MEASURES OF ASSOCIATION FOR CROSS CLASSIFICATIONS. II: FURTHER DISCUSSION AND REFERENCES*

LEO A. GOODMAN AND WILLIAM H. KRUSKAL
University of Chicago

---

* This research was supported in part by the Army, Navy, and Air Force through the Joint Services Advisory Committee for Research Groups in Applied Mathematics and Statistics, Contract No. N6ori-02035; and in part by the Office of Naval Research. This paper, in whole or in part, may be reproduced for any purpose of the United States Government. Part of Mr. Kruskal's work was done at the Department of Statistics, University of California, Berkeley.

We wish to thank the following persons for their generous suggestions and criticisms: J. Berkson (Mayo Clinic), C.E. Bonferroni (University of Florence), G. W. Brier (U. S. Weather Bureau), H. Byers (University of Chicago), F. R. Dodge (Berkeley, Calif.), H. E. Driver (Indiana University), O. D. Duncan (University of Chicago), C. Eisenhart (National Bureau of Standards), L. Faleschini (University of Milan), R. Féron (University of Paris), M. Fréchet (University of Paris), C. Gini (University of Rome), I. I. Gringorten (Air Force Cambridge Research Center), L. Guttman (Israel Institute of Applied Research), C. Jordan (University of Budapest), A. L. Kroeber (University of California), P. F. Lazarsfeld (Columbia University), H. V. Roberts (University of Chicago), T. Salvemini (University of Rome), I. R. Savage (University of Minnesota), R. H. Somers (Columbia University), J. W. Tukey (Princeton University), M. Uridge (University of California), W. van der Bijl (Kansas State College), E. W. Wahl (Air Force Cambridge Research Center), H. Walker (Teachers College, Columbia University), W. A. Wallis (University of Chicago), M. A. Woodbury (New York University).

Our earlier discussion of measures of association for cross classifications [66] is extended in two ways. First, a number of supplementary remarks to [66] are made, including the presentation of some new measures. Second, historical and bibliographical material beyond that in [66] is critically surveyed; this includes discussion of early work in America by Doolittle and Peirce, early work in Europe by Körösy, Benini, Lipps, Deuchler and Gini, more recent work based on Shannon-Wiener information, association measures based on latent structure, and relevant material in the literatures of meteorology, ecology, sociology, and anthropology. New expressions are given for some of the earlier measures of association.

1. INTRODUCTION AND SUMMARY

This paper has two purposes. First, we wish to present a supplementary discussion to problems considered in our first paper on cross classifications [66], including presentation of some new measures; this is Section 2 of the present paper. Second, we wish to extend the brief historical and bibliographical remarks in [66] to include a number of publications, many of them little-known, that may be of interest to those working with cross classifications; this is done in Sections 3 and 4 of the present paper.

We have in preparation a paper on approximate distributions for the sample analogues of the measures of association described in [66], but it seems desirable to bring the present remarks, virtually none of which deal with sampling distributions, to the reader's attention in a separate report.

The literature on measures of association for cross classifications is vast, it is poorly integrated, and seldom in this literature are meaningful interpretations of measures adduced. One finds the same questions discussed in papers on meteorology, anthropology, ecology, sociology, etc. with hardly any cross references and with considerable duplication. In surveying this literature, we have been selective, although the length of this paper may not suggest it. Discussion of a measure of association here does not mean *ipso facto* that it has an operational interpretation, a very desirable characteristic for which we argued in [66], but may simply reflect some other interesting aspect of the measure, for example its historical role.

One may organize the historical and bibliographical material in various ways, classifying by date, by type of measure, by substantive field, and so on. We have used a gross chronological division, but within it we have classified in several ways, as seemed most appropriate. Material from [66] has not been repeated here.

2. SUPPLEMENTARY DISCUSSION TO PRIOR PAPER

2.1. *Cross classifications in which the diagonal is not of interest.* Herbert Goldhamer (Rand Corporation) has been concerned with measuring association for $\alpha \times \alpha$ tables where the classes are the same for the two polytomies, as in Section 8 of [66], but where the diagonal entries are of little or no interest. For example, one might tabulate occupation of father against occupation of son, and investigate the association between the two occupations only in the off-diagonal subpopulation where they are not the same. Thus the situation, while similar to that of reliability measures, as in Section 8 of [66], differs from it in that the diagonal entries must not play a part; and hence λ_r of [66] would not be suitable.

It seems to us that reasonable measures of association in this situation would be provided by λ_a, λ_b, or λ in the unordered case, and by γ in the ordered case, when these measures are applied to the conditional classification with all $\rho_{aa}=0$. Hence, replacing ρ_{ab}, for $a \neq b$, by $\rho_{ab}/(1-\sum \rho_{aa})$, and taking all $\rho_{aa}=0$, we would get a new table for which the λ's or γ would have direct conditional interpretations. This kind of simple modification is often easy to make for measures with operational interpretations, whereas it is not at all clear how one might usefully alter a chi-square-like measure to fit Goldhamer's problem. A similar point is made in another context in Section 4.13.

2.2. *A relation between the λ measures and Yule's Y.* Suppose that in the 2×2 case we make a transformation of form $\rho_{ab} \rightarrow s_a t_b \rho_{ab}$ so that all the marginals become .5 [66, Sec. 5]. Then, for the altered table, $\lambda_a = \lambda_b = \lambda$, and all three are equal to the absolute value of Y, where

$$Y = \frac{\sqrt{\rho_{11}\rho_{22}} - \sqrt{\rho_{12}\rho_{21}}}{\sqrt{\rho_{11}\rho_{22}} + \sqrt{\rho_{12}\rho_{21}}} \qquad (\rho\text{'s of original table})$$

as described in Section 4 of [66]. The actual transformation is that for which

$$(s_1 \colon s_2 \colon t_1 \colon t_2) = (\sqrt{\rho_{21}\rho_{22}} \colon \sqrt{\rho_{11}\rho_{12}} \colon \sqrt{\rho_{12}\rho_{22}} \colon \sqrt{\rho_{11}\rho_{21}}).$$

Thus we have another formal identity in the 2×2 case between a classical measure of association and one with an operational interpretation.

2.3. *Symmetrical variant of proportional prediction.* In Section 9 of [66], we mentioned a measure of association based, not on optimal prediction, but on proportional prediction in a manner there explained. If one predicts polytomy B half the time and polytomy A the other half, always using proportional prediction, then the relative decrease in the proportion of incorrect predictions, as one goes from the nothing-given situation to the other-polytomy-category-given situation, is

$$\frac{\frac{1}{2} \sum_a \sum_b \left\{ (\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2 (\rho_{a\cdot} + \rho_{\cdot b}) / (\rho_{a\cdot}\rho_{\cdot b}) \right\}}{1 - \frac{1}{2} \sum_a \rho_{a\cdot}^2 - \frac{1}{2} \sum_b \rho_{\cdot b}^2}.$$

In the 2×2 case this quantity, together with the asymmetrical τ_b of [66], reduces to

$$\frac{(\rho_{11}\rho_{22} - \rho_{12}\rho_{21})^2}{\rho_{1\cdot}\rho_{2\cdot}\rho_{\cdot 1}\rho_{\cdot 2}},$$

or ϕ^2, the mean square contingency.

2.4. *Association with a particular set of categories.* In Section 10 of [66], we described a simple way to consider association between a *particular* A category and the B polytomy; namely coalescence of the $\alpha\times\beta$ table into a $2\times\beta$ table whose rows correspond to the particular A category and its negation respectively. A similar suggestion was made by Karl Pearson in 1906 [112].

We now discuss association between a particular set of A categories and the B polytomy. Suppose that we want to consider the association between $A_{a_1}, A_{a_2}, \cdots, A_{a_s}$, a specific set of A classes, and the B polytomy. One possible approach is to condense all the A_a rows that are not in the specific set of A classes (*i.e.*, all the A_a rows where a is not equal to any a_k, $k=1, 2, \cdots, s$) into a single row, thus obtaining an $(s+1)\times\beta$ table, and then apply whatever measure of association is thought appropriate. This approach might be used if the entire original population is of interest, and we are only concerned with association for the specific set of A categories and their pooled remainder. If, however, the population of interest consists *only* of those individuals who are in the specific set of A categories, $A_{a_1}, A_{a_2}, \cdots, A_{a_s}$, $(s\geq 2)$, then we would apply whatever measures of association are thought appropriate (e.g., λ_a, λ_b, λ, γ, etc.) to the conditional classification with $\rho_{ab}=0$ for all a that are not equal to any a_k, $k=0, 1, \cdots, s$. That is, we would delete all rows except those corresponding to $A_{a_1}, \cdots, A_{a_s}$, and in those rows we would replace ρ_{ab} by $\rho_{ab}/\sum_{k=1}^{s}\sum_{b=1}^{\beta}\rho_{a_k b}$. We would then have an $s\times\beta$ table, and the λ's or γ would have direct conditional interpretations.

The association between a particular set of A categories and a particular set of B categories, or a particular set of combined (grouped) A categories and a particular set of combined B categories, can be treated in an analogous manner.

2.5. *Comparison of degrees of association exhibited by two cross classifications.* Sometimes one wishes to compare the degrees of association shown by two cross-classified populations. This question is particularly likely to arise when the two classifications are the same for both populations. It was discussed briefly on page 740 of [66]; a bit more detail may be of interest here.

Suppose, for example, that we are considering two populations, each cross classified by the same pair of polytomies and such that λ_b is the appropriate measure of association. That is, the relative decrease in probability of error for optimum prediction of column, as we go from the case of row unknown to that of row known, is the relevant population characteristic. Then the difference between the λ_b's of the two populations gives a simple comparison with a clear meaning. Sometimes the *relative* difference between the λ_b's might be of more interest.

If the pairs of classifications for the two populations are not identical, as will necessarily be the case when the two cross classification tables are of different sizes, the purpose of comparison may not be clear. However, the absolute or relative differences described above may still be used and have perfectly definite interpretations. Of course, the above comments are applicable, not only to λ_b, but to any other measure of association that has an operational meaning.

When we are concerned with sampling problems, the question may arise whether two sample values of λ_b (say) from two different populations differ with statistical significance. This question, together with other questions relating to sampling, will be considered in a paper now in preparation. In that paper we shall also discuss the question of whether K sample values of λ_b from K different populations ($K \geq 2$) differ with statistical significance.

2.6. *A new measure of association in the latent structure context.* Several measures of association discussed in Sections 3 and 4 are based upon probabilistic models of a latent structure nature. This kind of model is explained and discussed in Section 4.9, and there we suggest a new measure in addition to those already suggested by others.

2.7. *Two corrections.* The second and third sentences of the second paragraph of [66], p. 758, are essentially correct, but may be misleading. It would have been clearer to have written

> It [λ_r] takes the value -1 if and only if (i) all ρ_{ab}'s not in the row or column of the modal class are zero, and (ii) ρ_{aa} for the modal class is not one. It takes the value 1 if and only if (i) $\Sigma\rho_{aa}=1$ (i.e. the two methods always agree), and (ii) ρ_{aa} for the modal class is not one.

Formula (6) on p. 740 of [66] should have contained a radical in the denominator, so that the correct formula is

$$T = \sqrt{[\chi^2/\nu]/\sqrt{(\alpha-1)(\beta-1)}}.$$

We thank Vernon Davies (Washington State) for calling this to our attention, and we apologize to him and to other readers for an erroneous corrigendum about this point on page 578 of the December 1957 issue of this *Journal*, in which a solidus was missing before the inner radical.

3. WORK ON MEASURES OF ASSOCIATION IN THE LATE NINETEENTH AND EARLY TWENTIETH CENTURIES

3.1. *Doolittle, Peirce, and contemporary Americans; Köppen.* In the 1880's, interest arose in American scientific circles regarding measures of association. Such eminent men as M. H. Doolittle, of Doolittle's method, and C. S. Peirce, the well-known logician and philosopher, took part in the discussion.

Apparently it began with the publication [47] by J. P. Finley, Sergeant, Signal Corps, U.S.A., of his results in attempting to predict tornadoes. During four months of 1884, Finley predicted whether or not one or more tornadoes would occur in each of eighteen areas of the United States. The predictions generally covered certain eight-hour periods of the day. One of Finley's summary tables is given below as an example.

COMPARISON OF FINLEY TORNADO PREDICTIONS AND OCCURRENCES, APRIL, 1884. SOURCE: [47, p. 86]

TABLE SHOWS FREQUENCIES OF TIME PERIOD—GEOGRAPHICAL AREA COMBINATIONS IN EACH CELL

		Occurrence: Tornado	Occurrence: No Tornado	Totals
Prediction	Tornado	11	14	25
	No Tornado	3	906	909
	Totals	14	920	934

Thus, for example, in 14 out of the 934 time period-geoegraphical area combinations considered, one or more tornadoes occurred; out of these 14, Finley predicted 11. Since Finley's predictions were correct in 917 out of 934 cases he gave himself a percentage score of 100 (917/934) = 98.18 per cent.* Thus he used the diagonal sum mentioned in Section 8 of [66].

This score, as a measure of association between prediction and occurrence, is wholly inappropriate for Finley's study. A completely ignorant person could always predict "No Tornado" and easily attain scores equal to or greater than Finley's; in the above example, always predicting "No Tornado" would give rise to a score of 100(920/934) = 98.50 per cent. (Of course it is clear that Finley did appreciably better than chance; the question is that of measuring his skill by a single number.)

It was not long before Finley was taken to task. G. K. Gilbert [55] pointed out the fallacy and suggested another procedure, prefacing his suggestion, with commendable humility, in the following words:

> "It is easier to point out an error than to enunciate the truth; and in matters involving the theory of probabilities the wisest are apt to go astray. The following substitute for Mr. Finley's analysis is therefore offered with great diffidence, and subject to correction by competent mathematicians." [55, p. 167]

If Finley's table is written in terms of proportions rather than frequencies, and in the notation of [66], it is of form

		Occurrence: Tornado	Occurrence: No Tornado	Total
Prediction	Tornado	ρ_{11}	ρ_{12}	$\rho_{1\cdot}$
	No Tornado	ρ_{21}	ρ_{22}	$\rho_{2\cdot}$
	Total	$\rho_{\cdot 1}$	$\rho_{\cdot 2}$	1

Gilbert suggests that a sensible index of prediction success would be the quantity

* Finley actually obtained such percentage scores for each geographical area separately and then averaged the scores. For April the average was 98.51 per cent.

$$\frac{\rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}}{\rho_{1\cdot} + \rho_{\cdot 1} - \rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}},$$

and he lists a number of formal properties that this index has. For example, it is ≤ 1; it is zero when $\rho_{11} = \rho_{1\cdot}\rho_{\cdot 1}$; it has desirable monotonicities; etc. Finally Gilbert mentions the difficulties of extending his index to prediction problems with more than two alternatives. H. A. Hazen [77] criticized Gilbert's paper, and suggested an alternative index of predictive success based upon a weighted scoring scheme that gave decreasing credit to occurring tornadoes as they fell further from the center of the predicted region.

In the same year that Gilbert's paper appeared, C. S. Peirce [115] suggested a much less ad hoc index of prediction success. Peirce pointed out that one could think of the observed results as obtained by using an infallible predictor a proportion θ of the time, and a completely ignorant predictor the remaining proportion $1-\theta$ of the time. The infallible predictor predicts "Tornado" if and only if a tornado will occur. The ignorant predictor uses an extraneous chance device that precicts "Tornado" with frequency ψ and "No Tornado" with frequency $1-\psi$. Thus what we are asked to contemplate is a mixture of the two 2×2 sets of probabilities

$\rho_{\cdot 1}$	0
0	$\rho_{\cdot 2}$

,

$\rho_{\cdot 1}\psi$	$\rho_{\cdot 2}\psi$
$\rho_{\cdot 1}(1-\psi)$	$\rho_{\cdot 2}(1-\psi)$

with weights θ and $1-\theta$ respectively. The meanings of the four cells in these tables are the same as in the preceding tables.

The mixed table is, therefore,

$\theta\rho_{\cdot 1}+(1-\theta)\rho_{\cdot 1}\psi$	$(1-\theta)\rho_{\cdot 2}\psi$
$(1-\theta)\rho_{\cdot 1}(1-\psi)$	$\theta\rho_{\cdot 2}+(1-\theta)\rho_{\cdot 2}(1-\psi)$

and Peirce inquires what values of θ and ψ will reproduce the actually observed 2×2 table. (Note that for any θ and ψ the column marginals of the mixed table are $\rho_{\cdot 1}$ and $\rho_{\cdot 2}$.)

For this approach to make sense, θ and ψ must be uniquely defined in terms of the actual ρ_{ab} table. From the (1, 2) cell, we require

$$(1 - \theta)\psi = \rho_{12}/\rho_{\cdot 2},$$

whence from the (1, 1) cell

$$\theta = \frac{\rho_{11}\rho_{22} - \rho_{12}\rho_{21}}{\rho_{\cdot 1}\rho_{\cdot 2}} = \frac{\rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}}{\rho_{\cdot 1}\rho_{\cdot 2}},$$

and

$$\psi = \frac{\rho_{12}\rho_{\cdot 1}}{\rho_{12}\rho_{\cdot 1} + \rho_{21}\rho_{\cdot 2}}.$$

Substitution shows that these values form a unique solution. The only difficulty occurs when θ is negative, for then it can scarcely be a probability. θ itself is suggested as the measure of association in the sense of prediction success. Note that

$$\theta = \frac{\rho_{11}}{\rho_{\cdot 1}} - \frac{\rho_{12}}{\rho_{\cdot 2}},$$

or the difference between the conditional columnwise probabilities of a tornado prediction.

If $\theta = 1$, prediction is considered as good as possible, since it is equivalent to infallible prediction. If $\theta = 0$, prediction is as poor as it can be without being perverse, since it is equivalent to randomized prediction using the row marginal frequencies of the table under investigation; that is, it corresponds to independence. Further, the θ that makes $\theta\rho_{\cdot 1} + (1-\theta)\rho_{\cdot 1}\psi$ equal to ρ_{11} has an operational interpretation in terms of a hypothetical, if perhaps far-fetched, model of activity. As θ increases, prediction improves.

This proposal by Peirce is of a kind that may be called latent structure measures. We discuss this kind of measure later on in Section 4.9. Peirce's measure, θ, was independently proposed and differently motivated by W. J. Youden in 1950 [66, p. 745, footnote].

Peirce mentions the extension of his approach to larger tables but gives no details. He concludes by suggesting another index that takes into account the "profit, or saving, from predicting a tornado, and . . . the loss from every unfulfilled prediction of a tornado (outlay in preparing for it, etc.). . . ." Thus Peirce, writing in 1884, is the first person of whom we know to discuss the measure of association problem with the intent of giving operationally meaningful measures. Of course, further study might bring earlier proposals to light.

Very soon after Peirce's letter appeared, M. H. Doolittle [35] discussed the topic at the December 3, 1884, meeting of the Mathematical Section of the Philosophical Society of Washington. Doolittle argued for a symmetrized version of Peirce's index, suggesting on rather ad hoc grounds the product of the two possible asymmetrical Peirce quantities

$$\frac{\rho_{11}\rho_{22} - \rho_{12}\rho_{21}}{\rho_{\cdot 1}\rho_{\cdot 2}}, \qquad \frac{\rho_{11}\rho_{22} - \rho_{12}\rho_{21}}{\rho_{1\cdot}\rho_{2\cdot}}.$$

This product is simply the mean square contingency, and may be the first occurrence of this chi-square-like index. Doolittle also alluded to the difficulty of extending such measures beyond the 2×2 case.

At a subsequent meeting of the Mathematical Section (February 16, 1887), Doolittle [36] continued his discussion in more general terms than those of measures of prediction success alone. His discussion is similar at points to that of Yule's 1900 paper [149] and he attempts to develop a rationale for the quantity we call the mean square contingency; Doolittle called it the discriminate association ratio. At a third meeting (May 25, 1887), Doolittle [36] concluded his discussion with a criticism of Gilbert's criticism of Finley.

We cannot forbear presenting a quotation from Doolittle in which he struggles to state verbally the general approach he favors.

"The general problem may be stated as follows: Having given the number of instances respectively in which things are both thus and so, in which they are thus but not so, in which they are so but not thus, and in which they are neither thus nor so, it is required to eliminate the general quantitative relativity inhering in the mere thingness of the things, and to determine the special quantitative relativity subsisting between the thusness and the soness of the things." [36, p. 85]

What is a reasonable measure of prediction success for Finley's tables in terms of our λ measures? In this case, λ_b is zero, reflecting the fact that knowledge of Finley's prediction would be no better than ignorance of it in predicting a tornado. If, however, we adjust Finley's table so that the column marginals are equal, while conditional column frequencies remain unchanged, we obtain $\lambda_b^* = .67$. This means that if Finely's prediction method were used in a world in which tornadoes occur half the time, we could reduce the error of prediction 67% by knowing Finley's prediction as against not knowing it. We might go further and make both column and row marginals equal, obtaining $\lambda_b^* = .88$. The interpretation of this is the same as before except that now Finley is allowed to use the knowledge that tornadoes occur half the time, so that he will predict a tornado half the time.

It may, of course, be cogently argued that in situations such as Finley's it is misleading to search for a single numerical measure of predictive success; and that rather the whole 2×2 table should be considered, or at least two numbers from it, the proportions of false positives and false negatives.

We conclude this section by mentioning briefly some suggestions made by German meteorologists at about the same time. As early as 1870, W. Köppen had considered association measures in connection with his study of the tendency of meteorological phenomena to stay fixed over time. This is related to the problem of measuring prediction, although it is not quite the same. Köppen's basic article on the topic appears to be [91]; an exposition is given by H. Meyer [108, Chapters 11 and 13] together with further references. Köppen and Meyer discuss the question of measuring constancy in various contexts; one relates to a 2×2 table with both classifications the same but referring to different times, and with the two marginal pairs of frequencies the same. For example, the table might be of the following form:

		Wind at 2 P.M. at an observation station		
		North	Not North	
Wind at preceding 8 A.M. at the observation station	North	ρ_{NN}	$\rho_{N\bar{N}}$	ρ_N
	Not North	$\rho_{\bar{N}N} = \rho_{N\bar{N}}$	$\rho_{\bar{N}\bar{N}}$	$1 - \rho_N$
		ρ_N	$1 - \rho_N$	1

In this case Köppen (as we interpret his discussion) suggests measuring constancy of wind direction between 8 A.M. and 2 P.M., with respect to the dichotomy North vs. Not North, by

$$\frac{\rho_N(1-\rho_N)-\rho_{N\bar{N}}}{\rho_N(1-\rho_N)},$$

or the difference between the probability of a change from North under independence and the same actual probability, this difference taken relative to the probability under independence.

In 1884, an article either by Köppen or someone probably influenced by him [92] suggested a measure of reliability between meteorological prediction and later occurrence in the 3×3 ordered case. The measure was

$$\sum \rho_{aa} + \frac{1}{2}\sum_{|a-b|=1}\sum \rho_{ab},$$

as in Section 8.3 of [66]. The next year, H. J. Klein [89] discussed the simpler measure $\sum \rho_{aa}$ in the general $\alpha\times\alpha$ reliability case.

Bleeker [10] presents an analytical survey of the above early American and German suggestions in the field of meteorological prediction, together with a discussion of many other papers. In Section 4.10 of this paper we survey more recent uses of association measures in meteorology.

3.2. *Körösy, Jordan, and Quetelet.* In [85], Charles Jordan discusses measures of association introduced by József Körösy in the late nineteenth century. Körösy wrote extensively on the effectiveness of smallpox vaccination, and he was led to introduce various measures of association for 2×2 tables in order to summarize and interpret his large quantities of data. Among the several measures discussed by Körösy for 2×2 tables, at least one is equivalent to Yule's Q and hence to our γ (see [66].)

Jordan [85] extends one of Körösy's measures to $\alpha\times\beta$ tables. In our notation, the extended measure is found as follows. For a 2×2 table, Körösy had proposed $(\rho_{11}\rho_{22})/(\rho_{12}\rho_{21})$ as a natural measure of association. Jordan suggests forming all possible $\alpha\beta$ pooled 2×2 tables out of an $\alpha\times\beta$ table, each of form

ρ_{ab}	$\rho_{a\cdot}-\rho_{ab}$
$\rho_{\cdot b}-\rho_{ab}$	$1-\rho_{a\cdot}-\rho_{\cdot b}+\rho_{ab}$

and averaging the corresponding 2×2 measures to obtain an over-all measure.

Jordan states in [85] the maximum value for the mean square contingency coefficient, ϕ^2. (Jordan also gives this maximum value in another related paper, [86]. The same maximum value has also been given by Cramér [66, p. 740].) Jordan further discusses Körösy's proof and use of the fact that, if in a 2×2 table we observe only a proportion of individuals in a column (i.e., if there is a probability of selection), then, providing the selection probabilities in the two cells of the column are equal, Yule's Q and Körösy's equivalent measure are unaffected. This property of Q is emphasized by Yule in [149] and [150]. Finally, Jordan asserts priority for Körösy's work in the following terms: "Le mérite de Körösy consiste a avòir introduit et utilisé on 1887, c.-à-d. avant l'avènement de la Statistique Mathématique, des grandeurs, mesurant l'association, en bon accord avec les coefficients δ et Q de Yule et ϕ^2 de Pearson utilisés aujourd'hui."

Körösy's writings are not readily available, and we have consulted only one of them [93]. This is a very interesting and sophisticated discussion of statistical material on the efficacy of smallpox vaccination, in which Körösy uses extensively 2×2 table coefficients of association. Emphasis is on the interpretation of such material and on the many ways in which vaccination and smallpox statistics might be consciously or unconsciously distorted, falsified, and biased. (On page 221 of the same volume in which [93] appears, there begins a fascinating discussion of a case of falsification of smallpox-vaccination data. The culprit was an anti-vaccinationist, and the detective work was done by Körösy.)

The question of priority in the use of simple measures of association for 2×2 tables scarcely seems very important. However, it may be of historical interest to note that Yule, in his first (1900) paper on the subject [149] speaks of Quetelet's use of a measure of association in 2×2 tables: $(\rho_{11}-\rho_{1\cdot}\rho_{\cdot1})/(\rho_{1\cdot}\rho_{\cdot1})$, in our notation.[1] In fact, Yule named his coefficient "Q" after Quetelet [150, p. 586]. The work by Quetelet of which Yule writes is not accessible to us, but in another place [119] Quetelet uses another very natural measure for comparing (say) the two rows of a 2×2 table in a case wherein they correspond to two binomial populations. He simply takes the ratio of the two binomial p's: $(\rho_{11}/\rho_{1\cdot})/(\rho_{21}/\rho_{2\cdot})$. This ratio probably has been used since nearly the beginning of arithmetic. Of course, neither of the two measures last mentioned have the symmetry of chi-square or of Yule's Q, so that perhaps Jordan would say that they are not "en bon accord" with the measures of Yule and Pearson.

Biographical, bibliographical, and appreciative material on Körösy may be found in a book by Saile [121] and in an obituary by Thirring [134]. A more recent paper by Jordan on the general question of association measures is [87].

3.3. *Benini.* In 1901, the Italian demographer and statistician, R. Benini [4, pp. 129 ff.] suggested measures of attraction and repulsion for 2×2 tables in which the categories of the two dichotomies were the same, or closely related. Benini was mainly concerned at this time with the association between dichotomous characteristics of husband and wife among married couples, for example the association between categories of civil status. Among marriages in Italy during 1898, Benini gives the following 2×2 breakdown of premarital civil status (in relative frequencies):

		Wife: Unmarried	Wife: Widow	Totals
Husband	Unmarried	.8668	.0275	.8943
	Widower	.0742	.0315	.1057
	Totals	.9410	.0590	1

[1] Note that this is the same as the suggestion by Köppen mentioned in Section 3.1.

Comparing this with the corresponding "chance" table obtained by multiplying marginal frequencies, Benini observed that there clearly was association between the premarital civil statuses of husband and wife. To measure the attraction between similar premarital civil statuses, he suggested the following measure (our notation):

$$\frac{\rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}}{\text{Min}\,(\rho_{1\cdot}, \rho_{\cdot 1}) - \rho_{1\cdot}\rho_{\cdot 1}} = \frac{\rho_{22} - \rho_{2\cdot}\rho_{\cdot 2}}{\text{Min}\,(\rho_{2\cdot}, \rho_{\cdot 2}) - \rho_{2\cdot}\rho_{\cdot 2}}$$

on the grounds that, when the numerator is nonnegative, the denominator gives the maximum possible value of the numerator for fixed marginals. The numerator is the usual quantity on which 2×2 measures of association are based. When the numerator is negative, a slight revision of the formula provides Benini's measure of repulsion. In the above example Benini's measure of attraction has the value

$$\frac{.8668 - .8415}{.8943 - .8415} = \frac{253}{528} = .479.$$

In 1928, Benini [5] extended his method of analysis by suggesting a separation of the 2×2 population into two 2×2 subpopulations, one with two cells empty, and the other with all marginal frequencies equal to 1/2. Then his measure of attraction (or repulsion) would be computed only for the second sub-population. This represents one way of eliminating the effect of unequal marginals in comparing several 2×2 populations. (In Section 5.4 of [66] another way of attaining this goal was briefly discussed.) A variation of this point of view, much akin to latent structure analysis (see Section 4.9), was applied by Benini to sex-ratios in twins in order to estimate the fractions of fraternal and identical twins in the population.

Benini's work has been discussed by a number of Italian statisticians. An early discussion was by Bresciani in 1909 [15]. A. Niceforo [110, pp. 383–91] and [111, pp. 462–8] also considers Benini's suggestions, and provides an entertaining discussion, with many examples, of several aspects of cross classifications. We refer in particular to Chapter 16 of [111]. A lengthy critical analysis of Benini's suggestions, as applied to matrimonial association, was given by R. Bachi [3]. Some further articles dealing with Benini's work are those of G. de Meo [31], F. Savorgnan [126], G. Andreoli [2], and C. E. Bonferroni [13]. Benini's first measure of attraction was independently suggested by Jordan [87] in 1941, by H. M. Johnson [84a] in 1945, and by L. C. Cole [28] in 1949. No doubt there have been many other independent suggestions of this measure. It has been frequently used by psychologists and sociologists in recent years and called, descriptively enough, $\phi/\phi_{\max}$.

Benini's first measure has recently been critically reviewed by D. V. Glass, J. R. Hall, and R. Mukherjee [63a, pp. 195–96, 248–59] in a book by these writers and others on social mobility in Britain. Glass et al. deal mostly with $\alpha\times\alpha$ cross classifications of father vs. son occupational status; their general approach is to construct a number of 2×2 condensed cross classifications from a larger $\alpha\times\alpha$ one, with the condensed dichotomies of form father (son) in occupational status a vs. not in status a. Then the 2×2 condensations are examined by looking at three of the ratios $\rho_{ab}/(\rho_{a\cdot}\rho_{\cdot b})$.

3.4. *Lipps.* In 1905, G. F. Lipps [100] discussed various ways of describing dependence in a two-way cross classification. For the 2×2 case, Lipps independently proposed Yule's Q. For larger tables, Lipps points out that $(\alpha-1)\cdot(\beta-1)$ numbers are required to describe the dependence in full; he argues againt use of a single numerical measure of association in these words: "Es ist demzufolge nicht zulässig (ausser wenn $r=s=2$ [$\alpha=\beta=2$ in our notation]) einen einzigen Wert als schlechthin gültiges Mass der Abhängigkeit aufzustellen" [100, p. 12]. However, he refers, in a footnote, to articles on correlation by Galton and K. Pearson in contradistinction.

It is interesting to notice that, in the last section of his paper, Lipps proposed a quantity equivalent to Kendall's rank correlation coefficient, τ. The quantity Lipps proposed is Kendall's $P=n(n-1)(\tau+1)/4$ where n is sample size. Lipps suggested testing for independence by this quantity, and to implement this he computed its mean and variance under the hypothesis of independence. A year later Lipps [101] discussed $2P-\binom{n}{2}=\binom{n}{2}\tau$. Material on Lipps' work, and on other early ranking methods, is presented by Wirth [147, particularly Chapter 4, Section 28]. A discussion of the history of Kendall's τ is given by Kruskal [96].

3.5. *Tönnies.* The German sociologist, F. Tönnies, suggested in 1909 [137] a measure of association for square cross classifications in which both polytomies are ordered. A later paper is [138]. Tönnies presents his measure, which is related to the so-called Spearman foot-rule, in terms of continuous, rather than grouped, variates, but he immediately collects them into groups on the basis of their relative magnitudes.

The measure, in our terminology, is found by first adding all ρ_{aa}'s, i.e. all ρ_{ab}'s in the main diagonal, and multiplying this sum by 2. To this is added the sum of all ρ_{ab}'s in the two diagonals neighboring the main diagonal. Then an analogous weighted sum is computed for the counter-diagonal and its two neighbors, and this is subtracted from the first sum. In terms of a formula, Tönnies looks at

$$\left[2\sum_{a-b=0}\sum \rho_{ab}+\sum_{a-b=\pm1}\sum \rho_{ab}\right]-\left[2\sum_{a+b-1=\alpha}\sum \rho_{ab}+\sum_{a+b-1=\alpha\pm1}\sum \rho_{ab}\right].$$

He compares this quantity with $2-(2/\alpha)$, its maximum possible absolute value. Thus Tönnies' measure is of the kind discussed briefly by us in Section 8.3 of [66].

H. Striefler [130] provides an exposition of Tönnies' measure and suggests an extension.

3.6. *Deuchler.* In 1914, the German educational psychologist, Gustav Deuchler [32], continued the earlier work of Lipps (Section 3.4) on the quantity now called Kendall's τ. Deuchler worked on the distribution of τ, both under the null hypothesis of independence and under alternative hypotheses, on methods of computing τ, and on modifications when ties are present.[2]

[2] For further discussion of Deuchler's work on τ itself we refer to [96]. For information about other aspects of Deuchler's work, and for remarks about an unpublished monograph by Deuchler, we refer to [95]. A microfilm of this unpublished manuscript is in our hands, and we will try to make it available on request.

A few years later, Deuchler [33] returned to the question of multiple ties in both coordinates, so that he was really concerned with cross classifications having meaningful order for both polytomies. For this situation Deuchler suggested as a measure of association (in our notation)

$$\mathfrak{R} = \frac{\Pi_{s\leq} - \Pi_{d\leq}}{1 - \Pi_{t(\text{both})}} = \frac{\Pi_s - \Pi_d}{1 - \Pi_{t(\text{both})}} .$$

Here $\Pi_{s\leq}(\Pi_{d\leq})$ is the probability that two randomly chosen individuals from the cross classified population will have their A and B categories similarly (dissimilarly) ordered, with a tie *in one polytomy alone* always counting as similarity (dissimilarity), but a tie in *both* categories—i.e. both individuals in the same cell—not counting in either case. $\Pi_{t(\text{both})}$ is the probability that two randomly chosen individuals fall into the same cell, i.e. are tied in both polytomies. Thus $\mathfrak{R}$ is much like our γ [66, Sec. 6] except that Deuchler has $\Pi_{t(\text{both})}$ where we have Π_t.

Actually Deuchler's presentation is in terms of choosing two individuals at random *without* replacement from a finite cross classified population, whereas we in [66] give an interpretation in terms of random choice with replacement. For $\mathfrak{R}$, one obtains the same value of the measure in either interpretation, while γ changes slightly as one goes from the with-replacement to the without-replacement interpretation.

Deuchler develops his $\mathfrak{R}$ by the same scoring scheme as that later used by Kendall. $\mathfrak{R}$ does not have quite as direct an interpretation as γ, but it possesses one characteristic that γ does not have: $\mathfrak{R}$ is 1 (its maximum value) if and only if at most one ρ_{ab} in each row and column is positive *and* the positive ρ_{ab}'s are all concordant. This last means that, denoting the positive ρ_{ab}'s by $\rho_{a_1b_1}$, $\rho_{a_2b_2}, \cdots$, with $a_1 < a_2 < \cdots$, then $b_1 < b_2 < \cdots$. The examples on p. 750 of [66] show that this property is not true for γ. Note that $|\mathfrak{R}| \leq |\gamma|$.

Deuchler observes that $\mathfrak{R}$ varies as contiguous categories are pooled and he discusses the magnitude of this effect at length. He also compares his $\mathfrak{R}$ with Spearman's rank correlation coefficient, and with the mean square contingency coefficient in the 2×2 case. The applications that Deuchler has in mind, and for which he uses his measure, are to the association between the grades of school children in two subjects or traits. He discusses briefly the situation in which one wishes to analyze such joint gradings on more than two such characteristics. In [34], Deuchler discusses in more detail the 2×2 case.

3.7. *Gini.* In 1914–1916, Corrado Gini [56, 57, 58, 59, 60] examined in detail many distinctions between relationships within a bivariate distribution, and proposed a great variety of measures of association and disassociation.[3] Examples were given to indicate the circumstances under which the various proposed measures might be appropriate.

Many of Gini's measures of association relate to cases in which the bivariate distribution is quantitative or can easily be made so by the use of relevant ordinal scores. For the qualitative case without ordering among the categories

[3] We wish to thank Sebastian Cassarino, Department of Italian, University of California, Berkeley, for his assistance in examining Gini's papers.

(*sconesse* categories), and where both polytomies are the same ($A_a = B_a$), Gini [57] proposed as a measure of association the quantity (in our notation)

$$\frac{\sum \rho_{aa} - \sum \rho_{a\cdot}\rho_{\cdot a}}{\sqrt{(1 - \sum \rho_{a\cdot}^2)(1 - \sum \rho_{\cdot a}^2)}}$$

This is based on a sort of indirect scoring scheme, suggested by divergences of cell frequencies from the corresponding marginal products. In the 2×2 case, the above quantity is the appropriately signed square root of the mean square contingency.

In [58], Gini proposed the following variant of the above measure:

$$\frac{\sum \rho_{aa} - \sum \rho_{a\cdot}\rho_{\cdot a}}{1 - \frac{1}{2}\sum |\rho_{a\cdot} - \rho_{\cdot a}| - \sum \rho_{a\cdot}\rho_{\cdot a}},$$

and a number of other variations were discussed systematically in [58] and [60].

We have not found in Gini's papers operational interpretations of his proposed measures. They all seem to be of a formal nature in which consideration of absolute or quadratic differences, followed by averaging, is taken as reasonable without argument. Special attention is paid to denominators so as to make the indices range between 0 and 1 (or -1 and 1) within appropriate limitations for variation in the joint distribution.

In [57, p. 598], Gini briefly discussed polytomies in which the categories are *cyclically* ordered, as for example the months of the year. This type of polytomy was not discussed by us in [66]. Gini suggested the possibility of a measure of association in this case, but he gave little detail. Ten years later Pietra [116] considered the cyclical case in great detail, and since then other Italian authors have written on this topic.

The measures proposed by Gini have formed the basis of a large literature, mostly in Italian. We now cite several publications outlining and discussing Gini's work in this area. First, Gini himself [60, pp. 1458 ff.] gave a systematic outline of his measures. An exposition in English of some of the Gini material was given by Weida [144], and a more detailed exposition by Pietra in the introduction of [116]. A general article on the work of the Italian school is that of Gini [61]; another, of a critical nature, is by Thionet [132]. (The reader of this last article should also refer to subsequent correspondence by Galvani [54] and Thionet [133].) Two recent expositions by Gini are [62] and [63, Chap. 9].

Some further references to the recent Italian literature appear in Section 4.7. In Section 4.4 a measure proposed by Gini in the $\alpha\times 2$ case is discussed in detail.

4. MORE RECENT PUBLICATIONS

4.1. *Textbook discussions. Guilford, Dornbusch and Schmid, Wallis and Roberts.* In [73], J. P. Guilford discusses association in an $\alpha\times\beta$ table from the viewpoint of optimal prediction in a manner essentially equivalent to that of Guttman (see comment and reference in [66, p. 742]), and to that in which we introduced λ_a and λ_b in [66]. This discussion appears in Chapter 10 of the 1942 edition and is amplified in Chapter 14 of the 1950 edition.

In a recent textbook [37, p. 215], S. M. Dornbusch and C. F. Schmid discuss a "coefficient of relative predictability" for $\alpha\times 2$ tables, their G. It is equal to λ_a for $\alpha\times 2$ tables.

W. A. Wallis and H. V. Roberts present the λ measures and γ in their book [142, Chap. 9]. Their notation corresponds to ours as follows:

Wallis-Roberts	$g_{c\cdot r}$	$g_{r\cdot c}$	g	h	S	D	T
Goodman-Kruskal	λ_b	λ_a	λ	γ	$\frac{n^2}{2}\Pi_s$	$\frac{n^2}{2}\Pi_d$	$\frac{n^2}{2}\Pi_t$

and their discussion is in terms of sample frequencies.

4.2. *Reliability measures.* We describe now some papers on measures of association in the reliability context, that is when both polytomies of a cross classification are the same and refer to two methods of assignment. Other papers that deal with reliability measures will be discussed elsewhere, particularly in Sections 4.9 through 4.12, under other classifications.

Wood. In 1928, K. D. Wood [148] suggested several variations of the kind of measure of association described in Section 8.3 of [66] where reliability for ordered polytomies was discussed. Wood's suggestions related to a 4×4 table with $\rho_{a\cdot}=\rho_{\cdot b}=.25$ for all a and b; they were

$$\sum \rho_{aa}, \quad \sum_{|a-b|\le 1}\sum \rho_{ab}, \quad \sum \rho_{aa}-\sum_{a+b=5}\sum \rho_{ab}, \quad \text{and} \quad \sum_{|a-b|\le 1}\sum \rho_{ab}-\sum_{|a-b|\ge 2}\sum \rho_{ab}.$$

Actually, Wood's discussion is in terms of sample analogs, and it is wholly motivated by the desire to find sample functions that approximate well to the sample correlation coefficient. To investigate this he divides a sample into 16 parts via its marginal quartiles, computes the above measures, and compares them with the sample correlation coefficient.

Reuning. H. Reuning [120] has recently suggested a new measure of reliability in the case of ordered polytomies. Reuning compares the actual ρ_{ab} table with the table that would result if (a) there were independence between rows and columns, and (b) the marginal distributions were rectangular—he calls this the case of pure chance; its meaning is that each $\rho_{ab}=1/\alpha^2$. Further, in order to use the natural ordering, Reuning suggests pooling all cells such that $|a-b|=$constant. There are α cells such that $|a-b|=0$, $2(\alpha-1)$ cells such that $|a-b|=1$, $2(\alpha-2)$ cells such that $|a-b|=2, \cdots$, and 2 cells such that $|a-b|=\alpha-1$, the maximum difference. Thus Reuning is led to compare

$$\begin{array}{rcl} \sum \rho_{aa} & \text{with} & \alpha(1/\alpha^2) = 1/\alpha \\ \sum_{|a-b|=1} \rho_{ab} & \text{with} & 2(\alpha-1)/\alpha^2 \\ \sum_{|a-b|=2} \rho_{ab} & \text{with} & 2(\alpha-2)/\alpha^2 \\ & \vdots & \\ \rho_{1\alpha}+\rho_{\alpha 1} & \text{with} & 2/\alpha^2 \end{array}$$

In order to obtain a measure of reliability, Reuning in effect considers the following χ^2-like quantity

$$\sum_{k=0}^{\alpha-1} \frac{\left\{ \sum_{|a-b|=k} \rho_{ab} - \frac{\text{No. of summands in } \sum_{|a-b|=k}}{\alpha^2} \right\}^2}{\frac{\text{No. of summands in } \sum_{|a-b|=k}}{\alpha^2}} .$$

Reuning also considers $\Sigma\rho_{aa}$, a measure mentioned in [66].

The above presentation differs slightly from that given by Reuning, first, because he works with sample instead of population quantities, and, second, because he emphasizes testing rather than estimation. If we regard the population characteristic in the above display as a general measure of reliability (and it is not wholly clear from Reuning's paper whether he so regards it) some problems of interpretation arise, stemming from the comparison with the "pure chance" cross classification. For one thing, if $\Sigma\rho_{aa}=1$, so that reliability in the ordinary sense is perfect, Reuning's measure takes the value $\alpha-1$, which is by no means its maximum possible value. On the other hand, if $\rho_{1\alpha}+\rho_{\alpha 1}=1$, so that reliability in the ordinary sense is about as poor as can be, Reuning's measure takes the value $(\alpha^2-2)/2$, which is actually greater than its value for $\Sigma\rho_{aa}=1$ (unless $\alpha=2$, when the two values are equal).

The "pure chance" or uniform table as a basis of comparison had been put forward by Andreoli [1] in 1934. H. F. Smith [126a] uses the same device of pooling along diagonals as does Reuning, but in the context of a comparative test of two square cross classifications.

Cartwright. D. S. Cartwright, for the case of unordered polytomies, has recently [19] suggested a measure of interreliability when there are two *or more* classifications, each with the same polytomy. He thinks of the common polytomy as possible judgments about members of the population on the part of J judges, so that

$$\rho_{a_1 a_2 \cdots a_J}$$

is that fraction of the population allocated by judge 1 to class a_1, by judge 2 to class a_2, etc., where $a_j=1, 2, \cdots, \alpha$. His measure of reliability, in our notation, may be written as

$$\frac{2}{J(J-1)} \sum_j \sum_{k>j} \sum_{a_j=a_k} \rho_{\cdots a_j \cdots a_k \cdots} ,$$

or the probability that two different randomly chosen judges out of the J judges will allocate a random member of the population to the same class. For $J=2$, this becomes just $\Sigma\rho_{aa}$.

Cartwright's presentation of his measure differs superficially from the above. He also considers distribution theory for the sample analogue of the above measure under special restrictive conditions.

4.3. *Measures that are zero if and only if there is independence.* The traditional χ^2-like measures of association, unlike the λ and γ measures discussed by us in

[66], have the property that they take a particular value, zero, if and only if there is independence in the cross classification, i.e., $\rho_{ab}=\rho_{a\cdot}\rho_{\cdot b}$. This property has seemed important to a number of workers, and they have proposed measures of association with the property but different from the traditional measures. In some cases, other formal properties have also been emphasized. We now discuss several such proposals that do not fit more naturally into other sections of this survey.

So far as we know, none of the measures discussed here have operational interpretations of the kind we have argued for in [66], and indeed this is not surprising. For a measure with an operational interpretation measures, so to speak, one aspect or dimension of association. Hence, if a given cross classification exhibits no association along this aspect or dimension one would expect a zero value for the measure, even if there is association in other senses. That is why we are not troubled by the fact that the λ and γ measures can be zero even though there is dependence. Note that if there is independence the λ and γ measures are zero. This is to be expected, since independence should correspond to lack of association in *any* sense.

Cramér. In 1924, H. Cramér [29] suggested for an $\alpha\times\beta$ table the measure

$$\min \sum_a \sum_b (\rho_{ab} - u_a v_b)^2$$

where the minimum is computed over all numbers $u_1, \cdots, u_\alpha; v_1, \cdots, v_\beta$. This quantity is zero if and only if there is independence, and is always $\leq .25$. It suffers from having no definite value in the case of complete dependence.

Cramér says [29, p. 226] that " . . . there is no absolutely general measure of the degree of dependence. Every attempt to measure a conception like this by a single number must necessarily contain a certain amount of arbitrariness and suffer from certain inconveniences."

Steffensen. In 1933, J. F. Steffensen [127] proposed the following measure of association for cross classifications:

$$\psi^2 = \sum_a \sum_b \rho_{ab} \frac{(\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2}{\rho_{a\cdot}(1 - \rho_{a\cdot})\rho_{\cdot b}(1 - \rho_{\cdot b})}$$

in our notation. (See Lorey [105] for a discussion.) Apparently Steffensen's motivation was to avoid certain formal inadequacies of previously suggested measures. For example, Steffensen points out that his ψ^2 attains its upper limit of 1 if and only if the two classifications are functionally related, i.e. if and only if exactly one ρ_{ab} in each row and in each column is positive. Steffensen gives no operational interpretation for ψ^2. Note that ψ^2 is an average of all 2×2 mean square contingencies formed from each of the $\alpha\beta$ cells of the cross-classification and its complement; in this it resembles the measure proposed by Jordan [85] that we discussed in Section 3.2.

The next year, Steffensen [128] returned to ψ^2 in greater detail. (In [127] the measure had appeared only in a nonnumbered page of errata, as a better version of a similar measure, given in the article proper, that Steffensen later decided was unsatisfactory.) Then Steffensen suggested a variant,

$$\omega = \frac{2 \overline{\sum} \overline{\sum} (\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})}{\overline{\sum} \overline{\sum} (\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b}) + 1 - \overline{\sum} \overline{\sum} \rho_{ab}^2},$$

where $\overline{\sum} \overline{\sum}$ means summation over those cells for which $\rho_{ab} > \rho_{a\cdot}\rho_{\cdot b}$. He showed that ω, along with ψ^2, (1) lies between 0 and 1, (2) is 0 if and only if independence obtains, and (3) is 1 if and only if exactly one ρ_{ab} in each row and column is positive. Finally, an extension to the case of continuous bivariate distributions was suggested.

Immediately following [128] an editorial [114] (presumably by Karl Pearson) criticized Steffensen's suggestions with arguments based on the assumption of an underlying continuous distribution. First, the editorial said that the continuous analogue of ψ^2 would be identically zero because of the presence of squared differentials. Then it argued that a measure of association for cross classifications should *not* be able to attain the value unity, because, while complete dependence might exist between the two polytomies, it could well be the case that a finer cross classification would show that *within* the original cells complete association did not exist. These arguments were used to contrast Steffensen's suggestions with the coefficient of mean square contingency, to the latter's favor. The editorial concluded with a numerical comparison of ψ^2 and the coefficient of mean square contingency for a number of artificial cross classifications, and it stated that ψ^2 tends to be too low, with values crowded in the interval [0, .25], even for quite sizable intuitive association.

In 1941, Steffensen [129] returned to his discussion of ω. He presented a natural generalization to the density function case and showed that the three properties mentioned above still essentially held. A lengthy discussion of the generalized ω in the bivariate normal case was given, and the paper concluded with a rebuttal to the arguments of [114].

This discussion reinforces our beliefs that it is essential to give operational interpretations of measures of association and that the mere fact that a measure can range from 0 to 1 (say) is of little or no use in understanding it.

Pollaczek-Geiringer. In 1932 and 1933, Hilda Pollaczek-Geiringer [117, 118], motivated by considerations similar to those adduced by Steffensen, suggested a measure of association for any bivariate distribution, continuous or discrete. The measure may also be applied, as Pollaczek-Geiringer suggested, to a cross-classification in which both polytomies are ordered. In our notation, the suggested measure for this case is

$$\frac{\sum_a \sum_b (A_{ab}D_{ab} - B_{ab}C_{ab})}{\sum_a \sum_b (A_{ab}D_{ab} + B_{ab}C_{ab})},$$

where

$$A_{ab} = \sum_{a' \le a} \sum_{b' \le b} \rho_{a'b'} \qquad B_{ab} = \sum_{a' > a} \sum_{b' \le b} \rho_{a'b'}$$

$$C_{ab} = \sum_{a' \le a} \sum_{b' > b} \rho_{a'b'} \qquad D_{ab} = \sum_{a' > a} \sum_{b' > b} \rho_{a'b'}$$

Pollaczek-Geiringer gives no operational interpretation. Her measure has a certain similarity to our γ [66, Section 6] especially if it is modified by replacement of the summations with weighted sums, having ρ_{ab}'s as weights.

Höffding. In 1941 and 1942, W. Höffding (now Hoeffding) presented two very interesting papers bearing on measures of association for cross classifications. Höffding's first paper on cross classifications [79] was based on a prior paper of his [78] that had dealt solely with the bivariate density function case. In [78], it was urged that measures of association should be invariant under transformations, monotone in the same direction, of the associated random variables. Several measures having this invariance were presented and their properties discussed. The cross classifications of [79] were considered as arising from underlying density function distributions by rounding. Hence their cumulative distribution functions are only known at points of a rectangular lattice, and their density functions are only known via averages over cells. In order to apply the suggestions of [78], Höffding replaced a cross classification by a density function distribution with constant density within each cell, proportional to its ρ_{ab}. (This might appear to make matters depend on metrics for the two classifications, but any such dependence is a notational artifact, disappearing later because of invariance.) Then Höffding applied to this "step-function" density the measures of [78]. The first was the correlation coefficient between the probability integral transforms of the marginal random variables (this is the so-called grade correlation, or population analogue of Spearman's rank correlation coefficient). Höffding obtained

$$\bar{\rho} = 3 \sum_a \sum_b \rho_{ab} \left[2\left(\sum_{a'<a} \rho_{a'\cdot} \right) + \rho_{a\cdot} - 1 \right] \left[2\left(\sum_{b'<b} \rho_{\cdot b'} \right) + \rho_{\cdot b} - 1 \right].$$

A slight modification gave him the more satisfactory

$$\rho^* = \bar{\rho} / \sqrt{(1 - \sum \rho_{a\cdot}{}^3)(1 - \sum \rho_{\cdot b}{}^3)}.$$

Höffding then discussed the extrema that $\bar{\rho}$ and ρ^* can reach, and their values for 2×2 tables. In the 2×2 case, ρ^{*2} is just the mean square contingency.

Höffding then pointed out that his ρ^* is the same as Student's modification of Spearman's rank correlation coefficient [131], provided that appropriate notational translations are made. The article continued with a discussion of mean square contingency and related coefficients, including one that is a function of the quantities

$$\left(\sum_{a' \leq a} \sum_{b' \leq b} \rho_{a'b'} \right) - \left(\sum_{a' \leq a} \rho_{a'\cdot} \right) \left(\sum_{b' \leq b} \rho_{\cdot b'} \right),$$

thus giving a measure of departure from independence as defined in terms of cumulative distributions.

In the later portion of [80], Höffding returned to these questions. He distinguished between those cases in which a continuous distribution is considered as underlying the discrete distribution of interest, and those cases in which the discrete distribution itself is of primary interest. For this second situation he suggested a measure of association by analogy with one for density-function distributions suggested earlier in the article. It is simply

$$\tfrac{1}{2}\sum\sum\left|\rho_{ab}-\rho_{a\cdot}\rho_{\cdot b}\right|.$$

A modification was then put forward, namely, division by $1-\overline{\sum}\,\overline{\sum}\rho_{ab}{}^2$, where $\overline{\sum}\,\overline{\sum}$ means summation over those (a, b) such that $\rho_{ab}>\rho_{a\cdot}\rho_{\cdot b}$. The result is simply related to Steffensen's ω.

Eyraud. H. Eyraud [43] suggested for the 2×2 table the measure of association $(\rho_{11}-\rho_{1\cdot}\rho_{\cdot 1})/(\rho_{1\cdot}\rho_{\cdot 1}\rho_{2\cdot}\rho_{\cdot 2})$. He discussed its extreme values, its interpretation, and, briefly, its extension to $\alpha\times\beta$ tables. In addition he considered the $2\times 2\times 2$ case.

Fréchet, Féron. M. Fréchet has discussed measures of association in a series of articles (e.g. [50] and [51]) that deal mostly with cases in which a meaningful metric exists for both polytomies. In some more recent articles, [52] and [53], he has studied the extent to which knowledge of the marginals restricts the probabilities of a cross classification. Fréchet's work discusses the extent to which measures of association satisfy a set of formal criteria such as those mentioned earlier in this section.

In two recent publications, [45] and [46], R. Féron has discussed measures of association, again with emphasis on the case when metrics are present, but with some consideration of the purely qualitative case. Several of the measures described in this section are discussed by Féron.

4.4. *Measures of dissimilarity, especially in the $\alpha\times 2$ case.* In considering an $\alpha\times 2$ cross classification, it is natural to approach the question of association by asking about the degree of dissimilarity between the two conditional multinomial populations in the two columns, when compared row by row. This approach has often been taken in the social sciences when columns refer to a dichotomy of interest (Negro-White, Male-Female, etc.) and rows correspond to places, times, or the like. It is, of course, equivalent to speak of a $2\times\beta$ cross classification by simply interchanging rows and columns.

Gini, Florence, Hoover, Duncan and Duncan, Bogue. A measure of dissimilarity in the $\alpha\times 2$ case that has been proposed a number of times, often in variant forms, is the following:

$$D=\frac{1}{2}\sum_{a=1}^{\alpha}\left|\frac{\rho_{a1}}{\rho_{\cdot 1}}-\frac{\rho_{a2}}{\rho_{\cdot 2}}\right|,$$

or half the sum of absolute differences between corresponding conditional probabilities in the first and second columns. The use of D appears to have been first suggested by C. Gini (see [56], [57], [61a]); some more recent publications about this measure are by P. S. Florence [48], E. M. Hoover [82] and [83], O. D. Duncan and B. Duncan [42], and D. J. Bogue [12].

Since the summation in D, if the absolute value signs were omitted, would be $1-1=0$, we see that

$$\sum_a{}^{+}\left\{\frac{\rho_{a1}}{\rho_{\cdot 1}}-\frac{\rho_{a2}}{\rho_{\cdot 2}}\right\}+\sum_a{}^{-}\left\{\frac{\rho_{a1}}{\rho_{\cdot 1}}-\frac{\rho_{a2}}{\rho_{\cdot 2}}\right\}=0$$

where $\sum_a{}^{+}$ indicates summation over nonnegative values of the summand, and $\sum_a{}^{-}$ indicates summation over negative values of the summand. Thus

$$D = \sum_a{}^{+} \left\{ \frac{\rho_{a1}}{\rho_{\cdot 1}} - \frac{\rho_{a2}}{\rho_{\cdot 2}} \right\} = - \sum_a{}^{-} \left\{ \frac{\rho_{a1}}{\rho_{\cdot 1}} - \frac{\rho_{a2}}{\rho_{\cdot 2}} \right\},$$

and we see that D is the difference between the proportion of the population in column 1 appearing in rows for which $\rho_{a1}/\rho_{\cdot 1} > \rho_{a2}/\rho_{\cdot 2}$ and the proportion of the column 2 population appearing in these rows. A similar verbal statement, with the difference taken in the opposite sense, for rows with $\rho_{a1}/\rho_{\cdot 1} < \rho_{a2}/\rho_{\cdot 2}$, corresponds to the second equality of the above display.

Now suppose we think of redistributing the (conditional) column 1 population among its cells so that it becomes equal to the (conditional) column 2 population. This means moving probability mass from the column 1 cells with $\rho_{a1}/\rho_{\cdot 1} > \rho_{a2}/\rho_{\cdot 2}$ to those with the opposite inequality holding, and clearly the minimum proportion of the column 1 population that we must shift to achieve this goal is D. A similar interpretation may be given in terms of redistributing the column 2 population so that it becomes (conditionally) equal to the column 1 population. After such a redistribution, the two cells in each row would have equal conditional probabilities, each conditional on its fixed column marginals. Also, the proportion of the population in a given row that is in column 1 will be the same for each row. Thus D has a useful operational interpretation for some purposes; for example see [42].

The construction of D suggests an ordering of the rows that may be of substantive interest in some contexts. Rearrange the rows so that the row with maximum $(\rho_{a1}/\rho_{\cdot 1}) - (\rho_{a2}/\rho_{\cdot 2})$ becomes the first row, the row with next largest $(\rho_{a1}/\rho_{\cdot 1}) - (\rho_{a2}/\rho_{\cdot 2})$ becomes the second row, and so on. If there are α_* rows with $\rho_{a1}/\rho_{\cdot 1} \geq \rho_{a2}/\rho_{\cdot 2}$, D may then be expressed as

$$\sum_{a=1}^{\alpha_*} \left\{ \frac{\rho_{a1}}{\rho_{\cdot 1}} - \frac{\rho_{a2}}{\rho_{\cdot 2}} \right\} = - \sum_{a=\alpha_*+1}^{\alpha} \left\{ \frac{\rho_{a1}}{\rho_{\cdot 1}} - \frac{\rho_{a2}}{\rho_{\cdot 2}} \right\}$$

in terms of the reordered cross classification.

Some other easily obtained expressions for D are

$$\begin{aligned} D &= \sum_{a=1}^{\alpha} \left| \frac{\rho_{a1}}{\rho_{\cdot 1}} - \rho_{a\cdot} \right| \Big/ [2\rho_{\cdot 2}] = \sum_{a=1}^{\alpha} \left| \frac{\rho_{a2}}{\rho_{\cdot 2}} - \rho_{a\cdot} \right| \Big/ [2\rho_{\cdot 1}] \\ &= \sum_{a=1}^{\alpha} \sum_{b=1}^{2} \left| \frac{\rho_{ab}}{\rho_{\cdot b}} - \rho_{a\cdot} \right| \Big/ [4(1 - \rho_{\cdot b})] \\ &= \sum_{a=1}^{\alpha} \sum_{b=1}^{2} \left| \rho_{ab} - \rho_{a\cdot}\rho_{\cdot b} \right| / [4\rho_{\cdot 1}\rho_{\cdot 2}]. \end{aligned}$$

The first three of these describe D in terms of absolute differences of form $(\rho_{ab}/\rho_{\cdot b}) - \rho_{a\cdot}$, while the last describes D in terms of the most conventional measure of deviations from cell independence, $\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b}$. This last expression for D resembles the traditional χ^2 kind of measure, but differs from such measures in that the absolute differences are used rather than the squared differences, and the weightings of the terms are different.

Still another mode of description for D may be given in terms of absolute differences between the column conditional probabilities, $\rho_{ab}/\rho_{a\cdot}$, and the column marginals, $\rho_{\cdot b}$. It is easily checked that

$$D = \sum_{a=1}^{\alpha} \left| \frac{\rho_{a1}}{\rho_{a\cdot}} - \rho_{\cdot 1} \right| \rho_{a\cdot}/[2\rho_{\cdot 1}\rho_{\cdot 2}]$$

$$= \sum_{a=1}^{\alpha} \left| \frac{\rho_{a2}}{\rho_{a\cdot}} - \rho_{\cdot 2} \right| \rho_{a\cdot}/[2\rho_{\cdot 1}\rho_{\cdot 2}]$$

$$= \sum_{a=1}^{\alpha} \sum_{b=1}^{2} \left| \frac{\rho_{ab}}{\rho_{a\cdot}} - \rho_{\cdot b} \right| \rho_{a\cdot}/[4\rho_{\cdot 1}\rho_{\cdot 2}].$$

The traditional χ^2-like measures may, of course, also be expressed in analogous equivalent ways in the special case of two columns. For example, $\phi^2 = \chi^2/\nu$ may be expressed as

$$\sum_{a=1}^{\alpha} \sum_{b=1}^{2} (\rho_{ab} - \rho_{a\cdot}\rho_{\cdot b})^2/\rho_{a\cdot}\rho_{\cdot b} = \sum \sum \left(\frac{\rho_{ab}}{\rho_{\cdot b}} - \rho_{a\cdot}\right)^2 \rho_{\cdot b}/\rho_{a\cdot}$$

$$= \sum_{a=1}^{\alpha} \left(\frac{\rho_{a1}}{\rho_{\cdot 1}} - \frac{\rho_{a2}}{\rho_{\cdot 2}}\right)^2 \frac{\rho_{\cdot 1}\rho_{\cdot 2}}{\rho_{a\cdot}}$$

$$= \sum \sum \left(\frac{\rho_{ab}}{\rho_{a\cdot}} - \rho_{\cdot b}\right)^2 \rho_{a\cdot}/\rho_{\cdot b}$$

$$= \sum_{a=1}^{\alpha} \left(\frac{\rho_{a1}}{\rho_{a\cdot}} - \rho_{\cdot 1}\right)^2 \rho_{a\cdot}/\rho_{\cdot 1}\rho_{\cdot 2}$$

$$= 1 - \frac{1}{\rho_{\cdot 1}\rho_{\cdot 2}} \sum_{a=1}^{\alpha} \frac{\rho_{a1}\rho_{a2}}{\rho_{a\cdot}}.$$

The possibilities of expressing a measure in terms of the deviation of ρ_{ab} from $\rho_{a\cdot}\rho_{\cdot b}$, in terms of the deviation of $\rho_{a1}/\rho_{\cdot 1}$ from $\rho_{a2}/\rho_{\cdot 2}$, or in terms of the deviation of $\rho_{ab}/\rho_{a\cdot}$ from $\rho_{\cdot b}$, etc., may give added insight into the nature of the measure by suggesting interpretations and approaches to it from different directions. On the other hand, the same possibility of variant expression may cause confusion in communication and may mislead authors to think that symbolically different expressions correspond to different measures, when in fact the measures are the same. Duncan and Duncan [42] and J. Williams [145] discuss a number of articles where this difficulty seems to exist. The last form given above for ϕ^2 has been discussed by E. Katz and P. Lazarsfeld [87a, p. 373].

Measures of association for the $\alpha \times \beta$ case may be based on the idea of dissimilarity between two columns by averaging in some way the $\beta(\beta-1)/2$ possible values of an $\alpha \times 2$ measure of dissimilarity obtained from pairs of columns in the larger cross classification. Alternatively, one might average the β values of an $\alpha \times 2$ measure obtained by comparing each column of the $\alpha \times \beta$ table with the column of row marginals, $\rho_{a\cdot}$. This approach has been used by Gini and by Fréchet, in references cited earlier.

Boas. In 1922, Franz Boas [11, pp. 432–4] suggested a measure of dissimilarity between one specific column of a cross classification and the column of row marginals, that is between one multinomial population and the (weighted) average of a group of multinomial populations to which the one in question belongs. Boas's suggestion, in our notation, seems to be the following:

Suppose that an individual is chosen at random from the bth column of a cross classified population in accordance with the conditional distribution for that column. That is, the individual falls in the (a, b) cell with probability $\rho_{ab}/\rho_{\cdot b}$. Now suppose that we are told the row in which he falls but *not* told that he came from the bth column. If we guess his column, based on knowledge of his row, in a random manner reproducing the population (as discussed in Section 9 of [66]), we shall guess column b with conditional probability $\rho_{ab}/\rho_{a\cdot}$, where a is the row in which he has fallen. Thus the probability of correctly guessing the cell in which the individual falls, when (i) he is in fact drawn from the bth column, and (ii) we guess his column, knowing only his row, in a random manner reproducing the population, is

$$\sum_a \left(\frac{\rho_{ab}}{\rho_{\cdot b}}\right)\left(\frac{\rho_{ab}}{\rho_{a\cdot}}\right) = \sum_a \rho_{ab}^2/(\rho_{a\cdot}\rho_{\cdot b}).$$

That is, as we undertand it, Boas's measure of dissimilarity between column b and the column of row marginals.

Boas also considers the possibility of changing the table so that it has equal column marginals (see Section 5.4 of [66]).

Long and Loevinger. In working with psychological tests made up of yes-no questions, one may wish to consider association between a particular question and the whole test. This situation may be viewed in the framework of an $\alpha\times 2$ table in which the columns refer to the two possible responses and the rows make up an ordered classification based on the whole test. The ρ_{ab}'s are the proportions of individuals in the population falling into one of the whole-test score classes and responding to the individual question in one of the two possible ways. For this special psychometric situation, measures of association have been proposed and discussed by Long [105] and by Loevinger [102, Chap. 5] and [103].

4.5. *Measures based on Lorenz or cost-utility curves*. For the $\alpha\times 2$ cross classification, where the α rows have a meaningful order (determined from the cross classification itself, as discussed in Section 4.4, or determined from external considerations) the following approach has been suggested. Consider the partial sums

$$X_a = \sum_{i=1}^{a} \frac{\rho_{i1}}{\rho_{\cdot 1}} \quad \text{and} \quad Y_a = \sum_{i=1}^{a} \frac{\rho_{i2}}{\rho_{\cdot 2}},$$

and consider the points (X_a, Y_a) for $a=1, \cdots, \alpha$ in the unit square. The underlying thought is that these are points on a smooth curve expressing a functional relationship between X and Y, but that we only know this curve at the α points (X_a, Y_a). If there is independence in the cross classification, then $Y_a = X_a$ for each a; i.e., the points (X_a, Y_a) lie on the straight line segment going diagonally from (0, 0) to (1, 1). But if there is association, the general shape of the underlying curve suggested by the (X_a, Y_a)'s, and its "distance" from the diagonal line, will describe it. Several measures of association, based on this idea, have been suggested in the literature (see, e.g., [42], [65], [6], [41]), but we shall not discuss them here. In some cases, a structural assumption or smoothing procedure (e.g., the use of straight line segments) is used to obtain a curve from the α points.

4.6. *Measures based on Shannon-Wiener information. McGill, Holloway, Woodbury, Wahl, Linfoot, Halphen.* Some time ago it was suggested to us by J. W. Tukey that measures of association based on the Shannon-Wiener information function might be useful. Since we were unable to satisfy ourselves that such measures would have reasonable interpretations for many contexts in which cross classifications appear, we did not discuss the possibility in [66]. We wish, however, to mention here a few papers in which the information concept is used as the basis of measures of association, although we continue to reserve our opinions about the utility of these proposals outside the area of communication theory.

Perhaps the first such paper is by W. J. McGill [107]. Soon after it, the approach was suggested in a meteorological setting by J. L. Holloway, Jr. and M. A. Woodbury [81]. E. W. Wahl [141] summarizes some of the material of [81]. The measure has been used in meteorology, notably by I. I. Gringorten and his colleagues, [72] and [69]. Two quite recent papers on this general theme are by E. H. Linfoot [99] and E. Halphen [73a].

4.7. *Recent proposals by Italian authors other than Gini.* We have already discussed the early suggestions of Benini (Section 3.3) and the extensive publications by Gini (Section 3.7). Since then, the Italian statistical literature has been replete with articles about one aspect or another of the measurement of association. Nearly all of this literature has been derivative from Gini's 1914–16 publications; the interested reader can find some key references in Section 3.7. We shall not attempt to give a complete outline of this literature, but some of the more interesting articles that have come to our attention will now be listed.

Salvemini. A prolific writer on the theme of measures of association has been T. Salvemini. In [122], he surveyed parts of the field, and suggested some new expressions for Gini's measures in the asymmetrical and unordered qualitative case. In [123], Salvemini discussed the calculation and application of measures of association; the case in which one polytomy is ordered, while the other is not, received consideration. More recently, he has presented [125] an extensive discussion of the whole field of measures of association. References to many other papers by Salvemini may be found in the three articles cited above.

Bonferroni and Brambilla. C. E. Bonferroni has given [12a] a detailed discussion of a number of measures of association, emphasizing relations between the ρ_{ab}'s, $\rho_{a\cdot}$'s and $\rho_{\cdot b}$'s, and pointing out problems and concepts that arise in the three-way cross classification. Another article by Bonferroni in this area is [13]. Closely associated is the work of F. Brambilla [14] who presented a systemic discussion of the field giving particular emphasis to the effects of holding marginals fixed or not and to three-way cross classifications.

Faleschini. Particularly interesting for us is an article by L. Faleschini [44]. His approach is to consider the most probable cell in the bth column, and to compare its conditional probability with some kind of average of the column conditional probabilities in the same row. Thus, if $a^*(b)$ is defined by

$$\rho_{a^*(b)b} \geq \rho_{ab} \qquad \text{(all } a\text{)},$$

Faleschini considers the differences

$$D_b = \frac{\rho_{a^*(b)b}}{\rho_{\cdot b}} - \text{some average of } \frac{\rho_{a^*(b)b'}}{\rho_{\cdot b'}} \qquad (b' = 1, \cdots, \beta).$$

Finally the D_b's are averaged in some way. Thus two averages can be rather arbitrarily introduced. If in the first (that of the conditional probabilities) we weight by $\rho_{\cdot b'}$ $(b' \neq b)$ and 0 $(b' = b)$, and if in the second (that of the D_b's)we weight by $\rho_{\cdot b}$, we obtain, following Faleschini,

$$\sum_b \frac{\rho_{a^*(b)b} - \rho_{a^*(b)} \cdot \rho_{\cdot b}}{1 - \rho_{\cdot b}}.$$

Faleschini appears to feel that this kind of measure should only be used when $\rho_{a^*(b)b}/\rho_{\cdot b} \geq \rho_{a^*(b)b'}/\rho_{\cdot b'}$ for each b and b', but we are not wholly clear about his intent. One difficulty with Faleschini's suggestion is that of interpreting averages of conditional probabilities. Nonetheless, Faleschini's discussion [44] is in terms of a probability model, the drawing of colored balls from urns.

Andreoli. Finally, we wish to mention two articles by G. Andreoli, [1] and [2]. Among the topics discussed is that of association between characteristics of one individual and a *group* of individuals, for example between occupation of father and occupations of his *several* sons.

4.8. *Problems of inference discussed by Wilson, Berkson, and Mainland.* We should like to call attention to three papers in the medical literature that are of interest in connection with measures of association, especially with respect to the very difficult problem of inference from one population to another.

The first is by E. B. Wilson [146]. Wilson emphasizes the importance of specifying the population carefully. For example, consider the 2×2 table

	Dead with evidence of cancer	Not [dead with evidence of cancer]
Dead with evidence of tuberculosis		
Not [dead with evidence of tuberculosis]		

If this table is filled in from the data of a large number of autopsies (so that all individuals represented in the table are dead) one may obtain a very different picture than if the table is filled in from the entire population, alive at a given time and observed one year later.

The second paper is by Joseph Berkson [7]. It considers examples like the above with emphasis on differential selection as a cause of confusion. Berkson proposes a specific mechanism for differential selection in the case of one study of the relation between smoking and lung cancer.

The third paper is by Donald Mainland [106]. He gives in considerable detail an example showing how differential selection can lead to a grossly fallacious inference.

4.9. *Measures based on latent structures.* We have already discussed the 2×2 case measures of association based on latent structures that have been suggested by Peirce (Section 3.1), and Benini (Section 3.3). Both authors suggested that the observable 2×2 cross classification might be regarded as an average or mixture of two or more underlying cross classifications having special characteristics, e.g., independence. The underlying cross classifications are those of the latent classes. One may then take as a measure of association a numerical characteristic of the latent class probabilities together with the averaging or mixing weights, provided that this characteristic is expressible as a function of the four probabilities in the observable cross classification. The latent class structure, which may be considered as either real or fanciful, then provides an interpretation for the proposed measure of association.

Lazarsfeld and Kendall. More recently, Paul F. Lazarsfeld has written extensively about latent class structures; it was indeed Lazarsfeld who introduced the term "latent structure." Although much of Lazarsfeld's work on latent structures has been concerned with much broader problems, he and Patricia Kendall [88, Appendix A] have discussed measures of association based on latent classes in the 2×2 case. We describe first their "index of turnover."

The sort of 2×2 cross classification that Lazarsfeld and Kendall discuss might result from asking people the same yes-or-no question at two different times. The supposed latent structure is that there are really two classes of people in the population of interest, those whose latent attitude towards the question is "Yes," in proportion K_1, and those whose latent attitude is "No," in proportion $K_2=1-K_1$. The actual answers that people give do not, however, always express their latent attitudes, since they may be temporarily swayed in the other direction, may misunderstand, and so on. Suppose that the "Yes" people answer "No" with probability x, and that the "No" people answer "Yes" with probability y. Responses are supposed independent for the people in a given class. Further, in order that the latent structure make sense, we suppose that x and y are $\leq\frac{1}{2}$.

If, now, we choose at random a member of the population, the following four probabilities, arranged in 2×2 form, describe the distribution of his two responses:

Second answer

		Yes	No	Totals
First Answer	Yes	$\rho_{11}=K_1(1-x)^2+K_2y^2$	$\rho_{12}=K_1(1-x)x+K_2y(1-y)$	$\rho_{1\cdot}=K_1(1-x)+K_2y$
	No	$\rho_{21}=K_1x(1-x)+K_2(1-y)y$	$\rho_{22}=K_1x^2+K_2(1-y)^2$	$\rho_{2\cdot}=K_1x+K_2(1-y)$
	Totals	$\rho_{\cdot1}=K_1(1-x)+K_2y$	$\rho_{\cdot2}=K_1x+K_2(1-y)$	1

This is the observable 2×2 cross classification. Following our general approach, we suppose it known and postpone discussion of sampling problems. Note that $\rho_{12}=\rho_{21}$ and that $\rho_{i\cdot}=\rho_{\cdot i}$ $(i=1, 2)$. There are two independent probabilities among the four of the 2×2 table, and three independent parameters of the latent structure, so one cannot hope to express these parameters in terms of the probabilities. If, however, one assumes that $x=y$, i.e., that the probability of a deviant response is the same for both the "Yes" and "No" latent classes, then the core of the above table simplifies to

$\rho_{11}=x^2-2K_1x+K_1$	$\rho_{12}=x(1-x)$	$\rho_{1\cdot}=K_1(1-2x)+x$
$\rho_{21}=x(1-x)$	$\rho_{22}=x^2-2x(1-K_1)+(1-K_1)$	$\rho_{2\cdot}=-K_1(1-2x)+1-x$

Hence $x=\frac{1}{2}[1\pm\sqrt{1-4\rho_{12}}]$. Since we have assumed $x\leq\frac{1}{2}$, the minus sign should be chosen. Thus $x=\frac{1}{2}[1-\sqrt{1-4\rho_{12}}]$ measures an aspect of association that has a real interpretation in the context of the stated latent structure, since x is the probability of a deviant response. Also $2x(1-x)$ is the probability that a random person answers the question differently at the two times; whence the descriptive term "turnover." And $1-2x(1-x)$ is the probability that a random person answers the question similarly.

One can also easily express K_1 in terms of the ρ's, since

$$K_1 = (\rho_{1\cdot} - x)/(1 - 2x)$$
$$= \frac{1}{2} + \frac{2\rho_{1\cdot} - 1}{2\sqrt{1 - 4\rho_{12}}}\,.$$

Further, independence obtains if and only if either $K_1=0$ or 1, or $x=\frac{1}{2}$. Thus x measures an aspect of association, unless $K_1=0$, 1.

A serious difficulty with the above latent structure is that it places severe limitations on the ρ's; only a limited set of 2×2 cross classifications can be fit by it. In fact, it is necessary and sufficient that

$$(1)\ \rho_{12} \leq \tfrac{1}{4}, \qquad (2)\ \rho_{21} = \rho_{21}, \qquad \text{and} \qquad (3)\ \rho_{11} \geq \rho_{1\cdot}\rho_{\cdot1}$$

for a 2×2 cross classification to be describable in terms of the above latent structure.

Kendall and Lazarsfeld also discuss a more general measure, appropriate to some cases in which $\rho_{12}\neq\rho_{21}$, by enlarging the model to embrace three, rather than two, latent classes with special characteristics. In order to exemplify the possibilities, we should like to suggest a new measure that may be more appropriate to some cases in which $\rho_{12}\neq\rho_{21}$. Which measures to use, if any, depends of course on context. The measure we shall now describe might be appropriate when two closely related questions are both asked once, rather than when the same question is asked twice, and we describe it in these terms.

Suppose that on question 1 people give deviant answers (e.g. a "yes" person answers "no") with probability $x_1\leq\frac{1}{2}$, and that on question 2 they give deviant answers with probability $x_2\leq\frac{1}{2}$. The probabilities of deviant response do not depend on the class to which a person belongs. In all other respects the latent structure is the same as before. We then have three independent parameters, K_1, x_1, and x_2 for describing our structure, and the 2×2 table becomes

		Answer to question 2: Yes	No	Totals
Answer to question 1	Yes	$\rho_{11}=K_1(1-x_1)(1-x_2)+K_2x_1x_2$	$\rho_{12}=K_1(1-x_1)x_2+K_2x_1(1-x_2)$	$\rho_{1\cdot}=K_1(1-x_1)+K_2x_1$
	No	$\rho_{21}=K_1x_1(1-x_2)+K_2(1-x_1)x_2$	$\rho_{22}=K_1x_1x_2+K_2(1-x_1)(1-x_2)$	$\rho_{2\cdot}=K_1x_1+K_2(1-x_1)$
	Totals	$\rho_{\cdot1}=K_1(1-x_2)+K_2x_2$	$\rho_{\cdot2}=K_1x_2+K_2(1-x_2)$	1

We may now express K_1, x_1, and x_2 in terms of the ρ's; and x_1 and x_2—thus expressed—are interpretable measures of association in terms of the supposed latent structure. They are the probabilities of deviant responses to the two questions. In order to get a single measure, one might take the average of x_1 and x_2; that is, the probability of deviant response to one of the two questions, which one to be decided by the toss of a fair coin. Or one might use $x_1x_2 + (1-x_1)(1-x_2)$, the probability that a random person answers the two questions similarly.

It is easily seen from the above table that

$$x_1 = \frac{\rho_{1\cdot} - K_1}{1 - 2K_1}, \qquad x_2 = \frac{\rho_{\cdot 1} - K_1}{1 - 2K_1}$$

and that

$$\frac{\rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}}{1 - 2(\rho_{12} + \rho_{21})} = K_1(1 - K_1) = R \quad \text{(say).}$$

Hence

$$K_1 = \tfrac{1}{2}[1 \pm \sqrt{1 - 4R}]$$

and we see that, for our latent structure to hold, R, as a function of the ρ's, must be $\leq \frac{1}{4}$. Substituting in the above expressions for x_1 and x_2, we obtain

$$x_1 = \frac{1}{2} \mp \frac{2\rho_{1\cdot} - 1}{2\sqrt{1 - 4R}}$$

$$x_2 = \frac{1}{2} \mp \frac{2\rho_{\cdot 1} - 1}{2\sqrt{1 - 4R}}$$

There remains the question about sign choice in the solution of the quadratic for K_1. We want to be able to make the same choice for both x_1 and x_2 so that x_1 and x_2 are $\leq \frac{1}{2}$. This means that $\rho_{1\cdot} - \frac{1}{2}$ and $\rho_{\cdot 1} - \frac{1}{2}$ must have the same sign in the sense that $(\rho_{1\cdot} - \frac{1}{2})(\rho_{\cdot 1} - \frac{1}{2}) \geq 0$. The necessary conditions thus far suggested come to (1) $\rho_{12} + \rho_{21} \leq \frac{1}{2}$, (2) $(\rho_{1\cdot} - \frac{1}{2})(\rho_{\cdot 1} - \frac{1}{2}) \geq 0$, and (3) $\rho_{11} \geq \rho_{1\cdot}\rho_{\cdot 1}$.

Note that if $\rho_{12} = \rho_{21}$, then $\rho_{1\cdot} = \rho_{\cdot 1}$, $x_1 = x_2$, and

$$1 - 4R = \frac{1 - 4\rho_{12} - 4\rho_{11} + 4\rho_{1\cdot}^2}{1 - 4\rho_{12}} = \frac{(1 - 2\rho_{1\cdot})^2}{1 - 4\rho_{12}}.$$

Hence

$$x_1 = x_2 = \frac{1}{2} \mp \frac{2\rho_{1\cdot} - 1}{2(1 - 2\rho_{1\cdot})}\sqrt{1 - 4\rho_{12}} = \frac{1}{2}[1 \pm \sqrt{1 - 4\rho_{12}}]$$

and the minus sign must be chosen, to obtain the same result as in the earlier structure. So the structure now being discussed does generalize the earlier one, giving us two turnover indexes.

For the present structure, independence obtains if and only if K_1 is 0 or 1, or if either x_1 or $x_2 = \frac{1}{2}$. Thus we see again that x_1 and x_2 measure aspects of association, unless $K_1 = 0, 1$.

Necessary and sufficient conditions for the present structure to be possible may be expressed in various ways. One such set of conditions is the following pair:

$$0 \leq \frac{\rho_{11} - \rho_{1\cdot}\rho_{\cdot 1}}{1 - 2(\rho_{12} + \rho_{21})} \leq \text{Min } [\rho_{1\cdot}\rho_{2\cdot}, \rho_{\cdot 1}\rho_{\cdot 2}] \tag{1}$$

$$(\rho_{1\cdot} - \tfrac{1}{2})(\rho_{\cdot 1} - \tfrac{1}{2}) \geq 0. \tag{2}$$

4.10. *More recent work on measures of association in meteorology. Gringorten, Bleeker, Brier, and others.* In Section 3.1, we discussed measures of association suggested by Peirce, Doolittle, Köppen and others for meteorological problems. Meteorologists have of course long been interested in the accuracy of weather forecasts, and they have suggested many measures of association between the predicted weather and the weather that actually occurred.

We shall not attempt to survey the large literature of this field in detail, especially since three relatively recent articles provide extensive reviews of it. The first, by R. H. Muller [109], gives abstracts of some 55 relevant publications prior to 1944, including most of those described in Section 3.1. (See Clayton [23] for criticism of Muller's abstracts of Clayton's work.) The second, by W. Bleeker [10], includes references to a number of continental articles not mentioned by Muller, and analyzes a number of proposed measures in detail, especially as regards the behavior of a predictor who knows that his predictions will be compared with actuality by a particular measure. The third, by G. W. Brier and R. A. Allen [17] discusses key publications appearing up to 1951. In the following paragraphs, we want to mention a few articles of particular interest to us, especially some published since the three surveys cited above.

The simplest case of interest to the meteorologists is where there is no order in the classifications and an asymmetrical interest in the two classifications. Sometimes the classifications are different, as when one is considering a particular qualitative variable as a predictor of qualitative weather. For this case, a measure of association based on the Shannon-Wiener information notion has been suggested by J. L. Holloway, Jr. and M. A. Woodbury [81] and has been used by several meteorologists, notably I. I. Gringorten and his colleagues. We have referred to it in Section 4.6. Gringorten [70, pp. 69–70] also suggests independently the same proportional prediction measure described in [66, Section 9]. This measure is very natural if we think of the possibility of making probabilistic, rather than categorical, forecasts, a possibility to which we shall recur in a few paragraphs. Gringorten's article also gives a brief general survey of measures of association in the meteorological context.

Sometimes the two classifications are the same, as when one is considering association between a categorical forecast and a categorical event, with both forecast and event classified in the same way. In this case of "forecast verification" both the above measures may be used, as well as others that take the identity of the two classifications explicitly into account. The use of association measures in connection with meteorological prediction, both with and without order taken into account, is considered by van der Bijl [140].

A more complex situation is that in which some third classification is brought into the picture. One important example is the three-way classification: forecast weather—observed weather—weather at time of forecast. Here interest is usually centered in the extent to which the forecaster can improve on persistence forecasting or on forecasting based on climatic information conditional upon weather at forecast time. Some materials referred to in the above paragraphs bear upon this situation; we should also like to cite two articles by Gringorten, [68] and [71a], and a closely related report by Gringorten, Lund, and Miller [69]. These references use scoring schemes with scores based on probabilities. Gringorten [68] makes it very clear that the appropriate measure depends upon the question being asked. In [71], Gringorten works on the sampling problem for measures based on scores.

An interesting problem is that of the construction of meaningful measures of association when the forecast is not categorical, but rather is itself a discrete probability distribution over a set of weather categories. Thus, for example, a prediction might be

No rain (probability .1)
Light rain (probability .6)
Heavy rain (probability .3)

and this prediction would be compared with that one of the three possibilities that later actually occurred. Suggestions for this kind of forecast prediction appear to go back at least to World War I, but it seems to have become of general interest only recently. Two recent articles relating to probabilistic forecasts are by G. W. Brier [16] and W. G. Leight [98].

If we attempt to construct a measure of association between probabilistic forecasts and the actual events later observed, we are faced with association between an essentially continuous distribution on a $k-1$ dimensional simplex (k categories, probabilities for each that sum to one) and a discrete distribution on k points (for the actual events).

Several articles take up Peirce's 1884 theme relating to economic losses as an important factor in evaluating forecast utility. For the 2×2 case, we refer to E. G. Bilham [9], H. C. Bijvoet and W. Bleeker [8], J. C. Thompson [135], J. C. Thompson and G. W. Brier [136], and G. W. Brier [18]. Gringorten [68 and 71a] considers more general cases by means of scores based directly on net losses.

4.11. *Association between species. Forbes, Cole, Goodall.* In the ecological literature there is a series of articles dealing with 2×2 cross classifications of the following kind:

NUMBERS OF AREAS IN WHICH SPECIES A AND SPECIES B ARE OR ARE NOT FOUND

		B: Found	B: Not Found	
A	Found	N_{11}	N_{12}	
	Not Found	N_{21}	N_{22}	
				n

Thus, for example, in N_{11} out of n marshes examined, grasses of species A and B are both found, while in N_{12} out of n marshes, species A is found but not species B.

A review and bibliography of ecological articles dealing with measures of association in this context is given by Goodall [64, pp. 221–3]. The series seems to have started with an article by Forbes [49] in 1907, followed by a long gap, and then a number of more recent articles. Of these, a particularly extensive one is by Cole [28], in which Benini's measure (see Section 3.3) was independently proposed.

4.12. *Association between anthropological traits. Tylor, Clements, Wallis, Driver, Kroeber, Chrétien, Kluckhohn, and others.* We have already discussed (Section 4.4) a proposal by the anthropologist, F. Boas. We now turn to a more special case than the one discussed by Boas, the 2×2 cross classification. Writers in the fields of anthropology and linguistics have long been concerned with 2×2 cross classifications similar to those discussed in the last section. The earliest paper of which we know that deals at all with measures of association in these fields is by Edward B. Tylor [139] in 1889. Tylor discussed many examples of association between cultural traits, some dichotomous and some trichotomous, but he contented himself with observing sizable apparent deviations from independence and did not suggest any numerical measures of association. In the ensuing discussion Francis Galton said [139, p. 270] that "... the degree of interdependence might with advantage be expressed in terms of a scale in which 0 represented perfect independence and 1 complete concurrence." We now list and discuss briefly those subsequent papers of which we know in this area that seem to us most germane to our survey.

In 1911, Jan Czekanowski [30], explicitly carrying Tylor's work forward, discussed the use of Yule's Q in ethnology and anthropology. Czekanowski also published a number of further papers dealing with 2×2 classifications.

In 1926, Forrest E. Clements and others [24] used the values of χ^2 and the resulting P-values in an examination of traits held in common by various Polynesian societies. An interesting controversy between Clements and Wilson D. Wallis [25, 143] followed. Wallis attacked Clements and his coauthors for using oversimplifying statistical methods and for drawing unjustified anthropological conclusions by these methods. Another article by Clements [26], discussing Q and ϕ prefixed by the appropriate $\pm$ sign, appeared in 1931. A quite recent article [27] by Clements goes over the same ground with added comments on subsequent literature.

In 1932, H. E. Driver and A. L. Kroeber [38] commented on the Clements-Wallis controversy, and used the following three measures in analyzing association between various pairs of societies:

$$\frac{\rho_{11}}{2}\left(\frac{1}{\rho_{1\cdot}}+\frac{1}{\rho_{\cdot1}}\right),\qquad \frac{\rho_{11}}{\sqrt{\rho_{1\cdot}\rho_{\cdot1}}},\qquad \frac{\rho_{11}}{1-\rho_{22}}.$$

The 2×2 cross classifications to which these were applied referred to populations of traits, and took the following form:

		Society B Has	Society B Has not	
Society A	Has	ρ_{11}	ρ_{12}	$\rho_{1\cdot}$
	Has not	ρ_{21}	ρ_{22}	$\rho_{2\cdot}$
		$\rho_{\cdot 1}$	$\rho_{\cdot 2}$	

so that ρ_{12}, for example, is the proportion of traits observed present in Society A and absent in Society B.

In 1937, A. L. Kroeber and C. D. Chrétien [94] applied 2×2 measures of association to linguistic classification. Several measures were discussed and compared. Such application to linguistics continued in several articles, notably [20]. A recent article by Chrétien in this line is [22]. It is interesting to observe that the article immediately following [22], by Joseph H. Greenberg [67], is one of the few instances we know in which descriptive statistics are constructed so as to have operational interpretations in the sense that we have discussed. Greenberg's suggestions relate to measuring concentration in a single classification, or multinomial, population.

In 1939, a critical survey of the application of measures of association to ethnological data was published by Clyde Kluckhohn [90]. This very interesting article contains an extensive bibliography, and it marshals many arguments for and against the use of measures of association in anthropological contexts.

Driver [39], in the same year, compared in detail formal properties and relations between some eight 2×2 measures of association. He was much concerned with the effect of nonuniform marginal distributions on comparisons between values of 2×2 measures.

In 1945, Chrétien [21] discussed a number of basic points, including several analyzed by Kluckhohn, regarding the use of measures of association. Here, for almost the first and only time in this line of papers, we find the problem of interpretation raised as Chrétien says (p. 488): "Primary in importance, it seems to me, is the need to determine more precisely the meaning of the scale of association. All association studies to date have confined their attention to the high positive values."

Finally, we wish to cite a 1953 survey article by Driver [40]. In its section on ethnology and social anthropology, there appears a discussion of measures of association for the 2×2 case.

4.13. *Other suggestions.* We conclude by listing a few other suggestions relating to measures of association that do not fall naturally into the above classification.

Harris, Pearson. In a number of articles by J. A. Harris and others, [74], [75] and [76], there is a discussion of the following situation: Sometimes the existence of observations (individuals) in certain cells of a cross classification table is arithmetically, physically, or otherwise impossible. Harris and his coauthors discuss the effect of this inherent emptiness of some cells on certain traditional measures of association, and suggest modifications of these measures. K. Pearson, commenting on Harris' papers in [113], discusses the computation of the coefficient of mean square contingency when careful *a priori* consideration indicates that for certain cells the appearance of individuals in those cells is impossible. With the use of measures of association that have operational meaning, rather than the coefficient of mean square contingency, the occurrence of zero frequencies in certain cells does not seem to us to be of special significance. See Sec. 2.1. The *a priori* considerations leading to the belief about zero frequencies may, however, suggest alternative ways of setting up the classifications that are more meaningful.

Irwin. In 1934, J. O. Irwin [84] commented on measures of association and emphasized the importance of relating the use of such measures to the goals of the particular investigation at hand. He says (p. 87) that "... we should [not] do away with correlation coefficients or other measures of association, but should try to make the end point of our statistical analysis not a single coefficient which may be hard to interpret, but a result bearing a 'physical' meaning; the more easily the result may be understood by an intelligent layman, the better we should regard it as expressed." Irwin ends his article by describing a particular case of careful and useful analysis based on measures of association applied to the data in various ways.

It seems to us that, when the operational interpretation viewpoint towards association measures is taken, one is automatically influenced away from sterile arguments about which measure is "best." For if different measures reflect different aspects of the population, no one is best in any abstract sense (although one may be most appropriate in a given case) and there is no reason why more than one should not be used. An analogy is to ask about measures of size for human beings. One might suggest weight, height, volume, girth, etc., but no one of these is best except perhaps in a particular context.

Lakshmanamurti. In [97], M. Lakshmanamurti suggested a rather complex measure of association for the 2×2 case and compared it with Yule's Q.

Fairfield Smith. In a recent article [126a] H. Fairfield Smith has complained entertainingly about the difficulty of interpreting conventional measures of association. Most of his article shows by example how one may compare two sample cross classifications by forming simple chi-square tests that emphasize some specific aspect of possible difference between the cross classifications.

We end this paper with a quotation [126a, pp. 72–3] that expresses Smith's dismay about the vague or nonexistent meaning of most association measures.

> "What can be the use to know that ghosts in my lord's and lady's chambers each wore a sash with the symbol .6 if we do not know how the sash or its decoration may reflect the more earthy bodies from which the ghosts have been supposed to emanate?"

REFERENCES

[1] Andreoli, Giulio, "Sulla definizione di certi indici, relativi a caratteri di omogamia, in problemi statistici," *Rendiconto dell'Accademia delle Scienze Fisiche e Matematiche (Classe della Società Reale di Napoli)*, Ser. 4, 4 (1934), 36–41.

[2] Andreoli, Giulio, "Teoria generale di certi indici nei fenomeni statistici (omagamia, endogamia, diffusione, etc.)," *Rendiconto dell'Accademia delle Scienze Fisiche e Matematiche (Classe della Società Reale di Napoli)*, Ser. 4, 5 (1935), 108–24.

[3] Bachi, Roberto, "Gli indici della attrazione matrimoniale," *Giornale degli Economisti e Rivista di Statistica*, 69 (1929), 894–938.

[4] Benini, Rodolfo, *Principii di Demografia*, Firenzi, G. Barbèra, 1901. No. 29 of *Manuali Barbèra di Scienze Giuridiche Sociali e Politiche.*

[5] Benini, Rodolfo, "Grupi chiusi e gruppi aperti in alcuni fatti colletivi di combinazioni," *Bulletin de l'Institut International de Statistique*, 23 (1928), 362–83.

[6] Berkson, Joseph, "Cost utility as a measure of the efficiency of a test," *Journal of the American Statistical Association*, 42 (1947), 246–55.

[7] Berkson, Joseph, "The statistical study of association between smoking and lung cancer," *Proceedings of the Staff Meetings of the Mayo Clinic*, 30 (1955), 319–48.

[8] Bijvoet, H. C., and Bleeker, W., "The value of weather forecasts," *Weather*, 6 (1951), 36–9.

[9] Bilham, E. G., "A problem in economics," *Nature*, 109 (1922), 341–2.

[10] Bleeker, W., "The verification of weather forecasts," *Mededeelingen en Verhandelingen*, Ser. B, Deel 1, No. 2 (1946); Koninklijk Nederlandsch Meteorologisch Instituut, No. 125.

[11] Boas, Franz, "The measurement of differences between variable quantities," *Journal of the American Statistical Association*, 18 (1922–23), 425–45.

[12] Bogue, Donald J., *The Structure of the Metropolitan Community*. University of Michigan, Ann Arbor, 1949.

[12a] Bonferroni, C. E., "Nuovi indici di connessione fra variabili statistiche," *Pubblicazioni dell'Istituto di Statistica*, Università Commerciale "Luigi Bocconi," Milan. Vol. I (1942) of *Studi Sulla Correlazione e Sulla Connessione*, 57–100.

[13] Bonferroni, Carlo Emilio, "Indici unilaterali e bilaterali di connessione," 49–60 in: Mazzoni, Pacifico, and Lasorsa, Giovanni (ed.), *Studi in Memoria di Rodolfo Benini*, Facoltà di Economia e Commercio, Università degli Studi di Bari, Bari, 1956.

[14] Brambilla, Francesco, *La Variabilità Statistica a Due Dimensioni. La Teoria Statistica della Correlazione e della Connessione.* Università Commerciale "Luigi Bocconi," Istituto di Statistica, Casa Editrice Ambrosiana, Milan, 1949. This volume is part 2 of Brambilla's *Manuale di Statistica.*

[15] Bresciani, Costantino, "Sui metodi per la misura delle correlazioni," *Giornale degli Economisti*, 38 (1909), 401–14, 491–522.

[16] Brier, Glenn W., "Verification of forecasts in terms of probability," *Monthly Weather Review*, 78 (1950), 1–3.

[17] Brier, Glenn W., and Allen, Roger A., "Verification of weather forecasts," 841–8 in: Malone, Thomas F. (ed.), *Compendium of Meteorology*, American Meteorological Society, Boston, 1951.

[18] Brier, G. W., "The effect of errors in estimating probabilities on the usefulness of probability forecasts," *Bulletin of the American Meteorological Society*, 38 (1957), 76–8.

[19] Cartwright, Desmond S., "A rapid non-parametric estimate of multi-judge reliability," *Psychometrika*, 21 (1956), 17–29.

[20] Chrétien, C. Douglas, "The quantitative method for determining linguistic relationships. Interpretation of results and tests of significance," *University of California Publications in Linguistics*, 1 (1943–8), 11–19. Published in 1943.

[21] Chrétien, C. Douglas, "Culture element distributions: XXV. Reliability of statistical procedures and results," *University of California Publications in Anthropological Records*, 8 (1942–5), 469–90. Published in 1945.

[22] Chrétien, C. Douglas, "Word distributions in Southeastern Papua," *Language*, 32 (1956), 88–108.

[23] Clayton, H. H., "On verification of weather forecasts," *Bulletin of the American Meteorological Society*, 25 (1944), 368.

[24] Clements, Forrest E., Schenck, Sara M., and Brown, T. K., "A new objective method for showing special relationships," *American Anthropologist*, 28 (1926), 585–604.

[25] Clements, Forrest E., "Quantitative method in ethnography," *American Anthropologist*, 30 (1928), 295–310.

[26] Clements, Forrest, "Plains Indian tribal correlations with sun dance data," *American Anthropologist*, 33 (1931), 216–27.

[27] Clements, Forrest E., "Use of cluster analysis with anthropological data," *American Anthropologist*, 56 (1954), 180–99.

[28] Cole, La Mont C., "The measurement of interspecific association," *Ecology*, 30 (1949), 411–24.

[29] Cramér, Harald, "Remarks on correlation," *Skandinavisk Aktuarietidskrift*, 7 (1924), 220–40.

[30] Czekanowski, Jan, "Objective Kriterien in der Ethnologie," *Mitteilungen der Anthropologischen Gesellschaft in Wien*, 42 (1912), Sitzungsberichte (1911–12) Section, 17–21.

[31] de Meo, G., "Su di alcuni indici atti a misurare l'attrazione matrimoniale in classificazioni dicotome," *Rendiconto dell'Accademia delle Scienze Fisiche e Matematiche* (*Classe della Società Reale di Napoli*), Ser. 4, 4 (1934), 62–77.

[32] Deuchler, Gustav, "Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie," *Zeitschrift für Pädagogische Psychologie und Experimentelle Pädagogik*, 15 (1914), 114–31, 145–59, and 229–42.

[33] Deuchler, Gustav, "Über die Bestimmung von Rangkorrelationen aus Zeugnisnoten," *Zeitschrift für Angewandte Psychologie*, 12 (1916–17), 395–439.

[34] Deuchler, Gustav, "Über die Bestimmung einseitiger Abhängigkeit in pädagogisch-psychologischen Tatbeständen mit alternativer Variabilität," *Zeitschrift für Pädagogische Psychologie und Experimentelle Pädagogik*, 16 (1915), 550–66.

[35] Doolittle, M. H., "The verification of predictions" (abstract), *Bulletin of the Philosophical Society of Washington*, 7 (1885), 122–7. A summary of this article appears in *The American Meteorological Journal*, 2 (1885–6), 327–9.

[36] Doolittle, M. H., "Association ratios" (abstract), *Bulletin of the Philosophical Society of Washington*, 10 (1888), 83–7 and 94–6.

[37] Dornbusch, Sanford M., and Schmid, Calvin F., *A Primer of Social Statistics*, McGraw-Hill, New York, 1955.

[38] Driver, H. E., and Kroeber, A. L., "Quantitative expression of cultural relationships," *University of California Publications in American Archaeology and Ethnology*, 31 (1931–3), 211–56.

[39] Driver, Harold E., "Culture element distributions: X. Northwest California," *University of California Publications in Anthropological Records*, 1(6) (1939), 297–433.

[40] Driver, Harold E., "Statistics in anthropology," *American Anthropologist*, 55 (1953), 42–59.

[41] Duncan, Otis Dudley, Ohlin, Lloyd E., Reiss, Albert J., Jr., and Stanton, Howard R., "Formal devices for making selection decisions," *The American Journal of Sociology*, 58 (1953), 573–84.

[42] Duncan, Otis Dudley, and Duncan, Beverly, "A methodological analysis of segregation indexes," *American Sociological Review*, 20 (1955), 210–7.

[43] Eyraud, Henri, "Les principes de la mesure des corrélations," *Annales de l'Université de Lyon*, Ser. A (3), 1 (1936), 30–47.

[44] Faleschini, Luigi, "Indici di connessioni," *Contributi del Laboratorio di Statistica, Ser. 6, Università Cattolica del Sacro Cuore* [*Milan*], *Publicazioni, Nuova Serie*, 21 (1948), 112–51.

[45] Féron, Robert, "Mérites comparés des divers indices de corrélation," *Journal de la Société Statistique de Paris*, 88 (1947), 328–50. Discussion on pp. 350–2.

[46] Féron, Robert, "Information, Régression, Corrélation," *Publications de l'Institut de Statistique de l'Université de Paris*, 5(3–4) (1956), 113–215.

[47] Finley, Jno. P., "Tornado predictions," *The American Meteorological Journal*, 1 (1884), 85–8.

[48] Florence, P. Sargant, *Investment, Location, and Size of Plant*. Cambridge University Press, Cambridge, 1948.

[49] Forbes, S. A., "On the local distribution of certain Illinois fishes: an essay in statistical ecology," *Bulletin of the Illinois State Laboratory of Natural History*, 7 (1904–9), 273–303, plus 32 plates.

[50] Fréchet, Maurice, "A general method of constructing correlation indices," *Proceedings of the Mathematical and Physical Society of Egypt*, 3 (2) (1946), 13–20. (Additional note 3 (4) (1948), 73–4.)

[51] Fréchet, Maurice, "Anciens et nouveaux indices de corrélation. Leur application au calcul des retards économiques," *Econometrica*, 15 (1947), 1–30. (Errata pp. 374–5.)

[52] Fréchet, Maurice, "Sur les tableaux de corrélation dont les marges sont données," *Annales de l'Université de Lyon*, Sec. A (3), 14 (1951), 53–77.

[53] Fréchet, Maurice, "Sur les tableaux de corrélation dont les marges sont données," *Comptes Rendus Hebdomadaires des Scéances de l'Académie des Sciences, Paris*, 242 (1956), 2426–8.

[54] Galvani, L., "A propos de la communication de M. Thionet: L'école moderne de statisticiens italiens," *Journal de la Société de Statistique de Paris*, 88 (1947), 196–203.

[55] Gilbert, G. K., "Finley's tornado predictions," *The American Meteorological Journal*, 1 (1884), 166–72.

[56] Gini, Corrado, "Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Series 8, 74 (2) (1914–5), 185–213.

[57] Gini, Corrado, "Indice di omofilia e di rassomiglianza e loro relazioni col coefficiente di correlazione e con gli indici di attrazione," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Series 8, 74 (2), (1914–5), 583–610.

[58] Gini, Corrado, "Nuovi contributi alla teoria delle relazioni statistiche," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Series 8, 74 (2), (1914–5), 1903–42.

[59] Gini, Corrado, "Sul criterio di concordanza tra due caratteri," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*, Series 8, 75 (2), (1915–6), 309–31.

[60] Gini, Corrado, "Indici di concordanza," *Atti del Reale Istituto Veneto di Scienze, Lettere ed Arti*. Series 8, 75 (2), (1915–6), 1419–61.

[61] Gini, Corrado, "The contributions of Italy to modern statistical methods," *Journal of the Royal Statistical Society*, 89 (1926), 703–24.

[61a] Gini, Corrado, "Die Messung der Ungleichheit zweier Verteilungen, angewendet auf die Untersuchung von qualitativen Rassenmerkmalen," *Archiv für Mathematische Wirtschafts- und Sozialforschung*, 3 (1937), 167–84, plus two appendixes by Vittorio Castellano.

[62] Gini, Corrado, *Metodologia Statistica: la Misura dei Fenomeni Collettivi*, Vol. III, Part 2a, Chap. 55, 245–321, *Enciclopedia delle Matematiche Elementari*, edited by Luigi Berzolari, Ulrico Hoepli, Milan, 1948.

[63] Gini, Corrado, *Corso di Statistica*, a cura di S. Gatti e Benedetti, nuova edizione aggiornata, Università degli Studi di Roma, Facoltà di Scienze Statistiche, Demografiche ed Attuariali, Libreria Eredi Virgilio Veschi, Rome, 1954–5.

[63a] Glass, D. V. (ed.), *Social Mobility in Britain*, Routledge and Kegan Paul, London, 1954.

[64] Goodall, D. W., "Quantitative aspects of plant distribution," *Biological Reviews of the Cambridge Philosophical Society*, 27 (1952), 194–245.

[65] Goodman, Leo A., "On urbanization indices," *Social Forces*, 31 (1953), 360–2.

[66] Goodman, Leo A., and Kruskal, William H., "Measures of association for cross classifications," *Journal of the American Statistical Association*, 49 (1954), 723–64.

[67] Greenberg, Joseph H., "The measurement of linguistic diversity," *Language*, 32 (1956), 109–15.

[68] Gringorten, Irving, I., "The verification and scoring of weather forecasts," *Journal of the American Statistical Association*, 46 (1951), 279–96.

[69] Gringorten, Irving I., Lund, Iver A., and Miller, Martin A., "A program to test skill in terminal forecasting," Air Force Surveys in Geophysics, No. 80, Geophysics Research and Development Command, June, 1955.

[70] Gringorten, Irving I., "Methods of objective weather forecasting," pp. 57–92 in: Landsberg, H. E. (ed.), *Advances in Geophysics*, Vol. 2, Academic Press, New York, 1955.

[71] Gringorten, Irving I., "Tests of significance in a verification program," *Journal of Meteorology*, 12 (1955), 179–95.

[71a] Gringorten, Irving I., "On the comparison of one or more sets of probability forecasts," *Journal of Meteorology*, 15 (1958), 283–7.

[72] Gringorten, Irving I., Lund, Iver A., and Miller, Martin A., "The construction and use of forecast registers," Geophysical Research Papers, No. 53, Geophysics Research Directorate, Air Force Cambridge Research Center, Air Research and Development Command, June, 1956.

[73] Guilford, J. P., *Fundamental Statistics in Psychology and Education*, McGraw-Hill, New York, First edition, 1942, second edition, 1950.

[73a] Halphen, Étienne, "L'analyse intrinsèque des distributions de probabilité," *Publications de l'Institut de Statistique de l'Université de Paris*, 6 (2) (1957), 79–159.

[74] Harris, J. Arthur, and Treloar, Alan E., "On a limitation in the applicability of the contingency coefficient," *Journal of the American Statistical Association*, 22 (1927), 460–72.

[75] Harris, J. Arthur, Treloar, Alan E., and Wilder, Marian, "On the theory of contingency II. Professor Pearson's note on our paper on contingency," *Journal of the American Statistical Association*, 25 (1930), 323–7.

[76] Harris, J. Arthur, and Tu, Chi, "A second category of limitation in the applicability of the contingency coefficient," *Journal of the American Statistical Association*, 24 (1929), 367–75.

[77] Hazen, H. Allen, "Verification of tornado predictions," *American Journal of Science*, Ser. 3, 34 (1887), 127–31.

[78] Höffding, Wassilij, "Massstabinvariante Korrelationstheorie," *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin*, 5 (3) (1940), 181–233.

[79] Höffding, Wassilij, "Massstabinvariante Korrelationsmasse für diskontinuienliche Verteilungen," *Archiv für Mathematische Wirtschafts- und Sozialforschung*, 7 (1941), 49–70.

[80] Höffding, Wassilij, "Stochastische Abhängigkeit und funktionaler Zusammenhang," *Skandinavisk Aktuarietidskrift*, 25 (1942), 200–27.

[81] Holloway, J. Leith, Jr., and Woodbury, Max A., "Application of information theory and discriminant function analysis to weather forecasting and forecast verification," Technical Report No. 1, Meteorological Statistics Project, Institute for Cooperative Research, University of Pennsylvania, Feb., 1955.

[82] Hoover, Edgar M., Jr., "The measurement of industrial localization," *Review of Economic Statistics*, 18 (1936), 162–71.

[83] Hoover, Edgar M., Jr., "Interstate redistribution of population, 1850–1940," *Journal of Economic History*, 1 (1941), 199–205.

[84] Irwin, J. O., "Correlation methods in psychology," *British Journal of Psychology, General Section*, 25 (1934–5), 86–91.

[84a] Johnson, H. M., "Maximal selectivity, correctivity and correlation obtainable in a 2×2 contingency table," *American Journal of Psychology*, 58 (1945), 65–8.

[85] Jordan, Charles, "Les coefficients d'intensité relative de Körösy," *Revue de la Société Hongroise de Statistique* (*Magyar Statisztikai Társaság, Revue*), 5 (1927), 332–45.

[86] Jordan, Charles, "Critique de la corrélation au point de vue des probabilités," *Actes du Colloque Consacré à la Théorie des Probabilités*, No. 740 of *Actualités Scientifiques et Industrielles*, Hermann, Paris, 1938.

[87] Jordan Károly, "A Korreláció számitása (Sur le calcul de la corrélation) I," *Magyar Statisztikai Szemle Kiadványai* (Központi Statisztikai Hivatal), (1941) 1 Szám. Hungarian with French summary.

[87a] Katz, Elihu, and Lazarsfeld, Paul E., *Personal Influence*, The Free Press, Glencoe, Illinois, 1955.

[88] Kendall, Patricia, *Conflict and Mood*, The Free Press, Glencoe, Illinois, 1954.

[89] Klein, Herman J., "Ergebnisse rationeller Prüfungen von Wetterprognosen und deren Bedeutung für die Praxis," *Wochenschrift für Astronomie, Meteorologie und Geographie*, NF 28 (1885), 9–14, 20–2, 25–8, 35–7, 41–4, and 57–61.

[90] Kluckhohn, Clyde, "On certain recent applications of association coefficients to ethnological data," *American Anthropologist*, 41 (1939), 345–77.

[91] Köppen, Wladimir, "Die Aufeinanderfolge der unperiodischen Witterungserscheinungen nach den Grundsätzen der Wahrscheinlichkeitsrechnung," *Repertorium für Meteorologie*, Akademiia Nauk, Petrograd, 2 (1870–1), 189–238.

[92] [Köppen, W.?], "Eine rationelle Methode zur Prüfung der Wetterprognosen," *Meteorologische Zeitschrift*, 1 (1884), 397–404.

[93] Körösi, Joseph, "Kritik der Vaccinations-Statistik und neue Beiträge zur Frage des Impfschutzes," *Transactions of the Ninth International Medical Congress*, Washington, D. C., 1887, Vol. 1, 238–418.

[94] Kroeber, A. L., and Chrétien, C. D., "Quantitative classification of Indo-European language," *Language* 13 (1937), 83–103.

[95] Kruskal, William H., "Historical note on the Wilcoxon unpaired two-sample test," *Journal of the American Statistical Association*, 52 (1957), 356–60.

[96] Kruskal, William H., "Ordinal measures of association," *Journal of the American Statistical Association* 53 (1958), pp. 814-61.

[97] Lakshmanamurti, M., "Coefficient of association between two attributes in statistics," *Proceedings of the Indian Academy of Sciences*, Ser. A, 22 (1945), 123–33.

[98] Leight, Walter G., "The use of probability statements in extended forecasting," *Monthly Weather Review*, 81 (1953), 349–56.

[99] Linfoot, E. H., "An informational measure of association," *Information and Control*, 1 (1957), 85–9.

[100] Lipps, Gottl. Friedr., "Die Bestimmung der Abhängigkeit zwischen den Merkmalen eines Gegenstandes," *Berichte über die Vorhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, 57 (1905), 1–32.

[101] Lipps, G. F., *Die Psychischen Massmethoden*, F. Vieweg und Sohn, Braunschweig, 1906. Vol. 10 of the monograph series, *Die Wissenschaft.*

[102] Loevinger, Jane, "A systematic approach to the construction and evaluation of tests of ability," *Psychological Monographs*, 61 (4), (1947).

[103] Loevinger, Jane, "The technic of homogeneous tests compared with some aspects of 'scale analysis' and factor analysis," *Psychological Bulletin*, 45 (1948), 507–29.

[104] Long, John A., "Improved overlapping methods for determining the validities of test items," *Journal of Experimental Education*, 2 (1934), 264–8.

[105] Lorey, W., "Neue Abhängigkeitsmasse für zufällige Grossen," *Deutsches Statistisches Zentralblatt*, 25 (1933), 153–4.

[106] Mainland, Donald, "The risk of fallacious conclusions from autopsy data on the incidence of diseases with applications to heart disease," *American Heart Journal*, 45 (1953), 644–54.

[107] McGill, William J., "Multivariate information transmission," *Psychometrika*, 19 (1954), 97–116.

[108] Meyer, Hugo, *Anleitung zur Bearbeitung Meteorologischer Beobachtungen für die Klimatologie*, Julius Springer, Berlin, 1891.

[109] Muller, Robert Hans, "Verification of short-range weather forecasts (a survey of the literature)," *Bulletin of the American Meteorological Society*, 25 (1944), 18–27, 47–53, and 88–95.

[110] Niceforo, Alfredo, *La Misura della Vita*, Fratelli, Bocca, Torino, 1919.

[111] Niceforo, Alfredo, *La Méthode Statistique et ses Applications aux Sciences Naturelles aux Sciences Sociales et à l'Art*, Marcel Giard, Paris, 1925. Translation by R. Jacquemin, with some changes, of Niceforo's *Il Metodo Statistico* (1923).

[112] Pearson, Karl, "On a coefficient of class heterogeneity or divergence," *Biometrika*, 5 (1906), 198–203.

[113] Pearson, Karl, "On the theory of contingency. I. Note on Professor J. Arthur Harris' paper on the limitation in the applicability of the contingency coefficient," *Journal of the American Statistical Association*, 25 (1930), 320–3.

[114] [Pearson, Karl?], "Remarks on Professor Steffensen's measure of contingency. Editorial," *Biometrika*, 26 (1934), 255–60.

[115] Peirce, C. S., "The numerical measure of the success of predictions" (letter to the editor), *Science*, 4 (1884), 453–4.

[116] Pietra, G., "The theory of statistical relations with special reference to cyclical series," *Metron,* 4 (3–4), (1924–5), 383–557.

[117] Pollaczek-Geiringer, H., "Bemerkungen zur Korrelations-theorie," *Verhandlungen des Internationalen Mathematiker-Congress,* Zurich (1932), 2, 229–30.

[118] Pollaczsk-Geiringer, H., "Korrelationmessung auf Grund der Summenfunktion," *Zeitschrift für Angewandte Mathematik und Mechanik,* 13 (1933), 121–4.

[119] Quetelet, M. A., *Letters addressed to H.R.H. the Grand Duke of Saxe Coburg and Gotha on the Theory of Probabilities as Applied to the Moral and Political Sciences* (translated from the French by Olinthus Gregory Downs), Charles and Edwin Layton, London, 1849.

[120] Reuning, H., "Evaluation of square contingency tables. A simple method of condensation for small samples," *Bulletin, National Institute for Personnel Research,* (South African Council for Scientific and Industrial Research, Johannesburg), 4 (1952), 160–7.

[121] Saile, Tivadar Antal, *Influence de Joseph de Körösy Sur l'Évolution de la Statistique,* Kiadja a Magyar Tudományos Akadémia, Budapest, 1927. Hungarian with French summary.

[122] Salvemini, Tommaso, "Su alcuni indici usati per la misura delle relazioni statistiche," *Statistica,* 7 (1947), 200–18.

[123] Salvemini, T., "Nuovi procedimenti di calcolo degli indici di dissomigliaza e di connessione," *Statistica,* 9 (1949), 3–27.

[124] Salvemini, Tommaso, "Su alcuni aspetti della dissomiglianza e della concordanza con applicazione alle distribuzioni degli sposi secondo l'età," *Bulletin de l'Institut International de Statistique,* 34 (2), (1954), 283–300.

[125] Salvemini, Tommaso, "Fondamenti razionali, schemi teorici e moderni aspetti delle relazioni tra due o più variabili," *Atti della XV Riunione Scientificà, Societa Italiana di Statistica,* 1955.

[126] Savorgnan, Franco, "La misura dell'endogamia e dell'omogamia," in *Proceedings of the International Congress for Studies on Population (Rome, 7th–10th September 1931–IX),* Vol. X, 9–37. Published by the Comitato Italiano per lo Studio dei Problemi della Popolazione, through the Istituto Poligrafico dello Stato, Rome, 1934.

[126a] Smith, H. Fairfield, "On comparing contingency tables," *The Philippine Statistician,* 6 (1957), 71–81.

[127] Steffensen, J. F., "Deux problèmes du calcul des probabilités," *Annales de l'Institut Henri Poincaré,* 3 (1932–3), 319–44. Plus one non-numbered page of errata.

[128] Steffensen, J. F., "On certain measures of dependence between statistical variables," *Biometrika,* 26 (1934), 250–5.

[129] Steffensen, J. F., "On the ω test of dependence between statistical variables," *Skandinavisk Aktuarietidskrift,* 24 (1941), 13–33.

[130] Striefler, Heinrich, "Zur Methode der Rangkorrelation nach Tönnies," *Deutsches Statistisches Zentralblatt,* 23 (1931), 128–35 and 160–7.

[131] "Student" [W. S. Gosset], "An experimental determination of the probable error of Dr Spearman's correlation coefficients," *Biometrika,* 13 (1920–1), 263–82. Reprinted in: Pearson, E. S., and Wishart, John (ed.), *"Student's" Collected Papers,* Biometrika Office, University College, London, 1942. Pp. 70–89.

[132] Thionet, Pierre, "L'école moderne de statisticiens italiens," *Journal de la Société de Statistique de Paris,* 86 (1945), 245–55, and 87 (1946), 16–34.

[133] Thionet, Pierre, "Résponse de l'auteur de la communication du 21 novembre 1945," *Journal de la Société de Statistique de Paris,* 88 (1947), 203–8.

[134] Thirring, Gustave, obituary of Joseph de Körösy, *Bulletin de l'Institut International de Statistique,* 16 (1907–8), 150–4.

[135] Thompson, J. C., "On the operational deficiencies in categorical weather forecasts," *Bulletin of the American Meteorological Society,* 33 (1952), 223–6.

[136] Thompson, J. C., and Brier, G. W., "The economic utility of weather forecasts," *Monthly Weather Review,* 83 (1955), 249–54.

[137] Tönnies, Ferdinand, "Eine neue Methode der Vergleichung Statistischer Reihen (im Anschluss an Mitteilungen über kriminalstatistische Forschungen)," *Jahrbuch für Gesetzgebung, Verwaltung und Volkswirtschaft im Deutschen Reich*, 33 (1909), 699–720.

[138] Tönnies, Ferdinand, "Korrelation der Parteien in Statistik der Kieler Reichstagswahlen," *Jahrbücher für Nationalökonomie und Statistik*, 122 (1924), 663–72.

[139] Tylor, Edward B., "On a method of investigating the development of institutions; applied to laws of marriage and descent," *Journal of the Anthropological Institute of Great Britain and Ireland*, 18 (1888–9), 245–69. Discussion, 270–2.

[140] van der Bijl, Willem, "Statistische Bemerkungen zu H. Müller-Annen: 'Versuche von Langfristvorhersagen mit einer Kontingenzmethode' (Ann. Met. 6, S. 257) nebst weiteren Betrachtungen über den prognostischen Wert von Kontingenztabellen," *Annalen der Meteorologie*, 7 (1955–6), 289–302.

[141] Wahl, Eberhard W., "Das statistisch Entropieverhältnis, ein Hilfsmittel zur Lösung von Vorhersageproblemen," *Meteorologische Rundschau*, 8 (1955), 51–4.

[142] Wallis, W. Allen, and Roberts, Harry V., *Statistics, A New Approach*, The Free Press, Glencoe, Illinois, 1956.

[143] Wallis, Wilson D., "Probability and the diffusion of culture traits," *American Anthropologist*, 30 (1928), 94–106.

[144] Weida, Frank M., "On various conceptions of correlation," *Annals of Mathematics*, 29 (1928), 276–312.

[145] Williams, Josephine J., "Another commentary on so-called segregation indices," *American Sociological Review*, 13 (1948), 298–303.

[146] Wilson, Edwin B., "Morbidity and the association of morbid conditions," *The Journal of Preventive Medicine*, 4 (1930), 27–38.

[147] Wirth, Wilhelm, *Spezielle Psychophysische Massmethoden*, Urban and Schwarzenberg, Berlin and Vienna, 1920. Part of Abderhalden, Emil (ed.), *Handbuch der Biologischen Arbeitsmethoden;* Lieferung 4, Abt. 6, Methoden der Experimentellen Psychologie, A. Heft 1.

[148] Wood, Karl D., "Rapid correlation by an empirical method," *Journal of Educational Psychology*, 19 (1928), 243–51.

[149] Yule, G. Udny, "On the association of attributes in statistics: with illustrations from the material from the Childhood Society, &c," *Philosophical Transactions of the Royal Society of London, Ser. A*, 194 (1900), 257–319.

[150] Yule, G. Udny, "On the methods of measuring association between two attributes," *Journal of the Royal Statistical Society*, 75 (1912), 579–642. Discussion 643–52