

## Chapter 831

# Tolerance Intervals for Any Data (Nonparametric)

---

### Introduction

This routine calculates the sample size needed to obtain a specified coverage of a  $\beta$ -content tolerance interval at a stated confidence level for data without a specified distribution. These intervals are constructed so that they contain at least  $100\beta\%$  of the population with probability of at least  $100(1 - \alpha)\%$ . 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 that 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

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

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

## Tolerance Intervals for Any Data (Nonparametric)

### Solving for N

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

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

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.

## 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 $\delta$

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

$$\alpha' = \Pr(p \geq P + \delta)$$

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

## Tolerance Intervals for Any Data (Nonparametric)

---

### Coverage Probabilities Calculation Method

#### Method

Select the method to be used to calculate the coverage probabilities. When the sample sizes are reasonably large (i.e. greater than 100) and the coverage proportions are between 0.2 and 0.8, the two methods will give similar results. For smaller sample sizes and more extreme proportions (less than 0.2 or greater than 0.8), the normal approximation is not as accurate so binomial enumeration is more appropriate.

The choices are

#### Binomial Enumeration

Each coverage probability is computed using exact binomial enumeration of all possible outcomes when  $N \leq \text{Max } N$  for Binomial Enumeration (otherwise, the normal approximation is used). Binomial enumeration of all outcomes is possible because of the discrete nature of the data. This method is a little slower for larger  $N$ , but the answers are exact!

#### Normal Approximation

Approximate coverage probabilities are computed using the normal approximation to the binomial distribution. This method is faster, but less accurate.

#### Max N for Binomial Enumeration (Method = None)

When  $N$  is less than or equal to this value, probabilities are calculated using the binomial distribution and enumeration of all possible outcomes. This is possible because of the discrete nature of the data.

When  $N$  is greater than this value, the normal approximation to the binomial is used when calculating probabilities.

We have found that with the speed of modern computers, this value can be set very large. We have put the default at 100,000.

---

### 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.

## Tolerance Intervals for Any Data (Nonparametric)

### Proportion of Population Covered

#### Proportion Covered (P)

Enter the proportion of the population that is covered by the tolerance interval. This is 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 - $\alpha$ )

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 ( $\delta$ )

$\delta$  is added to  $P$  to set an upper bound of  $P' = P + \delta$  on the coverage. Hence,  $\delta$  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*

#### $\alpha' = \Pr(p \geq P + \delta)$

$\alpha'$  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*.

## Tolerance Intervals for Any Data (Nonparametric)

## Example 1 – Calculating Sample Size

Suppose a study is planned to determine the sample size required to compute a two-sided 95% tolerance interval that covers 90% of the population without making specific assumptions about the data distribution. The researcher wants to investigate using a  $\delta$  of 0.01, 0.025, or 0.05 with an  $\alpha'$  of 0.05.

Suppose a study is planned to determine the sample size required to compute a two-sided 95% normal tolerance interval that covers 90% of the population. The researcher wants to investigate using a  $\delta$  of 0.01, 0.02, or 0.05 with an  $\alpha'$  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 Any Data (Nonparametric)** 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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Method .....	<b>Binomial Enumeration</b>
Max N for Binomial Enumeration .....	<b>100000</b>
Interval Type .....	<b>Two-Sided Tolerance Interval</b>
Proportion Covered (P) .....	<b>0.9</b>
Confidence Level (1 - $\alpha$ ) .....	<b>0.95</b>
Coverage Proportion Exceedance ( $\delta$ ) ....	<b>0.01 0.025 0.05</b>
$\alpha' = \Pr(p \geq P + \delta)$ .....	<b>0.05</b>

### Annotated Output

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

### Numeric Results

#### Numeric Results for Nonparametric Two-Sided Tolerance Interval

Confidence Level $1 - \alpha$	Sample Size N	Proportion Covered P	Proportion Covered Exceedance Margin $\delta$	P + $\delta$	Pr( $p \geq P + \delta$ ) $\alpha'$	Two-Sided Tolerance Interval
0.950	9309	0.900	0.010	0.910	0.050	X(441), X(8867)
0.950	1387	0.900	0.025	0.925	0.050	X(60), X(1327)
0.950	298	0.900	0.050	0.950	0.050	X(11), X(288)

#### 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.

#### Report Definitions

- Confidence Level (1 -  $\alpha$ ) 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. It is valid for any distribution.

## Tolerance Intervals for Any Data (Nonparametric)

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

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

$\alpha' = \Pr(p \geq P + \delta)$  is the probability that the coverage computed from a random sample ( $p$ ) is greater than  $P + \delta$ . It is set to a small value such as 0.05 or 0.01.

$X(i)$  and  $X(j)$  are the  $i$ th and  $j$ th 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 two-sided nonparametric tolerance interval computed from a sample of 9309 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  $\geq 0.010$  is 0.050. The order statistics that become the limits of the tolerance interval are  $X(441)$ ,  $X(8867)$ .

### Dropout-Inflated Sample Size

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

### 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").

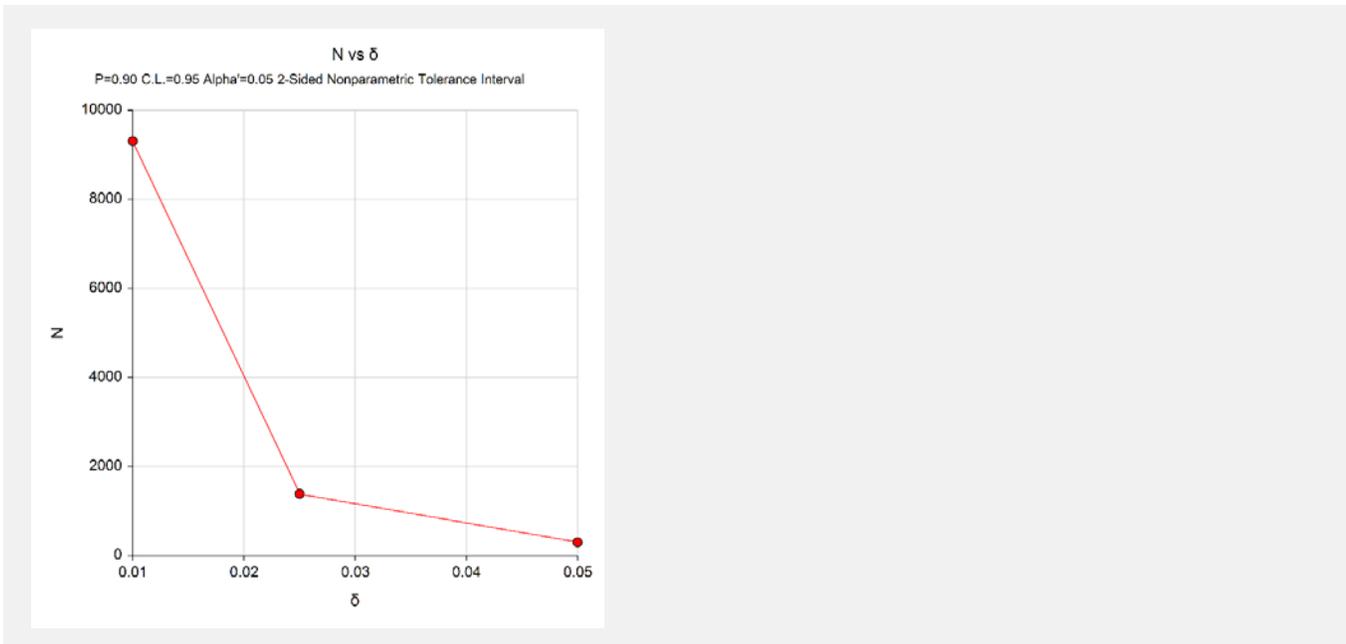
$N$  is the evaluable sample size at which the tolerance interval is computed. If  $N$  subjects are evaluated out of the  $N'$  subjects that are enrolled in the study, the design will achieve the stated tolerance interval.

$N'$  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 value of  $\delta$ .

## Tolerance Intervals for Any Data (Nonparametric)

## Example 2 – Calculating $\alpha'$

Continuing Example 1, the researchers want to show the impact of various sample sizes on  $\alpha'$ . They decide to determine the value of  $\alpha'$  for various values of  $N$  between 600 and 2200, keeping the other values the same except that they set  $\delta$  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 Any Data (Nonparametric)** 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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	$\alpha' = \Pr(p \geq P + \delta)$
Method .....	<b>Binomial Enumeration</b>
Max N for Binomial Enumeration .....	<b>100000</b>
Interval Type .....	<b>Two-Sided Tolerance Interval</b>
N (Sample Size) .....	<b>600 to 2200 by 200</b>
Proportion Covered (P) .....	<b>0.9</b>
Confidence Level ( $1 - \alpha$ ) .....	<b>0.95</b>
Coverage Proportion Exceedance ( $\delta$ ) ....	<b>0.025</b>

### Output

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

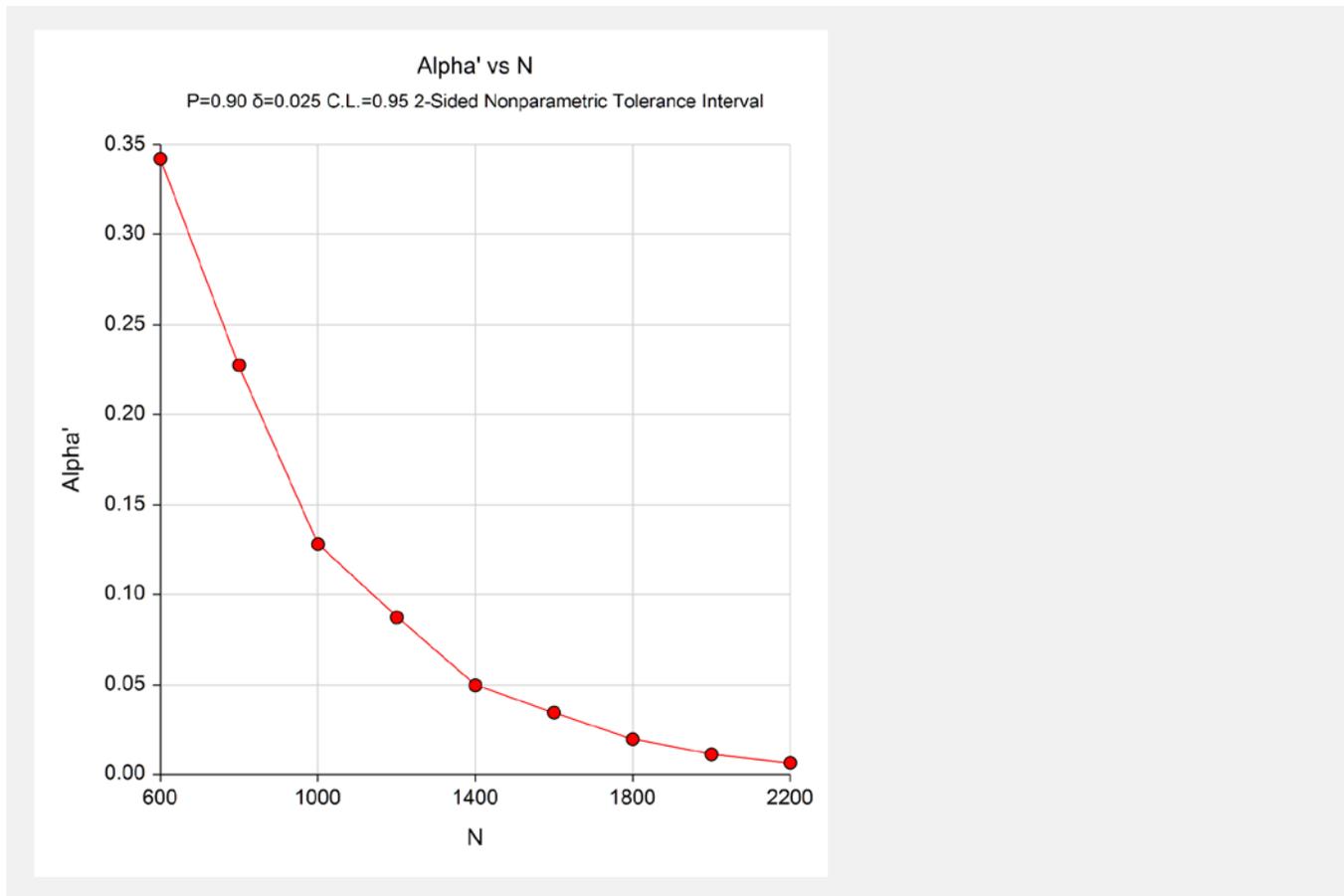
### Numeric Results

Numeric Results for Nonparametric Two-Sided Tolerance Interval							
Confidence Level $1 - \alpha$	Sample Size N	Proportion Covered P	Proportion Covered Exceedance Margin $\delta$	$P + \delta$	$\Pr(p \geq P + \delta)$ $\alpha'$	Two-Sided Tolerance Interval	
0.950	600	0.900	0.025	0.925	0.342	X(23), X(576)	
0.950	800	0.900	0.025	0.925	0.228	X(33), X(768)	
0.950	1000	0.900	0.025	0.925	0.128	X(42), X(958)	
0.950	1200	0.900	0.025	0.925	0.087	X(51), X(1149)	
0.950	1400	0.900	0.025	0.925	0.049	X(61), X(1340)	
0.950	1600	0.900	0.025	0.925	0.034	X(69), X(1530)	
0.950	1800	0.900	0.025	0.925	0.020	X(79), X(1721)	
0.950	2000	0.900	0.025	0.925	0.011	X(89), X(1912)	
0.950	2200	0.900	0.025	0.925	0.006	X(98), X(2102)	

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

## Tolerance Intervals for Any Data (Nonparametric)

## Plots Section



This plot shows the sample size versus  $\alpha'$ .

## Tolerance Intervals for Any Data (Nonparametric)

**Example 3 – Validation using Guenther (1977)**

Guenther (1977) page 161 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 38 with lower limit  $X(2)$  and upper limit  $X(36)$ .

**Setup**

This section presents the values of each of the parameters needed to run this example. First load the **Tolerance Intervals for Any Data (Nonparametric)** 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**.

<u>Option</u>	<u>Value</u>
<b>Design Tab</b>	
Solve For .....	<b>Sample Size</b>
Method .....	<b>Binomial Enumeration</b>
Max N for Binomial Enumeration .....	<b>100000</b>
Interval Type .....	<b>Two-Sided Tolerance Interval</b>
Proportion Covered (P) .....	<b>0.8</b>
Confidence Level ( $1 - \alpha$ ) .....	<b>0.9</b>
Coverage Proportion Exceedance ( $\delta$ ) ....	<b>0.15</b>
$\alpha' = \Pr(p \geq P + \delta)$ .....	<b>0.05</b>

**Output**

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

**Numeric Results**

Numeric Results for Nonparametric Two-Sided Tolerance Interval						
Confidence Level	Sample Size	Proportion Covered	Proportion Covered Exceedance Margin		$\Pr(p \geq P + \delta)$	Two-Sided Tolerance Interval
$1 - \alpha$	N	P	$\delta$	$P + \delta$	$\alpha'$	
0.900	38	0.800	0.150	0.950	0.050	$X(2), X(36)$

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