On exact two-sided statistical tolerance intervals
for normal distributions
with unknown means and unknown common variability

Ivan Janiga and Ivan Garaj

Slovak University of Technology in Bratislava

Slovak Republic

Email: ivan.janiga@stuba.sk

In the paper we deal with a **derivation** as well as a **computation** of exact tolerance factors used in construction of exact two-sided statistical tolerance intervals

Paper Goals

- Definition of $100p\%$ statistical tolerance interval with confidence $1 - \alpha$.

- Approximate computation of tolerance factors (k).

- Derivation of an Exact Equation (5).

- Computation of tolerance factors using Exact Equation (5).

- Simultaneous computation of tolerance factors for m distributions.

- Examples

- Conclusion

Definition of $100p\%$ statistical tolerance interval with confidence $1-\alpha$

Let measurements $x_1, x_2, \cdots, x_n$ be values of a random sample $X_1, X_2, \cdots, X_n$ of size n from a normal distribution with unknown mean μ and unknown variance σ^2 that is

$$X_i \sim N(\mu, \sigma^2), \ i = 1, 2, ..., n; \quad \mu \text{ and } \sigma - \text{unknown.}$$

The $100p\%$ two-sided statistical tolerance interval with confidence level $1-\alpha$ is constructed by

$$(\bar{x} - k\,s, \ \bar{x} + k\,s) \tag{1}$$

for which the following equation is valid

$$P[P(\bar{x} - k\,s < X < \bar{x} + k\,s) \geq p] = 1 - \alpha \tag{2}$$

where

$\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ is sample mean (estimate of the μ), $s = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ is sample standard deviation (estimate of the σ) and $k = k(n, v = n-1, p, 1-\alpha)$ is tolerance factor.

Although the definition of a $100p\%$ statistical tolerance interval with confidence level $1-\alpha$ is simple, the computation of precise values of tolerance factors k from (2) is fairly difficult, particularly without the use a computer.

Approximate computation of tolerance factors (k)

Analytical derivation of the solution of Equation (2) with respect to k is difficult, so approximate methods for the computation of a factor k have been used in the past.

WALD, A., WOLFOWITZ, J. (1946) Tolerance Limits for a Normal Distribution, In *Annals of Mathematical Statistics*, **1946, vol. 17, p. 208-215.**

They proposed an approximate computation of the factor k for the case $v = n - 1$ by using formulas:

$$k = r \sqrt{\frac{v}{\chi_{\alpha}^{2}(v)}} \qquad (3)$$

where

r — the root of the equation $\Phi\left(\frac{1}{\sqrt{n}} + r\right) - \Phi\left(\frac{1}{\sqrt{n}} - r\right) = p$,

$\chi_{\alpha}^{2}(v)$ — α-quantile of the χ^{2}-distribution with $v = n - 1$ degrees of freedom,

Φ — standard normal distribution function.

TAGUTI, G. (1958) Tables of tolerance coefficients for normal populations. In *Reports of Statistical Application Research*, **JUSE, 1958, vol. 5, p. 73-118.**

He was the first to advice some situations when

the degrees of freedom differs from $n - 1$ ($v \neq n-1$).

His tables of factors k are computed by using of the **Wald and Wolfowitz** approximation.

HOWE, W. G. (1969) Two-sided tolerance limits for normal populations. In *Journal of the American Statistical Associantion*, **1969, vol. 64, p. 610-620.**

He proposed more exact approximation in the form

$$
k \approx
\begin{cases}
\sqrt{\dfrac{v\,(2n^2 + 4n + 2 - \chi_\alpha^2(v) + v - 2)}{2n\,(n+1)\,\chi_\alpha^2(v)}}\;\; u_{\frac{1+p}{2}} & \text{for} \quad 1 \le v \le n^2 + \dfrac{n^2}{u_{1-\frac{\alpha}{2}}^2} \\[4ex]
\sqrt{A + \dfrac{n\,A^2}{2v}\,(1 + \dfrac{1}{u_{1-\frac{\alpha}{2}}^2})}\;\; u_{\frac{1+p}{2}} & \text{for} \quad v > n^2 + \dfrac{n^2}{u_{1-\frac{\alpha}{2}}^2}
\end{cases}
\qquad (4)
$$

where $u_{\frac{1+p}{2}}$, $u_{1-\frac{\alpha}{2}}$ are quantiles of a distribution $N(0,1)$ and $A = 1 + \dfrac{u_{1-\frac{\alpha}{2}}^2}{n} + \dfrac{(3 - u_{\frac{1+p}{2}}^2)\,u_{1-\frac{\alpha}{2}}^4}{6\,n^2}$.

This approximation is simpler than the Wald and Wolfovitz´s, the computation may be realized by a programmable calculator and formula is a good approximation of factors k when **degrees of freedom** $v = n-1$ **(the 1ˢᵗ part of formula)** and $v \ne n-1$ **(the 2ⁿᵈ part of formula)**.

Derivation of an Exact Equation (5)

Nowadays the approximation method is out of date in spite of the fact that a lot of books and applications in practice continually use this method.

More recently, the Exact Equation (5) for computation of the tolerance factor k has been derived from Equation (2). The derivation has been done independently by authors:

JÍLEK, M. (1988): *Statistické toleranční meze*. Praha: SNTL, 1988. 275 p.

EBERHARDT, K.R., MEE, R.W., REEVE, C.P. (1989) Computing factors for exact two-sided tolerance limits for a normal distribution. In *Communications in Statistics*, 1989, Part B, vol. 18, p. 397-413.

FUJINO, T. (1989) Exact two-sided tolerance limits for a normal distribution. In *Japanese Journal of Applied Statistics*, 1989, vol. 18, p. 29-36.

JANIGA, I., MIKLÓŠ, R. (2001) Statistical Tolerance Intervals for a Normal Distribution. In *Measurement Science Review*. 2001, vol. 1, no. 1, p. 29-32.

Hence for given n, $v = n-1$, α, p, the tolerance factor k is the root of the **Exact Equation**

$$\sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} F(x,k)\, e^{-\frac{nx^2}{2}}\, \mathrm{d}x - 1 + \alpha = 0 \qquad (5)$$

where

$$F(x,k) = \int_{\frac{vR^2(x)}{k^2}}^{\infty} \frac{t^{\frac{r}{2}-1}\, e^{-\frac{t}{2}}}{2^{\frac{r}{2}}\, \Gamma\left(\frac{n}{2}\right)}\, \mathrm{d}t \,,$$

$R(x)$ is the solution of the equation $\Phi(x+R) - \Phi(x-R) - p = 0$.

Φ is the standard normal distribution function.

Computation of tolerance factors using Exact Equation (5)

was carried out by means of a computer program that uses numerical integration. By means of an iterative process the factors $k = k(n, v = n - 1, p, 1 - \alpha)$ for different n, p and $1 - \alpha$ were **computed** and **published** in the book of extensive tables

> **GARAJ, I., JANIGA I. (2002)** *Two-sided tolerance limits of normal distribution for unknown mean and variability*. **Bratislava: Vyd. STU, 2002, p 147.**

which is cited in Bibliography of the international standard

> **ISO 16269: 2005 Statistical interpretation of data ─ Part 6: Determination of statistical tolerance intervals.**

as well as on the page before last, where is written:

"**Extensive tables of the factor** k for two-sided statistical tolerance interval for the normal distribution with unknown μ and σ **have been published by Garaj and Janiga**[8]. These tables **correspond to Annex E in this part of ISO 16269,** but the **number of entries and the ranges of** n, p **and** α **are larger than in the tables in Annex E. Introduction** to the tables **is given in English, French, German and Slovak**."

Description of the tables given in GARAJ, I., JANIGA I. (2002)

The values of tolerance factors k are **rounded up to four decimal places for combinations of**

$1-\alpha = 0{,}50;\ 0{,}75;\ 0{,}90;\ 0{,}95;\ 0{,}975;\ 0{,}99;\ 0{,}995;\ 0{,}999;$

$p = 0{,}50(0{,}05)0{,}90(0{,}01)0{,}99(0{,}001)0{,}999$

$n = 2(1)200(20)500(50)1000(500)10000(10000)100000;\ \infty.$

In table 9 is $1-\alpha = 0{,}9999$ and $p = 0{,}9991(0{,}0001)0{,}9999.$

In the last line (∞) are the values of $\dfrac{1+p}{2}$ -quantiles of the standard normal distribution.

Remark

As far as we know in the most complex tables published so far

ODEH, R.E., OWEN, D.B. (1980) *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*. New York, Marcel Dekker, 1980.

there are factors k computed **only for three decimal places** and **only for**

$$1 - \alpha = 0,5; \ 0,75; \ 0,90; \ 0,95; \ 0,975; \ 0,99; \ 0,995;$$

$$p = 0,75; \ 0,90; \ 0,95; \ 0,975; \ 0,99; \ 0,995; \ 0,999;$$

$$n = 2(1)100(2)180(5)300(10)400(25)650(50)1000;$$
$$1500; \ 2000; \ 3000; \ 5000; \ 10000; \ \infty.$$

Simultaneous computation of tolerance factors for $m \geq 2$ distributions

Later we found out that the **Exact Equation (5)** can be used for **simultaneous computation** of **more than one sample**. Let us go into details.

Let measurements $(x_{i1}, x_{i2}, \cdots, x_{in})$, $i = 1, 2, ..., m$; $m \geq 2$ be values of m random samples $(X_{i1}, X_{i2}, \cdots, X_{in})$ of size n drown from m normal distributions with unknown means μ_i and unknown common variability σ^2.

Hence
$$X_{ij} \sim N(\mu_i, \sigma^2), \ i = 1, 2, ..., m; \ j = 1, 2, ..., n$$

$$\mu_i - \text{unknown but may differ from each other}$$

$$\sigma - \text{unknown but common standard deviation}$$

Then in the construction of statistical tolerance intervals for m samples the **pooled standard deviation** s_P can be used.

The $100p$ % two-sided statistical tolerance intervals with confidence $1-\alpha$ are intervals

$$(\bar{x}_i - ks_P, \bar{x}_i + ks_P), \ i = 1, 2, ..., m; \ m \geq 2 \tag{6}$$

for which the following equations are valid

$$P[P(\bar{x}_i - ks_P < X_i < \bar{x}_i + ks_P) \geq p] = 1-\alpha, \ i = 1, 2, ..., m \tag{7}$$

where

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij} \qquad\qquad - \textbf{\textit{i}th sample mean} \text{ (estimate of } \mu_i)$$

$$k = k(n, v = m(n-1), p, 1-\alpha) \qquad - \textbf{tolerance factor}$$

$$s_P = \sqrt{\frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2} \qquad - \textbf{pooled standard deviation} \text{ (estimate of } \sigma).$$

The tolerance factors ($k = k(n, v = m(n-1), p, 1-\alpha)$), given in (6), (7), were **also computed from Exact Equation (5)** and **published** in the book

GARAJ, I., JANIGA I. (2004) *Two-sided tolerance limits of normal distributions with unknown means and unknown common variability.* **Bratislava: Vyd. STU, 2004, p 218.**

The tables 1─25, given in **GARAJ, I., JANIGA I. (2004)**, contain the values of tolerance factors k **rounded up to four decimal places** for **all combinations of couples** between

$$1-\alpha = \textbf{0,90; 0,95; 0,99; 0,995; 0,999;}$$

$$p = \textbf{0,90; 0,95; 0,99; 0,995; 0,999.}$$

In each table from 1 to 25 the factors k are computed for

$$n = \textbf{2(1) 40; 45(5) 100; 200(100) 1000; 5000; 10000, } \infty$$

$$v = \textbf{1(1) 60; 65(5) 100; 200(100) 1000; 5000; 10000, } \infty.$$

The values of tolerance factors k are computed for $n = 10^6$ in the last row (∞) and for $v = 10^6$ in the last column (∞).

Remark

As far as we know in the most complex tables published so far

TAGUTI, G. (1958) Tables of tolerance coefficients for normal populations. In *Reports of Statistical Application Research*, JUSE, 1958, vol. 5, p. 73-118.

there are factors k computed for $v \neq n - 1$ and **only for two decimal places**.

In computation the **Wald and Wolfovitz** approximation method is used **for all combinations** of

$$1 - \alpha = 0{,}90;\ 0{,}95;\ 0{,}99$$

$$p = 0{,}90;\ 0{,}95;\ 0{,}99$$

$$n = 2(1)10(2)20(5)30(10)60(20)100;\ 200;\ 500;\ 10000;\ \infty$$

$$v = 1(1)20(2)30(5)100(100)1000;\ \infty.$$

Example 1. The pressure of combustion gases in the engine is normally distributed. Twenty independent measurements of the pressure were made. On the basis of these measurements the sample mean and standard deviation were computed, that is $\bar{x} = 10$ (MPa) and $s = 0{,}5$ (MPa). It is required to compute the 99 % two-sided tolerance interval with the confidence level 90 %.

For $n = 20$, $p = 0{,}99$ and $1 - \alpha = 0{,}90$

Approximate computation by Wald and Wolfovitz (1946) gives

$$k = k(n, v = n - 1, p, 1 - \alpha) = k(20, 19, 0{,}99, 0{,}90) = \mathbf{3{,}3682}$$

$$(\bar{x} - k\,s,\ \bar{x} + k\,s) = (10 - 3{,}3682 \times 0{,}5;\ 10 + 3{,}3682 \times 0{,}5) = \mathbf{(8{,}316;\ 11{,}684)}$$

Computation using Exact Equation (5),

see **TABLE 3b** from **GARAJ, I., JANIGA I. (2002),**

gives

$$k = k(n, v = n - 1, p, 1 - \alpha) = k(20, 19, 0{,}99, 0{,}90) = \mathbf{3{,}3716}$$

$$(\bar{x} - k\,s,\ \bar{x} + k\,s) = (10 - 3{,}3716 \times 0{,}5;\ 10 + 3{,}3716 \times 0{,}5) = \mathbf{(8{,}314;\ 11{,}686)}$$

Example 2.

Suppose the percentage of solids in each of four batches of wet brewer's yeast (A, B, C and D), each from a different supplier, was to be determined. The percentages of the four batches are normally distributed with unknown means μ_i $i = A, B, C, D$ and unknown but common variance σ^2. The researcher wants to determine whether the suppliers differ so that decisions can be made regarding future orders.

For comparing the suppliers there was decided to use

95 % two-sided statistical tolerance intervals with confidence 95 %

The random samples of size $n = 10$ from each batch were collected. From the data the values of sample means and standard deviations were computed:

$$\bar{x}_A = 18,4; \quad s_A = 1,7127; \qquad \bar{x}_B = 14,1; \quad s_B = 2,76687;$$

$$\bar{x}_C = 10,7; \quad s_C = 2,05751; \qquad \bar{x}_D = 10,1; \quad s_D = 2,60128.$$

Case 1 ($m = 1$): We compute the statistical tolerance interval **for each batch particularly**.

For $\quad\quad\quad\quad n = 10$, $v = m(n-1) = 1(10-1) = 9$, $p = 0,95$ and $1 - \alpha = 0,95$

the value of the two-sided statistical tolerance factors for unknown common variability σ^2 can be found in

TABLE 4b of the book GARAJ, I., JANIGA I. (2002)

and equals

$$k = k(n, v = n-1, p, 1-\alpha) = k(10, 9, 0,95, 0,95) = \textbf{3,3935}$$

Then tolerance intervals for batches A, B, C, D are as follows

A: $\bar{x}_A \mp 3,3935 \times s_A = 18,40 \mp 3,3935 \times 1,7127 \Rightarrow$ **(12,59; 24,21)**

B: $\bar{x}_B \mp 3,3935 \times s_B = 14,10 \mp 3,3935 \times 2,76687 \Rightarrow$ **(4,71; 23,49)**

C: $\bar{x}_C \mp 3,3935 \times s_C = 10,70 \mp 3,3935 \times 2,05751 \Rightarrow$ **(3,72; 17,68)**

D: $\bar{x}_D \mp 3,3935 \times s_D = 10,10 \mp 3,3935 \times 2,60128 \Rightarrow$ **(1,27; 18,93)**

Case 2 ($m > 1$): Now we compute the statistical tolerance intervals **simultaneously for all four batches**. In this case we can use the estimate of the pooled standard deviation

$$s_P = \sqrt{\frac{1}{4}(s_A^2 + s_B^2 + s_C^2 + s_D^2)} = \sqrt{\frac{1}{4}(2{,}9333 + 7{,}6556 + 4{,}2333 + 6{,}7667)} = \textbf{2,3232}$$

For $n = 10$, $\nu = m(n-1) = 4(10-1) = 36$, $p = 0{,}95$ and $1 - \alpha = 0{,}95$ the value of the two-sided statistical tolerance factors for unknown common variability σ^2 can be found in

TABLE 7 of the book GARAJ, I., JANIGA I. (2004)

and equals $k = k(n, \nu = m(n-1), p, 1 - \alpha) = k(10, 36, 0{,}95, 0{,}95) = \textbf{2,5964.}$

Then statistical tolerance intervals for batches A, B, C, D are as follows

$$\text{A:} \quad \bar{x}_A \mp 2{,}5964 \times s_P = 18{,}40 \mp 2{,}5964 \times 2{,}3232 \;\Rightarrow\; \textbf{(12,36; 24,43)}$$

$$\text{B:} \quad \bar{x}_B \mp 2{,}5964 \times s_P = 14{,}10 \mp 2{,}5964 \times 2{,}3232 \;\Rightarrow\; \textbf{(8,07; 20,13)}$$

$$\text{C:} \quad \bar{x}_C \mp 2{,}5964 \times s_P = 10{,}70 \mp 2{,}5964 \times 2{,}3232 \;\Rightarrow\; \textbf{(4,67; 16,73)}$$

$$\text{D:} \quad \bar{x}_D \mp 2{,}5964 \times s_P = 10{,}10 \mp 2{,}5964 \times 2{,}3232 \;\Rightarrow\; \textbf{(4,07; 16,13)}$$

Conclusion

When comparing the result of the both cases it can be declared that the statistical tolerance intervals for batches B, C, D are **significantly much smaller** in the **Case 2** than in the **Case 1**. But the statistical tolerance interval for batch A is **significantly a little larger** in the **Case 1**.

We can conclude that the tolerance intervals **computed simultaneously for several populations** can **yield intervals shorter** than the tolerance intervals **computed for each random sample separately**, provided that the underlying normal populations have the same variance. **This nice property** follows from the fact that on the average the **estimate of the variance computed from several random samples** is **"better"** than the estimate computed from one random sample, because this is based on smaller number of observations.