371	.89517	.88322	.87766
.30	.94681	.92130	.88684	.84829	.81509	.79405	.78439
.35	.91490	.87166	.81419	.75186	.70013	.66843	.65422
.40	.86759	.79953	.71256	.62342	.55383	.51331	.49579
.45	.80016	.70106	.58263	.47124	.39158	.34840	.33061
.50	.70852	.57636	.43333	.31410	.23842	.20106	.18660
.55	.59120	.43246	.28280	.17677	.11921	.09398	.08494
.60	.45229	.28499	.15411	.07923	.04596	.03339	.02930
.65	.30434	.15589	.06530	.02606	.01253	.00828	.00704
.70	.16862	.06487	.01934	.00557	.00213	.00126	.00104
.75	.06880	.01780	.00338	.00064	.00018	.00010	.00008
.80	.01681	.00250	.00026	.00003	.00001	.00000	.00000
.85	.00159	.00011	.00000	.00000	.00000	.00000	.00000
.90	.00002	.00000	.00000	.00000	.00000	.00000	.00000

TABLE 4.

CONFIDENCE COEFFICIENT ETA WHEN N= 50 AND R_T=.50

WØ RHO*	0.000	-0.500	-1.000	-1.500	-2.000	-2.500	-3.000
.05	.99980	.99978	.99975	.99972	.99970	.99968	.99968
.10	.99958	.99947	.99934	.99919	.99907	.99899	.99895
.15	.99913	.99880	.99835	.99783	.99736	.99704	.99689
.20	.99824	.99736	.99609	.99452	.99303	.99200	.99150
.25	.99651	.99440	.99115	.98699	.98293	.98010	.97871
.30	.99322	.98847	.98092	.97103	.96130	.95450	.95117
.35	.98707	.97704	.96085	.93966	.91902	.90481	.89794
.40	.97580	.95586	.92387	.88295	.84440	.81870	.80659
.45	.95568	.91838	.86039	.78991	.72734	.68773	.66974
.50	.92079	.85563	.76035	.65389	.56744	.51663	.49471
.55	.86259	.75765	.61852	.48119	.38279	.33047	.30936
.60	.77053	.61803	.44295	.29766	.20976	.16873	.15352
.65	.63547	.44232	.26144	.14295	.08562	.06297	.05542
.70	.45827	.25695	.11496	.04737	.02297	.01516	.01287
.75	.26263	.10631	.03200	.00900	.00334	.00194	.00160
.80	.09975	.02473	.00425	.00071	.00019	.00010	.00008
.85	.01699	.00200	.00015	.00001	.00000	.00000	.00000
.90	.00047	.00002	.00000	.00000	.00000	.00000	.00000