ONE-SIDED SCREENING PROCEDURE USING MULTIPLE NORMALLY DISTRIBUTED VARIABLES

by

Lazar Boskov

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Industrial and Systems Engineering

APPROVED:

Joel A. Nachlas, Chairman

Golde I. Holtzman

John E. Kobza

August, 1994

Blacksburg, Virginia

ONE-SIDED SCREENING PROCEDURE USING MULTIPLE NORMALLY DISTRIBUTED VARIABLES

by

Lazar Boskov

Joel A. Nachlas, Chairman

Industrial and Systems Engineering

(ABSTRACT)

In the situations in which the proportion of acceptable products from the output of a production process is below the required level, it is essential to screen out the products of unacceptable quality. By eliminating the products of low quality, the proportion of products of the acceptable quality in the remaining population of products is raised.

In some instances it is not possible to make quality assessment by measuring directly on the variable of interest (performance variable). The measure is not possible because it destroys or degrades the product. In such cases, auxiliary variables which are correlated with the performance variable can be used to indirectly determine the quality of the product. These variables are called screening variables.

Under the assumption that the performance and the screening variables follow a

multivariate normal distribution, a regression model is used to predict the value of the performance variable. Using Monte Carlo simulation, the performance of the regression model is evaluated. The evaluation is done for two cases: (1) the parameters of the underlying distribution are known, (2) the parameters are not known. The results show that the efficiency of the screening depends highly on the value of the correlation coefficient between the performance variable and a linear combination of the screening variables. Furthermore, the comparison with the previously developed models is performed.

Findings show that the regression model is very useful, especially in the cases in which multiple screening variables are available and the parameters of the underlying distribution are not known.

ACKNOWLEDGMENTS

I would like to thank Dr. Joel A. Nachlas, my advisor, for all of his help, guidance and time. I would also like to thank Dr. Golde I. Holtzman and Dr. John E. Kobza for serving on my committee.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

1. 1. Introduction

A screening procedure is a statistical quality control method used in industrial production to increase the proportion of products meeting defined quality criteria by eliminating nonconforming products. These quality criteria may be expressed as the values of a performance variable. Often the product's performance characteristic is either too expensive to measure or it cannot be measured directly. The lifetime of a product and the hardness of steel are instances in which the performance variable cannot be directly measured because the measurement would cause destruction or degradation of the product. On the other hand, some internaly located performance characteristics such as the properties of an electronic component in an electric circuit can be measured without destroying the product but are too expensive to measure. In these instances, screening variables which are correlated with the performance variable can be used to select products of acceptable quality.

The screening variable is a measurable characteristic of a product which indirectly indicates the quality of the product as expressed by the performance variable. Through 100% inspection performed on the screening variable, selecting only the products within given a specification of the screening variable, the proportion of acceptable products in the population before screening is raised to a higher level in the reduced population. Screening does not improve the quality of products. It just increases the proportion of acceptable products by reducing the size of the total population. The specification for the screening variable depends on the required level for the proportion of acceptable products in the reduced population. The products not selected could be repaired, sold at a discount price or scrapped.

In this chapter, the screening procedure in the ideal situation is described and the essential conditions for the success of the screening procedure are given. Also, difficulties encountered in the practical applications of the screening as well as the relevance of the statistical tolerance intervals are discussed. In conclusion, a research problem is formulated.

1.2. Background Review

1.2.1. Screening Procedure - Ideal Situation

Ideally, the measurements on screening variables are error free and the total population before screening is infinite. Since all the products are inspected and nonconforming products are eliminated, output from the production process is of constant quality at the level δ which is higher than before screening.

To illustrate a screening procedure, the following assumptions are needed:

The first assumption is that the performance variable Y is normally distributed with known mean μ_Y, and standard deviation σ_Y. However, it should be noted that if the proportion, γ, of acceptable products in the whole population is known, then μ_Y and σ_Y are not needed.

Secondly, all products with performance variable values above a given lower specification limit L_Y are considered acceptable.

In addition, a single screening variable X is available. Suppose further that the screening variable X is normally distributed with known mean, μ_X, and standard deviation, σ_X. The correlation coefficient, ρ_{YX}, is positive and known also.

Under these assumptions, the following equalities hold (Owen, McIntire and Seymour, 1975):

i) The proportion of acceptable products in the whole population before screening is:

$$P(Y > L_Y) = P(Y > \mu_Y - K_\gamma \sigma_Y) = \gamma \qquad (1.1.)$$

ii) The proportion of products obtained by selecting the products satisfying the criterion for the screening variables is:

$$P(X > L_X) = P(X > \mu_X - K_\psi \sigma_X) = \psi \qquad (1.2)$$

iii) The proportion of acceptable products in the selected group after screening is:

$$P(Y > \mu_Y - K_\gamma \sigma_Y / X > \mu_X - K_\psi \sigma_X; \gamma, \rho_{YX}) = \delta > \gamma \quad (1.3)$$

In order to acquire a desired proportion δ of acceptable products in the screened population, the size of the original population should be reduced by a factor ψ.

The value of ψ is obtained from the widely available tables of the normal distribution with entries δ, γ, and correlation coefficient, ρ_{YX} (eg. Odeh and Owen, 1980). When the ψ is known, the $100\psi\%$ point K_ψ is found in the tables of the standard normal distribution and a lower limit for the screening variable is:

$$L_X = \mu_X - K_\psi \sigma_X \quad (1.4)$$

Thus, by selecting only the products for which the value of screening variable is above its lower limit the proportion ψ of products is selected.

In the proportion of the selected products ψ, which is obtained by measurement of the screening variable as explained above, $100\delta\%$ of products have performance variable values above desirable specification limit L_Y. On the other hand, $100(1\text{-}\delta)\%$ of the products have values of the performance variable below L_Y. This means that these products are incorrectly selected as acceptable. In the proportion of products not selected by measurement of the screening variable $1\text{-}\psi$, $100\frac{\gamma\text{-}\delta\psi}{1\text{-}\psi}\%$

of the products have values of the performance variable above L_Y (i.e. incorrectly screened products), and $100\frac{1-\gamma-\psi+\delta\psi}{1-\psi}\%$ of products have performance variable values below L_Y (i.e. correctly selected products).

If the correlation between the performance variable Y and the screening variable X is negative and/or an upper specification limit is given (the performance variable Y has to be less than some prespecified value U_Y) as well as in the combinations of these two cases, the criterion for selection must be adjusted as stated in Owen, McIntyre, and Seymour (1975). For example, if the correlation between the performance variable and the screening variable is negative the criterion for selection is:

$$X < \mu_X + K_\psi \sigma_X \tag{1.5}$$

1.2.2. Essential Conditions for a Success of the Screening Procedure

When performance and screening variables are jointly normally distributed, the success of the screening depends on the value of correlation coefficient ρ_{YX}, and on the proportion of selected products ψ, using the screening variable X as discussed in Kocherlakota, Kocherlakota, and Balakrishnan (1987):

i) If $\rho_{YX} = 0$, then

$$P(Y > L_Y / X > L_X) = \frac{P(Y > L_Y, X > L_X)}{P(X > L_X)} \tag{1.6}$$

$$= \frac{P(Y > L_Y)P(X > L_X)}{P(X > L_X)} \tag{1.7}$$

$$= P(Y > L_Y) = \gamma \tag{1.8}$$

Therefore, as the screening variable X does not provide any information regarding performance variable Y, the screening procedure will not be effective.

ii) If $\rho_{YX} > 0$, then by the inequality given in Tong (1990)

$$P(Y > L_Y, X > L_X) > P(Y > L_Y)P(X > L_X) = \gamma\psi \tag{1.9}$$

Hence,

$$P(Y > L_Y/X > L_X) = \frac{P(Y > L_Y, X > L_X)}{P(X > L_X)} > \frac{\gamma\psi}{\psi} = \gamma \tag{1.10}$$

resulting in $\delta > \gamma$.

In this instance, the proportion of acceptable products is higher than the proportion without screening which means that the screening procedure will be effective.

iii) If $\rho_{YX} = 1$, then given $X > L_X$, it follows that $P(Y > L_Y) = 1$. Hence

$$P(Y > L_Y, X > L_X) = P(X > L_X) = \gamma \tag{1.11}$$

$$P(Y > L_Y / X > L_X) = \frac{P(Y > L_Y, X > L_X)}{P(X > L_X)} = \frac{\gamma}{\gamma} = \delta = 1 \tag{1.12}$$

Accordingly, if the performance variable Y and a screening variable X are perfectly correlated, the screening variable X provides all the needed information regarding performance variable Y and the screening will be 100% successful.

iv) If $\psi \to 1$, then $\delta \to \gamma$.

In this case the screening will have no practical purpose as all the products are selected.

1.2.3. Difficulties in Practice

The screening methodology developed under the ideal conditions cannot be applied when all or some of the parameters of the underlying distributions of the performance variable Y and the screening variable X are not known. Depending on which of the parameters from either of the following two groups: μ_X, σ_X, ρ_{YX}, γ, or μ_X, σ_X, μ_Y, σ_Y, ρ_{YX}, are not available, the following proportions are random variables with their own distributions:

i) The proportion of acceptable products in the whole population $P(Y \geq L_Y) = \gamma$

and the proportion of acceptable products $P(Y > L_Y/X > L_X)=\delta$ in the selected group after screening (unknown γ, or any or both parameters μ_Y, σ_Y);

ii) The proportion of products selected by the measurement on the screening variable $P(X > L_X)=\psi$ and the proportion of acceptable products $P(Y > L_Y/X > L_X)=\delta$ in the selected group after screening (any or both parameters μ_X, σ_X, is unknown);

iii) The proportion of acceptable products in the whole population $P(Y > L_Y)=\gamma$, the proportion of products selected by the measurement on the screening variable $P(X > L_X)=\psi$, and the proportion of acceptable products $P(Y > L_Y/X > L_X)=\delta$ in the selected group after screening (any or all of the parameters from any of the following groups γ, μ_X, μ_Y; γ, μ_X, σ_Y; γ, σ_X, μ_Y; γ, σ_X, σ_Y; γ, σ_X, σ_Y).

iv) The proportion of acceptable products $P(Y > L_Y/X > L_X)=\delta$ in the selected group after screening (ρ_{YX} unknown).

In these situations, it is appropriate to construct a statistical tolerance interval which will cover at least a required proportion of acceptable products with a specified degree of confidence. The following example (Owen and Boddie, 1976) illustrates the application of the statistical tolerance interval in the screening procedures. If μ_X and σ_X are not known the proportion of acceptable products after screening

$$P(Y > L_Y = \mu_Y - K_\gamma \sigma_Y / X > \hat{L}_X = \bar{x} - k_1 s_X) = \delta \tag{1.13}$$

is a random variable and the desired interval at the confidence level η is of the form:

$$P(P(Y > L_Y = \mu_Y - K_\gamma \sigma_Y / X > \hat{L}_X = \bar{x} - k_1 s_X) = \delta) = \eta. \tag{1.14}$$

The factor k_1 depends upon the sampling errors in $\bar{x}$ and s_X as well as the population fraction δ and confidence level η. The factor k_1 can be obtained from the quantity $\sqrt{n}k_1$ which is the $100\eta\%$ point of the noncentral t-distribution with n-1 degrees of freedom and noncentrality parameter $K_\psi\sqrt{n}$. The coefficient K_ψ is the $100\psi\%$ point of the standard normal distribution. ψ is the proportion to be selected by the screening procedure from the whole population so that the proportion of acceptable products in the selected group is increased from γ to δ. Therefore, using the factor k_1, the proportion of acceptable products δ is covered by the tolerance interval at the confidence level η.

1.2.4. Statistical Tolerance Intervals

Formally, a statistical tolerance interval can be defined as follows: A statistical tolerance interval $S(Y_1, Y_2, \ldots, Y_n)$ is a random interval defined over measurable space $(\mathfrak{Y}, \mathfrak{U})$. The sample space of Y (that is S) is a random set function (or statistic) which is a measurable transformation from $(\mathfrak{Y},\mathfrak{U})$ into some $(\mathfrak{S},\mathfrak{B})$ with probability measure P defined over it (Billingsley, 1986; Guttmann, 1970; Lehman, 1991).

The above definition asserts that S is both a function defined over a sample space $(\mathfrak{Y},\mathfrak{U})$ and a random variable since it has a distribution over $(\mathfrak{S},\mathfrak{B})$. The term "statistic" indicates that the tolerance interval is a function of random variable Y.

A two-sided statistical tolerance interval constructed using n independent observations $Y_1,Y_2,..,Y_n$, is:

$$S(Y_1,Y_2,...,Y_n) = (L_1(Y_1,Y_2,...,Y_n),\ L_2\ (Y_1,Y_2,...,Y_n)) \tag{1.15}$$

In some cases L_1 or L_2 may not be finite. For example, the lifetime of the product must be greater than some prespecified length of time W_1. Then the statistical tolerance interval will be one-sided:

$$S(Y_1,Y_2,..,Y_n) = (L_1(Y_1,Y_2,...,Y_n),\ \infty) \tag{1.16}$$

Statistical tolerance limits can be selected in accordance with the following criteria:

i) Statistical tolerance interval of δ - content at confidence level η

$$P(P[S(Y_1,Y_2,...,Y_n)] \geq \delta) \geq \eta \qquad (1.17)$$

A δ-content tolerance interval $S(Y_1,Y_2,...,Y_n)$ contains at least $100\delta\%$ of the population with the confidence level of at least η. If the statistical tolerance interval is one sided, the criterion is:

$$P[\int_{L_1}^{\infty} f_Y(y)dy \geq \delta] \geq \eta \qquad (1.18)$$

where $f_Y(y)$ is the probability density function. Therefore, L_1 must be determined so that at least $100\delta\%$ of the population is contained in the tolerance interval at the confidence level of at least η.

The probability content, $P[S(Y_1,Y_2,...,Y_n)]$ of the tolerance region $S(Y_1,Y_2,...,Y_n)$ which is based on n independent observations from the probability distribution of Y is called the coverage of $S(Y_1,Y_2,...,Y_n)$, abbreviated as $C(S)$. $C(S)$ is simply the proportion of the population being sampled that is "covered" by S (Guttman, 1970).

ii) Statistical tolerance interval of δ - expectation

$$E(P[S(Y_1,Y_2,...,Y_n)]) = E[C(S)] = \delta \qquad (1.19)$$

If the statistical tolerance interval is of δ-expectation, then on the average, a proportion δ of the population being sampled will be covered by $S(Y_1,Y_2,...,Y_n)$.

Furthermore, the tolerance region of δ-expectation is identical to prediction interval for a single future observation Y_0 (Guttman, 1970):

$$E[C(S)] = \delta = P(L_1(Y_1,Y_2,...,Y_n) \leq Y_0 \leq L_2(Y_1,Y_2,...,Y_n)) \tag{1.20}$$

In this work, the screening procedure is applied to select large numbers of products so that the proportion of acceptable products after screening is δ. Therefore, all tolerance intervals are of the first type.

It should be noted that the first criterion is more restrictive. For example, for normally distributed random variables an approximate confidence level η that the δ-expectation tolerance interval has a coverage which exceeds δ is between 0.5 and 0.55 (Guttmann, 1970, p. 76). This is because the variance of $C(S)$ is an important factor in constructing δ-content tolerance interval while the variability of $C(S)$ plays no role in constructing δ-expectation tolerance region.

1.3. Problem Statement

The success of the screening procedure is directly related to the strength of correlation between the performance variable Y and the screening variable X (Tang and Tang, 1989). Accordingly, errors of rejecting products of good quality and accepting products of an inferior quality will be reduced if the strength of correlation is increased. The strength of the correlation can be increased by using more than one screening variable.

Two questions arise if several candidates for screening variables are available:

i) How to combine selected screening variables ?

ii) How to select screening variables which will give the most information regarding performance variable Y?

Moreover, a problem occurs if any of the parameters of the underlying distributions are not known.

In this research, a regression model is applied to select the smallest subset from the set of available screening variables and to estimate the linear combination of selected variables. When more than one screening variable is available, the regression model maximizes the correlation between the linear combination of screening variables and performance variable. For a given proportion of acceptable items in the whole population γ and for a given proportion of selected items ψ, the stronger correlation leads to improved screening accuracy resulting in lower errors of accepting defective and rejecting nondefective products.

The case in which the parameters of the underlying multivariate distribution are not known is considered. Next, the cut-off point for the combination of selected screening variables is given so that the proportion of acceptable products is increased from γ to δ with confidence level of some prespecified η.

1.4. Summary of the Results

A regression model is used to screen the products using auxiliary variables. The model is applied to the case in which the parameters of the underlying multivariate normal distribution are known and to the case in which the parameters are not known. The one-sided tolerance interval developed by Liberman and Miller (1963) is used for selecting the products in the cases in which the parameters of the underlying multivariate normal distributions are not known. If the value of its lower (upper) tolerance limit is not below (above) the acceptance limit, the product is selected.

A simulation study is performed in order to evaluate and compare the regression model with the two models developed by Owen, Seymour and McIntyre (1975) and with the model of Odeh and Owen (1980). The first two models, which cover the case in which the parameters of the underlying multivariate distribution are known, are applicable in the situations in which one and two screening variables are available respectively. The model of Odeh and Owen covers the case in which parameters are not known and a single screening variable is available.

In the case in which two screening variables are available and the parameters of the trivariate normal distribution are not known, the simulation study is used to examine the efficiency of the screening procedure when the regression model is applied. The two screening variables case is chosen to show how the strength of correlation between the performance variable and a linear combination of screening variables can be increased and the benefits of doing so.

In the simulation study, 300 samples of size 600 each are generated from the bivariate and trivariate normal distributions. The values of the variances and the covariances are selected so that the value of the correlation coefficient between the performance variable and the single screening variable or a linear combination of screening variables is in the range between 0.5 and 0.98.

The results of the analysis show that no screening procedure can be recommended for practical applications if the value of the correlation coefficient is below 0.8 and the parameters are known. If the parameters are not known, the correlation coefficient should be at least 0.95.

The regression method has the advantage over other models as it allows any number of the screening variables to be used. By using multiple screening variables the value of the correlation coefficient is increased and therefore the efficiency of screening is increased.

CHAPTER 2 - LITERATURE REVIEW

All of the previous work in the area of screening using correlated variables is concerned with developing procedures for the cases in which all parameters are known or for the cases in which only one screening variable is available. The case in which multiple screening variables are available and the parameters are not known is not considered in any of the references and to the best of the knowledge of the author, in any other sources.

Owen, McIntyre, and Seymour (1975) give procedures for two cases:

i) One screening variable is available.

ii) Two screening variables are available.

The underlying distribution is assumed to be bivariate normal in the first and trivariate normal in the second case. All parameters of the distributions and the proportion of acceptable items before screening γ are assumed to be known. Tables are provided for various values for δ, γ, and ρ and give the proportions of the population to be in the selected group. The proportion of acceptable items in the selected group is then guaranteed to be δ $(\delta > \gamma)$. For the case in which one screening variable is available, values of the performance variable Y above a lower specification limit are considered acceptable and the correlation coefficient ρ_{XY} is positive. The criterion for screening is to accept all items for which $X > \mu_X - K_\psi \sigma_X$, where K_ψ is the $100\psi\%$ point of the standard normal distribution.

Two procedures are given for the case in which the performance variable Y and two screening variables, X_1 and X_2 are available. In each, the performance variable values above a lower specification limit are considered acceptable and all correlations are positive.

1. The first procedure determines the proportion ψ to be selected to achieve a prespecified proportion of acceptable items in the selected group for both screening variables X_i, i=1,2 using the tables in Owen et al. (1975). Next, all items for which both $X_1 > \mu_{X_1} - K_\psi \sigma_{X_1}$ and $X_2 > \mu_{X_2} - K_\psi \sigma_{X_2}$ are accepted. K_ψ is the $100\psi\%$ point of the standard normal distribution. Correlations between Y and X_1 and between Y and X_2 are assumed to be the same or nearly the same. The actual percentage of accepted products is $\geq 100\psi^2\%$ depending on the strength of correlation between X_1 and X_2.

2. The second procedure involves a linear combination of the screening variables, $V = a_1 X_1 + a_2 X_2$, where a_1 and a_2 are computed to maximize the proportion of acceptable product after screening. The procedure is the same as that for one screening variable with V replacing X.

These procedures cannot be extended to the case in which the parameters of the underlying distribution are not known. Also, the method for the selection of a subset of the screening variables, if several are available, is not developed. The second procedure is also given in Thomas, Owen and Gunst (1977).

Kocherlakota, Kocherlakota and Balakrishnan (1987) consider a selection procedure with multiple screening variables in which the underlying distribution is assumed to be multivariate normal and all parameters of the distribution are known. Their approach is to calculate a correlation coefficient ρ_{Y1}, between the first principal component of the screening variables $X_1, X_2, \ldots, X_n$ and the performance variable Y. Next, they use the tables based on bivariate normal distribution, from Taylor and Russell (1939), with entries ρ_{Y1}, γ and δ to determine the proportion of products to be selected ψ so that the proportion of acceptable products in the reduced population is δ. Knowing the proportion which needs to be selected ψ, the cut-off point for the confidence limit for the first principal component is determined by multiplying the 100ψ percentile of the standard normal distribution with the standard error of the first principal component. The decision rule is to reject all products for which the value of the first principal component is outside its confidence limits.

If more than one screening variable is available, they consider the selection of screening variables. The selection is based on the strength of correlation between the performance variable and the first principal component with and without each candidate screening variable.

The procedure cannot be applied if any of the parameters of the distributions of the performance and screening variables are not known. In the case in which all parameters are known, each principal component follows normal distribution, $Z_i = e_i' X \sim N(e_i'\mu,\ e_i'\Sigma\, e_i = \lambda_i)$ where λ_i and e_i, are an eigenvalue - eigenvector pair of variance-covariance matrix Σ of random vector X. Nevertheless, if Σ is estimated from the data, the distribution of $\hat{Z}_i$ is not known and the confidence interval for $\hat{Z}_i$ cannot be computed. Therefore, the tolerance limits for the proportion of acceptable products cannot be calculated.

Several papers consider the situations in which one screening variable is available and some or all parameters of the underlying bivariate normal distribution are not known. Owen and Boddie (1976) give procedures for approximations based on a method for constructing a tolerance interval of δ-expectation developed by Wilks (1941) and procedures based on a method for constructing a tolerance region of δ-content.

The procedures based on Wilks' idea are developed for three cases:

i) The mean μ_X of the population of screening variables is not known.

ii) Both the mean μ_X and the standard deviation σ_X of the population of screening variables are not known.

iii) The mean μ_X and the standard deviation σ_X of the population of screening variables and the mean μ_Y and the standard deviation σ_Y of the population of the performance variable are not known. The only known parameter is correlation coefficient ρ_{YX}.

In all cases, the proportion of acceptable items before screening γ is assumed to be known. If the mean μ_X is unknown, the procedure calls for $k_1 = K_\psi\sqrt{\frac{n+1}{n}}$. Furthermore, if both the mean μ_X and the standard deviation σ_X are unknown, $k_2 = t_\psi\sqrt{\frac{n+1}{n}}$, where k_1 is:

$$E[P(Y > \mu_Y - K_\gamma \sigma_Y / X > \bar{x} - k_1 \sigma_X;\ \rho_{YX})] = \delta \tag{2.1}$$

and k_2 is:

$$E[P(Y > \mu_Y - K_\gamma \sigma_Y / X > \bar{x} - k_2 s_X;\ \rho_{YX})] = \delta \tag{2.2}$$

K_ψ and t_ψ are the $100\psi\%$ percentile of the standard normal and t-distribution with n-1 degrees of freedom, respectively. If all parameters except ρ_{YX} are estimated, the procedure gives $k_1^* = \sqrt{\frac{n+1}{n}}\, t_{2n-2,\delta\psi;\rho}$, and $k_2^* = \sqrt{\frac{n+1}{n}}\, t_{2n-2,\psi}$, where k_1^* and k_2^* are solutions to:

$$E[P(X > \bar{x} - k_2^* s_p)] = \psi, \text{ and} \tag{2.3}$$

$$E[P(Y > \bar{y} - k_1^* s_p,\ X > \bar{x} - k_2^* s_p;\ \rho_{YX})] = \delta\psi \tag{2.4}$$

$t_{2n\text{-}2,\psi}$ is the $100\psi\%$ percentile of the t-distribution with $2n$-2 degrees of freedom, $t_{2n\text{-}2,\delta\psi;\rho}$ is the $100\delta\psi\%$ percentile of the bivariate t-distribution with $2n$-2 degrees of freedom, and s_p is the pooled variance as it is assumed that the variances of X and Y are same.

Owen and Boddie (1976) develop two procedures based on a method for constructing δ-content tolerance regions. The following cases are considered:

i) Both the mean μ_X and the standard deviation σ_X of the population of screening variables are not known.

ii) The mean μ_X and the standard deviation σ_X of the population of screening variables and the mean μ_Y and the standard deviation σ_Y of the population of performance variable are not known. The only known parameter is the correlation coefficient ρ_{YX}.

Again, the proportion of acceptable items before screening is assumed to be known. Two solutions are given for the first case. The first solution employs the noncentral t-distribution and the second employs the χ^2 distribution. A solution for the second case requiring equal variances is based on the bivariate noncentral t-distribution.

Owen and Su (1977) consider the case in which the proportion of acceptable products in the unscreened population γ and parameters μ_Y, μ_X, σ_Y, are not known and the case in which the parameters ρ_{YX}, μ_Y, μ_X, and σ_Y are not known. In both cases, the underlying distributions are assumed to be bivariate normal. A conservative solution guaranteeing that the proportion of acceptable items in the screened population is at least δ is given. The solutions are based on the lower one-sided tolerance limit γ^* for unknown γ, and on the lower one-sided confidence limit for positive correlation coefficient ρ_{YX}, or upper one-sided confidence limit ρ^* for negative ρ_{YX}. Tables from Owen, McIntyre and Seymour (1975) are used with estimates instead of unknown parameters. This solution cannot be extended to cover the case where both γ and ρ_{YX} are not known.

Odeh and Owen (1980) extend the procedure to cover the case in which both γ and ρ_{YX} are not known. The joint confidence statement

$$P(\gamma > \gamma^* \text{ and } \rho_{YX} > \rho^*) \geq 2\xi - 1 \qquad (2.5)$$

for both $P(\gamma > \gamma^*)=\xi$, and $P(\rho > \rho^*)=\xi$, by the Bonferroni inequality ensures that the proportion of acceptable products after screening is at least δ at the confidence level of at least 2ξ-1. Measured by the number of items rejected during the screening 1-ψ, the approach can yield extremely conservative results (Mee, 1990).

Mee (1990) calculates an approximate lower bound for

$$P(Y > \mu_Y - z_\gamma \sigma_Y / X > \mu_X - z_\psi \sigma_X) = \delta \qquad (2.6)$$

in the case in which all parameters are unknown. From the sample of n (Y_i, X_i) pairs, estimates of the unknown parameters are calculated. A lower one-sided $100\eta\%$ interval is computed using the noncentral t-distribution with n-1 degrees of freedom.

Jensen and Moltoft (1987) propose the use of Fisher's sample linear discriminant function for selecting products of good quality from the total population of products. The assumption is that good and bad products come from two populations π_1 and π_2 and that the screening variables $X_1, X_2, \ldots, X_K$ follow multivariate distributions with different mean vectors $\boldsymbol{\mu}_{\pi_1}$ and $\boldsymbol{\mu}_{\pi_2}$ depending on the population to which a product belongs and the same variance-covariance matrix Σ.

The Fisher's linear discriminant function $Z=\boldsymbol{t}'\boldsymbol{X}$ with mean $\mu_{\pi_1 Z}=E(Z/\pi_1)=\boldsymbol{t}'\boldsymbol{\mu}_{\pi_1}$ or $\mu_{\pi_2 Z}=E(Z/\pi_2)=\boldsymbol{t}'\boldsymbol{\mu}_{\pi_2}$, and variance $\sigma_Z^2=\boldsymbol{t}'\Sigma\boldsymbol{t}$, converts multivariate populations of screening variables into univariate populations in such way that the means $\mu_{\pi_1 Z}$ and $\mu_{\pi_2 Z}$ are separated as much as possible. If the parameters of multivariate distributions are known, the vector of coefficients in a linear combination that gives maximum separation is $\boldsymbol{t}'=(\boldsymbol{\mu}_{\pi_1}-\boldsymbol{\mu}_{\pi_2})\Sigma^{-1}$. Upon taking a sample $X_{10}, X_{20}, \ldots,$ X_{K0}, the decision rule is

classify product to π_1 (good) if $\boldsymbol{t}'\boldsymbol{X}_0 > \frac{1}{2}(\mu_{\pi_1 Z} + \mu_{\pi_2 Z})$ (2.7)

classify product to π_2 (bad) if $\boldsymbol{t}'\boldsymbol{X}_0 < \frac{1}{2}(\mu_{\pi_1 Z} + \mu_{\pi_2 Z})$ (2.8) if the product is acceptable for values of the performance variable above the lower limit and

classify product to π_1 (good) if $t'X_0 < \frac{1}{2}(\mu_{\pi_1 Z} + \mu_{\pi_2 Z})$ (2.9)

classify product to π_2 (bad) if $t'X_0 > \frac{1}{2}(\mu_{\pi_1 Z} + \mu_{\pi_2 Z})$ (2.10)

if the product is acceptable for values of the performance variable below upper limit.

The probability of misclassification (the proportion of not acceptable products in the selected group 1-δ) can be calculated if the multivariate distributions have known functional form (eg. multivariate normal distributions).

Fisher's sample linear discriminant function is obtained when parameters μ_{π_1}, μ_{π_2} and Σ are replaced with their estimates, $\bar{x}_{\pi_1}$, $\bar{x}_{\pi_2}$ and S^{-1}. The shortcoming is that the probability of misclassification cannot be calculated without knowing both the functional form and the parameters of the multivariate distributions.

Hartler (1992) considers the optimization procedure for selecting the screening variables for reliability prediction of electronic components. Before a stress test, the measurements on very large number of screening variables are taken and their values are ordered in increasing order. After the test, a subset of the screening variables suitable for screening out nonreliable components is determined. The tolerance limits for these variables are also determined. That subset is determined by forming the critical sets of screening variables and calculating the Bayes factors (conditional likelihood quotients) for the number of failed and not failed components. The first critical set has the screening variable with maximum rank. The next critical set is formed by adding the variable with second largest rank. The search for the subset of the screening variables ends when the value of the Bayes factor reaches its maximum. The tolerance limits for the screening variables are the lowest ranks for the variables in the selected subset of the screening variables.

Comments on literature review:

Procedures developed for the cases in which the parameters of the underlying distributions are known are supported with the tables needed for calculating specification limits for the screening variables. The tables are required if the procedure is to be accepted for practical applications. The procedures developed for the cases in which the parameters are not known are too complicated and no tabulated values exist.

The importance of the strength of correlation between the performance variable and a screening variable or a linear combination of screening variables is not emphasized adequately. The efficiency of screening depends primarily on the strength of correlation. Therefore, an increase in the value of the correlation coefficient is an absolute necessity.

CHAPTER 3 - THE MODEL

The amount of information contained in the screening variable X regarding the performance variable Y depends on the strength of association between Y and X. The strength of association, measured by the correlation coefficient ρ_{YX}, can be increased by including more than one screening variable. All the screening procedures considering more than one screening variable developed so far require knowledge of all the parameters of the underlying multivariate distribution.

As a screening procedure is nothing more than the selection of products in which the judgment of the quality is based on the predicted value of the performance variable determined by the measuring screening variables, the form of the prediction function is crucial for that procedure to be successful.

Under the assumption that the relation between the performance variable Y and available screening variables $X_1, X_2, \ldots, X_K$ is linear, a regression function $E(W)=\beta_0+\sum_{i=1}^{K}\beta_i x_i$ is used in this work to make predictions regarding the performance variable Y. If the underlying multivariate distribution is the multivariate normal, the correlation between W and Y is stronger than for any other linear combination W^*. The proof is given in Tong (1990, pg 37).

The regression function is selected because of its desirable properties:

i) The regression function is the best linear prediction function under the assumption that the multivariate distribution of Y and $X_1, X_2, \ldots, X_K$ has mean $\boldsymbol{\mu}=(\mu_Y, \mu_{X_1}, \ldots, \mu_{X_K})$ and positive definite covariance matrix Σ.

ii) The regression function is the best "overall" prediction function if the underlying distribution of Y and $X_1, X_2, \ldots, X_K$ is a K+1 variate normal distribution. The regression function is the mean of the conditional distribution of $Y / X_1, X_2, \ldots, X_K$.

The criterion according to which the regression function is the best prediction function is the mean square error:

$$E(Y - \beta_0 - \sum_{i=1}^{K} \beta_i X_i) < E(Y - g(\boldsymbol{\theta}, \boldsymbol{X}))^2 \tag{3.1}$$

where:

Y - performance variable

$\beta_0 + \sum_{i=1}^{K} \beta_i x_i$ - regression function

$g(\boldsymbol{\theta}, \boldsymbol{X}))$ - any other prediction function

$\boldsymbol{\theta}$ - vector of parameters of the prediction function

$\boldsymbol{X}$ - vector of screening variables

The proof is given in Graybill (1976) and in Tong (1990).

For a K+1 variate distribution, $[Y,X] \sim G_{K+1}(\mu = \begin{bmatrix} \mu_Y \\ \mu_X \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{YX} \\ \sigma_{XY} & \Sigma_{XX} \end{bmatrix})$, with known mean vector μ, and known covariance matrix Σ, the regression coefficients are:

$$\beta = (\beta_1, \beta_2, \ldots, \beta_K) = \sigma_{YX}\Sigma_{XX}^{-1} \tag{3.2}$$

$$\beta_0 = \mu_Y - \sigma_{YX}\Sigma_{XX}^{-1}\mu_X \tag{3.3}$$

Under the assumption that the distribution is a K+1 variate normal, the regression function is:

$$W = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + \varepsilon \tag{3.4}$$

$$\varepsilon \sim N(0,\ \sigma_W^2 = \sigma_Y^2 - \sigma_{YX}\Sigma_{XX}^{-1}\sigma_{XY})$$

with correlation coefficient:

$$\rho_{YW} = \sqrt{\frac{\sigma_{YX}\Sigma_{XX}^{-1}\sigma_{XY}}{\sigma_Y^2}} \tag{3.5}$$

If a product is considered acceptable when the value of the performance variable is above some prespecified lower acceptance limit, the point P_δ above which the proportion of the conditional distribution of Y given X=x is at least δ is:

$$P_\delta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K - z_\delta \sigma_W \tag{3.6}$$

where z_δ is upper percentile point of standard normal distribution and the decision rule for selecting the products is:

accept the product i if the lower limit $L_Y \leq P_{\delta i}$, for $\boldsymbol{X}_i = \boldsymbol{x}_i$. (3.7)

If a product is considered acceptable when the value of the performance variable is below some prespecified upper acceptance limit, the point P_δ below which the proportion of the conditional distribution of Y given $\boldsymbol{X}=\boldsymbol{x}$ is at least δ is given by:

$$P_\delta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K + z_\delta \sigma_W \tag{3.8}$$

and the decision rule for selecting the products is:

accept the product i if the lower limit $U_Y \geq P_{\delta i}$, for $\boldsymbol{X}_i = \boldsymbol{x}_i$. (3.9)

The major drawback of the regression function is that prediction is inefficient if the value of the multiple correlation coefficient between the performance variable Y and the linear combination of screening variables W, ρ_{YW}, is less than 0.9, and should not be used when $\rho_{YW} < 0.8$. However, since the ρ_{YW} is increased by adding screening variables, this problem does not need to be very serious.

3.1. Simulation Study

Simulation study is done in order to evaluate screening procedures. For the models in which a single screening variable is used to select the products of acceptable quality from the bivariate normal distribution $N_2(\boldsymbol{\mu}, \Sigma)$, with:

$$\text{mean vector } \boldsymbol{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 7.0 \\ 10.5 \end{bmatrix}$$

$$\text{covariance matrix } \Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{YX} \\ \sigma_{YX} & \sigma_X^2 \end{bmatrix} = \begin{bmatrix} 0.05 & \sigma_{YX} \\ \sigma_{YX} & 0.2 \end{bmatrix}$$

and correlation coefficient ρ_{YX} from 0.5 to 0.98

300 samples of 600 pairs of variables X and Y are generated using computer program (source code is given in appendix A).

For the models in which linear combinations of the two screening variables are used to select the products of acceptable quality, 300 samples of size 600 are generated from the 3-variate normal distribution $N_3(\boldsymbol{\mu}, \Sigma)$ with:

$$\text{mean vector } \boldsymbol{\mu} = \begin{bmatrix} \mu_Y \\ \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} = \begin{bmatrix} 7.0 \\ 10.5 \\ 4.0 \end{bmatrix}$$

$$
\text{Covariance matrix } \Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{YX_1} & \sigma_{YX_2} \\ \sigma_{X_1Y} & \sigma_{X_1}^2 & \sigma_{X_1X_2} \\ \sigma_{X_2Y} & \sigma_{X_2X_1} & \sigma_{X_2}^2 \end{bmatrix} = \begin{bmatrix} 0.05 & \sigma_{YX_1} & \sigma_{YX_2} \\ \sigma_{X_1Y} & 0.2 & 0.03549 \\ \sigma_{X_2Y} & 0.03549 & 0.07 \end{bmatrix}
$$

$$
\text{and, correlation matrix } \boldsymbol{R} = \begin{bmatrix} 1 & \rho_{YX_1} & \rho_{YX_2} \\ \rho_{X_1Y} & 1 & \rho_{X_1X_2} \\ \rho_{X_2Y} & \rho_{X_2X_1} & 1 \end{bmatrix}
$$

$$
= \begin{bmatrix} 1.0 & 0.1 \le \rho_{YX_1} \le 0.60 & 0.507 \le \rho_{YX_1} \le 0.93 \\ 0.1 \le \rho_{X_1Y} \le 0.60 & 1.0 & 0.3 \\ 0.507 \le \rho_{X_2Y} \le 0.93 & 0.3 & 1.0 \end{bmatrix}
$$

The performance of the procedures depends only on the strength of correlation between the performance and the screening variable or a linear combination of screening variables. The strength of correlation is expressed through the value of the correlation coefficient. The relative magnitudes of variances and covariances do not affect the performance of the screening procedures as long as the value of the correlation coefficient remains same. Also, as location parameters only, the relative magnitude of means have no effect on the performance of the screening procedures.

3.2. Single Screening Variable

Besides regression, the procedures applicable to this case are the procedure developed by Owen, McIntire and Seymour (1975) if the parameters of underlying distributions of the performance variable and a single screening variable are known and the procedure developed by Odeh and Owen (1980) if all of the parameters are not known.

In the following sections comparisons between a regression model and these procedures are made for the situations in which the parameters of the underlying multivariate distributions are known and in which the parameters are not known.

3.2.1. Known Parameters

3.2.1.1. Comparison with the procedure developed by Owen et al.

In this section a comparison of the regression method with the method developed by Owen, McIntyre and Seymour (1975) is made. As noted in Chapter 2, the method of Owen et al. is based on the conditional normal distribution of Y given X. From knowledge of either the proportion of acceptable products in the total population γ, or the parameters μ_Y and σ_Y^2 of the normal distribution of the performance variable, the correlation coefficient ρ_{YX}, and the parameters μ_X and σ_X^2 of the normal distribution of screening variable, the proportion which needs to be selected ψ is given in the tables in Owen et al. (1975) and in the tables in Odeh and Owen (1980). The proportion of acceptable products in the selected group is then guaranteed to be $\delta > \gamma$.

The decision rule for the method of Owen et al. is (for positively correlated variables Y and X and a lower acceptance limit):

$$\text{accept the product } i \text{ if } x_i \geq \mu_X - z_\psi \sigma_X \tag{3.10}$$

The decision rule for the regression method is:

$$\text{accept the product } i \text{ if } L_Y \leq P_{\delta i} = \beta_0 + \beta_1 x_i - z_\delta \sigma_W \tag{3.11}$$

where z_ψ and z_δ are the upper ψ-percentile point and upper δ-percentile point, respectively, of the standard normal distribution.

The data generated by computer from the bivariate normal distribution $N_2(\boldsymbol{\mu}, \Sigma)$ with parameter values given in section 3.2. are used for evaluation and comparison of these screening procedures.

The product is considered acceptable if $Y > L_Y = \mu_Y - 0.675\sigma_Y = 6.849$. Therefore the proportion of acceptable products in each lot is 0.75 before screening. The comparison is made for two values of δ, δ=0.95 and δ=0.99. Thus, the required proportion of acceptable products after screening is 0.95 and 0.99 respectively.

The efficiency of screening increases with an increase in the value of the correlation coefficient. The relation between the value of the correlation coefficient and the number of products selected when the regression method is used and when method of Owen et al. is used are shown in Figures 3.2.1. and 3.2.2..

The proportion of unacceptable products must not exceed the required level 1-δ in any lot. Figures 3.2.3. and 3.2.4. show the quality of screening, expressed through number of lots in which the proportion of unacceptable products after screening exceeds the required levels of 1-δ=0.05 and 1-δ=0.01.

A detailed comparison between these two methods is done for the value of the correlation coefficient $\rho_{YX}=\rho_{YW}=0.8$ which is the lowest value recommended for the regression method. With respect to both the number of acceptable products selected and quality, screening using regression becomes better with increases in the value of ρ_{YW} with great improvement as ρ_{YW} approaches 1. When the method of Owen et al. is used, the number of products selected increases with an increase in ρ_{YX} but the quality of screening remains poor. Therefore the results of the comparison can be extended to cover all the cases in which the correlation coefficient has a value $0.8 \leq \rho_{YX}=\rho_{YW} < 1$. If the correlation coefficient is 1, the two methods are same.

The summary results from 300 samples of size 600 are (calculations of decision rules and detailed results for each sample are given in appendix B):

total number of products - 180000

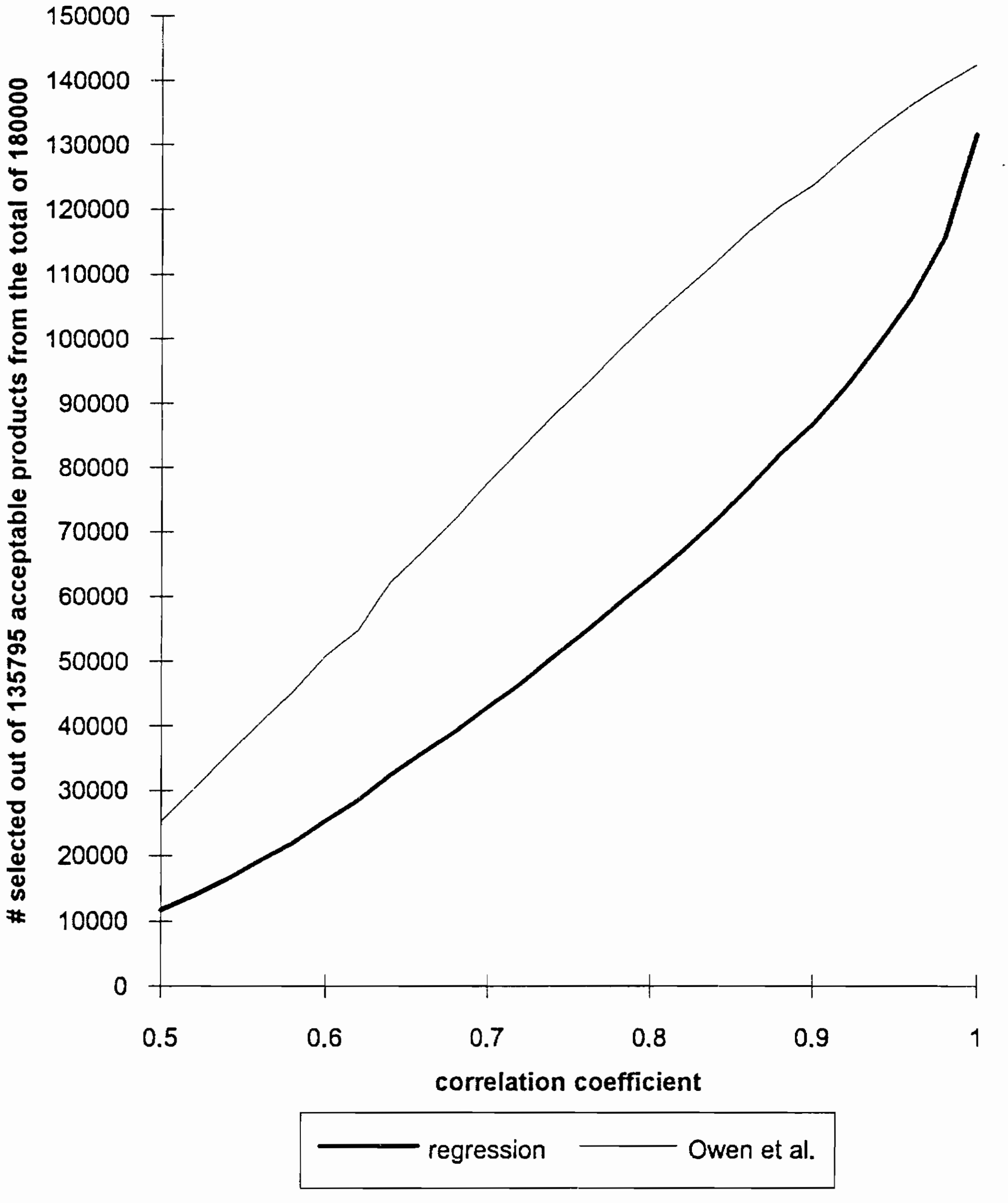

Figure 3.2.1.: $\delta = 0.95$, 1 screening variable

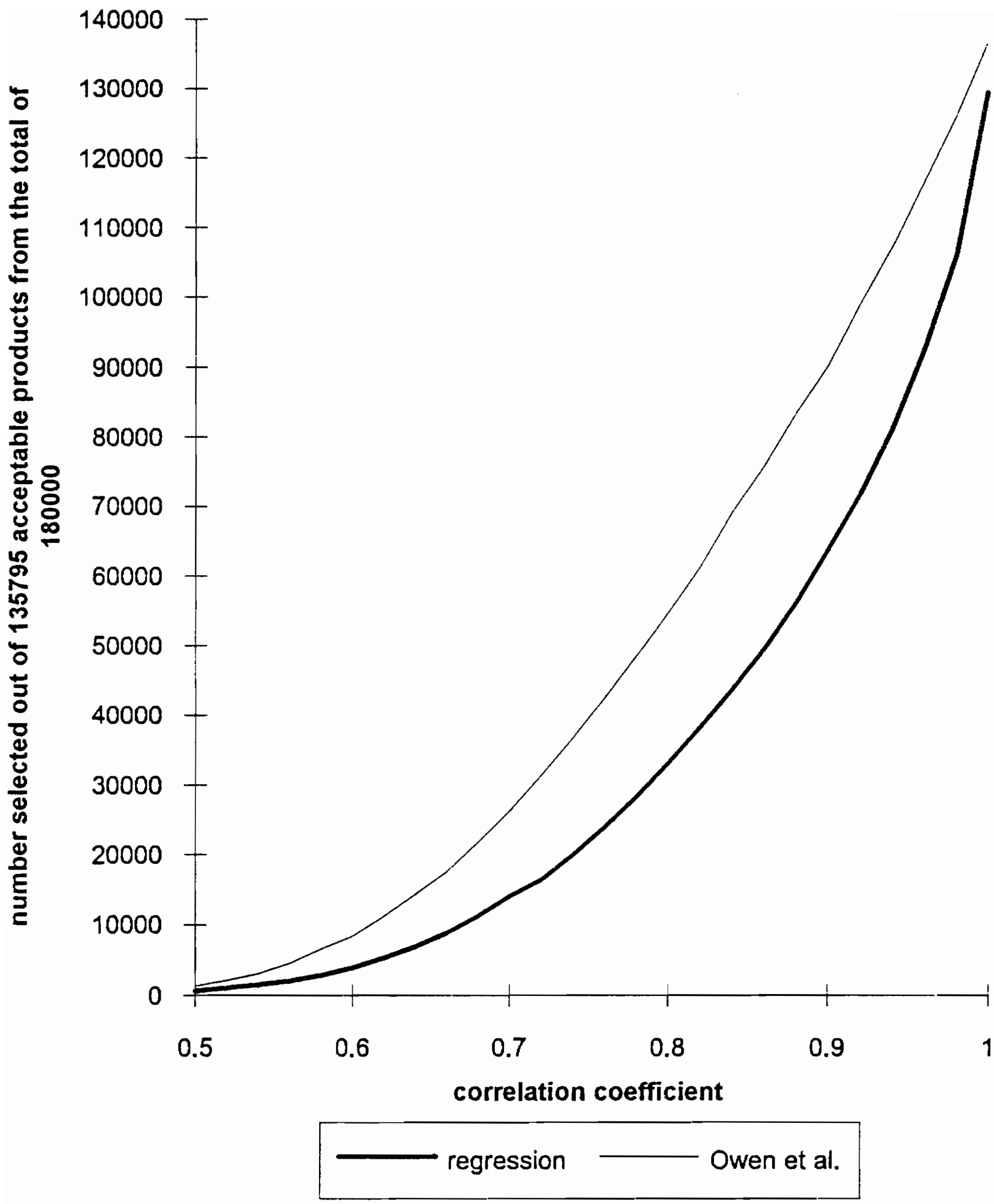

Figure 3.2.2.: $\delta = 0.99$, 1 screening variable

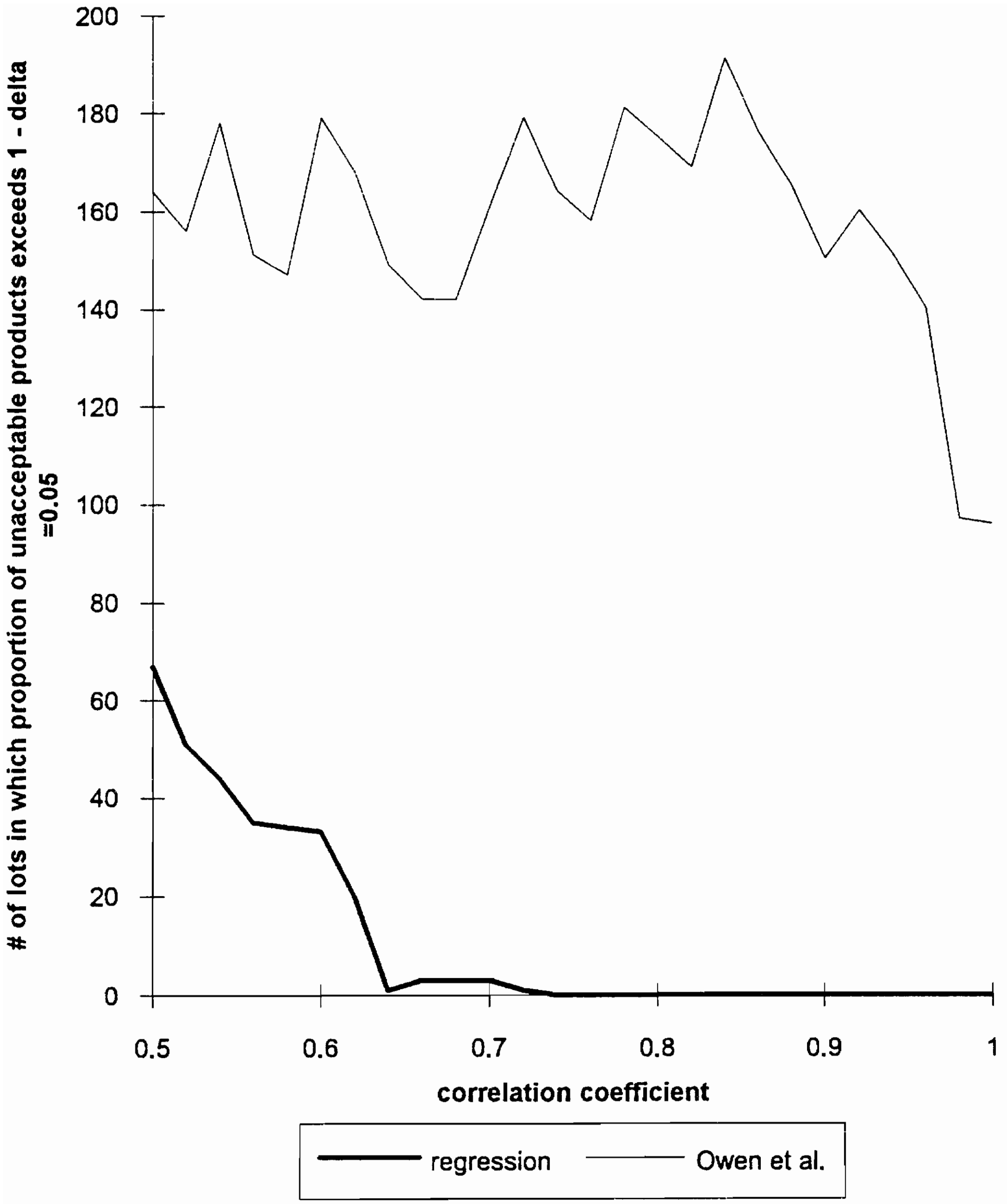

Figure 3.2.3.: $\delta = 0.95$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.05, out of 300 lots

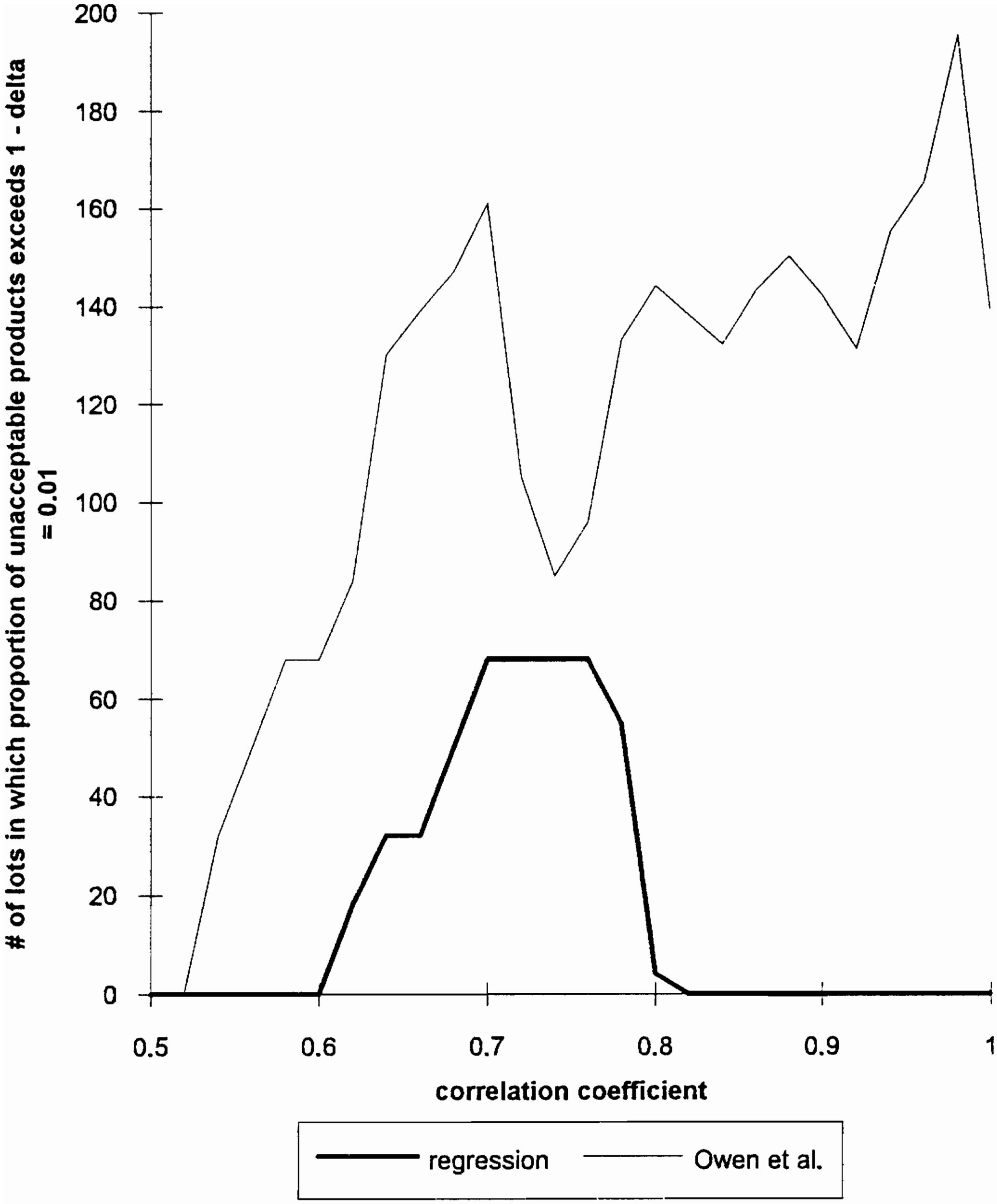

Figure 3.2.4.: $\delta = 0.99$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.01, out of 300 lots

total number of acceptable products before screening	- 135795
average proportion of acceptable products before screening	- $\frac{135795}{180000} = 0.7544$

	Method of Owen et al.		Regression method	
	δ=0.95	δ=0.99	δ=0.95	δ=0.99
total number selected	103000	54813	62869	33072
total number of acceptable products in selected proportion	97339	54278	61927	32976
average proportion of acceptable products in the selected proportion	0.945	0.9903	0.985	0.9971
proportion of acceptable products selected	0.7168	0.399	0.456	0.243

The number of samples (lots) in which proportion of unacceptable products exceeds the nominal level of 1-δ=0.05:.

	Method of Owen et al.	Regression method
$0.05 \le p < 0.06$	59	0
$0.06 \le p < 0.075$	98	0
$0.075 \le p < 0.085$	18	0
$0.085 \le p$	0	0
total (out of 300)	175	0

The number of samples (lots) in which proportion of unacceptable products exceeds the nominal level of 1-δ = 0.01:

	Method of Owen et al.	Regression method
$0.01 \leq p < 0.015$	74	4
$0.015 \leq p < 0.02$	49	0
$0.02 \leq p < 0.025$	20	0
$0.025 \leq p$	1	0
total (out of 300)	144	4

Comments:

i) δ=0.95

The number of unacceptable products is substantially reduced. The reduction is from 44205, or 24.56% of the total population to 5661 or 5.5% of the selected proportion when the method of Owen et al. is used and to 942 or 1.5% of the selected proportion when the regression method is used.

The average proportion of unacceptable products in the groups selected using the method of Owen et al. is around the nominal level of 0.05. However, that proportion exceeds the required level of 0.05 in 175 selected groups or 58.3% of all selected groups and in 20 or 6% of the groups it exceeds 0.075. On the other hand, using the regression method, the required level of 0.05 is not exceeded in any of the selected groups. The price for this high reliability is that many fewer products are selected, 45.6% compared to 71.68%, from 135795 acceptable products from the total population.

From the above comparison, it follows that the regression method should be used in cases in which it is required that every single lot be of high quality. This is probably the best application of any screening method. The method of Owen et al. cannot guarantee that the proportion of acceptable products in every single lot is 0.95.

ii) δ=0.99

The percentage of not acceptable products is highly reduced. The reduction is from 24.56% in the total population, to 0.97% and 0.29% when method of Owen et al. and the regression method are used respectively.

Again, the proportion selected using the method of Owen et al. is much higher and average proportion of not acceptable products in the selected groups is around the nominal level of 0.01. However, that proportion exceeds the nominal level in 144 or 48% of all groups. Also, the variability from sample to sample in the proportion of not acceptable products in the selected groups is high. The proportion of not acceptable products is above 0.02 in 20 groups. When the regression method is used, the number of groups in which the proportion of not acceptable products exceeds 0.1 is only 4 and even there the proportion of not acceptable products is just 0.0102. Again, the price for having lots with guaranteed high quality is that only 24.3% of all acceptable products are selected.

The above comparisons show that screening using the regression method should be used if the percentage of not acceptable products cannot exceeds 1% in any group (lot), but only in cases in which it is absolutely important to save every possible good product. Otherwise the percentages of selected products are too low for practical use. The proportion of products selected can be increased increasing the value of the correlation coefficient. That topic is discussed in section 3.3..

3.2.2. Unknown Parameters Case

If the parameters of the underlying multivariate distribution are not known, the decision rules given above are no longer valid. The proportion δ of acceptable products in the selected groups becomes a random variable. The tolerance interval which covers the proportion δ with a certain probability is required.

3.2.2.1. Procedure Developed by Odeh and Owen

Odeh and Owen (1980) develop a screening procedure to handle the case in which the parameters of the bivariate normal distribution of performance variable Y and screening variable X are not known. The new procedure guarantees the proportion δ is above lower specification limit L_Y, with confidence of η:

$$P(P(Y > L_Y/X > L_X) \geq \delta) = \eta \tag{3.12}$$

The procedure is as follows:

i) A sample correlation coefficient r is calculated from a sample of size n and depending on direction of dependency, lower for positive and upper for negative, a 100ξ% confidence limit ρ^* on the correlation coefficient ρ is determined.

ii) A 100ξ% lower confidence limit γ^* for the proportion of acceptable products before screening $\gamma = P(Y > L_Y)$, is determined.

iii) From the tables in Odeh and Owen (1980) with entries n-1,γ^*, $\rho^*\sqrt{\frac{n}{n-1}}$, and δ, the multiplying coefficient t_ψ for the sample standard deviation s_X is obtained.

The procedure then guarantees that:

$$P(P(Y > L_Y / X > L_X = \bar{x} - t_\psi s_x \sqrt{\tfrac{1}{n} + 1}\ ) \geq 2\xi\text{-}1) = \eta \tag{3.13}$$

The procedure is based on a normal distribution conditioned on a t-distribution. The lower confidence limit γ^* is calculated using the noncentral t-distribution. The details are given in Owen and Haas (1978) and Owen (1968). Lower confidence limits on γ^* and lower and upper confidence limits on ρ^* are tabulated in Odeh and Owen (1980).

3.2.2.2. Regression Method

Replacing the parameters with their estimates changes the coefficients of the regression equation from fixed values to random variables. The standard regression technique for calculating the prediction interval such that the single new observation Y for the given observed values $\boldsymbol{X}=\boldsymbol{x}$ will lie in it with the desired confidence of η cannot be used. The number of products in a lot may be arbitrarily large so that the number of predictions from the repeated sampling of $\boldsymbol{X}$ calculated using the originally fitted model is also arbitrarily large. Then the probability that the proportion δ of all future observations Y will lie in its prediction interval is no longer η.

Liberman and Miller (1963) develop a tolerance interval $((L(\vartheta),U(\vartheta))$ with the property that the proportion ϑ of all future observation lie in it for any value of $\boldsymbol{X}=\boldsymbol{x}$ using the originally fitted model with the confidence level of φ. Liberman and Miller use the Bonferroni inequality to combine the Working-Hotelling-Scheffe simultaneous confidence interval for the regression plane with an upper one-sided confidence interval on the standard deviation σ_W.

The lower and upper tolerance limits are obtained as follows:

i) The Working-Hotelling-Scheffe two-sided confidence region for the entire regression plane is given by (Liberman and Miller, 1963):

$$P(|\boldsymbol{x}'\boldsymbol{b} - \boldsymbol{x}'\boldsymbol{\beta}| \leq s_W\sqrt{(k + 1)F_{1-\varphi/2,\, k+1,\, n-k-1}\boldsymbol{x}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{x}}, \text{ for all } \boldsymbol{x}) = 1 - \varphi/2 \qquad (3.14)$$

where

$\boldsymbol{x}$ - k x 1 vector of observations

$\boldsymbol{b}$ - $(k + 1)$ x 1 vector of estimated regression coefficients

$\boldsymbol{\beta}$ - $(k + 1)$ x 1 vector of true regression coefficients

s_W - estimated standard deviation

$\boldsymbol{X}$ - n x $(k + 1)$ matrix of observations

$F_{1-\varphi/2,\ k+1,\ n-k-1}$ - upper 1 - $\varphi/2$ percentile point of the F- distribution with $k + 1$ and n - k - 1 degrees of freedom

ii) An upper bound for σ_W is given by:

$$P\left(\frac{(n - k - 1)s_W^2}{\sigma_W^2} \geq \chi^2_{\varphi/2,\ n-k-1}\right) = 1 - \varphi/2 \tag{3.15}$$

$$P\left(\sigma_W \leq s_W \sqrt{\frac{n-k-1}{\chi^2_{\varphi/2,\ n-k-1}}}\right) = 1 - \varphi/2 \tag{3.16}$$

where:

$\chi^2_{\varphi/2,\ n-k-1}$ - lower $\varphi/2$ percentile point of the χ^2 - distribution with n - k - 1 degrees of freedom

iii) The Bonferonni inequality:

$$P(AB) \geq 1 - P(A^c) - P(B^c) \tag{3.17}$$

Using the Bonferroni inequality to combine 3.15 and 3.16 it follows that:

$$P\left(|x'\beta \pm z_{\vartheta}\sigma_W - x'b| \le s_W\left(\sqrt{(k+1)F_{1-\varphi/2,\ k+1,\ n-k-1}x'X'Xx}\right.\right.$$

$$\left.\left. + \ z_{\vartheta}\sqrt{\frac{n-k-1}{\chi^2_{\varphi/2,\ n-k-1}}}\ ,\ \text{for all } x,\ \vartheta\right) \ge 1-\varphi \qquad (3.18)$$

where z_{ϑ} is defined by:

$$\vartheta = \Phi(z_{\vartheta}) - \Phi(-z_{\vartheta}) \qquad (3.19)$$

Either of the two limits of the tolerance interval can be taken to be infinity. The tolerance interval controls both sides of the proportion ϑ inside the tolerance interval with equal probabilities and the Working-Hotelling-Scheffe confidence interval for the entire regression plane is two-sided. Therefore, the one-sided tolerance interval can be obtained by changing the percentile points of the standard normal distribution to z_{δ} such that $\delta=\Phi(z_{\delta})$, the F-distribution to $F_{1-\eta}$, and the χ^2 distribution to $\chi^2_{1-\eta}$.

The one-sided tolerance interval is used for screening products. Acceptability of a product is determined by comparing the lower (upper) tolerance limit with the acceptance level. The product is accepted if its lower (upper) tolerance limit is not below (above) value of the acceptance limit:

For a lower acceptance limit,

$$L_Y \leq (L(\delta)) = \boldsymbol{b}'\boldsymbol{x} - s_W\left(\sqrt{(k+1)F_{1-\eta,\, k+1,\, n-k-1}\boldsymbol{x}'X'X\boldsymbol{x}} + z_\delta\sqrt{\frac{n-k-1}{\chi^2_{\eta,\, n-k-1}}}\right) \tag{3.20}$$

or, for an upper acceptance limit,

$$U_Y \geq (U(\delta)) = \boldsymbol{b}'\boldsymbol{x} + s_W\left(\sqrt{(k+1)F_{1-\eta,\, k+1,\, n-k-1}\boldsymbol{x}'X'X\boldsymbol{x}} + z_\delta\sqrt{\frac{n-k-1}{\chi^2_{\eta,\, n-k-1}}}\right) \tag{3.21}$$

The procedure then guarantees that:

$$P(\text{proportion of acceptable products} \geq \delta) \geq \eta \tag{3.22}$$

3.2.2.3. Comparison of Regression Method With Method of Odeh and Owen

The same assumptions are made as for the previous comparisons. A product is acceptable if the value of the performance variable Y is above a prespecified lower acceptance limit and the correlation between the performance and the screening variable is positive. The proportion of acceptable products before screening is 0.75. Again comparisons are made for δ=0.95, and δ=0.99. The confidence with which the proportion of acceptable products is δ is η=0.95, and η=0.99.

The efficiency and quality of screening of both methods is shown in Figures 3.2.5. - 3.2.12.

The same 300 samples of size 600 are used as in the cases where the parameters were known. From each of the 300 samples (lots), a sample of size 40 is randomly

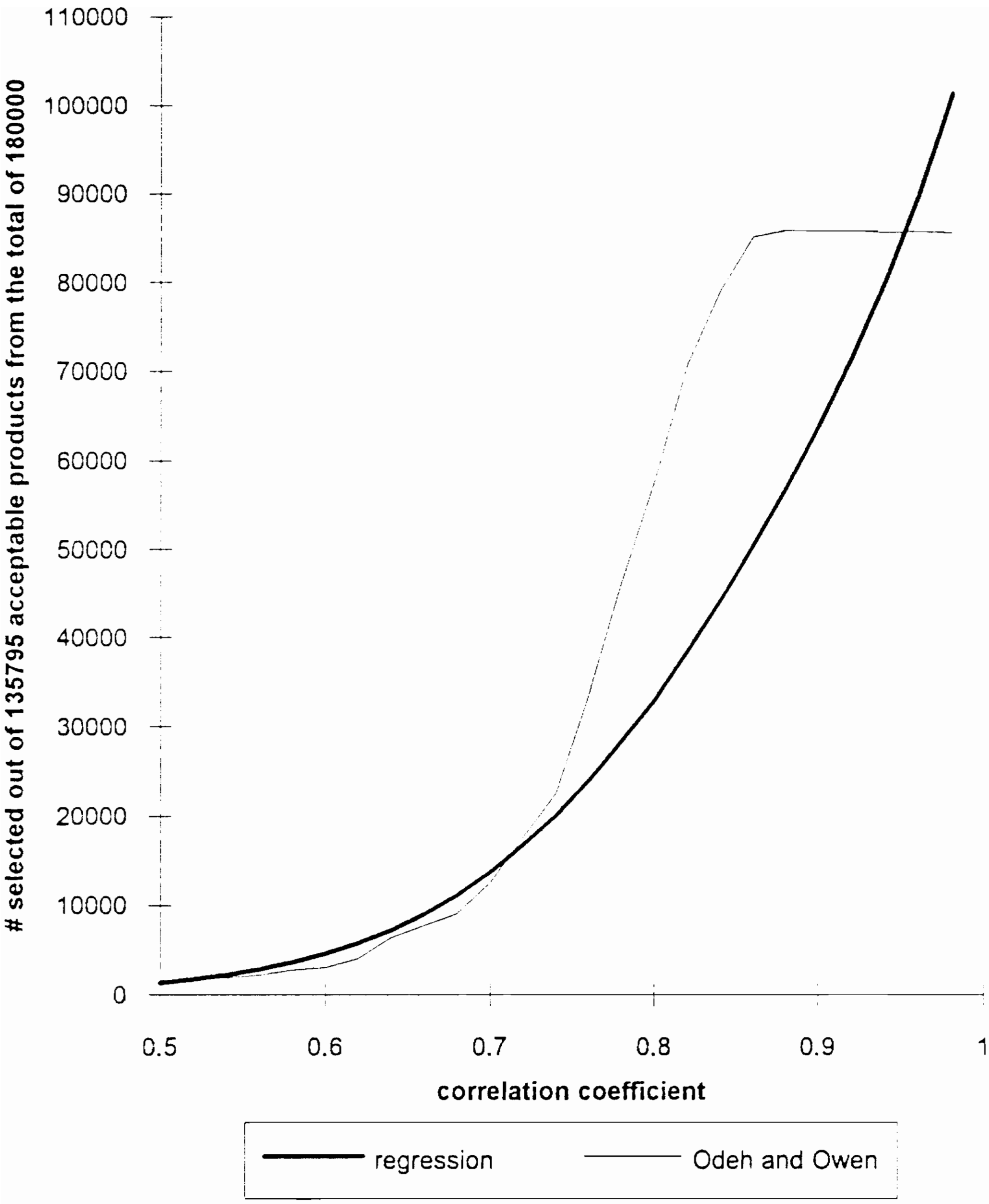

Figure 3.2.5.: $\eta = 0.95$, $\delta = 0.95$, 1 screening variable

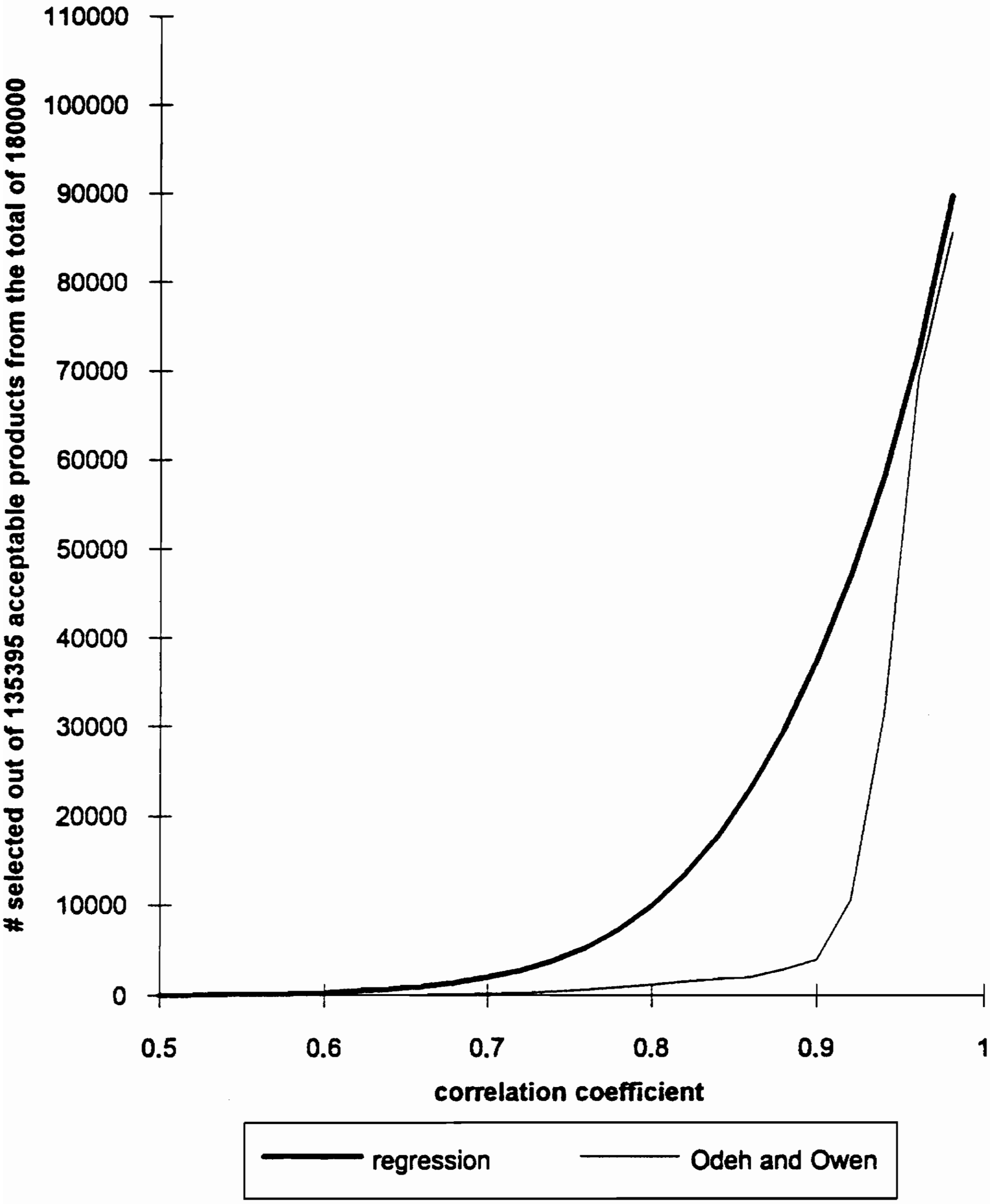

Figure 3.2.6.: $\eta = 0.95$, $\delta = 0.99$, 1 screening variable

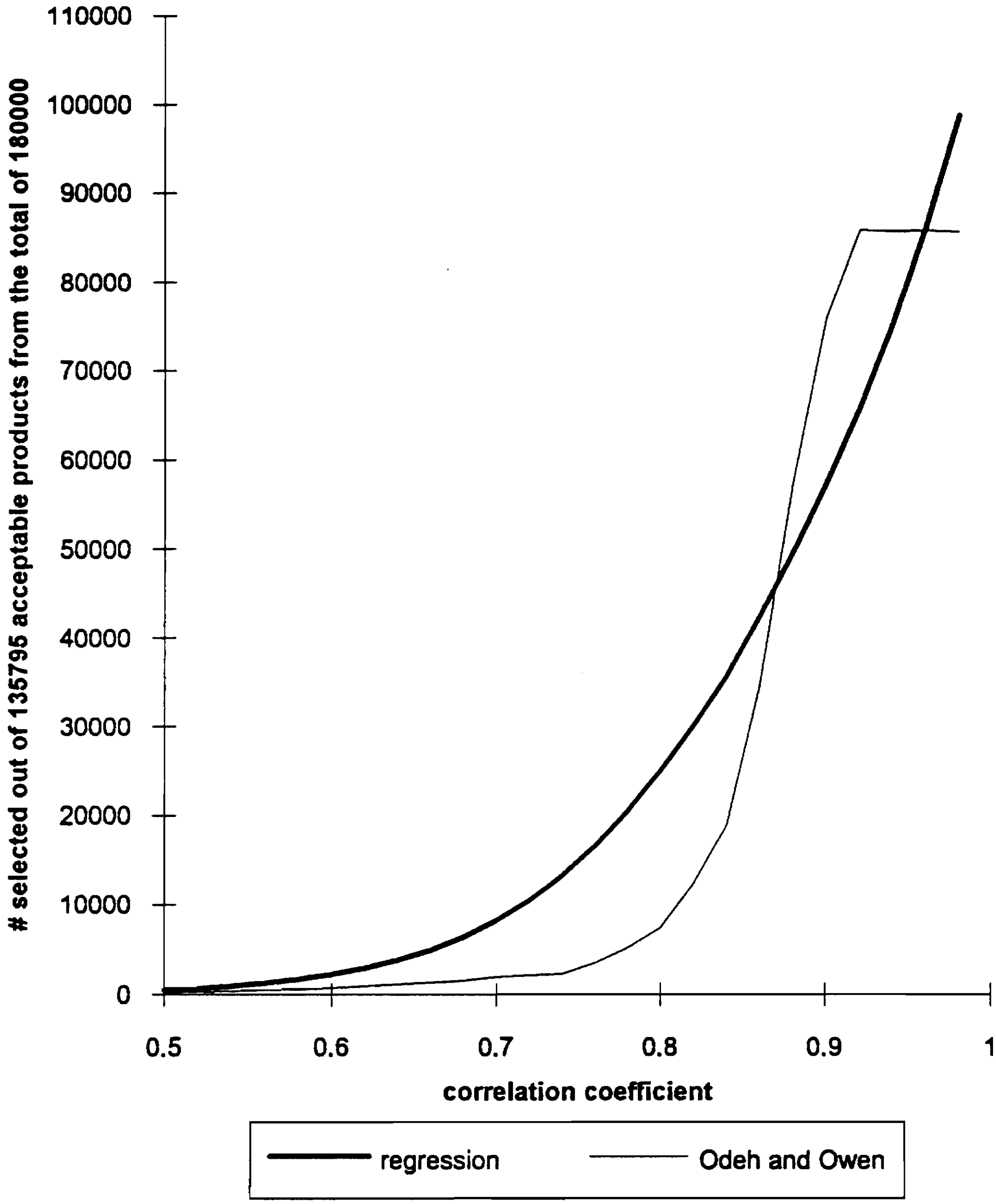

Figure 3.2.7.: $\eta = 0.95$, $\delta = 0.99$, 1 screening variable

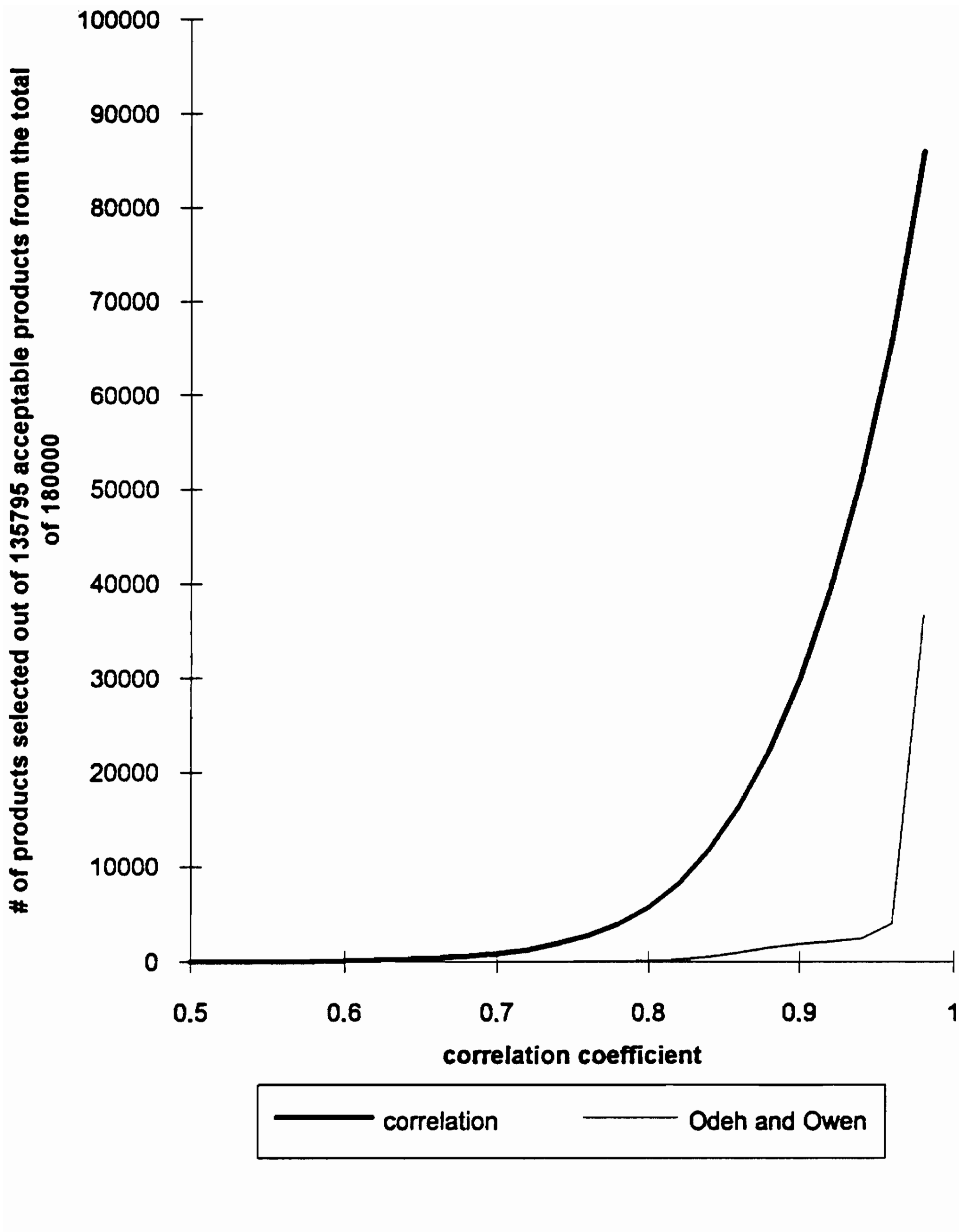

Figure 3.2.8.: $\eta = 0.99$, $\delta = 0.99$, 1 screening variable

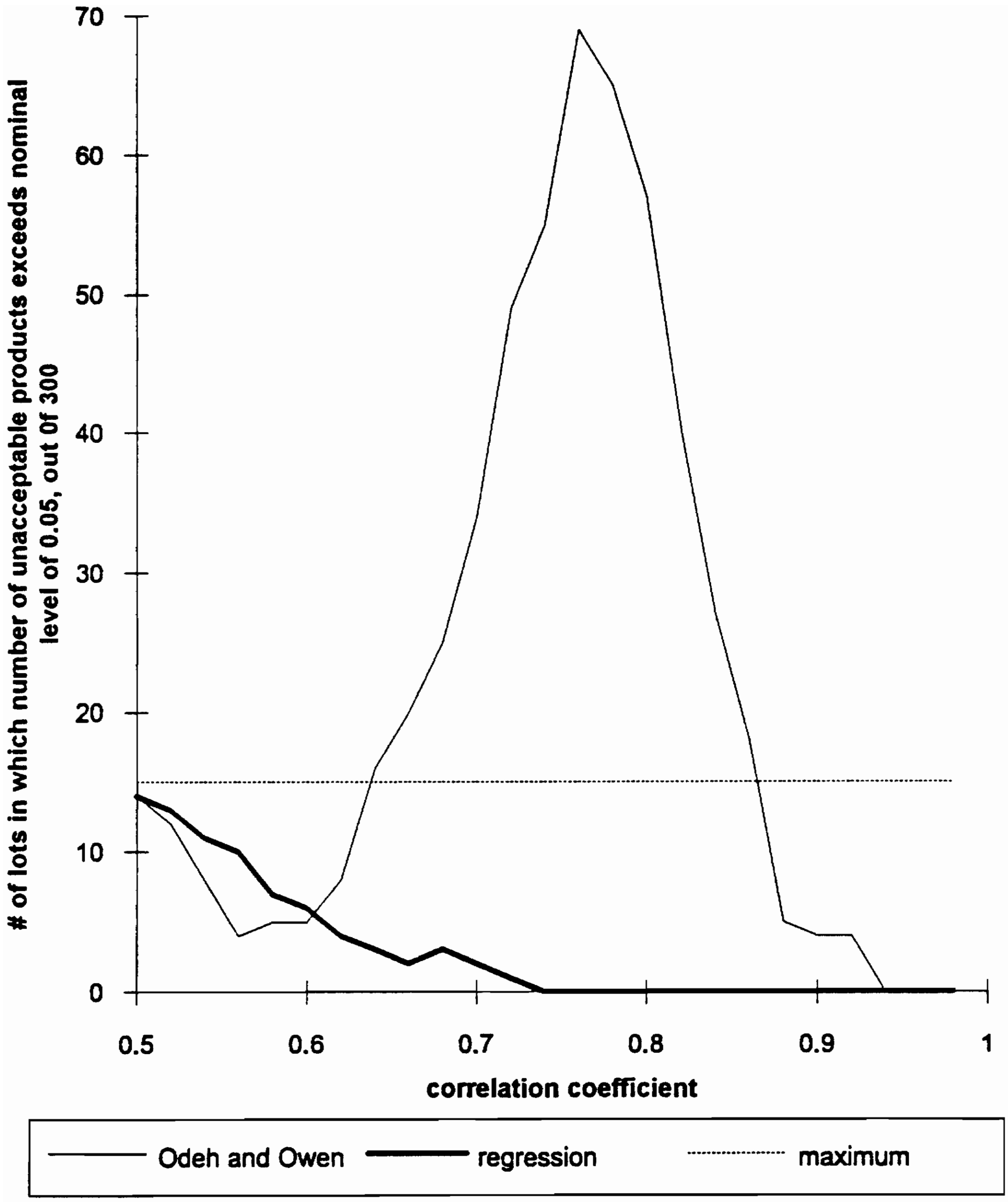

Figure 3.2.9.: $\eta = 0.95$, $\delta = 0.95$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.05, out of 300 lots

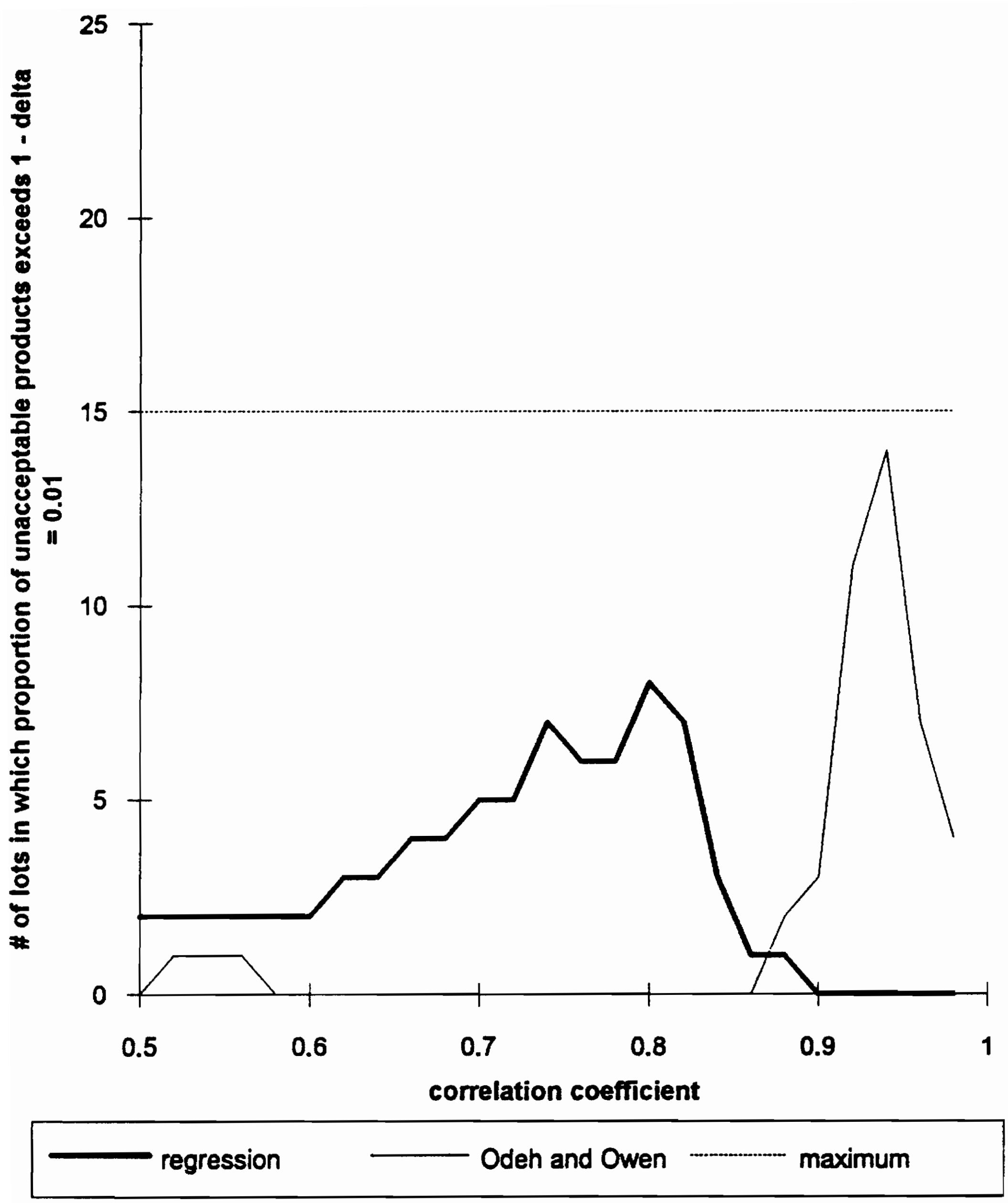

Figure 3.2.10.: η = 0.95, δ = 0.99, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.01, out of 300 lots

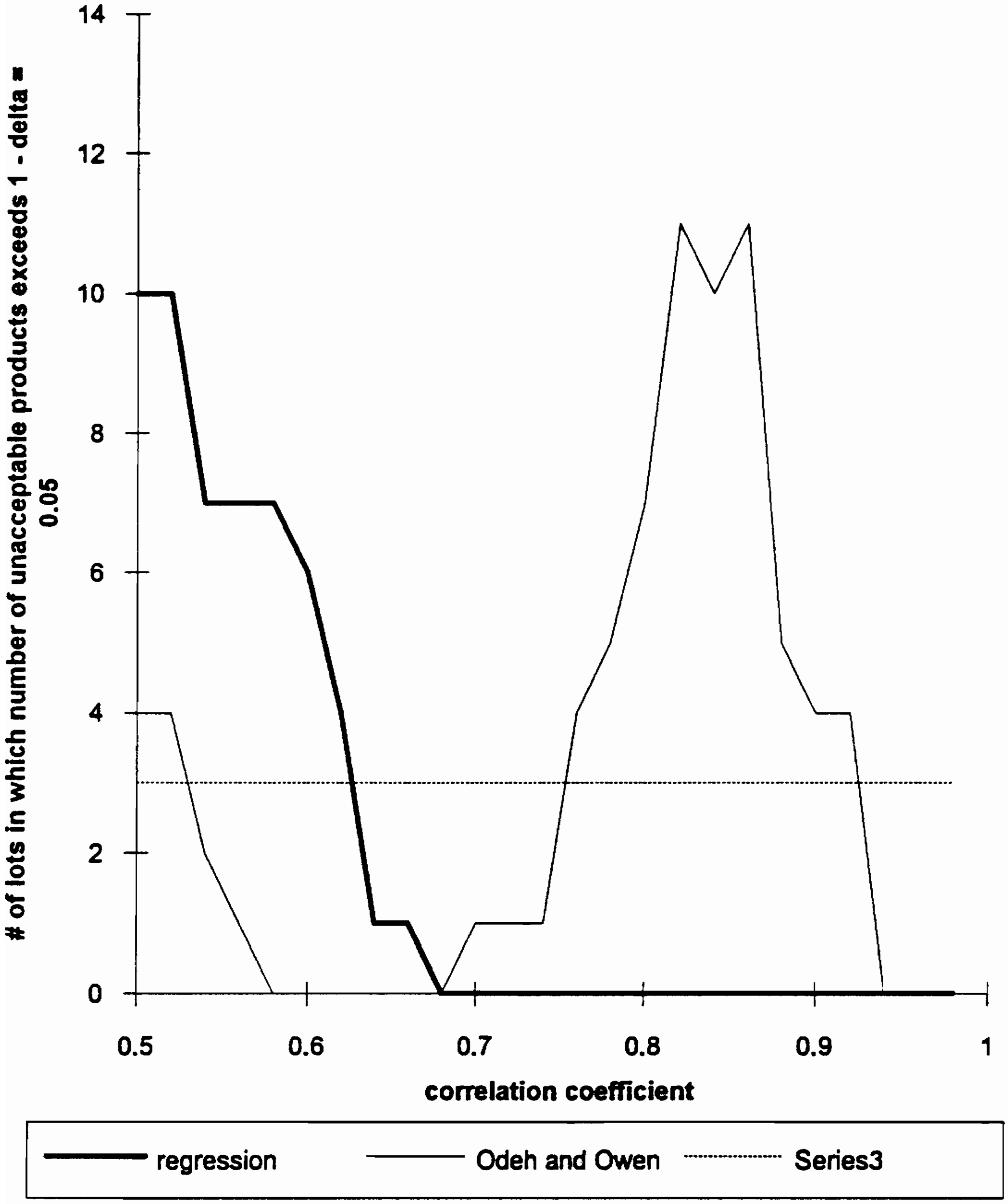

Figure 3.2.11.: $\eta = 0.99$, $\delta = 0.95$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.05, out of 300 lots

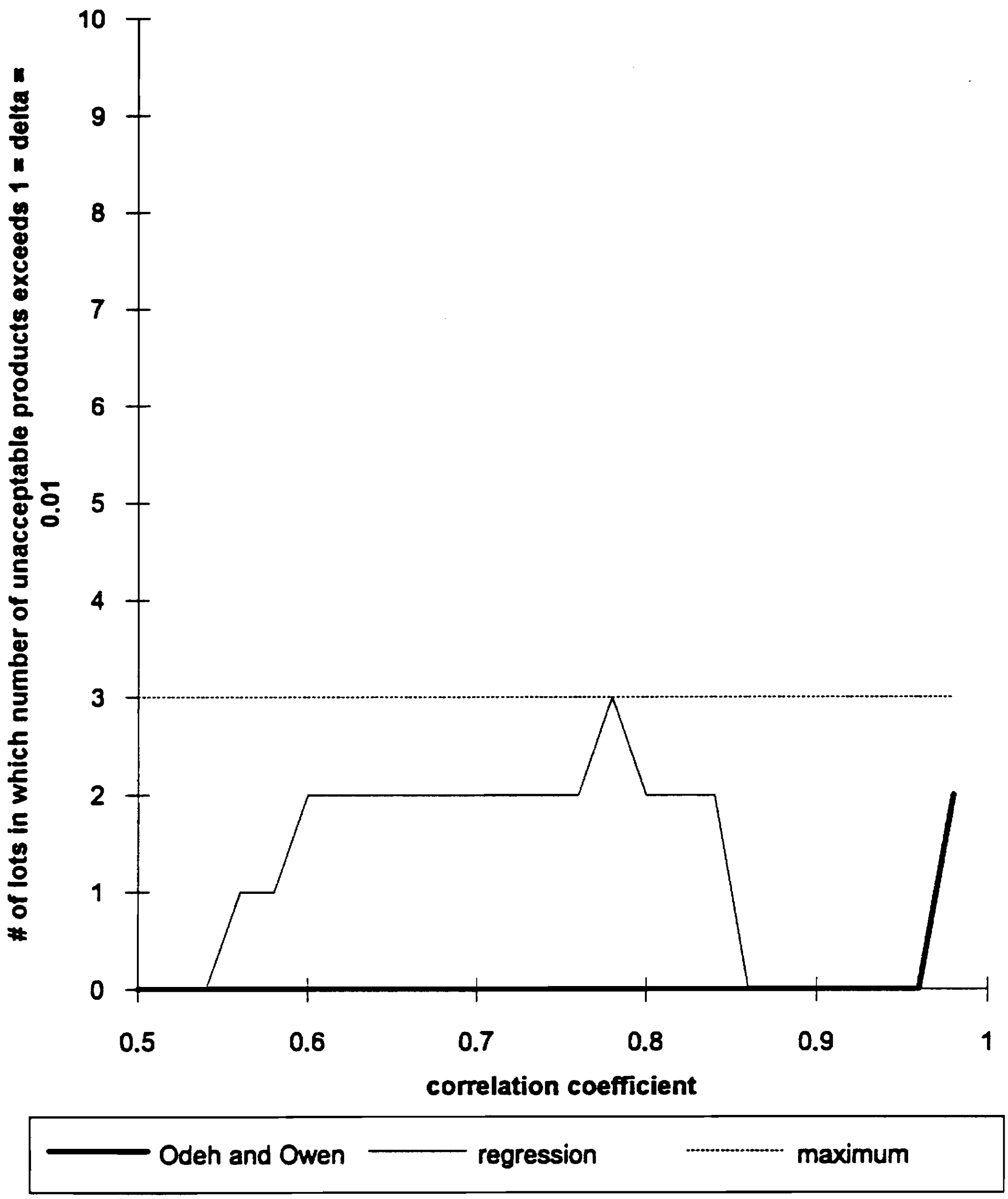

Figure 3.2.12.: η = 0.99, δ = 0.99, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.01, out of 300 lots

selected and maximum likelihood estimates for the unknown parameters are calculated for each lot separately. Therefore, the estimates of the parameters in both decision functions are the same in the same lot but are different across the lots. All proportions in the tables below are calculated on the basis of 180000 products. The total number is not reduced for the 12000 products in the samples used for calculating estimates. Otherwise the very large samples would artificially increase the efficiency of both methods. For example, using 150000 products to estimate parameters and reducing the total number of products to 30000 would make the proportion of accepted products to be around 90%.

The decision rule for selection when the method of Odeh and Owen is used is:

accept the product i if

$$x_i \geq \bar{x} - t_\psi s_X \sqrt{1 + \frac{1}{n}} \tag{3.23}$$

where $\bar{x}$ and s_X^2 are maximum likelihood estimates of μ_X and σ_X^2 adjusted for bias. The coefficient t_ψ is obtained from the tables in Odeh and Owen (1980) with interpolation and extrapolation when needed.

The decision rule for selection when the regression method is used is based on a lower one-sided tolerance interval, $((L(\delta),\infty)$. All products for which the lower one-sided tolerance interval is above the acceptance limit for the performance variable Y are considered acceptable. Thus the decision rule is:

accept product i if

$$L_Y \leq L_i(\delta) = b_0 + b_1 x_i - s_W\left(\sqrt{2F_{1-\eta,\,2,\,n-2} \frac{(x_i - \bar{x})^2}{\sum_{j=1}^{K} (x_j - \bar{x})^2}} + z_\delta \sqrt{\frac{n-2}{\chi^2_{\eta,\,n-2}}} \right) \tag{3.24}$$

where b_0, b_1, and s_W^2 are maximum likelihood estimates of β_0, β_1, and σ_W^2 respectively, adjusted for bias where needed.

The summary results from 300 samples of size 600 are (the detailed results for each sample are given in appendix C):

total number of products	- 180000
total number of acceptable products before screening	- 135795
average proportion of acceptable products before screening	- $\frac{135795}{180000} = 0.7544$
total number sampled for estimation	- 12000

i) comparisons when confidence level is $\eta = 0.95$:

	Method of Odeh and Owen		Regression method	
	$\delta = 0.95$	$\delta = 0.99$	$\delta = 0.95$	$\delta = 0.99$
total number selected	57853	2284	32970	9922
total number of acceptable products in selected proportion	55343	1226	32785	9912

	Method of Odeh and Owen $\delta = 0.95$	$\delta = 0.99$	Regression method $\delta = 0.95$	$\delta = 0.99$
average proportion of acceptable products in the selected proportion	0.9566	1.0	0.9943	0.999
proportion of acceptable products selected	0.4075	0.009	0.2414	0.0729

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - $\delta = 0.05$:

	Method of Odeh and Owen	Regression method
$0.05 \leq p < 0.06$	24	0
$0.06 \leq p < 0.075$	20	0
$0.075 \leq p < 0.085$	9	0
$0.085 \leq p$	4	0
total (out of 300)	57	0

actual confidence level = $1 - \frac{57}{300} = 0.81 < \eta = 0.95, \; 1 - \frac{0}{300} = 1.0 > \eta = 0.95$

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - $\delta = 0.01$:

	Method of Odeh and Owen	Regression method
$0.01 \leq p < 0.015$	0	6
$0.015 \leq p < 0.02$	0	2
$0.02 \leq p < 0.025$	0	0
$0.085 \leq p$	0	0
total (out of 300)	0	8

actual confidence level = 1 - $\frac{0}{300}$ = 1.0 > η = 0.95 , 1 - $\frac{8}{300}$ = 9.73 > η = 0.95

Comments:

The proportion of not acceptable products is greatly reduced again.

When the proportion δ = 0.95 of acceptable products is required, the confidence obtained when the method of Odeh and Owen is used is well below the required 0.95. Moreover, the number of products selected varies greatly from lot to lot. In 94 lots the number of selected products is 15 or less and in 198 the number of selected products is higher than 200 (detailed results are given in appendix C). The reason is that the procedure depends highly on the precision of estimates of the correlation coefficient ρ and of the proportion of acceptable products before screening γ. A lower one-sided confidence limit is used instead of the true values of these two parameters (upper for the correlation coefficient if correlation is negative). In certain combinations this leads to extremely conservative screening or to very liberal screening. When both parameters are underestimated or overestimated, respectively. Moreover, the required sample size for estimation of ρ with precision of $\pm$0.05 is 1080, Owen, McIntire and Seymour (1975). This sample size is prohibitively large.

Using regression, the selected proportions are evenly distributed from sample to sample and the observed confidence level exceeds the nominal level of 0.95. Therefore, the regression method has an advantage over the method of Odeh and

Owen.

When the proportion of acceptable products required is $\delta = 0.99$, no procedure performs satisfactory. Only 0.9% and 7.2% of all acceptable products are selected when the procedure of Odeh and Owen and the regression procedure are used respectively. Although the confidence level is exceeded in both procedures, neither procedure should be recommended for application for the given value of correlation coefficient (0.8). For that value of δ, the correlation coefficient should be at least 0.95 (Figure 3.2.6.).

ii) comparison when confidence level is $\eta = 0.99$

	Method of Odeh and Owen		Regression method	
	$\delta = 0.95$	$\delta = 0.99$	$\delta = 0.95$	$\delta = 0.99$
total number selected	7426	162	25019	5757
total number of acceptable products in selected proportion	7139	162	24940	5753
average proportion of acceptable products in the selected proportion	0.961	1.0	0.997	0.9993
proportion of acceptable products selected	0.0525	0.0012	0.1836	0.0423

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of $1 - \delta = 0.05$:

	Method of Owen et al.	Regression method
$0.05 \leq p < 0.06$	2	0
$0.06 \leq p < 0.075$	2	0
$0.075 \leq p < 0.085$	2	0
$0.085 \leq p$	1	0
total (out of 300)	7	0

actual confidence level = 1 - $\frac{7}{300}$ = 97.7 $< \eta = 0.99$, 1 - $\frac{0}{300}$ = 1.0 > 0.99

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - $\delta = 0.01$:

	Method of Owen et al.	Regression method
$0.01 \leq p < 0.015$	0	0
$0.015 \leq p < 0.02$	0	2
$0.02 \leq p < 0.025$	0	0
$0.085 \leq p$	0	0
total (out of 300)	0	2

actual confidence level = 1 - $\frac{0}{300}$ = 1.0 > $\eta = 0.99$, 1 - $\frac{2}{300}$ = 0.99 = 0.99

Comments:

Both methods are extremely conservative. Although, the proportion of acceptable products in the selected groups satisfy the requirement of $\delta \geq 0.99$ with a confidence of $\eta \geq 0.99$ for both methods and the requirement of $\delta \geq 0.95$ for the regression method, the selected proportions are too low to use any of the above procedures in practice.

The screening can be substantially improved by increasing the value of the correlation coefficient. The screening using the regression method is satisfactory for $\rho_{YX} > 0.95$, and it is much better than the method of Odeh and Owen, especially for δ=0.99 (Figures 3.2.7. and 3.2.8.).

3.3. Multiple Screening Variables

The efficiency of screening increases with increases in the value of the correlation coefficient. The value of the correlation coefficient can be increased by using more than one screening variable. The multiple screening variables case is considered in the following sections.

3.3.1. Known Parameters Case

When the parameters of the underlying 3-variate normal distribution are known, two procedures are applicable. The procedure developed by Owen, McIntire and Seymour (1975) and the regression procedure which remains the same regardless of number of screening variables.

3.3.1.1. Method of Owen et al.

Owen, McIntire and Seymour (1975) develop a procedure which uses a linear combination of the screening variables. The coefficients in the linear combination $V=a_1X_1+a_2X_2$ are calculated so that the proportion of acceptable products in a selected sub-population is maximized. These coefficients are:

$$a_1=\frac{\rho_{YX_1}-\rho_{YX_2}\rho_{X_1X_2}}{\sqrt{(1-\rho^2_{X_1X_2})(\rho^2_{YX_1}+\rho^2_{YX_2}-2\rho_{X_1X_2}\rho_{YX_1}\rho_{YX_2})}} \tag{3.25}$$

$$a_2=\frac{\rho_{YX_2}-\rho_{YX_1}\rho_{X_1X_2}}{\sqrt{(1-\rho^2_{X_1X_2})(\rho^2_{YX_1}+\rho^2_{YX_2}-2\rho_{X_1X_2}\rho_{YX_1}\rho_{YX_2})}} \tag{3.26}$$

As the procedure requires that the parameters of the underlying 3-variate normal distribution are known, the linear combination follows the normal distribution with mean $\mu_V=a_1\mu_{X_1}+a_2\mu_{X_2}$ and variance $\sigma^2_V=a_1\sigma^2_{X_1}+a_2\sigma^2_{X_2}+2\rho_{X_1X_2}a_1a_2\sigma_{X_1}\sigma_{X_2}$. Furthermore, the joint distribution of performance variable Y and the linear combination V is a bivariate normal with correlation coefficient

$$\rho_{YV}=\frac{a_1\sigma_{X_1}\rho_{YX_1}+a_2\sigma_{X_2}\rho_{YX_2}}{\sqrt{(a_1^2\sigma^2_{X_1}+a_2^2\sigma^2_{X_2}+2\rho_{X_1X_2}a_1a_2\sigma_{X_2})}}, \tag{3.27}$$

and the conditional distribution of Y given V is univariate normal. Thus, the method is identical to that for the single screening variable (Owen, McIntire and Seymour, 1975) with V replacing that variable:

$$P(Y > L_Y / V > \mu_V - z_\psi\sigma_V) = \delta \tag{3.28}$$

The same tables from Odeh and Owen (1980) or from Owen, McIntire and Seymour (1975) can be used to determine the proportion ψ which needs to be selected in order to have the proportion of acceptable products after screening be $\delta > \gamma$.

3.3.1.2. Comparison of Method of Owen et al. With the Regression Method

In this section, comparisons for the cases in which, the required proportion of acceptable products is δ=0.95 and δ=0.99 are made. The data from the 3-variate normal distribution $N_3(\mu, \Sigma)$ are generated by a computer. Parameter values are given in section 3.2..

A product is acceptable if the value of the performance variable Y is above the lower acceptance limit. The proportion of acceptable products before screening is 0.75 so the lower acceptance limit is L_Y=6.849. The comparisons are made for δ=0.95 and δ=0.99.

The relation between the value of the correlation coefficient and the number of products selected when the regression method is used and when the method of Owen et al. is used are shown in figures 3.3.1. and 3.3.2..

The proportion of not acceptable products must not exceed the required level 1 - δ in any lot. Figures 3.3.3. and 3.3.4. show the quality of screening expressed as the number of lots in which the proportion of not acceptable products after screening

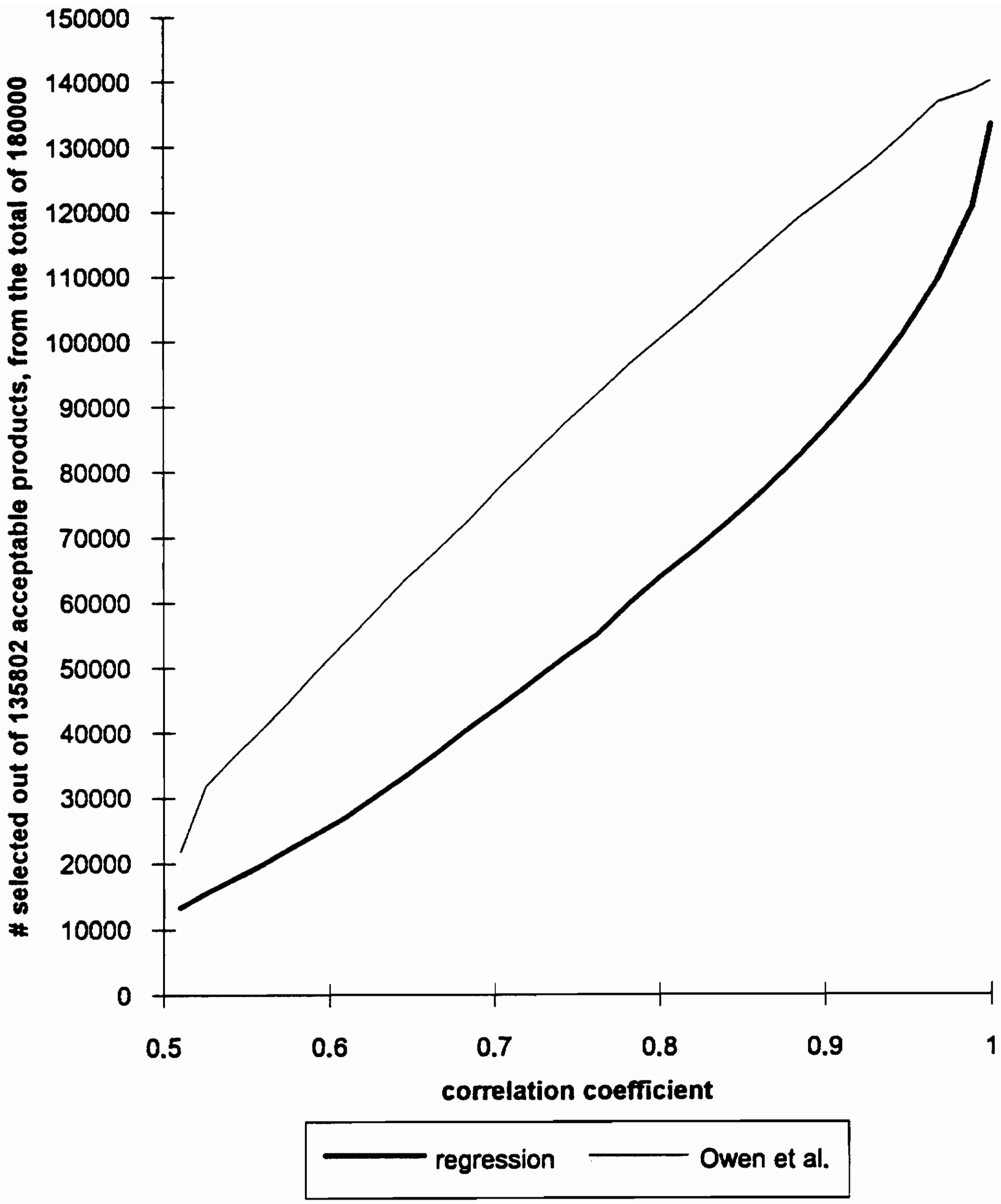

Figure 3.3.1.: $\delta = 0.95$, 2 screening variables

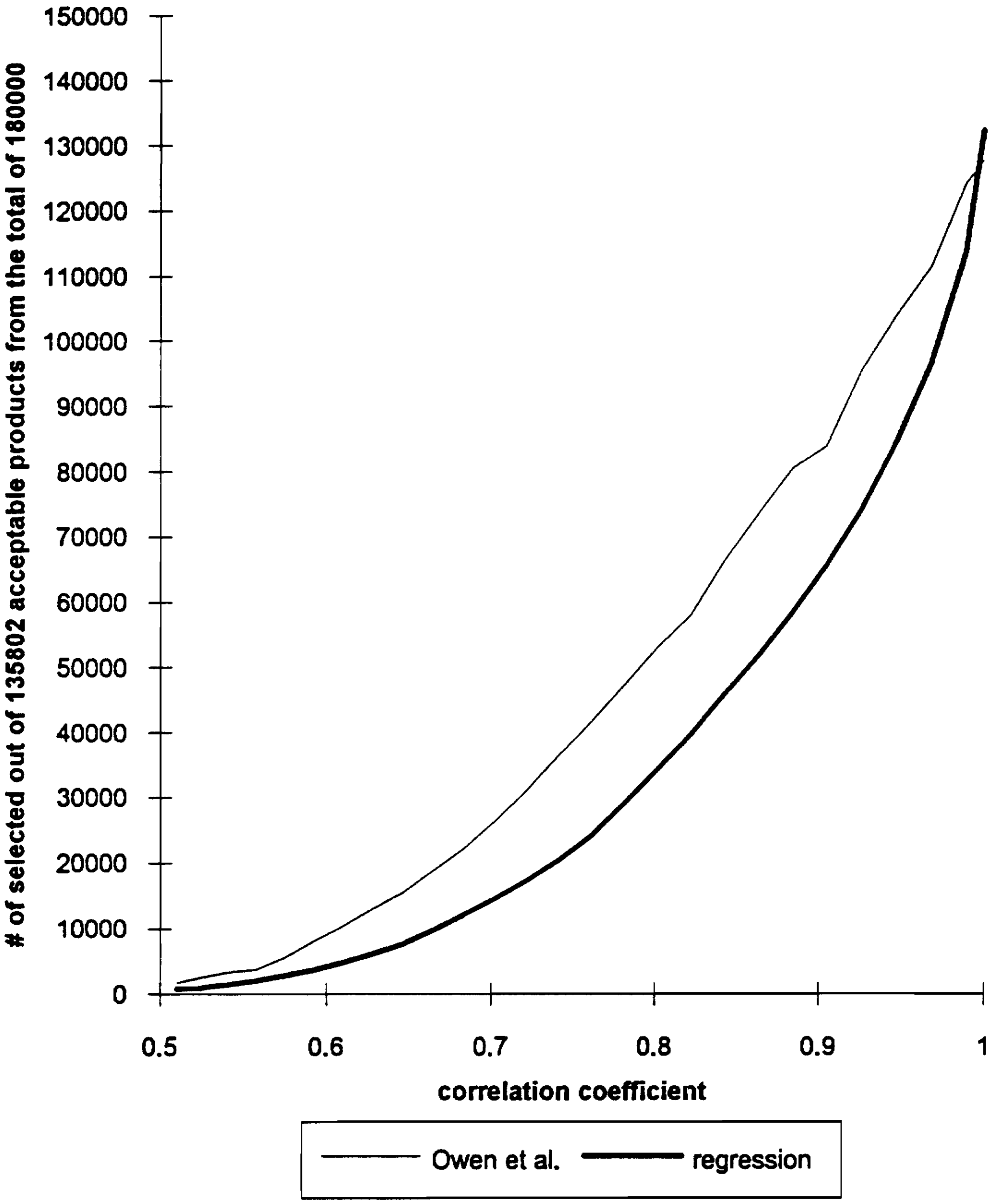

Figure 3.3.2.: $\delta = 0.99$, 2 screening variables

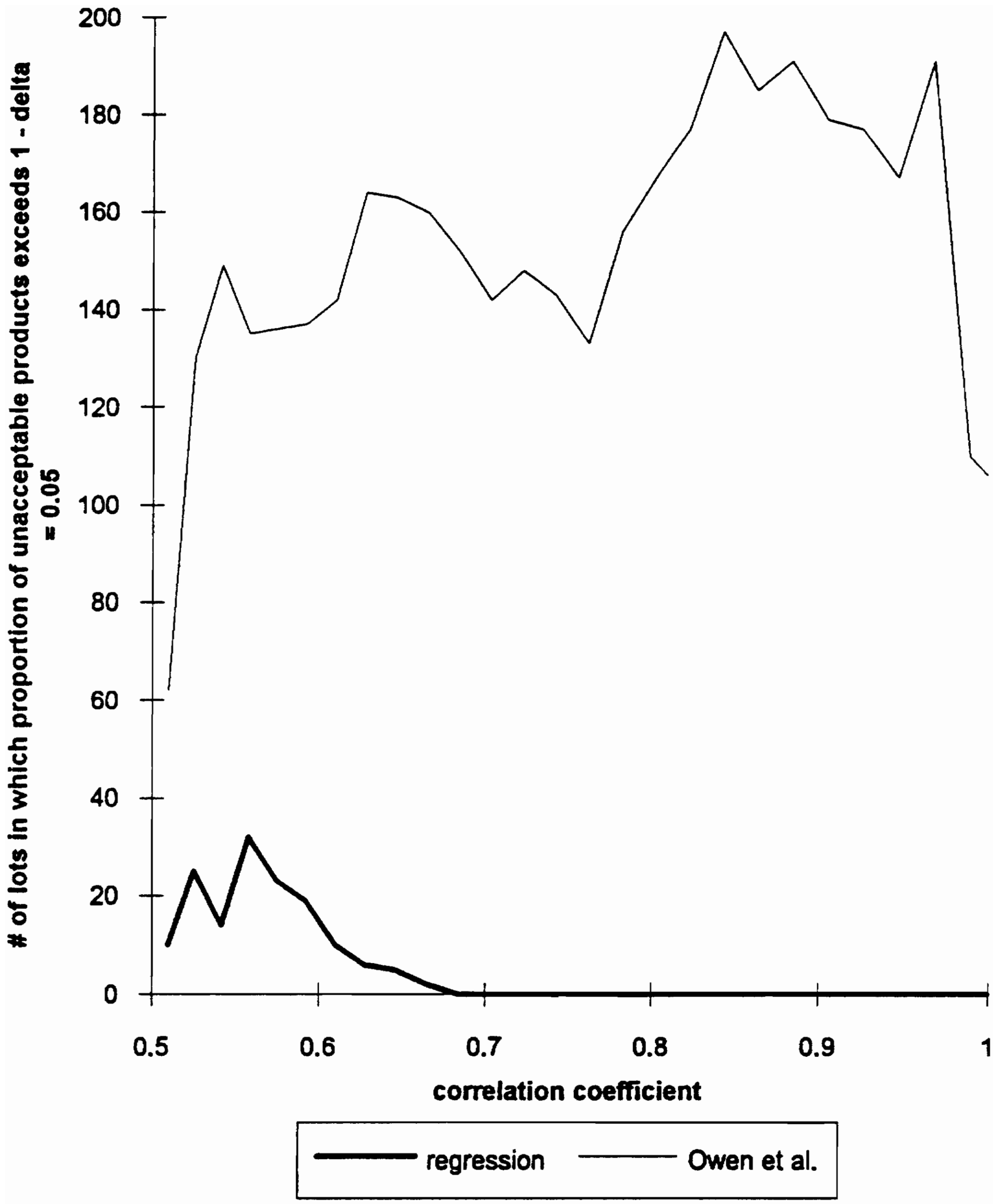

Figure 3.3.3.: $\delta = 0.95$, 2 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.05, out of 300 lots

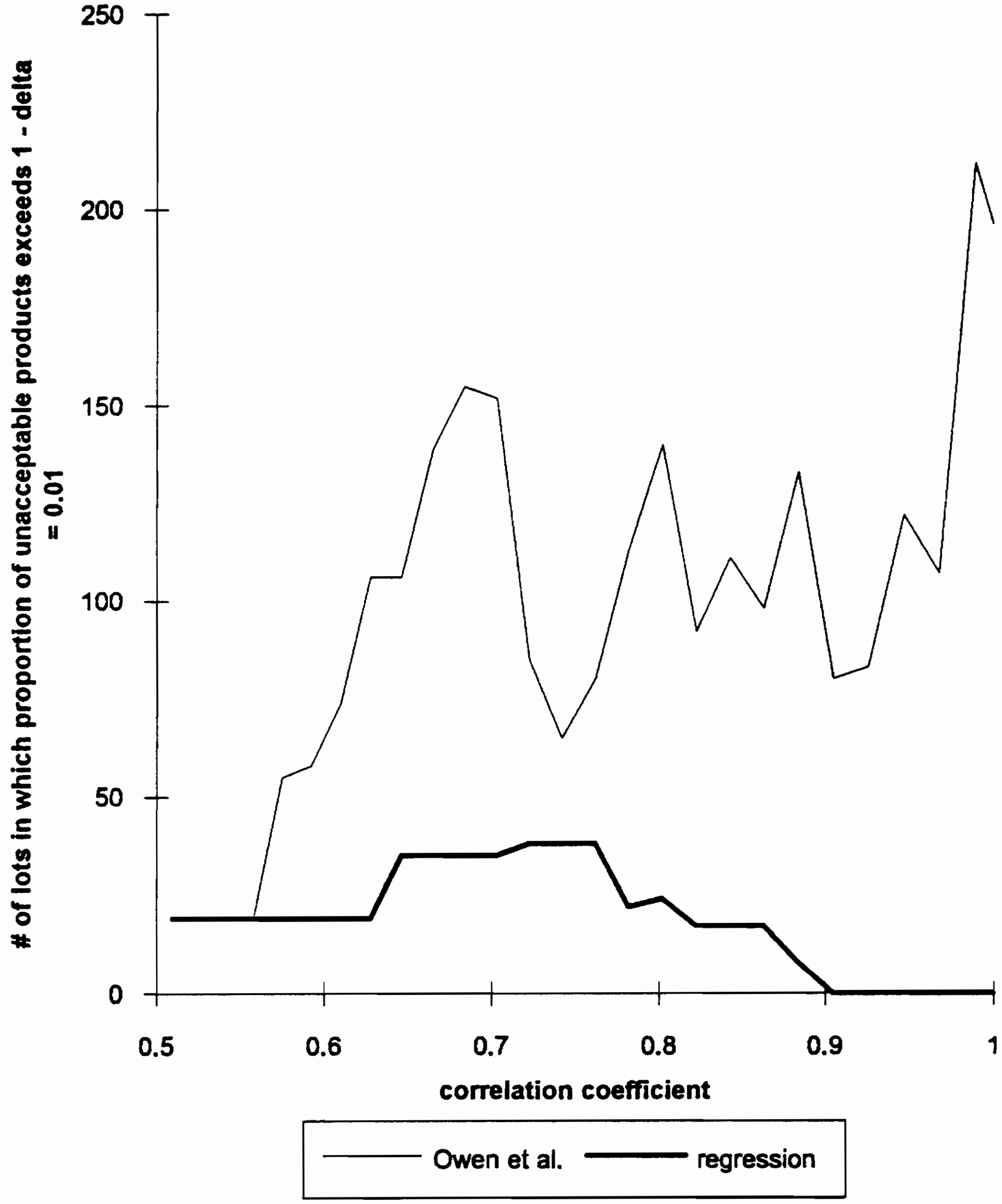

Figure 3.3.4.: $\delta = 0.99$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level 0.01, out of 300 lots

exceeds the required levels of 1 - $\delta = 0.05$ and 1 - $\delta = 0.01$.

A detailed comparison between these two methods is done for the value of the correlation coefficients:

$\rho_{YX_1} = 0.80$

$\rho_{YX_2} = 0.80$

These values of the correlation coefficients give:

$\rho_{YW} = 0.992$ for the regression equation

$\rho_{YV} = 0.975$ for the linear combination of Owen et al.

The summary results from 300 samples of size 600 (results for each sample are given in appendix D):

total number of products	-	180000
total number of acceptable products before screening	-	135802
average proportion of acceptable products before screening	-	$\frac{135802}{180000} = 0.7544$

	Method of Owen et al. δ=0.95	δ=0.99	Regression method δ=0.95	δ=0.99
total number selected	138109	123144	122970	117591

	Method of Owen et al. δ=0.95	δ=0.99	Regression method δ=0.95	δ=0.99
total number of acceptable products in selected proportion	131592	121872	122863	117591
average proportion of acceptable products in the selected proportion	0.953	0.9897	0.9991	1.0
proportion of acceptable products selected	0.9689	0.897	0.904	0.866

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - δ = 0.05:

	Method of Owen et al.	Regression method
$0.05 \leq p < 0.06$	77	0
$0.06 \leq p < 0.075$	34	0
$0.075 \leq p < 0.085$	0	0
$0.085 \leq p$	0	0
total (out of 300)	111	0

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - δ = 0.01:

	Method of Owen et al.	Regression method
$0.01 \leq p < 0.015$	87	0
$0.015 \leq p < 0.02$	32	0
$0.02 \leq p < 0.025$	6	0
$0.025 \leq p$	0	0
total (out of 300)	125	0

Comments:

Using a linear combination of two screening variables, X_1 and X_2, the screening of products is greatly improved. If the required proportion of acceptable products is $\delta = 0.99$, the proportion of acceptable products selected is increased from 0.399 to 0.897 using the method of Owen et al. and from 0.243 to 0.866 when the regression method is used. For $\delta = 0.95$, the increase is from 0.7168 to 0.9689 and from 0.456 to 0.904 respectively. The improvement is due to higher correlation between linear combinations of screening variables and the performance variable Y. Both correlation coefficients, $\rho_{YV} = 0.975$ and $\rho_{YW} = 0.992$, are much higher than ρ_{YW} $= 0.8$ when one screening variable is used.

The proportion of selected products is slightly higher when the method of Owen et al. is used but the required levels of 0.05 and 0.01 of nonconforming products are exceeded in 37% and 41.7% of all lots which is unacceptable. Therefore, the regression method which guarantees at least the required proportion of acceptable products in every single lot has a clear advantage over the method of Owen et al., and the advantage remains for all $\rho_{YV} \geq 0.8$ and $\rho_{YW} \geq 0.8$ (Figures 3.3.1. - 3.3.4.).

3.3.2. Unknown Parameters Case

If the parameters of the underlying multivariate normal distribution are not known, the decision rules are no longer valid. Replacing the parameters with their estimates makes the proportion selected a random variable. Again, the tolerance interval is needed. The regression method is the only method applicable in this case.

3.3.2.1. Regression Method

In this section, the case in which the parameters are not known is considered. The parameters are estimated using the method of maximum likelihood. All the products for which a lower one-sided tolerance limit, $((L(\delta),\infty)$, is above an acceptance limit are considered acceptable. Hence, the same methodology is used as for the single screening variable case described in section **3.2.2.2.**.

The same 300 samples are used as for the known parameters case. The parameters and regression coefficients are estimated from samples of size 40 drawn independently from each sample (lot) of size 600. Therefore, the screening of each lot is performed using the linear regression function with coefficients estimated from the sample from that lot. Evaluation of the regression procedure is done for two proportions of acceptable products after screening, δ=0.95 and δ=0.99 with two levels of confidence, η=0.95 and η=0.99.

Summary results from 300 samples of size 600 (detailed results are given in appendix E):

	γ=0.95		γ=0.99	
	δ=0.95	δ=0.99	δ=0.95	δ=0.99
total number selected	109460	104992	109460	104992
total number of acceptable products in selected proportion	109447	104991	109447	104991
average proportion of acceptable products in the selected proportion	0.9998	0.9999	0.9998	0.9999
proportion of acceptable products selected	0.8059	0.7731	0.8059	0.7731

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - δ = 0.05:

$0.05 \leq p < 0.06$	0
$0.06 \leq p < 0.075$	0
$0.075 \leq p < 0.085$	0
$0.085 \leq p$	0
total (out of 300)	0

actual confidence level = $1 - \frac{0}{300} = 1.0 > \eta = 0.99$

The number of samples (lots) in which proportion of not acceptable products exceeds nominal level of 1 - δ = 0.01

$0.01 \leq p < 0.015$	0
$0.015 \leq p < 0.02$	0
$0.02 \leq p < 0.025$	0
$0.025 \leq p$	0
total (out of 300)	0

actual confidence level = 1 - $\frac{0}{300} = 1.0 > \eta = 0.99$.

Comments:

The procedure performs extremely well. The proportions of acceptable products selected are 0.8059 and 0.7731. Both the confidence level and the proportion of acceptable products greatly exceed required levels. The improvement is due to the increased value of the correlation coefficient. It is interesting to note that confidence level η is 0.99 for $\rho_{YW} > 0.83$, and for $\rho_{YW} > 0.92$, $\eta = 1.0$ (figure 3.3.6.).

3.3.2.2. Selection of Screening Variables

The regression method can be easily extended to cover any number of screening variables. All previous procedures and rules remain the same.

If many screening variables are available and the parameters of the underlying multivariate distribution are not known, the standard methods can be used to select only those variables which significantly increase ρ_{YW}. These standard methods are (Graybill, 1976):

- all regression method
- forward selection method
- backward elimination method.

All of these methods are based on testing the significance of the contribution of an added or removed variable to a multiple correlation coefficient ρ_{YW}. Stepwise regression can be added to the above methods (Draper and Smith, 1981).

From the simulation study, it follows that a great increase in the number of products selected is achieved for any increase in the multiple correlation coefficient, especially for $\rho_{YW} > 0.95$ (figures 3.3.5. and 3.3.6.). Therefore, it is recommended to keep all the screening variables in the model regardless of how small the contribution of any variable is, or at least to use a very large p-value, say 0.5, for the level of significance. The decision for the number of screening variables used should be based primarily on economy considering the sample size needed to estimate parameters.

If the parameters are known, all available screening variables should be used.

Colinearity between screening variables is not a problem. Even if colinearity between screening variables is present, it does not have a great effect on the prediction. For the purpose of screening, prediction is made in the range of the data and close to the fitted plane where colinearity does not have a large impact on the stability of the estimated regression coefficients.

If extrapolation is required, colinearity may cause a problem. Colinearity can be detected by analyzing the eigenvalues of the $X'X$ matrix (Myers, 1990) and if needed, biased estimation or principal component regression can be employed.

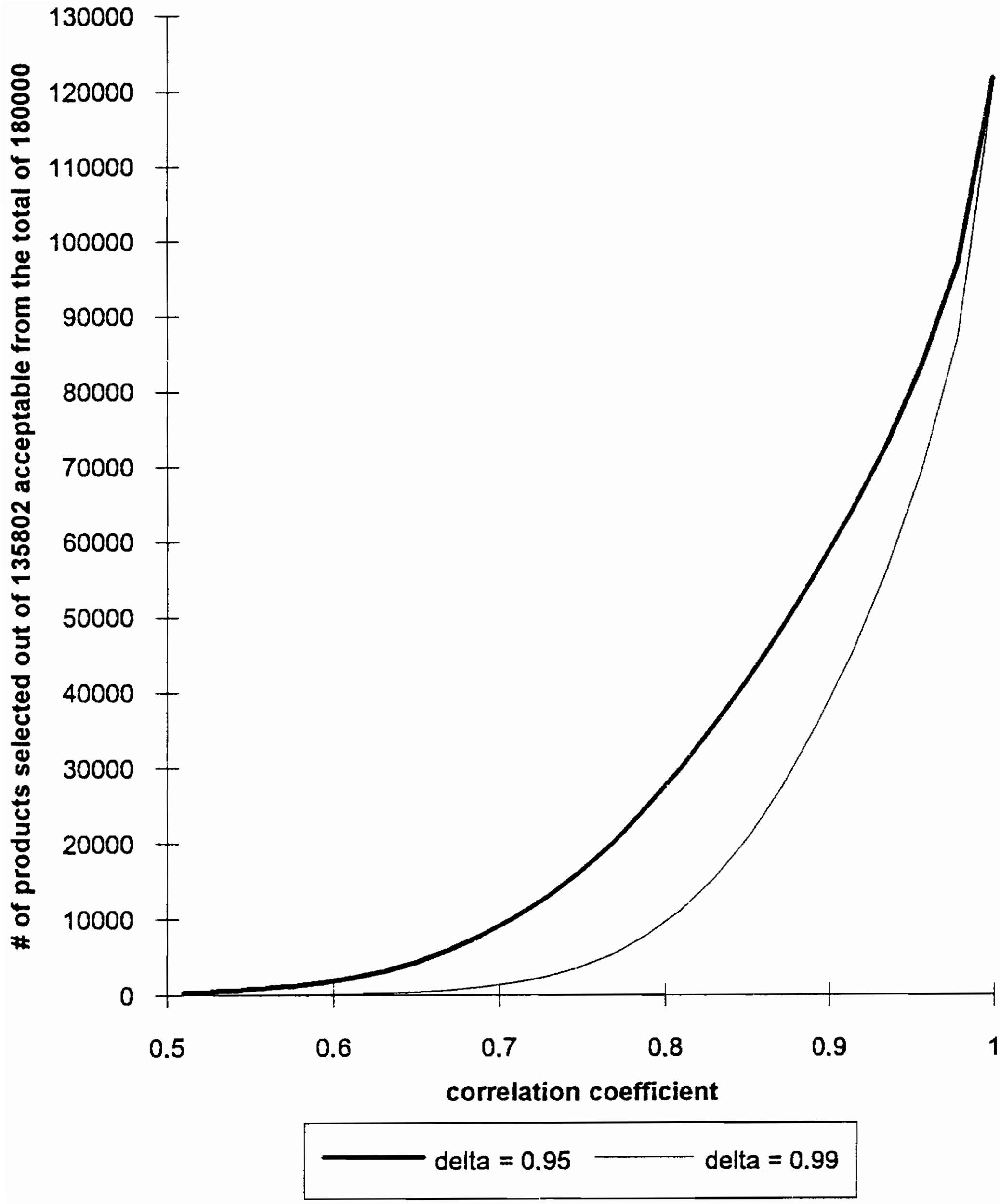

Figure 3.3.5.: $\eta = 0.95$ and $\eta = 0.99$, 2 screening variables

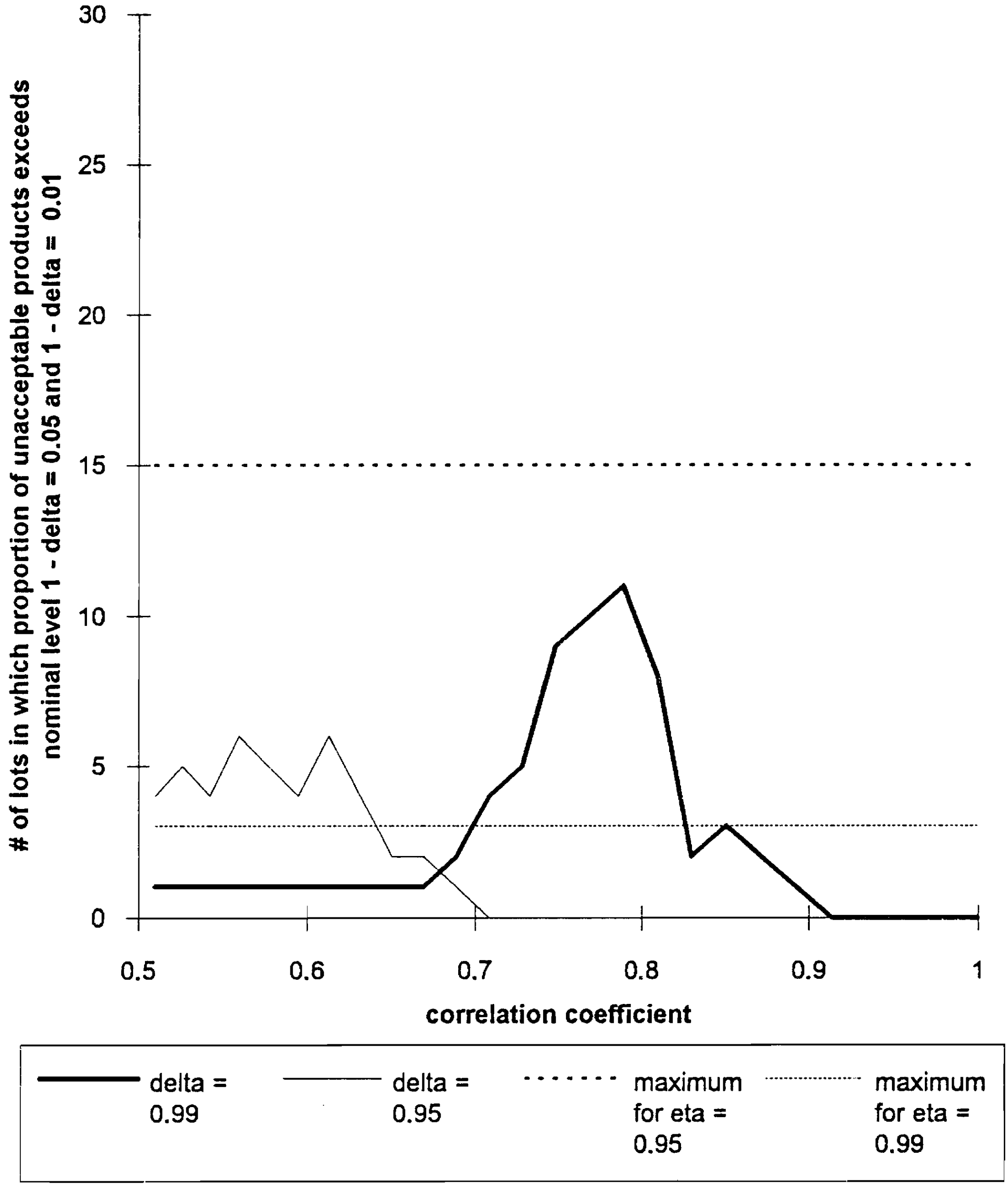

Figure 3.3.6.: $\eta = 0.95$ and $\eta = 0.99$, 1 screening variable, # of lots in which proportion of unacceptable products exceeds nominal level, out of 300 lots

3.3.2.3. Normality Assumption

The regression method does not require that the performance variable and the screening variables be jointly normally distributed. The above analysis is valid if the conditional distribution $W = Y / X_1, X_2, \ldots X_K$ is $N(\beta_0 + \sum_{k=1}^{K} \beta_k x_k, \sigma_W^2)$, the joint marginal distribution $f(\boldsymbol{\theta}, X_1, X_2, \ldots X_K)$ is not known and the parameters $\boldsymbol{\theta}$, $\beta_0, \beta_1, \ldots, \beta_K$ and σ_W^2 are functionally independent. However, in that case, the estimators $b_0, b_1, \ldots, b_K$ and s_W^2 are not necessarily UMVUE estimators (Graybill, 1976) as the properties of estimators depend on an unknown $f(\boldsymbol{\theta}, X_1, X_2, \ldots X_K)$.

CHAPTER 4 - CONCLUSIONS

A screening procedure that ensures that the required proportion of good quality products is contained in each lot is very valuable. In situations in which the performance variable cannot be measured directly, it can be used to separate good from bad quality products either as a primary means of quality assurance or it can be used to "save" good products from the rejected lots if a classical acceptance sampling methodology is employed. This is an alternative to selling products from the whole lot at a discount price, reworking or scrapping the products. The products which do not pass screening can be sold at the discount price, reworked or scraped. The screening procedure can be employed at different stages in a production process.

The major problem with the screening procedures developed so far is that they are too conservative in situations in which parameters of the underlying distributions of the performance and the screening variables are not known (Figures 3.2.5.-3.2.8.). If the parameters are known, they cannot guarantee the required level of quality for each single lot (Figures 3.2.3.,3.2.4., 3.3.3. and 3.3.4.). Furthermore, no procedure has been developed to cover cases in which more than one screening variable is available and the parameters are not known.

The efficiency of the screening procedure heavily depends on the strength of correlation between the performance variable and a single screening variable or a linear combination of screening variables. Also, it is highly unlikely that the parameters of the underlying multivariate distribution are known. Therefore, a model which allows the possibility of increasing correlation, especially in situations when the parameters are not known, is a necessity.

The regression method satisfies the above requirements. In section 3.3.2.1., an evaluation of the regression method is performed. With two screening variables producing a multiple correlation coefficient of 0.992, 77.31% of all products of acceptable quality are selected. Thus, the following probability statement holds:

$$P(\text{proportion of acceptable products } \geq \delta = 0.99) \geq \eta = 0.99 \qquad (4.1.)$$

Moreover, the regression method requires neither special tables nor special software. Tables of the χ^2 and F distributions and regression routines, which are available in almost every spreadsheet are all that is needed.

However, it should be emphasized that dependence of the proportion of acceptable products selected depends heavily on the value of the correlation coefficient. In the application of the regression method, one should try to achieve as high a correlation coefficient ρ_{YW} as possible by adding more screening variables to the model.

4.1. Areas for Future Study

One area of future study is to explore properties of the regression method in the cases in which a performance variable and the screening variables do not follow a multivariate normal distribution or if the conditional distribution $W=Y/X_1,X_2,...X_K$ is not $N(\beta_0+\sum_{k=1}^{K}\beta_k x_k,\sigma_W^2)$. In these situations suitable transformation of the performance variable and/or screening variables to achieve either multivariate normality or normality of conditional distribution may be a remedy. However, the impact of transformation on the efficiency and the quality of screening needs further research.

The next area for future study is to make the confidence level η and the proportion of acceptable products in the selected proportion δ exact. Now, both η and δ are lower limits. By making η and δ exact, more products would be selected leading to a more economical procedure.

REFERENCES

Bai, D.S., Kim, S.B. and Riew, C. (1990). "Economic Screening Procedures Based on Correlated Variables," Metrika, 37, 263-280.

Billingsley, P. (1986). "Probability and Measure," John Wiley & Sons, Inc., New York.

Draper, N.R. and Smith, H. (1980). "Applied Regression Analysis," 2nd ed., John Wiley & Sons, Inc., New York.

Duncan, Acheson J. (1986). "Quality Control and Industrial Statistics," 5th ed., Irwin, Homewood, Illinois.

Graybill, F. A. (1976). "Theory and Application of the Linear Model," Wadswort Publishing Company, Inc., Belmont, California.

Guttman, I. (1970). "Statistical Tolerance Regions: Classical and Bayesian," Charles Griffin & Company Limited, London, England.

Guttman, I. (1970). "Costruction of β-content Tolerance Regions at Confidence Level γ for Large Samples From the k-Variate Normal Distribution," Ann. Math. Statist., 41, 376-400.

Hartler, G. et al. (1992). "Mehrparameter-Screening von Komponenten," Tagungsband zur Forshungstagung Qualitatssicherung 1992, FQS-Schrift.

Jensen, F. and Moltoft, J. (1987). "Reliability Indicator Methods -A Feasibility Study-," Danmarks Ingeniorakademi, Elektroafdeligen (DIA-E), Lyngby, Danmark.

Kocherlakota, S., Kocherlakota, K., and Balakrishnan, N. (1987). "Selection Procedures Based on Normal and Nonnormal Variables with Single and Multiple Characters," Commun. Statist.-Theory Meth., 16(10), 3007-3023.

Law, Averill M. and Kelton, David W. (1991). "Simulation Modeling and Analysis," McGraw-Hill, Inc. New York, New York.

Lehman, E.L. (1991). "Testing Statistical Hypotheses," Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California.

Li, L. and Owen, D.B. (1979). "Two Sided Screening Procedures in the Bivariate Case," Technometrics, Vol. 21, No. 1, 79-85.

Liberman, G.J. and Miller, R.G., Jr. (1963). "Simultaneous Tolerance Intervals in Regression," Biometrika, 50, 1 and 2, 155-168.

Madsen, R.W. (1982). "A Selection Procedure Using a Screening Variate," Technometrics, Vol. 24, No. 4, 301-306.

Mee, R.W. (1990). "An Improved Procedure for Screening Based on Correlated, Normally Distributed Random Variable," Technometrics, Vol. 32, No. 3, 331-337.

Odeh, R.E. and Owen, D.B. (1980). "Tables for Normal Tolerance Limits, Sampling Plans, and Screening," Marcel Dekker, Inc., New York.

Owen, D. B. (1968). "A Survey Of Properties and Applications of the Noncentral t-Distribution," Technometrics, Vol. 7, No. 3, 445-478.

Owen, D.B., McIntire, and D., Seymour, E. (1975). "Tables Using One or Two Screening Variables to Increase Acceptable Product Under One-Sided Specifications," Journal of Quality Technology, Vol. 7, No. 3, 127-138.

Owen, D.B. and Bodie, J.W. (1976). "A screening Method For Increasing Acceptable Product with Some Parameters Unknown," Technometrics, Vol. 18, No. 2, 195-199.

Owen, D.B. and Yueh-ling Hsiao Su (1977). "Screening Based on Normal Variables," Technometrics, Vol. 19, No. 1, 65-68.

Owen, D.B. and Hass, R.W. (1978). "Tables of the Normal Conditioned on t-Distribution," Contribution to Survey Sampling and Applied Statistics, ed. H.A. David, Academic Press, New York.

Owen, D.B., Li, L., and Chou,Y.M. (1981). "Prediction Intervals for Screening Using a Measured Correlated Variate," Technometrics, Vol. 23, No. 2, 165-170.

Paulson, E. (1943). "A Note On Tolerance Limits," Ann. Math. Statist., 14, 90-93.

Seber, G.A.F. (1984). "Multivariate Observations," John Wiley & Sons, Inc., New York.

Shewhart W.A. (1939). "Statistical Method From the Viewpoint of Quality Control," Ed. W. Edwards Deming, The Graduate School, The Department of Agriculture, Washington.

Tang, K. (1987). "Economic Design of a One-Sided Screening Procedure Using a Correlated Variable," Technometrics, Vol. 29, No. 4, 477-485.

Tang, K. (1988). "Economic Design of a Two-Sided Screening Procedure Using a Correlated Variables," Applied Statistics, 37, 2, 231-241.

Tang, J. and Tang, K. (1989). "A Two-Sided Screening Procedure Using Several Correlated Variables," IIE Transactions, Vol. 21, No. 4, 333-336.

Taylor, H.C. and Russel, J.T. (1939). "The Relationship of Validity Coefficients to the Practical Effectiveness of Tests in Selection: Discussion and Tables," Journal of Applied Psychology, 23, 565-578.

Tang, K. and Tang, J. (1994). "Design of Screening Procedures: A Review," Journal of Quality Technology, Vol. 26, No. 3, 209-226.

Thompson, James R. and Koronacki, Jacek (1993). "Statistical Process Control For Quality Improvement," Chapman & Hall, Inc., New York, New York.

Tong, Y.L. (1980). "Probability Inequalities in Multivariate Distributions," New York, Academic Press, Inc.

Tong, Y.L. (1990). "The Multivariate Normal Distribution," New York, Springer-Verlag.

Thomas, J.G., Owen, D.B., and Gunst, R.F. (1977). "Improving the Use of Educational Tests as Selection Tools," Journal of Educational Statistics, Vol. 2, no. 1, 55-77.

Wald, A. (1942). "Setting Tolerance Limits When Sample Size is Large," Ann. Math. Statist., 13, 389-399.

Wallis, W.A. (1951). "Tolerance Intervals for Linear Regression," Proc. Second Berkeley Symp., Berkeley: University of California Press.

Weissberg, A. and Beatty, G.H. (1960). "Tables of Tolerance-Limit Factors for Normal Distributions," Technometrics, Vol. 2, No. 4, 483-500.

Wilks, S.S. (1941). "Determination of Sample Sizes for Setting Tolerance Limit," Ann. Math. Statist., 12, 91-96.

Appendix A

Fortran programs procedures and functions used for random number generation sampling, parameter estimation and comparisons:

1CORR2.FOR - compares procedure of Owen et al. with regression procedure, when parameters of 2-variate normal distribution are known, for value of correlation coefficient from 0.5 to 0.98

1CORR2E.FOR - compares procedure of Odeh and Owen with regression procedure, in the case when parameters are not known, for value of correlation coefficient from 0.5 to 0.98

1CORR3.FOR - compares procedure of Owen et al. with regression procedure in the case when two screening variables are available, and parameters of 3-variate normal distribution are known, for value of correlation coefficient from 0.51 to 0.9991

1CORR3E.FOR - evaluates regression procedure in the case when two screening variables are available, and the parameters are not known.

FUNCTION RAND - generates uniform U(0,1) random numbers, Law and Kelton, (1991)

SUBROUTINE NORM - generates standard normal random variable, using polar method

```
* PROGRAM:1CORR2.FOR
*
*       GENERATES 2-VARIATE NORMAL RANDOM VECTORS WITH
*       CORRELATION COEFFICIENTS FROM 0.5 TO 0.98 BY 0.02
*       CALCULATES REGRESSION COEFFICIENTS, AND STANDARD DEVIATION
*       SELECTS 300 SAMPLES OF SIZE 600
*       COMPARES METHOD OF OWEN ET AL. WITH REGRESSION METHOD, WHEN
*       PARAMETERS ARE KNOWN FOR EACH VALUE OF CORRELATION
*       COEFFICIENT

*       COEFFICIENTS FOR METHOD OF OWEN ET AL., FOR EACH VALUE OF
*       CORRELATION COEFFICIENT (FROM TABLES IN ODEH AND
*       OWEN, 1980) ARE IN FILE OWEN.KOF

*       STREAMS 11 AND 12, FOR RANDOM NUMBER GENARATOR
*       AND STREAM 31 FOR RANDOM SAMPLING


      REAL SIGMA(2,2),T(2,2),STNORM(9650,6),MLTNRM(9650,2),
     +    MEAN(10,1),X1,X2,SUM,YHAT,LCLM,VAR,OWKOF(26,2),
     +    STD,KSE,BO,B1,START,START1,Z,OWC,LY,OWK1,OWK2,
     +    PNREG,PNOWN,E,F,PROB1,PROB2,PROB3,PROB4,
     +    SAMPLE(610,4),TRREG,TWREG,TROWN,TWOWN,DELTA
      INTEGER ROW,COLUMN,S,NUMVAR,START2(610),SAMPLS,SAMS,TOTAL,
     +      NSAMPL,A,B,C,D,TRREG1,TWREG1,TROWN1,TWOWN1,NONA,TNONA,
     +      PREG1,PREG2,PREG3,PREG4,POWN1,POWN2,POWN3,POWN4,
     +      PREG11,PREG22,PREG33,POWN11,POWN22,POWN33,CON

      OPEN (1,FILE = 'B:\OWEN.KOF')
      OPEN (2,FILE = 'B:\ESTIM.COR')
      OPEN (3,FILE = 'C:\EFFIC.COR')
      OPEN (4,FILE = 'B:\TACREG2.COR')
      OPEN (5,FILE = 'B:\TACOWN2.COR')

      ROW    = 2
      COLUMN = 2
      NUMVAR = 600
      SAMPLS = 9650
      NSAMPL = 300

* FOR DELTA 0.95 Z = 1.645,  PROB1 = 0.05, PROB2 = 0.06
*                            PROB3 = 0.075, PROB4 = 0.085

* FOR DELTA 0.99 Z = 2.325,  PROB1 = 0.01, PROB2 = 0.015
*                            PROB3 = 0.02, PROB4 = 0.025

      CONN = 1
      DELTA = 0.95
      Z = 1.645
```

```
      LY = 6.849
      TRREG = 0
      TWREG = 0
      TROWN = 0
      TWOWN = 0
      PROB1 = 0.05
      PROB2 = 0.06
      PROB3 = 0.075
      PROB4 = 0.085
      PREG1 = 0
      PREG2 = 0
      PREG3 = 0
      PREG4 = 0
      POWN1 = 0
      POWN2 = 0
      POWN3 = 0
      POWN4 = 0
      NONA  = 0
      NONA = 0

      DO 2 I = 1,4
        DO 1 J = 1,NSAMPL
        SAMPLE(J,I) = 0.0
1        CONTINUE
2     CONTINUE

      DO 3 M = 1,NSAMPL
        START2(M) = 0
3     CONTINUE

* DEFINE MEAN VECTOR

        MEAN(1,1) = 7.0
        MEAN(2,1) = 10.5

* DEFINE COVARIANCE MATRIX

      SIGMA(1,1)  = 0.05
      SIGMA(1,2)  = 0.08
      SIGMA(2,1)  = 0.08
      SIGMA(2,2)  = 0.2

      DO 4 KK = 1,NSAMPL
        START = (RAND(31))*9
        START1 = START*1000
        START2(KK) = INT(START1)
4     CONTINUE

* GENERATE 6-VARIATE STANDARD NORMAL DISTRIBUTION
      DO 6 K = 1,SAMPLS
```

```
      DO 5 J = 1,3
      CALL NORM(X1,X2)
      STNORM(K,J) = X1
      S = J + 3
      STNORM(K,S) = X2
5      CONTINUE
6    CONTINUE

    DO 9 JIJ = 1,25
      READ(1,7) OWK1,OWK2
7      FORMAT(F7.4,2X,F7.4)
      OWKOF(JIJ,1) = OWK1
      WKOF(JIJ,2) = OWK2
      WRITE(2,8) OWKOF(JIJ,1),OWKOF(JIJ,2)
8      FORMAT(F7.4,2X,F7.4)
9    CONTINUE

    DO 200 NKN = 1,25
      OWC = 10.5 - (OWKOF(NKN,CON))*0.447213
      NKN1 = NKN - 1
      PRINT *,OWC,NKN1
      SIGMA(1,2) = 0.05 + NKN1*0.002
      SIGMA(2,1) = SIGMA(1,2)
      CORR = (SIGMA(1,2))/(SQRT((SIGMA(1,1))*(SIGMA(2,2))))
      VAR = (SIGMA(1,1)) - ((SIGMA(1,2))**2)/(SIGMA(2,2))
      STD = SQRT(VAR)
      BO = (MEAN(1,1)) - (((SIGMA(1,2))/(SIGMA(2,2)))*(MEAN(2,1)))
      B1 = (SIGMA(1,2))/(SIGMA(2,2))
      KSE = STD*Z

      WRITE(2,11) SIGMA(1,2),CORR,VAR,BO,B1,KSE
11    FORMAT(F6.4,2X,F6.4,2X,F6.4,2X,F6.4,2X,F6.4,2X,F6.4)

      DO 20 I = 1,ROW
        DO 15 K = 1,COLUMN
15        T(K,I) = 0.0
20    CONTINUE

      T(1,1)  = SQRT(SIGMA(1,1))

      DO 30 I = 2,COLUMN
30        T(1,I) = (SIGMA(1,I))/(T(1,1))

      T(2,2) = SQRT((SIGMA(2,2)) - ((T(1,2))**2))

* GENARATE 2-VARIATE NORMAL DISTRIBUTION WITH GIVEN
* MEAN VECTOR AND GIVEN COVARIANCE MATRIX

      SUM = 0.0
```

```
      DO 75 K = 1,SAMPLS
         DO 70 I = 1,ROW
            DO 65 J = 1,I
65             SUM = SUM + (T(J,I))*(STNORM(K,J))
               MLTNRM(K,I) = (MEAN(I,1)) + SUM
               SUM = 0.0
70          CONTINUE
75       CONTINUE

      WRITE(3,80) DELTA
80    FORMAT('DELTA = ',F4.3,/)


      DO 150 KKK = 1,NSAMPL
         MMM = START2(KKK)
         SAMS = MMM + NUMVAR - 1

* COMPARE METHOD OF OWEN ET AL. WITH REGRESSION METHOD

      DO 100 J = (START2(KKK)),SAMS
         YHAT = BO + B1*(MLTNRM(J,2))
         LCLM = YHAT - KSE

         IF((MLTNRM(J,1)).LT.LY) THEN
            NONA = NONA + 1
         END IF

         IF(LCLM.GE.LY) THEN
            SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
               IF((MLTNRM(J,1)).LT.LY) THEN
               SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.O
            END IF
         END IF

         IF((MLTNRM(J,2)).GE.OWC) THEN
            SAMPLE(KKK,3) = (SAMPLE(KKK,3)) + 1.0
            IF((MLTNRM(J,1)).LT.LY) THEN
               SAMPLE(KKK,4) = (SAMPLE(KKK,4)) + 1.0
            END IF
          END IF

100   CONTINUE

      IF((SAMPLE(KKK,1)).LE.0.2) THEN
         E = 0.0
      ELSE
         E = (SAMPLE(KKK,2))/(SAMPLE(KKK,1))
      END IF

      IF((SAMPLE(KKK,3)).LE.0.2) THEN
```

```
   F = 0.0
ELSE
   F = (SAMPLE(KKK,4))/(SAMPLE(KKK,3))
END IF

A = INT(SAMPLE(KKK,2))
B = INT(SAMPLE(KKK,1))
C = INT(SAMPLE(KKK,4))
D = INT(SAMPLE(KKK,3))

IF(E.GT.PROB1) THEN
   PREG1 = PREG1 + 1
END IF

IF(E.GT.PROB2) THEN
   PREG2 = PREG2 + 1
END IF

IF(E.GT.PROB3) THEN
   PREG3 = PREG3 + 1
END IF

IF(E.GT.PROB4) THEN
   PREG4 = PREG4 + 1
END IF

IF(F.GT.PROB1) THEN
   POWN1 = POWN1 + 1
END IF

IF(F.GT.PROB2) THEN
   POWN2 = POWN2 + 1
END IF

IF(F.GT.PROB3) THEN
   POWN3 = POWN3 + 1
END IF

IF(F.GT.PROB4) THEN
    POWN4 = POWN4 + 1
 END IF

     WRITE(3,102) KKK,NONA,B,A,E,D,C,F
102  FORMAT(1X,I3,5X,I5,5X,I5,5X,I5,7X,F6.4,5X,I5,5X,I5,7X,F6.4)

TRREG = TRREG + (SAMPLE(KKK,1))
TWREG = TWREG + (SAMPLE(KKK,2))
TROWN = TROWN + (SAMPLE(KKK,3))
TWOWN = TWOWN + (SAMPLE(KKK,4))
TNONA = TNONA + NONA
```

```
      NONA = 0

150   CONTINUE

      PNREG = TWREG/TRREG
      PNOWN = TWOWN/TROWN
      TRREG1 = INT(TRREG)
      TWREG1 = INT(TWREG)
      TROWN1 = INT(TROWN)
      TWOWN1 = INT(TWOWN)
      TOTAL = NUMVAR*NSAMPL
      PREG11 = PREG1 - PREG2
      PREG22 = PREG2 - PREG3
      PREG33 = PREG3 - PREG4
      POWN11 = POWN1 - POWN2
      POWN22 = POWN2 - POWN3
      POWN33 = POWN3 - POWN4

      WRITE(3,151)CORR,TOTAL,TNONA,TRREG1,TWREG1,PNREG,PREG11,PREG22
     +        PREG33,PREG4,TROWN1,TWOWN1,PNOWN,POWN11,POWN22,POWN33,
     +        POWN4
151   FORMAT(/'CORRELATION COEFFICIENT                 ',F6.4,
     +      /'TOTAL # 0F PRODUCTS                 ',I6,
     +      /'TOTAL # OF NOT ACCEPTABLE PRODUCTS      ',I6,
     +     //'TOTAL # SELECTED BY REG             ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY REG.     ',I5,
     +      /'PROP. OF NOT ACCEPTABLE IN SEL. PROP.   ',F8.6,
     +      /'TOTAL # OF NOT ACCEPTABLE EXCEEDING:',
     +      /'(0.01  < P <= 0.015), 0.05  < P <= O.06  ',I5,
     +      /'(0.015 < P <= 0.02 ), 0.06  < P <= 0.075 ',I5,
     +      /'(0.02  < P <= 0.025), 0.75  < P <= 0.085 ',I5,
     +      /'(O.O25 < P          , 0.085 < P          ',I5,
     +     //'TOTAL # SELECTED BY OWEN              ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY OWEN      ',I5,
     +      /'PROP. OF NOT ACCEPTABLE IN SEL. PROP.   ',F8.6,
     +      /'TOTAL # OF NOT ACCEPTABLE EXCEEDING:',
     +      /'(0.01  < P <= 0.015), 0.05  < P <= O.06  ',I5,
     +      /'(0.015 < P <= 0.02 ), 0.06  < P <= 0.075 ',I5,
     +      /'(0.02  < P <= 0.025), 0.75  < P <= 0.085 ',I5,
     +      /'(O.O25 < P          , 0.085 < P          ',I5)

      WRITE(4,180)CORR,TRREG1,PNREG,PREG1
180   FORMAT(F6.4,',',I6,',',F5.4,',',I5)

      WRITE(5,181)CORR,TROWN1,PNOWN,POWN1
181   FORMAT(F6.4,',',I6,',',F5.4,',',I5)

      DO 190 IKI = 1,4
         DO 185 JKJ = 1,NSAMPL
            SAMPLE(JKJ,IKI) = 0.0
```

```
185      CONTINUE
190    CONTINUE

       NONA  = 0
       TNONA = 0
       TRREG = 0.0
       TWREG = 0.0
       TROWN = 0.0
       TWOWN = 0.0
       PREG1 = 0
       PREG2 = 0
       PREG3 = 0
       PREG4 = 0
       POWN1 = 0
       POWN2 = 0
       POWN3 = 0
       POWN4 = 0

200  CONTINUE

     CLOSE (2)
     CLOSE (3)
     CLOSE (4)
     CLOSE (5)

    STOP
    END
```

```
* PROGRAM: 1CORR2E.FOR
*
*        GENERATES 2-VARIATE NORMAL RANDOM VECTORS
*        FOR CORRELATION COEFFICIENT FROM 0.5 TO 0.98 BY 0.02
*        SELECTS 300 SAMPLES (LOTS) OF SIZE 600
*        SELECTS SAMPLE OF SIZE 40 WITHIN EACH SELECTED SAMPLE (LOT)
*        ESTIMATES PARAMETERS FROM EACH SAMPLE OF SIZE 40
*        COMPARES METHOD OF OWEN ET AL. WITH REGRESSION METHOD,
*        WHEN PARAMETERS ARE NOT KNOWN

*        COEFFICIENTS FOR METHOD OF ODEH AND OWEN, FOR EACH VALUE OF
*        CORRELATION COEFFICIENT, FOR EACH ONE OF 300 SAMPLES, (FROM
*        TABLES IN ODEH AND OWEN, 1980) ARE IN FILES T9595.ET2, T9995.ET2,
*        T9599.ET2 AND T9999.ET2


*        STREAMS 11 AND 12 FOR 300 RANDOM NUMBER GENERATOR
*        STREAM 31 FOR RANDOM SAMPLING FOR 300 SAMPLES OF SIZE 600
*        STREAM 41 FOR RANDOM SAMPLING FOR SAMPLES OF SIZE 40
*        WITHIN EACH OF 300 SAMPLES OF SIZE 600


      REAL SIGMA(2,2),T(2,2),STNORM(9650,6),MLTNRM(9650,2),
     +    MEAN(10,1),X1,X2,SUM,SUM1,SUM2,SUM3,SUM4,SUM5,
     +    YHAT,LCLM,ESTIM(300,7),START3(610),SMLSS1,GAMMA,DELTA,
     +    STD1,BO,B1,START,STARTS,Z,OWC,LY,KOF2,TB1(7500),T1,
     +    PNREG,PNOWN,E,F,PROB1,PROB2,PROB3,PROB4,TB95,TB99,
     +    SAMPLE(610,4),TRREG,TWREG,TROWN,TWOWN,VAR,FD,CHISQ,TB,KOF1
      INTEGER ROW,COLUMN,S,NUMVAR,START2(610),SAMPLS,
     +      NSAMPL,A,B,C,D,TRREG1,TWREG1,TROWN1,TWOWN1,NONA,TNONA,
     +      PREG1,PREG2,PREG3,PREG4,POWN1,POWN2,POWN3,POWN4,
     +      SMLSS,SAMS,SAMS1,SIG,TOTAL,PREG11,PREG22,PREG33,
     +      POWN11,POWN22,POWN33

      OPEN (2,FILE = 'C:\LAZAR\ESTIMATE.ET2')
      OPEN (3,FILE = 'C:\LAZAR\KA99ET2.ET2')
      OPEN (4,FILE = 'C:\LAZAR\T9999.ET2')
      OPEN (7,FILE = 'C:\LAZAR\RET299.ET2')
      OPEN (8,FILE = 'C:\LAZAR\OET299.ET2')

* FOR GAMMA = 0.95, DELTA 0.95:
*              FD = 4.0713, Z = 1.645, FILE #4 = T9595.ET2

* FOR GAMMA = 0.95, DELTA = 0.99:
*              FD = 4.0713,  Z = 2.325, FILE #4 = T9599.ET2
```

```
* FOR GAMMA = 0.99, DELTA = 0.95:
*                 FD =  6.1108, Z = 1.645, FILE #4 = T9995.ET2
* FOR GAMMA = 0.99, DELTA = 0.99:
*                 FD = 6.1108, Z = 2.325, FILE #4 = T9999.ET2

      FD = 6.1108
      Z = 2.325
      LY = 6.849

      ROW    = 2
      COLUMN = 2
      NUMVAR = 600
      SAMPLS = 9650
      NSAMPL = 300
      SMLSS1 = 40.0
      SMLSS = INT(SMLSS1)

      IF(Z.LE.2.0)THEN
         DELTA = 0.95
         PROB1 = 0.05
         PROB2 = 0.06
         PROB3 = 0.075
         PROB4 = 0.085
      ELSE
         DELTA = 0.99
         PROB1 = 0.01
         PROB2 = 0.015
         PROB3 = 0.02
         PROB4 = 0.025
      END IF

      IF(FD.LE.5.0)THEN
         CHISQ = 22.879
         GAMMA = 0.95
      ELSE
         CHISQ = 19.289
         GAMMA = 0.99
      END IF

      SIG = 1
      OWC = 0.0
      TRREG = 0
      TWREG = 0
      TROWN = 0
      TWOWN = 0
      PREG1 = 0
      PREG2 = 0
      PREG3 = 0
      PREG4 = 0
```

```
      POWN1 = 0
      POWN2 = 0
      OWN3 = 0
      POWN4 = 0
      NONA = 0
      TNONA = 0
      BO = 0.0
      B1 = 0.0

      DO 2 I = 1,4
        DO 1 J = 1,NSAMPL
          SAMPLE(J,I) = 0
1       CONTINUE
2     CONTINUE

      DO 3 M = 1,NSAMPL
        START2(M) = 0
3     CONTINUE

* DEFINE MEAN VECTOR

      MEAN(1,1) = 7.0
      MEAN(2,1) = 10.5

* DEFINE COVARIANCE MATRIX

      SIGMA(1,1)  = 0.05
      SIGMA(1,2)  = 0.08
      SIGMA(2,1)  = 0.08
      SIGMA(2,2)  = 0.2

      DO 6 K = 1,300
        DO 5 J = 1,6
          ESTIM(K,J) = 0.0
5       CONTINUE
6     CONTINUE

      DO 7 NN = 1,7500
        TB1(NN) = 0.0
7     CONTINUE


* READ MULTIPLICATIVE COEFFICIENTS FOR METHOD OF ODEH AND OWEN
* FROM FILE #4

      DO 12 NNN = 1,7500
        READ(4,10) T1
10      FORMAT(F8.5)
        TB1(NNN) = T1
```

```
12  CONTINUE

    DO 20 KK = 1,NSAMPL
      START = (RAND(31))*9000
      START2(KK) = INT(START)
      STARTS = (RAND(41))*550
      START3(KK) = INT(STARTS)
20  CONTINUE

* GENERATE 6-VARIATE STANDARD NORMAL DISTRIBUTION

    DO 30 K = 1,SAMPLS
      DO 25 J = 1,3
        CALL NORM(X1,X2)
        STNORM(K,J) = X1
        S = J + 3
        STNORM(K,S) = X2
25    CONTINUE
30  CONTINUE

    DO 200 NKN = 1,25
      NKN1 = NKN - 1
      SIGMA(1,2) = 0.05 + NKN1*0.002
      SIGMA(2,1) = SIGMA(1,2)
      CORR = (SIGMA(1,2))/(SQRT((SIGMA(1,1))*(SIGMA(2,2))))
      NRN = NKN1*300
      PRINT *,NKN,CORR,NRN

      DO 40 I = 1,ROW
        DO 35 K = 1,COLUMN
35        T(K,I) = 0.0
40    CONTINUE

      T(1,1)  = SQRT(SIGMA(1,1))

      DO 45 I = 2,COLUMN
        T(1,I) = (SIGMA(1,I))/(T(1,1))
45    CONTINUE

      T(2,2) = SQRT((SIGMA(2,2)) - ((T(1,2))**2))

* GENERATE 2-VARIATE NORMAL DISTRIBUTION WITH GIVEN
* MEAN VECTOR AND GIVEN COVARIANCE MATRIX

      SUM = 0.0

      DO 75 K = 1,SAMPLS
        DO 70 I = 1,ROW
          DO 65 J = 1,I
```

```
65              SUM = SUM + (T(J,I))*(STNORM(K,J))
              MLTNRM(K,I) = (MEAN(I,1)) + SUM
             SUM = 0.0
70         CONTINUE
75      CONTINUE

        WRITE(3,80) GAMMA,DELTA
80       FORMAT('GAMMA = ',F4.3,'  DELTA = ',F4.3,/)

* CALCULATE STATISTICS FOR EACH SAMPLE OF SIZE 600

         DO 150 KKK = 1,NSAMPL
           NRN1 = KKK + NRN
           MMM = START2(KKK)
           MMMS = START3(KKK)
           SAMS1 = MMM + MMMS + SMLSS - 1
           MMMS1 = MMM + MMMS
           SAMS = MMM + NUMVAR - 1
           TB = TB1(NRN1)
           SUM1 = 0.0
           SUM2 = 0.0
           SUM3 = 0.0
           SUM4 = 0.0
           SUM5 = 0.0

* SELECT SMALL SAMPLE OF SIZE 40 FROM BIG SAMPLE OF SIZE 600
* AND CALCULATE STATISTICS FROM EACH SMALL SAMPLE

           DO 90 J = MMMS1, SAMS1
              SUM1 = SUM1 + ((MLTNRM(J,1))*(MLTNRM(J,2)))
              SUM2 = SUM2 + (MLTNRM(J,1))
              SUM3 = SUM3 + (MLTNRM(J,2))
              SUM4 = SUM4 + ((MLTNRM(J,1))**2)
              SUM5 = SUM5 + ((MLTNRM(J,2))**2)
90         CONTINUE

           ESTIM(KKK,1) = SUM2/SMLSS
           ESTIM(KKK,2) = (SUM4 - (SMLSS*(ESTIM(KKK,1))**2))/(SMLSS - 1)
           ESTIM(KKK,3) = SUM3/SMLSS
           ESTIM(KKK,4) = (SUM5 - (SMLSS)*((ESTIM(KKK,3))**2))/(SMLSS - 1)
           ESTIM(KKK,5) = (SUM1 - (SMLSS)*(ESTIM(KKK,1))*(ESTIM(KKK,3)))
        +              /(SMLSS - 1)
           ESTIM(KKK,6) = (ESTIM(KKK,5))/(SQRT( (ESTIM(KKK,2))*(ESTIM(KKK,4)
        +              )))
           ESTIM(KKK,7) = ((ESTIM(KKK,1)) - LY)/(SQRT((ESTIM(KKK,2))))

* CALCULATE REGRESSION COEFFICIENTS FOR EACH SAMPLE OF 600 FRQM
* SMALL SAMPLE OF 40
```

```
      B1 = (ESTIM(KKK,5))/(ESTIM(KKK,4))
      BO = (ESTIM(KKK,1)) - B1*(ESTIM(KKK,3))
      VAR = ((ESTIM(KKK,2))*(SMLSS - 1)
     +        - (((ESTIM(KKK,5))*(SMLSS - 1))**2)
     +      /((ESTIM(KKK,4))*(SMLSS - 1)))/(SMLSS - 2)
      STD1 = SQRT(VAR)
      KOF1 = SQRT(2*FD)
      KOF2 = SQRT((SMLSS - 2)/(CHISQ))

* OWEN DECISION FUNCTION

      KOFOW = (SQRT((ESTIM(KKK,4))))*TB*(SQRT((SMLSS1 + 1.0)/SMLSS1))
      OWC = ESTIM(KKK,3) - KOFOW

*   COMPARE EACH VALUE IN THE SAMPLE OF 600 EXCLUDING 40 FOR SMALL
*   SAMPLE WITH DECISION RULE OF ODEH AND OWEN AND REGRESSION RULE

      MMMM = MMMS1 - 1

      DO 100 J = MMM,MMMM
         YHAT = BO + B1*(MLTNRM(J,2))
         LCLM = YHAT - STD1*(KOF1*(SQRT(1/SMLSS + ((MLTNRM(J,2))
     +      - (ESTIM(KKK,3)))**2)/((ESTIM(KKK,4))*(SMLSS - 1)))
     +      + Z*KOF2)

         IF((MLTNRM(J,1)).LT.LY) THEN
            NONA = NONA + 1
         END IF

         IF(LCLM.GE.LY) THEN
            SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
            IF((MLTNRM(J,1)).LT.LY) THEN
               SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.O
            END IF
         END IF

         IF((MLTNRM(J,2)).GE.OWC) THEN
            SAMPLE(KKK,3) = (SAMPLE(KKK,3)) + 1.0
            IF((MLTNRM(J,1)).LT.LY) THEN
               SAMPLE(KKK,4) = (SAMPLE(KKK,4)) + 1.0
            END IF
         END IF
100    CONTINUE

      IIII = SAMS1 + 1

      DO 101 II = IIII, SAMS
         YHAT = BO + B1*(MLTNRM(II,2))
         LCLM = YHAT - STD1*( KOF1*( SQRT(1/SMLSS + ((MLTNRM(II,2))
```

```
     +        - (ESTIM(KKK,3)))**2) /( (ESTIM(KKK,4))*
     +      (SMLSS - 1))) + Z*KOF2)
       IF((MLTNRM(II,1)).LT.LY) THEN
          NONA = NONA + 1
       END IF

       IF(LCLM.GE.LY) THEN
          SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
          IF((MLTNRM(II,1)).LT.LY) THEN
             SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.0
          END IF
       END IF

       IF((MLTNRM(II,2)).GE.OWC) THEN
          SAMPLE(KKK,3) = (SAMPLE(KKK,3)) + 1.0
          IF((MLTNRM(II,1)).LT.LY) THEN
             SAMPLE(KKK,4) = (SAMPLE(KKK,4)) + 1.0
          END IF
       END IF
101      CONTINUE

    IF((SAMPLE(KKK,1)).LT.0.5) THEN
       E = 0.0
    ELSE
       E = (SAMPLE(KKK,2))/(SAMPLE(KKK,1))
    END IF

    IF((SAMPLE(KKK,3)).LT.0.5) THEN
       F = 0.0
    ELSE
       F = (SAMPLE(KKK,4))/(SAMPLE(KKK,3))
    END IF

    A = INT(SAMPLE(KKK,2))
    B = INT(SAMPLE(KKK,1))
    C = INT(SAMPLE(KKK,4))
    D = INT(SAMPLE(KKK,3))

    IF(E.GT.PROB1) THEN
       PREG1 = PREG1 + 1
    END IF

    IF(E.GT.PROB2) THEN
       PREG2 = PREG2 + 1
    END IF

    IF(E.GT.PROB3) THEN
       PREG3 = PREG3 + 1
    END IF
```

```
      IF(E.GT.PROB4) THEN
        PREG4 = PREG4 + 1
      END IF

      IF(F.GT.PROB1) THEN
        POWN1 = POWN1 + 1
      END IF

      IF(F.GT.PROB2) THEN
        POWN2 = POWN2 + 1
      END IF

      IF(F.GT.PROB3) THEN
        POWN3 = POWN3 + 1
      END IF

      IF(F.GT.PROB4) THEN
        POWN4 = POWN4 + 1
      END IF


      IF(NKN.EQ.16)THEN
        WRITE(3,103) KKK,NONA,B,A,E,D,C,F
103       FORMAT(1X,I3,5X,I5,5X,I5,5X,I5,7X,F6.4,5X,I5,5X,I5,7X,F6.4)
      END IF

      TRREG = TRREG + (SAMPLE(KKK,1))
      TWREG = TWREG + (SAMPLE(KKK,2))
      TROWN = TROWN + (SAMPLE(KKK,3))
      TWOWN = TWOWN + (SAMPLE(KKK,4))
      TNONA = TNONA + NONA
      NONA = 0
150   CONTINUE

    IF(TRREG.LT.0.5) THEN
      PNREG = 0.0
    ELSE
      PNREG = TWREG/TRREG
    END IF

    IF(TROWN.LT.0.5) THEN
      PNOWN = 0.0
    ELSE
      PNOWN = TWOWN/TROWN
    END IF

    TRREG1 = INT(TRREG)
    TWREG1 = INT(TWREG)
```

```
      TROWN1 = INT(TROWN)
      TWOWN1 = INT(TWOWN)
      TOTAL = NUMVAR*NSAMPL
      PREG11 = PREG1 - PREG2
      PREG22 = PREG2 - PREG3
      PREG33 = PREG3 - PREG4
      POWN11 = POWN1 - POWN2
      POWN22 = POWN2 - POWN3
      POWN33 = POWN3 - POWN4

      WRITE(3,151) CORR,TOTAL,TNONA,TRREG1,TWREG1,PNREG,PREG11,
     +             PREG22,PREG33,PREG4,TROWN1,TWOWN1,PNOWN,POWN11,
     +             POWN22,POWN33,POWN4
151   FORMAT(/'CORRELATION                              ',F5.3,
     +      /'TOTAL # OF PRODUCTS                       ',I6,
     +      /'TOTAL # OF NOT ACCEPTABLE PRODUCTS        ',I6,
     +     //'TOTAL # SELECTED BY REG                   ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY REG        ',I5,
     +      /'PROP. OF NOT ACCEPTABLE IN A SEL. PROP.   ',F8.6,
     +      /'TOTAL # OF NOT ACCEPTABLE EXCEEDING: ',
     +      /'(O.O1  < P <= 0.015), 0.05  < P <= 0.06   ',I5,
     +      /'(O.O15 < P <= 0.02),  0.06  < P <= 0.075  ',I5,
     +      /'(0.02  < P <= 0.025), 0.075 < P <= 0.085  ',I5,
     +      /'(0.025 < P        ),  0.085 < P           ',I5,
     +     //'TOTAL # SELECTED BY OWEN                  ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY OWEN       ',I5,
     +      /'PROP. OF NOT ACCEPTABLE IN A SEL. PROP.   ',F8.6,
     +      /'TOTAL # OF NOT ACCEPTABLE EXCEEDING: ',
     +      /'(O.O1  < P <= 0.015), 0.05  < P <= 0.06   ',I5,
     +      /'(O.O15 < P <= 0.02),  0.06  < P <= 0.075  ',I5,
     +      /'(0.02  < P <= 0.025), 0.075 < P <= 0.085  ',I5,
     +      /'(0.025 < P        ),  0.085 < P           ',I5)


      WRITE(7,180)CORR,TRREG1,PNREG,PREG1
180   FORMAT(F6.4,',',I6,',',F5.4,',',I5)

      WRITE(8,181)CORR,TROWN1,PNOWN,POWN1
181   FORMAT(F6.4,',',I6,',',F5.4,',',I5)

      DO 190 IKI = 1,4
        DO 185 JKJ = 1,NSAMPL
          SAMPLE(JKJ,IKI) = 0.0
185     CONTINUE
190   CONTINUE

      NONA = 0
      TNONA = 0
      TRREG = 0.0
```

```
      TWREG = 0.0
      TROWN = 0.0
      TWOWN = 0.0
      PREG1 = 0
      PREG2 = 0
      PREG3 = 0
      PREG4 = 0
      POWN1 = 0
      POWN2 = 0
      POWN3 = 0
      POWN4 = 0
200 CONTINUE

    CLOSE (7)
    CLOSE (8)
    CLOSE (2)
    CLOSE (3)
    CLOSE (4)

    STOP
    END
```

```
* PROGRAM: 1CORR3.FOR
*
*     FOR CORRELATION COEFFICIENT FROM 0.5 TO 0.98, BY 0.02:
*
*         GENERATES 3-VARIATE NORMAL RANDOM VECTORS
*         SELECTS 300 RANDOM SAMPLES OF SIZE 600
*         CALCULATES REGRESSION COEFFICIENTS AND STANDARD DEVIATION
*         CALCULATES COEFFICIENTS FOR LINEAR FUNCTION OF OWEN ET.AL
*         COMPARES METHOD OF OWEN ET AL. WITH REGRESSION METHOD,
*         WHEN PARAMETERS ARE KNOWN

*         COEFFICIENTS FOR METHOD OF OWEN ET AL., FOR EACH VALUE OF
*         CORRELATION COEFFICIENT (FROM TABLES IN ODEH AND OWEN, 1980)
*         ARE IN FILE OWKOF3.COR


*        STREAMS 11 AND 12 FOR RANDOM NUMBER GENERATOR
*        STREAM 31 FOR RANDOM SAMPLING


      REAL SIGMA(3,3),T(3,3),STNORM(10000,6),MLTNRM(10000,3),
     +    MEAN(10,1),SAMPLE(610,4),
     +    X1,X2,SUM,DT,DET,IS22,IS33,IS23,IS32,
     +    LCI,B0,B1,B2,YHAT,VAR,KSE,STD,V,LCV,A1,A2,RY1SQ,
     +    RY2SQ,R12SQ,DEN,VV,KOEF,SEV,SESQ,RY1,RY2,R12,RYV,
     +    STD2,STD3,RSQR,RR,LY,DELTA,OWKOF(30,2),OWK1,OWK2,
     +    PNREG,PNOWN,E,F,PROB1,PROB2,PROB3,PROB4
      INTEGER ROW,COLUMN,S,NUMVAR,START2(610),SAMPLS,CON,
     +       NSAMPL,A,B,C,D,TRREG1,TWREG1,TROWN1,TWOWN1,NONA,TNONA,
     +       PREG1,PREG2,PREG3,PREG4,POWN1,POWN2,POWN3,POWN4,
     +       SAMS,PREG11,PREG22,PREG33,POWN11,POWN22,
     +       POWN33,TOTAL

      OPEN (5,FILE = 'B:\PROBACOF.COR')
      OPEN (3,FILE = 'B:\EFFIC.COR')
      OPEN (4,FILE = 'B:\OWKOF3.COR')
      OPEN (7,FILE = 'B:\TACREG3.COR')
      OPEN (8,FILE = 'B:\TACOWN3.COR')

      ROW    = 3
      COLUMN = 3
      NUMVAR = 600
      SAMPLS = 9999
      NSAMPL = 300

* FOR DELTA = 0.95 Z = 1.645, CON = 1
*
```

```
*
* FOR DELTA = 0.99 Z = 2.326, CON = 2
   Z     = 1.645
   DELTA = 0.95
   CON   = 1
   LY    = 6.849

   TRREG = 0
   TWREG = 0
   TROWN = 0
   TWOWN = 0

* FOR 0.95 PROB1 = 0.05  FOR 0.99 PROB1 = 0.01
*         PROB2 = 0.06          PROB2 = 0.015
*         PROB3 = 0.075         PROB3 = 0.02
*         PROB4 = 0.085         PROB4 = 0.025

   PROB1 = 0.05
   PROB2 = 0.06
   PROB3 = 0.075
   PROB4 = 0.085

   PREG1 = 0
   PREG2 = 0
   PREG3 = 0
   PREG4 = 0
   POWN1 = 0
   POWN2 = 0
   POWN3 = 0
   POWN4 = 0
   NONA = 0
   TNONA = 0

* DEFINE MEAN VECTOR

   MEAN(1,1) = 7.0
   MEAN(2,1) = 10.5
   MEAN(3,1) = 4.0

* DEFINE COVARIANCE MATRIX

   SIGMA(1,1)  = 0.05
   SIGMA(1,2)  = 0.08
   SIGMA(1,3)  = 0.04732
   SIGMA(2,1)  = 0.08
   SIGMA(2,2)  = 0.2
   SIGMA(2,3)  = 0.035496
   SIGMA(3,1)  = 0.04732
   SIGMA(3,2)  = 0.035496
```

```
      SIGMA(3,3)   = 0.07

      DO 2 I = 1,4
         DO 1 J = 1,NSAMPL
            SAMPLE(J,I) = 0
1        CONTINUE
2     CONTINUE

      DO 3 M = 1,NSAMPL
         START2(M) = 0
3     CONTINUE

      DO 15 M = 1,30
         DO 14 N = 1,2
            OWKOF(M,N) = 0.0
14       CONTINUE
15    CONTINUE

      DO 16 KK = 1,NSAMPL
         START = (RAND(31))*9300
         START2(KK) = INT(START)
16    CONTINUE


      DO 20 M = 1,26
         READ(4,18) OWK1,OWK2
18       FORMAT(F9.6,2X,F9.6)
         OWKOF(M,1) = OWK1
         OWKOF(M,2) = OWK2
         WRITE(3,19) OWKOF(M,1),OWKOF(M,2)
19        FORMAT(F9.6,2X,F9.6)
20    CONTINUE


* GENERATE 6-VARIATE STANDARD NORMAL RANDOM VECTOR

      DO 35 K = 1,SAMPLS
         DO 30 J = 1,3
            CALL NORM(X1,X2)
            STNORM(K,J) = X1
            S = J + 3
            STNORM(K,S) = X2
30        CONTINUE
35    CONTINUE

      WRITE(7,36)DELTA
36    FORMAT('DELTA = ',F4.3,/)

      WRITE(8,37)DELTA
```

```
37  FORMAT('DELTA = ',F4.3,/)

     DO 200 NKN = 1,26
        NKN1 = NKN - 1
        SIGMA(1,2) = 0.01 + NKN1*0.002
        SIGMA(1,3) = 0.03 + NKN1*0.001
        SIGMA(2,1) = SIGMA(1,2)
        SIGMA(3,1) = SIGMA(1,3)

* CALCULATE REGRESSION COEFFICIENTS, AND STANDARD DEVIATION

     DET = (SIGMA(2,2))*(SIGMA(3,3)) - (SIGMA(2,3))*(SIGMA(3,2))
     DT = 1/DET
     IS22 = (SIGMA(3,3))*DT
     IS33 = (SIGMA(2,2))*DT
     IS23 = (-1)*(SIGMA(2,3))*DT
     IS32 = (-1)*(SIGMA(3,2))*DT
     VAR = (SIGMA(1,1)) -  ((SIGMA(1,2))*IS22*(SIGMA(2,1))
    +                      + (SIGMA(1,3))*IS32*(SIGMA(2,1))
    +                      + (SIGMA(1,2))*IS23*(SIGMA(3,1))
    +                      + (SIGMA(1,3))*IS33*(SIGMA(3,1)))
     STD = SQRT(VAR)
     B1 = (SIGMA(1,2))*IS22 + (SIGMA(1,3))*IS32
     B2 = (SIGMA(1,2))*IS23 + (SIGMA(1,3))*IS33
     B0 = (MEAN(1,1)) - B1*(MEAN(2,1)) - B2*(MEAN(3,1))
     RSQR = B1*(SIGMA(1,2)) + B2*(SIGMA(1,3))
     RR = SQRT(RSQR/(SIGMA(1,1)))
     KSE = STD*Z

* COEFFICIENTS OWEN ET AL. (1975), AND THOMAS ET AL. (1977)

     RY1SQ = ((SIGMA(1,2))**2)/((SIGMA(1,1))*(SIGMA(2,2)))
     RY2SQ = ((SIGMA(1,3))**2)/((SIGMA(1,1))*(SIGMA(3,3)))
     R12SQ = ((SIGMA(2,3))**2)/((SIGMA(2,2))*(SIGMA(3,3)))
     RY1 = SQRT(RY1SQ)
     RY2 = SQRT(RY2SQ)
     R12 = SQRT(R12SQ)
     STD2 = SQRT(SIGMA(2,2))
     STD3 = SQRT(SIGMA(3,3))
     DEN = SQRT((1 - R12SQ)*(RY1SQ + RY2SQ - 2*R12*RY1*RY2))
     A1 = (RY1 - RY2*R12)/DEN
     A2 = (RY2 - RY1*R12)/DEN
     VV =  A1*(MEAN(2,1)) + A2*(MEAN(3,1))
     SESQ = (A1**2)*(SIGMA(2,2)) + (A2**2)*(SIGMA(3,3)) +
    +         2*R12*A1*A2*STD2*STD3
     SEV =  SQRT(SESQ)

* CALCULATE LOWER LIMIT FOR PROCEDURE OF OWEN ET AL.
```

```
      KOEF = OWKOF(NKN,CON)
      LCV = VV - KOEF*SEV
      DEN1 = SQRT((A1**2)*(SIGMA(2,2)) + (A2**2)*(SIGMA(3,3)) +
     +        2*R12*A1*A2*(SQRT(SIGMA(2,2)))*(SQRT(SIGMA(3,3))))
      RYV = (A1*RY1*(SQRT(SIGMA(2,2))) + A2*RY2*(SQRT(SIGMA(3,3))))
     +    /DEN1

      PRINT *,NKN1,LCV

      DO 41 INI = 1,ROW
         DO 40 KIK = 1,COLUMN
            T(KIK,INI) = 0.0
40       CONTINUE
41    CONTINUE

      T(1,1)  = SQRT(SIGMA(1,1))

      DO 45 I = 2,COLUMN
45    T(1,I) = (SIGMA(1,I))/(T(1,1))

      T(2,2) = SQRT((SIGMA(2,2)) - ((T(1,2))**2))
      T(2,3) = (1/(T(2,2)))*((SIGMA(2,3)) - ((T(1,2))*(T(1,3))))
      T(3,3) = SQRT((SIGMA(3,3)) - (((T(1,3))**2) + ((T(2,3))**2)))

* GENERATE 3-VARIATE NORMAL RANDOM VECTOR WITH GIVEN
* MEAN VECTOR AND GIVEN COVARIANCE MATRIX

      SUM = 0.0

      DO 75 K = 1,SAMPLS
         DO 70 I = 1,ROW
            DO 65 J = 1,I
65          SUM = SUM + (T(J,I))*(STNORM(K,J))
            MLTNRM(K,I) = (MEAN(I,1)) + SUM
            SUM = 0.0
70       CONTINUE
75    CONTINUE

         DO 150 KKK = 1,NSAMPL
            MMM = START2(KKK)
            SAMS = MMM + NUMVAR - 1

            DO 95 I = MMM,SAMS
               YHAT = B0 + B1*(MLTNRM(I,2)) + B2*(MLTNRM(I,3))
               LCI = YHAT - KSE
               V   = A1*(MLTNRM(I,2)) + A2*(MLTNRM(I,3))
               LY = 6.849
               IF((MLTNRM(I,1)).LT.LY) THEN
                  NONA = NONA + 1
```

```
    END IF

    IF(LCI.GE.LY) THEN
      SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
      IF((MLTNRM(I,1)).LT.LY) THEN
        SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.0
      END IF
    END IF

    IF(V.GE.LCV) THEN
      SAMPLE(KKK,3) = (SAMPLE(KKK,3)) + 1.0
      IF((MLTNRM(I,1)).LT.LY) THEN
        SAMPLE(KKK,4) = (SAMPLE(KKK,4)) + 1.0
      END IF
    END IF

95  CONTINUE

  IF((SAMPLE(KKK,1)).LT.0.5) THEN
    E = 0.0
  ELSE
    E = (SAMPLE(KKK,2))/(SAMPLE(KKK,1))
  END IF

  IF((SAMPLE(KKK,3)).LT.0.5) THEN
    F = 0.0
  ELSE
    F = (SAMPLE(KKK,4))/(SAMPLE(KKK,3))
  END IF

  A = INT(SAMPLE(KKK,2))
  B = INT(SAMPLE(KKK,1))
  C = INT(SAMPLE(KKK,4))
  D = INT(SAMPLE(KKK,3))

  IF(E.GT.PROB1) THEN
    PREG1 = PREG1 + 1
  END IF

  IF(E.GT.PROB2) THEN
    PREG2 = PREG2 + 1
  END IF

  IF(E.GT.PROB3) THEN
    PREG3 = PREG3 + 1
  END IF

  IF(E.GT.PROB4) THEN
    PREG4 = PREG4 + 1
```

```
      END IF

      IF(F.GT.PROB1) THEN
        POWN1 = POWN1 + 1
      END IF
      IF(F.GT.PROB2) THEN
        POWN2 = POWN2 + 1
      END IF

      IF(F.GT.PROB3) THEN
        POWN3 = POWN3 + 1
      END IF

      IF(F.GT.PROB4) THEN
        POWN4 = POWN4 + 1
      END IF

      WRITE(3,101)KKK,NONA,B,A,E,D,C,F
101    FORMAT(1X,I3,5X,I5,5X,I5,5X,I5,7X,F6.4,5X,I5,5X,I5,7X,F6.4)

      TRREG = TRREG + (SAMPLE(KKK,1))
      TWREG = TWREG + (SAMPLE(KKK,2))
      TROWN = TROWN + (SAMPLE(KKK,3))
      TWOWN = TWOWN + (SAMPLE(KKK,4))
      TNONA = TNONA + NONA
      NONA = 0

150   CONTINUE

      IF(TRREG.LT.0.5) THEN
        PNREG = 0.0
      ELSE
        PNREG = TWREG/TRREG
      END IF

      IF(TROWN.LT.0.5) THEN
        PNOWN = 0.0
      ELSE
        PNOWN = TWOWN/TROWN
      END IF

      TRREG1 = INT(TRREG)
      TWREG1 = INT(TWREG)
      TROWN1 = INT(TROWN)
      TWOWN1 = INT(TWOWN)
      TOTAL = NUMVAR*NSAMPL
      PREG11 = PREG1 - PREG2
      PREG22 = PREG2 - PREG3
      PREG33 = PREG3 - PREG4
```

```
      POWN11 = POWN1 - POWN2
      POWN22 = POWN2 - POWN3
      POWN33 = POWN3 - POWN4

      WRITE(3,155) RR,RYV,TOTAL,TNONA,TRREG1,TWREG1,PNREG,PREG11,
     +             PREG22, PREG33,PREG4,TROWN1,TWOWN1,PNOWN,POWN11,
     +             POWN22,POWN33,POWN4

155   FORMAT(/'REG RSQ  = ',F5.4,' OWEN CORR. COEF. = ',F5.4,
     +      /'TOTAL # OF PRODUCTS                        ',I6,
     +      /'TOTAL # OF NOT ACCEPTABLE PRODUCTS         ',I6,
     +     //'TOTAL # SELECTED BY REG.                   ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY REG.        ',I6,
     +      /'PROP. OF NOT ACCEPTABLE IN SEL. PROP.      ',F8.6,
     +      /'TOTAL # OF SAMPLES IN WHICH PROP. OF',
     +      /'(0.01  < P <=  0.015), 0.05  < P <= 0.06  ',I5,
     +      /'(0.015 < P <=  0.02),  0.06  < P <= 0.075 ',I5,
     +      /'(0.02  < P <=  0.025), 0.075 < P <= 0.085 ',I5,
     +      /'(0.025 < P          ), 0.085 < P          ',I5,
     +     //'TOTAL # SELECTED BY OWEN                   ',I6,
     +      /'TOTAL # OF INCORRECTLY SEL. BY OWEN        ',I6,
     +      /'PROP. OF NOT ACCEPTABLE IN SEL. PROP.      ',F8.6,
     +      /'TOTAL # OF SAMPLES IN WHICH PROP. OF',
     +      /'NOT ACCEPTABLE PRODUCTS EXCEEDES:',
     +      /'(0.01  < P <=  0.015), 0.05  < P <= 0.06  ',I5,
     +      /'(0.015 < P <=  0.02),  0.06  < P <= 0.075 ',I5,
     +      /'(0.02  < P <=  0.025), 0.075 < P <= 0.085 ',I5,
     +      /'(0.025 < P          ), 0.085 < P          ',I5)

      WRITE(7,180) RR,TRREG1,PNREG,PREG1,PREG2
180   FORMAT(F5.4,2X,I6,2X,F5.4,2X,I5,2X,I5)

      WRITE(8,181) RYV,TROWN1,PNOWN,POWN1,POWN2
181   FORMAT(F5.4,2X,I6,2X,F5.4,2X,I5,2X,I5)

      DO 190 IKI = 1,4
        DO 185 JKJ = 1,NSAMPL
          SAMPLE(JKJ,IKI) = 0.0
185     CONTINUE
190   CONTINUE

      NONA  = 0
      TNONA = 0
      TRREG = 0.0
      TWREG = 0.0
      TROWN = 0.0
      TWOWN = 0.0
      PREG1 = 0
      PREG2 = 0
```

```
      PREG3 = 0
      PREG4 = 0
      POWN1 = 0
      POWN2 = 0
      POWN3 = 0
      POWN4 = 0

200   CONTINUE

      CLOSE (4)
      CLOSE (5)
      CLOSE (3)
      CLOSE (8)
      CLOSE (7)

      STOP
      END
```

```
* PROGRAM: 1CORR3E.FOR
*
*        FOR CORRELATION COEFFICIENT FROM 0.5 TO 0.98 BY 0.02:
*
*            GENERATES 3-VARIATE NORMAL RANDOM VECTORS
*            SELECTS SAMPLE OF SIZE 40 WITHIN EACH SELECTED SAMPLE
*            ESTIMATES PARAMETERS FROM EACH SAMPLE OF SIZE 40
*            TESTS REGRESSION METHOD WHEN PARAMETERS ARE NOT KNOWN

*       STREAMS 11 AND 12, FOR RANDOM NUMBER GENERATOR
*       STREAM 31 FOR RANDOM SAMPLING FOR 300 SAMPLES OF SIZE 600
*       STREAM 41 FOR RANDOM SAMPLING FOR SAMPLES OF SIZE 40
*       WITHIN EACH SAMPLE OF SIZE 600


      REAL SIGMA(3,3),T(3,3),STNORM(10000,6),MLTNRM(10000,3),
     +     MEAN(10,1),SAMPLE(610,4),ESTIM(300,9),XX(3,3),TCFACT(3,3),
     +     X1,X2,SUM,DT,DET,DT1,DET1,IA11,IA22,IA12,IA21,
     +     LCLM,BO,B1,B2,YHAT,VAR1,IS22,IS33,IS23,IS32,DT2,DET2,
     +     STD1,LY,DELTA,GAMMA,B11,B22,RR,RSQR,
     +     IXX11,IXX12,IXX13,IXX21,IXX22,IXX23,IXX31,IXX32,IXX33,
     +     PNREG,E,PROB1,PROB2,PROB3,PROB4
      INTEGER ROW,COLUMN,NUMVAR,START2(610),START3(610),SAMPLS,
     +        NSAMPL,A,B,TRREG1,TWREG1,NONA,TNONA,
     +        PREG1,PREG2,PREG3,PREG4,TOTAL,
     +        SMLSS,SAMS,SAMS1,PREG11,PREG22,PREG33

      OPEN (5,FILE = 'B:\PROBACOF.COR')
      OPEN (3,FILE = 'B:\EFFIC.COR')
      OPEN (2,FILE = 'B:\ESTIMATE.COR')
      OPEN (1,FILE = 'B:\TCRG993.COR')

      ROW    = 3
      COLUMN = 3
      NUMVAR = 600
      SAMPLS = 9999
      NSAMPL = 300
      SMLSS1 = 40.0
      SMLSS = 40

* FOR DELTA = 0.95 USE Z = 1.645
* FOR DELTA = 0.99 USE Z = 2.325

* FOR GAMMA = 0.95 USE FD = 3.4934
* FOR GAMMA = 0.99 USE FD = 5.0383

      FD = 5.0383
      Z = 2.325
```

```
    LY = 6.849

    IF(Z.LE.2.0)THEN
        DELTA = 0.95
         PROB1 = 0.05
         PROB2 = 0.06
         PROB3 = 0.075
         PROB4 = 0.085
    ELSE
       DELTA = 0.99
       PROB1 = 0.01
       PROB2 = 0.015
       PROB3 = 0.02
       PROB4 = 0.025
    END IF

    IF(FD.LE.4)THEN
       CHISQ = 22.879
       GAMMA = 0.95
    ELSE
       CHISQ = 19.289
       GAMMA = 0.99
    END IF

    TRREG = 0
    TWREG = 0
    PREG1 = 0
    PREG2 = 0
    PREG3 = 0
    PREG4 = 0
    NONA = 0
    TNONA = 0

* DEFINE MEAN VECTOR

    MEAN(1,1) = 7.0
    MEAN(2,1) = 10.5
    MEAN(3,1) = 4.0

* DEFINE COVARIANCE MATRIX

    SIGMA(1,1)  = 0.05
    SIGMA(1,2)  = 0.08
    SIGMA(1,3)  = 0.04732
    SIGMA(2,1)  = 0.08
    SIGMA(2,2)  = 0.2
    SIGMA(2,3)  = 0.035496
    SIGMA(3,1)  = 0.04732
    SIGMA(3,2)  = 0.035496
```

```
      SIGMA(3,3)   = 0.07

      DO 2 I = 1,4
         DO 1 J = 1,NSAMPL
            SAMPLE(J,I) = 0.0
1        CONTINUE
2     CONTINUE
      DO 3 M = 1,NSAMPL
         START2(M) = 0
         START3(M) = 0
3     CONTINUE

      DO 4 KK = 1,NSAMPL
         START = (RAND(31))*9300
         START2(KK) = INT(START)
         STARTS = (RAND(41))*550
         START3(KK) = INT(STARTS)
4      CONTINUE

* GENERATE 6-VARIATE STANDARD NORMAL RANDOM VECTOR

      DO 6 K = 1,SAMPLS
         DO 5 J = 1,3
            CALL NORM(X1,X2)
            STNORM(K,J) = X1
            I = J + 3
            STNORM(K,I) = X2
5        CONTINUE
6      CONTINUE


      DO 36 K = 1,300
        DO 35 J = 1,9
           ESTIM(K,J) = 0.0
35      CONTINUE
36    CONTINUE

      DO 200 NKN = 1,26
         NKN1 = NKN - 1
         SIGMA(1,2) = 0.01 + NKN1*0.002
         SIGMA(1,3) = 0.03 + NKN1*0.00103
         SIGMA(2,1) = SIGMA(1,2)
         SIGMA(3,1) = SIGMA(1,3)
         DET2 = (SIGMA(2,2))*(SIGMA(3,3)) - (SIGMA(2,3))*(SIGMA(3,2))
         DT2 = 1/DET2
         IS22 = (SIGMA(3,3))*DT2
         IS33 = (SIGMA(2,2))*DT2
         IS23 = (-1)*(SIGMA(2,3))*DT2
         IS32 = (-1)*(SIGMA(3,2))*DT2
```

```
      B11 = (SIGMA(1,2))*IS22 + (SIGMA(1,3))*IS32
      B22 = (SIGMA(1,2))*IS23 + (SIGMA(1,3))*IS33
      RSQR = B11*(SIGMA(1,2)) + B22*(SIGMA(1,3))
      RR = SQRT(RSQR/(SIGMA(1,1)))
      PRINT *,RR,NKN1
      DO 45 I = 1,ROW
        DO 40 K = 1,COLUMN
          T(K,I) = 0.0
40      CONTINUE
45    CONTINUE
      T(1,1)  = SQRT(SIGMA(1,1))
        DO 50 I = 2,COLUMN
50        T(1,I) = (SIGMA(1,I))/(T(1,1))

      T(2,2) = SQRT((SIGMA(2,2)) - ((T(1,2))**2))
      T(2,3) = (1/(T(2,2)))*((SIGMA(2,3)) - ((T(1,2))*(T(1,3))))
      T(3,3) = SQRT((SIGMA(3,3)) - (((T(1,3))**2) + ((T(2,3))**2)))

* CALCULATE 3-VARIATE RANDOM VECTOR WITH GIVEN MEAN VECTOR
* AND GIVEN COVARIANCE MATRIX

      SUM = 0.0

      DO 75 K = 1,SAMPLS
        DO 70 I = 1,ROW
          DO 65 J = 1,I
65         SUM = SUM + (T(J,I))*(STNORM(K,J))
         MLTNRM(K,I) = (MEAN(I,1)) + SUM
         SUM = 0.0
70      CONTINUE
75    CONTINUE

      WRITE(3,80)GAMMA,DELTA
80    FORMAT('GAMMA = ',F4.3,'  DELTA = ',F4.3,/)

      DO 150 KKK = 1,NSAMPL
        MMM = START2(KKK)
        SAMS = MMM + NUMVAR - 1
        MMMS = START3(KKK)
        SAMS1 = MMM + MMMS + SMLSS - 1
        MMMS1 = MMM + MMMS
        MMMM = MMMS1 - 1
        IIII = SAMS1 + 1

* ESTIMATE VARIANCES AND COVARIANCES FROM EACH SMALL SAMPLE

        SUM1 = 0.0
        SUM2 = 0.0
        SUM3 = 0.0
```

```
      SUM4 = 0.0
      SUM5 = 0.0
      SUM6 = 0.0
      SUM7 = 0.0
      SUM8 = 0.0
      SUM9 = 0.0

      DO 90 J = MMMS1,SAMS1
        SUM1 = SUM1 + ((MLTNRM(J,1))*(MLTNRM(J,2)))
        SUM2 = SUM2 + ((MLTNRM(J,1))*(MLTNRM(J,3)))
        SUM3 = SUM3 + ((MLTNRM(J,2))*(MLTNRM(J,3)))
        SUM4 = SUM4 + ((MLTNRM(J,1))**2)
        SUM5 = SUM5 + ((MLTNRM(J,2))**2)
        SUM6 = SUM6 + ((MLTNRM(J,3))**2)
        SUM7 = SUM7 + (MLTNRM(J,1))
        SUM8 = SUM8 + (MLTNRM(J,2))
        SUM9 = SUM9 + (MLTNRM(J,3))
90     CONTINUE

      ESTIM(KKK,1) = SUM7/SMLSS1
      ESTIM(KKK,2) = SUM8/SMLSS1
      ESTIM(KKK,3) = SUM9/SMLSS1
      ESTIM(KKK,4) = SUM4 - SMLSS1*((ESTIM(KKK,1))**2)
      ESTIM(KKK,5) = SUM5 - SMLSS1*((ESTIM(KKK,2))**2)
      ESTIM(KKK,6) = SUM6 - SMLSS1*((ESTIM(KKK,3))**2)
      ESTIM(KKK,7) = SUM1 - SMLSS1*((ESTIM(KKK,1))*(ESTIM(KKK,2)))
      ESTIM(KKK,8) = SUM2 - SMLSS1*((ESTIM(KKK,1))*(ESTIM(KKK,3)))
      ESTIM(KKK,9) = SUM3 - SMLSS1*((ESTIM(KKK,2))*(ESTIM(KKK,3)))

* CALCULATE INVERSE OF MATRIX A, AND INVERSE OF X'X

      DET1 = (ESTIM(KKK,5))*(ESTIM(KKK,6)) - ((ESTIM(KKK,9))**2)
      DT1 = 1/DET1
      IA11 = (ESTIM(KKK,6))*DT1
      IA22 = (ESTIM(KKK,5))*DT1
      IA12 = (-1)*(ESTIM(KKK,9))*DT1
      IA21 = IA12
      XX(1,1) = SMLSS1
      XX(1,2) = SUM8
      XX(1,3) = SUM9
      XX(2,1) = SUM8
      XX(2,2) = SUM5
      XX(2,3) = SUM3
      XX(3,1) = SUM9
      XX(3,2) = SUM3
      XX(3,3) = SUM6
      TCFACT(1,1) = (XX(2,2))*(XX(3,3)) - (XX(3,2))*(XX(2,3))
      TCFACT(2,1) = (-1)*((XX(2,1))*(XX(3,3)) - (XX(3,1))*(XX(2,3)))
      TCFACT(3,1) = (XX(2,1))*(XX(3,2)) - (XX(3,1))*(XX(2,2))
```

```
        TCFACT(1,2) = (-1)*((XX(1,2))*(XX(3,3)) - (XX(3,2))*(XX(1,3)))
        TCFACT(2,2) = (XX(1,1))*(XX(3,3)) - (XX(3,1))*(XX(1,3))
        TCFACT(3,2) = (-1)*((XX(1,1))*(XX(3,2)) - (XX(3,1))*(XX(1,2)))
        TCFACT(1,3) = (XX(1,2))*(XX(2,3)) - (XX(2,2))*(XX(1,3))
        TCFACT(2,3) = (-1)*((XX(1,1))*(XX(2,3)) - (XX(2,1))*(XX(1,3)))
        TCFACT(3,3) = (XX(1,1))*(XX(2,2)) - (XX(2,1))*(XX(1,2))
        DET = (XX(1,1))*(TCFACT(1,1)) + (XX(1,2))*(TCFACT(2,1))
     +       + (XX(1,3))*(TCFACT(3,1))
        DT = 1/DET
        IXX11 = DT*(TCFACT(1,1))
        IXX12 = DT*(TCFACT(1,2))
        IXX13 = DT*(TCFACT(1,3))
        IXX21 = DT*(TCFACT(2,1))
        IXX22 = DT*(TCFACT(2,2))
        IXX23 = DT*(TCFACT(2,3))
        IXX31 = DT*(TCFACT(3,1))
        IXX32 = DT*(TCFACT(3,2))
        IXX33 = DT*(TCFACT(3,3))

* CALCULATE REGRESSION COEFFICIENTS, AND STANDARD DEVIATION
* FOR EACH SAMPLE

        B1 = (ESTIM(KKK,7))*IA11 + (ESTIM(KKK,8))*IA21
        B2 = (ESTIM(KKK,7))*IA12 + (ESTIM(KKK,8))*IA22
        A1 = (ESTIM(KKK,2))*B1
        A2 = (ESTIM(KKK,3))*B2
        A3 = ESTIM(KKK,1)
        BO = A3 - A1 - A2
        VAR1 = ((ESTIM(KKK,4)) - B1*(ESTIM(KKK,7)) - B2*(ESTIM(KKK,8)))/37.0
        STD1 = SQRT(VAR1)
        KOF1 = SQRT(3*FD)
        KOF2 = SQRT((SMLSS1 - 3.0)/(CHISQ))

        DO 100 J = MMM,MMMM
          YHAT = BO + B1*(MLTNRM(J,2)) + B2*(MLTNRM(J,3))
           HTDIG = IXX11 + (MLTNRM(J,2))*IXX21 + (MLTNRM(J,3))*IXX31
     +          + (MLTNRM(J,2))*IXX12 + ((MLTNRM(J,2))**2)*IXX22
     +          + (MLTNRM(J,2))*IXX32*(MLTNRM(J,3)) +
     +            (MLTNRM(J,3))*IXX13 + (MLTNRM(J,2))*IXX23*(MLTNRM(J,3))
     +          + ((MLTNRM(J,3))**2)*IXX33


          LCLM = YHAT - STD1*(KOF1*(SQRT(HTDIG)) + Z*KOF2)

          IF((MLTNRM(J,1)).LT.LY) THEN
            NONA = NONA + 1
          END IF

          IF(LCLM.GE.LY) THEN
```

```
          SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
             IF((MLTNRM(J,1)).LT.LY) THEN
               SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.0
             END IF
        END IF

100     CONTINUE

      DO 104 II = IIII,SAMS
        YHAT = BO + B1*(MLTNRM(II,2)) + B2*(MLTNRM(II,3))
        HTDIG = IXX11 + (MLTNRM(II,2))*IXX21 + (MLTNRM(II,3))*IXX31
     +          +  (MLTNRM(II,2))*IXX12 + ((MLTNRM(II,2))**2)*IXX22
     +          + (MLTNRM(II,2))*IXX32*(MLTNRM(II,3)) +
     +             (MLTNRM(II,3))*IXX13 + (MLTNRM(II,2))*IXX23*(MLTNRM(II,3))
     +          + ((MLTNRM(II,3))**2)*IXX33
        LCLM = YHAT - STD1*(KOF1*(SQRT(HTDIG)) + Z*KOF2)
        IF((MLTNRM(II,1)).LT.LY) THEN
          NONA = NONA + 1
        END IF

        IF(LCLM.GE.LY) THEN
          SAMPLE(KKK,1) = (SAMPLE(KKK,1)) + 1.0
             IF((MLTNRM(II,1)).LT.LY) THEN
               SAMPLE(KKK,2) = (SAMPLE(KKK,2)) + 1.0
             END IF
        END IF
104   CONTINUE

      IF((SAMPLE(KKK,1)).LT.0.5) THEN
        E = 0.0
      ELSE
        E = (SAMPLE(KKK,2))/(SAMPLE(KKK,1))
      END IF

      A = INT(SAMPLE(KKK,2))
      B = INT(SAMPLE(KKK,1))


      IF(E.GT.PROB1) THEN
        PREG1 = PREG1 + 1
      END IF

      IF(E.GT.PROB2) THEN
        PREG2 = PREG2 + 1
      END IF

      IF(E.GT.PROB3) THEN
        PREG3 = PREG3 + 1
      END IF
```

```
         IF(E.GT.PROB4) THEN
            PREG4 = PREG4 + 1
         END IF

         WRITE(3,105)KKK,NONA,B,A,E
105       FORMAT(1X,I3,5X,I5,5X,I5,5X,I5,8X,F6.4)

         TRREG = TRREG + (SAMPLE(KKK,1))
         TWREG = TWREG + (SAMPLE(KKK,2))
         TNONA = TNONA + NONA
         NONA = 0

150   CONTINUE

      IF(TRREG.LT.0.5) THEN
      PNREG = 0.0
      ELSE
      PNREG = TWREG/TRREG
      END IF

      TRREG1 = INT(TRREG)
      TWREG1 = INT(TWREG)
      TOTAL = NUMVAR*NSAMPL
      PREG11 = PREG1 - PREG2
      PREG22 = PREG2 - PREG3
      PREG33 = PREG3 - PREG4

      WRITE(3,155)TOTAL,TNONA,TRREG1,TWREG1,PNREG,PREG11,PREG22
     +          ,PREG33,PREG4
155   FORMAT(/'TOTAL NUMBER OF PRODUCTS                    ',I6,
     +      /'TOTAL # OF NOT ACCEPTABLE PRODUCTS          ',I6,
     +     //'TOTAL # SELECTED                    ',I6,
     +      /'TOTAL # OF INCORRECTLY SELECTED             ',I6,
     +      /'PROP. OF NOT ACCEPTABLE IN SELECTED PROP.   ',F8.6,
     +      /'TOTAL # OF SAMPLES IN WHICH PROP. OF',
     +      /'NOT ACCEPTABLE PRODUCTS EXCEEDES:',
     +      /'(0.01  < P <= 0.015), 0.05  < P <= 0.06      ',I5,
     +      /'(0.015 < P <= 0.02),  0.06  < P <= 0.075     ',I5,
     +      /'(O.O2  < P <= 0.025), 0.075 < P <= 0.085     ',I5,
     +      /'(0.025 < P         ), 0.085 , P              ',I5)


      WRITE(1,180)RR,TRREG1,PNREG,PREG1
180   FORMAT(F5.4,2X,I6,3X,F5.4,3X,I5)

      DO 182 JIJ = 1,4
      DO 181 JKJ = 1,NSAMPL
      SAMPLE(JKJ,JIJ) = 0.0
181   CONTINUE
```

```
182    CONTINUE

        NONA  = 0
        TNONA = 0
        PREG1 = 0
        PREG2 = 0
        PREG3 = 0
        PREG4 = 0
        POWN1 = 0
        POWN2 = 0
        POWN3 = 0
        POWN4 = 0
        SUM1 = 0.0
        SUM2 = 0.0
        SUM3 = 0.0
        SUM4 = 0.0
        SUM5 = 0.0
        SUM6 = 0.0
        SUM7 = 0.0
        SUM8 = 0.0
        SUM9 = 0.0
        TRREG = 0.0
        TWREG = 0.0

200 CONTINUE

    CLOSE (1)
    CLOSE (5)
    CLOSE (3)
    CLOSE (2)

    STOP
    END
```

```
      SUBROUTINE NORM(Y1,Y2)

      REAL Y1,Y2,U1,U2,W,C,RAND
1000  U1 = RAND(11)
      U2 = RAND(12)
      V1 = 2*U1 - 1
      V2 = 2*U2 - 1
      W  = (V1**2) + (V2**2)
      IF(W.GT.1.0) THEN
         GO TO 1000
      ELSE
         F = LOG(W)*(-2)
         C = SQRT(F/W)
         Y1 = V1*C
         Y2 = V2*C
      END IF

      RETURN
      END
```

```
      REAL FUNCTION RAND(ISTRM)
      INTEGER B2E15,B2E16,HI15,HI31,ISTRM,IZSET,LOW15,LOWPRD,
     +        MODLUS,MULT1,MULT2,OVFLOW,ZI,ZRNG(100)
      INTEGER IRANDG,RANDST
      SAVE ZRNG
      DATA MULT1,MULT2/24112,26143/
      DATA B2E15,B2E16,MODLUS/32768,65536,2147483647/
      DATA ZRNG/1973272912,281629770,20006270,1280689831,2096730329,
     +          1933576050,913566091,246780520,1363774876,604901985,
     +          1511192140,1259851944,824064364,150493284,242708531,
     +          75253171,1964472944,1202299975,233217322,1911216000,
     +          726370533,403498145,99322223,1103205531,762430696,
     +          1922803170,1385516923,76271663,413682397,726466604,
     +          336157058,1432650381,1120463904,595778810,877722890,
     +          1046574445,68911991,2088367019,748545416,622401386,
     +          2122378830,640690903,1774806513,2132545692,2079249579,
     +          78130110,852776735,1187867272,1351423507,1645973084,
     +          1997049139,922510944,2045512870,898585771,243649545,
     +          1004818771,77368062,403188473,372279877,1901633463,
     +          498067494,2087759558,493157915,597104727,1530940798,
     +          1814496276,536444882,1663153658,855503735,67784357,
     +          1432404475,619691088,119025595,880802310,176192644,
     +          1116780070,277854671,1366580350,1142483975,2026948561,
     +          1053920743,786262391,1792203830,1494667770,1923011392,
     +          1433700034,1244184613,1147297105,539712780,1545929719,
     +          190641742,1645390429,264907697,620389253,1502074852,
     +          927711160,364849192,2049576050,638580085,547070247/

      ZI = ZRNG(ISTRM)
      HI15 = ZI/B2E16
      LOWPRD =(ZI - HI15*B2E16)*MULT1
      LOW15 = LOWPRD/B2E16
      HI31 = HI15*MULT1 + LOW15
      OVFLOW = HI31/B2E15
      ZI = (((LOWPRD - LOW15*B2E16) - MODLUS) +
     +    (HI31 - OVFLOW*B2E15) * B2E16 ) + OVFLOW
      IF(ZI.LT.0) ZI = ZI + MODLUS
      HI15 = ZI/B2E16
      LOWPRD = (ZI - HI15*B2E16) * MULT2
      LOW15 = LOWPRD/B2E16
      HI31 = HI15*MULT2 + LOW15
      OVFLOW = HI31/B2E15
      ZI = (((LOWPRD - LOW15 * B2E16) - MODLUS) +
     +    (HI31 - OVFLOW*B2E15) * B2E16) + OVFLOW
      IF ( ZI.LT.0) ZI = ZI + MODLUS
      ZRNG(ISTRM) = ZI
      RAND = (2*(ZI/256) + 1)/16777216.0
      RETURN
      ENTRY RANDST(IZSET,ISTRM)
      ZRNG(ISTRM) = IZSET
```

```
RETURN
ENTRY IRANDG(ISTRM)
IRANDG = ZRNG(ISTRM)
RETURN
END
```

Appendix B

Known Parameters

In section **3.1.1.1.** comparison of methods of Owen et al. and regression method is made, for $\rho_{YX} = \rho_{YW} = 0.8$. Screening is based on following decision rules:

From the tables given in Odeh and Owen (1980) the proportion which needs to be selected is:

i) for $\delta = 0.95$

$\psi = 0.5661$, giving $z_\psi = 0.165$, then the decision rule is:

accept the product i if $x_i \geq 10.5 - 0.165 \times 0.4472. = 10.4262$

ii) for $\delta = 0.99$

$\psi = 0.3029$, giving $z_\psi = -0.515$, and the decision rule is:

accept the product i if $x_i \geq 10.5 - (-0.515) \times 0.4472 = 10.7303$

The regression coefficients, and the standard deviation are:

$$\beta_0 = \mu_Y - \sigma_{YX}\Sigma_{XX}^{-1}\mu_X = \mu_Y - \frac{\sigma_{YX}}{\sigma_X^2}\mu_X = 7.0 - \frac{0.08}{0.2}10.5 = 2.80$$

$$\beta_1 = \sigma_{YX}\Sigma_{XX}^{-1} = \frac{0.08}{0.2} = 0.40$$

$$\sigma_W = \sqrt{\sigma_Y^2 - \sigma_{YX}\Sigma_{XX}^{-1}\sigma_{XY}} = \sqrt{\sigma_Y^2 - \frac{\sigma_{YX}^2}{\sigma_X^2}} = \sqrt{0.05 - \frac{0.08^2}{0.2}} = 0.134$$

and the decision rule is:

i) for $\delta = 0.95$

accept the product i if $P_{\delta i} = 2.80 + 0.40x_i - 1.645 \times 0.134 \geq L_Y = 6.849$

ii) for $\delta = 0.99$

accept the product i if $P_{\delta i} = 2.80 + 0.40x_i - 2.325 \times 0.134 \geq L_Y = 6.849$.

List of figures:

i) $\delta = 0.95$

Figure B3 - proportion of unacceptable products after screening

i) $\delta = 0.99$

Figure B4 - number of product selected from each lot out of 452 acceptable on average

Figure B5 - proportion of unacceptable products in each sample, before and after screening

Figure B6 - proportion of unacceptable products after screening

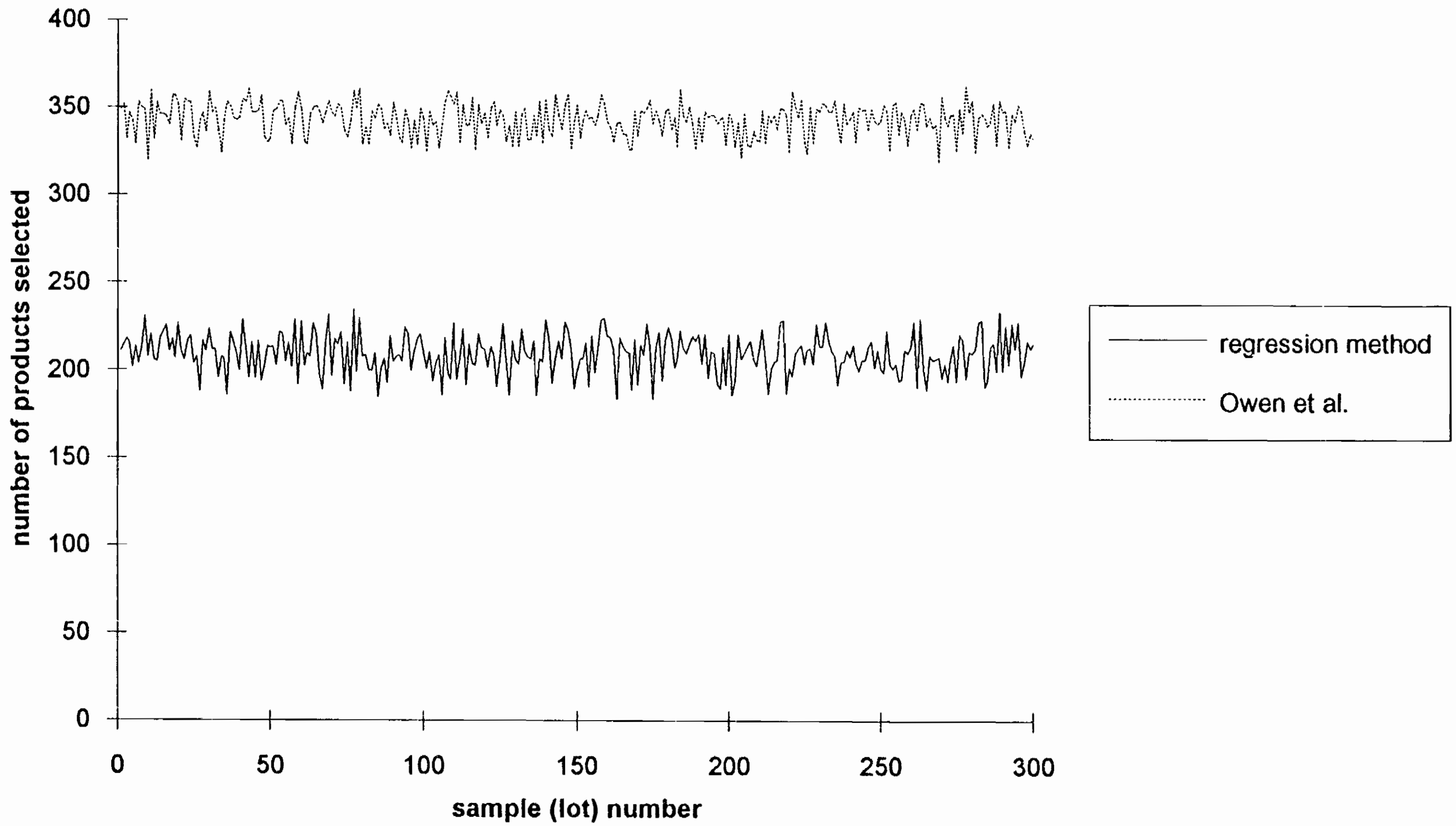

Figure B1: $\delta = 0.95$

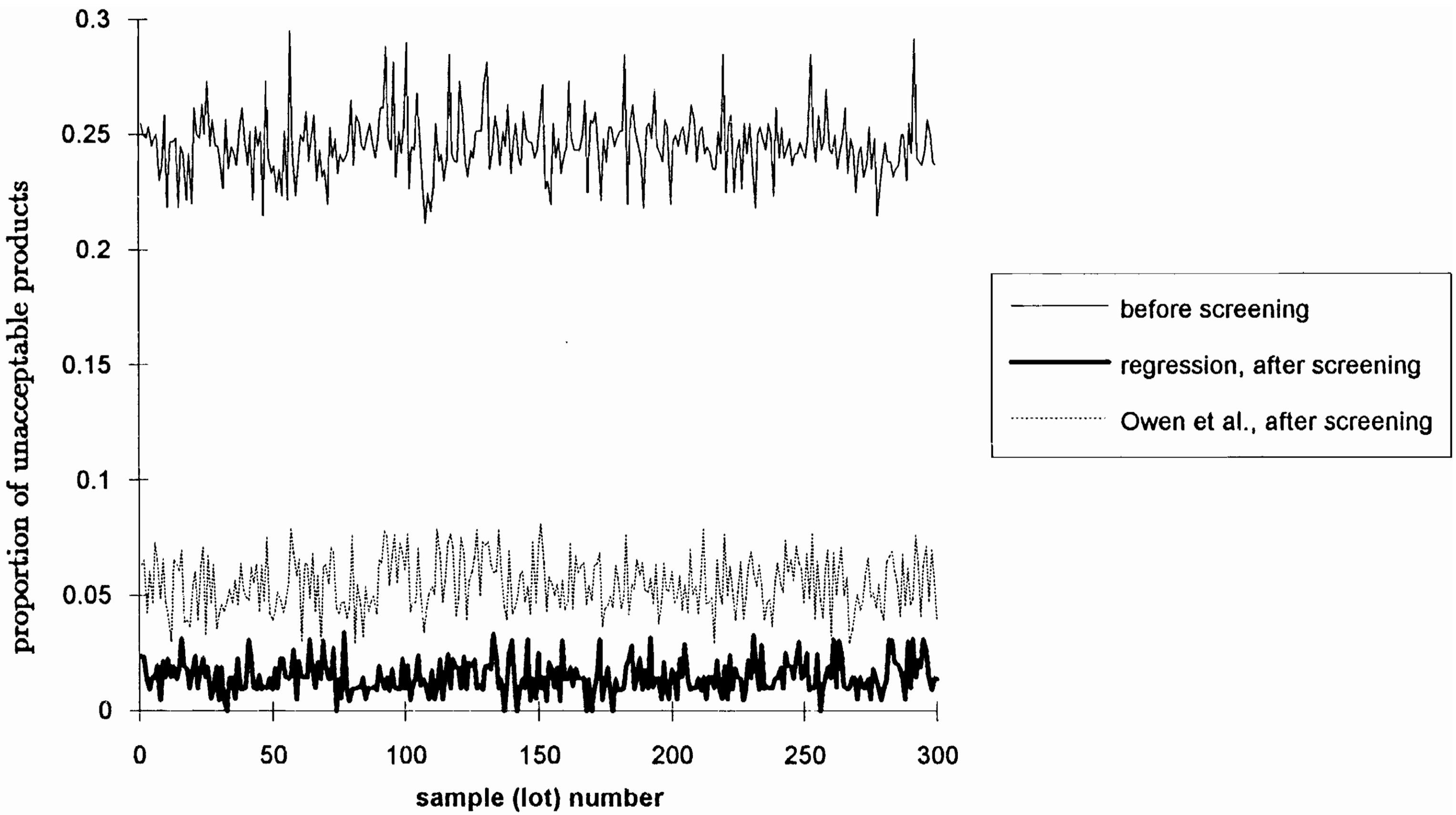

Figure B2: $\delta = 0.95$

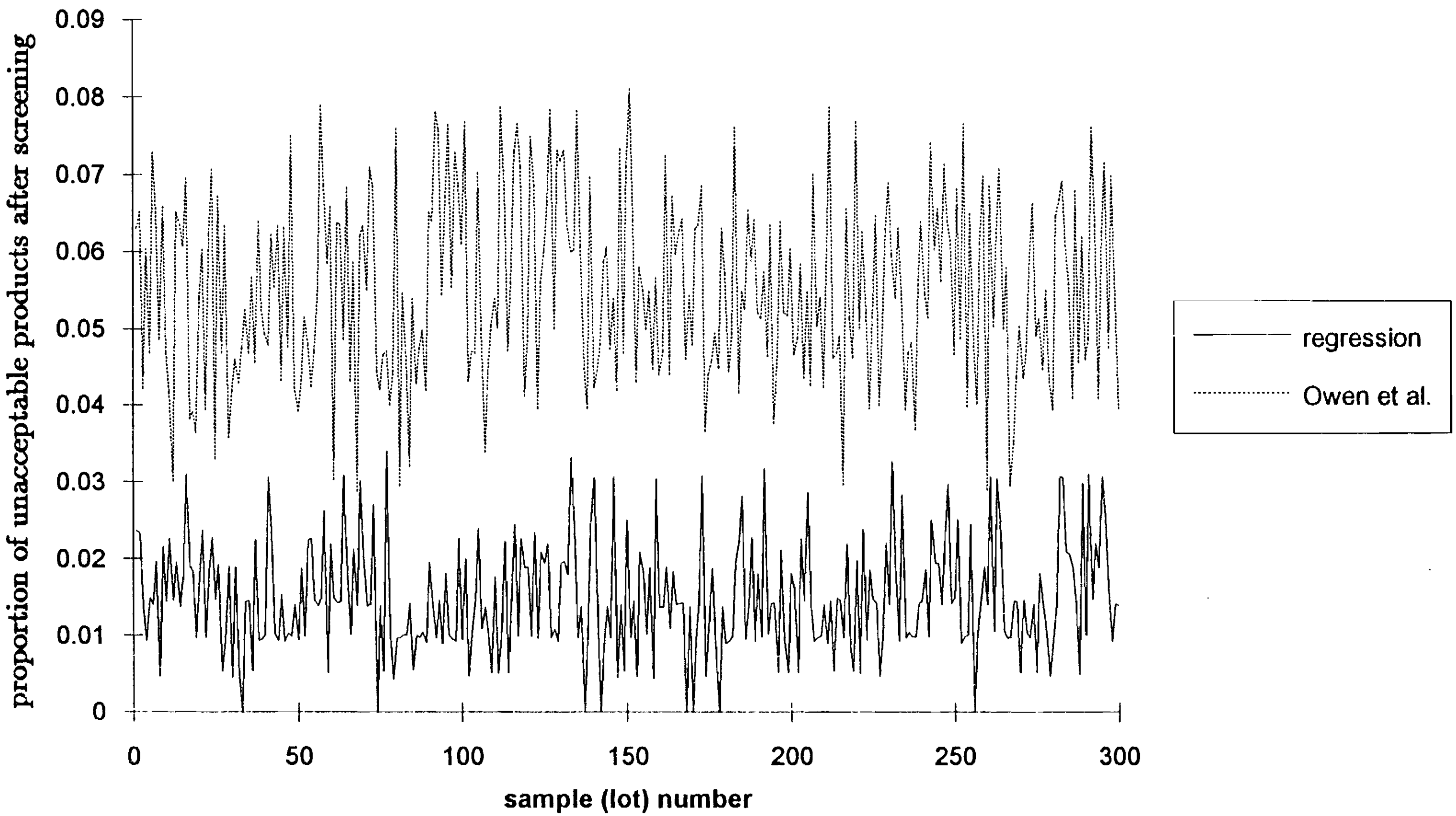

Figure B3: $\delta = 0.95$

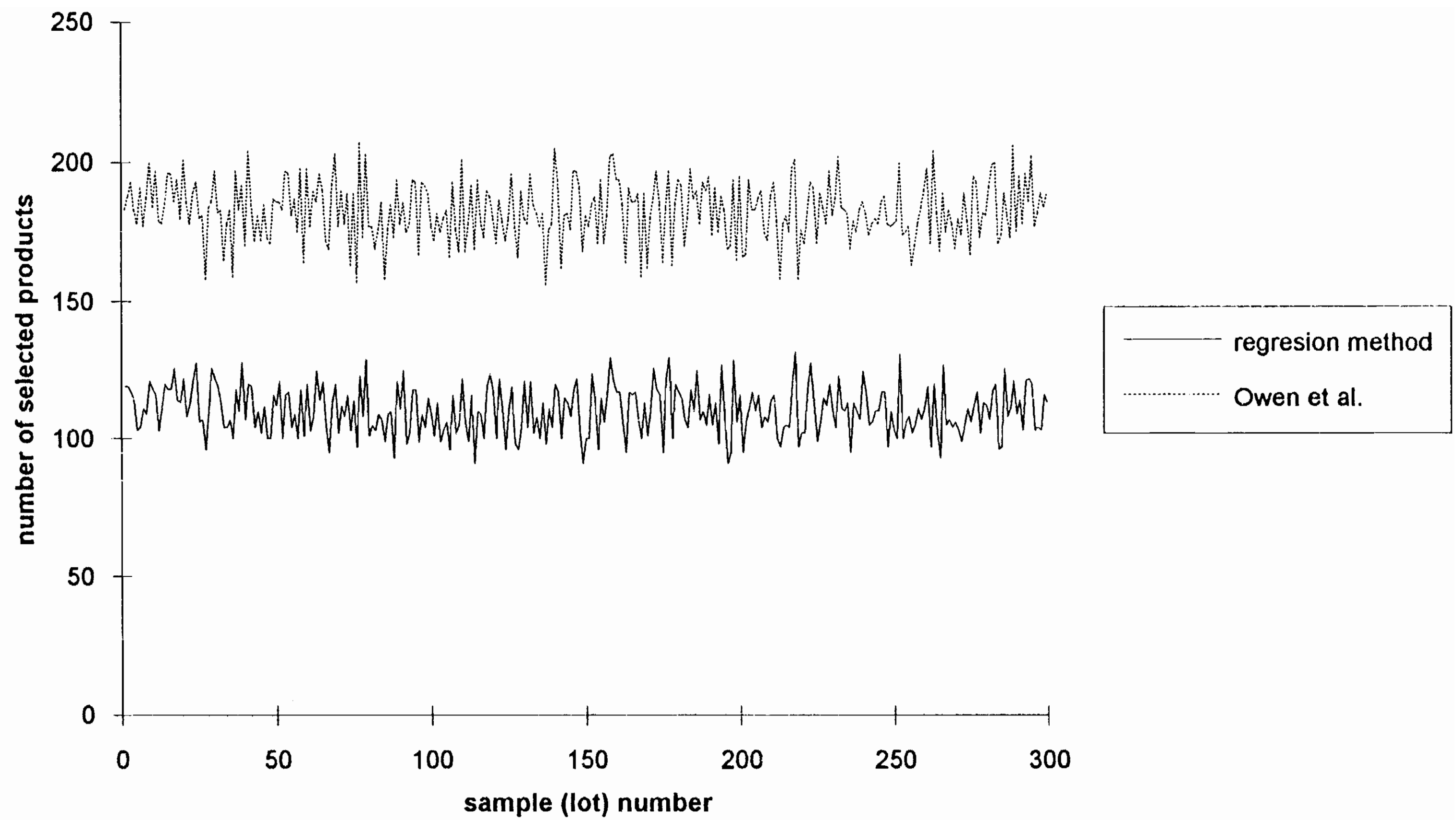

Figure B4: $\delta = 0.99$

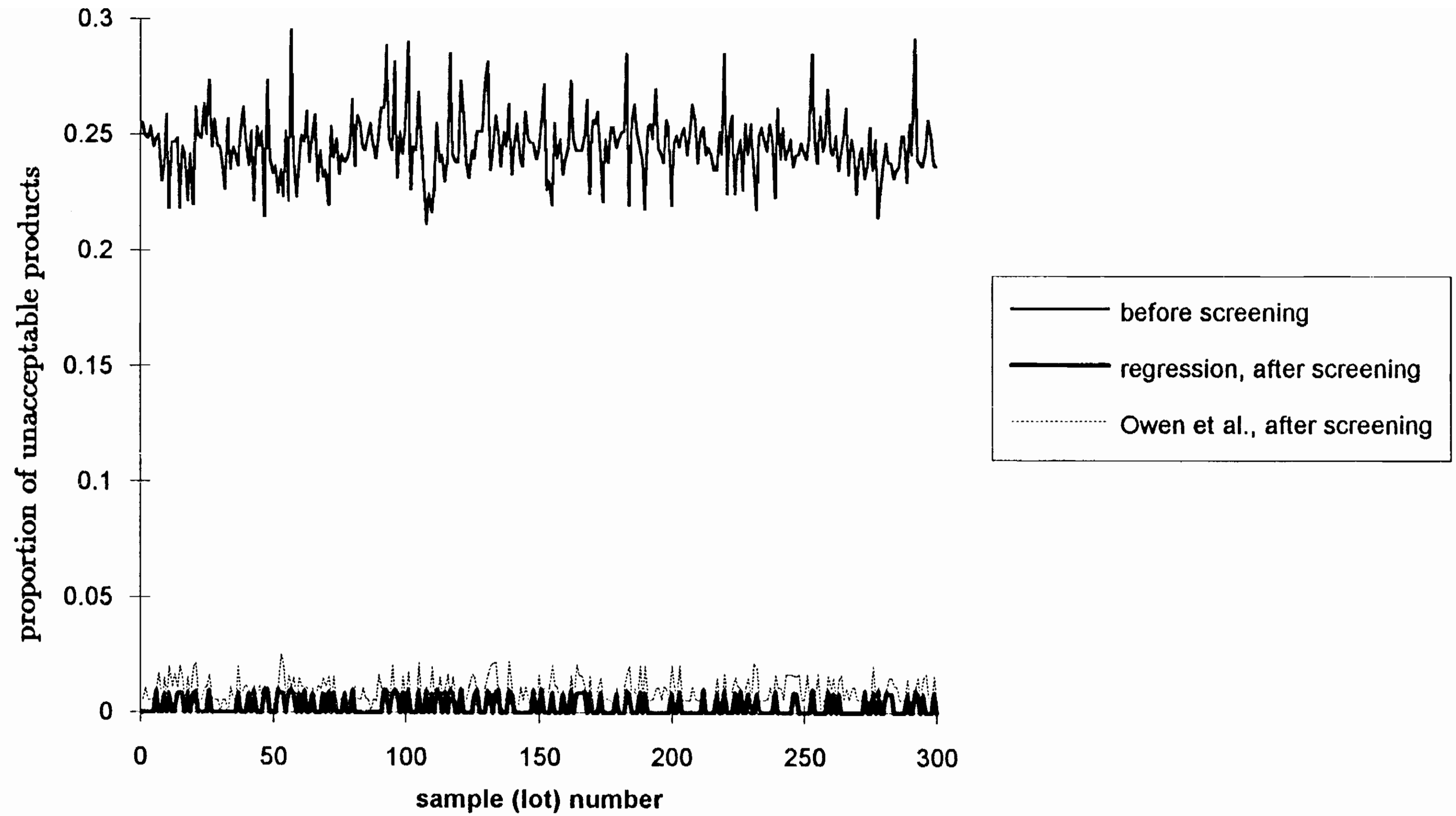

Figure B5: $\delta = 0.99$

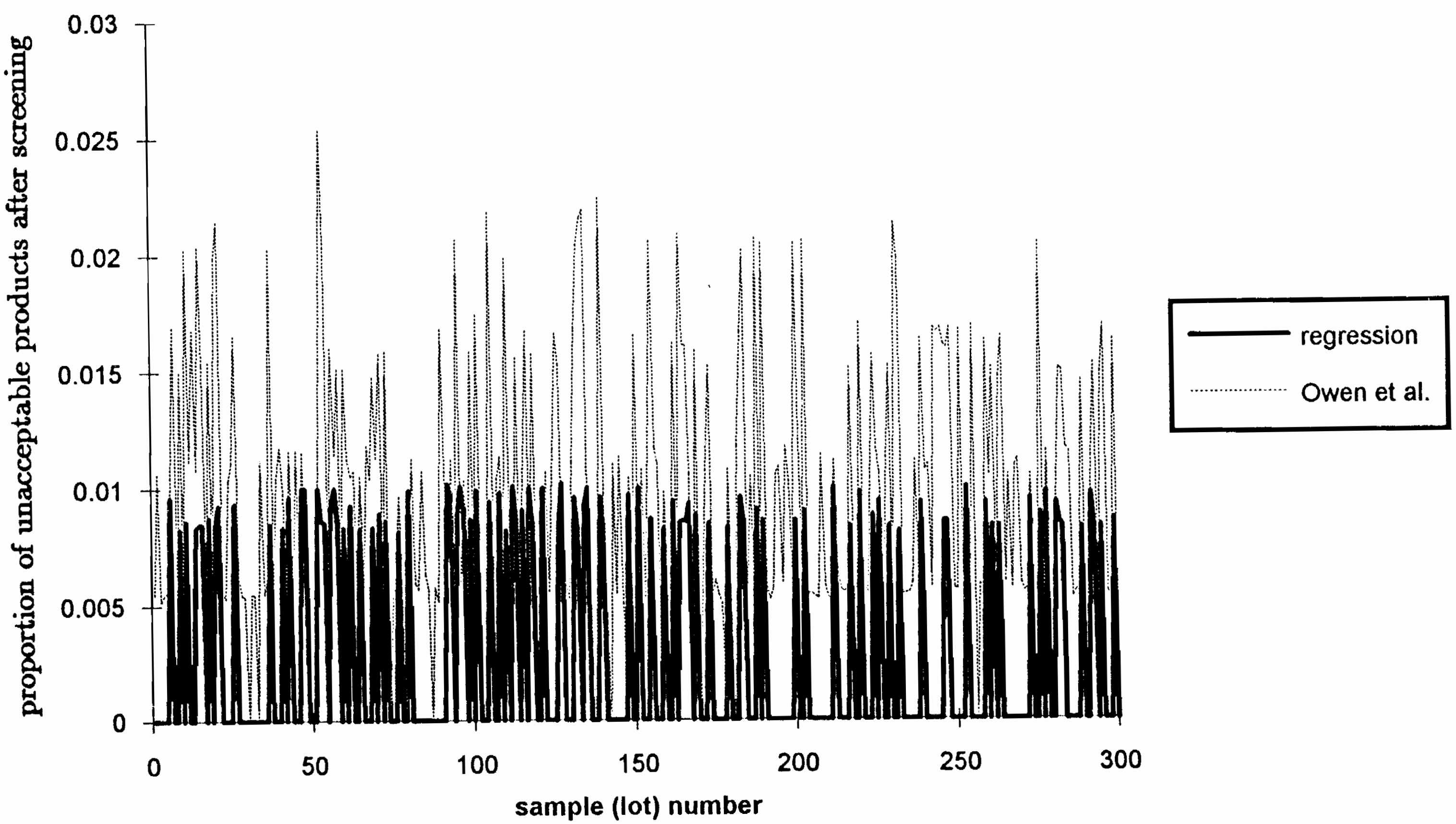

Figure B6: $\delta = 0.99$

Appendix C

Unknown Parameters

detailed result from section **3.1.2.3.** for each sample.

List of figures:

$\eta = 0.95$

i) $\delta = 0.95$

ii) $\delta = 0.99$

Figure C7 - proportion of unacceptable products after screening

$\eta = 0.99$

i) $\delta = 0.95$

Figure C7 - number of product selected from each lot out of 452 acceptable on average

Figure C8 - proportion of unacceptable products in each sample, before and after screening

Figure C9 - proportion of unacceptable products after screening

ii) $\delta = 0.99$

Figure C10 - number of product selected from each lot out of 452 acceptable on average

Figure C11 - proportion of unacceptable products in each sample, before and after screening

Figure C12 - proportion of unacceptable products after screening

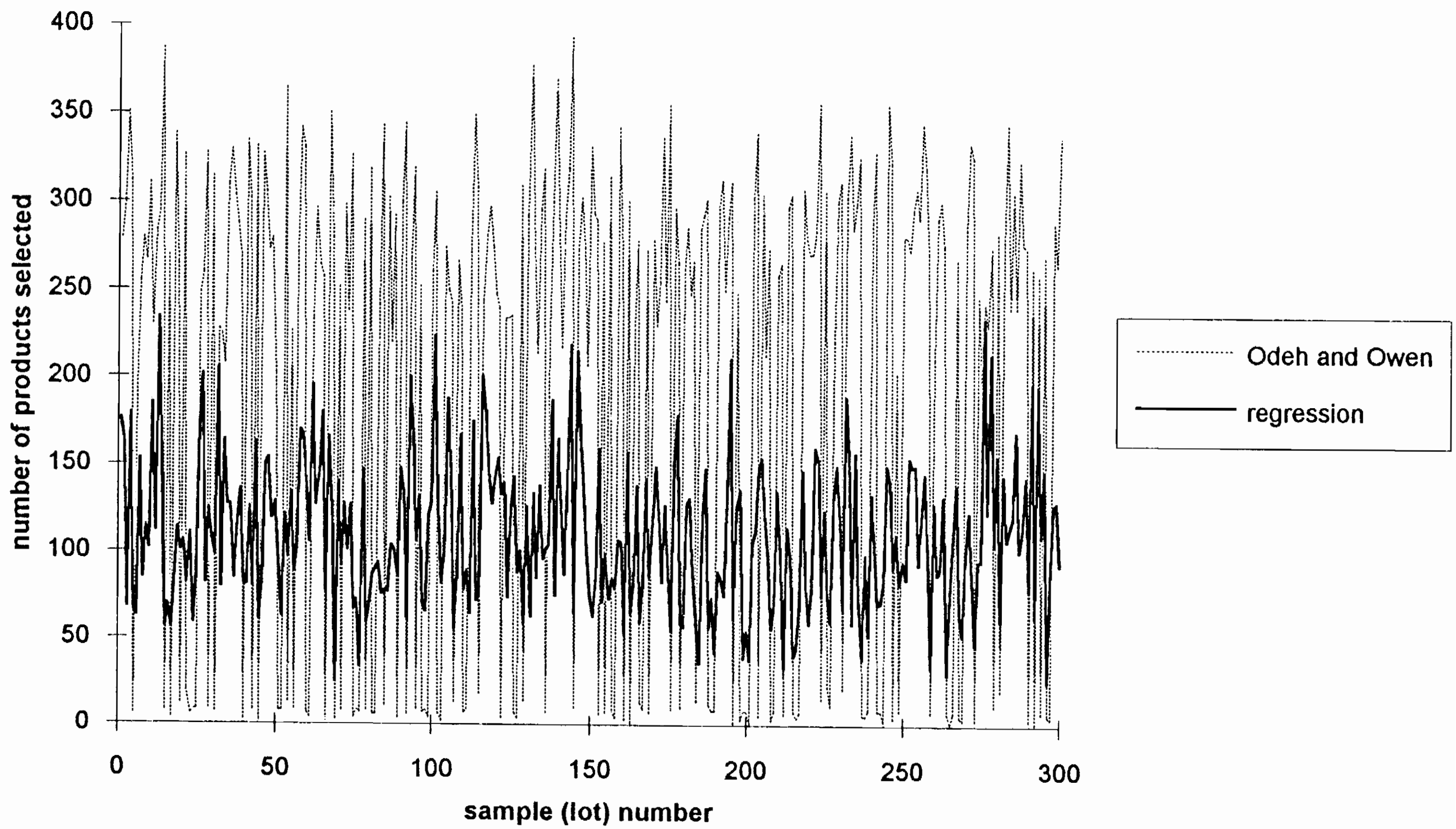

Figure C1: $\eta = 0.95$, $\delta = 0.95$

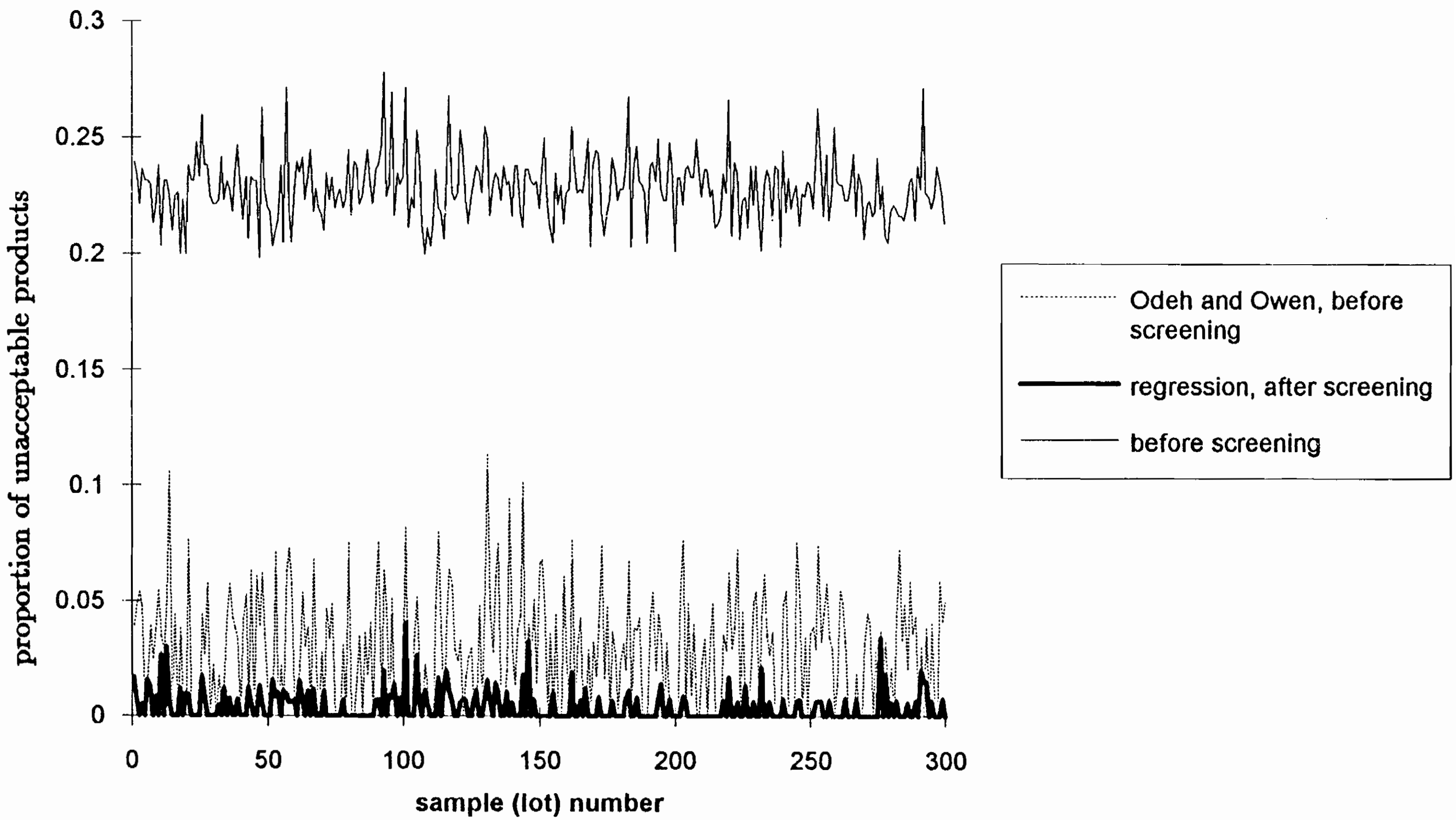

Figure C2: $\eta = 0.95$, $\delta = 0.95$

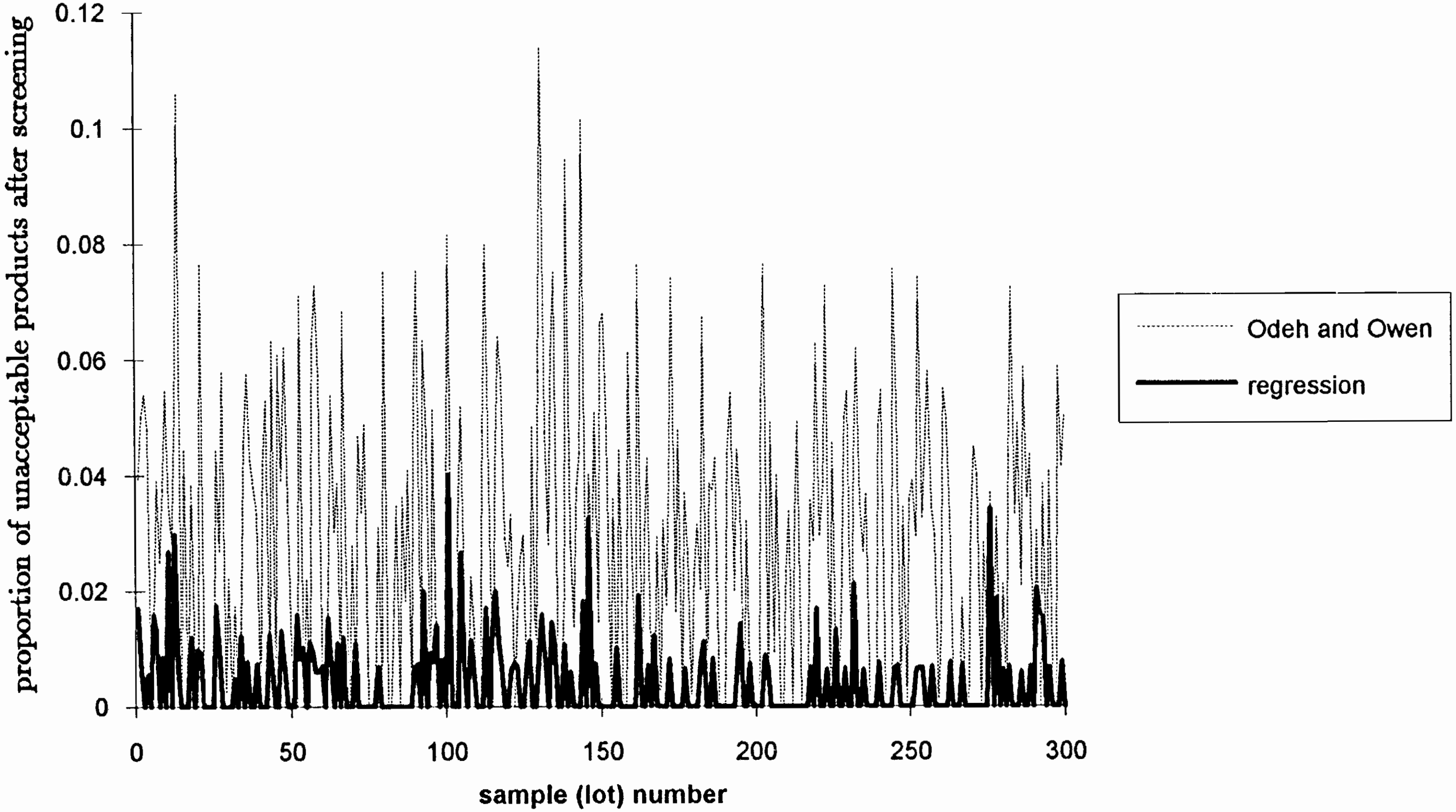

Figure C3: $\eta = 0.95$, $\delta = 0.95$

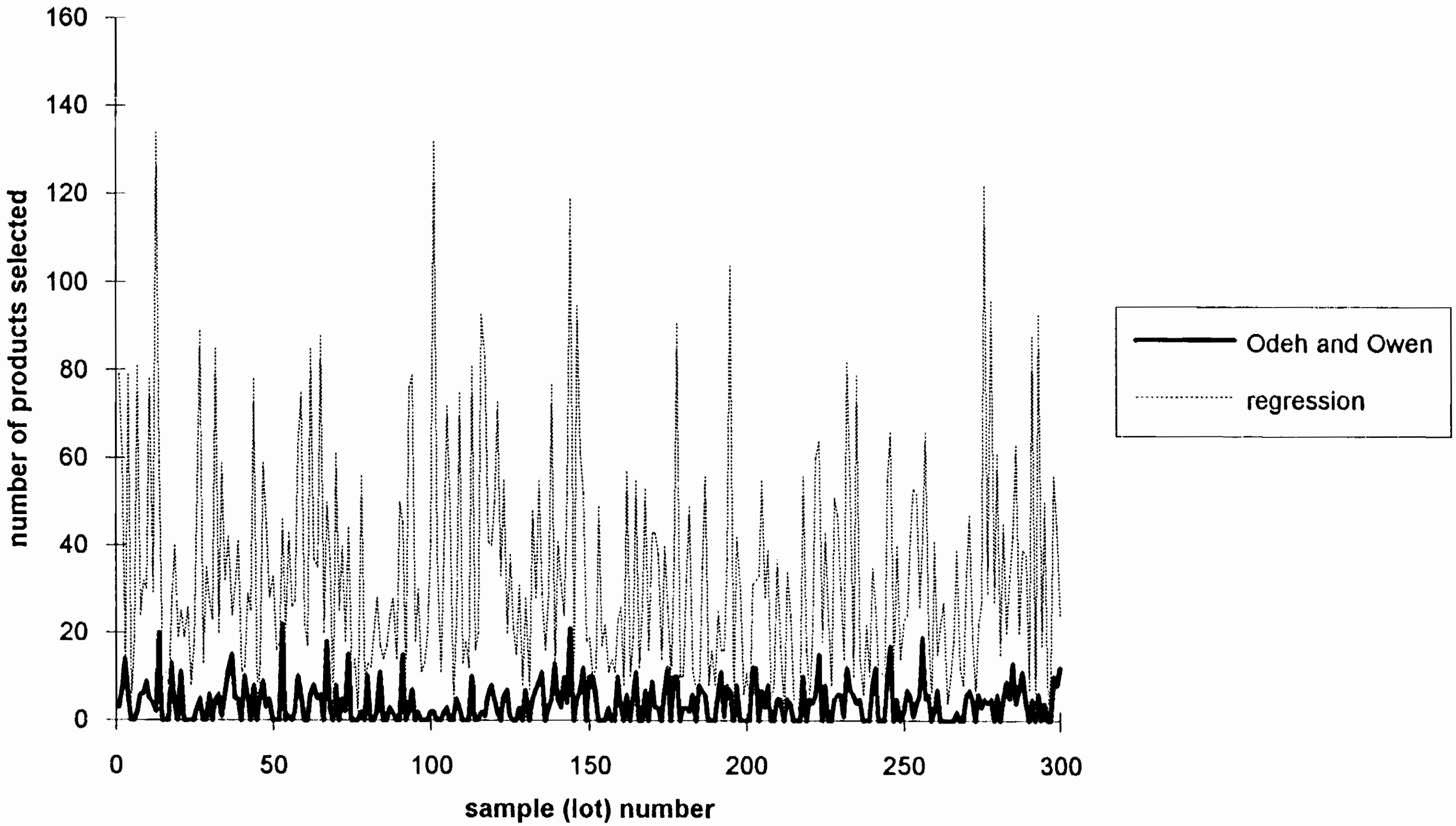

Figure C4: $\eta = 0.95$, $\delta = 0.99$

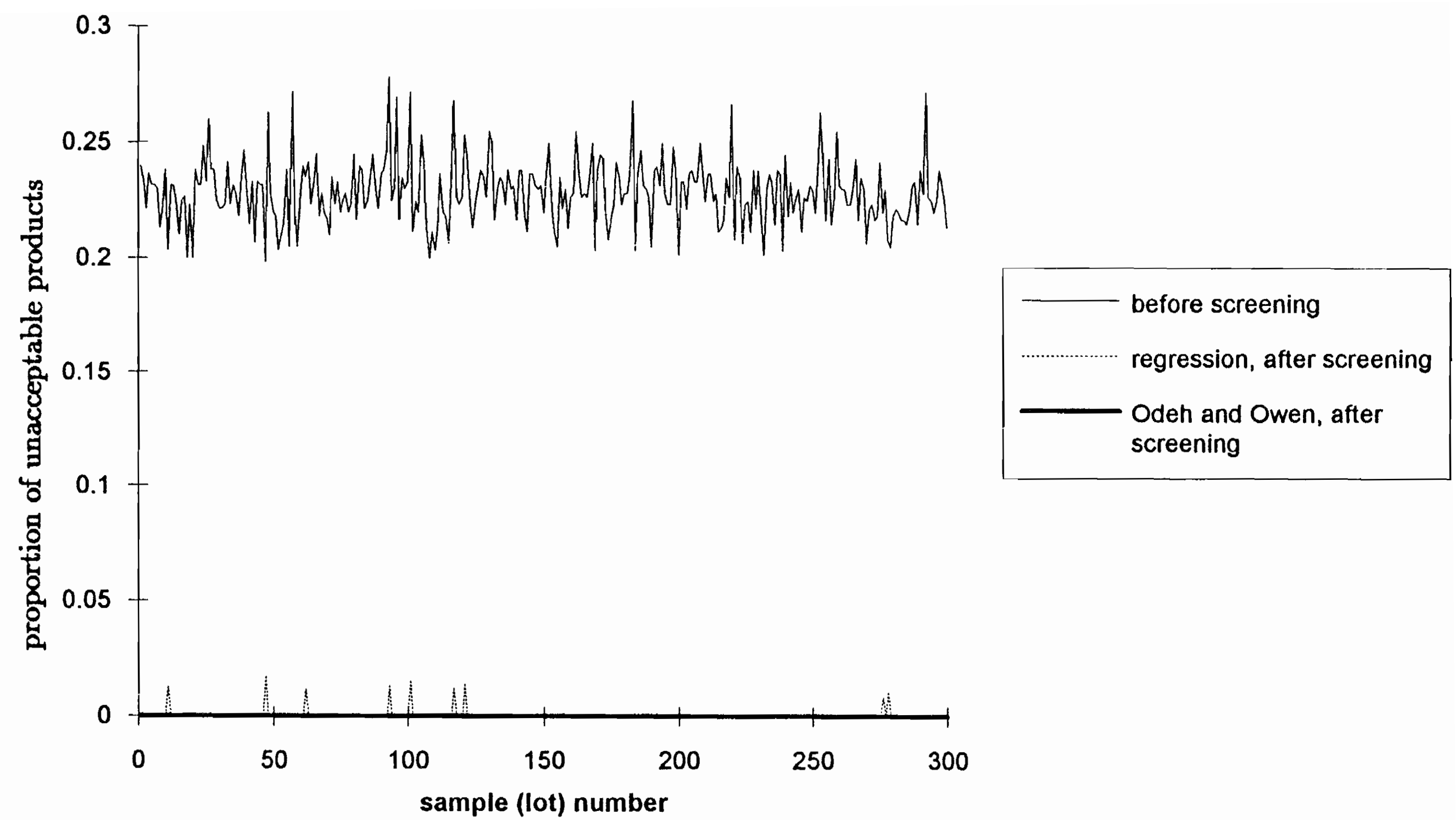

Figure C5: $\eta = 0.95$, $\delta = 0.99$

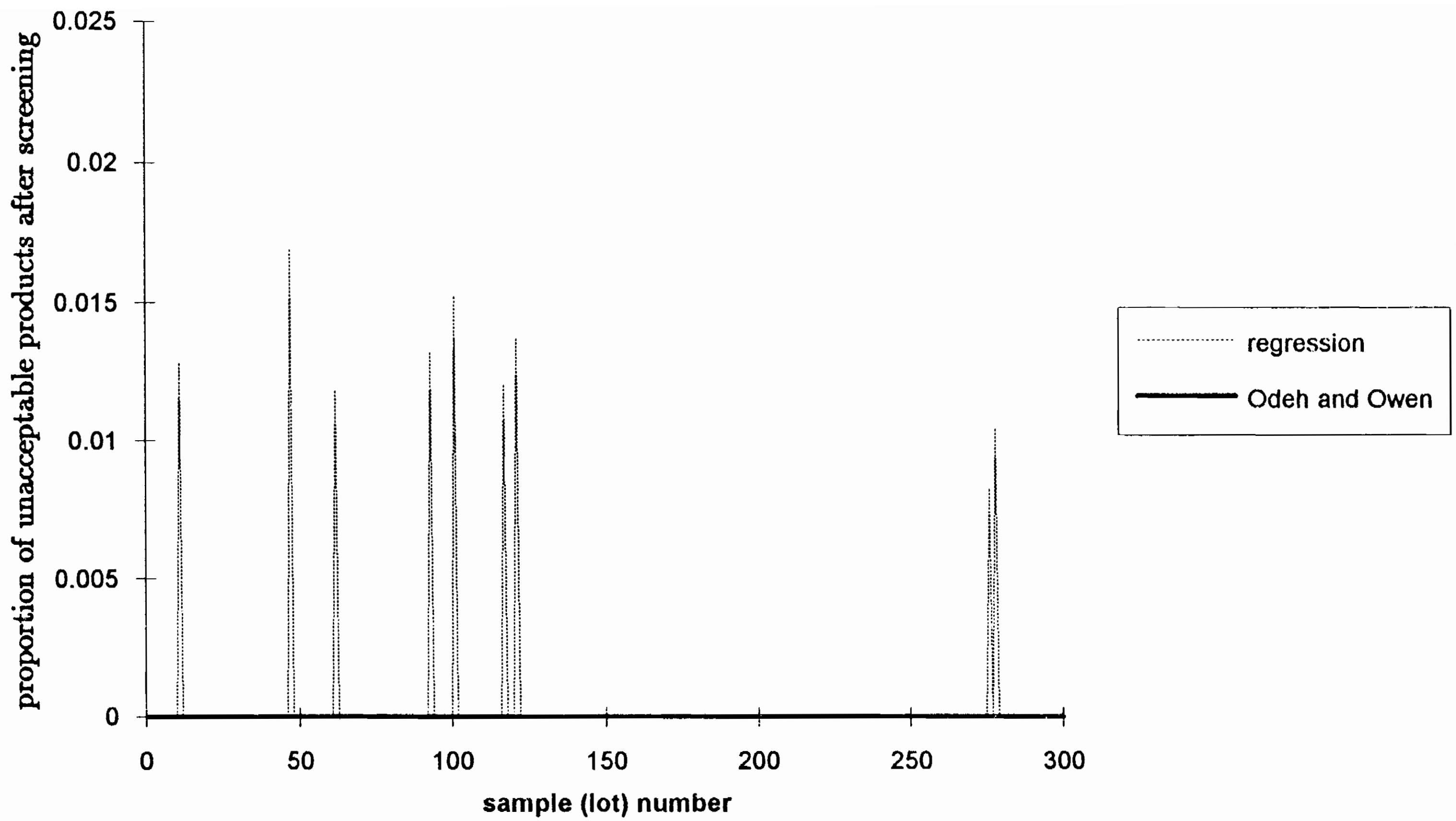

Figure C6: $\eta = 0.95$, $\delta = 0.99$

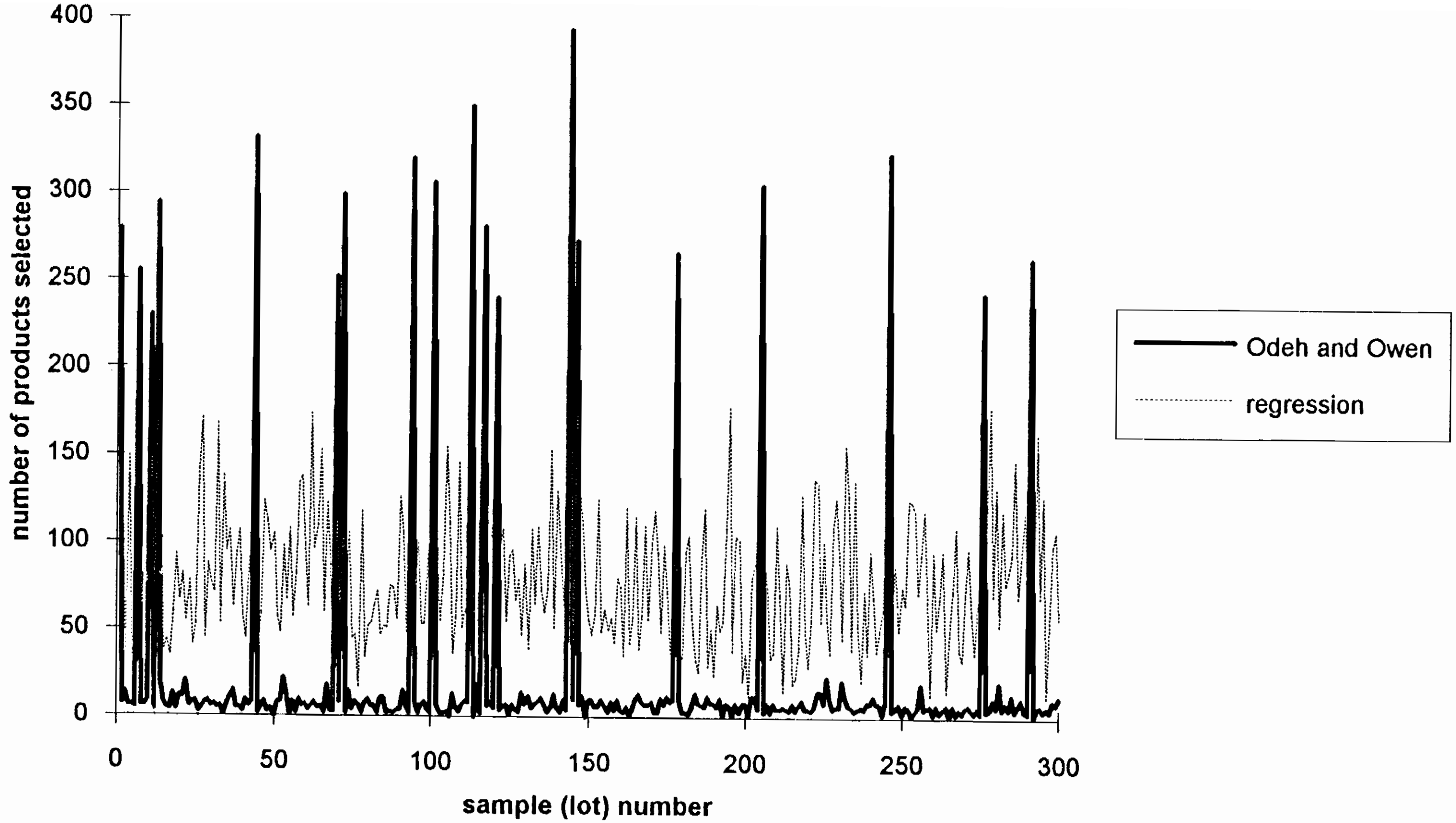

Figure C7: $\eta = 0.99$, $\delta = 0.95$

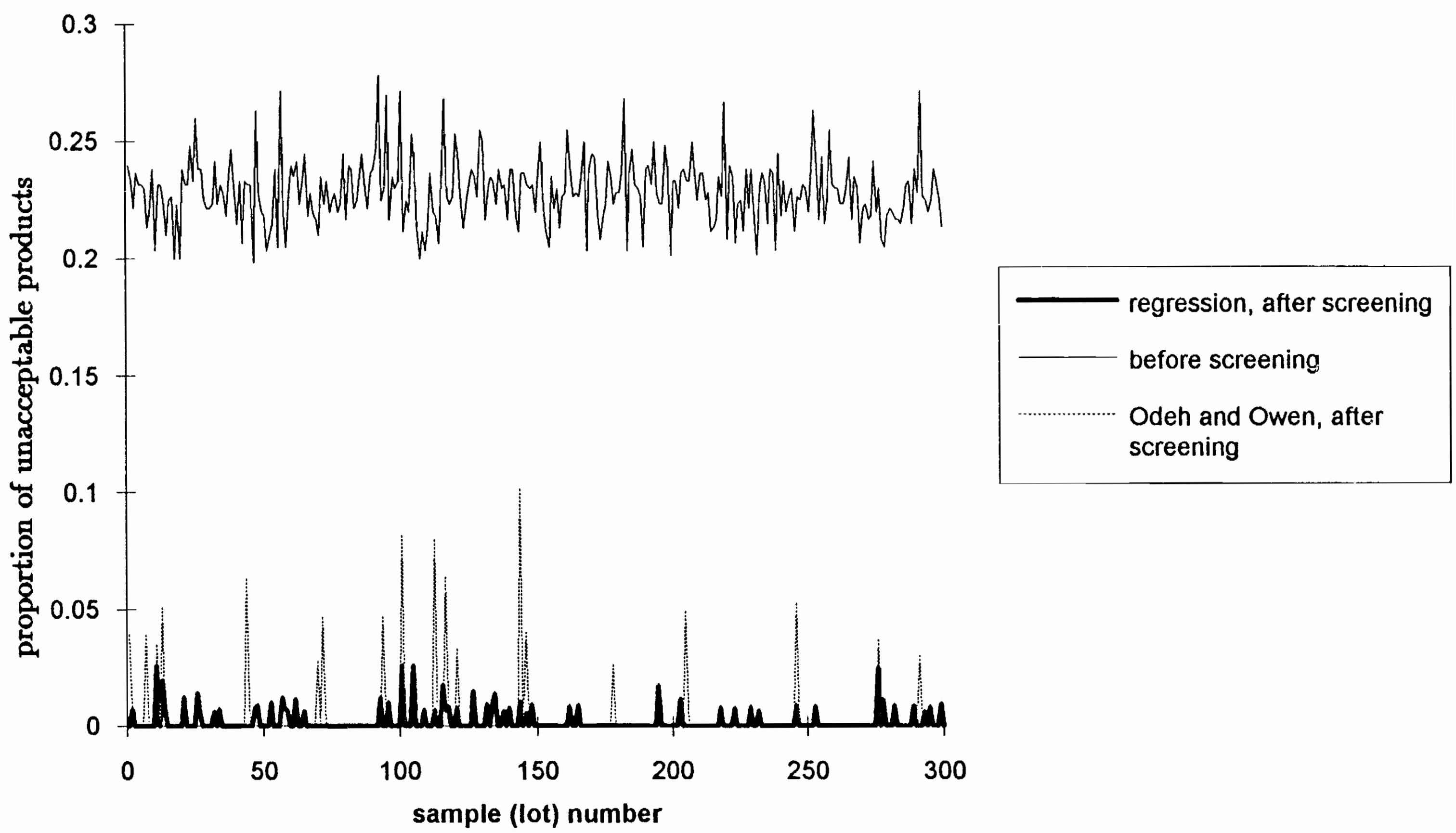

Figure C8: $\eta = 0.99$, $\delta = 0.95$

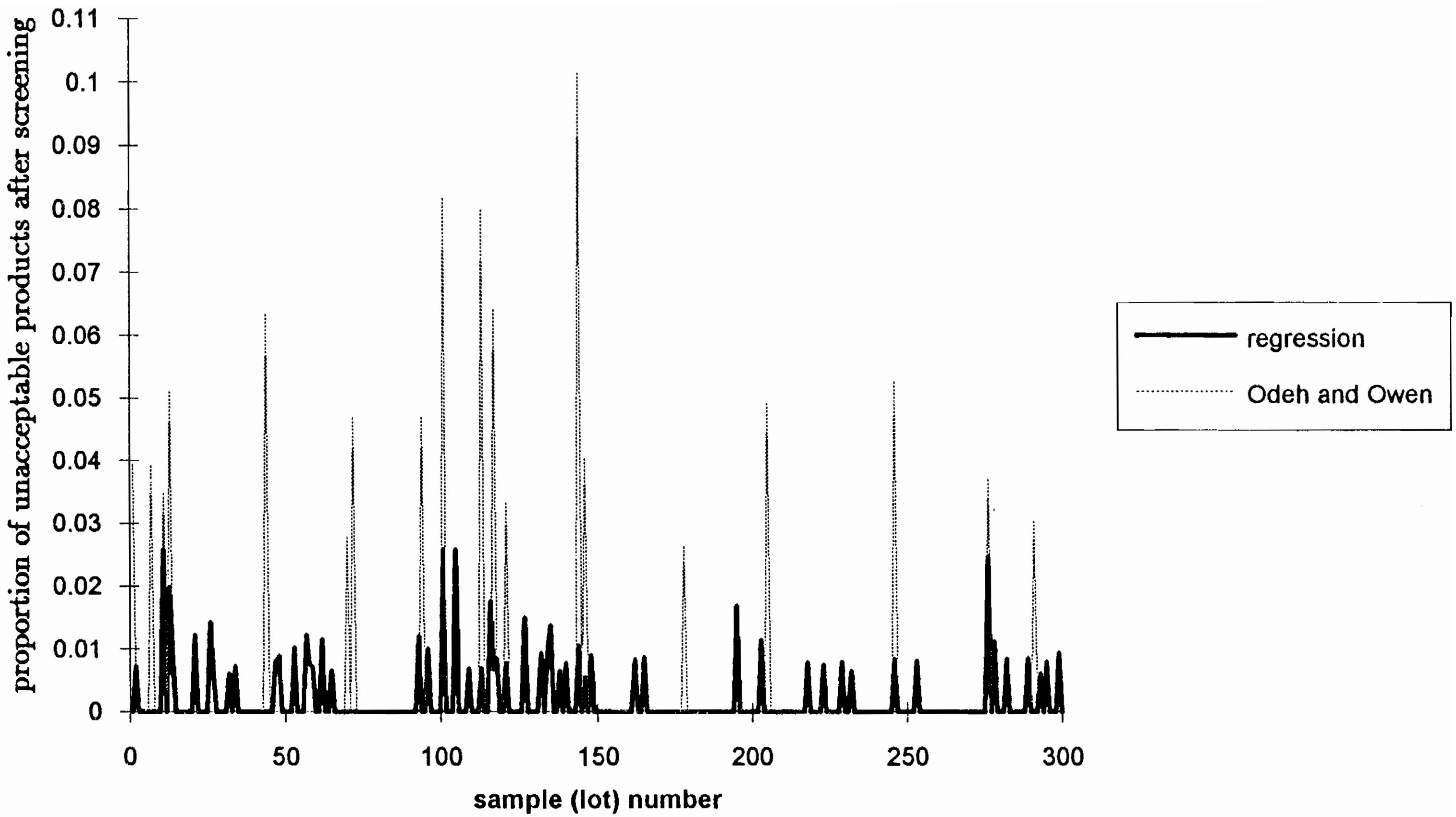

Figure C9: $\eta = 0.99$, $\delta = 0.95$

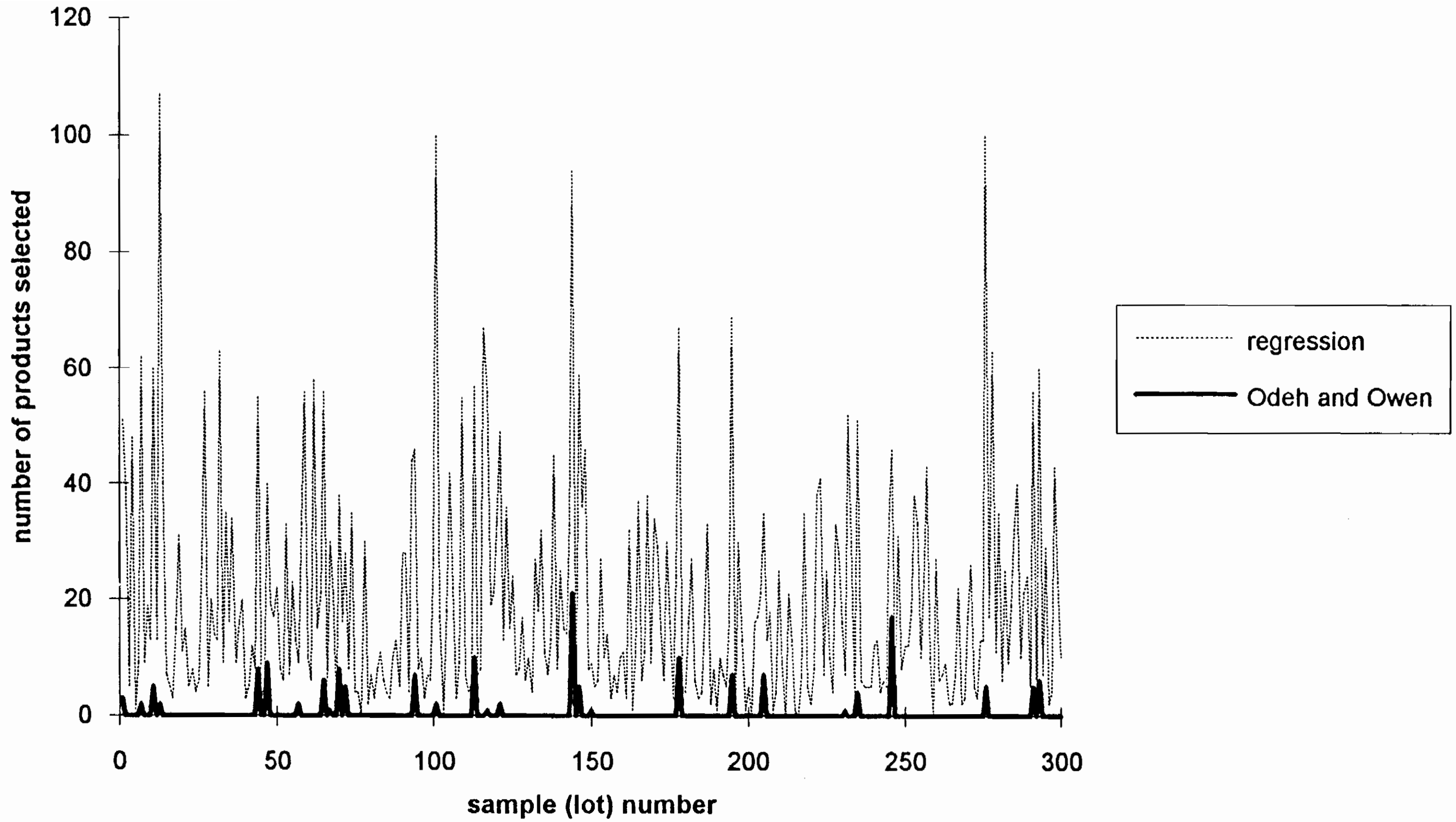

Figure C10: $\eta = 0.99$, $\delta = 0.99$

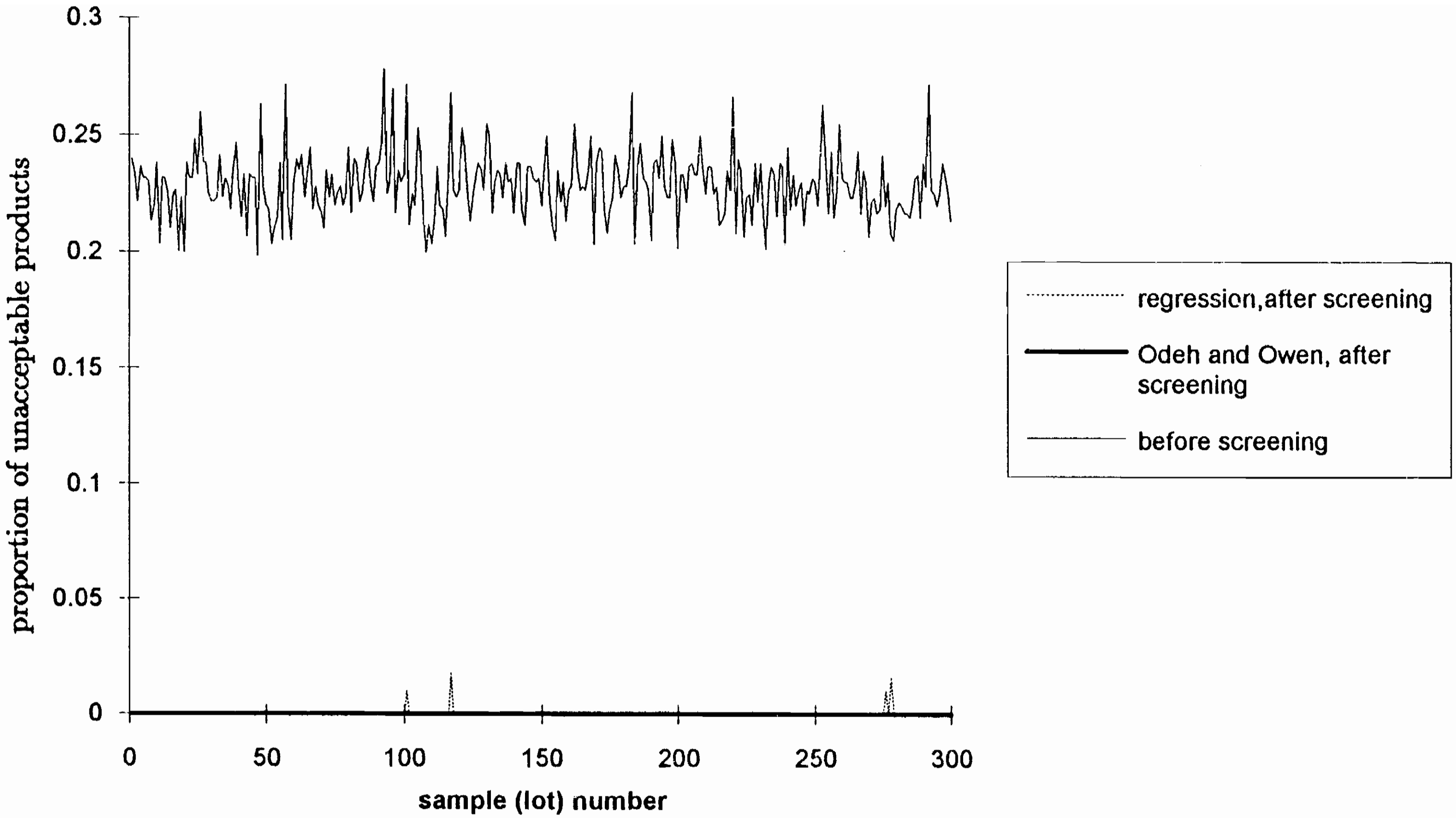

Figure C11: $\eta = 0.99$, $\delta = 0.99$

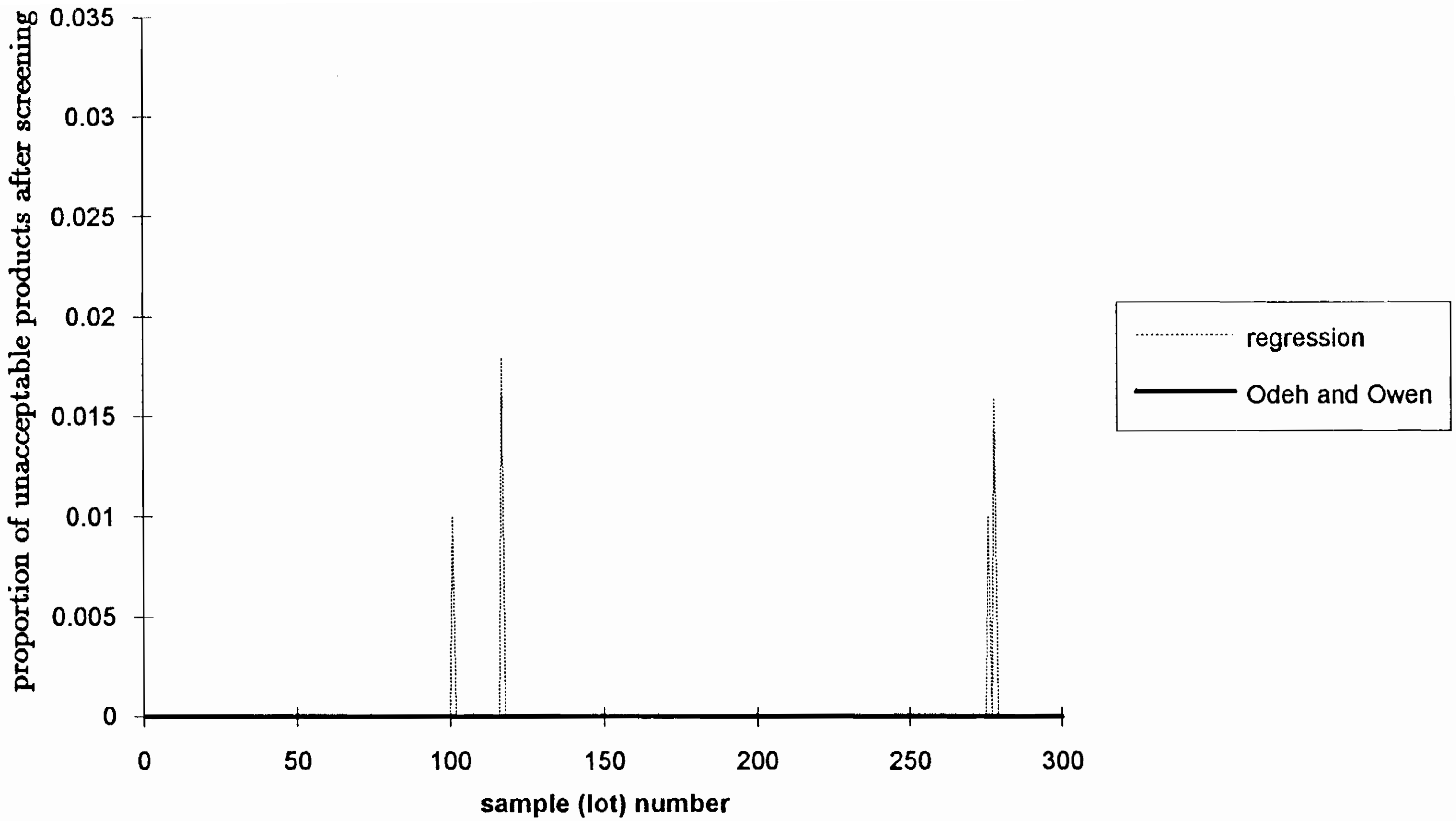

Figure C12: $\eta = 0.99$, $\delta = 0.99$

Appendix D

Known Parameters

In section **3.2.1.3.** comparison of method of Owen et al. with regression method is performed. Screening is based on following decision functions:

The coefficients and parameters for linear function of Owen et al. are:

$$a_1 = \frac{\rho_{YX_1} - \rho_{YX_2}\rho_{X_1X_2}}{\sqrt{(1 - \rho^2_{X_1X_2})(\rho^2_{YX_1} + \rho^2_{YX_2} - 2\rho_{X_1X_2}\rho_{YX_1}\rho_{YX_2})}} = 0.62$$

$$a_2 = \frac{\rho_{YX_2} - \rho_{YX_1}\rho_{X_1X_2}}{\sqrt{(1 - \rho^2_{X_1X_2})(\rho^2_{YX_1} + \rho^2_{YX_2} - 2\rho_{X_1X_2}\rho_{YX_1}\rho_{YX_2})}} = 0.62$$

$$\rho_{YV} = \frac{a_1\sigma_{X_1}\rho_{YX_1} + a_2\sigma_{X_2}\rho_{YX_2}}{\sqrt{(a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + 2\rho_{X_1X_2}a_1a_2\sigma_{X_2})}} = 0.975$$

$$\sigma^2_V = a_1^2\sigma^2_{X_1} + a_2^2\sigma^2_{X_2} + 2\rho_{X_1X_2}a_1a_2\sigma_{X_2} = 0.131$$

$$V = a_1\mu_{X_1} + a_2\mu_{X_2} = 0.62 \text{ x } 10.5 + 0.62 \text{ x } 4.0 = 8.99$$

Proportions which need to be selected are found in Owen et al. (1975):

i) for $\delta = 0.95$

$\psi = 0.766$, giving $z_\psi = 0.72$, then the decision rule is:

accept a product i if $V_\delta \leq V_{\delta i} = 8.99 - 0.72 \times 0.3619$

ii) for $\delta = 0.99$

$\psi = 0.685$, giving $z_\psi = 0.47$, and decision rule is:

accept a product i if $V_\delta \leq V_{\delta i} = 8.99 - 0.47 \times 0.3619$

The regression coefficients and standard deviation are:

$\beta_0 = \mu_Y - \sigma_{YX}\Sigma_{XX}^{-1}\mu_X = 1.689$

$[\beta_1, \beta_2] = \sigma_{YX}\Sigma_{XX}^{-1} = [0.308, 0.520]$

$\sigma_W = \sqrt{\sigma_Y^2 - \sigma_{YX}\Sigma_{XX}^{-1}\sigma_{XY}} = 0.0316$

$\rho_{YW} = 0.992$

and the decision rule is:

i) for $\delta = 0.95$

accept a product i if:

$P_{\delta_i} = 1.689 + 0.308x_{1_i} + 0.520x_{2_i}$ - 1.645 x 0.0316 $\geq L_Y = 6.849$

ii) for $\delta = 0.99$

accept a product i if:

$P_{\delta_i} = 1.689 + 0.308x_{1_i} + 0.520x_{2_i}$ - 2.325 x 0.0316 $\geq L_Y = 6.849$

List of figures:

i) $\delta = 0.95$

Figure D1 - number of product selected from each lot out of 452.6 acceptable on average

Figure D2 - proportion of unacceptable products in each sample, before and after screening

Figure D3 - proportion of unacceptable products after screening

i) $\delta = 0.99$

Figure D4 - number of product selected from each lot out of 452.6 aceptable on average

Figure D5 - proportion of unacceptable products in each sample, before and after screening

Figure D6 - proportion of unacceptable products after screening

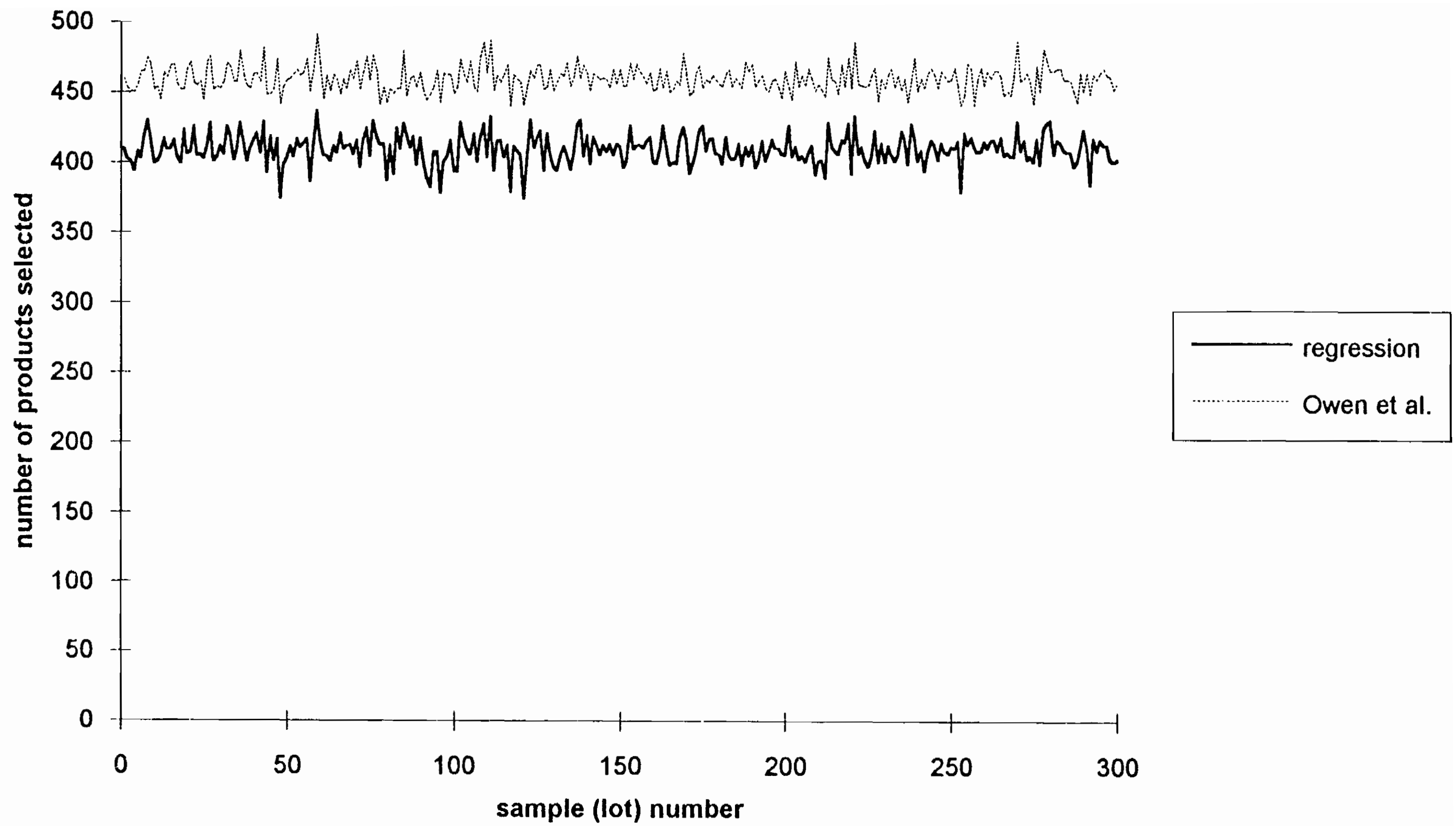

Figure D1: $\delta = 0.95$

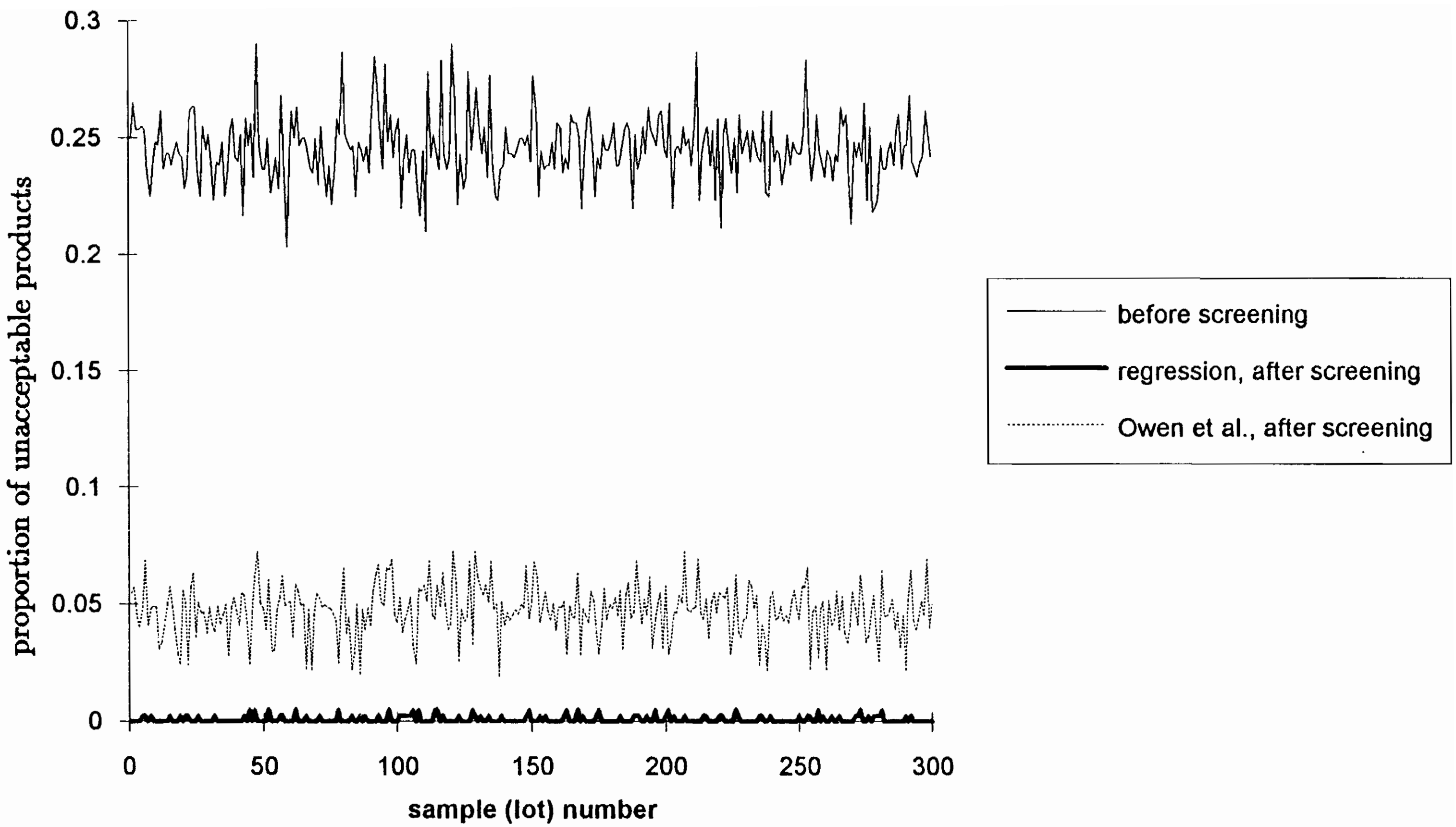

Figure D2: $\delta = 0.95$

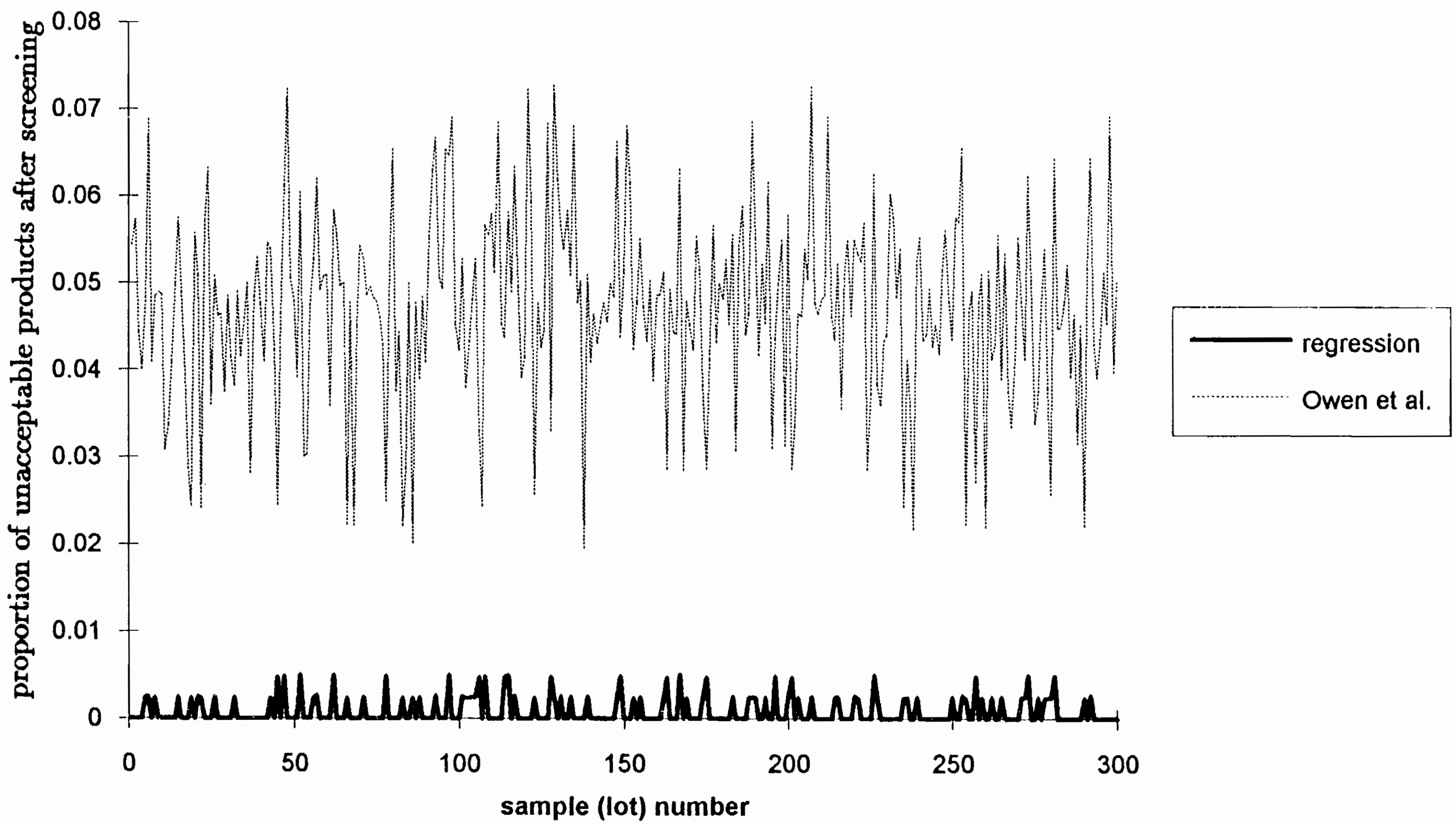

Figure D3: $\delta = 0.95$

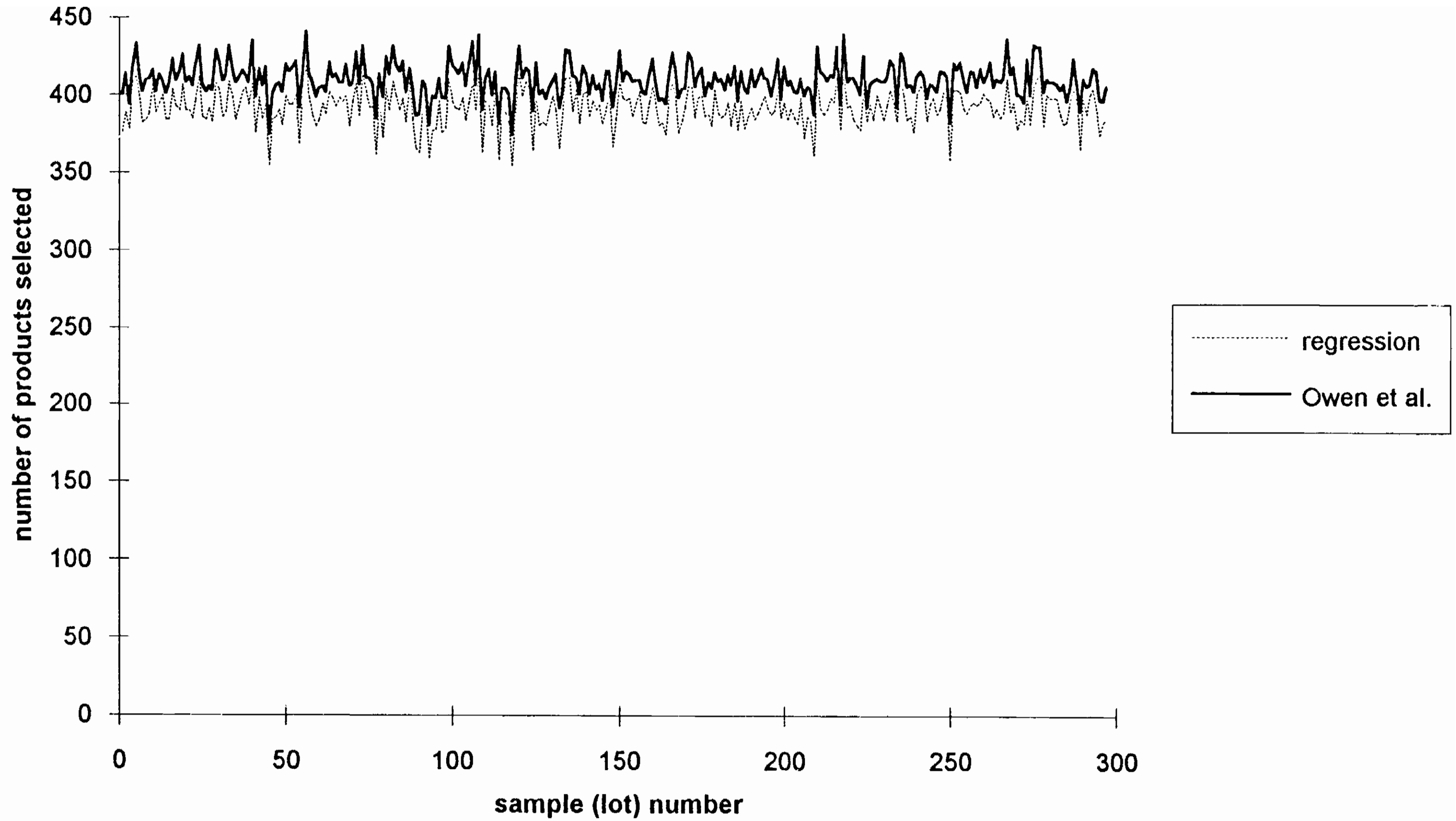

Figure D4: $\delta = 0.99$

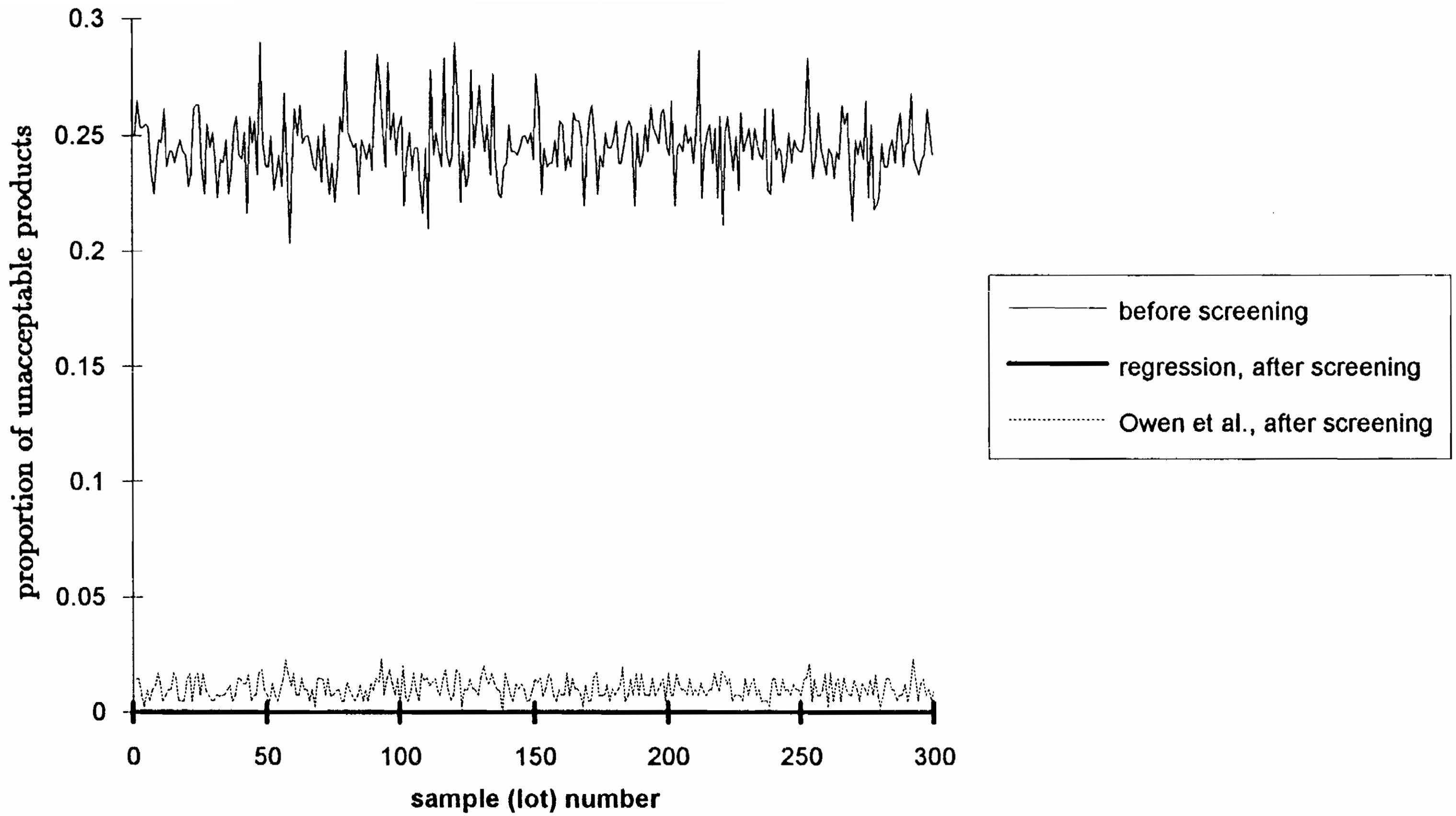

Figure D5: $\delta = 0.99$

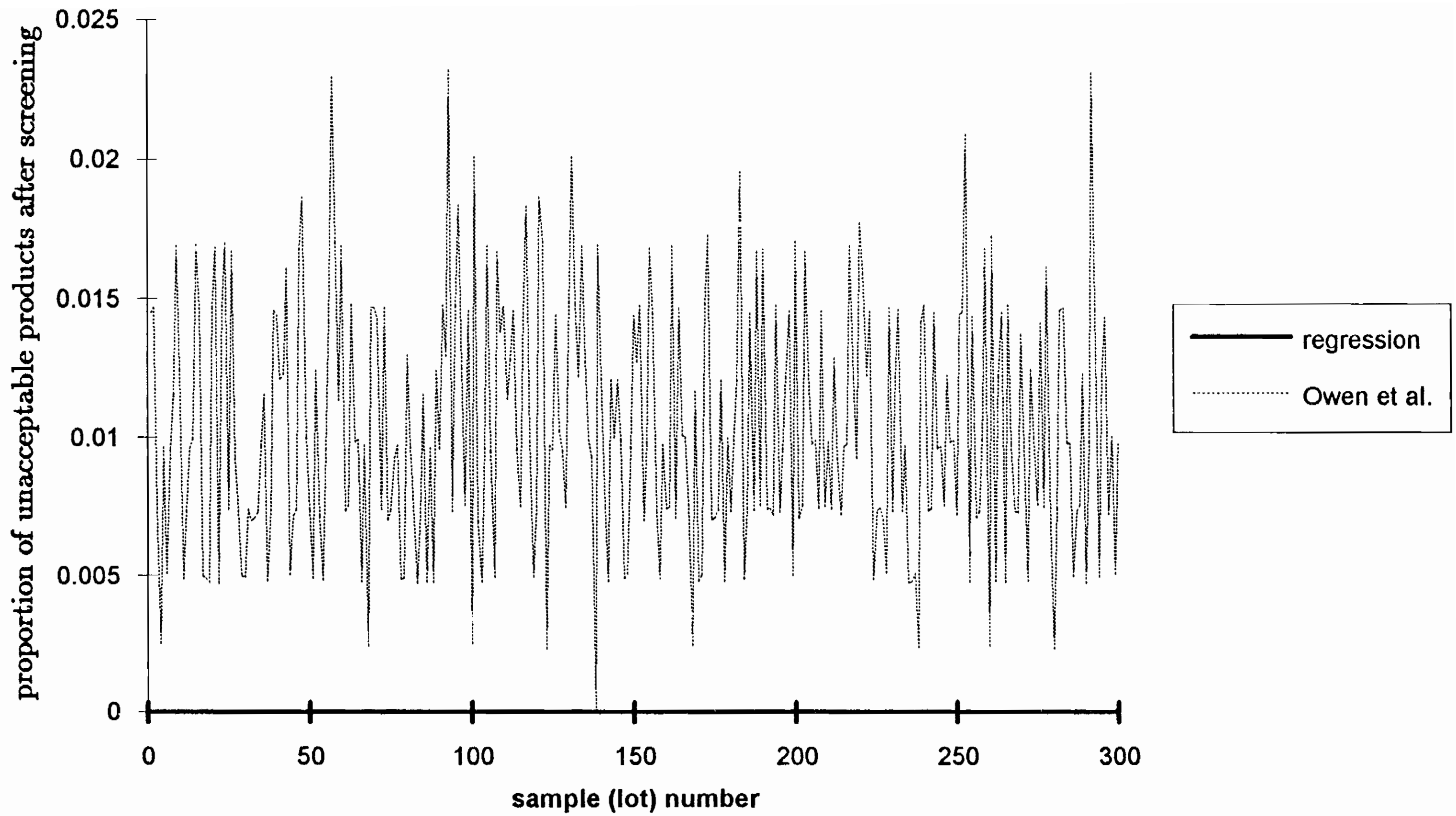

Figure D6: $\delta = 0.99$

Appendix E

Unknown Parameters

detailed results from section **3.2.2.1.**

List of figures:

$\eta = 0.95$, and $\eta = 0.99$

i) $\delta = 0.95$

ii) $\delta = 0.99$

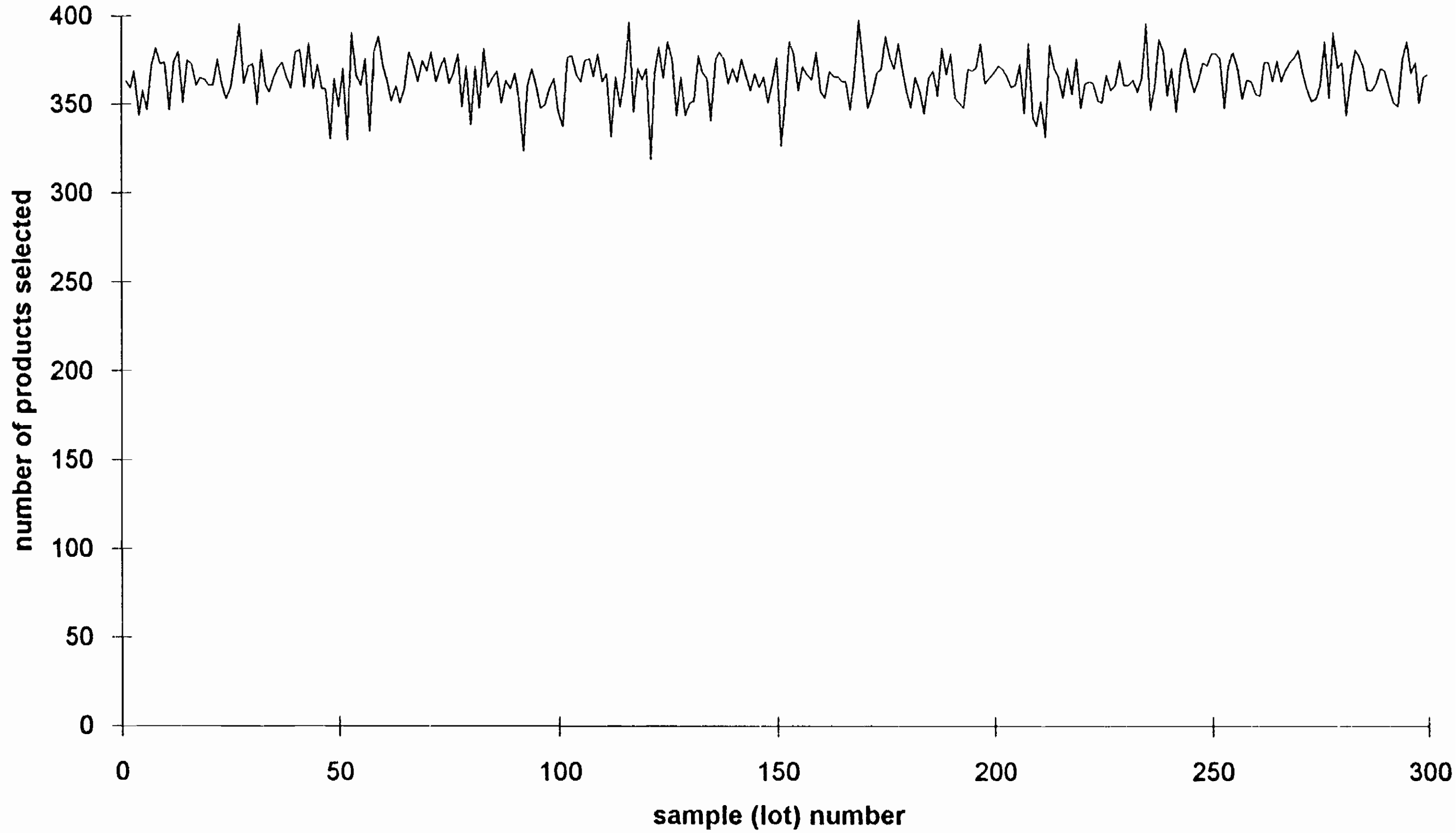

Figure E1: $\eta = 0.95$ and $\eta = 0.99$, $\delta = 0.95$

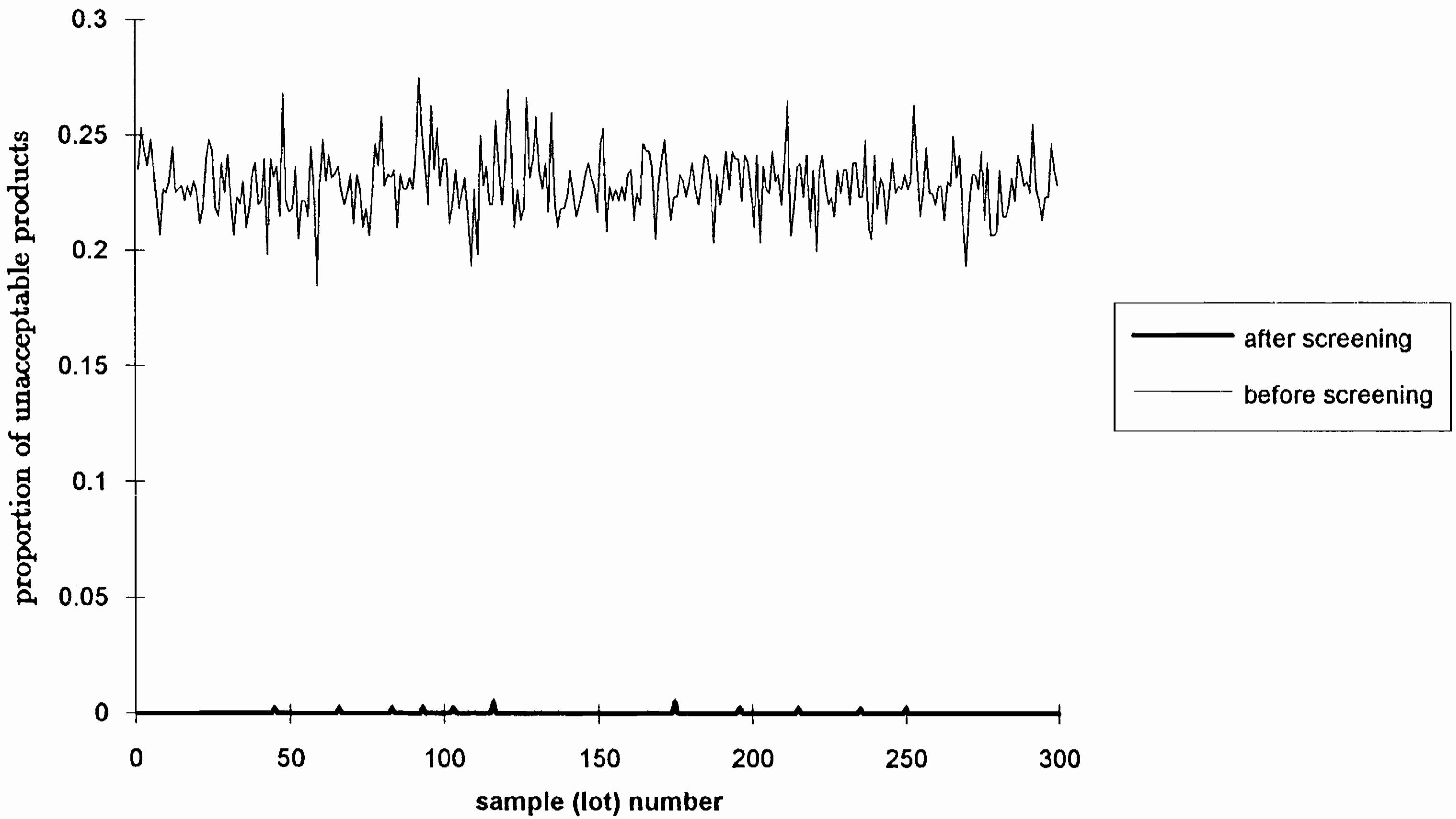

Figure E2: η= 0.95 and η = 0.99, δ = 0.95

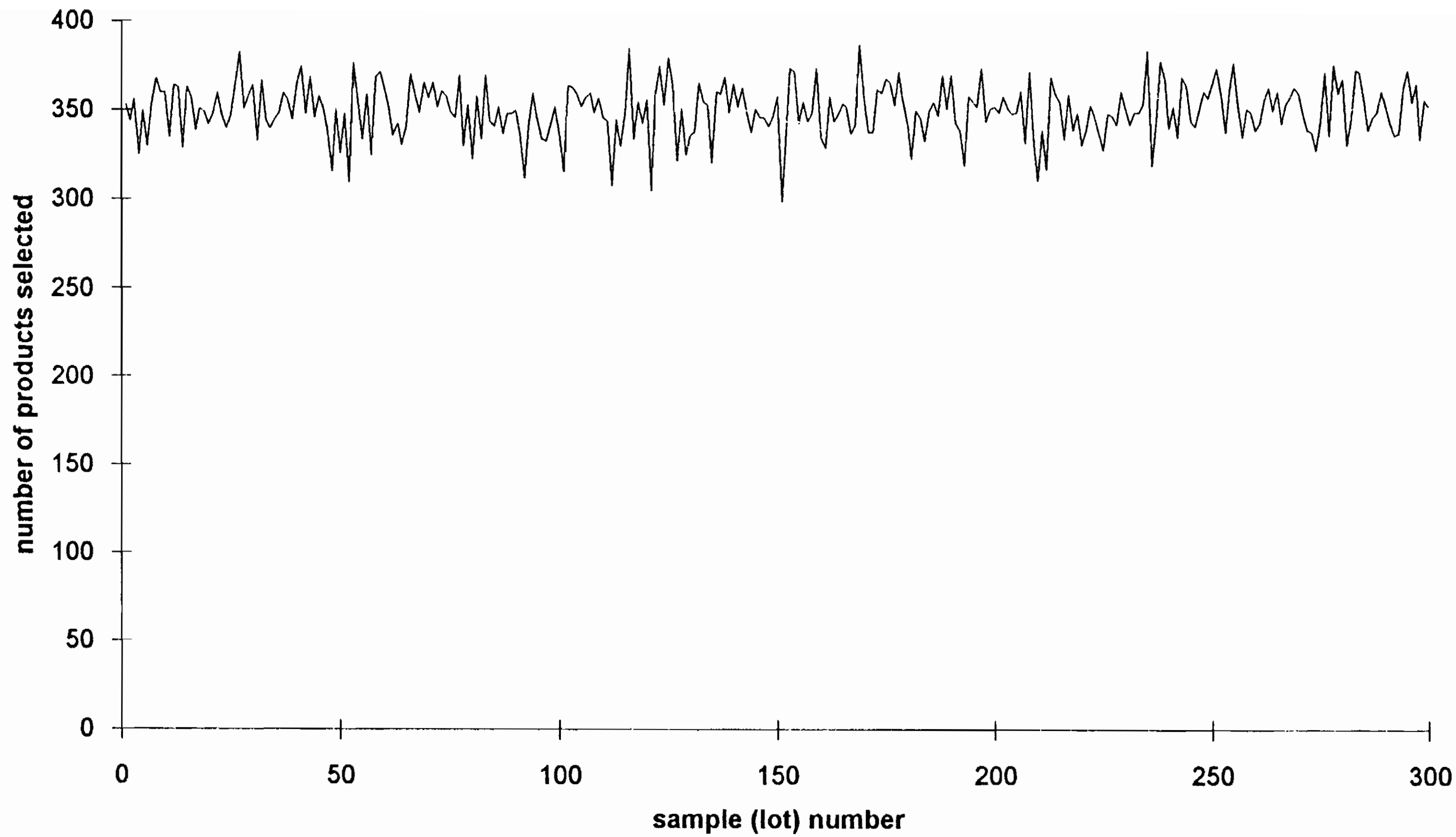

Figure E3: $\eta = 0.95$ and $\eta = 0.99$, $\delta = 0.99$

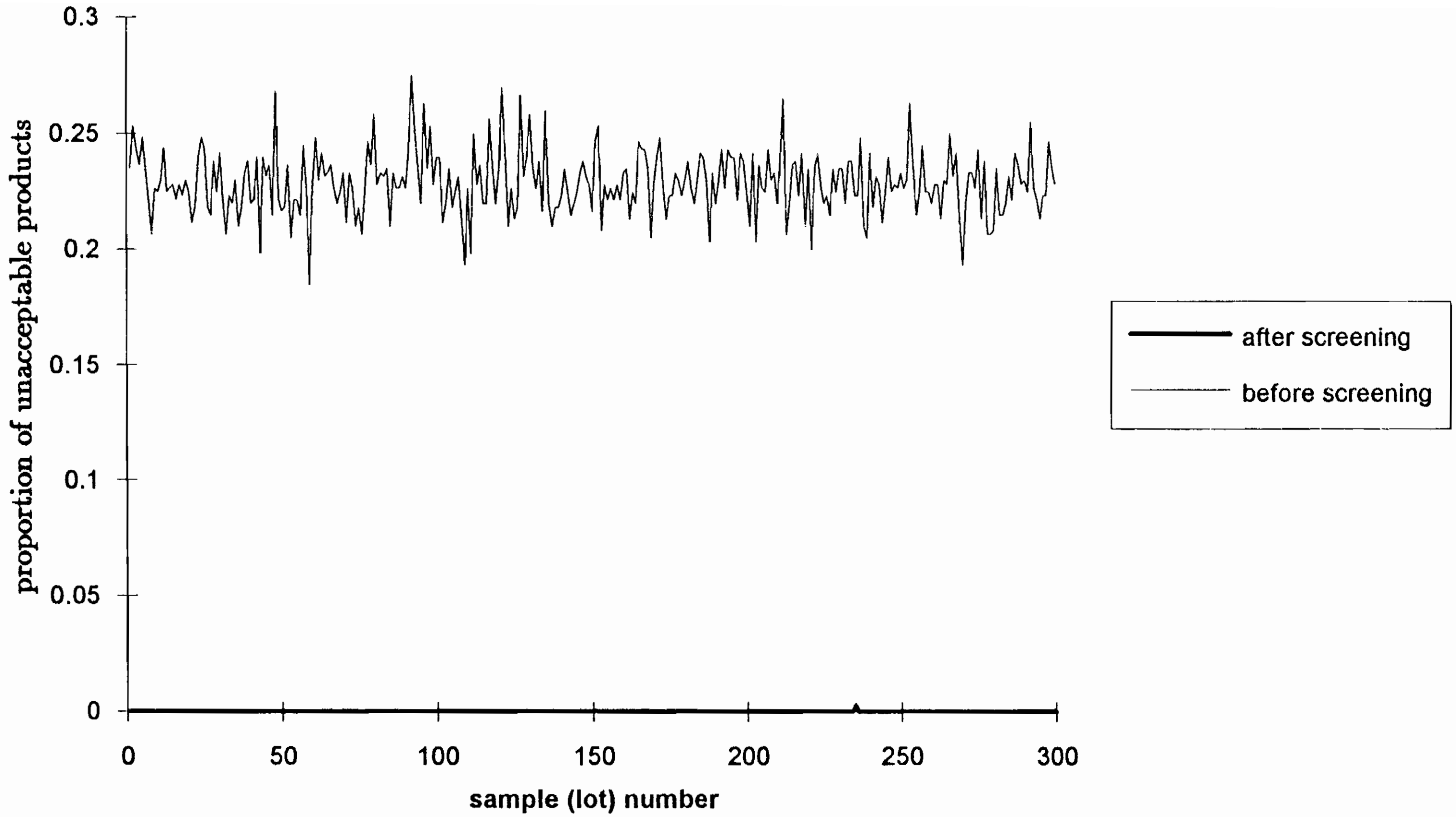

Figure E4: $\eta = 0.95$ and $\eta = 0.99$, $\delta = 0.99$

VITA

Lazar Boskov was born in Novi Sad, Yugoslavia on April 18, 1958. He obtained his B.S. in Economics from University of Novi Sad, Yugoslavia in March, 1987. In December, 1990, he received his M.S. in Statistics from Virginia Polytechnic Institute and State University. He has enrolled in graduate studies in Industrial and Systems Engineering with concentration in operations research at Virginia Polytechnic Institute and State University in January, 1993.