Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Report No. PR-65-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.



Reports from the Research Laboratories
of the
Department of Psychiatry
University of Minnesota

# Detecting Latent Clinical Taxa by Fallible Quantitative Indicators Lacking an Accepted Criterion[1]

**Paul E. Meehl**

Report No. PR-65-2 May 25, 1965

NB(1981): Sections 4-5-6 and Appendix are now obsolete, being replaced by the procedure in Meehl: *Pscyhodiagnosis* (1973) Chapter 12 "MAXCOV-HITMAX" where the hitmax interval covariance test (pp. 28-29 of this tech report) is elevated from the role of a consistency test to that of a main estimator. See also Meehl and Golden "Taxometric methods" in P.C. Kendall and J. N. Butcher (eds) *Handbook of Research Methods in Clinical Psychology* (1982) and references cited therein.

**Pagination in this posted version does not match original publication.**

[1] This research was done in connection with a long-term program on "The Skilled Clinician's Assessment of Personality," supported by grants from the Ford Foundation and the National Institute of Mental Health, United States Public Health Service (Research Grant M-4465)

CONTENTS

**Detecting Latent Clinical Taxa by**
**Fallible Quantitative Indicators Lacking an Accepted Criterion**
**by**
**Paul E. Meehl**

## 1. The Problem

In the early stages of investigating a subject-matter domain, it often happens that we have plausible reasons for supposing that the entities under study belong to two or more qualitatively distinct classes or categories, although we have a very imperfect understanding as to the intrinsic nature of the categories, and consequently do not possess theoretically powerful means for sorting the individuals into the putative categories on the basis of their presently observable characteristics. For example, the first published account of a new disease entity in medicine is often little more than a kind of syndrome description, in which the physician calls attention to a set of signs or symptoms which have a tendency to occur jointly, and (relying upon the history of medicine) he postulates the existence of an underlying pathological condition—a new disease—to account for the observed clustering of clinical signs. It is misleading to call such a syndrome an "operational definition" of the disease entity, since the usual situation is one in which the syndrome consists merely of a statistical cluster rather than an invariable compresence; and the clinical investigator, if pressed, will be reluctant to stipulate any one sign or symptom, or any specified combination of signs or symptoms, as the necessary and sufficient condition for an individual patient to "have the disease." This refusal, while irksome to most psychologists, reflects a clinical wisdom on the part of the physician, because it is well known in medicine that very few symptoms or signs are truly pathognomonic. And, on the other hand, many pathological processes are, at least at certain stages of their development, not attended by clinical indications. (See e.g., Peppard 1949). Since the diseased state exists in individual organisms, and the clinical signs are fallible, it follows that the linkage between the syndrome S and the diseased state D is, from a logical and mathematical standpoint, only a probabilistic linkage, $S \overset{P}{\longleftrightarrow} D$, as applied to individuals. But this means, of course, that the conjunctions $(\sim S \cdot D)$ and $(S \cdot \sim D)$ are possible (and actual) states of nature. Hence, as a matter of formal logic, we see that S cannot be considered a definition of D.

Situations of this sort are, of course, not confined to medicine. The tentative postulation of a "type" in personality theory, or the tentative identification of a "species" in biology, or of an "ideology" in political life, while they deal with very different sorts of phenomena and kinds of causal explanation, do nevertheless share important logical and mathematical features. While the existence and scientific utility of so-called open concepts has been generally recognized by logicians and methodologists of science for some years (see, e.g., Pap 1953) methods of dealing with such open concepts, and especially methods for tightening them up a bit, have not received as much attention either in the social or biological sciences as they probably deserve in terms of their ubiquity and importance in these fields. See Cronbach and Meehl (1955), Meehl (1959) and methodological references cited therein.

I shall not enter here into the difficult question of precisely what one means by the expression "qualitatively distinct classes," a phrase which, I believe, has several distinct alternative explications, each possessing methodological power in appropriate contexts. (For an insightful analysis of the meanings of "type," see Cattell 1957.) Since the words "class" or "category" have what might be called a rather weak logical meaning in general, namely, any set whose membership can be defined by reference to the individuals' possession of a certain property, or conjunction or disjunction of properties; for present purposes I prefer to employ a term with somewhat stronger connotations, such that it becomes an empirical question whether or not a certain class exists. Roughly speaking, the empirical existence of a class of the kind exemplified above is, ultimately, a matter of theoretical understanding (e.g., that the individuals belonging to the contrasted classes differ in regard to a dichotomous, "present-or-absent" etiology, such as a germ or a gene); and, associated with this dichotomous etiology, there is a non-chance patterning of the signs or symptoms such as would not be automatically forthcoming if one were to define a category by arbitrary stipulation of statistically unrelated symptomatic characteristics. For designating the non-arbitrary kind of empirical class in which we aim to "carve nature at its joints," I shall employ the term taxon, with apologies to the zoologists. As a first approximation, we may think of a population being dichotomously taxonomic with respect to a set of measurable characteristics if there is a marked statistical clustering of individuals about two points in the space defined by the measurable characteristics. But the chief specification of the meaning of 'taxon' for present purposes is contained implicitly in the

mathematical development to follow. I.e., a taxon in the present sense is a class (of individuals) which behaves approximately as the model supposes.

Since none of the observable signs or symptoms has a clearly privileged status, in the early stages of research in a taxonomic field it is not fruitful to stipulate any one of them, or any conjunction of them, as "the criterion." For this reason the traditional terminology of 'predictor' and 'criterion' is misleading, and a terminology which does more justice to the methodological realities is to refer to each sign or symptom in the alleged cluster as an indicator or indicator-variable (See Meehl 1959). The indicators are probabilistic predictors of one another, singly and jointly; but, what is more important from the standpoints of theory and long-term research strategy, they are purportedly—singly and jointly—also probabilistic indicators of the postulated state, condition, or etiological factor which gives rise to their syndromic patterning. Since we do not know by any acceptable (stipulative) criterion which individuals belong to which taxon, and since in the very early stages of research we have only the crudest theory-sketch as to the theoretical nature of the taxonomic condition which gives rise to these indicators, it is appropriate to refer to such postulated taxa as latent. The reader who is operationally puristic need not be offended by this terminology, since he can reformulate it (without reference to inferred hypothetical causal entities) in terms of a problem of sorting individuals—who are after all observables!—into classes so as to achieve certain desired statistical relationships among the observable indicators.

Suppose, for example, clinical experience leads us to hypothesize that a population of psychiatric outpatients consists of two latent taxa which we shall call schizotypes and nonschizotypes (Meehl 1962, 1964). It is not, of course, supposed that the nonschizotypic population is homogeneous, or that a further taxonomy could not be discerned by the use of appropriate indicator-clusters within this group or, for that matter, even within the schizotypic group. There are as many classes identifiable within any taxon as there are properties and conjunctions of properties. And there are as many taxa as there are property-sets which, once specified, will reveal the right kind of statistical covariation. This fact will be set aside for the moment. Since we do not have a received theory of schizotypy, our basis for postulating such a diagnostic taxonomy must lie in observed relationships among the indicators. But as yet we do not know which indicators have the greater validity for discriminating the latent taxa; we may not be in a position to rank order the indicators as to their validity; and, worst of all, we may not

have an accurate estimate, or even professional consensus, as to the relative frequency of schizotypes and nonschizotypes in our clinical population, i.e., we do not know the latent base-rate P of schizotypy (Meehl and Rosen 1955). In short, we start with a small amount of clinical knowledge and a vast amount of clinical and theoretical ignorance. The problem is to corroborate or discorroborate the theory that such a latent taxonomy exists, to arrive at improved estimates of the unknown base rate P, and to make some sort of assessment as to the construct validities of the purported indicators.

The psychologist in reading the development of disease concepts in organic medicine is struck by a paradox. It seems that physicians over the course of time change their views as to the diagnostic validity of certain signs or symptoms, prior to reaching that stage of research in which identification of pathology or etiology provides an acceptable "criterion" against which the facets of the syndrome can be "validated." It seems strange that such a revision of indicator-weights occurs, but unquestionably it does. Thus a physician writing on the basis of his clinical experience with "Fisbee's syndrome," may state that his experience to date shows Fisbee to have erred in considering "Fisbee's sign" as pathognomonic, whether as an inclusion or as an exclusion test. It is evident that prior to elucidating the pathology and etiology of whatever disease entity underlies the majority of cases currently being diagnosed as Fisbee's syndrome, the only basis upon which such a claim (or denial) of pathognomicity for a given sign could be supported would be facts about its relationships to other signs belonging to the cluster. Of course the psychologist does have a statistical analog to this kind of cumulative clinical experience in such procedures as factor analysis, where (a) a set of indicators is used to provide a partial specification of meaning for those open concepts called 'factors,' and (b) the factor loading of an indicator, while treated until further notice as an estimate of the indicator's construct validity for the inferred latent variable, is in fact computed, ultimately, in terms of the relationships between each indicator and the others together with the interrelationships among the others.

## 2. The Idealized Taxonomic Model

We assume as an admitted idealization, not fully attainable in practice, that there exists a set of quantitative indicator-variables x, y, z, v, **…,** each of which has considerable construct validity in the sense of discriminating between the two latent taxa but which are independent of one another within each taxon. We assume further that the frequency distribution for each indicator is

unimodal within taxa. We do not make further assumptions regarding the distributions, such as normality, symmetry, homogeneity of variance, or the like.

The density-function within each taxon, i.e., the ordinate of the frequency-function fixing each taxon's distribution area = 1, will be designated by the letter "$\phi$" with subscripts "s" or "n" indicating that the function is that for schizotypes or nonschizotypes, respectively. Thus we have,

$$\int_{-\infty}^{\infty} \phi_s(x)dx = 1$$

$$\int_{-\infty}^{\infty} \phi_n(x)dx = 1$$

I shall use "density-function" for the $\phi$-functions, and "frequency-function" for the functions f(x) obtained when the density-functions are multiplied by their appropriate base-frequencies (i.e., by the <u>number</u>—not rate!—of schizotypes and nonschizotypes in the population under study). That is, the (<u>absolute</u>) <u>frequency-function</u> (i.e., the ordinate of each taxon's distribution when the two areas are set equal to their base-frequencies in the population) will be indicated by the English letter <u>f</u>, again with the appropriate subscripts indicating which latent taxon is involved. Then the (absolute) ordinates of the two frequency-functions are given by the products of the respective (within-taxon) density-functions and the associated base-frequencies as follows:

$f_s(x) = N_s \phi_s(x)$ $\quad$ $N_s$ = Number of schizotypes

$F_n(x) = N_n \phi_n(x)$ $\quad$ $N_n$ = Number of nonschizotypes

$$\int_{-\infty}^{\infty} f_s(x)dx = N_s$$

$$\int_{-\infty}^{\infty} f_n(x)dx = N_n$$

And the base-<u>rates</u> are

$$P = \frac{N_s}{N} = 1 - Q = 1 - \frac{N_n}{N}$$

We shall have no occasion to use a density-function for the total (mixed) population.

Suppose we locate a particular cut $x = x_c$ on an indicator-variable, and classify all patients falling above that cut as schizotypic and those falling below the cut as nonschizotypic. We

thereby define a latent group of "hits" and "misses" (i.e., Omniscient Jones knows what hits and misses are yielded by this cut $x_c$, although we don't know the latent situation yet), as follows:

$H_s(x_c) = \int_{x_c}^{\infty} f_s(x)dx$ — Total hits (= absolute frequency of schizotypes) above cut $x_c$

$M_s(x_c) = \int_{-\infty}^{x_c} f_s(x)dx$ — Total misses (= absolute frequency of schizotypes) below cut $x_c$

$H_n(x_c) = \int_{-\infty}^{x_c} f_n(x)dx$ — Total hits (= absolute frequency of nonschizotypes) below cut $x_c$

$M_n(x_c) = \int_{x_c}^{\infty} f_n(x)dx$ — Total misses (= absolute frequency of nonschizotypes) above cut $x_c$

The probability within a taxon of being correctly [or incorrectly] classified by a given cut on the indicator-variable will be referred to throughout as the 'valid [or false] rate', positive or negative as the case may be. We can write these rates as functions of x [= $x_c$] as follows:

Valid positive rate = $p_s(x) = \int_{x}^{\infty} \phi_s(x)dx$

False negative rate = $q_s(x) = 1 - p_s(x) = \int_{-\infty}^{x} \phi_s(x)dx$

Valid negative rate = $q_n(x) = \int_{-\infty}^{x} \phi_n(x)dx$

False positive rate = $p_n(x) = 1 - q_n(x) = \int_{x}^{\infty} \phi_n(x)dx$

It is important to keep in mind that these rates, when viewed as fractions, are based upon denominators equal to the true taxon frequencies $N_s$ and $N_n$, rather than denominators equal to above-cut and below-cut frequencies $N_a$ and $N_b$.

I shall assume that each indicator is "everywhere valid," in the (weak) sense that the valid positive rate exceeds the false positive rate at any cut which divides the population. That is, we assume:

$P_s(x) > p_n(x)$ for all x within range

$P_s(y) > p_n(y)$ for all y within range

$P_s(z) > p_n(z)$ for all z within range

which of course implies that the valid negative rate everywhere exceeds the false negative rate,

$q_s(x) < q_n(x)$

$q_s(y) < q_n(y)$

$q_s(x) < q_n(z)$

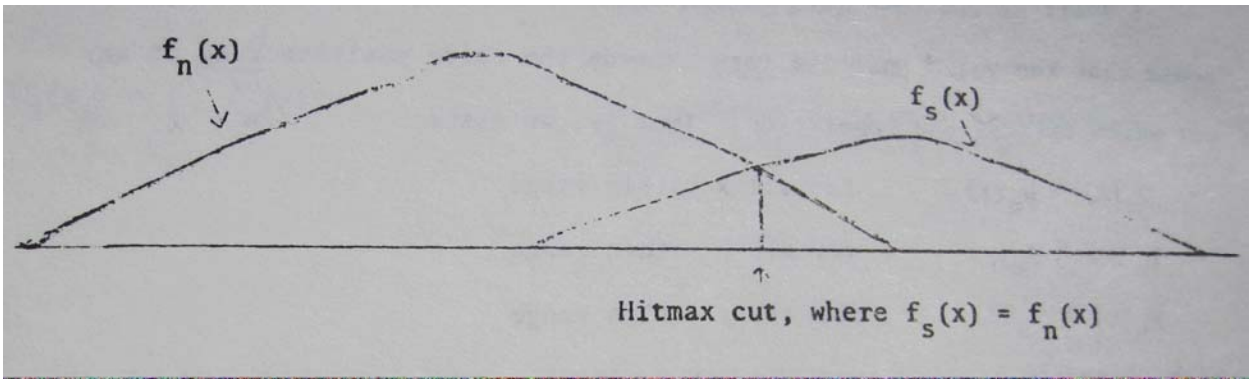
Actually this assumption is only used once, non-crucially, and could be dropped. The increased variability and kurtosis of schizotypic patients on many variables reported in the literature makes the "everywhere valid" postulate less safe than it would be in most research areas, and it is quite possible that some indicators may be "valid backward" in selected regions of the distribution.

It will be convenient at times to simplify the notation by writing the indicator-variable merely as a subscript instead of in functional form, thus,

$H_{sx} = H_s(x_c) = N_s p_s(x) = NPp_s(x)$

$P_s(x) = p_{sx} = 1 - q_{sx}$

A cut of particular interest is that cut on an indicator-variable which minimizes the misclassifications. This cut is easily seen from geometry, or by differentiating the expression for total hits, to be the abscissa value at which the absolute ordinates are equal, i.e., the position on the distribution above which the two latent absolute frequency functions intersect:




To minimize total misclassifications, we set

$M_s'(x) + M_n'(x) = 0$

and

$M_s''(x) + M_n''(x) > 0$

which occurs at

(1) $f_s(x) = f_n(x)$

We shall refer to this position on each indicator as the hitmax cut for the indicator in question. Subscript "h" on a variable will denote the hitmax cut.

## 3. Locating the Hitmax Cut

We first show that the hitmax cut on one indicator, say w, can be located by finding the w-interval within which the covariance of two other indicators, say x and y, is maximal.

Assuming that the correlation between a pair of indicators is zero within taxa, any observed correlation (or covariance) between two indicators in a given sub-sample is attributable solely to the "mixture" of schizotypes and nonschizotypes in that sub-sample. Intuitively, it seems obvious that the more "mixed" a sub-sample is, the larger should be the correlation between an indicator-pair. In a "pure" sample, i.e., one in which all of the individuals are schizotypes, or in which none of the individuals are schizotypes, the xy-covariance will be zero. Symmetry suggests that if we arranged sub-samples in order from a sub-sample with no schizotypes (at the low end of the w-distribution) to a sample composed solely of schizotypes (at the high end) the covariance of x and y would rise from zero to some larger value and then decline to zero again, with the highest value of the xy-covariance occurring in that sample which was "most mixed," that is, in which the incidence of schizotypes and nonschizotypes is equal. Hence, we can locate the hitmax interval on indicator w by computing the xy-covariance for the sub-samples of patients lying within each successive w-interval, plotting this xy-covariance as a function of the w-value, and locating this graph's maximum. Because the hitmax cut on w is the value of w for which the absolute ordinates $f_s(w)$ and $f_n(w)$ of the schizotypic and nonschizotypic frequency-functions are equal, we see that taking an interval sufficiently small surrounding the hitmax ordinates will be equivalent to choosing a sub-sample within which the schizotypic and nonschizotypic cases are approximately equal in absolute frequency.

### Proof

Since covariance is mean cross-product of deviations, the size $N_i$ of any subsample has no influence on the maximizing procedure.

We want to choose the w-interval within which

$$\frac{\sum^{N_i} xy}{N_i} - \bar{x}_i \bar{y}_i = \text{Max}$$

Let s = (Absolute) schizotype frequency in sub-sample $\underline{i}$

n = (Absolute) nonschizotype frequency in sub-sample $\underline{i}$

Then

$p_i = \frac{s}{N_i}$ = Relative frequency of schizotypes in the sub-sample lying within interval $\underline{w}_i$.

$q_i = 1 - p_i = \frac{n}{N_i}$ = Relative frequency of nonschizotypes in the sub-sample lying within interval $\underline{w}_i$.

Since $r_{xy} = 0$ within taxa, the mean cross-product within a taxon equals the product of the taxon's means.

$\overline{(xy)}_s = \bar{x}_s \bar{y}_s$ among schizotypes

$\overline{(xy)}_n = \bar{x}_n \bar{y}_n$ among nonschizotypes

The mean cross-product for the mixed sample is therefore

$$\frac{\sum^{N_i} xy}{N_i} = \frac{\sum^{s} xy + \sum^{n} xy}{N_i} = \frac{s\overline{(xy)}_s + n\overline{(xy)}_n}{N_i} = \frac{s\bar{x}_s\bar{y}_s + n\bar{x}_n\bar{y}_n}{N_i} = p_i\bar{x}_s\bar{y}_s + q_i\bar{x}_n\bar{y}_n$$

The product of the means in the mixed sample is

$$\bar{x}\bar{y} = \left(\frac{s\bar{x}_s + n\bar{x}_n}{N_i}\right)\left(\frac{s\bar{y}_s + n\bar{y}_n}{N_i}\right) = \left(p_i\bar{x}_s + q_i\bar{x}_n\right)\left(p_i\bar{y}_s + q_i\bar{y}_n\right)$$

Then the mixed-sample covariance is

$$\begin{aligned} \text{cov}(xy) &= \left(p\bar{x}_s\bar{y}_s + q\bar{x}_n\bar{y}_n\right) - \left(p\bar{x}_s + q\bar{x}_n\right)\left(p\bar{y}_s + q\bar{y}_n\right) \\ &= pq\left(\bar{x}_s - \bar{x}_n\right)\left(\bar{y}_s - \bar{y}_n\right) \end{aligned}$$

(2) $\text{cov}(xy) = pq\,\Delta\bar{x}\,\Delta\bar{y}$

That is, the covariance of two indicators in a mixed sample is the product of the sample base-rates times the product of the differences of the latent means. These latter being fixed, the mixed-sample covariance depends only upon the amount of "mixture,"

(3) $\text{cov}(xy) = Kpq$

which is maximal when $p = q = 1/2$ as our intuitions indicated.

Case of non-quantitative signs.

It is worth mentioning that this method can be used in a pinch, provided we have three indicators available even when only one of them is quantitative, the other two being clinical signs scored only in dichotomous form (e.g., "ever married?"). It is more illuminating in the proof, however, to treat these two signs as determined by sliding cuts; then we show that even if the cuts were non-optimally chosen, hitmax on the quantitative indicator can nevertheless be located. The sample must be large enough so that the hitmax interval contains sufficient cases to provide a stable phi-coefficient numerator, and that several adjacent w-intervals yield stable phi-numerators, enabling us to discern a clear maximum in the hitmax interval.

The input indicator w is dealt with quantitatively; the two output indicators x and y, assumed for the moment to be quantitative, are cut at various arbitrary points to yield dichotomous "signs" $x^+$ and $y^+$. We choose an arbitrary w-interval, and consider the sub-sample of cases lying within it. We then choose arbitrary pairs of cuts on x and y, and for each cut-pair we compute the phi-numerator $(p_{xy} - p_x p_y)$ on the (xy)-4-fold table determined by these cuts. We find the maximum of these (xy)-cut-pairs. We repeat this process within all w-intervals. Then the w-interval whose $(p_{xy} - p_x p_y)$-maximum is a maximum is the interval containing the hitmax cut on w.

Proof

Let $p_{xy}$ = Probability of cases falling in Cell I of the (xy)-4-fold table, in any sub-sample.

$p_x$ = Marginal probability (cases > x-cut) in sub-sample

$p_y$ = Marginal probability (cases > y-cut) in sub-sample

The latent quantities are

$p_s$ = Schizotypic base-rate in a sub-sample

$p_n = 1 - p_s$ = Nonschizotypic base-rate in a sub-sample

$p_{sx}$ = Valid positive rate determined by an x-cut

$p_{sy}$ = Valid positive rate determined by a y-cut

$p_{nx}$ = False positive rate determined by an x-cut

$p_{ny}$ = False positive rate determined by a y-cut

Given any sub-sample chosen without reference to $\underline{x}$ or $\underline{y}$, in which the sub-sample base-rates are $p_s$ and $p_n$, the observed statistics can be written in terms of these latent quantities thus:

$$p_{xy} = p_s\ p_{sx}\ p_{sy} + p_n\ p_{nx}\ p_{ny}$$

$$= p_s\ \ p_{sx}\ \ p_{sy} + q_s\ p_{nx}\ p_{ny}$$

$$p_x\ = p_s\ p_{sx} + q_s\ p_{nx}\ p_y$$

$$p_y\ = p_s\ p_{sy} + q_s\ p_{ny}$$

Then the phi-coefficient numerator on any sub-sample, say, one lying within the $\underline{w}$-interval $\underline{i}$, is

$$(p_{xy} - p_x p_y)w_i = (p_{sw_i} p_{sx} p_{sy} + q_{sw_i} p_{nx} p_{ny})$$

$$- (p_{sw_i} p_{sx} + q_{sw_i} p_{nx})(p_{sw_i} p_{sy} + q_{sw_i} p_{ny})$$

$$= (p_{sw_i} q_{sw_i})(p_{sy} - p_{nx})(p_{sy} - p_{ny}) \qquad (4)$$

[This expression is, of course, the dichotomous form of equation (2).]

The indicators being everywhere valid, all three factors of (4) are > 0, and since they are functions of $\underline{w}$, $\underline{x}$, and $\underline{y}$ only, respectively, this quantity is maximized when each factor is maximized. So we want

(5) Max $(p_{sx} - p_{nx})$

(6) Max $(p_{sy} - p_{ny})$

(7) Max $(p_{sw_i} q_{sw_i})$

Conditions (5) and (6) are satisfied when the separate sign-"validities" are greatest, in the special sense of having the maximum difference between the sign-positive <u>rate</u> among schizotypes and that among nonschizotypes. Condition (7), a function of the $\underline{w}$-interval chosen, is met when $p_{swi} = q_{swi} = 1/2$. But this is in the interval surrounding the hitmax cut $\underline{w}_h$, since it is the interval in which the absolute ordinates $f_s(w) = f_n(w)$.

But we have maximized more than we need to, if our aim is merely the locating of $\underline{w}_h$. Equation (4) shows that even if $\underline{x}_c$ and $\underline{y}_c$ were non-optimally chosen, the phi-numerator $(p_{xy} - p_x p_y)$ will always be maximal in a sample for which $p_{sw} = q_{sw}$. Hence if we are forced to begin with already ("observationally") dichotomized signs $x^+$ and $y^+$, we can still use their fourfold-table statistics, as computed within each of the $\underline{w}$-intervals, to locate $\underline{w}_h$.

This possibility is important because in the early stages of taxonomic research we may have quite persuasive reasons for tentatively regarding a certain indicator as highly valid and negligibly correlated with others intra-taxon, and yet we may possess no psychometric basis for expressing it numerically.

## [**]4. Locating a Cut Which Equates the Valid Positive and Valid Negative Rates

In the preceding section we showed that the xy-covariance of any sub-sample is given by $Kp_sq_s = Kp_sp_n$, where K = product of latent mean differences. We now use this fact to locate a cut different from hitmax, namely, that cut which equates the valid positive rate $p_s$ and the valid negative rate $q_n$. It turns out that this latent condition is fulfilled by a cut so chosen that the product of the covariance of another indicator-pair, calculated upon all cases lying above the cut, by the square of the frequency above the cut, is equal to the corresponding product for all of the cases lying below the cut. We consider a sliding cut on input indicator y and its influence upon the covariance of output indicators u and v.

Proof

The (uv)-covariances of the two sub-samples falling above and below any arbitrary y-cut are expressible in latent terms thus, from Equation (3), where H and M are hits and misses, respectively:

$$\mathrm{cov}_{ay}(uv) = Kp_{say}q_{say} = K\frac{H_{sy}}{N_{ay}}\frac{M_{ny}}{N_{ay}}$$

$$\mathrm{cov}_{by}(uv) = Kp_{sby}q_{sby} = K\frac{M_{sy}}{N_{by}}\frac{H_{ny}}{N_{by}}$$

We impose the observable condition that y be cut so that

(8) $$N_{ay}^2\,\mathrm{cov}_{ay}(uv) = N_{by}^2\,\mathrm{cov}_{by}(uv)$$

[**]This section superseded. See note on title page.

which in latent terms says that

$$N_{ay}^2 K \frac{H_{sy}}{N_{ay}} \frac{M_{ny}}{N_{ay}} = N_{by}^2 K \frac{M_{sy}}{N_{by}} \frac{H_{ny}}{N_{by}}$$

$$H_{sy}\, M_{ny} = M_{sy}\, H_{ny}$$

Dividing by product of total base-frequencies, a constant over all cuts,

$$\frac{H_{sy} M_{ny}}{N_s N_n} = \frac{M_{sy} H_{ny}}{N_s N_n}$$

or, in terms of latent rates,

(9) $\quad p_{sy}\, p_{ny} = q_{sy}\, q_{ny}$

i.e., product of valid and false positive rates equals product of valid and false negative rates. Hence

$$p_{sy} = q_{ny}$$

and $\quad p_{ny} = q_{sy}$

So that what we do computationally is to start at one end of the y-distribution, locate an arbitrary y-cut, calculate the (uv)-covariance for the $N_{ay}$ cases lying above that cut, multiply this covariance by $N_{ay}^2$, and do the same for the $N_{by}$ cases that fall below this arbitrary y-cut. We then move up to the next y-interval and recalculate these quantities. We continue this process and find—if the latent model is approximately fulfilled—that the absolute value of the difference between $N_{ay}^2$ cov(uv) and $N_{by}^2$ cov(uv) declines to zero at some intermediate y-cut and then rises again from zero as the y-cut is moved along. That y-cut which approximately equates the two products is the cut which equates the latent hit-rates on y and, of course, therefore also equates the latent miss-rates on y.

## An Alternative Method for Locating the Hitmax Cut

Determining a cut which equates the valid positive and valid negative rates on one indicator makes it possible to locate the hitmax cut on a second indicator by a method alternative to that of Section 3. Suppose a cut on x has been located by the method of the present section so as to equate the valid positive rate $p_{sx}$ with the valid negative rate $q_{nx}$. Holding this x-cut fixed, we generate a series of 4-fold tables by choosing an arbitrary y-cut, computing the concordance (defined as the proportion of cases in the first and third quadrants, i.e., the proportion of tallies in the (+ +) and (– –) cells, where the "signs" defined by the cuts are "concordant"), and we plot

this concordance measure as a function of the sliding y-cut. The y-value which produces a maximum on this graph is the hitmax cut on y.

Proof

We have initially fixed a cut $x_c$ such that the latent hit-rates are equal for schizotypes and nonschizotypes, that is, we know that for this cut

| | | |
|---|---|---|
| $p_{sx} = q_{nx}$ | Valid rates equal | From Equation (9) |
| $p_{nx} = q_{sx}$ | False rates equal | |

The "concordance" C(x,y) is defined as the sum of probabilities in Cells I and III of the 4-fold table determined by any pair of cuts, and is expressible in latent terms as

$$C(x,y) = \underbrace{P\, p_s(x)\, p_s(y) + Q\, p_n(x)\, p_n(y)}_{\text{Cell I}} + \underbrace{P\, q_s(x)\, q_s(y) + Q\, q_n(x)\, q_n(y)}_{\text{Cell III}}$$

To simplify notation write latent rates on x, being fixed throughout, with subscripts instead of as functions of x. Also multiplying by N, we use absolute frequencies instead of base-rates. Then we want to maximize a function of y alone,

$$C(y) = p_{sx}\, N_s\, p_s(y) + p_{nx}\, N_n\, p_n(y) + q_{sx}\, N_s\, q_s(y) + q_{nx}\, N_n\, q_n(y)$$

Differentiating with respect to y and setting derivative at zero,

$$\frac{dc}{dy} = p_{sx} N_s p_s'(y) + p_{nx} N_n p_n'(y) + q_{sx} N_s q_s'(y) + q_{nx} N_n q_n'(y) = 0$$

which can be written in terms of the intra-taxon ordinates [=density-functions] as

$$\frac{dc}{dy} = p_{sx} N_s[-\phi_s(y)] + p_{nx} N_n[-\phi_n(y)] + q_{sx} N_s[\phi_s(y)] + q_{nx} N_n[\phi_n(y)] = 0$$

But since the absolute ordinates are given by the intra-taxon ordinates times the base-frequencies, this says that

$$\frac{dc}{dy} = -p_{sx} f_s(y) - p_{nx} f_n(y) + q_{sx} f_s(y) + q_{nx} f_n(y) = 0$$

$$f_n(y)[q_{nx} - p_{nx}] = f_s(y)[p_{sx} - q_{sx}]$$

(10)
$$\frac{f_n(y)}{f_s(y)} = \frac{p_{sx} - q_{sx}}{q_{nx} - p_{nx}} = 1 \text{ since } p_{sx} = q_{nx}$$

Hence

$$f_n(y) = f_s(y)$$

which occurs at the hitmax cut on y, as shown in Section 2, Equation (1).

## **5. Estimating the Unknown Latent Base-Rate P

Suppose that cuts have been located on two indicators x and y by the method in Section 4 such that the latent valid rates $p_s(x) = q_n(x)$ and $p_s(y) = q_n(y)$. Thus if, say, four indicator-variables x, y, u, v, are available to us, we may have chosen a cut on x and a cut on y utilizing cov(uv) as the output measure. The resulting cuts $x_c$ and $y_c$ determine a 4-fold table, and it is intuitively obvious that the distribution of case-tallies over the cells of this table will now depend upon the unknown base-rate P.

The observed probabilities in the 4-fold table determined by the x and y cuts for equated latent valid rates can be expressed in terms of the latent variables as follows:

| | |
|---|---|
| $p_I = P\, p_{sx}\, p_{sy} + Q\, p_{nx}\, p_{ny}$ | Observed probability in Cell I (+ +) |
| $p_{II} = P\, q_{sx}\, p_{sy} + Q\, q_{nx}\, p_{ny}$ | Observed probability in Cell II (– +) |
| $p_{III} = P\, q_{sx}\, q_{sy} + Q\, q_{nx}\, q_{ny}$ | Observed probability in Cell III (– –) |
| $p_{IV} = P\, p_{sx}\, q_{sy} + Q\, p_{nx}\, q_{ny}$ | Observed probability in Cell IV (+ –) |

of which only three are freely assignable. Take, say, Cells I-III. The latent values on the right side are all expressible in terms of the two valid rates and the base-rate, thus:

| | | |
|---|---|---|
| $p_I$ | $= P\, p_{sx}\, p_{sy} + (1 - P)\,(1 - p_{sx})\,(1 - p_{sy})$ | 3 unknowns |
| $p_{II}$ | $= P\,(1 - p_{sx})\, p_{sy} + (1 - P)\, p_{sx}\,(1 - p_{sy})$ | No new unknowns |
| $p_{III}$ | $= P\,(1 - p_{sx})\,(1 - p_{sy}) + (1 - P)\, p_{sx}\, p_{sy}$ | No new unknowns |

**This section superseded. See note on title page.

This appears to be a somewhat discouraging system of three cubics, but it is not, because the third-power terms vanish. The resulting system of quadratics (only cross-products responsible for the degree) and its solution is easier with a simplified notation.

Let $P = v$

$p_{sx} = x$

$p_{sy} = y$

Then we write, for Cell I,

$$v\,x\,y + (1 - v)(1 - x)(1 - y) = p_I$$

And for Cell II,

$$v(1 - x)\,y + (1 - v)\,x\,(1 - y) = p_{II}$$

And for Cell III,

$$v(1 - x)(1 - y) + (1 - v)\,x\,y = p_{III}$$

So the system we have to solve is

$$(11) \quad v + x + y - vx - vy - xy = 1 - p_I$$

$$(12) \quad x - vx + vy - xy = p_{II}$$

$$(13) \quad v - vx - vy + xy = p_{III}$$

The solution of this system (see Appendix) gives us the latent valid rates and, more importantly, the unknown base-rate P.

## [**]6. Estimating the Latent Means

Combining the information yielded by the procedure in Section 3 (locating the hitmax cut on an indicator) with the results of Section 5 (estimating the base-rate of schizotypy), we are in a position to solve for the latent means on any other indicator. The hitmax cut on an indicator w being a point $w_h$ at which the absolute ordinates, $f_s(w) = f_n(w)$, by considering the w-interval which contains $w_h$ (the "hitmax interval" on w), we will be dealing with a sub-sample of cases defined by their being found within the w-interval where the frequency is evenly split between schizotypes and nonschizotypes. Since the indicators are independent except by virtue of the latent taxa, it follows that the mean value of some other indicator, say x, calculated upon the

---

[**]This section superseded. See note on title page.

sub-sample of patients falling within the hitmax interval on w, depends solely upon the relative proportions of schizotypes and nonschizotypes in that sub-sample. That is, this average value of x is not influenced by the fact that the cases all have w-values = $w_h$, except insofar as this fact determines that the proportion of schizotypes and nonschizotypes is in the ratio of the f(w) ordinates at $w_h$. For the sub-sample of patients lying in the hitmax interval on w, therefore, the observed mean of the x-indicator will lie midway between the two unknown latent means $\bar{x}_s$ and $\bar{x}_n$.

The grand mean of x is, of course, merely the weighted mean of the latent means, the weights being the base rates, P and Q = (1 – P), so we have the two equations in the two unknown latent means,

(14) $P\bar{x}_s + (1-P)\bar{x}_n = \bar{x}_t$ Grand mean of x (observable)

(15) $\frac{1}{2}\bar{x}_s + \frac{1}{2}\bar{x}_n = \bar{x}_{hw}$ Mean x of cases lying in hitmax interval on w (observable)

We solve for the latent means $\bar{x}_s$ and $\bar{x}_n$.

## 7. Drawing the Latent Frequency Functions

Possessing estimates of the latent means on an indicator puts us, of course, in a rather powerful position insofar as these estimates are accurate, because on the independence assumption the mean value of an indicator for the sub-sample of cases lying within a certain interval on some other indicator depends only upon the relative proportions of schizotypes and nonschizotypes within that sub-sample. Thus, for example, having estimated by Equations (14) and (15) the latent means $\bar{x}_s$ and $\bar{x}_n$, we write, for the subsample of cases falling within any arbitrary w-interval, the simple relation:

(16) $$\bar{x}_{w_i} = \frac{n_{sw_i}}{n_{w_i}}\,\bar{x}_s + \frac{n_{nw_i}}{n_{w_i}}\,\bar{x}_n$$

in which $\bar{x}_{w_i}$, $\bar{x}_s$, $\bar{x}_n$, and $n_{w_i}$ are known, $n_{nw_i} = n_{w_i} - n_{sw_i}$, so we solve for the single unknown $n_{sw_i}$.

By this operation, which is repeated within each w-interval, we are in effect having the computer draw the latent frequency function $f_s(w)$ and $f_n(w)$. The cumulative frequencies under

each taxon's function f(w), when divided by the total estimated base-frequencies $N_s$ and $N_n$, give us directly the latent hit-rates ($p_s$, $q_n$) and miss-rates ($p_n$, $q_s$) achieved by any arbitrary $\underline{w}$-cut.

## 8. Locating the Interval Within Which the Schizotype:Nonschizotype Frequencies Are Proportional to the Base Rates P:Q.

A point analogous to the hitmax cut when we consider the relativized, intra-taxon density functions $\phi_s$ and $\phi_n$ (which point would be identical with the hitmax cut if the base rate P = Q = 1/2) is that cut at which the relativized ordinates $\phi_s$ and $\phi_n$ are equal. I have not been able to make direct use of the fact that this interval can be found by a simple maximizing procedure, but it is of some theoretical interest, and also provides one among several methods for checking the total consistency of the data (Section 9).

The procedure for locating the interval on an indicator-distribution within which the ratio of schizotypes to nonschizotypes is the same as it is in the entire sample relies upon the fact that the absolute ordinates of the latent frequency functions are, for each taxon, given by the product of the within-taxon density-function (i.e., the ordinate when each taxon's frequency distribution has area = 1) and the base-frequency of that taxon. That is, we know that $f_s(x) = N_s\ \phi_s(x)$ and $f_n(x) = N_n\ \phi_n(x)$ for each indicator $\underline{x}$. Consequently if we find the indicator-value for which the intra-taxon density-functions $\phi_s$ and $\phi_n$ are equated, this will be the point at which the absolute ordinates $f_s$ and $f_n$ are proportional to the base-rates, and therefore, the absolute <u>frequencies</u> of schizotypes and nonschizotypes falling within the interval immediately surrounding this cut will be in proportion to the grand base-rates.

Without prejudging any issues, but simply for convenience of expression, let us define the 'validity of an indicator at a point' as the difference between the valid positive rate $p_s$ and the false positive rate $p_n$ at that point. Such a definition of 'validity' would not, I think, be very useful for most clinical purposes although it is an interesting concept in one way, since its maximum is invariant with respect to the base-rates. The searching procedure involves making use of the fact that the numerator of the phi-coefficient calculated on the (xy)-4-fold table is a maximum when the table is determined by cuts on $\underline{x}$ and $\underline{y}$ which each separately maximize the "validities" of the signs so determined. This has a certain intuitive obviousness about it, because

the expression in the numerator of the phi-coefficient is of course algebraically identical with the covariance in a Pearson r, when the variables substituted in the usual formula for r are variables which in fact take on only the two values 1 and 0.

The searching procedure then consists of sliding arbitrary cuts around on x and y jointly in various combinations until we find that pair of cuts for which the numerator of the phi-coefficient, computed on the 4-fold table resulting from the particular pair of cuts, is a maximum.

Proof

Let "p" without s or n subscript denote the observed probabilities, whether within cells or marginals. In latent terms we write

| | |
|---|---|
| $p_{xy} = P\, p_{sx}\, p_{sy} + Q\, p_{nx}\, p_{ny}$ | Observed probability above both cuts (Cell I) |
| $p_x = P\, p_{sx} + Q\, p_{nx}$ | Observed marginal probability above x-cut |
| $p_y = P\, p_{sy} + Q\, p_{ny}$ | Observed marginal probability above y-cut |

Then the numerator θ of the phi-coefficient on this (xy)-table is

$$\theta(x,y) = p_{xy} - p_x\, p_y = (P\, p_{sx}\, p_{sy} + Q\, p_{nx}\, p_{ny}) - (P\, p_{sx} + Q\, p_{nx})\,(P\, p_{sy} + Q\, p_{ny})$$

$$= P\, Q\, (p_{sx} - p_{nx})\,(p_{sy} - p_{ny}) \qquad (17)$$

That is, the phi-numerator is the product of the base-rates times the product of the sign-validities at the two cuts.

Differentiating, the base-rate being fixed (PQ = K),

$$\frac{\partial \theta}{\partial x} = K(p_{sy} - p_{ny})(p'_{sx} - p'_{nx}) = 0$$

$$\frac{\partial \theta}{\partial y} = K(p_{sx} - p_{nx})(p'_{sy} - p'_{ny}) = 0$$

Since each indicator is "everywhere valid," $P_{sx} \neq P_{nx}$ and $P_{sy} \neq P_{ny}$ regardless of cuts chosen, so these equations have no solutions within the observed range except at the points

$$p'_{sx} = p'_{nx}$$

and $$p'_{sy} = p'_{ny}$$

which are the points of intersection of the intra-taxon density-functions, to wit,

(18) $\phi_s(x) = \phi_n(x)$

(19) $\phi_s(y) = \phi_n(y)$

At each of these points the absolute ordinates are in proportion to the base-frequencies, since

$f_s(x) = N_s\ \phi_s(x)$

$f_n(x) = N_n\ \phi_n(x)$

$f_s(y) = N_s\ \phi_s(y)$

$f_n(y) = N_n\ \phi_n(y)$

And therefore the frequencies of schizotypes and nonschizotypes within the intervals containing these cuts are (approximately) proportional to the base-rates P and Q.

Let us call these abscissa values the phimax cuts on x and on y.

## 9. Some Suggested Consistency Tests for the Postulated Latent Model

Recognizing that the postulated model of the latent situation is an idealization which is very unlikely to be completely in accord with the unknown state of nature, a desideratum for the method's justifiable application is the possibility of detecting departures of the state of nature from the idealization when these departures are so gross as to vitiate the method. Note that I speak of vitiating the method by gross departures. We are oriented toward research contexts characterized by the presence of a rather small amount of purported knowledge and a very large amount of admitted ignorance. In attempting, for example, to arrive at an estimate of the unknown base-rate P in a specified clinical population (such as general psychiatric outpatients), we do not have as our goal a highly precise estimate of P with exact confidence limits attached; and I do not maintain that the estimate of P is a maximum likelihood estimate (although I have no reason for thinking that it is not). In the case of the substantive problem which gave rise to my interest in this approach, it is well known that competent and experienced clinicians in psychiatry and clinical psychology will offer considered judgments as to the value of P in general outpatient practice which vary from as little as 10% schizotypes to as high as 90% schizotypes. Nor is this

entirely a matter of theoretical controversy regarding the status of the 'schizotypy' concept as a fruitful or legitimate one. Even among clinicians who agree that only a minority of schizoid persons decompensate sufficiently to be diagnosable as 'clinical schizophrenia,' one still finds a wide range of estimates as to the schizoid base-rate in a specified clinical population. I cannot emphasize too strongly that, in proposing the present method as worthy of clinical and research trial, I am not talking about a means of sharpening such estimates as that of the unknown base rate P by a few percentage points, as if we were trying to settle a dispute between practitioners who believe that 60% of outpatients are schizotypic and those who opine that only 50% are such. When we start with expert disagreement varying over some 4/5 of the numerically possible range of incidence, any method which will materially reduce this spread would seem to be worth a try.

From this point of view, it would seem methodologically inappropriate to think of any so-called "consistency tests" in the way we customarily think of "significance tests," i.e., of aiming to answer the question whether the data are statistically compatible with the model, where the alternative statistical hypothesis is that the state of nature departs by some amount, however small, from the idealization provided by the model. We begin with the recognition that some departure from the idealization is antecedently so probable that the refutation of this null hypothesis is really of very little scientific interest.

Admittedly this methodological orientation cannot free one of the responsibility for attempting some sort of estimate of the range of parameter values reasonably compatible with the estimates made via the method. Since analytic derivation of such ranges is beyond my competence, and since it is by no means clear whether such analytic derivations are possible given a model deliberately chosen to be weak (except for the independence postulate), perhaps the most fruitful approach at this stage lies in Monte Carlo study of the method's "robustness." Pending the carrying out of such Monte Carlo investigations, I have thought it worthwhile to publish this preliminary discussion of the method in the present series of research reports.

In this section, I shall propose a set of "consistency tests" based upon certain internal relationships among the computable statistics which follow from the idealization and which, therefore, if sufficiently badly unsatisfied by the observational facts, would justify a strong suspicion that the idealization represents a rather poor approximation to the state of nature, and therefore that the estimates of the latent quantities are untrustworthy.

**a. Phimax interval value test:** When the method of Section 8 has been employed to locate the phimax cut on an indicator x, we infer that the cases lying within the interval $x_i$ containing this phimax cut, if the interval is taken sufficiently small around the cut-value, will consist of schizotypes and nonschizotypes in the same ratio as the absolute ordinates $f_s(x_i):f_n(x_i)$, and that this is the ratio of the grand base-rates P:Q. Thus we write, for the y-mean of the cases lying in the phimax x-interval:

$$\bar{y}_{x_i} = \frac{n_{sx_i}}{n_{x_i}}\,\bar{y}_s + \frac{n_{nx_i}}{n_{x_i}}\,\bar{y}_n = P\bar{y}_s + Q\bar{y}_n = \bar{y}_t \tag{20}$$

that is, the mean y of cases lying within the phimax interval of x should equal the grand y-mean.

The situation is similar proceeding in the reverse direction, i.e., the mean x of the subsample of cases lying within the phimax interval on y should equal the grand x-mean.

$$\bar{x}_{y_i} = \frac{n_{sy_i}}{n_{y_i}}\,\bar{x}_s + \frac{n_{ny_i}}{n_{y_i}}\,\bar{x}_n = P\bar{x}_s + Q\bar{x}_n = \bar{x}_t$$

**b. Grand covariance test:** In Section 3 we found a general expression, Equation (2),

$$\mathrm{cov}(xy) = pq\,\Delta\bar{x}\,\Delta\bar{y}$$

for the covariance of an indicator-pair within any mixed sample, namely, it is the product of the schizotypic and nonschizotypic rates in that sample by the product of the differences of the latent means. This expression also holds for the entire population, and therefore should permit us to calculate the grand covariance from our estimates of the latent quantities. So far as I can see, the grand covariance has not entered, even implicitly, into our estimates of the grand latent values, and therefore a good agreement between the estimated grand covariance (calculated from the latent values) and the observed grand covariance provides a consistency test for the model. Suppose we have, for example, been working with three indicator-variables x, y and z. We then write:

| | Calculated | | Observed |
|---|---|---|---|
| (21) | $PQ\Delta\bar{x}\Delta\bar{y}$ | = | $\mathrm{cov}_t(xy)$ |
| (22) | $PQ\Delta\bar{x}\Delta\bar{z}$ | = | $\mathrm{cov}_t(xz)$ |
| (23) | $PQ\Delta\bar{y}\Delta\bar{z}$ | = | $\mathrm{cov}_t(yz)$ |

**c. Hitmax interval covariance test:** This procedure, like the immediately preceding one, employs Equation (2) of Section 3 to calculate an observable covariance, namely, that between two indicators on a subsample lying within the hitmax interval on a third indicator, from the inferred ratio of schizotypes:nonschizotypes within that subsample and the latent means of the two output indicators. As a consistency test, this might appear wholly circular, since the method employed to locate the hitmax interval involved maximizing a covariance. However, locating the abscissa value which yields a maximum on the covariance curve is not the same as actually computing the numerical value of the covariance at this maximum. So that while a certain element of redundancy and therefore perhaps of "circularity" does exist here, it does not seem possible to infer directly, from the fact that the xy-covariance is maximal for cases lying within a certain w-interval, that this maximum will have a specified numerical value. However, if the latent model is a good approximation to the state of nature, this numerical value should be calculable from the inferred latent values. We have estimates of the four latent means $\overline{x}_s$, $\overline{x}_n$, $\overline{y}_s$, and $\overline{y}_n$. Within the hitmax interval on a third indicator, say w, the subsample thus defined consists of schizotypes and nonschizotypes in approximately equal numbers. Hence, the observed (xy)-covariance of this subsample should be theoretically calculable from the formula

$$(24) \qquad \mathrm{cov}_{w_h}(xy) = \frac{1}{4}(\overline{x}_s - \overline{x}_n)(\overline{y}_s - \overline{y}_n)$$

Again, this consistency test can be run in three ways for any three indicators, in ten ways using any set of four indicators, and so on.

**d. Mean difference ratio test:**[*] If the postulated latent model is satisfied, the ratio of the mean differences of a pair of indicators on cases above and below any arbitrary cut taken on a third indicator should be invariant with respect to this cut. We choose an arbitrary cut on x. We compute the observed means of y for cases lying above this cut, and for cases lying below this cut, and record their difference $(\overline{y}_{xa} - \overline{y}_{xb})$. We do the same for z-means above and below this x-cut. We then compute the ratio of these mean differences. Recording this, we move to the next arbitrary x-cut and repeat the procedure. This ratio of mean differences on x and y should not

[*] [Described here as a consistency test, the MAMBAC (Mean Above minus Mean Below A sliding Cut) was subsequently elevated to an additional procedure in the Coherent Cut Kinetics method for detecting taxonicity and generating estimates of base rates and other parameters (Meehl & Yonce, 1994).]

fluctuate, except as a result of random sampling error, as we slide our x-cut along the x-continuum. Failure of this ratio of means to remain constant over the successive splits of the total sample, while I am not prepared to say precisely what departure it indicates, at least suggests that the requirements of the latent model are not being satisfied.

Proof

We first require an expression for the difference of manifest means on one indicator in terms of the latent hit-rates yielded by a given cut on another indicator. Because of the intra-taxon independence condition, the mean y in a subsample demarcated by any arbitrary x-cut is a function of this x-cut only insofar as that cut determines a certain proportion of x-hits and x-misses above and below the cut. Thus, we can write, in terms of the latent hit-rates and miss-rates, the latent means $\bar{y}_s$ and $\bar{y}_n$, and the observed frequencies above and below an x-cut:

$$\bar{y}_{xa} = \frac{H_{sx}\bar{y}_s + M_{nx}\bar{y}_n}{N_a} \tag{25}$$

$$\bar{y}_{xb} = \frac{H_{nx}\bar{y}_n + M_{sx}\bar{y}_s}{N_b} \tag{26}$$

Then the y-mean difference above and below an arbitrary x-cut is given in latent terms by

$$\begin{aligned}\bar{y}_{xa} - \bar{y}_{xb} &= \frac{H_{sx}\bar{y}_s + M_{nx}\bar{y}_n}{N_a} - \frac{H_{nx}\bar{y}_n + M_{sx}\bar{y}_s}{N_b} \\ &= \frac{(H_{nx} + M_{sx})(H_{sx}\bar{y}_s + M_{nx}\bar{y}_n) - (H_{sx} + M_{nx})(H_{nx}\bar{y}_n + M_{sx}\bar{y}_s)}{N_a N_b} \\ &= \Delta\bar{y}\left(\frac{H_{sx}H_{nx} - M_{sx}M_{nx}}{N_a N_b}\right) \qquad (27)\end{aligned}$$

where $\Delta\bar{y} = (\bar{y}_s - \bar{y}_n) =$ a constant, the difference between the latent taxa means.

Similarly for a second output indicator z we have

$$\bar{z}_{xa} - \bar{z}_{xb} = \Delta\bar{z}\left(\frac{H_{sx}H_{nx} - M_{sx}M_{nx}}{N_a N_b}\right)$$

Hence for any arbitrary x-cut the ratio of the mean differences on y and z is

$$\frac{\bar{y}_{xa} - \bar{y}_{xb}}{\bar{x}_{xa} - \bar{x}_{xb}} = \frac{\Delta\bar{y}}{\Delta\bar{z}} = C \tag{28}$$

**e. <u>Phimax means-by-frequencies test</u>:** If the latent model is satisfied, the mean difference on an output indicator between the subsamples above and below a cut on an input indicator, when multiplied by the product of the absolute frequencies above and below the cut, will be a maximum when this cut is the phimax cut.

Computationally, we start at some arbitrarily low value on the manifest x-distribution and choose an x-cut. We calculate the mean on an output indicator, say y, for the two subsamples of cases falling above and below this x-cut. We then count the number of cases $N_{ax}$ lying above the x-cut and the number of cases $N_{bx}$ lying below the x-cut, and multiply the product ($N_{ax}$ $N_{bx}$) of these absolute frequencies by the difference between the observed y-means ( $\overline{y}_{ax} - \overline{y}_{bx}$ ).We record this value, and then move up to the next arbitrary x-cut and repeat the procedure. Plotting the graph of these products as determined by consecutive arbitrary x-cuts, we look to see whether this graph shows a fairly clear maximum. If it does, the x-cut yielding this maximum should correspond closely to the phimax cut on x as previously determined by the method of Section 8.

<u>Proof</u>

In subsection (d) just preceding we saw that the difference between y-means above and below any x-cut is (Equation 27)

$$\overline{y}_{xa} - \overline{y}_{xb} = \Delta\overline{y}\left( \frac{H_{sx}H_{nx} - M_{sx}M_{nx}}{N_a N_b} \right)$$

So

$$N_a N_b(\overline{y}_{xa} - \overline{y}_{xb}) = \Delta\overline{y}\left(H_{sx}H_{nx} - M_{sx}M_{nx}\right)$$

Since $\Delta\overline{y} = \overline{y}_s - \overline{y}_n =$ a constant, the left-hand expression is maximized when

$$\text{MAX}(H_{sx}\, H_{nx} - M_{sx}\, M_{nx})$$

Differentiating,

$$\frac{d}{dx}(H_{sx}H_{nx} - M_{sx}M_{nx}) = 0$$

$$H'_{sx}H_{nx} + H_{sx}H'_{nx} - M'_{sx}\,M_{nx} - M_{sx}M'_{nx} = 0$$

$$-f_{sx}H_{nx} + H_{sx}f_{nx} - f_{sx}M_{nx} + M_{sx}f_{nx} = 0$$

$$f_{nx}(H_{sx} + M_{sx}) = f_{sx}(H_{nx} + M_{nx})$$

(29) $$\frac{f_{sx}}{f_{nx}} = \frac{H_{sx} + M_{sx}}{H_{nx} + M_{nx}} = \frac{N_s}{N_n} = \frac{P}{Q}$$

But since $f_{sx} = N_s \phi_s(x)$

$f_{nx} = N_n \phi_n(x)$

The maximum occurs at

$$\phi_s(x) = \phi_n(x)$$

the phimax cut on x.

**f. <u>Indicator combinations test</u>:** Since the sequence of procedures described in Sections (3)-(4)-(5)-(6)-(7) can be carried out by putting different indicators from the available set in the several roles of "input" and "output" indicators, it is obvious that there are alternative routes, employing only partially redundant information, for arriving at estimates of the same hypothetical cuts and the same latent quantities. Thus, for example, if only three indicators are available, this sequence of operations can be carried on beginning with a determination of the hitmax cut on x, where we employ cov(yz) as the output measure to be maximized; but we could equally justifiably begin by locating the hitmax cut on y, employing cov(xz) as the output to be maximized, or the hitmax cut on z via the maximization of cov(xy). Again, we may draw the latent frequency-functions as outlined in Section 7 for indicator z by studying the y-means of consecutive z-intervals, and also by studying the behavior of the x-means for successive z-intervals. Obviously the number of alternative routes to estimating the same latent quantities or frequency functions rises in a positively accelerated manner as the number of antecedently plausible indicators available for study increases. The location of the hitmax and phimax cuts, the estimates of the grand base-rates P and Q, and the shapes of the latent frequency functions should all be in approximate agreement when arrived at by employing different patterns of

indicators in different input and output roles, and if several indicators are available the element of redundancy as a source of "spurious consistency" should be considerably reduced. It would seem, although I am not prepared to demonstrate this rigorously, that the most exacting test upon the model would be a close conformity throughout the range between two or more nonredundant drawings of the latent frequency functions as described in Section 7. Unless I am mistaken, this is possible if we have five indicators to work with.

We note that each indicator has a theoretically calculable mean, a theoretically calculable variance, and a theoretically calculable covariance on subsamples lying within the hitmax interval on each other indicator; and similar statistics are calculable for subsamples lying within the phimax intervals on each other indicator. The subsample of cases identified by the fact that they lie within the phimax interval on any indicator should, on the independence postulate, constitute a little representative group whose properties with respect to other indicators taken singly or in pairs essentially "duplicates" the statistical properties of the entire group as to mean, variance, and covariance because the representation of schizotypes and nonschizotypes in this phimax-located subsample is in proportion to the grand base-rates P:Q. Similarly, the subsample of cases defined by their falling within the hitmax interval of an indicator must have statistical properties arising from an even mixture of the latent taxa.

**g. General comment on consistency tests:** I am acutely aware of the absence of any precise statement as to how much "consistency" should be required or expected from the tests proposed in the preceding subsections (a)-(f). It does not seem that any straightforward test of "significant inconsistency" could be easily developed, but this hardly seems appropriate anyhow because, as stated before, we are not very hopeful that literal fulfillment of the latent model will be possible, so that here again the really important question in practice is one of robustness. What we need to know is how much disagreement in these consistency tests is still compatible with the hypothesis that the state of nature is adequately approximated by the idealization where 'adequate' means 'close enough to justify employing the method to improve our guesses as to the latent values by a worthwhile amount in relation to our initial state of ignorance.' Even allowing for some considerable redundancy, we might hope that, taking all of the consistency tests together, a crude "inspectional" examination of the numerical agreement between different estimates of the same latent quantity would provide a modicum of assurance that the idealization was being

approximately satisfied and that some averaged value of the several estimates of the same latent quantity or frequency function could be taken as an improved approximation to the unknown parameter.

## 10. Inverse Probability of Schizotypy for Each Sign-Combination

If the results of the previous procedures and consistency-tests lead us to accept the latent model as an adequate approximation to the state of nature and consequently put us in possession of estimates of the latent parameters, we can employ Bayes' Rule and derive the inverse probability of a patient's membership in the latent taxa, given his values on the several indicators. In any clinical population, it is this inverse probability that is of the most immediate clinical importance. While in practice we frequently are forced to rely only upon the conditional probabilities, or in many situations (e.g., using MMPI scales) only upon the conditional probabilities within the "normal" group, the inverse probability that a patient who shows a pattern of symptoms or signs does so on the basis of a particular latent condition is the most interesting and useful number. In ordinary practice of clinical medicine such inverse probabilities, while not formally computed using Bayes' Rule, are adjusted subjectively by the practitioner on the basis of his rough knowledge of the prior probabilities. For example, an internist draws a certain diagnostic inference from a given set of complaints partly because his recent practice shows that "there is a lot of a new kind of flu bug going around these days." Or, a positive Wasserman in Minnesota (if it is not a technical mistake) constitutes pretty strong evidence for lues; whereas in Puerto Rico this inference is much more shaky because of the non-negligible prior probability that the patient is suffering from yaws.

Let us suppose that on the basis of our inferences concerning the latent situation we have assigned hitmax cuts on all three indicators $\underline{x}$, $\underline{y}$, $\underline{z}$. Then if each indicator were considered singly, every patient falling above its hitmax cut would be classified as a schizotype, and every patient falling below the hitmax cut would be classified as a nonschizotype. This dichotomous use of the cut is what I mean by the word 'sign.' That is, the $\underline{x}^{+}$-sign is manifested by a patient if his observed score on the $\underline{x}$-indicator exceeds $x_h$. A patient is '$\underline{x}$-positive' if the $\underline{x}$-sign is present, and '$\underline{x}$-negative' if the $\underline{x}$-sign is absent.

Such a situation gives rise to 8 possible (and, since we are dealing with fallible indicators throughout, actually realized) sign patterns among a group of patients. Our information at this

point permits us to apply Bayes' Rule and calculate the inverse probability that a patient who shows a given sign-pattern is a schizotype. Consider, for example, a patient who shows the sign-pattern ($x^+ y^- z^+$). That is, this patient is above $x_h$, below $y_h$, and above $z_h$. What is the probability of his being schizotypic? We write

$$p(s/x^+y^-z^+) = \frac{Pp_{sx}q_{sy}p_{sz}}{Pp_{sx}q_{sy}p_{sz} + Qp_{nx}q_{ny}p_{nz}} \tag{30}$$

and similarly for each of the 8 cells in the xyz-cube, defined by the 8 possible indicator sign-patterns.

Assuming invariance of the relativized frequency functions $\phi_s$ and $\phi_n$ when they are found in differently defined populations (e.g., in-patients, out-patients, private practice, college sample, "general population") such a table of inverse probabilities can be constructed for various values of P/Q. Whether any decision about an individual patient should be made on the basis of a sign-pattern which yields an inverse probability close to one-half depends, of course, upon utilities, possibility of collecting further information, and the like. For research purposes, especially research devoted to theoretical interests (such as testing a specific genetic model), it may be more legitimate and fruitful to confine attention to groups falling in those cells for which the inverse probabilities are fairly close to zero or one, leaving out of consideration the subsamples found in highly doubtful cells where the inverse probability is closer to one-half. It goes without saying that whether the hitmax or some other cut should be used for defining a sign depends upon utilities, and of course with a sufficiently large N and fairly smooth inferred latent frequency distributions (Equation 16) one might look not at areas above and below a fixed cut but instead at areas above and below the individual patient's scores, or even the probability-density associated with his particular score-combinations.

## 11. Proving That an Indicator is Infallible by Using Fallible Indicators: The Super-Bootstraps Effect.

In our paper on construct validity, Cronbach and Meehl (1955) introduced the term 'bootstraps effect' to designate a paradoxical feature of the growth of knowledge concerning theoretical entities, namely, that the construct validity of an observational measure as an indicator of an inferred or latent variable is, over the course of time, sometimes elevated to a

higher status than that status accorded to the indicators by means of which the construct was initially identified and the (now high-status) indicator was originally "validated." The classic example in psychology is, of course, the intelligence test, which was originally validated against such crude criteria as chronological age, age-grade placement, course grades, teacher or peer-group ratings, and the like. But the net impact of all this evidence, plus intra-test and inter-test relationships, has resulted in a bootstraps effect such that informed persons today will view a combination of high IQ with poor school performance as a problem in identifying the "interfering factors" (i.e., what factors lead a bright student to underachieve academically?). There is no question about the occurrence of this phenomenon in the history of science, and it is of course not confined to psychology. It is probably safe to say that no single logical or mathematical model for the carrying out of a bootstraps effect on an indicator-variable can be specified, and we should not expect this methodologically because there are as many legitimate avenues to construct validity as there are ways to corroborate or discorroborate a theory (including, of course, as strictly speaking part of the theory, a postulated causal relation between an indicator and the construct variable of interest). While the methodological necessity of some such notion as construct validation seems to have become generally accepted among psychologists, detailed analysis of its various modes, which were discussed only in a rather general and sketchy fashion by Cronbach and myself, have not yet been forthcoming.

In the present section, I present a proof of a simple theorem in construct validation which is, at least to most psychologists of my acquaintance, an extreme form of the paradoxicality of the bootstraps effect and strikes most of us trained in the orthodox tradition of validation procedures as counter-intuitive. I call it the 'super-bootstraps effect' for reasons which will be apparent.

The counter-intuitive claim I wish to make is that if an infallible sign (i.e., a sign which is pathognomonic both as an inclusion and as an exclusion test) exists, and we are clever or lucky enough to try it out, its infallibility can be demonstrated by studying its statistical relationship to a set of signs which are admittedly fallible, individually and jointly. If the infallible sigh is observationally dichotomous (e.g., certain biochemical reactions, a neurological phenomenon clinically judged as either present or absent, a single test item or life-history fact) its infallibility can be corroborated. More generally if a quantitative indicator-variable is "potentially infallible," in the sense that there exists a hitmax cut on its manifest distribution—or, as would presumably rather be the case in practice, a hitmax region—for which the number of latent misclassifications

is zero, this perfect hitmax cut or region can be located by a systematic searching procedure and, when found, its infallibility can be corroborated. To distinguish this method of validation from the orthodox validation in which the validity of an indicator is investigated by studying its presence as a function of the (known or assumed) taxon membership of individuals, whereas in the proposed method the validity of the new indicator is investigated by studying its incidence as a function of the (known or inferred) base-rate over a varying set of sub-populations, Dawes and Meehl (1966) have proposed a designation of the orthodox approach as criterion group validation and the new method as mixed group validation. This terminology must not be understood to suggest that the fundamental methodological meaning of the word 'validation' is being surreptitiously changed so as to be (illegitimately) applied to the proposed method. The methodological aim in both situations is still that of classifying individuals; and the estimated validity, in both situations, is a number purporting to state the accuracy with which individuals can be classified by the given sign. The validity being estimated is the same; the method of validation is different. However, logically analyzed, orthodox criterion validation is a special case of mixed group validation, namely, that in which the base-rates of two "criterion groups" initially known or assumed to be "pure" groups of the latent taxa are therefore taken to have base-rates of zero and one respectively. Needless to say, this criterion purity is almost impossible to realize in practice except in pure prediction situations (e.g., industrial or military selection) where the pragmatic context stipulates a truly "operational" criterion which is in itself the criterion of interest.

Viewing mix group validation as the general case of which orthodox criterion validation is a special case helps to reduce somewhat the counter-intuitive flavor of the super-bootstraps theorem. One might look at it this way: Since it is possible to locate an infallible cut on a potentially infallible indicator by showing that the resulting sign is positive for all individuals who belong to the taxon of interest, and that this same sign is negative for all individuals not belonging to the taxon; then, generalizing from the limiting case, it ought to be possible to corroborate infallibility by showing that the proportions of positive and negative signs yielded by an infallible cut can be made to match the proportions of individuals belonging to the latent taxon over a set of subpopulations for which these proportions differ. While it is fairly obvious intuitively that the infallibility of the new sign is a sufficient condition for the proportion of sign-positives to match the varying base-rates over subpopulations, it is somewhat surprising to

realize that this is—of course setting aside sampling errors—also a necessary condition for such a match. But it is easy to show by a little algebra that the infallibility condition is both necessary and sufficient, assuming invariance of the latent $\phi$-functions over populations.

Proof

Consider two populations i and j for which the base-rates $P_i \neq P_j$ are known or somehow inferable. Let $p_i^+$ be the probability of sign-positives among individuals in population i, and $p_j^+$ be the corresponding probability in the other population. The valid positive rate is $p_s$ [= proportion of schizotypes who show sign (+)] and the false positive rate is $p_n$ [= proportion of nonschizotypes who show sign (+).] We assume that the latent density-functions $\phi_s$ and $\phi_n$ are unchanged in moving from one population to another, although of course the unrelativized frequency-functions $f_s$ and $f_n$ will shift with the base-rates.

Then the sign-positive rates for the two populations can be written in latent terms,

(31) $$p_i^+ = P_i p_s + Q_i p_n$$

(32) $$p_j^+ = P_j p_s + Q_j p_n$$

Suppose we impose the observational condition that the difference in sign-positive rates between the two populations should equal the known base-rate difference, that is, that

(33) $$p_i^+ - p_j^+ = P_i - P_j$$

which in latent terms says that

$$(P_i p_s + Q_i p_n) - (P_j p_s + Q_j p_n) = (P_i - P_j)$$

$$p_s(P_i - P_j) - p_n(P_i - P_j) = (P_i - P_j) \neq 0$$

and, dividing by the base-rate difference,

$$p_s - p_n = 1$$

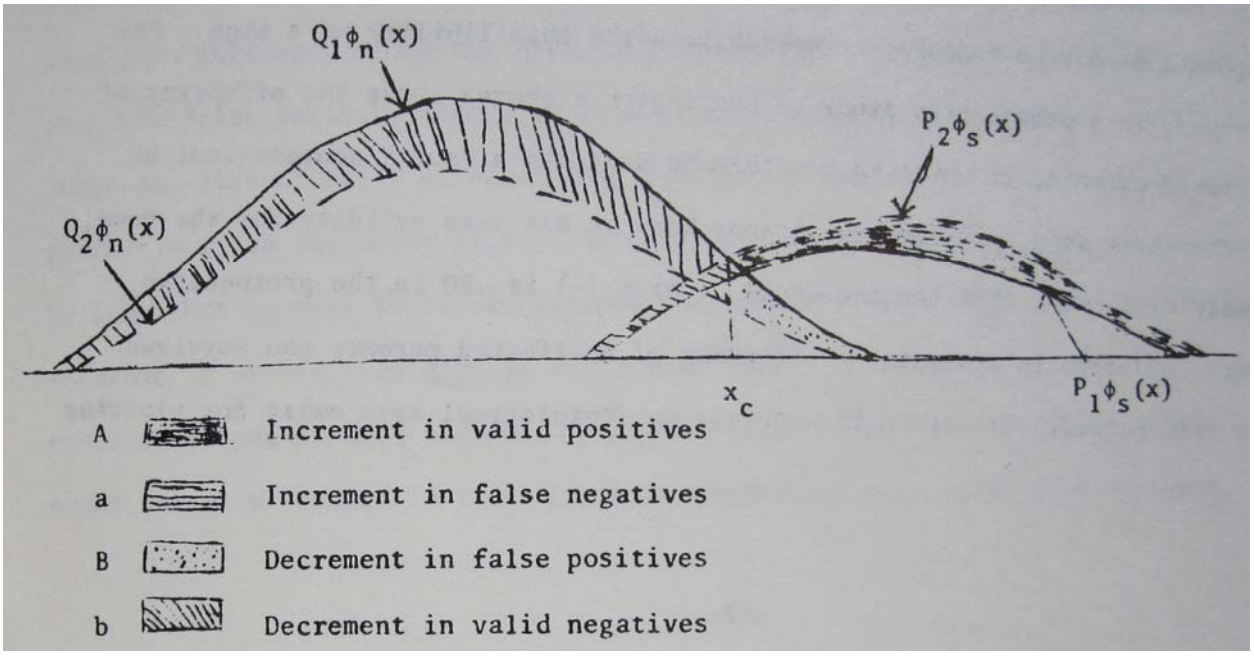
$$p_s + q_n = 2$$

which, since $p_s \leq 1$ and $q_n \leq 1$, can be true only if

$$p_s = q_n = 1$$

i.e., the sign is infallible.

The effect of moving from a population with base-rate $P_1$ to a different population with a larger base-rate $P_2$ is to multiply the ordinates $\phi_s$ and $\phi_n$ of the intra-taxon density-functions by $P_2$ and $(1 - P_2)$ respectively, instead of by $P_1$ and $(1 - P_1)$. Graphically (see figure), if $x_c$ were fallible, holding $x_c$ fixed and moving from base-rate $P_1$ to base-rate $P_2$, and setting <u>total</u> area at unity, how are the areas related?




What happens to the observed $x^+$-rate when we move from a population distributed like the solid-curve pair to one distributed like the dashed curve-pair? The $x^+$-rate rises, attributable to the large area A. But this rise cannot be as large as the true schizotype-increment $\Delta P = (P_2 - P_1)$, because area <u>a</u> is below the cut, and area <u>b</u> is also being subtracted from the observed $x^+$-rate as the nonschizotypic curve $Q\phi_n(x)$ shrinks from solid to dashed level. It is geometrically obvious that the net change observed in the $x^+$-rate cannot be as large as the true shift in schizotype-rate, because

$$\Delta P > A - b$$

Even if the sign were unidirectionally pathognomonic, say, perfect as an exclusion test ($q_s = 0$) while fallible as an inclusion test ($p_n > 0$), we still could not match the new base-rate $P_2$ with our new observed $x^+$-rate, since although

$$\Delta P = A \qquad (a = 0)$$

the increment in $x^+$-rate is

$$\Delta x^+ = A - b \qquad (b > 0)$$

so

$$\Delta x^+ = A - b = \Delta P - b < \Delta P$$

Any source of information as to the base-rates of a set of two or more populations may be employed to corroborate the infallibility of a sign. For example, in a prospective study of Huntington's chorea among the offspring of affected parents, it would be possible to validate a "soft" neurological or psychometric sign, provided we already know it has some validity for the gene, simply by showing that the incidence of sign (+) is .50 in the prospective sample and zero in a sample of offspring of unaffected parents who survived the risk period. Or, again if accurate epidemiological data exist for plotting the curve of morbidity risk as a function of life-age and we are studying a biochemical indicator which allegedly is an infallible indicator of the initial, subclinical development of the pathological condition prior to its production of clinical manifestations, a proportional relation between incidence of positive reactions and chronological age in a population of persons, none of whom is known to be affected when seen, would strongly corroborate the claim of validity. It does not matter by what means or from what source we get knowledge of the base rates in a set of sub-populations; as long as such base-rates can be estimated, the base-rate validation procedure can be used.

One's immediate impulse is to object that surely some <u>other</u> factor, unrelated to the taxon of interest, could yield matching sign-positive rates. Why couldn't tuberculosis, for example, be latent in half of the offspring of choreic parents, and in none of the offspring of controls? This objection is very forceful intuitively, and I persist in feeling its force myself; but it is not sound, given the assumed invariance of the intra-taxon density-functions. Consider the Huntington's chorea example, keeping in mind that we are ignoring sampling errors and discussing sub-population parameters throughout. The allegedly irrelevant tuberculosis sign (say, a positive Mantoux) is observed to have zero incidence among the offspring of non-choreic parents. Hence we know that the false positive rate $p_n = 0$ (i.e., the Mantoux test is, oddly enough, at least one-directionally pathognomonic for the Huntington chorea gene, functioning perfectly as an inclusion test). Hence among the offspring of affected parents we know that <u>none</u> of those who escaped the gene will show a positive Mantoux. (Because if any of them <u>did</u>, it would mean that

$p_n > 0$ in this sub-population, contrary to the assumed ($\phi$-function invariance.) From which it follows that every positive Mantoux in the second sub-population must arise from the gene. But the only way this can be true, if there are exactly 50% positive Mantouxs and exactly 50% carrying the gene, is for the Mantoux to be also pathognomonic as an exclusion test. If it were not, some of the 50% positives would have to be false positives, to make up for the false negatives. But there aren't any false positives. Hence there aren't any false negatives either. The same reasoning, mutatis mutandis, applies if base-rates other than .00 and .50 are used. One tends to imagine some sort of counterbalanced jugglery of invalid components due to nuisance variables that will match the two base-rates. But it just can't be done without changing the intra-taxon positive rates $p_s$ and $p_n$. If my general proof leaves the reader unconvinced, I can only invite him to try cooking up numerical counter-examples, which he will find it impossible to do.

The statistics do not tell us why the sign is infallible, of course. Why all, and only, chorea-gene-carriers should have a positive Mantoux would present a baffling theoretical problem. But the super-bootstraps theorem is concerned with sign-pathognomicity, whatever may be the causal analysis of this fact.

Returning to the schizotype-identification problem, suppose that the methods described in the preceding sections have been employed on a mixed sample with the result that we are reasonably satisfied that the latent model is adequately enough approximated to use the method, and that we are therefore in possession of inverse probabilities of schizotypy from Equation (30).

Sign-patterns of the three fallible indicators x, y, z define samples from 8 sub-populations for which the inverse probability figures estimate the 8 latent base-rates of schizotypy, cell by cell. Consider a fourth indicator-variable v which is potentially infallible, i.e., there exists a hitmax cut $v_h$ which divides schizotypes and nonschizotypes with zero misclassifications resulting. If that cut is chosen, thereby determining for every patient whether he is v-positive or v-negative, the percentage of $v^+$ cases in the subsample defined by each of the 8 xyz-patterns will match the inverse probabilities. Of course even if such a potentially infallible cut on v exists, we do not immediately know where the cut is located. However, this presents no new problem in the light of what has gone before, since we proceed systematically from low to high cuts on v and record the discrepancies between the observed $v^+$-rates and the inverse probabilities. The discrepancies should approximate zero over all cells as we approach a particular value of the v-cut, and thereafter should again increase. Except for sampling errors in the inverse probability estimates

and, of course, discrepancies due to imperfect realization of the model, there would be a v-cut (or a set of cuts in a certain v-region) such that when any of these cuts is used to define the sign $v^+$, the $v^+$-proportion exactly matches the inverse probability in every cell.

Of course we have assumed throughout that the integrals of the separate and joint density functions (i.e., the valid and false positive and negative rates) remain fixed as we move over the different populations. This is the same assumption made ordinarily in estimating validity by the orthodox method of criterion validation, as pointed out by Dawes and Meehl (1966). However, it should be noted that whereas this assumption is required in order to show that sign-infallibility is both sufficient and necessary if the incidence of sign (+) is to match the base-rates, we can afford to settle for something less than this deductive necessity. Suppose that the shapes of the relativized frequency functions $\phi_s(x)$, $\phi_n(x)$, $\phi_s(y)$, etc. do undergo significant changes in moving from one subpopulation to another. (Of course if this is true the consistency tests should not be satisfied.) Then it is of course algebraically possible for the resulting shift in the valid and false positive rates ($p_s$, $p_n$) to be such as to compensate precisely for the shift in base-rates, in such a manner that the incidence of observed sign-positives follows the base-rates even though latent misclassifications are occurring. But this kind of systematic counterbalancing, while it might happen with some frequency when only two populations are under study, obviously has a negligibly low probability over any appreciable number of populations. For example, consider a situation in which, to make the arithmetic easy, we assume that each of the three fallible indicators x, y, z achieves an 80% hit-rate when optimally cut, and that this hit-rate is symmetrical as between the two kinds of misclassifications. A latent hit-rate of this size is far from infallible and yet not absurdly optimistic, since it is about as good as we can get with individual MMPI scales employing formal psychiatric diagnosis as the "criterion." Let us assume the overall base-rate is P = l/2. Then the inverse probabilities over the 8 cells of the (xyz)-table will vary from a low of .02 (for patients showing the sign-pattern $x^-y^-z^-$) to a high of .98 (for patients showing the sign-pattern $x^+y^+z^+$). This gives us a very nice range of base-rate values estimated from the inverse probability formula, and a close matching of the incidence of the new sign $v^+$ with these 8 cell-probabilities from .02 to .98, cell by cell, would require—while not algebraically impossible—a remarkable set of compensatory coincidences to achieve. If, as Einstein opines, "Nature is not malicious," a good matching will indicate infallibility almost as strongly as if infallibility were a rigorously necessary condition. And, of course, we have utilized

only a small portion of our total information. If we can find a $\underline{v}$-cut which behaves "infallibly," the $\underline{v}^+$-rate employing that cut should also follow the inferred latent proportions on the indicators $\underline{x}$, $\underline{y}$, and $\underline{z}$ taken separately when these latent frequency functions are drawn as described in Section 7. That is the $\underline{v}^+$-rate will match the schizotype-rate inferred from Equation 16, interval by interval. So that instead of the $\underline{v}^+$-rate merely matching the 8 inverse probabilities for the 8 cells of the sign-combination table, we require of $\underline{v}^+$ that its incidence should match the inferred schizotypic rates over, say, the $3 \times 10 = 30$ class-intervals on the $\underline{x}$, $\underline{y}$, and $\underline{z}$ distributions collectively.

## 12. Mixed-Group Validation of a Fallible Sign

A generalization of the super-bootstraps theorem due to Dawes will be published elsewhere (Dawes and Meehl, 1966), in the context of a discussion of the mixed group validation method. Dawes' proof, which will be briefly reproduced here, shows that the super-bootstraps theorem is a special case of the more general principle that the validity (in the sense of hit-rates) of a dichotomous sign can be estimated from the observed incidence of sign-positives in two populations for which the base-rates are known.

Proof

Consider two populations within which the schizotype base-rates are $P_1$ and $P_2$. In both populations, let the valid positive rate be $p_s$ and the false positive rate be $p_n$.

In latent terms,

$$P_1 p_s + Q_1 p_n = p_1^+ \tag{34}$$

$$P_2 p_s + Q_2 p_n = p_2^+ \tag{35}$$

Solving for the unknown latent positive rates $p_s$ and $p_n$,

$$p_s = \frac{p_1^+ Q_2 - p_2^+ Q_1}{P_1 Q_2 - P_2 Q_1} \tag{36}$$

(37) $$p_n = \frac{p_2^+ P_1 - p_1^+ P_2}{P_1 Q_2 - P_2 Q_1}$$

expressing the latent "validity" values $p_s$ and $p_n$ in terms of the known base-rates and the observed sign-positive rates. (Since $P_1 = P_2$ and $P = 1 - Q$, the determinant of the system cannot vanish.)

The super-bootstraps theorem is seen as a limiting case of this, substituting $P_1 = p_1^+$ and $P_2 = p_2^+$ to get

(38) $$p_s = \frac{P_1 Q_2 - P_2 Q_1}{P_1 Q_2 - P_2 Q_1} = 1$$ i.e., perfect exclusion test

(39) $$p_n = \frac{P_2 P_1 - P_1 P_2}{P_1 Q_2 - P_2 Q_1} = 0$$ i.e., perfect inclusion test

So we see that the more closely the sign-positive rates approximate the known base-rates, the more validity the sign must be inferred to possess. Of course there are really two "validity" measures, one positive and one negative. Different functions of the known quantities P, Q, $p_s$ and $p_n$ are constructible as measures of "overall, two-way validity," depending upon a somewhat arbitrary explication of this explicandum. In a truly taxonomic situation, the most useful specification of "validity" is probably provided by stating the two latent rates $p_s$ and $p_n$ rather than by a single number that is a function of them. In clinical medicine, it has usually been deemed more useful to think in terms of both the inclusion and exclusion power of a sign or cut, rather than in terms of a "correlational" number computed over the fourfold table. And I do not believe that this tradition is properly viewed as a lack of statistical sophistication among physicians.

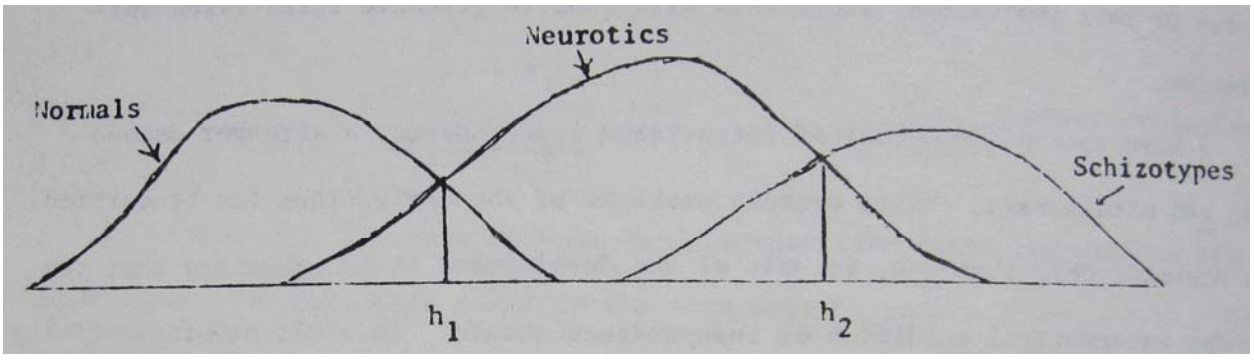
## 13. The Assumptions

With regard to the assumptions made throughout the preceding sections, we note the freedom from such familiar demands as normality and homogeneity of variance—assumptions which are quite unwarranted when schizotypy is the subject-matter. It should also be emphasized that imperfect reliability of each indicator is allowed for, it being one causal source of the overlap between the latent distributions. However, the systematic "error" component of an indicator's observational measure, whether included in reliable or unreliable variance, may vitiate the

method if such systematic error reflects nuisance variables shared by two or more indicators, since this will tend to generate intra-taxon correlation.

I have spoken throughout of intra-taxon independence, a stronger demand than $\underline{r} = 0$ within taxa. While certain portions of the derivations can be carried out assuming only that $\underline{r} = 0$, in most of the development it is necessary that the broader mathematical condition of independence obtain. Thus all proofs involving the latent hit rates $p_s$ and $q_n$ as multipliers within two sub-populations presuppose that the latent density-functions $\phi_s$ and $\phi_n$ for selected "output" indicators are invariant over the "input" indicator being studied.

If there are more than two latent taxa, the critical question is, of course, whether we shall be able to discover this fact so as to avoid making erroneous inferences by presupposing a two-taxon situation. (I believe that certain features of the method are generalizable to the multiple taxon case, but the situation seems to be pretty complicated.) Of course the first effect of multiple taxa is to generate intra-taxon correlations. Thus, if sex were (carelessly) allowed to vary in a study of schizotypy, and two or more of the proposed schizotype-indicators were appreciably sex-related, the existence of a male/female "mixture" within the schizotypic and nonschizotypic sub-populations would generate latent indicator-correlations within these sub-populations; but hopefully, the resulting violation of the intra-taxon independence postulate would be manifested by unsatisfactory outcomes of the consistency tests (Section 9). In addition, the presence of multiple taxa will, under a considerable variety of parametric assumptions, give rise to multiple maxima when the searching procedures (e.g., hitmax cut location) are tried. For example, a latent situation like the one diagrammed will yield two hitmax cuts:




I hope to present a general treatment for $m > 2$ taxa subsequently.*

* [For subsequent discussion of multiple taxa, see Meehl (1999, 2001).]

Quite unsolved at present is the problem of sample size. As is well-known, correlational statistics such as the $cov_w(xy)$ used in hitmax location are excessively subject to sampling instability. Pending Monte Carlo study, I shall only say that some crude rule-of-thumb thinking suggests that the method is probably pointless unless one has several hundred cases to work with. In clinical research as presently conducted this sampling limitation is, of course, a grave disadvantage.

The first reaction I have elicited from most psychologists is concern over the independence postulate. The reader may satisfy himself that with regard to hitmax cut location (Section 3), a considerably weaker assumption suffices for the result, namely, that the two intra-taxon covariances must be equal. The generalization of Equation 2, if $p_w$ is the schizotype-rate in interval w, is

$$(40) \quad cov_w(xy) = p_w\, cov_s(xy) + q_w\, cov_n(xy) + p_w q_w\, (\bar{x}_s - \bar{x}_n)(\bar{y}_s - \bar{y}_n)$$

so that, if $cov_s(xy) = cov_n(xy) \neq 0$, the maximum value of the observable $cov_w(xy)$ will still be located at w-hitmax. This result is very important, because in practice it will usually be easier to achieve approximately equal covariances at some moderate value than to hold both covariances at near-zero. If one or more of the indicators is a "constructed" psychometric variable (e.g., an MMPI anhedonia key), item-analytic procedures can be employed to "rig" the scale's properties so that its covariances with another indicator are approximately equal within taxa, the taxa being clinically identified (e.g., "unquestionably schizotypic" versus "normal controls") for this preliminary purpose.

A numerical example may perhaps be somewhat reassuring, pending Monte Carlo results, as to the robustness against non-zero intra-taxon correlation. How far off will our results be, due to systematic bias in our hitmax location, if the intra-taxon covariances are non-zero and unequal? Differentiating Equation 40 with respect to $p_w$ to maximize the (xy)-covariance,

$$(41) \quad \frac{d}{dp} cov_w(xy) = cov_s(xy) + cov_n(xy)\frac{dq}{dp} + \Delta\bar{x}\,\Delta\bar{y}\left(q + \frac{dq}{dp}p\right) = 0$$

$$(42) \quad \frac{cov_s(xy) - cov_n(xy)}{\Delta\bar{x}\,\Delta\bar{y}} = p - q$$

(So we see that if the intra-taxon covariances are equal, p = q = 1/2, i.e., we find the hitmax interval on w, as in the independence situation.)

Our error in locating $w_h$ if these covariances are unequal arises because the maximum condition (42) is satisfied by a sub-sample lying in an interval where $p \neq q$, and the $|p - q|$ discrepancy will be proportional to the difference between the two intra-taxon covariances on the left. The constant of this proportionality will be half as large as the reciprocal of $\Delta\bar{x}\,\Delta\bar{y}$.

Suppose all variances equal. Let $r_{xy} = .40$ among schizotypes but only half this value among nonschizotypes. Assume that the latent validities are such as to yield one sigma separation between means. Then, substituting these assumed values in (42), we find

(43) $$.60 - .40 = p - q$$
$$p = .60$$

Hence our maximizing the observed cov(xy) has located, instead of $w_h$, a $\underline{w}$-interval within which the schizotype:nonschizotype ratio is 60:40, instead of 50:50 as at hitmax. [That an extremum is a maximum is shown by the fact that

$$\frac{d^2}{dp^2}\mathrm{cov}(xy) = \Delta\bar{x}\,\Delta\bar{y}(q-p)' < 0$$

throughout the range.]

What does this systematic mislocation of $w_h$ do to our estimates of the latent means? Assume for the moment that we had an accurate estimate of the grand base-rate P. The latent rates in the pseudo-hitmax interval are actually $p_w = .60$ and $q_w = .40$, instead of $p_w = q_w = 1/2$ as we are erroneously assuming them to be when we apply Equation 15. Suppose our indicator is an MMPI scale for which the latent mean T-scores are, say, $\bar{x}_s = 70$ and $\bar{x}_n = 60$. Assume the grand base-rate to be P = .80. Then the (observed) grand mean is

$$\bar{x}_t = P\bar{x}_s + Q\bar{x}_n = 68$$

The mean observed for cases lying within the pseudo-hitmax interval is

$$\bar{x}_{w(h)} = (.6)70 + (.4)60 = 66$$

whereas the cases within the true hitmax interval have, of course, a mean of 65, halfway between the latent values.

Solving for the latent means by plugging our (erroneous) values in Equations 14-15 of Section 6, we have

$$\frac{1}{2}\bar{x}_s + \frac{1}{2}\bar{x}_n = 66$$

$$.8\bar{x}_s + .2\bar{x}_n = 68$$

yielding the erroneous latent means

$$\bar{x}_s = 69.33 \qquad \text{(Instead of 70)}$$

$$\bar{x}_n = 62.67 \qquad \text{(Instead of 60)}$$

which are in error by around 1% and 4% of the true values (or by .07 sigma and .27 sigma, or by 1% and 5% of the 5σ range) respectively.

Looking at the matter another way, several derivations involve terms in which we take the joint-probability of two "signs" being positive (i.e., of a case falling above both indicator-cuts) as given by the product of the separate sign(+) probabilities within a taxon. Let us suppose that the intra-taxon $\underline{r}_{xy} = .39$ (a correlation which most clinical investigators would be pleased to get up to in their research, so a little care and ingenuity should enable us to hold it down that low!). Then from the usual $\underline{r} - \phi$ equivalence formula, the ϕ-coefficient treating these indicators dichotomously will be

$$\phi = .637\underline{r} = .25$$

What will this intra-taxon correlation do to a joint-probability value occurring in one of our latent equations? We have, from the ϕ-coefficient formula,

$$p_{xy} = p_x p_y + \phi_{xy}\sqrt{p_x q_x p_y q_y} \tag{44}$$

Suppose, to make the arithmetic easy, we assume that the valid positive rates are .80 for both indicators (i.e., not much better than single MMPI scales do against conventional manifest criteria). Then

$$p_{xy} = p_x^2 + \phi_{xy} p_x q_x \tag{45}$$

and the proportional error in approximating the joint probability $p_{xy}$ by the independence assumption $p_{xy} = p_x\, p_y$ is

$$\frac{\text{Error}}{\text{TrueValue}} = \frac{-\phi_{xy} p_x q_x}{p_x^2 + \phi_{xy} p_x q_x} \tag{46}$$

$$= -\frac{(.25)(.8)(.2)}{(.8)^2 + (.25)(.8)(.2)} = .059$$

only a six percent error, in spite of the intra-taxon $\underline{r} = .39$.

If the base-rate asymmetry P « Q is too great, the frequency function $f_s(w)$ will be less than $f_n(w)$ throughout the range, hence no hitmax cut exists. Such a situation should be manifested by the output quantity $cov_w(xy)$ increasing monotonically to the upper tail of w where frequencies are too small for stability. But it is obviously foolish even to try the method in a population with extreme base-rate asymmetry. Thus it would be pointless to study schizotypy in a college population not otherwise selected, since the gene-frequency can hardly be above .10 at the outside (a "clinical penetrance" of less than 10%.)

The assumption that an indicator is "everywhere valid" (i.e., $p_s(x) > p_n(x)$ for all realized x) may at times be incorrect. Appearance of stable, cross-validating multiple maxima in the hitmax-location stage should lead to suspicion that reversals in validity-direction are latently present.

## 14. Alternative methods

Several alternative paths to estimating the same latent quantities are available as partially independent checks on the results reached by the sequence described in Sections 3-7 preceding. For reasons of space, and since the sequence described above is essentially self-contained, detailed development of these alternative methods will be reserved for a subsequent publication, and only the essential ideas presented here:

a. Once the hitnax interval $w_h$. has been located, the median cut on another indicator x for the sub-sample lying within $w_h$ is the x-cut that equates the latent valid rates $p_{sx}$ and $q_{nx}$. This procedure, if N is very large, provides an alternative to the method of Section 4.

b. Given an arbitrary x-cut, there is a system of six equations in six unknowns relating (1) the y-sums above and below the x-cut, (2) the z-sums above and below the x-cut, and (3) the corresponding frequencies $N_{ax}$ and $N_{bx}$, to the latent quantities $H_{sx}$, $H_{nx}$, $\overline{y}_s$, $\overline{y}_n$, $\overline{z}_s$, $\overline{z}_n$, and P. Solution of this system then gives us the hit-rates on x, the base-rate, and the latent means on the other two indicators.

c. If four indicators are available and independent within taxa, we can bypass location of the hitmax cuts or the latent means. Instead, we can study the overlap of two sub-samples chosen by a symmetry criterion. Using indicators x and y, we locate an x-cut such that the y-means for cases above and below this x-cut are symmetrical about the grand y-mean. We do the same thing using the other pair of indicators, u and v. Then the number of cases shared between the

sample above $x_c$ and the sample above $\mu_c$ will be a simple function of the latent quantities. This method demands four independent indicators and we must solve a system of three quadratics.

d. With a very large sample, there should be an extreme "tail" region on an indicator w which is sufficiently far out that the intervals in this region consist solely of schizotypes (or nonschizotypes at the extreme low end.) On the independence assumption, the mean observed value of another indicator x among these extreme, "unmixed" cases is an estimate of the latent schizotype mean $\bar{x}_s$. A test of the independence assumption is that the regression line of y on x should be flat when computed on the sample of cases in the region where the x-means and the y-means cease to change with increments in w. Even if no really flat region occurs, it may be possible to extrapolate a fitted function (defined as a sun of two sufficiently general latent frequency functions) and to estimate $\bar{x}_s$ as its asymptote.

If four indicators are available we can estimate the grand base-frequency $N_s$ [=NP] from the overlap of cases in samples defined as lying above a y-cut (chosen to "flatten" output indicator u) and above an x-cut (chosen to "flatten" output indicator v), the formula being

$$N_s = \frac{n_{ax} n_{ay}}{n_{a(xy)}} \tag{47}$$

e. Finally, one can express the plus-rates on x, y, z (three equations), the joint plus-rates by pairs (three more), and the triple-plus rate (one more) in terms of the latent quantities, at any triplet of cuts $x_c$, $y_c$, $z_c$. We have the following at each (x, y, z)-point:

$$Pp_{sx} + (1-P)p_{nx} = p_x^+$$

$$Pp_{sy} + (1-P)p_{ny} = p_y^+$$

$$Pp_{sz} + (1-P)p_{nz} = p_z^+$$

$$Pp_{sx}p_{sy} + (1-P)p_{nx}p_{ny} = p_{x,y}^+$$

$$Pp_{sx}p_{sz} + (1-P)p_{nx}p_{nz} = p_{x,z}^+$$

$$Pp_{sy}p_{sz} + (1-P)p_{ny}p_{nz} = p_{y,z}^+$$

$$Pp_{sx}p_{sy}p_{sz} + (1-P)p_{nx}p_{ny}p_{nz} = p_{x,y,z}^+$$

This system (quadratics, cubics, and a quartic!) could be solved repeatedly at successive arbitrary cuts to yield the latent values. But such an approach would involve solving the 7-equation quartic system at 1000 cut-combinations even if cuts were taken at a coarseness of half-sigma steps, and can hardly be advocated for practical use[*]. However, with a mammoth N and plenty of computer time, it would be interesting to compare its results with the method proposed.

[*] I am indebted to Dr. David Lykken for a critical suggestion on this approach, and to Dr. Richard McHugh for pointing out that it is merely a generalization of Lazarsfeld's latent structure analysis (Lazarsfeld, 1959) to the continuous indicator case, in which we solve Lazarsfeld's accounting equations repeatedly at many cut-combinations.

## **Appendix

The system of quadratics at the end of Section 5 is solved as follows:

(48) $v + x + y - vx - vy - xy = K_1 = 1 - p_I$

(49) $x - vx + vy - xy = K_2 = p_{II}$

(50) $v - vx - vy + xy = K_3 = p_{III}$

where $p_I$, $p_{II}$, and $p_{III}$ are the observed probabilities of cases lying in the $(x^+ y^+)$ , $(x^- y^+)$, and $(x^- y^-)$ quadrants, respectively.

Subtracting (49) from (48),

(51) $v + y - 2vy = K_1 - K_2 = A$

Subtracting (50) from (48),

(52) $x + y - 2xy = K_1 - K_3 = B$

And adding (49) to (50)

(53) $v + x - 2vx = K_2 + K_3 = C$

Solving (51) for y,

(54) $$y = \frac{A - v}{1 - 2v}$$

Solving (53) for x,

(55) $$x = \frac{C - v}{1 - 2v}$$

Substituting (54) and (55) in (52),

(56) $$\frac{C - v}{1 - 2v} + \frac{A - v}{1 - 2v} - 2\frac{C - v}{1 - 2v}\frac{A - v}{1 - 2v} = B$$

which simplifies to a quadratic in v

(57) $(2 - 4B)v^2 - (2 - 4B)v + (A + C - 2AC - B) = 0$

When the constants A, B, C are expressed in terms of the K's and these replaced by $1 - p_I$, $p_{II}$, and $p_{III}$ and the result divided by 2, we have

(58) $(p_{II} + p_{III} - p_I - p_{IV})v^2 - (p_{II} + p_{III} - p_I - p_{IV})v + (p_{II}^2 + p_I p_{II} + p_I p_{III} + p_{II} p_{III} - p_{II}) = 0$

---

** **Appendix has been superseded. See note on title page.**

The coefficients of $v^2$ and v are equal, and a simple function of the quadrant probabilities. The constant term simplifies further, thus,

$$\begin{aligned} p_{II}^2 + p_I p_{II} + p_I p_{III} + p_{II} p_{III} - p_{II} &= p_{II}(p_{II} - 1) + p_{II}(p_I + p_{III}) + p_I p_{III} \\ &= p_{II}(p_I + p_{II} + p_{III} - 1) + p_I p_{III} \\ &= p_{II}(-p_{IV}) + p_I p_{III} \\ &= p_I p_{III} - p_{II} p_{IV} \end{aligned} \tag{59}$$

which we recognize as proportional to the phi-coefficient of this 4-fold table.

So, computationally, our actual procedure after locating the two cuts that determine this xy-table, is to solve for v in

$$(p_{II} + p_{III} - p_I - p_{IV})v^2 - (P_{II} + P_{III} - p_I - p_{IV})v + (p_I p_{III} - p_{II} p_{IV}) = 0 \tag{60}$$

choosing the root which lies between zero and one. We then substitute this value in (54) and (55) to obtain

$$y = \frac{p_{III} + p_{IV} - v}{1 - 2v} \tag{61}$$

$$x = \frac{p_{II} + p_{III} - v}{1 - 2v} \tag{62}$$

(I am indebted to Professor William D. Munro of the University of Minnesota Numerical Analysis Center, and to Dr. Richmond E. McQuistan of Honeywell Research Center, for advice as to the handling of these equations. I am also grateful to Professor Robert C. Cameron, Dr. Richard Darlington, and Dr. James Kiefer who severally deflected me from wasting time trying to prove an invalid theorem concerning hitmax location (Section 3), by their cooking up of strong counter-examples.)

## References


Cattell, R.B. (1957) *Personality and motivation structure and measurement*, Yonkers-on-Hudson: World Book Company.

Cronbach, L.J. and Meehl, P.E. (1955) Construct validity in psychological tests. *Psychol. Bull.*, *52*, 281-302.

Dawes, R.M. and Meehl, P.E. (1966) Mixed group validation: a method for determining the validity of diagnostic signs without using criterion groups. *Psychological Bulletin*, *66*, 63-67. Reprinted in Meehl, *Psychodiagnosis: Selected papers* (pp. 156-164). Minneapolis: University of Minnesota Press, 1973.

Lazarsfeld, P. (1959) Latent structure analysis. In Koch, S. (Ed) *Psychology: a study of a science*, Vol.3, *Formulations of the person and the social context* (pp. 476-543). New York: McGraw-Hill.

Meehl, P.E. (1959) Some ruminations on the validation of clinical procedures. *Canad. J. Psychol.*, *13*, 102-128.

Meehl, P.E. (1962) Schizotaxia, schizotypy, schizophrenia. *Amer. Psychologist., 17*, 827-838.

Meehl, P.E. (1964) *Manual for use with checklist of schizotypic signs*. Minneapolis: Psychiatric Research Unit, University of Minnesota Medical School.

Meehl, P.E. and Rosen, A. (1955) Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, *52*, 194-216.

Pap. A. (1953) Reduction sentences and open concepts. *Methodos*, *5*, 3-30.

Peppard, T.A. (1949) Mistakes in diagnosis. *Minnesota Medicine*, *32*, 510-511.


**Subsequent publications by Paul Meehl concerning taxometrics:**


Meehl, P. E. (1968). *Detecting latent clinical taxa, II: A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems* (Report No. PR-68-4). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Meehl, P. E., Lykken, D. T., Burdick, M. R., & Schoener, G. R. (1969). *Identifying latent clinical taxa, III. An empirical trial of the normal single-indicator method, using MMPI Scale 5 to identify the sexes* (Report No. PR-69-1). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1973). *Detecting latent clinical taxa, IV: Empirical study of the maximum covariance method and the normal minimum chi-square method, using three MMPI keys to identify the sexes* (Report No. PR-73-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1973). *Detecting latent clinical taxa, V: A Monte Carlo study of the maximum covariance method and associated consistency tests* (Report No. PR-73-3). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

— (1973). MAXCOV-HITMAX: A taxonomic search method for loose genetic syndromes. In Meehl, *Psychodiagnosis: Selected papers* (pp. 200-224). Minneapolis: University of Minnesota Press.

Golden, R., Tyan, S., & Meehl, P. E. (1974). *Detecting latent clinical taxa, VI: Analytical development and empirical trials of the consistency hurdles theory* (Report No. PR-74-4). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., Tyan, S., & Meehl, P. E. (1974). *Detecting latent clinical taxa, VII: Maximum likelihood solution and empirical and artificial data trials of the multi-indicator multi-taxonomic class normal theory* (Report No. PR-74-5). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1974). *Detecting latent clinical taxa, VIII: A preliminary study in the detection of the schizoid taxon using MMPI items as indicators* (Report No. PR-74-6). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., Tyan, S., & Meehl, P. E. (1974). *Detecting latent clinical taxa, IX: A Monte Carlo method for testing taxometric theories* (Report No. PR-74-7). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.

Golden, R., & Meehl, P. E. (1978). Testing a single dominant gene theory without an accepted criterion variable. *Annals of Human Genetics London*, *41*, 507-514.

Golden, R., & Meehl, P. E. (1979). Detection of the schizoid taxon with MMPI indicators. *Journal of Abnormal Psychology*, *88*, 217-233.

— (1979). A funny thing happened to us on the way to the latent entities. *Journal of Personality Assessment*, *43*, 563-581.

Golden, R., & Meehl, P. E. (1980). Detection of biological sex: An empirical test of cluster methods. *Multivariate Behavioral Research*, *15*, 475-496.

Meehl, P. E., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127-181). New York: Wiley.

— (1986). Diagnostic taxa as open concepts: Metatheoretical and statistical questions about reliability and construct validity in the grand strategy of nosological revision. In T. Millon & G. L. Klerman (Eds.), *Contemporary directions in psychopathology* (pp. 215-231). New York: Guilford.

— (1992) Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, *60*, 117-174. Reprinted in *A Paul Meehl Reader: Essays on the practice of scientific psychology* (pp. 331-370). (N.G. Waller, L.J. Yonce, W.M. Grove, D. Faust, & M.F. Lenzenweger, Eds.). Mahwah, NJ: Erlbaum, 2006.

Grove, W. M., & Meehl, P. E. (1993) Simple regression-based procedures for taxometric investigations. *Psychological Reports*, *73*, 707-737.

Meehl, P. E., & Yonce, L. J. (1994) Taxometric analysis: I. Detecting taxonicity with two quantitative indicators using means above and below a sliding cut (MAMBAC procedure). *Psychological Reports*, *74*, 1059-1274.

— (1995) Bootstraps taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266-275. Reprinted in *A Paul Meehl Reader: Essays on the practice of scientific psychology* (pp. 371-387). (N.G. Waller, L.J. Yonce, W.M. Grove, D. Faust, & M.F. Lenzenweger, Eds.). Mahwah, NJ: Erlbaum, 2006.

— (1995) Extension of the MAXCOV-HITMAX taxometric procedure to situations of sizeable nuisance covariance. In D. Lubinski and R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 81-92). Palo Alto, CA: Davies-Black.

Meehl, P. E., & Yonce, L. J. (1996) Taxometric analysis: II. Detecting taxonicity using covariance of two quantitative indicators in successive intervals of a third indicator (MAXCOV procedure). Psychological Reports, 78, 1091-1227.

— (1996) MAXCOV pseudotaxonicity. *American Psychologist*, *51*, 1184-1186. [Reply to Miller]

Waller, N. G., & Meehl, P. E. (1998) *Multivariate taxometric procedures: Distinguishing types from continua*. Newbury Park, CA: Sage.

— (1999) Clarifications about taxometric method. *Journal of Applied and Preventive Psychology*, *8*, 165-174. Reprinted in *A Paul Meehl Reader: Essays on the practice of scientific psychology* (pp. 389-404). (N.G. Waller, L.J. Yonce, W.M. Grove, D. Faust, & M.F. Lenzenweger, Eds.). Mahwah, NJ: Erlbaum, 2006.

— (2001) Comorbidity and taxometrics. *Clinical Psychology: Science and Practice*, *8*, 507-519.

— (2004) What's in a taxon? *Journal of Abnormal Psychology*, *113*, 39-43.

pdf by ljy September 2010