On Two-Sided Statistical Tolerance Intervals for Normal Distributions with Unknown Parameters

I. Janiga, I. Garaj

Slovak University of Technology, Bratislava, Slovak Republic
Email: ivan.janiga@stuba.sk

Abstract. *The equation for computing factors for exact two-sided tolerance limits for a normal distribution with unknown mean and variability was developed in [1] and realised in [2]. Later it was discovered that the equation can be used also in computing factors for more than one normal distribution with unknown means and unknown common variability. The results are presented in [3]. In the paper is given unified approach to two-sided statistical tolerance intervals for normal distributions with unknown parameters.*

Keywords: Statistical Tolerance Intervals, Tolerance Factor, Estimate of Common Variability

1. Introduction

The paper deals with a unified approach to exact two-sided statistical tolerance limits for one normal distribution with unknown parameters (see [2]) and for $m \geq 2$ normal distributions with possible different unknown means and unknown common variability (see [3]).

2. Statistical tolerance intervals for m distributions with common variability

Let measurements $(x_{i1}, x_{i2}, \cdots, x_{in})$ be values of m random samples of the same sizes n drawn from m populations. We assume that measured values x_{ij} are realizations of independent normally distributed random variables X_{ij} with mean value μ_i and common variability σ^2, that is $X_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1,2,...,m$; $j = 1,2,...,n$. The parameters μ_i and σ^2 are supposed to be unknown.

We are looking for two-sided intervals, which with confidence $1-\alpha$ ($0 < \alpha < 1$) cover at least the fraction p ($0 < p < 1$) of values of the distributions $N(\mu_i, \sigma^2)$, $i = 1,2,...,m$. These intervals are named $100\,p$ % statistical tolerance intervals.

If $X_i \sim N(\mu_i, \sigma^2)$, $i = 1,2,...,m$ then two-sided statistical tolerance intervals for distributions $N(\mu_i, \sigma^2)$, $i = 1,2,...,m$ are intervals

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

where

$\bar{x}_i$ value of a sample mean of the *i*th random sample

k tolerance factor

s_P^2 estimate of the common variability (sometimes called “pooled”).

The value $\bar{x}_i$ is an unbiased estimate of unknown parameter μ_i computed by the formula

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij} \tag{2}$$

and the value s_P^2 is an unbiased estimate of the unknown common parameters σ^2

$$s_P^2 = \frac{1}{m(n-1)} \sum_{i=1}^{m} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 \tag{3}$$

The intervals defined in (1) have to fulfil the conditions

$$P\left[P\left(\bar{x}_i - ks_P < X_i < \bar{x}_i + ks_P\right) \geq p\right] = 1 - \alpha\,, \quad i = 1, 2, \ldots, m \tag{4}$$

from which the exact formula for computation of tolerance factors was derived (see [1]).

3. Exact formula for computation of statistical tolerance factors

The value of a tolerance factor is the solution of the following equation (derived in [1])

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

where

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

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

The values of statistical tolerance factors $k = k(n, \nu, p, 1-\alpha)$ are computed from (5) for $\nu = m(n-1)$ and selected $n, p, 1-\alpha$. Let us suppose two cases.

Case 1 ($m = 1$): It means that we consider only one sample. Then we obtain from the equation (5) a value of statistical tolerance factor $k = k(n, \nu, p, 1-\alpha)$, where $\nu = n - 1$.

Case 2 ($m > 1$): We consider m samples and the equation (5) give us the value of $k = k(n, \nu, p, 1-\alpha)$, where $\nu = m(n-1) \neq n - 1$.

4. Comparison of both cases

The comparison was performed by using data. 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 researcher wants to determine whether the suppliers differ so that decisions can be made regarding future orders.

The random samples from each batch were collected and data are presented in Table 1.

Table 1. Percentages of total solids in four batches of brewner´s yeast

n	1	2	3	4	5	6	7	8	9	10
Batch A	20	18	16	21	19	17	20	16	19	18
Batch B	19	14	17	13	10	16	14	12	15	11
Batch C	11	12	14	10	8	10	13	9	12	8
Batch D	10	7	11	9	6	11	8	12	13	14

For comparing the suppliers there was decided to use 95 % two-sided statistical tolerance intervals with confidence equals $1-\alpha = 0{,}95$.

A good fit with normal distributions with unknown parameters μ_i and σ_i^2 $i = \mathrm{A,B,C,D}$ was confirmed by Shapiro-Wilks W statistics. P-values of all batches are greater than or equal to 0,10 so we can conclude that batches come from normal distributions with 90% or higher confidence (see Table 2).

Table 2. P-values of W tests for normality of batches

Batch	A	B	C	D
P value	0,61	0,99	0,69	0,95

On the basis of the data the values of sample means and standard deviations were computed, that are $\bar{x}_\mathrm{A} = 18{,}4$; $s_\mathrm{A} = 1{,}7127$; $\bar{x}_\mathrm{B} = 14{,}1$; $s_\mathrm{B} = 2{,}76687$; $\bar{x}_\mathrm{C} = 10{,}7$; $s_\mathrm{C} = 2{,}05751$; $\bar{x}_\mathrm{D} = 10{,}1$ and $s_\mathrm{D} = 2{,}60128$.

Table 3. Two-sided statistical tolerance factors for unknown common variability σ^2

$1-\alpha = 0{,}95$; $p = 0{,}95$; $\nu = m(n-1)$					
n \ m	1	2	3	4	5
8	3,7456	3,0609	2,8357	2,7201	2,6488
9	3,5459	2,9541	2,7548	2,6515	2,5873
10	**3,3935**	2,8700	2,6904	**2,5964**	2,5377

Case 1 ($m = 1$): It is required to compute the 95 % two-sided tolerance interval with the confidence level 95 %. The value of $k = k(n, \nu, p, 1-\alpha) = k(10,\ 9,\ 0{,}95,\ 0{,}95) = 3{,}3935$ can be found in Table 3. Then statistical tolerance intervals for batches A, B, C, D are as follows

$$\begin{aligned}
&\mathrm{A:}\ 18{,}40 \mp 3{,}3935 \times 1{,}7127 \Rightarrow (12{,}59;\ 24{,}21)\\
&\mathrm{B:}\ 14{,}10 \mp 3{,}3935 \times 2{,}76687 \Rightarrow (4{,}71;\ 23{,}49)\\
&\mathrm{C:}\ 10{,}70 \mp 3{,}3935 \times 2{,}05751 \Rightarrow (3{,}72;\ 17{,}68)\\
&\mathrm{D:}\ 10{,}10 \mp 3{,}3935 \times 2{,}60128 \Rightarrow (1{,}27;\ 18{,}93)
\end{aligned}$$

Case 2 ($m > 1$):It is also required to compute the 95 % two-sided tolerance interval with the confidence level 95 %. First of all it is needed to check variances of the batches A, B, C, D. In testing the null hypothesis H_0: $\sigma_\mathrm{A}^2 = \sigma_\mathrm{B}^2 = \sigma_\mathrm{C}^2 = \sigma_\mathrm{D}^2$ were used Cochran's C test (P-Value = 0,605823), Bartlett's test (P-Value = 0,496912) and Levene's test (P-Value = 0,559791). Since the smallest of the P-values is greater than or equal to 0,10 it can be concluded that batches have the common variability σ^2. Then estimates of the unknown common variability and standard deviation computed from data are $\sigma_\mathrm{est}^2 = s_P^2 = 4{,}6463846$ (see (3)) and $\sigma_\mathrm{est} = s_P = 2{,}3231923$.

Now the value of $k = k(n, \nu, p, 1-\alpha) = k(10,\ 36,\ 0{,}95,\ 0{,}95) = 2{,}5964$ can be found in Table 3. Then statistical tolerance intervals for batches A, B, C, D are as follows

$$\begin{aligned}
&\mathrm{A:}\ 18{,}40 \mp 2{,}5964 \times 2{,}3231923 \Rightarrow (12{,}36;\ 24{,}43)\\
&\mathrm{B:}\ 14{,}10 \mp 2{,}5964 \times 2{,}3231923 \Rightarrow (8{,}07;\ 20{,}13)
\end{aligned}$$

$$\text{C:}\ \ 10{,}70 \mp 2{,}5964 \times 2{,}3231923 \ \Rightarrow\ (4{,}67; 16{,}73)$$
$$\text{D:}\ \ 10{,}10 \mp 2{,}5964 \times 2{,}3231923 \ \Rightarrow\ (4{,}07; 16{,}13)$$

5. Discussion and Conclusions

The unified approach to two-sided statistical tolerance intervals for normal distribution with unknown parameters was developed in chapters 2 and 3. The equation (5) was used for computation of the values of statistical tolerance factors $k = k(n, v, p, 1-\alpha)$ for the selected $n = 2$ (1) 10; $m = 1$ (1) 4; $v = m(n-1)$; $p = 0{,}95$ and $1-\alpha = 0{,}95$ (see Table 3).

In the first case ($m = 1,\ n = 10$) the 95 % two-sided tolerance intervals with the confidence level 95 % were computed for all batches. In the computation there were used the value of tolerance factor 3,3935 given in Table 3 and the values of standard deviation $s_A = 1{,}7127$; $s_B = 2{,}76687$; $s_C = 2{,}05751$; $s_D = 2{,}60128$.

Then in the second case ($m = 4,\ n = 10$) the 95 % two-sided tolerance intervals with the confidence level 95 % were also estimated for all batches. But in this case the value of tolerance factors 2,5964 from Table 3 and the estimate of the unknown common standard deviation $s_P = 2{,}3231923$ were used.

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 second case than in the first one. But the statistical tolerance interval for batch A is significantly a little larger in the first case.

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.

Acknowledgements

This contribution was granted by the scientific projects VEGA 1/0437/08 and VEGA 1/4091/07 as well as VEGA 1/0374/08.

References

[1] Janiga I., Miklóš R. Statistical tolerance intervals for a normal distribution. n: Measurement Science Review. ISSN 1335-8871. - Vol. 1, Sec. 1, No. 1 (2001), s. 29-32.

[2] Garaj I., Janiga, I. Dvojstranné tolerančné medze pre neznámu strednú hodnotu a rozptyl normálneho rozdelenia. Vydavateľstvo STU, Bratislava, 2002, ISBN 80-227-1779-7.

[3] Garaj I., Janiga, I. Two sided tolerance limits of normal distributions with unknown means and unknown common variability. Vydavateľstvo STU, Bratislava, 2004, ISBN 80-227-2019-4.

[4] Terek M., Hrnčiarová Ľ. Štatistické riadenie kvality. IURA EDITION, Bratislava, 2004, ISBN 80-89047-97-1.