2014 ▪ vol. 10 ▪ no. 1

DOI: 10.20982/tqmp.10.1.p029

An SPSS Macro to Compute Confidence Intervals for Pearson's Correlation

Bruce Weaver [a], Ray Koopman [b]

[a] Human Sciences Division, Northern Ontario School of Medicine, and Centre for Research on Safe Driving, Lakehead University

[b] Department of Psychology, Simon Fraser University

Abstract ▪ In many disciplines, including psychology, medical research, epidemiology and public health, authors are required, or at least encouraged to report confidence intervals (CIs) along with effect size estimates. Many students and researchers in these areas use IBM-SPSS for statistical analysis. Unfortunately, the CORRELATIONS procedure in SPSS does not provide CIs in the output. Various work-around solutions have been suggested for obtaining CIs for *rho* with SPSS, but most of them have been sub-optimal. Since release 18, it has been possible to compute bootstrap CIs, but only if users have the optional bootstrap module. The !rhoCI macro described in this article is accessible to all SPSS users with release 14 or later. It directs output from the CORRELATIONS procedure to another dataset, restructures that dataset to have one row per correlation, computes a CI for each correlation, and displays the results in a single table. Because the macro uses the CORRELATIONS procedure, it allows users to specify a list of two or more variables to include in the correlation matrix, to choose a confidence level, and to select either listwise or pairwise deletion. Thus, it offers substantial improvements over previous solutions to the problem of how to compute CIs for rho with SPSS..

Keywords ▪ Correlation, confidence interval, SPSS

bweaver@lakeheadu.ca

Introduction

One of the recommendations from the American Psychological Association's Task Force on Statistical Inference (TFSI) was that "interval estimates [i.e., confidence intervals] should be given for any effect sizes involving principal outcomes" (Wilkinson & APA Task Force on Statistical Inference, 1999). The report goes on to mention correlations specifically: "Provide intervals for correlations and other coefficients of association or variation whenever possible."

Many medical and health-related journals also require authors to report confidence intervals (CIs) for effect sizes, including correlations. For example, the website of the International Committee of Medical Journal Editors (ICMJE) includes Uniform Requirements for Manuscripts (URM) submitted to biomedical journals. The URM includes this statement: "When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals)" (International Committee of Medical Journal Editors, n.d.).

IBM-SPSS is a very popular statistical package that is used by many students and researchers in psychology, epidemiology, public health and related disciplines. Unfortunately, the CORRELATIONS procedure in SPSS does not include CIs as part of the output. Therefore, how to obtain the CI for *rho* is a rather frequently asked question in SPSS discussion forums.[1] One solution is to use the syntax found in this Technote on the IBM-SPSS website: http://www-01.ibm.com/support/docview.wss?uid=swg21478368. The Technote syntax works insofar as it provides correct results, but is less than ideal for many SPSS users. The main issue is that it does not work on raw data. Rather, users must compute the correlations for which they want CIs, and then supply the values of r and n, either by entering them into a dataset, or by using a DATA LIST command in a syntax window.

Two other methods that have been proposed were discussed in this thread from the SPSSX-L mailing list: http://spssx-discussion.1045642.n5.nabble.com/SPSS-syntax-for-CI-for-correlation-td1069009.html. One suggestion was to use Andrew Hayes' "rci" macro. It uses the SPSS MATRIX language to compute the 95% CI for *rho*. Unlike the SPSS Technote syntax, it has the advantage of working on raw data. But it does have

[1] The Greek letter *rho* (ρ) is used to symbolize the *parameter* corresponding to Pearson's r. CIs are for parameters, not for the statistics that estimate those parameters. Therefore, the CI for Pearson's correlation is properly described as the CI for *rho*, not the CI for r.

some limitations. First, it takes only two variables at a time. Therefore, if one wants CIs for a correlation matrix with three or more variables, the macro must be called once for each pair of variables. And to obtain listwise deletion, one must first set a filter to ensure that only cases with valid data for all of the variables are included. Second, Hayes' macro is hard coded to compute the 95% CI. Users comfortable with programming could modify the syntax easily enough to use other confidence levels, but it would be preferable to have a macro argument that specifies the desired confidence level. And finally, the URL for Hayes' macro shown in the SPSSX-L discussion is no longer active, so finding it may be difficult.

The IBM-SPSS Technote syntax and Hayes' rci macro both compute the CI for *rho* using the conventional method (see Beaulieu-Prévost, 2006 and Fisher, 1925). First, Fisher's *r*-to-*Z* transformation is applied to the sample correlation (see Equation 1).

$$Z_r = \operatorname{arctanh}(r) = (0.5)\log_e\left(\frac{1+r}{1-r}\right) \quad (1)$$

Then the limits of the CI are computed on the transformed scale (see Equation 2, where $z_{\alpha/2}$ is the critical value of z for a two-tailed α-level test).

$$100(1-\alpha)\%\text{ CI for } Z_\rho = Z_r \pm \frac{z_{\alpha/2}}{\sqrt{n-3}} \quad (2)$$

And finally, the lower and upper limits of the CI are transformed back to the original scale (Equation 3).[2]

$$r = \tanh(Z_r) = \frac{e^{Z_r} - e^{-Z_r}}{e^{Z_r} + e^{-Z_r}} \quad (3)$$

Another option discussed in the SPSSX-L thread was to perform simple linear regression using *z*-scores rather than the original raw scores—see this video, for example: http://www.youtube.com/watch?v=-dSoWqDyT4E. This approach is based on the fact that conceptually, Pearson's r is equal to the slope from a simple linear regression model using standardized X and Y values. And of course, the REGRESSION procedure *does* provide a confidence interval for the slope. The CI from regression with standard scores is $r \pm t_{\alpha/2}\sqrt{(1-r^2)/(n-2)}$. But those limits are necessarily wrong for two reasons: First, they are *symmetric* around *r*; second, they can exceed 1 and −1. The correct interval, on the other hand is always wider on the side of *r* toward zero than on the side away from zero; and it always has limits that fall within the range −1 to 1. Therefore, the CI for the slope from a simple linear regression with standard scores must not be used as the CI for *rho*.

The aforementioned video also shows how to compute bootstrap CIs for *rho* using SPSS Amos, a program for structural equations modeling (SEM). However, the main SPSS Statistics program can also compute bootstrap CIs for *rho*, provided one has access to the optional bootstrap module (first available in release 18). As suggested in the video, bootstrap CIs are legitimate, and may be preferable to CIs computed using standard asymptotic methods when distributional assumptions for the latter are too grossly violated. However, the bootstrap module is optional, so not all SPSS users have access to it. The macro described below, on the other hand, requires only the statistics base module and the ability to work with multiple datasets. Therefore, it is accessible to all SPSS users who have release 14 or later.

Description of the !rhoCI macro[3]

The limitations of currently available methods for computing CIs for *rho* motivated us to write a new, more flexible and user-friendly SPSS macro. The !rhoCI macro uses the /MATRIX OUT subcommand of the CORRELATIONS procedure to write the correlations and sample sizes to another dataset. Using matrix output from CORRELATIONS has several advantages: 1) The macro works on raw data; 2) one can have as many variables as desired; and 3) one can choose either LISTWISE or PAIRWISE deletion. (If LISTWISE deletion is selected, each correlation is computed using only

[2] The *tanh* in Equation 3 is the hyperbolic tangent, and the *arctanh* in Equation 1 is the *inverse* hyperbolic tangent (sometimes symbolized as *atanh* or $tanh^{-1}$). These functions are not directly available in SPSS, so it is customary to use expressions to the right in Equations 1 and 3 when computing the CI for *rho*. However, the idf.logistic and cdf.logistic functions, which are available in SPSS, can be used to compute *arctanh* and *tanh* as follows: arctanh(r) = .5×idf.logistic((1+r)/2,0,1); tanh(Z_r) = 2×cdf.logistic(2Z_r,0,1) − 1. More details about the idf.logistic and cdf.logistic functions can be found in the IBM-SPSS *Command Syntax Reference* manual.

[3] Macro names do not have to begin with an exclamation mark, so we could have called our macro rhoCI. We called it !rhoCI in accordance with a practice recommended by Raynald Levesque, the owner of the well-known SPSS Tools website (www.spsstools.net). Levesque's reason for starting macro names with an exclamation mark is that this makes it very easy for someone viewing a syntax file to recognize macros. Furthermore, all macros that follow this naming convention can be found quickly by searching for the explanation point.

Table 1 ■ Description of the five !rhoCI macro arguments.

Argument[a]	Mandatory?	Function
DataSetName	Yes	To specify the name of the dataset containing the raw data used for computing the correlations (e.g., DataSetName = RawData)
Vars	Yes	To provide a list of variables to be correlated; can use the TO convention (e.g., /Vars = v1 TO v10)
ConfidenceLevel	No	To specify the desired confidence level (e.g., /ConfidenceLevel = 99); if omitted, the default value is 95
ListWise	No	An indicator variable for LISTWISE deletion; e.g., /ListWise=1 to select listwise deletion, or /Listwise=0 to select pairwise deletion; if ListWise is omitted, the default value is 0 (pairwise)
CorrMat	No	An indicator variable for display of the correlation matrix produced by the CORRELATIONS procedure; e.g., /CorrMat=1 to display the correlation matrix, /CorrMat=0 to suppress its display; if CorrMat is omitted, the default value is 0 (no correlation matrix)

[a] Note that like other SPSS syntax, the macro arguments are not case-sensitive.

cases that have valid data for *all* variables in the list, so the sample size is the same for all correlations. If PAIRWISE deletion is selected, all cases with valid data for each pair of variables are used, so the sample size is not necessarily the same for each correlation.)

The macro has two mandatory and three optional arguments (see Table 1). The first mandatory argument, *DataSetName*, is the name of the dataset containing the raw data to be used for the computing the correlations (e.g., DataSetName = raw).[4] (Dataset names are shown in square brackets beside the file name in the top left corner of the Data Editor window.) The second mandatory argument, *Vars*, is a list of variables to include in the correlation matrix. It can be set to a list of individual variable names (e.g., Vars = educ paeduc maeduc speduc), or one can use the TO convention if the variables are contiguous in the data file (e.g.,Vars = educ TO speduc).[5] The third argument, *ConfidenceLevel*, allows the user to specify the desired confidence level (e.g., ConfidenceLevel = 95, or ConfidenceLevel = 99, etc.). The fourth argument, *ListWise*, is an indicator variable for LISTWISE deletion (ListWise = 1 to choose LISTWISE, ListWise = 0 to choose PAIRWISE). And the final argument, *CorrMat*, is an indicator for showing the matrix produced by the CORRELATIONS command in the output (CorrMat = 1 to show the correlation matrix, CorrMat = 0 to suppress it).[6] As noted above, the first two macro arguments (DataSetName and Vars) are mandatory. The other arguments can be omitted, but will then default to the following values: ConfidenceLevel = 95, ListWise = 0 and CorrMat = 0.

$$\begin{aligned} 100(1-\alpha)\%\text{ CI for }\rho = \tanh(Z_r \pm d) \\ = \frac{\tanh(Z_r) \pm \tanh(d)}{1 \pm \tanh(Z_r) \times \tanh(d)} \\ = \frac{r \pm \tanh(d)}{1 \pm r \times \tanh(d)} \end{aligned} \tag{4}$$

The macro computes the CI for *rho* as shown in Equation 4, where $d = z_{\alpha/2}\big/\sqrt{n-3}$. This simple variation on the usual method uses an algebraic shortcut based on the standard hyperbolic identity, tanh(a + b) ≡ [tanh(a) + tanh(b)] / [1 + tanh(a) × tanh(b)] (see Equation 4.5.26 in Abramowitz & Stegun, 1972). Note that the rightmost expression in Equation 4 works even when r = 1 or −1. Equation 1, on the other hand, breaks down in both of those cases: When r = 1, $(1+r)/(1-r)$ results in division by zero; and when r =

[4] It is necessary to have a dataset name associated with the file containing the raw data if one wishes that dataset to remain open for subsequent use after calling the macro.

[5] The key word WITH cannot be used, however, because CORRELATIONS cannot write matrix output in that case.

[6] We included the option of displaying the correlation matrix from CORRELATIONS so that readers who wish to do so can verify that the correlations and p-values in the macro's summary table are correct.

Pearson correlations with 95% confidence intervals*

	X	Y	r	Lower	Upper	p	n	Notes
1	age	age	1.000				497	
2	age	educ	-.049	-.136	.040	.281	496	
3	age	speduc	.034	-.105	.172	.634	200	
4	age	childs	.319	.238	.396	.000	496	
5	educ	age	-.049	-.136	.040	.281	496	
6	educ	educ	1.000				499	
7	educ	speduc	.598	.501	.680	.000	200	
8	educ	childs	-.256	-.336	-.172	.000	498	
9	speduc	age	.034	-.105	.172	.634	200	
10	speduc	educ	.598	.501	.680	.000	200	
11	speduc	speduc	1.000				201	
12	speduc	childs	-.120	-.254	.018	.089	201	
13	childs	age	.319	.238	.396	.000	496	
14	childs	educ	-.256	-.336	-.172	.000	498	
15	childs	speduc	-.120	-.254	.018	.089	201	
16	childs	childs	1.000				499	

* With PAIRWISE deletion.

Figure 1 ■ Output for Example 1. The three optional macro arguments were omitted, so they took their default values: ConfidenceLevel = 95, ListWise = 0 and CorrMat = 0. ListWise = 0 indicates use of pairwise deletion, and CorrMat =0 suppresses display of the correlation matrix generated by the CORRELATIONS procedure.

-1, $\log_e((1+r)/(1-r)) = \log_e(0)$, which is undefined.

Examples

Before the !rhoCI macro can be used, the macro definition (shown in the appendix) must be run. A convenient way to do this is via the INSERT FILE command, as follows:

```
INSERT FILE =
"C:\bw\SPSS\Syntax\Confint\rhoCI.SPS".
```

Users will have to modify the path to match where the *rhoCI.SPS* syntax file is stored on their computers. Note too that macro definitions remain active until you exit SPSS, so you do not have to run the macro definition before each call to the macro. Running it once per SPSS session will suffice.

The examples shown below were generated using syntax file *rhoCI_examples.SPS*, which can be downloaded from the first author's website (see the link provided in the appendix) and as part of the supplementary material on the journal's website. [7] We used only the first 500 records from the *survey_sample.sav* data file that can be found in the *Samples* sub-folder where IBM-SPSS is installed. Often, one uses all of the data in a file when computing correlations. We used only the first 500 records in this case because when we used all records, the sample size was so large that all *p*-values were less than .001, and all of the CIs were extremely narrow. By using only the first 500 records, we obtained a situation that is more typical of what most readers will see when analyzing their own data—i.e., some correlations are statistically significant, whereas others are not, and variation of width in the CIs is more easily seen.

The first example shown below uses only the first two mandatory macro arguments, DataSetName and Vars. Because the optional arguments are omitted, they take their default values: ConfidenceLevel = 95, ListWise = 0 and CorrMat = 0. The output for Example 1 is shown in Figure 1.

```
* Example 1.
!rhoCI DataSetName = raw
 /Vars = age educ speduc childs .
```

The second example computes uses the same variable list as Example 1, but computes 99% rather than 95% CIs, and it uses listwise deletion rather than the default pairwise deletion. The results for this macro call are shown in Figure 2.

```
* Example 2.
!rhoCI DataSetName = raw
 /Vars = age educ speduc childs
```

[7] The *rhoCI_examples.SPS* file includes other examples not shown here. For example, it shows how to compute bootstrap CIs for *rho*.

Pearson correlations with 99% confidence intervals*

	X	Y	r	Lower	Upper	p	n	Notes
1	age	age	1.000				199	
2	age	educ	.057	-.126	.237	.421	199	
3	age	speduc	.038	-.144	.219	.590	199	
4	age	childs	.324	.151	.478	.000	199	
5	educ	age	.057	-.126	.237	.421	199	
6	educ	educ	1.000				199	
7	educ	speduc	.597	.465	.702	.000	199	
8	educ	childs	-.172	-.344	.010	.015	199	
9	speduc	age	.038	-.144	.219	.590	199	
10	speduc	educ	.597	.465	.702	.000	199	
11	speduc	speduc	1.000				199	
12	speduc	childs	-.120	-.295	.063	.092	199	
13	childs	age	.324	.151	.478	.000	199	
14	childs	educ	-.172	-.344	.010	.015	199	
15	childs	speduc	-.120	-.295	.063	.092	199	
16	childs	childs	1.000				199	

* With LISTWISE deletion.

Figure 2. Output for Example 2. ConfidenceLevel was set to 99 and ListWise to 1. Thus 99% CIs are shown, and the sample size is the same for all correlations.

```
 /ConfidenceLevel = 99
 /ListWise = 1.
```

Example 3 demonstrates the use of the TO convention on the Vars argument. And by setting the CorrMat argument to 1, it tells the macro to display the usual correlation matrix generated by the CORRELATIONS procedure. ConfidenceLevel and ListWise ore omitted, so they take their default values (i.e., 95 and 0). The results are show in Figure 3.

```
* Example 3.
!rhoCI DataSetName = raw
 /Vars = educ TO speduc
 /CorrMat = 1 .
```

Example 4 demonstrates the use of the macro when the sample size is quite small. This was done to illustrate the various warning messages that may appear in the *Notes* field of the summary table when the sample size is small. The results are show in Figure 4.

```
* Example 4.
* Test with n = 4 records to show
* warnings in the "Notes" field.
COMPUTE f = ANY(id,10,20,30,40).
* f = 1 when id is 10, 20, 30 or 40;
* f = 0 otherwise.
* use only records where f = 1.
FILTER by f.
EXECUTE.
!rhoCI DataSetName = raw
 /Vars = age educ TO speduc .
* Turn off the filter to use all records
* in the file again.
USE ALL.
FILTER OFF.
```

As shown in Figure 4, the three possible warning messages that can appear in the *Notes* field are:

– $n < 3$: p and CI not computed
– $n < 4$: CI not computed
– $n < 10$: Normal approximation is poor

Pearson correlations with 95% confidence intervals*

	X	Y	r	Lower	Upper	p	n	Notes
1	educ	educ	1.000				499	
2	educ	paeduc	.379	.286	.464	.000	358	
3	educ	maeduc	.290	.201	.374	.000	433	
4	educ	speduc	.598	.501	.680	.000	200	
5	paeduc	educ	.379	.286	.464	.000	358	
6	paeduc	paeduc	1.000				359	
7	paeduc	maeduc	.644	.577	.703	.000	338	
8	paeduc	speduc	.411	.273	.533	.000	159	
9	maeduc	educ	.290	.201	.374	.000	433	
10	maeduc	paeduc	.644	.577	.703	.000	338	
11	maeduc	maeduc	1.000				434	
12	maeduc	speduc	.323	.185	.449	.000	179	
13	speduc	educ	.598	.501	.680	.000	200	
14	speduc	paeduc	.411	.273	.533	.000	159	
15	speduc	maeduc	.323	.185	.449	.000	179	
16	speduc	speduc	1.000				201	

* With PAIRWISE deletion.

Correlations

		Highest year of school completed	Highest year school completed, father	Highest year school completed, mother	Highest year school completed, spouse
Highest year of school completed	Pearson Correlation	1	.379**	.290**	.598**
	Sig. (2-tailed)		.000	.000	.000
	N	499	358	433	200
Highest year school completed, father	Pearson Correlation	.379**	1	.644**	.411**
	Sig. (2-tailed)	.000		.000	.000
	N	358	359	338	159
Highest year school completed, mother	Pearson Correlation	.290**	.644**	1	.323**
	Sig. (2-tailed)	.000	.000		.000
	N	433	338	434	179
Highest year school completed, spouse	Pearson Correlation	.598**	.411**	.323**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	200	159	179	201

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 3 ■ Output for Example 3. CorrMat was set to 1, so the correlation matrix generated by the CORRELATIONS procedure is displayed. ConfidenceLevel and ListWise were omitted, so they took their default values of 95 and 0 respectively.

Computation of the *p*-value entails using a *t*-distribution with degrees of freedom equal to $n-2$; and the standard error of Z_r (which is used in computing the CI for *rho*) has $\sqrt{n-3}$ in the denominator. Thus computation of the *p*-value requires $n \geq 3$ and computation of the CI for *rho* requires $n \geq 4$. The warning about the normal approximation being poor for $n < 10$ is based on Kirk's (2007) discussion of CIs for *rho*.

Summary

In many disciplines, including psychology, medical research, public health and epidemiology, authors are asked to report CIs for effect size estimates. IBM-SPSS does not provide CIs for Pearson's correlation in the output from its CORRELATIONS procedure. Various

Pearson correlations with 95% confidence intervals*

	X	Y	r	Lower	Upper	p	n	Notes
1	age	age	1.000				4	
2	age	educ	-.159	-.972	.947	.841	4	n < 10: Normal approximation is poor
3	age	paeduc	.012	-.960	.962	.988	4	n < 10: Normal approximation is poor
4	age	maeduc	-1.000			.011	3	n < 4: CI not computed
5	age	speduc	1.000				2	n < 3: p and CI not computed
6	educ	age	-.159	-.972	.947	.841	4	n < 10: Normal approximation is poor
7	educ	educ	1.000				4	
8	educ	paeduc	.971	.153	.999	.029	4	n < 10: Normal approximation is poor
9	educ	maeduc	.982			.121	3	n < 4: CI not computed
10	educ	speduc	1.000				2	n < 3: p and CI not computed
11	paeduc	age	.012	-.960	.962	.988	4	n < 10: Normal approximation is poor
12	paeduc	educ	.971	.153	.999	.029	4	n < 10: Normal approximation is poor
13	paeduc	paeduc	1.000				4	
14	paeduc	maeduc	.866			.333	3	n < 4: CI not computed
15	paeduc	speduc	1.000				2	n < 3: p and CI not computed
16	maeduc	age	-1.000			.011	3	n < 4: CI not computed
17	maeduc	educ	.982			.121	3	n < 4: CI not computed
18	maeduc	paeduc	.866			.333	3	n < 4: CI not computed
19	maeduc	maeduc	1.000				3	
20	speduc	age	1.000				2	n < 3: p and CI not computed
21	speduc	educ	1.000				2	n < 3: p and CI not computed
22	speduc	paeduc	1.000				2	n < 3: p and CI not computed
23	speduc	speduc	1.000				2	

* With PAIRWISE deletion.

Figure 4 ■ Output for Example 4. Only four records were used for this example. This was done to illustrate the three types of warning messages that can appear in the *Notes* field of the macro output table when sample sizes are small.

work-around methods for computing those CIs have been used in the past, but they are all sub-optimal in one way or another. Since release 18, it has been possible to compute bootstrap CIs for *rho*, but the bootstrap module is optional, and therefore not available to all SPSS users. The !rhoCI macro described in this article is accessible to all SPSS users who have release 14 or later. It allows users to specify a list of variables to include in the correlation matrix, to choose a confidence level, and to select either listwise or pairwise deletion. It also summarizes the output in a single table with one row per correlation. We hope it will prove useful to SPSS users, and that it will help lead to improved reporting of CIs for *rho*.

Authors' notes and acknowledgments

The first author thanks Sacha Dubois for reminding him of the usefulness of reading the fine manual (RTFM) when things are not working quite as desired. In particular, Sacha helped me realize that I needed to use variable name index rather than a simple numeric index in the VARSTOCASES command that is a crucial part of the !rhoCI macro. We also thank two anonymous reviewers for their thoughtful and constructive comments on an earlier draft of this article.

References

Abramowitz, M., & Stegun, I. A. (1972). *Handbook of mathematical functions with formulas, graphs, and mathematical tables (10th printing, with corrections)* National Bureau of Standards Applied Mathematics Series 55.

Beaulieu-Prévost, D. (2006). Confidence intervals: From tests of statistical significance to confidence intervals, range hypotheses and substantial effects. *Tutorials in Quantitative Methods for Psychology*, 2, 11.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

International Committee of Medical Journal Editors. (n.d.) *Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication*. Retrieved 31-May-2013, from http://www.ICMJE.org.

Kirk, R. E. (2007). *Statistics: An introduction* Wadsworth Publishing Company.
Wilkinson, L., & APA Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.

Appendix: The !rhoCI Macro Definition

Before using the !rhoCI macro, users must run the macro definition shown below. (As noted in the article, a convenient way to do that is via the INSERT FILE command.) After running the macro definition, the macro remains available until the SPSS session is terminated. This macro definition, the syntax used to generate the examples reported in the text of the article and a file containing full output for the examples are available as part of the supplementary material on the journal's website. They can also be downloaded from the first author's website. Use the first link below to download the macro definition, the second link to download syntax for the examples, and the third to see the full output from the syntax file.

– https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/rhoCI.txt?attredirects=0
– https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/rhoCI_examples.txt?attredirects=0
– https://sites.google.com/a/lakeheadu.ca/bweaver/Home/statistics/files/rhoCI_examples.pdf?attredirects=0&d=1

```
* ========================================================================
* File:  rhoCI.SPS .
* Date:  3-Jun-2013 .
* Author: Bruce Weaver, bweaver@lakeheadu.ca .
* ======================================================================== .
* This file contains an SPSS macro to generate confidence intervals
* for rho, where rho is the parameter corresponding to Pearson's r.
* Examples of how to use the macro can be found in syntax file
* "rhoCI_examples.SPS".
* The macro uses the following basic equations:
* Zr = arctanh[r] <==> r = tanh[Zr]
* d = z_alpha / sqrt[n-3]
* t = tanh[d]
*                        tanh[Zr] +- tanh[d]     r +- t
* ci = tanh[Zr +- d] = --------------------- = --------
*                       1 +- tanh[Zr]*tanh[d]   1 +- r*t
* The rightmost term in the last equation works even when r = +1 or -1.
* SPSS has no tanh and arctanh functions.
* HOWEVER, one can use IDF.LOGISTIC and CDF.LOGISTIC as follows:
* arctanh[r] = .5*ln((1+r)/(1-r)) = .5*idf.logistic((1+r)/2,0,1)
* tanh[d] = (exp(2*d)-1)/(exp(2*d)+1) = 2*cdf.logistic(2*d,0,1)-1
* SOURCES:
* http://people.math.sfu.ca/~cbm/aands/page_83.htm (see Equation 4.5.26).
* http://people.math.sfu.ca/~cbm/aands/intro.htm#001 .

* Note that the first two macro arguments, DataSetName and Vars
* are required. The others arguments can be omitted, in which
* case they will take the default values specified below.
* Note too that the macro syntax is NOT case sensitive.

DEFINE !rhoCI
 ( DataSetName = !CHAREND('/') /
  Vars = !CHAREND('/') /
  ConfidenceLevel = !DEFAULT(95) !CHAREND('/') /
  ListWise = !DEFAULT(0) !CHAREND('/') /
  CorrMat = !DEFAULT(0)!CMDEND ).
```

```
* Suppress all output.
OMS /DESTINATION VIEWER=NO /TAG='suppressall'.

DATASET DECLARE @corrmat.
DATASET ACTIVATE !DataSetName.

* Use MATRIX=OUT option of CORRELATIONS procedure
* to send the desired correlations to another dataset.
CORRELATIONS
 /VARIABLES=!vars
 /PRINT=TWOTAIL NOSIG
 /MATRIX=OUT('@corrmat')
!IF (!ListWise !EQ 1) !THEN
 /MISSING=LISTWISE.
!ELSE
 /MISSING=PAIRWISE.
!IFEND

DATASET ACTIVATE @corrmat.
RENAME VARIABLES (varname_ = X).
SELECT if not any(rowtype_,"MEAN","STDDEV").
EXECUTE.

* Use VARSTOCASES to restructure from a square correlation matrix
* to a long file format.

************************************************************************.
* Thanks to Sacha Dubois for pointing out the "Variable Name" Index
* option for VARSTOCASES, and for reminding that when all else fails,
* one should RTFM.
************************************************************************.

VARSTOCASES
 /ID=id
 /MAKE V FROM !Vars
 /INDEX=Y(V)
 /KEEP=ROWTYPE_ X
 /NULL=KEEP.

* Some of the data management steps differ depending on
* whether one selected LISTWISE or PAIRWISE deletion.

***** LISTWISE *****.
!IF (!ListWise !EQ 1) !THEN

IF rowtype_ EQ "CORR" r = V.
IF rowtype_ EQ "N" n = V.
IF missing(n) n = lag(n).
FORMATS n (f8.0).
EXECUTE.

!IFEND
```

```
***** PAIRWISE *****.
!IF (!ListWise !NE 1) !THEN

IF ($casenum EQ 1) OR
  ( rowtype_ EQ "CORR" and LAG(rowtype_) eq "N") MySorter = 1.
IF missing(MySorter) MySorter = LAG(MySorter) + 1.
FORMATS MySorter(f5.0).
EXECUTE.

SORT CASES by MySorter rowtype_ .
DO IF rowtype_ EQ "N".
+ COMPUTE r = Lag(V).
+ COMPUTE n = V.
END IF.
EXECUTE.
!IFEND

***** COMMON TO LISTWISE & PAIRWISE *****.

SELECT if not missing(r). /* Keep only 1 row per correlation.
STRING Notes (a45).
COMPUTE Notes = "".
EXECUTE.

DO IF X NE Y. /* Omit main diagonal.
+ DO IF n GT 3.
-  COMPUTE z = idf.normal(!ConfidenceLevel/200 + .5,0,1).
-  COMPUTE tanh_d = 2*cdf.logistic(2*z/sqrt(n-3),0,1) - 1.
-  COMPUTE Lower = ( r - tanh_d ) / ( 1 - r*tanh_d ) .
-  COMPUTE Upper = ( r + tanh_d ) / ( 1 + r*tanh_d ) .
+ END IF.
+ DO IF (n GT 2) and ABS(r) LT 1.
-  DO IF ABS(r) LT 1.
+   COMPUTE df = n-2.
+   COMPUTE t = r*SQRT(df/(1-r**2)).
+   COMPUTE p = cdf.t(-abs(t),df)*2.
-  ELSE IF ABS(r) EQ 1.
+   COMPUTE p = 0.
-  END IF.
+ END IF.
+ DO IF (n LE 2).
-  COMPUTE Notes = "n < 3: p and CI not computed".
+ ELSE if (n LE 3).
-  COMPUTE Notes = "n < 4: CI not computed".
+ ELSE if (n LT 10).
-  COMPUTE Notes = "n < 10: Normal approximation is poor".
+ END IF.
END IF.
FORMATS n df (f5.0) / r Lower Upper (f6.3).
!IF (!ListWise !EQ 1) !THEN
!LET !method = "LISTWISE"
!ELSE
```

```
!LET !method = "PAIRWISE"
!IFEND
!LET !title = !CONCAT("Pearson correlations with ",!ConfidenceLevel,"% confidence
intervals*")
!LET !footnote = !CONCAT("* With ",!method," deletion.")

OMSEND. /* Turn output on again.

* Suppress the "Case Processing Summary" for SUMMARIZE.

OMS
 /SELECT TABLES
 /IF COMMANDS=['Summarize'] SUBTYPES=['Case Processing Summary']
 /DESTINATION VIEWER=NO.
SUMMARIZE
 /TABLES= X Y r Lower Upper p n Notes
 /FORMAT=VALIDLIST NOCASENUM TOTAL
 /TITLE=!quote(!title)
 /FOOTNOTE=!quote(!footnote)
 /MISSING=VARIABLE
 /CELLS=NONE.

DATASET ACTIVATE !DataSetName.
DATASET CLOSE @corrmat.

* Display CORRELATIONS table for comparison with my table containing CIs.
OMSEND.

!IF (!CorrMat !EQ 1) !THEN

CORRELATIONS
 /VARIABLES=!vars
 /PRINT=TWOTAIL NOSIG
!IF (!ListWise !EQ 1) !THEN
 /MISSING=LISTWISE.
!ELSE
 /MISSING=PAIRWISE.
!IFEND
!IFEND
!ENDDEFINE.

* ======================================================================.
```

Citation

Weaver, B., & Koopman, R. (2014). An SPSS Macro to Compute Confidence Intervals for Pearson's Correlation. *The Quantitative Methods for Psychology, 10* (1), 29-39.

Received: 28/05/13 ~ Accepted: 4/06/13