Statistical tolerance intervals for a normal distribution

Ivan Janiga - Roman Miklóš

Dept. of Mathematics, Mechanical Engineering Faculty, Slovak University of Technology
janiga@sjf.stuba.sk - rmiklos@cvt.stuba.sk

Abstract

In this paper the theory underlying statistical tolerance intervals for a normal distribution with unknown standard deviation is given

1. Introduction

Tolerance intervals have important role in application of statistical methods in technical practice, especially in statistical quality control (more about SQC see in [2], [3], [4], [5], [6], [7], [8], [10]). As far as we know from the literature, tolerance factors for a normal distribution (normality test see in [9]) have been computed by using approximate methods. These values of factors do not fulfil current requirements as for accuracy (see [1]), which is why the numerical methods should be used.

2. One-sided statistical tolerance limit

We consider the case of an upper tolerance limit, $T_U = \bar{x} + ks$, where $\bar{x}$ and s are the sample mean and sample standard deviation respectively for a random sample of size *n* from a normal distribution. The problem is to determine *k* such that one may have $100(1-\alpha)\,\%$ confidence that the area under the normal distribution between $\bar{x} = -\infty$ and $\bar{x} = T_U$ is at least $1-\beta$.

This can be alternatively stated as solving for *k* the equation

$$P\left(\int_{-\infty}^{\bar{x}+ks} f(y)\,\mathrm{d}y \geq 1-\beta\right) = 1-\alpha \tag{1}$$

where $f(y)$ is the density function of a normal distribution with mean μ and standard deviation σ. Standardizing the limits of the integral, equation (1) can be rewritten

$$P\left(\int_{-\infty}^{(\bar{x}+ks-\mu)/\sigma} \varphi(y)\,\mathrm{d}y \geq 1-\beta\right) = 1-\alpha \tag{2}$$

where $\varphi(y)$ is the standard normal density function. Inverting the probability on the left-hand side of (2), we have

$$P\big((\bar{x}+ks-\mu)/\sigma \geq u_{1-\beta}\big) = 1-\alpha \tag{3}$$

where $u_{1-\beta} = \Phi^{-1}(1-\beta)$, i.e. $u_{1-\beta}$ is the quantile of the standard normal distribution. Note that $\big((\bar{x}+ks-\mu)/\sigma \geq u_{1-\beta}\big)$ simply defines a region in the $\bar{x}, s$ plane bounded by a straight line. For any value of *k* we could find the value of the left hand side of (3) by numerical integration of the joint probability density of $\bar{x}$ and s, and then iteratively adjust the value of *k* and reintegrate until the integral equalled $1-\alpha$ within a specified accuracy. However, a more straightforward method is available.

Expression (3) may be written

$$P\left((\bar{x}+ks-\mu)/\sigma \geq \Phi^{-1}(1-\beta)\right)=P\left((\bar{x}+ks-\mu)/\sigma \geq u_{1-\beta}\right)$$

$$=P\left(\frac{\left(\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)}{s/\sigma}+\sqrt{n}\,k \geq \sqrt{n}\,\frac{u_{1-\beta}}{s/\sigma}\right) \tag{4}$$

$$=P\left(\frac{U+\delta}{s/\sigma} \geq -\sqrt{n}\,k\right)=1-\alpha$$

where U is a standard normal variable and $\delta=-\sqrt{n}\,u_{1-\beta}$ is a scalar constant. The quotient $\frac{U+\delta}{s/\sigma}$ has a non-central *t*-distribution with $n-1$ degrees of freedom and non-centrality parameter δ. Hence (4) can be solved directly from the fact that $-\sqrt{n}\,k$ must be the α-quantile of this distribution, i.e. $k=-\left(1/\sqrt{n}\right)t_{\alpha}(n-1,-\sqrt{n}\,u_{1-\beta})$. It follows from the properties of the non-central *t*-distribution that the formula for *k* can be rewritten as

$$k=\left(1/\sqrt{n}\right)t_{1-\alpha}(n-1,\sqrt{n}\,u_{1-\beta}) \tag{5}$$

The reflected problem is to determine *k* such that one may have $100(1-\alpha)$ % confidence that the area under the normal distribution between $T_L=\bar{x}-ks$ and $\bar{x}=\infty$ is at least $1-\beta$. The required value of *k* is the same, i.e. $k=\left(1/\sqrt{n}\right)t_{1-\alpha}(n-1,\sqrt{n}\,u_{1-\beta})$.

3. Two-sided statistical tolerance limits

We now consider the case of a pair of tolerance limits, i.e. a simultaneous upper tolerance limit $T_U=\bar{x}+ks$ and lower specification limit $T_L=\bar{x}-ks$. The problem is to determine *k* such that one may have $100(1-\alpha)$% confidence that the area under the normal distribution between $x=T_L$ and $x=T_U$ is at least $1-\beta$.

This can be alternatively stated as solving for *k* the equation

$$P\left(\int_{\bar{x}-ks}^{\bar{x}+ks} f(y)\,\mathrm{d}y \geq 1-\beta\right)=1-\alpha \tag{6}$$

where *f*(.) is the density function of a normal distribution with mean μ and standard deviation σ. Standardizing the limits of the integral, equation (6) can be rewritten

$$P\left(\int_{(\bar{x}-ks-\mu)/\sigma}^{(\bar{x}+ks-\mu)/\sigma} \varphi(y)\,\mathrm{d}y \geq 1-\beta\right)=1-\alpha \tag{7}$$

where $\varphi(y)$ is the standard normal density function. Without loss of generality, we can set $\mu = 0$ and $\sigma = 1$, so that the problem becomes one of determining k such that

$$P_{\overline{X},S^2}\left(\Phi(\overline{x}+ks)-\Phi(\overline{x}-ks)\geq 1-\beta\right)=1-\alpha \tag{8}$$

We introduce the auxiliary function

$$\zeta(\overline{x},s,k)=\Phi(\overline{x}+ks)-\Phi(\overline{x}-ks) \tag{9}$$

and denote

$$C(k)=P_{\overline{X},S^2}(\zeta(\overline{x},s,k)\geq 1-\beta)$$

$$F(x,k)=P_{S^2}(\zeta(\overline{x},s,k)\geq 1-\beta \mid \overline{x}=x)\,. \tag{10}$$

Then obviously

$$C(k)=\int_{-\infty}^{\infty}F(x,k)\,f_{\overline{X}}(x)dx \tag{11}$$

Further, because of $\overline{X}$ is distributed as $N(\mu,\sigma^2/n)$ then $f_{\overline{X}}(x)=\sqrt{\frac{n}{2\pi}}e^{-nx^2/2}$. Furthermore because of the function $\zeta(\overline{x},s,k)$ is increasing (strictly) with respect to variable s, then the equation $\zeta(\overline{x},s,k)=1-\beta$ has for given values of $\overline{x},k,\beta$ the unique solution s_0. Let us denote $R = ks_0$, then according to formula (9) the next is valid:

$$\Phi(\overline{x}+R)-\Phi(\overline{x}-R)=1-\beta\,. \tag{12}$$

From the increasing property of $\zeta(\overline{x},s,k)$ it follows, that the condition of $\zeta(\overline{x},s,k)\geq 1-\beta$ is equivalent to the condition $s > s_0$, i.e. Then from independence of $\overline{X}$ and S and according to (10) it follows that:

$$F(x,k)=P_{S^2}(s>R/k)\,.$$

Applying the fact, that vS^2 is distributed as $\chi^2(v)$ and from the property, that $s > R/k$ is equivalent to $vs^2 > v\,R^2/k^2$ it follows:

$$F(x,k)=P_{\chi^2(v)}(\chi^2(v)>vR^2/k^2)=$$

$$=1-F_{\chi^2}(vR^2/k^2,v)=$$

$$=\int_{\frac{vR^2}{k^2}}^{\infty}\frac{t^{v/2-1}e^{-t/2}}{2^{v/2}\Gamma(v/2)}dt \tag{13}$$

Finally according to formulae (11), (12), (13), the numerical process for computation of two sided tolerance limit k is given as follows:

For given values of n, ν, α, β compute the value of k to be the solution of equation

$$\sqrt{\frac{n}{2\pi}}\int_{-\infty}^{\infty} F(x,k)e^{-\frac{nx^2}{2}}dx - 1 + \alpha = 0 \qquad (14)$$

where $F(x,k)$ is the upper tail area of the $\chi^2(\nu)$ distribution beyond the point $\frac{\nu R^2}{k^2}$, i.e.

$$F(x,k) = \int_{\frac{\nu R^2}{k^2}}^{\infty} \frac{t^{\nu/2-1}e^{-t/2}}{2^{\nu/2}\Gamma(\nu/2)}dt\,. \qquad (15)$$

The quantity R is the solution of equation

$$\Phi(x+R) - \Phi(x-R) - 1 + \beta = 0 \qquad (16)$$

where $\Phi(.)$ is the standard normal distribution function. The above computation of k using formulae (14), (15), (16) can be performed using the suitable numerical methods for computing the zero of a function and for quadrature.

Acknowledgement

This contribution was supported by the Grant No. 1/8094/01 as well as the Grant No. 1/7077/20 both from VEGA

Bibliography

[1] Janiga I., Miklóš R.: O jednom probléme výpočtu tolerančných činiteľov, 9. Medzinárodný seminár Výpočtová štatistika, SŠDS, Bratislava, 2000, str. 30 - 32

[2] Palenčár R.: Metódy a prostriedky kontroly akosti, STU Bratislava, 1999

[3] Palenčár R., Halaj M.: Metrologické zabezpečenie systémov riadenia kvality, STU Bratislava, 1998

[4] Terek M., Hrnčiarová Ľ.: Possibilities of Economic Design of 3σ - Control Charts. Scientific Papers of the University of Pardubice 1998

[5] Terek M., Peller F.: Some Possibilities of the Analysis of the Costs of SPC by Control Charts. Scientific Papers of the University of Pardubice 1999

[6] Terek M., Hrnčiarová Ľ.: Štatistické riadenie kvality, Ekonom Bratislava 1999

[7] Hrnčiarová Ľ., Terek M.: Použitie nástrojov na zlepšenie kvality. Kvalita č.1/1998

[8] Hrnčiarová, Ľ., Terek, M.: Štatistické riadenie kvality: Indexy spôsobilosti a ich testovanie. Kvalita, Inovácia, Prosperita č. 1-2, 1999

[9] Garaj I.: W-kritérium a jeho spracovanie na počítači, 3, seminár Výpočtová štatistika, SŠDS, Bratislava, 1990, str. 8-9

[10] Rublík F., Bognárová M.: Tables for a statistical control test, Applications of mathematics 37, No. 6, 1992