**Comparing Correlated but Nonoverlapping Correlation Coefficients**

 Raghunathan, Rosenthal, and Rubin (1996, *Psychological Methods*, *1*, 178-183) discuss techniques for comparing correlated but nonoverlapping correlations. Let me first make the distinction between overlapping and nonoverlapping correlations.

Suppose that you wish to compare the correlation between one pair of variables with that between a second, **overlapping** pair of variables (for example, when comparing the correlation between one IQ test and grades with the correlation between a second IQ test and grades), and you have data on the same set of subjects for all three variables. **H∅: *ρWX*= *ρWY.*** Notice that variable W is common to both correlations. You can use Williams’ procedure explained in our textbook -- pages 261-262 of Howell, D. C. (2007), *Statistical Methods for Psychology*, 6th edition, Thomson Wadsworth -- or you can use Hotelling’s more traditional solution, available from Wuensch and elsewhere. Should you get seriously interested in this sort of analysis, consult this reference: Meng, Rosenthal, & Rubin (1992) Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*: 172-175.

Now suppose that you wish to compare the correlation between one pair of variables and a second **nonoverlapping** pair of variables, that is, **H∅: *ρAB*= *ρXY***. Assume that you have data on the same set of subjects for all four variables. Raghunathan et al. give several examples of research questions involving such nonoverlapping correlations. One is “comparing the correlation between two variables at two different points in time, for example, to learn whether, as children develop, their cognitive functions become less highly intercorrelated. Another of their examples involves comparing cross-lagged panel correlations, and their numerical example involves such correlations. Here is the correlation matrix they provided for a psychological (Psy) and a psychophysiological (PP) variable measured at both time one and time 2. Please note that the comparison they make is (corr between PSY1 and PP2) versus (corr between PP1 and Psy2), a so-called “crossed-lagged panel correlation).

|  |  |
| --- | --- |
|  | Variable |
| Variable | 1 | 2 | 3 | 4 |
| 1 = Psy, time1 | 1 | .45 | .53 | .38 |
| 2 = PP, time 1 |  | 1 | .25 | .31 |
| 3 = Psy, time 2 |  |  | 1 | .55 |
| 4 = PP, time 2 |  |  |  | 1 |

 Raghunathan et al. first present the Pearson-Filon statistic

For comparing *r*14 with *r*23, the Pearson-Filon statistic is:

, where

*k* = (*r*12 − *r*24\**r*14)\*(*r*34−*r*24\**r*23) + (*r*13 − *r*12\**r*23)\*(*r*24 − *r*12\**r*14)

 + (*r*12 − *r*13\**r*23)\*(*r*34 − *r*13\**r*14) + (*r*13 − *r*14\**r*34)\*(*r*24 − *r*34\**r*23)

Here I shall modify the notation better to fit the situation where we want to compare the correlation between two variables at time 1 with the correlation between the same two variables at time 2. Suppose that the two variables are EEG recordings at a pair of sites, one on the left, one on the right. Let ‘A’ represent the left site at time 1, ‘B’ the right site at time 1, ‘X’ the left site at time 2, and ‘Y’ the right site at time 2. Let ‘AB’ represent the correlation between left and right site at time 1, ‘XY’ the correlation between left and right site at time 2 ,and so on. We want to compare AB with XY. Dropping the ‘r’ symbols and replacing ‘1’ with ‘A,’ ‘4’ with ‘B,’ ‘2’ with ‘X,’ and ‘3’ with ’Y,’ we obtain:

, where

*k* = (AX − BX\*AB)\*(BY − BX\*XY) + (AY − AX\*XY)\*(BX − AX\*AB)

 + (AX − AY\*XY)\*(BY − AY\*AB) + (AY − AB\*BY)\*(BX − BY\*XY)

 Below I show the correlation coefficients used in the example presented by Raghunathan et al., adapted for use in the hypothetical research described in the preceding paragraph. My modification will result in the comparison between (corr between left and right at time 1) and (corr between left and right at time 2) in this modified correlation matrix being identical to the cross-lagged comparison presented by Raghunathan et al., (corr between PSY at time 1 and PP at time two) versus (corr between PP at time 1 and Psy at time 2) using their correlation matrix. This enabled me to verify that I had programmed the calculations correctly. The solution obtained from my modified correlation matrix is not identical to the solution you would get for comparing (PSY1 corr w PP1) with (PSY2 corr with PP2) using Raghunathan et al.’s correlation matrix.

|  |  |
| --- | --- |
|  | Variable |
| Variable | A | B | X | Y |
| A = left, time 1 | 1 | .38 | .45 | .53 |
| B = right, time 1 |  | 1 | .31 | .55 |
| X = left, time 2 |  |  | 1 | .25 |
| Y = right, time 2 |  |  |  | 1 |

*k* = (.45 − .31\*.38)\*(.55 − .31\*.25) + (.53 − .45\*.25)\*(.31 − .45\*.38)

 + (.45 − .53\*.25)\*(.55 − .53\*.38) + (.53 − .38\*.55)\*(.31 − .55\*.25) = .38105.

.

Raghunathan et al. demonstrate that a modified Pearson-Filon statistic, the *ZPF* statistic is superior to the PF statistic:



where *k* is the same quantity used in the *PF*, and *Z*AB and *Z*XY are the correlations being compared after Fisher’s *Z*r transformation, which is defined as:

.

 



 Raghunathan et al. present an approximation procedure that can be used to simplify the hand-calculation of the ZPF, but I find no need to use an approximation procedure, given that I have written an SAS program that will compute the exact PF and ZPF (it also gives you the value of *k* and the two-tailed *p* values for PF and ZPF).

 The program is named ZPF.SAS and can be found at <http://core.ecu.edu/psyc/wuenschk/SAS/SAS-Programs.htm>. It is written to allow you to make more than one comparison. Each comparison requires one line of data. Each data value must be separated from the next by a blank space. Each data line must contain the following information:

|  |  |
| --- | --- |
| Pair | An identification number for the pair of correlations being compared |
| N | The number of subjects. |
| AB | Correlation between variables A and B at time 1. |
| AX | Correlation between variable A at time 1 and A at time 2. |
| AY | Correlation between variable A at time 1 and B at time 2. |
| BX | Correlation between variable B at time 1 and A at time 2. |
| BY | Correlation between variable B at time 1 and B at time 2. |
| XY | Correlation between variables A and B at time 2. |

 In the program, pair 1 uses the data I borrowed from Raghunathan et al. If you read the article from which I borrowed these data, you will notice that my program produces answers slightly different from those in the article. My answers are better, since I computed with greater precision and I did not make a little rounding error that Raghunathan et al. made. Pair 2 uses data from an undergraduate statistics class where A is cumulative homework score at the time of the first exam, B is score on the first exam, X is cumulative homework score at the time of the second exam, and Z is score on the second exam. The SAS output shows that the correlation between homework score and exam score is significantly higher at the time of the second exam (*r* = .98) than at the time of the first exam (*r* = .64), *ZPF* (*N* = 10) = 3.77, *p* < .001.

 Bruce Weaver, at the Northern Ontario School of Medicine, has translated the SAS syntax to SPSS. You can find the syntax file [here](http://core.ecu.edu/psyc/wuenschk/SPSS/ZPF.sps).

Karl L. Wuensch
Dept. of Psychology
East Carolina University
Greenville, NC 27858
USA

May, 2012

[Return to the Stat Help Page](http://core.ecu.edu/psyc/wuenschk/StatHelp/StatHelp.htm)