Removing the Effects of Measurement Errors in Constructing Statistical Tolerance Intervals

JOHN L. JAECH

Exxon Nuclear Company, Inc., Bellevue, Washington 98009

Items are selected at random from a large lot and measured for some quality characteristic. The producer wishes to make the following type of statement about the lot: "With 100γ percent confidence, the quality characteristic of at least 100*P* percent of the items in the lot exceed *L*." Although this statement is expressed as a statistical tolerance statement, in effect it is also an acceptance sampling plan whenever the lot is to be rejected when *L* is less than some lower specification limit, and the results in this paper apply in either of these two situations. This paper considers how the effect of measurement error may be taken into account when calculating the tolerance limit.

Introduction

In a recent paper, Hahn (1982) presented a method for determining whether a required proportion of a product meets specifications when the observations are subject to measurement error. His presentation centered around a specific example in which a household product is packaged in containers that are labeled as providing twelve ounces net weight of product. The actual net weight of the product was assumed to vary randomly, following a normal distribution with an unknown mean μ_A and standard deviation σ_A. Under this assumption, 95 percent of the product net weights exceed $L_A = \mu_A - 1.645\sigma_A$. A consumer protection agency requires that L_A must be 12 ounces or more.

Containers of the product are selected at random from the lot and weighed. A known average container weight is subtracted from the measured weight to obtain the estimated product net weight. This process introduces two possible errors of measurement: (1) the error in the observed weight of the filled container; and (2) the difference between the average container weight and the individual container weight. The simple model that describes this measurement process is given as

$$M = A + R + C \tag{1}$$

where M is the measured net weight of the product, A is the unknown actual net weight, R is the unknown measurement error in measuring the gross weight, and C is the unknown container weight. Thus, the actual net weight is

$$A = M - R - C \tag{2}$$

and for uncorrelated A, R, and C, it follows that

$$\sigma_A^2 = \sigma_M^2 - \sigma_R^2 - \sigma_C^2. \tag{3}$$

Hahn (1982) considers the case where σ_R^2 and σ_C^2 are known quantities, and where σ_M^2 is the variance of the observed gross weights. He then applies (3) to calculate σ_A^2, denoting this calculated quantity by $\hat{\sigma}_A^2$. The estimated five percent point of the distribution of actual weights is then

$$\hat{L}_A = \hat{\mu}_A - 1.645\ \hat{\sigma}_A \tag{4}$$

where $\hat{\mu}_A$ is the average measured net weight, that is, $\hat{\mu}_A = \hat{\mu}_M$.

This approach is reasonable as long as the sample size is sufficiently large so that sampling variability in the estimates can be ignored. For small samples, sampling variability must be taken into account.

Mr. Jaech is a Staff Consultant for Statistics/Safeguards in the Fuel Engineering and Technical Services Department. He is a Senior Member of ASQC.

KEY WORDS: Measurement Error, Tolerance Limits

An approximate method suggested by Hahn (1982) is to calculate the lower confidence bound on the population percentage of *measured* values meeting the specifications limit. The interest, of course, is not in the measured values, but in the actual values. Hahn (1982) suggests that an approximate bound can be found by using the estimate $\hat{\sigma}_A$ as if it were an observed standard deviation, having the same number of degrees of freedom as $\hat{\sigma}_M$, that is, one less than the sample size. The purpose of this paper is to examine the quality of this approximation. An alternate approach is also suggested and evaluated.

Although the problem is treated with respect to a lower specification limit, the results clearly apply to any one-sided specification, upper or lower.

Calculation of Tolerance Limit

The problem solution discussed by Hahn (1982) is modified slightly. Rather than finding the 95 percent (say) confidence bound on the percentage of actual values that exceed the specification limit of 12 ounces, a 95/95 lower tolerance bound is calculated and compared with the lower specification limit. The expression, "95/95" is used here to describe a bound above which 95 percent of the population lies with 95 percent confidence. In general, the value above the slash refers to the degree of confidence and the value below the slash to the percent of the population lying above the bound. The problem treated by Hahn (1982) is generalized as described in the following paragraphs, with the results then applied to his specific example.

In the generalized problem, the random variables R and C in (1) are combined to form a single variable $V = R + C$. This does not change the essence of the problem in any way. Thus, (3) is rewritten

$$\sigma_A^2 = \sigma_M^2 - \sigma_V^2. \tag{5}$$

The steps in constructing the lower tolerance limit following Hahn's (1982) approximation are

(a) For the sample of size N, estimate σ_M^2, calling the estimate

$$\hat{\sigma}_M^2$$

(b) For given σ_V^2, calculate

$$\hat{\sigma}_A^2 = \hat{\sigma}_M^2 - \sigma_V^2 \tag{6}$$

(c) Calculate the 95/95 lower tolerance limit:

$$\mathrm{L} = \hat{\mu}_M - k_N \hat{\sigma}_A \tag{7}$$

where k_N is the 95/95 one-sided tolerance limit factor for sample size N found in Table A-7 of Natrella (1963) or in Table 2.4 of Owen (1963).

(d) Letting L_0 be the lower specification limit, if $L > L_0$, the specification is met.

In Hahn's (1982) specific example, $N = 30$, $k_{30} = 2.220$, and $\sigma_V/\hat{\sigma}_A = 0.59$. The key question is whether or not the $k_{30} = 2.220$ factor is warranted, that is, whether or not $\hat{\sigma}_A$ has approximately the 29 degrees of freedom $\hat{\sigma}_M$ does. Intuitively, if σ_V^2 is small relative to $\hat{\sigma}_A^2$, one would expect that using 29 degrees of freedom to obtain $k_{30} = 2.220$ is an approximate tolerance limit factor, for in this instance $\hat{\sigma}_A$ is nearly equal to $\hat{\sigma}_M$. The question is, how small is small? Before examining this question by a simulation study, an alternate approach to calculating the lower limit is presented. This alternate approach requires finding an improved approximation of the degrees of freedom associated with $\hat{\sigma}_A^2$ and is given in a later section. First, however, the results of this paper are summarized.

Results

An improved approximation to the degrees of freedom (df) is given by Satterthwaite (1946):

$$df\,(2) = (N - 1)(1 - \sigma_V^2/\hat{\sigma}_M^2)^2. \tag{8}$$

Our simulation study shows that one should use $df\,(2)$ rather than $(N-1)$ as the degrees of freedom. The use of $df\,(2)$ will give conservative results as long as the following inequality is satisfied:

$$\sigma_V/\sigma_A < -0.4 + 0.5 \ln N. \tag{9}$$

A result is defined to be conservative if $100P$ in the calculated $95/100P$ tolerance bound exceeds 95. The estimate of $100P$ is given in Figures (1)–(5) as a function of σ_V/σ_A and $N = 5$, 10, 17, 37, and 59. In these figures, P_i/n is the same as P.

Three curves are drawn in each figure. The three curves correspond to $a_i = 0$, 1.5, and 2 in the general expression for degrees of freedom:

$$(N - 1)(1 - \sigma_V^2/\hat{\sigma}_M^2)^{a_i}. \tag{10}$$

Note that $a_i = 0$ gives $(N-1)$ degrees of freedom, and $a_i = 2$ gives the recommended formula (8). As an alternative to (8), $a_i = 1.5$ may be used in those instances where (8) is too conservative.

The inequality (9) was derived from Figures (1)–(5) by noting those regions where $100P$ exceeds 95. One may also use Figures (1)–(5) to estimate P when inequality (9) is not satisfied, as will be illustrated later in an example.

Before giving the basis for these summary results, a numerical example is considered.

Example

In his example, Hahn (1982) used the following values when computing the 95/95 lower tolerance limits:

$$N = 30$$
$$\hat{\sigma}_M = 0.22$$
$$\sigma_V = \sqrt{0.0125}$$
$$\hat{\mu}_A = 12.27.$$

If the sample size used to estimate μ_A were $N = 30$ and the degrees of freedom used to estimate $\hat{\sigma}_A^2$ were $(N - 1) = 29$, then the 95/95 one sided tolerance limit factor would be 2.220. With $\hat{\sigma}_A = \sqrt{0.0484 - 0.0125} = 0.1895$, the lower limit would be

$$12.27 - (2.220)(0.1895) = 11.849 \text{ oz.}$$

If the sample size used to estimate μ_A were $N = 30$ and the degrees of freedom used to estimate $\hat{\sigma}_A^2$ were calculated by Satterthwaite's formula, (8) then

$$df\ (2) = 29(1 - 0.0125/0.0484)^2 = 16.0.$$

The tolerance limit factor would be 2.486 and the limit would be

$$12.27 - (2.486)(0.1895) = 11.799 \text{ oz.}$$

In this example, $\sigma_V/\hat{\sigma}_A = 0.59$ and $N = 30$. From (9), df (2) may be applied if $\sigma_V/\sigma_A < -0.4 + 0.5 \ln 30$, or $\sigma_V/\sigma_A < 1.30$.

Since $0.59 < 1.30$, df (2) will give conservative results. On the other hand, if the 95/95 limit is applied based on 29 degrees of freedom, it is noted from Figure (3) or (4) that $100P$ will be somewhere between 89 ($N = 17$) and 91 ($N = 37$) percent. We now provide the basis for these results.

Degrees of Freedom for $\hat{\sigma}_A^2$

Satterthwaite's (1946) formula may be used to calculate the approximate degrees of freedom associated with $\hat{\sigma}_A^2$ where σ_V^2 is assumed to be a known quantity, that is, to have infinite degrees of freedom.

$$df\ (2) = \frac{(\hat{\sigma}_M^2 - \sigma_V^2)^2}{\hat{\sigma}_M^4/(N-1) + \sigma_V^2/\infty} = (N-1)(1 - \sigma_V^2/\hat{\sigma}_M^2)^2. \quad (11)$$

For the number of degrees of freedom calculated using (11), the k factor may be calculated by interpolation in Owen's (1963) Table 4 or Natrella's (1963) Table A.7 for given sample size N.

Satterthwaite's (1946) formula that leads to the expression for df (2) in (11) is derived by assuming that $\hat{\sigma}_A^2$ has the chi-square distribution. The second moment of $\hat{\sigma}_A^2$, expressed as a function of df (2), is equated to the second moment of $(\hat{\sigma}_M^2 - \sigma_V^2)$ and solved for df (2) in deriving (11). As part of the simulation study, the respective chi-square distribution third moments were also equated to one another and solved for df (3) yielding the result

$$df\ (3) = (N-1)(1 - \sigma_V^2/\hat{\sigma}_M^2)^{3/2}. \quad (12)$$

The simulation study conducted to investigate the effect of replacing

$$df\ (1) = (N-1) \quad (13)$$

by df (2) or df (3) is now discussed. Note that df (1), df (2), and df (3) may all be expressed in the same form

$$df\ (i) = (N-1)(1 - \sigma_V^2/\hat{\sigma}_M^2)^{a_i} \quad (14)$$

where $a_1 = 0$, $a_2 = 2$, and $a_3 = 3/2$.

Simulation Study Design

In the simulation study, the sample size N was 5, 10, 17, 37, and 59 (chosen to correspond to sample sizes tabled by Owen (1963)), and σ_V/σ_A was set at 0.25, 0.50, 0.75, 1.00, 1.25, and 1.50. All 30 combinations of N and σ_V/σ_A were run with a minimum of 5,000 simulation runs per case. With the tolerance limit factor calculated by using degrees of freedom $(N - 1)$ and those given by (11) and (12), the lower limit was calculated and compared with the lower 5 percent point of the population. For $i = 1$, 2, and 3 corresponding to the three methods of calculating degrees of freedom, P_i was the number of times the lower 5 percent point was exceeded in n simulation runs. In order for the 95/95 statement to be valid, $100P_i/n$ must equal or exceed 95 percent.

Simulation Study Results

The results of the simulation study are displayed in a series of figures.

Note that the plot of $100P_i/n$ for $a_i = 0$ is always below the 95 percent line as one would expect. This means that by assigning $(N - 1)$ degrees of freedom to $\hat{\sigma}_A^2$, the calculated 95/95 limit will actually cover less than 95 percent of the population 95 percent of the time. The actual coverage will be considerably smaller than 95 percent when the measurement error is relatively large.

On other other hand, regions can be identified in which the use of Satterthwaite's approximation (a_i

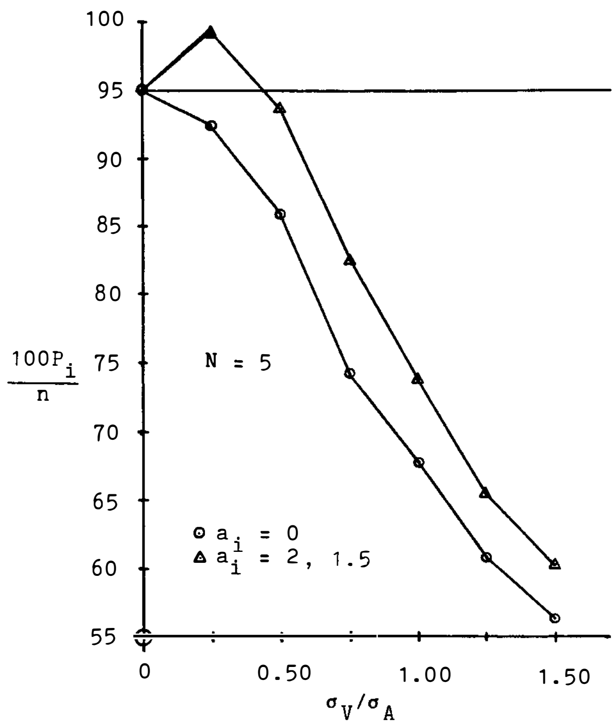

FIGURE 1. **$100P_i/n$ versus σ_V/σ_A for $N = 5$ (Actual percentage above tolerance limit for sample size 5)**

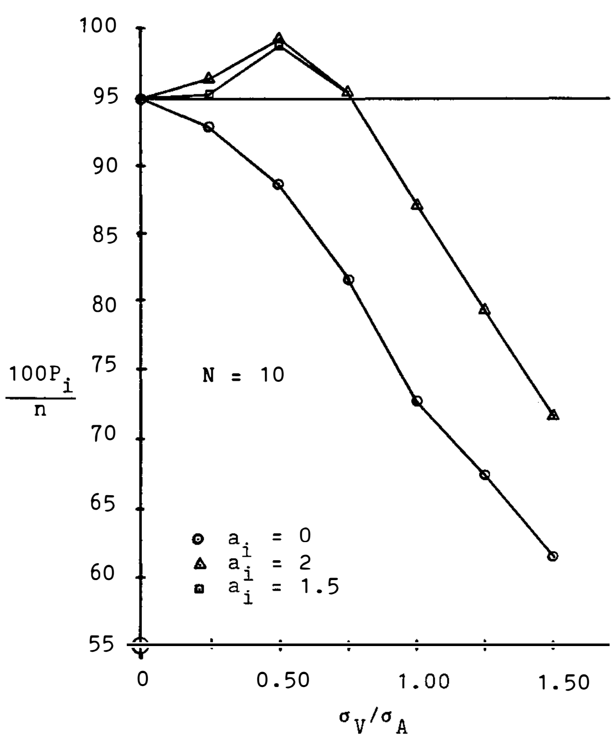

FIGURE 2. **$100P_i$/n versus σ_V/σ_A for $N = 10$ (Actual percentage above tolerance for sample size 10)**

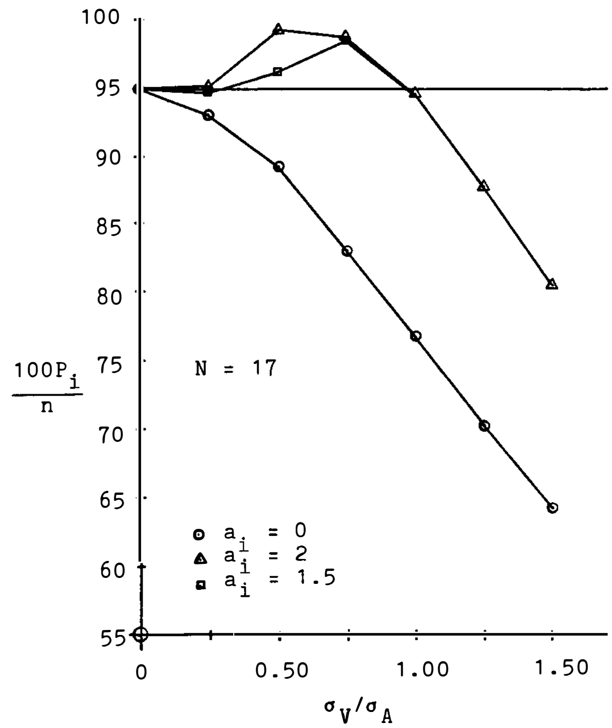

FIGURE 3. **$100P_i$/n versus σ_V/σ_A for $N = 17$ (Actual percentage above tolerance limit for sample size 17)**

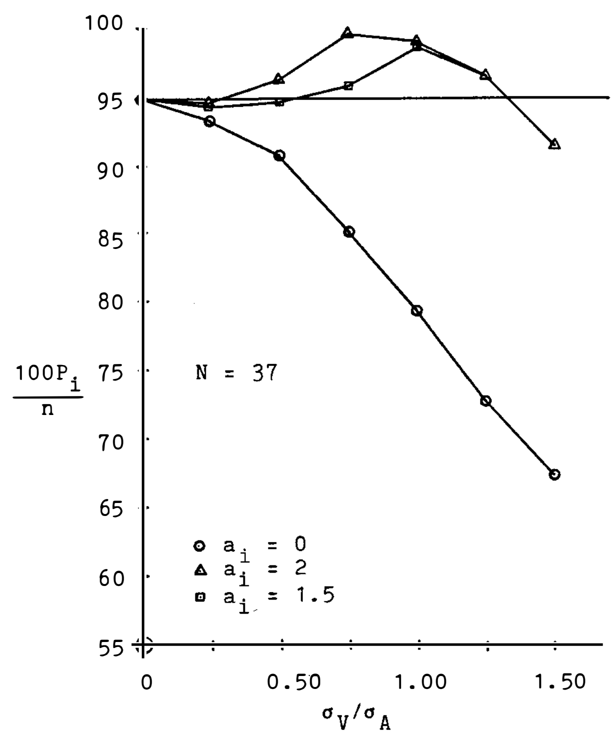

FIGURE 4. **$100P_i$/n versus σ_V/σ_A for $N = 37$ (Actual percentage above tolerance limit for sample size 37)**

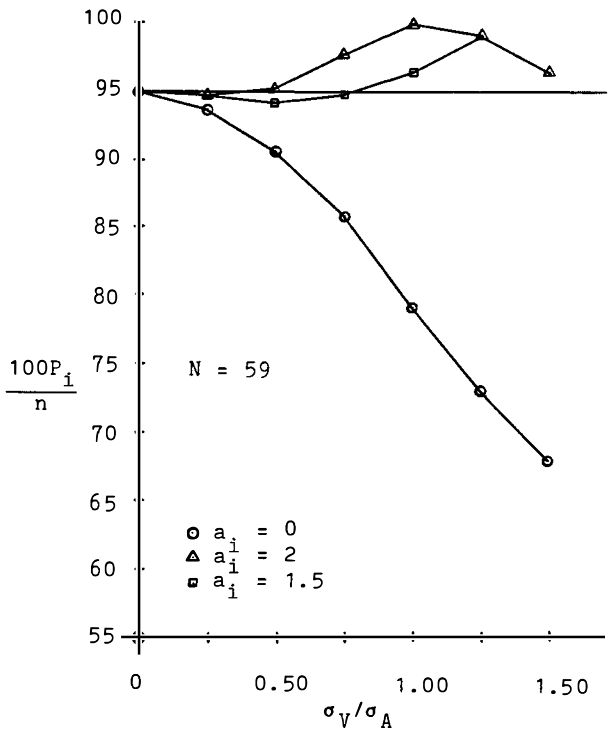

FIGURE 5. $100P_i/n$ versus σ_V/σ_A for $N = 59$ (Actual percentage above tolerance limit for sample size 59)

= 2), or the modified approximation found by equating third moments rather than second moments ($a_i = 1.5$) result in conservative tolerance limits, that is $100\ P_i/n > 95$ percent. Generally speaking, there is less conservatism with $a_i = 1.5$ (equation (12)) than with $a_i = 2$ (equation (11)), but both give essentially identical results for increasing σ_V/σ_A with fixed N.

By noting the locations at which the plots of $100P_i/n$ versus σ_V/σ_A fall below 95 percent for either $a_i = 2$ or $a_i = 1.5$, the following empirical rule is derived:

> Use equation (8) to calculate the degrees of freedom. This will give conservative results when inequality (9) holds. Determine the true coverage, conservative or not, resulting from the use of (8) (or also by using $(N - 1)$ degrees of freedom) from Figures (1)–(5). (An alternative is provided by using equation (12) instead of (8). The resulting coverage is also obtained from Figures (1)–(5). Equation (12) may be used when it gives more appropriate coverage than (8)).

It is noted that in the general application of the rule, σ_V/σ_A is the ratio of the known standard deviation due to measurement error to the *unknown* standard deviation σ_A. In application, σ_A would have to be replaced by $\hat{\sigma}_A$, and the effect of making that replacement has not been studied. Nevertheless, the rule should provide guidance as to when Satterthwaite's formula may be applied.

Additional Comments

In studying linear combinations of mean squares, Gaylor and Hopper (1969) developed a criterion for determining when one could apply Satterthwaite's formula for degrees of freedom. Their criterion for adequacy was a goodness-of-fit criterion in which the distribution of

$$MS = MS_1 - MS_2$$

was required to be approximated satisfactorily by the chi-square distribution. In expressing their criterion in the terminology of this paper, it is required that

$$\sigma_V/\sigma_A < (F - 1)^{-0.5} \tag{15}$$

where F is the 0.975 percentile of the F distribution with degrees of freedom ∞ and $(N - 1)$. A comparison of the criteria given by (9) and (15) is shown in Table 1 for selected values of N. The criterion (9) is less restrictive than the criterion (15) since it addresses the specific problem of constructing 95/95 tolerance limits.

TABLE 1. Comparison of Two Criteria on σ_V/σ_A

N = Sample Size	$-0.4 + 0.5\ell nN$	$(F-1)^{-0.5}$
5	0.40	0.37
10	0.75	0.66
16	0.99	0.85
21	1.12	0.96
25	1.21	1.03
31	1.32	1.13
41	1.46	1.25
61	1.66	1.44

References

GAYLOR, D. W. and HOPPER, F. N. (1969). "Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula". *Technometrics* 11, pp. 691–706.

HAHN, G. J. (1982). "Removing Measurement Error in Assessing Conformance to Specifications". *Journal of Quality Technology* 14, pp. 117–121.

NATRELLA, M. G. (1963). *Experimental Statistics* (NBS Handbook 91). U.S. Department of Commerce, Washington, D.C.

OWEN, D. B. (1963). "Factors for One-Sided Tolerance Limits and for Variables Sampling Plans". *Sandia Corp. Monograph*, SCR-607, Sandia Laboratories, Albuquerque, NM.

SATTERTHWAITE, F. E. (1946). "An Approximate Distribution of Estimates of Variance Components". Biometrics Bulletin 2, pp. 110–14.