Psychol. Belg.
1982, XXII-2, 87-91.

University of Ghent
Laboratory for Experimental, Differential and Developmental Psychology

ON THE USE OF THE JACKKNIFE AND THE BOOTSTRAP FOR ESTIMATING A CONFIDENCE INTERVAL FOR THE PRODUCT-MOMENT CORRELATION COEFFICIENT

GEERT DE SOETE* & ANDRÉ VANDIERENDONCK
**Aspirant N.F.W.O.*

The purpose of the article is to demonstrate the use of the jackknife and the bootstrap for interval estimation of the product-moment correlation coefficient. After a general exposition, the two techniques are applied to a real data set. The relative merits of these methods as compared to the normal theory are discussed.

In psychological research the product-moment correlation coefficient, r, is a very popular and widely used statistic. In many data analyses correlations are routinely calculated and lay at the basis of inferences about the population correlation ρ.

One method for arriving at such inferences consists of estimating a confidence interval. This allows the data analyst to assert that ρ is contained in an interval with a predefined confidence or probability.

The purpose of the present article is to highlight some recently developed distribution-free methods of interval estimation and to compare them to the classical approach based on the assumption of bivariate normality. First, we present the classical approach and apply it to an example. Next, the jackknife and the bootstrap are described in general. In the following two sections these methods are used to estimate a confidence interval for ρ. Finally, in the discussion section, the three methods are compared, and some suggestions are formulated.

CONFIDENCE INTERVALS BASED ON NORMAL THEORY

As an example, we borrow a data set from Glass and Stanley (1970). The data consist of the raw scores of 40 high-school juniors on two 50-item tests, one on abstract reasoning and one on verbal reasoning. A scatter plot of these data is presented in Figure 1. The product-moment correlation between these two sets of scores amounts to 0.675.

Having computed a sample correlation, the data analyst is usually interested in the "true" value of the correlation, i.e., the value one would

The authors are indebted to Professor W. De Coster for providing the necessary computational facilities at the Laboratory for Experimental, Differential, and Developmental Psychology of the University of Ghent, Belgium.

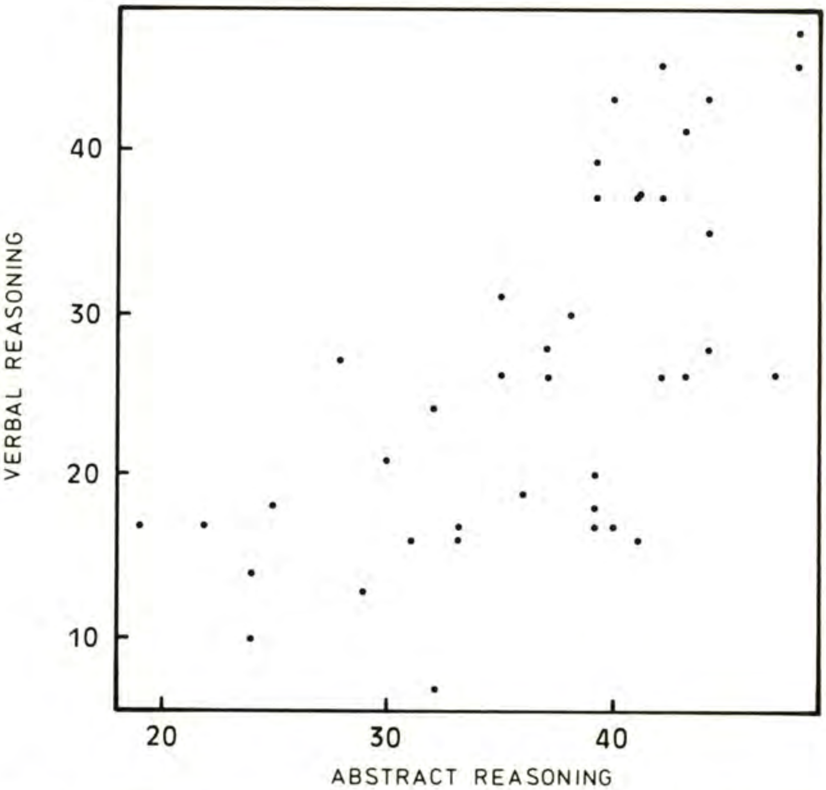

FIG. 1. SCATTER PLOT OF THE GLASS AND STANLEY DATA (Glass and Stanley, 1970, Table 7.1)

obtain if one took an infinitely large sample. The theory of inferential statistics (e.g., Glass and Stanley, 1970; Hays, 1973) allows one to infer a likely value for ρ. Such inferences are based on the sampling distribution of the statistic. This is the distribution of the value of the statistic over a very large set of samples of size n. If this sampling distribution is normal, and if it is possible to estimate its variance from the sample data, then it is quite straightforward to estimate a confidence interval.

In general, when $\hat{\theta}$ is a statistic, having a normal sampling distribution with mean θ and variance σ^2, the quantity $z = (\theta - \hat{\theta})/\sigma$ is contained within the interval $(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}})$ with probability $1-\alpha$, where $z_{\frac{\alpha}{2}}$ is defined as the standard normal deviate corresponding to $1-\frac{\alpha}{2}$. The confidence interval is then

$$\text{Prob}(-z_{\frac{\alpha}{2}} \leqslant (\theta - \hat{\theta})/\sigma \leqslant z_{\frac{\alpha}{2}}) = 1 - \alpha,$$

or

$$\text{Prob}(\hat{\theta} - z_{\frac{\alpha}{2}}\sigma \leqslant \theta \leqslant \hat{\theta} + z_{\frac{\alpha}{2}}\sigma) = 1 - \alpha.$$

In the case of the correlation coefficient, it is not possible to apply this theoretical scheme directly, because the sampling distribution of r is not normal. But it is possible to transform r to a new variable z_r,

$$z = \tanh^{-1}(r) = 0.5 \ln[(1+r)/(1-r)],$$

whose sampling distribution is approximately normal for almost any value of ρ, given that the sample is drawn from a bivariate normal distribution. The variance of this sampling distribution depends almost exclusively on the number of observations:

$$\mathrm{Var}(z_r) \approx 1/(n-3),$$

such that the estimate of the standard error of z_r based on the normal theory becomes

$$s_N = 1/\sqrt{n-3}.$$

The 99% confidence interval for z_ρ for the data depicted in Figure 1 is

$$0.819 \pm 0.424.$$

Applying the inverse transformation[1] tanh to the bounds of the confidence interval yields a 99% confidence interval for ρ

$$\mathrm{Prob}(0.375 \leqslant \rho \leqslant 0.846) = 0.99.$$

Calculating the 95% and 68% confidence intervals proceeds in a similar way. The results are summarized in Table 1.

TAB. 1. ESTIMATES OF THE STANDARD ERROR OF z_r AND SELECTED CONFIDENCE INTERVALS FOR THE GLASS AND STANLEY DATA BASED ON THE NORMAL THEORY, THE JACKKNIFE AND THE BOOTSTRAP

	normal theory		jackknife		bootstrap	
estimated standard error of z_r	0.164		0.128		0.124	
68% confidence interval	0.575	0.754	0.598	0.739	0.606	0.738
95% confidence interval	0.460	0.815	0.508	0.792	0.531	0.792
99% confidence interval	0.375	0.846	0.441	0.822	0.482	0.818

The example demonstrates the usefulness of the normal theory in statistics. We only have to assume that the sample is drawn from a bivariate normal distribution to obtain a $(1-\alpha)100\%$ confidence interval for the population correlation in a quite straightforward way. However, the assumption we had to make is a strong one, which is difficult to test when the sample size is small. In many practical situations, one can doubt its validity. This seems also to be true for the data plotted in Figure 1. Like many theoretical multivariate

[1] The inverse of Fisher's r-to-z transformation is $\tanh(z) = (e^{2z}-1)/(e^{2z}+1)$.

distributions, the bivariate normal distribution requires the marginal distributions to be identical up to location and scale parameters. One way to assess how well this requirement is fulfilled, is to plot the ordered standardized observations on the first variable against the ordered standardized observations on the second variable, yielding a standardized component probability plot (Gnanadesikan, 1972, 1977). If the two marginal distributions are identical (up to origin and scale), the points will show up as a straight line through the origin with unit slope. A standardized component probability plot for the Glass and Stanley data is presented in Figure 2. The systematic curvature of the

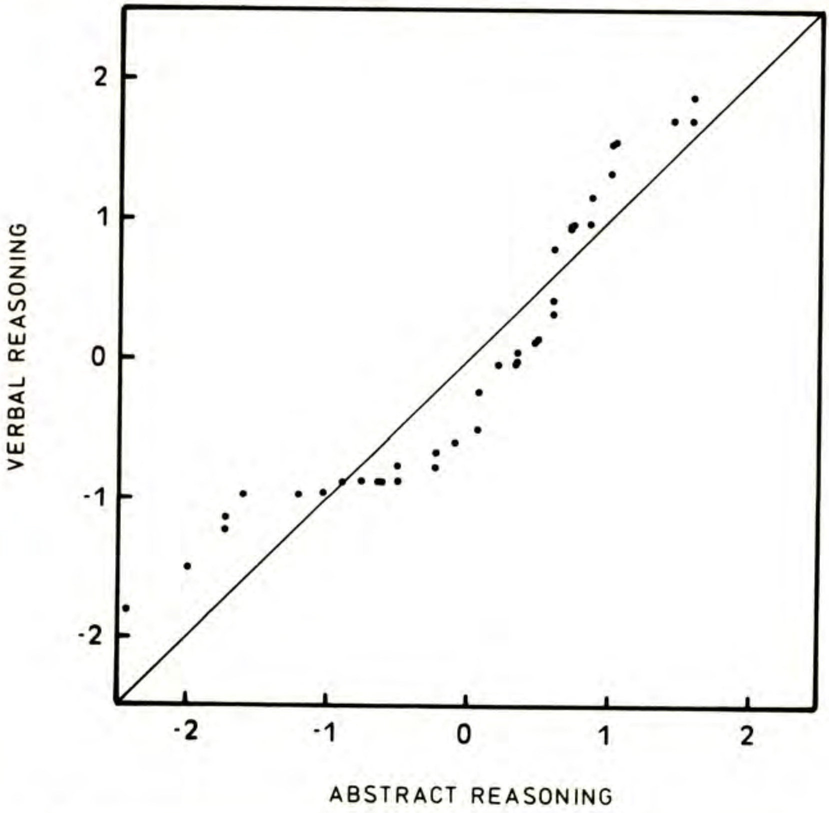

FIG. 2. STANDARDIZED COMPONENT PROBABILITY PLOT OF THE GLASS AND STANLEY DATA

point scatter suggests that the two marginal distributions are not very similar. Therefore, it is useful to consider techniques for interval estimation which do not require such strong and questionnable parametric assumptions. In the next section two such methods are introduced, namely the jackknife and the bootstrap.

The jackknife and the bootstrap are most easily described in a a univeriate one-sample situation. Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ be a sample of n independent random variables, each having the same unknown probability distribution F, i.e.,

$$X_i \sim F.$$

If $\theta(F)$ is some parameter of interest, such as the mean or the variance of F, and $\hat{\theta}(\underset{\sim}{X})$ is an estimate of $\theta(F)$ based on $\underset{\sim}{X}$, then we are interested in the sampling distribution of

$$R(\underset{\sim}{X}, F) = \hat{\theta}(\underset{\sim}{X}) - \theta(F) \tag{1}$$

since its mean gives us the bias of $\hat{\theta}(\underset{\sim}{X})$ and its variance can be used to construct a confidence interval for $\theta(F)$. In this paper we are primarily concerned with the latter use of the sampling distribution of $R(\underset{\sim}{X}, F)$.

The jackknife [2] which was originally introduced by Quenouille (1956), has been advocated by Tukey (1958; Mosteller and Tukey, 1968) as a rough-and-ready statistical tool for nonparametric interval estimation. Let $\underset{\sim}{X}_{(-i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ be constructed from $\underset{\sim}{X}$ by deleting its i'th component, then the i'th pseudo-value is defined by

$$\tilde{\theta}_i = n\,\hat{\theta}(\underset{\sim}{X}) - (n-1)\,\hat{\theta}(\underset{\sim}{X}_{(-i)}).$$

The jackknife estimate of the bias of $\hat{\theta}(\underset{\sim}{X})$ is

$$\widehat{\text{Bias}}(\hat{\theta}) = \hat{\theta} - \tilde{\theta}$$

where

$$\tilde{\theta} = \frac{1}{n} \sum_{i=1}^{n} \tilde{\theta}_i,$$

while the jackknife estimate of the variance of $\hat{\theta}(\underset{\sim}{X})$ is

$$\widehat{\text{Var}}(\hat{\theta}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\tilde{\theta}_i - \tilde{\theta})^2.$$

Tukey (1958) argued that the n pseudo-values can often be treated as approximately independent and identically distributed random variables. Therefore, the distribution of

$$(\tilde{\theta} - \theta)/\sqrt{\widehat{\text{Var}}(\hat{\theta})} \tag{2}$$

can be approximated by a t-distribution with $n-1$ degrees of freedom. This result can be used to construct an approximate confidence

[2] This technique is named after the boy scout's rough-and-ready jackknife which serves in a variety of situations where highly specialized tools are not available.

interval for $\theta(F)$. Further details on the jackknife can be found in a review by Miller (1974).

To illustrate the use of the jackknife more fully, a common example is discussed. Suppose that we have a sample of n observations x_i and that we are interested in the mean of the underlying distribution μ. As an estimate of μ, we use $\bar{x}$, the arithmetic mean of the n observations. The i'th pseudo-value $\tilde{\mu}_i$ becomes

$$\tilde{\mu}_i = n\bar{x} - (n-1)\bar{x}_{(-i)} = x_i.$$

Consequently, the jackknife estimate of the bias of $\bar{x}$ is zero since $\hat{\mu} = \tilde{\mu}$. The jackknife estimate of $\mathrm{Var}(\bar{x})$ becomes

$$\widehat{\mathrm{Var}}(\bar{x}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

which is the same as the estimate of $\mathrm{Var}(\bar{x})$ based on the normal theory. Therefore, the jackknife and the normal theory yield — in this case — identical confidence intervals.

Recently Efron (1979a) proposed the bootstrap as an alternative to the jackknife. If $x = (x_1, \ldots, x_n)$ is the observed realization of $\underset{\sim}{X}$ and $\hat{F}$ the sample probability distribution constructed by putting mass $1/n$ at $x_1, \ldots, x_n$, then a random sample $\underset{\sim}{X}^* = (X_1^*, \ldots, X_n^*)$ drawn from $\hat{F}$ is called a bootstrap sample and the distribution of

$$R(\underset{\sim}{X}^*, \hat{F}) = \hat{\theta}(\underset{\sim}{X}^*) - \theta(\hat{F}) \tag{3}$$

is called the bootstrap distribution. Efron (1979a) suggested that the sample distribution of (1) can be very well approximated by the bootstrap distribution. Consequently the variance of (3) can be used to construct an approximate confidence interval for $\theta(F)$. In practice, the distribution of (3) can often only be determined by Monte Carlo methods. This means that a large number (say N) of realizations of $\underset{\sim}{X}^*$ are generated by sampling n values with replacement from the set $\{x_1, \ldots, x_n\}$. For each realization, $R(\underset{\sim}{X}^*, \hat{F})$ is evaluated, such that after N repetitions, an accurate estimate of the variance of $R(\underset{\sim}{X}^*, \hat{F})$ can be made. This is the bootstrap estimate of $\mathrm{Var}(\hat{\theta})$.

Using the bootstrap to arrive at an estimate of $\mathrm{Var}(\bar{x})$, we would have to draw with replacement say 100 samples of size n from the set $\{x_1, \ldots, x_n\}$ and compute the arithmetic mean of each of those samples. The variance of these 100 arithmetic means would be a bootstrap estimate of $\mathrm{Var}(\bar{x})$.

INTERVAL ESTIMATION OF THE PRODUCT-MOMENT CORRELATION BY JACKKNIFING

In this section we use the jackknife to estimate a confidence interval for the population product-moment correlation ρ. In order to improve the approximation of the distribution of (2) by a t-distribution with $n-1$ degrees of freedom, it is better to jackknife $z_r = \tanh^{-1}(r)$ rather than r.

Applying the jackknife theory to estimate a confidence interval for z_ρ implies that with n pairs of observations, $\tanh^{-1}(r)$ is calculated n times, each time deleting one pair of data values. If $z_{r_{(-i)}}$ denotes the value of $\tanh^{-1}(r)$ obtained after deleting the i'th pair of observations, then the i'th jackknife pseudo-value can be written as

$$\tilde{z}_i = nz_r - (n-1)z_{r_{(-i)}},$$

where z_r is the value of $\tanh^{-1}(r)$ obtained for the complete data set. Assuming that z_r is an unbiased estimate[3] of z_ρ, an approximate $(1-\alpha)100\%$ confidence interval for z_ρ is given by

$$z_r \pm t_{n-1;\frac{\alpha}{2}}\, s_J$$

where s_J is the jackknife estimate of the standard error of z_r, i.e.,

$$s_J = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} (\tilde{z}_i - \tilde{z})^2}$$

with

$$\tilde{z} = \frac{1}{n} \sum_{i=1}^{n} \tilde{z}_i.$$

Applying the tanh transformation to the bounds of the confidence interval for z_ρ, yields a $(1-\alpha)100\%$ confidence interval for ρ.

The jackknife estimate of the standard error of z_r for the Glass and Stanley data is listed in the second column of Table 1, together with a 68%, 95%, and 99% confidence interval for ρ. Since (2) will have an exact t-distribution when the pseudo-values are normally distributed, it is interesting to see how well the distribution of the pseudo-values fits a normal distribution. This can be done by plotting the ordered pseudo-values against the corresponding quantiles of the standard normal distribution, constituting a so-called Q-Q plot (cf. Wilk and Gnanadesikan, 1968). If the pseudo-values are exactly normally distributed, all points will fall on a straight line. Such a normal probability plot of the pseudo-values obtained for the Glass and Stanley data is presented in Figure 3. As can be inferred from the plot, the distribution of the pseudo-values is fairly normal. This is mainly due to the fact that we jackknifed $\tanh^{-1}(r)$ instead of r.

INTERVAL ESTIMATION OF THE PRODUCT-MOMENT CORRELATION BY BOOTSTRAPPING

The bootstrap estimate of the standard error of a sample statistic is based on the variance of the bootstrap distribution. When, like in the

[3] This is not exactly true, but the bias is small enough that it can be safely ignored for all practical purposes.

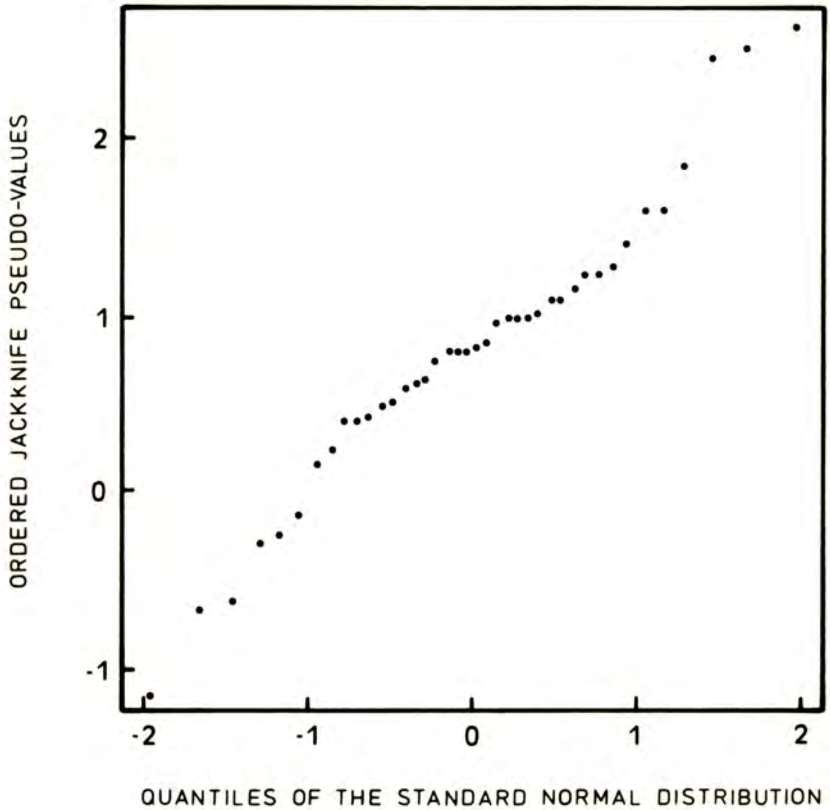

FIG. 3. NORMAL PROBABILITY PLOT OF THE JACKKNIFE PSEUDO-VALUES FOR THE GLASS AND STANLEY DATA

present case, the bootstrap distribution itself is not known, we can approximate it by Monte Carlo sampling. This is done by generating a large number N of bootstrap samples and by evaluating for each sample $\tanh^{-1}(r_j^*) - z_r$, where r_j^* is the product-moment correlation of the j'th bootstrap sample. The distribution of $\{\tanh^{-1}(r_j^*) - z_r | j = 1, N\}$ can be called the empirical bootstrap distribution and its standard deviation can be used as a bootstrap estimate of the standard error of z_r (written as s_B).

However, in order to estimate a confidence interval for z_ρ, we do not need to calculate s_B explicitly, since the easiest way to construct a confidence interval is to use the appropriate quantiles of the empirical bootstrap distribution. Denoting the quantile of the empirical bootstrap distribution corresponding to a cumulative probability of p by q_p, the upper and lower bounds of a $(1-\alpha)100\%$ confidence interval for z_ρ can be written respectively as $q_{\frac{\alpha}{2}} + z_r$ and $q_{(1-\frac{\alpha}{2})} + z_r$.

The value of s_B for the Glass and Stanley data is given in column three of Table 1, as well as a 68%, 95%, and 99% confidence interval. These estimates are based on N = 1000 bootstrap samples.

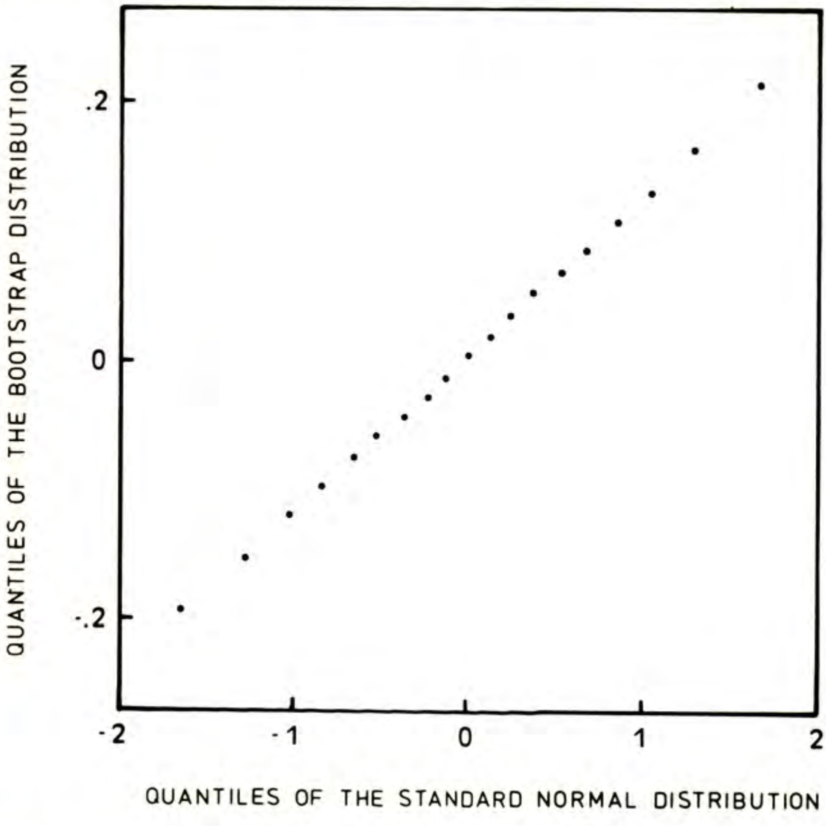

FIG. 4. NORMAL PROBABILITY PLOT OF THE EMPIRICAL BOOTSTRAP DISTRIBUTION FOR THE GLASS AND STANLEY DATA

In Figure 4, we plotted the vigintiles of the empirical bootstrap distribution against the vigintiles of the standard normal distribution. From the linearity of the point scatter we can conclude that the empirical bootstrap distribution is quite normal.

DISCUSSION

The most striking result in Table 1 is that the confidence intervals obtained by the jackknife and the bootstrap are smaller than those based on normal theory[4]. Moreover, the jackknife and bootstrap confidence intervals are more trustworthy since they are not based on strong distributional assumptions. Of course, when the necessary parametric requirements are met, one can safely rely on the normal theory.

Furthermore, the bootstrap confidence intervals are consistently smaller than those obtained by jackknifing. This confirms some previous results (Efron, 1979a, 1979b). However, the difference between the jackknife and the bootstrap confidence intervals is not as large as that found by Efron (1979b). This is due to the fact that we jackknifed $\tanh^{-1}(r)$ whereas Efron jackknifed r.

Although the bootstrap technique seems to be better than the jackknife, it has the disadvantage of requiring more computational effort. The bootstrap method roughly requires N times as much computations as the normal theory, whereas the jackknife only requires n times that many calculations. While jackknifing small data sets is still feasible on a pocket calculator, bootstrapping almost always requires a computer.

Although the jackknife and the bootstrap seem to be very promising methods for interval estimation, they always should be applied with care. For instance, it has been remarked upon that the jackknife yields extremely large confidence intervals when applied to data containing *outliers* (Miller, 1974; Wainer and Thissen, 1975). However, all estimation methods based on the ordinary sample product-moment correlation will fail on such data. The only way out in such a situation is to use so-called robust statistics, a topic we will cover in a future paper.

REFERENCES

EFRON, B. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 1979, *7*, 1-26. (a)

EFRON, B. Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review*, 1979, *21*, 460-480. (b)

GLASS, G.V., & STANLEY, J.C. *Statistical methods in educational research*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

GNANADESIKAN, R. Methods for evaluating similarity of marginal distributions. *Statistica Neerlandica*, 1972, *26*, 69-78.

GNANADESIKAN, R. *Methods for statistical data analysis of multivariate observations*. New York: Wiley, 1977.

HAYS, W.L. *Statistics for the social sciences* (2nd edition). New York: Holt, Rinehart & Winston, 1973.

[4] This does not imply that confidence intervals based on the bootstrap are always smaller than those based on the normal theory. Extensive application of both the normal theory and the bootstrap revealed that sometimes the bootstrap yielded larger intervals than the normal theory. Very often the two intervals differed from each other only at one of the two boundaries, suggesting important differences between the sampling distributions.

MILLER, R.G. The jackknife — a review. *Biometrika*, 1974, *61*, 1-15.
MOSTELLER, F., & TUKEY, J.W. Data analysis, including statistics. In G. LINDZEY & E. ARONSON (Eds.), *Handbook of social psychology* (2nd edition). Reading, Massachusetts: Addison-Wesley, 1968.
QUENOUILLE, M.H. Notes on bias in estimation. *Biometrika*, 1956, *43*, 353-360.
TUKEY, J.W. Bias and confidence in not-quite large samples. *Annals of Mathematical Statistics*, 1958, *29*, 614. (Abstract)
WAINER, H., & THISSEN, D. When jackknifing fails (or does it?). *Psychometrika*, 1975, *40*, 113-114.
WILK, M.B., & GNANADESIKAN, R. Probability plotting methods for the analysis of data. *Biometrika*, 1968, *55*, 1-17.

Department of Psychology
University of Ghent
H. Dunantlaan 2
9000 Gent

Received January 1982