Chapter 830

Tolerance Intervals for Normal Data

Introduction

This routine calculates the sample size needed to obtain a specified coverage of a β-content tolerance interval at a stated confidence level for data from the normal distribution and data without a specified distribution. These intervals are constructed so that they contain at least 100β% of the population with probability of at least 100(1 - α)%. For example, in water management, a drinking water standard might be that one is 95% confident that certain chemical concentrations are not exceeded more than 3% of the time.

Difference between a Confidence Interval and a Tolerance Interval

It is easy to get confused about the difference between a *confidence interval* and a *tolerance interval*. Just remember than a *confidence interval* is usually a probability statement about the value of a distributional parameter such as the mean or proportion. On the other hand, a *tolerance interval* is a probability statement about a proportion of the distribution from which the sample is drawn.

Technical Details

This section is primarily based on results in Guenther (1977). Other useful references are Guenther (1972), Hahn and Meeker (1991), Howe (1969), Krishnamoorthy and Mathew (2009), Odeh, Chou, and Owen (1987), and Young, Gordon, Zhu, and Olin (2016).

A tolerance interval is constructed from a random sample so that a specified proportion of the population is contained within the interval. The interval is defined by two limits, L_1 and L_2, which are constructed using

$$L_1 = \bar{x} - ks, L_2 = \bar{x} + ks$$

where $\bar{x}$ is the sample mean, s is the sample standard deviation, and k is calculated as described below.

Proportion of the Population Covered

An important concept is that of *coverage*. Coverage is the proportion of the population distribution that is between the two limits. For the normal distribution, these population limits are defined by

$$\mu \pm z_{\alpha/2}\sigma$$

The coverage is the area under the standard normal distribution between these limits.

Solving for N

The tolerance factor k is constructed so that

$$\Pr(L_2 - L_1 \geq P) = 1 - \alpha$$

Guenther (1977) provides the following approximation for k in the **two-sided case**

$$k = uvw$$

where

$$u = z_{(1+P)/2}\sqrt{\frac{N+1}{N}}$$

$$v = \sqrt{\frac{N-1}{\chi^2_{N-1,\alpha}}}$$

and

$$w = \sqrt{1 + \frac{N - 3 - \chi^2_{N-1,\alpha}}{2(N+1)^2}}$$

Using this value of k, the minimum value of N is found by increasing N by one at each iteration and stopping when

$$\frac{\chi^2_{N-1,\alpha\prime}}{\chi^2_{N-1,\alpha}} \leq \left[\frac{z_{(1+P\prime)/2}}{z_{(1+P)/2}}\right]^2$$

Guenther (1977) also provides an algorithm for finding N in the **one-sided case**.

$$k_1 = \frac{T'_{N-1,1-\alpha,z_P\sqrt{N}}}{\sqrt{N}}$$

$$k_2 = \frac{T'_{N-1,\alpha\prime,z_{P\prime}\sqrt{N}}}{\sqrt{N}}$$

Where T' is the noncentral T distribution.

Starting at $N = 2$, increase N by one until $k_1 \leq k_2$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Design tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Design Tab

The Design tab contains most of the parameters and options that you will be concerned with.

Solve For

Solve For

Select the parameter you want to solve for. The parameter you select here will be shown on the vertical axis in the plots.

Sample Size N

A search is conducted for the minimum sample size that adheres to the parameter values. Remember that the search is based on two inequalities, so the particular values of the confidence level and alpha' may not be met exactly.

Exceedance Margin δ

A search is conducted for the value of δ that meets the requirements of the confidence level and alpha' inequalities.

$α' = \Pr(p \geq P + \delta)$

α' is the probability that the sample coverage is $\geq P + \delta$. A search is conducted for the value of α' that meets the requirements of the other settings.

One-Sided Limit or Two-Sided Interval

Interval Type

Specify whether the tolerance interval is two-sided or one-sided. A one-sided interval is usually called a **tolerance bound** rather than a tolerance interval because it only has one limit.

Two-Sided Tolerance Interval

A two-sided tolerance interval, defined by two limits, will be used.

Upper Tolerance Limit

An upper tolerance bound will be used.

Lower Tolerance Limit

A lower tolerance bound will be used.

Sample Size

N (Sample Size)

Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

You can enter a single value or a range of values.

Proportion of Population Covered

Proportion Covered (P)

Enter the proportion of the population that is covered by the tolerance interval. This the desired coverage proportion between the tolerance limits. It is the probability (area) between the tolerance limits based on the normal distribution. This is the proportion of the population that lies between the two limits.

If a two-sided interval (L1, L2) is specified, this is the desired area of the probability distribution between L1 and L2. If F(x) is the CDF of the distribution, P = F(L2) - F(L1).

If a lower limit (L1) is specified, this is the desired area of the probability distribution above L1. If F(x) is the CDF of the distribution, P = 1 - F(L1).

If an upper limit (L2) is specified, this is the desired area of the probability distribution below L2. If F(x) is the CDF of the distribution, P = F(L2).

The possible range is $0 < P < 1 - \delta$. Usually, one of the values 0.80, 0.90, 0.95, or 0.99 is used.

You can enter a single value such as *0.9*, a series of values such as *0.8 0.9 0.99*, or a range such as *0.8 to 0.98 by 0.02*.

Confidence Level (1 - α)

Specify the proportion of tolerance intervals (constructed with these parameter settings) that would have the same coverage. The absolute range is $0 < 1 - \alpha < 1$. The typical range is $0.8 < 1 - \alpha < 1$.

Usually, one of the values 0.90, 0.95, and 0.99 is used.

You can enter a single value such as *0.95*, a series of values such as *0.9 0.95 0.99*, or a range such as *0.8 to 0.95 by 0.01*.

Proportion Covered Exceedance

Proportion Covered Exceedance Margin (δ)

δ is added to P to set an upper bound of $P' = P + \delta$ on the coverage. Hence, δ represents a precision value for P. The value of P' is set to occur with a low probability (0.05 or 0.01).

For example, if $P = 0.9$ and $\delta = 0.01$, then the parameters are set so that a coverage of 0.9 occurs with high probability, but a coverage of 0.91 occurs with low probability.

The possible range is $0 < \delta < 1 - P$. Usually 0.1, 0.05, or 0.01 is used.

You can enter a single value such as *0.02* or a series of values such as *0.01 0.02 0.05* or a range such as *0.01 to 0.05 by 0.01*

α' = Pr(p ≥ P + δ)

α' is the probability that the sample coverage p is greater than $P + \delta$. It is set to a small value such as 0.05 or 0.01.

The range is $0.0 < \alpha' < 0.5$. Usually 0.1, 0.05, or 0.01 is used.

You can enter a single value such as *0.05*, a series of values such as *0.01 0.02 0.05*, or a range of values such as

0.01 to 0.05 by 0.01.

Example 1 – Calculating Sample Size

Suppose a study is planned to determine the sample size required to compute a two-sided 95% normal tolerance interval the covers 90% of the population. The researcher wants to investigate using a δ of 0.01, 0.025, or 0.05 with an α' of 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First load the **Tolerance Intervals for Normal Data** procedure window. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**Sample Size**
Interval Type	**Two-Sided Tolerance Interval**
Proportion Covered (P)...........................	**0.9**
Confidence Level (1 – α)	**0.95**
Coverage Proportion Exceedance (δ)	**0.01 0.025 0.05**
α' = Pr(p ≥ P + δ)	**0.05**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Tolerance Interval of Normal Data

Confidence Level 1 - α	Sample Size N	Proportion Covered P	Proportion Covered Exceedance Margin δ	P + δ	Pr(p ≥ P + δ) α'	Tolerance Factor k
0.950	5910	0.9000	0.0100	0.9100	0.050	1.6703
0.950	866	0.9000	0.0250	0.9250	0.050	1.7138
0.950	179	0.9000	0.0500	0.9500	0.050	1.8084

References

Faulkenberry, G.D. and Weeks, D.L. 1968. 'Sample Size Determination for Tolerance Limits.' Technometrics, Vol. 10, No. 2, Pages 343-348.

Guenther, William C. 1972. 'Tolerance Intervals for Univariate Distributions.' Naval Research Logistics Quarterly, Vol. 19, No. 2, Pages 309-333.

Guenther, William C. 1977. Sampling Inspection in Statistical Quality Control. Griffin's Statistical Monographs, Number 37. London.

Hahn, G. J. and Meeker, W.Q. 1991. Statistical Intervals. John Wiley & Sons. New York.

Howe, W.G. 1969. 'Two-Sided Tolerance Limits for Normal Populations-Some Improvements.' Journal of the American Statistical Association, Vol 64, No. 326, Page 610-620.

Krishnamoorthy, K. and Mathew, T. 2009. Statistical Tolerance Regions. John Wiley, New York.

Odeh, R.E., Chou, Y.M., and Owen, D.B. 1987. 'The Precision for Coverages and Sample Size Requirements for Normal Tolerance Intervals.' Communications in Statistics – Simulation and Computation, 16(4), Pages 969-985.

Young, D.S., Gordon, C.M, Zhu, S., and Olin, B.D. 2016. 'Sample size determination strategies for normal tolerance intervals using historical data.' Quality Engineering, Volume 28(3), Pages 337-351.

Report Definitions

Confidence Level (1 - α) is the proportion of studies with the same settings that produce tolerance intervals with
a proportion covered of at least P.

N is the number of subjects.

Proportion Covered P is the proportion of the population covered. It is the probability between the tolerance
interval limits based on the normal distribution.

Proportion Covered Exceedance Margin δ is the value that is added to P to set an upper bound on the coverage at P +
δ.

P + δ is the upper limit of the proportion covered P. It is a measure of the precision (closeness) of the actual
coverage to P.

α' = Pr(p ≥ P + δ) is the probability that the coverage computed from a random sample (p) is greater than P + δ. It
is set to a small value such as 0.05 or 0.01.

k is the tolerance factor used in calculating the tolerance interval from a sample. The sample tolerance interval
is Mean ± k (SD).

Summary Statements

A two-sided normal tolerance interval computed from a sample of 5910 observations has a target
coverage of 0.900 at a 0.950 confidence level. The probability that the coverage exceeds the
target value by an amount greater than or equal to 0.010 is 0.050. The factor k for computing
the tolerance limit(s) is 1.6703.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	5910	7388	1478
20%	866	1083	217
20%	179	224	45

Definitions

Dropout Rate (DR) is the percentage of subjects (or items) that are expected to be lost at random during the
course of the study and for whom no response data will be collected (i.e. will be treated as "missing").

N is the evaluable sample size at which the tolerance interval is computed. If N subjects are evaluated out of
the N' subjects that are enrolled in the study, the design will achieve the stated tolerance interval.

N' is the total number of subjects that should be enrolled in the study in order to end up with N evaluable
subjects, based on the assumed dropout rate. After solving for N, N' is calculated by inflating N using the
formula N' = N / (1 - DR), with N' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C.,
Shao, J., and Wang, H. (2008) pages 39-40.)

D is the expected number of dropouts. D = N' - N.

This report shows the calculated sample size for each of the scenarios.

Plots Section
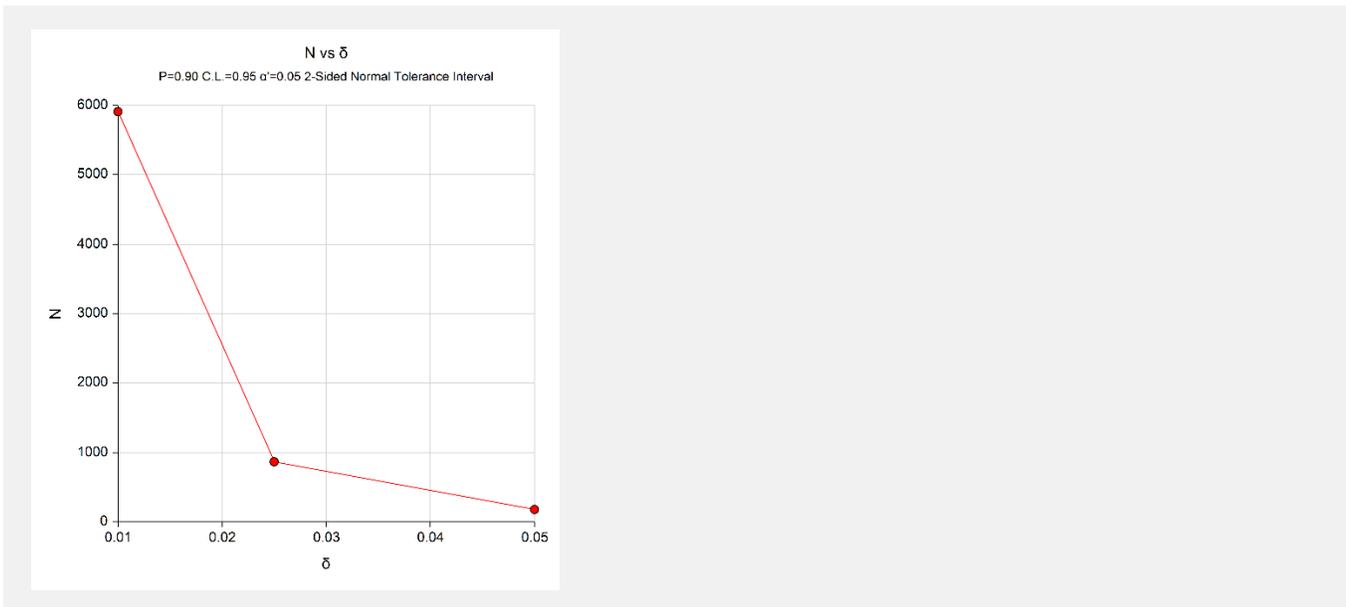
This plot shows the sample size versus the three values of δ.

Example 2 – Calculating α'

Continuing Example 1, the researchers wants to show the impact of various sample sizes on α'. They decide to determine the value of α' for various value of *N* between 200 and 1600, keeping the other values the same except that they set δ to 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First load the **Tolerance Intervals for Normal Data** procedure window. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

Option	Value
Design Tab	
Solve For	**α' = Pr(p ≥ P + δ)**
Interval Type	**Two-Sided Tolerance Interval**
N (Sample Size).....................................	**200 to 1600 by 200**
Proportion Covered (P)...........................	**0.9**
Confidence Level (1 – α)	**0.95**
Coverage Proportion Exceedance (δ)	**0.025**

Annotated Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Tolerance Interval of Normal Data

Confidence Level 1 - α	Sample Size N	Proportion Covered P	Proportion Covered Exceedance Margin δ	P + δ	Pr(p ≥ P + δ) α'	Tolerance Factor k
0.950	200	0.9000	0.0250	0.9250	0.545	1.7984
0.950	400	0.9000	0.0250	0.9250	0.287	1.7493
0.950	600	0.9000	0.0250	0.9250	0.140	1.7287
0.950	800	0.9000	0.0250	0.9250	0.065	1.7168
0.950	1000	0.9000	0.0250	0.9250	0.029	1.7088
0.950	1200	0.9000	0.0250	0.9250	0.012	1.7029
0.950	1400	0.9000	0.0250	0.9250	0.005	1.6984
0.950	1600	0.9000	0.0250	0.9250	0.002	1.6948

This report shows the impact on α' of various sample sizes. Since the value of the *Tolerance Factor k* is not related to α' or δ, this report allows you to calculate *k* for use with sample data.
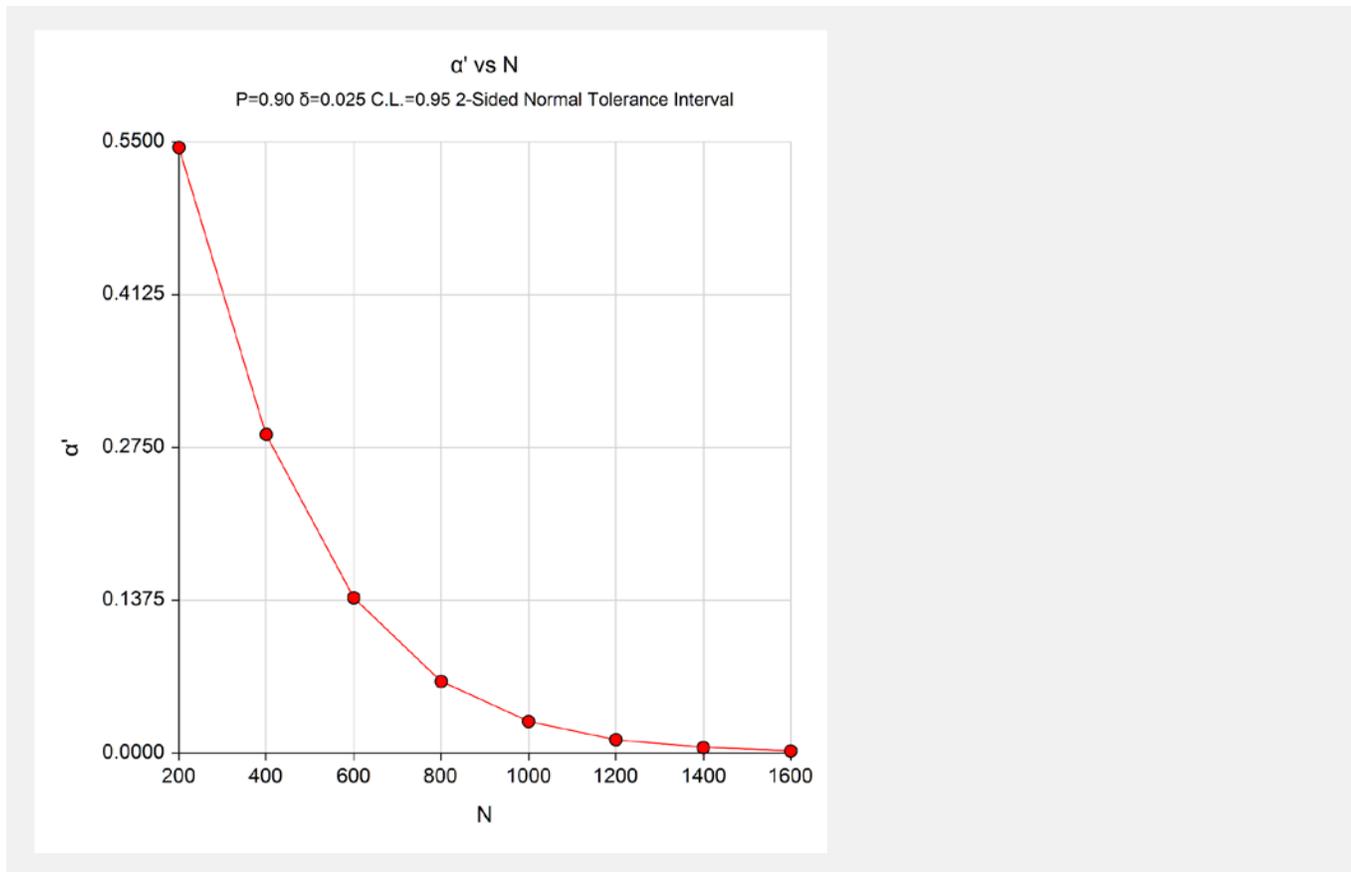
Plots Section

This plot shows the sample size versus α'.

Example 3 – Validation using Guenther (1977)

Guenther (1977) page 174 gives an example in which $P = 0.8$, $1 - \alpha = 0.9$, $P + \delta = 0.95$, and $\alpha' = 0.05$. He obtains a sample size of 26 and a k of 1.613.

Setup

This section presents the values of each of the parameters needed to run this example. First load the **Tolerance Intervals for Normal Data** procedure window. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

Option   **Value**

Design Tab
Solve For .... **Sample Size**
Interval Type .... **Two-Sided Tolerance Interval**
Proportion Covered (P) **0.8**
Confidence Level (1 – α) **0.9**
Coverage Proportion Exceedance (δ) **0.15**
α' = Pr(p ≥ P + δ) **0.05**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Two-Sided Tolerance Interval of Normal Data

Confidence Level 1 - α	Sample Size N	Proportion Covered P	Proportion Covered Exceedance Margin δ	P + δ	Pr(p ≥ P + δ) α'	Tolerance Factor k
0.900	26	0.800	0.150	0.950	0.050	1.6124

PASS also calculates the sample size to be 26. The value of k matches to within rounding.