J. King Saud Univ., Vol. 2, *Eng. Sci.* (1), pp. 195-200, (1410H/1990)

LETTER TO THE EDITOR

A Computer Package for Computing the Confidence Limits of a Measured Factor

Saad Haj Bakry and Mustafa Shatilla
College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia.

Abstract. This paper is concerned with the development of a computer algorithm for computing the range of values of the confidence limits for a repeatedly measured factor. The paper briefly reviews the theory of the confidence limits for both large and small sample sizes. The presented algorithm uses the theory to compute confidence limits, for both large and small sample sizes. The algorithm is simple to program in any computer language, and it saves its users from having to resort to using statistical tables.

Keywords: Confidence Limits; Measurements; Algorithms.

1. Introduction

Practical experiments, whether on real systems or on models of systems, usually produce a set of different values for a specific factor. Such a set of values does not include all possibilities that can be produced, but it includes a sample of the possibilities; and therefore the set can be described as a sample set. In many cases, the numerical value of a measured factor is usually given by the mean value of the sample set. However, statistical estimation theory suggests that, the value of a measured factor should be given as a range of values, usually known as the confidence interval, depending on the confidence level required [1,2].

The literature dealing with estimation theory provides formulae accompanied with statistical tables for computing the required range of values, for a factor, from its sample set of measurements [1,2]. Using such formulae and tables may be considered by many people working in experimental fields as a tedious and inconvenient job. In addition, in many cases, the available statistical tables may not include the required ranges [2].

This short note describes a computer algorithm, that computes, for any sample set of a measured factor, the range of values of the factor for any required confidence level. The program is based on estimation theory, and can deal with cases where the sample size is large, and also where the sample size is limited [2].

2. Review of the Basic Theory

A sample size of a measured factor is considered to be large, if the number of values in the sample set, N is greater or equal to a specific value, usually given as 30 [2]. If the value of N is less than this specific value, the sample size is considered to be small.

For a large sample, the measured values of a factor (X1, ..., Xn), are considered to be normally distributed. For this case, the confidence limits that specify the confidence level, L (%), are given as follows [1,2]

$$[\,(\bar{X} - Z_{\frac{a}{2}} \frac{S}{\sqrt{N}}),\ (\bar{X} + Z_{\frac{a}{2}} \frac{S}{\sqrt{N}})\,] \tag{1}$$

$\bar{X}$: is the mean value of the sample set, and is given as follows;

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} Xi \tag{2}$$

S : is the standard deviation of the sample and is given as follows;

$$S = \sqrt{\left(\frac{1}{N-1}\right) \sum_{i=1}^{N} (Xi - \bar{X})^2} \tag{3}$$

$Z_{\frac{a}{2}}$ is the value of the standard normal distribution leaving area of (a/2) to the right. (a) is related to the confidence level L (%) as follows;

$$L = 100\,(1\text{-}a) \tag{4}$$

The equation from which $Z_{\frac{a}{2}}$ can be computed is given below.

$$(0.5 - \frac{a}{2}) = \frac{1}{\sqrt{\pi 2}} \int_0^{Z\frac{a}{2}} e^{-(u^2/2)}\, du \tag{5}$$

For a small sample, the measured values are considered to be based on the t-distribution. The confidence limits for this case are given as follows [1,2].

$$[(\bar{X} - t_{\frac{a}{2}}\frac{S}{\sqrt{N}}), (\bar{X} + t_{\frac{a}{2}}\frac{S}{\sqrt{N}})] \tag{6}$$

$t_{\frac{a}{2}}$ is the value of the t-distribution leaving an area of (a/2) to the right. The equation from which $t_{\frac{a}{2}}$ can be computed is given as follows.

$$(0.5 - \frac{a}{2}) = \frac{\Gamma(N/2)}{\Gamma[(N-1)/2]\sqrt{\pi(N-1)}} \int_0^{t_{\frac{a}{2}}} \left(1 + \frac{u^2}{N-1}\right)^{-(N/2)} du \tag{7}$$

The gama function (V) is given as follows [2]

$$\Gamma(V) = \int_0^{\infty} u^{V-1} e^{-u} du = (V-1)! \tag{8}$$

3. The Proposed Algorithm

A computer algorithm that implements the theory reviewed above has been developed, and is shown in Fig. 1. The algorithm consists of the following four major stages:

- The first stage is concerned with the inputing of data including: the sample size N, the given values (X1,...,Xn), and the required confidence level L (%).
- The second stage is concerned with computing the basic needed information including: the mean value, X, the standard deviation, S, and the area (a/2). The computation required for this stage can be easily performed using formulae (2), (3) and (4).
- The third stage is concerned with computing the confidence coefficient, and consequently, the confidence limits for the given sample, whether large or small. This stage is the most important stage in the algorithm, and further discussion on this stage is given below.
- The last stage in the algorithm is concerned with reporting the results in the output.

In the third stage, computing the confidence coefficients $Z_{\frac{a}{2}}$, and $t_{\frac{a}{2}}$ from the formulae (5) and (7), for large and small samples, requires the use of numerical iteg-

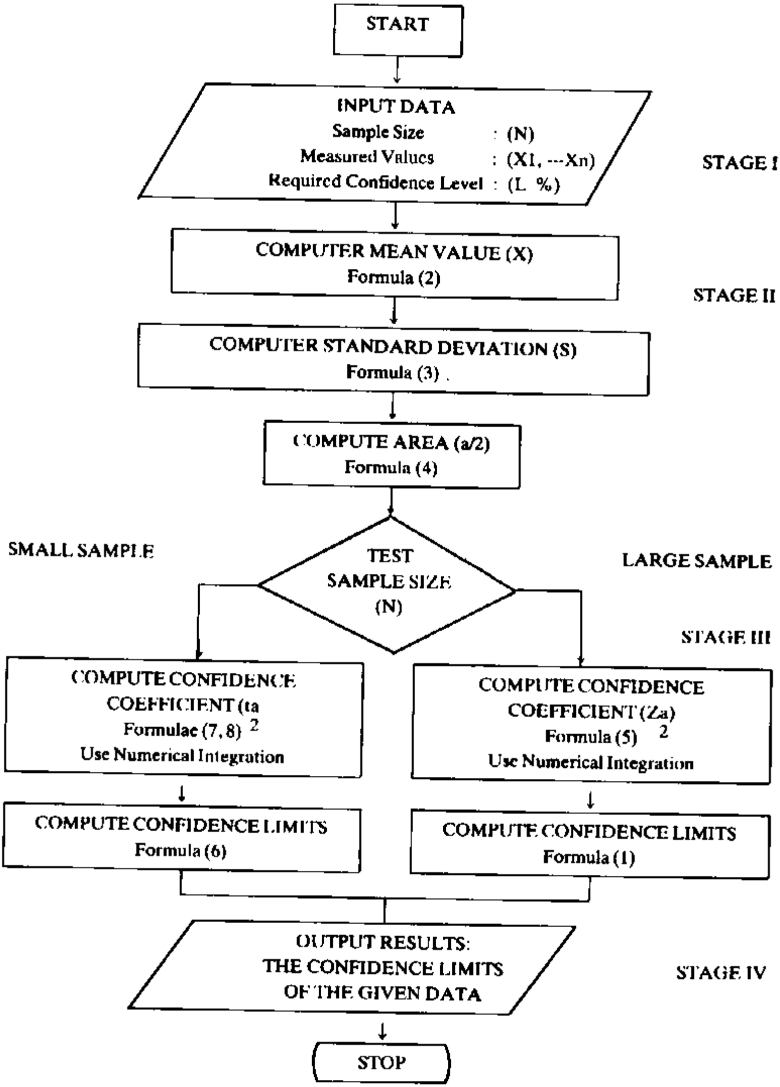

Fig. 1. A general computer algorithm for computing the confidence limits of a repeatedly measured factor

ration techniques. Various basic and composite numerical integration rules are available in the literature [3, 4]. Such rules differ in their accuracy and complexity. A reasonable and easy to use numerical integration rule that has been used in programming the proposed algorithm is the Simpson's 1/3 rule [3, 4]. The developed program allows its users to set the value of the integration increment depending on the required accuracy.

In addition to the integration part in formula (7), used for computing $t_{\frac{a}{2}}$, gama function $\Gamma(V)$ also needs to be computed. This can either be done, as above, using numerical integration rules, or by computing the factorial of (V-1), as suggested by formula (8) [2, 4]. Computing the factorial of an integer value can be easily programmed by a multiplication loop. However, if the factorial is required for a real value, Sterling's formula for factorials can be used [5]. In programming the gama function, both the numerical integration and the factorial methods have been tested, and the factorial method was chosen, as it requires less computer time.

In order to ensure the validity of the resulted computer program, which is based on the suggested algorithm, the program was tested for two cases, that have been previously considered in [2], using available statistical tables. The first case has a large sample size of $N = 36$, while the second has a small size of $N = 7$. The first case is given in page (189) of reference [2], and the second is in page (192) of the same reference. The results produced by the computer program of the algorithm are exactly the same results given in the reference.

4. Conclusion

The computer package presented in this short paper would help engineers and scientists, working in experimental fields, in computing the necessary confidence limits for their results. The algorithm is easy to program, and it saves its users from having to look for, and look into statistical tables.

References

[1] Kobayashi, H. *Modeling and Analysis: An Introduction to System Performance Evaluation Methodology*. Reading, Mass: Addison-Wesley, 1978.

[2] Walpole, R. and Myers, R. *Probability and Statistics for Engineers and Scientists*. U.K.: Collier Macmillan, 1972.

[3] Conte, S.D. and De Boor, C. *Elementary Numerical Analysis: An Algorithmic Approach*, New York: McGraw-Hill, 1972.

[4] Gerald, F.C. *Applied Numerical Analysis*. Reading, Mass: Addison-Wesley, 1978.

[5] Hughes, W.E. and Gaylord, E.W. *Basic Equations of Engineering Science*, Schaum's Series, New York: McGraw-Hill, 1964.

(*Manuscript Received:* 13.3.1988; *Accepted:* 25.4.1988)

مجموعة حاسوبية لاستخراج حدود الثقة في قياسات العوامل

سعد الحاج بكري ومصطفى شاتيلا

كلية الهندسة، جامعة الملك سعود، ص.ب ٨٠٠، الرياض ١١٤٢١، المملكة العربية السعودية

ملخص البحث: يهتم هذا البحث بتطوير مجموعة حاسوبية تقوم باستخراج حدي مجال القيم الموثوقة لعامل تم قياسه مرات متكررة. وفي سبيل ذلك يقدم البحث مراجعة مختصرة لنظرية حدود القيم الموثوقة لعينات كبيرة أو صغيرة من القيم المقاسة، حيث تعتمد المجموعة الحاسوبية المطورة على هذه النظرية. وتتميز المجموعة الحاسوبية الناتجة بسهولة البرمجة، كما أنها توفر على مستخدميها مصاعب الاستعانة بالجداول الإحصائية.