Another Look at the Robustness of the Product-Moment Correlation Coefficient to Population Non-Normality

R. Clifford Blair
University of South Florida
and
Stephanie Bear Lawson
Hillsborough County Schools

Historical Perspective

Studies of the effects of population nonnormality on the sampling distribution of the product-moment correlation coefficient date from the 1920's. For example, E. S. Pearson (1929), after studying samples taken from two "considerably non-normal distributions" concluded that "the normal bivariate surface can be mutilated and distorted to a remarkable degree without affecting the frequency distribution of r." (Quotations taken from Kowalski, 1972). Other studies of the period resulted in essentially the same conclusion. (See Kowalski 1972 for a review of this literature.) Later researchers tend to agree with Pearson's (1929) conclusion provided that $|\rho|$ is near zero. Norris and Hjelm (1961) state, "When there was essentially no correlation in the population, the shape of the sampling distributions for the product-moment correlation coefficients did not vary markedly as a function of nature or extent of nonnormality in the bivariate distribution. In general, these obtained sampling distributions conform very closely to their theoretical distributions." This point of view has become so popular that Kowalski (1972), after reviewing the literature, states "Everyone seems to agree that the distribution of r is quite robust to nonnormality when $\rho = 0$…."

Background For the Present Study

Bradley (1977) states "...the strength of the evidence for robustness [of the independent means *t* test] appears to derive partly from selectivity in investigating only the more familiar population shapes-which may be far less prevalent than their familiarity would suggest." Bradley (1977) goes on to argue that data gathered in actual research contexts may be sampled from populations that are much more radically nonnormal than are the familiar population shapes commonly used in robustness studies. As evidence for this contention, Bradley (1977) presents population distributions (or more accurately, large sample estimates of population distributions) that were generated in the contexts of psychological and medical experiments. Bradley (1968; 1976) has also shown that the *t* test is less robust to population nonnormality when sampling is from the first of the two aforementioned distributions than it is when sampling is from distributions traditionally used in robustness studies. As a final point, Bradley (1977) contends that distributions of the general type discussed above may occur in a wide variety of research contexts. Examples include studies dealing with (1) human vigilance, (2) the time required for rats to run a maze, (3) the time required to type letters, (4) annual medical payments to the aged, (5) social conformity and (6) the time required for pilots to operate ejection mechanisms in their aircraft, as well as other topics.

The Present Study

The purpose of the present study is to test the validity of the commonly held belief that population nonnormality has minimal effect on the sampling distribution of the product-moment correlation coefficient as long as the population correlation is zero. For reasons outlined above, it is believed that the Bradley (1977) distribution will provide a more stringent test of the robustness of r than has been provided by distributions found in many other studies of this type. A second reason for choosing this distribution for study lies in Bradley's (1977) assertion that this general type of distribution is quite prevalent in a variety of research contexts.

The Bradley Distribution

The Bradley distribution used in this study is essentially a mixture of three normal distributions. The first normal component of this distribution has mean 96.5, variance 3.1, and is sampled with probability .900. The second normal component has mean 130, variance 69.6 and is sampled with probability .095, while the third normal component has mean 160, variance 350.8 and is sampled with probability .005. The result is an "L" shaped distribution with a skew slightly greater than 3 and kurtosis of approximately 17.

Method of Study

Computer generated Monte Carlo methods were employed as the primary means of study. Via this technique, two independent samples of size n were drawn from the Bradley distribution described above. The Pearson product-moment correlation coefficient was calculated on the samples; this value was compared with the appropriate critical values and the decision as to whether or not the hypothesis of $\rho = 0.0$ was to be rejected at various levels of significance was recorded. This procedure was repeated 5,000 times for each sample size. Sample sizes studied were 5, 30, 50 and 100.

The Bradley distribution was simulated by means of the GGUBS and GGNML subroutines of the International Mathematical and Statistical Libraries (1980) and a FORTRAN program written specifically for this study. To verify the correctness of the FORTRAN program, all simulations were conducted on a normal distribution as well as the Bradley distribution.

Results

Table 1 shows the lower and upper tail cumulative probabilities of the product-moment correlation coefficients calculated on samples drawn from normal and Bradley distributions. Column headings indicate the population sampled, the sample size employed, the proportion of lower tail rejections of the hypothesis expected under ormal theory, the proportion of upper tail rejections of the null hypothesis expected under normal theory and the mean of the correlation coefficients.

As Table 1 indicates, proportions of rejections obtained with samples drawn from a normal distribution correspond closely with normal theory expectations at all significance levels for all sample sizes. On the other hand, samples drawn from the Bradley distribution produced results reasonably close to normal theory expectations only for lower tail tests with samples of size 5. Otherwise, lower tail tests resulted in very substantial deflations of Type I error rates while upper tail tests produced very sizable inflations of Type I error rates did not meet Bradley's (1978) liberal criterion for robustness. (According to this criterion, a statistic is said to be robust if in a given set of circumstances $.5 \quad < \rho < 1.5$ where p represents the actual Type I error rate and represents the Type I error rate obtained under normal theory.) It is noteworthy that even with samples as large as 100, obtained probabilities do not meet the liberal criterion for robustness except at the .05 level in the upper tail. It is also relevant to point out that increases in sample size may greatly exacerbate deviations of obtained probabilities from normal theory probabilities. (This latter point has been previously noted by Duncan and Layard, 1973.)

Conclusions

The exuberance of some previous researchers concerning the robustness of the product-moment correlation coefficient to population nonnormality should be tempered. As has been shown here, this coefficient may demonstrate alarmingly nonrobust properties when sampling is from populations that are sufficiently nonnormal. Also troubling is the fact that the usual remedy of increasing sample sizes may, in this instance, add to rather than ameliorate the problem.

TABLE 1

Lower And Upper Tail Cummulative Probabilities Of Product-Moment Correlation Coefficients Calculated On Samples Drawn From Normal And Bradley Distributions

		Lower Tail				Upper Tail				
Population	n	.005	.010	.025	.050	.050	.025	.010	.005	Avg. Corr.
Normal	5	.006	.010	.024	.049	.052	.028	.011	.006	.005
Bradley	5	.003	.007	.020	.042	.071	.046*	.030*	.024*	-.002
Normal	30	.006	.011	.025	.050	.047	.023	.009	.005	.001
Bradley	30	.000*	.000*	.000*	.004*	.080*	.058*	.041*	.031*	.000
Normal	50	.005	.010	.026	.049	.056	.027	.011	.007	.003
Bradley	50	.000*	.000*	.000*	.001*	.086*	.062*	.038*	.029*	.002
Normal	100	.006	.012	.027	.050	.055	.023	.010	.005	.000
Bradley	100	.000*	.000*	.000*	.003*	.068	.048*	.025*	.017*	.001

References

Bradley, J. V. (1968) Distribution-Free Statistical Tests. Englewood Cliffs, N. J.: Prentice-Hall.

Bradley, J. V. (1976) Probability; Decision; Statistics. Englewood Cliffs, N. J.: Prenctice-Hall.

Bradley, J. V. (1977) "A common situation conducive to bizarre distribution shapes," The American Statistician 31: 147-150.

Bradley, J. V. (1978) "Robustness?" British Journal of Mathematical and Statistical Psychology 31: 144-152.

Duncan, G. T. and Layard, M. W. J. (1973) "A Monte-Carlo study of asymptotically robust tests for correlation coefficients." Biometrickia 60: 551-558.

International Mathematical and Statistical Libraries (1980) Reference Manual. Houston, TX: Author.

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.

Norris, R. C. and Hjelm, H. F. (1961) "Nonnormality and product-moment correlation." Journal of Experimental Education 29: 261-270.

Pearson, E. S. (1929) "Some notes on sampling tests with two variables." Biometrikia 21: 337-360.