Chapter 586

Reference Intervals

Introduction

A reference interval contains the middle 95% of measurements of a substance from a healthy population. It is a type of prediction interval. This procedure calculates one-, and two-, sided reference intervals using three different methods promoted by CLSI EP28-A3c: normal distribution, nonparametric-percentiles, or robust percentile estimators.

Horn and Pesce (2005) state that "The reference interval is the most widely used medical decision-making tool. It is central to the determination of whether or not an individual is healthy." Not only does this procedure calculate a reference interval for a set of data, it also allows one to study whether the sample meets the various assumptions needed for an accurate reference interval to be formed.

Technical Details

Let $X_1, X_2, \ldots, X_n$ be a random sample of size *n* from a population with distribution function $F(X)$. A two-sided, 100(1-*α*)% reference interval (R_L, R_U) for a new observation X_{new} is defined as

$$P[R_L \leq X_{new} \leq R_U] = 1 - \alpha$$

One-sided intervals are defined similarly.

CLSI EP28-A3c discusses three methods of computing these limits along with their confidence intervals. These are presented next using this document as well as Horn and Pesce (2005).

Normal-Theory Method

This method is based on traditional normal theory. If the data are not normally distributed, you can try the *Box-Cox Transformation* procedure to determine if a power transformation will bring the distribution closer to normal.

The following formulation is given by Horn and Pesce (2005). The lower and upper limits of the reference interval are defined as

$$R_L = \bar{x} + t_{\frac{\alpha}{2},n-1} s \sqrt{1 + \frac{1}{n}}$$

$$R_U = \bar{x} + t_{1-\frac{\alpha}{2},n-1} s \sqrt{1 + \frac{1}{n}}$$

where $\bar{x}$ is the sample mean and *s* is the sample standard deviation.

CLSI recommends 90% confidence intervals be calculated for the two reference limits. The formulas for these confidence intervals are

$$R_L \pm z_{\gamma/2} s_{\alpha/2}$$

and

$$R_U \pm z_{\gamma/2} s_{\alpha/2}$$

where

$$s_{\alpha/2} = s\sqrt{\frac{2 + z_{\alpha/2}^2}{2n}}$$

$$\gamma = 0.90$$

Here, *z* is the standard normal variate.

Percentile (Nonparametric) Method

The following formulation for the percentile method is given by Horn and Pesce (2005). In this case, the lower and upper limits of the reference interval are defined as the $100(\alpha/2)$ and $100(1-\alpha/2)$ percentiles of the sorted data values.

There is some controversy over the definition of a percentile. **NCSS** provides you with five choices. CLSI recommend

$$\hat{F}(p) = (1-r)X_{(j)} + rX_{(j+1)}$$

where $Y_{(j)}$ is the j^{th} ordered value, $j = [(n+1)p]$, $r = (n+1)p - j$, [z] is the integer part of z, and $X_{(n+1)} = X_{(n)}$.

CLSI recommends γ% confidence intervals be calculated for the two reference limits where $\gamma = 0.9$. The formula for the confidence interval of the lower reference limit is the interval $\left(x_{(l)}, x_{(r)}\right)$ where

$$\sum_{i=l}^{r-1}\binom{n}{i}\left(\frac{\alpha}{2}\right)^i\left(1-\frac{\alpha}{2}\right)^{n-i} = \sum_{i=0}^{r-1}\binom{n}{i}\left(\frac{\alpha}{2}\right)^i\left(1-\frac{\alpha}{2}\right)^{n-i} - \sum_{i=0}^{l-1}\binom{n}{i}\left(\frac{\alpha}{2}\right)^i\left(1-\frac{\alpha}{2}\right)^{n-i} \geq \gamma$$

Note that this is the difference between two cumulative binomial probabilities.

The confidence interval for the upper reference limit is the interval $\left(x_{(n-r+1)}, x_{(n-l+1)}\right)$ were *l* and *r* are defined above.

The values of *l* and *r* are found using a search procedure that has three goals:

1. Meet the confidence coefficient γ requirement. Usually set to 0.90.
2. Keep the limits as symmetric as possible.
3. Keep the width of the interval as narrow as possible.

A popular table of solutions to this problem for various sample sizes is given in Table 8 of CLSI EP28-A3c. We have found that occasionally a better solution can be found than the one given in Table 8 in the sense that it is narrower, more symmetric, or closer to the nominal value of 0.90. Be default, **NCSS** presents this optimum solution. It does, however, provide an option that forces the use of the less-optimal Table 8 solution when the settings allow this.

Here is an example that shows how our algorithm converges to the optimum solution.

In this example, N = 388, α = 0.05, and γ = 0.90. The percentile of 9.725 results in P = 0.0250644 and NP = 9.7. Note that a symmetry measure is presented which is $(r - 9.725) - (l - 9.725)$. If the limits where perfectly symmetric, this value would be 0.

NCSS begins the search at the rounded percentile value. In this case, since the percentile is 9.725, we begin with observations 9 and 10.

r	**l**	**CIγ**	**Width**	**Symmetry**
10	9	0.12982	1	-0.45
11	9	0.25598	2	0.55
11	8	0.37589	3	-0.45
12	8	0.48705	4	0.55
12	7	0.58524	5	-0.45
13	7	0.67479	6	0.55
13	6	0.74497	7	-0.45
14	6	0.81137	8	0.55
14	5	0.85425	9	-0.45
15	5	0.89986	10	0.55
15	**4**	**0.92163**	**11**	**-0.45**
16	5	0.92902	11	1.55

Step 0. The search begins with 10 and 9, the two integers that surround the percentile of 9.725.

Step 1. The value of r is increased by one. The resulting confidence coefficient is 0.25598.

Step 2. The value of l is decreased by one. The resulting confidence coefficient is 0.37589.

The algorithm continues step by step. First, r is increased by one, then l is decreased by one.

The algorithm terminates with r = 15 and l = 4 since the confidence coefficient of 0.92163 is the first to be greater than 0.9.

For comparison, we present the results for r =16 and l = 5 which is the CLSE Table 8 value. Its confidence coefficient is close to the optimum, its width is the same, but its symmetry value is 1.55 (more lop-sided).

Robust Method

The robust algorithm is given in Appendix B of CLSI EP28-A3c. This is a rather long algorithm and it is not repeated here.

Confidence intervals for the two limits are calculated using the percentile bootstrap method. This method requires a medium to large (not small!) sample size.

Data Structure

The data are contained in a single column.

Example 1 – Generating Percentile Reference Intervals

This section presents a detailed example of how to generate nonparametric-percentile reference intervals for the *Calcium* variable in the Calcium dataset. This dataset contains 120 calcium measurements from males and 120 calcium measurements from females. To run this example, take the following steps:

Setup

To run this example, complete the following steps:

1 Open the Calcium example dataset

- From the File menu of the NCSS Data window, select **Open Example Data**.
- Select **Calcium** and click **OK**.

2 Specify the Reference Intervals procedure options

- Find and open the **Reference Intervals** procedure using the menus or the Procedure Navigator.
- The settings 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.

Variables Tab

Variables..**Calcium**
Group Variable..**Gender**

Reports Tab

Descriptive Statistics...**Checked**
Normality ....**Checked**
Quantiles ....**Checked**
Reference Intervals using the Percentile Method ....**Checked**
Use the CLSI-EP28-A3C Percentile Method (Table 8) ......**Unchecked**
if available
Reference Intervals using the Normal-Theory Method ......**Checked**
Reference Intervals using the Robust Method..................**Checked**
Random Seed...**3883571** (for reproducibility)

3 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Descriptive Statistics

Descriptive Statistics of Calcium

Gender	Count	Mean	Median	Standard Deviation	IQR	Minimum	Maximum
Men	120	9.700	9.7	0.3272	0.5	9.0	10.6
Women	120	9.474	9.5	0.2926	0.4	8.7	10.3
Combined	240	9.587	9.6	0.3298	0.4	8.7	10.6

This report gives a statistical summary of the data. The “Combined” line gives the values for all groups combined.

Count

This is the number of nonmissing values. If no frequency variable was specified, this is the number of nonmissing rows.

Mean

This is the average of the data values.

Median

This is the median of the data values.

Standard Deviation

This is the standard deviation of the data values.

IQR

This is the interquartile range. It is the difference between the third quartile and the first quartile (between the 75th percentile and the 25th percentile). This represents the range of the middle 50 percent of the distribution. It is a very robust (not affected by outliers) measure of dispersion. In fact, if the data are normally distributed, a robust estimate of the sample standard deviation is IQR/1.35. If a distribution is very concentrated around its mean, the IQR will be small. On the other hand, if the data are widely dispersed, the IQR will be much larger.

Minimum

The smallest value in this variable.

Maximum

The largest value in this variable.

Normality Report

Normality Report of Calcium

						Normality Test P-Value	
Gender	Mean	Standard Deviation	COV	Skewness (Normal = 0)	Kurtosis (Normal = 3)	Anderson-Darling	Shapiro-Wilk
Men	9.700	0.3272	0.0337	0.0145	2.5092	0.0223	0.0548
Women	9.474	0.2926	0.0309	0.0481	3.0217	0.0434	0.3286
Combined	9.587	0.3298	0.0344	0.1247	2.7586	0.0029	0.0383

This report gives statistics that help you evaluate the normality assumption.

Mean

This is the average of the data values.

Standard Deviation

This is the standard deviation of the data values.

COV

The *coefficient of variation* is a relative measure of dispersion. It is most often used to compare the amount of variation in two samples. It can be used for the same data over two time periods or for the same time period but two different places. It is the standard deviation divided by the mean:

$$COV = s/\bar{x}$$

Skewness (Normal = 0)

This statistic measures the direction and degree of asymmetry. A value of zero indicates a symmetrical distribution. A positive value indicates skewness (longtailedness) to the right while a negative value indicates skewness to the left. Values between -3 and +3 are typical values of samples from a normal distribution.

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}}$$

Kurtosis (Normal = 3)

This statistic measures the heaviness of the tails of a distribution. The usual reference point in kurtosis is the normal distribution. If this kurtosis statistic equals three and the skewness is zero, the distribution is normal. Unimodal distributions that have kurtosis greater than three have heavier or thicker tails than the normal. These same distributions also tend to have higher peaks in the center of the distribution (leptokurtic). Unimodal distributions whose tails are lighter than the normal distribution tend to have a kurtosis that is less than three. In this case, the peak of the distribution tends to be broader than the normal (platykurtic). Be forewarned that this statistic is an unreliable estimator of kurtosis for small sample sizes.

$$b_2 = \frac{m_4}{m_2^2}$$

Anderson-Darling Test

This test, developed by Anderson and Darling (1954), is the most popular normality test that is based on EDF statistics. In some situations, it has been found to be as powerful as the Shapiro-Wilk test.

Unfortunately, both the Shapiro-Wilk and Anderson-Darling tests have small statistical power (probability of detecting nonnormal data) unless the sample sizes are large, say over 100. Hence, if the decision is to reject, you can be reasonably certain that the data are not normal. However, if the decision is to accept, the situation is not as clear. If you have a sample size of 100 or more, you can reasonably assume that the actual distribution is closely approximated by the normal distribution. If your sample size is less than 100, all you know is that there was not enough evidence in your data to reject the normality assumption. In other words, the data might be nonnormal, you just could not prove it. In this case, you must rely on the graphics and past experience to justify the normality assumption.

Shapiro-Wilk W Test

This test for normality has been found to be the most powerful test in most situations. It is the ratio of two estimates of the variance of a normal distribution based on a random sample of *n* observations. The numerator is proportional to the square of the best linear estimator of the standard deviation. The denominator is the sum of squares of the observations about the sample mean. The test statistic W may be written as the square of the Pearson correlation coefficient between the ordered observations and a set of weights which are used to calculate the numerator. Since these weights are asymptotically proportional to the corresponding expected normal order statistics, W is roughly a measure of the straightness of the normal quantile-quantile plot. Hence, the closer W is to one, the more normal the sample is.

The test was developed by Shapiro and Wilk (1965) for samples up to 20. **NCSS** uses the approximations suggested by Royston (1992) and Royston (1995) which allow unlimited sample sizes. Note that Royston only checked the results for sample sizes up to 5000 but indicated that he saw no reason larger sample sizes should not work.

The probability values for W are valid for samples greater than 3.

This test may not be as powerful as other tests when ties occur in your data.

Quantile Report

Quantile Report of Calcium

	Percentile						
Gender	**5th**	**10th**	**25th**	**50th**	**75th**	**90th**	**95th**
Men	9.105	9.2	9.5	9.7	10.0	10.1	10.200
Women	9.000	9.1	9.3	9.5	9.7	9.9	10.000
Combined	9.100	9.2	9.4	9.6	9.8	10.0	10.195

This report gives various percentiles of the data distribution.

Percentile Reference Interval

Two-Sided 95% Percentile Reference Interval for Calcium

		2.5% Lower Reference Limit			97.5% Upper Reference Limit		
			90% Confidence Interval Limits			90% Confidence Interval Limits	
Gender	Count	Value	Lower	Upper	Value	Lower	Upper
Men	120	9.1	9.0	9.2	10.300	10.2	10.6
Women	120	8.9	8.7	9.0	10.098	10.0	10.3
Combined	240	9.0	8.8	9.0	10.200	10.2	10.3

This report gives reference intervals and associated confidence intervals based on the percentile method.

Normal-Theory Reference Interval

Two-Sided 95% Normal-Theory Reference Interval for Calcium

		2.5% Lower Reference Limit			97.5% Upper Reference Limit		
			90% Confidence Interval Limits			90% Confidence Interval Limits	
Gender	Count	Value	Lower	Upper	Value	Lower	Upper
Men	120	9.049	8.965	9.133	10.351	10.267	10.435
Women	120	8.892	8.817	8.967	10.056	9.981	10.131
Combined	240	8.936	8.876	8.996	10.238	10.178	10.298

This report gives reference intervals and associated confidence intervals based on the normal-theory method.

Robust Reference Interval

Two-Sided 95% Robust Reference Interval for Calcium

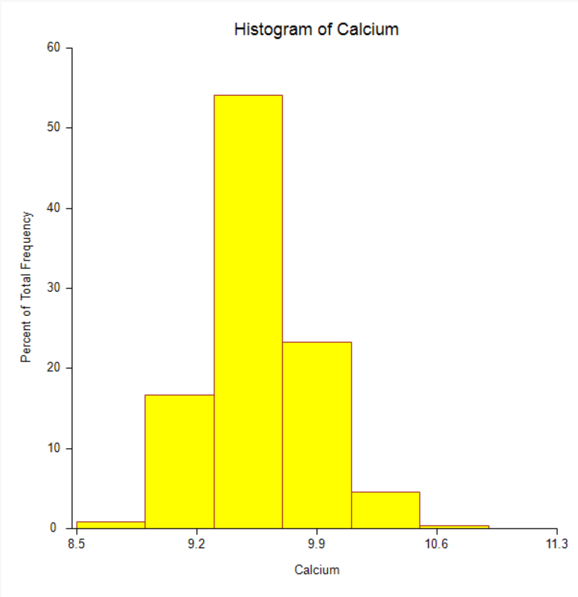
		2.5% Lower Reference Limit			97.5% Upper Reference Limit		
			90% Confidence Interval Limits			90% Confidence Interval Limits	
Gender	Count	Value	Lower	Upper	Value	Lower	Upper
Men	120	9.050	8.966	9.135	10.352	10.270	10.430
Women	120	8.888	8.809	8.969	10.057	9.981	10.132
Combined	240	8.927	8.869	8.988	10.230	10.172	10.289

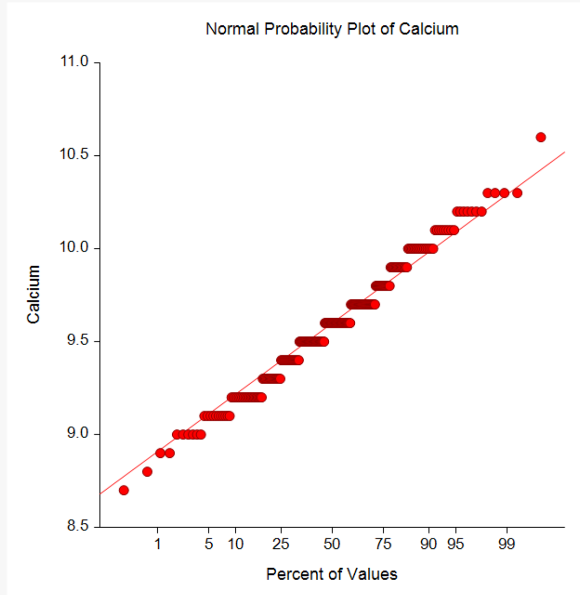
Robust Method Constants: c1 = 3.700, c2 = 205.408, MAD Scale Factor = 0.674500
Bootstrap C.I. Estimation: Number of Samples = 3000, User-Entered Random Seed = 3883571

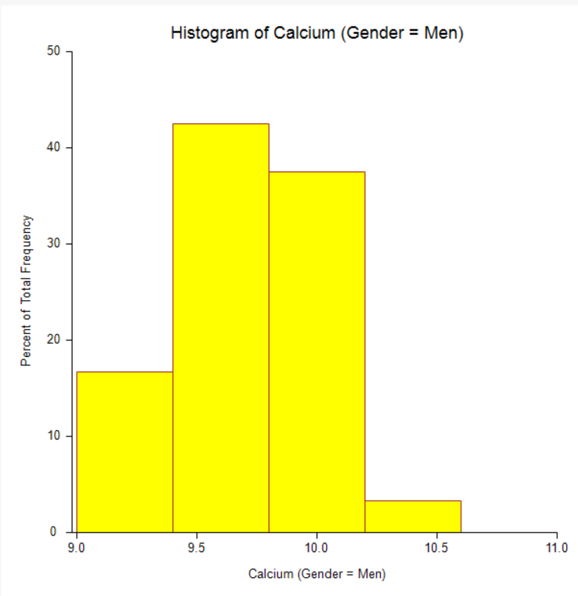
This report gives reference intervals and associated confidence intervals based on the robust method.

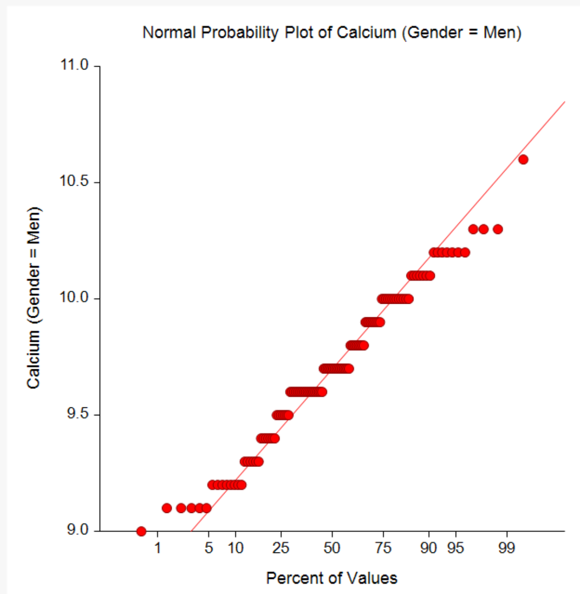
Plots Section

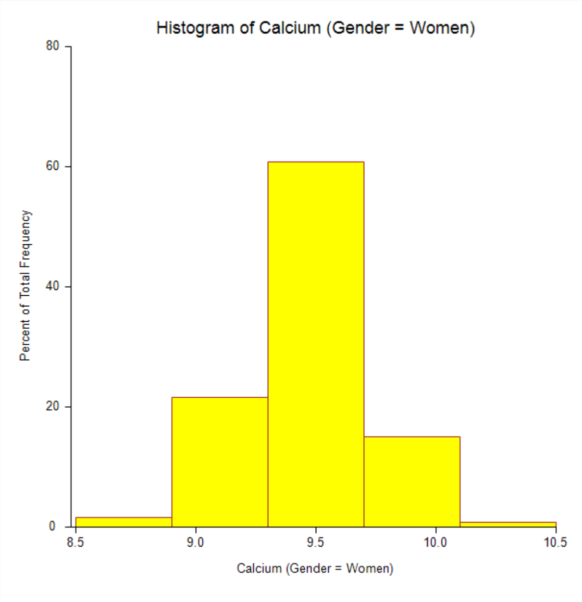
Plots Section

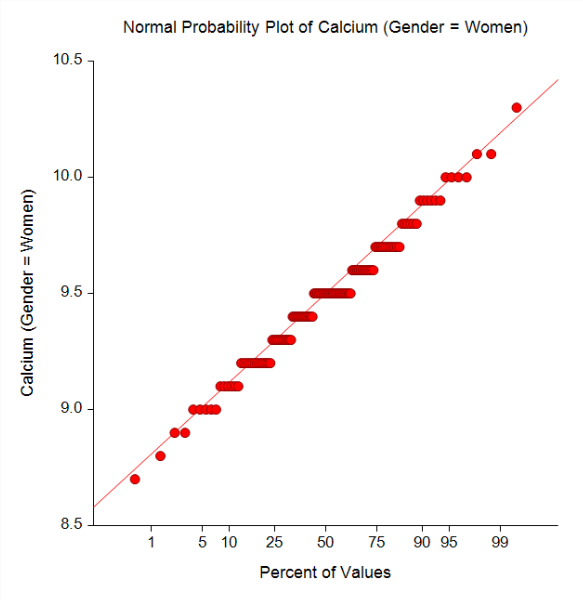












The plots section displays a histogram and a probability plot for each line of the reports that let you assess the accuracy of the normality assumption.