PREDICTION INTERVALS APPLIED TO SCREENING VARIATES

by

D. B. Owen, Youn-Min Chen and Loretta Li

Technical Report No. 130
ONR Contract

PREDICTION INTERVALS APPLIED TO SCREENING VARIATES

by

D. B. Owen, Youn-Min Chen and Loretta Li

Technical Report No. 130
Division of Mathematical Sciences
Department of Statistics
ONR Contract

June 18, 1979

Research sponsored by the Office of Naval Research
Contract N00014-76-C-0613
Project NR 042-341

DEPARTMENT OF STATISTICS
Southern Methodist University

PREDICTION INTERVALS APPLIED TO SCREENING VARIATES

D. B. Owen and Youn-Min Chen
Southern Methodist University
Dallas, TX 75275

Loretta Li
Bishop College
Dallas, TX 75241

Key Words & Phrases: acceptance inspection; bivariate normal probabilities; selection procedures; normal conditioned on t-distribution.

ABSTRACT

The effect of a finite number of items being screened is evaluated, both in the case where all parameters of a bivariate normal distribution are known and where all parameters are unknown. Some illustrative tables are included.

1. INTRODUCTION

A survey of prediction intervals and their applications was given by Hahn and Nelson (1973). More recent work has been done by Fertig and Mann (1975), (1977a) and (1977b). In the present paper we are interested only in normally distributed variables. This paper differs from previous papers on prediction intervals for the normal distribution in that we have available observations on a variable, X, which we will use to screen items for inclusion in a set of items which are required to meet a specification on a correlated variable, Y. That is, we will consider an item to be good if $Y \leq U$, and we want to be able to assure that the number of items meeting this specification is at least k in the group found acceptable by screening on X. The problem is to set a limit on X which will accomplish this with a given probability. We will treat two cases, one where all parameters are known, and one where all parameters are unknown.

2. CASE WHERE ALL PARAMETERS ARE KNOWN

Owen, et al. (1975) proposed a method of using a variable X correlated with a variable Y to screen items so that the proportion of $Y \leq U$ is raised from γ before screening to δ after screening. The variables (X, Y) are assumed to have a joint bivariate normal distribution with means (μ_x, μ_y), respectively; standard deviations, (σ_x, σ_y) respectively and correlation ρ which we will assume is positive. Owen, et al. (1975) give tables of a quantity β so that if all items are accepted for which $X \leq \mu_x + K_\beta \sigma_x$ then the goal of raising the proportion of Y's less than U from γ to δ is accomplished. The quantity K_β is the normal deviate that corresponds to a proportion β in the lower tail of a univariate normal distribution.

Owen, et al. (1975) point out that the proportion δ is an expectation achieved for an entire normal population. If only a finite number m of items are screened then the number of items, V, meeting the specification $Y \leq U$ is a random variable following the binomial distribution with parameters (m, δ). Hence, if we wanted the $P\{V \geq \ell\} = \zeta$, then we need to solve the equation

$$\sum_{j=\ell}^{m} \binom{m}{j} \delta^j (1 - \delta)^{m-j} = \zeta$$

where

$$\delta = P\{Y \leq U | X_i \leq \mu_x + k\sigma_x, \forall\, i = 1, \ldots, m\}$$

for k. Note that we can solve the first equation for δ and obtain

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

where $F_{\zeta,a,b}$ is $100(1 - \zeta)\%$ upper tail percentage point of the F-distribution with a degrees of freedom for the numerator and b degrees of freedom for the denominator. Hence, after computing the δ from this formula we proceed to compute k as in Owen, et al. (1975) from

$$P\{Y \le U | X \le \mu_x + k\sigma_x\} = \delta.$$

The result is shown in the accompanying tables under infinite degrees of freedom for a sample of size 100, and some selected values of ℓ.

Note that the adjustments can be made as in Owen et al. (1975) for both lower specification limits, L, and negative correlation. Hence, the procedures and tables given here apply to any case where there is a one-sided specification limit and either a positive or negative correlation. The steps given above also apply to two-sided specification limits down to the last expression for δ. We are then $100\zeta\%$ sure that at least ℓ of m future values of Y will be less than U when items are selected based on $X \le \mu_x + k\sigma_x$.

3. CASE WHERE ALL PARAMETERS ARE UNKNOWN

In Owen and Su (1977) the following procedure was given to take care of the case where $(\mu_x, \mu_y, \rho, \sigma_x, \sigma_y, \gamma)$ are unknown.

1. Take a preliminary sample $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and estimate $(\mu_x, \mu_y, \rho, \sigma_x, \sigma_y)$ in the usual way, i.e.

$$\bar{x} = \sum_{i=1}^{n} x_i/n, \ \bar{y} = \sum_{i=1}^{n} y_i/n,$$

$$s_x^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1), \ s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n - 1),$$

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)\ s_x\ s_y}.$$

2. Compute a $100\eta\%$ lower confidence limit γ^* for $\gamma = P\{Y \le U\}$ by finding the noncentrality parameter $\sqrt{n}\ K_{\gamma*}$ for a noncentral t-variate, T_f, that satisfies

$$P\{T_f \le k\sqrt{n}\} = \eta$$

where $k = (U - \bar{y})/s_y$. Since $K_{\gamma*}$ is the univariate normal deviate corresponding to γ^* in the lower tail we can easily obtain γ^* and we have

$$P\{\gamma \geq \gamma^*\} = \eta.$$

3. Compute a $100\eta\%$ lower confidence limit ρ^* for ρ. This may be done using David's (1954) tables or approximately from

$$\rho^* = \tanh\left\{\operatorname{arctanh} r - \frac{K_\eta}{\sqrt{n-3}}\right\}$$

4. Obtain the value of k from Owen and Haas (1978) by entering those tables with D.F. = $n - 1$, $D = \delta$, $R = \rho^*\sqrt{\frac{n}{n+1}}$, $G = \gamma^*$.

5. For all additional items accept those for which $X \leq \bar{x} + k\, s_x \sqrt{\frac{n+1}{n}}$.

6. We can then be $100(2\eta - 1)\%$ sure that at least $100\delta\%$ of the Y values in the screened population are less than U.

Again, if there is only a finite number, m, in the screened population then the number of items, V, meeting the specification $Y \leq U$ follows a binomial distribution with parameters (m, δ). Again we want $P\{V \geq \ell\} = \zeta$ where ℓ is some minimum number of items meeting specifications which we want to see among the m items which have been screened, and ζ is the probability of our seeing this result.

The procedure is the same as in Section 2 above except that this time

$$\delta = P\{Y \leq U | X \leq \bar{x} + k\, s_x\}$$

where δ is computed from the F-distribution, as before.

This time the quantity on the right must be obtained from the normal conditioned on t-distribution instead of the normal conditioned on normal distribution as in Section 2. Computational algorithms are given by Owen and Haas (1978) for this.

4. TABLES

In the accompanying tables we give values of k as defined in Sections 2 and 3 for $m = 100$; $\gamma = 0.4, 0.8, 0.9$; $\rho = 0.90, 0.99$, $\zeta = 0.90$ and 0.99, i.e., just a few illustrative values since the tables would be very massive to cover any reasonable range of uses.

TABLE I

Values of k When $m = 100$ and $\zeta = 0.90$

		n - 1 = 30		n - 1 = infinity	
		$\rho = 0.90$	$\rho = 0.99$	$\rho = 0.90$	$\rho = 0.99$
γ	ℓ				
.4	40	1.176	1.176	1.135	1.136
	60	.232	.285	-.227	-.278
	80	-.311	-.073	-.303	-.071
	90	—	—	-.620	-.237
.8	80	1.695	1.722	1.626	1.642
	90	.924	1.110	.909	1.084
.9	90	1.899	1.968	1.826	1.868

TABLE II

Values of k When $m = 100$ and $\zeta = .99$

		n - 1 = 30		n - 1 = infinity	
		$\rho = .90$	$\rho = .99$	$\rho = .90$	$\rho = .99$
γ	ℓ				
.40	40	.800	.805	.777	.781
	60	.092	.176	.090	.172
	80	-.427	-.129	-.377	-.126
	90	—	—	-.743	-.257
.80	80	1.341	1.412	1.303	1.357
	90	.739	.999	.729	.987
.90	90	1.520	1.669	1.486	1.606

5. CONCLUSION

We can be $100(2\eta - 2 + \zeta)\%$ sure that at least ℓ of m future observations on the variate Y will be below U if the m observations on X all have been screened so that

$$X \leq \bar{x} + k\, s_x \sqrt{\frac{n+1}{n}} .$$

The quantity k can be read from the accompanying table for some special cases. However, in most instances it will be necessary to compute k using the methods of Owen and Haas (1978) for the case where the normal conditioned on t probability is set equal to δ as defined above.

BIBLIOGRAPHY

Chew, V. (1968). Simultaneous prediction intervals. Technometrics. 10, 323-330.

David, F. N. (1954). Tables of the Ordinates and Probability Integral of the Correlation Coefficient in Small Samples. University Press, Cambridge.

Fertig, K. W. and Mann, N. R. (1975). A new approach to the determination of exact and approximate one-sided prediction intervals for normal and lognormal distributions, with tables. Reliability and Fault Tree Analysis, edited by R. E. Barlow, J. B. Fussell, and N. D. Singpurwalla, Philadelphia, PA; Society for Industrial and Applied Mathematics.

Fertig, K. W. and Mann, N. R. (1977a). One-sided prediction intervals for at least p out of m future observations from a normal distribution. Technometrics. 19, 167-178.

Fertig, K. W. and Mann, N. R. (1977b). A prediction interval approach to obtaining variables sampling plans for small lots: Single sampling from Gaussian processes. J. Amer. Statist. Assoc. 72, 585-592.

Hahn, G. J. (1969). Factors for calculating two-sided prediction intervals for samples from a normal distribution. J. Amer. Statist. Assoc. 64, 878-888.

Hahn, G. J. (1972). Simultaneous prediction intervals to contain the variability of future samples from a normal distribution. J. Amer. Statist. Assoc. 68, 938-942.

Hahn, G. J. and Nelson, W. (1973). A survey of prediction intervals and their applications. J. Qual. Tech. 5, 178-188.

Kabe, D. G. (1967). On multivariate prediction intervals for sample mean and covariance based on partial observations. J. Amer. Statist. Assoc. 62, 634-637.

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

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

Owen, D. B. and Haas, R. W. (1978). Tables of the normal conditioned on t-distribution. Contributions to Survey Sampling and Applied Statistics. Academic Press, New York.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 130	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PREDICTION INTERVALS APPLIED TO SCREENING VARIATES		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. B. Owen, Youn-Min Chen and Loretta Li, Bishop College, Dallas, TX 75241		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0613
9. PERFORMING ORGANIZATION NAME AND ADDRESS SOUTHERN METHODIST UNIVERSITY Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-341
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH Arlington, Virginia 22217		12. REPORT DATE June 15, 1979
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

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

18. SUPPLEMENTARY NOTES

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

acceptance inspection; bivariate normal probabilities; selection procedures; normal conditioned on t-distribution.

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

The effect of a finite number of items being screened is evaluated, both in the case where all parameters of a bivariate normal distribution are known and where all parameters are unknown. Some illustrative tables are included.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)