UNIVERSITY OF ARIZONA

INSTITUTE OF ATMOSPHERIC PHYSICS

SCIENTIFIC REPORT NO. 4

A CRITICAL EVALUATION OF CORRELATION METHODS IN CLIMATOLOGY AND HYDROLOGY

James E. McDonald

The work reported herein is part of a study supported by the Alfred P. Sloan Foundation.

January 25, 1957

TABLE OF CONTENTS

ABSTRACT

The sampling properties of three types of correlation coefficients are examined critically with reference to the practical problem of selecting a statistic optimally suited to correlation analyses in climatology and hydrology. Reasons are presented for concluding (i) that undue concern has been given the problem of non-normality of underlying frequency distributions, and (ii) that application of significance tests and determinations of confidence limits for the product-moment coefficient, r, are quite adequately achieved through simple use of the standard error of r. It is suggested that the growing tendency in climatology and hydrology to employ unnecessarily elaborate methods stems jointly from characteristic emphasis upon theoretical refinements found in most statistics references and from neglect on the part of users of statistics to examine the numerical magnitudes of these refinements in relation to basic data-precision and to desired precision of inference.

The properties of a tetrachoric and a rank correlation coefficient are discussed, and it is concluded that in many geophysical applications, especially in cases of moderate sample sizes, the Spearman rank-difference coefficient should be regarded as the preferred correlation statistic. Empirical comparison of these three coefficients are presented for a sample of precipitation data taken from a region (Arizona) where non-normality of precipitation frequency distributions is known to be extreme. The rank-difference coefficient is found to lie within one standard error of r for eleven of fourteen cases in this sample. The tetrachoric coefficient is found to be a much poorer estimate of r, yet in thirteen of fourteen cases it yields (on an approximation basis) the same inferences as does r with regard to existence of correlation.

I. INTRODUCTION.

The underlying thesis of this paper is the following: The writer believes that very real need exists for critiques of statistical methods in the geophysical sciences that can serve as guides to those who must employ these methods in their work but who have limited opportunity to make sufficiently exhaustive surveys of the statistical literature (or sufficiently comprehensive comparative tests) to develop a sense of proportion in selecting techniques of statistical analysis in given areas of geophysical research. Unfortunately the two or three existing references specifically concerned with statistical methods of geophysical research, though extremely useful, offer too little critical advice to the geophysicist trying to decide how far to carry statistical refinements in handling particualar types of data. To object that it is the business of each investigator to find these things out for himself does not fully satisfy the writer, for he feels that this endeavor on the part of the non-statistician is actually hampered by the viewpoints typical of texts and references in the field of statistics, the latter seem all too generally to have been written by authors somewhat too concerned with the intricacies of sampling theory developments to linger upon the unexciting, but eminently practical question of which of these developments make a difference that is truly significant in handling the usual kind of data that climatologists, hydrologists, and other geophysicists wish to subject to statistical analysis.

Indeed, as a result of a recent fairly extensive survey of the literature of correlation statistics which the writer had to make as part of a pilot study for climatic and hydrologic investigations in the Southwest, he has formed the conclusion that many of the more involved applications of statistics in the current geophysical literature are being made by workers who have made a point of becoming familiar with more than just the classical methods without ever having closely examined the question of whether the effort was justified.

Geophysicists who only occasionally use statistical techniques, but who have acquired vague apprehensions of commission of statistical sins by failure to deal elegantly with their data, and particularly younger workers in the field who must apportion their efforts between familiarizing themselves with advanced statistical methodology and familiarizing themselves with innumerable other matters would be benefited in at least a practical way if it could be shown that most geophysical data do not warrant use of anything more advanced than classical (say, pre-1920) sampling theory.

II. SAMPLING CHARACTERISTICS OF THREE CORRELATION STATISTICS.

The above prefatory generalizations will now be examined with reference to the particular (and methodologically important) case of linear correlation analysis, and the attempt will be made to show that elaborations beyond the simplest and most familiar techniques and tests cannot, save in rare cases, yield refinements of real geophysical significance.

A. Product-moment correlation coefficient, r. This statistic is so extensively discussed in so many statistics references that all that will be done here is to try to block out areas of the sampling theory of this statistic that ought not be given much deference by geophysicists (except, of course, on the proper grounds that the theory is interesting and represents an admirable achievement of mathematical statistics that may well prove efficacious in certain restricted areas of research).

Two references concerned with the applications of statistical methods in geophysical problems (chiefly climatological problems) have recently appeared, that of Conrad and Pollak (1950) and that of Brooks and Carruthers (1953). The first of these offers only the barest details of the sampling theory of the product-moment coefficient, yet it is the writer's view that, in most if not all climatological and hydrologic correlation analyses, the investigator is probably well-advised to remain little more sophisticated in his statistical treatment of his data than are Conrad and Pollak in their treatment of correlation. Specific reasons for this somewhat unorthodox view will next be summarized briefly.

1. The normality problem. First, it must be noted that neither Conrad and Pollak nor Brooks and Carruthers make clear to their readers the point that every bit of what theory they discuss applies rigorously only to normal bivariate populations. Now the non-statistician who consults the statistical literature is almost invariably disturbed at every turn by warnings against applications of normal-theory findings to non-normal populations, but is seldom given any inkling of how large an effect non-normality in his data may have. "Effect" here takes on two slightly different meanings associated with two statistical questions: First, the investigator wishes to have some notion of whether computed coefficient can be seriously misleading (i.e., gives a biased estimate of the population value of the correlation coefficient) if it is derived from non-normal data, and second he wishes (or at least should wish) to know whether non-normality can seriously interfere with significance tests (i.e., vitiates the sampling probability distribution assumptions) if these are made on the basis of normal-distribution theory. (Stidd (1953) has suggested that most precipitation data must be regarded as suspect, on grounds of non-normality, with respect to exactly such questions as these two.)

In an empirical test of the first of the above two questions which the writer has just completed, the numerical effects of three frequently suggested "normalizing" transformations (square root, cube root, logarithmic) on the magnitudes of product-moment coefficients of a sample of fourteen

Arizona seasonal precipitation totals were found to be negligible as regards the type of inferences one customarily draws from r, despite presence of typical arid-zone skewness in these data. (This work will be reported in a separate paper concerned specifically with the seemingly small effect of non-normality on a number of the principle statistics used in climatology, hydrology, and other branches of geophysics.)

With respect to the second question, that of making significance tests for correlation coefficients derived from non-normal bivariate populations, the geophysicist must clearly realize that theory is at present still incapable of yielding decisive answers. Statisticians have resorted to experimental sampling to gain estimates of this effect of non-normality, but it would seem that this work was done so many years ago as to have been lost to the present attention of most geophysicists, and hence deserves very brief recapitulation. A substantial amount of experimental sampling was done by E. S. Pearson and his collaborators early in the 1930's, with the objective of testing non-normality effects on a number of statistics, including the product-moment correlation coefficient. In one study relevant to the geophysicist's usual problem, Pearson (1931) set up four different parent populations of numbers distributed with varying degrees of non-normality. Although all of his distributions were much smoother than, say typical precipitation or streamflow frequency-distributions, they exhibited skewness and peakedness of a degree that the writer would say was as extreme as that of most precipitation data outside of the driest of desert regions. Pearson experimentally drew 250 successive samples of 10 pairs each and then 250 more of size 20 pairs each, except for one case where he used 395 samples of each of these two sizes. In each instance the sampling scheme corresponded to sampling from a parent bivariate non-normal population of zero correlation.

From his eight series of samples he formed frequency distributions of r and determined the experimental "standard error" for comparison with the corresponding value computed on the <u>trial</u> assumption that normal theory could be used. For samples of 10 items and a "normal theory" standard error of 0.3333 (the choice of using four digits is Pearson's, not the writer's), Pearson found experimental values of 0.3234, 0.3487, 0.3238, and 0.3181. Similarly for the samples of 20 items each and a "normal theory" standard error of 0.2294, he found experimental values of 0.2413, 0.2405, 0.2289, and 0.2308, so his experiment shows how slight the change in sampling distributions may be as a consequence of non-normality. His conclusion should be familiar to geophysicists who make much use of correlation techniques: "Taken as a whole, the extent of agreement has proved considerably greater than was anticipated before the experiment was commenced, and it seems to indicate that the coefficient of correlation is another of the criteria based on <u>ratios</u>, whose distribution even in very small samples is remarkably insensitive to changes in the form of the population." Indeed, any geophysicist would surely describe this agreement between normal-theory and observed standard errors as near-perfect.

In a somewhat similar experimental sampling study, Baker (1930) drew 50 samples of 40 items (pairs) each taken from a <u>highly</u> <u>skewed</u> population. This case, like the one just cited, constitutes a test of the case of <u>zero</u> correlation in the parent population, (since members of each pair were independently drawn). Baker drew the conclusion that his resultant distribution of r-values was "very skew", but inspection of his results reveals that 29 of the 50 lay within one theoretical standard error of zero, and 44 of 50 lay within two standard errors of zero, which constitutes discernible discrepancy but hardly <u>geophysically</u> alarming asymmetry. The remaining six of the fifty samples fell above (more positive than) two standard errors in excess of zero, presumably a consequence of the high probability of independently drawing two items such that both lie below the mean of a positively skewed parent population. The conclusion of relevance here is that Baker's study tends clearly to support the view that non-normality of even a marked degree is not so large that it would seriously mislead the geophysicist, even though Baker, as a statistician, took his results as empirical evidence of the seriousness of the non-normality problem.

Haldane (1949), in a note on the problem of non-normality in correlation, briefly summarizes his and others' experimental sampling studies as having led to the conclusion that the normal bivariate surface may be distorted and mutilated to a remarkable degree without seriously influencing the frequency distribution of r in the case of small parent-population correlation (i.e., in just the case of importance in making significance tests for mere existence of correlation).

To see the kind of results that arise in experimental sampling from parent bivariate non-normal populations with non-zero correlation, geophysicists should refer, for example, to Chesire, Oldis, and Pearson (1932). For the explanation of their sampling model, the original paper must be consulted; here it is only necessary to note that these workers set up two populations with known correlation coefficients equal to 0.5000 and 0.5462 (number of digits same as used by those authors) and drew 1000 samples of 5 pairs each, 500 samples of 10 pairs each, and 250 samples of 20 pairs each. One population had symmetric triangular frequency distributions of both variates, the other population had one triangular symmetric and one asymmetric distribution; hence two varieties of non-normality were tested in this experiment. "Normal theory" values of r and of its standard error were computed and compared with experimental values. Because the implications of these results are of practical concern to geophysicists and to the thesis of the present paper but are to be found only in a paper that might not be generally available, comparative values for just the cases of 10 and 20 pairs per sample are displayed in Table 1. The authors, as statisticians, conclude that there are "real differences between the distribution of r in samples from these two populations and those appropriate for the normal case", but do suggest that the question of whether these differences would be considered serious in practice depends upon the degree of accuracy required in particular investigations. What geophysicist would see any significance in the differences displayed by the pairs of observed and theoretical correlation coefficients in Table 1?

The writer's conclusion is this: It would appear that many of us who use statistics as a tool in geophysical studies may have accepted with too little scrutiny the warnings of many writers of statistics references concerning the problem of non-normality in our data and have overlooked the existence of older experimental studies which strongly suggest that, within the reasonable limits of non-normality satisfied by most geophysical data (though, of course, _not_ by _all_ such data), investigators could feel confident that _the kind of inferences that geophysicists wish to draw from most correlation studies can be entirely adequately drawn with the aid of the elementary sampling theory of the normal distribution_. Thus, whereas it might seem to some workers that Conrad and Pollak (1950) could scarcely be excused for omitting to explain to their readers that their brief discussion of correlation sampling error was strictly limited to the normal-distribution case, a strong basis exists for the viewpoint that there simply is no such _practical_ limitation in most geophysicists' applications of correlation analysis.

2. _Tests for existence of correlation_. The second basis for arguing that the geophysicist may, in most of his correlation analyses, be

Table 1.

Observed and theoretical values of correlation coefficient r, and standard error, $_r$ for 500 samples of 10 pairs and 250 samples of 250 pairs drawn by Chesire, et al (1932).

		Samples of 10 pairs		Samples of 20 pairs	
		I[1]	II	I	II
Mean r	Obs.	.4871	.5289	.4896	.5460
	Th.	.4787	.5242	.4900	.5359
σ_r	Obs.	.2371	.2383	.1701	.1503
	Th.	.2671	.2534	.1780	.1677

---

[1] I and II designate the two populations samples, II being composed of a triangular asymmetric and an asymmetric subpopulation (see text). I and II had parent-population correlation coefficients of 0.5000 and 0.5462, respectively. For deduction of "theoretical" sampling values see original paper.

satisfied with about as limited a treatment of theory as that given by Conrad and Pollak concerns actual tests for significance which the investigator may make once the question of normality has been decided in favor of believing that it may be ignored. The investigator always wishes to know whether a given computed product-moment coefficient may safely be taken to indicate that the parent population is characterized by non-zero correlation, i.e., he certainly wishes at least to test for existence of correlation (without regard to "true" magnitude). Conrad and Pollak (p. 245) quote a rule-of-thumb criterion which used to appear in the literature of meteorology and climatology fairly frequently, namely the rule that a coefficient must equal or exceed three times its probable error (or, closely enough, twice its standard error) if it is to be taken as evidence of the existence of correlation in the parent population.[1] Nowadays, we seem to regard it more fashionable to invoke arguments based on a t-test. To be sure, it takes little more time to compute t and look up a significance level in a t-table than to compute a standard error, yet with reference to the main thesis of the present report, it is of interest to observe the wholly unimportant degree of refinement (in all but rather small samples) which this modernization introduces into one's statistical inferences. In Table 2 are shown for several values of sample size N, values of the product-moment coefficient r that will be just significant at the 95 per cent confidence level based on the t-test, as described, for example, by Brooks

[1] The philosophical meaning of the phrase "parent population" is, of course, obscure when sampling time series, but is to be understood here in common-sense terms. There is evident need for closer grappling with this difficulty that underlies so much of climatology's and hydrology's use of statistics.

Table 2.

Values of r significant at the 5 per cent level as derived for two criteria.

Sample Size N	(1)[a]	(2)[b]
10	.63	.50
20	.44	.39
40	.31	.29
60	.25	.24
80	.22	.22
100	.20	.20

[a] Values of r in Column (1) are based on the t-test.

[b] Values of r in Column (2) equal twice their respective standard errors, i.e., are solutions of equation (1) for $\sigma_r = r/2$

and Carruthers (1953, p. 220). Also shown, for comparison, are corresponding values of r, for the same sample sizes, that would be regarded as significant at about the 95 per cent level as computed by the writer on the rule-of-thumb basis that r must then equal or exceed twice the standard error as given by

$$\sigma_r = (1-r^2) \Big/ (N-1)^{\frac{1}{2}}. \qquad (1)$$

It can be seen from Table 2 that if one used the rule-of-thumb criterion that to be "significant" r must equal or exceed twice its standard error, he would, for all but rather small sample-sizes be employing almost exactly the same threshold values that he would derive from the t-test for r applied at the 95 per cent significance level. For samples much smaller than 20 variate-pairs the numerical difference between the two criteria begins to assume some small importance, but two considerations render even these small-sample differences apparent in Table 2 of only secondary significance: First, one cannot, in most geophysical problems, feel safe in drawing definitive conclusions from such small samples of data, and second the entirely arbitrary step of selecting 95 per cent rather than 90 or 99 per cent as the working significance level introduces a difference about as large as the observed difference in threshold r-values of the two columns Table 2. The writer's conclusion is that the once commonly used rule of requiring r to exceed twice its standard error in order for r to be regarded as "significant" (at the 90-95 per cent level) should never have been abandoned and should be restored to good standing. Only in occasional cases where one feels that he has real basis for trying to draw an unusually refined inference from a quite small sample need even one resort to a t-test. That is, examination of the numerical implications of the t-test have here strongly indicated that the geophysicist can safely trust the unsophisticated treatment of

Conrad and Pollak on this second point too and this despite the fact that the latter treatment overlooks completely the theoretical developments of the past thirty years.

3. Determination of confidence limits. The third basis for suggesting that the niceties of sampling theory of product-moment correlation coefficients not found in a treatment such as that of Conrad and Pollak need not greatly trouble the geophysicist who has not been able to become thoroughly familiar with them concerns the final problem of placing confidence limits on a given computed correlation coefficient. It is theoretically important to distinguish such a process from that just discussed above. For the sampling distribution of correlation coefficients drawn from a parent bivariate population of zero correlation is itself normal, but the corresponding distribution for a parent population of non-zero (especially of high positive or negative) correlation is skewed, and (still speaking theoretically) this affects significance-testing procedures. In almost any reference the geophysicist may consult for aid in interpreting his correlation analyses, he will be confronted with disquieting admonishments against naively using any kind of standard error argument in setting his confidence limits on r. If he reads further, being sufficiently disturbed by these admonishments, he is lead to use the hyperbolic arctangent transformation, usually called simply the z'-transformation.

The technique whereby one transforms a computed r into equivalent z', places selected confidence limits above and below this value of z' by what amounts to a t-test, and then makes the inverse transformation from the two resulting limits back to their equivalents in terms of r is explained in many statistics texts. The interested reader will find particularly clear discussions given by Ostle (1954, pp. 181-185) and Mills (1955, pp. 297-309), and a handy graphic aid for effecting the r-z' transformation and its inverse is presented by Snedocor (1946, pp. 152-153). But first, one should ask, how uncomfortable should an investigator, correlating data of inherent precision typical of most geophysical data, feel if criticised for simply citing a conventional standard error computed by equation (1) rather than placing confidence limits via Fisher's z'-transformation? Table 3 reveals that he really need not feel remiss at all. In that table, values of the (asymmetric) upper and the lower half-widths of the 95 per cent confidence interval about r based on z' are compared with twice the standard error of r computed from equation (1) for two sample sizes, N, and for a number of values of r. It is evident from Table 3 that using plus and minus twice the standard error of r (strictly speaking one should use 1.96 times the standard error, but such a two per cent refinement is no more materially important than the other refinements here under discussion) yields estimates of the 95 per cent confidence limits that are, even in the cases of poorest agreement, close enough to those obtained by using Fisher's z'-transformation that geophysical inferences will never be seriously distorted by the discrepancy. Since one must always recognize that use of the 95 per cent rather than, say, the 90 or the 99 per cent confidence level is itself basically arbitrary, and since substituting either one of these levels for the 95 per cent level will alter the half-widths by as much as or more than the replacement of the z' method by the two-standard-errors rule, little practical justification can be given for use of this transformation in most geophysical studies. Note that, in general, double the standard error gives a half-width lying intermediate in value between the upper and the lower half-width, so that, in a sense, $2\sigma_r$ constitutes a <u>single</u> figure rather nicely summarizing the two asymmetric limits of refined sampling theory. The statistician will rightly object that this single figure is, however, a <u>biased</u> estimate of the confidence half-widths, but it is the small <u>magnitude</u> of this bias that is stressed here as noteworthy. In all, the foregoing quantitative comparison

Table 3.

Comparison of 95 per cent confidence half-widths of r for two sample-sizes N.

	N = 20			N = 80		
r	d[1]	d'	d"	d	d'	d"
0.10	.46	.42	.46	.22	.21	.22
0.20	.44	.39	.47	.21	.20	.22
0.30	.42	.36	.46	.20	.19	.21
0.40	.39	.31	.45	.19	.17	.20
0.50	.35	.27	.43	.17	.15	.20
0.60	.30	.22	.39	.14	.12	.16
0.70	.24	.17	.33	.11	.10	.13
0.80	.17	.12	.25	.08	.07	.10
0.90	.09	.06	.14	.04	.04	.05

---

[1] $d = 2\sigma_{\bar{r}}$. The quantity d' is the upper half-width of the 95 per cent confidence interval as derived from the z'-transformation, and d" is the lower half-width of the 95 per cent confidence interval as derived from the z' transformation.

between an old (in fact, virtually abandoned) criterion of simple form and a somewhat newer and more complex criterion of sampling uncertainty of the correlation coefficient should convince the investigator who has not yet had the opportunity to familiarize himself with Fisher's z'-transformation that he should surely employ the simple standard error argument unless he feels that he is handling geophysical data of such rare precision and freedom from observational-sampling variability that errors of estimate of a few hundredths in his confidence limits are going to lead him to erroneous inferences. And any such unusual data are likely to be of such functional simplicity as to obviate resort to statistics in the first place.

4. Historical interpretations. All that has been said above can be concisely summarized by saying that investigators dealing with climatological, hydrologic, or other complex geophysical data should take a common-sense view of correlation analysis as seen in the context of data-precision and should, except in the few unusual cases noted above, feel quite justified in using only the simplest of classical sampling theory of the Pearson product-moment correlation coefficient.

Since this thesis may seem unorthodox and retrogressive to some, it is important to attempt an explanation here of why so many non-statisticians have acquired an inclination towards statistical overelaboration that so often seems to underlie the statistical discussions in geophysical papers (especially those in climatology and meteorology), or worse, have acquired only a vague uncertainty as to the appropriateness of various statistical techniques. It seems to the writer, as a result of having recently had the curiosity to scrutinize some two dozen statistics texts and leading references, that these books chiefly reflect, for one reason or another, the interest of theoreticians, and that the latter have in recent decades (especially since the early work of R. A. Fisher) extended their sampling theories in directions and to degrees of refinement that frequently have borne only weak relation to the typical levels of data-precision of those very areas of research which, for reasons of inherent complexity and inherent variability of subject phenomena, must turn to statistical techniques to secure objective answers to many problems. (This historical speculation must not be taken too broadly, for within the present decade there has certainly been much theoretical work designed to meet more practical investigative needs. These very recent advances are not to be found in the references typically contained in geophysical bibliographies, so one senses a need for critiques on the application of these newer techniques in the geophysical sciences.)

One further observation may be made tending to support the above view that the statistics books most geophysicists are likely to consult for aid and advice reflect the viewpoints of theoreticians who have not always confronted the facts of data-precision in those very fields most needing statistical approaches: The point had not been previously noted by the writer, but in the present study he began to be aware of a general tendency of writers of statistics texts to give correlation coefficients to more than two significant digits, surely the limit that a sense of numerical proportion would seem to indicate. Systematic inspection of fifteen widely used and quoted statistics texts revealed, in fact, that eleven out of fifteen writers used three or more digits, and in all, fully a third gave coefficients to <u>four or more digits</u>! Such practices may offer partial explanation of an almost incredible but true instance wherein a non-statistician associated with a hydrologic project came to the writer with a basinwide rainfall-runoff correlation equal to 0.946 when all years were included, and 0.957 when a single very wet year was excluded, and asked if this might not be taken to indicate that a proposed major and very costly hydrologic experiment on that basin could be tested statistically by comparing the correlation coefficient for all years up to but not including the year of treatment with the coefficient found for that data plus the data for the immediately following treatment! That statisticians cannot collectively smile at such an interpretation of significant digits in r is shown not only by the curious results of the above-mentioned tally of fifteen textual practices on significant digits but also by many extreme cases such as those of the application of attenuation corrections to r, or adjustment of r via Sheppard's correction, or a particularly astonishing case which is to be found in a text by Smith and Duncan (1945, p. 303) where the reader is shown how an original correlation coefficient of 0.89576 can, by suitable algebraic manipulation, be converted into the Fisherian "maximum-likelihood estimate" of 0.89465. Clearly, statisticians who would theorize on a means of improving a correlation coefficient by one part in nine hundred are not addressing themselves to the problem of statistical manipulation of rainfall amounts read in 1908 under a rapidly growing shade tree or of streamflow data for a month containing a record flood which the observer was unable to gauge when cresting because the access roads were submerged.

The geophysicist who has been consciously or unconsciously influenced by any school of statistical thought divorced from awareness of physical and observational uncertainties inherent in the work of most of the geophysical sciences may be well-advised to review his own statistical practices along lines similar to those here followed for the product-moment correlation coefficient and its sampling properties, and to determine thereby whether simplifications may not be in order. If for no other reason than to reduce the burden of methodology to be assimilated by younger workers in the field of geophysical sciences, all unnecessary and superficial complexity of statistical methods in geophysics should be discouraged.

B. <u>Tetrachoric correlation coefficient, r_t</u>. This simple correlation statistic was originally developed by K. Pearson for the purpose of testing for association between two factors or attributes that could only be dichotomized; that is, r_t measures association in a two-by-two contingency table. From an original cumbersome quartic equation for r_t, Pearson has derived simpler approximations, of which one is given by Brooks and Carruthers (1953, p. 238, eqn. 170). It is the latter formulation that will here be called the tetrachoric correlation coefficient. This statistic is occasionally met in the geophysical literature and has certain useful applications that should be more widely known, so its properties and limitations deserve brief comment within this critique.

Perusal of even a modest sampling of references and texts in statistics discloses a curious disagreement as to whether an investigator is well advised to dichtomize numerical variates about their means and compute r_t; but there is no disagreement that statistical efficiency is therby lowered, for information is clearly discarded in that step. Dixon and Massey (1951, p. 234) give 0.40 for the efficiency of a tetrachoric coefficient closely similar to which, though not identical with, the r_t of the present discussion, which gives numerical values almost identical with those obtained using Brooks and Carruthers' equation (170). Dixon and Massey describe a modified scheme characterized by an efficiency of 0.52 which would be only slightly more trouble to apply in practice, but it is evident from such efficiencies that about twice as large a sample size must be used in tetrachoric schemes or in the product-moment method to get equal sampling uncertainties for the two schemes.

Use of r_t is limited, in theory, to bivariate normal distributions, a limitation Brooks and Carruthers do not cite, but which will surely not be dwelt upon here. Brooks and Carruthers offer an expression for the standard error of r_t due to K. Pearson, but Kendall (1948, p. 356) and other authors take pains to point out that such an expression has questionable theoretical basis.

Every one of the several extant tetrachoric coefficients can be computed so rapidly compared to the computation of r, that it is of real methodological interest to have some notion of how reliable a statistic r_t may be. Aside from the theoretical efficiency values cited above and some values to be given below for empirical comparisons involving precipitation data, the writer has found only a single additional bit of helpful evidence in the literature. Peters and Van Voorhis (1940, p. 380) cite an empirical comparison of 78 values of r_t and corresponding values of r. The nature of the data is not specified, sample-sizes are ambiguously stated, and the exact type of r_t that is involved is unclear, but it is stated that the mean deviation of r_t from r was only about 7 per cent, or, more informatively, that r_t values were prevailingly within one probable error of corresponding values of r. If it could be shown that r_t would usually estimate r that faithfully when applied to typical geophysical data, this statistic would deserve much wider use as a rapid approximative measure of correlation (notwithstanding the real theoretical difficulty imposed by present ignorance of its sampling errors).

The writer is of the opinion that in pilot correlation studies of large amounts of geophysical data where only first-approximation results are needed, r_t can serve a quite useful purpose. Thus, in exploratory studies of climatic homogeneity with respect to some given variable where large numbers of interstation correlations must be computed and their geographic distribution studied for dominant patterns of similarity and dissimilarity, tetrachoric schemes have much to offer, and the more so the more willing the investigator is to pay for large arrays of correlation estimates with reduced precision of individual estimates. An extensive empirical test with actual data would be a valuable contribution to knowledge of geophysical-statistical methods, and may become possible in the course of the studies that have led to the present report. The sample offered later in this report constitutes a step in that direction.

C. Spearman's rank-difference correlation coefficient, r_s. Just as the tetrachoric schemes of testing association were originally developed for application to merely dichotomized distributions, so Spearman developed what is called the rank-difference correlation coefficient primarily for use in objectively assessing the degree of concordance of pairs of attributes that can be ranked even if not given numerical expression. However, just as r_t can readily be computed for numerical data by first dichotomizing the bivariate distributions, so also r_s can be computed for numerical data by simply arranging the variates in numerical order, assigning integral ranks, and proceeding with the arithmetic. The practical question before the investigator preparing to examine some data for existence of (or degree of) correlation is that of whether r_s may be expected to provide a measure of correlation that is reliable enough and readily enough interpreted to meet his particular needs.

In the two most recently published works on the use of statistics in geophysics one finds so little information on r_s as to lead one to believe that this statistic has little to recommend it. Conrad and Pollak (1950) do not even mention r_s, while Brooks and Carruther (1953, pp. 235-237) go as far as to say that correlation by rank is not likely to have much application in meteorology. The writer cannot agree with that view, for after recently sifting the somewhat conflicting opinions and evidence to be found scattered throughout the statistics literature, he has concluded that Spearman's rank-difference coefficient should probably be regarded as the preferred coefficient for use in most geophysical applications involving samples of moderate size (say, less than about 40 pairs; see below). No comments on computational details are in order here, but in view of the potential importance of this little-used coefficient, it seems relevant to cite a few references that are particularly informative with respect to this statistic: Very extensive discussions of r_s and related rank correlation statistics have been presented by Kendall (1948a, pp. 388-437; 1948b) and a fairly complete treatment is given by Yule and Kendall (1950, pp. 258-270); Walker and Lev (1953, pp. 278-286) offer a shorter discussion that is useful but which omits mention of the tied-rank problem and which, like a number of other references the reader might come upon, leaves the impression that it really ought not be used for continuous variates (such as the geophysicist almost always deals with); Moran (1951) has treated partial and multiple rank correlation; finally, for simple computational instructions, see the two compact but very useful paperbacks by Arkin and Colton (1955, pp. 85-87) and Moroney (1953, pp. 334-336).

It must be recognized that the ranking step, which comes first in computing Spearman's r_s, ultimately becomes very laborious in the limit of large samples, unless one uses punchcard methods in which sorting techniques are extremely rapid. It was observed many years ago by K. Pearson that for samples of less than about forty paired variates one can compute r_s faster than r; while for samples much in excess of forty r is easier to compute, manual (desk calculator) methods being assumed. The writer's colleague,

R. B. DesJardins found that to compute r_s for the fourteen samples (each sample comprising an average of about 65 pairs) of precipitation data discussed below, using an IBM 602A calculator, required about one-half hour following sorting that took approximately fifteen minutes; and wiring and listing on an IBM 419 tabulator consumed an additional hour, or a total of about eight minutes per coefficient. (Independent time estimates made by DesJardins for product-moment calculations with this same equipment yield rates that are slightly, but not significantly, slower; so there is apparently little to choose between r and r_s on a time basis in punchcard analysis with this type of equipment. It should be noted that standard deviations and regression equations are useful by-products of an r-calculation, while medians, percentiles, and simple skewness measures are useful by-products of the ranking step.) For investigators limited to manual methods, it seems certain that for anything under about 30-40 variate-pairs, r_s will be much easier to compute, partly because in its computation one squares only integers and this can be done mentally unless r_s is large and negative. These time considerations can become decisive in studies where large numbers of correlations are to be computed and deserve further scrutiny if a given study is to involve an extremely large series of correlations.

In random sampling from the statistical literature, the geophysicist must be prepared to find widely divergent opinions concerning the limits of utility of r_s, expecially as a substitute for r. Much of this becomes irrelevant when one weighs the points of numerical disagreement against the inherent errors and observational-sampling fluctuations characteristic of geophysical data, i.e., again in this instance it seems that one needs to put in its proper place a certain amount of theoretical hair-splitting.

Nevertheless, an interesting theoretical argument favoring use of r_s in the commonly-occurring case of non-normality of the basic data was given twenty years ago by Hotelling and Pabst (1936) with reference to the statistical efficiency of r_s versus r. Those investigators were able to show that in the large-sample limit, and for normal bivariate populations characterized by zero correlation, r_s has an efficiency of 0.91; that is, in order to obtain rank correlation coefficients as sensitive (in the sense of equal sampling variance) as product-moment coefficients, one must use sample-sizes about ten per cent larger for r_s than for r, a consequence of information lost through converting from continuous variates to ranks. Hotelling and Pabst then suggest that this theoretical nine per cent loss of information in using r_s instead of r to test for existence of correlation in the normal case is probably larger than the corresponding loss for any non-normal case, and this because the most efficient estimator of correlation will no longer be r when one turns to non-normal populations. Their final conclusion bears strongly on correlation analysis in geophysical studies: "Certainly where there is complete absence of knowledge of the form of the bivariate distribution, and especially if it is believed not to be normal, the rank correlation coefficient is to be strongly recommended as a means of testing the existence of relationship."

(Another interesting property of any rank-correlation coefficient was called to the writer's attention by Dr. R. A. Bryson in discussing this manuscript: He points out that one sometimes wishes to correlate physical quantities which are themselves non-linearly related to the independent variables governing them. Relative humidities, vapor pressures, kinetic energies of flow, etc., are examples. In undertaking product-moment correlations of such quantities, doubt often arises as to whether some non-linear transformation of the quantities should precede correlation analysis.

Inasmuch as ranks are invariant under any kind of transformation that only stretches scales, this question entirely disappears in using a ranking correlation technique.)

The sampling errors of r_s were for a long time unknown. In 1938 two independent solutions for samples of less than eight (for which the combinatorial algebra was regarded as still just barely manageable) were given. For samples of more than about twenty-five pairs, theory indicates very satisfactory convergence toward the familiar standard error of r. Olds has recently filled the remaining gap from eight to twenty-five, and a summary of his results is given in convenient form by Dixon and Massey (1951, p. 261). (In using the latter summary the one-tailed nature of the table must be noted.) It is informative to note how close the sampling errors of r_s follow those of r even for small samples. Thus a one-tailed test at the 95 per cent significance level gives critical values for r_s and r, respectively, as follows; for samples of 10, 0.56 vs. 0.55; for samples of 20, 0.38 vs. 0.38. Indeed, from the point of view of the geophysicist, it is entirely adequate to put r_s in equation (1) and neglect Old's results except inasmuch as they serve to show the very facts that one's sampling uncertainties for r_s do not depart from those of r by amounts that are worth bothering about. For those tiny samples where 95 per cent values of r_s are different from those for r, one ought not infer that resort to Old's figures render inferences "exact".

At present, theoretical basis exists only for tests of the null hypothesis that a given r_s is non-zero only by virtue of fluctuations associated with sampling from a parent population of zero correlation. No theory yet exists for placing confidence limits on non-zero values of r_s, though the writer strongly suspects that the investigator who wishes, merely for his own information, to have some estimate of how closely he can trust an observed non-zero r_s will seldom draw a wrong geophysical inference by simply making appropriate use of σ_r given by (1) above. A long-standing but apparently controversial functional relation between r_s and r derived by Pearson (see, for example, the discussion of Kelley, 1947, p. 367) gives the result that r_s should never be more than a few hundredths different from (specifically, smaller than) r for the same data.

It would be most helpful, from the practical standpoint, to have experimental sampling results for r_s using a variety of synthetic populations, especially for the case of non-zero correlation in non-normal populations, but the writer in his search of the literature found only a single experiment of this type. Snedecor (1946, p. 166) drew ten samples, each of 20 variate-pairs, from a normally distributed parent population whose product-moment correlation was known to be 0.56, and found the average value (averaged via z') for the ten r_s-values to be 0.51, with the greatest individual discrepancy between r_s and r amounting to 0.16 and over half of the discrepancies amounting to less than 0.07. Snedecor draws only the conclusion that this small experiment gives no evidence <u>against</u> using r_s instead of r. The comparative data given below, for Arizona precipitation data, point in the same direction, namely that r_s is a geophysically adequate estimator of r.

It is to be noted that, in addition to Spearman's rank-correlation coefficient, there exist a number of other rather similar rank-correlations. Kendall's <u>tau</u> coefficient is one of these which the geophysicist may find described as having certain advantages over r_s. In references over ten years old, the advantages asserted for tau will usually refer to the fact that the sampling distribution tau was then more satisfactorily known than was that of r_s. However, that difference has been eliminated through recent work of Olds cited earlier here. Kendall (1948a, p. 393) summarizes a discussion of these two rank correlations by saying that one appears to be as good as the other, but that

r_s is easier to compute than tau. Siegel (1956, p. 239) points out that both of these coefficients have the same theoretical efficiency.

In summary of the foregoing critique of Spearman's r_s, it should be said that a correlation coefficient with theoretical efficiency estimated to be as high as 0.91 for the one solved case, a statistic whose distribution-free nature gives it conceptual advantages over r for exactly the kind of distributions (non-normal) customarily encountered by the geophysicist, a statistic which can, for samples-sizes that commonly arise, be computed more easily than can r, would indeed seem to be the preferred measure of correlation in many geophysical studies. Hence, Spearman's coefficient r_s deserves much wider attention than has been given it despite its relative antiquity. To the above considerations may be added a further one of philosophical nature: The hydrologist, say, who converts into a pair of series of merely integral ranks an original pair of series of conventional raingauge reports and streamgauge readings is really not discarding even as much real "information" on rainfall-runoff relations as Hotelling and Pabst's surprisingly low nine per cent information-loss estimate would suggest. Candid appraisal of the overall representativeness and accuracy of such original data would often, in the writer's opinion, make ranking seem a more intellectually honest procedure than calculating product-moments with several digits of which at least one or two must be admitted to be only questionably significant in the first place.

III. EMPIRICAL COMPARISON OF r, r_t, and r_s.

In the preceding sections the general properties of r, r_t, and r_s have been considered, and in a few instances, experimental sampling comparisons have been cited to indicate the approximate extent to which the last two of these coefficients serve as estimates of the first. None of the comparisons have dealt with actual geophysical data, and in fact in most instances the sampled data were entirely hypothetical data. It is of real interest to obtain some notion as to whether actual climatological or hydrologic data will support the view that both r_t and r_s, and especially the latter, are acceptable estimators of r.

Since there are, to the writer's knowledge, no published comparative figures on this question, the following sample, though admittedly too small to permit broad conclusions to be drawn from it, is presented as indicative evidence based on some recent pilot studies carried out by the writer. Inasmuch as the samples comprise seasonal precipitation data from an arid region (Arizona) and inasmuch as the frequency distributions of precipitation amounts for arid regions are well-known to be notoriously non-normal, it is believed that these data will provide particularly useful indicators of the upper limit to the extent to which non-normality leads to quantitative differences between the several correlation coefficients.

Five long-record Weather Bureau stations in Arizona were used, with records ranging from 50 to 84 years. All data were taken from Weather Bureau Bulletin W and were grouped into seasonal totals for each year at each station, with "winter" taken as the six-month period from November 1st to April 30th, and "summer" the remainder of the year. Correlations between Tucson and the other four stations were obtained, and then correlations between Natural Bridge (a station in the mountainous central part of Arizona) and the remaining three stations were obtained, yielding seven correlation coefficients of each type for each of two seasons, or a grand total of forty-two coefficients.

The numerical results are presented in Table 4, along with three additional sets of numbers, the standard error of r as computed from equation (1), the difference r_s-r, and the difference r_t-r. It is chiefly these last three sets of numbers that reveal the representativeness of the tetrachoric and the rank-difference correlation coefficients.

Table 4.

Comparative values of fourteen sets of r_s, r_t, and r for Arizona seasonal precipitation data.

		Station-pairs[1]						
		A-B	A-C	A-D	A-E	B-C	B-D	B-E
	N[2]	64	76	50	83	64	50	63
	r_s	.70	.76	.53	.58	.83	.59	.66
	r_t	.69	.74	.46	.55	.85	.78	.65
Winter[3]	r	.78	.82	.58	.65	.85	.57	.68
	σ_r	.05	.04	.09	.06	.03	.10	.07
	r_s-r	-.08	-.06	-.05	-.07	-.02	-.02	-.02
	r_t-r	-.09	-.08	-.12	-.10	.00	.21	-.03
	N	63	76	52	84	63	51	63
	r_s	.57	.46	.49	.26	.64	.68	.34
	r_t	.64	.58	.37	.29	.72	.68	.10
Summer[4]	r	.54	.38	.44	.19	.61	.63	.22
	σ_r	.09	.10	.12	.11	.08	.08	.12
	r_s-r	.03	.08	.05	.07	.03	.05	.12
	r_t-r	.10	.20	-.07	.10	.09	.05	-.12

[1] Station code letters as follows: A-Tucson, B-Natural Bridge, C-Phoenix, D-Flagstaff, E-Yuma.

[2] N is the number of years of record entering into each correlation.

[3] Winter is here the period from November 1st to the following April 30th.

[4] Summer is here the period from May 1st to the following October 31st.

Note that 11 of the 14 cases exhibit absolute differences between r_s and r that are equal to or smaller than the computed standard error of r, and that none of the remaining three is even as large as twice σ_r. This constitutes, for actual geophysical data of non-normal distribution characteristics and for the case of almost certainly non-zero "parent-population correlation" very encouraging evidence that the geophysicist can expect to get reliable estimates of degree of correlation by the use of the Spearman coefficient, thus supporting the conclusions reached in the preceding section on other grounds. (The climatologist will note the seasonal reversal in sign of $r_s - r$, associated with seasonal skewness differences.)

Next, with reference to the reliability of the tetrachoric coefficient, note that 7 of the 14 cases exhibit an absolute difference between r_t and r that is equal to or less than one standard error of r, but that 3 of the 14 exhibit differences equal to or in excess of two standard errors. This must be taken as confirmatory evidence that r_t is not so satisfactory an estimator of degree of correlation as is r_s, at least for use in the Southwest, where skewness imposes obvious difficulties in using r_t with dichotomy at the mean. It would be of interest to have similar comparisons for less arid regions, but it must be noted that even in this present rather stringent test, one could use r_t as only an indicator of existence of correlation (say, in conjunction with a highly approximative standard error obtained from (1) by replacing r by r_t for rough estimation purposes or by using Pearson's approximation) and would reach conclusions identical to those based on r in all of the fourteen instances of Table 4 except for the case of the Tucson-Yuma summer correlation, which would be just significant if only r_t were employed but is not quite significant at about the 95 per cent level when r and its standard error are employed. Yuma's summer data are so highly skewed that one is facing here the most extreme disturbing effect of non-normality likely to be found in the entire country, so this single failure is not alarming.

In Figures 1 and 2 the degree of concordance of r with respect r_t and r_s is graphically displayed in such a way that the departures of the latter two from the former are readily measured in terms of the standard error of r as a yardstick. Also displayed in each figure are the limits of the band within which r_t and r_s are not over 10 per cent of r away from r, and it can be seen that nine of the fourteen values of r_s fall within this band, though only three of fourteen values of r_t satisfy this condition.

Table 4 and Figures 1 and 2 exhibit the results of what is, of course, a quite small sample from the statistician's viewpoint. Nevertheless, it refers to actual geophysical data of a type (arid-zone precipitation data) known to depart from normality to a larger degree than many other geophysical data. Hence, the fact that r_s and r have been found to agree so closely (above all in the important r-region near 0.5 to 0.7) is taken by the writer to imply that there is probably no reason to see superiority in r over r_s in most geophysical correlation analyses. It should be remembered that, in view of the position taken by Hotelling and Pabst (cited above), one may even suspect that r_s is, for data such as these, an even <u>more</u> efficient estimator of correlation than r. And, because r_s is much more easily calculated than r for small samples, the writer suggests that for such cases r_s be used in preference to r.

IV. SUMMARY AND RECOMMENDATIONS.

A. Geophysical applications of product-moment correlation. A critique of certain essential sampling characteristics of Pearson's r, the most widely used linear correlation statistic, has been presented above as an outgrowth of a rather detailed study of the correlation problem which the writer has

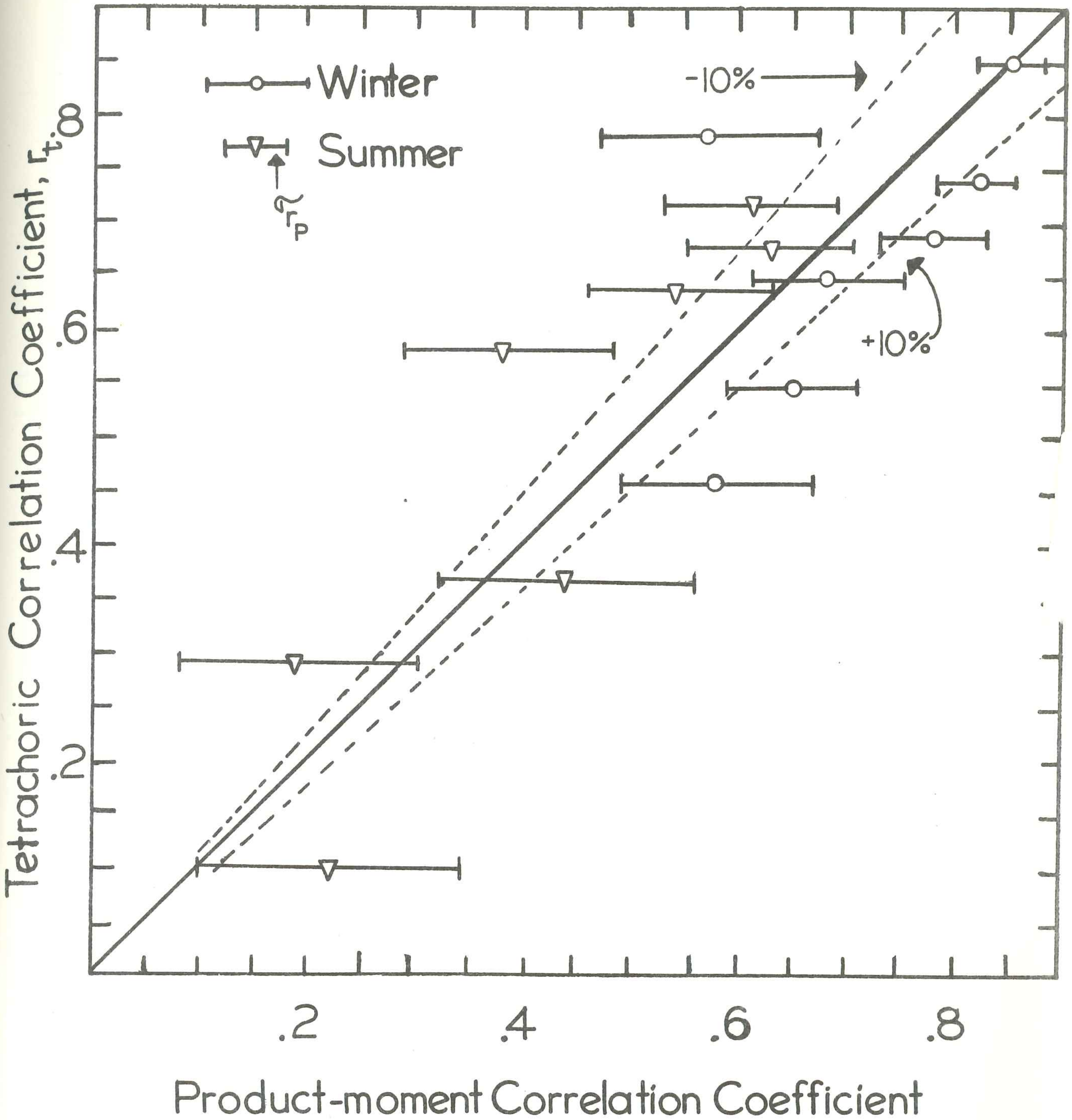

Figure 1. Comparison of r_t and r for Arizona seasonal precipitation data.

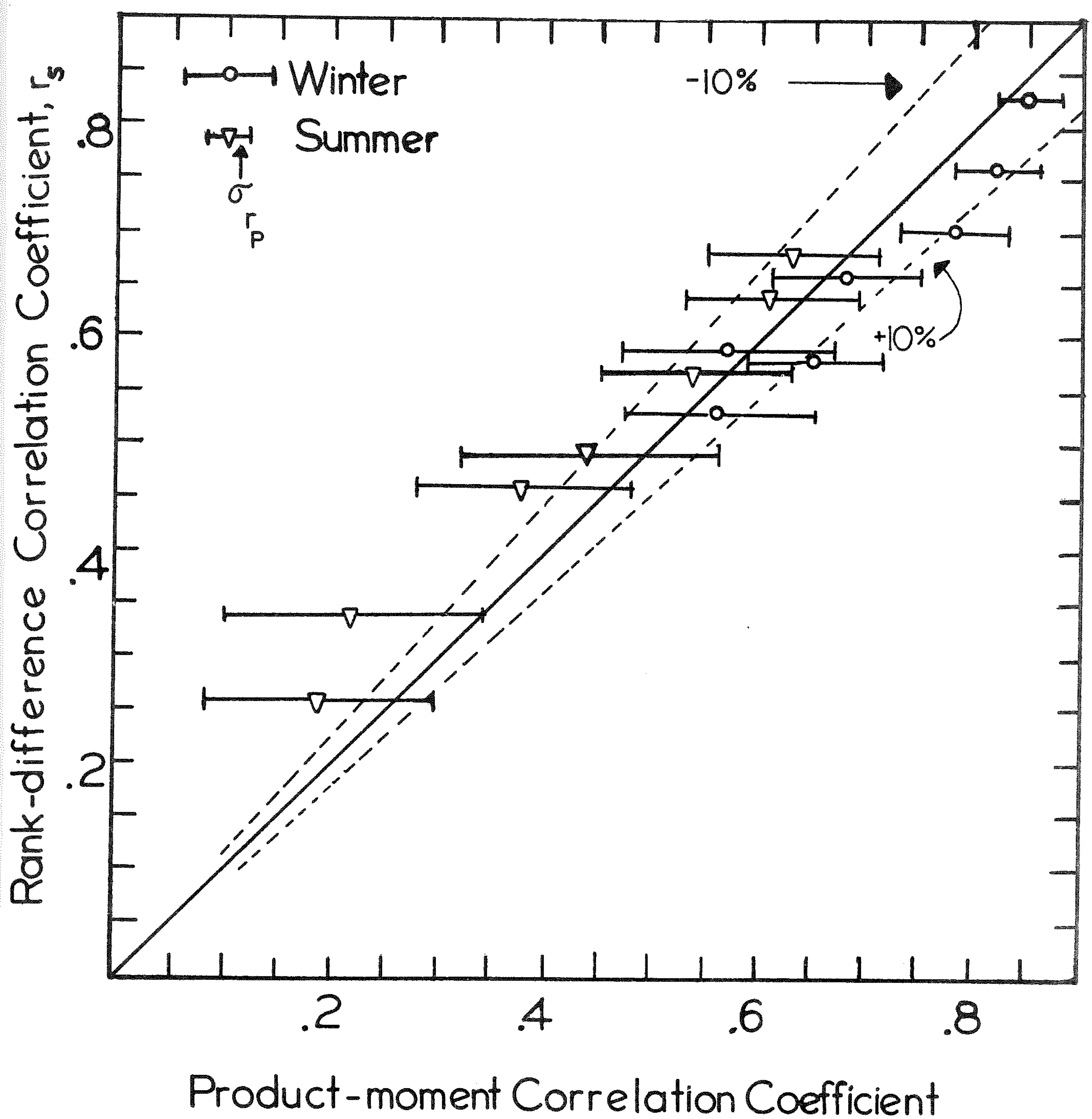

Figure 2. Comparison of r_s and r for Arizona seasonal precipitation data.

recently made. Whereas this survey began as an effort to make an optimum selection from among the several sampling-theory refinements that one finds appearing from time to time in the geophysical (especially climatological) literature, the final conclusions of the survey proved to be of surprisingly different nature. These conclusions were:

i) The problem of normality has probably, for most types of geophysical problems, been grossly overemphasized, for experimental sampling studies yield evidence of non-normality effects which, though they have sometimes been regarded as disturbing by the statisticians, are of geophysically quite inconsequential magnitude.

ii) The geophysicist who wishes to test for mere existence of correlation seldom need feel strong obligation to use even so easily applied a refinement as the t-test, but may as well simply quote the sample estimate of the standard error of r, for the differences in inferences derivable by turning to a t-test in place of the old rule-of-thumb cited here are not differences that carry appreciable meaning in the face of the usual level of precision of the basic data in most problems where statistical techniques are used.

iii) If a substantially non-zero value of r is found in a study, and it is desired to suggest confidence limits, use of the Fisher z'-transformation will usually be a wholly unwarranted touch of elegance. Merely cite the sample value of σ_r, and draw relevant inferences on the basis of normal-distribution arguments, realizing that all such considerations are aimed only at providing a rough guess of how uncertain the sample value of r may be by virtue of the vagaries of sampling fluctuations and that z' leads to quantitatively unimportant modifications.

Put in a few words, the recommendations made above are a plea for elimination of what can often be shown to be irrelevant refinements but which when cast in the language of significance testing give what will ordinarily be an unreal appearance of precision. The most objectionable consequence of unnecessary use of statistical elaboration is that it forces younger investigators to learn about concepts that will seldom be of real assistance to them in their work. One surely needs to know the elements of normal-distribution theory for profitable exploitation of statistics in geophysics, but that plus related elements of classical statistics should be sufficient until one turns to much more recent developments in non-parametric statistics.

The geophysicist who wishes to make use of these recent developments of non-parametric statistics may consult the somewhat overwhelming bibliography of Savage (1953), or should examine a recent general treatment by Siegel (1956); but as with other statistical-methods questions he must, regrettably, be prepared for a good deal of bewilderment as to which of four or five alternative tests will be most appropriate. A critique of non-parametric tests of particular geophysical utility is sorely needed.

B. Rank-correlation. One interesting non-parametric statistic (though not a new one) is the Spearman rank-difference correlation coefficient, r_s. It has been suggested above that since r_s is only slightly less efficient than r for the normal-distribution case, since r is known to be not the most efficient correlation statistic for the non-normal case the geophysicist almost always deals with,since ranks may be almost a more rational choice than three- or four-digit numbers in many geophysical analyses, since ranking methods obviate all perplexities arising from correlating quantities that are themselves non-linear in functional nature, and since a modest sample of comparisons presented here for actual precipitation data show r_s to be very close to r, it would seem to follow that wider use should be made of rank correlation in geophysics. This is expecially true for sample sizes under about thirty or forty, for which the manual labor of computation is sure to be much less for r_s than for r; and even with one type of punchcard equipment r_s was found to be just slightly more easily calculated than r.

C. Tetrachoric correlation. Finally, the tetrachoric correlation coefficient has been shown to be a substantially poorer estimator of r than is r_s for the actual climatological data here considered, but its utility, above all for very rapid testing of merely the null hypothesis of independence, must be recognized.

D. Other statistics. Although correlation statistics have been exclusively dealt with here, it is suggested that objections to almost meaningless overrefinement of statistical techniques of other types are readily found. The geophysicist who tests significance of the difference between means via the t-test or of the comparative sizes of variances via the F-test is altogether correct, of course, but he should ask whether he is unwittingly elaborating numerical manipulations that yield but slight practically important difference from what would be obtained by an old-fashioned standard-error argument that will be better understood by many readers. One must admit that there is a true gain in rigor in replacing standard-error approximations by t-tests and related tests, but the writer objects that this rigor is almost always out of place in the very situations where theory calls for it.

The kind of statistical elaborations which are truly needed in many branches of geophysical research would seem to be those in which new and revealing syntheses of several different statistical techniques are achieved. A single case from the field of hydrology will suffice to illustrate the point: The development of statistical rainfall-runoff prediction relationships discussed by Kohler and Linsley (1949) seems to the writer to exemplify the type of statistical syntheses and refinements of procedure that can yield practically significant gains. More of this and less of empty refinement is surely to be encouraged.

ACKNOWLEDGEMENTS.

The present paper is one portion of a study being carried out with the support of the Alfred P. Sloan Foundation, whose aid is gratefully acknowledged.

Mr. Robert P. DesJardins did the punchcard calculations of the product-moment and rank correlations reported here. Mr. R. W. Mitchell carried out a number of miscellaneous computations contributing to this study. The writer is indebted to Dr. Henry Tucker for a number of very helpful discussions of general statistical aspects of the correlation problem.

REFERENCES.

Arkin, H., and R. R. Colton, 1955: An outline of statistical methods (4th ed., rev.). New York, Barnes and Noble, 226 pp.

Brooks, C. E. P., and N. Carruthers, 1953: Handbook of statistical methods in meteorology. M. O. 538, Air Ministry. London, H. M. Stationery Off., 412 pp.

Chesire, L., E. Oldis, and E. S. Pearson, 1932: Further experiments on the sampling distribution of the correlation coefficient. J. Am. Stat. Assn., 27, 121-128.

Conrad, V. and L. W. Pollak, 1950: Methods in climatology (2d ed.). Cambridge, Harvard Univ. Press, 459 pp.

Dixon, W. J., and F. J. Massey, 1951: Introduction to statistical analysis. New York, McGraw-Hill, 370 pp.

Haldane, J. B. S., 1949: A note on non-normal correlation. Biometrika, 36, 467-468.

Hotelling, H., and M. R. Pabst, 1936: Rank correlation and tests of significance involving no assumptions of normality. Ann. Math. Stat., 7, 29-43.

Kelley, T. L., 1947: Fundamentals of statistics. Cambridge, Harvard Univ. Press, 755 pp.

Kendall, M. G., 1948a: The advanced theory of statistics, Vol. I. London, Charles Griffin, 457 pp.

Kendall, M. G., 1948b: Rank correlation methods. London, Charles Griffin, 160 pp.

Kohler, M. A., and R. K. Linsley, Jr., 1949: Recent developments in water supply forecasting from precipitation. Trans. Am. Geophys. Un., 30, 427-436.

Mills, F. C., 1955: Statistical methods (3d ed.). New York, Henry Holt, 842 pp.

Moran, P. A. P., 1951: Partial and multiple rank correlations. Biometrika, 38, 26-32.

Moroney, M. J., 1953: Facts from figures (2d ed.). London, Penguin Books, 427 pp.

Ostle, B., 1954: Statistics in research. Ames, Iowa, Iowa State Coll. Press 487 pp.

Pearson, E. S., 1931: The test of significance for the correlation coefficient. J. Am. Stat. Assn., 26, 128-134.

Peters, C. C., and W. R. Van Voorhis, 1940: Statistical procedures and their mathematical bases. New York, McGraw-Hill, 516 pp.

Savage, I. R., 1953: Bibliography of non-parametric statistics and related topics. J. Am. Stat. Assn., 48, 844-906.

Siegel, S., 1956: Nonparametric statistics. New York, McGraw-Hill, 312 pp.

Smith, J. G., and A. J. Duncan, 1945: Sampling statistics and applications. New York, McGraw-Hill, 498 pp.

Snedecor, G. W., 1946: Statistical methods (4th ed.). Ames, Iowa, Iowa State Coll. Press, 485 pp.

Stidd, C. K., 1953: Cube-root normal precipitation distributions. Trans. Am. Geophys. Un., 34, 31-35.

Walker, H. M., and J. Lev, 1953: Statistical inference. New York, Henry Holt, 510 pp.

Yule, G. U., and M. G. Kendall, 1950: An introduction to the theory of statistics (14th ed.). New York, Hafner, 701 pp.