Chapter 801

Confidence Interval for Pearson's Correlation

Introduction

This routine calculates the sample size needed to obtain a specified width of a Pearson product-moment correlation coefficient confidence interval at a stated confidence level.

Caution: This procedure requires a planning estimate of the sample correlation. The accuracy of the sample size depends on the accuracy of this planning estimate.

Technical Details

This procedure is based on the results of Bonett and Wright (2000). Assuming a bivariate normal population with population correlation ρ, the transformation of the sample product moment correlation from r to z_r

$$z_r = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$

is approximately normally distributed with variance $1/(n-3)$ (Fisher, 1921). The lower and upper confidence limits for ρ are obtained by computing

$$z_r \pm z_{1-\alpha/2}\sqrt{\frac{1}{n-3}}$$

to obtain z_L and z_U. The values of z_L and z_U are then transformed back to the correlation scale using the inverse transformations

$$r_L = \frac{\exp(2z_L) - 1}{\exp(2z_L) + 1}$$

and

$$r_U = \frac{\exp(2z_U) - 1}{\exp(2z_U) + 1}$$

One-sided limits may be obtained by replacing $\alpha/2$ by α.

For two-sided intervals, the distance from the sample correlation to each of the limits may be different. Thus, instead of specifying the distance to the limits we specify the width of the interval, W.

The basic equation for determining sample size for a two-sided interval when W has been specified is

$$W = r_U - r_L$$

For one-sided intervals, the distance from the sample correlation to limit, D, is specified.

The basic equation for determining sample size for a one-sided upper limit when D has been specified is

$$D = r_U - r$$

The basic equation for determining sample size for a one-sided lower limit when D has been specified is

$$D = r - r_L$$

Each of these equations can be solved for any of the unknown quantities in terms of the others.

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population correlation is $1 - \alpha$.

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

This option specifies the parameter to be solved for from the other parameters.

One-Sided or Two-Sided Interval

Interval Type

Specify whether the confidence interval for the population correlation is two-sided or one-sided. A one-sided interval is often called a **confidence bound** rather than a confidence interval because it only has one limit.

Two-Sided

The two-sided confidence interval is defined by two limits: an upper confidence limit (UCL) and a lower confidence limit (LCL).

These limits are constructed so that the designated proportion (confidence level) of such intervals will include the true population value.

Upper One-Sided

The upper confidence interval (or bound) is defined by a limit above the estimated parameter value. The limit is constructed so that the designated proportion (confidence level) of such limits has the true population value below them.

Lower One-Sided

The lower confidence interval (or bound) is defined by a limit below the estimated parameter value. The limit is constructed so that the designated proportion (confidence level) of such limits has the true population value above them.

Confidence

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of n items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population correlation is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as *0.90, 0.95* or *0.90 to 0.99 by 0.01*.

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.

Precision

Confidence Interval Width (Two-Sided)

This is the distance from the lower confidence limit to the upper confidence limit.

You can enter a single value or a list of values. The value(s) must be between 0 and 2.

Distance from r to Limit (One-Sided)

This is the distance from the sample correlation to the lower or upper limit of the confidence interval, depending on the Interval Type.

You can enter a single value or a list of values. The value(s) must be between 0 and 2.

Estimated Sample Correlation

r (Sample Correlation)

Enter an estimate of the sample Pearson product-moment correlation calculated from the usual formula. This value can be obtained from prior studies, expert opinion, or as a reasonable guess. The sample size and width calculations assume that the value entered here is the actual correlation estimate obtained from the sample. The accuracy of your results will depend on the accuracy of this estimate.

The range of the values of the sample correlation that can be entered is -1 to 1.

You can enter a range of values such as 0.1 0.3 0.5 or 0.1 to 0.5 by .2.

Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the population Pearson correlation such that the width of the interval is no wider than 0.08. The researcher would like to examine a large range of sample correlation values to determine the effect of the correlation estimate on necessary sample size. Instead of examining only the interval width of 0.08, widths of 0.06 and 0.10 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Interval for Pearson Correlation** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Confidence Interval**, and then clicking on **Confidence Interval for Pearson Correlation**. 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**
Confidence Level (1 – Alpha) **0.95**
Confidence Interval Width (Two-Sided).. **0.06 0.08 0.10**
r (Sample Correlation) .... **-0.9 to 0.9 by 0.1**

Annotated Output

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

Numeric Results

Numeric Results for Two-Sided Confidence Intervals for Pearson Correlation

Confidence Level	Sample Size N	Target Width	Actual Width	Sample Correlation r	C.I. Lower Limit	C.I. Upper Limit	Width if r = 0.0
0.950	161	0.060	0.060	-0.900	-0.926	-0.866	0.309
0.950	559	0.060	0.060	-0.800	-0.828	-0.768	0.166
0.950	1115	0.060	0.060	-0.700	-0.729	-0.669	0.117
0.950	1752	0.060	0.060	-0.600	-0.629	-0.569	0.094
0.950	2404	0.060	0.060	-0.500	-0.529	-0.469	0.080
0.950	3014	0.060	0.060	-0.400	-0.430	-0.370	0.071
0.950	3536	0.060	0.060	-0.300	-0.330	-0.270	0.066
0.950	3935	0.060	0.060	-0.200	-0.230	-0.170	0.062
0.950	4184	0.060	0.060	-0.100	-0.130	-0.070	0.061
0.950	4269	0.060	0.060	0.000	-0.030	0.030	0.060
0.950	4184	0.060	0.060	0.100	0.070	0.130	0.061
0.950	3935	0.060	0.060	0.200	0.170	0.230	0.062
0.950	3536	0.060	0.060	0.300	0.270	0.330	0.066
0.950	3014	0.060	0.060	0.400	0.370	0.430	0.071
0.950	2404	0.060	0.060	0.500	0.469	0.529	0.080
0.950	1752	0.060	0.060	0.600	0.569	0.629	0.094
0.950	1115	0.060	0.060	0.700	0.669	0.729	0.117
0.950	559	0.060	0.060	0.800	0.768	0.828	0.166
0.950	161	0.060	0.060	0.900	0.866	0.926	0.309
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

Confidence Interval for Pearson's Correlation

Report Definitions
Confidence level is the proportion of confidence intervals (constructed with this same confidence level,
    sample size, etc.) that would contain the true correlation.
Sample Size N is the size of the sample drawn from the population.
Width is the distance from the lower limit to the upper limit.
Target Width is the value of the width that is entered into the procedure.
Actual Width is the value of the width that is obtained from the procedure.
r is the estimate of Pearson's product moment correlation coefficient.
Lower and Upper Limit are the lower and upper limits of the confidence interval.
Width if r = 0.0 is the maximum width for a confidence interval with sample size N.

References
Bonett, D. G. and Wright, T. A. 2000. 'Sample Size Requirements for Estimating Pearson, Kendall and Spearman
    Correlations.' Psychometrika, Vol 65, No 1 (March), 23-28.
Looney, S. W. 1996. 'Sample size determination for correlation coefficient inference: Practical problems and
    practical solutions.' American Statistical Association 1996 Proceedings of the Section on Statistical
    Education, 240-245.
Cook, R. D. and Weisburg, S. 1999. Applied Regression Including Computing and Graphics. John Wiley and Sons,
    Inc.
Ostle, B. and Malone, L.C. 1988. Statistics in Research. Iowa State University Press. Ames, Iowa.
Zar, J. H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.
Fisher, R. A. 1921. 'On the probable error of a coefficient of correlation deduced from a small sample.'
    Metron, i (4), 1-32.

Summary Statements
A sample size of 161 produces a two-sided 95% confidence interval with a width equal to 0.060
when the estimate of Pearson's product-moment correlation is -0.900.

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

Plots Section
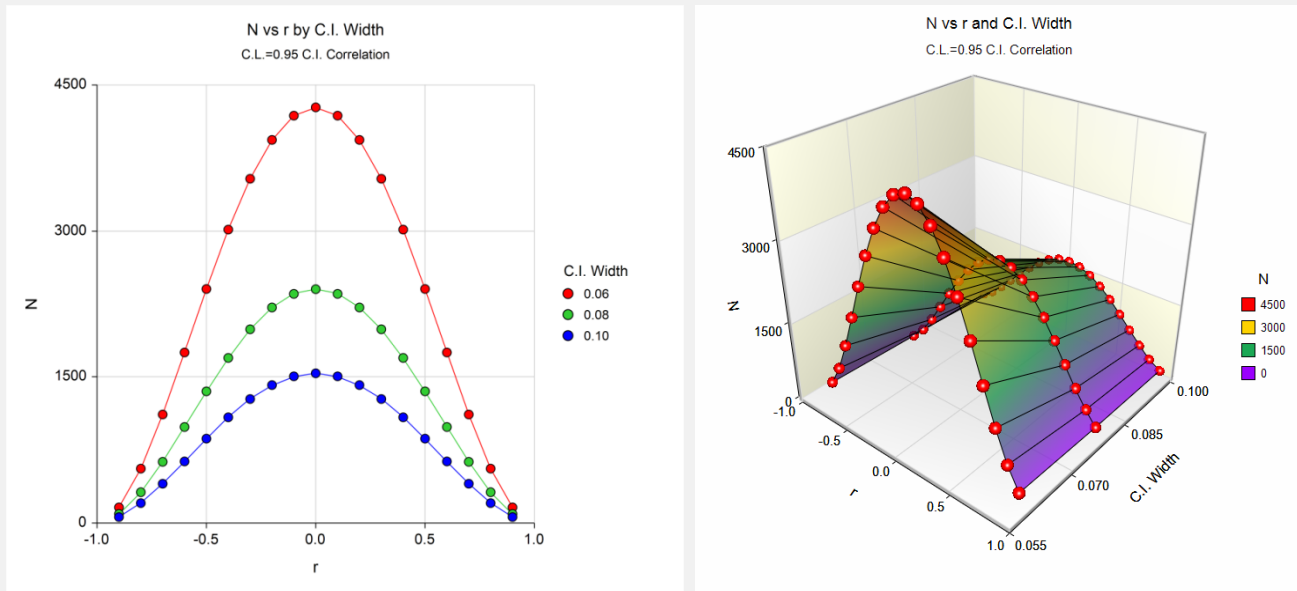
These plots show the sample size versus the sample correlation for the three confidence interval widths.

Example 2 – Validation using Bonett and Wright

Bonett and Wright (2000), page 26, give an extension table of sample sizes for two-sided confidence intervals for Pearson correlations when the confidence levels are 95% and 99%. When the sample correlation is 0.3 and the interval width is 0.2, they obtain sample sizes of 320 and 550, respectively.

Note that we checked our results with this table and found a view differences which are obvious typos.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Interval for Pearson Correlation** procedure window by expanding **Correlation**, then **Correlation**, then clicking on **Confidence Interval**, and then clicking on **Confidence Interval for Pearson Correlation**. 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 .... **Sample Size**
Interval Type .... **Two-Sided**
Confidence Level (1 – Alpha) **0.95 0.99**
Confidence Interval Width (Two-Sided).. **0.2**
r (Sample Correlation) **0.3**

Output

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

Numeric Results

Confidence Level	Sample Size N	Target Width	Actual Width	Sample Correlation r	C.I. Lower Limit	C.I. Upper Limit	Width if r = 0.0
0.950	320	0.200	0.200	0.300	0.197	0.397	0.219
0.990	550	0.200	0.200	0.300	0.197	0.397	0.219

PASS also calculated the sample sizes to be 320 and 550.