Chapter 831

Tolerance Intervals for Any Data (Nonparametric)

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 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 procedure is primarily based on results in Guenther (1977) and Hahn and Meeker (1991).

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 order statistics

$$L_1 = X_{(i)},\ \ L_2 = X_{(j)}$$

where $X_{(i)}$ is the i^{th} order statistic found by sorting the data in ascending order and selecting the i^{th} sorted value.

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. In the nonparametric case, these population limits are defined by quantiles of the distribution.

The coverage is the area under the (unknown) distribution between these limits.

Solving for N

The tolerance limits are found by selecting the appropriate order statistics: $X_{(i)}$ and $X_{(j)}$ so that

$$\Pr\left(X_{(j)} - X_{(i)} \geq P\right) = 1 - \alpha$$

Guenther (1977) provides the following two inequalities that can be solved simultaneously for *i*, *j*, and minimum *N* in the two-sided case.

$$E(N - j + i + 1; N, 1 - P) \geq 1 - \alpha$$

$$E(N - j + i + 1; N, 1 - P - \delta) \leq \alpha'$$

where

$$E(r; n, p) = Pr(X \geq r) = \sum_{x=r}^{n} b(x; n, p) = \sum_{x=r}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

It turns out that the solution is for minimum *N* and for the difference $N - j + i + 1$. For example, if the solution to a particular problem turns out to be $N = 38$ and $N - j + i + 1 = 5$, then any of the index pairs (1, 35), (2, 36), (3, 37), and (4, 38) will work. That is, any pair for which $j - i = 34$. Note that here (2, 36) represents the order statistics: $X_{(2)}$ and $X_{(36)}$.

In this example, any of the four pairs are solutions to the two inequalities. Usually, a reasonable choice is to pick one of the central pairs. In the program, we arbitrarily pick (2, 36). But remember that this choice is not unique.

The solution is found using a smart searching algorithm that we developed for the one-proportion test which solves a similar problem.

Example 1 – Calculating Sample Size

Suppose a study is planned to determine the sample size required to compute a two-sided 95% tolerance interval the covers 90% of the population without making specific assumptions about the data distribution. The researchers want to investigate using a *δ* of 0.01, 0.025, or 0.05 with an *α'* of 0.05.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For .... **Sample Size**
Method.. **Binomial Enumeration**
Max N for Binomial Enumeration **100000**
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**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size N
Interval Type: Two-Sided Tolerance Interval

		Proportion of Population Covered				
Confidence Level 1 - α	**Sample Size N**	**Value P**	**Exceedance Margin δ**	**Upper Limit P + δ**	**Pr(p ≥ P + δ) α'**	**Two-Sided Tolerance Interval**
0.95	9309	0.9	0.010	0.910	0.05	X(441), X(8867)
0.95	1387	0.9	0.025	0.925	0.05	X(60), X(1327)
0.95	298	0.9	0.050	0.950	0.05	X(11), X(288)

1 - α	Confidence Level. The proportion of studies with the same settings that produce tolerance intervals with a proportion covered of at least P.
N	The number of subjects.
P	The proportion of the population covered. It is the probability between the tolerance interval limits. It is valid for any distribution.
δ	Proportion Covered Exceedance Margin. The value that is added to P to set an upper bound on the coverage at P + δ.
P + δ	The upper limit of the proportion covered, P. It is a measure of the precision (closeness) of the actual coverage to P.
p	The value of P computed from a random sample.
α'	The probability that the sample value p is greater than P + δ. It is set to a small value such as 0.05 or 0.01. α' = Pr(p ≥ P + δ).
X(i) and X(j)	The ith and jth sample order statistics (i < j). The values within the parentheses are the indices of the order statistics. For example, X(53) means the 53rd observation after the N values are sorted in ascending order. These two values form the limits of the tolerance interval.

Summary Statements

A single-group design will be used to obtain a two-sided 95% tolerance interval where the target proportion of the population covered is 0.9. No distribution is assumed for the underlying data and nonparametric methods will be used. To produce a tolerance interval where the probability that the coverage (0.9) is exceeded by more than 0.01 is 0.05, 9309 subjects will be needed. The order statistics that become the limits of the tolerance interval are X(441), X(8867).

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	9309	11637	2328
20%	1387	1734	347
20%	298	373	75

Dropout Rate	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"). Abbreviated as DR.
N	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'	The total number of subjects that should be enrolled in the study in order to obtain 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., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D	The expected number of dropouts. D = N' - N.

Dropout Summary Statements

Anticipating a 20% dropout rate, 11637 subjects should be enrolled to obtain a final sample size of 9309 subjects.

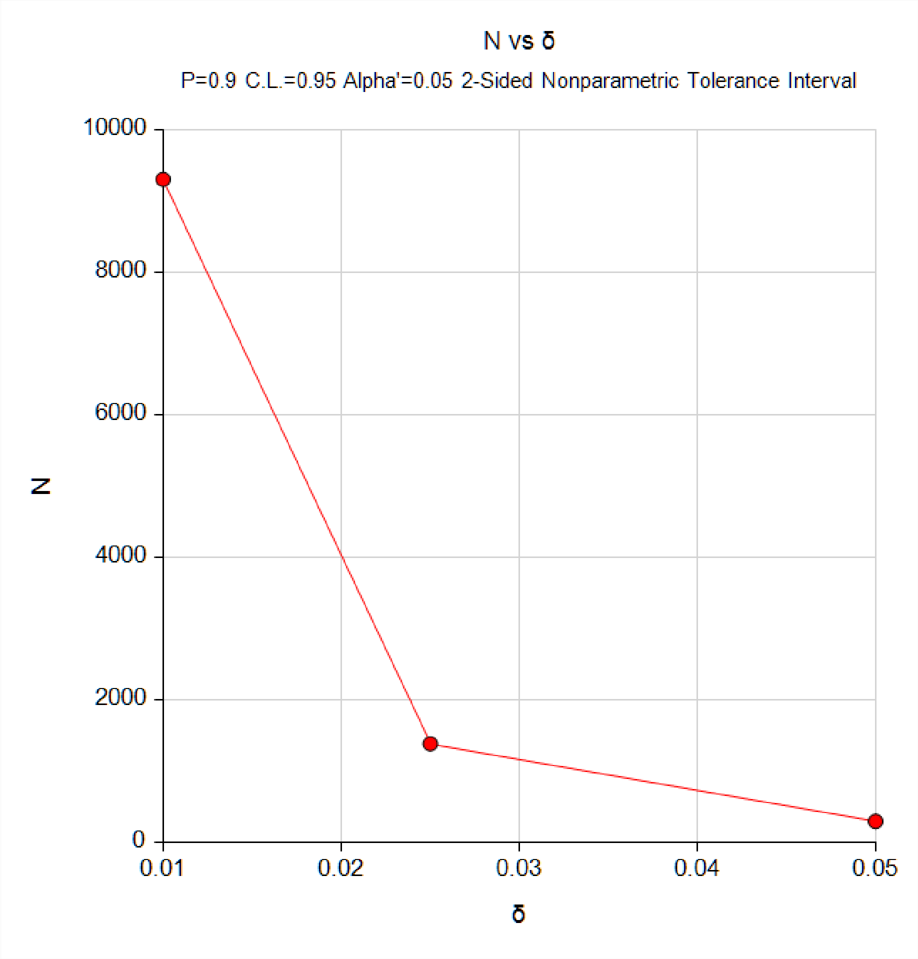
References

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

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

Plots Section

Plots



This plot shows the sample size versus the three value of δ.

Example 2 – Calculating α'

Continuing Example 1, the researchers want to show the impact of various sample sizes on *α*'. They decide to determine the value of *α*' for various value of *N* between 600 and 2200, keeping the other values the same except that they set *δ* to 0.025.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**α' = Pr(p ≥ P + δ)**
Method	**Binomial Enumeration**
Max N for Binomial Enumeration	**100000**
Interval Type	**Two-Sided Tolerance Interval**
N (Sample Size)	**600 to 2200 by 200**
Proportion Covered (P)	**0.9**
Confidence Level (1 – α)	**0.95**
Coverage Proportion Exceedance (δ)	**0.025**

Output

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

Numeric Reports

Numeric Results

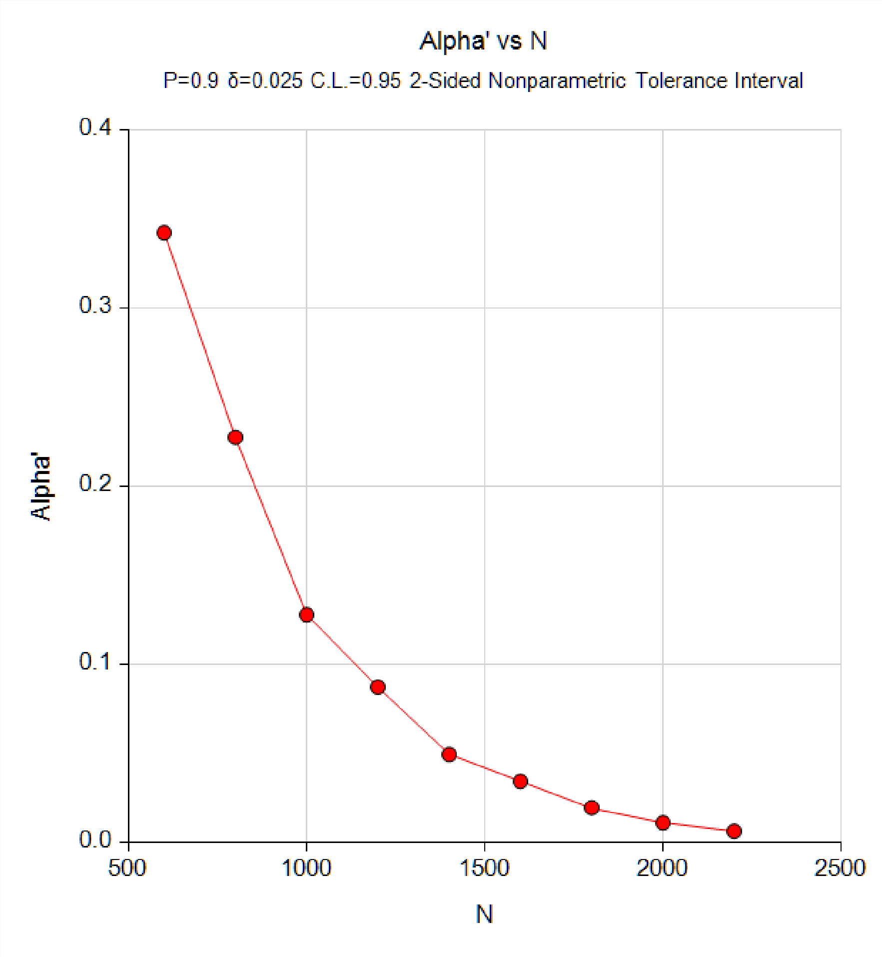
Solve For: α' = Pr(p ≥ P + δ)
Interval Type: Two-Sided Tolerance Interval

		Proportion of Population Covered				
Confidence Level 1 - α	Sample Size N	Value P	Exceedance Margin δ	Upper Limit P + δ	Pr(p ≥ P + δ) α'	Two-Sided Tolerance Interval
0.95	600	0.9	0.025	0.925	0.342	X(23), X(576)
0.95	800	0.9	0.025	0.925	0.228	X(33), X(768)
0.95	1000	0.9	0.025	0.925	0.128	X(42), X(958)
0.95	1200	0.9	0.025	0.925	0.087	X(51), X(1149)
0.95	1400	0.9	0.025	0.925	0.049	X(61), X(1340)
0.95	1600	0.9	0.025	0.925	0.034	X(69), X(1530)
0.95	1800	0.9	0.025	0.925	0.020	X(79), X(1721)
0.95	2000	0.9	0.025	0.925	0.011	X(89), X(1912)
0.95	2200	0.9	0.025	0.925	0.006	X(98), X(2102)

This report shows the impact on α' of various sample sizes. Since the values of the *Tolerance Factor indices* are not related to α' or δ, this report allows you to calculate appropriate indices for use with sample data.

Plots Section

Plots



This plot shows the sample size versus α'.

Example 3 – Validation using Guenther (1977)

Guenther (1977) page 161 gives an example in which P = 0.8, 1 – α = 0.9, $P + \delta$ = 0.95, and α' = 0.05. He obtains a sample size of 38 with lower limit X(2) and upper limit X(36).

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab

Solve For	**Sample Size**
Method	**Binomial Enumeration**
Max N for Binomial Enumeration	**100000**
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

Solve For: Sample Size N
Interval Type: Two-Sided Tolerance Interval

		Proportion of Population Covered				
Confidence Level 1 - α	**Sample Size N**	**Value P**	**Exceedance Margin δ**	**Upper Limit P + δ**	**Pr(p ≥ P + δ) α'**	**Two-Sided Tolerance Interval**
0.9	38	0.8	0.15	0.95	0.05	X(2), X(36)

PASS also calculates a sample size of 38. The values of X(i) and X(j) match as well.