Computer Programs for Comparing Independent Correlations

Programas informáticos para comparaciones entre correlaciones: Muestras independientes

Programas de informática para comparações entre correlações: Amostras independentes

N. Clayton Silver*
Department of Psychology, University of Nevada, Las Vegas, NV, USA
César Merino-Soto**
Instituto de Investigación de Psicología, Universidad de San Martín de Porres, Lima, Peru

ABSTRACT. There are a variety of techniques for testing the differences among independent correlations that are not available using the standard statistical software packages. Examples of these techniques for examining different hypotheses within the independent correlational realm are presented along with the output and interpretation from easily attainable, user-friendly, interactive software.

RESUMEN. Existe una variedad de técnicas para probar las diferencias entre correlaciones independientes que no están disponibles en los programas estadísticos familiares para el investigador. Se presentan ejemplos de estas técnicas para evaluar diferentes hipótesis dentro del contexto de correlaciones en muestras independientes, junto con programas informáticos interactivos, amigables y libre distribución.

RESUMO. Existe uma variedade de técnicas para provar as diferencias entre correlações independentes que não estão disponíveis nos programas estatísticos familiares para o investigador. Apresentam-se exemplos destas técnicas para avaliar diferentes hipóteses dentro do contexto de correlações em amostras independentes, junto com programas de informáticas interativos, amigáveis e livre distribuição.

Palavras chave: correlações, estatística, metodologia, amostras independentes, software.

There are occasions when researchers are interested in determining if two or more correlations are different. For these situations, there are various techniques that evaluate the hypotheses of correlation differences. However, these procedures have not been implemented in papers published in Spanish. Some reasons may explain this limitation. First, these techniques are not usually addressed in statistics books. In fact, a partial review of the books circulating in Latin America (e.g. Coolican, 2005; Guardia, Freixa, Peró, & Turbany, 2008) indicated that these methods (that will be presented in this manuscript) are not included in the contents.

Second, because these techniques are not readily available in the standard statistical analysis programs (e.g., SPSS), researchers have not implemented them because they either do not know about them or they are too mathematically complex. Finally, these techniques have their foundations in the statistical and quantitative methodology literature, which might not only be difficult for researchers to read, but also researchers would not gravitate toward. Some studies have used these procedures (e.g., Caldwell, Silver, & Strada, 2010; Caldwell-Gunes, Silver, Smith, & Norton, 2016; Meijs, Cillessen, Scholte, Segers, & Spijkerman, 2010; Merino, 2011a; Merino, Calderón, & Manzanares, in print; Merino & Grimaldo, 2015), but compared to the frequency with which studies use bivariate correlations, their application can still be considered emerging in the Hispanic context. Hence, this paper aims to familiarize the reader with a number of these techniques and describes the software that calculates them.

SPECIFIC PROCEDURES

Let us suppose a researcher is interested in determining if the correlations between the job satisfaction and salary scores are different between elementary and secondary school teachers. As these correlations come from different samples, they are considered independent. The z formula which would test this hypothesis, common to some statistics books, is the following:

$$z = \frac{z'_1 - z'_2}{\sqrt{\frac{1}{(N_1 - 3)} + \frac{1}{(N_2 - 3)}}}$$

In this formula, z' refers to the Fisher (1921) transformation of the correlation and N corresponds to the sample size in each group. If the correlation is .75 (N=100) in the elementary group and .30 (N=100) in the secondary group, then the result is: $z = 5.34$, $p < .0001$. This would indicate that, compared to the secondary school teachers, there is a significantly higher correlation between job satisfaction and salary for the elementary teachers. Although this is easy to calculate, it is rarely found in some standard statistical software programs (e.g., SPSS). One of the authors of this research developed a user-friendly, interactive program (INDCOR) to carry out this test. The user simply enters the correlations and the sample size of each group. The output (INDCOR.OUT) responds with the correlations, computed z, and its associated probability value.

Recently, Zou (2007) developed a method based on confidence intervals to examine the difference between two independent correlations. The premise is that the confidence interval provides the magnitude and precision of the relations, whereas the standard hypothesis test intertwines these two characteristics (Zou, 2007). A user-friendly interactive program called COMPCOR (Silver, Ullman, & Picker, 2015), carries out this calculation by requiring the sample size for each group, each group's correlation, and the confidence interval percentage (generally, 95%) from the user. The results appear in Figure 1.

Using Zou's method (2007) in the example, the confidence interval bounds are .2534 and .6542; because the confidence interval does not include zero, then the result obtained would be statistically significant (p < .05). Both methods are useful when there are two independent correlations.

However, let us suppose that we also wanted to include technical education and university instructors, then one could test the null hypothesis that the four independent population correlations are equivalent. As the previous z test only evaluated two groups at a time; six z tests would be needed in this instance. Of course, this would dramatically increase Type I error. To fight this problem, there are several omnibus tests to keep Type I error rate at the nominal level and yet still show reasonable statistical power. Rao (1970) described one of these procedures based on the Fisher transformation (1921) that is evaluated as X^2 with k-1

degrees of freedom (k is the number of independent correlations). The corresponding formula is found in Rao (1970) and Silver and Burkey (1991).

If we have 100 university and technical instructors with correlations of .60 and .40, respectively combined with the elementary and secondary school teacher correlations (.75 and .30 respectively), then the omnibus test would be statistically significant X^2 of 25.53, p < .0001. Because there are statistically significant differences among the correlations, a range test can be applied (Levy, 1976) to examine all pairwise differences. The program INDEPCOR (Silver & Burkey, 1991) computes both the omnibus test and the range test using the Student Newman-Keuls approach. Figure 2 demonstrates INDEPCOR's output with the data from the example used.

The range test tells us that there is a significantly higher correlation between job satisfaction and salary for the elementary teachers than for the secondary teachers, p < .0001, and technical instructors, p < .0005. Moreover, university instructors had a significantly higher correlation between the variables examined than did the secondary school teachers, p < .022. Similarly, another omnibus test is proposed, $X^2_{c(F)}$, , based on Neyman's (1959) C (a) statistic. The formula for this method can be found in Paul (1989). Paul's (1989) simulation results indicated that the test $X^2_{c(F)}$ kept a reasonable control of Type I error whereas the statistic X^2 based on Fisher's z transformation was conservative. Both procedures showed similar power

```
     FOR THE INDEPENDENT CORRELATIONS
      r1 = .7500 AND r2 = .3000

      THE SAMPLE SIZE FOR r1 = 100.0000
      THE SAMPLE SIZE FOR r2 = 100.0000

      THE .9500 CONFIDENCE INTERVAL FOR .7500 HAS A LOWER BOUND OF .6492 AND AN UPPER BOUND
     OF .8249

      THE .9500 CONFIDENCE INTERVAL FOR .3000 HAS A LOWER BOUND OF .1101 AND AN UPPER BOUND
     OF .4688


     *** TESTING THE DIFFERENCE BETWEEN INDEPENDENT rs ***
      THE LOWER BOUND IS .2534 AND THE UPPER BOUND IS .6542
      FOR THE .9500 CONFIDENCE INTERVAL

      IF THE INTERVAL CONTAINS 0, THEN IT IS NONSIGNIFICANT
```

Figure 1. *COMPCOR output, as per the example presented (Silver, Ullman, & Picker, 2015).*

(Paul, 1989). In this case, x2c(F) is 24.60, p <.0001. The program INCOR (Silver, Zaikina, Hittner, & May, 2008) calculates the test $X^2{}_{c(F)}$ together with an a posteriori range test (Levy, 1976), using the degrees of freedom of the Fisher-Hayter (Hayter, 1986) approach. The Fisher-Hayter range test tries to maintain an adequate control of Type I error rate while demonstrating reasonable power. Similar to INDEPCOR, the user must enter the number of groups, each correlation, and the sample size for each group. Figure 3 shows the INCOR output.

The conclusions are analogous to the ones obtained using INDEPCOR. In a new situation, let us suppose that a researcher wanted to examine the difference in the correlations between job satisfaction and salary, but remove the effects of the supervisor's effectiveness ratings from both variables for each of the groups. Fisher (1924) demonstrated that the distribution of a k order partial correlation based on n data points is equal to the bivariate correlation based on $n-k$ data points. Therefore, a modified version

```
THE GLOBAL TEST FOR EQUALITY = 25.5353
 AND IT HAS A PROBABILITY = .0000

 4. STEP DIFFERENCES
  1ST CORR 2ND CORR  RSTAT  PROB
 ( 1) .300 ( 4) .750 6.534 .00003

 3. STEP DIFFERENCES
  1ST CORR 2ND CORR  RSTAT  PROB
 ( 1) .300 ( 3) .600 3.778 .02141
 ( 2) .400 ( 4) .750 5.410 .00045

 2. STEP DIFFERENCES
  1ST CORR 2ND CORR  RSTAT  PROB
 ( 1) .300 ( 2) .400 1.124 .42715
 ( 2) .400 ( 3) .600 2.654 .06129
 ( 3) .600 ( 4) .750 2.756 .05206
```

Figure 2. *INDEPCOR results and application of the post hoc Student Newman-Keuls approach.*

```
THIS PROGRAM COMPUTES THE OMNIBUS C-ALPHA TEST
 DESCRIBED BY PAUL IN 1989 FOR TESTING THE DIFFERENCE AMONG
 MORE THAN TWO INDEPENDENT CORRELATIONS AND THE SUBSEQUENT
 RANGE TEST PUBLISHED BY LEVY IN 1976

 THE OMNIBUS TEST FOR EQUALITY = 24.6078
 AND IT HAS A PROBABILITY = .0000

 4. STEP DIFFERENCES
  1ST CORR  2ND CORR RSTAT  PROB
 ( 1) .300 ( 4) .750 6.534 .00002

 3. STEP DIFFERENCES
  1ST CORR 2ND CORR  RSTAT  PROB
 ( 1) .300 ( 3) .600 3.778 .02141
 ( 2) .400 ( 4) .750 5.410 .00045

 2. STEP DIFFERENCES
  1ST CORR  2ND CORR RSTAT  PROB
 ( 1) .300 ( 2) .400 1.124 .70634
 ( 2) .400 ( 3) .600 2.654 .14678
 ( 3) .600 ( 4) .750 2.756 .12662
```

Figure 3. *INCOR output.*

of Rao's (1970) formula could be used to evaluate the independent partial correlations. The modified formula, based on the work of Levy and Narula (1978), can be found in Silver, Wadiak, and Massey (1995). For example, the correlations for the variables of job satisfaction, supervisor ratings, and salary for each of the educational groups, is shown in Table 1.

The program INDPART allows for the comparison of partial correlations (Silver, Wadiak, & Massey, 1995) by computing an omnibus test followed by an a posteriori range test using the Student Newman-Keuls approach, similar to the procedure in INDEPCOR. The results are shown in Figure 4.

The partial correlation is computed for each group together with the F value and its significance level. In this example, for the elementary teachers and university instructors, there was a significant positive relationship between job satisfaction and salary after removing the effects of the supervisor ratings from each. Next, the omnibus test indicated that there was a statistically significant difference among the four groups of educators with regard to their partial correlations. Finally, the range test demonstrated that the correlation between salary and job satisfaction after removing the supervisor ratings was higher for the elementary teachers than for the other groups of educators ($ps < .0001$).

Table 1.
Example of correlations among the three variables.

	JOB SATISFACTION - SALARY	JOB SATISFACTION – SUPERVISOR RATINGS	SALARY – SUPERVISOR RATINGS
Elem. teacher (n=100)	.75	.40	.30
Tech. instructor (n=100)	.40	.50	.60
Univ. instructor (n=100)	.60	.70	.50
Sec. teacher (n=100)	.30	.40	.30

```
  PARTIAL CORRELATION FOR GROUP 1 = .7206
E F VALUE = 104.76000 WITH A PROBABILITY = .0000

E PARTIAL CORRELATION FOR GROUP 2 = .1443
E F VALUE = 2.06383 WITH A PROBABILITY = .1540

E PARTIAL CORRELATION FOR GROUP 3 = .4042
E F VALUE = 18.94532 WITH A PROBABILITY = .0000

E PARTIAL CORRELATION FOR GROUP 4 = .2059
E F VALUE = 4.29344 WITH A PROBABILITY = .0409

E GLOBAL TEST FOR EQUALITY = 34.4658
D IT HAS A PROBABILITY = .0000

  STEP DIFFERENCES
T PARTIAL CORR 2ND PARTIAL CORR RSTAT PROB
1) .1443        ( 4) .7206    7.4807 .00000

  STEP DIFFERENCES
T PARTIAL CORR 2ND PARTIAL CORR RSTAT PROB
1) .1443        ( 3) .4042    2.7761 .12287
2) .2059        ( 4) .7206    6.8584 .00001

  STEP DIFFERENCES
T PARTIAL CORR 2ND PARTIAL CORR RSTAT PROB
1) .1443        ( 2) .2059     .6223 .65667
2) .2059        ( 3) .4042    2.1538 .12858
3) .4042        ( 4) .7206    4.7045 .00096
```
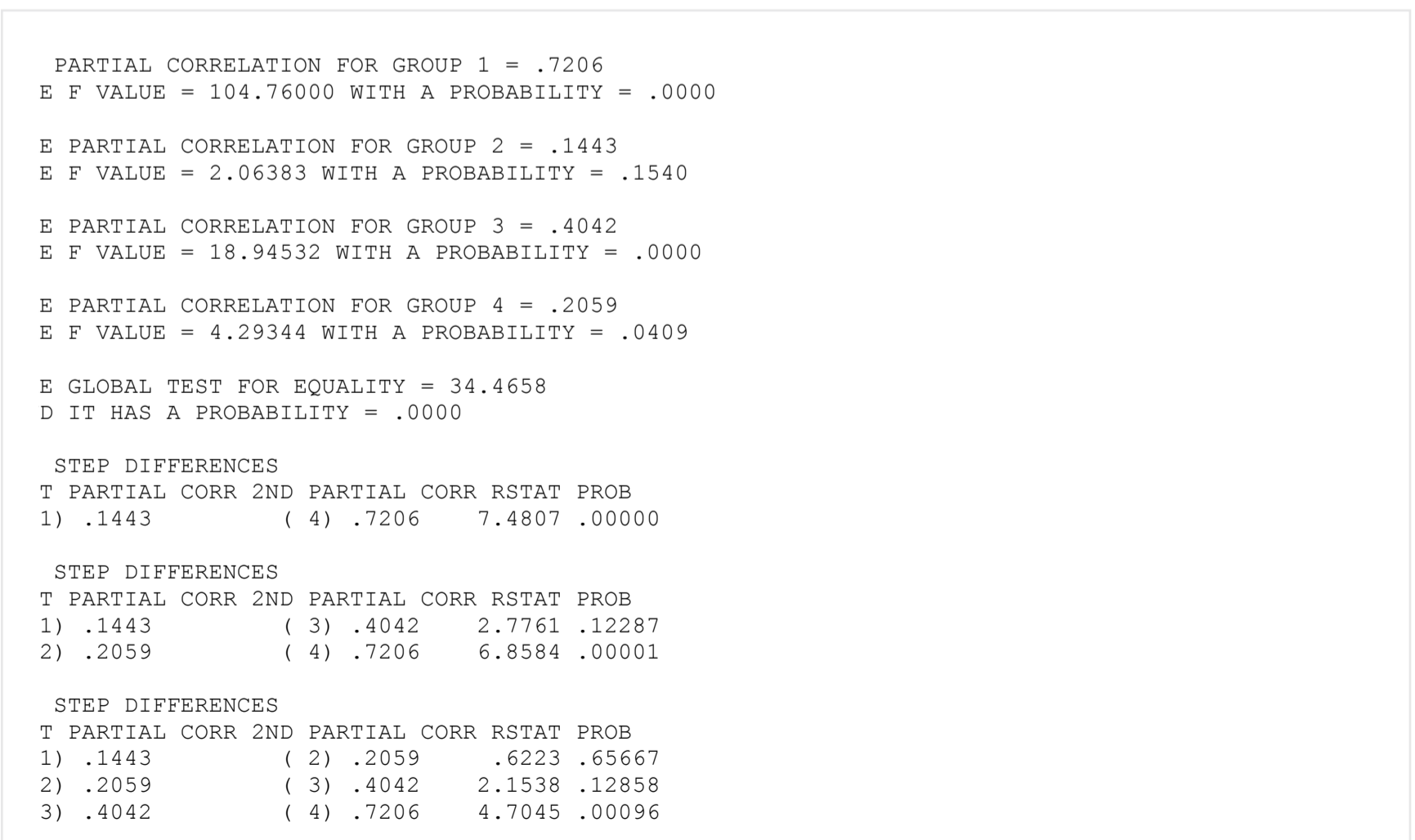
Figure 4. INDPART output.

FINAL COMMENTS

This article has considered various hypothesis tests for independent correlations. Although the differences between dependent correlations (correlations obtained from the same sample) could also be examined, that issue would best be served in a separate article. The methods presented expand the questions that may be considered in a research study beyond the simple testing of the null hypothesis that $\rho = 0$. One may also use the estimator q for the statistical tests presented to determine the magnitude of the effect (Cohen, 1992), by taking into consideration the following levels: trivial ($< \pm 0.20$), low ($\geqslant \pm 0.20$), moderate ($\geqslant \pm 0.50$), high ($\geqslant \pm 0.80$). These tools, coupled with the confidence intervals, may enhance the quantitative findings (Merino, 2011b). Finally, the programs cited are not readily available in standard statistical software programs (e.g., SPSS, SAS). These programs, that are user-friendly (i.e., need no programming skill from the user) and compatible with Windows technology, are distributed for free by writing to either author.

REFERENCES

Caldwell, R. M., Silver, N. C., & Strada, M. (2010). Substance Abuse, Familial Factors, and Mental Health: Exploring Racial and Ethnic Group Differences among African American, Caucasian, and Hispanic Juvenile Offenders. *American Journal of Family Therapy, 38*(4), 310-321. doi:10.1080/01926187.2010.493438

Caldwell-Gunes, R., Silver, N.C., Smith, K.M., & Norton, K. A. (2016). Racial and ethnic differences in factors related to substance use among adult female offenders: Implications for treatment. *Journal of Forensic Psychology Practice, 16*(1), 39-48. doi:10.1080/15228932.2016.1124595

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Coolican, H. (2005). *Métodos de investigación y estadística en psicología. [Research Methods and Statistics in Psychology].* México, DF: El Manual Moderno.

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

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron, 3,* 329-332.

Guardia, J., Freixa, M., Peró, M., & Turbany, J. (2008). *Análisis de datos en Psicología [Data Analysis in Psychology]* (2da ed.). Barcelona: Delta publicaciones.

Hayter, A. J. (1986). The maximum familywise error rate of Fisher's least significant difference test. *Journal of the American Statistical Association, 81*(396), 1000-1004.

Levy, K. L. (1976). A multiple range procedure for independent correlations. *Educational and Psychological Measurement, 36*(1), 27-31. doi:10.1177/001316447603600103

Levy, K. L., & Narula, S. C. (1978). Testing hypotheses concerning partial correlations: Some methods and discussion. *International Statistical Review, 46*(2), 215-218. doi:10.2307/1402814

Meijs, N., Cillessen, A. H. N., Scholte, R. H. J., Segers, E., & Spijkerman, R. (2010). Social intelligence and academic achievement as predictors of adolescent popularity. *Journal of Youth and Adolescence, 39*(1), 9373-9379. doi:10.1007/s10964-008-9373-9

Merino, C. (2011a). Test Gestáltico Visomotor de Bender Modificado y Test de Caras: Una evaluación de la validez de constructo. [Bender Gestalt Visualmotor Test and Caras Test: a exam of construct validity]. *Cuadernos de Neuropsicología, 5*(2), 129-143.

Merino, C. (2011b). Carta editorial [Letter to the Editor]. *Avances en Psicología Latinoamericana, 29*(2), 7-10.

Merino, C., Calderón, G., & Manzanares, E., (En prensa) Estudio comparativo del acuerdo y consistencia intercalificadores en el test gestáltico visomotor de Bender 2a edición. [Comparative study of inter-rater agreement and consistency in the Bender visual motor gestalt test, 2.nd edition]. *Revista Latinoamericana de Psicología.*

Merino, C., & Grimaldo, M. (2015). Análisis no paramétrico de rankings de tolerancia en comportamientos moralmente cuestionables. [Rankings of tolerance in morally debatable questionable behaviors: A nonparametric analysis]. *Revista Digital de Investigación en Docencia Universitaria, 9*(2), 3-48. doi:10.19083/ridu.9.448 http://dx.doi.org/10.1016/j.rlp.2015.09.011

Neyman, J. (1959). Optimal asymptotic tests of composite hypotheses. In U. Grenander (Ed.), *Probability and Statistics: The Harold Cramer Volume* (pp. 213-234). New York: Wiley.

Paul, S. R. (1989). Test for the equality of several correlation coefficients. *The Canadian Journal of Statistics, 17*(2), 217-227. doi:10.2307/3314850

Rao, R. (1970). *Advanced statistical methods in biometric research.* New York: Hafner.

Silver, N. C. (2009). A *computer program for comparing correlations using confidence intervals.* Unpublished manuscript.

Silver, N. C., & Burkey, R. T. (1991). A FORTRAN 77 program for testing the differences among independent correlations. *Educational and Psychological Measurement, 51* (3), 641-644. doi:10.1177/0013164491513011

Silver, N. C., Wadiak, D. L., & Massey, C. J. (1995). A Microsoft FORTRAN 77 program for testing the differences among independent first-order partial correlations. *Educational and Psychological Measurement, 55 (*2), 245-248. doi:10.1177/0013164495055002008

Silver, N. C., Ullman, J. R., & Picker, C. (2015). COMPCOR: A computer program for comparing correlations using confidence intervals. *Psychology and Cognitive Sciences, 1,* 26-28.

Silver, N. C., Zaikina, H., Hittner, J. B., & May, K. (2008). INCOR: A computer program for testing differences among independent correlations. *Molecular Ecology Resources, 8* (4), 763-764. doi:10.1111/j.1755-0998.2008.02107.x

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods, 12* (4), 399-413. doi:10.1037/1082-989X.12.4.399