

## THE VALIDITY OF THE SUCCESSIVE INTERVALS METHOD OF PSYCHOMETRIC SCALING\*

WILLIAM W. ROZEBOOM AND LYLE V. JONES  
UNIVERSITY OF CHICAGO

The degree to which scale values computed by the method of successive intervals diverge from theoretically "true" values is seen to be due to three types of error: error due to inequalities in variances of the distributions from which the scale values are computed, error due to non-normality of the distributions, and sampling error. The contribution of each type of error to the total error is evaluated; the latter is seen to be surprisingly small under appropriate conditions. Certain aspects of the formal methodology underlying scaling procedures are also briefly considered.

One of the most popular and perhaps the simplest of all methods by which stimuli can be assigned values for some psychological variable is the rating scale technique. Basically, a rating scale is some set of categories that partition sets of events into mutually exclusive classes. For example, a rating scale might be defined by the categories *high*, *medium*, and *low*, and a set of events generated by "the evaluation of the esthetic value of art object  $i$  by judge  $j$ ," where  $i$  and  $j$  range over specified classes of art objects and judges. That is, each judge  $j$  assigns each art object  $i$  to a category of the rating scale, such an assignment constituting an event. Corresponding to each event designated by the coordinate pair  $i, j$  there is one and only one category: *high*, *medium*, or *low*.

Usually, the localization of each event on the scale is only a means to a representation of various subclasses of the events by a single value of the scale. This can be done by taking the most representative scale value of the distribution of scores in a subclass as the scale value for the subclass as a whole. Thus, for the example already given, we may be less concerned with the rating given to a particular art object by a given judge than we are in a rating representative of the values assigned to that object by the various judges. Since a specific art object defines a subclass of ratings, the most representative rating (however defined) can be taken as the value of this stimulus on the esthetic scale, not dependent upon any particular judge.

In its most elementary form, a rating scale imparts no measure of quantity to the events rated by it, merely being comprised of a set of mutually exclusive

\*This paper reports research undertaken in cooperation with the Quartermaster Food and Container Institute for the Armed Forces, and has been assigned number 475 in the series of papers approved for publication. The views or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the views or indorsement of the Department of Defense.

categories. An example of this simplest type is the standard color chart by which colors are classified. Stevens (9) has named this kind of scale a *nominal* scale. With appropriate additional data, assumptions, or definitions, however, the rating scale can be utilized as an ordinal, or even an interval scale.

If a relation can be obtained which orders the categories, then the rating scale has become an ordinal scale for that relation. One of the more customary ordering relations employed by psychologists in generating ordinal scales is that of preference, or choice. If the categories are such that, for a given judge, (i) an item assigned to category *A* is always chosen over any item assigned to category *B* (at least at the time of the assignment) and any item assigned to category *B* is always chosen over an item assigned to category *C*, and (ii) no item assigned to *C* is ever chosen over items assigned to *A*, then categories *A*, *B*, and *C* are ordered by the relation of preference.

It is customary at this point either to define or to hypothesize the existence of a psychological continuum underlying the categories of the rating scale, such that each category covers a range of the continuum, these ranges being exhaustive, mutually exclusive, and in the same ordinal relation as the corresponding categories. In short, the rating scale is interpreted as a gross technique by which the values of events are estimated on a similar, but much more discriminating underlying scale. Thus, art objects evaluated in terms of a three-category scale are assumed to be much more finely dis-

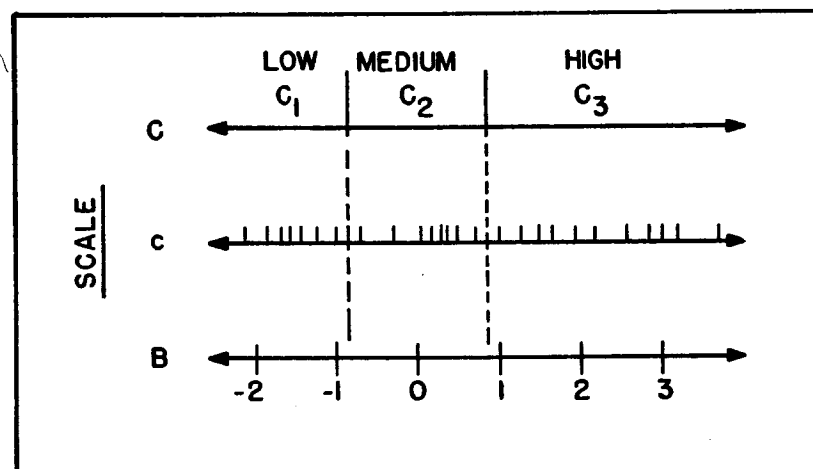


FIGURE 1

Conversion of a Sequence of Ordered Categories into an Interval Scale. Scale *C* Represents the Observed Rating Categories; Scale *c*, the Assumed Underlying Continuum; and Scale *B*, the Metric (with Arbitrary Origin and Unit Distance) Assigned to Scale *c*.

tinguishable esthetically. This is illustrated in Figure 1. Scale *C* is comprised of the categories  $C_1$ ,  $C_2$ ,  $C_3$ . Events falling in a category  $C_i$  are ordered as higher or lower than events in  $C_j$  ( $i \neq j$ ) for the property being rated, but no distinction is made among the events falling in  $C_i$ . Scale *c* is the continuum which is hypothesized or defined to underlie scale *C*, the smaller categories

indicating much finer differences in degree of the rated property. Strictly speaking, the underlying scale need not be an actual continuum; it may be conceptualized as a finite number of subdivisions of each category of the rating scale, as long as these subcategories are ordered by the same relation that orders the rating scale.

If numbers can now be assigned to the various positions on the underlying scale in such a manner that an interpretation can be defined, discovered, or assumed for the relative differences between positions on the scale, the theoretical underlying scale becomes an interval scale indicated in Figure 1 by scale *B*. Then the assignment of an event to a category of *C* is interpreted as an estimation of the score of the event on the underlying scale *B*. We shall refer to theoretical metrics such as *B*, defined or inferred from cruder empirical measures, as *base scales*. It has been customary to define the base scale (more rigorously, a set of base scales, linear transformations of one another) for a particular set of categories and distributions over the categories as that assignment of numbers to the continuum which normalizes the distributions (10). There may be other equally valid ways of defining the base scale, e.g., the counting of just-noticeable-differences, or defining the base scale so as to normalize distributions other than the ones being dealt with in the given study. It is not always possible, given more than one distribution over the same continuum, to find a numbering of the continuum that simultaneously normalizes all the distributions. Although the method of successive intervals, as described in the literature, has assumed normality for all distributions used in the analysis, we shall demonstrate that the validity of the method as a computational technique need not assume normal distributions.

Once the existence of a base scale has been defined over the categories of a rating scale, each event classified by the scale is considered to have a value on the base scale. Since each category corresponds to an interval of the base scale, the assignment of an event to a specific category determines a range in which its base scale value falls. If, now, the shape of a distribution of scores on the base scale is known, values for the widths of the various category intervals can be computed in terms of the standard deviation of that distribution as the unit of measurement. These are computed by tabulating the cumulative proportion of scores at each boundary of the interval and calculating the width in sigmas corresponding to such a percentile difference for that type of distribution. If the distribution is known or assumed to be normal, then the interval width will be the standard deviation of the distribution multiplied by the difference between the normal deviates corresponding to the cumulative proportions at the lower and upper boundaries of the interval. If a number of distributions are available, i.e., a group of judges rates a set of stimuli so that each stimulus determines a class of ratings, a number of measures of each interval width will be obtained in terms of the sigmas of the various distributions. If these are pooled,

estimates of the interval widths in terms of a common unit of measurement are obtained. Finally, if the median of a distribution be taken as the base scale score representing the distribution, the exact base scale value of this point can be estimated as follows: observe the cumulative proportion at the boundaries of the interval in which the median falls, compute from the assumed distribution function the proportion of the distance from the lower boundary, multiply this proportion by the interval width, and add the product to the base scale value for the lower boundary.

This and similar techniques for conversion of distributions of scores of a set of rating scale categories into points along an interval scale have been variously described in the literature, most frequently under the title, *the method of successive intervals* (1, 2, 4, 5, 6, 7, 8). The general computational steps usually given for evaluation of the interval widths are: (i) for each distribution, compute each interval width in terms of the sigma of that distribution by taking the difference between the normal deviates corresponding to the boundaries of the interval, assuming each distribution to be normal; (ii) let the average value of the computed widths for a given interval be taken as the best estimate of the width of that interval in terms of a unit of measurement common to all intervals. When the cumulative proportion at a boundary of an interval is nearly 0 or 1, the estimate of interval width given by that distribution for the interval is too unreliable for use, so the average width for an interval must be a weighted average; the weights of 0 or 1 have been employed in all past applications of the method.

Previous advocates of the method of successive intervals have attempted to validate the technique by demonstrating its extremely high correlation with the method of paired comparisons (8), and its internal consistency (3). It is our present aim to evaluate the method in terms of the degree to which results of the computations from empirical data can be expected to diverge from theoretically "true" values as determined from the definition of the base scale. That is, we propose to evaluate the *absolute* validity of the method.

The primary scores which are determined by the method of successive intervals are the widths of the category intervals relative to some arbitrary unit of measurement. The location of medians of the various distributions is secondary to estimation of the interval widths, since once the latter are known the former are easily determined. It is obvious that (i) if all the distributions used to measure the interval widths have equal variances, (ii) if all the distributions are normal, and (iii) if there are no sampling errors, then the computed values of relative interval widths are identical with the theoretical values. For, if each experimentally obtained distribution were normal, and for every distribution the proportion of cases falling within each interval showed no sampling errors, then the interval widths computed from a given distribution would be identical with the theoretical values as measured by the standard deviation of that distribution. If the variances

of all the distributions were equal, then each distribution would give the same computed value for a given interval width. Thus there are three possible sources of error in computation of base scale values by means of the method of successive intervals: type (a) errors due to unequal variances of the distributions used to compute the interval widths, type (b) errors due to non-normality of the distributions, and type (c) sampling errors, i.e., errors due to the estimation of cumulative proportions of the interval boundaries from finite samples of the measuring distributions.

As a tool for evaluation of the contributions of these sources of error to a total error of estimate of relative interval width, it is convenient to define a *coefficient of error*. Let a quantity  $\chi$  be estimated by a quantity  $X$ . Then the coefficient of error,  $\xi$ , for the estimation of  $\chi$  by  $X$  is  $\xi = (X - \chi)/\chi$  or  $X = (1 + \xi)\chi$ . The magnitude of the coefficient of error gives the discrepancy between  $X$  and  $\chi$  as a proportion of  $\chi$  and is nothing more than 1/100 of the percentage error in the approximation of  $\chi$  by  $X$ .

Relative interval widths computed by the method of successive intervals are estimates of true *relative* interval widths on the base scale. By relative interval widths, we recognize that the unit of measurement is arbitrary, so that the ratio of one interval width to another (which is invariant under transformation of the unit of measurement) is the critical quantity by which relative interval width is expressed. We can evaluate the coefficient of error for the estimation of true interval width ratios from computed ratios as follows: (i) find an expression for the computed interval widths,  $L_j$  and  $L_k$ , for categories  $j$  and  $k$ , in terms of the three types of errors that influence the computed widths, and of the true interval widths,  $\lambda_j$  and  $\lambda_k$ ; (ii) set

$$L_j/L_k = (1 + \xi_{jk})(\lambda_j/\lambda_k). \quad (1)$$

Solving for  $\xi_{jk}$ , we obtain the coefficient of error for the estimation of relative interval widths by the method of successive intervals as a function of the different types of error; we shall be able to see explicitly the manner and extent to which each kind of error contributes to the total error.

Let  $\lambda_j$  be the true width on the base scale of an interval  $j$  in terms of some arbitrary unit of measurement  $U$ , and  $L_j$  the width of the interval as computed by the method of successive intervals. Let  $\eta_i$  (measured in terms of  $U$ ) be the standard deviation of the  $i$ th distribution over the base scale. If the  $i$ th distribution is normal and displays no sampling errors, then the cumulative proportions at the upper and lower boundaries of interval  $j$  permit an exact computation (through use of a table of the normal probability integral) of the magnitude of  $\lambda_j$  in terms of  $\eta_i$  as a unit of measurement. Specifically,

$$l_{ij} = \lambda_j/\eta_i = v_i\lambda_j,$$

where  $l_{ij}$  is the width of interval  $j$  as computed from distribution  $i$ ;  $\lambda_j$  and  $\eta_i$  are the true magnitudes of interval  $j$  and the standard deviation of distri-

bution  $i$ , respectively, in terms of the arbitrary unit of measurement  $U$ ; and  $\nu_i = 1/\eta_i$ . However, to the extent that distribution  $i$  diverges from normality and contains sampling errors,  $l_{ij}$  will differ from  $\nu_i\lambda_j$ . In general,  $l_{ij} = \nu_i\lambda_j + \zeta_{ij}$ , where  $\zeta_{ij}$  is the discrepancy between the computed  $l_{ij}$  and the theoretical  $\nu_i\lambda_j$ .  $\zeta_{ij}$  can be analyzed into two additive components  $E_{ij}$  and  $\epsilon_{ij}$ , where  $E_{ij}$  is a *constant* bias due to non-normality of the distribution and  $\epsilon_{ij}$  is a random sampling error. Thus,

$$l_{ij} = \nu_i\lambda_j + E_{ij} + \epsilon_{ij}. \quad (2)$$

It should be noted that the unit of measurement for the error terms  $E_{ij}$  and  $\epsilon_{ij}$  is  $\eta_i$ , the standard deviation of the  $i$ th distribution; while the unit of measurement for  $1/\nu_i$  and  $\lambda_j$  is the arbitrary  $U$ , which is the same for all distributions.

The computed width  $L_j$  of interval  $j$  is a weighted average of the estimates of widths contributed by the various distributions. That is,

$$L_j = \sum_i w_{ij}l_{ij} = \lambda_j \sum_i w_{ij}\nu_i + \sum_i w_{ij}E_{ij} + \sum_i w_{ij}\epsilon_{ij}, \quad (3)$$

where  $\sum_i w_{ij} = 1$ . Defining the quantities  $A_j$ ,  $\beta_j$ , and  $\gamma_j$  by

$$A_j = \sum_i w_{ij}\nu_i, \quad (4)$$

$$\beta_j = (\sum_i w_{ij}E_{ij})/(A_j\lambda_j), \quad (5)$$

$$\gamma_j = (\sum_i w_{ij}\epsilon_{ij})/(A_j\lambda_j), \quad (6)$$

we obtain

$$L_j = A_j\lambda_j(1 + \beta_j + \gamma_j). \quad (7)$$

Since  $\lambda_j$  is inversely proportional to, and  $A_j$  proportional to the magnitude of the base unit of measurement  $U$ ,  $A_j\lambda_j$  is invariant for transformations of  $U$ . Therefore  $\beta_j$  and  $\gamma_j$ , which are also invariant under transformations of  $U$ , may be interpreted as error per unit length of interval due to non-normality of distribution and to sampling error, respectively.  $L_j$  may be interpreted as an estimate of  $\lambda_j$  with  $1/A_j$  as the unit of measurement. It will be seen below that  $1/A_j$  is approximately the harmonic mean of the standard deviations of the measuring distributions.

We are now able to evaluate the coefficient of error,  $\xi_{jk}$ , for the computed ratio  $L_j/L_k$  as an estimate of the true ratio  $\lambda_j/\lambda_k$  of the widths of intervals  $j$  and  $k$ . Finding  $L_k$  by substitution of  $k$  for  $j$  throughout (7) and solving for  $\xi_{jk}$  in (1) we find that

$$\xi_{jk} = \frac{A_j}{A_k} \left( \frac{1 + \beta_j + \gamma_j}{1 + \beta_k + \gamma_k} \right) - 1,$$

which may be written

$$\xi_{jk} = \alpha_{jk} \left( \frac{1 + \beta_j + \gamma_j}{1 + \beta_k + \gamma_k} \right) + \frac{\beta_j - \beta_k}{1 + \beta_k + \gamma_k} + \frac{\gamma_j - \gamma_k}{1 + \beta_k + \gamma_k}, \quad (8)$$

where

$$\alpha_{jk} = (A_j/A_k) - 1. \quad (9)$$

Since  $\alpha_{jk}$  reflects the difference between units of measurement within interval  $j$  and within interval  $k$ , and vanishes (as shown below) when the variances of all the distributions are equal,  $\alpha_{jk}$  may be regarded as the error in relative interval width due to unequal variances of the distributions which are used in estimating the interval widths. Thus, of the three sources of error in the method of successive intervals, type (a) is represented quantitatively by  $\alpha$ , type (b) by  $\beta$ , and type (c) by  $\gamma$ .

#### $\gamma$ -Error

It will be recalled from (3) that each distribution  $i$  was assigned a weight  $w_{ij}$  for its contribution  $l_{ij}$  in the computation of  $L_j$ . It is now possible to assign these weights in a manner that minimizes the sampling error  $\gamma_j$ . Assuming the various distributions to be essentially independent of one another in their sampling errors, we find from (6) the mean and variance for  $\gamma_j$  (under repeated sampling with a fixed set of weights) to be

$$\mu_{\gamma_j} = \sum_i w_{ij} \mu_{\epsilon_{ij}} / A_j \lambda_j; \quad (10)$$

$$\sigma_{\gamma_j}^2 = \sum_i w_{ij}^2 \sigma_{\epsilon_{ij}}^2 / (A_j \lambda_j)^2. \quad (11)$$

But  $\epsilon_{ij} = \delta_{U_{ij}} - \delta_{L_{ij}}$ , where  $\delta_{U_{ij}}$  and  $\delta_{L_{ij}}$  are the sampling errors for the standardized deviates of the normal probability distribution corresponding to the cumulative proportions of distribution  $i$  at the upper and the lower boundaries of interval  $j$ , and  $\delta \simeq \Delta P/y$ , where  $\Delta P$  is the sampling error of a cumulative proportion at an interval boundary and  $y$  is the ordinate of the normal probability distribution at  $P$ . Since the sampling mean of  $\Delta P$  is zero,  $\mu_{\epsilon_{ij}} \simeq 0$  and hence, from (10),

$$\mu_{\gamma_j} \simeq 0; \quad (12)$$

while

$$\begin{aligned} \sigma_{\epsilon_{ij}}^2 &= \sigma_{\delta_{U_{ij}}}^2 + \sigma_{\delta_{L_{ij}}}^2 - 2 \text{cov}(\delta_{U_{ij}}, \delta_{L_{ij}}) \\ &\simeq \frac{1}{n_i} \left[ \frac{P_{U_{ij}}(1 - P_{U_{ij}})}{y_{U_{ij}}^2} + \frac{P_{L_{ij}}(1 - P_{L_{ij}})}{y_{L_{ij}}^2} - \frac{2P_{L_{ij}}(1 - P_{U_{ij}})}{y_{L_{ij}} y_{U_{ij}}} \right], \end{aligned} \quad (13)$$

where  $n_i$  is the sample size of distribution  $i$ ,  $P_{L_{ij}}$  ( $P_{U_{ij}}$ ) is the parametric cumulative proportion of distribution  $i$  falling at the lower (upper) boundary

of interval  $j$ , and  $y_{L_{ii}}$  ( $y_{U_{ii}}$ ) is the ordinate of the normal probability distribution at  $P_{L_{ii}}$  ( $P_{U_{ii}}$ ). Since the sample size  $n_i$  is known, and the sample cumulative proportions provide sufficiently close approximations to the parametric cumulative proportions, very close approximations of  $\sigma_{\epsilon_{ii}}^2$  may be computed from empirical data by use of formula (13). Furthermore,  $\sigma_{\epsilon_{ii}}^2$  may be made as small as desired by choosing the sample size  $n_i$  sufficiently large.

The assumption of (11) might at first seem gratuitous; for many experimental situations the sampling errors of one distribution will not be strictly independent of the next. Thus, if two sample distributions are obtained from judgments for two stimuli by the same judges, the sampling errors of the two distributions would probably be correlated. However, the disturbing effects of such a lack of strict independence are vitiated by the following considerations: (i) factors linking the sampling errors of two distributions usually comprise only a small portion of the total factors determining the outcome of the observed cumulative proportions; (ii) linear correlations among the sampling errors may be negligible even when significant non-linear correlations exist; and (iii) the intercorrelations may assume both positive and negative values, so that even when their absolute magnitudes are significant their net effect may be negligible. Thus the assumption of (11) involves little loss of generality.

Since by (12) the average value of  $\gamma_i$  is approximately zero, the expected absolute magnitude of  $\gamma_i$  is less than (though on the order of)  $\sigma_{\gamma_i}$ , so the expected (absolute) size of  $\gamma_i$  will be minimal when  $\sigma_{\gamma_i}$  is minimal. By differentiation of (11) it will be found that  $\sigma_{\gamma_i}$  is minimal when, for each  $i$ ,  $w_{ij}\sigma_{\epsilon_{ii}}^2 = k_j$ , where  $k_j$  is a constant of proportionality. Since  $\sum_i w_{ij} = 1$ ,  $k_j = (\sum_i \sigma_{\epsilon_{ii}}^{-2})^{-1}$ , so

$$w_{ij} = (\sigma_{\epsilon_{ii}}^2 \sum_i \sigma_{\epsilon_{ii}}^{-2})^{-1}. \quad (14)$$

Equations (14) and (13) provide the steps for computation of the proper weights. (Except for those distributions for which the cumulative proportion at one of the boundaries of an interval is close to 0 or 1, the weights assigned to the various distributions for that interval are very similar. Hence, the customary procedure of giving zero weight to those distributions for which the sampling reliability of the interval estimate is small and of giving the remaining distributions equal weight in the computation of the interval width should be acceptable for most purposes.)

Substitution of (14) in (11) gives

$$\sigma_{\gamma_i}^2 = (A_i \lambda_i)^{-2} (\sum_i \sigma_{\epsilon_{ii}}^{-2})^{-1}. \quad (15)$$

Since  $\sigma_{\epsilon_{ii}}^2$  is usually on the order of  $1/n_i$ , letting  $n$  be the average size of the sample distributions and  $N$  the number of distributions,  $\sigma_{\gamma_i}^2$  is roughly on



the order of  $(A_i \lambda_i)^{-2} (nN)^{-1}$ . Thus the expected order of magnitude for  $\gamma_i$  is roughly  $(A_i \lambda_i)^{-1} (nN)^{-\frac{1}{2}}$ . This value may be made as small as desired by taking sufficiently large  $n$  and  $N$ . For example, if  $A_i \lambda_i = .5$ ,  $N = 50$  and  $n = 500$ , then the expected order of magnitude for  $\gamma_i$  is  $10^{-2}$ . Thus for an empirical study of any substantial proportions an expected order of magnitude for  $\gamma_i$  of  $10^{-2}$  should not be difficult to obtain. In order to maintain a fixed order of magnitude for  $\gamma_i$ , a decrease of interval widths must be compensated for by an increase in (i) the sample sizes, (ii) the number of distributions, or (iii) both. For with  $\sigma_{\gamma_i}$  held constant,  $\sqrt{nN}$  is inversely proportional to  $A_i \lambda_i$ , while the latter, as shown below, is the width of category  $j$  in units of measurement given by the harmonic mean of the standard deviations of the measuring distributions. This has direct implications for the design of rating scales, for it shows that the number of categories into which a scale can be reliably decomposed is limited by the number of stimuli and the size of the population upon which the scale is to be standardized.

Thus, if the width of an interval relative to  $1/A_i$  is not too small, and if the study by which the scale is being standardized is of reasonably substantial dimensions, the error in estimation of  $\lambda_i$  due to sampling will be insignificant—generally on the order of  $10^{-2}$ . In light of this, the mean and sampling error of  $\xi_{ik}$  can be evaluated. Any reciprocal,  $1/s_i$ , from a distribution of  $s$  with mean  $M_s$  can be replaced by the expression  $(2/M_s - s_i/M_s^2)$  with an error coefficient of  $-[(s_i - M_s)/M_s]^2$ . Since from (12), the mean of  $(1 + \beta + \gamma)$  is  $(1 + \beta)$ ,  $[1/(1 + \beta_k + \gamma_k)]$  may be replaced by  $[(1 + \beta_k - \gamma_k)/(1 + \beta_k)^2]$ , with an error coefficient of  $-[(\gamma_k)/(1 + \beta_k)]^2$ , the error of the replacement being negligible so long as  $\beta_k$  does not approach  $-1$ . With this replacement we find from (8) and (12) that the sampling mean of  $\xi_{ik}$  is

$$\mu_{\xi_{ik}} \simeq \alpha_{ik} [(1 + \beta_i)/(1 + \beta_k)] + (\beta_i - \beta_k)/(1 + \beta_k), \quad (16)$$

while, disregarding second-order terms,

$$\sigma_{\xi_{ik}} \simeq [(1 + \alpha_{ik})/(1 + \beta_k)^2] \sqrt{(1 + \beta_k)^2 \sigma_{\gamma_i}^2 + (1 + \beta_i)^2 \sigma_{\gamma_k}^2}. \quad (17)$$

#### *$\alpha$ -Error*

For evaluation of  $\alpha_{ik}$ , the error due to inequality of variances, it is convenient to employ the identity

$$\begin{aligned} A_i &= \sum_{i=1}^N w_i \nu_i = N \sigma_{w_i \sigma_{\nu}} r_{w_i \nu} + N \bar{w}_i \bar{\nu} \\ &= N \sigma_{w_i \sigma_{\nu}} r_{w_i \nu} + \bar{\nu}, \end{aligned}$$

where  $N$  is the total number of distributions,  $\sigma_{w_i}$  is the standard deviation of the weights for the  $j$ th interval over the  $N$  distributions,  $\sigma_{\nu}$  is the standard deviation of the  $\nu_i$  over the  $N$  distributions,  $r_{w_i \nu}$  is the product-moment

correlation between  $w_{ij}$  and  $\nu_i$  over the  $N$  distributions, and  $\bar{\nu}$  is the mean value of  $\nu_i$  over the  $N$  distributions. From (9), this gives for  $\alpha_{jk}$

$$\begin{aligned}\alpha_{jk} &= \frac{N\sigma_{w_j}\sigma_{\nu}r_{w_j\nu} - N\sigma_{w_k}\sigma_{\nu}r_{w_k\nu}}{N\sigma_{w_j}\sigma_{\nu}r_{w_k\nu} + \bar{\nu}} \\ &= \left( \frac{C_j r_{w_j\nu} - C_k r_{w_k\nu}}{C_k r_{w_k\nu} V_{\nu} + 1} \right) V_{\nu},\end{aligned}\quad (18)$$

where  $C_j = N\sigma_{w_j}$ ,  $C_k = N\sigma_{w_k}$ , and  $V_{\nu} = \sigma_{\nu}/\bar{\nu}$ .

The value of  $C_j$  depends only on the shape of the distribution of weights for the interval  $j$ ; this value will be of an order higher than  $10^{-1}$  only when a relatively small proportion of the distributions receive a significant weight for interval  $j$ . In the case where a proportion,  $k$ , of the distributions receive equal weights and the rest receive 0 weight,  $C = \sqrt{(1/k) - 1}$ , which exceeds 1 only when  $k < .5$  and is no larger than 3 when  $k = .1$ . The correlation,  $r_{w_j\nu}$ , between the weights assigned to the distributions for interval  $j$  and the reciprocals of the standard deviations of the distributions can be expected to assume some small negative value (with a chance divergence which vanishes as  $N$  grows large), since as  $\eta_i$  increases the boundaries of the interval draw closer to the center of the distribution, yielding an increase in  $w_{ij}$ . However, we should expect this correlation to be equal for both intervals  $j$  and  $k$ . Thus, the maximum value of  $\alpha_{jk}$  would be approximately  $V_{\nu} \times 10^{-1}$ .

But  $V_{\nu}$  is the coefficient of variation for the reciprocals of the standard deviations of the measuring distributions and is approximately equal to the coefficient of variation for the  $\eta_i$ . We shall, as a rule, expect to find  $V_{\nu}$  on the order of  $10^{-1}$ , which makes  $\alpha_{jk}$  on the order of  $10^{-2}$ . Thus only when the variances of the distributions by which the interval widths are computed differ widely among themselves is the error contributed by the inequality of variances of any significance. In such cases, the data can be reanalyzed using the correction for inequalities in variance suggested by Attneave (1).

It should be noted that if all distributions receive equal weights for two intervals  $j$  and  $k$ , then  $\alpha_{jk} = 0$ , regardless of the magnitude of  $V_{\nu}$ . Even when the  $\eta_i$  differ widely,  $\alpha_{jk}$  will be negligible if the  $w_{ij}$  and  $w_{ik}$  are sufficiently homogeneous. It should also be noted that

$$A_i = (C_i r_{w_i\nu} V_{\nu} + 1) \bar{\nu} \simeq \bar{\nu}, \quad (19)$$

where  $\bar{\nu}$  is the reciprocal of the harmonic mean of the  $\eta_i$ . This substantiates our earlier contention that the computed interval widths are expressed in units of measurement determined by the harmonic mean of the standard deviations of the measuring distributions.

#### *$\beta$ -Error*

Of the three sources of error in the method of successive intervals, evaluation of  $\beta$ -error is the most difficult. We can replace  $\sum_{i=1}^N w_{ij} E_{ij}$  by  $(N\sigma_{w_j}$

$\sigma_{E_i} r_{w_i E_i} + \bar{E}_i$ ), where  $\bar{E}_i$  and  $\sigma_{E_i}$  are the mean and standard deviation of the errors introduced into the estimation of  $\lambda_i$  by the non-normality of the  $N$  distributions. Then from (5)

$$\beta_j = (C_i \sigma_{E_i} r_{w_i E_i} + \bar{E}_i) / (A_i \lambda_i). \quad (20)$$

Note that  $E_{i,j}$  is measured in terms of the standard deviation,  $\eta_i$ , of distribution  $i$ . In particular, when  $P_{L,i,j}$  and  $P_{U,i,j}$  are the cumulative proportions of distribution  $i$  at the lower and upper boundaries of interval  $j$ ,  $E_{i,j}$  is the difference between the number of sigmas spanned between  $P_{L,i,j}$  and  $P_{U,i,j}$  by the actual distribution  $i$  and the number of sigmas spanned between  $P_{L,i,j}$  and  $P_{U,i,j}$  by a normal distribution. Let  $D_{i,j}$  be the number of sigmas spanned by distribution  $i$  between these two cumulative proportions, and let  $d_{i,j}$  be the corresponding number of sigmas spanned by a normal distribution. Then  $E_{i,j} = d_{i,j} - D_{i,j} = \omega_{i,j} D_{i,j}$ , where  $\omega_{i,j} = (d_{i,j}/D_{i,j}) - 1$  and is thus the coefficient of error for the approximation of the distance in sigmas spanned between  $P_{U,i,j}$  and  $P_{L,i,j}$  by distribution  $i$  by the corresponding distance spanned by a normal distribution. Since  $D_{i,j} = \lambda_i / \eta_i = \nu_i \lambda_i$ , (20) may be rewritten as

$$\begin{aligned} \beta_j &= (C_i r_{w_i E_i} \sigma_{\omega_i \nu_i \lambda_i} + \overline{\omega_i \nu_i \lambda_i}) / (A_i \lambda_i) \\ &= C_i r_{w_i E_i} \sigma_{\omega_i (\nu/A_i)} + \overline{\omega_i (\nu/A_i)}. \end{aligned}$$

But when  $V_i$  is small,

$$\sigma_{\omega_i (\nu/A_i)} \simeq \sigma_{\omega_i}$$

and

$$\overline{\omega_i (\nu/A_i)} \simeq \sigma_{\omega_i} V_i r_{w_i \nu} + \bar{\omega}_i.$$

Therefore,

$$\begin{aligned} \beta_j &\simeq \sigma_{\omega_i} (C_i r_{w_i E_i} + V_i r_{w_i \nu}) + \bar{\omega}_i \\ &\simeq \beta'_j + \bar{\omega}_i, \end{aligned} \quad (21)$$

where

$$\beta'_j = \sigma_{\omega_i} (C_i r_{w_i E_i} + V_i r_{w_i \nu}). \quad (22)$$

In general, while there may be some small non-linear correlation between  $w_{i,j}$  and  $E_{i,j}$ , the linear  $r_{w_i E_i}$  will be close to zero as  $N$  increases and the chance fluctuation of  $V_{w_i E_i}$  thus diminishes. A similar argument holds for  $r_{w_i \nu}$ ; because of the small expected values of  $C_i$  and  $V_i$ ,  $\beta'_j$  should be on the order of  $\sigma_{\omega_i} \times 10^{-1}$  at maximum. It will be shown below that even when a distribution is markedly non-normal, the expected order of magnitude for  $\omega_{i,j}$  is only  $10^{-1}$ , so  $\sigma_{\omega_i}$  will be on the order of  $10^{-1}$  at maximum. Thus,  $\beta'_j$  will be on the order of  $10^{-2}$  at maximum and is more likely to be of order  $10^{-3}$ .

It follows that the only likely significant component of  $\beta_i$  is  $\bar{\omega}_i$ , the latter comprising the average value of  $\omega_{ij}$  over the  $N$  distributions used to measure the width of interval  $j$ . These  $N$  distributions may be conceived as a sample of size  $N$  from an infinite population of potential distributions over the scale. Then  $\bar{\omega}_i$  has a sampling mean,  $\mu_{\bar{\omega}_i}$ , and variance,  $\sigma_{\bar{\omega}_i}^2$ , of its own. Similarly, the  $\omega_{ij}$  for the infinite potential population of distributions over interval  $j$  have a mean,  $\mu_{\omega_j}$ , and variance,  $\sigma_{\omega_j}^2$ . Finally, since  $\bar{\omega}_i$  is the mean of a sample of size  $N$  from the  $\omega_{ij}$ ,  $\mu_{\bar{\omega}_i} = \mu_{\omega_j}$  and  $\sigma_{\bar{\omega}_i}^2 \simeq \sigma_{\omega_j}^2/N_j$ . (More generally,  $\sigma_{\omega_j}^2/N \leq \sigma_{\bar{\omega}_i}^2 \leq \sigma_{\omega_j}^2$ , depending upon the extent to which the  $\omega_{ij}$  for the sample of  $N$  distributions are independent of one another. In most situations, we will expect to find that the  $\omega_{ij}$  are not wholly independent, but, for the same reasons advanced to justify equation (11), we shall expect that the sum of the covariances will be negligible.) Let  $\beta_i''$  be the extent to which  $\bar{\omega}_i$  diverges from its mean. Then

$$\bar{\omega}_i = \beta_i'' + \mu_{\omega_j}, \quad (23)$$

and thus, from (21),

$$\beta_i = \beta_i' + \beta_i'' + \mu_{\omega_j}. \quad (24)$$

Since  $\beta_i''$  is of order  $\sigma_{\omega_j}/\sqrt{N}$ , and as already mentioned, the expected order of magnitude for  $\sigma_{\omega_j}$  is  $10^{-1}$  or smaller, then if  $N$  is reasonably large the maximum expected order of magnitude for  $\beta_i''$  is  $10^{-2}$ . This leaves  $\mu_{\omega_j}$  in (21) as the only component of  $\beta_i$  likely to be significant. But  $\mu_{\omega_j}$  is merely the expected value of  $\omega_{ij}$  on the interval  $j$ . As indicated below, the absolute magnitude of  $\omega_{ij}$  is only of expected order  $10^{-1}$  even when the distributions are quite non-normal. While it is impossible to make any definite statement about the average,  $\mu_{\omega_j}$ , for an interval  $j$ , it would seem unlikely that it could exceed .10 except under cases of extreme, persistent, and positively correlated non-normalities among the population of distributions over interval  $j$ . Thus, except under unusual circumstances,  $\beta_i$  is of expected order of magnitude  $10^{-1}$  or less, and we may simplify (16) and (17) to

$$\begin{aligned} \mu_{\xi_{ik}} &\simeq \alpha_{ik} + \beta_i - \beta_k \\ &\simeq \alpha_{ik} + (\beta_i' - \beta_k') + (\beta_i'' - \beta_k'') + \mu_{\omega_j} - \mu_{\omega_k} \end{aligned} \quad (25)$$

and

$$\sigma_{\xi_{ik}} \simeq \sqrt{\sigma_{\gamma_i}^2 + \sigma_{\gamma_k}^2}. \quad (26)$$

Of the terms in (25), only  $\mu_{\omega_j}$  and  $\mu_{\omega_k}$  are of expected order larger than  $10^{-2}$ .

It yet remains to determine the anticipated order of magnitude for  $\omega$ . Since the population of potential distributions over a rating scale cannot be specified, it is impossible to assign a mathematical expectation to this term. However,  $\omega$  may be computed as a function of the degree of non-normality of the distribution being approximated. One may then select a range of distributions within which an empirically encountered distribution reasonably may be anticipated to fall and hence obtain reasonable bounds

for the magnitude of  $\omega$ . What we shall illustrate here is a technique by which a distribution of any given shape readily may be inspected for its values of  $\omega$ . By this technique, the reader may select what he considers to be fair examples of empirically anticipated non-normal distributions and easily convince himself that  $\omega$  is unlikely to be of an order greater than  $10^{-1}$ .

It will be recalled that  $\omega = (d/D) - 1$ , where  $d$  and  $D$  are the distances, measured in terms of the standard deviations of the distributions, spanned between the cumulative frequencies at the upper and lower boundaries of the interval by a normal distribution and the empirical distribution, respectively. Let  $P_U$  and  $P_L$  be the cumulative proportions at the upper and lower boundaries of the interval, let  $y(x)$  and  $Y(x)$  be the ordinates at  $x$  of the unit normal distribution and of the empirical distribution standardized to  $\sigma = 1$ , respectively, and let  $x_P$  and  $X_P$  be the distance of cumulative proportion  $P$  from the means of the unit normal distribution and the standardized empirical distribution, respectively. Then  $d = x_{P_U} - x_{P_L}$  and  $D = X_{P_U} - X_{P_L}$ . But

$$P_U - P_L = \int_{x_{P_L}}^{x_{P_U}} y(x) dx = \bar{y}d,$$

where  $\bar{y}$  is the mean value of the ordinate to the unit normal distribution over the interval. Similarly,

$$P_U - P_L = \int_{X_{P_L}}^{X_{P_U}} Y(x) dx = \bar{Y}D.$$

Hence  $d/D = \bar{Y}/\bar{y}$ , so  $\omega = (\bar{Y}/\bar{y}) - 1$  and is thus the coefficient of error for the approximation of the average height of the unit normal distribution between two cumulative proportions by the corresponding average height of the standardized empirical distribution. The magnitude of  $\omega$  is then readily seen by an inspection of the graphs of  $y$  and  $Y$  against  $P$ . That is, let  $y(P) = y(x_P)$  and  $Y(P) = Y(X_P)$ . It is computed without difficulty that  $\bar{y}(x) = \bar{y}(P)$ , where  $\bar{y}(P)$  is the harmonic mean of  $y(P)$  between  $P_L$  and  $P_U$ , and similarly  $\bar{Y}(X) = \bar{Y}(P)$ . Also, except for those intervals over which the coefficient of variation for  $y(P)$  or  $Y(P)$  is large,  $\bar{y}(P) \simeq \bar{y}(P)$  and  $\bar{Y}(P) \simeq \bar{Y}(P)$ . Thus, given any empirical distribution, the magnitude of the approximation error can be determined readily by standardizing the distribution to unit variance, graphing the height of the distribution against cumulative proportion, and superimposing the corresponding graph of the unit normal distribution. One may then select two cumulative proportions, estimate the average difference between the curves over the interval visually, and divide this by the estimated average ordinate of the normal distribution over the interval.

We illustrate the method through its application to two arbitrary distributions, a rectangular distribution and a triangular distribution skewed

so that the projection of the apex divides the base in a ratio of 1:3. These are shown with unit variance in Figure 2, together with the unit normal distribution by which they are to be approximated. Both distributions represent departures from the normal that, in an empirical distribution, would be considered severe. Figure 3 shows the same distributions in terms

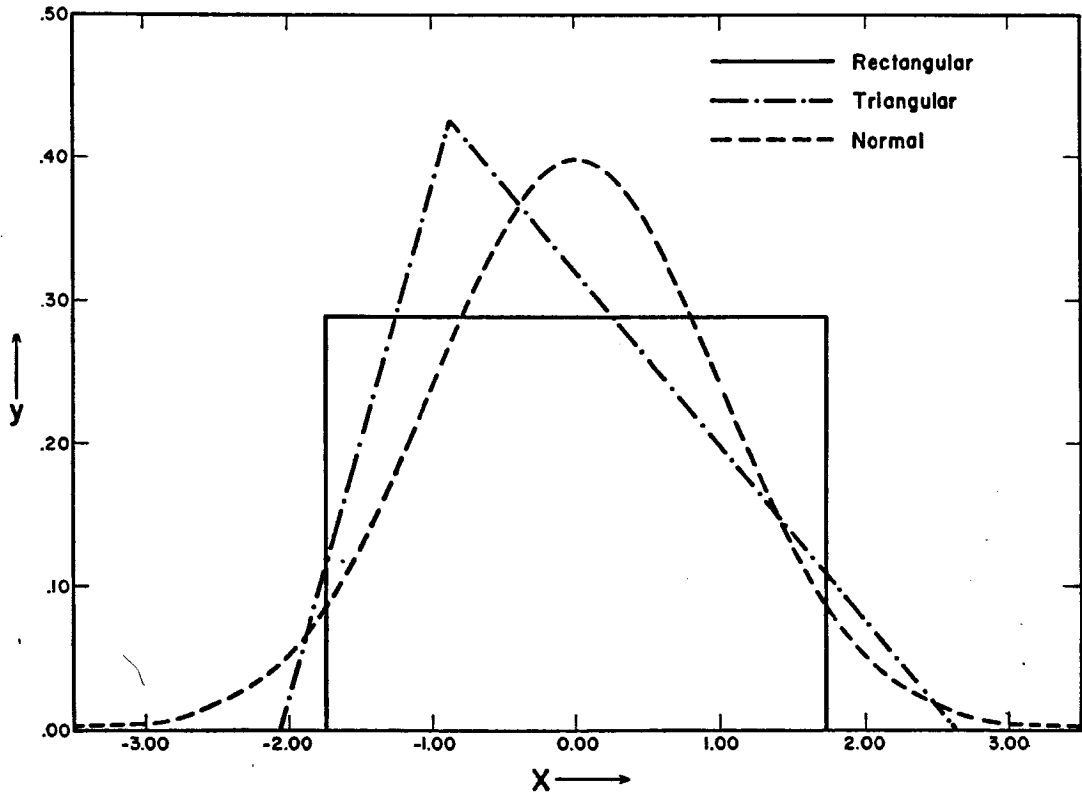


FIGURE 2

Normal, Rectangular, and Triangular Probability Distributions for Which  $\mu = 0$ ,  $\sigma = 1$ .  
Projection of Triangular Apex Divides the Base into 1:3 Ratio.

of the ordinates of Figure 2 plotted against the corresponding cumulative proportions. If various pairs of cumulative proportions are selected and  $\omega$  estimated, it is seen that  $|\omega|$  has a modal value in the range .25 to .30 for the two distributions and grows much larger than this only when one of the proportions approaches 0 or 1 (due, here, to the finite ranges of both illustrative distributions). This is typical for most distributions;  $\omega$  is likely to exceed the order of  $10^{-1}$  only when one of the cumulative proportions at an interval boundary approaches the upper or lower limit. But it is precisely in this case that the error variance of a proportion obtained through finite sampling becomes so large as to give negligible weight to the contribution to the total estimate by an interval width estimate based on such a proportion.

In those few cases where an empirical distribution is likely to show large approximation errors (such as the case of multimodal distribution in which

the modes are well separated and the intervening troughs deep) the severe non-normality of the distribution should be painfully apparent when the distribution is plotted on the successive intervals scale as finally computed. The non-normal distribution then may be discarded and a new analysis of the remaining data performed if the investigator sees fit.

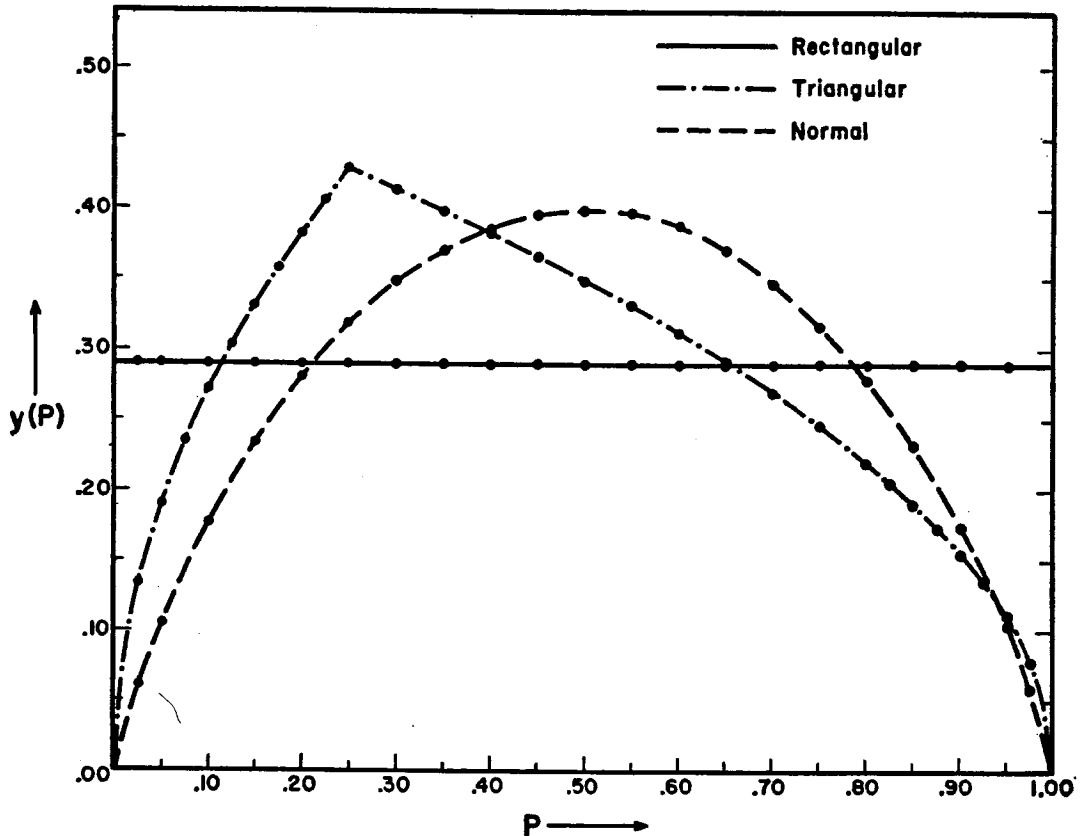


FIGURE 3  
Ordinates of the Distributions of Figure 2 as Function of Cumulative Proportions.

#### *$\beta$ -Error and Its Relation to the Base Scale*

So far we have found it unnecessary to make any comments concerning the base scale which supposedly underlies the rating scale except to hypothesize its existence.

Necessary and sufficient conditions for the existence of an *interval* base scale underlying a successive interval rating scale are: (i) There must (potentially) exist a numbering of all the potentially infinite number of events classifiable by the rating scale such that there is no overlap, for any two categories of the rating scale, of the ranges of the numbers corresponding to the events falling within each category. (ii) The positions of the ranges corresponding to the various categories must be in the same ordinal relation as are the categories. (iii) For any set of events so numbered, there must exist some interpretation of (a) the ordinal relations among the numbers

assigned to members of the set and of (b) the ratio of the difference between the numbers of any pair of the set to the difference between the numbers of any other pair, in terms of some properties of the events of the set. (If the base scale provides interpretations for additional properties of the numbers assigned to events, it may become a *ratio* or even an *absolute* scale.) There may be many different numberings satisfying conditions (i) and (ii), and many different interpretations in accordance with condition (iii). Hence there may be many different base scales underlying a given successive intervals scale. In fact, any assignment of numbers satisfying conditions (i) and (ii) is a potential base scale for the rating scale since we can never know for certain that there exists no interpretation of a numbering in conformance with condition (iii). Such potential base scales for a given rating scale need be correlated only to the extent that the values of a given event on the various potential base scales must all fall within ranges corresponding to the same rating scale category. In particular, any order-preserving transformation of a potential base scale is also a potential base scale.

The essential result of a method of successive intervals analysis is the derivation of a set of numbers corresponding to the boundaries of the intervals of the rating scale; these numbers, when paired and the ratio of differences between members of pairs taken, give the ratio of the base scale intervals corresponding to these pairs. The ratio of intervals for a potential base scale is always the same as the corresponding ratio for any *linear* transformation of that scale. However, this is not uniformly true for any other transformation. Let all potential base scales be separated into classes, any member of a given class being a linear transformation of any other member of that class. These classes are, in general, characterized by different values for the ratio of two intervals corresponding to two pairs of points on the rating scale; the classes of potential base scales most closely approximated by the scale computed through the successive intervals technique will be those classes whose ratios for interval widths are most similar to the ratios displayed by the computed scale. That is, the classes of potential base scales most closely approximated by the computed scale are the classes of potential base scales minimizing the  $\xi_{jk}$ .

Since the exact magnitudes of the  $\xi_{jk}$  are unknown in applications of the method of successive intervals, it is impossible to determine the class of potential base scales most closely approximated in a specific instance. The classes of potential base scales *most likely* to minimize the  $\xi_{jk}$ , however, are those which minimize the *expected* values of the  $\xi_{jk}$ . Now, the data of a specific successive intervals analysis are obtained by sampling of two kinds: a sample of size  $N$  from possible distributions over the rating scale, and a sample of  $n_i$  individuals from each distribution  $i$  ( $i = 1, 2, \dots, N$ ). The expected value of  $\xi_{jk}$  for a specific sample of distributions is given by (25). But the terms  $\alpha_{jk}$ ,  $(\beta'_i - \beta'_k)$ ,  $\beta'_i$ ,  $\beta'_k$  are dependent upon the specific sample of



distributions chosen;  $\beta_i''$  and  $\beta_k''$ , by definition, have an expected value of 0, while both  $\alpha_{jk}$  and  $(\beta_j' - \beta_k')$  are determined essentially by differences of the form  $X_j r_{y_i z_i} - X_k r_{y_k z_k}$ , where  $X$ ,  $y$ , and  $z$  are various specified properties of the  $N$  distributions. These differences should be negative as often as positive, so that the expected values of  $\alpha_{jk}$  and  $(\beta_j' - \beta_k')$  should be 0. Thus, the expected value of  $\xi_{jk}$  is approximately  $\mu_{\omega_j} - \mu_{\omega_k}$ , and hence the classes of potential base scales most closely approximated by the expected computed scale are those base scales showing the smallest differences among the  $\mu_{\omega_j}$ ,  $\mu_{\omega_k}$ ,  $\dots$  for the various intervals  $j$ ,  $k$ ,  $\dots$  of the scale. This important conclusion may be rephrased as: *the classes of potential base scales expected to be most closely approximated by the method of successive intervals are the classes for which the average coefficient of error (for the estimation of interval widths under assumptions of normality) is most nearly the same for all intervals of the rating scale.*

In particular, if, as implicitly assumed by previous psychometric analyses wherein the base scale remained unidentified, there exists a class of potential base scales which simultaneously normalize all distributions, then  $\mu_{\omega} = 0$  for all intervals of these scales; there is no class of base scales more closely approximated by the expected computed scale.

Thus, we see that there is no single answer to the question of the magnitude of error involved in the approximation of an unidentified base scale by the method of successive intervals; the magnitude of error is relative to that base scale for which the computed scale is considered an approximation. If we wish, however, we may define the base scale to be approximated as that scale which simultaneously equalizes the  $\mu_{\omega}$  for all intervals. A class of such scales can always be found, and further, the set of all such classes includes all base scales which simultaneously normalize all distributions over the rating scale if such scales exist. If the base scale is so defined, then from (25)

$$\mu_{\xi_{jk}} = \alpha_{jk} + (\beta_j' - \beta_k') + (\beta_j'' - \beta_k''). \quad (27)$$

Only when the measuring distributions are extraordinarily non-normal are any of the terms on the right side of (27) of expected magnitude greater than  $10^{-2}$ , and thus  $\xi_{jk}$  has an expected order of magnitude of no greater than  $10^{-2}$ . This, in conjunction with (26), shows that if the sample sizes of the distributions have been taken sufficiently large (say, large enough to make  $\sigma_{\gamma}$  on the order of  $10^{-2}$ ), then *the extent to which interval ratios computed by the method of successive intervals diverge from the corresponding theoretically "true" values should not exceed 10 per cent of the latter, and may be much smaller if the experimental study has been well designed.*

### Conclusions

Abstracting the essentials of the foregoing analysis, three major points are of significance—the first, a contribution to the computation technique of the method of successive intervals; the second, an evaluation of the validity

of the method; and the third, the significance of the method for the basic methodology of psychophysical measurement.

The contribution to computational technique is given by equations (13) and (14); it involves the computation of weights for the estimates of a given interval width so as to minimize the sampling errors for the composite estimate of the interval width. Except for the more exacting studies, however, or unless suitable tables have been obtained, the improvement of this exact method of weighing over the more rough and ready techniques now in use will scarcely be worth the extra computational labor. Of greater potential application in the design of empirical studies is the determination of the relations among width of interval, the number of measuring distributions and their sample sizes for the maintenance of a fixed level of freedom from sampling error.

The validity and reliability of the method of successive intervals do not depend upon normality of distributions or equality of their variances. The reliability, as attested by (26), may be made as high as desired. If the base scale is suitably defined (i.e., defined so as to equalize, for the various intervals, the error due to estimation of interval width from a table of the normal probability integral) and if the reliability is made sufficiently high, then the validity, as implied by (27), is so high as to lead to an expected coefficient of error for relative interval widths of no more than a few parts in a hundred. Further, this validity is in reference to the theoretical values of the interval ratios. It is thus an *absolute* validity in contrast to past validation of psychophysical scaling techniques, where validation is attempted only in terms of internal consistency or consistency among different techniques purported to compute the same base scale. It would appear, then, that until similar analyses can be constructed for other psychophysical scaling techniques, the method of successive intervals should be accepted as the basic standard against which other techniques are to be validated.

Finally, and probably most important of all, we consider the implications of this analysis for the methodology of psychophysical measurement. It has been shown that it is unnecessary for psychophysical measurement (or for that matter, for any form of measurement) to assume any specific form of distributions over a measuring scale. The only assumption required is that certain properties of the measurements obtained by the measuring technique have some potential interpretative significance. The major premise of psychometric scaling in the past has been that if (a) a scale can be obtained which normalizes the distributions over it, then (b) that scale, or another very similar to it, has interpretive significance as an interval scale. We may now replace this premise with another: if (a') a scale can be obtained which equalizes, for all intervals, the average coefficient of error for the approximation of interval width by the distance which normal distributions of equal standard deviations would span between corresponding percentiles,

then (b) that scale, or another very similar to it, has interpretive significance as an interval scale. The latter premise is both weaker and stronger than the former: weaker in that a scale satisfying (a') can always be found, and such a scale also satisfies (a) when scales satisfying (a) exist; stronger in that the latter premise demands a meaningful scale to underlie every psychophysical measuring technique, whereas the former demands such a meaningful understructure only if a psychometric scale can be found to normalize simultaneously all distributions over it. Actually, the (b) clause of these premises is not so strong as it might appear. In a certain sense, the mere act of defining a scale in terms of the distributions over it imparts a meaning to the scale values so defined. Essentially, what our present analysis has shown is that it is always possible to give a distributional definition to a base scale ~~which simultaneously normalizes all distributions regardless of whether or not a scale exists.~~ *which simultaneously normalizes all distributions.*

Since interpretation of psychometric scales has been sought in actual practice, regardless of whether simultaneous normalization could be realized, it is essential, if psychometric custom now current is to be justified, that a way be found to define psychometric scales in terms of properties other than such normalization. It is our belief that such justification has now been furnished.

## REFERENCES

1. Attneave, F. A. A method of graded dichotomies for the scaling of judgments. *Psychol. Rev.*, 1949, 56, 334-340.
2. Edwards, A. L. The scaling of stimuli by the method of successive intervals. *J. appl. Psychol.*, 1952, 36, 118-122.
3. Edwards, A. L. and Thurstone, L. L. An internal consistency check for scale values determined by the method of successive intervals. *Psychometrika*, 1952, 17, 169-180.
4. Garner, W. R. and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, 58, 446-459.
5. Guilford, J. P. The computation of psychological values from judgments in absolute categories. *J. exp. Psychol.*, 1938, 22, 32-42.
6. Guilford, J. P. *Psychometric methods*, 2nd ed. New York: McGraw-Hill, 1954 (pp. 223-262).
7. Mosier, C. I. A modification of the method of successive intervals. *Psychometrika*, 1940, 6, 101-107.
8. Saffir, M. A. A comparative study of scales constructed by three psychophysical methods. *Psychometrika*, 1937, 2, 179-198.
9. Stevens, S. S. Mathematics, measurements, and psychophysics. In S. S. Stevens (Ed.): *Handbook of experimental psychology*. New York: Wiley, 1951.
10. Thurstone, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, 34, 273-286.

*Manuscript received 3/10/54*

*Revised manuscript received 3/18/55*