

Sample size requirements for interval estimation of the strength of association effect sizes in multiple regression analysis

Gwonen Shieh
National Chiao Tung University

Abstract

Background: Effect size reporting and interpreting practices have been extensively recommended in academic journals when analyzing primary outcomes of all empirical studies. Accordingly, the sample squared multiple correlation coefficient is the commonly reported strength of association index in practical applications of multiple linear regression. **Method:** This paper examines the sample size procedures proposed by Bonett and Wright for precise interval estimation of the squared multiple correlation coefficient. **Results:** The simulation results showed that their simple method for attaining the desired precision of expected width provides satisfactory results only when sample sizes are large. Moreover, the suggested sample size formula for achieving the designated assurance probability is inaccurate and problematic. **Conclusions:** According to these findings, their sample size procedures are not recommended.

Keywords: Assurance probability, expected width, squared multiple correlation.

Resumen

Requisitos del tamaño de la muestra para la estimación por intervalo de la fuerza de asociación de tamaños de efecto en análisis de regresión.

Antecedentes: la práctica al presentar e interpretar el tamaño de efecto ha sido recomendada extensivamente en revistas académicas al analizar resultados primarios en estudios empíricos. En consecuencia, el coeficiente de correlación múltiple al cuadrado de la muestra es el índice de fuerzas de asociación que se presente con más frecuencia en aplicaciones prácticas de regresión lineal múltiple. **Método:** este trabajo examina el procedimiento del tamaño de la muestra que Bonett y Wright propusieron para una precisa estimación por intervalos de coeficiente de correlación múltiple al cuadrado. **Resultados:** el resultado de esta simulación señala que su método simple para alcanzar la deseada precisión de la amplitud esperada proporciona el resultado satisfactorio solamente cuando el tamaño de la muestra sea extensivo. Además, la fórmula del tamaño de la muestra sugerida para lograr la designada probabilidad garantizada es inexacta y problemática. **Conclusiones:** de acuerdo con estos descubrimientos, no se recomienda el procedimiento del tamaño de la muestra.

Palabras clave: probabilidad garantizada, anchura esperada, correlación múltiple al cuadrado.

Multiple regression analysis is one of the major methods of statistical analysis in applied research across many scientific fields. For descriptive purpose, the sample squared multiple correlation coefficient, usually denoted by R^2 , is commonly employed to assess the strength of association between the response variable and the predictor variables in many applications. See Bobko (2001) and Cohen et al. (2003) for operational guidelines and practical implications in areas of management and behavioral sciences. A primary concern of regression analysis is the conception of the two distinct scenarios of fixed (conditional) and random (unconditional) modeling formulations that ultimately lead to different inferential procedures. One must have a clear understanding of the respective setups and how they can be utilized before the issues involved in the construction of an appropriate regression model can be fully

explained. Notably, Sampson (1974) gave an excellent and thorough description of the two modeling formulations in which the random setting adopts the convenient assumption that all variables have a joint multivariate normal distribution. The procedures for power calculation, interval estimation, and sample size determination under the fixed regression models are well known; see Murphy and Myers (2004) and Smithson (2003) and the references therein for further details. However, the statistical properties of corresponding inferential procedures are more complex under the random model.

Although the underlying normality assumption provides a convenient and useful setup, the resulting probability density function of the sample squared multiple correlation coefficient R^2 is notoriously complicated in form. The complexity incurs numerous investigations to give various expressions, approximations and computing algorithms for the distribution of sample squared multiple correlation coefficient. See Johnson et al. (1995, Chapter 32) and Stuart and Ord (1994, Chapter 16) for further details. For the purpose of point estimation, it is well known that R^2 is a positively biased estimator of the population squared multiple correlation coefficient ρ^2 . To reduce the bias, several shrinkage estimators have been suggested in the literature. See Raju et al.

(1997), Shieh (2008), and Yin and Fan (2001) for further details. On the other hand, Helland (1987) suggested a simple approximate confidence interval using ordinary F distribution and the accuracy of the approximation is remarkably good for practical purposes. Moreover, exact confidence interval procedures were presented in Mendoza and Stafford (2001), Shieh (2006), Shieh and Kung (2007), and Steiger and Fouladi (1992). Unlike the approximate method, the exact approach employs an inversion technique of R^2 distribution and is called the “cumulative distribution function” pivotal method in Casella and Berger (2002, Section 9.2.3) and Mood, Graybill and Boes (1974, Section 4.2). Therefore, the calculations of exact confidence intervals for ρ^2 are methodologically and computationally more involved than those for the standard interval procedures of treatment contrasts in ANOVA. Consequently, the calculation of confidence intervals requires a special purpose computer program for performing the necessary computations of the probability distribution function of R^2 .

Instead of a direct accept-or-reject conclusion in a simple hypothesis test, confidence intervals are more informative about location and precision of the statistic, and they should be the best reporting strategy according to the recommendations of Wilkinson and the American Psychological Association Task Force on Statistical Inference (1999), as well as the *Publication Manual of the American Psychological Association* (2009). In addition, the editorial guidelines and methodological recommendations of several prominent educational and psychological journals stress that it is necessary to include some measures of effect size and confidence intervals for all primary outcomes. For example, see Alhija and Levy (2009), Dunst and Hamby (2012), Fritz, Morris and Richler (2012), Odgaard and Fowler (2010), and Sun, Pan and Wang (2010). The emphasis on reporting effect sizes and confidence intervals implies that researchers should plan studies not only to select practically meaningful effect size indices but also to have sufficiently accurate interval estimates of effect sizes. Thus it is prudent to facilitate this research practice by determining the necessary sample sizes to satisfy the desired precision of interval estimation in the planning stage of research design.

It follows from the general review of effect size estimates in Breaugh (2003), Ferguson (2009), Fern and Monroe (1996), Kirk (1996), Richardson (1996), and Vacha-Haase and Thompson (2004) that the squared multiple correlation coefficient is one the most commonly used strength of association measures in social science research. Accordingly, there is a considerable recent literature pertaining to the sample size determinations for precise interval estimation of squared multiple correlation coefficient within the linear regression framework. Due to the complexity of the exact probability density distribution of R^2 , the calculations of required sample size are extremely complicate to perform both efficiently and reliably. Therefore, Kelley (2008) and Krishnamoorthy and Xia (2008) utilized the simulation-based or trial-and-error approach to circumvent the difficulties in calculating the necessary sample sizes for adequate interval precision with respect to the control of expected width, and to the assurance probability of interval width within a designated value. Both computer programs and tabular sample sizes are provided in Kelley (2008) and Krishnamoorthy and Xia (2008) for constructing precise confidence intervals under the selected precision criterion. Although the difficulty of exact sample size computations has been avoided in the Kelley (2008) and Krishnamoorthy and Xia (2008), their suggested simulation procedures are still computationally intensive.

In view of the importance of accurate sample size formulas for a precise confidence interval of the squared multiple correlation coefficient and computational demands of the current methods, Bonett and Wright (2011) proposed a simple procedure of approximating the sample size requirement for obtaining a squared multiple correlation confidence interval with desired precision. The suggested sample size formula is derived from the approximate confidence interval of the squared multiple correlation coefficient using the asymptotic normal distribution of the sample multiple correlation coefficient. It is noted in Bonett and Wright (2011) that the resulting technique is attractive in its simplicity and is surprising accurate for controlling the expected width for the nearly-exact confidence intervals of Helland (1987). Moreover, they also presented a closed-form sample size formula for computing the necessary sample size that will yield a confidence interval that is not wider than the designated bound with a nominal assurance probability. Numerical illustration and practice recommendation are described to illustrate and enhance the practical usefulness of their procedures.

Despite the appealing advantage of simplicity for the sample size procedures of Bonett and Wright (2011), two obvious caveats in their arguments and expositions should be noted. First, it is well known that the distribution of sample squared multiple correlation is generally skewed. Hence, the equidistant confidence interval derived from the asymptotic normal distribution for transformation of sample squared multiple correlation is therefore presumably inappropriate and is not likely to be accurate. Second, the justification for the accuracy of their formulas is only based on the computed relative precision of confidence limits for the approximate interval procedure of Helland (1987) under selected values of model configurations. Accordingly, the lack of rigorous assessment of the accuracy of sample size formulas through comprehensive simulation study is an obvious drawback of the current explication in Bonett and Wright (2011). The actual performance of the suggested sample size procedures should be extensively evaluated before it can be adopted as a general methodology in practice. To this end, the article aims to conduct detailed numerical investigations to assess the adequacy of the sample size methods in Bonett and Wright (2011). We respectively suggest that our article serves as an updating and clarification of their recent work.

The remainder of the paper is organized in the following manner. In the next section, the fundamental results of the approximate interval procedure of Helland (1987) and sample size techniques for precise confidence intervals of Bonett and Wright (2011) are described. Then, Monte Carlo simulation studies were performed to appraise the accuracy the sample size formulas under a variety of model and precision configurations. The accuracy of the approximate techniques of Bonett and Wright (2011) are evaluated by the computed Helland's (1987) confidence intervals corresponding to the control of expected width, and to the tolerance probability of interval width within a designated value. Finally, some concluding remarks are provided.

Confidence intervals and sample size calculations

Consider the standard multiple linear regression model with criterion variable Y and p predictor variables (X_1, \dots, X_p) for N independent sets of these jointly multivariate normal variables. The sample squared multiple correlation coefficient R^2 is a

prevailing strength of association effect size measure for the population squared multiple correlation coefficient ρ^2 between the criterion variable and the set of predictor variables. For practical use, Helland (1987) presented an approximate interval estimation procedure for ρ^2 . Specifically, the approximate $100(1 - \alpha)\%$ confidence interval for ρ^2 is

$$(\hat{\rho}_L^2, \hat{\rho}_U^2) \tag{1}$$

where

$$\hat{\rho}_L^2 = \frac{(N-p-1)R^2 - (1-R^2)pF_L}{(N-p-1)\{R^2 + (1-R^2)F_L\}}, \hat{\rho}_U^2 = \frac{(N-p-1)R^2 - (1-R^2)pF_U}{(N-p-1)\{R^2 + (1-R^2)F_U\}}$$

F_L is the $100(1 - \alpha/2)$ percentile of the F distribution with v_L and $N - p - 1$ degrees of freedom, and $v_L = \{(N - p - 1)\hat{\rho}_L^2 + p\}^2 / \{N - 1 - (N - p - 1)(1 - \hat{\rho}_L^2)^2\}$, whereas F_U is the $100(\alpha/2)$ percentile of the F distribution with v_U and $N - p - 1$ degrees of freedom, and $v_U = \{(N - p - 1)\hat{\rho}_U^2 + p\}^2 / \{N - 1 - (N - p - 1)(1 - \hat{\rho}_U^2)^2\}$. Essentially, since F_L and F_U , or v_L and v_U also depend on the confidence limits $\hat{\rho}_L^2$ and $\hat{\rho}_U^2$, the optimal values of $\hat{\rho}_L^2$ and $\hat{\rho}_U^2$ in Equation 1 need to be found by a simple iterative search. It follows from the numerical comparison with the exact results, Helland (1987) concluded that the accuracy of the approximate interval estimates is surprisingly good because the error is negligible for practical purposes. Moreover, the approximate confidence intervals for ρ^2 can be computed with the SAS procedure PROC CANCORR (SAS Institute, 2011).

To ensure the precision of Helland’s (1987) confidence intervals for the squared multiple correlation coefficient, two methods were considered in Bonett and Wright (2011). The width of the $100(1 - \alpha)\%$ confidence interval $(\hat{\rho}_L^2, \hat{\rho}_U^2)$ is denoted by $W = \hat{\rho}_U^2 - \hat{\rho}_L^2$. One formula gives the minimum sample size, such that the expected confidence interval width $E[W]$ is within the designated bound. The other provides the sample size needed to guarantee, with a given assurance probability $P\{W \leq \omega\}$, that the width of a confidence interval will not exceed the planned range. Specifically, for a given value $\rho^2 = \tilde{\rho}^2$, the sample size N_{EW} needed for the expected width of a $100(1 - \alpha)\%$ confidence interval $(\hat{\rho}_L^2, \hat{\rho}_U^2)$ to fall within the designated bound ω is the minimum integer N such that

$$N \geq 16\tilde{\rho}^2 \{z_{\alpha/2} / \ln(\tilde{\epsilon})\}^2 + p + 2, \tag{2}$$

where $z_{\alpha/2}$ is the upper $100(\alpha/2)$ percentile of the standard normal distribution and $\tilde{\epsilon} = (1 - \tilde{\rho}^2 + \omega/2)/(1 - \tilde{\rho}^2 - \omega/2)$. On the other hand, for a given value $\rho^2 = \tilde{\rho}^2$, the sample size N_{Ap} required to guarantee with a given assurance probability $(1 - \gamma)$ that the width of a $100(1 - \alpha)\%$ confidence interval $(\hat{\rho}_L^2, \hat{\rho}_U^2)$ will not exceed the planned range ω is the smallest integer N such that

$$N \geq 16\tilde{\rho}_U^2 \{z_{\alpha/2} / \ln(\tilde{\epsilon})\}^2 + p + 2, \tag{3}$$

where $\tilde{\rho}_U^2 = 1 - \exp[\ln(1 - \tilde{\rho}^2) + z_\gamma \{4\tilde{\rho}^2/(N - p - 2)\}^{1/2}]$, and z_γ is the upper $100-\gamma$ percentile of the standard normal distribution. Bonett and Wright (2011) suggested repeating the calculations of N and $\tilde{\rho}_U^2$ two or three times for better approximation. Analytical arguments and theoretical justifications can be found in Bonett and Wright

(2011). It is clear from Equations 2 and 3 that the sample sizes required to attain the designated precision of expected width and assurance probability can be readily computed without complex algorithm. For illustration, consider the model and precision settings with $\rho^2 = 0.2$, $p = 5$, $1 - \alpha = 0.95$, and $\omega = 0.3$. It follows from Equation 2 that the sample size N_{EW} needed for the expected width of a 95% confidence interval $(\hat{\rho}_L^2, \hat{\rho}_U^2)$ to fall within the designated bound 0.3 is $N_{EW} = 93$ when the underlying population $\rho^2 = 0.2$. Likewise, the corresponding sample size is $N_{EW} = 43$ for the configurations of $\rho^2 = 0.7$, $p = 5$, $1 - \alpha = 0.95$, and $\omega = 0.3$. In contrast, the necessary sample size computed with the simulation-based approach of Kelley (2008, Table 4, p. 547) is 87 and 50, respectively. Whereas, Krishnamoorthy and Xia (2008, Table 2, p. 401) yielded the respective values 84 and 49 for $\rho^2 = 0.2$ and 0.7 according to their trial-and-error procedure. It can be readily seen from these results that there are discrepancies between the sample sizes computed by the different techniques of Bonett and Wright (2011), Kelley (2008), and Krishnamoorthy and Xia (2008). It is presumably that the computational intensive methods of Kelley (2008), and Krishnamoorthy and Xia (2008) are more accurate than the simplified formula of Bonett and Wright (2011). However, the prescribed exemplified sample size calculations for precise confidence intervals are not detailed enough to elucidate whether the trade of accuracy for simplicity is a wise bargain. To our best acknowledge, no research to date has examined the performance of Bonett and Wright’s (2011) simple formulas in greater detail. Consequently, it is worthwhile to clarify the issue surrounding the adequacy of their techniques because computational simplicity is not the only concern in sample size planning. For pedagogical and practical purposes, the accuracy of their sample size procedures is demonstrated in the next numerical investigation.

Numerical study

In order to demonstrate the features of Bonett and Wright’s (2011) sample size procedures in Equations 2 and 3, empirical examinations were performed for precise interval estimation of the squared multiple correlation coefficient. The numerical study was carried out in two stages. The first stage involved extensive sample size calculations for the two precision principles of expected width and assurance probability across a variety of model configurations. In the second stage, Monte Carlo simulation studies were conducted to assess the actual precision outcome for the suggested sample sizes under the design characteristics described in the first stage.

Sample size calculations

The determination of sample sizes needed for the chosen precision of the confidence intervals requires the specification of the confidence level, the magnitude of squared multiple correlation coefficient, and the number of predictor variables. It is evident that the influence of each of these components on the precision behavior not only differs but also depends on the concurrent impact of other factors. To provide a concise explication, the numerical assessments are specified by fixing the number of predictors $p = 5$ and confidence level $1 - \alpha = 0.95$, and varying the squared multiple correlation coefficient ρ^2 from 0.1 to 0.9 with an increment of 0.1 in the appraisals. Moreover, the interval bound $\omega = 0.2, 0.3$, and 0.4, and assurance probability $1 - \gamma = 0.90$ are selected for the two precision criteria of expected width and assurance probability.

These levels were chosen to reflect common sample sizes used in typical research settings. Accordingly, the computed sample sizes N_{EW} and N_{AP} with respect to the selected precision requirements are listed in Tables 1 and 2 for the expected width and assurance probability principle, respectively. As expected, the sample sizes vary with the parameter and precision specifications of ρ^2 and ω in the two tables. But it is evident from the reported results that the sample size is increasing with decreasing value of ω when all other factors are fixed. Also, the sample size is a concave function of ρ^2 with a maximum around 0.3 when all other factors are fixed. Since the bound of interval width is identical in Tables 1 and 2, the consistent magnitude difference in the sample sizes N_{EW} and N_{AP} indicates that it typically requires a larger sample size to meet the necessary precision of assurance probability than the control of a designated expected width. In other words, the sample sizes computed by the expected width consideration tend to be inadequate to guarantee the desired assurance level of interval width.

Simulation study

We then evaluate the accuracy of the sample size calculations through the following Monte Carlo simulation study. Under the computed sample sizes, parameter configurations, and precision settings described in Table 1 and 2, estimates of the true expected width or assurance probability are computed through Monte Carlo simulation of 10,000 independent data sets. Note that the exact probability density function of R^2 is extremely complex and therefore, it is difficult to generate a pseudo random variable with the common expression of R^2 in terms of the hypergeometric and beta functions. However, it is well known that there is a direct connection between the correlation model with multinormal variables and the multivariate normal regression model. Hence, inferences for ρ^2 can be accomplished with the usual F^* statistic:

$$F^* = \frac{R^2 / p}{(1 - R^2) / (N - p - 1)}$$

Additionally, there is an important correspondence between the derived F^* distribution and the following generic form suggested by Gurland (1968), namely

$$\frac{\{Z + (\Lambda \cdot W_1)^{1/2}\}^2 + W_2}{W_3}$$

where $\Lambda = \rho^2 / (1 - \rho^2)$, Z has the standard normal distribution $N(0, 1)$, $W_1 \sim \chi^2(N - 1)$, $W_2 \sim \chi^2(p - 1)$, $W_3 \sim \chi^2(N - p - 1)$ where $\chi^2(df)$ denotes a chi-square distribution with df degree(s) of freedom, and the random variables Z , W_1 , W_2 and W_3 are mutually independent. Consequently, the pseudo F^* random variable or, equivalently, the pseudo random variable $R^2 = pF^* / \{ (N - p - 1) + pF^* \}$, can be generated by employing the provided random number functions of standard normal and chi-square distributions in most modern statistical packages.

For each replicate of R^2 , the confidence limits and corresponding interval width of the two-sided 95% confidence intervals ($\hat{\rho}_L^2, \hat{\rho}_U^2$) of ρ^2 are calculated. Then the simulated expected width is the mean of the 10,000 replicates of interval widths, whereas the simulated assurance probability is the proportion of the 10,000 replicates whose values of interval width are less than or equal to the specified bound ω . The adequacy of the sample size procedure for precise interval estimation is determined by one of the following formulas: error = simulated expected width – nominal expected width or error = simulated assurance probability – nominal assurance probability. Both the simulated expected width and simulated tolerance probability along with the associated errors are summarized in Tables 1 and 2 as well.

Results and discussions

For the simulated results of expected width and assurance probability in Tables 1 and 2, there exists some disturbing behavior for the sample size procedures of Bonett and Wright (2011). First, the simulated expected width and corresponding error in Table 1 show that the computed sample sizes by Equation 2 are not

Table 1
Computed sample size and simulated expected width for the nearly exact 95% two-sided confidence interval of strength of association effect size ρ^2 when the number of predictors $p = 5$

ω	0.2			0.3			0.4		
	N_{EW}	Simulated E[W]	Error	N_{EW}	Simulated E[W]	Error	N_{EW}	Simulated E[W]	Error
0.1	131	0.1920	-0.0080	82	0.2390	-0.0610	56	0.2861	-0.1139
0.2	202	0.1955	-0.0045	93	0.2846	-0.0154	55	0.3612	-0.0388
0.3	230	0.1972	-0.0028	105	0.2899	-0.0101	61	0.3764	-0.0236
0.4	225	0.1982	-0.0018	102	0.2939	-0.0061	59	0.3853	-0.0147
0.5	194	0.1999	-0.0001	88	0.2984	-0.0016	50	0.3985	-0.0015
0.6	149	0.2017	0.0017	67	0.3058	0.0058	38	0.4164	0.0164
0.7	97	0.2067	0.0067	43	0.3250	0.0250	24	0.4658	0.0658
0.8	48	0.2237	0.0237	20	0.4088	0.1088	8	0.8872	0.4872
0.9	8	0.8458	0.6458	8	0.8458	0.5458	8	0.8461	0.4461

Table 2
Computed sample size and simulated assurance probability for the nearly exact 95% two-sided confidence interval of strength of association effect size ρ^2 when the number of predictors $p = 5$ and assurance probability $1 - \gamma = 0.9$

ω	0.2			0.3			0.4		
	N_{AP}	Simulated $P\{W \leq \omega\}$	Error	N_{AP}	Simulated $P\{W \leq \omega\}$	Error	N_{AP}	Simulated $P\{W \leq \omega\}$	Error
0.1	195	0.9745	0.0745	129	1.0000	0.1000	93	1.0000	0.1000
0.2	257	1.0000	0.1000	127	1.0000	0.1000	79	1.0000	0.1000
0.3	273	1.0000	0.1000	132	1.0000	0.1000	80	1.0000	0.1000
0.4	256	1.0000	0.1000	122	1.0000	0.1000	73	1.0000	0.1000
0.5	216	0.8987	-0.0013	101	0.9062	0.0062	60	1.0000	0.1000
0.6	163	0.6750	-0.2250	75	0.6155	-0.2845	44	0.5755	-0.3245
0.7	105	0.5355	-0.3645	48	0.4662	-0.4338	27	0.3718	-0.5282
0.8	51	0.3946	-0.5054	22	0.2800	-0.6200	8	0.0527	-0.8473
0.9	8	0.0494	-0.8506	8	0.0752	-0.8248	8	0.1048	-0.7952

uniformly accurate for the 27 combined settings of ρ^2 and ω . The absolute error is less than 0.01 for $\rho^2 \leq 0.7$ when $\omega = 0.2$, and for $0.4 \leq \rho^2 \leq 0.6$ when $\omega = 0.3$. Whereas all the resulting absolute errors are greater than 0.01 except for the single case of $\rho^2 = 0.5$ when $\omega = 0.4$. Moreover, it is noteworthy that the error is increasing with ρ^2 for each fixed value of three interval bounds. The distinct pattern reveals the potential deficiency in the sample size computation of Bonett and Wright (2011). Unfortunately, this undesirable property was not addressed in their numerical assessment.

On the other hand, the assurance performance in Table 2 also demonstrates the underlying drawback of the simple formula in Equation 3. Due to the highly skewed distribution of R^2 and the underlying metric of integer sample sizes, some of the simulated assurance probabilities are 1 for the calculated sample sizes. Hence the corresponding errors have the value of 0.1 for three, four and five cases when $\omega = 0.2, 0.3$ and 0.4 , respectively. The only two cases giving acceptable result are associated with the simulated assurance 0.9745 and 0.9062 for $N_{AP} = 195$ and 101 under the settings of $\rho^2 = 0.1$ and $\omega = 0.2$, and $\rho^2 = 0.5$ and $\omega = 0.3$, respectively. However, there are five, four, and three occurrences that the computed sample sizes do not guarantee the desired assurance probability level for $\omega = 0.2, 0.3$ and 0.4 , respectively. All but one of these achieved or simulated assurance probabilities are not substantially lower than the nominal probability 0.9, and the only exception is associated with the simulated value 0.8987 and error -0.0013 for $\rho^2 = 0.5$ and $\omega = 0.2$. Consequently, the sample size formula has the serious disadvantage of underestimating the necessary sample size for achieving the specified assurance level. Nonetheless, our extended calculations also confirm that this phenomenon continues to exist in other model configurations. Overall, the presented numerical evidence suggests that the sample size formulas of Bonett and Wright (2011) are not accurate enough to serve as a general method for computing the sample sizes for ensuring precise confidence intervals of squared multiple correlation.

Conclusions

There is a considerable recent literature pertaining to the illuminating applications of effect sizes and confidence intervals

in quantitative study. Accordingly, the desirability of achieving required precision in effect size estimation and the importance of sample size planning in constructing precise confidence intervals are repeatedly emphasized in applied research across many scientific fields. Researchers should become methodologically conscious that most rules of thumb for sample size calculations are inadequate to warrant the conclusion that the resulting confidence interval is of statistical precision and practical importance. Due to the computational complexity in sample size computation for precise interval estimation of strength of association effect sizes, Bonett and Wright (2011) presented alternative and simple sample size techniques in two distinct aspects. One method gives the minimum sample size, such that the expected confidence interval width is within the designated bound. The other provides the sample size needed to guarantee, with a given assurance probability, that the width of a confidence interval will not exceed the planned range. To justify the usefulness of the suggested methodology, numerical investigations were preformed here to evaluate the accuracy of their sample size procedures. In view of the conducted comprehensive empirical assessments, the approximate formulas of Bonett and Wright (2011) are not accurate enough to give optimal sample sizes in achieving the desired precision. Therefore, their procedures are not recommended for precise interval estimation of squared multiple correlation coefficient in multiple regression analysis.

In order to enhance the applicability of confidence intervals for strength of association effect sizes, in the present article, we present a comprehensive and update account of the corresponding sample size techniques. It is important to realize that the simplicity of an explicit formula may be appealing for inducing computational shortcuts but it does not involve all of the key factors in sample size calculation and, thus, is generally error prone. Without our appraisal and demonstration in this paper, applied researchers and practitioners will unknowingly adopt their sample size formulas for its advantage of simplicity. This may lead to miscomputed sample size, distorted precision performance and unsatisfactory research outcome for the planned study. Consequently, instead of the simplified formulas, it is prudent to consider a more sophisticated approach such as the prescribed simulation-based method.

References

- American Psychological Association (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Alhija, F.N.A., & Levy, A. (2009). Effect size reporting practices in published articles. *Educational and Psychological Measurement, 69*, 245-265.
- Bobko, P. (2001). *Correlation and regression: Applications for Industrial Organizational Psychology and Management* (2nd. ed.). Thousand Oaks, CA: Sage.
- Bonett, D.G., & Wright, T.A. (2011). Sample size requirements for multiple regression interval estimation. *Journal of Organizational Behavior, 32*, 822-830.
- Breaugh, J.A. (2003). Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management, 29*, 79-97.
- Casella, G., & Berger, R.L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury.
- Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Dunst, C.J., & Hamby, D.W. (2012). Guide for calculating and interpreting effect sizes and confidence intervals in intellectual and developmental disability research studies. *Journal of Intellectual & Developmental Disability, 37*, 89-99.
- Ferguson, C.J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practices, 40*, 532-538.
- Fern, E.F., & Monroe, K.B. (1996). Effect-size estimates: Issues and problems in interpretation. *Journal of Consumer Research, 23*, 89-105.
- Fritz, C.O., Morris, P.E., & Richler, J.J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*, 2-18.
- Gurland, J. (1968) A relatively simple form of the distribution of the multiple correlation coefficient. *Journal of the Royal Statistical Society, Series B, 30*, 276-283.
- Helland, I.S. (1987). On the interpretation and use of R^2 in regression analysis. *Biometrics, 43*, 61-69.
- Johnson, N.L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed., Vol. 2). New York: Wiley.
- Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43*, 524-555.
- Kirk, R. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746-759.
- Krishnamoorthy, K., & Xia, Y. (2008). Sample size calculation for estimating or testing a nonzero squared multiple correlation coefficient. *Multivariate Behavioral Research, 43*, 382-410.
- Mendoza, J.L., & Stafford, K.L. (2001). Confidence interval, power calculation, and sample size estimation for the squared multiple correlation coefficient under the fixed and random regression models: A computer program and useful standard tables. *Educational and Psychological Measurement, 61*, 650-667.
- Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.
- Murphy, K.R., & Myers, B. (2004). *Statistical Power Analysis: A simple and general model for traditional and modern hypothesis tests* (2nd ed.). Hillsdale, NJ: Laurence Erlbaum Associates.
- Odgaard, E.C., & Fowler, R.L. (2010). Confidence intervals for effect sizes: Compliance and clinical significance in the *Journal of Consulting and Clinical Psychology, 78*, 287-297.
- Raju, N.S., Bilgic, R., Edwards, J.E., & Flier, P.F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement, 21*, 291-305.
- Richardson, J.T.E. (1996). Measures of effect size. *Behavior Research Methods, Instruments, & Computers, 28*, 12-22.
- Sampson, A.R. (1974). A tale of two regressions. *Journal of the American Statistical Association, 69*, 682-689.
- SAS Institute (2011). *SAS/STAT User's Guide, Version 9.3*. Cary, NC: SAS Institute Inc.
- Shieh, G. (2006). Exact interval estimation, power calculation and sample size determination in normal correlation analysis. *Psychometrika, 71*, 529-540.
- Shieh, G. (2008). Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient. *Organizational Research Methods, 11*, 387-407.
- Shieh, G., & Kung, C.F. (2007). Methodological and computational considerations for multiple correlation analysis. *Behavior Research Methods, 39*, 731-734.
- Smithson, M. (2003). *Confidence intervals*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-140. Thousand Oaks, CA: Sage.
- Steiger, J.H., & Fouladi, R.T. (1992). R2: A computer program for interval estimation, power calculations, sample size estimation, and hypothesis testing in multiple regression. *Behavioral Research Methods, Instruments and Computers, 24*, 581-582.
- Stuart, A., & Ord, J.K. (1994). *Kendall's advanced theory of statistics* (6th ed., Vol. 1). New York: Halsted Press.
- Sun, S., Pan, W., & Wang, L.L. (2010). A comprehensive review of effect size reporting and interpreting practices in academic journals in education and psychology. *Journal of Educational Psychology, 102*, 989-1004.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51*, 473-481.
- Wilkinson, L., & the Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.
- Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods. *Journal of Experimental Education, 69*, 203-224.