

Chapter 813

Nonparametric Reference Intervals for Non-Normal Data

Introduction

This routine calculates the sample size needed to obtain a specified relative width of a confidence interval about the limits of a *nonparametric* reference interval based on ranks of the data.

Consider a measurement made on a population of individuals (usually healthy patients). A **reference interval** (RI) of this measurement gives the boundaries between which a typical measurement is expected to fall. When a measurement occurs that is outside these reference interval boundaries, there is cause for concern. That is, the measurement is unusually high or low. The reference interval is often presented as percentiles of a reference population, such as the 2.5th percentile and the 97.5th percentile. Of course, the choice of the reference population is important, and you would expect that there is often differences according to age, size, and so on. Hence, the population must be partitioned into homogeneous subpopulations and individual reference intervals determined for each.

This procedure calculates sample sizes based on the methods shown in Machin, Campbell, Tan, and Tan (2018). These methods are for data that is non-normal and can't be easily transformed to normal. The method is nonparametric in that only the ranks of the data are used.

Technical Details

Reference Interval

Assume that a sample of N observations of an i.i.d. random value Y are ranked from smallest to largest. Refer to these ranked values as $y[1], y[2], \dots, y[N]$. The $100(1 - \alpha)\%$ *reference interval* is given by the *reference limits* R_L and R_U where $R_L = y[N\alpha/2]$ and $R_U = y[N(1 - \alpha/2)]$.

Actually, the limits are usually formed by some form of interpolation since $N\alpha/2$ will seldom result in an integer.

Confidence Intervals of Reference Limits

The $100(1 - \gamma)\%$ confidence intervals for each of the reference limits are each estimated from the following ranks based on quantiles q

$$r_q = Nq - \left\lceil z_{1-\frac{\gamma}{2}} \sqrt{Nq(1-g)} \right\rceil$$

$$s_q = 1 + Nq + \left\lceil z_{1-\frac{\gamma}{2}} \sqrt{Nq(1-g)} \right\rceil$$

where q is set to $\alpha/2$ for the C.I. of R_L and to $1-\alpha/2$ for the C.I. of R_U .

Relative Margin of Error

The relative width (ignoring the standard deviation) of the above confidence interval is approximated by

$$W_{CI} = 2 \eta z_{1-\frac{\gamma}{2}} / \sqrt{N}$$

where

$$\eta = \frac{\sqrt{\frac{\gamma}{2} \left(1 - \frac{\gamma}{2}\right)}}{\phi}$$

$$\phi = \frac{\exp\left(-\frac{1}{2} z_{1-\frac{\gamma}{2}}^2\right)}{\sqrt{2\pi}}$$

By rearranging the sample size formula on page 343 of Machin, Campbell, Tan, and Tan (2018), the margin of error is given by

$$ME = 100 \sqrt{\frac{\sqrt{3}\eta}{N} \left(\frac{z_{1-\frac{\gamma}{2}}}{z_{1-\frac{\alpha}{2}}}\right)^2}$$

Finally, the sample size is obtained through a binary search of possible values.

Example 1 – Calculating Sample Size

Suppose a study is being planned to obtain 90% confidence intervals of the 95% reference limits of blood pressure (BP) for a certain gender and age range. Sample sizes are needed for relative margin of errors from 10% to 30%. The goal is to determine the necessary sample size for each value of the relative margin of error.

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 **N (Sample Size)**
 RI (Reference Interval Percentage) **95**
 CL (Confidence Level Percentage) **90**
 ME (Percentage Margin of Error) **10 20 30**

Output

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

Numeric Reports

Numeric Results

Solve For: **N (Sample Size)**

Sample Size N	Relative Margin of Error ME	Percentage of the Population in the Reference Interval RI	Confidence Level of the Confidence Intervals for the Reference Limits CL	Width of the Confidence Intervals for the Reference Limits WCL*	Width of the Reference Interval WRI*
258	10.0	95	90	0.433	4.33
65	19.9	95	90	0.862	4.33
29	29.8	95	90	1.291	4.33

* The standard deviation was assumed to be 1.0.

- N The total sample size of the study.
- ME The margin of error, which is a measure of imprecision. It is the percentage that the width of the confidence interval of the reference limits is of the width of the reference interval. $ME = 100 WCL / WRI$.
- RI The reference interval percentage. It is the percentage of data values that are anticipated to be between the reference limits assuming a normal distribution.
- CL The confidence level of the confidence intervals made for each reference limit. It is a percentage.
- WCL The width of the confidence interval made for each reference limit. The standard deviation was assumed to be 1.0.
- WRI The width of the reference interval. The standard deviation was assumed to be 1.0.

Nonparametric Reference Intervals for Non-Normal Data

Summary Statements

A single-group design will be used to obtain a two-sided reference interval based on nonparametric estimation methods. To obtain a 95% reference interval with 90% confidence intervals of the limits, and with a target relative margin of error of 10, 258 subjects will be needed. The relative margin of error is the width of the confidence intervals of the limits as a percentage of the width of the reference interval. For this scenario, if the standard deviation is assumed to be 1, the width of the confidence intervals of the reference limits is 0.433, and the width of the reference interval is 4.33.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout- Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	258	323	65
20%	65	82	17
20%	29	37	8

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 reference 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 reference 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, 323 subjects should be enrolled to obtain a final sample size of 258 subjects.

References

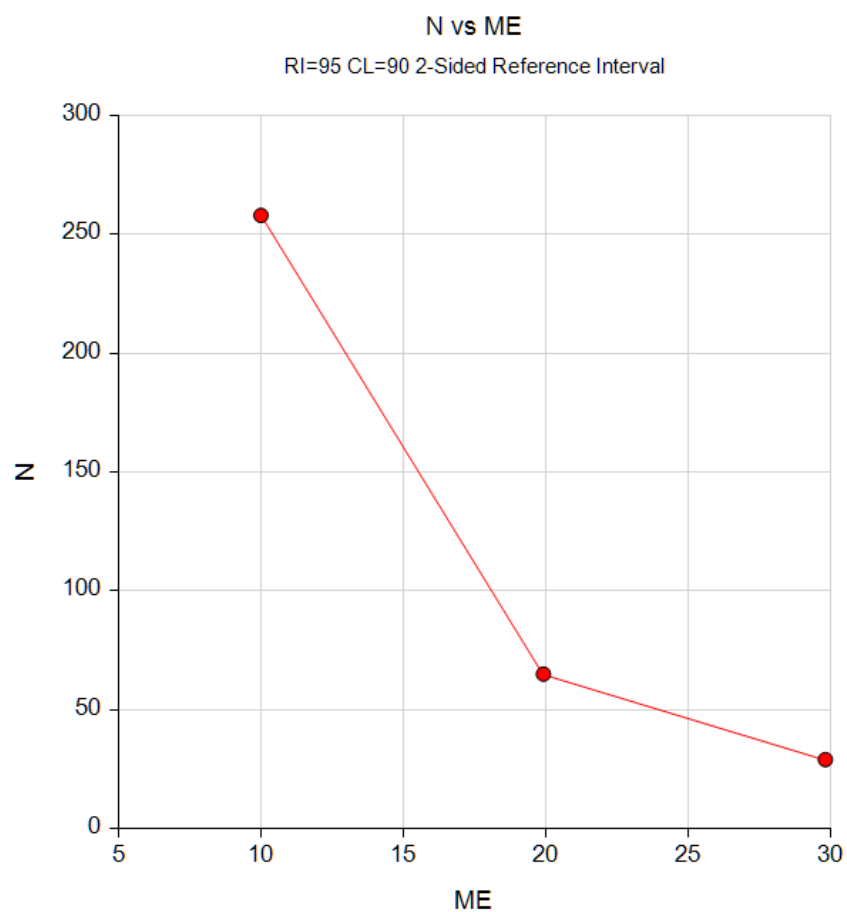
Machin, D, Campbell, M.J., Tan, S.B, Tan, S.H. 2018. 'Sample Sizes for Clinical, Laboratory and Epidemiology Studies, Fourth Edition'. John Wiley and Sons. Hoboken, New Jersey.

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

Nonparametric Reference Intervals for Non-Normal Data

Plots Section

Plots



This plot shows the sample size versus the margin of error (MoE).

Example 2 – Validation using Machin et al. (2018)

Machin *et al.* (2018) page 351 give Table 21.2 which contains sample sizes for various scenarios. We will duplicate the first row of this table as a validation example. In this example RI = 90; CI = 80, 90, 95; and ME = 2.5. The sample sizes are computed as 2876, 5857, and 10512.

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 **N (Sample Size)**
 RI (Reference Interval Percentage) **90**
 CL (Confidence Level Percentage) **80 90 95**
 ME (Percentage Margin of Error) **2.5**

Output

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

Numeric Results

Solve For: **N (Sample Size)**

Sample Size N	Relative Margin of Error ME	Percentage of the Population in the Reference Interval RI	Confidence Level of the Confidence Intervals for the Reference Limits CL	Width of the Confidence Intervals for the Reference Limits WCL*	Width of the Reference Interval WRI*
2876	2.5	90	80	0.082	3.268
5857	2.5	90	90	0.091	3.634
10512	2.5	90	95	0.102	4.085

* The standard deviation was assumed to be 1.0.

PASS matches the first row of Table 21.2 on page 351 exactly.