United States
Department of
Agriculture

Forest Service

Forest
Products
Laboratory

Research
Paper
FPL 458

An Equation for One-Sided Tolerance Limits for Normal Distributions

Carol L. Link, Mathematical Statistician
Forest Products Laboratory, Madison, WI

Abstract

An equation that does not require tables is given to determine a one-sided tolerance limit for the 100 p^{th} percentile of a normal distribution with confidence $1-\gamma$ for any sample size n. This equation gives accuracy to approximately three or more significant digits when compared to tabled values. Thus it is possible to develop an automated procedure for determining tolerance limits that is not restricted to tabled values.

Keywords: Tolerance limits, normal distribution.

August 1985

Link, Carol L. An equation for one-sided tolerance limits for normal distributions. Res. Pap. FPL 458. Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory; 1985. 4 p.

A limited number of free copies of this publication are available to the public from the Forest Products Laboratory, One Gifford Pinchot Drive, Madison, WI 53705. Laboratory publications are sent to over 1,000 libraries in the United States and elsewhere.

The Laboratory is maintained in cooperation with the University of Wisconsin.

In quality control applications, the practitioner may wish to estimate an extreme (lower) percentile of the distribution from a sample. For example, daily samples are made from a machine's production. If the estimated 100 p^{th} percentile of some characteristic is too low, the machine is adjusted. Since overestimating the 100 p^{th} percentile defers needed machine adjustment, it is often desirable to be conservative and use a lower tolerance limit, t_0, such that the true population 100 p^{th} percentile is above t_0 with probability or confidence $1-\gamma$: $P[P(X \geqq t_0) \geqq 1 - p] \geqq 1 - \gamma$. If the data come from a normal distribution, t_0 has the form $\overline{x}$ - Ks where $\overline{x}$ is the sample mean, and s is the sample standard deviation. Tables of K for various values of p, γ, and n have been developed (e.g. Guttman 1970). However, when using daily samples, sample sizes may vary and it may be desirable to vary the percentile being estimated or the desired level of confidence. If the process of determining tolerance limits is to be automated, a method of determining K needs to be found that does not require the use of tables and is not restricted in sample size and confidence levels to tabled values. Lieberman (1958) gives a formula for K that could be used for sample sizes larger than 50 (the extent of his table). This formula tends to underestimate K, which leads to an overestimation of the tolerance limit, t_0, and is contrary to our goal. This underestimation of K becomes more extreme as the same size decreases, especially if the formula is used for sample sizes less than 50. Lieberman's equation has been used in other popular references (e.g. Natrella 1966). The purpose of this paper is to show how this formula can be improved to give reasonably accurate values of K for any sample size n, percentile 100 p, and confidence $1 - \gamma$.

Given a sample size n from a normal population $N(\mu, \sigma^2)$, in theorem 4.4 Guttman (1970) shows that

$$P\left[P(X \geqq t_0) \geqq 1 - p\right] = P\left[T_{n-1}^*\left(\sqrt{n}\, z_p\right) \leqq \sqrt{n}\, K\right]$$

where $t_0 = \overline{x}$ - Ks, z_p is the (1 - p) 100th percentile of the standard normal distribution and $T_v^*(\delta)$ is the noncentral Student's t distribution with v degrees of freedom and noncentrality parameter δ. Abramowitz and Stegun (1972, equation 26.7.10) show that the noncentral t distribution may be approximated by the standard normal distribution Z.

$$P\left[T_v^*(\delta) \leqq t\right] \approx P\left\{Z \leqq \left[t\left(1 - \frac{1}{4v}\right) - \delta\right] \Big/ \left[1 + \frac{t^2}{2v}\right]^{1/2}\right\} \tag{1}$$

Therefore K should be found such that

$$1 - \gamma = P\left[T_{n-1}^*\left(\sqrt{n} z_p\right) \leqq \sqrt{n} K\right]$$

$$\approx P\left\{Z \leqq \left[\sqrt{n} K\left(1 - \frac{1}{4(n-1)}\right) - \sqrt{n} z_p\right] \Big/ \left[1 + \frac{nK^2}{2(n-1)}\right]^{1/2}\right\}$$

or

$$z_\gamma = \left[\sqrt{n} K\left(1 - \frac{1}{4(n-1)}\right) - \sqrt{n} z_p\right] \Big/ \left[1 + \frac{nK^2}{2(n-1)}\right]^{1/2}$$

since $1 - \gamma = P[Z \leqq z_\gamma]$.

Solving for K yields the following equation:

$$K = \frac{z_p(1-f) + \left\{z_p^2(1-f)^2 - \left[(1-f)^2 - z_\gamma^2/(2(n-1))\right]\left(z_p^2 - z_\gamma^2/n\right)\right\}^{1/2}}{(1-f)^2 - z_\gamma^2/(2(n-1))} \qquad (2)$$

where $f = 1/(4(n-1))$. Lieberman's formula is based on equation (2), but ignores the factor f. As n gets larger this factor is negligible; but for small n, leaving out this factor underestimates K.

In order to determine K without the use of tables, z_P and z_γ must be determined. An approximation for these quantities may be found in Abramowitz and Stegun (1972, equation 26.2.23):

$$z_p \approx t - (c_0 + c_1 t + c_2 t^2)/(1 + d_1 t + d_2 t^2 + d_3 t^3) \qquad (3)$$

where

$t = (\ln(1/p^2))^{1/2}$,
$c_0 = 2.515517$,
$c_1 = 0.802853$,
$c_2 = 0.010328$,
$d_1 = 1.432788$,
$d_2 = 0.189269$,
and $d_3 = 0.001308$.

The error of the approximation of z_p or z_γ is less than 4.5×10^{-4} in absolute value.

A FORTRAN program was written in double precision to compare equation (2) with Lieberman's formula and tabled values (see Guttman 1970, table 4.6). When computing K in Lieberman's formula and equation (2), approximations for z_p and z_γ from equation (3) were used. The chosen values of the parameters were: p = .25, .10, .05, .01, .001; γ = .25, .10, .05, .01; and n = 10 to 200 by 10. Some typical results can be seen in tables 1 and 2. Lieberman's formula will always give a value for K that is less than equation (2) because factor f is missing. The difference in equation (2) and Guttman's tabled values will be due to the error in the two approximations used: the approximation of the noncentral t by the normal (equation (1)) and the approximation of the normal quantiles (equation (3)). Of these, the error in equation (1) has the most effect especially for small sample sizes. Of course a more accurate determination of K could be made using tabled values of z_P and z_γ. This changes the estimates of equation (2) by .001 or less in absolute value. Because of the goal of tolerance limits, a conservative approach would be to overestimate K. For this reason equation (2) is preferable to Lieberman's formula as it either overestimates K (due to the approximation of the noncentral t distribution by the normal) or the underestimation is less severe. As n gets large, either formula will give reasonable estimates of K since the factor $f = 1/(4(n-1))$ goes to zero. If greater accuracy is desired, and the IMSL library of subroutines is available, the noncentral t cumulative distribution function routine MDTN can be used with the root finding routine ZREAL2. The estimate of K from equation (2) can be used as an initial estimate. This procedure gives results that agree with Guttman's tables.

Thus tolerance limits, $t_0 = \bar{x} - Ks$, may be estimated for samples from normal populations without the use of tables. Using equations (2) and (3), K may be determined with approximately three or more significant digits of accuracy. When the estimation of K has fewer than three significant digits of accuracy, the problem can be traced to the poor approximation of a noncentral t distribution for small degrees of freedom by the normal distribution. If greater accuracy is desired, tables or IMSL subroutines are needed.

Table 1.—Values of K for one-sided tolerance limits. $p = .05$, $g = .25$

n	Lieberman	Equation (2)	Guttman
10	2.0322	2.0995	2.104
20	1.9021	1.9300	1.932
30	1.8501	1.8674	1.869
40	1.8203	1.8329	1.834
50	1.8005	1.8103	1.811
60	1.7862	1.7941	1.795
70	1.7751	1.7819	1.782
80	1.7663	1.7721	1.772
90	1.7590	1.7642	1.764
100	1.7529	1.7575	1.758
110	1.7477	1.7519	1.752
120	1.7431	1.7469	1.747
130	1.7392	1.7426	1.743
140	1.7356	1.7388	1.739
150	1.7324	1.7354	1.735
160	1.7296	1.7324	1.732
170	1.7270	1.7296	1.730
180	1.7246	1.7271	1.727
190	1.7224	1.7248	1.725
200	1.7204	1.7226	1.723

Table 2.—Values of K for one-sided tolerance limits for selected values of p, g, and n

p	g	n	Lieberman	Equation (2)	Guttman
.10	.25	10	1.6154	1.6683	1.671
.10	.25	50	1.4174	1.4250	1.425
.10	.25	100	1.3760	1.3796	1.380
.10	.25	200	1.3477	1.3494	1.349
.10	.05	10	2.3215	2.4231	2.355
.10	.05	50	1.6401	1.6497	1.646
.10	.05	100	1.5243	1.5285	1.527
.10	.05	200	1.4485	1.4504	1.450
.05	.25	10	2.0322	2.0995	2.104
.05	.25	50	1.8005	1.8103	1.811
.05	.25	100	1.7529	1.7575	1.758
.05	.25	200	1.7204	1.7226	1.723
.05	.05	10	2.8758	3.0047	2.911
.05	.05	50	2.0590	2.0713	2.065
.05	.05	100	1.9239	1.9293	1.927
.05	.05	200	1.8362	1.8386	1.837
.01	.25	10	2.8235	2.9182	2.927
.01	.25	50	2.5233	2.5371	2.538
.01	.25	100	2.4627	2.4692	2.470
.01	.25	200	2.4215	2.4246	2.425
.01	.05	10	3.9412	4.1224	3.981
.01	.05	50	2.8553	2.8725	2.862
.01	.05	100	2.6808	2.6883	2.684
.01	.05	200	2.5684	2.5719	2.570

Literature Cited

Abramowitz, M.; Stegun, I. A., eds. Handbook of mathematical functions. New York: Dover Publications: 1972: 933,949.

Guttman, I. Statistical tolerance regions. Darien, CT: Hafner Publishing Company; 1970.

IMSL Library Reference Manual. Houston TX: International Mathematical and Statistical Libraries. Inc.; 1979.

Lieberman, G. J. Tables for one-sided statistical tolerance limits. Industrial Quality Control. 14(10): 7-9: 1958.

Natrella, M. G. Experimental statistics. Handbook 91. National Bureau of Standards; 1966: 2-15.

.9-8/85