Chapter 801

Confidence Intervals 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 – α, 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 – α.

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

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**
Interval Type	**Two-Sided**
Confidence Level (1 - Alpha)	**0.95**
Confidence Interval Width	**0.06 0.08 0.10**
r (Sample Correlation)	**-0.9 to 0.9 by 0.1**

Output

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

Numeric Reports

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Confidence Interval Width			Sample Pearson's Correlation	Confidence Interval Limits	
		Target	Actual	If r = 0.0	r	Lower	Upper
0.95	161	0.06	0.06	0.309	-0.9	-0.926	-0.866
0.95	559	0.06	0.06	0.166	-0.8	-0.828	-0.768
0.95	1115	0.06	0.06	0.117	-0.7	-0.729	-0.669
0.95	1752	0.06	0.06	0.094	-0.6	-0.629	-0.569
0.95	2404	0.06	0.06	0.080	-0.5	-0.529	-0.469
0.95	3014	0.06	0.06	0.071	-0.4	-0.430	-0.370
0.95	3536	0.06	0.06	0.066	-0.3	-0.330	-0.270
0.95	3935	0.06	0.06	0.062	-0.2	-0.230	-0.170
0.95	4184	0.06	0.06	0.061	-0.1	-0.130	-0.070
0.95	4269	0.06	0.06	0.060	0.0	-0.030	0.030
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.

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

Summary Statements

A single-group design will be used to obtain a two-sided 95% confidence interval for a single Pearson product-moment correlation coefficient. The sample estimate of the Pearson correlation is assumed to be -0.9. To produce a confidence interval with a width of no more than 0.06, 161 subjects will be needed.

Dropout-Inflated Sample Size

Dropout Rate	Sample Size N	Dropout-Inflated Enrollment Sample Size N'	Expected Number of Dropouts D
20%	161	202	41
20%	559	699	140
20%	1115	1394	279
20%	1752	2190	438
20%	2404	3005	601
20%	3014	3768	754
20%	3536	4420	884
20%	3935	4919	984
20%	4184	5230	1046
20%	4269	5337	1068
.	.	.	.
.	.	.	.
.	.	.	.

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 confidence 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 confidence 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, 202 subjects should be enrolled to obtain a final sample size of 161 subjects.

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.

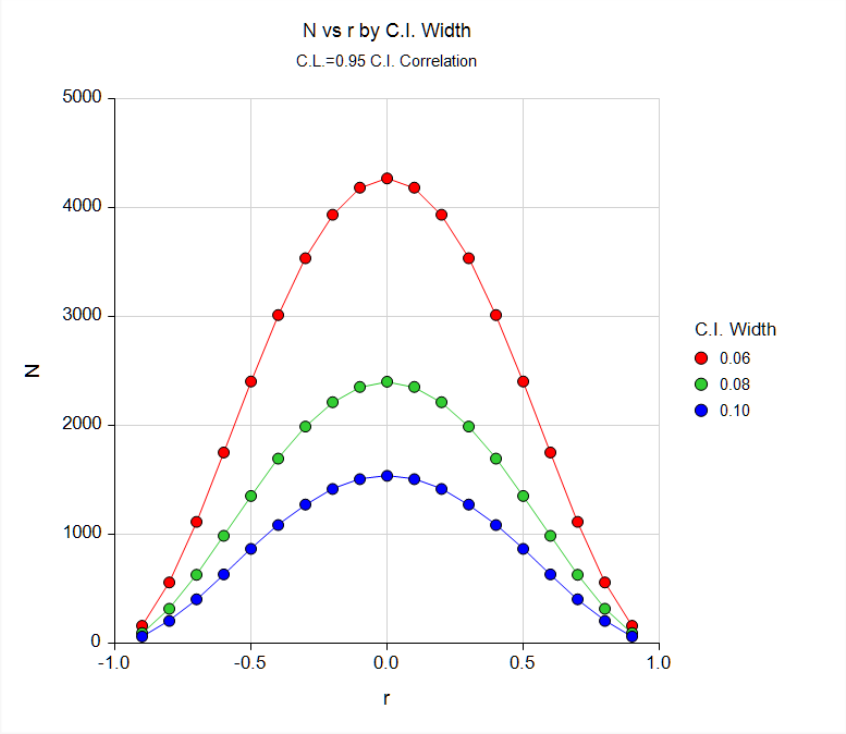
Zar, J. H. 1984. Biostatistical Analysis. Second Edition. Prentice-Hall. Englewood Cliffs, New Jersey.

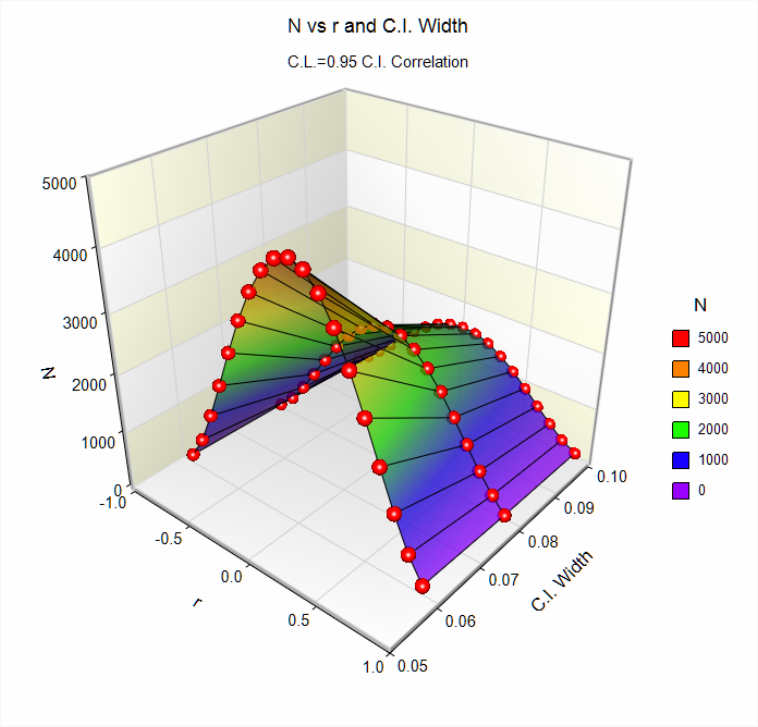
Fisher, R. A. 1921. 'On the probable error of a coefficient of correlation deduced from a small sample.' Metron, i (4), 1-32.

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

Plots Section

Plots





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

Example 2 – Validation using Bonett and Wright (2000)

Bonett and Wright (2000), page 26, give an extensive 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 few differences which are obvious typos.

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	**Sample Size**
Interval Type	**Two-Sided**
Confidence Level (1 - Alpha)	**0.95 0.99**
Confidence Interval Width	**0.2**
r (Sample Correlation)	**0.3**

Output

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

Numeric Results

Solve For: Sample Size
Interval Type: Two-Sided

Confidence Level	Sample Size N	Confidence Interval Width			Sample Pearson's Correlation r	Confidence Interval Limits	
		Target	Actual	If r = 0.0		Lower	Upper
0.95	320	0.2	0.2	0.219	0.3	0.197	0.397
0.99	550	0.2	0.2	0.219	0.3	0.197	0.397

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