PSYCHOMETRIKA—VOL. 65, NO. 1, 23–28
MARCH 2000

SAMPLE SIZE REQUIREMENTS FOR ESTIMATING PEARSON, KENDALL AND SPEARMAN CORRELATIONS

DOUGLAS G. BONETT

IOWA STATE UNIVERSITY

THOMAS A. WRIGHT

UNIVERSITY OF NEVADA, RENO

Interval estimates of the Pearson, Kendall tau-*a* and Spearman correlations are reviewed and an improved standard error for the Spearman correlation is proposed. The sample size required to yield a confidence interval having the desired width is examined. A two-stage approximation to the sample size requirement is shown to give accurate results.

Key words: sample size, interval estimation, correlation, rank correlation.

1. Introduction

Pearson, Kendall tau-*a* and Spearman correlations, which will all be denoted by the symbol θ, are used frequently in behavioral research. Although hypothesis testing is common, interval estimation may be more appropriate in applications where the magnitude of a correlation is of primary interest.

Cohen (1988), Desu and Raghavarao (1990), Odeh and Fox (1991), and several intermediate level statistics texts, such as Cohen and Cohen (1975) and Zar (1984), give formulas that can be used to determine the sample size required to test a hypothesis regarding the value of a population Pearson correlation with desired power. To date, sample size formulas to determine the sample size required for interval estimation of Pearson, Kendall tau-*a* and Spearman correlations are not available. Recently, Looney (1996) produced a table of sample sizes needed to obtain a 95% confidence interval for the Pearson correlation.

2. Confidence Intervals

To keep notation simple, let $\hat{\theta}$ denote both the estimator and the estimate of a population Pearson, Kendall tau-*a* or Spearman correlation. Define $\zeta = \tanh^{-1}\theta$ and $\hat{\zeta} = \tanh^{-1}\hat{\theta}$. The population variance of $\hat{\zeta}$ and its estimate are both denoted as $\sigma^2_{\hat{\zeta}}$. For the Pearson correlation, $\sigma^2_{\hat{\zeta}} \simeq 1/(n-3)$ for bivariate normal random variables (Fisher, 1925). For absolute values of Kendall correlations less than .8, $\sigma^2_{\hat{\zeta}} \simeq .437/(n-4)$ for any monotonic transformation of the bivariate normal random variables (Fieller, Hartley, & Pearson, 1957). We show that for absolute values of Spearman correlations less than .95, $\sigma^2_{\hat{\zeta}} \simeq (1+\hat{\theta}^2/2)/(n-3)$ for any monotonic transformation of the bivariate normal random variables.

Assuming asymptotic normality of $\hat{\theta}$, a large-sample $100\%(1-\alpha)$ confidence interval (Hahn & Meeker, 1991, p. 238) for θ may be defined as

$$\hat{\theta} \pm z_{\alpha/2}\left[\frac{d(\tanh\hat{\zeta})}{d\hat{\zeta}}\right]\sigma_{\hat{\zeta}}, \tag{1}$$

Requests for reprints should be sent to Douglas G. Bonett, Department of Statistics, Iowa State University, Ames, IA 50011.

0033-3123/2000-1/1997-0607-A $00.75/0

where $z_{\alpha/2}$ is the point on the standard unit normal distribution exceeded with probability $\alpha/2$. Note that $[d(\tanh\hat{\zeta})/d\hat{\zeta}]^2\sigma^2_{\hat{\zeta}} = (\text{sech}^2\,\hat{\zeta})^2\sigma^2_{\hat{\zeta}} = (1 - \tanh^2\hat{\zeta}^2)\sigma^2_{\hat{\zeta}} = (1-\hat{\theta}^2)^2\sigma^2_{\hat{\zeta}}$ is the approximate variance of $\hat{\theta}$ obtained by the delta method (Stuart & Ord, 1994, p. 351) which simplifies to $(1-\hat{\theta}^2)^2/(n-3)$, $.437(1-\hat{\theta}^2)^2/(n-4)$, and $(1+\hat{\theta}^2/2)(1-\hat{\theta}^2)^2/(n-3)$ for the Pearson, Kendall and Spearman correlations, respectively.

If $|\theta|$ is large and n is small, (1) may have a coverage probability that is quite different from $1-\alpha$. A better confidence interval, originally proposed by Fisher (1925) for the Pearson correlation, is defined as

$$\text{Lower Limit:} \quad \frac{[\exp(2L_1) - 1]}{[\exp(2L_1) + 1]}$$

$$\text{Upper Limit:} \quad \frac{[\exp(2L_2) - 1]}{[\exp(2L_2) + 1]}$$

where (2)

$$L_1 = .5[\ln(1+\hat{\theta}) - \ln(1-\hat{\theta})] - \frac{c(z_{\alpha/2})}{(n-b)^{1/2}}$$

$$L_2 = .5[\ln(1+\hat{\theta}) - \ln(1-\hat{\theta})] + \frac{c(z_{\alpha/2})}{(n-b)^{1/2}},$$

with $c = 1, (.437)^{1/2}$, and $(1+\hat{\theta}^2/2)^{1/2}$, $b = 3, 4$, and 3 for the Pearson, Kendall and Spearman correlations, respectively.

David (1938) recommends the use of (2) for Pearson correlations only if $n \geq 25$. The Fisher confidence interval for the Pearson correlation also assumes bivariate normality, and the effects of violating this assumption deserve careful consideration. Pearson (1929) concluded that "the normal bivariate surface can be mutilated or distorted to a remarkable degree" without affecting the sampling distribution of the Pearson correlation estimator. Subsequent simulations by Pearson (1931), Dunlap (1931), Rider (1932), and Gayen (1951) led to similar conclusions. However, Haldane (1949), Kowalski (1972), and Duncan and Layard (1973) have shown that the robust properties of the Pearson estimator apply only under independence and that marginal kurtosis can have a serious effect on the asymptotic sampling distribution of the Pearson estimator in the non-null case. If the assumption of bivariate normality cannot be justified, Kendall or Spearman correlations should be considered. The Kendall and Spearman correlations are attractive because (2) can be used to generalize from the sample to the population correlation for any monotonic transformation of bivariate normal variables.

As noted previously, the approximate variance of $\hat{\zeta}$ for a Kendall correlation is accurate only for $|\theta| < .8$. Under bivariate normality, a Kendall correlation is equal to $2/\pi$ times the inverse sine of the Pearson correlation so that a Kendall correlation of .8 corresponds to a Pearson correlation of about .95. Long and Cliff (1997) found that (2) works reasonably well for Kendall correlations if $n > 10$.

Fieller et al. (1957) claim that $\sigma^2_{\hat{\zeta}} \simeq 1.06/(n-3)$ for absolute values of Spearman correlations less than .8. We claim that $(1+\hat{\theta}^2/2)/(n-3)$ is a more accurate estimate of $\sigma^2_{\hat{\zeta}}$. The results of a computer simulation are summarized in the Appendix and provide support for our claim.

3. Sample Size Determination

The sample size required to obtain a $100(1-\alpha)\%$ Fisher confidence interval with a desired width (Upper Limit minus Lower Limit) can be obtained by first solving for n in (1). This gives a first-stage sample size approximation, denoted as n_o, equal to

$$n_o = 4c^2(1 - \tilde{\theta}^2)^2 \left(\frac{z_{\alpha/2}}{w}\right)^2 + b, \tag{3}$$

where w is the desired width of the Fisher confidence interval (2) and $\tilde{\theta}$ is a planning estimate of θ obtained from previous research or expert opinion. Round (3) up to the nearest integer and set $n_o = 10$ if $n_o < 10$. Note that $c^2 = 1 + \tilde{\theta}^2/2$ for the Spearman correlation.

Let w_o denote the width of the Fisher confidence interval (2) for a sample of size n_o and $\hat{\theta}$ set equal to $\tilde{\theta}$. Let n denote the sample size that yields a Fisher confidence interval having the desired width. Assume $w_o = k_1(n_o - b)^{-1/2}$ in the neighborhood of n_o and $w = k_2(n - b)^{-1/2}$ in the neighborhood of n where k_1 and k_2 are the constants of proportionality. For n_o close to n, assume $k_1 = k_2$ and define

$$\frac{w_o}{w} = \frac{(n - b)^{1/2}}{(n_o - b)^{1/2}}. \tag{4}$$

Solving for n gives a second-stage approximation to the required sample size

$$n = (n_o - b)\left(\frac{w_o}{w}\right)^2 + b, \tag{5}$$

which is rounded up to the nearest integer.

A planning estimate of θ is often obtained from a range of possible values based on expert opinion or confidence intervals from previous research. All other factors held constant, the sample size requirement is inversely proportional to $|\tilde{\theta}|$. Given a range of possible values for θ, some researchers will want to compute (5) for both minimum and maximum values of $\tilde{\theta}$ to obtain maximum and minimum sample size requirements.

4. Example

The following example illustrates the computation of (5) for a Pearson correlation. For $\tilde{\theta} = .8$, $w = .2$, and $\alpha = .05$, use (3) to compute $n_o = 4(1 - .8^2)^2(1.96/.2)^2 + 3 \simeq 52.8$ and round up to 53. Setting $n = n_o = 53$, $c = 1$, $b = 3$, and $\hat{\theta} = \tilde{\theta} = .8$ in (2) gives lower and upper Fisher interval limits of about .6758 and .880 with an interval width of $w_o \simeq .2042$. Use (5) to compute $(53 - 3)(.2042/.2)^2 + 3 \simeq 55.1$ and round up to $n = 56$. If a sample of size $n = 56$ is taken from the population and the Pearson correlation estimate is close to the planning estimate of .8, the Fisher confidence interval width should be close to the desired width of .2.

5. Accuracy of the Two-stage Sample Size Approximation

The sample size given by (5) approximates the correct sample size for the Fisher confidence interval. The correct sample size is defined as the smallest value of n which yields a Fisher confidence interval width that is less than or equal to the desired width. The accuracy of the two-stage approximation is evaluated by comparing the value obtained by (5) with the correct sample size for eight values of $\tilde{\theta}$, three values of w, and two values of α. The correct sample size was obtained by systematically incrementing n by 1 until the width of the Fisher confidence interval attained the desired width. The results are summarized in Table 1.

It can be seen from Table 1 that (5) gives a value of n that is exactly equal to the correct sample size or exceeds the correct sample size by a small amount. Although not shown in Table 1, (5) tends to overstate the correct sample size to a slightly greater degree if $|\tilde{\theta}| > .9$ and $w > 2(1 - |\tilde{\theta}|)$.

If the correct sample size must be obtained (e.g., in a commercial software package), the value given by (5) can be systematically decreased by 1 and the width of (2) can be checked at each step. Given the accuracy of (5), only one or two checks will be required in most cases.

TABLE 1.
Accuracy of Sample Size Approximation

			Pearson		Spearman		Kendall	
$\tilde{\theta}$	w	α	Eq. 5	Correct n	Eq. 5	Correct n	Eq. 5	Correct n
.10	.1	.05	1507	1507	1517	1517	661	661
.10	.1	.01	2601	2601	2614	2614	1139	1139
.10	.2	.05	378	378	382	382	168	168
.10	.2	.01	650	650	653	653	269	269
.10	.3	.05	168	168	169	169	77	77
.10	.3	.01	288	288	290	290	129	129
.30	.1	.05	1274	1274	1331	1331	560	560
.30	.1	.01	2198	2198	2297	2297	963	963
.30	.2	.05	320	320	334	334	143	143
.30	.2	.01	550	550	574	574	243	243
.30	.3	.05	143	143	149	149	65	65
.30	.3	.01	245	244	255	255	110	110
.40	.1	.05	1086	1086	1173	1173	448	448
.40	.1	.01	1874	1874	2024	2024	822	822
.40	.2	.05	273	273	295	295	122	122
.40	.2	.01	469	469	507	507	208	208
.40	.3	.05	123	123	132	132	57	57
.40	.3	.01	209	209	226	226	94	94
.50	.1	.05	867	867	975	975	382	382
.50	.1	.01	1495	1495	1682	1682	656	656
.50	.2	.05	219	219	246	246	99	99
.50	.2	.01	376	376	422	422	167	167
.50	.3	.05	99	99	111	111	46	46
.50	.3	.01	168	168	189	189	76	76
.60	.1	.05	633	633	746	746	280	280
.60	.1	.01	1091	1091	1287	1287	480	480
.60	.2	.05	161	161	189	189	73	73
.60	.2	.01	276	276	325	325	123	123
.60	.3	.05	74	74	86	86	35	35
.60	.3	.01	125	125	146	146	57	57
.70	.1	.05	404	404	503	503	180	180
.70	.1	.01	696	696	866	866	307	307
.70	.2	.05	105	105	129	129	49	49
.70	.2	.01	178	178	221	221	81	81
.70	.3	.05	49	49	60	60	24	24
.70	.3	.01	82	82	101	101	39	39
.80	.1	.05	205	205	269	269	93	93
.80	.1	.01	352	352	463	463	157	157
.80	.2	.05	56	56	72	72	27	27
.80	.2	.01	94	93	122	122	44	44
.80	.3	.05	28	28	36	35	15	15
.80	.3	.01	46	45	59	59	23	23
.90	.1	.05	63	62	87	86	30	30
.90	.1	.01	106	105	147	147	49	49
.90	.2	.05	21	20	28	27	12	11
.90	.2	.01	34	33	46	45	18	17
.90	.3	.05	13	12	18	16	8	8
.90	.3	.01	21	20	28	25	11	11

TABLE 2.
Empirical Coverage Rates for Spearman Correlations ($\alpha = .05$)

		Variance Estimate		
θ	n	A	B	C
.1	20	95.4	95.8	94.8
	50	95.7	95.5	95.1
	100	95.6	95.3	95.2
	200	95.3	95.0	95.0
.3	20	95.2	95.3	94.6
	50	95.2	95.3	94.9
	100	95.3	95.3	95.0
	200	95.3	95.2	95.1
.5	20	94.8	95.6	94.2
	50	95.0	95.6	95.0
	100	94.7	95.4	94.9
	200	94.8	95.5	95.1
.7	20	94.5	95.7	94.4
	50	94.1	95.5	94.4
	100	94.0	95.8	94.5
	200	94.0	95.8	94.9
.8	20	93.8	95.8	94.1
	50	93.2	95.5	94.1
	100	93.0	95.6	94.1
	200	93.1	95.6	94.4
.9	20	92.4	95.1	92.6
	50	92.3	95.5	93.6
	100	92.1	95.6	94.1
	200	92.1	95.6	94.2
.95	20	89.5	94.0	90.8
	50	90.3	94.5	92.3
	100	90.5	94.6	93.0
	200	91.0	95.5	93.7

Key: $A = 1.06/(n-3)$ (Fieller, et al., 1957)
$B = (1 + \hat{\theta}^2/2)/(n-3)$
$C = 1/(n-2) + |\hat{\zeta}|/(6n + 4n^{1/2})$ (Caruso & Cliff, 1997)

6. Conclusion

Testing the null hypothesis that a population correlation is equal to zero may not always be interesting—a confidence interval may be more informative as suggested by Gardner and Altman (1986), Schmidt (1996), and many others. When designing a study to estimate a Pearson, Kendall or Spearman correlation, the sample size required to obtain a Fisher confidence interval with the desired width will be a primary concern. An accurate sample size approximation can be obtained using (5).

Appendix

The accuracy of (2) for a Spearman correlation is investigated using three different estimates for $\sigma^2_{\hat{\zeta}}$: 1) $1.06/(n-3)$, 2) $(1 + \hat{\theta}^2/2)/(n-3)$, and 3) $1/(n-2) + |\hat{\zeta}|/(6n + 4n^{1/2})$. The third estimate was recently proposed by Caruso and Cliff (1997). All three variance estimates were determined empirically.

A computer simulation (20,000 random samples per condition) of the empirical coverage of a 95% Fisher confidence interval (2) for Spearman correlations under bivariate normality was performed for $\theta = [.1\ .3\ .5\ .7\ .8\ .9\ .95]$ and $n = [20\ 50\ 100\ 200]$. Column A of Table 2 shows that the empirical coverage rate with the Fieller et al. (1957) variance estimate is liberal for $\theta \geq .7$. Column B of Table 2 shows that the empirical coverage rate with $(1 + \hat{\theta}^2/2)/(n - 3)$ is close to 95% for $\theta \leq .9$ and slightly liberal for $\theta = .95$ with small n. Column C of Table 2 shows that the empirical coverage rate with the Caruso and Cliff (1997) variance estimate is close to 95% for $\theta \leq .7$ and has liberal tendencies for $\theta > .7$ that are most pronounced with small n. The results of Table 2 hold for any monotonic transformation of bivariate normal random variables.

References

Caruso, J.C., & Cliff, N. (1997). Empirical size, coverage, and power of confidence intervals for Spearman's rho. *Educational and Psychological Measurement, 57*, 637–654.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., & Cohen, P. (1975). *Applied multiple regression/Correlation analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

David, F.N. (1938). *Tables of the ordinates and probability integral of the distribution of the correlation coefficient in small samples*. Cambridge: Cambridge University Press.

Desu, M.M., & Raghavarao, D. (1990). *Sample size methodology*. Boston, MA: Academic Press.

Duncan, G.T., & Layard, M.W.J. (1973). A Monte-Carlo study of asymptotically robust tests for correlation coefficients. *Biometrika, 60*, 551–558.

Dunlap, H.F. (1931). An empirical determination of the distribution of means, standard deviations, and correlation coefficients drawn from rectangular distributions. *Annals of Mathematical Statistics, 2*, 66–81.

Fieller, E.C., Hartley, H.O., & Pearson, E.S. (1957). Tests for rank correlation coefficients. I. *Biometrika, 44*, 470–481.

Fisher, R.A. (1925). *Statistical methods for research workers*. London: Hafner Press.

Gardner, M.J., & Altman, D.G. (1986). Confidence intervals rather than p-values: Estimation rather than hypothesis testing. *British Medical Journal, 292*, 746–750.

Gayen, A.K. (1951). The frequency distribution of the product moment correlation in random samples of any size drawn from non-normal universes. *Biometrika, 38*, 219–247.

Hahn, G.J., & Meeker, W.Q. (1991). *Statistical intervals: A guide for practitioners*. New York, NY: Wiley.

Haldane, J.B.S. (1949). A note on non-normal correlation. *Biometrika, 36*, 467–468.

Kowalski, C.J. (1972). On the effects of non-normality on the distribution of the sample product moment correlation coefficient. *Applied Statistics, 21*, 1–12.

Long, J.D., & Cliff, N. (1997). Confidence intervals for Kendall's tau. *British Journal of Mathematical and Statistical Psychology, 50*, 31–41.

Looney, S.W. (1996). Sample size determination for correlation coefficient inference: Practical problems and practical solutions. *American Statistical Association 1996 Proceedings of the Section on Statistical Education*, 240–245.

Odeh, R.E., & Fox, M. (1991). *Sample size choice* (2nd ed.). New York, NY: Marcel Dekker.

Pearson, E.S. (1929). Some notes on sampling tests with two variables. *Biometrika, 21*, 337–360.

Pearson, E.S. (1931). The test of significance for the correlation coefficient. *Journal of the American Statistical Association, 26*, 128–134.

Rider, P.R. (1932). The distribution of the correlation coefficient in small samples. *Biometrika, 24*, 382–403.

Schmidt, F. (1996). Statistical significance testing and cumulative knowledge in Psychology: Implications for training of researchers. *Psychological Methods, 1*, 115–119.

Stuart, A., & Ord, J.K. (1994). *Kendall's advanced theory of statistics, Vol. 1, Distribution theory*. New York, NY: Halsted Press.

Zar, J.H. (1984). *Biostatistical analysis* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Manuscript received 20 OCT 1997
Final version received 23 FEB 1999